Monitor: An Abnormality Detection Approach in Buildings Energy Consumption

Haroon Rashid, Pushpendra Singh
{*haroonr, psingh*}@*iiitd.ac.in*
Department of Computer Science and Engineering
IIIT Delhi, New Delhi, India

***Abstract*—With the growth of smart cities, more and more buildings are now being instrumented with smart meters for providing better energy efficiency for sustainable development. Buildings consume around 39% of electrical energy worldwide and studies report that wasteful consumer behavior such as forgetting to switch off an appliance after use or using an appliance with misconfigured settings adds about one-third to buildings consumption. These instances result in deviations in energy consumption as compared to its normal consumption and are called as *abnormalities*. Detecting such abnormalities is important for reducing energy wastage. Existing methods detect abnormalities by analyzing smart meter data, however, they result in a high number of false positive alarms. This inaccuracy results in ignoring the alarms by building administrators which also affects genuine alarms. Thus, reducing the false positive alarms and making detection algorithms more accurate is a major aim.**

In this paper, we present our novel approach, called *Monitor*, which first identifies patterns in past consumption data and then uses these patterns to detect abnormalities. Our approach requires smart meter data only and reduces the rate of false positive alarms considerably. We have evaluated our approach on 16 weeks smart meter data of real world buildings. The comparison of this approach with existing approaches shows that our approach improves the accuracy by up to 24% in best scenario and on average by 14%. This improvement in accuracy reduces the rate of false positive alarms significantly and makes it more suitable for real-world deployments.

***Index Terms*—Smart cities, Energy Monitoring, Abnormality Detection, Smart Meters**

I. Introduction

As urbanization continues, energy-efficient buildings are being constructed. At the same time, existing buildings are being instrumented with smart meters to make them smart [1]. Unlike traditional billing meters, smart meters allow online control to electrical utilities and logging of energy data (current, power factor, etc) at the required frequency. Analyzing this data allows providing tailored feedback to customers through various energy efficiency programs. The United States Energy Information Administration (EIA) reports that nearly half of its electricity customers have smart meters installed[1].

Buildings consume 39% of electrical energy globally, which is expected to increase up to 43% by 2040 [2]. Around one-third of this consumption is attributed to instances of wasteful behavior by consumers, e.g., forgetting to switch off water heater after use or using AC with mis-configured settings [3]. Such instances result in deviation from normal consumption pattern in the energy data that is recorded by smart meters. This deviation is referred as *abnormality*. For example Figure 1 illustrates power consumption scenario of a residential apartment, where the x-axis represents days and the y-axis represents the power consumption. Each boxplot shows the distribution of power consumption of a specific day. As shown, the median power consumption is approximately same (= 0.18 kW) for all days except day 9, which has higher (= 0.36 kW) consumption. Apart from high median value, day 9 also contains larger abnormal values as shown by × (crosses). A subplot in the same figure on top left shows detailed hourly power consumption pattern of day 9 along with 2 neighboring days. The rectangular region in this subplot describes the abnormal values of day 9, i.e., the power consumption was significantly high from 1300 to 1700 hours. This significantly higher power usage on day 9 is an abnormality and it makes day 9 consumption as abnormal.

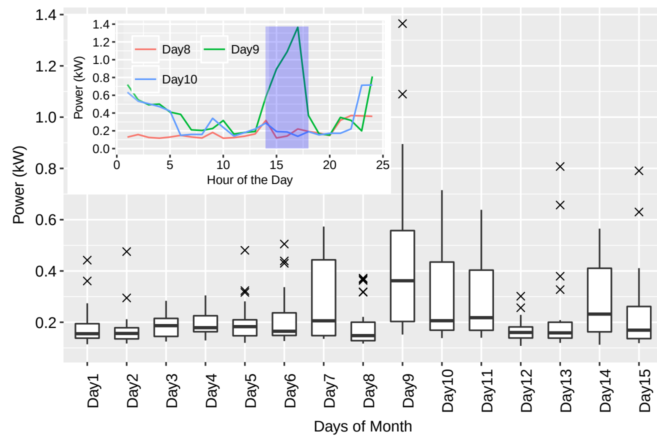
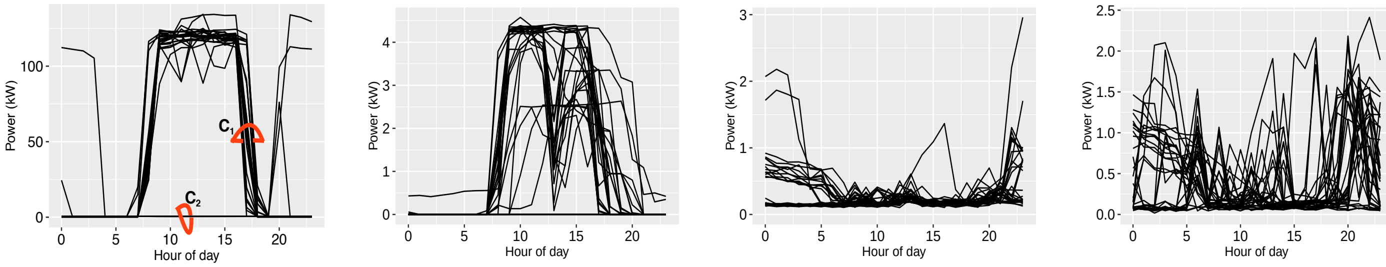


Fig. 1. Boxplots show power consumption distribution on 15 consecutive days of a month. A subplot on top left shows detailed power consumption pattern of three days. Marked region in the subplot highlights the abnormal readings of day 9.

The identification of abnormalities is essential for reducing energy[2] wastage. Timely identification can be used to prompt customers for corrective actions. As the scale of data grows, automated methods are needed to analyze and provide actionable feedback to customers as well as organizations. A study shows that giving such feedback to customers results in up to

[1]https://www.eia.gov/todayinenergy/detail.php?id=34012

[2]words "power", and "energy" are used interchangeably in this paper



(a) Commercial Bldg 1 **(b) Commercial Bldg 2** **(c) Residential Bldg1** **(d) Residential Bldg 2**

Fig. 2. Hourly average energy usage of different commercial and residential buildings for a duration of one month. Each wiggly line represents energy usage of a different day of month. The thick dark bands represent different dominant usage patterns, where each band consist of similar usage days.

8.4% energy savings [4]. However, for feedback to become effective, it is important to accurately identify instances of abnormality so as to avoid cases of false positive alarms.

A simple approach to detect abnormalities is to use threshold-based monitoring. In this approach, an alarm generates when the energy usage is higher than the defined threshold. But, this approach does not cope well with dynamic energy usage behavior of buildings. Balakrishnan et al. report that in the campus of the University of California, San Diego, this approach resulted in more than 10,000 alarms per day [5]. The high number of alarms were then ignored by the building administrators thus defeating the purpose of identifying abnormalities. Thus it is important to develop novel approaches that adapt with dynamic energy usage and are more accurate i.e. the rate of false positive alarms is low. Existing methods detect an abnormality on a particular day by considering data from all previous days. However, in doing so, these methods ignore the consumption patterns and treat all days *equally*. We show in Section VI that ignoring patterns existing in consumption affects the accuracy and leads to high false positive alarms.

