Non-intrusive load monitoring by using active and reactive power in additive Factorial Hidden Markov Models

Non-Intrusive Load Monitoring by Using Active and Reactive Power in Additive Factorial Hidden Markov Models

Roberto Bonfigli*, Emanuele Principi, Marco Fagiani, Marco Severini, Stefano Squartini, Francesco Piazza

Department of Information Engineering, Università Politecnica delle Marche, Via Brecce Bianche, 60131, Ancona, Italy

Abstract

Non-intrusive load monitoring (NILM) is the task of determining the appliances individual contributions to the aggregate power consumption by using a set of electrical parameters measured at a single metering point. NILM allows to provide detailed consumption information to the users, that induces them to modify their habits towards a wiser use of the electrical energy. This paper proposes a NILM algorithm based on the joint use of active and reactive power in the Additive Factorial Hidden Markov Models framework. In particular, in the proposed approach, the appliance model is represented by a bivariate Hidden Markov Model whose emitted symbols are the joint active-reactive power signals. The disaggregation is performed by means of an alternative formulation of the Additive Factorial Approximate Maximum a Posteriori (AFAMAP) algorithm for dealing with the bivariate HMM models. The proposed solution has been compared to the original AFAMAP algorithm based on the active power only and to the seminal approach proposed by Hart (1992), based on finite state machine appliance models and which employs both the active and reactive power. Hart's algorithm has been improved for handling the occurrence of multiple solutions by means of a Maximum A Posteriori technique (MAP). The experiments have been conducted on the AMPds dataset in noised and denoised conditions and the performance evaluated by using the F_1-Measure and the normalized disaggregation metrics. In terms of F_1-Measure, the results showed that the proposed approach outperforms AFAMAP, Hart's algorithm, and Hart's with MAP respectively by $+14.9\,\%$, $+21.8\,\%$, and $+2.5\,\%$ in the 6 appliances denoised case study. In the 6 appliances noised case study, the relative performance improvement is $+25.5\,\%$, $+51.1\,\%$, and $+6.7\,\%$.

Keywords: Non-Intrusive Load Monitoring (NILM), Disaggregation, Active and Reactive power, Factorial Hidden Markov Model (FHMM), Quadratic Programming, Constrained Optimisation

*Corresponding author

Email addresses: r.bonfigli@univpm.it (Roberto Bonfigli), e.principi@univpm.it (Emanuele Principi), m.fagiani@univpm.it (Marco Fagiani), m.severini@univpm.it (Marco Severini), s.squartini@univpm.it (Stefano Squartini), f.piazza@univpm.it (Francesco Piazza)

1. Introduction

In the recent years, the public awareness on energy saving themes has been constantly increasing. Indeed, the consequences of global warming are now tangible and studies have demonstrated that they are directly related to humans activities and their inefficient use of energy and natural resources [1–3]. The response of governments and public institutions to counteract this trend is to promote policies for reducing energy waste and intelligently use natural resources. The electricity grid is a key component in this scenario: the original electromechanical grid, where the information flow was one-directional, is transforming into the new digital *smart grid* [4] where the information flows from the energy provider to distributed sensors and generator stations and vice-versa. Part of this change involves the integration of smart meters in the grid in order to provide detailed consumption information both to the consumers and to the energy provider.

Indeed, recent studies demonstrated that this fine-grained information is able to provide significant energy savings [5]. On the consumers side, the knowledge of the energy consumption of individual appliances establishes a virtuous behaviour towards a wiser use of electric energy [6, 7]. Studies showed that this can lead to savings greater than 12% with specific appliance feedback and personalised recommendations [5, 8–10]. On the energy provider side, fine-grained information enables the prediction of the power demand, the application of management policies and the prevention of overloading or blackouts over the energy network [11].

Providing detailed consumption information without installing several dedicated meters requires intelligent methods able to infer the energy consumed by individual appliances with minimal metering points. Non-intrusive load monitoring (NILM) denotes the class of methods and algorithms able to perform this task by using the electrical parameters measured in a single-point [5, 12, 13]. Originally developed in the seminal work by Hart [14], NILM has been an active area of research in the last years. The most promising approaches recently presented in the literature are based on machine learning algorithms, and their general scheme consists in extracting significant features from the measured electrical parameters and then estimating the appliance specific active power signal by using a supervised or unsupervised algorithm [12, 15, 16].

The electrical parameters are usually represented by active power, reactive power, voltage or current. The majority of appliances is characterised by a finite number of operating states [14] and they are analysed by observing the values of the signals *after* a state transition has completed (*steady state* approach [14, 17–28]) or *during* a state transition (*transient state* approach [29–34]). Usually, the latter requires higher sampling rates with respect to steady state analysis (respectively in the order of kHz and of Hz) and more complex and costly hardware equipment [16]. This explains why the scientific community devoted particular attention to steady state approaches.

The necessity of the user intervention for creating appliance models distinguishes supervised from unsupervised approaches. The first implies the availability of the individual signals of each appliance. In a real

operating scenario, this translates into requiring support by the user, that should sequentially switch on the appliance of interest and switch off the remaining [14]. In a recent work by the authors [35], this requirement has been partially reduced by allowing selected appliances (e.g., the fridge) to remain operational while signatures of the other appliances are being created.

Unsupervised techniques provide the means to automate the learning process, thus being completely transparent to the user [15]. Furthermore, they are capable of dynamically adapting to the power system changes over time (i.e., addition, removal, or substitution of appliance) [36]. However, their major shortcoming is represented by the inability to apply an appropriate label to the disaggregated signals. Different approaches try to overcome these limitations by exploiting the information contained on a generic labelled dataset and generalising to unseen household data by using an unsupervised algorithm [21].

As aforementioned, machine learning techniques have become a popular choice for NILM, since they showed significant disaggregation performance: in particular Factorial Hidden Markov models (FHMMs) [17–22, 26, 28, 37–41], Neural Networks (NN) [23–25, 32], graph-based signal processing [27], Support Vector Machines (SVM) [42], k-Nearest Neighbours [42], and Decision Trees [43] have been successfully employed for NILM.

This paper proposes a disaggregation algorithm based on FHMMs and active and reactive power measured at low sampling rates. The paper describes the HMM models of the appliances and the proposed solution for obtaining their parameters from a training dataset. Load disaggregation is performed by proposing a reformulated version the Additive Factorial Approximate Maximum a Posteriori (AFAMAP) algorithm [18] that allows a straightforward extension to the bivariate case. The experimental evaluation has been conducted on the Almanac of Minutely Power dataset (AMPds) dataset [44] in noised and denoised scenarios, and the proposed solution has been compared to AFAMAP based on the active power only and to two variants of Hart's algorithm [14] both based on active and reactive power. The results show that in terms of F_1-Measure the proposed approach provides a significant performance improvement with respect to the comparative methods.

The remainder of this section provides an overview of recent works on NILM based on FHMMs and illustrates the contribution of this paper with respect to them.

1.1. Related work

Among unsupervised approaches, the ones based on FHMMs have been devoted particular attention in the last years. One of the earliest work on the topic has been presented in [17] by Kim and colleagues. The key idea is to model each appliance with independent parallel HMM each contributing to the aggregate power. The framework is assessed by using the steady-state real power signal, but it allows multidimensional features as input. In [21, 22], the authors employ HMMs in a Bayesian framework in order to combine multiple models and form a general model of an appliance. Labelled data are required in the training phase

and then appliance specific models are tuned on aggregate data without requiring user intervention. In the literature, particular attention has been devoted to the algorithm proposed by Kolter and Jaakkola [18], since it showed noteworthy performance with a reasonable computational complexity. The Additive Factorial Approximate Maximum a Posteriori (AFAMAP) algorithm is an efficient method, based on an optimization problem, for the inference of the working states combination in the Factorial Hidden Markov Model framework. The authors introduced the AFAMAP algorithm, where they constrain the posterior probability to require only one HMM change state at any given time. Semi-Markov models are combined with Hierarchical Dirichlet Process in [40] for inferring both the state complexity of the models and the duration of the distributions. The authors use the active power as input feature and evaluate the performance on the five most consuming appliances of the REDD dataset [41]. Makonin and colleagues in [39] proposed the sparse Viterbi algorithm for disaggregating the active power online and in real-time. Sparse Viterbi exploits the matrix sparsity in HMMs and it was evaluated on the AMPds [44] and REDD [41] datasets. Aiad and Lee [45] augmented FHMMs with additional chains for modelling possible interactions among the appliances. The algorithm operates on the active power input feature and it was evaluated on the REDD dataset. The work in [26] introduces an FHMM model with unbounded number of chains, and states for each chain as well. In [38] the authors introduce Hierarchical FHMM with the aim of overcoming the device independence assumption and the one-at-time condition. The algorithm operates on the steady-state active power signal by clustering the signals of correlated devices and then by training HMM models on the identified clusters (denoted as "super devices"). In the disaggregation phase, inference is performed with AFAMAP on the super devices, and the result is mapped back to the original device by using the state relation table learned during the training phase. Compared to the original AFAMAP algorithm on the REDD and Pecan datasets, the method proposed by the authors provides significant performance improvements. Zhong *et al.*, [19] incorporate domain knowledge in the FHMM in the form of signal aggregate constraint. In the NILM scenario, this translates into constraining the total energy consumed in a day by an appliance to be close to a predefined value. The algorithm was assessed on the Household Electricity Survey dataset and compared to the Additive Factorial HMM and the AFAMAP algorithms. The results showed that the method indeed achieves better performance in terms of disaggregation error. In a different work [20], the same authors introduce interleaved factorial non-homogeneous hidden Markov model (IFNHMM), where the transition probabilities of the models are supposed time variant in order to represent the different pattern of usage of an appliance during the day. In addition, at each time step only one chain is allowed to change. The algorithm presented in [28] combine FHMM and Subsequence Dynamic Time Warping (SDTW). The FHMM is employed in the first stage to identify only the ON and OFF state of each appliance. SDTW, then, is applied iteratively to extract the final output. The authors propose both a supervised and semi-supervised version of the algorithm, with the latter employing the aggregate signal and consumption diaries to extract

the appliance signatures.

The works presented so far perform load disaggregation by using the active power as the only input feature. Differently, in [37], the authors propose a structural variational approximation method and they evaluated the combination of five features: active and reactive power, power factor, and the active and reactive power standard deviation calculated in a window of five samples. The algorithm is evaluated in a "denoised scenario", for different combinations of low-power appliances (e.g., laptop, desk lamp, LCD monitor). Instead of using only electrical parameters, in [46] the authors proposed the inclusion of contextual information represented by the timing-usage statistics and the presence of the user in the house. The disaggregation algorithm is based on of AFAMAP and Conditional FHMMs, and the experiments are conducted on the Tracebase dataset augmented with synthetic contextual information.

1.2. Contribution

A part from [37], the aforementioned approaches employ the active power as the sole electrical parameter for NILM, despite some algorithmic frameworks have been formulated for operating on multidimensional feature vectors [18]. The use of reactive power has been employed since the very first work by Hart [14] and in more recent works based on the same principles [47–51] or on transient-state analysis [29–31, 33, 34]. However, up to the authors' knowledge, the only work that employs both the active and the reactive power in the FHMM framework is the work by Zoha and colleagues [37].

Following a similar philosophy, in this paper we propose a disaggregation algorithm based on FHMMs that uses both the active and reactive power. However, differently from [37], where the disaggregation algorithm is based on the structural variational approximation method and on the Viterbi algorithm, in the proposed approach the active power is disaggregated by reformulating the AFAMAP algorithm for the bivariate case. As demonstrated in [18], this allows the introduction of a Differential FHMM (DFHMM) that improves the performance and reduces the computational cost. Thus, differently from [37], here the reactive power component is introduced also in the DFHMM. More in details, the proposed solution belongs to the family of supervised approaches based on steady state signals acquired from low frequency measurements. The reactive power is introduced in the FHMM framework by employing bivariate hidden Markov appliance models whose emitted symbols are represented by active and reactive power pairs. Differently from [37], here we describe the entire procedure for obtaining the bivariate HMM appliance models. The parameters are estimated by clustering the appliance disaggregate signals and the bivariate optimisation problem is solved by proposing an alternative formulation of AFAMAP [18] for disaggregating appliances consumption profiles. The proposed approach differs from the one presented in [52] by some of the authors, since there the reactive power was employed alone in an initial disaggregation stage whose output served as a constraint for the subsequent disaggregation of the active power only. The proposed approach has been compared to the original AFAMAP algorithm [18], which employs the active power only, and to Hart's algorithm

[14], which employs both the active and reactive power. Two implementations of Hart's algorithm have been developed for dealing with the occurrence multiple appliance combinations: in the first, the final combination is selected randomly. In the second, it is selected by choosing the most probable combination calculated on a training set. The experiments have been conducted on the Almanac of Minutely Power dataset (AMPds) [44], containing recordings of consumption profiles belonging to a single home for a period of two years at 1 minute sampling rate. Both the "noised" and the "denoised" scenarios have been addressed, and the results show that the proposed approach outperforms both AFAMAP and Hart's algorithm.

The outline of the paper is the following: Section 2 describes the proposed bivariate HMM appliance model. Section 3 provides the details of the proposed disaggregation algorithm, whereas Section 4 describes the Hart's approach in the bivariate problem. The experimental setup and the related results are discussed in Section 5. Finally, Section 6 concludes this contribution and presents future developments.

2. Appliance modelling

This section firstly presents the bivariate HMM appliance model employed in the disaggregation algorithm, then it describes the procedure adopted to estimate its parameters.

2.1. Bivariate hidden Markov model

Each appliance is modelled as a bivariate HMM, i.e., an HMM whose emitted symbols are represented by active-reactive power pairs. More in details, each HMM is represented by the following parameters [53]:

- the number of states $m \in \mathbb{Z}_+$;
- the hidden states $x \in \{1, 2, \ldots, m\}$;
- the symbols emitted $\boldsymbol{\mu}_j \in \mathbb{R}^n$, where $j = 1, \ldots, s$;
- the symbol emission probability matrix $\boldsymbol{M}^{s \times m}$;
- the state transition probability matrix $\boldsymbol{P} \in [0, 1]^{m \times m}$;
- the starting state probability vector $\boldsymbol{\phi} \in [0, 1]^m$.

In the algorithm, it is assumed that each state of the HMM corresponds to a working state of the appliance, i.e., $x \in \{\mathrm{ON}_1, \mathrm{ON}_2, \ldots, \mathrm{OFF}\}$, so that the number of states m is equal to the number of symbols s and $\boldsymbol{M} \equiv \boldsymbol{I}^{m \times m}$ (*degenerate HMM*). In the proposed approach, $n = 2$ and for the sake of clarity in the remainder of this paper it will be omitted since the individual active and reactive power components will be made explicit. For example, each symbol is defined as $\boldsymbol{\mu}_j = [\mu_{a,j}\, \mu_{r,j}]^T$, where the subscripts a and r distinguish the active and reactive components. For each appliance, the quantities to be estimated are

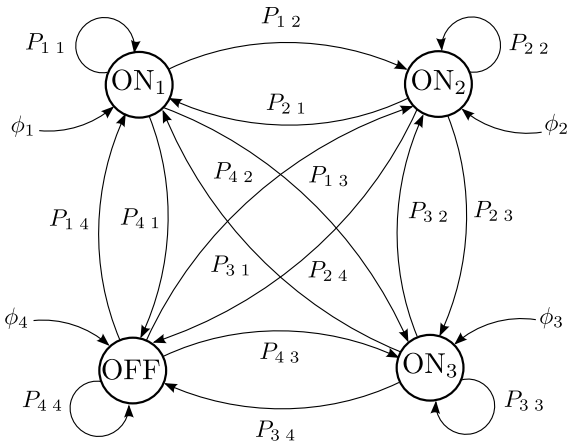


Figure 1: An example of a four states HMM.

the number of states m, the values of $\boldsymbol{\mu}_j$ for each state, the state transition probability matrix $\boldsymbol{P}$, and the starting state probability vector $\boldsymbol{\phi}$. Estimation of m and of $\boldsymbol{\mu}_j$ will be addressed in Section 2.2.

