Contributions to Electrical Energy Disaggregation in a Smart Home

Marisa Batalha Figueiredo

PhD Thesis submitted to fulfill the requirements of the Doctoral Program in Information Science and Technology

Department of Informatics Engineering
Faculty of Sciences and Technology
University of Coimbra
Portugal

Coimbra, September 2013

Advisor

Prof. Doutora Bernardete Martins Ribeiro

Associate Professor with Aggregation
Informatics Engineering Department
Faculty of Sciences and Technology of University of Coimbra

Co-Advisor

Prof. Doutora Ana Maria Carvalho de Almeida

Assistant Professor
Sciences and Technologies of Information Department
ISCTE - University Institute of Lisbon

Dedicated to the ones that

were always there,
are always there,
and
will always be there...

even when they are not around anymore...

Abstract

Tomorrow's main energy resource may well be energy efficiency. Over the last years, society awareness on environmental changes and high energy costs has been increasing. Nevertheless, the improper use of electrical devices still represents a substantial slice of the electrical energy consumption. Continuous and detailed electricity monitoring has been demonstrated an essential tool to ensure energy efficient in buildings as our homes. Appliance-specific consumption information empowers consumers, leading to informed choices and change of behaviours.

Non-Intrusive Load Monitoring (NILM) systems, aiming at energy monitoring, load forecasting and improved control of residential appliances, are an attractive solution to bring detailed consumption at device-level to end-users. Using only the aggregated electricity consumption data acquired at a single-point, usually the utility-customer interface, NILM discerns appliances' power usage data employing machine learning and pattern recognition algorithms. Due to its possible low cost, easy installation and easy integration into the future smart grids, which would enable consumers to participate in the electricity market, NILM has become an active area of research.

This Thesis is concerned with energy disaggregation as the key part of a NILM framework. Given the whole-home electrical consumption data it aims at investigating and exploring methodologies not yet applied to tackle the correct disaggregation of this signal into the detailed usage of each appliance, or groups of devices, connected to the home electrical circuit.

Widespread NILM approaches usually explore the disaggregation of single-point acquired data as a classification problem for which appliances signatures are required. Yet, no set of distinctive characteristics able to accurately describe each appliance has been found. Thereby, this thesis reinforces the search for the set of features used as appliances signatures. Namely, a rule for steady-state

identification and its mathematical proof are introduced. This rule was applied for detection of step-changes occurring in the active and reactive power signals and the power factor measurements. The step-changes identified comprised a new appliance signature posteriorly used by the 5-Nearest Neighbours and the Support Vector Machines classification methods in order to obtain the appliance identification. The computational experiments yielded in real-world dataset showed the effectiveness of the proposed signature for distinguishing the different loads in study.

The disaggregation and extraction of meaningful information from the aggregated electricity consumption can alternatively be interpreted in the light of signal processing analysis. In this sense, signal processing and time series analysis strategies arise as suitable tools for the extraction of information from the whole-home signal. Before aiming at the calculation of consumption estimates for each appliance, a previous study concerning the extraction of variations in the aggregated electrical signal associated with devices that work automatically without any human intervention is performed. In this context, a technique based on Wavelet Shrinkage and signal processing operations, designed to extract information from the aggregated signal considering several of its segments that can be analysed by distinct mother wavelets, is proposed.

Following this path, a novel way to look into the issue of energy disaggregation is its interpretation as a single-channel source separation problem. To this end, the performance of source modelling based on multi-way arrays (tensors) and correspondent factorization is analysed. With the proviso that a tensor composed by the data for the several devices in the house is given, non-negative tensor factorization is performed in order to extract the most relevant components. The outcome is later embedded in the test step, where only the whole-home measured consumption is available. Inference of individual consumptions is then achieved by matrix factorization using the learned models. The approaches based on signal processing, for the extraction and disaggregation of information from the whole-home electrical signal, were successfully evaluated on a real-world dataset, as illustrated by the favourable performance and statistical evidence.

Overall, this Thesis contributes with electrical energy disaggregation approaches, successfully validated on real-world data, which - as we hope - will have a positive impact in solving efficiency problems in a smart home.

Resumo

A eficiência energética pode ser o futuro em termos de recursos energéticos. Nos últimos anos, a consciencialização da sociedade relativamente às mudanças ambientais e aos elevados custos energéticos tem aumentado. Contudo, o uso desajustado dos aparelhos eléctricos ainda representa uma fatia substancial do consumo de energia. A monitorização contínua e detalhada, como já demonstrado, é essencial para assegurar a eficiência energética em edifícios como as nossas casas. O consumo detalhado por electrodoméstico é uma mais-valia para o consumidor, conduzindo-o a escolhas informadas e mudanças de comportamentos.

Sistemas de monitorização não intrusiva de cargas (NILM) possibilitam a monitorização da energia, a previsão dos consumos e o controlo dos aparelhos eléctricos residenciais, constituindo uma solução atractiva para munir o consumidor final com o gasto detalhado por equipamento. Acedendo apenas aos dados relativos ao consumo agregado de electricidade, adquiridos num único ponto (quadro eléctrico geral), estes sistemas discriminam os gastos de cada aparelho através de algoritmos de aprendizagem computacional e de reconhecimento de padrões. Considerando o baixo custo associado, a instalação fácil e o potencial associado às redes energéticas inteligentes do futuro, pois facilitará a participação dos consumidores finais no mercado da electricidade, o NILM tornou-se uma área activa de investigação.

A presente Tese foca-se na desagregação de energia, enquadrado num sistema NILM. Assentando na disponibilidade dos dados de consumo agregado, esta Tese tem por objectivo investigar e explorar metodologias ainda não aplicadas à resolução do problema de separação do referido consumo nos gastos dos diversos equipamentos, ou grupos, ligados ao circuito eléctrico da casa.

Uma abordagem comum considera a questão da desagregação como um problema de classificação, requerendo a definição de assinaturas dos vários equipamentos. Todavia, ainda não foi encontrado um conjunto de características

distintivas apto a descrever com precisão cada aparelho. Assim sendo, esta Tese procura reforçar a procura de um tal conjunto de características. Para tal, é introduzida uma regra para a identificação de estados estáveis, bem como a sua demonstração matemática, utilizada para o reconhecimento de *step-changes* nos sinais de potência activa e reactiva e factor potência. As *step-changes* identificadas definem as assinaturas dos equipamentos posteriormente utilizadas por dois métodos de classificação para o reconhecimento dos electrodomésticos, a saber: 5-Vizinhos Mais Próximos e Máquinas de Vectores de Suporte. As experiências computacionais, baseadas em dados reais, mostraram a eficácia da assinatura proposta para a distinção das diferentes cargas em estudo.

A desagregação e extracção de informação relevante dos dados agregados de consumo eléctrico podem ser interpretadas à luz da área de processamento de sinal. Neste sentido, as abordagens de análise de sinal e de séries temporais adequam-se à extracção de informação do consumo agregado. Com efeito, previamente ao cálculo das estimativas dos gastos de cada um dos aparelhos, o estudo foca-se na extracção de variações contidas no sinal agregado e associadas a equipamentos cuja operação não requer qualquer tipo de interacção humana. Neste contexto, é proposto um método que tem por base a técnica de *Wavelet Shrinkage* em conjunto com operações de processamento de sinais para a extracção de informação do sinal agregado. Note-se que o referido método proposto assenta no pressuposto de que vários segmentos podem ser analisados por funções *wavelet* distintas.

No mesmo contexto, a desagregação de energia pode ser interpretada como um problema de separação de fontes quando apenas um sinal de mistura é conhecido. Assim sendo, foi analisado o desempenho da modelação de fontes com recurso a vectores multidimensionais e correspondente método de factorização. Assumindo que um vector multidimensional composto pelos dados de gastos dos vários equipamentos da casa pode ser definido, a sua factorização não-negativa é executada de forma a extrair os componentes mais relevantes. Os factores resultantes são incorporados no processo de inferência das fontes, no qual apenas o consumo agregado da casa está disponível. As estimativas dos consumos associados a cada equipamento são, então, obtidas pela factorização não-negativa de matrizes. Tanto a abordagem para a extracção de variações como o método proposto para a desagregação de sinal foram avaliados com sucesso num conjunto de dados real, para o qual um desempenho favorável foi observado e validado através de análise estatística.

Em suma, esta Tese contribui com abordagens para a desagregação de energia eléctrica, testadas e validadas em dados reais, pelo que esperamos que venha a ter um impacto tangível na resolução de problemas de eficiência numa casa inteligente.

Acknowledgments

Every human being at any given instant of time is the outcome of a 'mixing process' of several 'sources'. Some related to his/her past experiences, others related to his/her surroundings. During this path, I have been surrounded by amazing people who, in one way or another, helped me to reach this point and who are a valuable input for my 'mixing process'. In some cases, their contribution was beyond the scientific point of view and indeed helped my personal and professional growth. After all these years what started as a purely scientific guidance somehow ended up in friendship.

This adventure would not have started and it would not have come this far without my advisors Prof. Ana Maria de Almeida and Prof. Bernardete Ribeiro. I would like to express them my deep gratitude. Prof. Ana Maria, for bringing me to this adventure, for all the trust and for all the advices along these years, thank you! Prof. Bernardete, for joining me in this journey, for the precious guidance in all its possible dimensions, for all the energy, 'pushes' and all 'the slaps on the wrist' which led me to these pages, thank you!

Coming this far would not be possible without my 'primary sources', my parents and sister, who believe in me, support all my options and put up with all my bad moments. Thank you!

Last but not least, my 'mixing process' is never completed without friends: the ones from a lifetime and the ones that joined me along the way. You know who you are and I know you will be there! For the friendship in good and bad moments, thank you!

I would like to express special thanks to all the AmILab colleagues for the fellowship during these years, to ISA-Intelligent Sensing Anywhere, S.A. and to Joana Abreu for the support provided in specific parts of this adventure.

Finally, the research that led to this Thesis would not have been possible without the funding and support provided by the iTeam project, the Fundação para a Ciência e Tecnologia under the scholarship SFRH / BD / 68353 / 2010 and by the Centre for Informatics and Systems of University of Coimbra (CISUC).

Marisa Batalha Figueiredo
Coimbra, September 2013

Contents

List of Figures

List of Tables

List of Algorithms

List of Abbreviations

AAL Ambient Assisted Living 6

AC Alternating Current 15

ALS Alternating Least Squares xix

ANN Artificial Neural Network 44

ANNs Artificial Neural Networks 30

BLUED Building-Level fUlly-labeled dataset for Electricity Disaggregation 23

CANDECOMP Canonical Decomposition 125

CART Classification and Regression Trees 40

CFHMM Conditional Factorial Hidden Markov Model 47

CWT Continuous Wavelet Transform 86

DDSC Discriminative Disaggregation Sparse Coding xv

DWT Discrete Wavelet Transform xiv

ECG Electrocardiogram 82

EEG Electroencephalography 82

EMI Electromagnetic Interference 33

EU European Union 1

EWD Embedding, Wavelet Shrinkage and Diagonal Averaging xiv

EWD-2 Embedding, Wavelet Shrinkage and Diagonal Averaging performed with 2 levels of decomposition 101

EWD-3 Embedding, Wavelet Shrinkage and Diagonal Averaging performed with 3 levels of decomposition 102

FFT Fast Fourier Transform 30
FHMM Factorial Hidden Markov Model xiv
FN False Negative 51
FP False Positive 51
FPR False Positive Rate 51
FT Fourier Transform 30
GMM Gaussian Mixture Model 44
GPU Graphics Processing Units 45
HDP-HSMM Hierarchical Dirichlet Process Hidden Semi-Markov Model . . 48
HMM Hidden Markov Model 45
HMMs Hidden Markov Models 45
ICA Independent Component Analysis 119
K-NN K-Nearest Neighbours 7
LIFO Last-In-First-Out 28
MDL Minimum Description Length 94
MIT Massachusetts Institute of Technology 14
MRA Multiresolution Analysis 87
NALM Nonintrusive Appliance Load Monitoring 26
NILM Non-Intrusive Load Monitoring i
NMF Non-Negative Matrix Factorization xix
NNSC Non-Negative Sparse Coding 122
NTF Nonnegative Tensor Factorization 127
PARAFAC Parallel Factor Analysis xix
PCA Principal Component Analysis 82
RBF Radial Basis Function 30
REDD Reference Energy Disaggregation Dataset xvii
RES Renewable Energy Sources 1
RMSE Root-Mean-Square-Error xv

ROC Receiver Operating Characteristic 51

SMPS Switch Mode Power Supplies 33

SNR Signal to Noise Ratio xv

SSA Singular Spectrum Analysis 7

STFT Short-Time Fourier Transform 35

STMF Source Separation via Tensor and Matrix Factorization xv

SV Support Vectors 43

SVD Singular Value Decomposition 82

SVM Support Vector Machine xvii

SVMs Support Vector Machines 7

TP True Positive 51

TPR True Positive Rate 51

UK United Kingdom III

WDT Wavelet Decomposition 80

WDT-2 Wavelet Shrinkage performed with 2 levels of decomposition 101

WDT-3 Wavelet Shrinkage performed with 3 levels of decomposition 102

WT Wavelet Transform 31

5-NN 5-Nearest Neighbours xvii

Nomenclature

A_i, B_i	Outcome of the X_i factorization
$\hat{A}, A^*$	Activations associated with the aggregated signal
A, B, C	Outcome of the PARAFAC factorization
Ap_j	Approximations at level j
$\tilde{B}$	Disaggregation bases
C	Regularization parameter in SVMs
D	Dimension of the feature space
$\mathcal{D}$	Denoising operator
D_i	Diagonal matrix associated with the i-th row of C
Dt_j	Details at level j
$\mathcal{DA}(\cdot)$	Diagonal averaging operators
$\mathcal{E}(\cdot)$	Embedding operator
F_β, F_1	F-measure
H	Kruskal-Wallis test statistic
$\bar{H}$	Hankel matrix
$\mathcal{H}$	Hankelization operation
I	Current
I_n	Number of disjoint groups
J	Levels of decomposition
K	Number of neighbours
L_w	Embedding dimension (window length)
$\tilde{L}$	Number of directions
N	Number of dimensions on a tensor
$\bar{N}$	Number of samples
$\check{N}$	Number of lagged vectors
$\tilde{N}$	Total sample size

P	Active power measured in Watts
Q	Reactive power measured in Volt-ampere Reactive
R	Number of components for the PARAFAC factorization
R_i	Sum of ranks of group i
S	Apparent power measured in Volt-ampere
$\bar{S}$	Covariance matrix
S^*	Steady-state with start at t_a and composed by t^* samples
$SE_{\bar{U}}$	Standard error of the U statistic
T	Length of a signal
$\tilde{T}$	Toeplitz matrix
U	Mann-Whitney test statistic
$\bar{U}$	Mean of the U statistic
U'	Eigenvectors matrix of $\bar{S}$
V	Voltage
V'	Eigenvectors matrix of $\check{X}^T\check{X}$
$W, H, \tilde{W}, \tilde{H}$	Outcome of matrix factorization
$\mathcal{W}, \mathcal{W}^{-1}$	Forward and inverse wavelet transform operator
$\underline{X}$	3-order tensor
X_i, X_i'	Consumption matrix of appliance i
$\check{X}$	Trajectory matrix
$\bar{X}, \bar{X}'$	Aggregated consumption matrix
$\tilde{X}_i, \tilde{X}_i'$	Estimated consumption of appliance i
$\hat{\bar{X}}$	Estimated aggregated consumption matrix
Y	Matrix of shrunk lagged vectors
$Y_{i_1,:,:}$	Horizontal slices of a 3-order tensor
$Y_{:,i_2,:}$,	Lateral slices of a 3-order tensor
$Y_{:,:,i_3}, Y_{i_3}$	Frontal slices of a 3-order tensor
$Y_{(i)}$	Mode-i unfolding matrix
$\tilde{Y}$	Output of the Hankelization operation
$\underline{Y}, \underline{\check{Y}}$	Tensor
a	Dilatation
a_0	Fixed dilatation step
b	Translation along the time axis
$\tilde{b}$	Bias
$\bar{c}$	Class
c, d	Wavelet coefficients

$e_i, \tilde{v}_i$	Signal extracted
h	High-pass filter coefficients
k	Number of devices or circuits
$\tilde{k}$	Number of groups in the statistical analysis
l	Low-pass filter coefficients
m	Number of signals available of training set
m'	Number of signals available of the test set
n_i	Sample size of group i
p_f	Filter length
q	Set of k state sequences z^i
r	Number of bases used by the sparse coding
$\tilde{r}$	Synthetic generated noise
$s_{x,y}$	Cross-covariance between signals x and y
$s_{x,x}, s_{y,y}$	Autocovariance
s_{min}	Minimum number of consecutive samples in a stable-state
t_n	Appliance label associated with the v_n (target)
th_{ws}, th_{ssa}	Threshold
v	Vector of features associated with a new load
v_n	n-th vector of features
v_i	Vector of features associated with load i
w	Weight vector that defines a model
x	Measured signal (sequence of observations)
x_i	Electrical consumption signal for device or circuit i
x_M	Maximum value within S^*
x_m	Minimum value within S^*
x'_1	Aggregated signal without variations of 'automatic' equipment
x'_2	Variations associated with 'automatic' equipment
$\bar{x}$	Whole-home electrical consumption (aggregated) signal
$\hat{x}_i$	Predicted electrical consumption signal for device or circuit i
$\check{x}_{\check{n}}$	Lagged vector of length L_w
$y, \hat{y}$	Output of a model
$\tilde{y}$	Recovered signal
z^i	Sequence of operational states associated with device i
Σ	Diagonal matrix of singular values $\tilde{\lambda}$
Λ_t, Λ^*	Rectangular areas
α	Lagrange multiplier in SVMs optimization

$\tilde{\alpha}$	Stepsize parameter
$\alpha_A, \alpha_B, \alpha_C$	Nonnegative regularization parameters
γ, σ	Free parameter within the RBF kernel
$\epsilon, \epsilon_{active}$	Tolerance value
$\epsilon_{reactive}, \epsilon_{pfactor}$	Tolerance value
$\kappa(\cdot), \bar{\phi}(\cdot)$	Kernel functions
λ	Sparseness degree
$\tilde{\lambda}$	Singular values of S
$\check{\lambda}$	Vector of weights
μ_W, μ_H	Gradient stepsizes
ξ	Positive slack variables in SVMs optimization
$\rho_{x,y}$	Cross-correlation between signals x and y
σ^2	Variance
$\tau_0, \tilde{n}$	Translation factors
$\tilde{\phi}$	Phase angle
ϕ, φ	Wavelet functions
χ	Typical rank of a tensor
$.^T$	Transpose
$\|\cdot\|_F$	Frobenius norm
$\|\cdot\|$	L_2 - norm
$\circledast$	Hadamard product (elementwise multiplication)
$\oslash$	Elementwise division
$\odot$	Khatri-Rao product
$\otimes$	Kronecker product

CHAPTER 1

Introduction

"If you can't measure it, you can't manage it"
- Kaplan and Norton (1996)-

1.1. Motivation

Today, energy efficiency – the ability to maintain the productivity while reducing the energy consumption – is a major society concern. Saving energy through behavioural changes is highly desirable: not only it is financially profitable but also ambient friendly, due to the reduction of pollutants emissions. Moreover, saving energy reduces the countries' dependency on energy imports, which is a critical vulnerability for any country, since it exposes them to oil price fluctuations and energy supply interruptions.

Over the last decade, the European Union (EU) energy dependence has increased significantly. Overall, the energy imports were increased by 6% (from 46.7% in 2000 to 52.7% in 2010 (Eurostat, 2012a)). To overcome this problem, the EU energy policy has focused on the use of Renewable Energy Sources (RES) to

replace part of the fuel imports. These measures decrease the carbon emissions and dissociate the energy prices from the oil prices which in 2008 have reached their highest peak. From beginning 2007 till May 2008, the Euro price per barrel of Brent crude oil increased 90% (European Commission, 2008). Another essential part of the EU energy policy is the promotion of energy efficiency at the end-use sector. This policy is seen as the "Europe's biggest energy resource" (European Commission, 2011). In fact, the EU aims at increasing it by 20% by 2020, which is equivalent to cut the primary energy consumption by 20% (Eurostat, 2012b).

In order to achieve the defined goals, an Energy Efficiency Plan focusing on the main sectors of final energy consumption (transport, residential and industry) was developed. According to Eurostat (2012a), these three sectors represented in 2010 84% of the final energy consumption of the 27 EU countries: transport sector consumed 32% of final energy, followed by the residential sector with 27% and by the industrial sector with 25%. While the energy for the transportation sector is mainly composed of fossil-based fuels, the residential and industrial sectors have a diversified set of energy sources: oil products, gases (*e.g.*, natural gas), electrical energy and renewable energy. As illustrated in Figure 1.1, the electrical energy consumption is substantial, both in residential and industrial sectors. In fact, it represents 66.2% of the total electricity consumption in the EU. Nevertheless, from 2000 to 2010, the electricity consumption decreased slightly in the industrial sector (-2%) in contrast to the household sector where its consumption increased by 18% (Eurostat, 2012a). Consequently, to achieve the 20% cut in the primary energy consumption, the efforts for improving the energy efficiency must consider the usage of this source, in particular concerning the household sector.

In this context, consumers' awareness about their electrical consumption plays an important role. Despite the fact that appliances are becoming more efficient, precise and up to date detailed information about consumption would encourage a better usage.

During the last five years, the market has been offering an increasing number of devices, so-called smart meters, which are able to measure electricity consumption of an appliance or an electrical system in real time, usually providing some kind of interface for the end-user. The device usually communicates that information, via network, back to the local utility for monitoring and billing purposes (telemetering) or displays it on a monitor directly to the user.

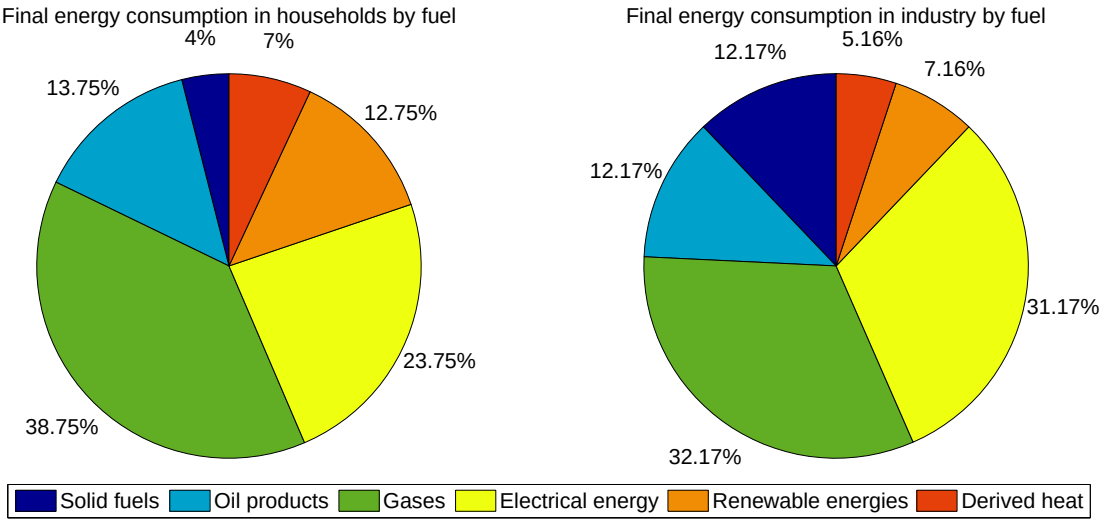

Figure 1.1.: Final energy consumption in residential (at left) and industrial (at right) sectors by fuel according to data supplied by Eurostat (2012a) for 2010.

Although smart meters are helpful displaying the overall consumption and leading to a 15% reduction (Darby, 2006), they do not yet show the individual consumption associated with the appliances. Instead, they provide an aggregate measurement. That is, when connected to the electrical system, they can not specify which devices are switched on or off. At household level, feedback involving appliance-specific information would help to capture the consumer's attention and to change particular behaviours. As illustrated by Fischer (2008); Karjalainen (2011); Sundramoorthy et al. (2011), this would lead to the desired energy efficiency.

1.2. Challenges and Research Questions

The need for gaining understanding of the characteristics of appliance electrical consumption not only has potential to drive the choices made by individuals in the privacy of their homes but also would be useful for utilities and retailers to tailor their service offers. As suggested by Abreu and Pereira (2012), it would help to improve the general efficiency of the electrical energy system.

Detailed electrical consumption at device-level could be obtained as the outcome of an intrusive 'sub-metering' solution although that would require

several individual meters spread out by the different appliances. Despite remote access to information and control inputs may be obtained easily and inexpensively via networking, access does not provide useful information without installation of a potentially expensive and intrusive sensor array. As a consequence, a less invasive and more practical monitoring system arises as a requisite for an efficient framework to bring detailed information to the end-users. A viable solution for this purpose is the Non-Intrusive Load Monitoring (NILM) framework.

Non-Intrusive Load Monitoring (NILM) can determine the operating schedule of electrical loads in a target system from measurements made at a centralized location, such as the electric utility service entry. In contrast to other systems, NILM reduces sensor costs by using relatively few sensors. The current advances in the related areas as sensing technology, data communication, machine learning and pattern recognition have enhanced the exploration and investigation of methodologies to tackle the electrical consumption disaggregation from a single sensing device (Zoha et al., 2012). Accordingly, the associated hardware complexity would be almost irrelevant, since only one sensor is needed to gather all the data in the electrical network. As a result, one of the main challenges addresses how to treat this problem, either by classification or source separation techniques aiming at the loads identification.

Over the last years, single-point sensing associated research has had significant advances. Still, several challenges must be addressed to provide NILM as a viable electrical consumption disaggregation method (Froehlich et al., 2011). Another main challenge concerns the correct disaggregation of appliances, given the whole-home data, when more than two are switched on in the same electrical circuit with very similar electrical signatures. In fact, the energy disaggregation problem, part of the NILM framework, can be seen as a pattern recognition problem. In this widely studied approach, inspired by the work of Hart (1992), distinctive characteristics for each appliance forming the appliance signature are defined and identified. Nevertheless, after two decades of research and exploration mainly focused on the appliance signature, a robust set of features able to effectively describe the appliances regardless their characteristics (*e.g.* manufacturer) have not yet been defined (Zeifman and Roth, 2011; Zoha et al., 2012). Notice that most of the multi-state appliances, *e.g.* washing machines, have particular user settings and their power consumption is not static. In addition, low power appliances have similar power usage

and similar electrical features enhancing the difficulty of their disaggregation. Moreover, depending on the features selected to describe the appliance, even the house electrical network can influence the captured signature (Patel et al., 2007; Gupta et al., 2010). The lack of a reference dataset, until recently, has also limited the comparisons and evaluations of the several set of features, along with the absence of consensual and standard performance evaluation metrics (Zoha et al., 2012).

In this context, and given that a definitive answer was not yet found for an accurate energy disaggregation as part of a NILM system, it remains a relevant problem to be investigated. As mentioned before, no robust set of features is able to describe accurately each appliance (Zeifman and Roth, 2011; Zoha et al., 2012). Therefore, the analysis of other electrical features would reinforce the search for this set of distinctive characteristics used as appliance signatures when the energy disaggregation is seen as a classification problem. Nonetheless, other paths, rather than the classification approaches, could be explored (Kolter et al., 2010).

The single-point sensed data corresponding to the whole-home electrical consumption is the signal from which information associated with the consumption of each device or groups of appliances should be extracted. In this sense, alternative techniques, namely signal processing and time series analysis, arise as potential 'tools' for the disaggregation and extraction of meaningful components from the original signal. If we group the devices that can be found in our homes into two distinct sets accordingly to their operation mode (manual or automatic), these techniques would be able to separate information associated with automatic appliances from the manual ones.

Following this path, the energy disaggregation can be formulated as a single-channel source separation problem (Kolter et al., 2010). This aims at the estimation of individual sources from a single observed mixed signal, which in this case is the aggregated electrical consumption. In the same direction of research, approaches based on data-adaptive representations, usually employed for solving these source separation problems by source modelling, would be an alternative to the usual classification methods for the energy disaggregation. These approaches learn a model for each source, which are used for the posterior separation of the mixture signal. For energy disaggregation, an intermediate step to adjust independent source models considering the aggregated signal may be required (Kolter et al., 2010). Therefore, other data representations,

namely, multi-dimensional arrays and associated factorization methods could be an alternative to explore.

Within the NILM framework both at an initial stage, gathering data for training algorithms in a supervised approach, and at a posterior phase, when a new equipment is added to the household electrical network (Froehlich et al., 2011), the dataset of signatures is also an issue. In a supervised approach, a dataset of labelled signatures is required for further training of the classification/disaggregation methods. An optimal solution, yet impractical, would consider a dataset composed by signatures of all possible appliances of every single manufacturer. Nevertheless, and as mentioned, if the signatures are network dependent, such training database must be built for each house (Froehlich et al., 2011). Still, if the signatures are invariant from house to house, it would be possible to develop a collaborative training dataset. Another challenge regards the signature dataset update when a new load is added to the electrical network. A possible approach would be an interactive system that in the presence of unknown loads, in an unsupervised learning mode, would label the new appliance, prompting the user for validation, and next update the dataset, as suggested by Froehlich et al. (2011).

Last but not the least, note that, once the electrical consumption has been disaggregated, information can be extracted and used for several purposes. Examples are consumer detailed profiles for utilities (Abreu and Pereira, 2012), and home-activity monitoring either for children or elderly persons within an Ambient Assisted Living (AAL) framework. Despite all associated benefits, this leads to the question of user privacy and the question of ownership with regard to power consumption information since it could be used for surveillance purposes as discussed by Hart (1989).

1.3. Thesis Contributions

This Thesis contributes with a partial answer to the challenges addressed in the previous section. Specifically, it aims at energy consumption disaggregation, investigating and exploring methodologies not yet applied to solve the problem and proposing new approaches for the separation of information from the whole-home electrical usage signal. In more detail, in this Thesis:

- a new appliance signature composed by the step changes in active, reactive power signals and the power factor measurements was explored. In

addition, a rule is provided for the extraction of step changes in the electrical signals, and we prove a theorem with a deductive argument for our mathematical statement. This technique was applied for the detection of step changes in active, reactive powers and power factor measurements, which were then used to define appliance signatures. The steady-state signatures were used to build device identification models, using the learning methods K-Nearest Neighbours (K-NN) and Support Vector Machines (SVMs) after considering both clean input data as well as measurements with associated noise (Figueiredo et al., 2010, 2011a, 2012a) (see Chapter 3 Section 3.2).

- two datasets composed by measurements of active power, voltage, current and power factor signals for a set of selected appliances were gathered using a sensing meter prototype. In particular, data was collected for only one appliance in the circuit (Dataset A) and for a pair of devices plugged into the electrical network (Dataset B). The former dataset is composed by information of six classes, a microwave, a coffee machine, a toaster, an incandescent lamp and two LCD's (different models of the same manufacturer); while for the latter two distinct pairs of appliances were considered: an incandescent lamp and one of the LCD's, a microwave and a toaster (Figueiredo et al., 2012a) (see Chapter 3 Section 3.3).

- a new strategy for the extraction of variations in the aggregated consumption signal, associated with devices that work automatically without any human intervention, based on wavelet decomposition, was investigated. This technique, based on Wavelet Shrinkage, allows for the extraction of information from the aggregated signal considering several of its segments, which can be analysed by possible distinct mother wavelets. In addition, the suitability of the signal processing tool wavelet transform (and associated mother wavelets) and the time series method Singular Spectrum Analysis (SSA) was investigated for the above goal. Moreover, these approaches were successfully evaluated on a real-world dataset (Figueiredo et al., 2011d,b,c) (see Chapter 4 Sections 4.4, 4.5).

- a new approach for solving the energy disaggregation problem, formulated as a single-channel source separation problem, and based on the multi-dimensional representation of sources was studied and investigated. This strategy, based on the non-negative factorization of multi-way arrays

(tensors) and on non-negative matrix factorization, learns bases for each source (appliances) such that the error between these sources and their estimation is minimum. The learning process is 'global', *i.e.* the model is learned for every source at the same time and intends to gather relevant information across the three domains in study: time of the day, day of the week, and home devices (Figueiredo et al., 2012b, 2013b) (see Chapter 5 Section 5.4).

- in a novel study we empirically investigated the need of imposing a sparsity degree to the source models of appliances in order to solve the energy disaggregation problem in a sparse coding based technique. This methodology considers that the consumption of each appliance is independently modelled imposing a certain degree of sparsity for each learned model. The validation of our findings was successfully tested on a real-world dataset for energy disaggregation (Figueiredo et al., 2013a) (see Chapter 5 Section 5.5).

- a new strategy is proposed for solving the energy disaggregation problem based on multi-dimensional arrays and their non-negative factorization using in the three dimensions (day time, day of the week , and devices). The validity of our model has been successfully evaluated on a real-world dataset for energy disaggregation. The proposed model improves the performance with respect to the sparse coding based method (Figueiredo et al., 2013b) (see Chapter 5 Section 5.5).

1.4. Thesis Outline

The present Thesis is organized into six chapters.

Following the general introduction in this chapter, in **Chapter 2** a review of the state-of-the-art on NILM research is presented. The chapter begins with a brief introduction of notions revisiting concepts on the electrical power in alternating current and its particularities, necessary for the description of specific details within the NILM framework. In the sequel, the formal description of the energy disaggregation problem and the main steps of a NILM system (data acquisition, signal analysis and feature extraction, and model learning and classification) are presented. Next, each NILM main step is revised. The chapter

addresses the data acquisition process, focusing on the relation between the sampling rate and the appliance signatures. The background on the electrical signatures of appliances is provided, in particular steady-state and transient signatures, which play a crucial role for an accurate appliance identification. The former can be deduced from the difference between two steady-states in a signal while the latter are composed by features extracted from the transitory period between two steady-states. The research on these different types of signatures is revised as well as the alternative signatures found on the related literature. The model learning and classification process, reviewing both supervised and unsupervised approaches found on other scientific works is also presented. In fact, the NILM research has been focused on supervised strategies rather than on unsupervised methods, which eliminates the need of using labelled data for building the learning models as in former approaches. The overview over the research on NILM systems ends with a discussion on the main disaggregation metrics usually applied to assess the performance of the research methodologies.

Chapter 3 deals with the exploration of steady-state appliance signatures within a NILM framework. The 'classic' NILM approach regards the energy disaggregation as a classification problem, requiring then appliances' signatures in order to perform an accurate identification of several devices. Among the possible alternatives, the steady-states signatures have become widely explored due to its simplicity and undemanding requirements, in terms of sampling rate and specific hardware for the data collection. This chapter begins by describing the requirements that define a steady part of a signal. A rule for steady-state detection is introduced together with a Theorem and its proof, which accounts for the veracity of our mathematical argument. The developed approach was used for the identification of above mentioned stable signal parts. Next, the steady-states occurring in the active and reactive power signals and the power factor measurements are proposed as distinctive features composing then an appliance signature and the K-NN and SVM classification methods are applied for appliance identification. The chapter presents two different experiments which were drawn to assess the performance of this signature and classification methods: the first considers only one device on the electrical circuit being switched on, and the second considers that the previous devices were connected simultaneously to the network. For the latter, as the data was gathered in a controlled environment, Gaussian white noise was added to collected signals.

The data was collected for only one appliance in the circuit (Dataset A) and for a pair of devices plugged into the electrical network (Dataset B). The former dataset is composed by information of six classes, a microwave, a coffee machine, a toaster, an incandescent lamp and two LCD's (different models of the same manufacturer); while for the latter set two distinct pairs of appliances were considered: an incandescent lamp and one of the LCD's, a microwave and a toaster. Finally, the chapter ends by presenting the computational experiments and discussing the results. A summary of the main aspects tackled in the chapter is presented with the main findings and conclusions.

Chapter 4 addresses an exploratory study on extraction of variations from the aggregated power signal consumption, before aiming at the disaggregation of the signal. In this chapter, the home appliances are categorized into main groups: (i) equipment that must be turned on and off with manual intervention and (ii) appliances as refrigerators, that, once that they are on, work automatically without any human interference. The latter group will be denominated as 'automatic' appliances. This chapter starts introducing the problem statement: the extraction of variations associated with the consumption of a set of 'automatic' appliances from the whole-home electricity measurements. Next, signal processing operations for multivariable signals that are required by the explored techniques for the problem at hand, are presented, in particular the embedding that maps one-dimensional signals into multi-dimensional data and its reverse operation. Additionally, the explored methods are described namely the wavelet transform and associated Wavelet Shrinkage method, the Singular Spectrum Analysis method and an approach base on Wavelet Shrinkage, in which the signal is analysed considering several segments. In the sequel, the chapter describes the two different experiments drawn to evaluate the performance of the studied and proposed approaches for the extraction of information from signals. The first considers synthetic generated data and the removal of added noise and the second comprises real-world electrical consumption signals and the extraction of variations associated with 'automatic' appliances from the aggregated measurements. The computational experiments are presented and discussed, namely the analysis concerning the most appropriate set of mother wavelets for the extraction of such variations. Finally, the main findings are summarized and directions of future work are pointed out.

Chapter 5 describes the energy disaggregation problem as a single-channel source separation problem and explores data-adaptive representations approaches for its resolution. This chapter initiates with the description of the problem at study (energy disaggregation) in the light of the single-channel source separation followed by an overview over the usual approaches for solving this type of problem, source modelling approaches in particular. In the sequel, matrix factorization is reviewed, namely the non-negative and sparse approaches as the non-negative sparse coding and the non-negative matrix factorization with sparseness conditions. Additionally, multi-dimensional arrays (also known as tensors), its related background and its non-negative factorization, required for the energy disaggregation algorithms, are introduced. Following in this chapter, the explored energy disaggregation algorithms are presented. A strategy based on the sparse coding in which each source (the consumption of each appliance) is modelled independently and on imposing a degree of sparsity for each learned model is described. Moreover, an approach based multi-way arrays representation and its associated non-negative factorization is also addressed and proposed for solving the energy disaggregation problem, such that the several sources are modelled in order to capture the relevant components in the three domains in study (time of the day, day of the week, devices), forming a source model used to infer the energy usage of appliances when a new aggregated signal is provided. Next, this chapter presents extensive experiments yielded on a reference energy disaggregation dataset, regarding the study of the sparsity condition for the former method and concerning the performance evaluation of the proposed approach based on multi-dimensional source representation. This chapter ends summarizing the main conclusions and presenting directions of future work.

In **Chapter 6**, the research outcomes are assessed bearing in mind the goals established in the first chapter. The main conclusions of each chapter are reviewed leading to possible directions of practical implementations of the ideas presented in the Thesis. Additionally, lines of future work regarding the scientific domain worthwhile of investigating for the disaggregation of electrical consumption are also addressed.

CHAPTER 2

Non-Intrusive Load Monitoring Overview

2.1. Introduction

Society awareness on environmental changes and high energy costs has been increasing. Nowadays, energy consumption reduction is a constant concern. In the residential sector, apart from the environmental issues, consumers are

also interested in financial savings associated with cuts in energy consumption which are dependent on a change of habits. In this sector, one of the main sources of energy is electricity (Eurostat, 2012a). Nevertheless, electrical savings may be difficult to achieve due to the lack of available information. Nowadays, consumers do not have easily a reference of their appliances power consumption (electrical machines used to accomplish household functions) or even of their overall consumption with exception of the information provided by the monthly power bill. Studies have shown that real-time power consumption information, at the aggregate level, would empower consumers, leading to savings of up to 15% (Darby, 2006). Detailed appliance consumption would enable users to change behaviours, using the electrical appliances efficiently, and save more (Fischer, 2008; Karjalainen, 2011; Sundramoorthy et al., 2011).

A possible solution to provide consumers with power consumption feedback of each appliance plugged into a particular network is the use of a so-called Non-Intrusive Load Monitoring (NILM) system. NILM describes a set of techniques to compute estimates of the device electrical consumptions in a particular network, based on single-point measurements of current and/or voltage or aggregated consumption from a building's electrical network. Since the use of invasive measurement equipment is avoided, this is considered to be a non-intrusive monitoring solution.

The rationale behind the concept was proposed in the eighties by Hart, Kern and Schweppe and was developed at Massachusetts Institute of Technology (MIT) (Hart, 1989, 1992), sponsored by Electric Power Research Institute. This work was succeeded in the patent number US4858141 approved in 1989 (Hart et al., 1989). During the same period, a similar idea has been investigated at Electricité de France by Sultanem (Sultanem, 1991). With the increasing environmental and economic issues, the interest in NILM architectures has been growing, particularly in view of the progresses in the related areas namely sensing technology, data communication, machine learning and pattern recognition methods. Efforts of researchers and companies are focused on the development of a suitable solution for non-intrusively monitor appliances loads (Zeifman and Roth, 2011). An example of the required and the outcome information in a NILM system is illustrated in Figure 2.1 where the aggregate consumption (input) is depicted by the graphic at the top and the signals from the several devices in the network (output) at the bottom.

This chapter presents a review of the main concepts and techniques associated

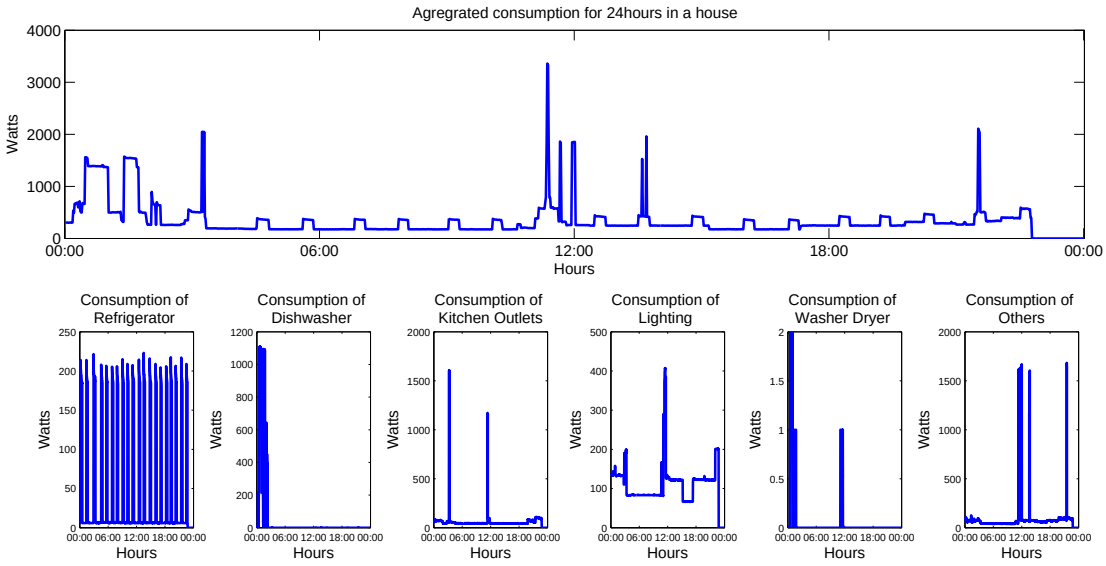

Figure 2.1.: The home aggregated (and device) consumption signals during 24hours.

with NILM. First, the necessary notions and elementary concepts of electrical power in alternating current circuits are briefly introduced. Next, NILM is properly described and its main modules and related work are reviewed. In particular, this chapter addresses: the data acquisition process; the possible electrical signatures of appliances that are seen as a key element for an accurate appliance identification (steady-state and transient signatures); and the model learning and classification process for load identification, reviewing both supervised and unsupervised approaches found in the related literature. Finally, the disaggregation metrics usually employed to assess the performance of the NILM approaches are also reviewed.

2.2. Electric Power in Alternating Current

The electric network built in our houses consists of an Alternating Current (AC) circuit. This section intends to provide the essential definitions necessary to understand some key elements of NILM and electrical energy. For further reading and details the reader is directed to (El-Sharkawi, 2012).

In an AC circuit the current does not maintain the same polarity (positive or negative) along the time. On the contrary, it oscillates from positive to negative at a given frequency *e.g.* 50Hz in Europe, which means that its polarity

changes 50 times *per* second. This oscillatory behaviour can be described by a sinusoidal function with three parameters: amplitude, frequency, and phase. The amplitude gives the distance from the maximum value to the neutral position (the x axis). The frequency, as referred, indicates the number of oscillatory movements *per* unit time and the phase ($\tilde{\phi}$) designates the starting point of the sinusoid that graphically can be identified as the shift of the entire waveform.

By definition, the power (P) is the amount of energy (consumed or produced) by unit of time, measured in Watts (W) and calculated as the product of the voltage (V) by the current (I). In fact, regarding the definitions of current and voltage,

$$P = IV = \frac{Charge}{Time} \times \frac{Energy}{Charge} = \frac{Energy}{Time}. \tag{2.1}$$

In AC circuits, the current and voltage change over time therefore the definition of power is not straightforward. In this case, power, current and voltage are complex quantities and

$$S = I^* V \tag{2.2}$$

is the complex power or apparent power (S), where * denotes the complex conjugate of the current I.

The previous definition of power as the amount of energy *per* unit of time can only be interpreted as instantaneous power, *i.e.*, at a given instant t the power is the product of the instantaneous current by the instantaneous voltage:

$$P(t) = I(t) \cdot V(t). \tag{2.3}$$

Still, for defining the average power over the cycles of the alternating current and voltage, a more suitable equation is:

$$\bar{P} = I_{rms} V_{rms} \cos(\tilde{\phi}) \tag{2.4}$$

where $\bar{P}$ is the average power, I_{rms} and V_{rms} are, respectively, the root mean square of the current and voltage waveforms, and $cos(\tilde{\phi})$ is the power factor for a phase angle $\tilde{\phi}$ that represents the phase shift between the current and voltage waveforms as illustrated in Figure 2.2. This definition holds for any type of load either resistive or reactive. For the former, current and voltage are in phase (*i.e.*, $\tilde{\phi} = 0^\circ$) while for the latter current and voltage are 90° out of phase (*i.e.*,

$\tilde{\phi} = 90°$). Note that when $\tilde{\phi} = 90°$ then no energy is dissipated since $\bar{P} = 0$.

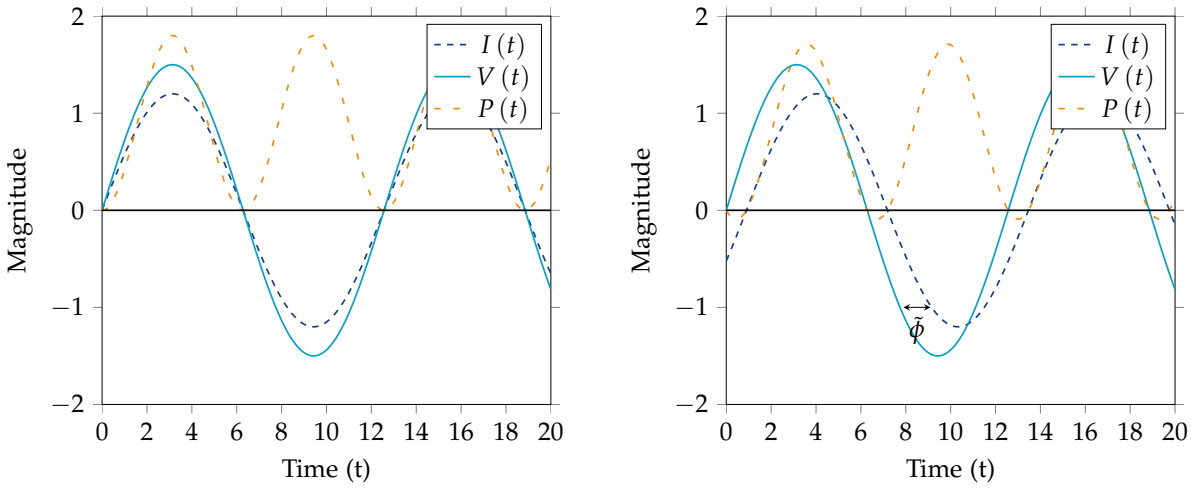

Figure 2.2.: Illustration of instantaneous power (P) as a product of current (I) and voltage (V) in phase (a) and out of phase (b).

The average power $\bar{P}$, also known as real power, active power or true power and hereafter denoted as P, measured in Watts (W), corresponds to the power being actually consumed (or transmitted). The power regardless the shift phase between current and voltage is defined as apparent power, and measured in volt-amperes (VA), whose magnitude is

$$S = I_{rms} V_{rms} \tag{2.5}$$

where I_{rms} and V_{rms} correspond to the root mean square of current and voltage values, respectively. The oscillating energy in the electrical network that does not get dissipated is called reactive power. Its magnitude is

$$Q = I_{rms} V_{rms} \sin(\tilde{\phi}) \tag{2.6}$$

and it is measured in volt-ampere reactive (VAR, Var or VAr).

These three kinds of powers (active, reactive and apparent) are related in the complex plane as illustrated in Figure 2.3 and thus S is the vector sum of P and Q:

$$S = P + iQ \tag{2.7}$$

where i is the imaginary unit. The apparent and active powers are also related

by the power factor $\cos(\tilde{\phi})$. This value represents the amount of real power in the measured apparent power since $\cos(\tilde{\phi}) = \frac{P}{S}$. Notice that when only real power is being transferred, current and voltage are in phase ($\tilde{\phi} = 0$), no reactive power exists in the electrical network and so the apparent power would be the real power ($\cos(0) = 1$). On the contrary, if only reactive power is being transferred, $\tilde{\phi} = 90$, $\cos(90) = 0$ and as $P = 0$, $\frac{P}{S} = 0$. Notice that low values of $\cos(\tilde{\phi})$ are associated with the existence of reactive energy on the network in higher quantity than active power leading to low efficiency of the electrical distribution system.

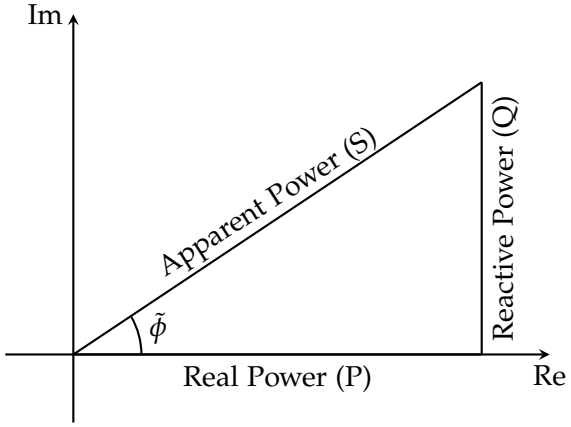

Figure 2.3.: The relation between the apparent power S, the active power P and the reactive power Q in the complex plane, where $\tilde{\phi}$ denotes the phase angle.

2.3. NILM Preliminaries

Non-Intrusive Load Monitoring (NILM) determines the electrical consumption of home appliances being turned on and off in a given residential network, based on the analysis of single-point measurements of current and voltage of the aggregated load (the total consumption of the network). This problem can be formulated as the separation of the aggregated signal corresponding to the consumption during a period of time T, $\bar{x} \in \mathbb{R}^T$,

$$\bar{x} = [\bar{x}(1), \bar{x}(2), \ldots, \bar{x}(T)]^T, \tag{2.8}$$

into the signals associated with each device, or circuit, $x_i \in \mathbb{R}^T$, $i = 1, \ldots, k$,

$$x_i = \left[x_i\left(1\right), x_i\left(2\right), \ldots, x_i\left(T\right)\right]^T. \tag{2.9}$$

Under the assumption that $\bar{x}$ is obtained by a linear mixing process (Zoha et al., 2012) corresponding to the sum of the signals x_i with $i = 1, \ldots, k$, then,

$$\bar{x}\left(t\right) = \sum_{i=1}^{k} x_i\left(t\right). \tag{2.10}$$

The NILM framework proposed by Hart (1992) in his influential work requires that (i) signals from the aggregate consumption of an electrical network are acquired; (ii) features of important events, as changes in the electrical power measurements from one nearly constant value (steady-state) to another and/or characteristics, are extracted; and (iii) identification of these events is performed. The NILM block diagram is shown in Figure 2.4. These events, associated with steady-state changes in the signal and characterized by their magnitude and sign in real and reactive power, correspond to the turning on or turning off of an appliance in the network. The two-dimensional space defined by the active and reactive powers, illustrated in Figure 2.5, is designated as "signature space". In this approach, the energy consumption of individual appliances is achieved by a pairing step of the detected events with equal magnitudes and opposite signs.

The original NILM algorithm is a five-step approach for energy disaggregation based on the step changes detected in the aggregated power signals as described in Algorithm 1, which was successfully applied in the residential sector (Laughman et al., 2003). Still, this technique, based on the "signature space", assumes that the events associated with distinct loads are unique and no overlap in the two-dimensional space defined by the active and reactive power occurs. Due to its simplicity, similar approaches were employed in further works (Norford and Leeb, 1996; Laughman et al., 2003). Nevertheless, the cost of its low complexity is expressed by the inability to recognize different appliances with overlapping features in the defined two-dimensional space. In fact, as the number of different appliances of interest increases, the number of events in this plane expands accordingly. Consequently, the loads become indistinguishable due to the overlap in the "signature space" (Laughman et al., 2003). In addition, this method is also sensitive to power drifts (Zoha et al.,

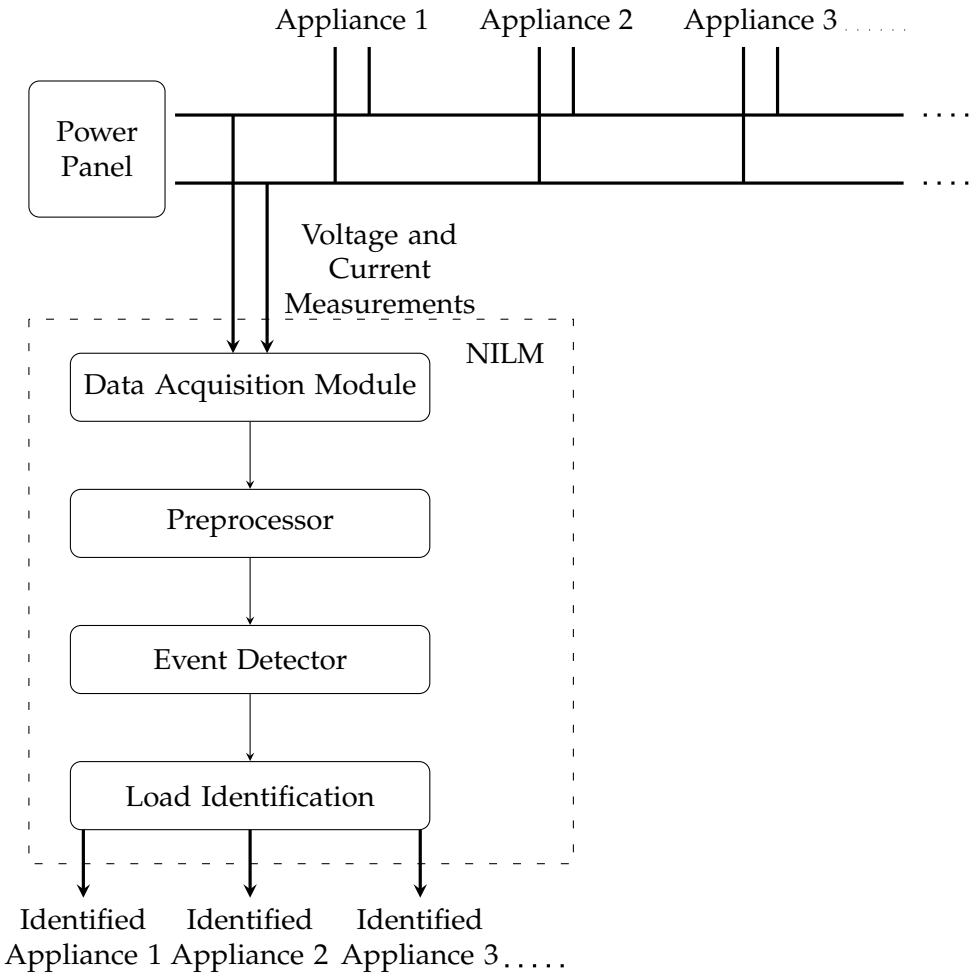

Figure 2.4.: NILM Block Diagram. The Data Acquisition module gathers, at the power panel of a building, the voltage and current measurements of an electrical circuit to which appliances are connected. The data is fed into a Prepocessor module, if needed, and then events are detected. Next, events are the input of the Load Identification module whose output are the appliances connected to the circuit.

