Energy & Buildings 199 (2019) 562–578

Contents lists available at ScienceDirect

Energy & Buildings

journal homepage: www.elsevier.com/locate/enbuild

Systematic data mining-based framework to discover potential energy waste patterns in residential buildings

Jun Li [a,b], Karthik Panchabikesan [a], Zhun Yu ("Jerry") [b,*], Fariborz Haghighat [a,*], Mohamed El Mankibi [c], David Corgier [d]

[a] Energy and Environment Group, Department of Building, Civil and Environmental Engineering, Concordia University, Montreal, H3G 1M8, Canada
[b] College of Civil Engineering, National Center for International Research Collaboration in Building Safety and Environment, Hunan University, Changsha, Hunan, 410082, China
[c] Université de Lyon, Ecole Nationale des Travaux Publics de l'Etat (ENTPE), Laboratoire de Tribologie et de Dynamique des Systèmes (LTDS), UMR 5513, 3 rue Maurice Audin, Vaulx en Velin, 69120, France
[d] MANASLU Ing., Savoie Technolac, BP209, Le Bourget du Lac, 73374, France

ARTICLE INFO

Article history:
Received 20 April 2019
Revised 7 June 2019
Accepted 18 July 2019
Available online 18 July 2019

Keywords:
Residential buildings
Occupant behavior
Data mining
Energy savings
Feedback

ABSTRACT

Energy feedback systems are recently proposed to help occupants understand and improve their energy use behavior. Despite many potential benefits, the question remains, whether useful and straightforward knowledge are transferred to the occupants about their energy use patterns. In this context, the key is to develop methodologies that can effectively analyze occupants' energy use behavior and distinguish their energy-inefficient behavior (if any). Previous studies seldom considered the dynamics of occupancy, which may result in misleading information to the occupants and inefficacy in recognizing the actual wasteful behavior. To fill this gap, this study proposes a data mining framework with a combination of change point analysis (CPA), cluster analysis, and association rule mining (ARM) to explore the relationship between occupancy and building energy consumption, aiming at identifying potential energy waste patterns and to provide useful feedback to the occupants. To demonstrate the capability of the developed framework, it was applied to datasets collected from two different apartments located in Lyon, France. Results indicate that different energy waste patterns can be effectively discovered in both apartments through the proposed framework and a substantial amount of energy savings can be achieved by modifying occupants' energy use behavior. The proposed framework is flexible and can be adaptive to households with different occupancy patterns and habitual energy-use behavior. Nevertheless, the discovered energy saving potentials and benchmark values are limited to the apartments considered in this study and similar analysis based on the proposed framework are needed in wider building stocks to explore its generalizability.

1. Introduction

The building sector is undergoing a transition to become smarter and more actively interact with the grid network, thereby reducing its energy consumption. Since occupants play a vital role in building energy consumption, they are expected to effectively manage their energy use behavior to reduce the potential energy wastes. The involvement of occupants in building energy efficiency programs is a useful, cost-effective supplement and also an alternative for traditional physical-based strategies (which usually require high economic investments) [1]. Especially, in residential buildings, occupants are directly responsible for their energy bills and thus there is plenty of room to motivate them to save energy by different strategies such as individual feedback, economic incentives and social norms [1]. Among the many strategies, energy feedback has been attracting increasing interests as it can provide information regarding the energy use data to occupants in a direct and accessible way (e.g. real-time in-home display through home energy management system (HEMS)) [2–6]. Prior studies [3–5] indicated that approximately 6–10 % energy consumption could be saved through different feedback methods (e.g. direct or indirect).

Abbreviations: ARM, Association rule mining; BEMS, Building energy management system; BR, Bedroom; BTR, Bathroom; CPA, Change point analysis; DM, Data mining; EC, Energy consumption; FP-growth, Frequent pattern growth; HEMS, Home energy management system; HVAC, Heating, ventilation, and air conditioning; KIT, Kitchen; LR, Living room; OM, Occupant movement; PS, Power sensor.

* Corresponding authors.
E-mail addresses: zhunyu@hnu.edu.cn (Z. Yu), Fariborz.Haghighat@concordia.ca (F. Haghighat).

https://doi.org/10.1016/j.enbuild.2019.07.032
0378-7788/

Although the energy feedback has been reported to have significant energy saving potentials on intervening occupant behavior, research studies [1,3,5] indicated that the easily interpretable feedback has not yet been transferred to the occupants as useful knowledge. There are still unclear directions for the occupants to modify and improve their energy-use behavior (e.g. which type of behavior needs to be modified and what actions they should take to reduce energy) in practice, which is the main barrier to enlarge occupant-related energy saving potentials. A promising solution to address this issue is to develop effective methodologies that can directly analyze occupants' energy-consuming activities from energy use data and transfer it into organized knowledge and feasible recommendations. As a result, several methods have been proposed to evaluate energy use behavior and identifying their energy saving potentials [7–11]. For example, Yu et al. [7] applied clustering analysis to group buildings with similar occupant behavior and annual energy-saving potential for a high-energy consumption building was estimated by comparing it with an identified reference building. To further analyze and improve the occupant behavior, Yu et al. [8] developed a three-step data mining (DM) methodology in which cluster analysis, decision tree and ARM were applied to find the occupant's energy-inefficient behavior in residential buildings. Similarly, Milad et al. [9] developed a DM-based procedure based on daily end use load to quantitatively evaluate the potential for energy savings in residential buildings. Milad et al. [10] developed a ranking procedure involving DM techniques to evaluate the building energy performance based on occupant behavior which can motivate residents to reduce their energy consumption and improve their rank when compared to the others.

