22nd International Conference on Renewable Energies and Power Quality (ICREPQ'24)
Bilbao (Spain), 26th to 28th June 2024
Renewable Energy and Power Quality Journal (RE&PQJ)
ISSN 2172-038 X, Volume No.22, September 2024

Optimising energy efficiency enhancing NILM through high-resolution data analytics

C. Rodriguez-Navarro[1], F. Portillo[1], L. Castro-Santos[2], A. Filgueira-Vizoso[3], F. G. Montoya[1] and A. Alcayde[1]

[1] Departamento de Ingeniería
Escuela Superior de Ingeniería, Universidad de Almería
Carretera Sacramento s/n, La Cañada de San Urbano, 04120 Almeria (Spain)

[2] Departamento de Enxeñaría Naval e Industrial
Centro de Investigación en Tecnoloxías Navais e Industriais (CITENI), Escola Politécnica de Enxeñaría de Ferrol, Universidade da Coruña
Campus Industrial de Ferrol, Esteiro, 15471 Ferrol (Spain)

[3] Departamento de Química
Centro de Investigación en Tecnoloxías Navais e Industriais (CITENI), Escola Politécnica de Enxeñaría de Ferrol, Universidade da Coruña
Campus Industrial de Ferrol, Esteiro, 15471 Ferrol (Spain)

Abstract. This study introduces a novel approach to enhancing energy efficiency by building a high-resolution dataset for Non-Intrusive Load Monitoring, addressing the challenges of monitoring a wide range of devices. To achieve optimal energy efficiency, it is essential to have advanced monitoring of electrical variables. In this context, the second version of openZmeter is presented, which supports up to 4 devices, each capable of 160 measurements, including voltage, current and power harmonics on each channel. To carry out this purpose, the Non-Intrusive Load Monitoring Toolkit is adapted for the new openZmeter v2. The main objective of this study is to offer a new pragmatic approach to generating artificial intelligence models to optimise energy use, analysing the results through an exhaustive experimental analysis under various casuistry and conditions. Numerous comparative studies are provided using classical disaggregation algorithms with different requirements. Conclusively, the research emphasises the transformative potential of artificial intelligence for energy efficiency strategies, offering insights for scholars and practitioners.

Key words. Energy efficiency, NILM, high-resolution dataset.

List of Abbreviations. ANN (Artificial Neural Networks); CO (Combinatorial Optimization); DDSC (Direct Discriminant Component Analysis); EAE (Error in Allocated Energy); FHMM (Factorial Hidden Markov Model); HMM (Hidden Markov Models); ILM (Intrusive Load Monitoring); MNEAP (Mean Normalized Error in Assigned Power); NILMTK (Non-Intrusive Load Monitoring Toolkit); OBSA (Online Binary Search Algorithm); oZm (openZmeter); OMPM (Open Multi Power Meter); PSO (Particle Swarm Optimisation); RMSE (Root Mean Square Error); SVM (Support Vector Machines).

1. Introduction

In the current energy transition scenario, Non-Intrusive Load Monitoring (NILM) [1] has seen an increase in demand compared to Intrusive Load Monitoring (ILM). NILM efficiently optimises energy consumption and makes it possible to estimate the individual consumption of electrical devices connected to an installation from a single centralised meter. It eliminates the need for individualised meters for each application, a feature in ILM techniques.

The Non-Intrusive Load Monitoring Toolkit (NILMTK) [2] has been adapted to integrate the new oZm v2, an advanced three-phase power meter and power quality analyser with IoT capabilities. Therefore, implementing this technique is facilitated by using the NILMTK and the latest version of the oZm, which can support up to 4 channels per meter, simplifying the acquisition process.

This procedure may seem simple, thanks to the tools implemented at the software and hardware level. Still, it is crucial to consider the various variables that can influence the quality of the generated model [3]. This article aims to provide a clear overview of the generation of energy consumption disaggregation models, regardless of the algorithm used and its associated parameters.

