UNIVERSITA' DEGLI STUDI DI PADOVA

Dipartimento di Ingegneria Industriale DII

Corso di Laurea Magistrale in Ingegneria dell'Energia Elettrica

Non – intrusive load monitoring: Use of low resolution steady state features to disaggregate household appliances

Relatore

Prof. Ing. Pietro Fiorentin

Supervisor

Dr. Barry Hayes

Studente: Avogaro Jacopo

1178720

Anno Accademico 2019/2020

INDEX

SUMMARY

The possibility to recognize single appliances usage and power consumption from a single measuring device in the house is referred as Non-Intrusive Load Monitoring. Since the middle 80s algorithms and methods have been proposed, but a satisfactory solution is still not available. This thesis is focused on appliances disaggregation using low resolution power measurement, specifically active and reactive power provided with 1 s granularity. Many researches are focused on this kind of data, because low resolution active power is the only feature available today from domestic smart-meters. Unfortunately from this power signal it is not possible to extract many features that require higher resolution to be computed (harmonics, current waveform, etc.). This reduces the capability to tell one appliance from another.

Many methods have been proposed in literature, but experimental evaluation is often limited. This work focuses on understanding capability and limits of low resolution steady state features to disaggregate significant domestic appliances. Three algorithms have been used, both supervised and unsupervised. Two of them (EICCA-NILM and Weiss) are recent and, at the moment of writing, lack of significant experimental validation in literature, regarding our knowledge. The third one (Hart algorithm) is the first NILM algorithm proposed in history.

A good number of experiments have been run using different datasets in order to understand which appliances show consistent results in disaggregation metrics. This thesis details how only for some appliance groups results are satisfactory, whereas for other significant devices more refinements seem necessary, and probably features used are not enough to disaggregate them.

In addition to already mentioned features, time of occurrence of events can be extracted from low resolution power measures. This has been investigated in this thesis, showing how performances of supervised methods can increase with time of usage, but with limitations here exposed.

The thesis is structured as follows. Chapter 1 gives general definitions and presents state of the art of NILM research. Chapter 2 is about unsupervised algorithm used (EICCA-NILM) and experimental results achieved. In Chapter 3 supervised work is exposed and results are discussed. Several appendices are present with detailed tables and charts about experiments.

(ITALIAN VERSION)

La possibilità di riconoscere l'uso dei singoli elettrodomestici e il loro consumo energetico a partire da un solo dispositivo di misura è l'obiettivo delle tecniche di Non – Intrusive Load Monitoring. A partire dalla metà degli anni 80 diversi algoritmi e metodi sono stati proposti, ma una soluzione completa e soddisfacente non è ancora disponibile. Questa tesi si concentra sulla disaggregazione dei carichi componenti un' utenza domestica utilizzando quella che in letteratura si definisce misura di potenza a bassa risoluzione, nello specifico potenza attiva ed eventualmente reattiva campionate con periodo di uno o alcuni secondi. Un gran numero di ricerche si concentra su questa premessa poiché la potenza attiva a bassa risoluzione è l'unica grandezza attualmente ottenibile dagli smartmeter installati nelle abitazioni di molti stati. Purtroppo da questo segnale di potenza non è possibile estrarre molte informazioni che richiederebbero risoluzione maggiore (armoniche, forma d'onda della corrente, ecc). Questo riduce di molto la capacità di distinguere tra loro i dispositivi.

Molti metodi sono stati proposti negli anni, ma la loro validazione sperimentale è quasi sempre limitata. Questa tesi vuole investigare la capacità delle misure a bassa risoluzione di distinguere tra loro gli elettrodomestici più significativi. Si è scelto di utilizzare tre algoritmi, comprendendo soluzioni supervisionate e non supervisionate. Due di questi (EICCA-NILM e Weiss) sono recenti e, al momento della scrittura, non è a nostra conoscenza una loro validazione sperimentale significativa. L'algoritmo di Hart è storicamente il primo ad essere stato proposto, ed ha fornito la struttura base per i lavori successivi.

In questo lavoro sono stati eseguiti numerosi esperimenti, utilizzando dataset differenti, con lo scopo di capire quali elettrodomestici mostrano risultati positivi e consistenti, avvalendosi dei parametri di valutazione più usati in letteratura. Questa tesi mostra come solo per certi gruppi di elettrodomestici i risultati si possono considerare soddisfacenti. Al contrario per altri dispositivi risulta necessario investigare ulteriormente. Probabilmente le informazioni utilizzate non sono sufficienti al fine di un loro riconoscimento efficace.

Nella parte conclusiva del lavoro si è provato a migliorare i risultati introducendo l'informazione temporale. Essa è facilmente estraibile dal segnale analizzato. Il lavoro qui svolto mostra come le performance dei metodi supervisionati possono migliorare con quest'informazione aggiuntiva, in presenza però di limitazioni qui esposte.

La tesi è strutturata come segue. Il Capitolo 1 propone le definizioni generali e presenta lo stato dell'arte della ricerca in campo NILM. Il Capitolo 2 riguarda l'algoritmo non supervisionato (EICCA - NILM) e i risultati sperimentali ottenuti. Nel Capitolo 3 il lavoro sugli algoritmi supervisionati è esposto e i risultati sono discussi. Molte appendici sono state inserite con tabelle e grafici dettagliati relativi agli esperimenti eseguiti.

1 INTRODUCTION

This chapter reports NILM general definitions, in order to give the reader an overview of this field of research. General categories and structure of algorithms are presented, in addition to features proposed in literature to detect loads. Most common evaluation metrics are presented too. Some publicly available datasets for NILM research are described, and in the last part a description of some significant algorithms is provided.

1.1 NILM DEFINITION

The acronym NILM (Non-Intrusive Load Monitoring) refers to techniques that extract the power consumption of single appliances out of aggregated power data. That means that the measuring device has to be installed only in a single point in the household, with no metering devices installed at individual appliance level.

1.2 ENERGY SAVING

The residential sector represents a significant part of the total energy demand of a country. [1] states that in the EU 30 % of the total electricity is used for residential buildings. In the U.S. buildings account for 40 % of primary energy and 73 % of electricity consumption [2]. Considering that the global energy demand will double by the end of 2030 [3], It results crucial to investigate methodologies that can improve energy saving in the residential sector. Several studies suggest that an effective saving can be obtained by giving to the consumer real-time appliance level consume information (direct feedback), as opposed to Indirect feedback given by monthly bills [3]. It is demonstrated that Energy Consumption Awareness (which appliances are operating at a certain time instant and how much electrical energy they are consuming) influences users to moderate their energy consumption. The results are economic benefits for consumers and reduction of energy required to the provider [4]. Possible reduction of building electricity consumption is predicted in the range 10-15% [2][3].

1.3 APPLIANCES CLASSIFICATION

For NILM’s purposes, the most complete classification of Appliances is the following [2]:

1. <u>Type I</u> – ON/OFF appliances
   Appliances with only 2 states of operation: ON and OFF. These appliances consume only one specific amount of power when active. Usually these appliances are purely resistive
   Examples are toaster or light bulb.

2. <u>Type II</u> – Multi-state appliances
   These are appliances with a finite number of operating states, also referred as Finite State Machines (FSM).
   A FSM model represents this kind of devices. Graphically it consists of several circles, representing the possible operating states and edges that connect the circles, representing the

possible state transitions. The magnitude of these transition is given by the difference between the states that they connect. An example is given in Figure 1.1.
Washing machine and stove burner are Multi-state appliances.

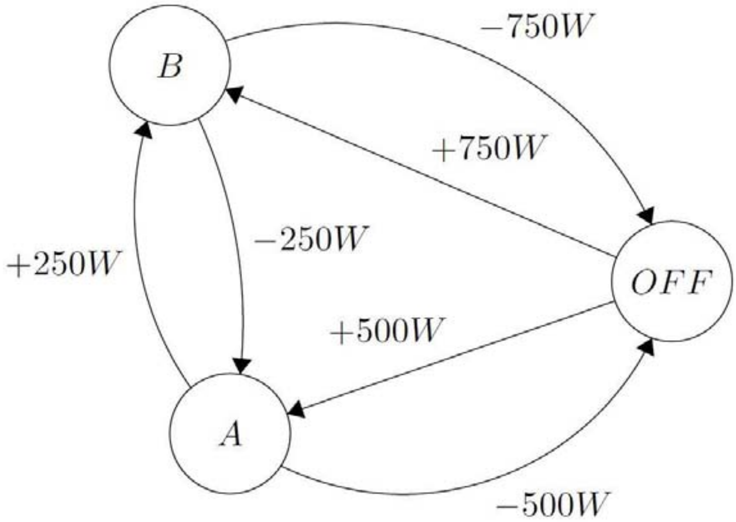


Figure 1.1: An example of a FSM model, taken from [1]

3. Type III – Continuously variable devices
   Appliances whose power draw characteristic is variable with no fixed number of states. They are also referred as Continuously Variable Devices (CVD).
   An example is a light dimmer.

4. Type IV – Permanent consumer devices
   Devices that remain active for days or weeks consuming energy at a constant rate. Some examples are hardwired smoke detector, telephone sets and cable TV receivers.

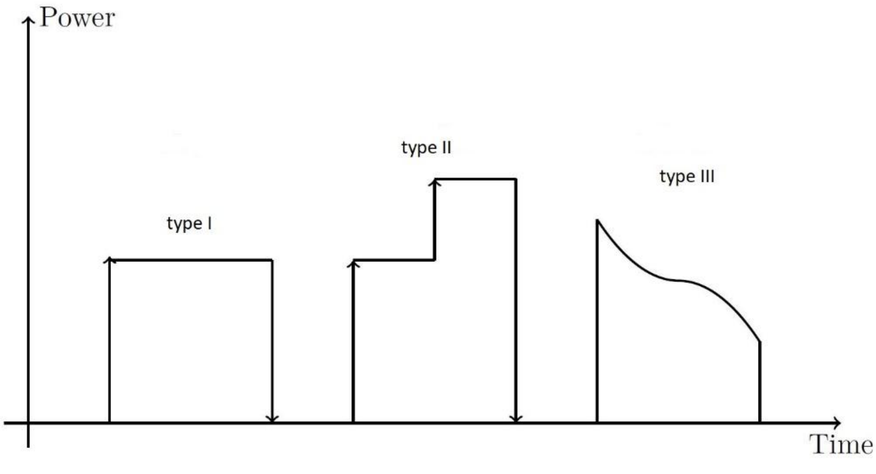


Figure 1.2: Schematic examples of type I, II and III appliance models, taken from [1]

1.4 NILM GENERAL FRAMEWORK

The problem can be formulated as follows. Given the aggregate power signal at the entry point of the meter $P(t)$, the task is to decompose it into his $p_i(t)$ component, with $i = 1,2,3,\dots\dots\dots,N$ and N is the number of active appliances at the moment t.

$$P(t) = p_1(t) + p_2(t) + p_3(t) + \cdots + p_N(t) \tag{1.1}$$

To achieve the disaggregation task, is important to discern and recognize appliance operations from the aggregate load measurements. For a perfect disaggregation, each load has to show a consumption pattern different from the others. This pattern is referred as "load signature" [3].
We will now describe the general framework of NILM.

Figure 1.3: General Framework of NILM approach

1.4.1 DATA ACQUISITION MODULE

The role of this module is to acquire measurement from the aggregate load at a rate that make possible to identify distinctive load patterns. A generic classification of the devices is given in [3]:

<u>LOW FREQUENCY</u> Energy meters
They are available in a range of sampling frequencies. According to the Nyquist-Shannon sampling criteria, the sampling rate of the meter has to be the double of the frequency of the highest harmonic of the signal that has to be measured. So an energy meter with a sampling rate of 500 Hz is able to capture up to the 5^{th} harmonic of the signal, if the fundamental is 50 Hz. If the purpose of the meter is the evaluation of traditional metrics such as active, reactive power or RMS voltage or current, the sampling rate can be lower, for example 100-120 Hz.

<u>HIGH FREQUENCY</u> Energy meters
These devices are capable to capture transient events or electrical noise. Their sampling rate is in the order of 10-100 MHz.

A problem often underlined by researchers is that commercially available low cost meters show a variation of 10-20% in data measurements.

1.4.2 APPLIANCE FEATURE EXTRACTION

a) Processing of raw data to compute power metrics, for example active and reactive power.
b) Detecting of events in processed data that represent state transition of appliances.

The methods used in detecting events are Steady-state or Transient based. The difference regards the features used to identify loads. Only steady-state methods are feasible if the low cost of the solution is relevant.

1.4.3 LOAD IDENTIFICATION (LEARNING AND INFERENCE)

SUPERVISED LEARNING APPROACHES

This part of the framework is manly based on supervised machine learning approaches, where labelled data are used to train a classifier to recognize loads from the feature extracted from the aggregate signal. There are two main categories of supervised learning approaches [1].

OPTIMISATION
This approach seeks a combination of appliances, present in a database, that better approximate the observed power signal. The combination of appliances chosen for every time step is the one that minimize an objective function. This approach suffer of an increasing complexity with the increase of appliances number, with consequent computational costs. Moreover performances decrease rapidly with the presence of unknown appliances. Considering that usually the complete set of appliances is unknown, this approach isn’t usually regarded as promising for the disaggregation task.

PATTERN RECOGNITION
In this approach the extracted features from the power signal are matched with power states in the database one by one. Usually state changes on the aggregate signal are represented in a feature space, which is a plane if the feature is a 2-dimensional vector, e.g. active and reactive power. Features in the plane are grouped into clusters and all the clusters are then compared with the features present in the database and obtained from training or from other sources.

UNSUPERVISED LEARNING APPROACHES

Recently researchers have been focused on developing methods that don’t need a priori information about the appliances present in the household that they are going to meter. The task is ambitious but these methods, if reliable, are very interesting because they don’t require any set up procedure, and that ease of installation can be very appealing for consumers. Well suited for this task are Hidden Markov Models methods.

1.5 FEATURES

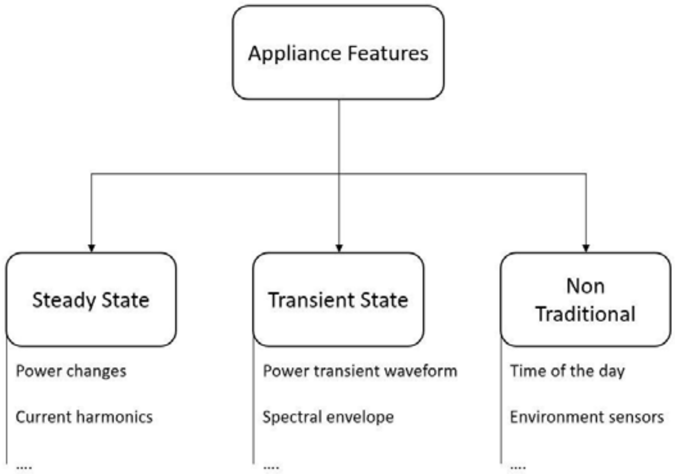


Figure 1.4: General Classification of Appliance Features

Appliance features can be classified as shown in Figure 4. Steady state analysis consider the stable state operation of the appliance, whereas transient state features are based on transitional states where appliance power consumption is unstable. Non-traditional features are additional environmental or behaviour information that can improve the disaggregation performances if added to the traditional features.

1.5.1 STEADY STATE FEATURES

- REAL AND REACTIVE POWER
  Most of NILM literature rely on real power only, because it is the only feature that can be provided by low cost smart meter that are present or are likely to be installed in households. For these meters the real power reading is provided with a frequency below 1 Hz. The use of both active and reactive power can improve the performance of the disaggregation algorithm, allowing to place the features on a P-Q plane. These features are more common used to track ON/OFF operation of appliances. Problems occur when different appliances features overlap, so there is no way to separate them. Generally this features alone can be used to detect only distinct big loads, well separated by the rest of the house’s appliances in terms of power draw.

- RMS VOLTAGE AND CURRENT, POWER FACTOR
  These time-domain features have been used in some studies to overcome some limitation of the previous features. The power factor can make easier to discriminate loads. Resistive loads have PF close to 1, whereas motor driven loads have lower values.

- STEADY STATE CURRENT HARMONICS
  Different loads are well characterized by current harmonics. This helps the disaggregation of non-linear loads, and of type III appliances that are not detectable with power-only features. In Figure 5, taken from [5], an example of different load waveforms is shown. The difference in harmonics content is clear.

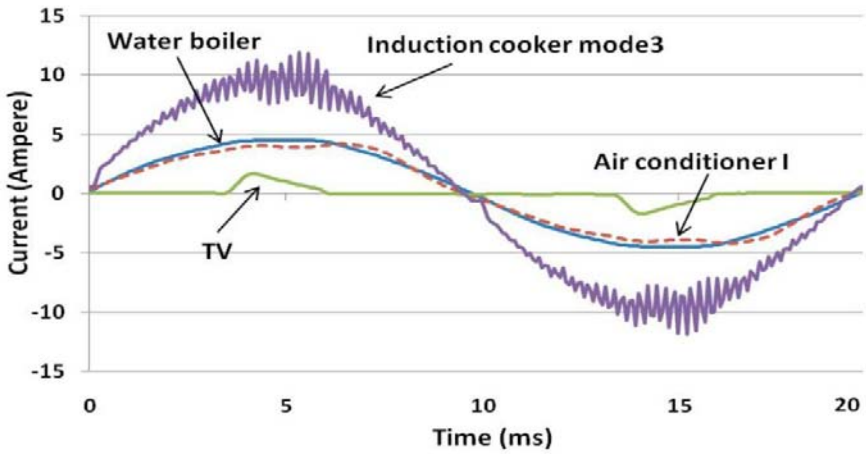


Figure 1.5: Different Loads Current Waveforms, taken from [5]

- STEADY STATE VOLTAGE NOISE
  It has been shown that appliances equipped with Switch Mode Power Supply can be characterized by analysing steady-state voltage noise generated during their operation.

- CURRENT WAVEFORM
  The current waveform is a very complete and distinctive feature to characterize different loads, as visible in Figure 5. High resolution in measurement is necessary for this analysis.

- INSTANTANEUS ADMITTANCE WAVEFORM
  Defined as the ratio between the instantaneous current and voltage waveforms.

$$IAW(t) = \frac{i(t)}{v(t)} \tag{1.2}$$

- INSTANTANEOUS POWER WAVEFORM (IPW)
  Defined as the product of instantaneous current and voltage waveform.

$$IPW = i(t) \times v(t) \tag{1.3}$$

- V – I TRAJECTORY
  The shape of this trajectory shows useful characteristics of loads. Features extracted by V-I Trajectories are asymmetry, looping direction and enclosed area.

1.5.2 TRANSIENT STATE FEATURES

Transient features are extracted during switching transients.

- SWITCHING TRANSIENT WAVEFORM

In [5] authors propose to compute instantaneous power for every half cycle and use the waveform obtained with these values. This feature can help to discriminate loads with different transient behaviours. For example a water boiler reaches the steady state in power consumption without any peak. Different is the case for an air conditioner, as shown in Figure 1.6.

- SPECTRAL ENVELOPE BASED ON SHORT TIME FOURIER TRANSFORM
  This feature is considered useful to characterize transient events. Some researcher proposed the use of Wavelet transform instead of the Fourier transform.

- HIGH FREQUENCY VOLTAGE NOISE DURING TRANSIENT
  Appliances, especially ones equipped with SMPS, emit voltage noise back to the main line. This noise can be measured from any outlet inside the home.

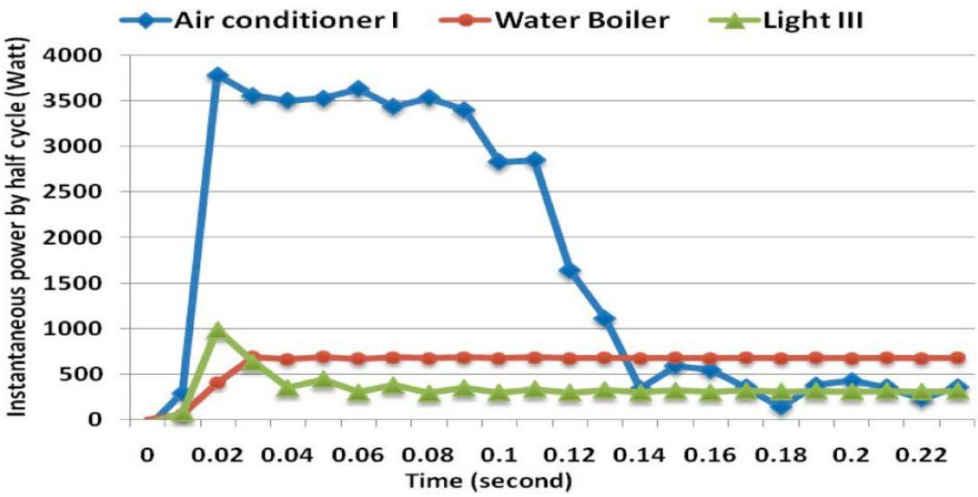


Figure 1.6: Switching Transient Waveforms of Different Loads, taken from [5]

1.5.3 NON TRADITIONAL FEATURES

- TIME OF THE DAY
  Some appliances can show a pattern of usage along the day. This additional features can be used to discriminate devices whose feature, for example active and reactive power, overlap.

- FREQUENCY OF APPLIANCE USAGE
  If there is a frequency of use of some appliances, this can help to decide if an eligible event belongs to that appliance or not.

- CORRELATION BETWEEN APPLIANCES
  Some appliances, such as washing machine and dryer, can show correlation between their usages.

- SENSORS IN THE HOME ENVIROMENT
  Some sensor could be installed in the house to collect information about heat dissipation, light or noise emission of certain appliances.

- WEATHER PATTERNS OF THE LOCATION
  Weather or temperature information of the location can influence usage of certain appliances.

1.6 EVALUATION METRICS

A problem often underlined by researchers is that there is not a “standard” set of metrics, shared by everyone who propose NILM approaches. This create problems in comparisons between different methods. Moreover results are often difficult to replicate, so there are not a lot of instruments to evaluate the performance of methods and to compare them. The following metrics are the ones presented by J. D. Kelly in [6], and are used in literature.

1.6.1 EVENT METRICS

These definitions are preliminary to compute event metrics. They are obtained by comparison of events assigned by algorithms to one appliance to ground truth events.

TRUE POSITIVE: correct claim that the appliance was used

TRUE NEGATIVE: correct claim that the appliance wasn’t used

FALSE POSITIVE: incorrect claim that an appliance was used

FALSE NEGATIVE: event present in ground truth data but not detected by the algorithm used

Following symbols are used in equations to compute metrics.

TP = number of true positives

TN = number of true negatives

FP = number of false positives

FN = number of false negatives

P = number of positives in ground truth

N = number of negatives in ground truth

RECALL

Recall metric gives a quantification of how many correct events have been detected in comparison to the total number of appliance events (TP + FN).

$$recall = \frac{TP}{TP+FN} \tag{1.4}$$

PRECISION

Precision gives information of how many correct events are present in assigned events. If it is high that means that most of assigned events are correct.

$$precision = \frac{TP}{TP+FP} \tag{1.5}$$

F-SCORE

F-score combines precision and recall in order to give an overall metric about disaggregation quality.

$$F - score = 2 \times \frac{precision \times recall}{precision + recall} \tag{1.6}$$

ACCURACY

Accuracy quantify how inferred events are close to represent ground truth data.

$$accuracy = \frac{TP+TN}{P+N} \tag{1.7}$$

1.6.2 ENERGY METRICS

Preliminary definitions for metric computation:

$$E = total\ actual\ energy$$

$$\hat{E} = total\ predicted\ energy$$

$$y_t^{(i)} = appliance\ i\ actual\ power\ at\ time\ t$$

$$\hat{y}_t^{(i)} = appliance\ i\ estimated\ power\ at\ time\ t$$

$$\bar{y}_t = aggregate\ actual\ power\ at\ time\ t$$

RELATIVE ERROR IN TOTAL ENERGY

This metric simply quantify how much estimated energy consumption is far from ground truth value.

$$relative\ error\ in\ total\ energy = \frac{\hat{E}-E}{\max(E,\hat{E})} \tag{1.8}$$

DEVIATION IN PERCENTAGE

Deviation in percentage is similar in meaning to relative error in total energy.

$$deviation\ in\ percentage = \frac{abs(\hat{E}-E)}{\hat{E}} \qquad (1.9)$$

MEAN ABSOLUTE ERROR

Mean absolute error is about how estimated power profile is different to the actual one in each sample for one appliance.

$$mean\ absolute\ error = {}^{1}/_{T} \sum_{t=1}^{T} |\hat{y}_t - y_t| \qquad (1.10)$$

TOTAL ENERGY CORRECTLY ASSIGNED

TECA metric quantify how much energy is assigned to correct appliances out of the total consumption.

$$total\ energy\ correctly\ assigned = 1 - \frac{\sum_{t=1}^{T}\sum_{i=1}^{n}\left|\hat{y}_t^{(i)} - y_t^{(i)}\right|}{2\sum_{t=1}^{T}\bar{y}_t} \qquad (1.11)$$

1.7 DATASETS

In order to make possible evaluation of NILM techniques, some datasets have been made available for researcher. Moreover this allow to make comparisons between different algorithms. We will now describe some of the existing datasets.

REDD [7]
Released by MIT researchers in 2011, it consists of whole-home and circuit/device specific electricity consumption for 6 houses in the U.S.A. Duration of measures is approximately 2 weeks, different for each house.
The whole home electric signal (current of both phase and voltage of one phase) is recorded at 15 kHz. Aggregate power on both phases is provided also at 1Hz. Power of each monitored appliance is given at 1/3 Hz. For each house the number of channels span between 11 and 26.

REFIT [8]
A UK- based dataset with recordings from 20 households. Each household has active power measured for 9 appliances and the main aggregate power, all recorded at 8-second intervals. This dataset aim to closely mimic the active power information that is given by common smart meter in the UK. Measurements span approximatively a 1-2 year long period between 2013 and 2015, with variations between each house.

UK DALE [9]
A UK- based dataset with recordings from 5 households. For houses 1, 2 and 5 the aggregated signal is recorded both at 16 kHz and at 1 Hz. For the 1 Hz reading the dataset provide voltage, real and apparent power. For house 3 and 4 the aggregate active power is recorded with frequency of 1/6 Hz. Number of appliances metered for each house is between 5 and 54. The individual appliances power signal is given at 6 seconds intervals.

Measures cover a period between 39 (House 3) and 786 days (House 1). Data were taken between 2012 and 2015.

GREEND [10]
This dataset claims to be the first power consumption dataset for Austria and Italy measured at 1 Hz sampling frequency. It consists of active power measurements of nine houses for a continuous duration of 1 year collected in Carinthia (Austria), and Friuli Venezia Giulia (Italy). There’s an average of 9 appliances metered for each house.

ECO [11]
This dataset contains data collected over 8 months in 6 houses in Switzerland. I provides 1 Hz measurements of voltage, current and phase shift for the 3 phases of each house. Six to ten smart plugs per household have been employed to provide ground truth data for appliances.

The following table, presented in [4], shows several available datasets with some of their characteristics.

Contribution	Dataset	Location	Duration per house	Number of houses	Appliance sample resolution	Aggregate sample resolution
[11]	REDD	USA	3-19 days	6	3 sec	1 sec & 15 kHz
[12]	BLUED	USA	8 days	1	state transition label	12 kHz
[13]	UMass Smart	USA	3 months	3	1 sec	1 sec
[14]	Tracebase	DE	N/A	15	1-10 sec	N/A
[15]	Pecan Street Sample	USA	7 days	10	1 min	1 min
[16]	HES	UK	1 or 12 months	251	2 or 10 min	2 or 10 min
[17]	AMPds	CDN	1 year	1	1 min	1 min
[18]	iAWE	IND	73 days	1	1 or 6 sec	1 sec
[19]	UK-DALE	UK	3-17 months	4	6 sec	1-6 sec & 16 kHz
[20]	GreenD	AT/IT	1 year	9	1 sec	1 sec
[21]	COMBED	IND	18 months	8	30 sec	30 sec
[22]	ECO	CH	8 months	6	1 sec	1 sec
[23]	BERDS	USA	1 year	N/A	20 sec	20 sec
[24]	SustData	PT	5 years	50	50 Hz	50 Hz

Table 1.1: Summary of Some NILM Datasets, taken from [4]

1.8 ALGORITHMS

Many algorithms have been proposed in literature. This section only describes some of them, without claiming to be complete. Overview of NILM algorithms can be found in [1] – [4].

1.8.1 SUPERVISED ALGORITHMS

PATTERN RECOGNITION ALGORITHMS

HART ALGORITHM

This algorithm was the one that started NILM research, in the middle 80's. It is described deeply in section 3.1.1.

WEISS ALGORITHM

Proposed in 2012, this algorithm rooted in Hart's work proposes a new approach to event detection. It is detailed in section 3.1.1.

DECISION TREE - BASED ALGORITHM [26]

This low complexity algorithm is based on low frequency active power data only. During training phase a Decision Tree is built using instructions presented in [27] and known appliances signatures. A DT is made by nodes that are connected to other two nodes. Terminal nodes not connected to other nodes are called *leaves*. Each *leaf node* is associated with an appliance or an undefined value. Classification process for each event consists of going across the tree starting from the *root node*, and ending up in a *leaf node*. If the final *leaf* is associated to an appliance, the event is assigned to that appliance. If the *leaf* is not associated to an appliance, a classification error is declared. In this method appliances are detected one at each time, moving from the bigger to the smallest one. Once one appliance is detected, it is removed from the aggregated signal, facilitating detection of following appliances.

OPTIMIZATION ALGORITHMS

INTEGER PROGRAMMING ALGORITHMS

These algorithms are based on optimization using integer programming. First proposed application in [28] uses as feature current waveforms of appliances. A database with waveforms and high sampling rate measuring system is needed. More recently other methods have been proposed using active or apparent power at low sampling rate, 1 Hz or lower. In this case the input required is the set of power states of analysed appliances. In [29] several constraints have been introduced in order to increase performances, i.e. to treat appliances with constant base consumption and to avoid unrealistic transitions from one set of appliances to another caused by small variations in power. Another method in this family can be found in [30]. In all these cases results presented are not obtained on real aggregate signal, but on an artificial one, created by addition of plug level power measurements of considered appliances. This makes results unimportant for real world applications, unless nearly all the devices connected to household mains are known.

MULTI- ALGORITHMS FRAMEWORKS

In [5] and [31] multi-scale, multi-algorithmic frameworks are presented. They extract several features from raw data, both transient and steady state. More algorithms are used to disaggregate loads producing candidates for each event. Then a decision mechanism evaluate which of the candidates is the most suitable. Method in [31] focuses on each load in order to determine which feature/method better identify it, Introducing also statistical methods for certain loads with variable behaviour. Case study for this method wasn’t a household, but a coast guard cutter ship.

1.8.2 UNSUPERVISED ALGORITHMS

DTW ALGORITHM [26]

This low sampling/active power based algorithm rely on Dynamic time warping, a time series based approach used to compare vectors of different length and with non-identical values. In that way not only turn on and turn off events characterize an appliance, but all power samplings during activity. A library of signatures is needed, and they are compared with events detected via pattern matching to classify them.

GSP ALGORITHMS

In [33] Graph signal processing is employed for NILM applications. Active Power at resolution lower than 1s is used as feature. GSP is an emerging field successfully used in image processing and signal filtering. A supervised use of GSP is presented in [32].

ANN ALGORITHMS

In [34] a method based on deep learning has been proposed. First an Edge detection algorithm extract potential appliance profiles from aggregate low resolution active power signal. Then an Artificial Neural Network is trained to classify these profiles and recognize appliance activity in testing period. ANN are used in [35] too but for a supervised algorithm.

HMM ALGORITHMS

In literature Hidden Markov Models – based algorithms are often referred as *state of the art* or benchmark of NILM applications. HMM are statistical models used with success in speech recognition, audio separation and other signal applications. The qualitative idea is that each appliance is a HMM, characterized by its possible power states. All the HMM develop in parallel during time producing as output their states. The only observable quantity is the aggregate signal, which is made by the sum of model's outputs. For that reason models are hidden. With these algorithms it is possible, from aggregate signal, to train models that are representative of real appliances.

> KOLTER'S ALGORITHM [36]
> This work proposed in 2012 is completely unsupervised and infer models only from aggregate signal, without any general prior knowledge about appliances.

> PARSON'S ALGORITHM [37]
> This method starts from general models of appliances, and then adapt them to real appliances in the household during training using only aggregate power signal, collected at 1/60 Hz.

Other HMM methods have been presented in literature, a review can be found in [4].

1.9 NILM - EVAL TOOLKIT [25]

Authors of [11] made publicly available *NILM – eval*, a Matlab-based framework for NILM research. It provides four algorithms and useful tools to evaluate results metrics. For this thesis this framework has been used as a starting point for all the experiments. In particular Weiss algorithm used in chapter 3 was already included in the framework. It has been expanded and adapted in certain parts in order to perform experiments better, because the version provided has been tested by its authors only in house 2 of ECO dataset. The other algorithms described in chapters 2 and 3 have been implemented by author of this thesis in *NILM - eval* environment. Evaluation tools provided by the framework have been adapted and used for all result metrics presented in this work. In particular section 3.2.5, where evaluation methodology for supervised work is presented, is rooted on *NILM-eval* evaluation tools, with adaptations necessary to treat correctly all analysed appliances.

2 UNSUPERVISED NILM WORK

Chapter 2 is organized as follows. Section 2.1 presents theoretically some data clustering algorithms and the way EICCA algorithm has been employed for NILM purposes. A brief presentation of results achieved in literature is included. Section 2.2 is about experiments that have been done for this work. First experiments goals are defined, then their structure and actual implementation is described. Results are then presented, focusing first on clusters emerged, then on introduction of reactive power. Finally some appliances are disaggregated and event metrics are discussed. Appendices related to this chapter are the following:

- Appendix A: Some other data clustering algorithms.
- Appendix B: Dates of days used for experiments in relative datasets.
- Appendix C: Parameters, result tables for experiments and relevant pictures.

2.1 METHODOLOGY

2.1.1 EICCA BASED NILM ALGORITHM

This section of the thesis is focused on NILM algorithm proposed in [15], [16] and [17]. The basic characteristic of this method is the usage of a clustering algorithm on aggregate data, in order to isolate clusters that can likely represent states of single appliances.

Competitive Agglomeration algorithm (abbreviated in CA, [13]) has been used in [15], whereas in [16] and [17] authors used the Entropy Index Constraints Competitive Agglomeration algorithm (abbreviated in EICCA, [14]), which has been used in this thesis.

The main advantage of using Competitive Agglomeration algorithms is that the number of clusters has not to be specified *a priori*. This condition is essential in NILM applications, because the number of appliances in households is hardly ever known, both in supervised and unsupervised work. The algorithm only needs an over specified number of clusters that are reduced automatically along iterations.

EICCA algorithm is a generalization of CA. As CA is built on Fuzzy C-means Algorithm (FCM), an Entropy Index Constraints Fuzzy C-means Algorithm has been used in [14] to obtain EICCA. FCM algorithms are also necessary for the initialization phase of CA. In this section EIC-FCM and EICCA are presented. The other two algorithms are shown in Appendix A.

In following lines the definitions of main symbols are given.

- INITIAL DEFINITIONS

Dataset to be clustered X:

$$X = \{x_j | x_j \in R^D, j = 1,2,\dots,N\},$$
$$N\ number\ of\ data\ points\ in\ the\ dataset, D\ number\ of\ dimensions\ of\ data$$

Number of Clusters C:

$$C, \quad (2 \le C < N)$$

Fuzzy Partition Matrix U:

$$U = (u_{ij})_{C\times N} \in R^{CN}, where\ R^{CN}\ represents\ the\ set\ of\ all\ possible\ C \times N\ real\ matrixes$$

In *Figure 2.1* a representation of *U* matrix is given. The element *Uij* of the matrix represent for the dataset j-th element his “degree of belonging” to the i-th cluster.

Vector of centroids V:

$V = \{v_i | v_i \in R^D, i = 1,2, \dots, C\}$

Elements of V are the centres assigned by the algorithm to each cluster, updated for each iteration.

Euclidean Norm || • ||: This symbol is used here for the Euclidean norm or 2 - norm.

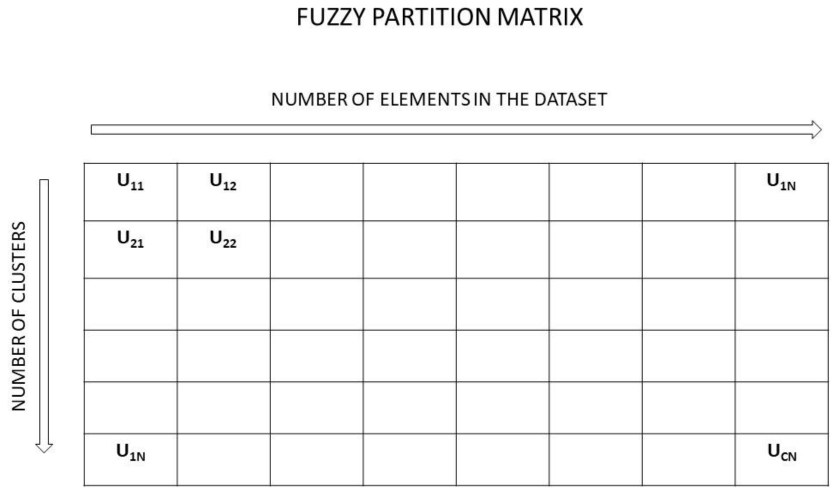


Figure 2.1: Graphic representation of Fuzzy Partition Matrix

2.1.2 EIC - FCM ALGORITHM

The objective function for EIC – FCM is the following:

$$\min J_{EIC-FCM}(U,V) = min \sum_{i=1}^{C} \sum_{j=1}^{N} u_{i,j} \|x_j - v_i\| \qquad (2.1)$$

With constraints:

$$\sum_{i=1}^{C} (u_{ij})^r = 1,\ 0 < r < 1,\ \forall j;\ u_{ij} \in [0,1], \forall\, i,j \qquad (2.2)$$

It is shown in [14] that from (2.1) and (2.2) it is possible to obtain the update equations for cluster centres and membership:

$$u_{ij} = \frac{(\|x_j - v_i\|^2)^{\frac{1}{r-1}}}{(\sum_{k=1}^{C} (\|x_j - v_k\|^2)^{\frac{r}{r-1}})^{\frac{1}{r}}},\ i = 1, \dots, C \qquad (2.3)$$

$$v_i = \frac{\sum_{j=1}^{N} u_{ij} x_j}{\sum_{j=1}^{N} u_{ij}}, i = 1, \dots, C \tag{2.4}$$

- ALGORITHM DESCRIPTION

As indicated in [14], the following are the 5 steps of the algorithm.

1) Given the Dataset X, the number of clusters C and the threshold ε, initialize $V^{(0)}$ (arbitrary centres for the iteration 0)
2) Compute $U^{(0)}$ using eq. (2.3)
3) Update $V^{(k+1)}$ using eq. (2.4)
4) Update $U^{(k+1)}$ using eq. (2.3)
5) If $\|V^{(k+1)} - V^{(k)}\| < \varepsilon$, then output U and V and exit, otherwise $k = k+1$ and go back to step 3

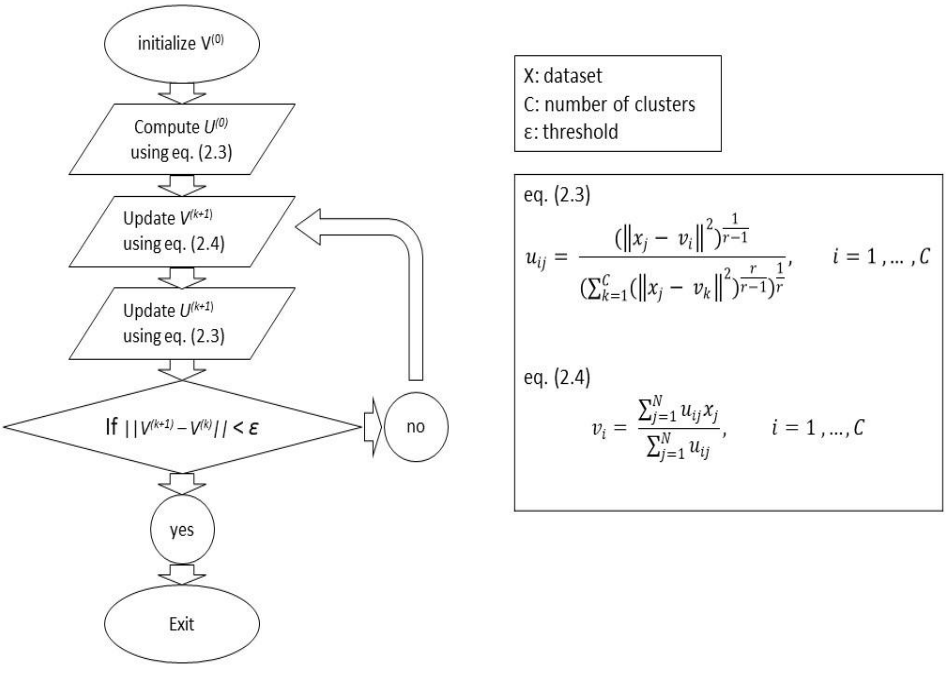


Figure 2.2: Flow chart of EIC-FCM algorithm

- APPLICATION ON BENCHMARK DATASETS

Figures *2.3* and *2.4* show behaviour of the algorithm with some 2D benchmark datasets for clustering. Datasets are available in [18].

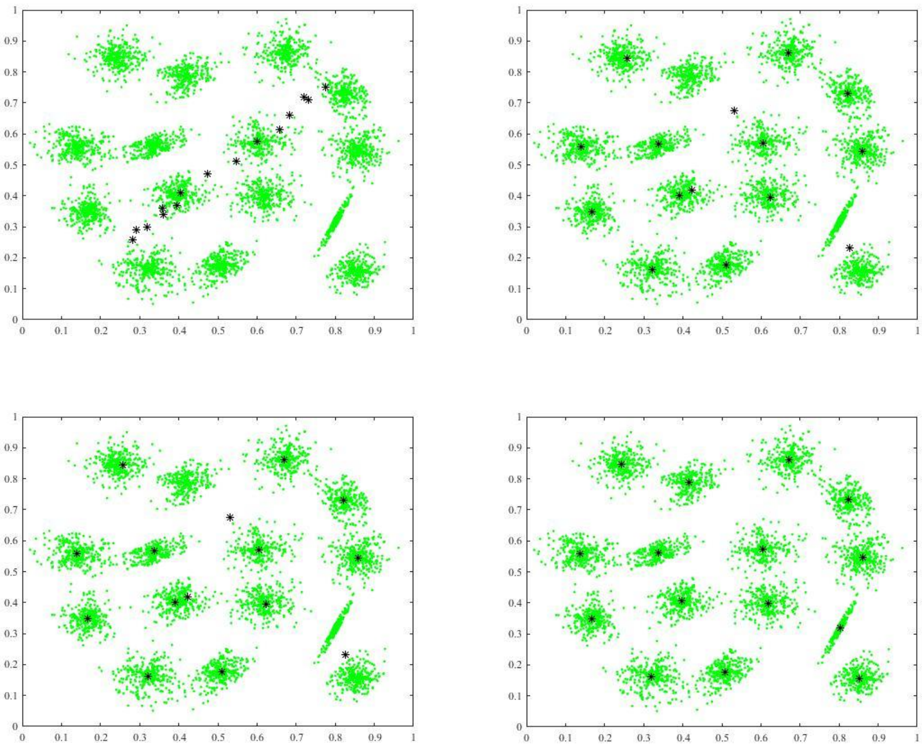


Figure 2.3: Progression of EIC FCM applied for S1 dataset. C = 15, r = 0.5, ε = 10^-5. Pictures of iterations 1, 4, 13, and 75

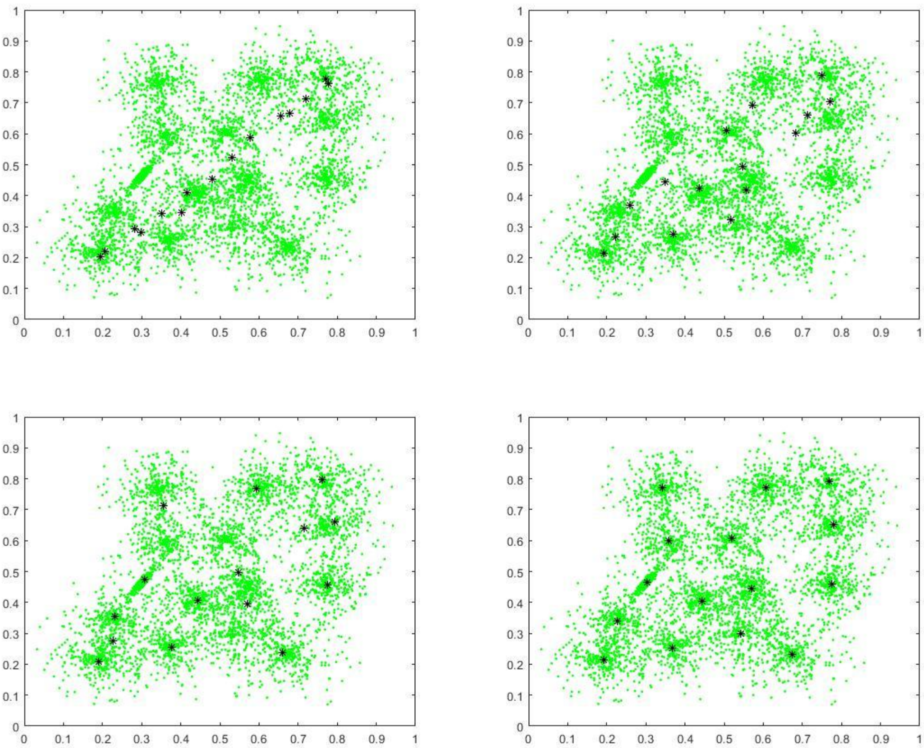


Figure 2.4: Progression of EIC FCM applied for S3 dataset. C = 15, r = 0.5, ε = 10^-5. Picture of iterations 1, 4, 13, and 137

2.1.3 EICCA ALGORITHM

The objective function for EICCA is the following:

$$\min J_{EICCA}(U,V) = min \sum_{i=1}^{C} \sum_{j=1}^{N} u_{ij} \left\| x_j - v_i \right\|^2 - \alpha \sum_{i=1}^{C} \sum_{j=1}^{N} u_{ij} \quad (2.5)$$

With constraints:

$$\sum_{i=1}^{C} (u_{ij})^r = 1,\ 0 < r < 1,\ \forall j;\ u_{ij} \in [0,1], \forall\, i,j \quad (2.6)$$

It is shown in [14] that from *eq.* (2.5) and (2.6) it is possible to obtain the update equations for cluster centres and membership:

$$u_{ij} = \frac{(\left\| x_j - v_i \right\|^2 - \alpha)^{\frac{1}{r-1}}}{(\sum_{k=1}^{C} (\left\| x_j - v_k \right\|^2 - \alpha)^{\frac{r}{r-1}})^{\frac{1}{r}}},\ i = 1, \dots, C \quad (2.7)$$

$$v_i = \frac{\sum_{j=1}^{N} u_{ij} x_j}{\sum_{j=1}^{N} u_{ij}}, i = 1, \dots, C \quad (2.8)$$

The value of α in eq. (2.7), if too high, may give rise to non-positive u_{ij}. To avoid that α is updated dynamically using eq. (2.9), where η_0 and τ are parameters determined by user.

$$\alpha(k) = \eta_0 \exp\left(-\frac{k}{\tau}\right) \frac{\sum_{i=1}^{C} \sum_{j=1}^{N} u_{ij} \left\| x_j - v_i \right\|^2}{\sum_{i=1}^{C} \sum_{j=1}^{N} u_{ij}} \quad (2.9)$$

N_i is defined as the cardinality of the i-th cluster. It is computed as follows:

$$N_i = \sum_{j=1}^{N} u_{ij} \quad (2.10)$$

For each iteration In EICCA algorithm, clusters with cardinality too low are discarded, in order to converge from C_{max} (number of initial clusters) to the optimal number of clusters for the analysed dataset.