This paper presents *Monitor*, which is an extension of our earlier work presented in the poster session of ACM e-Energy 2016 [6]. In our approach, we first identify prominent patterns present in past consumption; thereafter we cluster days in different groups as per their pattern; at the end, we classify an instance as abnormal if that instance differs significantly from the prominent pattern of the cluster. Our approach of first identifying patterns resembles closely with the real-world usage, e.g., in commercial settings often working weekdays have a very different and distinct pattern than non-working weekends or holidays and even in residential apartments we can see distinct patterns when the occupants are in house whole day or while they are out for regular work (say office hours). For clustering days as per their patterns, we use Local Outlier Factor (LOF) [7]. The LOF first clusters all past days into several clusters having different energy consumption patterns. Next, it computes abnormality score in the range [0 – 1] for each day within a specific cluster by using other days of the same cluster only. Therefore, days in other clusters do not affect abnormality score calculation. For Example, Figure 2(a) shows energy consumption on different days of a commercial building. Days of a different type (working, non-working) result in different clusters where days of one type usually go into one cluster. For example, Cluster C_1 represent energy consumption of working days and C_2 that of non-working days only. Both of these clusters differ due to inherent day type context. Existing methods would detect an abnormality on a particular day with respect to all past days which results in a large number of false alarms. Our proposed approach handles this limitation by first dividing days into relevant clusters. Afterward, the abnormality score is calculated locally and days in the remaining clusters do not affect abnormality score calculation.

For evaluation, We use a dataset comprising of different *loads* at IIIT-Delhi Institute campus that show different usage patterns, thus allowing us the opportunity to test our approach on diverse energy consumption patterns. For generalizability, we refer loads to different energy consuming sources such as residential apartments or HVAC chiller. The energy usage pattern of commercial and residential buildings (loads) is different as shown in Figure 2. The commercial building shows 1–3 distinct patterns while among the residential buildings only one of them show a pattern. Further, we correlate the performance of *Monitor* with two existing Abnormality Detection Methods (ADM), namely, ADM-I [8] and ADM-II [9]. Our approach, *Monitor* improves the accuracy in detecting abnormality up to 24% in best scenario and on average by 14%. The improved accuracy of *Monitor* results in lesser false alarms as compared to the existing approaches. The contributions of this paper are:

- We propose a novel approach *Monitor* to detect abnormal instances of energy usage using smart meter data only.
- We evaluate *Monitor* using commercial and residential buildings energy consumption data (Section VI).
- We analyze the performance of *Monitor* with other two existing abnormality detection approaches, ADM-I and ADM-II (Section VI-D), and show an improvement in accuracy by 24% in best scenario and on average by 14%.

II. Related work

Abnormality detection is being studied actively in various application domains such as network intrusion detection, bank fraud detection, safety critical systems [10], [11], [12]. Specific to energy domain, Srinivas et al. explain different abnormality detection approaches [13].

Among statistical approaches, Seem uses similar power consuming days of a week to detect abnormal instances; such days are identified using features like daily average, and peak energy usage [14]. While as, in [15], Seem uses statistical features, mean and standard deviation, to detect abnormal instances. This algorithm uses historical data to decide whether consumed energy is abnormal or not. Similarly, Chou et al. flag power usage of a day as abnormal if it deviates two standard deviations from the predicted usage [16].

Machine learning approaches were introduced to improve the state-of-the-art. These approaches mostly build adaptive models according to the energy consumption data; and the energy usage instances which do not follow model behavior are labeled as abnormal. Among machine learning approaches, Chen et al. have used clustering approach [17]. They first transformed numeric energy data to symbolic representation, and later used Suffix trees to find different patterns in energy usage [18]. Finally, clustering is used to identify abnormal instances. While Li et al. have used classification approach, where a day's consumption is flagged as abnormal if it does not follow the classification model built from the training data [19]. Zhang et al. have used regression to identify abnormal energy consumption days before energy prediction [20]. Bellala et al. have proposed a density based abnormality detection approach for commercial buildings [8]. Hourly power consumption readings of each day for n consecutive days were used to identify abnormal instances.

Among various contextual based abnormality detection approaches [21], [22], [9], Arjunan et al. have used temporal information to detect abnormal instances [9]. They grouped meters from a same building or from a same geographical location and finally used the group information to adjust abnormality score for each day. While as, Fontugne et al. have proposed Strip, Bind and Search (SBS) – an unsupervised abnormality detection method [23]. SBS explores the relationship among different appliances (AC, fan, lights, computer) usage behavior, *i.e.*, it groups the devices used at the same time in a particular setting. These intrinsic device relationships are uncovered using Empirical Mode Decomposition (EMD) [24]. These relationships change with seasonality and hence are purely contextual. A difference observed in any group behavior is flagged as an abnormality. Similarly, Balakrishnan et al. proposed Model, Cluster and Compare, another technique for fault diagnosis and detection [5]. It explores the relationship between different HVAC zones and builds models for similar zones to detect faults.

The proposed approach *Monitor* matches with works [8], [9], as both of these use meter data only and compute abnormality score day wise. Further, previous works have been extensively tested either under commercial [19], [8], [25], [16], [9], [26], [27], [14] or residential settings [17], [20], [9] only. We evaluate our approach under both settings and show that *Monitor* reduces FPR significantly as compared to [8], [9].

III. Methodology

Monitor takes time series power readings from a smart meter as input, and then for each day, abnormality score is computed separately. The abnormality score varies between 0 and 1. The higher the score, the higher are the chances of the day being abnormal. *Monitor* consists of three steps. Algorithm 1 shows stepwise execution of *Monitor*.

Algorithm 1: Steps of *Monitor*

Input: $X[M]$: Time series power consumption data of M days.

Output: $A[M]$: Abnormality score for each day in the range [0 – 1].

1 Transform input time series sequence X into Matrix $Y[M,T]$ where M represents number of days and T represents number of power readings per day.

2 Calculate DFT for each row of Y separately as, $Y^{'}[i,\ ] = DFT(Y[i,\ ]), \forall i \in \{1, \cdots, M\}$.

3 Calculate dissimilarity matrix $\Delta[M,M]$ for all pairs of M days with Euclidean distance measure, i.e., $\Delta[i,j] = [\sum_{k=1}^{T}(Y^{'}[i,k] - Y^{'}[j,k])^2]^{1/2}$ where $i,j \in \{1, \cdots, M\}$

4 Reduce dimensionality of Δ from M to 2 using MDS technique, i.e., $\widehat{\Delta}[M,2] = MDS(\Delta[M,M])$

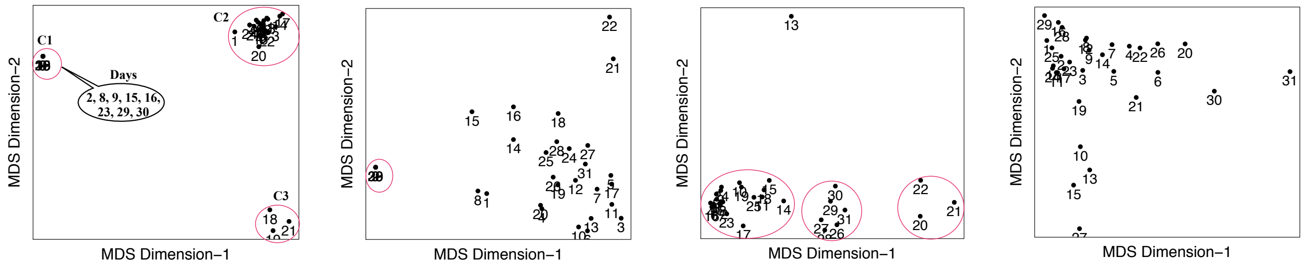
5 Calculate LOF $L[M,K]$ using $\widehat{\Delta}[M,2]$ for various K, where K refers to the number of neighbors

