EVALUATION OF NON-INTRUSIVE LOAD MONITORING ALGORITHMS FOR APPLIANCE-LEVEL ANOMALY DETECTION

Haroon Rashid[⋆†] *Vladimir Stankovic*[⋆] *Lina Stankovic*[⋆] *Pushpendra Singh*[†]

[⋆] Dept. Electronic & Electrical Engineering, University of Strathclyde [†] IIIT Delhi

ABSTRACT

Appliance fault in buildings resulting in abnormal energy consumption is known as an *anomaly*. Traditionally, anomaly detection is performed either at aggregate, i.e., meter-level, or at appliance level. Meter-level anomaly detection does not identify the anomaly-causing appliance, while appliance-level detection requires submetering each appliance in the building. Non-Intrusive Load Monitoring (NILM) has been proposed as an alternative to submetering to detect when appliances are running as well as estimate the appliance energy consumption. So far, applications have revolved around meaningful energy feedback. In this paper, we assess whether NILM can indeed be used for anomaly detection, as an alternative to submetering. We propose a supervised anomaly detection approach, **AEM**, and evaluate the effectiveness of NILM for anomaly detection. The proposed approach first learns an appliance's normal operation and then monitors its energy consumption for anomaly detection. We resort to real data, aggregate and submetered data from the two-year long REFIT dataset. We explain why anomaly detection performs worse with NILM data as compared to submetered data, highlighting the need for new, anomaly-aware NILM approaches.

Index Terms— NILM, energy disaggregation, anomaly detection, smart metering

1. INTRODUCTION

Whenever appliance energy consumption is statistically different from usual, expected consumption, we say that load *anomaly* has occurred. Some reasons for these anomalies include: the appliance is not switched off after usage, the appliance experiences genuine problems, such as malfunctioning due to age and/or wear and tear, or mis-configured settings. For example, Fig. 1 shows two anomalies in freezer usage. In each of these cases an appliance consumes more energy than necessary, which could be due to, e.g., not closing the freezer door appropriately or worn-out seal, and thus their timely detection is important to provide appropriate energy saving advice, such as appliance retrofit, replacement, or change of settings.

The anomaly detection problem is well investigated, and with the emergence of smart grids and widespread use of smart meters, load anomaly detection continues to remain in the research focus [1, 2, 3]. Smart meters, measuring aggregate household consumption, allow online billing, facilitate demand response measures, and home automation by logging energy consumption data at frequencies often in the order of seconds. Existing smart meter-based anomaly detection approaches only detect anomalies at aggregate, household-level and does not identify the anomaly causing appliance [1, 2, 3]. Identifying timely anomalous appliances can reduce energy wastage and appliance breakdown time [4]; however, so far, it has required the

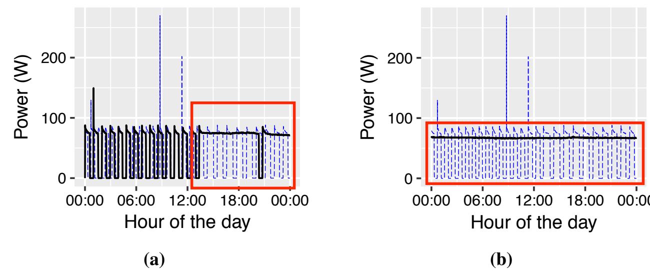

Fig. 1: Power consumption signatures of a freezer. Thin dashed blue lines show appliance's normal working pattern and thick solid black lines show consumption on an anomalous day. On an anomalous day, either appliance took longer ON cycles (Fig. a) or remained ON throughout the day (Fig. b) as highlighted by red rectangles.

use of submetered data at appliance level, which is a massive overhead due to a combination of installation, maintenance, communications, storage, and data validity checks. While an argument can be made for IoT-enabled appliances providing condition monitoring statistics, only a minority of the world population will have access to these and therefore extracting (appliance-level) information from the smart-meter is more viable as a sustainable solution.

Non-Intrusive Load Monitoring (NILM) [5], an algorithmic energy disaggregation approach, has been used to infer individual appliance's consumption using smart metered energy data. Over the years, the NILM research community has shown improvement in the appliance classification accuracy demonstrating that NILM is suitable for numerous applications, ranging from demand response and energy feedback to activity recognition [5, 6, 7, 8].

