

RESEARCH

Open Access



Robust energy disaggregation using appliance-specific temporal contextual information

Pascal Alexander Schirmer^{1*} , Iosif Mporas¹ and Akbar Sheikh-Akbari²

Abstract

An extension of the baseline non-intrusive load monitoring approach for energy disaggregation using temporal contextual information is presented in this paper. In detail, the proposed approach uses a two-stage disaggregation methodology with appliance-specific temporal contextual information in order to capture time-varying power consumption patterns in low-frequency datasets. The proposed methodology was evaluated using datasets of different sampling frequency, number and type of appliances. When employing appliance-specific temporal contextual information, an improvement of 1.5% up to 7.3% was observed. With the two-stage disaggregation architecture and using appliance-specific temporal contextual information, the overall energy disaggregation accuracy was further improved across all evaluated datasets with the maximum observed improvement, in terms of absolute increase of accuracy, being equal to 6.8%, thus resulting in a maximum total energy disaggregation accuracy improvement equal to 10.0%.

Keywords: Non-intrusive, Load monitoring, Energy disaggregation, Contextual temporal information, Two-stage energy disaggregation

1 Introduction

In the last decades, rising energy consumption needs within residential and industrial environments have become a crucial issue with nowadays consumer households accounting for approximately 40% of the total worldwide consumed energy [1, 2]. With the development of information and communication technologies (ICT) and the increasing usage of electrical appliances and automation of tasks, the electric power needs will grow further and the number of electrical appliances per household will significantly increase within the next 20 years [1, 2]. Despite the expected increase in total energy consumption, studies estimate that 20% of households' consumed energy could be saved by changing consumers' behaviour and improving the existing poor operational strategies [3, 4]. Furthermore, the establishment of smart grids and demand management as well as the fluctuation of power generation due to an increasing percentage of renewable energies are enhancing the issue of

increasing energy needs [5, 6]. These changes in energy demand and generation are challenging for network operators and power generation facilities, since power needs are becoming less stable and unpredictable while rising at the same time [7, 8]. To address those challenges, accurate and fine-grained monitoring of electrical energy consumption within residential environments is needed [2, 9] as well as proper demand management [10]. However, nowadays, energy monitoring is mostly done via an aggregated measure of energy consumption in the form of monthly bills and therefore does not address the above-mentioned issues.

To measure energy consumption, smart meters are used. A smart meter, also referred to as a smart plug, is a device used to measure electrical power/energy consumption with resolution in the order of seconds to minutes. Smart meters measure the voltage drop over the device/circuit and the current flowing through the device/circuit with an arbitrary sampling frequency f_s , which usually varies from 1/60 Hz to 30 kHz [11]. Higher sampling frequencies are usually preferred, since they contain more detailed information about the energy consumption; however, they increase linearly the amount of acquired data

* Correspondence: p.schirmer@herts.ac.uk

¹Communications and Intelligent Systems Group, School of Engineering and Computer Science, University of Hertfordshire, Hatfield AL10 9AB, UK
Full list of author information is available at the end of the article

and exponentially the cost of hardware [12]. With the sampling rate in the order of seconds, data handling for several months/years becomes feasible and hardware costs are relatively low. However, with the ability to provide real-time information through smart metering and determining detailed household energy consumption, consumer privacy concerns are arising and energy data protection becomes prominent [7, 13]. To address these issues, energy monitoring must be carried out cost-effectively and under the consideration of privacy concerns.

According to [14], the largest improvements in terms of energy savings can be made when monitoring energy consumption on a device level to detect faulty device operation and inefficient or suboptimal operational strategies. To measure energy consumption on a device level, energy has to be measured either for each device separately using one sensor per device or the aggregated energy (combined energy of several devices measured at one central point, e.g. the power inlet of a household) has to be disaggregated into a device level using computational algorithms. When only using one sensor to disaggregate the total consumed energy and extract energy consumption on the appliance level, the task is referred to as non-intrusive load monitoring (NILM) as introduced in [15]. NILM formulates the energy disaggregation problem as a single-channel source separation problem, where the smart meter is the only input channel measuring the total power consumption, and the goal is to find the inverse of the aggregation function to calculate consumption per device. Comparing with intrusive load monitoring (ILM), NILM has the advantage of requiring less hardware (ILM uses one smart meter per device) as well as meets consumers' acceptability with respect to privacy conserving [7, 13].

In general, NILM assumes that there is a single observation (smart meter measurements) and multiple unknowns (electrical devices) making the disaggregation problem highly under-determined and difficult to solve without any further constraints. Therefore, several approaches for disaggregation have been proposed, which can be briefly split into methods with and without source separation (SS). Approaches without SS are based on the decomposition of the aggregated signal to a sequence of feature vectors, which will be classified to device labels by a machine learning (ML) algorithm (e.g. artificial neural networks (ANN) [16], decision trees (DT) [17], hidden Markov models (HMM) [18], K -nearest neighbours (KNN) [19], support vector machines (SVM) [20]), or by a pre-defined set of rules and thresholds [21, 22]. Furthermore, recent research in deep learning and big data has led to a significant increase of use of data-driven approaches using large-scale datasets (e.g. AMPd [23]). Approaches based on convolutional neural networks (CNNs) [24–26], recurrent neural networks (RNNs) [27, 28] and long short-time memories (LSTMs) [27, 29] have been proposed in the

literature, while denoising autoencoders (dAEs) [30] and gate recurrent units (GRUs) [26] have also been used. Approaches with SS are based on single-channel source separation algorithms (e.g. non-negative matrix factorization [31], sparse component analysis [32]) to extract the consumption of each device from the aggregated signal by using additional constraints (e.g. sparseness or sum-to-one [33]) during the optimization procedure. The features extracted from the aggregated signal in approaches with and without SS strongly depend on the sampling frequency, with either macroscopic (for low sampling frequency) or microscopic (for high sampling frequency) features being extracted. Macroscopic features are mainly active and reactive power, while statistical values from the active or reactive power (e.g. mean, median, variance or energy) can be estimated as well [34]. Microscopic features can be current harmonics or transient energy [21, 35] and require high-sampling frequency to be calculated (1 kHz and above).

Several NILM approaches with and without SS have been proposed in the literature. In these approaches, one- or multi-state electrical devices have been modelled by finite-state machines, i.e. with steady energy consumption behaviour per operational state [15]. In contrast to one/multi-state devices, there is no established approach in detecting appliances with continuous power consumption or with non-linear behaviour and highly varying power signature [36, 37]. Researchers have addressed this issue by using high-frequency features or wavelets to detect transient device behaviour, which however have the drawback of higher cost in hardware and increased computational power needed [12, 37, 38]. Therefore, most approaches use disaggregation algorithms with sampling rates in the order of seconds to minutes, in addition to temporal information (e.g. factorial hidden Markov models (FHMM) [18, 39]) to identify appliances with varying power consumption [12, 40]. Furthermore, special filtering techniques (e.g. Kalman filters [41]) with time-varying coefficients and probabilistic approaches using appliance grouping [42] have been proposed to address the issue of modelling devices with continuous or non-linear characteristics.

In this paper, we propose the integration of temporal contextual information for each electrical appliance in the form of concatenation of adjacent feature vectors within a device-dependent time window to improve device detection performance in NILM. The remainder of this paper is organized as follows: in Section 2, the proposed NILM approach using temporal contextual information per device is presented. In Section 3, the experimental set-up is described, and in Section 4, the evaluation results are presented. Finally, the paper is concluded in Section 5.

2 Methods

NILM energy disaggregation can be formulated as the task of determining the power consumption on a device level based on the measurements of one sensor, within

the time window (frame or epoch). Specifically, for a set of $M - 1$ known devices each consuming power p_m with $1 \leq m \leq M$, the aggregated power P_{agg} measured by the sensor will be:

$$P_{\text{agg}} = f(p_1, p_2, \dots, p_{M-1}, p_g) = \sum_{m=1}^{M-1} p_m + p_g = \sum_{m=1}^M p_m \quad (1)$$

where $p_g = p_M$ is a ‘ghost’ power consumption usually consumed by one or more unknown devices. In NILM, the goal is to find estimations $\hat{P} = \{\hat{p}_m, \hat{p}_g\}$ of the power consumption of each device m using an estimation method f^{-1} with minimal estimation error and $\hat{p}_M = \hat{p}_g$, i.e.:

$$\hat{P} = \{\hat{p}_1, \hat{p}_2, \dots, \hat{p}_{M-1}, \hat{p}_g\} = f^{-1}(P_{\text{agg}})$$

$$s.t. \quad \underset{f^{-1}}{\operatorname{argmin}} \{(P_{\text{agg}} - \hat{P})^2\} = \underset{f^{-1}}{\operatorname{argmin}} \left\{ (P_{\text{agg}} - \sum_{m=1}^M \hat{p}_m)^2 \right\} \quad (2)$$

2.1 Baseline NILM architecture

As a baseline NILM approach, we consider a data-driven energy disaggregation methodology without the use of SS techniques, adopted in several publications found in the literature [39, 43–46]. The baseline NILM consists of pre-processing of the aggregated signal P_{agg} , then decomposition of the sequence of frames to a sequence of feature vectors followed by processing from a classification/regression algorithm using pre-trained appliances’ models to determine device operation as shown in Fig. 1.

