

A low-level set of stationary features dedicated to non-intrusive load monitoring

Farid Dinar, Romain Chayla, Sébastien Paris, Eric Busvelle

▶ **To cite this version:**

Farid Dinar, Romain Chayla, Sébastien Paris, Eric Busvelle. A low-level set of stationary features dedicated to non-intrusive load monitoring. International Conference on Systems and Control, Nov 2022, Marseille, France. hal-03855164

HAL Id: hal-03855164
https://hal.science/hal-03855164

Submitted on 16 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A low-level set of stationary features dedicated to non-intrusive load monitoring.

Farid Dinar[1,2], Romain Chayla[1], Sébastien Paris[2] and Éric Busvelle[2,3]

***Abstract*— We study a NILM (Non-intrusive load monitoring) problem by considering the classification in two steps: a first step centered on the short duration signal of an electrical device (around one second) and a second step centered on the sequence of operating modes of this electrical device (separated by several seconds). The second step is mainly based on an RNN (recursive neural network) and will be described in future work. In this article, we study the first step which consists of a simple classification of certain electrical devices in stationary mode. We show that two parameters and some harmonics constitute a characteristic set of features and we look for a minimal set of features that preserve good results. We compare (both theoretically and practically) the meaning of odd and even harmonics. We illustrate our results using the PLAID database and some of our own electrical devices, in order to validate our real-time embedded system.**

I. INTRODUCTION

Local production and storage of "green" electrical energy are among the possible paths towards ecological transition, making it possible to partially overcome the all-centralized model favored until now. From a few homes to the scale of a building or a small village, the approach consists of pooling the production of energy consumed by the community, and optimizing its use by minimizing the share of conventional production. The advantages of local production are numerous and indisputable: empowerment of consumers, reduction of losses by Joule effect due to the transport of energy, improvement of the overall efficiency of installations, and reduction of the use of more contested production methods (see for example [5]).

However, carbon-free production tools (photovoltaic, wind, hydraulic) are intermittent and often out of sync with the real needs of consumers. Storage is expensive and inefficient and is not a solution on its own. It is therefore necessary to consume more intelligently in order to reduce the need for storage while guaranteeing an optimal quality of service. It has been shown (see for example the meta-analysis [1]) that simple personalized feedback on energy-consuming devices can help reduce energy consumption by more than 12%. It should also be noted that a more global return, with general advice, does not make it possible to achieve such levels of savings. Coupled with a local green energy production system, accompanied by personalized advice based on optimization approaches (whether conventional or from machine learning), the expected benefit will be even greater. The more precise estimation of possible savings is currently unknown, because the proposed approach is new enough that the benefits have not yet been precisely quantified. The interest of pooling information between several households, but in a limited number, makes it possible to consider more advanced optimizations, because taking into account a greater number of devices, to better manage constraints (such as consumption peaks), without, however, arriving at a situation that is too complex to manage. In addition, the community approach, when it has been tested, has often shown much better results than the approach consisting in delegating the installation of renewable energy production means to large groups. An example is the development of wind power in Denmark, which was largely initiated (under the impetus of state guarantees) by individual initiatives and cooperatives (see [5]). Wind power generation in Denmark is over 6000 MW in 2019.

Corresponding author : farid.dinar@lis-lab.fr

Partially supported by ANRT/MESRI and INDEWATT

[1] INDEWATT, Pépinière #Cleantech, Avenue Louis Philibert, 13545 Aix-en-Provence, France

[2] LIS, UMR CNRS 7020, Université de Toulon, Aix Marseille Univ., France

[2] IUT de Toulon, Electrical Engineering and Computer Science Dept, Av. de l'Université, 83130 La Garde

II. THE SUNLEAVS PROJECT

The level of technological maturity of local energy production solutions on a household scale is quite high, especially when it is based on a single means of production (generally photovoltaic or wind power). The use of a real energy mix on the scale of a household (micro-grid) is already rarer. However, the major specificity of this project is the number of consumers, between ten and fifty households. The problems introduced by this scale factor have not been studied extensively. The starting point of this work is non-intrusive measurement (NILM, non-intrusive load monitoring), the only approach fully compatible with domestic use. Since the work of George William Hart in 1992 ([9]), energy disaggregation has been the subject of numerous academic studies (see [1], [21] for example) and applications ("e.quilibre" service, with the very controversial French counter Linky). In our case, on the scale of several households, the problem becomes a semi-intrusive measurement problem, since we have several measurement points for all the electrical appliances, taken as a whole. In addition, it is well known that characterization attempts in the domestic network, non-intrusive and at high frequency, can be contaminated by electrical devices upstream of the meter, on the distribution network. We can therefore consider improving detection by using all the information from several neighbors, members of the community.

