Rimor: Towards Identifying Anomalous Appliances in Buildings

Haroon Rashid
IIIT Delhi, India
haroonr@iiitd.ac.in

Nipun Batra
University of Virginia, USA
nb2cz@virginia.edu

Pushpendra Singh
IIIT Delhi, India
pushpendra@iiitd.ac.in

ABSTRACT

Buildings across the world contribute about one-third of the total energy consumption. Studies report that anomalies in energy consumption caused by faults and abnormal appliance usage waste up to 20% of energy in buildings. Recent works leverage smart meter data to find such anomalies; however, such works do not *identify* the appliance causing the anomaly. Moreover, most of these works are not real-time and report the anomaly at the end of the day. In this paper, we propose a technique named Rimor that addresses these limitations. Rimor predicts the energy consumption of a home using historical energy data and contextual information and flags an anomaly when the actual energy consumption deviates significantly from the predicted consumption. Further, it *identifies* anomalous appliance(s) by using easy-to-collect appliance power ratings. We evaluated it on four real-world energy datasets containing 51 homes and found it to be 15% more accurate in detecting anomalies as compared to four other baseline approaches. Rimor reports an appliance identification accuracy of 82%. In addition, we also release an anomaly annotated energy dataset for the research community.

CCS CONCEPTS

• **Computing methodologies** → **Anomaly detection**; • **Hardware** → *Energy metering*;

KEYWORDS

Anomalous appliance identification, Real-time anomaly detection, smart meter, smart buildings

ACM Reference Format:
Haroon Rashid, Nipun Batra, and Pushpendra Singh. 2018. Rimor: Towards Identifying Anomalous Appliances in Buildings. In *The 5th ACM International Conference on Systems for Built Environments (BuildSys '18), November 7–8, 2018, Shenzen, China.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3276774.3276797

1 INTRODUCTION

Globally, buildings consume one-third of the total energy consumption [13]. Studies report that anomalies in buildings' energy[1] consumption caused due to faults (such as air conditioner duct leakage) and abnormal appliance usage (such as leaving lights on after usage) waste up to 20% of energy in buildings [29, 32, 34]. Studies also show that detecting such anomalies in real-time and identifying the anomalous appliance can result in more than 12% energy savings [2].

Unfortunately, the current anomaly detection techniques using smart meter[2] data only detects anomalies but does not identify the appliance causing the anomaly [9, 20, 26]. Identification of such appliances in near real-time will allow consumers to take prompt action and results in energy savings. Identifying an anomalous appliance among n appliances of a home using single a smart meter is more economical than using n separate monitors for each appliance. Furthermore, current techniques are not real-time and report the anomaly at the end of the day. These techniques use power data over the duration of an entire day to detect whether the day has anomalous usage or not [1, 8, 27]. In related areas, such as energy prediction [14, 16, 37], contextual information (such as day of week or external weather conditions) have been used, but none of the existing anomaly detection techniques in the energy domain uses such information. We believe that adding contextual information should increase the anomaly detection accuracy because the energy consumption of buildings is a function of these contextual factors.

In this paper, we propose a novel technique called Rimor[3] - that can *identify* the appliance causing the anomaly by using aggregate smart meter data and contextual information in near real-time. The basic intuition behind Rimor is that homes should have similar temporal energy patterns if we account for variations in context, such as weather. Rimor works in three steps. Firstly, it predicts the energy consumption of a target day by using historical energy data and contextual information. Secondly, it flags the energy consumption to be anomalous if there is a significant difference between the predicted energy consumption and the actual energy consumption. Lastly, it *identifies* the anomalous appliance by comparing the difference between the predicted and the actual energy consumption to the rated power consumption of appliances present in the household. The rated power consumption of different home appliances is publicly available and can be easily collected [38].

We evaluate Rimor on 51 homes from four different real-world publicly available datasets namely Dataport [12], AMPds [21], ECO [7] and REFIT [22]. All these datasets contain appliance level consumption data in addition to aggregate home consumption collected using smart meters. None of these datasets are annotated with anomalies. Moreover, no anomaly-annotated dataset is publicly available in the energy domain. Previous works [1, 8, 20, 25–27] have manually annotated datasets using domain experts, but they have not been released. For the first time, we annotate these datasets manually and make them available to the research community. We

[1] power and energy are used interchangeably

[2] A smart meter measures aggregate power usage of a home in real-time. Apart from billing, it allows data logging and can be controlled remotely by power utilities

[3] In Latin, Rimor means investigator, explorer

BuildSys '18, November 7–8, 2018, Shenzen, China

ACM ISBN 978-1-4503-5951-1/18/11…$15.00
https://doi.org/10.1145/3276774.3276797

compare Rimor with four other baselines: Multi-User Anomaly Detection (MUAD) [1], Collect, Compare and Score (CCS) [27], Twitter Anomaly Detection (TAD) [28, 36] and Real-time Anomaly Detection (RAD) [10]. We find that Rimor reports an improvement of **15%** in F-score in the detection of anomalies as compared to the baseline approaches. The anomalous appliance identification accuracy is found as **82%**. Our analysis shows that contextual information accounts for a 16% improvement in energy prediction. We produce Rimor as an open-source web application[4] that can potentially *identify* anomalous appliance in any home that has a smart meter installed. Smart meters are now becoming ubiquitous IoT devices to manage countries' energy demands[5].

The contributions of this paper are:

- We propose a near real-time anomaly detection technique, Rimor, which identifies the anomalous appliance by using easily available appliance's power ratings. Further, we show that adding contextual information improves anomaly detection accuracy.
- We evaluate Rimor with four other existing techniques. All these techniques can only detect the anomaly but do not identify the anomalous appliance. The evaluation shows that our technique improves detection accuracy by 15%.
- We make the anomaly-annotated dataset public, along with a web application of Rimor.

In section 2, we discuss the existing anomaly detection work in the building's energy consumption domain. Section 3 explains the methodology of Rimor. Section 4 explains the experimental setup. Section 5 presents the results obtained. Section 7 presents insights and the potential impact of Rimor. Section 8 discusses the future work of Rimor, and concludes the paper.

