Model-based and data-driven anomaly detection for heating and cooling demands in office buildings

Ashouri, Araz; Hu, Yitian; Gunay, H. Burak; Newsham, Guy R.; Shen, Weiming

Publisher’s version / Version de l'éditeur:

ASHRAE Winter Conference 2019, 2019

Model-based and Data-driven Anomaly Detection for Heating and Cooling Demands in Office Buildings

Araz Ashouri, PhD
Associate Member ASHRAE

Yitian (Vera) Hu

H. Burak Gunay, PhD
Associate Member ASHRAE

Guy R. Newsham, PhD

Weiming Shen, PhD

ABSTRACT

A considerable portion of total energy loss within the built environment originates from operational errors during the actual lifespan of a building. With the rise of fully automated commercial buildings, a large amount of sensory data is becoming available that can be leveraged to detect and predict such errors. However, processing these data on-site requires significant knowledge and effort by building operators. In this work, a combination of model-based and data-driven approaches are employed to facilitate the analysis of historical energy demand data. Using change-point models and symbolic quantisation techniques, a large dataset of heating and cooling demand profiles collected from several office buildings are transformed into a format that is easily interpreted by the building operator and is suitable for actionable anomaly detection. Further quantification of anomalies and calculation of potential savings are drawn from the results.

INTRODUCTION

Given the buildings' considerable contribution to total energy use among different sectors, any effort in improving their operation and energy efficiency can scale to a large impact in sustainability and greenhouse gas emissions (Omer 2008). Studies show that many instances of energy inefficiencies in buildings are introduced by operational errors and faulty devices, although identification of such malfunctions appears to be a big challenge in modern office buildings with thousands of sensors and actuators (Gunay and Shen 2017). As a result, automated fault detection tools - that can replace in-person walk-arounds and reduce costs associated with human resources - may serve as a promising solution to this problem.

One of the challenges in such tools is associated with anomaly detection: a process in which energy data is scanned to find abnormal behaviour that can potentially lead to a fault. As revealed by a recent review (Miller, Nagy, and Schlueter 2018), many researchers are targeting the topic of anomaly detection in buildings' energy demand, using various model-based and data-driven methods. Data-driven methods, which require less in-depth knowledge of the thermophysical properties of building systems, proved to be efficient for anomaly detection based on the analysis of electricity demands (Seem 2007). However, the approach is not as effective for heating and cooling demands, mainly due to their additional correlation to weather parameters; hence requiring an underlying model to capture the behaviour explained by thermal parameters. To this end, model-based anomaly detection method create an inverse model of a subsystem in the building (for example a thermal zone) to learn its normal operational behaviour, such as heating and cooling demands (Gunay, Shen, and Yang 2017). A common type of such models is a change-point model that is employed in several studies to establish a relationship between average outdoor temperature and energy

Araz Ashouri, **Yitian (Vera) Hu, Guy R Newsham**, and **Weiming Shen** are researchers at National Research Council Canada, Ottawa, Ontario, Canada. **H. Burak Gunay** is an assistant professor in Department of Civil and Environmental Engineering at Carleton University, Ottawa, Ontario, Canada.

demand of the building (Kissock, Haberl, and Claridge 2003).

In previous work (Ashouri et al. 2018), the present authors applied a data-driven method to historical electricity demand of several buildings to detect abnormal patterns, potentially due to a fault in the building energy system. In this study, a similar methodology is employed for heating and cooling demand profiles in commercial buildings; these demands are correlated to the number of occupants in the building, day-of- week, and time-of-day as demonstrated by standards such as ASHRAE (ASHRAE 2016) and the National Energy Code of Canada for Buildings (NRC Canada 2011). In order to capture the weather dependencies, change-point models are developed for each building and each type of service. In the rest of the paper, first a brief introduction of the buildings and their hourly heating and cooling demand is provided. The employed method is explained in several sections while the input and output of each step is illustrated using data from the case study buildings. In the last step, potential energy savings identified by the proposed anomaly detection method are calculated and discussed. The paper concludes with discussion of the results and recommendations for future work.

Studied Buildings

This study is conducted using archived hourly heating and cooling demand data collected from two federal office buildings located in Ottawa, Canada. Table 1 presents a summary of the buildings' information. Hourly steam and chilled water demand (supplied by a central heating and cooling plant) of these buildings were monitored for three consecutive cooling and heating seasons as shown in Figure 1.

