

Article

A Non-Intrusive Load Monitoring Algorithm Based on Non-Uniform Sampling of Power Data and Deep Neural Networks

Marco Fagiani ^{1,*} , Roberto Bonfigli ² , Emanuele Principi ¹ , Stefano Squartini ¹ 
and Luigi Mandolini ²

¹ Department of Information Engineering (DII), Università Politecnica delle Marche, Via Brecce Bianche 1, 60131 Ancona, Italy; e.principi@univpm.it (E.P.); s.squartini@univpm.it (S.S.)

² Research and Development Area, MAC Srl, Via XX Settembre 23, 62019 Recanati (MC), Italy; roberto.bonfigli@mac-italia.com (R.B.); mandolini@mac-italia.com (L.M.)

* Correspondence: m.fagiani@staff.univpm.it; Tel.: +39-071-2204-146

Received: 12 February 2019; Accepted: 5 April 2019; Published: 9 April 2019



Abstract: Nowadays, measurement systems strongly rely on the Internet of Things paradigm, and typically involve miniaturized devices on purpose. In these devices, the computational resources and signal acquisition rates are limited in order to preserve battery life. In addition, the amount of streamed data is affected by the network capacity strictly related to the transmission protocol constraints and the environmental conditions. All those limitations are in contrast with the need of exploiting all possible signal details for the task under study. In the specific application of interest, i.e., Non-Intrusive Load Monitoring (NILM), they could lead to low performance in the energy disaggregation process. To overcome these issues, an ad hoc data reduction policy needs to be adopted, in order to reduce the acquisition and elaboration burden of the device, and, at the same time, to ensure compliance with network bandwidth limits while maintaining a reliable signal representation. Moved by these motivations, an extended evaluation study concerning the application of data reduction strategy to the aggregate signal is presented in this work. In particular, a non-uniform subsampling (NUS) scheme is defined together with a uniform subsampling (US) strategy and compared, in terms of disaggregation performance, with the use of data at original sampling (OS) rate. A Deep Learning based technique is used for disaggregation, having the aggregate active power signal sampled according to diverse sampling schema mentioned above as input. The approaches are tested on the UK-DALE and REDD datasets, and the combination of US+NUS configurations allows for achieving a good performance in terms of F_1 -score, even superior than the one obtained with the OS rate, and a remarkable data reduction at the same time.

Keywords: non-intrusive load monitoring; energy disaggregation; deep learning; neural networks; non-uniform sampling; active power; computational energy management

1. Introduction

In the last several years, an increasing interest on the energy themes focused on the reduction of emissions from fossil sources has led the researchers to develop solutions oriented to improve the users' energy awareness in everyday life actions, as well as to assist them to better schedule these actions. Among these solutions, Non-Intrusive Load Monitoring (NILM) is surely one of the most studied [1], in addition to other useful ones, such as energy management and analytic [2,3], load task scheduling [4], or behaviour-based consumption [5]. These algorithms require high computational burden, since their results could be requested in real time to provide an immediate feedback to the user. In order to limit the computational burden of the algorithms, strategies to reduce the amount of

data required by the applications could be taken into account. Focusing on the energy consumption in a domestic scenario, in particular on the measurement of the appliance power signal, many time instants do not present occurrences of noteworthy events, such as the absence of power consumption (or consumption above a minimal threshold), i.e., at night and/or holidays, or time slots with a reduced variation (low variance) of the power consumption profile (steady state of an appliance). Therefore, in such scenarios, a data reduction policy is a possible solution. In contrast, during the active periods, a higher level of detail of the power consumption trace is needed in order to guarantee the reliability of the advanced elaboration service.

Nowadays, data acquisition systems are more and more oriented towards the use of cloud computing resources, where centralized services can handle multiple requests. Such infrastructure topology is directly applicable in the residential energy scenario, where a smart meter samples the power consumption trace on the domestic power supply line and, by means of an Internet connection, sends the data to the elaboration service on cloud. This scenario lies among the Internet of Things (IoT) paradigm, which points toward the development of minimal devices capable of acquiring and pre-process signals keeping a low elaboration power. Specifically, IoT devices are designed to meet low power requirements, in order to have an improved battery life and longer life-span. To pursue this goal, these devices will demand the main data elaborations to dedicated cloud centres with a high computational power. Therefore, the introduction of application-oriented data reduction policies results beneficial also from the point of view of the acquisition infrastructure, i.e., reduce the overall data stream in order to limit the bandwidth requirements and support more data streams towards the cloud service.

State-of-the-Art

In the scientific literature, several techniques have been proposed to deal with the NILM problem, with different purposes, based on the steady states or transient states [6,7].

The kind of signal trait used for addressing the task depends on the signal detail needed, which reflects on different sampling frequencies. The authors of [8] propose a qualitative performance trend of the disaggregation algorithms for different sampling frequencies. Generally, despite the task to be solved depending on the algorithm formulation, the disaggregation accuracy decreases proportionally with the sampling frequency. A detailed study of Hidden Markov Model-based (HMM) algorithms has been conducted in [9], where the authors evaluated the algorithms performance for sampling rates ranging to 1 s to 6 min by using the REDD dataset. Generally, they found that higher sampling rates provide more accurate disaggregation results. Similarly, the work presented in [10] evaluates the effects of sampling rate reduction on the performance of Factorial Hidden Markov Model (FHMM) and sparse-matrix-based algorithms for NILM [11,12]. The authors evaluated different sampling periods, from 6 s to 15 min, and they showed that the performance degrades nonlinearly as the sampling rate decreases. Basu et al. [13] evaluated the performance of two event-based algorithms operating at different sampling rates 10 s and 15 min, and they observed that the latter achieves the lowest performance.

These outcomes highlight the need to pursue ad hoc strategies for data reduction in order to maintain details on the signal, otherwise it is reasonable to expect a deterioration in performance with low sampling rates. Indeed, this assumption is confirmed by the strategies of data reduction proposed in the literature that are focused either on the transmission of data only when significant changes are detected, or on the compression of data before transmission. In the former group, the acquired samples are transmitted to the NILM algorithm only when relevant events occur in the aggregate power signal [14–20]. Generally, these methods reduce the amount of data during the transmission preserving the details of the original signal, but they depend on the reliability of the event detection method applied to the aggregated power signal. In addition, many state-of-the-art approaches rely on the profile reconstruction before the application of the NILM algorithm. In this sense, the algorithm does not deal with the data reduction of the aggregate power consumption, but it is applied to the

signal at original frequency. The reliability of this procedure is strictly dependent on the relevant samples detection, and on the profile reconstruction technique as well.

In the second class of approaches, data reduction is achieved through lossless or lossy compression algorithms, and they are used depending on the transmission requirements to be satisfied. A study on sampling rate reduction of aggregated power signal and its effects on NILM algorithms has been conducted in [21]. The authors presented a method based on compressed sensing (CS) to reach a lower-than-Nyquist sampling rate. Moreover, the work investigated how the restrictions to the CS sensing matrix associated with random filtering and demodulation affects signal recovery and NILM performance. The experiments have been conducted on the BLUED [22] dataset, exploiting the voltage and current waveforms of the aggregate signals, and reporting the percentage error rate (PER) of the working state prediction of each appliance. The results have shown that the proposed approaches can give better NILM performance than direct subsampling with a NILM algorithm based on the one proposed in [23].

As discussed above, the state-of-the-art highlights the lack of direct evaluations of NILM algorithms with data reduction techniques. Indeed, in event-based approaches [14–20], the focus is mainly on the evaluation of the transient event, with a resolution on the NILM problem using a classification approach on the events detected or by reconstructing the signal at original rate. On the other hand, in the data-compression based approach [21], the goal is to analyse the entire signal and to compress the information exploiting its sparsity. Despite the study presenting a similar aim to the one of the authors' work, its peculiarity lies in the application with a high frequency sampled waveform, with the reconstruction of the signal at the original rate before the application of the NILM algorithm.

The authors' interest here is on the resolution of the NILM problem with an ad hoc data reduction strategy for a residential environment, i.e., where power consumption signals are acquired by means of smart meters, thus signals at sub-Hz frequencies. The work focus is on the application of an algorithm which allows to easily manage various sampling frequencies of the power consumption signal, as well as without a significant loss of performance. Specifically, up to the authors' knowledge, none of the studies have evaluated reduction strategies on the Neural NILM approach, proposed in [24]. Indeed, Neural NILM provides a flexible solution, from the point of view of the subsampling schema, e.g., uniform and non-uniform, which can take advantage of the possibility of characterizing each appliance with a different and dedicated network topology. In addition, this algorithm exploits both the transient and the steady states information for the disaggregation aim, which allows for managing the NILM problem with a higher information level with respect to other approaches. Specifically, the disaggregated outputs carry on the total information of the appliances contribution, i.e., the events occurred and the detailed energy consumption in the observation period.

The outline follows. Problem statement and work motivations are discussed in Section 2. Section 3 proposes an overview of the uniform and non-uniform subsampling strategies, and in Section 3.1 details on the adopted network topology are provided. Section 4 reports the experimental setup and the adopted evaluation methods. Results' discussion and advanced considerations are reported in Section 5. Finally, Section 6 concludes the paper.

2. Problem Statement and Motivations

The consumption power signal of any appliance is generally composed of a set of different working states (including power off/on states), and it can be represented in terms of portions of signal based on the presence of more or less rapid changes. Therefore, generally speaking, the signal portions can be classified in either steady states or transient states [6]. Moreover, it is considered that the analogue power signal is sampled at a specific frequency, denoted as original sampling rate from here on.

Under this assumption, it is natural to consider the possibility to apply an ad hoc subsampling strategy in order to preserve a higher level of detail during the transients, e.g., keep the original sampling rate, whereas they downsample the signal during the steady phases. In fact, aiming to

create models as generic as possible to represent the different typologies of appliances, the adopted information cannot refer to the consumption levels only, i.e., the steady states, but has to exploit the details contained in the transition states as well. In this way, it is possible to guarantee the ability to discriminate the type of appliance (transient-based features widely use to appliance identification [6]), as well as to reduce the overall amount of data processed and transmitted to the elaboration service in compliance with the IoT paradigm. For example, in a real system, the smart meter will store in a buffer the data acquired at the original sampling rate, whereas only the sub-sampled data are continuously sent, after proper pre-processing. Whenever a transition state is automatically detected, the appropriate amount of data in the buffer, at the original sampling rate, will be forwarded to the remote system that performs the disaggregation.

Considering the application of this idea to the NILM paradigm, where the system receives as input the aggregated signal and disaggregates it to produce the signals related to each appliance, two main problems arise. On the one hand, the data acquisition system should be capable of exactly discerning between steady and transient states by analysing the aggregated signal only, in order to modulate the subsampling activity. On the other hand, the disaggregation algorithm should be capable of properly operating with subsampled data (thus reduced information), in order to either reach performance equivalent to the system without subsampling (reference system), or to achieve a marked reduction in the data required by the elaboration system against a modest deterioration in performance.

The state-of-the-art discussed in the previous section confirms the practicability to apply a non-uniform subsampling strategy at the aggregated data, keeping a high level of signal detail. On the other hand, up to the authors' knowledge, the state-of-the-art discussion highlighted a lack of techniques for the data reduction, both based on event-detection or data compression, directly applied and evaluated with NILM algorithms, in order to understand how the data reduction itself can affect the NILM performance. Specifically, neither uniform or non-uniform subsampling approaches have been evaluated in combination with the Neural NILM.

The main objective of this paper is the evaluation of the latter issue by considering the Neural NILM [24] approach as the disaggregation algorithm. In order to validate the ad hoc subsampling strategy, an extended evaluation study is presented, providing a comparison among the application of different sampling rate on different datasets. In particular, the signal processed by the ad hoc subsampling strategy, called non-uniform subsampling (NUS), can be decomposed in a signal uniformly subsampled (thus applying a uniform subsampling—US), and portions of signal sampled at the original sampling rate (OS). Therefore, in order to provide a comprehensive evaluation, it is needed to compare the NUS performance against the one achieved by applying only a US and the performance at the OS. Moreover, the evaluations will also consider the application of different subsampling rates and different lengths of the transition windows for the NUS.

