On the Impact of Temporal Data Resolution on the Accuracy of Non-Intrusive Load Monitoring

Jana Huchtkoetter
TU Clausthal, Germany
jana.huchtkoetter@tu-clausthal.de

Andreas Reinhardt
TU Clausthal, Germany
reinhardt@ieee.org

ABSTRACT

Many approaches to perform Non-Intrusive Load Monitoring, i.e., to disaggregate electrical load curves collected at a single measurement point, have been presented in literature. The largely different characteristics of the datasets used to evaluate newly proposed disaggregation algorithms, however, complicate an objective and comparable assessment of their capabilities. Different temporal resolutions of the input data (i.e., different sampling rates) are a major impediment to the comparative evaluation of load disaggregation methods in particular. We hence investigate the impact of the temporal data resolution on the disaggregation results of three state-of-the-art algorithms in this work. Our study not only confirms that temporal resolution has an impact on load disaggregation accuracy, but also highlights that a favourable low-frequency sampling rate exists for the appliances under consideration, and generally falls within the range from 1 Hz to 1/30 Hz.

CCS CONCEPTS

• **Hardware** → **Power estimation and optimization**.

KEYWORDS

Non-Intrusive Load Monitoring, Dataset Characteristics, Sampling Rate, Dataset Comparability, Comparative Evaluation

ACM Reference Format:

Jana Huchtkoetter and Andreas Reinhardt. 2020. On the Impact of Temporal Data Resolution on the Accuracy of Non-Intrusive Load Monitoring. In *The 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '20), November 18–20, 2020, Virtual Event, Japan.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3408308.3427974

1 INTRODUCTION

Non-Intrusive Load Monitoring (NILM) is an active and growing research area. The scientific challenge of disaggregating the load curve collected at a single metering point, however, has not been fully solved to date. Comparative studies of newly proposed disaggregation methods (like [3, 16]) are thus vital to identify the underlying algorithmic methods that bear the greatest potential. Comparative assessments of NILM performance often rely on published electrical load signature datasets (such as REDD [12] or BLUED [1]). However, such datasets are often heterogeneous in multiple dimensions, e.g., their sampling rates, the degrees of building instrumentation, and/or their geographical diversity. These differences have been shown to impede cross-dataset evaluations [11], and motivated studies that systematically consider the impact of such parameters on the disaggregation process [3, 13, 14].

The impact of the data sampling rate on NILM performance has been evaluated in both theoretical models and empirical studies. The authors of [4] explore the fundamental limits of NILM by providing real-world data to a theoretically optimal disaggregation algorithm. A monotonous decline of recognition accuracy for decreasing sampling rates is determined, and confirmed by means of three practical studies. The authors of [8] evaluate the impact of temporal resolutions using data sampled at frequencies greater than the AC mains frequency, revealing a favourable sampling rate between 922 Hz and 1.2 kHz for high-frequency NILM. The impact of data granularity has also been studied in the context of user privacy protection. Results presented in [5] confirm that increased sampling intervals lead to decreasing accuracy levels, and thus a better privacy protection. A supplementary study in [6] identifies varying levels of disaggregation accuracy for different devices, with the area of greatest sensitivity located between 3 s and 60 s. The surveyed works unambiguously conclude that there is a reduction in appliance recognition accuracy with decreasing sampling rates. The extent to which NILM is impacted by such reductions, however, depends on the trade-off between the experienced information loss due to the lower temporal resolution and the potential information gain that emerges as a result of the increased time span covered by the same number of samples.

Accordingly, the study presented in this paper seeks to evaluate the relation between data sampling rates and the corresponding load disaggregation results, irrespective of the input dataset used. By running a set of from several load disaggregation studies and interpreting the results, we target to determine a favourable sampling rate range for load disaggregation. Note that we only consider *low-frequency* NILM in our evaluations, i.e., the use of data collected using sampling rates below the mains frequency.

2 EVALUATION SETUP

The objective of this study is to investigate how data sampling rates impact NILM results. As such, we consider 168 combinations of disaggregation algorithms, input datasets, and sampling rates to enable a generalisation of the findings. The experimental setup and methodology we have followed are documented as follows.