Table 1. Summary of the characteristics of the studied buildings.

Building identifier	Number of floors	Floor area m^2 (ft^2)	Construction year	Heating demand intensity kWh/m^2-yr ($kBTU/ft^2$-yr)	Cooling demand intensity kWh/m^2-yr ($kBTU/ft^2$-yr)
B1	4	39,000 (420,000)	1952	121 (38)	67 (21)
B2	13	61,000 (657,000)	1979	33 (10)	56 (18)

The dataset includes data between April 23, 2013, and April 24, 2016, although some data may be missing during short periods. At a first glance, abnormal behavior is observed for building B1 in terms of simultaneous heating and cooling: heating during summer, and cooling during winter. However, these behaviours may be justifiable for a variety of reasons, such as multi-zone conditioning, that is when some thermal zones need heating while others require cooling. In addition, some buildings incorporate server rooms or laboratories that have to be maintained at a fixed temperature setpoint throughout the year, resulting in continuous heating/cooling demand.

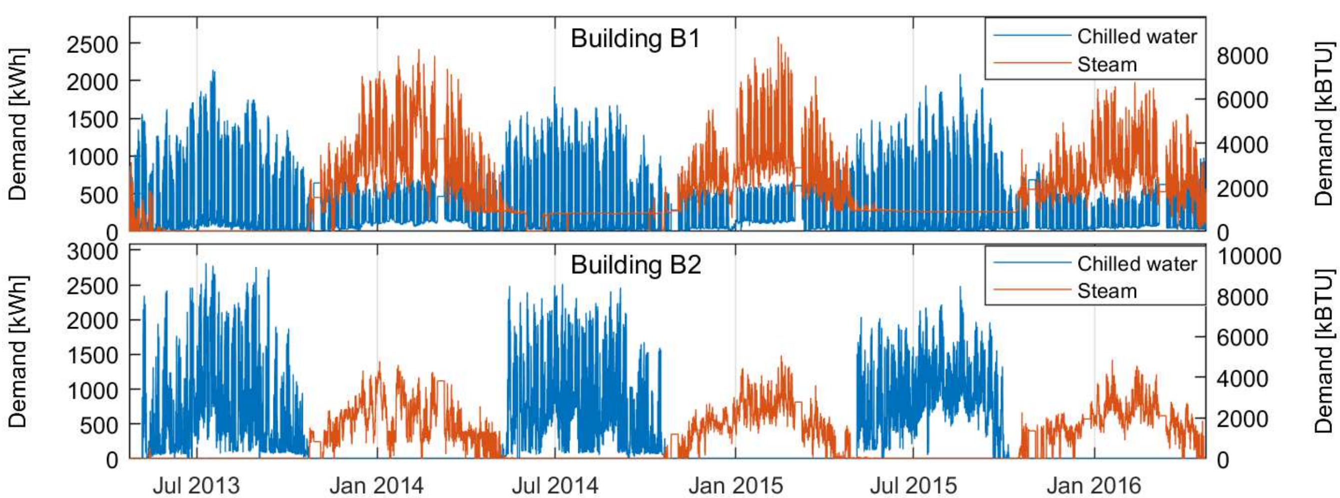

Figure 1 Hourly data for chilled water (cooling) demand and steam (heating) demand for the studied buildings.

Therefore, the anomaly detection proposed in this study first identifies a typical demand profile for each building to capture its own unique characteristics, then the actual demand is compared to the typical profile and the deviations from this typical profile are quantified.

PROPOSED METHOD

The anomaly detection method used in this study consists of several modules that are shown in Figure 2. The input to the tool is the raw data sourcing from the installed sensors; i.e. energy meters measuring heating and cooling, via a database. The data is treated in two stages responsible for model-based and data-driven processing. In the model-based stage, a change-point model is trained to remove the trend introduced by weather parameters. In the data-driven stage, the demand data is quantised into symbols in order to identify a motif and a corresponding typical demand profile. Finally, the actual weekly demand profiles are compared with the typical profile to detect the anomalies and resulting deviations are used in calculation of potential savings. In the following sections, each module is described in more detail.

