Benefits of Three-Phase Metering for Load Disaggregation

Apostolos Vavouris
apostolos.vavouris@strath.ac.uk
University of Strathclyde
Electronic & Electrical Engineering
Glasgow, UK

Lina Stankovic
lina.stankovic@strath.ac.uk
University of Strathclyde
Electronic & Electrical Engineering
Glasgow, UK

Vladimir Stankovic
vladimir.stankovic@strath.ac.uk
University of Strathclyde
Electronic & Electrical Engineering
Glasgow, UK

Jiufeng Shi
jfs@discovergy.com
Discovergy GmbH
Heidelberg, Germany

ABSTRACT

With the ever increasing pace of introduction of energy intensive devices and services, such as electric vehicle (EV) charging and heat pumps, the transition to smart metering for three-phase electric installations for nationwide smart meter roll-outs is underway. In this paper, we explore how three-phase metering can benefit non-intrusive load monitoring (NILM), especially for those appliances that are difficult to disaggregate and not widely reported in the literature. Traditionally, the NILM literature tends to tackle three-phase metering by summing the three phases, without exploiting the potential benefits of load disaggregation per phase. Emphasis is placed on the disaggregation performance and loss that is introduced when using different levels of granularity of low-frequency data. Finally, we augment a public dataset with which phase the appliance is connected to, and release a three-phase electric vehicle dataset from three-phase aggregate measurements.

CCS CONCEPTS

• **Computing methodologies → Neural networks**; **Machine learning**; • **General and reference → Reliability**; **Validation**; **Verification**.

KEYWORDS

non-intrusive load monitoring (NILM), energy disaggregation, deep NILM, 3-phase load

ACM Reference Format:

Apostolos Vavouris, Lina Stankovic, Vladimir Stankovic, and Jiufeng Shi. 2022. Benefits of Three-Phase Metering for Load Disaggregation. In *6th International Workshop on Non-Intrusive Load Monitoring (NILM '22), November 9–10, 2022, Boston, MA, USA*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3563357.3566149

NILM '22, November 9–10, 2022, Boston, MA, USA

ACM ISBN 978-1-4503-9890-9/22/11…$15.00
https://doi.org/10.1145/3563357.3566149

1 INTRODUCTION

Three-phase (3ϕ) installations, predominant in commercial and industrial settings as well as in residential settings in the majority of the central and northern Continental Europe [20]—due to the traditionally lower limits of power availability per phase—are slowly becoming the new standard for residential buildings [6] to support the surge in end-user demand from high loads, such as EV chargers, and meeting emission targets with the additional inclusion of renewable energy sources. The introduction of 3ϕ installations enables the installation of larger domestic solar photovoltaic (PV) systems, when compared to current ones that are capped at a lower maximum power levels [14], imposed by the maximum load that single-phase installations' can handle. 3ϕ residential installations can also benefit the stability of the grid.

Recent standardisation for 3ϕ smart metering in UK and Europe [5, 13], are making it obligatory for manufactures to produce metering devices that are capable to measure and transmit the power, voltage, current and angle between the different phases of a 3ϕ installation. Distribution of residential loads across the three phases is expected to reduce the disaggregation noise on a per phase basis. This provides an opportunity in potentially improving disaggregation accuracy of non-intrusive load monitoring (NILM) algorithms by exploiting the load distribution across the three phases.

In this paper, a detailed and robust methodology for evaluation of load disaggregation of energy intensive appliances from 3ϕ household installations is presented. The main contributions of this paper are: 1) adapting sequence-to-subsequence (seq2subseq) [15] deep learning (DL)-based NILM algorithm from [7] providing full details of the proposed pre-processing, hyper-parameter tuning and post-processing steps for different appliances; 2) quantifying gain in disaggregation accuracy when using per phase and aggregate signal, taking into account noisiness and sparsity metrics, and different data granularities; 3) proposing a method of appliance phase identification and releasing information regarding the phase on which each appliance of the ECO dataset [3] is connected to [1]; 4) labelling the EV usage of a single 3ϕ household in Germany for a period of 1-year via transfer learning with manual verification, and releasing the labelled dataset [2].

