Energy performance based anomaly detection in non-residential buildings using symbolic aggregate approximation
Ashouri, Araz; Hu, Yitian; Newsham, Guy R.; Shen, Weiming

This publication could be one of several versions: author’s original, accepted manuscript or the publisher’s version. / La version de cette publication peut être l’une des suivantes : la version prépublication de l’auteur, la version acceptée du manuscrit ou la version de l’éditeur.

For the publisher’s version, please access the DOI link below./ Pour consulter la version de l’éditeur, utilisez le lien DOI ci-dessous.

Publisher’s version / Version de l’éditeur:
https://doi.org/10.1109/CASE.2018.8560433
2018 IEEE 14th International Conference on Automation Science and Engineering (CASE), pp. 1400-1405, 2018-12-06

NRC Publications Archive Record / Notice des Archives des publications du CNRC :
https://nrc-publications.canada.ca/eng/view/object/?id=4a2c0581-f009-4692-a91c-c6c91a1f71d9
https://publications-cnrc.canada.ca/fra/voir/objet/?id=4a2c0581-f009-4692-a91c-c6c91a1f71d9

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at
https://nrc-publications.canada.ca/eng/copyright
READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L’accès à ce site Web et l’utilisation de son contenu sont assujettis aux conditions présentées dans le site
https://publications-cnrc.canada.ca/fra/droits
LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D’UTILISER CE SITE WEB.

Questions? Contact the NRC Publications Archive team at
PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the
first page of the publication for their contact information.

Vous avez des questions? Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la
première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n’arrivez
pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.
Energy Performance Based Anomaly Detection in Non-Residential Buildings Using Symbolic Aggregate Approximation*

Araz Ashouri, Yitian Hu, Guy R. Newsham, and Weiming Shen

Abstract—Building system faults in commercial and office buildings can result in a reduced occupant comfort and increased utility bills. Energy performance-based anomaly detection helps operators efficiently identify faults. In this work, a data-driven method for anomaly detection is presented. Using a symbolic aggregate method, the weekly energy demand profiles are statistically quantised and labeled to determine normal and abnormal building behaviours. A case study with three federal office buildings has been conducted to demonstrate the proposed method. The resulting technology provides building operators with easily-interpreted and actionable information for optimised building performance.

I. INTRODUCTION

Buildings are responsible for about 40% of end-use energy consumption in developed countries [1]. This implies that increasing energy efficiency of buildings and improving how they are operated can greatly contribute to reduction of both energy costs and greenhouse gas emissions. Although such efforts are usually addressed during the design phase, many studies show that buildings do not perform as expected, and that a majority of energy loss comes from errors occurring either in the construction phase or during the actual lifespan of the building, introduced by deteriorating materials, malfunctioning devices, and non-optimal operating commands [2], [3].

Traditionally, stakeholders tackle the problem of anomalous building systems behaviour by engaging technical consultants to inspect the building and to review energy consumption data. In reality, many non-residential building portfolios such as university campuses, military bases, banks, post offices, etc. are maintained by a limited number of facility operators, who do not have the resources to detect every system malfunction, such as equipment failures and unnecessary lighting during unoccupied periods [4]. Furthermore, with the rise of fully automated buildings, these systems are becoming more sophisticated; a modern non-residential building is often equipped with thousands of energy-related sensors, actuators, and controllers [5]. The resulting streams of raw data from these sensors are typically collected and temporarily stored by the building automation system (BAS) at regular time intervals. Ideally, a building energy management system (BEMS) is placed on top of the BAS, and will archive such data for long-term analysis purposes. At a first glance, this seems to be a great opportunity for building operators and energy data scientists, given the availability of data-driven tools for building monitoring, benchmarking, and control. However, these systems accumulate a vast amount of data over time, creating new challenges in efficient data interpretation. As a result, building energy researchers have been paying closer attention to methods that can enable them to analyse large building datasets for the purpose of remote auditing and automated anomaly detection. A recent review [6] indicates that researchers are widely targeting the topic of anomaly detection using analyses of the whole building energy demand. The two major approaches used in such analyses are data-driven methods [7] and model-based methods [8]. While model-based approaches are capable of simulating the behaviour of buildings under various boundary conditions, they rely on a relatively detailed level of knowledge about the building structure, parameters, and occupant behaviour. On the contrary, data-driven methods can be deployed more quickly by data analysts, though some knowledge of building physics and systems is essential for meaningful interpretation. In addition, the recent advancements in machine learning and data mining techniques have led to several methods that may be leveraged by building energy scientists [9].