- ALGORITHM DESCRIPTION

As indicated in [14], these are the 8 steps of the algorithm.

1) Given the Dataset *X*, the over specified number of clusters C_{max}, iterative threshold ε and competitive threshold ε_1, initialize $U^{(0)}$ and set the iteration number *k* to be 0.

2) Compute $V^{(0)}$ using eq. (2.8); Compute the cardinality N_i of every current cluster *i*, using eq. (2.10).

3) Update α_k using eq. (2.9).

4) Update $U^{(k+1)}$ using eq. (2.7).

5) Update N_i using eq. (2.10), if $N_i < \varepsilon_1$, then give up the cluster and its centre v_i.

6) Update the number C of clusters.

7) Update $V^{(k+1)}$ using eq. (2.8).

8) If $||V^{(k+1)} - V^{(k)}|| < \varepsilon$, then output U, V and C and exit, otherwise $k = k+1$ and go back to step 3

The flowchart of the algorithm is presented in *Figure 2.7.*

The initialization of the Fuzzy Partition matrix (U) is performed running EIC – FCM algorithm for a specified number of iterations k_0 with number of clusters equal to C_{max}.

- APPLICATION ON BENCHMARK DATASETS

Figures 5 and 6 show the behaviour of the algorithm with some 2D benchmark datasets for clustering. As before, datasets are available in [18].

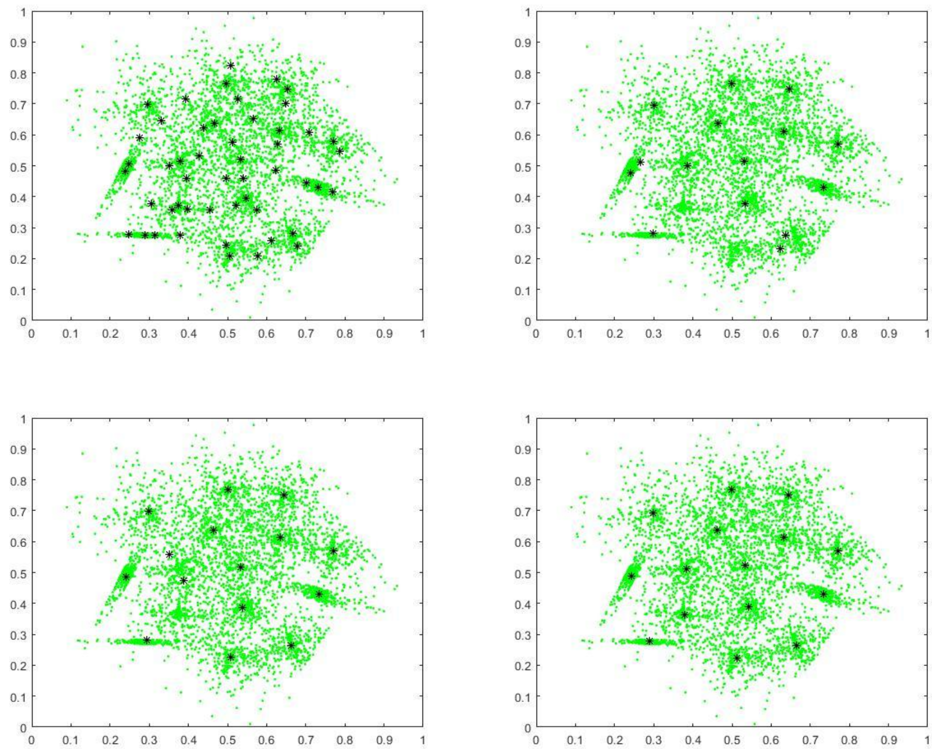


Figure 2.5: Progression of EICCA applied for S4 dataset. Cmax = 50, r = 0.5, ε = 10^-5. Pictures of iterations 0, 4, 22, and 44

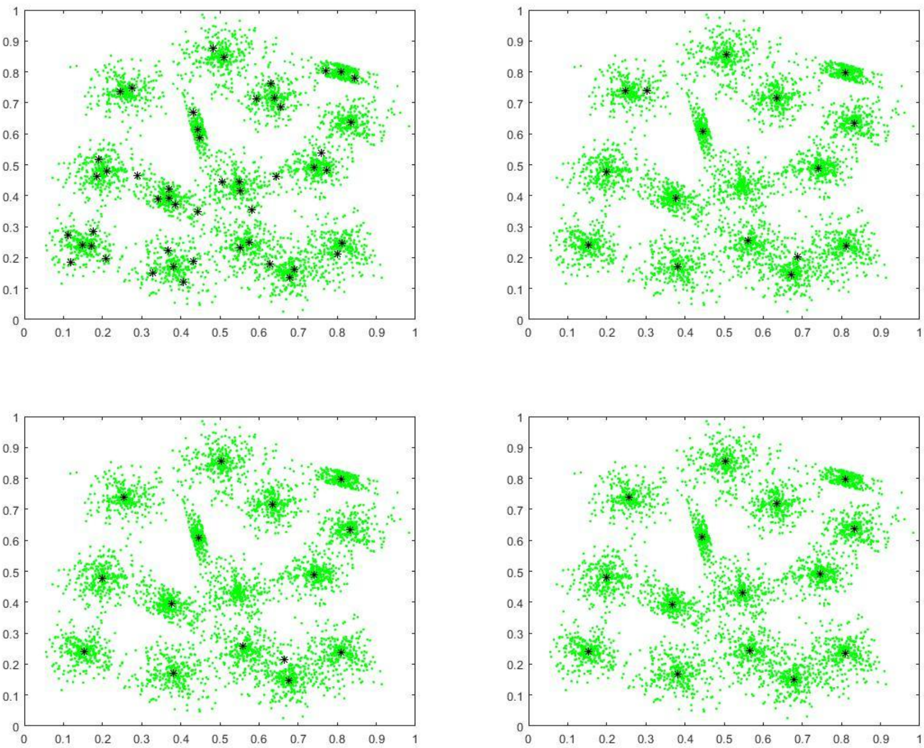


Figure 2.6: Progression of EICCA applied for S2 dataset. Cmax = 50, r = 0.5, ε = 10^-5. Pictures of iterations 0, 11, 21, and 44

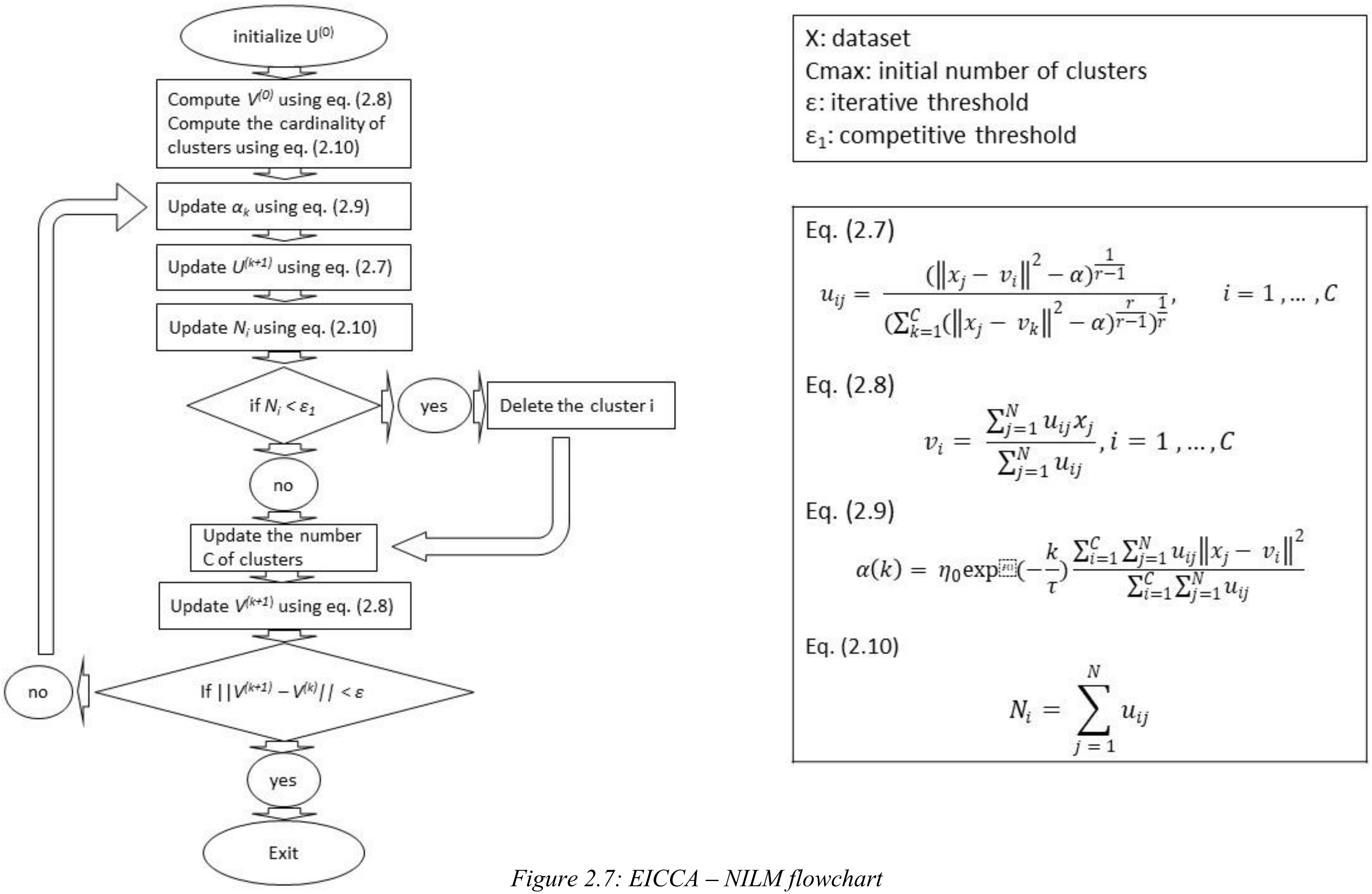


Figure 2.7: EICCA – NILM flowchart

2.1.4 EICCA BASED NILM ALGORITHM FRAMEWORK

As stated at the beginning of the chapter, the main element that make EICCA algorithm suitable for NILM problems is the fact that number of clusters has not to be specified. This is important because number of appliances and consequently of their power states is usually unknown. If we consider a practical and/or commercial application, it is clear that the non-necessity to count all the devices present in one household is a clear advantage for installation time and cost. Authors in [16] and [17] have made use of the algorithm described in 2.1.3, adapted to this use with a modification now explained.

In step 5 of EICCA algorithm the cardinality N_i of each cluster is compared to the value ε_I, and only clusters with cardinality greater than this value are kept for the next iteration. The value of ε_I is decided at the beginning and is constant along all the iterations. In [16] the concept of aggregation threshold percentage is introduced. For every Iteration, between steps 5 and 6, the value of ε_I is updated as shown in the following equation.

$$\varepsilon_1 = avg(N) * aggregation\ threshold\ percentage \qquad (2.11)$$

It means that for every iteration the average of the cardinality of all the clusters is computed and all the clusters with cardinality lower than a percentage of this average are discarded. This allow the algorithm to adapt better the number of clusters, considering information about the actual state of them.

As stated in [15], goals of the algorithm are the following:

1. Being able to decompose individual energy usage using only active power feature
2. Group similar features together and use these groups to build appliance models
3. Use the appliance models to recognize real appliances
4. Reconstruct the total load at time *t* by matching it with a set of appliance models

Some assumptions are made:

1. The total load is sampled at a low rate
2. Only one device will change within the given time window t.

- ALGORITHM OVERVIEW

The algorithm can be synthetized in the following parts, as shown in *Figure 2.8*.

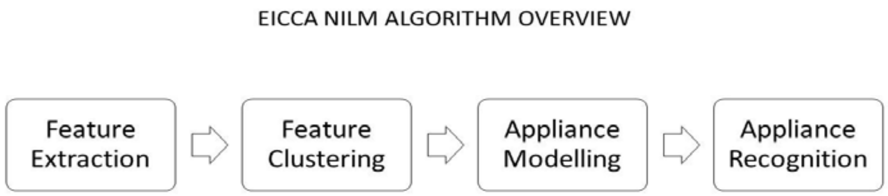


Figure 2.8: EICCA NILM Algorithm Overview

FEATURE ESTRACTION

For the first part of the algorithm aggregated data are analysed and features are extracted. In this case active power is the only feature used. Most of NILM algorithms proposed in the last years rely on this feature only, given its availability from conventional smart meters. Changes of active power between 1s intervals are recorded if they are greater in absolute value than a *significance threshold*, as reported in Eq. (2.12)

$$abs(\Delta P_{t_i}) = abs\left(P_{t_{i+1}} - P_{t_i}\right) > significance\ threshold \qquad (2.12)$$

Authors of the method in [16] propose to save only the absolute value of the transition, getting rid of the sign.

FEATURE CLUSTERING

Features obtained in the previous point are grouped together in sets composed by events similar in magnitude, using EICCA clustering algorithm (2.1.3). Clustering can be done on both positive and negative transitions or on the absolute values of them. The centres of clusters emerged from the clustering technique are called power states.

APPLIANCE MODELLING

In chapter 1 appliances have been categorized in different types. Authors of the method have mainly used Type 1 appliances, with some references to Type 2. For the purposes of this work only Type 1 models are considered. Models are generated as follows.

If Feature Clustering has been performed on absolute values only, the ON/OFF model for a power state is created if in the aggregate signal events that match this power state in magnitude are both positive and negative.

If Power states are both positive and negative, the procedure described in [15] is this:

1. Clusters are divided in positive (C_p) and negative (C_n) in sign.
2. For each positive cluster search the set of negative clusters for a cluster that is similar in magnitude.
3. If there's a match, the two clusters are taken away from the 2 sets.

The process can be summarized in following equation:

$$M_i = \left\{C_{P_i}, \min\left(\left\|C_{P_i} - \left|C_{N_j}\right|\right\|^2\right)\right\},\ i = 1, \dots, C_{P_n};\ ,\ j = 1, \dots, C_{N_n} \qquad (2.13)$$

Since only Type 1 models are considered, all the unmatched clusters are then discarded.

APPLIANCE RECOGNITION

Appliance recognition is performed in this way, according to [17].

Events extracted from aggregate signal (ΔP_{ti}) are compared to power states of appliance models. If an event match one power state within a 5% error plus a further 5W tolerance, the event is labelled as a member of that power state.

2.1.5 EXPERIMENTAL VALIDATION IN LITERATURE

According to our knowledge, at the moment of writing experimental validation on this method has been limited. The bigger limit found in literature is that the method hasn't been validated yet in relation to real appliances. The main goal of a NILM framework is to provide insight of single appliance usage. Lack of appliance - based metrics limits the significance of experimental validations. Evaluation in literature is now briefly presented.

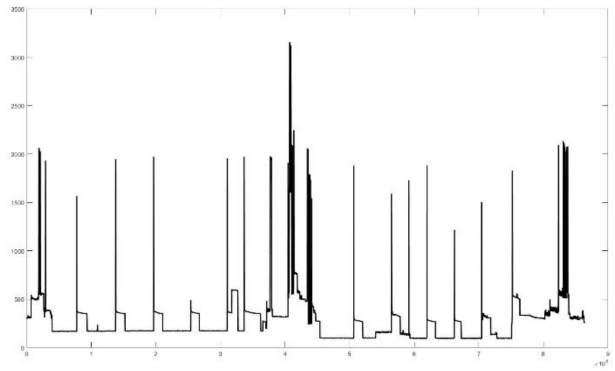


Figure 2.9: Example of 24 hours power consumption profile. 25/04/2011, house 1 REDD dataset

ALGORITHM PERFORMANCES

The first effort has been focused to validate the algorithm capability to detect events and to extract features from them. Authors in [17] used houses 1-6 of REDD dataset for periods of 1 and 3 days. Once events have been clustered, the clusters have been used to create appliance Type 1 models. They used features in absolute value form. Then common NILM metrics have been computed. Usually metrics are in relation with specific appliances. A detailed description of metrics can be found in section 1.6. In *Figure 2.10* results achieved in [17] are reported.

TABLE I
NILM METRICS FOR SINGLE DAY ENERGY USAGE

REDD House	Acc. (%)	P (%)	R (%)	f_1 (%)	TECA (%)
1	94.71	93.71	95.58	94.63	81.40
2	98.36	98.58	97.80	98.18	91.73
3	97.90	97.74	97.92	97.83	91.95
4	98.98	99.02	98.86	98.94	93.65
5	99.20	99.36	98.73	99.04	98.84
6	97.93	97.69	98.14	97.91	83.25

TABLE II
NILM METRICS FOR THREE DAY ENERGY USAGE

REDD House	Acc. (%)	P (%)	R (%)	f_1 (%)	TECA (%)
1	95.96	95.67	96.17	95.92	63.61
2	98.90	98.97	98.60	98.79	93.89
3	88.99	85.39	90.44	87.84	77.16
4	97.80	97.32	98.22	97.77	90.82
5	95.04	94.46	94.59	94.52	75.13
6	97.37	97.01	97.61	97.31	83.10

TABLE IV
NILM METRICS FOR SUBSEQUENT DAY ENERGY USAGE

REDD House	Acc. (%)	P (%)	R (%)	f_1 (%)	TECA (%)
2	96.20	96.79	94.71	95.74	87.95

Figure 2.10: NILM metrics on houses 1 – 6 in REDD dataset.
Taken from [17]

These results have been obtained by comparison of inferred events with models emerged from Appliance Modelling (2.1.4). The goal of section 2.2 will be to determine how models are representative of real appliances.

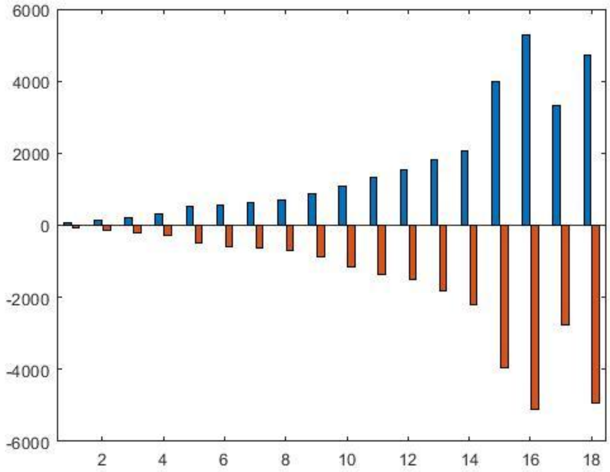


Figure 2.11: Type 1 appliance models (W) generated for house 1 REDD dataset, 28-30/04/2019.

USE OF MODELS FOR FUTURE ENERGY USAGE RECOGNITION

For house 2 in REDD dataset models inferred from 1 day of consumption have been used to disaggregate the load in the following day. Results are shown on the third table in *Figure 2.10.* This concept will be expanded in section 2.2, using significantly longer time periods.

2.2 EXPERIMENTS AND RESULTS

2.2.1 EXPERIMENT GOALS

Experiments that have been run using EICCA NILM are focused to investigate these topics:

1. Ability of EICCA algorithm to converge to clusters that are representative of real appliance power states, focusing on high power appliances.

2. Evaluation of NILM metrics on some appliances.

3. Behaviour of NILM algorithm with different time windows for training and evaluation. It means that feature extraction, feature clustering and appliance modelling (2.1.4) is performed on a set of days, and evaluation on another one, longer and consequent to the first.

4. Introduction of reactive power feature in the algorithm.

5. From a general point of view consider if it is realistic for NILM methods to rely only on active (and eventually reactive) power transitions.

2.2.2 DATA DESCRIPTION

Experiments have been performed on data from REDD and UKDALE datasets. For both sets power measures are provided with 1s resolution.

REDD has been chosen because it has been widely used in literature, for a lot of active power – based methods. The main drawback of this dataset is the limited time length of its measures. For these experiments houses 1, 2, and 3 have been chosen. The choice is driven by the presence of some very high power appliances in these houses.

UKDALE dataset has been chosen because for houses 1, 2 and 5 reactive power is provided in addition to active power. Moreover the time span of measures is wider than in REDD. This allow to simulate more realistic situation, with longer training and evaluation period of few months.

Table 2.1 details data used for experiments. In Appendix B all the dates of days used are reported.

DATASET	HOUSE	TRAINING DAYS	EVALUATION DAYS	FEATURES
REDD	1	9	///	P
REDD	2	9	///	P
REDD	3	8	///	P
UKDALE	1	30	170	P/Q
UKDALE	2	20	80	P/Q
UKDALE	5	30	50	P/Q

Table 2.1: Days used for experiments

2.2.3 FEATURE EXTRACTION AND CLUSTERING

EVENT DETECTION/FEATURE EXTRACTION

Feature extraction method used for experiments is different than the simplistic one described in 2.1.4. Here we adopt the event detection method from [19]. This method is described in 3.1.2 (points B, C and D). In this section only the reasons of this choice are detailed.

1. Power edges longer than one second
   Experience shows that several events in power profile take more than 1 second to occur. Let suppose that the event of interest has magnitude of 1500 W. If it take 2 seconds, the event could be split in two edges, for example 500 and 1000 W. This is problematic because we want the clustering phase to produce clusters from groups of similar events connected to real appliances. So a more stable method is preferable. The method in 3.1.2 overcomes this problem being steady state based. *Figure 2.13* summarizes this point.

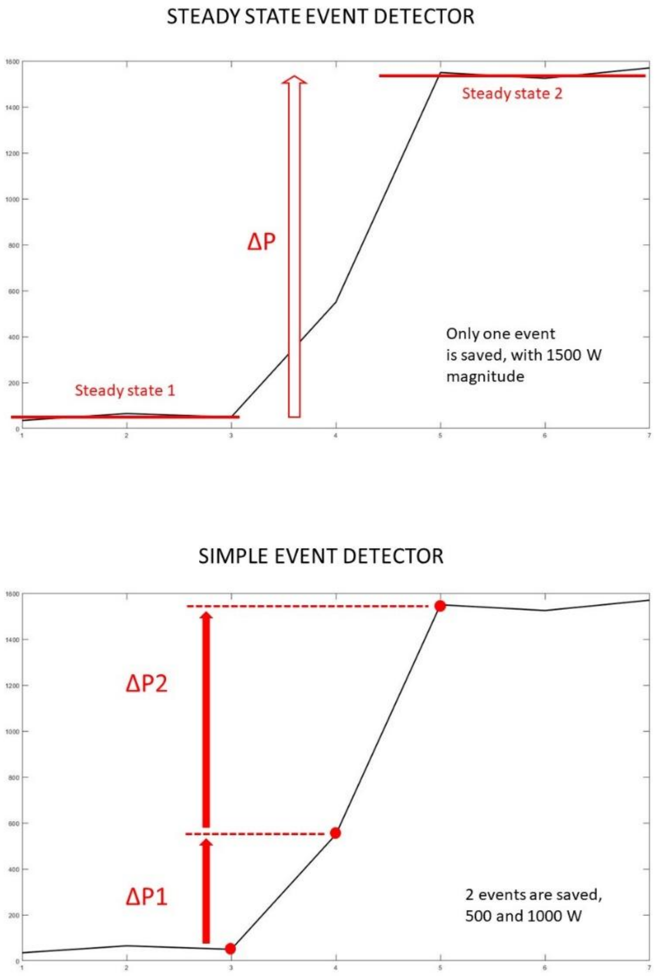


Figure 2.13: Graphic comparison of steady state vs simple event detector

2. Filtering
   Adopted event detection method smooths the power profile, in order to eliminate noisy fluctuations.

FEATURE CLUSTERING

Clustering phase is performed on extracted features as follows

1. An over specified number of cluster is selected (i.e. 50). The set of features is normalized in order to fit in the interval [-1, 1].

2. EIC-FCM algorithm (2.1.2) is performed on the set of features, for a specified number of iterations (i.e. 30). It outputs the fuzzy partition matrix $U^{(0)}$ necessary for the next point.

3. EICCA algorithm (2.1.3) is performed on the set of features. The output will be the "optimal" set of clusters, representing the power states emerged from the algorithm.

2.2.4 CLUSTERING EVALUATION METHOD

In order to understand the capability of the method to converge to clusters that are representative of actual appliances power states, as far as possible events clustered have been assigned to the appliance that generated them. To do that ground truth data related to appliances have been used. Main focus is on high power appliances. A table have been computed for each experiment, with the following fields:

1. Number of Cluster
   Progressive number of clusters.
2. ΔP (W)
   Active Power signature of the cluster.
3. ΔQ (VAR)
   Reactive Power signature of the cluster.
4. Cardinality
   Number of events in the cluster.
5. Appl. 1
   Appliance with the bigger number of events in the cluster.
6. % Appl. 1
   Percentage of the events of appliance 1 over the whole number of events in the cluster.
7. Appl. 2
   Appliance with the second bigger number of events in the cluster.
8. % of Appl. 2
   Percentage of the events of appliance 2 over the whole number of events in the cluster.
9. Appl. 3
   Appliance with the third bigger number of events in the cluster.
10. % of Appl. 3
    Percentage of the events of appliance 3 over the whole number of events in the cluster.

All the tables obtained are collected in appendix C. Usually for each house there is more than one experiment, so the numbers of experiments in this chapter are referred to the numeration presented in that appendix.

2.2.5 CLUSTERING EVALUATION RESULTS

In this section results obtained with experiments are presented. For further details see appendices B and C.

Parameters in the algorithm are numerous so a big variability of results is possible. Values chosen and reported in appendices B and C are the ones that have shown significant results. Parameter *r* (entropy index) is kept constant to 0.5, which is the same value that have been used in literature in all reported experiments.

Following evaluation is done by comparing emerged clusters to appliances detailed in Table C.1. They have been chosen as the most significant for this study. They are mainly appliances with an electric power draw higher than 1 kW. The only exception are cooling appliances (fridges and refrigerators). Even if some appliances are not represented by a type 1 model, for this study they have been considered as they were, taking into account only the biggest of their power state (3.2.3 and *Figure 3.12*).

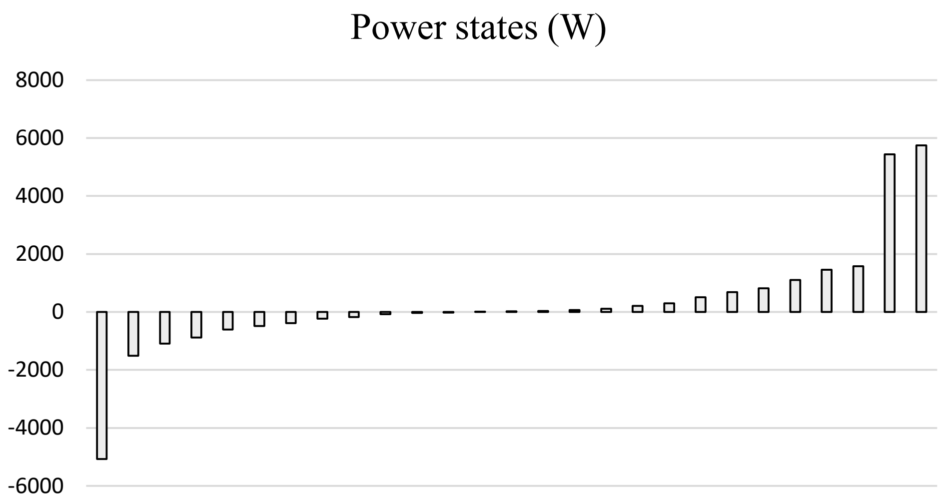


Figure 2.14: Power states emerged from clustering in REDD 1, experiment 4

DATASET/ HOUSE	EXPERIMENT (Appendix C)	NUMBER OF CLUSTERS	CLUSTERS WITH ONE APPLIANCE * (**)	CLUSTER WITH 2 OR MORE APPLIANCES* (**)
REDD/1	4	27	7 (5)	8 (1)
REDD/2	2	17	1 (0)	10 (8)
REDD/3	1	26	14 (10)	6 (1)
UKDALE/1	1, P only	3	///	///
UKDALE/1	2, P and Q	44	5 (3)	4 (0)
UKDALE/2	1, P only	2	///	///
UKDALE/2	2, P only	11	6 (3)	2
UKDALE/2	1, P and Q	45	9 (5)	2
UKDALE/5	1, P and Q	32	6 (3)	2

*a cluster is counted in this column only if the percentage of the appliance's events is relevant on cluster cardinality. Where it is very low (few percentage points) the cluster is not considered.

**Experience shows that fridge is present in several clusters, because it has variability in behaviour. Moreover, being working all the day, it contributes to total number of events in significant percentage. In brackets the number of clusters where fridge is the predominant appliance is reported.

Table 2.2: Relevant clusters emerged from experiments. Compare the table with results in appendix C

FRIDGES CONSIDERATIONS

Table 2.2 shows how many clusters have an important number of events generated by fridges. Let's consider house 1 in REDD dataset as an example. In Table 2.3 clusters of interest for fridge are reported. Full set of clusters is available in *Table C.7.*

Cluster	ΔP (W)	ΔQ (VAR)	Cardinality	Appl. 1	% of Appl. 1	Appl. 2	% of Appl. 2	Appl. 3	% of Appl. 3
6	-175	0.00	194	'Fridge'	69.07	'///'	0.00	'///'	0.00
12	212	0.00	103	'Fridge'	30.10	'///'	0.00	'///'	0.00
16	297	0.00	24	'Fridge'	62.50	'///'	0.00	'///'	0.00
18	815	0.00	52	'Fridge'	78.85	'Dishwasher'	1.92	'///'	0.00
26	685	0.00	44	'Fridge'	84.09	'///'	0.00	'///'	0.00

Table 2.3: Clusters generated mainly by fridge events, in REDD 1, exp. 4 (see Table C.7)

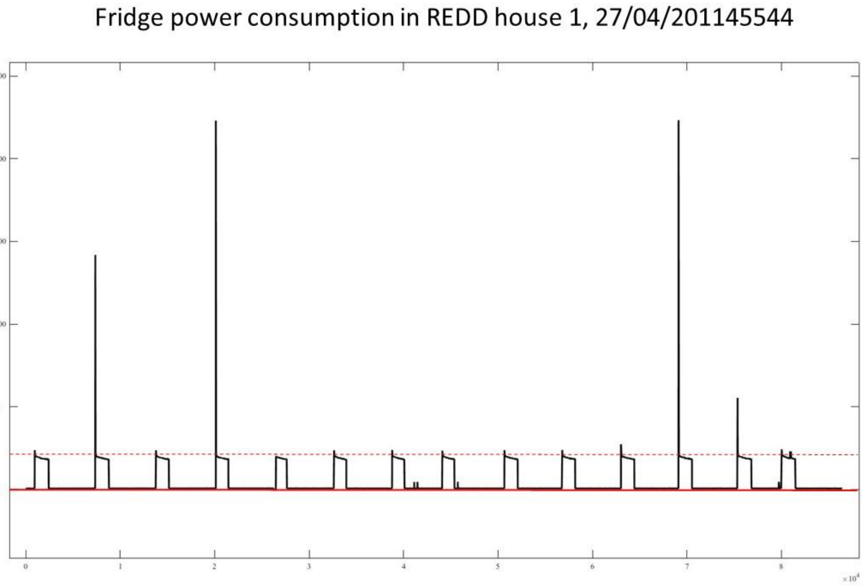


Figure 2.14: some hours of active power consumption of fridge in REDD house 1, 27/04/2011

The power profile of a fridge is shown in Figure 2.14. It is clear that sometimes positive transitions correspond with significant spikes in power. Clusters 16, 18 and 26 are likely produced by these events. Anyway in chapter 3 it is clearly shown that only 2 power states are necessary to characterize well cooling appliances activity. In this case clusters 6 and 12 are probably good candidates. In the situation of an unsupervised method, it is not trivial that clusters 6 and 12 are paired together in one model. In the simplistic case of Type 1 modelling (as described in section 2.1.4) what is going to happen is that cluster 12 will be paired with cluster 13, reported here.

Cluster	ΔP (W)	ΔQ (VAR)	Cardinality	Appl. 1	% of Appl. 1	Appl. 2	% of Appl. 2	Appl. 3	% of Appl. 3
13	-230	0.00	64	'Fridge'	1.56	'///'	0.00	'///'	0.00

It is clear from percentages that cluster 13 represent the fridge worse than 6. Similar considerations can be done for other experiments.

In general the big amount of clusters produced by fridge could create problems in appliance modelling, leading to models that are not connected with real devices.

HIGH POWER APPLIANCES CONSIDERATIONS

In this section we will understand if, for bigger appliances, feature clustering converge to distinct clusters.

Table 2.4 details big appliances taken into account for experiments, and if their signatures have been isolated by the method. Otherwise, if two appliances have been mixed together, that information is provided. For each house the experiment chosen is the one with better results.

DATASET/HOUSE	EXPERIMENT (Appendix C)	H.P. APPLIANCES	APPLIANCE ISOLATED	OTHER INFORMATION
REDD 1	4	Kettle		Washing machine and oven, despite being very powerful, are confused in the same clusters. In all the 1500 W clusters microwave, kettle, dishwasher and bathroom GFI are confused. In general this house is quite difficult, because a lot of appliances have similar power signatures.
		Microwave		
		Dishwasher		
		Washing m.		
		Oven		
		Bathroom GFI		
		Cooker		
REDD 2	2	Microwave		Only stove has been identified. For microwave a positive cluster is emerged, but not the negative one.
		Dishwasher		
		Cooker		
		Stove	✓	
REDD 3	1	Microwave		Microwave has been mixed with bathroom GFI.
		Washing m.	✓	
		Bathroom GFI		
		Electronics	✓	
UKDALE 1	2, P and Q	Kettle	✓	Hairdryer is partially mixed with microwave.
		Microwave	✓	
		Dishwasher		
		Hairdryer		
UKDALE 2	1, P and Q	Kettle	✓	Washing machine is the second appliance in dishwasher clusters.
		Microwave	✓	
		Dishwasher	✓	
		Washing m.		
UKDALE 5	1, P and Q	Kettle	✓	Otherwise dishwasher is predominant in 2 clusters, Its percentage is quite low. (Table C.19)
		Dishwasher		
		Oven	✓	

Table 2.4: Effectiveness of clustering on high power appliances

In general results are not promising. When reactive power is involved performances are better.

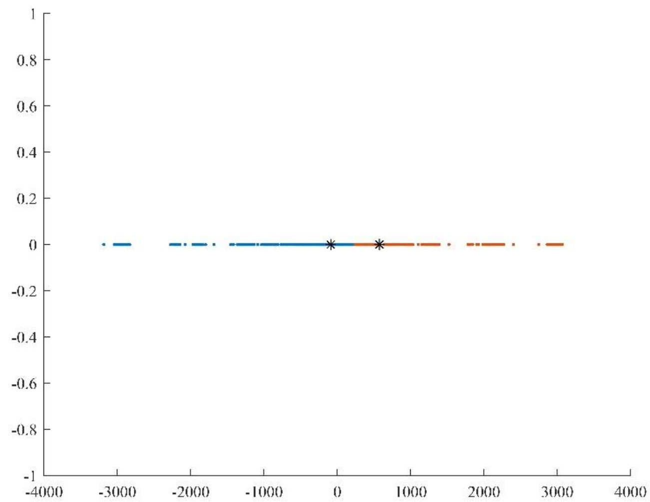


Figure 2.15: Graphic representation of clusters, UKDALE 2, exp. 1, P only

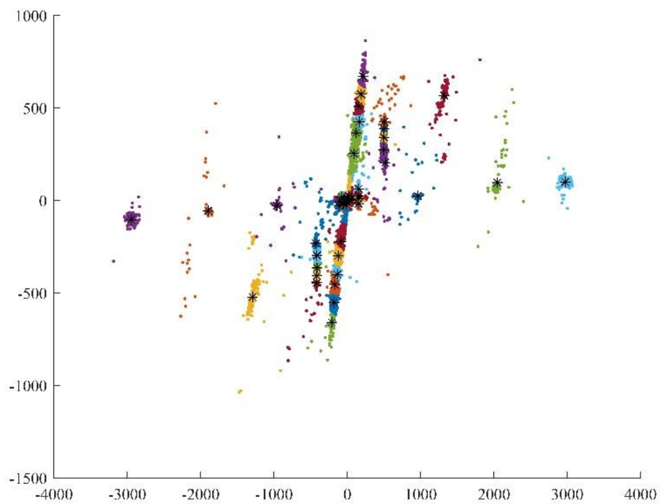


Figure 2.16: Graphic representation of clusters, UKDALE 2, exp. 1, P and Q. Look at Figure C.2 for better definition

This NILM method, as described in literature, is based on active power only. In this work, for UKDALE dataset, we have extended clustering also to reactive power. It is clear by comparing Figure *2.15* and *2.16* that the added feature allows to distinguish better certain groups of events. It is relevant that adding the feature number of clusters changes from 2 to 45. Figure *2.17* shows a detail of that, in the positive part of the P/Q plane. Reactive power in microwave events place its cluster high in the plane, allowing to distinguish it from other events with similar active power.

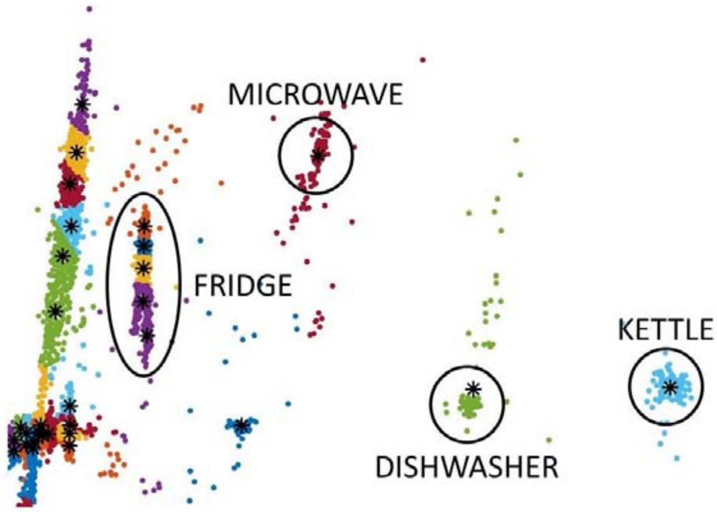


Figure 2.17: detail of Figure 16, with appliance events underlined

A similar consideration can be done with house 1 in UKDALE dataset, experiment 2. *Figure C.1* shows the full set of clustered events in the P/Q plane. In *Figure 2.18* a detail is provided. It can be clearly seen that with reactive power a distinction between microwave and hairdryer is possible.

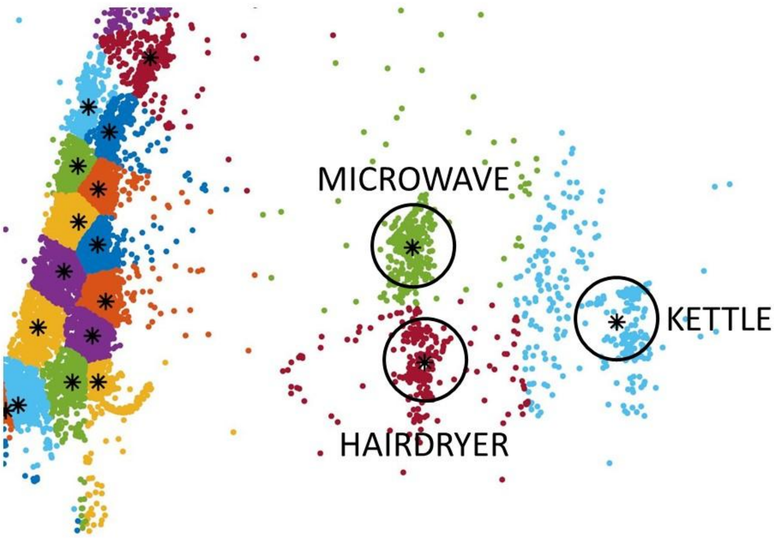


Figure 18: detail of Figure C.1, with appliance events underlined

2.2.6 APPLIANCE MODELLING AND APPLIANCE RECOGNITION RESULTS

Type 1 appliance modelling and appliance recognition (as described in 2.1.4) have been performed with houses 2 and 5 of UKDALE dataset, using P and Q. Results are shown here.

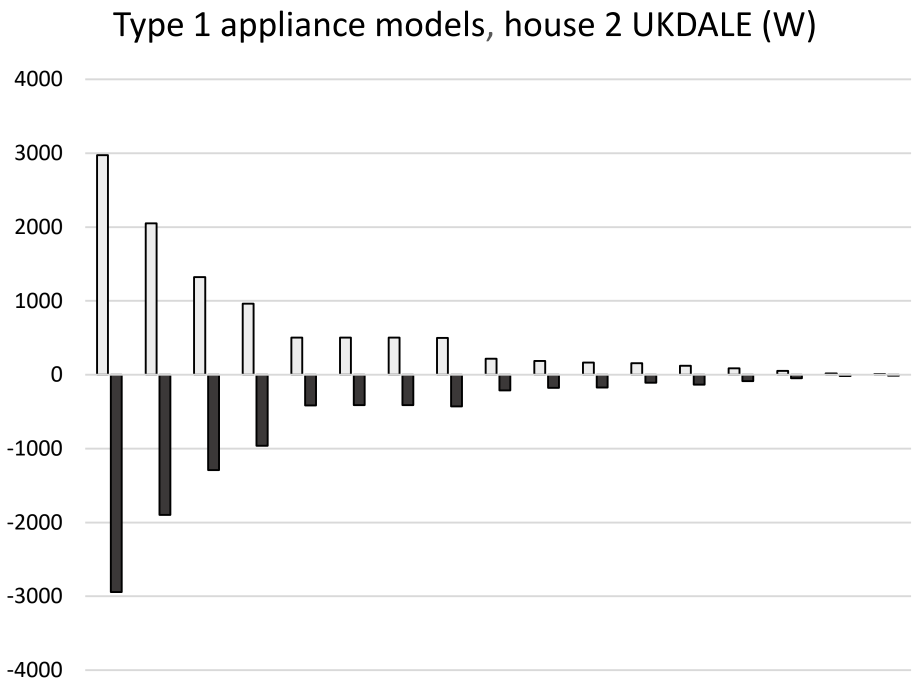


Figure 2.19: Type 1 appliance models, UKDALE 2 exp. 1, P and Q. Only active power feature is shown.

For both households appliance models have been generated using data in training phase, and appliance recognition has been performed in evaluation period. More details are in *Table 2.1* and appendix B. In *Table 2.5* emerged models are reported. Each model is characterized by a positive and a negative transition.

Models for *UKDALE 2 exp. 1, P and Q*				
Num.	+ΔP(W)	+ΔQ(Var)	-ΔP(W)	-ΔQ(Var)
1	156.86	24.39	-107.72	-27.79
2	53.69	6.83	-48.67	-14.14
3	19.40	24.18	-17.84	-21.86
4	9.64	10.43	-11.47	-8.61
5	2048.72	97.47	-1899.00	-54.47
6	1320.84	568.38	-1291.28	-521.60
7	2972.20	100.80	-2943.56	-101.62
8	501.38	341.27	-413.22	-364.46
9	122.07	366.02	-135.36	-398.88
10	88.26	254.24	-87.01	-220.62
11	498.01	274.21	-428.89	-232.13
12	961.68	24.28	-960.17	-26.56
13	215.11	671.83	-210.37	-658.04
14	504.37	426.26	-413.77	-444.09
15	187.62	574.74	-178.41	-551.17
16	162.98	425.85	-171.05	-454.31
17	503.31	386.24	-413.65	-407.00

a)

Models for *UKDALE 5 exp. 1, P and Q*				
Num.	+ΔP(W)	+ΔQ(Var)	-ΔP(W)	-ΔQ(Var)
1	26.40	7.46	-31.95	-8.80
2	69.89	-94.22	-52.95	72.68
3	11.44	3.31	-15.68	-4.86
4	46.61	12.40	-51.18	-13.17
5	138.94	11.01	-155.16	-17.58
6	246.53	21.53	-236.61	-23.25
7	214.88	122.70	-212.19	-118.72
8	385.08	290.52	-268.23	-315.78
9	267.52	187.82	-285.34	-205.30
10	797.29	69.43	-796.92	-68.57
11	1089.31	82.05	-1088.19	-84.58
12	1566.88	109.45	-1559.65	-109.06
13	2090.72	177.18	-2041.68	-176.10
14	2772.34	261.57	-2780.50	-262.73

b)

Table 2.5: Type 1 models generated from cluster centroids (power states) in UKDALE 1 and 5

By manual comparison, taking into account information obtained in *Tables C.19* and *C.17*, these models have been assigned to appliances:

UKDALE 2 exp. 1, P and Q		*UKDALE 5 exp. 1, P and Q*	
Fridge	8,11,14,17	Fridge	8
Kettle	7	Kettle	14
Microwave	6	Oven	13
Dishwasher	5	Dishwasher	12

Table 2.5: Models assigned to appliances

Table 2.6 shows event metrics achieved disaggregating these loads from aggregate signal. In 3.2.5 metrics computation is detailed.