[1]Appliance phase identification data for the ECO dataset can be accessed at: https://doi.org/10.15129/deddd9a7-0cff-4db2-8478-42abc93fba9f

[2]The research data supporting this publication can be accessed at: https://doi.org/10.15129/c41a6a02-5df5-4ed7-b8e6-6488895d43f7

2 NILM APPROACHES FOR 3ϕ DATA

With plans for 3ϕ smart metering in residential settings only emerging, NILM approaches for 3ϕ data are scarce for residential buildings. However, different NILM approaches are proposed for industrial buildings, e.g., an event detection approach based on composite window analysis for 3ϕ industrial metering [22], concluding that a significant improvement in classification performance is obtained compared to detection on the sum of phases approach.

Due to the inherent complexity of measuring electricity consumption in 3ϕ installations and the requirement of additional hardware, 3ϕ datasets are limited. In an in-depth review [8] of DL based approaches for low-frequency NILM, only three 3ϕ datasets were identified, namely iAWE [2], ECO [3] and BLOND [11]. Of these, iAWE [2] consists of a single house in Delhi where electricity, gas and water were monitored for 73 days, ECO [3] contains electricity readings from 6 houses in Switzerland for a period of 8 months, and, BLOND [11] contains energy readings from an office for 213 days. This paper uses the ECO one since it is the only that contains measurements from multiple houses over several months.

ECO [3] dataset is widely used and contains information about voltage, current, active power per phase etc. as well as sub-metering of several house appliances. A comparison of performance of four different NILM algorithms on this dataset is presented in [3], where it was concluded that to achieve adequate results, a supervised approach was required. An Artificial Neural Networks (ANNs)-based approach for disaggregation of the ECO dataset is proposed in [10], where data were resampled to 10 min intervals. Again, only the aggregate signal of the phases was used as an input. Multi Layer Perceptron ANNs performed poorly and therefore only DNNs were further explored. Results, varied greatly between the appliances that were studied—i.e. fridge, freezer, PC and washing machine, with fridges performing best and PCs worst. Lastly, authors in [17, 18] proposed the use of a 4-layered bidirectional long short-term memory (LSTM) model to disaggregate several datasets, including the aggregate signal of ECO dataset. Signal Aggregate Error (SAE) ranged from 1.3% up to 64.9%, with dishwashers having an SAE of 28.8%, fridges 12.1%, washing machines 64.1% and microwaves 43.7%. To the best of our knowledge, there are no NILM methods designed to exploit the three phase by disaggregating per phase.

3 METHODOLOGY

Following the recent NILM review papers [1, 4, 8, 16, 19], we adapt the DL seq2subseq NILM approach of [15], shortlisted in [8] as one of the best performing on standard household appliances and demonstrated on the PECAN [9] dataset in [21].

3.1 Data Selection & Preparation

This study focuses on the following, regularly used, energy intensive appliances in the ECO dataset: 1) House 1: coffee machine (CM)–113 days; dryer (TD), freezer (FRZ). fridge (FRD) and washing machine (WM)–231 days; 2) House 2: dishwasher (DW), FRZ and FRD–240 days; 3) House 4: FRZ–192 days; FRD and microwave (MW)–194 days; 4) House 5: CM and FRD–218 days. Measurements were resampled to granularities of 10, 30 and 60 sec—by averaging the power consumption over the aforementioned duration—to investigate disaggregation accuracy when using low-and very-low frequency data, considering data collection and storage limitations as well as end-users' privacy, as per smart meter standards [5, 13].