2. Related Work

Many disaggregation algorithms and datasets are available for load disaggregation in the literature. Regarding the algorithms for NILM [4], these methods are categorised into three main groups: optimisation methods (such as SVM, OBSA [5], genetic algorithms [6] and PSO [7]), supervised methods (such as Bayesian Classifiers [8], SVM [9], DDSC [10], ANN [11] and their extensions) and, finally, unsupervised methods (including CO [12], HMM and its extensions such as FHMM [13]), the latter being the approach chosen for the present work.
Datasets for power disaggregation can be diverse, with adequate sampling resolution, and accessible to the research community. Here are some of the most current datasets:

- DEPS (Higher Polytechnic School of the University of Seville): Offers power, voltage, and current readings at 1 Hz through six devices present in a classroom, taken during a month [14].
- DSUALMH (University of Almeria): offers up to 160 data of electrical measurements captured with the oZm v1 of six commonly used devices plus the aggregate in about two hours [15].
- DSUALM10H (University of Almeria): offers up to 160 pieces of data referring to ten devices plus the aggregate with a time of about four hours.
- iAWE (Indraprastha Institute of Information Technology): provides aggregated and submetered electricity and gas data from 33 household sensors, with a 1-second resolution for 73 days from a single household [16].
- UALM2 (University of Almeria): dataset generated with the measurements of the free OMPM hardware based on an RS485 bus that allows up to 127 devices. In this implementation, the design is carried out for six devices: five are low-power applications, and one corresponds to the aggregate [17].

3. Methodology

This study takes advantage of the features of the new oZm v2 [18] that complies with standards such as IEC 61000-4-30 and EN 50160 and can measure single or three-phase systems (85-264VAC) in a single device with an accuracy of 0.1%. Acquisition features include 12- or 13-bit resolution and 24 kHz sampling rate. Various current probes, such as current transformers, Hall effect, Rogowski and others, can be used as they comply with the input voltage levels.

In the experiments between 2022 and 2023, three oZm v2 were used to acquire data from twelve measurement channels, with one channel dedicated to aggregate measurements, as shown in Fig. 1. Ten electrical devices were used in different pseudo-random situations, such as different on/off sequences, different time intervals, different time slots, and other conditions.

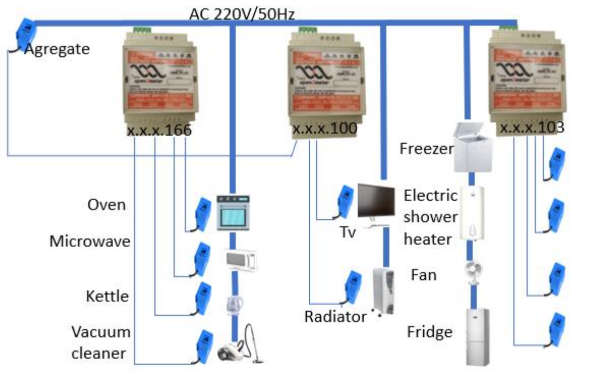


Fig. 1. oZm’s connections with the applications.

The models use several hours of device operation logs obtained through the oZm API. The data collected by the oZm are stored in files consisting of 160 data fields with a 13-digit timestamp mark, which will be used in successive stages of the NILM.

A. *NILMTK flowchart*

NILMTK is a popular free, open-source tool that simplifies NILM research using converters, evaluation metrics, algorithms, and essential resources. For this reason, in the disaggregation process, NILMTK was used together with the oZm v2. A flow diagram is shown in Fig. 2.

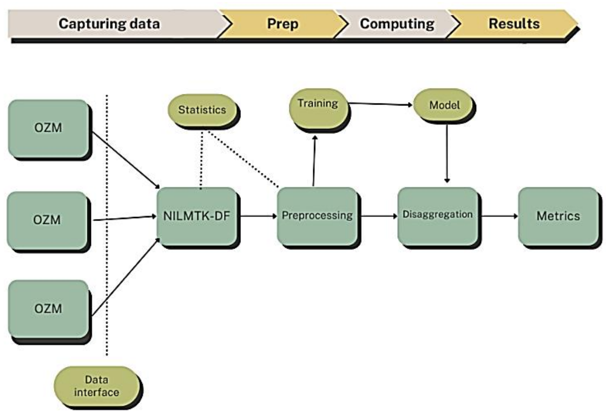


Fig. 2. NILMTK flow diagram.

New preconverters and converters have been developed specifically for the new oZm v2 measurements, and the associated metadata has been integrated into new HDF5-formatted datasets.

