EMBED: A Dataset for Energy Monitoring through Building Electricity Disaggregation

Farrokh Jazizadeh
Virginia Tech
Blacksburg, VA, 24061
USA
jazizade@vt.edu

Milad Afzalan
Virginia Tech
Blacksburg, VA, 24061
USA
afzalan@vt.edu

Burcin Becerik-Gerber
University of Southern
California
Los Angeles, CA, 90089
USA
becerik@usc.edu

Lucio Soibelman
University of Southern
California
Los Angeles, CA, 90089
USA
soibelman@usc.edu

ABSTRACT

Electricity disaggregation, the task of inferring appliances' energy consumption in a building from a few sensing points, has received attention in the energy community. In this paper, we introduce EMBED, a publically available dataset for Energy Monitoring through Building Electricity Disaggregation. EMBED is the most comprehensive fully labeled disaggregation dataset to date containing the aggregate power measurement as well as plug load data of different appliances for three residential units, each was collected for at least two weeks. Therefore, this dataset could be used for comparing the performance for both event-based and non-event-based methodologies. In this dataset, all the events were uniquely labeled by identifying separate labels for different transition states for each appliance. This feature enables comprehensive evaluation of classification or clustering algorithms for disaggregation. The paper discusses the dataset, hardware setup, data collection procedure, and unique characteristics compared to the similar works. In addition, it presents preliminary analyses on the dataset as a benchmark to encourage future research in the disaggregation domain.

CCS CONCEPTS

• **General and reference** → Evaluation; Performance • **Hardware** → Power and energy; Energy metering • **Computer systems organization** → Embedded and cyber-physical systems

KEYWORDS

Electricity disaggregation, public dataset, algorithm evaluation, event-based, non-event-based

ACM Reference format:

Jazizadeh, F., Afzalan, M., Becerik-Gerber, B., and Soibelman, L. 2018. A Dataset for Energy Monitoring through Building Electricity Disaggregation. In *Proceedings of eEnergy'18, Karlsruhe, Germany, June 12-15, 2018*, 4 pages.

DOI: <https://doi.org/10.1145/3208903.3208939>

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

1 INTRODUCTION

Electricity disaggregation or non-intrusive load monitoring (NILM) comprises a category of techniques for identifying the contribution of individual loads from aggregate electricity consumption metrics (either raw signals or power metrics). These techniques rely on a combination of signal processing, machine learning, and statistical inference methodologies to achieve a drill-down pattern of consumption for individual appliances. In the broad sense, disaggregation methods can be categorized as event based and non-event based [5]. Event-based methods track the variations of electricity consumption that are caused by changes in operational states of appliances. They commonly use a process of event detection, classification, and energy estimation algorithms to quantify individual load contribution. On the other hand, non-event-based methods mainly apply Hidden Markov Models (HMM) to decompose the aggregate time series into appliance-level components. These two categories of methods call for different requirements in terms of data resolution and algorithm training. In event-based approaches, it has been demonstrated that higher data resolution could result in better classification accuracies [5]. However, energy estimation and non-event-based methods have mainly relied on low resolution data (e.g., with 1Hz sampling frequency and lower) [9].

Given the critical role that artificial intelligence methodologies play in energy disaggregation, contextual performance evaluation for the algorithms is an important step in research and development efforts. Until 2011 that the first public dataset [1] was released, researchers tend to collect/generate aggregate data for evaluating their proposed algorithms. Although valuable, this trend was a barrier in evaluating the performance of methods in reference to a common benchmark. Therefore, in recent years, a number of datasets have been collected, processed, and published for public access (e.g., [1-4, 6-8]). These datasets provide a common context for benchmarking the algorithms so that researchers could incrementally improve methodologies without repeating similar research. Publically

eEnergy'18, June 12-15, 2018, Karlsruhe, Germany
© 2018 Association for Computing Machinery.
ACM ISBN 978-1-4503-5767-8/18/06...\$15.00
DOI: <https://doi.org/10.1145/3208903.3208939>