2012). Extension works tried to overcome these limitations, for example, with the definition of a three-dimensional feature space or considering additional features (Laughman et al., 2003; Farinaccio and Zmeureanu, 1999).

The subsequent NILM approaches found in the related literature explored different appliance signatures (a set of features that can be extracted from the total load which supplies information about the device's nature or operation state), mainly derived from the step changes between steady-states (steady-state signatures), as Hart (1992), or derived from characteristics associated with the

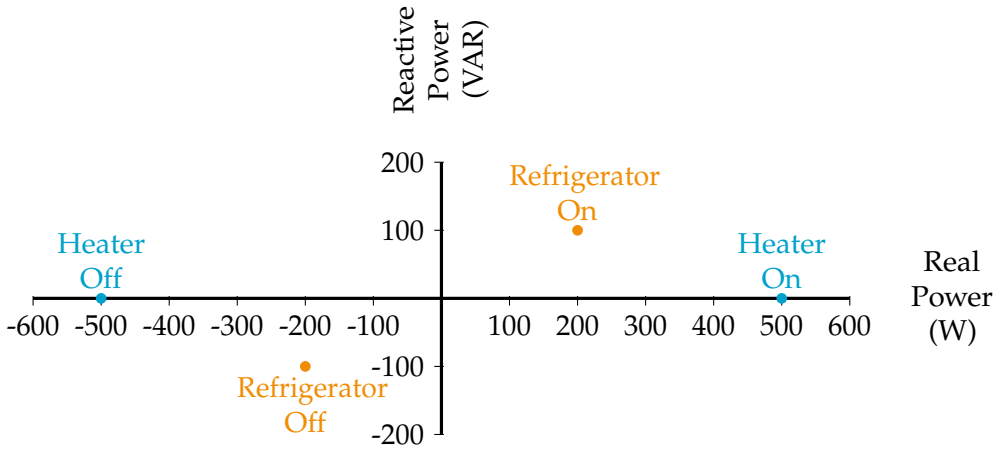

Figure 2.5.: Changes in the two-dimensional space defined by the active and reactive powers ("signature space"). A heater and a refrigerator are represented based on the active and reactive powers. Source: Hart et al. (1989).

Algorithm 1: Original Hart's NILM algorithm.

Data: Single-point aggregated energy measurements
Result: Identification of each appliance in the electrical network
1 Identification of changes in the steady-state ("edge detector")
2 Performance of cluster analysis to locate these changes in the two-dimensional space defined by the active and reactive powers ("signature space", see Figure 2.5)
3 Matching of clusters with similar magnitude and opposite sign
4 Association of unmatched clusters with existing or new clusters according to a best likelihood algorithm ("anomaly resolution")
5 Assignment of human-readable labels, according to a database of known appliance electrical consumption learned at a training phase

transient period between two steady-states (transient signatures). In terms of framework, these approaches followed a similar design to the original NILM. Roughly, the general framework of these approaches considers the existence of appliance signatures and is composed by tree main steps:

- the acquisition of an electrical signal (Data Acquisition);
- the extraction of important events and/or characteristics to define an appliance electrical signature (Feature Extraction and Appliance Electrical Signature);
- the classification of events for appliance identification (Model Learning

and Classification for Load Identification).

An overview of each one of the three highlighted steps is presented in the following sections.

2.4. Data Acquisition

This section addresses the process of data acquisition that is extremely relevant in a NILM system, namely for signature definition as will be described in Section 2.5. On the one hand, given the data and its sampling rate, the type of characteristics being used to describe the appliances can be defined. On the other hand, the kind of appliance signatures to be studied can determine the data sampling rate.

Traditional power signals, as the real and reactive power, needed for the step changes signatures, as studied by Hart (1992), can be computed using the current and voltage waveforms sampled at twice the highest frequency component of interest in order to capture it, according to the Nyquist-Shannon sampling theorem (Shannon, 1949). In the United States of America this sampling rate would be of 120Hz, since the alternating current is delivered at 60Hz (Matthews et al., 2008), while in Europe it is delivered at 50Hz. However, as it will be described in Section 2.5.1, harmonic analysis can also be useful for appliance identification and therefore current harmonics should be captured with a sampling of about 1.2-2kHz (considering that no harmonic higher than the 11^{th} is of interest) (Zeifman and Roth, 2011). If instead of harmonic analysis, the distinctive features of interest belong to the transient period between two steady-states (transient signatures) then the sampling frequency should be in the order of kHz or even of MHz. The latter was used by Patel et al. (2007) to capture the turn-on (start-up) transient noise to define appliances' signatures.

The commercial power meters, as the one illustrated in Figure 2.6, have an internal high sampling frequency adequate to capture the start-up transients. However, the power signals are reported at a small sampling rate of 1Hz or less (Matthews et al., 2008). The low-frequency meters are an 'on-budget' metering solution able to provide at least one signal of interest as the real power (Berges et al., 2011). For a finer granularity report a measuring instrument able to provide data at MHz rate is required. These high-frequency sensors are usually expensive, hence, more suitable for professional use and for industrial

monitoring rather than for residential systems (Berges et al., 2011; Zoha et al., 2012).

Figure 2.6.: A power meter from the ISA company.

Although the low frequency meters are more interesting for a household NILM solution they do not produce consistent measurements, often showing variations of 10% to 20% (Matthews et al., 2008; Berges et al., 2008). Other limitations, as the low resolution associated with the Analog to Digital converter and small storage capacity, have encouraged researchers to build their own meter sensors (Patel et al., 2007; Berges et al., 2011; Zoha et al., 2012).

Regarding the related literature, in an early stage of NILM a sampling rate of 1Hz was often used, as in (Cole and Albicki, 1998a; Norford and Leeb, 1996). Also, samples spaced by 16 seconds period were collected as by Farinaccio and Zmeureanu (1999); Marceau and Zmeureanu (2000). These values increase when the goal is to capture start-up transient and they can range from 8kHz (Laughman et al., 2003) to 50kHz (Ting et al., 2005).

Recently, to overcome the lack of public data, two datasets were made freely available: the REDD (Kolter and Johnson, 2011) and the Building-Level fUlly-labeled dataset for Electricity Disaggregation (BLUED) (Anderson et al., 2012b). The former reports the power usage for several houses, in particular, the whole-home consumption and the detailed demand by appliance (or circuit) over several months' time. The whole-home signal was recorded at high frequency (15Hz) while each labelled circuit or appliance signal was recorded at 1Hz. This dataset is specially suitable for non-event based approaches in order to solve the energy disaggregation problem (Anderson et al., 2012b). This is the main difference between REDD and BLUED datasets. The latter is composed by measurements of almost fifty appliances in one home during a week. Current, voltage signals measured at 12kHz and active power signals computed at

60Hz are reported and fully labelled according to the operating status of each appliance (turned on or off). This particular dataset is suitable for the development and testing of event-based approaches for the resolution of the energy disaggregation problem.

At the same time, a dataset with a broader purpose was also presented recently by Barker et al. (2012). This dataset is part of the Smart* project and includes data from the average whole-home electrical consumption at 1Hz rate. It includes also data for every circuit and corresponding appliance. In addition, other information, as outdoor weather data, temperature and humidity are reported, which make this dataset suitable for other potential uses beyond NILM and for the analysis between non-power events and power events (Zoha et al., 2012).

2.5. Feature Extraction and Appliance Electrical Signature

This section surveys feature extraction and appliance electrical signatures. After collecting the raw data, the following step is to process it in order to compute, for instance, the real and reactive powers, and next event detection takes place. The changes in the power measurements are analysed and then the associated events of switching on or off an equipment are described in terms of steady-states or transients. For the steady-states, a change in the real power corresponding to an increase in the signal may be associated with the switching on of a device. For the transients, it is necessary to extract features, as its shape, size and duration, which require high sampling rates.

From a conceptual point of view, both kinds of signatures have their advantages and disadvantages in terms of appliance identification performance and also in regard to the hardware specifications for the data collection. If the priority is to incorporate transient features for enabling the disaggregation of distinct appliances with a similar steady-state description then the use of expensive hardware is necessary. On the other hand, if a low cost solution is the goal the system to be developed should consider steady-state signatures.

The key element for accurate appliance identification within a NILM framework is the electrical signature of each appliance composed by the extracted features from the input data. Under this paradigm, the first question about

intrusiveness degree is the way the needed information is gathered (Hart, 1992). A manual setup requires an intrusive period during which the appliances are manually turned on and off and their electrical signatures and identification are collected. An automatic setup uses *a priori* information about the features and possible appliances, without requiring any manual setup (Hart, 1992). The manual setup would be the most accurate. Still, due to its complexity, mainly in homes with a large number of devices, it can be considered as undesirable from the consumers' point of view. However, this setup can be interpreted as a step towards automatic NILM since the information thus gathered would constitute a useful set of signatures for the automated NILM (Hart, 1992).

The definition of a signature is related to the operational nature of each appliance. Its operation behaviour can supply important features for the identification. Devices can be divided in three distinct types (Hart, 1992):

- on/off appliances that only have two distinct states of operations, like lamps and toasters;
- multi-state appliances, usually represented as Finite State Machine (FSM), that have a finite set of operating states, like washing machines;
- continuous variable appliances, which can be seen as a generalization of the previous ones, that have variable power levels and infinite number of states, like light dimmers and power tools.

The repeated behaviour associated with the multi-state appliances make their operation easier to identify. On the other hand, the continuous variable appliances due to their non-repeated states present a challenge for NILM methods. A fourth group characterization is referred by Zeifman and Roth (2011) composed by 'always on' appliances: once they are on they remain on 24/7 with constant power consumption, as smoke detectors and telephone sets. Figure 2.7 exemplifies the loads graphics for each type of device.

Residential appliances can also be separated by their basic power consumption unit: induction coil or motors, as in washing machines, fans and air conditions; heating resistance, as in incandescent lamps and toasters; electronic circuit, as in PC and TV sets. Based on the power types and on the working styles of devices, Wang and Zheng (2012) recently presented another categorization of appliances that the authors used for the extraction of relevant features regarding their identification.

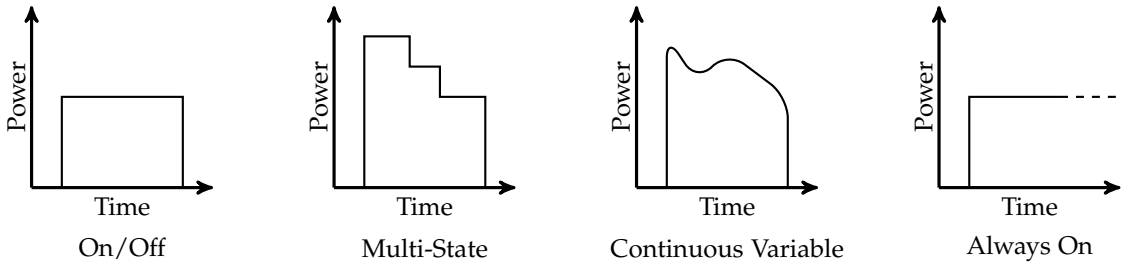

Figure 2.7.: Illustration of the different types of operation behaviours.

Device signatures "are the essence of the Nonintrusive Appliance Load Monitoring (NALM)" (Hart, 1992). In the following, the different types of signatures are presented in more detail.

2.5.1. Steady-States Signatures Review

A segment of a given sampled signal is considered as stable if the sampled values, associated with consecutive samples, are within a given threshold. A steady-state signature can be deduced from the difference between two steady-states in a signal. For instance, the change in the power measurements from a lower value (before turning on the device) to a higher value may indicate that an appliance was switched on. Current, real (or active) and reactive power and admittance can be analysed in search for these variations (Sultanem, 1991; Hart, 1992; Pihala, 1998). Steady-state signatures derive from the difference between steady-states in these signals and can be divided into fundamental frequency properties and harmonic frequency properties (Hart, 1992; Pihala, 1998).

Fundamental Frequency

The fundamental analysis consists in the observation of a signal and the search for step changes. The original prototype presented by Hart (1992) seeks for changes in the active and reactive power. The method successfully identified the on/off appliances tested, still it had difficulties identifying multi-state and continuous variable appliances running below 150W. Moreover, appliances with similar consumptions (similar changes in the active and reactive powers) can be misidentified, *e.g.* a computer and incandescent bulb as presented in (Hart, 1992; Laughman et al., 2003), since their features 'overlap' (Figure 2.8).

The initial described method therein was further explored by considering either additional features to the defined signature or only state-changes in the

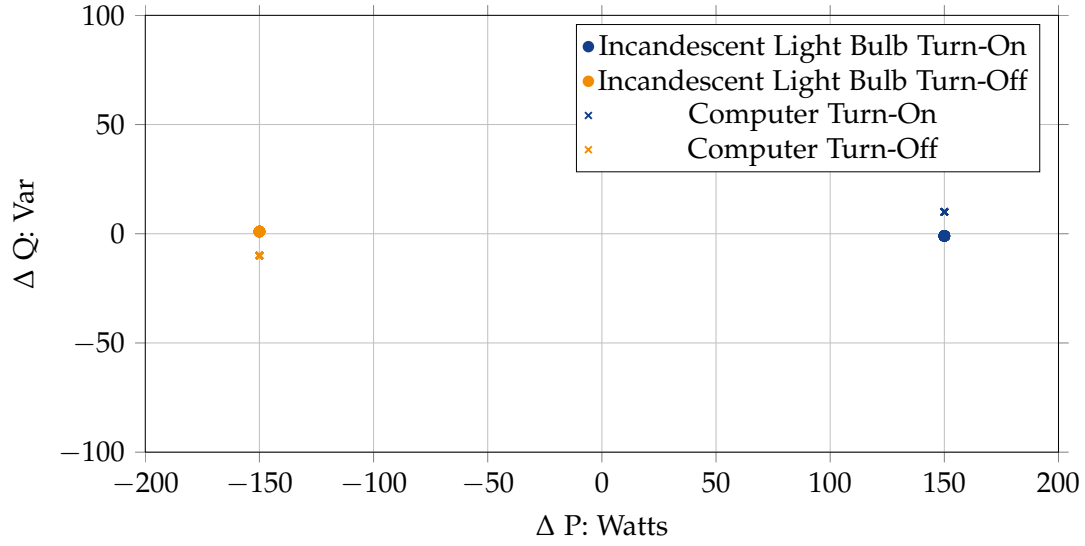

Figure 2.8.: Illustration of the Real-Reactive plane, the steady-state power consumption of a computer and a bank of incandescent lights. Source: Laughman et al. (2003).

real power for appliance recognition. For instance, Cole and Albicki (1998a,b) focus on finding additional features to characterize complex loads as heat pumps, dishwashers or refrigerators. These type of loads do not usually present a discrete change of the signal, they are more complex at the switching on moment than during their own operation mode. Therefore, in addition to the step changes in power, these appliance signatures should also consider the initial upward spike on the power, *edges*, and slower changing variations, *slopes*. According to the authors, and since they are monitoring multiple loads, the power level is not unique during the steady-state of a particular load thus this is not a distinctive feature. Only the edges and slopes are unique and enable for load distinction (Cole and Albicki, 1998a,b).

Despite the disadvantage of not producing a unique ID, researchers explored steady-state signatures due to their simplicity, in particular using steady-states in the real power signal, as in (Farinaccio and Zmeureanu, 1999; Marceau and Zmeureanu, 2000; Baranski and Voss, 2004a; Bijker et al., 2009). These studies demonstrated that major-use appliances, such as a water heater or a refrigerator, can be identified from the aggregated load since their energy consumption is distinctive enough from the electrical demand of the remaining appliances. The common factor between the above mentioned research is how the appliance signature is defined (step changes in the real power signal) while the differences

lie on how to perform the identification.

Decision rules specific for each appliance are required by the approach presented by Farinaccio and Zmeureanu (1999) and Marceau and Zmeureanu (2000). The definition of these particular rules requires a training period that, as suggested by the authors, should be of one week and during which each appliance of interest is individually metered. The error associated with the evaluation of the daily energy consumption ranges from 10.5% to 15.9%. Two main limitations are related to this method: the need to design appliance-specific rules and the non-consideration of appliances with similar power consumptions. Later, the improvements described by Marceau and Zmeureanu (2000) included preprocessing involving detection of on/off events and their consistency by examining duration statistics. This approach is able to identify two distinct appliances with similar consumptions as long as they present different operation durations. For the major appliances as water heater, refrigerator, clothes washer, stove and dishwasher the approach yielded a test accuracy of 90% for the estimes of their energy consumption.

The disaggregation of electrical consumption into its major end-uses is also the focus of the work presented by Baranski and Voss (2004a,b). Despite the signature being also based on active power, this approach differs from the previous one in the fact that no *a priori* individual knowledge is needed. Instead, frequency patterns are detected in the active power signal acquired at a sampling rate of 1Hz. The steady-state changes are detected, treated as events and clustered by fuzzy clustering. As a consequence, each cluster is composed by events with similar structures. A large set of events is then matched to an appliance, while for the method proposed by Hart (1992) only a pair of on and off events is matched each time. As this approach is based on the frequency patterns only significant or frequent appliances are detected thereby devices that are only occasionally used cannot be identified.

A Last-In-First-Out (LIFO) approach is proposed by Bijker et al. (2009). Every appliance before being switched off needs to have been switched on, thus the matching algorithm computes the step changes in chronological order. Any positive step-change is saved in memory and every negative step-change is matched with the corresponding positive step-change. The method was able to distinguish main end-uses between the major appliances as stove and oven. However, on average 42% of the total energy consumption was not detected. In fact, it presents similar limitations to the original approach presented in (Hart,

1992): first, difficulties identifying appliances other than those with on/off states; second, the lack of capacity to distinguish devices presenting similar consumptions. Identical restrictions were found by Berges et al. (2009). In this work, the state changes in the real power signal are used as signatures of on/off appliances and a similar strategy based on the sequential ordering of the apliances was also followed by the authors to help the identification of the several devices.

To overcome the limitations of using only the step changes in the power consumption, the signature could be enriched with specific features extracted from the current and voltage signals, for instance, peaks and root mean square values of these signals and the values of the power factor. These features were explored for a Real Time Recognition and Profiling of Appliances system, in particular for energy disaggregation of kitchen appliances (Ruzzelli et al., 2010). The system presented 84% of accuracy in the identification of on/off appliances. Nevertheless, multi-state appliances were not included and it is conceded that further experimentation is needed. The inclusion of the root mean square values was also investigated by Kato et al. (2009), showing that these features are more discriminative in comparison with peak values, regarding non-continuous variable appliances and non-simultaneous activation of devices.

Harmonic Frequency

In addition to the fundamental analysis the harmonic analysis can be used to define an appliance signature, mainly of small devices (portable or semi-portable machines, usually standing on tabletops, to accomplish a household task) (Hart, 1992), and to perform power quality analysis (Chan et al., 2000). The additional information provided by the harmonic currents is useful to distinguish loads that 'overlap' on the real and reactive power domains, in particular, non-linear loads that trace non-sinusoidal current waveforms during their operation (a linear load would correspond to a sinusoidal waveform) (Hart, 1992; Najmeddine et al., 2008). For instance, a water boiler (resistive components) presents a sinusoidal waveform while TV set (electronic components) describes a non-sinusoidal waveform and an induction cooker has high-order harmonics (Liang et al., 2010a) as illustrated in Figure 2.9.

The current waveforms are the basis of the integer programming problem proposed by Suzuki et al. (2008) to solve NILM. The technique does not require a learning process when new devices are installed in the home, since only

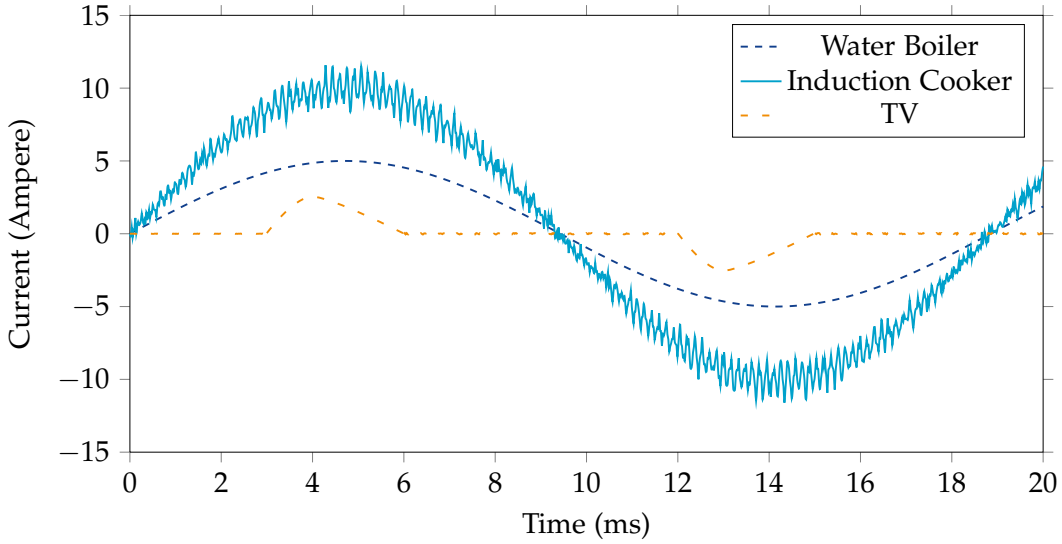

Figure 2.9.: Illustration of different current waveforms. Source: Liang et al. (2010a).

a period of current waveform is needed. Nevertheless, continuous variable appliances are not considered. This waveform is also used in (Yi-xin et al., 2008; Liang et al., 2010a) as part of a multi-feature signature composed also by fundamental and harmonic currents (steady-state voltages and currents) and by the active/reactive power and the harmonics, respectively.

The harmonic analysis is usually performed by calculating the Fourier Transform (FT) either short or fast (Laughman et al., 2003; Srinivasan et al., 2006; Zeifman and Roth, 2011; Li et al., 2012) or the Wavelet Decomposition (Chan et al., 2000) of the current signal. For example, in the work presented by Srinivasan et al. (2006), the important information contained in the harmonic components of the current waveform is obtained by Fast Fourier Transform (FFT). This information is then used to train Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs) based models for the appliance identification. The system needed to be trained with the greatest number of possible combinations of appliances, which turn it into an unpractical solution for a real-world home (Zeifman and Roth, 2011). SVMs were also selected by Nakano et al. (2007) to infer the on/off states of the devices based on the harmonics analysis in addition to Radial Basis Function (RBF) Networks to deduce their power consumption. Posterior work (Li et al., 2012) has followed similar steps, employing the FFT to transform the current signal in the frequency domain and using it for power decomposition based on SVM regression. This representation is particularly

distinctive for resistive loads, as toasters, and switching loads, as PC or Laptops. On the other hand, the Wavelet Transform (WT) is an alternative to the FT since simultaneous time and frequency information is provided. Hence, harmonic analysis can be performed applying wavelets as explored by Chan et al. (2000), showing that the wavelet features were distinct for several devices.

These 'microscopic' features (harmonics and signal waveforms) used as a complement to 'macroscopic' ones (power changes) require high sampling rates. Admittedly, regarding the Nyquist-Shannon sampling theorem (Shannon, 1949), and that no harmonic higher than the 11^{th} is of interest, thus the ideal minimum value should be of about 1.2-2kHz. However this minimum may lead to transmission and storage limitations (Zeifman and Roth, 2011).

In summary, the signatures based on power changes require a small sampling rate (usually 1Hz) (Zeifman and Roth, 2011), allow for the identification of the switch off, and allow for the disambiguation of two simultaneous events since these signatures are additive (Hart, 1992). Two different devices with the same steady-state signature, on the other hand, can not be distinguished although harmonic features can bring valuable additional information to overcome this 'overlap' of signatures. Also, the small sampling rate can represent a limitation since sequences of turn on loads during a period smaller than the sampling rate are not possible to identify.

2.5.2. Transient Signatures Review

The switching on (or off) of an appliance corresponds to a change between two steady-states. During this transition, modifications on the electrical signals occur, which can characterize a given device (Figure 2.10). In fact, this transitory behaviour is deeply related to the task performed by the specific device: charging the capacitors in a computer power supply is different from heating the filaments of an incandescent lamp (Laughman et al., 2003). The distinctive features extracted from the signals analysis during this transition compose the transient signature of that device. The processing requirements associated with this kind of signature are more demanding than for the step changes, especially regarding the needed sampling rate. Moreover, these transients can be less informative than the steady-state signatures (Hart, 1992). Once again, we are looking for 'microscopic' features (Zeifman and Roth, 2011).

The transient behaviour can be described by its shape, size, duration and time

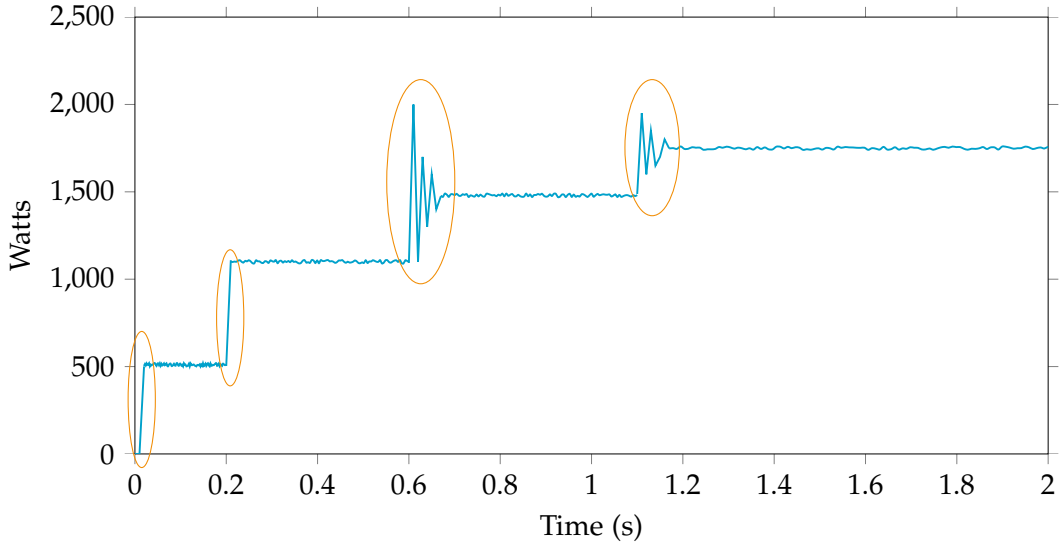

Figure 2.10.: Illustration of active power and transient delimitation. Source: Najmeddine et al. (2008).

constants (Hart, 1992). In the early stages of NILM, these were explored for an extension of the original Hart's prototype regarding its use in commercial and industrial buildings. In such structures, efforts to control the power factor, to reduce the reactive power and to balance loads are made in order to homogenize the steady-states of the different loads (Leeb et al., 1995; Norford and Leeb, 1996). Therefore, transient signatures represent a suitable alternative under this particular context and in (Leeb et al., 1993; Leeb and Kirtley Jr, 1993; Leeb et al., 1995; Norford and Leeb, 1996) techniques to develop, test and implement a transient event detector are described. The initial presented method by Leeb et al. (1993); Leeb and Kirtley Jr (1993) searches for transient sections with significant variations, called of *v-sections*, rather than the entire transient waveform. A filter is then applied to extract distinctive features for each appliance. In the tests, four types of loads (an instant-start fluorescent lamp bank, a rapid-start fluorescent lamp bank and two induction motors differing in the horsepower rating) were correctly identified. Moreover, two simultaneous loads were correctly distinguished though no *v-sections* were overlapped. The extraction of features from these *v-sections* was posteriorly improved in (Leeb et al., 1995) with the use of spectral envelopes (curve in the frequency-amplitude plane derived from a Fourier magnitude spectrum) of the current signal, which represent an extension of the harmonics analysis since it consists of the first several coefficients obtained by FT (Zeifman and Roth, 2011). Notice that

this type of signature is able to detect variable load appliances, which were addressed by Lee et al. (2005); Wichakool et al. (2009), regarding the estimation of associated power consumption, another requirement of NILM. A recent work of Shaw et al. (2008) describes the addition of a diagnostic module to a NILM system. The studied procedure requires a dataset of signatures (spectral envelopes) and the identification is performed by searching for the closest signature in terms of distance. According to Zeifman and Roth (2011), further significant developments concerning this concept of the spectral envelope and its application to NILM are not expected to occur.

The transient signatures can also be used in a residential non-intrusive load monitoring. The approach proposed by Patel et al. (2007) makes use of the electric noise produced by the switching of appliances (transient noise) and during its operation (continuous noise). The noise sensor is able to detect the transient signals occurring in the electrical circuit of the house when plugged in to a central socket. The duration of this transient noise can be in the order of microseconds and consists of a spectrum rich in frequencies that can vary from 10Hz to 100kHz. Features were extracted from this signal by FFT and used as the devices' signature. The system requires a training phase where the features of each appliance and their combinations are learned. Later on, Gupta et al. (2010) points out four drawbacks of the previous work (Patel et al., 2007). First, the computational process needed to capture and to analyse the transient noise is expensive. Second, the extrinsic properties of the signature used do not allow for the identification of another similar appliance. Third, these signatures are dependent on the wiring system of the house in study. Fourth, the transient events, which are relatively weak due to their broadband distribution of signal energy, do not occur very often.

Patel et al. (2007) and Gupta et al. (2010) aim at identifying different types of appliances. The former work was focused on resistive and inductive electrical loads while the latter author is interested in Switch Mode Power Supplies (SMPS), specifically modern consumer electronics and fluorescent lighting. Thereby, the authors state that the techniques are complementary. Indeed, Gupta et al. (2010) is centred on modern households devices that have a *soft-switch* which minimizes the transient generated at the moment of activation. This kind of devices uses SMPS that continuously generate high frequency Electromagnetic Interference (EMI). In the described approach, the EMI signals are analysed employing FFT. Features are then extracted to form the devices'

signatures. The authors claimed that these signatures are able of balancing two apparent opposite tasks. On the one hand, they are general enough to define a given appliances' model (stability across homes). On the other hand, there is enough variability to distinguish between similar appliances in the same home (Gupta et al., 2010). The reported test mean accuracy is 93.82%. In spite of appearing to be more suitable for some modern equipment than the work of Patel et al. (2007), this approach raises some issues (Zeifman and Roth, 2011). Like in (Patel et al., 2007), the technique proposed by Gupta et al. (2010) may also be dependent of the home wiring system hence a change of the socket may lead to device misidentification. As the authors showed, some modern appliances may not have EMI, like dryers and electric stoves. Some other important questions raised by Zeifman and Roth (2011) are: (i) although the signatures are general and variable enough, it is not clear how the load identification would occur in an apartment building; or (ii) how overlapped EMI signatures would be analysed to perform identification. Another relevant question (iii) is that this type of signature is dependent on the manufactures' specifications of the SMPS used in a class of devices, which can change anytime (Zeifman and Roth, 2011). Lastly, (iv) the energy consumption of each device - a NILM request - is not addressed.

The transient signature has advantages and disadvantages. Its main advantage is the fact that the transient behaviour is related to the internal operation of each appliance and therefore this signature is able to distinguish two different appliances with similar steady-state description. The main disadvantage is related with the duration of the transient behaviours, which is of few seconds, and therefore high sampling rates are needed in order to detect them. Moreover, the expensive process for the analysis of transients (Cole and Albicki, 1998a; Patel et al., 2007; Gupta et al., 2010) may turn this type of signature into an uninteresting solution for residential non-intrusive load monitoring.

2.5.3. Alternative Signatures Review

The types of signatures used for the identification of appliances in a NILM system are not restricted to those described above. In fact, steady-state and transient signatures can be joined to form a single appliance signature. Either way, non-traditional features can be extracted to define a signature. For instance, Sultanem (1991), at the same time of Hart, Kern and Schweppe's research,

proposes the use of active and reactive power changes in addition to the transient duration as signatures. Besides, the shape of the current transient and of the harmonic currents were also suggested yet not included in the Sultanem's prototype. Sultanem (1991) concluded that, in most cases and regarding the appliances in study, the active and reactive changes are enough to identify the appliances. Nevertheless, details about the algorithm and corresponding accuracy are not disclosed.

As the transient signatures were initially studied to implement the NILM system in industrial and commercial buildings expected progress would involve the two types of signatures. In (Norford and Leeb, 1996), the subsequent work of Leeb et al. (1993), transient patterns of the real and reactive power and also of harmonics from current and voltage signals are proposed as signature. Nevertheless, the authors intended to include the harmonic analysis only in a posterior prototype. This work is a sequence of (Leeb et al., 1993, 1995) described in the previous section, thereby, the transient analysis is also based on the definition of *v-sections*.

Chang et al. also explored the transient signatures for industrial and commercial buildings (Chang et al., 2008b,a; Chang, 2012). Chang et al. (2008b) show that the transient energy associated with the switching on of a device (turn-on transient) exhibits repeatability. It is also shown therein that the efficiency of load identification and computational time improved. Moreover, this type of signature, when used in addition to the steady-state signatures, also improves load identification when multiple operations occur in the same circuit. In the latter work (Chang, 2012), the transient response time was studied, using Short-Time Fourier Transform (STFT) and also Discrete Wavelet Transform (DWT), and it was added to the turn-on transient energy as load signature. This signature is suitable to describe different loads with the same real and reactive power. Furthermore, as shown by the experiences, it is appropriated to distinguish loads with simultaneous starting time for industrial and commercial loads in particular. Chang et al. (2008a) investigated a signature considering turn-on transients to distinguish loads with the same real and reactive power both simulating data from a commercial building as well as lab appliances. Later, the study was extended to industrial loads (Chang and Yang, 2009; Chang et al., 2010).

Despite that the above referred research is focused on commercial and industrial NILM frameworks, most of the work in this field addresses mainly

residential solutions where the signatures include steady-states and transient features. For instance, Lee et al. (2004) explored start-up transients, steady-states and operational pattern signatures (the continuous changes in the appliance mode of operation along the time domain). The experiments showed that the voltage waveform does not change significantly for different devices. Notwithstanding, the electrical circuit current was distinctive enough to represent a load signature. Thereby, Ting et al. (2005) investigated only the current waveforms and their start-up, the steady-state, the transient and, in addition, the voltage-current trajectory for the taxonomy presented. The V-I (voltage-current) trajectory, also studied by Lam et al. (2007), is a two dimensional load signature characterized by the nonexistence of time reference, the asymmetry, the looping direction, the area, the curvature of mean line, the self-intersection, the slope of middle segment, the area of left and right segments and the peak of middle segment. The signature was compared with traditional power based signatures. It was shown that the latter promotes the clustering of appliances with distinct loads (as resistive power electronic and motor driven appliances) while Lam's approach was able to distinguish between them. Nevertheless, no algorithm for non-intrusive load monitoring had yet been described based on these features (Zeifman and Roth, 2011). Liang et al. (2010a) proposed the use of several features to characterize the load such as: the current waveform, changes in real and reactive power, harmonics, instantaneous admittance and power waveforms, eigenvalues and switching transient waveform. The authors described two approaches to solve the disaggregation problem via optimization and pattern recognition, and computational experiments were provided in a later work (Liang et al., 2010b).

Alternative and not obvious features have been proposed recently by Wang and Zheng (2012). Their approach is based on the separation of the power consumption signal into two main parts: the fast switching events (*triangles*) and the steady working events (*rectangles*), while smaller fluctuations are not taken into consideration. These two basic units can be described by their start time, peak time, peak value in addition to end time for the *triangle* unit, and to steady time and steady power for the *rectangle* as illustrated in Figure 2.11. The authors argue that its main advantage is the reduction of overlaps regarding the data features. These comprise the average active power, the power factor and the active power changes to characterize each appliance. Other features related to the usage pattern of the appliance, as the time of the day and the day

of the week, have been studied together with traditional features to improve the identification accuracy (Kim et al., 2011). In the same direction, time-on and time-off durations additionally to historical data were also proposed as additional features by Zeifman (2012).

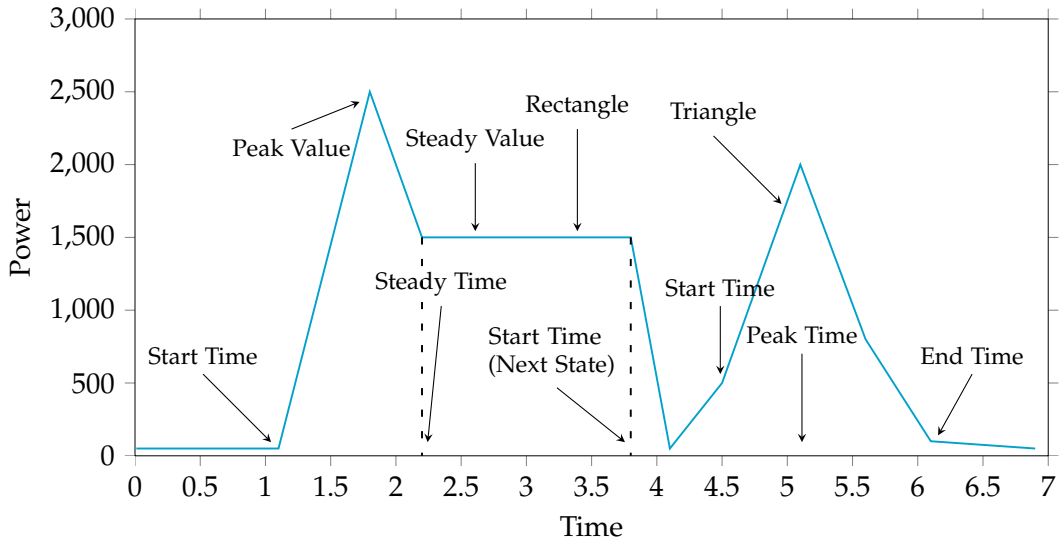

Figure 2.11.: Illustration of triangles and rectangles. Source: Wang and Zheng (2012).

In short, additional features to traditional signatures intend to increase the appliance identification accuracy. The solution of a composed signature by steady-states and transient events is mainly directed to commercial and industrial buildings. Note that the sampling rates required to detect the step changes for the steady-states signatures are less demanding than the ones needed to capture spikes in the waveforms for the transients since their duration can be of only few milliseconds (Hart, 1992; Matthews et al., 2008). Once again, transient events do not represent a suitable solution for a residential system (Cole and Albicki, 1998a). As a consequence, researchers have explored alternative features from waveforms to time and days of usage in order to achieve a performance as accurate as possible in what concerns load identification.

2.6. Model Learning and Classification for Load Identification

In this section, model learning and classification techniques for the load identification are addressed. Once features are extracted from the collected data, comprising the appliance electrical signature, the identification of the load associated with the changes in the measured aggregated signal can be performed. For this identification, research within NILM contexts has been mainly focused on supervised methods rather than unsupervised techniques. The former requires the use of labelled data for learning a model able to correctly identify the events or loads while the latter approach eliminates the need of labelled data. According to Liang et al. (2010a); Zeifman and Roth (2011); Zoha et al. (2012), the supervised strategies explore optimization approaches that seek for a combination of appliances for which the associated load demand would be as close as possible to the observed load on the aggregated signal. Beyond the optimization techniques, pattern recognition approaches able to recognize the appliance signatures (the extracted features) are also worthwhile of investigating.

Optimization approaches require the existence of datasets with all the possible combinations of power demands regarding the several appliances. This is a drawback since the presence of unknown loads in the aggregated signal would interfere with the identification based on the combination of known devices (Hart, 1992; Liang et al., 2010a). Therefore, pattern recognition methods are generally preferred by researchers (Zoha et al., 2012). Nevertheless, the supervised approaches need labelled data which may not be always available. Thereby, unsupervised approaches have been received increasing attention (Zoha et al., 2012). These are usually non-event based, in contrast with the supervised techniques, and seek for straightforward disaggregation directly from the aggregated signal without event detection.

2.6.1. Supervised Approaches

Supervised approaches for load identification require a training set of $\bar{N}$ examples $\{(v_n, t_n)\}, n = 1, \ldots, \bar{N}$, where $v_n \in \mathbb{R}^D$ is the nth vector of features and t_n is its appliance label. Thereby, the labelling process of this training set is an important step for supervised techniques that can be performed during

some period of time, usually previously to the model learning process (off-line approach) or almost in parallel with real-time detection and features learning (on-line approach) as in (Berges et al., 2011).

The manual labelling of data (for an event based dataset) is demanding when only the aggregated data is observed as well as when all the individual appliances are being measured. In particular, event labelling is a complex task when only the aggregated signal is used (Zoha et al., 2012). Thereby, recent research has been focused on this issue, proposing alternatives and (almost) automatic solutions (Schoofs et al., 2010; Rowe et al., 2010; Giri and Berges, 2012) to simplify all the data collection needed for the training, event labelled data in particular. Note also that a sub-metering approach, as used by Berges et al. (2011); Farinaccio and Zmeureanu (1999), is not the most practical alternative since it requires the installation (and setup) of sensors in every device of interest, which can be numerous in a modern household. Nevertheless, no standard automated solutions for data collection processes are known (Zoha et al., 2012).

Optimization Methods

Load identification regarded as an optimization problem requires the definition of an objective function. Considering a single load and given the training set, the objective is to minimize the residue between a training instance, v_n, whose label is known and a new load v. Formally, the new load is assigned to the class of v_n if

$$\min_n \|v_n - v\| \tag{2.11}$$

where v_n is the example n from the known dataset and v is the description of the new load. The problem complexity increases as the number of different loads in the aggregated signal increases, since the algorithm should take into account all the possible combinations of appliances contained into the training set instead of one-to-one matching (Zoha et al., 2012). However, some researchers deal with the disaggregation problem employing optimization methods (Baranski and Voss, 2004a,b; Suzuki et al., 2008; Liang et al., 2010a) including genetic algorithms (Baranski and Voss, 2004a; Liang et al., 2010a) and integer programming (Suzuki et al., 2008; Liang et al., 2010a). Apart from the complexity issue, two other factors hinder an adequate solution: (i) the presence of new loads in the aggregated signal, not included in the dataset, (ii) appliances with similar signatures are difficult to distinguish by this approach (Zoha et al., 2012).

Machine Learning and Pattern Recognition Methods

To overcome the mentioned issues, an often alternative found in the related literature rely on the use of pattern recognition methods for load disaggregation (Zoha et al., 2012). On a supervised strategy, statistical models are trained with input examples (*e.g.*, appliance signature) and their target labels (appliance name), *i.e.* with examples of the training set. The goal of supervised learning methods is to learn a statistical model from labelled signatures so that it can be generalized to new unseen loads and predict correctly the associated appliance.

The original NILM algorithm pioneered by Hart (1992) describes a clustering based approach for the load identification as presented in Algorithm 1. Nevertheless, its simplicity is expressed by the inability to recognize different appliances with similar features in the signature space. This limitation was tackled using a three dimensional (3D) feature space instead of the P-Q plane by adding the changes in the 3^{rd} Harmonic (Laughman et al., 2003). Another alternative consists on using additional features such as: the variation of electric current in time or the frequency of activation and on the definition of rules for recognition of objects (Farinaccio and Zmeureanu, 1999). Other works propose the use of distinct classification methods, as for instance, the 1-Nearest Neighbour (Berges et al., 2009; Gupta et al., 2010; Berges et al., 2010), Decision Trees, Multiclass Adaboost and Gaussian Naïve Bayes while using signatures only comprising information of the real consumption (Berges et al., 2009, 2010). Berges et al. (2009) illustrates that the simple 1-Nearest Neighbour using a Euclidean distance measure between the signature vectors achieves 90% of accuracy.

The K-Nearest Neighbours is a non-parametric, lazy learning classification method, usually employed as baseline technique (Silva and Ribeiro, 2009). No representation is built for each appliance. Instead, when $k = 1$ this method finds the nearest example in the training set and assign its label to the test appliance. Notwithstanding, the nearest neighbour may not be particularly representative of its class. Thereby, increasing k, the number of neighbours used to decide the label of a new device, would mitigate the misclassification occurrences.

Another method used in machine learning for classification is Decision Trees, also known as Classification and Regression Trees (CART), that sort the instances from the root to a leaf node that provides the label for the new example. Given the training set, a decision tree is built such that each node specifies a test of some attribute and each descending branch corresponds to one possible

value for that attribute. A label is assigned to a new appliance by, starting at the root of the tree, testing the attribute condition at each node, and moving downward in the tree accordingly, until a leaf is reached. The optimal partitioning of the training data in order to grow the decision tree is a NP-complete problem and, therefore, a greedy procedure is usually employed in the common implementations of CART (Murphy, 2012).

The Multiclass AdaBoost is a multi-class classification technique based on the Boosting method for solving binary classification problems via, for example, the AdaBoost algorithm (Webb, 2000). The Boosting method is based on the observation that finding many rough rules (weak learner or base learner) is less difficult than finding a single highly accurate prediction rule (Schapire, 2001). The boosting algorithm assigns different importance weights to different training examples. First, equal weights are given to all the training examples. In each round of the algorithm these are updated: more weight is given to training examples hard to classify, while easier training examples get lower weights. After many rounds, the generated weak prediction rules are combined by the boosting algorithm into a single prediction rule, which is expected to be more accurate than the single base learners.

Probabilistic approaches are also widely used for classification, as the Naïve Bayes classifiers (Murphy, 2012). Let $v \in \{1, \ldots, K\}^D$, where K is the number of values for each feature and D is the number of features, be the new load to classify, described by discrete features. The probabilistic models determine the label of a load v using the probability of v belong to a given class $\bar{c}$, $p(\bar{c} \mid v)$. This probability can be determined by the Bayes' theorem

$$p(\bar{c} \mid v) = \frac{p(\bar{c})\, p(v \mid \bar{c})}{p(v)}, \tag{2.12}$$

where $p(v)$ is the probability of a randomly selected load be described by the feature vector v and $p(\bar{c})$ is the probability of a randomly chosen load belong to the class $\bar{c}$. For the specification of the class conditional distribution, $p(v \mid \bar{c})$, this approach assumes that the features describing the load v are conditionally independent given the class. The class conditional density is the product of one dimensional densities, leading to

$$p(v \mid \bar{c}) = \prod_{j=1}^{D} p\left(v_j \mid \bar{c}\right). \tag{2.13}$$

As a consequence, the searched probability is given by

$$p\left(\bar{c} \mid v\right) = p\left(\bar{c}\right) \prod_{j=1}^{D} \left(v_j \mid \bar{c}, \right), \tag{2.14}$$

where an estimation of the $p\left(\bar{c}\right)$ can be computed considering the training set (Silva and Ribeiro, 2009). If instead of discrete features, the instance v to be classified is described by continuous features, the Naïve Bayes classification uses Gaussian distributions (Gaussian Naïve Bayes), *i.e.*, each feature is defined by a Gaussian probability density function (Murphy, 2012).

The Naïve Bayes classifier was also studied for energy disaggregation by Marchiori et al. (2011). The probabilistic approach described computes the most likely state of each device given the measured aggregate signal, and the detected state changes. This classifier assumes that the states of each device are completely independent. For some appliances as TV sets and DVD players this is not true. In addition, the load disaggregation is performed at circuit-level, rather than at whole-home consumption, since the number of devices associated with a circuit is small. Therefore, the number of devices with 'overlapped' features should also be smaller.

Another successfully applied method for energy disaggregation is the Support Vector Machine (SVM), a kernel-based model. The SVM is a decision machine, primarily developed to solve binary classification problems, which predicts labels of input examples based on projection of the examples into a decision hyperplane. The usual notation considers the targets (appliance labels) $t_n \in \{-1, 1\}$ associated with the examples $v_n, n = 1, \ldots, \bar{N}$ (appliance signatures) in the training set. In the simplest case, the training data is linearly separable and a linear model

$$y\left(v\right) = w^T v + \tilde{b}, \tag{2.15}$$

can be used. In this case, there exists, at least, one hyperplane (a weight vector w_0 and a $\tilde{b}_0$) that correctly defines the model. Thereby, the function defined by Equation 2.15 satisfies $y\left(v_n\right) > 0$ for instances with target $t_n = 1$ and $y\left(v_n\right) < 0$ for instances with $t_n = -1$ so that $t_n y\left(v_n\right) > 0$ for each pair $\left(v_n, t_n\right)$ in the training set. From the several solutions to separate the classes that satisfy the above conditions, the SVM searches for the hyperplane that will have the largest margin ρ, *i.e.* the smallest distance between the decision boundary and any of the training examples (Bishop, 2006). The training examples that have an exact

distance of ρ to the hyperplane are called Support Vectors (SV). The margin is defined as the perpendicular distance between the closest point v_n of the dataset and the defined hyperplane. Thereby, the maximum margin solution is found by solving

$$\underset{w,\tilde{b}}{\operatorname{argmin}} \left\{ \frac{1}{\|w\|} \min_n \left(t_n \left(w^T v_n + \tilde{b} \right) \right) \right\}. \tag{2.16}$$

An equivalent optimization problem is

$$\text{minimize } \frac{1}{2} \|w_n\|^2, \tag{2.17}$$

subject to:

$$t_n \left(w^T v_n + \tilde{b} \right) \geq 1, \quad n = 1, \ldots, \bar{N}. \tag{2.18}$$

In order to solve this optimization problem, Lagrange multipliers $\alpha_n \geq 0$ are introduced and the solution has the form

$$w = \sum_{n=1}^{\bar{N}} \alpha_n t_n v_n \tag{2.19}$$

and

$$\tilde{b} = t_{SV} - w.v_{SV} \tag{2.20}$$

where w are the weights, $\alpha \neq 0$ only for the support vectors and $\tilde{b}$, the bias, can be determined using an arbitrary v_{SV} and its correspondent target t_{SV} (Silva and Ribeiro, 2009).

The SVMs can be easily generalized for the nonlinear case (Silva and Ribeiro, 2009). The training data is mapped by $\bar{\phi}(v)$ into a higher-dimensional feature space V'. In general, the computation of such mapping is not efficient. However, during the training and test it is only required to compute dot products in the feature space, $\bar{\phi}(v_n) \cdot \bar{\phi}(v_{n'})$, which can be replaced by a kernel function, a real-valued function with two arguments $\kappa(v_n, v_{n'}) \in \mathbb{R}$, symmetric and non-negative. This is known as the *kernel trick*. As a consequence, at the test time, the prediction for a new instance v is

$$\hat{y}(v) = sign \left(\sum_{n=1}^{\bar{N}} \alpha_n t_n \kappa(v_n, v) + \tilde{b} \right). \tag{2.21}$$

In the particular context of NILM, SVMs were applied to classification of appliances' signatures composed by harmonic signatures and low frequency features and composed by current signals, as reported in (Srinivasan et al., 2006; Kato et al., 2009), respectively. SVMs are also proposed for load classification by Lai et al. (2013) as part of an hybrid approach comprising a Gaussian Mixture Model (GMM). The SVM is used to classify the extracted power features of the different appliances, while the GMM, an extension of a single Gaussian probability density function, is employed to describe the distribution of the current waveforms in order to find power similarity.

The Artificial Neural Networks (ANNs) have also been used in the energy disaggregation context. ANNs are inspired by the biological systems and how they process information (Bishop, 2006). In a similar manner to the brain, an Artificial Neural Network (ANN) is a densely interconnected processing structure whose elements called units (neurons) determine the behaviour of the network. The most usual setting of an ANN is a feedforward neural network with one hidden layer (Figure 2.12) (Silva and Ribeiro, 2009) where the information flows only in the forward direction, *i.e.*, from the input nodes to the output nodes, through the hidden nodes. Within the NILM framework,

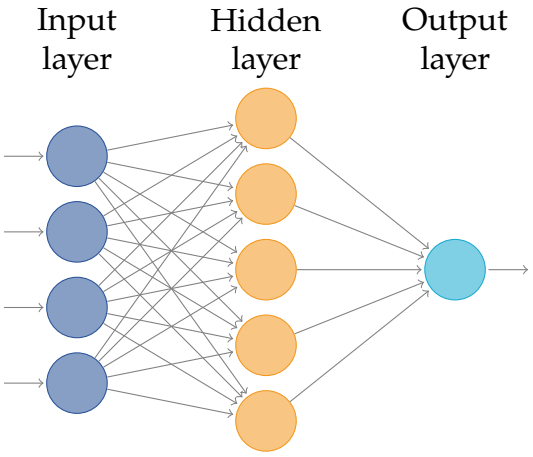

Figure 2.12.: Illustration of a feedforward neural network with one hidden layer.

classification of a new load consists of feeding the ANN with the vector of extracted features - the signature - via the input units forward through the hidden layers. In the hidden layers, the inputs are weighted, summed and transformed by an activation function, usually non-linear, in each layer, before processing takes place in the next layer. The output layer determines the identification of the new load. The most popular learning technique for a neural

network is backpropagation. In this method, the training examples are fed into the network as described. The network output is compared to the target, the error is then calculated and is backpropagated by the network in order to adjust the network parameters leading to a reduction or elimination of the error.

Srinivasan et al. (2006); Ruzzelli et al. (2010) studied ANNs for load classification. In particular, ANN handles any type of data, is easy extensible to a higher number of inputs, can be improved using feedback from the user and can handle with multiple simultaneous appliances states. Still, the training step can be time consuming (Ruzzelli et al., 2010). Nowadays, with Graphics Processing Units (GPU) the processing can be fasten up and easily applied in the NILM framework (Lopes, 2013).

State transition of devices are also handled by the Hidden Markov Models (HMMs). HMMs are a statistical tool based on Markov chains which build probabilistic models for sequences of observations, $x = [x_1, x_2, \ldots, x_T] \in \mathbb{R}^T$, of length T. A Markov chain assumes that $x_t, t = 1, \ldots, T$ captures all the relevant information for predicting the future and then, considering discrete-time steps, the joint distribution, known as Markov chain, is

$$p(x_{1:T}) = p(x_1)\, p(x_2 \mid x_1)\, p(x_3 \mid x_2) \ldots p(x_T \mid x_{T-1}) = p(x_1) \prod_{t=2}^{T} p(x_t \mid x_{t-1}). \tag{2.22}$$

When the observed variables are also discrete, $x_t, \in \{1, \ldots, K\}$, where K is the number of possible discrete values, it is called of discrete-state Markov chain. The Hidden Markov Model (HMM) consists of a discrete-time, discrete-state Markov chain with hidden states $z_t \in \{1, \ldots, K\}$ together with an observation model $p(x_t \mid z_t)$ (Murphy, 2012). In this extension of Markov chains the observation is a probabilistic function of a state. The model has a non-observable stochastic process that is 'accessible' via another set of stochastic processes that produce the sequence of observations (Rabiner, 1989).

For energy disaggregation, given the aggregated power readings, $\bar{x}(t), t = 1, \ldots, T$, (the discrete sequence of observations), it is required to determine the sequence of operational states $z^i = z_1^i, \ldots, z_T^i$ of each appliance $i \in \{1, \ldots, k\}$ which corresponds to turn on or turn off event (Parson et al., 2011). The suitability of the HMM for solving the load disaggregation problem was illustrated in (Zia et al., 2011; Parson et al., 2011). Their ability to handle temporal data (daily operational schedule), as well as the information about state transition of

devices, leads to the good performance of these approaches. However, HMMs have some drawbacks such as (i) the exponential complexity explosion associated with the increasing number of appliances to identify and (ii) the model re-training that is needed when a new group of appliances is added (Zoha et al., 2012).