Building on the success of NILM over the past few years [7, 8], this paper explores the usability of state-of-the-art NILM methods for appliance's anomaly detection. Towards this, we propose an anomaly detection approach, namely, **A**ppliance **E**nergy **M**onitor (**AEM**) for detecting anomalies at an appliance level, which first builds a model of an appliance load by training on normal operation electrical measurements, and then monitors energy consumption of appliance for anomalies using a built-in training model. We test AEM's usability on UK's publicly available dataset, REFIT [9].

Contributions are: (i) an anomaly detection approach that works on NILM and submetered data. Testing on submetered data reports its baseline performance whereas testing on NILM data shows the usability of NILM for identifying anomalous appliances. (ii) a post-processing algorithm for improving anomaly detection capability of traditional NILM. (iii) release of a publicly annotated anomalies of the REFIT dataset [10]. Presently, no such detailed annotations are publicly available for any electrical load measurement datasets.

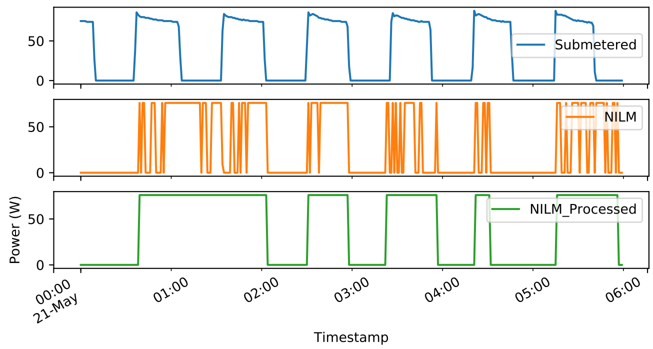

Fig. 2: Submetered (**top**), NILM (**middle**) and NILM post-processed (**bottom**) data.

2. RELATED WORK

Detecting anomalies by monitoring the energy consumption of an appliance over some time has been the subject of research. For example, Ganu et al. [11] first create appliance specific energy models of several high energy consuming appliances and then use built models to identify anomalies. Similarly, Pereira et al. [12] use clustering to identify anomalies by comparing the usage of an appliance along a period. However, such approaches are cumbersome as separate intrusive data collection kit is required for each appliance in a home. [1, 2, 3] propose different anomaly detection approaches using aggregate smart meter data. These approaches detect anomalies but are unable to identify the anomaly causing appliance.

NILM techniques provide the breakdown of smart meter data into individual appliances consumption. Previously, NILM has been used for identifying faults in isolated systems such as waste-disposal and HVAC [13, 14, 15]. Over the years, NILM's accuracy improved significantly, and according to [16], NILM outputs can be used for anomaly detection, but empirical evaluation has not been performed yet on building's aggregate meter data.

To the best of our knowledge, this paper is the first to evaluate the suitability of building's smart metered NILM data for anomaly detection. We presented the preliminary version of this work at ACM BuildSys [17, 18]. This work differs from the previous version in the following ways: (i) It is done on actual anomalies present in the dataset while [17, 18] was done on artificially injected anomalies, (ii) It uses a real-world energy dataset (REFIT [9]) for evaluation, while the previous version uses synthesized anomalies on USA energy Dataport [19], (iii) This work targets three appliances (Fridge, Freezer and Heater) as compared to two appliances (Air conditioner & Fridge) used in the previous version (iv) Current work proposes post-processing of NILM result for improved anomaly detection, while the previous version only directly applied the anomaly detection on NILM data.

3. METHODOLOGY

We focus on appliances that periodically pass through two cycles during their operation, i.e., appliances that have ON cycle followed by an OFF cycle, such as refrigerator, freezer, electric heater etc.

The analysis of cyclical appliances shows that anomalies are reflected in the appliance signature in two ways: (i) the appliance has a longer on duration, possibly due to appliance malfunctioning, fridge/freezer door left open, cracked door, or mis-configured settings [11]; (ii) the appliance goes through a significantly higher number of cycles as compared to normal operation [20]. This is due to appliance malfunctioning, due to age or a fault.