It should be noted that the detection the appliance states, i.e., steady and transient phases of the power signal, lies outside the scope of this work. Specifically, the aim here is to evaluate the NILM algorithm performance in different US and NUS conditions, minimizing external causes of errors, such as the erroneous detection of a state. However, requiring the knowledge of the appliance state, the algorithm relies on the clustering approach presented in [25].

Finally, the overall data reductions achievable with different NUS strategies, in contrast to the US ones, are reported and discussed.

3. NUS in Neural NILM

The aggregate power in a household electrical circuit is composed of contributions due to known loads, i.e., the power consumption related to loads modelled in the NILM algorithm, and to unknown

contributions or noise components, i.e., the power consumption related to unmodeled loads and noise present in the circuit. Specifically, the total active power, $y[n]$, can be expressed as:

$$y[n] = \sum_{i=1}^{N_A} y_i[n] + e[n], \quad (1)$$

where N_A is the number of modelled appliances—known contributions, $y_i[n]$ denotes the active power signal of the i -th appliance, and $e[n]$ is the overall noise component—unknown contributions and circuit noise. The NILM paradigm relies on the extraction of the power consumption of the i -th appliance, $y_i[n]$, from the aggregate power signal, removing the remaining components $n_i[n]$, and can be formulated as:

$$y_i[n] = y[n] - n_i[n], \text{ where } n_i[n] = \sum_{j=1, j \neq i}^{N_A} y_j[n] + e[n]. \quad (2)$$

Under this assumption, a *denoised* condition is expressed as:

$$y_i[n] = y[n] - n_i[n], \text{ where } n_i[n] = \sum_{j=1, j \neq i}^{N_A} y_j[n], \quad (3)$$

where it is assumed that the aggregated power consumption signal is composed of known appliances only, without additional unknown contributions—loads—or sources of noise. Therefore, the noise component $e[n]$ is set equal to zero, and Equation (2) can be reformulated as shown in Equation (3). The assumption of a *denoised* scenario allows for avoiding undesired signals from masking the small signal fluctuations of a specific appliance out of the aggregated signal. Therefore, working in a *denoised* scenario allows the acquisition system to ideally perform an exact detection of the variations in the aggregated signal and discerns the occurrences of steady and transient states. The *noised* scenario, on the other hand, represents a more realistic case study and it has been considered in the experimental evaluation. Assuming that the power signal is acquired by a measurement system capable of providing samples at a sub-Hz sampling rate, the expectation is that it is possible to further “downsample” the signal during the appliance steady states without losing essential information, while the sampling rate is kept unchanged in the transition states, in order to preserve the information. Specifically, the sampling rate reduction in a specific portion of the signal (the steady state) allows for reducing the amount of data to be processed by the algorithm, thus lowering the computational burden. On the other hand, this operation decreases the data transmitted from a smart meter to the disaggregation service host.

As discussed in the previous section, in order to keep the original sampling rate during the transition samples, and to apply the downsampling only in the steady phases, an ad hoc non-uniform subsampling strategy is applied. The proposed US and NUS processing operations are used on the aggregated data, already sampled at the OS rate; therefore, the approach differs from the nonuniform sampling theory applied to analogue signals [26]. Specifically, as depicted in Figure 1, during the subsampling processing, if a transition state is encountered, the algorithm collects a predetermined number of samples at the original sampling rate, creating a selection window centred at the middle of the transition edge, and thus providing a set of equal distributed samples centred on this event. As a result, considering a vector of constant length as network input, depicted in Figure 2, an overall shorter time interval with respect to the case of uniform subsampled points only, i.e., the green ones, will be represented.

The selection window, which marks the signal portion to be acquired at the original sampling rate, is called expansion window, and its length is expressed in terms of samples number at the original rate, called expansion rate (ER). In fact, considering the majority of the signal being subsampled, from the receiver point of view, the signal is expanded/upsampled, providing additional level of details when a transition phase is met.

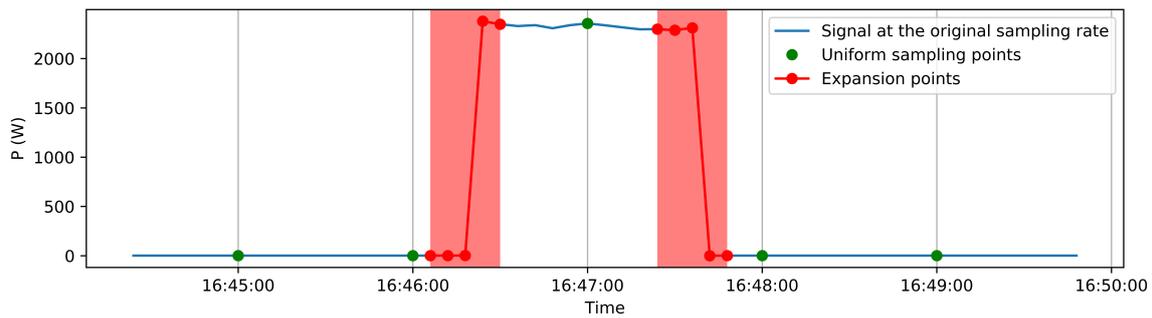


Figure 1. Example of expansion window (red regions) with 1 min subsampling period and 30 s expansion window length (NUS10-5). The NUS vector is composed of green and red points.

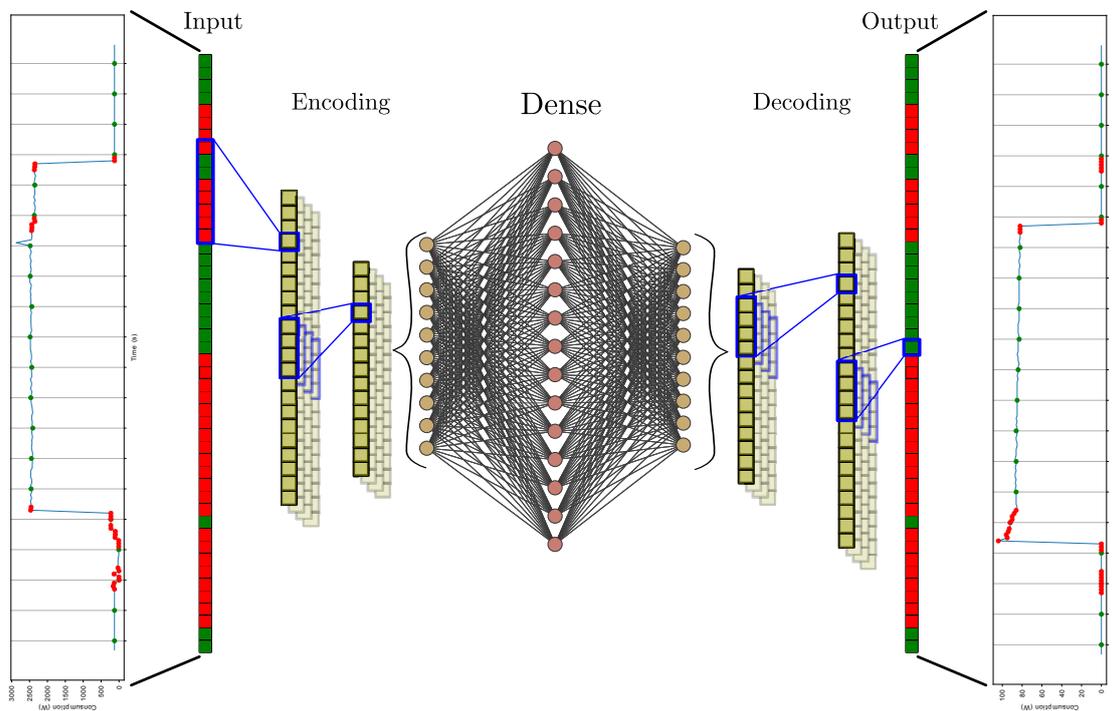


Figure 2. Overview of the base network structure. In the input/output vectors, the *green* points/blocks denote the samples at uniform subsampling, whereas the *red* ones mark the samples within the expansion window.

Together with the selection window, a uniform subsampling is performed for the remaining data, i.e., the steady states. The general subsampling period applied is expressed by means of the parameter subsampling rate (SR) that specifies the multiplication factor applied to the original sampling rate, which corresponds to the sampling period increasing factor.

Denoting with $y[n]$, with $n = 0, \dots, N - 1$ where N is the number of samples, the n -th sample at the original sampling rate for a given active power signal, the vector \mathbf{Y}^{OS} of the original sampling samples is defined as:

$$\mathbf{Y}^{OS} = \{y[i] | i \in \{0, 1, \dots, N - 1\}\}. \tag{4}$$

Accordingly, the vector \mathbf{Y}^{US} of uniform subsampled samples at $SR = C$ is obtained as:

$$\mathbf{Y}^{US} = \left\{ y[i] | i \in \left\{ 0, C, 2C, \dots, \left\lfloor \frac{N-1}{C} \right\rfloor C \right\}, C \in \mathbb{N} \right\}, \tag{5}$$

where $\lfloor x \rfloor$ is the integer part of x . Finally, the vector \mathbf{Y}^{NUS} of non-uniform subsampled samples at $SR = C$ and $ER = G$ is expressed as:

$$\mathbf{Y}^{NUS} = \mathbf{Y}^{US} \cup \left\{ y[i] \mid i = k + w, k \in \mathbf{E}, w \in \left\{ -\left\lfloor \frac{G}{2} \right\rfloor, -\left\lfloor \frac{G}{2} \right\rfloor + 1, \dots, 0, \dots, \left\lfloor \frac{G}{2} \right\rfloor - 1, \left\lfloor \frac{G-1}{2} \right\rfloor \right\}, G \in \mathbb{N} \right\}, \quad (6)$$

where $\mathbf{E} = \{e_0, e_1, \dots, e_{L-1}\}$, with L the number of expansion windows, denotes the set of the indices corresponding to each transient state within the N samples, i.e., the centres of the expansion windows.

The number of elements of \mathbf{Y}^{OS} , \mathbf{Y}^{US} , and \mathbf{Y}^{NUS} are denoted, respectively, as K^{OS} , K^{US} , and K^{NUS} . Essentially, the number of samples at original sampling, K^{OS} , is equal to the total number of samples N , in the US case $K^{US} = \lfloor (N-1)/C \rfloor + 1$, whereas, in the NUS case, K^{NUS} depends on the parameters used in the subsampling procedure.

The detection of transient states is precomputed for each appliance independently, after which all the detections are gathered over a common time base. This information is provided to the pre-processing algorithm which performs the US and/or NUS on the aggregated data and returns the data to be use as network input. In the transients' detection, the working state changes as well as the power-on/off switching activities are taken into account as transient states.

3.1. Neural NILM

The neural network architecture used for load disaggregation uses CNN layers, and it is based on the architecture proposed by Kelly et al. [24], and further studied by the authors [27,28].

As depicted in Figure 2, the structure is based on the auto-encoder topology, where both encoding and decoding portions are composed of convolutional layers concatenated to a linear activation function and a max pooling layer. The encoding stage ends with a fully connected network based on ReLU activation function [29], denoted as dense layer in Figure 2. Symmetrically, in the decoding stage, the upsampling layers replace the max pooling ones.

The introduction of the max pooling operation allows the network to develop an independent behaviour with respect to location of the activation inside the input window. Moreover, a reduction of the features maps size is achieved, with a consequent reduction of the input neurons number in the dense layer. Additionally, in order to respect the constraint of non-negative active power, a ReLU activation has been adopted.

The Stochastic Gradient Descent (SGD) algorithm with Nesterov momentum [30] is used in the training phase, and an early stopping technique is adopted in order to prevent overfitting. The network is trained providing an aggregated signal (window) as input and the corresponding disaggregated signal of a specific appliance as output. The loss between input and output is quantifies in terms of mean squared error (MSE).

In the disaggregation phase, the input provided to the network corresponds to a sliding window portion of the aggregated power signal with a specific stride. Therefore, the output/disaggregated signal has to be properly reconstructed, since it presents more or less overlapping windows based on the stride values. Specifically, the lower the stride, the higher the number of windows overlapped. In order to produce the output signal, the samples are recombined by calculating a mean or a median operation at each time instant. Both operations are evaluated because they both present shortcomings. In the case of the mean operation, averaging the overlapped portions could produce an overall underestimated signal. On the contrary, the median operation produces a better estimate, erasing the outliers that are typically near to zero, only in the case of reliable samples statistic.