As the different recognition algorithms have different performance accuracies, depending on the different type of features used, a multi-feature and multi-algorithm disaggregation framework is proposed in (Liang et al., 2010a,b). In Liang's work, the features such as the changes in the P-Q plan, the current waveform and also the harmonics, among others, are extracted (Section 2.5.3). The processing takes place both by optimization algorithms, such as integer programing and genetic algorithm, and by pattern recognition methods, such as ANNs. The final outcome on the load identification is computed by a committee decision mechanism. The committees are based on the opinion of several experts (*e.g.*, classification methods) before the decision about a given task (load identification) is made. The decision is then found by combining the outputs of the several experts by a certain function. Usually a voting strategy like the majority voting or weighted majority voting is adequate although more sophisticated schemes may be used (Silva and Ribeiro, 2009).

Note that a meaningful comparison between the described methods is not possible since their performance is highly correlated with the features being extracted and also with the type and number of appliances to be disaggregated in the experiments (Zoha et al., 2012).

2.6.2. Unsupervised Approaches

In order to find a NILM solution with practical installation and minimum setup cost, unsupervised learning methods started to be explored recently. They try to achieve the load disaggregation without the need of *a priori* knowledge, being an interesting alternative to the supervised approaches which require model training datasets that are not practical to collect. Bearing this in mind, Gonçalves et al. (2011) propose an unsupervised approach. The work presented therein describes the appliances in terms of step changes in the active and reactive power. Given the aggregated signal, these features are extracted and clusters are computed in the P-Q plane by Genetic k-Means and by Hierarchical Agglomerative Clustering. The former method is initialized with randomly

clustering solutions which are evolved and evaluated via a fitness function based on Euclidean distance between points in the same cluster. The latter seeks to build a hierarchy of clusters, where initially each observation defines its own cluster, and at each step of this bottom-up approach the most similar pairs of clusters are merged. The authors state that the better results were obtained with Genetic k-Means. From this step forward, the number of appliances included in the aggregated load is determined due to the assumption that each cluster corresponded to one appliance state-transition. The last step, a matching pursuit algorithm is used for source reconstruction. At each iteration, this greedy algorithm tries to match the new event with the closest source (the centroid of each defined cluster), in terms of Euclidean distance, in order to determine which sources are active. As highlighted by the authors, small and similar appliances with low consumption are not correctly disaggregated. Furthermore, some limitations were found for multi-state appliances for whose several states form distinct clusters leading to misclassified appliances. Still, large devices are correctly identified.

HMMs and its variants were also explored in an unsupervised manner for the disaggregation of signals acquired at low frequency (Kim et al., 2011). In the illustrated approach, each appliance is described by its real power consumption in addition with other useful features as duration of the on and off periods, time of the day, day of the week and dependencies between appliances since, for instance, a gaming console is not used without a TV set. In the defined probabilistic model, the hidden variables are the states of appliances and the observation corresponds to the aggregated load. Thereby, at an instant of time t the aggregated consumption $\bar{x}(t)$ depends on the states of the several appliances. The model seeks for the combination of appliances states that better describe the observation. The base model is a Factorial Hidden Markov Model (FHMM) which models multiple independent hidden states at each instant of time t (Figure 2.13), *i.e.*, for a given sequence of observations $x_1, \ldots, x_T$, $q = \{q^1, q^2, \ldots, q^k\}$ represents the set of k (number of devices) state sequences $z^i = z_1^i, \ldots, z_T^i$. To incorporate additional features, as time of day, or day of week, and dependencies between appliances, the model was extended to a Conditional Factorial Hidden Markov Model (CFHMM), where the transition probabilities are not constant but are conditioned by the extra features. The computational experience using real-world data shows that the inclusion of the dependencies between appliances in particular leads to an improvement of

performance regarding the energy disaggregation. On the other hand, as the number of appliances increases, this performance decreases. Only on and off operations were explored and no method to estimate the number of appliances to disaggregate is presented.

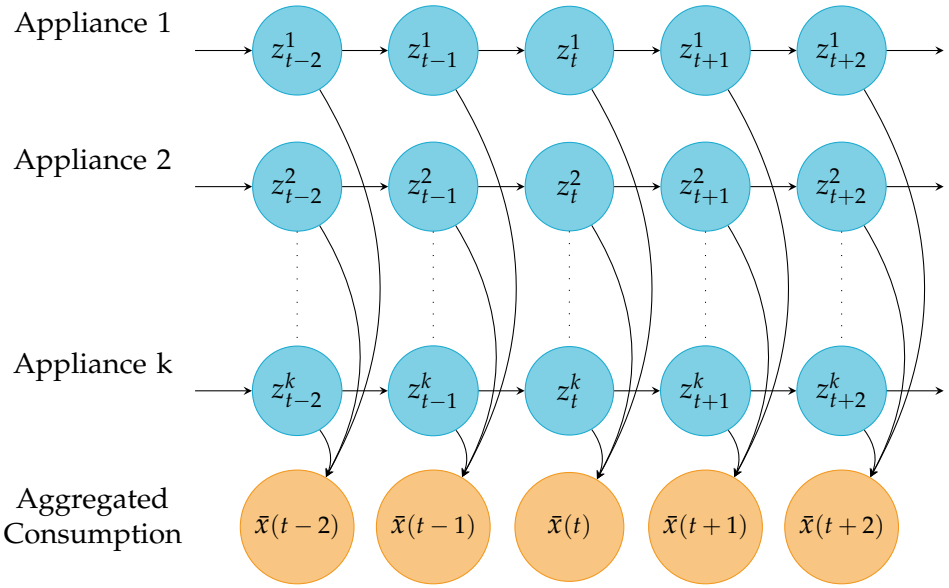

Figure 2.13.: Illustration of FHMM.

Although the HMM is a powerful technique, the methods for the inference of hidden states are often affected by local minima. To overcome this limitation, a convex quadratic programing relaxation of the inference problem is proposed by Kolter and Jaakkola (2012). This method was employed to infer the hidden states of a FHMM in order to solve the unsupervised load disaggregation and separate the several appliances from the aggregated data. In this work the authors follow the original FHMM concept which considers that the output is an additive function of all the hidden states. The comparison to competitor methods has shown that the new approach is more suitable for this problem in terms of precision and recall at circuit level. The recall measures which portion of energy (of a given circuit) is correctly classified, while the precision measures, from the overall energy assigned to a circuit, the portion of energy correctly assigned (*i.e.*, truly belonging to that circuit). In particular, the method achieved a good performance for some of the studied appliances however, for the groups 'electronics' and 'kitchen outlets', both metrics attained less than 50%.

Another extension of the HMM was recently proposed by Johnson and Willsky (2013). The presented technique, an explicit-duration Hierarchical Dirichlet

Process Hidden Semi-Markov Model (HDP-HSMM), is able to learn complex sequential data and, when applied for unsupervised energy disaggregation, intends to overcome some limitations found in previous works (Kim et al., 2011; Kolter and Jaakkola, 2012). In particular, details about the device models are learned during the inference process employing 'simple' prior information, as the rough power draw levels and duration statistics of the modes for several devices, rather than previously learned. In addition, the method allows for more than more only two states therefore the method is not limited to on/off appliances. The method has shown favourable accuracies (defined by Equation 2.32) when compared with Expectation-Maximization based methods. The Expectation-Maximization algorithm finds the maximum likelihood estimates of parameters in statistical models, where the model depends on unobserved latent variables. Thereby, Expectation-Maximization based methods fix the device models parameters, hence, they may not be consistent across all the houses, while the method described in (Johnson and Willsky, 2013) is flexible at this point.

2.7. Disaggregation Evaluation Metrics

This section reviews the disaggregation metrics for performance assessment. The diversity of metrics found in the related literature covers a broad range of evaluation methods. Regardless the fact that almost every NILM approach report a distinct measure of its performance, no meaningful comparison can be drawn since different datasets were tested. Furthermore, no agreement about the most suitable metric exists (Zeifman and Roth, 2011; Zoha et al., 2012). The performance of load disaggregation approaches is intrinsically related to the type of appliances used in the experiment and to the hardware used to collect the load data (Zoha et al., 2012).

Initially, the events correctly identified and the percentage of time that a given appliance has been used were representative of the approach performance (Hart, 1992). Later work of the same research group proposes a comparison between the measured and estimated power, although without describing precise metrics or results (Norford and Leeb, 1996). Similarly, Cole and Albicki (1998a) evaluate their approach regarding the percentage of power consumption correctly associated with each device in study, which can be enriched with the error associated with the load duration and time intervals of the 'on' operation

		True label	
		Positive	**Negative**
Assigned label	**Positive**	True Positives (TP)	False Positives (FP)
	Negative	False Negatives (FN)	True Negatives (TN)

Table 2.1.: Contingency matrix for binary classification.

modes (Farinaccio and Zmeureanu, 1999).

Within the related literature, the selection of the most suitable metric depends on how the NILM problem is being solved. For instance, Srinivasan et al. (2006) and Berges et al. (2009) regard it as a classification problem and, therefore, the performance evaluation was reported as classification accuracy:

$$Accuracy = \frac{\text{Correctly Assigned Events or Signatures}}{\text{Total of Events or Signatures in Study}}, \tag{2.23}$$

i.e., the fraction of appliances being correctly associated with their classes. The performance of a binary classification model is often assessed by the accuracy, precision, recall and F-measure metrics, expressed as percentages (Sokolova and Lapalme, 2009). These are based on a contingency matrix (Table 2.1), also known as confusion matrix. The accuracy, defined by

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}, \tag{2.24}$$

provides only global information. In fact, high values are not related to correct identification of the positive examples, especially in unbalanced data. In imbalanced data, the accuracy may report an almost perfect classification as a reflection of a poor one. For example, in a binary classification scenario where the number of negative examples is remarkably larger than the positive ones, if all examples are only assigned to the negative class, the accuracy value will be close to 1. Nevertheless, no positive examples were detected. This means that despite the high accuracy, a bad classifier worst than random is found since no element of the positive class was identified. To overcome this issue, more detailed performance metrics are the

$$Precision = \frac{TP}{TP + FP} \tag{2.25}$$

and the

$$Recall = \frac{TP}{TP + FN}. \tag{2.26}$$

The first is the ratio of True Positive (TP) in the universe of the all examples assigned as positive (emphasis on False Positive (FP)), while the second is the ratio of TP in the universe of all positive examples in the dataset (emphasis on False Negative (FN)). Recall has also shortcomings since, if a classifier assigns to all examples the same positive class, no FN exists and then its value would be maximum. The trade-off between recall and precision is provided by the F-measure, also known as F_β or F-score,

$$F_\beta = \left(1 + \beta^2\right) \frac{Precision \times Recall}{\beta^2 \times (Precision + Recall)}. \tag{2.27}$$

Usually $\beta = 1$, *i.e.*, F_β is the harmonic mean of the precision and recall. This metric was selected to report the performance of the approach described in (Zeifman, 2012). These metrics will be further explored in Chapter 3 Section 3.3.1 concerning the appliance classification.

Beyond classification accuracy, recall and precision, an approach can be evaluated regarding its event detection performance if it is considered relevant for identification. An event can be falsely detected, a false positive, or missed, or a false negative. Therefore, the accuracy should consider the number of true, wrong and missed events, as illustrated by Liang et al. (2010a). In the same direction, metrics as True Positive Rate (TPR)

$$TPR = \frac{TP}{TP + FN}, \tag{2.28}$$

which is the recall value also known as sensitivity, and the False Positive Rate (FPR)

$$FPR = \frac{FP}{FP + TN}, \tag{2.29}$$

which provides the ratio of misclassified negative event examples on the universe of all negative event examples in the dataset, based on the confusion matrix, were employed for evaluation of event detection by Anderson et al. (2012a). The metric defined as the trade-off between TPR and FPR is a Receiver Operating Characteristic (ROC) curve also suggested by Zeifman and Roth (2011). However, note that none of the described metrics provides information about the accurateness of the power consumption computation for each

appliance.

More general metrics are (i) the percentage of energy correctly identified (Hart, 1992); (ii) the distance between the predicted energy and the energy actually consumed by the appliances

$$\text{Disaggregation Error} = \sum_{i=1}^{k} \frac{1}{2} \left\| x_i - \hat{x}_i \right\|^2, \tag{2.30}$$

where x_i is the vector of length T of the measured consumption for device i, $\hat{x}_i$ is its predicted version and thus, it provides a global measure of the distance between the prediction and the measured consumption (Kolter et al., 2010); and (iii) the normalised error in the total energy assigned to an appliance over all days

$$\text{Normalized Disaggregation Error} = \frac{\left| \sum_{t=1}^{T} x_i(t) - \sum_{t=1}^{T} \hat{x}_i(t) \right|}{\sum_{t=1}^{T} x_i(t)} \tag{2.31}$$

where $x_i(t)$ is the power consumed at time t by device i and $\hat{x}_i(t)$ is the estimated power consumption (Parson et al., 2012). In addition, accuracy defined as the average (or total) amount of energy predicted correctly over a given period of time

$$Accuracy = 1 - \frac{\sum_{t=1}^{T} \sum_{i=1}^{k} \left| \hat{x}_i(t) - x_i(t) \right|}{2 \sum_{t=1}^{T} \bar{x}(t)} \tag{2.32}$$

where $\bar{x}(t)$ refers to the measured total power consumption at time t, can also be computed, as proposed by Kolter and Johnson (2011); Johnson and Willsky (2013). For an overview of the error for each appliance in each instant of time, the root mean square error

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^{T} \left(x_i(t) - \hat{x}_i(t) \right)^2}. \tag{2.33}$$

can be additionally computed (Parson et al., 2012).

In short, no global or consensual performance metric is known for the energy disaggregation problem. The selection of the most appropriate one seems to depend on the approach being followed for the problem resolution. Although it is consensual that classification approaches can be evaluated employing common metrics as accuracy, nevertheless, these do not allow for the assessment of the electrical consumption assigned to each device. Metrics concerning the distance between the measured and predicted amount of consumption should also be applied for a more complete evaluation.

2.8. Summary

In this chapter, energy disaggregation problem and NILM systems were introduced and a review of the state-of-the-art was presented. First, a brief introduction to concepts related to the electrical power in alternating current and its particularities, necessary for the description of specific details within a NILM framework was provided. Next, the energy disaggregation problem was formally described and the original NILM algorithm and framework was surveyed. A review was presented for each main step of a NILM system, data acquisition, feature extraction and appliance electrical signatures, and model learning and classification for load identification.

Regarding the data acquisition process, the chapter addressed the relation between the sampling rate and the appliance signatures, the sensors required and the known and freely available datasets for on the energy disaggregation problem. In the sequel, the background and research on the electrical signatures of appliances, in particular steady-state and transient signatures, was revised. The former are deduced from the step changes in a signal while the latter are composed by features extracted from the transitory period between two steady-states. The alternative signatures found in the related literature were also addressed. According to Zeifman and Roth (2011), the main research effort has been focused on signature exploration, nevertheless, it remains to identify a robust set of features able to correctly describe all types of signals, for low-power appliances in particular. To overcome this issue, non-power features may be combined to the previously to power characteristics to allow a more accurate disambiguation (Zoha et al., 2012).

The research on model learning and classification process for load identifica-

tion was also surveyed. Both supervised and unsupervised approaches can be found in scientific works. The main focus of NILM research has been directed to supervised strategies. Still this type of approaches requires a training dataset, composed by information of each device in the house electrical network. In addition, the turn on and off events may also be needed. The collection of this dataset can be expensive and unpractical due to the increasing number of appliances that can be found in a house nowadays (Anderson et al., 2012b). Thereby, the research concerning unsupervised methods, which eliminates the need of using labelled data for building the learning models, has been raising.

This chapter ended with a discussion on the main disaggregation metrics usually employed to assess the performance of the methodologies. In the related literature no global or consensual performance metric for the energy disaggregation problem was found. The selection of the most appropriate performance assessment method seems to depend on the approach developed.

CHAPTER 3

Steady-State Signature Recognition

3.1. Introduction

Steady-state signatures for non-intrusive energy disaggregation are the focus of this chapter. The 'classical' approach regards NILM disaggregation task as a classification problem where appliances' signatures are fundamental. These signatures should be recognized in order to perform an accurate appliances identification. Each one of these signatures is a distinctive characterization for a device, representing its nature and state of operation. Among the possibilities, as described in Chapter 2 Section 2.5, steady-states based identification,

the fundamental analysis in particular, has been studied in several research works mainly due to its simplicity and undemanding requirements in terms of sampling rate and specific hardware for data collection.

A device characterization, based on the analysis and recognition of steady-states occurring in the active and reactive power signals and the power factor measurements, is next proposed and explored. The definition of such a signature requires the identification of segments of signals named as steady-states. For the recognition of a steady-state, a mathematical rule based on the ratios between rectangular areas, which is then further simplified, is proposed.

The NILM approach described in this chapter investigates two learning techniques, SVM and K-NN, for the classification of the extracted steady-state signatures. In order to evaluate the proposed approach, two experimental sets are drawn. First, data from a set of devices is collected considering that only one appliance in the electrical circuit is switched on, operating and switched off after a given amount of time. Second, data is gathered using two pairs of devices, from the previous dataset, connected simultaneously to the network. In this particular case, Gaussian white noise is also added to the set of signals collected that are then cleaned using the Wavelet Shrinkage method (Donoho and Johnstone, 1994; Donoho, 1995; Donoho and Johnstone, 1995).

This chapter is organized as follows. Next section presents a new mathematical result that enables steady-state recognition (MinMaxSteady-State), the features used as distinctive signatures of appliances in the proposed approach, and the classification methods that will be explored for event classification. Then, Section 3.3 describes the experimental setup for both experiences, where the MinMaxSteady-State rule is used for the identification of steady segments regarding the extraction of relevant step changes to be classified using the SVM and K-NN methods, as well as the computational results analysis and discussion. Finally, conclusions and lines of future work are pointed out.

3.2. Steady-States Recognition for Energy Disaggregation

This section introduces a rule for steady-state recognition, the electrical features proposed to define a steady-state appliance signature, and the pattern recognition methods used to identify loads characterized by this specific steady-state

signature. When a set of consecutive samples has values within a given threshold it is considered to be a stable state. The difference between two of these consecutive stable states defines a steady-state signature for a given device. These steady-state changes, known as events, nominally correspond to the load either turning on or turning off and are characterized by their magnitude and signal, *e.g.* on real and reactive powers. For example, when an appliance is turned on, an increase in the power consumption occurs with a magnitude that will correspond to the consumption of the given device. In particular, the steady-state characterization permits the recognition of (i) turning off states and (ii) simultaneous events. In fact, being additive in nature, makes it possible to verify whether two appliances are switched on at the same time. Thus, steady-state signatures were chosen and explored by several authors (Zoha et al., 2012), mainly for residential load monitoring systems. Nevertheless, as mentioned in the previous chapter, this characterization has disadvantages: two different appliances may have similar steady-state signatures making it impossible to distinguish them just from the recognition of the steady-state signature, as illustrated in Figure 2.8 in Chapter 2. The small sampling rate can also be considered a disadvantage since a sequence of turning on a load and turning it off in a smaller period than the sampling rate makes the identification impossible.

The first step in the definition of a steady-state characterization is the recognition of a stable value sequence in the sampled signal. In the following, a rule is proposed for the identification of stable segments of a signal. The rule tests whenever a new sample belongs to a posterior identified stable state. The approach is initially based on the difference between the rectangular area produced by aggregating a new sample and the one already defined by the previous values in the stable state. However and, as next proved, the approach can be simplified by keeping only the extreme values already in the stable state and testing the new sample value against the previous ones. This rule is applied for the detection of step changes in power signals, namely, active, reactive powers and power factor measurements, which are proposed as device signatures in this approach. The extracted step changes are further classified by the learning methods SVM and K-NN for device identification.

3.2.1. A Rule for Steady-States Recognition

A sequence of consecutive samples from a signal represents a stable-state if the difference between any two samples of this sequence does not exceed a given tolerance value. Thereby, the definition of two parameters is required: (i) the tolerance value, ϵ, and (ii) the minimum number of consecutive samples necessary to specify a stable-state, s_{min}. The first depends on the input signal and on the required granularity of identification. For instance, if the tolerance is higher than the consumption of small appliances then these will not be identified. The second parameter, the minimum number of consecutive samples needed to identify a stable state, depends on the sampling frequency: when it is low, a small number of samples is enough, otherwise, a larger number is needed. For instance, Hart (1992) considered $s_{min} = 3$ for a sampling frequency of 1Hz. Figure 3.1 illustrates an example where the steady-states are identified with two dashed line segments, while the remaining samples belong to a change state.

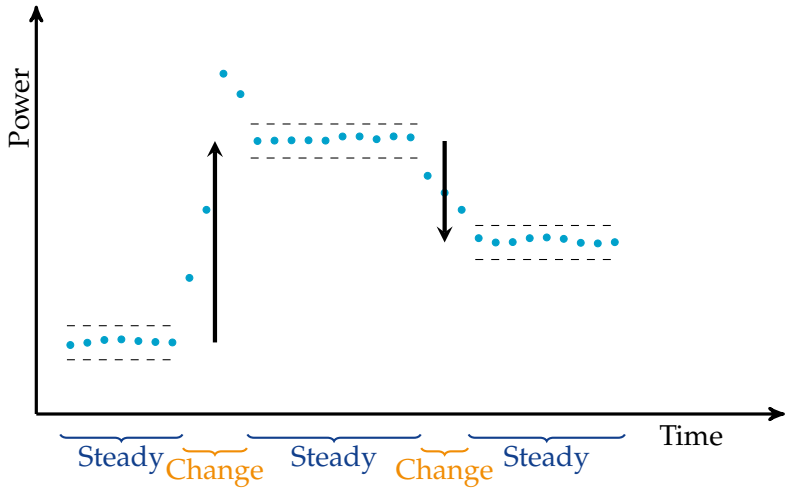

Figure 3.1.: Illustration of three steady-states.

Let's consider a sequence of samples from the measured signal of interest, $x = [x(1), \ldots, x(T)]$, with length T. A steady-state of minimum length, *i.e.*, composed by s_{min} successive values $x(t)$, can be identified verifying if the condition

$$\left| x\left(t'\right) - x\left(t''\right) \right| < \epsilon \tag{3.1}$$

is held for each pair $x(t'), x(t'')$ in the set of s_{min} consecutive samples. A sequence of null sample values is considered as a steady-state on its own. Given

such a minimal steady-state, the inclusion of a consecutive sample could be based on the sum of successive rectangular areas. In detail, for each sample $x(t')$ of a given steady-state, let's consider the rectangular area $\Lambda_{t'}$ whose height equals $x(t')$ and whose width is the difference between t' and its predecessor. In a similar manner and given a new sample at the instant t of value $x(t)$, an area Λ_t for the new sample can be defined. Additionally, let's consider the rectangular area Λ_t^*, whose height equals $x(t)$ and whose width equals the sum of widths of the previous defined rectangular areas. The new $x(t)$ would belong to the sequential steady-state if Λ_t^* is equal (up to a given tolerance value) to the sum of the sequential areas $\Lambda_{t'}$ in addition to Λ_t.

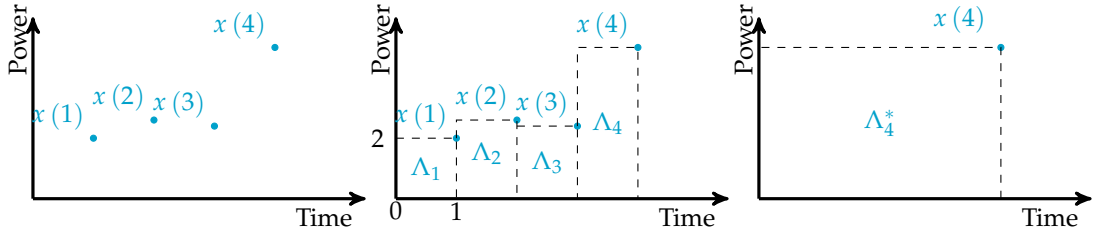

Figure 3.2.: Illustration of the rectangular areas for steady-state recognition.

To clarify, let the discrete signal in analysis be as illustrated in Figure 3.2 for which the samples $x(t), t = 1, \ldots, 3$, define a steady-state and ϵ be the given tolerance. Notice that, for each pair $t, t' \in \{1, \ldots, 4\}, t \neq t'$, if $|x(t) - x(t')| < \epsilon$ the sum of the areas Λ_1, Λ_2, Λ_3 and Λ_4 would be similar to the Λ_4^* area. Formally,

$$\left| \frac{\sum_{t=1}^{4} \Lambda_t}{\Lambda_4^*} \right| \approx 1, \tag{3.2}$$

Hence, a sample point x_4 could be considered as part of the steady-state if Equation 3.2 would be held. Still, notice that for each pair of consecutive samples, $|t - t'|$ is a constant value and this reasoning can be simplified by keeping only the extreme values already in the stable state and testing the new sample value against the previous ones as described in the following.

Let $S^* = \{x(t) : t = t_a, \ldots, t_a + t^* - 1\}$ be a segment of signal with beginning at instant t_a and composed by t^* consecutive sampling values, and already identified as a steady-state. By definition, the value of any pair of samples composing it does not differ more than the tolerance ϵ, *i.e.*,

$$\left| x(t') - x(t'') \right| \leq \epsilon, \;\; \forall\, t', t'' = t_a, \ldots, t_a + t^* - 1, \;\; t' \neq t'', \tag{3.3}$$

where $\epsilon > 0$ is the defined tolerance. Let $x_M = \max\{x(t')\}$ and $x_m = \min\{x(t')\}, \forall\, t' = t_a, \ldots, t_a + t^* - 1$, be the maximum and minimum values, respectively, in S^*. Additionally, let $x(t)$ be the next sample value in study $(t = t_a + t^*)$ for which is possible to define a limited range of values that would preserve the stable behaviour of S^*. The range of possible values is illustrated in Figure 3.3.

Theorem 3.2.1. *In the conditions above, the $t^* + 1$ consecutive values form a steady-state iff $x_M - \epsilon \leq x(t) \leq x_m + \epsilon$, i.e.,*

$$\left|x\left(t'\right) - x\left(t''\right)\right| \leq \epsilon, \;\; \forall t', t'' = t_a, \ldots, t_a + t^*, \;\; t' \neq t''. \tag{3.4}$$

Proof. In fact, if $x_m \leq x(t) \leq x_M$ then $|x(t') - x(t)| \leq |x_m - x_M| \leq \epsilon$, $\forall t' = t_a, \ldots, t_a + t^* - 1$.

Consider now that, $x_M < x(t) \leq x_m + \epsilon$. For any $x(t') \in [x_m, x_M]$, $t' = t_a, \ldots, t_a + t^* - 1$, it is verified that

$$|x\left(t'\right) - x(t)| \leq |x_m - x(t)| \leq |x_m - x_m + \epsilon| = \epsilon. \tag{3.5}$$

Thus, the sequence of the $t^* + 1$ values, $x(t'), t' = t_a, \ldots, t_a + t^*$, composes a steady-state with a new maximum value: $x_M = x(t_a + t^*) = x(t)$.

By a similar reasoning, if $x_M - \epsilon < x(t) \leq x_m$, then, $x(t) = x(t_a + t^*)$ preserves the stability of the state and the steady sequence $x(t'), t' = t_a, \ldots, t_a + t^*$, has a new minimum value: $x_m = x(t_a + t^*) = x(t)$.

Regarding the remaining cases, $x(t) < x_M - \epsilon$ and $x(t) > x_m + \epsilon$, $x(t)$ does not belong to the steady-state S^* since the maximum tolerance value is compromised. Indeed, if $x(t) < x_M - \epsilon$, then

$$x(t) < x_M - \epsilon \leq x_m \leq x\left(t'\right) \leq x_M, \;\; \forall t' = t_a, \ldots, t_a + t^* - 1. \tag{3.6}$$

Thereby, the difference between x_M and $x_M - \epsilon$ is smaller than between $x(t)$ and x_M, *i.e.*, $|x(t) - x_M| > |x_M - \epsilon - x_M|$ equivalently, $|x(t) - x_M| > \epsilon$.

The latter case, $x(t) > x_m + \epsilon$, can be proved similarly. □

In conclusion, a consecutive sample point $x(t)$ belongs to the steady-state immediately before if $x_M - \epsilon \leq x(t) \leq x_m + \epsilon$, such that x_m and x_M are the minimum and the maximum values in that state (MinMaxSteady-State rule). Otherwise, the previous sample, $x(t_a + t^* - 1)$, is considered as the end of the

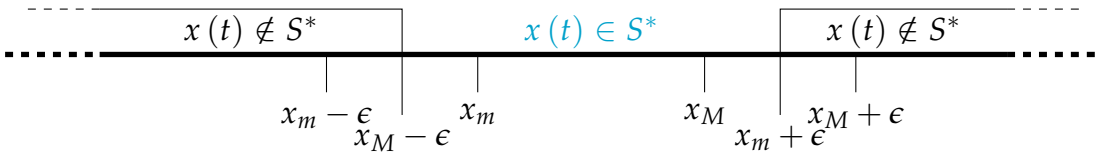

Figure 3.3.: Range of acceptable values for inserting $x(t)$ in a previous identified steady-state S^*.

steady-state. After all the samples have been tested, the differences between consecutive states are computed, representing the signature of on-off appliances.

3.2.2. A Steady-State Signature

Active power signals and their associated steady-state changes are not sufficient to define an unique appliance signature for an accurate appliance recognition, as mentioned in Chapter 2 Section 2.5.1. The signature can be enriched by other electrical parameters, as the step changes in the reactive power and power factor measurements, in addition to the active power information. These three complement each other. In fact, this holds since the active power represents the power actually being consumed, the reactive power measures the power going back from the load to the supply (*i.e.* it does not get consumed) and the power factor is a ratio between the active power and the apparent power. As observed in Chapter 2 Section 2.2 $\cos(\tilde{\phi})$ provides the portion of real within the apparent power.

The amount of each of these powers is related to the type of load in the AC circuit, as presented in Chapter 2 Section 2.2. In practice, loads can be purely resistive (*e.g.* incandescent lamps), motors (*e.g.* pumps) or electronics (*e.g.* battery charges). For resistive loads, the current and voltage waveforms would be in phase and, therefore, only real power would be transferred in the circuit, the reactive power would be 0 and the power factor would be 1. On the opposite side, if the load is purely reactive, these two signals would be out of phase and only reactive energy would exist in the circuit: the active power and power factor would be 0. However, most of the household loads can be characterized as motors and electronics, for which a shift between the current and voltage signals occur and, therefore, both the active and reactive power exist and also the power factor. See an example for a LCD screen 20" in Figure 3.4.

Accordingly, in this Thesis, a signature composed of the step changes occur-

ring in the active, reactive power and power factor is explored. The incorporation of the steady-state changes in the power factor measurements aims at the inclusion of proportional information in the appliance signature in addition to the absolute values provided by the active and reactive power measurements.

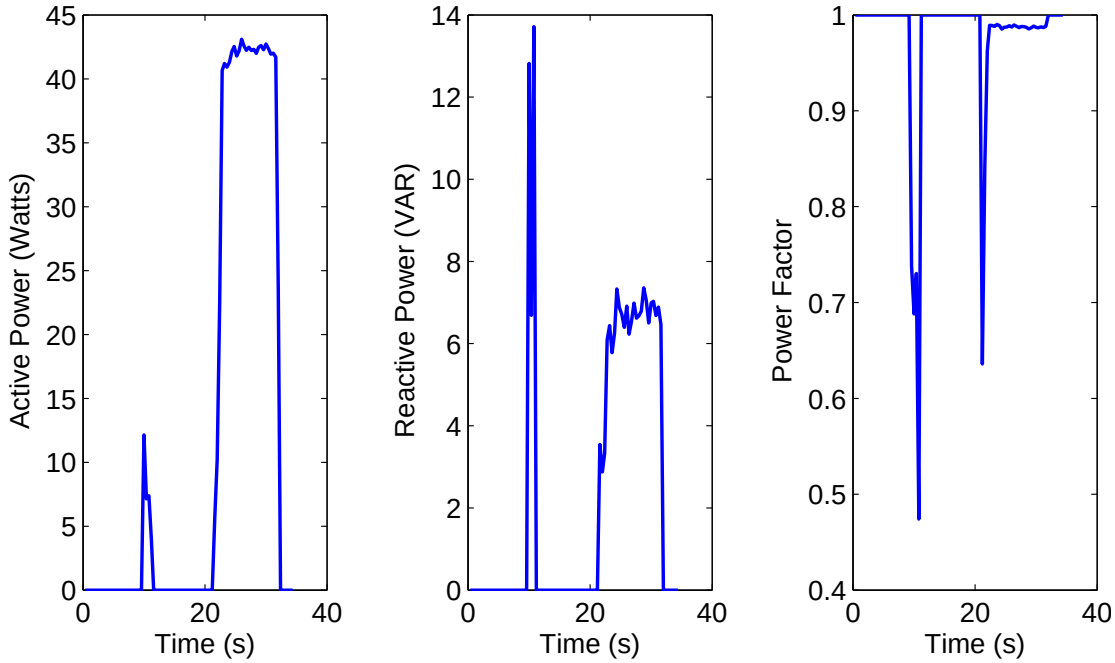

Figure 3.4.: Active, reactive power and power factor for a LCD screen of 20 in.

3.2.3. Classification Models for Steady-State Signature Recognition

In this subsection the basics of models for identification of device signatures are introduced. After the extraction of features in accordance with the defined steady-state signature (the step changes on real and reactive powers and power factor) the next step is appliance identification. The appliances are identified via a classification model. The main goal of the classification problem is to learn a model able to correctly assign an appliance class label to a given feature vector (appliance signature). In this Thesis, these extracted features are classified by a baseline and a kernel method, namely, the K-Nearest Neighbours (K-NN) and Support Vector Machines (SVMs). The former has achieved good performance for the classification of appliances characterised by the changes in real and reactive powers and by the shape of their transient profile (Berges et al., 2011),

while the latter was explored for the classification of harmonic components from the current waveform (Srinivasan et al., 2006), as revised in Chapter 2 Section 2.6.1. In the following, these methods are briefly described.

K-Nearest Neighbours

K-Nearest Neighbours (K-NN), as described in Chapter 2 Section 2.6.1, is a simple example-based method for classification, usually employed to build a baseline against other methods (Silva and Ribeiro, 2009). Instead of building a representation for each appliance category, this method finds the nearest example in the training set and uses its class to label the test appliance. This simple label assignment has a particular drawback: the nearest neighbour may not be especially representative of its class. In order to overcome it, the number of neighbours used to decide the label of a new device can be increased to k, decreasing then the misclassification occurrences.

The K-NN classification algorithm calculates the dissimilarity between the test device signature and all the appliances signatures in the training set (Barber, 2012). The dissimilarities are ranked and the most frequent label of the k most similar neighbours is then associated with the new device. The method's simplicity is an advantage, nevertheless, there are two main issues. The commonly used dissimilarity function is the Euclidean distance, yet not always the most suitable one. This metric does not consider how data is distributed which can be a drawback. Another issue is related with the algorithm efficiency. No true training phase exists for the K-NN, no representation is built and, then, at each appliance classification, dissimilarities must be calculated and ranked, which can be expensive.

Support Vector Machines

Support Vector Machines (SVMs) are a popular example of a kernel-based model. These approaches map the data into a higher-dimensional feature space. However, rather than computing the mapping of a feature vector, the learning algorithms are kernelized, *i.e.*, the algorithm is modified to account for linearity separation in the feature space. As mentioned in Chapter 2 Section 2.6.1, the existent dot products of feature vectors are replaced by the kernel function κ. Usually, these functions are symmetric and non-negative which can be interpreted as a measure of similarity. This adjustment is known as the *kernel*

trick.

The SVM was primarily developed to solve binary classification problems (Cortes and Vapnik, 1995). However, the device classification is usually a multi-class problem. Still, the linear SVM described in Chapter 2 Section 2.6.1 can be extended to solve multi-classification problems by two main approaches: one-against-all (one-against-the-rest) and one-against-one (pairwise classification) (Hsu and Lin, 2002). Let's consider a set of k different appliances (classes), $\mathcal{A} = \{a_1, \ldots, a_k\}$, and a set of $\bar{N}$ observations, consisting in the pairs $(v_n, t_n), n = 1, \ldots, \bar{N}$, such that v_n is a vector of D features with K elements each and $t_n \in \{-1, 1\}$ is 1 if the observation belongs to the class $a_i, i = 1, \ldots, k$, and -1, otherwise.

In the one-against-all strategy, a binary problem is defined by fixing each class against the remaining ones and k binary classifiers ($k > 2$) are defined. In particular, for a given class a_i, the correspondent i^{th} SVM is trained, considering that all the samples of this class are represented by positive labels and the remaining examples by negative ones. Therefore, the training data is $\{(v_1, t_1), \ldots, (v_{\bar{N}}, t_{\bar{N}})\} \in \mathbb{R}^D \times \{-1, 1\}$, and the associated SVM solves the following optimization problem:

$$\text{minimize}_{w_i, \tilde{b}_i, \xi_i} \quad \frac{1}{2} \|w_i\|^2 + C \sum_{n=1}^{\bar{N}} \xi_i^n, \tag{3.7}$$

subject to:

$$(w_i)^T \bar{\phi}(v_n) + \tilde{b}_i \geq 1 - \xi_i^n, \text{ if } t_n = 1, \tag{3.8}$$

$$(w_i)^T \bar{\phi}(v_n) + \tilde{b}_i \leq -1 + \xi_i^n, \text{ if } t_n \neq 1, \tag{3.9}$$

$$\xi_i^n \geq 0, \; n = 1, \ldots, \bar{N}, \tag{3.10}$$

where v_n is mapped into a higher dimensional space by the function $\bar{\phi}$, w is a normal vector to the hyperplane that separates the positive samples from the negative ones, $\frac{\tilde{b}}{\|w\|}$ is the perpendicular distance from the referred hyperplane to the origin, C is the penalty parameter and ξ_i^n are positive slack variables. This optimization problem corresponds to the maximization of the margin between the two classes by the minimization of $\frac{1}{2} \|w_i\|^2$. For the non-linear separable case, the slack variables are used to penalize the number of training errors

whose upper bound is $\sum_{n=1}^{\bar{N}} \xi$. Solving this problem for each class leads to k decision functions

$$\hat{y}_i(v) = \tilde{b}_i + \sum_{n=1}^{\bar{N}} \alpha_n^i t_n \kappa_i(v_n, v), \quad i = 1, \ldots, k, \tag{3.11}$$

(considering the nonlinear case, the kernel trick as in Equation 2.21, and that $\alpha \neq 0$ only occurs for the support vectors (SV)), which are later used to define the class of a new v. The class with a greater value of its decision function is associated with vector v.

In the one-against-one (pairwise) strategy, each pair of classes is compared by employing $\frac{k(k-1)}{2}$ binary classifiers. For each pair of classes i and j, the SVM optimizes the problem

$$\text{minimize}_{w_{ij}, \tilde{b}_{ij}, \xi_{ij}} \quad \frac{1}{2} \left\| w_{ij} \right\|^2 + C \sum_{n=1}^{\bar{N}} \xi_{ij}^n, \tag{3.12}$$

subject to:

$$\left(w_{ij}\right)^T \bar{\phi}(v_n) + \tilde{b}_{ij} \geq 1 - \xi_{ij}^n, \quad \text{if} \quad t_n = 1, \tag{3.13}$$

$$\left(w_{ij}\right)^T \bar{\phi}(v_n) + \tilde{b}_{ij} \leq -1 + \xi_{ij}^n, \quad \text{if} \quad t_n \neq 1, \tag{3.14}$$

$$\xi_{ij}^n \geq 0, \; n = 1, \ldots, \bar{N}. \tag{3.15}$$

Then, $\frac{k(k-1)}{2}$ decision functions are computed, one for each pair of classes. The class label of a test sample v is assigned by a voting strategy, as suggested in (Friedman, 1996). This "Max Wins" method increases by one the voting of class i, if the decision function

$$\hat{y}_{ij}(v) = sign\left(\tilde{b}_{ij} + \sum_{n=1}^{\bar{N}} \alpha_n^{ij} t_n \kappa_{ij}(v_n, v)\right) \tag{3.16}$$

indicates that v belongs to the ith class, otherwise, the voting of class j is increased by one. The class with the highest number of votes is assigned to the vector v.

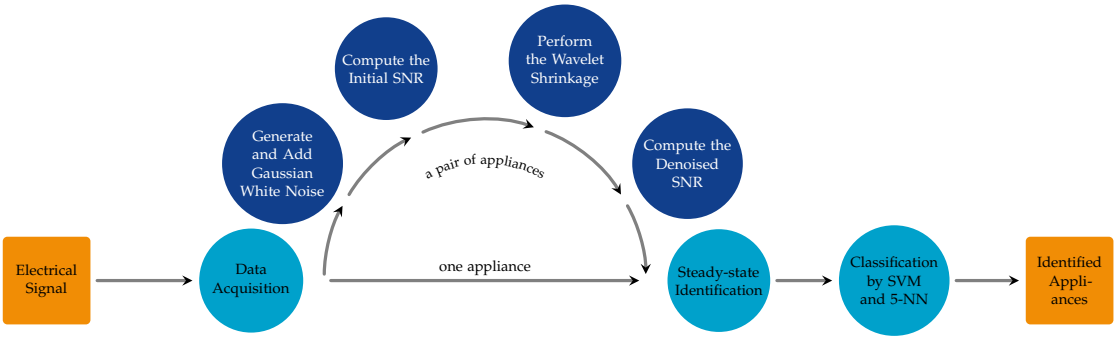

Figure 3.5.: High-level description of the computational experience, considering one appliance (Experience A) and a pair of devices (Experience B) in the circuit.

3.3. Computational Experiments

We evaluate the proposed approach (the MinMaxSteady-State rule, the steady-state signature and the above described learning methods) on acquired real datasets described in what follows. Our experiments in this section comprise two goals: first, to evaluate the effectiveness of the defined steady-state signature based on active and reactive powers and power factor changes; second, to assess the performance of identification when two devices were switched on at the same network. Figure 3.5 illustrates the two experiments performed considering: first, one device in the network (Experience A); and, second, two devices in the circuit (Experience B). For both cases, data is acquired, steady-state identification is performed, by the MinMaxSteady-State rule, and classification is carried through the kernel method SVM and the baseline technique K-NN, whose outcome is the identification of the devices in study. As the data is collected in a controlled environment, no interference of other appliances occurred. Hence, in Experience B, Gaussian white noise signals with a noise level of 25dB is generated and added to each gathered signal. At last, a denoising method is applied to reduce the interference so that features are extracted.

3.3.1. Experimental Setup

In this section, the designed computational experience is detailed, namely, we describe the datasets, evaluation metrics and experimental free model parameters. The proposed approach was implemented in Matlab software and using LIBSVM (Chang and Lin, 2001) as will be detailed.

Figure 3.6.: Hardware used to collect the datasets: a smart meter, a laptop and the required converter. Apart from the laptop, the apparatus was supplied by ISA (ISA-Intelligent Sensing Anywhere, S. A., 2012).

Datasets Description

The active power, voltage, current and power factor signals were recorded using a sensing meter prototype provided by ISA-Intelligent Sensing Anywhere (ISA-Intelligent Sensing Anywhere, S. A., 2012) illustrated in Figure 3.6. When sampling several signals, this prototype supplies only one sample value of one signal of interest at each point in time. Consequently, a delay between the sample values of different signal types exists in temporal terms. In addition, measurement errors can eventually occur, resulting in the failure of deliverance of the expected value. To overcome this shortcoming, data from electrical appliances were acquired considering a 100 milliseconds delay between the samples of the four distinct signals. Therefore, the time distance between each sample of each type of signal was of 400 milliseconds (a sampling frequency of 2.5Hz).

Measurements of active power, voltage, current and power factor signals were collected for (i) only one appliance in the circuit (Dataset A) and for (ii) a pair of devices plugged into the network (Dataset B) to yield Experiences A and B, respectively. The collection procedure of Dataset A was composed by four steps:

1. the signal samples are acquired without the appliance being plugged in to the socket during 10 to 15 seconds;

2. the device is plugged in and samples are collected for 15 seconds;

3. the apparatus is switched on and it runs for a period of 1 minute;

4. the appliance is switched off followed by a sampling period of 15 seconds.

The process was repeated fifty times for each one of the six appliances illustrated in Figure 3.7: an incandescent bulb of 60W, two LCD's (from the same manufacturer but different models), a microwave, a toaster, and a coffee machine. Note that for the coffee machine, the running time is less than one minute corresponding to the time needed for an expresso. In addition, the machine heating system was ready for making the coffee.

Figure 3.7.: Illustration of appliances used in the drawn experiments.

A similar routine was performed to collect data with a pair of appliances here discriminated: (i) incandescent bulb and 32 in. LCD screen, (ii) microwave and toaster in the same circuit. In this collection process however there are different steps. In this case, the devices are plugged in at the same time and samples are collected for 15 seconds. The first apparatus (*e.g.* incandescent bulb) is switched on and, after 20 seconds, the other (*e.g.* 32 in. LCD screen) is turned on. The samples are then acquired during 1 minute for each appliance. This procedure was repeated twenty five times for the two pairs of appliances considered.Then, two Gaussian white noise signals with a noise level of 25dB were generated and added to each collected signal. Therefore, Dataset B is composed by fifty signals for each pair of devices. Table 3.1 summarizes the monitored household devices and the number of signals available in each dataset.

Evaluation Metrics

The performance evaluation of the binary decision tasks, as described in Chapter 2 Section 2.7, requires the definition of a contingency matrix, which allows the calculation of accuracy, precision and recall metrics. Moreover, the harmonic mean of the precision and recall is provided by the F-measure, also known as F_1 when $\beta = 1$,

$$F_1 = 2\frac{Precision \times Recall}{Precision + Recall}. \tag{3.17}$$

	Dataset A	**Dataset B**
#Signals	250	100
#Class	6	4
Devices	Incandescent bulb of 60W 22 in. LCD screen 32 in. LCD screen Microwave Toaster Coffee machine	Incandescent bulb of 60W and 32 in. LCD screen Microwave and Toaster

Table 3.1.: Description of devices datasets A and B used respectively for Experiences A and B.

Still, in this approach a multi-classification problem is solved. To evaluate these problems, the binary classification metrics are replaced by other assessment measures able to consider the performance of all classes. Indeed, macro-average and micro-average metrics can be calculated, namely the macro-average F_1 and micro-average F_1. The macro-average of the trade-off between the recall and the precision is calculated as the average of the F_1 associated with each class (Silva and Ribeiro, 2009), *i.e.*,

$$macroF_1 = \frac{\sum_{i=1}^{k} F_1^i}{k} \tag{3.18}$$

where F_1^i is the F_1 measure for class i.

The micro-average, a global F-measure, can also be calculated considering global confusion matrix computed from the sum of all the confusion matrices related to the binary problems. To allow the comparison between one-against-all and pairwise classification strategies, k confusion matrices associated with the binary problems one-against-all are determined based on the achieved pairwise classification results. The aforementioned metrics were then evaluated.

To assess the models' performance for the classification of unseen data, the datasets are usually divided in two sets: (i) training data and (ii) testing data. The training partition is used to build the models while the test partition is used to validate the developed models. This strategy for estimating the models generalization performance is known as cross-validation. Among the common types of this statistical method, one can find the K-Fold Cross Validation, which

is used in these computational experiments. In this validation scheme, first, the data is partitioned into K sets of equally sizes (or nearly equal) known as folds. Next, the models are built considering K-1 folds as training set and the remaining fold for testing. In fact, K models are built each with one distinct fold for testing. For example, when $K = 3$ we define folds f_1, f_2 and f_3. First, f_1 and f_2 are considering as training set and f_3 as test set. In the second run, f_1 and f_3 are used as training set while f_2 is set as test partition. Lastly, the model is built using the samples in the folds f_2 and f_3 and the test is performed with f_1. The overall model performance is reported in terms average of all the experiments.

The performance of the different learning methods in these experiments is accessed by 3-Fold Cross-Validation whose folds are a representative subset of the original dataset, *i.e.*, the 3 folds contain approximately the same number of samples *per* class.

Experimental Parameters

Parameter selection plays a central role in model building in machine learning and pattern recognition. The main idea of parameter selection is to choose a subset of relevant parameters for building robust learning models. The importance of parameter selection lies on their potential to facilitate data visualization and data understanding, to reduce measurement and storage requirements, to decrease training times and to defy the curse of dimensionality in order to improve performance prediction. In this context, in the following, the parameter selection namely for K-NN and SVM is presented.

For the load identification by the lazy learning model K-NN, the only parameter to set is the number of neighbours required to decide the label of a new appliance. As mentioned, considering only a neighbour may lead to misclassification since this particular neighbour may not be the best representative of its class. On the other hand, if a large number of neighbours is considered, classes with more examples in the dataset will overwhelm the other. For these experiments, and after trials on the datasets, K is set to 5.

The other classification method explored, SVM, requires the setup of the kernel function and associated parameters, in addition to the regularization constant C. The proposed approach is evaluated considering two different kernels: the linear kernel and the RBF kernel. The simplest kernel function, the

linear kernel, is

$$\kappa_{linear}(v_i, v_j) = v_i^T v_j, \tag{3.19}$$

i.e., the inner products between two vectors v_i and v_j without any free parameter. The other kernel considered, the RBF kernel, is defined by

$$\kappa_{rbf}(v_i, v_j) = e^{\left(-\gamma \left\| v_i - v_j \right\|^2\right)}, \tag{3.20}$$

and it has a free parameter γ, that when considering $\gamma = \frac{1}{2\sigma^2}$ an equivalent RBF is obtained. The pair of parameters (C, γ) for the appliance classification is set to $\left(1, \frac{1}{2}\right)$. This was the pair with the highest cross-validation accuracy. Thereby the SVM classifier trained with this parameter combination was used for classifying unseen data.

The multi-class classification by SVM is carried out by the one-against-one (pairwise) and one-against-all strategies. The former was implemented using the LIBSVM (Chang and Lin, 2001), a publicly available software package, while the latter and 5-NN were implemented in Matlab software.

With regard to MinMaxSteady-State rule for the identification of steady-states in the signals, the tolerance value and the minimum number of consecutive samples necessary to specify a stable-state are also set by trial-and-error. In particular, we set $s_{min} = 3$, $\epsilon_{active} = 15$, $\epsilon_{reactive} = 15$ and $\epsilon_{pfactor} = 0.05$, respectively, the minimum number of samples required to define a stable segment of the signal, the tolerance value for the active, reactive and reactive power signals. These values were found adequate to identify the start and the end of stable states in the signals by the rule described in Section 3.2.1.

Concerning the particular case of Dataset B, as mentioned, Gaussian white noise was added to the signals. Noise reduction, in order to identify the steady-states, was then performed. Consequently, a denoise method based on wavelet decomposition was employed to reduce the noise. This procedure, the Wavelet Shrinkage (Donoho, 1995) method, is further explained in Chapter 4 Section 4.4.1. Still, to clarify the experience setup, its parameters are introduced briefly. The Wavelet Shrinkage is based on wavelet decomposition which is generated from a chosen basis function (also known as mother wavelet or wavelet function). In the Wavelet Shrinkage procedure, firstly, the discrete wavelet transform is employed to the noisy signal, converting it into the wavelet domain and decomposing it into J given levels of approximations and details. At each level, an approximation of the original signal and a signal composed by

high frequencies (details) are provided. Next, a chosen threshold is applied to the wavelet coefficients. Then, by the inverse discrete wavelet transform, the denoised signal is transformed back to the time domain. For Experience B, we explored $J = 2$ and $J = 3$ and tested the wavelet functions listed in Table 4.1 presented in Chapter 4. The list is composed by five main families of wavelets: Daubechies, Symlets, Coiflet, BiorSplines and ReserveBior. Further details will be presented in Chapter 4.

3.3.2. Results and Performance Evaluation

Given both datasets and the corresponding measured signals, the reactive power signals were calculated. Next, Theorem 3.2.1 from Section 3.2.1 was implemented, yielding the MinMaxSteady-State recognition algorithm, and steady-states were identified for the active, reactive and power factor signals. First, for each device on Dataset A, three steady-states were computed: a stable part before the switching on; another one corresponding to the appliances' operation phase; and a last one occurring after switching off. The step-change correspondent to the standby of a particular LCD screen was also possible to identify, when considering a proper small value for s_{min}. Second, similar steady-states for the signals of Dataset B were recognized including those associated with the switching on and switching off of the second device in the circuit, namely the 32 in. LDC screen and the toaster. Finally, the difference between the steady-states (step changes) was calculated so that positive/negative values could be associated with the switching on and switching off (Figure 3.1), respectively, which composed the feature vector of each appliance to identify. To assess the performance of the composed signature, the features for the six-class problem and four-class problem associated with the switching on were normalized by dividing each element of each feature vector by its norm.

Experience A: One Appliance in the Electrical Circuit (six-class problem)

Table 3.2 displays the F_1 measure for each class, the mean accuracies, the macro-averages (mean values of the F_1) and the micro-averages F_1 obtained by each method (SVM one-against-all and one-against-one and K-NN), from the experiments with Dataset A. In the pairwise strategy, the performance yielded by the two kernels was alike, therefore, only the results for the linear kernel are reported. Thereby, it can be concluded that in this experience and

	SVM						K-NN	
	One-against-all				Pairwise		K=5	
	Linear Kernel		RBF Kernel		Linear Kernel			
	F_1 (%)	Acc. (%)	F_1 (%)	Acc. (%)	F_1 (%)	Acc. (%)	F_1 (%)	Acc. (%)
Incandescent bulb	95.2 ± 1.7	98.1 ± 0.0	97.1 ± 2.8	98.5 ± 0.1	96.3 ± 4.1	98.7 ± 1.5	99.4 ± 1.1	99.7 ± 0.6
22 in. LCD screen	*n.d.*	83.2 ± 0.1	49.1 ± 21.2	89.6 ± 1.2	97.0 ± 5.3	99.0 ± 1.7	99.4 ± 0.0	99.0 ± 0.0
32 in. LCD screen	96.0 ± 0.0	99.3 ± 0.5	95.9 ± 1.9	98.4 ± 0.4	95.8 ± 2.2	98.7 ± 0.6	96.0 ± 4.6	97.9 ± 0.8
Microwave	96.7 ± 1.6	98.7 ± 0.3	97.9 ± 3.6	99.4 ± 0.5	67.7 ± 8.6	91.3 ± 1.5	96.8 ± 5.6	98.1 ± 1.0
Toaster	97.2 ± 2.8	99.2 ± 0.3	98.2 ± 3.2	99.8 ± 0.4	95.8 ± 1.8	98.7 ± 0.6	100.0 ± 0.0	100.0 ± 0.0
Coffee Machine	*n.d.*	99.2 ± 0.3	86.9 ± 4.9	99.8 ± 0.9	97.3 ± 1.8	99.0 ± 1.7	98.0 ± 1.2	100.0 ± 0.0
Average	*n.d.*	96.3 ± 6.4	87.5 ± 19.3	97.5 ± 4.1	91.7 ± 4.7	97.6 ± 3.1	98.3 ± 1.6	99.1 ± 0.9
Micro-average	76.8		90.0		91.0		98.9	

Table 3.2.: The mean accuracies (Acc.) and F_1 measure for SVM (with linear kernel and RBF kernel) and 5-NN approaches in Dataset A.

considering this classification scheme, the identity mapping, $\bar{\phi}$, for the feature space associated with the linear kernel simply gives the decision boundary (Equation 2.21), *i.e.* the kernel trick is unnecessary, which indicates that the classes (one-against-one) are linearly separable. The performance was quite effective for the one-against-all strategy. The mean accuracy was 97.5% for the SVM with RBF kernel which shows an increase of 1.2% over the performance of the SVM with linear kernel. A similar high accuracy value was also reached by the SVM pairwise approach, differing only by 0.1% from the performance achieved by the one-against-all SVM with RBF kernel. In terms of average accuracy, both SVM multi-class classification achieved similar high performance. Still, the highest was yielded by the 5-NN, which was of 99.1%.