A fundamental factor that affects the accuracy of NILM algorithms is the amount of activations of each appliance available to train models. We estimate this via *sparsity* measure for each appliance in a dataset, calculated as: $S = T_{\text{On}}/T_{\text{Total}}$, where T_{On} is the duration that the appliance is on and T_{Total} is the total duration. Another metric used to assess how difficult it is for a NILM method to disaggregate an appliance is Noisiness measure [12] that is positively correlated with the disaggregation performance [21], and is defined as: $NM^{(T)} = \left(\sum_{t=1}^{T}\left|y_t - \sum_{m=1}^{M} y_t^{(m)}\right|\right)/\left(\sum_{t=1}^{T} y_t\right)$, where T is the total monitoring duration in the number of samples, y_t is the aggregated load measured at sampling instant t and $y_t^{(m)}$ is the submetered measurement of load/appliance m at sampling instant t. M denotes the number of appliances that are disaggregated. In Table 1, the sparsity and the noisiness metrics for each appliance both for per-phase (NM_{ϕ}) and aggregated (NM_{Agg}) signal is presented, where all other loads are considered as noise. ϕ denotes the phase that the appliance is drawing current from.

3.2 Sequence-to-Subsequence Parameters

The seq2subseq NILM algorithm with conditional Generative Adversial Network (GAN) is implemented as in [21], using the same approach on data splitting between training (60%), validation (10%) and testing (30%) as well as on the post-processing steps. Seq2subseq targets the middle part of a sequence and therefore offers a trade-off between the computational load of sequence-to-point—where the algorithm targets a single point—and the convergence speed of the sequence-to-sequence—where the algorithm targets the whole sequence. Training and testing was performed on the same house. The window size (WS) ω is set to the first element in the set $\{2^0, 2^1, ..., 2^n, ...\}$ that is larger than $2 \times L \times f$, where L [in sec] is the usual length of appliance cycle period and f [in Hz] is the frequency of the samples. The WS used are: 1) for CM, FRZ, FRD and MW: 256, 128 and 64; 2) for DW and TD: 2048, 512 and 256; 3) for WM: 1024, 512 and 128, for granularities of 10, 30 and 60 sec, respectively. The rest of the hyperparameters were chosen based on the performance on the validation set. L1 loss was used in all setups. Stochastic Gradient Descent (learning rate = 0.001) and the ADAM optimiser (learning rate = 0.0005, momentum term =0.5), were used for the discriminator and generator filters, respectively. The weights on L1 and GAN term for the generator gradient were 100 and 1, respectively. The layers used were: 1) 7 for $256 \leq \omega \leq 2048$; 2) 6 for $\omega = 128$; and 3) 5 for $\omega = 64$. The number (n) of generator and discriminator filters in the first convolutional layer was given as $\text{n} = \omega/4$, i.e., $\text{n} \in [16, 512]$. Using the early stopping criterion on the validation set, the number of epochs was chosen to be 120.

3.3 Accuracy Improvement & Granularity Loss

The improvement in accuracy, when using the signal only from the phase where the appliance is connected, is given as: $G_{\text{phase}} = Acc_{\text{phase}}/Acc_{\text{aggregate}} - 1$, where Acc, is the standard *Accuracy* metric [21]. Furthermore, the loss, introduced by lower-sampled data, was used in order to correlate the disaggregation loss with the use of less granular data, given as: $Loss = Acc_i/Acc_j - 1$, where Acc_i and Acc_j

Table 1: Phase ϕ, Sparsity S [%] and relative disaggregation performance in accuracy [%] due to per-phase disaggregation.

						Aggregated Phases			Appliance Phase			Gain G_{phase}		
House	Appliance	ϕ	S	NM_{Agg}	NM_{ϕ}	10 sec	30 sec	1 min	10 sec	30 sec	1 min	10 sec	30 sec	1 min
1	CM	2	0.64	98.56	96.18	75.15	58.54	-	79.77	71.97	59.81	6.15	22.94	-
	TD	3	3.25	91.98	56.30	40.73	36.47	16.82	79.27	78.06	75.49	94.62	114.04	348.8
	FRZ	1	54.45	93.54	86.32	85.32	83.59	81.28	93.53	92.69	90.69	9.62	10.89	11.58
	FRD	2	36.92	92.66	78.76	69.52	69.05	68.00	83.79	82.06	80.51	20.53	18.84	18.40
	WM	1	6.70	91.79	82.64	80.33	67.70	47.12	88.02	83.30	72.23	9.57	23.04	53.29
2	DW	1	1.43	92.73	87.78	62.70	43.22	81.56	67.17	60.27	87.80	7.13	39.45	7.65
	FRZ	1	50.16	87.36	78.80	91.09	87.89	83.14	92.05	89.64	85.32	1.05	1.99	2.62
	FRD	1	34.09	88.63	80.92	85.48	82.49	76.76	87.32	84.72	79.58	2.15	2.70	3.67
4	FRZ	1	28.90	96.76	94.64	81.80	80.05	80.60	85.32	83.61	83.98	4.30	4.45	4.19
	FRD	1	84.83	79.80	66.59	50.12	47.73	46.98	54.70	54.98	54.19	9.14	15.19	15.35
	MW	1	1.39	98.12	96.90	64.88	60.45	58.13	74.28	74.22	69.70	14.49	22.78	19.90
5	CM	3	1.44	99.32	98.48	74.56	66.41	54.52	83.48	79.67	72.57	11.96	19.97	33.11
	FRD	3	35.70	94.39	87.37	75.10	74.54	70.20	89.42	89.31	87.13	19.07	19.81	24.12

