Can Non-intrusive Load Monitoring be used for Identifying an Appliance's Anomalous Behaviour?

Haroon Rashid*, *PushpendraSingh*

IIIT Delhi, Okhla Phase III, India

Vladimir Stankovic, Lina Stankovic

Dept. of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, UK

Abstract

Identification of faulty appliance behaviour in real time can signal energy wastage and the need for appliance servicing or replacement leading to energy savings. The problem of appliance fault or anomaly detection has been tackled vastly in relation to submetering, which is not scalable since it requires separate meters for each appliance. At the same time, for applications such as energy feedback, Non-intrusive load monitoring (NILM) has been recognised as a scalable and practical alternative to submetering. However, the usability of NILM for anomaly detection has not yet been investigated. Since the goal of NILM is to provide energy consumption estimate, it is unclear if the signal fidelity of appliance signatures generated by state-of-the-art NILM is sufficient to enable accurate appliance fault detection. In this paper, we attempt to determine whether appliance signatures detected by NILM can be used directly for anomaly detection. This is carried out by proposing an anomaly detection algorithm which performs well for submetering data and evaluate its ability to identify the same faulty behaviour of appliances but with NILM-generated appliance power traces. Our results on a dataset of six residential homes using four state-of-the-art NILM algorithms show that, on average, NILM traces are not as robust to identification of faulty behaviour as compared to using submetered data. We discuss in detail observations pertaining to the reconstructed appliance signatures following NILM and their fidelity with respect to noise-free

submetered data.

---

1. Introduction

In buildings, electrical appliance's faulty behaviour can happen either due to a fault in any appliance part or user negligence, e.g., refrigerant loss in a refrigerator or keeping the refrigerator door open. An instance of faulty behaviour can result in higher energy consumption than its normal behaviour and/or can lead to permanent damage of the appliance. Mostly, such faulty instances are intermittent; identifying them promptly improves appliance maintenance and lifespan, and results in energy savings. In this paper, hereafter, we call an appliance showing faulty behaviour as "*anomalous appliance*" and the anomalous instance as an "*anomaly*".

Identifying faulty behaviour of appliances in buildings has traditionally used submetered data, i.e., measuring energy consumption at appliance level individually, as in [1, 2, 3]. However, the number of submeters or individual appliance monitors increases with the number of appliances or loads, and therefore anomaly detection based on submetering is not a scalable solution, especially in modern households with over 40 electric appliances[1].

On the other hand, Non-intrusive load monitoring (NILM) estimates the individual consumption of an appliance within a building from the aggregate meter reading obtained from a smart meter, measuring total household electricity consumption at each sampling point; effectively eliminating the need for submetering. The effectiveness of NILM has been demonstrated in providing appliance [4, 5] and activity-based feedback [6] to consumers, utilities, and policy makers (see [7, 8, 9, 10, 11] for recent reviews).

NILM research has received an increased boost since 2010, primarily due to

---

[1]https://tinyurl.com/yc4frb7f

the roll-out of smart meters worldwide [12] and evidence that the appliance-level feedback to consumers can result in energy savings of up to 15% [13]. Many algorithms have been proposed to improve the disaggregation performance of NILM [14]. High disaggregation accuracy has been reported in the literature (in some cases, around 90% [15, 16, 17]), and presently, more than 30 companies are offering NILM-based solutions [18], e.g., EnerTalk (`https://www.enertalk.com/product`) from *Encorded* and SPEED (`https://bit.ly/2NLP1Yu`) from *Enetics* provide appliance-level consumption details to households from a bespoke smart meter fitted at the mains. There are other somewhat meter-agnostic offerings too that work on smart meter data from national roll-outs. However, in NILM literature and industrial offerings, the algorithms are not always tested at scale on real, noisy datasets typical of smart meter actual measurements from buildings and households. Furthermore, many NILM solutions are limited to disaggregating few appliances accurately, use multiple features (e.g., active and reactive power, voltage, current) and sampled measurements at $\gg$ 1Hz that are generally not available from national smart meter deployments, and offer either good classification accuracy (i.e., which appliance was running) or good consumption estimation (i.e., how much the detected appliance consumed in watts) accuracy. Current EU and national law and smart meter deployments do not make data available remotely (e.g., utility) at rates higher than 15-60 minutes. Furthermore, the feature available is mostly restricted to active power. However, as per the UK Smart Meter specifications [19] and other home energy management providers on the market (as discussed above) with bespoke higher resolution smart meters, the data available to the customer or data owner within the Home Area Network (HAN) is at higher granularity, e.g, 1-60 seconds, and therefore NILM can provide useful energy feedback directly to the customer.

While the ability of NILM in removing the need of submetered data for itemized billing is well recognized, so far NILM has not been tested for detection of appliance's faulty behaviour in buildings. To ensure accurate appliance anomaly detection, it is not sufficient to produce an accurate energy consumption estimate, but also to reconstruct with high fidelity the appliance load signature.

3

In this paper, we assess the accuracy of reconstructed appliance load signatures using state-of-the-art NILM methods and therefore the possible impact of NILM on anomaly detection, that depends on these load signatures being replicated accurately. That is, we evaluate whether NILM-generated power traces can be used directly in identifying anomalous appliances.

To identify faulty appliances from a single smart meter, first we use four publicly available, well-established and popular NILM techniques of [20, 21, 22, 16] to obtain disaggregated appliance power traces, and then attempt anomaly detection on these appliance power traces. Given the exploratory nature of the work, and to gain deep insights, we focus our study on the anomaly detection of two major energy consuming appliances in residential buildings, i.e., Air Conditioner (AC) and refrigerator. Typically, an AC runs for limited hours of a day, but often consumes significantly high amount of energy. On the other hand, a refrigerator remains operational 24x7, which causes it to consume energy (usually around 7% of the total energy consumption [23]) continuously.

Anomaly detection is performed using a new rule-based proposed algorithm, which we term **UNUM**[2] that first learns the appliance's ON-OFF cycle frequency and duration during normal operation and then monitors the appliance's consumption and flags an anomaly whenever a deviation is found.

Our study consists of two steps: (i) Perform energy disaggregation using existing techniques to get NILM data (i.e., appliance-level traces); (ii) Apply the proposed **UNUM** on both NILM data and submetered appliance data, where testing on submetered data provides the baseline performance of **UNUM**.

We use energy consumption data of six homes from three different publicly available datasets (REDDs [24], iAWE [25], Dataport [26]) to perform experiments. These datasets provide both aggregate smart meter measurements at 1 minute (Dataport) and 1 second (REDD, iAWE) sampling rates, and submetered data at the same rates (which is used purely for baseline performance evaluation).

---

[2]**UNUM** means "one" in Latin. It uses one appliance-level power trace.

Contributions of this paper are summarised as:

1. A rule-based **UNUM** algorithm is proposed for detecting anomalies, which uses appliance-level power traces of an AC or a refrigerator.

2. An in-depth methodological evaluation of the viability of NILM power traces is provided through careful insertion of well-established AC and refrigerator anomalies and through multiple metrics of assessment, to determine the correlation between NILM accuracy and resulting anomaly detection based on NILM power traces. The generated annotated appliance anomaly dataset is made publicly available.

3. Anomaly detection is performed directly on NILM-generated power traces obtained from the smart meter aggregate measurements instead of circuit-level measurements or appliance submetering.

