HELD1: Home Equipment Laboratory Dataset for Non-Intrusive Load Monitoring

Pirmin Held, Steffen Mauch, Alaa Saleh, Dirk Benyoucef
Furtwangen University, Germany
Mechanical and Medical Engineering
Email: {pirmin.held,steffen.mauch,
alaa.saleh,dirk.benyoucef}@hs-furtwangen.de

Djaffar Ould Abdeslam
Université de Haute-Alsace, France
IUT de Mulhouse
Email: djaffar.ould-abdeslam@uha.fr

***Abstract*—Non-Intrusive Load Monitoring (NILM) can be split into event detection, classification and energy tracking. Different algorithms have already been proposed for the respective tasks. Each algorithm has been verified based on publicly available data sets to assess its performance. The two types of data sets that currently exist can be distinguished into two types: laboratory measurements and data sets from real world environment. In general, the available laboratory measurements provide data of individual devices; these are only of limited use for overall benchmark tests. Measurements, in which several devices have been active simultaneously, only exist in real scenario datasets. Nevertheless, the assignment of reference data in real scenarios is somehow problematic: issues are, for example, the synchronization between reference data and measured data, absence or excess of events and the number of on and off cycles of each device respectively. Furthermore, the probability distribution of the devices, as well as long measurement cycles with correspondingly large amounts of data, but low number of events, are challenging. Therefore, it is very difficult to compare the current NILM algorithms. Home Equipment Laboratory Dataset (HELD1) has multiple switching on and off events of several devices acting individually and/or simultaneously. Since the individual devices can be controlled separately, the reference data is available in a very high quality. Thus, high number of events can be generated within a short measuring time. In addition, the dataset contains different complex scenarios of various numbers of appliances. The objective of this data set is to offer a better basis to enhance the comparability between the individual NILM approaches.**

Keywords–NILM dataset; feature extraction; feed forward neural net; supervised classification.

I. INTRODUCTION

In the past years, Non-Intrusive Load Monitoring (NILM) has gained more and more attention. The initial idea of Non-Intrusive Load Monitoring (NILM) has been to determine the consumption of individual devices. The visualization of the amount of power that is consumed by individual devices [1] should raise the consumers consciousness. NILM in the Ambient Assisted Living (AAL) context focuses on analyzing the behavior of persons in need of care to detect critical situations [2]. NILM can be split into the following aspects: event detection, classification, and tracking. Depending on the application, the complexity of the task is quite different. In the AAL context, a classification of the devices is necessary. To measure the power consumption of each single device, which was the primary idea of NILM, it is necessary to track the power consumption of the individual devices additionally.

Thus, the requirements for data sets vary from task to task. Apart from fundamental aspects, such as resolution, sampling rate, and waveform capturing, the reference data are of utmost importance. The reference data are needed for a successful training and the evaluation of the algorithms results. It is extremely important that the reference data are reliable, otherwise the training of the algorithms will be incorrect. For example a supervised classifier needs information about the device and the position of the event. During the evaluation of the tests, the results will be compared with the reference data. Inconsistent reference data leads to an unreliable assessment of the tested algorithms.

Most of the public data sets, e.g., Blued [3], Redd [4] or UK-Dale [5] have been recorded in real scenarios. In Table I several public data sets have been compared. The advantage of such an approach is that the different modalities for the use of appliances are realistic. However, the use of real scenarios has drawbacks, too: extremely long periods of time have to be measured in order to obtain sufficient switching-on and switching-off cycles for each consumer. And, if the measuring period is to short, the risk persists that some consumers have not been switched on or off. Such a situation is particularly problematic with respect to classification. In the case of measuring over a very long period like months or years, the danger, that errors will occur during the generation of reference data rises. Having measurement campaigns where data have been sampled at a high rate, the amount of data being generated over a period of years is enormous. Additional difficulties in the generation of reference data, rises from complex devices such as washing machines or dishwashers. It is difficult to map the reference data of the internal consumers of complex devices. Therefore, in general, events of these devices are not marked in the data sets.