During the pre-processing step, filtering and/or down-sampling is performed, and then the signal is frame blocked. Framing can be done either with constant or with variable frame length [35, 47]. In the state-based baseline NILM approach, in order to estimate the device consumption on a state level, a regression algorithm instead of a classification algorithm is used [48, 49], while

classification is used in event-based approaches to detect devices’ on/off states [39, 45, 46].

2.2 Proposed NILM architecture

The proposed methodology uses a two-stage disaggregation scheme, with the first stage performing power consumption estimation for each device by extending the baseline NILM architecture to using temporal contextual information (TCI) and the second stage fusing the estimation results of each device using a regression model. The block diagram of the proposed two-stage NILM architecture using TCI is illustrated in Fig. 2.

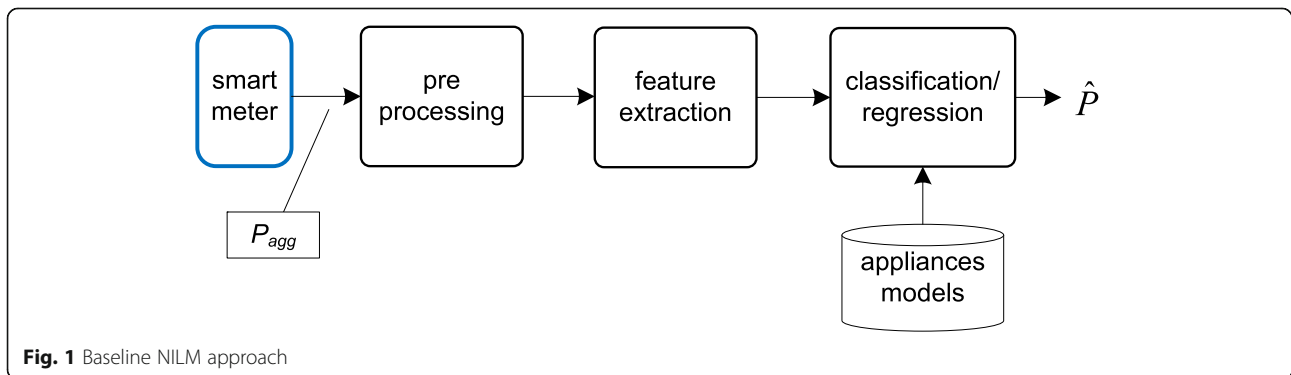
Similarly to the baseline NILM, the aggregated power consumption signal P_{agg} is initially pre-processed, and a feature vector v_t , $v_t \in \mathbb{R}^L$ is extracted for every frame h_t with $1 \leq t \leq T$, where T is the total number of frames. During stage 1, the feature vectors are expanded to C_m using their N adjacent ones, thus creating a temporal contextual window w of length equal to $w = 2N + 1$ concatenated frames, i.e.:

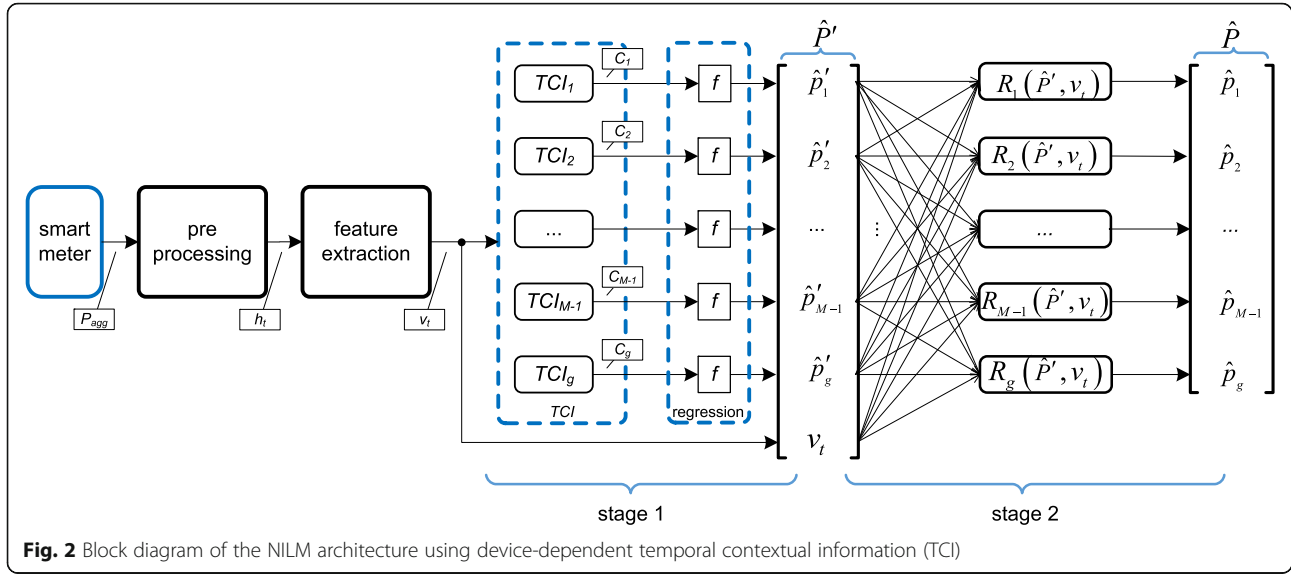
$$C_{m_t} = \text{TCI}_m(v_t, w_{\text{opt}}^m) = [v_{t-N_{\text{opt}}^m}, \dots, v_t, \dots, v_{t+N_{\text{opt}}^m}] \quad (3)$$

where TCI_m is the temporal contextual information expansion function for the m -th device and C_{m_t} is the expansion for the m -th device and the t -th frame. The TCI expansion is performed separately for each device m using its optimal temporal contextual information $w_{\text{opt}} = \{w_{\text{opt}}^m\}$, with w_{opt} being calculated offline on a bootstrap training dataset. The expanded feature vector C_m of each device m is then processed by a regression model $f()$, and the output of stage 1, \hat{p}_m' , is an initial estimation of the power consumption of each device:

$$\hat{p}_m' = f(C_m) \quad (4)$$

The power consumption estimations, $\hat{P}' \in \mathbb{R}^M$ of the M devices from stage 1, are used together with the feature vector,





v_t in order to calculate enhanced estimations of the power consumptions of the M devices. In detail, in the second stage M regression models are receiving as input the power consumption estimates \hat{P}' from stage 1 and the initial feature vector v_t . The use of the device estimates \hat{P}' allows the second-stage regression model estimators to model power consumption correlations between different devices. In both stages 1 and 2, the regression models of the M devices operate in parallel and separately for each device. The proposed methodology combines the integration of temporal contextual information with the device-specific operation of each of the M appliances, thus capturing temporal information individually for each appliance and learning it by the regression model.

3 Experimental set-up

The proposed two-stage NILM architecture with the device-dependent temporal contextual information presented in Section 2 was evaluated using a number of publicly available datasets and a deep learning algorithm for regression. The datasets and parameters set for deep learning regression are presented below.

3.1 Databases

Three different publicly available databases were used, namely the ECO [50], the REDD [51], and the iAWE [52] database. The ECO and REDD databases consist of different datasets with each of them containing power consumption recordings from different houses, while iAWE database consists of recordings from one house. The evaluated datasets are tabulated in Table 1 with the number of appliances denoted in column '#App'. In the same column, the number of appliances in brackets is the number of appliances after excluding devices with

power consumption below 25 W (italic entries), which were added to the power of the 'ghost device', similarly to the experimental set-up followed in [53, 54]. The next three columns in Table 1 are tabulating the sampling period T_s , the duration T and the appliance types of each evaluated dataset.