4. Robust, methodological evidence is provided via four NILM algorithms and three datasets for experiments. Using publicly available NILM techniques and datasets allows reproducibility of presented results.

5. We discuss further steps needed to facilitate effective anomaly detection using NILM-outputs, i.e., appliance-level power traces obtained from NILM.

The remaining paper is organized as follows: Section 2 discuss the related work in the anomaly detection domain. Section 3 discusses the proposed anomaly detection algorithms. Section 4 explains the dataset, baseline algorithms and the evaluation metrics used. Section 5 mentions the results obtained. Section 6 discusses results obtained and Section 7 concludes the paper.

2. Related work

Related work can be broadly divided into two groups: work on anomaly detection and on NILM.

Anomaly detection: Anomaly detection in energy domain has become a popular research topic with the introduction of smart meters (aggregate load measurements), circuit-level and plug monitors (latter two providing submetering data), which enable logging and analysis of power consumption data. Therefore anomaly detection approaches target either aggregate smart-meter or at submetered load level energy measurements. First, we will discuss anomaly detection methods at the smart-meter level and then at the appliance level.

Seem [27] first proposed an anomaly detection approach, from smart meter data, that grouped different days consumption data into clusters according to various criteria (e.g., weekday/weekend) and then used statistical measures, such as mean and standard deviation, for each cluster separately to identify anomalies. An alternative, unsupervised learning approach is proposed [28] to identify anomalous days by first creating a lower-dimensional representation of high-dimensional energy data and then using a density-based algorithm to find anomalies. An enhanced unsupervised algorithm, which in addition to smart meter data uses context information (e.g., homes in the same locality should get affected similarly by weather), is used to improved anomaly detection [29]. The above algorithms detect household's total energy consumption anomalies, but cannot pinpoint the appliance causing the anomaly. More recently, a rule-based approach for identifying anomalous appliances using smart meter data only is proposed [30], but the approach falls short whenever several appliances with similar power rating are present in a home. Therefore, identifying an anomalous appliance from smart meter data only, is still an open challenge.

There have been recent attempts to use appliance-level consumption to identify appliance anomalies. Submetered data is used to build models specific for AC, washing machine, and refrigerator which track appliance's consumption over time and flag anomalous usage instances [1]. A self-adaptive stream clustering algorithm [2] is proposed to detect anomalies in the previous appliances as well as in electronic loads (TV, Laptop, Tablet, Mobile phone) using submetered data from these appliances. However, multiple appliance-level monitors are needed, which impacts the scalability of these approaches [1, 2, 3].

6

NILM: The ability to obtain appliance-level load measurements from smart meter aggregate data, using purely computational software methods with improved accuracy (see [14, 31, 32, 11, 10] for recent surveys of methods) has also ignited broader applications beyond energy feedback such as device scheduling, recommendation engine, demand response capacity estimation, itemized bills [33, 6], appliance mining [4], consumer studies [5], etc.(see [7, 8, 9] for surveys of NILM applications), that either relied on submetered appliance-level power traces or appliance models, which do not represent actual usage patterns. Primarily, effectiveness of NILM is evaluated along three dimensions: (*i*) *reducing sensing hardware cost*, i.e., submetering, by minimizing sensing installation and maintenance costs and reducing infrastructural change. (*ii*) *improving disaggregation accuracy*: every year numerous disaggregation approaches are being proposed to improve the disaggregation accuracy. Broadly, this includes both state-based and event-based approaches. Event-based NILM approaches, e.g., supervised and unsupervised Graph Signal Processing-based NILM [10, 32], DT-based [34] estimate ON-OFF timings of appliances from the aggregate household signal whereas state-based approaches, such as those of [35, 36, 37, 38], estimate combination of different appliance states from the aggregate signal. Pre-processing methods, e.g. [11] to improve NILM algorithms also improve accuracy. State-based approaches model consumption of each appliance with a finite state machine (FSM) [21, 39, 24, 16, 40, 41, 42]. Ideally, aggregate consumption is the combination of the appliance FSMs, and state-based algorithm should identify contributing appliances with inferencing algorithms. Mostly, state-based methods are Hidden Markov Model (HMM)-based and often outperform existing event-based methods. (*iii*) *minimizing the need of training data*: proposed NILM approaches are classified as supervised, such as those in [16, 24, 22, 43, 10, 34], or unsupervised [40, 35, 44, 45, 32]. Supervised approaches require labeled training data to do the disaggregation, while, unsupervised methods do not require labeled training data. In general, supervised methods are more accurate than unsupervised ones, but labeled training data requirement impedes their scalability.

NILM for anomaly detection: NILM has been explored in [46, 47, 48] for identifying faults in stand-alone closed systems such as waste-disposal systems, air conditioner and refrigeration systems, using circuit-level measurements, but not for identifying anomalous appliances in an overall building using aggregate smart meter data.

A NILM-based approach is used to provide appliance-level feedback by identifying appliances that consume more energy than expected [49, 50]. Focusing on refrigerator and HVAC, appliance models, in terms of energy consumed per cycle are built for different operating states (e.g., defrost, baseline, etc). Then, the NILM output is compared to the models to test if it can provide appliance-specific advice (e.g., high defrost energy). Unfortunately, it is concluded that tested NILM methods do not provide sufficient level of accuracy for such energy feedback.

Compared to the previous approaches of [46, 47, 48, 49, 50], this paper's originality can be summarised as: (*i*) we perform anomaly detection on the NILM-generated power traces obtained directly from the smart meter aggregate measurements instead of circuit-level measurements for sub-systems as in [46, 47, 48]. As expected, the more appliances contributing to the true aggregate measurements, the "noisier" the measurements due to multiple unknown appliances, and appliances with similar power ratings, which is a more realistic scenario but also rendering NILM problem more challenging [32] with potentially less accurate appliance-level power traces. (*ii*) the proposed anomaly detection algorithm is based on clear rules to estimate both the type of anomaly, as well as when the anomaly occurred, (*iii*) we use extra state-of-the-art NILM methods ([22, 16]) which show improved disaggregation performance compared to the ones used in [49, 50] and help in providing a more robust evaluation of using NILM for anomaly detection within specific appliances, (*iv*) a much more in-depth methodological evaluation of the viability of NILM power traces is provided through careful insertion of known anomalies and through multiple metrics of assessment, to determine the correlation between NILM accuracy and resulting anomaly detection based on NILM power traces.
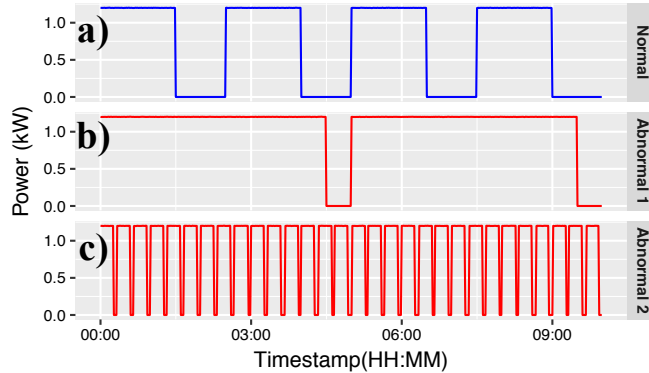
8

Figure 1: Power consumption signature of AC in three different modes for 10 hours: a) Normal mode, b) Abnormal mode when compressor takes long duty cycles, c) Abnormal mode when compressor duty cycles frequently.

