# What's Up for the Weekend? Exploiting Day Type Information in Non-Intrusive Load Monitoring


Mazen Bouchur
TU Clausthal
Germany
mazen.bouchur@tu-clausthal.de

Daniel Szafranski
TU Clausthal
Germany
daniel.szafranski@tu-clausthal.de

Andreas Reinhardt
TU Clausthal
Germany
reinhardt@ieee.org



## ABSTRACT

Interactions with most electrical appliances in buildings are governed by temporal patterns. Daily routines and activities have an inherent impact on (1) the types of appliances being used, (2) at what time they are operated, and (3) for what duration. Similarly, most commercial and industrial energy consumers exhibit consumption characteristics that depend on time and the day of the week. Almost all published load disaggregation methods are, however, agnostic to the time and day during which their input data has been collected. In this work, we hence study whether an improved load disaggregation is possible when input data are partitioned by the day type, and separate models are created for each partition. We specifically consider two methods to partition daily energy consumption into two sets. The first method is based on separating weekends and holidays (from weekdays), while the second method uses $k$-means clustering to autonomously separate the data into clusters. We evaluate the disaggregation performance of each clustering method using three datasets that were collected in different settings. Our results show that an improvement up to 40 % can be reached in the disaggregation accuracy, as measured by the MAE, when temporal information is considered in the disaggregation process.



## CCS CONCEPTS

• **Hardware → Power estimation and optimization**; • **Computing methodologies → Neural networks**.

## KEYWORDS

Non-Intrusive Load Monitoring, Neural networks

**ACM Reference Format:**
Mazen Bouchur, Daniel Szafranski, and Andreas Reinhardt. 2022. What's Up for the Weekend? Exploiting Day Type Information in Non-Intrusive Load Monitoring. 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/10.1145/3563357.3566154


## 1 INTRODUCTION

Electrical energy demand is not only tightly correlated with the time of day, but also depends on the type of the day (weekday, weekend, public holiday). The extent of this effect varies between different environments, as can be seen in Fig. 1, where the daily average power demand is compared between buildings from an industrial plant, a university, and a residential home. We can observe that the consumers in industrial and educational environments do not exhibit significant appliance activity on weekends, while following regular weekday routines. In contrast, the residential data have discernible patterns on weekdays, hinting at the varying presence of inhabitants in the building. Using data from industrial plants that have been collected on weekends (where there was no activity) to train Non-Intrusive Load Monitoring (NILM) models incurs a significant computational demand, at no benefit to the disaggregation of weekday data.

Temporal data annotations are present in all NILM datasets, usually in the form of timestamped data collection times, and can be easily extracted into new features and utilized to improve the accuracy of the disaggregation algorithms. In Fig. 1, we observe two main pattern types in the data, that are specific to *off-days* (weekends and holidays) and to *on-days*, respectively. In the remainder of this work we use two different clustering methods based on $k$-means clustering and a static separation of weekdays and weekends to train distinct models for each group. We explore whether this grouping of the input data leads to better results in terms of disaggregation accuracy compared to a baseline model trained on unpartitioned data.