Among the various techniques of data-driven anomaly detection, a method called Symbolic Aggregate approximation (SAX), first introduced by Lin et al. in 2003 [10], shows superior performance with regards to larger datasets. The SAX method incorporates a value quantisation for given intervals within a data time series, and facilitates large-scale evaluations by reducing the dimension of data from \( N \) (length of time series) to \( D \) (number of “letters” in the SAX “word”) where \( D << N \). As a result, many applications of SAX method can be found in the literature for motif selection and anomaly detection [11], [12]. Given these advantages, in this study SAX has been selected as the method to identify anomalous energy behaviour in a group of non-residential buildings.

The rest of the paper is organized as follows. Section II discusses the proposed method and demonstrates the adjustments that distinguished our method from similar works, while Section III is dedicated to the analysis of results for the studied cases and discusses the capabilities of the labeling system for various applications. Three federal government office buildings, referred to as buildings B1, B2, and B3 are investigated in the case study.

Although the proposed method is applicable to various forms of energy demand, our focus in this study is solely on electricity. Furthermore, demand data from building B2 are used to elaborate the deployment of techniques in Section II.
II. METHODOLOGY

The anomaly detection method used in this article consists of several modules that were progressively developed and fine-tuned. The modules are responsible for data cleaning, processing, and post-processing analysis. Figure 1 shows a block diagram describing the modules and the connections among them.

The four main modules of the proposed method are

1) **Data cleaning**: the block where pre-processing of raw data takes place;
2) **Statistical quantisation**: the block where demand profiles are transformed into “words”; in this case words are a string of letters than concisely summarize the weekly energy load shape of a building;
3) **Motif selection**: the block responsible for finding the “Normal” word and the corresponding normal annual demand; and
4) **Anomaly detection**: the final block where a comparison of normal and actual demands is performed, followed by a qualitative labeling of results.

In the followings, each aforementioned block is described in more details.

A. **Data cleaning**

The anomaly detection process relies on a historical dataset of demand profiles (also referred to as load profile, energy consumption, etc.) corresponding to a single or a dataset of demand profiles (also referred to as load profile, energy consumption, etc.) corresponding to a single or a combination of collection points in one or several buildings. In the case of data collection for electricity usage, a meter is installed either at the transformer level or right after entering the building. In modern commercial and office buildings, the sensory data are collected and integrated in the BAS which can be access directly, or through a high-level platform such as the BEMS. The latter provides a user-friendly environment to enable building operators to access energy information.

The federal buildings studied in this project are monitored using a BEMS called Kaizen, a commercial product from CopperTree Analytics [13]. Figure 2 shows a sample log of the hourly electricity demand profile for building B2. A brief visual inspection of the data reveals a gap in the first week of the demand profile, plotted as a horizontal line in the figure.

The purpose of the data cleaning module is to handle such gaps in the data, preventing the anomaly detection tool from flagging the period as abnormal while it could simply correspond to a temporary power outages in the sensor or an out-of-memory buffer.

![Figure 1. Block diagram showing the main four modules (in color) and their connections.](image)

![Figure 2. Electricity demand for building B2 for two weeks in May 2014.](image)

The module involves two main tasks of Outlier Elimination and Missing-Value Imputation. A detailed description of the algorithms used for these tasks is out of the scope of this article and will be discussed elsewhere.

B. **Statistical Quantisation**

Once the demand profiles are pre-processed and cleaned, the next step is to convert them into a form suitable for anomaly detection. The demand profiles are time series with an hourly resolution. The anomaly detection method incorporates a statistical quantisation technique based on the SAX method to transform profiles into words. The choice of the SAX method for anomaly detection in commercial buildings was inspired by the work described in [14]. However, we propose a more sophisticated approach at the quantisation stage to improve the performance.

Our implementation of the statistical quantisation consists of the following main stages: normalization, temporal division, sampling, and SAX transformation. Each stage is briefly explained in the following sections.

**Normalization**: This stage involves calculating a standard score (or z-score) normalized version of the cleaned electricity demand profile $p^\text{elec}$, using

$$p^\text{elec}_\text{normal} = \frac{1}{\sigma} \cdot (p^\text{elec} - \mu)$$

where, $\sigma$ is the standard deviation and $\mu$ is the mean of the electricity demand.