Other data sets, such as PLAID [6], WHITED [7] and COOLL [8]) have been recorded under laboratory conditions. Here, individual loads are measured independent of each other. In these data sets, there are no scenarios prevalent where several devices are operated simultaneously. Nevertheless, the advantage of these data sets is that many on / off cycles are available in a short period of time. These data sets enable a good examination of device-specific properties, especially since a very high sampling rate of up to 100 kHz has been applied. However, these data sets are far removed from real conditions, since only one load was activated at the same time. Due to fluctuations in the grid and also phase shifts, it is not possible to simply combine the individual measurements by superpositioning. Thus, there is no straight forward method to simulate multiple consumers being active at the same time.

Due to the afore mentioned aspects, our goal has been to generate a new data set based on real measurements from real consumers, which are switched on and off randomly by a personal computer. In the proposed experimental setup, up to ten different consumers can be active at the same time. The advantage of this laboratory setup is that the appliances are switched on and off at a higher frequency than in an ordinary utilization. Additionally, it is possible to adjust the frequency distribution of the switch-on and switch-off cycles of the individual consumers individually. Combinations of devices with different power consumption and scenarios, combined with various numbers of appliances being active at the same time, were recorded.

TABLE I. COMPARISON OF NILM DATASETS

dataset	samplerate aggregate measurement	samplerate ground truth	duration	single/ multiple devices/ laboratory data set
REDD [4]	1sec / 15kHz	3sec	3 to 19 days	-/yes/no
BLUED [3]	12kHz	event based	8 days	-/yes/no
PLAID [6]	30kHz	-	5sec	yes/no/yes
WHITED [7]	44kHz	-	5sec	yes/no/yes
COOLL [8]	100kHz	-	6sec	yes/no/yes
UK-DALE [5]	16kHz	6sec	3-17 months	-/yes/no
ECO [9]	1sec	1sec	8 months	-/yes/no
SustData [10]	50Hz	50Hz	5 years	-/yes/no
HELD1	4kHz	event based	scenario	yes/yes/yes

Since the new data set HELD1 has been recorded under laboratory conditions, the reference data are available in very high quality. Current and voltage data are available, along with reference data consisting of on and off events of the corresponding device ID. The data set is hosted public and free to use for anyone; it is hosted at [11].

The paper is structured as follows: The dataset including the measurement system and scenarios is described in Sec. II. Feature extraction is presented in Sec. III followed by the classification in Sec. IV, which introduces a feedforward neural net. Sec. V presents the results followed by the conclusion in Sec. VI.

II. HELD1 Dataset

This section describes the measurement system, and the different measurement scenarios of the HELD1 dataset. In addition, the power distribution of the existing consumers in the record is presented.

A. Measurement system

The measuring system, see Fig. 1, has already been described in detail in [12]. The sampling rate for current and voltage is 4 kHz. To avoid aliasing, a low-pass filter with a cutoff frequency of 1.3 kHz is applied before sampling. The analog-digital converter operates with 16 bits which corresponds to a theoretical resolution of $\sqrt{2} \cdot 63\,\mathrm{A}/2^{16} = 0.961\,\mathrm{mA}$ being equal to $0.61\,\mathrm{W}$. In reality, current and voltage measurements have a noise of $\approx 16\,\mathrm{mA}$ and $\approx \pm 2\,\mathrm{V}$ when the inputs shorted to ground. This corresponds to $\pm 8\,\mathrm{mA} \cdot 230\,\mathrm{V} = \pm 1.84\,\mathrm{W}$.

In the HELD1 data set, up to 18 different consumers are used. Their characteristics are given in Table II. The appliances are selected simple consumers, which do not have any major internal load switching.

Some appliances did not work reliably during the measurement campaign, thus, their measurements were excluded from the dataset. Therefore, the numbering of the devices in Table II is not continuous.

Some of the used devices are shown in Fig. 2.