The appliances' type categorization is based on their operation as described in [55, 56], i.e. one-state devices have only on/off status (e.g. resistive lamps, kettles, or fridges without significant power spikes), multi-state devices have several discrete power consumption states (e.g. washing machines including different washing cycles), non-linear loads (e.g. electronics) and devices with continuous power consumption signature, which are controlled by power electronics (e.g. air condition) and usually have an exponential decay pattern. In all appliance types, a peak might appear at the beginning of their signature, e.g. in refrigerators. Characteristic examples of the power consumption signatures of each of the four appliance types are illustrated in Fig. 3. The ECO-3 and REDD-5 datasets were excluded as ECO-3 contains only the aggregated signal and not the power consumptions per device; thus, there is no ground truth to evaluate NILM approaches [50], and REDD-5 has significantly short-monitoring duration [57]. Regarding the size of the evaluated data, the whole REDD database was used (ignoring the gaps in the measurements as in [58]), while 1 week of data was chosen for the ECO and iAWE databases to have similar amounts of training samples as in the REDD dataset. In detail, the week from 5 July until 11 July 2012 was selected from the ECO database while the week from 8 June until 14 June was selected for the iAWE database, respectively. These particular weeks were selected in order as many as possible devices to appear in the aggregated signal, and since in previous

Table 1 List of evaluated datasets and their properties

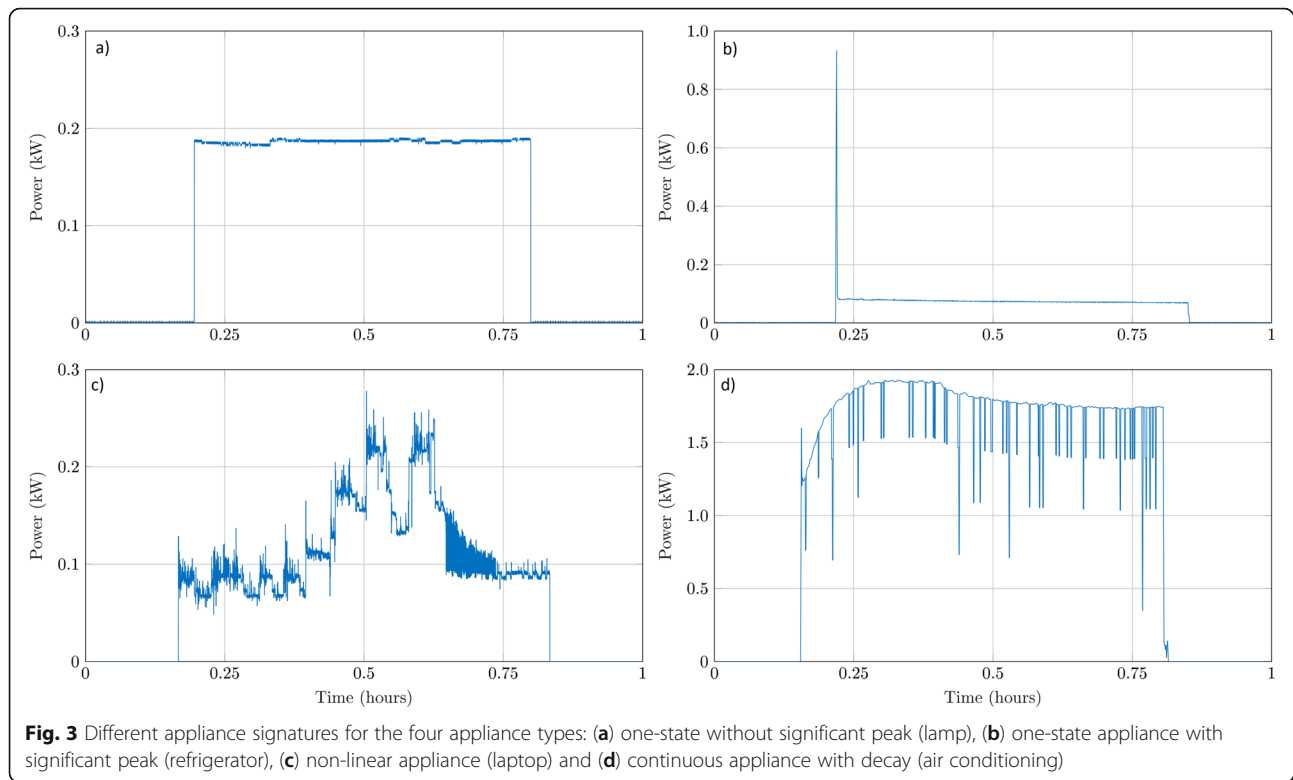
Dataset	#App	T_s (s)	T (days)	Appliance type	Appliances
ECO-1	7 (6)	1	7	One state/multi-state	(1) Fridge, (2) dryer, (3) <i>coffee machine</i> , (4) kettle, (5) washing machine, (6) <i>PC</i> , (7) freezer
ECO-2	12 (9)	1	7	One state/multi-state/non-linear	(1) <i>Tablet</i> , (2) <i>dishwasher</i> , (3) air exhaust, (4) fridge, (5) entertainment, (6) freezer, (7) <i>kettle</i> , (8) lamp, (9) laptop, (10) <i>stove</i> , (11) TV, (12) stereo
ECO-4	8 (8)	1	7	One state/multi-state/non-linear	(1) Fridge, (2) kitchen appliances, (3) lamp, (4) stereo and laptop, (5) freezer, (6) <i>tablet</i> , (7) entertainment, (8) microwave
ECO-5	8 (6)	1	7	One state/multi-state/non-linear	(1) <i>Tablet</i> , (2) coffee machine, (3) <i>kettle</i> , (4) microwave, (5) fridge, (6) entertainment, (7) <i>PC</i> , router and printer, (8) <i>fountain</i>
ECO-6	7 (6)	1	7	One state/multi-state/non-linear	(1) <i>Lamp</i> , (2) laptop and printer, (3) <i>routers</i> , (4) coffee machine, (5) entertainment, (6) fridge, (7) kettle
REDD-1	18 (17)	3	14	One state/multi-state/continuous	(1) Oven, (2) oven, (3) refrigerator, (4) dishwasher, (5) kitchen outlets, (6) kitchen outlets, (7) lighting, (8) washer-dryer, (9) microwave, (10) bathroom, (11) electric heat, (12) <i>stove</i> , (13) kitchen outlets, (14) kitchen outlets, (15) lighting, (16) lighting, (17) <i>washer-dryer</i> , (18) washer-dryer
REDD-2	9 (10)	3	11	One state/multi-state	(1) Kitchen outlets, (2) lighting, (3) <i>stove</i> , (4) microwave, (5) washer-dryer, (6) kitchen outlets, (7) refrigerator, (8) dishwasher, (9) disposal
REDD-3	20 (18)	3	14	One state/multi-state/non-linear	(1) Outlets unknown, (2) <i>outlets unknown</i> , (3) lighting, (4) electronics, (5) refrigerator, (6) disposal, (7) dishwasher, (8) furnace, (9) lighting, (10) outlets unknown, (11) washer-dryer, (12) washer-dryer, (13) <i>lighting</i> , (14) microwave, (15) lighting, (16) <i>smoke alarms</i> , (17) lighting, (18) bathroom, (19) kitchen outlets, (20) kitchen outlets
REDD-4	18 (16)	3	14	One state/multi-state/continuous/non-linear	(1) Lighting, (2) furnace, (3) kitchen outlets, (4) outlets unknown, (5) washer-dryer, (6) <i>stove</i> , (7) air conditioning, (8) air conditioning, (9) <i>miscellaneous</i> , (10) <i>smoke alarms</i> , (11) lighting, (12) kitchen outlets, (13) dishwasher, (14) bathroom, (15) bathroom, (16) <i>lighting</i> , (17) lighting, (18) air conditioning
REDD-6	15 (14)	3	12	One state/multi-state/continuous/non-linear	(1) Kitchen outlets, (2) <i>washer-dryer</i> , (3) <i>stove</i> , (4) electronics, (5) bathroom, (6) refrigerator, (7) dishwasher, (8) outlets unknown, (9) outlets unknown, (10) electric heat, (11) kitchen outlets, (12) lighting, (13) air conditioning, (14) air conditioning, (15) air conditioning
iAWE	10 (9)	1	7	One state/multi-state/continuous	(1) Fridge, (2) air condition, (3) air condition, (4) washing machine, (5) laptop, (6) iron, (7) <i>kitchen</i> , (8) television, (9) <i>water filter</i> , (10) water motor

Italic entries indicate power consumption below 25 W

papers using the ECO and iAWE databases [44, 50], the time interval used has not been reported.