2 RELATED WORK

Anomaly detection in buildings energy consumption at aggregate smart meter level started with the use of threshold-based techniques. These techniques use thresholds defined by statistical features like daily average and peak energy usages to find anomalies [30, 31]. These techniques do not model the effect of dynamic factors affecting energy usage such as seasonality and other user contexts. In contrast, Chen et al. converted raw power readings to symbolic representation, and then used clustering to identify anomalies [9]. Moreover, Li et al. introduced a different technique, which flags anomalies if the energy consumption does not follow the classification model built from the historical consumption data [20]. Recent works [1, 8, 27] compute the anomaly score in the range [0 - 1] for the entire day's energy consumption. A major limitation of these techniques is that they detect anomalies at the end of the entire day consumption. As a result, an anomaly in the early morning hours remains undetected till day-end, which results in energy wastage for a long duration. Moreover, on detection of an anomaly at the day-end, the building administrator has to look through the entire day's usage logs to identify the exact anomalous time interval, which is a time-consuming process. All of these techniques work on offline historical energy data sets to find anomalous days. Running these techniques at smaller duration is challenging because then they have fewer data points compared to a full-day duration.

To overcome the lag in reporting anomalies, a few real-time techniques detect anomalies by comparing the actual real-time usage with the predefined allowable energy limits, usually calculated via prediction approaches. In this direction, a work by Nadai et al. [11] proposed a hybrid forecasting model with auto-regressive integrated moving average and artificial neural networks to find anomalies in gas consumption. Authors flag usage as anomalous if it is greater than some user-defined threshold. This technique is *not adaptive* to dynamic energy usage, as the defined threshold is *static*. Energy usage varies with time, and it is important to adjust the threshold adaptively. The proposed technique, Rimor, differs from this work as it does not use any such type of static threshold. Another technique by Chou et al. works in two stages, i.e., prediction and the anomaly detection stage [10]. Prediction is done on a weekly basis, and the actual usage is considered as anomalous if it is outside two standard deviations of the predicted usage. The two main limitations of this technique are: (*i*) predictions are made a week ahead, which means predictions do not include the effect of recent historical days; (*ii*) detected anomalous observations are not removed from data for future predictions. Therefore, anomalous observations affect predictions in this technique.

A major limitation of all existing techniques is that they do not account for the effect of contextual factors (weather, occupancy) in anomaly detection. Several works have shown that energy consumption is severely affected by these contexts [16, 18, 23, 37]. While looking at the limitations of the existing approaches, we propose a near-real-time anomaly detection approach, Rimor , which not only finds the anomalies in near real-time but also *identifies* the anomaly-causing appliance. Rimor uses contextual factors to detect anomalies in energy consumption data.

ALGORITHM 1: Steps in proposed anomaly detection approach, Rimor

Input: Historical power consumption Y_{train} of N train days; Contextual variables K_{real}, K_{binary}, and actual energy consumption Y_{test} of a test day; and power ratings a_l^u of n appliances present in home

Output: Anomalous appliance

1 Predict energy consumption $\widehat{Y_{test}}$ and prediction band $\widehat{Y_{test,error}}$ for the test day using Y_{train}, K_{real}, and K_{binary}

2 Compare Y_{test} with $\widehat{Y_{test}}$ and $\widehat{Y_{test,error}}$ to flag anomalous time instances of the test day

3 Identify anomalous appliance if any anomalous time instances found with

$$\arg\min_{a_l} \ (abs(\widehat{Y_{test}} - Y_{test}) - a_l^u), \forall l \in \{1, \cdots, n\}$$

return a_l /* i.e., the anomalous appliance,if any */

3 METHODOLOGY

The overall goal of Rimor is to detect anomalies and *identify* the anomaly-causing appliances in near real-time. To achieve this goal, Rimor uses contextual information for a more accurate detection of anomalies. For example, homes can have different energy patterns on weekdays and weekends, and accounting for those differences can lead to better anomaly detection. Rimor uses two types of

[4]https://github.com/loneharoon/AnomAppliance
[5]https://goo.gl/5GuaUJ

contextual information: real-valued (such as temperature and humidity) and binary (such as: is it the weekend?) denoted by K_{real} and K_{binary} respectively. Algorithm 1 provides an outline of steps in RIMOR.

Since RIMOR aims to detect anomalies in near real-time, we divide a day into W equally spaced chunks (for example $W = 24$ means 24 chunks of an hour each). Now, our problem reduces to detecting each of these $w_i \in W$ (where $i \in \{1, \ldots, |W|\}$) chunks as anomalous or not, and for anomalous chunks identifying the anomaly causing appliance. An important consideration in RIMOR is that the number of samples (T) collected from the smart meter should be greater than the chunk duration. Higher T will provide more points in a chunk and would likely be better for deciding whether the chunk has anomalous usage or not. RIMOR uses following three steps to detect anomalies and identify anomalous appliance:

I. Energy prediction: This step takes energy consumption data of the previous N days ($Y^d_{train}, \forall d \in \{1, \ldots, N\}$), the K_{real} values, and the K_{binary} contextual data points of a test[6] day as input variables to predict the energy consumption ($\widehat{Y_{test}}$) of test day.

The relation between the input variables and the energy consumption of test day can be either linear or non-linear. So two different regression approaches are used, i.e., linear regression and neural networks. Linear regression models the linear relationship between input variables ($Y^d_{train}, K_{real}, K_{binary}$) whereas neural networks model non-linear relationship, if exists [19].

• • • The proposed regression model is summarized as follows:

$$\widehat{Y^t_{test}} = C + \sum_{d=1}^{N} \alpha^d Y^{d,t}_{train} + \sum_{j=1}^{K_{real}} \beta^j Z^{j,t}_{real} + \sum_{j=1}^{K_{binary}} \gamma^j Z^{j,t}_{binary} + \epsilon^t_{test}, \forall t \in \{1, \ldots, T\} \quad (1)$$

where $\widehat{Y^t_{test}}$ represents predicted energy consumption at the t^{th} time instant of a *test* day, C represents the intercept term. α^d is the coefficient corresponding to energy consumption (Y^d_{train}) of d^{th} historical day; $Z^{j,t}_{real}$ and $Z^{j,t}_{binary}$ represent the value of the j^{th} real and binary external context variables at the t^{th} time instant of the day. β^j and γ^j correspond to the coefficients for the j^{th} real and binary contextual variables, and ϵ^t_{test} represents modelling error.

In addition to $\widehat{Y^t_{test}}$, the prediction error band is computed as:

$$\widehat{Y^t_{test,error}} = \pm\delta * \sigma^t, \forall t \in \{1, \ldots, T\} \quad (2)$$

where σ^t represents standard deviation at the t^{th} time instant of a day, and $\delta \in \{1, \ldots, 3\}$ represents number of standard deviations.

• • • Neural networks take same variables ($Y^d_{train}, K_{real}, K_{binary}$) as input and output $\widehat{Y^t_{test}}$ and $\widehat{Y^t_{test,error}}$. The exact relationship between input variables keeps changing so we cannot represent any specific form of the equation here. The set parameters of neural networks are defined in Section 4.5 (Experimental Setup) in detail.