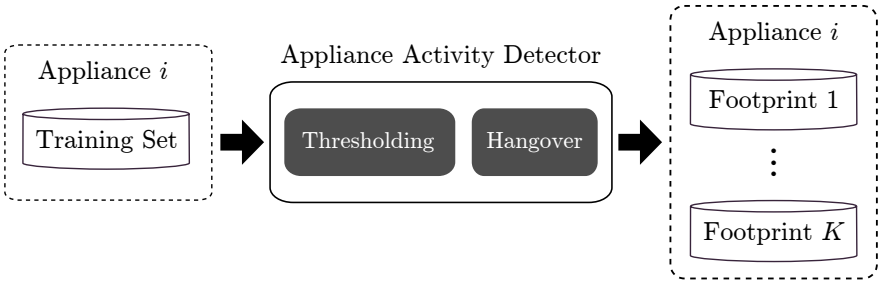
Regarding the state transition probability matrix $\boldsymbol{P}$, each entry P_{ij} represents the probability of transitioning from state i to state j. Thus, P_{ij} can be estimated with a Maximum Likelihood criterion by calculating the number of times state i transitions to state j and normalising by the total number of transitions from state i. Formally:

$$P_{ij} = \frac{C_{ij}}{\sum_{j'=1}^{m} C_{ij'}}, \tag{1}$$

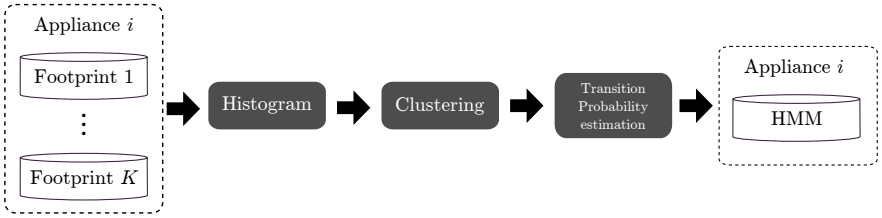
where C_{ij} is the number of transitions from state i to state j. Typically, the greatest values in the matrix are located in the diagonal, meaning that the probability of remaining in the same state is higher compared to the probability of transitioning to another state. As might be expected, the greatest value of the transition matrix is the self-transition probability of the OFF state, since the activation of an appliance occurs after a long time in which it is turned off. In addition, the OFF state corresponds to the initial state, since the footprint starts just before the turning on instant, thus $\boldsymbol{\phi} = [0\,0\cdots 0\,1]^T$.

An example of a four states appliance model is shown in Figure 1, where the arc between two states is the probability of transition P_{ij}, while the arc starting e closing on the same state represent the probability P_{ii} of permanence in each state.

A probability value close to zero denotes that the transition is very unlikely. In practice, it is recommended to avoid such low probability values, since evaluating it in the log domain as usually done in the disaggregation algorithm would result in numeric problems. As so, it is recommended to fix the value to a little quantity, e.g.,$\simeq 10^{-5}$.



(a) Footprint extraction block scheme.



(b) Model training block scheme.

Figure 2: Diagram of the footprint extraction procedure (a) and of the training phase of the appliance models (b).

2.2. Model training

The working states power level estimation consists in obtaining representative power level distributions related to each appliance state, i.e., the values of the emitted symbols $\boldsymbol{\mu}_j$. In a realistic scenario, this is obtained by using a set of examples of an appliance typical consumption cycle. This information can be extracted by observing the aggregate power signal, under the assumption that only one appliance at time is operating [14].

In particular, this stage involves the extraction of a *footprint* of the appliance, i.e., the active and reactive power signals comprised between the power on (transition from the OFF state to an ON state) and the power off (transition from an ON state to the OFF state). This is performed by firstly identifying these instants by means of an *Appliance Activity Detector* (AAD). Basically, it consists in detecting when the active power level signal exceeds a certain threshold or not (typical values are in the order of 20 W). Isolated occurrences of power levels below the threshold are managed by employing a *hangover* technique: it is a counter, which decreases its value for each sample the signal is below the threshold. If the signal returns over the threshold before the end of the counter, the footprint is considered continued. The typical value is 5-10 minutes. The diagram of the footprint extraction stage is shown in Figure 2a.

The power value and the temporal information of the OFF state cannot be obtained by analysing the signal extracted with the AAD. The value is reasonably assumed 0 W and 0 VAR for the active and reactive power signals, respectively. The temporal information, i.e., the typical interval intercurring between the OFF state and a ON state, has to be specified a-priori for each appliance based on the typical usages (e.g., once in an hour, three times in a day, etc.).

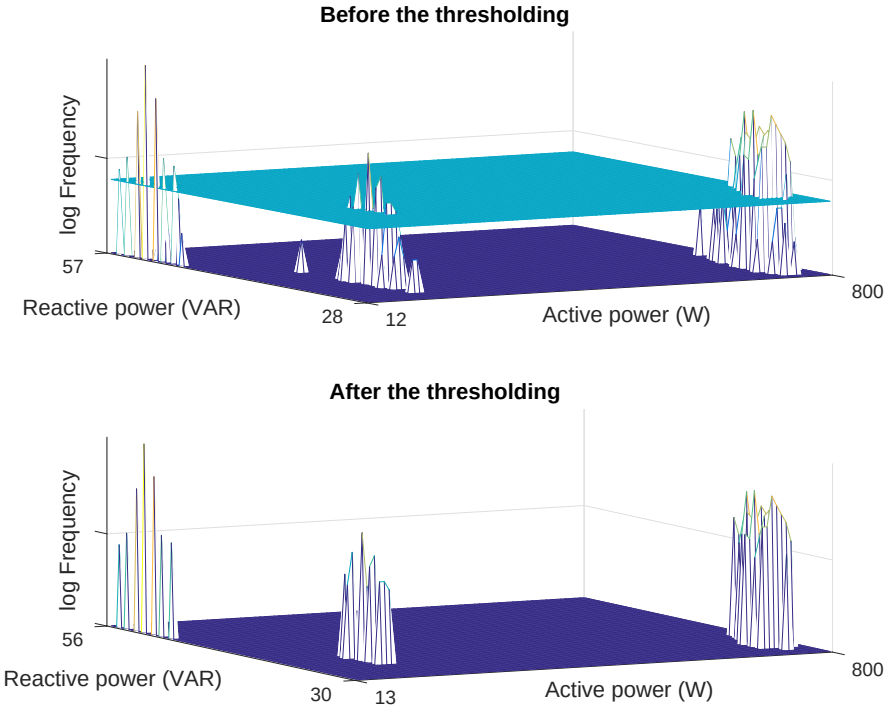


Figure 3: An example of a two-dimensional histogram of the active and reactive power signals related to the dishwasher in the dataset AMPds.

Complex appliances (e.g., washing machines, dishwashers) are characterised by several working cycles and the extraction of a single footprint might not be completely representative of its operation. This motivates the need to acquire several footprints for each appliance. Furthermore, even though only one footprint is sufficient to explore all the working states of an appliance, multiple footprints allow to employ more data for the power level extraction phase, particularly useful for those power levels characterised by a short duration.

The estimation of the power level associated to a state of the HMM relies on the appliance consumption data, which is not composed of discrete values of consumption, but it presents a continuous variability in the values. In order to find the averaged values of the signal, within the period of permanence in the same working state, a clustering procedure is adopted, and the k-means [54] has been selected as the algorithm.

Since the OFF state information is not present in the data, the number of clusters is set to $(m-1)$. After identifying the clusters, the power levels associated to each HMM state are represented by their centroids.

The clustering operation is not directly performed on the footprints extracted with the AAD. Indeed, after extracting the footprint, a bivariate histogram composed of 100 bins per kW and per kVAR is used to analyse the probability distribution of the active and reactive power signals. The number of bins is empirically chosen after analysing some footprints of the training set in order to obtain a sufficiently detailed

histogram able to provide a good trade-off between variance and bias of the density estimate. Additionally, power levels with a low number of occurrences are excluded from the successive processing. More in details, bins having a number of occurrences below the threshold are considered of lower relevance, thus the related observations are discarded. This technique allows to obtain the number of working states m, which is determined by observing the number of clusters obtained in the final bivariate histograms. An example is shown in Figure 3, where the histogram before and after the thresholding operation is shown. It refers to the dishwasher consumption in the AMPds dataset. Additionally, it reduces a limitation of the clustering algorithm: k-means does not employ the information on the samples distribution in the cluster, since it selects the centroid which satisfies the rule of convergence over all data. Discarding bins with low occurrences forces k-means to select the centroids with higher probability and to discard local clusters with lower probability, that could result in erroneous centroids. Furthermore, it allows to discriminate close clusters which can be confused as a single one: indeed, transients between near clusters produce samples comprised between the cluster with higher occurrences, which merge the two clusters in a single one. The diagram of the clustering and of the model training stage is shown in Figure 2b.

In general, clusters present different characteristics depending on the magnitude of their centroid. Typically, the ones characterised by high values (e.g., 3000 W) are highly variable, since they depend on the appliance usage by the user, e.g., the water temperature chosen in the washing machine or the rinsing cycle of the dishwasher affect the maximum power consumption. On the other hand, clusters characterised by low power value (e.g., 300 W) have lower variability, since deviation from the centroid is mainly caused by intermediate working stages of the appliance, and they do not depend on the usage.

Figure 4 shows an example of the inference procedure conducted on the active power signal only, denoted as P_a, and on the joint active-reactive power signals, denoted as (P_a, P_r). The signals are related to the washing machine in the AMPds dataset. In particular, Figure 4a shows the active power signal and the cluster membership of each sample when k-means operates on the P_a signal only. Figure 4b and Figure 4c show respectively the same active power signal and the reactive power signal, but the cluster membership is related to the outcome of k-means operating on the joint (P_a, P_r). Figure 4d shows at the bottom the 1-D P_a line with the clusters obtained when k-means operates on the P_a signal only and at the top the (P_a, P_r) plane with the clusters obtained when k-means operates on the joint (P_a, P_r) signals. In the figure, each cluster is depicted as an interval or as an ellipse whose size is twice the standard deviation of the cluster centered at its centroid. The number of clusters is different between the active power and the active and reactive power cases: in the first case 4 cluster can be identified, whereas the addition of the reactive power allows to distinguish 5 clusters. As shown in the figure, 2 clusters share the same value of active power, but differ in the reactive component. Using the reactive power, thus, allows to have a better representation of the working states of the appliance, therefore reducing the admissible combination of working states in the

aggregated data.

2.3. Noise model

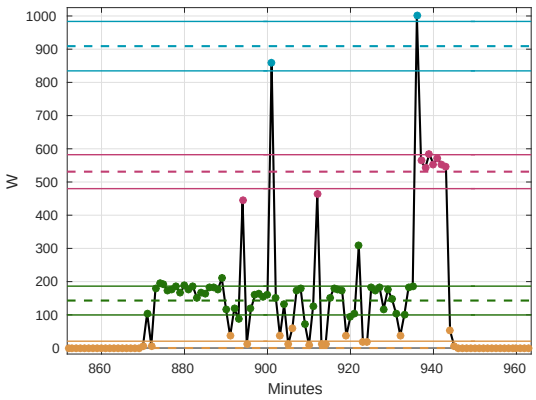
In a real case scenario, a noise contribution can be observed on the aggregated signal, due to electrical noises in the system, very small loads, leakages. This contribution can be considered as a source of power consumption, additionally to the appliances which we want to monitor, therefore it can be modelled with a HMM, as described in Section 2.2, leading to a *noise model*. The number of working states is a parameter which could depends on the application scenario, therefore it has to be explored in the experimental phase, nevertheless it would be grater than the number of states defined for the appliances, since it represents a set of multiple load working at the same time. The data used for training this model can be extracted by observing the aggregate power signal, when all the appliances of interest are switched off and all the remaining equipments in the house are working. Since no operating cycle or footprint is defined in this case, this model will no be provided by the OFF state.

3. Disaggregation algorithm

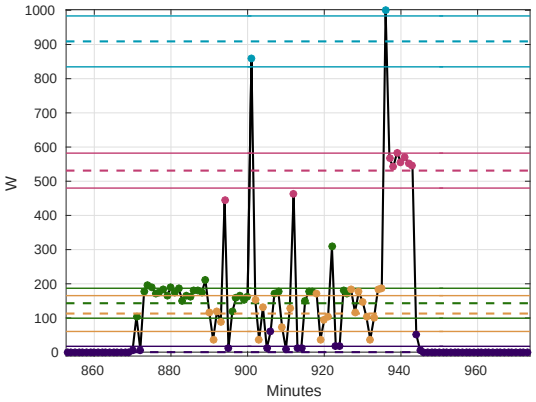
The general scheme of the disaggregation phase is shown in Figure 5. The algorithm is based on the work proposed by Kolter and Jaakkola [18], where the problem is modelled in the Additive Factorial Hidden Markov Model (AFHMM) framework.

Basically, this consists in modelling the value of each aggregated power sample as a combination of working states of the appliances. In [18], an assumption is made that at most one HMM changes its state at any given time, which holds true if the sampling time is reasonably short. In this case, a transition on the aggregate power, when moving from a sample to the next, corresponds to a state change of a particular HMM. As a consequence, a differential signal can be modelled as the result of a Differential Factorial Hidden Markov Model (DFHMM), which relies on the same HMM models composing the AFHMM. The DFHMM models the observation output as the difference between the states combination of the HMMs in two consecutive time instants. By combining the additive and differential models, the inference on the set of states of multiple HMMs can be computed through the Maximum A Posteriori (MAP) algorithm, which takes the form of a Mixed Integer Quadratic Programming (MIQP) optimisation problem. One of the shortcomings of this approach is the non-convex nature of the problem, due to the integer nature of the variables: therefore, a relaxation towards real values is taken into account, which allows the solution to assume any value in the range [0, 1], instead of the binary solution, leading to a convex Quadratic Programming (QP) optimisation problem.