In Table 2, the appliances from each dataset are categorized according to the four different appliance types mentioned above. The categorization is done with respect to the electrical properties of the appliances and their corresponding power consumption signatures. In addition, the percentage of the total energy per appliance type in each dataset is given. The ID number of appliances (columns 'App') corresponds to the appliances of each dataset as denoted in Table 2.

As can be seen in Tables 1 and 2, the number of appliances as well as the appliance type in the evaluated datasets is varying. In particular, the number of appliances varies from 6 (ECO-1) to 18 (REDD-3) while the number of appliance types varies from 2 (REDD-2) to 4 (REDD-4/6); thus, the 11 evaluated datasets include different device combinations and characteristics, which are representative of modern households. Common in all datasets is their relatively low sampling period (1–3 s) and the consideration of active power samplings only, resulting to computational simplicity and runtime advantages

**Table 2** Distribution of four appliance types

Dataset	One state/multi-state without power peak (A)		One state/multi-state with power peak (B)		Non-linear (C)		Continuous (D)	
	Energy (%)	App	Energy (%)	App	Energy (%)	App	Energy (%)	App
ECO-1	57.3	(2), (4), (5)	42.7	(1), (7)	0	–	0	–
ECO-2	4.7	(3), (8)	46.1	(4), (6)	49.2	(5), (9), (11), (12)	0	–
ECO-4	6.2	(3), (8)	74.1	(1), (5)	19.7	(2), (4), (7)	0	–
ECO-5	12.7	(2), (4)	45.6	(5)	41.7	(6), (7)	0	–
ECO-6	16.6	(4), (7)	21.0	(6)	62.4	(2), (5)	0	–
REDD-1	13.1	(1), (2), (4), (9), (11)	67.5	(3), (7), (8), (15), (16), (18)	19.4	(5), (6), (10), (13), (14)	0	–
REDD-2	11.4	(3), (4), (9)	77.4	(2), (5), (7), (8)	11.2	(1), (6)	0	–
REDD-3	6.6	(6), (14), (18)	64.0	(1), (3), (5), (7)–(12), (15), (17)	29.4	(4), (19), (20)	0	–
REDD-4	3.3	(6)	71.7	(1), (2), (5), (11), (13)–(15), (17)	24.0	(3), (4), (12)	1.1	(7), (8), (18)
REDD-6	7.2	(3), (10)	52.7	(6), (7)	17.5	(1), (4), (5), (8), (9), (11), (12)	22.6	(13)–(15)
iAWE	1.1	(4), (6)	14.7	(1)	10.9	(5), (8)	73.3	(2), (3), (10)

A one state/multi-state appliances without significant power peak, B appliances with significant power peak, C non-linear appliances, D continuous appliances, across the 11 evaluated datasets

[59]. Furthermore, all three databases were recorded within the last decade meaning that the households used were equipped with recent device technology [50, 51].

In our experimental set-up, the real aggregated signal (which includes ghost power from unknown devices) was used to evaluate the performance of the proposed NILM methodology, thus making the experimental set-up identical to real-life conditions. Specifically, the input aggregated power consumption signal we used was the originally measured by the smart meter (one sensor only) during data acquisition (similarly to [60]) and not an artificially generated aggregated signal created by adding the power consumption signals from a manually selected closed set of devices (synthesized data), as in [29, 61–63], which was criticized in [64] for not corresponding to real-world conditions.

3.2 Pre-processing and feature extraction

During pre-processing, the aggregated signal was frame blocked in frames of ten samples with overlap between successive frames equal to 50% (i.e. five samples). For every frame, a feature vector consisting of the mean, root mean square, standard deviation and peak to root mean square value was calculated, similarly to [65], resulting to feature vectors of dimensionality equal to four. In detail, the mean value is used as the most general information about the energy consumption in each frame, while the root mean square value is used as a filtered version of the mean value smoothing outliers and small changes (noise) in the power consumption signal [65]. Moreover, the standard deviation is used in order to capture sudden changes of the power signal within a frame, i.e. changes of device states, while the peak to root mean square value is selected to capture the maximum change in power normalized to the root mean square value of the frame in order to have a quantitative measure of change in power within each frame [65]. In order to consider temporal contextual information, expanded feature vectors were extracted by concatenating to each feature vector, the N preceding and the N succeeding vectors as described in Section 2.

For the regression models of stage 1, feed-forward deep neural networks (DNNs) were used. In detail, the DNN consisted of 3 hidden layers with 32 sigmoid nodes per layer. The number of layers and nodes was empirically selected after evaluation on a bootstrap training subset with artificially generated aggregated data (removed ghost power) as shown in Table 3. A ‘one vs. all’ regression approach was followed; thus, the output layer consisted of one regression node only predicting the power of the m -th appliance. In order to avoid overlap between training and test data, each of the evaluated datasets was equally split into two subsets, one for training the DNN models and one for evaluating the proposed architecture.

Table 3 DNN regression performance for different numbers of hidden layers and nodes per layer

Layers/nodes	4	8	16	32	64	128
1 layer	80.40%	87.50%	87.90%	83.70%	86.40%	81.70%
2 layers	70.10%	86.40%	86.90%	87.50%	82.70%	83.60%
3 layers	80.40%	86.70%	87.90%	88.70%	88.40%	84.20%
4 layers	75.40%	87.95%	87.02%	87.15%	85.32%	x

The best performing length of the temporal contextual window w for each of the evaluated datasets is indicated in *italics*

4 Results and discussion

The architecture presented in Section 2 was evaluated according to the experimental set-up described in Section 3. The performance was evaluated in terms of estimation accuracy (E_{ACC}), as proposed in [51], taking into account the estimated power \hat{p}_m where T is the number of disaggregated frames and M is the number of disaggregated devices including the ghost power, i.e.:

$$E_{ACC} = 1 - \frac{\sum_{t=1}^T \sum_{m=1}^M |\hat{p}_m^t - p_m^t|}{2 \sum_{t=1}^T \sum_{m=1}^M |p_m^t|} \quad (5)$$

For evaluating the estimation accuracy on the device level, Eq. 5 was modified and the summation over M appliances was eliminated resulting in Eq. 6

$$E_{ACC}^i = 1 - \frac{\sum_{t=1}^T |\hat{p}_m^t - p_m^t|}{2 \sum_{t=1}^T |p_m^t|} \quad (6)$$

The NILM architecture with temporal contextual information (TCI) was tested for a set of temporal contextual windows of different length. The experimental results of the TCI architecture (i.e. the output of stage 1 in Fig. 2) for different temporal contextual window lengths w , with the same w for all devices and $1 \leq N \leq 6$, are shown in Table 4. The best performing length of the temporal contextual window w for each of the evaluated datasets is indicated in *italics*. In the first column ($w = 1$), the performance without TCI is given. In the last column (w_{opt}), the estimation accuracy E_{ACC} when using the optimal temporal contextual window separately for each device is shown.

As can be seen in Table 4, the use of TCI improves energy disaggregation performance when compared with the baseline NILM system ($w = 1$) across all evaluated datasets. In the case of using a temporal contextual window of the same length for all devices, i.e. $w = 3$ up to $w = 13$, the best performing set-up varies from $w = 5$ to $w = 11$. In general, the datasets with optimal w in low lengths ($w \leq 5$) mostly have one/multi-state types of devices, while datasets with higher optimal TCI lengths ($w \geq 9$) are dominated by devices of non-linear/continuous type. The NILM performance using TCI is further improved when the optimal temporal contextual window

Table 4 Energy disaggregation performance in terms of estimation accuracy (E_{ACC}) for different temporal contextual window lengths w

Dataset	$w = 1$	$w = 3$	$w = 5$	$w = 7$	$w = 9$	$w = 11$	$w = 13$	w_{opt}
ECO-1	70.0	70.6	70.6	72.8	72.0	71.2	71.1	74.2
ECO-2	75.0	76.0	76.0	76.1	77.3	76.1	75.1	79.5
ECO-4	79.7	79.9	80.0	80.1	81.1	80.2	79.2	83.3
ECO-5	84.5	84.6	84.6	85.7	86.8	85.8	84.9	87.9
ECO-6	80.8	81.1	81.3	81.5	81.4	80.7	79.7	82.3
REDD-1	69.2	69.2	71.2	70.2	69.4	69.6	69.7	72.7
REDD-2	73.8	75.9	76.9	76.9	76.0	75.9	74.9	78.2
REDD-3	62.5	62.5	63.7	63.6	63.0	63.1	62.8	64.6
REDD-4	70.7	71.0	71.3	73.5	73.8	72.9	72.1	74.3
REDD-6	77.9	78.9	79.1	79.1	79.0	78.1	77.1	80.7
iAWE	63.1	63.8	65.9	66.1	67.9	68.9	67.7	70.4

The best performing length of the temporal contextual window w for each of the evaluated datasets is indicated in *italics*

length per device is used (w_{opt}). Specifically, the use of an optimized w value for each device instead of a flat value for all devices improves the performance from 0.5 (REDD-4) up to 2.2% (ECO-2/REDD-1), in terms of absolute improvement. The use of device-dependent TCI was found to improve the performance across all evaluated datasets and especially in the datasets with approximately equal energy consumption distribution of the appliances types, like datasets ECO-2 and REDD-1.