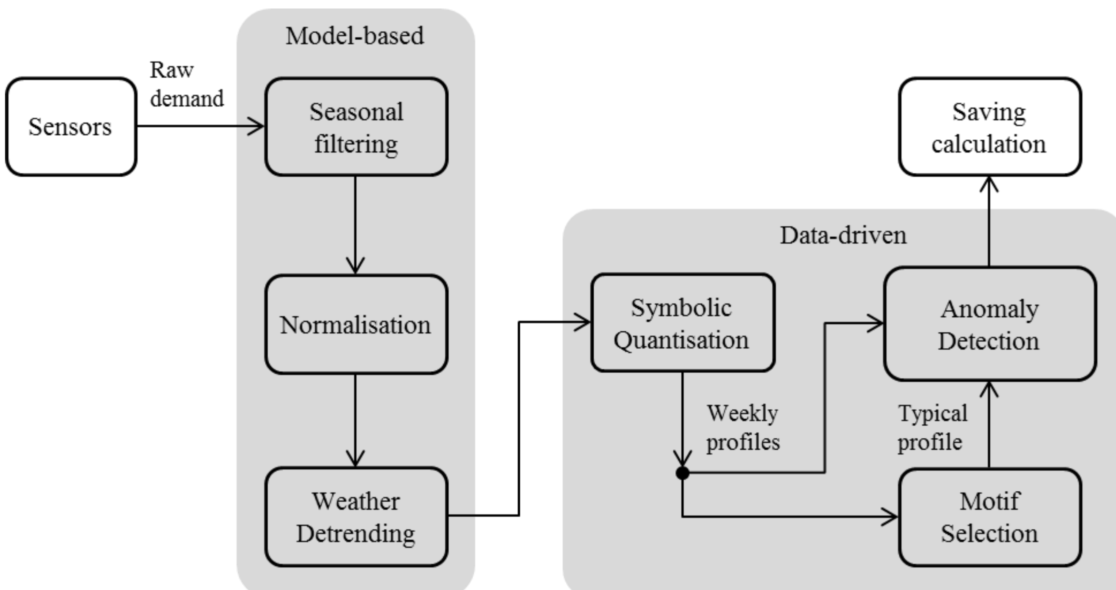

Figure 2 Block diagram showing the main modules and their connections.

Seasonal Filtering

The proposed method is designed for application to large dataset of demand data, in this case 25320 data-points (1055 days × 24 hours = 25320) per studied building. In order to increase the accuracy of evaluation and improve computational efficiency, dimension-reduction techniques are used to extract the most representative and meaningful features from the data. A first step towards this end is to divide data based on temporal characteristics. This is performed based on categorizing the dataset into heating/cooling seasons, day/night sections, and high/low demand periods.

Heating and cooling seasons. Air conditioning in commercial and office buildings typically takes place based on pre-defined schedules defined by building operators. Based on the historical weather data and occasionally depending on short-term forecasts, heating and cooling systems are turned ON and OFF. However, as seen in Figure 1, simultaneous heating and cooling can happen in non-residential buildings, as the two systems are operated independently. In order to detected anomalous behaviours, one may identify and distinguish the heating season from the cooling season within the same year, based on local weather data. In addition, for a building cluster located in the same campus, fed by the same central heating/cooling plant, the seasons can be detected from data collected at one of the buildings and applied to the rest of the cluster. This is indeed the situation for the buildings studied in this paper. Hence, as building B2 shows a clear separation of heating and cooling seasons (refer to Figure 1), it is used to define the seasons for both buildings. Once the seasons are defined, off-season air conditioning (i.e. heating in summer or cooling in winter) can be flagged as anomalous behavior. Such anomalies are observed in demand profiles of building B1 (for cooling during winter).

Day and night sections. Another major factor in scheduling HVAC demand is hour of the day. In office buildings, specifically, a routine occupancy pattern is followed by the building users, mainly due to regulated working hours. However, the HVAC schedules may vary dramatically from one building to another. In the case of the office buildings studied in this project, the morning ramp-up of HVAC system varied between 4am and 6am, and the ramp-down after working hours took place between 5pm and 7pm. Because of such statistical observations, each day is divided into two sections, namely Day and Night, with the day section being between 6:00am to 5:59pm and the night section being between 6:00pm and 5:59am. For the purpose of anomaly detection, such separation helps compare the demand among the 7 day sections and 7 night sections within a week, to see whether sequential patterns exist and to identify anomalous behaviour such as a lack of night-time setback in HVAC operation. The day/night sections also contribute to high and low demand periods, as explained below.