[6]Test day refers to a day for which prediction is to be done, and train days refer to historical days

Mostly, the energy consumption throughout a day does not follow complete *stationarity*[7], so we created four separate models corresponding to every six consecutive hours of a day in both regression and neural network approaches. In this way, day-times at which complete stationarity is violated do not affect the prediction accuracy of the remaining times of the day.

II. Anomaly Detection: This step detects whether the actual energy consumption Y_{test} on a test day of any chunk w_i ($w_i \in W$) is anomalous or not. It uses a single rule - if Y_{test} deviates significantly from $\widehat{Y_{test}}$ for a duration S ($S <= \texttt{timeduration}(w_i)$) of chunk w_i, then w_i is considered as anomalous. Programmatically,

```
count = 0
S = x                          // User defined value
for t = 1 to length(w_i)
   if  Y^t_test > (Ŷ^t_test + Ŷ^t_test,error):
       count = count + 1
if count >= S:
   print (Chunk w_i is anomalous)
```

Note that the first `if` statement in the above snippet considers only the upper prediction band because appliances mostly consume higher energy in an anomalous state than in their normal state.

III. Anomalous Appliance Identification: Once an anomalous chunk is detected, the next step is to *identify* the anomalous appliance a_l which resulted in an anomaly from the set of n appliances $A = \{a_1, \ldots, a_n\}$ present in household. This step uses all n appliance's power ratings (in Watts) to identify the anomalous appliance.

Let a^u_l denotes the power rating of appliance a_l where $a_l \in \{A\}$. Among n appliances, the anomalous appliance is identified with the following equation

$$\arg\min_{a_l} (abs(\widehat{Y_{test}} - Y_{test}) - a^u_l), \forall l \in \{1, \ldots, n\} \quad (3)$$

i.,e., the appliance a_l which *minimizes* the difference between $abs(Y_{test} - \widehat{Y_{test}})$ and its power rating (a^u_l) is flagged as anomalous.

4 EVALUATION

4.1 Dataset

Energy data: We use four different publicly available datasets spanning four countries, for the evaluation of RIMOR. The datasets include Dataport [12], AMPds [21], ECO [7] and REFIT [22]. Table 1 shows the details of each dataset. We used consecutive three months of energy data with minimal missing values from each home for anomaly detection. The selected months were: June to August 2014 from Dataport and REFIT, January to March 2014 from AMPds, and August to October 2012 from ECO. While the Dataport dataset has data from more than 500 homes, we chose 24 homes having a consistent set of appliances. These datasets logged power readings at rates varying from 1 to 60 readings a minute. However, most electrical utilities log data once every 10 to 15 minutes[8]. Thus, we downsampled the data from all the datasets to 10 minutes.

[7]Mostly, at morning and evening times, energy consumption is higher and different than remaining times of a day

[8]https://www.eia.gov/consumption/residential/reports/smartmetering/pdf/assessment.pdf

Dataset	Homes used (#)	Sampling rate (s)	Appliances per home (#)	Country
Dataport	24	60	09	USA
AMPds	01	60	20	Canada
ECO	06	01	08	Switzerland
REFIT	20	08	09	UK

Table 1: Dataset characteristics

Dataset	Min # per day	Max # per day	Mean # per day	Total # of anomalies
Dataport	0	3	0.35	480
AMPds	0	1	0.25	18
ECO	0	2	0.08	30
REFIT	0	2	0.21	280
Total	0	3	0.22	808

Table 2: Statistics of anomalies present in datasets of three months duration.

Weather data: Weather data, i.e., temperature and humidity required for predictions is obtained from a publicly available weather service, WeatherUnderground[9]. The data is available at the sampling rate of thirty minutes to one hour. We up-sampled weather data to the same rate of 10 minutes as energy data via interpolation.
Appliance ratings: Power ratings of any appliance can be easily obtained by knowing the make and model of the appliance. However, in our datasets, only Dataport contained appliance make and model, and that too for a limited number of homes and appliances. Thus, we obtained the appliance power ratings using the appliance power traces made available in these datasets. Most appliances can be modeled as finite state machines (FSMs), where each state corresponds to a mode of operation (e.g. off, on, intermediate) and has an associated power draw (e.g. 0 Watts when off, 200 Watts when on) [3]. We learned the appliance ratings, i.e. the amount of power draw in different states of operations using standard clustering procedures [15].

4.2 Ground Truth Collection

Collecting ground truth information about anomalies in the building energy domain is considered a challenging and tedious task [8]. Very few studies [35] had access to the necessary infrastructure to automate the process of collecting ground truth partially. This ground truth labeling task is inherently manual and requires a domain expert to analyze (through visualizations) the total building energy data, appliance energy, weather parameters, and service logs. Most prior works [1, 8, 20, 25–27] use domain expertise to label anomalies. We followed the same approach to label the ground truth in these four datasets by consulting with a domain expert. Due to space constraints, we explain the labelling process separately at *https://github.com/loneharoon/PowerViz* in detail. Here, we briefly mention the steps followed.

(1) Plot power consumption for each month in a subplot pattern, where each subplot shows aggregate and appliances consumption for separate days.
(2) Plot weather variables temperature and humidity of the same duration in subplot style.
(3) Analyze all these plots at once and identify the deviations, if a significant deviation was observed in aggregate consumption and weather variables do not explain the deviation then we labeled that an anomaly.

Datasets used in prior works are not publicly available, but we release our anomaly labeled dataset for public reuse[10].

In total, 808 anomalies were observed across the four datasets as shown in Table 2. Dataport and ECO had the highest and the lowest mean number of anomalies per day, respectively. Table 3 shows the top-3 anomaly causing appliances across four datasets. Appliances, such as dryer, dishwasher, and washing machine cause the maximum number of anomalies. All these three appliances generally contribute significantly to the overall energy consumption in a home[11].

Dataset	Top 3 anomalous appliances (% contribution)
Dataport	Dryer (35%), Air conditioner (9%), Dishwasher (7%)
AMPds	Dryer (58%), Heat pump (29%), Oven (5%)
ECO	Fridge (54%), Dryer (27%), Washing machine (18%)
REFIT	Dishwashwer (34%), Dryer (25%), Washing machine (7%)

Table 3: Statistics on the top-3 anomaly causing appliances and the percentage of anomalies contributed by them.