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

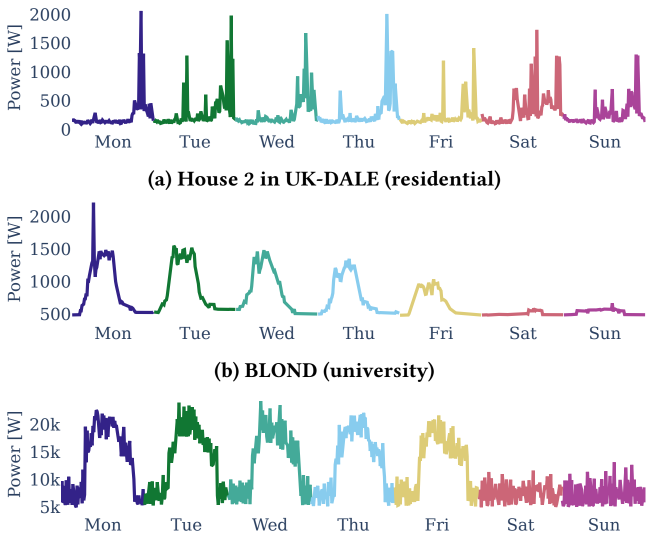


ACM ISBN 978-1-4503-9890-9/22/11…$15.00
https://doi.org/10.1145/3563357.3566154


## 2 RELATED WORK

Considering the time of collection in NILM is not an entirely novel concept. Dinesh et al. have used time-of-day usage patterns of single appliances to improve the appliance identification accuracy in [6]. Similar patterns along with recent activity/inactivity of appliances have been used in [18] to increase the identification accuracy. In [17], the authors identified three temporal features (month, day of week, hour) and provided them along with an Recurrent Neural Network (RNN) into a Feedforward Neural Network (FNN) classifier to improve the detection accuracy of their ensemble model. While the majority of the research efforts in NILM have been focusing on the residential sector [9], studies have also been conducted in industrial settings. They show that it is more challenging to apply typical NILM algorithms to industrial buildings [8, 11]. However, Holmegaard and Kjærgaard show that the disaggregation accuracy in industrial settings can be improved by utilizing day specific training in [8].

Neither of the aforementioned works has conducted analyses of the applicability of NILM methods across datasets from different types and studied the effect of temporal daily clustering on modern




(a) House 2 in UK-DALE (residential)

(b) BLOND (university)

(c) HIPE (industrial)

**Figure 1: Average total consumption for each day of the week in each considered dataset.**

deep NILM models, which motivates our decision to conduct such an in-depth analysis in this work.

## 3 DATA SELECTION

To cater to the generalizable nature of our results, we include three datasets captured in fundamentally different settings (industrial, residential, university), as listed in Table 1. As a side effect of this decision, however, it was not possible to choose the same set of appliances for analysis across all datasets. Therefore, the appliances have been selected based on their activity and average consumption per day in each dataset.

We have selected UK-DALE [13] for the residential dataset because it is among the most-utilized datasets in the community and has been employed in more than 40 studies [9]. The options for datasets from other sectors are limited, although energy consumption in industrial environments accounts for more than a third of the total consumption in the US and around 45 % in Germany, according to [4, 7] respectively. We have hence selected the BLOND [14] and HIPE [5] datasets to represent non-residential settings. BLOND was captured in a university building and HIPE in a factory. Both datasets offer continuous measurements across several consecutive weeks. The selected appliances and their instance ID as defined in their metadata [2] are given as follows. **UK-DALE**: Computer monitor, Rice cooker, Laptop computer, Kettle. **BLOND**: Desktop computer 5, Computer monitor 3, Projector, Laptop computer 3. **HIPE**: Motor 3, Printer, Washing machine, Oven 2.

In order to use the data for our analysis, we have first split the data into separate traces for each day. For each day, we have extracted separate files for each individual appliance under consideration as well as the building aggregate. Afterwards, a harmonization step is applied to remove days on which the aggregated power is lower than the data from any single appliance (physically impossible) or a large number of samples are missing. Through a manual inspection of each resulting trace, days with obviously irregular patterns (e.g., constant power draw for more than one hour) are moreover eliminated. Subsequently, all data have undergone a resampling step and were aligned to regular timestamps, in order to avoid distortion effects in the clustering step.

**Table 1: Selected datasets and their main characteristics.**

| Dataset | Type | Sampling frequency | Duration |
|---|---|---|---|
| UK-DALE [13] (house 2) | Residential | 6s | 234 days |
| BLOND [14] | University | 1s | 49 days |
| HIPE [5] | Industrial | 5s | 92 days |