In the first case, the energy consumed in an anomalous cycle is usually significantly higher than in the normal cycles and in the second case, the number of cycles in a specific time duration will be significantly higher. With these two scenarios, we create a rule-based anomaly detection algorithm, namely **AEM**, for cyclical appliances.

AEM works in two phases - training and testing. In the training phase, it learns an appliance's power consumption signature during normal operation and computes average energy consumed by an appliance's ON cycle and the number of cycles by the appliance in a specific time. In the testing phase, it firstly computes the mentioned parameters of the power consumption of the test day and then compares these with the statistics computed during the training phase. A deviation observed in any value is flagged as an anomaly. Next, we formally describe each of these two phases.

Training Phase: First, power consumption readings of an appliance for D normal days are collected. For each day, D_i, we count the number of cycles taken by an appliance as c_i, and compute energy consumption of different cycles as vector $\mathbf{e}_i$, whose size is c_i. Next, using the computed statistics of D days, average number of cycles C_{train} and energy per cycle E_{train} are computed as:

$$C_{train} = mean(c_i), i \in \{1, \dots, D\} \tag{1}$$

$$\sigma^C_{train} = std(c_i), i \in \{1, \dots, D\}, \tag{2}$$

$$E_{train} = mean(\mathbf{e}_i), i \in \{1, \dots, D\}, \tag{3}$$

$$\sigma^E_{train} = std(\mathbf{e}_i), i \in \{1, \dots, D\}, \tag{4}$$

where σ^C_{train} and σ^E_{train} represent standard deviation of the number of cycles and energy consumption per cycle, respectively.

Testing Phase: In this phase, **AEM** takes power consumption signature of the appliance during the target test day and computes statistics C_{test} and E_{test} as above. Next, it uses the following set of rules to flag anomalies:

Rule # 1: If the average energy consumption of the test day cycles is significantly greater than the train day cycles, i.e.,

$$E_{test} > \alpha * (E_{train} + n * \sigma^E_{train}), \tag{5}$$

where n is the number of standard deviations and α is the number of times energy consumption on a test day deviates from the 'normal'.

Rule # 2: If the number of cycles taken by appliance on the test day is significantly greater than the train day cycles

$$C_{test} > C_{train} + n * \sigma^C_{train}. \tag{6}$$

We use five different well-known NILM algorithms to obtain the appliance-level NILM data from the aggregate smart meter data. These include Combinatorial Optimization (CO) [5], Factorial Hidden Markov Model (FHMM) [21], Latent Bayesian Modeling (LBM) [22], Super-state Hidden Markov Model (SSHMM) [23], and Graph-based Signal Processing (GSP) [8]. All of these are publicly available and are considered state of the art.

3.1. NILM signal post-processing (Algorithm 1)

The top two panels of Fig. 2 show submetered and NILM data obtained with SSHMM [23]. We can see that NILM often detects events but gets confused with the ON-OFF cycle frequency. **AEM** detects anomalies by monitoring the average energy consumption of each cycle. So, using NILM data as such will result in wrong anomaly results. To avoid this, we post-process NILM signal first and then **AEM** uses post-processed data for anomaly detection.

Case	Action 1	Action 2
• Appliance's consumption found significantly different from its historical normal consumption	Flagged as anomalous and marked as S (sure)	Noted time-duration of anomaly
• Appliance's consumption found significantly different from its historical normal consumption, but anomalous duration seems to be due to sensor malfunctioning	Flagged as anomalous and marked as NS (not-sure)	Noted time-duration of anomaly
• Appliance's ON cycle duration found significantly longer than its historical consumption and the predecessor OFF cycle also found longer.	Not marked as anomaly as it is normal in considered appliances	Nothing

Table 1: Rules for marking anomalies in REFIT dataset.

Input: Appliance's submetered power consumption $Y_{metered}$ of D normal days and NILM data of the appliance Y_{nilm}
Output: Post-processed NILM data, $Y_{processed}$

Compute duration d_i of each OFF cycle in $Y_{metered}$, where $i \in \{1, \cdots, H\}$; H is the number of cycles in D consumption days
Compute mean, $d_{metered}$, and standard deviation, $\sigma^{d}_{metered}$, over all d_i
for $j \leftarrow 1$ **to** *number of OFF cycles in* Y_{nilm} **do**
  Compute duration d^{j}_{nilm} of j^{th} OFF cycle
  if $d^{j}_{nilm} < (d_{meterd} - 2 * \sigma^{d}_{metered})$ **then**
    Find first and last readings of d^{j}_{nilm} as o_f and o_l
    $d^{intpol}_{nilm} \leftarrow$ Interpolate linearly all readings between o_f & o_l
    Update Y_{nilm} with d^{intpol}_{nilm}
  end
end
$Y_{processed} \leftarrow Y_{nilm}$
return $Y_{processed}$

Algorithm 1: Steps in post-processing NILM data.

