UNIVERSIDAD DE ALCALÁ

ESCUELA POLITÉCNICA SUPERIOR

Programa de Doctorado en Electrónica
Sistemas Electrónicos Avanzados. Sistemas Inteligentes

Doctoral Thesis

Non-Intrusive Load Monitoring techniques for Activity of Daily Living recognition

José Manuel Alcalá Orzáez

2016

UNIVERSIDAD DE ALCALÁ

ESCUELA POLITÉCNICA SUPERIOR

Programa de Doctorado en Electrónica
Sistemas Electrónicos Avanzados. Sistemas Inteligentes

Non-Intrusive Load Monitoring techniques for Activity of Daily Living recognition

Author

José Manuel Alcalá Orzáez

Supervisors

Dr. Jesús Ureña Ureña

Dr. Álvaro Hernández Alonso

2016

Doctoral Thesis

Resumen

Entre los grandes desafíos de nuestra era, los que despiertan mayor interés, se puede destacar aquellos que son consecuencia directa del desarrollo de nuestra sociedad y de su sobrepoblación: el cambio climático, falta de recursos energéticos y el envejecimiento de la población.

Respecto a los dos primeros, la demanda de energía es cada vez mayor y, en pocos años, ésta será superior a la producida. Por tanto, no sólo es necesario la sustitución de combustibles fósiles por energías limpias y renovables, sino también su consumo y distribución eficiente. De esta idea nace en la primera década de este siglo el concepto de Smart Grids. Numerosos países comienzan a realizar despliegues masivos de contadores inteligentes ("Smart Meters"), lo que despierta el interés de un gran número de investigadores que comienzan a desarrollar nuevas técnicas para predecir la demanda y promover la eficiencia. Entre ellas destacan los sistemas NILM (Non-Intrusive Load Monitoring) por el gran interés que han despertado. Éstos tratan de predecir qué dispositivos hay conectados y qué están consumiendo a partir de un único sensor: el contador inteligente. Sin embargo, el interés que despierta NILM entre investigadores y compañías eléctricas no es correspondido por los consumidores finales; quienes no encuentran beneficio al ahorro energético en sus facturas, en su mayor parte estancadas por tasas y tarifas fijas.

Por otra parte, los grandes avances en la medicina moderna han permitido que nuestra esperanza de vida aumente considerablemente. No obstante, esta longevidad, junto con la baja fertilidad en los países desarrollados, tiene un efecto secundario: el envejecimiento de la población. Unos de los grandes avances es la incorporación de la tecnología en la vida cotidiana, lo que ayuda a los más mayores a llevar una vida independiente. El despliegue de una red de sensores dentro de la vivienda permite su monitorización y asistencia en las tareas cotidianas. Sin embargo, son intrusivos, no escalables y, en algunas ocasiones, de

alto coste, por lo que no están preparados para hacer frente al incremento de la demanda de esta comunidad.

Esta tesis doctoral nace de la motivación de afrontar estos problemas y tiene dos objetivos principales: lograr un modelo de monitorización sostenible para personas mayores y, a su vez, dar un valor añadido a los sistemas NILM que despierte el interés del usuario final. Con este objetivo, se presentan nuevas técnicas de monitorización basadas en NILM, aunando lo mejor de ambos campos. Esto supone un ahorro considerable de recursos en la monitorización, ya que únicamente se necesita un sensor: el contador inteligente; lo cual da escalabilidad a estos sistemas.

Las contribuciones de esta tesis se dividen en dos bloques principales. En el primero se proponen nuevas técnicas NILM optimizadas para la detección de la actividad humana. Así, se desarrolla una propuesta basada en detección de eventos (conexiones de dispositivos) en tiempo real y su clasificación a un dispositivo. Con el objetivo de que pueda integrarse en contadores inteligentes. Cabe destacar que el clasificador se basa en modelos generalizados de dispositivos y no necesita conocimiento específico de la vivienda.

El segundo bloque presenta tres nuevas técnicas de monitorización de personas mayores basadas en NILM. El objetivo es proporcionar una monitorización básica pero eficiente y altamente escalable, ahorrando en recursos. Los procesos Cox, log Gaussian Cox Processes (LGCP), monitorizan un único dispositivo si la rutina está estrechamente ligada a este. Así, se propone un sistema de alarmas si se detectan cambios en el comportamiento. LGCP tiene la ventaja de poder modelar periodicidades e incertidumbres propias del comportamiento humano. Cuando la rutina no depende de un único dispositivo, se proponen dos técnicas: una basada en gaussianas mixtas, Gaussian Mixture Models (GMM); y la otra basada en la Teoría de la Evidencia de Dempster-Shafer (DST). Ambas monitorizan y detectan deterioros en la actividad, causados por enfermedades como la demencia y el alzhéimer. Únicamente DST usa incertidumbres que simulan mejor el comportamiento humano y, por tanto, permite alarmas en caso de un repentino desvío.

Finalmente, todas las propuestas han sido validadas mediante la evaluación de métricas y la obtención de resultados experimentales. Para ello, se han usado medidas de escenarios reales que han sido recopiladas en bases de datos. Los resultados obtenidos han sido satisfactorios, demostrando que este tipo de monitorización es posible y muy beneficioso para nuestra sociedad. Además, se ha dado a lugar nuevas propuestas que serán desarrolladas en el futuro.

Abstract

Among the great challenges of this era, those that raise the most interest, it can be highlighted those that are direct consequence of the development of our society and its overpopulation: the climate change, the lack of energy resources and the ageing population.

With regard the two first challenges, the energy demand is dramatically increasing and it is envisaged that, in a matter of a few years, it will overcome the produced one. It is, hence, not only necessary to replace fossil fuels with clean and renewable energy, but also to reach an efficient energy consumption and distribution. Thus, the Smart Grid concept emerges from this idea during the first decade of this century and, since then, many countries have been largely rolling out Smart Meters to achieve this milestone. This fact is bringing in the interest of a large number of researchers who are devising new techniques to predict the demand and to promote efficiency. NILM (Non-Intrusive Load Monitoring) is among them and it is raising a great interest. It tries to foresee what devices are plugged and how much they are consuming from a single sensor: the Smart Meter. However, the motivation in NILM between researchers and electricity companies is not corresponded by end-consumers, who find no benefit on their bills from energy savings, mostly leveraged by taxes.

Moreover, great achievements in the modern medicine is leading to a significant life expectancy incresing. Nonetheless, this longevity, along with low fertility in developed countries, entails a side effect: the ageing population. One of the greatest reached milestones is the technology integration in the daily life, which supports the elderlies to lead an independent life. The deployment of a sensor network within the household allows monitoring and assistance in everyday tasks. Although, they are not prepared to cope with the increasing demand of this community as they are intrusive, not scalable and, in cases, expensive.

This doctoral thesis is born from the motivation to address these issues. It has two main objectives: the achievement of a sustainable monitoring model for elderlies and, in turn, to give an added value to NILM that awakes the interest of the end-consumer. To this end, novel monitoring techniques based on NILM are presented, combining the best of both domanins. This represents considerable savings in monitoring resources since it only uses a sensor: the Smart Meter. Hence, scalability is at reach.

The contributions of this Thesis are organised in two blocks. The first introduces new event-based NILM algorithm optimised for human activity detection. The proposal detects events in real time, such as appliance swithing-on; and it classifies them into appliance types, all aimed to be integrated in a Smart Meter. It is worth noting that the classifier is based on general models of appliance types and, hence, it does not need specific knowledge about household. It is an unsupervised training method and, consequently, it can be trained offline and loaded into Smart Meter memories at once for all households.

Concerning the second block, three novel activity monitoring techniques for elderlies are presented. They are based on NILM and they use the appliance usage patterns to infer the human activity. The aim is to provide basic but efficient and highly scalable monitoring, saving resources during the deployment. Thus, the log Gaussian Cox Processes (LGCP) approach monitors a single appliance, in case the daily routine is strongly correlated to it. It implements an alarm system whether a rapid pattern deviation is detected. LGCP has the advantage of being able to model periodicities and uncertainties inherent to the human behavior. Whenever the routine does not depend on a single appliance, two techniques are proposed: the Gaussian Mixture Model (GMM) and the Dempster-Shafer Theory (DST) approach. Both monitor and detect deteriorations in the activity, possibly caused by diseases such as dementia and Alzheimer's disease. Nevertheless, only the DST approach simulate uncertainties, due to the human's free will, and, therefore, it allows to trigger alarms in case of a sudden pattern deviation in the behavior.

All proposals have been validated by the evaluation of metrics and the obtaining of experimental results. To this end, real household measurements have been used, which have been collected in datasets. Satifactory results have been obtained, proving that this type of monitoring is feasible and very beneficial to our society. In addition, this thesis has yielded new proposals to be addressed in the future.

A mis padres y hermano, y en especial, a Paloma y al pequeño ratoncín que está por llegar.

You do not really understand
something unless you can explain it to
your grandmother

Ernest Rutherford

Agradecimientos

En primer lugar, me gustaría agradecer a mis directores de tesis Jesús Ureña y Álvaro Hernández por su inestimable apoyo a lo largo de toda la tesis. Desde mis comienzos como becario, Jesús me ha dado la oportunidad de realizar este doctorado y ha sabido guiarme con maestría a la vez que darme la libertad para explorar nuevos campos. Álvaro ha sido clave en los momentos difíciles, donde me encontraba atascado y donde gracias a sus consejos y ánimos he conseguido avanzar. Sin lugar a dudas, he tenido suerte, pues no se me ocurren mejores directores para formarme como investigador.

A mis padres Sebastián y Guadalupe quienes gracias a su apoyo y a su cariño incondicional me encuentro aquí. Podría decir sin equivocarme que la mitad del mérito es suyo. Me gustaría parafrasear la siguiente expresión: "con fuerza de voluntad y constancia todo se consigue".

A mi hermano Jorge quien me ha visto crecer y ha influido tan positivamente en mi vida dándome tan buenos momentos.

A Paloma, simplemente decirte: "bendito el lugar y el motivo de estar ahí, bendita la coincidencia..."

To Victor Cionca who introduces me to Python and the Linux world during my first international research stay in Cork.

To Alex Rogers and Oliver Parsons who collaborate to devise the LGCP algorithm presented in Chapter 5 and provide an excellent guidance during my international research stay in Southampton. I learned all I know of Pandas from Oli. Remember: *"Do the hat!"*

A mis compañeros y al profesorado de la Universidad de Alcalá. En especial a Juan Jesús García por introducirme en la teoría de Dempster-Shafer y a David Gualda por ayudarme

siempre con el papeleo, por las numerosas veces que me ha recogido a mi llegada del AVE y por todos esos buenos ratos en el fondo 31 junto con Pablo. También a Jorge por instruirme en el uso del cluster y otras herramientas muy útiles. Al resto de mis compañeros: siempre recordaré las míticas cenas de los viernes.

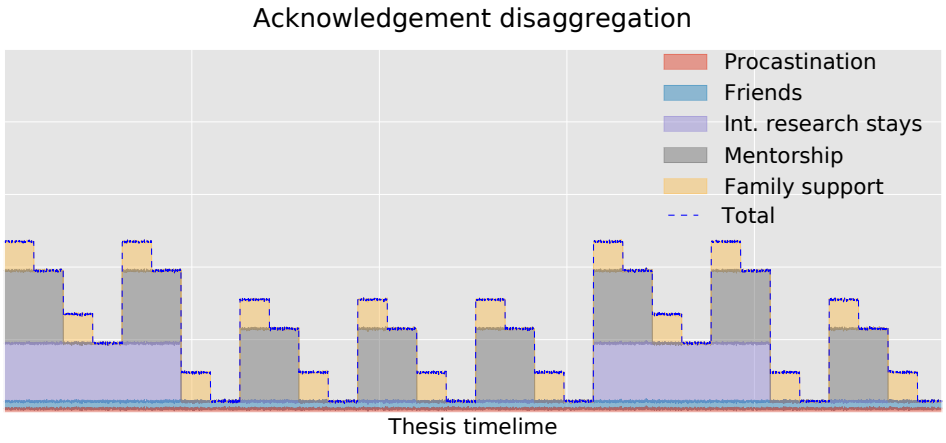
A Lucía Palacios por ayudarme con la bibliografía, por ser una gran amiga durante todos estos años y por saber sacarme siempre una sonrisa.

A mi abuela y a mi tío Fali, quien me ayudó con el montaje del sistema de monitorización. Aunque no he conseguido explicar la tesis a mi abuela, al menos he conseguido convencerla de que dejar "la lucecita" encendida durante la noche no va a quemar la casa.

A Chelu Martin por introducirme en la Universidad de Alcalá y por ser el mejor amigo y compañero de piso que se puede tener. También a Lubna por esos buenos ratos bebiendo un Verdejo.

Finalmente, aunque no menos importante, a todos aquellos amigos que me han soportado y me han dado tanto buenos momentos. En granada en las fiestas de pre-nochevieja con Lucía, Carlos, Macius, Marina y demás "criaturas". En los reencuentros con Javi, Juan, David, Víctor y demás "colegas". En Madrid con Lizette, Víctor y Mariajosé. En el extranjero con Bruno, Lorna, Levi, Eoin, Laura y Cecilia.

Para concluir, me gustaría incluir la siguiente figura que dice mucho de la temática de esta tesis y resume estos duros pero gratificantes años de trabajo.



Contents

List of Figures

List of Tables

List of Acronyms

UK-DALE	The UK Domestic Appliance-Level Electricity dataset
NILMTK	Open source NILM ToolKit in python
NILM	Non Intrusive Load Monitoring
SIA	Semiconductor Industry Association
SG	Smart Grids
SM	Smart Meter
EV	Electric Vehicle
HVAC	Heating, Ventilation and Air Conditioning
EDR	Elderly Dependency Ratio
CO	Combinatorial Optimisation
HMM	Hidden Markov Models
DNN	Deep Neural Network
UK	United Kingdom
LGCP	Log Gaussian Cox Processes
GMM	Gaussian Mixture Model
DST	Dempster-Shafer Theory
PQD	Active, Reactive and Distortion power trajectories
NIALM	Non Intrusive Appliance Load Monitoring
PLC	Power Line Communication
HAN	Home Area Network
NALM	Nonintrusive Appliance Load Monitoring
DR	Demand Response

LF	Low Frequency
HF	High Frequency
EMI	ElectroMagnetic Interference
P	Active Power
D	Distortion Power
HAR	Harmonics
SMPS	Switch Mode Power Supply
ILM	Intrusive Load Monitoring
DDSC	Disaggregation via Discriminative Sparse Coding
FHMM	Factorial Hidden Markov Model
CFHSMM	Conditional Factorial Hidden Semi Markov Model
dHMM	Difference Hidden Markov Model
ANN	Artificial Neural Network
LS	Generalised Likelihood Ratio
GLR	Load Signature
FFT	Fast Fourier Transform
CL	Composite Load
Q	Reactive Power
WS	Wave-Shaped feature
CW	Current Waveform
IAW	Instantaneous Admittance Waveform
IPW	Instantaneous Power Waveform
PCA	Principal Component Analysis
MS-NILM	Manual Setup NILM
AS-NILM	Automatic Setup NILM
NN	Neural Network
MLP	MultiLayer Perceptron
RBF	Radial Basis Function
SVM	Support Vector Machine

LR	Least Residue
CDM	Committee Decision Mechanism
MLE	Maximum Likelihood Estimator
LUR	Least Unified Residue
MCO	Most Common Occurrence
PRIME	PoweRline Intelligent Metering Evolution
NILM-eval	NILM evaluation framework in Matlab
AAL	Ambient Assisted Living
AmI	Ambient Intelligence
ADL	Activities of Daily Living
BSN	Body Sensor Network
MEMS	Microelectromechanical Systems
ADLs	Activities of Daily Living
IADL	Instrumental Activities of Daily Living
CS	Guttman Coefficient
COPD	Chronic Obstructive Pulmonary Disease
AR	Activity Recognition
HHSMMs	Hierarchical Hidden semi-Markov Models
NB	Naive Bayesian
LHMMs	Layered Hidden Markov Models
LDA	Latent Dirichlet Allocation
rms	Root mean square value
S	Apparent Power
PLAID	The Plug-Level Appliance Identification Dataset
REDD	The Reference Energy Disaggregation Data Set
BLUED	Building-Level fUlly-labeled dataset for Electricity Disaggregation
TPR	True Positive Rate

FPR	False Positive Rate
TPP	True Positive Percentage
ROC	Receiver Operating Characteristic
AUC	Area Under Curve
TP	True Positives
FN	False Negatives
FP	False Positives
TN	True Negatives
FPP	False Positive Percentage
P	Number of positives or events
CRC	Cumulative Ranking Curve
kNN	k-Nearest-Neighbors
GNB	Gaussian Naive Bayes
LGC	Logistic Regression Classifier
DTree	Decision Tree
RForest	Random Forest
GPy	Gaussian Process framework in Python
HES	Household Electricity Survey
BBA	Basic Belief Assignment
MLE	Maximum Likelihood Estimator

Nomenclature

f_s	Sampling frequency.
ΔP	Delta values of the active power: difference between two consecutives samples.
v_n	Vote value for item n in the knapsack problem algorithm.
w_n	Weight for item n in the knapsack problem algorithm.
$\hat{\boldsymbol{x}}$	Estimated values for vector $\boldsymbol{x}$ in a HMM algorithm.
$x_t^{(n)}$	Unobserved variable at instant t for appliance or item n.
$y_t^{(n)}$	Observed variable at instant t for appliance or item n.
$\overline{y}_t$	Sum of observed variables along appliances or items N at instant t.
V_{rms}	Root mean square value of voltage.
V_{norm}	Regularised voltage in mains (e.g. $230V_{rms}$ in Europe).
$v(t)$	Voltage values over time t.
$i(t)$	Current values over time t.
$P_{norm}(t)$	Normalized active power over t.
$IAW(t)$	Instantaneous Admittance Waveform over t.

$IPW(t)$	Instantaneous Power Waveform over t.
$\boldsymbol{v}_s$	Voltage value vector for cycle s in the event detector.
$\boldsymbol{i}_s$	Current value vector for cycle s in the event detector.
$v_{rms,s}$	Voltage root mean square value in cycle s.
$\boldsymbol{v}_{rms}$	Voltage root mean square value vector over all cycles S.
$v_{s,j}$	The *j-th* voltage value in the vector-cycle $\boldsymbol{v}_s$.
$\bar{i}_{rms,s}$	The normalised current root mean square value in cycle s.
$\boldsymbol{\nabla \bar{i}_{rms}}$	The derived vector of consecutive $\bar{i}_{rms,s}$ values over all cycles S.
$\boldsymbol{\rho}$	Output of the change detection block in the event detector algorithm.
W_d	Peak detector parameter of minimum distance between peaks.
θ_{th}	Peak detector parameter of minimum peak amplitude.
$\boldsymbol{\tau}$	Timestamp vector of event occurrences.
$\boldsymbol{\gamma}$	Set of w aggregated current vector-cycles centered at the event e.
$\boldsymbol{I}_e$	The de-energised current signal centered at event e and composed by w vector-cycles.
$I_{rms,1}$	The root mean square value of the current first harmonic.
φ	The displacement phase between voltage v and current i.
$[\boldsymbol{PQD}]^{\boldsymbol{a}}_{\boldsymbol{m}}$	Matrix of evaluated power P, Q and D for a certain sample m of the appliance type a.
$\boldsymbol{\tau}^a_m$	Matrix of evaluated power P, Q and D for a certain sample m of the appliance type a, rearranged in 1 column vector.
$\boldsymbol{\Psi}^a$	Averaged vector for the set of training M of the appliance type a in the PQD-PCA algorithm.

$\mathbf{\Phi}_m^a$	The variance vector of the training sample $\boldsymbol{\tau}_m^a$ after subtracting the average vector $\mathbf{\Psi}^a$.
$\boldsymbol{S}_T^a$	The Scatter matrix along all training samples $\boldsymbol{\tau}_m^a$ in the set M.
$\mathbf{\Omega}_m^a$	The feature vector of the training sample $\boldsymbol{\tau}_m^a$.
$\boldsymbol{U}$	A transformation matrix.
$\boldsymbol{U}_{opt}^a$	The optimised transformation matrix of l eigenvalues for appliance type a.
$\boldsymbol{u}_k$	The *k-th* eigenvalue of an optimised transformation matrix
$\hat{\boldsymbol{\tau}_e}$	Estimation of the arrival test sample $\boldsymbol{\tau}_e$.
$\widehat{\mathbf{\Phi}}_e^a$	Estimation of the variance vector of the arrival test sample $\boldsymbol{\tau}_e$ after substracting the average vector $\mathbf{\Psi}^a$ of a appliance type a
ε_e^a	Error reconstruction for arrival test sample e evaluated by classifier appliance type a.
γ_a	Maximum error reconstruction in the classifier appliance type a.
C	Appliance class that labels the arrival test sample e in the PCA-PQD algorithm.
ψ_{opt}	The optimal binary detector.
Ψ	The set of all binary detectors in the event detector performance.
P_i	Active power feature for sample i.
S_i	Apparent power feature for sample i.
w_{recall}	Weight of metric *recall*.
$w_{precision}$	Weitgh of metric *precision*.
β	The ratio between weights w_{recall} and $w_{precision}$.
F_β	The weighed mean between *recall* and *precision*.

F_1	The harmonic mean between *recall* and *precision*.
$\lambda(x)$	Intensity function over the temporal points.
$m(x)$	Mean of a Gaussian Process over the temporal points in the LGCP algorithm.
x'	Predicted temporal points in a temporal point space X.
$K(x, x')$	Kernel function of a Gaussian Proccess over the temporal points x and given the predicted points x'.
$\mathrm{GP}(m(x), K(x, x'))$	Gaussian Proccess given a mean and a kernel function.
σ	Variance of a Kernel function.
ℓ	The lenght-scale window of a Kernel funcion within where neighbors have effect on the regression.
ρ	The period of a Kernel function that models the regression in a Gaussian Proccess.
$p(a_{0:t})$	The cumulative probability that an appliance a was used at least once by time of day t.
E_j	The *j-th* event of m event occurrences within a day for an appliance.
$P(E_j)$	The probability of the event E_j occuring at a certain time T_J.
A_i	The set of all events m of the appliance i occurring within the day.
$P(A_i)$	The probablity of occurring the set of events A_i within a day.
h_x	The hypothesis x.
H_x	The subset of hyphotesis x.
Ω	The universe of hpythesis.
2^Ω	The power set of hypothesis in the Dempster-Shafer Theory.

$m_x(h_x, T_i)$	The mass function for appliance x in the hypothesis h_x at the observation window T_i in the DST algorithm.
$m_{1,2}(A)$	The resulting mass function for appliance 1 and 2 in hypothesis A in the DST algorithm.
$bel(A)$	The belief in hypothesis A.
$pl(A)$	The plausibility in hypothesis A.
C_0	The used uncertainty constant whenever a event occurs in the DST algorithm.
C_1	The used uncertainty constant in the absence of events in the DST algorithm.
$m_{T_i}(h_x)$	In the DST algorithm, the total mass function for hypothesis h_x after merging the all appliance mass functions in the observed window T_i.

Chapter 1

Introduction

1.1 Overview

We face harsh global challenges in the forthcoming years that are much related to the thriving of our species: climate change, lack of energy resources and ageing of the population. This dissertation encompasses our tiny contribution to cope with these challenges. These three concepts may be well studied separately in several thesis as they are not directly bonded, specially the last one related to the two others: "ageing of the population". Although, we have found a synergy where improvements in one field may help to strengthen the weakest point of the others. Thus, this thesis devises technical solutions for monitoring the health of elderlies through the employ NILM (Non-Intrusive Load Monitoring) techniques, which are typically used to save energy resources, in relation to climate change.

In an overpopulated World like ours, the potential danger of resource shortages is real. Concerning energy resources, this danger is doubled up as we tend to increase the energy consumption per capita. An increasing power demand means to increase the energy generation, whose main source is currently fossils fuels; and consequently, this leads to increase the CO2 emissions. Thus, the concern about the climate change is pushing forward the research of new renewable and clean energy sources. Nevertheless, this will not be enough as the power demand will shortly be greater than the world's power generation and power shortages will eventually happen. Could you imagine the consequences of that in a digitalised world like ours? We may sound alarming and apocalyptic, nonetheless, there is not lack of studies that foresee this scenario. For instance, a report by the

Semiconductor Industry Association (SIA) predicts that, following the actual pattern of production and consumption, the total energy employed for computing (i.e. the total amount of energy per bit of all computers around the World) will exceed the World's energy production by 2040 [Beall, 2016], (see Figure 1.1). Of course, it is not feasible that the whole energy production will be dedicated to computing, therefore, this shortage may be far before 2040, whether actions are not to be taken.

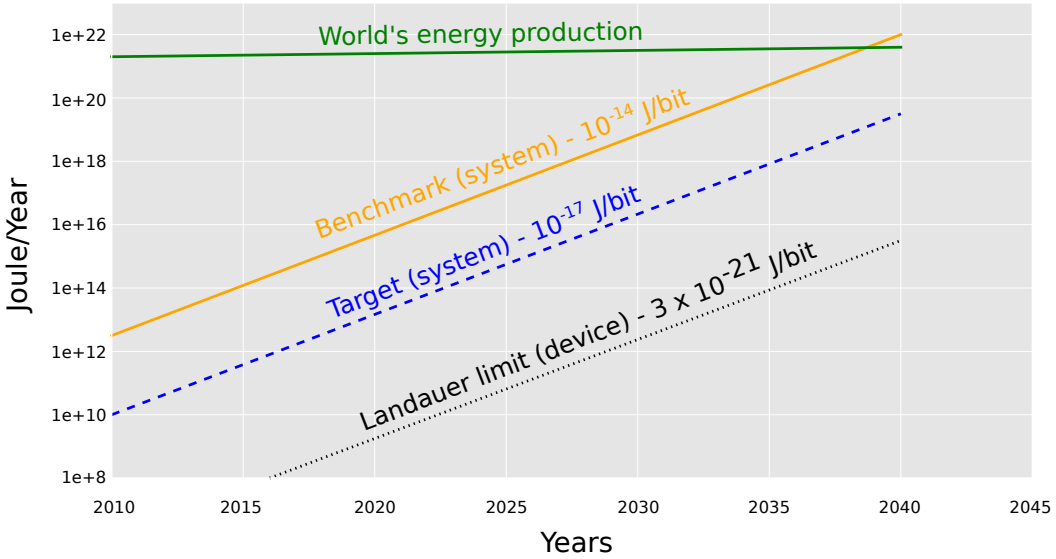


Figure 1.1: Total energy of computing versus World's energy production, adapted from [Beall, 2016].

From SIA, a realistic action is to improve the energy efficiency on transistors to reduce the amount of energy per bit; instead of reducing the World's computation, which would affect our development as society. Analogically, we can translate the energy efficient problem to our whole society: we must be energy efficient in industry and residential and commercial buildings if we are to reach energy sustainability. Thus, the Smart Grid concept (SG) is born with the major goal of achieving the energy efficiency of our society. It changed the concept of a centralised distribution grid to a decentralised one to help the inclusion of small heterogeneous renewable sources (e.g. solar panels in households). Furthermore, it enables a better and a bidirectional communication (i.e. from the central station to the household and back) thanks to the use of Smart Meters that help gain insights and to anticipate great power demands whenever the electric vehicles (EV) are charged. The power grid has seldom changed since its inception and many appliances are

highly inefficient, therefore, one might think it may be easier to achieve energy efficiency than in transistors. In fact, buildings are highly inefficient and its thermal isolation can be far improved: they count for 41% of the total power demand and more than 63% is due to Heating, Ventilation and Air Conditioning (HVAC) systems. However, there is another factor that highly influences on the energy consumption: the pattern consumption of inhabitants. Therefore, we also need to be educated in healthy energy efficiency habits.

Thanks to the breakthrough of the Smart Meters, a new research domain has born and it is gathering the interest from researches: Non-Intrusive Load Monitoring (NILM). This is the process of itemising the total energy consumption of a household, measured from the SM, into the individual consumption of all appliances connected to the mains. We will extensively cover this in the Section 2.1, nevertheless, we would like to anticipate that NILM pursues energy efficiency through the "energy awareness" and it has the major advantage of being non-intrusive. Thus, it may use the itemised energy information to give feedback to the inhabitants so they can consequently take actions to reduce their consumption. At large scale, NILM may contribute much to energy efficiency spreading healthy energy consumption habits within the population, however, it might be of less interest for individuals: changing their routines is hard and this effort might not be reflected on their annual bill (usually leveraged by taxes). Section 2.1.4 will give more insights about this, also other thesis dissertations address this issue [Kelly, 2016]; however, we would like to highlight here that the lack of motivation from inhabitants is preventing NILM from a widespread adoption that would benefit to our society in terms of energy efficiency.

On the other hand, current advances on the modern medicine have made possible to us to have a longer and a quality life. Specially thanks to the development of antibiotics and vaccines, we have overcome many epidemics diseases (e.g. influenza, bubonic plague, smallpox and cholera) that decimated the world population; and now, the life expectancy has doubled up. Nevertheless, this prosperity is also reflected into profound demographic changes that arise new challenges. Likewise, in our modern world, the elderly dependency ratio (EDR), which is the ratio of the number of old age people (older than 65) to the number of productive people (between 20 to 64); is rapidly increasing since 2013 (see Figure 1.2), where the ratio is multiplied by 100. An ageing of the population inevitably entails consequences: this population is prone to suffer from diseases (e.g. alzheimer and dementia) and needs more medical and care assistance. If we are to maintain the wellness within the elderly community, more investment in health care is needed.

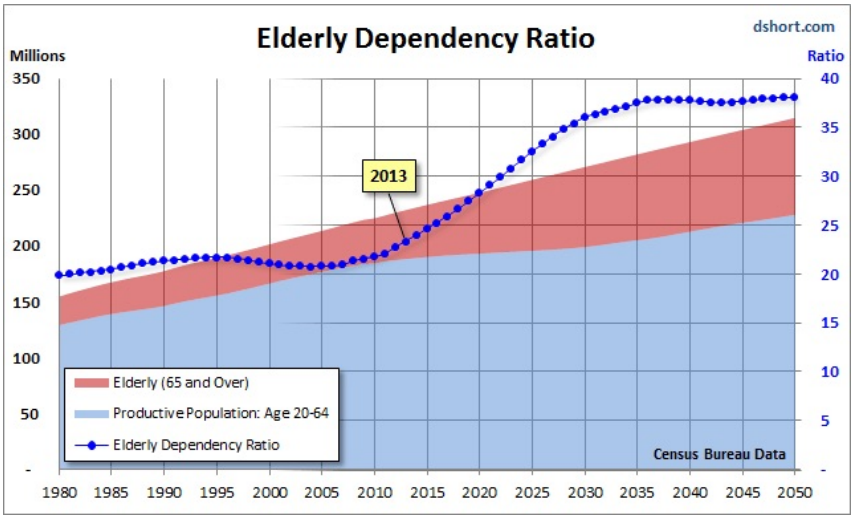


Figure 1.2: Elderly dependency ratio by 2050 in the United States [Short, 2013].

Although they are more dependant in terms of assistance, they are eager to maintain their independent live as long as they can, living by themselves in their own houses. Hence, there has been a blooming of health care tele-monitoring systems (i.e. remotely monitoring their activity) that try to meet their requirements. These health care monitoring systems have been much welcomed by this community, as well as their relatives, who feels more secured knowing that they are being monitored; however, they often need to adapt their surroundings and to deploy sensors along the house. The deployment and maintenance of these systems are usually expensive and not feasible and, hence, this raises a sustainability problem where the majority of this increasing community has no access to the basis health care tele-monitoring systems (this is more deeply discussed in Section 2.2.1).

In summary, in a near future (i.e. by 2050) we might experiment shortages in the energy distribution system and in the healthcare system, caused by an overpopulated and thriving world, that could severely affect the development of our society. Many studies have addressed these issues separately and solutions are being devised, but few have seen how these two worlds can be combined. How solutions for one problem can be applied to the other. Thus, healthcare monitoring systems may be a solution to avoid a collapse into the healthcare system, however, they are out of reach for the majority of the elderly community. If NILM approaches are to be applied to healthcare monitoring systems, the latter could much benefit from the low cost and non-intrusiveness to enable massive deployments. In turn, healthcare monitoring systems can give an added value to NILM to motivate individuals to adopt these approaches.

This thesis dissertation aims to contribute to achieve energy efficiency and sustainability in healthcare systems. Our goal is to open new ways to monitor elderlies using NILM techniques so we can use the best of both world to achieve massive deployments and to bring both technical and social values to this thesis. This may be summarised on the abstract diagram depicted in Figure 1.3. Although, our contribution might be tiny into a global scale, we hope that more studies will follow us and, together, we make huge changes.

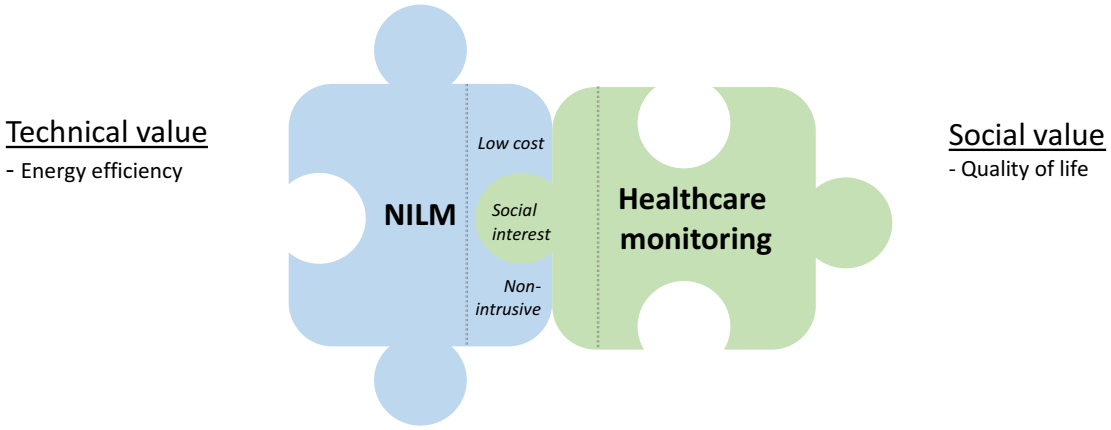


Figure 1.3: Abstract diagram of the synergy pursued on this thesis.

In the remaining of this Chapter, we give the framework into which this thesis has been developed (i.e. fundings and research projects). Then, the thesis outline gives a summary of the forthcoming chapters.

1.2 Thesis framework

The work developed in this thesis has been supported by the Spanish Ministry of Economy and Competitiveness and it has been developed under the following research projects:

- ENERGOS (UAH-24/2010),

- LORIS (ref. TIN2012- 38080-C04-01),
- DISSECT-SOC(ref. TEC2012- 38058-C03-03),
- TARSIUS (ref. TIN2015-71564-C4-1-R),
- SOC-PLC (ref. TEC2015-64835-C3-2-R MINECO/FEDER).

Likewise, the LORIS project seeks for continuous tracking in large environments by using ultrasound and radio frequency as well as the integration of these Local Positioning Systems with other networks such as the Smart Grids and the Body Sensor Networks. Moreover, the TARSIUS project pursues the construction of ambient intelligence-based spaces where it is essential to detect the presence of agents (i.e. people and autonomous vehicles) and to provide local based services (LBS). Thus, the project's objectives that influence more in this thesis deal with the exploration of new non-intrusive approaches (i.e. device-free) and opportunistic signals to the issue of people location, such as tomography, TDV cameras and the disaggregation of the electricity consumption. Apart from the position, additional signals, related to the human activity, can be obtained (e.g. appliance usage patterns) to give support during the activity performance. Besides, another remarkable objective is the development of a case study with the monitoring of elderly and diagnosis of sleep disorders (with support of clinicians) by using positioning data, BSN and disaggregation of energy, prioritising non-intrusive systems.

Furthermore two research stays have been possible thanks to the fundings from the University of Alcalá fellowship program (FPI 2011) and its mobility program: a three month research stay in Tyndall National Institute (Cork, Ireland) and another three month research stay in the Agents, Interaction and Complexity Research Group (AIC) from the University of Southampton (Southampton, United Kingdom).

This thesis has been also partially supported by other international institutions under the following projects:

- The ORCHID project at the University of Southampton):

  http://www.orchid.ac.uk/

- The ITOBO project (Information and Communication Technology for Sustainable and Optimized Building Operation) at the University College Cork:

  http://careers.tyndall.ie/myprojects/project/11303

1.3 Thesis outline

From the ground to the roof, the content of this thesis has been arranged in six different chapters, following a natural flow, that guides the readers from the basis knowledge to the conclusions, passing through the low-level contributions in the NILM field and the high-level contributions in activity monitoring. Because Chapter 4 and Chapter 5 present different proposals to overcome different challenges, it makes more sense to present results and conclusions just after each proposal. Hence, we have made these chapters self-contained to facilitate the reading: the proposal, results and conclusions are sequenced in the chapter. The content of each chapter is as follows:

- **Chapter 2: "State of the Art"**

A thorough survey of literature is presented in this Chapter. Always keeping in mind our final goal: the use of NILM techniques for activity monitoring for elderlies. Therefore, the reader will be guided through the NILM basis methods that may differ according to the used sampling rate. A low sample rate allows a better integration in detriment to the accuracy and eventless NILM algorithm are predominant: Combinational Optimisation, Hidden Markov Models and Deep Neural Networks are the trend in the field. Higher sampling rates allow the implementation of event-based NILM algorithms that may have greater accuracy and are thought to work on real time. These algorithms are often divided into three stages: "event detection", "load signature training" and "load signature classification". Then, a discussion about the need of new applications for NILM will open the door to the survey of a high-level application: activity monitoring for elderlies. The importance of activity monitoring in healthcare monitoring methods is highlighted and different methods (i.e. direct and indirect ones) are evaluated according to its degree of intrusiveness and scalability.

- **Chapter 3: "General overview of the proposal"**

It serves as a nexus between the thesis contributions: Chapter 4 and Chapter 5; and the state of the art (Chapter 2). Thus, having identified the weakness and the strengths in NILM and activity monitoring, we proceed to present our approaches based on the acquired knowledge. A discussion about the chosen method is presented based on literature

and the goals are enumerated. Furthermore, we present a diagram that summarises all contributions as a whole, given a complete solution.

- **Chapter 4: "Energy disaggregation algorithm proposal"**

We present our contribution to the NILM field. In other words, this is what we call our low-level application. To adapt NILM to activity monitoring purposes, we have chosen to devise a new event-based NILM algorithm that has a high accuracy. The event detector is based on the envelope extraction through the evaluation of the root mean square value of the current and voltage. A novel load signature is presented, which has much information about the occurred event and minimise the signature overlapping. This method is based on the extraction of the active, reactive and distortion power over the transient and the steady states. To learn the signatures and to discriminate between them, a classifier based on Principal Component Analysis has been devised. This technique is often used in image processing to compress the variance and to reduce the file size. Thus, we train our classifier to obtain general compressions of different appliance types. Given a new instance, the model that generates the best compression will be the one which labels the instance. The general models only need to be trained once from a dataset, hence, we called this approach "unsupervised" method. Furthermore, experimental results are also presented in this Chapter.

- **Chapter 5: "Activity Monitoring proposals"**

Given the disaggregated data from a Smart Meter, what can we do to monitor the health of elderlies? This Chapter presents our three proposals. We have found that some appliances may have a strong correlations with daily patterns. For instance, the kettle is strongly correlated with the daily routines in the United Kingdom (UK). The Log Gaussian Cox Processes (LGCP) approach learns the usage pattern of a single appliance including an uncertainty, which is strongly correlated with the daily routing, and it infers the activity monitoring of the inhabitant. It triggers an alarm whenever the subject departs from his daily routine. Multi-appliance monitoring allows to detect pattern deviations in short-term monitoring (i.e. within a day) and long-term monitoring (i.e. within months), this may be useful to detect degenerative diseases such as Dementia and Alzheimer. Therefore, we present two methods to evaluate someone's routine over time based on learning

the usage pattern of several appliances: a Gaussian Mixture Model (GMM) approach and a Dempster-Shafer Theory (DST) approach. Furthermore, empirical results are also presented in this Chapter: LGCP approach is suitable when there is a strong correlation of the routine with one appliance; otherwise, the LGCP approach performs better than the GMM does because it models the uncertainty. This last one is important; it models the free will of the inhabitant and the sparsity of events, and, therefore, less false alarms are produced.

- **Chapter 6: "Conclusions"**

This last Chapter summarises the whole work of this thesis dissertation. What are the strengths and weakness of our approach? What are the benefits to the society from using our approach? What are the limitations? These and other questions will be answered in this Chapter. Furthermore, contributions to the research community will be presented in terms of publications, collaborations and etcetera. Important remarks about future work will be made to guide all research to come in this topic.

Chapter 2

State Of the Art

Through this chapter, we intend to give to the reader a through and detailed view of the current state of the art with regard to energy disaggregation and activity monitoring. This sets the scenario where developments of this thesis start from. Following the natural flow from low to high level application, we firstly present the current state on the energy disaggregation domain. Then, we delve into the last advances in activity monitoring, specially focused on elderly monitoring.

2.1 Energy disaggregation

2.1.1 Definition of energy disaggregation

The notion of energy disaggregation begins from a simple quest: given the total energy consumption from a building as the only input, can we unravel the energy consumption of every appliance plugged into the mains? Therefore, energy disaggregation embraces a collection of techniques (e.g. those based on machine learning and signal processing) aimed to answer that question. To illustrate the process, we select a single day from a random household in the UK-DALE dataset [Kelly and Knottenbelt, 2015b] (see also Apendix A) using the NILMTK toolkit [Batra et al., 2014]. Figure 2.1.a depicts the smart meter data where, from a visual inspection, we can recognise power steps due to connection and disconnection of appliances. Nevertheless, we do not know a-priori which appliances are responsible for this signal. Applying energy disaggregation algorithms and

techniques, this aggregated consumption can be broken down to discover what appliances lay underneath and what their respective energy consumption are, as seen in Figure 2.1.b

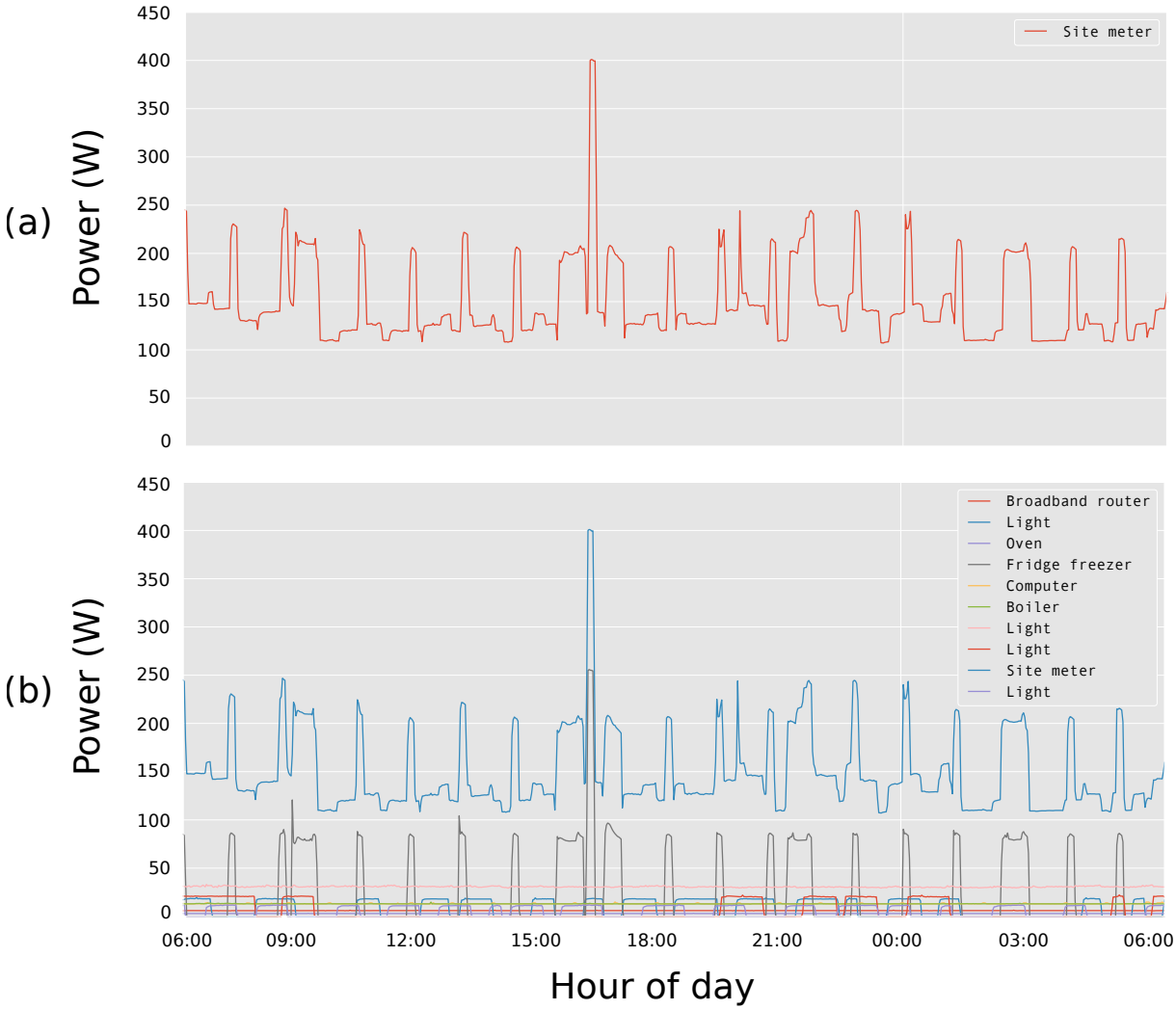


Figure 2.1: The energy disaggregation process in an UK-DALE household during a single day: (a) The Smart Meter data; (b) The appliance-level consumption.

Energy disaggregation is most commonly known as Non-Intrusive Load Monitoring (NILM) or Non-Intrusive Appliance Load Monitoring (NIALM) to highlight its most compelling attribute: the appliance monitoring inside a household without the need of an inner installation. In terms of non-intrusiveness, the ideal source of information is the smart meter data, which would have been already deployed by the utility company, accessed through Power Line Communication (PLC). Nevertheless, the sample rate is about one sample every 15 minutes and, consequently, only a coarse-grained disaggregation can be achieved.

Likewise, researches rely on the access to the Smart Meter data through either directly reading the sensor measurements, through a Home Area Network (HAN) or single-point sensors developed by third companies to achieve a greater accuracy on the disaggregation (see Section 2.1.5). Henceforth, we refer to energy disaggregation as NILM.

2.1.2 Evolution of NILM

Energy disaggregation is not a new concept, the first appearance dated from 1992 in [Hart, 1992]. George Hart was the first to coin the term NALM (Non-intrusive Appliance Load Monitoring) and to set the basis of a number of NILM concepts, they will be following discussed in Section 2.1.3. However, it has not been until 2010 that NILM has arisen the interest within the research community. To support this statement, Figure 2.2, created by Oliver Parson ([Parson, 2015a]), shows the number of publications around NILM since early 90's up to mid 10's (it does not specify the range limits). It can be clearly observed a disruptive point in 2010 when the number of publications begins to increase exponentially.

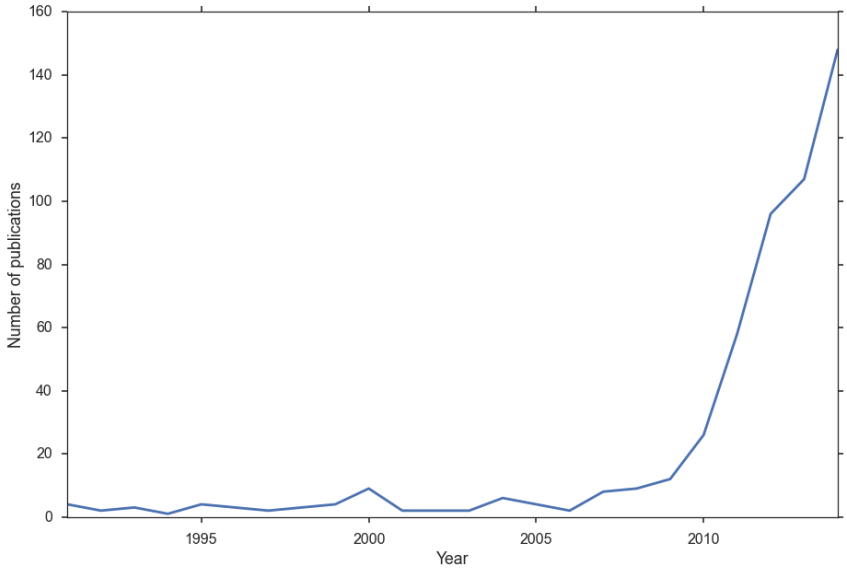


Figure 2.2: Evolution of NILM publications by year [Parson, 2015a].