High- and low-demand periods. The final dividing of demand data is done based on day-of-week and type of the day. The idea is to identify periods that are expected to show high or low demands, which is mostly driven by building occupancy in non-residential buildings. As a result, weekends and night sections of weekdays (i.e. Monday to Friday) are considered as expected-low (demand) periods, while day sections during weekdays are categorised as expected-high (demand) periods. In addition, national and regional holidays are treated the same as weekends. This separation is a crucial part of the method, as it affects both the weather detrending and statistical quantisation, as will be explain in the next sections.

Normalisation and Aggregation

Characteristics of heating and cooling demands such as mean, minimum, and maximum values fluctuate by various temporal factors such as time of day, day of week, and season of year. As a result, a holistic anomaly detection method that takes a long historical data period as an input must work on normalised data. Various normalisation schemes are employed by data analysts and statisticians, the most popular ones being the standard score and zero-to-one normalizations. For this study, the incorporated approach is the zero-to-one normalisation that is more appropriate for visualization since negative demand values cannot exist. The normalised demand is calculated as follows:

$$p_{\text{normal}}[\text{t}] = \frac{p[t]-p_{\min}}{p_{\max}-p_{\min}} \tag{1}$$

where p is original demand, t is time, and p_{min} and p_{max} represent the minimum and maximum values of original demand, respectively. Following the normalization, the demand data are aggregated within each day/night section, achieved by replacing the demand at each section with its averaged value; as shown below:

$$p_{\text{agg}}\,[\text{t}] = \sum_{t=12\cdot k-11}^{12\cdot k}(p_{\text{normal}}\,[\text{t}]) \tag{2}$$

As a result, a week of demand data (comprising of 168 hours) is compressed into 14 averaged day/night sectional demands, drastically reducing the dimension of data.

Weather Detrending

Heating and cooling demands in office buildings can be described by two major components. The first is a sequential pattern as a function of time-of- day and day-of-week and is mainly affected by occupants' behaviour.

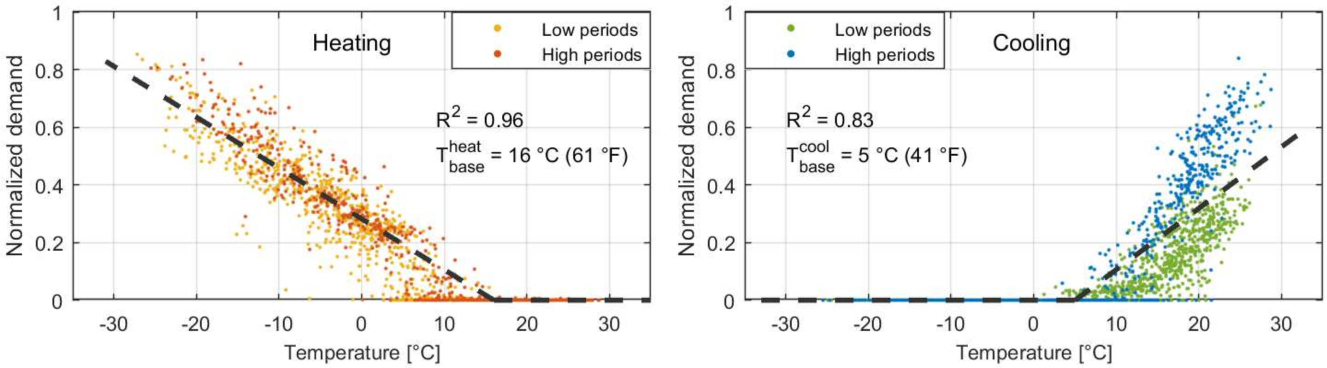

Figure 3 Normalized heating and cooling demand versus outdoor temperature and the corresponding change-point models, for building B2. The R-squared values and base temperatures are shown for each linear fit.