4.3 Baseline Techniques

We compare the performance of Rimor with four other techniques:
I. Multi-User Anomaly Detection (MUAD)[1]: It breaks down the daily power consumption in W chunks as we did in our technique. For each chunk, this technique has two steps. First, it computes anomaly score for energy consumption by using *k-medoid* clustering algorithm for each home separately. Next, it compares anomaly scores of all homes in the same geographical locality to adjust the final anomaly score.
II. Collect, Compare, and Score (CCS) [27]: This technique takes as an input the same feature vector as MUAD. However, instead of *k-medoid* clustering, it uses Local Outlier Factor (LOF), a density-based approach to compute the anomaly score for the energy consumption in each user-defined time interval. It assigns an outlier score, i.e., the degree of outlierness, to each object based on the neighboring objects only.
III. Twitter Anomaly Detection (TAD) [28, 36]: In 2015, Twitter released an open-source anomaly detection package. This uses seasonal hybrid Extreme Studentized Deviate (ESD) algorithm, which builds upon Generalized ESD test.

[9] https://www.wunderground.com/

[10] https://goo.gl/SjozdF

[11] http://www.eia.gov/consumption/residential/data/2009/

IV. Real-time Anomaly Detection (RAD) [10]: This technique firstly uses an auto-regressive neural network-based approach for prediction, and later observation which lies outside two standard deviations from the predicted consumption are flagged as anomalies.

4.4 Evaluation Metrics

The metrics used to measure the accuracy of RIMOR are defined as:
F-score: It is interpreted as a harmonic average of precision and recall as

$$\text{F score} = 2 * \frac{precision * recall}{precision + recall} \tag{4}$$

Precision measures the percentage of true anomalies against the total number of reported anomalies. Recall measures the percentage of true reported anomalies to the total number of anomalies present. The value of the F-score ranges between [0, 1]. The higher the value, the better is the performance.
Appliance identification accuracy: Anomaly detection results in true positives, true negatives, false positives, and false negatives. To compute the true appliance identification accuracy, we consider only true positive cases by using the following formula

$$Identification\ Accuracy = \frac{Total\ \#\ of\ correct\ identified\ appliances}{Total\ \#\ of\ true\ positive\ anomalies} \tag{5}$$

Symmetric mean absolute percentage error: The anomaly detection performance of RIMOR depends on the accuracy of the prediction (Step 2: Anomaly detection). We use a standard performance metric, symmetric mean absolute percentage error (SMAPE) [33], to measure the prediction accuracy of regression and neural network models. SMAPE is defined as

$$SMAPE = \sum_{t=1}^{T} \frac{|\widehat{Y^t_{test}} - Y^t_{test}|}{|\widehat{Y^t_{test}}| + |Y^t_{test}|} \tag{6}$$

Where, $\widehat{Y^t_{test}}$ and Y^t_{test} are predicted and actual values of power consumption data at time instant t. The lesser the value of SMAPE, the better is the prediction accuracy.

4.5 Reproducible Experimental Setup

For RIMOR technique: We use leave-one-out cross-validation to find the optimal number of historical train days N using Bayesian Information Criteria (BIC). N was varied between 1 to 21 and was finally set at $N = 4$ as it resulted in the lowest BIC value. Anomaly detection rate, $W = 24$, is done at an hourly rate, and the S parameter is set to 30 minutes. We will present the sensitivity analysis of all these parameters in Section 6. Note that there is no division of training and testing dataset, RIMOR works in a rolling window manner, where the current day for which prediction (or anomaly detection) is done is considered as a test day and historical days are considered as train days. It must be noted that an outlier or anomaly in the train days can lead to learning incorrect models. Thus, we check and remove such outliers in train days before training the regression model or the neural networks using Rlof R package[12].

[12]https://cran.r-project.org/web/packages/Rlof/index.html

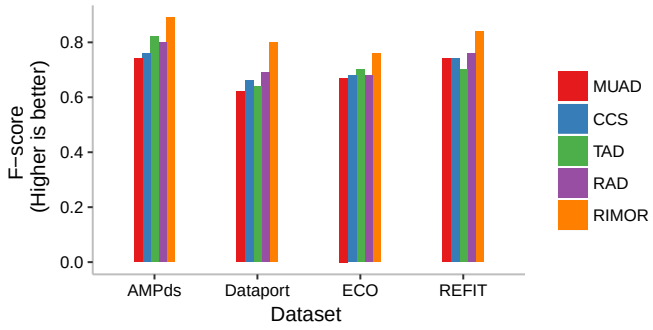


Figure 1: Our approach (RIMOR) gives 15% better accuracy in detecting anomalies.

	SMAPE (Lower is better)	
Dataset	**Regression**	**Neural Networks**
Dataport	0.60	**0.33**
AMPds	0.42	**0.29**
ECO	0.62	**0.39**
REFIT	0.54	**0.31**

Table 4: Prediction accuracy using SMAPE metric.

Existing implementation of neural networks in the `Caret` package of R was used to get prediction results. With cross-validation, the optimal neural parameters were found as: number of hidden nodes = 10, weights = 0.05, and the number of repeats as 500. For anomaly detection in Step 2 of RIMOR, two standard deviation rule was used to flag the anomalies, i.e. $\delta = 2$. All parameters were set empirically, due to space constraints we will show the analysis of important ones only. The implementation code of RIMOR is publicly available in the form of R shiny application at Github[13].
For baseline techniques: The R code of MUAD and CCS were provided by their authors. Implementation of TAD is publicly available as an `AnomalyDetection`[14] R package. RAD results were obtained using the `R Forecast` package as used by the authors of the paper. For all the four baseline techniques, we used the parameter values as suggested by their authors for optimal performance.

5 RESULTS

5.1 Anomaly Detection Performance

Our main result in Figure 1 shows that our approach, RIMOR, reports an average 15% better anomaly detection accuracy compared to the four baselines. For both our approach and the baselines, we found that the false positive rate (FPR) was generally higher than the false negative rate (FNR). For our approach, most of these false positives result due to the underprediction of actual energy consumption, which, in turn, results from irregularity in pattern across several consecutive days. In contrast, the high FPR in the baseline

[13]https://github.com/loneharoon/AnomAppliance
[14]https://github.com/twitter/AnomalyDetection

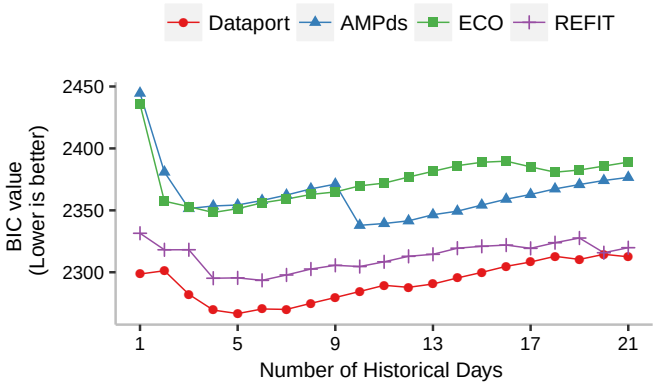


Figure 2: Effect of number of historical days N on energy consumption prediction. Using 4-5 historical days gives the best model accuracy.