More precise information about the classes classified as positive is reported by the macro and micro-averages F_1. Notice, that no F_1 was computed, for the 22 in. LCD screen and Coffee Machine which indicates that no true positives (TP) were found and these appliances were never correctly labelled. Regarding the macro F_1, the results yielded by the SVM model are near 90%. In fact, the performance for the pairwise strategy surpasses by 4.2% the one-against-all approach, when a RBF kernel is employed. Moreover, the macro F_1 associated 5-NN is 98.3% which represents an increase of 10.8% when compared to the performance of the one-against-all SVM model. Similar conclusions can be drawn regarding the micro-averages performance. In general, all the methods yielded high micro-averages F_1 values, around 90%, with exception of the SVM with linear kernel performed with the one-against-all strategy. In this case, the micro-average F_1 was, at least, 13.2% lower than for the other defined strategies. Once again, the performance achieved by the 5-NN was the highest, 98.9%, which represents an increase of, at least, 7.9% when compared with

the remaining methods. In a closer look to the F_1 by appliance (see most left columnn of Table 3.2), as mentioned, the SVM model with linear kernel, using the one-against-all strategy, was not able to identify any example of 22 in. LCD screen and Coffee Machine. Notice also that the F_1 of these devices yielded by the RBF kernel, and using the same multi-class classification type, is smaller than the F_1 for the remaining devices. This holds also for the other classification approaches. Still, the lowest F_1 for the pairwise strategy is associated with the Microwave.

Finally, the good performance in terms of micro-average, for each method, indicates that the composed signature can be an accurate description for each one of the appliances in the database. Nevertheless, a six-class problem is too small to uncover all the subtleties associated with in a larger multi-class problem where multiple devices should be discovered as it occurs in our homes. A dataset composed by measurements of a higher number of devices would be required to evaluate the performance of the proposed approach in more complex environments.

Experience B: A Pair of Appliances in the Electrical Circuit (four-class problem)

In Experience B, Dataset B, which reports signals for a pair of appliances in the same circuit, is used. As described in Section 3.3.1, for each acquired signal, two Gaussian white noises were generated and added to it, resulting in fifty sampled signals. Then, the Wavelet Shrinkage method was applied in order to obtain noise reduction. This experiment seems to be adequate to inspect the effect of presence of other information in the signal rather than loads of interest, and its corresponding reduction, regarding the appliance identification performance. Its benefits are clarified below.

Table 3.3 presents the average of the maximum improvement obtained for each type of signal in the Database B, in terms of Signal to Noise Ratio (SNR), which provides a measure between the meaningful information and the unwanted signal (further details in Chapter 4 Section 4.5.1). Both decomposition levels considered by the Wavelet Shrinkage achieved similar results: for the reactive power, the noise reduction yielded a smaller decrease than for the active power and the power factor. In fact, on average, the active power and power factor improvements exceeded by 189% and 166%, respectively, those of the reactive power. This might be due to the waveform of the reactive power signal and this

aspect should be further investigated.

	TV and Lamp		Microwave and Toaster	
	2 levels	**3 levels**	**2 levels**	**3 levels**
Active Power	4.6 ± 0.6	4.8 ± 0.7	5.2 ± 0.6	5.6 ± 0.7
Reactive Power	2.8 ± 1.0	2.2 ± 1.2	2.1 ± 0.9	1.3 ± 1.1
Power Factor	4.3 ± 1.1	4.0 ± 1.8	5.8 ± 0.9	6.3 ± 1.3

Table 3.3.: The mean improvements in the SNR value for two and three levels of wavelet decomposition for Dataset B.

From these denoised signals, steady-states were identified in the active, reactive and power factor signals by the proposed MinMaxSteady-State rule. Once again, these were used to calculate the step changes that composed the required feature vectors. For the classification of the appliance signatures, in order to identify them, both the SVM (with linear kernel and RBF kernel) and the 5-NN classifiers were used. Moreover, and as performed in Experience A, the SVM multi-class classification was carried out considering the one-against-all SVM and one-against-one SVM schemes. Table 3.4 presents the macro-averages, micro-averages and mean accuracies of the classification by one-against-all strategy considering the two distinct levels of wavelet decomposition in addition to the results yielded by the 5-NN. Table 3.5 presents the corresponding evaluation metrics for the one-against-one SVM classification.

		SVM				**K-NN**	
		One-against-all				**K=5**	
		Linear Kernel		**RBF Kernel**			
		F_1 (%)	**Acc.** (%)	F_1 (%)	**Acc.** (%)	F_1 (%)	**Acc.** (%)
2 Levels	**Incandescent bulb**	56.7 ± 25.2	86.2 ± 3.0	64.7 ± 5.1	84.5 ± 0.3	100.0 ± 0.0	100.0 ± 0.0
	32 in. LCD screen	63.8 ± 9.4	80.1 ± 2.7	64.9 ± 5.9	77.5 ± 4.0	100.0 ± 0.0	100.0 ± 0.0
	Microwave	87.1 ± 3.5	93.5 ± 0.8	90.8 ± 2.8	95.7 ± 0.4	90.7 ± 3.1	95.7 ± 0.4
	Toaster	91.6 ± 12.1	98.2 ± 2.4	91.5 ± 10.3	92.8 ± 3.0	92.3 ± 3.5	95.1 ± 1.1
	Average	74.8 ± 17.2	89.5 ± 8.0	78.0 ± 15.2	87.6 ± 8.3	95.7 ± 5.0	97.7 ± 2.6
	Micro-average	74.6		78.3		96.0	
3 Levels	**Incandescent bulb**	51.9 ± 14.5	77.8 ± 1.1	59.3 ± 14.6	86.0 ± 3.2	100.0 ± 0.0	100.0 ± 0.0
	32 in. LCD screen	68.7 ± 4.8	80.1 ± 1.2	73.0 ± 0.3	84.3 ± 0.7	100.0 ± 0.0	100.0 ± 0.0
	Microwave	94.9 ± 3.4	97.7 ± 0.8	96.9 ± 3.1	99.2 ± 0.8	96.9 ± 3.1	98.7 ± 0.4
	Toaster	97.0 ± 0.1	98.5 ± 0.0	97.0 ± 3.1	98.8 ± 0.4	97.1 ± 2.8	98.8 ± 0.4
	Average	78.1 ± 21.7	88.5 ± 11.1	81.6 ± 18.6	92.1 ± 8.0	98.5 ± 18.6	99.4 ± 0.7
	Micro-average	79.4		80.7		98.5	

Table 3.4.: The mean accuracies (Acc.) and F_1 for the tests performed using one-against-all SVM (linear kernal and RBF kernel) and 5-NN for Dataset B denoised with two and three levels of wavelet decomposition.

A notorious superiority is observed for the results achieved by the 5-NN and

the pairwise SVM with linear kernel in terms of accuracy and F_1 metrics. In fact, the mean accuracy was superior to 97% and the macro-average F_1 was of at least 95.7%. Notice that this macro-average F_1 value is 30.3% superior to the macro-average F_1 yielded by the pairwise SVM with RBF kernel, when considering two levels of decomposition at the denoising procedure. In this case, at the denoising step, the signals were decomposed into a subsignal of low frequency components, approximations, and two distinct subsignals of high frequencies components, details. The 5-NN improved the micro-average F_1 over the one-against-all SVM, for linear and RBF kernels, by 21.4% and 17.7%, respectively, for two levels of decomposition. Regarding the pairwise SVM with RBF kernel, the 5-NN improved its micro-average F_1 by 28.5%. Considering the three levels of decomposition, where three subsignals of high frequencies components were considered by the denoising procedure, similar observations can be drawn. The 5-NN micro-average F_1 is 19.1% and 17.8% higher than for one-against-all SVM with linear and RBF kernel, respectively. The micro-average F_1 yielded by the pairwise SVM with RBF kernel is 33.5% lower than the correspondent value achieved by the 5-NN. Similar improvements are achieved by pairwise classification with the linear kernel. Regarding the F_1 by appliance class, it can be observed that the classes with lower F_1 values for the SVM one-against-all strategy, the incandescent bulb and 32 in. LCD screen, are the ones with highest F_1 values by the 5-NN. Notice also that there is no significant improvement of the evaluation metrics by increasing the decomposition level for the denoising step. Actually, for the Toaster class results, associated to the pairwise approach with a RBF kernel, F_1 could not be calculated, since no TP were found. Notwithstanding, the associated average accuracy was 82.6%.

Under these circumstances, we empirically show that the performed methods are suitable for appliance classification and able to accomplish the of the steady-state signature recognition, even considering noise addition to the signals. Moreover, the proposed signature can be an accurate description for each one of the appliances in the Database B. Nevertheless, and once again, the explored 4-class problem is a small representation of a real-world problem and a dataset with electrical signals of a higher number of appliances should be further considered in foreseen studies.

		SVM			
		Pairwise			
		Linear Kernel		RBF Kernel	
		F_1 (%)	Acc. (%)	F_1 (%)	Acc. (%)
2 Levels	Incandescent bulb	100.0 ± 0.0	100.0 ± 0.0	88.7 ± 1.7	95.0 ± 0.8
	32 in. LCD screen	100.0 ± 0.0	100.0 ± 0.0	86.2 ± 6.6	94.0 ± 2.6
	Microwave	91.7 ± 5.0	96.0 ± 2.3	60.1 ± 1.5	67.3 ± 0.8
	Toaster	92.5 ± 4.0	96.0 ± 2.3	26.6 ± 14.6	78.4 ± 3.1
	Average	96.0 ± 4.6	98.0 ± 2.3	65.4 ± 28.9	83.7 ± 13.7
	Micro-average	96.5		67.5	
3 Levels	Incandescent bulb	100.0 ± 0.0	100.0 ± 0.0	94.7 ± 4.9	97.5 ± 2.3
	32 in. LCD screen	100.0 ± 0.0	100.0 ± 0.0	88.1 ± 13.1	95.0 ± 5.2
	Microwave	97.0 ± 0.0	98.5 ± 0.0	84.0 ± 22.4	87.6 ± 19.0
	Toaster	97.0 ± 0.1	98.5 ± 0.0	*n.d.*	82.6 ± 13.7
	Average	98.5 ± 1.7	99.3 ± 0.9	*n.d.*	90.7 ± 6.9
	Micro-average	98.5		65.0	

Table 3.5.: The mean accuracies (Acc.) and F_1 for the tests performed using one-against-all SVM (linear kernel and RBF kernel) for Dataset B denoised with two and three levels of wavelet decomposition.

3.4. Summary

In this chapter, the steady-state signatures as distinctive characteristics of devices considering the appliance identification as part of a NILM system were explored. As described in Chapter 2 Section 2.5, these signatures are inferred from the difference between two steady-states thereby, a rule for the identification of stable segments in a signal was described and mathematically proved in Section 3.2.1. This rule was further used to define the steady-states on the active, reactive signals and power factor. Indeed, an appliance electrical signature composed by the step changes on these three signals was proposed to be used by the classification process for load identification. For this identification process, models learned by the SVM and 5-NN techniques were studied.

To assess the performance of the proposed approach based on steady-state signatures on the active, reactive power and power factors, two datasets composed by measurements on the active power, voltage, current and power factors were gathered. First, a dataset considering that only one device, among an incandescent bulb and two LCD's, a microwave, a toaster and a coffee machine, was connected to the circuit. Second, a dataset considering that a pair of devices, an incandescent lamp and a LCD or a microwave and a toaster, was linked in

the network.

The multi-class SVM was performed by the two common approaches: one-against-all and one-against-one (pairwise classification). One of the biggest drawback for a system consisting of k (in this case $k = 6$) classifiers trained with one-against-all is that the classification boundary of each classifier is drawn independently from others due to the separate training processes. This may result in a situation that a portion of the feature space is not covered by any SVM, which is referred to as an uncovered region in the feature space, or a portion of the space is covered by more than one class, which is mention as an overlapped region. Since one-against-one modeling approach presents redundancies in classification, the posterior decision function can make a significant impact on the final system performance.

The computational experiments showed that the simplest methods are able to tackle the recognition issue accurately and the proposed signature can describe properly each appliance. We observed in the results thus far obtained that the pairwise approach performed better than the one-against-all. In addition, the 5-NN yielded high F_1 results. The devices in the datasets were treated as on/off devices and this experimental case study allowed for verifying the effectiveness of the proposed approach for these appliances. A larger number of devices on the datasets, in order to perform a more robust and more accurate evaluation over real conditions, is required. In addition, further research should consider the computation of energy consumption estimates of each device, which is a NILM request yet it was not addressed in the proposed approach.

Before aiming at the energy disaggregation into the consumption of each appliance, we will firstly investigate the extraction of information from the aggregated signal. In the preliminary study, described in the next chapter, the devices operation modes are divided into 'automatic' and 'manual'. Given this description, we will explore the separation of variations, associated to appliances which do not require human intervention for state-switching, from the remaining signal.

CHAPTER 4

Power Consumption Extraction from an Aggregated Signal

4.1. Introduction

In the previous chapter, the electrical devices used on the computational experiments compose only a small sample of equipment that can be found within nowadays houses. Notice that all the previously considered appliances share

a common feature: the equipment must be turn on and turn off with manual intervention. Still, not all the devices in a household are switched on and off manually. In fact, equipment as refrigerators, once they are connected to the network, work automatically and without any associated human intervention (as long as they don't breakdown). Henceforth, these will be as considered as 'automatic' equipment. Consequently, the electrical devices in a household can be roughly divided in two main groups ('automatic' and 'manual' appliances) and the separation of information corresponding to these two types of appliances from the aggregated electrical consumption signal is object of an exploratory study.

Extraction of variations associated only with the power consumption of 'automatic' appliances from the aggregated electrical signal is the focus of this chapter. The problem is approached employing widely used signal processing and time series analysis techniques, namely Wavelet Decomposition (WDT) and Singular Spectrum Analysis (SSA). The former is an alternative to Fourier analysis (Weeks, 2007) while the latter is a standard analysis tool for climatic and meteorological time series, and also explored for the analysis of physiological time series (Golyandina et al., 2001; Teixeira et al., 2005). The wavelet analysis requires the selection of a basis function such that signal decomposition can be performed. The usual approach uses only one mother wavelet for the entire signal. However, in this work we investigate the extraction of information from the aggregated signal considering that distinct segments may be better decomposed if different basis functions are used. Towards this end, we begin by exploring the most appropriate set of wavelet functions required to perform the wavelet decomposition, regarding the context of information extraction from whole-home electrical consumption measurements.

This chapter is organized as follows. First, the problem statement and background are introduced, followed by a description of two signal processing operations, embedding and diagonal averaging, in Section 4.3. Next, the decomposition techniques based on wavelet transform and SSA as well as an alternative technique proposed for the analysis of segments of the signal are presented. The computational experiments, performed with synthetic generated data and real-world electrical consumption signals, and the correspondent results and statistical evidence are presented and discussed in Section 4.5. Last, conclusions and future work are drawn from the empirical work.

4.2. Problem Statement and Background

Given the aggregated consumption during a period of time T, $\bar{x} \in \mathbb{R}^T$, our goal is to decompose it into signals $x'_1 \in \mathbb{R}^T$ and $x'_2 \in \mathbb{R}^T$ such

$$\bar{x} = x'_1 + x'_2 \tag{4.1}$$

where x'_1 corresponds to the aggregated consumption without the variations associated with 'automatic' equipment while these variations compose the signal x'_2. This problem can also be described as recovering source x'_1 from the mixture $\bar{x}$ which includes the additive 'noise' x'_2. Accordingly, strategies for signal recovering are studied.

In this chapter, we focus on extracting information from the signal by either transforming it by the Wavelet Transform (WT) or decomposing it by Singular Spectrum Analysis (SSA). Transforms are a powerful tool in signal processing, in particular the Fourier Transform (FT) (Weeks, 2007). In this classical analysis, a signal is transformed from the time domain into the frequency domain where it is represented by sinusoidal components, *i.e.* sine and cosine functions. This representation is based on periodic signals, thereby, the Fourier transform is well-suited for the study of signals on the frequency domain and for which temporal information is not relevant (as for stationary signals, whose frequencies do not change over time) (Rioul and Vetterli, 1991). When time dependent information is required, as for the analysis of non-stationary signals like speech, an alternative to FT is required. In fact, to overcome this limitation Short-Time Fourier Transform (STFT) was developed. In this alternative, the signal is divided into several windows of fixed size along the time axis and, for each of the windows, the signal is considered stationary and the Fourier transform is applied to each window.

Another alternative for the analysis of non-stationary signals is the Wavelet Transform (WT). This signal decomposition method provides both time and frequency analysis, making it more adequate for the exploration of non-stationary signals. In fact, it has been used in several areas of signal processing like compression and denoising of natural signals, speech and images (Rioul and Vetterli, 1991; Mallat, 2008). Wavelet analysis decomposes a signal into a set of basis functions called wavelets. These wavelets are the outcome of scalings and shifts of a single prototype wavelet. For this reason, the basis functions can be thought

as filters (Rioul and Vetterli, 1991). The diversity of application domains such as physiological signals (like Electrocardiogram (ECG) or Electroencephalography (EEG)) or speech where the results are quite satisfactory, encouraged us to explore it for the extraction of variations on electrical consumption (Alfaouri and Daqrouq, 2008; Kumar et al., 2008; Rocha, 2012; Mallat, 2008).

Rather than analyse the aggregated signal with a transform procedure, another possible strategy for 'breaking' the mixture signal $\bar{x}$ into the consumption x'_1 and the variations x'_2 would be a decomposition approach (see Chapter 5 Section 5.2) as Principal Component Analysis (PCA) (Hotelling, 1933) or Singular Value Decomposition (SVD) (Eckart and Young, 1936). These mathematical procedures factorize a given matrix into its principal components and are closely related. The main difference between the two techniques lies on the fact that SVD provides the principal components in both row and column space while PCA provides only the principal components associated with the column space (Teixeira, 2011; Shlens, 2009). These methodologies are an instrument in signal processing, namely for multivariate data analysis, *i.e.*, when a set of M signals of length T is provided. However, these techniques can also be applied to a single-channel signal, as the aggregated electrical consumption, by employing particular procedures that are described in the following section (Teixeira, 2011). A particular method based on SVD for dealing with one-dimensional signals is the SSA which is mainly applied to time series analysis forecasting and detection of structural changes in time series models in climatic, meteorological and geophysical areas (Golyandina et al., 2001). This method separates the initial mixture into a sum of interpretable components, oscillatory and noise (Hassani, 2007), and has been successfully employed to noise elimination in EEG signals (Teixeira et al., 2005). In this chapter we also investigate SSA for electrical signal decomposition into 'noise' and 'meaningful' components.

4.3. Signal Processing Operations for Multivariable Signals

Signal processing applications often address one-dimensional signals like the whole-home electricity measurements. Still, methods as PCA or SVD allow for the factorization of a matrix, which in multi-channel signal problems is naturally formed by the M signals of length T gathered at each instant of time

by the set of sensors. In one-dimensional signal processing applications, and in order to apply such factorization methods, one-dimensional data is mapped into a multi-dimensional signal considering vectors formed by the application of a window frame to the signal. The technique is applied to the resultant matrix and then the multi-dimensional data is mapped again into a one-dimensional signal. These procedures of time series analysis are known as embedding and diagonal averaging (which is embedding reverse operation) and will be briefly presented in the following.

4.3.1. Embedding

An embedding procedure can be simply described as a transformation of a one-dimensional signal into a sequence of multi-dimensional lagged vectors (Golyandina et al., 2001). Given a signal $x = [x(1), \ldots, x(T)] \in \mathbb{R}^T$ of length T, let L_w be a window length (defined by the user) such that $1 < L_w < T$. This window is composed of L_w signal elements, forming vectors with L_w components, where each successive vector starts with the second component of the previous (the lag) until all the samplings have been exhausted. Formally, the process considers $\check{N} = T - L_w + 1$ lagged vectors $\check{x}_{\check{n}}$ of length L_w,

$$\check{x}_{\check{n}} = [x(\check{n}), \ldots, x(\check{n} + L_w - 1)]^T, 1 \leq \check{n} \leq \check{N}. \tag{4.2}$$

The $\check{N}$ lagged vectors represent the columns of a $L_w \times \check{N}$ matrix, usually named as a *trajectory matrix*, defined as $\check{X} = \left[\check{x}_1 \ldots \check{x}_{\check{N}}\right]$, *i.e.*,

$$\check{X} = \begin{bmatrix} x(1) & x(2) & \ldots & x\left(\check{N}\right) \\ x(2) & x(3) & \ldots & x\left(\check{N}+1\right) \\ \vdots & \vdots & \ddots & \vdots \\ x(L_w) & x(L_w+1) & \ldots & x(T) \end{bmatrix}. \tag{4.3}$$

For example, consider the signal of length $T = 6$, $x = [5,7,3,2,1,6]$ and a window length $L_w = 3$. The trajectory matrix $\check{X}$ composed by $\check{N} = 4$ lagged vectors would be

$$\check{X} = \begin{bmatrix} 5 & 7 & 3 & 2 \\ 7 & 3 & 2 & 1 \\ 3 & 2 & 1 & 6 \end{bmatrix}. \tag{4.4}$$

Observe that each anti-diagonal line in the trajectory matrix presents the

exactly same elements and therefore this is an Hankel matrix (Golyandina et al., 2001). Another strategy for the definition of this matrix would be to impose that the diagonal entries are identical obtaining a Toeplitz matrix (Teixeira, 2011). Thus, by definition, the matrix entries verify $\bar{h}_{i,j} = \bar{h}_{i-1,j+1}$ in a Hankel matrix $\bar{H}$ and $\tilde{t}_{i,j} = \tilde{t}_{i-1,j-1}$ in a Toeplitz matrix $\tilde{T}$.

4.3.2. Diagonal Averaging

The reverse procedure, the diagonal averaging, transforms a matrix Y into a Hankel matrix that can then be converted into a time series. For that purpose, the averages of the corresponding anti-diagonals of this matrix Y are computed. Let Y be a $L_w \times \check{N}$ random matrix with entries $y_{ij}, i = 1, \ldots, L_w,\ j = 1, \ldots, \check{N}$. The Hankelization operation ($\mathcal{H}$), which transforms an arbitrary matrix into a Hankel matrix, calculates the elements $\tilde{y}_{ij}$ of $\mathcal{H}Y$ such that

$$\tilde{y}_{ij} = \begin{cases} \frac{1}{t-1} \sum_{l=1}^{t-1} y_{l,t-l} & \text{for} \quad 2 \leq t \leq L_w \\ \frac{1}{L_w} \sum_{l=1}^{L_w} y_{l,t-l} & \text{for} \quad L_w + 1 \leq t \leq \check{N} + 1 \\ \frac{1}{\check{N}+L_w-t+1} \sum_{l=t-\check{N}}^{L} y_{l,t-l} & \text{for} \quad N + 2 \leq t \leq \check{N} + L_w \end{cases} \tag{4.5}$$

assuming that $L_w \leq \check{N}$ and $t = i + j$, otherwise the elements $\tilde{y}_{ij}$ are calculated analogously by replacing L_w by $\check{N}$ and using the transpose matrix Y^T. This procedure is optimal since $\mathcal{H}Y$ is the nearest to Y in what concerns the matrix norm (Golyandina et al., 2001). For example, given a random matrix with 3×4 elements,

$$Y = \begin{bmatrix} 1 & 5 & -2 & 7 \\ 3 & 9 & 6 & 8 \\ 2 & 3 & 4 & 1 \end{bmatrix} \tag{4.6}$$

and applying the described procedure the outcome matrix is

$$\mathcal{H}Y = \begin{bmatrix} 1 & 4 & 3 & \frac{16}{3} \\ 4 & 3 & \frac{16}{3} & 6 \\ 3 & \frac{16}{3} & 6 & 1 \end{bmatrix}, \tag{4.7}$$

where

$$\tilde{y}_{ij} = \frac{5+3}{2}, \quad \forall i,j : i+j = 3, \tag{4.8}$$

$$\tilde{y}_{ij} = \frac{-2+9+2}{3}, \quad \forall i,j : i+j = 4, \tag{4.9}$$

$$\tilde{y}_{ij} = \frac{7+6+3}{3}, \quad \forall i,j : i+j = 5, \tag{4.10}$$

$$\tilde{y}_{ij} = \frac{8+4}{4+3-6+1}, \quad \forall i,j : i+j = 6, \tag{4.11}$$

$$\tilde{y}_{ij} = \frac{1}{4+3-7+1}, \quad \forall i,j : i+j = 7, \tag{4.12}$$

then $\tilde{Y} = \mathcal{H}Y$ is transformed into a time series, since its columns would correspond to the lagged vectors of the one-dimensional signal.

4.4. Explored Methods for Electrical Load Extraction

This section introduces the methods that will be explored to perform the extraction of variations from a given aggregated electrical signal. We begin by describing the two particular techniques of interest: Wavelet Shrinkage and Singular Spectrum Analysis. Next, an approach that considers the possibility of using distinct mother wavelets to analyse distinct signal segments is proposed.

4.4.1. Wavelet Decomposition and Wavelet Shrinkage

Wavelet Transform

Fourier transform (FT), an extensively used technique in signal processing, represents a given signal by oscillatory components, sines and cosines functions, allowing it to be analysed over the frequency domain. However, when the properties of the signal in study change over time, *i.e.*, a non-stationary signal, wavelet transform offers a more appropriate decomposition since temporal information is also taken into account. For this reason, the use of wavelet decomposition for the analysis of non-stationary signal has received great attention (Graps, 1995; Weeks, 2007). In this section we present a concise description of wavelet transforms. Further mathematical treatment of this subject can be found in (Mallat, 2008; Rioul and Vetterli, 1991).

Analogously with the Fourier analysis, the wavelet decomposition makes use of orthogonal bases, known as wavelets, to represent the signal of interest. On the contrary to the sines and cosines used on the FT, the wavelet functions have a irregular shape and only a finite time interval have non-zero values, *i.e.*, they are compactly supported. These characteristics make them adequate for the study of non-stationary signals. In particular, their irregular waveform is important to analyse signals with discontinuities or abrupt changes which can be temporally localised due to the compact support.

For the decomposition, a basic wavelet function, or *mother wavelet*, must be selected and the set of wavelets used to transform the data in study derives from this mother wavelet by 'shifting' it along the time axis (translations) and 'sketching' or 'compressing' it (dilatations). The outcome of the transformation is a set of wavelet coefficients showing how similar the signal and a basis function are (Rioul and Vetterli, 1991). Its selection is then a major step for a successful analysis. In order to better illustrate this transform and its adequacy for the analyses of non-stationary signals, in the following an overview of some of the methodological details of wavelet transforms for signal analysis is presented.

The Continuous Wavelet Transform (CWT) provides a representation based on the multiplication of a signal by the set of basis functions: translated and scaled versions of the chosen wavelet function $\phi(t)$,

$$\phi_{a,b}(t) = \frac{1}{\sqrt{a}}\phi\left(\frac{t-b}{a}\right) \tag{4.13}$$

where $a \neq 0 \in \mathbb{R}^+$ is the dilatation or scale which is inversely related with the frequency, $b \in \mathbb{R}$ is the translation along the time axis and $\frac{1}{\sqrt{a}}$ ensures the energy normalization (at each scale, the wavelet have the same energy). Then, for a given signal x,

$$CWT(a,b) = \int x(t)\,\phi^*_{a,b}(t)\,dt, \tag{4.14}$$

where * denotes the complex conjugation. The CWT calculates the correlation between a given signal and wavelets that are defined from continuously varying the scale a and the translation b. Consequently, this process provides redundant information (scales and translations are continuous quantities) which is abundant to reconstruct the original signal (Weeks, 2007; Rocha, 2012).

Since CWT provides redundant information due to the continuous scales and translations a sampled and computationally efficient version of it, the Discrete Wavelet Transform (DWT), was introduced. The time-scale parameters a and b are discretized such $a = a_0^j$ and $b = \tilde{n}a_0^j\tau_0$ where $j, \tilde{n} \in \mathbb{Z}, a_0 > 1$ is a fixed dilatation step and τ_0 is a translation factor that depends on a_0. Thereby, this transform uses discrete wavelets

$$\phi_{j,\tilde{n}}(t) = \frac{1}{\sqrt{a_0^j}}\phi\left(\frac{t - \tilde{n}a_0^j\tau_0}{a_0^j}\right) \tag{4.15}$$

$$= a_0^{\frac{-j}{2}}\phi\left(a_0^{-j}t - \tilde{n}\tau_0\right), \tag{4.16}$$

to transform the signal, where $j, \tilde{n} \in \mathbb{Z}$ indicate the wavelet's width or dilation (j) and its position by the translation factor $\tilde{n}$, *i.e.*, the scales and translations are discrete values. The CWT is usually sampled in a dyadic sampling grid[1] in the time-scale plane where each node is defined such that $a_0 = 2^j$ and $\tau_0 = 1$ (Rioul and Vetterli, 1991; Gargour et al., 2009). Therefore, the correspondent wavelets in Equation 4.15 become

$$\phi_{j,\tilde{n}}(t) = 2^{-\frac{j}{2}}\phi\left(2^{-j}t - \tilde{n}\right) \tag{4.17}$$

and the resultant wavelet coefficients are

$$c_{j,\tilde{n}} = \int x(t)\,\phi^*_{j,\tilde{n}}(t)\,dt, \tag{4.18}$$

where * denotes the complex conjugation, and that are used to describe the signal x as

$$x(t) = \sum_j \sum_{\tilde{n}} c_{j,\tilde{n}}\phi_{j,\tilde{n}}(t). \tag{4.19}$$

The DWT can be implemented in the form of filter banks employing the Multiresolution Analysis (MRA) strategy proposed by Mallat (1989). The decomposition is then accomplished by quadrature mirror filters, a pair of low and high pass filters, applied in a pyramidal form which provides scales with different time and frequency resolutions. Firstly, the signal in study x is separated

[1] A dyadic sampling grid consists of tiles of different width and length depending on actual time and frequency resolution of each partial DWT spectra component (Rioul and Vetterli, 1991).

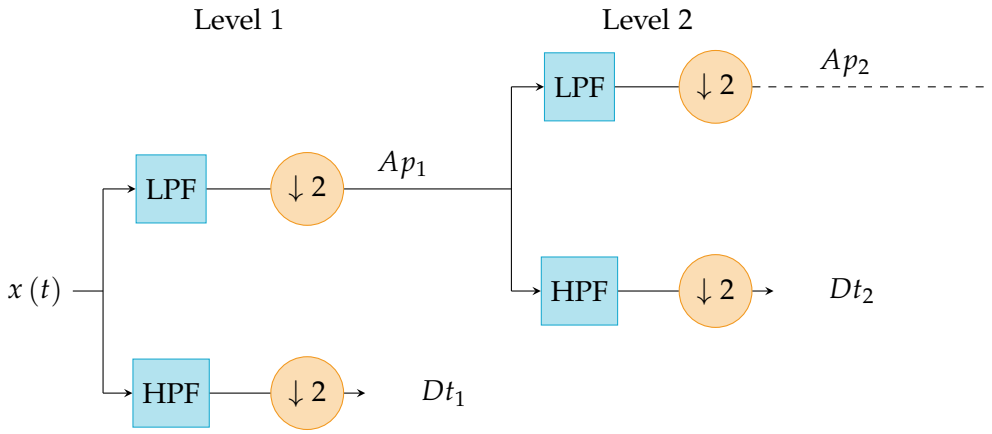

Figure 4.1.: The schematic DWT tree for 2 levels of decomposition.

into a sequence

$$Ap_1(t) = \sum_{\tilde{n}} l_{t-\tilde{n}} x(t) \tag{4.20}$$

resulting from the convolution of x with the low-pass filter with coefficients $l_1, \ldots, l_{p_f}$, $LPF = \left[l_1, \ldots, l_{p_f}\right]$, of length p_f, *i.e.*, with p_f no-zero elements. Next, a sequence

$$Dt_1(t) = \sum_{\tilde{n}} h_{t-\tilde{n}} x(t) \tag{4.21}$$

is computed by convolving x with the high-pass filter $HPF = \left[h_1, \ldots, h_{p_f}\right]$. Both Ap_1 and Dt_1 are composed by $\frac{T}{2}$ elements where T is the number of samples in signal x. The Ap_1 represents the smooth, low-frequency, information (approximations) of the signal while the Dt_1 contains the extracted details (high frequency information). The achieved approximations are then decomposed by the same approach, leading to the approximations Ap_2 and details Dt_2 for the second level of decomposition. The process is repeated for J defined levels and at each level $j \neq 1$, Ap_{j-1} is decomposed as described. This decomposition is illustrated in Figure 4.1. At each level j, the original signal can be re-written in terms of the current approximations Ap_j and of the sequence of details $Dt_1, \ldots, Dt_j$:

$$x(t) = f\left(Ap_j, Dt_j, Dt_{j-1}, Dt_1\right), \quad j = 1, \ldots, J. \tag{4.22}$$

The reconstruction process follows the reverse order of the decomposition. At each level, details and approximations are upsampled by two, passed through the correspondent reconstruction filters and then added (Rocha, 2012).

The described decomposition procedure can represent the signal in terms of wavelets and scale functions: the high frequencies components derived from the high-pass filter are represented by the mother wavelet (ϕ) while approximations associated with the low-pass filter are represented by the 'father' wavelet, or scaling function, φ, which are described similarly the mother wavelets of Equation 4.17,

$$\varphi_{j,\tilde{n}}(t) = 2^{-\frac{j}{2}} \varphi\left(2^{-j}t - \tilde{n}\right). \tag{4.23}$$

The scaling function φ and the mother wavelet ϕ are the required basis functions, in addition to the correspondent wavelet transform coefficients $c_{J,\tilde{n}}, d_{J,\tilde{n}}, \ldots, d_{1,\tilde{n}}$, to rewrite the decomposed signal at level J in terms of approximation and details. The required wavelet transform coefficients are the outcome of the inner product between the signal in study and the correspondent basis function:

$$c_{j,\tilde{n}} = \sum_t x(t)\, \varphi_{j,\tilde{n}}(t), \tag{4.24}$$

$$d_{j,\tilde{n}} = \sum_t x(t)\, \phi_{j,\tilde{n}}(t) \tag{4.25}$$

since the bases are orthogonal (Rocha, 2012).

Although wavelets have a similar structure they can be further grouped by families. The families can be characterize by the size of the support of the mother wavelet, the smoothness, the orthogonality and other related properties (Graps, 1995; Mallat, 2008; Rocha, 2012). The most known are Haar, Daubechies, Symlets and Coiflet, illustrated in Figure 4.2, among the possible groups (Rocha, 2012).

Wavelet Shrinkage

A denoising procedure attempts to remove noise contained in a signal regardless its frequency and retains the important information instead of only removing the existing high frequencies in the signal (Taswell, 2000). In order to perform signal denoising, Donoho and his co-authors proposed the Wavelet Shrinkage method which performs a thresholding (shrinking) in the wavelet domain (Donoho and Johnstone, 1994; Donoho, 1995; Donoho and Johnstone, 1995). It is composed of three main steps. First, the original signal is transformed into the wavelet domain and decomposed into a given level J of approximations and details by the discrete wavelet transform. Then, a chosen threshold is applied to the wavelet coefficients. Finally, the inverse discrete wavelet transform is

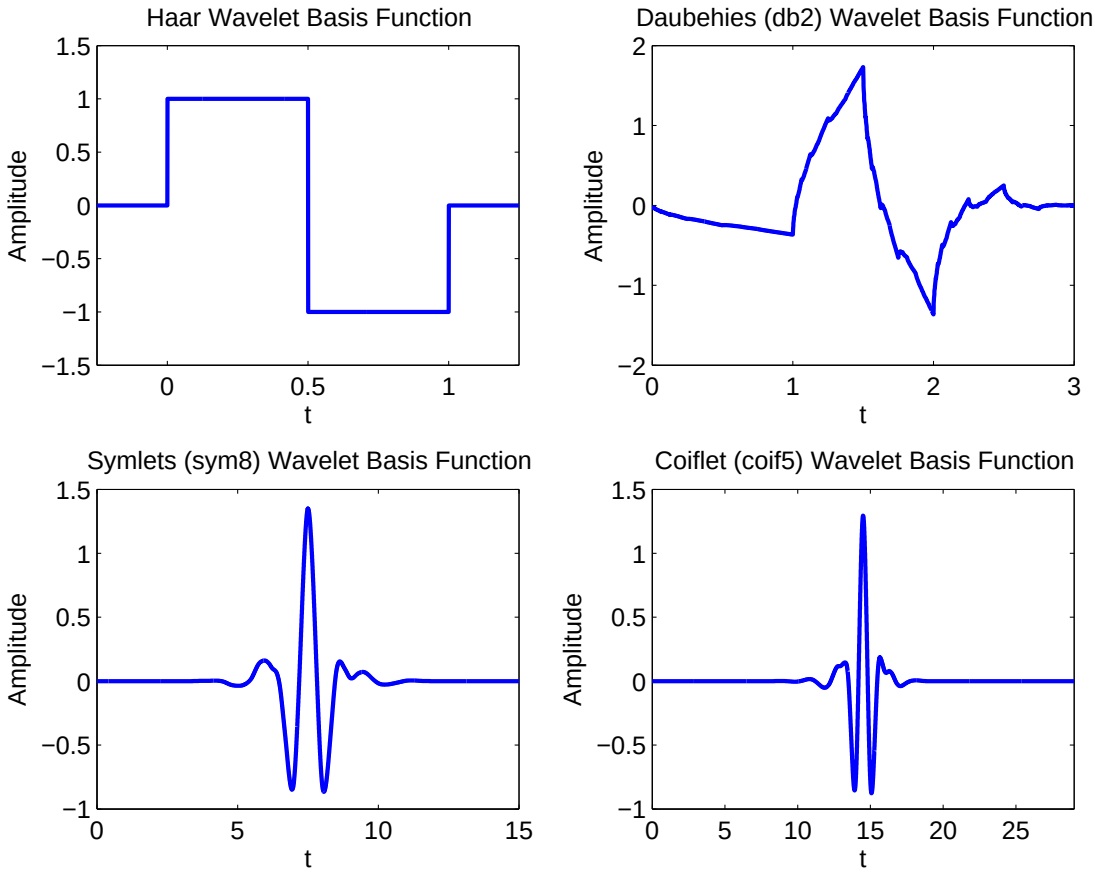

Figure 4.2.: Examples of the most known orthogonal wavelets (mother wavelet function).

performed in order to transform the denoised signal back to time domain. Formally, given a signal x, as defined in Equation 4.1, the Wavelet Shrinkage procedure can be described as

$$W = \mathcal{W}(x) \tag{4.26}$$

$$Z = \mathcal{D}(W, th_{ws}) \tag{4.27}$$

$$\tilde{y} = \mathcal{W}^{-1}(Z) \tag{4.28}$$

where $\mathcal{W}(\cdot)$ and $\mathcal{W}^{-1}(\cdot)$ correspond to the forward wavelet transform operator and to the inverse wavelet transform operator, respectively, $\mathcal{D}(\cdot, th_{ws})$ denotes the denoising operator with a threshold th_{ws}, and $\tilde{y}$ is the recovered signal.

The threshold value plays an important role in this method and several methods were introduced and discussed by Dohono et al. like the SURE

threshold, the James-Stein threshold (Donoho and Johnstone, 1995), the minimax threshold and the universal threshold where both hard and soft thresholding were analysed for wavelet and Fourier domain (Donoho and Johnstone, 1994). Indeed, this universal thresholding method and its two variants (hard and soft thresholding) are generally chosen to perform the denoising (Gargour et al., 2009). In the hard method, absolute coefficients values smaller than the given threshold are set to zero:

$$c = \begin{cases} c & \text{if } |c| \geq th_{ws}, \\ 0 & \text{if } |c| < th_{ws}, \end{cases} \tag{4.29}$$

where c is the wavelet coefficient. For the soft-threshold, the coefficients are set according to:

$$c = \begin{cases} c - th_{ws} & \text{if } c \geq th_{ws}, \\ 0 & \text{if } |c| < th_{ws}, \\ c + th_{ws} & \text{if } c \leq -th_{ws}, \end{cases} \tag{4.30}$$

which modifies the magnitude of all the wavelet coefficients.

Further research has proposed other wavelet shrinkage methods. These can be considered either probabilistic or selective methods, based on the performed modification of the wavelet coefficients. For probabilistic approaches, the level of reduction applied to the magnitude of the wavelet coefficient is continuous between 0 and 1 while the selective methods, the coefficients are either selected or removed (as in the hard-threshold) (Balster et al., 2005).

4.4.2. Singular Spectrum Analysis

Singular Spectrum Analysis (SSA), a method for time series analysis, aims at the separation of the signal into basic and meaningful sub-components. This methodology uses no previous statistical assumptions concerning the signal or noise, despite applying elements of classical time series analysis (Golyandina et al., 2001). Due to its capabilities and according to Golyandina et al., this technique has been successfully used in the analysis of climatic, meteorological and geophysical time series.

The decomposition via SSA is carried out by a two stage procedure: decomposition and reconstruction. Each stage is composed by two steps. The decomposition stage starts with the embedding where a one-dimensional sig-

nal is transformed, as described in Section 4.3, in a multi-dimensional signal represented by the trajectory matrix for which the SVD is computed. The SVD constitutes the second step of the decomposition stage. The following stage, reconstruction, starts with the grouping step. The matrices computed by the SVD are split into groups and, within each group, they are summed. At this point, the trajectory matrix can be represented as a sum of matrices. At last, each resulting matrix of each group is transformed into a time series by diagonal averaging (as described in Section 4.3).

An important step in SSA is the SVD factorization. This technique decomposes a given data matrix $\check{X} \in \mathbb{R}^{L_w \times \check{N}}$ into matrices $U' \in \mathbb{R}^{L_w \times L_w}$, $\Sigma \in \mathbb{R}^{L_w \times \check{N}}$ and ${V'}^T \in \mathbb{R}^{\check{N} \times \check{N}}$ such that

$$\check{X} = U' \Sigma {V'}^T \tag{4.31}$$

where the left singular vectors compose the columns of U' (the eigenvectors of the correlation matrix $\bar{S} = \check{X}\check{X}^T$), the right singular vectors compose the rows of ${V'}^T$ (V' is composed by the eigenvectors of the matrix $\check{X}^T\check{X}$) and the singular values $\tilde{\lambda}_1 \geq \ldots \geq \tilde{\lambda}_{L_w}$, ordered from high-to-low value, are the diagonal entries of the diagonal matrix Σ (Wall et al., 2003; Teixeira, 2011).

In the decomposition stage of SSA, after the embedding procedure which transforms the original signal x into the trajectory matrix $\check{X}$, the covariance matrix $\bar{S} = \check{X}\check{X}^T$ is computed and is decomposed using SVD resulting in the extraction of L_w eigenvalues $\tilde{\lambda}_1 \geq \ldots \geq \tilde{\lambda}_{L_w}$ and L_w eigenvectors (the columns ${U'}_1, \ldots, {U'}_{L_w}$ of matrix U'). By fixing $d = \max_i \{i : \tilde{\lambda}_i > 0\}$, for $i = 1, \ldots, L_w$, and considering the matrices ${V'}_i = \check{X}^T \frac{{U'}_i}{\sqrt{\tilde{\lambda}_i}}$ and matrices $\check{X}_i = \sqrt{\tilde{\lambda}_i {U'}_i {V'}_i^T}$, $i = 1, \ldots, d$, then the SVD of matrix $\check{X}$ can be written as

$$\check{X} = \check{X}_1 + \ldots + \check{X}_d. \tag{4.32}$$

The matrices $\check{X}_i$ are then split into I_n disjoint groups and, within each group they are added, and $\check{X}$ can be rewritten as

$$\check{X} = \check{X}_{I_1} + \ldots + \check{X}_{I_n}, \tag{4.33}$$

where $\check{X}_{I_m}$ corresponds to the resultant matrix for group I_m, $I_m = 1, \ldots, I_n$. The number of considered groups is directly connected with the goal to be achieved. For instance, for signal denoising only two groups are set for the reconstruction

process, *i.e.*, $I_n = 2$. Finally, for each group, this 'reconstructed trajectory matrix' is transformed into a one-dimensional signal by the diagonal averaging method (Section 4.3).

Variations of SSA have been introduced in the related literature, like the approach described in (Teixeira, 2011). The author describes therein the Local SSA in which the columns of the trajectory matrix are clustered, and for each cluster the SVD decomposition and grouping steps are performed. After computing all the reconstructed matrices for each cluster, the clustering process is reversed, as well as the embedding process.

In order to accomplish SSA, several parameters must be set. For embedding, the window length L_w must be assigned. Nevertheless no general rule exists to define it. According to Golyandina et al. (2001), L_w should be large enough so that important parts of the initial time series are present in each lagged vector. However, if no additional information is provided, the window length is usually set to approximately half of the dimension of the time series. Another strategy, applied in (Teixeira, 2011), considers the sampling frequency and the minimum frequency to be extracted in order to define the window length.

Also the number of disjoint groups considered at the grouping step, I_n, must be predefined. The process of forming I_n groups of matrices is not easily formalized (Golyandina et al., 2001). Regarding the problem described in Section 4.2, the recover of x'_1 from the aggregated signal $\bar{x}$ which contains x'_2 (the variations of 'automatic' appliances) two disjoint groups, *i.e.* $I_n = 2$ should be considered. As said before, x'_2 was regarded as 'noise' in the signal $\bar{x}$. Thereby, the matrices $\check{X}_i, i = 1, \ldots, d$, must be split into two groups. One group represents the 'denoised' signal while the other contains the 'noise'. Consequently, the number of relevant directions to be maintained for the reconstruction process, *i.e.*, the most significant eigenvalues and corresponding eigenvectors, must be selected. The number of relevant directions can be fixed regarding the variance of the data that should be kept associated with x'_1. This criterion assumes that the eigenvalues are ordered by decreasing order, a threshold th_{ssa} (usually set between 85% and 90%) is assigned and the number of directions $\tilde{L}$ must satisfy

$$\frac{\tilde{\lambda}_1 + \ldots + \tilde{\lambda}_{\tilde{L}}}{\tilde{\lambda}_1 + \ldots + \tilde{\lambda}_{\tilde{L}} + \ldots + \tilde{\lambda}_{L_w}} \times 100 > th_{ssa}. \tag{4.34}$$

In this popular criterion, the threshold keeps 85% to 90% of variance of the

initial information (Shawe-Taylor and Cristianini, 2004; Teixeira et al., 2005). Alternatively, and was used in (Teixeira et al., 2005), the number of directions could be determined in order to minimize the Minimum Description Length (MDL) criterion[2] is minimized (Rissanen, 1978).

4.4.3. Embedding, Wavelet Shrinkage and Diagonal Averaging

This section proposes an approach denominated Embedding, Wavelet Shrinkage and Diagonal Averaging (EWD), whose general architecture is illustrated in Figure 4.3. In short, a embedding step and correspondent inverse procedure are introduced in the Wavelet Shrinkage method. For a given one-dimensional aggregated signal $\bar{x}$, a window of length L_w is defined and a multi-dimensional signal $\check{X} \in \mathbb{R}^{L_w \times \check{N}}$ is computed using an embedding procedure (Section 4.3). The resulting $\check{N}$ columns of matrix $\check{X}$ correspond to the $\check{N}$ lagged segments of the original signal $\bar{x}$. Next, each one of these vectors is transformed into the wavelet domain and decomposed into a given level J of approximations and details. At this point, the wavelet coefficients are modified by a threshold and then the inverse discrete wavelet transform is performed. These three steps compose the shrinkage procedure summarized by Equations 4.26, 4.27 and 4.28. Thereby, for each column of $\check{X}$ we have the corresponded shrunk lagged vectors, y, that form the columns of the matrix $Y = \left[y_1 \dots y_{\check{N}}\right]$. Finally, a one-dimensional signal $\tilde{y}$ results from reversing the embedding process by the application of the diagonal averaging procedure (Section 4.3) to the matrix Y. Formally, the EWD can be described as

$$\check{X} = \mathcal{E}\left(\bar{x}\right) \tag{4.35}$$

$$W_{\check{n}} = \mathcal{W}\left(x_{\check{n}}\right), \forall\, x_{\check{n}}, \check{n} = 1, \dots, \check{N} \tag{4.36}$$

$$Z_{\check{n}} = \mathcal{D}\left(\check{W}_n, th_{ws}\right), \forall\, \check{W}_{\check{n}}, \check{n} = 1, \dots, \check{N} \tag{4.37}$$

$$y_{\check{n}} = \mathcal{W}^{-1}\left(Z_{\check{n}}\right), \forall\, Z_{\check{n}}, \check{n} = 1, \dots, \check{N} \tag{4.38}$$

$$\tilde{y} = \mathcal{DA}\left(Y\right) \tag{4.39}$$

[2]The minimum description length principle is a general approach for statistical modeling and inference from observed data which states that the model selection among a class of possible models should prefer the one that allows the shortest description of the data.

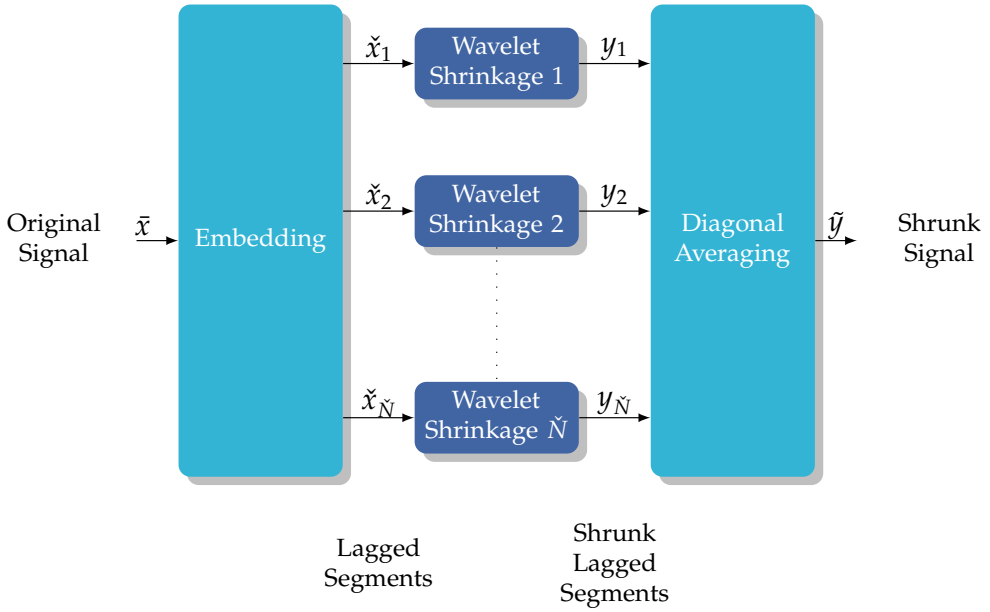

Figure 4.3.: The EWD approach with $\bar{x}$ the original one-dimensional signal, $\check{x}_{\check{n}}, \check{n} = 1, \ldots, \check{N}$, the lagged vectors and corresponding shrunk versions ($y_{\check{n}}$ and $\tilde{y}$).

where $\check{X} = \left[\check{x}_1 \ldots \check{x}_{\check{N}}\right]$, $Y = \left[y_1 \ldots y_{\check{N}}\right]$, and $\mathcal{E}(\cdot)$ and $\mathcal{DA}(\cdot)$ denote the embedding and diagonal averaging operators, respectively. As defined in Section 4.4.1, $\mathcal{W}(\cdot)$ and $\mathcal{W}^{-1}(\cdot)$ represent the forward and inverse wavelet transform operators, respectively, and $\mathcal{D}(\cdot, th_{ws})$ corresponds to the denoising operator with a threshold th_{ws}. Notice that for each lagged segment $\check{x}_{\check{n}}$ a distinct mother wavelet can be employed when performing the step associated with the Equation 4.36 (Figure 4.1) since there may be functions mode adequate than others for each $\check{x}_{\check{n}}$.

In fact, the Wavelet Shrinkage method is based on the wavelet decomposition, thereby the most suitable wavelet to decompose the signal $\bar{x}$ into the several approximations and details is required (Section 4.4.1). Nevertheless, the EWD allows for the use of different wavelets to decompose distinct segments of the original signal, since the wavelet shrinkage is applied to each of the lagged vectors that compose the trajectory matrix. This differs from the Wavelet Shrinkage method described in Section 4.4.1 where a single chosen wavelet is used to decompose the whole signal.

The EWD requires the set up of the window length L_w, the mother wavelet used to decompose each column of the matrix $\check{X}$ and the threshold th_{ws} being applied that the wavelet coefficients in order to perform the Wavelet Shrinkage

procedure. Computational experiences were done changing the window length L_w. The assignment of L_w to half of the signal length has shown the better performance for the EWD method among the L_w values tested. Further details are presented in Appendix A.

As mentioned and considering that no previous knowledge about the most appropriated wavelet for the decomposition of the signal $\bar{x}$ in study exists, several wavelets are explored, by trial-and-error, for each one of the lagged vectors and, in each case, the one with best performance results is selected. The choice is based on the similarity measured using the correlation between the extracted information from $\bar{x}$ and prototype signals of the information to be separated from $\bar{x}$. The mother wavelet achieving the highest similarity between the described signals is selected as the most adequate for the signal in study. This process is performed for each lagged vector, *i.e.*, for each column of the trajectory matrix $\check{X}$ a different mother wavelet can be used depending on the lagged vector signal. Further details are provided in Section 4.5.1.

Last, the threshold th_{ws} being applied to the wavelet coefficients is part of the Wavelet Shrinkage step, thereby the options for its assignment were described in Section 4.4.1.

4.5. Computational Experiments

This section describes the experimental setup and the results of the performed computational experiments. First, the previous described methods (Wavelet Shrinkage, Singular Spectrum Analysis, and Embedding, Wavelet Shrinkage and Diagonal Averaging) are applied in a synthetic generated dataset. The goal is to compare the performance of the three approaches on removing additive noise from the signals. Second, a study is carried out using real-world data gathered in a domestic environment where the energy consumption was daily recorded during a period of four months. For this dataset we explore the 'best' wavelet function to transform and denoise the aggregated electrical signal by the Wavelet Shrinkage. Third, the three approaches are compared regarding the extraction of information from the signal in study. In other words, the performance comparison concerns the separation of two types of information contained in the household electrical signal: the variations of 'automatic' appliances loads (considered here as 'noise') from the consumption of manually operated appliances and stand-by consumptions.

4.5.1. Experimental Setup

In the following, the designed computational experiences are detailed by presenting the respective datasets, evaluation metrics to assess the techniques performance, the experimental parameters and the statistical tests applied to validate the conclusions of the empirical evaluation. The three above described methods were implemented in Matlab software.

Dataset Description

Two experiences were performed considering (i) a synthetic generated signal and (ii) the aggregated consumption of a household.