The previous work is the first step to address in order to improve power consumption. The goal is to consume as

much energy as possible locally, and therefore to minimize the use of energy distributed by large operators. Incidentally, we can resell any surplus production to them. This is a problem that has mainly been studied on the scale of a household or on the scale of national production. Having a motivated community, gathered around a local production and use project, makes it possible to envisage unprecedented optimizations. Indeed, on this scale, many uses can be deferred over time, which is not possible on smaller or larger scales (washing machines and dishwashers, recharging of electric vehicles, water balloons hot,...). An even more convincing example concerns the well-known problem of consumption peaks. The solution sometimes envisaged is to set up an RTP (Real Time Pricing) type protocol which dissuades consumers from consuming at certain times. Local optimizers that take this protocol into account inevitably lead to strong instabilities on the network and a rebound in peak consumption [3]. Distributed optimization can solve this problem at the scale of the production micro-grid. However, this promising approach requires solving several difficulties, and we can identify two main ones:

- establish an optimal strategy for triggering electrical devices. On this scale, the number of large household appliances is quite large. This confers an interest in optimization, but also certain difficulties. It is indeed a problem of optimal scheduling (triggering of the devices), under a weak maximum constraint (of production capacity, but which can be violated at times during inevitable peaks of consumption), and subject to disturbances some of which are measured and predictable (weather) and others not (sporadic use of electrical appliances, breakdowns);
- implement this strategy in the community. Here again, we want to favor a non-intrusive method as much as possible. Indeed, in addition to a financial investment for the purchase of receiver and control modules, the automatic triggering of electrical appliances remotely and without the knowledge of consumers in the home poses technical and human problems. If this approach is easily possible for a hot water tank, it is less so for noisy devices requiring human intervention.

The recent article [18] offers a survey of currently known methods in a framework quite close to what we are looking for.

In this paper, we will focus on NILM, the first part of the SUNLEAVS project. An important part of our study is the development of an electronic device which will be described below.

III. Non-intrusive load monitoring

Non-intrusive load monitoring is a well-known problem and there exists a lot of papers dealing with this subject (see [20] for recent survey paper). Several studies use available databases (Table I present some of these databases) in order to compare various approaches (see e.g. [6], [8]) but an important part of our work is the improvement of an electronic device already developed by INDEWATT. The actual version of this device is a data logger which samples the voltage and current with a sample time of one second and which store several parameters in the cloud : root mean square (RMS) voltage U_{RMS}, RMS current I_{RMS}, frequency F, active power P and reactive power Q (and therefore the complex power $S = P + i\,Q$ and phase of voltage relative to the current, $\cos\varphi = \frac{P}{|S|}$). These parameters are used as information for the user on his energy consumption. It will be later used in order to help the user to manage his energy demand according to its community.

Database	Type	sampling frequency	reference
PLAID	single	30 kHz	[7], [16]
WHITED	single	44 kHz	[12]
UK-DALE	multiple	16 kHz	[13]
REDD	multiple	15 kHz	[14]
COOLL	single	100 kHz	[19]

TABLE I
SOME LABELED DATASET FOR NILM

These five parameters are redundant, hence only two parameters give useful information. The RMS Voltage and the fundamental frequency of the voltage being regulated (and more or less constant, $U_{RMS} \approx 220\,\mathrm{V}$ and $F \approx 60\,\mathrm{Hz}$), only two parameters give independent information, I_{RMS} and $\cos\varphi$ (or other combinations among active and reactive power).

In order to distinguish several electric devices operating stationary, it is clear that these two parameters are not sufficient. Therefore, we plan to use high order harmonics of the current $I(t)$. Indeed, high order harmonics give information on the linearity of the device and this information will allow us to distinguish, for instance, a classical on/off lamp from a dimmer controlled lighting system which change the voltage waveform applied to the lamp (hence the current).

Remark 1: Our device is based upon a specific electronic component from ATMEL to measure root mean square current/voltage, phase and power, and an ESP32 in charge of basic computations and communication with the cloud via domestic WiFi. In order to estimate harmonics at sufficient order, we sample the current at a frequency $F_s = 1\,\mathrm{MHz}$ for one second, using capabilities of ESP32 and perform Fourier transform using FFT in ESP32.This allow us to estimate the spectrum of the signal up to Nyquist frequency $F_n = 500\,\mathrm{kHz}$ and therefore to estimate square current/voltage, phase, power and harmonics at a sample time of approximately $T_s = 1\,\mathrm{s}$.

At this point, it is important to distinguish two kinds of harmonic phenomenon.