are the *Accuracy* metrics when using lower and higher frequency data, respectively. Here, $i = 30 \veebar 60$ sec and $j = 10$ sec.

3.4 Phase Identification & Load Labelling

As ECO does not contain information on which phase each appliance is connected to, this information was inferred from the data. To that end, for each appliance, we train and test the seq2subseq model on each phase separately, with parameters described in 3.2. The models are trained and tested on a small subset—30% of the total dataset—of known timestamped activations of each appliance for each phase—training and testing sets do not contain the same set of activations. We expect that the models on the phase where the appliance is connected will result in high estimated consumption output with high confidence whereas in all other phases the model predicted consumption output will be low. This assumption was further validated through manual inspection of the signal. In order to compare the performance of the three different models, Match Rate (MR) [21] was used since it measures how "well" the model's output matches the ground truth, when compared to other commonly used metrics that are only averaging the output. The resulting MR on the phase where the appliance is connected will be relatively higher than on the other two phases, thus indicating which phase (ϕ) the appliance is connected to.

Furthermore, the 3ϕ meter readings from an unlabelled German household were also used to demonstrate the value of per-phase disaggregation for dataset labelling. Measurements spanned a period of one year: 01/01/2021—31/12/2021, where an EV symmetrical load charger was installed. In the absence of sub-metering data for EV charging, the dataset was labelled via transfer learning, i.e., training the seq2subseq algorithm as per [21] with the one-phase PECAN Dataport [9], and disaggregating the load per phase. The training houses from PECAN were chosen such that they had a similar power level (~3kW) as one phase of the German household. The symmetric EV load distribution across the 3 phases enabled the elimination of false positives, due to other loads, as activations. These activations were validated by manually inspecting the entire period of the dataset and annotating symmetrical 3ϕ loads.

4 EXPERIMENTAL RESULTS

Results obtained on the ECO dataset are summarised in Table 1. *Accuracy* metric is presented for data sampled at 10, 30 and 60 sec when using the aggregate and the per-phase signal. *Accuracy* metric, also referred as Total Energy Correctly Assigned, is demonstrated [21] to be a more accurate measure of the performance of a regression network when compared to traditional used metric such as the Mean Absolute Error (MAE) and normalised Signal Aggregate Error (SAE). The Gain as introduced in Section 3 is also presented. In house 1, as the CM signal consisted of very short pulses, in the range of 30–60 sec, the algorithm was unable to disaggregate the signal when using a sampling rate of 1 minute. In addition, when comparing the CM of house 1 to the one of house 5—where more activations, as indicated by the *Sparsity* metrics, were present—disaggregation accuracy for 10 sec data is similar, whereas in 30 sec data sparser signal lead to increased performance. As expected, due to the reduced noise in the signal—summarised in Table 1—results demonstrate that accuracy is increased when using the per-phase signal. In house 1, per-phase disaggregation accuracy of the dryer is greatly increased, especially for lower granular data. This is partially due to the reduced noise, confirmed by NM_{ϕ} in the signal and manual observation, which indicated false positives due to an appliance with a multi-state load profile—connected to a difference phase than dryer—in the aggregate of 3 phases.