Anomaly Type	Causes
Elongated duty cycle	Clogged air filter in AC, set point misconfiguration in AC, dry and cracked door gas kit in refrigerator [1]
Frequent cycling	Refrigerant leak, electrical problem(short circuit or damaged wire in compressor or thermostat), compressor damage [51]

Table 1: Causes for different anomaly types.

3. Methodology

In this paper, we focus our analysis on AC and refrigerator, which are common household appliances. They are both compressor-based and high energy consuming appliances, with the primary contributor to their energy consumption being their compressor. Any fault in the compressor itself or in any other part affecting the compressor gets reflected in the power consumption trace of the appliance. Figure 1(a) shows the normal functioning of such appliances where each cycle consists of ON and OFF states. In different types of faults, the duty-cycling ON-OFF nature of an appliance deviates significantly from its normal operation. Either it remains ON for longer durations, or it switches frequently between ON, OFF states as shown in Figures 1(b) and (c), respectively. The power trace shown in Figure 1 is for 10 hours. Table 1 reports various reasons which result in these type of abnormal behaviour.
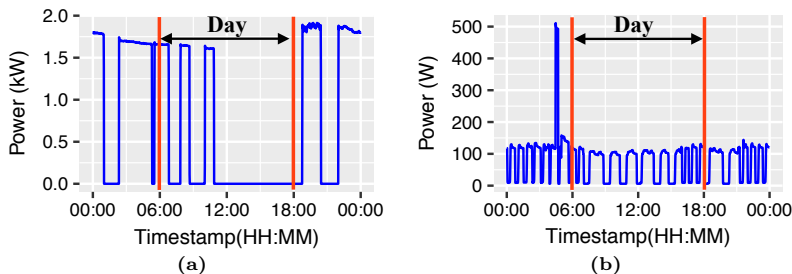
9

Figure 2: Different power consumption signatures during day and night: (a) AC, (b) Refrigerator.

The total energy consumed by the AC operating effectively, as shown in Figure 1(a), is 7.2 kWh, while the total energy consumed during anomalous operation, as shown in Figures 1 (b) and (c), for the same duration, are 10.8 kWh, and 9.6 kWh, respectively. This is a 3.6 kWh and 2.4 kWh increased energy use during anomalous conditions. However, the AC is not used continuously like a refrigerator, and the usage pattern varies significantly. We also show in Figure 2 that the consumption pattern also varies depending on time of day, both for AC and refrigerator. Therefore, using the relative total consumption per day or a period of a day to estimate anomalies would not be accurate. This is why we propose a rule-based approach, that takes into account the appliance characteristics, and focuses on the average energy consumption of the ON-cycle. For instance, in Figure 1(b), each ON cycle is of longer duration with energy consumption of 5.4 kWh, while as in Figure 1(c) each ON cycle is of shorter duration with energy consumption of 0.32 kWh. This is to be compared with the average energy consumption corresponding to the ON cycle for a normal operation, as in Figure 1(a), of 1.8 kWh. That is 3.6 kWh increase as above for anomaly (b) and 1.48 kWh decrease for anomaly (c). While the total energy consumption normalized by duration of use could provide an indicator of anomalies, it would not explain the type of anomaly, i.e., whether elongated or frequent cycling issue occurred. This is why a rule-based approached is proposed.

Our study consists of two steps: (i) First, apply existing NILM techniques on the smart meter data to get appliance power traces (ii) Next, use proposed
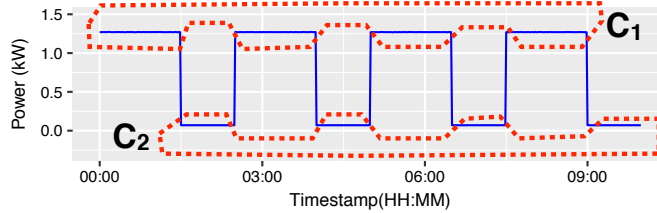
Figure 3: Clustering of ON and OFF states into two clusters C_1 and C_2.

anomaly detection algorithm, **UNUM**, on both AC and refrigerator disaggregated power consumption traces to identify anomalies. We build **UNUM** upon our preliminary work presented in a poster paper [52] by including more detailed rules. This has resulted in improvement in accuracy and the scope; the algorithm has also been evaluated thoroughly since its inception. Next, we explain **UNUM**, in detail.

UNUM consists of training and testing phases. In training phase, the statistical model of an appliance is built from T days of the appliance's historical power consumption trace, and in the testing phase, with the appliance's power consumption trace during a day as input, the algorithm outputs whether the appliance's consumption is anomalous or not. The following steps are used in the training phase:

1. Input appliance's power consumption trace of T days. The selected days are from the period when the appliance worked normally and also showed varied duty cycles according to different load conditions. For example, compared to higher settings, at lower set-point settings, the AC operates with considerably longer cycles.

2. Identify ON and OFF compressor states by using k-means clustering algorithm [53]. This results into two clusters corresponding to ON and OFF states as shown in Figure 3. Label all power consumption readings according to the cluster label. Cluster labels C_1 and C_2 are assigned to readings of ON and OFF states, respectively. Note that power consumption of OFF state is not 0 because an appliance still consumes energy when its compressor is off.

11

3. For each ON and OFF state, identify timestamps of first and last power consumption reading as $first$ and $last$. Calculate the duration D_s of each state as $D_s = last - first$.

4. Compute energy E_s of each state using all power readings of a state between $first$ and $last$.

5. For all ON states, compute mean over D_s and E_s, as $\overline{D}$ and $\overline{E}$, respectively. Also, compute standard deviation over E_s denoted by σ_e. Similarly, repeat all these statistics for OFF state.

Therefore, training model is a tuple $(\overline{D}, \overline{E}, \sigma_e)$ containing various parameter values corresponding to ON and OFF states separately. During analysis, we found that AC and refrigerator consume energy differently at different times of a day. Figure 2 shows that during the day time, appliance's duty cycle signature is distinct from that of the night. So, we build separate models: one for the day (0600 - 1800 hours) consumption and the other for night consumption.

During the testing phase on a test day, **UNUM** first takes power consumption data of an appliance and computes all mentioned parameters as defined above. Next, it uses the following set of rules to decide whether the test day consumption is anomalous, and if yes, then which type, i.e., elongated duty-cycle or frequent cycling:

Rule # 1: If an appliance switched between ON and OFF states frequently, then it is frequent anomaly type as shown in Figure 1(c). In this case, on average, energy consumed in any of the cycles is lower than a normal cycle, because ON cycles are interrupted frequently by OFF cycles and hence are of shorter durations leading to lower energy consumption than in normal operation case.

$$E^i_{testday} < \overline{E}^i - n * \sigma^i_e, \forall_{i \in \{ON, OFF\}} \tag{1}$$

where $\{n \in \mathbb{R} | n > 0\}$.