- the first one is due to each device when used in continuous mode, such as a lamp, a motor, a radiator, a water heater, or a washing machine in a given mode of operation, an electric oven when it is heating, and so on. In all these cases, the process is stationary and

a harmonic analysis using the fast Fourier transform is a classical way to summarize the frequency content of the signal $I(t)$. Moreover, the fundamental frequency F and its harmonics characterize the current as a time function since $I(t)$ can be written as a Fourier series of these harmonics.

- the second one is due to the starts and stops of each device, and to the change of operating mode of some complicated devices (washing machines, thermostats, fridge or any regulated systems,...). In this case, the current is not a stationary process and Fourier analysis is no more suitable.

In the following, we will focus only on the first case and we will explain the reason of this choice in this section. Indeed, our objective is to classify each electric device accordingly to its electric consumption profile. We plan to operate at different time scales. Since our device operate with a sample time of one second, we distinguish micro behavior (at a scale where the device is in stationary mode) from macro behavior (succession of cycles). The macro behavior will be handled with a recurrent neural network. The task of this RNN will be to identify each device by learning the succession of cycles of this device. More precisely, parameters such as I_{RMS}, $\cos\varphi$ and harmonics will be the inputs of the RNN, with a sample time between 1 and 60 seconds, and the RNN will use its memory to classify each device.

Remark 2: In our approach, we separate the numerical processing into two main parts: one is devoted to the classification considering only the stationary property of the current (amplitude, phase, harmonics) and a second part devoted to the analysis of the events of the device: switching on/off a thermostat, cycles of a washing machine, etc. At this time, we are not planning to analyze transient features such as starting an engine. It has been noticed (see [6]) that the transient behavior can bring very interesting information but it supposes to sample the signal very quickly in order to capture the starting and the stopping of the processes. Our electronic device is not capable to react fast enough for this purpose. Moreover, we use the fast Fourier transform to estimate the harmonics and this spectral analysis only makes sense for stationary signals. In order to analyze the transient behavior of $I(t)$, one would have to use wavelet analysis or short-time Fourier transform (see [17] for interesting discussions on this topic, [20] for an overview and comparisons, and [10] for an implementation), but this approach requires too many operations, network transfers and cloud storage.

From now, we will therefore focus on micro-level parameters, I_{RMS}, $\cos\varphi$ and harmonics of F denoted $(c_i)_{i\in\mathbb{Z}}$, that is

$$y(t) = \sum_{j=-\infty}^{\infty} c_j \mathrm{e}^{2i\pi jFt} \tag{1}$$

(this notation is the classical notation of Fourier expansion, it is clear that in our case, $c_0 = 0$ and $c_{-i} = \bar{c}_i$)

Moreover, we will always consider our device in stationary mode, since transient mode will be handled later (and in future work) by the RNN.

IV. The PLAID database

Parameter selection is an important preliminary task before classification. In order to select parameters among RMS voltage, current and phases, including harmonics (available from c_1 to c_{9999}), we need a database of steady-state systems. We choose the well-known PLAID database ([7], [16]) among other interesting databases (see Table I) for two reasons because of its fast sampling rate, the large set of electrical devices (many of them being common with our own devices) and the open access of this database, both to use the database and to populate it.

Since it is growing thanks to user contributions, there is several versions of PLAID. We used mainly the version described in [16] containing 1876 records of single appliances (the dataset also contains records of aggregated devices, we will not use these recordings in this study).

Since we are only interested in stationary signals, we didn't use these records directly : we removed from each signal the non-stationary part. Indeed, as described in the documentation of the database, appliances activation is included in records and we don't want to use this part of the signal (although it is an important information, it will be missed by our measuring device and transient behavior will be handled with an RNN).

In order to select the stationary part of each signal from the database, we need to separate the transient and stable behavior of the appliance on each record of the database. We will only consider weak stationarity. Indeed, the transient behavior is due to the activation of the device, so it can be detected by considering only the second order moment of the signal. More precisely, we use a moving window to calculate the moving standard deviation over time (1000 sample times). We find the largest part of the signal with a constant and non-zero standard deviation (to avoid the stationary part before activation of the device). For example, the figure 1 shows the current consumption of an air conditioner at startup. At the beginning, the signal is very close to zero then increases rapidly and falls again to stabilize in a stationary mode. It is this last part (in red) that will be used to calculate the features.

We removed appliances appearing too rarely in the database. At the end of this process, our database contains 1824 records concerning 11 devices, each device being represented by a minimum of 75 records and a maximum of 246 records. Classes of appliances are air conditioner, compact fluorescent lamps, fans, fridges, hairdryers, heaters, incandescent light bulbs, laptops, microwave, vacuum, washing machines. After removing non-stationary part, the minimum size of signals is 6000 sample times.