UKDALE 2 exp. 1, P and Q				*UKDALE 5 exp. 1, P and Q*			
Appliance	F - score	Precision	recall	Appliance	F - score	Precision	recall
Fridge	0.621	0.999	0.451	Fridge	0.204	0.938	0.114
Kettle	0.904	0.995	0.828	Kettle	0.570	0.694	0.484
Microwave	0.770	1	0.626	Oven	0.647	0.574	0.742
Dishwasher	0.902	0.956	0.853	Dishwasher	0.006	0.004	0.012

Table 2.6: Disaggregation Results

Results for UKDALE 2 are promising, we can't say the same for UKDALE 5. If models used in UKDALE 5 are compared with table *C.19,* it is clear that percentages of selected appliances are low in clusters used for modelling. That means that they are not really representative of appliances, especially for dishwasher. For example cluster 9 in Table C.19 would have fit dishwasher better, but it hasn't been used in modelling because cluster 11 was closer to cluster 30, so they have been matched. In general these results indicate that a high percentage of events from one appliance in a cluster (in training phase) are indicative of good disaggregation performances in evaluation (test phase). It is difficult to determine a threshold, because size of clusters and so variability inside them vary (large or narrow clusters). *Table 2.7* gives proof of that.

If appliance events are mainly close to the centroid of the cluster, the 5% margin introduced in 2.1.4 can help removing events that are not of interest. It is the case of Dishwasher in UKDALE 2. A good percentage of washing machine events was present in clusters used for the appliance model, but this hasn't deteriorated disaggregation results.

UKDALE 2 exp. 1, P and Q			*UKDALE 5 exp. 1, P and Q*		
Appliance	% on	% off	Appliance	% on	% off
Fridge	~ 90 %	~ 0 %	Fridge	80.71	0.29
Kettle	97.94	98.97	Kettle	34.21	30.77
Microwave	84.26	80.99	Oven	45.68	41.18
Dishwasher	61.54	64.41	Dishwasher	///	4.44

Table 2.7: Percentage of appliance events in clusters used for models

2.2.7 DISCUSSION

Work described here has focused on ability of the method to represent real devices. At the time of writing, as far as we know, this is the first time that something similar is done with this NILM method. 5 points structure in 2.2.1 is here used again to develop final discussion.

1. *Ability of EICCA algorithm to converge to clusters that are representative of real appliance power states, focusing on high power appliances.*

   As described in 2.2.5 and in particular in *Table 2.4,* that doesn't happen for a good number of cases. It means that usually clustering space hasn't got well distinct set of points for every appliance. This is due to 3 facts:
   - i. Overlapping of signatures of different appliances.
   - ii. Variable behaviour of devices (power transitions can be quite different for the same appliance).
   - iii. Number of events of one appliance can condition clustering phase. If they are only few, it is more difficult that a cluster is centred close to these events.

   In general results presented here indicate that convergence of clustering to single appliance power states is a hypothesis too optimistic.

2. *Evaluation of NILM metrics on some appliances.*

   This has been done in 2.2.6, especially in *Table 2.6.* When a cluster is representative of an appliance in good percentage results are good. In the majority of cases that doesn't happen. Many models are not representation of real devices.

3. *Behaviour of NILM algorithm with different time windows for training and evaluation. It means that feature extraction, feature clustering and appliance modelling (2.1.4) is performed on a set of days, and evaluation on another one, longer and consequent to the first.*

   In 2.2.6 this has been done. By comparison of Tables *2.6* and *2.7* it is clear that good clustering in training involve good performances in evaluation.

4. *Introduction of reactive power feature in the algorithm.*

   Use of reactive power is promising for this application. In fact experiments with only active power achieved poor results. *Table 2.4* shows that in REDD dataset, where reactive power is not provided, significant appliances haven't been recognized, with few exceptions.

5. *From a general point of view consider if it is realistic for NILM methods to rely only on active (and eventually reactive) power transitions.*

   This work has clearly shown that active power is not sufficient for EICCA - NILM to perform well. The introduction of Reactive power has been crucial to achieve results in 2.2.6. On the other hand it seems that more features have to be included in this method to increase performances. Comparison between figures *C.2* and *C.3* clarify this point. Where for UKDALE 2 some groups of events are very separated and visible, in UKDALE 5 they are more confused without distinct sets.
   This NILM method in itself needs distinct sets of events to produce significant clustering. Experiments show that in real applications that doesn't happen. This approach doesn't seem exploitable for practical use. More information is necessary in order to distinguish appliances from aggregate power consumption. Variability of appliances and behaviours between different households is a problem in order to develop a method that can show consistency in a wide, commercial use. More significant contribution to this point is given in 3.3.8.

3 SUPERVISED NILM WORK

Chapter 3 is organized as follows. In section 3.1 Hart and Weiss algorithms are theoretically presented. In 3.2 experiment goals are defined, then data used are described. Implementation of both algorithms for experiments is then detailed. Finally metrics computation is presented for all the evaluation metrics used in this chapter. Section 3.3 is focused on results and discussion. The 2 algorithms are compared, use of active power and addition of reactive is discussed starting from experiments. Loads analysed are divided in groups and results are presented. Finally time of occurrence is discussed as a possible feature for disaggregation. Appendices related to this chapter are here reported.

- Appendix D: Dates of days used for experiments and relative datasets.
- Appendix E: Event metrics for all experiments.
- Appendix F: Algorithms comparison tables.
- Appendix G: P vs P/Q features.
- Appendix H: Cooling appliances algorithms and features comparisons.
- Appendix I: Mid-high power appliances event results.
- Appendix J: Time patterns bar charts.

3.1 METHODOLOGY

Supervised NILM methods are those methods where information about appliances are provided to algorithms by prior knowledge (i.e. from a database) or by a training phase, where each appliance is monitored on its own in order to extract needed features. In this chapter Hart and Weiss algorithms are used.

3.1.1 HART ALGORITHM

In the middle 80s' George Hart worked to the first Non-Intrusive Appliance Load Monitoring algorithm. His prototype has been the basis for all later studies. The algorithm, as described in publications [22 – 25], was developed for use with both active and reactive power features. Moreover it considers also the two legs of the usual American residential wiring, reported in *Figure 3.1*.

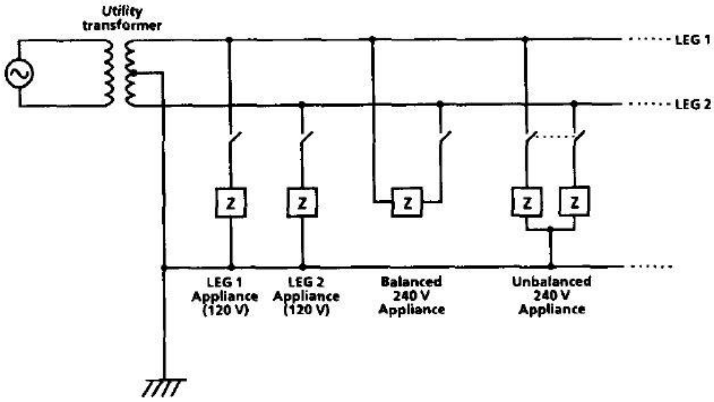


Figure 3.1: Usual American residential wiring, taken from [22]

- ALGORITHM OVERVIEW

Figure 3.2 summarizes building blocks of the algorithm, as presented in [22]. Hart presents two versions of his algorithm, a supervised and an unsupervised one. Acronym NALM stands for Non-Intrusive Appliance Monitoring.

1. MS-NALM (Manual setup)
   Requires an intrusive period to observe appliance signatures, and classify them.

2. AS-NALM (Automatic Setup)
   Automatically classify appliances from *a priori* information and only by analysing aggregate signal.

In this work only MS-NALM has been taken into account, so blocks of the algorithm dedicated to AS-NALM won't be presented deeply.

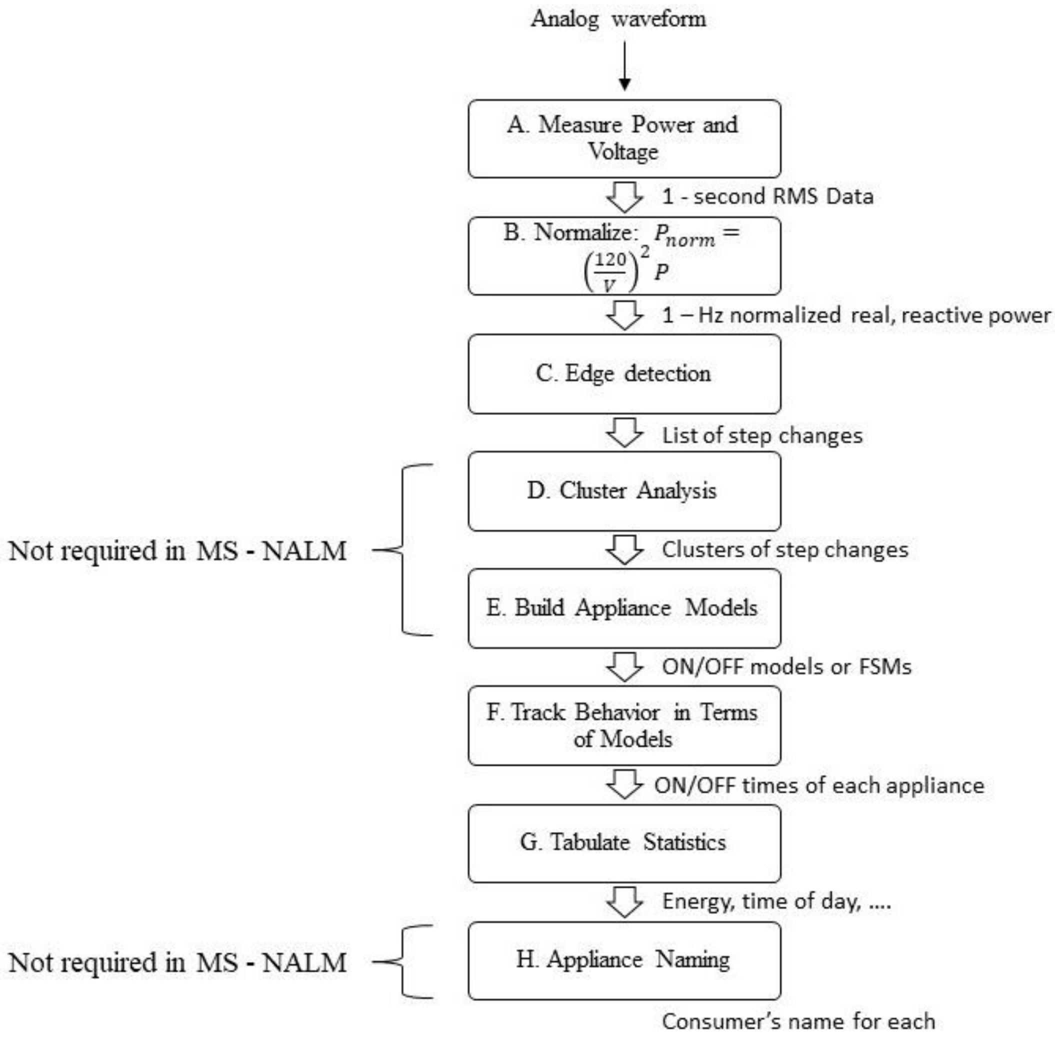


Figure 3.2: Hart algorithm structure

A. MEASURE POWER AND VOLTAGE

Average power and RMS voltage are measured at 1 second resolution, for both legs. Use of 1 second granularity has first been proposed by Hart, and it has been employed widely in literature. The lower resolution is, the more probable it is that two different events are combined together in one. For this work 1 Hz power signal has been used in every case.

B. CALCULATE NORMALIZED POWER

Normalization is performed to provide more consistent set of features to event detection block. Voltage has his own variability and that affects power draw of certain appliances. If voltage provided by distributors has ±10 variability, a linear device will change its current absorption by ±10% too. That produces ± 20% variation in power. This is the reason why the feature used for edge detection is admittance in place of power. Load admittance can be calculated from power and RMS voltage as follows.

$$Y(t) = \frac{P(t)}{V^2(t)}. \quad (3.1)$$

Use of admittance in that form is not appropriate because it is somewhat unfamiliar, especially compared to power. That is the reason why it is converted to *normalized power*.

$$P_{Norm}(t) = 120^2 Y(t) = \left(\frac{120}{V(t)}\right)^2 P(t) \quad (3.2)$$

Value of 120 V suits for the United States. It has to be replaced by the nominal voltage of the country of use, if different.

Exponent *2* in eq. (3.2) makes sense for linear loads. However many loads are not linear. A more general formulation is the following.

$$P_{Norm}(t) = \left(\frac{120}{V(t)}\right)^\beta P(t) \quad (3.3)$$

Exponent β, known as Optimal Normalizing Exponents, is variable for appliances and, in general, between active and reactive power. In [22] it is stated that for an optimal normalization the exponent has to be kept below 2 for active power and above 2 for reactive. The entity of these differences are not clear, so for this algorithm β is kept fixed to 2.

C. EDGE DETECTION

This section of the algorithm finds times and sizes of step changes in power.

Edge detector divides normalized power signal in periods where it is steady and others where it is changing. A steady period is defined as a period longer than a minimum length (3 samples) in which power vary less than 15 W or VAR. Samples in steady periods are averaged, and the difference between steady power levels are saved as events. *Figure 3.3* shows graphically this concept. Only events bigger in magnitude than a specified threshold are saved, because the algorithm is not expected to detect small appliances.

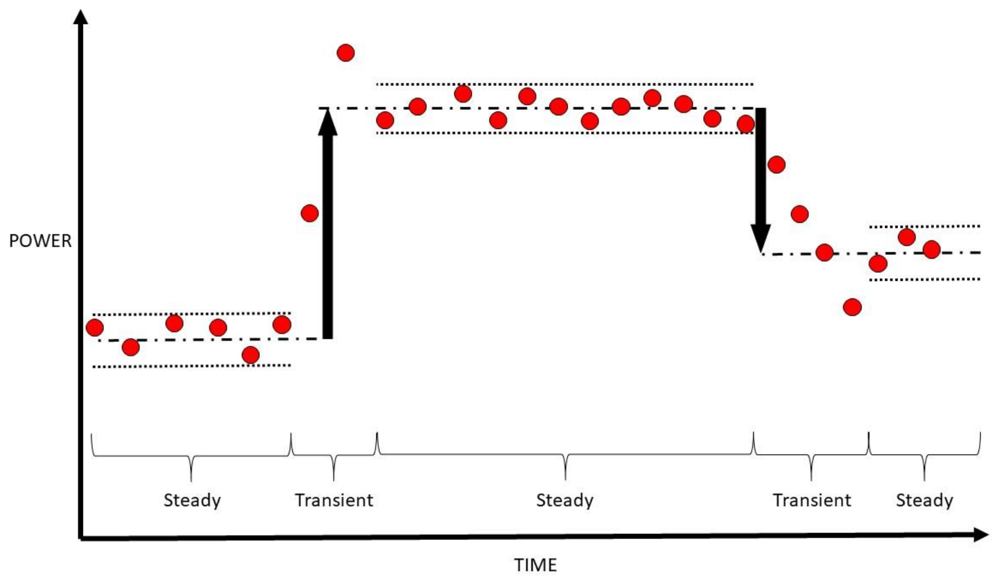


Figure 3.3 Edge detection graphic representation

Edge detection algorithm is now reported, following instructions in [24].

1. Initialize following vectors:

   E: estimate of actual power level during this steady period, based on averaging measurements.

   L: power level of the previous steady period.

   P: power measurement of the previous second.

   T: time of the output event.

   M: current power reading from aggregate data.

   N: counter used to compute average.

   Dimension of these vectors can be:
   i. 4 component if active and reactive power are available for a two phase supply (P1, Q1, P2, Q2).
   ii. 2 components if active and reactive power are available for a single phase supply (P, Q).
   iii. 1 component if only active power is provided for a single phase supply (P).

   Two flags are also necessary:

   A: It is set if the power level is changing in this second

   C: It is set if a change is in progression over a number of seconds

   Select a threshold power level for defining steady state periods. The value usually used is 15 W (or VAR)

   Select a noise level (NL) for defining significant events. Transitions below this value will be ignored. Hart suggests 70W.

Repeat steps 2 to 8 for every second in normalized measurement vector.

2. Get the Measurement M, then compute M - P. If one of the components in M – P exceed in absolute value the steady state threshold, set A, otherwise clear it.

3. If A is set and C is clear, a transition in power has just begun. So perform 3a, 3b and 3c. Otherwise skip them.

   a. Compute E-L. It is the size of the previous transition. If abs(E-L) > NL save E-L and the time stored in T.
   b. Set L to E for the next computation.
   c. Set T to the current time, for the next computation.

4. If A is set, zero the counter N.

5. Update E, the average of all measurements during the steady state period, using:

$$E = (N \times E + M)/(N + 1)$$

6. Add 1 to N, representative of the number of measurement incorporated in E.

7. Set C to the value of A.

8. Set P to the value of M.

9. Go back to Step 2.

Output of edge detector is the set of events (positive and negative transitions) and their time occurrence that will be used for the next phases.

D. CLUSTER ANALYSIS – only for AS-NILM

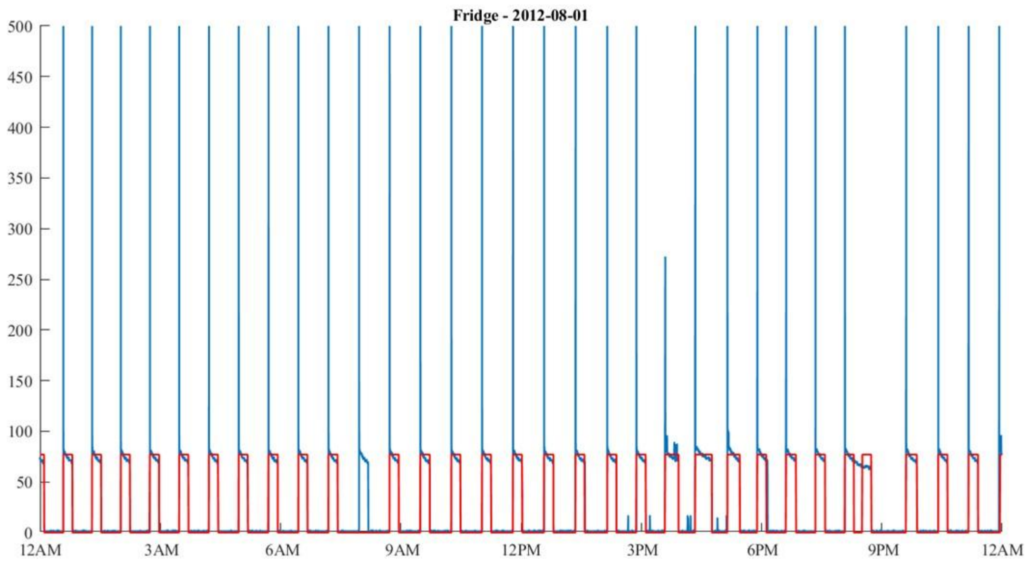
Events emerged from block C, if represented in P/Q plane, define a scatter plot. Figures *C.1*, *C.2* and *C.3* in Appendix C are good representation of that. A clustering algorithm has to be run on these data in order to group together events. Hart points out that a clustering algorithm with fixed number of clusters is not suitable for NILM use. Competitive Agglomeration algorithms, presented in 2.1.3 and A.2 could be used. A clustering method have been developed by Hart in [24].

E. BUILD APPLIANCE MODELS – only for AS-NILM

The algorithm, in its unsupervised form, is supposed to generate automatically ON/OFF and FSM appliance models. For ON/OFF appliances modelling the goal is to find clusters in the plane symmetrical with respect to the origin. Learning of FSM models is discussed in [22].

F. TRACK BEHAVIOUR IN TERMS OF MODELS

At this point events have been assigned to a specific appliance. They are then used to track appliance behaviour. For Type 1 appliances we expect a sequence of on and off events, with sporadic anomalies where some events are missing or wrong events have been assigned to the appliance. Tracking behaviour is necessary to compute energy estimation, which is done by integration of the reconstructed appliance signal.



*Red signal is the tracked behaviour, Blue one is ground truth profile. Reconstructed signal is done on the basis of events assigned by the algorithm to the appliance.

Figure 3.4: reconstruction of fridge signal in house 2, ECO dataset

G. TABULATE STATISTICS

For each appliance useful statistic are provided to users. Some examples are:

a. Energy consumed by appliances, in relation with whole consumption.
b. Number of usages of a certain appliance for each period.
c. Energy broke down by time of the day, consumption weekday vs weekends, differences in consumption in seasons, etc…

H. APPLIANCE NAMING – only for AS-NILM

AS – NALM arrives at this point without knowing names of appliances, steps E and F are done with auto generated models. Each model has to be named in relation to its characteristics referring to a database of general models for many different appliances.

3.1.2 WEISS ALGORITHM

M. Weiss and others proposed this method in 2012 [19].

This algorithm is rooted on Hart original algorithm, but it mainly differs from it for employing distortion power as a feature, smoothening of the consumption curve and a different event detector. The algorithm extracts switching events from the household aggregate consumption and assign them to the best matching feature presented in the database.

Aggregate signal considered by authors in [19] is given at 1 Hz frequency, like for Hart's framework.

CLASSIFICATION OF APPLIANCES

Physical quantities (signatures) that characterize loads are apparent, real, reactive and distortion power. Loads can be classified in resistive, inductive or capacitive. Reactive power is null in resistive loads (for example a light bulb). In inductive and capacitive loads there is a consumption of reactive power. Distortions in current and voltage in the network produces another component in power, called distortion power. Equation of power is the following:

$$|S| = \sqrt{P^2 + {Q_{trans}}^2 + D^2} \tag{3.4}$$

Where S is apparent power, P real power, Q reactive translative power and D distortion power.

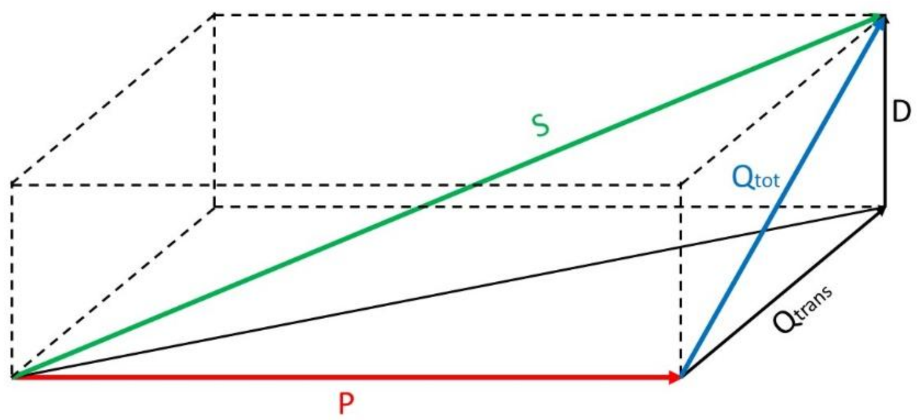


Figure 3.5: Relation between different power quantities

BASIC CONCEPT

The algorithm, as designed by its authors, require as input the total load consumed by household at each time step, divided into power components described in eq. (3.4).

The algorithm starts identifying time points where significant changes between two levels of power consumption occur in power curve. Then it computes the difference of physical quantities between consecutive levels and save these values. Finally values are compared to appliance signatures already present in a signature database. If there is a match, the edge is classified to a specific appliance.

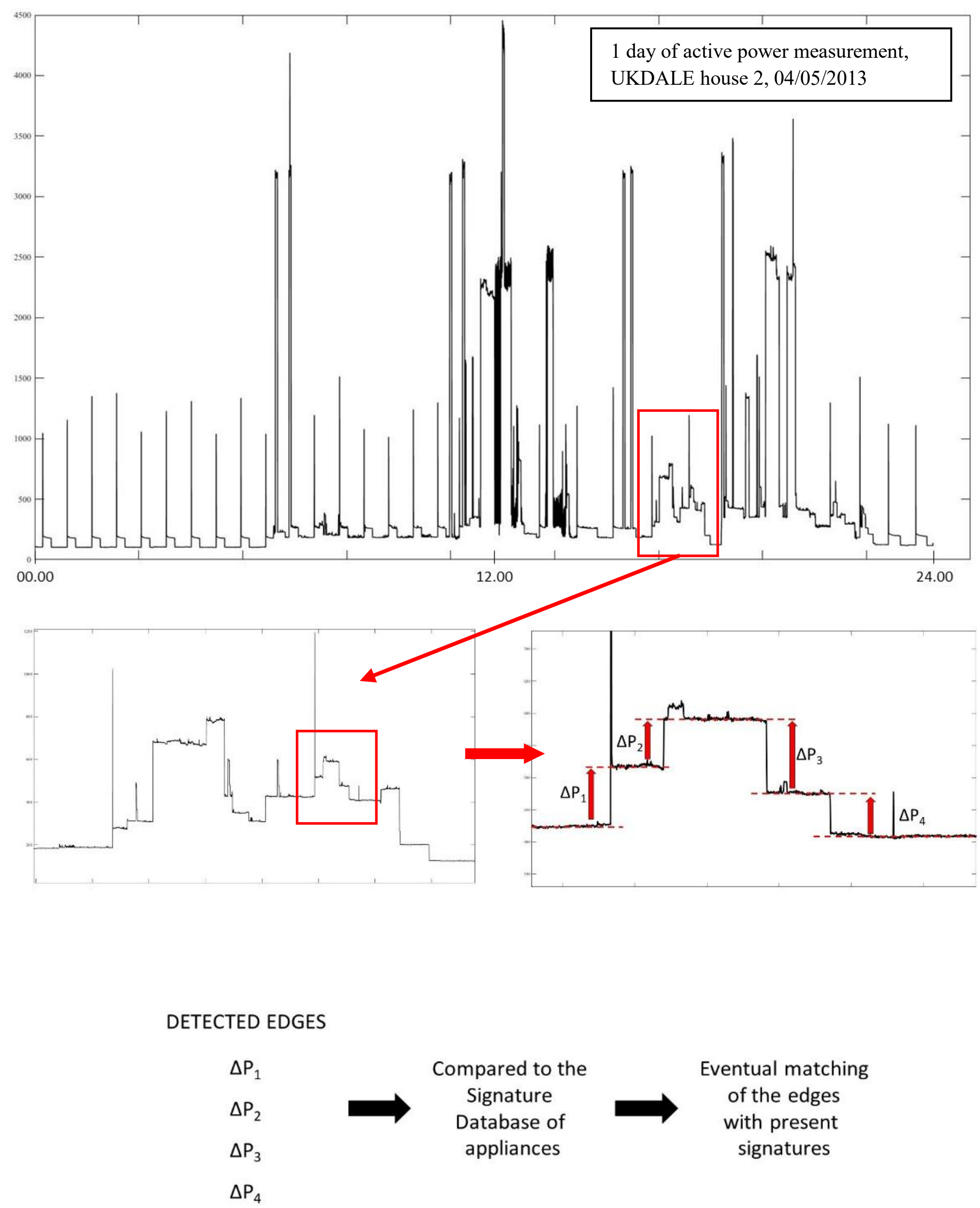


Figure 3.6: Basic concept of Weiss algorithm

- ALGORITHM OVERVIEW

The algorithm is composed by 6 steps [19], now described. *Figure 3.7* summarizes its building blocks.

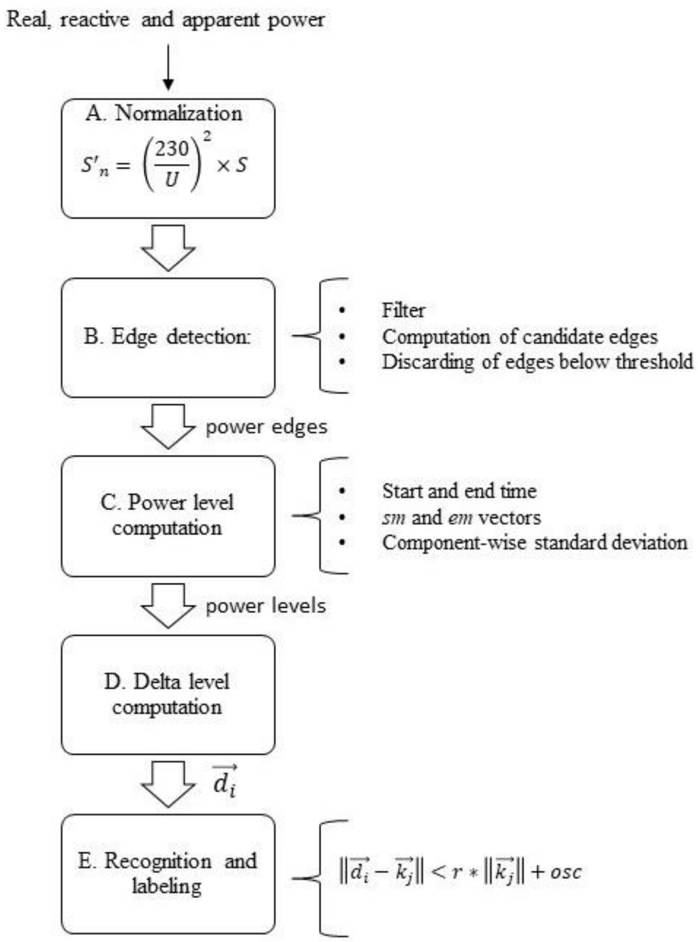


Figure 3.7: Weiss algorithm structure

A. NORMALIZATION

The concept is the same as in 3.1.1, point B. Equation presented in [19] is the following. In this method apparent power is normalized as input vector:

$$S'_n = \left(\frac{230}{U}\right)^2 \times S \tag{3.5}$$

B. EDGE DETECTION

The algorithm computes difference between consecutive values of normalized apparent power S'_n. If the absolute value of one of these transitions is larger than a pre-defined threshold (*f_th*), the value potentially belongs to an edge. Value of *f_th* is essential in order to avoid to save small edges that are meaningless for disaggregation purposes. Authors in [19] suggest to use 2 VA.

Another problem is presence in power profile of peaks related to transient behaviour of appliances that result in a consistent number of spurious events detected. In order to reduce this number a filtering of apparent power signal is performed. The filter chosen is a combination of median and mean filter, both computed in a time window of 5 samples.

A mean filter is a sliding window filter that replaces the central value of the window with the mean of all values in the window.

A median filter is a sliding window filter too. It replaces the central value with the median of all values in the window.

The two filters are applied in series, in the order MEAN – MEDIAN. After this it is possible to extract edges from the signal. *Figure 3.8* summarizes Edge Detection block.

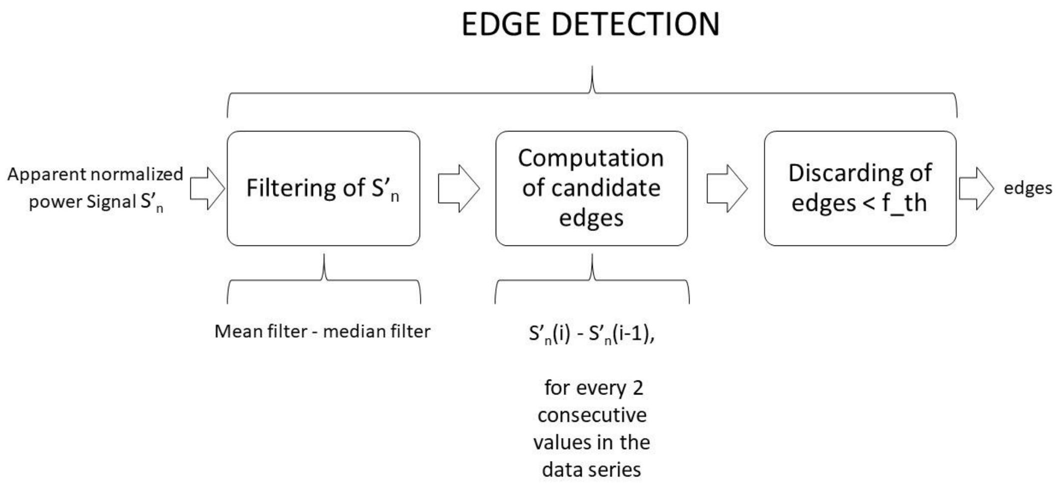


Figure 3.8: Edge detection in Weiss algorithm

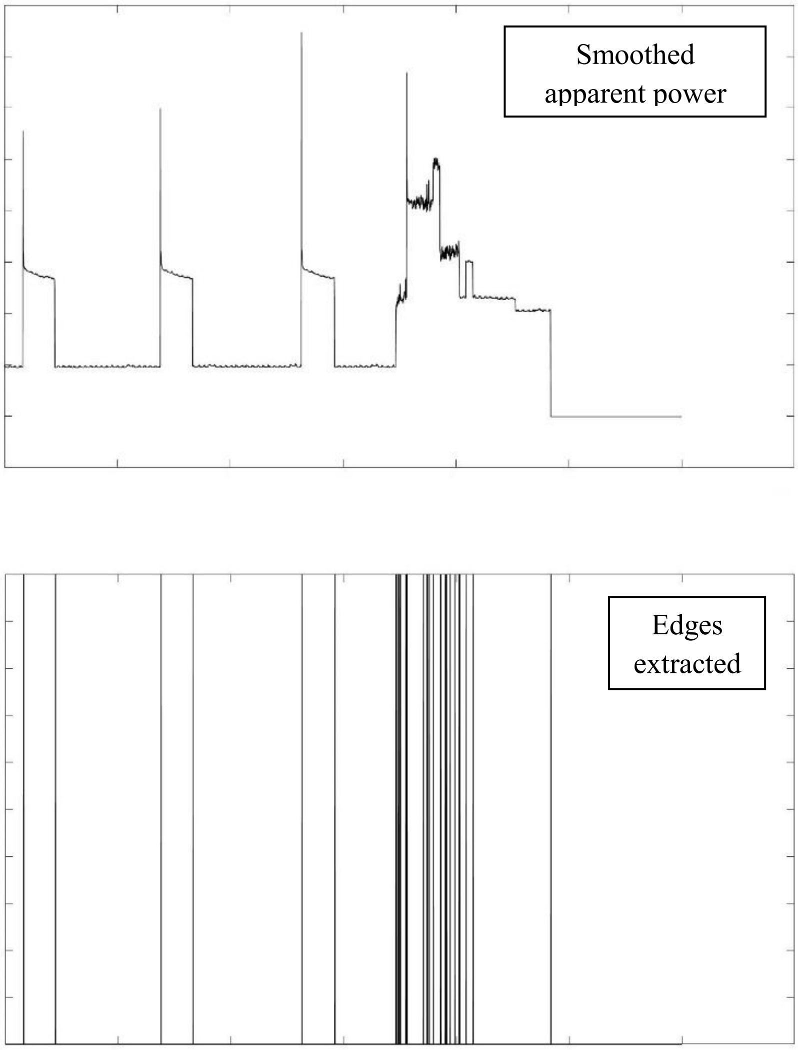


Figure 3.9: Edges extracted from smoothed signal

C. POWER LEVEL COMPUTATION

Now that edges are extracted, the goal is to determine power levels that separate one edge to one another in smoothened signal. For each power level the algorithm determines:

a. START and END time
b. Component – wise mean of REAL, REACTIVE and DISTORTION power for first five measurement at start and last five at the end of each power level. The two vectors are called *sm* (start mean) vector and *em* (end mean) vector.
c. Component-wise standard deviation of all power values in each power level.

D. DELTA LEVEL COMPUTATION

Now that power levels have been computed, consecutive power levels are used to determine variations in power (delta levels) that occur in power signal. To quantify them, the algorithm computes differences of real, reactive and distortive power between two consecutive power levels.

In order to take start up and shut down oscillation of appliances into account, four different difference vectors are computed between consecutive power levels *i* and *i+1*.

$$\overrightarrow{d_{s_\iota,s_{\iota+1}}} = \overrightarrow{sm_\iota} - \overrightarrow{sm_{\iota+1}}, \tag{3.6}$$

$$\overrightarrow{d_{s_\iota,e_{\iota+1}}} = \overrightarrow{sm_\iota} - \overrightarrow{em_{\iota+1}},$$

$$\overrightarrow{d_{e_\iota,s_{\iota+1}}} = \overrightarrow{em_\iota} - \overrightarrow{sm_{\iota+1}},$$

$$\overrightarrow{d_{e_\iota,e_{\iota+1}}} = \overrightarrow{em_\iota} - \overrightarrow{em_{\iota+1}},$$

The four differences are between mean vectors of:

(Start of power state i)	–	(start of power state i+1)
(Start of power state i)	–	(end of power state i+1)
(End of power state i)	–	(start of power state i+1)
(End of power state i)	–	(End of power state i+1)

For each delta level these differences are added to a matrix that will be used in recognition phase.

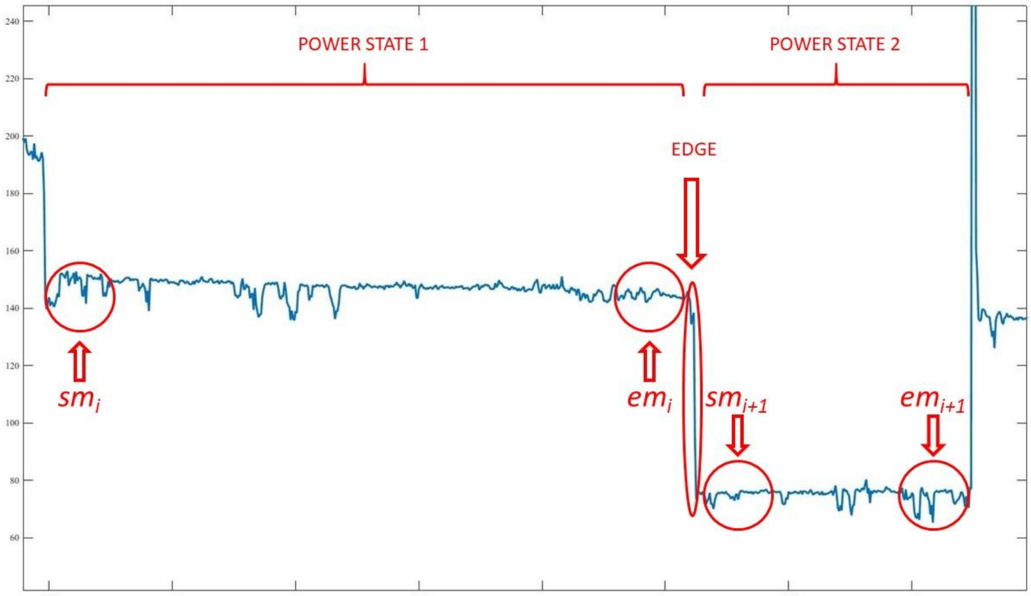


Figure 3.10: Edge, power states and delta level computation

E. RECOGNITION AND LABELING

Known appliance signatures $\overrightarrow{k_j}$ are saved in the signature database as *a priori* information. All the $\overrightarrow{d_i}$ (delta levels) saved from previous phase are compared with signatures, performing a nearest neighbour search in the 2 dimensional ΔP/ΔQ space. Recognition phase is organized as follows:

a. For all the $\overrightarrow{d_i}$, Euclidean distance to each $\overrightarrow{k_j}$ is computed.
b. If the distance is lower than a predefined value (*r*) of the length of $\overrightarrow{k_j}$ plus an oscillation value (*osc*), a potential matching is identified.

$$\left\|\overrightarrow{d_i} - \overrightarrow{k_j}\right\| < r * \left\|\overrightarrow{k_j}\right\| + osc \begin{cases} if\ true, there\ is\ potential\ matching \\ if\ false, there\ is\ no\ matching \end{cases} \tag{3.7}$$

osc term is computed as the length of a two element vector. First element is the maximum of standard deviation in active power between power levels *i* and *i+1*, second one is the same but for reactive power.

$$osc = \begin{pmatrix} \max\left(std(P\ at\ level\ i), std(P\ at\ level\ i+1)\right) \\ \max\left(std(Q\ at\ level\ i), std(Q\ at\ level\ i+1)\right) \end{pmatrix} \tag{3.8}$$

c. At this stage each $\overrightarrow{d_i}$ can be associated with one or more $\overrightarrow{k_j}$ (or not associated at all). If candidates $\overrightarrow{k_j}$ are more than one, the closer one is chosen.
d. Each event is labelled with the name of the appliance related to matched $\overrightarrow{k_j}$.

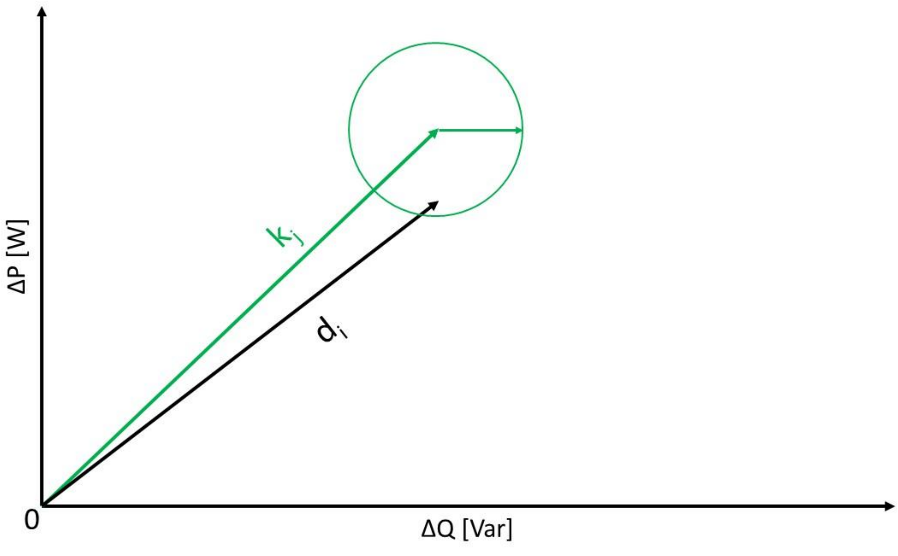


Figure 3.11: Search in the 2 dimensional ΔP/ΔQ space.

3.2 EXPERIMENTS DESCRIPTION

3.2.1 EXPERIMENT GOALS

Experiments with Weiss and Hart algorithms have been performed to investigate these topics:

1. Understand which algorithm performs better, using different datasets and features.
2. Understand which appliances among the analysed ones are well disaggregated by algorithms.
3. Introduction of time of the day feature in load disaggregation.
4. Critical discussion of power disaggregation using P and eventually Q only as features.

3.2.2 DATA DESCRIPTION

Three datasets have been employed for this study. They all share 1 second resolution of aggregate signal.

REDD

Widely used in literature, only active power is provided at 1Hz. As said before, time length of measures is relatively short. Houses 1, 2, 3 and 4 have been used.

UKDALE

Houses 1, 2 and 5 have been chosen. For these houses active, apparent power and voltage are provided at 1 Hz. For plug level data active power only is provided at 1/6 Hz.

ECO

Houses 1, 2, 4 and 5 have been chosen. For these houses active power, current, voltage and power factor are provided at 1 Hz. For plug level data active power only is provided at 1 Hz.

Table 2.1 details time length of experiments. In Appendix D all dates of days used are reported.

DATASET	HOUSE	TRAINING DAYS	EVALUATION DAYS	FEATURES
REDD	1	6	9	P
REDD	2	5	9	P
REDD	3	5	8	P
REDD	4	5	8	P
UKDALE	1	30	170	P/Q
UKDALE	2	20	80	P/Q
UKDALE	5	30	50	P/Q
ECO	1	15	75	P/Q
ECO	2	15	75	P/Q
ECO	4	15	75	P/Q
ECO	5	15	75	P/Q

Table 2.1: Time length of experiments

Data have been organized as indicated in NILM-EVAL toolkit documentation, available in [25]. For each channel of each house, whether it is a smart meter or a plug meter, data have been divided into one file for each day, with exactly 86400 samples. Missing values have been replaced by -1, in order to have the same length for each file. In all the experiments evaluation is performed only when in both smart meter and plug data samples of the quantity needed are not -1.

3.2.3 APPLIANCE DATABASE COMPUTATION

First step in experiments is signature database computation. For each house a set of plug-level power measurements of certain appliances is provided. Some of these channels have been chosen in this work, following 3 principles:

1. Appliances with high power draw and significant for households.
2. Chosen channels have to cover nearly all the time period of aggregate signal analysed. This is relevant in UKDALE where some channels are very limited in time.
3. Appliances representable, at least partially, as Type 1 devices.

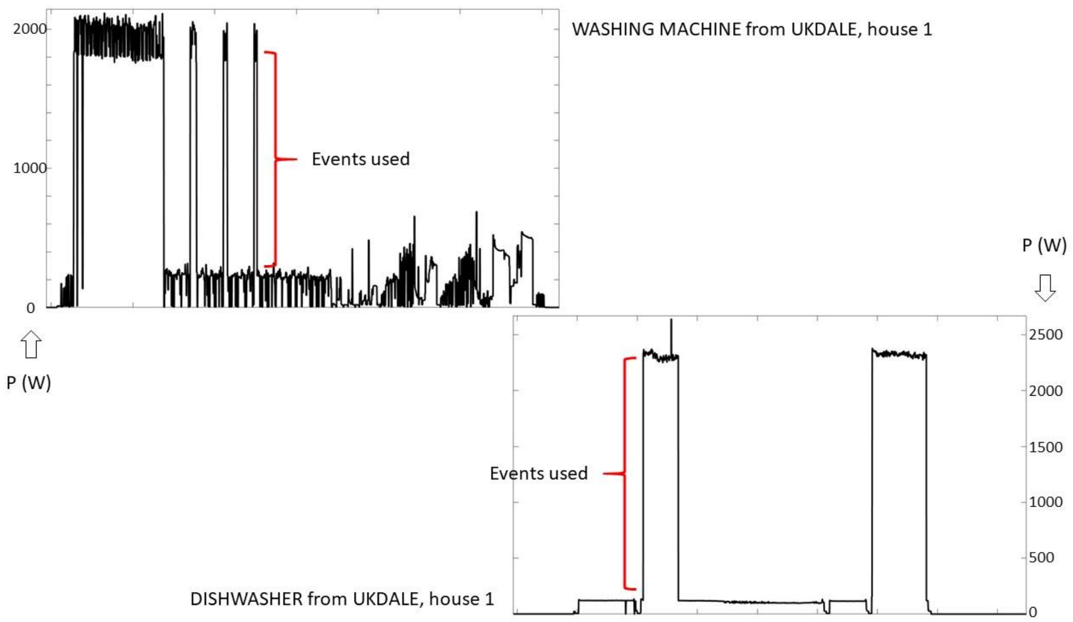


Figure 3.12: Simplification of non-Type 1 Appliances, Washing machine and Dishwasher. Events underlined in red are the only ones used to recognize these loads.

Database computation follows these steps for both algorithms. All measures considered are in training time span

For each appliance:

A. Event detection in plug level data.

   For each algorithm its specific event detector is applied on plug data (3.1.1 step C for Hart and 3.1.2 steps B, C and D for Weiss).