Next, we evaluated the performance of the two-stage methodology presented in Section 2. The evaluation results of the proposed NILM architecture are shown in Table 5. For the purpose of direct comparison of the two-stage architecture with the TCI approach (stage 1), the same training and test subset division was used in all evaluated datasets. The best achieved performance of

the TCI approach for each of the evaluated datasets shown in Table 4 is repeated in Table 5 as well.

As can be seen in Table 5, the proposed two-stage methodology outperforms the TCI NILM architecture (stage 1) in all evaluated datasets. In detail, the highest performance improvement (when considering temporal contextual window of the same length for all devices) in terms of E_{ACC} values was observed in the REDD-3 dataset (+ 5.2% for $w = 5$) followed by the REDD-2/ECO-5 dataset (+ 3.0%, for $w = 5$), while the lowest improvement was found in the REDD-6 dataset (+ 0.1%, for $w = 3$), when compared with the TCI NILM. Moreover, the best energy disaggregation performance for 10 out of 11 datasets was observed for temporal contextual window lengths between $3 \leq w \leq 11$ with the majority of the datasets having an optimal temporal contextual window length between $5 \leq w \leq 9$. In the case of the ECO database (with only 6–9 appliances per dataset), the two-stage NILM methodology offered an improvement of 0.5–3.0% in terms of E_{ACC} , while the REDD database (with 10–18 appliances per dataset) offered an improvement of 0.1–5.2%. When considering the optimal temporal contextual window length per device (column ' w_{opt} ' in Table 5), the energy disaggregation improvement offered by the two-stage NILM architecture is even higher. In particular, the highest performance improvement was observed in ECO-2 and ECO-4 datasets (+ 5.2% and + 3.0%, respectively), while the lowest improvement was observed in ECO-5 dataset (+ 0.1%), when compared with the TCI NILM. When compared with the baseline NILM, the highest performance improvement is + 10.0% (iAWE) and the lowest one is + 2.0% (ECO-6).

To further compare the results with the NILM methods proposed in the literature, the very recent work of [66] was used, which includes a summary of NILM performances for the REDD database for different setups. Approaches using the most popular experimental

Table 5 Energy disaggregation performance in terms of estimation accuracy for the two-stage NILM methodology

Dataset	Baseline NILM	TCI (stage 1)	$w = 1$	$w = 3$	$w = 5$	$w = 7$	$w = 9$	$w = 11$	$w = 13$	w_{opt}
ECO-1	70.0	72.8	70.6	72.7	73.2	73.7	67.6	67.1	67.5	76.1
ECO-2	75.0	77.3	75.1	76.6	76.5	76.4	79.9	76.1	76.9	84.1
ECO-4	79.7	81.1	82.3	83.1	82.9	83.4	83.2	82.9	82.1	86.4
ECO-5	84.5	86.8	87.3	87.3	87.4	87.3	89.8	87.4	87.6	89.9
ECO-6	80.8	81.5	81.1	82.0	80.7	80.5	80.5	80.4	80.3	82.8
REDD-1	69.2	71.2	70.0	72.9	73.6	69.2	70.1	69.4	67.2	73.9
REDD-2	73.8	76.9	74.3	74.4	77.9	76.4	75.3	71.9	73.0	80.1
REDD-3	62.5	63.7	66.6	68.7	68.9	67.1	63.2	62.8	60.8	69.7
REDD-4	70.7	73.8	73.9	73.2	74.3	74.7	73.0	73.1	72.9	76.3
REDD-6	77.9	79.1	78.5	79.2	79.1	77.4	77.5	77.7	76.7	81.3
iAWE	63.1	68.9	63.9	64.0	66.3	64.7	66.9	71.4	64.1	73.1

The best performing length of the temporal contextual window w for each of the evaluated datasets is indicated in *italics*

set-up using houses 1, 2, 3, 4, and 6 with all devices and measuring performance using the E_{ACC} metric were considered. Moreover, the results from [66] were extended by including recently published results [67, 68] on the same experimental set-up. It is worth mentioning that although the same data and the same accuracy metric was used, direct comparison is not assured as data splits or pre-processing might vary between the compared approaches (such information is not provided in most papers found in the bibliography). The results are tabulated in Table 6.

As can be seen in Table 6, the proposed fusion methodology outperforms all other reported approaches on the REDD-1/2/3/4/6 dataset set-up. In detail, the proposed approach outperforms the Powerlets approach [67] by 4.3%, while it performs 1.7% better than supervised GSP proposed in [68]. However, it must be noted that the approach in [68] uses a reduced number of appliances and thus cannot be directly compared with the other NILM approaches.

Analysis of the proposed two-stage NILM methodology on a device level was performed. In Table 7, the energy disaggregation improvement in terms of absolute increase of device estimation accuracy (E_{ACC}^i) and the corresponding optimal temporal contextual window length per device, respectively are presented. The first column in Table 7 denotes the type of each appliance as defined in Tables 1 and 2.

As can be seen in Table 7, appliances belonging to type A (i.e. single- or multi-state appliances with their power consumption signature not varying in time, like air exhaust, disposal, electric heat, iron, lamp) are not significantly benefiting by the two-stage NILM methodology with temporal contextual information since the energy disaggregation improvement for type A devices ranges between 0.0 and 3.4% with an average improvement of 1.6%. Type B appliances (i.e. devices without strong temporal behaviour but with significant peak power at the beginning of their power signature, like

dishwasher, freezer, fridge, washer-dryer) were found to benefit from the proposed methodology with the energy disaggregation improvement for type B appliances ranging between 0.4 and 17.8% with an average improvement of 8.6%. In the case of non-linear appliances (appliances type C, e.g. electronic devices, entertainment, laptops), the power signature is usually strongly varying with time and the temporal contextual information can capture well their dynamic characteristics, with the energy disaggregation improvement for type C appliances ranging between 0.2 and 12.7% with an average improvement of 3.8%. As regards continuous devices (appliances type D, like air-conditioner and water motor), their power signature appears in the form of an exponential rise or decay including significant power peaks at the onset of their signature. Due to their slowly but strongly time-varying behaviour, their amplitude variation can be captured by temporal contextual information and misclassification with multi-state appliances of the similar consumption amplitude levels can be reduced, with the energy disaggregation improvement for type D devices ranging between 1.4 and 44.7% with an average improvement of 28.6%. The effect of the two-stage temporal contextual information NILM methodology proposed in Section 2 on each of the four appliance types is summarized in Table 8.

As can be seen in Table 8, the energy disaggregation performance in type D devices improves by almost 30%, followed by type B benefiting by almost 10%. Also, the average optimal temporal contextual window length for appliance types D and B is $w = 9.00$ and $w = 7.38$, respectively. For the case of non-linear appliances (type C), the performance improvement is almost 4%; however, the average optimal window length is greater than the one of type B, which is most probably owed to the longer duration of patterns as well as the non-repetitive micropatterns within non-linear appliances. Furthermore, the two-stage architecture improves the detection of continuous or non-linear appliances as they can be highly related to the daily routine of the users/consumers or even be related/dependent to each other as for example, in the case of TV and entertainment appliances which are usually interconnected. For such devices, with inter-device dependencies or daily routine patterns, the a priori knowledge of the power consumption of other devices they operate together with or devices with similar daily routine (i.e. usually operating or not operating simultaneously) can be beneficial for the estimation of their power consumption. Such devices can benefit from the fusion stage of the proposed architecture in which estimates of the power consumption of the other appliances (calculated from the first stage) are used as input. Except this, detection of devices with power spikes, i.e. peaks that appear during the switching on of electrical motors, e.g. in fridges or freezers, was found to benefit from the fusion stage of the proposed methodology, since the presence of a power spike within a frame affects the distribution of energy

Table 6 Comparison of power disaggregation accuracy values (E_{ACC}) for recently proposed NILM methodologies. The reported performance is the average E_{ACC} across houses 1, 2, 3, 4, and 6

