Multivariate Event Detection Methods for Non-Intrusive Load Monitoring in Smart Homes and Residential Buildings

Sarra Houidi, François Auger, Houda Ben Attia Sethom, Dominique Fourer, Laurence Miègeville

HAL Id: hal-02386782

https://hal.science/hal-02386782

Submitted on 21 Dec 2021

Multivariate Event Detection Methods for Non-Intrusive Load Monitoring in Residential Buildings

Sarra Houidi[a,b], François Auger[a], Houda Ben Attia Sethom[b,c], Dominique Fourer[d], Laurence Miègeville[a]

[a] *University of Nantes, Institut de Recherche en Énergie Électrique de Nantes Atlantique (IREENA), EA 4642, Saint-Nazaire, France*
[b] *Université de Tunis El Manar, École Nationale d'Ingénieurs de Tunis, LR11ES15, Laboratoire des Systèmes Électriques 1002, Tunis, Tunisia*
[c] *Université de Carthage, École Nationale d'Ingénieurs de Carthage, 2035, Tunis, Tunisia*
[d] *University of Évry-Val-d'Essonne/Paris-Saclay, Informatique, Bio-informatique et Systèmes Complexes (IBISC), EA 4526, Courcouronnes, France*

Abstract

Non-Intrusive Load Monitoring (NILM) approaches refer to the analysis of the aggregated electrical signals of Home Electrical Appliances (HEAs) in order to identify their operating schedules. It has emerged as a promising solution to help residential consumers to reduce their electricity bills through a breakdown of energy consumption. NILM methods are either event-based or non event-based. This categorization depends on whether or not they rely on the detection of HEAs' significant state transitions (e.g., On/Off or state change) in power consumption signals. This paper focuses on event-based approaches and especially in multivariate change detection algorithms. It aims at highlighting the benefits brought by a multivariate approach for change detection using the appropriate electrical features. We first suggest to extend four existing change detection algorithms in the multidimensional case. The studied detection algorithms are first detailed and compared to each other and to their existing scalar versions through numerical simulations. Then, a new feature selection

Email addresses: `sarra.houidi@etu.univ-nantes.fr` (Sarra Houidi),
`francois.auger@univ-nantes.fr` (François Auger), `houda.benattia@enicarthage.rnu.tn`
(Houda Ben Attia Sethom), `dominique.fourer@univ-evry.fr` (Dominique Fourer),
`laurence.miegeville@univ-nantes.fr` (Laurence Miègeville)

algorithm for change detection is presented and assessed when combined with the most efficient detector among the four investigated ones. Finally, the feature selection method for detection purposes is applied to two different NILM case studies. The first one uses power features derived from the Building-Level fully-labeled Dataset for Electricity Disaggregation (BLUED) current and voltage measurements and the second one is based on current and voltage measurements acquired using our own acquisition system. Compared to the classical scalar approach, the results show that the multivariate approach brings a significant performance improvement when the features selected by the proposed algorithm are used.

1. Introduction

In the current context of energy transition, it is essential to better monitor the domestic consumption. Feedback on electricity consumption appears to be a major tool to save energy through more conscious user behaviour [1]. NILM
5 has emerged as a promising solution to provide a breakdown of the residential energy consumption without instrumenting HEAs [2, 3, 4]. The appliances identification is made through the analysis of the current and voltage signals using a single measuring device connected to the house electrical panel at the Point of Common Coupling (PCC) [5]. This information is relayed back to the consumers,
10 who can therefore make an informed decision about energy savings. Throughout literature, a large number of NILM approaches are reported [2, 6]. Most of the methods can be classified into supervised or unsupervised methods. The former rely on classification algorithms that require a sufficient amount of labeled data for a training process [6, 7]. On the opposite, the latter methods are based on
15 clustering procedures which do not need a labeled training dataset [6]. Finally, NILM approaches can be categorized into non event-based and event-based ap-

proaches [8]. Non event-based methods disaggregate the power consumption using techniques such as the Hidden Markov models (HMM) [9, 10] which identify repetitive HEAs' patterns in the load curve based on HEAs previous states.

20 The event-based approaches assume that each event in the total household power consumption is a response to a Home Electrical Appliance (HEA) state change [8]. Fig. 1 shows an overview of event-based NILM approaches. In this research work, a specific interest is given to HEAs' event detection which is an elementary step in the NILM pipeline [11].
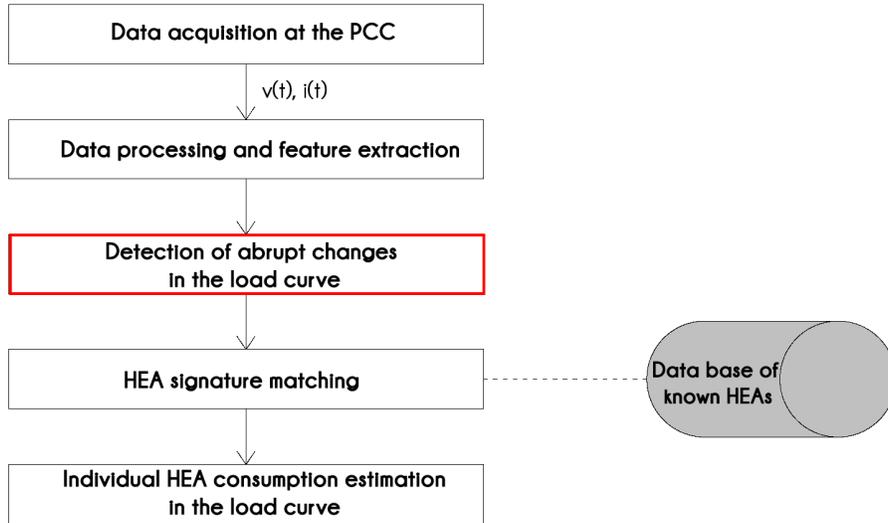
Figure 1: Block diagram of the event-based NILM approaches.

25 According to the existing literature on detection algorithms for event-based NILM methods, two main approaches are considered: the heuristic and the probabilistic methods [12]. Algorithms under the first category analyze a time series of data, looking for changes above a given threshold. In [13], the events are detected in the total active power signal by computing the absolute value of

30 the difference between two consecutive samples, which is then compared to a pre-defined threshold. M.N. Meziane *et al.* in [14], propose an event detector

3

called High Accuracy NILM Detector (HAND) that tracks the standard deviation variation of the current signal envelope using a sliding window. A threshold separates the events characterized by high standard deviations from the steady states defined by much smaller standard deviations. They obtained a probability of detection of 96.7 % on simulated data.

Algorithms in the second category of probabilistic methods comprise three entities: a stochastic process under observation, a change point at which the statistical properties of the process undergo a change, and a decision maker that observes and detects the change of the process statistical properties [15, 16]. In the NILM context, three algorithms are mainly used to detect changes in active power signals: the Generalized Likelihood Ratio (GLR) test, the Chi-squared Goodness Of-Fit (χ^2 GOF) and the CUmulative Sum (CUSUM) detectors. Authors in [17, 18] use the GLR approach to test if two consecutive time frames share a common distribution by deriving a decision function from the log-probability distribution ratio before and after a potential change of the mean value. In [18, 19], a χ^2 test statistic is used to assess if two neighboring windows share a common distribution. If this is not the case, an event is then detected. Researchers in [20, 21] apply a CUSUM algorithm for the detection of both the beginning and the end of a HEA transient-state active power signal. Z. Zhu *et al.* [21] reached a probability of detection of 90 % by applying their approach on real data comprising 200 events of eight different HEAs switched on and off.

The common thread in all these research works is that the detection is done in a univariate context by only considering the active power signal, whereas prior literature on NILM focuses on identifying an effective features' set that defines a unique HEA signature [22, 23, 24]. Fig. 2 illustrates three power time series related to the same scenario of several HEAs switched on and off using our own acquisition system [25], and three power time series derived from BLUED dataset [26]. It can be observed that changes are present in the active power time series, the reactive power time series and their 4th and 5th order harmonic. The information provided by these time series could be used together to improve the detection robustness. To the best of our knowledge, there is no existing liter-

ature on multivariate change detectors designed for NILM event-based methods. Consequently, the goal of this paper is to demonstrate the benefit brought by

using a multivariate approach for change detection with the most relevant features. Reference [27] presents our preliminary results related to the application of univariate change detectors on active power signals. In this paper, the pool of studied change detectors is enlarged and extended to the multivariate case by considering several power features such as the active power, the reactive power

and their respective harmonics. Four algorithms suited for mean change detection are considered in both the univariate and the multivariate cases. These change detectors are the Bayesian Information Criterion (BIC), the CUSUM, the Hotelling T^2 test (equivalent to the GLR statistic test for normal distributions) and an updated version of the Effective Residual algorithm described in [27].