B. Search for correspondent events in aggregate signal.

Events are compared to aggregate signal in order to find out correspondent ones and use them for next part. This is important especially to determine other features as reactive power which is present only in aggregate signal (*Figure 3.13*). In order to facilitate next step, a threshold is specified *a priori* to discard small events that occur in the appliance behaviour but are not relevant for his recognition. This is important for appliances like washing machine, which produces a big variety of events. This can be clearly seen in *Figure 3.12.* All the minor events have to be filtered before next step.

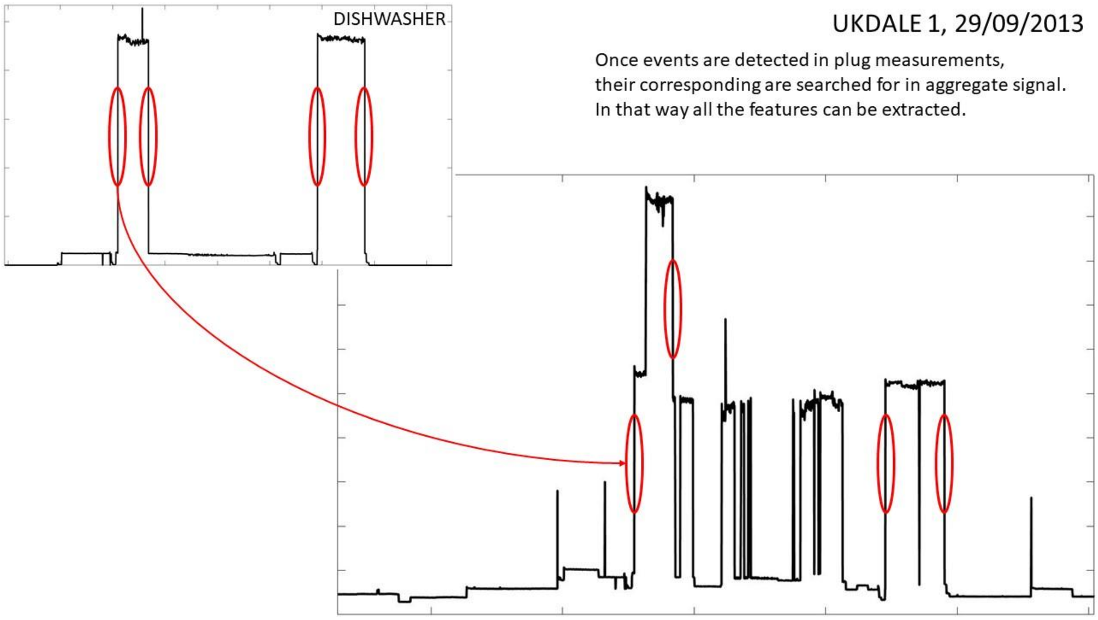


Figure 3.13: Events extraction for database computation

C. Determination of appliance signatures from selected events in aggregate signal.

   Events are divided between positive and negative in active power. A k-means clustering is performed in both groups to produce the centroids that will be saved as signatures of appliances.

3.2.4 ALGORITHM RUN

Algorithms are run in evaluation period, using their specific characteristics. In general output of this section is the list of events detected in aggregate signal and their eventual match in the signature database.

PARAMETER *r*

Parameter *r* has been used in both algorithms to represent how far an event can be from a signature in the database to be assigned to that signature. In addition to *r,* also parameter *osc* is used for Weiss. *Figure 3.11* summarizes the concept.

3.2.5 EVALUATION AND METRICS

Three kinds of evaluation are performed in this work. Main metrics used here have been described in section 1.6.

EVENT EVALUATION

This is in general the main evaluation done in NILM work. Events detected and assigned by algorithms to appliances in the database are compared to plug-level data. Following steps are computed for each appliance that has to be evaluated:

A. The difference vector is computed on plug level data between each P_{i+1} and P_i samples in evaluation period

$$diff_i = (P_{i+1} - P_i), i = 2, \ldots. N, \ \ N\ number\ of\ elements\ in\ P \qquad (3.9)$$

B. Edges vector is computed as follows:

$$edges = abs(diff) > edge\ treshold \qquad (3.10)$$

It is important to set the threshold high enough in order to consider only significant events. This is especially important for Appliances with multiple states that have been simplified as single state. *edges* is a logic vector, with 1 and 0 values.

C. Start and end time vectors have to be computed. A start occurs when in *edges* vector there is a transition from 0 to 1. An end when the transition is from 1 to 0. Indexes of starts and ends of events in *edges* are saved in the vectors.

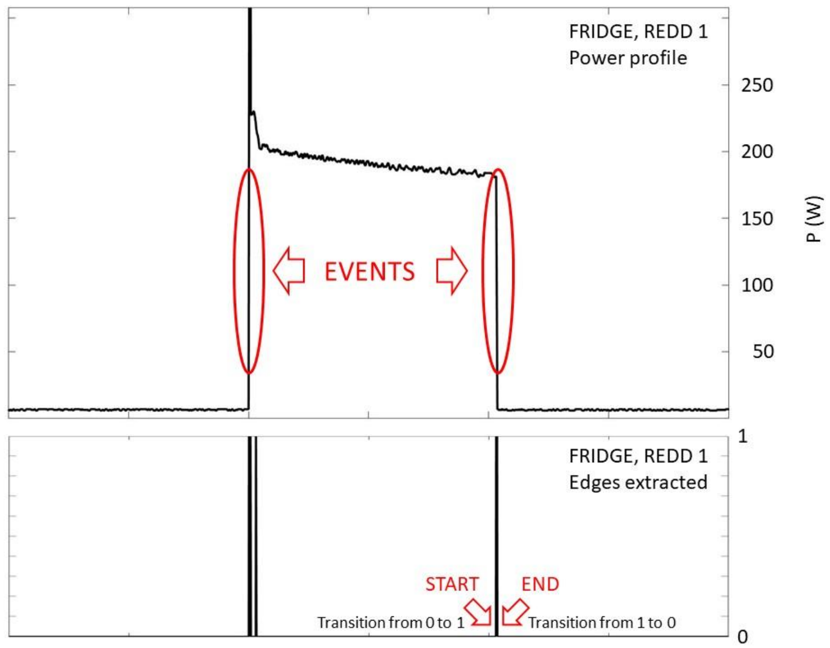


Figure 3.14: Edges in event evaluation, steps A, B and C

At this point time of events in plug data (ground truth events) is known. It is time to compare them with events in aggregate signal assigned by the disaggregation algorithm to the appliance in question (assigned events).

D. True positive, False positive and False negative computation

TRUE POSITIVE
For each assigned event, it is considered a true positive if its time of occurrence is between the start time minus 20 seconds and the end time plus 20 seconds of one ground truth event:

$$assigned\ event\ time\ \in [start\ time - 20,\ end\ time\ + 20]\ of\ a\ ground\ truth\ event$$

Figure 3.15 represents that.

FALSE POSITIVE
Each time the condition for true positive is not satisfied, the event is classified as false positive:

$$assigned\ event\ time \notin [start\ time - 20, end\ time\ + 20]\ of\ any\ ground\ truth\ event$$

FALSE NEGATIVE
Each time one ground truth event has no matching in inferred events, this count as a false negative:

$$\nexists\ assigned\ event\ time\ \in [start\ time - 20, end\ time\ + 20]\ of\ a\ ground\ truth\ event$$

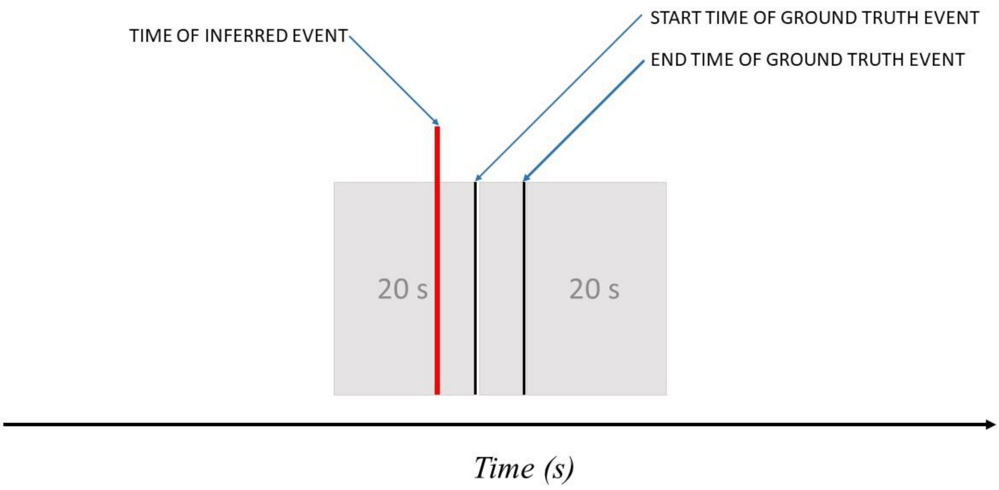


Figure 3.15: Concept of true positive computation

E. Metrics are computed using equations (1.4), (1.5) and (1.6). F-score, precision and recall are used in this work.

ENERGY EVALUATION

Main purpose of NILM work is to provide users energy consumption of appliances, in order to increase people awareness of how each device specifically contributes to their total energy demand.

Task of energy evaluation is somewhat more appliance-specific than event evaluation. What is described here is related to Type 1 appliances with regular behaviour. If necessary, extension to other devices will be specified later.

Two quantities are necessary at the beginning: *average runtime* and *average power* of the appliance.

AVERAGE RUNTIME

Using plug data in training days, ON and OFF transitions times are detected. The difference between each OFF and ON time is computed. The average of all these differences becomes *average runtime*.

AVERAGE POWER

Same ON and OFF transition times are used. The integral of plug level consumption in training is computed and divided by the whole time when the appliance is running, which is the sum of all the differences used for *average runtime*.

Now it is possible to compute *energy consumption* vector as follows. It is a vector with length equal to the whole evaluation period's number of samples. For each assigned event *e(i)* and its time *t(i)*:

A. If event *e(i)* is positive, potential off events are searched between *t(i)* and {*t(i) + val • average runtime*}. *val* is a parameter that can change between each appliance. If the appliance has a consistent behaviour, *val* can be set slightly higher than 1. That means that each usage has more or less the same length. If length of a single usage is variable, *val* has to be increased.

   a. If there is a match, *energy consumption* is set equal to average power for all the samples between the two events. If there are more matches, the closer one is kept.
   b. Otherwise *energy consumption* is set equal to average power for all samples between *t(i)* and {*t(i) + average runtime*}.

B. If else *e(i)* is negative, *energy consumption* is set equal to average power for all samples between {*t(i) - average runtime*} and *t(i).*

Vector *energy consumption* represents now estimated power profile, computed by the algorithm on the basis of assigned events. This vector is used to evaluate metrics related to energy.

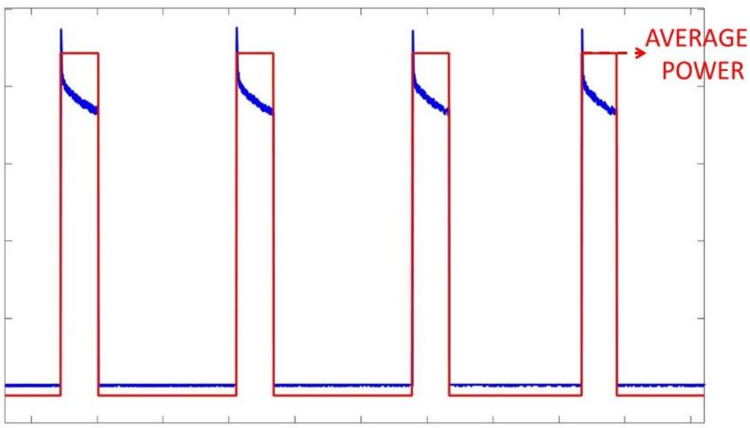


Figure 3.16: energy consumption vector for fridge in REDD 1. In blue the ground truth profile

ENERGY METRICS

Metrics used are F-score, Precision, Recall and Deviation in Percentage. First it is necessary to compute True positives, False positives and False negatives numbers. This is done comparing *energy consumption* vector to ground truth active power measures. A threshold is set in order to define whether the appliance is ON or OFF in ground truth and estimated power profile. For each sampling *i*:

$$\begin{Bmatrix} if\ energy\ consumption(i) < threshold\ \rightarrow estimated\ profile\ is\ OFF\ in\ i \\ if\ energy\ consumption(i) \geq threshold \rightarrow estimated\ profile\ is\ ON\ in\ i \end{Bmatrix} \tag{3.11}$$

$$\begin{Bmatrix} if\ groud\ truth\ power(i) < threshold\ \rightarrow ground\ truth\ profile\ is\ OFF\ in\ i \\ if\ groud\ truth\ power(i) \geq threshold \rightarrow ground\ truth\ profile\ is\ ON\ in\ i \end{Bmatrix} \tag{3.12}$$

TRUE POSITIVES

A true positive occurs at the *i* –th sample if both profiles are ON in it.

FALSE POSITIVES

For each *i* where estimated profile is ON and ground truth is OFF.

FALSE NEGATIVE

For each *i* where ground truth profile is ON and estimated one is OFF.

With TP, FP and FN is possible to compute F-score, Precision and Recall. To compute Deviation in Percentage integrals of estimated and ground truth power profiles are used to compute estimated and ground truth energies.

$$Deviation\ in\ Percentage = \frac{abs(estimated\ energy - ground\ truth\ energy)}{ground\ truth\ energy} \tag{3.13}$$

USAGE EVALUATION

Usage detection's goal is to determine how many times a specific appliance runs each day. Computation is performed as follows.

USAGE DETECTION

For each day assigned events of appliance in question are considered. It is necessary to determine a parameter called *usage duration,* specific for each appliance and indicative of the runtime of it. The algorithm runs through events $e_{day}(i)$ in that way, until there are still events in considered day:

A. If there is an event $e_{day}(i)$, the number of usages is increased by 1.

B. Algorithm then skips all the events after $e_{day}(i)$ and until $\{t(i) + usage\ duration\}$.

C. If another event is found after that time window, number of usages is increased by 1. Then repeat from step B.

That is iterated for each day, until there are events in it.

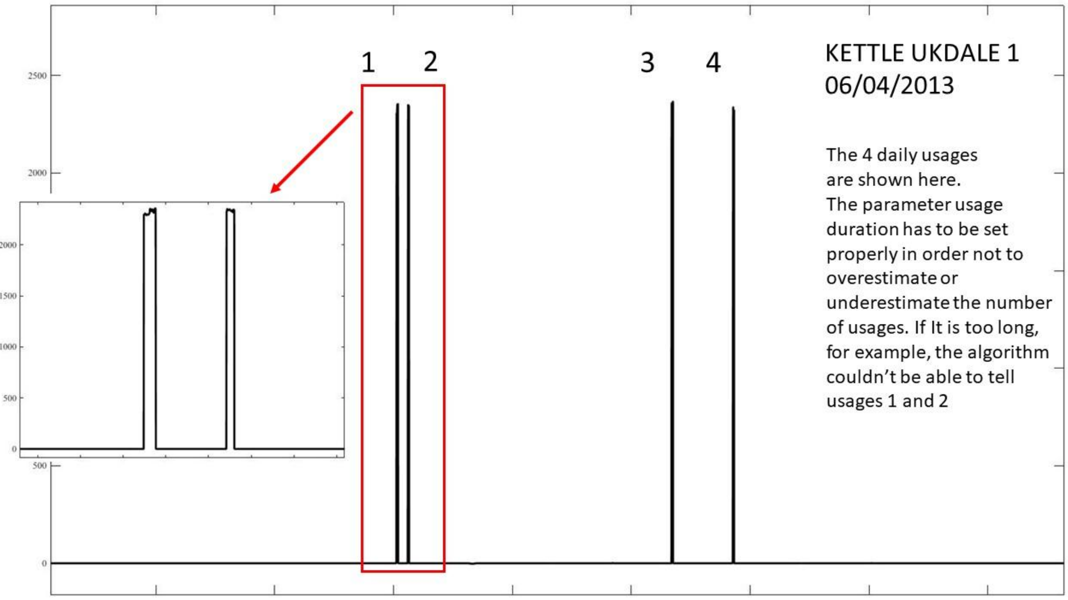


Figure 3.17: Example of usages in one day of a kettle

Same routine is then applied to ground truth data in evaluation period. This gives information about actual number of usages of the appliances. At this point for both ground truth data and assigned events number of usages for each day is available. TP, FP and FN can be computed.

TRUE POSITIVES

For each day the number of true positives is the minimum between ground truth usages and inferred usages

FALSE POSITIVES

For each day, if inferred usages are more than ground truth ones, their difference is the number of false positives.

FALSE NEGATIVES

When instead ground truth usages are more than inferred ones, their difference is the number of false negatives for that day

All TP, FP and FN of each day are then summed up. It is now possible to compute F-score, Precision and Recall.

3.3 RESULTS AND DISCUSSION

The two algorithms have been used for several experiments on datasets described in 3.2.2. A good number of these experiments is reported in Appendix E.

3.3.1 ALGORITHMS COMPARISON

Nearly all the experiments in Appendix E have been performed for both algorithms. It is therefore possible to compare them.

ACTIVE POWER – BASED EXPERIMENTS

Using Tables *F1 – F10* in Appendix F, Tables *3.2* and *3.3* are composed. For each household *Table 3.2* details for how many appliances one algorithm outperform the other in that specific metric. Only differences higher than 3 % in absolute value are taken into account.

	F-SCORE		PRECISION		RECALL	
HOUSE	WEISS	HART	WEISS	HART	WEISS	HART
REDD1	4	2	4	2	3	2
REDD2	3	0	2	0	2	1
REDD3	2	0	3	1	2	1
REDD4	2	1	3	0	2	1
UKDALE 1	2	3	1	1	2	3
UKDALE 2	1	3	1	2	1	3
ECO1	1	2	1	3	2	2
ECO2	3	0	3	0	4	0
ECO4	2	0	2	1	2	0
ECO5	2	0	2	0	1	0
TOTAL	22	11	22	10	21	13

Table 3.2: Weiss vs Hart performances, P only

In *Table 3.3* cooling appliances only are taken into account. Symbol '•' is placed on the column of the algorithm that performs better.

	F-SCORE		PRECISION		RECALL	
HOUSE	WEISS	HART	WEISS	HART	WEISS	HART
REDD1	•		•		•	
REDD2	•		•		•	
REDD3	•		•		•	
REDD4	•		•		•	
UKDALE 1	•		•		•	
UKDALE 2	•		•		•	
ECO1/Freezer		•	•		•	
ECO1/Fridge	•			•	•	
ECO2/Freezer	•		•		•	
ECO2/Fridge	•		•		•	
ECO4/Freezer	•		•		•	
ECO4/Fridge	•		•		•	
ECO5/Fridge	•		•		•	
TOTAL	12	1	12	1	13	

Table 3.3: Weiss vs Hart performances on cooling appliances

P – Q BASED EXPERIMENTS

Using Tables *F11 – F17* in Appendix F, Tables *3.4* and *3.5* are composed. For each household *Table 3.4* details for how many appliances one algorithm outperform the other in that specific metric. Only differences higher than 3 % in absolute value are taken into account.

	F-SCORE		PRECISION		RECALL	
HOUSE	WEISS	HART	WEISS	HART	WEISS	HART
UKDALE 1	3	3	1	1	2	2
UKDALE 2	2	1	2	0	2	2
UKDALE 5	1	3	0	3	1	3
ECO1	2	2	2	2	2	2
ECO2	3	1	1	0	3	1
ECO4	2	1	0	1	2	1
ECO5	2	0	1	0	1	1
TOTAL	15	11	7	7	13	12

Table 3.4: Weiss vs Hart performances, P and Q

In *Table 3.5* cooling appliances only are taken into account. Symbol '•' is placed on the column of the algorithm that performs better. '//' stands for equality, when the difference is less than 3%.

	F-SCORE		PRECISION		RECALL	
HOUSE	WEISS	HART	WEISS	HART	WEISS	HART
UKDALE 1	•		•		//	//
UKDALE 2	//	//	//	//	•	
UKDALE 5	•		//	//	•	
ECO1/Freezer		•	•			•
ECO1/Fridge	•		•		•	
ECO2/Freezer	•		•		•	
ECO2/Fridge	•		//	//	•	
ECO4/Freezer	•		//	//	•	
ECO4/Fridge	•		//	//	•	
ECO5/Fridge	•		//	//	•	
TOTAL	8	1	4	0	8	1

Table 3.5: Weiss vs Hart performances on cooling appliances

A general sight shows that Weiss algorithm is never outperformed by Hart in all the metrics, as shown by rows TOTAL in all four tables above. For cooling appliances this trend is even more accentuated. *Table 3.4* shows more uniform behaviour between the 2 algorithms. This could be due to absence of REDD dataset in this table. *Table 3.2* shows that in this dataset Weiss algorithm outperforms Hart's nearly everywhere. More appliance-specific information will be given later if necessary.

3.3.2 ACTIVE AND REACTIVE POWER

For UKDALE houses 1 and 2 and for all households in ECO dataset experiments have been performed both with active power only and with addiction of reactive power. In Appendix G comparison of metrics between the two cases is provided specifically at appliance level. In this section a summary of that is provided. *Table 3.6* summarizes information from Tables *G.1 – G.6*. Numbers in cells represent the number of appliances for which the metric obtained with specified features (P/Q or P) outperform the metric with the other feature. Only Differences higher than 3% in absolute value are taken into account.

	F-SCORE				PRECISION				RECALL			
	Weiss		Hart		Weiss		Hart		Weiss		Hart	
HOUSE	P/Q	P	P/Q	P	P/Q	P	P/Q	P	P/Q	P	P/Q	P
UKDALE 1	3	2	2	2	3	2	3	1	1	3	1	3
UKDALE 2	2	1	3	2	5	0	4	1	1	4	1	3
ECO1	1	1	1	2	3	0	3	0	0	2	0	4
ECO2	1	2	2	0	4	0	4	0	0	3	2	1
ECO4	2	1	2	1	3	0	3	0	1	1	2	1
ECO5	0	2	0	1	1	0	1	0	0	2	0	2
TOTAL	9	9	10	8	19	2	18	2	3	15	6	14

Table 3.6: P/Q vs P, performances

From F-score it is not possible to discriminate one set of features as better in term of disaggregation performances. As it can be seen in TOTAL row, 9 appliances are better disaggregated with P/Q features, and 9 with P only using Weiss algorithm. They are 10 and 8 for Hart algorithm.

Precision metric scores in *Table 3.6* show clearly that adding Q as feature precision increases in the majority of cases.

Recall generally decreases using P/Q features. It is expected because the addition of a feature increases selectivity of algorithms on events.

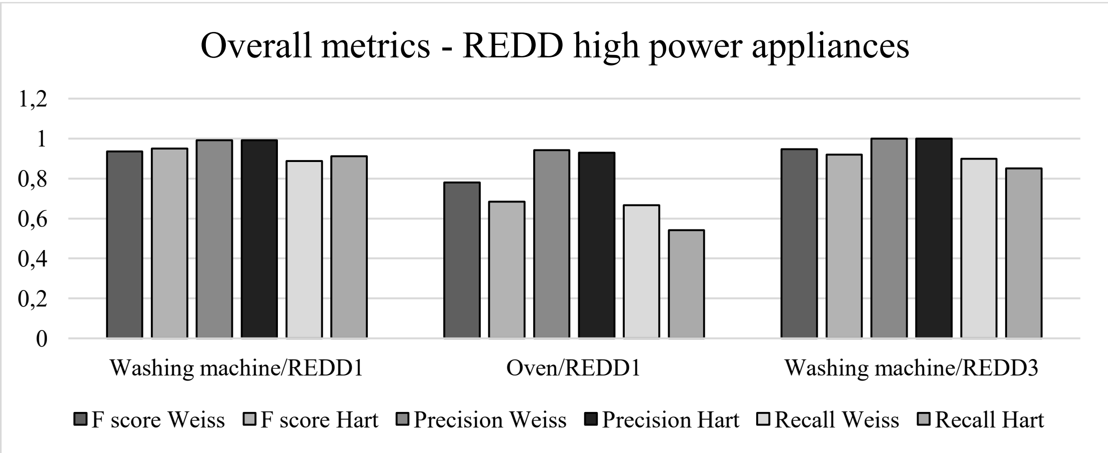
According to these results, it is not possible to determine that reactive power increases significantly results of employed algorithms. What can be said is that it boosts precision at the expense of recall. For energy evaluation purposes good precision is very important, while recall can be not too high. As it has been explained in 3.2.5, use of *average runtime* parameter allow to predict consumption profile quite well even if some events are missing. If otherwise false positive events are present, they are used to add sequences of samples where the appliance is estimated ON. That can change value of predicted energy significantly, especially in presence of many false positives (low precision).

3.3.3 HIGH POWER APPLIANCES

In this section results obtained with major appliances are presented. High power appliances are defined here as appliances characterized by events with magnitude in active power generally higher than 2400 W. *Table 3.7* details results in event detection of these appliances.

Dataset/ household	Appliance	F-score Weiss	F-score Hart	Precision Weiss	Precision Hart	Recall Weiss	Recall Hart	Parameter r	Features
REDD 1	Washing machine	0.9361	0.9495	0.9909	0.9912	0.8871	0.9113	0.1	P
REDD 1	Oven	0.7805	0.6842	0.9412	0.9286	0.6667	0.5417	0.1	P
REDD 3	Washing machine	0.9465	0.9194	1.0000	1.0000	0.8985	0.8507	0.1	P
UKDALE 1	Kettle	0.8888	0.9093	0.8706	0.8969	0.9079	0.9221	0.1	P/Q
		0.9433	0.9557	0.9769	0.9810	0.9119	0.9316	0.1	P
UKDALE 2	Kettle	0.9174	0.9183	0.9952	0.9937	0.8509	0.8536	0.1	P/Q
		0.9141	0.9303	0.9874	0.9969	0.8509	0.8720	0.1	P
UKDALE 5	Kettle	0.8199	0.8862	0.8924	0.9440	0.7582	0.8352	0.05	P/Q

Table 3.7, Results in event detection for High Power appliances



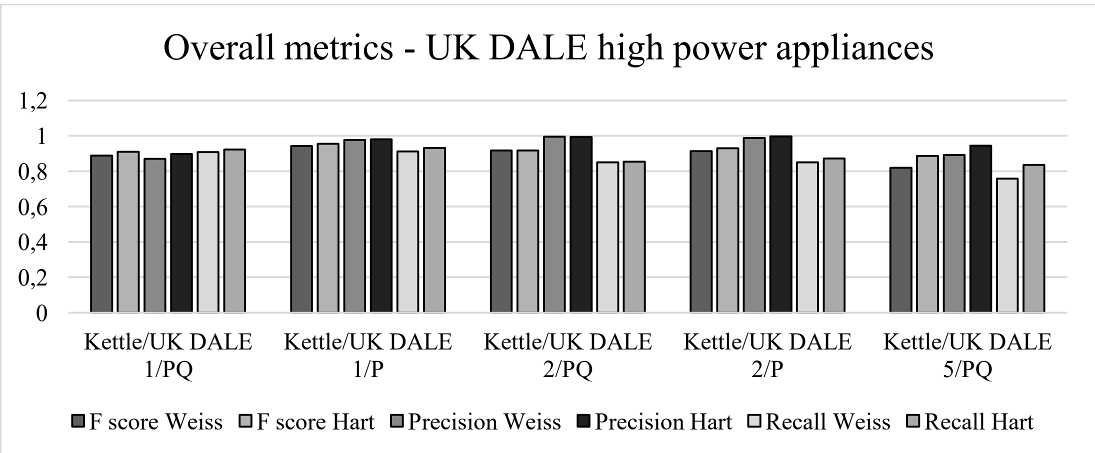


Figure 3.18: Overall event metrics for high power appliances

As can be seen from *Figure 3.18*, metrics are usually higher than 0.8 for all appliances. Only in REDD1/oven there is a drop in performances, with still high precision and F-score not far from 0.8, using Weiss algorithm. In general it seems that for all these appliances both algorithms and set of features can be used for energy disaggregation. Following disaggregation is performed using Weiss algorithm and P/Q when available.

Table 3.8 details energy metrics (described in 3.2.5) obtained for appliances in question, and using events obtained from experiments in *Table 3.7*.

Dataset/ household	Appliance	F score	Precision	Recall	Deviation %
UKDALE 1	Kettle	0.7446	0.6893	0.8095	0.1888
UKDALE 2	Kettle	0.9173	0.9370	0.8984	0.0493
UKDALE 5	Kettle	0.7758	0.7818	0.7699	0.0273
REDD 1	Washing machine	0.7437	0.9468	0.6123	0.0704
REDD 3	Washing machine	0.4851	0.9992	0.3203	0.1106
REDD 1	Oven	0.5613	0.7407	0.4519	0.3326

Table 3.8: Energy metrics for High power appliances

Deviation in percentage has been underlined in the table. That metric describes how energy predicted is far from ground truth value. As far as we know, there are no guidelines that state an acceptable error threshold for Deviation % for this kind of applications. In this work 20% of error is considered acceptable. For REDD1/Oven, the only appliance with problems in event metrics, Deviation in percentage is still very high. In particular analysing other metrics Recall is quite low. On the contrary for kettle in UKDALE 1 Precision is lower than in all other cases. To increase recall for Oven the experiment is run again increasing parameter *r* up to 0.2, in order to catch more events. For kettle in UKDALE 2 *r* is decreased to 0.05 in order to discard false positive events from assigned ones. That is likely going to increase precision. *Table 3.9* shows updated results in energy metrics.

Dataset/ household	Appliance	F score	Precision	Recall	Deviation %
REDD 1	Oven	0.7329	0.7899	0.6836	0.0532
UK DALE 1	Kettle	0.7467	0.7313	0.7627	0.0559

Table 3.9: Energy metrics for High power appliances, updates

As expected, Recall is increased for Oven and precision for Kettle too. Deviation in % is better now for both cases.

Figure 3.19 summarizes results achieved in Deviation for all the appliances used in this section.

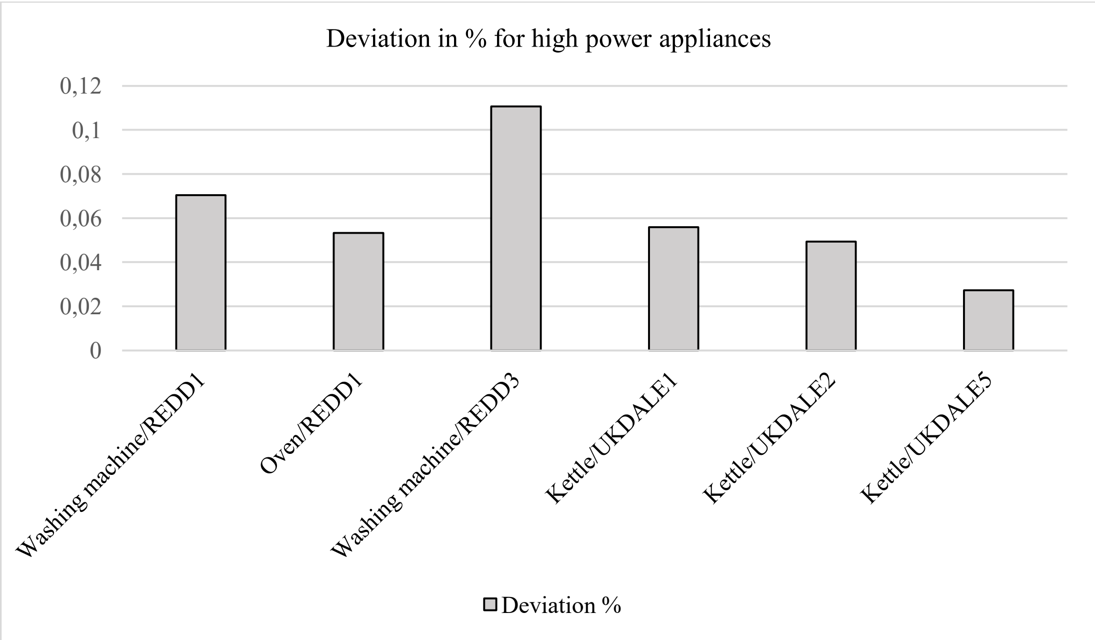


Figure 3.19: Deviation in % for high power appliances

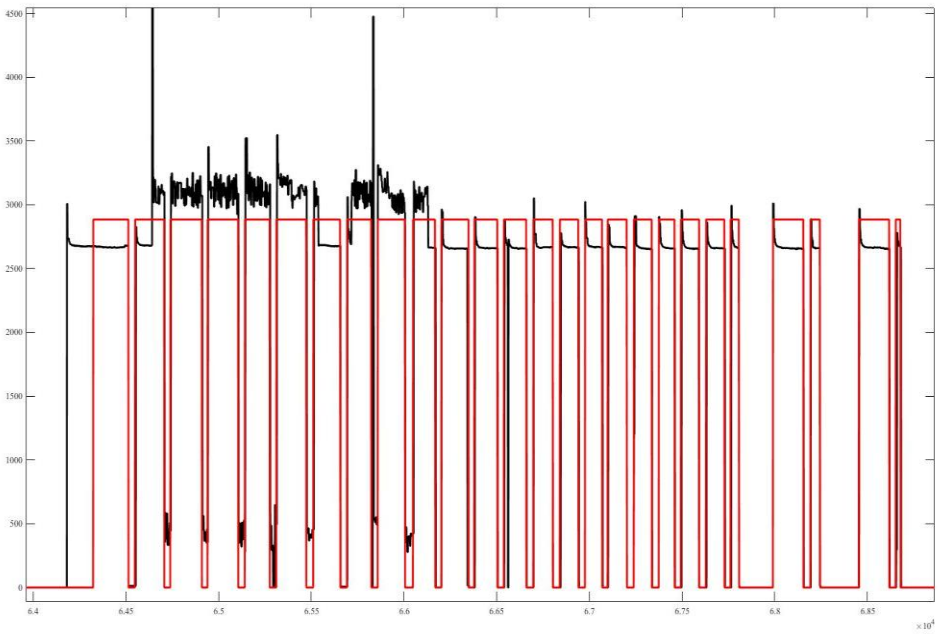


Figure 3.20: Example of actual consumption and predicted one for Washing machine, REDD 1. Time covered in this picture is about one hour and a half

3.3.4 COOLING APPLIANCES

The most represented appliances in experiments (reported in Appendix E) are Fridges and Freezers. Despite being appliances with low power draw, they show relatively good performances, especially in term of Precision. *Table 3.10* summarizes event metrics for cooling appliances.

Appliance	HOUSE/ DATASET	F-score Weiss	F-score Hart	Precision Weiss	Precision Hart	Recall Weiss	Recall Hart	Param. r	Features
Fridge	REDD 1	0.5971	0.5198	0.8297	0.7517	0.4663	0.3973	0.1	P
Fridge	REDD 2	0.5411	0.4757	0.9204	0.7254	0.3832	0.3539	0.1	P
Fridge	REDD 3	0.6087	0.5250	0.8463	0.6214	0.4753	0.4544	0.1	P
Fridge	REDD 4	0.6068	0.4223	0.8988	0.6796	0.4580	0.3063	0.1	P
Fridge	UKDALE 1	0.7639	0.7250	0.9433	0.8605	0.6418	0.6265	0.2	P/Q
		0.4448	0.3353	0.6354	0.3273	0.3421	0.3437	0.2	P
Fridge	UKDALE 2	0.7865	0.7641	0.9787	0.9779	0.6573	0.6269	0.2	P/Q
		0.7744	0.6959	0.8522	0.6972	0.7096	0.6946	0.2	P
Fridge	UKDALE 5	0.8965	0.7571	0.9939	0.9812	0.8166	0.6163	0.1	P/Q
Freezer	ECO 1	0.4922	0.6884	0.8904	0.8012	0.3401	0.6034	0.1	P/Q
		0.5259	0.5914	0.7170	0.5310	0.4152	0.6673	0.1	P
Fridge	ECO 1	0.4526	0.2447	0.9696	0.7578	0.2952	0.1459	0.1	P/Q
		0.4384	0.3432	0.4330	0.5457	0.4439	0.2503	0.1	P
Freezer	ECO 2	0.8379	0.5747	0.9961	0.7973	0.7230	0.4492	0.1	P/Q
		0.8634	0.4485	0.8246	0.7386	0.9060	0.3220	0.1	P
Fridge	ECO 2	0.9302	0.5033	0.9901	0.9793	0.8772	0.3387	0.1	P/Q
		0.9297	0.4910	0.9324	0.8599	0.9271	0.3436	0.1	P
Fridge	ECO 4	0.6505	0.6141	0.5322	0.5244	0.8363	0.7408	0.1	P/Q
		0.5883	0.4131	0.4497	0.3431	0.8502	0.5191	0.1	P
Freezer	ECO 4	0.6134	0.5580	0.9984	0.9991	0.4427	0.3871	0.1	P/Q
		0.5529	0.3159	0.8789	0.3709	0.4033	0.2751	0.1	P
Fridge	ECO 5	0.4826	0.2235	0.9961	0.9796	0.3184	0.1261	0.1	P/Q
		0.5969	0.3940	0.9492	0.6320	0.4354	0.2862	0.1	P

Table 3.10: Results in event detection for High Power appliances

In Appendix H comparison between different algorithms and features for cooling appliances is detailed. Conclusion achieved is that Weiss algorithm with P/Q features is the best solution, so it will be employed for remainder of this section.

Results obtained for energy disaggregation, using events of experiments in *Table 3.10,* are reported in *Table 3.11.*

Appliance	Dataset/ House	F score	Precision	Recall	Deviation in percentage
Fridge	UK DALE 1	0.8844	0.9436	0.8323	0.1342
Fridge	UK DALE 2	0.9043	0.9116	0.8971	0.0252
Fridge	UK DALE 5	0.9323	0.9626	0.9039	0.0587
Fridge	REDD 1	0.8387	0.9344	0.7607	0.1905
Fridge	REDD 2	0.8797	0.9712	0.8039	0.1775
Fridge	REDD 3	0.8292	0.8878	0.7779	0.1244
Fridge	REDD 4	0.8799	0.9239	0.8398	0.0223
Fridge	ECO 1	0.7809	0.9787	0.6496	0.3589
Freezer	ECO 1	0.7963	0.9513	0.6848	0.2731
Fridge	ECO 2	0.9060	0.9698	0.8501	0.0740
Freezer	ECO 2	0.8993	0.9352	0.8660	0.0939
Fridge	ECO 4	0.7180	0.6026	0.8881	0.5177
Freezer	ECO 4	0.8968	1.0000	0.8128	0.0597
Fridge	ECO 5	0.8388	0.9927	0.7261	0.2821

Table 3.11: Energy metrics for cooling appliances

Results in energy disaggregation are promising in many cases. Problems are in ECO 4/fridge, ECO 5/fridge and ECO 1/fridge and freezer. Fridge in ECO 4 is the only Cooling appliance with low precision in energy metrics (see *Table 3.10*). It means that a good percentage of assigned events are false positives. That produces low precision in energy metrics too.

The other problematic appliances are characterized by low recall in event metrics, as shown in *Figure 3.21.*

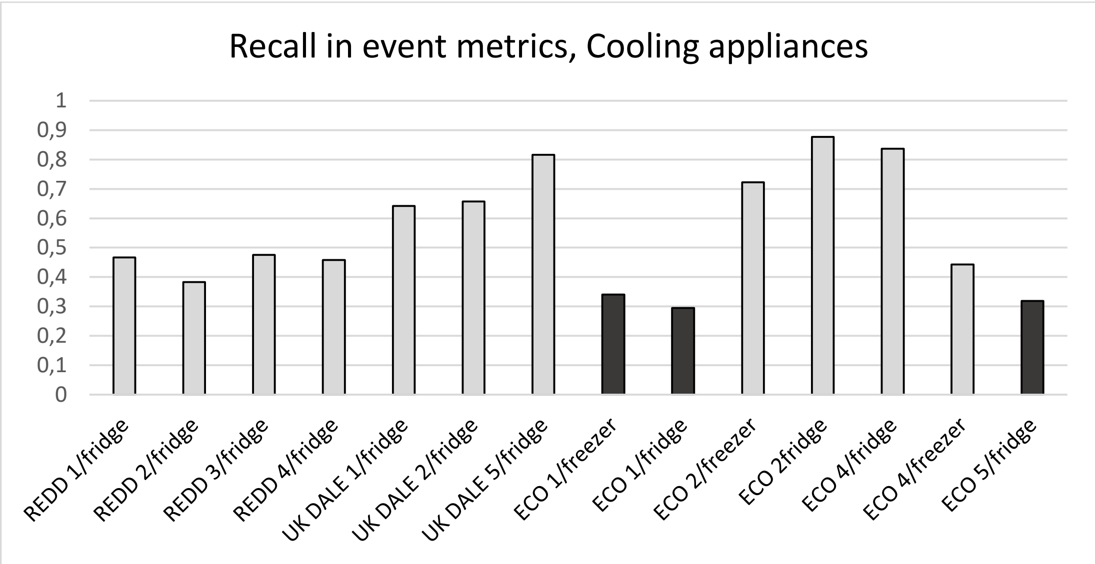


Figure 3.21: Recall in event metrics, cooling appliances

For these three appliances new experiments are run with parameter *r* changed from 0.1 to 0.2. This is supposed to increase recall in event metrics. New results both in events and energy metrics are reported in *Table 3.12 and 3.13.* It is clear that improvement is significant compared to previous results. *Figure 3.22* compare results with the two values of *r*.

Appliance	House	F score	Precision	Recall	Parameter r
Freezer	ECO 1	0.5760	0.7774	0.4574	0.2
Fridge	ECO 1	0.4509	0.6541	0.3440	0.2
Fridge	ECO 5	0.6059	0.9824	0.4381	0.2

Table 3.12: Events metrics for new experiments, cooling appliances

Appliance	Dataset/ House	F score	Precision	Recall	Deviation in percentage
Freezer	ECO 1	0.9100	0.9377	0.8838	0.0481
Fridge	ECO 1	0.7929	0.8259	0.7624	0.1083
Fridge	ECO 5	0.9630	0.9880	0.9392	0.0670

Table 3.13: Energy metrics for new experiments, cooling appliances

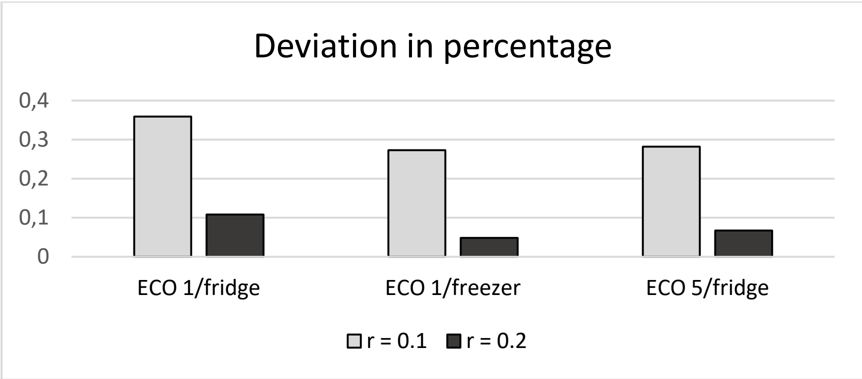


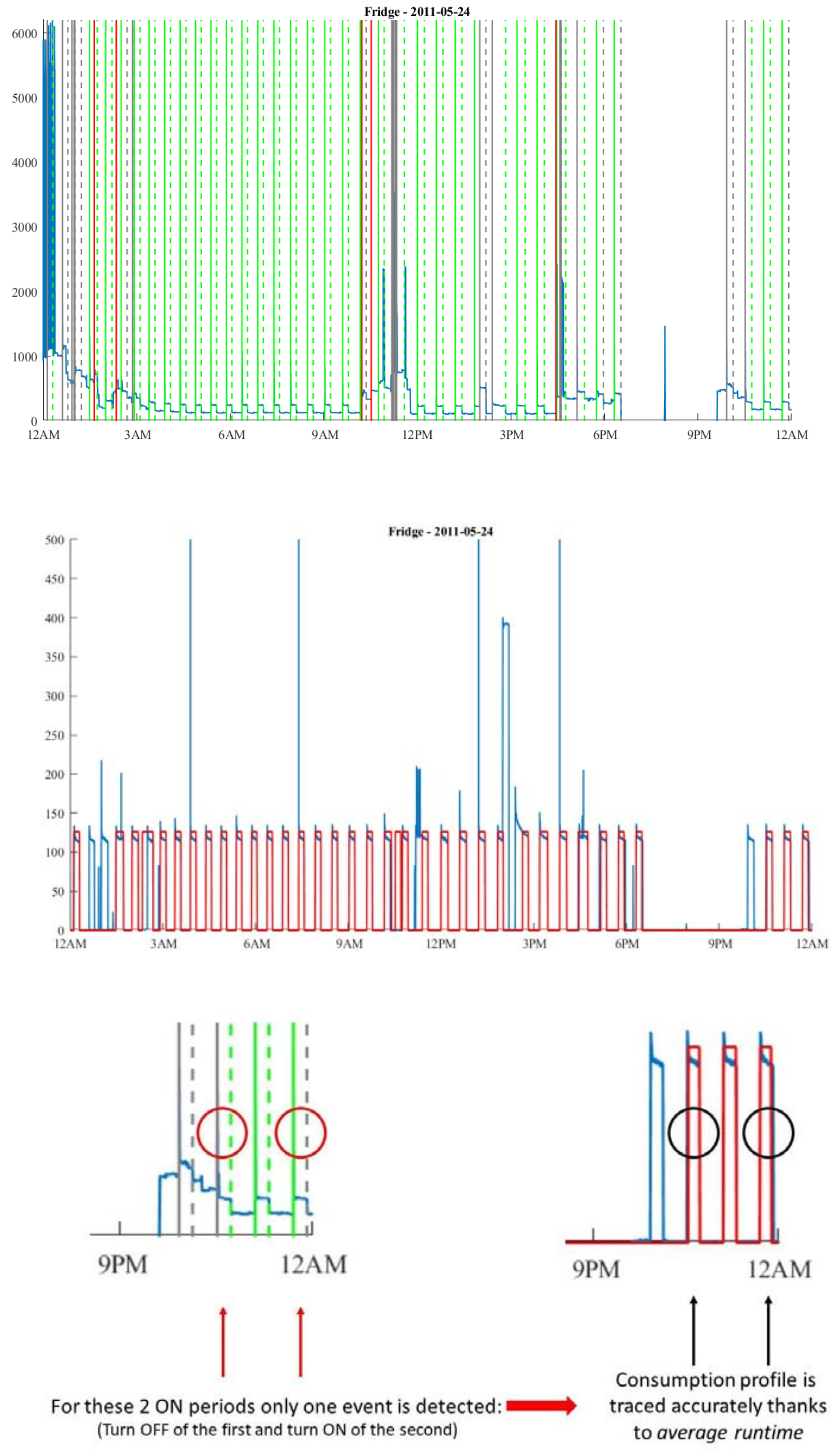
Figure 3.22: Performances increase moving from r = 0.1 to r = 0.2

After this adjustment, nearly all deviations are within 20%, so energy disaggregation can be considered good. As said before, fridge in ECO 4 is the only problematic appliance. A reason for that could be the similarity between features of fridge and freezer in that household, as can be seen in *Table 3.14.* Moreover that table shows that Weiss algorithm has achieved good results in disaggregation of cooling appliances that are, in certain cases, very small, like fridge and freezer in ECO 1. A reason for good performances with cooling appliances could be the fact that they are running all the day, with a good number of events for each day. If there is an error with some events unassigned or with false positives, it is less relevant if the number of correctly assigned events is quite high, which is the case here.