NILM method	Year	Dataset	E_{ACC} (%)
General sparse coding [69]	2010	REDD-1/2/3/4/6	56.4
Discriminating sparse coding [69]	2010	REDD-1/2/3/4/6	59.3
Temporal ML [70]	2011	REDD-1/2/3/4/6	53.3
Powerlets-PED [67]	2015	REDD-1/2/3/4/6	72.0
Greedy deep sparse coding [71]	2017	REDD-1/2/3/4/6	62.6
Exact deep sparse coding [71]	2017	REDD-1/2/3/4/6	66.1
Supervised GSP* [68]	2018	REDD-1/2/3/4/6	67.8
Unsupervised GSP* [68]	2018	REDD-1/2/3/4/6	74.6
Proposed TCI method	2020	REDD-1/2/3/4/6	76.3

*Not directly comparable due to a reduced number of devices

The best performing length of the temporal contextual window w for each of the evaluated datasets is indicated in italics

Table 7 Energy disaggregation performance increase for each device in terms of estimation accuracy E_{ACC}^i when using the optimal temporal contextual window length w per device

Type	Appliance	ECO					REDD					iAWE
		1	2	4	5	6	1	2	3	4	6	1
D	Air conditioner									1.4 (3)	9.7 (11)	26.5 (9)
A	Air exhaust		0.0 (1)									
A/B/C	Bathroom-Gfi						0.9 (13)		13.1 (3)	0.1 (11)	2.4 (3)	
A	Coffee maker				1.8 (3)	0.2 (3)						
B	Dishwasher						4.8 (5)	4.9 (5)	0.0 (1)	3.6 (13)	0.5 (3)	
A	Disposal							0.5 (3)	0.0 (1)			
A	Dryer	0.6 (7)										
A	Electric heat						1.9 (3)			0.0 (1)		
C	Electronics								4.3 (11)		10.5 (7)	
C	Entertainment		2.3 (5)	1.8 (7)	1.2 (7)	3.2 (3)						
B	Freezer	3.4 (5)	3.4 (9)	0.4 (7)								
B	Fridge	2.8 (5)	2.8 (5)	17.1 (5)	0.6 (3)	5.1 (3)	4.8 (11)	5.6 (5)	3.5 (13)		2.0 (3)	0.6 (5)
B	Furnace								60.8 (11)	5.4 (13)		
	Ghost	1.1 (3)	2.9 (3)	1.0 (3)	1.0 (9)	0.3 (11)		0.8 (3)	0.0 (1)	0.1 (11)	0.0 (1)	0.4 (3)
A	Iron											0.0 (1)
A	Kettle	3.4 (7)				2.4 (3)						
A/B/C	Kitchen			0.0 (1)			6.1 (7)	8.4 (5)	7.7 (3)	2.8 (5)	14.2 (3)	
A	Lamp		0.2 (5)	32.2 (13)			0.1 (3)					
C	Laptop		12.7 (9)									1.1 (7)
B	Lighting						17.8 (7)	4.9 (5)	8.4 (7)	2.0 (7)	5.8 (9)	
A/B	Microwave			0.5 (9)	0.0 (1)		5.7 (7)	0.7 (3)	9.6 (5)			
B/C	Out-unknown								4.6 (7)	1.1 (9)	3.0 (11)	
A	Oven						0.0 (1)					
C	PC + printer				2.3 (5)	3.3 (13)						
C	Stereo		1.9 (7)	0.2 (7)								
A	Stove							7.6 (3)		6.6 (7)	9.4 (3)	
C	TV		0.9 (9)									0.4 (13)
B	Washer-dryer						7.2 (7)	5.7 (11)	14.1 (7)	17.4 (7)		
A	WM	0.0 (1)										0.0 (1)
D	Water motor											44.7 (11)

Table 8 Average E_{ACC} improvement and temporal contextual window length for four appliance types (A, B, C and D)

Appliance type	Average optimal temporal contextual window length w	Average E_{ACC} improvement (%)
A (one state/multi-state without power peak)	2.92	1.6
B (one state/multi-state with power peak)	7.38	8.6
C (non-linear)	8.30	3.8
D (continuous)	9.00	28.6

among the set of devices to be disaggregated which is implicitly expressed by the power consumption estimates of each device detector computed at the first stage of the proposed architecture. The power signature for each appliance type was illustrated in Fig. 3.

5 Conclusion

A two-stage methodology for energy disaggregation using temporal contextual information was presented. The methodology extends the baseline non-intrusive load monitoring (NILM) approach by employing a two-stage disaggregation and using a temporal expansion of the feature vectors within a time window of variable length. The proposed methodology was evaluated using the real-aggregated signal as measured by the smart meter across various datasets of different sampling frequency, number, and types of appliances, demonstrating improvement of performance across all datasets. The maximum improvement in terms of absolute increase of accuracy was equal to 10.0% when using appliance-driven temporal contextual information lengths and two-stage disaggregation. In detail, the most significant improvements were observed for devices with power peaks and exponential decay power consumption signatures such as refrigerators and air conditions. Moreover, improvements in energy disaggregation performance were observed for appliances with strong time-varying power signatures like electronic devices, e.g. stereos, laptops or entertainment electronics. With the use of the fusion stage inter-device dependencies or daily routine patterns can be modelled and power spikes can be found, thus resulting in further improvement of the disaggregation accuracy.

Abbreviations

ANN: Artificial neural network; CNN: Convolutional neural network; DAE: Denoising autoencoders; DNN: Deep neural network; DT: Decision trees; FHMM: Factorial hidden Markov model; GRU: Gate recurrent units; HMM: Hidden Markov model; ICT: Information and communication technologies; ILM: Intrusive load monitoring; KNN: K-nearest -Neighbours; LSTM: Long short-time memory; NILM: Non-intrusive load monitoring; RNN: Recurrent neural network; SS: Source separation; SVM: Support vector machines; TCI: Temporal contextual information

Acknowledgements

This work was partially supported by the UA Doctoral Training Alliance (<https://www.unialliance.ac.uk/>) for Energy in the UK.

Authors' contributions

PAS and IM developed the methodology. PAS carried out the simulations. All authors prepared, read and approved the final manuscript.

Funding

This work was partially funded by the UA Doctoral Training Alliance (<https://www.unialliance.ac.uk/>) for Energy in the UK.

Availability of data and materials

The utilized datasets are available online and free to download under common creative licence (ECO [50] (10.5905/ethz-1007-35), REDD (<http://redd.csail.mit.edu/>) [51] and iAWE [52] (<https://iawe.github.io/>)).

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Author details

¹Communications and Intelligent Systems Group, School of Engineering and Computer Science, University of Hertfordshire, Hatfield AL10 9AB, UK. ²School of Built Environment, Engineering and Computing, Leeds Beckett University, Leeds LS6 3QS, UK.