This paper is organized as follows. In Section 2, the four change detectors are presented. Their performances are assessed through Monte Carlo experiments using detection performance evaluation metrics, considering both univariate and multivariate cases. In Section 3, the proposed algorithm called Feature Selection Algorithm for Detection Purposes (FSADP) is introduced. Section 4 focuses on

the experimental results when applying the FSADP on real-world power signals corresponding to a scenario of HEAs switched on and off. Firstly, power signals are obtained from the BLUED dataset current and voltage measurements [26], and secondly, power signals are computed from current and voltage measurements using our own acquisition system. The subsequent discussion addresses

the selection of the power features that are best suited for the detection task. Finally, the main conclusions and contributions of the current work are provided in Section 5.
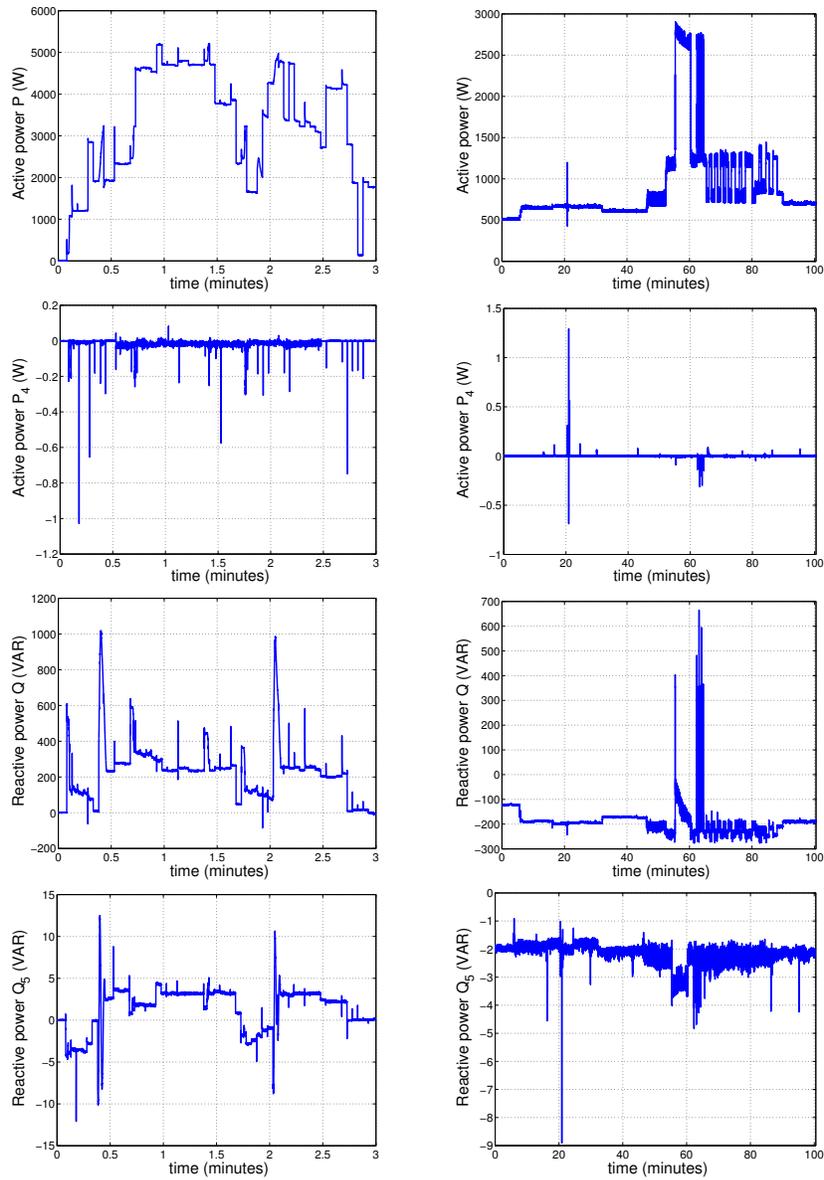
Figure 2: Active power signal, its 4th harmonic, reactive power signal and its 5th harmonic related to the same scenario of HEAs switched on and off (using our own acquisition system at the left and an extract from the phase B of BLUED dataset at the right).

6

2. Change detectors and statistical assessment

2.1. Hypothesis testing framework

In this subsection, the mathematical formulation of the detection problem is introduced. Let $x_m \in \mathbb{R}^p$ be a multidimensional time series such as $x_m = (x_{m,1}, \ldots, x_{m,j}, \ldots, x_{m,p})^{\mathrm{T}}$, where $x_{m,j}$ is the value of feature $j \in \{1, \ldots, p\}$ at time instant m. Let $X_{n,p} = (x_{n_a}, x_{n_a+1}, \ldots, x_m, \ldots, x_n)$ with $m \in \{n_a, \ldots, n\}$, be the $p \times (w+u)$ matrix (w being an even integer) of the last $n - n_a + 1 = w + u$ samples of x_m until the current time instant n. Since a HEA transient state is not necessarily an abrupt change, we use the "Window with Margins" approach [28], where: separate computations are applied to the first $w/2$ samples (located at the left) and to the last $w/2$ samples (located at the right), and u samples (located at the middle) are unused, as shown in Fig. 3. This approach [28] is adjusted to the dealt with multivariate case. It has two parameters subject to user choice: the number of useful samples w and the number of unused samples u.
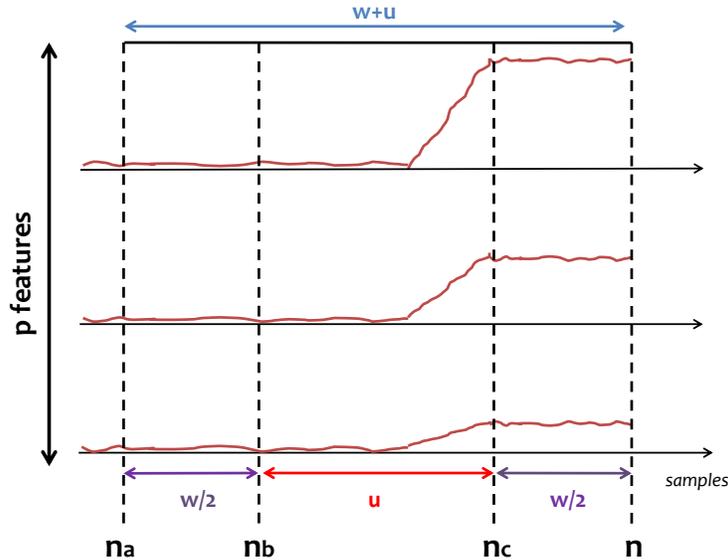

Figure 3: Sliding "Window with Margins" method adjusted to the multivariate case (p=3 features).

Each vector x_m is assumed to follow a multivariate Gaussian distribution $\mathcal{N}_p(\mu, \Sigma)$ where $\mu \in \mathbb{R}^p$ is the mean vector and $\Sigma \in \mathbb{R}^{p \times p}$ is the semidefinite covariance matrix. A possible change is modeled by a modification of the statistical parameters (i.e., mean vector and/or covariance matrix) occurring between n_b and n_c. Two hypotheses are then considered:

$$H_0 \quad : \quad x_{n_a}, \ldots, x_{n_b}, x_{n_c}, \ldots, x_n \sim \mathcal{N}_p(\mu_0, \Sigma_0) \tag{1}$$

$$H_1 \quad : \quad x_{n_a}, \ldots, x_{n_b} \sim \mathcal{N}_p(\mu_{1a}, \Sigma_{1a})$$

$$x_{n_c}, \ldots, x_n \sim \mathcal{N}_p(\mu_{1b}, \Sigma_{1b}) \tag{2}$$

Under the "without change" hypothesis H_0, the signal samples x_m follow a multivariate Gaussian distribution with a mean vector μ_0 and a covariance matrix Σ_0. Under the "with change" hypothesis H_1, between n_a and n_b, the signal samples x_m follow a multivariate Gaussian distribution with a mean vector μ_{1a} and a covariance matrix Σ_{1a}. Between n_c and n, a change has occured, the signal samples follow a multivariate Gaussian distribution with a mean vector $\mu_{1b} \neq \mu_{1a}$ and/or a covariance matrix $\Sigma_{1b} \neq \Sigma_{1a}$. Therefore, a decision between H_0 and H_1 has to be made at each time instant n. This is done by comparing a decision function g_n to an adimensional threshold h (i.e., reject H_0 in favor of H_1 if $g_n > h$) [15, 29].

2.2. Detectors' decision functions

In this subsection, the decision function of each detector is presented. The details of their mathematical derivation are given in a publicly available report [30]. Each decision function is proposed in the multivariate case. For the univariate case, the expressions can simply be inferred by considering $p = 1$. For each detector, we assume that the samples x_m are taken from independent and identically distributed ($i.i.d$) Gaussian random vectors. We also assume that the estimated covariance matrices are invertible and that $w > 2(p + 1)$.

2.2.1. Bayesian Information Criterion (BIC)

The BIC algorithm is commonly used for acoustic change detection [31]. Under hypothesis H_i, $i \in \{0, 1\}$, the BIC of $X_{n,p}$ is defined as a maximum

8

likelihood criterion penalized by the model complexity [31], that is related to the number M_i of free parameters in the probabilistic model to be estimated,

$$\text{BIC}_n(\text{H}_\text{i}) = \max_{\mu,\Sigma} \ln(\mathcal{L}_{n,i}) - \frac{\lambda}{2} M_i \ln(w) \tag{3}$$

where λ is a penalty factor, ideally equal to 1 [31] and $\mathcal{L}_{n,i}$ is the likelihood function under hypothesis H_i, which corresponds to the joint Probability Density Function (PDF) of the observed data [32]:

$$\mathcal{L}_{n,0} = \prod_{m=n_a}^{n_b} p_{\mu_0,\Sigma_0}(x_m) \prod_{m=n_c}^{n} p_{\mu_0,\Sigma_0}(x_m) \tag{4}$$

$$\mathcal{L}_{n,1} = \prod_{m=n_a}^{n_b} p_{\mu_{1a},\Sigma_{1a}}(x_m) \prod_{m=n_c}^{n} p_{\mu_{1b},\Sigma_{1b}}(x_m) \tag{5}$$

$$p_{\mu,\Sigma}(x_m) = \frac{\exp\left(-\frac{1}{2}(x_m - \mu)^T \Sigma^{-1}(x_m - \mu)\right)}{(2\pi)^{\frac{p}{2}} \det(\Sigma)^{\frac{1}{2}}} \tag{6}$$

The BIC decision function is:

$$g_n = \text{BIC}_\text{n}(\text{H}_1) - \text{BIC}_\text{n}(\text{H}_0) \tag{7}$$