The second is a non-sequential trend that is a correlated to weather conditions, especially to outdoor air temperature. Therefore, prior to processing the demands with the symbolic quantisation method, the part of demand explained by the outdoor temperature has to be removed. This is done by a process known as Weather Detrending. The process involves building change-point models between the heating/cooling demand and the outdoor air temperature. The models are created for each studied building during the heating/cooling season and high/low demand periods.

Figure 3 shows the aggregated hourly heating and cooling demands for building B2 as a function of average outdoor temperature. It can be seen that for the data-points where demand is non-zero, a linear correlation between the demand and the temperature can be drawn. In case of cooling demand, it is even possible to fit separate lines for low and high periods. However, as we intend to explain the behaviour of cooling system with a single model, a single line is used instead.

This correlation, known as a trend, is presented with a change-point model by fitting a line to the data using a minimization of least square errors. The resulting model is described as

$$\hat{p}_{\text{agg}} = \begin{cases} \alpha \cdot (T_{\text{base}} - \bar{T}_{\text{out}}) & \bar{T}_{\text{out}} \leq T_{\text{base}} \\ \beta \cdot (\bar{T}_{\text{out}} - T_{\text{base}}) & \bar{T}_{\text{out}} > T_{\text{base}} \end{cases} \tag{3}$$

where $\hat{p}_{\text{agg}}$ is the expected estimation of aggregated power, $\bar{T}_{\text{out}}$ is the average outdoor temperature within a day/night section, T_{base} is the base temperature at which point the heating or cooling is triggered, and (α, β) are the heating/cooling rate represented as the slope of the line in the linear fit. It is trivial that $\alpha = 0$ for cooling and $\beta = 0$ for heating. In the case of building B2, the change-point models describe the trend with a confidence of $R^2 = 0.96$ for heating demand and $R^2 = 0.83$ for cooling demand. In addition, the base temperature is $16\,^{\circ}C$ $(61\,^{\circ}F)$ for heating and $5\,^{\circ}C$ $(41\,^{\circ}F)$ for cooling. The same approach of change-point modelling is used to identify the trend in buildings B1 and B2. Once the trends are found, detrending is achieved by extracting the power estimation of the change-point model from the actual power:

$$p_{\text{detrend}} = p_{\text{agg}} - \hat{p}_{\text{agg}} \tag{4}$$

An example of how such weather detrending can affect the demand is shown in Figure 4. During this week of June, the aggregate cooling demand demonstrates an upward trend that rises from Monday to Friday, which, in raw form, may be viewed by the building operator as an anomaly. However, once detrended, it shows a consistent normalized demand for all weekdays, revealing that the increasing demand is in fact due to a rise in outdoor temperature instead of a fault in the building. Therefore, one concludes that the weather-detrended demand is a better representation of the seasonal (i.e. periodic) behaviour of the building that is mostly correlated to building occupancy.

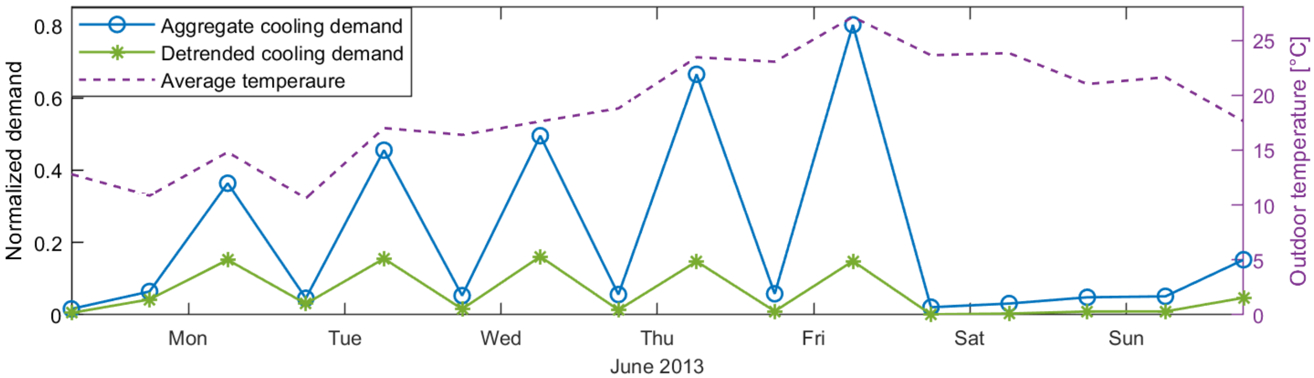

Figure 4 The effect of weather detrending on a week of aggregate cooling demand.