6 Calculate final LOF score for each day M as, $\widehat{L}[i] = max(L[i,k]), k \in \{1, \cdots, K\}, \forall i \in \{1, \cdots, M\}$

7 Standardize $\widehat{L}[M]$ in the range $[0-1]$ to calculate abnormality score for each day as $A[M] = Standardise(\widehat{L}[M])$

Step-I: In this step, firstly, it transforms input power consumption data X of M days into a Matrix Y such that each row of Y contains power consumption of a consecutive day consisting of T readings. Next, it computes frequency density vector corresponding to each row of Y separately as Matrix $Y^{'}$, since the data is periodic and it is found that frequency representation is more sensitive to abnormalities [28], [29]. Frequency representation is obtained using Discrete Fourier Transform (DFT) function [30]. It considers only magnitude part of DFT for further calculations. This contains all the necessary information of power consumption pattern.

Step-II: This step compares power consumption across all the input days of Step-I (Lines 3-4). Firstly, it computes the dissimilarity matrix $\Delta[M,M]$ by calculating the Euclidean distance between each pair of M days of $Y^{'}$. Euclidean distance reflects the underlying similarity in energy usage data. Next, it reduces the dimensionality of $\Delta[M,M]$ to $\widehat{\Delta}[M,2]$ by using Multidimensional Scaling (MDS) algorithm as the following step of *Monitor* works on lower-dimensional data [31]. MDS essentially provides a lower dimensional representation of higher (T) dimensional data. Also, MDS plots result in better visualization of energy usage among different days. Figure 3 shows corresponding MDS plots of different loads shown in Figure 2. Each • on the plot represents power consumption of a different day and a number below each • represents the respective day of the month. Each plot



(a) HVAC Chiller **(b) Lecture Block** **(c) Apartment_1** **(d) Apartment_2**

Fig. 3. Multidimensional Scaling (MDS) plots of selected commercial and residential loads. Each • represents a different day and the distribution of • shows the variation in the average power consumption on different days of a month. Circles represent clusters, i.e., days with similar power consumption pattern.

shows a different distribution of •, meaning each load follows a different energy consumption pattern. The distribution does not reflect the lower or higher power consumption days but shows the variation in power usage on various days. Days within a cluster consume energy in the similar pattern and different from the days in remaining clusters.

Step-III: This step calculates the abnormality score for each day power consumption (Lines 5-7). It uses LOF to calculate the abnormality score, which has been extensively used in various domains such as network intrusion detection [32]. LOF takes $\widehat{\Delta}[M, 2]$ as input and represents each row with a single observation thus $\widehat{\Delta}[M, 2]$ reduces to M observations. Next, it computes density for each observation with its *k* neighbors. Finally, it computes abnormality score for each observation by comparing its density with neighboring observations. Observations with significantly lower density as compared to neighbors get higher abnormality score.

LOF scores do not have a fixed range, so to interpret the scores *Monitor* standardizes output LOF scores using the approach in [33]. This approach first performs regularization and then normalization to output scores between 0 and 1.

To compute abnormality score, *k* value is specified (Line 5) in LOF. A single specific value of *k* cannot be used for all the loads as each load follows different energy usage pattern and hence result in different data distribution. Therefore, range of *k* values are used to calculate LOF scores, similar to the approach in [34]. The final LOF score considered is the maximum of LOF scores calculated with different k's as suggested in [7]. Please refer section VI-C for detailed information about selection of k and the final LOF score.

IV. Dataset

We use a real-world dataset to evaluate the performance of the *Monitor*. This dataset contains power consumption readings of both the residential and commercial loads at the IIIT-D campus.

The residential dataset contains energy consumption data of faculty apartments on the campus. Each apartment consists of three bedrooms, a hall, and a kitchen. These apartments are equipped with common home appliances, such as lighting systems, refrigerator, heaters, and air-conditioners. A smart meter is installed for *each apartment* separately. The residential apartments remain operative throughout the week (24 x 7). The apartments for our experiments were selected on the basis of: *i*) apartments with no missing readings, *ii*) apartments showing *diversity*. Diversity (Figure 2) refers to different power consumption patterns and we focus on this to check the competency of *Monitor*.

For the commercial dataset, we chose HVAC chiller, Lecture block due to two reasons: *i*) these loads mostly follow regular working hours, and *ii*) these are high energy consuming loads. The Lecture block comprises of 12 classrooms spread over the three floors. The Lecture block and HVAC chiller run usually from 0800 - 1700 hours on weekdays.

Our dataset represents energy usage across buildings because it includes all the four possible usage scenarios: (*i*) commercial loads following specific limited patterns across days (Figure 3(a)), (*ii*) commercial loads following approximate diverse usage across days (Figure 3(b)), (*iii*) residential loads following specific limited patterns across days (Figure 3(c)), and (*iv*) residential loads following approximate diverse usage across days (Figure 3(d)).

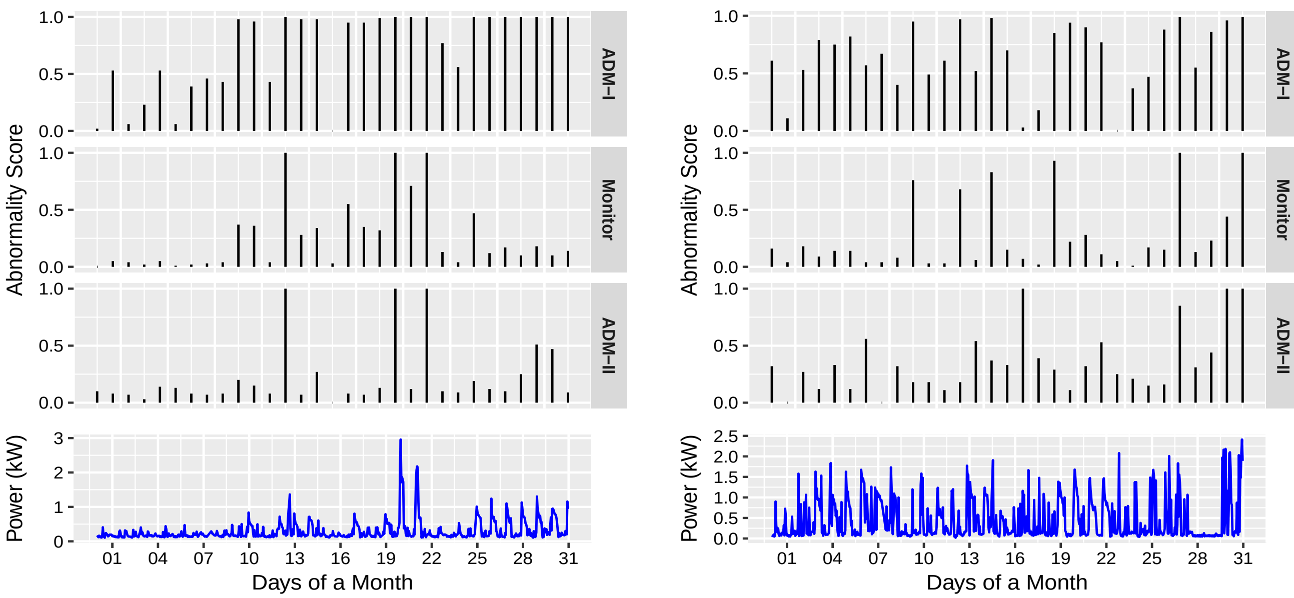
The sampling rate of smart meters is 30 seconds to get the fine-grained power consumption details. The data from all the smart meters on campus are stored in a centralized meter data aggregation system using an open source system, sMAP [35]. This dataset consists of continuous sixteen weeks of data starting from I^{st} Aug. till 29^{th} Nov. 2015. We chose this period of the year primarily for two reasons: (*i*) This is a period of the academic semester, so everyone is on campus and every electric consuming equipment remains operational. (*ii*) This period of the year shows seasonality effect, where the months of August and September are hot and are most energy consuming while the remaining two months do not normally require AC and are low energy consuming months.