This expression is maximized using the Maximum Likelihood Estimators (MLEs) of the covariance matrices and the mean vectors [32], which are:

$$\hat{\Sigma}_{1a} = \frac{1}{n_b - n_a + 1} \sum_{m=n_a}^{n_b} (x_m - \hat{\mu}_{1a})(x_m - \hat{\mu}_{1a})^{\text{T}} \tag{8}$$

$$\hat{\Sigma}_{1b} = \frac{1}{n - n_c + 1} \sum_{m=n_c}^{n} (x_m - \hat{\mu}_{1b})(x_m - \hat{\mu}_{1b})^{\text{T}} \tag{9}$$

$$\hat{\Sigma}_0 = \frac{1}{w} \left(\sum_{m=n_a}^{n_b} (x_m - \hat{\mu}_0)(x_m - \hat{\mu}_0)^{\text{T}} + \sum_{m=n_c}^{n} (x_m - \hat{\mu}_0)(x_m - \hat{\mu}_0)^{\text{T}} \right)$$
$$= \frac{1}{2}\left(\hat{\Sigma}_{1a} + \hat{\Sigma}_{1b}\right) + \frac{1}{4}\left(\hat{\mu}_{1b} - \hat{\mu}_{1a}\right)\left(\hat{\mu}_{1b} - \hat{\mu}_{1a}\right)^{\text{T}} \tag{10}$$

$$\text{with}\quad \hat{\mu}_{1a} = \frac{1}{n_b - n_a + 1} \sum_{m=n_a}^{n_b} x_m, \quad \hat{\mu}_{1b} = \frac{1}{n - n_c + 1} \sum_{m=n_c}^{n} x_m \tag{11}$$

$$\text{and}\quad \hat{\mu}_0 = \frac{n_b - n_a + 1}{w}\hat{\mu}_{1a} + \frac{n - n_c + 1}{w}\hat{\mu}_{1b} = \frac{1}{2}(\hat{\mu}_{1a} + \hat{\mu}_{1b}) \tag{12}$$

9

Using Eqs. (3) and (7), the BIC decision rule is [30]:

$$g_n \underset{H_0}{\overset{H_1}{\gtrless}} h \text{ with } g_n = \frac{\det(\hat{\Sigma}_0)^2}{\det(\hat{\Sigma}_{1a})\det(\hat{\Sigma}_{1b})} \text{ and } h \geq 1 \tag{13}$$

2.2.2. Hotelling T^2 test

The Hotelling T^2 test is also widely used for NILM purposes [20]. It is derived from the ratio Λ_n of the likelihood functions when considering H_0 and H_1 [33]:

$$\Lambda_n = \frac{\mathcal{L}_{n,1}}{\mathcal{L}_{n,0}} \tag{14}$$

This likelihood ratio is maximized using the MLEs expressed in Eqs. (8)-(12), and a development of its expression which is outlined in detail in our online report [30], leads to:

$$\left(\frac{\det(\hat{\Sigma}_1)}{\det(\hat{\Sigma}_0)}\right)^{-\frac{w}{2}} = (1 + g_n)^{\frac{w}{2}}, \text{ with } \hat{\Sigma}_1 = \frac{1}{2}\left(\hat{\Sigma}_{1a} + \hat{\Sigma}_{1b}\right) \tag{15}$$

where $\hat{\Sigma}_1$ is called the pooled covariance matrix. The derivation of the decision function expression g_n can be found using a mathematical development such as:

$$g_n = \frac{1}{4}\left(\hat{\mu}_{1b} - \hat{\mu}_{1a}\right)^{\mathrm{T}} \hat{\Sigma}_1^{-1} \left(\hat{\mu}_{1b} - \hat{\mu}_{1a}\right) \tag{16}$$

If the likelihood ratio is small, H_1 is rejected. According to the Neyman-Pearson Lemma [33], this is possible when Λ_n is lower than $\gamma \in [0,1]$, which leads to $g_n > \gamma^{\frac{2}{w}} - 1$. The decision function is finally compared to a threshold $h \in \mathbb{R}_+$ such as:

$$g'_n \underset{H_0}{\overset{H_1}{\gtrless}} h \quad \text{with} \quad g'_n = 4\,g_n \tag{17}$$

2.2.3. Cumulative Sum (CUSUM)

The CUSUM is a probabilistic detection method used in a large extent for event-based NILM approaches [20, 27]. It is assumed that under H_0, $\mu_0 = \mu_{1a}$

10

and $\Sigma_0 = \Sigma_{1a}$, and that under H_1, $\Sigma_{1b} = \Sigma_{1a}$. The CUSUM decision rule is derived from the log-likelihood ratio $\ln(\Lambda_n)$:

$$\ln\left(\Lambda_n\right) = \ln\left(\frac{\mathcal{L}_{n,1}}{\mathcal{L}_{n,0}}\right) = \sum_{m=n_c}^{n} \ln\left(\frac{p_{\mu_{1b},\Sigma_{1a}}(x_m)}{p_{\mu_{1a},\Sigma_{1a}}(x_m)}\right) \tag{18}$$

that is maximized using the MLEs $\hat{\mu}_{1a}$, $\hat{\mu}_{1b}$ and $\hat{\Sigma}_{1a}$ and compared to a threshold $h \in \mathbb{R}_+$,

$$g_n \underset{H_0}{\overset{H_1}{\gtrless}} h \quad \text{with} \quad g_n = (\hat{\mu}_{1b} - \hat{\mu}_{1a})^T \hat{\Sigma}_{1a}^{-1} (\hat{\mu}_{1b} - \hat{\mu}_{1a}) \tag{19}$$

Both CUSUM and Hotelling T^2 methods have very similar decision functions with quadratic forms.

2.2.4. Effective Residual

The Effective Residual [34, 35] uses the gradient of the signal and makes a stationarity assumption in the margins of the window. We propose herein an updated version of this detector and a multivariate version. The goal is to build a function that reaches its maximum at time n when a change arises between n_b and n_c. The absolute variation between two consecutive signal samples of the same feature $j \in \{1, \ldots, p\}$ is first computed as follows:

$$\delta_{m,j} = |x_{m,j} - x_{m-1,j}| \quad \forall m \in [n_a + 1, n_b] \cup [n_c + 1, n] \tag{20}$$

$$\delta_{n_c,j} = |x_{n_c,j} - x_{n_b,j}| \tag{21}$$

Then the residual $r_{m,j}$ representing the difference between two consecutive variations of the same feature j is computed as:

$$r_{m,j} = |\delta_{m,j} - \delta_{m-1,j}| \quad \forall m \in [n_a + 2, n_b] \cup [n_c + 1, n] \tag{22}$$

$$r_{n_c,j} = |\delta_{n_c,j} - \delta_{n_b,j}| \tag{23}$$

11

Finally, the decision function g_n is compared to a threshold h, and leads to:

$$g_n \underset{H_0}{\overset{H_1}{\gtrless}} h, \quad \text{with} \quad g_n = \sum_{j=1}^{p} \frac{g_{n,j}}{d_j} \tag{24}$$

$$g_{n,j} = \left(\frac{w}{2} - 2\right) \sum_{i=\frac{w}{2}-2}^{\frac{w}{2}-1} r_{n-i,j} - \sum_{i=0}^{\frac{w}{2}-3} r_{n-i,j} - \sum_{i=u+\frac{w}{2}}^{u+w-3} r_{n-i,j} \tag{25}$$

where d_j are normalization factors set by the user in order to sum features with possibly different physical dimensions.

2.3. Detectors' statistical assessment

We propose a comparative statistical assessment of the four detectors in strictly the same conditions. We aim at comparing the detectors to each other and to their univariate version to demonstrate the performance improvement brought by the multivariate case.

2.3.1. Monte Carlo experiment description

A one-dimensional signal slice $X_{n,1}$ and a two-dimensional signal slice $X_{n,2}$ (with $w = 6$ and $u = 0$) are filled $100\,000$ times with random samples drawn from a Gaussian process with a zero mean and a standard deviation $\sigma = 0.1$. This corresponds to $100\,000$ realizations of the H_0 case where no mean change occurs and the signal is only made of noise. For the H_1 case, one mean change of height $\Delta\mu$ occurring at the $(w/2+1)^{th}$ sample is added to the signal slices $X_{n,1}$ and to the first line of the signal slice $X_{n,2}$. The Step height-to-Noise Ratio is defined as SNR $= \Delta\mu/\sigma$. A mean change $\Delta\mu' = q \times \Delta\mu$ (with $q \in [0,1]$) is also added at the $(w/2+1)^{th}$ sample of the second line of $X_{n,2}$. Fig. 4 shows an example of a realization with a two-dimensional signal where the mean change of the second feature corresponds to 60% of the first one ($q = 0.6$).

The decision functions expressed in Eqs. (13), (17), (19) and (24) are computed for each realization of the both cases H_0 and H_1, and then compared to logarithmically spaced values of the threshold.
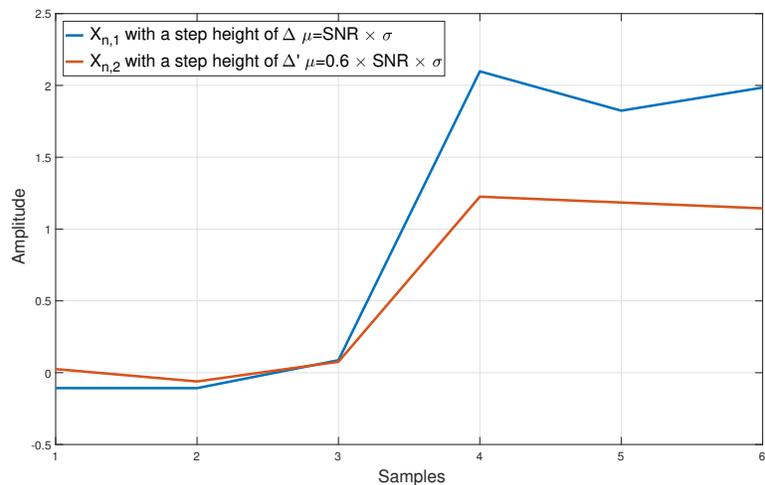
12

Figure 4: Example of a realization obtained in the Monte Carlo experiment with SNR=20, $\sigma = 0.1$ and $\Delta\mu' = 0.6 \times \Delta\mu$.

The assessment is made for SNR $= 2$, which corresponds to the smallest increase of the active power obtained when switching on a Compact Fluorescent Lamp (CFL). The goal of this Monte Carlo experiment is first to observe the effect on the four studied detectors performances when adding a spurious dimension that only contains noise (i.e., $q = 0$, corresponding to the case where a resistive load is switched on or off, leading to a mean change observed in the active power but not in the reactive power). Second, it allows to observe the effect on detection performance when adding under H_1 a feature whose step height is larger than zero. It should be noted that for the Effective Residual, all the normalization factors d_j mentioned in (25) are chosen to be equal to 1.