Synthetic Generated Dataset A sinusoid signal consisting of 504 samples ranging from $-\pi$ to π was generated forming a time series x. In addition, 50 Gaussian white noise signals for three different SNR levels (9dB, 11dB and 13dB), $\tilde{r}$, were also generated. For each SNR level, this dataset is composed by 50 signals $\tilde{x}$ resulting from the addition of the x to the noise $\tilde{r}$. Furthermore, six Gaussian white noise signals, for each SNR level, were generated to be used as the prototype noise set needed for EWD in order to select the must appropriate mother wavelet.

Household Electrical Signal Dataset The electrical consumption dataset is part of a larger collection of data composed by the electrical consumption of 15 monitored households collected by Abreu et al. (2010). The dataset in study includes the aggregated electrical consumption measured during four months using a smart meter for each one of these households. The collected data was transmitted, via power line, to a communication module, where every fifteen minutes, the second by second data was aggregated and sent to a central server (Abreu et al., 2010). As referred by the authors, errors may occur leading to omissions in data. Thereby, after a data preparation step where uncompleted days were excluded, the daily signals with 96 samples were divided in two sets: one with the load consumption of 107 regular days and another set of 6 vacation days (Figure 4.4). During the vacation period no human activity interference was observed. Consequently, the measured electrical consumption can be considered as corresponding only to the 'automatic' appliances and stand-by modes. The variations within these signals are associated with the 'au-

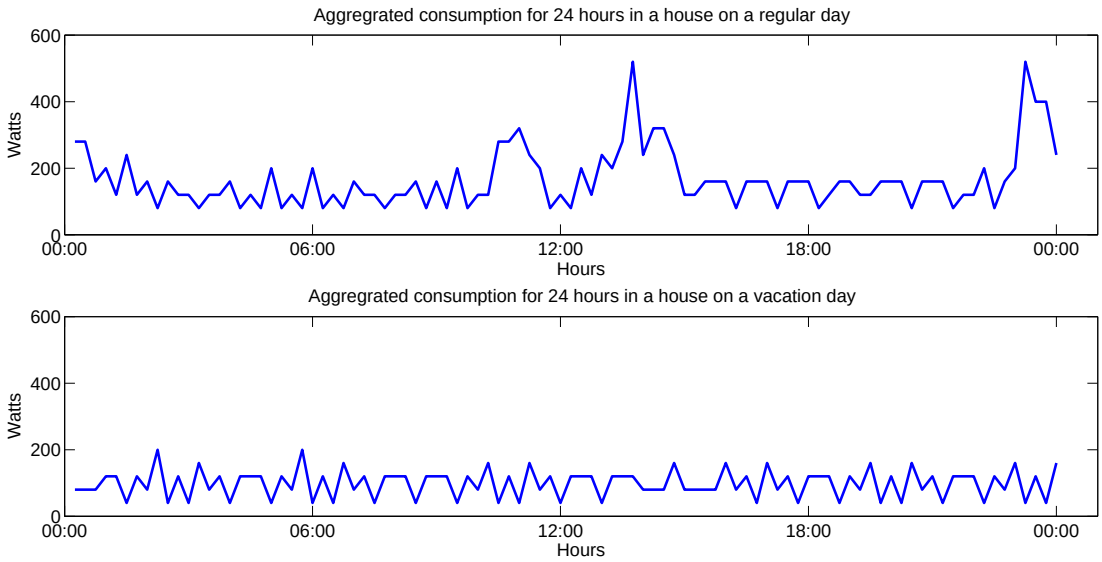

Figure 4.4.: The home aggregated consumption signals during 24 hours for a regular (at top) and a vacation day (at bottom).

tomatic' appliances with a periodical behaviour as refrigerators: it automatically switches on for a certain period of time when it is refrigerating and turns off when the desired inside temperature is reached. For this reason, the variations extracted from these signals by the described methods were used for performance evaluation and, in particular, as the prototype information required by the EWD. In other words, the extracted information from the vacation signals are examples of the information that should be extracted from the aggregated signal. Figure 4.5 illustrates one vacation day and the correspondent version after the variation extraction by the Wavelet Shrinkage procedure, considering two levels of decomposition and the Haar wavelet as mother wavelet.

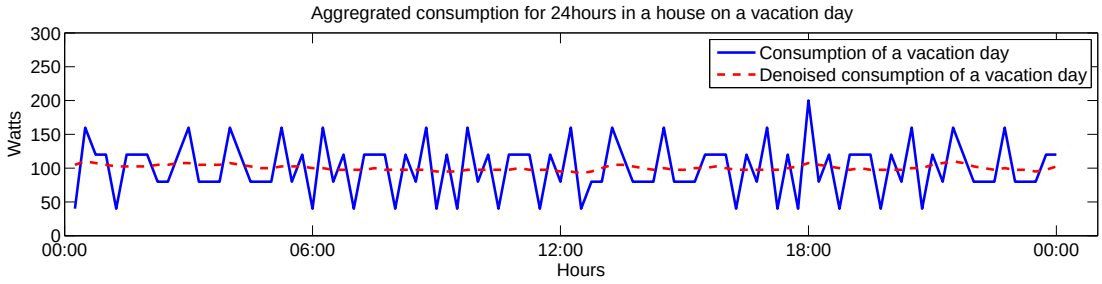

Figure 4.5.: The home aggregated consumption signals during a 24 hours vacation day and correspondent denoised version.

Evaluation Metrics

The performance of each approach was measured considering the Signal to Noise Ratio (SNR) and the correlation for the synthetic and real datasets, respectively. In the first experience, using the synthetic dataset, the objective is to compare the performance of the three methods to denoise a given signal, and thereby, for this assessment, the SNR

$$SNR = 10 \log_{10} \left(\frac{\sigma_x^2}{\sigma_e^2} \right) \tag{4.40}$$

where σ_x^2 is the mean square of the signal in study and σ_e^2 is the mean square of the noise, was computed (Fgee et al., 1999).

In the second experience we aim at the extraction of variations in the regular days signals associated with the electrical consumption of 'automatic' appliances. Therefore, the performance was evaluated by measuring the similarity between the extracted signal with the prototype signals associated with the vacation dataset as described previously. For this evaluation, the correlation, which computes how much alike two signals are by convolution (reversing one signal), was calculated (Weeks, 2007). The outcome value, the correlation coefficient, ranges within $[-1, 1]$. A positive large number indicates a strong relation between the signals and a negative number reflects an inverse relation (one signal can be increasing along the time while the other is decreasing (Weeks, 2007)). A coefficient value close to zero suggests that the two signals are not alike. However, a weak relation can be associated with an inappropriate alignment of the signals. To overcome this issue, one of the signals is shifted T times (with T being the signal length) and, for each shift, the correlation is computed. The final value for correlation between the two signals is then set to the maximum of the values that were calculated. This process is known as cross-correlation. Formally, the estimated cross-correlation between the signals $x \in \mathbb{R}^T$ and $y \in \mathbb{R}^T$ is

$$\rho_{x,y}(k) = \frac{s_{x,y}(k)}{\sqrt{s_{x,x} s_{y,y}}} \tag{4.41}$$

where

$$s_{x,y}(k) = \sum_{t=1}^{T} x(t)\, y(t+k) - \frac{\left(\sum_{t=1}^{T} x(t) \right) \left(\sum_{t=1}^{T} y(t) \right)}{T}, \tag{4.42}$$

$$s_{x,x} = \sum_{t=1}^{T} \left(x\left(t\right)\right)^2 - \frac{\left(\sum_{t=1}^{T} x\left(t\right)\right)^2}{T}, \tag{4.43}$$

such that k stands for the index of the reversed y and $s_{y,y}$ comes out from Equation 4.43 by replacing x with y (Weeks, 2007).

In short, the performance of each method for the household dataset (a set of 107 regular days and a set of 6 vacation days) is evaluated by correlation computed by the cross-correlation process. The extracted signal from each regular day $e_i, i = 1, \ldots, 107$, is compared with the extracted information from each vacation signal $\tilde{v}_i, i = 1, \ldots, 6$ (see Figures 4.4 and 4.5). Thereby, for each regular day, a set of 6 correlation values is calculated. The highest value is assigned as the performance measure for that given regular day. In the case of the Wavelet Shrinkage, this comparison is directly performed by considering the same mother wavelet for the extraction of both e_i and $\tilde{v}_i$. However, a set of wavelets is tested and the highest correlation value between the described signals is reported as the performance associated with the given regular day. The associated wavelet is considered as the most suitable for the task of the extraction of variations associated with the 'automatic' devices from the regular day signal.

A similar strategy is used to assess the EWD performance. Still, the comparison occurs between the extracted variations from segments of signals. In particular, the extracted information from a segment of a regular day is compared with the extracted signal from all possible segments of the vacation days. Thereby, the best correlation for two given distinct segments within the same regular day may be achieved for extracted information of different vacation segments corresponding to two different vacation days. This evaluation scheme results in a correlation value for each segment (the best among the calculated ones) of a given regular day. Then, the mean of the highest correlation values yielded for each segment is reported as the performance measure.

Experimental Parameters

As previously mentioned, the approaches in study require the setup of distinct parameters. The wavelet based approaches need a wavelet function and the number of levels J being used for the signal decomposition as well as the

threshold th_{ws} being applied to the wavelet coefficients to perform the required extraction of 'noise'. Regarding the Singular Spectrum Analysis, the number of relevant directions and the embedding dimension L_w must be set. This last parameter is also required by the EWD.

No previous knowledge about what function would constitute the most suitable wavelet for the decomposition of the aggregated electrical consumption is established, thereby an extended set, presented in Table 4.1, is considered for a trial-and-error selection. With respect to the Wavelet Shrinkage approach, the list of fifty-two wavelets is tested for each signal in the dataset. Based on the performance results it is possible to select a sub list of 'better' wavelets. The sub list is composed by wavelet functions that were chosen at least one time as yielding the highest correlation value for the Wavelet Shrinkage approach, in any of the decomposition levels considered. As the EWD approach is evaluated over the same dataset, and in order to save computation time, this sub list of wavelets is used.

Wavelet Family	**Wavelet function**
Daubechies	db1, db2, db3, db4, db5, db6, db7, db8, db9, db10
Symlets	sym2, sym3, sym4, sym5, sym6, sym7, sym8
Coiflet	coif1, coif2, coif3, coif4, coif5
BiorSplines	bior1.1, bior1.3, bior1.5, bior2.2, bior2.4 bior2.6, bior2.8, bior3.1, bior3.3, bior3.5 bior3.7, bior3.9, bior4.4, bior5.5, bior6.8
ReverseBior	rbio1.1, rbio1.3, rbio1.5, rbio2.2, rbio2.4, rbio2.6, rbio2.8, rbio3.1, rbio3.3, rbio3.5 rbio3.7, rbio3.9, rbio4.4, rbio5.5, rbio6.8

Table 4.1.: Wavelet functions list used by the wavelet decomposition based approaches.

The selection of J, the levels of decomposition used to compute the discrete wavelet transform, is also crucial. This definition must concern the specific signal in analysis. In the case that the noise is hardly noticeable, more resolution is needed, *i.e.*, more level of detail. Accordingly, a higher number of levels of decomposition would be appropriated. For the extraction of the required variations from the aggregated electrical signal, and after trials with the dataset, $J = 2$ and $J = 3$ were chosen for the wavelet based approaches. In the following, the performed methods will be denoted as Wavelet Shrinkage performed with 2 levels of decomposition (WDT-2) and as Embedding, Wavelet Shrinkage

and Diagonal Averaging performed with 2 levels of decomposition (EWD-2). Likewise, for 3 levels of decomposition we denote: Wavelet Shrinkage performed with 3 levels of decomposition (WDT-3) and Embedding, Wavelet Shrinkage and Diagonal Averaging performed with 3 levels of decomposition (EWD-3).

The wavelet based techniques also require the assignment of the threshold applied to the wavelet coefficients resulting from the decomposition. In this Thesis, in accordance to (Donoho, 1995), the universal soft-threshold method was employed considering $th_{ws} = \sqrt{2\sigma^2 \log(T)}$, where σ^2 is the variance of the signal and T is the number of samples.

Regarding the SSA, the number of relevant directions being kept in order to denoise the signal was set based on the variance of the data. In this work $th_{ssa} = 90\%$ is chosen, and the number of relevant directions is defined such that Equation 4.34 holds. The SSA also requires the setting of the embedding dimension L_w. As mentioned on Section 4.4.2, no general rule is known to assign a value to L_w and usually it is set to half of the signal length, if no additional information is provided (Golyandina et al., 2001). Thereby, for this computational experience, L_w is equal to half of the dimension of the signals in study for each dataset. For the same reasons, this value was also used as the embedding dimension required by the EWD. Still, we empiricaly have shown that the EWD performance is better when this L_w value is used (Appendix A).

Statistical Validation

The performance assessment and its validation via the experimental analysis is essential when comparing several methods. In this line, statistical inference rises as a necessary tool to support the conclusions achieved by the empirical evaluation and their generalization beyond the performed tests. In accordance, statistical tests are performed to validate the conclusions drawn for both designed computational experiences. In the following we describe briefly two particular tests that are required to validate the observations for the experiences with (i) the synthetic generated dataset and (ii) the household electrical consumption dataset. Notice also that we assume the independence of the sample of results being analysed, *i.e.*, the performance for a given signal is not influenced by another signal.

The one-way ANOVA is a statistical test used to compare means of two or more independent samples. In particular, it will be used to test the following null hypothesis: there is no statistical difference between the results of the three

methods and its versions. In a situation that its outcome lead us to reject the null hypothesis, that is, there is a low evidence that the null hypothesis holds, multiple comparison of the means, also known as post-hoc analysis, can be performed in order to identify the specific groups whose means are significantly different from others. In such analysis, pairwise comparisons are performed. The selection of the most appropriated one from the set of tests available such as the Tukey, the Bonferroni, or the Tamhane is based on the homogeneity of the variances. The first two tests are adequate when the homogeneity of the variances is assumed, otherwise the Tamhane's T2 test, a pairwise comparison based on t-tests, is applicable (Hochberg and Tamhane, 1987).

When the assumptions to perform the ANOVA test, such as normality, are not met, the alternative statistical analysis lies on non-parametric tests. The equivalent non-parametric test to ANOVA is the Kruskal-Wallis test (Kruskal and Wallis, 1952). In this test, data is ranked from all groups (in this case, methods) together regardless the method to which the performance value belong. The performance values are ordered from the lowest, assigning rank 1 to it, to the highest. In case of ties, the average of the ranks they would have received if they had not been tied is calculated and assigned to them. For instance, given the sample $\{150, 100, 120, 130, 120\}$, its order version is $100, 120, 120, 130, 150$ and its potential rank would be $1, 2, 3, 4, 5$. Nevertheless, there are ties and thereby the actual rank is $1, 2.5, 2.5, 4, 5$. Next, the ranks within each group are added. This sum for a particular method i is denoted as R_i. Then, the test statistic,

$$H = \frac{12}{\tilde{N}\left(\tilde{N}+1\right)} \sum_{i=1}^{\tilde{k}} \frac{R_i^2}{n_i} - 3\left(\tilde{N}+1\right), \tag{4.44}$$

where $\tilde{N}$ is the total sample size, $\tilde{k}$ is the number of groups in analysis and n_i is the sample size of group i, can be calculated. Under the null hypothesis, the performance of the methods came from the same population, the test statistic H is distributed according to a chi-square law, $\chi^2_{\tilde{k}-1}$, with $\tilde{k}-1$ degrees of freedom. The null hypothesis can be or not be accepted at a given significance level according to the correspondent *p-value* computed. In the case that the test is significant, a difference exists between the performance of the methods, at least, between two of them. In such scenario, a step-down follow-up analysis can be performed in order to identify each groups differ (Field, 2013). It starts by

ordering the methods based on the sum of ranks, from the smallest to largest. In case of ties, the median value is used for the ordering. Next, it is verified if a significant difference exists between the first group and the second. If no significant difference is found, the third group is compared with the previous two. The procedure is repeated for the following ranked methods until a significant difference is found. In such case, a set of homogeneous groups is found, composed by all the previous methods except the last group compared. The process re-starts for the excluded group and for the remaining ranked groups. The process ends when all the ranked groups have been tested (Field, 2013).

4.5.2. Results and Performance Evaluation

For the previously described datasets, the Wavelet Shrinkage, Singular Spectrum Analysis and Embedding, Wavelet Shrinkage and Diagonal Averaging were applied in order to extract information from the signals: noise from the signals comprising the synthetic dataset and variations associated with 'automatic' appliances from the whole-home electricity consumption signals. In the following, the performance evaluation, according to the metrics defined in Section 4.5.1, is reported. First, the results for the synthetic dataset are presented. Second, the 'best' mother wavelet for the decomposition and shrinkage of the aggregated electrical signal is explored. Finally, the results for the extraction of information from these electrical measurements by the three methods considered under the conditions defined in Section 4.5.1, *i.e.*, the necessary parameters setup, are described.

Synthetic Generated Dataset

The results report the improvement in terms of SNR values (the difference between the SNR associated with the denoised signal and the one associated with the initial signal). Figure 4.6 shows the box plots associated with the results accomplished with each defined scenario. Note that, considering the lower and upper quartile and for all the levels of noise tested, the EWD and Wavelet Shrinkage yielded very similar results, independently of the used level of decomposition. In fact, the average improvements for all the levels of initial noise were very similar for both approaches that are based on wavelet decomposition (Table 4.2).

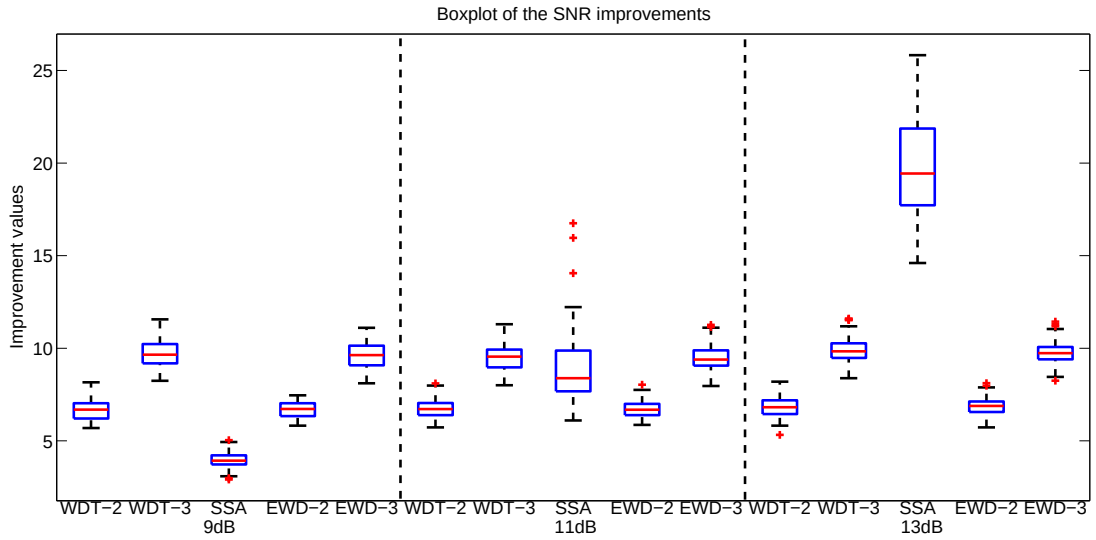

Figure 4.6.: Box plots of the improvements in the SNR achieved by the 3 different methods considering the 3 noise levels defined.

	9dB	**11dB**	**13dB**
WDT-2	6.663 ± 0.606	6.777 ± 0.568	6.819 ± 0.567
WDT-3	9.659 ± 0.755	9.569 ± 0.805	9.909 ± 0.804
SSA	3.972 ± 0.470	9.048 ± 2.180	19.530 ± 2.498
EWD-2	6.696 ± 0.430	6.745 ± 0.478	6.880 ± 0.497
EWD-3	9.621 ± 0.718	9.517 ± 0.769	9.770 ± 0.688

Table 4.2.: Descriptive statistics, mean and standard deviation, of the improvement values.

Considering the SSA performance, the computed arithmetic mean improvement for 9dB (3.92 ± 0.40dB) was smaller than for all the remaining ones and a small difference between the lower and the upper quartile is observed (*i.e.* 50% of the results have very similar values). However, for a noise level of 13dB, SSA was able to improve by 182.87% over EWD-2, 188.48% over WDT-2, 99.90% over EWD-3, and 97.09% over WDT-3. In this case, it achieves the highest average improvement: 19.530 ± 2.498dB. Moreover, the observed interquartile range indicates that the achieved improvements are very distinct, in contrast with the other levels of noise for which 50% of the results have very similar values. Note that for a initial noise level of 11dB and $J = 3$, all the methods presented similar results (SSA obtained an average improvement of 9.048 ± 2.180dB). The performance of the EWD and Wavelet Shrinkage appears not to be influenced by

	9dB	11dB	13dB
WDT-2	0.071	0.085	0.06
WDT-3	0.065	0.087	0.102
SSA	0.088	0.169	0.117
EWD-2	0.07	0.078	0.094
EWD-3	0.068	0.103	0.113

Table 4.3.: Kolmogorov-Smirnov test statistics $D\,(50)$ for the three levels of initial noise.

the initial noise level, while the SSA results vary accordingly to this parameter variation. Notice that the threshold value th_{ssa} was set to 90% and thereby 90% of variance of the initial signal is being kept which may lead to the low improvements observed for an initial noise of 9dB.

A similar behaviour between both methods based on wavelet decomposition was observed in all the cases tested and also between EWD-3 and SSA for 11dB. Then, statistical tests were performed using the SPSS statistical tool to analyse the significance of these similarities. The conditions under the tests validity were verified. For each initial noise level the results of EWD and WDT are normally distributed, according to the Kolmogorov-Smirnov test (Field, 2013) at a significance level of 0.01. The test statistics are presented in Table 4.3. The lower bound obtained by this test for its true significance was of 0.20 (*i.e.* $p\text{-}value > 0.20$) for all the cases, with the exception of EWD-3 and SSA for an initial noise level of 13dB for which the respective values were $p\text{-}value = 0.152$ and $p\text{-}value = 0.085$. These *p-values* indicate significant evidence in favour of the null hypothesis in test: the results follow a normal distribution. Still, for an initial noise level of 11dB, the results of SSA are far from normal since evidence in favour of the null hypothesis was low ($p\text{-}value = 0.001$). However, with such a small deviation from normal distribution we use the one-way ANOVA analysis (Field, 2013), which is robust test, considering no homogeneity of the variances, to validate the differences of averages between EWD, Wavelet Shrinkage and SSA.

The one-way ANOVA analysis considering no homogeneity of the variances indicates that for each initial level of noise a significant difference between the performances of the methods exists. In detail, at an initial noise of 9dB the test statistics is $F\,(4, 121) = 804.319$, $p\text{-}value < 0.001$, at an initial noise of 11dB, $F\,(4, 119.09) = 220.367$, $p\text{-}value < 0.001$, and at a noise level of 13dB, $F\,(4, 119.75) = 528.340$, $p\text{-}value < 0.001$. Since there are more than two groups

being analysed by the one-way ANOVA, further tests (post-hoc tests) are used to determine which pairs of groups would be significantly different. Thereby, the Tamhane's T2 post-hoc test (Hochberg and Tamhane, 1987) was run. According to the post-hoc test there is no statistically significant difference between the results produced at each decomposition level by the EWD and WDT, at a significance level of 0.01: $p\text{-}value = 1$ for each pairwise comparison EWD-2 vs. WDT-2 and EWD-3 vs. WDT-3, for all the different initial levels of noise. The exception was the pair EWD-3 vs. WDT-3 considering a initial noise level of 13dB ($p\text{-}value = 0.987$). Additionally, in the case of the 11dB initial noise, no statistical significant difference between the results produced by the SSA and EWD-3 ($p\text{-}value = 0.819$) and between SSA and WDT-3 ($p\text{-}value = 0.71$) is found. We can then conclude that the difference in the results of EWD and WDT is not significant. This indicates that the new method performs at least as good as WDT for this type of sinusoidal signal.

Household Electrical Signal Dataset

The same methods were also applied to a dataset composed by measurements of real energy consumption in a household. In the following, the 'best' mother wavelet for the Wavelet Shrinkage process of the aggregated electrical signal is empirically investigated. The performance results regarding the extraction of information from the electrical consumption by the three methods are presented and discussed.

Choosing the 'Best' Wavelet Function Figure 4.7 presents the number of occurrences of each wavelet function considered as best for information extraction from the regular days when compared with the information extracted from the vacation days via the Wavelet Shrinkage method. The wavelet functions that were selected as best less than five times were grouped together under the label "others". For both decomposition levels, the rbio3.1 function presents the highest frequency, which is significantly higher than the rest. We also observed that functions boir3.1 and rbio3.3 are appropriate for the designed task at both decomposition levels. Still, their number of occurrences is less than half of the rbio3.1 frequency. In addition, the wavelets db1.1 and coif5 were selected more than 5 times only for $J = 2$. These frequencies point out the existence of a particular set of functions showing a better adequacy for the extraction of information from the aggregated electrical signal in study. This may be related with

the wavelets waveform and its similarity with the waveform of the variations being extracted. Further exploration would be required to clarify this point.

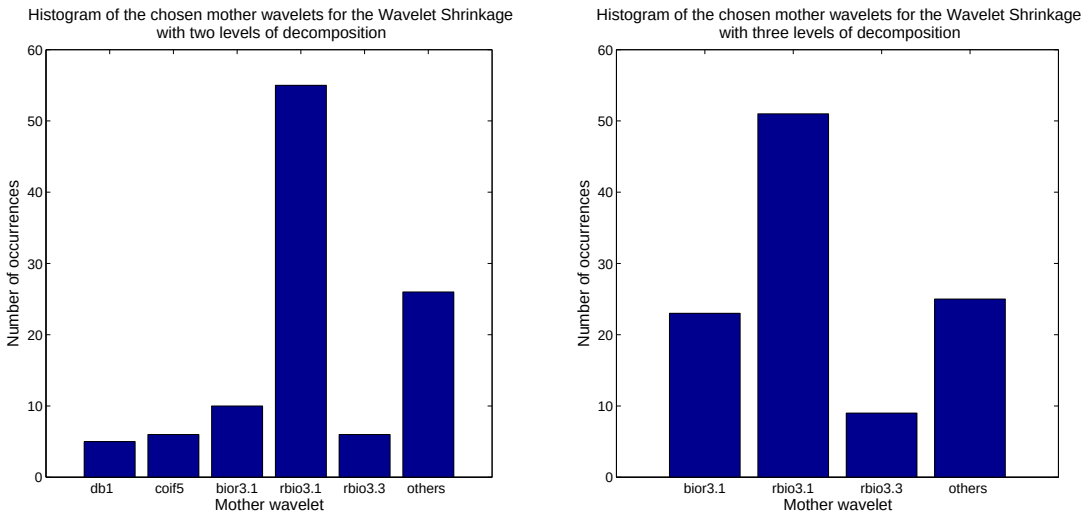

Figure 4.7.: Number of occurrences of each wavelet function considering the Wavelet Shrinkage with 2 and 3 decomposition levels at left and right, respectively.

A similar study was carried out for the EWD. To evaluate the initial intuition that different segments of the aggregated signals would require distinct mother wavelets to extract the variations associated with the 'automatic' appliances, the number of different mother wavelets selected *per* signal is analysed and reported in Figure 4.8. Indeed, the median number of distinct functions was of 10 and 8 for EWD-2 and EWD-3, respectively. Note that each signal was divided in 49 segments: the signals in study have length $T = 96$ and window length defined for the embedding procedure is of $L_w = 48$. Recalling Section 4.3.1, $T - L_w + 1 = 49$ lagged vectors were defined for each signal by the embedding procedure. Thereby, the median number of distinct wavelets indicates that each of the selected functions was employed at 5 or 6 segments of signal. For 50% of the signals, the number of distinct wavelets *per* signal ranged in $[7, 12]$ (EWD-2) and $[6, 11]$ (EWD-3).

Regarding the most selected wavelets, the observations are similar to the comments above drawn for the WDT-2 and WDT-3. We observe that the most selected functions by the Wavelet Shrinkage also occur with the highest frequency in EWD as illustrated in the histogram in Figure 4.9. Once again, the rbio3.1 function presents a significant high frequency indicating its adequacy for

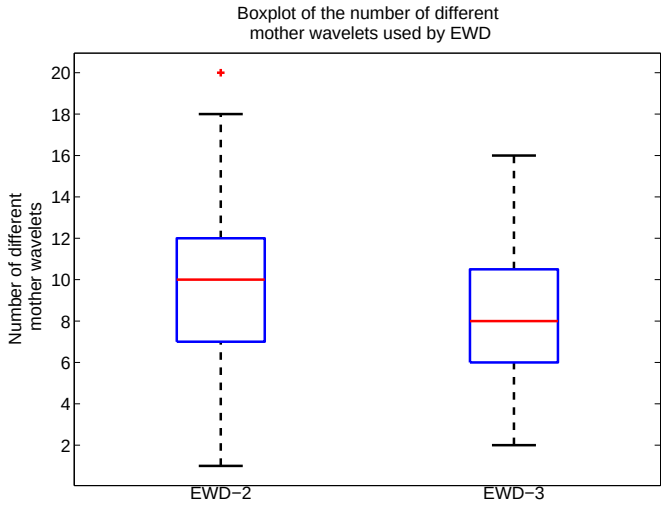

Figure 4.8.: The number of different mother wavelets used for each signal by the EWD with 2 and 3 decomposition levels at left and right, respectively.

the analysis of the aggregated consumption signal. In this case, the group 'others' in Figure 4.9 represents the mother wavelets that were chosen for less than one hundred and fifty distinct segments. Further investigation would be required to complement this empirical experimentation, namely the waveform similarity between the residual information of the vacation days and the wavelets functions selected.

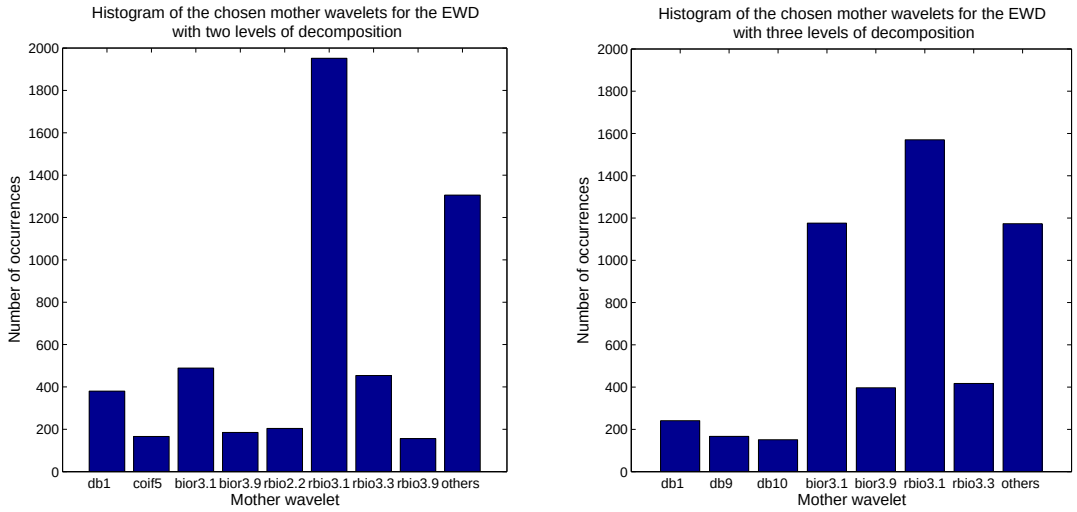

Figure 4.9.: Number of occurrences of each wavelet function considering the EWD with 2 and 3 decomposition levels at left and right, respectively.

Performance Evaluation As previously mentioned, this experiment is carried out using electrical consumption signals. Each sample reported the average active power in the house electrical circuit over 15 minutes. The results next presented report the best correlation values computed between the extracted information for a given regular day and the residual associated with the vacation days as described in Section 4.5.1. Figure 4.10 presents the box plots expressing the results achieved by each implemented method and version.

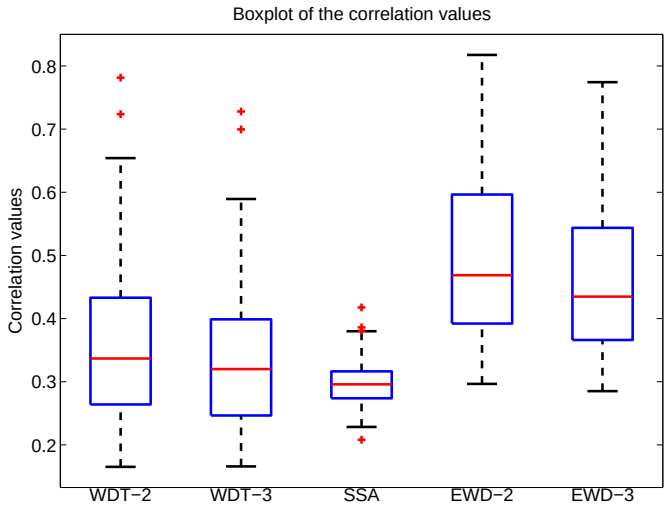

Figure 4.10.: Box plots of the correlation values achieved by the 3 different methods and corresponding versions.

The highest median correlation values are achieved by the EWD approach while the WDT-2, WDT-3 and SSA median correlation values are very similar. Note also that, the third quartile of WDT-2 is very close to the first quartile of EWD-2. A similar observation can be drawn for the correspondent quartiles of WDT-3 and EWD-3. As observed to the previous experience using the sinusoidal signal, the SSA present performance values that are very close across the several signals of the dataset as expressed by the small interquartile range. This difference between the first and third quartile is, in fact, the smallest among all the results yielded by the different methods. Moreover, the wavelet based approaches (Wavelet Shrinkage and EWD) a similar difference for 50% of the correlation values (*i.e.* the interquartile range) is observed.

In regard to the average correlation, EWD-2 achieved 0.492 ± 0.127 against 0.359 ± 0.123 for WDT-2. Note that the median value for WDT-2 is identical to the minimum yielded by the EWD-2. A similar remark can be made regarding

the EWD-3 minimum correlation and the median value achieved by the WDT-3. Moreover, EWD-3 average correlation was 0.125 higher than for WDT-3 (0.457 ± 0.116 against 0.332 ± 0.112). Within each wavelet based approach, the difference between the mean values for each decomposition levels is not significant: a value of 0.035 (0.027) differentiates the mean correlation of EWD-3 (WDT-3) from EWD-2 (WDT-2, respectively). In what concerns SSA, note that the threshold described in Section 4.4.2 set to 90% might influence negatively the achieved average correlation (0.299 ± 0.037) since 90% of the variance of the initial data is being kept. This average correlation value which, although very near to the mean value performance of WDT-3, was the smallest performance value of all the methods in study.

The response of the correlation values along the complete time series composed by the several days in the dataset was also observed. The performance of the approaches based on wavelet decomposition associated with the signals in the month of August is higher than for the other months, particularly for the Winter season. Note that, the selection of the most appropriate mother wavelet for extracting variations on regular days is based on information extracted from measurements of vacation days that occurred in August (Summer season). Figure 4.11 shows the correlations achieved by the WDT-2, SSA and EWD-2 along the four months of measurements. Here, only the values yielded by the WDT-2 and EWD-2 are presented since, and analogously with the previous results, it was observed that their performance is slightly better than for $J = 3$. As illustrated in the figure, with the arrival of Fall, the correlation between extracted information from regular days and the prototype signals decreases. This might be explained by behavioural changes in terms of energy consumption of the household family. Nevertheless, more information about the consumption, equipment or changes of behaviour would be needed to explore these assumptions, which was not available in the provided dataset.

With respect to the performance analysis of the proposed method, EWD, a statistical analysis was designed to validate the observations drawn. Again, the distributions of the results of the three methods and versions implemented were compared using SPSS. The Kolmogorov-Smirnov test was performed in order to validate the initial conditions of the performed tests. This test indicated that the results of EWD are far from a normal distribution, at a significance level of 0.01, since $D\,(107) = 0.104$, $p\text{-}value = 0.006$ concerning the EWD-2 and $D\,(107) = 0.108$, $p\text{-}value = 0.004$ for the results of EWD-3. In regard to the

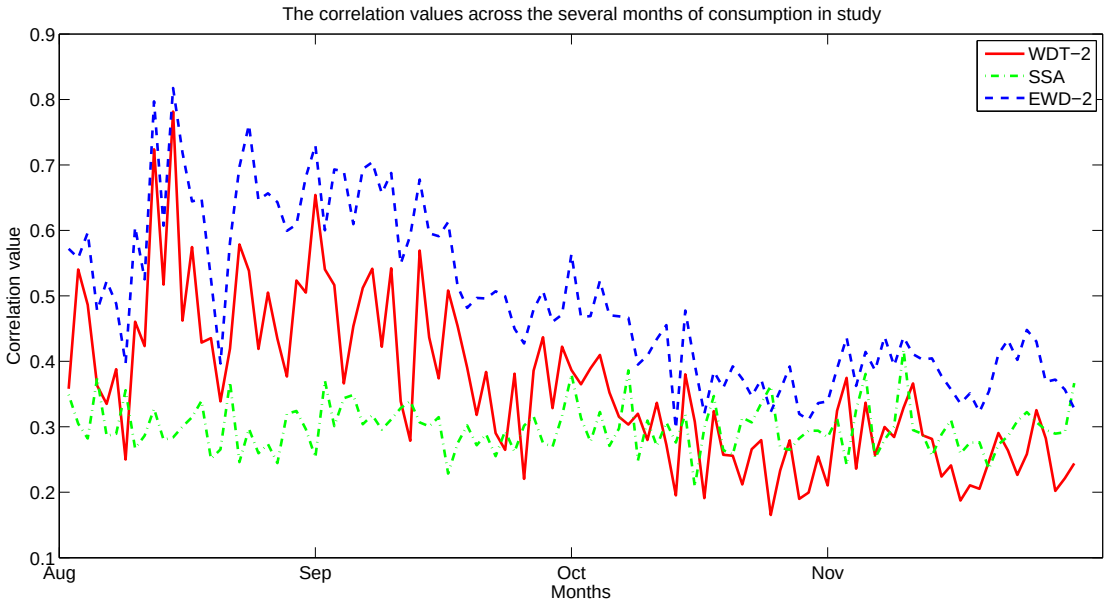

Figure 4.11.: The correlation values achieved by the 3 different methods along the four months of measurements.

Wavelet Shrinkage, there is a small evidence, at the same significance level, that the results follow a Gaussian distribution since $D\,(107) = 0.09$, $p\text{-}value = 0.03$ for both levels of decomposition. On the contrary, the SSA results follow a normal distribution, $D\,(107) = 0.07$, $p\text{-}value > 0.20$.

Given these normally test results, the non-parametric Kruskal-Wallis test was applied, followed by a multiple-comparison analysis. This test assesses if two or more samples come from the same population. The results indicate the existence of a statistical difference between the performance of EWD, Wavelet Shrinkage and SSA, $H\,(4) = 194.453$, $p\text{-}value < 0.001$, *i.e.*, the performance results come from different populations. The step-down follow-up analysis showed that no statistical difference exists between the results achieved by the different decomposition levels for the same method, $p\text{-}value = 0.211$ for the results of Wavelet Shrinkage and a $p\text{-}value = 0.095$ as obtained for the results of EWD. Additionally, no significant difference between the results of SSA and WDT-3 was found ($p\text{-}value = 0.40$). As a final conclusion, we observe that EWD yielded a moderate correlation between the residuals extracted from the signal and the vacation signals while both Wavelet Shrinkage and SSA achieved a weak correlation. Indeed, there exists a statistical difference between the results yielded by the EWD and the remaining methods. These statistical results support that EWD is a suited alternative to Wavelet Shrinkage and SSA for the

extraction of variations from electrical consumption signals.

4.6. Summary

In this chapter, the extraction of variations associated with 'automatic' equipment from the aggregated consumption electrical signal was investigated. For this preliminary separation of information, approaches based on wavelet decomposition and singular value decomposition were explored. The former is a widely studied transform in signal processing, for non-stationary signals in particular, due to its capability of providing temporal information which is a Fourier analysis shortcoming. The latter constitutes an important step of the SSA technique, usually applied for time series analysis. Based on wavelet transform and corresponding Wavelet Shrinkage method, the aggregated signal was also analysed considering several time windows where information was extracted.

Our experiments led to the identification of a small set of mother wavelets better suited for the extraction of the intended variations from the aggregated signal. However, further work would be required to complement this empirical experimentation by verifying the waveform similarity between the residual information of the vacation days and the wavelets functions selected. Additionally, the proposed approach accomplished the expected separation and its performance was superior to the Wavelet Shrinkage without time windows by 37% (considering two levels of decomposition) in terms of correlation between the separated variations and the prototypes of information extracted. A performance decrease along the four-months of measurements in a real-household was observed. An interesting research question would be that of providing some insight about how the correlations would decrease with time. To understand this phenomenon, prototypes for the electrical seasonal variations would be needed. This would open the way to establish a ground truth consumption for similar households. Future work directions would include the exploration of a source modelling framework to solve the problem at hand.

The investigated approaches, for the separation of the defined variations correspondent to 'automatic' operated devices (*e.g.* refrigerator), are relevant for areas of energy management or in-Home Activity Tracking casted in, for instance, an Ambient Assisted Living (AAL) framework. The outcome of such extraction would be a signal where the only variations are associated with

'manual' operated equipment. Thereby, the study and modelling of behaviours associated with the consumption of energy would be facilitated. In short, the extraction of appropriated information from the analysis of the loads occurring in the electrical network might represent a relevant step towards activity behavioural models. Furthermore, in a scenario here no full disaggregation of the whole-home signal would possible due to, for instance, the lack of required data, the proposed methodology would be a feasible alternative.

In the next chapter we will explore signal disaggregation into the loads associated with each device in the house, regardless of their operation mode as defined in this chapter. Rather than solving the energy disaggregation as a classification task as in Chapter 3, the problem is explored as a single-channel source separation and correspondent methods are investigated.

CHAPTER 5

Single-Channel Source Separation Problems

5.1. Introduction

Bearing in mind the literature review presented in Chapter 2, previous approaches of NILM were based on identification of electrical appliances' signatures from aggregated load monitored in residential buildings. In fact, the signatures based on the recognition of events (the step changes occurring in the active and reactive power associated with the switch on and switch off of the

devices), which were explored in the Chapter 3, have attracted a great deal of research interest so far. In the original NILM algorithm, the events, defined by the mentioned changes, were then clustered regarding the identification of the several appliances. Thereby, energy disaggregation is interpreted as a classification task. Notwithstanding, the whole-home consumption disaggregation and single-channel source separation problems coincide in their goals. Indeed, single-channel source separation aims at the estimation of individual sources from a single observed mixed signal which, in a NILM context, corresponds to the calculation of consumption estimates associated with each appliance connected to the electrical network, provided only that the whole-home electrical consumption measurements are given.

This chapter focus on non-intrusive energy disaggregation, casted in a NILM system, as a single-channel source separation problem. This reinterpretation of the problem was recently proposed by Kolter et al. (2010) and encouraged by this direction of research, approaches based on data-adaptive representations often employed for solving source separation problems by source modelling will be study in the following. Actually, Kolter et al. (2010) proposes a non-negative sparse coding based approach for electrical disaggregation. The methodology learns a sparse model for each device, which is followed by an adjustment phase to take into consideration the aggregated consumption measured. Finally, the learned models are used for the posterior separation of an aggregated signal. However, as these models are independently learned, dependencies among appliances may not be rightfully captured by the modelling process. Indeed, some devices are switched on simultaneously at particular instances of time, when users are performing particular activities (*e.g.* cooking), which could be relevant information for the disaggregation of the electricity consumption.

An alternative approach for solving energy disaggregation seen as single-channel source separation problem is proposed in this Thesis. Multi-way array (tensor) representation and correspondent factorization algorithms are explored for electrical source modelling. Tensors arise in a wide range of applications as signal analysis, computer vision, neuroscience and source separation (Kolda and Bader, 2009; Cichocki et al., 2009; Fitzgerald et al., 2005). However, no particular application to energy disaggregation is known so far. For this purpose, the collected consumption data of each electrical device for a given house is represented as a tensor and the data source model results from its non-negative factorization. This analysis intends to gather information along a given day and

across several days and appliances, and included it into a unique source model, able to be discriminative enough such that the energy consumption of each appliance is accurately computed. This model is then used to predict the power consumption of each device over a period of time where only the whole-home electrical consumption signal (aggregated signal) is measured.

This chapter is organized as follows. Next section, the formulation of energy disaggregation as a single-channel source separation problem is provided as well as background related literature. Additionally to the non-negative matrix factorization, multi-way arrays and their non-negative factorizations are described in Section 5.3. Next, single-channel source separation approaches for energy disaggregation are presented, in particular the sparse coding based approach and the alternative here proposed using tensor factorization. Following, the designed computational experiments, performed with real-world electrical consumption reference dataset, are detailed. Furthermore, in Section 5.5, results and statistical evidence are presented and discussed. At last, conclusions and directions of future work are addressed.

5.2. Single-Channel Source Separation and Energy Disaggregation

Source separation attempts to extract individual sources from observed inputs composed by a mixture of the several sources. Particular cases occur when (i) no *a priori* information is known about the source signals or the mixing process, named blind source separation (Bell and Sejnowski, 1995; Belouchrani et al., 1997; Cardoso, 1997; Gelle et al., 2003); or when (ii) only a single observed input exists, which is called of single-channel source separation (Jang and Lee, 2003; Schmidt and Mørup, 2006). As mentioned previously, the energy disaggregation can be interpreted as a single-channel source separation problem.

Formally, given a signal $\bar{x} \in \mathbb{R}^T$ which corresponds to the aggregated consumption during a period of time T, as defined in Chapter 2 Section 2.3, we intend to rewrite it as

$$\bar{x}(t) = f(x_1(t), \ldots, x_k(t)), \tag{5.1}$$

where $t = 1, \ldots, T$, and f is the mixing process of sources $x_i \in \mathbb{R}^T, i = 1, \ldots, k$, corresponding to the consumption of each device or circuit i (Equation 2.9). We

assume that the electrical consumption $\bar{x}$ is obtained by a linear mixing process, therefore, f is a combination of sources x_i and, consequently,

$$\bar{x}(t) = \sum_{i=1}^{k} x_i(t). \tag{5.2}$$

For a set of m daily observed signals, each column of $\bar{X} \in \mathbb{R}^{T\times m}$ represents the d-th aggregated consumption over the m-th day and each column of $X_i \in \mathbb{R}^{T\times m}$ expresses the m-th daily consumption signal associated with the device i. Thus, the aggregated consumption verifies

$$\bar{X} = \sum_{i=1}^{k} X_i. \tag{5.3}$$

In the above free-noise formulation, take the number of devices to be $k = 2$. Then, the resolution of $\bar{x} = s_1 + s_2$, where s_1 and s_2 are unknown, will require additional information about the sources. Single-channel source separation problems are thus under-determined and consequently, machine learning methods are suitable for their resolution. In this framework, specific knowledge associated with sources is formulated, namely, in terms of probability distribution, and inference of the most probable solution is performed (Schmidt, 2008).

According to Schmidt (2008), the related literature approaches usually rely on:

- filtering approaches, where estimates of the sources are produced by a set of filters (functions) applied to the mixture signal ($\bar{x}$);
- decomposition and grouping approaches, where the mixture signal is decomposed into several components that are properly grouped in a subsequent phase, resulting into estimates of the sources;
- source modelling approaches, where statistical models are formulated for both sources and mixing process which are posteriorly used to separate the sources.

Single-channel source separation by a modelling framework requires the mixed signal $\bar{x}$ and, for a supervised learning, training data, if available. Then, from the training set or from previous knowledge, models can be learned/defined

for both sources and mixing process (denoted before as f and considered here as a linear process in the energy disaggregation context). Posteriorly, given a new mixture signal, inference takes place: source models and mixing models are used to compute estimates of the separated sources. Additionally, for the disaggregation process a suitable representation of the signals may be required to enable separation. In such scenario, the source separation is not achieved directly from the measured signal. Instead, the signal is transformed into another representation and then source and mixing models are built. Consequently, in this case, signal reconstruction methods are needed for the inference of the sources (Schmidt, 2008).

Source and mixing models are defined in order to capture the essential properties of sources and mixing process to effectively allow its separation in the inference step. Matrix factorization is among the usual approaches for source modelling. Another is the Hidden Markov Models described in Chapter 2 Section 2.6. In matrix factorization approaches each source x_i, at a certain instant t, is modeled by a combination of a set of basis vectors

$$x_i(t) \approx \sum_{l=1}^{r} h_l w_l(t), \tag{5.4}$$

where w_l represents the bases collecting the main characteristics of the source and h_l is the correspondent activation (Schmidt, 2008). In general, given a matrix $X_i \in \mathbb{R}^{T \times m}$, the goal is to represent X_i by a factorization WH, such that $W \in \mathbb{R}^{T \times r}$ is a matrix of r bases and $H \in \mathbb{R}^{r \times m}$ is an m-dimensional set of activations. For that aim, X_i and WH must be as similar as possible. Therefore, methods as Independent Component Analysis (ICA) (Hyvärinen and Oja, 2000), NMF (Lee and Seung, 1999, 2000) and Sparse Coding (Olshausen and Field, 1996) are suitable for solving the problem.

ICA requires the assumption of statistical independence of the components. Thus, the linear representation of nongaussian data computed by the ICA comprises statistically independent (or as independent as possible) components (Hyvärinen and Oja, 2000). A growing interest has, in the last decade, raised the usage of ICA for the resolution of source separation problems, namely of the blind type (Hyvärinen and Oja, 2000; Choi et al., 2005). Moreover, ICA based approaches for achieving source separation when only single-channel measurements are given have also been studied in the related literature (Davies and James, 2007), aiming at blind single-channel source separation (Jang et al.,

2003).

Non-Negative Matrix Factorization (NMF) has also evoked researchers' interest due to its suitability and intuitive interpretation for real-world problems that usually deal with non-negative data (Cichocki et al., 2009). This method computes non-negative representations of non-negative matrices without additional requirements on data or on the computed components. Consequently, the related work covers a broad range of fields from the mathematical exploration (de Almeida, 2011) to the application in several fields as financial data mining (Ribeiro et al., 2009), text mining (Pauca et al., 2004), image classification (Guillamet et al., 2002), blind source separation (Cichocki et al., 2009) or single-channel source separation (Schmidt, 2008) among others.

Sparse Coding, a method inspired by the sensory information processing, *i.e.*, how brain represents the enormous amount of information provided at instant of time by our senses, is also a powerful tool in analysis and processing of signals (Olshausen and Field, 2004; Elad, 2012). In sparse models, information is represented by a relatively small number of simultaneously non-zero components, referred to as 'sparse coding'. In other words, the goal of a linear sparse coding is to find a decomposition for which activations are sparse and such that any given mixture signal can be well represented using only a few significantly non-zero activations (Hoyer, 2002). ICA and sparse coding are related when the computed coefficients are mainly zero and, then, sources can be represented by a sparse code (Schmidt, 2008; Hyvärinen and Oja, 2000). Furthermore, sparse extensions of the above described NMF have also been proposed (Hoyer, 2002, 2004). The approaches therein add a sparsity restriction to the intrinsic non-negative conditions and have been successfully applied to single-channel speech separation (Schmidt and Olsson, 2006).

Regarding the energy disaggregation context, approaches with non-negativity restrictions, as NMF, come up as the most suitable to solve the problem. As a matter of fact, electrical consumption is a non-negative quantity, thereby all the elements of $\bar{X}$, $\bar{X}'$, X_i and $X_i', i = 1, \ldots, k$, should be either zero or positive. Note also that for a supervised source modelling, matrices $\bar{X}$ and X_i are available at the training step, however, at the test step, only a new set of m' aggregated signals forming $\bar{X}' \in \mathbb{R}^{T \times m'}$ is accessible. At this point, the goal is to decompose the latter matrix into $X_i', i = 1, \ldots, k$, the components associated with each device (inference step). In the following, further details are presented regarding the NMF technique.

5.3. Matrices, Tensors and their Non-Negative Factorizations

Matrix factorization is the transformation of a given matrix into a canonical form (Weisstein, 2013). The two major applications of matrix factorization are (i) the resolution of a linear system of equations, and (ii) the presentation of the inherent structure of given data collected for objects observed over a set of variables and represented in form of a matrix (Hubert et al., 2000). The range of application of this mathematical powerful tool is wide: from time series analysis and denoise as described in Chapter 4, image processing (Lee and Seung, 1999) to dimensionality reduction (Jolliffe, 2002). Among the possible methods that differ in the constrains imposed to the resulting factors, *e.g.* PCA, SVD, ICA mentioned heretofore, NMF and its non-negative constraints suit many real-world data whose analysis rely on non-negative factorization (Cichocki et al., 2009).

Given a non-negative matrix $X \in \mathbb{R}_+^{T \times m^*}$, the NMF computes non-negative matrices $W \in \mathbb{R}_+^{T \times r}$ and $H \in \mathbb{R}_+^{r \times m^*}$, such that the reconstruction error between WH and X is minimum. Formally,

$$\begin{aligned} X &\approx WH && (5.5)\\ \textit{s.t.:}& \\ w_{tl} &\geq 0, \quad \forall\, t = 1, \ldots, T, \ \forall\, l = 1, \ldots, r, && (5.6)\\ h_{ld} &\geq 0, \quad \forall\, l = 1, \ldots, r, \ \forall\, d = 1, \ldots, m^*. && (5.7) \end{aligned}$$

To quantify the error the most commonly used measures are the Euclidean distance,

$$E\left(W, H\right) = \|X - WH\|^2 = \sum_{ij} \left(X_{ij} - \left(WH\right)_{ij}\right)^2, \tag{5.8}$$

and the divergence

$$D\left(X \,\|\, WH\right) = \sum_{ij} \left(X_{ij} \log \frac{X_{ij}}{\left(WH\right)_{ij}} - X_{ij} + \left(WH\right)_{ij}\right), \tag{5.9}$$

as proposed by Lee and Seung (1999, 2000). This optimization problem is not convex for both W and H. To overcome this issue, a usual approach is to optimize over one matrix while the other is fixed such as in the multiplicative

alternating algorithm proposed by Lee and Seung (1999, 2000).