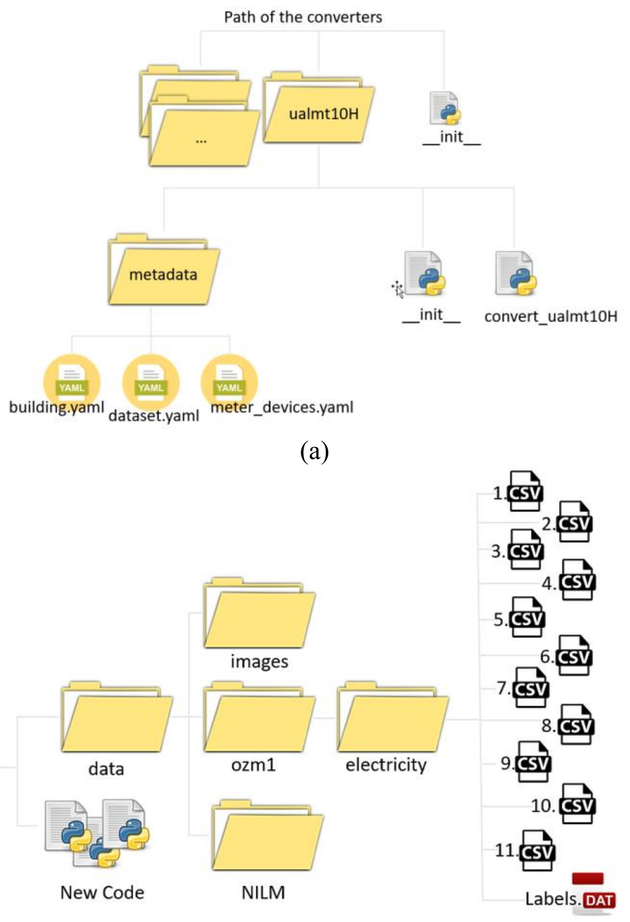
B. *Generation of the DSUALM10H dataset*

The initial stage involves a preliminary analysis of the data files, decompressing measurement files from parquet format to CSV format, adding headers, and converting angular values from harmonics to modules. The discrepancy in the date and time fields returned by the oZm is addressed. The CSV files are then rearranged, and extra characters are removed.

The next step involves converting the pre-processed files and their metadata into a unified HDF5 file stored in the runtime directory. Due to the unique nature of oZm data, a new converter integrated into NILMTK is created, with a subdirectory for metadata in YAML format.

Locating data files involves calling the converter, with the metadata path and the new name of the dataset as parameters. The converter performs several stages for each measurement: reading the numbered file, converting the timestamp field to date format, loading the remaining columns, sorting and indexing, resampling, and re-indexing the file. After processing all the files, an attempt is made to combine them in YAML format, add metadata, and generate a new dataset in HDF5 format.

Fig. 3a shows the converter configuration, and Fig. 3b shows the directory structure. Each CSV file derived from oZm in the previous phase is numbered. The new function accesses measured data files in the "/electricity/" input folder using .csv file tags.



(a)

(b)

Fig. 3. (a) Data file structure. (b) Metadata file structure.

C-. Analysis, training, and validation

After generating a new dataset, the initial diagnosis uses NILMTK implementations, focusing on the power profile. At this stage, the voltage profile is also determined, possible missing sections are identified, or samples with shallow values are filtered. After data analysis, the set is segmented into training, validation, and testing.

D-. Disaggregation

In the NILMTK framework, two disaggregation models, specifically the CO and FHMM algorithms, are employed to disaggregate new datasets based on the active power data of devices. This process entails, first and foremost, the loading of essential libraries, followed by the definition of the dataset structure, the association of labels with the respective devices, and the delineation of the training subset. Upon establishing the training model, the CO and FHMM algorithms are executed across various time intervals (10", 30", 60", 5', 10', 15') using three different methods (First, Mean, and Median), with the resulting models being saved in H5 format. The subsequent step involves deploying the highest-rated model in the validation phase to facilitate a comparison between the actual signal and the predictions made by the top-performing model from each dataset.

4. Results

When disaggregating several applications, an evaluation is carried out with some NILMTK tools, considering metrics such as F1-score, EAE, MNEAP, and RMSE. Significant challenges are recognised with the increase of household appliances, in our case, up to 10 applications plus the aggregate, in contrast to DSUALMH, which, indeed, with only six more applications, the aggregate achieved excellent metrics [15].