The approach based on denoising autoencoders used here for disaggregation employs an individual neural network for each appliance of interest [24]. As will be evident in the experimental evaluation (Section 5), the best performing subsampling method can be appliance-dependent, and using an individual network for each appliance allows for combining the US and NUS techniques. An example of this solution applied to the fridge and dishwasher appliances is depicted in Figure 3. In the example, the fridge network takes as input the US samples (the green points in Figure 3), while

the dishwasher network uses the NUS samples, i.e., both the green and the red points in Figure 3. Clearly, each network provides an estimate only for the samples at its input. Assuming that the most performing subsampling method for the fridge is US and for the dishwasher is NUS, this solution allows for achieving a higher disaggregation performance compared to the use of a single technique.

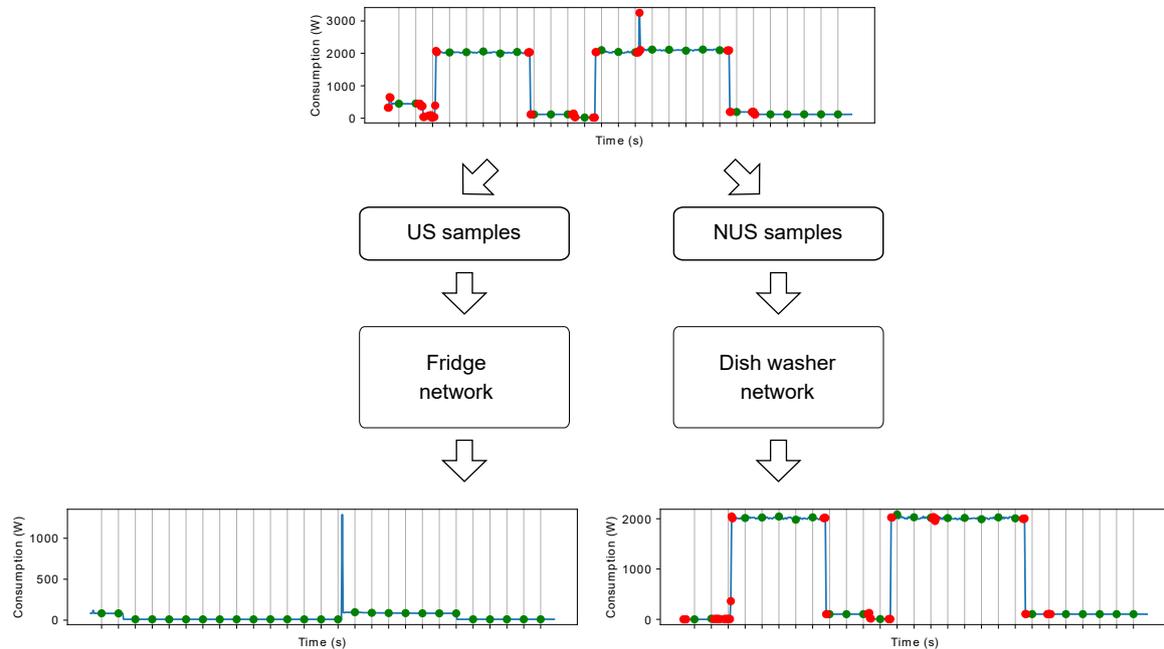


Figure 3. An example scheme of the combination of the NUS and US methods for the fridge and dishwasher appliances.

4. Experimental Setup

The experiments have been conducted on active power signals contained in the UK-DALE [31] and the REDD [11] datasets both in the *denoised* and *noised* conditions. The *denoised* condition has been considered since it allows for evaluating the algorithms avoiding undesired signals from masking the small signal fluctuations of a specific appliance out of the aggregated signal. On the other hand, the *noised* condition represents the common situation encountered in real application scenarios.

Train and test data are extracted using the time intervals adopted in [28] for the UK-DALE dataset, Table 1, and the ones assumed in [27] in the case of REDD dataset, Table 2. Specifically, within the given intervals, the activations and the corresponding *silence* are extracted for each appliance. The first 20% of the activations are adopted as validation data, whereas the remaining 80% are combined with *silence* to compose the batches for the model training. A mean and variance normalization is applied to each batch, whose parameters (mean and variance) are computed from a random sample of the training set. Differently, a min-max normalization is performed on the target data, adopting the maximum power consumption for the corresponding appliance.

Table 1. Details of train and test time intervals for UK-DALE dataset.

Building	Train		Test	
	Interval	Days	Interval	Days
1	12 April 2013–21 October 2014	555	22 October 2014–15 December 2014	56
2	22 May 2013–26 September 2013	128	27 September 2013–10 October 2013	13
5	29 June 2014–01 September 2014	64	01 September 2014–07 September 2014	6

Table 2. Details of train and test time intervals for REDD dataset.

Build.	Train		Test	
	Interval	Days	Interval	Days
1	18 April 2011–20 May 2011	31	21 May 2011–24 May 2011	3
2	17 April 2011–29 April 2011	11	30 April 2011–01 May 2011	1
3	16 April 2011–28 April 2011	11	21 May 2011–27 May 2011	6

The availability of signals acquired in several buildings allows the evaluation of the algorithm performance in *seen* and *unseen* conditions. For each appliance, in the case of *seen* condition, two buildings are used both for training and testing, whereas, in the case of *unseen* condition, the model is trained on the same data used in the *seen* condition, then tested over the data related to a different building. The number of buildings used for each dataset has been limited according to the specifications of [27].

The experiments presented in the following sections are conducted both in US and NUS conditions, as well as at OS. The sampling rate adopted as OS rate is 6 s. In particular, the UK-DALE dataset already provides data at this rate, whereas the REDD dataset has been downsampled. In the case of US, a fixed, reduced sampling rate is applied at the whole data without discrimination of the appliance working state. On the other hand, in case of NUS, both subsampling and expansion window, as discussed above, are applied. As a general notation, for the experiments conducted with US, the subsampling rate is expressed with a number concatenated to the string “US”, for example with a SR = 10 the experiment is denoted as US10. In the case of NUS, both SR and ER are concatenated to the string “NUS” and separated from each other by the symbol “-”, i.e., for SR = 10 and ER = 10, the NUS experiment is marked as NUS10-10.

As discussed above, the NUS experiments require the knowledge of all the possible state transitions of all the appliance in the aggregated signal. However, being the scope of the work to provide an extensive validation of the NUS idea, the evaluations have been performed by assuming an a priori knowledge of the transition points (oracle scenario), therefore minimizing the occurrence of errors (expansion windows badly positioned) due to wrong estimations of the state transition points. A detection as accurate as possible of the state changes has been carried out by processing each appliance separately. The information generated for each appliance, composed of pairs timestamp—state label, is then combined in an aggregated ground-truth. The process relies on the estimation of the different power levels, i.e., working states, by applying a k-means algorithm [32]. At first, the clustering procedure is executed over the whole train set by setting a predefined number of clusters for the appliance to be evaluated. Gaussian variables, mean and variance, are inferred for each cluster, then exploited in the state classification over the test set. At this point, to reveal the transition instants, the difference between each label and the previous one is performed over the output time series that contains the detected states. More details concerning the whole procedure are available in [25].

The hyperparameters of the neural networks have been determined by conducting a grid search separately for the two datasets. This procedure allows for determining the most performing configurations, that, however, are specific for the target dataset. This represents a general problem for neural networks-based algorithms [33,34], and here it has not been taken into account, since the main focus of this paper is the evaluation of different sampling strategies regardless the network topology.

Regarding the UK-DALE dataset, the experiments have been performed by conducting a grid search within the following sets: $[1 \times 4, 1 \times 16, 1 \times 32]$ for the kernels dimension, $[32, 128]$ for the number of features maps, $[2, 4]$ for the pool size of the pooling layer, $[512, 4096]$ for the number of neurons in the dense layer. Moreover, 2 CNN layers have been adopted in both encoding and decoding stages: in the former stage, the number of feature maps of the second layer is twice the number of feature maps in the first layer, while it is the opposite for the decoder. In the case of REDD

dataset, after a preliminary evaluation, only one CNN layer has been taken into account, whereas the grid search has been performed over a wider set for each parameters with respect to the UK-DALE. Specifically, the set $[1 \times 8, 1 \times 16, 1 \times 32, 1 \times 128]$ has been assumed for the kernels dimension, $[8, 16, 32, 128]$ for the number of features' maps, $[1, 2, 4]$ for the pool size of the pooling layer, and $[128, 512, 4096]$ for the number of neurons in the dense layer.

For both datasets, in the pooling layers, the pooling stride has been fixed to 1, the max epochs have been set to 20,000, the validation in train has been performed every 10 epochs, the early stopping condition has been evaluated every 2000 validations (thus 20,000 epochs), the batch size is composed of 64 sequences, and the adaptive learning rate has started from 0.1 and reduced by a factor of 10 if the improvement in validation is lower than 0.01.

For each appliance, thus for each dataset, the length of the input window has been set equal to the one in the reference work [27] in the case of NUS experiments, whereas, in the case of US, the window length is reduced proportionally to the SR factor.

Additionally, the tests for each possible configuration, both network parameters and US/NUS approaches, have been executed adopting different strides of the sliding window (or hop size) of the input data and two different types of data reconstruction: $[1, 8, 16, 32]$ and $[\text{mean}, \text{median}]$, respectively. Therefore, given nine different sampling configurations (OS, US5, US10, US20, NUS5-5, NUS10-5, NUS10-10, NUS20-10 and NUS20-20) for the UK-DALE dataset a total of 1080 models have been generated, whereas 5184 models for the REDD dataset. Specifically, considering the different sliding window strides and reconstruction techniques, an overall total of 8640 and 41,472 evaluations have been carried out in the case of UK-DALE and REDD datasets, respectively. Furthermore, in contrast with [27,28], the *data augmentation* procedure, proposed in [24], has not been adopted during the generation of the batches. Therefore, a comparison against the achieved OS results cannot be carried out.

The experiments have been performed exploiting both a local cluster and an HPC resource. Specifically, the former is composed of two PCs equipped respectively with Intel i7-4930K@3.40 GHz, 32 GB RAM, GTX TITAN X 12 GB, GTX 1080 8 GB, and Intel i7-6850K@3.60 GHz, 32 GB RAM, TITAN X (Pascal) 12 GB, TITAN Xp 12 GB. The latter system is the HPC GALILEO at CINECA, and the experiments have relied on a maximum of 6 nodes, each one equipped with 2x8-cores Intel Haswell@2.40 GHz, 128 GB RAM, and 2 Nvidia K80 GPUs.

The project is developed in Python and the neural network is based on the Keras library [35] using the TensorFlow backend [36]. All the code is publicly available [37].

Evaluation Methods

The performance has been evaluated by relying on the metrics proposed in [38] specific to energy disaggregation. In particular, for the i -th appliance, with $i = 1, \dots, N_A$ and N_A number of appliances, $\hat{y}_i[k]$ denotes the disaggregated power signal, $y_i[k]$ is the ground-truth power signal, and K is the overall samples number, the energy-based precision and recall are defined as:

$$P_i = \frac{\sum_{k=0}^{K-1} \min(\hat{y}_i[k], y_i[k])}{\sum_{k=0}^{K-1} \hat{y}_i[k]}, \quad R_i = \frac{\sum_{k=0}^{K-1} \min(\hat{y}_i[k], y_i[k])}{\sum_{k=0}^{K-1} y_i[k]}. \quad (7)$$

Information about the power consumption that has been correctly classified is given by the recall, whereas information about the power correctly assigned to an appliance is given by the precision. The F_1 -score is a geometric mean between precision and recall, and is given as (for the i -th appliance):

$$F_{1i} = 2 \frac{P_i \cdot R_i}{P_i + R_i}. \quad (8)$$

Finally, once the precision and the recall of each appliance are computed, the averaged values of precisions and recalls over the appliances are exploited to compute the overall F_1 -score.

A first set of evaluations has been conducted by comparing the outputs generated for each possible configuration of OS, US, and NUS, against the corresponding ground-truth, where also the ground-truth has been re-sampled accordingly. This evaluation will be denoted as *specific-rate* evaluation from here on.