2.1 Disaggregation Algorithm Selection

In this study, we compare the disaggregation performance of the Recurrent Neural Network (RNN), Denoising Autoencoder (DAE),

BuildSys '20, November 18–20, 2020, Virtual Event, Japan

ACM ISBN 978-1-4503-8061-4/20/11…$15.00
https://doi.org/10.1145/3408308.3427974

Table 1: Appliances evaluated for each building

Appliance	UK-Dale 1	UK-Dale 5	REDD	DRED
fridge	✓	✓	✓	✓
washer dryer	✓	✓	✓	✓
microwave	✓	✓	✓	✓
dish washer	✓	✓	✓	*X*
lights	✓	*X*	✓	*X*
laptop computer	✓	*X*	*X*	✓

and Sequence-to-Point (Seq2Pt) algorithms, which were originally presented in [9, 18]. The methods were chosen due to their good performance in recent evaluative studies [2, 16]. The Combinatorial Optimization (CO) method [7] was chosen as a baseline algorithm for comparison. All evaluations are conducted using the NILMTK disaggregation framework [2]. The RNN, DAE, and Seq2Pt were trained for 50 *epochs* using a *batch size* of 1024.

2.2 Metric Selection

To measure disaggregation performance, two metrics are used in this work. The F1 score is the harmonic mean of precision and recall. It allows for a direct comparison of disaggregation results and is independent of the power consumption of the appliances under consideration. The Mean Absolute Error (MAE), in turn, allows us to determine to what extent the energy consumed by an appliance was accurately detected. While a F1 score of 1.0 indicates a perfect disaggregation result, a minimal MAE indicates the best disaggregation success.

2.3 Input Data Selection

For our study, we rely on data from existing load signature datasets, more specifically buildings 1 and 5 of UK-Dale [10], building 1 of REDD [12], and building 1 of DRED [17]. These buildings were chosen as they all offer data from similarly developed countries (USA, UK, NL), collected during the same timeframes. All three datasets provide mains readings collected at a rate of 1 Hz, which is further reduced as described in Section 2.5.

2.4 Appliance Selection

Related research has shown that different disaggregation results can be expected for different devices [2, 5, 16]. Thus, it is necessary to consider multiple appliances in order to get comparable and generalisable results. Hence, devices common to at least three of the buildings are selected. As shown in Table 1, these devices are the fridge, washer dryer, microwave, and dish washer. The buildings also include smaller consumers, two of which we have additionally included in our study, as marked in the table.

2.5 Reduction of Temporal Resolution

We apply downsampling through value omission to reduce the temporal resolution of the data. This is the default mode of operation of metering devices reporting low-frequency data. We have selected a total number of 14 sampling periods from 1 Hz to 1/900 Hz, with increasing stride lengths in-between successive values (cf. Fig. 1).

3 EVALUATION RESULTS

In the following paragraphs we first present the baseline results when using the *native sampling rate*, i.e., 1 Hz and consider the impact of general differences in the input data. This is followed by the presentation of the disaggregation results across the chosen temporal resolutions.

3.1 Baseline Values

In a first evaluation run, the F1 scores and MAE values for the baseline analysis (i.e., the use of data at the native sampling rate) were determined. The microwave in UK-Dale 5 has been excluded from the table and all further experiments, as it was found to exhibit a constant power draw. We have added these baseline values as the leftmost element to all evaluations results (cf. Fig. 1) in order to show the impact of reduced sampling rates more clearly.

3.2 Impact of Temporal Data Resolution

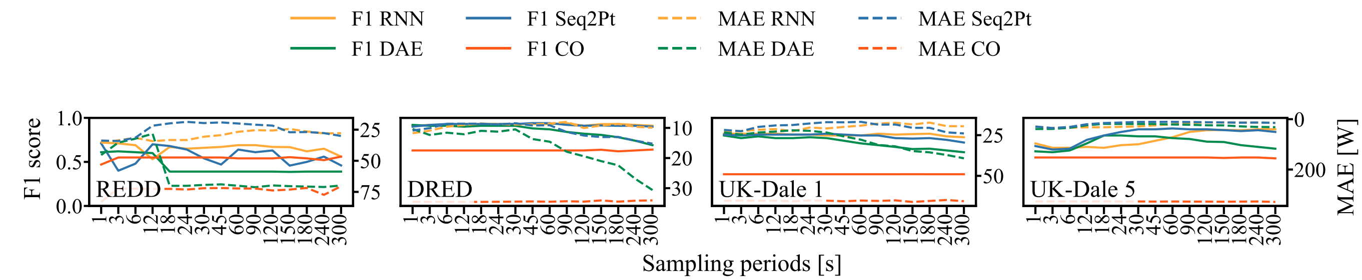
We visualise the disaggregation results for the changing temporal resolution in Fig. 1 By plotting the MAE score on an inverted (right-hand side) axis, the best disaggregation performance is given when both metrics yield the highest value on the y-axis. For the sake of visual clarity, we exclude the MAE in the graphs unless a notably different trend exists. Three main observations can be distinguished.