Rule # 2: If an appliance remains in ON state for an extended period, then it is an elongated duty-cycle anomaly as shown in Figure 1(b). In this case, on

average, energy consumed in any of the cycles is higher than in a normal cycle.

$$E_{testday}^{ON} > \overline{E}^{ON} + n * \sigma_e^{ON} \tag{2}$$

Rule # 3: If an appliance remains in ON state for an extended period, and the OFF duration is also longer as compared to normal, then it means that the appliance has been switched ON after a long time, as usual, hence it is not an anomaly.

$$\begin{aligned} E_{testday}^{ON} &> \overline{E}^{ON} + n * \sigma_e^{ON}, \quad \wedge \\ D_{testday}^{OFF} &> \overline{D}^{OFF} + n * \sigma(D) \end{aligned} \tag{3}$$

With these set of rules **UNUM** outputs anomaly status informatively to enable prompt decision-making: (i) which type of anomaly is present in the appliance, and (ii) which part of the day resulted in anomaly, as separate models are being used for day and night.

4. Evaluation

4.1. Dataset

We use energy consumption data of six homes from three different publicly available datasets (four from Dataport, one from iAWE and one from REDD [24, 25, 26]) for the evaluation. Other publicly available datasets (ECO [54], DRED [55], Smart [56]*, GREEND [57], REFIT [58], UK-DALE [59], AMPds [60], Dataport [26], REDD [24], PLAID [61], tracebase [62]) either do not have both AC and refrigerator or do not have data of considerable duration required for the experiments. Only one home in REDD has both AC and refrigerator data. All these six homes selected for our experiments have both aggregate and submetered data available. Sampling rate of Dataport is one sample per minute while REDD and iAWE have 60 readings per minute. Homes 1 - 4 are from Dataport, Homes 5 and 6 correspond to iAWE and REDD, respectively.

We select only four distinct homes from the 500 homes of Dataport. The selection methodology was: (i) We ran disaggregation techniques - FHMM [24] and CO [20] of NILMTK on all 500 homes and selected homes having minimal

13

disaggregation error for AC and refrigerator. We assume that small disaggregation error will result in higher fidelity appliance power traces and thus essential to evaluate the effectiveness of NILM in identifying an anomalous appliance. (ii) All homes are *distinct* in their energy consumption patterns thus enabling us to robustly evaluate our algorithms with a variety of consumption patterns. For Dataport, the period June 2014 - August 2014 was chosen for evaluation due to the following: (i) these three months are high energy consuming months due to extreme heat, and (ii) there is almost no seasonal variation during these months. This enables us to train our models on some data and then evaluate the built model on remaining data since there is *no seasonal drift*.

For iAWE, the period July 13, 2013 - August 04, 2013 was selected as suggested[3] by the authors of the dataset and for REDD, data from May 22, 2011 - June 13, 2011 was selected as this duration has minimal missing data.

We did a meticulous manual inspection of the energy consumption data and for these six homes, we found that the patterns were consistent and *"no anomalous instances were already present"*. We inserted anomalies, explained in detail in the next section, in these homes to check how effective NILM is in detecting the inserted anomalies. Anomalies were inserted following the methodology of [29]. The annotated dataset, with inserted anomalies, is publicly available at this[4] link.

---

[3]http://iawe.github.io
[4]`http://bit.ly/2wibl4L`

Table 2:

Home	Ratings (W)		Insertions(#)		Duration (H)		Duty cycle		Frequency	
	AC	Refrig.	AC	Refrig.	AC	Refrig.	AC	Refrig.	AC	Refrig.
1	2500	150	6	4	6 - 12	6 - 12	0.5 - 0.9	0.3 - 0.9	1/3, 3, 6	1/3, 1/4, 10, 12
2	1800	140	7	4	6 - 12	6 - 08	0.3 - 0.9	0.3 - 0.8	1/3, 1/4, 6, 10	1/2, 1/3, 10, 12
3	4000	090	6	4	6 - 12	6 - 10	0.5 - 0.9	0.3 - 0.9	1/3, 1/4, 1/7, 3, 6	1/3, 10, 12
4	1200	110	6	5	7 - 09	4 - 12	0.4 - 0.9	0.5 - 0.9	1/4, 1/6, 1/8, 16	1/7, 1/9, 1/12, 8
5	1800	100	5	4	5 - 12	5 - 10	0.2 - 0.9	0.8 - 0.9	2, 4, 140, 150	9, 12, 150, 200
6	1200	150	5	6	6 - 08	3 - 11	0.2 - 0.9	0.2 - 0.9	4/5, 4, 12	4/7, 1, 6, 90

Table 2: Statistics of inserted anomalies in AC and refrigerator of six different homes. Homes 1-4, 5, and 6 are Dataport, iAWE and REDD homes respectively.

Table 3:

Appl.	Home1				Home2				Home3				Home4				Home5				Home6			
	CO	FHMM	LBM	SSHMM	CO	FHMM	LBM	SSHMM	CO	FHMM	LBM	SSHMM	CO	FHMM	LBM	SSHMM	CO	FHMM	LBM	SSHMM	CO	FHMM	LBM	SSHMM
1	0.1	0.1	0.3	0.4	0.2	0.1	1.3	0.3	0.1	0.1	0.9	1.0	0.2	0.4	0.3	0.4	0.3	0.1	0.3	0.9	0.2	0.5	0.6	0.2
2	0.4	0.4	0.6	0.6	0.5	0.6	0.9	0.9	0.6	0.5	9.9	0.9	0.5	0.6	0.9	1.0	0.7	0.4	0.5	0.5	0.4	0.5	0.8	0.9
3	0.3	0.2	0.8	0.4	0.2	0.2	1.1	0.3	0.6	0.5	1.1	0.1	0.2	0.5	0.4	0.5	1.6	0.9	1.3	1.1	1.1	1.9	1.8	0.7
4	0.1	0.4	0.2	0.4	0.5	0.7	0.8	0.6	5.2	3.2	2.7	1.1	9.4	9.7	7.9	1.0	2.9	1.9	1.4	1.0	3.1	8.3	9.9	0.6
5	0.9	0.3	0.7	0.2	0.6	0.9	5.2	0.4	5.2	6.2	2.6	1.7	0.3	2.5	2.1	1.0	9.2	6.4	0.8	1.1	9.0	9.7	9.00	0.2

Table 3: Appliance Normalized Error for appliances using NILM techniques: CO, FHMM, LBM and SSHMM. Appliance mapping is shown in Table 4. Homes 1- 4 are from Dataport, Home5 and Home6 correspond to iAWE and REDD respectively.

Table 4:

Appliance	Home1	Home2	Home3	Home4	Home5	Home6
1	AC	AC	AC	AC	AC	AC
2	Refrigerator	Refrigerator	Refrigerator	Refrigerator	Refrigerator	Refrigerator
3	Furnace	Furnace	Furnace	Furnace	Laptop	Electric heater
4	Meter 1	Meter 1	Dishwashwer	Microwave	TV	Stove
5	Meter 2	Meter 2	Clotheswasher	Water heater	Water filter	Bathroom GFI

Table 4: Appliance mapping of different homes. Meter 1 and Meter 2 represent aggregate consumption of separate rooms comprising lighting and fan loads only. The Dataport Furnace refers to a blower. It is a part of Air Conditioner (AC) and is submetered separate to the AC. It is used to move the air over the condenser coils of AC.

15

4.2. Anomaly Insertion

To evaluate the performance of NILM in identifying anomalous appliances, two types of anomalies were inserted in the AC and refrigerator power traces of all six homes: (i) Elongated duty cycle and (ii) Frequent cycling. These anomalies appear in AC and refrigerator operation due to various reasons as reported in Table 1. Signatures of these anomalies are shown in Figure 1(b) and (c), respectively. Table 2 shows the statistics of all inserted anomalies. The following steps were used to insert anomalies:

1. Extract appliance's duty-cycle statistics manually for each home. As the separate models are built for daytime (06:00 - 18:00) and nighttime (00:00 - 06:00 & 18:00 - 24:00) hours, we note the minimum and maximum duty-cycle of an appliance for both day and night hours separately.