This reborn interest might well be related to the recent apparition of Smart Grids (SG). One of the major commitments that our society faces since the new century is to reach energy sustainability and this has motivated the breakthrough of Smart Grids (SG) as a

new paradigm to achieve Energy Efficiency. Special attention is drawn over buildings and households that account for more than 41% of energy consumption in the European Union (EU), whereas residential use represents 63% of total energy consumption in the buildings sector, [Balaras et al., 2007]. Studies state that providing customers with real-time information at aggregate level may induce them to adopt energy-conserving behaviours that could lead to save 10-15% on power. Thus, Smart Meters (SM) are called to play an important role for customer, helping them to manage their Demand Response (DR) (changes in the pattern consumption in response to the electricity market) in a more efficient way. Consequently, many countries are massively deploying smart meters to achieve a finer-granularity on the DR. Figure 2.3 shows the numbers in Europe: around 35, 32, 32, 27 and 23 millions of total smart metering installations forecasted by 2020 in France, Germany, Italy, United Kingdom and Spain, respectively. This new scenario, where a fine-grained DR is needed and the presence of smart metering data is greater, has awaken a renew interest in NILM.

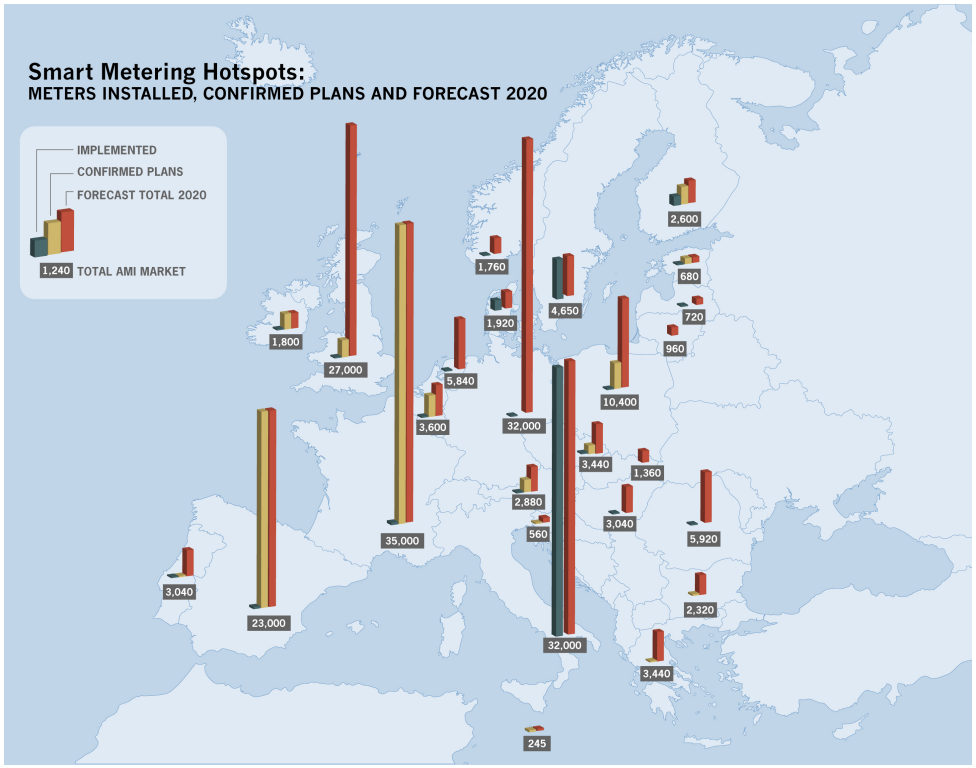


Figure 2.3: Smart Metering installations in Europe by countries: installed, planned and forecasted by 2020. Source: [van der Zanden, 2011].

2.1.3 NILM algorithms

The state of the art in NILM is often organised into event and eventless based algorithms. The former pursues the detection and classification of transient states (i.e. switching on/off events or state changes of appliances), whereas the latter aims to foresee the state of a set of appliances (i.e. those to be disaggregated) at each sample based on previous observations. Another popular classification falls into supervised and unsupervised algorithms. However, we prefer another type of classification, based on the frequency sampling: Low-Frequency (LF) and High-Frequency (HF) algorithms. This feels a more natural process for us to discuss the different features of NILM algorithms and their development. After all, NILM algorithms are designed to meet some sampling requirements either to be integrated into smart meters or third-party devices. The sampling rate also defines the resources that NILM algorithm can play with to achieve a better accuracy.

With concern to the supervised and unsupervised methods, it is worth to note that these concepts may differ a bit from the classical machine learning approaches. In the machine learning domain, supervised algorithms are those that need labelled data during the training process, whereas unsupervised algorithms are only fed with unlabelled data. Back to the NILM domain, there is a subtle difference about the "unsupervised" concept. An unsupervised NILM algorithm is such that does not use labelled data from houses to be disaggregated (i.e. test houses), however, it may use unlabelled data from test houses and labelled data from training houses [Parson, 2015b]. For instance, eventless algorithms presented in this Section such as [Kelly and Knottenbelt, 2015a, Parson et al., 2014] are unsupervised.

Back to the classification based on the frequency sampling, there is a trade-off between integration (also called intrusiveness) and the disaggregation accuracy, and, furthermore, both are strongly correlated with the sampling frequency. We depict this correlation in Figure 2.4, which illustrates the state of the art in terms of frequency. The disaggregation accuracy proportionally increases as the sampling frequency does, whereas the integration decreases. Besides, to facilitate the comprehension of this Figure, we provide the Table 2.1 that lists the acronyms used and whose studies will be detailed in the following subsections.

With regard to the integration and intrusiveness, utility companies often use Power Line Communication (PLC) to transmit load curves that have been downsampled to 1 sample every 15 or 30 minutes. This sampling rate allows the lowest degree of intrusiveness,

obtaining measurements directly from the central station, which collects the SM data through Power Line Communication (PLC). In some countries, such as UK, Smart Meters may include a wireless port (e.g. ZigBee) to enable higher resolutions (i.e. load curves sampled every 1 or 10 seconds) via the Home Area Network (HAN). Thus, a dummy device may capture this data and send it to the cloud for disaggregation. We enter in a higher intrusiveness degree, placing that device within the household. Nevertheless, communication through the HAN is not guaranteed in most countries. Finally, higher sample rates (i.e. higher than 1 Hz) may need a hardware plugged into the mains (i.e. as those presented in section 2.1.5) or to integrate the energy disaggregation algorithm into the smart meter, which could entail more complex integration. The smart metering standardization regulates the measurement of harmonics to evaluate distortion in the mains, [Sanduleac et al., 2015, Terzija et al., 2007]. That means regularised smart meters sample up to 8kHz, sometimes up to 1MHz, [Chen et al., 2011].

With concern to the disaggregation accuracy, at the very lowest position on the sampling frequency scale (i.e. a sampling interval of 15 or 30 minutes), the disaggregation is very coarse and only major appliances can be disaggregated ([Kolter et al., 2010]). In a medium scale, we have what we call LF NILM algorithms, which include the majority of the literature: [Kolter and Johnson, 2011, Hart, 1992, Kelly and Knottenbelt, 2015a, Kim et al., 2011, Kolter et al., 2012, Bonfigli et al., 2015, Parson et al., 2014]. Most recent efforts on NILM have been focused on this sampling rate, since measurements can be easily collected and sent to a cloud service where disaggregation is performed. The use of load curves sampled every 1 or 10 seconds makes the event detection (i.e. appliances switching on and off) impractical and, consequently, they are eventless methods in their majority, but for some exceptions [Baranski and Voss, 2004]. They are able to disaggregate the majority of appliances presented in a house and their disaggregation accuracy has been measured up to 87%.

Finally, HF NILM algorithms are placed at the highest scale of the sampling frequency, where the recorded disaggregation accuracy has been up to 99.9%: [Liang et al., 2010a, Hassan et al., 2014, Srinivasan et al., 2006, Gao et al., 2015]. They are mostly event-based algorithms, which enables real-time activity monitoring applications. Furthermore, a few studies explore the potential of very high frequency sampling (i.e. higher than 100kHz) where ElectroMagnetic Interference (EMI) noise allows even to discriminate between same appliance types located in different rooms ([Patel et al., 2007, Gupta et al., 2010]).

We would like to state that Figure 2.4 is a biased vision of the state in NILM based on this survey. It serves to the purpose of structuring the state of the art, rather than establishing a strict classification of it. The disaggregation accuracy is also very subjective: one may say that discriminating between same appliances located in different rooms does not mean disaggregation accuracy; or there is not a fair comparison between LF NILM algorithms and HF NILM algorithms as most HF algorithms presented in the Figure only evaluate the load identification, which will be explained in the following subsections. However, it does not exist a complete unbiased comparison as yet, and assumptions are to be made. From our opinion and as a big picture, HF algorithms may disaggregate better than LF algorithms when both are optimised.

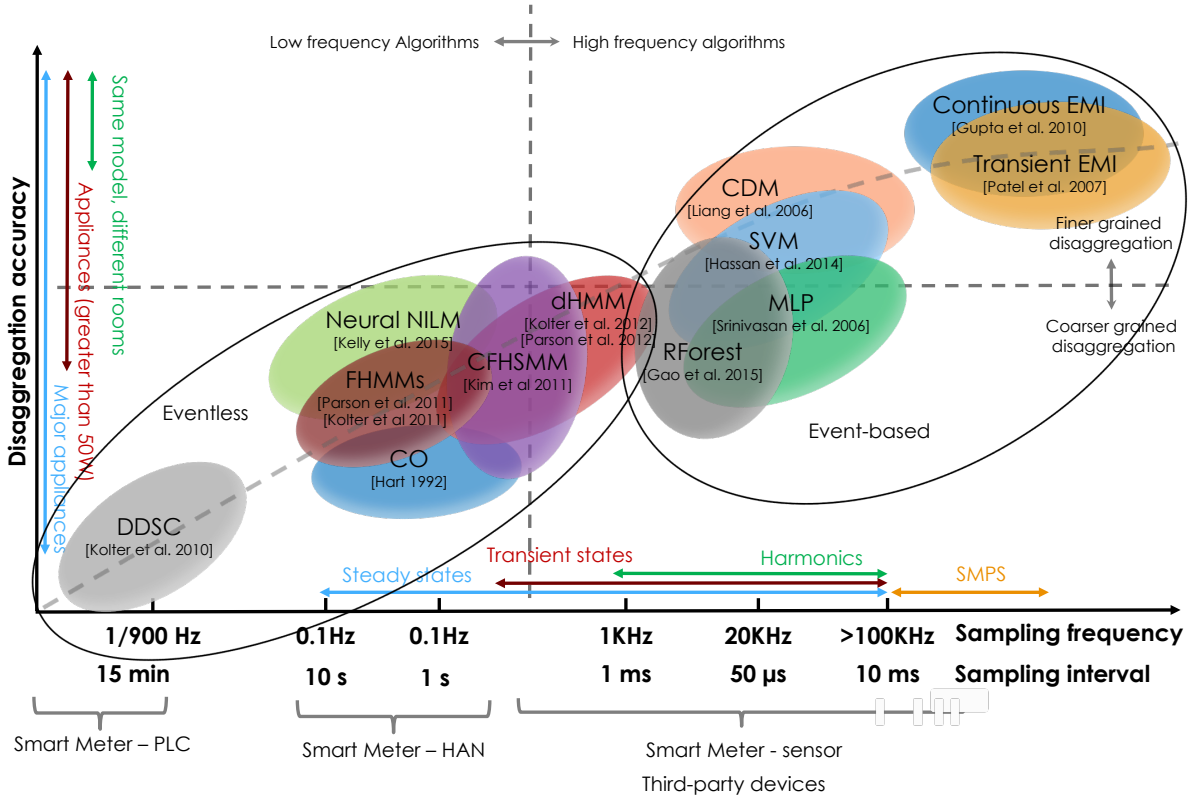


Figure 2.4: Disaggregation accuracy related to the sampling frequency.

Acronym	**Stand for**
DDSC	The Disaggregation via Discriminative Sparse Coding
FHMMs	Factorial Hidden Markov Models
CO	Combinatorial Optimisation
CFHSMM	Conditional Factorial Hidden Semi Markov Model
dHMM	difference Hidden Markov Model
RForest	Random Forest
MLP	Multilayer Perceptron
SVM	Support Vector Machine
CDM	Committee Decision Mechanism
EMI	ElectroMagnetic Interferences

Table 2.1: Table of acronyms for Figure 2.4.

Figure 2.5 is another example of why HF algorithms performs better, where Table 2.2 indicates the name of some features that will be discussed in the following subsections. Transient states cannot be characterised for LF (i.e. 0.1Hz $< f_s <$ 1Hz where f_s is the frequency sampling), only the active power consumption P during steady states (i.e. normal operational mode of an appliance) and the difference between two steady states ΔP. However, many features can be extracted from a range of 1Hz $< f_s <$ 20Hz, such as the Distortion power D and harmonics (HAR). Higher frequency rates allow to detect EMI noise produced by Switch Mode Power Supply (SMPS) that are unique for each appliance. Therefore, the richer feature extraction is the larger space to discriminate between appliances is. Figure 2.5 also serves to frame NILM into the energy disaggregation domain. Likewise, energy disaggregation can be also conducted based on Intrusive Load Monitoring methods (ILM) using multi-point sensors as in [Kato et al., 2009], or using an intrusive signature as in [McWilliam and Purvis, 2006]. However, these methods are out of scope in this thesis as they are much intrusive.

Following the discussion, we proceed to detail LF and HF NILM algorithms. Then a discussion is open about the practical application of NILM in a real scenario (Section 2.1.4).

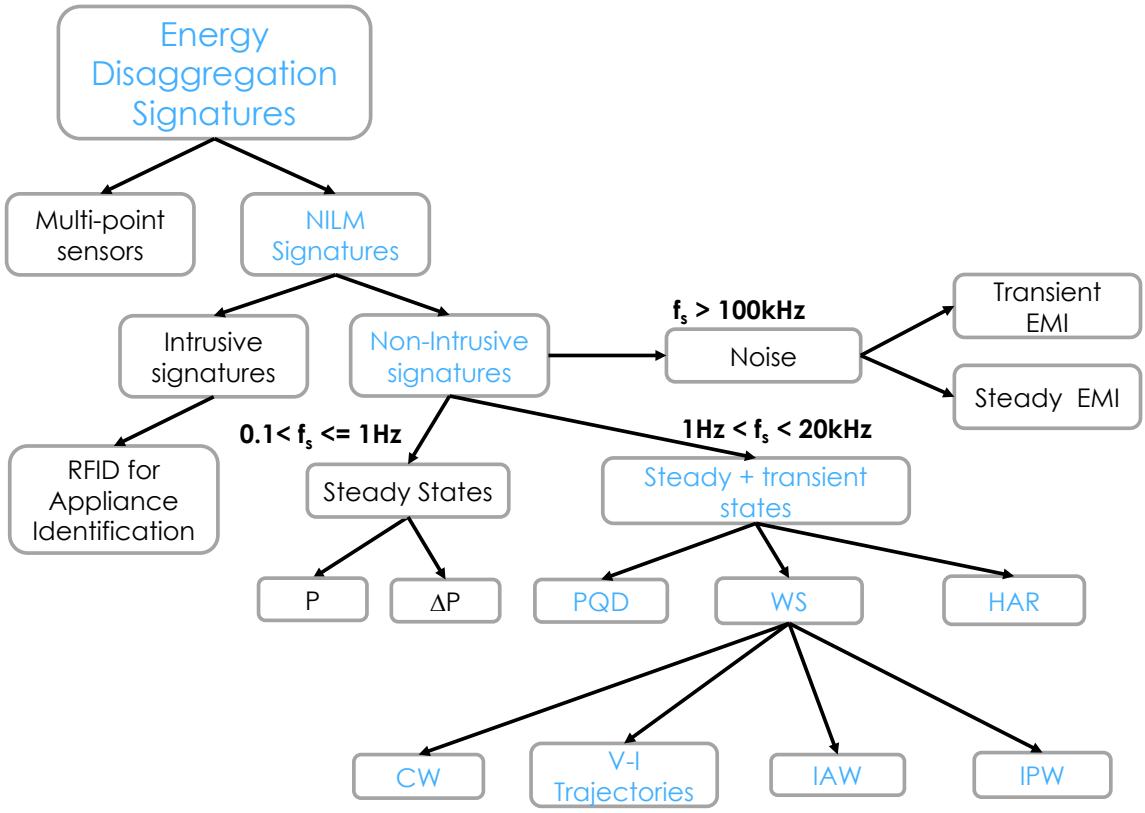


Figure 2.5: Available signatures by frequency sampling.

Acronym	**Stand for**
P	Active Power feature
ΔP	Delta value of active Power feature
PQD	Active, reactive and distortion power trajectories
HAR	Harmonics
WS	Wave-Shaped features
CW	Current Waveform feature
V-I	Voltage-Current trajectories
IAW	Instantaneous Admittance Waveform feature
IPW	Instantaneous Power Waveform feature

Table 2.2: Table of acronyms for Figure 2.5.

2.1.3.1 Low-Frequency NILM algorithms

The only use of smart meters as the hardware unit is the most compelling goal. Therefore, most recent efforts lead to the use of a low sampling rate to meet smart meter standardization.

The ideal scenario is such that the data are received through the power line. Nevertheless, this approach is almost impractical: the sampling interval is about 30-15 minutes and this means that only a very coarse-grained disaggregation is possible. A coarse-grained disaggregation is conducted in [Kolter et al., 2010], where sparse coding methods are used to disaggregate data collected at 15-minute intervals. The Disaggregation via Discriminative Sparse Coding (DDSC) learns general appliance model for a large number of houses (i.e. 590 homes in total). They claim this method has a classification accuracy between 47% to 59%. However, the results may be biased since the test were carried out only on simulated data.

To achieve a better accuracy, researches need higher resolutions. Most Smart Meters can deliver their data through a HAN with a sampling interval that ranges from 1 to 10 seconds. This seems a suitable trade-off between intrusiveness and disaggregation accuracy and that is why most of studies are focused on this sample rate. In [Baranski and Voss, 2004] a genetic algorithm along with a dynamic programming approach are used to detect and classify edges from smart meter data with 1-second resolution. Although non empirical results are provided, this event-based method seems to be enough to disaggregate the major appliances such as the fridge and heating. Nevertheless, non event-based or eventless based method are more popular regarding this sampling interval. Typically, transient states last between 100-500ms and, hence, at least 100Hz are needed to analyse them [Froehlich et al., 2011]. This does not allow to extract many features from events with the aforementioned sampling rate, only the power change between the previous and the next steady states, [Parson, 2014].

Eventless NILM algorithms predict the state (e.g. ON or OFF) of a set N of appliances at each sample time. If the set N is large, the number of potential combinations exponentially increases and so the complexity does. Furthermore, the number of iterations proportionally increases as the sampling rate does and this why these methods are not very popular for high sampling frequency. From [Batra et al., 2014], we can highlight two

classical non-event based methods that are widely used as benchmarks: Combinatorial Optimisation (CO) and Factorial Hidden Markov Model (FHHM).

Combinatorial Optimisation was firstly proposed by George Hart in his seminal work ([Hart, 1992]). This method solves a optimisation problem at each sampling time t that can be illustrated by the Knapsack problem in Figure 2.6. That is, given the knapsack in the Figure with capacity up to $15kg$ and the five items with different weights and values, what is the optimal combination of items that maximises the value on the knapsack without overloading it? The problem is formulated following (2.1), where $n = 1, ..., N$ is a specific item from a total of N, v_n and w_n are the value and weight for item n, $\boldsymbol{x} = [x_1, x_2, \ldots, x_N]$ is a sequence of activation that maximises the problem, and W is the knapsack maximum weight. George Hart adapted the knapsack problem to the energy disaggregation domain relaxing the constraint of overloading and simplifying to a knapsack problem 0-1 where $v_n = w_n$. Likewise, the aggregated power demand $P(t)$ is defined as in (2.2) at each sampling time t. $P(t)$ is composed by the addition of the power demand P_n for every activated appliance $n = 1, 2, \ldots, N$, and an assumed error $e(t)$, where the vector $\boldsymbol{x} = [x_1, x_2, \ldots, x_N]$ refers to the activation sequence. Thus, the knapsack problem finds the most likely sequence $\boldsymbol{x}$ that minimises the error $e(t)$, as in (2.3).

As indicated in [Hart, 1992, Batra et al., 2014], there is a number of difficulties applying CO: firstly, the single power demand of each appliance P_n needs to be a-priori known, which could be a cumbersome task if many appliances are presented in the household; furthermore, small variations on the aggregated power demand $P(t)$ can draw big changes on the activation sequence $\boldsymbol{x}$. To overcome these disadvantages, many studies have pointed out the use of Hidden Markov Models (HMM).

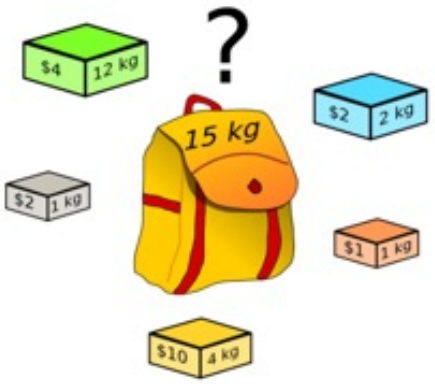


Figure 2.6: The knapsack problem.

$$\arg\max_{\boldsymbol{x}}(\sum_{n=1}^{N} v_n x_n)|\sum_{n=1}^{N} w_n x_n < W \tag{2.1}$$

$$P(t) = \sum_{n=1}^{N} x_n(t) P_n + e(t) \tag{2.2}$$

$$\hat{\boldsymbol{x}}(t) = \arg\min_{\boldsymbol{x}}|P(t) - \sum_{n=1}^{N} a_n P_n| \tag{2.3}$$

Factorial Hidden Markov Model is a temporal probabilistic method that is graphically illustrated in Figure 2.7. We try to simplify the notation to describe this method because it is very dense and out of scope in this thesis, refer to [Parson, 2014] for a more detailed explanation. Each of the n appliances can be described as an independent HMM, where $x_t^{(n)}$is the unobservable variable (e.g. ON or OFF) at a certain sampling time t. Likewise, $y_t^{(n)}$ denotes the observed variable, which is the power demand for the appliance n, at the sampling time t, and for a given state $x_t^{(n)}$. Actually, the appliance power demand is not observed on a real case scenario, since only the aggregated power demand can be observed. Therefore, all n-HMM must be unified in the FHHM shown in Figure 2.7. The observed variable (i.e. the aggregated power demand) is denoted as the sum of the all appliance power demands: $\overline{y}_t = \sum_{n=1}^{N} y_t^{(n)}$. Hence, the problem is summarised in finding the posterior probability of every appliance consumption $y_t^{(n)}$ at any time $t = 1, 2, \ldots, T$ and for any given aggregated power demand $\overline{y}_t = \overline{y}_1, \overline{y}_2, \ldots, \overline{y}_T$. FHHM-based methods in NILM may disaggregate better than CO, but they have a higher complexity which makes them difficult to scale, [Batra et al., 2014]. Hence, the inference is conducted by approximate methods such as the Viterbi algorithm or the Gibbs sampling [Parson, 2014].

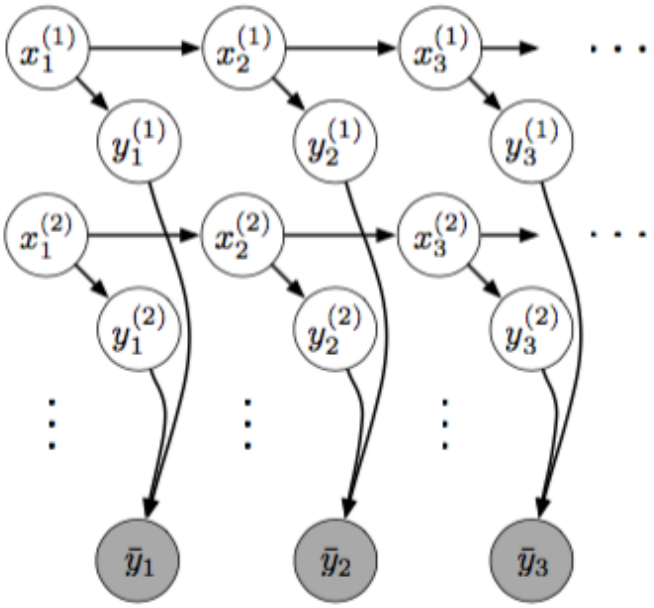


Figure 2.7: Graphical representation of a Factorial Hidden Markov Model, [Kolter and Johnson, 2011].

The first use of FHHM in NILM can be found in [Kolter and Johnson, 2011]. This approach achieved an accuracy of 47.7%; however, it has the disadvantage that all the appliances in the household need to be manually labelled, what is impractical in a real scenario. [Kim et al., 2011] proposed an extension of FHHM and Hidden semi-Markov Model known as Conditional Factorial Hidden Semi Markov Model (CFHSMM). This approach increases the accuracy by introducing rules of activations between appliances (e.g. if the game console is ON, then the TV must be on) and modelling the duration of appliance states by a probability distribution function. Only disaggregated appliances need to be manually labelled. The F-measure score (see Appendix B) raises up to 87% using 10 appliances; however, this value drops according the number of disaggregated appliance increases and, for a typical house containing up to 20 appliances, the F-score would be quite low. Furthermore, the difference Hidden Markov Model (dHMM) is proposed in [Kolter et al., 2012], where the observed variable is composed by "delta values" (i.e. differences between two consecutive samples) from the aggregated power. This approach allows to ignore all appliances but the ones that need to be disaggregated, those needing manual labelling on the data. The performance is high: 87.2% and 60.3% for precision and recall, respectively (see Appendix B); nevertheless, this has been measured at a high frequency sampling.

Moreover, a variant of the aforementioned dHMM is proposed in [Parson et al., 2012], where the aggregate power is also measured and tracked. This fact discards ON states of a certain appliance whenever the aggregated power is lower than the mean consumption of that appliance, avoiding missed event happens. Manual labelling seems to be a non-desirable feature since it implies to firstly take appliance-level measurements within the household and, in most cases, that is not possible. Hence, general models of appliance types have been developed to avoid that. Thus, given an appliance type, a general model can be extracted from test houses and apply to a new house without the need of new labelling; the general model adapts itself to the particular appliance inside the house. This has been conducted in [Parson et al., 2014, Parson et al., 2011]: it disaggregates the three highest energy consumption appliances (i.e. the refrigerator, the clothes dryer and the microwave) using FHHM and prior knowledge of the general models to avoid labelling. This proposal has an accuracy 83% and disaggregates the 35% of the total energy. The same author also proposed a variant of dHMM where we also measure and track the aggregated power

Although most efforts in NILM have been focused on HMM methods, there is a very new trend: the Deep Learning. This is a branch of machine learning that aims to learn a high level of abstraction in the data. Since 2006, the Deep Learning has made huge improvements in domains such as speech recognition and vision computing. As happened with HMM models, the Deep Learning can be another example of a well-stablished technique from another domain that can benefit much to NILM. Likewise, Jack Kelly ([Kelly and Knottenbelt, 2015a]) proposes what he calls Neural NILM based on Deep Neural Networks. This is an Artificial Neural Network (ANN) with many layers able to perform Deep Learning from the data. He argues that, in contrast to High-Frequency NILM algorithm, Low-Frequency NILM algorithms are poor in feature extraction and only learn from steady state changes. Nevertheless, many features are still present on a Low-Frequency data and human eyes are able to detect them. Neural NILM can extract those many features on low frequency and use them to achieve a better accuracy. In fact, experimental results show that Neural NILM outperforms CO and FHMM, as can be observed in Figure 2.8 where three neural network are used: autoencoder, rectangles architecture and Long Short-Term Memory (LSTM). It is worth to note that algorithms have not been optimized and results could be improved in all sides.

As in FHMM approaches, Neural NILM is an unsupervised method (i.e. it can be trained with labelled data from different houses than the one from the training data). However,

the computational cost for the training is considerably higher (i.e. days and weeks of training per appliance). Still, this training process is carried out just once and it does not have to be repeated. Furthermore, Neural NILM needs a lot of data to be trained and this data can be synthetic data composed out of real data. Deep Learning is still very recent in NILM and there is much to explore. The performance can be improved, although for that it is needed to optimise the 1 to 150 millions parameters present in the Neural NILM methods in [Kelly and Knottenbelt, 2015a]. An interesting feature of Neural NILM is the possibility of doing a "pre-training" on the unlabelled data before the training process using the labelled data. This may not have much sense on another domains such as vision computing where there is a high availability of labelled data, but, in NILM, the amount of unlabelled data is much higher than the labelled one. The "pre-training" would not help to disaggregate but can extract singular features from data such as peaks of consumption or patterns.

In summary, Low-Frequency (LF) NILM algorithms are the trend on the domain because they are less intrusive and the hardware cost is lower (i.e. considering that Smart Meter data can be easily accessed via HAN, which still is not clear in all countries (see Section 2.1.5). They may have less accuracy than High-Frequency NILM algorithms but, in terms of energy-savings, this is good approximation as main appliances can be easily detected. However, for other purposes apart from energy-savings, it may be interesting to disaggregate other appliances that have smaller power demands. For instance, manually operated appliances may play a main role for human activity monitoring, although, they often have a low power consumption. Hence, High Frequency (HF) NILM algorithms may benefit other approaches where energy-savings and hardware low cost are not predominant factors. In fact, HF NILM algorithms are more focused on events, what is more indicated for real-time diagnostic and activity recognition applications, whereas eventless algorithms are more appropriate for medium to long term energy savings, as described in [Hassan et al., 2014]

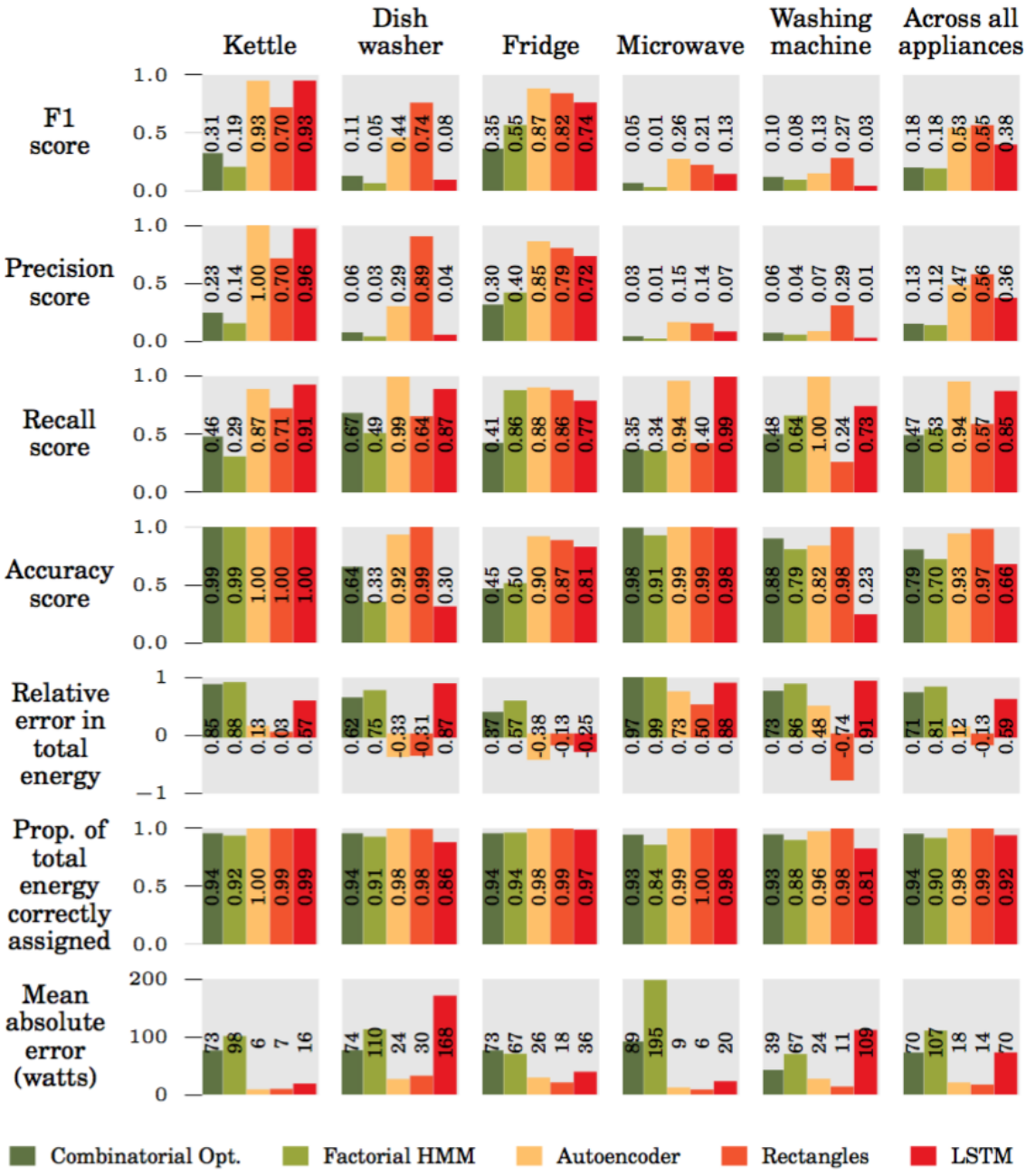


Figure 2.8: Comparison between CO, FHMM and Neural NILM using NILMTK, by [Kelly and Knottenbelt, 2015a].

2.1.3.2 High-Frequency NILM algorithms

We grouped in this Section those studies that assume higher frequency rates than 1 Hz. All studies presented here are event-based approaches. Consequently, more features can be extracted and the accuracy increases. However, their integration with smart meters is worse and they need third-parties meters. In term of disaggregation, these algorithms are less dependent on the appliance power demand. This means that the performance does not improve on major appliances as in eventless algorithms, however, it does for those appliances with an unique Load Signature (LS).

Event-based NILM algorithms may be more complex than eventless algorithms and they are often composed by three different stages: event-detection, Load Signature extraction and load identification. We have divided the survey into different sections because most studies only address one of these challenges at a time.

2.1.3.2.1 Normalization

The voltage signal along the mains tends to be a sinusoidal signal with a constant frequency and amplitude (e.g. in the European Union the frequency is 50Hz and the amplitude is $230 * \sqrt{2}$ V). However, there are small changes due to noise, voltage drops and distortions produced by appliances. The current signal is susceptible to these changes and it might vary its amplitude for the same transient state produced at a different time. This can affect the performance and, consequently, most studies perform a normalization of the signal before the event detection and classification process. Equation (2.4) denotes a normalization of the power where $V_{norm} = 230V_{rms}$ (i.e. in European Union), $v(t)$ and $P(t)$ are the measured voltage and power from the mains [Hart, 1992, Weiss et al., 2012]. To normalise the current signal, the process is the same than in (2.4) changing $P(t)$ for the measured current signal $i(t)$.

$$P_{norm}(t) = (\frac{V_{norm}}{v(t)})^2 * P(t) \tag{2.4}$$

2.1.3.2.2 Event detection

Event detectors aim to detect transient states and start-up features of appliances from the aggregated power demand that are next used to extract and synchronise signatures. They

can be distinguished into three types of approaches: (a) expert heuristics; (b) probabilistic models; and (c) matched filters, [Anderson et al., 2012a].

Expert heuristics methods involve a set of different techniques and a certain a-priori knowledge about appliances to be detected. George Hart ([Hart, 1992]) proposed the first event detector that is based on the active power changes: that is, two consecutive measurements are considered in the same steady states unless their active power difference is higher than a certain threshold (15 Watts) in which case it is marked as a transient state. Nevertheless, no results were presented in this work and a power threshold of 15 Watts is an ambitious goal that may not work well in a real scenario. An extension of Hart's detector is proposed in [Farinaccio and Zmeureanu, 1999] where, apart from the power step changes, other rules are taken into account. This study focuses on the detection of the major appliances such as hot water heater and refrigerator. For each appliance, a set of rules is created, such as the number of data points during the transient state, total power demand and power variation. Expert heuristic approaches often need to initialise some parameters and, therefore, they need a-priori knowledge.

Probabilistic models provide a probability that is used to make decisions about the occurrence of events. They perform well, although a training process is needed to set parameters and to learn the statistics models in different scenarios (i.e. different household mains). That is the case in [Anderson et al., 2012a, Anderson et al., 2012b], where, apart from a detailed survey on the state of the art, a Generalised Likelihood Ratio (GLR) approach is used. It requires a training process to tune five parameters, whose values change depending on the metric, in order to optimise performance. Furthermore, in [Pereira et al., 2014] another GLR variant is used to detect events and to label a data set. The study claims to detect the 95% of events in their tests with the constrains of a power threshold and minimum time separation between events of 30 Watts and 5 seconds, respectively. Both aforementioned studies evaluated the likelihood at every sample.

Matched filters may require higher data acquisition rates to extract the signal waveform and to correlate it with a known pattern; however, no previous knowledge is needed. Signal normalization is usually used to shield the analysis again voltage drops and to generalise their approaches. In [Weiss et al., 2012], a similar approach to the one in [Hart, 1992] is carried out, where the use of average and derivative filters is remarked to filter noise out. The study detects events and performs the disaggregation at the same time, obtaining a recognition rate (i.e. appliance events correctly classified) of 87% and the power threshold

is 10 VA (apparent power). However, this was tested on a small private dataset of 80 switching events. The analysis of the current signal envelope is a commonly used method. Thus, a signal processing method is presented in [Alcalá et al., 2014], where the use of the Hilbert Transform extracts the envelope of the current waveform. Subsequently, by using average and derivation filters and thresholding the signal, transient events can be detected. Other relevant studies using envelope extraction is presented in [Shaw and Laughman, 2007], where spectral envelopes from several harmonics are extracted and a Kalman Filter is used to track events.

Most event detectors measure either the current or the power because voltage barely varies over time. However, switching-on events do affect to the voltage by introducing noise in the spectrum from 10 kHz to 100 kHz. In [Patel et al., 2007, Gupta et al., 2010], the event detector sweeps the high frequencies to detect events. A Fast Fourier Transform (FFT) of 2048 points is performed on a timing window of a few microseconds, then an average value is obtained over the 20248 points. The timing window moves and the process is repeated. If the new difference between the new average value and the previous one is higher than a threshold, then the beginning of the timing window is marked as an event.

Finally, the trend in recent years is the use of machine learning techniques and frequency domain analysis to achieve high performance results. Likewise, in [Barsim et al., 2014, Barsim and Yang, 2015] a clusterization method achieving up to 98% of True Positive Percentage (TPP). In [De Baets et al., 2016] the cepstrum components are obtained through a FFT analysis and used to detect events, achieving a maximum performance of 98% on the F1 score. Similarly, in [Wild et al., 2015] the authors apply a kernel fisher discriminant analysis to the current harmonics to obtain high performances of 98.78% and 99.66% for the recall and precision, respectively.

2.1.3.2.3 Load Signature(LS)

Once the events have been detected, we need to classify them and, for that, we need to extract LS from the electrical signal. Load signatures must be unique for each appliance in the household and they can be extracted either during steady states (i.e. appliance operational modes) or transient states (i.e. when an appliance switches between two operational modes). In any case, they should be carefully selected to avoid overlapping, so the best possible results can be achieved [Hassan et al., 2014, Zeifman and Roth, 2011]. Signatures based on transient states may be more rich than those based on steady

states, however they still have some limits to be taken into account. Transient states are divided into four types: on/off appliances; finite state machines, i.e. those with many consuming states, such as a washing machine; continuous variable consumer appliances, i.e. those with variable power draw that cannot be classified into states, such as dimmer lights; and permanent consumer appliances, i.e. those which are continuously on, also called vampire loads. Thus, signatures based on transient states cannot detect continuous variable consumer appliances and vampire loads [Zeifman and Roth, 2011]. Furthermore, transient states last a tiny portion of the steady states and they are difficult to set aside, hence, some studies extract signatures from both together [Srinivasan et al., 2006]. From [Liang et al., 2010b], we can extract several concepts about LS. Thus, we can speak of Composite Load (CL) if the signature results from two different appliances operating at the same time; macro-level signature, which is a signature or composite load sampled lower than 1 sample per cycle; and micro-level signature that is sampled faster than 1 sample per cycle. Likewise, we can distinguish between snapshot-form LS, which is a direct measurement of the variable, such as the aggregated power demand, and it can be a LS or a CL; and "delta form" LS, which is the difference between two consecutive snapshot form signatures. Therefore, a delta-form signature is more similar to the real LS of a single appliance. Furthermore, the feature-additive criterion is a composite load that can be break down into the sum of the appliance load signatures. For instance, the active power P meets the feature-additive criterion because the active power of two different appliances (i.e. their Load Signatures) can be summed up together and the result (i.e. the composite load) is the same as the aggregated active power measured by the smart meter.

The nature of LS can be diverse and there is not a predominant load signature yet that performs well for all households and appliances. In [Liang et al., 2010b] it is stated that active power P and reactive power Q are the most frequently features used to model load behaviour. In fact, the very first work in NILM addresses the evaluation of the P-Q plane (delta values for P and Q powers) [Hart, 1992]. These signatures characterise well the load appliance behaviour and they discriminate appliances on the P-Q plane (see Figure 2.9). For instance, a purely resistive load has active power and a null reactive power. Likewise, inductive loads have a delay factor in the current with respect to the voltage, which implies a positive reactive power. Capacitive loads are characterized by a leading current and, hence, they draw a negative reactive power. Mainly, incandescent appliances, such as kettles and light bulbs, are mostly resistive, whereas motors, such

as fans and heaters, and devices with power electronic converters, such as laptops, are inductive and capacitive, respectively.

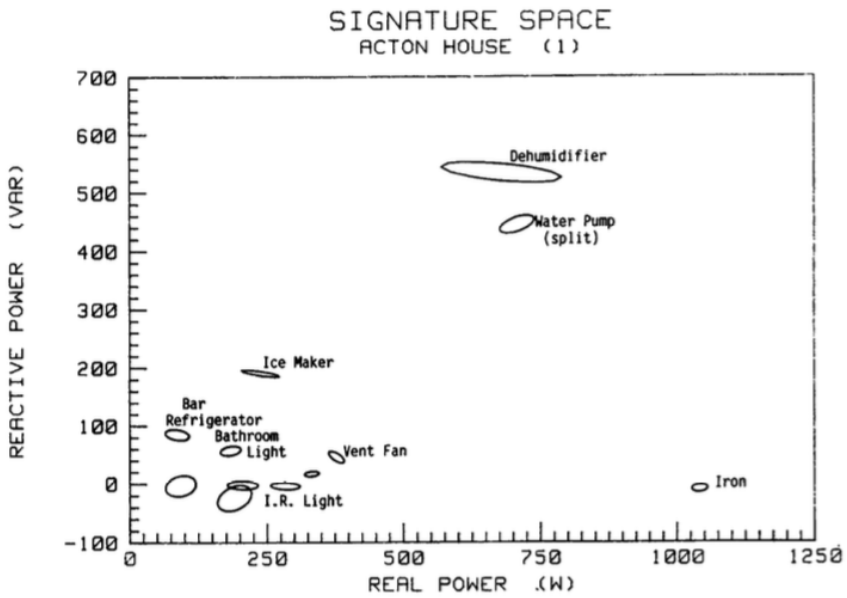


Figure 2.9: PQ plane and cluster of appliances, by [Hart, 1992].

Most studies based on the P and Q analysis to model load behaviours only take into account linear effects. However, within households, there is a high number of non-linear devices generating harmonic components that produce distortion on the sinusoidal waveform of the current. Michael Zeifman ([Zeifman and Roth, 2011]) states that the harmonic signatures (HAR) contain meaningful information up to the 11th harmonics and, therefore, micro-level signatures should be sampled at least at 1.2-2 kHz. For instance, the 3rd harmonic contains information about the computer power supply and help to discriminate computers from incandescent bulbs. Figure 2.10 shows the harmonic analysis of four appliances. It could be observed that the TV can be distinguished by its third and the fifth harmonic. The harmonic component analysis is often conducted by Fourier Transform, nonetheless, due to the nature of these signatures (i.e. they are non-stationary signals and have a limited duration), it may be better to perform a wavelet analysis that provides both frequency and time resolution [Chan et al., 2000]. [Weiss et al., 2012, Zeifman and Roth, 2011] introduce the distortion power D to model non-linear behaviours in loads achieving a high performance measured by the metrics.

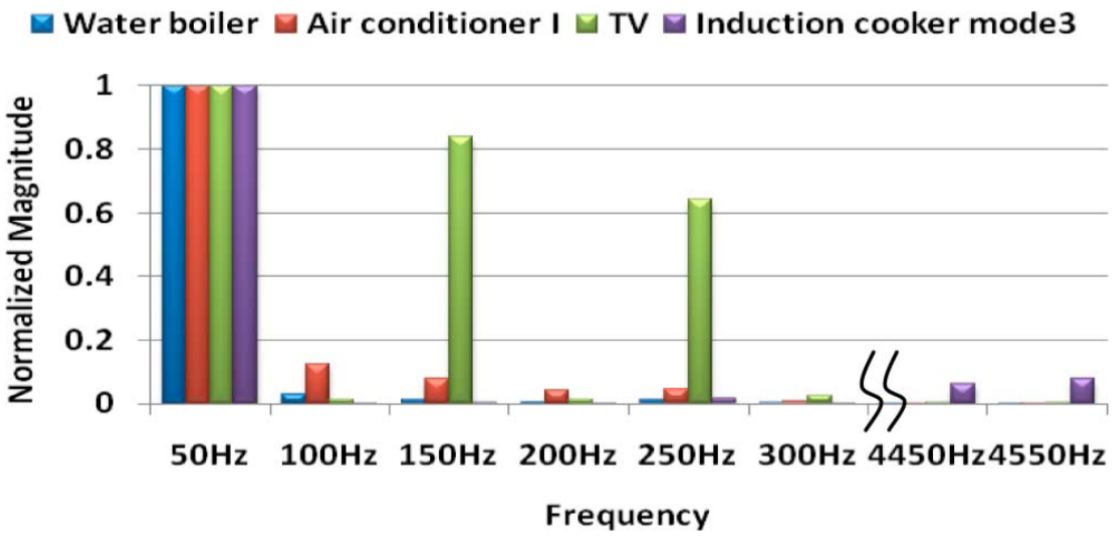


Figure 2.10: Harmonics signatures for four appliances: water boiler, air conditioner, TV set and induction cooker; by [Liang et al., 2010b].

Other most used load signatures are grouped together into the Wave-Shaped features (WS). For instance, the Current Waveform signature (CW), basically the aggregated current measured by the sensor $i(t)$, is very meaningful to describe the load behaviour. In contrast, the aggregated voltage $v(t)$ is merely constant, apart from small distortion produced by the plugged appliances. Thus, the CW signature is completed with values of $v(t)$ composing the Instantaneous Admittance Waveform (IAW) and the Instantaneous Power Waveform (IPW), both in (2.5) and (2.6), respectively [Liang et al., 2010b]. Current and Voltage (V-I) trajectories (see Figure 2.11), first introduced by [Lam et al., 2007], are becoming more popular as LS. In [Hassan et al., 2014] different features of the V-I trajectories are used to discriminate between appliances, such as the area enclosed into the trajectory and the non-linearity of the mean curve to measure the distortion. Moreover, in [Gao et al., 2015] the V-I trajectories are treated like binary images that fed the posterior classification algorithm. The latter also proposes the dimension reduction of signatures by applying a Principal Component Analysis (PCA).

$$IAW(t) = \frac{i(t)}{v(t)} \tag{2.5}$$

$$IPW(t) = i(t) \times v(t) \tag{2.6}$$

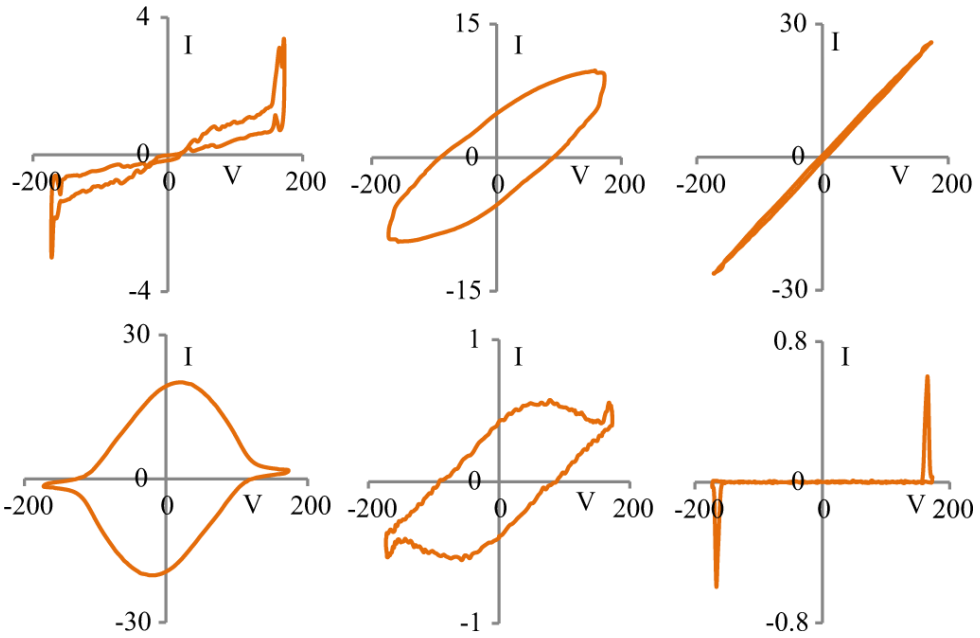


Figure 2.11: V-I trajectories for six different appliances, by [Hassan et al., 2014].