In this configuration, the metrics in Equation (7) are calculated using the signal samples taken from the related samples vector:

$$y_i[k] \in \mathbf{Y}_i^{OS/US/NUS}, \quad \hat{y}_i[k] \in \hat{\mathbf{Y}}_i^{OS/US/NUS}, \quad (9)$$

$$\text{with } k = 0, 1, \dots, K^{OS/US/NUS} - 1.$$

Moreover, a so-called *max-rate* evaluation has been performed, in order to produce evaluations for each tested configuration by reporting the data to the condition of the original sampling. Specifically, the output generated by each US configuration has been up-sampled to the original sampling rate (that is the higher maximum frequency for the signal) by applying the zero-insertion followed by an interpolation filter. In the case of NUS, the same upsampling procedure has been performed over the uniform sampling points (the ones acquired applying the SR parameters), then the points within an expansion window (therefore at original sampling rate) will replace the interpolated data within the same temporal window. In this configuration, the metrics in Equation (7) are calculated using the signal samples taken from the OS samples vector:

$$y_i[k] \in \mathbf{Y}_i^{OS}, \quad \hat{y}_i[k] \in \hat{\mathbf{Y}}_i^{OS}, \quad (10)$$

$$\text{with } k = 0, 1, \dots, N - 1.$$

5. Experimental Results

5.1. Data Reduction Evaluation

The application of US and NUS approaches allows for achieving an overall reduction of the elaborated data. Detailed information about the reductions achieved for each configuration is reported in Tables 3 and 4, in the case of UK-DALE and REDD datasets, respectively.

Table 3. Details of data reduction achieved at different configurations of US and NUS with UK-DALE dataset.

Sampling Method	Building 1		Building 2		Building 5		Overall	
	Samples	$\Delta\%$	Samples	$\Delta\%$	Samples	$\Delta\%$	Samples	$\Delta\%$
OS	8,798,999	100%	2,016,000	100%	997,561	100%	11,812,560	100%
US5	1,759,800	20%	403,200	20%	199,512	20%	2,362,512	20%
US10	879,900	10%	201,600	10%	99,756	10%	1,181,256	10%
US20	439,950	5%	100,800	5%	49,878	5%	590,628	5%
NUS5-5	2,178,932	24.76%	457,193	22.68%	237,281	24.79%	2,873,406	24.33%
NUS10-5	1,351,455	15.36%	262,420	13.02%	142,247	14.26%	1,756,122	14.87%
NUS10-10	1,604,546	18.24%	308,998	15.33%	166,290	16.67%	2,079,834	17.61%
NUS20-10	1,204,890	13.69%	214,125	10.62%	120,083	12.04%	1,539,098	13.03%
NUS20-20	1,616,211	18.37%	305,631	16.16%	158,487	15.89%	2,080,329	17.61%

Table 4. Details of data reduction achieved at different configurations of US and NUS with REDD dataset.

Sampling Method	Building 1		Building 2		Building 3		Overall	
	Samples	$\Delta\%$	Samples	$\Delta\%$	Samples	$\Delta\%$	Samples	$\Delta\%$
OS	498,378	100%	171,884	100%	252,266	100%	922,528	100%
US5	99,676	20%	34,377	20%	50,453	20%	184,506	20%
US10	49,838	10%	17,188	10%	25,227	10%	92,253	10%
US20	24,919	5%	8594	5%	12,613	5%	46,126	5%
NUS5-5	111,978	22.47%	37,547	21.84%	58,288	23.11%	207,813	22.53%
NUS10-5	63,687	12.78%	20,758	12.08%	34,033	13.49%	11,8478	12.84%
NUS10-10	74,597	14.97%	24,074	14.01%	42,067	16.68%	140,738	15.26%
NUS20-10	51,084	10.25%	15,862	9.23%	30,390	12.05%	97,336	10.55%
NUS20-20	70,338	14.11%	22,491	13.086%	44,405	17.60%	137,234	14.88%

In the case of US configurations, regardless the dataset, data reduction is directly proportional to the SR parameter, whereas, regarding the combination of US+NUS, the achieved reduction will always be equal to the NUS one. In particular, the measurement system (smart meter) has to collect and to forward data to the NILM service at the OS sampling rate whatever a state change is detected, since the measurement system is not aware of which network relies on the US or NUS sampling scheme.

In the case of NUS configurations, the reduction is not directly proportional to the SR factor because additional samples (expansion windows) at OS are taken into account. However, the NUS approach allows for reaching an overall reduction between 24.33% and 13.03% for the UK-DALE dataset, and between 22.53% and 10.55% for the REDD dataset. Specifically, considering the best combinations for the UK-DALE dataset in both *seen* and *unseen* scenarios, US5+NUS5-5 and US20+NUS20-10, only the 24.33% and 13.03% of the original data is used, respectively. In the case of REDD, both *seen* and *unseen* scenario reach the maximum performance with US5+NUS5-5, thus using only 22.53% of the overall data.

Moreover, the application of the NUS by keeping an appliance window constant at the OS produces some side effects that could negatively affect the performance. The first side effect, with a minor impact on the performance, regards the reduction of the activations number usable by the algorithm, due to the increase of the SR and ER parameters. Specifically, in the train phase of a specific appliance, since each input window of aggregated signal is associated with the corresponding activation, if multiple activations appear within the same input window, one is discarded. Therefore, the SR and ER values' increase shrinks the activations together, increasing the discarding rate. The latter effect concerns the reduction of the *silences* number for a specific appliance, which is the temporal window without the footprint of the specific appliance in both input-aggregated and output-disaggregated data. As discussed above, the shrink of activations due to NUS will also reduce the available samples of silence among the activations. As shown in the next section, the fridge is heavily affected by this problem, since it is characterized by a long activation and a great number of activations.

5.2. Denoised Scenario

5.2.1. UK-DALE

In the case of UK-DALE dataset, the appliances taken into account in the experiments are the dishwasher, fridge, kettle, microwave, and washing machine, i.e., those with the major energy contribution.

Concerning the *seen* scenario, the results in Table 5 show that the lower reduction in the overall performance, considering *specific-rate* evaluations, is achieved by a subsampling rate of 5 for the

US configurations, and by a subsampling rate of 5 with an expansion window of 5 in case of NUS. Respectively, the US5 configuration achieves an overall degradation of just 1.9% with respect to OS, whereas the NUS configuration presents a significant degradation of about 11.5%. However, US10, NUS10-10 and NUS20-20 have a performance deterioration within the 3% from the best reference value.

Table 5. UK-DALE *specific-rate* evaluations (F_1 -score) for the different configurations in *seen* scenario. Appliance best score and overall best score are highlighted.

Sampling Method	Appliance					Overall
	Dishwasher	Fridge	Kettle	Microwave	Washing Machine	
OS	0.704	0.823	0.902	0.757	0.438	0.739
US5	0.641	0.825	0.886	0.755	0.439	0.726
US10	0.619	0.808	0.780	0.744	0.448	0.693
US20	0.608	0.799	0.529	0.652	0.426	0.608
NUS5-5	0.878	0.305	0.851	0.578	0.488	0.654
NUS10-5	0.804	0.229	0.824	0.505	0.488	0.607
NUS10-10	0.812	0.259	0.857	0.556	0.507	0.631
NUS20-10	0.762	0.238	0.845	0.531	0.538	0.614
NUS20-20	0.813	0.279	0.889	0.583	0.513	0.643

From Table 5, it is possible to note that the F_1 -score of some of the appliances improves. In the case of US5, the fridge and washing machine achieve a relative improvement of about 0.2%, whereas for US10, the washing machine reaches a 2.3% improvement.

In the case of NUS, the dishwasher presents a marked improvement for all the configurations, as well as the washing machine, with a maximum of 24.9% at NUS5-5 and 23.0% at NUS20-10, respectively. The kettle performs slightly worse than the OS, whereas both fridge and microwave seem to suffer of the NUS. As mentioned in the previous section, a possible cause of bad performance of the fridge is the shrinking of the activations that also reduces the number of silence samples among activations. Specifically, in the normal condition, the fridge presents 587 *silences*, whereas 0 *silences* are detected in the case of NUS10-10.

Therefore, as shown in Table 6 combining the US and NUS methods, for the same subsampling ratio, it is possible to achieve a better overall disaggregation behaviour. Specifically, for US5+NUS5-5 and US10+NUS10-10, the combinations allow for reaching higher performance with respect to the OS, with an increase of about 4.6% and 1.5%, respectively. Despite the other combination resulting in a performance degradation, the reduction is reduced, from 0.2% to 3%.

Table 6. UK-DALE *specific-rate* evaluations (F_1 -score) with combination of US and NUS configurations in *seen* scenario. Appliance best score and overall best score are highlighted.

Sampling Method	Appliance					Overall
	Dishwasher	Fridge	Kettle	Microwave	Washing Machine	
OS	0.704	0.823	0.902	0.757	0.438	0.739
US5+NUS5-5	0.878_n	0.825_u	0.886 _u	0.755 _u	0.488 _n	0.773
US10+NUS10-5	0.804 _n	0.808 _u	0.824 _n	0.744 _u	0.488 _n	0.738
US10+NUS10-10	0.812 _n	0.808 _u	0.857 _n	0.744 _u	0.507 _n	0.751
US20+NUS20-10	0.762 _n	0.799 _u	0.845 _n	0.652 _u	0.538_n	0.721
US20+NUS20-20	0.813 _n	0.799 _u	0.889 _n	0.652 _u	0.513 _n	0.738

_u: US applied; _n: NUS applied.

The evaluations adopting the *max-rate* approach executed for the combinations of US and NUS are reported in Table 7. These results confirm that, despite the disadvantageous conditions,

the US5+NUS5-5 allows for reaching good performance. Specifically, the same results of OS are achieved, with the exception of US20+NUS20-10, which reaches a degradation of 11.3%. For the remaining configurations, the degradation is from 5% to 6.8%. In order to provide a better insight, a comparison between OS and best combination US+NUS is reported in Figure 4.

Table 7. UK-DALE *max-rate* evaluations (F_1 -score) with combination of US and NUS configurations in *seen* scenario. Appliance best score and overall best score are highlighted.

Appliance						
Sampling Method	Dishwasher	Fridge	Kettle	Microwave	Washing Machine	Overall
OS	0.704	0.823	0.902	0.757	0.438	0.739
US5+NUS5-5	0.880_n	0.816 _u	0.803 _u	0.656 _u	0.507 _n	0.739
US10+NUS10-5	0.825 _n	0.788 _u	0.748 _n	0.567 _u	0.481 _n	0.689
US10+NUS10-10	0.816 _n	0.788 _u	0.745 _n	0.597 _n	0.515 _n	0.702
US20+NUS20-10	0.770 _n	0.757 _u	0.676 _n	0.524 _n	0.517_n	0.656
US20+NUS20-20	0.797 _n	0.757 _u	0.721 _n	0.608 _n	0.515 _n	0.690

u: US applied; *n*: NUS applied.

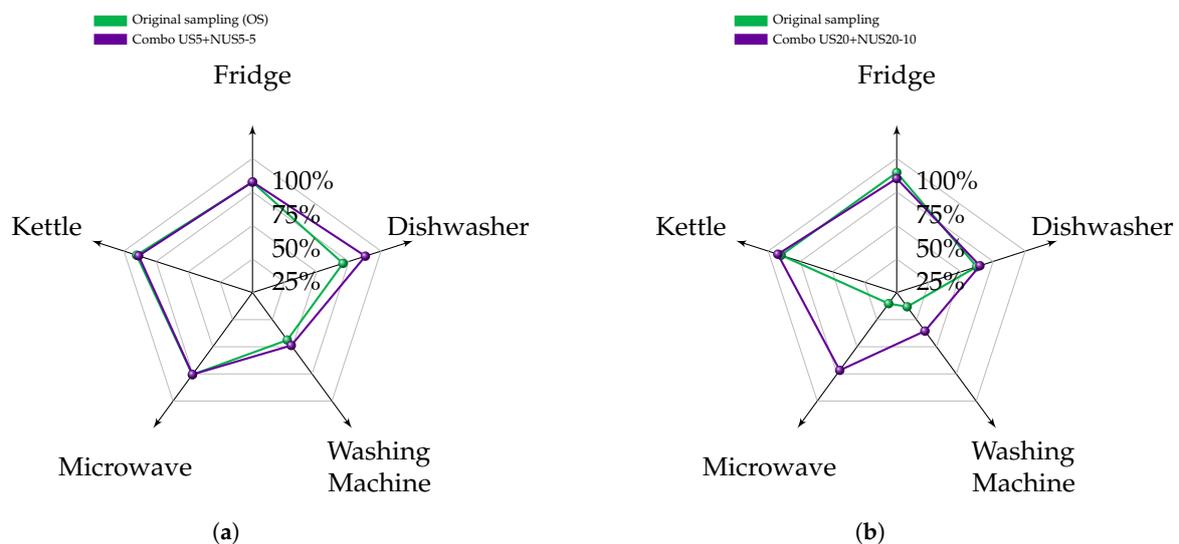


Figure 4. Comparison of disaggregation performance at *specific-rate* for UK-DALE dataset in *seen* (a) and *unseen* (b) scenario.