2.3.2. Performance evaluation tools

In order to compare the four detection algorithms, several performance measures are commonly used [2, 3, 27].

i) Classification metrics:

- True Positive (TP) which counts the times a test detects a change in

13

$X_{n,p}$ when there is really one,

- True Negative (TN) which counts the times a test does not detect a change in $X_{n,p}$ when there is really not,

- False Positive (FP) which counts the times a test detects a change in $X_{n,p}$ when there is really not,

- False Negative (FN) which counts the times a test does not detect a change in $X_{n,p}$ when there is really one.

Then, different rates or probabilities can be computed:

- The Probability of Detection P_D (or True Positive Rate) also called Recall, which measures the ability of a test to detect a truly present change. It can be estimated by the ratio of the correctly detected events over all the cases under H_1:

$$P_D = TP/(TP + FN) \tag{26}$$

- The Probability of False Alarm P_{FA} (or False Positive Rate), that can be estimated by the ratio of wrongly detected events over all the cases under H_0:

$$P_{FA} = FP/(FP + TN) \tag{27}$$

Other metrics can also be mentioned:

- the True Negative Rate $TNR = TN/(FP + TN)$
- the False Negative Rate $FNR = FN/(TP + FN)$

But since $TNR + P_{FA} = FNR + P_D = 1$, only two of the four metrics are necessary. P_D and P_{FA} are the most commonly used performance metrics for assessing detectors in the detection theory literature [16]. They are two intuitive and antagonist metrics meaningful for making a balance between the benefits, i.e., P_D, and costs, i.e., P_{FA}. The Probability of Detection is desired close to one and the Probability of False Alarm is desired close to zero.

14

ii) Receiver Operating Characteristic (ROC) curves:

In order to plot a ROC curve, a given detector decision function output is compared to a threshold h. For each value of the threshold, the performance metrics FP and TP are computed. This value of the threshold leads to one point in a plane at coordinates $(P_D(h), P_{FA}(h))$. A set of values of the threshold leads to a set of points called the ROC curve [36]. This curve indicates therefore the Probability of Detection for a given false alarm rate and gives the behavior of the detector with respect to a change of the threshold. The optimal point on the ROC curve is the nearest point to the upper left corner of the plot located at $(0, 1)$. An ideal detector produces a curve close to this point. On the opposite, a bad detector produces a curve close to or below the diagonal line from the origin $(0, 0)$ to the top right corner $(1, 1)$. Good detectors should lie somewhere between these extremes. The ROC curve of the best detector is the one which is always over the other curves in the ROC plane.

Fig. 5 depicts the ROC curves of the four studied detectors. It can be observed that the addition of a second spurious feature whose mean change is between 0 and 60% of the first feature one, slightly decreases all the detectors' performances when comparing to one dimensional signal performances. However, the detector performances are increased when the step height of the second added feature is above $0.6 \times \Delta\mu$.
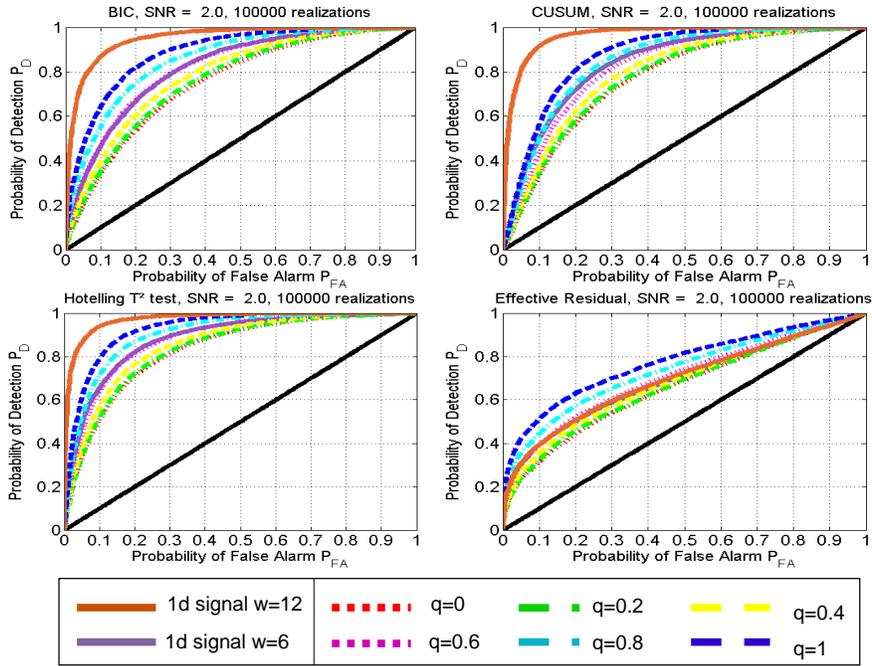
15

Figure 5: ROC curves of the four studied detectors considering one- and multi-dimensional cases with different step heights in the second dimension.

It can also be observed that the Hotelling T^2 is the detector whose ROC curves are the closest ones to the upper left corner, denoting the best performances when comparing to the other detectors. The Effective Residual performances are the worst according to the obtained ROC curves. However, unlike the three other detectors, its performances remain the same when applied to the one-dimensional signal using sliding windows of different sizes $w = 12$ or $w = 6$.

iii) Probability of Detection (P_D) for a fixed Probability of False Alarm (P_{FA}): Usually, the Area Under the ROC Curve (AUC) is used to summarize a detector's performance into a single quantity ranging between 0 and 1 [27, 36]. The larger the AUC, the better the detector. Detection rules with an AUC lower than 0.5 are worthless, while values close to

16

1 imply nearly perfect detection. However, the AUC may be somehow a misleading measure of the detector's performance. Indeed, sometimes ROC curves with equivalent AUCs can have very different shapes. To amend this deficiency, a trade-off consists in computing the P_D for different SNR values while considering a specific P_{FA} value [37]. This describes the minimal SNR required to achieve a specific operating point on the ROC curve. Usually, a constant and very small value of P_{FA} is set (often smaller than 5%) [16]. Fig. 6 plots P_D for each SNR for a specific P_{FA}. For all the detectors, the same observation as the one made for the ROC curves in Fig. 5 can be pointed out: for a given value of the SNR, the addition of a second feature, whose mean change is between 0 and 60% of the first feature one, decreases the Probability of Detection compared to a one-dimensional signal. The probabilities of detection are increased when $\Delta\mu' \geq 0.6\Delta\mu$. It can be noticed that the Hotelling T^2 detector performances are better in comparison to those of the three other detectors. Indeed, for the same SNR value, the reached P_D is the highest one for both the multivariate and univariate cases. In addition, for very small values of the SNR, which means that the step height is lost in the signal noise, the reached P_D is the greatest one. As mentioned previously, the Effective Residual method has the worst performances, i.e., for SNR=2, the reached P_D is between 20 % and 50 % for both univariate and multivariate signals, whereas for the other detectors such as the Hotelling T^2 test, the reached P_D is between 30 % and 70 %.
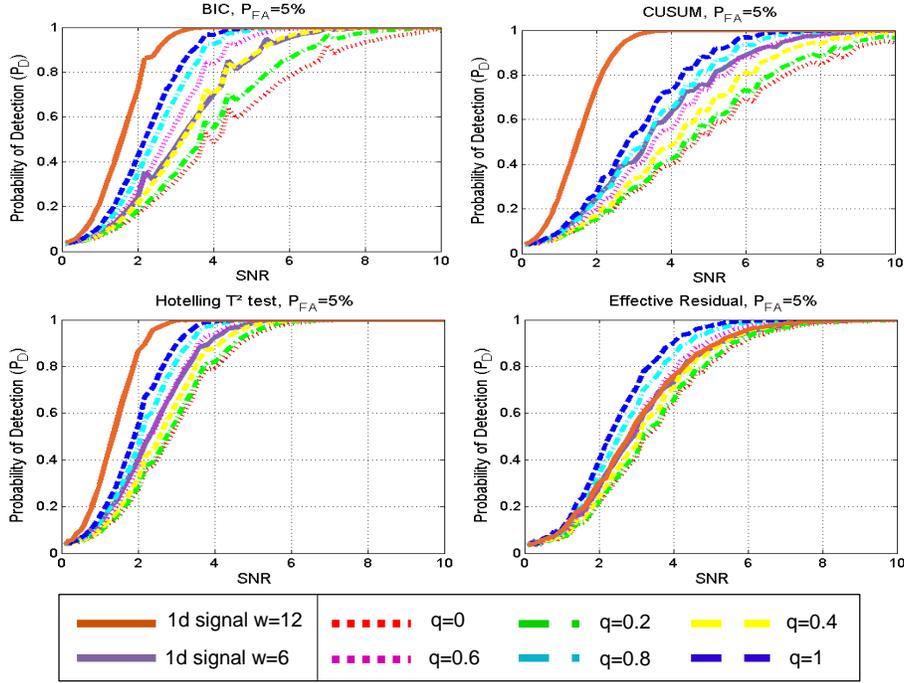
Figure 6: Probability of Detection P_D (Eq. 26) of the four studied detectors for a varying SNR value with a Probability of False Alarm $P_{FA} = 5\%$ (Eq. 27) for the one- and two-dimensional cases.

2.3.3. Optimal threshold value

The ROC analysis can provide the optimal threshold value by considering several metrics.