Table 1: Comparison of several disaggregation datasets

Dataset	Location	No. of units	Duration	Labeled	Aggregate Rate	Appliance Rate	REF
REDD	MA, USA	6	3-19 days	No	1 Hz & 15 kHz	1/3 Hz	[1]
BLUED	PA, USA	1	8 days	Yes ¹	12 kHz	N/A	[2]
Smart [*]	MA, USA	3	3 months	No	1 Hz	1 Hz	[3]
AMPds	BC, Canada	1	1 year	No	1/60 Hz	1/60 Hz	[4]
iAWE	Delhi, India	1	73 days	No ²	1 Hz	1 Hz	[6]
UK-DALE	London, UK	4	3- 17 months	No	16 kHz	1/6 Hz	[7]
Dataport by PecanStreet	TX, CA & CO, USA	~1400	Up to 7 years	No	1/60 & 1/900 Hz	1/60 & 1/900 Hz	[8]
EMBED	CA, USA	3	14-27 days	Yes	12 kHz	1 Hz	-

¹ Different transition states for each appliance are not separately labeled | ² Appliance label for only one day is included

available datasets have been usually designed for a specific category of the techniques. Out of the seven major public datasets (as described in Table 1), six have focused on non-event-based techniques as they do not require the provision of labels for individual events. Labels are required in event-based methods as they commonly use supervised learning such as classification algorithms.

In this paper, we introduce the EMBED public dataset for Energy Monitoring through Building Electricity Disaggregation. EMBED is a unique dataset in the sense that it enables the evaluation of algorithms for both categories of event-based and non-event-based methods. It contains data from three apartment units in Los Angeles, California, collected for at least two weeks (the data collection period varies between two to four weeks). The dataset contains both the fully labeled appliance event timestamps in addition to the ground truth (GT) plug loads for a variety of appliances, making it suitable for both categories of methods. The dataset is publicly available for download at <http://embed-dataset.org>¹.

Table 1 summarizes the characteristics of comparable disaggregation datasets. The EMBED's contributions compared to the previous datasets are as follows:

1) It contains fully labeled appliance events as well as plug load consumptions for appliances that could be connected to power sockets. This is the feature that makes this dataset a good fit for performance assessment of disaggregation algorithms across both categories of event-based and non-event-based algorithms for comparison.

2) EMBED is the most comprehensive dataset in the event-based domain. Among the event based datasets, BLUED [2] has been considered as the primary source. Compared to BLUED that collected data for one house during one week, EMBED covers three apartments, each containing between two to four weeks of data, which resulted in having 9 times more labeled appliance events compared to the BLUED.

3) Instead of assigning only one label to each appliance, in EMBED, different transition states for each appliance were separately labeled. This feature renders EMBED unique in

evaluating different classification and clustering algorithms since the dataset is fully labeled for different observable appliance state transition (i.e., appliance signature patterns).

The EMBED dataset characteristics are as detailed in Table 2:

Table 2: Characteristics of the EMBED dataset

Feature	Value
Current sampling rate	12 KHz
Voltage sampling rate	12 KHz
Power resolution	60 Hz
Plug load sampling rate	1-2 Hz
Power metrics	Real (P) Reactive (Q)
Calculated power harmonic content	Up to 9 th harmonics
Apt 1 Duration, # of occupants	14 days, 1 occupant
Apt 2 Duration, # of occupants	21 days, 1-2 occupants
Apt 3 Duration, # of occupants	27 days, 1-2 occupants

2 DATA COLLECTION AND PROCESSING

In the United States, most residential loads are connected to a two-phase circuit. Each phase is fed with AC 120V with 180° phase difference. Appliances are usually connected to one phase, except for two-phase appliances (e.g., some air conditioning units or dryers), which are fed from both phases. More details on the split phase system could be found in [10]. Power measurements could be carried out at different levels of granularity from individual appliances/plugs to building level. A building unit circuit breaker is a logical aggregation point to measure the aggregate consumption of multiple appliances. Given that in this dataset, we monitored electricity consumption in apartment units, the main electric panel at each unit was monitored as the point of aggregation.

In AC circuits, depending on the characteristics of the load, different load behavior patterns could be observed. The appliances' components could be purely resistive loads, such as heating elements, or partially reactive loads, such as refrigerator and washing machines, which in addition to resistive loads have inductive or capacitive components. The resistive loads consume all the energy that is provided, but reactive loads consume part of the provided energy and release some of the energy back to the

¹ Samples of code for basic analysis on the dataset are available at: <https://github.com/EMBEDdata/code>

source. Accordingly, the power in AC circuits is a complex quantity. Loads with reactive components cause a phase shift between the voltage and current waveforms, which results in reactive power. Moreover, modern appliances contain nonlinear loads, which are characterized by switching action which causes current interruptions (e.g., electronic appliances with components such as rectifiers and switched-mode power supply (SMPS)). Presence of the nonlinear loads results in current waveforms with higher harmonics, which are the integer multiples of the fundamental frequency (60 Hz).