For the *unseen* scenario, as shown in Table 8, the US configurations reach the worst performance with respect to OS. In particular, the minimum degradation is achieved with US5, about 2.9%, whereas US20 reaches about 14.3% of performance reduction. On the contrary, the NUS20-10 performs better than OS, achieving an improvement of about 6.8%. The remaining NUS configurations show a degradation from 3.9% to 8.5%.

However, in the case of US, both microwave and fridge show a good performance improvement, instead of the marked deterioration of the dishwasher, which falls from 18.6% to 23.3%. On the contrary, the NUS configurations report a marked improvement in the dishwasher performance, as well as for microwave and washing machine, but a severe degradation for the fridge, from 63.0% to 75.5%.

Therefore, adopting the combination of US and NUS, all the configurations outperform the OS, as reported in Table 9. Specifically, US20+NUS20-10 reaches an improvement of about 27.6%, and the remaining combinations show improvements from 12.5% to 14.5% with respect to the OS.

Table 8. UK-DALE *specific-rate* evaluations (F_1 -score) for the different configurations in *unseen* scenario. Appliance best score and overall best score are highlighted.

Sampling Method	Appliance					Overall
	Dishwasher	Fridge	Kettle	Microwave	Washing Machine	
OS	0.622	0.895	0.900	0.102	0.131	0.559
US5	0.487	0.871	0.894	0.121	0.167	0.543
US10	0.506	0.861	0.805	0.120	0.123	0.509
US20	0.477	0.850	0.521	0.120	0.248	0.479
NUS5-5	0.820	0.332	0.904	0.131	0.284	0.537
NUS10-5	0.762	0.251	0.892	0.173	0.310	0.513
NUS10-10	0.784	0.257	0.925	0.199	0.237	0.527
NUS20-10	0.647	0.220	0.925	0.717	0.355	0.597
NUS20-20	0.660	0.244	0.941	0.267	0.314	0.511

Table 9. UK-DALE *specific-rate* evaluations (F_1 -score) with combination of US and NUS configurations in *unseen* scenario. Appliance best score and overall best score are highlighted.

Sampling Method	Appliance					Overall
	Dishwasher	Fridge	Kettle	Microwave	Washing Machine	
OS	0.622	0.895	0.900	0.102	0.131	0.559
US5+NUS5-5	0.820_n	0.871 _u	0.904 _n	0.131 _n	0.284 _n	0.639
US10+NUS10-5	0.762 _n	0.861 _u	0.892 _n	0.173 _n	0.310 _n	0.630
US10+NUS10-10	0.784 _n	0.861 _u	0.925 _n	0.199 _n	0.237 _n	0.640
US20+NUS20-10	0.647 _n	0.850 _u	0.925 _n	0.717_n	0.355_n	0.713
US20+NUS20-20	0.660 _n	0.850 _u	0.941_n	0.267 _n	0.314 _n	0.629

u: US applied; *n*: NUS applied.

The evaluations adopting the *max-rate* approach executed for the combinations of US and NUS are reported in Table 10. These results confirm that, despite the disadvantageous conditions, the US20+NUS20-10 reaches good performance. Specifically, an increase of about 22.6% is achieved with respect to OS, as well as the remaining combinations, with an overall improvement from 7.0% to 12.0%.

Table 10. UK-DALE *max-rate* evaluations (F_1 -score) with combination of US and NUS configurations in *unseen* scenario. Appliance best score and overall best score are highlighted.

Sampling Method	Appliance					Overall
	Dishwasher	Fridge	Kettle	Microwave	Washing Machine	
OS	0.622	0.895	0.900	0.102	0.131	0.559
US5+NUS5-5	0.882_n	0.862 _u	0.788 _n	0.121 _u	0.292 _n	0.626
US10+NUS10-5	0.853 _n	0.842 _u	0.730 _n	0.115 _u	0.301 _n	0.598
US10+NUS10-10	0.865 _n	0.842 _u	0.737 _n	0.115 _u	0.272 _n	0.602
US20+NUS20-10	0.752 _n	0.810 _u	0.636 _n	0.793_n	0.409_n	0.685
US20+NUS20-20	0.786 _n	0.810 _u	0.694 _n	0.115 _u	0.423 _n	0.587

u: US applied; *n*: NUS applied.

These results show that, in both *seen/unseen* scenario, the sampling reduction allows an overall increment of the disaggregation performance.

The performance differences of each sampling method can be analysed more in detail by observing the related disaggregated profiles. Considering the seen scenario and Building 1, the disaggregated profiles of the best performing combo parametrisation is US5+NUS5-5.

The disaggregated profiles of the dishwasher are represented in Figure 5. The appliance behaviour is distinctive, with multiple working state changes and steady state with long permanence; therefore, the activation is well represented by exploiting the non-uniform sampling scheme. Indeed, the NUS5-5 provides a more accurate appliance profile, in particular during the transient phases, and it avoids low performance in the steady states that could result from the recombination of the outputs in the overlapped portion. Figure 6 shows the disaggregated profiles of the microwave. In this case, the activations in the dataset present a high variability in duration and shape; therefore, the analysis of transient phases only of the activation could remove critical information. Indeed, the NUS5-5 method reproduces an inaccurate profile, as it does not reach the peak value of the consumption. In this case, the uniform sub-sampling scheme provides a good reproduction of the activations, but the reduction of the sampling frequency implicates a loss of details, which are more relevant in short activations, with respect to other appliances with longer activation. For those reasons, the sampling scheme which gives the most accurate reconstructed profiles, with a consequent higher performance, is the original sampling method.

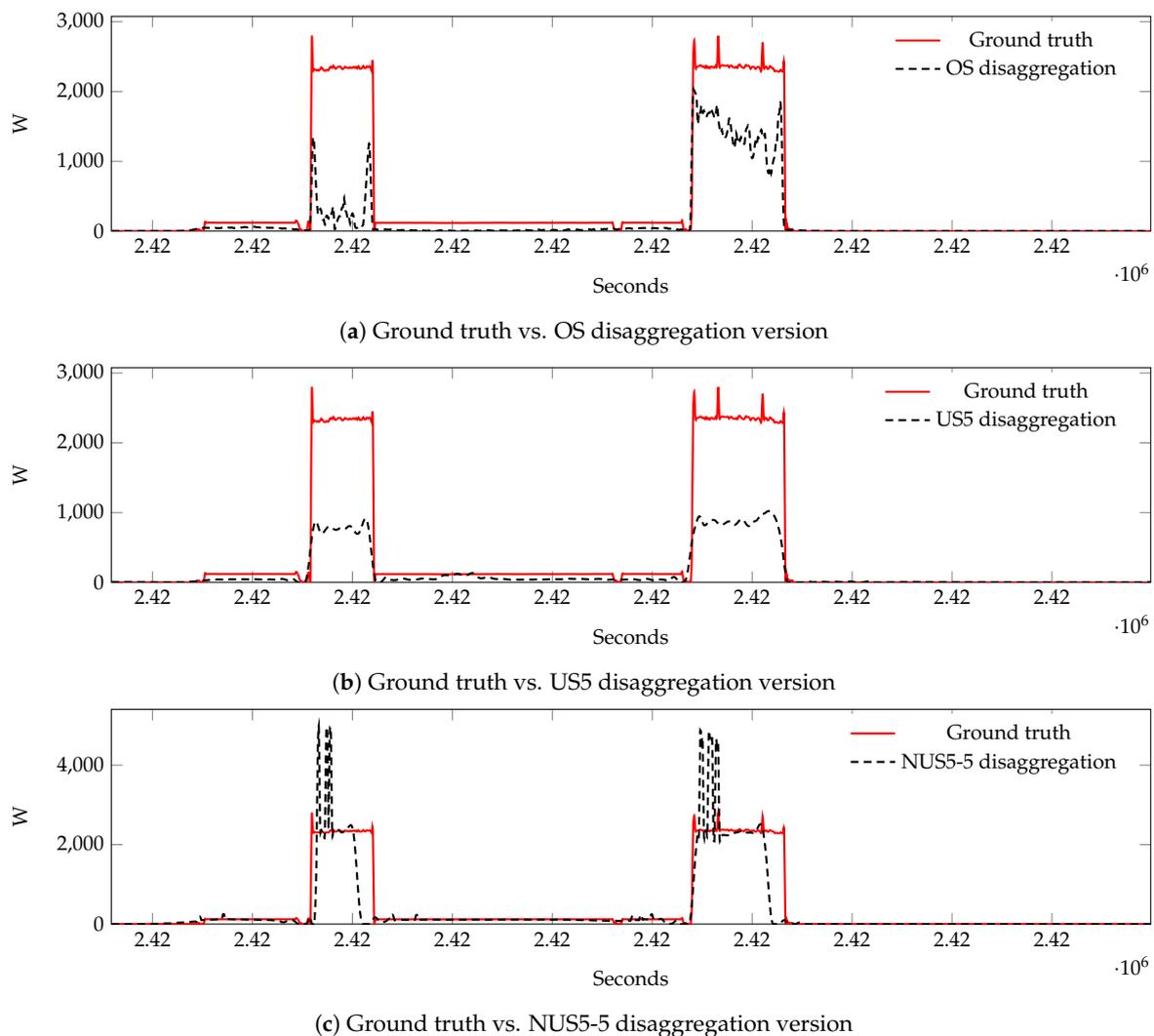


Figure 5. The ground truth and the disaggregated profile of the dishwasher in the UK-DALE dataset, Building 1, using the OS, US5, and NUS5-5 methods. The best disaggregation performance for the dishwasher is provided by the NUS5-5 configuration.