V. Features calculation : harmonics

Only a few parameters can be measured:

- The current I_{RMS} and the voltage U_{RMS}
- The phase $\cos\varphi$ between voltage and current
- Harmonics coefficients c_j in (1)

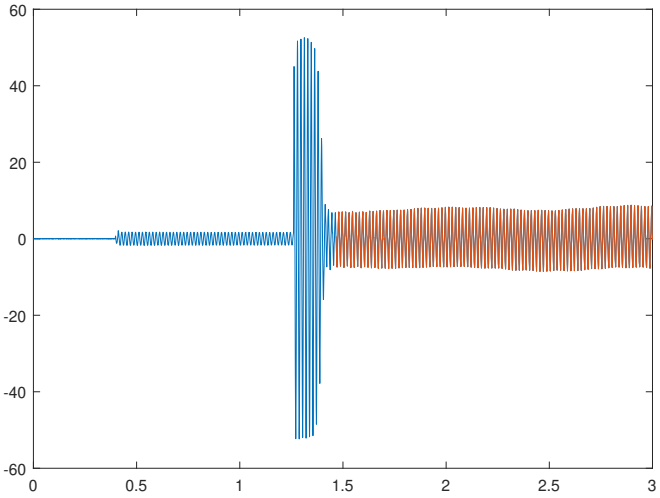


Fig. 1. Stationary part (red) in whole signal (blue)

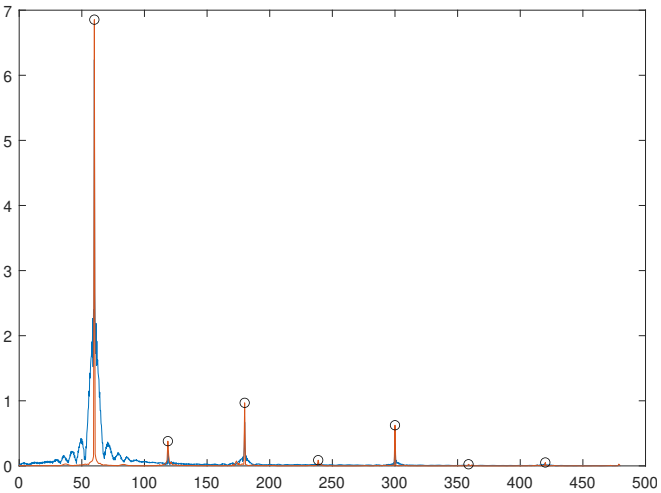


Fig. 2. FFT of stationary part (red) vs. whole signals (blue)

or some other equivalent combinations of measures such as active and reactive power, for instance. In stationary mode, these parameters are an exhaustive set of parameters since they are sufficient to reconstruct the current $I(t)$ using the inverse Fourier transform (and mathematically speaking, I_{RMS} should be redundant with the spectrum of the current signal because of the Parseval's identity).

Many authors have already studied this feature selection problem, see for instance the harmonic selection in [8], [11], [17]. However, our main concern is to validate our approach by selecting features only from the stationary mode of each device. Therefore, the conclusions of these authors apply only partially to our case. For instance, we will mainly focus on odd harmonics lower than 15. The importance of odd harmonics relatively to even harmonics has been shown in [11], [17] but without theoretical explanation. We propose below a simple calculation to explain this well-known remark.

A. The Fourier series decomposition

The Fourier series decomposition will be much more useful to us. To a periodic function $y(t)$ of period T, therefore of frequency $F = \frac{1}{T}$ defined on $\mathbb{R}$, it matches a sequence of coefficients $(c_j)_{j\in\mathbb{Z}}$ defined by

$$c_j = \frac{1}{T}\int_{\frac{-T}{2}}^{\frac{T}{2}} y(t)\mathrm{e}^{-2i\pi\, jF\, t}dt$$

The inverse passage, from the frequency domain to the time domain, is written:

$$y(t) = \sum_{j=-\infty}^{\infty} c_j \mathrm{e}^{2i\pi\, jF\, t}$$

The coefficients c_j are therefore (complex) amplitudes of periodic functions of frequency kF.