We downsampled the data to hourly average readings as the power utilities generally collect data at hourly scale [36]. Therefore, the motivation is to check the performance of the *Monitor* on any diverse dataset collected by various power utilities. We have released our dataset[3] for the public use.

V. Experimental setup

Baseline Methods: We correlate the performance of *Monitor* with two baseline methods, i.e., ADM-I [8] and ADM-II [9]. We chose these well-known methods because both of these (*i*) use aggregate hourly meter readings, and (*ii*) compute abnormality score corresponding to each day. Both ADM-I and

[3]https://drive.google.com/drive/folders/0ByK27OBInnBnR1N0WEZBbVV5aDQ



(a) Apartment_1 **(b) Apartment_2**

Fig. 4. The raw power consumption of two residential apartments for a month and abnormality scores computed using ADM-I, ADM-II and *Monitor*

ADM-II use energy readings of n consecutive days to identify abnormal instances. For each day energy consumption, ADM-I uses k-NN (nearest neighbor) to compute the density and finally assigns an abnormality score with a global day having highest density. ADM-II is a clustering based abnormality detection approach. It uses k-medoid clustering algorithm and the number of clusters in data are computed using Partitioning Around Medoids algorithm.

Physical Interpretation of k**:** In *Monitor* and the baseline methods, we need k-nearest neighbors to calculate the abnormality score for each observation that corresponds to entire day energy usage. k nearest neighbors of an observation represent k other observations which deviate less from the considered observation as compared to remaining observations. Figure 3(c) shows that energy usage on day 21 differs less from days 20 and 22 as compared to remaining days on the plot. Therefore for day 21, days 20 and 22 represent two nearest neighbors.

For *Monitor*, k value is set between 4 - 7, as suggested to use a range of k values in [7]. Please refer section VI-C for sensitive analysis of k. For ADM-I, k is set to 6 as computed with the formula in the paper [8]. In ADM-II, an optimal k value is determined automatically using Partitioning Around Medoids algorithm.

Interpretation of Final Abnormality Score: The *Monitor*, ADM-I, and ADM-II calculate abnormality score for the energy usage of each day. The abnormality score is interpreted in two different ways: *i*) Top-s: In this method, a ranked list of abnormality score is sorted in the decreasing order, and then the building administrator chooses upper s days as being abnormal, *ii*) Threshold-based: In this method, building administrator selects a threshold on abnormality score, and then the days with score $\geq$ threshold are declared as abnormal. The range of the threshold is [0 – 1], similar to the abnormality score defined in [11], [12].

We use the threshold-based method as it gives finer control to an administrator, i.e., an administrator can check abnormalities of any severity level which top-s does not allow. Further, in top-s, correct information is required to extract only abnormal instances [34]. We use a threshold value of 0.75 in all the three approaches to maintain consistency. Area Under Curve (AUC) plots in Section VI-D shows the effect of threshold as these plots are drawn using all possible thresholds dynamically and hence do not limit the applicability of *Monitor* performance to a static threshold of 0.75 [37].

Abnormality Verification: Collecting ground truth information in the energy domain is an open issue. There is no publicly available abnormality annotated dataset, so all existing works consult building administrator (BA) to mark the ground truth [8], [9]. Accordingly, we obtained ground truth information from our BA. He analyzed the raw power readings manually (visual inspection) and marked all the abnormal days using historical usage pattern and the service log book entries. We will discuss the reasons for abnormal behavior as mentioned by the BA on a case by case basis.

Experimental Settings: For all loads, abnormality score for each day was computed from August 1^{st} 2015 to November 29^{th} 2015 using all the three methods. All methods were run on a rolling basis, and the window size was set to one month. The `R` implementation of ADM-II is publicly[4] available, *Monitor* and ADM-I were implemented in `R` too. In *Monitor*, `Rlof` package was used for LOF, and for LOF score standardization `HighDimOut` `R` package was used [38], [39]. We have created a publicly-available implementation of *Monitor* on GitHub[5].

[4] https://github.com/pandarasamy/anomaly_detection
[5] https://github.com/loneharoon/LoF_Anomaly

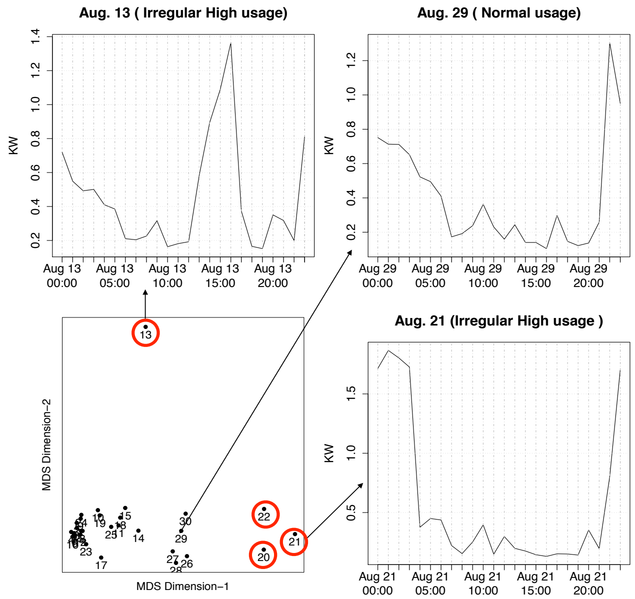


Fig. 5. MDS plot along with power consumption pattern of selected days of Apartment_1. Encircled days represent abnormal instances.

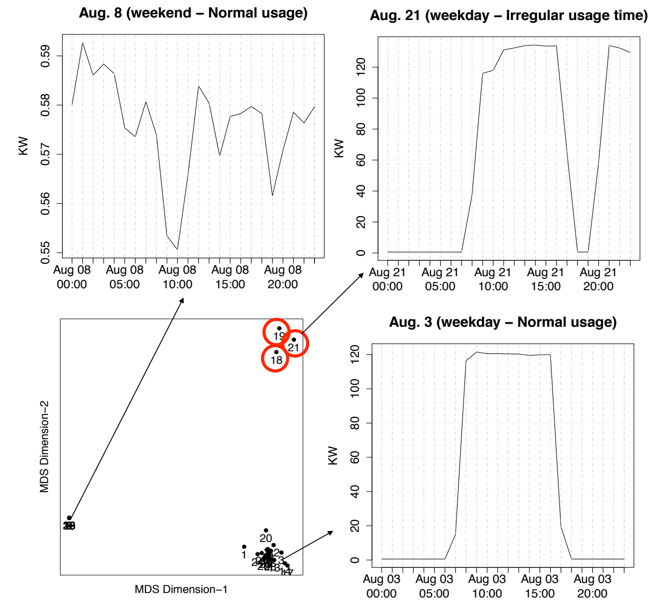


Fig. 6. MDS plot along with power consumption pattern of selected days of Chiller. Encircled days represent abnormal instances.