Post-processing comprises two steps: (1) Calculate average duration $d_{metered}$ and standard deviation $\sigma^{d}_{metered}$, as per Algorithm 1, of all OFF cycles in submetered data $Y_{metered}$ of target appliance. (2) Take and Calculate the OFF duration d^{j}_{nilm} of each j^{th} cycle from NILM data Y_{nilm}. If the OFF duration is significantly less than $d_{metered}$, the "first" and "last" readings of the OFF duration are identified and then in-between readings are linearly interpolated as explained in the algorithm, resulting in removal of high frequency cycles introduced by NILM and in a form now suited for anomaly detection. This is shown in the bottom panel of Fig. 2.

4. EVALUATION

All experiments were conducted on the REFIT dataset [9], comprising household aggregate and appliance's submetered power consumption data of 20 UK homes for about two years. The default sampling rate of the data is eight seconds, but was uniformly down-sampled to one-minute. Out of 20 homes, Houses 1, 10, 16, 18 and 20 had the largest number of detected anomalies and were selected. We selected four months of data from each of these homes in such way that one month did not contain any anomaly and remaining three contain anomalies. The entire REFIT dataset was searched for anomalies, using rules defined in Table 1, for a whole month. The identified anomalies for all affected appliances were labelled as per the latter rules in a separate CSV file [10].

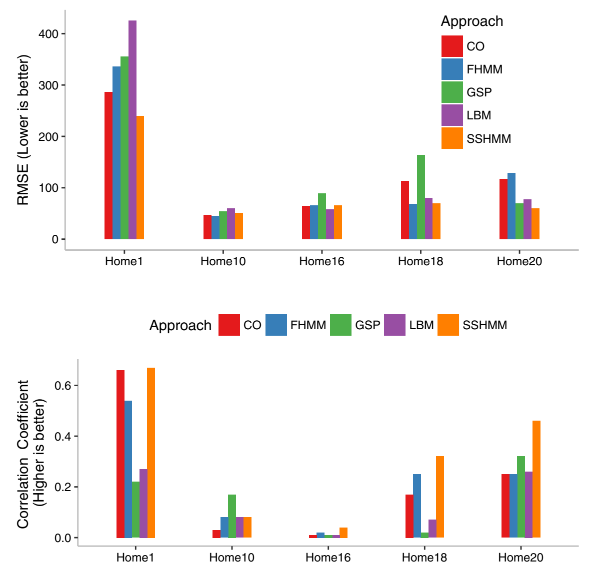

Fig. 3: Different metrics of NILM approaches on heater (Home1) and Freezer (Homes 10, 16, 18 and 20). RMSE (**top**), Pearson correlation coefficient (**bottom**). **[Best viewed in color]**

Disaggregation performance of NILM algorithms is presented using Root Mean Squared Error (RMSE) and Pearson Correlation Coefficient metrics. RMSE shows the difference between estimated and actual readings, and Pearson coefficient measures the correlation between appliance's submetered (s) and predicted (p) NILM readings.

$$\rho_{s,p} = \frac{cov(s,p)}{\sigma_s \sigma_p}, \tag{7}$$

where cov is covariance, σ_s and σ_p are the standard deviation of s and p. The value of $\rho_{s,p}$ varies in the range [-1, 1], where -1 means either s readings increase and p readings decrease or vice versa.

Experimental settings: For supervised NILM algorithms - CO, FHMM, LBM and SSHMM, one month of data was used for training and the remaining months for testing. Testing was done in a sliding-window manner with a window size of one day. Therefore, one month of training data was appropriate for one day of test data. Publicly available implementations of CO and FHMM from NILMTK toolkit [24], LBM [25], SSHMM [26], and GSP [27] were used to get disaggregation results.