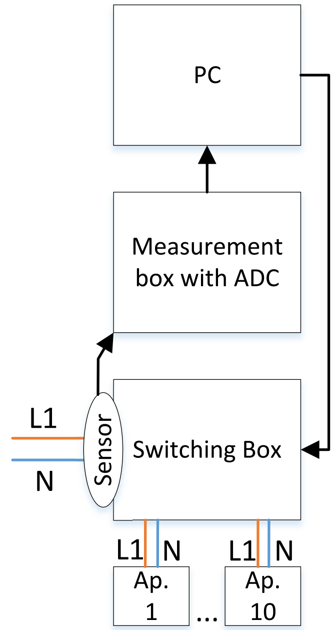

Figure 1. Block diagram of the measurement system; ADC means analog digital converter, Ap. appliance, L1 phase 1, N neutral line, PC personal computer

TABLE II. DEVICE LIST

ID	Name	P(W)
001	Toaster	998
002	Hairdryer (setting 2)	1155
003	Radio	6.2
005	Vacuum Cleaner (red)	424
007	Hair Straightener	56
009	Heat Gun (setting 1)	820
010	Router	9.2
011	Desk Lamp	20
013	Refrigerator (white)	170
014	Refrigerator (blue)	190
015	Fluorescent Lamp	40
016	Light Bulb Box	20
017	Kettle	2100
019	Hairdryer (setting 1)	500
020	Heat Gun (setting 2)	1603
021	Fan	22
022	Multifunction Tool (Dremel ®)	30
023	LED lamp	1

B. Measurement scenarios

The data set is divided into four consecutive measurements:

- Training data (consisting of individual measurements, 100 on / off events per device)
- Test measurement one (max. one device being active at the same time)
- Test measurement two (up to four active devices being active at the same time)
- Test measurement three (up to six devices being active at the same time)

In test measurements, each device was turned on and off in total 20 times. Additionally during the test measurements, the devices were randomly switched on and off under consideration of the corresponding scenario. For all scenarios, the minimum time distance between two events is three seconds.

Copyright (c) IARIA, 2018. ISBN: 978-1-61208-638-5

Figure 2. Illustration of random selected individual appliances with corresponding device ID

C. Power distribution of the appliances

In Fig. 3, different power distributions of the consumers are shown. A few individual consumers (e.g., refrigerator or kettle) have two different clusters of power distribution. The reason can be found in a different power consumption between the switching on and the switching off steady state in these appliances. Fig. 4 shows a zoomed extract of Fig. 3 in the lower power range, due to better visalization. Figures 3 and 4 result from the training data record, where only one device is active at the same time. If several devices are active simultaneously, the overall noise level is increased. This results in larger clusters of the individual appliances which leads to a higher probability of overlapping clusters in test dataset two and three.

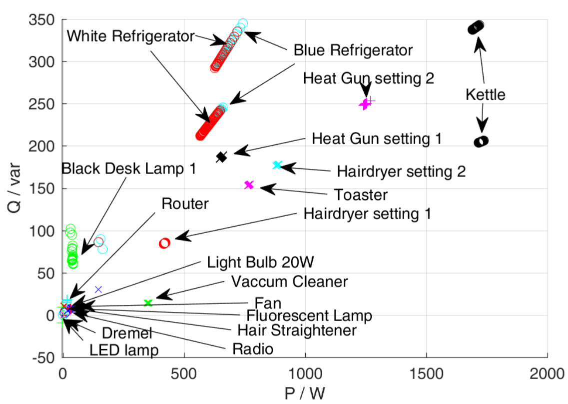

Figure 3. Power distribution of the individual consumers

Most devices can already be separated by P and Q (see Fig. 3 and 4). However, some clusters show overlapping data, e.g., the clusters of the hair straightener, multifunction tool (Dremel®, radio, LED lamp or the two refrigerators. The idea is that the distinctiveness of these devices can be improved by adding the harmonics to the feature space of P and Q.