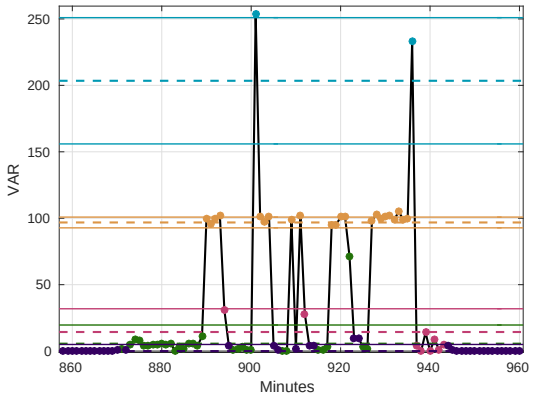
In a real case scenario, the modelled output may not match with the observed aggregated signal, due to electrical noises, very small loads, or leakages. In that case, the issue is addressed by defining a robust



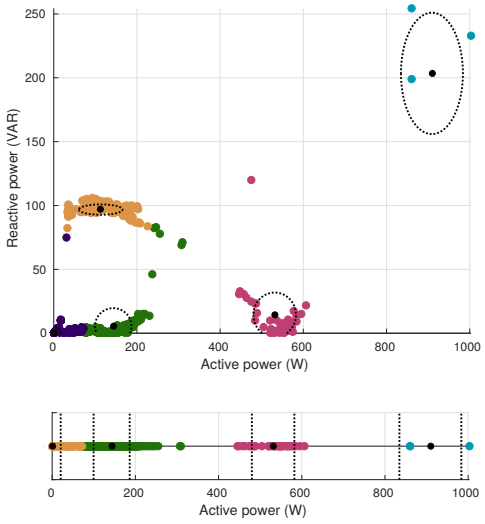
(a) Footprint (P_a) and cluster membership of each sample with k-means operating on P_a.



(b) Footprint (P_a) and related clusters with k-means operating on (P_a, P_r).



(c) Footprint (P_r) and related clusters with k-means operating on (P_a, P_r).



(d) Clusters in the (P_a, P_r) plane (above) and the P_a line (below).

Figure 4: Washing machine footprint and clusters in the dataset AMPds.

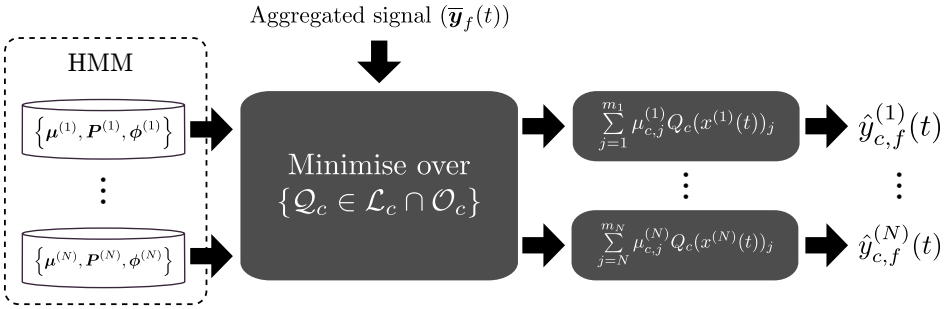


Figure 5: Diagram of the load disaggregation phase.

mixture component both in the AFHMM and in the DFHMM. This component is missing in this dissertation, since all the contributions to the aggregated power are modelled. Indeed, each appliance and the *noise* is represented by its HMM.

3.1. Algorithm formulation

In the following, the superscript (i) denotes terms related to HMM i, while subscripts a or r denote terms related to the active and reactive power component, respectively. The subscript $c \in \{a, r\}$ denotes a term related to the active or to the reactive power component. The parameters of the problem are the following:

- $N \in \mathbb{Z}_+$ is the number of HMMs in the system;
- $\overline{\boldsymbol{y}}(\tau) \in \mathbb{R}^n$ is the observed aggregate output, where $\tau = 1, 2, \ldots, \Upsilon$ is the sample index and Υ is the total number of samples;
- $\boldsymbol{\Sigma}_1 \in \mathbb{R}^{n \times n}$ is the observation covariance matrix related to the AFHMM;
- $\boldsymbol{\Sigma}_2 \in \mathbb{R}^{n \times n}$ is the observation covariance matrix related to the DFHMM;
- $\Delta\overline{\boldsymbol{y}}(\tau) = \overline{\boldsymbol{y}}(\tau) - \overline{\boldsymbol{y}}(\tau - 1)$ is the differential signal.

As aforementioned, all the contribution to the aggregated power are considered, thus:

$$\overline{\boldsymbol{y}}(\tau) = \sum_{i=1}^{N} \boldsymbol{y}^{(i)}(\tau), \tag{2}$$

where $\boldsymbol{y}^{(i)}(\tau)$ corresponds to the ground truth consumption of the appliances and the *noise*. Recalling the notation of Section 2, the parameters of the i-th HMM at the sample index τ are:

- $m_i \in \mathbb{Z}_+$ is the number of states;
- $x^{(i)}(\tau) \in \{1, \ldots, m_i\}$ is the HMM state at time instant τ, where m_i corresponds to the OFF state (if present);

- $\boldsymbol{\mu}_j^{(i)}$ is the emitted symbol in the j-th state, where $j = 1, 2, \ldots, m_i$;
- $\boldsymbol{\phi}^{(i)} \in [0,1]^{m_i}$ is the initial states probability distribution;
- $\boldsymbol{P}^{(i)} \in [0,1]^{m_i \times m_i}$ is the state transition probability matrix.

The aggregate signal $\overline{\boldsymbol{y}}(\tau)$ is analysed using non-overlapping frames of length T. Each frame $\overline{\boldsymbol{y}}_f(\tau)$, where $f = 1, 2, \ldots, F$, is defined as

$$\overline{\boldsymbol{y}}_f(\tau) = \begin{cases} \overline{\boldsymbol{y}}(\tau) & \text{if } \tau = (f-1)T + 1, \ldots, fT, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

After the analysis of all the $F = \Upsilon/T$ frames, the disaggregated signals $\hat{\boldsymbol{y}}^{(i)}(\tau)$ are reconstructed as follows:

$$\hat{\boldsymbol{y}}^{(i)}(\tau) = \sum_{f=1}^{F} \hat{\boldsymbol{y}}_f^{(i)}(\tau). \tag{4}$$

In the following, the algorithm is formulated for a single frame of the signal and for convenience, a new temporal variable t is defined with the relation $t = \tau - (f-1)T$, for $t = 1, 2, \ldots, T$, with $T \in \mathbb{Z}_+$.

In [18], the parameter n defines the problem dimensionality: the authors use only the active power data to characterise the observed aggregated signal $\overline{y}_a(t)$, therefore they assumed $n = 1$. In this work, both the active and the reactive power are used for disaggregation, therefore $n = 2$ and the problem variables are decomposed in two components:

$$\overline{\boldsymbol{y}}_f(t) = \begin{bmatrix} \overline{y}_{a,f}(t) \\ \overline{y}_{r,f}(t) \end{bmatrix}, \quad \boldsymbol{\mu}_j^{(i)} = \begin{bmatrix} \mu_{a,j}^{(i)} \\ \mu_{r,j}^{(i)} \end{bmatrix}, \tag{5}$$

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} \sigma_{a,1}^2 & \sigma_{a,r,1} \\ \sigma_{r,a,1} & \sigma_{r,1}^2 \end{bmatrix}, \quad \boldsymbol{\Sigma}_2 = \begin{bmatrix} \sigma_{a,2}^2 & \sigma_{a,r,2} \\ \sigma_{r,a,2} & \sigma_{r,2}^2 \end{bmatrix}. \tag{6}$$

Since the statistical independence between the active and reactive power components is supposed, the covariance terms $\sigma_{a,r}$ and $\sigma_{r,a}$ are zero in both $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$, and the same problem formalisation as the $n = 1$ case can be used, introducing additional variables and constraining them each other. For the generic power component c, the variables in the optimisation problem are defined as follows:

$$\mathcal{Q}_c = \left\{ \boldsymbol{Q}_c(x^{(i)}(t)) \in \mathbb{R}^{m_i}, \boldsymbol{Q}_c(x^{(i)}(t-1), x^{(i)}(t)) \in \mathbb{R}^{m_i \times m_i} \right\}. \tag{7}$$

In the vector $\boldsymbol{Q}_c(x^{(i)}(t))$, the element $Q_c(x^{(i)}(t))_j$ indicates the state assumed at time instant t, while in the matrix $\boldsymbol{Q}_c(x^{(i)}(t-1), x^{(i)}(t))$ the element $Q_c(x^{(i)}(t-1), x^{(i)}(t))_{jk}$ indicates the state transition from previous to the current time instant.

This problem statement is a reformulated version of the algorithm proposed in [18]: since the original algorithm allows to operates with multivariate dimension, the variables associated to the state represent

all the components. When only one dimension is considered, the variables $\mathcal{Q}_a$ is only associated at the active power level consumption. In this problem statement instead, the author started from the univariate formulation, and they extended the algorithm to $n = 2$ by using twice the optimisation variables, thus introducing the $\mathcal{Q}_r$ variable set, and an additional minimisation function. Moreover, the supplementary variables need to be constrained to the original ones in order to assume the same value during the optimisation process, representing the bivariate resolution problem with a univariate problem formalisation:

$$\begin{cases} Q_a(x^{(i)}(t))_j - Q_r(x^{(i)}(t))_j = 0, \\ Q_a(x^{(i)}(t-1), x^{(i)}(t))_{jk} - Q_r(x^{(i)}(t-1), x^{(i)}(t))_{jk} = 0. \end{cases} \tag{8}$$

A numerically safer definition of the constraints can be defined using a tolerance α and inequalities:

$$\begin{cases} -\alpha \leq Q_a(x^{(i)}(t))_j - Q_r(x^{(i)}(t))_j \leq \alpha, \\ -\alpha \leq Q_a(x^{(i)}(t-1), x^{(i)}(t))_{jk} - Q_r(x^{(i)}(t-1), x^{(i)}(t))_{jk} \leq \alpha, \end{cases} \tag{9}$$

where $j, k = 1, \ldots, m_i$.

Algorithm 1 The proposed disaggregation algorithm.

1: **Input**:

- $\overline{\boldsymbol{y}}_f(t)$, for $t = 1, 2, \ldots, T$;
- $\left\{\boldsymbol{\mu}^{(i)}, \boldsymbol{P}^{(i)}, \boldsymbol{\phi}^{(i)}\right\}$, for $i = 1, 2, \ldots, N$;
- $\sigma^2_{c,1}$, $\sigma^2_{c,2}$;
- λ: regularisation parameter, described in [18].

2: **Minimise over** $\{\mathcal{Q}_c \in \mathcal{L}_c \cap \mathcal{O}_c\}$

$$\sum_{c\in\{a,r\}} \left\{ \frac{1}{2\sigma^2_{c,1}} \sum_{t=1}^{T} E'_c(t) + \frac{1}{2\sigma^2_{c,2}} \sum_{t=2}^{T} E''_c(t) + \frac{1}{2} \sum_{t=2}^{T} E'''_c(t) + \right.$$
$$\left. + \sum_{t=2}^{T} \sum_{i=1}^{N} \sum_{\substack{j=1 \\ k=1}}^{m_i} \left\{ Q_c(x^{(i)}(t-1), x^{(i)}(t))_{jk} \left(-\log P^{(i)}_{kj} \right) \right\} + \sum_{i=1}^{N} \sum_{j=1}^{m_i} \left\{ Q_c(x^{(i)}(1))_j (-\log \phi^{(i)}_j) \right\} \right\} \tag{10}$$

3: **Output**:

$$\hat{y}^{(i)}_{c,f}(t) = \sum_{j=1}^{m_i} \mu^{(i)}_{c,j} Q_c(x^{(i)}(t))_j \tag{11}$$

where $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$.

The final algorithm is shown in Algorithm 1. In eq. (10), the error terms are defined as:

$$E'_c(t) = \left(\overline{y}_{c,f}(t) - \sum_{i=1}^{N} \sum_{j=1}^{m_i} \mu_{c,j}^{(i)} Q_c(x^{(i)}(t))_j \right)^2, \tag{12}$$

$$E''_c(t) = \sum_{i=1}^{N} \sum_{\substack{j=1 \\ k=1 \\ k \neq j}}^{m_i} \left\{ \left(\Delta \overline{y}_{c,f}(t) - \Delta \mu_{c,kj}^{(i)} \right)^2 Q_c(x^{(i)}(t-1), x^{(i)}(t))_{jk} \right\}, \tag{13}$$

$$E'''_c(t) = D\left(\frac{\Delta \overline{y}_{c,f}(t)}{\sigma_{c,2}}, \lambda \right) \left(1 - \sum_{i=1}^{N} \sum_{\substack{j=1 \\ k=1 \\ k \neq j}}^{m_i} Q_c(x^{(i)}(t-1), x^{(i)}(t))_{jk} \right). \tag{14}$$

The QP optimisation problem is defined as follows:

Minimise

$$\frac{1}{2} \boldsymbol{v}^T \boldsymbol{H} \boldsymbol{v} + \boldsymbol{f}^T \boldsymbol{v}, \tag{15}$$

subjet to the constraints:

$$\boldsymbol{A}_{eq}\, \boldsymbol{v} = \boldsymbol{b}_{eq}, \tag{16}$$

$$\boldsymbol{lb} \leq \boldsymbol{v} \leq \boldsymbol{ub}. \tag{17}$$

The variables of the problem are represented by the vector $\boldsymbol{v} = [\boldsymbol{v}_a\, \boldsymbol{v}_r]^T$ whose components are defined as follows:

$$\boldsymbol{v}_c = \begin{bmatrix} \boldsymbol{\Theta}(1) \\ \vdots \\ \boldsymbol{\Theta}(T) \end{bmatrix}, \; \boldsymbol{\Theta}(t) = \begin{bmatrix} \boldsymbol{\Psi}^{(1)}(t) \\ \vdots \\ \boldsymbol{\Psi}^{(N)}(t) \end{bmatrix}, \; \boldsymbol{\Psi}^{(i)}(t) = \begin{bmatrix} \boldsymbol{\xi}^{(i)}(t) \\ \boldsymbol{\beta}^{(i)}(t) \end{bmatrix}, \tag{18}$$

$$\boldsymbol{\xi}^{(i)}(t) = \begin{bmatrix} Q_c(x^{(i)}(t))_1 \\ \vdots \\ Q_c(x^{(i)}(t))_{m_i} \end{bmatrix}, \quad \boldsymbol{\beta}^{(i)}(t) = \begin{bmatrix} Q_c(x^{(i)}(t-1), x^{(i)}(t))_{1\,1} \\ \vdots \\ Q_c(x^{(i)}(t-1), x^{(i)}(t))_{1\,m_i} \\ \vdots \\ Q_c(x^{(i)}(t-1), x^{(i)}(t))_{m_i\,1} \\ \vdots \\ Q_c(x^{(i)}(t-1), x^{(i)}(t))_{m_i\,m_i} \end{bmatrix}, \tag{19}$$

where the variables for the state are represented in $\boldsymbol{\xi}^{(i)}(t)$, and the variables for the transition in $\boldsymbol{\beta}^{(i)}(t)$.