₂₉₀   i) The first metric is the F-measure $F_M(h) = 2\frac{P_R(h)P_D(h)}{P_R(h)+P_D(h)}$,

where $P_R = TP/(TP + FP)$ is called the precision and denotes the fraction of the correctly detected changes over the total number of times the decision rule favors H_1 [27]. The optimal threshold can be chosen as:

$$h_{1opt} = \arg\max_h F_M(h) \tag{28}$$

ii) The second metric represents the largest vertical distance from the ROC

₂₉₅   curve to the main diagonal. The optimal threshold is then defined as the

18

value that maximizes the difference between P_D and P_{FA} [38]:

$$h_{2opt} = \arg\max_h J_2(h), \ \text{with} \ J_2(h) = P_D(h) - P_{FA}(h) \tag{29}$$

iii) The last metric is known as "the closest point to the (0,1) corner in the ROC curve". The optimal threshold value is defined as the point that minimizes the Euclidean distance between the ROC curve and the (0,1) point [38], such as:

$$
\begin{aligned}
h_{3opt} &= \arg\min_h J_3(h), & (30) \\
\text{with} \quad J_3(h) &= \sqrt{(1 - P_D(h))^2 + P_{FA}(h)^2} & (31)
\end{aligned}
$$

Fig. 7 shows the considered metrics as a function of the threshold values. It can be seen that for the four detectors, the three metrics reach their extremes for very close optimal thresholds. It can also be noticed that these extremes are the highest for the Hotelling T^2. In addition, for the CUSUM and the Hotelling T^2 test detectors, the range of optimal h values for which J_2 is maximized and J_3 is minimized is larger in comparison to the BIC and the Effective Residual detectors. In particular, a score of $J_3 \leq 0.5$ is reached for several values of the threshold ranging from 10^{-1} to 10^1 for the Hotelling T^2 whereas for the Effective Residual detector, such a score is reached for $h \in [-0.05, 0.2]$. This shows that the Hotelling T^2 detector provides better results within a wider range of h values compared to the Effective Residual detector. The same observation can be made when comparing the Hotelling T^2 detector to the CUSUM for which $h \in [0.5, 50]$ and to the BIC for which $h \in [2, 5]$ when $J_3 \leq 0.5$. It can also be observed that for $h \in [0, 1]$, the BIC has no detection ability, since $P_D = P_{FA} = 1$.

Through these numerical simulations, we have evidenced the improvement brought by the multi-dimensional case for detection purposes. We also highlighted the Hotelling T^2 detector performances. Knowing that several electrical features can be extracted within a NILM framework and because there is no knowledge about a set of "ad hoc" electrical features set apart for HEAs detection, a Feature Selection (FS) process must be considered. Since the reached

19

extremes are almost the same for each metric regardless the used detector, we decide to consider the metric J_3 in what follows.
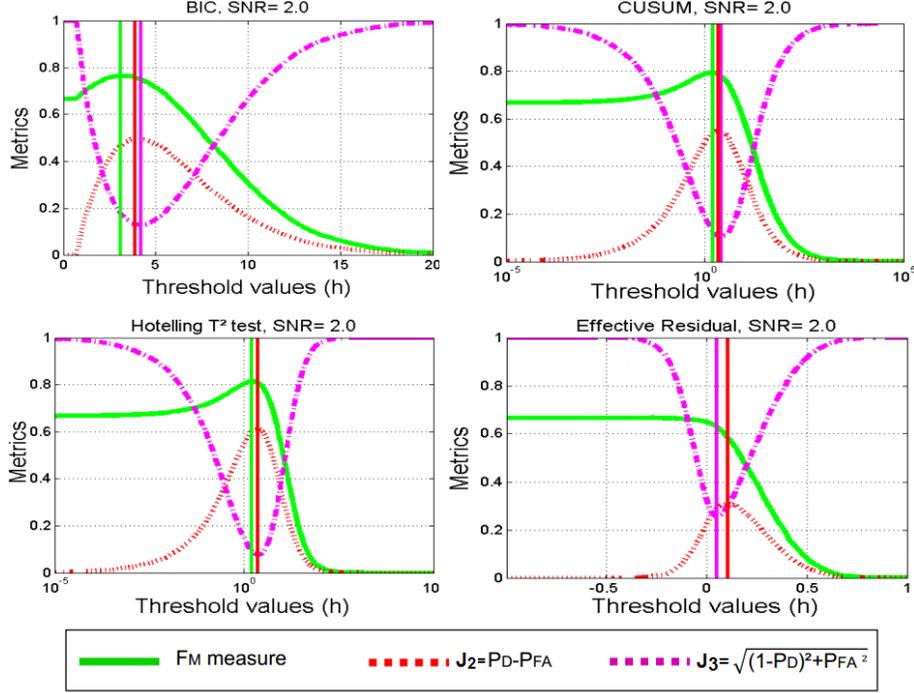
Figure 7: Three metrics values, F_M (Eq. 28), J_2 (Eq. 29) and J_3 (Eq. 31) against threshold values h of each detector for a fixed SNR value.

3. Feature Selection Algorithm for Detection Purposes (FSADP)

3.1. Motivations

In NILM approaches, the features are extracted attributes from current and voltage measurements. In Fig. 2, it can be well observed that an event induces a common change for different electrical features at different scales. In other words, some features might be informative and useful whereas others are not. FS is a process commonly used for classification tasks to reduce overfitting and to avoid the curse of dimensionality problem [39, 40]. The selected subset contains the minimal number of features that contribute most to the classification

20

accuracy [24, 40]. In this section, we aim at only using FS for event detection purposes and not for classification. The goal is to improve the detector's performance by removing irrelevant features and significantly reducing the initial p-dimensional feature space to a d-dimensional relevant features subspace, with $d \leq p$.

3.2. The proposed Feature Selection Algorithm for Detection Purposes (FSADP)

The proposed feature selection method is a forward method that starts with an empty set of features and iteratively increases its cardinality. On the opposite, a backward algorithm would start with the whole set of features and iteratively remove the least significant ones [40]. FSADP uses p iterations and requires a knowledge on the ground truth events [41]. This means that the set $\mathbb{N}_c$ of the timestamps indicating when events really occur has to be known, with $\text{Card}(\mathbb{N}_c) = N_c$. The proposed method is described by Algorithm 1. The first iteration starts with an empty set of selected features. Each feature $f \in F$, where F is the set of p features, is tested individually and the detector is applied to the signal by using a sliding window of $w+u$ successive values over the whole signal of length N. In each window, the detector decision function is computed and compared to logarithmically spaced values of the threshold $h \in H$. If the decision function is larger than h, TP is incremented when an event truly occurs, whereas FP increases when no event occurs. The optimal threshold h_{min} corresponding to the minimum value of the J_3 score is determined. The feature with the lowest minimum is then selected. At the subsequent iterations, each remaining feature is added to the set of previously selected features, and the same process is applied to find the lowest score J_3. The most relevant features are the ones that contribute to decrease the score J_3. FSADP can be applied to any dataset containing labeled events. It selects the most relevant features for which P_{FA} is the lowest and P_D is the highest. This couple (P_{FA}, P_D) is associated to a threshold h_{min} that is also the optimal one for the whole considered dataset. The time complexity of this algorithm is quadratic according to the number of features p, and proportionally increases with the cardinal of H and

21

the number of change events N_c.

Algorithm 1 FSADP

Input: $X_{n,p}$, whole features set F, signal length N, threshold values set H, window length

    $w + u$, true events time instants set $\mathbb{N}_c$, $N_c = \mathrm{Card}(\mathbb{N}_c)$

Output: Set of sorted features F_p

  1: Initialization: $F_0 \leftarrow \emptyset$

  2: **for** $k \leftarrow 0$ to $p - 1$ **do**

  3:     **for** $f \in F \backslash F_k$ **do**

  4:         $F_+ \leftarrow F_k \cup \{f\}$, $\mathrm{TP(h)} \leftarrow 0$, $\mathrm{FP(h)} \leftarrow 0$

  5:         **for** $h \in H$ **do**

  6:             **for** $n \leftarrow (w + u)$ to N **do**

  7:                compute $g(n)$ from $X_n(F_+)$ (cf. Section 2.2)

  8:                $n_0 \leftarrow n - w/2 - u + 1$

  9:                **if** $g(n) > h$ **then**

10:                    **if** $n_0 \in \mathbb{N}_c$ **then**

11:                      $\mathrm{TP(h)} \leftarrow \mathrm{TP(h)} + 1$

12:                    **else**

13:                      $\mathrm{FP(h)} \leftarrow \mathrm{FP(h)} + 1$

14:                    **end if**

15:                 **else**

16:                    **if** $n_0 \in \mathbb{N}_c$ **then**

17:                      $\mathrm{FN(h)} \leftarrow \mathrm{FN(h)} + 1$

18:                    **else**

19:                      $\mathrm{TN(h)} \leftarrow \mathrm{TN(h)} + 1$

20:                    **end if**

21:                 **end if**

22:             **end for**

23:             $\mathrm{P_{FA}}(F_+, h) \leftarrow \mathrm{FP(h)}/(\mathrm{FP(h)} + \mathrm{TN(h)})$

24:             $\mathrm{P_D}(F_+, h) \;\; \leftarrow \mathrm{TP(h)}/(\mathrm{TP(h)} + \mathrm{FN(h)})$

25:             $J_3(F_+, h) \;\;\; \leftarrow \sqrt{(1 - \mathrm{P_D}(F_+, h))^2 + \mathrm{P_{FA}}(F_+, h)^2}$

26:         **end for**

27:         Select the optimal threshold:

             $h_{\min}(F_+) \leftarrow \arg \min\limits_{h \in H} J_3(F_+, h)$

28:     **end for**

29:     Select the best remaining feature:

        $f_{\min} \leftarrow \arg \min\limits_{f \in F \backslash F_k} J_3(F_k \cup \{f\}, h_{\min}(F_k \cup \{f\}))$

30:     Update $F_{k+1} \leftarrow F_k \cup \{f_{\min}\}$

31: **end for**

---

4. Application to NILM

In this section, the Hotelling T^2 detector, which showed the best perfor-
mances, is combined to the proposed FSADP and applied to real data streams in
order to find the most relevant features for the detection task. We first introduce
the power features computed using current and voltage measurements. Then we
present the publicly available BLUED dataset dedicated to NILM event-based
approaches and a new proposed dataset built using our acquisition system. Both
datasets contain several events of small or large amplitudes that are represen-
tative of a house.