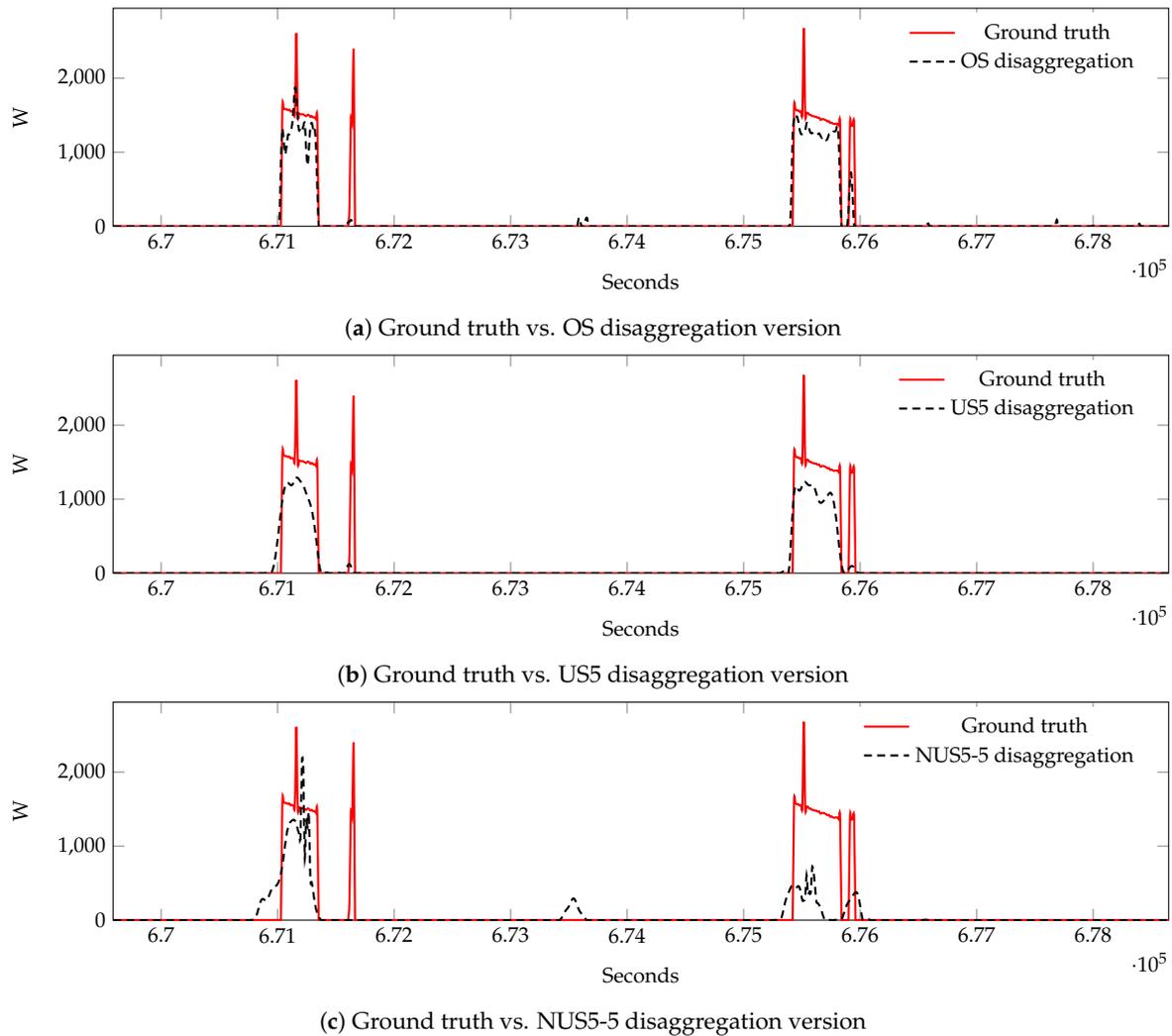


Figure 6. The ground truth and the disaggregated profile of the microwave in the UK-DALE dataset, Building 1, using the OS, US5, and NUS5-5 methods. The best disaggregation performance for the microwave is provided by the OS configuration.

5.2.2. REDD

In the case of REDD dataset, the appliances taken into account in the experiments are the dishwasher, fridge, microwave, and washer dryer, i.e., those with the major energy contribution.

The results achieved on the REDD dataset in the *seen* scenario are reported in Table 11. First of all, the sampling reduction in the case of US20 restricts too much the microwave window, therefore no valid inputs/outputs have been produced.

In the case of US, the best configuration is US5 and presents a slight degradation, about 1.5%, with respect to the OS, whereas the NUS configurations achieved worse degradations. However, the dishwasher achieves better performance in NUS configurations, with respect to both OS and US, by reaching the maximum for NUS5-5, as well as the microwave, whereas washer dryer presents better performance in all the US configurations than NUS ones. As a result, in the combinations US+NUS, all the configurations outperform the OS. Specifically, the best result is achieved by US5+NUS5+5, with a performance improvement of about 13.0%, and the remaining configurations present overall improvements from 3.5% to 10.7%. In order to provide a better insight, a comparison between OS and best combination US+NUS is reported in Figure 7.

Table 11. REDD *specific-rate* evaluations (F_1 -score) for the different configurations in *seen* scenario. Appliance best score and overall best score are highlighted.

Appliance					
Sampling Method	Dishwasher	Fridge	Microwave	Washer Dryer	Overall
OS	0.489	0.636	0.600	0.708	0.614
US5	0.510	0.555	0.519	0.796	0.605
US10	0.497	0.527	0.382	0.678	0.559
US20 (not valid)	0.530	0.487	-	0.666	0.580
NUS5-5	0.768	0.588	0.589	0.258	0.565
NUS10-5	0.726	0.442	0.582	0.217	0.522
NUS10-10	0.718	0.473	0.670	0.328	0.559
NUS20-10	0.624	0.381	0.674	0.278	0.515
NUS20-20	0.659	0.425	0.727	0.397	0.561
Combo US+NUS					
US5+NUS5-5	0.768 _n	0.588 _n	0.589 _n	0.796 _u	0.694
US10+NUS10-5	0.726 _n	0.527 _u	0.582 _n	0.678 _u	0.663
US10+NUS10-10	0.718 _n	0.527 _u	0.670 _n	0.678 _u	0.680
US20+NUS20-10	0.624 _n	0.487 _u	0.674 _n	0.666 _u	0.635
US20+NUS20-20	0.659 _n	0.487 _u	0.727 _n	0.666 _u	0.649

_u: US applied; _n: NUS applied.

Also evaluating the US+NUS combinations at the *max-rate* condition, Table 12, almost all configurations outperform the OS, with improvements from 1.3% to 13.0%, confirming the US5+NUS5-5 as best combination.

Table 12. REDD *max-rate* evaluations (F_1 -score) for the different configurations in *seen* scenario. Appliance best score and overall best score are highlighted.

Appliance					
Sampling Method	Dishwasher	Fridge	Microwave	Washer Dryer	Overall
OS	0.489	0.636	0.600	0.708	0.614
US5+NUS5-5	0.803_n	0.615 _n	0.564 _n	0.762_u	0.694
US10+NUS10-5	0.773 _n	0.520 _u	0.522 _n	0.611 _u	0.637
US10+NUS10-10	0.774 _n	0.520 _u	0.582 _n	0.611 _u	0.649
US20+NUS20-10	0.774 _n	0.502 _n	0.582 _n	0.563 _u	0.622
US20+NUS20-20	0.662 _n	0.473 _u	0.519 _n	0.563 _u	0.572

_u: US applied; _n: NUS applied.

The results achieved by the REDD dataset in the *unseen* scenario are reported in Table 13. First of all, the sampling reduction in the case of US20 restricts too much the microwave window, therefore no valid inputs/outputs have been produced. The best configuration for US, US5 perform a slight degradation, about 4.2% with respect to OS, whereas a worse degradation is achieved by US10, 14.3%. In the overall performance, the NUS configurations present degradation from 15.1% to 32.7%.

As for the *seen* case, the dishwasher achieves better performance in NUS configurations, with respect both to OS and US, reaching the max for NUS5-5; in addition, the microwave performs better in NUS with respect to US, whereas washer dryer presents better performance in US configurations than in NUS ones. As a result, in the combinations US+NUS, the US5+NUS5-5 outperforms OS with an improvement of about 3.2%, whereas US10+NUS10-5 and US10+NUS10-10 produce performance close to OS, with a slight degradation of about 2.2% and 0.4%, respectively.

Table 13. REDD *specific-rate* evaluations (F_1 -score) for the different configurations in *unseen* scenario. Appliance best score and overall best score are highlighted.

Appliance					
Sampling Method	Dishwasher	Fridge	Microwave	Washer Dryer	Overall
OS	0.243	0.809	0.671	0.755	0.646
US5	0.323	0.761	0.619	0.731	0.619
US10	0.323	0.723	0.424	0.705	0.554
US20 (not valid)	0.143	0.614	-	0.656	0.475
NUS5-5	0.535	0.671	0.621	0.358	0.549
NUS10-5	0.514	0.440	0.568	0.339	0.474
NUS10-10	0.476	0.515	0.662	0.398	0.514
NUS20-10	0.354	0.401	0.633	0.345	0.435
NUS20-20	0.444	0.472	0.676	0.440	0.512
Combo US+NUS					
US5+NUS5-5	0.535 _n	0.761 _u	0.621 _n	0.731 _u	0.666
US10+NUS10-5	0.514 _n	0.723 _u	0.568 _n	0.705 _u	0.632
US10+NUS10-10	0.476 _n	0.723 _u	0.662 _n	0.705 _u	0.644
US20+NUS20-10	0.354 _n	0.614 _u	0.633 _n	0.656 _u	0.566
US20+NUS20-20	0.444 _n	0.614 _u	0.676 _n	0.656 _u	0.599

u: US applied; *n*: NUS applied.

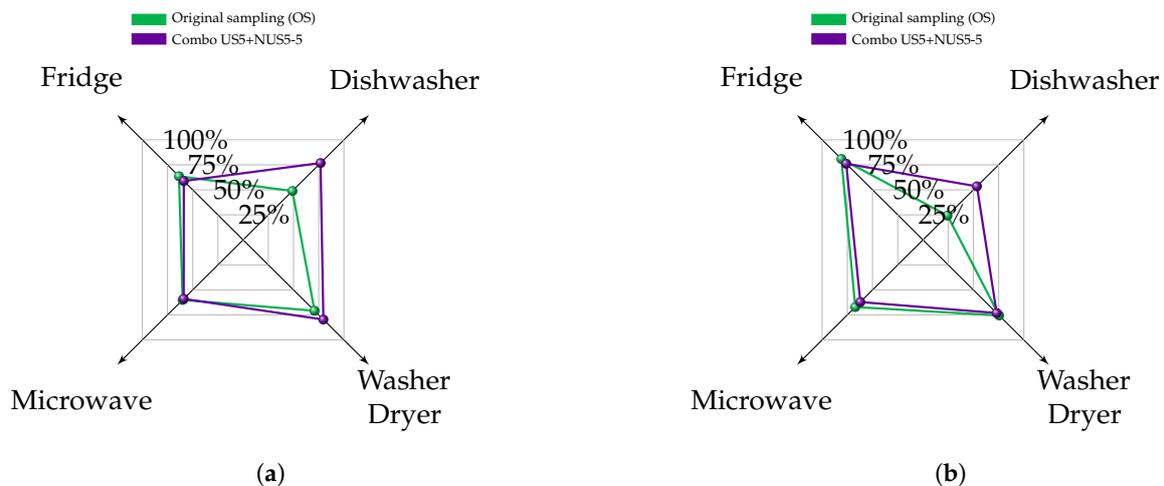


Figure 7. Comparison of disaggregation performance at *specific-rate* for REDD dataset in *seen* (a) and *unseen* (b) scenario.

Evaluating the US+NUS combinations at the *max-rate*, Table 14, the US5+NUS5-5 confirms the good performance, with an improvement with respect to OS of about 1.4%.

These results show that, in both *seen/unseen* scenarios, the sampling reduction allows an overall increment of the disaggregation performance.

Table 14. REDD *max-rate* evaluations (F_1 -score) for the different configurations in the *unseen* scenario. Appliance best score and overall best score are highlighted.

Sampling Method	Appliance				Overall
	Dishwasher	Fridge	Microwave	Washer Dryer	
OS	0.243	0.809	0.671	0.755	0.646
US5+NUS5-5	0.562_n	0.759 _u	0.642 _n	0.655 _u	0.655
US10+NUS10-5	0.560 _n	0.710 _u	0.571 _n	0.575 _u	0.611
US10+NUS10-10	0.518 _n	0.710 _u	0.630 _n	0.575 _u	0.619
US20+NUS20-10	0.393 _n	0.578 _u	0.630 _n	0.469 _n	0.522
US20+NUS20-20	0.523 _n	0.578 _u	0.619 _n	0.511 _n	0.565

_u: US applied; _n: NUS applied.

5.3. Noised Scenario

In order to evaluate the performance of the proposed approach in a realistic scenario, further experiments have been performed in *noised* condition for the UK-DALE dataset. Specifically, with reference to Equations (1) and (2), the noise component $e[n]$, which represents the contribution provided by unknown appliances and circuit noise, has not been removed. Therefore, the aggregated power consumption signal adopted during the disaggregation phase corresponds with the actual “main” signal, thus recorded at the general switch of the house.

The overall best US and NUS configurations prevailing in the combo comparison in both *seen* and *unseen* conditions for the *denoised* scenario have been selected for the evaluation also in the *noised* condition. Specifically, for the UK-DALE dataset, the best performance is achieved at $SR = 5$ and $ER = 5$, thus US5 and NUS5-5, in *seen* condition, and at $SR = 20$ and $ER = 10$, thus US20 and NUS20-10 in *unseen* condition. Thus, the best network resulting for each appliance has been re-trained with *noised* signals. The results of the *noised* evaluations are reported in Tables 15 and 16 for *seen* and *unseen* scenarios, respectively.