Dataport	AMPds	ECO	REFIT
0.85	0.88	0.76	0.81

Table 5: Anomalous appliance identification accuracy

Home #	Appliances	Dataset	Approx. ratings (W)
1	Clothes Washer & Microwave	REFIT	0450
2	Cooktop and AC	Dataport	1200
3	Water heater and AC	Dataport	1100
4	Heat pump and wall oven	AMPds	1800
5	Furance and kitchen plugs	Dataport	0450

Table 6: Appliances found in few homes with approximately similar power ratings. Due to similar power ratings, such appliances typically lower appliance identification accuracy.

approaches is because they do not account for contextual variables, and they do not remove the outliers in the historical data.

While, across all the datasets, our approach outperforms the four baselines, the performance for AMPds and REFIT is better than the other two datasets. This can be attributed to the better energy prediction accuracy on AMPds and REFIT as shown in Table 4. AMPds and REFIT show better accuracy since the power patterns are more regular when compared to the other two datasets. We believe that there can be additional contextual parameters to improve the energy consumption prediction for Dataport and ECO.

From Table 4, we can also see that neural networks modeling the non-linear relationship between the input features and the output power consumption outperform the linear regression model. This indicates the presence of non-linear relationships among the considered data variables. Since neural networks perform better than the regression model, we use only neural network prediction results for anomaly detection and appliance identification.

5.2 Appliance Identification Performance

The appliance identification accuracy is shown in Table 5. On average, Rimor reports 82% appliance identification accuracy. We observe that the ECO dataset shows the lowest appliance identification accuracy. This is because the dryer and the washing machine, which collectively account for 45% of the anomalies in the dataset, have very similar appliance power ratings. Since our identification approach currently leverages only appliance power ratings, it will show low accuracy in the presence of appliances having similar power ratings. Table 6 shows a few pairs of appliances having a similar power rating across the different datasets.

6 SENSITIVITY ANALYSIS OF PARAMETERS

6.1 Number of Historical Days

The accuracy of predictions in prediction models depends upon the number of historical days (N). Figure 2 shows the effect of the change in the number of historical days on BIC value across the different datasets. The lower the BIC value, the better the model is. As shown, prediction models achieve lower BIC values when the days are between 4 and 5. As we increase history from 1 to 4 days, the BIC decreases, signifying that the regression models work better with more historical data. However, beyond 6 or 7 days, adding more historical information increases BIC, signifying that the additional historical data is generally less representative of the recent energy consumption trends.

6.2 W and S Parameter Selection

The W parameter controls the chunk size; for example, $W = 24$ means each chunk consists of one-hour consumption data ($= \frac{\#\text{ of hours in a day}}{W}$). The S parameter defines the threshold on the number of observations within a chunk behaving abnormally. First, we discuss the effect of S on various metrics such as precision, recall, and F score while keeping W constant, and later, we discuss the effect of varying W on the mentioned metrics.

For $W = 24$, Figures 3(a) and (b) show as S increases, both precision and recall increases for all the four homes. This means that with increasing S, the number of false positives and false negatives decreases. Our data is at a 10-minute sampling rate, so $W = 24$ means that a chunk of hour size consists of 6 observations. For this W, the chance of flagging a chunk as erroneously anomalous is higher at $S = 10$ as compared to $S = 50$, because $S = 50$ defines the behavior of five observations as compared to a single observation ($S = 10$). This means a smaller S value results in more false positives as compared to a larger S value within the same chunk size. Figure 3(c) shows the combined effect of precision and recall in terms of F score. The higher the F score, the better the performance of the technique.

For $W = 12$, i.e., two hourly chunk size, Figures 3(d), (e), and (f) show precision, recall, and F-score, respectively. Figures 3(c) and (f) show that F-scores decreases as chunk size increases. F-score of 0.8 is achieved at 50 and 90 minutes in Figures 3(c) and (f), respectively. A chunk size of one hour contains six observations while a chunk size of two hours contains 12 observations. The chances of behaving

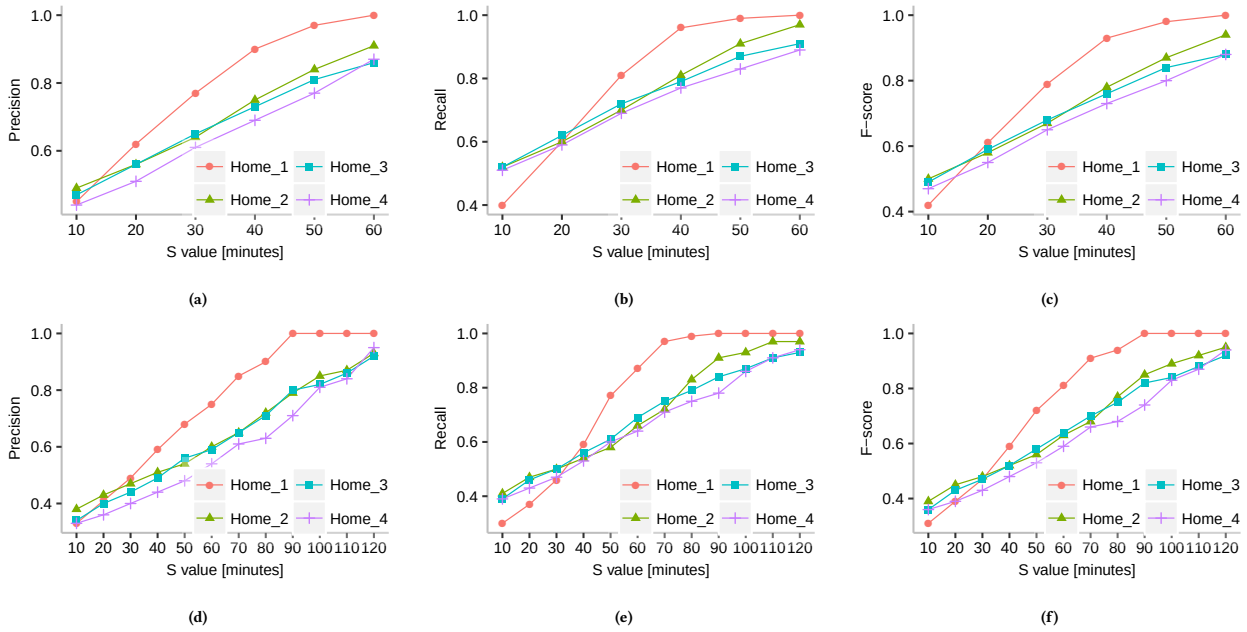


Figure 3: At W = one hour : (a) Precision, (b) Recall, (c) F-score. At W = two hours : (d) Precision, (e) Recall, (f) F-score. Figures show with increasing S both false positives and false negatives decrease. And with increasing chunk size, F-score decreases.

any three observations as erroneously anomalous are higher in 2-hour chunks as compared to one-hour chunks. As a result, the same value of S results in higher F-score in one hour chunk as compared to a two-hour chunk.