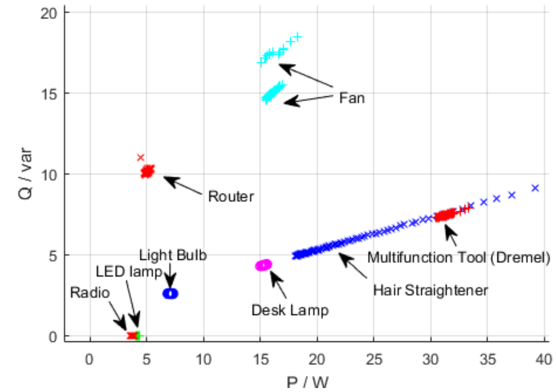

Figure 4. Detailed illustration of the active and reactive power of the different devices in the lower power range shown in Fig. 3

III. Feature Extraction

Before performing a classification, it is necessary to extract the different features out of the given data. Most frequently, active power P and the reactive power Q has been used as features in context of NILM. Harmonics are used as additional features to reduce the probability of overlapping features. Moreover, the use of harmonics benefits from the high frequency sampling. The harmonics $H \in \mathbb{R}^7$ are calculated for each device D as shown in Equ. (1).

$$H_D(\omega) = \int_{-\infty}^{+\infty} s(t)e^{-j\omega t} \;\Big|\; \omega = 50, 100, \ldots, 350\,\text{Hz} \quad (1)$$

Since more than one load can be switched on at the same time, the difference between the signal (P, Q, H) before and after the event is calculated. This principle is visualized in Fig. 5 with a real power signal. Two windows including 4000 sampling points are used. The 8000 sampling points before and after the event are not considered. The switching on and off events are handled as two independent features. The result is a doubling of the clusters in the feature space. In general, on and off events usually differ only in the sign. This utilization increases the feature space but does not result in a disadvantage

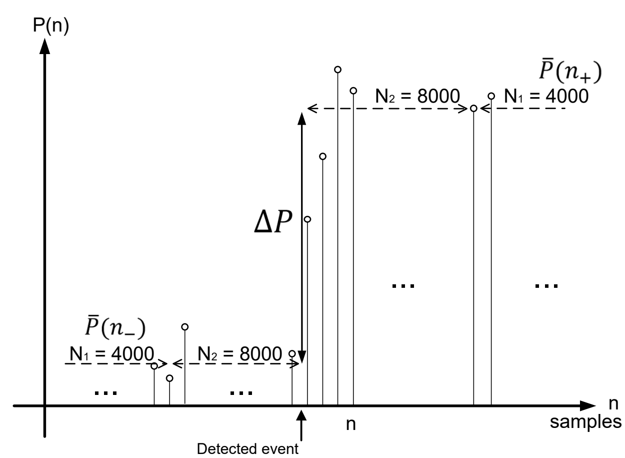

Figure 5. Representation of the feature extraction based on the real power

Copyright (c) IARIA, 2018. ISBN: 978-1-61208-638-5

during the classification, because switching on and off events can easily be distinguished.

$$P(n_-) = \frac{1}{N_1} \sum_{l=1}^{N_1} \left(P\left(n - l - N_2\right) \right) \tag{2}$$

$$P(n_+) = \frac{1}{N_1} \sum_{l=1}^{N_1} \left(P\left(n + l + N_2\right) \right) \tag{3}$$

$$\Delta P = P(n_+) - P(n_-) \tag{4}$$

The calculation of Q and H is analogous to the calculation of P which is given in Equ. (2), (3) and (4). The normalization of the signals is as follows:

$$\tilde{P}_D = \Delta P_D - \min(\Delta P_D) \tag{5}$$

$$\tilde{Q}_D = \Delta Q_D - \min(\Delta Q_D) \tag{6}$$

$$\tilde{H}_D(\omega) = \Delta H_D - \min(\Delta H_D(\omega)) \tag{7}$$

$$\Big| \quad \omega = 50, 100, ..., 350$$

The individual waveforms $\tilde{P}$, $\tilde{Q}$, $\tilde{S}$ are combined into the matrix

$$X = \begin{pmatrix} \tilde{P}_1 & \tilde{Q}_1 & \tilde{S}_1 \\ \vdots & \vdots & \vdots \\ \tilde{P}_{18} & \tilde{Q}_{18} & \tilde{S}_{18} \end{pmatrix}. \tag{8}$$