Finally, in a very small micro-level, Electromagnetic Interference (EMI) produced by the switch mode power supplies (SMPS) of modern appliances (e.g. laptops and TV) can be used as signatures [Zeifman and Roth, 2011, Patel et al., 2007, Gupta et al., 2010]. These may be the most unique signatures as SMPS are unique for each device (i.e. even uniques between appliances of the same type and different model). However, they have some disadvantages: the sampling rate must be very high, EMI signals occur in the range from 20kHz to 30MHz; some of them only last for a instant and are difficult to capture, that is, a few microseconds; they need much calibration as they are much influenced by the surrounding environment (e.g. the voltage noise in neighbouring houses).

2.1.3.2.4 Load signature identification

We dive into NILM classification algorithms based on the load signature detailed in the previous subsection. They often have two processes: training or learning load signatures; and classifying. Regarding the training process they may be divided into two groups: Manual-Setup NILM (MS-NILM) or supervised NILM algorithms; and Automatic-Setup NILM (AS-NILM) or unsupervised NILM algorithms [Hart, 1992, Hassan et al., 2014].

The earliest studies on the field [Hart, 1992, Hart, 1985] use active and reactive power signatures (P and Q respectively) clusterization during the steady states where each cluster is an appliance. Its supervised classification algorithm is based on a dynamic clusterization where, at every sample, cluster are merged or divided according to a set of rules. Clustering of signatures is a recurrent technique, however, the number of cluster may be specified a-priori, which is not always possible. Likewise, a variant of the approach in [Hart, 1992, Hart, 1985] is proposed in [Weiss et al., 2012], by capturing P, Q and D delta values. The supervised labelling process of clusters is solved by receiving feedback from users through a smartphone interface.

Moreover, a Neural Network (NN)-based approach is used in [Srinivasan et al., 2006] to learn and classify harmonics in the current waveform. NN-based classification models such as multilayer perceptron (MLP), radial basis function (RBF) network, and support vector machine (SVM), are evaluated on a data set of 8 appliances tested in the laboratory. MLP and SVM are the best signature identification, being MLP better because its low computational cost. The accuracy is high (around 99.9%) but it rapidly drops when non-ideal conditions are introduced (i.e. for some appliances it decreases down to 60%). Similarly in [Hassan et al., 2014], NN-based classification models are also used to evaluate the viability of V-I trajectory signature, since ANNs have high noise tolerance and non-linear functions to model distortions. The labelling process is carried out by a K-means algorithm and the performance of several disaggregation algorithms is evaluated: Artificial feed-forward Neural Network (ANN), hybrid neural network (ANN+EN), Support Vector Machine (SVM), and adaptive boost (AdaBoost); over two households in the REDD data set [Kolter and Johnson, 2011]. It shows satisfactory results when using Wave-Shape (WS) Features based on V-I trajectories, specially in the case of SVM and AdaBoost classification models (i.e. around 98% of precision). However, the result drastically decreases (around 78%) when Gaussian noise is added.

Furthermore, multi-signatures are used in [Liang et al., 2010a] where snapshots and delta form signatures are extracted from the current waveform CW, active and reactive power PQ, harmonics HAR, admittance waveform IAW and others. In this case, multi-algorithms are applied: Least Residue (LR), Integer Programming, Genetic algorithm and ANN. On the occurrence of a event, all the aforementioned signatures are extracted; each disaggregation algorithm chooses a candidate (i.e. an appliance class); and, finally a Committee Decision Mechanism (CDM) decides on the disaggregation based on the outcome from those multi-algorithms. Furthermore, three CDMs are presented: Max-

imum Likelihood Estimator (MLE), Least Unified Residue (LUR), and Most Common Occurrence (MCO). The results are presented for simulated data and the accuracy in the disaggregation is 92.7% for the MLE before adding noise to data. One of most recent studies ([Gao et al., 2015]) explores the feasibility of load identification for labelling plug-level collected data at real time. Therefore, it evaluates the computational cost of a number of load signatures and classification algorithms as well as their performance when the frequency is downsampled. Nine classifiers are evaluated: k-nearest-neighbors (KNN), Gaussian Naive Bayes (GNB), logistic regression classifier (LGC), SVM, linear/quadratic discriminant analysis (LDA/QDA), decision tree (DTree), random forest (RForest) and Adabost. The performance is evaluated over the PLAID data set, a real public dataset with more than 1074 appliances coming from 55 different houses [Gao et al., 2014]. The highest accuracy (i.e. 86.03%) is obtained for the RForest, combining all extracted LS.

Finally, EMI noise is used as signatures in [Patel et al., 2007, Gupta et al., 2010] to identify appliances. The former uses SVM to classify EMI transient noises produced by switching-on events that last a few microseconds. These EMI signatures depend on the mechanical connection of the appliance into the mains, hence, this approach can discriminate between similar light bulbs from different rooms. Its measure accuracy on test is between 85% - 90% of events correctly classified. The latter uses a KNN classifier with $k = 1$ to discriminate continuous EMI noise. This continuous noise depends on the appliance internal oscillator, which is unique for each appliance. The evaluated accuracy is 93.82%.

2.1.4 A new perspective for NILM

Since the breakthrough of the Smart Meters in mid-2010 (see Figure 2.2), NILM has really grabbed the attention of many well motivated researchers and the number of publications have bloomed to a hundred per year since 2013. We cannot say less about the industrial sector where NILM companies have also proliferated ([NILM-wiki, 2016]) and it has turned into a big business [Kelly and Knottenbelt, 2016, Parson, 2015a, Kelly, 2016]. However, NILM is still a small business for final customers (i.e. ordinary people living in a household) that do not find its profit. Jack Kelly ([Kelly and Knottenbelt, 2016, Kelly, 2016]) investigates the efficacy of energy feedback to costumers across twelve studies. It turned out that the electricity reduction after disaggregated feedback is up to 4.5% (i.e.

the case of "energy enthusiasts"). In contrast, the aggregated feedback counts up to 3% of electricity reduction. Hence, this arises a question: Is energy disaggregation actually valuable to regular costumers? The study ended up stating that "there is no robust evidence that disaggregated energy feedback is more effective than aggregated energy feedback". For sure, the 1.5% of energy-savings globally can make a big different for big companies as well as for the climate warning, however, does it for individuals? Energy-saving oriented vision of NILM fails to motivate users and, therefore, it does prevent it from wide-spreading across the population. Thereby, we must pursue the search of an added value that brings the interest of final customers. In [Kelly, 2016, Kelly and Knottenbelt, 2016], this added value may be an increasing concern about the climate change or a better real-time energy feedback with better recommendations. In this thesis, we contribute addressing this problem and proposing an activity monitoring system mostly oriented to supervise the elderly routine. To the best of our knowledge, the first use of disaggregation of smart meter data for health monitoring purposes was carried out in [Song et al., 2014]. They performed an iterative time-dependent hidden Markov model to disaggregate appliances based on a-priori knowledge of inhabitant's activities. Then, the disaggregation was carried out and every appliance was assigned to a certain activity that could be monitored.

Summarising the state of the art in NILM, Low-Frequency algorithms are well developed and they are energy-oriented. The disaggregation accuracy is biased and it improves in major appliances where the consumption is higher and states are easier to predict. However, in terms of human activity monitoring, the major appliances do not often reflect the human behaviour. Manually operated appliance events may be better detected from an event-based NILM approach at high sampling frequency where the disaggregation accuracy increases. However, it can be observed that there is not yet a predominant load signature or HF NILM algorithm that suitably performs over all houses and appliances. Therefore, finding a general energy disaggregation model and load signatures is still a challenge. As a weak point, HF NILM algorithms may be more difficult to integrate: they either need a specialised hardware to take the measurements or a modification on the Smart Meter firmware. Although, most LF NILM algorithms also need a device to gather measurements from the Smart Meter through the HAN, which may not be as accessible as one though. We have seen in [NILM-wiki, 2016] that there is already a number of companies offering High-Frequency resolution. Furthermore, the sampling rate could be dropped down to 1Hz if more measurements are provided. For instance, [Weiss

et al., 2012] was classified as a HF NILM algorithm because, to obtain PQD values, it is necessary to process the voltage and current waveform at high frequencies. However, it uses PQD samples every second. Hence, more event-based NILM algorithms could be run at low sampling frequency giving high disaggregation accuracy if the new smart meters and third-party companies provide more measurements at low frequency.

2.1.5 Resources for NILM

This small subsection points out to the important resources (i.e. hardware and software resources) to develop any NILM project. The purpose is to orientate, given a few examples, more than realising a survey, which is out of the scope.

2.1.5.1 Hardware resources

We refer to **smart-plugs** as the devices that enable the monitoring of individual appliances. The smart-plug is directly plugged to a socket and the appliance is plugged to it. Then, several measurements are taken (e.g. active and reactive power, current, voltage) and are wirelessly sent to a concentrator or a in-home display. Then, they may be sent to the cloud by an internet connection. Smart-plugs are typically used in NILM dataset to gather ground truth information about appliances and for training and labelling tasks of some of the algorithms presented in 2.1.3. There are a wide offer in the market.

For instance, the Smart-me company implements smart-plugs that can take several measurements: energy demand, active power, voltage, current, power factor and temperature; every second. The devices connect with a router via WiFi and send their measurements to a website. Then, customers can access to the site to check the real-time consumption and power load curves. Further, they can turn appliances on and off remotely, see Figure 2.12.

Then, the **Smart Meter** are the aim of the NILM community. They provide the aggregated data to be disaggregated. However, regulations are different for each country as well as specifications. Although, the trend is to replace electromechanical meters by new smart meters by 2020. Typically, the smart meter sensors can measure the current and voltage up to 8kHz, although this depends much on vendors: some measure up to 1kHz,

others up to 1MHz. The smart meters may communicate with central stations with a sample interval of 15 to 30 minutes, through PLC, although this does depend on the country regulations. The sampling rate is higher within the HAN (i.e. 0.1Hz or 1Hz), however, not many countries implement HAN. For instance, Spanish and UK regulations are following discussed.

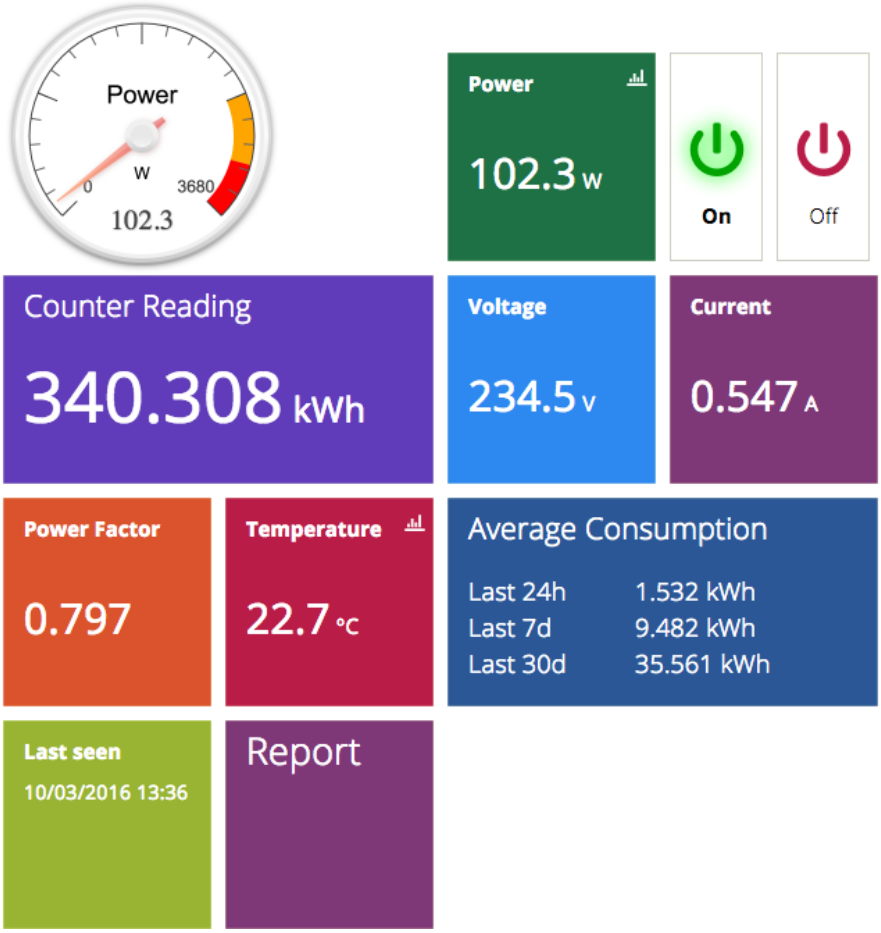


Figure 2.12: Parameters measured by an Smart-me smart plug.

In Spain, the Royal Decree and Order (i.e. RD1634/2006 and ORDEN ITC/3860/2007) establishes a meter substitution plan and the obligation to install new smart meters to all consumers under 15kW by 2018. Spain has 46 million inhabitants and approximately 26 million electricity customers. By 2013, Endesa replaced 30% of meters (i.e. 3.5 million customers). Although it does not exist standards about communication protocols, the communication of the smart meter with the central system is often divided into two stages: the communication from the smart meter to a concentrator and, then, from the

concentrator to the central office. The former is carried out by a PLC communication protocol (i.e. PoweRline Intelligent Metering Evolution, PRIME) and it allows a data rate transmission of 4.4 kbps, which allows a typical 15-minute load curves from smart meters [Wikipedia, 2016b, Vadacchino et al., 2013, Casellas et al., 2010, Sendin et al., 2012]. Regarding the communications with the HAN, a port should be enabled to communicate through Zigbee, WiFi or 6LoWPAN [MINETAD, 2011]; however, to the best of our knowledge, Spanish smart meters (e.g. from Endesa and Iberdrola) do not implement this port to reduce costs. Hence, LF NILM algorithms cannot take advantage of it.

Moreover, smart metering technical specifications are better regulated in UK where the most ambitious plan of smart meter roll-out is considered to be taken: all mechanical meters shall be replaced by smart meters by 2020 (i.e. 30 million homes and 2 million small business). The Department of Energy & Climate Change (DECC) regulates the smart metering specifications: each thirty minutes, smart meters shall be capable of measuring active and reactive power; also, active power shall be accessed every 10 seconds through HAN (typically Zigbee) [DECC, 2014].

Manufactures and companies that roll out Smart Meters in many countries do provide little access to the Smart Meter resources. Consequently, we can find in the market companies that offer metering solutions as well as NILM services [NILM-wiki, 2016]. The sampling rate ranges from 1 Hz (i.e. oriented to households) up to 100 kHz (e.g. the case of Green running, which is most oriented to business such as hotels and supermarkets). Furthermore, there are solutions such as Bidgely and Onzo that rely on the smart meter data sent through the PLC to perform a coarse-grained disaggregation. Other companies, such as Discovergy in Germany and Informetics in Japan, are reaching agreements with manufactures to get access to the full potential of Smart Meters at 8kHz and 100 MHz, respectively.

2.1.5.2 Open Software

Within the NILM community, there are two important framework that allows researches to evaluate their own NILM algorithms and compare them with others, using the same datasets and same metrics: one is implemented in Python, the NILMTK [Batra et al., 2014]; whereas the other is implemented in Matlab, The NILM-eval [Beckel et al., 2014, Cicchetti, 2014].

2.2 Activity Monitoring in Elderlies

2.2.1 Sustainability challenge in modern healthcare systems

Life expectancy is becoming higher and higher every year in most developed countries. Although this is clearly a benefit, there are also potential challenges to be faced. For instance, the old-age dependency ratio, the ratio between elderly people and those in working age, is rising and expected to be double within the next 35 years (from 11.7% to 25.4% of population). This ratio will exceed the 50% in countries such as Japan, Germany, Italy, Spain and Poland; and the elderlies ageing 80 or over will be triple in European Union and United States. In fact, this is a challenge on the sustainability of healthcare models that need to care this community. Furthermore, several studies reveal that the majority of elderlies, at 90%, tries to live independently in their own homes, and a shortage of caregivers is expected. Therefore, these two factors, ageing of the population and need of monitoring within the households, are the key issues that have motivated many studies in Ambient Assisted Living (AAL) domain [Debes et al., 2016, Muszyńska and Rau, 2012]. AAL is based on Ambient Intelligence (AmI) and it pursues to care and monitor people, by adapting their environment and using new technologies.

For instance, Figure 2.13 depicts the singular case of ageing in Spain where, for the first time in history, the community of elderlies (i.e. over 65 years old) has overtaken in numbers the community of infants (i.e. under 14 years old). Further, the number of elderly dependants has radically increased and the 71.4% of them receive their principal attention from the family, whose concerns on AAL technologies to monitor their relatives have increased. In other countries, they are already experimenting shortages on the public national health care services. Thus, one million people in the UK (one-third of the ageing population in need) struggle in their daily routines because the National Health Service (NHS) fails to meet their demands [Triggle, 2015].

In summary, modern healthcare systems face serious sustainability issues due to the rapid increase of the ageing population and this is more acute nowadays as this community is reticent to leave their houses. AAL can contribute much by developing monitoring strategies to serve this huge population within a cost-effective solution. Therefore, this Thesis explores the scalability of healthcare monitoring systems to serve this increasing community.

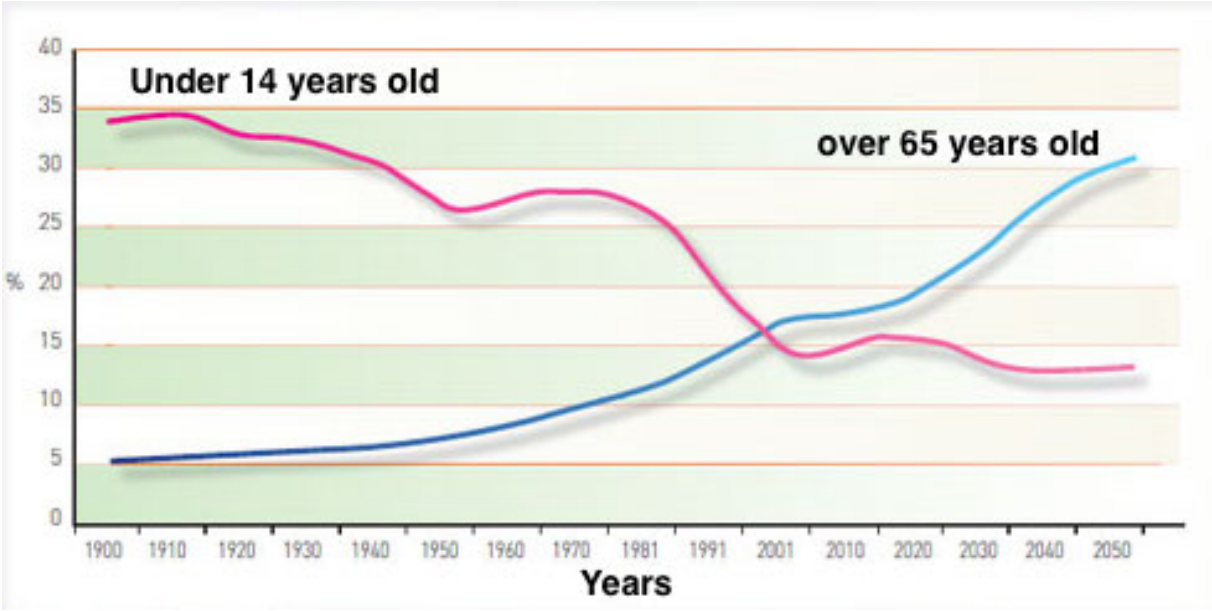


Figure 2.13: Inversion of the population in Spain [Patricio, 2011, INE, 2004].

2.2.2 Healthcare monitoring systems

The state of the art can be organised according to several factors: the type of sensor (e.g. wearables, ambient sensors), the monitoring source (e.g. physiological signals, movements or activities of daily livings ADL), the type of monitoring (e.g. monitoring the health deterioration at long term or producing alerts for short-term intervention), etc. Nevertheless, we divide the state of the art into direct and indirect methods. The formers include all approaches where the health is directly diagnosed through the evaluation of health parameters, whereas the laters comprise those methods where the health is inferred through the readings of an indirect parameter that may reflects the health status.

In pursuit of providing a better understanding from the frame of this thesis (see Section 2.2.1), we depict the state of the art in Figure 2.14. In the graph, the different studies, classified into direct and indirect methods, have been placed according to accuracy and scalability. We refer to accuracy as how reliable our model is in order to make a diagnostic. Thus, approaches based on physiological parameters are more reliables than those based on movements. Moreover, the latter are themselves more precise than activities. In turn, ADL approaches, which are more detailed in the following sections, may be oriented to create alerts if someone has deviated from his daily routine, or to follow someone's health

deterioration across a extended interval (e.g. months). We may consider short-term outcomes (i.e. alerts) more accurate than long-term monitoring in this study as they may be more critical for the health, even though they might not be more accurate.

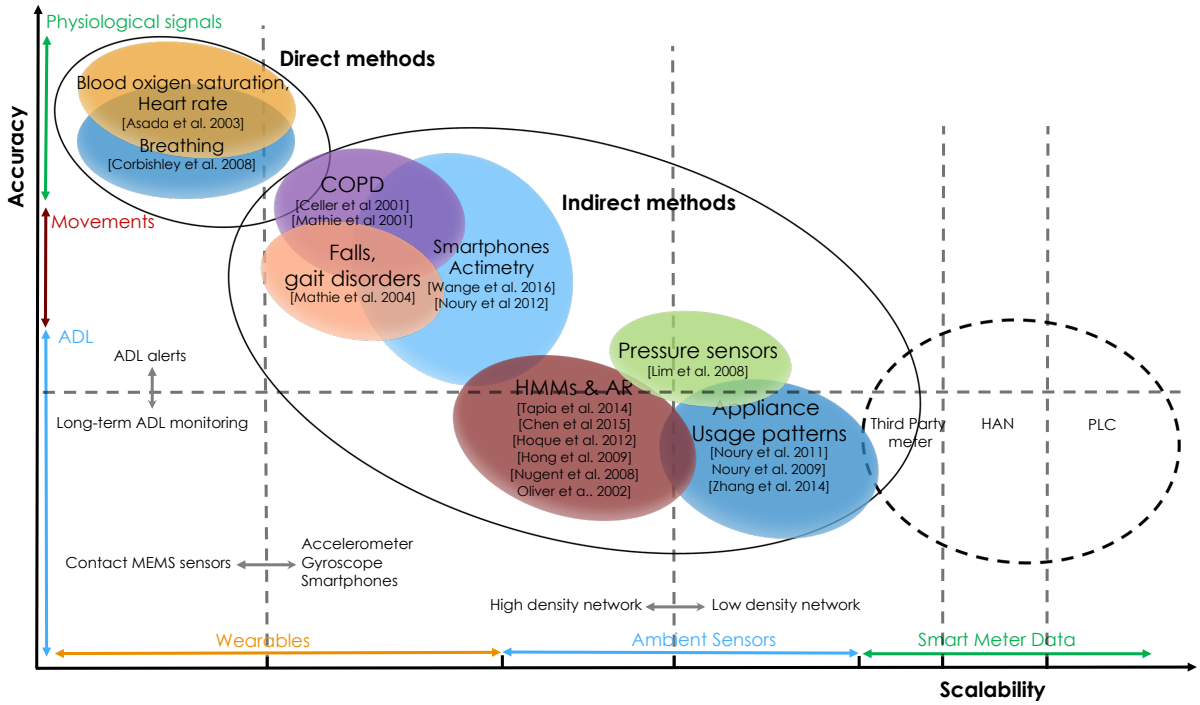


Figure 2.14: State of the Art for Activity monitoring in elderlies.

Regarding the scalability, it depends on several variables: the degree of acceptance, the degree of intrusiveness and the deployment and maintenance costs. In the highest scale of intrusiveness, non-acceptance and cost, there are wearable sensors that must be in contact with the body to measure vital signals, followed by those less intrusive that only need to be carried. Then, there are ambient sensors that are better accepted and have a lower cost. Some approaches rely on a high density WSN to sense as much as they can inside the house, whereas others pursue a low density WSN to achieve a better scalability. As we already did in Section 2.1.3, Figure 2.14 reflects our subjective view of the state of the art, based on accuracy and scalability concepts that we define. It serves to our purpose of structuring the survey. Furthermore, Figure 2.14 depicts a new type of data (i.e. smart meter data) with a even higher scalability than ambient sensors, which still do not have state of the art. This thesis aims to explore this new scalability to expand the state of the art in healthcare monitoring systems, which will be properly discussed at the end of this Chapter.

Then, we present a survey of the background in direct and indirect healthcare monitoring methods.

2.2.2.1 Direct methods

We refer to direct monitoring methods as those where the health is directly inferred from biomechanical and physiological measurements that are taken through the use of wearable sensors and Body Sensor Networks (BSN). Physiological signals may be blood pressure, blood oxygen saturation, heart rate and muscle activity [Patel et al., 2012]. These are valuable information and sufficient to make diagnostics, hence, these methods are very accurate. However, they might be very intrusive, which means low acceptance, and more expensive than indirect methods.

During the last decade, big steps in sensor miniaturisation and Microelectromechanical Systems (MEMS) have made feasible to embed sensors in clothing at a relative low price. This makes direct methods more attractive, however, they still need to be worn almost continuously and be placed close to the skin. In fact, one of the big issues is the motion artefact produced by a bad attachment of the sensors to the subject. For instance, in [Asada et al., 2003], sensors to measure the blood oxygen saturation and the heart rate have been embedded into a ring to be worn in a finger. The continuous monitoring of these variables can prevent illness such as hypertension and congestive heart failure. Furthermore, techniques to reduce the motion artefact are implemented. Likewise, in [Corbishley and Rodriguez-Villegas, 2008] the author proposes the use of a embedded microphone to be placed close to the neck to record acoustic signals related to the breathing. Using signal processing techniques artefact signals are removed and the envelope (which carries the breathing information) is kept. He also developed an algorithm for apnea detection based on this sensor. Other challenges in direct monitoring, apart from the cost, may be the reduction of the size in transmitters and receivers and the power consumption [Patel et al., 2012]

Direct methods may be of great help for patients that need to be strongly monitored at all times, however, this is not the case of the majority and these approaches are barely scalable. Consequently, it is out of the scope of this thesis whose purpose is to devise an scalable health monitoring system to sustain the ageing of the population.

2.2.2.2 Indirect methods

We call indirect monitoring methods those that pursue to infer the health through the inference of other parameters. For instance, the movement can be a good indicator of the health: someone that moves freely along the house might be in better health that other who restrains their movements to one or two rooms. Further, the activity in home plays a key role and it may contribute to reduce readmissions in hospitals and early detections of problems. A long-term monitoring is also beneficial as, if someone becomes to struggle to perform their daily's task, it might be an indicator of health deterioration [Scanaill et al., 2006]. Likewise, a commonly used concept in activity monitoring oriented to health care models is the term of Activities of Daily Livings (ADLs or ADL).

ADL refers to the daily activities and routines carried out by people for self-caring. This concept was first proposed by Sidney Katz in 1950 who stated that age-related diseases have a direct impact on ADLs. Few years later, he created ADLs Indexes to measure the level-dependence of a person in 1963 ([Katz et al., 1963]). In this work, he evaluated six different daily tasks: bathing, dressing, going to toilet, transferring (e.g. moving in/out of bed), continence and feeding; and based on them, he scored the someone's ADL performance in grades: A, B, C, D, E, F, G, or other; being A the maximum degree of independency. A more recent study completed the original Katz ADL and called it Instrumental Activities of Daily Living (IADLs) [LaPlante, 2010]. They proposed more complex activities than ADLs such as: doing light housework, using the phone, managing medication and money, etc; and the performance is scored based on different coefficients such as the Guttman coefficient (CS): CS values lower than 0.6 means that the individuals are dependant whereas higher values give independency, being the maximum value 1. Before the use of sensors, the evaluation of ADL was cumbersome and it was based on surveys done to each person. Nowadays, the task has been automated thanks to the technology and also ADL monitoring (i.e. the continuous evaluation of ADL over time) can be performed. Activity is a complex concept and it does not have a standard of how to measure it. Hence, we generalise the ADL monitoring concept to any performed task, through which the health can be inferred and someone's independency can be assessed. From [Debes et al., 2016], we can describe the main challenges in ADL monitoring: quality, how accurate and valuable the measurements are; cost, including the deployment and maintenance; acceptance, how intrusive the system is and its adaption to the environment; and privacy. See also Figure 2.15.

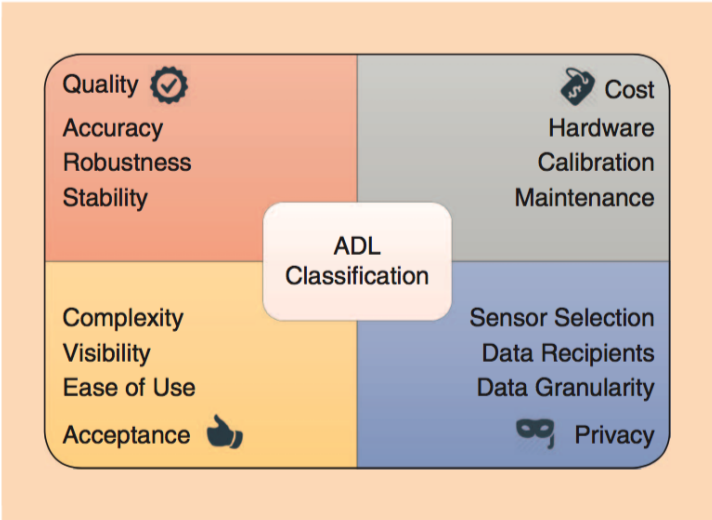


Figure 2.15: ADL challenges, in [Debes et al., 2016].

Concerning the involved sensors, there is a mixture of wearables and ambient sensors aiming to capture the human activity within a household. Wearables sensors may be more precise in registering activities related to movements, such getting up, sitting and going upstairs; however, they may fail in registering others such as watching TV or cooking, which are better obtained by ambient sensors. Therefore, there are many solutions that combine both approaches. Typically, heterogeneous sensor networks are deployed avoiding CCTV cameras and other invasive systems that could threaten the privacy of householders. Further, it is worth noting that ambient sensors are less intrusive than wearables and, consequently, better accepted within the community, whereas wearables can provide a better accuracy on someone's movements, see Figure 2.16.

Some wearable sensors may fit better than others depending on the activity nature. For instance, PDR can well register dynamic activities such as walking, however, they cannot record static activities such as postures. In contrast, accelerometers and gyroscopes can register both dynamic and static activities [Scanaill et al., 2006]. In [Mathie et al., 2004], accelerometers and gyroscopes are used to detect falls and gait disorders. Furthermore, the use of these sensors helps to detect a good number of pathological conditions such as COPD (Chronic Obstructive Pulmonary Disease) [Celler et al., 2001, Mathie et al., 2001].

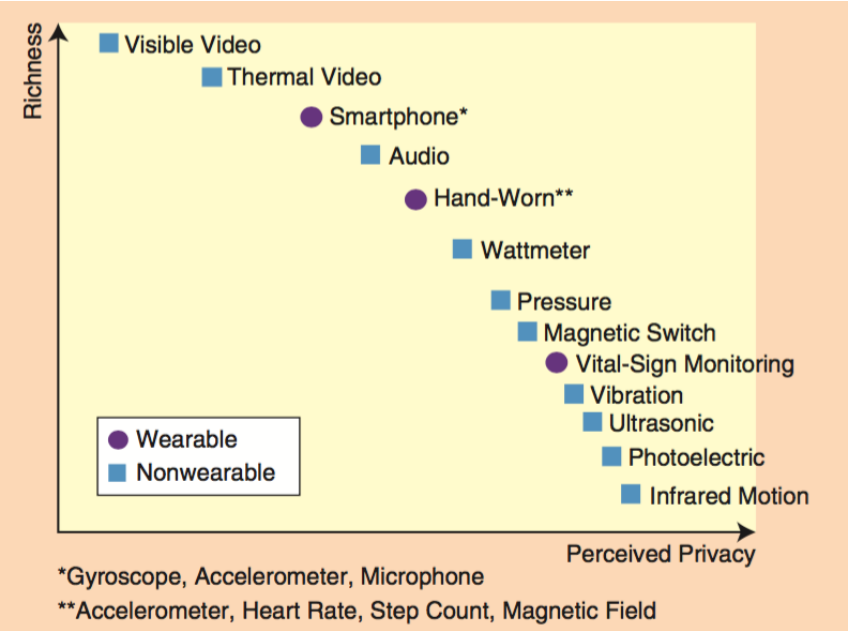


Figure 2.16: Valuable information against intrusiveness, in [Debes et al., 2016].

Wearables may enable to monitor the subject outside of the covered zone (i.e. the household) thanks to the use of data loggers. However, this has cons: the data analysis is only available once he has returned to the covered zones. Some wearables embed a processing unit inside to enable the real-time diagnostic, but this increases the power consumption and, consequently, the autonomy. Thus, wearables systems often use transmitters to send the data to an external processing unit to increase the autonomy and enable real-time diagnostic whether the subject does not need to be monitored outside [Scanaill et al., 2006]. Nowadays, thanks to the widespread adoption of smartphones, wearables can rely on them as external processing unit and, therefore, it enables the real-time diagnostic outside of the covered zone without detriment to the autonomy. In fact, smartphones are taking over wearables based on accelerometers and gyroscopes, because they are sufficient to detect most of the aforementioned activities thanks to their embedded inertial-unit [Wang et al., 2016]. Similarly, in [Noury et al., 2012] the actimetry of a subject (i.e. the quantity of movement per minute during sleep and asleep cycles) can be measured, which is directly correlated to homeostasis.

Regarding ambient sensors, a number of heterogeneous sensors is placed in the house creating a WSN to infer the human activity through its interaction with them. For instance, they are pressure sensors, infrared sensors, sound sensors, magnetic switches, optical and

ultrasonic sensors [Debes et al., 2016, Scanaill et al., 2006, Kautz et al., 2003]. They are better accepted because they do not have to be carried and, consequently, are less intrusive. Nevertheless, their accuracy depends on the number of deployed sensors, meaning that, if we want more accuracy, there will be more cost, less privacy and less acceptance. Besides, inferring the activity through the measurement of several heterogeneous sensors is a more complex process; and it can only be determined that the subject performed the activity if he is living along. For instance, in [Lim et al., 2008] a network of pressure sensors are placed in furniture and floors to detect 5 main activities: meal, sleep, excretion, go-out and rest. Based on this method, it can be reported to the family when the subject carries unhealthy habits such as skipping meals.

In most ambient sensor approaches, they require an Activity Recognition (AR) proceeding to group event sequences by activities [Massot et al., 2013]. Then, these activities are assigned to the major ADL activities and they are evaluated to get the ADLs Indexes, also known as Circadian Rhythms. The composite of major Activities based on minor activities, as well as their tracking, has been studied in [Rashidi et al., 2011, Kakde and Gulhane, 2015]. In [Tapia et al., 2004] a "tape and forget" method to deploy as many sensors as possible inside the household is presented. These sensors have been particularly designed to adapt themselves to many types of furnitures, items or objects inside the house using a small magnet (see Figure 2.17). The labelling process is huge, therefore, they use unsupervised methods as Hierarchical Hidden semi-Markov Models (HHSMMs) to avoid it. They can detect ADL and IADL activities with an accuracy that ranges from 25% to 89%. However the computational cost of using HHSMMs with so many observed variables is high. This is a suitable method to collect data for a data set but it is impractical in large deployments due to the quantity of sensors per house. Moreover, activities can be overlapped over time, which makes the AR clusterization process more difficult. The use of supervised algorithms such as Support Vector Machine (SVM) can help to group event sequences corresponding to ADL even if they are overlapped [Chen et al., 2015]. However, the labelling process can be tedious. Other relevant studies dealing with overlapped activities are [Hoque and Stankovic, 2012], where probabilistic models such as Naïve Bayesian (NB), Hidden Markov Model (HMM) and Hidden Semi Markov Model (HSMM) are adopted; and [Hong et al., 2009, Nugent et al., 2008] where sensor fusion techniques such as the Dempster-Shafer Theory are used. Determining the duration of an activity is also a highly complex task; in [Oliver et al., 2002] Layered Hidden Markov Models (LHMMs) are used to model the different durations of activities.

All the aforementioned methods based on AR are precise about inferring the classical ADL, which is suitable for long-term deployments, but they do not perform alerts for short-term monitoring as in [Lim et al., 2008].

Figure 2.17: Some of the 77 "tape of forget" sensors deployed in a house in [Tapia et al., 2004].

Overall, ambient sensor approaches based on AR and on heterogeneous high-density sensor networks present three major issues that lead to a low implementation in real scenarios [Massot et al., 2013]: overlapped activities, activity heterogeneous duration, and the deployment of a complex WSN, which, in most cases, is intrusive. Furthermore, this approaches need a supervised training process for each individual household: the sensor activation sequence may be different in order, duration and nature (e.g. they might use different appliances type) and it must be manually labelled as an activity. In [Noury et al., 2011, Noury et al., 2009] a new approach of monitoring ADLs through electrical signatures of appliances captured by plug-meters is proposed. This monitoring method has been well accepted by subjects, as well as by professionals, during experiments since it is less intrusive. Besides, daily activities are strongly connected to the usage pattern of appliances and, consequently, they are a good way to infer the human activity. Electrical events are mapped over Daily Activities using the classifier k-means neighbours, and ADL Indexes and Circadian Rhythms are obtained and presented as a decision support tool

for experts. Three main activities are monitored: food preparation and eating, hygiene and elimination. Although the author states that his approach is prone to be deployed massively, there are many labelling tasks such as the weight of appliances on the activity, which is different depending on the individual, and finding the activity duration is still an issue. Similarly, the appliance usage patterns are again proved to be correlated with ADLs [Zhang et al., 2014], where a method known as Latent Dirichlet Allocation (LDA) for text analysis is used to map appliance events with ADLs. Thus, appliance usage patterns seem to be a rich source of information about someone's activity, which allows to use a lower sensor density, lowering hardware-cost and complexity; whereas their monitoring is well accepted, enabling large deployment. However, there are still issues when trying to infer the major ADL because the overlapping and the complex heterogeneous duration. Besides, these methods are thought to be deployed for a long-term monitoring to measure the health deterioration, however, they do not perform alerts for a short-term monitoring of the health. Moreover, there are works where only the appliance usage pattern is used, instead of inferring ADLs. Thus, the monitoring of the kettle or the TV set has been considered as a relevant variable to detect changes in routines [Fleury et al., 2010], whereas in [Nikamalfard et al., 2012] the usage pattern of kettle and fridge during nighttime is used to detect sleep disorders, characteristic of Dementia.

2.2.3 Enabling massive deployments in modern healthcare monitoring systems

Summarising, our aim is to develop sustainable healthcare monitoring systems to support an increasing ageing population. After our survey of the state of the art, we found the direct monitoring methods and indirect monitoring methods.

Direct monitoring methods rely on wearable to take physiological measurements of the subjects that can lead to a high-accuracy diagnostic. However, only a few subjects within the community need that intensive care and these methods are expensive and not well accepted. Then, we discard these approaches as solutions for our concerning.

Within the indirect monitoring, there are also wearables that are less intrusive and expensive. They measure the movement and can trigger alarms when falling and monitor gait disorders. Special attention is drawn in smartphones, whose widespread adoption can be a factor to be considered in massive deployments. However, they are not so popular

within the elderly community; neither are carried at all times and, therefore, its capability to detect disorders is limited. Moreover, ambient sensors seem to be well accepted by subjects and they have a low cost. To achieve a better scalability, less sensors need to be deployed. Likewise, approaches that rely on low density WSN seek for a source rich in information. Studies have proved that appliance usage patterns are such source of information and someone's ADL can be well inferred through them. Furthermore, appliance monitoring seems to be well accepted both by subjects and doctors [Noury et al., 2011, Noury et al., 2009, Zhang et al., 2014].

Consequently, all indicates that we should focus our efforts on this last type of monitoring to overcome the sustainability issues and meet the challenges proposed in this thesis. Further, recognising ADL by using an indirect method and a low density of sensors still seems to be an issue: overlapping activities and heterogeneous duration of activities. We will further investigate alternatives to the classical activity recognition of ADL that simplify the process and also enable the long-term monitoring as well as short-term one. Thus, the dashed circle in Figure 2.14 indicates where we will contribute in the state of the art. We present a more detailed discussion about our solution in Chapter 5.

Chapter 3

General overview of the proposal

Although a short one, this Chapter is an essential component on this thesis. It interweaves the previous documentation work with the research itself and, hence, it can be placed nowhere else but here, halfway in the reading.

Along Chapter 1, we have stated our motivation: devise new solutions of activity monitoring for elderlies based on NILM that help overcome the new challenges for the forthcoming years. Based on that, a thorough survey of the state of the art has been conducted, Chapter 2, in order to identify the potential innovations that we may contribute with. Then, we identify the requirement to meet our Objectives, Section 3.1, and, from there, we propose our solutions, Section 3.2.

3.1 Requirements

The conducted survey on NILM algorithms presented in Chapter 2, points to the fact that a new NILM algorithm needs to be devised to release the full potential of energy disaggregation for activity monitoring. Therefore, we propose the development of a new NILM algorithm with the following features:

- The disaggregation must be done in **real time** to enable activity monitoring.
- Therefore, our algorithm must be **event-based.**

- The **sampling rate** must be higher enough to allow event detection but lower enough to enable embedding in Smart Meters and third-party devices. The typical internal sampling rate in Smart Meters is $1Hz < f_s < 8kHz$ (see Subsection 2.1.5), hence our proposal must be validated taking into account this constrain.
- To allow scalability, our NILM algorithm shall be **unsupervised**. The definition of unsupervised in NILM may differ from other machine learning fields and it is defined in Section 2.1.3: the algorithm may be trained with unlabelled data plus labelled data from different houses than the test.
- Also, in alignment to the aforementioned point, the **Load Signature must be per appliance type** and generalised along different houses.

Regarding the activity monitoring part, our solution must meet the following requirements:

- To support this increasing community of elderly residents, it is clear that it has to be a large amount of **resource savings**. Likewise, it shall be **scalable, non-intrusive and of low cost**; therefore, it will be based on the outcome of a NILM system.
- Short-term monitoring must be enabled to **trigger alarms** in case of an urgency. Although, this **alarms must be of very low cost** as the accuracy is quite low comparing with approaches based on BSN.
- Long-term monitoring must be enabled to **score the daily routine** and follow degenerative patterns over time.
- As discussed in Section 2.2, using Activity Recognition for ADL tracking might bring some scalability and robustness problems. First of all, AR are accurate only if a heterogeneous sensor network is deployed, which is not our case: we only have a single-point sensor to measure current and voltage. Furthermore, AR requires an supervised training process for each specific house: sensor activation sequences must be matched to a particular activity and this is a manual labelling process. In fact, sensor activation sequences may be different in order, duration and nature (e.g. they might use different appliance types). This conflicts with the requirement above of unsupervised approach. Also, issues such as overlapping activities or heterogeneous

duration are not yet solved. To give scalability and robustness to our approach, we shall discard AR methods and devise **generalise methods based on appliance usage pattern.**

- Our solution must follow **strong usage patterns** and **erratic usage patterns.**
- Also, related to the point above, **the uncertainty and sparsity of the appliance usage pattern must be modelled.**

3.2 Proposals

Based on the above requirements, Chapter 4 delves into the low-level proposal (i.e. our NILM algorithm). Moreover, three different alternatives are presented in Chapter 5 for the high-level application (i.e. activity monitoring). Each Chapter is self-contained: they present the implementation, experimental results and conclusions. Besides, it is worth noting that the three activity monitoring approaches work independently and can be implemented together or separately. Then, Figure 3.1 depicts an overview of the proposals and a potential implementation.

The proposed NILM algorithm is event-based and it has been validated with datasets that have signals with a maximum bandwidth of 1kHz, which is low enough to be implemented in smart meters and third-party devices. **The event detector** is a signal processing approach that extracts the envelope based on the current root mean square value every fundamental period. Furthermore, we introduced a pre-processing block to shield the signal against voltage drops and it enables the generalization along different households. Then, we use a peak detection block to identify the occurrences of events (i.e. appliance switching-on events). To infer the human activity inside a house, we need to extract the usage pattern from manually operated appliances. This is the reason why our NILM algorithm only disaggregates switching-on events: vampire loads and switching-off events may be automated and programmed and they do not reflect the human activity. Later on the activity monitoring process, automatic switching-on events (e.g. fridge cooling cycles) will be filtered out. Thus, the event detector outcome is the timestamp at the occurrence of an event.

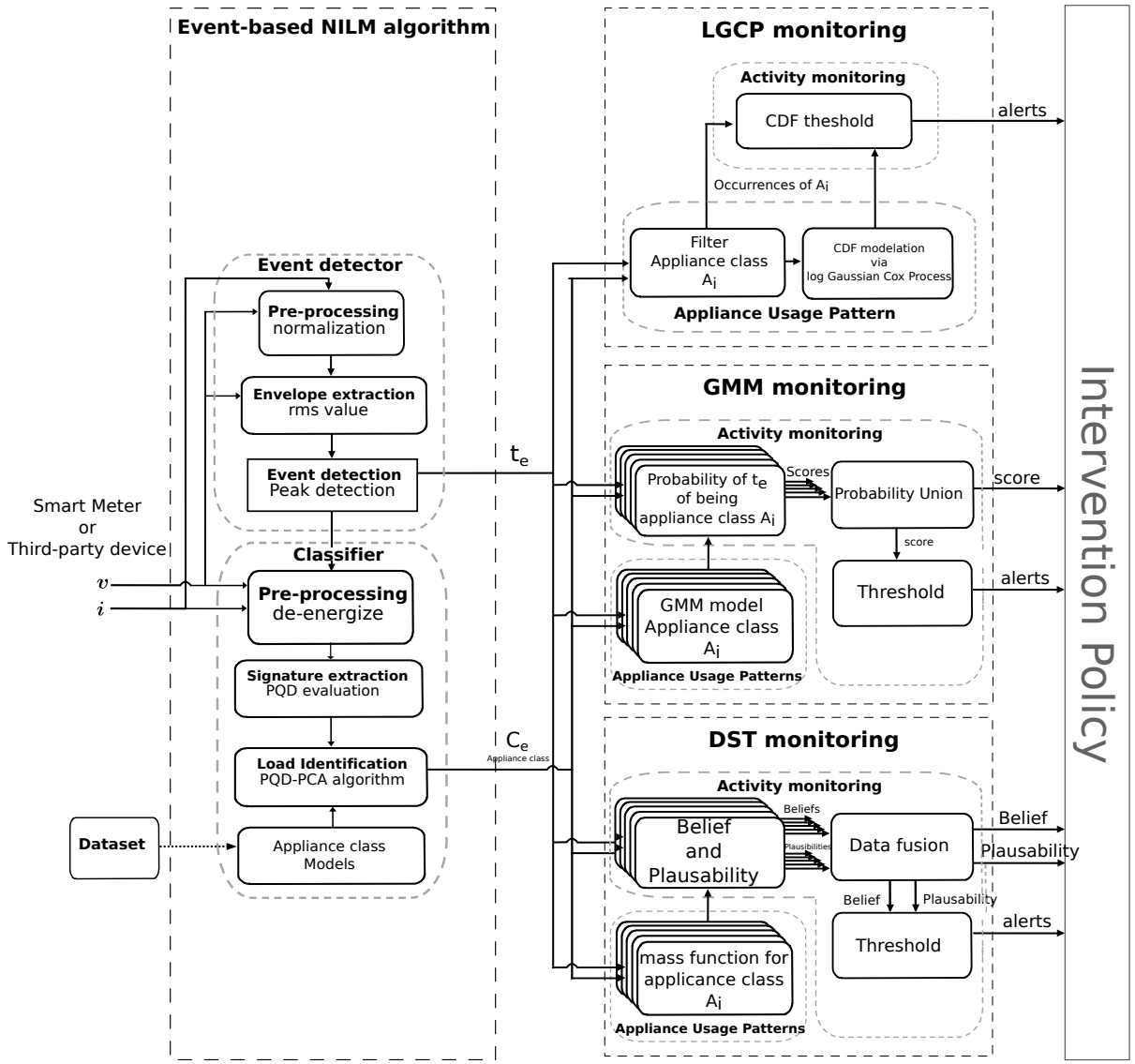


Figure 3.1: Block diagram of the proposals and workflow.