From the analysis of W and S, we conclude that for a chunk size of C minutes, the S value needs to be $> 0.5 * C$ minutes to achieve a better F-score. A domain expert can define both of these parameters according to the nature of energy consumption. If anomaly detection is critical, then W can be set as low as possible, say hourly.

6.3 Impact of Contextual Information

We hypothesized that adding contextual information would improve the anomaly detection accuracy as it would lead to a more accurate energy prediction. Figure 4 shows that adding the weekday/weekend context and weather information reduces the energy prediction errors compared to using energy data, by 6% and 11%, respectively. Adding the two contexts reduces the error by 16%, thus validating our hypothesis.

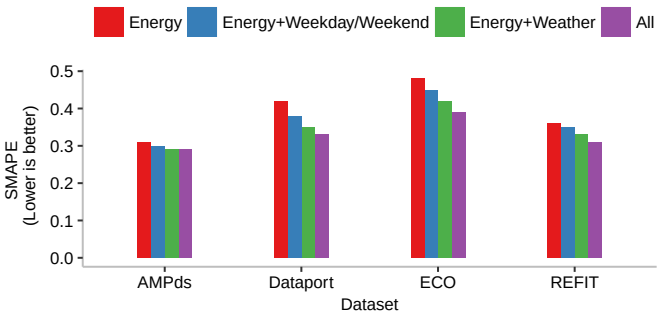


Figure 4: Adding more contextual features decreases SMAPE error and, hence, improves energy prediction.

7 INSIGHTS AND POTENTIAL IMPACT

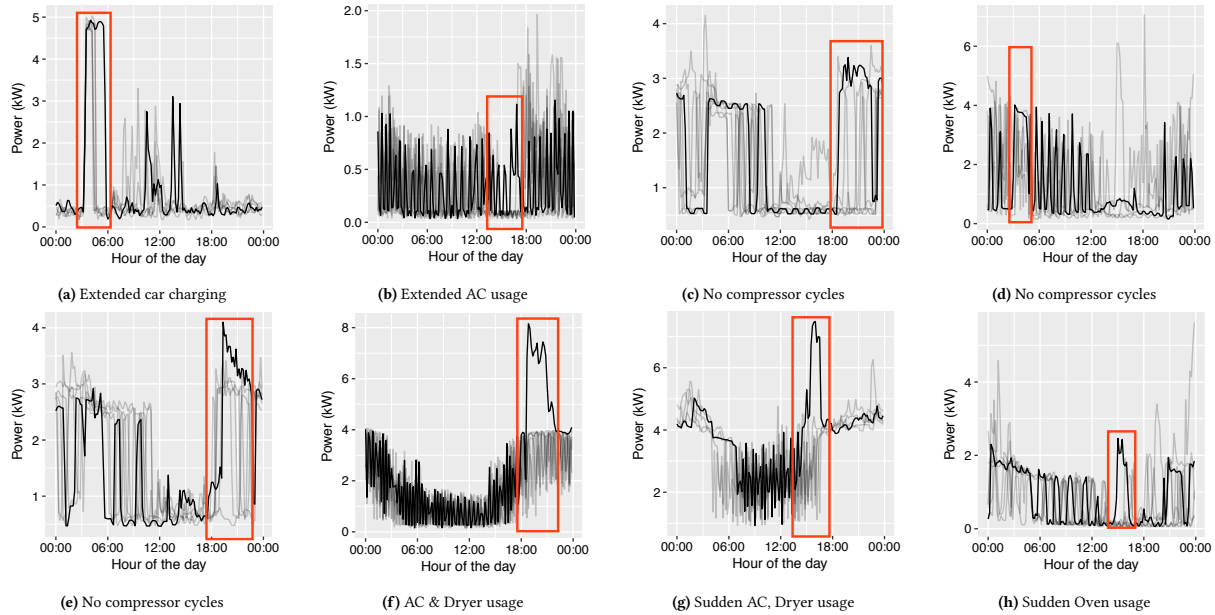
Having discussed Rimor and its evaluation, we now revisit the original motivation: can we help reduce the energy consumption by identifying anomalous usage? To answer this question, we now discuss the classification of flagged anomalies into actionable and non-actionable anomalies, followed by a discussion on the potential energy and monetary savings if these anomalies are timely rectified.

7.1 Actionable Anomalies

These anomalies represent energy wastage instances where timely action can lead to energy savings. The marked regions in Figure 5(c), (d) and, (e) show detected actionable anomalies. On these particular days, the air conditioner took a longer ON compressor cycle during the marked time duration. As a result, the aggregate usage deviated from the historical days, and this resulted in an anomaly. We observed similar external weather conditions during the previous non-anomalous days and these particular anomalous days. Thus, we believe that these anomalies can be attributed either to the air conditioner misconfiguration [5] or some fault, like air conditioner gas leakage. However, since the fault does not persist over time, we believe that the most probable cause is setpoint misconfiguration.

7.2 Non-Actionable Anomalies

Such anomalies arise when users change their energy consumption behavior as per their needs. For instance, if occupants have



(a) Extended car charging **(b)** Extended AC usage **(c)** No compressor cycles **(d)** No compressor cycles

(e) No compressor cycles **(f)** AC & Dryer usage **(g)** Sudden AC, Dryer usage **(h)** Sudden Oven usage

Figure 5: Aggregate power consumption of homes over several consecutive days. Transparent lines represent power consumption of non-anomalous days, and the opaque line that of an anomalous day. Red rectangular regions shows anomalous time intervals. The separate caption of each figure defines the cause of the anomaly.

guests at home, their energy consumption may increase. This raises an important question on the subjectivity of the definition of an anomaly. If we have additional contextual information, such as the presence of guests, this usage will not be treated as anomalous. Figure 5(a) shows a non-actionable anomaly from the dataset. We found these anomalies were flagged out due to the extended usage of the car charger. Normally, the charging used to happen for one hour over several days, but on this particular day, it took two hours. Figure 5(b) reports anomalies between 1300 - 1600 hours because the air conditioner was running continuously, which is unusual compared to historical days. Similarly, Figure 5(g) shows anomalies that occurred when the air conditioner and the dryer were operated simultaneously. Note that these time intervals are flagged as anomalous because these devices were never used simultaneously in the same time interval during historical days. Figure 5(h) shows anomalies that result from the sporadic untimely usage of the oven.