We have the famous Parseval equality (resulting from the Pythagorean theorem) which links the power of the signal to the coefficients of the Fourier series, therefore to the total spectral power:

$$\frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} |y\,(t)|^2\, dt = \sum_{j=-\infty}^{\infty} c_j^2$$

Discrete Fourier Transform: The discrete Fourier transform of a sampled signal $(y_n)_{n=0,\ldots,N-1}$ is a numerical sequence of the same length $(\widehat{y}_k)_{k=0,\ldots,N-1}$ defined by

$$\widehat{y}_k = \sum_{n=0}^{N-1} y_n \mathrm{e}^{\frac{-2i\pi k\, n}{N}}$$

and the inverse transform is given by

$$y_n = \frac{1}{N}\sum_{k=0}^{N-1} \widehat{y}_k \mathrm{e}^{\frac{2i\pi k\, n}{N}}$$

If the signal was sampled with a frequency F_e and by noting T_e the sampling period, $y_n = y(n\,T_e)$, the exponential is written $\mathrm{e}^{2i\pi k \frac{F_e}{N} nT_e}$ so coefficients $\widehat{y}_k$ correspond to frequencies that are multiples of $\frac{F_e}{N}$.

It is sometimes called FFT (Fast Fourier Transform), which is the famous algorithm used to calculate it (and therefore the name of the function in many software).

The link between the Fourier series decomposition and the discrete Fourier transform: Let's write the Fourier series decomposition of a periodic signal with period T, therefore frequency $F = \frac{1}{T}$ sampled with period T_e, therefore frequency $F_e = \frac{1}{T_e}$. It must be assumed that before sampling, an anti-aliasing filter has been used in order to cut off frequencies above $\frac{F_e}{2}$, therefore $|j|F < \frac{F_e}{2}$ We also assume, to simplify the interpretation of the calculations, that F is a multiple of $\frac{F_e}{N}$, and we note $F = \nu\frac{F_e}{N}$, so that $fracF_eF = \frac{N}{\nu}$:

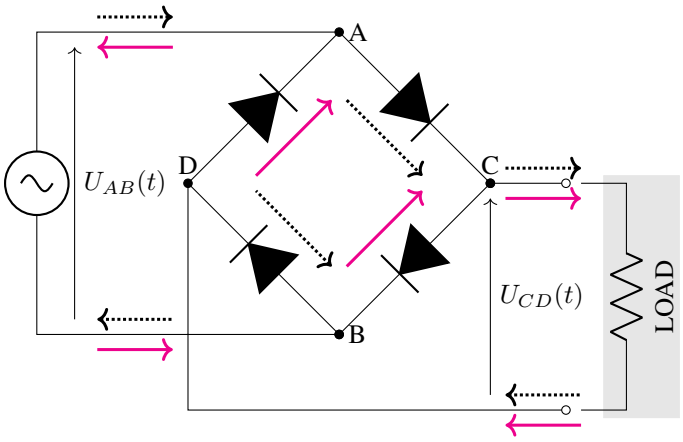


Fig. 3. Current direction in solid and dotted line (following the sign of the sinusoidal voltage) on either side of the diode bridge

$$y_n = y(n\,T_e) = \sum_{j=-\frac{1}{2}\frac{F_e}{F}}^{\frac{1}{2}\frac{F_e}{F}} c_j \mathrm{e}^{2i\pi\, j F\, n\, T_e}$$
$$= \sum_{j=-\frac{1}{2}\frac{N}{\nu}}^{\frac{1}{2}\frac{N}{\nu}} c_j \mathrm{e}^{2i\pi\, j\nu\frac{F_e}{N}\, n\, T_e} = \sum_{j=-\frac{1}{2}\frac{N}{\nu}}^{\frac{1}{2}\frac{N}{\nu}} c_j \mathrm{e}^{2i\pi\,\frac{\nu\, j\, n}{N}}$$

By comparing this writing with that of the inverse discrete Fourier transform, we see that

- if k is not a multiple of ν then $\widehat{y}_k = 0$
- if there exists j such that $k = j\nu$ then $\frac{1}{N}\widehat{y}_k = c_j$
- if $j = 1$, we see that the corresponding frequency is F and corresponds to the fundamental period of the signal $\mathrm{e}^{2i\pi\, F\, t}$. Strict multiples of ν correspond to harmonics.

A special case of a full-wave rectified signal: Most power supplies start with a rectifier diode bridge (possibly after a coil which is a pure inductor). Main voltage $U_{AB}(t)$, frequency $F \approx 60\,\mathrm{Hz}$, is applied to the diode bridge between A and B. This results in a rectified voltage $U_{DC}(t) = |U(t)|$ between C and D. This voltage generates a current $I_{DC}(t)$ which crosses the system, always in the same direction, from D to C. This current undergoes the effects of the charge, possibly non-linear, but it has a frequency $2F$ due to the rectification, i.e. $I_{DC}(t + \frac{T}{2}) = I_{DC}(t)$. The current $I_{AB}(t)$ measured on the generator side, upstream of the diode bridge, is not always in the same direction, but on each half-period $\frac{T}{2}$, it corresponds in absolute value current $I_{CD}(t)$. It follows that $I_{AB}(t + \frac{T}{2}) = -I_{AB}(t)$