4.1. Extracted features

As shown in Fig. 1, in a NILM context, current and voltage measurements
are first acquired and then features are extracted. We are interested in the
features related to the power characteristics including the active power, the re-
active power and their respective harmonics. Indeed, these features meet an
additive property [22, 42]. Thus, when a HEA is connected (resp. disconnected)
from the power system, an "additive" feature is added (resp. substracted) by an
amount equal to that produced by this HEA working individually. This prop-
erty is required for extracting HEAs' features from an aggregated signal and
for comparing them with the dataset of HEAs' signatures used by supervised
NILM classification methods. These features are computed using the mathemat-
ical formulations set out in [24] based on the latest IEEE 1459-2010 standard
for the definition of the single phase physical components under non-sinusoidal
conditions [43]:

- The kth harmonic components ($k \in [1, 15]$) of the active and reactive
  powers P_k, Q_k and their respective sums P_H, Q_H:

$$P_k[n] = v_{ak}[n]i_{ak}[n] + v_{bk}[n]i_{bk}[n] \tag{32}$$

$$Q_k[n] = v_{ak}[n]i_{bk}[n] - v_{bk}[n]i_{ak}[n] \tag{33}$$

$$P_H[n] = \sum_{k=2}^{15} P_k[n], \quad Q_H[n] = \sum_{k=2}^{15} Q_k[n] \tag{34}$$

23

- The active and reactive powers P and Q:

$$P = P_1 + P_H, \qquad Q = Q_1 + Q_H, \tag{35}$$

where v_{ak}, v_{bk}, i_{ak} and i_{bk} are the Fourier coefficients:

$$v_{ak}[n] = \frac{\sqrt{2}}{M} \sum_{m=n-M+1}^{n} v[m] \cos\left(2\pi \frac{mk}{M}\right) \tag{36}$$

$$v_{bk}[n] = \frac{\sqrt{2}}{M} \sum_{m=n-M+1}^{n} v[m] \sin\left(2\pi \frac{mk}{M}\right) \tag{37}$$

$$i_{ak}[n] = \frac{\sqrt{2}}{M} \sum_{m=n-M+1}^{n} i[m] \cos\left(2\pi \frac{mk}{M}\right) \tag{38}$$

$$i_{bk}[n] = \frac{\sqrt{2}}{M} \sum_{m=n-M+1}^{n} i[m] \sin\left(2\pi \frac{mk}{M}\right) \tag{39}$$

with $M = F_s/F_0$ is the number of samples per period and F_0 is the utility frequency.

4.2. BLUED data streams

The BLUED [26] is one of the most commonly used research dataset for event detection benchmarking [12, 19, 44, 45] due to its fine granularity of event labeling. It contains time stamped current and voltage signals of two phases A and B with ground truth events [2] obtained from a house in the United States (with a utility frequency of $F_0 = 60$ Hz) during a whole week. The sensors used for measuring the electrical current in the mains have a cutoff frequency of approximately 300 Hz. This means that from the sampled current and voltage signals, although sampled at $F_s = 12$ kHz, only the active and reactive powers up to the 5th harmonic of the current and the voltage can be computed. Due to some data download issues, only the first segments of both phases A and B of the BLUED dataset are used, which corresponds to almost 13 hours of recording. In the data segment of phase A, 461 events are involved. In phase B, there are 1132 events. Several types of HEAs in the dataset are monitored [26]: lights, laptops, air conditionner, hair dryer, etc. leading to a large diversity of HEAs' transients. A total of $p = 14$ electrical power features are extracted from

24

the current and the voltage measurements. They correspond to the active and reactive powers P and Q in addition to their harmonics P_k and Q_k up to the

410   5th order and the sum of their harmonics P_H and Q_H.

4.3. Proposed experimental data streams

A typical home electrical network was reproduced by randomly switching on and off several HEAs of different kinds: resistive loads, motors and nonlinear loads such as electronic devices. The aim is to cover an extensive range

415   of HEAs' changes. For this purpose, a test bench made of an Arduino Nano microcontroller (8-bit processor ATMega328 microcontroller at 16 MHz) and twelve 5V relay modules is used [25]. Each relay is connected to a HEA and to the microcontroller that randomly turns them on or off. The HEAs in this experiment are a microwave, a DVD player, a fan, a screen, a vacuum, a waffle

420   iron, a hair dryer, an iron, a flat iron, a mixer, a CFL and a LED lamp. For greater clarity, Fig. 8 shows the electrical assembly and instrumentation for only four HEAs. During more than 1 hour, each of these HEAs is randomly activated and shut down every 3 seconds. The current $i(t)$ and voltage $v(t)$ signals are measured using our own acquisition system that was specially designed for this

425   research work, based on an Arduino MKR Zero microcontroller (32 bit processor SAMD21 Cortex-M0 microcontroller, 48 MHz) with a sampling frequency $F_s = 6.25$ kHz [25]. A list of events was then generated by visually inspecting and hand-labeling transitions in the power signals as done for the BLUED dataset generation [26]. Timestamps of the labeled events were adjusted to

430   match the transitions. Almost 1200 events were labeled in the dataset allowing an accurate evaluation of an arbitrary event detection method.
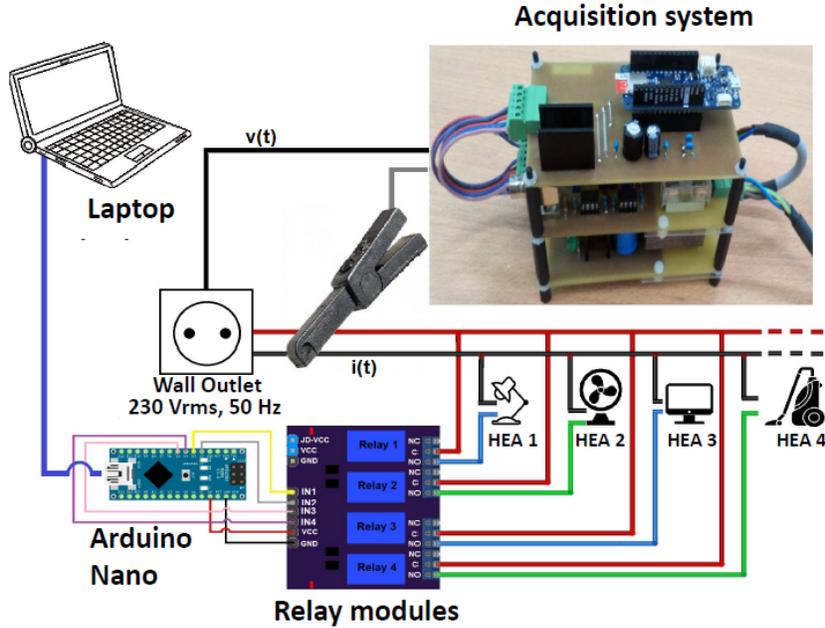
Figure 8: Electrical assembly and instrumentation of four HEAs using our acquisition system [25]

.

A total of $p = 34$ electrical power features are then extracted from the current and the voltage measurements. They correspond to the active and reactive powers in addition to their harmonics up to the 15th order and the sum of their respective harmonics. Fig. 2 depicts a part of the signals obtained for the different HEAs that are randomly switched on and off in this experiment. This data set will be freely available once this paper published.

4.4. Application of the Feature Selection Algorithm for Detection Purposes (FSADP): results and discussion

The proposed FSADP with the Hotelling T^2 detector is applied on both BLUED data streams and on our own extracted power time series. For both phases of BLUED dataset, a sliding window of $w = 8$ successive values of the power data stream is used and computed every M samples with $u = 4$. In our experimental data streams, a sliding window of $w = 8$ successive values of the

26

power data stream is also used and computed every M samples with $u = 0$. This setting is the result of a necessary tuning due to the event labelling differences between the BLUED dataset and the experimental one. Hotelling T^2 decision functions are computed for each window and compared to 500 logarithmically spaced values of the threshold $h \in [10^{-10}, 10^{10}]$. The performances obtained
with the selected features subset using the FSADP are compared to the detection performances using the active power signal only, as commonly considered for detection in NILM event-based methods [18, 21]. They are also compared to the multivariate detection performances with the P,Q feature subset and the whole features subset. Tables 1, 2 and 3 show the obtained results. Several
observations can be addressed:

Table 1: Detection performances of the Hotelling T^2 detector computed on different power features subsets for a 13 hours long excerpt of the BLUED dataset (phase A). These detection performances are J_3 (Eq. 31), the Probability of False Alarm P_{FA} (Eq. 27) and the Probability of Detection P_D (Eq. 26).

Features subset	Subset size	Detection score $\mathbf{J_3}$ (%)	P_{FA} (%)	P_D (%)
Active power P [26, 44]	1	5.78	$1.15.10^{-6}$	94.3
Active and reactive powers P,Q	2	13.70	$1.09\ 10^{-6}$	86.3
Features selected by FSADP for phase B (P_1, P, Q_H, Q_3)	4	28.30	$8.70\ 10^{-6}$	71.7
FSADP ($\mathbf{P_1}$)	**1**	**3.21**	$\mathbf{1.04\ 10^{-6}}$	**96.8**
Complete set	14	50.40	$8.20\ 10^{-7}$	49.6

27

Table 2: Detection performances of the Hotelling T^2 detector computed on different power features subsets for a 13 hours long excerpt of the BLUED dataset (phase B). These detection performances are J_3 (Eq. 31), the Probability of False Alarm P_{FA} (Eq. 27) and the Probability of Detection P_D (Eq. 26).

Features subset	Subset size	Detection score J_3 (%)	P_{FA} (%)	P_D (%)
Active power P [26]	1	16.7	0.04	83.4
Active and reactive power P,Q	2	16.2	0.03	83.8
Feature selected by FSADP for phase A (P_1)	1	16.5	0.04	83.5
FSADP ($\mathbf{P_1, P, Q_H, Q_3}$)	**4**	**15.0**	**0.03**	**85.0**
Complete set	14	52.5	0.005	47.5

Table 3: Detection performances of the Hotelling T^2 detector computed on different power features subsets for the proposed experimental dataset containing 1203 events. These detection performances are J_3 (Eq. 31), the Probability of False Alarm P_{FA} (Eq. 27) and the Probability of Detection P_D (Eq. 26).

Features subset	Subset size	Detection score J_3 (%)	P_{FA} (%)	P_D (%)
Active power P	1	12.04	5.12	89.10
Active and reactive power P,Q	2	11.96	5.08	89.17
FSADP ($\mathbf{P, P_H, Q_3, Q_9}$)	**4**	**11.25**	**4.67**	**89.78**
Complete set	34	12.86	5.86	88.55

1. When using only the active power, the proposed tuning of the Hotelling T^2 detector (minimizing the J_3 criterion) allowed to outperform the performance results obtained in [26], where authors used the same detector and found $P_D = 69.2$ % for phase B and $P_D = 94.5$ % for phase A. In addition, the results that we obtained for phase A of BLUED dataset in an univariate context are better than those in [44], for which a Probability of Detection of 90 % is reached when using an excerpt of the active power signal containing 115 events.