Whenever a timestamp is received from the event-detector, **the classifier** extracts the voltage and the current and, during the pre-processing block, de-energises the current signal (i.e. it isolates the quantity of the current signal that is due to the appliance producing the event). This is equivalent to the delta-values explained in Section 2.1. From there, the active, reactive and distortion power trajectories (PQD) are evaluated within a time window. Finally, using Principal Component Analysis (PCA) and based on the general models of appliance types (i.e. appliance classes), the event is labelled based on the model that most compresses the signature (this will fully discussed in Subsection

4.1.2.2). One important requirement was the unsupervised training: our general models are pre-loaded on the system and may be trained offline in an independent dataset. From time to time, general models will be updated; however, no training process is required in the system. This meets our requirements.

On a high-level stage, our activity monitoring proposals receive from a NILM algorithm the event occurrence and its label (i.e. appliance type). Thus, scalability, non-intrusiveness and low cost are enabled. Then, there is a training process based on these occurrences. This time, the training is done on the system and with labelled data from the same house. Each individual performs their daily routine differently and, hence, his/her particular routine must be learned. However, this does not affect the scalability as no labelling process is needed because this has been already performed by the NILM algorithm. It only needs an training period before starting the monitoring process. From our experience, one may have a strong correlated routine, using one specific appliance. Therefore, we take advantage of that, and trigger alarms when someone has departed from his normal routine. We present a **log Gaussian Cox Process approach (LGCP)** to monitor a single appliance, and to trigger alarms if that appliance has not been used by a time of the day. The novelty of this approach is that repetitive pattern (i.e. daily and weekly periodicity) and uncertainties can be successfully modelled by Cox Processes. In our studies, we found that elderlies in UK follow a strict routine using the kettle [Alcalá et al., 2015a].

Let assume that, in most cases, the human activity cannot be inferred only from a single appliances. Then, we also introduced two more approaches based on multi-appliance monitoring: the **Gaussian Mixture Model (GMM)** approach and the **Dempster-Shafer Theory (DST)** approach. Both approaches model the manually operated appliances based on the list of event occurrences and their labels. The former uses the Bayesian theory to score the use of appliances over the day, based on the GMM model. It merges the score to obtain a single score for the day. The use of a threshold enables to trigger an alarm if there has been a deviation in the normal daily pattern. Therefore, this approach provides long-term and medium-term (i.e. a day) monitoring. It was devised as a first approach to multi-appliance monitoring [Alcalá et al., 2015b]. However, it does not model uncertainties and the performance is lower than the DST, drawing a larger number of false alarms. The Dempster-Shafer theory is an generalization of the Bayesian theory, widely applied in sensors and data fusion. The uncertainty is successfully modelled as the difference between the plausibility and the belief. The mass functions (equivalents to distribution function in the Bayesian theory) model the appliance usage patterns during

the training. Furthermore, DST defines the rule of combination to merge the score of each appliance on a general score of the activity performance. The time interval to obtain the score can be changed. For the experimental results obtained in Subsection 5.4.2.3, this time interval is six hours, which seems to be a good trade-off for enabling long-term and short-term monitoring.

Summarising the high-level proposals, the LGCP may be more robust and faster when strong routines are stablished using a single appliance. GMM also performs alerts and uses multi-appliance monitoring. Although it performs short-term and long-term monitoring, it has many false alarms. The DST approach successfully merges multi-appliance usage patterns and it performs short-term and long-term monitoring. The outcome of our proposals will be alerts, and scores of the performance with a certain time resolution (i.e. six hours or a day). It is worth noting that our proposals use very little information from the environment: only the smart meter data. Therefore, they may not have the accuracy of other well-defined healthcare monitoring systems (refer to Section 2.2). We have sacrificed accuracy in the pursue of scalability. Hence, the intervention policy (i.e. the action to be taken in case of emergencies) must be consistent. This intervention policy may well have a penalty of low profile: a call from a relative or healthcare worker to check in the status of the subject. This is more detailed in Section 5.4.3, and more details about all proposals are following in Chapter 4 and Chapter 5, from low-level to high-level application.

Chapter 4

Energy Disaggregation Algorithm Proposal (PCA-PQD disaggregation)

This chapter delves into the Energy Disaggregation algorithm proposed in this Thesis. As discussed in Chapter 2, one trend in the NILM domain is the pursue of eventless low sampling disaggregation algorithms (i.e. LF-NILM algorithms where the sample rate ranges from 0.1Hz to 1Hz) for enabling potential compatibility with Smart Meters into the HAN. Nevertheless, LF-NILM algorithms, as defined in the previous Chapter, does not detect events and are more appropriated for energy saving oriented applications. Thus, they may disaggregate better the major appliances than others. Moreover, event-based HF-NILM algorithms are more appropriate for real time activity recognition applications such as the one pursued in this manuscript. Thus, we propose a novel event-based NILM algorithm feeding on a feasible high sampling rate (i.e. low enough to enable its integration in Smart Meters and third-company devices) in order to detect and classify the highest number of events. This sampling rate also allows the use of signatures based on transient and steady states, which prevents from overlapping. These signatures are trajectories of the active, reactive and distortion powers (PQD). A Principal Component Analysis (PCA) transformation has been applied to learn those signatures, where a general model for every appliance type is obtained. This makes possible the installation of the proposed system in new houses with no further calibration or training process. Furthermore, the event detection is focused on the appliance switching-on events because it is when we can assure a human interaction (i.e. if it is manually operated appliance) and, hence, to infer the human behaviour using the proposals presented in Chapter 5.

Firstly, the model of the proposal is described in Section 4.1, followed by a thorough evaluation of the event detector and the disaggregation algorithm in Section 4.2 using two different datasets; finally, conclusions and potential future works are discussed in Section 4.3.

4.1 Model description

The proposed NILM algorithm is described hereinafter, split into two major blocks. In Section 4.1.1 the event detection process is detailed, whereas in Section 4.1.2 the classification algorithm performing the disaggregation is explained.

4.1.1 Event Detection Algorithm

The main contribution by the proposed event-detection algorithm is the detection of the highest number of switching-on events. Those events will be later classified by the disaggregation algorithm in Section 4.1.2. Furthermore, it has a high performance without a training process required. This proposal is based on the detection of changes in the envelope of the acquired signal for the electrical intensity (current signal hereinafter). Analyzing this signal over time instead of its corresponding power has the advantage of discarding network noise due to voltage drops and other effects, as the electrical current is the immediate cause of appliance activities and power also includes voltage.

Following the block diagram depicted in Figure 4.1, the process of the event detection is based on three stages: pre-processing, change detection and peak detection. Furthermore, Figure 4.2 illustrates a case of detection to facilitate the explanation.

4.1.1.1 Pre-processing

The main purpose of the pre-processing stage is to isolate the current signal produced by the transient states from the rest (i.e steady states, noises, 50/60 Hz component, etc.). Therefore, we average the root mean square (rms) value from the current signal $\boldsymbol{i}$ every cycle of the utility frequency for filtering the 50 or 60Hz component and other information

or noise, keeping only the transient states information. Using this block, it is possible to extract that information and to the signal against skews in frequency and drops in voltage.

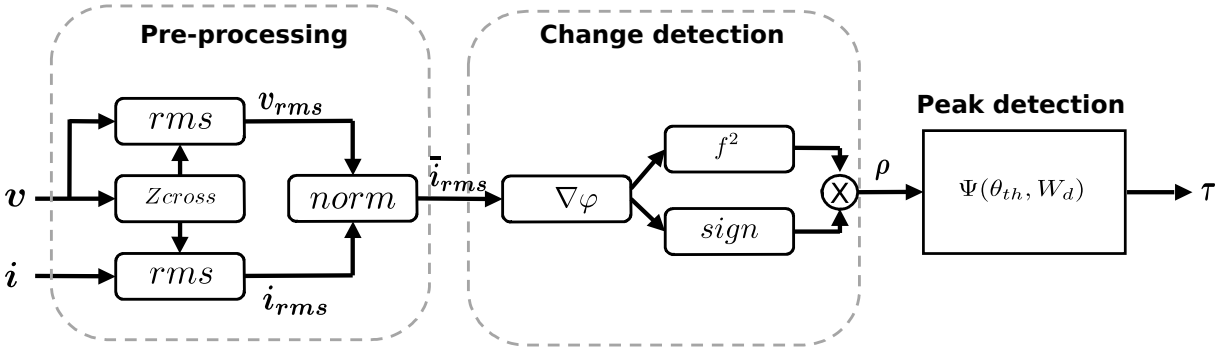


Figure 4.1: Block diagram for the proposed event detector

Let us consider the inputs described in (4.1) and (4.2) as the voltage and current signals from the mains, $\boldsymbol{v}$ and $\boldsymbol{i}$ respectively, where N is the number of samples over the monitoring time. The event-detector is a real-time algorithm, thus, the length for N does not influence on the performance and the monitoring time could be from minutes to weeks. However, for the sake of computation, results on Subsection 4.2.1 are evaluated through a windowing of about 2 minutes. Figure 4.2(a) depicts the current waveform $\boldsymbol{i}$ measured at a certain time from the mains in the BLUED data set (see Subsection 4.2 and Appendix A). A zooming window of 14 seconds is applied to facilitate the visualisation of graphs. The voltage waveform $\boldsymbol{v}$ has not been plotted since it has very little changes and does not contribute to understand the process. It is a 60Hz sinusoidal signal of 170V of amplitude (i.e. 120Vrms), that may have small amplitude variations due to voltage drops and a phase shift with respect to the current $\boldsymbol{i}$. Moreover, four events (ground-truth) have been marked with vertical dashed lines. These events correspond to state changes in one appliance, in this case the refrigerator.

$$\boldsymbol{v} = [v_0, v_1, \ldots, v_{N-1}] \tag{4.1}$$

$$\boldsymbol{i} = [i_0, i_1, \ldots, i_{N-1}] \tag{4.2}$$

The rms value of the signal is evaluated at 60Hz or 50Hz, depending on the utility frequency of the standardization. For that, the block "Zcross" detects the zero crossing of the n-dimensional vector $\boldsymbol{v}$ and it passes them to the block "rms", thus providing the starting indexes of $\boldsymbol{v}$ for every utility cycle. Afterwards, the block "rms" uses the signal divided into smaller vectors as defined in (4.3) and (4.4), where S is the number of cycles. The utility frequency is prone to suffer from skews and small variations due to external factors; consequently, it should be denoted that $\boldsymbol{v_s}$ must have the same number of samples as $\boldsymbol{i_s}$, but it may have a different number compared to $\boldsymbol{v_{s+1}}$. Finally, the block "rms" computes $\boldsymbol{v_{rms}}$ as follows in (4.5) and (4.6), where $v_{s,j}$ is the j-th element for the vector $\boldsymbol{v_s}$ described in (4.3), and N_s is the number of samples per cycle S. Likewise, $\boldsymbol{i_{rms}}$ is also evaluated in a similar way.

$$\boldsymbol{v} = \{\boldsymbol{v}, \ldots, \boldsymbol{v}_s, \ldots, \boldsymbol{v}_{S-1}\} \tag{4.3}$$

$$\boldsymbol{i} = \{\boldsymbol{i_0}, \ldots, \boldsymbol{i}_s, \ldots, \boldsymbol{i}_{S-1}\} \tag{4.4}$$

$$\boldsymbol{v_{rms}} = [v_{rms,0}, v_{rms,1}, \ldots, v_{rms,S-1}] \tag{4.5}$$

$$v_{rms,s} = \sqrt{\frac{1}{N_s}\sum_{j=0}^{N_s-1} v_{s,j}^2}, \forall s = 0, 1, 2, \ldots, S-1 \tag{4.6}$$

Ideally, the rms value $v_{rms,s}$ defined in (4.6) should be constant along cycles. Nevertheless, it may varies and this affects to the rms value of the current $i_{rms,s}$ (defined as in (4.6)) given different values for the very same transient produced at different times. A normalization prevents the detector against those variations and allows the generalization of this approach for different scenarios [Han and Lim, 2010, Hart, 1992]. The block "norm" carries out this normalization for each cycle as detailed in (4.7). The parameter V_{norm} is the normalized voltage (for instance $V_{norm} = 220V_{rms}$ in Europe). Thus, the outcome of this block is an N-dimensional vector $\bar{\boldsymbol{i}}_{\boldsymbol{rms}}$ plotted in Figure 4.2(b). The resulting signal is the proposed envelope extraction from the current waveform $\boldsymbol{i}$. It presents step-like changes due to transient states of the refrigerator. Likewise, it is worth to note that $\boldsymbol{i}_{rms}$

is susceptible to voltage drops and it may vary its amplitude for the same transient state produced at a different time. Nevertheless, the normalized signal $\bar{\boldsymbol{i}}_{\boldsymbol{rms}}$ guaranties the same amplitude always. This generalises the detection along different scenarios.

$$\bar{i}_{rms,s} = \frac{V_{norm}^2}{v_{rms,s}^2} * i_{rms,s,} \forall s = 0, 1, 2, \ldots, S-1 \tag{4.7}$$

4.1.1.2 Change Detection

The evaluation of changes in the aforementioned pre-processed signal $\bar{\boldsymbol{i}}_{\boldsymbol{rms}}$ allows the detection of connections and disconnections of appliances in the mains.

By performing a derivation, only changes in the signal are kept. This process filters steady states, leaving spikes corresponding to transient states. Then, the derived signal $\boldsymbol{\nabla}\bar{\boldsymbol{i}}_{\boldsymbol{rms}}$ is squared to increase the gap between small changes due to noise, and significant changes due to transient states.

It is worth noting that the aforementioned derived signal $\boldsymbol{\nabla}\bar{\boldsymbol{i}}_{\boldsymbol{rms}}$ draws positive or negative spikes to reflect the type of event. In the case of positive events, such as turning on or passing to a state of higher consumption, the rms value of the current signal $\bar{\boldsymbol{i}}_{\boldsymbol{rms}}$ increases and, therefore, a positive spike is obtained. Exactly the opposite happens with negative events such as turning off appliances, where negative spikes are reflected. In order to keep track of the type of event (positive or negative), the sign of the derived signal $\boldsymbol{\nabla}\bar{\boldsymbol{i}}_{\boldsymbol{rms}}$ is included after the squaring in the resulting signal $\boldsymbol{\rho}$. This process is shown in Figure 4.1 and described in (4.8) and (4.9). For instance, Figure 4.2(c) shows the signal $\boldsymbol{\rho}$ in blue line for the case of on/off connections of the refrigerator. Positive spikes in $\boldsymbol{\rho}$ are due to ON switching events, whereas OFF events provides negative ones. It can be observed that $\boldsymbol{\rho}$ only keep the transient states information from the current waveform presented in $\boldsymbol{i}$.

$$\boldsymbol{\rho} = [\rho_1, \rho_2, \ldots, \rho_{S-1}] \tag{4.8}$$

$$\rho_s = \begin{cases} -(\bar{i}_{rms,s} - \bar{i}_{rms,s-1})^2, if(\bar{i}_{rms,s} - \bar{i}_{rms,s-1}) < 0 \\ (\bar{i}_{rms,s} - \bar{i}_{rms,s-1})^2, if(\bar{i}_{rms,s} - \bar{i}_{rms,s-1}) \geq 0 \end{cases} , \forall s = 2, 3, \ldots, S-1 \tag{4.9}$$

4.1.1.3 Peak Detector

The output $\boldsymbol{\rho}$ in (4.8) presents small peaks produced by noise and high peaks due to events (i.e. switching events or changes in the consumption state). Further, the signal $\boldsymbol{\rho}$ allows to discriminate between switching-on and switching-off events, which gives extra information to the classifier. The algorithm described in [Duarte, 2015] has been applied to detect those positive peaks from events (i.e. switching-on events). The algorithm requires two parameters: a windowing that constrains peaks to be at a minimum distance, W_d; and a threshold θ_{th} that filters out any peak below it. Transient states may generate oscillations around the main peak and, therefore, the windowing W_d should be adjusted for this period. It is worth noting that a small windowing is desirable since it allows the detection of events occurring at similar time. For instance, if the utility frequency is 60Hz, using a windowing W_d of 20 cycles allows events separated more than 332ms to be differentiated. Likewise, constraining the amplitude of peaks by using the threshold θ_{th} filters out noise from the steady states and only permits passing those peaks coming from events. There is a significant gap between the amplitude of peaks generated by events and those due to noise, so the threshold θ_{th} should be set according to those intervals. This value θ_{th} has been empirically optimized, based on the ROC curve presented in Section 5.4. The corresponding output signal $\boldsymbol{\tau}$ expressed in (4.10) is a vector where every element t_k is the timestamp at which events take place, where K is the number of events.

$$\boldsymbol{\tau} = [t_o, t_1, \ldots, t_{K-1}] \tag{4.10}$$

Figure 4.2(c) shows the peak detection (red dots) where $\theta_{th} = 5(A^2)$ and $W_d = 5$ cycles of the utility frequency (i.e. 83ms in mains of 60Hz). In this case, the four events have been detected and the recorded timestamps for the detection compound the signal $\boldsymbol{\tau}$, defined in (4.10).

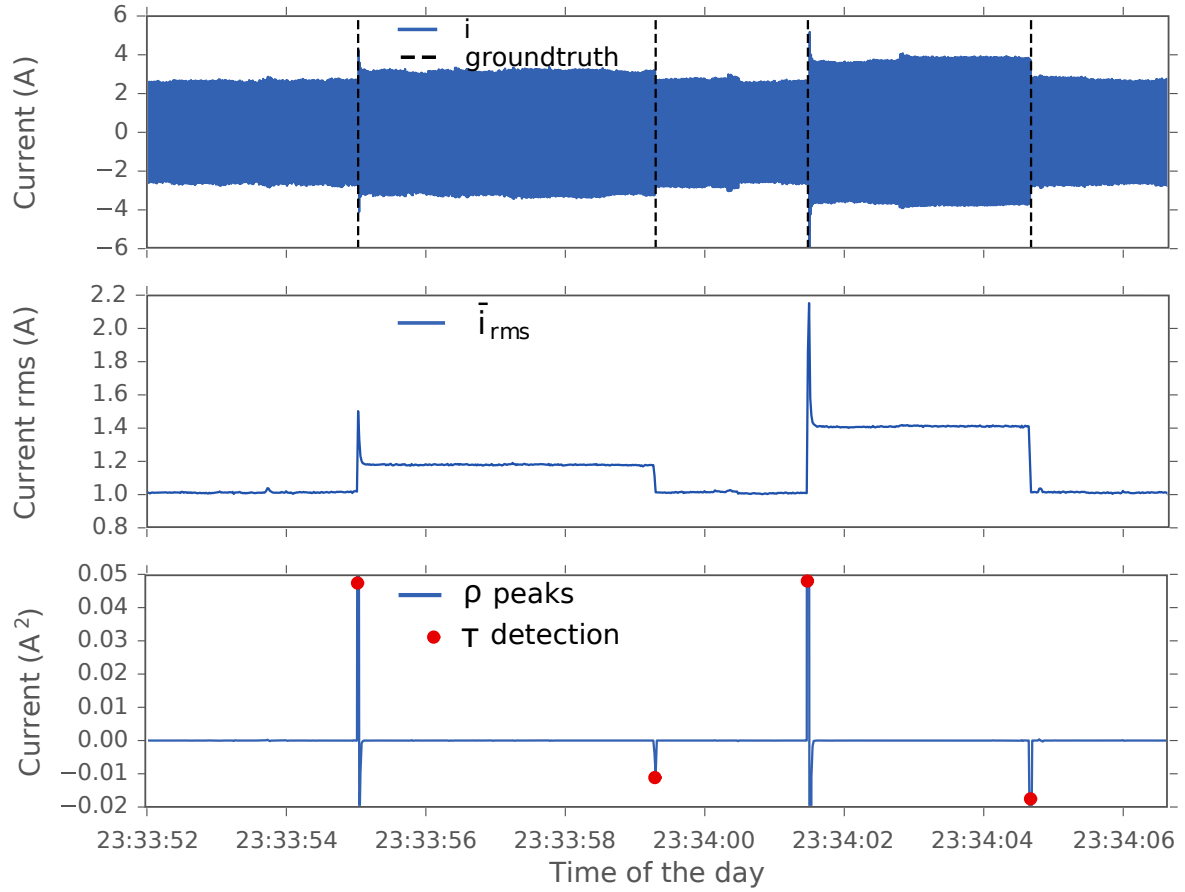


Figure 4.2: Illustration of the event detection algorithm. (a) $\boldsymbol{i}$ Current waveform from the mains and ground-truth of events; (b) Envelope extraction: $\bar{\boldsymbol{i}}_{\boldsymbol{rms}}$ normalized rms value for the current; (c) $\boldsymbol{\rho}$ peak signal and $\boldsymbol{\tau}$ event detection.

4.1.2 Event Classification Algorithm

Given the current and voltage waveforms measured from the mains in a household and applying the algorithm previously described in Section 4.1.1, the events can be tracked and the occurrence timestamp registered. Then, the next step is to classify or label them into appliance types. We aim to have a high accuracy and, therefore, active, reactive and distortion powers are used as signatures to avoid overlapping (see Subsection 2.1.3). In order to extract the most of the information from signatures, delta values (single appliance signatures) of *PQD* curve trajectories are characterized during the transient state and previous and next steady states. Using a mix of transient/steady states as load signatures is a novel approach which, to the best of our knowledge, it has been only presented in

[Cole and Albicki, 2000]. Hence, the proposed approach involves a pre-processing stage, followed by the proposed classification algorithm $PQD - PCA$, Principal Component Analysis of active, reactive and distortion (PQD) powers. This procedure is shown in Figure 4.3 and detailed in following sections.

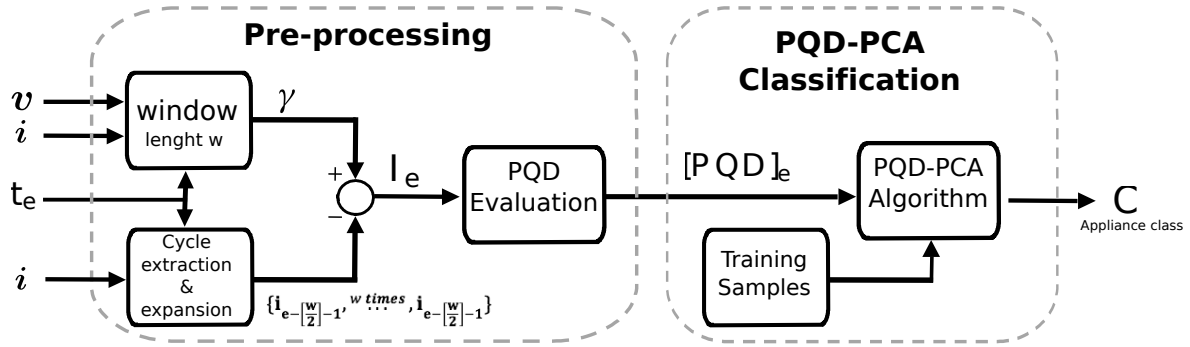


Figure 4.3: Block diagram for the proposed PCA-PQD Classification.

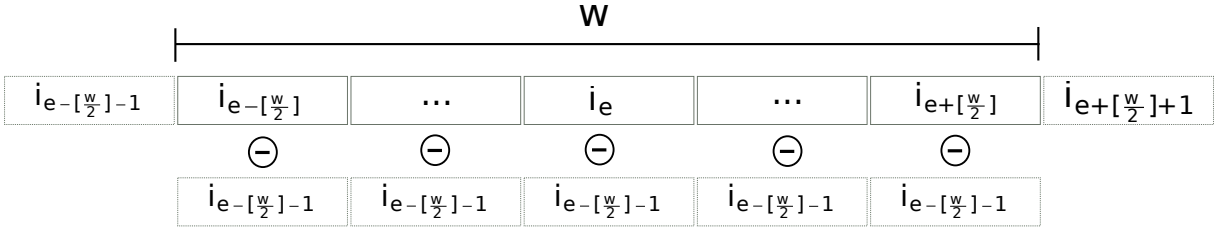


Figure 4.4: Extracting delta values from the aggregated current $\boldsymbol{i}$.

4.1.2.1 Signal Pre-processing: extraction of SPQD Power Signatures

In Figure 4.3, the inputs in the workflow are the following: the voltage $\boldsymbol{v}$ and the aggregated current $\boldsymbol{i}$ from the mains, defined in (4.1) and (4.2) respectively; and the occurrence time t_e of a specific event e that was already detected in Section 4.1.1 and, therefore, contained in the outcome defined in (4.10). Thus, this block extracts and sends the signatures to the $PQD - PCA$ classification block for a particular event e detected in the mains v and i.

The proposal deals with "delta form" signatures as those detailed in Subsection 2.1.3. Basically, a delta value or incremental value is the difference between two successive samples and a "delta form" signature is a collection of delta values. This process, what

we call "de-energising" the signal, is necessary to isolate the current due to the appliance that is causing the event from the rest of the aggregated current $\boldsymbol{i}$. Other approaches deal with incremental values of the active and reactive powers (P-Q) as in [Hart, 1992]. Nevertheless, we also work with the distortion power D and this cannot be easily added or subtracted, as it presents non-linear components. Thereby, we present a new isolation process to extract the "delta form" signatures from the aggregated current $\boldsymbol{i}$ that are directly produced for a specific appliance that triggers a specific event e in t_e. We make a subtle modification and use delta values that are the difference between samples inside a w-length window W and a fixed sample before that window. Note that the voltage waveform $\boldsymbol{v}$ is not affected by the number of appliances connected as all of them are connected to the same voltage, therefore, the isolation process is only needed on the current waveform $\boldsymbol{i}$.

Figure 4.3 shows this isolation process. Firstly, the w-length window is applied to the current waveform $\boldsymbol{i}$, where w is a number of cycles (i.e from the fundamental frequency 50/60 Hz in $\boldsymbol{i}$). The windowing is applied as in (4.11), resulting in $\boldsymbol{\gamma}$, a collection of vector-cycle vectors ("Window block" in Figure 4.3). This window is centred at the event occurrence t_e and should be large enough to hold a previous steady state, the transient state and the next steady state. Heuristically, we found that a suitable choosing of the window length is $w = 60$ cycles (this value has been used on the experimental results in Section 5.4). Nevertheless, it is tunable and could be changed regarding requirements. Then, the "Cycle extraction & expansion" block filters the current waveform i keeping only the vector-cycle previous to the windowing ($\boldsymbol{i}_{e-\left[\frac{w}{2}\right]-1}$) and it creates a collection of repeated vectors. Then, the latter is subtracted from $\boldsymbol{\gamma}$ in (4.12). This, windowing, expansion and subtraction process is depicted in Figure 4.4.

$$\boldsymbol{\gamma} = \left\{ \boldsymbol{i}_{e-\left[\frac{w}{2}\right]}, \ldots, \boldsymbol{i}_e, \ldots, \boldsymbol{i}_{e+\left[\frac{w}{2}\right]} \right\} \tag{4.11}$$

$$\boldsymbol{I}_e = \boldsymbol{\gamma} - \left\{ \underbrace{\boldsymbol{i}_{e-\left[\frac{w}{2}-1\right]}, \boldsymbol{i}_{e-\left[\frac{w}{2}-1\right]}, \ldots, \boldsymbol{i}_{e-\left[\frac{w}{2}-1\right]}}_{w\text{-times}} \right\} \tag{4.12}$$

The resulting collection of vectors $\boldsymbol{I_e}$ emulates the current waveform that the appliance causing the event e would draw during the previous and next steady states and the

transient state. Then, *PQD* trajectories are obtained in the "*PQD* evaluation" block, where every vector-cycle of $\boldsymbol{I_e}$ draws a point value on the *PQD* power cube depicted in Figure 4.5. The latter also gives the power relations that they are expressed in (4.13). Then, for every vector-cycle, the parameters I_{rms} and V_{rms} are evaluated according to (4.17), T is the fundamental period corresponding to the utility frequency. Hence, the apparent power S can be deducted as in (4.14). A FFT analysis is applied to obtain the first harmonic of the electrical current $I_{rms,1}$, also denoted as the lineal component since it presents the fundamental utility frequency. The rms of the current first harmonic is used in (4.15) and (4.16) to obtain the active and reactive powers, P and Q, respectively, where φ is the displacement phase between v and i, also depicted in 4.5. Finally, the distortion power D is obtained from (4.13). The resulting are three vectors $\boldsymbol{P}$, $\boldsymbol{Q}$ and $\boldsymbol{D}$ with length w. These are the incomes for the PQD-PCA classification.

$$S^2 = P^2 + Q^2 + D^2 \tag{4.13}$$

$$S = V_{rms} * I_{rms} \tag{4.14}$$

$$P = V_{rms} * I_{rms,1} * \cos(\varphi) \tag{4.15}$$

$$Q = V_{rms} * I_{rms,1} * sen(\varphi) \tag{4.16}$$

$$I_{rms} = \sqrt{\frac{1}{T}\int_0^T i(t)^2 dt} \tag{4.17}$$

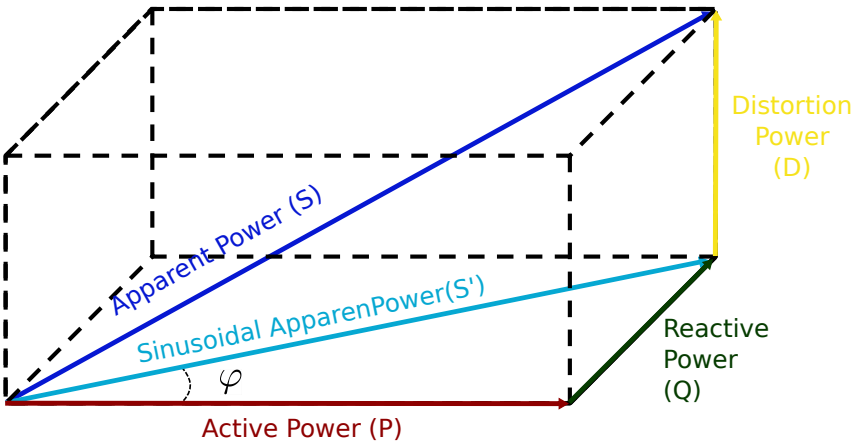


Figure 4.5: Representation of the PQD power cube used to model load behaviour.

4.1.2.2 PQD-PCA Classification Algorithm

Principal Component Analysis (PCA) is a widespread statistical procedure used in many machine learning domains such Vision Computing [Turk and Pentland, 1991] and Ultrasonic Localization [Jimenez et al., 2005]. Regarding the disaggregation domain, the decomposition in Principal Components has been applied in [Xu et al., 2010] to separate the gas-liquid two-phase flow rig through electrical resistance tomography. The goal is to convert a set of observations for correlated variables into a set of values for linearly uncorrelated variables, called principal components. This is carried out by performing an orthogonal transformation. The first principal component has the highest variance possible, whereas every following component is optimized to achieve the maximum allowed variance, always orthogonal to the previous components.

The set of observation PQD within the windowed current I_e in (4.12) provides significant information about the appliance type. Thus, before and after the event, the obtained powers PQD are steady and characteristic for the several operation modes causing them. Moreover, within a few cycles around the event occurrence, there is the transient state, such as start-ups, that are strongly influenced by the circuit configuration of every appliance. The features of these transient states are less susceptible of overlapping [Alcalá et al., 2014]. In contrast to other event-based energy disaggregation algorithms presented in Chapter 2 that are only focused on either steady or transient states, the proposal described here pursues to classify events using both steady and transient states as signatures. The PCA transforms these three correlated variables PQD into non-correlated

ones, which simplify the classification described below. This classification process was introduced for multi-sensory ultrasonic localization in [Jimenez et al., 2005].

Following the diagram in Figure 4.3 the PQD-PCA algorithm needs to be previously trained with *PQD* instances of those appliance types (hereinafter appliance classes) susceptible to be detected. Therefore, for a certain appliance class a, let us consider a training set of M samples, where every sample m is a w-dimensional vector of current measures I_e, as described in (4.12). Let us also consider the matrix $[\boldsymbol{PQD}]_m^a$ (4.18) as the evaluated powers P, Q and D in(4.13), (4.14), (4.15), (4.16) and (4.17) for a certain sample m (indicated as a subindex in the following expressions).

$$[\boldsymbol{PQD}]_{\boldsymbol{m}}^{\boldsymbol{a}} = \begin{bmatrix} P_{0,} & Q_{0,m} & D_{0,m} \\ \vdots & \vdots & \vdots \\ P_{w-1,m} & Q_{w-1,m} & D_{w-1,m} \end{bmatrix} \tag{4.18}$$

The matrix above can be rearranged as a $3w \times 1$ column vector $\boldsymbol{\tau}_m^a$ as follows:

$$\begin{aligned} &\boldsymbol{\tau}_{\boldsymbol{m}}^{\boldsymbol{a}} = \{\boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{D}\}^{\boldsymbol{T}} = \\ &= [P_{0,m}, P_{1,m}, ..., P_{w-1,m}, Q_{0,m}, Q_{1,m}, ..., Q_{w-1,m}, D_{0,m}, D_{1,m}, ..., D_{w-1,m}]^T \end{aligned} \tag{4.19}$$

Then, the averaged vector $\boldsymbol{\Psi}^a$ for the set of training can be evaluated as:

$$\boldsymbol{\Psi}^{\boldsymbol{a}} = \frac{1}{M} \sum_{m=0}^{M-1} \boldsymbol{\tau}_m^a \tag{4.20}$$

Where $\boldsymbol{\Psi}^a \in \mathbb{R}^{3w}$. Subtracting the averaged vector $\boldsymbol{\Psi}^a$ from the training set $\boldsymbol{\tau}_m^a$, only the changes $\boldsymbol{\Phi}_m^a$ remains:

$$\boldsymbol{\Phi}_m^a = \boldsymbol{\tau}_m^a - \boldsymbol{\Psi}_m^a, \forall m = 0, 1, \ldots, M-1 \tag{4.21}$$

And the total scatter matrix $\boldsymbol{S}_T^a$ is evaluated as:

$$\boldsymbol{S}_T^a = \sum_{m=0}^{M-1} \boldsymbol{\Phi}_m^a (\boldsymbol{\Phi}_m^a)^T \tag{4.22}$$

Considering a linear transformation where $3w$ column vectors of the training set above in (4.21) are mapped into a l-dimensional space, such as $l < 3w$, a new set of features $\boldsymbol{\Omega}_m^a \in \mathbb{R}^{\mathsf{l}}$ is obtained:

$$\boldsymbol{\Omega}_m^a = (U)^T \boldsymbol{\Phi}_{\boldsymbol{m}}^{\boldsymbol{a}}, \forall m = 0, 1, \ldots, M-1 \tag{4.23}$$

Where the matrix transformation $\boldsymbol{U} \in \mathbb{R}^{3w \times l}$ has orthogonal columns.

If we consider the scatter matrix $\boldsymbol{S}_T^a$ in (4.22) for the training set (4.21), then, for the new feature vectors $\boldsymbol{\Omega}_m^a$ in (4.23), the new scatter matrix is $(U)^T \boldsymbol{S}_T^a U$. In PCA this scatter matrix in the projected space should be chosen in order to maximize its determinant. This is done as shown in (4.24), where $\{\boldsymbol{u_k} | k = 0, 1, \ldots, l\}$ are the set of the l largest eigenvectors out of the $3w$ independent eigenvectors from the scatter matrix $\boldsymbol{S}_T^a$. Thus, the vector $\boldsymbol{u_1}$ is the first principal component and it has the largest variance.

$$\boldsymbol{U}_{opt}^a = \arg\max_U |(U)^T \boldsymbol{S}_T^a U| = [\boldsymbol{u_1}, \boldsymbol{u_2}, \ldots, \boldsymbol{u_l}] \tag{4.24}$$

So far, we have described the training process for a certain appliance class a, resulting in the matrix $\boldsymbol{U}_{opt}^a$. Then, assuming a new sample $\boldsymbol{\tau}_e = \begin{bmatrix} \boldsymbol{P} & \boldsymbol{Q} & \boldsymbol{D} \end{bmatrix}_m$ arrives for classification (see Figure 4.3), firstly, we translate it into the new feature space for the appliance class a as:

$$\boldsymbol{\Omega}_e^a = (\boldsymbol{U}_{opt}^a)^T \boldsymbol{\Phi}_e^a = (\boldsymbol{U}_{opt}^a)^T (\boldsymbol{\tau}_e - \boldsymbol{\Psi}^{\boldsymbol{a}}) \tag{4.25}$$

Where Ψ^a is the mean vector evaluated in (4.20). Then, over the new feature space $\boldsymbol{\Omega}_e^x$, we perform the inverse transformation in (4.26). This provides an estimation of the new feature space $\hat{\boldsymbol{\Phi}}_{\boldsymbol{e}}^{\boldsymbol{a}}$ from where an estimation of the sample $\hat{\boldsymbol{\tau}}_{\boldsymbol{e}}$ can also be obtained.

$$\widehat{\boldsymbol{\Phi}}_e^a = (\hat{\boldsymbol{\tau}}_e - \boldsymbol{\Psi}^{\boldsymbol{a}}) = \boldsymbol{U}_{opt}^a \boldsymbol{\Omega}_e^a \tag{4.26}$$

Since the linear transformation is $\boldsymbol{U}^a_{opt} \in \mathbb{R}^{3w \times l}$, the k smallest eigenvalues (those with the smallest variance), where $k = 3 \cdot w - l$, are discarded. This results in energy losses during the reconstruction $\widehat{\boldsymbol{\Phi}}^a_e$, expressed as:

$$\varepsilon^a_e = ||\boldsymbol{\Phi}^a_e - \hat{\boldsymbol{\Phi}}^a_e|| \tag{4.27}$$

Consequently, if the arrival sample $\boldsymbol{\tau}_e$ has a similar variance to the training set $\boldsymbol{\tau}^a_m$ in (4.19), then, the linear transformation $\boldsymbol{U}^a_{opt}$ arranges its variance on the first l eigenvalues in $\boldsymbol{\Omega}^a_e$ leaving out just a small quantity of variance. As a result, the reconstruction in (4.26) barely losses energy and the error reconstruction ε^a_e (4.27) should be small. Otherwise, a sample $\boldsymbol{\tau}_e$ belonging to a different appliance class does not have similar variance distribution, resulting in a bad re-arranging after the linear transformation and larger losses (i.e. big error in the reconstruction (4.27)).

With concern to the classification, the sample $\boldsymbol{\tau}_e$ can be classified into a certain appliance class x if the reconstruction error is $\varepsilon^a_e < \gamma_A$, where γ_a is a certain heuristic threshold for the appliance class x. Let us then consider a set of A appliance classes that have been trained to obtain their respective linear transformation as described before: $\boldsymbol{U}^a_{opt}, \forall a = 0, 1, 2, \ldots, A-1$. Thus, the reconstruction error in (4.27) for the sample $\boldsymbol{\tau}_e$ is evaluated for every appliance class a in order to obtain the corresponding errors $\varepsilon^a_e, \forall a = 0, 1, 2, \ldots, A-1$, and the sample τ_e is classified as appliance class C, if it satisfies (4.28). Otherwise, this sample $\boldsymbol{\tau}_e$ remains as unclassified (not belonging to any of the trained classes in A).

The threshold γ_a is used to discard outliers, that are situations that do not correspond with any of the classes considered, and it can be set automatically. Hence, its value is high compared with the reconstruction error ε^a_e. Let us say the average error for the appliance class a during the training process is 0.5. Then, we set γ_a ten times higher (i.e. $\gamma_a = 5$). On the arrival of a new sample $\boldsymbol{\tau}_e$, if the minimum reconstruction error obtained from 4.27 is $\varepsilon^a_e = 8$, the sample $\boldsymbol{\tau}_e$ is likely to be an outlier and, consequently, it remains unlabelled. Therefore, once the training process is carried out (as performed in Subsection 4.2.2), the parameter is set automatically. Eventually, it could be updated with new correctly classified samples from new houses. As for the results presented in Subsection 4.2.2, no outlier has been found.

$$C = \arg\min_{\forall a \in A} \varepsilon^a_e \text{ and } \varepsilon^a_e < \gamma_a \tag{4.28}$$

4.2 Experimental Results

We use two different public datasets to evaluate the performance of the proposal: the BLUED and the PLAID datasets [Anderson et al., 2012b, Gao et al., 2014]. Taking into account the particularities of each database, we tested the performance of the event detection algorithm presented in Section 4.1.1 over the first one. Elsewise, the second one has been used to evaluate the classification algorithm described in Section 4.1.2.

The BLUED dataset contains voltage and current measurements from a family household in the United States, sampled at 12kHz during a week. Nevertheless, the current sensor used to acquire data had an upper cutoff frequency of about 300Hz. This means that the signals used to validate the proposed event-detector have a bandwidth of 300Hz [Anderson et al., 2012b]. It comprises a large number of events (2845 events) labelled with timestamp, appliance identifier and phase. The number of appliances involved is also large: 50 different appliances and 30 different appliance types. Thus, this dataset is much convenience for the analysis of the proposal as it represents a widespread number of real scenarios. Besides, to the best of our knowledge it is the existing largest dataset with events labelled.

The Plug-Level Appliance Identification Dataset (PLAID) consists of current and voltage snapshots of a few seconds (i.e. containing both steady and transient states) from different appliances and different households. Currently, there are 1074 instances from 11 different appliance types present in 55 households, sampled up to 30kHz. For example, there are 176 instances of different laptops. This is a very convenient scenario for the PQD-PCA algorithm, since it can be trained with heterogeneous samples, in steady and transient states, for achieving an universal classifying model that works along different appliance brands and households. It is worth to note that the current clamp employed (the Fluke i200 AC model) had an upper cutoff frequency of 1kHz. Consequently, the actual bandwidth of the signals considered for the proposed classification algorithm is 1kHz.

4.2.1 Event detection performance

This Section presents the performance of the proposed event-detection algorithm using BLUED dataset. Typically, the event detection metrics are four (Appendix B): True

Positive Rate (TPR), True Positive Percentage (TPP), total power change and average power change. The last two weigh the event according to the corresponding power change. In this study, the TPR and the False Positive Rate (FPR) are evaluated, since the power demand should not be a key factor to measure the performance. For activity analysis purposes, all events weigh the same. Furthermore, a ROC curve and the Area Under Curve (AUC) can be drawn out from TPR and FPR pair values to optimise the detector. TPR and FPR are defined as:

$$TPR = \frac{TP}{TP + FN} \in [0, 1] \tag{4.29}$$

$$FPR = \frac{FP}{FP + TN} \in [0, 1] \tag{4.30}$$

Where TP is the number of True Positives, FN is the number of False Negatives (also known as missed), FP is the number of False Positives and is the number of True Negatives. Likewise, for a binary detector, a TP happens whenever an event occurs and this is detected, otherwise it is a FN whether the event is missed. Moreover, any time that events do not occur, there is a TN whether the detector does not detect any event or a FP otherwise. Hence, based on the event detector parameter settings we can obtain different TPR and FPR values. The best detector is such that and $TPR = 1$ and $FPR = 0$ and our optimal detector ψ_{opt} is such that minimises the Euclidean distance as in (4.31). Each pair of settings for parameters θ_{th} and W_d gives a detector ψ. The set of all settings in the test gives the set of detectors Ψ.

$$\psi_{opt} = \underset{\forall \psi \in \Psi}{\arg\min} \, \|(0, 1) - (FPR, TPR)\|^2 \tag{4.31}$$

The proposed event-detection algorithm has been carried out several times for different configurations of θ_{th} and W_d, over the whole BLUED dataset for both phases. Then, for every iteration, the TPR and the FPR as in (B.1) and (B.2) have been evaluated.

Assuming the window W_d constant, a set of θ_{th} draws different pairs of TPR and FPR values, which can be represented in a ROC curve as in Figure 4.6. Small values in W_d allow higher time resolution for detection and this is reflected on the ROC curve, whose

slope increases getting close to the ideal estimator $(TPR, FPR) = (1, 0)$. This is true until a certain point, at which the window is so short that oscillations in the transient states may be also identified as detections and, hence, they increase the false positive rate (FPR). After several tests, the chosen optimal windowing for the BLUED dataset and for the proposed algorithm is $W_d = 5$ cycles, which guaranties the transient states have already occurred. This also means that the achieved time resolution to distinguish an event is 83ms (5 cycles for a 60 Hz signal). Nevertheless, this accuracy cannot be verified through the dataset as, in most cases, there is a desynchronization between the actual event and the ground-truth, which forces to downsample to 1 second in order to draw the ROC curve. Then, events happening within a shorter interval than one second will be considered just as a single event.

The ROC curve is presented in Figure 4.6 for both phases A and B, as well as for a random guess detector which draws a straight line from (0,0) to (1,1). The zoom area gives more details about the performance for phase A. Furthermore, Table 4.1 shows the optimal detector (the one with the minimum Euclidean distance between the ideal detector and the respective curves). Table shows a high performance in phase A (i.e. 94% of events detected with less than 0.1% of FP). Although the performance is better in phase A, it is still high in B, where 88% of True Positives are correctly detected with less than 12% of False Positives. In a real deployment, it is often impractical to tune the detector to achieve an optimal one and, hence, the AUC is normally provided to give an idea of the overall detector performance whenever the parameters are not optimised. Thus, Table 4.1 shows a high performance for the AUC in phase A (i.e. 97%) and in phase B (i.e. 94%).

The outperforming in phase A can be explained in Table 4.2. Regarding the 2485 events, there is a larger number of events in phase B (1578) than in phase A (907). Furthermore, there is a larger number of appliances plugged to phase B than A, and also those appliances are characterized by a smaller step-change. All this makes phase B a more complex scenario for detecting events.

It is worth noting that the used dataset is large: two currents (phases A and B) and a voltage sampled at 12kHz during a week draws 56 Gb of compressed data. To evaluate the performance shown in Figure 4.6, where every value in the curve implies running the detector algorithm through the whole dataset, a computation cluster has been used to launch parallel tasks on several machines with highly computational features. This has

allowed to speed up the process of obtaining metrics when offline evaluation is possible since, in a real implementation, the event detector just runs in real time.

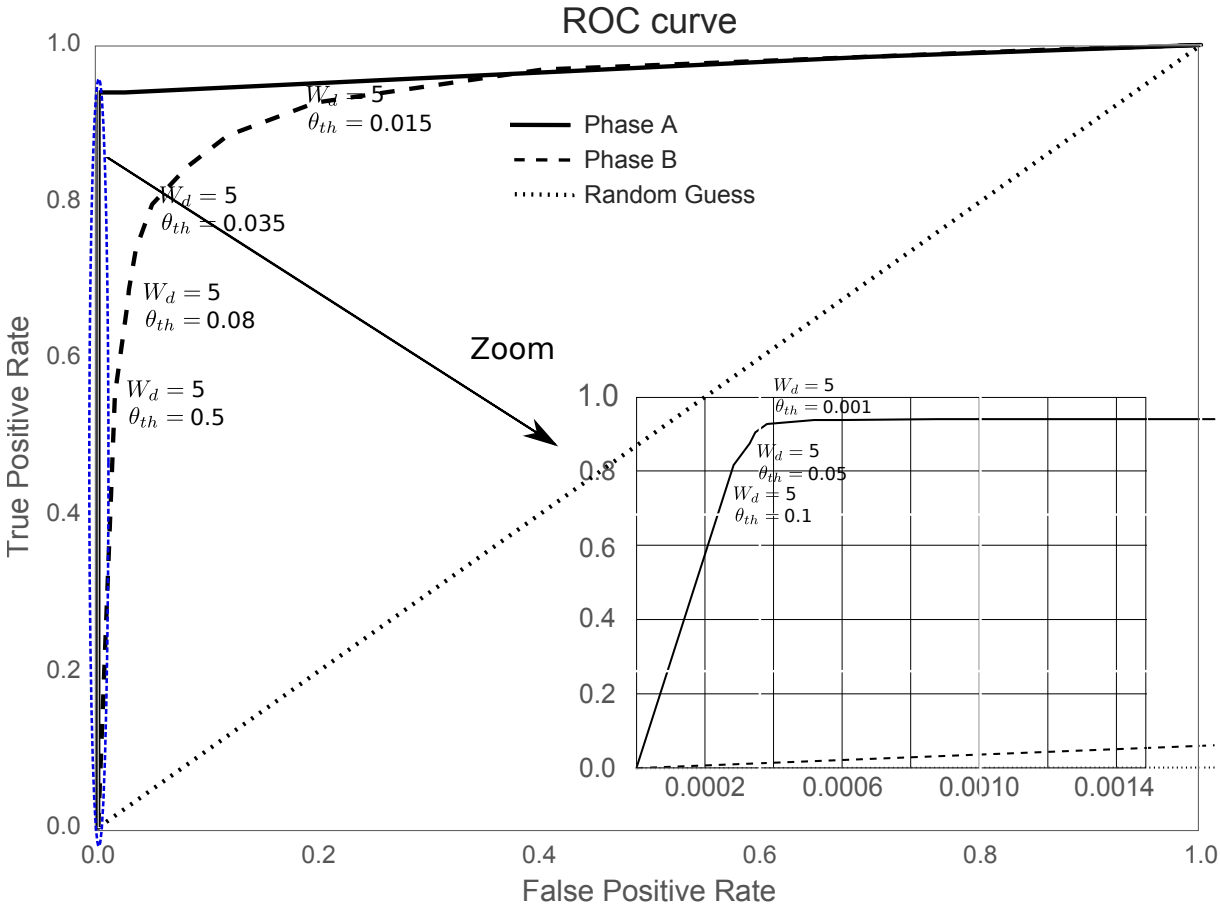


Figure 4.6: Event-detection algorithm performance evaluated on a ROC curve.