Symbolic Quantisation

Despite all the processing, the energy demands at this stage of the algorithm contain continuous values, which is not optimal for the pattern recognition and anomaly detection tasks. Therefore, value quantisation and symbolic conversion are applied to the data. The idea is to transform the detrended weekly demand (that is a time series with 14 values) into “words” comprised of 14 symbols or “letters”. The method, called symbolic quantisation, has been commonly deployed in research studies for anomaly detection (Miller, Nagy, and Schlueter 2015).

In order to select the thresholds at which the demand is quantised, statistical methods are used. More details on the statistical selection of thresholds and the corresponding quantised symbols can be found in (Ashouri et al. 2018). As shown on the left side of Figure 5, for the studied buildings, the six letters used for quantisation are “*L*” for very low, “*l*” for low, “*m*” for medium, “*h*” for high, “*H*” for very high, and “*X*” for extremely high demand. As a result, a 14-letter word comprised of these six letters represents a weekly demand profile.

Motif Selection

In order to detect anomalous behavior in building energy demand, all weekly demand profiles have to be compared with a “typical profile” which corresponds to the most frequent weekly word, called a motif. The motif is found using a motif-selection task that identifies the most frequent weekly word. A typical profile is unique to the building and the type of service. Figure 5 also shows a sorted histogram (i.e. frequency) of weekly words for the cooling demand of building B1 (for this visualization, *h* and *H* are shown as *H*, and *l* and *L* are shown as *L*).

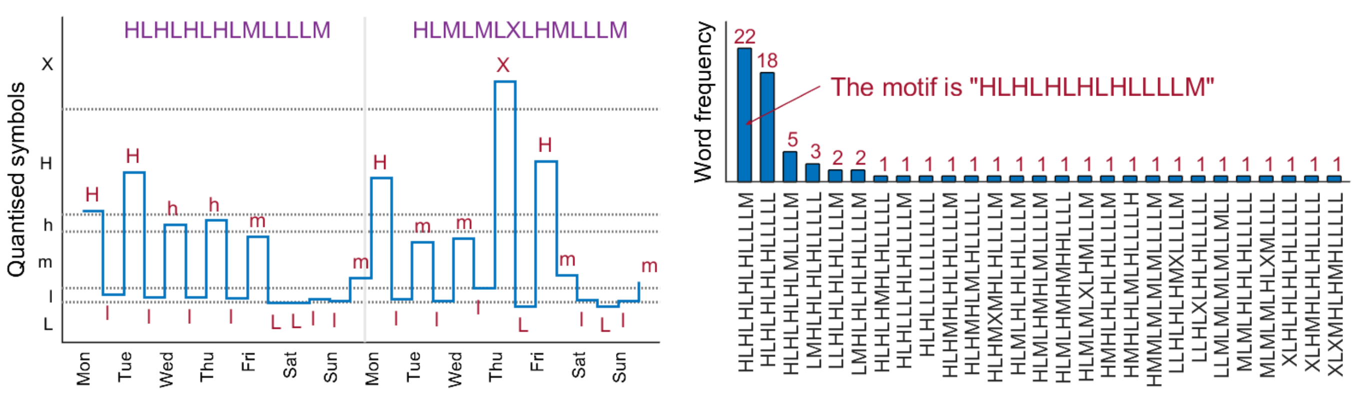

Figure 5 Symbolic quantisation for two weeks of detrended demand (left) and the word frequency for the cooling demand of building B1 (right).

There are 27 distinguished words and the word *HLHLHLHLHLLLLM* has the highest frequency of 22, which is selected as the motif. This word corresponds to a high demand (*H*) during the weekday day sections and a low demand (*L*) during weekday night sections as well as the whole weekend, except for the night section on Sundays (indicative of cooling beginning early in preparation for Monday morning occupancy). The second most frequent word is *HLHLHLHLHLLLLL* with the only difference being the last letter. It is important to mention that the latter word does not get flagged as abnormal, since it represents a lower demand ($L<M$). Therefore, it indicates that the anomaly detection is robust with respect to slight changes in quantisation thresholds that may change the ranking of the two words. In addition, the two words represent almost 55% of the weeks. Once the motif is selected for each building during a calendar year, the corresponding typical profile is calculated by averaging all the weekly demands whose weekly words are the same as the motif.