**Table 2: Number of days per cluster.**

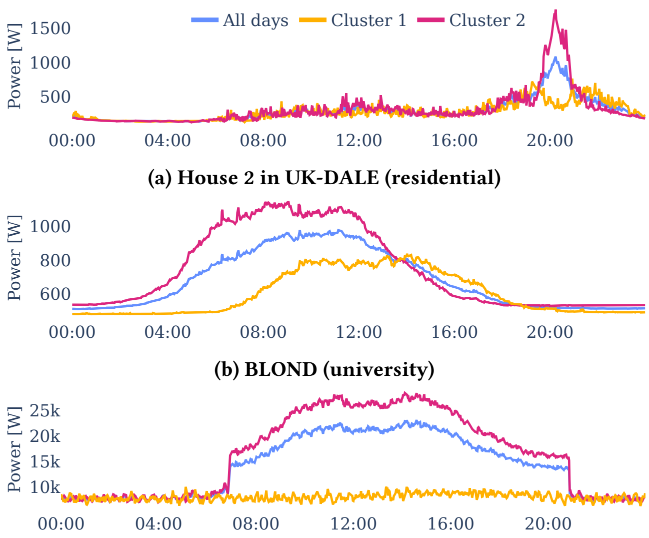
| Dataset | $k$-means | | On/Off | |
|---|---|---|---|---|
| | Cluster 1 (Minority) | Cluster 2 (Majority) | On-days | Off-days |
| UK-DALE | 36 | 68 | 94 | 40 |
| BLOND | 21 | 27 | 32 | 16 |
| HIPE | 16 | 74 | 61 | 29 |

Overall, 134 days worth of data have been used for UK-DALE, 48 days for BLOND, and 90 days for HIPE. In each dataset, the first 70 % of the available data is used for training and the rest for testing, i.e., we have used training sets of 93 (UK-DALE), 33 (BLOND), and 62 (HIPE) days, and considered 41 (UK-DALE), 15 (BLOND), and 28 (HIPE) days for testing.

## 4 DATA CLUSTERING

To cluster each day's aggregated readings into two disjoint groups, we considered two partitioning approaches: (1) $k$-means clustering: using the well-known $k$-means algorithm using $k = 2$, and the Euclidean distance between two traces as the similarity metric. (2) On/off-days: A simple partitioning by day type: Off-days combines weekends and holidays, and the rest are regarded as on-days. The locale of each dataset is used to determine holidays.

Inspecting the daily consumption mean in Fig. 1 indicates that while the activity in the BLOND and HIPE datasets is around the middle of the day, it shifts towards the end of the day in the residential UK-DALE dataset. This fact has been captured by both clustering methods, as we can see in Fig. 2 for $k$-means and Fig. 3 for on/off-days. It is clearly visible that one cluster in each of the three datasets is aligned with lower activity, while the second cluster captures higher activity. The $k$-means methods resulted in a greater number of days in the high activity cluster, as we can see in Table 2. On the other hand, the on/off-days partitioning resulted in a more prominent low activity off-days cluster in the BLOND and HIPE datasets due to their nature, and higher activity in the UK-DALE dataset, as energy consumption in this residential setting is higher on weekends and holidays.

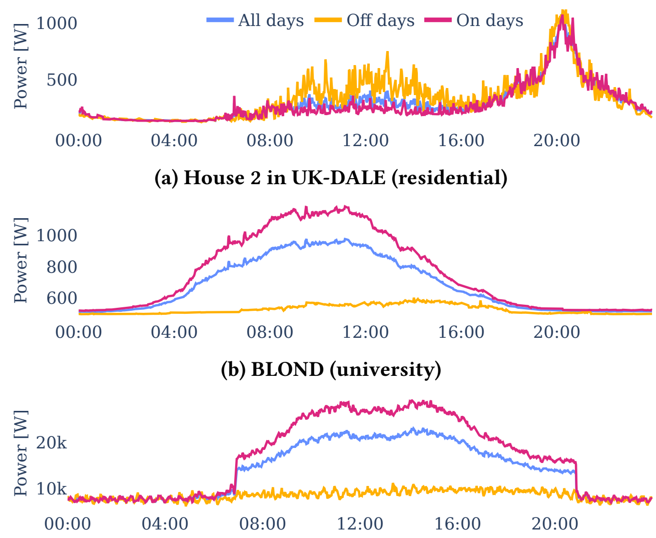



(a) House 2 in UK-DALE (residential)