Metric	Method	Apart-ment_1	Apart-ment_2	Lecture Block	Chiller
False Positives	ADM-I	15	9	7	20
	ADM-II	0	1	2	2
	Monitor	0	2	0	0
False Negatives	ADM-I	0	0	2	0
	ADM-II	1	1	2	2
	Monitor	1	1	3	1

TABLE I
FALSE POSITIVES AND FALSE NEGATIVES FOUND AT A THRESHOLD OF 0.75 ABNORMALITY SCORE WITH DIFFERENT METHODS.

VI. EXPERIMENTAL RESULTS

This section initially presents the analysis of *Monitor* on residential and commercial loads, and later explains the sensitivity analysis of *k*. At the end, performance comparison of *Monitor* with baseline methods ADM-I and ADM-II using standard Area Under Curve (AUC) [37] metric is presented.

A. Analysis of residential building data

As we observe in Figure 4(a), hourly average raw power readings of apartment_1 shows two distinct patterns: (*i*) power consumption is below 1000 watts from Aug. 1 to Aug. 25 excluding days 13, 20, 21, and 22, (*ii*) power consumption is above 1000 watts from Aug. 26 until Aug. 30.

It indicates that days of the considered month should result in two distinct clusters corresponding to two usage patterns found. This observation is confirmed by the MDS plot shown in Figure 5.

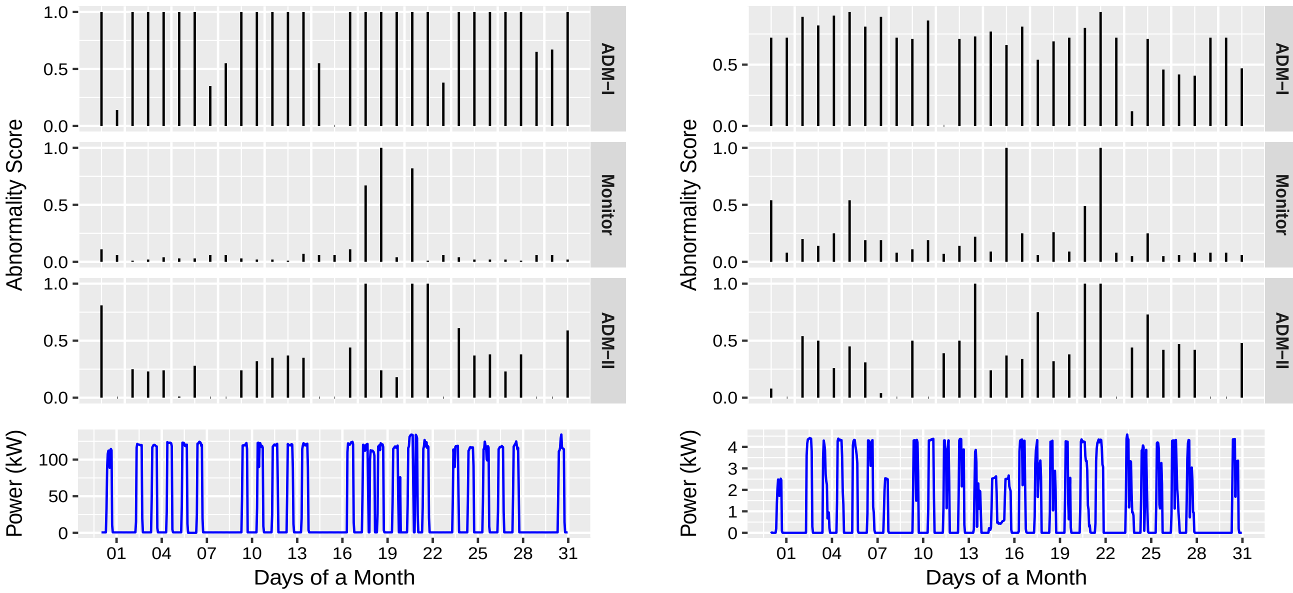
From raw power analysis, day 13 is abnormal due to unusual high power consumption between 1430 - 1620 hours as shown in a subplot of Figure 5. Similarly, days 20, 21, and 22 are abnormal due to high power usage during night hours. Discussions with the apartment owner revealed all these abnormal instances were due to AC usage during unusual day hours.

Contrarily, the raw power usage of apartment_2 in Figure 4(b) does not show any distinct pattern across days; as a result, the MDS plot in Figure 3(d) does not lead to any separate cluster. The absence of any pattern in power consumption makes it difficult to identify abnormal days accurately. But, in Figure 3(d) we find days 10, 13, 15, 27, and 31 to be different from the rest, as these days are far away from the remaining days in the plot. On manual analysis of raw power readings and discussions with the apartment owner, we found that on these days AC was running for a longer duration (day 31) continuously and at irregular intervals of the day (days 10, 13, 15, and 27).

Table I shows false positives and false negatives for both residential loads using *Monitor*, ADM-I and ADM-II. In both Apartments, *Monitor* resulted in a single false negative at a threshold of 0.75. These false negatives are day 21 of Apartment_1 and day 13 of Apartment_2. *Monitor* has assigned an abnormality score of 0.74 to both of these days as shown in Figure 4(a) and 4(b), but as the threshold is set to 0.75 both of them resulted as false negatives. Total number of abnormal days in apartment_1 and apartment_2 are 4 (days 13, 20, 21, 22) and 4 (days 15, 27, 30, 31) respectively. We discuss the cause of high false positives in the ADM-I in section VI-D.

B. Analysis of commercial buildings data

Figure 6 shows the MDS plot of an HVAC chiller. It depicts two clusters, which correspond to energy usage of weekends and weekdays. It also shows that days 18, 19, and 21 are far from both the clusters centers. On examining the raw energy usage of 21, we found that Chiller was working through late night hours as depicted in the upper right plot of Figure 6. During conversations with the building management system (BMS) operator, we were informed that in IIIT-Delhi campus two HVAC chillers are installed, which run in alternate time



(a) Chiller **(b) Lecture Block**

Fig. 7. The raw power consumption of two commercial loads for a month and abnormality scores computed using ADM-I, ADM-II and the *Monitor*

periods. But, on days 18, 19 and 21 both chillers were running simultaneously for a duration of 2 hours due to high demand, and this resulted in abnormalities on respective days.

In comparison to Chiller consumption, we find one tiny, dense cluster in the power consumption of Lecture block as shown in Figure 3(b). This cluster represents the weekend days as the power consumption is 0 for all these days. As the lecture block consists of 12 classrooms, it shows more variation in the power consumption due to the difference in the timetable for each day and due to dynamic usage patterns (either all fans/lights are on or a few). Furthermore, we found that on some days either lights/fan were switched off completely during lunch hours or hack-nights were organized. This results in dynamic usage behavior on the different days of a month.