Phase	Optimal detector		Area Under Curve
	FPR	**TPR**	**AUC**
A	0.88%	94%	97%
B	0.12%	88%	94%

Table 4.1: Event-detection algorithm performance. Metrics: False Positive Rate and True Positive Rate for the optimal detector (FPR, TPR) and Area Under Curve (AUC).

Phase	Appliance class	Events	Phase	Appliance class	Events
A	Air compressor	20	B	Basement receiver /DVR/blue ray player	34
	Backyard lights	16		Basement lights	39
	Bathroom upstairs lights	98		Circuit 10	99
	Bedroom lights	19		Circuit 11	394
	Circuit 7	38		Circuit 4	46
	Hair dryer	8		Circuit 9	46
	Kitchen aid chopper	16		Closet light	22
	Refrigerator	619		Computer 1	45
	Unknown appliances	67		Dining Room overhead light	32
	Washroom light	6		Garage door	24
				Hallway stairs light	58
				Iron	40
				Kitchen hallway light	6
				Kitchen overhead light	56
				LCD monitor 1	77
				Laptop 1	14
				Living room A/V system	8
				Living room desk lamp	26
				Living room empty socket	2
				Living room tall desk lamp	25
				Monitor 2	150
				Office lights	54
				Printer	150
				TV	54
				Unknown appliances	60
				Upstairs hallway light	17
	Total	**907**		**Total**	**1578**
Total (A+B)					**2485**

Table 4.2: Occurrence of events in the PLAID dataset per phase and appliance type.

There is a potential number of studies in literature to compare with. Before that, it is important to keep in mind two things:

Firstly, the proposed event detector has been devised to be as much generalisable as possible around households. This means that no training is required and the parameter settings have been reduced to one. This could lead to lower performances than other well trained and parametrised event detectors.

Secondly, the detector has been optimized according to the ROC curve because the aim is to achieve the highest number of TP positive whereas the FP has less relevant importance. Thus, FP can be lately filtered out by a high-performance event classifier (e.g. the PQD-PCA algorithm that is also proposed in this research). This means that our proposal may score lower in other metrics.

At the time of implementing the proposed event-detector, in [Barsim et al., 2014] it is presented the results of an event detector along with an energy disaggregation algorithm that, with a 93% accuracy, was placed between the top ten disaggregation algorithms in an energy disaggregation competition held by the company Belkin in 2014 [Belkin, 2013]. We evaluate the performance of our algorithm using TPP and False Positive Percentage (FPP) metrics as in (4.32) and (4.33), respectively; where P is the number of positives or occurred events. Then, we compare our results with the aforementioned study. In this case, the optimal detector is given by the closest point to $(FPP, TPP) = (0, 1)$. We draw the results in Table 4.3 and Table 4.4 for phase A and B, respectively. It can be observed that the TPP performance is similar in phase A and our algorithm outperforms in phase B. Regarding the FPP performance, there is a detriment to the performance in our algorithm. In fact, this was expected as we optimised the algorithm to get the highest number of TP in detriment to FP that can be later filtered out during the classification stage. It is worth noting than $P = TP + FN$ and, hence, $TPP = TPR$; nonetheless $FPP \neq FPR$ and this leads to a different optimal detector and slight changes in the TPP value in Table 4.3 and Table 4.4 related to Table 4.1.

$$TPP = \frac{TP}{P} \in [0, 1] \tag{4.32}$$

$$FPP = \frac{FP}{P} \in [0, \infty) \tag{4.33}$$

Phase A	**Proposed event detector**	**[Barsim et al., 2014]**
TPP	95%	98%
FPP	28%	0.55%

Table 4.3: The proposed event detector performance compared with [Barsim et al., 2014] in the TPP and FPP metrics for the phase A of BLUED dataset.

Phase B	**Proposed event detector**	**[Barsim et al., 2014]**
TPP	89%	70.5%
FPP	33%	8.75%

Table 4.4: The proposed event detector performance compared with [Barsim et al., 2014] in the TPP and FPP metrics for the phase B of BLUED dataset.

Then, if we are to compare with other recent methods, where training, constrains and parameter settings are required, the proposed event detector performance is lower than them but it is still able to compete. Table 4.5 and Table 4.6 draw the performance of our algorithm regarding recall (4.34), precision (4.35), F1 score (4.36) where $\beta = 1$ and F2 score (4.36) where$\beta = 2$ metrics for phase A and phase B, respectively. Our proposal has been compared with three other approaches presented in Chapter 2 (Section 2.1.3). [Anderson et al., 2012b] develops a statistic method based on GLR that needs training and parametrises two moving windows, a voting window and a threshold value. [Wild et al., 2015] is not a probabilistic model and, consequently, does not need a training process, however, it does need to parametrise two moving windows, a threshold and to set two constrains. Finally, in [De Baets et al., 2016] a frequency-domain analysis is performed and the cepstrum components are obtained where a training process is required. Moreover, our proposed event detector algorithm only needs to set a threshold (i.e. the θ_{th} parameter). Even though, the metrics presented in Table 4.5 and Table 4.6 show that performance is not far from other approaches. As explained before, our proposal prioritises TP and that makes it more sensitive to FP, drawing a low performance in the precision metric. Therefore, we also presented the F2 score metrics that double up the importance of recall against precision. This can be obtained from (4.36) and (4.36) where W_{recall} and $W_{precision}$ are the weight of recall and precision, respectively. Thus the F1 score is the harmonic mean between recall and precision (i.e. $\beta = 1$). Then, observing F2 we can obtain a more approximated view of the performance of our proposal, taking into account this biased importance of recall over precision.

Phase A	**Proposed event detector**	**[Wild et al., 2015]**	**[Anderson et al., 2012b]**	**[De Baets et al., 2016]**
Recall	94.7%	98.78%	98.16%	–
Precision	76%	99.66%%	97.94%	–
F1	83%	–	–	98.01%
F2	88.7%	–	–	–

Table 4.5: Comparison of the proposed event detector with three other approaches ([Wild et al., 2015, Anderson et al., 2012b, De Baets et al., 2016]) in recall, precision and F-scores metrics for the phase A of BLUED dataset.

Phase A	The proposed event detector	[Wild et al., 2015]	[Anderson et al., 2012b]	[De Baets et al., 2016]
Recall	77%	92.17%	70.4%	–
Precision	52%	86.32%	87.29%	–
F1	63%	–	–	80.03%
F2	81%	–	–	–

Table 4.6: Comparison of the proposed event detector performance with three other approaches ([Wild et al., 2015, Anderson et al., 2012b, De Baets et al., 2016]) in recall, precision and f-scores metrics for the phase B of BLUED dataset.

$$recall = \frac{TP}{TP + FN} \in [0, 1] \tag{4.34}$$

$$precision = \frac{TP}{TP + FP} \in [0, 1] \tag{4.35}$$

$$F_\beta = (1 + \beta^2) * \frac{precision * recall}{\beta^2 * precision + recall} \in [0, 1] \tag{4.36}$$

$$\beta = \frac{W_{recall}}{W_{precision}} \tag{4.37}$$

4.2.2 PQD-PCA algorithm performance

Firstly, we proceed to extract the signatures P, Q and D from every single appliance in the PLAID dataset as was detailed in Subsection 4.1.2. For instance, the signature extraction process for two appliances (a microwave and a heater) is shown in Figure 4.7. The *PQD* signatures from capacitive appliances (i.e the microwave) provide a negative reactive power, whereas the *PQD* signatures from resistive appliances (i.e the heater) have a null reactive power. Besides, the distortion power provides another discriminant factor, as some appliances generate more harmonics than others. Microwaves produce a high level of harmonics and, consequently, the distortion is high, whereas heaters are purely resistive and their produced distortion is negligible. The transient states also have

a different duration, and the curves for powers *PQD* within this period are unique and characteristic from the circuit configuration [Alcalá et al., 2014].

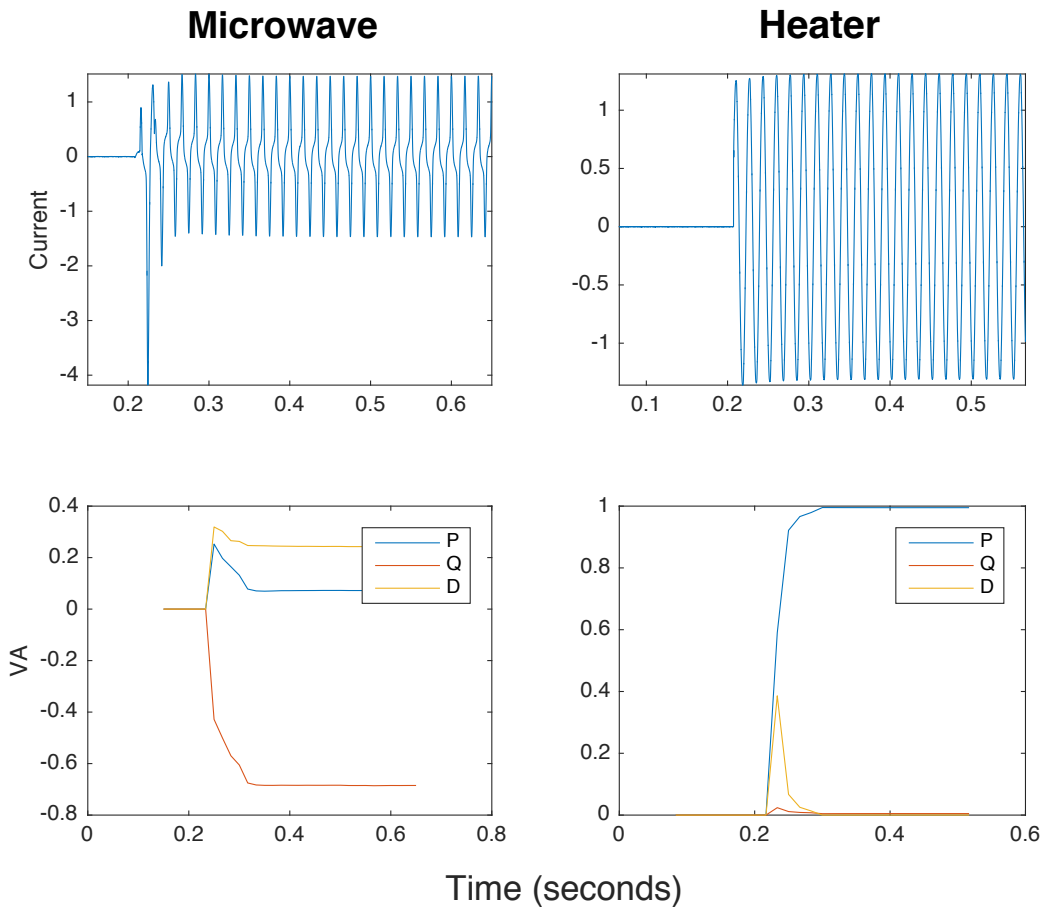


Figure 4.7: Two instances from a Microwave and a Heater: (a) current waveform for microwave; (b) PQD signatures for microwave; (c) current waveform for heater; (d) PQD signatures for heater.

The curves for powers *PQD* seem to be a suitable choice to distinguish the appliance type. Nevertheless, the *PQD* curves for a single appliance involve some correlations, and this is why they need to be processed as described in section 4.1.2, in order to rule out redundancies and compress the variance on the principal components. We carry out this task by means of a linear transformation of the scatter matrix learned from several *PQD* signature curves from instances of the same appliance class. Assuming that the instances from the same appliance class have similar variance distribution, the same linear transformation leads to a similar distribution over the Principal Components. Then, by performing the

inverse linear transformation over the Principal Components, where the smallest eigenvalues are discarded, small energy losses are obtained because most energy information is packed into the largest Principal Components. Nevertheless, the same procedure over instances from different classes results in higher scattering variances, which imply larger energy losses after the inverse linear. We set the number of eigenvalues for the inverse transformation of each appliance class to be such that the 97% of the variance in the Principal Components from the scatter matrix is kept. For instance, if the 97% of the scatter matrix variance is compressed over the two first Principal Components after the linear transformation, only the two first eigenvalues are used for the inverse transformation. This results in about 3% energy losses for instances from the same appliance class and much more energy losses for instances from different classes. Furthermore, to prevent the *PQD-PCA* algorithm from learning abrupt changes on the *PQD* signature curve due to errors in measurements, we evaluate the powers through (4.13), (4.14), (4.15), (4.16) and (4.17) every 4 cycles of the utility frequency with an overlapping of 3 cycles. Also, in order to avoid voltage drops, the following normalization is presented.

$$P_i = \text{sgn}(P_i) * \frac{P_i^2}{S_i^2}, \forall i = 0, 1, \ldots, w-1 \tag{4.38}$$

Where w is the window length in number of cycles detailed in Subsection 4.1.2 and, consequently, the number of samples for the signature curve $\boldsymbol{P}$ in *(4.18)* whereas S_i is the sample i in the apparent power signature $\mathbf{S}$ (4.18). This is also evaluated, in a similar manner, for $\boldsymbol{Q}$ and $\boldsymbol{D}$. This $\boldsymbol{PQD}$ normalization is presented in Figure 4.8(b) and 4.8(d).

We conducted the learning process for all appliance types in the PLAID dataset except for the fridge that has been discarded due to the lack of instances. Figure 4.8 depicts three different instances of a fridge from the PLAID dataset that seem to correspond to different operational modes. It can be observed how the *PQD* curve signatures are completely different, so they should be considered as different appliance classes in the learning process. Furthermore, there are only a few instances of fridges in the dataset (35 instances) which imply a reduced set to distinguish among the operational modes (less than 10 instances per operational mode). Hence, we had to discard the fridge in the learning process. In this way, for every appliance class the 60% of instances are used for training, whereas 40% of them are dedicated to test. As shown in Table 4.7, the washing machine has only 14 instances for training. This number would be less than the constraining 10 instances per operational mode whether the training rate is less than 60%.

Also, there is another appliance type in the dataset which has to be divided into several classes (operational modes). The microwave has been divided into "preheat" and "heat" classes.

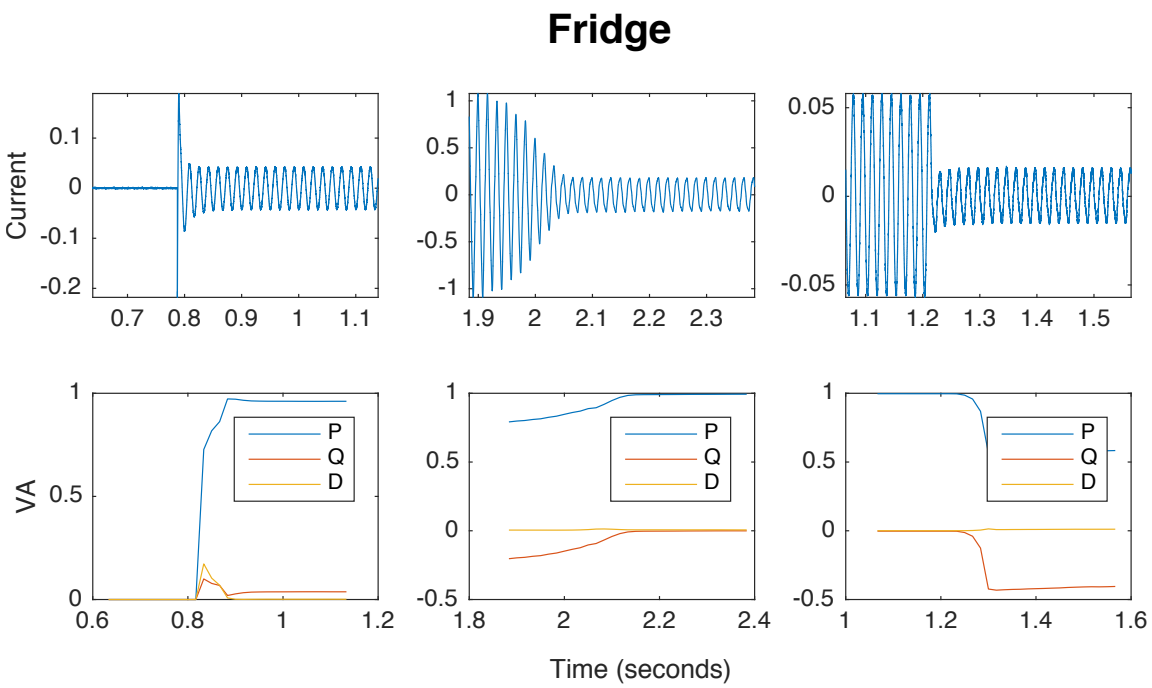


Figure 4.8: Three different instances of fridges corresponding to different transient states.

We evaluate the performance of our *PQD-PCA* algorithm on the *PQD* curve signatures from PLAID by means of the metrics presented in Table 4.7 for every appliance class. Furthermore, to achieve more reliable results, we perform a Monte Carlo Cross Validation. Thus, the values in Table 4.7 are the resulted mean values of metrics throughout ten conducted experiments. For every experiment, training instances have been randomly selected. Likewise, overfitting during the training process is avoided. Finally, an average value for metrics along all appliances classes is presented. Also, it is worth remarking that the "Test instances" column only indicates the number of instances left for testing in every appliance class. The classification performance has been computed for the whole test instance population (i.e 446, the "Total" in "Test instances" columns in 4.7).

Regarding the metrics, there are four commonly used to evaluate NILM algorithms performance: the recall (4.39, the precision (4.40), the F1 score (4.36) and the accuracy

(4.41). They are also defined in Appendix B. Recall (also known as sensibility) represents the ratio of positives detected out of the total number of positives. For instance, the total number of Laptop instances is 67 and our detector is able to correctly detect and classify 63 out of them (93.1%), discerning from a population of 446. Likewise, precision represents the ratio between True Positives and False Positives. Therefore, the precision on the example before is 100% as no instance from another appliance class was classified as Laptop. Both recall and precision are sufficient to evaluate the performance of our detector. From Table 4.7, it can be seen that the proposed algorithm presents a high performance on both metrics with an average for all the appliances of 83% and 98.6% for recall and precision, respectively.

$$recall = \frac{TP}{TP + FN} \tag{4.39}$$

$$precision = \frac{TP}{TP + FP} \in [0, 1] \tag{4.40}$$

$$accuracy = \frac{TP + TN}{P + N} \in [0, 1] \tag{4.41}$$

These two metrics are often represented as one on F1 score, which is the harmonic mean between them. The average of the F1 score along all the appliances is 90.6%. In general, it can be observed that the appliances classes with more training instances are best detected than others. For instance, the F1 score is 100% (ideal) for the microwave on "heat" mode (74 training instances) whereas the heater (20 training instances) draws a 64.7%. However, some appliances classes need less training examples than others because the singularity of their signatures overlaps less with other classes. In this way, the washing machine scores higher than the heater with less training instances: F1 score is 85.4% with 14 training instances.

Precision and recall often have the same importance on the performance. Nevertheless, one might be more important than the other in some scenarios. Specifically, using a NILM algorithm for Activity Recognition applications, where the usage pattern needs to be learnt, it becomes more important precision than recall. Otherwise, having a high sensibility and a poor precision could lead to a bad usage pattern learning, feeding the

modelling with many false positives (i.e instances from a different appliance class). Therefore, we also present the F_β score in (4.36), which is the weighed mean between recall and precision. Likewise, $\beta = \frac{w_{recall}}{w_{precision}}$ (4.37) is the ratio between the weight of the recall (w_{recall}) and the weight of the precision ($w_{precision}$). Thus, $\beta = 1$ is the harmonic mean.

Table 4.7 shows $F_{0.5}$ score to illustrate the prominent importance of precision against recall on applications where the usage pattern needs to be learnt. Here, the precision weights double. It can be seen now how the performance increases, specially for the heater that, having a recall of 47.7%, the $F_{0.5}$ score draws a performance of 99.7%. Although less relevant, the accuracy metrics (a measure of the systematic error) is also evaluated. The averaged accuracy of this proposal is 98%. In general, it can be seen that the PCA-PQD algorithm performs quite well for all the metrics, being slightly better on precision (98.6% on average). The best performance is for Compact Fluorescent Lamps (98% in F1 score), hence, this proposal could be beneficial in Smart Lightning Systems for large smart buildings where accuracy is required and the impact of Wireless interference is intended to be reduced [Li and Lin, 2015].

Appliance class	Recall	Precision	F1	F0.5	Accuracy	Test instances	Training instances
Air conditioner	0.923	0.985	0.953	0.984	0.967	26	38
Compact Fluorescent Lamp	0.994	1	0.997	1	0.992	67	100
Fan	0.825	0.981	0.896	0.981	0.958	44	66
Hairdryer	0.815	0.996	0.896	0.995	0.959	61	91
Heater	0.477	1	0.646	0.997	0.971	13	20
Incandescent Light bulb	0.795	1	0.887	0.999	0.973	44	67
Laptop	0.931	1	0.964	0.999	0.986	68	102
Microwave “preheat”	0.984	1	0.992	1	0.996	49	74
Microwave “heat”	1	1	1	1	0.998	50	74
Vacuum Cleaner	0.7	0.934	0.8	0.934	0.986	15	23
Washing Machine	0.778	0.95	0.854	0.946	0.992	9	14
Average	**0.838**	**0.986**	**0.906**	**0.985**	**0.98**		
Total						**446**	**669**

Table 4.7: Metrics to evaluate the performance of the *PQD-PCA* algorithm over the PLAID data set.

Although results are promising, there is a significant decrease in the classification performance for the heater regarding the recall metric. We would like to further investigate the cause of that. Hence, we employ a new metric commonly used in computer vision: the Cumulative Ranking Curve (CRC) or Cumulative Matching Characteristics [Gray and Tao, 2008].

For instance, for a set of 11 different appliance classes, there will be 11 different CRC curves, each one corresponding to a class a. The x-axis corresponds to the ranking, whereas the y-axis represents the correct classification rate, ranking one is equivalent to recall. Thus, having 100 instances from a certain appliance class, if it is obtained a ranking value of 3 and a classification rate of 0.5, it means that, for the 50 instances out of the 100, the correct appliance class was chosen as first, second or third option. Thus, the ranking is the number of possible appliance options to select and the y-axis indicates the instance rate where the appliance class was considered as an option within the ranking interval. For instance, if a CRC curve has these sequence of values: [(1,0.5), (2,0.7), (3, 0.8) (4, 1)]; this means that the appliance class a was correctly classified 50% of cases (also known as recall); 70% was selected as first and second options; 80% as first, second and third options; and, at the worst case, it was selected between the fourth best options. This gives an idea about how similar this class is with regard to other classes, and the number of classes alike as well.

Figure 4.9 presents the worst case scenario obtained from our experiments. As can be observed at the values corresponding to ranking no 1, all the appliances are correctly classified with a rate higher than 65%, except for the heater whose ranking is 7.6%. Whether considering the ranking no 2, the classification rate for the heater significantly increases, reaching the 80%. This means that 92.4% of the heater instances look like another class and, therefore, the heater class has similar *PQD* signatures. From a manual inspection of the CRC classification, we noticed that this other appliance class is the fan (i.e 10 out of 13 heater test instances are classified as Fan class as first option). This itself slightly affects to the fan precision decreasing its performance. Thus, the general model for fan appliances also models the heater and this overlapping is causing the recall performance degradation. Conversely, this overlapping does not affect to the fan recall performance because the number of training instances for the fan class is large enough to avoid this effect.

We compare with the study in [Gao et al., 2015] where several classifiers are evaluated considering several features over the PLAID dataset. Thus, from every instance in the dataset, a representative current and voltage cycle are extracted. During the steady state, 100 cycles are extracted, synchronised and averaged along samples over the same relative time occurrence in the cycle. Then eight features are evaluated: current (i.e. the raw current from the representative cycle), real and reactive power, harmonics (i.e. up to the 21st harmonic), quantised (i.e. the normalized current amplitude digitalised in 20 bins),

VI image (i.e. VI feature normalized and recorded in an binary image), PCA current and quantised (i.e. PCA reduction of the raw current and the quantised feature) and PCA IV (i.e. PCA reduction to the VI binary image). Furthermore, the "combined" feature means the use of all aforementioned features for the classifier. Then, five classifiers are evaluated using all the features: k-nearest-neighbors (kNN), Gaussian Naive Bayes (GNB), logistic regression classifier (LGC), decision tree (DTree) and Random Forest (RForest). The metric used in [Gao et al., 2015] and, hence, in Table 4.8 is called accuracy (Acc) and it is defined in (4.8). This metric differs from the one defined in (4.41) and it is evaluated in Table 4.7. It does not take into account the scenario where unknown appliances are plugged and the classifier have to discard samples (i.e. it does not take into account the TN and N).

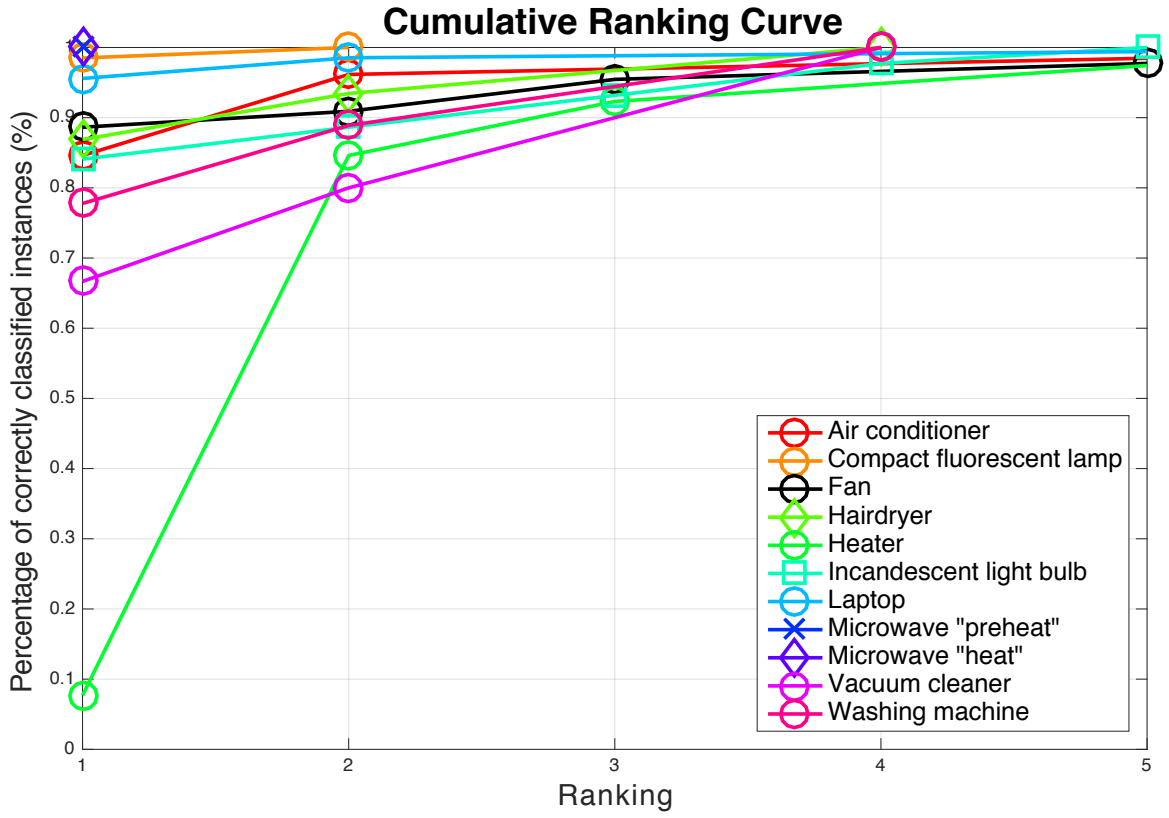


Figure 4.9: Cumulative Ranking Curve of the *PQD-PCA* algorithm performance over the 11 appliance classes on the PLAID data set.

To compare the performance of our study with the one in [Gao et al., 2015], the accuracy *Acc* (4.42) has been evaluated over the PQD-PCA algorithm and the results are presented

in Table 4.8. It can be observed that the best result from [Gao et al., 2015] is using RForest over the VI binary image features (81.75%). Our approach outperforms the best combination of all features (86.03%) with 88% accuracy *Acc*. This is because we used all the information from every instance and they can be precisely compressed and decompressed obtaining a reconstruction error that helps classify. It is worth noting that our training is different. In [Gao et al., 2015], the training strategy is to use 54 out of 55 of the households as training samples and the remaining house for test. This process is repeated 55 times changing the test house at each time. Thereby, it can be guaranteed that the test is performed over different appliances than the training. In our case, we need to guarantee a minimum number of samples per appliance and that is why we randomly split the data in 60% for training and 40% for test without considering the household origin.

Feature	kNN(1)	GNB	LGC	DTree	RForest	PCA error reconstruction
Current	75.98%	61.73%	69.83%	70.67%	76.26%	–
Real/Reactive	55.4%	27.19	29.14%	49.07	51.58	–
Harmonics	45.25%	18.72	30.45%	42.18%	49.63%	–
Quantised	60.06%	57.17%	60.06%	73.09%	80.63%	–
VI image	78.96%	51.96%	74.49%	76.07%	81.75%	–
PCA Current	44.13%	52.14%	46.37%	48.14%	45.07%	–
PCA Quantized	24.30%	18.06%	11.08%	25.98%	27.28%	–
PCA IV	69.93%	60.34%	64.53%	70.67%	77.65%	–
Combined	69.93%	59.22%	49.44%	74.49%	86.03%	–
PQD power trajectories	–	–	–	–	–	88%

Table 4.8: Extended comparison in accuracy from [Gao et al., 2015] with the PQD-PCA performance.

$$Acc = \frac{\text{number of correct preditions}}{\text{total number of predictions}} \in [0, 1] \tag{4.42}$$

Figure 4.10 draws the confusion matrix of the RForest classifier over the VI binary image features from [Gao et al., 2015] obtained from the python code in [Jingkun, 2015]. This figure also reflects the diversity between fridge instances depicted in Figure 4.8. This leads to a lower performance classifying the fridge shown in the confusion matrix: 14 instances are correctly classified as fridge and 24 are wrongly classified as something else.

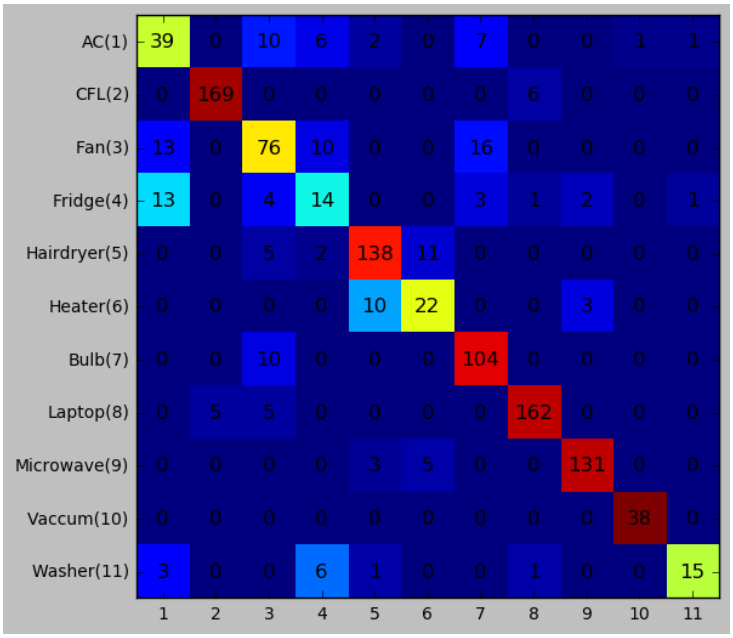


Figure 4.10: Confusion Matrix for RForest using VI binary images in [Gao et al., 2015, Gao et al., 2014].

4.3 Conclusions

This Chapter introduces a novel event-based NILM algorithm oriented to real-time activity monitoring and based on an "unsupervised" approach of disaggregation. Therefore, requirements from Chapter 3 are accomplished. Likewise, a thorough evaluation of the performance has been conducted, using well-known standard metrics and real measurements from two different datasets showing promising results and a high accuracy.

Regarding the event detector, we present a signal processing method where isolation from drop voltages and noises has been implemented. We have optimised our approach to detect the highest number of switching-on events. Our aim is to detect manually operated appliances that, at last instance, will be used to infer the human behaviour. Thus, the human activity can only be assured whenever a switching-on event of these appliances is triggered. After the normalization, the current envelope is extracted using the root mean square value, which is also synchronised by the voltage zero-crossings. Finally, the peak detector described in [Duarte, 2015] has been applied, where the time window W_d (see Subsection 4.1.1) is reasonably the same for other households, and the threshold θ_{th}

is generalised along different mains thanks to the normalization. The performance has been analyzed, by means of standard metrics, over a dataset with 2485 events from 30 different appliance types, providing satisfactory results, which are shown in Figure 4.6 and Table 4.1. The event detector performs a 94% of True Positive while keeping the False Positives less than 0.08% in phase A (with 11 different appliance types and 907 events). In phase B, where the activity significantly increases (26 appliance types and 1578 events), it detects 88% of events against 12% false positives. The AUC metrics provide a more real evaluation whenever the detector cannot be optimised. Thus, the proposed event detection algorithm highly performs at 97% and 94% in phase A and phase B, respectively. Furthermore, a comparison has been carried out, which shows that our proposal has similar results to other relevant studies but with the advantage than only one parameter needs to be set.

The PQD-PCA classifier presents the novelty of using the PQD trajectories as Load Signatures and a learning process carried out offline and, therefore, "unsupervised" disaggregation can be performed. The *PQD* trajectories proved to be a suitable choice, as they minimise the overlapping. They are unique for each appliance type and, hence, the approach can be scalable for more houses without further training. Besides, another important novelty is that PQD trajectories does not need to identify transient and steady states, which is a time-consuming process in Load Identification and they carry much information about both states. Moreover, the PCA classification based on error reconstruction is, for the best of our knowledge, new in NILM. It is inspired from other domains, such as indoor localization and image processing. Instances from the same appliance type have a similar energy distribution and, the devised technique successfully learnt the optimised transformation matrix to compress the energy distribution on the Principal Components. When a new instance arrives, it is compressed using all learned transformation matrix. The one who better compresses the data is the one to label the instance. The training process is conducted offline, by using a dataset. This allows the algorithm to be run over new households without any further training. The PQD-PCA algorithm has been also evaluated according to standard metrics over the PLAID dataset with up to 170 different instances per appliance type. The obtained metrics show the PQD-PCA algorithm successfully classifies 446 different instances from 11 different appliance classes, achieving an averaging value in the recall and the precision metrics of 83.8% and 98.6%. For applications where the precision and the recall have the same importance, the harmonic mean of both (F1 score) gives a value of 90.6%. Elsewise, for applications where usage patterns

need to be learnt (i.e the one sought in this Thesis), precision becomes more important than the recall. For those cases, the F0.5 score is evaluated and the PQD-PCA algorithm performs 98.5%. Besides, a comparison has been conducted where the proposed classifier outperforms other classifiers even though a combination of features is used in other approaches.

Our results are based on datasets that have signals with a bandwidth of 300Hz and, whenever an event occurred, a bandwidth of 1kHz to improve the classification accuracy. This sampling frequency is still acceptable to enable integration of our approach in Smart Meters and third-party devices (see Section 2.1 for more details about Smart Meters sampling capacity), which is a requirement of the Thesis.

This research opens a number of future studies that we would like to point out. Likewise, it would be interesting to evaluate the performance using more downsampled signals. A low sampling rate may reduce implementation cost and the trade-off between performance and cost-implementation could be worth investigating further. Furthermore, it is worth evaluating the performance of the PCA-PQD classifier when more appliance types are involved in order to test if there is overlapped effect. In this term, a bigger dataset is needed where not only more appliance types are present but also more instances for each type. The latter may increase the performance. For instance, in our study, the fridge had to be discarded because of the lack of samples (see Figure 4.8 and discussion in Section 4.2.2).

Chapter 5

Activity Monitoring Proposals

Along this Chapter, we introduce the major contributions of this Thesis regarding the activity monitoring approached from a NILM perspective of view.

Following the discourse presented in Chapter 3, our approaches rather than directly tracking and evaluating ADLs they use the appliance usage patterns to infer the human activity as a proxy for the health and welfare. We have developed three different methods classified into two streams: mono-appliance and multi-appliance monitoring.

Mono-Appliance: The log Gaussian Cox Process (LGCP) model presented in Section 5.1 models the usage pattern of a single appliance and it infers the human activity through this. The uncertainty is encoded into the usage pattern, which also takes into account daily and weekly fluctuations. An alarm is triggered whether the monitored person deviates from his normal routine. The monitoring of a single appliance is preferable to the multi-appliance monitoring in those cases where a very strict routine is followed. Thus, the complexity of merging several appliance usage patterns is not necessary as we have a very reliable appliance usage pattern.

Multi-Appliance: The Gaussian Mixture (GMM) Model and the Dempster-Shafer Theory (DST) model present our approaches for monitoring and merging several appliance usage patterns. There are more information to support decisions than in the Mono-Appliance approach but it is more influenced by the non-deterministic human behaviour and, consequently, more prone to false alarms. Therefore, appliance usage patterns must be properly merged to obtain a good index that scores the human activity. The GMM model is based on a Bayesian probability union whereas the DST presents a generalization of this theory and encodes uncertainties.

We have focused our evaluations on single pensioner households because we believe this is the community that could benefit most from our studies. Our two main reasons are:

(i) The usage pattern of a single occupant is more stable and predictable than several people living together.

(ii) It is crucial to intervene as early as possible when elderly people are living alone as no other monitoring is possible.

5.1 Log Gaussian Cox Process

The study presented in this Section, as well as the experimental results in Subsection 5.4.1, has been developed in the *AIC ("Agents, Interaction and Complexity") Research Group* from *University of Southampton,* under the supervision of Prof. *Alex Rogers* and *Dr. Oliver Parson,* as a collaboration between *University of Alcalá (PhD program)* and *University of Southampton.* The outcome of this collaboration has been published in [Alcalá et al., 2015a]. The first author, José Alcalá, is responsible for the writing (i.e. this Section and Subsection 5.4.1) and the experimental results, whereas all authors devised the LGCP technique for activity monitoring. We would like to remark that, although applicable to other countries, produced results in this Section are particularized to the United Kingdom.

In this Section, we propose the used of the smart meter data to monitor the Activities of Daily Living of single elderly residents, which is used as a proxy to the elderly resident's health. This proposal is based on the monitoring of a single appliance. Figure 5.1 shows an schematic of the workflow. The first diagram block applies a NILM algorithm to disaggregate the power consumption of the single chosen appliance. Then, the event detector draws the occurrences, powering-on events of such appliance. We use the powering-on events because it is the only moment we can guarantee the interaction of the human with the appliance and, hence, those are the events we want to learn for our usage pattern model. In the Usage Pattern block, we model daily and weekly appliance usage with a periodic log Gaussian Cox process, this allows to automatically learn the usage pattern for each household. This model helps predict the appliance usage with a measure of uncertainty, which in turns helps detect unusual patterns and trigger alerts that may require

an intervention (e.g. warning that on 90% of previous days the kettle would have been used by this time).

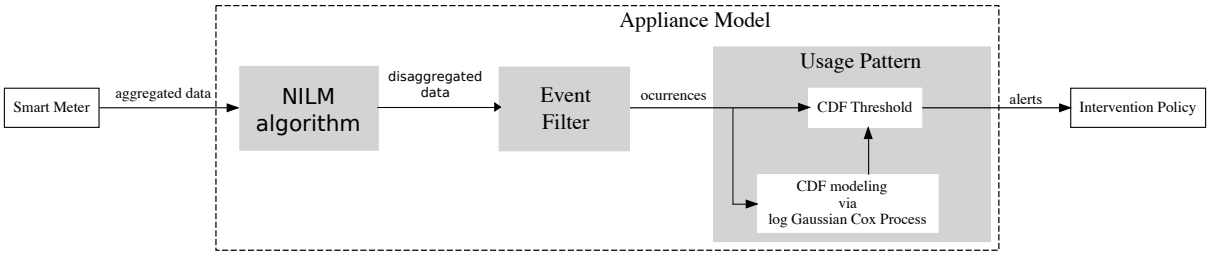


Figure 5.1: LGCP Usage Pattern Model for one appliance.

5.1.1 Selection of a single appliance with a strong routine

The monitoring of a single appliance to infer the human activity can be more beneficial than the multi-appliance monitoring approach in those cases where the usage pattern routine is very strict and short-time intervention is desired. Thereby, there is no need to tackle with more erratic appliance usage patterns that can blurry the real human behaviour.

The kettle has been chosen as an example for this study and for the evaluation in Section 5.4.1 due to the following reasons:

First, kettles are present in almost all households in the UK. Thus, Figure 5.2 depicts the 30 most common appliances in UK homes, taken from the Household Electricity Survey (Appendix A). The kettle is represented as the second most common appliance in the country. This dataset is very representative of the routine in UK because it entails 250 households from this country.

Second, kettles play a main role in the daily routine of UK households and they are regularly used. Especially those of the elderly residents.

Third, kettles exhibit standard power consumption and non-linearities, which make them easily detectable by NILM algorithms.

Although this proposal can model other appliances, none provides the same combination of benefits of the kettle. For instance, major appliances (e.g. such as the fridge and the

heater) are present in many households and they are relatively easy to detect [Parson et al., 2014], although their automatic operation means their usage is not indicative of the routines of the household occupants. Alternatively, washing machines are also present in many households and easy to detect, however, they are used infrequently, which makes them less useful to provide early interventions whether normal routines are not followed.

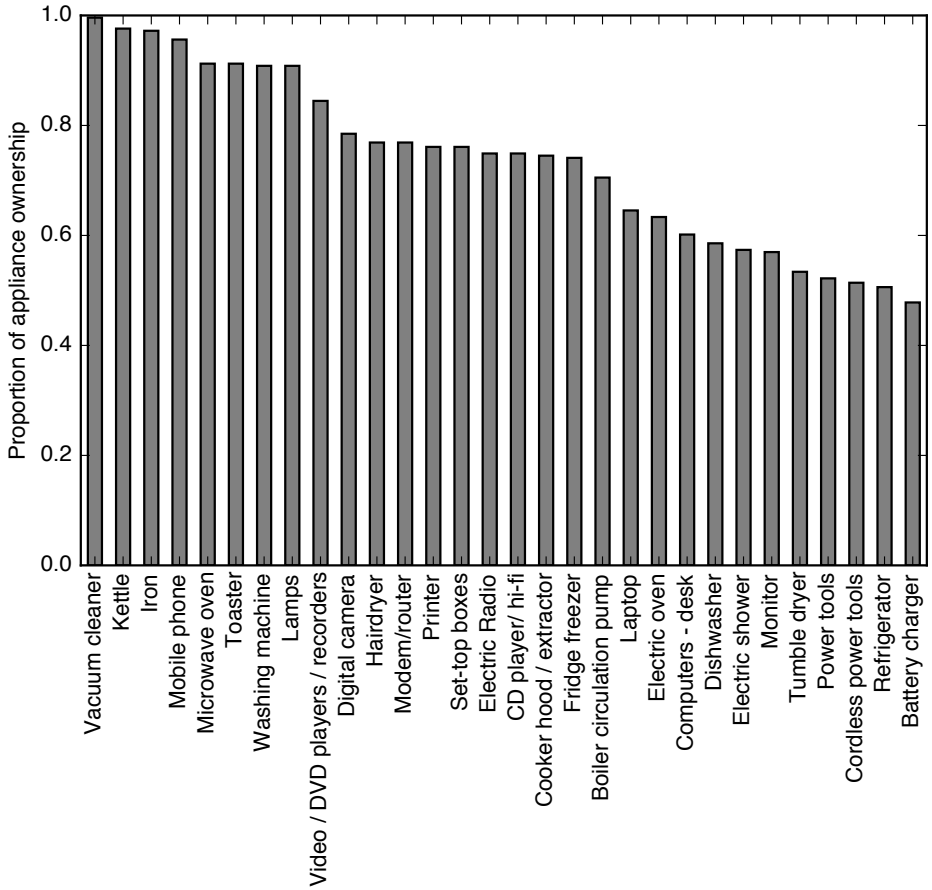


Figure 5.2: Appliance ownership in the UK [Alcalá et al., 2015a].

The routines followed by elderly single-occupant households are unique to each individual household, and as such must be learned individually. Figures 5.3 and 5.4 show histograms of kettle usage over a day, week and year for two households in the HES data set. This dataset is the largest one concerning electricity consumption and single elderly residents (see Appendix A). It should be noted that the gap in the histogram in Figure 5.3.(c) is due to missing sensor data. It can be seen that both households exhibit a clearly daily

routine, though both routines are subtly different. In addition, it can be seen that the house in Figure 5.3 shows no difference in routine between different days of the week, while the house in Figure 5.4 shows a clear difference, with frequent kettle usage from Monday to Thursday, and reduced usage from Friday to Sunday. From a manual inspection of the data, we found that 31 out of the 32 houses present some daily periodicity, while 25 out of 32 houses present weekly periodicity. Last, similar to the daily routines, it can be seen that the kettle usage varies over the year for both households, although again the routine is unique for each household. These are exactly the routines that our approach learns for individual households by considering kettle uses detected from smart meter data.

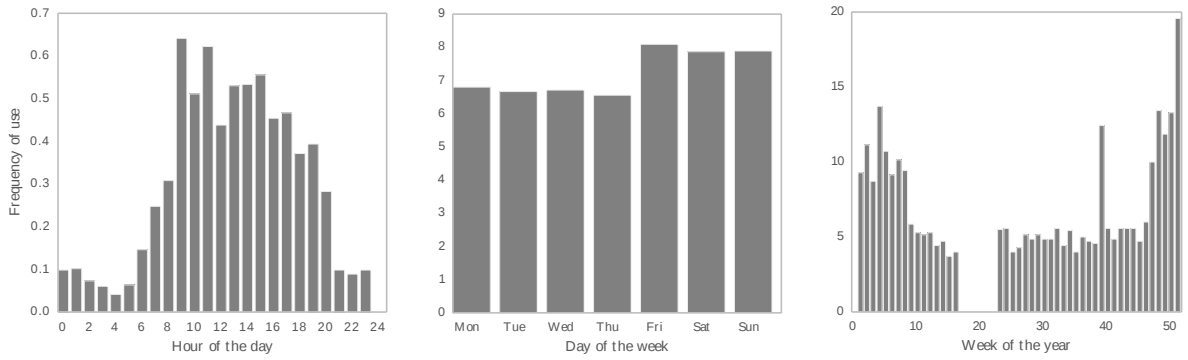


Figure 5.3: Frequency of kettle usage in house no. 101017 from HES data set: (a) per hour of day; (b) per day of week; and (c) per week of year.

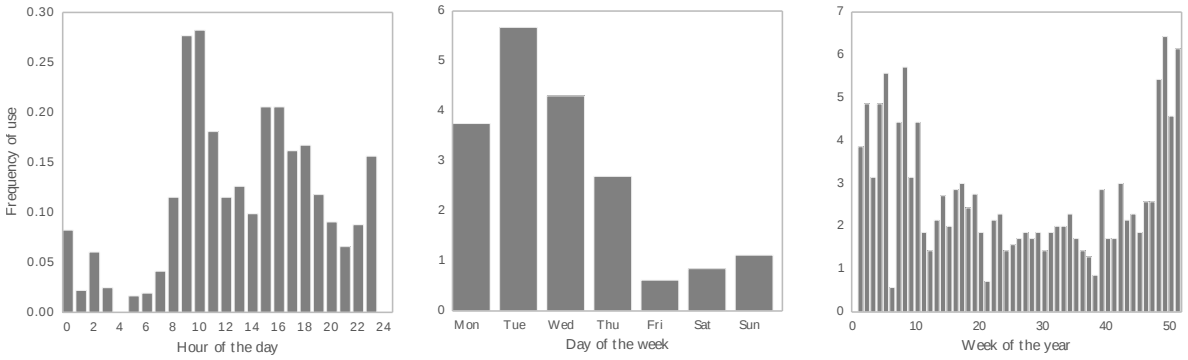


Figure 5.4: Frequency of kettle usage in house no. 102003 from HES data set: (a) per hour of day; (b) per day of week; and (c) per week of year.