Similar performance on refrigeration appliances, across houses, is observed with improved performance for freezers relative to fridges, which can be attributed to the higher load of freezers, and their more constant current draw when compared to the more variable load of fridges that are used more often—opening/closing the door. An exception exists in house 4, where, despite the noisiness metric indicating the signal of the fridge is less noisy than other houses, the disaggregation performance appears to be poor. This was the result of a second fridge with an almost identical signature that was present, not sub-metered [3] and connected to the same phase. Therefore, the algorithm was also detecting the second fridge that was installed in the house which led to a decrease in accuracy. It is also worth noting, that although fridges and freezers present a very similar consumption pattern, the algorithm discriminate these devices due to the small difference in the peak energy levels as well as to the differences in the duration of the activations of the two appliances. It is also worth noting that refrigeration appliances that demonstrate a lower *Sparsity* metric tend to be disaggregated more easily especially for less granular data.

Table 2 presents the loss in accuracy, for the aggregate and the appliance phase signals, when using 30-sec and 1-min instead of 10-sec data. Negative values indicate drop in accuracy. In general, there is a positive correlation between the loss and the granularity. Accuracy of refrigeration appliances in all houses, except house 2, when using the phase signal is almost invariant w.r.t. granularity levels. This is expected as these appliances tend to have a constant periodic signal. Furthermore, a significant decrease in accuracy of the dryer and the washing machine in house 1 when using lower granularity data is observed. This is especially visible when using the aggregate, as the combination of higher levels of noise and the reduction in the granularity led to about 50% decrease in accuracy.

Therefore, it can be concluded that, in general, disaggregation of refrigeration appliances can be highly effective when using data with granularities in the range of 1 min, whereas devices with sparse activations and multi-state load profiles require a higher sampling frequency to achieve similar performance. Furthermore, appliances with sparse activations and variable multi-state current draws with less granular data, benefit from per-phase disaggregation, as is the case with the dryer in house 1. However, as observed from both Tables 1 and 2, disaggregation accuracy of dishwasher does not follow the same pattern as performance improved when using less granular data. This is because the dishwasher's signal consists of one or more high energy pulses with lower energy levels around the main pulses. There were some activations with low energy—probably corresponding to rinse cycles—whose energy was spread across many samples for more granular data. The algorithm was unable to disaggregate the pattern of the energy consumption outside the main pulses when using data of higher granularity as well as these low-energy uses. On the contrary, when using 1-min data, as these low-powered level signals were aggregated over a minute, the algorithm could assign the disaggregated energy to the dishwasher. Lastly, disaggregation results for both coffee machines in houses 1 and 5 as well as microwave in house 4 indicate that a higher sampling frequency can greatly increase accuracy for appliances with sparse loads (see Table 1) and short activations.

To demonstrate applicability of the proposed approach to appliance phase labelling, an EV charger on an unseen 3ϕ metered dataset was labelled using the methodology presented in Section 3.4. In Table 3, estimated Accuracy results both for 1-and 15-min data are presented. The estimation of the EV's charger signal as described was used to juxtapose the signal obtained from the algorithm. This example of labelling a dataset where ground truth data are absent, underlines the importance of using the per phase signal when compared to the phase-aggregate, as otherwise the elimination of the false positives from similar loads would be difficult.

Table 2: Granularity loss (*Loss*) [%] disaggregating on 30-and 60-sec measurements, compared to 10-sec.

		Aggregated Phases		Appliance Phase	
House	Appliance	30 sec	1 min	30 sec	1 min
1	CM	-22.10	-	-9.78	-25.02
	TD	-10.46	-58.70	-1.53	-4.77
	FRZ	-2.03	-4.74	-0.90	-3.04
	FRD	-0.68	-2.19	-2.06	-3.91
	WM	-15.72	-41.34	-5.36	-17.94
2	DW	-31.07	30.08	-10.27	30.71
	FRZ	-3.51	-8.73	-2.62	-7.31
	FRD	-3.50	-10.20	-2.98	-8.86
4	FRZ	-2.14	-1.47	-2.00	-1.57
	FRD	-4.77	-6.26	0.51	-0.93
	MW	-6.83	-10.40	-0.08	-6.17
5	CM	-10.93	-26.88	-4.56	-13.07
	FRD	-0.75	-6.52	-0.12	-2.56

Table 3: Estimated accuracy [%] of EV labelling.