Appliance	Dataset/ House	Active Power signature 1 (W)	Active Power signature 2 (W)
Fridge	UK DALE 1	96	-80
Fridge	UK DALE 2	85	-74
Fridge	UK DALE 5	124	-95
Fridge	REDD 1	221	-171
Fridge	REDD 2	253	-229
Fridge	REDD 3	129	-113
Fridge	REDD 4	138	-116
Fridge	ECO 1	60	-45
Freezer	ECO 1	29	-30
Fridge	ECO 2	80	-69
Freezer	ECO 2	62	-50
Fridge	ECO 4	112	-87
Freezer	ECO 4	106	-94
Fridge	ECO 5	142	-103

Table 3.14: Active power signatures of cooling appliances

Another highlight has to be done on Recall. While in event metrics it is quite low, it increases in energy metrics in all cases. The explanation is simple. One refrigerator cycle is detected with two events, turn on and turn off. If hypothetically only one event is detected for each cycle, the overall recall will be 50%. When energy consumption is computed, each ON event is paired with an OFF event, if it is found within a time window following this event, as described in 3.2.5. In that way consumption profile is easily traced. If instead one ON or one OFF event is unmatched with other near events opposite in sign, consumption profile is predicted placing a hypothetic event at a time distance equal to parameter *average runtime,* described in 3.2.5 too. That works quite well in the case of cooling appliances, due to their regular behaviour. *Figure 3.22* claims to give a graphical explanation of that. The first graph details assigned and unassigned events, the second one the reconstructed power profile for the appliance.



*continuous lines are positive events, dashed lines are negative ones. Green events are true positives, red false positives and grey false negatives.

Figure 3.22: Reconstruction of Fridge's consumption profile, REDD 3

3.3.5 KETTLES AND STOVES

All kettles have been disaggregated with good results (*Table 3.15*). In this section energy metrics are reported. Weiss algorithm with P/Q features has been used for this computation.

Dataset/Household	F-score	Precision	Recall	Parameter r
UKDALE 1	0.8888	0.8706	0.9079	0.1
UKDALE 2	0.9174	0.9952	0.8509	0.1
UKDALE 5	0.8199	0.8924	0.7582	0.05
ECO 1	0.9095	0.9179	0.9013	0.1
ECO 2	0.8184	0.9754	0.7049	0.1

Table 3.15: Events metrics for Kettles

Results in event metrics are consistent, so energy metrics are expected to be positive to. *Table 3.16* confirms that. Deviation is below 20% for all the cases.

Dataset/Household	F score	Precision	Recall	Deviation in %
UKDALE 1	0.7446	0.6893	0.8095	0.1888
UKDALE 2	0.9173	0.9370	0.8984	0.0493
UKDALE 5	0.7758	0.7818	0.7699	0.0273
ECO 1	0.8920	0.8490	0.9396	0.1230
ECO 2	0.9635	0.9546	0.9726	0.0223

Table 3.16: Energy metrics for Kettles

All stoves considered in experiments have been disaggregated with good results (*Table 3.17*). They are only three so they can't show a significant trend. Event metrics are characterized by excellent precision in all cases. Only recall in REDD 4 is relatively low. Energy metrics reported in *Table 3.18* proof that a good energy prevision can be done with these appliances.

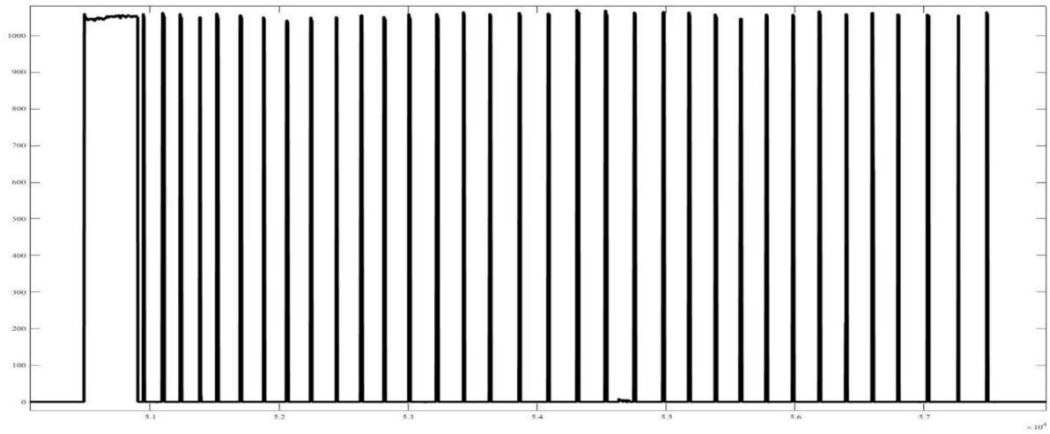
Dataset/Household	F score	Precision	Recall	Parameter r
REDD 2	0.9803	0.9836	0.9770	0.1
REDD 4	0.7804	0.9677	0.6538	0.1
ECO 2	0.9981	1.0000	0.9962	0.1

Table 3.17: Events metrics for Stoves

Dataset/Household	F score	Precision	Recall	Deviation
REDD 2	0.7854	0.7596	0.8129	0.0898
REDD 4	0.7985	0.8120	0.7855	0.0222
ECO 2	0.7367	0.7999	0.6828	0.1389

Table 3.18: Energy metrics for Stoves

To achieve results in *Table 3.18* a modification in energy evaluation routine presented in 3.2.5 has been done. It can be seen from *Figure 3.23* that the power behaviour is characterized by a first long period of constant consumption and by several ON – OFF cycles, in order to to keep the temperature constant. If the system miss the first period, a significant error is produced. In that case the reconstruction using *average runtime* parameter will set the power for a period too short, comparable to the minor ones that follow the first.



a) Ground truth active power of one usage

b) Reconstructed active power of one usage

c) overlap of the two profiles

Figure 3.23: Ground truth and reconstructed active power profiles for stove, REDD 2

To solve this problem the system check the beginning of each usage and, if power is not constant for a long enough time window, all samples between the first 2 ON periods in consumption profile are set to *average power*. This solution worked well in the examined cases.

3.3.6 MID/HIGH POWER APPLIANCES

If we consider the other appliances with active power draw between 1000 and 2400 W, three categories are well represented: microwaves, dishwashers and washing machines. For these appliances results obtained are variable. They are gathered together in *Table I.1.* Focus of this section is on these 3 appliances.

MICROWAVES

Event results in microwave are presented in *Table 3.19.* Experiments reported here are the ones with P/Q features, with exception of REDD dataset, where only P is available.

DATASET/HOUSE/ APPLIANCE	F-score Weiss	F-score Hart	Precision Weiss	Precision Hart	Recall Weiss	Recall Hart	Param. r
REDD 1/Microwave	0.6557	0.6084	0.9610	0.8200	0.4977	0.4836	0.1
	0.6641	0.6214	0.9620	0.8269	0.5070	0.4977	0.8
REDD 2/Microwave	0.9036	0.7556	0.9785	1.0000	0.8393	0.6071	0.1
REDD 3/Microwave	0.7715	0.7314	0.7667	0.6543	0.7763	0.8289	0.1
	0.7548	0.6935	0.7231	0.5895	0.7895	0.8421	0.5
UKDALE 1/Microwave	0.8812	0.8159	0.9865	0.9857	0.7962	0.6960	0.1
UKDALE 2/Microwave	0.6673	0.9086	0.9518	0.9809	0.5138	0.8462	0.1
	0.7857	0.8760	0.8263	0.8754	0.7489	0.8766	0.3
	0.8121	0.8237	0.7390	0.7176	0.9013	0.9666	1
ECO 4/Microwave	0.5529	0.7001	0.9327	0.9911	0.3929	0.5412	0.1
ECO 5/Microwave	0.7182	0.6744	0.6735	0.5530	0.7692	0.8643	0.1

Table 3.19: Event metrics in microwaves

Performances are good but some problem are present in few cases. A good energy evaluation doesn't seem possible for all the cases so the focus of the following parts will be on usage detection. Comparison between algorithms and features is not reported here, but can be done starting from *Table 3.19.* In majority of cases results of different algorithms are similar, with some experiments in favour of both algorithms. In continuity with 3.3.3 – 3.3.5, Usage detection has been performed with results from Weiss algorithm and P/Q features when available. Table 3.20 report results.

Dataset/house	F score	Precision	Recall	Parameter r
REDD 1	0.757576	1.000000	0.609756	0.8
REDD 2	1.000000	1.000000	1.000000	0.1
REDD 3	0.800000	0.666667	1.000000	0.05
UK DALE 1	0.966486	0.984581	0.949045	0.1
UK DALE 2	0.950108	0.948052	0.952174	0.1
ECO 4	0.928571	0.951220	0.906977	0.1
ECO 5	0.733333	0.666667	0.814815	0.055

Table 3.20: Usage metrics in microwaves

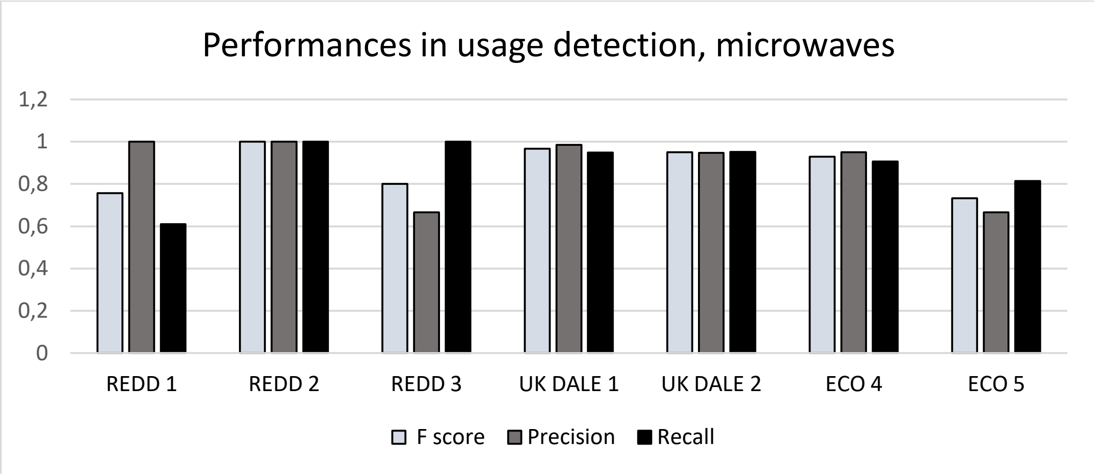


Figure 3.24: Usage metrics in microwaves

In 4 out of 7 examples results are excellent in all metrics. In REDD 1 recall is low, while for REDD 2 and ECO 5 there are precision problems.

Household 1 in REDD is problematic because three appliances in it have features that are very close, as can be seen from *Table 3.21.*

Appliance	ON event signature	OFF event signature
Microwave	1524.60	-1520.03
Water kettle	1530.13	-1527.96
Bathroom gfi	1505.96	-1536.66

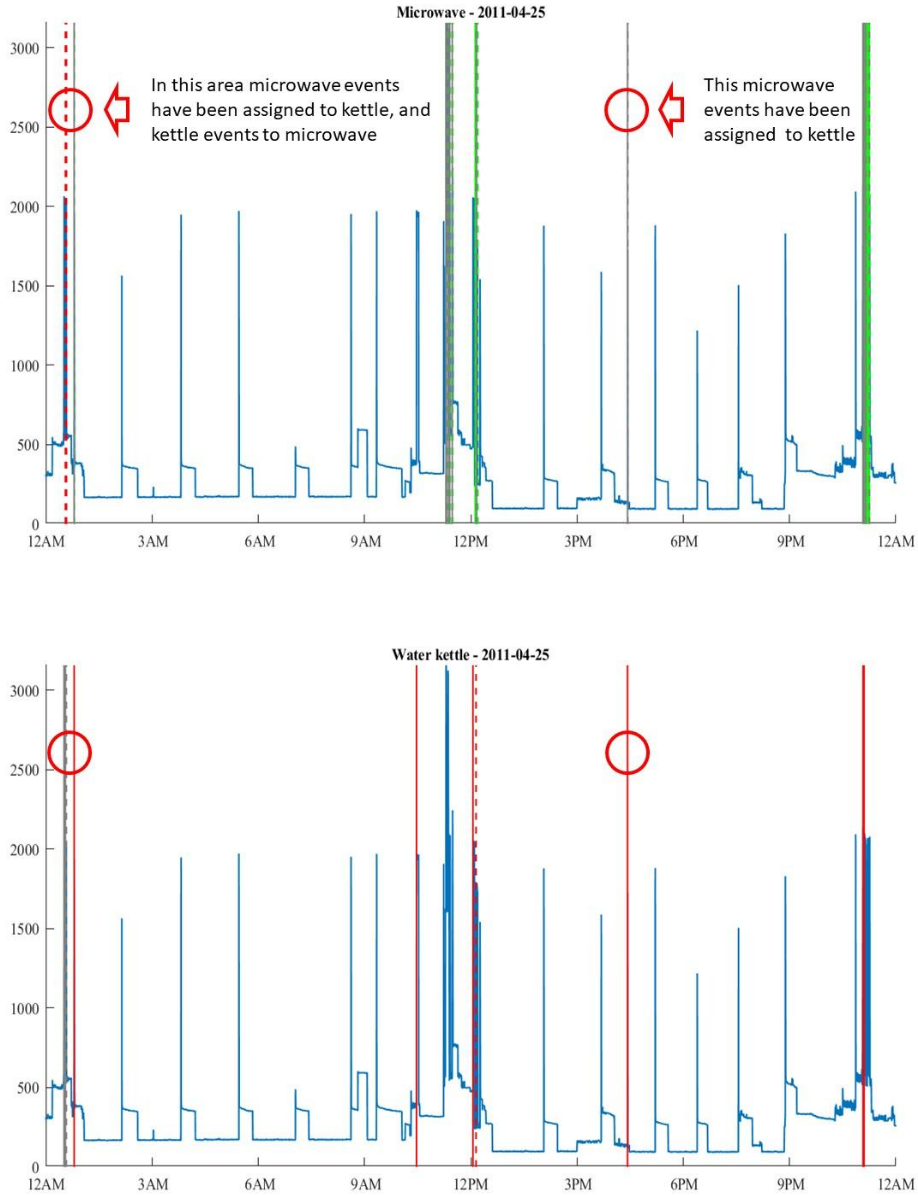
Table 3.21: Features of mid/high power appliances in REDD 1

If one event is detected, it is assigned to the correct appliance only if it is extremely close to the signature. Considering that power variability of events of a single appliance is certainly higher than the small differences between these signatures, the result is a random assignment to one of the three appliances. *Figure 3.25* shows how some events are wrongly assigned between microwave and kettle in REDD 1. Reactive power is not available so it is not possible to see if it differentiate signatures of these appliances better. In 3.3.7 considerations about introduction of time usage patterns will be done.

In REDD 3 and ECO 5 low precision in event detection is reflected in low precision in usage metrics.

DISHWASHERS AND WASHING MACHINES

In *Table I.2* event results for dishwasher and washing machines are reported. With some exception they are quite low. Next section will consider introduction of time patterns to increase disaggregation metrics. Washing machines considered in 3.3.3 are not reported here. Actually in REDD they are labelled as washer dryers, and their power draw is significantly higher than other washing machines. They have been considered washing machines in event detection only for simplicity.



**Events are represented by vertical bars. Red bars are false positives, green true positives and grey false negatives.*

Figure 3.25: Confusion in assignment of events between kettle and microwave, REDD 1.

3.3.7 MID/HIGH POWER APPLIANCES AND TIME OF THE DAY PATTERNS

For all the appliances considered in 3.3.6 time patterns have been studied. All the graphs are reported in Appendix J. Goal of this section is to evaluate if time information can improve disaggregation of loads.

TRAINING LENGTH

Whereas a relatively short training phase is sufficient to infer power signatures, deduction of time patterns needs more time. Many events are necessary to build a time pattern graph during training phase that is similar to what is going to happen during evaluation. *Figure 3.25* compares Dishwashers in REDD 1 and UKDALE 2. A day has been divided in 48 time windows, and for each window graphs show the number of events detected. In REDD 1 during training phase activity is detected mainly between 16.30 and 22. In evaluation period events are before and after this time window, but no activity is recorded in it. If training information is used to select only events that occur in the same time windows as in training, most of the good events will be discarded, producing significant error. In UKDALE 1, on the other hand, profiles are very similar. In this case training phase can be used to improve performances. Unfortunately for this appliance event metrics are already excellent. The main difference between the two appliances is the number of events detected in training phase: they are a few dozens for REDD 1 and some hundreds for UKDALE 2. Time covered in training phase is 6 days for REDD and 20 for UKDALE. Probably the order of magnitude of a significant training is one or few months.

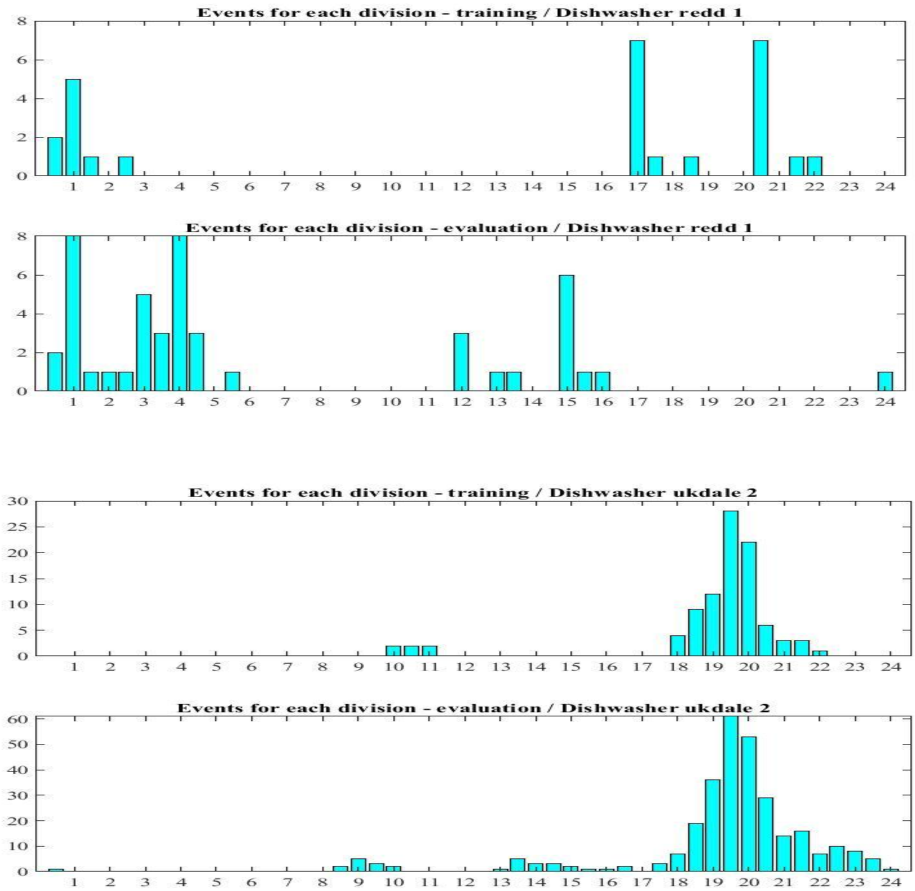


Figure 3.26: Events time patterns of Dishwashers in REDD 1 and UKDALE 2

SIMILARITIES FOR THE SAME APPLIANCE

Even if cases studied are few, no recurrent trends are evident for the same appliance in different households. In general dishwashers and microwaves should be connected to meals times. For dishwasher there are great differences. *Figure 3.26* show the case of an activity that is mainly focused in the evening. In UKDALE 1 (*Fig J.11*) activity is mainly during morning and evening time. In REDD 1 and in REDD 4 activity is mainly during night time (*Fig J.8 and J.10*).

Microwave activity is differently distributed between morning, lunch and dinner. In ECO 4 and ECO 5 major peaks are during lunch time (*Figures J.6* and *J.7*). In UKDALE 1 on the other hand lunch time is the moment of lower activity (*Fig. J.4*). In REDD 2 there's plenty of night time activity (*J.2*).

In general the indication provided by Figures *J.1 – J.17* is that habits of different households affect patterns of usages of these appliances significantly, so general trends can't be determined, at least with data in our possession.

METRICS IMPROVEMENTS

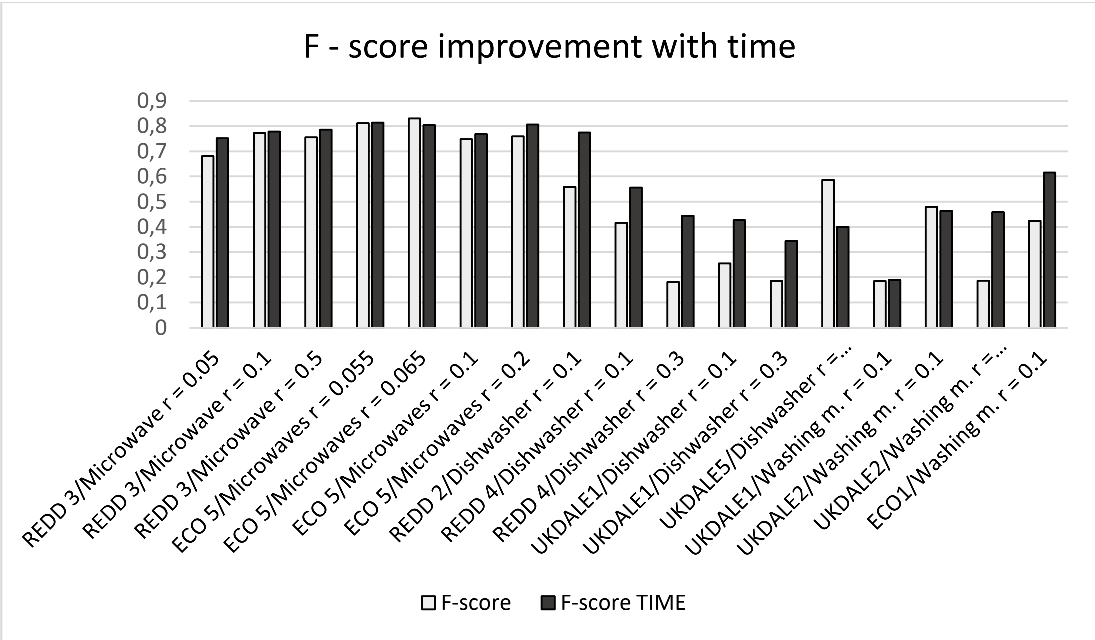
Using information provided by time analysis during training phase, number of events for each of the 48 divisions of the day is produced. A threshold is determined. All the divisions that have number of events higher than this threshold are considered allowed divisions. 1 or 2 divisions adjacent to allowed ones are considered allowed too. The algorithm is run as usually, but in the end results are refined by discarding events that don't occur during allowed divisions. Results and comparisons with metrics obtained without time information are reported in *Table 3.21.* Weiss algorithm with P/Q features has been used when available.

DATASET/HOUSE/ APPLIANCE	F-score	F-score TIME	Precision	Precision TIME	Recall	Recall TIME	Parameter r
REDD 3/Microwave	0.6803	0.7513	0.7200	0.9000	0.6447	0.6447	0.05
	0.7715	0.7780	0.7667	0.7797	0.7763	0.7763	0.1
	0.7548	0.7864	0.7231	0.7833	0.7895	0.7895	0.5
ECO 5/Microwaves	0.8111	0.8132	0.9772	0.9834	0.6932	0.6932	0.055
	0.8301	0.8038	0.9560	0.9756	0.7335	0.6834	0.065
	0.7472	0.7687	0.6730	0.7551	0.8397	0.7827	0.1
	0.7586	0.8062	0.6194	0.7246	0.9784	0.9086	0.2
REDD 2/Dishwasher	0.5581	0.7742	0.3871	0.6316	1.0000	1.0000	0.1
REDD 4/Dishwasher	0.4167	0.5556	0.4167	0.8333	0.4167	0.4167	0.1
	0.1818	0.4444	0.1111	0.4000	0.5000	0.5000	0.3
UKDALE1/Dishwasher*	0.2554	0.4264	0.1622	0.3525	0.6000	0.5393	0.1
	0.1851	0.3435	0.1066	0.2379	0.7036	0.6179	0.3
UKDALE5/Dishwasher	0.5871	0.40000	0.6774	0.2778	0.5181	0.7143	0.028
UKDALE1/Washing m.	0.1857	0.1894	0.1107	0.1134	0.5741	0.5741	0.1
UKDALE2/Washing m.	0.4793	0.4637	0.7105	0.7083	0.3616	0.3446	0.1
	0.1868	0.4587	0.1173	0.4718	0.4576	0.4463	0.25
ECO1/Washing m.	0.4238	0.6163	0.3169	0.6346	0.6396	0.5991	0.1

**These results are different compared to the ones presented in Appendix E. In this case kettle hasn't been considered, in order to get higher recall with dishwasher. Features were similar, so a lot of washing machine events were assigned to kettle.*

Table 3.21: Event results, with and without time

Microwave and dishwasher in REDD 1 are not present because of the difference between training and evaluation in time patterns. Appliances with results good already without time information are not included. With the only exception of Dishwasher in UKDALE 5, time information has increased or kept steady F-score metrics. The main effect of time is increase in precision metric. *Figure 3.27* shows that there aren't significant drops in recall. For dishwashers in REDD 4 and in UKDALE 1 and washing machines in UKDALE 2 and ECO 1 increase in performance is significant.



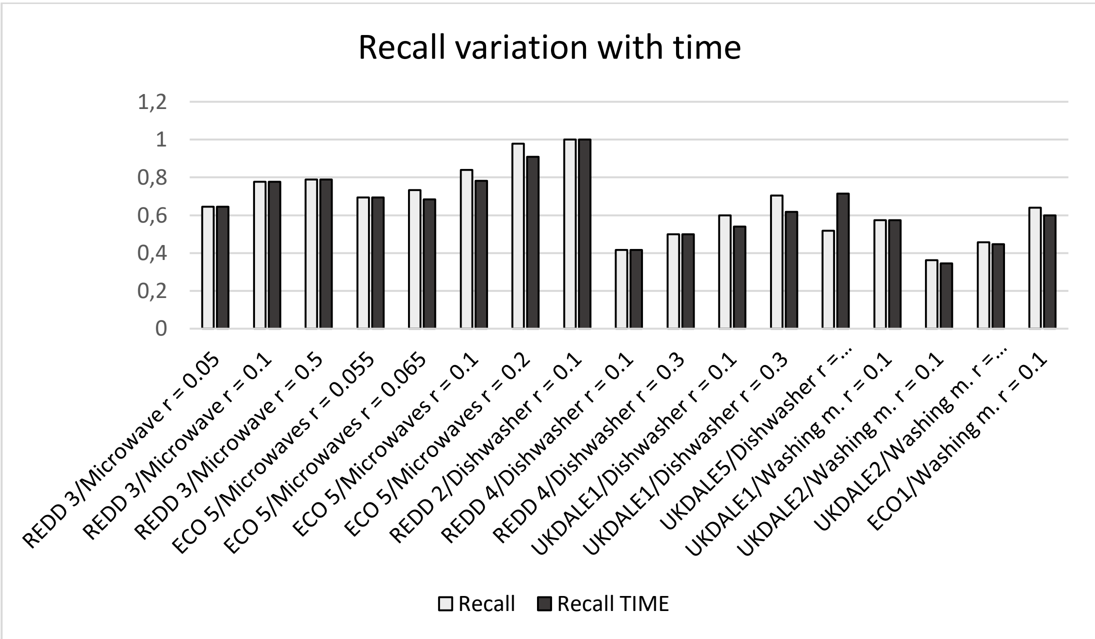
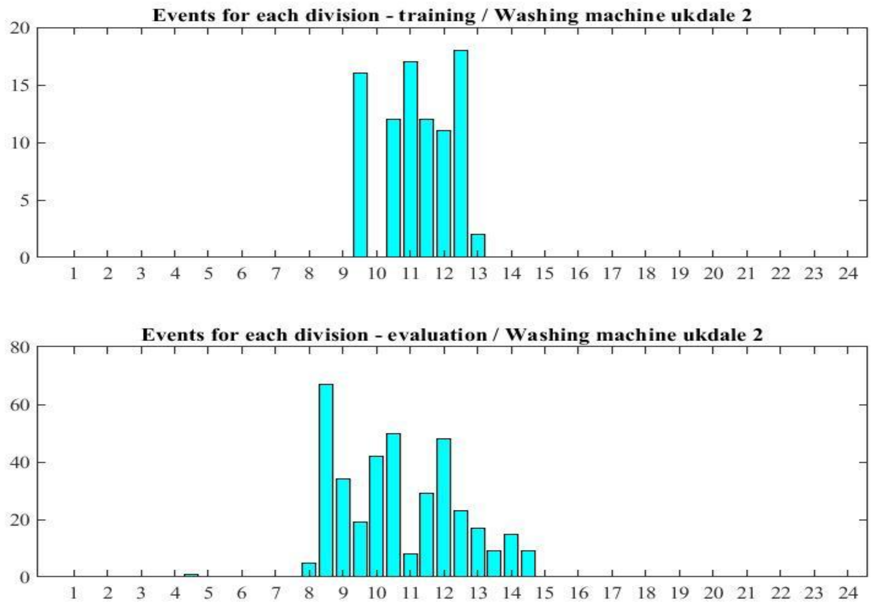


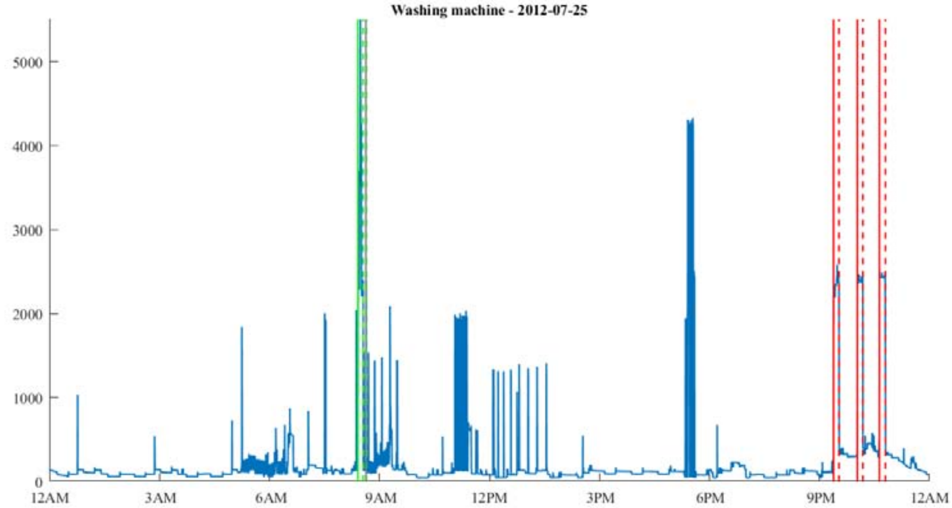
Figure 3.27: Metrics variation with time.

Washing machine in ECO 1 is a good example of how and when time pattern can increase performances significantly, because its time of usage is focused in few hours of the day (*Figure 3.28*). That allow to discard events in a broad portion of daytime. *Figure 3.29* shows how some false positives events are not present anymore after use of time information.

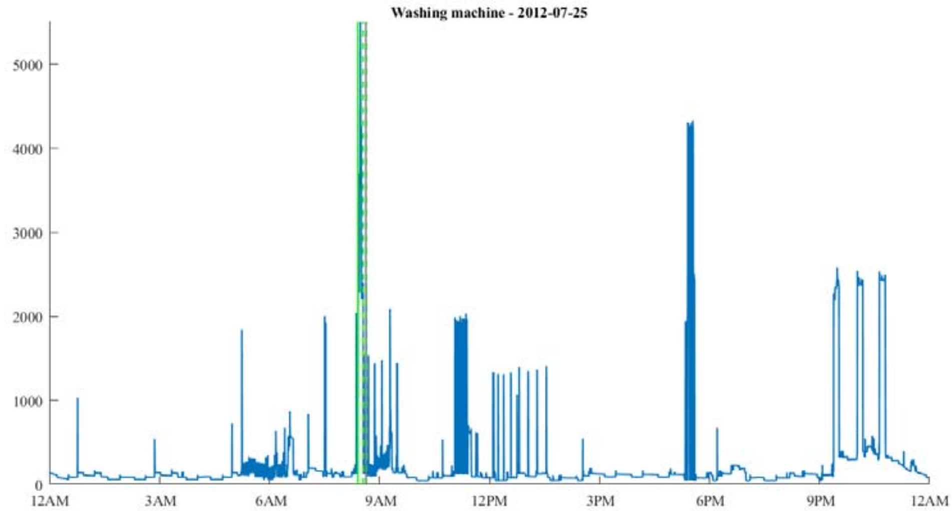


Events for each division - training / Washing machine ukdale 2
20
15
10
5
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Events for each division - evaluation / Washing machine ukdale 2
80
60
40
20
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

Figure 3.28



a) without time information



b) with time information

Figure 3.29: Effect of time pattern on disaggregation, washing machine ECO 1

3.3.8 DISCUSSION

In 3.2.1 four goals for these experiments have been listed. The discussion will follow the same structure.

1. *Understand which algorithm performs better, using different datasets and features.*

   In section 3.3.1 it is showed that from a generic point of view Weiss algorithm seems to perform better, especially for cooling appliances, where it outperforms Hart nearly in all the cases. For other appliances the difference is more subtle, and in certain cases Hart performs better.

2. *Understand which appliances among the analysed ones are well disaggregated by algorithms.*

   Energy computation has been possible with good results for cooling appliances (3.3.4), with the exception of one device only. Given the significant number of appliances analysed with very different power ratings that study gives a good indication that Weiss algorithm seems consistent in disaggregation of fridges and freezers. Appliances with high power draw are well disaggregated too with both algorithms, because their features are well separated from the others in features space. Section 3.3.3 details that. In general kettles and stoves showed good performances too (3.3.5). Other three great categories analysed are microwaves, dishwashers and washing machines.
   Better results have been achieved with microwaves, but in some cases metrics aren't good enough to hope to get good energy evaluation. Dishwasher and washing machines share inconstant results, with some peaks of excellent disaggregation and appliances with poor metrics. From a generic point of view the features of these 3 categories of appliances are in a region of the features space that seems to be crowded by many events for many houses. These events belong to different appliances, so that produces confusion in assignment. Another problem should be the simplification explained in *Figure 3.12.* For Dishwasher and washing machines only a certain category of their events have been involved in disaggregation process, as it has been done in other studies in literature ([11]). Further work has to be done on these devices, in order to create a more complex appliance model and to verify if it is usable in real world applications.

3. *Introduction of time of the day feature in load disaggregation.*

   Section 3.3.7 gives considerations about employing time patterns for mid/high power appliances. As far as we know, this is the first time that this has been done for supervised NILM work. Effectiveness of including time patterns in evaluation is specific for each appliance and dependent to behaviours of users, which can be more or less regular. The main drawback is the long training phase that is necessary to pick up enough information to produce a significant time profile. Results presented in *Table 3.21* and *Figure 3.27* show that increases of performance are possible, and only in one case metrics decrease using time. That shows that time feature can be used to boost quality of disaggregation, but its applicability has to be investigated more. In none of the cases studied the increase of event performances is good enough to applicate energy computation of good quality.

4. *Critical discussion of power disaggregation using P and eventually Q only as features.*

   Section 3.3.2 compares use of active power only and the introduction of reactive. Only for cooling appliances it can be safely said that P/Q features produce better results in general. For the other cases using P/Q precision is almost ever equal or better than using P only. P based experiments are characterized in general by better recall. Use of P only can't be excluded for further studies, especially for appliances with medium/high power rating.

4 CONCLUSIONS

4.1 SUMMARY

This work has focused on three algorithms, one unsupervised (EICCA-NILM) and two supervised (HART and WEISS). EICCA-NILM has been proposed in literature but without significant experimental validation ([15] – [17]), that have been done here (chapter 2). Extension to both real and reactive power has been done, which is a novelty for this method.

Poor results with unsupervised algorithm led us to focus on two supervised work, a well-known algorithm (3.1.1) and a relatively new one (3.1.2). They have been compared step by step, with a good number of experiments (Appendix E), and with different features availability. Evaluation methodology has been deeply explained (3.2.5), so it is clear how results have been obtained. Considerations on groups of appliances have been done, to test strengths and weaknesses of these methods. Use of time of the day information have been tested to investigate its utility in supervised NILM.

4.2 IMPLICATIONS OF RESULTS

Low resolution - steady state based features are the only features currently available from domestic smart meters. Many works in NILM field investigate the possibility to provide complete disaggregation of domestic loads using these features only. Many methods have been proposed, but experimental validation is quite limited. Sometimes it is done with few loads in a controlled environment, like in [19]. In these cases results are almost ever excellent, but not significant for a real world implementation, where problems are way more. Other times evaluation is done on public datasets, with complete arbitrary choice of what data, what households and what days to use. This makes comparisons between methods nearly impossible. Moreover algorithms described in papers are not easy to be replicated by other researchers, so to perform parallel comparisons. In last years the introduction of publicly available toolkits ([25] and [38]), with more methods already implemented, is an advantage for new researches.

4.2.1 UNSUPERVISED RESULTS

Experiments and results presented in chapter 2 point out the difficulty of EICCA - NILM to produce good results. In most cases it is not possible to detect appliances. Complete resolution of problems seems impossible. In other works unsupervised methods are criticized, like in [11], where Kolter algorithm [36] is not able to produce any correct results. This work reinforces the thesis that unsupervised algorithms, despite very appealing for no need of initial intrusive period, are very susceptible to the environment where they are supposed to work, and difficult to apply with consistent results on different houses. In specific case of EICCA-NILM, many parameters have to be determine so the setup is not easy.

This algorithm in particular rely on the assumption that events of appliances produce, in 1 dimensional P space or, as in cases studied, in 2D P/Q space, a scatter plot with distinct sets of events. These sets are supposed to represent specific appliance power states, allowing clustering and correct modelling. In experiments performed here it is clear that this happens only in certain cases. This is a general indication also for other algorithms: load events are not completely distinguishable using these features only.

4.2.2 SUPERVISED RESULTS

In chapter 3 many experiments with supervised algorithms have been presented. They have been performed on real world data, with evaluation periods up to 170 days. As far as we know Weiss algorithm has been tested before only in house 2 of ECO dataset [11]. A close comparison between the 2 algorithms is well detailed too. The result is that for cooling appliances Weiss outperforms Hart, whereas for more powerful devices results are more comparable.

Comparison between use of active power only and addition of reactive power gave us the indication that cooling appliances are usually better disaggregated with the full set of features, but for higher power appliances active power could be enough in many cases. These supervised algorithms seem to be less susceptible to the change of features compared to EICCA-NILM.

Results have been consistently good with appliances with power draw higher than 2400 W, cooling appliances, Kettles and stoves. Appliances with heating elements seem to be easier to detect. Microwaves are not always well disaggregated, and for dishwashers and washing machines results are inconsistent. These are multi-state appliances that will probably need more sophisticated treatment.

Use of time of the day of events to increase performances looks promising. Drawbacks are long training and applicability not always possible.

In general these methods are robust only on certain appliances. The core problem is the same as before: load events are not completely distinguishable using these features only.

Moreover a big limit of supervised algorithms is the great level of intrusiveness necessary for the first part, united to the presence of parameters (like *r*) that have to be determined, with great difference in results changing them.

4.3 LIMITATIONS AND FUTURE WORK

To understand better fields of applicability of methods more experiments are necessary, in order to increase the number of devices analysed for each appliance category and to introduce more high power appliances. Some examples are: electric boilers, electric showers, electric heaters, ovens, toasters, air conditioning systems, hairdryers and vacuum cleaners. Some of them are present in experiments (Appendix E), but a bigger number of devices is necessary in order to have a more general sight.

EICCA-NILM and Hart algorithms have been implemented using only indications found in relative papers. In that way their structure could be different in some parts compared to methods used by their creators.

Only Type 1 appliance models have been used for this work. To test effectiveness of methods, especially for some complex loads, Type 2 models have to be created and tested.

Data used for experiments are from datasets that have been developed in three different countries, with different wiring layouts. That makes the study general but maybe focusing on specific countries would allow to be more specific and exploit some characteristics, like multi-phase supply, to separate some load groups. Moreover appliances that are common in one country can be unused in another one.

After years of studies a reliable NILM solution hasn't been found. Research has to continue but the task seems ambitious, especially if every electric load present in one household has to be recognized. More efforts have to be made to develop cost-effective intrusive solutions that, combined with NILM algorithms, can give consistent results for wide-commercial applications.

BIBLIOGRAPHY

[1] Klemenjak, Christoph and Goldsborough, Peter. (2016). Non-Intrusive Load Monitoring: A Review and Outlook.

[2] M. Zeifman and K. Roth: "Nonintrusive Appliance Load Monitoring: Review and Outlook", in IEEE Transaction on Consumer Electronics, Vol 57, No. 1, February 2011, pp 76-84.

[3] A. Zoha, A. Gluhak, M. A. Imran and S. Rajasegarar, "Non-Intrusive Load Monitoring Approaches for Disaggregated Energy Sensing: A Survey", submitted to Sensors, November 28, 2012.

[4] R. Bonifigli, S. Squartini, M. Fagiani, and F. Piazza, "Unsupervised Algorithms for Non-Intrusive Load Monitoring: an Up-to-Date Overview", in 2015 IEEE 15th International Conference on Environment and Electrical Engineering (EEEIC), June 2015, Rome, Italy.

[5] J. Liang, S. K. K. Ng, G. Kendall and J. W. M. Cheng, "Load Signature Study-Part I: Basic Concept, Structure, and Methodology", in IEEE Transaction on Power Delivery, Vol. 25, No. 2, April 2010, pp 551-560.

[6] J. D. Kelly, "Disaggregation of Domestic Smart Meter Energy Data", University of London, Imperial College of Science, Technology and Medicine, Department of Computing

[7] J. Zico Kolter and Matthew J. Johnson. REDD: A public data set for energy disaggregation research. In proceedings of the SustKDD workshop on Data Mining Applications in Sustainability, 2011.

[8] D. Murray, L. Stankovic and V. Stankovic, "An electrical load measurement dataset of United Kingdom households from a two-year longitudinal study", in Scientific Data, 5 January 2017

[9] J. Kelly and W. Knottenbelt, "The UK-Dale dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes", in Scientific Data, 31 March 2015

[10] A. Monacchi, D. Egarter, W. Elmenreich, S. D'Alessandro and A. M. Tonello, "GREEND: An Energy Consumption Dataset of Households in Italy and Austria". Available: https://arxiv.org/abs/1405.3100

[11] C. Beckel, W. Kleiminger, R. Cicchetti, T. Staake and S. Santini, "The ECO Data Set and the Performance of Non-Intrusive Load Monitoring Algorithms", available at https://www.vs.inf.ethz.ch/publ/papers/beckel-2014- nilm.pdf

[12] H. Najmeddine, K. E. K. Drissi, C. Pasquier and Others, "State of art of load monitoring methods", 2nd IEEE International Conference on Power and Energy (PECon 08), December 1-3, 2008, Johor Baharu, Malaysia

[13] H. Frigui and R. Krishnapuram, "Clustering by competitive agglomeration", in Pattern Recogn. 30 (7), 1109–1119 (1997)

[14] Huang, C., Chung, F., Wang, S.: Generalized competitive agglomeration clustering algorithm. Int. J. Mach. Learn. Cybernet. 8(6), 1945–1969 (2015)

[15] K. M. Kamoto, Q. Liu, and X. Liu, "Unsupervised energy disaggregation of home appliances," in Proc. 3rd Int. Conf. Cloud Comput. Security, Nanjing, China, 2017, pp. 398–409, doi: 10.1007/978-3-319-68505-2_34

[16] K. M. Kamoto and Q. Liu, "Monitoring home energy usage using an unsupervised NILM algorithm based on entropy index constraints competitive agglomeration (EICCA)," in *Proc. 4th Int. Conf. Cloud Comput. Security*, Haikou, China, 2018, pp. 478–490, doi: 10.1007/978-3-030-00018-9_42

[17] K. M. Kamoto, Q. Liu, X. Liu, M Sun and N. Linge, “Low-Complexity Non-Intrusive Load Monitoring Using Unsupervised Learning and Generalized Appliance Models”, in IEEE Transactions on Consumer Electronics, Vol. 65, NO. 1, February 2019, pp 28 – 37

[18] P. Fränti and S. Sieranoja K-means properties on six clustering benchmark datasets *Applied Intelligence*, 48 (12), 4743-4759, December 2018 https://doi.org/10.1007/s10489-018-1238-7BibTex

[19] M.Weiss, A. Helfenstein, F. Mattern, and T. Staake. Leveraging smart meter data to recognize home appliances. In Proc. PerCom. IEEE, 2012

[20] J. C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York (1981)

[21] T. C. Havens, J. C. Bezdek, C. Leckie, L. O. Hall, M. Palaniswami, “Fuzzy c – Means Algorithms for Very Large Data”, in IEEE Transaction on Fuzzy Systems, Vol 20, No 6, December 2012, pp 1130-1146

[22] G. W. Hart, “Nonintrusive Appliance Load Monitoring”, in Proceedings of the IEEE, vol. 80, No 12, December 1992, pp 1870-1891

[23] G. W. Hart, E Kern, and F. Schweppe, “Nonintrusive appliance load monitor”, US Patent 4 858 141

[24] G. W. Hart, “Prototype nonintrusive appliance load monitor”, MIT Energy Laboratory Technical Report, and Electric Power Research Institute Technical Report, September 1985

[25] NILM-EVAL toolkit URL: https://github.com/beckel/nilm-eval

[26] J. liao, G. Elafoudi, L. Stankovic, V. Stankovic, “Non – intrusive appliance load monitoring using low – resolution smart meter data”, in IEEE International Conference on Smart Grid Communications, 2014, pp 535 – 540

[27] J.R. Quinlan,“Induction of decision tree,” *Machine Learning*, vol.1, no.1, pp.81-106, 1986

[28] K. Suzuki, S. Inagaki, T. Suzuki, H. Nakamura, and K. Ito, “Nonintrusive appliance load monitoring based on integer programming,” in *Proc. 47th Annu. Conf. Soc. Instrum. Control Eng. Jpn*, Tokyo, Japan, 2008, pp. 2742–2747

[29] M. Z. A. Bhotto, S. Makonin and I. V. Bajic, “Load Disaggregation Based on Aided Linear Integer Programming”, in IEEE Transaction on Circuits and Systems – II: Express briefs, Vol. 64, no 7, July 2017, pp 792 - 796

[30] F. M. Wittmann, J. C Lopez and M. J. Rider, “Nonintrusive Load Monitoring Algorithm Using Mixed-Integer Linear Programming”, in IEEE Transactions on Consumer Electronics, vol. 64, no 2, May 2018, pp 180 – 187

[31] D. H. Green, S R. Shaw, P. Lindahl, T. J. Kane, J. S. Donnal and S. B. Leeb, “A Multi-Scale Framework for Nonintrusive Load Identification”, accepted for publication in IEEE Transactions on Industrial Informatics, doi 10.1109/TII.2019.2923236

[32] V. Stankovic, J. Liao, and L. Stankovic, “A graph-based signal processing approach for low-rate energy disaggregation,” in IEEE Symposium Series on Computational Intelligence (SSCI), Orlando, FL, December 2014

[33] B. Zhao, L. Stankovic and V. Stankovic, “Blind Non – intrusive Appliance Load Monitoring using Graph-based Signal Processing”, in Symposium on Signal Processing Applications in Smart Buildings, Orlando, FL, December 2015

[34] M. A. Devlin and B. P. Hayes, "Non-Intrusive Load Monitoring and Classification of Activities of Daily Living using Residential Smart Meter Data", in IEE Transactions on Consumer Electronics, May 2019

[35] A. G. Ruzzelli, C. Nicolas, A. Schoofs, and G.M.P O'Hare, "Real-Time Recognition and Profiling of Appliances through a Single Electricity Sensor", in 7th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON), Boston, MA, June 2010, doi: 10.1109/SECON.2010.5508244

[36] J. Z. Kolter and T. Jaakkola, "Approximate inference in additive factorial HMMs with application to energy disaggregation", In Proc. AISTATS, 2012

[37] O. Parson, S. Ghosh, M. Weal, and A. Rogers, "Nonintrusive load monitoring using prior models of general appliance types", In Proc. AAAI, 2012

[38] NILMTK framework URL: https://github.com/nilmtk/nilmtk.