The last step in the normalization procedure is as follows:

$$Y = \begin{pmatrix} \frac{1}{\max(X(\cdot,1))} & 0 & \dots & 0 \\ \vdots & \ddots & & \vdots \\ 0 & 0 & \dots & \frac{1}{\max(X(\cdot,9))} \end{pmatrix} \tag{9}$$

$$Z = X \cdot Y \tag{10}$$

Every column m of X is divided by the maximum of the m-th column of X (Equ. (9) and (10)). $X(\cdot, m)$ means that the m-th column of the matrix X is used. By multiplication of X and Y, the matrix Z is calculated.

IV. Classification

In order to generate a reference value for the classification of HELD1, the individual measurements were used as training values. The features have been chosen as described in Sec. III. As classification method, a feedforward net, as shown in Fig. 6, is applied. During the training phase, the values for weights w and bias b have been determined. For the training, scaled conjugate gradient back-propagation [13] has been chosen.

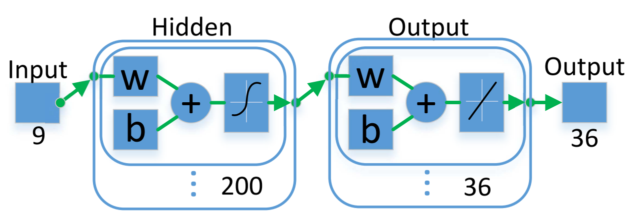

Figure 6. Overview of the applied feedforward neural network, whereas weights are depicted with w and bias with b.

ΔP, ΔQ and the first seven harmonics are used as input of the neural network. Thus, the total input is nine dimensions. The number of consumers to be distinguished defines the corresponding number of dimensions of the output. Additionally, as on and off events are treated as separate events, the number of outputs is doubled. In order to find the required size of the neural network, the number of hidden layers has been varied between one and 500. The accuracy of the achieved result is visualized in Fig. 7. The results were obtained by application of the test measurements ‘0003’, ‘0116’ and ‘0201’. In order to determine the lower, upper, and average values of the results for the corresponding number of hidden layers, each neural net has been trained in total 100 times.

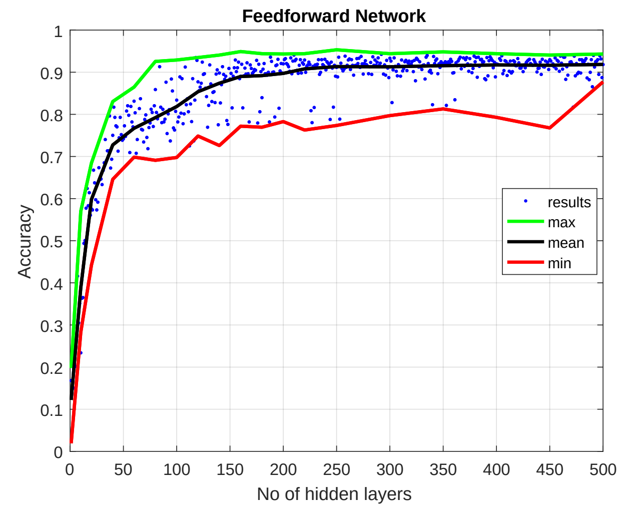

Figure 7. Result of the classification according to different numbers of hidden layers

An accuracy of $\approx 90\,\%$ can be achieved with about 200 hidden layers (Fig. 7). A further enlargement of the net has no positive effect on the achieved accuracy and the result of classification. The variance of results can be explained by the random initialization of each neural net for every new initiated training. The optimization problem is not convex, therefore local minima prevail.