5.1.2 The LGCP model description

Regarding the log Gaussian Cox process, this modelling allows arbitrarily complex usage patterns to be learned from point data. Furthermore, it provides a model to encode daily and weekly periodicity in the usage model. The model also provides estimates of the uncertainty around each prediction. It is important to bear in mind that the choice of the appliance is crucial to gain an insight into the human behaviour behind the usage pattern. The appliance should be one that is manually operated, and also exhibits daily and predictable use. We now describe the model in more detail.

The list of appliance uses is mapped for a given household, as produced by the disaggregation model, such that only time of week information is retained. This produces a temporal point pattern covering a 7 day period, as shown in Figure 5.5. We assume that the temporal distribution of these points is generated by a random process with an intensity function $\lambda(x)$, where x is a continuous variable representing a time throughout the week. Furthermore, if we bin the data into fixed length bins, the expected number of appliance uses is $E[\lambda(X)]$ and the number of points within that bin is distributed by a Poisson distribution [Cressie and Wikle, 2011]:

$$\rho(E[\lambda(X)]) = \text{Poisson}\left(\int_X \lambda(X)\right) \tag{5.1}$$

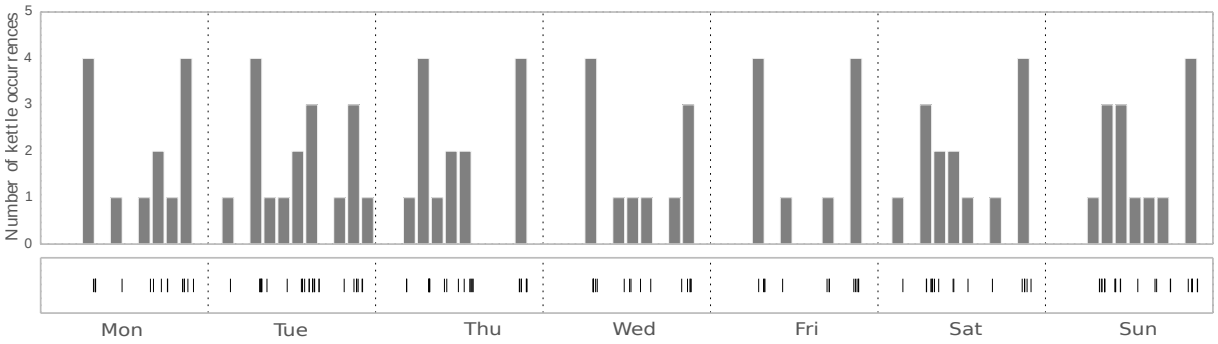


Figure 5.5: Occurrences and histogram of kettle usage during a week.

In fact, the non-homogeneity of the process over the whole week is due to a time-dependency of λ that follows an unknown function and is itself generated by another

stochastic process. This double stochastic process is known as a Cox process [Cox and Isham, 1980]. Therefore, it is necessary to learn the hyper-parameters of this second stochastic process. To do so, we use a non-parametric statistical method referred as the log Gaussian Cox process. This method uses a Gaussian process to learn the temporal point processes [Møller et al., 1998]. In this case, the log of the intensity function is assumed to be generated by a Gaussian process:

$$z(x) = \log(\lambda(x)) \sim \mathrm{GP}(m(x), K(x, x')) \tag{5.2}$$

where:

$$m(x) = E[z(x)] \tag{5.3}$$

$$K(x, x') = \mathrm{Cov}(z(x), z(x')) \tag{5.4}$$

Equations (5.3) and (5.4) are the mean and the kernel function of a Gaussian process and describe the Gaussian process regression, while x is a variable in the temporal points and x' refers to the prediction point. We also want to use a kernel that ensures the model predictions will change smoothly over the course of the day, while also allowing both the daily and weekly periodicity to be explicitly encoded in the model. To do so, we use two periodic-exponential kernels, each of which have three hyper-parameters:

(i) The variance σ that represents the strength of the dependency on other temporal points.

(ii) The length-scale ℓ that represents a window in which the points have the greatest effect on the regression.

(iii) The period ρ which represents the daily or weekly repeating patterns in the temporal point process:

$$K(x, x') = \sigma^2 \exp\left(\frac{-2\sin^2(\pi|x - x'|/\rho)}{\ell^2}\right) \tag{5.5}$$

It should be noted that a yearly periodicity (e.g. seasonal variations) could be also trivially included in this model. However, the HES data set (the used for our evaluation) only monitors three elderly single occupant households for one year, while the remaining households were monitored for one month. To the best of our knowledge, there is no other longer historical data of single pensioners energy consumption than this one. As such, we focus on only daily and weekly repeating patterns.

The next step is to obtain a posterior distribution $\rho(z|X_i)$ to allow predictions to be made. This is a computationally complex problem which requires the integration of the Gaussian prior (5.2) over a Poisson likelihood distribution. We solve this equation by using the Laplace approximation [Friston et al., 2007], which provides an efficient approximation at a small cost in accuracy. We implement the log Gaussian Cox process using the GPy Gaussian Process framework in python [GPy, 2012].

Figure 5.6(a) shows the posterior distribution of kettle usage over the week, while Figure 5.6(b) shows the covariance function of the relationship between midnight on Monday and each other point during the week. It can be seen that points close to each other in time of day share a strong correlation via the exponential part of the kernel, while points that occur at a similar time but on different days also share strong correlation via the first periodic kernel. Similarly, points that are separated by exactly one week also share a strong a correlation via the second periodic kernel.

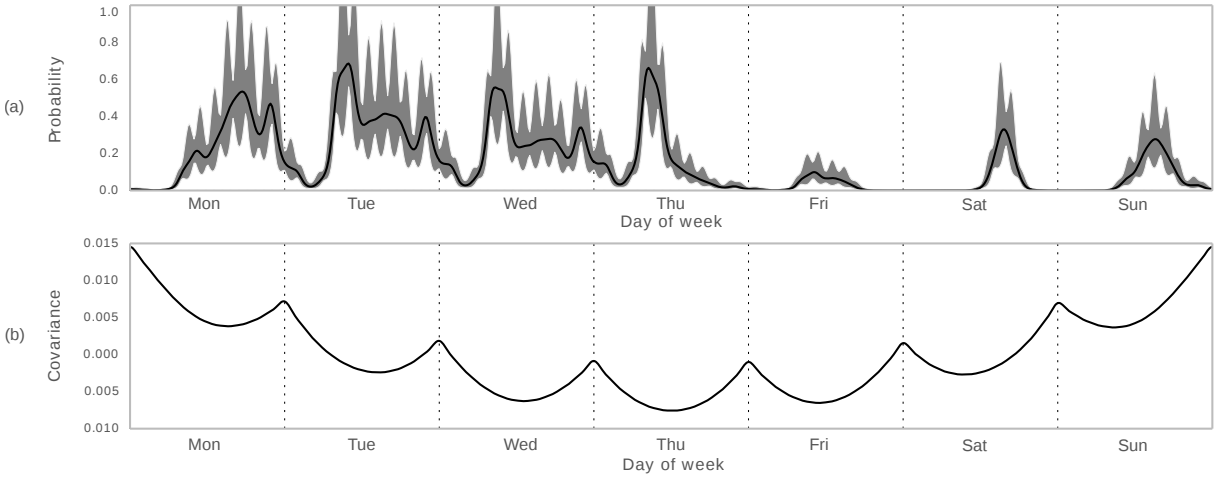


Figure 5.6: (a) The probability of a kettle being used at each point in the week for a single household. (b) The covariance function between each point in the week and midnight on Monday.

We initialise the kernel with a length-scale of six hours and a variance of one hour, before optimising these parameters given the disaggregated appliance usage data, while holding the periodicity constant at both one day and one week. This ensures that both the model parameters and the regression fit is learned for each household. Figure 5.7(a) shows an example regression after optimising the model parameters and returning the exponential shape to the intensity function for the kettle in a single household. Figure shows the mean of the regression and the 95% confidence interval. It can be seen that the usage pattern for this household shows a high chance of the kettle being used at around 7am when the occupant wakes up and another at roughly 9pm before the occupant goes to bed, and also three smaller peaks during the mid-morning, lunchtime and mid-afternoon periods.

Finally, we transform the regression fit to represent the cumulative probability that an appliance a was used at least once by time of day t. This is calculated using:

$$p(a_{0:t}) = 1 - \Big[(1 - p(a_{0:t-1}))(1 - p(a_t))\Big] \tag{5.6}$$

Figure 5.7(b) shows an example of this cumulative function for the same household as in Figure 5.7(a) over the course of one day. It can be seen that, while the probability of kettle usage within each interval is not necessarily high, the probability of a single usage quickly accumulates over the day. For instance, the graph shows that there is a 80% of probability of the kettle being used at least once by 11am, which could be a favourable threshold for intervention in the case that the kettle has not been used by this point of the day. Our approach applies this threshold to all households, in order to obtain a unique intervention time for each household. As a result, it is possible to recommend an intervention in any household which would have likely seen the modelled appliance being used by that time of the day, yet no usage was detected by the disaggregation algorithm.

So far, we have described the usage pattern modelling and the intervention policy. The performance of this proposal is presented in Section 5.4.1

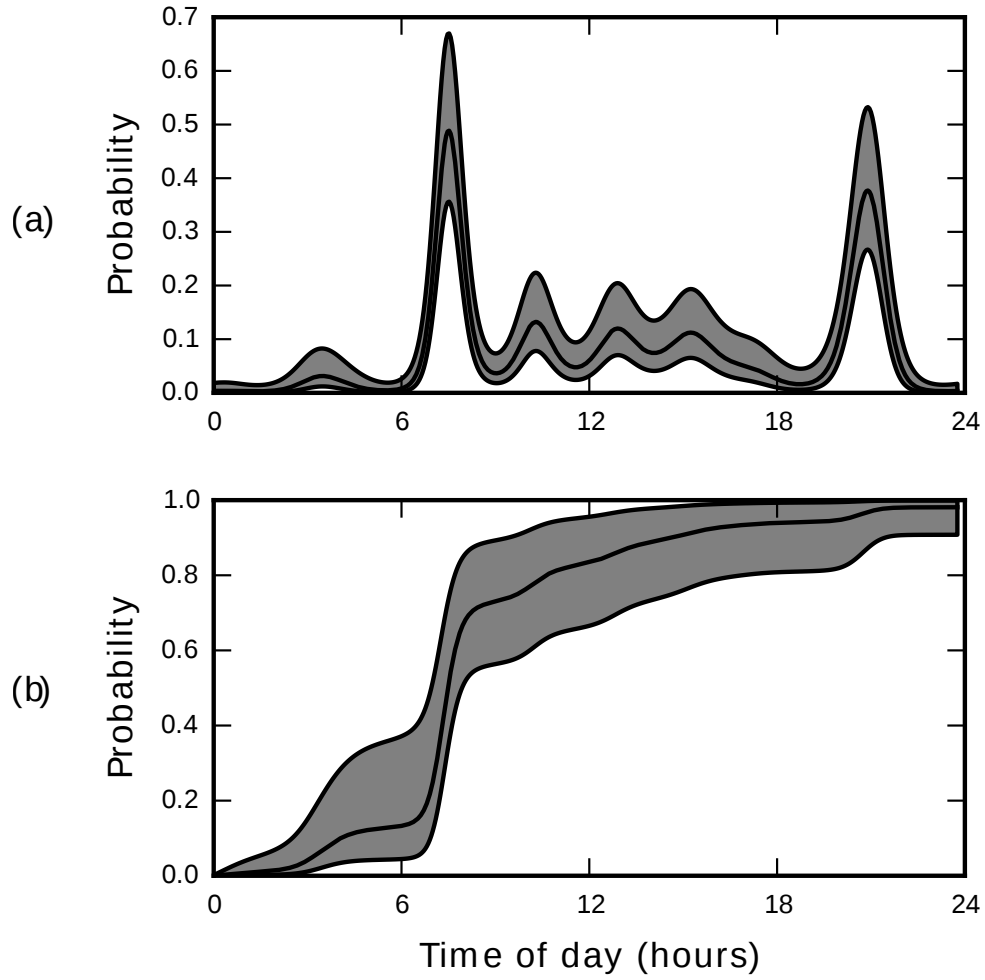


Figure 5.7: (a) The probability of a kettle being used at each point in the week for a single household. (b) The covariance function between each point in the week and midnight on Monday.

5.2 The GMM-Score Monitoring Algorithm

Here, we present a method for learning the usage pattern of different appliances and scoring the underlying human activity based on the Bayesian Probability. This approach uses Gaussian Mixture Models (GMM) to learn repetitive patterns. For instance, a most repetitive pattern can be found in a day such as the use of 5 appliances within an interval of a hour (around 8am at morning) and then, another 3 different appliances around 17:30pm within an hour. We then look up for that pattern in every day and give a score of how similar is that day according to the pattern. In case that a day follows the pattern, the

score will be close to 1; otherwise, if it deviates from what we consider a "normal" day, the score will tend to 0. This allows to trigger an alarm for the caregivers in case that something happens out of the ordinary.

In order to infer the human activity, we focus exclusively on the powering-on events of manually operated appliances. This is the only time we can guarantee that the tenant is interacting with the appliance. Then, the event data are divided into training data (to model the usage pattern) and test (to evaluate the performance of the algorithm). For instance, Figure 5.8 shows on-power connections for some appliances in a household from the HES dataset (Appendix A) over a week, y-axis differentiates between appliances (one at each horizontal level and the number is the ID in the data set). From a manual inspection of the data, it can be observed how the human activity ins inferred from the events: it decreases over the week and each appliance usage pattern is unique and can be learnt.

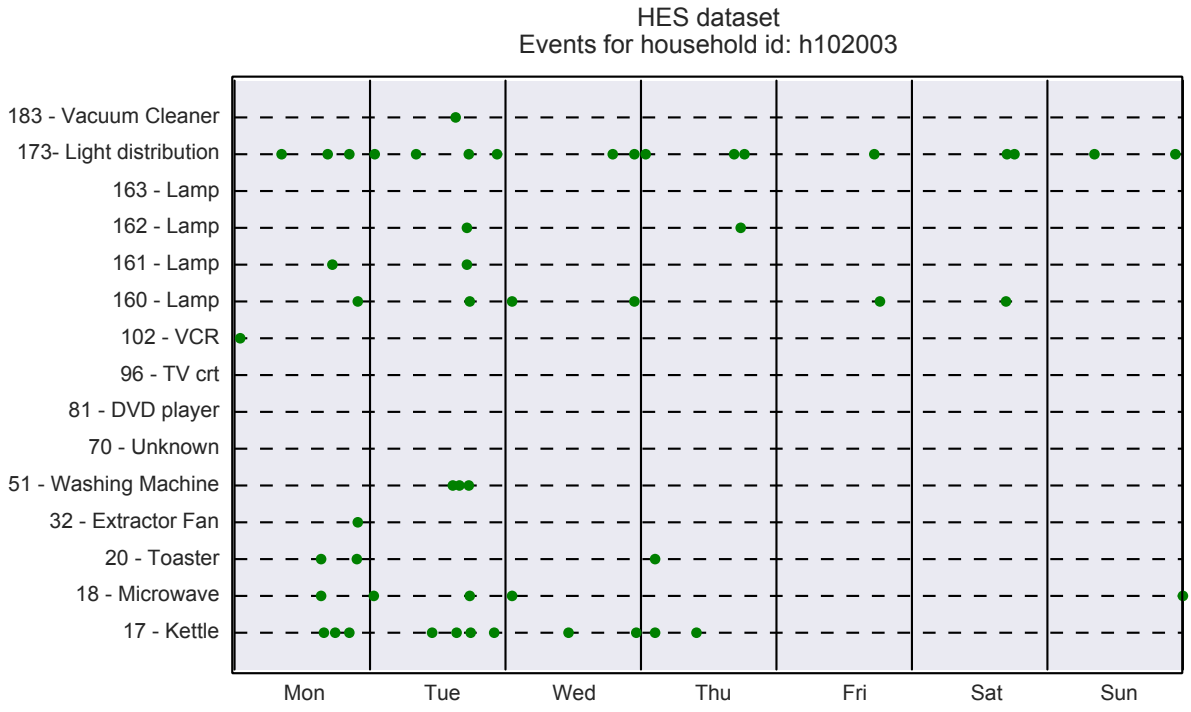


Figure 5.8: A week of switching-on events for manual appliances in household no. h102003 from HES dataset.

5.2.1 Modelling using Gaussian Mixture Models

We use the training data to model the usage pattern with Gaussian Mixture Models (GMM). Initially, a filtering is needed in order to prevent the model to learn any possible erratic behaviour. Thus, days with a number of events less than 10 (empirically we observed the mean number of events is 30) are discarded. The next step is, for each appliance, to map all events in the training data into a single week. This is done by keeping only the day of the week, hour and minute information for each event. Mapping along a week is desired to accurately model the usage pattern: it was proved in Section 5.1 (see Figure 5.3 and Figure 5.4) that not only the time of the day but also the day of the week affects the usage pattern. Hence, a temporal sequence of points over a week space minutely sampled is obtained. We filter this point sequence by appliance and day of the week, then a Gaussian Mixture Model is applied to learn the usage pattern of that specific appliance during that specific day of the week.

For every Gaussian Mixture Model eight components have been used. This number has been chosen empirically: using less components leads to less flexibility and prevents some usage pattern model to adjust to the 24 hour pattern of the appliance; whereas a higher number of components could lead to overfitting. Once the number of components is set, a Expectation-maximization algorithm is used to fit parameters: mean, covariance and weights. This is an iterative algorithm that tunes the three parameters (i.e. the mean, the covariance and weights) while the samples arrive [Wikipedia, 2016a]. We have used the scikit-learn python module to model the GMMs and to tune them using the Expectation-maximization algorithm in [Scikit, 2016].

Figure 5.9 shows the usage pattern model for three main appliances in a household: a lamp located in the bedroom, a microwave located in the kitchen and a DVD player in the lounge. The major uses are at 19:00pm and 1:00am for the lamp, 1:00am and 16:00pm for the microwave and 14:00pm and 21:00pm for the DVD player. Out of those intervals the density of temporal points decays so the probability density function does.

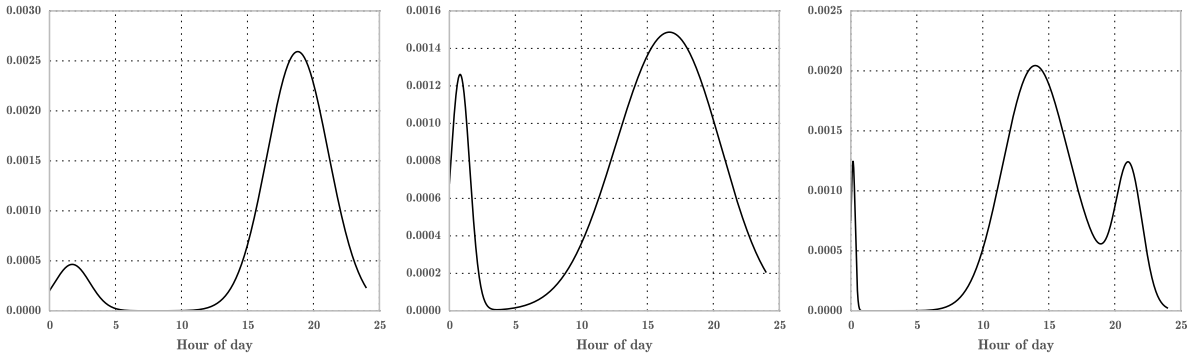


Figure 5.9: Examples of Gaussian Mixture Models for three different appliances during Mondays: (a) GMM for a lamp; (b) GMM for a microwave; (c) GMM for a DVD player.

5.2.2 Scoring the activity

Given some temporal events for a day, we now want to score how usual the day is based on the usage pattern models obtained with GMMs (Section 5.3). Therefore, the likelihood of receiving that sequence of events is evaluated. Probabilities are evaluated in three steps: first, the interval probability of each event within the day, as in (5.7); then, the probability intersection of all events for every appliance, which provides a measure of how usual the appliance usage is, (5.8); and finally, the probability union of the intersection probabilities for all appliances (5.9).

Firstly, events are grouped by appliances and, for every event E_j with a time of arrival T_j, where $1 \leq j \leq m$ is the number of events for each appliance, we evaluate its probability of occurring within an interval of time ($T_1 \leq T \leq T_2$) as:

$$P(E_j) = p(T_1 \leq T_j \leq T_2) = \int_{T_1}^{T_2} GMM(t)dt \tag{5.7}$$

Where GMM in the Gaussian Mixture Model evaluated in Section 5.3 (each day of the week has its corresponding GMM) and ($T_1 \leq T_j \leq T_2$) is an interval of an hour around the arrival time, T_j. Sometimes, there are intersections between events, occurring within the same interval of time. Therefore, we need to take into account this to evaluate the probability $P(A_i)$ which is our measure of how usual the usage of the appliance i is, where $1 \leq i \leq n$ is the number of appliances. Following the equation:

$$P(A_i) = P(\cup_{j=1}^{m} E_j) = \sum_{j=1}^{m} P(E_j) - \sum_{j \neq k} P(E_j \cap E_k) \tag{5.8}$$

Then, the probability union between appliances is evaluated, which states for the probability of having a normal behaviour, according to the use of all the appliances. The equation applied is:

$$P(\cup_{i=1}^{n} A_i) = 1 - (P(A_1 \cup \cdots \cup A_n)^c \tag{5.9}$$

Likewise, (5.9) can be decomposed as in (5.10) because each appliance is modelled with a different GMM and, in terms of probabilities, they are independent process.

$$P(\cup_{i=1}^{n} A_i) = 1 - P(A_1 \cup \cdots \cup A_n)^c = 1 - P(A_1^c) \ldots P(A_n^c) \tag{5.10}$$

Where $P(A_i^c)$ is the complementary $1 - P(A_i)$.

Therefore, in (5.10) is eventually evaluated to score how usual a period of time has been (e.g. 12h, a day, etc). It is worth noting that the use of the probability union instead of the intersection gives more flexibility in the mix of usage patterns. In this case, if a event is missed from an appliance, which is quite common due to the human behaviour, this probability does not drop to zero; instead, it will simply decrease. Therefore, if the probability drops below a certain value, it means that many appliances are not following their pattern and a alert should be triggered.

5.3 The DST-Score Monitoring Algorithm

The Dempster-Shafer Theory (DST), also known as the Evidence theory, is widely used in sensor data fusion. It is a generalisation of the Bayesian Theory [Rakowsky, 2007], where, instead of probability distribution functions, belief functions or mass functions are handled. Our approach considers the disaggregated consumption data from each appliance as a reading from an independent sensor, which, using a mass function, provides a belief about the *normality* in the usage pattern for that specific appliance. By using the Evidence

theory, we merge all beliefs, thus we obtain a general belief about the *normality* in the use of all appliances. This allows to score daily routines and to detect deviations.

The hypotheses 2^{Ω} in the DST are increased, compared to the hypotheses Ω in the Bayesian Theory. For instance, for a presence detector, the hypotheses in the Bayesian domain is the universe $\Omega = \{h_1(Presence), h_2(Non-presence)\}$. Unlike, in the DST domain the hypotheses are the power set $2^{\Omega} = \{\emptyset, h_1(Presence), h_2(Non-presence), H_x\}$, where $\emptyset$ is the empty set and H_x is a subset for all the potential combinations, in this case $H_x = \{h_1(Presence) \cup h_2(Non-presence)\}$. Thus, the parameter x models the uncertainty whenever there is a lack of information. This is hardly represented in the Bayesian Theory by assigning a probability of 0.5.

Applying DST to our case, the power set is as denoted in (5.11). Then, a belief mass function for every element h_x in the power set 2^{Ω} has to be defined according to (5.12). A belief mass function is formally called Basic Belief Assignment (BBA) whether it meets (5.13) and (5.14). The BBA for a certain hypothesis $m(h_1(\text{normal pattern}))$ represents the proportion of evidence that support the claim that the actual state is $h_1(\text{normal pattern})$.

$$\begin{aligned} 2^{\Omega} = \{&\emptyset, h_1(\text{normal pattern}), h_2(\text{abnormal pattern}), \\ &h_3 = \{h_1 \cup h_2\}\,(\text{normal or abnormal pattern})\} \end{aligned} \tag{5.11}$$

$$m : 2^{\Omega} \rightarrow [0, 1] \tag{5.12}$$

$$m(\emptyset) = 0 \tag{5.13}$$

$$\sum_{A \subseteq 2^{\Omega}} m(A) = 1 \tag{5.14}$$

Belief (or support) and plausibility define the minimum and maximum of the confident interval for the hypothesis A, respectively. Hence, the probability for the current state to be A is given by an uncertainty interval expressed in (5.15), where the belief *bel* and the

plausibility pl are defined in (5.16) and (5.17). The belief in A is evaluated by adding the belief masses from all the subsets of A, whereas the plausibility is the sum of all belief masses of subsets where there is an intersection with the set A. For instance, the belief of a pattern to be normal or abnormal is $bel(H_x = h_1 \cup h_2) = m(h_1 \cup h_2) + m(h_1) + m(h_2)$.

$$bel(a) \leq P(A) \leq pl(a) \tag{5.15}$$

$$bel(A) = \sum_{B|B \subseteq A} m(B) \tag{5.16}$$

$$pl(A) = \sum_{B|B \cap A \neq \emptyset} m(B) \tag{5.17}$$

Let consider two appliances whose hypotheses are to be merged. For that, the Dempster's rules of combination are applied as it follows for the hypothesis A.

$$m_{1,2}(\emptyset) = 0 \tag{5.18}$$

$$m_{1,2}(A) = (m_1 \oplus m_2)(A) = \frac{1}{1-K} \sum_{B \cap C = A \neq \emptyset} m_1(B) m_2(C) \tag{5.19}$$

$$K = \sum_{B \cap \emptyset} m_1(B) m_2(B) \tag{5.20}$$

Where (5.18) denotes that the mass function assigned to the empty set must be zero; and (5.19) defines the new mass function for the hypothesis A resulting from merging appliances 1 and 2, where $1 - K$ is a normalization factor. Finally, (5.20) describes the amount of conflict between the two mass sets.

For instance, considering two appliances X and Y, a certain observation window T_i, and the power set $2^{\Omega} = \{\emptyset, h_1, h_2, h_3\}$, their BBAs are defined in Table 5.1. We apply the Dempster's rule of combination, (5.18), (5.19) and (5.20) to merge the observation of both appliances. The resulting fusion of BBAs is depicted in Table 5.2. The combination of

$h_1 \cap h_2$ and $h_2 \cap h_1$ are conflicts because the current state cannot be *normal* and *abnormal* at the same time. Consequently, we assign them to the empty set and the proportion of belief corresponding to the conflict states is assigned to K as in (5.20). Lately, the parameter K is distributed to the remaining compatible hypotheses according to (5.19). Table 5.3 is obtained by adding equal hypotheses in Table 5.2, normalising by K and rearranging. The belief and plausibility are evaluated in (5.16) and (5.17) from the mass functions of the fusion and presented in Table 5.3. The resulting *bel* and *pl* values show that, for the beliefs of appliances X and Y in Table 5.1, the probability of a pattern to be *normal* can be found in the interval $0.89 \leq P(h_1) \leq 0.92$, defined by 5.15 within the window T_i.

BBAs for X in T_i	$\mathbf{2^\Omega}$	**BBAs for Y in T_i**
$m_x(h_1, T_i) = 0.8$	$\boldsymbol{h_1}$	$m_y(h_1, T_i) = 0.6$
$m_x(h_2, T_i) = 0.1$	$\boldsymbol{h_2}$	$m_y(h_2, T_i) = 0.2$
$m_x(h_3, T_i) = 0.1$	$\boldsymbol{h_1 \cup h_2}$	$m_y(h_3, T_i) = 0.2$

Table 5.1: Basic Belief Assignments for appliances X and Y (example).

$\cap$	$\boldsymbol{X_{h1}}$	$\boldsymbol{X_{h2}}$	$\boldsymbol{X_{h3}}$
$\boldsymbol{Y_{h1}}$	$h_1 \cap h_1 = h_1$	$\emptyset$	$h_3 \cap h_1 = h_1$
$\boldsymbol{Y_{h2}}$	$\emptyset$	$h_2 \cap h_2 = h_2$	$h_3 \cap h_2 = h_2$
$\boldsymbol{Y_{h3}}$	$h_1 \cap h_3 = h_1$	$h_1 \cap h_1 = h_1$	$h_3 \cap h_3 = h_3$

Table 5.2: BBA fusion for appliances X and Y.

$\mathbf{2^\Omega}$	$\boldsymbol{m_{T_i}}$	$\boldsymbol{bel_{T_i}}$	$\boldsymbol{pl_{T_i}}$
$\boldsymbol{h_1}$	$h_1 \cap h_1 = h_1$	$\emptyset$	$h_3 \cap h_1 = h_1$
$\boldsymbol{h_2}$	$\emptyset$	$h_2 \cap h_2 = h_2$	$h_3 \cap h_2 = h_2$
$\boldsymbol{h_3}$	$h_1 \cap h_3 = h_1$	$h_1 \cap h_1 = h_1$	$h_3 \cap h_3 = h_3$

Table 5.3: Mass functions, Beliefs and Plausibility after fusion.

The example above can be applied to more appliances by accumulating their evidence iteratively. Therefore, considering N appliances, we merge the mass of evidence from appliance no. 1 and no. 2 to obtain a new mass: $\boldsymbol{m}_{T_i}$, as shown in Table 5.3. Then, repeating the process, this new mass is merged with the mass from appliance no. 3, and so on up to the appliance N. This iterative process results in a general belief and plausibility based on the accumulative evidential from the N appliances in a certain time interval T_i. This assures that, over that time, the probability of a pattern to be *normal* is higher than the belief and lower than the plausibility.

Likewise, if we are to evaluate a daily pattern, we should proceed accumulating evidences over the time intervals. Thus, if a daily time $\boldsymbol{T}$ is divided into I periods of time T_i (5.21), a general mass function $\boldsymbol{m}_{T_i}$ for each interval i can be obtained as explained above. These are accumulated following the same process over time, instead of appliances. This results in the belief and plausibility for the performed daily routine within the household.

$$\boldsymbol{T} = [T_1, \ldots, T_i, \ldots, T_I] \tag{5.21}$$

5.3.1 Basic Belief Assignments and Weighing

We model the Basic Belief Functions (BBAs) as follows. For every appliance, a BBA is obtained taking into account the day of the week and time interval T_i of the day. It has been empirically proved that the appliance usage patterns not only vary during hours of the day, but they also do in a significant way depending on the day of the week (e.g. Figure 5.3 and Figure 5.4). Likewise, during the training process, the occurrences for each appliance is binned by the time interval T_i and the day of the week. The number of occurrences in each bin for a certain day of the week is divided into the total number of that specific day of the week in the training set. This denotes the probability of the appliance to be used at that time of the day for that day of the week. Following this process, we obtain similar BBAs to those in Figure 5.10 where the time interval T_i is fixed to three hours. Note, for instance, that here the probability $P(T_i)$, where $T_i \in [9, 12)$ is the time interval of the day when using the kettle between 9h and 12h, is roughly 0.85. Besides, there is another time interval with high probability of using the kettle between 18h and 21h (0.6 approximately). During the test, BBAs are weighed depending on the presence or absence of events within the considered interval T_i as in (5.22), (5.23) and (5.24). Thus, the constant C_o represents a certainty on the probability for an arrival event, whereas the constant C_1 shows the certainty in absence of event. The value of these two constants ranges from 0 to 1, and they depend on the application. More details about the certainty constants will be discussed in Section 5.4.2.2. In order to meet the requirements in (5.12), (5.13) and (5.14), the mass should be defined as in 5.24 to hold the uncertainty produced by C_0 and C_1 in (5.22) and (5.23).

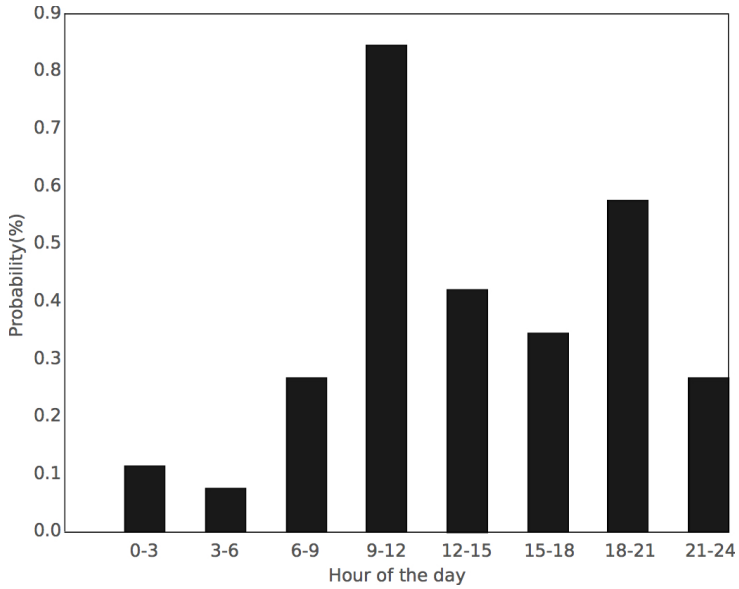


Figure 5.10: . Kettle's BBA on Mondays with a time interval T_i of 3 hours.

$$\boldsymbol{m}_{T_i}(h_1) = \begin{cases} P(T_i) * C_0, & \text{if event} \\ (1 - P(T_i)) * C_1, & \text{if not event} \end{cases} \quad \forall T_i \in [0, 24) \tag{5.22}$$

$$\boldsymbol{m}_{T_i}(h_2) = \begin{cases} (1 - P(T_i)) * C_0, & \text{if event} \\ P(T_i) * C_1, & \text{if not event} \end{cases} \quad \forall T_i \in [0, 24) \tag{5.23}$$

$$\boldsymbol{m}_{T_i}(h_3) = 1 - (\boldsymbol{m}_{T_i}(h_1) - \boldsymbol{m}_{T_i}(h_2)) \tag{5.24}$$

Most criticisms about DST is that the generation of mass functions and basic belief assignments functions is subjective. However, the DST has a main advantage regarding the classical Bayesian theory: it is able to tackle with uncertainties.

For the purpose of our study, it is very important to properly infer the non-deterministic behaviour of humans and this implies to encode uncertainty into the usage pattern of appliances. Thus, the presence of appliance events could be an indicator of somebody´s healthy routine, but its absence does not necessarily imply the opposite. We perform this by using the constants C_0 and C_1 in the aforementioned equation (5.22), (5.23) and (5.24).

The following Section shows the benefits of using the DST instead of the classical Bayesian theory, comparing the performance of the GMM algorithm presented against our DST algorithm.

5.4 Experimental Results

Along this Section, experimental results obtained for the three proposals are presented. We evaluate the performance of the single appliance monitoring using the LGCP algorithm in Section 5.4.1 whereas a evaluation and comparison of multi appliance approaches (GMM and DST models) are presented in Section 5.4.2.

5.4.1 Single Appliance Monitoring: LGCP algorithm

We evaluate the monitoring of usage patterns using the Household Electricity Survey (HES) data set (Appendix A). It contains kettle electricity data that covers 250 households in the UK. We evaluate the LGCP algorithm in 32 households that are known to be occupied by elderly residents living alone.

For each household, the usage pattern is extracted as described in Section 5.1. The same kernel was used for all households, except for the hyper-parameters (variance and length-scale) which were optimised for individual households while the periodicity was constrained to one day and one week. We use a bin size of 120 minutes given the limited usage data that was available.

To evaluate the importance of an adjusted intervention time regarding the kettle usage pattern, Figure 5.11 shows in solid lines the false alarm rate against the hour of intervention for every day of week and for each household (i.e. 224 curves, 32 households by 7 days of week). We define a false alarm at a given time of day as a recommended intervention in a household which has not used the kettle by that time but uses the kettle at a later point during that day. Conversely, a correctly raised alarm is a recommended intervention for a given day in which the kettle is not used at all. Thus, the false alarm rate is the number of false alarm divided by the number of days.

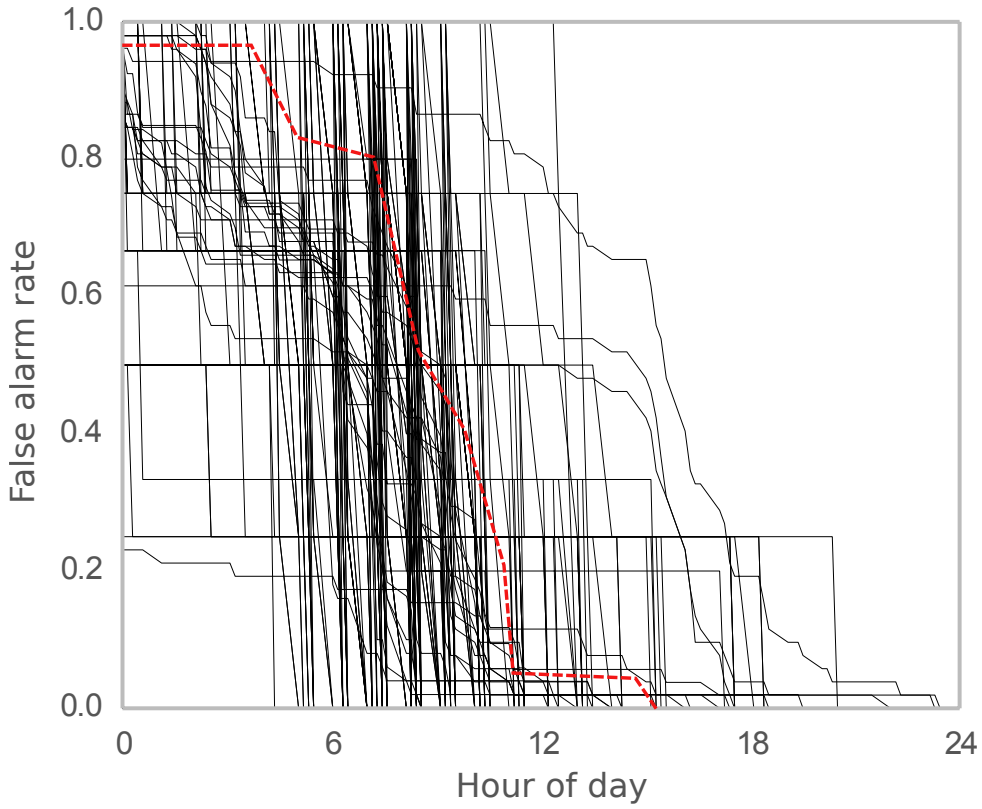


Figure 5.11: False positive count for various intervention times. Each line represents one day of week for one house.

For instance, the dashed red line in Figure 5.11 correspond to a single elderly household on Saturdays. It can be observed that the false alarm rate decreases while the intervention time is delayed on time. Thus, whether an intervention time is set at 6:00am on Saturdays, the false alarm rate is high (i.e. more than 80% of Saturdays a false alarm will be triggered). If we want to keep the false alarm rate on Saturdays bellow than 10%, then the intervention time shall be set at 11:00am. This varies depending on the household and the day of week. Interventions earlier in the day allow the elderly resident to receive care as soon as possible but can only be achieved at the cost of a high false alarm rate, while interventions later in the day are necessary to achieve a low false alarm rate. As a result, the hour of day of the intervention can be determined for every individual house given an acceptable false alarm rate. It can be observed that the false alarm rate drops rapidly below 0.1 for most houses and days of week between the hours of approximately 5:00am and 5:00pm. However, an early intervention within this 12 hour window could be crucial to ensure that an elderly person receives the attention they need as soon as possible. It

is exactly this characteristic that our approach exploits, allowing early interventions in households when a low false alarm rate can be kept, whereas allowing later interventions in houses which use the kettle less frequently or later in the day.

The LCGP monitoring approach is compared against a fixed threshold method of intervention, which will intervene in any household if there has not been any use of the disaggregated appliance before a fixed time of day. This benchmark was chosen as it represents the optimal intervention point when no individual information about household routines are provide. We optimised the intervention time to achieve the earliest possible intervention while averaging a false alarm rate of 10% across all households.

Figure 5.12 shows the intervention time of the LGCP monitoring approach against the fixed-time benchmark across all single occupant elderly households in the HES data set, where the x-axis is the hour of day and the y-axis the day of the week. It can be observed that the fixed-time benchmark is the same for every day of the week (i.e. the dashed black line at 10:00am) whereas the LGCP proposal adapts the intervention time to the usage pattern. In Figure, the grey shade map indicates how many houses have that intervention time at that day of the week. For instance, on Wednesday there are around 10 houses whose its intervention time is 8:00am (i.e. x-axis is 8:00am, y-axis is "Wed" and the colour is black). On average, our approach allows earlier interventions in 25 out the 32 households, relative to the fixed-time benchmark. Furthermore, LGCP monitoring approach intervenes later in the remaining households, in order to ensure the tolerable false alarm rate of 10% in every household, rather than on average across all households as is achieved by the benchmark. As a result, the LGCP monitoring is able to identify the households which really need an intervention, as opposed to those that are just following their normal schedule of using the kettle later in the day.

Finally, it is important to note that the LGCP monitoring aims to provide intuitive information to a friend, relative or health care worker. For example, the system could produce an alert such as: 'on 90% of days, the kettle would have been used by now in this household.' Such alerts could be delivered by text message, allowing the recipient to make an informed decision on what action to take. For instance, if the text message is received by a relative who had taken the elderly resident out for lunch, there is relatively little cost of generating this false positive. Alternatively, if the text message is received by a care worker, the care worker could make a phone call to the elderly occupant to check in, and raise the case as an emergency if no contact can be made. Such a system clearly requires

less resources to monitor elderly residents living independently than making daily phone calls or in-person visits to all such households, as interventions only occur due to unusual deviations from standard routines.

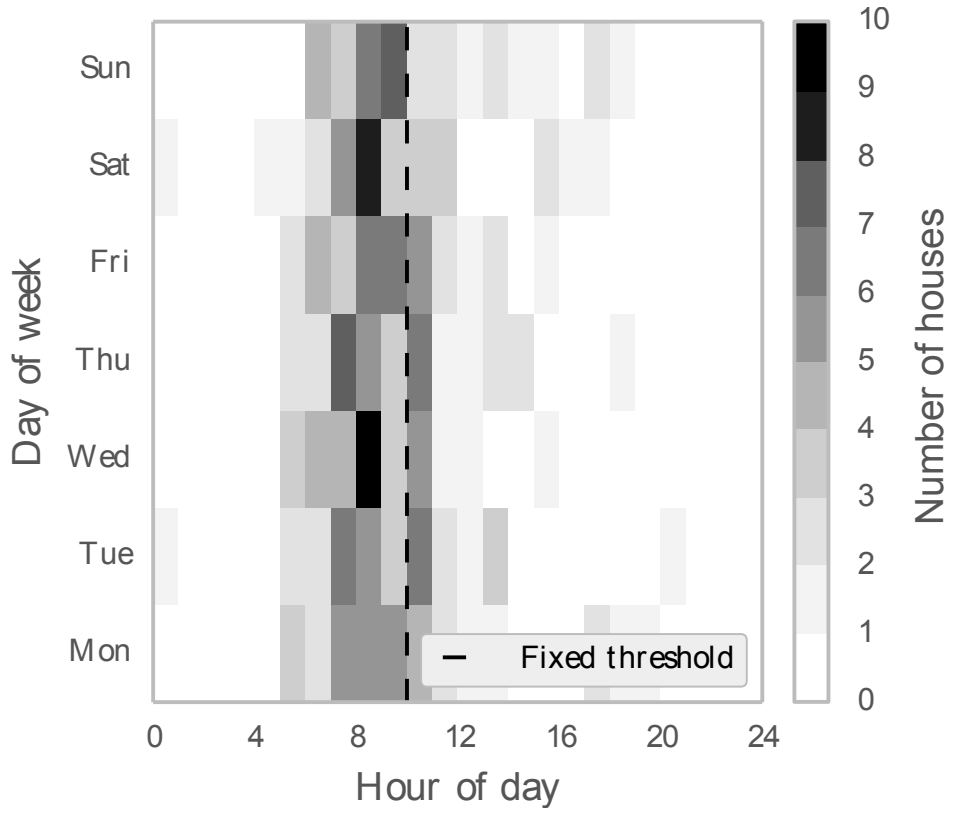


Figure 5.12: Intervention times of our approach (LGCP) against fixed-time intervention benchmark.

5.4.2 Multi-Appliance Monitoring: GMM and DST algorithms

5.4.2.1 Datasets, preprocessing and selection of the training

In order to evaluate the performance of the multi-appliance monitoring algorithms developed in this Thesis, two different datasets have been considered: the HES dataset and the UK Domestic Appliance-Level Electricity (UK-DALE) dataset (Appendix A). Both are real collected data from the aggregated and disaggregated energy consumption from UK households. From the former we use a year data from three single pensioner

households, which are the targeted community in this study; whereas the latter one is a two-year collection data from a family household (two adults, two children and a dog), so it can be useful to enhance the performance of the proposed algorithm over the first community.

The approaches presented here are based on the employment of a NILM algorithm and, therefore, only the disaggregated data from datasets are used. In Chapter 4, we presented a novel NILM algorithm for disaggregation and its performance. However, any NILM algorithm can be used as far as it provides the events and appliance labels, which are the inputs for the GMM and the DST algorithm.

Likewise, the relevant events are extracted from the appliance-level consumption. Relevant events are those whose evaluation can result in the inferring of any human activity inside the household. Consequently, only manually activated appliances with a repetitive pattern over time are considered here, discarding appliances such as the fridge with a continuous and automatic consumption. The relevant events are the recorded timestamp of switching-on for these appliances. Although, the time duration of appliance usages is also a good indication about the routines, this implies a more complex pattern to infer the real human activity, as they could be overlapped. Furthermore, the duration does not really imply any human activity as the appliance could have been left on. Nevertheless, switching-on events do require a human activity and this is the reason why we have chosen this unique events as incomes for our high-level application algorithms (GMM and DST). For instance, Figure 5.13 shows the switching-on events in household no. 101017 from the HES dataset over time. Each spot is an event and the y-axis represents the appliance code, where non-relevant appliances have been already filtered out.

The first 60% of total days are used for training whereas the remaining 40% is for test. It is worth noting that it is expected some deviation from the routine over time. This is why the model should be trained with earlier days than those used for test, in order to learn the starting normal routine. As can be observed in Figure 5.13, if the ratio between training and test samples is less than 60%, then some appliances do not have enough samples for training.

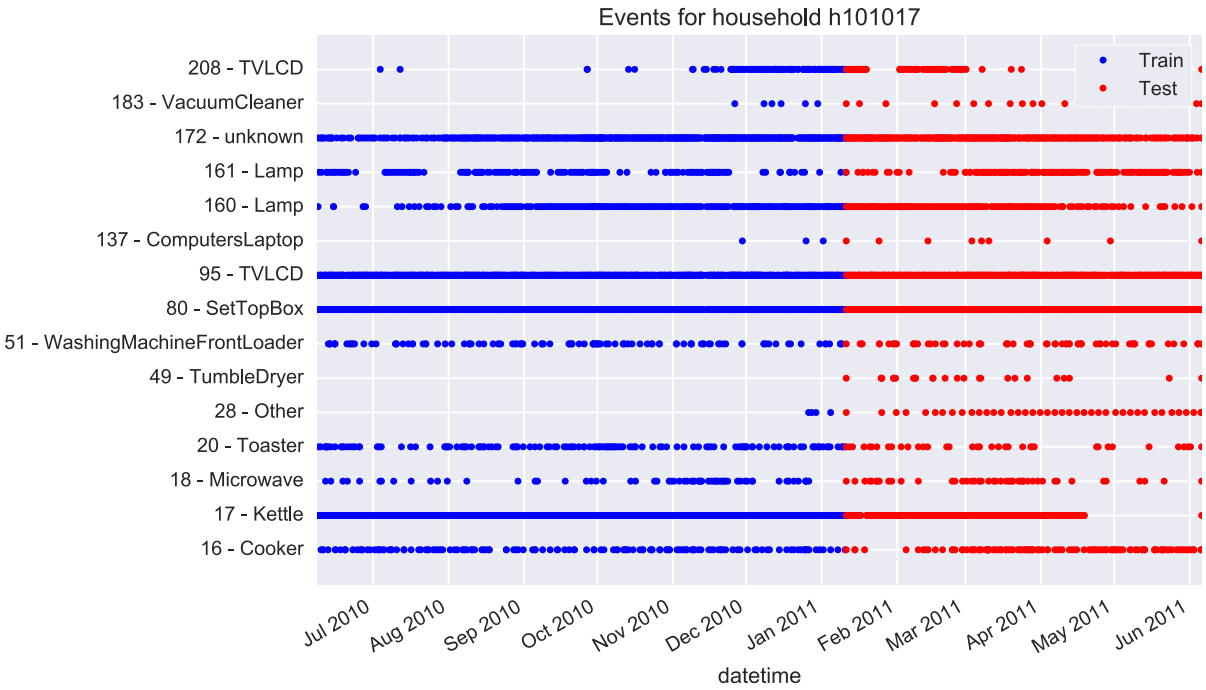


Figure 5.13: Training and test event samples for household no. 101017 in HES dataset.