Anomaly Detection

In the final step of the algorithm, the actual weekly demands of each building are compared to the corresponding typical demand and anomalies are quantified by calculating the divergence.

This is done by evaluating the energy performance of a building using the total energy-cost saving, calculated as:

$$S_{\text{total}} = \frac{\sum_w \sum_i (p_{i,w} - \boldsymbol{P}_{i,w})}{\sum_w \sum_i p_{i,w}}, \quad w \in \{1, \dots, \mathrm{W}\}, \quad i \in \{1, \dots, 14\} \cap \{p_i > \boldsymbol{P}_i\} \tag{5}$$

where S is the energy saving, P is the typical demand, and p is the actual weekly demand. Index w iterates over the weeks with complete data availability. Index i enumerates the 12-hour daily sections, and iterates only over the letters in the actual demand that are higher than the typical demand. For example, if the motif is *HLHLHLHLHLLLLM* and the actual weekly word is *HMHMHLHLMLLLLL*, then index i takes the values $\{2,4\}$ because only the second and the forth letters are different to, and higher than, the corresponding letter in the motif (since $M>L$).

RESULTS AND DISCUSSIONS

The results of anomaly detection for the two studied buildings are shown in Table 2. For the energy cost savings, unit cost of heating and cooling in Canadian dollars is extracted from actual billing reports, amounting to 113 C$/MWh ($33$ C$/MBTU) for heating and 110 C$/MWh ($32$ C$/MBTU) for cooling. The different identified motifs and corresponding number of abnormal weeks for each building and service type are translated to potential savings in the range of 1.3% to 8.8%. Further analysis of Table 2 reveals several noteworthy points.

First, from the four identified motifs, three can be considered as regular and one as irregular. A regular motif contains a certain amount of setback when shifting from day to night and from weekdays to weekends and holidays. In other words, it is expected that low- and high-demands are well separated. For building B1 and for cooling demand of building B2, the motifs have regular formats: H (High) demand for day sections during weekdays and *L* (low) or *M* (medium) during the night sections and weekends. Looking at the potential savings, these three categories show a higher promise compared to the last category (i.e. heating of building B2). This is logical in the sense that the motif *HHHHHHHHHHHHHH* does not correspond to a consumption pattern triggered by time-of--day and day-of-week, hence it is not representing the behaviour of building occupants or any other sequential factor. In such cases, the presented anomaly detection tool provides limited information for the operators regarding fault detection.

Second, the results indicate that the initial assumption about the severity of anomaly in demand of building B1 versus B2 may not have been accurate. If the overall savings (for both heating and cooling) are calculated for the two buildings, one finds out that building B1 has an overall potential cost saving of 5.4% while it is 7.2% for B2. This indeed points to the importance of data analytic tools, such as the one presented in this paper, as their ability in processing large data is likely superior to the visual inspection of time-pressed building operators. In fact, finding a typical pattern in large dataset by solely monitoring a visualisation of energy demand is cumbersome, if not impossible, for a human operator.

Finally, the importance of quantifying the divergence is highlighted by comparing the potential savings between cooling and heating demands in building B1. Although the typical profiles have very similar motifs and the number of abnormal weeks is also similar, the heating demand shows a much bigger potential in terms of percentage (6.6% versus 2.5%). In fact, monetising the divergence from the typical profile is critical in understanding the severity of the anomalies, which can help building operators or maintenance personnel prioritize their work.