Received: 25 September 2019 Accepted: 21 January 2020

Published online: 11 February 2020

References

1. L. Pérez-Lombard, J. Ortiz, C. Pout, A review on buildings energy consumption information. *Energy and Buildings*. **40**, 394–398 (2008). <https://doi.org/10.1016/j.enbuild.2007.03.007>
2. Mostafavi S, Cox RW. An unsupervised approach in learning load patterns for non-intrusive load monitoring. In: Fortino G, editor. [Piscataway, NJ]: IEEE; 2017. p. 631–636. doi:10.1109/ICNSC.2017.8000164 .
3. Ogwumike C, Short M, Denai M. Near-optimal scheduling of residential smart home appliances using heuristic approach. In: Piscataway, NJ: IEEE; 2015. p. 3128–3133. doi:<https://doi.org/10.1109/ICIT.2015.7125560>.
4. J. Froehlich, E. Larson, S. Gupta, G. Cohn, M. Reynolds, S. Patel, Disaggregated end-use energy sensing for the smart grid. *IEEE Pervasive Computing*. **10**, 28–39 (2011). <https://doi.org/10.1109/MPRV.2010.74>
5. Chis A, Rajasekharan J, Lunden J, Koivunen V. Demand response for renewable energy integration and load balancing in smart grid communities. In: 2016 24th European Signal Processing Conference (EUSIPCO); Budapest, Hungary. Piscataway, NJ: IEEE; 2016. p. 1423–1427. doi: <https://doi.org/10.1109/EUSIPCO.2016.7760483>.
6. L.R.M. Silva, C.A. Duque, P.F. Ribeiro, Smart signal processing for an evolving electric grid. *EURASIP J. Adv. Signal Process.* **2015**, 210 (2015). <https://doi.org/10.1186/s13634-015-0229-7>
7. Li Z, Oechtering TJ, Skoglund M. Privacy-preserving energy flow control in smart grids. In: Piscataway, NJ and Piscataway, NJ: IEEE; 2016. p. 2194–2198. doi:<https://doi.org/10.1109/CASSP.2016.7472066>.
8. L. Alfieri, Some advanced parametric methods for assessing waveform distortion in a smart grid with renewable generation. *EURASIP J. Adv. Signal Process.* **2015**, 41 (2015). <https://doi.org/10.1186/s13634-015-0195-0>
9. J.A. Cortés, A. Sanz, P. Estopiñán, J.J. García, Analysis of narrowband power line communication channels for advanced metering infrastructure. *EURASIP J. Adv. Signal Process.* **2015**, 18 (2015). <https://doi.org/10.1186/s13634-015-0211-4>
10. J. Xu, M. van der Schaar, Incentive-compatible demand-side management for smart grids based on review strategies. *EURASIP J. Adv. Signal Process.* **2015**, 34 (2015). <https://doi.org/10.1186/s13634-015-0235-9>
11. Gao J, Kara EC, Giri S, Berges M. A feasibility study of automated plug-load identification from high-frequency measurements. In: Piscataway, NJ and Piscataway, NJ: IEEE; 2015. p. 220–224. doi:<https://doi.org/10.1109/GlobalSIP.2015.7418189>.
12. G.C. Koutitas, L. Tassiulas, Low cost disaggregation of smart meter sensor data. *IEEE Sensors Journal*. **16**, 1665–1673 (2016). <https://doi.org/10.1109/JSEN.2015.2501422>
13. K. Buchanan, N. Banks, I. Preston, R. Russo, The British public's perception of the UK smart metering initiative: threats and opportunities. *Energy Policy*. **91**, 87–97 (2016). <https://doi.org/10.1016/j.enpol.2016.01.003>
14. S. Katipamula, M. Brambley, Review article: methods for fault detection, diagnostics, and prognostics for building systems—a review, part II. *HVAC&R Research*. **11**, 169–187 (2005). <https://doi.org/10.1080/10789669.2005.10391133>
15. G.W. Hart, Nonintrusive appliance load monitoring. *Proceedings of the IEEE*. **80**, 1870–1891 (1992). <https://doi.org/10.1109/5.192069>
16. Y.-H. Lin, M.-S. Tsai, An advanced home energy management system facilitated by nonintrusive load monitoring with automated multiobjective power scheduling. *IEEE Transactions on Smart Grid*. **6**, 1839–1851 (2015). <https://doi.org/10.1109/TSG.2015.2388492>
17. Bilski P, Winiński W. Generalized algorithm for the non-intrusive identification of electrical appliances in the household. In: 2017 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS); Bucharest,

- Romania. Piscataway, NJ: IEEE; 2017. p. 730–735. doi:<https://doi.org/10.1109/IDAACS.2017.8095186>.
18. Zoha A, Gluhak A, Nati M, Imran MA. Low-power appliance monitoring using factorial hidden Markov models. In: Palaniswami M, editor. Piscataway, NJ and Piscataway, NJ: IEEE; 2013. p. 527–532. doi:<https://doi.org/10.1109/ISSNIP.2013.6529845>.
 19. Kim Y, Kong S, Ko R, Joo S-K. Electrical event identification technique for monitoring home appliance load using load signatures. In: 2014 IEEE International Conference on Consumer Electronics (ICCE); Las Vegas, NV, USA. Piscataway, NJ: IEEE; 2014. p. 296–297. doi:<https://doi.org/10.1109/ICCE.2014.6776012>.
 20. T. Hassan, F. Javed, N. Arshad, An empirical investigation of V-I trajectory based load signatures for non-intrusive load monitoring. *IEEE Transactions on Smart Grid*. **5**, 870–878 (2014). doi:<https://doi.org/10.1109/TSG.2013.2271282>
 21. Bilski P, Winiński W. The rule-based method for the non-intrusive electrical appliances identification. In: 2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS); Warsaw, Poland. Piscataway, NJ: IEEE; 2015. p. 220–225. doi:<https://doi.org/10.1109/IDAACS.2015.7340732>.
 22. Zhou Y, Zhai Q, Li X, Yang Y. A method for recognizing electrical appliances based on active load demand in a house/office environment. In: [Piscataway, NJ] and [Piscataway, NJ]: IEEE; 2017. p. 3584–3589. doi:10.1109/CAC.2017.8243403 .
 23. S. Makonin, F. Popowich, L. Bartram, B. Gill, I.V. Bajic, editors. AMPDs: a public dataset for load disaggregation and eco-feedback research; 2013.
 24. Q. Wu, F. Wang, Concatenate convolutional neural networks for non-intrusive load monitoring across complex background. *Energies*. **12**, 1572 (2019). doi:<https://doi.org/10.3390/en12081572>
 25. Barsim KS, Yang B. On the feasibility of generic deep disaggregation for single-load extraction; 2018/02/05.
 26. Murray D, Stankovic L, Stankovic V, Lulic S, Sladojevic S. Transferability of neural network approaches for low-rate energy disaggregation. In: ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); Brighton, United Kingdom: IEEE; 2019. p. 8330–8334. doi:<https://doi.org/10.1109/ICASSP.2019.8682486>.
 27. He W, Chai Y. An empirical study on energy disaggregation via deep learning. In: 2016 2nd International Conference on Artificial Intelligence and Industrial Engineering (AIIE 2016); Beijing, China. Paris, France: Atlantis Press; 2016. doi:<https://doi.org/10.2991/aiie-16.2016.77>.
 28. İ. ÇAVDAR, V. FARYAD, New design of a supervised energy disaggregation model based on the deep neural network for a smart grid. *Energies*. **12**, 1217 (2019). doi:<https://doi.org/10.3390/en12071217>
 29. Mauch L, Yang B. A new approach for supervised power disaggregation by using a deep recurrent LSTM network. In: Piscataway, NJ and Piscataway, NJ: IEEE; 2015. p. 63–67. doi:<https://doi.org/10.1109/GlobalSIP.2015.7418157>.
 30. F.C.C. Garcia, C.M.C. Crealy, E.Q.B. Macabebe, Development of an intelligent system for smart home energy disaggregation using stacked denoising autoencoders. *Procedia Computer Science*. **105**, 248–255 (2017). doi:<https://doi.org/10.1016/j.procs.2017.01.218>
 31. A. Rahimpour, H. Qi, D. Fugate, T. Kuruganti, Non-intrusive energy disaggregation using non-negative matrix factorization with sum-to-k constraint. *IEEE Transactions on Power Systems*. **32**, 4430–4441 (2017). doi:<https://doi.org/10.1109/TPWRS.2017.2660246>
 32. Makonin S, Bajic IV, Popowich F. Efficient sparse matrix processing for nonintrusive load monitoring (NILM). Makonin 2014 EfficientSM.
 33. Pathak N, Roy N, Biswas A. Iterative signal separation assisted energy disaggregation. In: Piscataway, NJ: IEEE; 2015. p. 1–8. doi:<https://doi.org/10.1109/IGCC.2015.7393701>.
 34. Gislser C, Ridi A, Zufferer D, Khaled OA, Hennebert J. Appliance consumption signature database and recognition test protocols. In: Piscataway, NJ: IEEE; 2013. p. 336–341. doi:<https://doi.org/10.1109/WoSSPA.2013.6602387>.
 35. Meziane MN, Abed-Meraim K. Modeling and estimation of transient current signals. In: 2015 23rd European Signal Processing Conference (EUSIPCO); Nice. [Piscataway, New Jersey]; 2015. p. 1960–1964. doi:10.1109/EUSIPCO.2015.7362726.
 36. Zhu Y, Lu S. Load profile disaggregation by blind source separation: a wavelets-assisted independent component analysis approach. In: Piscataway, NJ and Piscataway, NJ: IEEE; 2014. p. 1–5. doi:<https://doi.org/10.1109/PESGM.2014.6938947>.
 37. H.-H. Chang, K.-L. Lian, Y.-C. Su, W.-J. Lee, Power-spectrum-based wavelet transform for nonintrusive demand monitoring and load identification. *IEEE Transactions on Industry Applications*. **50**, 2081–2089 (2014). doi:<https://doi.org/10.1109/TIA.2013.2283318>
 38. H.-H. Chang, Non-intrusive demand monitoring and load identification for energy management systems based on transient feature analyses. *Energies*. **5**, 4569–4589 (2012). doi:<https://doi.org/10.3390/en5114569>
 39. Y. Li, Z. Peng, J. Huang, Z. Zhang, J.H. Son, *Energy disaggregation via hierarchical factorial HMM* (2014)
 40. W. Wichakool, Z. Remscrib, U.A. Orji, S.B. Leeb, Smart metering of variable power loads. *IEEE Transactions on Smart Grid*. **6**, 189–198 (2015). doi:<https://doi.org/10.1109/TSG.2014.2352648>
 41. S.R. Shaw, C.R. Laughman, A Kalman-filter spectral envelope preprocessor. *IEEE Transactions on Instrumentation and Measurement*. **56**, 2010–2017 (2007). doi:<https://doi.org/10.1109/TIM.2007.904475>
 42. Y. Liu, G. Geng, S. Gao, W. Xu, Non-intrusive energy use monitoring for a group of electrical appliances. *IEEE Transactions on Smart Grid*. **9**, 3801–3810 (2018). doi:<https://doi.org/10.1109/TSG.2016.2643700>
 43. A. Cominola, M. Giuliani, D. Piga, A. Castelletti, A.E. Rizzoli, A hybrid signature-based iterative disaggregation algorithm for non-intrusive load monitoring. *Applied Energy*. **185**, 331–344 (2017). doi:<https://doi.org/10.1016/j.apenergy.2016.10.040>
 44. van Cutsem O, Ilis G, Kayal M. Automatic multi-state load profile identification with application to energy disaggregation. In: [Piscataway, NJ] and [Piscataway, NJ]: IEEE; 2017. p. 1–8. doi:10.1109/ETFA.2017.8247684 .
 45. A. Marchiori, D. Hakkarinen, Q. Han, L. Earle, Circuit-level load monitoring for household energy management. *IEEE Pervasive Computing*. **10**, 40–48 (2011). doi:<https://doi.org/10.1109/MPRV.2010.72>
 46. M. Figueiredo, A. de Almeida, B. Ribeiro, Home electrical signal disaggregation for non-intrusive load monitoring (NILM) systems. *Neurocomputing*. **96**, 66–73 (2012). doi:<https://doi.org/10.1016/j.neucom.2011.10.037>
 47. Jin Y, Tebekaemi E, Berges M, Soibelman L. Robust adaptive event detection in non-intrusive load monitoring for energy aware smart facilities. In: Staff I, editor. [Place of publication not identified]: IEEE; 2011. p. 4340–4343. doi:10.1109/ICASSP.2011.5947314 .
 48. Kaselimi M, Doulamis N, Doulamis A, Voulodimos A, Protopapadakis E. Bayesian-optimized bidirectional LSTM regression model for non-intrusive load monitoring. In: ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); Brighton, United Kingdom: IEEE; 2019. p. 2747–2751. doi:<https://doi.org/10.1109/ICASSP.2019.8683110>.
 49. P. Vrablecová, A. Bou Ezzeddine, V. Rozinajová, S. Šárík, A.K. Sangaiah, Smart grid load forecasting using online support vector regression. *Computers & Electrical Engineering*. **65**, 102–117 (2018). doi:<https://doi.org/10.1016/j.compeleceng.2017.07.006>
 50. Beckel C, Kleiminger W, Cicchetti R, Staake T, Santini S. The ECO data set and the performance of non-intrusive load monitoring algorithms. In: Srivastava M, editor. New York: ACM; 2014. p. 80–89. doi:<https://doi.org/10.1145/2674061.2674064>.
 51. Kolter JZ, Johnson MJ, editors. REDD: a public data set for energy disaggregation research; 2011.
 52. Batra N, Gulati M, Singh A, Srivastava MB. It's different. In: Unknown, editor. the 5th ACM Workshop; Roma, Italy. New York, NY: ACM; 2013. p. 1–8. doi: <https://doi.org/10.1145/2528282.2528293>.
 53. H. Wang, W. Yang, An iterative load disaggregation approach based on appliance consumption pattern. *Applied Sciences*. **8**, 542 (2018). doi:<https://doi.org/10.3390/app8040542>
 54. S.M. Tabatabaei, S. Dick, W. Xu, Toward non-intrusive load monitoring via multi-label classification. *IEEE Transactions on Smart Grid*. **8**, 26–40 (2017). doi:<https://doi.org/10.1109/TSG.2016.2584581>
 55. S.R. Shaw, S.B. Leeb, L.K. Norford, R.W. Cox, Nonintrusive load monitoring and diagnostics in power systems. *IEEE Transactions on Instrumentation and Measurement*. **57**, 1445–1454 (2008). doi:<https://doi.org/10.1109/TIM.2008.917179>
 56. A. Zoha, A. Gluhak, M.A. Imran, S. Rajasegarar, Non-intrusive load monitoring approaches for disaggregated energy sensing: a survey. *Sensors (Basel, Switzerland)*. **12**, 16838–16866 (2012). doi:<https://doi.org/10.3390/s121216838>
 57. M. Gaur, A. Majumdar, Disaggregating transform learning for non-intrusive load monitoring. *IEEE Access*. **6**, 46256–46265 (2018). doi:<https://doi.org/10.1109/ACCESS.2018.2850707>
 58. Batra N, Kelly J, Parson O, Dutta H, Knottenbelt W, Rogers A, et al. NILMTK. In: Crowcroft J, Penty R, Le Boudex J-Y, Shenoy P, editors. New York, New York, USA: ACM Press; 2014. p. 265–276. doi:<https://doi.org/10.1145/2602044.2602051>.