Let us calculate an even harmonic :

$$\begin{aligned} c_{2k} &= \frac{1}{T}\int_0^T I(t)\mathrm{e}^{-2i\pi\times 2kFt}dt \\ &= \frac{1}{T}\left(\int_0^{\frac{T}{2}} I(t)\mathrm{e}^{-2i\pi\times 2kFt}dt + \int_{\frac{T}{2}}^{T} I(t)\mathrm{e}^{-2i\pi\times 2kFt}dt\right) \\ &= \frac{1}{T}\left(\int_0^{\frac{T}{2}} I(t)\mathrm{e}^{-2i\pi\times 2kFt}dt \right. \\ &\qquad \left. + \int_0^{\frac{T}{2}} i_m(t+\frac{T}{2})\mathrm{e}^{-2i\pi\times 2kF(t+\frac{T}{2}))}dt\right) \\ &= \frac{1}{T}\left(\int_0^{\frac{T}{2}} I(t)\mathrm{e}^{-2i\pi\times 2kFt}dt \right. \\ &\qquad \left. - \int_0^{\frac{T}{2}} I(t)\mathrm{e}^{-2i\pi\times (2kFt+k))}dt\right) = 0 \end{aligned}$$

This also explains why at least one even harmonic is a good feature for selection since it will allow us to separate devices with bridge rectifier from others. The other relevant remark is that it is not necessary to keep high order odd harmonics. This was confirmed from our study.

Harmonics are calculated for stationary part of signals, as described before. Figure 2 shows a Fourier transform of both the whole part of the current signal from Figure 1 (in blue) and the stationary part (in red). Of course, the side lobes of the sinc function due to signal discontinuities disappeared when only the stationary part is used to calculate the harmonic coefficients.

Remark 3: A first experimentation was done with the classical SCT-013 current clamp associated to a Voltage divider and a filter Capacitor in order to measure the current. We sample the signal using an USB-1208FS from Measurement Computing (capable of sampling at a frequency of 1 MHz. When we perform Fourier analysis, we denoted some energy around frequency 3 kHz. We didn't succeed to explain this phenomenon which disappear when using a more accurate current clamp (namely Fluke i30s AC/DC Current Clamp).

VI. Features selection

In order to decide what features will be used as input parameters to our RNN, we perform several tests with several classifiers (using `scikit-learn`). However, our objective is not to choose the best classifier (since this is not the final stage of our approach) but to determine a minimal set of parameters which characterize as well as possible appliances in a steady state while simplifying the next part of the classification process. In our problem, random forest (RF) classifier usually performs better than the other classifiers. Indeed, RF classifiers are known to offer small generalization error upper bounds (see [15]). Therefore, for our comparisons, we keep the RF classifier and a simple support-vector machine (SVM).

Our database has 1824 entries (11 classes of appliances) , each of them being a set of features, namely : U_{RMS}, I_{RMS}, $\cos\varphi$, active power, reactive power, apparent power and harmonic coefficients.

Test	Set of features	Score	
		SVM (%)	RForest (%)
1	I_{rms}, $\cos\varphi$	36.80±1.50	74.80±1.60
2	I_{rms}, $\cos\varphi$, $(c_{2k})_{k=0\ldots7}$	38.70±1.30	81.30±1.40
3	I_{rms}, $\cos\varphi$, $(c_{2k+1})_{k=0\ldots7}$	48.20±2.18	91.10±1.54
4	I_{rms}, $\cos\varphi$, c_1	41.67±2.50	77.06±2.34
5	I_{rms}, $\cos\varphi$, c_1, c_2	43.48±2.21	83.37±1.30
6	I_{rms}, $\cos\varphi$, c_1, c_2, c_3	50.22±3.33	90.11±1.20
7	I_{rms}, $\cos\varphi$, c_1, c_2, c_3, c_4	49.26±1.92	89.15±1.61
8	I_{rms}, $\cos\varphi$, c_1, c_2, c_3, c_4, c_5	48.60±1.55	88.90±1.58
9	I_{rms}, $\cos\varphi$, $c_1, c_2, c_3, c_4, c_5, c_6$	50.49±1.57	89.15±1.58
10	I_{rms}, $\cos\varphi$, $c_1, c_2, c_3, c_4, c_5, c_6, c_7$	50.76±1.78	89.86±1.79
11	I_{rms}, $\cos\varphi$, $(c_{2k+1})_{k=0\ldots7}, c_2$	49.61±1.97	92.41±1.91