These anomalies have limited the scope of energy saving, but their detection is still important because utilities may want to target users with more consistent usage for programs such as Demand Response (DR) [24]. Having information about abnormal usage may improve the efficiency of DR programs.

7.3 Potential Energy Savings

Table 7 shows the statistics on energy consumed due to anomalies in our dataset. From each dataset, we report only homes with minimum and maximum amount of anomalous energy consumed. We also report the average amount of anomalous energy over all the homes of a dataset. In summary, this table shows that Rimor has the potential to save up to 63.5 units of anomalous energy. In relative terms, this would amount to an energy saving of around 8% per

Dataset	Min. units (kWh)	Max. units (kWh)	Avg. units (kWh)
Dataport	04	101	48.8
AMPds	93	93	93.0
ECO	53	74	64.0
REFIT	09	96	48.4
	04	101	**63.5**

Table 7: Statistics of anomalous energy units consumed over the duration of two months.

home. This shows that Rimor can save energy by using existing smart meter data and publicly available weather information. Moreover, it does not require any new hardware to be installed in the home.

7.4 Application Prototype of Rimor

We show the utility of Rimor by developing a web application[15], which requires energy consumption and weather data. Figure 6 shows a screenshot of our prototype application. Energy consumption data can be readily obtained from installed home-level smart meters, and weather data can be obtained from the same home if available or from any publicly available weather services, such as Weatherunderground[16] or openweathermap[17]. Also, the two

[15] https://github.com/loneharoon/AnomAppliance
[16] https://www.wunderground.com
[17] http://openweathermap.org

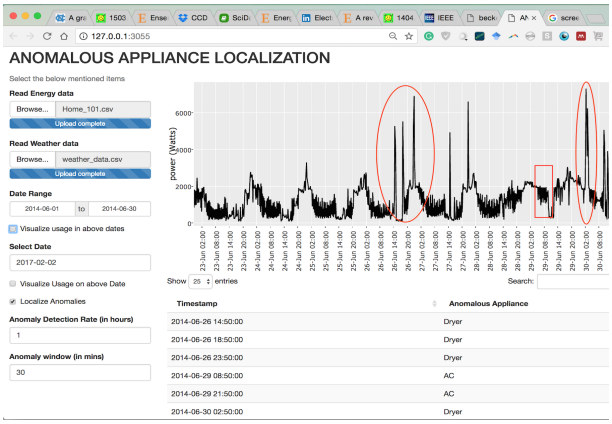


Figure 6: Screenshot of prototype application

user-defined parameters W and S are provided to tune the anomaly detection rate (chunk size) and anomaly duration, respectively. Before deploying RIMOR, we plan to:

- Develop a dashboard which will show energy consumption in real-time to a user and provide potential anomaly alerts while running RIMOR in the background. This dashboard will also include an appliance registry for a one-time registration of appliances present in a home. A homeowner needs to add only the make and model of the appliances.
- Obtain consent from the homeowners and the power utility to interface RIMOR with their smart meter data.

8 FUTURE WORK AND CONCLUSION

In the future, we plan to extend RIMOR in the following ways:

- RIMOR currently identifies appliances by using just appliance power ratings. However, as we saw in our analysis, many appliances can have similar appliance ratings. For solving such ambiguities, we plan to use the appliance time-series power signal in addition to the rated power. The appliance power signal can be obtained via non-intrusive disaggregation approaches from the aggregate smart meter data [4, 6].
- Currently, we assume that collecting appliance rating is a one-time step, but certain cases have been found where new appliances were introduced after a certain time. This can be solved by maintaining a proper appliance registry portal which can be updated.
- Currently, RIMOR cannot differentiate between actionable and non-actionable anomalies. In the future, we plan to use active learning approaches [17] to differentiate anomalies.

In this paper, we presented RIMOR to *identify* anomalous appliances in near real-time. We evaluated RIMOR on four publicly available datasets from different geographical locations and found it to be 15% better in detecting anomalies. Our results showed that adding contextual information helped to improve anomaly detection by up to 16%. Given that the data required to run our approach – smart meter data and external weather data – is readily available, we believe our application can be scaled to a large number of homes. Additionally, we make our anomaly-annotated dataset publicly available and release RIMOR in the form of a web application.

ACKNOWLEDGMENTS

Authors would like to acknowledge the support provided by ITRA project, funded by DEITy, Government of India, under a grant with Ref. No. ITRA/15(57)/Mobile/HumanSense/01. We thank Mr. Manoj Gulati for helping us in identifying anomalies in the datasets. Haroon Rashid is a TCS Ph.D. Fellow, and Pushpendra Singh is a Visvesvaraya Young Faculty Fellow.