As mentioned, NMF decomposition differs from other matrix decomposition techniques such as Principal Component Analysis (PCA) or Independent Component Analysis (ICA) in terms of the imposed requirements for factorization. All these methods compute matrices W and H, such that $X \approx WH$. However, while NMF imposes the non-negativity of the W and H, PCA demands the orthonormality of the columns of W and the orthogonality of the rows of H and ICA asks for the statistically independence of the rows of H. The combination of requirements leads to hybrid methods as the non-negative ICA (Plumbley, 2002, 2003; Plumbley and Oja, 2004), Non-Negative Sparse Coding or NMF with sparseness conditions as mentioned in the previous section. Actually, NMF *per se* enforces some sparseness while the sparse coding requires the definition of an intended sparsity degree. The combination of these techniques results in the Non-Negative Sparse Coding (NNSC) (Hoyer, 2002; Eggert and Korner, 2004) which demands W and H to be mostly composed by null elements. Formally, the non-negative sparse coding of a non-negative matrix X derive from the resolution of

$$\begin{aligned}
\min \quad & \frac{1}{2}\|X - WH\|^2 + \lambda \sum_{tl} H_{tl} && (5.10)\\
s.t.: & \\
w_{tl} &\geq 0, \quad \forall\, t = 1,\ldots,T, \ \ \forall\, l = 1,\ldots,r, && (5.11)\\
h_{ld} &\geq 0, \quad \forall\, l = 1,\ldots,r, \ \ \forall\, d = 1,\ldots,m^*, && (5.12)\\
\|w_{:l}\| &= 1 \quad \forall l = 1,\ldots,r, && (5.13)\\
\lambda &\geq 0 && (5.14)
\end{aligned}$$

where $w_{:l}$ denotes the *l-th* column of matrix W. The trade-off between sparseness and accurate reconstruction is controlled by the parameter λ (Hoyer, 2002). Moreover, the combination of non-negativity and sparseness constrains can be also thought as NMF with sparseness constraints (Hoyer, 2004). In this case, the associated optimization problem is

$$\begin{aligned}
\min \quad & \|X - WH\|^2 && (5.15)\\
s.t.: & \\
w_{tl} &\geq 0, \quad \forall\, t = 1,\ldots,T, \ \ \forall\, l = 1,\ldots,r, && (5.16)\\
h_{ld} &\geq 0, \quad \forall\, l = 1,\ldots,r, \ \ \forall\, d = 1,\ldots,m^*, && (5.17)
\end{aligned}$$

$$sparseness\,(w_{:l}) = S_w, \quad \forall\, l = 1, \ldots, r, \tag{5.18}$$

$$sparseness\,(h_{l:}) = S_h, \quad \forall\, l = 1, \ldots, r, \tag{5.19}$$

such that $h_{l:}$ the *l-th* row of H, S_w and S_h are the desired sparseness degrees for of W and H; using the objective function described by Equation 5.8 for the minimization. Hoyer (2004) proposes a projected gradient algorithm to solve the NMF with sparseness constraints presented in Algorithm 2. The algorithm takes a step in the direction of the negative gradient, followed by a projection on the constraint space, ensuring that the taken step is small enough that the error described by Equation 5.8 decreases. The parameters μ_W and μ_H used in the algorithm are small positive constraints, gradient step-sizes, updated by the implementation proposed by Hoyer (2004). Symbols $\circledast$ and $\oslash$ denote the elementwise multiplication (or Hadamard product) and division, respectively.

Although the decomposition usually considers a matrix (or a two-dimensional array), it can be extended to a multi-dimensional array also known as tensor or N-way tensor, such that N is the number of dimensions in the tensor, also known as modes. Formally, $\underline{Y} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ is a tensor of order N and the element $(i_1, i_2, \ldots, i_N)$ is denoted as $y_{i_1, i_2, \ldots, i_N}$. When a symbol of variable denotes a tensor, it will always be underscored.

In the case of matrices we refer columns and rows and, likewise, we can mention fibers and slices of a tensor. A fiber is the higher-order equivalent to rows and columns and is defined by fixing every tensor index excluding one. For instance, in a 3-order tensor, we can extract three different fibers: (i) columns $y_{:,i_2,i_3}$, (ii) rows $y_{i_1,:,i_3}$, and (iii) tubes $y_{i_1,i_2,:}$. Then, a fiber is a one-dimension section of a tensor. Similarly, two-dimension sections can be defined. These sections are obtained by fixing every tensor index except two of them. Again, in a 3-order tensor, we have three different sections: (i) horizontal slices $Y_{i_1,:,:}$, (ii) lateral slices $Y_{:,i_2,:}$ and (iii) frontal slices $Y_{:,:,i_3}$. In order to simplify the notation, the i_3-th frontal slice of a 3-order tensor will be denoted as Y_{i_3}. Additionally, the elements of a N-order tensor can be rearranged as a matrix. This process is known as unfolding, matricization or flattening. The unfolding associated with the n-th dimension is known as mode-n unfolding. By definition, the mode-n unfolding of a tensor $\underline{Y} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ re-arranges the mode-n fibers into columns of a matrix (Cichocki et al., 2009). The mode-n unfolding is denoted as

Algorithm 2: The Hoyer's NMF with sparseness constraints algorithm.

Data: $X \in \mathbb{R}_{+}^{T \times m'}$, r the number of components, S_w, S_h
Result: $W \in \mathbb{R}_{+}^{T \times r}$, $H \in \mathbb{R}_{+}^{r \times m'}$

Initialize W and H with positive numbers;
if *sparseness constraints on W apply* **then**
    Project each column of W to be non-negative, have unchanged $L2$ norm, but $L1$ norm set to achieve desired sparseness
end
if *sparseness constraints on H apply* **then**
    Project each row of H to be non-negative, have unit $L2$ norm, and $L1$ norm set to achieve desired sparseness
end
repeat
    if *sparseness constraints on W apply* **then**
        Set $W \leftarrow W - \mu_W \left(WH - X\right) H^T$
        Project each column of W to be non-negative, have unchanged $L2$ norm, but $L1$ norm set to achieve desired sparseness
    end
    else
        $W \leftarrow W \circledast \left(VH^T\right) \oslash \left(WHH^T\right)$
    end
    if *sparseness constraints on H apply* **then**
        Set $H \leftarrow H - \mu_H \left(WH - X\right)$
        Project each row of H to be non-negative, have unit $L2$ norm, and $L1$ norm set to achieve desired sparseness
    end
    else
        $H \leftarrow H \circledast \left(W^T X\right) \oslash \left(W^T WH\right)$
    end
until *no improvement occurs in the objective function*;

$Y_{(n)}$. For a 3-order tensor $\underline{\hat{Y}} \in \mathbb{R}^{3\times4\times2}$, whose frontal slices are

$$\hat{Y}_1 = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{bmatrix} \tag{5.20}$$

and

$$\hat{Y}_2 = \begin{bmatrix} 13 & 14 & 15 & 16 \\ 17 & 18 & 19 & 20 \\ 21 & 22 & 23 & 24 \end{bmatrix}, \tag{5.21}$$

the three mode-n unfolding matrices would be

$$\hat{Y}_{(1)} = \begin{bmatrix} 1 & 2 & 3 & 4 & 13 & 14 & 15 & 16 \\ 5 & 6 & 7 & 8 & 17 & 18 & 19 & 20 \\ 9 & 10 & 11 & 12 & 21 & 22 & 23 & 24 \end{bmatrix}, \tag{5.22}$$

$$\hat{Y}_{(2)} = \begin{bmatrix} 1 & 5 & 9 & 13 & 17 & 21 \\ 2 & 6 & 10 & 14 & 18 & 22 \\ 3 & 7 & 11 & 15 & 19 & 23 \\ 4 & 8 & 12 & 16 & 20 & 24 \end{bmatrix}, \tag{5.23}$$

$$\hat{Y}_{(3)} = \begin{bmatrix} 1 & 5 & 9 & 2 & 6 & 10 & 3 & 7 & 11 & 4 & 8 & 12 \\ 13 & 17 & 21 & 14 & 18 & 22 & 15 & 19 & 23 & 16 & 20 & 24 \end{bmatrix}. \tag{5.24}$$

Further details about tensor properties and bases of tensor algebra can be found in (Kolda and Bader, 2009; Cichocki et al., 2009).

As mentioned, like matrices also tensors can be decomposed. Tucker's decomposition (Tucker, 1966) and PARAFAC (Carroll and Chang, 1970; Harshman, 1970), also known as Canonical Decomposition (CANDECOMP), are the most used decompositions. Tucker (1966) describes a decomposition of a N-way array into a core tensor multiplied by factor matrices $A^{(n)}, n = 1, \ldots, N$, with a possible different number of components for each mode. A more restrictive approach is the PARAFAC factorization (Kolda and Bader, 2009; Cichocki et al., 2009). The PARAFAC technique decomposes a N-order tensor $\underline{Y} \in \mathbb{R}^{I_1\times I_2\times\ldots\times I_N}$ into factors $A^{(n)} \in \mathbb{R}^{I_n\times R}$ for $n = 1, \ldots, N$, for a given fixed number of components R. In particular for a 3-order tensor, $\underline{X} \in \mathbb{R}^{T\times m\times k}$, the method decomposes it

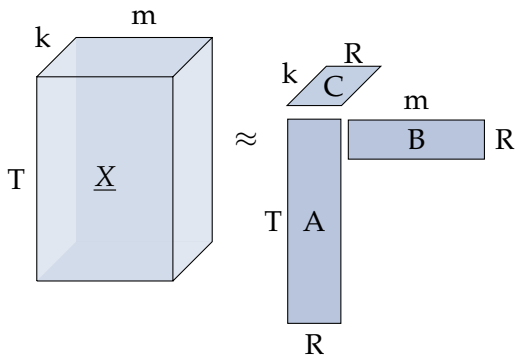

Figure 5.1.: Illustration of the PARAFAC decomposition method for a three-dimensional tensor.

into factors $A \in \mathbb{R}^{T\times R}$, $B \in \mathbb{R}^{m\times R}$ and $C \in \mathbb{R}^{k\times R}$, given $R \in \mathbb{N}$ such that

$$\underline{X} \approx \sum_{l=1}^{R} a_l \circ b_l \circ c_l = [\![A, B, C]\!], \tag{5.25}$$

or likewise

$$\underline{X} \approx \sum_{l=1}^{R} \breve{\lambda}_l a_l \circ b_l \circ c_l, \tag{5.26}$$

where $a_l \in \mathbb{R}_+^T$, $b_l \in \mathbb{R}_+^m$, $c_l \in \mathbb{R}_+^k$, for $l = 1, \ldots, R$, $\breve{\lambda} \in \mathbb{R}^R$ is a vector of weights and $\circ$ represents the vector outer product. Equivalently, by this decomposition, each element of $\underline{X}$ can be re-written as

$$x_{t,d,i} \approx \sum_{l=1}^{R} a_{tl} b_{dl} c_{il}, \tag{5.27}$$

for $t = 1, \ldots, T$, $d = 1, \ldots, m$ and $i = 1, \ldots, k$. Moreover, the i-th frontal slice of $\underline{X}$ can be approximated by

$$X_i \approx A D_i B^T, \tag{5.28}$$

where D_i is a diagonal matrix based on the i-th row of C. Nevertheless, this representation in terms of tensor slices is not possible for N-order tensors when $N > 3$ (Cichocki et al., 2009). Note also that each column of $\tilde{X}_i = AD_iB^T$ contains the m-th reconstructed signal for the source i. The PARAFAC decomposition of a 3-order tensor is illustrated in Figure 5.1.

The most usual approach for the computation of the factors $A \in \mathbb{R}^{T\times R}$, $B \in \mathbb{R}^{m\times R}$ and $C \in \mathbb{R}^{k\times R}$, is the ALS method as proposed in (Carroll and Chang, 1970; Harshman, 1970). This technique fixes two of the matrices, *e.g.*, B

and C, and outputs an A matrix admissible for the problem; then, other two are fixed, for example, A and C and the problem is solved for B and, lastly, A and B are fixed to solve for C. This process is repeated until convergence is achieved.

Regarding Nonnegative Tensor Factorization (NTF), the optimization problem is modified such that non-negative constrains are added, this is, factors $A \in \mathbb{R}_+^{T\times R}$, $B \in \mathbb{R}_+^{m\times R}$ and $C \in \mathbb{R}_+^{k\times R}$ are computed. Typically, the objective function is

$$\min \|\underline{X} - [\![A, B, C]\!]\|_F^2 + \alpha_A \|A\|_F^2 + \alpha_B \|B\|_F^2 + \alpha_C \|C\|_F^2 , \tag{5.29}$$

where $\|\bullet\|_F$ represents the Frobenius norm, and α_A, α_B and α_C are nonnegative regularization parameters (Cichocki et al., 2009). Again, the prevalent approach to compute the non-negative factors associated with the PARAFAC decomposition is based on an ALS technique whose update rules of a three-order array are shown in Algorithm 3 (Cichocki et al., 2009). In Algorithm 3, I denotes the identity matrix, $\odot$ denotes the Khatri-Rao product and $\circledast$ denotes the elementwise multiplication. By definition, the Khatri-Rao product of matrices

Algorithm 3: The ALS algorithm to compute the PARAFAC decomposition with non-negative constrains of a three-dimensional tensor.

Data: $\underline{X} \in \mathbb{R}_+^{T\times m\times k}$, R the number of components
Result: $A \in \mathbb{R}_+^{T\times R}$, $B \in \mathbb{R}_+^{m\times R}$, $C \in \mathbb{R}_+^{k\times R}$
1 Initialize A, B and C randomly;
2 **repeat**
3 Set $A \leftarrow \left(X_{(1)} (C \odot B) \left((C^T C) \circledast (B^T B) + \alpha_A I\right)^{-1}\right)_+$,
4 Set $B \leftarrow \left(X_{(2)} (C \odot A) \left((C^T C) \circledast (A^T A) + \alpha_B I\right)^{-1}\right)_+$,
5 Set $B \leftarrow \left(X_{(3)} (B \odot A) \left((B^T B) \circledast (A^T A) + \alpha_C I\right)^{-1}\right)_+$,
6 **until** *no improvement occurs in the objective function or maximum iterations exhausted*;

$A = [a_1 \dots a_k] \in \mathbb{R}^{T\times k}$ and $B = [b_1 \dots b_k] \in \mathbb{R}^{m\times k}$ results into a matrix of size $(Tm) \times k$ described as

$$A \odot B = [a_1 \otimes b_1 \;\; a_2 \otimes b_2 \;\; \dots \;\; a_k \otimes b_k] , \tag{5.30}$$

where $\otimes$ denotes the Kronecker product. Once again, given matrices $A \in \mathbb{R}^{I\times J}$ and $B \in \mathbb{R}^{T\times L}$, their Kronecker product is a matrix of size $(IT) \times (JL)$ defined

by

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \dots & a_{1J}B \\ a_{21}B & a_{22}B & \dots & a_{2J}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{I1}B & a_{I2}B & \dots & a_{IJ}B \end{bmatrix} \tag{5.31}$$

$$= \begin{bmatrix} a_1 \otimes b_1 & a_1 \otimes b_2 & a_1 \otimes b_3 & \dots & a_J \otimes b_{L-1} & a_J \otimes b_L \end{bmatrix}. \tag{5.32}$$

The Khatri-Rao product can be seen as the 'matching columnwise' Kronecker product (Kolda and Bader, 2009).

In this Thesis we use a PARAFAC model with non-negativity constraints, *i.e.*, the factors A, B and C are composed by non-negative elements. Nevertheless, other approaches to perform the non-negative three-order factorization have been proposed, like an extension of the 2-way positive matrix factorization method to the three-way PARAFAC model (Paatero, 1997a,b; Cichocki et al., 2009).

5.4. Single-Channel Source Separation Approaches for Energy Disaggregation

In this section, a method from the related literature to solve the energy disaggregation task as a single channel-source separation problem that makes use of the source modelling framework is presented. The source models are then adjusted considering the overall energy consumption. The source models are independently learned which may lead to the fact that dependencies among appliances that are jointly operated at specific periods of the day may not be captured. This chapter proposes an alternative approach for solving the energy disaggregation that, also in light of source separation problem, learns a 'global' source model (in the sense that no independent appliances' source models are computed) by means of multi-way arrays.

5.4.1. Sparse Coding for Energy Disaggregation

The algorithmic method proposed by Kolter et al. (2010), known as Discriminative Disaggregation Sparse Coding (DDSC), relies on a NNSC defined in Equations 5.10 – 5.14 (Hoyer, 2002; Eggert and Korner, 2004). The DDSC tech-

nique is composed by three main steps: (i) sparse coding pre-training, (ii) discriminative disaggregation training and (iii) test. The pre-training consists in modelling each source by NNSC, *i.e.*, consists in solving

$$\min \frac{1}{2} \|X_i - B_i A_i\|_F^2 + \lambda \sum_{p=1,q=1}^{r,m} (A_i)_{pq}, \quad A_i \in \mathbb{R}_+^{r \times m}, B_i \in \mathbb{R}_+^{T \times r}, \tag{5.33}$$

where the columns of $B_i \in \mathbb{R}_+^{T \times r}$ represent an r basis (dictionary), the columns of $A_i \in \mathbb{R}_+^{r \times m}$ represent the activations (sparse codes) of this dictionary, the regularization parameter λ represents the sparseness degree presented in the solution and $\|\bullet\|_F$ represents the Frobenius norm. This optimization problem is solved using a coordinate descent approach described in (Friedman et al., 2007, 2010) for the activations and the multiplicative NMF update proposed by Eggert and Korner (2004) for the basis calculations. Next, a re-optimization of the source bases is performed at an intermediate step. It is based on structured prediction methods and it is introduced since the bases of each model were not computed in order to minimize the disaggregation error

$$\text{Disaggregation Error} = \sum_{i=1}^{k} \frac{1}{2} \left\| X_i - \hat{X}_i \right\|_F^2, \tag{5.34}$$

where X_i is the matrix of measured signals for equipment i, $\hat{X}_i$ is its predicted version and $\|\bullet\|_F$ is the Frobenius norm (Kolter et al., 2010). For that purpose, the discriminative disaggregation training incorporates data supplied by $\bar{X}$ in the bases $B_i, i = 1, \ldots, k$, considering

$$\hat{A}_{1:k} = \arg\min_{A_{1:k}} \left\| \bar{X} - [B_1 \ldots B_k] \begin{bmatrix} A_1 \\ \vdots \\ A_k \end{bmatrix} \right\|_F^2 + \lambda \sum_{i=1,p=1,q=1}^{k,r,m} (A_i)_{p,q}, \tag{5.35}$$

where $X_{1:k}$, $B_{1:k}$ and $A_{1:k}$ represent $X_1, \ldots, X_k$, $B_1, \ldots, B_k$ and $A_1, \ldots, A_k$, respectively. $\hat{A}_1, \ldots, \hat{A}_k$ are the activations associated with the aggregated signal. The best value for $\hat{A}_i$ is $A_i^* = A_i$, thereby, $\hat{A}_{1:k}$ and $A_{1:k}^*$ should be as close as possible. For this purpose, $B_{1:k}$ is re-calculated by

$$\tilde{B}_{1:k} \leftarrow \tilde{B}_{1:k} - \tilde{\alpha} \left(\left(\bar{X}_{1:k} - \tilde{B}_{1:k} \hat{A}_{1:k} \right) \hat{A}_{1:k}^T - \left(\bar{X}_{1:k} - \tilde{B}_{1:k} A_{1:k}^* \right) {A_{1:k}^*}^T \right), \tag{5.36}$$

resulting in the disaggregation bases $\tilde{B}_{1:k}$, where $\tilde{\alpha}$ is the step size. Moreover, the reconstruction bases $B_{1:k}$, learned in the pre-training, are replaced by the disaggregation bases $\tilde{B}_{1:k}$ in Equation 5.35. Lastly, given a set of aggregated signals $\bar{X}'$ and the computed bases, the correspondent activations are calculated, followed by the consumption associated with each device. Algorithm 4 illustrates the approach proposed by Kolter et al. (2010).

Algorithm 4: The DDSC algorithm. Source Kolter et al. (2010).

Data: $X_i, i = 1, \ldots, k,\ \bar{X}, \bar{X}',\ \lambda \in \mathbb{R}_+,\ \tilde{\alpha} \in \mathbb{R}_+\ r \in \mathcal{N}, e \in \mathbb{R}_+$
Result: $\hat{X}'_1, \ldots, \hat{X}'_k \in \mathbb{R}^{T \times m'}$
`/* Pre-training:` */
Initialize B_i and A_i with positive values ;
Columns of B_i must have unit norm;
for *each* $i = 1, \ldots, k$ **do**
  repeat
    Set $A_i \leftarrow \arg\min_{A_i \geq 0} \|X_i - B_i A\|_F^2 + \lambda \sum_{p,q} A_{p,q}$;
    Set $B_i \leftarrow \arg\min_{B_i \geq 0} \|X_i - B A_i\|_F^2$;
  until *no improvement occurs in the objective function*;
end
`/* Discriminative disaggregation training:` */
Set $A^*_{1:k} \leftarrow A_{1:k}, \tilde{B}_{1:k} \leftarrow B_{1:k}$;
repeat
  Set $\hat{A}_{1:k} \leftarrow \arg\min_{A_{1:k}} F\left(\bar{X}, \tilde{B}_{1:k}, A_{1:k}\right)$;
  Set $\tilde{B} \leftarrow \left[\tilde{B} - \tilde{\alpha}\left(\left(\bar{X} - \tilde{B}\hat{A}\right)\hat{A}^T - \left(\bar{X} - \tilde{B}A^*\right)A^{*T}\right)\right]_+$;
  for *all the columns of* $\tilde{B}$, $\tilde{b}_i^{(j)}$ **do**
    $\tilde{b}_i^{(j)} \leftarrow \frac{\tilde{b}_i^{(j)}}{\left\|\tilde{b}_i^{(j)}\right\|_2}$
  end
until *no improvement in the objective function*;
`/* Test:` */
$\hat{A}'_{1:k} \leftarrow \arg\min_{A_{1:k}} F\left(\bar{X}', \tilde{B}_{1:k}, A_{1:k}\right)$;
Predict $\hat{X}'_i = B_i \hat{A}'_i$;

5.4.2. Non-Negative Tensor Factorization for Energy Disaggregation

Motivated by the multi-dimensional representation, the rationale behind the approach next proposed, Source Separation via Tensor and Matrix Factorization (STMF), is to explore a 'global' source model considering all the devices at once. This framework takes advantage of a multi-way array representation and associated factorization methods in addition to NMF for the disaggregation. In particular, a source model is learned by tensor decomposition and, at the inference phase, an aggregated signal is separated employing a NMF method and the factors resulting from the previous step.

Consider that the consumption associated with k appliances is represented as an 3-order tensor where each frontal slice of $\underline{X} \in \mathbb{R}^{T \times m \times k}$ is a matrix X_i representing the electrical consumption of device i, during m days, with T samples by day, as illustrated in Figure 5.2. The data in tensor $\underline{X}$ ranges in three different domains: time, days and devices. The non-negative tensor factorization of $\underline{X}$ via PARAFAC with non-negativity constrains and with R components models each source i at a certain instant t for a given day d as

$$x_{t,d,i} \approx \sum_{l=1}^{R} a_{tl} b_{dl} c_{il}. \tag{5.37}$$

At this point, the electrical consumption tensor is described by multiple components in time, days and device domains.

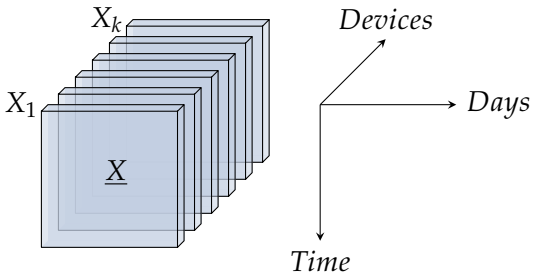

Figure 5.2.: Illustration of the three-way tensor in the STMF approach.

Recall that the aggregated power consumption is a linear combination of sources (Equation 5.3) and that NTF can be re-written in terms of frontal slices (Equation 5.28), which in this case correspond to the estimated consumption of

each device,

$$\tilde{X}_i = AD_iB^T. \tag{5.38}$$

Then, the aggregated energy $\bar{X}$ can be expressed in terms of factors the A, B and C. In detail,

$$\bar{X} \equiv \sum_{i=1}^{k} X_i \approx \sum_{i=1}^{k} \tilde{X}_i = \sum_{i=1}^{k} \left(AD_iB^T \right) = A \left(\sum_{i=1}^{k} D_i \right) B^T. \tag{5.39}$$

To achieve the separation of new m' aggregated signals previously, $\bar{X}' \in \mathbb{R}_+^{T \times m'}$, into the consumption of each device, $\hat{X}'_1, \ldots, \hat{X}'_k \in \mathbb{R}_+^{T \times m'}$, we need to decompose it accordingly. Since $\bar{X}'$ is the only available information at this point, NMF techniques are the most suitable. Non-negative matrices $W \in \mathbb{R}_+^{T \times R}$ and $H \in \mathbb{R}_+^{R \times m'}$ are computed in order to minimize the reconstruction error between WH and $\bar{X}'$ as described in Section 5.3.

The mere non-negative factorization of $\bar{X}'$ without further knowledge about the sources could not lead to $\hat{X}'_i, i = 1, \ldots, k$. Hence, the previously learned factors should be included in the decomposition calculation, in particular, the factors associated with the time and device domains, *i.e.*, matrices A and C and the correspondent matrices, $D_i, i = 1, \ldots, k$. Similarly to Expression (5.39), a factor $\tilde{H}$ should be computed such that

$$\bar{X}' \approx A \left(\sum_{i=1}^{k} D_i \right) \tilde{H}. \tag{5.40}$$

Still A was computed in order to minimize the error described in Equation 5.29. Additionally, a dependency between the factor associated with the 'time' domain, A, and the factor associated with the 'days' domain exists. Thereby, when optimizing over $\tilde{H}$ the factor A should be 're-adjusted' while the factor associated with 'devices' remains fixed. Consequently, the factors $\tilde{W}$ and $\tilde{H}$ are computed such that

$$\min E' \left(\tilde{W}, \tilde{H} \right) = \min \left\| \bar{X}' - \tilde{W} \left(\sum_{i=1}^{k} D_i \right) \tilde{H} \right\|^2, \tag{5.41}$$

with respect to $\tilde{W}$ and $\tilde{H}$, subject to $\tilde{W}, \tilde{H} \geq 0$, where $\tilde{W}$ and $\tilde{H}$ were initialized as A and as a random matrix with positive values, respectively. The strategy to solve this problem was similar to the usual approach for solving NMF:

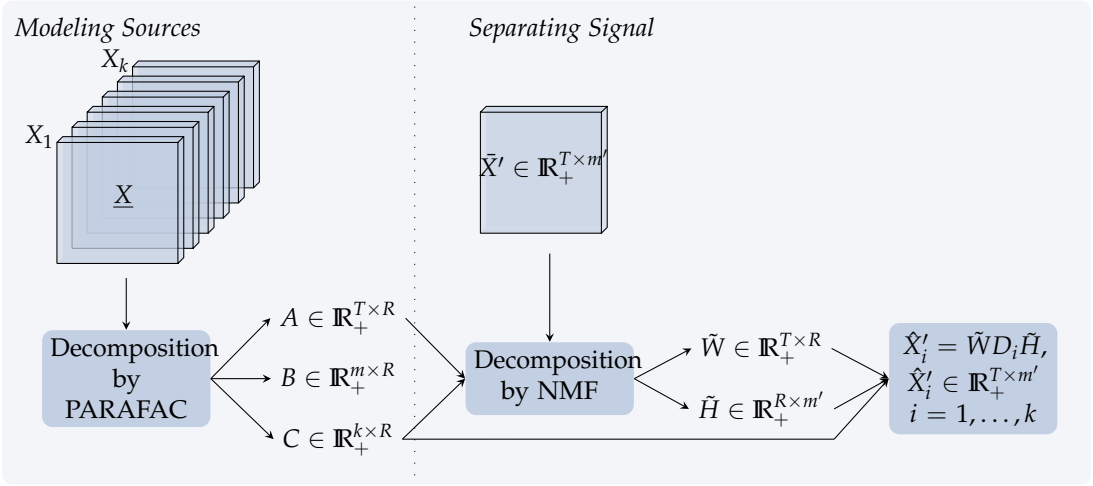

Figure 5.3.: Illustration of the STMF approach.

optimize over one matrix, while the other is fixed. While optimizing over $\tilde{W}$, the factor $\left(\sum_{i=1}^{k} D_i\right) \tilde{H}$ remains fixed and, while optimizing over $\tilde{H}$, the factor $\tilde{W} \left(\sum_{i=1}^{k} D_i\right)$ remains fixed. Finally, the consumption of appliance i is calculated as $\hat{X}_i' = \tilde{W} D_i \tilde{H}$ for $i = 1, \ldots, k$.

To keep the sparseness degree of $\tilde{W}$ and $\tilde{H}$ close to those of A and B^T, sparseness constraints were added to the problem. Thereby, the STMF approach incorporates a non-negative matrix factorization with sparseness restrictions, described in Section 5.3. The sparseness of a given vector $\bar{v} \in R^{\bar{n}}$ can be measured by

$$sparseness\,(\bar{v}) = \frac{\sqrt{\bar{n}} - \frac{(\sum |\bar{v}_i|)}{\sqrt{\sum \bar{v}_i^2}}}{\sqrt{\bar{n}} - 1} \tag{5.42}$$

as proposed by Hoyer (2004). This measurement is used in Algorithm 2, whose updates are employed in the STMF approach to solve the problem with sparseness conditions, *i.e.*, the calculation of $\tilde{W}$ and $\tilde{H}$. The proposed Source Separation via Tensor and Matrix Factorization approach is summarized in Algorithm 5 and illustrated in Figure 5.3.

5.5. Computational Experiments

This section describes the experimental setup and the accomplished results of the performed computational experiments. The approaches described above, the Discriminative Disaggregation Sparse Coding (DDSC) method and the

Algorithm 5: The STMF algorithm.

Data: $X_i \in \mathbb{R}_+^{T\times m}, i = 1,\ldots,k,\ \bar{X} \in \mathbb{R}_+^{T\times m},\ \bar{X}' \in \mathbb{R}_+^{T\times m},\ R \in \mathbb{N}, \epsilon \in \mathbb{R}_+$

Result: $\hat{X}_i' \in \mathbb{R}_+^{T\times m'}, i = 1,\ldots,k$

`/* Training:` */

Set $\underline{X}$ as a tensor such frontal slices are $X_i, i = 1,\ldots,k$;

Compute the non-negative tensor factorization of $\underline{X}$ via PARAFAC model (A, B and C such $x_{t,d,i} \approx \sum_{l=1}^{R} a_{tl} b_{dl} c_{il}$ for $t = 1,\ldots,T$, $d = 1,\ldots,m$ and $i = 1,\ldots,k$): Initialize A, B and C randomly;

repeat

  Set $A \leftarrow \left(X_{(1)}\left(C \odot B\right)\left(\left(C^T C\right) \circledast \left(B^T B\right) + \alpha_A I\right)^{-1}\right)_+$,

  Set $B \leftarrow \left(X_{(2)}\left(C \odot A\right)\left(\left(C^T C\right) \circledast \left(A^T A\right) + \alpha_B I\right)^{-1}\right)_+$,

  Set $B \leftarrow \left(X_{(3)}\left(B \odot A\right)\left(\left(B^T B\right) \circledast \left(A^T A\right) + \alpha_C I\right)^{-1}\right)_+$,

until *no improvement occurs in the objective function or maximum iterations exhausted*;

`/* Test:` */

Initialize $\tilde{H}$ with random positive values;

Set $\tilde{W} \leftarrow A$;

Set $v_0 \leftarrow \left\|\bar{X}' - \tilde{W}\left(\sum_{i=1}^{k} D_i\right)\tilde{H}\right\|$;

Compute vector $sparsenessA$, the sparseness for each column of A;

Set $sparsenessW = mean\{sparsenessA\}$;

Compute vector $sparsenessB$, the sparseness for each column of B^T;

Set $sparsenessH = mean\{sparsenessB\}$;

repeat

  Set $W \leftarrow \tilde{W}\left(\sum_{i=1}^{k} D_i\right)$;

  $\tilde{H} \leftarrow argmin_{\tilde{H}\geq 0}\left\|\bar{X}' - W\tilde{H}\right\|$;

  Set $H \leftarrow \left(\sum_{i=1}^{k} D_i\right)\tilde{H}$;

  $\tilde{W} \leftarrow argmin_{\tilde{W}\geq 0}\left\|\bar{X}' - \tilde{W}H\right\|$;

  $v_j = \left\|\bar{X}' - \tilde{W}\left(\sum_{i=1}^{k} D_i\right)\tilde{H}\right\|$;

until *no improvement in the objective function* $\left(\left|v_j - v_{j-1}\right| < \epsilon\right)$;

Predict $\hat{X}_i' = \tilde{W} D_i \tilde{H}$ such D_i is a diagonal matrix based on the i-th row of C

Source Separation by Tensor and Matrix Factorization (STMF) approach are evaluated for the task of estimate the electrical consumption of each appliance in a measured household electrical circuit. The experiments are yielded using a real-world data from a domestic household environment, where the energy consumption was daily recorded during a given period of time. First, a study regarding the sensibility analysis of each method is carried out. In particular, the influence of the regularization parameter in the performance of the DDSC approach is explored. Second, the two approaches are compared regarding the computed estimates of each source, *i.e.*, the consumption associated with the electrical devices in study.

5.5.1. Experimental Setup

In the following, the designed computational experience is detailed: the dataset is described, evaluation metrics and experimental parameters are presented and, additionally, the statistical test used to validate the performance comparison is introduced. Both DDSC and STMF were implemented using MATLAB. The DDSC was implemented according to the description found in (Kolter et al., 2010) and, for the STMF implementation, we used the N-way Toolbox (Andersson and Bro, 2000). For both approaches, the maximum number of iterations was set to 1000 and the error threshold for the condition stop ϵ to 0.00001.

Dataset Description

For the evaluation of both methods performance, we used the publicly available Reference Energy Disaggregation Dataset (REDD) (Kolter and Johnson, 2011; Kolter, 2011) consisting of on-site measurements obtained from sensors that measured the aggregated and circuit/device specific electricity consumption of 6 real houses, over several months' time. For each monitored house, the whole-home electricity signal and up to 24 individual circuits in the home, each labelled with its appliance category, were recorded. Since errors may occur leading to missing data, a preprocessing phase that selected only the mutual sampling periods to both aggregated and individual signals was carried out. The chosen signals were then down sampled by a median filter to enforce a sampling rate of $\frac{1}{60}Hz$.

Regarding the measured devices, data of similar labelled circuits have been

		House 1	House 2	House 3	House 4	House 5	House 6
#Days		23	16	23	25	8	16
#Training		16	11	16	17	6	11
#Test		7	5	7	8	2	5
Groups	Refrigerator	✓	✓	✓		✓	✓
	Dishwasher	✓	✓	✓	✓		
	Kitchen outlets	✓	✓		✓		
	Lighting	✓	✓	✓	✓	✓	✓
	Washer Dryer	✓		✓			
	Microwave		✓				
	Electronics			✓		✓	
	Furnace				✓	✓	
	Stove				✓		
	Subplane					✓	
	Outlet unknown						✓
	Electric Heat						✓
	Air Conditioning						✓
	Others	✓	✓	✓	✓	✓	✓

Table 5.1.: Post-processed REDD Dataset. Note that the cardinality of the test set associated with House 5 is two therefore it was not considered for analysis.

added and, based on the percentage that each group represents in the total electricity consumed, the five ones presenting the highest impact were selected. A sixth group contains the data of the remaining equipment. These signals were normalised in order to maintain the relative importance of each group in the aggregated signal. Hence, the normalisation of the signals in the dataset was performed using the aggregated time series norm. For each house, $\frac{2}{3}$ of the signals are used for learning the source models and the remaining third for test. Table 5.1 describes the post-processed REDD dataset and indicates the available signals for each house. Observe that House 5 does not have significant data for testing, therefore, it will not be considered in the following analysis.

Notice that an exploratory comparison on the STMF results from using the post-processed REDD with different sampling rates is also performed to evaluate the influence of the sampling rate on the energy disaggregation. This particular study is presented in Appendix B.

Evaluation Metrics

As mentioned in Chapter 2 Section 2.7, a reasonable metric to evaluate the performance of each approach would be the disaggregation error:

$$\sum_{i=1}^{k} \frac{1}{2} \left\| X_i - \hat{X}_i \right\|_F^2 , \tag{5.43}$$

where X_i is the matrix of measured signals for equipment i, $\hat{X}_i$ is its predicted version and $\|\bullet\|_F$ is the Frobenius norm. Thus, this error provides a global measure of the distance between the prediction and the measured consumption (Kolter et al., 2010).

Additionally, root-mean-square errors (RMSE) can also be calculated. The RMSE is a measure between the values predicted and the values actually observed. For a general overview of the error, regarding the m days in study ($d = 1, \ldots, m$, for training and $d = 1, \ldots, m'$, for test), the RMSE associated with the aggregated signals $\bar{X}$ and their predicted versions $\hat{\bar{X}}$,

$$RMSE(\bar{X}, \hat{\bar{X}}) = \sqrt{\frac{\sum_{t=1}^{T} \sum_{d=1}^{m} \left(\bar{X} - \hat{\bar{X}} \right)^2}{T * m}} \tag{5.44}$$

can be calculated. To allow for a more detailed analysis about the performance of the algorithms, the RMSE values associated with each group of devices $i\,(i = 1, \ldots, k)$ can also be computed by making the proper modifications in the previous equation. Both disaggregation and root-mean-square errors are calculated for performance assessment in this experiment. Moreover, electricity consumption profiles, as suggested by Hart (1992), are also computed.

Experimental Parameters

The DDSC requires that several parameters are set: the parameter λ (Equation 5.33), which corresponds to the desired sparsity degree associated with source models; the step size $\tilde{\alpha}$, representing the perception update for the discriminative disaggregation training (Equation 5.36); and r, the number of bases being used at the sparse coding pre-training (Equation 5.33). Since no indication about the parameter selection is made in (Kolter et al., 2010), the algorithm's performance is explored using different set of values for λ and r. The influence of these

parameters is studied by a grid search, with the hyper-parameter $\tilde{\alpha}$ fixed. As later described, this experiment showed that the set of parameters chosen will ultimately have impact in the success of the energy load disaggregation. The setup is then: $\tilde{\alpha} = 0.0001$, $\lambda \in \{0.0001, 0.0005, 0.001, 0.005, 0.01\}$ and $r \in \{10, 15, 20, 25, 30\}$ and all the 25 possible pairs of parameters (r, λ) were tested. In the sparse pre-training each source is represented by a factorization of rank r, which is not necessarily a compressed version of that source since the rank of $\bar{X}$ and $X_i, i = 1, \ldots, k$, is always less or equal to $\min\{T, m\}$, where T corresponds to the number of samples within one day signal (1440) and m is the number of available signals for the source modelling in each house. Table 5.1 shows that the largest number of days available from the set of houses is 17 and the minimum is 11, which occur, respectively, in Houses 4 and 2 (and 6). Therefore, number of available m signals is within the interval $[11, 17]$. Hence, we chose to use values of r from 10 to 30 with a step-size of 5.

Regarding the STMF parameters, only the number of components being used at the tensor decomposition R needs to be setup (Equation 5.37). Since STMF handles 3-order tensors and the DDSC addresses matrices, the methods deal with different dimensionality and, therefore, the values for the number of components differs for both approaches. The dimensions of these 3-order tensors are $T = 1440$, m (the number of daily training signals) and $k = 6$ (the number of groups in study). The condition $T > m \times k$ that characterises a very tall tensor is verified since $m \leq 17$ and $k = 6$. Despite the rank computation of a three-dimensional tensor is a NP-complete problem (Håstad, 1990) and no straightforward algorithm is known to calculate the rank of a given tensor (Kolda and Bader, 2009), Berge (2000) has shown that the typical rank of a very tall tensor is $\chi = m \times k$. Therefore, for this dataset, the typical χ belongs to the set $\{66, 96, 102\}$. Note that the rank of a tensor is the smallest number of components in the factors A, B and C that makes of Equation 5.25 an equality. Thus, the performance of STMF is investigated when setting $R < \chi$ and $R = \chi$, namely: $R \in \{5, 10, 20, 30, 40, 50, 60, \chi_h\}, h = 1, \ldots, 6$, where χ_h represents the χ for the tensor associated with House h. Note that at the test, $5 \leq m' \leq 8$ and therefore the test matrices rank belongs to this interval. Hence $R = 5$ is included in the R set of values to study a case where the NMF decomposition would be a compressed version of the $\bar{X}'$. The experimental setup is summarized in Table 5.2.

	DDSC	STMF
#bases	$\{10, 15, 20, 25, 30\}$	$\{5, 10, 20, 30, 40, 50, 60, \chi_h, \}, h = 1, \ldots, 6$
$\boldsymbol{\lambda}$	$\{0.0001, 0.0005, 0.001, 0.005, 0.01\}$	—
$\tilde{\boldsymbol{\alpha}}$	0.0001	—
error	0.00001	0.00001
#iterations	1000	1000

Table 5.2.: The experimental setup for the experiments with DDSC and STMF.

Statistical Validation

The statistical validation is a necessary step to support conclusions drawn from performance comparison of several methods based on empirical evaluation, as mentioned in Chapter 4 Section 4.5.1. In particular, the statistical inference allows for the generalization of conclusions beyond the experimental analysis carried out. In accordance, in the following a particular test required to validate the performance assessment outcome of the described approaches for energy disaggregation is presented. Note also that we assume the independence of the samples of results being analysed.

When the normality assumption required by the t-test test is not met the alternative statistical analysis lies on non-parametric tests. The equivalent non-parametric tests are the Mann-Whitney test (Mann and Whitney, 1947) and the Wilcoxon's rank-sum test (Wilcoxon, 1945). For the statistical analysis of this chapter, the former is chosen. Both tests are equivalent and their null hypothesis is, in this context, that the performance of the methods came from the same population (Field, 2013). These methods, in accordance with the Kruskal-Wallis test described in Chapter 4 Section 4.5.1, are based on ranks. The samples are arranged in ascending order of its value, regardless of the method associated with that performance value. Ranks are then assigned in increasing order: to the lowest performance value is attributed the rank 1. In case of ties, the attached rank is the average of the potential ranks received if they had not been tied. The procedure continues with the sum of the ranks within each method (DDSC or STMF) and this value is denoted as R_1 for method 1 and R_2 for method 2. Next, the Mann-Whitney test statistic,

$$U = n_1 n_2 + \frac{n_1 (n_1 + 1)}{2} - R_1 \tag{5.45}$$

where n_1 and n_2 are the sample size of method 1 and 2, respectively, is calculated for each technique, by exchanging the method role in Equation 5.45. The lower

U between the two computed values is set to the test statistic. To determine if the test statistic is significant, it is converted to the so-called z-score,

$$z = \frac{U - \bar{U}}{SE_{\bar{U}}} \tag{5.46}$$

where $\bar{U}$ is the mean, $SE_{\bar{U}}$ the standard error of the U statistic, and such that the resultant distribution has mean 0 and standard deviation 1.

5.5.2. Results and Performance Evaluation

In this section, the results regarding the Discriminative Disaggregation Sparse Coding (DDSC) and Source Separation by Tensor and Matrix Factorization (STMF) methods obtained for each one of the houses in the post-processed REDD dataset are presented and discussed. The performance evaluation reports the mean value of 30 runs. First, the study of the parameters' influence in both algorithms based on the disaggregation error and on the RMSE for the aggregated signal is considered. In particular, this empirical experience focus on the sparseness regularization parameter and its importance for successful electrical source separation using a sparse coding based approach. Second, the best set of parameters for each method is selected and a more detailed analysis is reported, in particular, both approaches are compared regarding the RMSE for each appliance and the electrical consumption profiles.

Sensibility Analysis for DDSC method

The DDSC is evaluated for the set of parameters defined on Section 5.5.1. Figure 5.4 shows the disaggregation error for this method regarding each pair of parameters, after the training (on the right) and test (on the left) steps, for each one of the five houses under consideration.

These results revealed two major decreasing trends in the training results. An increase on the number of bases r or a reduction on the enforced sparsity is translated by a decrease in the disaggregation error, *i.e.*, the lower the sparsity parameter the better the error. However, this reduction is not equally noticeable in both parameters. An increment in the number of bases r has a bigger impact in the disaggregation error than decreasing the degree of sparseness λ, with exception of House 2. For instance, considering House 1 and fixing $\lambda = 0.0001$, there was a considerable reduction of the error values for the highest values of r.

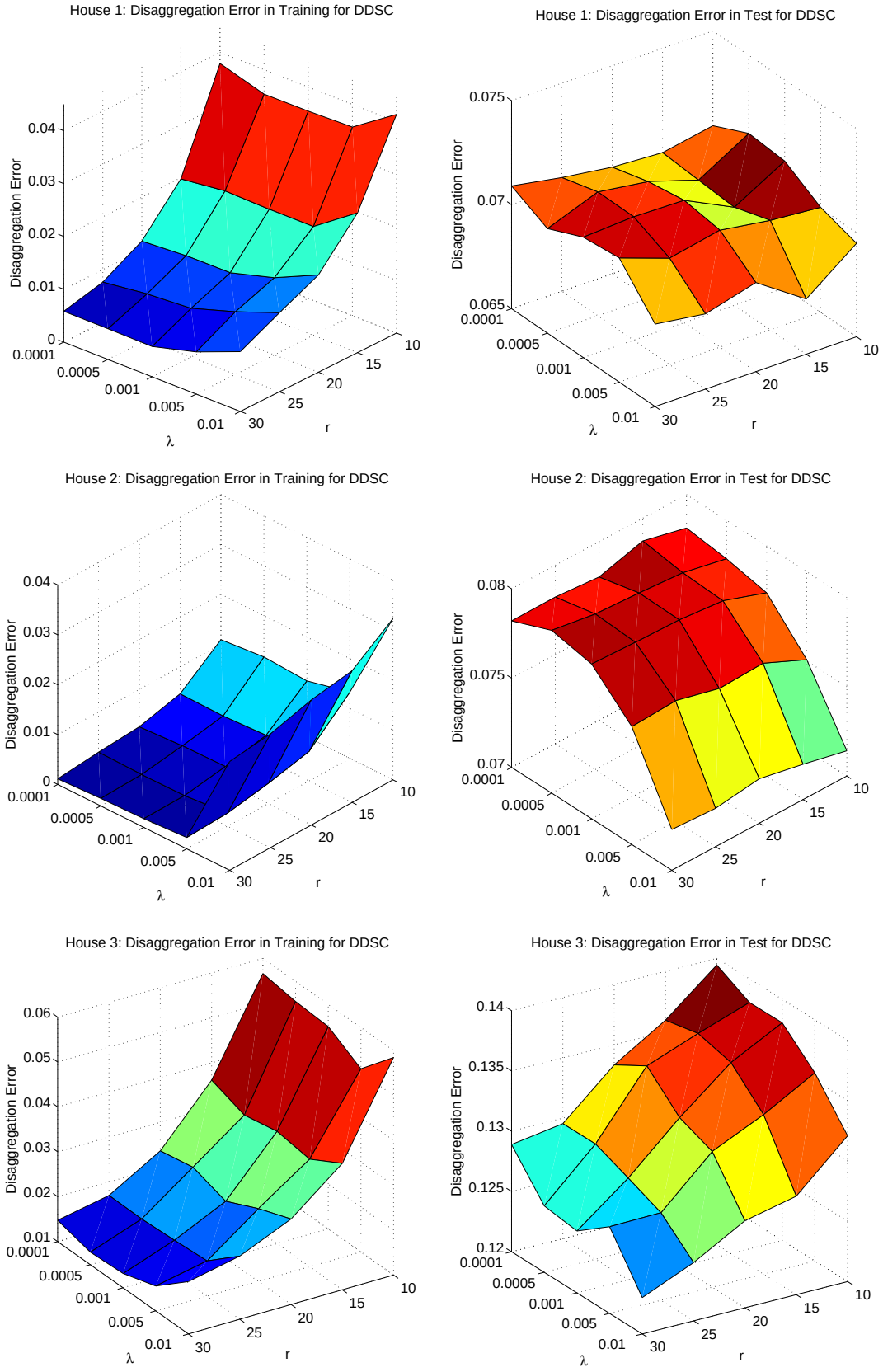

Figure 5.4.: Disaggregation Error for the DDSC approach – Houses 1, 2 and 3.

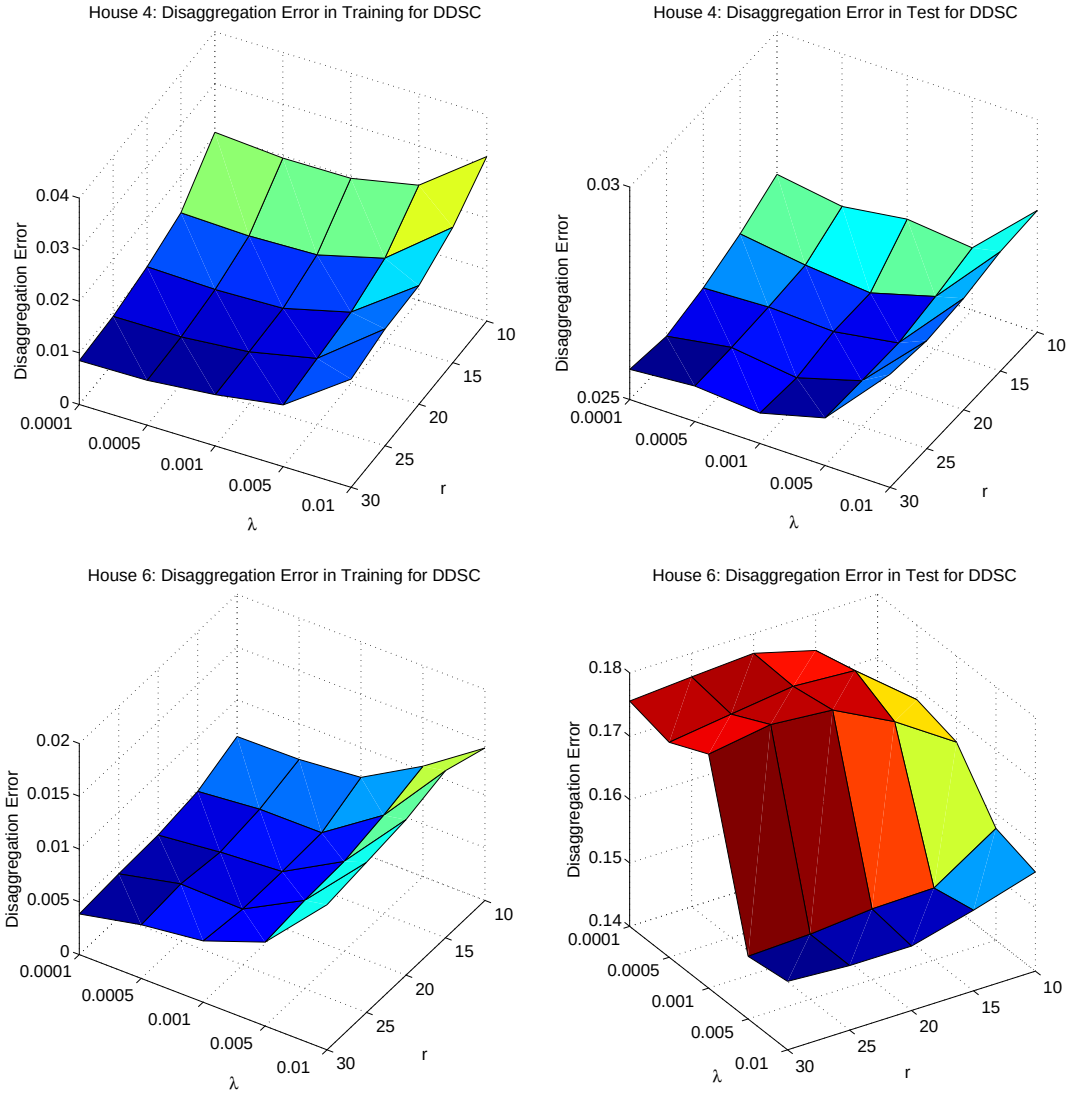

Figure 5.4.: Disaggregation Error for the DDSC approach – Houses 4 and 6.

In fact, the error associated with $r = 30$ diminished by 84.7% when comparing it with the correspondent for $r = 10$, that is, the sum of the distances between the predicted consumption and the one measured for each appliance has decreased substantially with the increase of the number of bases used for source modelling. On the other hand, the variation over the sparseness parameter λ was about 8.5% from the highest ($\lambda = 0.01$) to the lowest sparseness degree ($\lambda = 0.0001$), considering $r = 30$, achieving the lowest error for the $\lambda = 0.0001$. In this case, imposing a higher degree of sparseness in Equation 5.33 does not correspond to matrices $\hat{X}_i$ significantly closer to $X_i, i = 1, \ldots, k$ (Equation 5.43). As a consequence, the lowest disaggregation error occurred for the pair $(30, 0.0001)$. Yet, for House 2 and regarding the larger number of bases ($r = 30$), a decrease of the imposed sparsity from $\lambda = 0.01$ to $\lambda = 0.0005$ represented a reduction of 88.53% in the error.

Interesting to notice is the fact that, in the training phase, the higher the values of λ, the worse are the disaggregation errors' results. According to Figure 5.4 all the best results were obtained for the lowest value of λ, 0.0001, with the exception of House 3, where the best result is precisely for the second lowest sparsity enforcement, 0.0005. In this case, the error ranged in the interval $[0.0128, 0.0585]$, which is the largest interval among the results for all the houses in analysis. On the other hand though presenting the same trend, for House 6 the results do not show such a significant variation for the changes in the parameters values: the minimal error was of 0.0039 while the maximum was of 0.0159.

Similar trends were observed for the test results as the increasing number of bases r and decreasing λ's for Houses 3 and 4. For the latter, the higher the degree of sparseness is, the higher the disaggregation errors are. Still, for Houses 2 and 6, the higher enforced sparsity degree achieved the best results. In fact, for House 6, using $r = 30$, the error for $\lambda = 0.01$ is 14.08%, lower than considering a sparsity of $\lambda = 0.0001$. With exception of Houses 1 and 2, the smallest distance between the predicted and the measured energy occurred when considering $r = 30$, regardless of the enforced sparsity degree. Apart from House 4, which has similar error ranges for the training and test results, the test disaggregation errors ranged in $[0.06, 0.18]$, while the training ones varied from 0 to 0.06. Notice that Houses 3 and 6 achieved errors significantly higher than the other houses. This might be due to the type of appliances as it will be inspected later in the detailed discussion of results. In conclusion, especially in

the testing phase, the disaggregation error performance is mostly dependent on the conjugation pair of values r and λ. The performance values are also highly dependent on the house being studied, in the sense of what are the devices groups to achieve disaggregation.

The most commonly used metric for comparison of results, RMSE, was also employed to analyse the DDSC performance. Figure 5.5 presents the RMSE values obtained for the aggregated signal by this method.

Once again, the trends either at training or at test are quite straightforward: as the λ decreased or as r increased, the RMSE values decreased. Taking House 1 for an example and fixing the number of bases to $r = 30$, when changing the λ from 0.0001 to 0.01, the RMSE increased 79.29% at training and 4.82% at test. Likewise, fixing $\lambda = 0.0001$, the RMSE for the highest number of bases, $r = 30$, represents only 38.94% of the error associated with $r = 10$ at training, and 94.83% at test. Another example is House 6: when the number of bases is set to $r = 30$, the RMSE associated with a sparsity of $\lambda = 0.01$ is 2.31 times greater than for $\lambda = 0.0001$ at training and 1.12 times at test. In addition, given $\lambda = 0.0001$, the RMSE is 1.56 times greater for $r = 10$ than for $r = 30$ at training and 1.07 times at test. Apart from House 4 (in which similar results were obtained for both training and test), it can be observed that the error at test is larger than the error at training. Hence, the minimum RMSE is observed again for the pair $(30, 0.0001)$, with the exception of House 4 for which the minimum is reached by setting the parameter to $\lambda = 0.001$. This RMSE analysis clarified the idea that imposing a higher sparsity degree to the source models does not imply that the DDSC accomplishes better results for energy disaggregation as highlighted by the statistical analysis presented in Appendix C. Similarly, we observed that the best RMSE values were not associated with the source models corresponding to compressed versions of the matrices $X_i, i = 1, \ldots, k$, ($r < 20$, see Section 5.5.1).

Sensibility Analysis for STMF method

The performance of the Source Separation by Tensor and Matrix Factorization (STMF) was also measured in terms of the disaggregation error with respect to each R. Remember that the tensor approach parameter R is not related to the DDSC approach parameter r and no sparsity enforcement is used.