The parameters of the problem, e.g., the HMMs parameters and the aggregated power signal, compose the elements of $\boldsymbol{H}$ and $\boldsymbol{f}$, according to the structure of the $\boldsymbol{v}$ vector. In a QP problem, the coefficient of the

quadratic terms in the cost function are defined in $\boldsymbol{H}$, as a symmetric matrix. In the proposed approach, since the independence between the active and reactive power is assumed, there are no joint quadratic terms, therefore $\boldsymbol{H}$ is structured as follow:

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{H}_a & \mathbf{0} \\ \mathbf{0} & \boldsymbol{H}_r \end{bmatrix}. \tag{20}$$

Differently, the coefficients of the linear terms are expressed in $\boldsymbol{f} = [\boldsymbol{f}_a \ \boldsymbol{f}_r]^T$. Whereas $\boldsymbol{A}_{eq}$ and $\boldsymbol{b}_{eq}$ are used to represent the consistent constraints between the state and the transition variables. The vectors $\boldsymbol{lb}$ and $\boldsymbol{ub}$ define the lower and upper boundaries of the solution: because of the nature of the variables [18], the lower boundary is equal to 0, whereas the upper boundary to 1, for all the elements in $\boldsymbol{v}$.

Additional constraints to QP problem need to be considered, in order to impose the inequality constraints between the optimisation variables. Duplicating the constraints of eq. (9):

$$\begin{cases} -\alpha \leq Q_a(x^{(i)}(t))_j - Q_r(x^{(i)}(t))_j, \\ Q_a(x^{(i)}(t))_j - Q_r(x^{(i)}(t))_j \leq \alpha, \end{cases} \tag{21}$$

$$\begin{cases} -\alpha \leq Q_a(x^{(i)}(t-1), x^{(i)}(t))_{jk} - Q_r(x^{(i)}(t-1), x^{(i)}(t))_{jk}, \\ Q_a(x^{(i)}(t-1), x^{(i)}(t))_{jk} - Q_r(x^{(i)}(t-1), x^{(i)}(t))_{jk} \leq \alpha, \end{cases} \tag{22}$$

results in the following optimisation constraint:

$$\boldsymbol{A}_{ineq}\, \boldsymbol{v} \leq \boldsymbol{b}_{ineq}. \tag{23}$$

This is needed only for the joint active-reactive problem, since, solving only for the active power, the related unique variable is not constrained to other variables. Indeed, in eq. (10) only the the active power terms need to be considered. Further details on the terms $\boldsymbol{H}$, $\boldsymbol{f}$, $\boldsymbol{H}$, $\boldsymbol{A}_{eq}$, $\boldsymbol{b}_{eq}$, $\boldsymbol{lb}$, $\boldsymbol{ub}$, $\boldsymbol{A}_{ineq}$, and $\boldsymbol{b}_{ineq}$ are provided in Appendix A.

3.2. Algorithm operation

As aforementioned, the aggregate signal is analysed in frames of length T. In the first frame, the value of starting probability vector $\boldsymbol{\phi}^{(i)} = [0\ 0\ \cdots\ 0\ 1]$, i.e., the appliance is initially assumed in the OFF state. In the subsequent frames, the value of $\boldsymbol{\phi}^{(i)}$ depends on the last state assumed in the previous frame in order to ensure the contiguity of the solution at the border. Thus, if the last state assumed in the previous frame is j, the corresponding element of $\boldsymbol{\phi}^{(i)}$ is set to 1, while the others are set to 0. This information is represented by the value of the solution $\boldsymbol{\xi}^{(i)}(t)$ in the last sample $t = T$.

4. Comparative method

The proposed approach has been compared with the algorithm presented by Hart in [14], since it employs both the active and the reactive power to model the appliance working behaviour and it employs those

electrical parameters for disaggregation. This section provides an overview of its basic operating principles as well as additional details on its implementation. In addition, the algorithm originally presented in [14] has been improved for handling the occurrence of multiple solutions by means of a MAP technique.

Hart's algorithm models each appliance as a Finite State Machine (FSM). Each FSM is represented by the following parameters:

- the number of states $m \in \mathbb{Z}_+$;
- the finite states $x \in \{1, 2, \ldots, m\}$;
- the symbols emitted $\boldsymbol{\mu}_j \in \mathbb{R}^n$, where $j = 1, \ldots, m$;
- state transition matrix $\boldsymbol{T} \in \{0, 1\}^{m \times m}$.

As in the proposed approach, each state of the FSM corresponds to a working state of the appliance and $n = 2$, i.e., the symbol emitted in the j-th state is defined as $\boldsymbol{\mu}_j = [\mu_{a,j}\, \mu_{r,j}]^T$. A tolerance parameter $\boldsymbol{\beta}_j = [\beta_{a,j}\, \beta_{r,j}]^T$ is associated to the emitted symbol in the j-th state, in order to define the effectiveness interval for the emitted symbol. The interval width is $2\,\boldsymbol{\beta}_j$ and it is centred in $\boldsymbol{\mu}_j$. For each appliance, the quantities to be estimated are the number of states m, the values of $\boldsymbol{\mu}_j$ and $\boldsymbol{\beta}_j$ for each state, and the state transition matrix $\boldsymbol{T}$.

In order to model the power consumption of an appliance as a stochastic process, under the assumption of multiple independent causes to the circuital power dissipation, the central limit theorem might be invoked. Therefore, the power consumption $\boldsymbol{y}^{(i)}(t)$ of the i-th appliance at time instant t, related to the working state $x^{(i)}(t)$, can be modelled as a bivariate Gaussian variable, described by a mean vector $\boldsymbol{\mu}_{x^{(i)}(t)}$ and a covariance matrix $\boldsymbol{\Sigma}_{x^{(i)}(t)}$:

$$\boldsymbol{y}^{(i)}(t) | x^{(i)}(t) \sim \mathcal{N}\left(\boldsymbol{\mu}_{x^{(i)}(t)}, \boldsymbol{\Sigma}_{x^{(i)}(t)}\right). \tag{24}$$

Following this approach, the consumption signal is replaced by a simplified model that represents a constant power consumption, corresponding to the mean value of the working state power value, with a superimposed noisy contribution, described by the variance value in the working state. Under the assumption of statistical independence between the active and reactive power components, the covariance matrix $\boldsymbol{\Sigma}_{x^{(i)}(t)}$ is diagonal:

$$\boldsymbol{\Sigma}_{x^{(i)}(t)} = \begin{bmatrix} \sigma^2_{a,x^{(i)}(t)} & 0 \\ 0 & \sigma^2_{r,x^{(i)}(t)} \end{bmatrix}, \tag{25}$$

where $\sigma^2_{a,x^{(i)}(t)}$ and $\sigma^2_{r,x^{(i)}(t)}$ represent respectively the variance of the active and reactive power in the cluster. The inference procedure is carried out independently for the two components. Therefore, at each state,

$$y_c^{(i)}(t) | x^{(i)}(t) \sim \mathcal{N}\left(\mu_{c,x^{(i)}(t)}, \sigma^2_{c,x^{(i)}(t)}\right). \tag{26}$$

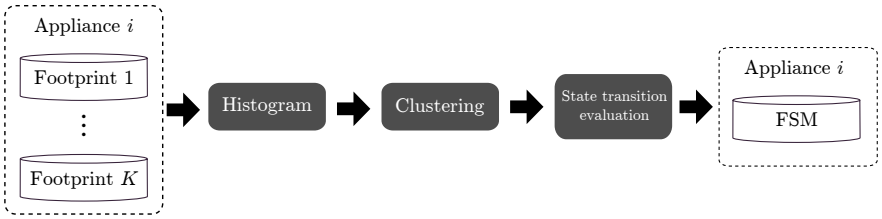


Figure 6: Block diagram of the clustering and of the model training stages of Hart's algorithm.

The number of states m_i is defined in the clustering phase, described in Section 2.2, assuming that each cluster corresponds to a state in the FSM model: the estimation of the mean and the variance values for each component is performed with the Maximum Likelihood criterion on the clusters data. Each component of the tolerance parameter $\beta_{c,j}$, associated to the respective component of the emitted symbol $\mu_{c,j}$, is set equal to the standard deviation $\sigma_{c,j}$ of the Gaussian distribution.

Regarding the state transition matrix $\boldsymbol{T}$, each entry T_{ij} represents the admissibility of the transition from state i to state j, using the value $T_{ij} = 1$ if the transition is allowed and $T_{ij} = 0$ otherwise. This value is inferred from the ground truth state evolution of each appliance consumption. Since this model does not represent the evolution in time of a signal, the permanence in the state is not represented, therefore the variable T_{ii} is set to 1. The diagram of the clustering and of the model training stage is shown in Figure 6.

Since the aggregated data $\overline{y}_c(t)$ is assumed to correspond with the sum of the power consumption of each appliance, it can be modelled as a Gaussian variable, described by a mean value and a variance value equivalent to the sum of the corresponding values of each appliance, under the assumption of statistical independence among the appliances:

$$\overline{y}_c(t)|x^{(1:N)}(t) \sim \mathcal{N}\left(\sum_{i=1}^{N} \mu_{c,x^{(i)}(t)}, \sum_{i=1}^{N} \sigma^2_{c,x^{(i)}(t)}\right). \tag{27}$$

This variable represents the Probability Density Function (PDF) of the working states combinations and it allows to evaluate which combination of working states fit the power value for each sample of the aggregated data. The number of admissible combinations of working states is equal to $\prod_{i=1}^{N} m_i$.

Following the same rule defined for each appliance symbol, the effectiveness interval for each combination is centred in mean value, and its width is twice the value of the standard deviation. For some combinations, which have similar mean value or great variance, the effectiveness intervals are overlapped: for those cases, if the power value falls in this region, both the combinations are considered valid.

The aggregate power data is analysed sample by sample: for each value, the effectiveness intervals in which the sample falls are selected. The related state combination might be admissible or not, depending on the previous state combination selected. Therefore, for each FSM, from the knowledge of the previous state selected, the admissible transition are evaluated through the transition matrix T_{ij}: the FSMs which do not make any variation in the state from the previous combination are not evaluated, then if the transition

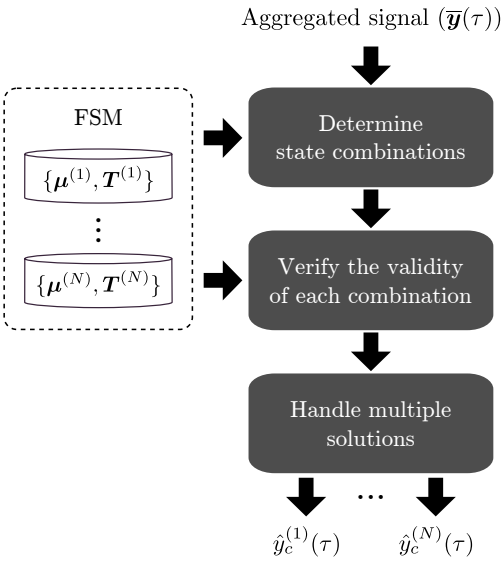


Figure 7: Diagram of the load disaggregation phase.

is not admissible for at least one FSM, the selected combination is discarded. The starting combination is evaluated on the first sample, without the evaluation on the transition from any previous state. If no combination is admissible, the previous state in maintained for each FSM. If the aggregated data sample does not fall within any combination interval, the previous state is maintained for each FSM.

In this way, the time series of the state evolution is reconstructed for each FSM. The disaggregation consists in using the related power level consumption assigned to each state of the FSM, thus reconstructing the power consumption profile for each appliance. The general scheme of the disaggregation phase is shown in Figure 7.

In order to deal with the noise presence in the aggregated data, an FSM version of the *noise* model defined in Section 2.3 is considered, additionally to the FSM models representing the appliances.

4.1. Handling multiple solutions

In [14], the author did not describe the technique adopted for dealing with the occurrence of multiple solutions during the disaggregation phase. In this paper, we adopted two different approaches for dealing with the problem. The first consists in supposing that each combination of appliances is equally probable, thus the ambiguity is solved by choosing a random combination sampled from a uniform distribution. This algorithm will be denoted as "Hart" in the remainder of this paper.

A second approach, consists in adopting a MAP technique [44]: the *posterior probability* of each combination is calculated from the training data, and it is multiplied to the Gaussian PDF, resulting in the *posterior PDF*. The value of the posterior PDF in the aggregate data sample is denoted as the *posterior likelihood*. The combination with the higher posterior likelihood value is then chosen as the most probable

combination. This alternative of Hart's algorithm will be denoted as "Hart w/ MAP" in the remainder of this paper.

5. Experiments

This section firstly presents the metrics employed to evaluate the performance of the proposed approach and of the comparative methods. Then it describes the experimental procedure and discusses the obtained results.

5.1. Performance metrics

According to the study carried out by the authors [15], the metrics chosen to evaluate the performance represent both the aspects of the disaggregation problem: the classification of the switching activity of the appliances and the accuracy of the disaggregated profiles compared to the ground truth appliance consumption. The chosen metrics are defined by Kolter and Jaakkola [18]: the "Recall" measures the part of the power consumption that has been correctly classified, whereas the "Precision" measures the amount of power assigned to an appliance truly belonged to it. The proposed approach and Hart's algorithm are able to disaggregate both the active and the reactive power, however the performance metrics have been calculated on the active power only in order to compare it with the univariate formulation of AFAMAP. Furthermore, the active power is the physical quantity directly related to the cost in the bill, therefore it is the most relevant component to be analysed. Considering the i-th appliance, Precision $\left(P^{(i)}\right)$ and Recall $\left(R^{(i)}\right)$ are calculated as follows:

$$P^{(i)} = \frac{\sum_{\tau=1}^{\Upsilon} \min\left(\hat{y}_a^{(i)}(\tau), y_a^{(i)}(\tau)\right)}{\sum_{\tau=1}^{\Upsilon} \hat{y}_a^{(i)}(\tau)}, \tag{28}$$

$$R^{(i)} = \frac{\sum_{\tau=1}^{\Upsilon} \min\left(\hat{y}_a^{(i)}(\tau), y_a^{(i)}(\tau)\right)}{\sum_{\tau=1}^{\Upsilon} y_a^{(i)}(\tau)}, \tag{29}$$

where $\hat{y}_a^{(i)}(\tau)$ is the appliance active power signal estimated by the disaggregation algorithm, Υ is the total number of samples, and $y_a^{(i)}(\tau)$ is the ground truth appliance active power signal.

Finally, in order to consider the total performance of the disaggregation system, the metric average across the appliances is computed:

$$P = \frac{1}{N}\sum_{i=1}^{N} P^{(i)}, \quad R = \frac{1}{N}\sum_{i=1}^{N} R^{(i)}. \tag{30}$$

As unique evaluation metric, the F_1-Measure is chosen and it is calculated as the geometric mean between Precision and Recall:

$$F_1 = 2\frac{P\,R}{P+R}. \tag{31}$$

Additionally to Precision, Recall, and F_1-Measure, the performance has been evaluated by using the Normalized Disaggregation Error (NDE) [18, 55], defined as: defined as:

$$NDE = \sqrt{\frac{\sum_{\tau,i}\left(y_a^{(i)}(\tau) - \hat{y}_a^{(i)}(\tau)\right)^2}{\sum_{\tau,i}\left(\hat{y}_a^{(i)}(\tau)\right)^2}}. \tag{32}$$

NDE provides a direct measure of the ability of the algorithm of reconstructing the active power profiles.