V. Results

The individual classification result of each measurement file is listed in Table III to VII. Table III shows the results of test measurement two, where only one device was active at one time. The overall recognition rate is $\approx 93.15\,\%$. In order to identify the occurring errors, the results of measurement ‘0076’ are exemplified in Table IV. Target values are plotted in the columns, whereas the actual values are given in corresponding row. Each device in this measurement had in total 40 events. For ID nine, there are only 39 events, since the last event was immediately before the end of the recorded measurement; thus no feature could be calculated. In this measurement, six different consumers were used. The devices with ID 3, 11 and 21 were detected incorrectly. These devices were not included in the measurement file ‘0076’ but the neural net had been trained with all available devices.

Since both refrigerators (ID 13 and 14) are showing almost similar consumption characteristics, these devices have been randomly mixed, which leads to an increase of wrong

TABLE III. CLASSIFICATION PERFORMANCE OF TEST MEASUREMENT ONE

Measurement File	No. of Events	Accuracy
0076	240	88.285%
0077	240	87.342%
0078	240	86.611%
0079	240	89.121%
0080	240	87.029%
0117	400	91.729%
0118	400	90.727%
0119	400	88.446%
0120	400	88.722%
0202	400	98.496%
0203	400	97.744%
0204	400	99.499%
0205	400	99.749%

TABLE IV. CONVERSION MATRIX OF THE CLASSIFICATION RESULTS OF MEASUREMENT ‘0076’

Dev. ID (actual / set)	3	5	9	11	13	14	16	17	19	21
3	0	0	0	0	0	0	0	0	0	0
5	0	39	0	0	0	0	1	0	0	0
9	0	0	39	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0
13	0	0	0	1	33	3	0	0	0	3
14	0	0	0	0	17	23	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	39	0	1
19	1	0	0	0	0	0	0	0	38	1
21	0	0	0	0	0	0	0	0	0	0

assignments. All other devices have been identified correctly with an accuracy of $\approx 97\,\%$ in measurement ‘0076’.

TABLE V. CLASSIFICATION PERFORMANCE OF TEST MEASUREMENT TWO

Measurement File	No. of Events	Accuracy
0005	400	57.789%
0008	400	52.393%
0009	400	57.868%
0011	400	55.138%
0171	400	59.25%
0172	400	62.907%
0173	400	56.391%
0175	400	59.649%
0192	400	59.649%
0193	400	61.905%
0194	400	66.165%
0195	400	64.16%
0206	400	73.434%
0207	400	68.672%
0208	400	45.614%
0209	400	67.92%
0210	400	66.667%

Table V contains scenario test measurement two with up to four devices active at the same time. The total recognition rate is $60.92\,\%$. Table VI shows the results of measurement ‘0194’ exemplarily. Again, the two refrigerators are the major source of error. The detection rate for the nine remaining devices in measurement ‘0194’ is $72\,\%$. As several devices are operated simultaneously, the probability of mis-detection increased, especially for small consumers.

Table VII presents the results of scenario test measurement three with up to six devices active at the same time. The overall recognition rate reveals $58.85\,\%$.

TABLE VI. CONVERSION MATRIX OF THE CLASSIFICATION RESULTS OF MEASUREMENT ‘0194’

Dev. ID (actual / set)	3	5	7	9	10	11	13	14	15	16	17	19	21	23
3	20	0	1	0	0	1	1	0	0	0	0	0	3	14
5	0	36	0	0	0	0	3	1	0	0	0	0	0	0
7	0	0	33	0	0	5	1	0	0	0	0	0	1	0
9	2	0	0	29	0	0	1	0	0	4	0	0	4	0
10	1	0	0	0	25	2	7	0	0	1	0	0	0	4
11	1	0	2	0	0	28	1	0	0	4	0	0	1	3
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	6	22	6	1	0	0	0	4	1
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	1	1	0	0	29	0	0	0	9
17	0	0	1	0	0	4	1	0	0	0	32	0	2	0
19	3	0	0	0	0	0	0	0	0	4	0	26	6	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0

TABLE VII. CLASSIFICATION PERFORMANCE OF TEST MEASUREMENT THREE

Measurement File	No. of Events	Accuracy
0241	400	58.897%
0242	400	58.145%
0243	400	61.654%
0244	400	58.647%
0245	400	56.892%