5.4.2.2 Definition of Parameters and Constants in the DST algorithm

Regarding the configuration of the DST algorithm, the observation window T_i and the certainty constants C_0 and C_1 have been empirically fixed. A six-hour interval has been used for the observation window T_i, so every six hours the presence of each manual appliance (switching-on event) is evaluated and the mass scores in (5.22), (5.23) and (5.24) are obtained. Likewise, we set the certainty constants as follows: $C_0 = 0.9$ and $C_1 = 0.1$. This means that there is a 10% uncertainty $((1 - C_0) * 100)$ in case that a switching-on event of a certain appliance is presented; and a 90% of uncertainty otherwise. Because the *normality* of a pattern is evaluated by switching-on events, it cannot be stated that, if an appliance is not used during a certain time interval, there is an *abnormal* pattern. Note that the human behaviour is non-deterministic and, therefore, a high uncertainty is assigned in absence of events. Indeed, there is lack of information, which the DST algorithm models as uncertainty, whereas a Bayesian approach can seldom model it by using only probabilities. The inverse logic can be also applied: the occurrence of an event does not necessarily mean that the pattern is *normal*, but there is more information and

that leads to assign only a 10% of uncertainty. These values are empirically obtained and could be changed regarding the confidence on the usage pattern for appliances.

5.4.2.3 Analysis of DST and GMM scores

Figures 5.14, 5.15 and 5.16 depict the analysis of the human activity for the three single household pensioners over the test data (145 days roughly in each case). Similarly, Figure 5.18 shows the family household activity during the test (315 days). Apart from some exceptions, the family house keeps a strict routine and, consequently, the proposed algorithms generates fewer alarms. Plots in Figure 5.14.a, Figure 5.15.a, Figure 5.16.a, Figure 5.17.a, Figure 5.18.a and Figure 5.19.a are similar to the one in Figure 5.13, where the x-axis represents time and the y-axis means the different manual appliances, whereas the switching-on events of appliances are drawn as red dots. Table 5.4 shows the correspondence between the y-axis values and the appliance labels. A long-term observation of the score (several months) can help to detect deteriorations on someone's activity performance. These might be a symptom about the apparition of degenerative diseases as dementia, whose early detection could be very beneficial. On the other hand, short-term deviations (several hours or days) are more difficult to interpret, although they should be also watched as they might denote an emergency. Following, we present some potential cases of deterioration where the performance of the GMM and the DST algorithms are evaluated.

Appliance ID	**Appliance Label**			
	Household no. 101017	**Household no. 103034**	**Household no. 102003**	**Family House**
1	Cooker	Kettle	Cooker	Boiler
2	Kettle	Microwave	Kettle	Solar thermal station
3	Microwave	Toaster	Microwave	Washer drier
4	Toaster	Cooker extractor fan	Toaster	Dish washer
5		Washing machine	Cooker extractor fan	Television
6	Tumble dryer		Washing machine	Light 1
7	Washing Machine	DVD player	TV-LCD	HTPC
8	Set Top Box	TV-CRT	TV-LCD 2	Kettle
9	TV-LCD	VCR	Computer desktop	Toaster
10	Laptop	Lamp 1	Computer monitor	Fridge freezer
11	Lamp 1	Lamp 2	Printer inkjet	Microwave
12	Lamp 2	Lamp 3	Lamp 1	Computer monitor
13		Lamp 4	Lamp 2	Bread maker
14	Vacuum cleaner		Lamp 3	Audio amplifier
15	TV-LCD	Vacuum cleaner	Lamp 4	Light 2
16			Lamp 5	Soldering iron
17				Ethernet switch
18			Vaccuum cleaner	Vaccum cleaner
19				Light 3
20				Light 4
21				Light 5
22				Light 6
23				Active subwoofer
24				Light 7
25				Radio
26				Light 8
27				Phone charger
28				light 9
29				Phone charger 2
30				Light 10
31				Coffee maker
32				Radio 2
33				Phone charger 3
34				Hair dryer
35				Hair straighteners
36				Clothes iron
37				Oven
38				Light 11
39				Baby monitor
40				Light 12
41				Light 13
42				Computer desktop
43				Fan
44				Printer

Table 5.4: Appliance labelling in Y-axis by house.

5.4.2.3.1 Single pensioner household no. 101017 in HES dataset

Figure 5.14.a shows the test event samples of manually operated appliances that show a repetitive pattern: cooker, kettle, microwave, toaster, lamps and others. Figure 5.14.b represents the score provided by the DST algorithm with a 6-hour observation window T_i, whereas Figure 5.14.c depicts the date score by accumulating evidences over time as was explained in Section 5.3. In both plots, there is a green filled area limited by the *belief* and *plausibility* curves. This is the uncertainty area and it becomes larger as the number of events decreases. The uncertainty is clear in Figure 5.14.b, but the accumulative evidence over the day makes the uncertainty area thinner in Figure 5.14.c, thus obtaining a more accurate score.

Likewise, Figure 5.14.d and Figure 5.14.e show the results after applying the GMM algorithm (Section5.2) to score a day, and the average score over the week, respectively.

The red line in Figure 5.14.b, Figure 5.14.c, Figure 5.14.d and Figure 5.14.e are the empirical thresholds for the DST and the GMM algorithms, where any score bellow those thresholds is considered an anomaly in the behaviour. For this specific household, a deviation from the regular routine over the months is described with a lower score in both algorithms. These kinds of long-term deviations are better observed in the DST day score and in the GMM week score. Nevertheless, the former seems to be more sensitive to this kind of pattern, thus rapidly dropping to zero what allows an earlier detection of pattern deviations than the one achieved by the GMM algorithm. The DST algorithm is also able to detect three isolated days of *abnormal* activity.

Furthermore, it is worth noting that the oscillation in the daily score for the GMM (see Figure 5.14.c) is high and it makes difficult to configure the threshold to avoid false alarms. From a manual inspection of data, it has been verified that this deviation is due to those appliances related to cooking, such as the kettle, microwave and toaster, whose occurrences decrease during that period. Consequently, the ADL of feeding might be deteriorated.

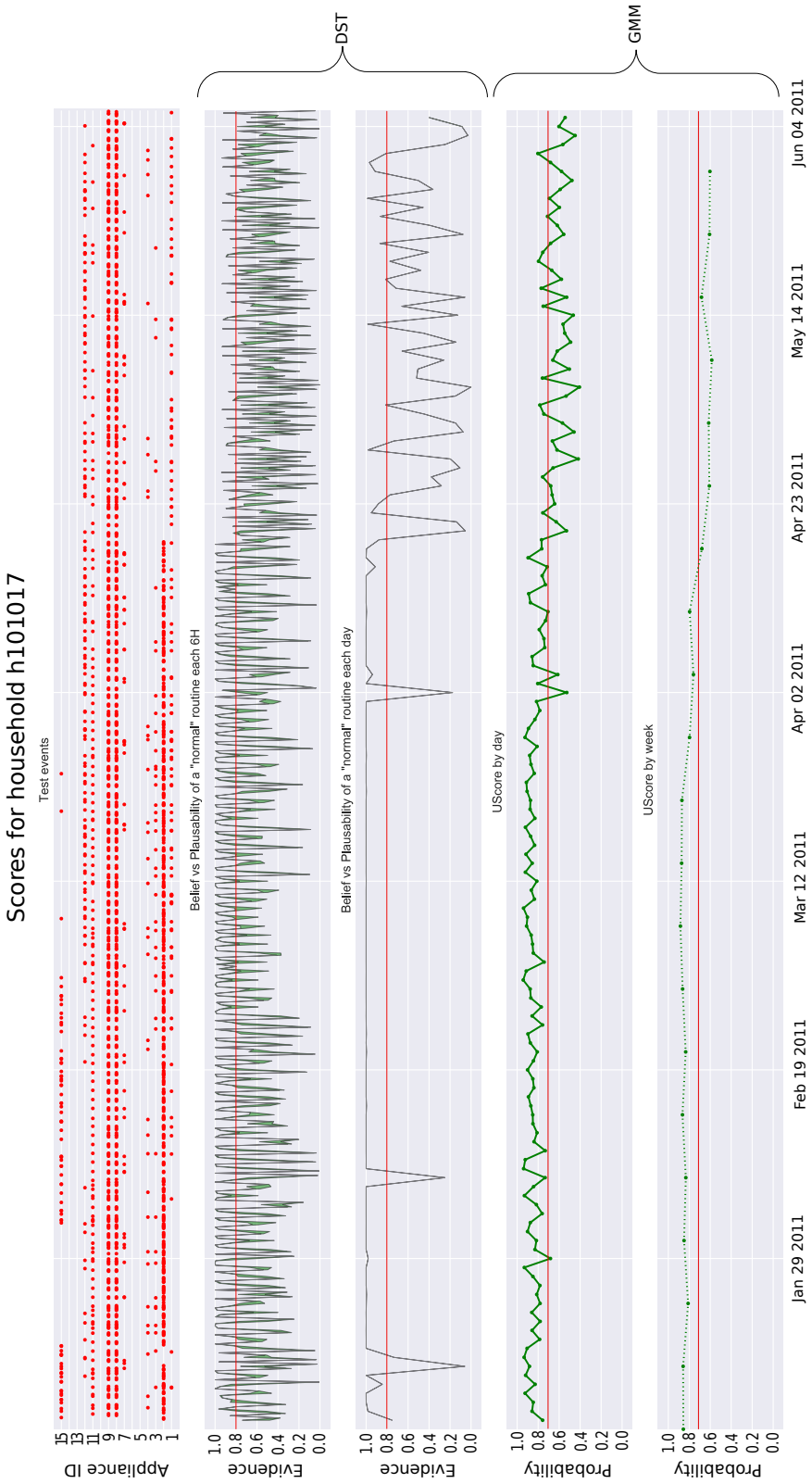


Figure 5.14: Score for test events in household no. 101017 from HES dataset: a) Test events; b) DST score every 6 hours; c) DST score by day; d) Union probability score by day; e) Union probability score by week.

5.4.2.3.2 Single pensioner household no. 103034 in HES dataset

We found similar deviations in Figure 5.15.c where the deterioration is due to those appliances related to a range of activities: some days caused by cooking appliances as seen before, others due to the use of the laptop, TV set or the lighting system. This implies more noise for the GMM approach (see Figure 5.15.d) and the decrease of the mean per weeks becomes very slight (see Figure 5.15.e). Again, the threshold in the daily score for the GMM (Figure 5.15.d) is difficult to set a priori. Furthermore, looking into month intervals such as the one from January 29th to February 19th of 2011 in Figure 5.14, short-term anomalies can be detected in the DST algorithm, whereas the GMM algorithm does not show these deviations.

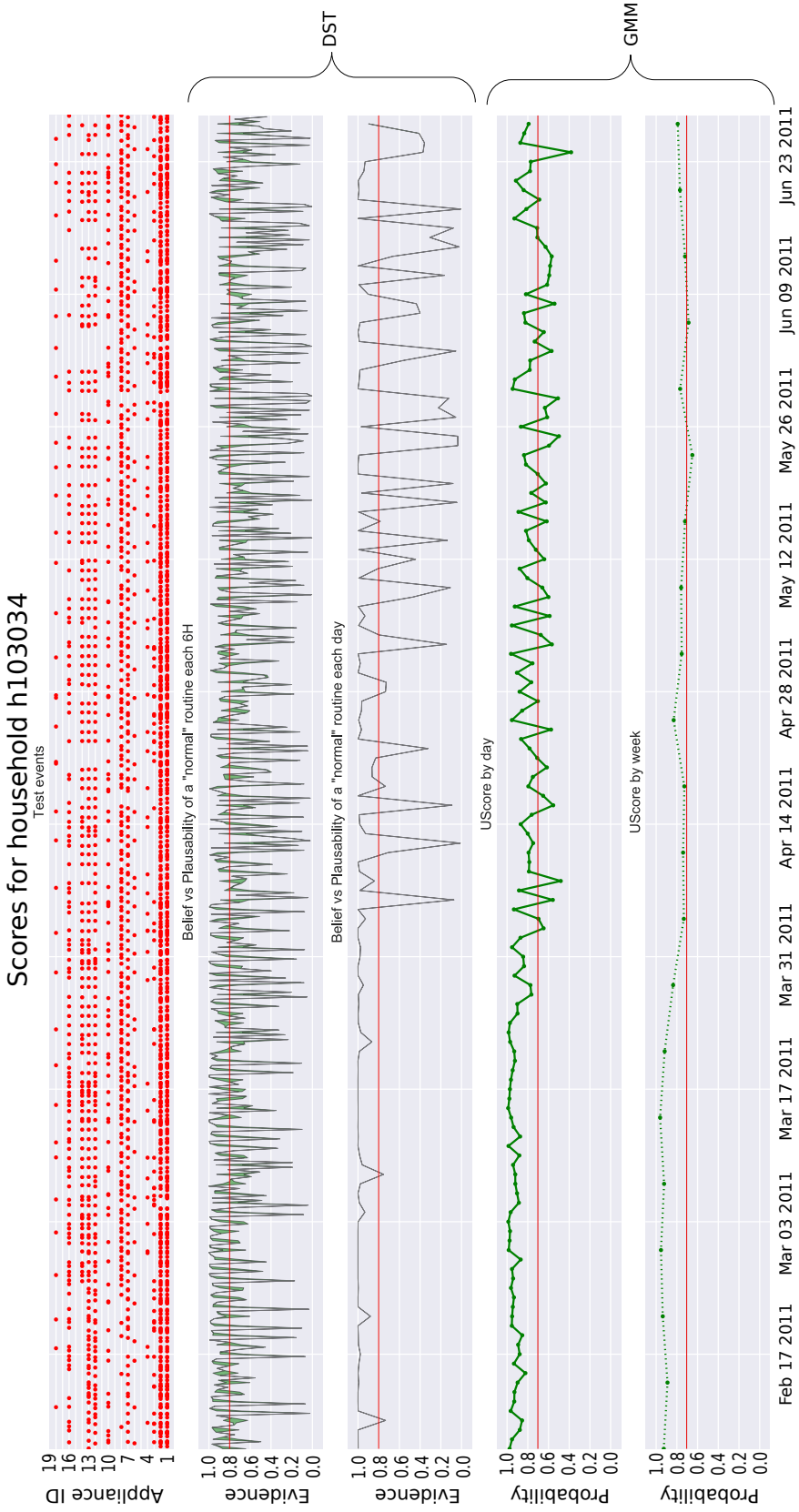


Figure 5.15: Score for test events in household no. 103034 from HES dataset: a) Test events; b) DST score every 6 hours; c) DST score by day; d) Union probability score by day; e) Union probability score by week.

5.4.2.3.3 Single pensioner household no. 102003 in HES dataset

Conversely, there is a case in Figure 5.16.c where a single pensioner has recovered his routine after a deviation. The pattern routine carried by the person involves days of inactivity, which cannot be properly modelled by the Bayesian Theory resulting in oscillations from *normality* to an anomalous behaviour score provided in Figure 5.16. This can be closely observed in Figure 5.17, a zooming zone of a week from Figure 5.16. Starting on Monday 4th April of 2011, this inhabitant spends most of his time out during Fridays, Saturdays and Sundays. This has been previously learned by the DST algorithm from the training data and it is modelled as an increasing of the uncertainty in Figure 5.17.b, which decreases in Figure 5.17.c due to the accumulated evidence. For instance, in 9th April of 2011 in Figure 5.17.b, the uncertainty is such that most time the threshold lies into the area between the belief and the plausibility and, hence, one cannot state whether the pattern is *normal* or *abnormal*. Only between the 12:00pm and the 18:00pm it is feasible to conclude that the period is *normal*, since the threshold lies bellow both belief and plausibility. The GMM algorithm, which is a classical Bayesian network, cannot handle uncertainties, so it assigns a probability that decreases with the lack of activity. This leads to a larger number of false alarms.

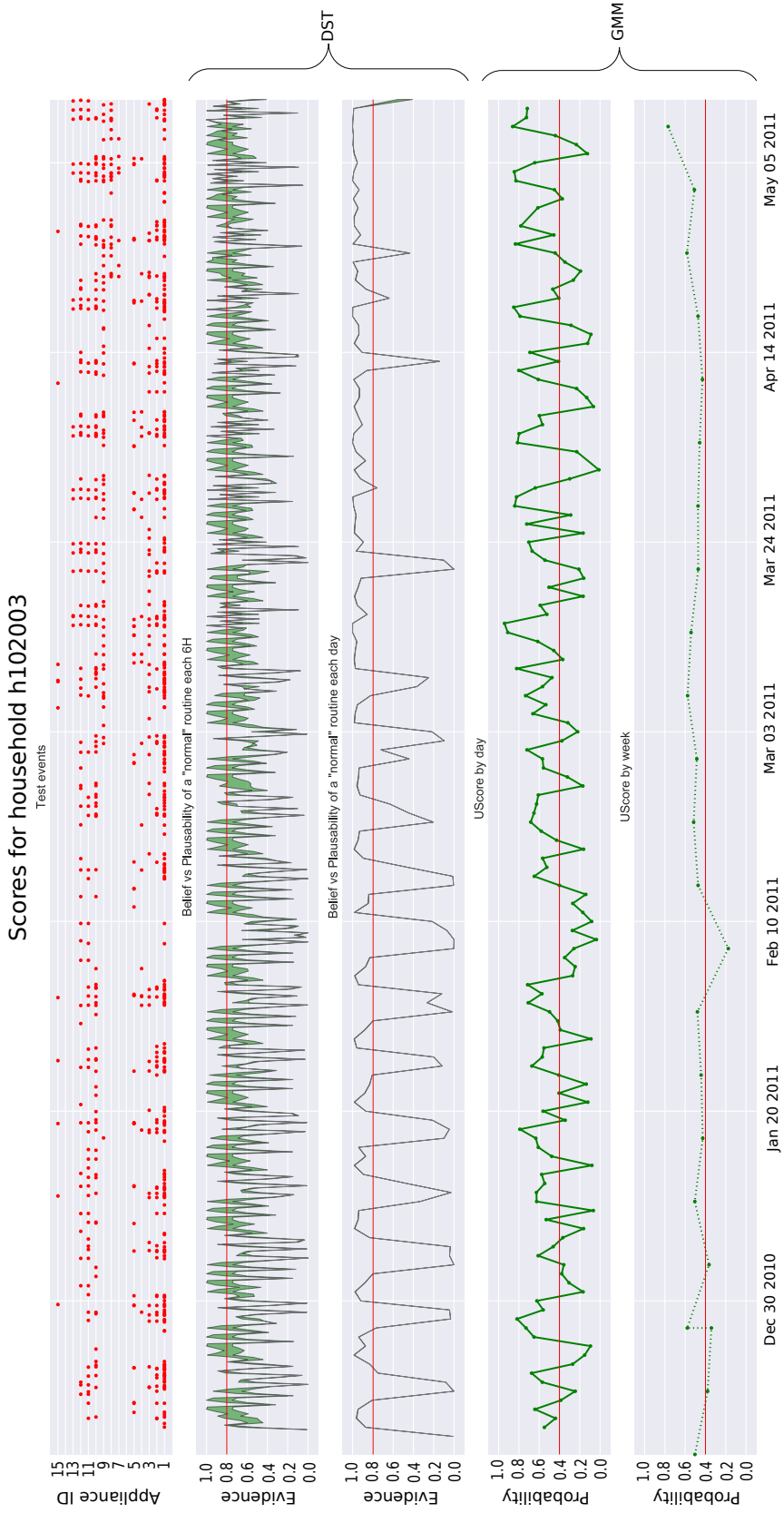


Figure 5.16: Score for test events in household no. 102003 from HES dataset: a) Test events; b) DST score every 6 hours; c) DST score by day; d) Union probability score by day; e) Union probability score by week.

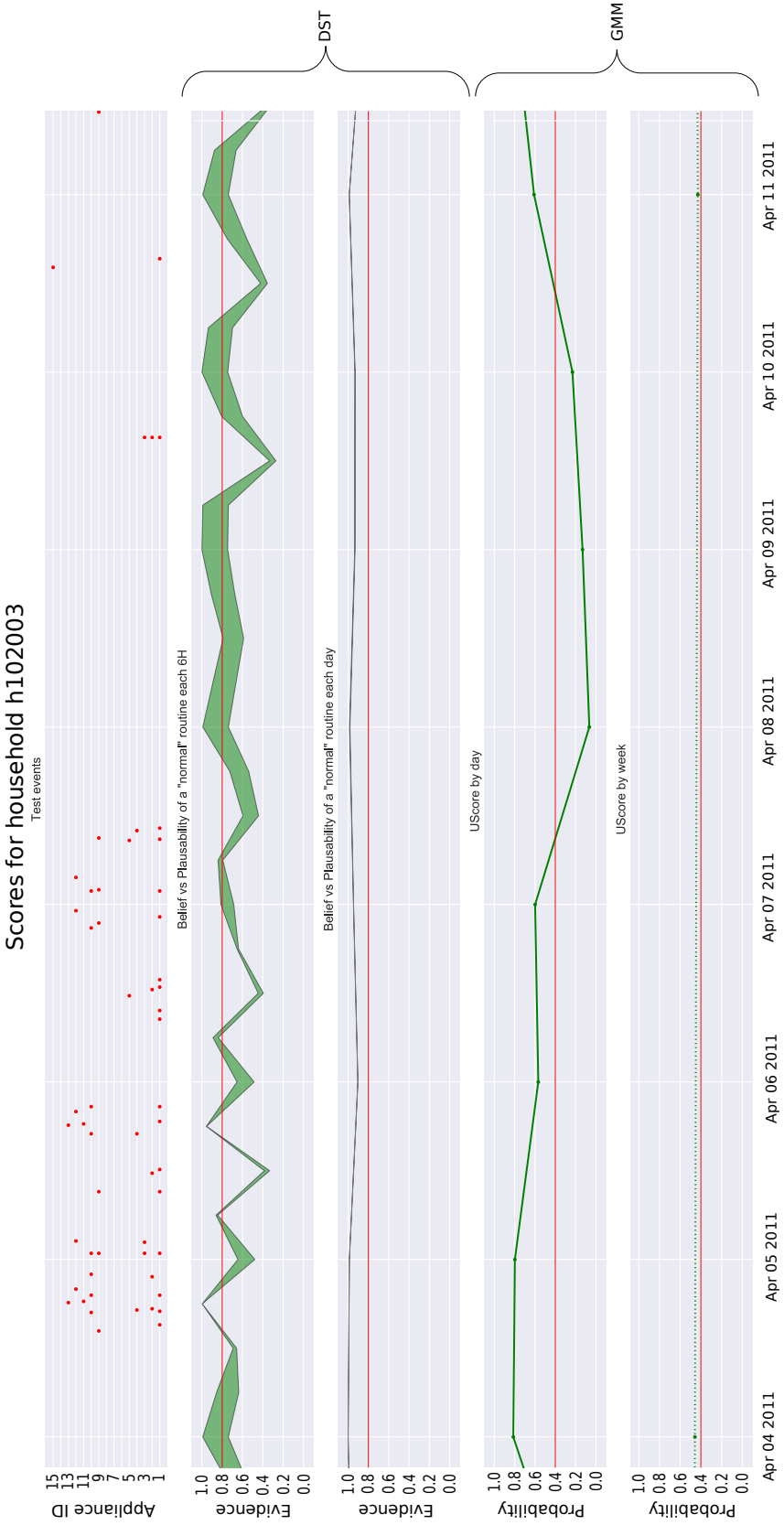


Figure 5.17: Score for test events in household no. 102003 from HES dataset during a week: a) Test events; b) DST score every 6 hours; c) DST score by day; d) Union probability score by day; e) Union probability score by week.

5.4.2.3.4 Family house in UK-DALE dataset

Finally, a case formed by a family of two adults, two children and a dog is analyzed in Figure 5.18. It is worth noting the increasing number of events coming from different (manually operated) appliances. Furthermore, the scores provided by Figure 5.18.c, Figure 5.18.d and Figure 5.18.e reflect a very strict routine. This routine does not deteriorate over time and only shows *abnormal* intervals that may correspond with holidays or absences in the house. However, these deviations pass unnoticed by the GMM algorithm. Regarding the threshold, this can be easily set for the DST algorithm, whereas it becomes more difficult for the GMM one. Anyway, it should be close to 1 in order to detect some anomalies.

Zooming into a week period as in Figure 5.19, it is shown that the DST algorithm is sensitive to the anomalies, whereas the GMM cannot detect them. Since the score in GMM is an union probability, a single appliance which keeps a strict routine can saturate the score and, consequently, masks other anomalies in the remaining appliances. This does not happen in DST where the accumulative evidence of the hypotheses with *abnormal* pattern is higher than the evidence of *normal* patterns. Therefore, it properly models these effects. Furthermore, we would like to enhance that the uncertainty is smaller in Figure 5.19.b than its corresponding in Figure 5.17.b due to the increasing number of appliances per 6-hour observation window.

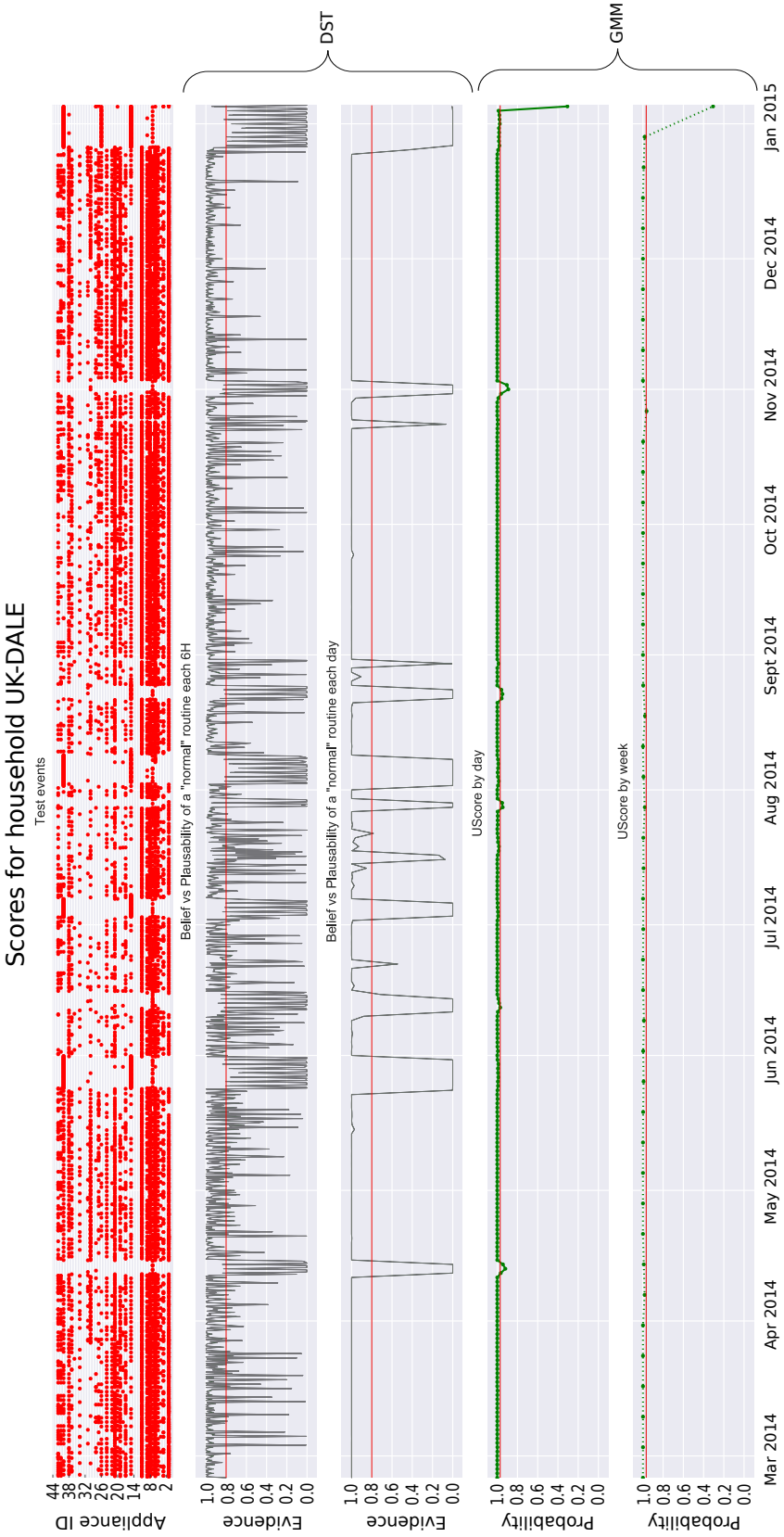


Figure 5.18: Score for test events in household no. 1 from UKDALE dataset: a) Test events; b) DST score every 6 hours; c) DST score by day: d) Union probability score by day; e) Union probability score by week.

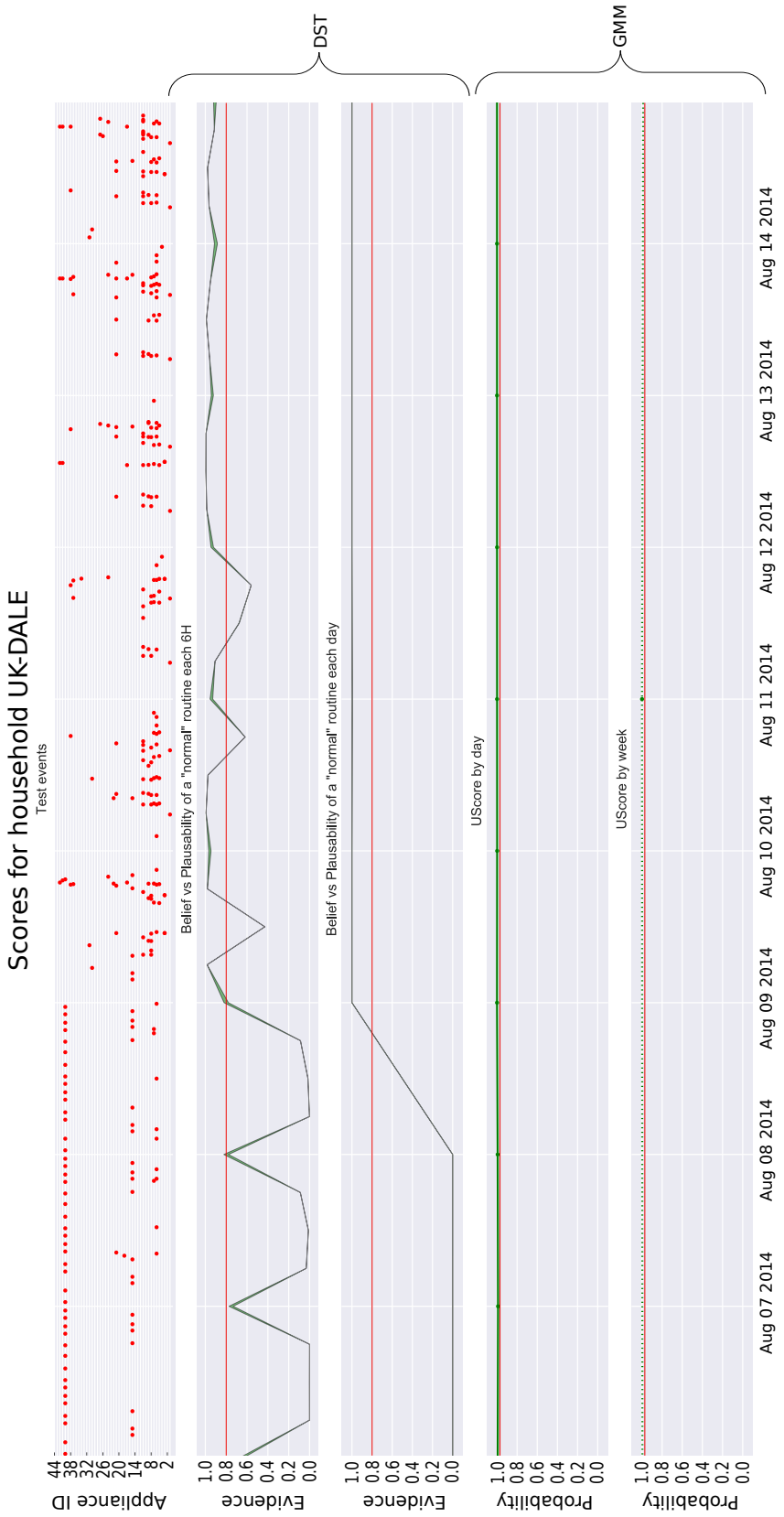


Figure 5.19: Score for test events in household no. 1 from UKDALE dataset: a) Test events; b) DST score every 6 hours; c) DST score by day; d) Union probability score by day; e) Union probability score by week.

5.4.3 Conclusions

We disclosed in this Chapter three proposals to monitor the activity of humans within a household. The spotted community is elderly residents who live alone and our main goal is to provide basic non-intrusive health monitoring that can be deployed at large scale and at low cost. Crucially, they do not require additional sensors to be installed within each household because they rely on a previous NILM disaggregation algorithms. We do not intend to obtain an accurate *index* of ADLs because tracking them over time often requires more heterogeneous sensor data to deal with overlapped activities and activity heterogeneous duration, which is much intrusive (see Section 2.2). Instead, we disaggregate the smart meter data and demonstrate that the human activity can be inferred through its interaction with household appliances. That can be evaluated to detect deviations on the normal daily routine and to monitor the health of the household's occupant. Thus, the independence of elderly people can be maintained whereas non-intrusively monitoring their activity. We look into the problem from two perspectives:

- *Single appliance monitoring:* If the usage pattern of a single appliance is strong enough (i.e it is used regularly and it is an important part on someone's routine as explained in Section 5.1), a single-appliance monitoring approach is preferable. That is, less variables are involved in, which means that the analysis is more robust and less dependent on disaggregation errors and less affected by someone's freewill. The strong coordination between the appliance usage pattern and the human activity indicates that a system alarm to detect short-term deviations in the normal routine is the best approach in this case. Therefore, we model the usage of the appliance using a log Gaussian Cox process, and show that our algorithm is able to learn the usage pattern unique to each household and efficiently trigger an alarm when any deviation from this pattern is detected by not using the specific appliance. Our evaluation indicates that our approach allows earlier interventions in households with a consistent routine (25 out of 32 households on average) and fewer false alarms in the remaining households, relative to a fixed-time intervention benchmark. Furthermore, a log Gaussian Cox process allows the usage pattern to change over time and provide a measure of uncertainty.

- *Multi appliance monitoring:* There is not a predominant appliance within the household and, consequently, the human activity is inferred from the combination of all

manually operated appliances. We propose two models based on scoring the appliance usage patterns (Section 5.2 and Section 5.3) and compare them in Section 5.4.2. We apply the Bayesian Theory in the GMM monitoring model where appliance usage patterns are learnt by Gaussian Mixture Models and the score is obtained from the probability union. In Section 5.4.2, we have demonstrated how a deterioration in the human activity is reflected in this model with a decrease of the score. However, we have not been able to set a automatic threshold to trigger an alarm and human intelligence is still necessary. Therefore, we believe this approach is more convenient to detect long-term deviations in the human activity. Furthermore, we developed the DST monitoring model in order to encode uncertainty to meet the freewill and non-deterministic human behaviour. The Dempster-Shafer Theory is a generalization of the Bayesian Theory and it admits certain subjectivity on the design of the *Mass Functions,* which has brought criticism. However, we proof in Section 5.4.2 that the DST monitoring model outperforms the GMM monitoring model in detecting either long-term or short-term deviations. Likewise, the DST monitoring model allows to set threshold to trigger alarms as it is more sensitive to pattern deviations. Besides, it less susceptible to false alarms because it successfully models the sparsity of events through uncertainty. In fact, the lack of uncertainty in the GMM monitoring model makes more difficult the learning process as unusual days with less than 10 events have to be filtered out (see Section 5.2).

Due to the lack of similar studies, a benchmarking could not be stablished with literature. In single appliance monitoring, we devised the fixed-time intervention benchmark to compare the performance of the LGCP. Likewise, in multi-appliance monitoring, we have compared the performance of the DST approach against the GMM approach. Most studies as in [Rashidi et al., 2011, Fleury et al., 2010, Chen et al., 2015] score their performance on the classification of event sequences into ADLs using public dataset such as the CASAS dataset ([Cook et al., 2013]). Instead of aiming to detect ADLs as other works, we pursue the monitoring of the human behaviour using only the smart meter data and, hence, there is a lack of available datasets with labelled data where the performance can be benchmarked.

We enhance the importance of including the daily and weekly periodicity, as well as the uncertainty, in our models to properly infer the human activity. Furthermore, it is worth noting that individuals have completely different routines and, hence, activity

monitoring approaches must be trained specifically for each household. If we are to base our approaches on Activity Recognition, we should perform a supervised learning process for each household, and this would affect the scalability. Because we only need to learn appliance usage patterns, the learning process is subjected to the NILM learning process to label the data. Hence, the proposed activity monitoring approaches are unsupervised as long as the NILM algorithm that lies beneath is so.

Finally, we would like to remark the main advantage over other health care systems: there is not need of previous calibrations and sensor network installations, as we only need the information from smart meters, which are being widespread installed. Therefore, there is a significant resource savings compared to other technologies where daily phone calls and in-person visits are needed. This fact makes our approaches prone to be deployed in a large scale to provide service to the increasing elderly community. Many health problems and welfare issues are directly related to the deviations detected in Section 5.4.2.3. For instance, dementia leads to a deterioration on someone's performance on the use of appliances [Nikamalfard et al., 2012]. Thus, it is a powerful tool for experts, caregivers and relatives to keep up with someone's activity and give basic assistance or check in. Moreover, the cost of a false alarm is relatively low, which would result in a phone call to the elderly occupant to check in. However, under non circumstances it is a replacement in cases of severe risk, where 24 hours assistance and invasive methods for measuring vital signs are still needed.

Consequently, we believe that the findings in this Chapter can be of large benefit to current health problems and welfare issues of our society pointed out in Chapter 1 and Section 2.2.1 regarding the rapid increase of the elderly community and basis assistance services.

Chapter 6

Conclusions and Future Works

This thesis is the result of our concern about the new challenges to be faced in the forthcoming years: lack of energy resources, energy efficiency, ageing of the population and sustainability of healthcare services. Along Chapter 1, the critical need of reaching energy efficiency and sustainability of healthcare monitoring models has been pointed out. A belief that these issues are correlated to an overpopulation and they may benefit from a common solution has pushed the research in this thesis. Likewise, the research has been focused on developing no-intrusive activity monitoring techniques for elderlies that can be scalable and are based on NILM, which is an energy solution to reach efficiency. Chapter 2 gained us the knowledge about NILM and healthcare monitoring from this particular perspective. It leads to identify the needs and requirements to achieve our purposes, Chapter 3. From here, we are able to present our contributions as well as the results along the Chapter 4 and Chapter 5.

To close this manuscript, we would like to summarise the major achievements and innovations of this thesis in Subsection 6.1. Following by the communication of this work to the scientific community in form of scientific publications (Subsection 6.2), other types of communications (Subsection 6.3) and stays in international centers to generate collaborations and knowledge exchanges (Subsection 6.4).

6.1 Conclusions

The main research contributions of this thesis is to devise new NILM techniques, which are mainly energy saving oriented, into activity monitoring approaches in order to add a social

value (i.e. monitoring single elderly residents) to this field and, consequently, to bring the interest of end-customers. Likewise, we aimed to the development of novel activity monitoring techniques that, thanks to the use of NILM, can be scalable, non-intrusive and of low cost and, thereby, able to support an increasing need (Chapter 1). We have fulfilled all the requirements in Chapter 3 and an unsupervised activity monitoring systems, to be integrated in Smart Meters, has been devised. Our achievements have been respectively presented in Chapter 4 and Chapter 5, along with a thorough evaluation. However, let us summarise them in this Section in order to picture the importance of them.

- The complete proposal is scalable and integrable into Smart Meters: it is unsupervised and the sampling frequency is feasible. This must be higher than 1Hz to detect events but lower enough to enable Smart Meter integration, which is typically up to 8kHz (see Section 2.1). Experimental results presented in Chapter 4 validate our proposal in datasets that have signals with a bandwidth of 300Hz and 1kHz, whenever an event occurs. Hence, it can be easily integrated in Smart Meters and in third-party devices.

- Our NILM algorithm is event-based and the event detector can discriminate between switching-on and switching-off events, which is the basis input in our activity monitoring proposals. It normalises the current signal to avoid voltage drifting and voltage drops so the noise in the mains does not affect the evaluation of the current signal. In fact, this make possible its generalization along all houses as it will reasonably draw a similar current signature for the same appliance in different houses.

- A novel Load Signature based on the Active, Reactive and Distortion power trajectories (PQD) is presented, which has proved to a suitable choice as many features of the appliance type can be extracted from both the transient and steady states. In fact, the cumbersome task of labelling steady and transient states has been eliminated by using "curves", which is a novelty and give more scalability.

- Then, the proposed PCA-PQD classifier has the novelty of producing general models of appliance types that can be trained offline ("unsupervised training"). This is clearly an advantage and a step forward to scalability because it makes our proposal directly deployable in new scenarios with no further training.

- With regards to monitoring process, three different techniques for elderly activity monitoring based on NILM have been proposed. Likewise, the main goal has been

achieved: huge resources savings in the hypothetical deployment (i.e. only the smart meter is needed to monitor), low-cost and scalability to support the increasing numbers of elderlies residents living alone.

- Activity Recognition is a well developed technique and there is not lack of attempt to use AR to detect ADLs and monitor elderly people. However, along Chapter 2, Chapter 3 and Chapter 5 we have addressed the complexity of tracking ADL in terms of duration and action using just one sensor (i.e. the smart meter): overlapped activities and heterogeneous duration issues. Furthermore, these approaches are yet "supervised", and specific for every single house, which conflicts with the unsupervised requirement and poses scalability issues. Thereby, our approaches only use appliance usage patterns to infer the human behaviour. Therefore, they are unsupervised, whether an unsupervised NILM algorithm is to be used.

- Someone's daily routine can be inferred from a single appliance usage pattern whether there is a strong correlation between the appliance and the human activity. In this case, a single appliance monitoring is preferable as the analysis become more robust and less dependent on disaggregation errors and other variables. Thus, the LGCP approach has been proposed to monitor the daily routine and to early trigger alarms if someone deviates from his normal routine. Furthermore, the daily and weekly periodicities, as well as uncertainties, are encoded to the model and, hence, it successfully infers the human behaviour. This technique has been devised to provide early interventions in case of urgencies and, because only one appliance needs to be disaggregated, is the most scalable approach.

- Moreover, two multi-appliance monitoring approaches are presented when the human activity becomes more erratic and cannot be inferred just from one appliance. The GMM monitoring approach implements long-term monitoring using the Bayesian Theory. However, this theory does not enable uncertainties and higher number of false alarms are produced whenever a threshold for triggering alerts is set. Uncertainties are the key to infer the freewill and non-deterministic human behaviour behind the appliance usage patterns that also causes sparsity. Thereby, we devised the DST approach with outperforms to GMM approach and works well for multi-appliance monitoring in long term and in short term (i.e. using alerts).

These contributions have been validated by using datasets collected from real scenarios and standard metrics. This a common practice in the NILM community as the process

of collecting data is long, cumbersome, expensive and becomes more complex with the number of households. Furthermore, the use of real public datasets allows to compare works. In our work:

- The event-detector performance has been fully evaluated with one of the most comprehensive existed datasets for event-detection and disaggregation in NILM (i.e. the BLUED [Anderson et al., 2012b]). It entails real voltage and current aggregated data consumption and up to 2845 real events coming from 30 different appliance types, which makes a comprehensive case study. Real standard metrics have been used to evaluate the performance obtaining promising results: up to 94% of TPR with less than 0.08% of TPR with a optimal detector. In case of heavy event density (i.e. the case in the phase B in the dataset), the proposal detects up to 88% of TPR with less than 12% of TPR. As we perform a normalization, the optimal detector should be the same for all houses, however, in real deployments, it may be difficult to achieve a complete optimal one. Thus, we also evaluated the AUC that provides an estimated value of the overall performance obtaining high performance of 95% and 94% in phase A and phase B, respectively.

- To evaluate the PCA-PQD performance, the PLAID dataset ([Gao et al., 2014]) has been selected as the most suitable choice and, to the best of our knowledge, it is the most comprehensive to perform LS classification. It comprises up to 170 different instances per appliance type (i.e. in total 11 appliance types) coming from 55 different households. Hence, having 170 real measurements of voltage and current from the same appliance type, it seems to be a reasonably good scenario to evaluate the generalization and scalability. The PCA-PQD proposal, along with the PQD trajectories, is able to extract the information signature from these instances to obtain a very accurate general model of the appliance type. In fact, our results points at a high performance of 98% accuracy, 83% recall and 90.6% F1 score (i.e. the harmonic mean between accuracy and recall). Furthermore, we provide another metric that becomes more important whenever the daily usage pattern has to be learnt: the F0.5 score that performs 98.5%; see Section 4.2.

- We have showed a strong correlation between the daily routine and some appliances. Thus, the HES dataset, which is the largest existing dataset that specifically contains energy consumption from single elderly residents, has been used for this purpose. The kettle has been identified as having one of the strongest usage pattern of all

appliances in the UK. In fact, it has been validated that most of the appliances have a daily and weekly pattern. Therefore, the LGCP monitoring proved to be a successful method to establish an alarm system based on "the first use of the kettle within the day" as the intervention policy. Using the HES dataset, we compare this LGCP approach, with its intervention policy, against a fixed-time intervention threshold. Our evaluation indicates that this approach enables early intervention in 25 out of 32 households.

- The GMM and the DST approaches have been validated using three single elderly households from the HES dataset and one family household from the UK-DALE dataset, which at last instance serves to highlight results in the elderly households. After a training period, both the GMM and the DST score the daily routine and, using a threshold they trigger alarms. In both approaches the deterioration can be observed in the elderly resident household scores whereas the family household score remains strong. Hence, this is an indication of the importance in monitoring single elderly residents. Both approaches show deterioration in a long-term monitoring, which could be well related to degenerative diseases such as dementia and alzheimer. Power, only the DST approach successes in triggering alarms for early interventions. This is due to the sparsity that is well modelled in the DST as uncertainty and consequently, less false alarms occurs.

The thesis proposals also have some limitations. Thus, the proposed NILM algorithm may have more difficulties in term of integration than other LF-NILM algorithms that are eventless. Particularly in the case of a HAN connection at 1Hz or 0.1Hz sample rate. However, this is something that strongly depends on the Smart Meter and the smart metering rolling-out country policies (as seen in Section 2.1). Moreover, the activity monitoring proposals do not have the same accuracy and reliability than other well developed healthcare systems based on BSN or a high-density heterogeneous WSN, as seen in Section 2.2. There is a trade-off between scalability and accuracy and, therefore, more false positives are expected in our models, which are highly scalable. However, it is worth noting that we pursue to provide a powerful tool for experts, caregivers and relatives to keep up with someone's activity and give basic assistance. In this case, a false alarm would result in a phone call to check in and. Then, false alarms are of relative low cost, which contrasts to the huge benefit from a non-intrusive monitoring method.

6.2 Publications

The following publications have been derived from this thesis:

6.2.1 International Journals

- **José M. Alcalá**, Jesús Ureña, Álvaro Hernández, David Gualda, "Event-based Energy Disaggregation Algorithm for fine-grained Activity Monitoring from a Single-Point sensor", *IEEE Transactions on Instrumentation and Measurement, ISSN: 0018-9456, in review.*

- **José M. Alcalá**, Jesús Ureña, Álvaro Hernández, David Gualda, Assessing Human Activity in Elderly People using Non-Intrusive Load Monitoring", *Sensors, ISSN: 1424-8220, in review.*

6.2.2 International Conferences

- **José M. Alcalá**, Victor Cionca, Michael Hayes, Brendan O'Flynn, and Álvaro Hernández. *"An energy consumption model for a wsn node based solely on the duty cycle*". In Proceedings of the 7th International Conference on Sensor Technologies and Applications (SENSORCOMM 2013), ISBN: 978-1-61208-296-7, pages 113-118, Barcelona (Spain), August 2013.