5.2. Experimental setup

The dataset used for the experiments is the Almanac of Minutely Power dataset (AMPds) [44]: it contains recordings of consumption profiles belonging to a single home in Canada for a period of two years, at 1 minute sampling rate. Additionally to the aggregated power consumption, it provides active and reactive power at appliance level, unlike most of the dataset, in which the appliances consumption is described by the only active power [15]: this information is crucial in order to create the appliance models and test the new approach.

The experiments are conducted by using the six appliances which contribute the most to the power consumption: dryer, washing machine, dishwasher, fridge, electric oven, and heat pump. Regarding the significance of the reactive components of the appliances taken into consideration, the following values have been extracted from the datasets: (128.25 W, 7.96 VAR) for the fridge, (4545.91 W, 413.75 VAR) and (248.11 W, 408.94 VAR) for the dryer, (909.11 W, 203.44 VAR), (531.10 W, 14.37 VAR),(146.80 W, 3.60 VAR) and (137.54 W, 96.47 VAR) for the washing machine, (753.07 W, 33.31 VAR), (137.96 W, 35.86 VAR) and (14.42 W, 52.55 VAR) for the dishwasher, (3187.67 W, 136.63 VAR),(125.68 W, 121.67 VAR) and (89.54 W, 50.62 VAR) for the electric oven, (1798.83 W, 320.95 VAR) and (37.23 W, 17.03 VAR) for the heat pump. As shown by these values, it is authors' opinion that the appliances evaluated in the experiments have a significant contribution of reactive power that make them suitable for evaluating the performance of the proposed approach. Analysing the contents of the dataset, the usage of the appliances proves to be homogeneous throughout the entire period, therefore the experiments are evaluated on 6 months of data, which can be considered representative of the entire dataset. A subset of the data, spanning over 14 days, has been considered sufficient to collect all the signatures required to train all the HMMs. This represents the training set in the Figure 2b.

Two different scenario are defined in this work, according to [56]. The *noised* scenario employs the aggregated power consumption in the dataset as the aggregated signal, therefore it includes the noise term. In this case, the training data used to create the *noise model* are obtained subtracting the ground truth consumption signals, related to the appliances of interest, from the aggregated power. Whereas, in the *denoised* scenario the aggregated data are synthetically composed by summing the ground truth appliance power signals in the dataset, determining the absence of the noise term.

Table 1: Number of states m_i related to each class of appliance.

Problem dimensionality	**Dryer**	**Washing machine**	**Dishwasher**	**Fridge**	**Electric oven**	**Heat pump**
Univariate	3	4	3	2	3	3
Bivariate	3	5	4	2	4	3

The frame size is set to $T = 60$ minutes, which is an interval sufficiently large to include a complete activation for the most of appliances under study. This value is considered within the *Windowing* operation in the Figure 5. For the ones which have a longer activation, this value allows to include a complete operating sub cycle, for which the HMM is still representative. The variance parameters are set to $\sigma^2_{c,1} = \sigma^2_{c,2} = 0.01$ according to the variance of the experimental data, and the regularisation parameter is set to $\lambda = 1$.

The algorithm has been implemented in Matlab and the CPLEX[1] solver has been used to solve the QP problem. The amount of time required to disaggregate a frame of 60 minutes on a personal computer equipped with an Intel i7 CPU running at 3.3 GHz and 32 GB of RAM is about 30 s. The performance is compared to the univariate formulation of AFAMAP and to Hart's algorithm presented in Section 4. The tolerance parameter is set $\alpha = 10^{-6}$.

Table 1 presents the number of states, defined a-priori for each class of appliance.

The number of states in the *noise* model has been varied in the range $\{4, 6, 8, 10\}$, both in the univariate and bivariate approaches, in order to find the most performing model.

5.3. Denoised scenario

In this section, the results of the experiments related to the *denoised* scenario will be shown. Since the aggregated power signal depends on which and how many appliances are considered, the experiments have been conducted by varying the number of appliances, in order to evaluate the disaggregation performance for different problem complexities. In particular, different test sets, each composed of every combination of N appliances have been created. For each test set, the total number of experiments is $\binom{6}{N}$, with $N = 2, \ldots, 5$ and the final metrics are calculated averaging between the single experiments overall performance. Before calculating the final F_1-Measure, the Precision and Recall are averaged between the experiments. Differently, the final NDE is the average between the single experiment value.

5.3.1. Results

In Figure 8, the disaggregated appliances active power (D) are compared to the corresponding ground truth (GT): in the figure, for each appliance, an adequate time span is considered, in order to evaluate

[1] `http://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/`

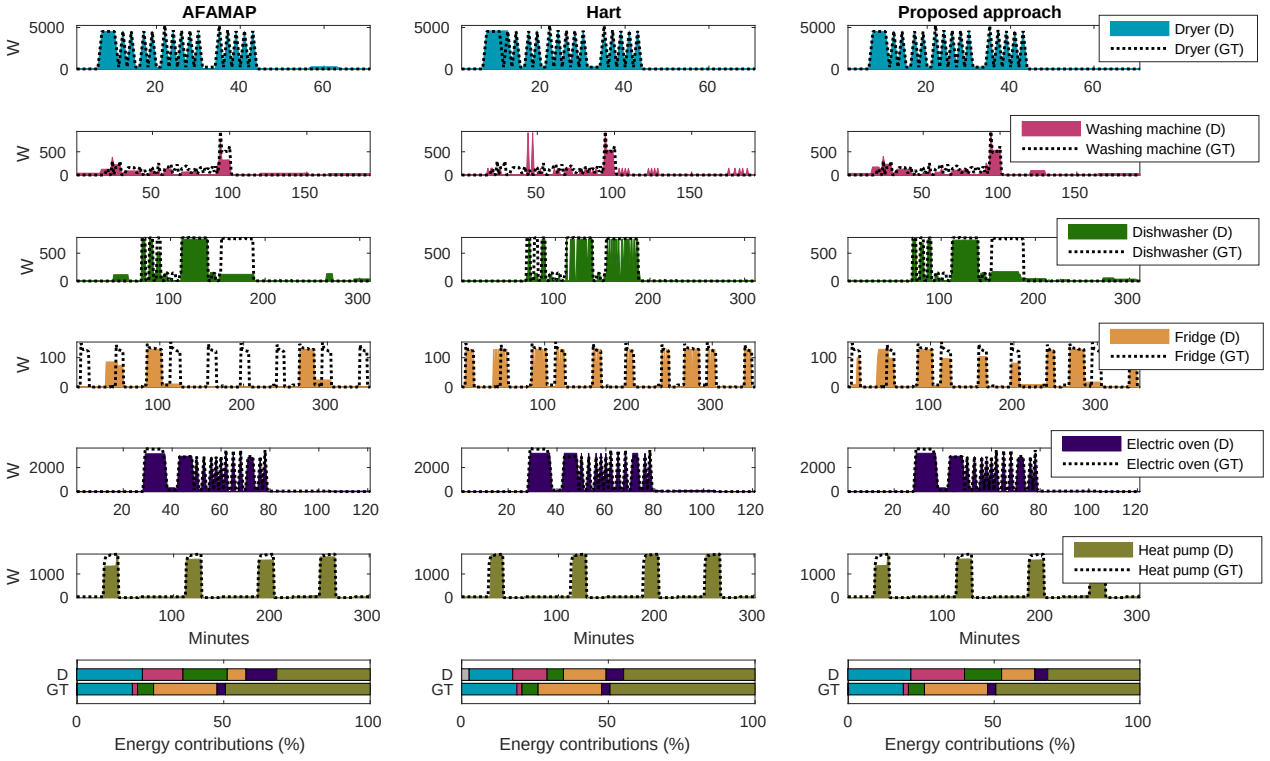
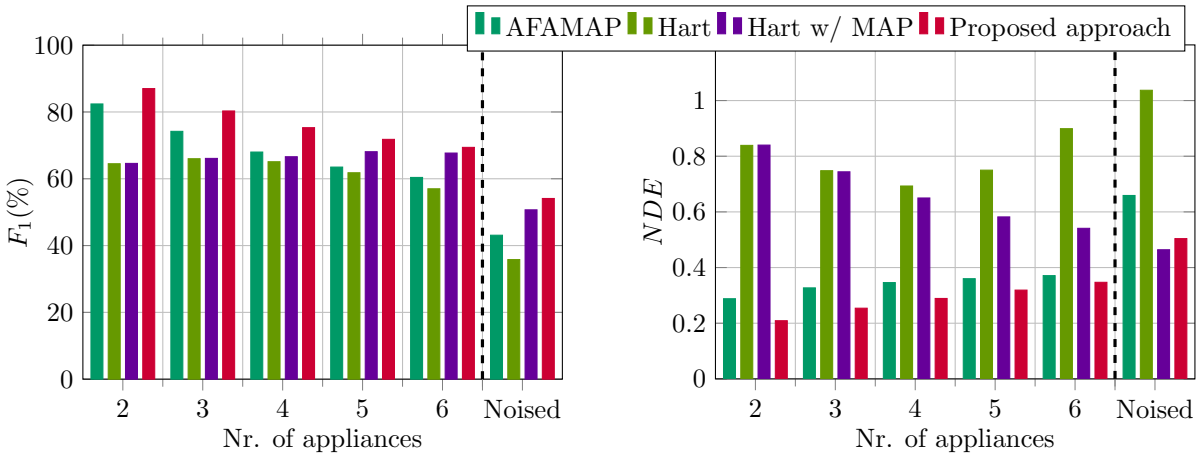


Figure 8: Algorithms comparison: AFAMAP vs Hart vs proposed approach. For each algorithm, the disaggregation output (D) is compared against the ground truth (GT) signals.

the performance on a single or multiple activations. The bottom of the figure shows the comparison of the appliance contribution to the total energy in the aggregated signal, between the disaggregation outputs and the ground truth consumptions. The left side of the figure shows the disaggregation profiles resulting from the univariate formulation of the AFAMAP algorithm, the central shows the active power component resulting from the Hart's algorithm, and the right side shows profiles related to the proposed approach.

The overall disaggregation results are reported in Figure 9, where the F_1-Measure is reported in the Figure 9a and the NDE in the Figure 9b. The values are related to Table 3, where the absolute improvements of the proposed approach with respect to the AFAMAP and the Hart's algorithm are shown. The proposed approach reaches the best performance in each case study, with F_1-Measure of 87.0 and NDE equal to 0.209 in the 2 appliances case, and with F_1-Measure of 69.4 and NDE equal to 0.347 in the 6 appliances case, The proposed approach reaches the best performances in each case study, with F_1-Measure of 87.0 and NDE equal to 0.209 in the 2 appliances case, and with F_1-Measure of 69.4 and NDE equal to 0.347 in the 6 appliances case.

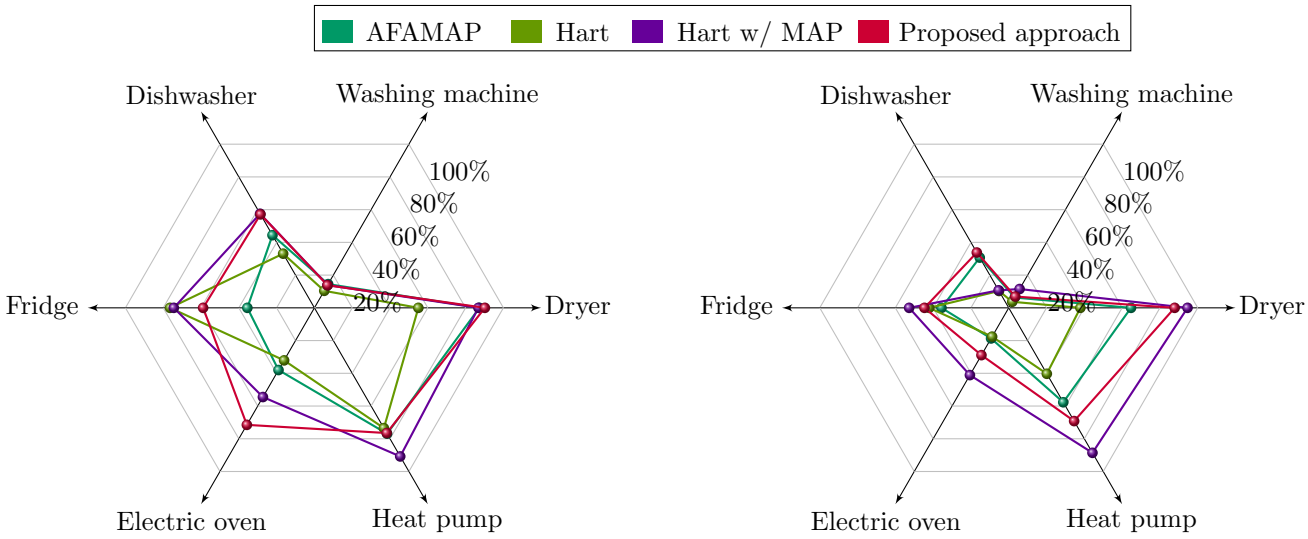
The radar chart in Figure 10 shows the F_1-Measure for each appliance the experiment including all the 6 appliances and the area of each coloured line is proportional to the F_1-Measure of the related algorithm



(a) Comparison of the disaggregation performance in terms of F_1-Measure for different number of appliances.

(b) Comparison of the disaggregation performance in terms of NDE for different number of appliances.

Figure 9: Disaggregation performance on AMPds dataset for all the addressed algorithms.



(a) Denoised scenario.

(b) Noised scenario.

Figure 10: Performance in terms of F_1-Measure (%) for the different appliances in the "6 appliances" case study: (a) denoised scenario, (b) noised scenario.

Table 2: Performance improvement in the "6 appliances" case study (denoised scenario).

Algorithm		**Metric**	**Dryer**	**Washing machine**	**Dishwasher**	**Fridge**	**Electric oven**	**Heat pump**
AFAMAP	(I)	F_1 (%)	87.3	14.5	44.4	35.5	38.0	76.9
Hart	(II)		54.9	10.4	33.1	76.4	32.1	73.4
Hart w/ MAP	(III)		86.8	14.0	57.5	74.4	54.5	90.8
Proposed approach	(IV)		90.2	13.8	57.1	58.9	71.5	76.5
Improvement	(IV)-(I)	$\Delta F_1/F_1(\%)$	+3.3	-4.8	+28.6	+65.9	+88.2	-0.5
	(IV)-(II)		+64.3	+32.7	+72.5	-22.9	+122.7	+4.2
	(IV)-(III)		+3.9	-1.4	-0.7	-20.8	+31.2	-15.8
AFAMAP	(I)	NDE	0.215	2.279	0.685	0.878	0.478	0.388
Hart	(II)		0.798	4.383	1.282	0.670	1.725	0.739
Hart w/ MAP	(III)		0.481	3.003	0.768	0.685	1.026	0.411
Proposed approach	(IV)		0.229	2.384	0.446	0.735	0.286	0.377
Improvement	(IV)-(I)	ΔNDE	+0.014	+0.104	-0.239	-0.143	-0.192	-0.011
	(IV)-(II)		-0.569	-1.999	-0.836	+0.065	-1.439	-0.363
	(IV)-(III)		-0.252	-0.620	-0.322	+0.049	-0.740	-0.034

averaged across the appliances. The values are related to Table 2, where the absolute improvements of the proposed approach with respect to the AFAMAP and the Hart's algorithm are shown.