REFERENCES

[1] P. Arjunan, H. D. Khadilkar, T. Ganu, Z. M. Charbiwala, A. Singh, and P. Singh. 2015. 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, 35–44.
[2] K Carrie Armel, Abhay Gupta, Gireesh Shrimali, and Adrian Albert. 2013. Is disaggregation the holy grail of energy efficiency? The case of electricity. *Energy Policy* 52 (2013), 213–234.
[3] Sean Barker, Sandeep Kalra, David Irwin, and Prashant Shenoy. 2013. Empirical characterization and modeling of electrical loads in smart homes. In *Green Computing Conference (IGCC), 2013 International*. IEEE, 1–10.
[4] Nipun Batra, Jack Kelly, Oliver Parson, Haimonti Dutta, William Knottenbelt, Alex Rogers, Amarjeet Singh, and Mani Srivastava. 2014. NILMTK: an open source toolkit for non-intrusive load monitoring. In *Proceedings of the 5th international conference on Future energy systems*. ACM, 265–276.
[5] Nipun Batra, Amarjeet Singh, and Kamin Whitehouse. 2015. If you measure it, can you improve it? exploring the value of disaggregation. In *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*. ACM.
[6] Nipun Batra, Amarjeet Singh, and Kamin Whitehouse. 2016. Gemello: Creating a detailed energy breakdown from just the monthly electricity bill. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 431–440.
[7] Christian Beckel, Wilhelm Kleiminger, Thorsten Staake, and Silvia Santini. 2014. The ECO data set and the performance of non-intrusive load monitoring algorithms. In *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*. ACM.
[8] G. Bellala, M. Marwah, M. Arlitt, and G. Lyon. 2011. 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.
[9] C. Chen and D. J. Cook. 2011. Energy Outlier Detection in Smart Environments. *Artificial Intelligence and Smarter Living* 11 (2011), 07.
[10] J.-S. Chou and A. S. Telaga. 2014. Real-time Detection of Anomalous Power Consumption. *Renewable and Sustainable Energy Reviews* 33 (2014), 400–411.
[11] M. D. Nadai and M. V. Someren. 2015. Short-term Anomaly Detection in Gas Consumption through ARIMA and Artificial Neural Network Forecast. In *IEEE Workshop on Environmental, Energy and Structural Monitoring Systems (EESMS),*.
[12] Dataport. 2018. https://dataport.cloud. (2 2018).
[13] EIA. 2016. International Energy Outlook. *Energy Information Administration (EIA)* DOE/EIA-0484(2016) (2016).
[14] Marcelo Espinoza, Caroline Joye, Ronnie Belmans, and Bart De Moor. 2005. Short-term load forecasting, profile identification, and customer segmentation: a methodology based on periodic time series. *IEEE Transactions on Power Systems* 20, 3 (2005), 1622–1630.
[15] George William Hart. 1992. Nonintrusive appliance load monitoring. *Proc. IEEE* 80, 12 (1992), 1870–1891.
[16] Melissa Hart and Richard de Dear. 2004. Weather sensitivity in household appliance energy end-use. *Energy and Buildings* 36, 2 (2004), 161–174.
[17] Dezhi Hong, Hongning Wang, and Kamin Whitehouse. 2015. Clustering-based active learning on sensor type classification in buildings. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM, 363–372.
[18] Amir Kavousian, Ram Rajagopal, and Martin Fischer. 2013. Determinants of residential electricity consumption: Using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior. *Energy* 55 (2013).
[19] Max Kuhn and Kjell Johnson. 2013. *Applied predictive modeling*. Vol. 26. Springer.
[20] X. Li and T. Schnier. 2010. Classification of Energy Consumption in Buildings with Outlier Detection. *IEEE Transactions on Industrial Electronics* 57, 11 (2010).
[21] Stephen Makonin, Bradley Ellert, and Fred Popowich. 2016. Electricity, water, and natural gas consumption of a residential house in Canada from 2012 to 2014.

Scientific data (2016).

[22] David Murray, Lina Stankovic, and Vladimir Stankovic. 2017. An electrical load measurements of United Kingdom households from a two-year longitudinal study. *Scientific Data* 4 (2017).

[23] Frauke Oldewurtel, Alessandra Parisio, Colin N Jones, Dimitrios Gyalistras, Markus Gwerder, Vanessa Stauch, Beat Lehmann, and Manfred Morari. 2012. Use of model predictive control and weather forecasts for energy efficient building climate control. *Energy and Buildings* 45 (2012), 15–27.

[24] Peter Palensky and Dietmar Dietrich. 2011. Demand side management: Demand response, intelligent energy systems, and smart loads. *IEEE transactions on industrial informatics* 7, 3 (2011), 381–388.

[25] M. Peña, F. Biscarri, J. I. Guerrero, I. Monedero, and C. León. 2016. Rule-based system to detect energy efficiency anomalies in smart buildings, a data mining approach. *Expert Systems with Applications* 56 (2016), 242–255.

[26] J. Ploennigs, B. Chen, A. Schumann, and N. Brady. 2013. Exploiting Generalized Additive Models for Diagnosing Abnormal Energy use in Buildings. In *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*. ACM, 1–8.

[27] H. Rashid, P. Arjunan, P. Singh, and A. Singh. 2016. Collect, Compare, and Score: A Generic Data-driven Anomaly Detection Method. In *Proceedings of the Seventh International Conference on Future Energy Systems Poster Sessions (e-Energy '16)*. ACM, New York, NY, USA, 2. DOI: http://dx.doi.org/10.1145/2939912.2942354

[28] Bernard Rosner. 1983. Percentage points for a generalized ESD many-outlier procedure. *Technometrics* 25, 2 (1983), 165–172.

[29] K. W. Roth, D. Westphalen, M. Y. Feng, P. Llana, and L. Quartararo. 2005. Energy Impact of Commercial Building Controls and Performance Diagnostics: Market Characterization, Energy Impact of Building Faults and Energy Savings Potential. *TAIX LLC for the US Department of Energy. November. 412pp* (2005).

[30] John E. Seem. 2005. Pattern Recognition Algorithm for Determining Days of the Week with Similar Energy Consumption Profiles. *Energy and Buildings* 37, 2 (2005).

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

[32] William Sisson, Constant van Aerschot, Christian Kornevall, Roger Cowe, Didier Bridoux, Thierry Braine Bonnaire, and James Fritz. 2009. Energy efficiency in buildings: Transforming the market. *Switzerland: World Business Council for Sustainable Development (WBCSD* (2009).

[33] SMAPE. 2017. http://goo.gl/ihYBj5. (12 2017).

[34] K. Srinivas and M. Brambley. 2005. Methods for Fault Detection, Diagnostics, and Prognostics. *HVAC and R Research* 11 (2005), 3–25.

[35] Shravan Srinivasan, Arunchandar Vasan, Venkatesh Sarangan, and Anand Sivasubramaniam. 2015. Bugs in the freezer: Detecting faults in supermarket refrigeration systems using energy signals. In *Proceedings of the 2015 ACM Sixth International Conference on Future Energy Systems*. ACM, 101–110.

[36] Owen Vallis, Jordan Hochenbaum, and Arun Kejariwal. 2014. A Novel Technique for Long-Term Anomaly Detection in the Cloud.. In *HotCloud*.

[37] Liping Wang, Paul Mathew, and Xiufeng Pang. 2012. Uncertainties in energy consumption introduced by building operations and weather for a medium-size office building. *Energy and Buildings* 53 (2012), 152–158.

[38] Hung-Chia Yang, Sally M Donovan, Scott J Young, Jeffery B Greenblatt, and Louis-Benoit Desroches. 2015. Assessment of household appliance surveys collected with Amazon Mechanical Turk. *Energy Efficiency* 8, 6 (2015), 1063–1075.