2.1 Data Acquisition System

The data acquisition was carried out by using general purpose Analogue to Digital (A/D) cards from National Instruments (NI), with high sampling frequency (≥ 100 kHz) and the capability of simultaneous sampling. According to the Nyquist sampling theorem, to capture at least the first nine current harmonics, a sampling frequency of more than 1 KHz is required. Therefore, the selected A/D card enabled us to capture a wide range of harmonic contents.

Figure 1: The data collection system set-up

Figure 1 shows the data collection system set-up. Voltage and current transformers were used as the aggregate sensing technology. Since the voltage waveforms in the split phase system have 180° phase difference, the voltage measurement was performed on one phase only through a socket. Pico TA041 - 25MHz ± 700 v differential probe was used as the voltage sensor. For current waveforms, the measurement was performed on each phase using Fluke i200 AC current clamps. The analogue signals, captured by these sensors are digitized through NI DAQ and were stored/processed on a local PC computer.

Another important aspect in the data acquisition process was the collection of the ground truth (GT) data. The most accurate approach in collecting GT data is through sensing at the individual appliance (load) level in parallel to aggregate measurements. In this way, the isolated log of operational modes

for each appliance could be obtained. The isolated logs could then be used for individual state transitions' labeling and energy ground truth collection. The ground truth data were obtained through plug level metering for appliances that could plugged into a power socket. Enmetric Powerports [11], which are off-the-shelf plug level meters for high resolution data collection were used. These sensors provide 1Hz sampling rate for power measurement through their integrated Application Programming Interface (API). Collecting the ground truth for lighting was performed using WiFi-enabled LinkSprite DiamondBack™ microcontrollers, equipped with AMBI™ light intensity sensors.

2.2 Data Processing Approach

The collected data could be used in different forms for electricity disaggregation processes: raw current and voltage waveforms as well as processed RMS or power metrics. The EMBED dataset includes the raw current and voltage waveforms as well as processed power metrics. Due to the presence of non-linear loads and their corresponding harmonics, the spectral envelope coefficient computation approach [10], which is an approximation method, was used to convert the current $i(t)$ and voltage $v(t)$ waveforms into power metrics. This approach uses Short-time Fourier Transform (STFT) on current and voltage waveforms which results in $I(t)$ and $V(t)$. The STFT window is usually an integer multiple of fundamental frequency (i.e., 60Hz). To account for the phase of the voltage relative to the window of the STFT, the phase shift ($\theta(t)$) between the first harmonic component of the current waveform and the voltage signal is obtained after the application of the STFT. Therefore, the real and reactive power components are obtained as follows [12]:

$$P_k(t) = |I_k(t)| \cdot \sin(\theta(t)) \cdot |V_1(t)|$$

$$Q_k(t) = |I_k(t)| \cdot \cos(\theta(t)) \cdot |V_1(t)|$$

in which P_k and Q_k are k^{th} real and reactive power quantities, I_k is the k^{th} harmonic component of the transformed current waveform, and V_1 is the first harmonic component of the transformed voltage waveform. The current and voltage sampling rates were set to 12kHz and the processed power metrics have a 60Hz resolution.

3 DATA LABELING PROCESS

Events are the reflection of the changes in appliances' operational states on the electricity consumption time series. In the EMBED dataset, these events have been fully labeled. As noted, the labels in this dataset group the events based on appliance type and operational states of that appliance type. The labels are five digit numbers, in which the first three represent the type of the appliance and the last two digits represent the mode of operation. For example 11101, 11102, 11103, and 11104 represent a refrigerator compressor on and off as well as defrost on and off events, respectively. The labeling process was carried out manually by human observers. The events on the aggregate time series were compared to the events on the power/light intensity time series for individual loads. Figure 2 illustrates the user-interface of the labeling platform, which was used in labeling the

events in EMBED. In this context, events refer to sharp changes of the power values on the aggregate times series. The details of the labeling is unique in the EMBED dataset as it identifies all possible states of a specific appliance in an apartment unit. The similarity between events from different states of an appliance has been identified by the user by considering both real and reactive power signature of those states. Considering the high resolution of the power time series (i.e., 60Hz), the transient trajectories of events facilitate the visual identification of similarities.