2. Keeping in view the normal operation of an appliance as recorded, multiple instances of the anomalies, as reported in Table 2, were inserted in AC and refrigerator power traces by replacing the measured signature (Figure 1(a)) with the anomalous signature, e.g., either Figure 1(b) or Figure 1(c). Four parameters define an anomaly signature: (i) Power rating, (ii) duration, (iii) duty cycle, and (iv) frequency. An appliance's power rating defines its power consumption in the ON state of duty cycle. For example, for Home#1, for the AC, six anomalies of varying durations (6 - 12 hours) were inserted on different days. Figure 1(b) shows an anomalous signature with a frequency of 1/2.5 meaning one ON and OFF cycle completes in 2.5 time units (= one hour) and with a duty cycle of 0.9 means 90% of cycle is in the ON state and remaining 10% is in the OFF state.

3. Later, the corresponding anomalies were also inserted in aggregate consumption to maintain synchronization between appliance's consumption and the aggregate consumption.

16

4.3. Disaggregation Techniques

For ease of reproducibility, we use publicly available NILM techniques: classical Combinatorial Optimization (CO) [20], Factorial-Hidden Markov Model (FHMM) [21], Latent Bayesian Melding (LBM) [22] and Super-state Hidden Markov Model (SSHMM) [16].

CO: The aggregate power consumption of a home accounts for the sum of individual appliances consumption at time instant t. Mathematically: $Y_t = \sum_{i=1}^{n} y_t^i + e_t$, where Y_t represents aggregate power consumption at time t, y represents appliance consumption, n the total number of appliances contributing to Y_t and e represents the residual. Hence, NILM can be formulated as a combinatorial optimization problem:

$$e_t = \arg\min_{e_t} |Y_t - \sum_{i=1}^{n} y_t^i|. \tag{4}$$

At every time instant t, CO ensures that the optimal combination of the set of ON appliances and their power consumption is found.

FHMM: FHMM is an extension of Hidden Markov Models. Each state consists of multiple independent chains corresponding to the number of appliances, and the output is represented as an addition function of all hidden states.

LBM: LBM is an extension of additive FHMM. It adds extra constraints in the form of the appliance's summary statistics including total energy consumption, duration of use, and usage frequency.

SSHMM: Unlike FHMM, each state is computed as the Cartesian product of all possible states of particular household appliances. It uses sparse Viterbi algorithm to reduce the computational overhead.

4.4. Baseline performance of **UNUM**

To evaluate the efficacy of **UNUM**, we run **UNUM** on raw submetered data available in all six homes. We refer the process of running **UNUM** on submetered and NILM disaggregated data as **UNUM_S** and **UNUM_D** respectively. Comparing their performance will indicate how effective NILM power traces are in identifying anomalous behaviour.

Home (#)	Training duration	Testing duration
1 - 4	June 01, 2014 - June 30, 2014	July 01, 2014 - August 31, 2014
5	July 13, 2013 - July 20, 2013	July 21, 2013 - August 04, 2013
6	May 24, 2011 - May 27, 2011	May 28, 2011 - June 13, 2011

Table 5: Division of data for **UNUM** and NILM approaches

4.5. Performance Metrics

Following metrics were used to report the NILM and anomaly detection accuracies:

(1) **Appliance Normalized Error** [63]: This metric captures the deviation of NILM readings from the submetered power readings for an appliance

$$\text{Appliance Normalized Error (ANE)} = \frac{\sum_t |y_t^a - \hat{y}_t^a|}{\sum_t y_t^a} \tag{5}$$

where y_t^a represents submetered power readings of appliance a at time t, $\hat{y}_t$ represents estimated power readings from NILM. The lower the ANE, the higher is the disaggregation accuracy for a.

(2) **F-score**: F-score is interpreted as weighted average of precision and recall.

$$\text{F-score} = 2 * \frac{precision * recall}{precision + recall} \tag{6}$$

Specific to anomaly detection methods, precision measures the percentage of correct anomalies to the total number of reported anomalies and recall measures the percentage of correct anomalies reported to the total number of anomalies present in a dataset. F-score varies in the range [0 - 1]. The higher the score, the better is the anomaly detection performance of algorithm.

4.6. Experimental Settings

Experiments were conducted under the following settings:

- **UNUM**: It is implemented in `Python` and the value for n was empirically set to 1.5. We present a sensitivity analysis of n in Section 5. Table 5 shows the data used for training and testing of **UNUM**.

- **Disaggregation techniques**: We use publicly available implementations, NILMTK[5](for CO and FHMM), LBM[6], and SSHMM[7] to get disaggregation results. Table 5 shows the division of training and testing data used in the techniques. Further, we ensured that training data duration was adequate for the disaggregation techniques as there was no instance in the testing data which was not in the training data.

5. Results

In this section, first, we report disaggregation performance of various existing NILM techniques. Then, with **UNUM**, we show how effective NILM data is for anomaly detection as compared to submetered appliance data.

Table 3 reports ANE for different appliances of six homes using CO, FHMM, LBM, and SSHMM separately. Appliance mapping of these homes is given in Table 4. In Table 3, few entries are > 1 meaning that the disaggregation technique predicted an appliance consumed more energy in total (sum) than it did. Considering ANE of AC (Appliance 1) and refrigerator (Appliance 2) across all homes, overall, FHMM performs better than remaining techniques. So, we chose FHMM for further steps.

Analyzing disaggregated data for anomaly detection: During testing phase, for each test day, first, we use FHMM technique to get disaggregated appliance level data. Next, we use **UNUM** on each appliance's data separately to identify anomalous instances.

Bottom row of Figure 4 shows F-score, precision and recall for refrigerator with **UNUM_D** and **UNUM_S**. Overall, precision of **UNUM_D** is found lower than **UNUM_S** because of high false positives as ANE is higher for refrigerators as compared to ACs shown in Table 3. Higher ANE in refrigerator results because NILM techniques find it difficult to track small changes due to a refrigerator ($\pm$ 90 $-$ 150 W approx.) as compared to AC ($> \pm 1$ kW)

---

[5]https://github.com/nilmtk/nilmtk
[6]https://github.com/MingjunZhong/LatentBayesianMelding
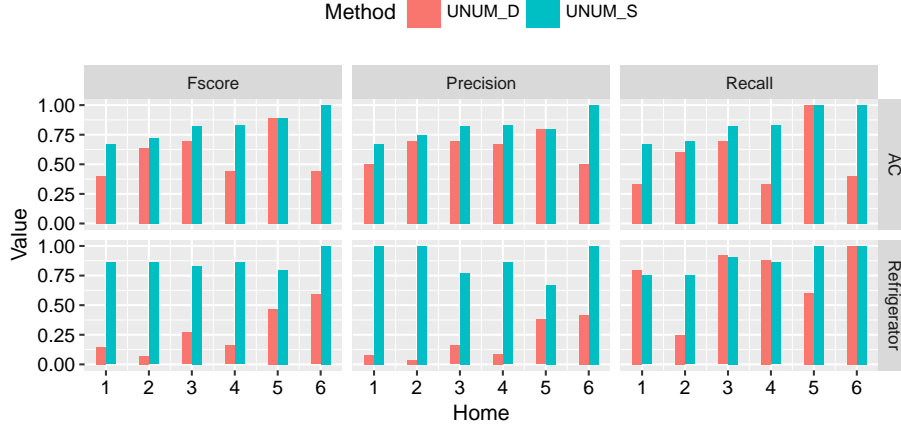[7]https://github.com/smakonin/SparseNILM

Figure 4: Anomaly detection accuracy metrics – F-score, Precision, and Recall – of **UNUM S** and **UNUM D** on AC and refrigerator data.

in the aggregate smart meter data. On the other hand, **UNUM D** recall is found to be better than **UNUM S** because high number of false positives in **UNUM D** results a drop in the false negatives.