Table 15. UK-DALE *specific-rate* evaluations (F_1 -score) in *noised* condition for *seen* scenario.

Sampling Method	Appliance					Overall
	Dishwasher	Fridge	Kettle	Microwave	Washing Machine	
OS	0.621	0.717	0.879	0.601	0.464	0.669
US5	0.631	0.679	0.674	0.542	0.348	0.598
US20	0.536	0.639	0.381	0.359	0.268	0.440
NUS5-5	0.860	0.294	0.792	0.277	0.492	0.571
NUS20-10	0.575	0.237	0.725	0.001	0.463	0.422
Combo US+NUS						
US5+NUS5-5	0.860 _n	0.679 _u	0.792 _n	0.542 _u	0.492 _n	0.679
US20+NUS20-10	0.575 _n	0.639 _u	0.725 _n	0.359 _u	0.463 _n	0.557

_u: US applied; _n: NUS applied.

With respect to the *denoised* scenario in *seen* condition, Table 5, the disaggregation performance at OS achieved an overall degradation of about 9.5%, whereas US5 and US20 configurations reached an overall loss of 17.6% and 27.6%, and NUS5-5 and NUS20-10 configurations of 12.6% and 31.2%. As for the *denoised* condition, the combination of US5+NUS5-5 confirmed to achieve better performance than OS, with a slight degradation with respect to the *denoised* scenario of about 12.1%.

Concerning the *unseen* performance, compared to the *denoised* scenario results in Table 8, the configurations in *noised* condition present an overall more significant degradation than the one

in *seen* scenario. Specifically, a performance worsening of 13.2% is reached by the OS, whereas US5 and US20 configurations present an overall degradation of 20.6% and 31.1%, respectively, NUS5 and NUS20-10 of 20.8% and 40.8%, respectively. As shown, the lower performance deteriorations are achieved with US5 and NUS5; indeed, the combo US5+NUS5-5 achieves slight better performance than OS, instead of US20+NUS20-10 as in the *denoised* scenario.

Table 16. UK-DALE *specific-rate* evaluations (F_1 -score) in *noised* condition for *unseen* scenario.

Appliance						
Sampling Method	Dishwasher	Fridge	Kettle	Microwave	Washing Machine	Overall
OS	0.470	0.696	0.798	0.098	0.128	0.485
US5	0.426	0.637	0.734	0.084	0.117	0.431
US20	0.364	0.559	0.273	0.073	0.230	0.330
NUS5-5	0.541	0.249	0.710	0.135	0.236	0.425
NUS20-10	0.344	0.144	0.546	0.006	0.250	0.353
Combo US+NUS						
US5+NUS5-5	0.541 _n	0.637 _u	0.734 _u	0.135 _n	0.236 _n	0.487
US20+NUS20-10	0.364 _u	0.559 _u	0.546 _n	0.073 _u	0.250 _n	0.400

u: US applied; *n*: NUS applied.

5.4. Discussion

The results obtained from the application of the proposed sub-sampling techniques allow elaboration on the capability to recognize an appliance mainly from their steady or transient state information, therefore to discriminate appliances whose information is in a steady state or a transient state dominant. The following considerations have been carried out by observing the appliances behaviours exposed in *denoised* scenarios reported in Section 5.2.1 and Section 5.2.2. The considerations are expressed per appliance typology in order to provide insights independently from the adopted dataset and related only to the appliance category. Moreover, for each appliance in the datasets, the main characteristics regarding number and length of the activations, as well as number and power consumption of the working states, are reported in Table 17.

A first and clear case of appliance with dominant information in the transient state is the dishwasher. Despite this being the appliance with the longest activations (Table 17), thus having a higher number of steady state samples with respect to the transient state ones, all the NUS configurations produced better performance than OS, as shown in Tables 5, 8, 11 and 13. The main information is concentrated on the transition states, and the reduction of the data during the steady state (the US reduce the data for both steady and transition states) helps the network to efficiently acquire a clear understanding of the appliance behaviour. On the other hand, a uniform reduction of the number of samples produces a reduction of all the samples in both steady and transient phases, without providing an advantage during the training process.

On the contrary, the fridge information seems to be principally correlated with the steady states and their power levels. In fact, applying a significant data reduction ($SR = 20$) results in a modest degradation of the performance as shown in Tables 5, 8, 11 and 13. On the other hand, the application of the NUS reduces the number of steady state samples compared to the transient state ones, resulting in an alteration of the steady/transient states ratio, which affects the training phase negatively. Moreover, an additional side effect caused by the NUS sub-sampling affects the fridge by producing a further reduction of the performance. In particular, a windowing procedure is applied at the network input to select the portions of the dataset with the target appliance activations and the portions without the target appliance activations, the latter used to train the network to recognize the silence for the specific

appliance. In the case of the fridge, its activations are present for almost 50% of the dataset duration, since the fridge is continuously active. The NUS method reduces the length of the silence portions, and almost zero examples are provided at the network during the training phase.

Table 17. Appliances details.

Appliance	Activations		Working States	
	Nr.	Length *	Nr.	Power + (W)
UK-DALE				
Dishwasher	~271	915/±86	3	2.3k/±60–120/±10–0
Fridge	~18k	256/±1853	2	90/±20–0
Kettle	~3k	29/±262	2	2.3k/±45–0
Microwave	~3k	22/±20	2	1.4k/±210–0
Washing machine	~520	910/±145	4	1.9k/±150–360/±85–180/±45–0
REDD				
Dishwasher	~18	1076/±150	3	1.1k/±35–240/±15–0
Fridge	~800	223/±139	2	210/±70–0
Microwave	~350	18/±24	2	1.5k/±65–0
Washer dryer	~520	533/±419	2	2.7k/±40–0

*: average length and standard deviation as number of samples at OS. +: average and standard deviation for each working state.

Differently from the previous appliances, the kettle and the microwave present a heterogeneous behaviour. In the working state, the kettle is characterized by a high power consumption (Table 17), which allows it to be easily recognizable. On the other hand, its activations are rather short compared to the ones of other appliances, thus, increasing the SR, the number of valid samples is strongly reduced, in some cases to just two samples, and the high consumption power helps the disaggregation process. On the contrary, in the case of NUS, even if the SR increases, for values of ER equal to 10 and 20, essentially all the activation samples are used; therefore, the target activations are kept unaltered, whereas the activations of the remaining appliances are fully affected by the NUS sub-sampling. Indeed, good performance is achieved from NUS10-10 to NUS20-10 in both *seen* and *unseen* scenarios: a slight degradation with respect to OS in the *seen* case, and slight better results in the *unseen* one, as reported in Tables 5 and 8, respectively.

The microwave presents short activations similarly to the ones of the kettle, and it exhibits the same behaviour for both US and NUS methods in the REDD dataset as shown in Tables 11 and 13. In the case of UK-DALE, despite the performance increment with the NUS method for increasing SR, the overall performance presents a significant degradation compared to the OS of the kettle (see Tables 5 and 8). As reported in the appliance details, Table 17, the microwave in the UK-DALE dataset presents a high standard deviation of the power level for the working state. This high variability depends on random spikes during the activations that reach instant power levels between 2.2 kW and 2.7 kW, thus almost double the average consumption power reported in Table 17. These spikes in the footprints are mainly present in Building 1, used to train the network and perform the tests in *seen* condition, whereas no spikes are present in Buildings 2 and 5, as depicted in the traces in Figure 8. Specifically, Building 2 data have been adopted in both training and test phases for the *seen* scenario, whereas Building 5 data have been exclusively used in the test phase for *unseen* condition.

Therefore, the strong heterogeneity in the footprints negatively affects the training phase of the network by producing a final model both unable to fully represent the appliance behaviour in *seen* tests, i.e., lowering the achievable performance, and completely inappropriate to disaggregate the *unseen* data, i.e., traces with a marked incongruity with respect to the ones used in the training phase.

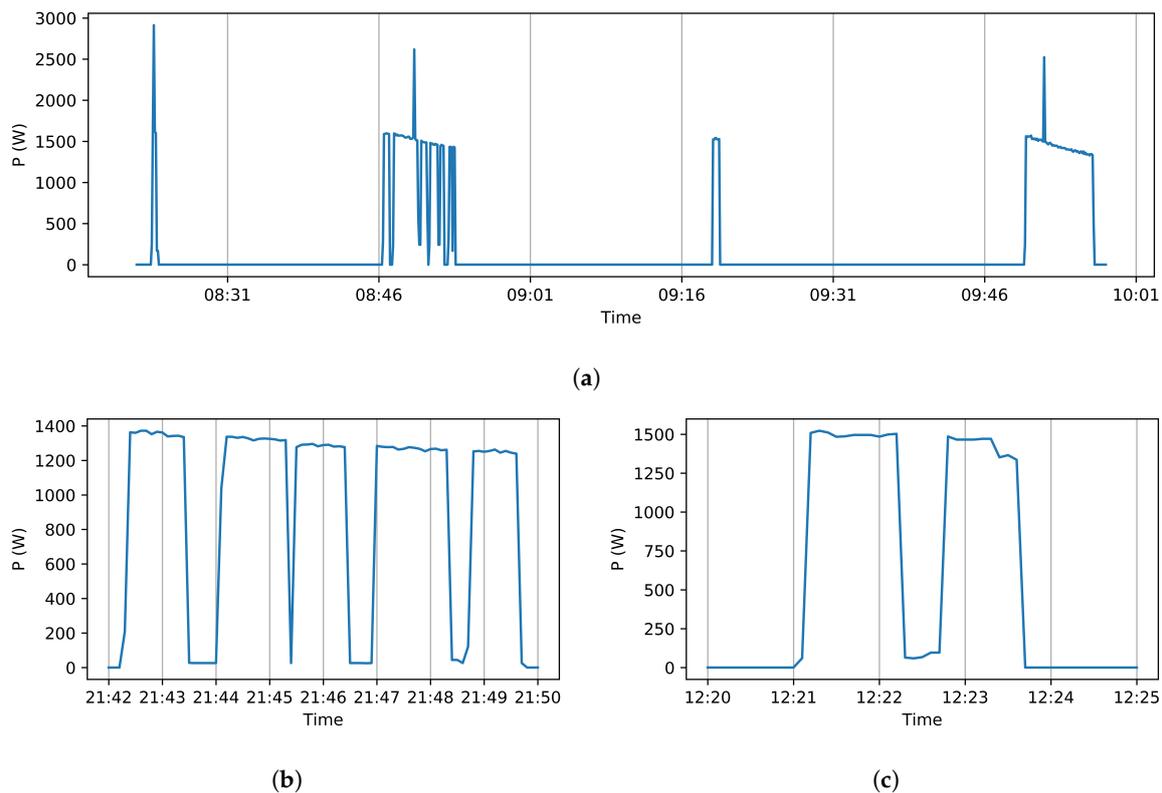


Figure 8. Microwave footprints in the UK-DALE dataset for Building 1 (a), 2 (b), and 5 (c).

Based on the data reported in Table 17, the washing machine should be the most difficult appliance to recognize due to the low number of activations, their significant length, the large number of working states (four) and the presence of two working states with close power consumption. Indeed, the results related to the washing machine are lower than the ones of the other appliance. However, the high number of working states, thus of transition phases, take full advantage of the NUS method and better performance than OS and US is archived for all NUS configurations.

Compared to the washing machine, the washer dryer presents only two working states without any intermediate states. Indeed, differently from the washing machine, the NUS method does not provide any benefit for the performance. Specifically, the high F_1 -score is achieved either in OS or US with a slight decimation of the data ($SR = 5$). Therefore, the network exploits the information associated with the power levels of the steady states in accordance with what reported in Table 17. The washer dryer has the highest power level among all the appliances, therefore the power information of the state is enough to produce a good disaggregation of the appliance.