Figure 3(b) shows that days 21 and 22 are far from the remaining days, and Figure 7(b) indicates that the *Monitor* assigns high abnormality scores to days 16 and 22. On analyzing the raw power traces of these days we found: *i*) on day 16, there was a continuous usage of 500 watts (see raw power in Figure 7(b)) from 0000 - 0800 hours, which is unusual and hence high abnormality score is assigned. During discussions with the BA, we found that on some floor lights and fans were not switched off completely. *ii*) on day 22, usual power usage was extended from 1700 - 2100 hours. From the institute calendar, we found day 22 as Institute's technical festival day and hence resulted in extra usage. This analysis shows that usage on day 16 during morning hours was energy wastage, and the *Monitor* detected it.

Table I shows false positives and false negatives for both loads with *Monitor*, ADM-I and ADM-II. For these loads, we do not find a single false positive with *Monitor* while ADM-I contains false positives in higher proportion. Total number of abnormal days in chiller and lecture block are 4 (days 18, 19, 20, 21) and 5 (days 14, 15, 16, 21, 22) respectively.

We differentiate the abnormalities detected by various methods as *actionable* and *non-actionable*. Actionable abnormalities represent energy wastage instances that can be eliminated upon timely detection. While as non-actionable represent significantly high energy usage instances that result due to some genuine cause. Although non-actionable abnormalities cannot be eliminated but their detection helps in understanding the cause of high electricity bills and thus provide the knobs to consumers for optimal appliance settings. Our results show both actionable (Day 16, Lecture block) and non-actionable abnormalities (Apartments 1 & 2, Chiller), but without discriminating the nature of abnormality detected we proved that *Monitor* is capable to detect both type of abnormalities.

C. Sensitivity Analysis

Optimal value of k: The performance of the *Monitor* depends upon k value, where k refers to the numbers of neighbors of an observation. We remove the sensitivity of *Monitor* on k by using *ensemble* approach as suggested in [34], [7]. In this approach, abnormality score is computed for a range of k values instead on a single value of k. This approach is important as it ensures that not a specific k value affects the abnormality score of an object. We used k values in the range [4 – 7]. To find the lower limit, we varied k in the range [2 – 10] for all the loads. On analyzing the results, we found that abnormality scores approximately remain same for the k range [2 – 4], so to reduce the computing cycles we set lower limit = 4. And we set higher limit = 7 owing to the reason that in most cases energy usage follows a weekly pattern due to seasonality or occupancy behavior. This means that on average the energy usage should follow similar usage pattern for 7 days.

Figure 8 explains the effect of diversity on abnormality score with different k's. We observe two patterns in the Figure: *i*) low variability score with different k, e.g., days 1 - 9, and

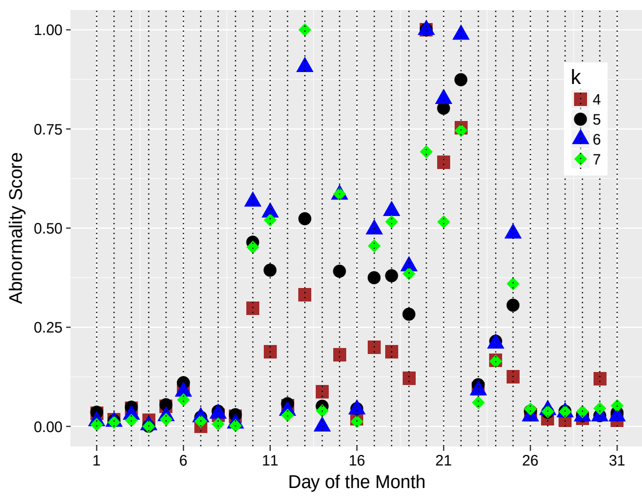


Fig. 8. Comparison of abnormality scores within each day with different values of k (4 to 7) for the Apartment_1. Days 10 – 25 show distinct variability in score due to variation in calculated density with different k neighbors.

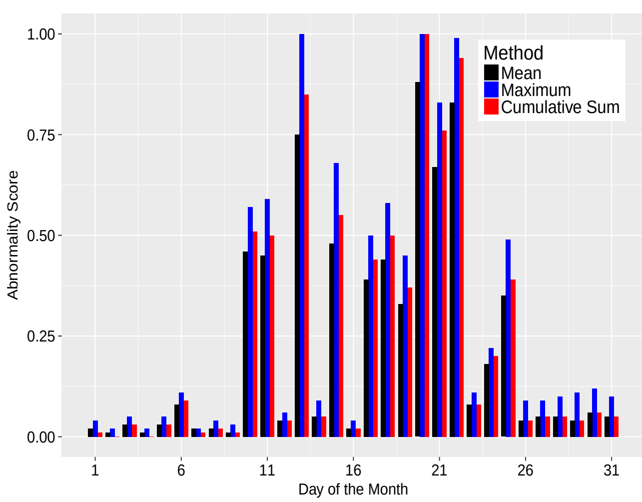


Fig. 9. Abnormality scores computed using different aggregation methods (Mean, Maximum, and Cumulative) for Apartment_1 dataset. Each method aggregates the abnormality scores calculated using different k values in the range [4 – 7].

26 - 31, *ii*) high variability in score with different k, e.g., mostly days between 10 - 25. We illustrate the effect of k on abnormality score as: assume abnormality threshold is set to 0.75, then days with abnormality score $\geq$ 0.75 are considered as abnormal. With the assumed threshold and the k value of 4 or 5, day 13 is non-abnormal while as day 20 is abnormal; on the other hand with the k value of 7, day 13 is abnormal while as day 20 is normal. But, we know both of these days are abnormal from the ground truth information (Section V). This ambiguity shows that we can not rely on a particular k value, hence we should use ensemble approach as suggested by [7], [34] to find the outlying nature of an object.

Final Abnormality Score: We find different methods to combine the abnormality scores corresponding to different k, which include: *i*) Breadth First Scheme (BFS) [40], *ii*) Cumulative sum [40], *iii*) Mean [7], and *iv*) Maximum value [7]. In both the BFS and the Cumulative sum, abnormality scores on different k are sorted and ranked according to decreasing order of their magnitudes. We empirically computed abnormality scores with all the four, and we chose Maximum value method over the rest as it always highlights the outlying instance, if an object behaves as an abnormal with any k. Also, the maximum value method is least sensitive to the cut-off value considered for the objects to be abnormal. For days 13, 21 and 22, Figure 9 shows the effect of maximum value method on abnormality score in comparison with other methods. The mean and cumulative sum dilutes the outlying nature of an object as shown in Figure 9. While as, in some datasets we found BFS is sensitive to the order of abnormality detection method, and this is highlighted in [40].