Top row of Figure 4 shows F-score, precision and recall for AC. The figure shows **UNUM S** was not able to achieve a precision of one because of our inability to find a unique value of n (in Equations 1, 2, 3) for all homes, as energy consumption pattern of each home is distinct. Similarly, **UNUM S** recall was not able to reach a score of one because some of the defined anomalies (Table 2) were not considered as anomalous due to smaller anomaly duration parameter. Table 6 shows the number of missed anomalies in different homes using **UNUM S** both appliance-wise (AC, refrigerator) and anomaly type-wise (elongated and frequent). It shows that **UNUM S** often misses "frequent type" of anomalies in AC in all the homes. The primary reason for all these missed anomalies was their shorter duration. Duration is an important parameter in defining anomaly as shown in the Table 2. All the missed anomalies have duration < 8 hours; as a result, in a full test day's duration (day or night context) the nature of anomaly gets diluted with the normal behaviour of appliance. So, **UNUM S** fails to detect such anomalies. Similarly, for refrigerator, the duration of missed anomalies was found ≤ 7 hours. However, inability of

Home #	# of Inserted Anomalies		# of Missed Anomalies	
	AC ($= E^a$+ F)	Refrig. ($= E$+ F)	AC ($= E$+ F)	Refrig. ($= E$+ F)
1	3 + 3	2 + 2	1 + 1	1 + 0
2	4 + 3	2 + 2	0 + 1	0 + 1
3	4 + 2	2 + 2	0 + 1	0 + 1
4	5 + 1	3 + 2	0 + 1	1 + 0
5	3 + 2	2 + 2	0 + 0	0 + 0
6	2 + 3	2 + 2	0 + 0	0 + 0

[a]E = Elongated duty-cyle, F= Frequent cycling

Table 6: Number of missed anomalies at appliance level with **UNUM_S** in different homes.
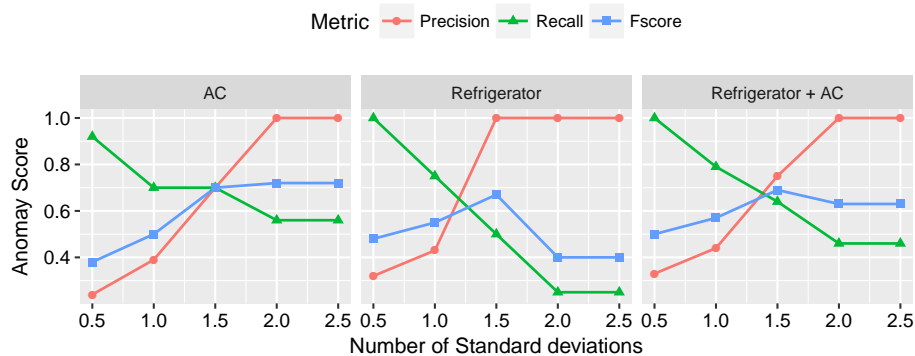
Figure 5: Effects of the number of standard deviations on Precision, Recall and F-score in AC, Refrigerator, Both Refrigerator and AC

UNUM_S to reach precision and recall of one should not affect our findings as both **UNUM_S** and **UNUM_D** are run under similar conditions and our aim is to compare their performance under such conditions.

Lower **UNUM_D**'s F-score in all homes indicates that it is difficult for existing NILM techniques to find anomalous refrigerator instances.

5.1. Sensitive Analysis of the number of Standard Deviations in **UNUM**

In Equations (1), (2), and (3), we find only one controlling parameter, n, which determines how many standard deviations from the historical consumption should be labeled as an anomaly. Intuitively, n controls the granularity of anomaly − small n means that an anomaly gets flagged if a minor deviation is observed and vice-versa. Figure 5 shows the effect of a change in the
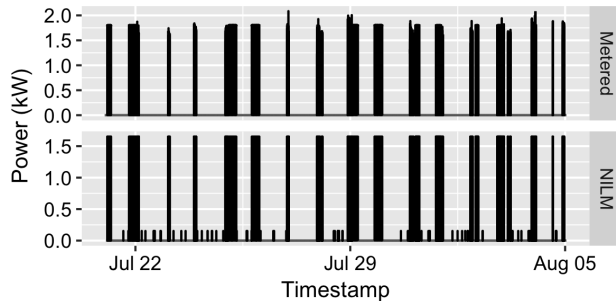
Figure 6: AC submetered & NILM data of home 5. Mirror images show NILM method has performed better.

number n of standard deviations on F-score, Precision, Recall while considering anomalies of AC, refrigerator, and both (refrigerator+AC). This figure shows that as n increases, recall decreases. This decrease happens because as n increases, anomalies with a minor deviation get treated as normal observations and hence false negatives increase. Also, with the increase in n the chances for minor deviations getting reflected as anomalies decrease which reduces false positives and as a result precision increases. As mentioned before, the trade-off between Precision and Recall can be used to set a particular value for n. Higher precision ensures fewer false alarms.

6. Discussion

In this section, we discuss our findings through key research questions.

1. *How do we know which NILM technique will perform better for anomaly detection without using* **UNUM***?*

Our experiments show that a good number of anomalous instances can be identified correctly if the ANE for an appliance is < 0.1 as reported in Table 3. Overall, ANE for AC is lowest as compared to remaining appliances, and the top row of Figure 4 shows that AC anomalies can be detected with a precision of 0.7 and recall of 0.5, approximately. On the other hand, the bottom row of the Figure 4 shows that precision is pretty low (0.12 approx.) for refrigerator due to its higher ANE as reported in Table 3.

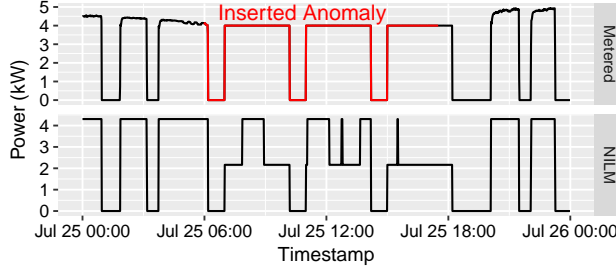This study demonstrates that we cannot directly use NILM techniques in

22

Figure 7: AC submetered & NILM data of home 3. This shows NILM method did not recover anomaly signature fully.

identifying anomalies correctly for AC or refrigerator if ANE is high. Therefore, the choice of NILM technique is determined by ANE of appliances of interest.