TABLE II
CHOICE OF FEATURES

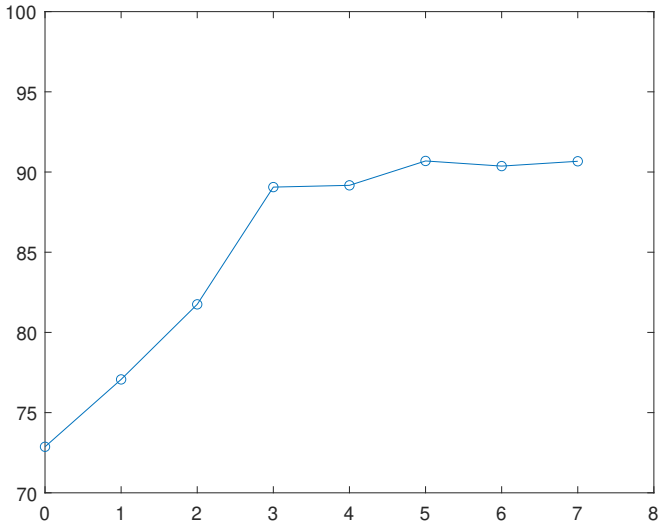


Fig. 4. Score vs. number of harmonics

In order to calculate a score, we randomly shuffle and split our database into a learning set (1459 set of features, 80% of the database) and a test set (365 elements, 20% of the database). We train the classifier on the learning set (the meta-parameters being already fixed) and estimate a score on the test set. This score is the ratio of appliances being correctly classified among the whole test set.

We extract several subset of parameters and for each of them we compare the performance of the classification. Results are presented in Table II. In order to estimate a standard deviation on the score, we repeat the learning and testing phase ten times.

Our electronic device is able to sample the voltage and current with a frequency equal to 1 Hz and give us I_{rms} and $\cos\varphi$ thanks to its polyphase energy metering integrated circuit ATMEL M90E32AS. Using only these measurements allow us a score of 73%. Adding some few harmonics allows us to increase the score to more than 90% but requires to sample the currents at a frequency equal to at least 2 kHz (if the current has a frequency equal to 60 Hz and one want to estimate harmonics at order 15 for instance). Our integrated circuit is not capable of estimating harmonics but this is easy with our ESP32 microcontroller unit which can sample the current at this frequency (up to 1 MHz using direct memory access) with 12 bits.

On Figure 4, we plot the evolution of the score vs. the number of harmonics. It appears that harmonics higher than order 5 become less useful for a good classification using only stationary features.

We have also studied odd harmonics vs. even harmonics and our results confirm some papers conclusion (see [8]). For instance, the fact that a set of odd harmonics plus the second harmonic is a good characterization of an appliance has been explained in section V and is confirmed by our experiment. The confusion matrix obtained for the best case (line 11 in Table II) is shown Table III. Confusion occurs between light bulb, air conditioner and fan, or fridge and washing machine,

APPLIANCES	air cond.	lamp	fan	fridge	hairdryer	heater	bulb	laptop	microwave	vacuum	wash. machine
air cond.	46	0	2	0	0	0	1	0	0	1	0
lamp	0	42	0	0	0	0	0	1	0	0	0
fan	3	0	44	1	0	0	2	0	0	0	0
fridge	0	0	1	12	0	0	0	0	1	0	2
hairdryer	1	1	0	0	47	1	0	0	0	0	0
heater	0	0	0	1	2	13	0	0	0	0	0
bulb	0	0	2	0	0	0	30	0	0	0	0
laptop	0	0	0	0	0	0	0	44	0	0	1
microwave	0	0	0	0	0	1	0	0	30	0	0
vacuum	0	0	0	0	0	0	0	0	1	17	0
wash. machine	0	1	0	1	0	0	0	0	0	0	12

TABLE III
CONFUSION MATRIX

for example. These confusions will be removed by the upper stage of our classifier using the non-stationary signal on the long-term behavior.

VII. CONCLUSION

This study is only a preliminary study. Our result cannot be compared with other results in the literature devoted to the PLAID database since we decided to ignore the transient behavior although it helps a lot to characterize certain devices. We explained that we will defer the use of transient behavior to the later stage of recursive neural network classification. Nevertheless, we obtained rather good results and, moreover, we were able to determine a set of good characteristics and eliminate redundant or unnecessary parameters.

REFERENCES

[1] Armel, K.C., Gupta, A., Shrimali, G. and Albert, A., 2013. Is disaggregation the holy grail of energy efficiency ? The case of electricity. Energy Policy, 52, pp.213-234.