Fig. 4 compares the key metrics with incremental improvements made over the measurements, including increased sampling time, reduction in electrical noise, change in transducers, and progress in the alignment of active phases in split-core transducers, revealing clear signs of improvement.

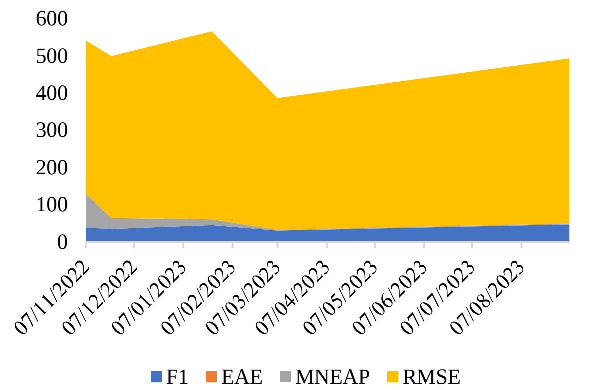


Fig. 4. Evolution of metrics in successive experiments.

When increasing the number of devices in training with the CO algorithm, there are few problems, even with minimal sampling times. Still, with FHMM, running the algorithm with the new dense dataset can be challenging, for which it is necessary to reduce the training time or sampling time (or even the number of devices).

A. Influence of sample time extension and harmonics inclusion

On the other hand, extending the training period by two hours does not necessarily improve the basic metrics, as evidenced in Table I, where the results show a slightly lower F1-score (45.20), a much worse EAE value (mean 0.477, while with less time we obtain a perfect value of 0), a worse MNEAP value (average of 2.444) and above all a much worse RMSE value (average of 431.515).

Table I.- Metrics of CO for ten appliances with extended training.

	F1-score	EAE	MNEAP	RMSE
Electric furnace	0.514	0.385	0.916	963.602
Microwave	0.471	0.498	1.671	566.094
TV	0.753	0.026	0.579	24.426
Incandescent lamp	0.487	0.520	3.157	406.878
Vacuum cleaner	0.229	0.514	2.403	322.363
Electric heater	0.152	0.889	2.570	872.400
Electric shoer heater	0.529	0.548	1.092	614.549
Fan	0.662	0.049	0.686	22.566
Fridge	0.406	0.431	5.639	273.104
Freezer	0.317	0.412	5.724	249.175

Excluding odd harmonics while retaining even harmonics has been tested, and the results suggest that this approach generally does not lead to improvements in overall performance across various applications (mean F1-score of 44.01, EAE of 0, lower MNEAP of 2.2877, and RMSE of 401.3312).

While excluding just odd harmonics offers mixed results, removing all harmonics hurts performance across metrics (Table II). Despite similar F1-score (0.472), EAE (worse), MNEAP (much worse), and RMSE (slightly worse) suggest it's generally less effective.

Table II.- Summary metrics without harmonics

	F1-score	EAE	MNEAP	RMSE
Electric furnace	0.576	0.767	0.849	880.820
Microwave	0.518	0.399	1.645	570.959
TV	0.767	0.027	0.595	24.598
Incandescent lamp	0.479	0.082	1.095	54.838
Vacuum cleaner	0.261	0.594	2.617	333.217
Electric heater	0.189	0.895	2.112	806.931
Electric shoer heater	0.431	0.681	1.196	657.188
Fan	0.651	0.054	0.713	23.501
Fridge	0.434	0.422	4.893	269.037
Freezer	0.423	0.351	4.400	201.280

B. Influence of the number of samples

The CO and FHMM algorithms were applied to the datasets using NILMTK for disaggregation to obtain the common metrics. Since significantly different results were observed, it is hypothesised that the number of samples per application could influence the metrics obtained. The average number of samples for each application in each dataset was calculated and related to the F1-score, EAE, and RMSE metrics to investigate this possibility.

As shown in Fig.5, there is no clear trend in the relationship between F1-score and the number of samples, with F1-score values in a relatively narrow range, between 0.4658 and 0.5982.