(1) Results for some devices meet the expectations (i.e., information loss with decreased temporal resolution): All microwaves, all but one washing machine, and two fridges exhibit declining disaggregation results (cf. Fig. 1).
(2) Some devices show no or very low impact from the changing temporal resolution, such as the lights or laptops (cf. Figs. 1e and 1f). They are largely resilient to sampling rate variations.
(3) Appliances exist that exhibit improvements when the sampling rates are lowered. This is the case for all considered dish washers (cf. Fig. 1d, the fridges in REDD and UK-Dale (cf. Fig. 1a), and the washing machine in DRED (cf. Fig. 1c). For the Seq2Pt and DAE these improvements continue until sampling periods between 30 s and 90 s, after which they begin to decrease again. The improvement for the RNN even continues to the longest considered sampling period of 300 s.

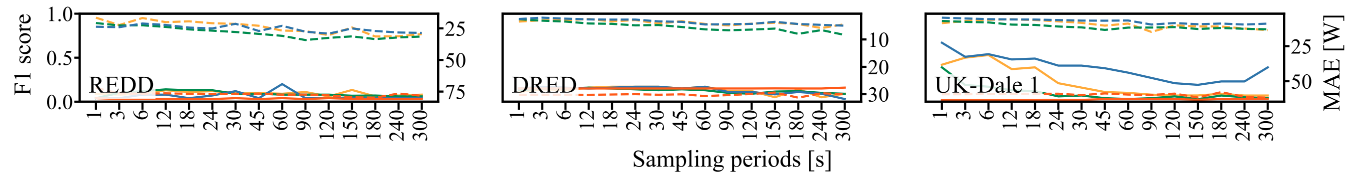
3.3 Discussion of the Results

Let us look at possible causes for increasing F1 scores (or decreasing MAE levels) next. As stated in Section 1 and reported in [15], the duration of historical data available to the NILM algorithms is increased through the application of downsampling. The neural network-based NILM methods we have considered have an input size of 1024 samples. This represents 17 min at the *native* resolution. Reducing the resolution increases the time window to a multiple of this value (from 51 min for a downsampling factor of 3, up to an effective input window length of 85 h when retaining only every 300^{th} sample). It thus appears to be advantageous for some long-running appliances to provide the disaggregation algorithm with data containing information for up to a day of data.

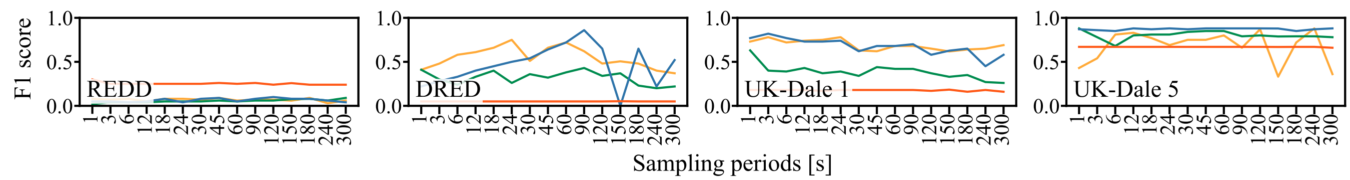
While this insight provides one possible explanation for the unexpected behaviour, no general trend across all device types becomes apparent. To investigate this further, we have manually compared characteristics of the washer dryer appliances in UK-Dale