**Temporal division**: After normalization, a temporal division of the data according to primary (weekly) and secondary (half-daily) periods is performed. The available dataset has a resolution of 1 hour; hence a weekly profile is presented as a vector with 168 elements. A week, in our definition, starts on a Monday at 6am and ends on the following Monday at 5:59am. The half-daily sections are defined as 6am – 5:59pm (day time) and 6pm – 5:59am (night time).

**Sampling**: In order to reduce the dimension of data, the weekly demand profiles with 168 hours are averaged for each half-day section, resulting in a vector with a length of 14.
Representing extreme high demand and high demand, are taken from the expected-high demand density, demand density, respectively. Similarly, the letters are assigned to a letters (nine values per week). Furthermore, public holidays are values during weekends and night time during weekdays expected-low demands, comprised of the averaged demand day time during weekdays (five values per week), and the demands, comprised of the averaged demand values for the divide the demand set into two groups: The expected-high meaningful approach to select the cut-off thresholds. First we occurrence for letters are equal. However, we use a more letters (nine values per week). Furthermore, public holidays are cutting the probability density of the average demands (i.e. the orange lines in Figure 3) into disjoint sets.

The method used for the cutting in most studies (such as [14]) is an equiprobable cut, implying that the probability of occurrence for letters are equal. However, we use a more meaningful approach to select the cut-off thresholds. First we divide the demand set into two groups: The expected-high demands, comprised of the averaged demand values for the day time during weekdays (five values per week), and the expected-low demands, comprised of the averaged demand values during weekends and night time during weekdays (nine values per week). Furthermore, public holidays are treated the same as weekends. These two demand sets are not disjoint but are usually separated well enough so that they can be used to define the cut-off thresholds.

Figure 4 shows the normalized probability density of the average electricity demand for building B2 with the expected high and low demand densities. The cut-off thresholds used for SAX quantisation are calculated as shown in Table 1. The letters \( L \) (for extreme low demand) and \( l \) (for low demand) are assigned to a 50% and 40% portion of the expected low demand density, respectively. Similarly, the letters \( H \) and \( h \) are taken from the expected-high demand density, representing extreme high demand and high demand, respectively.

Table 1. The cut-off thresholds used for SAX quantisation.

<table>
<thead>
<tr>
<th>Letter</th>
<th>Share in the expected-low demand</th>
<th>Share in the expected-high demand</th>
</tr>
</thead>
<tbody>
<tr>
<td>( L )</td>
<td>50%</td>
<td>0</td>
</tr>
<tr>
<td>( l )</td>
<td>40%</td>
<td>0</td>
</tr>
<tr>
<td>( m )</td>
<td>10%</td>
<td>10%</td>
</tr>
<tr>
<td>( h )</td>
<td>0</td>
<td>40%</td>
</tr>
<tr>
<td>( H )</td>
<td>0</td>
<td>50%</td>
</tr>
</tbody>
</table>

As an example, a weekly demand profile for building B2 is shown in Figure 3 with the 12-hour sections and the corresponding averaged demands.

**SAX transformation:** At the final stage, the SAX method is applied to the sampled data. As described in [14], SAX is performed by quantising the values in data and transforming them into letters. This means that the demand vector of length 14 that represents a weekly profile will be transformed into a word with 14 letters. The mapping of average values into letters is done by a statistical approach that involves cutting the probability density of the average demands (i.e. the orange lines in Figure 3) into disjoint sets.

The method for cutting in most studies (such as [14]) is an equiprobable cut, implying that the probability of occurrence for letters are equal. However, we use a more meaningful approach to select the cut-off thresholds. First we divide the demand set into two groups: The expected-high demands, comprised of the averaged demand values for the day time during weekdays (five values per week), and the expected-low demands, comprised of the averaged demand values during weekends and night time during weekdays (nine values per week). Furthermore, public holidays are treated the same as weekends. These two demand sets are not disjoint but are usually separated well enough so that they can be used to define the cut-off thresholds.

Figure 4 shows the normalized probability density of the average electricity demand for building B2 with the expected high and low demand densities. The cut-off thresholds used for SAX quantisation are calculated as shown in Table 1. The letters \( L \) (for extreme low demand) and \( l \) (for low demand) are assigned to a 50% and 40% portion of the expected low demand density, respectively. Similarly, the letters \( H \) and \( h \) are taken from the expected-high demand density, representing extreme high demand and high demand, respectively.