5.3.2. Analysis

As shown in the plots, the appliances presenting a high steady power consumption are easily recognised, whereas the appliances with complex working cycles, or with several power levels, are more difficult to detect. For instance, the dryer, the electric oven, and the heat pump are successfully reconstructed, whereas the washing machine, the dishwasher and the fridge are partially erroneously reconstructed. Indeed, in the univariate formulation, whenever several appliances present similar consumption levels, many combinations may satisfy the problem constraints and the algorithm chooses an erroneous solution for disaggregation. Comparing the results with the proposed bivariate approach, the multiple combinations of the solution are reduced due to the component constraint to be satisfied by the algorithm, which leads to the correct solution and, consequently, to a better profile disaggregation of the active power component. For instance, although the appliances with higher power level maintain a successful disaggregation, the fridge and the dishwasher improve the correspondence with the ground truth signals. The washing machine partially improve the disaggregation performance in the activation period, whereas introduces some false energy assignation. The disaggregated profiles of Hart's method show that, for some appliances, the FSM is a modelling technique which allows a better representation for the appliances with sharply defined steady states, e.g., the fridge

Table 3: Comparison of the disaggregation performance for different number of appliances (denoised scenario).

Algorithm		**Metric**	**2 appl.**	**3 appl.**	**4 appl.**	**5 appl.**	**6 appl.**
AFAMAP	(I)	F_1 (%)	82.4	74.2	68.0	63.5	60.4
Hart	(II)		64.5	66.0	65.1	61.8	57.0
Hart w/ MAP	(III)		64.6	66.1	66.6	68.1	67.7
Proposed approach	(IV)		**87.0**	**80.3**	**75.3**	**71.8**	**69.4**
Improvement	(IV) - (I)	$\Delta F_1/F_1$ (%)	+ 5.6	+ 8.2	+ 10.7	+ 13.1	+ 14.9
	(IV) - (II)		+ 34.9	+ 21.7	+ 15.7	+ 16.2	+ 21.8
	(IV) - (III)		+ 34.7	+ 21.5	+ 13.1	+ 5.4	+ 2.5
AFAMAP	(I)	NDE	0.288	0.327	0.346	0.360	0.371
Hart	(II)		0.839	0.748	0.693	0.750	0.899
Hart w/ MAP	(II)		0.840	0.744	0.650	0.582	0.541
Proposed approach	(IV)		**0.209**	**0.254**	**0.289**	**0.319**	**0.347**
Improvement	(IV) - (I)	ΔNDE	- 0.079	- 0.073	- 0.057	- 0.041	- 0.024
	(IV) - (II)		- 0.630	- 0.494	- 0.404	- 0.431	- 0.552
	(IV) - (III)		- 0.631	- 0.490	- 0.361	- 0.263	- 0.194

and the heat pump, but a worse representation for appliances with highly variable activity, e.g., the electric oven.

The more confident are the disaggregated profiles with respect to the ground truth signal, the better is the estimation of the energy consumption percentage distribution among the appliances: indeed, for the proposed approach, the consumption distribution have a better correspondence with the ground truth ones, with respect to the AFAMAP algorithm. For instance, the disaggregated profiles related to the fridge results to be more confident, which reflects on the increase of the energy assignation, whereas the dishwasher and the electric oven ones results to have a false energy assignation during the OFF period, corresponding to a decrease of the related energy contributions. Regarding the washing machine, some errors are introduced, therefore the energy assignation is erroneously increased. Regarding the dryer and the heat pump the energy contributions are maintained, because of the correspondence between the algorithms disaggregation performance. In the Hart's method, the improvements in the heat pump and the fridge are reflected on a better correspondence between the energy contributions, but the absence of the constraint between the aggregate power amount and the sum of the disaggregated profiles leads to an unassigned percentage of the total energy (represented as the *grey* portion).

Regarding the performance of the individual appliances, the major improvements with respect to AFAMAP are observed in the electric oven, the fridge and the dishwasher, with an relative increase of the F_1-Measure of $+88.2\,\%$, $+65.9\,\%$ and $+28.6\,\%$, and a variation in the NDE of -0.192, -0.143, -0.239 respectively.

This is due to a more accurate correspondence between the disaggregated output and the ground truth, as already shown in the disaggregation output plots. On the contrary, the performance is almost unchanged for the washing machine, the dryer and the heat pump. With respect to the Hart's algorithm, the proposed approach shows an high improvement additionally for the dryer, with an absolute increase of F_1-Measure equal $+64.3\,\%$ and a variation in the NDE of -0.569, whereas, it shows a substantial loss for the fridge, with a decrease of F_1-Measure equal $-22.9\,\%$ and a variation in the NDE of $+0.065$. This demonstrates that the HMM modelling results more effective with a higher number of states. Since moving from the univariate to the bivariate model leads to a greater number of states, this also demonstrates the effectiveness of the proposed approach. Compared to the Hart's algorithm with the MAP stage, the performance on each appliance reduce their gain, particularly for the dishwasher and the dryer, with a decrease of F_1-Measure equal to $-0.7\,\%$ and an increase of $+3.9\,\%$ and a variation in the NDE of -0.322 and -0.252 up to the heat pump, where a loss of performance is shown, with an absolute increase in the F_1-Measure of $-15.8\,\%$ and a variation in the NDE of -0.034. The washing machine remains the appliance with the worst disaggregation performance: the reason is the model complexity, since it is the appliance with the highest number of states, both in the univariate and bivariate representation. Observing the radar chart, the area under the curve related to the proposed approach is increased with respect to AFAMAP and Hart's algorithm, resulting in an average performance improvement, whereas it is slightly higher with respect to the Hart's algorithm version with the MAP stage. The average performance of the system increases, resulting in a relative improvement of F_1-Measure equal to $+14.9\,\%$, $+21.8\,\%$ and $+2.5\,\%$, and a variation in the NDE of -0.024, -0.552, -0.194 with respect to AFAMAP, the Hart's algorithm and the version with MAP stage, respectively.

Concerning the experiments with for different number of appliances, the results shows that, lowering the number of appliances, the performance improve in the FHMM-based algorithms, while in the Hart's algorithm it reaches a peak with 4 appliance, after that the performance decrease. Regarding the Hart's algorithm version with the MAP stage, the performance decrease gradually with a lower number of appliance.

Compared to AFAMAP and to Hart's algorithm, the proposed approach provides a significant performance improvement also when the problem complexity is minimal, i.e., when the number of appliances is 2. The higher absolute increase from AFAMAP occurs with 6 appliances, whereas it decreases lowering the complexity of the problem: this demonstrates that the proposed approach resolves more ambiguities in the NILM solution when the number of combinations of working states is higher.

Regardless the number of appliances, the performance of Hart's algorithm is lower compared to the proposed approach, because of the less descriptive capabilities of the FSM appliance model with respect to the HMM one. The comparative evaluation with the Hart's version with the MAP stage proves that, even if this approach exploits the information on the most probable solution in case of ambiguity, which is an ideal condition, the proposed approach reaches better performance. Furthermore, the proposed algorithm

Table 4: Appliances performance improvement in the "6 appliances" case study (noised scenario).

Algorithm		**Metric**	**Dryer**	**Washing machine**	**Dishwasher**	**Fridge**	**Electric oven**	**Heat pump**	**Overall**
AFAMAP	(I)	F_1 (%)	64.6	6.3	30.7	35.6	18.7	57.7	43.1
Hart	(II)		37.9	3.7	10.3	42.1	17.6	40.3	35.8
Hart w/ MAP	(III)		94.6	11.6	10.7	52.7	41.1	88.6	50.7
Proposed approach	(IV)		87.8	6.9	33.8	44.6	28.9	69.2	**54.1**
Improvement	(IV)-(I)	$\Delta F_1/F_1$ (%)	+ 35.9	+ 9.5	+ 10.1	+ 25.3	+ 54.5	+ 19.9	+ 25.5
	(IV)-(II)		+ 131.7	+ 86.5	+ 228.2	+ 5.9	+ 64.2	+ 71.7	+ 51.1
	(IV)-(III)		- 7.2	- 40.5	+ 215.9	- 15.4	- 29.7	- 21.9	+ 6.7
AFAMAP	(I)	NDE	0.305	4.395	0.888	0.909	0.939	0.787	0.659
Hart	(II)		0.882	5.960	1.714	0.982	1.593	0.929	1.037
Hart w/ MAP	(III)		0.254	1.965	1.110	0.942	0.974	0.432	**0.464**
Proposed approach	(IV)		0.272	4.055	0.829	0.861	0.930	0.467	0.504
Improvement	(IV)-(I)	ΔNDE	- 0.033	- 0.340	- 0.058	- 0.048	- 0.010	- 0.319	- 0.155
	(IV)-(II)		- 0.610	- 1.905	- 0.884	- 0.121	- 0.663	- 0.462	- 0.533
	(IV)-(III)		+ 0.019	+ 2.090	- 0.280	- 0.081	- 0.044	+ 0.036	+ 0.040

provides an optimum solution on a frame of T samples, which takes into account both the short-term and long-term dependencies of the signal. This differs in Hart's algorithm that finds the solution by processing the aggregate signal sample-by-sample. For this method, the performance decreases reducing the number of the appliances: a motivation behind this phenomenon can reside in the fact that the MAP stage of the Hart's algorithm chooses a solution with higher probability, but which results incorrect for the majority of the experiments, specially with few combinations.

5.4. Noised scenario

In this section, the results of the experiments related to the *noised* scenario will be shown. Differently from the *denoised* scenario, the aggregated power signal does not vary with the appliances considered, therefore only the results with all the appliances will be shown. Regarding the number of states of the *noise model*, the experiments demonstrated that, for each approach, the best value is 4, except for the Hart's algorithm with the MAP stage, for which the best results are reached with 10 states. For the sake of conciseness, only the results for the best configuration will be reported in this section.

5.4.1. Results

The overall disaggregation results are reported in Figure 9, on the last column, in order to make a comparative evaluation with the *denoised* scenario. The values are related to Table 4 on the *Overall* column, where the absolute improvements of the proposed approach with respect to the AFAMAP and the Hart's

algorithm are shown. The proposed approach reaches the best overall performances, with F_1-Measure of 54.1 and NDE equal to 0.504, despite of the Hart's algorithm version with the MAP stage shows an higher NDE value. This discordance will be motivated in the analysis. The radar chart in Figure 10b shows the F_1-Measure for each appliance. The values are related to Table 4.

5.4.2. Analysis

Differently from the *denoised* scenario, the major improvement, with respect to AFAMAP, is observed for the dryer, with a F_1-Measure relative improvement of $+35.9\,\%$, and a variation in the NDE of -0.033, whereas the improvements are reduced for the remaining appliances. This proves the effectiveness of the transition from the univariate to the bivariate formulation of the problem, even in the presence of noise.

With respect to Hart's algorithm, the proposed approach shows a higher improvement for the dryer, the dishwasher, and the heat pump with an improvement of $+131.7\,\%$, $+228.2\,\%$, $+71.7\,\%$, and a variation in the NDE of -0.610, -0.884, -0.462. Differently, Hart's algorithm with the MAP stage achieves a higher F_1-Measure, and the relative difference of F_1-Measure for the heat pump, the electric oven and the dryer is $-19.4\,\%$, $-12.2\,\%$, $-6.8\,\%$, while in terms of NDE the difference is $+0.036$, -0.044, $+0.019$. This demonstrates that the HMM modelling leads to performance improvements with respect to the FSM modelling even in the presence of noise, but considering the MAP stage this improvements is substantially reduced. The washing machine is still the appliance with the worst disaggregation performance, following the trend of the *denoised* scenario. Observing the radar chart, the area under the curve related to the proposed approach is increased with respect to AFAMAP and Hart's algorithm, resulting in an average performance improvement, whereas it is comparable with respect to the Hart's algorithm version with the MAP stage, due to unbalancing between the appliances.

The average performance of the system increases, resulting in a F_1-Measure absolute improvement of $+25.5\,\%$, $+51.1\,\%$ and $+6.7\,\%$, and a variation in the NDE of -0.155, -0.533, $+0.040$ with respect to AFAMAP, the Hart's algorithm and the version with MAP stage, respectively.

Comparing those results to the *denoised* scenario ones, the overall performance is lower, due to the introduction of the noise contribution in the aggregated power, except for the Hart's algorithm with the MAP stage: despite the F_1-Measure shows a degradation of performance, the NDE decreases, meaning that this version of the algorithm maintains the trend showed with the increase of the number of appliances. In fact, the *noised* scenario can be defined as the *denoised* scenario using the *noise* model additionally to the appliances models, therefore the MAP stage introduces additional advantages, leading to a performance improvement. The MAP stage exploits additional information which are not introduced within the AFHMM, but represents an almost ideal FSM based case study.

6. Conclusion

In this paper, active and reactive power have been introduced in Additive Factorial HMM for non-intrusive load monitoring. The appliance models based on bivariate HMM have been introduced and the procedure for estimating their parameters has been described. This consists in the extraction of the footprint of the appliance by means of an Appliance Activity Detector and in the estimation of the power levels of each working state by clustering the appliance footprint with the k-means algorithm. The disaggregation algorithm is in an alternative formulation of the AFAMAP algorithm [18] developed in order to deal with the bivariate formulation of the problem. As results, the algorithm is able to output the disaggregated profiles in the active and reactive power components.

The proposed approach has been compared to the univariate formulation of AFAMAP and to the algorithm presented by Hart in [14]. The latter is based on Finite State Machine appliance models and it employs both the active and reactive power. The algorithm has been improved for handling the occurrence of multiple solutions by means of a MAP technique. The experiments have been conducted on the AMPds [44] dataset, which provides the ground truth appliance consumption both in the active and reactive power components. The results showed that, in a denoised scenario, the proposed approach outperforms both the comparative methods, with an absolute F_1-Measure improvement of $+\,14.9\,\%$ and $+\,2.5\,\%$ in the 6 appliances case study.

As future works, a more reliable appliance model will be considered in order to improve the representation of the working states, e.g., the usage of Gaussian Mixture Model (GMM) within the HMM allows the representation of a more suitable power level density distribution with respect to a simple Gaussian distribution. Furthermore, additional information about the working states duration will be introduced, allowing the discrimination of HMMs with similar transition probabilities but different time in the switching activity. This translates into a fully exploitation of the differential model. Regarding the appliance modelling stage, an unsupervised clustering technique will be introduced to automatically detect the number of power levels, e.g., regarding appliances which not belongs to the categories considered in this work. Regarding the disaggregation and solver algorithms, binary variables will be introduced in the problem formalisation, leading to a Mixed Integer Quadratic Program (MIQP), in order to impose the variable to assume binary results and not integer values as in fuzzy logic, which can lead to ambiguous evaluation in the HMM state evolution. Finally, further experiments will be conducted in order to compare the proposed solution to other approaches recently presented in the literature [27, 28, 39].