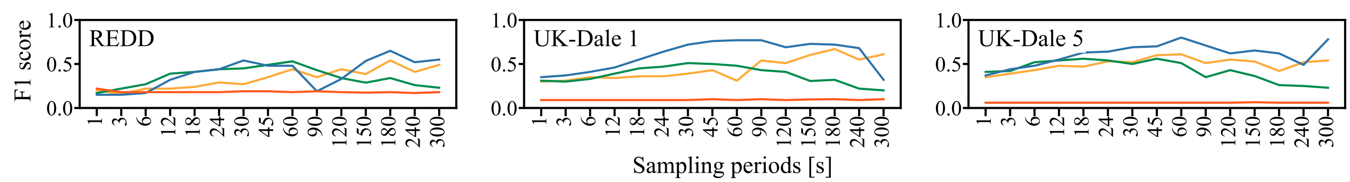
(a) MAE and F1 scores for the fridges



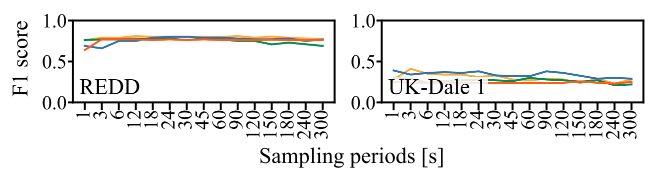
(b) MAE and F1 scores for the microwaves



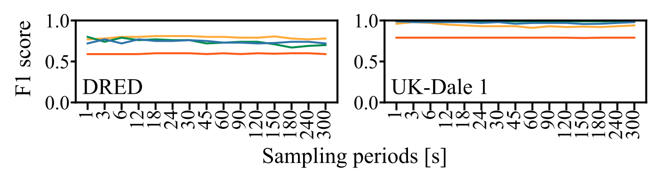
(c) F1 scores for the washer dryer machines



(d) F1 scores for the dish washers



(e) F1 scores for the lights



(f) F1 scores for the laptops

Figure 1: Disaggregation scores (F1 and MAE) for the electrical appliances under consideration

houses 1 and 5 to the characteristics of the dish washers in the same building. The amount of energy consumed is comparable, as well as the general pattern of consumption consisting of phases with high consumptions (greater than 1500 W) connected through phases of lower consumption (less than 750 W). The only difference common to both washer dryers is the higher fluctuation of consumption magnitude during both phases. Considering that improvements can be seen for the washer dryer in DRED, which in term presents itself with a level of fluctuation between the previously considered washer dryers and the dish washers, reveals an interdependence between improvement potentials due to the amount of historical data and further device characteristics.

Other than that, the overall trend of the F1 scores and MAE levels confirms that reduced temporal resolutions do not lead to significant improvements for most household appliances. At the same time, minor sampling rate reductions (up to 1/30 Hz) have no significant negative impact either.

3.4 Observations and Insights

The findings of this evaluation study provide relevant insights concerning load monitoring processes, which should be considered in future research on load disaggregation methods.

First, our results indicate that data resampled to 1/30 Hz might be sufficient to run NILM at a high accuracy. At lower sampling frequencies, however, the onset of strong F1 and MAE deteriorations becomes apparent. Furthermore, sampling rates in this range allow the NILM algorithms to profit from additional information pertaining to the past power consumption of some long-running appliances. Using a common sampling rate in NILM evaluations would allow for the improved comparability of results, and to profit from the correspondingly lowered execution times and memory requirements.

Second, our study confirms that large variations of disaggregation results can be observed when considering the same appliance, yet using data from different datasets. For example, disaggregation of the lights from UK-Dale house 1 showed consistently lower scores than disaggregating the same appliance from REDD, as visible in Fig. 1e. The same effect exists when considering the same appliance from the same dataset, yet using data of a different temporal resolution (as seen for, e.g., the fridge in REDD in Fig. 1a). Still, we could show that resampling data from multiple datasets to a common sampling rate retains power consumption features characteristic important for their successful disaggregation in a comparable way, as visible through the comparable reaction of the dish washers whose further characteristics are similar across the three considered devices (cf. Fig. 1d).

Third and lastly, concerning the algorithms under evaluation, this work determined that the DAE method is more sensitive to lower data resolutions, while the RNN and Seq2Pt methods partially even profited from it. We consider these different reactions to be a result of the aforementioned trade off: The algorithms can make use of the additional historical data to different degrees, whereas all algorithms are subjected to the same loss of information content through the lowered resolution. Additionally, the good results achieved on all devices for both UK-Dale buildings and the high disaggregation performance and resilience to sampling rate variations for laptop computers and light installations indicate that further data characteristics, such as the uniqueness of the consumption trace, very likely have a positive influence on the disaggregation results.