2. *Why is F-score for AC in Home 5 the same for submetered and NILM data as shown in Figure 4?*

Figure 6 shows AC submetered and NILM data for Home 5. We can see that FHMM was able to recover AC's consumption signature from the aggregate signal to an acceptable accuracy. The same F-score shows that NILM has worked successfully for AC in Home 5 and hence **UNUM** was able to recover all anomalous instances as compared to remaining homes. The primary reason for this good NILM performance is the remaining appliances (refrigerator, laptop, TV and water filter) of the home. All these appliances are distinct and low energy consuming appliances. As a result, FHMM resulted in better performance.

3. *Since AC of Homes 3 and 5 have the same FHMM ANE (0.07) as reported in Table 3, then why do they have different F-scores as shown in Figure 4?*

Having the same ANE does not mean that the recovered appliance signatures vary in the same pattern in both AC instances for the entire duration. For example, Figure 7 shows that in Home 3, on July 25, the inserted anomaly signature was not recovered to a required detection level in the NILM output. **UNUM** flags anomalies using tuple $(\overline{D}, \overline{E}, \sigma_e)$, but computing the values for this tuple does not flag the day's consumption as anomalous, since the signature was not accurate. On the other hand, no such case was found in Home 5 AC's NILM output. Thus, we conclude that if the appliance has an anomaly and the

23

NILM technique did recover the anomalous signature then **UNUM** will flag usage as anomalous, otherwise it will not.

7. Conclusion & Future Work

Submetering, i.e., using separate energy monitors for each appliance, to detect appliance-specific faulty behaviour is neither a scalable nor practical solution. Instead, NILM or non-intrusive load disaggregation using only as input, smart meter data, which has shown substantial progress in accurately estimating appliance level energy consumption, seems a good alternative to submetering for identifying faulty appliance behaviour at scale. In order to determine whether the reconstructed appliance-level signature generated by NILM is of sufficient fidelity to accurately detect anomalous appliance load behaviour, we used state-of-the-art NILM algorithms to generate appliance-level signatures first, and then use proposed anomaly detection algorithm on both obtained NILM as well as on submetered appliances' traces. Detailed evaluation shows that NILM output is often not accurate enough for identifying anomalies, and hence calls for proposing anomaly aware NILM methods, with some post-processing of NILM output signals to minimise the effect of noise.

We plan to extend the current work in following ways:

1. We evaluated **UNUM** on inserted anomalies due to the unavailability of anomaly annotated dataset. In future, we plan to build a system which will collect energy data and facilitate the collection of ground truth by allowing consumers to log their abnormal observations. Eventually, this will result in a more practical and large anomaly annotated dataset which can be used by the energy research community.

2. We did not consider homes with multiple instances of the same appliance. In future, we would like to evaluate NILM techniques on multiple appliance instances.

8. Acknowledgement

References

[1] T. Ganu, D. Rahayu, D. P. Seetharam, R. Kunnath, A. P. Kumar, V. Arya, S. A. Husain, S. Kalyanaraman, Socketwatch: an autonomous appliance monitoring system, in: Pervasive Computing and Communications (PerCom), 2014 IEEE International Conference on, IEEE, 2014, pp. 38–43.

[2] W. Pereira, A. Ferscha, K. 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.

[3] S. Ahmadi-Karvigh, A. Ghahramani, B. Becerik-Gerber, L. Soibelman, Real-time activity recognition for energy efficiency in buildings, Applied Energy 211 (2018) 146 – 160.

[4] D. Murray, J. Liao, L. Stankovic, V. Stankovic, Understanding usage patterns of electric kettle and energy saving potential, Elsevier Applied Energy 171 (2016) 231–242.

[5] D. Murray, L. Stankovic, V. Stankovic, N. Espinoza-Orias, Appliance electrical consumption modelling at scale using smart meter data, Elsevier Journal of Cleaner Production 187 (2018) 237–249.

[6] L. Stankovic, V. Stankovic, J. Liao, C. Wilson, Measuring the energy intensity of domestic activities from smart meter data, Elsevier Applied Energy 183 (2016) 1565–1580.

[7] D. Christensen, L. Earle, B. Sparn, Nilm applications for the energy-efficient home, Tech. rep., National Renewable Energy Lab.(NREL), Golden, CO (United States) (2012).

[8] S. Barker, S. Kalra, D. Irwin, P. Shenoy, Nilm redux: The case for emphasizing applications over accuracy, in: NILM-2014 workshop, Citeseer, 2014.

[9] J. Froehlich, E. Larson, S. Gupta, G. Cohn, M. Reynolds, S. Patel, Disaggregated end-use energy sensing for the smart grid, IEEE Pervasive Computing 10 (1) (2011) 28–39.

[10] K. He, L. Stankovic, J. Liao, V. Stankovic, Non-intrusive load disaggregation using graph signal processing, IEEE Transactions on Smart Grid 9 (3) (2018) 1739–1747.

[11] B. Zhao, K. He, L. Stankovic, V. Stankovic, Improving event-based non-intrusive load monitoring using graph signal processing, IEEE Access 6 (2018) 53944–53959.

[12] O. Parson, Overview of the nilm field, `http://blog.oliverparson.co.uk/2015/03/overview-of-nilm-field.html` (2015).

[13] K. C. Armel, A. Gupta, G. Shrimali, A. Albert, Is disaggregation the holy grail of energy efficiency? the case of electricity, Energy Policy 52 (2013) 213–234.

[14] A. Faustine, N. H. Mvungi, S. Kaijage, K. Michael, A survey on non-intrusive load monitoring methodies and techniques for energy disaggregation problem, arXiv preprint arXiv:1703.00785.

[15] Encorded, Enertalk NILM fusebook, `https://www.enertalk.com/product` (2018).

[16] S. Makonin, F. Popowich, I. V. Bajić, B. Gill, L. Bartram, Exploiting hmm sparsity to perform online real-time nonintrusive load monitoring, IEEE Transactions on Smart Grid 7 (6) (2016) 2575–2585.

[17] J. Kelly, W. 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.

[18] N. Community, Companies offering nilm products and services, `http://wiki.nilm.eu/companies.html` (2017).

[19] UK: Department of Energy & Climate Change, Smart metering equipment technical specifications: Second version (SMETS2). version 1.58., Tech. rep. (2014).

[20] G. W. Hart, Nonintrusive appliance load monitoring, Proceedings of the IEEE 80 (12) (1992) 1870–1891.

[21] J. Z. Kolter, T. S. Jaakkola, Approximate inference in additive factorial hmms with application to energy disaggregation., in: AISTATS, Vol. 22, 2012, pp. 1472–1482.

[22] M. Zhong, N. Goddard, C. Sutton, Latent bayesian melding for integrating individual and population models, in: Advances in Neural Information Processing Systems, 2015, pp. 3618–3626.

[23] EIA, International Energy Outlook, Energy Information Administration (EIA) (DOE/EIA-0484(2016)).

[24] J. Z. Kolter, M. J. Johnson, Redd: A public data set for energy disaggregation research, in: Workshop on Data Mining Applications in Sustainability (SIGKDD), San Diego, CA, Vol. 25, 2011, pp. 59–62.

[25] N. Batra, M. Gulati, A. Singh, 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, ACM, 2013, pp. 1–8.

[26] Dataport, https://www.pecanstreet.org/ (2017).

[27] J. E. Seem, Using Intelligent Data Analysis to Detect Abnormal Energy Consumption in Buildings, Energy and Buildings 39 (1) (2007) 52–58.