Appendix A. Algorithm details

This appendix provides further details on the algorithm formulation presented in Section 3.1. In particular, the following terms of the QP problem are described: $\boldsymbol{H}$, $\boldsymbol{f}$, $\boldsymbol{A}_{eq}$, $\boldsymbol{b}_{eq}$, $\boldsymbol{lb}$, $\boldsymbol{ub}$, $\boldsymbol{A}_{ineq}$, and $\boldsymbol{b}_{ineq}$.

As described in Section 3.1, the matrix $\boldsymbol{H}$ is structured as follows:

$$\boldsymbol{H} = \begin{bmatrix} \boldsymbol{H}_a & \mathbf{0} \\ \mathbf{0} & \boldsymbol{H}_r \end{bmatrix}, \tag{A.1}$$

where $\boldsymbol{H}_c \in \{\boldsymbol{H}_a, \boldsymbol{H}_r\}$ is given by:

$$\boldsymbol{H}_c = \begin{bmatrix} \frac{1}{\sigma^2_{c,1}} \boldsymbol{H}_{\boldsymbol{\Theta}(1)} & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & \frac{1}{\sigma^2_{c,1}} \boldsymbol{H}_{\boldsymbol{\Theta}(T)} \end{bmatrix}, \quad \boldsymbol{H}_{\boldsymbol{\Theta}(t)} = \begin{bmatrix} \boldsymbol{H}_{\boldsymbol{\Psi}^{(1\,1)}(t)} & \dots & \boldsymbol{H}_{\boldsymbol{\Psi}^{(1\,N)}(t)} \\ \vdots & \ddots & \vdots \\ \boldsymbol{H}_{\boldsymbol{\Psi}^{(N\,1)}(t)} & \dots & \boldsymbol{H}_{\boldsymbol{\Psi}^{(N\,N)}(t)} \end{bmatrix}, \tag{A.2}$$

and

$$\boldsymbol{H}_{\boldsymbol{\Psi}^{(i\,j)}(t)} = \begin{bmatrix} \boldsymbol{H}_{\boldsymbol{\xi}^{(i\,j)}(t)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \boldsymbol{H}_{\boldsymbol{\xi}^{(i\,j)}(t)} = \begin{bmatrix} \mu^{(i)}_{c,1}\,\mu^{(j)}_{c,1} & \dots & \mu^{(i)}_{c,1}\,\mu^{(j)}_{c,m_j} \\ \vdots & \ddots & \vdots \\ \mu^{(i)}_{c,m_i}\,\mu^{(j)}_{c,1} & \dots & \mu^{(i)}_{c,m_i}\,\mu^{(j)}_{c,m_j} \end{bmatrix}. \tag{A.3}$$

Regarding the vector $\boldsymbol{f}$, in Section 3.1 it has been defined as follows:

$$\boldsymbol{f} = \begin{bmatrix} \boldsymbol{f}_a & \boldsymbol{f}_r \end{bmatrix}^T, \tag{A.4}$$

where $\boldsymbol{f}_c \in \{\boldsymbol{f}_a, \boldsymbol{f}_r\}$ is given by the sum of five terms:

$$\boldsymbol{f}_c = -\boldsymbol{f}_{c,1} - \frac{1}{\sigma^2_{c,1}} \boldsymbol{f}_{c,2} - \boldsymbol{f}_{c,3} + \frac{1}{2} \frac{1}{\sigma^2_{c,2}} \boldsymbol{f}_{c,4} - \frac{1}{2} \boldsymbol{f}_{c,5}, \tag{A.5}$$

where

$$\boldsymbol{f}_{c,1} = \begin{bmatrix} \boldsymbol{f}_{1,\boldsymbol{\Theta}(1)} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}, \; \boldsymbol{f}_{1,\boldsymbol{\Theta}(1)} = \begin{bmatrix} \boldsymbol{f}_{1,\boldsymbol{\Psi}^{(1)}(1)} \\ \vdots \\ \boldsymbol{f}_{1,\boldsymbol{\Psi}^{(N)}(1)} \end{bmatrix}, \; \boldsymbol{f}_{1,\boldsymbol{\Psi}^{(i)}(1)} = \begin{bmatrix} \boldsymbol{f}_{1,\boldsymbol{\xi}^{(i)}(1)} \\ \mathbf{0} \end{bmatrix}, \boldsymbol{f}_{1,\boldsymbol{\xi}^{(i)}(1)} = \begin{bmatrix} \log \phi^{(i)}_1 \\ \vdots \\ \log \phi^{(i)}_{m_i} \end{bmatrix}, \tag{A.6}$$

$$\boldsymbol{f}_{c,2} = \begin{bmatrix} \boldsymbol{f}_{2,\boldsymbol{\Theta}(1)} \\ \vdots \\ \boldsymbol{f}_{2,\boldsymbol{\Theta}(T)} \end{bmatrix}, \; \boldsymbol{f}_{2,\boldsymbol{\Theta}(t)} = \begin{bmatrix} \boldsymbol{f}_{2,\boldsymbol{\Psi}^{(1)}(t)} \\ \vdots \\ \boldsymbol{f}_{2,\boldsymbol{\Psi}^{(N)}(t)} \end{bmatrix}, \; \boldsymbol{f}_{2,\boldsymbol{\Psi}^{(i)}(t)} = \begin{bmatrix} \boldsymbol{f}_{2,\boldsymbol{\xi}^{(i)}(t)} \\ \mathbf{0} \end{bmatrix}, \boldsymbol{f}_{2,\boldsymbol{\xi}^{(i)}(t)} = \begin{bmatrix} \overline{y}_{c,f}(t)\,\mu^{(i)}_{c,1} \\ \vdots \\ \overline{y}_{c,f}(t)\,\mu^{(i)}_{c,m_i} \end{bmatrix}, \tag{A.7}$$

$$\boldsymbol{f}_{c,3} = \begin{bmatrix} \mathbf{0} \\ \boldsymbol{f}_{3,\boldsymbol{\Theta}(2)} \\ \vdots \\ \boldsymbol{f}_{3,\boldsymbol{\Theta}(T)} \end{bmatrix}, \; \boldsymbol{f}_{3,\boldsymbol{\Theta}(t)} = \begin{bmatrix} \boldsymbol{f}_{3,\boldsymbol{\Psi}^{(1)}(t)} \\ \vdots \\ \boldsymbol{f}_{3,\boldsymbol{\Psi}^{(N)}(t)} \end{bmatrix}, \; \boldsymbol{f}_{3,\boldsymbol{\Psi}^{(i)}(t)} = \begin{bmatrix} \mathbf{0} \\ \boldsymbol{f}_{3,\boldsymbol{\beta}^{(i)}(t)} \end{bmatrix}, \boldsymbol{f}_{3,\boldsymbol{\beta}^{(i)}(t)} = \begin{bmatrix} \log P^{(i)}_{1\,1} \\ \vdots \\ \log P^{(i)}_{m_i\,1} \\ \vdots \\ \log P^{(i)}_{1\,m_i} \\ \vdots \\ \log P^{(i)}_{m_i\,m_i} \end{bmatrix}, \tag{A.8}$$

$$\boldsymbol{f}_{c,4} = \begin{bmatrix} \mathbf{0} \\ \boldsymbol{f}_{4,\boldsymbol{\Theta}(2)} \\ \vdots \\ \boldsymbol{f}_{4,\boldsymbol{\Theta}(T)} \end{bmatrix}, \ \boldsymbol{f}_{4,\boldsymbol{\Theta}(t)} = \begin{bmatrix} \boldsymbol{f}_{4,\boldsymbol{\Psi}^{(1)}(t)} \\ \vdots \\ \boldsymbol{f}_{4,\boldsymbol{\Psi}^{(N)}(t)} \end{bmatrix}, \ \boldsymbol{f}_{4,\boldsymbol{\Psi}^{(i)}(t)} = \begin{bmatrix} \mathbf{0} \\ \boldsymbol{f}_{4,\boldsymbol{\beta}^{(i)}(t)} \end{bmatrix}, \tag{A.9}$$

$$\boldsymbol{f}_{4,\boldsymbol{\beta}^{(i)}(t)} = \begin{bmatrix} k^{(i)}_{1\,1}(t) \\ \vdots \\ k^{(i)}_{m_i\,1}(t) \\ \vdots \\ k^{(i)}_{1\,m_i}(t) \\ \vdots \\ k^{(i)}_{m_i\,m_i}(t) \end{bmatrix}, \quad \boldsymbol{k}^{(i)}(t) = \begin{bmatrix} 0 & \dots & \left(\Delta\overline{y}_{c,f}(t) - \left(\mu^{(i)}_{c,1} - \mu^{(i)}_{c,m_i}\right)\right)^2 \\ \vdots & \ddots & \vdots \\ \left(\Delta\overline{y}_{c,f}(t) - \left(\mu^{(i)}_{c,m_i} - \mu^{(i)}_{c,1}\right)\right)^2 & \dots & 0 \end{bmatrix} \tag{A.10}$$

$$\boldsymbol{f}_{c,5} = \begin{bmatrix} \mathbf{0} \\ \boldsymbol{f}_{5,\boldsymbol{\Theta}(2)} \\ \vdots \\ \boldsymbol{f}_{5,\boldsymbol{\Theta}(T)} \end{bmatrix}, \ \boldsymbol{f}_{5,\boldsymbol{\Theta}(t)} = \begin{bmatrix} \boldsymbol{f}_{5,\boldsymbol{\Psi}^{(1)}(t)} \\ \vdots \\ \boldsymbol{f}_{5,\boldsymbol{\Psi}^{(N)}(t)} \end{bmatrix}, \ \boldsymbol{f}_{5,\boldsymbol{\Psi}^{(i)}(t)} = \begin{bmatrix} \mathbf{0} \\ \boldsymbol{f}_{5,\boldsymbol{\beta}^{(i)}(t)} \end{bmatrix}, \tag{A.11}$$

$$\boldsymbol{f}_{5,\boldsymbol{\beta}^{(i)}(t)} = \begin{bmatrix} d^{(i)}_{1\,1}(t) \\ \vdots \\ d^{(i)}_{m_i\,1}(t) \\ \vdots \\ d^{(i)}_{1\,m_i}(t) \\ \vdots \\ d^{(i)}_{m_i\,m_i}(t) \end{bmatrix}, \boldsymbol{d}^{(i)}(t) = \begin{bmatrix} 0 & \dots & D\left(\frac{\Delta\overline{y}_{c,f}(t)}{\sigma_{c,2}}, \lambda\right) \\ \vdots & \ddots & \vdots \\ D\left(\frac{\Delta\overline{y}_{c,f}(t)}{\sigma_{c,2}}, \lambda\right) & \dots & 0 \end{bmatrix}, \tag{A.12}$$

$$D(y, \lambda) = \min\left\{\frac{1}{2}y^2, \max\left\{\lambda|y| - \frac{\lambda^2}{2}, \frac{\lambda^2}{2}\right\}\right\}. \tag{A.13}$$

The matrix $\boldsymbol{A}_{eq}$ is defined as follows:

$$\boldsymbol{A}_{eq} = \begin{bmatrix} \boldsymbol{A}_{eq,a} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{A}_{eq,r} \end{bmatrix}, \boldsymbol{A}_{eq,c} = \begin{bmatrix} \boldsymbol{A}_{eq,\boldsymbol{\Theta}(1)} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \boldsymbol{A}_{eq,\boldsymbol{\Theta}(T)} \end{bmatrix}, \tag{A.14}$$

$$\boldsymbol{A}_{eq,\boldsymbol{\Theta}(t)} = \begin{bmatrix} \boldsymbol{A}_{eq,\boldsymbol{\Psi}_1^{(1)}(t)} & \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{A}_{eq,\boldsymbol{\Psi}_2^{(1)}(t)} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{A}_{eq,\boldsymbol{\Psi}_1^{(N)}(t)} & \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{A}_{eq,\boldsymbol{\Psi}_2^{(N)}(t)} \end{bmatrix}, \tag{A.15}$$

$$\boldsymbol{A}_{eq,\boldsymbol{\Psi}_1^{(i)}(t)} = \begin{bmatrix} \boldsymbol{A}_{eq,\boldsymbol{\xi}_1^{(i)}(t)} \\ \boldsymbol{A}_{eq,\boldsymbol{\beta}_{1b}^{(i)}(t)} \\ \boldsymbol{A}_{eq,\boldsymbol{\beta}_{1f}^{(i)}(t)} \end{bmatrix}, \boldsymbol{A}_{eq,\boldsymbol{\xi}_1^{(i)}(t)} = \begin{bmatrix} 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix}, \tag{A.16}$$

$$\boldsymbol{A}_{eq,\boldsymbol{\beta}_{1b}^{(i)}(t)} = \begin{bmatrix} -1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & -1 & 0 & \cdots & 0 \end{bmatrix}, \boldsymbol{A}_{eq,\boldsymbol{\beta}_{1f}^{(i)}(t)} = \begin{bmatrix} 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix} \tag{A.17}$$

$$\boldsymbol{A}_{eq,\boldsymbol{\Psi}_2^{(i)}(t)} = \begin{bmatrix} \boldsymbol{A}_{eq,\boldsymbol{\xi}_2^{(i)}(t)} \\ \boldsymbol{A}_{eq,\boldsymbol{\beta}_{2b}^{(i)}(t)} \\ \boldsymbol{A}_{eq,\boldsymbol{\beta}_{2f}^{(i)}(t)} \end{bmatrix}, \boldsymbol{A}_{eq,\boldsymbol{\xi}_2^{(i)}(t)} = \begin{bmatrix} 1 & \cdots & 1 & 0 & \cdots & 0 \end{bmatrix}, \tag{A.18}$$

$$\boldsymbol{A}_{eq,\boldsymbol{\beta}_{2b}^{(i)}(t)} = \begin{bmatrix} 0 & \cdots & 0 & 1 & \cdots & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 1 & \cdots & 1 \end{bmatrix}, \boldsymbol{A}_{eq,\boldsymbol{\beta}_{2f}^{(i)}(t)} = \begin{bmatrix} -1 & \cdots & 0 & 1 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & -1 & 0 & \cdots & 1 & \cdots & 0 & \cdots & 1 \end{bmatrix}, \tag{A.19}$$

The vector $\boldsymbol{b}_{eq}$ has the following form:

$$\boldsymbol{b}_{eq} = \begin{bmatrix} \boldsymbol{b}_{eq,a} & \boldsymbol{b}_{eq,r} \end{bmatrix}^T, \tag{A.20}$$

$$\boldsymbol{b}_{eq,c} = \begin{bmatrix} \boldsymbol{b}_{eq,\boldsymbol{\Theta}(1)} \\ \vdots \\ \boldsymbol{b}_{eq,\boldsymbol{\Theta}(T)} \end{bmatrix}, \ \boldsymbol{b}_{eq,\boldsymbol{\Theta}(t)} = \begin{bmatrix} \boldsymbol{b}_{eq,\boldsymbol{\Psi}^{(1)}(t)} \\ \vdots \\ \boldsymbol{b}_{eq,\boldsymbol{\Psi}^{(N)}(t)} \end{bmatrix}, \ \boldsymbol{b}_{eq,\boldsymbol{\Psi}^{(i)}(t)} = \begin{bmatrix} \boldsymbol{b}_{eq,\boldsymbol{\xi}^{(i)}(t)} \\ \boldsymbol{b}_{eq,\boldsymbol{\beta}^{(i)}(t)} \end{bmatrix}, \tag{A.21}$$

$$\boldsymbol{b}_{eq,\boldsymbol{\xi}^{(i)}(t)} = \begin{bmatrix} 1 \end{bmatrix}, \quad \boldsymbol{b}_{eq,\boldsymbol{\beta}^{(i)}(t)} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \tag{A.22}$$

$$\boldsymbol{lb} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \ \boldsymbol{ub} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \tag{A.23}$$