4 CONCLUSIONS

Non-Intrusive Load Monitoring follows the objective to infer the operation of individual appliance from the aggregate power consumption data of a building. There is still no agreement on a widely acknowledged methodology for the collection of the underlying consumption data, however. In particular, the sampling rates of publicly available datasets differ widely with regard to the temporal resolution at which data have been captured. Through a study that uses state-of-the-art algorithms implemented in the NILMTK framework and data from four different datasets, we have evaluated the impact of the temporal data resolution and gained two major insights of practical relevance. First, lower sampling rates may not always result in degraded disaggregation results, but in fact some results (chiefly those of long-running appliances) did improve. Second, favourable low-frequency sampling rates range between 1 Hz and 1/30 Hz. In fact, a sampling rate of 1/30 Hz seems to be sufficient to fully exploit the positive effects of lowered execution times and memory usage, in addition to offering more historical data to NILM algorithms, at the expense of only slightly degraded scores.

ACKNOWLEDGMENTS

This research was supported by Deutsche Forschungsgemeinschaft (DFG) grant no. RE 3857/2-1. We thank the Simulationswissenschaftliches Zentrum Clausthal-Göttingen (SWZ) for financial support.

REFERENCES

[1] Kyle Anderson, Adrian Ocneanu, Diego Benítez, Derrick Carlson, Anthony Rowe, and Mario Bergés. 2011. BLUED: A Fully Labeled Public Dataset for Event-based Nonintrusive Load Monitoring Research. In *2nd Workshop on Data Mining Applications in Sustainability.*

[2] 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.*

[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.*

[4] Roy Dong, Lillian Ratliff, Henrik Ohlsson, and Shankar Sastry. 2014. Fundamental Limits of Nonintrusive Load Monitoring. *Proceedings of the 3rd International Conference on High Confidence Networked Systems.*

[5] Günther Eibl and Dominik Engel. 2014. Influence of Data Granularity on Nonintrusive Appliance Load Monitoring. In *Proceedings of the 2nd ACM Workshop on Information Hiding and Multimedia Security.*

[6] Günther Eibl and Dominik Engel. 2014. Influence of Data Granularity on Smart Meter Privacy. *IEEE Transactions on Smart Grid* 6, 2 (2014).

[7] George Hart. 1985. *Prototype Nonintrusive Appliance Load Monitor.* Technical Report. MIT Energy Laboratory and Electric Power Research Institute.

[8] Jana Huchtkoetter and Andreas Reinhardt. 2019. A Study on the Impact of Data Sampling Rates on Load Signature Event Detection. *Energy Informatics* 2, Supplement 1 (2019).

[9] 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.*

[10] 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* 2, 1 (2015).

[11] Christoph Klemenjak, Andreas Reinhardt, Lucas Pereira, Mario Bergés, Stephen Makonin, and Wilfried Elmenreich. 2019. Electricity Consumption Data Sets: Pitfalls and Opportunities. In *Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation.*

[12] Zico Kolter and Matthew Johnson. 2011. REDD: A Public Data Set for Energy Disaggregation Research. In *Proceedings of the 2nd KDD Workshop on Data Mining Applications in Sustainability.*

[13] Christoforos Nalmpantis and Dimitris Vrakas. 2019. Machine Learning Approaches for Non-Intrusive Load Monitoring: From Qualitative to Quantitative Comparation. *Artificial Intelligence Review* 52 (2019).

[14] 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* (2018).

[15] Andreas Reinhardt and Mazen Bouchur. 2020. On the Impact of the Sequence Length on Sequence-to-Sequence and Sequence-to-Point Learning for NILM. In *Proceedings of the 5th International Workshop on Non-Intrusive Load Monitoring.*

[16] Andreas Reinhardt and Christoph Klemenjak. 2020. How does Load Disaggregation Performance Depend on Data Characteristics?: Insights from a Benchmarking Study. In *Proceedings of the 11th ACM International Conference on Future Energy Systems.*

[17] Akshay Uttama Nambi, Antonio Reyes Lua, and Venkatesha Prasad. 2015. LocED: Location-aware Energy Disaggregation Framework. *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments.*

[18] Chaoyun Zhang, Mingjun Zhong, Zongzuo Wang, Nigel Goddard, and Charles Sutton. 2016. Sequence-to-Point Learning with Neural Networks for Nonintrusive Load Monitoring. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence.*