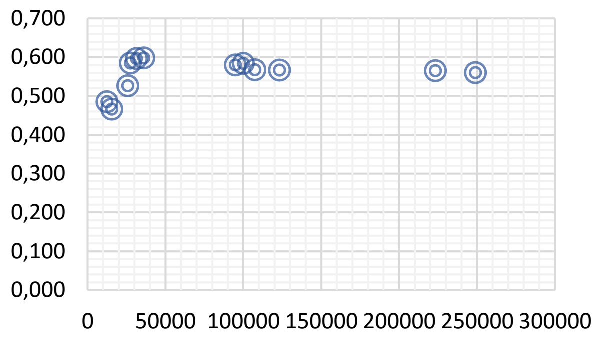


Fig. 5. Relationship between the number of samples and F1-score.

In most cases, the EAE metric shows a general downward trend as the number of samples increases, as shown in Fig. 6. This indicates that the accuracy of energy disaggregation improves with more measurement points.

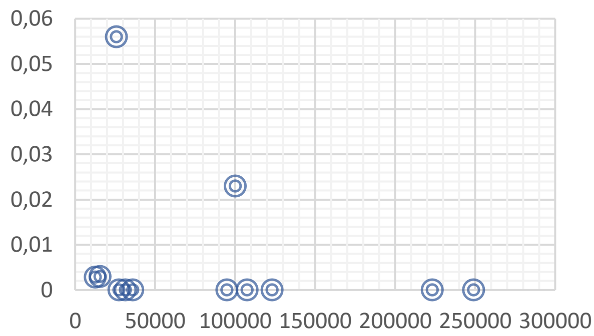


Fig. 6. Relationship between the number of samples and the EAE metric.

Regarding the RMSE metric, with the increase in samples, there is no clear trend in the relationship between RMSE and the number of samples (Fig. 7), oscillating the values from a minimum of 2.9964 to a maximum of 15.9096.

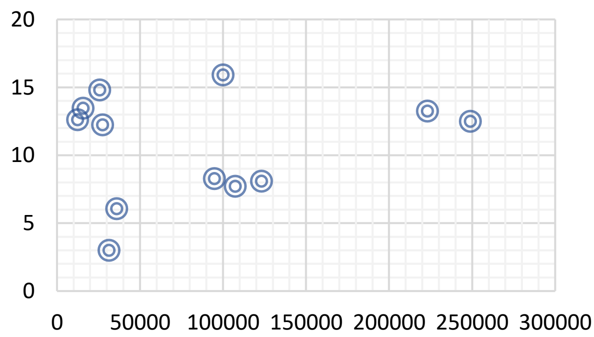


Fig. 7. Relationship between the number of samples and the RMSE metric.

.

Table III- Summary of metrics in six datasets using the mean, the best or the worst value.

	F1 mean	F1 best	F1 worst	EAE mean	EAE best	EAE worst	MNEAP mean	MNEAP best	MNEAP worst	RMSE mean	RMSE best	RMSE worst
DSUALM10H	0.454	0.812	0.253	0	0	0	1.981	0.553	4.930	372.847	23.354	920.380
DSUAMH	0.829	1	0.671	0	0	0	0.479	0.021	0.895	33.837	22.195	46.070
UALM2	0.632	0.789	0.420	0.006	0.001	0.012	0.725	0.348	1.138	14.782	7.339	17.417
DSUALM	0.851	0.996	0.679	0	0	0	0.666	0.660	1.815	33.490	20.600	62.253
DEPS	0.845	0.915	0.463	1.001	0.608	2.556	0.558	0.155	0.937	105.706	108.786	193.970

C. Influence of application type on metrics

Considering the arithmetic mean of the measurements, the total will be significantly affected depending on the type of applications considered in the dataset. Therefore, in addition to considering the arithmetic mean, the best and worst values across different datasets are considered by using the public data available from these datasets (Table III). This approach aims to provide a more comprehensive analysis by evaluating performance under different scenarios.

Examining F1-scores, a metric ideally valued at unity, analysing the best values yields superior metrics for DSUALM and DSUALMH datasets, in a similar way for the average value, which obtains worse values for DSUALM10H or iAWE datasets. On the other hand, selecting the worst values identifies iAWE, DSUALM10H, OMPM and DEPS as the least-performing datasets. For EAE values, these are excellent for all datasets (values close to zero) except for DEPS, which consistently performs poorly.