APPENDIX A

Two more clustering methods are here presented. They have been implemented to develop EIC-FCM and EICCA method. Symbols and notation are the same used in 2.1.

A.1 FCM ALGORITHM

Objective function for FCM (Fuzzy C Means) is the following:

$$\min J_{FCM}(U,V) = min \sum_{i=1}^{C} \sum_{j=1}^{N} {u_{i,j}}^{m} \left\| x_j - v_i \right\|^2 \quad \text{(A.1)}$$

With the constraints:

$$\sum_{i=1}^{C} (u_{ij}) = 1,\ \forall j;\ u_{ij} \in [0,1], \forall\, i,j \quad \text{(A.2)}$$

It is shown in [20] that from (2.1) and (2.2) it is possible to obtain the update equations for cluster centers and membership:

$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\left\| x_j - v_i \right\|}{\left\| x_k - v_i \right\|} \right)^{\frac{2}{m-1}}},\ i = 1, \dots, C \quad \text{(A.3)}$$

$$v_i = \frac{\sum_{j=1}^{N} {u_{ij}}^{m} x_j}{\sum_{j=1}^{N} {u_{ij}}^{m}}, i = 1, \dots, C \quad \text{(A.4)}$$

The parameter *m* is called fuzziness index, taken from [1:+∞].

- ALGORITHM DESCRIPTION

The following are the 5 steps of the algorithm.

1. Given the Dataset *X*, the number of clusters *C* and the threshold ε, initialize $V^{(0)}$ (arbitrary centres for the iteration 0)
2. Compute $U^{(0)}$ using eq. (A.3)
3. Update $V^{(k+1)}$ using eq. (A.4)
4. Update $U^{(k+1)}$ using eq. (A.3)
5. If $\|V^{(k+1)} - V^{(k)}\| < \varepsilon$, then output *U* and *V* and exit, otherwise $k = k+1$ and go back to step 3

A.2 CA ALGORITHM

Objective function for CA (Competitive Agglomeration) is the following:

$$\min J_{CA}(U,V) = min \sum_{i=1}^{C} \sum_{j=1}^{N} {u_{ij}}^2 \left\| x_j - v_i \right\|^2 - \alpha \sum_{i=1}^{C} \left[\sum_{j=1}^{N} u_{ij} \right]^2 \quad \text{(A.5)}$$

With constraints:

$$\sum_{i=1}^{C} \left(u_{ij} \right) = 1, \quad \forall j;\ u_{ij} \in [0,1], \forall\ i,j \quad \text{(A.6)}$$

It is shown in [13] that from (A.5) and (A.6) it is possible to obtain the update equations for cluster centres and membership:

$$u_{ij} = {u_{ij}}^{FCM} + u_{ij}^{Bias}, \qquad {u_{ij}}^{FCM}\ in\ eq. (A.3) \quad \text{(A.7)}$$

$$u_{ij}^{Bias} = \frac{\alpha}{\left\| x_j - v_i \right\|^2} \left(N_i - \overline{N}_j \right) \quad \text{(A.7.1)}$$

$$\overline{N}_j = \frac{\sum_{i=1}^{C} [1/\left\| x_j - v_i \right\|^2] N_i}{\sum_{i=1}^{C} [1/\left\| x_j - v_i \right\|^2]} \quad \text{(A.7.2)}$$

$$v_i = \frac{\sum_{j=1}^{N} (u_{ij})^2 x_j}{\sum_{j=1}^{N} (u_{ij})^2}, i = 1, \dots, C \quad \text{(A.8)}$$

The value of α in eq. (2.7), if too high, may give rise to non-positive u_{ij}. To avoid that α is updated dynamically using eq. (2.9), where η_0 and τ are parameters determined by user.

$$\alpha(k) = \eta_0 \exp\left(-\frac{k}{\tau}\right) \frac{\sum_{i=1}^{C} \sum_{j=1}^{N} {u_{ij}}^2 \left\| x_j - v_i \right\|^2}{\sum_{i=1}^{C} \left[\sum_{j=1}^{N} u_{ij} \right]^2} \quad \text{(A.9)}$$

N_i is defined as cardinality of the i-th cluster. It is defined as follows:

$$N_i = \sum_{j=1}^{N} u_{ij} \quad \text{(A.10)}$$

In CA algorithm, for each iteration clusters with cardinality too low are discarded, in order to converge from C_{max} (number of initial clusters) to the optimal number of clusters for the analysed dataset.

- ALGORITHM DESCRIPTION

As indicated in [13], the following are the 8 steps of the algorithm.

1. Given the Dataset X, the over specified number of clusters C_{max}, iterative threshold ε and competitive threshold ε_1, initialize $U^{(0)}$ and set the iteration number k to be 0.

2. Compute $V^{(0)}$ using eq. (A.8); Compute the cardinality N_i of every current cluster i, using eq. (A.10).

3. Update α_k using eq. (A.9).

4. Update $U^{(k+1)}$ using eq. (A.7).

5. Update N_i using eq. (A.10), if $N_i < \varepsilon_1$, then give up the cluster and its centre v_i.

6. Update the number C of clusters.

7. Update $V^{(k+1)}$ using eq. (A.8).

8. If $||V^{(k+1)} - V^{(k)}|| < \varepsilon$, then output U, V and C and exit, otherwise $k = k+1$ and go back to step 3

Initialization of Fuzzy Partition matrix (U) is performed running FCM algorithm for a specified number of iterations k_0 with number of clusters equal to C_{max}.

APPENDIX B

Here the dates of days used for experiments in chapter 2 are reported, in order to allow people to reproduce them.

Data have been organized for experiments following the guidelines presented in NILM-EVAL framework documentation (link available at [11]). The important thing to know is that for each meter and for each day a file was created containing only the 86400 readings for that day, if data are provided at one sample per second. Each missing sample is replaced by a -1. The framework recognizes missing readings so it avoids to do evaluation on them.

REDD 1	REDD 2	REDD 3
'2011-04-25' '2011-04-26' '2011-04-27' '2011-04-30' '2011-05-01' '2011-05-02' '2011-05-03' '2011-05-12' '2011-05-23'	'2011-04-23' '2011-04-24' '2011-04-25' '2011-04-26' '2011-04-27' '2011-04-28' '2011-04-29' '2011-04-30' '2011-05-01'	'2011-04-24' '2011-04-25' '2011-04-26' '2011-04-27' '2011-04-28' '2011-05-18' '2011-05-23' '2011-05-24'

Table B.1: Training days in REDD experiments

UKDALE 1	UKDALE 2	UKDALE 5
'2013-03-17' - '2013-03-30' '2013-04-02' - '2013-04-14' '2013-04-16' - '2013-04-18'	'2013-05-21' - '2013-06-09'	'2014-06-30' - '2014-07-29'

Table B.2: Training days in UKDALE experiments

UKDALE 1	UKDALE 2	UKDALE 5
'2013-04-19' - '2013-10-05'	'2013-06-10' - '2013-08-04' '2013-09-12' - '2013-10-05'	'2014-07-30' - '2014-09-06'

Table B.3: Evaluation days in UKDALE experiments

APPENDIX C

Clustering tables, as described in 2.2.4, for experiments conducted with EICCA method.

APPLIANCES TAKEN INTO ACCOUNT FOR EACH HOUSE

	REDD 1	REDD 2	REDD3	UKDALE 1	UKDALE 2	UKDALE 5
Fridge	•	•	•	•	•	•
Kettle	•			•	•	•
Microwave	•	•	•	•	•	
Dishwasher	•	•		•	•	•
Washing m.	•		•		•	
Oven	•					•
Bathroom GFI	•		•			
Cooker	•	•				
Stove		•				
Electronics*			•			
Hairdryer				•		

Table C.1: Appliances taken into account for each house, for the experiments in chapter 2

PARAMETERS SYMBOLS

r	Entropy index
η_0	Parameter used to compute α (equation 2.9)
τ	Parameter used to compute α (equation 2.9)
ε	Iterative threshold
ε_1	Competition threshold
$agtr$	Aggregation threshold percentage
C_{max}	Over-specified number of clusters
k_0	Number of iterations of EIC – FCM algorithm

Table C.2: Parameter symbols, for the experiments in chapter 2

PARAMETERS COMMON FOR ALL EXPERIMENTS

r :	0.5	ε :	$1x10^{-5}$
η_0 :	$1x10^{-6}$	C_{max} :	50
τ :	10	k_0 :	30

Table C.3: Parameter fixed for the experiments in chapter 2

REDD house 1

- EXPERIMENT 1

PARAMETERS:

$\varepsilon_1 = 10$

$agtr = 0.32$

Cluster	ΔP (W)	ΔQ (VAR)	Cardinality	Appl. 1	% of Appl. 1	Appl. 2	% of Appl. 2	Appl. 3	% of Appl. 3
1	-34	0.00	386	'///'	0.00	'///'	0.00	'///'	0.00
2	-79	0.00	111	'///'	0.00	'///'	0.00	'///'	0.00
3	27	0.00	515	'///'	0.00	'///'	0.00	'///'	0.00
4	-1048	0.00	54	'Dishwasher'	46.30	'Cooker'	31.48	'///'	0.00
5	-617	0.00	72	'///'	0.00	'///'	0.00	'///'	0.00
6	-439	0.00	86	'Fridge'	1.16	'Dishwasher'	1.16	'Washing m.'	1.16
7	-13	0.00	292	'Washing m.'	0.34	'///'	0.00	'///'	0.00
8	103	0.00	56	'Microwave'	1.79	'///'	0.00	'///'	0.00
9	228	0.00	135	'Fridge'	39.26	'Microwave'	0.74	'///'	0.00
10	-184	0.00	255	'Fridge'	52.94	'///'	0.00	'///'	0.00
11	5539	0.00	69	'Washing m.'	79.71	'Oven'	13.04	'///'	0.00
12	-5064	0.00	69	'Washing m.'	78.26	'Oven'	13.04	'///'	0.00
13	-1512	0.00	126	'Microwave'	71.43	'Bathroom g.'	11.90	'Water kettle'	2.38
14	1589	0.00	46	'Bathroom g.'	34.78	'Microwave'	19.57	'Dishwasher'	2.17
15	1089	0.00	72	'Dishwasher'	26.39	'Cooker'	23.61	'Fridge'	12.50
16	737	0.00	114	'Fridge'	73.68	'Microwave'	0.88	'Dishwasher'	0.88
17	1457	0.00	58	'Microwave'	48.28	'Water kettle'	3.45	'Dishwasher'	1.72

Table C.4: Clustering Table for REDD house 1, exp. 1

- EXPERIMENT 2

PARAMETERS:

$\varepsilon_1 = 10$

agtr = 0.29

Cluster	ΔP (W)	ΔQ (VAR)	Cardinality	Appl. 1	% of Appl. 1	Appl. 2	% of Appl. 2	Appl. 3	% of Appl. 3
1	-36	0.00	332	'///'	0.00	'///'	0.00	'///'	0.00
2	-81	0.00	109	'///'	0.00	'///'	0.00	'///'	0.00
3	15	0.00	289	'///'	0.00	'///'	0.00	'///'	0.00
4	106	0.00	45	'Microwave'	2.22	'///'	0.00	'///'	0.00
5	-1048	0.00	54	'Dishwasher'	46.30	'Cooker'	31.48	'///'	0.00
6	-617	0.00	72	'///'	0.00	'///'	0.00	'///'	0.00
7	-440	0.00	85	'Fridge'	1.18	'Dishwasher'	1.18	'Washing m.'	1.18
8	-17	0.00	333	'Washing m.'	0.30	'///'	0.00	'///'	0.00
9	36	0.00	252	'///'	0.00	'///'	0.00	'///'	0.00
10	223	0.00	128	'Fridge'	36.72	'///'	0.00	'///'	0.00
11	-184	0.00	255	'Fridge'	52.94	'///'	0.00	'///'	0.00
12	804	0.00	60	'Fridge'	76.67	'Dishwasher'	1.67	'///'	0.00
13	5540	0.00	69	'Washing m.'	79.71	'Oven'	13.04	'///'	0.00
14	-5066	0.00	69	'Washing m.'	78.26	'Oven'	13.04	'///'	0.00
15	-1512	0.00	126	'Microwave'	71.43	'Bathroom g.'	11.90	'Water kettle'	2.38
16	1587	0.00	46	'Bathroom g.'	34.78	'Microwave'	19.57	'Dishwasher'	2.17
17	1100	0.00	73	'Dishwasher'	26.03	'Cooker'	23.29	'Fridge'	12.33
18	1458	0.00	57	'Microwave'	49.12	'Water kettle'	3.51	'Dishwasher'	1.75

Table C.5: Clustering Table for REDD house 1, exp. 2

- EXPERIMENT 3

PARAMETERS:

$\varepsilon_1 = 10$

agtr = 0.25

Cluster	ΔP (W)	ΔQ (VAR)	Cardinality	Appl. 1	% of Appl. 1	Appl. 2	% of Appl. 2	Appl. 3	% of Appl. 3
1	-440	0.00	85	'Fridge'	1.18	'Dishwasher'	1.18	'Washing m.'	1.18
2	-37	0.00	332	'///'	0.00	'///'	0.00	'///'	0.00
3	-82	0.00	109	'///'	0.00	'///'	0.00	'///'	0.00
4	16	0.00	290	'///'	0.00	'///'	0.00	'///'	0.00
5	106	0.00	45	'Microwave'	2.22	'///'	0.00	'///'	0.00
6	-1049	0.00	54	'Dishwasher'	46.30	'Cooker'	31.48	'///'	0.00
7	-617	0.00	72	'///'	0.00	'///'	0.00	'///'	0.00
8	-17	0.00	333	'Washing m.'	0.30	'///'	0.00	'///'	0.00
9	37	0.00	251	'///'	0.00	'///'	0.00	'///'	0.00
10	223	0.00	128	'Fridge'	36.72	'///'	0.00	'///'	0.00
11	-185	0.00	255	'Fridge'	52.94	'///'	0.00	'///'	0.00
12	804	0.00	60	'Fridge'	76.67	'Dishwasher'	1.67	'///'	0.00
13	5541	0.00	69	'Washing m.'	79.71	'Oven'	13.04	'///'	0.00
14	-5067	0.00	69	'Washing m.'	78.26	'Oven'	13.04	'///'	0.00
15	-1512	0.00	126	'Microwave'	71.43	'Bathroom g.'	11.90	'Water kettle'	2.38
16	1101	0.00	73	'Dishwasher'	26.03	'Cooker'	23.29	'Fridge'	12.33
17	1459	0.00	57	'Microwave'	49.12	'Water kettle'	3.51	'Dishwasher'	1.75
18	643	0.00	62	'Fridge'	70.97	'Microwave'	3.23	'///'	0.00
19	1588	0.00	46	'Bathroom g.'	34.78	'Microwave'	19.57	'Dishwasher'	2.17

Table C.6: Clustering Table for REDD house 1, exp. 3

- EXPERIMENT 4

PARAMETERS:

$\varepsilon_1 = 10$

agtr = 0.2

Cluster	ΔP (W)	ΔQ (VAR)	Cardinality	Appl. 1	% of Appl. 1	Appl. 2	% of Appl. 2	Appl. 3	% of Appl. 3
1	-36	0.00	328	'///'	0.00	'///'	0.00	'///'	0.00
2	-78	0.00	109	'///'	0.00	'///'	0.00	'///'	0.00
3	11	0.00	175	'///'	0.00	'///'	0.00	'///'	0.00
4	37	0.00	160	'///'	0.00	'///'	0.00	'///'	0.00
5	-1087	0.00	33	'Cooker'	51.52	'Dishwasher'	36.36	'///'	0.00
6	-175	0.00	194	'Fridge'	69.07	'///'	0.00	'///'	0.00
7	-612	0.00	63	'///'	0.00	'///'	0.00	'///'	0.00
8	-484	0.00	50	'Dishwasher'	2.00	'Washing m.'	2.00	'///'	0.00
9	-385	0.00	43	'Fridge'	2.33	'///'	0.00	'///'	0.00
10	-17	0.00	333	'Washing m.'	0.30	'///'	0.00	'///'	0.00
11	24	0.00	176	'///'	0.00	'///'	0.00	'///'	0.00
12	212	0.00	103	'Fridge'	30.10	'///'	0.00	'///'	0.00
13	-230	0.00	64	'Fridge'	1.56	'///'	0.00	'///'	0.00
14	111	0.00	31	'Microwave'	3.23	'///'	0.00	'///'	0.00
15	-888	0.00	23	'Dishwasher'	56.52	'///'	0.00	'///'	0.00
16	297	0.00	24	'Fridge'	62.50	'///'	0.00	'///'	0.00
17	512	0.00	28	'Fridge'	50.00	'Microwave'	7.14	'///'	0.00
18	815	0.00	52	'Fridge'	78.85	'Dishwasher'	1.92	'///'	0.00
19	67	0.00	44	'///'	0.00	'///'	0.00	'///'	0.00

20	5749	0.00	20	'Washing m.'	80.00	'Oven'	5.00	'///'	0.00
21	5435	0.00	49	'Washing m.'	79.59	'Oven'	16.33	'///'	0.00
22	-5074	0.00	69	'Washing m.'	78.26	'Oven'	13.04	'///'	0.00
23	-1512	0.00	126	'Microwave'	71.43	'Bathroom g.'	11.90	'Water kettle'	2.38
24	1582	0.00	47	'Bathroom g.'	34.04	'Microwave'	19.15	'Dishwasher'	2.13
25	1101	0.00	72	'Dishwasher'	26.39	'Cooker'	23.61	'Fridge'	11.11
26	685	0.00	44	'Fridge'	84.09	'///'	0.00	'///'	0.00
27	1456	0.00	56	'Microwave'	50.00	'Water kettle'	3.57	'Dishwasher'	1.79

Table C.7: Clustering Table for REDD house 1, exp. 4

REDD house 2

- EXPERIMENT 1

PARAMETERS:

$\varepsilon_1 = 10$

agtr = 0.25

Cluster	ΔP (W)	ΔQ (VAR)	Cardinality	Appl. 1	% of Appl. 1	Appl. 2	% of Appl. 2	Appl. 3	% of Appl. 3
1	-1102	0.00	388	'Stove'	68.56	'Microwave'	10.57	'Dishwasher'	1.55
2	22	0.00	517	'///'	0.00	'///'	0.00	'///'	0.00
3	268	0.00	184	'Fridge'	29.35	'Stove'	0.54	'///'	0.00
4	1105	0.00	389	'Stove'	70.44	'Microwave'	6.17	'Fridge'	1.80
5	512	0.00	253	'Fridge'	52.57	'Stove'	0.40	'///'	0.00
6	154	0.00	281	'Fridge'	12.10	'///'	0.00	'///'	0.00
7	-246	0.00	684	'Fridge'	10.96	'///'	0.00	'///'	0.00
8	-54	0.00	774	'Fridge'	1.94	'Stove'	0.13	'///'	0.00

Table C.8: Clustering Table for REDD house 2, exp. 1

- EXPERIMENT 2

PARAMETERS:

$\varepsilon_1 = 10$

agtr = 0.15

Cluster	ΔP (W)	ΔQ (VAR)	Cardinality	Appl. 1	% of Appl. 1	Appl. 2	% of Appl. 2	Appl. 3	% of Appl. 3
1	-1088.33	0.00	383	'Stove'	69.19	'Microwave'	10.70	'Dishwasher'	1.57
2	-489.17	0.00	91	'Fridge'	2.20	'Stove'	1.10	'///'	0.00
3	-35.87	0.00	671	'Stove'	0.15	'///'	0.00	'///'	0.00
4	64.91	0.00	76	'Fridge'	1.32	'///'	0.00	'///'	0.00
5	249.87	0.00	155	'Fridge'	29.68	'///'	0.00	'///'	0.00
6	400.97	0.00	90	'Fridge'	20.00	'Stove'	1.11	'///'	0.00
7	1837.74	0.00	49	'Microwave'	44.90	'///'	0.00	'///'	0.00
8	1083.22	0.00	325	'Stove'	84.31	'Dishwasher'	1.23	'Microwave'	0.62
9	802.90	0.00	37	'Fridge'	37.84	'///'	0.00	'///'	0.00
10	485.46	0.00	119	'Fridge'	77.31	'///'	0.00	'///'	0.00
11	584.47	0.00	46	'Fridge'	52.17	'Stove'	2.17	'///'	0.00
12	179.05	0.00	139	'Fridge'	19.42	'///'	0.00	'///'	0.00

13	129.69	0.00	129	'Fridge'	4.65	'///'	0.00	'///'	0.00
14	24.57	0.00	403	'///'	0.00	'///'	0.00	'///'	0.00
15	-209.70	0.00	163	'Fridge'	9.82	'///'	0.00	'///'	0.00
16	-141.96	0.00	235	'Fridge'	14.04	'///'	0.00	'///'	0.00
17	-249.35	0.00	359	'Fridge'	10.86	'///'	0.00	'///'	0.00

Table C.9: Clustering Table for REDD house 2, exp. 2

REDD house 3

- EXPERIMENT 1

PARAMETERS:

$\varepsilon_1 = 10$

agtr = 0.2

Cluster	ΔP (W)	ΔQ (VAR)	Cardinality	Appl. 1	% of Appl. 1	Appl. 2	% of Appl. 2	Appl. 3	% of Appl. 3
1	18	0.00	305	'///'	0.00	'///'	0.00	'///'	0.00
2	85	0.00	292	'Fridge'	19.18	'///'	0.00	'///'	0.00
3	48	0.00	254	'Fridge'	0.39	'///'	0.00	'///'	0.00
4	-21	0.00	435	'Fridge'	0.23	'///'	0.00	'///'	0.00
5	-87	0.00	336	'Fridge'	30.95	'///'	0.00	'///'	0.00
6	-360	0.00	126	'Fridge'	0.79	'///'	0.00	'///'	0.00
7	-229	0.00	200	'Fridge'	3.00	'Washing m.'	0.50	'///'	0.00
8	-125	0.00	446	'Fridge'	45.52	'Bathroom g.'	0.22	'///'	0.00
9	-520	0.00	58	'Fridge'	8.62	'Microwave'	1.72	'Washing m.'	1.72
10	-52	0.00	304	'Fridge'	0.33	'///'	0.00	'///'	0.00
11	129	0.00	257	'Fridge'	15.18	'Bathroom g.'	0.39	'///'	0.00
12	187	0.00	91	'Fridge'	13.19	'///'	0.00	'///'	0.00
13	294	0.00	72	'Fridge'	56.94	'///'	0.00	'///'	0.00
14	356	0.00	84	'Fridge'	86.90	'///'	0.00	'///'	0.00
15	463	0.00	38	'Fridge'	68.42	'Bathroom g.'	2.63	'///'	0.00
16	645	0.00	39	'Fridge'	64.10	'Bathroom g.'	2.56	'///'	0.00
17	936	0.00	38	'Bathroom g.'	10.53	'Electronics'	7.89	'Fridge'	2.63
18	1135	0.00	69	'Electronics'	60.87	'Fridge'	2.90	'Microwave'	1.45
19	4817	0.00	101	'Washing m.'	97.03	'Bathroom g.'	0.99	'///'	0.00
20	5047	0.00	55	'Washing m.'	96.36	'///'	0.00	'///'	0.00
21	-915	0.00	36	'Bathroom g.'	8.33	'Electronics'	5.56	'///'	0.00
22	1773	0.00	41	'Microwave'	39.02	'Bathroom g.'	26.83	'Washing m.'	2.44
23	-1130	0.00	68	'Electronics'	63.24	'Bathroom g.'	5.88	'Microwave'	1.47
24	403	0.00	58	'Fridge'	56.90	'///'	0.00	'///'	0.00
25	-4427	0.00	156	'Washing m.'	95.51	'///'	0.00	'///'	0.00
26	-1740	0.00	50	'Microwave'	56.00	'Bathroom g.'	28.00	'Fridge'	2.00

Table C.10: Clustering Table for REDD house 3, exp. 1

- EXPERIMENT 2

PARAMETERS:

$\varepsilon_1 = 10$

agtr = 0.22

Cluster	ΔP (W)	ΔQ (VAR)	Cardinality	Appl. 1	% of Appl. 1	Appl. 2	% of Appl. 2	Appl. 3	% of Appl. 3
1	-84	0.00	1615	'Fridge'	19.32	'Washing m.'	0.06	'Bathroom g.'	0.06
2	394	0.00	300	'Fridge'	67.00	'Bathroom g.'	0.67	'///'	0.00
3	73	0.00	1203	'Fridge'	8.73	'Bathroom g.'	0.08	'///'	0.00
4	-373	0.00	290	'Fridge'	3.10	'Microwave'	0.34	'Washing m.'	0.34
5	-1352	0.00	144	'Electronics'	31.25	'Microwave'	20.14	'Bathroom g.'	13.89
6	1241	0.00	145	'Electronics'	31.03	'Microwave'	12.41	'Bathroom g.'	10.34
7	4903	0.00	156	'Washing m.'	96.79	'Bathroom g.'	0.64	'///'	0.00
8	-4423	0.00	156	'Washing m.'	95.51	'///'	0.00	'///'	0.00

Table C.11: Clustering Table for REDD house 3, exp. 2

UKDALE house 1 – active power only

- EXPERIMENT 1

PARAMETERS:

$\varepsilon_1 = 40$

agtr = 0.35

Cluster	ΔP (W)	ΔQ (VAR)	Cardinality	Appl. 1	% of Appl. 1	Appl. 2	% of Appl. 2	Appl. 3	% of Appl. 3
1	-383	0.00	8928	'Microwave'	2.35	'Water kettle'	1.65	'Fridge'	0.55
2	407	0.00	8908	'Fridge'	6.56	'Microwave'	2.84	'Water kettle'	1.67
3	-7	0.00	11924	'Microwave'	0.41	'Fridge'	0.18	'Hairdryer'	0.10

Table C.12: Clustering Table for UKDALE house 1, active power only, exp. 1

UKDALE house 1 – active and reactive power

- EXPERIMENT 1

PARAMETERS:

$\varepsilon_1 = 40$

agtr = 0.35

Cluster	ΔP (W)	ΔQ (VAR)	Cardinality	Appl. 1	% of Appl. 1	Appl. 2	% of Appl. 2	Appl. 3	% of Appl. 3
1	-17	-7	2663	'Hairdryer'	0.11	'Fridge'	0.08	'Water kettle'	0.08
2	22	8	4505	'Fridge'	0.18	'Microwave'	0.09	'Water kettle'	0.02

3	158	271	819	'Fridge'	0.49	'Microwave'	0.12	'///'	0.00
4	275	468	775	'Fridge'	4.26	'Microwave'	0.26	'Water kettle'	0.13
5	379	935	748	'Fridge'	0.40	'Microwave'	0.13	'///'	0.00
6	552	1231	470	'Fridge'	0.64	'///'	0.00	'///'	0.00
7	333	799	784	'Fridge'	0.26	'///'	0.00	'///'	0.00
8	377	689	1031	'Fridge'	0.97	'Washing m.'	0.39	'Microwave'	0.10
9	1701	380	889	'Microwave'	27.11	'Water kettle'	16.09	'Hairdryer'	3.60
10	340	575	1135	'Fridge'	13.30	'Microwave'	0.09	'Water kettle'	0.09
11	409	347	998	'Fridge'	35.67	'Washing m.'	0.40	'Microwave'	0.20
12	365	231	713	'Fridge'	2.52	'Washing m.'	0.56	'Water kettle'	0.14
13	328	99	1006	'Microwave'	0.40	'Fridge'	0.20	'Water kettle'	0.20
14	76	30	1173	'Fridge'	0.60	'Hairdryer'	0.26	'Microwave'	0.17
15	-86	-6	1322	'Microwave'	0.45	'Water kettle'	0.08	'Hairdryer'	0.08
16	-139	-254	905	'Microwave'	3.20	'Hairdryer'	0.33	'///'	0.00
17	-280	-471	1009	'Microwave'	0.50	'Washing m.'	0.20	'Hairdryer'	0.10
18	-317	-565	1110	'Fridge'	0.36	'Hairdryer'	0.27	'Microwave'	0.18
19	-368	-673	921	'Microwave'	0.54	'Water kettle'	0.11	'///'	0.00
20	-391	-921	671	'Water kettle'	0.30	'///'	0.00	'///'	0.00
21	-322	-781	800	'Water kettle'	0.25	'Microwave'	0.13	'Washing m.'	0.13
22	-553	-1230	443	'///'	0.00	'///'	0.00	'///'	0.00
23	-322	-345	917	'Fridge'	0.65	'Microwave'	0.11	'Water kettle'	0.11
24	-378	-236	810	'Microwave'	0.37	'Water kettle'	0.25	'Washing m.'	0.25
25	-43	-11	1746	'Microwave'	0.52	'Hairdryer'	0.11	'///'	0.00
26	-326	-92	1037	'Fridge'	3.76	'Microwave'	0.39	'Water kettle'	0.10
27	-1692	-342	864	'Microwave'	21.64	'Water kettle'	15.86	'Hairdryer'	2.78

Table C.13: Clustering Table for UKDALE house 1, active and reactive power, exp. 1

- EXPERIMENT 2

PARAMETERS:

$\varepsilon_1 = 40$

agtr = 0.3

Cluster	ΔP (W)	ΔQ (VAR)	Cardinality	Appl. 1	% of Appl. 1	Appl. 2	% of Appl. 2	Appl. 3	% of Appl. 3
1	-15	-6	2416	'Hairdryer'	0.12	'Fridge'	0.08	'Water kettle'	0.08
2	22	8	4409	'Fridge'	0.16	'Microwave'	0.09	'Water kettle'	0.02
3	147	261	707	'Fridge'	0.57	'Microwave'	0.14	'Hairdryer'	0.14
4	248	431	664	'Fridge'	2.26	'Microwave'	0.30	'Washing m.'	0.15
5	302	752	667	'Fridge'	0.15	'///'	0.00	'///'	0.00
6	343	931	549	'Fridge'	0.18	'///'	0.00	'///'	0.00
7	582	1081	288	'Fridge'	1.74	'Microwave'	0.35	'///'	0.00
8	424	855	444	'Fridge'	1.35	'///'	0.00	'///'	0.00
9	381	681	878	'Fridge'	0.57	'Washing m.'	0.46	'Microwave'	0.11
10	305	582	818	'Fridge'	0.24	'Water kettle'	0.24	'Washing m.'	0.12
11	527	1299	257	'///'	0.00	'///'	0.00	'///'	0.00
12	1600	504	348	'Microwave'	62.07	'Hairdryer'	0.57	'Water kettle'	0.29
13	2394	278	296	'Water kettle'	47.64	'Dishwasher'	4.05	'Washing m.'	4.05
14	1648	155	250	'Hairdryer'	11.60	'Microwave'	9.60	'Dishwasher'	0.80
15	378	513	613	'Fridge'	32.30	'Microwave'	0.33	'///'	0.00
16	410	340	882	'Fridge'	36.96	'Washing m.'	0.57	'Hairdryer'	0.11
17	380	95	463	'Fridge'	0.22	'Microwave'	0.22	'///'	0.00
18	358	236	706	'Fridge'	2.41	'Washing m.'	0.42	'Water kettle'	0.14
19	280	95	614	'Microwave'	0.49	'Water kettle'	0.33	'Hairdryer'	0.33
20	70	25	1182	'Fridge'	0.68	'Microwave'	0.17	'Hairdryer'	0.08
21	-83	-2	1116	'Microwave'	0.18	'Water kettle'	0.09	'///'	0.00
22	-63	-117	593	'Microwave'	3.37	'Hairdryer'	0.34	'///'	0.00
23	-129	-255	580	'Microwave'	3.10	'Hairdryer'	0.52	'///'	0.00
24	-215	-377	466	'Microwave'	0.86	'Hairdryer'	0.21	'///'	0.00
25	-284	-467	663	'Washing m.'	0.30	'Microwave'	0.15	'///'	0.00

26	-350	-606	636	'Microwave'	0.47	'Dishwasher'	0.16	'Hairdryer'	0.16
27	-385	-699	583	'Microwave'	0.51	'Water kettle'	0.17	'///'	0.00
28	-302	-535	797	'Fridge'	0.50	'Hairdryer'	0.38	'Microwave'	0.25
29	-275	-699	547	'Washing m.'	0.18	'///'	0.00	'///'	0.00
30	-319	-819	455	'Water kettle'	0.66	'///'	0.00	'///'	0.00
31	-345	-955	304	'///'	0.00	'///'	0.00	'///'	0.00
32	-440	-873	323	'Water kettle'	0.31	'///'	0.00	'///'	0.00
33	-320	-344	613	'Fridge'	0.98	'Microwave'	0.16	'Water kettle'	0.16
34	-411	-303	397	'Microwave'	0.76	'Washing m.'	0.50	'///'	0.00
35	-357	-223	559	'Water kettle'	0.36	'Fridge'	0.18	'///'	0.00
36	-315	-113	431	'Microwave'	0.46	'Water kettle'	0.23	'///'	0.00
37	-39	-7	1679	'Microwave'	0.18	'Hairdryer'	0.06	'///'	0.00
38	-249	-55	379	'Fridge'	10.03	'Microwave'	0.53	'///'	0.00
39	-398	-63	341	'Fridge'	0.29	'///'	0.00	'///'	0.00
40	-1628	-167	288	'Hairdryer'	8.33	'Microwave'	6.25	'Dishwasher'	0.69
41	-2346	-274	276	'Water kettle'	48.91	'Dishwasher'	7.61	'Washing m.'	1.09
42	-1591	-493	302	'Microwave'	55.96	'Water kettle'	0.66	'///'	0.00
43	-528	-1303	237	'///'	0.00	'///'	0.00	'///'	0.00
44	-593	-1086	248	'///'	0.00	'///'	0.00	'///'	0.00

Table C.14: Clustering Table for UKDALE house 1, active and reactive power, exp. 2

UKDALE house 2 – active power only

- EXPERIMENT 1

PARAMETERS:

$\varepsilon_1 = 20$

agtr = 0.15 (same results for 0.18 and 0.22)

Cluster	ΔP (W)	ΔQ (VAR)	Cardinality	Appl. 1	% of Appl. 1	Appl. 2	% of Appl. 2	Appl. 3	% of Appl. 3
1	-90	0	7665	'Fridge'	9.38	'Microwave'	1.46	'Water kettle'	1.34
2	575	0	1073	'Fridge'	64.31	'Water kettle'	9.13	'Microwave'	7.83

Table C.15: Clustering Table for UKDALE house 2, active power only, exp. 1

- EXPERIMENT 2

PARAMETERS:

$\varepsilon_1 = 20$

agtr = 0.1

Cluster	ΔP (W)	ΔQ (VAR)	Cardinality	Appl. 1	% of Appl. 1	Appl. 2	% of Appl. 2	Appl. 3	% of Appl. 3
1	-442	0	902	'Microwave'	10.64	'Fridge'	0.11	'Water kettle'	0.11
2	-2844	0	161	'Water kettle'	60.25	'Dishwasher'	15.53	'Washing m.'	3.73
3	155	0	1330	'Fridge'	0.90	'Washing m.'	0.38	'Microwave'	0.15
4	13	0	2377	'Microwave'	0.42	'Fridge'	0.25	'Water kettle'	0.13
5	-192	0	583	'Fridge'	1.37	'Washing m.'	0.51	'Microwave'	0.34
6	487	0	269	'Fridge'	89.96	'Microwave'	0.37	'Water kettle'	0.37
7	2970	0	97	'Water kettle'	97.94	'///'	0.00	'///'	0.00
8	2047	0	70	'Dishwasher'	57.14	'Washing m.'	18.57	'Water kettle'	2.86
9	1278	0	133	'Microwave'	61.65	'Fridge'	0.75	'///'	0.00
10	518	0	482	'Fridge'	92.53	'///'	0.00	'///'	0.00
11	-77	0	2334	'Fridge'	29.69	'Washing m.'	0.39	'Microwave'	0.09

Table C.16: Clustering Table for UKDALE house 2, active power only, exp. 2

UKDALE house 2 – active and reactive power

- EXPERIMENT 1

PARAMETERS:

$\varepsilon_1 = 20$

agtr = 0.22

Cluster	ΔP (W)	ΔQ (VAR)	Cardinality	Appl. 1	% of Appl. 1	Appl. 2	% of Appl. 2	Appl. 3	% of Appl. 3
1	-108	-28	233	'Washing machine'	0.43	'///'	0.00	'///'	0.00
2	-49	-14	441	'Water kettle'	0.23	'///'	0.00	'///'	0.00
3	19	24	595	'///'	0.00	'///'	0.00	'///'	0.00
4	-104	-5	287	'Microwave'	0.35	'///'	0.00	'///'	0.00
5	-74	19	742	'Microwave'	0.13	'///'	0.00	'///'	0.00
6	163	426	154	'Fridge'	1.30	'///'	0.00	'///'	0.00
7	157	511	145	'Fridge'	1.38	'Washing m.'	1.38	'Microwave'	0.69
8	503	386	193	'Fridge'	98.96	'///'	0.00	'///'	0.00
9	504	426	187	'Fridge'	88.24	'Water kettle'	0.53	'///'	0.00
10	188	575	137	'Washing m.'	0.73	'///'	0.00	'///'	0.00
11	215	672	78	'Fridge'	1.28	'///'	0.00	'///'	0.00
12	122	366	150	'Fridge'	1.33	'Washing m.'	1.33	'///'	0.00
13	-48	9	297	'Microwave'	0.34	'Dishwasher'	0.34	'Water kettle'	0.34
14	54	7	277	'Microwave'	0.36	'Water kettle'	0.36	'///'	0.00
15	-18	-22	662	'Microwave'	0.15	'Water kettle'	0.15	'///'	0.00
16	10	10	669	'Fridge'	0.45	'Microwave'	0.15	'///'	0.00
17	501	341	132	'Fridge'	99.24	'///'	0.00	'///'	0.00
18	498	274	110	'Fridge'	99.09	'///'	0.00	'///'	0.00
19	88	254	151	'Fridge'	3.31	'Washing m.'	2.65	'Microwave'	0.66
20	156	63	103	'Fridge'	0.97	'///'	0.00	'///'	0.00
21	157	24	123	'///'	0.00	'///'	0.00	'///'	0.00
22	-11	-9	965	'Microwave'	0.31	'Dishwasher'	0.10	'///'	0.00
23	-1899	-54	59	'Dishwasher'	64.41	'Washing m.'	6.78	'Water kettle'	1.69
24	-1291	-522	121	'Microwave'	80.99	'Water kettle'	0.83	'///'	0.00
25	-2944	-102	97	'Water kettle'	98.97	'///'	0.00	'///'	0.00
26	-413	-364	129	'///'	0.00	'///'	0.00	'///'	0.00
27	-135	-399	171	'Washing m.'	1.17	'Microwave'	0.58	'///'	0.00
28	-87	-221	174	'Washing m.'	1.72	'///'	0.00	'///'	0.00
29	-429	-232	118	'Microwave'	1.69	'///'	0.00	'///'	0.00
30	153	-17	113	'Fridge'	0.88	'Microwave'	0.88	'///'	0.00
31	155	9	141	'///'	0.00	'///'	0.00	'///'	0.00
32	518	205	104	'Fridge'	86.54	'///'	0.00	'///'	0.00
33	2049	97	65	'Dishwasher'	61.54	'Washing m.'	16.92	'Water kettle'	3.08
34	2972	101	97	'Water kettle'	97.94	'///'	0.00	'///'	0.00
35	1321	568	108	'Microwave'	84.26	'Fridge'	0.93	'///'	0.00
36	962	24	47	'Fridge'	2.13	'///'	0.00	'///'	0.00
37	28	3	608	'Fridge'	0.33	'Microwave'	0.16	'Water kettle'	0.16
38	-125	-299	110	'Washing m.'	2.73	'///'	0.00	'///'	0.00
39	-960	-27	47	'///'	0.00	'///'	0.00	'///'	0.00
40	-210	-658	57	'///'	0.00	'///'	0.00	'///'	0.00
41	-415	-297	114	'Microwave'	0.88	'///'	0.00	'///'	0.00
42	-414	-444	186	'///'	0.00	'///'	0.00	'///'	0.00
43	-178	-551	161	'Microwave'	0.62	'Washing m.'	0.62	'///'	0.00
44	-171	-454	145	'Dishwasher'	0.69	'Washing m.'	0.69	'///'	0.00
45	-414	-407	182	'///'	0.00	'///'	0.00	'///'	0.00

Table C.17: Clustering Table for UKDALE house 2, active and reactive power, exp. 1

- EXPERIMENT 2

PARAMETERS:

$\varepsilon_1 = 20$

agtr = 0.28

Cluster	ΔP (W)	ΔQ (VAR)	Cardinality	Appl. 1	% of Appl. 1	Appl. 2	% of Appl. 2	Appl. 3	% of Appl. 3
1	-108	-28	219	'///'	0.00	'///'	0.00	'///'	0.00
2	-49	-36	248	'Washing m.'	0.81	'Microwave'	0.40	'///'	0.00
3	19	24	598	'///'	0.00	'///'	0.00	'///'	0.00
4	-104	-5	297	'Microwave'	0.34	'///'	0.00	'///'	0.00
5	-74	19	817	'Microwave'	0.12	'///'	0.00	'///'	0.00
6	131	385	183	'Fridge'	1.09	'Washing m.'	0.55	'///'	0.00
7	167	458	160	'Fridge'	2.50	'Washing m.'	1.25	'Microwave'	0.63
8	504	383	206	'Fridge'	99.03	'///'	0.00	'///'	0.00
9	505	425	194	'Fridge'	88.66	'Water kettle'	0.52	'///'	0.00
10	172	550	194	'Washing m.'	0.52	'///'	0.00	'///'	0.00
11	209	649	103	'Fridge'	0.97	'///'	0.00	'///'	0.00
12	94	266	173	'Fridge'	3.47	'Washing m.'	2.89	'Microwave'	0.58
13	-47	-5	539	'Water kettle'	0.37	'Microwave'	0.19	'Dishwasher'	0.19
14	53	7	278	'Microwave'	0.36	'Water kettle'	0.36	'///'	0.00
15	-17	-20	664	'Water kettle'	0.15	'///'	0.00	'///'	0.00
16	10	10	666	'Fridge'	0.45	'Microwave'	0.15	'///'	0.00
17	502	335	130	'Fridge'	100.0	'///'	0.00	'///'	0.00
18	505	266	101	'Fridge'	98.02	'///'	0.00	'///'	0.00
19	156	64	100	'Fridge'	1.00	'///'	0.00	'///'	0.00
20	157	25	125	'///'	0.00	'///'	0.00	'///'	0.00
21	-12	-8	876	'Microwave'	0.34	'Dishwasher'	0.11	'///'	0.00
22	-1899	-54	59	'Dishwasher'	64.41	'Washing m.'	6.78	'Water kettle'	1.69
23	-1289	-518	129	'Microwave'	75.97	'Water kettle'	0.78	'///'	0.00
24	-2943	-102	97	'Water kettle'	98.97	'///'	0.00	'///'	0.00
25	-411	-338	115	'///'	0.00	'///'	0.00	'///'	0.00
26	-167	-520	172	'Microwave'	0.58	'Dishwasher'	0.58	'Washing m.'	0.58
27	-104	-256	229	'Washing m.'	0.87	'///'	0.00	'///'	0.00
28	-431	-221	126	'///'	0.00	'///'	0.00	'///'	0.00
29	153	-17	113	'Fridge'	0.88	'Microwave'	0.88	'///'	0.00
30	155	9	142	'///'	0.00	'///'	0.00	'///'	0.00
31	521	201	136	'Fridge'	59.56	'///'	0.00	'///'	0.00
32	2049	100	65	'Dishwasher'	61.54	'Washing m.'	16.92	'Water kettle'	3.08
33	2972	101	97	'Water kettle'	97.94	'///'	0.00	'///'	0.00
34	1319	565	113	'Microwave'	80.53	'Fridge'	0.88	'///'	0.00
35	28	3	609	'Fridge'	0.33	'Microwave'	0.16	'Water kettle'	0.16
36	-147	-404	272	'Washing m.'	2.21	'Microwave'	0.37	'///'	0.00
37	-429	-272	98	'Microwave'	3.06	'///'	0.00	'///'	0.00
38	-414	-440	220	'///'	0.00	'///'	0.00	'///'	0.00
39	-202	-612	110	'///'	0.00	'///'	0.00	'///'	0.00
40	-414	-396	212	'///'	0.00	'///'	0.00	'///'	0.00

Table C.18: Clustering Table for UKDALE house 2, active and reactive power, exp. 2

UKDALE house 5 – active and reactive power

- EXPERIMENT 1

PARAMETERS:

$\varepsilon_1 = 25$

agtr = 0.15

Cluster	ΔP (W)	ΔQ (VAR)	Cardinality	Appl. 1	% of Appl. 1	Appl. 2	% of Appl. 2	Appl. 3	% of Appl. 3
1	-32	-9	6092	'Fridge'	0.11	'Water kettle'	0.05	'Dishwasher'	0.02
2	26	7	4631	'Dishwasher'	0.02	'Oven'	0.02	'///'	0.00
3	-53	73	1401	'Fridge'	0.29	'Water kettle'	0.07	'///'	0.00
4	183	65	524	'Fridge'	0.19	'///'	0.00	'///'	0.00
5	247	22	693	'Oven'	0.29	'///'	0.00	'///'	0.00
6	215	123	484	'Fridge'	0.41	'///'	0.00	'///'	0.00
7	385	291	337	'Fridge'	80.71	'///'	0.00	'///'	0.00
8	2091	177	243	'Oven'	45.68	'Water kettle'	0.82	'Fridge'	0.41
9	1683	129	197	'Dishwasher'	10.15	'Water kettle'	1.02	'Oven'	1.02
10	2772	262	190	'Water kettle'	34.21	'Oven'	3.68	'Fridge'	0.53
11	1567	109	296	'Oven'	0.34	'///'	0.00	'///'	0.00
12	1089	82	437	'Oven'	1.83	'Dishwasher'	0.69	'Fridge'	0.23
13	797	69	451	'Fridge'	1.11	'Dishwasher'	1.11	'Oven'	0.44
14	415	175	577	'Fridge'	68.11	'Water kettle'	0.17	'Oven'	0.17
15	268	188	297	'Fridge'	1.35	'///'	0.00	'///'	0.00
16	139	11	839	'Fridge'	0.48	'Dishwasher'	0.12	'///'	0.00
17	70	-94	841	'Water kettle'	0.24	'Fridge'	0.12	'///'	0.00
18	47	12	2595	'Fridge'	0.04	'///'	0.00	'///'	0.00
19	11	3	5756	'Fridge'	0.02	'///'	0.00	'///'	0.00
20	-16	-5	6438	'Fridge'	0.05	'Oven'	0.03	'Water kettle'	0.02
21	-51	-13	2410	'Fridge'	0.50	'Water kettle'	0.04	'///'	0.00
22	-99	30	1117	'Fridge'	56.58	'Water kettle'	0.18	'///'	0.00
23	-155	-18	948	'Fridge'	0.42	'Water kettle'	0.11	'///'	0.00
24	-237	-23	921	'Fridge'	0.65	'Dishwasher'	0.22	'Oven'	0.11
25	-212	-119	778	'Fridge'	0.26	'///'	0.00	'///'	0.00
26	-268	-316	350	'Fridge'	0.29	'///'	0.00	'///'	0.00
27	-285	-205	771	'Fridge'	0.39	'///'	0.00	'///'	0.00
28	-797	-69	521	'Dishwasher'	0.58	'Water kettle'	0.19	'Oven'	0.19
29	-1088	-85	439	'Oven'	0.91	'Water kettle'	0.46	'Fridge'	0.23
30	-1560	-109	450	'Dishwasher'	4.44	'Fridge'	0.44	'Oven'	0.44
31	-2042	-176	238	'Oven'	41.18	'Dishwasher'	2.52	'Fridge'	0.42
32	-2780	-263	195	'Water kettle'	30.77	'Oven'	4.62	'///'	0.00

Table C.19: Clustering Table for UKDALE house 5, active and reactive power, exp. 1

UKDALE house 1 – active and reactive power, experiment 2

Graphic representation of clusters in P/Q plane, x axis in Watt, y axis in VAR.