[2] Cecati, C., Citro, C. and Siano, P., 2011. Combined operations of renewable energy systems and responsive demand in a smart grid. IEEE transactions on sustainable energy, 2(4), pp.468-476.

[3] Chang, T.H., Alizadeh, M. and Scaglione, A., 2012, July. Coordinated home energy management for real-time power balancing. In 2012 IEEE Power and Energy Society General Meeting (pp. 1-8). IEEE.

[4] Choi, S., Park, S., Kang, D.J., Han, S.J. and Kim, H.M., 2011, October. A microgrid energy management system for inducing optimal demand response. In 2011 IEEE international conference on smart grid communications (SmartGridComm) (pp. 19-24). IEEE.

[5] Dessus, B., Hélène Gassin, 2004. So watt ? L'énergie : une affaire de citoyens. Editions de l'Aube.

[6] Drouaz, M., Colicchio, B., Moukadem, A., Dieterlen, A. and Ould-Abdeslam, D., 2021. New time-frequency transient features for non-intrusive load monitoring. Energies, 14(5), p.1437.

[7] Gao, J., Giri, S., Kara, E.C. and Bergés, M., 2014, November. Plaid: a public dataset of high-resoultion electrical appliance measurements for load identification research: demo abstract. In proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings (pp. 198-199).

[8] Hacine-Gharbi, A., Ravier, P. and Nait-Meziane, M., 2020. Relevant harmonics selection based on mutual information for electrical appliances identification. International Journal of Computer Applications in Technology, 62(2), pp.102-115.

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

[10] Hoyo-Montaño, J.A., León-Ortega, N., Valencia-Palomo, G., Galaz-Bustamante, R.A., Espejel-Blanco, D.F. and Vázquez-Palma, M.G., 2018. Non-intrusive electric load identification using wavelet transform. Ingeniería e Investigación, 38(2), pp.42-51.

[11] Kang, H. and Kim, H., 2020. Household appliance classification using lower odd-numbered harmonics and the bagging decision tree. IEEE Access, 8, pp.55937-55952.

[12] Kahl, M., Haq, A.U., Kriechbaumer, T. and Jacobsen, H.A., 2016, May. Whited-a worldwide household and industry transient energy data set. In 3rd International Workshop on Non-Intrusive Load Monitoring (pp. 1-4).

[13] Kelly, J. and Knottenbelt, W., 2015. The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes. Scientific data, 2(1), pp.1-14.

[14] Kolter, J.Z. and Johnson, M.J., 2011, August. REDD: A public data set for energy disaggregation research. In Workshop on data mining applications in sustainability (SIGKDD), San Diego, CA (Vol. 25, No. Citeseer, pp. 59-62).

[15] Lyu, Shen-Huan and Yang, Liang and Zhou, Zhi-Hua, 2019, A Refined Margin Distribution Analysis for Forest Representation Learning, Proceedings of the 33rd International Conference on Neural Information Processing Systems.

[16] Medico, R., De Baets, L., Gao, J., Giri, S., Kara, E., Dhaene, T., Develder, C., Bergés, M. and Deschrijver, D., 2020. A voltage and current measurement dataset for plug load appliance identification in households. Scientific data, 7(1), pp.1-10.

[17] Nait-Meziane, M., Hacine-Gharbi, A., Ravier, P., Lamarque, G., Le Bunetel, J.C. and Raingeaud, Y., 2016, February. Hmm-based transient and steady-state current signals modeling for electrical appliances identification. In Proceedings of the 5th International Conference on Pattern Recognition Applications and Methods (pp. 670-677).

[18] Pallonetto, F., De Rosa, M., D'Ettorre, F. and Finn, D.P., 2020. On the assessment and control optimisation of demand response programs in residential buildings. Renewable and Sustainable Energy Reviews, 127, p.109861.

[19] Picon, T., Meziane, M.N., Ravier, P., Lamarque, G., Novello, C., Bunetel, J.C.L. and Raingeaud, Y., 2016. COOLL: Controlled on/off loads library, a public dataset of high-sampled electrical signals for appliance identification. arXiv preprint arXiv:1611.05803.

[20] Sadeghianpourhamami, N., Ruyssinck, J., Deschrijver, D., Dhaene, T. and Develder, C., 2017. Comprehensive feature selection for appliance classification in NILM. Energy and Buildings, 151, pp.98-106.

[21] Salem, H., Sayed-Mouchaweh, M. and Tagina, M., 2020. A Review on Non-intrusive Load Monitoring Approaches Based on Machine Learning. In Artificial Intelligence Techniques for a Scalable Energy Transition (pp. 109-131). Springer, Cham.