The matrix $\boldsymbol{A}_{ineq}$ is given by:

$$\boldsymbol{A}_{ineq} = \begin{bmatrix} \boldsymbol{A}_{ineq,\boldsymbol{\Theta}(1)} \\ \vdots \\ \boldsymbol{A}_{ineq,\boldsymbol{\Theta}(T)} \end{bmatrix}, \quad \boldsymbol{A}_{ineq,\boldsymbol{\Theta}(t)} = \begin{bmatrix} \boldsymbol{A}_{ineq,\boldsymbol{\Psi}^{(1)}(t)} \\ \vdots \\ \boldsymbol{A}_{ineq,\boldsymbol{\Psi}^{(N)}(t)} \end{bmatrix}, \tag{A.24}$$

$$\boldsymbol{A}_{ineq,\boldsymbol{\Psi}^{(i)}(t)} = \begin{bmatrix} \boldsymbol{A}_{ineq,\boldsymbol{\Psi}^{(i)}(t),a} & \boldsymbol{A}_{ineq,\boldsymbol{\Psi}^{(i)}(t),r} \end{bmatrix}, \quad \boldsymbol{A}_{ineq,\boldsymbol{\Psi}^{(i)}(t),r} = -\boldsymbol{A}_{ineq,\boldsymbol{\Psi}^{(i)}(t),a} \tag{A.25}$$

$$\boldsymbol{A}_{ineq,\boldsymbol{\Psi}^{(i)}(t),a} = \begin{bmatrix} 0 & \cdots & 0 & 1 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & -1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & & & \\ 0 & \cdots & 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & -1 & 0 & \cdots & 0 \end{bmatrix} \tag{A.26}$$

Finally, the vector $\boldsymbol{b}_{ineq}$ is given by:

$$\boldsymbol{b}_{ineq} = \begin{bmatrix} \boldsymbol{b}_{ineq,\boldsymbol{\Theta}(1)} \\ \vdots \\ \boldsymbol{b}_{ineq,\boldsymbol{\Theta}(T)} \end{bmatrix}, \ \boldsymbol{b}_{ineq,\boldsymbol{\Theta}(t)} = \begin{bmatrix} \boldsymbol{b}_{ineq,\boldsymbol{\Psi}^{(1)}(t)} \\ \vdots \\ \boldsymbol{b}_{ineq,\boldsymbol{\Psi}^{(N)}(t)} \end{bmatrix}, \ \boldsymbol{b}_{ineq,\boldsymbol{\Psi}^{(i)}(t)} = \begin{bmatrix} \alpha \\ \vdots \\ \alpha \end{bmatrix} \tag{A.27}$$

References

[1] D. Archer, Global warming: understanding the forecast, 2nd Edition, John Wiley & Sons, 2012.

[2] C. Rosenzweig, D. Karoly, M. Vicarelli, P. Neofotis, Q. Wu, G. Casassa, A. Menzel, T. Root, N. Estrella, B. Seguin, P. Tryjanowski, C. Liu, S. Rawlins, A. Imeson, Attributing physical and biological impacts to anthropogenic climate change, Nature 453 (7193) (2008) 353–357.

[3] N. Oreskes, The scientific consensus on climate change, Science 306 (5702) (2004) 1686–1686.

[4] H. Farhangi, The path of the smart grid, IEEE Power Energy Mag. 8 (1) (2010) 18–28.

[5] K. C. Armel, A. Gupta, G. Shrimali, A. Albert, Is disaggregation the holy grail of energy efficiency? The case of electricity, Energy Policy 52 (2013) 213–234.

[6] C. Fischer, Feedback on household electricity consumption: a tool for saving energy?, Energy efficiency 1 (1) (2008) 79–104.

[7] S. Darby, The effectiveness of feedback on energy consumption, Tech. rep., University of Oxford, Oxford, UK (2006).

[8] K. Ehrhardt-Martinez, K. A. Donnelly, J. A. Laitner, Advanced metering initiatives and residential feedback programs: a meta-review for household electricity-saving opportunities, Tech. Rep. E105, American Council for an Energy-Efficient Economy Washington, DC (2010).

[9] J. Laitner, K. Ehrhardt-Martinez, V. McKinney, Examining the scale of the behaviour energy efficiency continuum, in: American Council for an Energy Efficient Economy, European Council for an Energy Efficient Economy Conference, Cote d'Azur, France, 2009, paper ID 1367.

[10] G. Gardner, P. Stern, The short-list: the most effective actions us households can take to curb climate change, Environment: Science and Policy for a Sustainable Environment 50 (5) (2008) 12–24.

[11] I. Abubakar, S. Khalid, M. Mustafa, H. Shareef, M. Mustapha, Application of load monitoring in appliances' energy management – a review, Renewable and Sustainable Energy Reviews 67 (2017) 235–245.

[12] A. Zoha, A. Gluhak, M. A. Imran, S. Rajasegarar, Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey, Sensors 12 (12) (2012) 16838–16866.

[13] Z. Wang, G. Zheng, Residential appliances identification and monitoring by a nonintrusive method, IEEE Trans. Smart Grid 3 (1) (2012) 80–92.

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

[15] R. Bonfigli, S. Squartini, M. Fagiani, F. Piazza, Unsupervised algorithms for non-intrusive load monitoring: An up-to-date overview, in: Proc. of IEEE 15th Int. Conf. on Environment and Electrical Engineering (EEEIC), 2015, pp. 1175–1180.

[16] M. Zeifman, K. Roth, Nonintrusive appliance load monitoring: Review and outlook, IEEE Trans. Consum. Electron. 57 (1) (2011) 76–84.

[17] H. Kim, M. Marwah, M. F. Arlitt, G. Lyon, J. Han, Unsupervised disaggregation of low frequency power measurements, in: Proc. 11th SIAM Int. Conf. Data Mining, Mesa, AZ, USA, 2011, pp. 747–758.

[18] J. Kolter, T. Jaakkola, Approximate inference in additive factorial HMMs with application to energy disaggregation, Journal of Machine Learning Research 22 (2012) 1472–1482.

[19] M. Zhong, N. Goddard, C. Sutton, Signal aggregate constraints in additive factorial HMMs with application to energy disaggregation, in: Advances in Neural Information Processing Systems, 2014, pp. 3590–3598.

[20] M. Zhong, N. Goddard, C. Sutton, Interleaved factorial non-homogeneous hidden markov models for energy disaggregation, in: Proc. of Advances in Neural Information Processing System, Workshop on Machine Learning for Sustainability, Lake Tahoe, Nevada, USA, 2014, pp. 1–5.

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

[22] O. Parson, M. Weal, A. Rogers, A scalable non-intrusive load monitoring system for fridge-freezer energy efficiency estimation, in: Proc. of the 2nd Int. Workshop on Non-Intrusive Load Monitoring, Austin, TX, USA, 2014.

[23] J. Kelly, W. Knottenbelt, Neural NILM: Deep neural networks applied to energy disaggregation, in: Proc. of the 2nd ACM Int. Conf. on Embedded Systems for Energy-Efficient Built Environments, BuildSys '15, ACM, New York, NY, USA, 2015, pp. 55–64.

[24] L. Mauch, B. Yang, A new approach for supervised power disaggregation by using a deep recurrent LSTM network, in: Proc. of GlobalSIP, Orlando, FL, USA, 2015, pp. 63–67.

[25] L. Mauch, B. Yang, A novel DNN-HMM-based approach for extracting single loads from aggregate power signals, in: Proc. of ICASSP, Shanghai, China, 2016, pp. 2384–2388.

[26] I. Valera, F. Ruiz, F. Perez-Cruz, Infinite Factorial Unbounded-State Hidden Markov Model, IEEE Trans. Pattern Anal. Mach. Intell. 38 (9) (2016) 1816–1828.

[27] B. Zhao, L. Stankovic, V. Stankovic, On a Training-Less Solution for Non-Intrusive Appliance Load Monitoring Using Graph Signal Processing, IEEE Access 4 (2016) 1784–1799.

[28] A. Cominola, M. Giuliani, D. Piga, A. Castelletti, A. Rizzoli, A hybrid signature-based iterative disaggregation algorithm for non-intrusive load monitoring, Applied Energy 185, Part 1 (2017) 331–344.

[29] L. K. Norford, S. B. Leeb, Non-intrusive electrical load monitoring in commercial buildings based on steady-state and transient load-detection algorithms, Energy and Buildings 24 (1) (1996) 51 – 64.

[30] M. Berges, E. Goldman, H. S. Matthews, L. Soibelman, K. Anderson, User-centered nonintrusive electricity load monitoring for residential buildings, Journal of Computing in Civil Engineering 25 (6) (2011) 471–480.

[31] H.-H. Chang, C.-L. Lin, H.-T. Yang, Load recognition for different loads with the same real power and reactive power in a non-intrusive load-monitoring system, in: Proc. of the 12th Int. Conf. on Computer Supported Cooperative Work in Design (CSCWD), Xi'an, China, 2008, pp. 1122–1127.

[32] M.-S. Tsai, Y.-H. Lin, Modern development of an adaptive non-intrusive appliance load monitoring system in electricity energy conservation, Applied Energy 96 (2012) 55–73.

[33] H.-H. Chang, Non-intrusive demand monitoring and load identification for energy management systems based on transient feature analyses, Energies 5 (11) (2012) 4569–4589.

[34] H.-H. Chang, P. W. Wiratha, N. Chen, A non-intrusive load monitoring system using an embedded system for applications to unbalanced residential distribution systems, Energy Procedia 61 (2014) 146–150.

[35] R. Bonfigli, E. Principi, S. Squartini, M. Fagiani, M. Severini, F. Piazza, User-aided Footprint Extraction for Appliance Modelling in Non-Intrusive Load Monitoring, in: Proc. of the IEEE Symposium Series on Computational Intelligence, Athens, Greece, 2016.

[36] Y. F. Wong, Y. Ahmet Sekercioglu, T. Drummond, V. S. Wong, Recent approaches to non-intrusive load monitoring techniques in residential settings, in: Proc. of the IEEE Symp. on Computational Intelligence Applications In Smart Grid (CIASG), Singapore, Singapore, 2013, pp. 73–79.

[37] A. Zoha, A. Gluhak, M. Nati, M. Imran, Low-power appliance monitoring using Factorial Hidden Markov Models, in: Proc. of the 8th Int. Conf. on Intelligent Sensors, Sensor Networks and Information Processing: Sensing the Future (ISSNIP), Vol. 1, Melbourne, VIC, Australia, 2013, pp. 527–532.

[38] Y. Li, Z. Peng, J. Huang, Z. Zhang, J. H. Son, Energy Disaggregation via Hierarchical Factorial HMM, in: Proc. of the 2nd Int. Workshop on Non-Intrusive Load Monitoring, Austin, TX, USA, 2014.

[39] S. Makonin, F. Popowich, I. V. Bajić, B. Gill, L. Bartram, Exploiting HMM Sparsity to Perform Online Real-Time Nonintrusive Load Monitoring, IEEE Transactions on Smart Grid 7 (6) (2016) 2575–2584.

[40] M. Johnson, A. Willsky, Bayesian nonparametric hidden semi-Markov models, Journal of Machine Learning Research 14 (1) (2013) 673–701.

[41] J. Z. Kolter, M. J. Johnson, REDD: A public data set for energy disaggregation research, in: Proc. of the SustKDD Workshop on Data Mining Appliations in Sustainbility, San Diego, CA, USA, 2011, pp. 1–6.

[42] M. Figueiredo, A. De Almeida, B. Ribeiro, Home electrical signal disaggregation for non-intrusive load monitoring (NILM) systems, Neurocomputing 96 (2012) 66–73.

[43] J. M. Gillis, S. M. Alshareef, W. G. Morsi, Nonintrusive Load Monitoring Using Wavelet Design and Machine Learning, IEEE Trans. Smart Grid 7 (1) (2016) 320–328.

[44] S. Makonin, F. Popowich, L. Bartram, B. Gill, I. V. Bajic, AMPds: A public dataset for load disaggregation and eco-feedback research, in: Proc. of the IEEE Electrical Power and Energy Conference (EPEC), Halifax, NS, Canada, 2013.

[45] M. Aiad, P. H. Lee, Unsupervised approach for load disaggregation with devices interactions, Energy and Buildings 116 (2016) 96–103.

[46] F. Paradiso, F. Paganelli, D. Giuli, S. Capobianco, Context-based energy disaggregation in smart homes, Future Internet 8 (1) (2016).

[47] A. Cole, A. Albicki, Algorithm for nonintrusive identification of residential appliances, in: Proc. of the IEEE Int. Symp. on Circuits and Systems (ISCAS), Monterey, CA, USA, 1998, pp. 338–341.

[48] C. Laughman, K. Lee, R. Cox, S. Shaw, S. Leeb, L. Norford, P. Armstrong, Power signature analysis, Power and Energy Magazine, IEEE 1 (2) (2003) 56–63.

[49] A. Marchiori, D. Hakkarinen, Q. Han, L. Earle, Circuit-level load monitoring for household energy management, IEEE Pervasive Computing 10 (1) (2011) 40–48.

[50] M. Weiss, A. Helfenstein, F. Mattern, T. Staake, Leveraging smart meter data to recognize home appliances, in: Proc. of IEEE Int. Conf. on Pervasive Computing and Communications (PerCom), Lugano, Switzerland, 2012, pp. 190–197.

[51] H. Goncalves, A. Ocneanu, M. Berges, Unsupervised disaggregation of appliances using aggregated consumption data, in: Proc. of the 1st KDD Workshop on Data Mining Applications in Sustainability (SustKDD), San Diego, CA, USA, 2011.

[52] R. Bonfigli, M. Severini, S. Squartini, M. Fagiani, F. Piazza, Improving the performance of the AFAMAP algorithm for non-intrusive load monitoring, in: Proc. of the IEEE Congress on Evolutionary Computation (CEC), Vancouver, Canada, 2016, pp. 303–310.

[53] L. Rabiner, A tutorial on hidden markov models and selected applications in speech recognition, Proceedings of the IEEE 77 (2) (1989) 257–286.

[54] J. A. Hartigan, M. A. Wong, Algorithm as 136: A k-means clustering algorithm, Journal of the Royal Statistical Society. Series C (Applied Statistics) 28 (1) (1979) 100–108.

[55] O. Parson, S. Ghosh, M. Weal, A. Rogers, Non-intrusive load monitoring using prior models of general appliance types, in: Twenty-Sixth Conference on Artificial Intelligence (AAAI-12), 2012.

[56] S. Makonin, F. Popowich, Nonintrusive load monitoring (NILM) performance evaluation, Energy Efficiency 8 (4) (2014) 809–814.