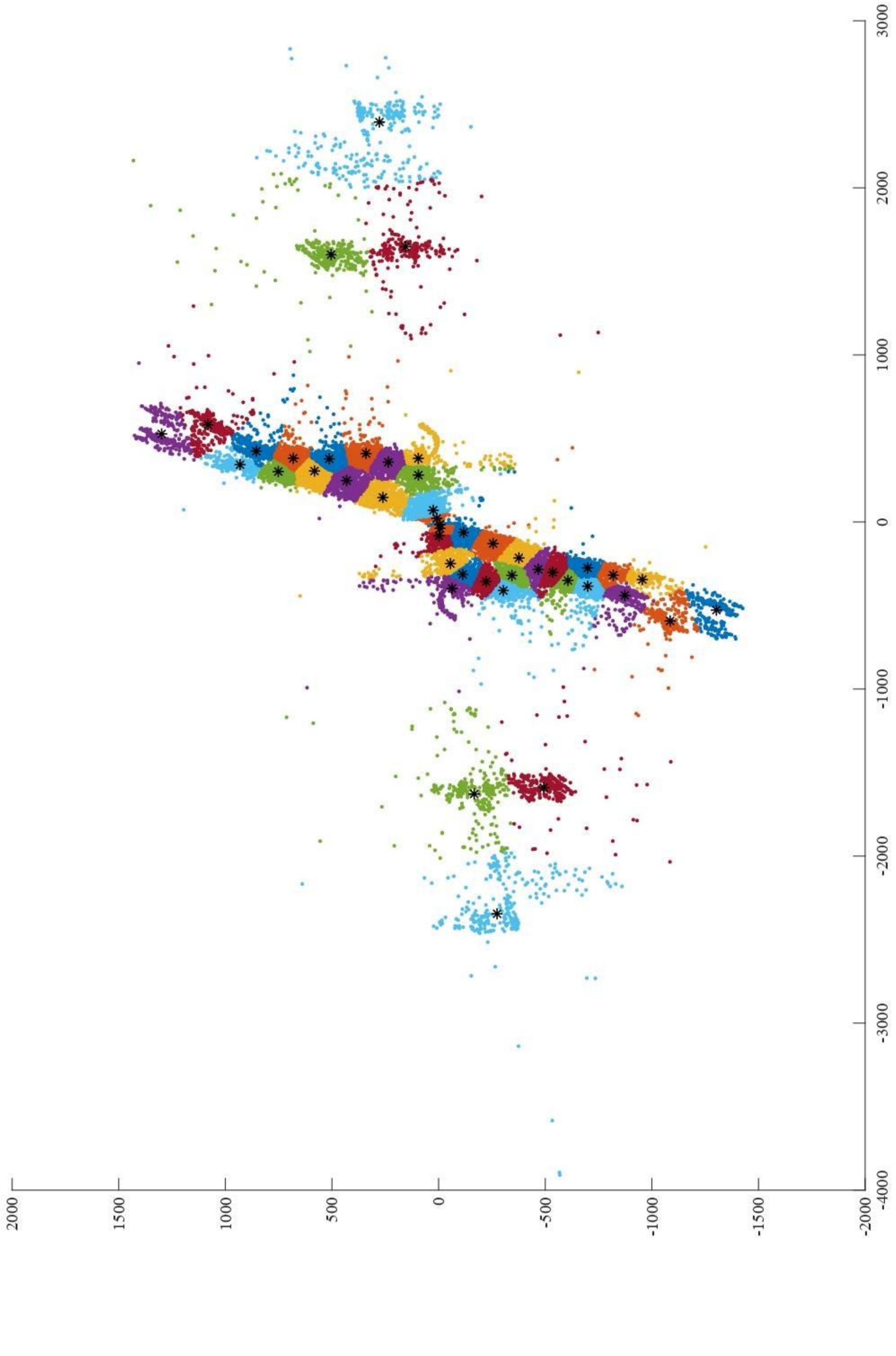


Figure C.1: Graphic representation of clusters *for UKDALE house 1, experiment 2*

UKDALE house 2 – active and reactive power, experiment 1

Graphic representation of clusters in P/Q plane, x axis in Watt, y axis in VAR.

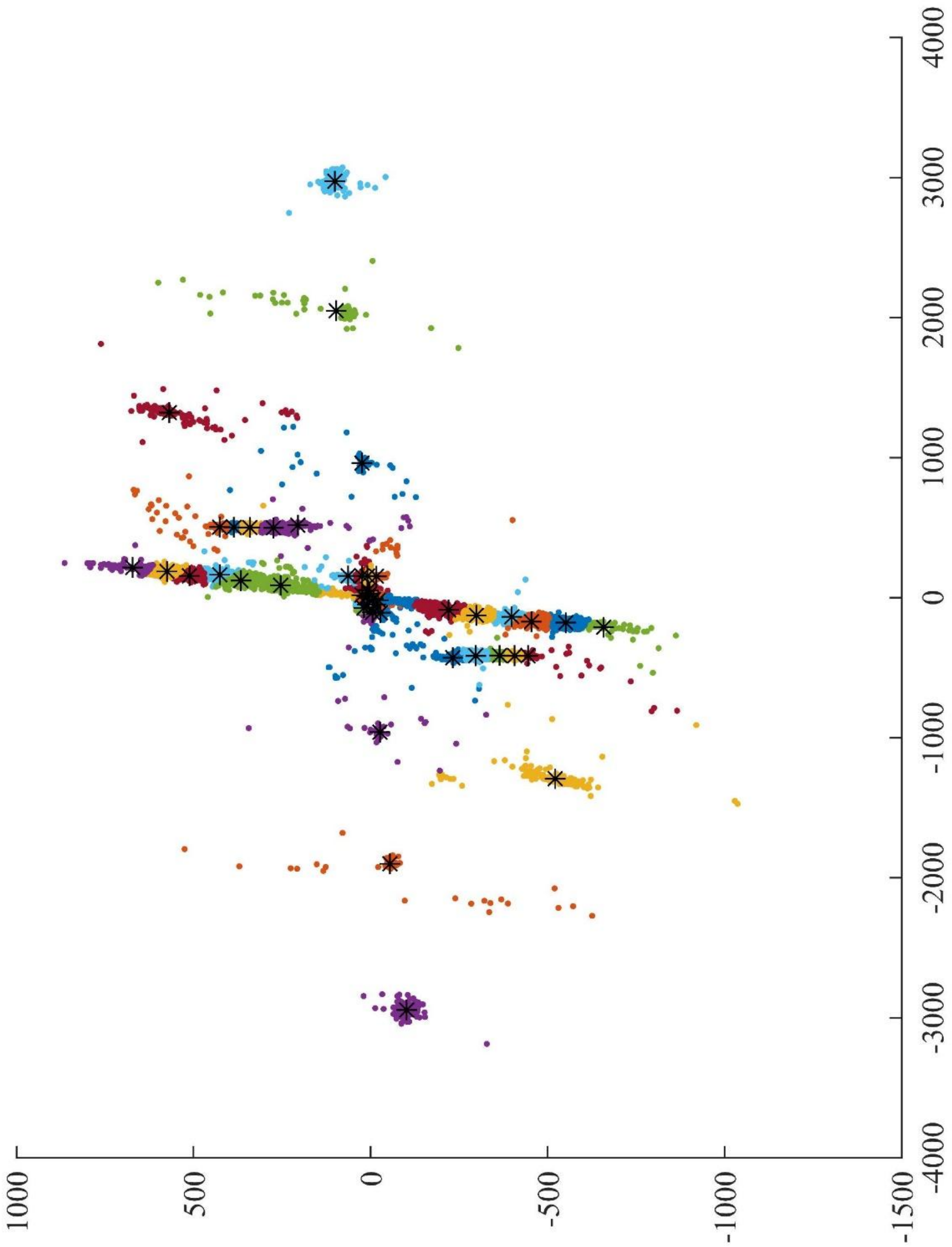


Figure C.2: Graphic representation of clusters *for UKDALE house 2, experiment 1*

UKDALE house 5 – active and reactive power, experiment 1

Graphic representation of clusters in P/Q plane, x axis in Watt, y axis in VAR.

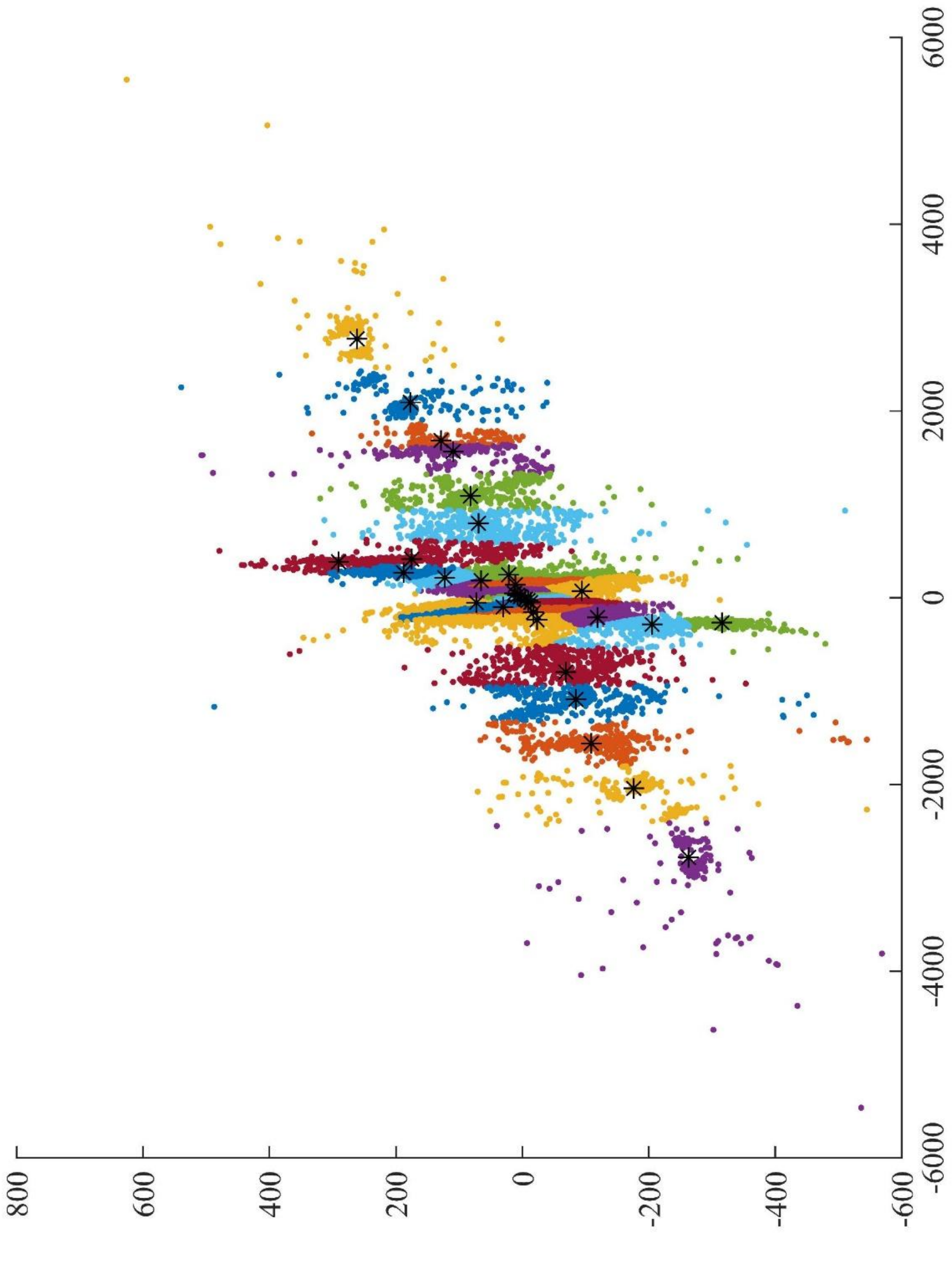


Figure C.3: Graphic representation of clusters *for UKDALE house 5, experiment 1*

APPENDIX D

Here dates of the days used for the experiments in chapter 3 are reported, in order to allow eventual reproduction.

Data have been organized for experiments following the guidelines presented in NILM-EVAL framework documentation (link available at [11]). The important thing to know is that for each meter and for each day a file was created containing only the 86400 readings for that day, if data are provided at one sample per second. Each missing sample is replaced by a -1. The framework recognizes missing readings so it avoids to do evaluation on them.

REDD 1	REDD 2	REDD 3	REDD 4
'2011-04-18'	'2011-04-18'	'2011-04-18'	'2011-04-21'
'2011-04-19'	'2011-04-19'	'2011-04-20'	'2011-04-22'
'2011-04-20'	'2011-04-20'	'2011-04-21'	'2011-04-23'
'2011-04-22'	'2011-04-21'	'2011-04-22'	'2011-04-24'
'2011-04-23'	'2011-04-22'	'2011-04-23'	'2011-04-26'
'2011-04-24'			

Table B.1: Training days in REDD experiments

REDD 1	REDD 2	REDD 3	
'2011-04-25'	'2011-04-23'	'2011-04-24'	'2011-04-27'
'2011-04-26'	'2011-04-24'	'2011-04-25'	'2011-04-28'
'2011-04-27'	'2011-04-25'	'2011-04-26'	'2011-04-29'
'2011-04-30'	'2011-04-26'	'2011-04-27'	'2011-04-30'
'2011-05-01'	'2011-04-27'	'2011-04-28'	'2011-05-01'
'2011-05-02'	'2011-04-28'	'2011-05-18'	'2011-05-02'
'2011-05-03'	'2011-04-29'	'2011-05-23'	'2011-05-23'
'2011-05-12'	'2011-04-30'	'2011-05-24'	'2011-05-24'
'2011-05-23'	'2011-05-01'		

Table B.2: Evaluation days in REDD experiments

UKDALE 1	UKDALE 2	UKDALE 5
'2013-03-17' - '2013-03-30'	'2013-05-21' - '2013-06-09'	'2014-06-30' - '2014-07-29'
'2013-04-02' - '2013-04-14'		
'2013-04-16' - '2013-04-18'		

Table B.3: Training days in UKDALE experiments

UKDALE 1	UKDALE 2	UKDALE 5
'2013-04-19' - '2013-10-05'	'2013-06-10' - '2013-08-04' '2013-09-12' - '2013-10-05'	'2014-07-30' - '2014-09-06'

Table B.4: Evaluation days in UKDALE experiments

ECO 1	ECO 2	ECO 4	ECO 5
'2012-07-01' - '2012-07-06' '2012-07-08' - '2012-07-13' '2012-07-16' - '2012-07-18'	'2012-06-01' '2012-06-03' '2012-06-06' '2012-06-11' - '2012-06-13' '2012-06-16' '2012-06-18' - '2012-06-20' '2012-06-24' - '2012-06-28'	'2012-06-27' - '2012-07-11'	'2012-06-27' - '2012- 07-11'

Table B.5: Training days in ECO experiments

ECO 1	ECO 2	ECO 4	ECO 5
'2012-07-19' - '2012-07-22' '2012-07-25' '2012-07-28' '2012-07-29' '2012-08-01' '2012-08-03' - '2012-08-06' '2012-08-10' '2012-08-13' '2012-08-14' '2012-08-16' '2012-08-19' - '2012-09-13' '2012-09-16' - '2012-09-28' '2012-10-01' '2012-10-02' '2012-10-04' - '2012-10-05' '2012-10-11' - '2012-10 -13' '2012-10-17' - '2012-10-22' '2012-10-24' - '2012-10-29' '2012-10-31'	'2012-06-30' '2012-07-04' - '2012-07-17' '2012-07-20' - '2012-08-05' '2012-08-08' - '2012-08-10' '2012-08-12' - '2012-08-18' '2012-08-20' - '2012-09-05' '2012-09-07' '2012-09-16' - '2012-09-24' '2012-09-28' '2012-09-29' '2012-10-01' - '2012-10-04'	'2012-07-12' '2012-07-13' '2012-07-15' - '2012-09-04' '2012-09-06' - '2012-09-10' '2012-09-12' '2012-09-13' '2012-09-16' - '2012-09-29'	'2012-07-12' - '2012-09-03' '2012-09-05' '2012-09-12' '2012-09-13' '2012-09-16' - '2012-10-03'

Table B.6: Evaluation days in ECO experiments

APPENDIX E

A summary of experiments run with Weiss and Hart algorithms is here reported. These and other experiments are referred to in section 3.3. More experiments have been performed, but they are not reported here if they are not relevant, i.e. if they have similar or identical results with other experiments. For some appliances more experiments have been performed with different values of parameter *r*, the relevant results are here reported.

Days selected for training and evaluation are reported in Appendix D.

E.1 REDD 1-4, APPLIANCES OF INTEREST, USING P ONLY

HOUSE 1

Appliance	F score Weiss	F score Hart	Precision Weiss	Precision Hart	Recall Weiss	Recall Hart	Parameter r	Threshold diff on off	Threshold diff on off
Fridge	0.5971	0.5198	0.8297	0.7517	0.4663	0.3973	0.1	80	80
Dishwasher	0.5210	0.4038	0.5345	0.3069	0.5082	0.5902	0.1	900	900
	0.5246	0.3648	0.5246	0.2552	0.5246	0.6393	0.3		
Kettle	0.0377	0.0455	0.0204	0.0455	0.2500	0.0455	0.1	1000	1000
Oven	0.7805	0.6842	0.9412	0.9286	0.6667	0.5417	0.1	2500	2500
Washing machine	0.9361	0.9495	0.9909	0.9912	0.8871	0.9113	0.1	1500	1500
Microwave	0.6557	0.6084	0.9610	0.8200	0.4977	0.4836	0.1	1000	1000
	0.6641	0.6214	0.9620	0.8269	0.5070	0.4977	0.8		
Bathroom GFI	0.2406	0.5435	0.1616	0.4310	0.4706	0.7353	0.1	500	500
Cooker	0.4144	0.6189	0.3239	0.5714	0.5750	0.6750	0.1	15	15

Table E.1

HOUSE 2

Appliance	F score Weiss	F score Hart	Precision Weiss	Precision Hart	Recall Weiss	Recall Hart	Parameter r	Threshold diff on off	Threshold diff on off
Fridge	0.5411	0.4757	0.9204	0.7254	0.3832	0.3539	0.1	80	80
Dishwasher	0.5581	0.1667	0.3871	0.1000	1.0000	0.5000	0.1	900	900
Microwave	0.9036	0.7556	0.9785	1.0000	0.8393	0.6071	0.1	1000	1000
Stove	0.9803	0.9711	0.9836	0.9668	0.9770	0.9754	0.1	500	500
Cooker	0.5263	0.5532	0.4268	0.4471	0.6863	0.7255	0.1	500	500

Table E.2

HOUSE 3

Appliance	F score Weiss	F score Hart	Precision Weiss	Precision Hart	Recall Weiss	Recall Hart	Parameter r	Threshold diff on off	Threshold diff on off
Fridge	0.6087	0.5250	0.8463	0.6214	0.4753	0.4544	0.1	80	80
Bathroom gfi	0.2097	0.1967	0.1518	0.1818	0.3393	0.2143	0.1	1000	1000
Washing machine	0.9465	0.9194	1.0000	1.0000	0.8985	0.8507	0.1	2500	2500
Microwave	0.7715	0.7314	0.7667	0.6543	0.7763	0.8289	0.1	1000	1000
	0.7548	0.6935	0.7231	0.5895	0.7895	0.8421	0.5		
Electronics	0.8762	0.8711	0.8852	0.8496	0.8673	0.8938	0.1	500	500

Table E.3

HOUSE 4

Appliance	F score Weiss	F score Hart	Precision Weiss	Precision Hart	Recall Weiss	Recall Hart	Parameter r	Threshold diff on off	Threshold diff on off
Fridge	0.6068	0.4223	0.8988	0.6796	0.4580	0.3063	0.1	80	80
Stove	0.5818	0.7600	1.0000	0.9333	0.4103	0.6410	0.1	500	500
Dishwasher	0.4167	0.2000	0.4167	0.1667	0.4167	0.2500	0.1	900	900

Table E.4

E.2 UK DALE 1, 2, 5, APPLIANCES OF INTEREST, USING P AND Q

HOUSE 1

Appliance	F score Weiss	F score Hart	Precision Weiss	Precision Hart	Recall Weiss	Recall Hart	Parameter r	Threshold diff on off	Threshold diff on off
Fridge	0.7639	0.7250	0.9433	0.8605	0.6418	0.6265	0.2	60	40
	0.7290	0.7216	0.8261	0.7730	0.6524	0.6767	0.3		
Kettle	0.8888	0.9093	0.8706	0.8969	0.9079	0.9221	0.1	2000	2000
Hairdryer	0.6714	0.8972	1.0000	0.9882	0.5054	0.8215	0.1	1500	1500
	0.8861	0.9692	0.9864	0.9863	0.8043	0.9527	1		
Microwave	0.8812	0.8159	0.9865	0.9857	0.7962	0.6960	0.1	1000	1000
Washing machine	0.1991	0.1576	0.1264	0.0984	0.4691	0.3951	0.1	1700	1700
Dishwasher	0.1660	0.6050	0.1303	0.7273	0.2286	0.5179	0.1	2000	2000
	0.0061	0.6050	0.0208	0.7273	0.0036	0.5179	0.05		

Table E.5

HOUSE 2

Appliance	F score Weiss	F score Hart	Precision Weiss	Precision Hart	Recall Weiss	Recall Hart	Parameter r	Threshold diff on off	Threshold diff on off
Fridge	0.6416	0.5315	0.9988	0.9960	0.4726	0.3624	0.1	40	40
	0.7865	0.7641	0.9787	0.9779	0.6573	0.6269	0.2		
Dishwasher	0.9115	0.8150	0.9349	0.7570	0.8893	0.8827	0.1	1000	1000
Kettle	0.9174	0.9183	0.9952	0.9937	0.8509	0.8536	0.1	2000	2000
Washing machine	0.4793	0.2985	0.7105	0.2051	0.3616	0.5480	0.1	1500	1500
	0.4881	0.4097	0.6374	0.2719	0.3955	0.8305	0.25		
Microwave	0.6673	0.9086	0.9518	0.9809	0.5138	0.8462	0.1	1000	1000
	0.7857	0.8760	0.8263	0.8754	0.7489	0.8766	0.3		
	0.8121	0.8237	0.7390	0.7176	0.9013	0.9666	1		

Table E.6

HOUSE 5

Appliance	F score Weiss	F score Hart	Precision Weiss	Precision Hart	Recall Weiss	Recall Hart	Parameter r	Threshold diff on off	Threshold diff on off
Fridge	0.8965	0.7571	0.9939	0.9812	0.8166	0.6163	0.1	40	40
Kettle	0.7245	0.6723	0.6798	0.5561	0.7756	0.8498	0.1	2500	2000
	0.8199	0.8862	0.8924	0.9440	0.7582	0.8352	0.05		
Dishwasher	0.2970	0.2850	0.1869	0.1715	0.7229	0.8434	0.05	1600	1600
	0.5871	0.7951	0.6774	0.7619	0.5181	0.8313	0.028		
Oven	0.6046	0.6807	0.5052	0.5504	0.7526	0.8919	0.1	1500	1500
	0.5869	0.6554	0.5116	0.5490	0.6881	0.8129	0.05		
	0.4977	0.7268	0.5459	0.7140	0.4574	0.7401	0.025		

Table E.7

E.3 UK DALE 1, 2, APPLIANCES OF INTEREST, USING P

HOUSE 1

Appliance	F score Weiss	F score Hart	Precision Weiss	Precision Hart	Recall Weiss	Recall Hart	Parameter r	Threshold diff on off	Threshold diff on off
Fridge	0.4448	0.3353	0.6354	0.3273	0.3421	0.3437	0.2	60	40
	0.5653	0.2597	0.5709	0.1994	0.5597	0.3723	0.3		
Kettle	0.9433	0.9557	0.9769	0.9810	0.9119	0.9316	0.1	2000	2000
Hairdryer	0.8918	0.9554	0.9972	0.9907	0.8065	0.9226	0.1	1500	1500
	0.8931	0.9578	0.9973	0.9908	0.8086	0.9269	1		
Microwave	0.7535	0.6932	0.6867	0.6568	0.8347	0.7340	0.1	1000	1000
Washing machine	0.1465	0.2001	0.0870	0.1179	0.4630	0.6605	0.1	1700	1700
Dishwasher	0.4538	0.5936	0.3400	0.4437	0.6821	0.8964	0.1	2000	2000
	0.4538	0.5936	0.3400	0.4437	0.6821	0.8964	0.05		

Table E.8

HOUSE 2

Appliance	F score Weiss	F score Hart	Precision Weiss	Precision Hart	Recall Weiss	Recall Hart	Parameter r	Threshold diff on off	Threshold diff on off
Fridge	0.7920	0.6866	0.9312	0.8041	0.6890	0.5991	0.1	40	40
	0.7744	0.6959	0.8522	0.6972	0.7096	0.6946	0.2		
Dishwasher	0.5444	0.7762	0.4746	0.6867	0.6384	0.8925	0.1	1000	1000
Kettle	0.9141	0.9303	0.9874	0.9969	0.8509	0.8720	0.1	2000	2000
Washing machine	0.3679	0.5818	0.2755	0.4289	0.5537	0.9040	0.1	1500	1500
	0.3033	0.4688	0.2000	0.3165	0.6271	0.9040	0.25		
Microwave	0.6968	0.8528	0.9098	0.9232	0.5646	0.7925	0.1	1000	1000
	0.6244	0.8700	0.6876	0.7929	0.5718	0.9637	0.3		
	0.5907	0.8248	0.6108	0.7169	0.5718	0.9710	1		

Table E.9

E.4 ECO 1, 2, 4, 5, APPLIANCES OF INTEREST, USING P AND Q

HOUSE 1

Appliance	F score Weiss	F score Hart	Precision Weiss	Precision Hart	Recall Weiss	Recall Hart	Parameter r	Threshold diff on off	Threshold diff on off
Freezer	0.4922	0.6884	0.8904	0.8012	0.3401	0.6034	0.1	15	15
Fridge	0.4526	0.2447	0.9696	0.7578	0.2952	0.1459	0.1	30	30
Kettle	0.9095	0.6662	0.9179	0.9722	0.9013	0.5067	0.1	1500	1500
Washing machine	0.4194	0.6228	0.3119	0.4838	0.6396	0.8739	0.1	1800	1800
	0.3826	0.5850	0.2705	0.4234	0.6532	0.9459	0.35		
	0.1603	0.2903	0.0910	0.1699	0.6712	0.9955	0.7		

Table E.10

HOUSE 2

Appliance	F score Weiss	F score Hart	Precision Weiss	Precision Hart	Recall Weiss	Recall Hart	Parameter r	Threshold diff on off	Threshold diff on off
Freezer	0.8379	0.5747	0.9961	0.7973	0.7230	0.4492	0.1	15	15
Fridge	0.9302	0.5033	0.9901	0.9793	0.8772	0.3387	0.1	15	15
Dishwasher	0.5012	0.8899	0.9891	1.0000	0.3356	0.8017	0.05	1000	1000
Stove	0.9981		1.0000		0.9962		1	500	500
Kettle	0.8184	0.6564	0.9754	0.9651	0.7049	0.4973	0.1	1500	1500

Table E.11

HOUSE 4

Appliance	F score Weiss	F score Hart	Precision Weiss	Precision Hart	Recall Weiss	Recall Hart	Parameter r	Threshold diff on off	Threshold diff on off
Fridge	0.6505	0.6141	0.5322	0.5244	0.8363	0.7408	0.1	15	40
Freezer	0.6134	0.5580	0.9984	0.9991	0.4427	0.3871	0.1	15	40
Microwave	0.5529	0.7001	0.9327	0.9911	0.3929	0.5412	0.1	1300	1300

Table E.12

HOUSE 5

Appliance	F score Weiss	F score Hart	Precision Weiss	Precision Hart	Recall Weiss	Recall Hart	Parameter r	Threshold diff on off	Threshold diff on off
Fridge	0.4826	0.2235	0.9961	0.9796	0.3184	0.1261	0.1	40	40
Microwave	0.7182	0.6744	0.6735	0.5530	0.7692	0.8643	0.1	1000	1000

Table E.13

E.5 ECO 1, 2, 4, 5, APPLIANCES OF INTEREST, USING P

HOUSE 1

Appliance	F score Weiss	F score Hart	Precision Weiss	Precision Hart	Recall Weiss	Recall Hart	Parameter r	Threshold diff on off	Threshold diff on off
Freezer	0.5259	0.5914	0.7170	0.5310	0.4152	0.6673	0.1	15	15
Fridge	0.4384	0.3432	0.4330	0.5457	0.4439	0.2503	0.1	30	30
Kettle	0.7937	0.7875	0.7090	0.7814	0.9013	0.7937	0.1	1500	1500
Washing machine	0.4285	0.6313	0.3167	0.4749	0.6622	0.9414	0.1	1800	1800
	0.3595	0.5118	0.2455	0.3473	0.6712	0.9730	0.35		
	0.1648	0.2697	0.0936	0.1560	0.6892	0.9955	0.7		

Table E.14

HOUSE 2

Appliance	F score Weiss	F score Hart	Precision Weiss	Precision Hart	Recall Weiss	Recall Hart	Parameter r	Threshold diff on off	Threshold diff on off
Freezer	0.8634	0.4485	0.8246	0.7386	0.9060	0.3220	0.1	15	15
Fridge	0.9297	0.4910	0.9324	0.8599	0.9271	0.3436	0.1	15	15
Dishwasher	0.8136	0.8018	0.8348	0.8585	0.7934	0.7521	0.05	1000	1000
Stove	0.9941		0.9919		0.9962		1	500	500
Kettle	0.7757	0.6792	0.8623	0.8000	0.7049	0.5902	0.1	1500	1500

Table E.15

HOUSE 4

Appliance	F score Weiss	F score Hart	Precision Weiss	Precision Hart	Recall Weiss	Recall Hart	Parameter r	Threshold diff on off	Threshold diff on off
Fridge	0.5883	0.4131	0.4497	0.3431	0.8502	0.5191	0.1	15	40
Freezer	0.5529	0.3159	0.8789	0.3709	0.4033	0.2751	0.1	15	40
Microwave	0.7561	0.7680	0.8658	0.9290	0.6711	0.6545	0.1	1300	1300

Table E.16

HOUSE 5

Appliance	F score Weiss	F score Hart	Precision Weiss	Precision Hart	Recall Weiss	Recall Hart	Parameter r	Threshold diff on off	Threshold diff on off
Fridge	0.5969	0.3940	0.9492	0.6320	0.4354	0.2862	0.1	40	40
Microwave	0.7867	0.7029	0.6628	0.5506	0.9676	0.9715	0.1	1000	1000

Table E.17

APPENDIX F

In this appendix the general comparison between metrics of the 2 algorithms is shown. When there are multiple experiments for one appliance (see Appendix D), those with better F-score have been selected.

The difference between metrics in Weiss and in Hart is computed and multiplied by 100. If the number is positive Weiss outperform Hart, if it is negative Hart is better.

F.1 P ONLY

REDD 1

Appliance	% diff. in F-score	% diff. in Precision	% diff. in Recall
Fridge	7.73	7.8	6.9
Dishwasher	11.72	22.76	-8.2
Kettle	-0.78	-2.51	20.45
Oven	9.63	1.26	12.5
Washing machine	-1.34	-0.03	-2.42
Microwave	4.27	13.51	0.93
Bathroom GFI	-30.29	-26.94	-26.47
Cooker	-20.45	-24.75	-10

Table F.1

REDD 2

Appliance	% diff. in F-score	% diff. in Precision	% diff. in Recall
Fridge	6.54	19.5	2.93
Dishwasher	39.14	28.71	50
Microwave	14.8	-2.15	23.22
Stove	0.92	1.68	0.16
Cooker	-2.69	-2.03	-3.92

Table F.2

REDD 3

Appliance	% diff. in F-score	% diff. in Precision	% diff. in Recall
Fridge	8.37	22.49	2.09
Bathroom gfi	1.3	-3	12.5
Washing machine	2.71	0	4.78
Microwave	4.01	11.24	-5.26
Electronics	0.51	3.56	-2.65

Table F.3

REDD 4

Appliance	% diff. in F-score	% diff. in Precision	% diff. in Recall
Fridge	18.45	21.92	15.17
Stove	-17.82	6.67	-23.07
Dishwasher	21.67	25	16.67

Table F.4

UKDALE 1

Appliance	% diff. in F-score	% diff. in Precision	% diff. in Recall
Fridge	30.56	37.15	18.74
Kettle	-1.24	-0.41	-1.97
Hairdryer	-6.47	0.65	-11.83
Microwave	6.03	2.99	10.07
Washing machine	-5.36	-3.09	-19.75
Dishwasher	-13.98	-10.37	-21.43

Table F.5

UKDALE 2

Appliance	% diff. in F-score	% diff. in Precision	% diff. in Recall
Fridge	10.54	12.71	8.99
Dishwasher	-23.18	-21.21	-25.41
Kettle	-1.62	-0.95	-2.11
Washing machine	-21.39	-15.34	-35.03
Microwave	-15.6	-1.34	-22.79

Table F.6

ECO 1

Appliance	% diff. in F-score	% diff. in Precision	% diff. in Recall
Freezer	-6.55	18.6	-25.21
Fridge	9.52	-11.27	19.36
Kettle	0.62	-7.24	10.76
Washing machine	-20.28	-15.82	-27.92

Table F.7

ECO 2

Appliance	% diff. in F-score	% diff. in Precision	% diff. in Recall
Freezer	41.49	8.6	58.4
Fridge	43.87	7.25	58.35
Dishwasher	1.18	-2.37	4.13
Stove	///	///	///
Kettle	9.65	6.23	11.47

Table F.8

ECO 4

Appliance	% diff. in F-score	% diff. in Precision	% diff. in Recall
Fridge	17.52	10.66	33.11
Freezer	23.7	50.8	12.82
Microwave	-1.19	-6.32	1.66

Table F.9

ECO 5

Appliance	% diff. in F-score	% diff. in Precision	% diff. in Recall
Fridge	20.29	31.72	14.92
Microwave	8.38	11.22	-0.39

Table F.10

F.2 P AND Q

UKDALE 1

Appliance	% diff. in F-score	% diff. in Precision	% diff. in Recall
Fridge	3.89	8.28	1.53
Kettle	-2.05	-2.63	-1.42
Hairdryer	-8.31	0.01	-14.84
Microwave	6.53	0.08	10.02
Washing machine	4.15	2.8	7.4
Dishwasher	-43.9	-59.7	-28.93

Table F.11

UKDALE 2

Appliance	% diff. in F-score	% diff. in Precision	% diff. in Recall
Fridge	2.24	0.08	3.04
Dishwasher	9.65	17.79	0.66
Kettle	-0.09	0.15	-0.27
Washing machine	7.84	36.55	-43.5
Microwave	-24.13	-2.91	-33.24

Table F.12

UKDALE 5

Appliance	% diff. in F-score	% diff. in Precision	% diff. in Recall
Fridge	13.94	1.27	20.03
Kettle	-6.63	-5.16	-7.7
Dishwasher	-20.8	-8.45	-31.32
Oven	-7.61	-4.52	-13.93

Table F.13

ECO 1

Appliance	% diff. in F-score	% diff. in Precision	% diff. in Recall
Freezer	-19.62	8.92	-26.33
Fridge	20.79	21.18	14.93
Kettle	24.33	-5.43	39.46
Washing machine	-20.34	-17.19	-23.43

Table F.14

ECO 2

Appliance	% diff. in F-score	% diff. in Precision	% diff. in Recall
Freezer	26.32	19.88	27.38
Fridge	42.69	1.08	53.85
Dishwasher	-38.87	-1.09	-46.61
Kettle	16.2	1.03	20.76

Table F.15

ECO 4

Appliance	% diff. in F-score	% diff. in Precision	% diff. in Recall
Fridge	3.64	0.78	9.55
Freezer	5.54	-0.07	5.56
Microwave	-14.72	-5.84	-14.83

Table F.16

ECO 5

Appliance	% diff. in F-score	% diff. in Precision	% diff. in Recall
Fridge	25.91	1.65	19.23
Microwave	4.38	12.05	-9.51

Table F.17

APPENDIX G

In this appendix the general comparison between use of P only and P/Q is shown, for both algorithms. The difference between metrics in the two cases is computed and multiplied by 100. For each appliance the experiment with better results has been kept, when there were multiple experiments

UKDALE 1

appliance	ΔF-sc. W	ΔF-sc. H	ΔPr. W	ΔPr. H	ΔRec. W	ΔRec. H	param. R
Fridge	31.91	38.97	30.79	53.32	29.97	28.28	0.2
Kettle	-5.45	-4.64	-10.63	-8.41	-0.4	-0.95	0.1
Hairdryer	-0.7	1.14	-1.09	-0.45	-0.43	2.58	1
Microwave	12.77	12.27	29.98	32.89	-3.85	-3.8	0.1
Washing machine	5.26	-4.25	3.94	-1.95	0.61	-26.54	0.1
Dishwasher	-28.78	1.14	-20.97	28.36	-45.35	-37.85	0.1

Table G.1

UKDALE 2

appliance	ΔF-sc. W	ΔF-sc. H	ΔPr. W	ΔPr. H	ΔRec. W	ΔRec. H	param. R
Fridge	-15.04	-15.51	6.76	19.19	-21.64	-23.67	0.1
	1.21	6.82	12.65	28.07	-5.23	-6.77	0.2
Dishwasher	36.71	3.88	46.03	7.03	25.09	-0.98	0.1
Kettle	0.33	-1.2	0.78	-0.32	0	-1.84	0.1
Washing machine	18.48	-5.91	43.74	-4.46	-23.16	-7.35	0.25
Microwave	-2.95	5.58	4.2	5.77	-5.08	5.37	0.1

Table G.2

ECO 1

appliance	ΔF-sc. W	ΔF-sc. H	ΔPr. W	ΔPr. H	ΔRec. W	ΔRec. H	param. R
Freezer	-3.37	9.7	17.34	27.02	-7.51	-6.39	0.1
Fridge	1.42	-9.85	53.66	21.21	-14.87	-10.44	0.1
Kettle	11.58	-12.13	20.89	19.08	0	-28.7	0.1
Washing machine	-0.91	-0.85	-0.48	0.89	-2.26	-6.75	0.1

Table G.3

ECO 2

appliance	ΔF-sc. W	ΔF-sc. H	ΔPr. W	ΔPr. H	ΔRec. W	ΔRec. H	param. R
Freezer	-2.55	12.62	17.15	5.87	-18.3	12.72	0.1
Fridge	0.05	1.23	5.77	11.94	-4.99	-0.49	0.1
Dishwasher	-31.24	8.81	15.43	14.15	-45.78	4.96	0.05
Stove	0.4	0	0.81	0	0	0	1
Kettle	4.27	-2.28	11.31	16.51	0	-9.29	0.1

Table G.4

ECO 4

appliance	ΔF-sc. W	ΔF-sc. H	ΔPr. W	ΔPr. H	ΔRec. W	ΔRec. H	param. R
Fridge	6.22	20.1	8.25	18.13	-1.39	22.17	0.1
Freezer	6.05	24.21	11.95	62.82	3.94	11.2	0.1
Microwave	-20.32	-6.79	6.69	6.21	-27.82	-11.33	0.1

Table G.5

ECO 5

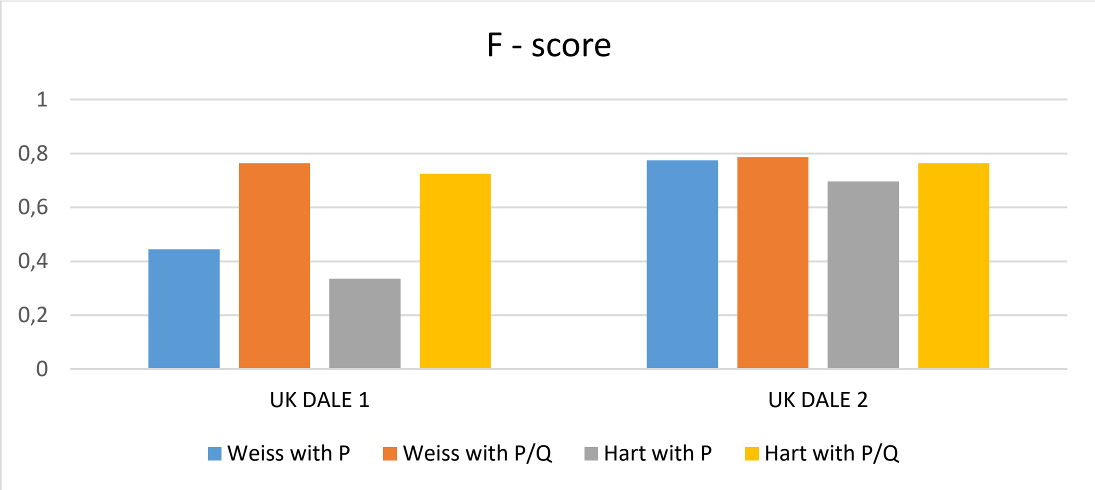
appliance	ΔF-sc. W	ΔF-sc. H	ΔPr. W	ΔPr. H	ΔRec. W	ΔRec. H	param. R
Fridge	-11.43	-17.05	4.69	34.76	-11.7	-16.01	0.1
Microwave	-6.85	-2.85	1.07	0.24	-19.84	-10.72	0.1

Table G.6

APPENDIX H

In this section the comparison between algorithms and features for cooling appliances is shown. Metrics used here are these from *Table 3.10* that summarizes all the Cooling appliances from experiments in Appendix E.