6. Conclusions

In this work, an extended experimental campaign, in order to perform an advanced analysis and to validate the Neural NILM approach in combination with an ad hoc non-uniform subsampling strategy, has been presented. Two subsampling policies, US and NUS, have also been evaluated using UK-DALE and REDD datasets in both *seen* and *unseen* scenarios by assuming a *denoised* environment, thus signals without contributions from unknown appliances and circuit noise. Exploiting the possibility of the Neural NILM to characterize each appliance with a dedicated network topology and subsampling strategy, different combinations of US+NUS have also been evaluated. Specifically, selecting for each appliance, the best policy between US and NUS, the overall disaggregation results have outperformed the ones achieved at OS, in terms of F_1 -score. Moreover, the application of US and NUS strategies achieves a significant reduction of the overall data, i.e., requiring less data to be collected and transmitted by a measurement system—smart meter.

In order to have an insight into the performance in a realistic scenario, additional evaluations have been carried out by assuming a *noised* scenario for the UK-DALE dataset. Specifically, the best US and NUS configurations in *denoised* conditions have been evaluated in *noised* ones as well. The achieved results confirmed the advantage provided, in the general results, by adopting a combination of US+NUS strategies.

As discussed, the NUS evaluations have been executed by assuming an a priori knowledge of the state transition points, aiming to reduce the disaggregation errors due to external causes, i.e., the application of expansion windows for erroneous detection of the signal transitions. Therefore, in future works, the effort will be toward the development of a pre-processing stage to automatically detect the transitions states in the aggregated power signal. Moreover, a further advancement to investigate will regard the possibility to adopt different networks to separately take care of steady and transient phases. The goal is to produce a network to work with slow changes, thus at a lower sampling rate, and another one to work with fast changes, thus at a higher sampling rate. Finally, more extended datasets, e.g., REFIT [39], will be taken into account for the experimental phase.

Author Contributions: Conceptualization, R.B., S.S., and L.M. Investigation and methodology, R.B., M.F., and S.S. Software development and Formal Analysis, M.F. and R.B. Supervision, M.F., R.B., E.P., and S.S. Writing-Original draft preparation and Review, M.F., R.B., E.P., and S.S. Project administration, S.S., and L.M.

Funding: This research received no external funding.

Acknowledgments: The authors would like to acknowledge the CINECA award under the ISCRA initiative, for the availability of high performance computing resources and support. Moreover, the authors also gratefully acknowledge the support of the Nvidia Corporation with the donation of the Titan Xp GPU used for this research.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Malik, F.H.; Lehtonen, M. A review: Agents in smart grids. *Electr. Power Syst. Res.* **2016**, *131*, 71–79. [[CrossRef](#)]
2. Mohsenzadeh, A.; Pang, C. Two stage residential energy management under distribution locational marginal pricing. *Electr. Power Syst. Res.* **2018**, *154*, 361–372. [[CrossRef](#)]
3. Mamounakis, I.; Efthymiopoulos, N.; Makris, P.; Vergados, D.J.; Tsaousoglou, G.; Varvarigos, E.M. A novel pricing scheme for managing virtual energy communities and promoting behavioral change towards energy efficiency. *Electr. Power Syst. Res.* **2019**, *167*, 130–137. [[CrossRef](#)]
4. Ahmed, M.S.; Mohamed, A.; Khatib, T.; Shareef, H.; Homod, R.Z.; Ali, J.A. Real time optimal schedule controller for home energy management system using new binary backtracking search algorithm. *Energy Build.* **2017**, *138*, 215–227. [[CrossRef](#)]
5. Ahmadi-Karvigh, S.; Becerik-Gerber, B.; Soibelman, L. A framework for allocating personalized appliance-level disaggregated electricity consumption to daily activities. *Energy Build.* **2016**, *111*, 337–350. [[CrossRef](#)]
6. Bonfigli, R.; Squartini, S.; Fagiani, M.; Piazza, F. Unsupervised algorithms for non-intrusive load monitoring: An up-to-date overview. In Proceedings of the 2015 IEEE 15th International Conference on Environment and Electrical Engineering (EEEIC), Rome, Italy, 10–13 June 2015; pp. 1175–1180. [[CrossRef](#)]
7. Zoha, A.; Gluhak, A.; Imran, M.A.; Rajasegarar, S. Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey. *Sensors* **2012**, *12*, 16838–16866. [[CrossRef](#)] [[PubMed](#)]
8. Alcalá, J.M.; Ureña, J.; Hernández, Á.; Gualda, D. Sustainable Homecare Monitoring System by Sensing Electricity Data. *IEEE Sens. J.* **2017**, *17*, 7741–7749. [[CrossRef](#)]
9. Lynch, S.; Longo, L. On the relationship between sampling rate and Hidden Markov Models accuracy in non-intrusive load monitoring. In Proceedings of the 25th Irish Conference on Artificial Intelligence and Cognitive Science (AICS), Dublin, Ireland, 7–8 December 2017; Volume 2086, pp. 180–192.

10. Huang, B.; Knox, M.; Bradbury, K.; Collins, L.M.; Newell, R.G. Non-intrusive load monitoring system performance over a range of low frequency sampling rates. In Proceedings of the 6th International Conference on Renewable Energy Research and Applications (ICRERA), San Diego, CA, USA, 5–8 November 2017; pp. 505–509.
11. Kolter, J.Z.; Johnson, M.J. REDD: A public data set for energy disaggregation research. In Proceedings of the SustKDD Workshop on Data Mining Applications in Sustainability, San Diego, CA, USA, 21–24 August 2011; pp. 1–6.
12. Makonin, S.; Bajic, I.V.; Popowich, F. Efficient Sparse Matrix Processing for Nonintrusive Load Monitoring (NILM). In Proceedings of the 2nd International Workshop on Non-Intrusive Load Monitoring, Austin, TX, USA, 27 August 2014.
13. Basu, K.; Hably, A.; Debusschere, V.; Bacha, S.; Ovalle, A. A comparative study of low sampling non intrusive load dis-aggregation. In Proceedings of the 42nd Annual Conference of the IEEE Industrial Electronics Society (IECON), Florence, Italy, 23–26 October 2016; pp. 5137–5142.
14. de Castro Tomé, M.; Nardelli, P.H.J.; Alves, H. Event-based Electricity Metering: An Autonomous Method to Determine Transmission Thresholds. In Proceedings of the 2018 IEEE 87th Vehicular Technology Conference (VTC Spring), Porto, Portugal, 3–6 June 2018.
15. Simonov, M.; Chicco, G.; Zanetto, G. Real-Time Event-Based Energy Metering. *IEEE Trans. Ind. Inform.* **2017**, *13*, 2813–2823. [[CrossRef](#)]
16. Jazizadeh, F.; Afzalan, M.; Wang, J. Self-configuring Event Detection for Electricity Disaggregation. In Proceedings of the 4th International Workshop on Non-Intrusive Load Monitoring, University of Texas, Austin, TX, USA, 7–8 March 2018.
17. Jazizadeh, F.; Becerik-Gerber, B.; Berges, M.; Soibelman, L. An Unsupervised Hierarchical Clustering Based Heuristic Algorithm for Facilitated Training of Electricity Consumption Disaggregation Systems. *Adv. Eng. Inform.* **2014**, *28*, 311–326. [[CrossRef](#)]
18. Anderson, K.D.; Bergés, M.E.; Ocneanu, A.; Benitez, D.; Moura, J.M.F. Event detection for Non Intrusive load monitoring. In Proceedings of the IECON 2012—38th Annual Conference on IEEE Industrial Electronics Society, Montreal, QC, Canada, 25–28 October 2012; pp. 3312–3317. [[CrossRef](#)]
19. Berges, M.; Goldman, E.; Matthews, H.S.; Soibelman, L.; Anderson, K. User-Centered Nonintrusive Electricity Load Monitoring for Residential Buildings. *J. Comput. Civ. Eng.* **2011**, *25*, 471–480. [[CrossRef](#)]
20. Quy, T.L.; Zerr, S.; Ntoutsand, E.; Nejd, W. Data augmentation for dealing with low sampling rates in NILM. In Proceedings of the 4th International Workshop on Non-Intrusive Load Monitoring, University of Texas, Austin, TX, USA, 7–8 March 2018.
21. Clark, M.; Lampe, L. Single-channel compressive sampling of electrical data for non-intrusive load monitoring. In Proceedings of the 2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Orlando, FL, USA, 14–16 December 2015; pp. 790–794. [[CrossRef](#)]
22. Anderson, K.; Ocneanu, A.; Benitez, D.; Carlson, D.; Rowe, A.; Bergés, M. 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), Beijing, China, 12 August 2012.
23. Wang, Y.; Filippi, A.; Rietman, R. Compressive Sampling for Non-intrusive Appliance Load Monitoring (NALM) Using Current Waveforms. In Proceedings of the Signal Processing, Pattern Recognition and Applications/779: Computer Graphics and Imaging (SPPRA 2012), Crete, Greece, 18–20 June 2012.
24. Kelly, J.; Knottenbelt, W. Neural NILM: Deep Neural Networks Applied to Energy Disaggregation. In *BuildSys '15, Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*, Seoul, Korea, 4–5 November 2015; ACM: New York, NY, USA, 2015; pp. 55–64.
25. Bonfigli, R.; Principi, E.; Fagiani, M.; Severini, M.; Squartini, S.; Piazza, F. Non-intrusive load monitoring by using active and reactive power in additive Factorial Hidden Markov Models. *Appl. Energy* **2017**, *208*, 1590–1607. [[CrossRef](#)]
26. Marvasti, F.A. *Nonuniform Sampling: Theory and Practice*; Kluwer Academic/Plenum: New York, NY, USA, 2001.
27. Bonfigli, R.; Felicetti, A.; Principi, E.; Fagiani, M.; Squartini, S.; Piazza, F. Denoising autoencoders for Non-Intrusive Load Monitoring: Improvements and comparative evaluation. *Energy Build.* **2018**, *158*, 1461–1474. [[CrossRef](#)]

28. Valenti, M.; Bonfigli, R.; Principi, E.; Squartini, S. Exploiting the Reactive Power in Deep Neural Models for Non-Intrusive Load Monitoring. In Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN), Rio, Brazil, 8–13 July 2018; pp. 1–8. [CrossRef]
29. Nair, V.; Hinton, G.E. Rectified Linear Units Improve Restricted Boltzmann Machines. In Proceedings of the 27th International Conference on International Conference on Machine Learning (ICML'10), Haifa, Israel, 21–24 June 2010; pp. 807–814.
30. Sutskever, I.; Martens, J.; Dahl, G.; Hinton, G. On the importance of initialization and momentum in deep learning. In *Machine Learning Research, Proceedings of the 30th International Conference on Machine Learning, Atlanta, GA, USA, 17–19 June 2013*; Dasgupta, S., McAllester, D., Eds.; PMLR: Atlanta, GA, USA, 2013; Volume 8, pp. 1139–1147.
31. Kelly, J.; Knottenbelt, W. The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes. *Sci. Data* **2015**, *2*. [CrossRef] [PubMed]
32. Hartigan, J.A.; Wong, M.A. Algorithm AS 136: A K-Means Clustering Algorithm. *J. R. Stat. Society. Ser. C (Appl. Stat.)* **1979**, *28*, 100–108. [CrossRef]
33. Murray, D.; Stankovic, L.; Stankovic, V.; Lulic, S.; Sladojevic, S. Transferability of Neural Network approaches for low-rate energy disaggregation. In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019.
34. DIncecco, M.; Squartini, S.; Zhong, M. Transfer Learning for Non-Intrusive Load Monitoring. *arXiv* **2019**, arXiv:1902.08835.
35. Chollet, F. Keras. Available online: <https://keras.io> (accessed on 1 November 2018).
36. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. Available online: <https://www.tensorflow.org/> (accessed on 1 November 2018).
37. A3NeuralNILM. Available online: <https://gitlab.com/a3labShares/A3NeuralNILM> (accessed on 1 November 2018).
38. Kolter, J.Z.; Jaakkola, T. Approximate Inference in Additive Factorial HMMs with Application to Energy Disaggregation. In *Machine Learning Research, Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics, La Palma, Canary Islands, 21–23 April 2012*; Lawrence, N.D., Girolami, M., Eds.; PMLR: La Palma, Spain, 2012; Volume 22, pp. 1472–1482.
39. Murray, D.; Stankovic, L.; Stankovic, V. An electrical load measurements dataset of United Kingdom households from a two-year longitudinal study. *Sci. Data* **2017**, *4*, 160122. [CrossRef] [PubMed]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).