Figure 2: Event labeling interface; (a) an isolated event on real power time-series @60Hz, (b) the isolated event on reactive power time-series @60Hz, (c) matching a longer segment of real power time-series with signal (power or light at 1Hz) time series from ground truth sensors

The appliances that were monitored in these three apartments have been listed in Table 3.

4 TYPICAL ALGORITHMS' PERFORMANCE

We have evaluated the performance of several algorithms on the EMBED data to demonstrate the applicability of several well-studied algorithms and provide performance benchmarks, rather than exploring the state-of-the-art methods. As noted, further investigations and development of different methodologies are the motivations of introducing the dataset.

Event detection performance: The Generalized Likelihood Ratio (GLR) event detector was used on the data. The GLR event detector uses the likelihood ratio between Gaussian distributions assigned to a number of points before and after each data point on the power time-series for event detection. We performed the analysis based on a modified GLR detector [13]. No specific consideration on parameters tuning was made, and the suggested parameters in [14] were employed. The performance results are as presented in Table 4 of the Appendix.

Classification performance: Classification algorithms are employed for classifying the appliance signatures, which are further employed for energy disaggregation. The performance of a set of well-known classifiers, namely, KNN, Naïve Bayes, and SVM were explored on two weeks of data for the Apartment 1 as presented in Table 5 of the Appendix. The feature vectors were constructed from the real and reactive power time series for 40

samples before and 60 samples after the event (corresponding to ~2 seconds).

Clustering performance: Clustering analyses are suitable to the task of disaggregation in different capacities [15]. We presented the result of a proposed automated spectral clustering for different Apartments in Table 6 of the appendix. The feature vectors characteristic was the same as the classification. More details on the clustering method could be found in [16].

Table 3: List of appliances in the EMBED dataset

Appliance	Apt 1	Apt 2	Apt 3
Refrigerator	X(P)**	X(P)	X(P)
Washing Machine	X(P)		
Central Air Conditioning (AC)			X
Window AC	X	X(P)	
Toaster	X(P)	X(P)	X(P)
Kettle	X(P)		X(P)
TV	X(P)	X(P)	X(P)
Iron	X(P)		
Hair Dryer	X(P)		X(P)
Hair Iron			X(P)
LCD Monitor	X(P)		
Laptop	X(P)	X(P)	X(P)
Light*	X	X	X
Grill		X	
Coffee Maker		X(P)	
XBOX		X(P)	
Bathroom Fan		X	X
Water Heater		X	
Surface Tablet			X(P)
Microwave		X(P)	X(P)
Electric Range		X	

* Individual ceiling and floor lights were separately labeled

** Indicates availability of plug load data

5 CONCLUSION

This paper presents EMBED, a publicly available dataset for electricity disaggregation. EMBED enables the evaluation of both event-based and non-event-based methodologies as it contains fully labeled events as well as plug load consumptions for a variety of appliances. In addition, EMBED contains labels for different transition states of each appliance, which facilitates the evaluation of different classification and clustering algorithms. The dataset was collected from three apartments at a course of 14-27 days at a high resolution (12kHz). The paper describes the characteristics of the dataset in terms of the hardware setup, data collection, and provides several preliminary performance benchmarking. As part of the future work, we plan to expand the dataset.

6 ACKNOWLEDGMENT

This material is based upon work supported by the National Science Foundation under Grant #0930868. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