To avoid evaluation bias, all NILM algorithms were run with default parameter settings as mentioned by their respective authors. For GSP, the appliance threshold was set to 40 Watts obtained empirically since the wattage was above 40 Watts for all appliances in the considered dataset.

Disaggregation performance: Fig. 3(top) shows RMSE of differ-

	Home 1						Home 10						Home 16						Home 18						Home 20					
	Submetered	CO	FHMM	LBM	SSHMM	GSP	Submetered	CO	FHMM	LBM	SSHMM	GSP	Submetered	CO	FHMM	LBM	SSHMM	GSP	Submetered	CO	FHMM	LBM	SSHMM	GSP	Submetered	CO	FHMM	LBM	SSHMM	GSP
Precis.	0.7	0.1	0.1	0.01	0.1	0.14	1	0	0.5	0	0.5	0.5	0.9	0	0.25	0.2	0.3	0.2	1	0	0.04	0.07	0.04	0.04	1	0.5	0.3	0.14	0.11	0.13
Recall	1.0	1.0	1.0	1.00	1.0	1	1	0	0.9	0	1	0.6	0.9	0	1	0.5	1	0.6	0.7	0	1	0.67	1	1	0.6	0.1	0.5	0.5	1	0.6
Fscore	0.8	0.2	0.2	0.15	0.2	0.25	1	0	0.7	0	0.7	0.5	0.9	0	0.4	0.3	0.4	0.3	0.8	0	0.08	0.13	0.08	0.08	0.7	0.17	0.3	0.22	0.2	0.21

Table 2: Precision, Recall and Fscore for **AEM** with post processed NILM data.

	Home 1						Home 10						Home 16						Home 18						Home 20					
	Submetered	CO	FHMM	LBM	SSHMM	GSP	Submetered	CO	FHMM	LBM	SSHMM	GSP	Submetered	CO	FHMM	LBM	SSHMM	GSP	Submetered	CO	FHMM	LBM	SSHMM	GSP	Submetered	CO	FHMM	LBM	SSHMM	GSP
Precision	0.7	0.1	0.1	0.08	0.1	0.14	1	0	0.6	0	0	0	0.9	0	0	0.23	0	0.24	1	0	0	0.06	0	0.04	1	0	0.33	0.13	0	0.33
Recall	1.0	1.0	1.0	1.00	1.0	1	1	0	0.27	0	0	0	0.9	0	0	0.43	0	0.48	0.7	0	0	0.33	0	1	0.6	0	0.3	0.4	0	0.1
Fscore	0.8	0.2	0.2	0.15	0.2	0.25	1	0	0.37	0	0	0	0.9	0	0	0.3	0	0.32	0.8	0	0	0.1	0	0.08	0.7	0	0.31	0.2	0	0.15

Table 3: Precision, Recall and Fscore for **AEM** without post processing of NILM data

ent NILM approaches on target (Heater in Home 1 and Freezer in remaining homes) appliances of five homes. Electric heater in Home1 has wattage > 1 kW, so RMSE values are higher than that of freezer of remaining homes. The appliance of Home 10 has lowest RMSE as compared to appliances of Homes 16, 18 and 20 because it has only one freezer as compared to two freezers found in remaining homes. Having appliances of distinct wattages decreases RMSE. Fig. 3(bottom) shows the Pearson correlation Coefficient of target appliances of five homes with different disaggregation approaches. Electric heater of Home 1 has highest correlation coefficient due to its distinct higher wattage as compared to appliances of other homes.

Overall, Fig. 3 show that NILM algorithms perform best in Home 10, followed by Home 20 and 16 in terms of lower RMSE and higher correlation coefficient. We expect to obtain better anomaly detection results for these houses.

Anomaly detection performance with NILM: For each target appliance (heater in home 1 and freezer in remaining homes), we use post processed NILM data (see Algorithm 1) of all approaches and submetered data to compute and compare the anomaly detection accuracies. Accuracy results obtained on submetered data are considered as baseline results. The NILM approach providing closest match to submetered data results is considered as the best NILM approach for appliance level anomaly detection.