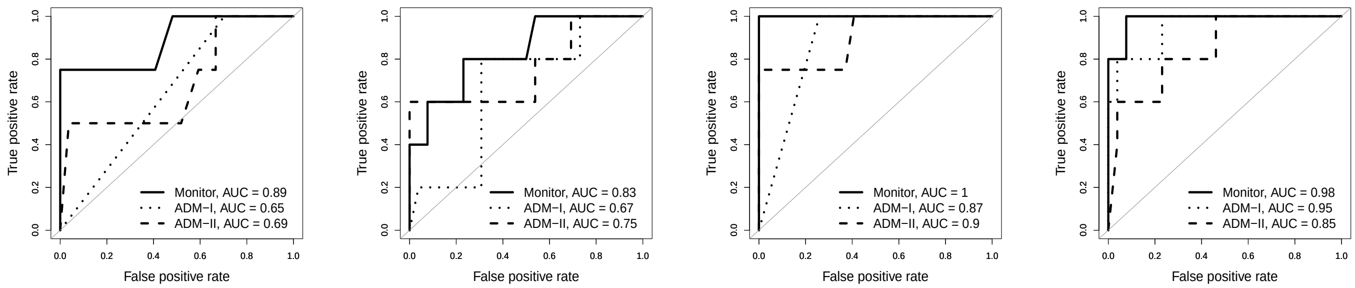
D. Performance Comparison

We use Area under curve (AUC) [37] to correlate the performance of *Monitor* with the baseline methods. Figure 10 shows receiver operating characteristic (ROC) curves for all loads with different methods. The true positive rate (TPR) in ROC curve shows the number of correctly labeled abnormalities while the false positive rate (FPR) depicts number of normal data points wrongly labeled as abnormal. The ideal ROC have TPR of 1 and FPR of 0. The performance is measured from ROC in terms of AUC value, which ranges between 0 and 1. The higher the value, the better is the performance [37].

Table II shows AUC values for *Monitor*, ADM-I and ADM-II on different loads. For instance, AUC value for *Monitor*, ADM-I and ADM-II in chiller is 0.89, 0.65, and 0.69 respectively. It is observed from the AUC values that the *Monitor* performs better than both ADM-I and ADM-II. Higher AUC of *Monitor* is attributed to its lower FPR as compared to the baseline methods. Note that ROC curve is not drawn at a particular threshold. Instead, it uses different thresholds automatically [37]. Table III summarizes percentage improvement in the AUC values of *Monitor* over the baseline methods.

1. Insights: Figures 4 and 7 show calculated abnormality scores by *Monitor*, ADM-I and ADM-II and Figure 3 shows the corresponding MDS plots. While looking at the MDS plots and the corresponding assigned abnormality scores by the ADM-I we infer that *"if there are clusters, i.e., dense clusters (e.g., Chiller, Apartment_1) found in power consumption, then the ADM-I results in high false positive rate"*. While as with the *Monitor* we do not find any impact of presence or absence of clusters on the abnormality score calculation. This inference shows that *Monitor* can be used in any similar type of energy consumption scenario as shown in Figure 3, ensuring minimal FPR.

2. High False Positive Rate in ADM-I: The abnormality score, a probability value is calculated by using a density estimation algorithm, namely k-NN. For each day's power consumption (observation), firstly density is calculated using



(a) Chiller **(b) LectureBlock** **(c) Apartment_1** **(d) Apartment_2**

Fig. 10. Comparison of ROC curves computed using baseline ADM-I, ADM-II and the proposed *Monitor* methods for different loads. It shows that the *Monitor* outperforms ADM-I and ADM-II as the AUC is high for all the loads.

Method	Chiller	Lecture Block	Apart-ment_1	Apart-ment_2
Monitor	0.89	0.83	1.00	0.98
ADM-I	0.65	0.67	0.87	0.95
ADM-II	0.69	0.75	0.90	0.85

TABLE II
AREA UNDER CURVE (AUC) VALUES FOR DIFFERENT LOADS USING *Monitor*, ADM-I, AND ADM-II

Chiller	Lecture Block	Apart-ment_1	Apart-ment_2	Average
24	14	10	8	**14**

TABLE III
AUC PERCENTAGE IMPROVEMENT IN *Monitor* OVER ADM-I & ADM-II

k-NN, and then probability value (abnormality score) is calculated with respect to the observation having highest density. This essentially means that a *single* observation with highest density influences the abnormality scores of all the remaining observations in a window.

We illustrate this with an example: In Chiller (Figure 3(a)), we found that the day 2 has highest local density of 13.7 and the average density of days (2, 8, 9, 15, 16, 23, 29, and 30) clustered in the same cluster as of day 2 is 9.17. Also, the average density of remaining 22 days of the month is 0.00020. With this statistics, it is clear that on average, each week/working day is given an abnormality score of 0.99 (= 1 - 0.00020/13.7) and each non-working day is given 0.33 (= 1-9.17/13.7). This example illustrates that ADM-I should not be used in scenarios where we have $1-2$ clear dense clusters, as in the case of Chiller. In such scenarios, an observation belonging to the densest cluster always affect the abnormality score calculation badly, and hence results in high FPR.

3. High Accuracy of *Monitor*: The clusters in data essentially depict that the points within a cluster behave similarly and different from the points in other clusters. The denser a cluster is, higher is the similarity between data points. Therefore, density is the most important metric to find similar data points.

Monitor takes advantage of this observation and uses well-known density-based LOF technique to identify abnormal days [7]. LOF uses only nearby points to calculate the abnormality score of any data point, *i.e.*, only points within a cluster are used to calculate score. We elucidate this further with an example as - consider Figure 3(a) in which we find three clusters $C1$ (represents weekend days of Aug. 2015) and $C2$ (working days of Aug. 2015) and $C3$ (abnormal days). There is no relation between $C1$ and $C2$ since both of these represent energy consumption of two different type of days. Hence, we should not use data points of one cluster to calculate abnormality score of some other data point in a different cluster. LOF uses the same concept and hence results in less FPR.

VII. LIMITATIONS AND FUTURE WORK

Monitor outperforms existing ADM-I and ADM-II by reducing the number of false alarms, but we believe there is further scope to improve the applicability of *Monitor* in real-world settings. The current implementation of *Monitor* has following limitations:

1) *Monitor* results in high false negative rate (FNR) at lower data frequencies, say at 4 hourly or daily average. This is because averaging smoothes out the effect of abnormal observations. Therefore, it is best to use *Monitor* on hourly or high-frequency data.
2) Number of detected abnormalities depends on the threshold chosen of abnormality score. Lower threshold increases FPR while higher threshold increases FNR. This is not a limitation but it gives a control to the administrator to filter abnormalities of different severity levels.
3) It computes abnormality score for each day after having observed the energy consumption trend for the day. This approach has two disadvantages: *i*) if a day gets high abnormality score, building administrator has to manually go through the entire day (24 hours) energy consumption log to find the exact abnormal usage time interval and to find the exact cause; *ii*) if abnormality occurs in the early hours of day, it results in energy wastage for the entire day. To mitigate these disadvantages, in the future, we plan to automate the labeling of abnormalities as well as reduce the resolution from 24 hours to one hour.

VIII. Conclusion

Automated abnormality detection approaches detect energy wastages timely. Unfortunately, building administrators often ignore the alarms due to their high false positive alarm rate. In this process, genuine (true positives) alarms may also get ignored leading to energy wastage. This unreliability makes the existing abnormality detection approaches useless. This paper presented *Monitor* – a reliable abnormality detection approach which reduces false alarms to a greater extent. The lesser rate of false alarms makes *Monitor* applicable for real-world building management systems. We evaluated *Monitor* on a diverse set of loads which exhibit different energy consumption patterns. *Monitor* improves the accuracy of abnormality detection by up to 24% in best scenario and on average by 14%.

References

[1] B. Foster et al., "Assessment of demand response and advanced metering," *Federal Energy Regulatory Commission, Staff Report*, 2017.

[2] EIA, "International Energy Outlook," *Energy Information Administration (EIA)*, no. DOE/EIA-0484(2016), 2016.