59. Welikala S, Dinesh C, Ekanayake MPB, Godaliyadda RI, Ekanayake J. A real-time non-intrusive load monitoring system. In: [Piscataway, NJ]: IEEE; 2016. p. 850–855. doi:10.1109/CIINF5.2016.8263057 .
60. S. Makonin, F. Popowich, I.V. Bajic, B. Gill, L. Bartram, Exploiting HMM sparsity to perform online real-time nonintrusive load monitoring. *IEEE Transactions on Smart Grid*. **7**, 2575–2585 (2016). <https://doi.org/10.1109/TSG.2015.2494592>
61. K. Chen, Q. Wang, Z. He, K. Chen, J. Hu, J. He, *Convolutional sequence to sequence non-intrusive load monitoring* (2018)
62. D. Piga, A. Cominola, M. Giuliani, A. Castelletti, A.E. Rizzoli, Sparse optimization for automated energy end use disaggregation. *IEEE Transactions on Control Systems Technology*. **24**, 1044–1051 (2016). <https://doi.org/10.1109/TCST.2015.2476777>
63. D. Egarter, V.P. Bhuvana, W. Elmenreich, PALDi: online load disaggregation via particle filtering. *IEEE Transactions on Instrumentation and Measurement*. **64**, 467–477 (2015). <https://doi.org/10.1109/TIM.2014.2344373>
64. 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**, e1265 (2018). <https://doi.org/10.1002/widm.1265>
65. K. Basu, V. Debusschere, S. Bacha, U. Maulik, S. Bondyopadhyay, Nonintrusive load monitoring: a temporal multilabel classification approach. *IEEE Trans. Ind. Inf.* **11**, 262–270 (2015). <https://doi.org/10.1109/TII.2014.2361288>
66. S. Welikala, C. Dinesh, M.P.B. Ekanayake, R.I. Godaliyadda, J. Ekanayake, Incorporating appliance usage patterns for non-intrusive load monitoring and load forecasting. *IEEE Transactions on Smart Grid*. **10**, 448–461 (2019). <https://doi.org/10.1109/TSG.2017.2743760>
67. Elhamifar E, Sastry S. Energy disaggregation via learning ‘powerlets’ and sparse coding. In: Proceeding AAAI’15 Proceedings of the Twenty-Ninth AAAI Conference on Artificial. Pages 629–635.
68. B. Zhao, K. He, L. Stankovic, V. Stankovic, Improving event-based non-intrusive load monitoring using graph signal processing. *IEEE Access*. **6**, 53944–53959 (2018). <https://doi.org/10.1109/ACCESS.2018.2871343>
69. J. Zico Kolter, Siddarth Batra, Andrew Ng, editor. *Energy disaggregation via discriminative sparse coding*; 2010.
70. V. Stankovic, J. Liao, L. Stankovic, A graph-based signal processing approach for low-rate energy disaggregation. In: 2014 IEEE Symposium on Computational Intelligence for Engineering Solutions (CIES), Orlando, FL. 2014. p. 81–87. <https://doi.org/10.1109/CIES.2014.7011835>.
71. S. Singh, A. Majumdar, Deep sparse coding for non-intrusive load monitoring. *IEEE Transactions on Smart Grid*. **1** (2017). <https://doi.org/10.1109/TSG.2017.2666220>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)