(b) BLOND (university)

(c) HIPE (industrial)

**Figure 2: Mean consumption by day for each dataset, as well as the daily mean by the $k$-means clustering. Cluster 2 is the majority cluster in all datasets (i.e. the cluster with greater number of days).**




(a) House 2 in UK-DALE (residential)

(b) BLOND (university)

(c) HIPE (industrial)

**Figure 3: Daily mean consumption for each dataset by day type based on on/off-day clustering. Off-days are weekends and holidays, and the rest is on-days.**

## 5 EVALUATION SETUP

### 5.1 Selected Models

As the proposed approach is agonistic to the actual NILM algorithm, we rely on five neural network-based approaches that have gained popularity in recent years [9]: RNN [12], Sequence-to-Sequence (S2S) [12, 19], Sequence-to-Point (S2P) [19], Residual Neural Network (ResNet) [10], and Load Disaggregation with Attention (LDwA) based on Bidirectional LSTM (BiLSTM) [15]. All considered methods were based on the implementations presented in [2, 3, 16].

### 5.2 Clustered Training Setup

Initially, the NILM models must be trained to correlate patterns in the aggregate data with the individual appliances' traces. We tested four different training settings based on the partitioning approaches discussed in Sec. 4:

- $k$-Majority: Only a single model is trained on the majority cluster, based on the $k$-means clustering. This model is then used to disaggregate all data in the test set (which contains data from both clusters). This model is motivated by the fact that days with lower activity (which are omitted in this case) are less likely to include useful patterns for disaggregation.
- $k$-both: Data from both clusters are fed into distinct models during training. During testing, the daily data are passed to the appropriate model based on their cluster. The predictions are then combined from both day-types to produce the final prediction, on which the metrics are computed.
- On-days: Same as $k$-Majority, but the majority cluster in this case is the on-days.
- On/Off-days: Same as $k$-both, but the two clusters are the on-days and off-days.

## 6 RESULTS AND DISCUSSION

In our evaluation, we use the Mean Average Error (MAE) as a metric that quantifies the prediction error at each time step. We chose this metric because it has been widely used in previous NILM studies [1]. The test results as the average MAE for each model and dataset over all appliances are summarized in Table 3. The *Baseline* column represents the average MAE between the ground truth and the predictions on the test set of a single model trained on all training data. The *$k$-Majority* and *On-days* columns represent the performance results of a single model trained only on days from a single group (the majority cluster of $k$-means and the on-days, respectively) and tested on all days in the test set. The *$k$-both* and *On/Off-days* columns correspond to the performance of two models, where one model is trained on the first day group and the second model is trained on the second one. For testing, each day's prediction is generated based on the day group and compared to the ground truth. For the final column, the Baseline model is used to generate the predictions for the on-days in the test set and the model trained on off-days is used to generate the predictions for the remaining test data.

It can be seen that the results are very different depending on the dataset used. Looking at the residential UK-DALE dataset, it can be seen that no clustering method offers consistently low MAE values for all tested models. However, the LDwA model benefits from the On/Off-days clustering by about 20 %, reducing the absolute MAE value from 11.29 W (Baseline) down to 8.98 W. For the BLOND dataset the results are more distinct, as all models trained on the majority $k$-means cluster consistently achieve the lowest MAE

**Table 3: MAE in Watts for each model under each tested setting in all considered datasets, averaged over all selected appliances in the respective dataset. The difference to the baseline model as percentage is provided in parentheses. The best result per setup is in bold. The mean row is the average MAE per setting and dataset.**