[3] W. Sisson, C. van Aerschot, C. Kornevall, R. Cowe, D. Bridoux, T. B. Bonnaire, and J. Fritz, "Energy efficiency in buildings: Transforming the market," *Switzerland: World Business Council for Sustainable Development (WBCSD*, 2009.

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

[5] B. Narayanaswamy, B. Balaji, R. Gupta, and Y. Agarwal, "Data Driven Investigation of Faults in HVAC Systems with Model, Cluster and Compare (MCC)," in *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings*. ACM, 2014, pp. 50–59.

[6] H. Rashid, P. Arjunan, P. Singh, and A. Singh, "Collect, compare, and score: a generic data-driven anomaly detection method for buildings," in *Proceedings of the Seventh International Conference on Future Energy Systems Poster Sessions*. ACM, 2016, p. 12.

[7] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying Density-based Local Outliers," in *ACM SIGMOD conference proceedings*, vol. 29. ACM, 2000, pp. 93–104.

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

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

[10] L. Akoglu, H. Tong, and D. Koutra, "Graph Based Anomaly Detection and Description: A Survey," *Data Mining and Knowledge Discovery*, vol. 29, no. 3, pp. 626–688, 2015.

[11] C. C. Aggarwal, *Outlier Analysis*. Springer Science + Business Media, 2013.

[12] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly Detection: A Survey," *ACM computing surveys (CSUR)*, vol. 41, no. 3, p. 15, 2009.

[13] K. Srinivas and M. Brambley, "Methods for Fault Detection, Diagnostics, and Prognostics," *HVAC and R Research*, vol. 11, pp. 3–25, 2005.

[14] J. E. Seem, "Pattern Recognition Algorithm for Determining Days of the Week with Similar Energy Consumption Profiles," *Energy and Buildings*, vol. 37, no. 2, pp. 127–139, 2005.

[15] ——, "Using Intelligent Data Analysis to Detect Abnormal Energy Consumption in Buildings," *Energy and Buildings*, vol. 39, no. 1, pp. 52–58, 2007.

[16] J.-S. Chou and A. S. Telaga, "Real-time Detection of Anomalous Power Consumption," *Renewable and Sustainable Energy Reviews*, vol. 33, pp. 400–411, 2014.

[17] C. Chen and D. J. Cook, "Energy Outlier detection in Smart Environments," *Artificial Intelligence and Smarter Living*, vol. 11, p. 07, 2011.

[18] J. Lin, E. Keogh, S. Lonardi, and B. Chiu, "A Symbolic Representation of Time Series, with Implications for Streaming Algorithms," in *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*. ACM, 2003, pp. 2–11.

[19] X. Li, C. P. Bowers, and T. Schnier, "Classification of Energy Consumption in Buildings with Outlier Detection," *IEEE Transactions on Industrial Electronics*, vol. 57, no. 11, pp. 3639–3644, 2010.

[20] Y. Zhang, W. Chen, and J. Black, "Anomaly Detection in Premise Energy Consumption Data," in *Power and Energy Society General Meeting, 2011 IEEE*. IEEE, 2011, pp. 1–8.

[21] V. M. Catterson, S. D. J. McArthur, and G. Moss, "Online Conditional Anomaly Detection in Multivariate Data for Transformer Monitoring," *IEEE Transactions on Power Delivery*, vol. 25, no. 4, pp. 2556–2564, 2010.

[22] M. Gupta, A. B. Sharma, H. Chen, and G. Jiang, "Context-aware Time Series Anomaly Detection for Complex Systems," in *WORKSHOP NOTES*, 2013, p. 14.

[23] R. Fontugne, J. Ortiz, N. Tremblay, P. Borgnat, P. Flandrin, K. Fukuda, D. Culler, and H. Esaki, "Strip, Bind, and Search: A Method for Identifying Abnormal Energy Consumption in Buildings," in *Proceedings of the 12th international conference on Information processing in sensor networks*. ACM, 2013, pp. 129–140.

[24] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The Empirical Mode Decomposition and the Hilbert Spectrum for Nonlinear and Non-stationary Time Series Analysis," in *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971. The Royal Society, 1998, pp. 903–995.

[25] J. Ploennigs, B. Chen, A. Schumann, and N. Brady, "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, 2013, pp. 1–8.

[26] A. Capozzoli, F. Lauro, and I. Khan, "Fault Detection Analysis using Data Mining Techniques for a Cluster of Smart Office Buildings," *Expert Systems with Applications*, vol. 42, no. 9, pp. 4324–4338, 2015.

[27] D. B. Araya, K. Grolinger, H. F. ElYamany, M. A. M. Capretz, and G. Bitsuamlak, "Collective Contextual Anomaly Detection Framework for Smart Buildings," in *IEEE International Joint conference on Neural Networks*, 2016.

[28] R. Agrawal, C. Faloutsos, and A. N. Swami, "Efficient similarity search in sequence databases," in *Proceedings of the 4th International Conference on Foundations of Data Organization and Algorithms*, ser. FODO '93. London, UK, UK: Springer-Verlag, 1993, pp. 69–84.

[29] X. Wu, R. Kotagiri, and K. B. Korb, *Research and Development in Knowledge Discovery and Data Mining: Second Pacific-Asia Conference, PAKDD'98, Melbourne, Australia, April 15-17, 1998, Proceedings*. Springer, 2006, vol. 1394.

[30] J. G. P. and D. G. M., *Digital signal processing: principles, algorithms, and applications*. Pearson Prentice Hall, 2007.

[31] T. F. Cox and M. A. Cox, *Multidimensional scaling*. CRC press, 2000.

[32] A. Lazarevic, L. Ertöz, V. Kumar, A. Ozgur, and J. Srivastava, "A comparative study of anomaly detection schemes in network intrusion detection." in *SDM*. SIAM, 2003, pp. 25–36.

[33] H.-P. Kriegel, P. Kröger, E. Schubert, and A. Zimek, "Interpreting and Unifying Outlier Scores." in *SDM*. SIAM, 2011, pp. 13–24.

[34] A. Zimek, R. J. G. B. Campello, and J. Sander, "Ensembles for Unsupervised Outlier Detection: Challenges and Research Questions," *ACM SIGKDD Explorations Newsletter*, vol. 15, no. 1, pp. 11–22, 2014.

[35] S. D.-Haggerty, X. Jiang, G. Tolle, J. Ortiz, and D. Culler, "sMAP: A Simple Measurement and Actuation Profile for Physical Information," in *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*. ACM, 2010, pp. 197–210.

[36] EIA, "https://www.eia.gov/tools/faqs/faq.php?id=108&t=1", 6 2017.

[37] T. Fawcett, "An Introduction to ROC Analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.

[38] Y. Hu, W. Murray, Y. Shan, and Australia., *Rlof: R Parallel Implementation of Local Outlier Factor(LOF)*, 2015, r package version 1.1.1. [Online]. Available: https://CRAN.R-project.org/package=Rlof

[39] C. Fan, *HighDimOut: Outlier Detection Algorithms for High-Dimensional Data*, 2015, r package version 1.0.0. [Online]. Available: https://CRAN.R-project.org/package=HighDimOut

[40] A. Lazarevic and V. Kumar, "Feature Bagging for Outlier Detection," in *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*. ACM, 2005, pp. 157–166.