- **José. M. Alcalá**, Jesús Ureña, and Álvaro Hernández. *"Event-based detector for non-intrusive load monitoring based on the Hilbert transform"*. In Proceedings of the IEEE 19th International Conference on Emerging Technology and Factory Automation (ETFA), doi: 10.1109/ETFA.2014. 7005320, ISBN: 978-1-4673-7929-8, pages 1-4, Barcelona (Spain), September 2014.

- **José M. Alcalá**, Jesús Ureña, and Álvaro Hernández. *"Activity supervision tool using non-intrusive load monitoring systems"*. In Proceedings of the IEEE 20th International Conference on Emerging Technologies Factory Automation (ETFA), doi: 10.1109/ETFA.2015.7301622, ISBN: 978-1-4673-7928-1, pages 1-4, Luxembourg, September 2015.

- **José M. Alcalá**, Oliver Parson, and Alex Rogers. *"Detecting Anomalies in Activities of Daily Living of Elderly Residents via Energy Disaggregation and Cox Processes"*. In Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments (BuildSys 2015),

  d.o.i: 10.11452821650.2821654, pages 225-234, 2015, Seoul (South Korea), November 2015.

- **José M. Alcalá**, Jesús Ureña, Álvaro Hernandez, David Gualda. *"Symbolic Localization using NILM from only Smart Meter Data",* In Proceedings of the Seventh International Conference on Indoor Positioning and Indoor Navigation (IPIN 2016), ISBN 978-1-5090-2424-7, Madrid (Spain), October 2016.

6.2.3 National Conferences

- **José M. Alcalá**, Jesús Ureña, Álvaro Hernandez, Juan Jesús García. *"Análisis no intrusivo de la actividad humana a través de la monitorización del consumo energético",* In XXXVI Jornadas de Automática, ISBN 978-84-15914-12-9, pages 732-736, Bilbao (Spain), September 2015.

6.2.4 International Workshops

The European Workshop on Non-Intrusive Load Monitoring, as well as The International Workshop on Non-Intrusive Load Monitoring, is the most important meeting for NILM researches and companies.

- **José M. Alcalá**, Jesús Ureña, Álvaro Hernandez, Juan Jesús García. *"Activity Supervision Tool using Non-Intrusive Load Monitoring Systems"*. The 2nd European Workshop on Non-Intrusive Load Monitoring (NILM European workshop), London (United Kingdom), July 2015.

- **José M. Alcalá**, Jesús Ureña, Álvaro Hernandez. *"Improving accuracy in Load Identification using Principal Component Analysis and power trajectories"*. The 3rd European Workshop on Non-Intrusive Load Monitoring (NILM European workshop), London (United Kingdom), October 2016.

6.3 Other research contributions

We have been also active within the NILM community contributing to the Open source NILM toolkit (NILMTK). Particularly, we have implemented the energy disaggregation algorithm called Maximum Likelihood Estimator algorithm (MLE). This is a variant of the Hart disaggregation algorithm where three features are evaluated over the active power: on-power, off-power and duration. For an appliance, these features are trained and a GMM is obtained for each feature. Later, the disaggregation is done by applying a threshold to the maximum likelihood estimated for the three features.

NILMTK is the most worldwide known open source tool written in python for NILM.

6.4 International Research Stays

This thesis has a remarkable profile in international research stays. Thus, there has been a total of two 7-month international stays:

- From August 2012 to March 2013: Wireless Sensor Network Group at Tyndall National Institute. Cork, Ireland. This stay has been much beneficial to gain basis knowledge in python, tinyOS and linux command line, which were the most developed tool to conduct the experiment during the thesis. Furthermore, an initial research conducted in autonomous WSN helps us to focus our research in NILM.
- From September 2014 to March 2015: Agents, Interaction and Complexity Research Group (AIC), University of Southampton. Southampton, United Kingdom. This stay was crucial to develop the LGCP monitoring approach and gain deep knowledge in the Pandas library to handle big datasets and basis in NILM techniques.

6.5 Future Works

There is a number of potential researches that can be taken from here, and we would like to point out some of them.

With concern to the NILM algorithm that has been presented in Chapter 4:

- Although the event detector and the classifier have been measured with standard metrics over public datasets that contain real measurements, the de-energise stage has not been evaluated. In further researches, a complete evaluation of the approach taking into account the de-energise process will be conducted.
- The BLUED dataset only contains a single household. In future work, the event detector will be evaluated in new dataset with higher number of houses.
- So far, the PQD-PCA classification algorithm has a very high performance (i.e. 98% accuracy) over 11 appliance types (which is fairly good in the NILM domain). It would be interesting to further evaluate the limits on the number of appliance type that this approach can disaggregate and see whether overlapping starts to appear between Load Signatures. However, it would be necessary a bigger dataset with more appliance types and, above all, many more instances.
- Experimental results have been presented based on signals with a bandwidth of 300Hz and 1kHz for the event detector and the classifier, respectively. In future works, it might be interesting to perform downsampling to observe how the metrics respond and find the limits for our proposal in terms of sampling rate.

Regarding the high-level proposals for activity monitoring in Chapter 5:

- There is a lack of real datasets that contain energy consumption data from single elderly households and certain metadata such as the performed activity or the health status. Thus, to further evaluate our approaches, this would be of great help to validate the results. We are in the process of collecting these data from a singular elderly household and creating the dataset. Once released, we believe it can be used to evaluate these approaches more easily.
- Besides, different intervention policies apart from "the first use of the kettle within the day" will be evaluated. Other potential candidates are: "times of use of a specific appliance during night time", to follow sleep disorders; and "times of use of an specific appliance within a short window time", to find short-term memory disorders.

Taking into account the whole system (i.e. NILM and the activity monitoring approaches):

- Feedback from users and specialists is always much valuable. Likewise, rolling out a real deployment in real households can give us much information about acceptability, interest and scalability. In this term, we believe that our approach can be easily integrated in real Smart Meters. However, a vendor partner is needed to carry out the deployment and the subsequent surveys.

Appendix A

Data Sets

Capturing and collecting real data from a large number of households is a cumbersome, time-consuming and a highly expensive task. Thus, it is a common practice between the NILM researches to use datasets. These are real collections from houses all over the World, which make them more comprehensive that any potential own collection of data. Besides, they are public, which enables comparison between researches. We have given descriptions of the used dataset in their respective chapters, however, a compilation of all used dataset in this thesis is provided bellow, along with an brief description and an explanation of its use. Furthermore, a list of the most comprehensive NILM datasets can be found in the NILM wiki (http://wiki.nilm.eu/index.php?title=NILM_datasets).

A.1 The BLUED dataset: Building-Level fUlly-labeled dataset for Electricity Disaggregation

The BLUED dataset is a fully labeled public dataset for event-based Non-Intrusive Load Monitoring [Anderson et al., 2012b]. It has been used for the obtained experimental results in Section 4.2, particularly for the event detector. This dataset has been selected because, unlike other datasets that are more appropriate for eventless algorithms (e.g. the REDD dataset [Kolter and Johnson, 2011] where the active power is sampled at 1Hz), it consists of voltage and current measurements sampled at 12kHz and where events are labelled. This is very convenient to evaluate the performance of event-based NILM algorithms and event detectors. The collection comes from a family household in the United States during a week. Although the sampling rate is 12kHz, the current sensor used to acquire

data had an upper cutoff frequency of about 300Hz, consequently, signals in this dataset have a bandwidth of 300Hz. It comprises a large number of events (2845 events) labelled with timestamp, appliance identifier and phase. The number of appliances involved is also large: 50 different appliances and 30 different appliance types. Besides, we have two characteristics scenarios: phase B where the event density is high and phase A with a lower event density. Thus, this dataset is much convenience for the analysis of the proposal as it represents a widespread number of real scenarios. In fact, to the best of our knowledge it is the existing largest and detailed dataset with events labelled.

A.2 The PLAID dataset: The Plug Load Appliance Identification Dataset

The Plug-Level Appliance Identification Dataset (PLAID, [Gao et al., 2014]) has been used to evaluate the performance of the *PQD-PCA* classifier presented in Section 4.2. It consists of current and voltage snapshots that last between 1 to 5 seconds, this is enough time to record both the transient and steady state whenever the switching-on happens. Currently, there are a total of 1074 different instances (i.e. number of appliances) coming from 11 different appliance types and 55 different households. For instance, there are recorded 176 different laptops, 167 different fluorescent lamps and more than 124 microwaves. This is a very convenient and consistent scenario for training a classification algorithm such as the *PQD-PCA* classifier where heterogeneous samples are available to achieve universal classifying models that works along different appliance brands and households. Besides, it is worth noting that, although the data resolution is 30 kHz, the current clamp employed (the Fluke i200 AC model) had an upper cutoff frequency of 1kHz. Consequently, signals in this dataset have a using a bandwidth of 1kHz.

A.3 The HES dataset: Household Electricity Survey dataset

This dataset has been collected in the village of Colden Common, Hampshire, UK [Zimmermann et al., 2012] (http://randd.defra.gov.uk/). It contains electricity data of 250

households from 32 households occupied by single elderly residents. The monitoring time varies between 1 month and 1 year (depending on the household) and the sampling interval is 2 minutes. We evaluated this data in Section 5.4 because it explicitly contains electricity data from elderlies from where we could extract the appliance usage patterns used in that Chapter. This dataset comes along with a rich metadata where surveys have been performed and aspects such as the daily routines of residents are available. Besides, to the best of our knowledge, this is the largest dataset that contains electrical consumption data from single elderly residents.

A.4 The UK-DALE dataset: UK Domestic Appliance-Level Electricity dataset

It records the power demand from six houses in UK during 2 years. The sampling interval is six seconds for both the aggregated and disaggregated data. However, three out the six households are sampling at 16kHz [Kelly and Knottenbelt, 2015b]. This dataset was used in Section 5.4.2 because it contains fully detailed metadata and it has valuable information of a family household. This house was compared to the single elderly resident households in the HES dataset to prove the utility of the adopted monitoring approaches in elderlies.

A.5 The Colden Common dataset

The LGCP approach presented in Section 5.1 was the result of an international research stay in University of Southampton and it was published in [Alcalá et al., 2015a]. In the aforementioned paper, we adopt the difference hidden Markov model (dHMM) [Parson et al., 2012] to disaggregate the kettle using a collecting dataset from the village of Colden Common in the UK. It consists of aggregated and disaggregated power data from 13 households. This dataset was suitable for our experiments as it contains data collected at 1Hz. We do not mention the Colden Common dataset previously because it was used to disaggregate using the dHMM and this is not a contribution of this thesis. However, we do mention it here because its outcome was used on the LGCP and it is part of the work done in this research.

Appendix B

Metrics

Metrics have been always at the table in discussions about NILM. So far, there has not been standardised metric used for everyone in the domain. This is currently changing and researches are trying to use the same metrics to allow comparison between works. We describe here the metrics used in this thesis and explain why their use, given a brief comparison with other NILM metrics. For more detailed about NILM metrics refer as to [Kelly and Knottenbelt, 2015a, Batra et al., 2014, Anderson et al., 2012a].

B.1 Evaluation of the event detector performance

Firstly, the following concepts must be defined in a test where P+N samples are provided.

- **Positives (P)** is the number of events.
- **Negatives (N)** is the number of samples that are not events.
- **True Positives (TP)** is the number of correctly detected events.
- **False Negatives (FN)** is the number of missed events.
- **False Positives (FP)** is the number of wrongly detected events.

There are two approaches to evaluate the event detection performance: all events have the same importance or they are weighed depending on their corresponding energy consumption. Regarding the latter, they are metrics specified in [Anderson et al., 2012a] that

evaluate the performance depending on how much power is mismatched. Thus, the total power change metric measures the power change ΔP_e due to an event e and it defines two variables: ΔP_M the total power change due to FN or missed events and ΔP_{FP} the total power change due to FP. The ideal detector is such that $(\Delta P_M, \Delta P_{FP}) = (0, 0)$ and the closest one to this point is the optimal detector. Similarly, the average power change presents the same metrics (i.e. ΔP_M and ΔP_{FP}) averaged by the numbers of FN and FP, respectively. Both the total power change and the average power change weigh the event detection performance depending on the power consumption. For instance, a fridge has a higher power consumption than a light and, hence, missing fridge events will result in a lower performance than missing light events. In this thesis the goal is to monitor the human activity that lies behind the appliance usage pattern and, consequently, all events weight the same.

Hence, we use the following metrics where all events weight the same:

True Positive Rate (TPR):

The ratio between the number of correctly detected events and all produced events. It gives a measure of the detector's sensitivity.

$$TPR = \frac{TP}{TP + FN} \in [0, 1] \tag{B.1}$$

False Positive Rate (FPR):

The ratio between the number of wrongly detected events and all samples except the produced events. It is also called fall-out and it is the inverse of the specificity (i.e. $1 - specificity$).

$$FPR = \frac{FP}{FP + TN} \in [0, 1] \tag{B.2}$$

Optimal detector:

Receiver Operating Characteristic (ROC) curve is a graphical tool that helps optimise the detector. It has the FPR on the x-axis and the TPR on the y-axis, consequently, the optimal detector is such that draws the closest point to the ideal detector (i.e. the point (0,1)) as in (B.3). Changing the parameters in the detector, different points are drawn in

the (FPR, TPR) plane and a ROC curve is drawn. A guess random detector would draw a straight line from the point (0,0) to the point (1,1).

$$\psi_{opt} = \arg\min_{\forall\psi\in\Psi} ||(0,1) - (FPR, TPR)||^2 \tag{B.3}$$

Where:

- ψ is a detector with a specific parameter setting.
- Ψ is the set of detectors, given a set of parameter settings.
- ψ_{opt} is the highest performance, given a parameter setting.

Area Under Curve (AUC):

The AUC is also measured in the ROC curve. It might be equivalent to the optimal detector, however, it is a better indicator of the overall detector performance (i.e. when it is not optimised). This value is obtained by integrating the area under the ROC curve.

True Positive Percentage (TPP):

We introduced this metric along with the False Positive Percentage (FPP) metric in Section 4.2 to compare the proposed event detector with other approach that evaluates the performance with these metrics. In fact, $TPR = TPP$ because $P = TP + FN$.

$$TPP = \frac{TP}{P} \in [0,1] \tag{B.4}$$

False Positive Percentage (TPP):

It gives a ratio of FP with regards to P. Although this is not a real percentage because the result in (B.5) can be higher than 1. An optimal detector can be also drawn as in (B.6) where the ideal detector is the point (0,1). It is worth noting that, even though $TPP = TPR$, the optimal detector obtained in (B.6) draws a slightly different value for TPP than for TPR in (B.3) because $FPR \neq FPP$ and, therefore, the optimisation is different. Likewise, an optimised detector in TPP and FPP gives more weights to FP.

$$FPP = \frac{FP}{P} \in [0, \infty) \tag{B.5}$$

$$\psi_{opt} = \arg \min_{\forall \psi \in \Psi} ||(0,1) - (FPP, TPP)||^2 \tag{B.6}$$

B.2 Evaluation of the Load Signature Classification

Given a classification algorithm and a set of test samples (i.e. P+N) we define the following concepts:

- **Positives (P)** is the total number of labelled samples.
- **Negatives (N)** is the total number of unknown label samples. For instance, if the classifier only knows about 2 appliance types (e.g. microwave or heating) and the arrival sample is from a fridge, then that is an unknown label sample.
- **True Positives (TP)** is the number of correctly labelled events by the classifier. For example, if the arrival sample is from a microwave and the classifier correctly labels it as a microwave.
- **False Negatives (FN)** is the number of unlabelled events. Following the previous examples, if the arrival sample is from a microwave and the classifier cannot label.
- **False Positives (FP)** is the number of wrongly labelled events. Following the previous examples, if the arrival sample is from a microwave and the classifier wrongly labels it as something else.

Recall:

It is evaluated as the TPR in (B.1)

$$recall = \frac{TP}{TP + FN} \tag{B.7}$$

Precision:

It indicates the ratio between correctly labelled events and all labeled events:

$$precision = \frac{TP}{TP + FP} \in [0, 1] \tag{B.8}$$

Accuracy:

It indicates how accurate the classifier is taking into account the correctly labelled events and the correctly discarded events, divided by the total number of events.

$$accuracy = \frac{TP + TN}{P + N} \in [0, 1] \tag{B.9}$$

F_β **score:**

It is a weighed mean between accuracy and precision where $\beta = \frac{w_{recall}}{wprecision}$ is the weighed ratio between recall and precision. If $\beta = 1$, precision and recall have the same weight. For appliance usage pattern learning, precision becomes more important that recall. Likewise, in the PCA-PQD classifier presented in Section 4.2 we use the $F_{0.5}$ score, where $\beta = 0.5$.

$$F_\beta = (1 + \beta^2) * \frac{precision * recall}{\beta^2 * precision + recall} \tag{B.10}$$

Accuracy (Acc):

This metric was introduced in [Gao et al., 2015] and was used in the comparative study for the PQD-PCA classifier in Section 4.2. It does not take into account the scenario where unknown appliances are plugged in and the classifier has to discard samples (i.e. it does not take into account the TN and N). Thus it can be seen as a percentage of correct predictions more than an accuracy metric.

$$Acc = \frac{\text{number of correct preditions}}{\text{total number of predictions}} \in [0, 1] \tag{B.11}$$

B.3 Others NILM metrics:

Apart from the detection and classification metrics, there are other metrics, mostly used in eventless NILM algorithms, that evaluate the performance of the NILM algorithm in terms of energy disaggregation. For instance, in [Kelly and Knottenbelt, 2015a] it is defined the relative error in total energy (B.12) and the mean absolute error (B.13). This type of metrics is not used because it is energy dependent and does not take into account events, however, they are worth mentioning.

$$\text{relative error in total energy} = \frac{|\hat{E} - E|}{max(E, \hat{E})} \tag{B.12}$$

where:

- E is the total energy consumption in the household.
- $\hat{E}$ is the total predicted energy consumption by the NILM algorithm.

$$\text{mean absolute error} = \frac{1}{N}\sum_{n=1}^{N}|\hat{y}[n] - y[n]| \tag{B.13}$$

where:

- y are the real measurements (e.g. the aggregated power consumption).
- $\hat{y}$ are the predicted values (e.g. the predicted power consumption).
- N is the total number of samples. For instance, $N = 24$ is the power consumption y is measured during a day with a sampling interval of one hour.
- n is the *n-th* sample of y.

Bibliography

[Alcalá et al., 2015a] Alcalá, J. M., Parson, O., and Rogers, A. (2015a). Detecting anomalies in activities of daily living of elderly residents via energy disaggregation and cox processes. In *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, pages 225–234, Seoul, South Korea. ACM.

[Alcalá et al., 2014] Alcalá, J. M., Ureña, J., and Hernández, Á. (2014). Event-based detector for non-intrusive load monitoring based on the Hilbert transform. In *Proceedings of the 19th IEEE International Conference on Emerging Technology and Factory Automation (ETFA)*, pages 1–4, Barcelona, Spain.

[Alcalá et al., 2015b] Alcalá, J. M., Ureña, J., and Hernández, Á. (2015b). Activity supervision tool using non-intrusive load monitoring systems. In *Proceedings of the 20th IEEE International Conference on Emerging Technologies Factory Automation (ETFA)*, pages 1–4, Luxembourg.

[Anderson et al., 2012a] Anderson, K. D., Bergés, M. E., Ocneanu, A., Benitez, D., and Moura, J. M. F. (2012a). Event detection for non intrusive load monitoring. In *Proceedings of the 38th Annual Conference of the IEEE Industrial Electronics Society (IECON 2012)*, pages 3312–3317, Montreal, QC, Canada.

[Anderson et al., 2012b] Anderson, K. D., Ocneanu, A., Benitez, D., Carlson, D., Rowe, A., and Berges, M. (2012b). BLUED: a fully labeled public dataset for Event-Based Non-Intrusive load monitoring research. In *Proceedings of the 2nd KDD Workshop on Data Mining Applications in Sustainability (SustKDD)*, pages 1–5, Beijing, China.

[Asada et al., 2003] Asada, H. H., Shaltis, P., Reisner, A., Rhee, S., and Hutchinson, R. C. (2003). Mobile monitoring with wearable photoplethysmographic biosensors. *IEEE Engineering in Medicine and Biology Magazine*, 22(3):28–40.

[Balaras et al., 2007] Balaras, C. A., Gaglia, A. G., Georgopoulou, E., Mirasgedis, S., Sarafidis, Y., and Lalas, D. P. (2007). European residential buildings and empirical assessment of the hellenic building stock, energy consumption, emissions and potential energy savings. *Building and Environment*, 42(3):1298 – 1314. `http://www.sciencedirect.com/science/article/pii/S0360132305004671`.

[Baranski and Voss, 2004] Baranski, M. and Voss, J. (2004). Detecting patterns of appliances from total load data using a dynamic programming approach. In *Proceedings of the 4th IEEE International Conference on Data Mining (ICDM '04)*, pages 327–330, Brighton, UK.

[Barsim et al., 2014] Barsim, K. S., Streubel, R., and Yang, B. (2014). An approach for unsupervised non-intrusive load monitoring of residential appliances. In *Proceedings of the 2nd International Workshop on Non-Intrusive Load Monitoring*, Austin, TX, USA.

[Barsim and Yang, 2015] Barsim, K. S. and Yang, B. (2015). Toward a semi-supervised non-intrusive load monitoring system for event-based energy disaggregation. In *Proceedings of the 3rd IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 58–62, Orlando, Florida, USA.

[Batra et al., 2014] Batra, N., Kelly, J., Parson, O., Dutta, H., Knottenbelt, W., Rogers, A., Singh, A., and Srivastava, M. (2014). Nilmtk: an open source toolkit for non-intrusive load monitoring. In *Proceedings of the 5th International Conference on Future Energy Systems*, pages 265–276, Cambridge, UK. ACM.

[Beall, 2016] Beall, A. (2016). Will computers run out of power? machines could use more than world's production of electricity by 2040. `http://www.dailymail.co.uk/sciencetech/article-3707040/Will-computers-run-power-Machines-use-world-s-production-electricity-2040-experts-claim.html`. Accessed: 12-11-2016.

[Beckel et al., 2014] Beckel, C., Kleiminger, W., Cicchetti, R., Staake, T., and Santini, S. (2014). The ECO Data Set and the Performance of Non-intrusive Load Monitoring Algorithms. In *Proceedings of the 1st ACM Conference on Embedded Systems*

for Energy-Efficient Buildings, BuildSys '14, pages 80–89, Memphis, TN, USA. ACM. http://doi.acm.org/10.1145/2674061.2674064.

[Belkin, 2013] Belkin (2013). Belkin energy disaggregation competition. https://www.kaggle.com/c/belkin-energy-disaggregation-competition. Accessed: 24-11-2016.

[Bonfigli et al., 2015] Bonfigli, R., Squartini, S., Fagiani, M., and Piazza, F. (2015). Unsupervised algorithms for non-intrusive load monitoring: An up-to-date overview. In *Proceedings of the 15th IEEE International Conference on Environment and Electrical Engineering (EEEIC)*, pages 1175–1180, Rome, Italy.

[Casellas et al., 2010] Casellas, F., Velasco, G., Guinjoan, F., and Piqué, R. (2010). El concepto de smart metering en el nuevo escenario de distribución eléctrica. *Actas del Seminario Anual de Automática y Electrónica Industrial, SAAEI*, pages 752–757. Bilbao. Spain.

[Celler et al., 2001] Celler, B. G., Lovell, N. H., Basilakis, J., Magrabi, F., and Mathie, M. (2001). Home telecare for chronic disease management. In *Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 4, pages 3586–3589 vol.4, Istambul, Turkey.

[Chan et al., 2000] Chan, W., So, A. T., and Lai, L. (2000). Harmonics load signature recognition by wavelets transforms. In *Proceedings of the International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (DRPT 2000)*, pages 666–671, City University London, UK,. IEEE.

[Chen et al., 2015] Chen, B., Fan, Z., and Cao, F. (2015). Activity recognition based on streaming sensor data for assisted living in smart homes. In *Proceedings of the 11th International Conference on Intelligent Environments (IE'15)*, pages 124–127, Prague, Czech Republic.

[Chen et al., 2011] Chen, F., Dai, J., Wang, B., Sahu, S., Naphade, M., and Lu, C.-T. (2011). Activity analysis based on low sample rate smart meters. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '11, pages 240–248, San Diego, CA, USA. ACM. http://doi.acm.org/10.1145/2020408.2020450.

[Cicchetti, 2014] Cicchetti, R. (2014). Nilm-eval: Disaggregation of real-world electricity consumption data. *Master's thesis, ETH Zurich*.

[Cole and Albicki, 2000] Cole, A. and Albicki, A. (2000). Nonintrusive identification of electrical loads in a three-phase environment based on harmonic content. In *Proceedings of the 17th IEEE Instrumentation and Measurement Technology Conference (IMTC 2000)*, volume 1, pages 24–29 vol.1, Baltimore, Maryland, USA.

[Cook et al., 2013] Cook, D. J., Crandall, A. S., Thomas, B. L., and Krishnan, N. C. (2013). Casas: A smart home in a box. *Computer*, 46(7).

[Corbishley and Rodriguez-Villegas, 2008] Corbishley, P. and Rodriguez-Villegas, E. (2008). Breathing detection: Towards a miniaturized, wearable, battery-operated monitoring system. *IEEE Transactions on Biomedical Engineering*, 55(1):196–204.

[Cox and Isham, 1980] Cox, D. R. and Isham, V. (1980). *Point processes*, volume 12. CRC Press.

[Cressie and Wikle, 2011] Cressie, N. and Wikle, C. K. (2011). *Statistics for spatio-temporal data*. John Wiley & Sons.

[De Baets et al., 2016] De Baets, L., Ruyssinck, J., Deschrijver, D., and Dhaene, T. (2016). Event detection in nilm using cepstrum smoothing. In *Proceedings of the 3rd International Workshop on Non-Intrusive Load Monitoring*, pages 1–4, Vancouver, Canada.

[Debes et al., 2016] Debes, C., Merentitis, A., Sukhanov, S., Niessen, M., Frangiadakis, N., and Bauer, A. (2016). Monitoring activities of daily living in smart homes: Understanding human behavior. *IEEE Signal Processing Magazine*, 33(2):81–94.

[DECC, 2014] DECC (2014). Department of energy & climate change: Smart metering equipment technical specifications: second version. `https://www.gov.uk/government/consultations/smart-metering-equipment-technical-specifications-second-version`. Accessed: 12-11-2016.

[Duarte, 2015] Duarte, M. (2015). Notes on scientific computing for biomechanics and motor control. `https://github.com/demotu/BMC`. Accessed: 12-11-2016.

[Farinaccio and Zmeureanu, 1999] Farinaccio, L. and Zmeureanu, R. (1999). 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.

[Fleury et al., 2010] Fleury, A., Vacher, M., and Noury, N. (2010). SVM-based multimodal classification of activities of daily living in health smart homes: Sensors, algorithms, and first experimental results. *IEEE Transactions on Information Technology in Biomedicine*, 14(2):274–283.

[Friston et al., 2007] Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., and Penny, W. (2007). Variational free energy and the Laplace approximation. *Neuroimage*, 34(1):220–234.

[Froehlich et al., 2011] Froehlich, J., Larson, E., Gupta, S., Cohn, G., Reynolds, M., and Patel, S. (2011). Disaggregated end-use energy sensing for the smart grid. *IEEE Pervasive Computing*, 10(1):28–39.

[Gao et al., 2014] Gao, J., Giri, S., Kara, E. C., and Bergés, M. (2014). Plaid: A public dataset of high-resoultion electrical appliance measurements for load identification research: Demo abstract. In *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*, BuildSys '14, pages 198–199, Memphis, USA. ACM. `http://doi.acm.org/10.1145/2674061.2675032`.

[Gao et al., 2015] Gao, J., Kara, E. C., Giri, S., Berg, M., et al. (2015). A feasibility study of automated plug-load identification from high-frequency measurements. In *Proceedings of the 3rd IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 220–224, Orlando, Florida, USA. IEEE.

[GPy, 2012] GPy (2012). GPy: A gaussian process framework in python. `http://github.com/SheffieldML/GPy`. Accessed: 12-11-2016.

[Gray and Tao, 2008] Gray, D. and Tao, H. (2008). Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *Proceedings of the 10th European Conference on Computer Vision (ECCV 2008)*, pages 262–275. Springer. `http://dx.doi.org/10.1007/978-3-540-88682-2_21`.

[Gupta et al., 2010] Gupta, S., Reynolds, M. S., and Patel, S. N. (2010). Electrisense: single-point sensing using emi for electrical event detection and classification in the

home. In *Proceedings of the 12th ACM international conference on Ubiquitous computing*, pages 139–148, Copenhagen, Denmark. ACM.

[Han and Lim, 2010] Han, D.-M. and Lim, J.-H. (2010). Smart home energy management system using ieee 802.15.4 and zigbee. *IEEE Transaction on Consumer Electronics*, 56(3):1403–1410. http://dx.doi.org/10.1109/TCE.2010.5606276.

[Hart, 1985] Hart, G. W. (1985). Prototype nonintrusive appliance load monitor. In *MIT Energy Laboratory Technical Report, and Electric Power Research Institute Technical Report*.

[Hart, 1992] Hart, G. W. (1992). Nonintrusive appliance load monitoring. *Proceedings of the IEEE*, 80(12):1870–1891.

[Hassan et al., 2014] Hassan, T., Javed, F., and Arshad, N. (2014). An empirical investigation of v-i trajectory based load signatures for non-intrusive load monitoring. *IEEE Transactions on Smart Grid*, 5(2):870–878.

[Hong et al., 2009] Hong, X., Nugent, C., Mulvenna, M., McClean, S., Scotney, B., and Devlin, S. (2009). Evidential fusion of sensor data for activity recognition in smart homes. *Pervasive and Mobile Computing*, 5(3):236–252.

[Hoque and Stankovic, 2012] Hoque, E. and Stankovic, J. (2012). Aalo: Activity recognition in smart homes using active learning in the presence of overlapped activities. In *Proeedings of the 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*, pages 139–146, San Diego, California, United States.

[INE, 2004] INE (2004). Instituto nacional de estadística: Inversión de la tendencia demográfica (1900-2050). http://www.ine.es/inebmenu/indice.htm. Accessed: 12-11-2016.

[Jimenez et al., 2005] Jimenez, J. A., Mazo, M., Urena, J., Hernandez, A., Alvarez, F., Garcia, J. J., and Santiso, E. (2005). Using PCA in time-of-flight vectors for reflector recognition and 3-D localization. *IEEE Transactions on Robotics*, 21(5):909–924.

[Jingkun, 2015] Jingkun (2015). Appliance identification code for global sip. http://nbviewer.ipython.org/github/jingkungao/PLAID/blob/master/Appliance%20Identification%20Code%20for%20GlobalSip.ipynb. Accessed: 25-11-2016.

[Kakde and Gulhane, 2015] Kakde, A. and Gulhane, V. (2015). Real time composite user activity modelling using hybrid approach for recognition. In *Proceedings of the 1st IEEE International Conference on on Computer and Communication Technologies (ICECCT)*, pages 1–6, Coimbatore, Tamilnadu, India.

[Kato et al., 2009] Kato, T., Cho, H. S., Lee, D., Toyomura, T., and Yamazaki, T. (2009). Appliance recognition from electric current signals for information-energy integrated network in home environments. In *Proceedings of the 7th International Conference on Smart Homes and Health Telematics (ICOST2009)*, pages 150–157, Tours, France. Springer.

[Katz et al., 1963] Katz, S., Ford, A. B., Moskowitz, R. W., Jackson, B. A., and Jaffe, M. W. (1963). Studies of illness in the aged: the index of adl: a standardized measure of biological and psychosocial function. *Jama*, 185(12):914–919. Basis of ADLs.

[Kautz et al., 2003] Kautz, H., Etzioni, O., Fox, D., Weld, D., and Shastri, L. (2003). Foundations of assisted cognition systems. *University of Washington, Computer Science Department, Technical Report*.

[Kelly, 2016] Kelly, J. (2016). *Disaggregation of Domestic Smart Meter Energy Data*. PhD thesis, Imperial College London.

[Kelly and Knottenbelt, 2015a] Kelly, J. and Knottenbelt, W. (2015a). Neural NILM. *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments - BuildSys '15*, pages 55–64. `http://dl.acm.org/citation.cfm?doid=2821650.2821672`.

[Kelly and Knottenbelt, 2015b] Kelly, J. and Knottenbelt, W. (2015b). The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes. *Scientific data*, 2:150007. `http://arxiv.org/abs/1404.0284`.

[Kelly and Knottenbelt, 2016] Kelly, J. and Knottenbelt, W. (2016). Does disaggregated electricity feedback reduce domestic electricity consumption? a systematic review of the literature. In *Proceedings of the 3rd International Workshop on Non Intrusive Load Monitoring*.

[Kim et al., 2011] Kim, H., Marwah, M., Arlitt, M. F., Lyon, G., and Han, J. (2011). Unsupervised disaggregation of low frequency power measurements. In *Proceedings*

of the Siam International Conference on Data Mining (SDM 11), volume 11, pages 747–758, Mesa, Arizona, USA. SIAM.

[Kolter et al., 2010] Kolter, J. Z., Batra, S., and Ng, A. Y. (2010). Energy disaggregation via discriminative sparse coding. In *Proceedings of Advances in Neural Information Processing Systems (NIPS 2010)*, pages 1153–1161, Vancouver, Canada.

[Kolter and Johnson, 2011] Kolter, J. Z. and Johnson, M. J. (2011). Redd: A public data set for energy disaggregation research. In *Proceedings of Workshop on Data Mining Applications in Sustainability (SIGKDD)*, volume 25, pages 59–62, San Diego, CA. Citeseer.

[Kolter et al., 2012] Kolter, Z., Jaakkola, T., and Kolter, J. Z. (2012). Approximate Inference in Additive Factorial HMMs with Application to Energy Disaggregation. *Proceedings of the International Conference on Artificial Intelligence and Statistics*, XX:1472–1482. `http://people.csail.mit.edu/kolter/lib/exe/fetch.php?media=pubs:kolter-aistats12.pdf`.

[Lam et al., 2007] Lam, H. Y., Fung, G. S. K., and Lee, W. K. (2007). A novel method to construct taxonomy electrical appliances based on load signaturesof. *IEEE Transactions on Consumer Electronics*, 53(2):653–660.

[LaPlante, 2010] LaPlante, M. P. (2010). The classic measure of disability in activities of daily living is biased by age but an expanded iadl/adl measure is not. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 65(6):720–732. Basis of ADLs.

[Li and Lin, 2015] Li, M. and Lin, H. J. (2015). Design and implementation of smart home control systems based on wireless sensor networks and power line communications. *IEEE Transactions on Industrial Electronics*, 62(7):4430–4442.

[Liang et al., 2010a] Liang, J., Ng, S. K., Kendall, G., and Cheng, J. W. (2010a). Load signature study part ii: Disaggregation framework, simulation, and applications. *IEEE Transactions on Power Delivery*, 25(2):561–569.

[Liang et al., 2010b] Liang, J., Ng, S. K. K., Kendall, G., and Cheng, J. W. M. (2010b). Load signature study part i: Basic concept, structure, and methodology. *IEEE Transactions on Power Delivery*, 25(2):551–560.

[Lim et al., 2008] Lim, J.-H., Jang, H., Jang, J., and Park, S. J. (2008). Daily activity recognition system for the elderly using pressure sensors. In *Proceedings of the 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5188–5191, Vancouver, Canada.

[Massot et al., 2013] Massot, B., Noury, N., Gehin, C., and McAdams, E. (2013). On designing an ubiquitous sensor network for health monitoring. In *Proceedings of the 15th IEEE International Conference on e-Health Networking, Applications and Services (Healthcom 2013)*, pages 310–314, Lisbon, Portugal.

[Mathie et al., 2001] Mathie, M. J., Basilakis, J., and Celler, B. G. (2001). A system for monitoring posture and physical activity using accelerometers. In *Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 4, pages 3654–3657 vol.4, Istanbul, TURKEY.

[Mathie et al., 2004] Mathie, M. J., Coster, A. C., Lovell, N. H., and Celler, B. G. (2004). Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement. *Physiological measurement*, 25(2):R1.

[McWilliam and Purvis, 2006] McWilliam, R. and Purvis, A. (2006). Potential uses of embedded RFID in appliance identification. *International appliance manufacturing.*, 10:100–107.

[MINETAD, 2011] MINETAD (2011). Ministerio de Energía Turismo y Agenda Digital: Smart Grids a Evolución de la Red Eléctrica. Technical report. `http://www.minetad.gob.es/industria/observatorios/SectorElectronica/Actividades/2010/Federaci%C3%B3n%20de%20Entidades%20de%20Innovaci%C3%B3n%20y%20Tecnolog%C3%ADa/SMART_GRIDS_Y_EVOLUCION_DE_LA_RED_ELECTRICA.pdf`. Accessed: 05/12/2016.

[Møller et al., 1998] Møller, J., Syversveen, A. R., and Waagepetersen, R. P. (1998). Log gaussian cox processes. *Scandinavian journal of statistics*, 25(3):451–482.

[Muszyńska and Rau, 2012] Muszyńska, M. M. and Rau, R. (2012). The old-age healthy dependency ratio in europe. *Journal of population ageing*, 5(3):151–162.

[Nikamalfard et al., 2012] Nikamalfard, H., Zheng, H., Wang, H., Jeffers, P., Mulvenna, M., McCullagh, P., Martin, S., Wallace, J., Augusto, J., Carswell, W., Taylor, B.,

and McSorley, K. (2012). Knowledge discovery from activity monitoring to support independent living of people with early dementia. In *Proceedings of the IEEE-EMBS International Conference on Biomedical and Health Informatics*, pages 910–913, Hong Kong, China.

[NILM-wiki, 2016] NILM-wiki (2016). Companies offering nilm products and services. `http://wiki.nilm.eu/index.php?title=Companies_offering_NILM_products_and_services`. Accessed: 12-11-2016.

[Noury et al., 2011] Noury, N., Berenguer, M., Teyssier, H., Bouzid, M. J., and Giordani, M. (2011). Building an index of activity of inhabitants from their activity on the residential electrical power line. *IEEE Transactions on Information Technology in Biomedicine*, 15(5):758–766.

[Noury et al., 2012] Noury, N., Quach, K. A., Berenguer, M., Bouzi, M. J., and Teyssier, H. (2012). A feasibility study of using a smartphone to monitor mobility in elderly. In *Proceedings of the IEEE 14th International Conference on e-Health Networking, Applications and Services (Healthcom 2012)*, pages 423–426, Beijing, China.

[Noury et al., 2009] Noury, N., Quach, K. A., Berenguer, M., Teyssier, H., Bouzid, M. J., Goldstein, L., and Giordani, M. (2009). Remote follow up of health through the monitoring of electrical activities on the residential power line - preliminary results of an experimentation. In *Proceedings of the 11th International Conference on e-Health Networking, Applications and Services (Healthcom 2009)*, pages 9–13, Sydney, Australia.

[Nugent et al., 2008] Nugent, C. D., Hong, X., Hallberg, J., Finlay, D., and Synnes, K. (2008). Assessing the impact of individual sensor reliability within smart living environments. In *Proceedings of the 4th Annual IEEE Conference on Automation Science and Engineering*, pages 685–690, Washington DC, U.S.A.

[Oliver et al., 2002] Oliver, N., Horvitz, E., and Garg, A. (2002). Layered representations for human activity recognition. In *Proceedings of the 4th IEEE International Conference on Multimodal Interfaces (ICMI 2002)*, pages 3–8, Pittsburgh, PA, USA.

[Parson, 2014] Parson, O. (2014). *Unsupervised Training Methods for Non-intrusive Appliance Load Monitoring from Smart Meter Data*. PhD thesis, University of Southampton. `http://eprints.soton.ac.uk/364263/`.

[Parson, 2015a] Parson, O. (2015a). Overview of the nilm field. http://blog.oliverparson.co.uk/2015/03/overview-of-nilm-field.html. Accessed: 12-11-2016.

[Parson, 2015b] Parson, O. (2015b). What even is supervised/unsupervised disaggregation? http://blog.oliverparson.co.uk/2015/05/what-even-is-supervisedunsupervised.html. Accessed: 12-11-2016.

[Parson et al., 2011] Parson, O., Ghosh, S., Weal, M., and Rogers, A. (2011). Using hidden markov models for iterative non-intrusive appliance monitoring. In *Proceedings of the 25th Annual Conference on Neural Information Processing Systems (NIPS 2011): Workshop on Machine Learning for Sustainability*, pages 1–4.

[Parson et al., 2012] Parson, O., Ghosh, S., Weal, M., and Rogers, A. (2012). Non-intrusive load monitoring using prior models of general appliances types. In *Proceedings of the 26th Conference on Artificial Intelligence (AAAI-12)*, pages 356–362. http://eprints.soton.ac.uk/336812/.

[Parson et al., 2014] Parson, O., Ghosh, S., Weal, M., and Rogers, A. (2014). An unsupervised training method for non-intrusive appliance load monitoring. *Artificial Intelligence*, 217:1–19.

[Patel et al., 2012] Patel, S., Park, H., Bonato, P., Chan, L., and Rodgers, M. (2012). A review of wearable sensors and systems with application in rehabilitation. *Journal of neuroengineering and rehabilitation*, 9(1):1.

[Patel et al., 2007] Patel, S. N., Robertson, T., Kientz, J. A., Reynolds, M. S., and Abowd, G. D. (2007). At the flick of a switch: Detecting and classifying unique electrical events on the residential power line. In *Proceedings of the 9th International Conference on Ubiquitous Computing (UbiComp 2007)*, pages 271–288, Innsbruck, Austria. Springer.

[Patricio, 2011] Patricio, B. (2011). Modelo de transicion demográfica. http://campus.belgrano.ort.edu.ar/cienciassociales/articulo/204483/modelo-de-transicion-demografica. Accessed: 12-11-2016.

[Pereira et al., 2014] Pereira, L., Quintal, F., Gonçalves, R., and Nunes, N. J. (2014). Sustdata: A public dataset for ict4s electric energy research. In *Proceedings of the*

2nd International Conference on ICT for Sustainability (ICT4S 2014), ISBN: 978-94-62520-22-6, Stockholm, Sweden.

[Rakowsky, 2007] Rakowsky, U. K. (2007). Fundamentals of the Dempster-Shafer theory and its applications to reliability modeling. *International Journal of Reliability, Quality and Safety Engineering*, 14(06):579–601.

[Rashidi et al., 2011] Rashidi, P., Cook, D. J., Holder, L. B., and Schmitter-Edgecombe, M. (2011). Discovering activities to recognize and track in a smart environment. *IEEE Transactions on Knowledge and Data Engineering*, 23(4):527–539.

[Sanduleac et al., 2015] Sanduleac, M., Albu, M., Martins, J., Alacreu, M. D., and Stanescu, C. (2015). Power quality assessment in lv networks using new smart meters design. In *Proceedings of the 9th International Conference on Compatibility and Power Electronics (CPE)*, pages 106–112, Caparica, Lisbon, Portugal.

[Scanaill et al., 2006] Scanaill, C. N., Carew, S., Barralon, P., Noury, N., Lyons, D., and Lyons, G. M. (2006). A review of approaches to mobility telemonitoring of the elderly in their living environment. *Annals of Biomedical Engineering*, 34(4):547–563.

[Scikit, 2016] Scikit (2016). Estimation algorithm expectation-maximization in the scikit-learn module for python. `http://scikit-learn.org/stable/modules/mixture.html#expectation-maximization`. Accessed: 01-12-2016.

[Sendin et al., 2012] Sendin, A., Berganza, I., Arzuaga, A., Pulkkinen, A., and Kim, I. H. (2012). Performance results from 100,000+ prime smart meters deployment in spain. In *Proceedings of the 3rd IEEE International Conference on Smart Grid Communications (SmartGridComm 2012)*, pages 145–150, Tainan City, Taiwan.

[Shaw and Laughman, 2007] Shaw, S. R. and Laughman, C. R. (2007). A kalman-filter spectral envelope preprocessor. *IEEE Transactions on Instrumentation and Measurement*, 56(5):2010–2017.

[Short, 2013] Short, D. (2013). Edge: Doug short. `https://rpseawright.wordpress.com/2013/03/04/edge-doug-short/`. Accessed: 12-11-2016.

[Song et al., 2014] Song, H., Kalogridis, G., and Fan, Z. (2014). Short paper: Time-dependent power load disaggregation with applications to daily activity monitoring.

In *Proceedings of the 2014 IEEE World Forum Internet of Things (WF-IoT)*, pages 183–184, Seoul, South Korea.

[Srinivasan et al., 2006] Srinivasan, D., Ng, W. S., and Liew, A. C. (2006). Neural-network-based signature recognition for harmonic source identification. *IEEE Transactions on Power Delivery*, 21(1):398–405.

[Tapia et al., 2004] Tapia, E. M., Intille, S. S., and Larson, K. (2004). Activity recognition in the home using simple and ubiquitous sensors. In *Proceedings of the International Conference on Pervasive Computing*, pages 158–175, Vienna, Austria. Springer.

[Terzija et al., 2007] Terzija, V. V., Stanojevic, V., Popov, M., and van der Sluis, L. (2007). Digital metering of power components according to IEEE standard 1459-2000 using the Newton-type algorithm. *IEEE Transactions on Instrumentation and Measurement*, 56(6):2717–2724.

[Triggle, 2015] Triggle, N. (2015). One million older people in need 'struggle alone'. `http://www.bbc.com/news/health-34576012`. Accessed: 12-11-2016.

[Turk and Pentland, 1991] Turk, M. A. and Pentland, A. P. (1991). Face recognition using eigenfaces. In *Proceedings of the 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'91)*, pages 586–591, Lahaina, Maui, Hawaii, USA. IEEE.

[Vadacchino et al., 2013] Vadacchino, V., Denda, R., and Fantini, G. (2013). First results in endesa smart metering roll-out. In *Proceedings of the 22nd International Conference Exhibition on Electricity Distribution (CIRED 2013)*, pages 1–4, Stockholm, Sweden.

[van der Zanden, 2011] van der Zanden, G.-J. (2011). The smart grid in europe 2012-2016: Technologies, market forecasts and utility profiles. `http://www.greentechmedia.com/research/report/the-smart-grid-in-europe-2012`. Accessed: 12-11-2016.

[Wang et al., 2016] Wang, A., Chen, G., Yang, J., Zhao, S., and Chang, C. Y. (2016). A comparative study on human activity recognition using inertial sensors in a smartphone. *IEEE Sensors Journal*, 16(11):4566–4578.

[Weiss et al., 2012] Weiss, M., Helfenstein, A., Mattern, F., and Staake, T. (2012). Leveraging smart meter data to recognize home appliances. In *Proceedings of the 2012*

IEEE International Conference on Pervasive Computing and Communications (PerCom 2012), pages 190–197, Lugano, Switzerland.

[Wikipedia, 2016a] Wikipedia (2016a). Expectation-maximization algorithm. `https://en.wikipedia.org/wiki/Expectation%E2%80%93maximization_algorithm`. Accessed: 01-12-2016.

[Wikipedia, 2016b] Wikipedia (2016b). Smart meter. `https://en.wikipedia.org/wiki/Smart_meter`. Accessed: 12-11-2016.

[Wild et al., 2015] Wild, B., Barsim, K. S., and Yang, B. (2015). A new unsupervised event detector for non-intrusive load monitoring. In *Proceedings of the 3rd IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 73–77, Orlando, Florida, USA.

[Xu et al., 2010] Xu, Y., Wang, H., Cui, Z., Dong, F., and Yan, Y. (2010). Separation of gas-liquid two-phase flow through independent component analysis. *IEEE Transactions on Instrumentation and Measurement*, 59(5):1294–1302.

[Zeifman and Roth, 2011] Zeifman, M. and Roth, K. (2011). Nonintrusive appliance load monitoring: Review and outlook. *IEEE Transactions on Consumer Electronics*, 57(1):76–84.

[Zhang et al., 2014] Zhang, X., Kato, T., and Matsuyama, T. (2014). Learning a context-aware personal model of appliance usage patterns in smart home. In *Proceedings of the 2014 IEEE Innovative Smart Grid Technologies - Asia (ISGT ASIA)*, pages 73–78, Kuala Lumpur, Malaysia.

[Zimmermann et al., 2012] Zimmermann, J.-P., Evans, M., Griggs, J., King, N., Harding, L., Roberts, P., and Evans, C. (2012). Household Electricity Survey A study of domestic electrical product usage. Technical report, Intertek. `https://www.gov.uk/government/collections/household-electricity-survey`. Accessed: 12-11-2016.