H.1 P vs P/Q



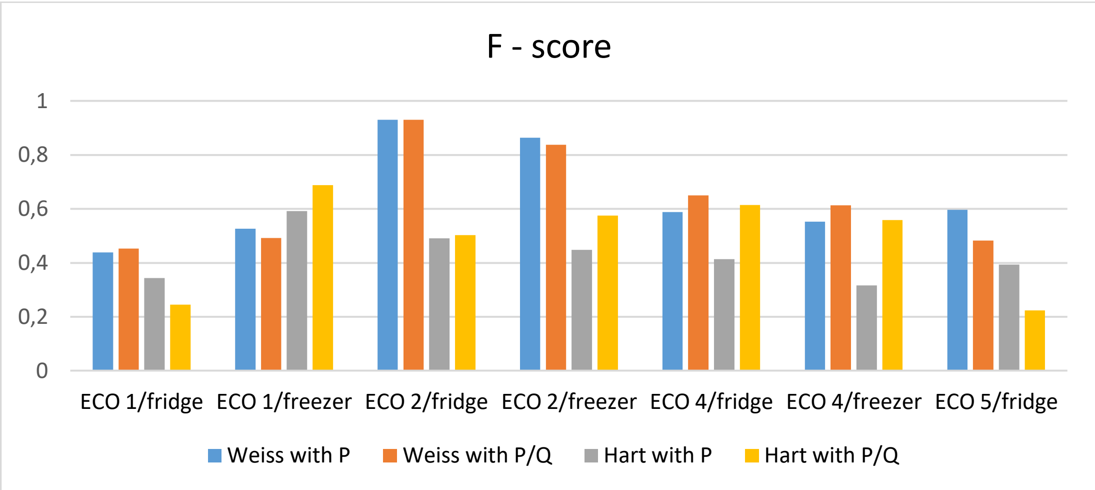
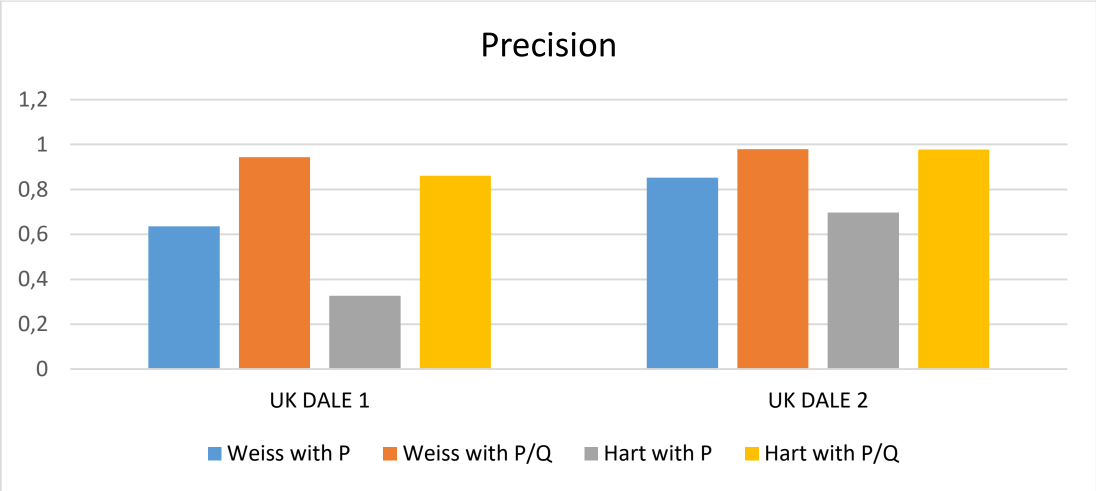


Figure H.1

In general F-score is better for experiments with P/Q features. There are some exceptions: ECO1/Fridge (Hart), ECO1/Freezer (Weiss) and ECO5/Fridge.



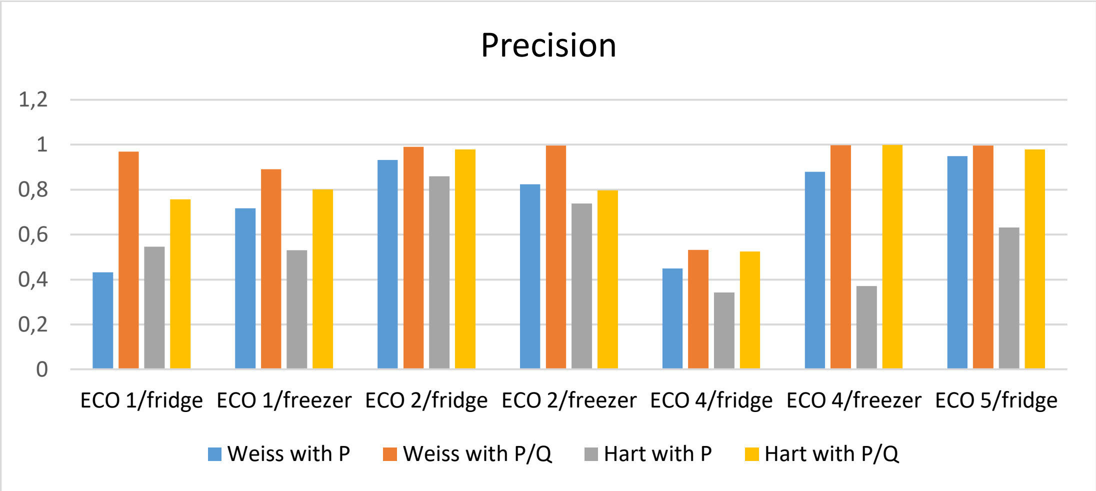


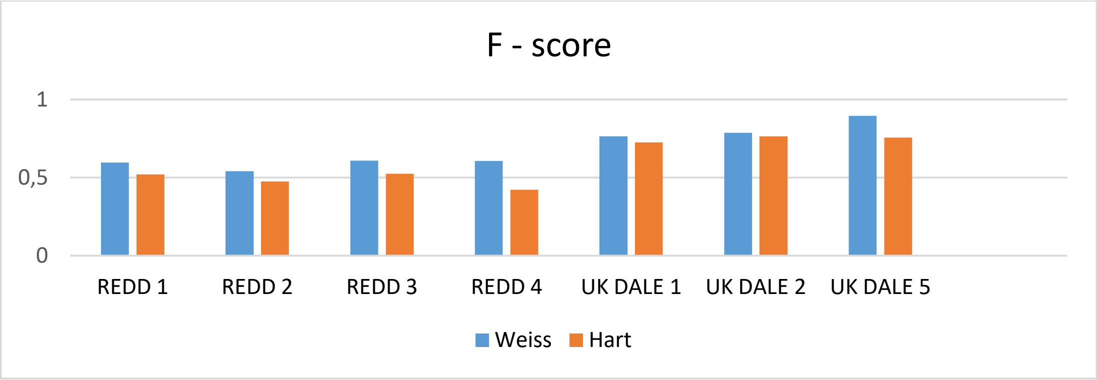
Figure H.2

In term of precision in all the cases the use of P/Q is preferable.

H.2 WEISS vs HART

From information of section H.2, use of P/Q is preferable. Now a comparison between the 2 algorithms is done, using P/Q features when available.

Apart from ECO 1/freezer, in all the other cases use of Weiss is preferable in term of F-score. Precision is higher using Weiss in all the cases. Charts are in Figures *H.3* and *H.4*.



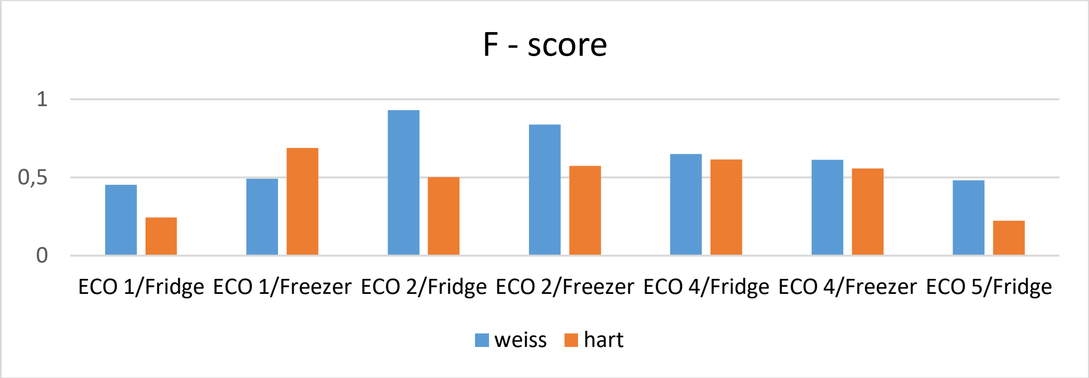
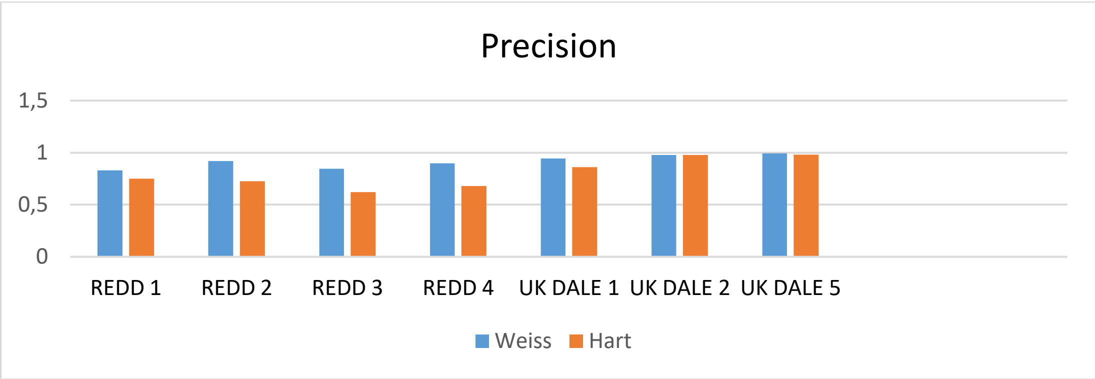


Figure H.3


Precision
1,5
1
0,5
0
REDD 1
REDD 2
REDD 3
REDD 4
UK DALE 1
UK DALE 2
UK DALE 5
Weiss
Hart

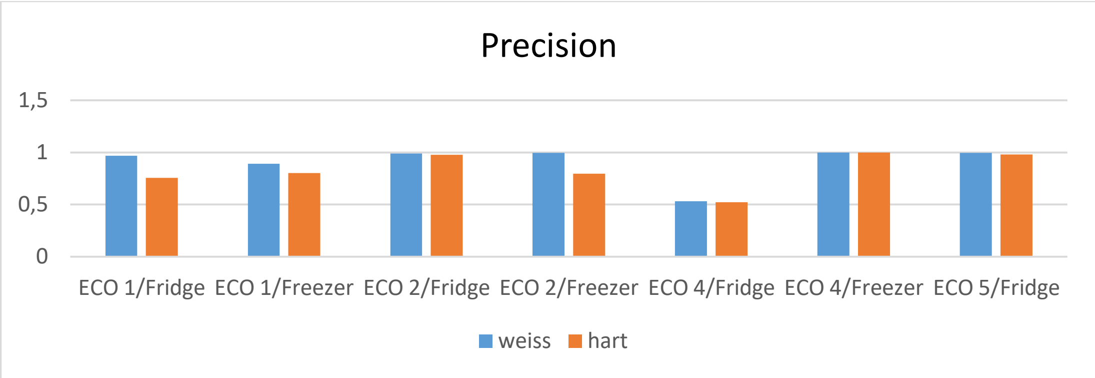

Precision
1,5
1
0,5
0
ECO 1/Fridge
ECO 1/Freezer
ECO 2/Fridge
ECO 2/Freezer
ECO 4/Fridge
ECO 4/Freezer
ECO 5/Fridge
weiss
hart

Figure H.4

APPENDIX I

Tables used for section 3.3.6 are gathered here.

I.1 MID/HIGH POWER APPLIANCES EVENT RESULTS

DATASET/HOUSE/APPLIANCE	F-score Weiss	F-score Hart	Precision Weiss	Precision Hart	Recall Weiss	Recall Hart	Parameter r
REDD 1/Dishwasher	0.5210	0.4038	0.5345	0.3069	0.5082	0.5902	0.1
	0.5246	0.3648	0.5246	0.2552	0.5246	0.6393	0.3
REDD 1/Microwave	0.6557	0.6084	0.9610	0.8200	0.4977	0.4836	0.1
	0.6641	0.6214	0.9620	0.8269	0.5070	0.4977	0.8
REDD 1/Bathroom GFI	0.2406	0.5435	0.1616	0.4310	0.4706	0.7353	0.1
REDD 2/Dishwasher	0.5581	0.1667	0.3871	0.1000	1.0000	0.5000	0.1
REDD 2/Microwave	0.9036	0.7556	0.9785	1.0000	0.8393	0.6071	0.1
REDD 2/Stove	0.9803	0.9711	0.9836	0.9668	0.9770	0.9754	0.1
REDD 3/Bathroom GFI	0.2097	0.1967	0.1518	0.1818	0.3393	0.2143	0.1
REDD 3/Microwave	0.7715	0.7314	0.7667	0.6543	0.7763	0.8289	0.1
	0.7548	0.6935	0.7231	0.5895	0.7895	0.8421	0.5
REDD 3/Electronics	0.8762	0.8711	0.8852	0.8496	0.8673	0.8938	0.1
REDD 4/Stove	0.5818	0.7600	1.0000	0.9333	0.4103	0.6410	0.1
REDD 4/Dishwasher	0.4167	0.2000	0.4167	0.1667	0.4167	0.2500	0.1
UKDALE 1/Hairdryer	0.6714	0.8972	1.0000	0.9882	0.5054	0.8215	0.1
	0.8861	0.9692	0.9864	0.9863	0.8043	0.9527	1
UKDALE 1/Microwave	0.8812	0.8159	0.9865	0.9857	0.7962	0.6960	0.1
UKDALE 1/Washing machine	0.1991	0.1576	0.1264	0.0984	0.4691	0.3951	0.1
UKDALE 1/Dishwasher	0.1660	0.6050	0.1303	0.7273	0.2286	0.5179	0.1
	0.0061	0.6050	0.0208	0.7273	0.0036	0.5179	0.05
UKDALE 2/Dishwasher	0.9115	0.8150	0.9349	0.7570	0.8893	0.8827	0.1
UKDALE 2/Washing machine	0.4793	0.2985	0.7105	0.2051	0.3616	0.5480	0.1
UKDALE 2/Microwave	0.6673	0.9086	0.9518	0.9809	0.5138	0.8462	0.1
	0.7857	0.8760	0.8263	0.8754	0.7489	0.8766	0.3
	0.8121	0.8237	0.7390	0.7176	0.9013	0.9666	1
UKDALE 5/Dishwasher	0.2970	0.2850	0.1869	0.1715	0.7229	0.8434	0.05
	0.5871	0.7951	0.6774	0.7619	0.5181	0.8313	0.028
UKDALE 5/Oven	0.6046	0.6807	0.5052	0.5504	0.7526	0.8919	0.1
	0.5869	0.6554	0.5116	0.5490	0.6881	0.8129	0.05
	0.4977	0.7268	0.5459	0.7140	0.4574	0.7401	0.025
ECO 1/Kettle	0.9095	0.6662	0.9179	0.9722	0.9013	0.5067	0.1
ECO 1/Washing machine	0.4194	0.6228	0.3119	0.4838	0.6396	0.8739	0.1
	0.3826	0.5850	0.2705	0.4234	0.6532	0.9459	0.35
	0.1603	0.2903	0.0910	0.1699	0.6712	0.9955	0.7
ECO 2/Dishwasher	0.5012	0.8899	0.9891	1.0000	0.3356	0.8017	0.05
ECO 2/Kettle	0.8184	0.6564	0.9754	0.9651	0.7049	0.4973	0.1
ECO 4/Microwave	0.5529	0.7001	0.9327	0.9911	0.3929	0.5412	0.1
ECO 5/Microwave	0.7182	0.6744	0.6735	0.5530	0.7692	0.8643	0.1

Table I.1: Mid/high power appliances event results, using P/Q when available

I.2 DISHWASHERS AND WASHING MACHINES EVENT RESULTS

DATASET/HOUSE/APPLIANCE	F-score Weiss	F-score Hart	Precision Weiss	Precision Hart	Recall Weiss	Recall Hart	Parameter r
REDD 1/Dishwasher	0.5210	0.4038	0.5345	0.3069	0.5082	0.5902	0.1
	0.5246	0.3648	0.5246	0.2552	0.5246	0.6393	0.3
REDD 2/Dishwasher	0.5581	0.1667	0.3871	0.1000	1.0000	0.5000	0.1
REDD 4/Dishwasher	0.4167	0.2000	0.4167	0.1667	0.4167	0.2500	0.1
UKDALE 1/Dishwasher	0.1660	0.6050	0.1303	0.7273	0.2286	0.5179	0.1
	0.0061	0.6050	0.0208	0.7273	0.0036	0.5179	0.05
UKDALE 2/Dishwasher	0.9115	0.8150	0.9349	0.7570	0.8893	0.8827	0.1
UKDALE 5/Dishwasher	0.2970	0.2850	0.1869	0.1715	0.7229	0.8434	0.05
	0.5871	0.7951	0.6774	0.7619	0.5181	0.8313	0.028
ECO 2/Dishwasher	0.5012	0.8899	0.9891	1.0000	0.3356	0.8017	0.05
UKDALE 1/Washing machine	0.1991	0.1576	0.1264	0.0984	0.4691	0.3951	0.1
UKDALE 2/Washing machine	0.4793	0.2985	0.7105	0.2051	0.3616	0.5480	0.1
ECO 1/Washing machine	0.4194	0.6228	0.3119	0.4838	0.6396	0.8739	0.1
	0.3826	0.5850	0.2705	0.4234	0.6532	0.9459	0.35
	0.1603	0.2903	0.0910	0.1699	0.6712	0.9955	0.7

Table I.2: Dishwashers and Washing Machines event results, using P/Q when available

APPENDIX J

Time patterns of appliances considered in 3.3.6 are reported here. The day have been divided in 48 time windows, each 30 minutes long. For both training and evaluation periods all the events have been assigned to their time window. Figures in this appendix detail that.

J.1 MICROWAVES

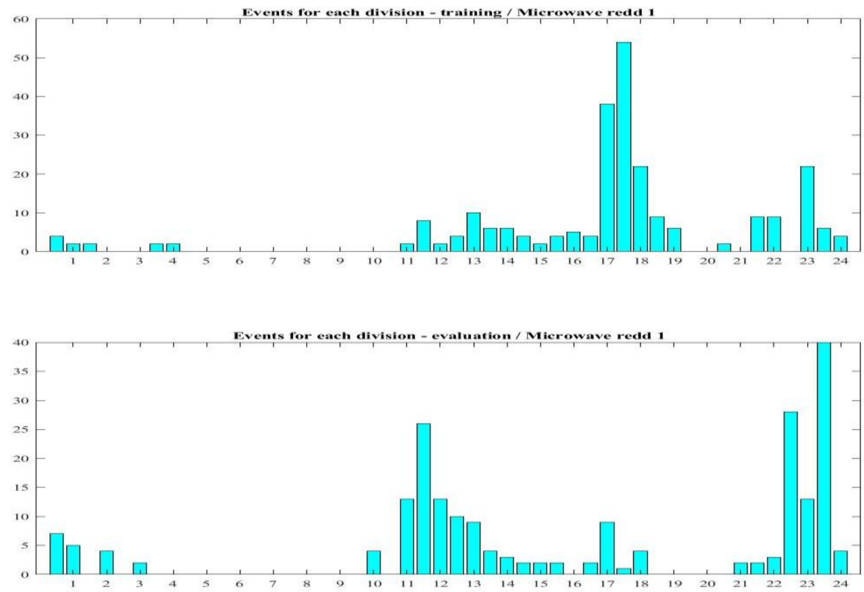


Figure J.1

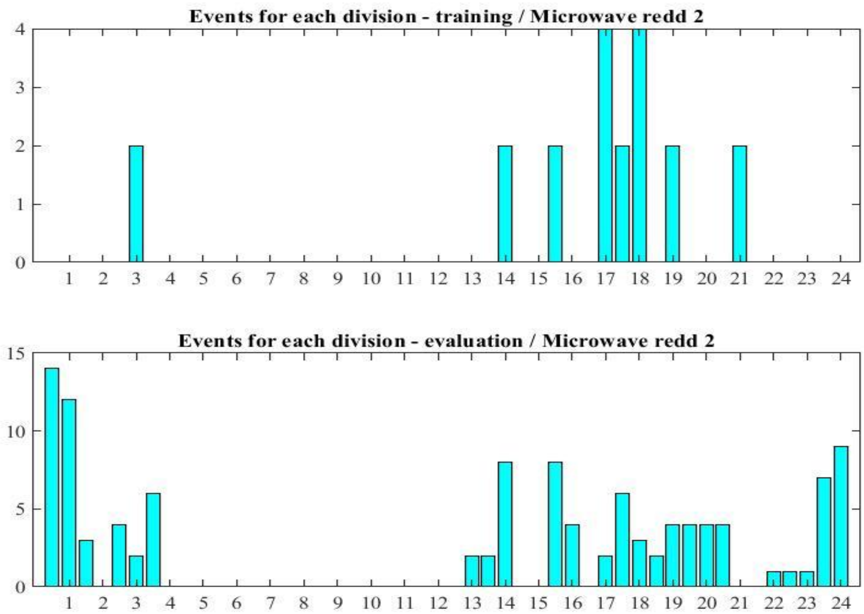


Figure J.2

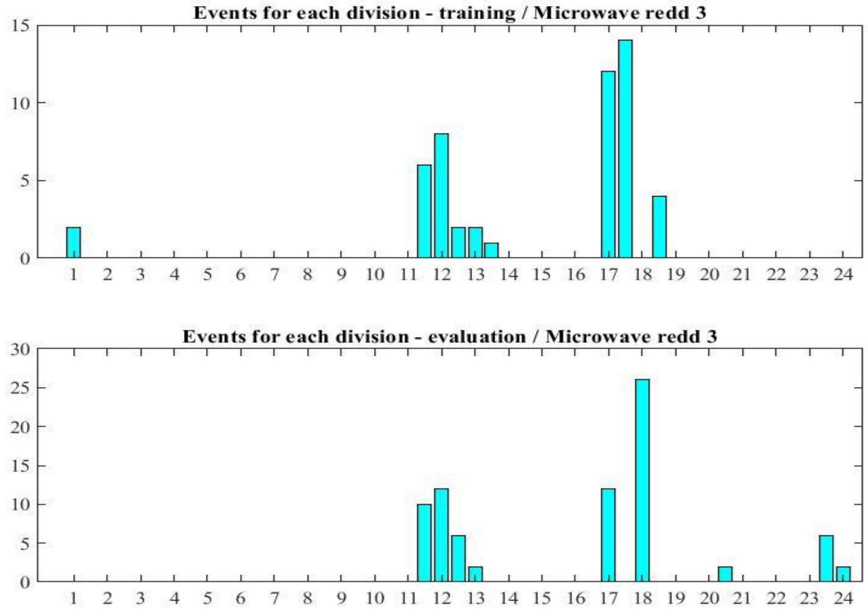


Figure J.3

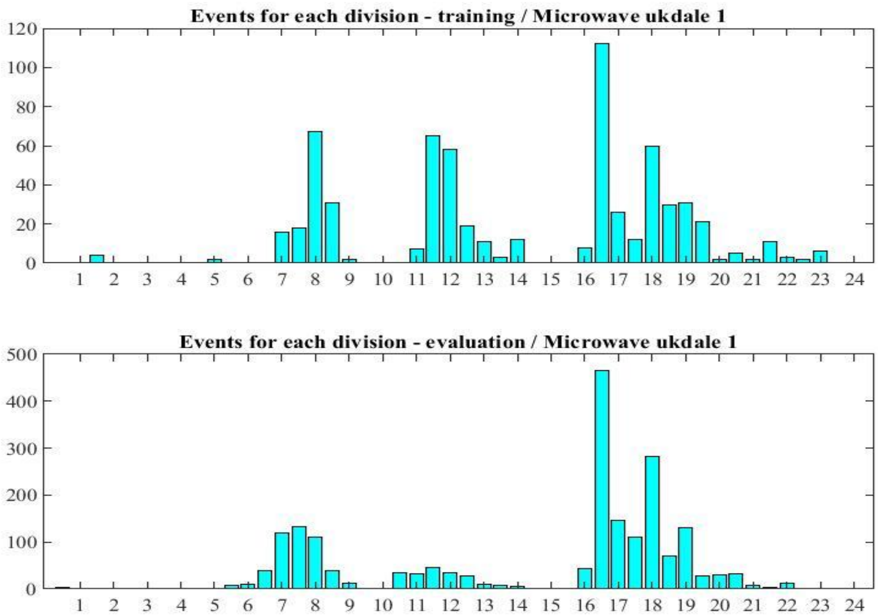


Figure J.4

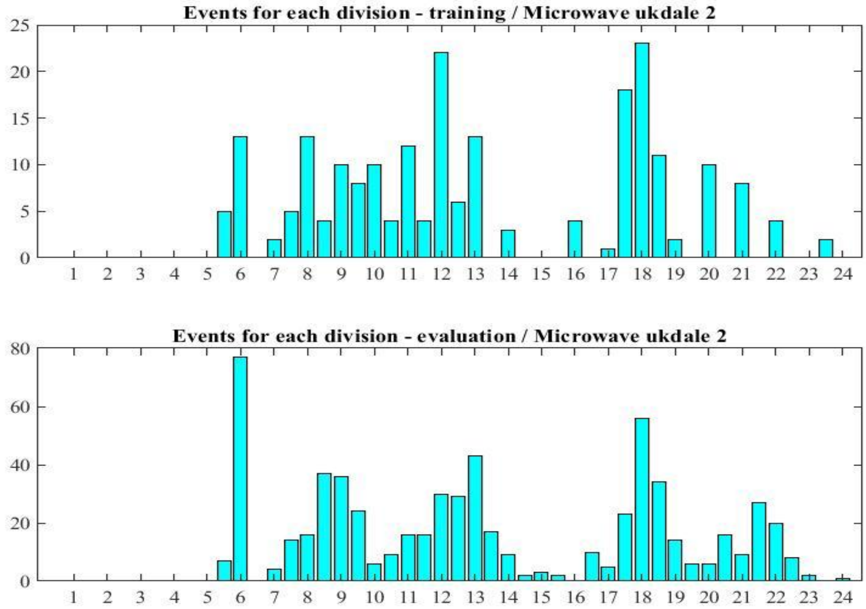


Figure J.5

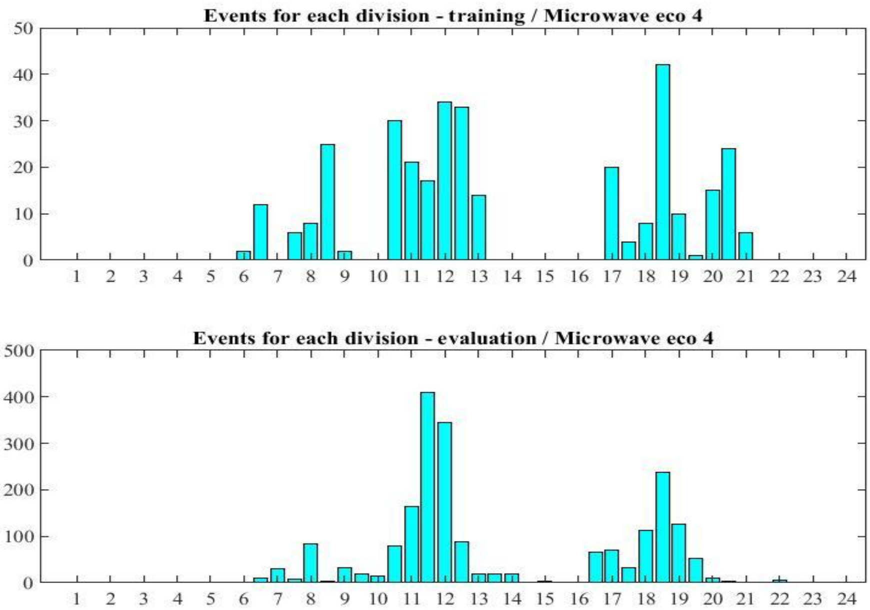


Figure J.6

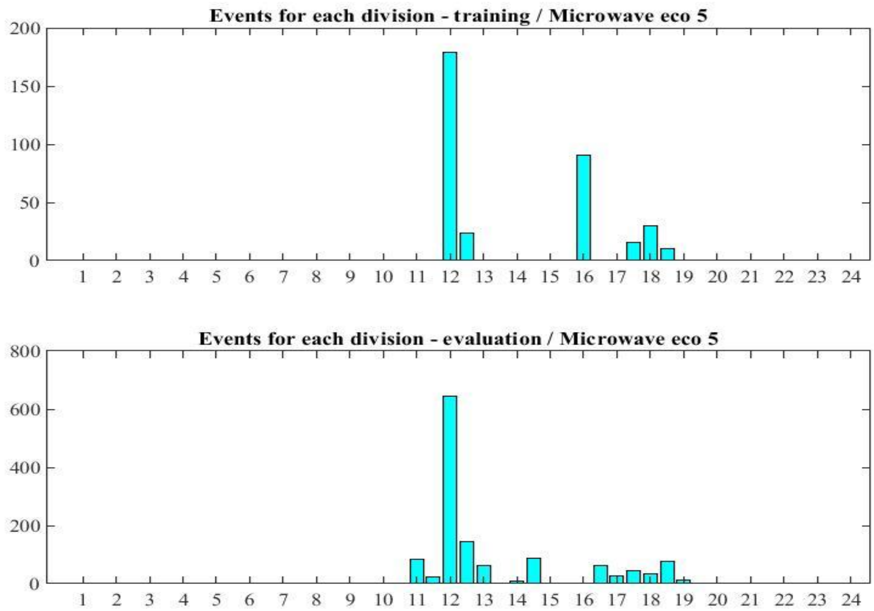


Figure J.7

J.2 DISHWASHERS

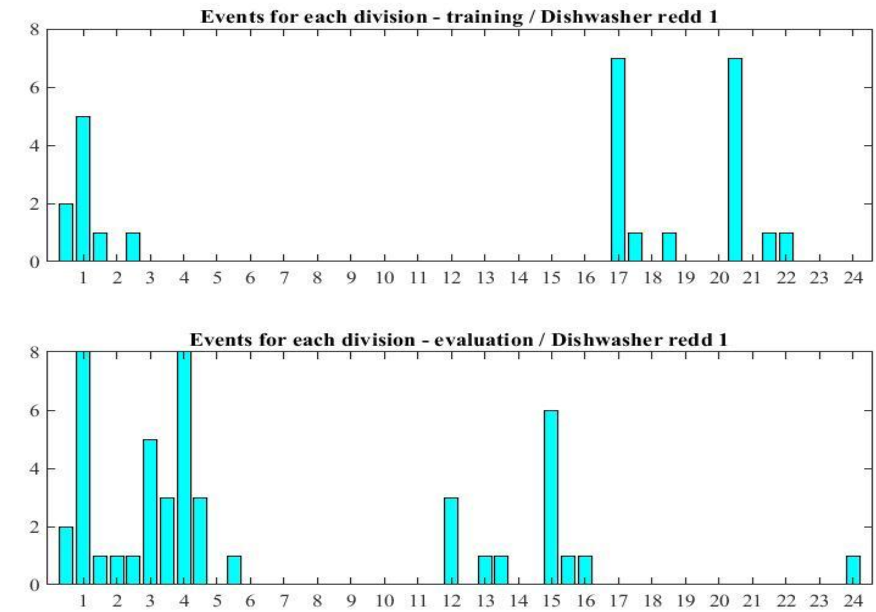


Figure J.8

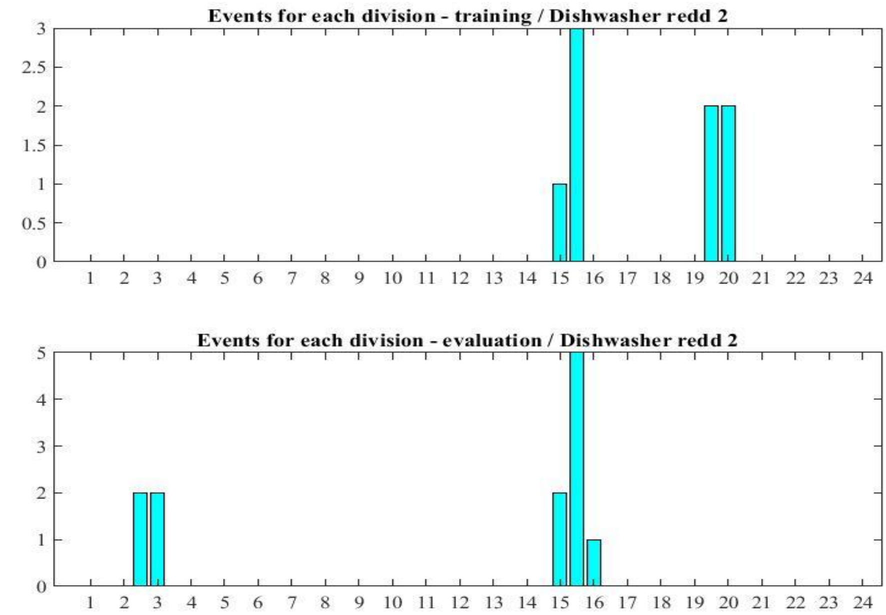


Figure J.9

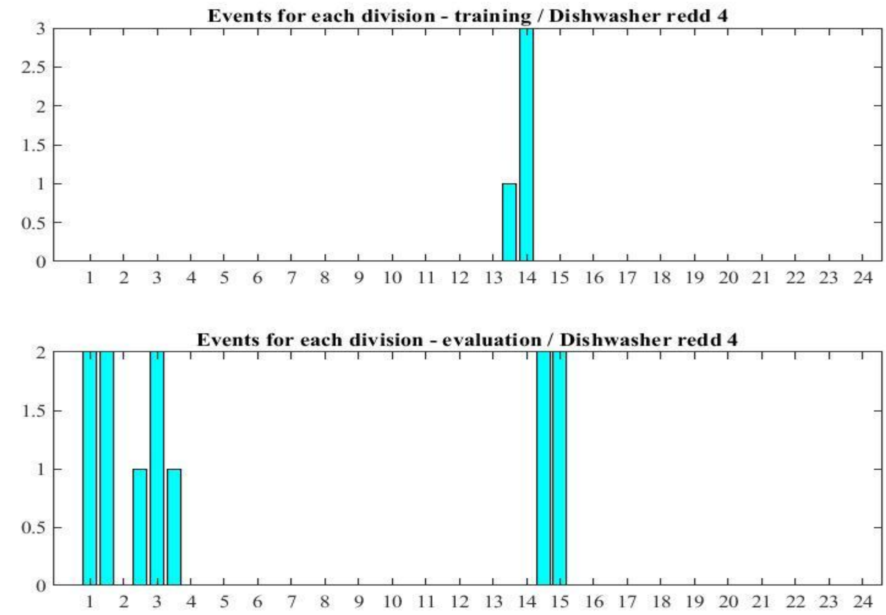


Figure J.10

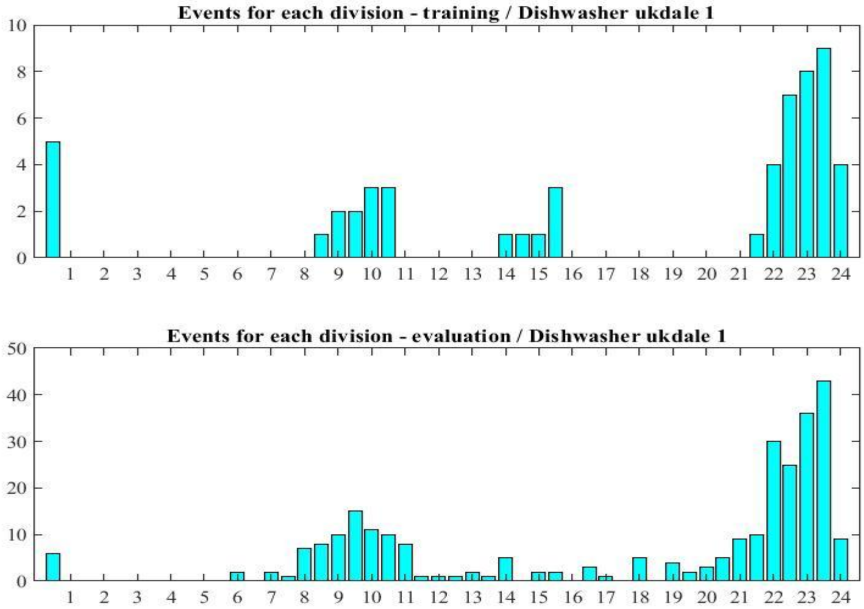


Figure J.11

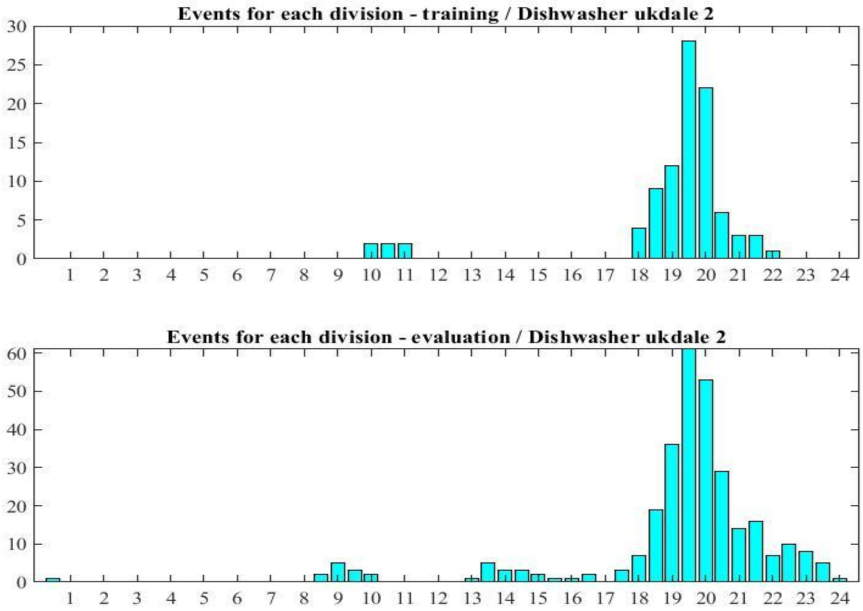


Figure J.12

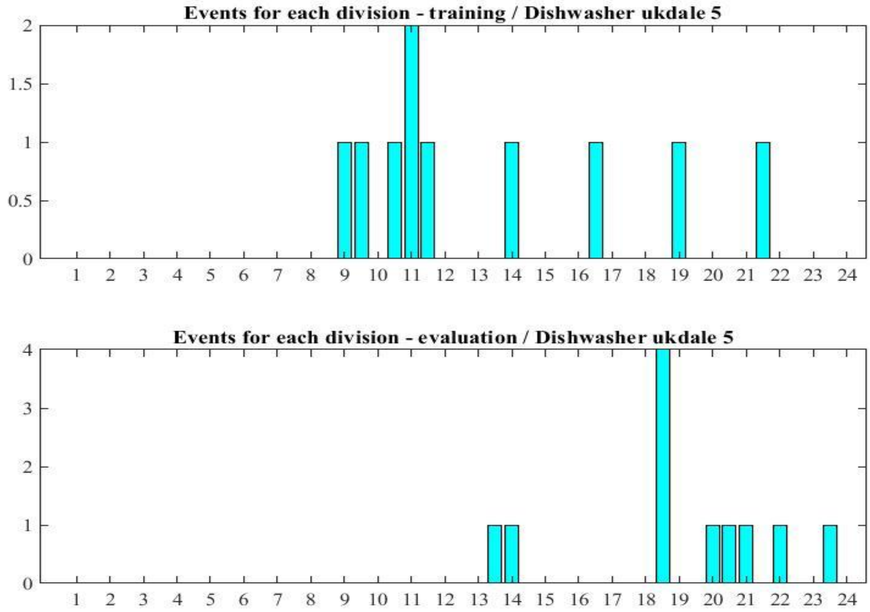


Figure J.13

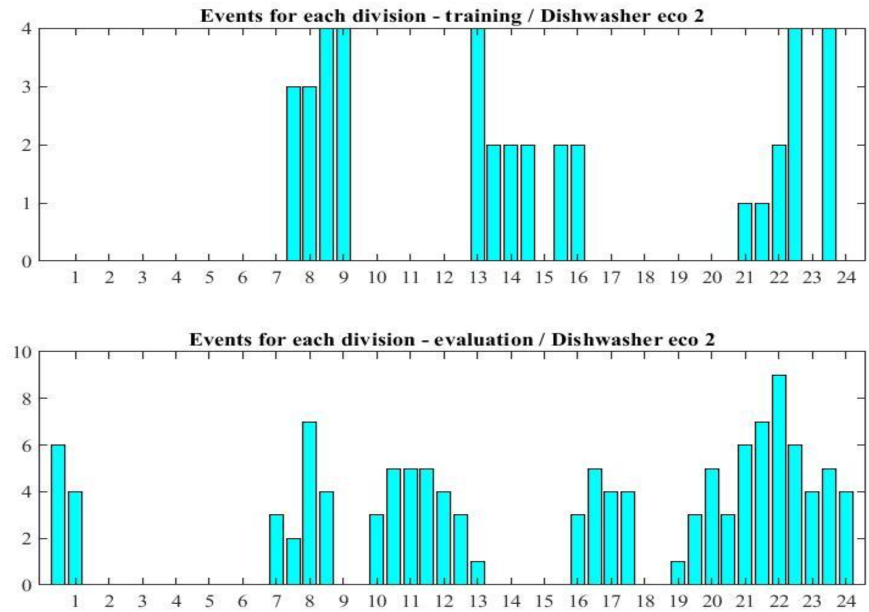


Figure J.14

J.3 WASHING MACHINES

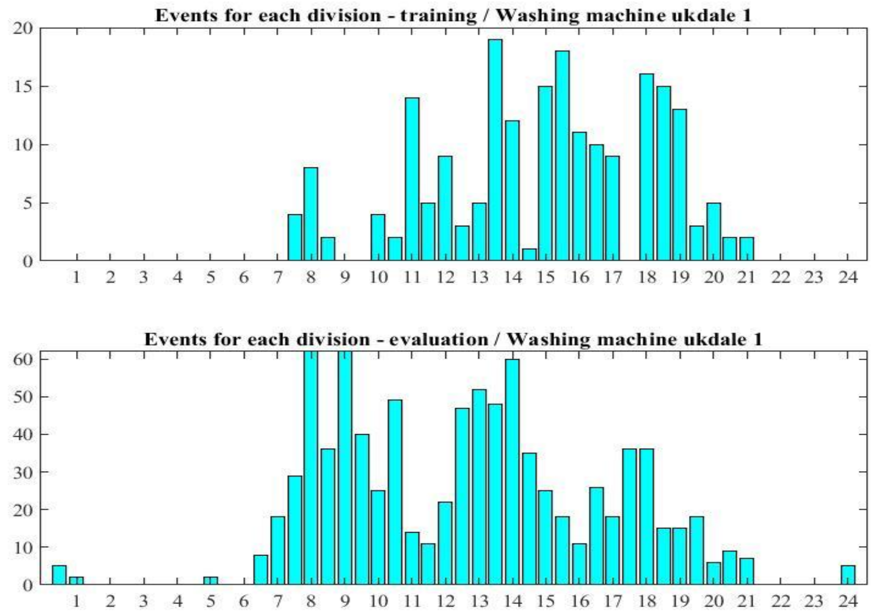


Figure J.15

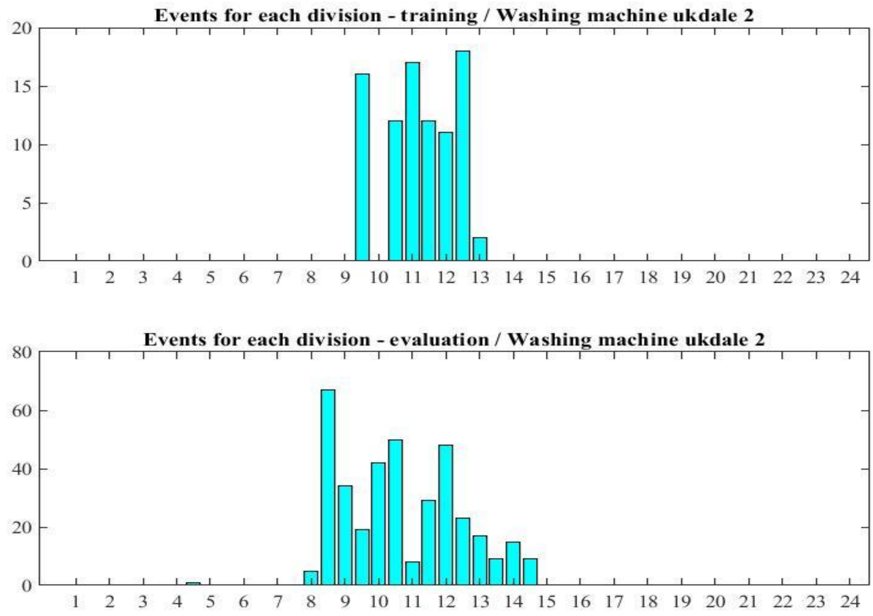
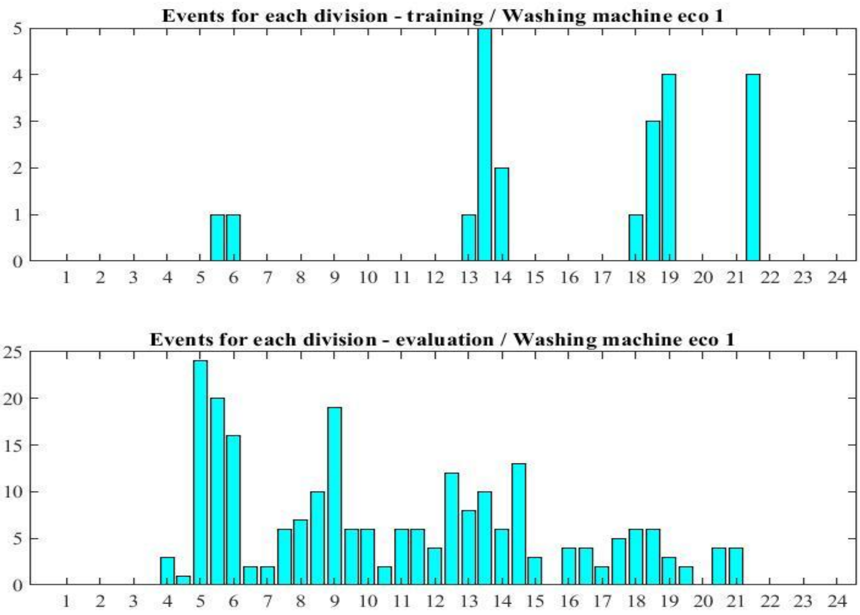


Figure J.16



Events for each division - training / Washing machine eco 1
Events for each division - evaluation / Washing machine eco 1

Figure J.17