Figure 5.6 shows these results for training (on the right) and test (on the left) for each one of the considered five houses. While, at the training phase,

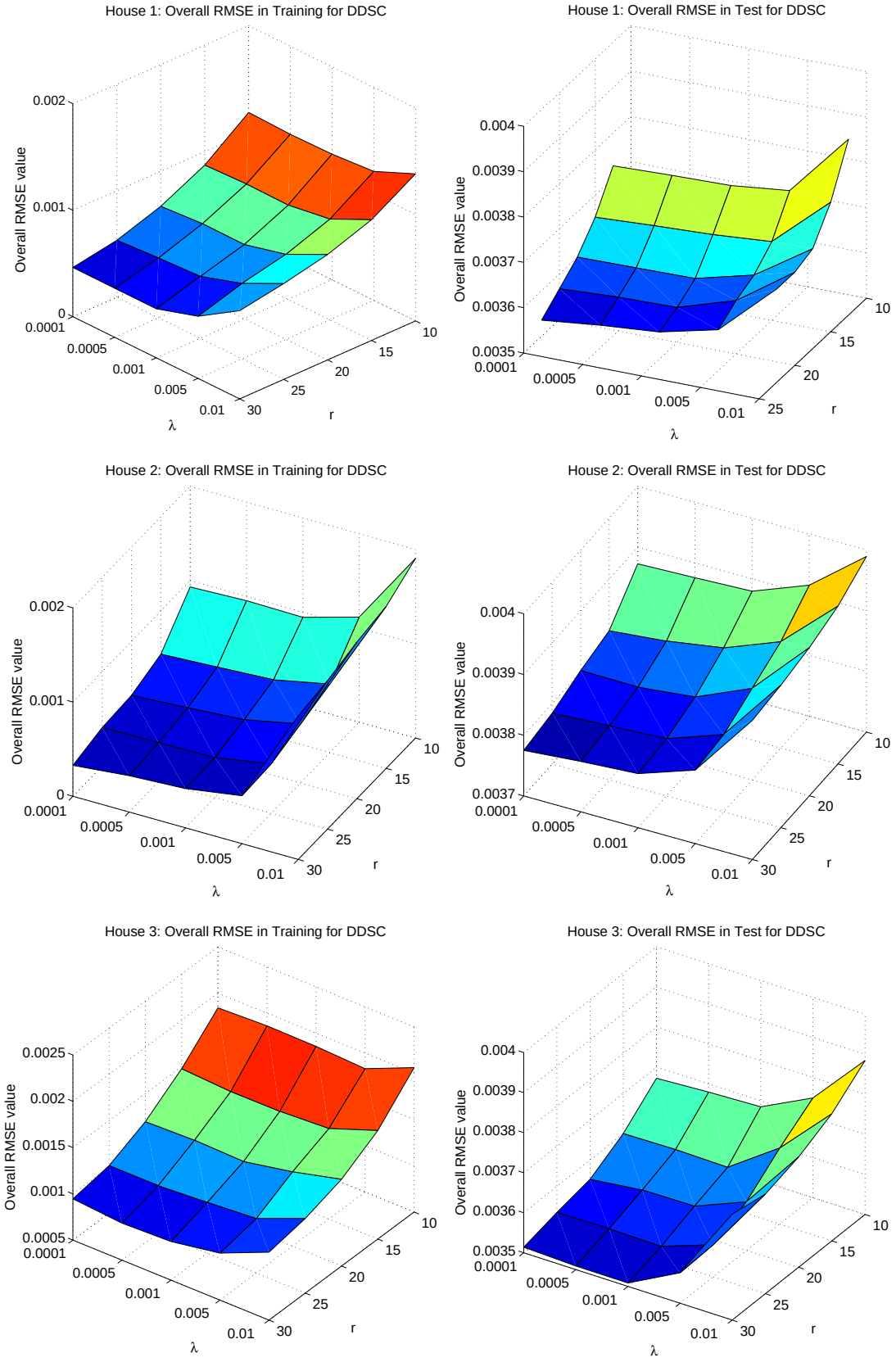

Figure 5.5.: Overall RMSE for the DDSC approach – Houses 1, 2 and 3.

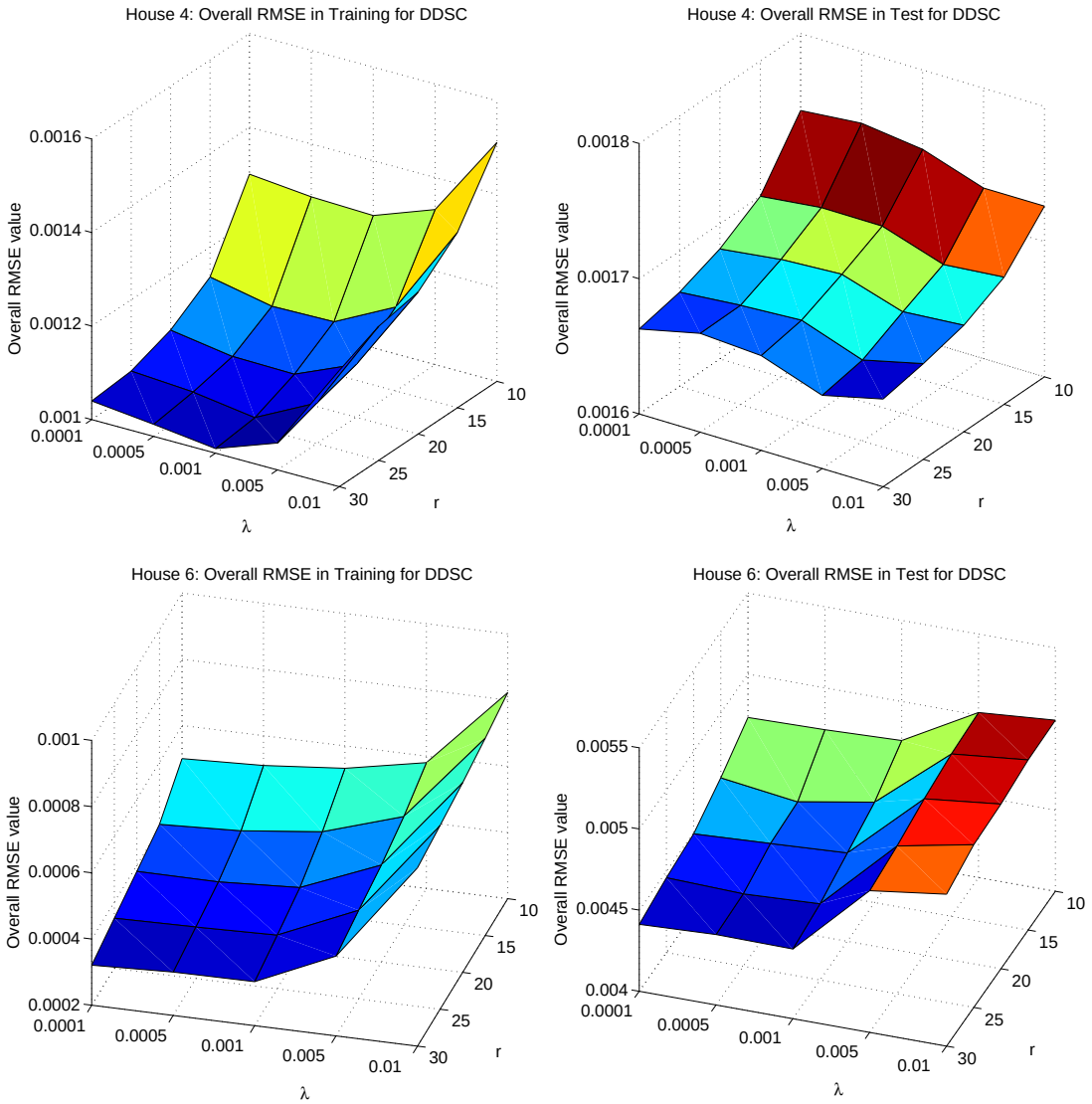

Figure 5.5.: Overall RMSE for the DDSC approach – Houses 4 and 6.

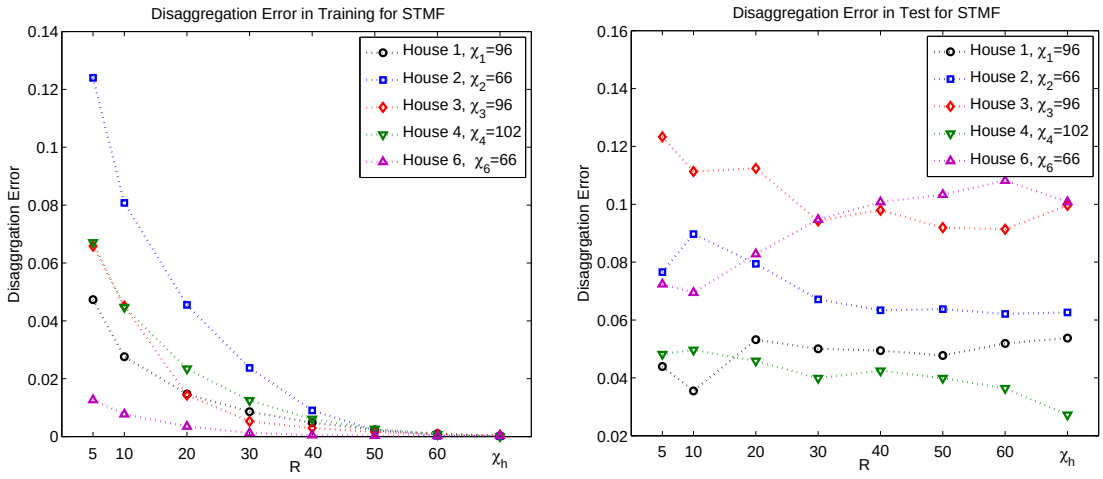

Figure 5.6.: Disaggregation Error for the STMF approach.

a noticeable decreasing trend is observed as the number of components R are incremented, no clear trend of all the houses is noticed at test. For the former, the best disaggregation error for any house is reached when we set $R = \chi_h$ rather than for smaller number of components. This was to be expected since χ is the typical rank of a very tall tensor, as stated in Section 5.5.1. In fact, for House 2, the disaggregation error associated with $R = \chi$ represents only 0.14% of the corresponding error for $R = 10$. Also noticeable is the fact that the slope decreases as R approaches at least half of the value χ_h. For instance, a small difference between the results for $R = 60$ and for $R = \chi$ exists. A higher number of components does not stand for a significant lower error with the possible exception of House 6, where the error was virtually null.

In the test results, no general trend for all the houses is observed. A decreasing trend is observed for Houses 2, 3 and 4 while for House 6 an increase in the disaggregation error occurs. For House 1, a decline in performance is observed for $R = 20$ in contrast with the disaggregation error yielded for $R = 10$. Still, the performance evolution remain mainly stable for $R \geq 20$. In a general overview, a change in the number of components used for the PARAFAC decomposition from $R = 5$ to $R = \chi$ correspond to a increase of 22% and of 39% in the disaggregation error for Houses 1 and 6, respectively. On the contrary, for Houses 2, 3 and 4, the change in the R value is associated with a decrease of 18%, 19% and 44%, respectively. Nevertheless, the variation is not expressive for $R \geq 30$ (in the order of 0.01). For Houses 2 and 3, the disaggregation error is the lowest when the number of components is $R = 50$ or 60 while for Houses 1 and 6

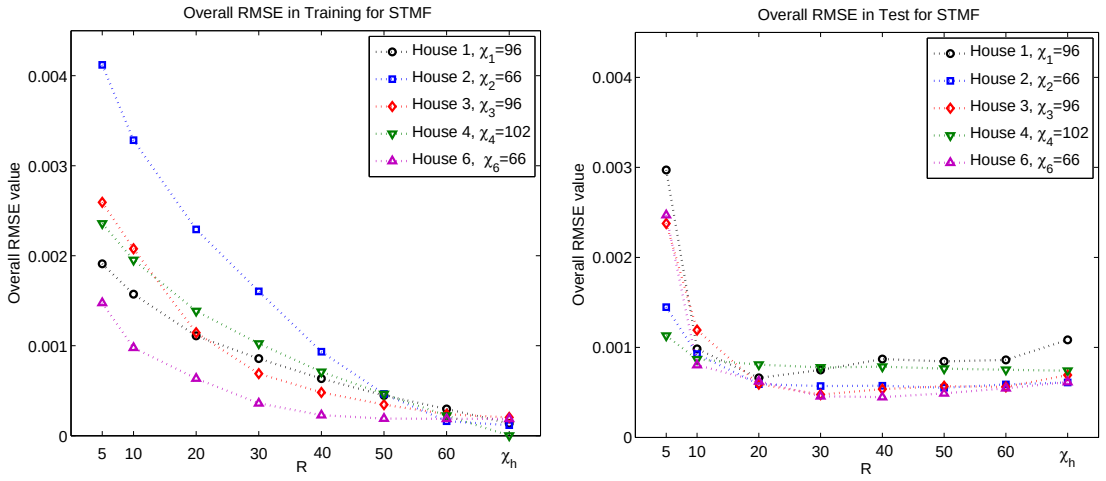

Figure 5.7.: Overall RMSE for the STMF approach.

the best performance occurs for $R = 10$ (disaggregation error of 0.0354 and 0.0694, respectively). In case of House 6, the worst performance (disaggregation error of 0.1081) is achieved for $R = 60$. Note also that, for Houses 3 and 6, the disaggregation errors are the highest particularly for $R \geq 20$, which again could be related to the particular devices of each house.

The STMF performance is also evaluated by the RMSE values for the aggregated signal. Figure 5.7 presents these RMSE values. Notice that the error scale has different ranges of values between both methods, DDSC and STMF. A noticeable decreasing trend in the RMSE values at training was observed: as the number of components R increased, the associated RMSE decreased. Regarding the test phase, although an improvement in performance (decreasing values of RMSE) is observed when R is incremented from 5 to 20, for the remaining R values it presented an almost constant value. The exception is House 1 at test. As illustrated in Figure 5.7, a slight decline in performance is observed for $R = \chi$. In this case, an increment from 0.0009 to 0.0011 was observed when the number of components increases from $R = 60$ to $R = 96$. Nevertheless, the slight increase observed can be seen as irrelevant. Notice that the test error is in average 3.03 times higher than the training error when R is set to 60. Actually, the training ranged in $[0.0002, 0.0045]$ while the test in $[0.0005, 0.003]$. Also, the best RMSE values are not associated with compress versions of matrix $\bar{X}'$ ($R = 5$). Since for the RMSE, the observed improvement from $R = 60$ to $R = \chi$ in training was not significant and in the test phase no relevant variation was observed (with exception of House 1) in the following we set $R = 60$ for all the houses.

House	Training		Test	
	DDSC	STMF	DDSC	STMF
	Disaggregation Error			
1	0.0058 ± 0.0021	0.0010 ± 0.0001	0.0709 ± 0.0030	0.0519 ± 0.0073
2	0.0011 ± 0.0004	0.0002 ± 0.0002	0.0782 ± 0.0018	0.0621 ± 0.0044
3	0.0148 ± 0.0046	0.0008 ± 0.0007	0.1289 ± 0.0031	0.0914 ± 0.0191
4	0.0099 ± 0.0007	0.0008 ± 0.0008	0.0257 ± 0.0004	0.0364 ± 0.0028
6	0.0039 ± 0.0006	0.0003 ± 0.0002	0.1755 ± 0.0031	0.1082 ± 0.0117
	RMSE			
1	0.0005 ± 0.0001	$0.0003 \pm 1.81 \times 10^{-5}$	0.0035 ± 0.0004	0.0009 ± 0.0001
2	0.0003 ± 0.0001	0.0002 ± 0.0001	$0.0038 \pm 1.62 \times 10^{-5}$	0.0006 ± 0.0001
3	0.0009 ± 0.0001	$0.0002 \pm 1.17 \times 10^{-5}$	$0.0035 \pm 3.54 \times 10^{-5}$	$0.0006 \pm 3.53 \times 10^{-5}$
4	$0.0010 \pm 1.21 \times 10^{-5}$	$0.0002 \pm 1.73 \times 10^{-5}$	$0.0017 \pm 0.66 \times 10^{-5}$	$0.0008 \pm 1.87 \times 10^{-5}$
6	$0.0003 \pm 1.64 \times 10^{-5}$	$0.0002 \pm 0.06 \times 10^{-5}$	$0.0044 \pm 4.82 \times 10^{-5}$	0.0005 ± 0.0001

Table 5.3.: Average Disaggregation Error and RMSE results for both methods.

Further results are provided in Appendix D, regarding the RMSE by appliance.

Discussion

Based on the results from the RMSE values for DDSC method, for the following we select $r = 30$ and $\lambda = 0.0001$ for each house. In case of House 4, we set $\lambda = 0.001$. For STMF, R was set to 60 as previously explained.

Table 5.3 displays the Disaggregation Error and RMSE for the aggregated signal obtained by both approaches for each house in the dataset. Apart from the disaggregation error in test for House 4 which is 41.63% higher than the value achieved by the DDSC, the STMF results are always considerably lower than the DDSC ones. For instance, for House 6, at the test step, the disaggregation error achieved by DDSC surpasses by 62.20% the STMF. At this point, it is worth mentioning that the source models in DDSC are computed in order to minimize the disaggregation error (Equations 5.35 and 5.36). Regarding the RMSE, the error associated with DDSC is 8.8 times greater than the error yielded by the new method. These results suggest that STMF would be more appropriate to solve the energy disaggregation task than DDSC.

For further exploration of the suitability of these methods to solve the disaggregation problem, the RMSE for each device is computed and shown in Figure 5.8. This detailed analysis allows us to identify the groups poorly disaggregated and support the observations made using the overall RMSE (*i.e.*, where no groups are discriminated). Notwithstanding that some appliances present similar RMSE values at the training phase for both methods (in particular for

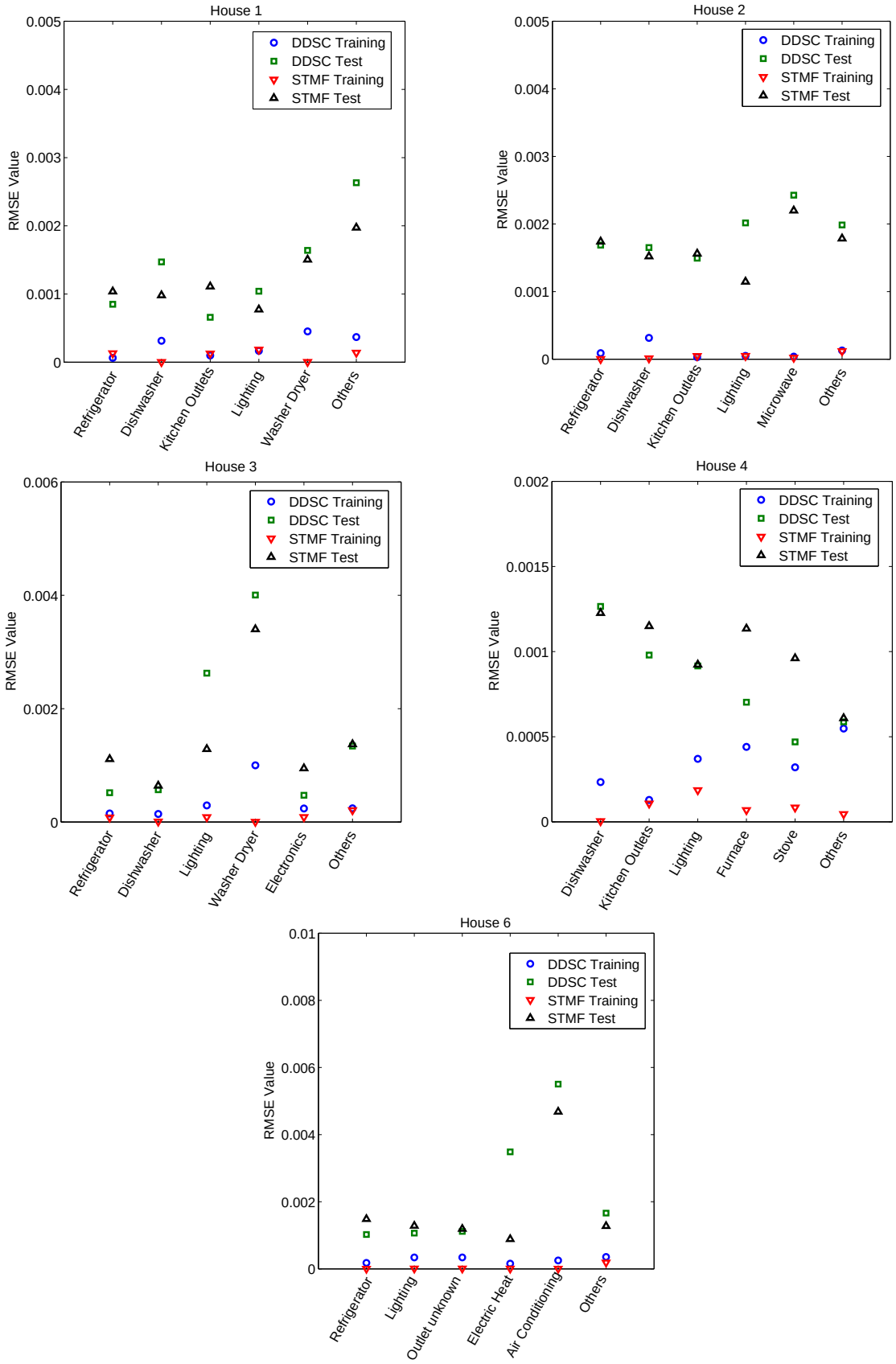

Figure 5.8.: RMSE by appliance for DDSC and STMF.

Houses 1 and 2), the STMF reaches much lower RMSE values, especially for House 4. However, this case does not correspond to a better performance in test since STMF yields either similar values or slightly higher than DDSC. Still, for all the other houses, the RMSE values at test achieved by STMF for some particular devices must be highlighted. For instance, the RMSE for "Lighting" in House 2 (and 3) represents 56.87% (49.05%, respectively) of the values yielded by DDSC. A similar observation can be drawn for the "Lighting" in House 4. Moreover, in House 6, the RMSE associated with "Electrical Heater" only corresponds to 22.95% of the value obtained with DDSC. It is interesting to notice that in House 4, except for "Furnace" and "Stove", the RMSE associated with each device for STMF is always smaller or similar to the correspondent value for DDSC. Note also the similar performance for the "Refrigerator" across the several houses where this device exists. On the other hand, observe that the measured consumption of the two particular groups, "Furnace" and "Stove", is mostly composed by zero-entries, since these devices usually only operate during short periods of time along one day. Hence, the sparsity degree imposed by the DDSC may be relevant in this particular case. To conclude, the RMSE associated with "Washer Dryer" and "Air Conditioning" in Houses 3 and 6, respectively, are the highest among all the appliances in study. Note that the observed disaggregation errors obtained for the latter houses by the STMF were higher than for the remaining ones. Still, they were consistently lower than the disaggregation errors for the DDSC.

To reinforce the analysis above, the consumption profiles in a schematic pie plot can be observed. In Figure 5.9, the computed and the consumed energy for the five houses in this experiment are shown considering only the test signals. These profiles are in line with the preceding observations: there is a general performance gain in favour of the STMF when compared to DDSC and regarding the actual measured consumption, with the exception of House 4. In fact, for Houses 1 and 3, STMF predicted the exact amount of consumed energy by the "Washer Dryer" while the DDSC considered that it only contributed with a small percentage (2% and 5% against 7%, respectively). This is not inconsistent with the observations made for the RMSE by appliance, which report the mean error for each sample in the signal.

To reinforce the above remark and regarding House 1, for "Lighting" the STMF produced a result that is 6% less than the ground truth values, while the DDSC algorithm responded with a value 36% higher than the actually

consumed electricity. In House 2, the method assigned 47% more of consumed energy than the exact amount. In this case, the STMF estimated a percentage only 8% higher than the real consumed energy. Similar observations can be drawn for House 3 and 6. Taking House 3 as example, DDSC produced a result that is 17% higher than the ground truth values, while the proposed approach considered that 22% of the energy was consumed by this group of appliances. Still, it was 3% less that the exact amount of used energy. Indeed, the estimates yielded by the sparse coding based approach for group "Lighting" in all the houses, with exception of House 4, are clearly excessive.

Regarding the "Refrigerator" in Houses 1 and 2 notice that both methods predicted a similar amount of energy. For House 1 the estimates were slightly higher than the electricity actually consumed (3%). On the other hand, for House 2, only half of the energy actually measured was estimated. For Houses 3 and 6, the STMF considered that this device consumed 20% of the total energy, an estimate 6% (5%, respectively) higher that the exact amount. In both houses, the DDSC yielded a prediction closer to the ground truth.

Concerning the remaining devices in House 1, the consumption predicted by STMF was considerably closer to the real than the estimate computed by the DDSC. In particular, this is verified for the group "Others" (indistinctive appliances), which corresponds to 34% of the energy and where the DDSC assigned only 10%. In a general overview, STMF improperly assigned 13% of energy while DDSC incorrectly computed 38%. The superiority of STMF for House 3, regarding the groups not yet mentioned, is also observed. Regardless of the group "Electronics", the energy usage predicted by STMF was closer to the exact amount than the estimates yielded by the DDSC. Again, 20% of energy was incorrectly assigned by the STMF, 4% less than the DDSC. Moreover, in what concerns the consumption profiles of House 6, with exception of the "Refrigerator", we also observe a performance gain in favour of the STMF when compared to DDSC and regarding the actual measured consumption. Once more, 20% of energy was incorrectly assigned by the STMF, against 33% by the DDSC.

An opposite observation is drawn for House 2. In this case, the consumption predicted by DDSC for the groups "Dishwasher", "Kitchen Outlets" and "Microwaves" was closer to the exact amount of consumed energy than the estimation of STMF. Notwithstanding, the DDSC improperly assigned 50% of the total energy, mainly due to the results achieved for "Lighting", a value 15%

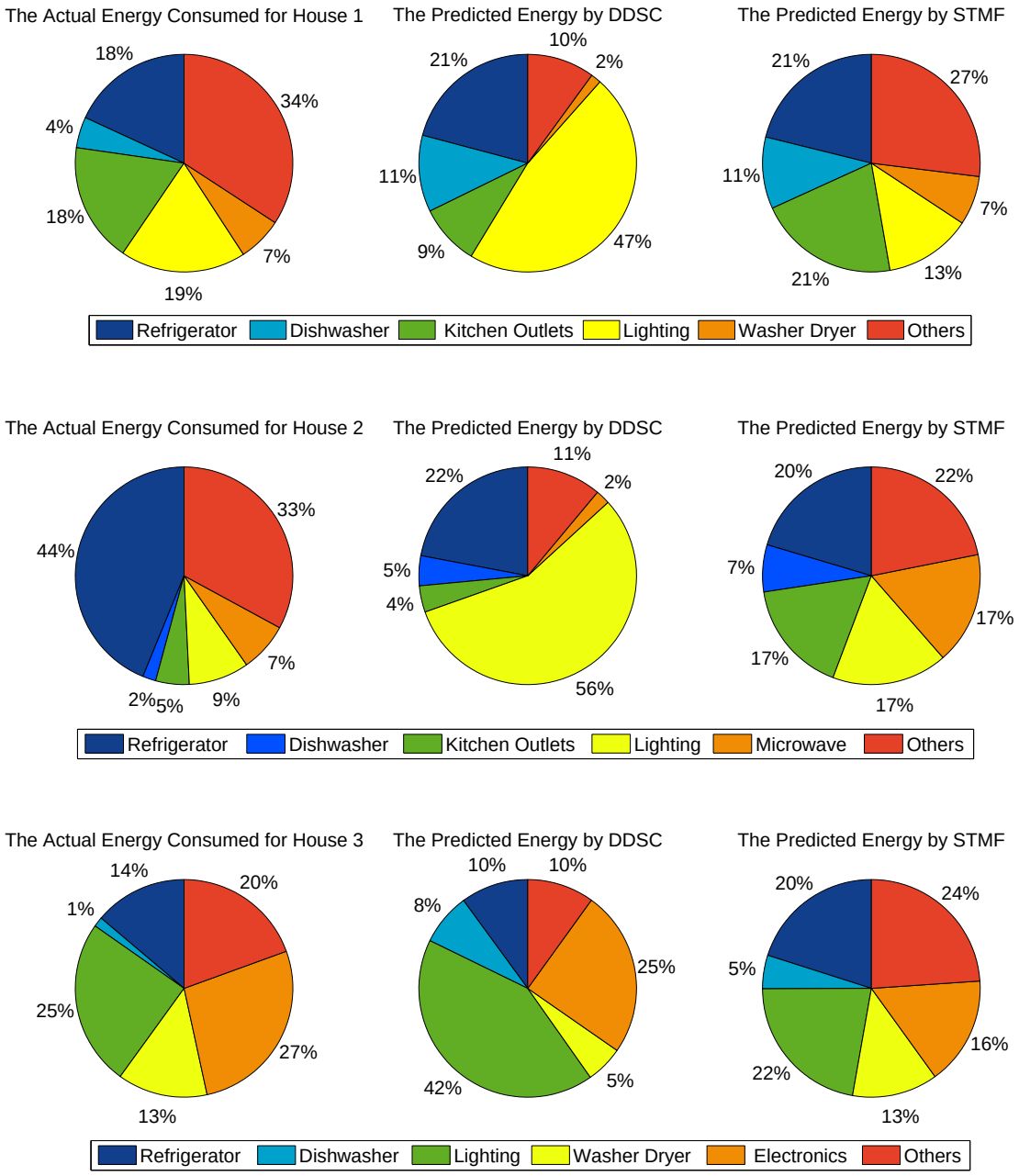

Figure 5.9.: Consumed and predicted energy by DDSC and STMF approaches for Houses 1, 2 and 3.

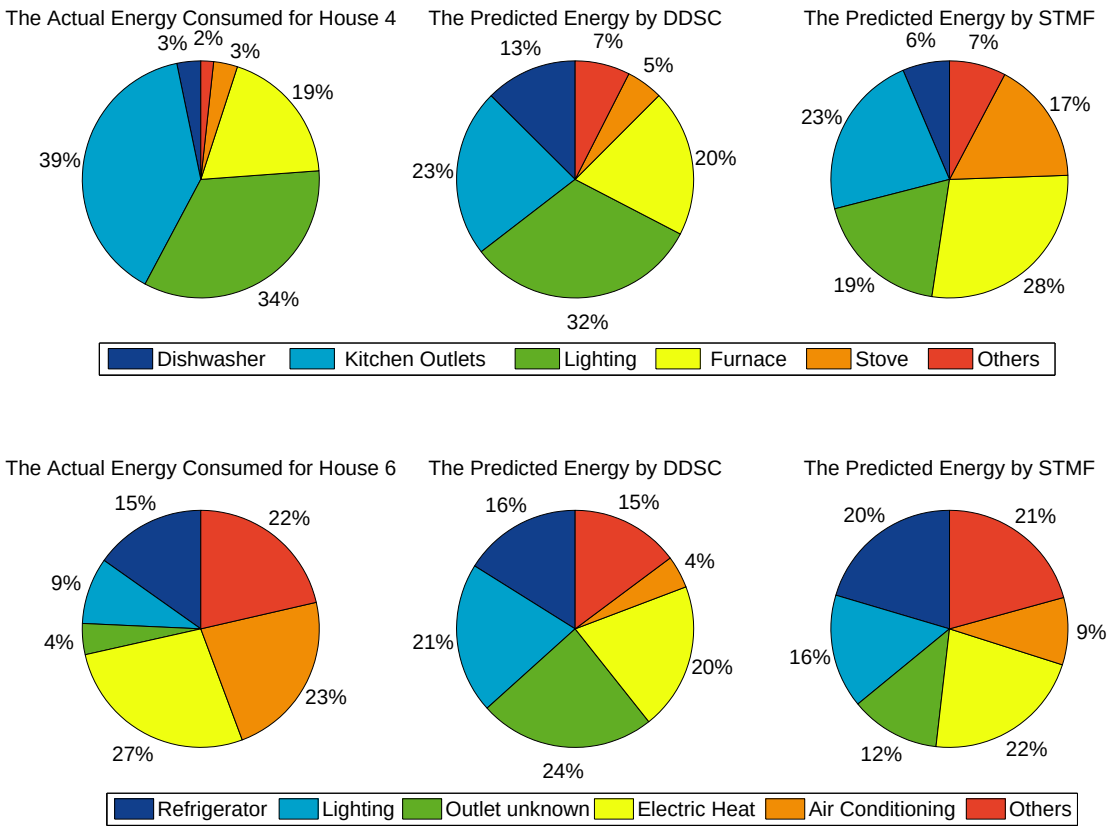

Figure 5.9.: Consumed and predicted energy by DDSC and STMF approaches for Houses 4 and 6.

higher than the energy incorrectly computed by the STMF.

Regarding the consumption profiles of House 4, both approaches considered that a similar amount of energy, 23%, was consumed by the "Kitchen Outlets". Nevertheless, the exact amount was 16% higher. A similar observation is drawn for the group "Others", which had assigned a consumption 5% higher than the ground truth. With exception of "Dishwasher", the consumption predicted by DDSC for the remaining devices was closer to the ground truth values than the estimates yielded by the STMF. Note also that for this house, the DDSC assigned a quantity of energy close to the exact amount used for "Lighting", in contrast with other houses. In fact, the performance of STMF for this house was not as good as DDSC which only improperly estimated 18% of energy in opposition to the 31% incorrectly appointed by the proposed approach.

Overall, the proposed approach STMF was more accurate than the DDSC regarding improperly energy consumption predictions. The STMF good per-

	Disaggregation Error				RMSE			
	Training		Test		Training		Test	
	DDSC	STMF	DDSC	STMF	DDSC	STMF	DDSC	STMF
Median (*Mdn*)	0.0059	0.0002	0.0780	0.0631	0.0005	0.0001	0.0035	0.0007

Table 5.4.: Median (*Mdn*) values for Disaggregation Error and RMSE regarding both methods using 150 samples in each test.

formance was particular observed for the estimates associated with the consumption of "Lighting". The exception is the House 4 for which the DDSC method was superior. It remains to be understood why the novel approach is better suited for certain houses than others, as can be seen in the pie plots of the electrical energy consumption. We claim that the main difference would be the type of devices in study for each house. Still, this is not a definitive answer.

To validate the observations regarding the suitability of STMF for the task at hand, statistical tests were again performed. The SPSS statistical tool was used to analyse the existence of similarities between the disaggregation errors for both DDSC and STMF. A similar study was carried out for the RMSE values associated with the aggregated signals.

To specify the statistical tests performed, the conditions under (their) validity must be verified. Thereby, we first performed a Kolmogorov-Smirnov test. This test was held with a significance level of 0.01. The *p-values* produced were, respectively, of 0.1×10^{-4} for the disaggregation errors at the training step regarding the DDSC, of 0.482×10^{-3} for disaggregation errors at the test for STMF and a *p-value* of 0.0 for the remaining cases. Therefore, we were led to conclude that the disaggregation errors and overall RMSE, for both training and test, were far from having a normal distribution, and, hence, a non-parametric test for comparison of both methods should be pursued. Since we want to compare two independent conditions, the Mann-Whitney test was chosen. The goal is to assess whether the performance of both methods differs.

Table 5.5 reports the U statistic associated with the Mann-Whitney test, the correspondent z-score (z) and asymptotic significance (Asymp. Sig.). The distribution associated with the DDSC disaggregation errors did differ significantly from the ones associated with the STMF, either at training or at test, at a significance level of 0.01 ($p\text{-}value < 0.001$). In fact, regarding the mean ranks and the median values (*Mdn*) (Table 5.4) for the training, the DDSC disaggregation errors ($Mdn = 0.0059$) are higher than for the STMF ($Mdn = 0.0002$) with $U = 90$ and $z = -14.86$. Furthermore, for the test, similar observations can be

	Disaggregation Error		RMSE	
	Training	**Test**	**Training**	**Test**
Mann-Whitney U	90.0000	8078.0000	30.0000	0.0000
z	-14.8553	-4.2223	-14.9351	-14.9751
Asymp. Sig. (2-tailed)	<0.001	<0.001	<0.001	<0.001

Table 5.5.: Test statistics for disaggregation error and RMSE regarding both methods.

drawn: the DDSC disaggregation errors ($Mdn = 0.0780$) differ from the ones for STMF ($Mdn = 0.0631$), $U = 8078$ and $z = -4.22$.

In addition, and according also to the Mann-Whitney test (Table 5.5), the distribution associated with RMSE values by the DDSC is not the same as for the RMSE results of STMF, both in training and test ($p\text{-}value < 0.001$). In particular, the DDSC obtained RMSE values ($Mdn = 0.0005$) higher than the STMF ($Mdn = 0.0001$) for training, $U = 30$ and $z = -14.94$. Likewise, for the test, the RMSE values of DDSC ($Mdn = 0.0035$) were higher than for the STMF ($Mdn = 0.0007$), $U = 0$ and $z = -14.98$. These statistical results clearly help to support the conclusion that the STMF is a reliable alternative to DDSC for the disaggregation of the electrical signals in this dataset of on-site home electrical measurements.

5.6. Summary

In this chapter, single-channel source separation approaches were explored for solving the disaggregation of electrical energy into its several components, corresponding to the consumption of appliances connected to the household network. This recent reinterpretation of NILM as a source separation problem differs from the usual approaches that rely on electrical appliances signatures identification and on event detection as presented in Chapter 2 and studied in Chapter 3. Following the source separation path, approaches based on data-adaptive representations arise as suitable for solving NILM and modelling the several intervening sources. In fact, sparse coding based methods for energy disaggregation have successfully been proposed. The approach described by Kolter et al. (2010) learns independent sparse sources models for each device in the electrical network which are used posteriorly for consumption separation. An alternative new approach was proposed in this chapter, based on multi-

array representation and non-negative tensor factorization for source modelling. Tensors are a natural representation of data in applications of signal analysis and source separation. Thereby, these were explored in this chapter for energy disaggregation. In the proposed STMF approach, a 'global' source model comprising the relevant information for each of the three domains in analysis (time of the day, days of the week and electrical appliances) is computed by non-negative tensor factorization. Then, given a set of unseen aggregated signals for which no tensor representation is available, the learned model is embedded in the non-negative matrix factorization procedure, in order to allow disaggregation.

Computational experiments for the different strategies evaluation were designed aiming at two goals. First, the empirical exploration of the parameters' influence in both algorithms based on the disaggregation error and on the RMSE for the aggregated signal was carried out. In particular, the importance of the sparseness regularization parameter for successful electrical disaggregation using the sparse coding based approach was studied. Second, the comparison between the proposed approach and the sparse coding based method for the task of electrical source separation was analysed.

Regarding the importance of the sparsity parameter in the sparse code modelling approach, the results indicated that a difference exists between the lower and the higher sparsity degree tested in terms of root-mean-square-error of the signals. Furthermore, in this context, the best results were obtained when imposing the lowest sparseness degrees. For example, in House 1 when the number of bases was fixed to $r = 30$, a change from $\lambda = 0.0001$ to $\lambda = 0.01$ corresponded to an increase of 79.29% at training and 4.82% at test in terms of overall RMSE.

For the second goal, the comparison between the proposed technique (STMF) and the method described by Kolter et al. (2010) from the related work, the computational experiment demonstrates the effectiveness of the novel approach on real-world data for energy disaggregation, as demonstrated by statistical evidence. Indeed, the overall RMSE associated with DDSC was 8.8 times greater than the error yielded by the new method. In terms of disaggregation error the results were always considerably lower for the proposed approach than for the sparse coding based method, apart from House 4 of the dataset. These conclusions were reinforced by the analysis of the consumption profiles which clarified that the energy incorrectly assigned by the STMF was lower than the

one by DDSC, again with exception of House 4. In fact, an interesting research question would be that providing some insight about why the STMF is better suited for certain houses than others, in particular, if the main difference lies in the type of devices in study for each house.

The dependencies between devices would be also worthwhile of exploring. Specially, it would be interesting to verify if and how the proposed technique based on multi-array factorization incorporates such information via the tensor factorization. In fact, information related to dependencies among devices that are used in specific home-activities may be of importance for achieving a more accurate disambiguation. For this purpose, the dataset should be refined such that appliances which intuitively appear to be related are discriminatively included in the study.

The performance assessment of the proposed technique against the supervised and unsupervised methods recently proposed in the related literature and mainly based on HMM, as revised in Chapter 2 Section 2.6.1, would also be worthwhile of investigating.

CHAPTER 6

Conclusions and Future Work

"To infinity and beyond!"
-Buzz Lightyear, Pixar Animation Studios (1995)-

6.1. Main Research Accomplishments and Conclusions

This Thesis aimed at the exploration of novel approaches for solving the energy disaggregation problem as part of a NILM framework. In such a system, the total electricity consumption is generally gathered at a single-sensing point, usually the utility-customer interface. Next, by means of machine learning and pattern recognition methods, this information is separated into the consumption of the several appliances connected to the monitored electrical network and, then, device-level detailed electricity usage is provided as feedback to the user. The appliance-specific information has been demonstrated as an essential tool for increasing energy efficiency at the household level which is of significant importance with the present-day climate change and environmental concerns.

Bearing in mind that the main challenge within NILM addresses the energy disaggregation, usually solved as a classification problem, the research work in this Thesis can be summarized in three main points. First, the classical approach based on classification methods and appliance electrical signatures was investigated, mainly focused on the exploration of electric features to reinforce the search for a robust set of distinctive characteristics used as device signatures. Second, in the view of signal processing and time series analysis, strategies for the extraction of meaningful components from the aggregated signal associated with the consumption of a particular group of devices were studied. Third, single-channel source separation approaches, based on source modelling were explored for the separation of single-point sensed data into the individual consumption of each appliance connected to the monitored electrical network.

Regarding the problem at hand as a classification task and the undemanding requirements associated with the steady-state signatures in terms of sampling rate and specific hardware for gathering the needed data, a distinctive appliance signature composed by the step changes (the difference between two steady-states) on active, reactive power signals and on the power factor measurements was proposed in Chapter 3 Section 3.2. This signature added proportional information to the usual features corresponding to the absolute quantities of energy. In more detail, the appliance was characterized by (i) the amount of energy actually consumed (active power), (ii) the amount of energy that does not get dissipated (reactive power), and with (iii) the ratio between the consumed energy and the measured apparent power. Additionally, a rule and correspondent mathematical foundation were proposed to identify the required steady-states on a signal. To assess the proposed signature, two datasets comprising measurements of active power, voltage, current and power factor signals were gathered, features were computed, and device identification was carried out with the 5-NN and SVM classification methods for both carried out experiments. This experimental evaluation showed that the simplest methods are able to tackle the device identification accurately: the yielded mean accuracies and F-measures were superior to 95%. Still, the studied multi-class problems are too small to represent all the complexity associated with a larger multinomial problem as a represention of our home. Then, for a more robust and more accurate signature evaluation, dataset composed by measurements of a larger number of devices would be required.

Alternatively to the classification approaches, the separation and extraction of relevant information from the whole-home electrical consumption signal could be explored from the signal processing point of view. In a preceding study concerning the disaggregation of the signal into consumption estimates of each appliance, the extraction of variations associated with a particular set of devices from the single-point sensed data was explored as detailed in Chapter 4. Home appliances were grouped into two distinct sets according to the need of human intervention to turn them on and off. For instance, a refrigerator once it is on, it operates automatically without any human related intervention. In this perspective, the goal was to extract the variations in the aggregated consumption associated with these 'automatic' devices. For this purpose, approaches based on wavelet decomposition and SVD were explored, in particular a new technique based on Wavelet Shrinkage, which allows the extraction of information from the aggregated signal considering several of its segments (analysed by possible distinct mother wavelets), was presented. The designed computational experience, yielded over real-world data, allowed for the selection of a small set of mother wavelets appropriated for the extraction of information associated with the 'automatic' devices from the aggregated signal. In particular, the experiments shown that the rbio3.1, rbio3.3 and bior3.1 are suitable wavelet functions for the analysis performed. For the new approach, it was possible to verify that for 50% of the signals in the dataset the new approach employed a median number of 10 and 8 distinct functions, considering the different levels of decomposition. Moreover, the performance of the new approach EWD was superior to the Wavelet Shrinkage in 37% (using two levels of decomposition) in terms of correlation between the separated variations and the prototypes of information being extracted. The favourable performance of the new approach with regard to the task at hand was further validated, and indeed a statistically significant difference exists between the performance of the new method and the other studied techniques.

Pursuing the formulation of energy disaggregation as signal processing problem, the disambiguation of the whole-home electrical consumption into the electricity usage of each appliance connected in the network was investigated as a single-channel source separation problem as detailed in Chapter 5. Regarding this re-interpretation of NILM, the several intervening sources can be modelled, in particular by data-adaptive representations, to allow the inference of consumptions over an unseen period of time and only provided the aggregated

signal. A new strategy based on multi-dimensional arrays and its non-negative factorization was explored for source modelling within the energy disaggregation problem. The described STMF approach computes a 'global' source model whose factors comprised the relevant information for each of the three domains: time of the day, days of the week and appliances. Provided a set of unseen aggregated signals, the inference of the electrical consumption for each device is calculated via non-negative matrix factorization into which the previously learned model is embedded. To assess the adequacy of the novel approach for solving the electrical disambiguation, computational experiments yielded in a reference dataset (REDD) were carried out. The effectiveness of STMF was demonstrated in comparison with a method from the related literature based on sparse coding (DDSC). In fact, the overall RMSE achieved by DDSC was 8.8 times greater than the error yielded by the new method. Regarding the disaggregation errors, apart from House 4 of the dataset, the new method achieved considerably lower values than the sparse coding based method. Moreover, the analysis of the consumption profiles and statistical validation, which showed the existence of a statistical significant difference between the performance of the methods, reinforced the effectiveness of STMF for solving the task at hand. Furthermore, the importance of the sparseness regularization parameter for a successful electrical disaggregation using sparse coding based approach was investigated. The empirical evaluation indicated that a difference exists between the lower and the higher sparsity degree in terms of root-mean-square-error. In detail, the best results were achieved for the lowest sparseness degrees imposed. For instance in House 1 of the dataset, an increase of 4.82% at test in terms of overall RMSE was observed, when the number of bases was fixed to $r = 30$, and the sparseness parameter was changed from $\lambda = 0.0001$ to $\lambda = 0.01$.

In short, the approaches proposed in this Thesis were demonstrated as efficient to tackle partially the challenges addressed in Chapter 1, namely algorithms for energy disaggregation. The proposed approaches could be further explored in a real-world application domain as well as in the scientific point of view.

6.2. Deployability

The ideas proposed in this Thesis can be extended to be incorporated in a commercial product comprising a NILM system. The market already presents

one operational NILM, commercialized by Enetics (Enetics Inc., 2010). The background research associated to the Enetics product is presented in (Drenker and Kader, 1999). The approach described therein employs a recorder which detects and stores in memory the occurring events. The data is stored during a period of two weeks and then it is downloaded to a master station where the pattern recognition takes place. The residential beta test program occurred in collaboration with utilities. These companies selected and installed the record device and the sub-metering sensors in a set of houses used then for performance assessment of the proposed NILM framework. The good results of beta testing led to the commercialization of the system by Enetics in 1996.

Furthermore, the separation and inference of electrical consumption information are essential for energy management, thereby, the work here proposed could be integrated into a household energy management system which additionally would provide personalized energy suggestions and, for instance, predictions about the energy usage of each appliance. Moreover, energy information, the device-level detailed electrical usage in particular, could also be included on in-Home Activity Tracking casted in an Ambient Assisted Living (AAL) framework as mentioned in Chapter 4 Section 4.6. The appliance specific usage would facilitate the study and modelling of behaviours associated with the consumption of energy. The launch of alerts in case of unusual occurrences or unexpected behavioural patterns would be enabled. In fact, an approach to detect activities of daily living using as input data the outcome of a NILM system is proposed in (Berenguer et al., 2008). The study therein used a Watteco (Watteco SAS, 2013) NILM system, which is no longer available on their internet page. Note that in-Home Activity Tracking as above described would be useful for helping both elderly caregivers and parents, who would have the opportunity to control their children in-home activities as suggested in the survey performed by Sundramoorthy et al. (2011).

6.3. Future Work

The study conducted along this Thesis led to interesting possibilities for further scientific investigation. In order that appliance detailed consumption information, *e.g.* provided by future smart meters, becomes a reality the exploration of disaggregation algorithms should pursue robustness, disaggregation accuracy and scalability of NILM for larger numbers of appliances. Furthermore, investi-

gation should seek for methods with undemanding requirements concerning the sampling rate and training.

In order to improve recognition accuracy and to minimize the training time, the parameters of classification methods could be optimized with specialized techniques. For instance, the appliance recognition via Adaptive Neural Networks (Palmer-Brown and Kang, 2005; Beliczynski, 2005) could benefit of a set of optimal parameters calculated with Particle Swarm Optimization. Additionally, other techniques could be explored as the Extreme Learning Machines which are shallow architectures with high potential in regression and classification problems (Huang et al., 2006). Also regarding classification methods, unsupervised approaches arise as a promising direction of investigation since the required learning step would be eliminated. In this line, Self-Organizing Trees have been effectively applied in other domains for unsupervised pattern classification. Thereby, its adequacy for the energy disaggregation solved as a classification task could be further investigated.

In fact, research has been focused on supervised classification methods as reviewed in Chapter 2 Section 2.6. Regarding the challenge of reducing the training requirements, investigation efforts should also consider single-channel blind source separation for NILM interpreted in the light of signal processing. In this case, non-negative matrix and tensor factorization techniques successfully applied for the task of blind source separation in other research domains could also be explored for unsupervised energy disaggregation.

In addition to the above described interpretations of NILM, the disaggregation of the whole-home electrical consumption could be solved as an optimization problem as mentioned in Chapter 2 Section 2.6. Under this context, the idea is to match a new appliance profile (signature) with the known appliance signatures with minimum error, for on/off appliances in particular. In this point of view, the energy disaggregation is related with the Knapsack problem, which is NP-hard, and the possible techniques to solve it arise as methods worthwhile of investigating, namely genetic and evolutionary optimization approaches.

Finally, notwithstanding that disaggregation algorithms are mainly focused on electrical energy usage measured at a single point, a possible direction of research would explore its adequacy for the disambiguation of other consumption measurements as water and gas usage.

APPENDIX A

Impact of Temporal Window-Size on the Performance of EWD

The approach EWD presented in Chapter 4 Section 4.4.3 requires the setup of the window length L. According to Golyandina et al. (2001), when no previous knowledge is available L should be set to half the length of the signal in analysis. For the performance evaluation presented in Chapter 4, this suggestion was followed. Still, the performance of EWD with distinct window sizes was further explored and is next presented.

For this experiment, the Household Electrical Signal Dataset described in Chapter 4 Section 4.5.1 is used. In fact, the experimental setup described therein is used in the following with the exception of the L values. The signals in the dataset have 96 samples then L values equal or higher than 48 are considered. Additionally, note that the length of the segment in analysis by the Wavelet Decomposition is downsampled by two at each level of decomposition until reach the level J. Thereby, in this experience, the selected L values verify

$$L \bmod 2^J = 0 \tag{A.1}$$

where $J \in \{2, 3\}$. In particular, the EWD was performed using $L \in \{48, 64, 72, 88\}$.

Figure A.1 presents the boxplots of the accomplished results for each window-size in each defined scenario, that is, applying the method with different levels of approximation and details, namely, EWD-2 and EWD-3. For all cases, the interquartile range indicates the existence of some dispersion in the achieved

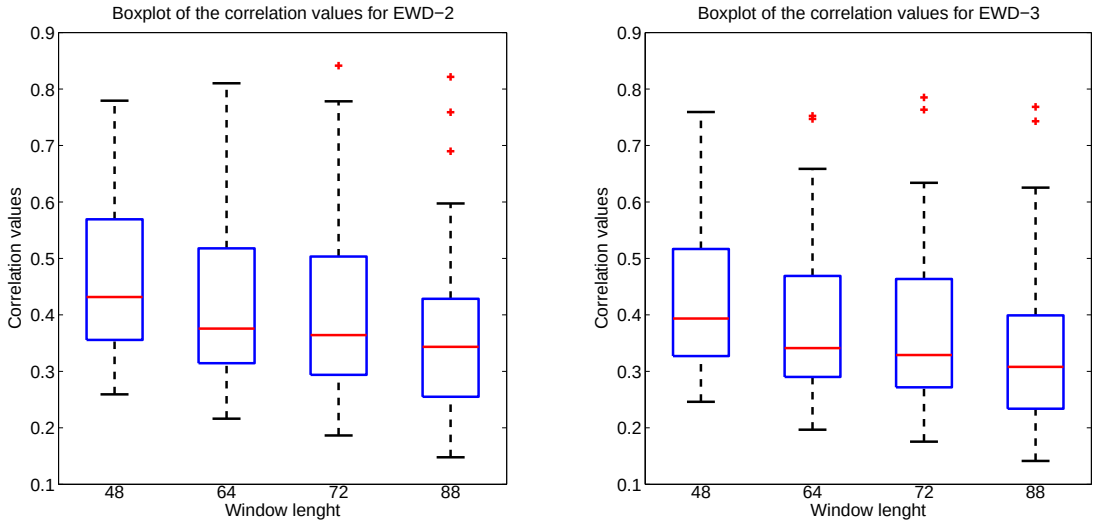

Figure A.1.: Box plots of the correlation values achieved by the EWD with 4 different window size, considering two and three levels of decomposition.

M	48	64	72	88
EWD-2	0.4317 ± 0.1302	0.3758 ± 0.1360	0.3642 ± 0.1340	0.3433 ± 0.1295
EWD-3	0.3936 ± 0.1219	0.3412 ± 0.1270	0.3289 ± 0.1248	0.3079 ± 0.1205

Table A.1.: The mean and standard deviation of the results.

results. Additionally, for both EWD-2 and EWD-3, the median correlation is very similar, for each L. Still, as the window size increases the median values decrease, independently of the decomposition level. A similar behaviour is observed for the mean values of EWD-2 and EWD-3, presented in Table A.1. Note that an increase from $L = 48$ to $L = 88$ led to a decrease of 0.10 (0.0928) in the mean correlation value for the EWD-2 (EWD-3, respectively). The best correlation values are achieved by the lowest window-size. A low L value allows the partition of the signal in more segments thus, more suitable mother wavelets can be employed to extract information of each part of the signal. On the other hand, for a high L value less and longer segments are defined and, therefore, the most suitable wavelet function for all the segment may not perform properly for specific parts of that segment. Note also that the two levels of decomposition are appropriate for the task at hand since the EWD-2 performance is superior to the EWD-3 in terms of correlation (for all the L values tested) in accordance with the observations drawn in Chapter 4 Section 4.5.2.

APPENDIX B

Exploratory Comparison of the Sampling Rate

As detailed in Chapter 2 Section 2.4, the sampling rate for state-of-art NILM frameworks depends on the type of signature used to characterize the devices in study. Within the suitable ranges indicated for each signature, no particular rate is known as most reliable for an accurate energy disaggregation. Likewise, for energy disaggregation interpreted as a single-source separation problem no related literature was found addressing this issue. Hence, we carried out an exploratory comparison on results using the post-processed REDD to evaluate the influence of the sampling rate on energy disaggregation.

Notice that the data acquisition sensor also plays an important role on the selection of the most appropriate rate. For instance, the smart meters in development in United Kingdom (UK) are able to report one measurement *per* minute (Parson et al., 2011), thereby, the downsampling performed in Chapter 5 Section 5.5.1 aimed at the simulation of the same context. Nevertheless, there are smart meters with lower rates, *e.g.* one measurement *per* fifteen minutes (Zoha et al., 2012). To mimic a similar condition we re-defined the median filter applied to the original REDD dataset described in Chapter 5 Section 5.5.1 such that signal samples became spaced five minutes apart.