In the MNEAP metric (Fig. 8), the highest values (MNEAP-WORST) above the average for all datasets stand out, the most striking being DSUALM10H (which takes the value of almost five points due to the fridge), but in any case, they are all very acceptable values below that threshold.

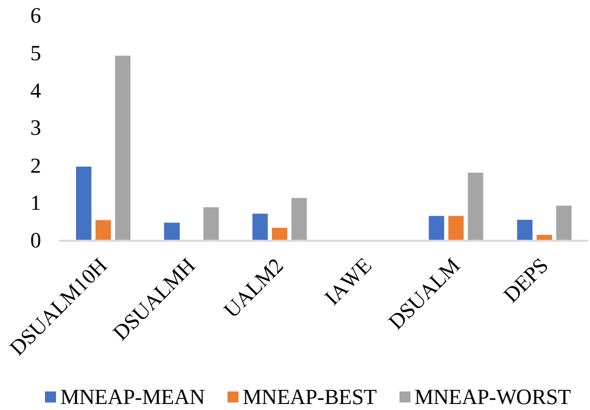


Fig. 8. MNEAP metric comparison using the mean, best or worst value.

Finally, Fig. 9 shows the same analysis for the RMSE metric, where again, the DSUALM10H dataset does not obtain too good values, undoubtedly impacted by the worst value, which in this case is for the oven. Similarly, both the worst case and the mean are highlighted; however, they are very good for the best value at the same level as the rest of the datasets, except for DEPS, which obtains a worse value.

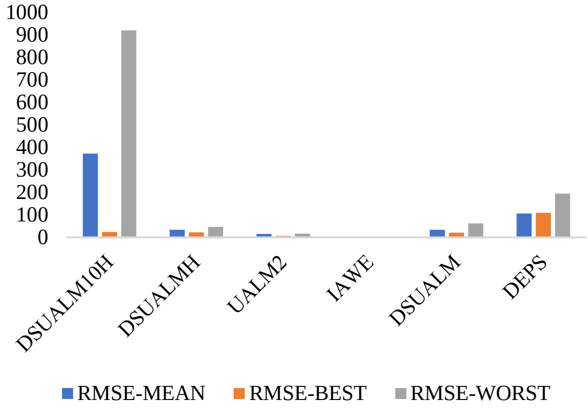


Fig. 9. RMSE metric comparison using the mean, best or worst value.

D. Only best values

Taking into account all the previous results and selecting the most optimistic value (i.e. the best value), Fig. 10 reveals a near uniformity in results, highlighting as better DSUAML and DSUALMH datasets (i.e. the dataset generated with oZm v1 and five applications), followed by DEPS and iAWE, and to end up with very acceptable values of 0.8 both DSUALM10H (taken with oZm v2) and UALM2 (generated with the OMPM meter).

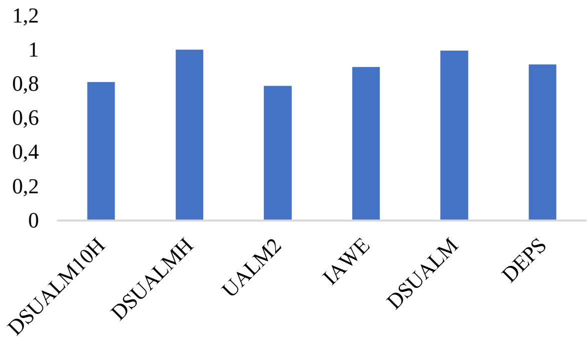


Fig. 10. F1-score results in the best case.

Regarding the EAE metric (which ideally should be as low as possible), Fig. 11 indicates that nearly all datasets have excellent values (since their value is null or almost zero), except for the DEPS dataset, which takes the worst value for this metric.

Regarding the MNEAP metric, the DSUALMH dataset (which includes harmonics) takes the best value followed by DEPS, UALM2 (generated with OMPM), DSUALM10H, and ends up as the worst, DSUALM (without harmonics). The notable improvement in results with harmonics, as DSUALMH demonstrates superior values, is highlighted in Fig. 12.