VI. CONCLUSION

This paper presents HELD1 as a data set for NILM which has been recorded under laboratory conditions. The primary advantage of this data set is the reliable reference data. Further advantages are the identical probability distribution for all consumers as well as a high density of events within the recorded data. This reduces the size of measurement data and the calculation effort. To the knowledge of the authors, this is the first data set being published in context of NILM, that provides measurements under laboratory conditions with several consumers being active at the same time. Therefore, the dataset reflects realistic scenarios. For supervised learning, individual measurements with 100 switch-on and switch-off cycles are available. A classification procedure using neural networks for the allocation of first reference values is presented. In the future, the data set will be supplemented with further measurements to offer even more different scenarios.

ACKNOWLEDGMENT

We would like to thank Steven Kyle Johnson and Fabian Suske for carrying out the measurements. This work was created as part of the iMon project (funding number 03FH001IX4).

REFERENCES

[1] G. Hart, “Residential energy monitoring and computerized surveillance via utility power flows,” vol. 8, no. 2, pp. 12–16.

[2] P. Klein, “Non-intrusive information sources for activity analysis in ambient assisted living scenarios.” [Online]. Available: https://tel.archives-ouvertes.fr/tel-01526695/

[3] K. Anderson, A. Ocneanu, D. Benitez, A. Rowe, and M. Berges, “BLUED: A fully labeled public dataset for event-based non-intrusive load monitoring research,” in Proceedings of the 2nd KDD Workshop on Data Mining Applications in Sustainability (SustKDD). [Online]. Available: http://inferlab.org/wp-content/uploads/2012/08/2012_anderson_SustKDD.pdf

[4] J. Z. Kolter and M. J. Johnson, “REDD: A public data set for energy disaggregation research,” in Workshop on Data Mining Applications in Sustainability (SIGKDD), San Diego, CA, vol. 25. Citeseer, pp. 59–62, 00406. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.454.5796&rep=rep1&type=pdf

Copyright (c) IARIA, 2018. ISBN: 978-1-61208-638-5

[5] J. Kelly and W. Knottenbelt, “The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes,” vol. 2, p. 150007, 00034.

[6] J. Gao, S. Giri, E. C. Kara, and M. Bergés, “PLAID: a public dataset of high-resoultion electrical appliance measurements for load identification research: demo abstract.” ACM Press, pp. 198–199, 00002.

[7] M. Kahl, A. U. Haq, T. Kriechbaumer, and H.-A. Jacobsen, “WHITED-a worldwide household and industry transient energy data set,” 00000. [Online]. Available: http://nilmworkshop.org/2016/proceedings/Poster_ID18.pdf

[8] T. Picon, M. N. Meziane, P. Ravier, G. Lamarque, C. Novello, J.-C. L. Bunetel, and Y. Raingeaud, “COOLL: Controlled on/off loads library, a public dataset of high-sampled electrical signals for appliance identification,” 00001. [Online]. Available: https://arxiv.org/abs/1611.05803

[9] C. Beckel, W. Kleiminger, R. Cicchetti, T. Staake, and S. Santini, “The ECO data set and the performance of non-intrusive load monitoring algorithms.” ACM Press, pp. 80–89, 00049. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2674061.2674064

[10] L. Pereira, F. Quintal, R. Gonçalves, and N. J. Nunes, “SustData: A public dataset for ICT4s electric energy research.” in ICT4S, 00012. [Online]. Available: http://www.atlantis-press.com/php/download_paper.php?id=13462

[11] P. Held, S. Mauch, A. Saleh, D. Ould Abdeslam, and D. Benyoucef, “Home equipment laboratory dataset (HELD1).” [Online]. Available: http://141.28.27.61/

[12] T. Bier, D. Benyoucef, D. Abdeslam Ould, J. Merckle, and P. Klein, “Smart meter systems measurements for the verification of the detection & classification algorithms,” 00003.

[13] F. M. Møller, “A scaled conjugate gradient algorithm for fast supervised learning,” 00000.