| | Model | Baseline | $k$-Majority | $k$-both | On-days | On/Off-days | Baseline & Off-days |
|---|---|---|---|---|---|---|---|
| UK-DALE | S2P | 6.58 | 7.13 (+8.35 %) | **6.22** (-5.47 %) | 6.71 (+1.98 %) | 6.56 (-0.30 %) | 6.63 (+0.78 %) |
| | S2S | 6.22 | **5.94** (-4.50 %) | 6.38 (+2.57 %) | 6.60 (+6.11 %) | 6.63 (+6.59 %) | 6.49 (+4.37 %) |
| | RNN | **6.03** | 6.54 (+8.64 %) | 7.32 (+21.39 %) | 7.28 (+20.73 %) | 7.14 (+18.41 %) | 6.25 (+3.65 %) |
| | ResNet | 5.78 | 6.73 (+16.44 %) | 6.71 (+16.09 %) | **5.65** (-2.25 %) | 5.76 (-0.35 %) | 5.79 (+0.11 %) |
| | LDwA | 11.29 | 11.30 (+0.09 %) | 11.10 (-1.68 %) | 9.13 (-19.13 %) | **8.98** (-20.46 %) | 11.02 (-2.34 %) |
| | Mean | 7.18 | 7.53 (+4.87 %) | 7.55 (+5.15 %) | 7.07 (-1.53 %) | **7.01** (-2.37 %) | 7.24 (+0.84 %) |
| BLOND | S2P | 10.45 | **9.26 (-**11.39 %**)** | 15.21 (+45.55 %) | 11.44 (+9.47 %) | 11.17 (+6.89 %) | 9.88 (-5.49 %) |
| | S2S | 10.04 | **9.12 (-**9.16 %**)** | 15.87 (+58.07 %) | 11.56 (+15.14 %) | 10.64 (+5.98 %) | 10.13 (+0.82 %) |
| | RNN | 9.68 | **8.97 (-**7.33 %**)** | 13.09 (+35.23 %) | 9.38 (-3.10 %) | 9.67 (-0.10 %) | 9.73 (+0.47 %) |
| | ResNet | 14.57 | **8.75 (-**39.95 %**)** | 14.88 (+2.13 %) | 16.80 (+15.31 %) | 14.10 (-3.23 %) | 12.05 (-17.26 %) |
| | LDwA | 10.11 | **9.47 (-**6.33 %**)** | 13.64 (+34.92 %) | 10.37 (+2.57 %) | 11.01 (+8.90 %) | 10.46 (+3.44 %) |
| | Mean | 10.97 | **9.11** (-16.96 %) | 14.54 (+32.54 %) | 11.91 (+8.57 %) | 11.32 (+3.19 %) | 10.45 (-4.74 %) |
| HIPE | S2P | 36.79 | 38.29 (+4.08 %) | 36.48 (-0.84 %) | 40.03 (+8.81 %) | 39.23 (+6.63 %) | **35.08 (-**4.65 %**)** |
| | S2S | 33.95 | 45.71 (+34.64 %) | 34.15 (+0.59 %) | 38.83 (+14.37 %) | 35.56 (+4.74 %) | **32.36 (-**4.68 %**)** |
| | RNN | 39.81 | 49.24 (+23.69 %) | 46.69 (+17.28 %) | 51.12 (+28.41 %) | 45.44 (+14.14 %) | **38.44 (-**3.44 %**)** |
| | ResNet | 38.27 | 33.29 (-13.01 %) | 31.39 (-17.98 %) | 35.90 (-6.19 %) | **31.38 (-**18.00 %**)** | 33.89 (-11.45 %) |
| | LDwA | 45.98 | 58.33 (+26.86 %) | 62.96 (+36.93 %) | 61.57 (+33.91 %) | 60.03 (+30.56 %) | **44.55 (-**3.11 %**)** |
| | Mean | 38.96 | 44.97 (+15.43 %) | 42.44 (+8.93 %) | 45.49 (+16.76 %) | 42.33 (+8.65 %) | **36.86** (-5.39 %) |

values. The highest reductions of up to 39.95 % can be achieved when the ResNet model is used. This is due to the effective model separation of the different daily patterns, which can be seen in Fig. 4. On the HIPE dataset, most models (S2P, S2S, RNN and LDwA) had the best results using the combined approach of the baseline and the off-days clustering. In the case of the ResNet model, the On/Off-days partitioning offers improvements of 18 %. For the On/Off-days partitioning, the best performance was achieved on the UK-DALE dataset in combination with the LDwA model (20.46 %) and on the HIPE dataset in combination with the ResNet model (18 %).