Table 1. The cut-off thresholds used for SAX quantisation.

<table>
<thead>
<tr>
<th>Letter</th>
<th>Share in the expected-low demand</th>
<th>Share in the expected-high demand</th>
</tr>
</thead>
<tbody>
<tr>
<td>( L )</td>
<td>50%</td>
<td>0</td>
</tr>
<tr>
<td>( l )</td>
<td>40%</td>
<td>0</td>
</tr>
<tr>
<td>( m )</td>
<td>10%</td>
<td>10%</td>
</tr>
<tr>
<td>( h )</td>
<td>0</td>
<td>40%</td>
</tr>
<tr>
<td>( H )</td>
<td>0</td>
<td>50%</td>
</tr>
</tbody>
</table>

Of course, many letters have been created, but the most frequent letters are equal. However, we use a more meaningful approach to select the cut-off thresholds. First we divide the demand set into two groups: The expected-high demands, comprised of the averaged demand values for the day time during weekdays (five values per week), and the expected-low demands, comprised of the averaged demand values during weekends and night time during weekdays (nine values per week). Furthermore, public holidays are treated the same as weekends. These two demand sets are not disjoint but are usually separated well enough so that they can be used to define the cut-off thresholds.

The method used for the cutting in most studies (such as [14]) is an equiprobable cut, implying that the probability of occurrence for letters are equal. However, we use a more meaningful approach to select the cut-off thresholds. First we divide the demand set into two groups: The expected-high demands, comprised of the averaged demand values for the day time during weekdays (five values per week), and the expected-low demands, comprised of the averaged demand values during weekends and night time during weekdays (nine values per week). Furthermore, public holidays are treated the same as weekends. These two demand sets are not disjoint but are usually separated well enough so that they can be used to define the cut-off thresholds.

Figure 4 shows the normalized probability density of the average electricity demand for building B2 with the expected high and low demand densities. The cut-off thresholds used for SAX quantisation are calculated as shown in Table 1. The letters \( L \) (for extreme low demand) and \( l \) (for low demand) are assigned to a 50% and 40% portion of the expected low demand density, respectively. Similarly, the letters \( H \) and \( h \) are taken from the expected-high demand density, representing extreme high demand and high demand, respectively.

### C. Motif Selection

Once the demand values are mapped into letters, each weekly demand profile is transformed to a 14-letter word. The next step in the statistical anomaly detection method is the selection of a motif in the SAX transformed demand profiles, referred to as a normal mot. This is done by counting the frequency of a weekly word appearing within the total observed period, in this case being February 2011 to June 2016.

Figure 5 depicts the word frequency for the electricity demand of building B2 (for this visualization, both \( h \) and \( H \) are shown as \( H \), and both \( l \) and \( L \) are shown as \( L \)). There are 78 distinguished words and, as expected, the word \( H L H L H L H L H L L L L \) has the highest frequency of 122. This word corresponds to a high demand (\( H \)) during the weekday mornings and a low demand (\( L \)) during weekday evenings as well as the whole weekends. The second most frequent word is \( L L H L H L H L L L L L \) and the third one is \( M L H L H L H L L L L L \) with the only difference being the first letter. These two words correspond to weeks when a holiday has fallen on a Monday; hence a lower demand is expected during the working hours. Finally, the fourth most frequent word of \( H L H L H L H L L L L L \) represents weeks when Friday is a holiday and has a lowered demand as a result. These four most frequent words (as highlighted in Figure 5) comprise almost 60% of the weeks. Regardless, only the most frequent word, being \( H L H L H L H L L L L \), is selected as the normal word or the "motif".

Note that the word frequency changes if different cut-off thresholds to those shown in Table 1 are selected. Therefore, in order to assure a robust selection of the normal word, one may perform a sensitivity analysis based on the cut-off values.
quantifying the divergence. The actual weekly demands are compared to the normal demand \( D \). This procedure is out of the scope of this article and will be discussed elsewhere.

Following the selection of a normal profile, the corresponding normal electricity demand needs to be calculated. In order to decide whether the normal demand is defined over the whole dataset or rather individually for each year, a data screening is performed. Figure 6 shows a compilation of demands on normal word weeks for the six studied years. It is observed that a downward trend in electricity demand exists in building B2, suggesting progressive improvements in the operation and efficiency of electrical devices, such as fans, lighting and plug loads. Similar trends can be observed for the electricity demand in the other two buildings. Therefore, an annual normal demand was calculated for each studied year, from the normal word weeks in each year.