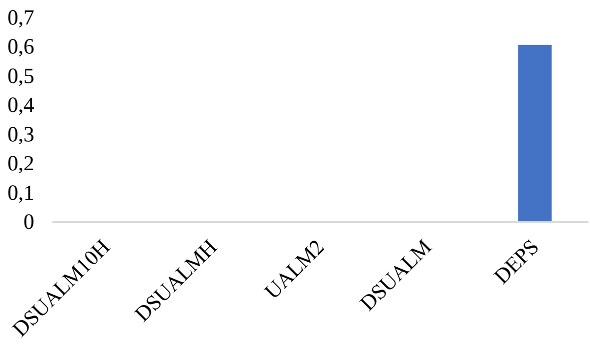


Fig. 11. EAE metric results in the best case.

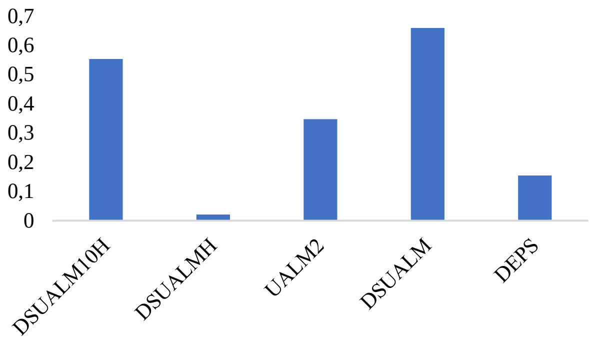


Fig. 12. MNEAP metric results in the best case.

Finally, the RMSE metric analysis is depicted in Fig. 13. The best value is obtained for the UALM2 dataset, followed by DSUALM (dataset of five applications without harmonics) and DSUALM10H (dataset of ten applications with harmonics). The results obtained for DSUALMH (dataset of six devices with harmonics) are already very far apart, and finally, the worst result is the one obtained with DEPS.

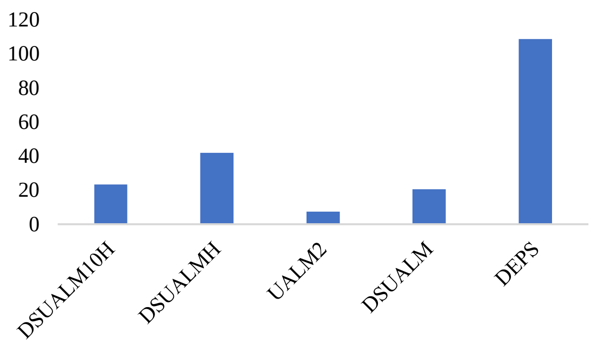


Fig. 13. RMSE metric results in the best case.

Conclusions

The analysis across different datasets (created with oZm, OMPM and others) highlights how the disaggregation algorithm, the sampling times, the filling method, the considered interval of the measurements, the number of samples, the composition of the meters, the sensors used, etc., influence as well as the inclusion or not of harmonics to offer an excellent disaggregation or not with the generated model.

Additionally, it has also been demonstrated through the analysis of the best or worst metrics how it is very interesting to exhaustively study the type of appliance used to generate datasets because it has been confirmed how, depending on the algorithm used, it can seriously harm the result of specific metrics and thus distort the generated model.

Only the oZm v2 meter has been used as a reference using the CO and FHMM algorithms in this work. In future research, it would be desirable to contrast the results with other disaggregation algorithms and use different open meters (such as the OMPM[17]) to capture the electrical measurements.

Acknowledgement

This research is part of the Project CPP2021-008949 “Artificial intelligence for the estimation of agricultural production in greenhouses”, financed by MCIN/AEI/10.13039/501100011033 and by the European Union “NextGenerationEU”/PRTR.

REFERENCES

[1] G. W. Hart, ‘Nonintrusive Appliance Load Monitoring’, *Proceedings of the IEEE*, vol. 80, no. 12, 1992, doi: 10.1109/5.192069.

[2] B. Nipun, K. Jack, P. Oliver, D. Haimonti, and K. William, ‘NILMTK: An Open Source Toolkit for Non-intrusive Load Monitoring’, Apr. 2014, doi: 10.1145/2602044.2602051.

[3] R. G. Pablo Salvador, ‘Implementación del algoritmo de análisis de demanda eléctrica por usos finales en las curvas de UVECASE’, 2020. doi: DOI: 10.13140/RG.2.2.27860.22405.