With respect to the two different methods of clustering training data that we considered in this study, we find that the MAE of predictions generated by models trained on some partitioning of the data between on-days and off-days indicated by the last three columns in Table 3 is, on average, 5.04 % lower than models trained on $k$-means clustered data. The average MAE of single models trained on a subset of the data in $k$-Majority and On-days increased by 4.98 % compared to two models trained on different data clusters in $k$-both, On/Off-days, and Baseline & Off-days settings. However, models trained on two clusters require twice as much storage space as a single model, which can be critical for memory-limited embedded applications.

These results indicate that clustered training by day on non-residential energy data improves the performance of popular deep NILM models by 11.18 % on average using the appropriate clustering methods, as measured by the MAE, compared to a model trained on the entire dataset. Also, in the case of office environments, the need to train a model on the entire data is eliminated, simultaneously increasing model accuracy and speeding up the training process.

In regard to the training duration, we found out that the simultaneous training of two models (as in on/off-days and $k$-both settings) is on average 1 % slower than training a single model on the same dataset, while training a single model on only the majority subset of the data was 16.11 % faster than training the same model on the full dataset in our evaluations.

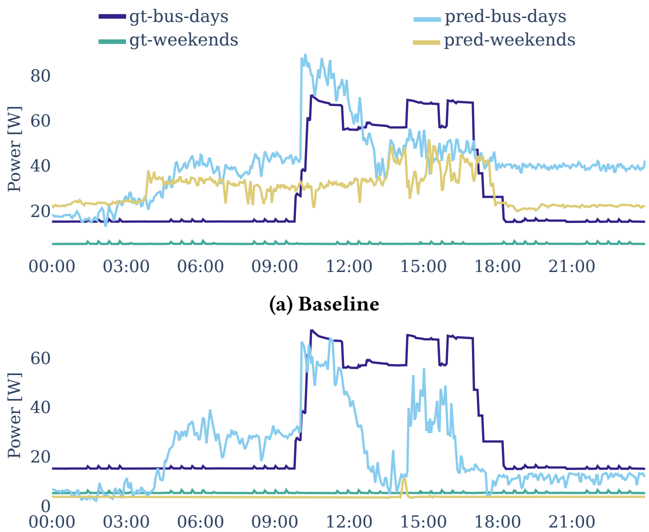
## 7 CONCLUSION

In this paper, we have examined the effect of clustering the training data based on daily patters on the performance of multiple state-of-the-art NILM models. We have shown that a partitioning of training data based on daily patterns can improve the disaggregation performance for energy data with different activity characteristics on weekends. The use of days in the majority group of $k$-means clustering to train models on data from office environments that follow a similar pattern to BLOND can improve the performance of any of the considered models. In the industrial setting, the model convergence can be accelerated and improved by utilizing different models for business and non-business days. Only residential data that show clear trends based on the time-of-day rather than the day-of-week (like UK-DALE) did not show any benefits from the partitioned training we have evaluated.

## 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* (2022).

[2] Nipun Batra, Jack Kelly, Oliver Parson, Haimonti Dutta, William Knottenbelt, Alex Rogers, Amarjeet Singh, and Mani Srivastava. 2014. NILMTK: an open source toolkit for non-intrusive load monitoring. In *Proceedings of the 5th international conference on Future energy systems.*




**(a) Baseline**

**(b) With $k$-Majority partitioning**

**Figure 4: Comparing the averaged predictions of the ResNet model for the computer monitor in BLOND (abbreviated as "pred") against the ground truth ("gt") between business days ("bus-days") and weekends for the two different training settings: Baseline and $k$-Majority, over all days in the corresponding group. We can clearly see, how the model trained on partitioned data has for the most part captured the fact that the appliance usage on weekends has little to no activity, while in the baseline setting, the model is still generating erroneous patterns.**