### D. Anomaly Detection

Once the annual normal demands are calculated, the actual weekly demands are compared to the normal demand of the corresponding building and anomalies are detected by quantifying the divergence.

The energy performance of a building within a week is evaluated using the three labels of Good, OK, and Bad. Obviously, an energy demand which is lower than expected is translated to a higher grade in performance. This implies that, if a letter \( H \) is assigned to a period in the normal profile, there is no need to evaluate that period in the actual profile as a poor performance anomaly, because any existent energy demand is better than or at least equal to a high (i.e. \( H \)) demand. Given this knowledge, the three labels are defined as follows:

**Good:** A Good label is assigned to a week that has a SAX word matching the pattern \( _{-}L_{-}L_{-}L_{-}L_{-}L_{-}L_{-}L_{-}L_{-}L_{-}L_{-}L_{-}L_{-}L_{-}L \) where the underscores can be replaced by any of the five letters (L, l, m, h, H). The Good pattern corresponds to a week when all expected-low periods do indeed have a low demand. Once the Good label is assigned to a week, no weekly saving \( (S_w) \) is calculated. Note that at this stage the algorithm is not case sensitive, therefore both \( h \) and \( H \) are treated as \( H \), and both \( l \) and \( L \) are treated as \( L \). In application, we do not wish to overly distract operators, and therefore choose not to alert operators to relatively small deviations from the ideal normal demand profile.

**OK and Bad:** If the SAX word of the week does not follow the Good pattern of \( _{-}L_{-}L_{-}L_{-}L_{-}L_{-}L \), the weekly demand is compared to the normal demand in order to calculate the potential weekly energy saving using the formula

\[
S_w = \frac{\sum (P_{\text{elec}} - P_{\text{elec}}^+)_{\text{week}}}{{\sum L_i} P_{\text{elec}}^+_{\text{week}}} \times 100, \quad i \in \{2,4,6,8,10,11,12,13,14\}
\]

where, \( S \) indicates the percentage of potential saving, \( P \) is the normal demand and \( p \) is the actual weekly demand. Index \( i \) enumerates the 12-hour sections, and in this case it iterates only for the letters that correspond to a low demand periods in the Good pattern. Index \( w \) indicates the weeks. Because of the way values of \( i \) are selected, the numerator of the fraction is always smaller that the denominator; hence \( S_w \) is properly defined as a percentage between 0 and 100. Note that the algorithm treats the holidays and weekend days the same.

The labeling process is based on the value of \( S_w \). If it exceeds 5%, the weekly performance is directly labeled as Bad. Otherwise, it is labeled as OK. The 5% threshold is a tunable parameter and should be adjusted according to the type of building and the deviation level tolerated by building operators. In the next section, a case study is carried out for the three office buildings in order to demonstrate the potential of the proposed labeling method.

### III. Case Study and Discussions

#### A. Buildings

The case study is based on the archived energy consumption data collected from three federal office buildings located in the city of Ottawa, Canada. Table 2 shows an overview of the buildings’ information.

The selected buildings demonstrate diversity in terms of floor area and the electricity demand which enables a more reliable evaluation of the method.
Table 2. Parameters of the studied buildings.

<table>
<thead>
<tr>
<th>Building</th>
<th>Number of floors</th>
<th>Floor area (m²)</th>
<th>Construction year</th>
<th>Average electricity demand (MWh/yr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1</td>
<td>1</td>
<td>2,000</td>
<td>1956</td>
<td>195</td>
</tr>
<tr>
<td>B2</td>
<td>13</td>
<td>61,000</td>
<td>1979</td>
<td>7,500</td>
</tr>
<tr>
<td>B3</td>
<td>2</td>
<td>4,400</td>
<td>1955</td>
<td>360</td>
</tr>
</tbody>
</table>

However, the diversity should not prevent a meaningful comparison of the results among the buildings, which reveals the importance of the normalization stage.

Figure 7 shows the normal demand profiles of the three studied buildings for the year 2015. As expected, the profiles are not alike in shape and magnitude, as they are bound to the characteristics of the buildings and the behaviour of the occupants. Therefore, z-score normalized versions of the profiles are also shown. The normal demands follow a very similar weekly pattern, suggesting similarity in working hours and the behavioural habits of the occupants.