	ϕ_1	ϕ_2	ϕ_3	ϕ_{Agg}
1 min	90.65	93.13	90.98	91.84
15 min	86.99	87.46	86.74	87.46

5 CONCLUSIONS AND RECOMMENDATIONS

In this paper, we demonstrate, quantitatively, the improvement of load disaggregation per phase in 3ϕ installations, over the traditional approach of disaggregating the sum of three phases. This is especially timely given the update in national smart meter 3ϕ roll-outs to provide for the growing number of households that include high power loads such as EVs. We also show that appliances that tend to be hard to disaggregate in the literature, due to sparse activations and variable load profiles, are more prone to noise from unknown appliances and therefore disaggregating per phase mitigates the effect of false positives. The appliances that benefit mostly from disaggregation per phase are washing machines, dryers, coffee machines and microwaves. Devices with similar load profiles, benefit from being spread across different phases. This demonstrates the importance of carefully picking the phase on which each appliance is connected to during a 3ϕ installation, to facilitate more accurate disaggregation. For example, it is recommended to distribute refrigeration appliances, when more than one exist, on different phases, as well as connecting resistive appliances with similar profiles such as kettles, coffee machines and stoves on different phases. The feasibility of accurately labelling an EV charger without the presence of ground truth data using the per-phase signals as well as knowledge transferred from another one-phase dataset was also demonstrated, and the annotated dataset released. Lastly, as smart metering is moving towards less granular data with the main concern being data privacy of the end users, disaggregation using lower sampling frequencies in the area of 15 min should be explored further.

ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 955422.

REFERENCES

[1] Georgios-Fotios Angelis, Christos Timplalexis, Stelios Krinidis, Dimosthenis Ioannidis, and Dimitrios Tzovaras. 2022. NILM applications: Literature review of learning approaches, recent developments and challenges. *Energy and Buildings* 261 (2022), 111951. https://doi.org/10.1016/j.enbuild.2022.111951

[2] Nipun Batra, Manoj Gulati, Amarjeet Singh, and Mani B. Srivastava. 2013. It's Different: Insights into Home Energy Consumption in India. In *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings* (Roma, Italy) *(BuildSys'13)*. Association for Computing Machinery, New York, NY, USA, 1—-8. https://doi.org/10.1145/2528282.2528293

[3] Christian Beckel, Wilhelm Kleiminger, Romano Cicchetti, Thorsten Staake, and Silvia Santini. 2014. The ECO Data Set and the Performance of Non-Intrusive Load Monitoring Algorithms. In *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings* (Memphis, Tennessee) *(BuildSys '14)*. Association for Computing Machinery, New York, NY, USA, 80–89. https://doi.org/10.1145/2674061.2674064

[4] Deepika R Chavan and Dagadu S More. 2022. A Systematic Review on Low-Resolution NILM: Datasets, Algorithms, and Challenges. *Electronic Systems and Intelligent Computing* (2022), 101–120. https://doi.org/10.1007/978-981-16-9488-2_9

[5] Comité Européen de Normalisation Électrotechnique. 2022. *Smart Grids and Meters.* Retrieved July 14, 2022 from https://www.cencenelec.eu/areas-of-work/cen-cenelec-topics/smart-grids-and-meters/smart-meters/

[6] The Association For Renewable Energy and Clean Technology. 2018. *The feasibility, costs and benefits of three phase power supplies in new homes.* Retrieved July 14, 2022 from https://www.r-e-a.net/new-homes-shouldnt-be-held-back-by-pre-wwii-electrical-standards/

[7] GitHub-DLZRMR/seq2subseq. 2019. *Seq2subseq Method for NILM.* Retrieved July 14, 2022 from https://github.com/DLZRMR/seq2subseq