[28] G. Bellala, M. Marwah, M. Arlitt, G. Lyon, 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.

[29] P. Arjunan, H. D. Khadilkar, T. Ganu, Z. M. Charbiwala, A. Singh, 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.

[30] H. Rashid, N. Batra, P. Singh, Rimor: Towards identifying anomalous appliances in buildings, in: Proceedings of the 5th Conference on Systems for Built Environments, BuildSys '18, ACM, New York, NY, USA, 2018, pp. 33–42.

[31] A. Zoha, A. Gluhak, M. A. Imran, S. Rajasegarar, Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey, Sensors 12 (12) (2012) 16838–16866.

[32] B. Zhao, L. Stankovic, V. Stankovic, On a training-less solution for non-intrusive appliance load monitoring using graph signal processing, IEEE ACCESS 4 (2016) 1784–1799. `doi:10.1109/ACCESS.2016.2557460`.

[33] L. Stankovic, V. Stankovic, D. Murray, J. Liao, Energy feedback enabled by load disaggregation, in: 1st Energy Feedback Symposium, 2016, pp. 78–84.

[34] J. Liao, G. Elafoudi, L. Stankovic, V. Stankovic, Non-intrusive appliance load monitoring using low-resolution smart meter data, in: Smart grid communications (SmartGridComm), 2014 IEEE international conference on, IEEE, 2014, pp. 535–540.

[35] K. S. Barsim, R. Streubel, B. Yang, An approach for unsupervised non-intrusive load monitoring of residential appliances, in: Proceedings of the 2nd International Workshop on Non-Intrusive Load Monitoring, 2014.

[36] Y. F. Wong, Y. A. Şekercioğlu, T. Drummond, V. S. Wong, Recent approaches to non-intrusive load monitoring techniques in residential settings, in: Computational Intelligence Applications In Smart Grid (CIASG), 2013 IEEE Symposium on, IEEE, 2013, pp. 73–79.

[37] K. D. Anderson, M. E. Bergés, A. Ocneanu, D. Benitez, J. M. Moura, Event detection for non intrusive load monitoring, in: IECON 2012-38th Annual Conference on IEEE Industrial Electronics Society, IEEE, 2012, pp. 3312–3317.

[38] L. Farinaccio, R. Zmeureanu, Using a pattern recognition approach to disaggregate the total electricity consumption in a house into the major end-uses, Energy and Buildings 30 (3) (1999) 245–259.

[39] A. Zoha, A. Gluhak, M. Nati, M. A. Imran, Low-power appliance monitoring using factorial hidden markov models, in: Intelligent Sensors, Sensor Networks and Information Processing, 2013 IEEE Eighth International Conference on, IEEE, 2013, pp. 527–532.

[40] O. Parson, S. Ghosh, M. Weal, A. Rogers, Non-Intrusive Load Monitoring Using Prior Models of General Appliance Types.

[41] L. Mauch, K. S. Barsim, B. Yang, How well can hmm model load signals, in: Proceeding of the 3rd International Workshop on Non-Intrusive Load Monitoring (NILM 2016), no. 6, 2016.

[42] R. Bonfigli, E. Principi, M. Fagiani, M. Severini, S. Squartini, F. Piazza, Non-intrusive load monitoring by using active and reactive power in additive factorial hidden markov models, Applied Energy 208 (2017) 1590–1607.

[43] C. Liu, A. Akintayo, Z. Jiang, G. P. Henze, S. Sarkar, Multivariate exploration of non-intrusive load monitoring via spatiotemporal pattern network, Applied Energy 211 (2018) 1106–1122.

[44] M.-S. Tsai, Y.-H. Lin, Modern development of an adaptive non-intrusive appliance load monitoring system in electricity energy conservation, Applied Energy 96 (2012) 55–73.

[45] B. Liu, W. Luan, Y. Yu, Dynamic time warping based non-intrusive load transient identification, Applied Energy 195 (2017) 634–645.

[46] M. 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).

[47] P. R. Armstrong, C. R. Laughman, S. B. Leeb, L. K. Norford, Detection of rooftop cooling unit faults based on electrical measurements, HVAC&R Research 12 (1) (2006) 151–175.

[48] R. W. Cox, Minimally intrusive strategies for fault detection and energy monitoring, Ph.D. thesis, Massachusetts Institute of Technology (2006).

[49] N. Batra, A. Singh, K. Whitehouse, If you measure it, can you improve it? exploring the value of energy disaggregation, in: Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments, ACM, 2015, pp. 191–200.

[50] N. Batra, A. Singh, K. Whitehouse, Exploring the value of energy disaggregation through actionable feedback, in: International Workshop on Non-Intrusive Load Monitoring, 2016.

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

[52] H. Rashid, P. Singh, Poster: Energy disaggregation for identifying anomalous appliance, in: Proceedings of the Fourth ACM International Conference on Embedded Systems for Energy-Efficient Built Environments, BuildSys '17, ACM, ACM, 2017.

[53] P. J. Rousseeuw, L. Kaufman, Finding Groups in Data, Wiley Online Library, 1990.

[54] W. Kleiminger, C. Beckel, S. Santini, Household occupancy monitoring using electricity meters, in: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, ACM, 2015, pp. 975–986.

[55] A. S. Uttama Nambi, A. Reyes Lua, V. R. Prasad, Loced: Location-aware energy disaggregation framework, in: Proceedings of the 2Nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments, ACM, 2015, pp. 45–54.

[56] S. Barker, A. Mishra, D. Irwin, E. Cecchet, P. Shenoy, J. Albrecht, Smart*: An open data set and tools for enabling research in sustainable homes, SustKDD, August 111 (2012) 112.

[57] A. Monacchi, D. Egarter, W. Elmenreich, S. D'Alessandro, A. M. Tonello, Greend: an energy consumption dataset of households in italy and austria, in: Smart Grid Communications (SmartGridComm), 2014 IEEE International Conference on, IEEE, 2014, pp. 511–516.

[58] D. Murray, L. Stankovic, V. Stankovic, An electrical load measurements dataset of united kingdom households from a two-year longitudinal study, Scientific data 4 (2017) 160122.

[59] J. Kelly, W. Knottenbelt, The uk-dale dataset, domestic appliance-level electricity demand and whole-house demand from five uk homes, Scientific data 2 (2015) 150007.

[60] S. Makonin, B. Ellert, I. V. Bajić, F. Popowich, Electricity, water, and natural gas consumption of a residential house in canada from 2012 to 2014, Scientific data 3.

[61] J. Gao, S. Giri, E. C. Kara, M. Berges, Plaid: a public dataset of high-resoultion electrical appliance measurements for load identification research: demo abstract, in: BuildSys@SenSys, 2014.

[62] F. Englert, T. Schmitt, S. Kossler, A. Reinhardt, R. Steinmetz, How to auto-configure your smart home?: High-resolution power measurements to the rescue, in: Proceedings of the Fourth International Conference on Future Energy Systems, e-Energy '13, ACM, New York, NY, USA, 2013, pp. 215–224.

[63] N. Batra, J. Kelly, O. Parson, H. Dutta, W. Knottenbelt, A. Rogers, A. Singh, M. Srivastava, Nilmtk: an open source toolkit for non-intrusive load monitoring, in: Proceedings of the 5th international conference on Future energy systems, ACM, 2014, pp. 265–276.