[4] K. Jack, B. Nipun, P. Oliver, D. Haimonti, and K. William, ‘NILMTK v0.2: A non-intrusive load monitoring toolkit for large scale data sets’, in *BuildSys 2014 - Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*, Association for Computing Machinery, Nov. 2014, pp. 182–183. doi: 10.1145/2674061.2675024.

[5] W. Huijuan, Y. Wenrong, C. Tingyu, and Y. Qingxin, ‘An Optimal Load Disaggregation Method Based on Power Consumption Pattern for Low Sampling Data’, doi: 10.3390/su11010251.

[6] C. Kwok Tai, L. Miltiadis D., and V. Anna, ‘Energy sustainability in smart cities: Artificial intelligence, smart monitoring, and optimization of energy consumption’, *Energies (Basel)*, vol. 11, no. 11, Nov. 2018, doi: 10.3390/en11112869.

[7] G. Feixiang, L. Chang, J. Linru, L. Hao, L. J Y, and Y. Bo, ‘Load Disaggregation in Non-Intrusive Load Monitoring Based on Random Forest Optimized by Particle Swarm Optimization’.

[8] Q. Bing, L. Liya, and W. Xin, ‘Low-Rate Non-Intrusive Load Disaggregation with Graph Shift Quadratic Form Constraint’, doi: 10.3390/app8040554.

[9] Y. Quanbo, W. Huijuan, W. Botao, S. Yaodong, and W. Hejia, ‘A fusion load disaggregation method based on clustering algorithm and support vector regression optimization for low

sampling data', *Future Internet*, vol. 11, no. 2, 2019, doi: 10.3390/fi11020051.

[10] F. Albert, 'Algorithms for Energy Disaggregation', 2016. Accessed: Apr. 21, 2022. [Online]. Available: https://upcommons.upc.edu/bitstream/handle/2117/89937/108724.pdf

[11] J. Lei, L. Jiaming, L. Suhuai, W. Sam, and P. Glenn, 'Power Load Event Detection and Classification Based on Edge Symbol Analysis and Support Vector Machine', *Applied Computational Intelligence and Soft Computing*, vol. 2012, 2012, doi: 10.1155/2012/742461.

[12] B. G, B. C., C. D., F. L., and M. G., 'Preliminary Sensitivity Analysis of Combinatorial Optimization (CO) for NILM Applications: Effect of the Meter Accuracy; Preliminary Sensitivity Analysis of Combinatorial Optimization (CO) for NILM Applications: Effect of the Meter Accuracy', 2021, doi: 10.1109/RTSI50628.2021.9597362.

[13] L. Hui, *Non-intrusive Load Monitoring*, 2020th ed., vol. 1. Central South University Chansha,China: Science press beijing, 2020. Accessed: Apr. 21, 2022. [Online]. Available: https://link.springer.com/book/10.1007/978-981-15-1860-7?noAccess=true

[14] S. Arias, 'AriasSilva/DEPS_NILM_Dataset: DEPS: Dataset de la Escuela Politénica Superior'. Accessed: Jun. 03, 2022. [Online]. Available: https://github.com/AriasSilva/DEPS_NILM_Dataset

[15] C. Rodriguez-Navarro, A. Alcayde, V. Isanbaev, L. Castro-Santos, A. Filgueira-Vizoso, and F. G. Montoya, 'DSUALMH- A new high-resolution dataset for NILM', *Renewable Energy and Power Quality Journal*, vol. 21, no. 1, pp. 238–243, Jul. 2023, doi: 10.24084/repqj21.286.

[16] C. Rodríguez-Navarro, F. Portillo, F. Martínez, F. Manzano-Agugliaro, and A. Alcayde, 'Development and Application of an Open Power Meter Suitable for NILM', *Inventions*, vol. 9, no. 1, p. 2, Dec. 2023, doi: 10.3390/inventions9010002.

[17] V. Eduardo, A. Alfredo, M. Francisco G., B. Raul, and A.-C. Francisco M., 'OpenZmeter: An efficient low-cost energy smart meter and power quality analyzer', *Sustainability (Switzerland)*, vol. 10, no. 11, Nov. 2018, doi: 10.3390/su10114038.