[8] Patrick Huber, Alberto Calatroni, Andreas Rumsch, and Andrew Paice. 2021. Review on Deep Neural Networks Applied to Low-Frequency NILM. *Energies* 14, 9 (2021). https://doi.org/10.3390/en14092390

[9] Pecan Street Inc. 2022. *Dataport.* Retrieved July 14, 2022 from https://www.pecanstreet.org/dataport/

[10] Tomasz Jasiński. 2020. Modelling the disaggregated demand for electricity in residential buildings using artificial neural networks (deep learning approach). *Energies* 13, 5, 1263. https://doi.org/10.3390/en13051263

[11] Thomas Kriechbaumer and Hans-Arno Jacobsen. 2018. BLOND, a building-level office environment dataset of typical electrical appliances. *Scientific data* 5, 1 (2018), 1–14. https://doi.org/10.1038/sdata.2018.48

[12] Stephen Makonin and Fred Popowich. 2015. Nonintrusive load monitoring (NILM) performance evaluation. *Energy Efficiency* 8, 4 (2015), 809–814. https://doi.org/10.1007/s12053-014-9306-2

[13] Department of Energy & Climate Change. 2014. *Smart Metering Equipment Technical Specifications: Second Version.* Retrieved July 14, 2022 from https://www.gov.uk/government/consultations/smart-metering-equipment-technical-specifications-second-version

[14] Operations Directorate of Energy Networks Association. 2012. *Engineering Recommendation G83 Issue 2.* Retrieved July 14, 2022 from https://www.ofgem.gov.uk/sites/default/files/docs/2012/08/er-g83-2-_v5--the-master-09-07-12-inc-ofgem-comments---clean-version_0.pdf

[15] Yungang Pan, Ke Liu, Zhaoyan Shen, Xiaojun Cai, and Zhiping Jia. 2020. Sequence-To-Subsequence Learning With Conditional Gan For Power Disaggregation. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 3202–3206. https://doi.org/10.1109/ICASSP40776.2020.9053947

[16] Lucas Pereira and Nuno Nunes. 2018. Performance evaluation in non-intrusive load monitoring: Datasets, metrics, and tools—A review. *WIREs Data Mining and Knowledge Discovery* 8, 6 (2018), e1265. https://doi.org/10.1002/widm.1265 arXiv:https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/widm.1265

[17] Hasan Rafiq, Xiaohan Shi, Hengxu Zhang, Huimin Li, and Manesh Kumar Ochani. 2020. A deep recurrent neural network for non-intrusive load monitoring based on multi-feature input space and post-processing. *Energies* 13, 9 (2020), 2195. https://doi.org/doi.org/10.3390/en13092195

[18] Hasan Rafiq, Hengxu Zhang, Huimin Li, and Manesh Kumar Ochani. 2018. Regularized LSTM Based Deep Learning Model: First Step towards Real-Time Non-Intrusive Load Monitoring. In *2018 IEEE International Conference on Smart Energy Grid Engineering (SEGE)*. 234–239. https://doi.org/10.1109/SEGE.2018.8499519

[19] Pascal A Schirmer and Iosif Mporas. 2022. Non-Intrusive Load Monitoring: A Review. *IEEE Transactions on Smart Grid* (2022). https://doi.org/10.1109/TSG.2022.3189598

[20] Thomas Allen Short. 2018. *Electric power distribution equipment and systems*. CRC press. https://doi.org/10.1201/9781315220741

[21] Apostolos Vavouris, Benjamin Garside, Lina Stankovic, and Vladimir Stankovic. 2022. Low-Frequency Non-Intrusive Load Monitoring of Electric Vehicles in Houses with Solar Generation: Generalisability and Transferability. *Energies* 15, 6 (2022). https://doi.org/10.3390/en15062200

[22] Shuhui Yi, Xiaodong Yin, Yinglong Diao, Binwu Wang, and Ping Wu. 2019. A New Event-detection Method Based on Composite Windows in NILM for Industrial Settings. In *2019 IEEE Sustainable Power and Energy Conference (iSPEC)*. 2768–2771. https://doi.org/10.1109/iSPEC48194.2019.8975265