Once the labeled weekly demand profiles are generated for each building and each studied year, the results can be investigated from two different angles: with a focus on label statistics (useful for fault detection and diagnosis (FDD)) and on demand reduction (useful for energy auditing purposes). Both aspects are explored in the following sections.

B. Labeling

Table 3 shows a summary of generated labels for the three buildings and each year of operation based on the SAX method. From an FDD perspective, the main concerns are the anomalies that are non-temporary and severe, hence a higher probability to identify at a systematic fault in the system. This implies that a persistent high number of OK or Bad weeks in any year would raise a red flag.

By reviewing Table 3c, one observes that the first criterion is satisfied by building B3, as a consistently large number of OK weeks can be found, especially in years 2012 to 2015 where a more complete dataset is available.

Using such information, a technician may be able to efficiently allocate their time towards fixing faulty equipment or aberrant operational parameters.

For the second criterion, that is the existence of less frequent but more severe anomalies, building B1 demonstrates an example. In years 2014 and 2015 about 20% of the weeks are labeled as Bad, pointing to the possibility of a severe fault, which might be behavioural, mechanical, or structural nature. Similarly, the Bad labels of building B2 for the year 2012 add up to 16%, pointing at a situation that requires further investigation by building operators. One can proceed with a seasonal, daily, or hourly examination of the labels to identify the potential sources of the faults.

C. Potential energy saving

The weekly labels can be used to estimate annual energy savings potential by comparing the actual scenario to one where all weeks are labeled as Good. This is valorized using the annual potential energy saving, calculated as

\[
S_{\text{annual}} = \frac{\sum_{w} \sum_{i} (p_{\text{act}}^{w,i} - p_{\text{low}}^{w,i})}{\sum_{w} \sum_{i} p_{\text{low}}^{w,i}} \times 100, \quad w \in \{1, ..., W\}
\]

where, \(w\) iterates over the weeks with complete data availability with a maximum of \(W\) and \(i\) is defined in the same way as Equation (2). Note that the \(S_{\text{annual}}\) is also equivalent to the average of \(S_{\text{weekly}}\) values.

Table 3. Results of anomaly detection for the three studied buildings. Only weeks with enough number of available data points are considered.

(a) Labeled weeks for building B1

<table>
<thead>
<tr>
<th>Year</th>
<th>Good</th>
<th>OK</th>
<th>Bad</th>
<th>Weeks</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>32</td>
<td>12</td>
<td>1</td>
<td>45</td>
</tr>
<tr>
<td>2012</td>
<td>36</td>
<td>12</td>
<td>3</td>
<td>51</td>
</tr>
<tr>
<td>2013</td>
<td>26</td>
<td>16</td>
<td>3</td>
<td>45</td>
</tr>
<tr>
<td>2014</td>
<td>26</td>
<td>14</td>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>2015</td>
<td>29</td>
<td>12</td>
<td>9</td>
<td>50</td>
</tr>
<tr>
<td>2016</td>
<td>17</td>
<td>7</td>
<td>0</td>
<td>24</td>
</tr>
<tr>
<td>Total</td>
<td>166</td>
<td>73</td>
<td>26</td>
<td>265</td>
</tr>
</tbody>
</table>

(b) Labeled weeks for building B2

<table>
<thead>
<tr>
<th>Year</th>
<th>Good</th>
<th>OK</th>
<th>Bad</th>
<th>Weeks</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>35</td>
<td>6</td>
<td>4</td>
<td>45</td>
</tr>
<tr>
<td>2012</td>
<td>35</td>
<td>8</td>
<td>4</td>
<td>51</td>
</tr>
<tr>
<td>2013</td>
<td>31</td>
<td>10</td>
<td>4</td>
<td>45</td>
</tr>
<tr>
<td>2014</td>
<td>38</td>
<td>7</td>
<td>5</td>
<td>50</td>
</tr>
<tr>
<td>2015</td>
<td>22</td>
<td>27</td>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td>2016</td>
<td>15</td>
<td>7</td>
<td>2</td>
<td>24</td>
</tr>
<tr>
<td>Total</td>
<td>176</td>
<td>65</td>
<td>24</td>
<td>265</td>
</tr>
</tbody>
</table>