Table 2 reports precision, recall and F-score of **AEM** with both submetered and five different post processed NILM data of target appliances. We infer the following from the table:

1. Recall is found to be better than precision for every home, meaning there are less false negatives as compared to false positives. A false positive results whenever NILM signature deviates significantly from the actual energy consumption and the deviation found matches to an anomaly signature, while false negative means missing the true anomaly.
2. The best results are obtained for Home 10, which is aligned with the fact that the performance of the NILM algorithms was the best for this house (see Fig. 3).
3. **AEM** results in an acceptable Fscore (≥ 0.8) on submetered data, and significantly lower F-score on using NILM output obtained with different approaches. Lower Precision results in the drop of Fscore and low Fscore on NILM data as compared to submetered data means **AEM** does not perform well on NILM output. This shows that NILM data of Freezer and Heater obtained with existing state-of-the-art NILM approaches cannot be used for anomaly detection.

We computed similar accuracy metrics on unprocessed NILM data too to show the improvement in F-score after post processing of NILM data. Table 3 shows precision, recall and F-score of different approaches on using unprocessed NILM data. Comparing Fscore of Table 2 and 3 we find post processing has improved results significantly, particularly for FHMM and SSHMM approaches. With post processing, **AEM** is able to flag all genuine anomalies.

5. CONCLUSION & FUTURE WORK

This paper reports detailed experiments to assess whether current state-of-the-art NILM algorithms outputs can be used for appliance-level anomaly detection. We conclude that appliance-level anomalies cannot be detected using NILM data directly because the appliance-level NILM signatures do not resemble submetered signatures since NILM algorithms are trained on normal appliance operation. While the obvious solution would be to train NILM algorithms with anomalous signatures, this is challenging because each appliance results in a different anomalous signature depending on the cause of the anomaly and knowing all anomalies signatures a priori is not always realistic. We also analyze the output of the NILM algorithms to understand the cause of the poor anomaly detection performance, and while the anomalies are detected, there is some confusion in detecting the events accurately with current metrics because of the frequency of the ON-OFF cycles. We propose a post-processing algorithm of NILM outputs to improve anomaly detection accuracy and show improved accuracy. There are two directions for future work: (1) developing novel anomaly detection rules that are suitable for NILM-based detection; (2) designing anomaly-aware NILM algorithms, without relying on learning normal operation load signatures.

Acknowledgement

This work was supported in part by the U.K. Engineering and Physical Sciences Research Council under Grant EP/R512898/1 EPSRC Global Challenges Research Fund Institutional Sponsorship Award 2017 (GCRF)/R171051-102 ENACT Project and in part by the ITRA project, funded by DEITy, Government of India, under a grant Ref. No. ITRA/15(57)/Mobile/HumanSense/01. H. Rashid is a TCS Ph.D. Fellow; P. Singh is a Visvesvaraya Young Faculty Fellow and an affiliate of Centre for Design and New Media. Annotated anomaly data underpinning this publication are openly available from University of Strathclyde at: https://doi.org/10.15129/9729a2a0-11ce-4cce-b0d0-144c483fcb33.

6. REFERENCES

[1] G. Bellala, M. Marwah, M. Arlitt, G. Lyon, and C. E. Bash, "Towards an Understanding of Campus-scale Power Consumption," in *Proceedings of the Third ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*. ACM, 2011, pp. 73–78.

[2] P. Arjunan, H. D. Khadilkar, T. Ganu, Z. M. Charbiwala, A. Singh, and P. Singh, "Multi-User Energy Consumption Monitoring and Anomaly Detection with Partial Context Information," in *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*. ACM, 2015, pp. 35–44.

[3] Haroon Rashid and Pushpendra Singh, "Monitor: An abnormality detection approach in buildings energy consumption," in *2018 IEEE 4th International Conference on Collaboration and Internet Computing (CIC)*. IEEE, 2018, pp. 16–25.

[4] K Carrie Armel, Abhay Gupta, Gireesh Shrimali, and Adrian Albert, "Is disaggregation the holy grail of energy efficiency? the case of electricity," *Energy Policy*, vol. 52, pp. 213–234, 2013.

[5] George William Hart, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, 1992.

[6] Jack Kelly and William Knottenbelt, "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*. ACM, 2015, pp. 55–64.