For the following analysis, two datasets are considered: (i) Dataset A, the post-processed REDD as described in Chapter 5 Section 5.5.1 whose samples within each signal are spaced one minute apart; and (ii) Dataset B which is the

outcome of the same pre-processing phase of Dataset A, differing on median filter applied, and whose signal samples are spaced five minutes apart.

The energy consumption for each house in the datasets was disaggregated applying the STMF method. The results next presented report means of 30 runs for each one of the datasets under the experimental parameters defined in Chapter 5 Section 5.5. The RMSE is the most commonly used metric for comparison of results, thereby we considered the difference between the achieved overall RMSE values for each dataset:

$$\Delta_{overallRMSE} = \text{Overall RMSE}_{\text{Dataset B}} - \text{Overall RMSE}_{\text{Dataset A}}. \tag{B.1}$$

As it can be observed in Figure B.1, the value of $\Delta_{overallRMSE}$ is virtually null, as the R is incremented for training, *i.e.*, the achieved overall RMSE performance was similar for both datasets, Houses 2 and 6 in particular. The highest difference is observed for House 2 with a $R = 5$, for which the computed aggregated signals of Dataset A were closer to the ground truth than for Dataset B. However, as R increases, the performance in terms of overall RMSE tends to be similar for both datasets. At test, the highest differences occur again for $R = 5$. Moreover, the achieved $\Delta_{overallRMSE}$ remain stable for $R > 10$, regarding the test results of each house, with the possible exception of House 1. In general, a better approximation of overall consumption is achieved by the $\frac{1}{60}$Hz sampling rate since $\Delta_{overallRMSE} \geq 0.0005$. Nevertheless, energy disaggregation aims at source separation thereby it is necessary to evaluate how accurately individual sources are predicted.

For more detail about the individual performance for each device, the differences for the RMSE by appliance were calculated, using the same reasoning,

$$\Delta_{RMSEbyappliance} = \text{RMSE by appliance}_{\text{Dataset B}} - \text{RMSE by appliance}_{\text{Dataset A}}. \tag{B.2}$$

Recalling the considerations drawn in Chapter 5 Section 5.5.2 for the overall RMSE, similar trends were observed for the corresponding values achieved in Dataset B. Thereby, in what follows, the difference between the results reached by the STMF using $R = 60$ for each datasets are presented.

Figure B.2 reports the computed $\Delta_{RMSEbyappliance}$ values under the defined conditions. Once again, either none or only slight variation was observed on the RMSE by appliance across the several houses for the training phase. The

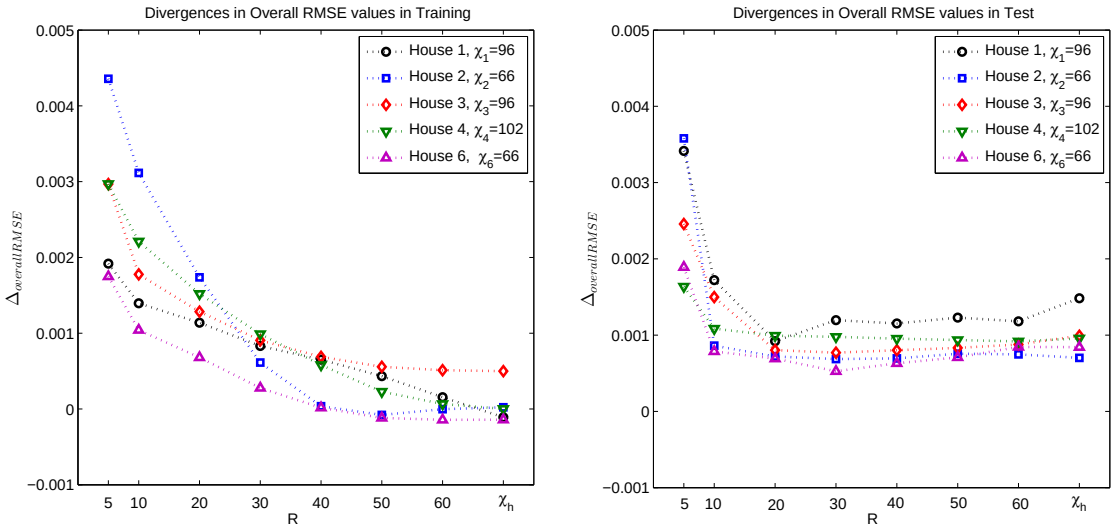

Figure B.1.: The $\Delta_{overallRMSE}$ between the overall RMSE performance of Dataset A and Dataset B.

only possible exception would be the group "Others" for House 3, presenting a higher value of 0.0005. Indeed, the RMSE associated to this group was of 0.0007 in Dataset B while for the remaining groups the error was always smaller than 0.0002, likewise to the results for Dataset A.

More significant variations were observed for the test results, the Houses 3 and 6 in particular. The $\Delta_{RMSEbyappliance}$ achieved for the "Dishwasher", "Others" and "Air Condition" in these houses is superior to 0.003. For House 3, the error associated to the group "Dishwasher" in Dataset B was 2.44 times superior than for Dataset A. Similar differences are presented for the group "Air Conditioning" in House 6. In general, the error observed for Dataset B across the several devices of the different houses is always higher than the one for the RMSE achieved for Dataset A.

In short, the sampling frequency influences the effectiveness of energy disaggregation. This empirical comparison showed that valuable information is possibly lost when considering an interval of 5 minutes between samples. A change from a rate of $\frac{1}{60}Hz$ to $\frac{1}{300}Hz$ led to an increase in the error associated to the computed signals of each device. Still, it remains to be understood if a sampling rate of $\frac{1}{60}Hz$ is the most adequate for the energy disaggregation performance.

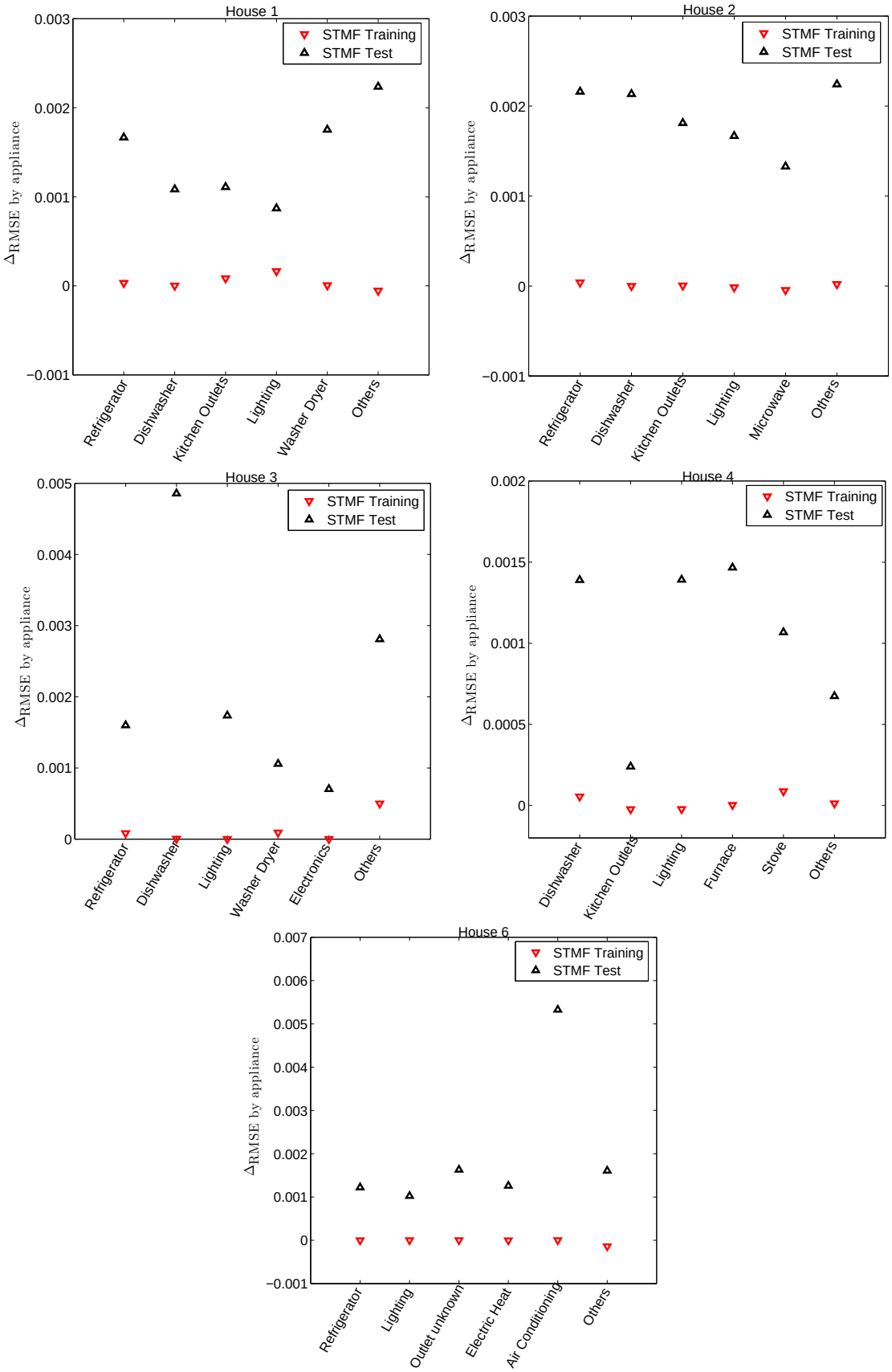

Figure B.2.: The differences on the RMSE by appliance between both datasets.

APPENDIX C

Statistical Analysis of the Sparseness Condition of DDSC

The true importance of imposing a higher degree of sparseness in the DDSC method, described in Chapter 5 Section 5.4.1, is next clarified by statistical tests. The existence of similarities between the overall RMSE results of all the houses in the dataset, presented in Chapter 5 Section 5.5, for $\lambda = 0.0001$ and $\lambda = 0.01$, considering $r = 30$ is analysed in the following.

In accordance with the statistical validation procedure along this Thesis, firstly a Kolmogorov-Smirnov test is performed. This test was held with a significance level of 0.01. The *p-values* produced were lower than 0.001 for all the tested cases. Therefore, we are led to conclude that the overall RMSE, for both training and test, is far from having a normal distribution. Consequently, a non-parametric test for comparison of both levels of sparseness considered as independent conditions was pursued. The Mann-Whitney test was chosen. The goal is to assess whether the performance of the method differs significantly for both λ values.

	RMSE			
	Training		**Test**	
	$\lambda = 0.0001$	$\lambda = 0.01$	$\lambda = 0.0001$	$\lambda = 0.01$
Median (*Mdn*)	0.0004	0.0012	0.0036	0.0038

Table C.1.: Median (*Mdn*) values for RMSE regarding both $\lambda = 0.0001$ and $\lambda = 0.01$ using 150 samples for each test.

	RMSE	
	Training	**Test**
Mann-Whitney U	18926	14498
z	10.218	4.323
Asymp. Sig. (2-tailed)	<0.001	<0.001

Table C.2.: Test statistics for RMSE regarding both λ values.

Table C.2 reports the U statistic associated with the Mann-Whitney test, the correspondent z-score (z) and the asymptotic significance (Asymp. Sig.). According to Mann-Whitney test results, the distribution associated to overall RMSE is not the same for both λ values of sparseness, neither in training nor in test, at a significance level of 0.01. In fact, in regard to the mean values (Mdn), the RMSE values for $\lambda = 0.01$ ($Mdn = 0.0012$) are higher than the ones for $\lambda = 0.0001$ ($Mdn = 0.0004$) for the training, $U = 18926$ and $z = 10.218$. Moreover, for the test similar observations can be drawn: the RMSE values associated $\lambda = 0.01$ ($Mdn = 0.0038$) differ from the ones for $\lambda = 0.0001$ ($Mdn = 0.0036$), $U = 14498$ and $z = 4.323$. Thus, the statistical results support the conclusion that a difference statistically significant exists between the DDSC performance when using a sparseness degree of $\lambda = 0.01$ and $\lambda = 0.0001$, in terms of the overall RMSE.

APPENDIX D

Analysis of RMSE by Appliance for the STMF

In this chapter, additional results are presented and discussed for the approach Source Separation by Tensor and Matrix Factorization (STMF) proposed in Chapter 5 Section 5.4.2.

In what follows, the RMSE by appliance as defined in Chapter 5 Section 5.5 is presented based on the overall RMSE achieved by STMF under the setup defined in Chapter 5 Section 5.5. In particular, we report the RMSE by appliance associated with the minimum and maximum overall RMSE values yielded by STMF at training and test steps, in accordance with the results presented in Chapter 5 Section 5.5.2.

From Figure 5.7 in Chapter 5 Section 5.5.2, we concluded that the maximum overall RMSE value was achieved by $R = 5$ in both training and test steps, considering all houses in the dataset. Additionally, from the decreasing trend observed for the training results, the minimum error was yielded for $R = \chi$, the typical rank of the tensor associated with each house. On the contrary, no clear trend was observed for the test results. Instead, it was observed that for $R \geq 20$ the overall RMSE was very similar. Considering Houses 2 and 4, the minimum overall RMSE was achieved for $R = 50$ while for Houses 1, 3 and 6 the error was the smallest when the tensor factorization was performed for $R = 20$, $R = 30$ and $R = 40$, respectively. Table D.1 presents the minimum and maximum overall RMSE values yielded by the STMF method for all the houses

	Training		Test	
	Minimum	**Maximum**	**Minimum**	**Maximum**
House 1	0.00014	0.00191	0.00066	0.00297
House 2	0.00012	0.00412	0.00055	0.00145
House 3	0.00020	0.00259	0.00048	0.00238
House 4	0.00001	0.00236	0.00074	0.00113
House 6	0.00019	0.00148	0.00045	0.00247

Table D.1.: Minimum and maximum overall RMSE values yielded by the STMF method.

in the dataset.

Figure D.1 illustrates the RMSE by appliance for the above described R values for all the houses in the dataset. As expected, the minimum overall RMSE at training corresponds to virtually null RMSE values for all the appliances across all the houses in study. Regarding the results when the overall error was the highest from the training step, the RMSE by appliances is always similar or smaller than the correspondent errors at test (in the interval $[0.0002, 0.0018]$) with few exceptions like the group "Dishwasher" in Houses 1 and 2, the "Furnace" and "Stove" for House 4.

Regarding the errors corresponding to the test step, when the overall RMSE was the highest, the associated RMSE by appliances was also the highest or similar to corresponding values concerning the minimum overall RMSE. Nevertheless, House 6 is an exception. In this case, apart from the "Lighting" and "Others", the reported RMSE by appliance, associated with the maximum overall RMSE of the aggregated signal, was always smaller than the corresponding values for the minimum overall RMSE. Similar observations can be drawn for the "Dishwasher" at Houses 1 and 3, for "Kitchen Outlets" and "Lighting" at House 2 and for the "Stove" and "Others" at House 4. For instance, for the group "Dishwasher", the error associated to the maximum RMSE corresponded to only 50% of the error achieved when the overall RMSE is minimum. On the other hand, for the maximum overall RMSE, the error associated with "Kitchen Outlets" at House 2 and "Others" at House 6 the was very similar to the value achieved when the overall RMSE is minimum. In fact, the former error represented 80% of the latter. Still, these results led us to conclude that in some cases, a low RMSE computed for the aggregated signal does not necessarily corresponds to low RMSE values at the device level.

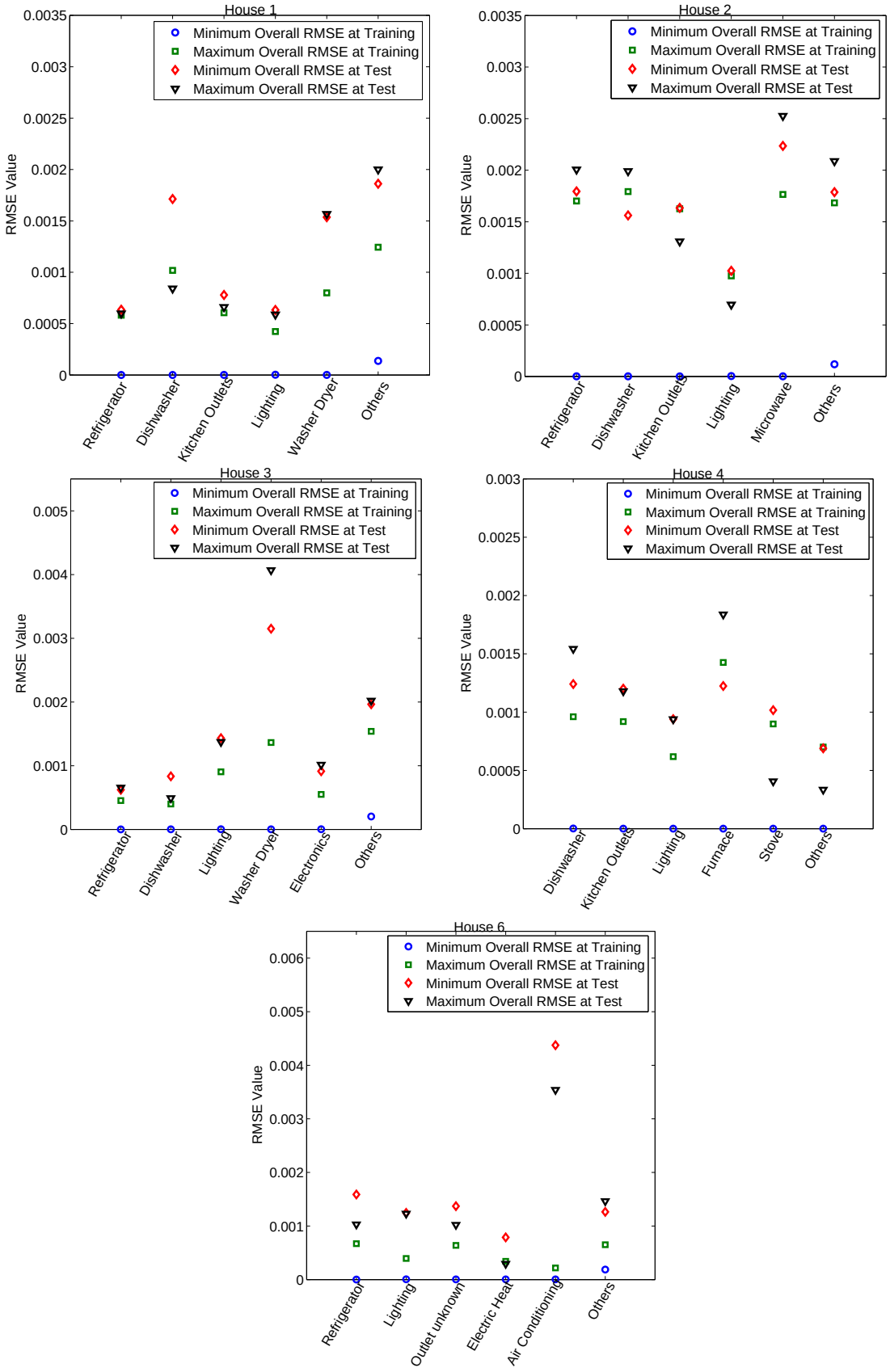

Figure D.1.: The RMSE by appliance for the minimum and maximum overall RMSE achieved by STMF at training and test steps.

Bibliography

J. Abreu and F. C. Pereira. Household electricity consumption routines and tailored feedback. In *ACEEE Summer Study on Energy Efficiency in Buildings*, 2012. [cited at page 3, 6]

J. Abreu, F. C. Pereira, J. Vasconcelos, and P. Ferrão. An approach to discover the potential for demand response in the domestic sector. In *Proceedings of IEEE Conference on Innovative Technologies for an Efficient and Reliable Electricity Supply*, pages 240–245, 2010. [cited at page 97]

M. Alfaouri and K. Daqrouq. ECG signal denoising by wavelet transform thresholding. *American Journal of Applied Sciences*, 5(3):276–281, 2008. [cited at page 82]

K. Anderson, M. Berges, A. Ocneanu, D. Benitez, and J. M. F. Moura. Event detection for non intrusive load monitoring. In *Proceedings of the Annual Conference on IEEE Industrial Electronics Society*. IEEE, 2012a. [cited at page 51]

K. Anderson, A. Ocneanu, D. Benitez, D. Carlson, A. Rowe, and M. Berges. BLUED: a fully labeled public dataset for event-based non-intrusive load monitoring research. In *Proceedings of the SustKDD Workshop on Data Mining Applications in Sustainability*, 2012b. [cited at page 23, 54]

C. A. Andersson and R. Bro. The n-way toolbox for matlab. *Chemometrics and Intelligent Laboratory Systems*, 52(1):1–4, 2000. [cited at page 135]

E. J. Balster, Y. F. Zheng, and R. L. Ewing. Feature-based wavelet shrinkage algorithm for image denoising. *IEEE Transactions on Image Processing*, 14(12): 2024–2039, 2005. [cited at page 91]

M. Baranski and J. Voss. Genetic algorithm for pattern detection in nialm systems. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, volume 4, pages 3462–3468, 2004a. [cited at page 27, 28, 39]

M. Baranski and J. Voss. Detecting patterns of appliances from total load data using a dynamic programming approach. In *Proceedings of the IEEE International Conference on Data Mining*, pages 327–330, 2004b. [cited at page 28, 39]

D. Barber. *Bayesian Reasoning and Machine Learning*. Cambridge University Press, 2012. [cited at page 63]

S. Barker, A. Mishra, D. Irwin, E. Cecchet, and P. Shenoy. Smart*: an open data set and tools for enabling research in sustainable homes. In *Proceedings of the SustKDD Workshop on Data Mining Applications in Sustainability*, 2012. [cited at page 24]

B. Beliczynski. Certain comments on data preparation for neural networks based modelling. In B. Ribeiro, R.. Albrecht, A. Dobnikar, D. Pearson, and N. Steele, editors, *Adaptive and Natural Computing Algorithms*, pages 5–8. Springer Vienna, 2005. [cited at page 164]

A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159, 1995. [cited at page 117]

A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines. A blind source separation technique using second-order statistics. *IEEE Transactions on Signal Processing*, 45(2):434–444, 1997. [cited at page 117]

M. Berenguer, M. Giordani, F. Giraud-By, and N. Noury. Automatic detection of activities of daily living from detecting and classifying electrical events on the residential power line. In *Proceedings of International Conference on e-health Networking, Applications and Services*, pages 29–32, 2008. [cited at page 163]

J. Berge. The typical rank of tall three-way arrays. *Psychometrika*, 65(4):525–532, 2000. [cited at page 138]

M. Berges, E. Goldman, H. S. Matthews, and L. Soibelman. Training load monitoring algorithms on highly sub-metered home electricity consumption data. *Tsinghua Science & Technology*, 13(1):406–411, 2008. [cited at page 23]

M. Berges, E. Goldman, H. S. Matthews, and L. Soibelman. Learning systems for electric consumption of buildings. In *Proceddings of ASCE International Workshop on Computing in Civil Engineering*, 2009. [cited at page 29, 40, 50]

M. Berges, E. Goldman, H. S. Matthews, and L. Soibelman. Enhancing electricity audits in residential buildings with nonintrusive load monitoring. *Journal of Industrial Ecology*, 14(5):844–858, 2010. [cited at page 40]

M. Berges, E. Goldman, L. Soibelman, H. S. Matthews, and K. Anderson. User-centered non-intrusive electricity load monitoring for residential buildings. *Journal of Computing in Civil Engineering*, 25(1), 2011. [cited at page 22, 23, 39, 62]

A. J. Bijker, X. Xia, and J. Zhang. Active power residential non-intrusive appliance load monitoring system. In *Proceedings of the IEEE AFRICON*, 2009. [cited at page 27, 28]

C. M. Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer, 2006. [cited at page 42, 44]

J. Cardoso. Infomax and maximum likelihood for blind source separation. *IEEE Signal Processing Letters*, 4(4):112–114, 1997. [cited at page 117]

J. D. Carroll and J. Chang. Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 35(3):283–319, 1970. [cited at page 125, 126]

W. L. Chan, A. T. P. So, and L. L. Lai. Harmonics load signature recognition by wavelets transforms. In *Proceedings of International Conference on Electric Utility Deregulation and Restructuring and Power Technologies*, pages 666–671, 2000. [cited at page 29, 30, 31]

C. Chang and C. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3):1–27, 2001. software available at (`http://www.csie.ntu.edu.tw/~cjlin/libsvm`). [cited at page 66, 71]

H. Chang. Non-intrusive demand monitoring and load identification for energy management systems based on transient feature analyses. *Energies*, 5(11): 4569–4589, 2012. [cited at page 35]

H. Chang and H. Yang. Applying a non-intrusive energy-management system to economic dispatch for a cogeneration system and power utility. *Applied Energy*, 86(11):2335–2343, 2009. [cited at page 35]

H. Chang, C. Lin, and H. Yang. Load recognition for different loads with the same real power and reactive power in a non-intrusive load-monitoring system. In *Proceedings of International Conference on Computer Supported Cooperative Work in Design*, pages 1122–1127, 2008a. [cited at page 35]

H. Chang, H. Yang, and C. Lin. Load identification in neural networks for a non-intrusive monitoring of industrial electrical loads. In W. Shen, J. Yong, Y. Yang, Jean-P. A. Barthès, and J. Luo, editors, *Computer Supported Cooperative Work in Design IV*, volume 5236 of *Lecture Notes in Computer Science*, pages 664–674. Springer Berlin Heidelberg, 2008b. [cited at page 35]

H. Chang, C. Lin, and J. Lee. Load identification in nonintrusive load monitoring using steady-state and turn-on transient energy algorithms. In *Proceedings of International Conference on Computer Supported Cooperative Work in Design*, pages 27–32, 2010. [cited at page 35]

S. Choi, A. Cichocki, H. Park, and S. Lee. Blind Source Separation and Independent Component Analysis: A Review. *Neural Information Processing - Letters and Reviews*, 6(1):1–57, 2005. [cited at page 119]

A. Cichocki, R. Zdunek, A. H. Phan, and S. Amari. *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*. John Wiley & Sons, Ltd, 2009. [cited at page 116, 120, 121, 123, 125, 126, 127, 128]

A. I. Cole and A. Albicki. Data extraction for effective non-intrusive identification of residential power loads. In *Proceedings of the IEEE Conference on Instrumentation and Measurement Technology*, volume 2, pages 812–815, 1998a. [cited at page 23, 27, 34, 37, 49]

A. I. Cole and A. Albicki. Algorithm for non-intrusive identification of residential appliances. In *Proceedings of the IEEE International Symposium on Circuits and Systems*, volume 3, pages 338–341, 1998b. [cited at page 27]

C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20(3): 273–297, 1995. [cited at page 64]

S. Darby. The effectiveness of feedback on energy consumption: a review for DEFRA of the literature on metering, billing and direct displays. Technical report, Environmental Change Institute, University of Oxford, 2006. [cited at page 3, 14]

M. E. Davies and C. J. James. Source separation using single channel ICA. *Signal Processing*, 87(8):1819–1832, 2007. [cited at page 119]

A. de Almeida. About nonnegative matrix factorization: On the posrank approximation. In A. Dobnikar, U. Lotric, and B. Ster, editors, *Adaptive and Natural Computing Algorithms*, volume 6594 of *Lecture Notes in Computer Science*, pages 295–304. Springer Berlin Heidelberg, 2011. [cited at page 120]

D. Donoho. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41(3):613–627, 1995. [cited at page 56, 71, 89, 102]

D. Donoho and I. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994. [cited at page 56, 89, 91]

D. Donoho and I. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90(432):1200–1224, 1995. [cited at page 56, 89, 91]

S. Drenker and A. Kader. Nonintrusive monitoring of electric loads. *IEEE Computer Applications in Power*, 12(4):47–51, 1999. [cited at page 163]

C. Eckart and G. Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, 1936. [cited at page 82]

J. Eggert and E. Korner. Sparse coding and NMF. In *Proceedings of the IEEE International Joint Conference on Neural Networks*, volume 4, pages 2529–2533, 2004. [cited at page 122, 128, 129]

M. A. El-Sharkawi. *Electric Energy: An Introduction*. Power Electronics and Applications Series. CRC Press, 2012. [cited at page 15]

M. Elad. Sparse and redundant representation modeling - what next? *IEEE Signal Processing Letters*, 19(12):922–928, 2012. [cited at page 120]

Enetics Inc. Enetics advanced metering and analyzers. `http://www.enetics.com/`, 2010. [Online; accessed 13-September-2013]. [cited at page 163]

European Commission. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the regions - facing the challenge of higher oil prices, 2008. [cited at page 2]

European Commission. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions - Energy Efficiency Plan, 2011. [cited at page 2]

Eurostat. Eurostat Pocketbook - Energy, transport and environment indicators - 2012 edition. Eurostat Pocketbook, 2012a. [cited at page xiii, 1, 2, 3, 14]

Eurostat. Eurostat Statistical books - Europe in figures - Eurostat yearbook 2012. Eurostat Statistical books, 2012b. [cited at page 2]

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*, 30(3):245–259, 1999. [cited at page 20, 23, 27, 28, 39, 40, 50]

E.-B. Fgee, W. J. Phillips, and W. Robertson. Comparing audio compression using wavelets with other audio compression schemes. In *Proceedings of IEEE Canadian Conference on Electrical and Computer Engineering*, volume 2, pages 698–701, 1999. [cited at page 99]

A. Field. *Discovering Statistics Using IBM SPSS Statistics*. SAGE Publications, 2013. [cited at page 103, 104, 106, 139]

M. Figueiredo, A. de Almeida, B. Ribeiro, and A. Martins. Extracting features from an electrical signal of a non-intrusive load monitoring system. In C. Fyfe, P. Tino, D. Charles, C. Garcia-Osorio, and H. Yin, editors, *Intelligent Data Engineering and Automated Learning*, volume 6283 of *Lecture Notes in Computer Science*, pages 210–217. Springer Berlin Heidelberg, 2010. [cited at page 7]

M. Figueiredo, A. de Almeida, and B. Ribeiro. An experimental study on electrical signature identification of non-intrusive load monitoring (NILM) systems. In A. Dobnikar, U. Lotric, and B. Ster, editors, *Adaptive and Natural Computing Algorithms*, volume 6594 of *Lecture Notes in Computer Science*, pages 31–40. Springer Berlin Heidelberg, 2011a. [cited at page 7]

M. Figueiredo, A. de Almeida, and B. Ribeiro. Smart home: A novel model for denoising an electrical signal. In *Proceedings of International Conference on Intelligent Systems Design and Applications*, pages 784–789, 2011b. [cited at page 7]

M. Figueiredo, A. de Almeida, and B. Ribeiro. Impact of temporal window-size on pattern denoising of a smart home electrical signal. In *Proceedings of Portuguese Conference on Pattern Recognition*, 2011c. [cited at page 7]

M. Figueiredo, A. de Almeida, and B. Ribeiro. Wavelet decomposition and singular spectrum analysis for electrical signal denoising. In *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*, pages 3329–3334, 2011d. [cited at page 7]

M. Figueiredo, A. de Almeida, and B. Ribeiro. Home electrical signal disaggregation for non-intrusive load monitoring (NILM) systems. *Neurocomputing*, 96:66–73, 2012a. [cited at page 7]

M. Figueiredo, A. de Almeida, and B. Ribeiro. Another perspective on NILM systems for informed energy consumer behavior. In *Proceedings of International Workshop on Energy Efficiency for a More Sustainable World*, 2012b. [cited at page 8]

M. Figueiredo, B. Ribeiro, and A. de Almeida. On the regularization parameter selection for sparse code learning in electrical source separation. In M. Tomassini, A. Antonioni, F. Daolio, and P. Buesser, editors, *Adaptive and Natural Computing Algorithms*, volume 7824 of *Lecture Notes in Computer Science*, pages 277–286. Springer Berlin Heidelberg, 2013a. [cited at page 8]

M. Figueiredo, B. Ribeiro, and A. de Almeida. Electrical signal source separation via non-negative tensor factorization using on-site measurements in a smart home. *IEEE Transactions on Instrumentation and Measurement (article in press)*, 2013b. [cited at page 8]

C. Fischer. Feedback on household electricity consumption: a tool for saving energy? *Energy Efficiency*, 1:79–104, 2008. [cited at page 3, 14]

D. Fitzgerald, M. Cranitch, and E. Coyle. Non-negative tensor factorisation for sound source separation. In *Proceedings of the Irish Signals and Systems Conference*, 2005. [cited at page 116]

J. Friedman, T. Hastie, H. Höfling, and R. Tibshirani. Pathwise coordinate optimization. *The Annals of Applied Statistics*, 1(2):302–332, 2007. [cited at page 129]

J. Friedman, T. Hastie, and R. Tibshirani. A note on the group lasso and a sparse group lasso. Technical report, Stanford University, Jan 2010. [cited at page 129]

J. H. Friedman. Another approach to polychotomous classification. Technical report, Department of Statistics, Stanford University, 1996. [cited at page 65]

J. Froehlich, E. Larson, S. Gupta, G. Cohn, M. S. Reynolds, and S. N. Patel. Disaggregated end-use energy sensing for the smart grid. *IEEE Pervasive Computing*, 10(1):28–39, 2011. [cited at page 4, 6]

C. Gargour, M. Gabrea, V. Ramachandran, and J.-M. Lina. A short introduction to wavelets and their applications. *IEEE Circuits and Systems Magazine*, 9(2): 57–68, 2009. [cited at page 87, 91]

G. Gelle, M. Colas, and C. Serviere. Blind source separation: a new pre-processing tool for rotating machines monitoring? *IEEE Transactions on Instrumentation and Measurement*, 52(3):790–795, 2003. [cited at page 117]

S. Giri and M. Berges. A study on the feasibility of automated data labeling and training using an EMF sensor in NILM platforms. In *Proceedings of the International EG-ICE Workshop on Intelligent Computing*, 2012. [cited at page 39]

N. E. Golyandina, V. V. Nekrutkin, and A. A. Zhigljavsky. *Analysis of Time Series Structure: SSA and Related Techniques*, volume 90 of *Monographs on Statistics and Applied Probability*. Chapman&Hall/CRC, 2001. [cited at page 80, 82, 83, 84, 91, 93, 102, I]

H. Gonçalves, A. Ocneanu, and M. Berges. Unsupervised disaggregation of appliances using aggregated consumption data. In *Proceedings of the SustKDD Workshop on Data Mining Applications in Sustainability*, 2011. [cited at page 46]

A. Graps. An introduction to wavelets. *IEEE Computational Science and Engineering*, 2:50–61, 1995. [cited at page 85, 89]

D. Guillamet, B. Schiele, and J. Vitria. Analyzing non-negative matrix factorization for image classification. In *Proceedings of the International Conference on Pattern Recognition*, volume 2, pages 116–119, 2002. [cited at page 120]

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 ACM International Conference on Ubiquitous computing*, pages 139–148. ACM, 2010. [cited at page 5, 33, 34, 40]

R. A. Harshman. Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis. *UCLA Working Papers in Phonetics*, 16(1):84, 1970. [cited at page 125, 126]

G. Hart, E. Kern, and F. Schweppe. Non-intrusive appliance monitor, August 1989. URL `http://www.google.com/patents?vid=4858141`. [cited at page xiii, 14, 21]

G. W. Hart. Residential energy monitoring and computerized surveillance via utility power flows. *IEEE Technology and Society Magazine*, 8(2):12–16, 1989. [cited at page 6, 14]

G. W. Hart. Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80: 1870–1891, 1992. [cited at page 4, 14, 19, 20, 22, 25, 26, 28, 29, 31, 32, 37, 38, 40, 49, 52, 58, 137]

H. Hassani. Singular spectrum analysis: Methodology and comparison. *Journal of Data Science*, 5:239–257, 2007. [cited at page 82]

J. Håstad. Tensor rank is NP-complete. *Journal of Algorithms*, 11:644–654, 1990. [cited at page 138]

Y. Hochberg and A. C. Tamhane. *Multiple comparison procedures*. John Wiley & Sons, Inc., 1987. [cited at page 103, 107]

H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6):417–441, 1933. [cited at page 82]

P. O. Hoyer. Non-negative sparse coding. In *Proceedings of the IEEE Workshop on Neural Networks for Signal Processing*, pages 557–565, 2002. [cited at page 120, 122, 128]

P. O. Hoyer. Non-negative matrix factorization with sparseness constraints. *Journal of Machine Learning Research*, 5:1457–1469, 2004. [cited at page 120, 122, 123, 133]

C. Hsu and C. Lin. A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, 13(2):415–425, 2002. [cited at page 64]

G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew. Extreme learning machine: Theory and applications. *Neurocomputing*, 70:489–501, 2006. [cited at page 164]

L. Hubert, J. Meulman, and W. Heiser. Two purposes for matrix factorization: A historical appraisal. *SIAM Review*, 42(1):68–82, 2000. [cited at page 121]

A. Hyvärinen and E. Oja. Independent component analysis: Algorithms and applications. *Neural Networks*, 13:411–430, 2000. [cited at page 119, 120]

ISA-Intelligent Sensing Anywhere, S. A. ISA Intelligent Sensing Anywhere. `http://www.isasensing.com/`, 2012. [Online; accessed 21-May-2013]. [cited at page xiv, 67]

G. Jang and T. Lee. A maximum likelihood approach to single-channel source separation. *Journal of Machine Learning Research*, 4:1365–1392, 2003. [cited at page 117]

G. Jang, T. Lee, T. Lee, J. Cardoso, E. Oja, and S. Amari. A maximum likelihood approach to single-channel source separation. *Journal of Machine Learning Research*, 4:1365–1392, 2003. [cited at page 119]

M. J. Johnson and A. S. Willsky. Bayesian nonparametric hidden semi-markov models. *Journal of Machine Learning Research*, 14:673–701, 2013. [cited at page 48, 49, 52]

I. T. Jolliffe. *Principal Component Analysis*. Springer Verlag, 2002. [cited at page 121]

R. Kaplan and D. Norton. *The Balanced Scorecard: Translating Strategy into Action*. Harvard Business Review Press, 1996. [cited at page 1]

S. Karjalainen. Consumer preferences for feedback on household electricity consumption. *Energy and Buildings*, 43(2-3):458–467, 2011. [cited at page 3, 14]

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. In *Proceedings of the International Conference on Smart Homes and Health Telematics: Ambient Assistive Health and Wellness Management in the Heart of the City*, pages 150–157, 2009. [cited at page 29, 44]

H. Kim, M. Marwah, M. F. Arlitt, G. Lyon, and J. Han. Unsupervised disaggregation of low frequency power measurements. In *Proceedings of the SIAM International Conference on Data Mining*, pages 747–758, 2011. [cited at page 37, 47, 49]

T. G. Kolda and B. W. Bader. Tensor decompositions and applications. *SIAM Review*, 51(3):455–500, 2009. [cited at page 116, 125, 128, 138]

J. Z. Kolter. The Reference Energy Disaggregation Data Set. `http://redd.csail.mit.edu/`, 2011. [Online; accessed 11-June-2013]. [cited at page 135]

J. Z. Kolter and T. Jaakkola. Approximate inference in additive factorial hmms with application to energy disaggregation. *Journal of Machine Learning Research*, 22:1472–1482, 2012. [cited at page 48, 49]

J. Z. Kolter and M. J. Johnson. REDD: A public data set for energy disaggregation research. In *Proceedings of the SustKDD Workshop on Data Mining Applications in Sustainability*, 2011. [cited at page 23, 52, 135]

J. Z. Kolter, S. Batra, and A. Ng. Energy disaggregation via discriminative sparse coding. In J. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R.S. Zemel, and A. Culotta, editors, *Advances in Neural Information Processing Systems*, volume 23, pages 1153–1161. MIT Press, 2010. [cited at page xix, 5, 52, 116, 128, 129, 130, 135, 137, 156, 157]

W. H. Kruskal and W. A. Wallis. Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, 47(260):583–621, 1952. [cited at page 103]

P. S. Kumar, R. Arumuganathan, K. Sivakumar, and C. Vimal. Removal of ocular artifacts in the EEG through wavelet transform without using an EOG reference channel. *International Journal of Open Problems in Computational Mathematics*, 1(3):188–200, 2008. [cited at page 82]

Y. Lai, C. Lai, Y. Huang, and H. Chao. Multi-appliance recognition system with hybrid SVM/GMM classifier in ubiquitous smart home. *Information Sciences*, 230:39–55, 2013. [cited at page 44]

H. Y. Lam, G. S. K. Fung, and W. K. Lee. A novel method to construct taxonomy of electrical appliances based on load signatures. *IEEE Transactions on Consumer Electronics*, 53(2):653–660, 2007. [cited at page 36]

C. Laughman, K. Lee, R. Cox, S. Shaw, S. Leeb, L. Norford, and P. Armstrong. Power signature analysis. *IEEE Power and Energy Magazine*, 1(2):56–63, 2003. [cited at page xiv, 19, 20, 23, 26, 27, 30, 31, 40]

D. D. Lee and H. S. Seung. Learning the parts of objects by nonnegative matrix factorization. *Nature*, 401:788–791, 1999. [cited at page 119, 121, 122]

D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *In Advances in Neural Information Processing Systems*, volume 13, pages 556–562. MIT Press, 2000. [cited at page 119, 121, 122]

K. D. Lee, S. B. Leeb, L. K. Norford, P. R. Armstrong, J. Holloway, and S. R. Shaw. Estimation of variable-speed-drive power consumption from harmonic content. *IEEE Transactions on Energy Conversion*, 20(3):566–574, 2005. [cited at page 33]

W. K. Lee, G. S. K. Fung, H. Y. Lam, F. H. Y. Chan, and M. Lucente. Exploration on load signatures. In *Proceedings of International Conference on Electrical Engineering*, 2004. [cited at page 36]

S. B. Leeb and J. L. Kirtley Jr. A multiscale transient event detector for nonintrusive load monitoring. In *IEEE International Conference on Industrial Electronics, Control and Instrumentation*, pages 354–359, 1993. [cited at page 32]

S. B. Leeb, M. S. LeVan, J. L. Kirtley Jr., and J. P. Sweeney. Development and validation of a transient event detector. *AMP Journal of Technology*, 3:69–74, 1993. [cited at page 32, 35]

S. B. Leeb, S. R. Shaw, and J. L. Kirtley Jr. Transient event detection in spectral envelope estimates for nonintrusive load monitoring. *IEEE Transactions on Power Delivery*, 10(3):1200–1210, 1995. [cited at page 32, 35]

J. Li, S. West, and G. Platt. Power decomposition based on SVM regression. In *Proceedings of International Conference on Modelling, Identification Control*, pages 1195–1199, 2012. [cited at page 30]

J. Liang, S. Ng, G. Kendall, and J. Cheng. Load signature study-part I: Basic concept, structure, and methodology. *IEEE Transactions on Power Delivery*, 25 (2):551–560, 2010a. [cited at page xiv, 29, 30, 36, 38, 39, 46, 51]

J. Liang, S. Ng, G. Kendall, and J. Cheng. Load signature study-part II: Disaggregation framework, simulation, and applications. *IEEE Transactions on Power Delivery*, 25(2):561–569, 2010b. [cited at page 36, 46]

N. Lopes. *Machine Learning for Adaptive Multi-Core Machines*. PhD thesis, University of Coimbra, 2013. [cited at page 45]

S. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):674–693, 1989. [cited at page 87]

S. Mallat. *A Wavelet Tour of Signal Processing: The Sparse Way*. Academic Press, 3rd edition, 2008. [cited at page 81, 82, 85, 89]

H. B. Mann and D. R. Whitney. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*, 18(1):50–60, 1947. [cited at page 139]

M. L. 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*, 41(13):1389–1403, 2000. [cited at page 23, 27, 28]

A. Marchiori, D. Hakkarinen, Q. Han, and L. Earle. Circuit-level load monitoring for household energy management. *IEEE Pervasive Computing*, 10(1):40–48, 2011. [cited at page 42]

H. S. Matthews, L. Soibelman, M. Berges, and E. Goldman. Automatically disaggregating the total electrical load in residential buildings: a profile of the required solution. In *Proceedings of International Workshop on Intelligent Computing in Engineering*, pages 381–389, 2008. [cited at page 22, 23, 37]

K. Murphy. *Machine Learning: A Probabilistic Approach*. The MIT Press, 2012. [cited at page 41, 42, 45]

H. Najmeddine, K. El Khamlichi Drissi, C. Pasquier, C. Faure, K. Kerroum, A. Diop, T. Jouannet, and M. Michou. State of art on load monitoring methods. In *Proceedings of IEEE International Conference on Power and Energy*, pages 1256–1258, 2008. [cited at page xiv, 29, 32]

Y. Nakano, H. Murata, K. Yoshimoto, K. Hidaka, M. Tadokoro, and K. Nagasaka. Non-intrusive electric appliances load monitoring system using harmonic pattern recognition - trial application to commercial building. In *Proceedings of International Conference on Energy Efficiency*, 2007. [cited at page 30]

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*, 24(1):51–64, 1996. [cited at page 19, 23, 32, 35, 49]

B. Olshausen and D. Field. Sparse coding of sensory inputs. *Current Opinion in Neurobiology*, 14(4):481–487, 2004. [cited at page 120]

B. A. Olshausen and D. J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583): 607–609, 1996. [cited at page 119]

P. Paatero. Least squares formulation of robust non-negative factor analysis. *Chemometrics and Intelligent Laboratory Systems*, 37(1):23–35, 1997a. [cited at page 128]

P. Paatero. A weighted non-negative least squares algorithm for three-way "PARAFAC" factor analysis. *Chemometrics and Intelligent Laboratory Systems*, 38(2):223–242, 1997b. [cited at page 128]

D. Palmer-Brown and M. Kang. Adfunn: An adaptive function neural network. In B. Ribeiro, R. Albrecht, A. Dobnikar, D. Pearson, and N. Steele, editors, *Adaptive and Natural Computing Algorithms*, pages 1–4. Springer Vienna, 2005. [cited at page 164]

O. Parson, S. Ghosh, M. Weal, and A. Rogers. Using hidden markov models for iterative non-intrusive appliance monitoring. In *Neural Information Processing*

Systems workshop on Machine Learning for Sustainability, 2011. [cited at page 45, III]

O. Parson, S. Ghosh, M. Weal, and A. Rogers. Non-intrusive load monitoring using prior models of general appliance types. In *Proceedings of Twenty-Sixth Conference on Artificial Intelligence*, 2012. [cited at page 52, 53]

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. In J. Krumm, G. D. Abowd, A. Seneviratne, and T. Strang, editors, *UbiComp 2007: Ubiquitous Computing*, volume 4717 of *Lecture Notes in Computer Science*, pages 271–288. Springer Berlin Heidelberg, 2007. [cited at page 5, 22, 23, 33, 34]

P. V. Pauca, F. Shahnaz, M. W. Berry, and R. J. Plemmons. Text Mining Using Non-Negative Matrix Factorizations. In M. W. Berry, U. Dayal, C. Kamath, and D. B. Skillicorn, editors, *Proceedings of SIAM International Conference on Data Mining*. SIAM, 2004. [cited at page 120]

H. Pihala. Non-intrusive appliance load monitoring system based on a modern kwh-meter. Technical report, VTT Technical Research Center of Finland, 1998. [cited at page 26]

Walt Disney Pictures. Pixar Animation Studios. Toy story. [Film: 81 min], 1995. [cited at page 159]

M. Plumbley. Conditions for nonnegative independent component analysis. *IEEE Signal Processing Letters*, 9(6):177–180, 2002. [cited at page 122]

M. Plumbley. Algorithms for nonnegative independent component analysis. *IEEE Transactions on Neural Networks*, 14(3):534–543, 2003. [cited at page 122]

M. Plumbley and E. Oja. A "nonnegative PCA" Algorithm For Independent Component Analysis. *IEEE Transactions on Neural Networks*, 15(1):66–76, 2004. [cited at page 122]

L. R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. In *Proceedings of the IEEE*, pages 257–286, 1989. [cited at page 45]

B. Ribeiro, C. Silva, A. Vieira, and J. Neves. Extracting discriminative features using non-negative matrix factorization in financial distress data. In M. Kolehmainen, P. Toivanen, and B. Beliczynski, editors, *Adaptive and Natural Computing Algorithms*, volume 5495 of *Lecture Notes in Computer Science*, pages 537–547. Springer Berlin Heidelberg, 2009. [cited at page 120]

O. Rioul and M. Vetterli. Wavelets and signal processing. *IEEE Signal Processing Magazine*, 8(4):14–38, 1991. [cited at page 81, 82, 85, 86, 87]

J. Rissanen. Modeling by shortest data description. *Automatica*, 14(5):465–471, 1978. [cited at page 94]

T. Rocha. *Similarity-based approaches for the analysis and prediction of physiological time series*. PhD thesis, University of Coimbra, 2012. [cited at page 82, 86, 88, 89]

A. Rowe, M. Berges, and R. Rajkumar. Contactless sensing of appliance state transitions through variations in electromagnetic fields. In *Proceedings of the ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, BuildSys '10, pages 19–24, 2010. [cited at page 39]

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 *Proceedings of IEEE Communications Society Conference on Sensor Mesh and Ad Hoc Communications and Networks*, pages 1–9, 2010. [cited at page 29, 45]

R. Schapire. The boosting approach to machine learning: An overview. In *Proceedings of MSRI Workshop on Nonlinear Estimation and Classification*, 2001. [cited at page 41]

M. N. Schmidt. *Single-channel source separation using non-negative matrix factorization*. PhD thesis, Technical University of Denmark, 2008. [cited at page 118, 119, 120]

M. N. Schmidt and M. Mørup. Nonnegative matrix factor 2-d deconvolution for blind single channel source separation. In *Proceedings of the International Conference on Independent Component Analysis and Blind Signal Separation*, pages 700–707, 2006. [cited at page 117]

M. N. Schmidt and R. K. Olsson. Single-channel speech separation using sparse non-negative matrix factorization. In *Proceedings of the International Conference on Spoken Language Processing*, 2006. [cited at page 120]

A. Schoofs, A. Guerrieri, D. T. Delaney, G. M. P. O''Hare, and A. G. Ruzzelli. Annot: Automated electricity data annotation using wireless sensor networks. In *Proceddings of IEEE Communications Society Conference on Sensor Mesh and Ad Hoc Communications and Networks*, pages 1–9, 2010. [cited at page 39]

C. E. Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, 1949. [cited at page 22, 31]

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*, 57(7):1445–1454, 2008. [cited at page 33]

J. Shawe-Taylor and N. Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004. [cited at page 94]

Jonathon Shlens. A tutorial on principal component analysis. Technical report, Systems Neurobiology Laboratory, Salk Institute for Biological Studies, 2009. [cited at page 82]

C. Silva and B. Ribeiro. *Inductive Inference for Large Scale Text Classification: Kernel Approaches and Techniques*. Springer Publishing Company, 2009. [cited at page 40, 42, 43, 44, 46, 63, 69]

M. Sokolova and G. Lapalme. A systematic analysis of performance measures for classification tasks. *Information Processing and Management*, 45(4):427–437, 2009. [cited at page 50]

D. Srinivasan, W. S. Ng, and A. C. Liew. Neural-network-based signature recognition for harmonic source identification. *IEEE Transactions on Power Delivery*, 21(1):398–405, 2006. [cited at page 30, 44, 45, 50, 63]

F. Sultanem. Using appliance signatures for monitoring residential loads at meter panel level. *IEEE Transactions on Power Delivery*, 6(4):1380–1385, 1991. [cited at page 14, 26, 34, 35]

V. Sundramoorthy, G. Cooper, N. Linge, and Qi Liu. Domesticating energy-monitoring systems: Challenges and design concerns. *IEEE Pervasive Computing*, 10(1):20–27, 2011. [cited at page 3, 14, 163]

K. Suzuki, S. Inagaki, T. Suzuki, H. Nakamura, and K. Ito. Nonintrusive appliance load monitoring based on integer programming. In *Proceedings of SICE Annual Conference*, pages 2742–2747, 2008. [cited at page 29, 39]

C. Taswell. The what, how, and why of wavelet shrinkage denoising. *Computing in Science Engineering*, 2(3):12–19, 2000. [cited at page 89]

A. R. Teixeira. *Subspace Learning Techniques and Applications*. PhD thesis, Department of Electronics, Telecommunications and Informatics, University of Aveiro, 2011. [cited at page 82, 84, 92, 93]

A. R. Teixeira, A. M. Tome, E. W. Lang, P. Gruber, and A. Martins da Silva. On the use of clustering and local singular spectrum analysis to remove ocular artifacts from electroencephalograms. In *Proceedings of IEEE International Join Conference on Neural Networks*, volume 4, pages 2514–2519, 2005. [cited at page 80, 82, 94]

K. H. Ting, Mark Lucente, G. S. K. Fung, W. K. Lee, and S. Y. R. Hui. A taxonomy of load signatures for single-phase electric appliances. In *Proceedings of the IEEE Power Electronics Specialist Conference*, 2005. [cited at page 23, 36]

L. R. Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3):279–311, 1966. [cited at page 125]

M. E. Wall, A. Rechtsteiner, and L. M. Rocha. Singular value decomposition and principal component analysis. *A Practical Approach to Microarray Data Analysis*, pages 91–109, 2003. [cited at page 92]

Z. Wang and G. Zheng. Residential appliances identification and monitoring by a nonintrusive method. *IEEE Transactions on Smart Grid*, 3(1):80–92, 2012. [cited at page xiv, 25, 36, 37]

Watteco SAS. Low power networking solutions for automation and smart grid applications. `http://www.watteco.com/index.php`, 2013. [Online; accessed 13-September-2013]. [cited at page 163]

G. I. Webb. Multiboosting: A technique for combining boosting and wagging. *Machine Learning*, 40(2):159–196, 2000. [cited at page 41]

M. Weeks. *Digital Signal Processing Using MATLAB & Wavelets*. Infinity Science Press, 2007. [cited at page 80, 81, 85, 86, 99, 100]

E. W. Weisstein. Mathworld - a wolfram web resource. `http://mathworld.wolfram.com/MatrixDecomposition.html`, 2013. [Online; accessed 01-September-2013]. [cited at page 121]

W. Wichakool, Al-T. Avestruz, R. W. Cox, and S. B. Leeb. Modeling and estimating current harmonics of variable electronic loads. *IEEE Transactions on Power Electronics*, 24(12):2803–2811, 2009. [cited at page 33]

F. Wilcoxon. Individual Comparisons by Ranking Methods. *Biometrics Bulletin*, 1(6):80–83, 1945. [cited at page 139]

Y. Yi-xin, L. Peng, and Z. Chun-liu. Non-intrusive method for on-line power load decomposition. In *Proceedings of China International Conference on Electricity Distribution*, pages 1–8, 2008. [cited at page 30]

M. Zeifman. Disaggregation of home energy display data using probabilistic approach. *IEEE Transactions on Consumer Electronics*, 58(1):23–31, 2012. [cited at page 37, 51]

M. Zeifman and K. Roth. Nonintrusive appliance load monitoring: Review and outlook. *IEEE Transactions on Consumer Electronics*, 57(1):76–84, 2011. [cited at page 4, 5, 14, 22, 25, 30, 31, 32, 33, 34, 36, 38, 49, 51, 53]

T. Zia, D. Bruckner, and A. Zaidi. A hidden markov model based procedure for identifying household electric loads. In *Proceedings of Annual Conference on IEEE Industrial Electronics Society*, pages 3218–3223, 2011. [cited at page 45]

A. Zoha, A. Gluhak, M. A. Imran, and S. Rajasegarar. Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey. *Sensors*, 12(12):16838–16866, 2012. [cited at page 4, 5, 19, 23, 24, 38, 39, 40, 46, 49, 53, 57, III]