2. The multivariate version of the Hotelling T^2 detector with appropriate

28

features improves the detection performances compared to the univariate version considering only the commonly used active power signal for phase B of BLUED dataset and our own dataset.

3. For BLUED data streams (phase B), the FSADP allows the selection of $d = 4$ power features $\{P_1, P, Q_H, Q_3\}$ that lead to a minimization of the criterion J_3. For phase A, the first order harmonic of active power P_1 is the only feature that is selected by the FSADP. For both phases, the whole set of $p = 14$ features makes the detection performances worse. This is an argument in favor of performing feature selection for detection tasks.

4. Since the segments of signals from phase A and phase B of the BLUED dataset are extracted from the same house, only one of the subset of features selected by the FSADP should be used. It appears that the feature P_1, which is the most relevant one for detection in phase A, is the suited one. Indeed the detection performances brought by feature P_1 applied to phase B are very close to the performances obtained by the subset of features $\{P_1, P, Q_H, Q_3\}$. On the opposite the subset of features selected by the FSADP for phase B sharply decreases the detection performances when applied to phase A.

5. For our own experimental data streams, the FSADP allows to identify a small subset composed of $d = 4$ power features $\{P, P_H, Q_3, Q_9\}$ that minimizes the criterion J_3 and allows a better detection performance than the whole set of $p = 34$ features.

6. In our own experimental data streams as well as BLUED data streams, the subset of features selected by the FSADP allows to reach a P_{FA} smaller than 5% [16] and a larger P_D when compared to the other subsets.

7. There are harmonics of odd order only in the set of selected features. This is due to the fact that even order Fourier coefficients are negligible. Indeed, HEAs current and voltage waveforms have half-wave symmetry, thus even harmonics have very small magnitude compared to odd harmonics, and

29

their step heights when a change occurs are hence scarcely noticeable (see active power 4th harmonic order signals in Fig. 2).

8. The selected features clearly depend on the HEAs types that are turned on/off (resistive, electronics, with motors, etc.). In phase A of BLUED dataset, air compressor, fridge, lights and kitchen aid chopper were monitored. In phase B of BLUED dataset, a much larger number of HEAs are monitored: printer, laptop, dryer, hair dryer, iron, microwave, washing machine, lights, TV, etc. [26]. In phase B of BLUED dataset, as well as in our own experimental dataset, some of the monitored HEAs' step heights, when turned on or off, might be lost in signal noise. This explains the need of several features for the detection task improvement.

Moreover, according to Fig. 9-11:

1. The use of some power features does not improve the detection performances since for some of them, the criterion J_3 remains constant (see Fig. 11 from feature Q_H to feature P_2 and in Fig. 10 from feature P_1 to Q). This means that these features do not bring relevant information detection purposes. The same observation was made for the features selected for classification purposes in [24].

2. The use of some features increases the criterion J_3. This means that these features are not discriminant. Therefore, their use has a detrimental effect on the detector performances.
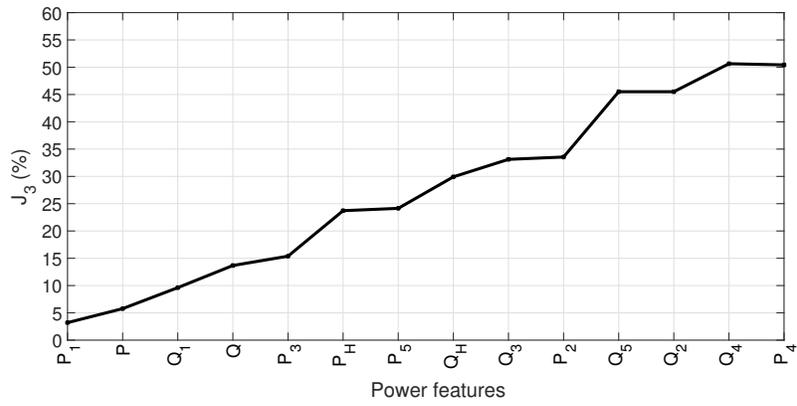
Figure 9: Power features sorted by the FSADP applied on BLUED dataset (phase A) based on the score J_3 (Eq. 31).
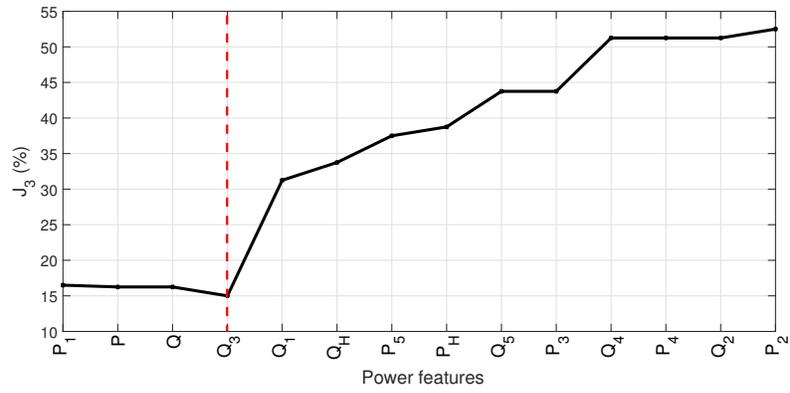

Figure 10: Power features sorted by the FSADP applied on BLUED dataset (phase B) based on the score J_3 (Eq. 31).
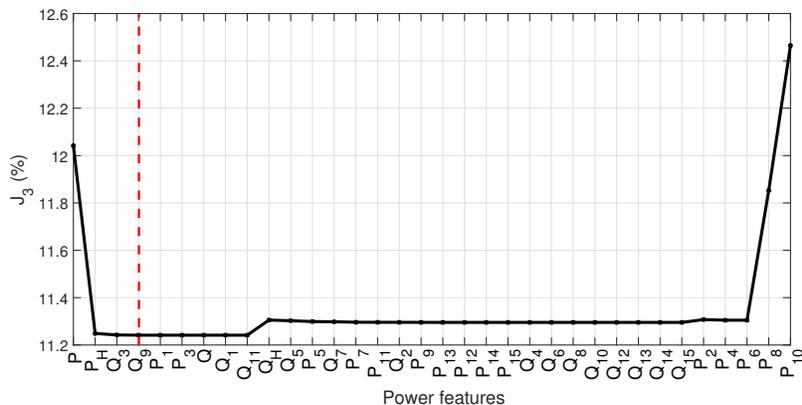
Figure 11: Power features sorted by the FSADP applied on our own dataset based on the score J_3 (Eq. 31).

5. Conclusion

This paper explores and extends to the multivariate case four methods for detecting HEAs state changes in a NILM context. Furthermore, a new labeled dataset is introduced and made available for the sake of reproducible research. We also propose a new method for properly selecting power features that improve detection performances over the existing univariate methods based on active power signal only. Our study further supports prior NILM studies that also emphasized the need of an effective set of electrical features for classification purposes. For this, the performances of four detectors in the univariate and multivariate cases are first assessed through numerical simulations and specific detection performance indicators on both synthetic and real-world data. According to our results, the Hotelling T^2 detector in its multivariate version provides the best performances. Then, the proposed FSADP method that aims at selecting the best features for detection purposes is combined with the multivariate Hotelling T^2 detector and applied to a 13 hours long excerpt of the BLUED dataset that comprises current and voltage measurements of two phases A and B. For phase A, only the feature P_1 was retained, whereas for phase B, four features were selected: $\{P_1, P, Q_H, Q_3\}$. Most of them are related to odd

32

order harmonics. The proposed FSADP method is also applied to other real-world experimental power signals containing 1203 hand-labeled events coming from several HEAs switched on and off, and extracted from current and voltage measurements acquired using our own system. Among the 34 features related to power characteristics, only a subset of four features corresponding to $\{P, P_H, Q_3, Q_9\}$ were found to be the most relevant since they improved the detector performances. The selected features clearly depend on the nature of the HEAs turned on/off. The detection performances improvement provided by the FSADP is observed by the increase of the Probability of Detection and the decrease of the False Alarm rate. This substantial improvement is essential in a NILM event-based pipeline since the success of HEAs identification depends on the detection step. Any error or missing in detection would lead to incorrect HEAs identification. For future work, other detectors in their multivariable version could be studied and assessed for HEAs purposes using the proposed method. In addition to dealing with mean changes, an interest could be given to variance changes under hypothesis H_1. Then, another extension may consist in investigating time-frequency analyzing methods using Short Time Fourier Transform (STFT) instead of using the time series signals for the event detection. For this purpose, the time signals must be transformed into the time-frequency domain. The idea would be to observe if changes in the signals have high-frequency components that can be used for event detection. Our long term perspective consists in setting up the whole block diagram shown in Fig. 1. With the detection and distinction between the switching On and Off, the HEAs can be identified and their energy consumption as well as their cost in a bill can be computed. Proposals for energy-saving can be finally made.

References

[1] A. Nilsson, M. Wester, D. Lazarevic, N. Brandt, Smart homes, home energy management systems and real-time feedback: Lessons for influencing

household energy consumption from a Swedish field study, Energy and Buildings 179 (15) (2018) 15–25. doi:10.1016/j.enbuild.2018.08.026.

[2] L. Pereira, N. Nunes, Performance evaluation in non-intrusive load monitoring: Datasets, metrics, and tools: A review, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 8 (6) (2018). doi:10.1002/widm.1265.

[3] S. Henriet, U. Simsekli, B. Fuentes, G. Richard, A generative model for non-intrusive load monitoring in commercial buildings, Energy and Buildings 177 (02 2018). doi:10.1016/j.enbuild.2018.07.060.

[4] D. Fischer, A. Hrtl, B. Wille-Haussmann, Model for electric load profiles with high time resolution for german households, Energy and Buildings 92 (02 2015). doi:10.1016/j.enbuild.2015.01.058.