[3] Nipun Batra, Rithwik Kukunuri, Ayush Pandey, Raktim Malakar, Rajat Kumar, Odysseas Krystalakos, Mingjun Zhong, Paulo Meira, and Oliver Parson. 2019. Towards reproducible state-of-the-art energy disaggregation. In *Proceedings of the 6th ACM international conference on systems for energy-efficient buildings, cities, and transportation.*

[4] BDEW 2021. *Stromverbrauch in Deutschland nach Verbrauchergruppen 2020.* https://www.bdew.de/service/daten-und-grafiken/stromverbrauch-deutschland-verbrauchergruppen/

[5] Simon Bischof, Holger Trittenbach, Michael Vollmer, Dominik Werle, Thomas Blank, and Klemens Böhm. 2018. HIPE – an Energy-Status-Data Set from Industrial Production. In *Proceedings of ACM e-Energy.*

[6] Chinthaka Dinesh, Stephen Makonin, and Ivan V Bajic. 2017. Incorporating time-of-day usage patterns into non-intrusive load monitoring. In *Proceedings of IEEE Global Conference on Signal and Information Processing (GlobalSIP).*

[7] EIA 2022. *Total Energy Monthly Data - U.S. Energy Information Administration.* https://www.eia.gov/totalenergy/data/monthly/index.php

[8] Emil Holmegaard and Mikkel Baun Kjærgaard. 2016. NILM in an Industrial Setting: A Load Characterization and Algorithm Evaluation. In *Proceedings of IEEE International Conference on Smart Computing (SMARTCOMP).*

[9] Patrick Huber, Alberto Calatroni, Andreas Rumsch, and Andrew Paice. 2021. Review on Deep Neural Networks Applied to Low-Frequency NILM. *Energies* (2021).

[10] Ziyue Jia, Linfeng Yang, Zhenrong Zhang, Hui Liu, and Fannie Kong. 2021. Sequence to point learning based on bidirectional dilated residual network for non-intrusive load monitoring. *International Journal of Electrical Power & Energy Systems* (2021).

[11] Florian Kalinke, Pawel Bielski, Snigdha Singh, Edouard Fouché, and Klemens Böhm. 2021. An Evaluation of NILM Approaches on Industrial Energy-Consumption Data. In *Proceedings of the Twelfth ACM International Conference on Future Energy Systems.*

[12] Jack Kelly and William Knottenbelt. 2015. Neural NILM: Deep Neural Networks Applied to Energy Disaggregation. In *Proceedings of the 2nd ACM international conference on embedded systems for energy-efficient built environments.*

[13] Jack Kelly and William Knottenbelt. 2015. The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes. *Scientific data* (2015).

[14] Thomas Kriechbaumer and Hans-Arno Jacobsen. 2018. BLOND, a building-level office environment dataset of typical electrical appliances. *Scientific data* (2018).

[15] Veronica Piccialli and Antonio M Sudoso. 2021. Improving Non-Intrusive Load Disaggregation through an Attention-Based Deep Neural Network. *Energies* (2021).

[16] Hetvi Shastri and Nipun Batra. 2021. Neural network approaches and dataset parser for NILM toolkit. In *Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation.*

[17] Junfei Wang, Samer El Kababji, Connor Graham, and Pirathayini Srikantha. 2019. Ensemble-Based Deep Learning Model for Non-Intrusive Load Monitoring. In *Proceedings of IEEE Electrical Power and Energy Conference (EPEC).*

[18] Shirantha Welikala, Chinthaka Dinesh, Mervyn Parakrama B Ekanayake, Roshan Indika Godaliyadda, and Janaka Ekanayake. 2017. Incorporating Appliance Usage Patterns for Non-Intrusive Load Monitoring and Load Forecasting. *IEEE Transactions on Smart Grid* (2017).

[19] Chaoyun Zhang, Mingjun Zhong, Zongzuo Wang, Nigel Goddard, and Charles Sutton. 2018. Sequence-to-Point Learning With Neural Networks for Non-Intrusive Load Monitoring. In *Proceedings of the AAAI conference on artificial intelligence.*