Table 2. Results of anomaly detection and potential savings in Canadian dollars

Building	Service type	Motif	Number of evaluated weeks (W)	Number of abnormal weeks	Energy cost for abnormal weeks C$	Potential total saving C$
B1	Heating	*HLHLHLHLHLLLLL*	85	65 (76%)	831,100	54,700 (6.6%)
	Cooling	*HLHLHLHLHLLLLM*	73	51 (68%)	328,400	8,200 (2.5%)
B2	Heating	*HHHHHHHHHHHHHH*	76	30 (39%)	163,800	2,100 (1.3%)
	Cooling	*HLHLHLHLHLLLLL*	72	46 (64%)	637,900	56,000 (8.8%)

Regarding the potential applications for the proposed approach, the authors believe that the tool can be widely applied to large centralized or distributed commercial and office building clusters where a limited number of facility operators are in charge of multiple buildings. Examples include schools and universities, military bases, hospitals, bank branches, postal outlets, airports, etc. The method will then aid the operators in efficiently allocating their limited resources. The main requirement for the deployment of the tool is reliable demand measurements at a reasonable resolution (finer than or equal to 1 hour). This is indeed a realistic assumption since the ASHRAE 90.1 standard now mandates archival of usage data for natural gas and electricity at 15-minute intervals for at least 36 months in large non-residential buildings (ASHRAE 2016).

CONCLUSION

Anomaly detection in heating and cooling demands requires a level of information about the building characteristics that cannot be captured by a pure data-driven method. In this work, the authors combined a change-point model and a symbolic quantisation to analyse a large dataset of hourly heating and cooling demand collected from office buildings in Ottawa, Canada. In summary, the case study results show a potential saving of $1\% - 9\%$ for the two studied buildings. The ability of the proposed anomaly detection method in detrending the heating/cooling demands and identifying the typical demand patterns makes it a capable multi-purpose energy-monitoring tool for both operators and stakeholders. The tool can compress large time series of demand data into low-dimension weekly words that increases computational efficiency. In order to make the algorithm more robust and the results more reproducible, a combination of normalisation and statistical techniques are employed, making a comparison of results among several buildings more meaningful.

For the future extension of this work, the authors intend to investigate the sensitivity of the results with respect to several parameters such as quantisation thresholds and number of symbols. Furthermore, an intelligent extension of the algorithm to enable automatic association of faults to certain anomalies will be developed.

ACKNOWLEDGMENTS

This research was supported by the Smart Building Project (A1-006050) of Public Services and Procurement Canada (PSPC) and by the National Research Council Canada (NRC).

REFERENCES

Ashouri, Araz, Yitian Hu, Guy Newsham, and Weiming Shen. 2018. “Energy Performance Based Anomaly Detection in Non-

Residential Buildings Using Symbolic Aggregate Approximation." In *IEEE International Conference on Automation Science and Engineering*. Munich, Germany: IEEE.

ASHRAE. 2016. *ASHRAE Standard 90.1*. Energy Standard for Buildings Except Low-Rise Residential Buildings.

Gunay, Burak, and Weiming Shen. 2017. "Connected and Distributed Sensing in Buildings: Improving Operation and Maintenance." *IEEE Systems, Man, and Cybernetics Magazine* 3 (4). IEEE:27–34.

Gunay, Burak, Weiming Shen, and Chunsheng Yang. 2017. "Characterization of a Building's Operation Using Automation Data: A Review and Case Study." *Building and Environment* 118. Elsevier:196–210.

Kissock, John Kelly, Jeff S Haberl, and David E Claridge. 2003. "Inverse Modeling Toolkit: Numerical Algorithms." *ASHRAE Transactions* 109. American Society of Heating, Refrigeration and Air Conditioning Engineers, Inc.:425.

Miller, Clayton, Zoltan Nagy, and Arno Schlueter. 2015. "Automated Daily Pattern Filtering of Measured Building Performance Data." *Automation in Construction* 49 (PA). Elsevier B.V.:1–17.

Miller, Clayton, Zoltán Nagy, and Arno Schlueter. 2018. "A Review of Unsupervised Statistical Learning and Visual Analytics Techniques Applied to Performance Analysis of Non-Residential Buildings." *Renew Sust Energ Rev* 81. Elsevier:1365–77.

NRC Canada. 2011. *NECB - National Energy Code of Canada for Buildings*. National Research Council Canada, Construction Research Centre.

Omer, Abdeen Mustafa. 2008. "Energy, Environment and Sustainable Development." *Renewable & Sustainable Energy Reviews* 12. Elsevier:2265–2300.

Seem, John E. 2007. "Using Intelligent Data Analysis to Detect Abnormal Energy Consumption in Buildings." *Energ Building* 39. Elsevier:52–58.