[5] L. Farinaccio, R. Zmeureanu, Using pattern recognition approach to disaggregate the total electricity consumption in a house into the major end-uses, Energy and Buildings 30 (1999) 245–259. doi:10.1016/S0378-7788(99)00007-9.

[6] A. Ruano, A. Hernandez, J. Urea, M. Ruano, J. Garcia, NILM techniques for intelligent home energy management and ambient assisted living: A review, Energies 12 (11) (2019). doi:10.3390/en12112203.

[7] C. Nalmpantis, D. Vrakas, Machine learning approaches for non-intrusive load monitoring: from qualitative to quantitative comparation, Artificial Intelligence Review 52 (1) (2019) 217–243. doi:10.1007/s10462-018-9613-7.

[8] W. Dan, H. X. Li, Y. S. Ce, Review of non-intrusive load appliance monitoring, in: IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conf. (IAEAC), 2018, pp. 18–23. doi:10.1109/IAEAC.2018.8577910.

[9] R. Bonfigli, E. Principi, M. Fagiani, M. Severini, S. Squartini, F. Piazza, Non-intrusive load monitoring by using active and reactive power in addi-

34

tive factorial hidden markov models, Applied Energy 208 (2017) 1590–1607. doi:10.1016/j.apenergy.2017.08.203.

[10] K. Basu, V. Debusschere, A. Douzal, B. Seddik, Time series distance-based methods for non-intrusive load monitoring in residential buildings, Energy and Buildings 96 (03 2015). doi:10.1016/j.enbuild.2015.03.021.

[11] L. K. Norford, S. B. Leeb, Non-intrusive electrical load monitor- ing in commercial buildings based on steady-state and transient load-detection algorithms, Energy and Buildings 24 (1) (1996) 51 – 64. doi:doi.org/10.1016/0378-7788(95)00958-2.

[12] K. D. Anderson, M. E. Bergs, A. Ocneanu, D. Benitez, J. M. F. Moura, Event detection for non intrusive load monitoring, in: IECON 2012 - 38th Annual Conf. on IEEE Industrial Electronics Society, 2012, pp. 3312–3317. doi:10.1109/IECON.2012.6389367.

[13] M. Weiss, A. Helfenstein, F. Mattern, T. Staake, Leveraging smart meter data to recognize home appliances, in: IEEE Int. Conf. on Pervasive Computing and Communications, 2012, pp. 190–197. doi:10.1109/PerCom.2012.6199866.

[14] M. Nait Meziane, P. Ravier, G. Lamarque, J. Le Bunetel, Y. Raingeaud, High accuracy event detection for non-intrusive load monitoring, in: IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), 2017, pp. 2452–2456. doi:10.1109/ICASSP.2017.7952597.

[15] M. Basseville, I. V. Nikiforov, Detection of abrupt changes: theory and application, Prentice Hall information and system sciences, Prentice Hall, 1993.

[16] I. V. Nikiforov, A generalized change detection problem, IEEE Trans. Intell. Transp. Syst. 41 (1) (1995) 171–187. doi:10.1109/18.370109.

[17] M. Berges, E. Goldman, M. H Scott, L. Soibelman, Learning systems for electric consumption of buildings, in: ASCE Int. Workshop on Computing in Civil Engineering, 2009, pp. 1–10. doi:10.1061/41052(346)1.

[18] C. C. Yang, C. S. Soh, V. V. Yap, Comparative study of event detection methods for non-intrusive appliance load monitoring, Energy Procedia 61 (2014) 1840 – 1843. doi:10.1016/j.egypro.2014.12.225.

[19] L. De Baets, J. Ruyssinck, C. Develder, T. Dhaene, D. Deschrijver, On the bayesian optimization and robustness of event detection methods in NILM, Energy and Buildings 145 (2017) 57–66. doi:10.1016/j.enbuild.2017.03.061.

[20] K. Nguyen, E. Dekneuvel, B. Nicoll, O. Zammit, C. Nguyen Van, G. Jacquemod, Event detection for non intrusive load monitoring, in: Int. Workshop on Non-Intrusive Load Monitoring (NILM), 2014, pp. 3312–3317.

[21] Z. Zhu, S. Zhang, Z. Wei, B. Yin, X. Huang, A novel cusum-based approach for event detection in smart metering, IOP Conf. Series: Materials Science and Engineering 322 (7) (2018) 1–6. doi:10.1088/1757-899X/322/7/072014.

[22] J. Liang, S. K. K. Ng, G. Kendall, J. W. M. Cheng, Load signature study part I: Basic concept, structure, and methodology, IEEE Transactions on Power Delivery 25 (2) (2010) 551–560. doi:10.1109/TPWRD.2009.2033799.

[23] N. Sadeghianpourhamami, J. Ruyssinck, D. Deschrijver, T. Dhaene, C. Develder, Comprehensive feature selection for appliance classification in NILM, Energy and Buildings 151 (2017) 98–106. doi:10.1016/j.enbuild.2017.06.042.

[24] S. Houidi, F. Auger, H. Ben Attia Sethom, D. Fourer, L. Miègeville, Relevant feature selection for home appliances recognition, in: Proc. of Electrimacs Conf., 2017, pp. 1–6.

[25] S. Houidi, F. Auger, P. Frétaud, D. Fourer, L. Miègeville, H. Ben Attia Sethom, Design of an electricity consumption measurement system for non

intrusive load monitoring, in: 2019 10th International Renewable Energy Congress (IREC), 2019, pp. 1–6. doi:10.1109/IREC.2019.8754586.

[26] K. Anderson, A. Ocneanu, D. R. Carlson, A. Rowe, M. Bergés, BLUED: A fully labeled public dataset for event-based non-intrusive load monitoring research, in: Workshop on Data Mining Applications in Sustainability (SustKDD), 2012, pp. 1–5.

[27] S. Houidi, F. Auger, H. Ben Attia Sethom, L. Miègeville, D. Fourer, X. Jiang, Statistical assessment of abrupt change detectors for non-intrusive load monitoring, in: IEEE Int. Conf. on Industrial Technology (ICIT), 2018, pp. 1314–1319. doi:10.1109/ICIT.2018.8352368.

[28] H. A. D. Azzini, R. Torquato, L. C. P. da Silva, Event detection methods for nonintrusive load monitoring, in: 2014 IEEE PES General Meeting Conf. Exposition, 2014, pp. 1–5. doi:10.1109/PESGM.2014.6939797.

[29] S. M. Kay, Fundamentals of Statistical Signal Processing: Detection Theory, Vol. II, Prentice Hall, 1998.

[30] S. Houidi, F. Auger, H. Ben Attia Sethom, L. Miègeville, D. Fourer, Multivariate abrupt change detectors, Research report, Université de Nantes (UNAM) ; Université de Tunis El Manar, Tunisia ; Université Evry-Val d'Essonne/Université Paris Saclay (Aug. 2019).
URL https://hal.archives-ouvertes.fr/hal-02264149

[31] G. Qian, Y. Wu, M. Xu, Multiple change-points detection by empirical bayesian information criteria and Gibbs sampling induced stochastic search, Applied Mathematical Modelling 72 (2019) 202–216. doi:10.1016/j.apm.2019.03.012.

[32] S. M. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory, Vol. I, Prentice Hall, 1997.

[33] N. Giri, On the likelihood ratio test of a normal multivariate testing problem, Ann. Math. Statist. 35 (1) (1964) 181–189. doi:10.1214/aoms/1177703740.

[34] H. Berriri, M. W. Naouar, I. Slama-Belkhodja, Easy and fast sensor fault detection and isolation algorithm for electrical drives, IEEE Power Electron. Lett. 27 (2) (2012) 490–499. doi:10.1109/TPEL.2011.2140333.

[35] F. Mouelhi, H. Ben Attia Sethom, I. Slama-Belkhodja, L. Miègeville, P. Guérin, A fast event detection algorithm for residential loads within normal and disturbed operating conditions, in: European Journal of Electrical Engineering, Vol. 18, 2016, pp. 95–116. doi:10.3166/ejee.18.95-116.

[36] S. Atapattu, C. Tellambura, H. Jiang, Analysis of area under the ROC curve of energy detection, IEEE Wireless Commun. 9 (3) (2010) 1216–1225. doi:10.1109/TWC.2010.03.091085.

[37] L. Sevgi, Hypothesis testing and decision making: Constant-false-alarm-rate detection, IEEE Antennas Propag. Mag. 51 (3) (2009) 218–224. doi:10.1109/MAP.2009.5251252.

[38] N. J. Perkins, E. F. Schisterman, The inconsistency of optimal cut-points using two ROC based criteria, American Journal of Epidemiology 163 (7) (2006) 670–675. doi:10.1093/aje/kwj063.

[39] J. Miao, L. Niu, A survey on feature selection, Procedia Computer Science 91 (2016) 919 – 926. doi:10.1016/j.procs.2016.07.111.

[40] M. Rong, D. Gong, X. Gao, Feature selection and its use in big data: Challenges, methods, and trends, IEEE Access 7 (2019) 19709–19725. doi:10.1109/ACCESS.2019.2894366.

[41] A. Y. Kim, C. Marzban, D. B. Percival, W. Stuetzle, Using labeled data to evaluate change detectors in a multivariate streaming environment, Signal Processing 89 (12) (2009) 2529–2536. doi:10.1016/j.sigpro.2009.04.011.

[42] Z. Zheng, H. Chen, X. Luo, A supervised event-based non-intrusive load monitoring for non-linear appliances, Sustainability 10 (4) (Mar. 2018). doi:10.3390/su10041001.

[43] IEEE standard definition for the measurement of electric power quantities under sinusoidal, non sinusoidal, balanced and unbalanced conditions, revision of IEEE standard 1459-2000 (2010).

[44] M. Liu, J. Yong, X. Wang, J. Lu, A new event detection technique for residential load monitoring, in: 18th International Conference on Harmonics and Quality of Power (ICHQP), 2018, pp. 1–6. doi:10.1109/ICHQP.2018.8378820.

[45] A. Yasin, S. A. Khan, Unsupervised event detection and on-off pairing approach applied to NILM, in: 2018 International Conference on Frontiers of Information Technology (FIT), 2018, pp. 123–128. doi:10.1109/FIT.2018.00029.