(c) Labeled weeks for building B3

<table>
<thead>
<tr>
<th>Year</th>
<th>Good</th>
<th>OK</th>
<th>Bad</th>
<th>Weeks</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>21</td>
<td>24</td>
<td>0</td>
<td>45</td>
</tr>
<tr>
<td>2012</td>
<td>16</td>
<td>33</td>
<td>2</td>
<td>51</td>
</tr>
<tr>
<td>2013</td>
<td>28</td>
<td>18</td>
<td>4</td>
<td>50</td>
</tr>
<tr>
<td>2014</td>
<td>27</td>
<td>23</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>2015</td>
<td>30</td>
<td>20</td>
<td>0</td>
<td>50</td>
</tr>
<tr>
<td>2016</td>
<td>17</td>
<td>6</td>
<td>1</td>
<td>24</td>
</tr>
<tr>
<td>Total</td>
<td>139</td>
<td>124</td>
<td>7</td>
<td>270</td>
</tr>
</tbody>
</table>

Figure 7. Normal weeks in 2015 for the three buildings. The last plot shows a combined overview of normalized profiles which reveals the similarities.
Table 4 lists the annual saving for the Bad and OK weeks as well as the average per-area demand which is a normalized measure for the energy intensity of buildings. The first observation is that the per-area demand values are in the same range for the three buildings, although the absolute electricity demand is much higher for B2 as shown in Table 2.

The second and the more interesting observation is that the saving potential for the OK weeks and the more critical Bad weeks is not proportional to the magnitude of the average per-area demand. For example, by comparing the average demand of Bad weeks between buildings B1 and B3, and energy auditor may conclude that they suffer from the same magnitude of severity in abnormal behaviour, as they demonstrate an increase of 16% and 12% in average weekly demand, respectively. However, the saving potential calculated using the anomaly detection method (Table 3) reveals that in fact, Bad weeks in building B1 are much frequent and require a more urgent investigation. Similar inappropriate wrong conclusions may be made for the OK weeks of buildings B2 and B3.

Table 4. Annual saving and demand per area for the three buildings.

<table>
<thead>
<tr>
<th>Building</th>
<th>$S_{\text{annual}}$</th>
<th>Average per-area demand (kWh/m²)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OK weeks</td>
<td>Bad weeks</td>
</tr>
<tr>
<td>B1</td>
<td>2.7%</td>
<td>12.0%</td>
</tr>
<tr>
<td>B2</td>
<td>2.5%</td>
<td>8.4%</td>
</tr>
<tr>
<td>B3</td>
<td>1.8%</td>
<td>7.3%</td>
</tr>
</tbody>
</table>

On the contrary, a comparison of average per-area demand for OK weeks in buildings B1 and B2 may bring the investigator to the conclusion that there is a much bigger saving potential in building B1, while the value of $S_{\text{annual}}$ for OK weeks indicates that it might not be the case.

IV. CONCLUSIONS

The data-driven anomaly detection methods are superior to the model-based methods in the sense that they do not rely on a detailed knowledge of the building structure and parameters in advance. In this work, a statistical data-driven approach using the SAX method was proposed to analyse years of collected data from three federal buildings with diversity in size and electricity demand profiles. In summary, the results of the proposed anomaly detection method reveals that the implemented SAX-based labeling approach can serve as a multi-purpose energy monitoring tool for both operators and stake holders, by transforming the large sets of time series data into the more comprehensible weekly words. As the method relies on normalization and statistical interpretation, a comparison of resulting saving potential among the buildings is facilitated and the conclusions are more meaningful since they are not bound to the specific building properties.

Regarding the potential applications for the proposed approach, we expect that it can be widely applied to large centralized or distributed campuses where a limited number of facility operators are in charge of a large number of buildings (such as universities/colleges, military bases, schools, hospitals, bank branches, postal outlets, etc.). The method will then aid the operators in efficiently allocating their limited resources. For the sake of anomaly detection in electricity demand, the main requirement for the deployment of the method is a reliable demand measurement at a reasonable resolution (finer than or equal to 1 hour). Such data collection is already available at many of the aforementioned locations.

For the future extension of this work, the authors intend to further develop the method and deploy it for anomaly detection in heating and cooling demand profiles. Furthermore, a deployment of the developed algorithm for “real time” anomaly detection in non-residential buildings is planned.

REFERENCES