[7] Ahmed Zoha, Alexander Gluhak, Muhammad Ali Imran, and Sutharshan Rajasegarar, "Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey," *Sensors*, vol. 12, no. 12, pp. 16838–16866, 2012.

[8] Bochao Zhao, Lina Stankovic, and Vladimir Stankovic, "On a training-less solution for non-intrusive appliance load monitoring using graph signal processing," *IEEE ACCESS*, vol. 4, pp. 1784–1799, 4 2016.

[9] David Murray, Lina Stankovic, and Vladimir Stankovic, "An electrical load measurements dataset of united kingdom households from a two-year longitudinal study," *Scientific data*, vol. 4, pp. 160122, 2017.

[10] "Annotated load anomalies from the REFIT dataset," `https://doi.org/10.15129/9729a2a0-11ce-4cce-b0d0-144c483fcb33`, 2018.

[11] Tanuja Ganu, Dwi Rahayu, Deva P Seetharam, Rajesh Kunnath, Ashok Pon Kumar, Vijay Arya, Saiful A Husain, and Shivkumar Kalyanaraman, "Socketwatch: an autonomous appliance monitoring system," in *Pervasive Computing and Communications (PerCom), 2014 IEEE International Conference on*. IEEE, 2014, pp. 38–43.

[12] Welma Pereira, Alois Ferscha, and Klemens Weigl, "Unsupervised detection of unusual behaviors from smart home energy data," in *International Conference on Artificial Intelligence and Soft Computing*. Springer, 2016, pp. 523–534.

[13] Michael R Brambley, "A novel, low-cost, reduced-sensor approach for providing smart remote monitoring and diagnostics for packaged air conditioners and heat pumps," Tech. Rep., Pacific Northwest National Lab.(PNNL), Richland, WA (United States), 2009.

[14] Peter R Armstrong, Christopher R Laughman, Steven B Leeb, and Leslie K Norford, "Detection of rooftop cooling unit faults based on electrical measurements," *HVAC&R Research*, vol. 12, no. 1, pp. 151–175, 2006.

[15] Robert Williams Cox, *Minimally intrusive strategies for fault detection and energy monitoring*, Ph.D. thesis, Massachusetts Institute of Technology, 2006.

[16] Huan Yang, "Residential energy data analysis using green button data," Tech. Rep., Leigh University, Bethlehem, PA(United States), 2014.

[17] Haroon Rashid and Pushpendra Singh, "Energy disaggregation for identifying anomalous appliance," in *Proceedings of the 4th ACM International Conference on Systems for Energy-Efficient Built Environments*, New York, NY, USA, 2017, BuildSys '17, pp. 41:1–41:2, ACM.

[18] Haroon Rashid, Pushpendra Singh, Vladimir Stankovic, and Lina Stankovic, "Can non-intrusive load monitoring be used for identifying an appliances anomalous behaviour?," *Applied Energy*, vol. 238, pp. 796–805, 2019.

[19] "Pecan street inc. dataport 2018," `https://dataport.cloud`, 2018.

[20] Berkeleyheating, "http://berkeleyheating.com/blog/why-does-my-air-conditioner-compressor-turn-on-and-off," 05 2017.

[21] J Zico Kolter and Tommi S Jaakkola, "Approximate inference in additive factorial hmms with application to energy disaggregation.," in *AISTATS*, 2012, vol. 22, pp. 1472–1482.

[22] Mingjun Zhong, Nigel Goddard, and Charles Sutton, "Latent bayesian melding for integrating individual and population models," in *Advances in Neural Information Processing Systems*, 2015, pp. 3618–3626.

[23] Stephen Makonin, Fred Popowich, Ivan V Bajić, Bob Gill, and Lyn Bartram, "Exploiting hmm sparsity to perform online real-time nonintrusive load monitoring," *IEEE Transactions on Smart Grid*, vol. 7, no. 6, pp. 2575–2585, 2016.

[24] "NILM toolkit," `https://github.com/nilmtk/nilmtk`, 2018.

[25] "LBM python implementation," `https://github.com/MingjunZhong/LatentBayesianMelding`, 2018.

[26] "SSHMM python implementation," `https://github.com/smakonin/SparseNILM`, 2018.

[27] "GSP python implmentation," `https://github.com/loneharoon/GSP_energy_disaggregator`, 2018.