1. Kolter, J.Z. and M.J. Johnson. *REDD: A public data set for energy disaggregation research*. in *Workshop on Data Mining Applications in Sustainability (SIGKDD)*, San Diego, CA. 2011.
2. Filip, A. *BLUED: A fully labeled public dataset for event-based non-intrusive load monitoring research*. in *2nd Workshop on Data Mining Applications in Sustainability (SustKDD)*. 2011.
3. Barker, S., A. Mishra, D. Irwin, E. Cecchet, P. Shenoy, and J. Albrecht, *Smart*: An open data set and tools for enabling research in sustainable homes*. *SustKDD*, August, 2012. **111**: p. 112.
4. Makonin, S., F. Popowich, L. Bartram, B. Gill, and I.V. Bajic. *AMPds: A public dataset for load disaggregation and eco-feedback research*. in *Electrical Power & Energy Conference (EPEC)*, 2013 IEEE. 2013. IEEE.
5. Zoha, A., A. Gluhak, M.A. Imran, and S. Rajasegarar, *Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey*. *Sensors*, 2012. **12**(12): p. 16838-16866.
6. Batra, N., M. Gulati, A. Singh, and M.B. Srivastava. *It's Different: Insights into home energy consumption in India*. in *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*. 2013. ACM.
7. Kelly, J. and W. Knottenbelt, *The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes*. *Scientific data*, 2015. **2**: p. 150007.
8. Dataport, S.P.S.I. 2017.
9. Kolter, J.Z. and T. Jaakkola. *Approximate inference in additive factorial hmms with application to energy disaggregation*. in *Artificial Intelligence and Statistics*. 2012.
10. Hart, G.W., *Nonintrusive appliance load monitoring*. *Proceedings of the IEEE*, 1992. **80**(12): p. 1870-1891.
11. enmetric. Available from: <https://www.enmetric.com/platform#Hardware>.
12. Gonzalez, M.E.B., *A framework for enabling energy-aware facilities through minimally-intrusive approaches*. 2010, Carnegie Mellon University.
13. Berges, M., E. Goldman, H.S. Matthews, L. Soibelman, and K. Anderson, *User-centered nonintrusive electricity load monitoring for residential buildings*. *Journal of computing in civil engineering*, 2011. **25**(6): p. 471-480.
14. Anderson, K.D., M.E. Bergés, A. Ocneanu, D. Benitez, and J.M. Moura. *Event detection for non intrusive load monitoring*. in *IECON 2012-38th Annual Conference on IEEE Industrial Electronics Society*. 2012. IEEE.
15. Jazizadeh, F., B. Becerik-Gerber, M. Berges, and L. Soibelman, *An unsupervised hierarchical clustering based heuristic algorithm for facilitated training of electricity consumption disaggregation systems*. *Advanced Engineering Informatics*, 2014. **28**(4): p. 311-326.
16. Afzalan, M. and F. Jazizadeh, *Automated spectral clustering for multi-scale data*. (In review), 2018.

APPENDIX

Table 4: GLR event detection performance

Apartment	Phase	# of events	True positive	False positive	False negative	Precision	Recall
1	A	1576	1021	466	555	0.69	0.65
	B	2758	1235	102	1523	0.92	0.45
2	A	4661	4585	2454	76	0.65	0.98
	B	4457	4415	1933	42	0.70	0.99
3	A	6864	6430	4153	434	0.61	0.94
	B	1204	1194	1789	10	0.40	0.99

Table 5: Performance evaluation of classification algorithms for apartment 1 (results averaged on 10-fold cross-validation)

Event type	Phase	Classifier	Precision	Recall	F-score
Turn-On	A	KNN	0.95	0.96	0.96
		Naïve Bayes	0.85	0.71	0.70
		SVM	0.50	0.68	0.56
Turn-Off		KNN	0.97	0.97	0.97
		Naïve Bayes	0.93	0.49	0.57
		SVM	0.60	0.77	0.67

Table 6: Performance evaluation of automated spectral clustering algorithm for different apartments (results averaged on 5-fold cross-validation)

Apartment	Event type	Phase	Accuracy	Precision	Recall	F-score
1	Turn-On	A	0.97	0.97	0.97	0.96
	Turn-Off		0.89	0.85	0.89	0.86
	Turn- On	B	0.97	0.96	0.97	0.96
	Turn-Off		0.96	0.95	0.96	0.95
2	Turn-On	A	0.88	0.82	0.88	0.84
	Turn-Off		0.84	0.77	0.84	0.80
	Turn- On	B	0.95	0.92	0.95	0.93
	Turn-Off		0.66	0.51	0.66	0.56
3	Turn-On	A	0.87	0.83	0.87	0.84
	Turn-Off		0.76	0.71	0.76	0.70
	Turn- On	B	0.98	0.97	0.98	0.98
	Turn-Off		0.92	0.92	0.92	0.91