The above studies demonstrated the effectiveness of using DM techniques in assessing occupant behavior related energy-saving potentials in residential buildings. However, the main disadvantage of the proposed methodologies is that they did not consider the dynamic variation in occupancy patterns (e.g. changes in occupancy status or the number of occupants) while estimating the energy savings. Indeed, without considering the occupancy patterns, it is very hard to distinguish whether the energy consumed by the appliances is caused by the actual usage or the energy inefficient behavior of the occupants. For example, when the amount of energy consumed by some appliances (e.g. computers, TV) are the same in two periods with different occupancy status (presence and absence), it is very hard to verify whether the appliances are in use or left on working status when occupants are not needed. Moreover, they face challenges in effectively recognizing irregular energy consumption patterns associated with infrequently occurred occupant activities. For example, when occupants perform one activity multiple times (e.g. doing laundry three times a day), such activity is unusual which might be considered as wasteful behavior in the previous works. This might lead to inaccurately identifying actual wastes and overestimation of energy saving potentials. Nowadays, due to the increasing growth in occupancy sensing technologies (e.g. non-intrusive sensors such as passive infrared (PIR) sensor) and their wide applications in the modern home environment, it is becoming more feasible to continuously monitor and collect detailed and high-resolution occupancy data. Accordingly, it is of vital importance to analyze the energy consumption data combining with dynamic and actual occupancy information, so as to extract more rational and practical energy-saving recommendations for occupants.

Very few researchers have attempted to consider occupant's actual occupancy behavior and developed activity-based inference methods to correlate the identified occupant activities with energy consumption [12,13]. Ahmadi-Karvigh et al. [13] proposed a framework using inductive and deductive reasoning to infer occupant activities and estimated the respective energy wastes. This method relies heavily on an accurate interpretation of the relationship between the sequence of appliance usage and corresponding occupant activities and thus would be particularly suitable for single-occupant families. When such a framework was applied to multiple-occupant households, the method may be too complicated as various activities might occur simultaneously and the respective sequence of appliance usage is sensitive to different occupants. This would result in difficulties in accurately differentiating activities associated with multiple occupants, thereby resulting in difficulties in identifying the energy wastes [14]. Hence, questions about how to effectively consider the dynamic variation of occupancy patterns in order to identify whether the variation of energy use is caused by occupants' energy-inefficient behavior or natural variability of occupancy patterns still remains unsolved. In this view, it is highly desirable that a general framework for assessing occupants' wasteful behavior taking into consideration both the occupancy and energy consumption patterns can be developed.

The main objective of this work is to develop a novel and general DM-based framework to identify the potential energy waste patterns in residential buildings by combining energy consumption and dynamic occupancy behavior. The proposed framework aims at (1) explicitly exploring the relationship between occupancy and energy consumption patterns; (2) developing benchmarking baselines that can discover actual energy waste patterns; (3) differentiating the irregular energy use from the energy waste patterns so as to give more practical feedback. To achieve the above said objectives, at first, a daily schedule is separated into different periods in order to approximately adapt to the routines of a given household and to represent the distinct characteristics. Then regular and irregular occupancy and energy use patterns are extracted for each period and are compared with each other to identify whether the energy consumption pattern is in accord with the occupancy level. At last, a set of benchmarking rules is created as baselines to identify energy saving potentials in line with occupant behavior. The proposed framework was implemented over the one-year dataset collected from two households with different occupancy and energy use behavior. Results show that different energy waste patterns can be effectively discovered in both the apartments and promising energy saving potentials are possible by improving occupants' wasteful behavior. Though the application of the proposed framework was validated in two different households, the discovered energy saving potentials and benchmark values were limited to buildings with similar contexts and further investigations are needed to explore the generalizability and robustness in wider building stocks.

The present work is organized as follows. Section 2 briefs the step-by-step procedure involved in the proposed framework. In Section 3, data collection, preprocessing and preparation are explained. The results obtained through the proposed framework are given in Section 4. Section 5 is a discussion about the advantages of using occupancy information for identifying occupants' energy-inefficient behavior. Section 4 describes some limitations and potential future studies. The main conclusions of this study are summarized in Section 7.

2. Proposed framework

The proposed framework consists of four steps (as shown in Fig. 1), including period estimation, identification of matching and mismatching patterns, creation of benchmarking rules and identification of energy wastes.

2.1. Step 1: *Period estimation*-change point analysis

In general, occupancy and energy consumption patterns represent the residents' daily activities and appliances usage. Hence, unlike the office building energy use analysis, where the periods in

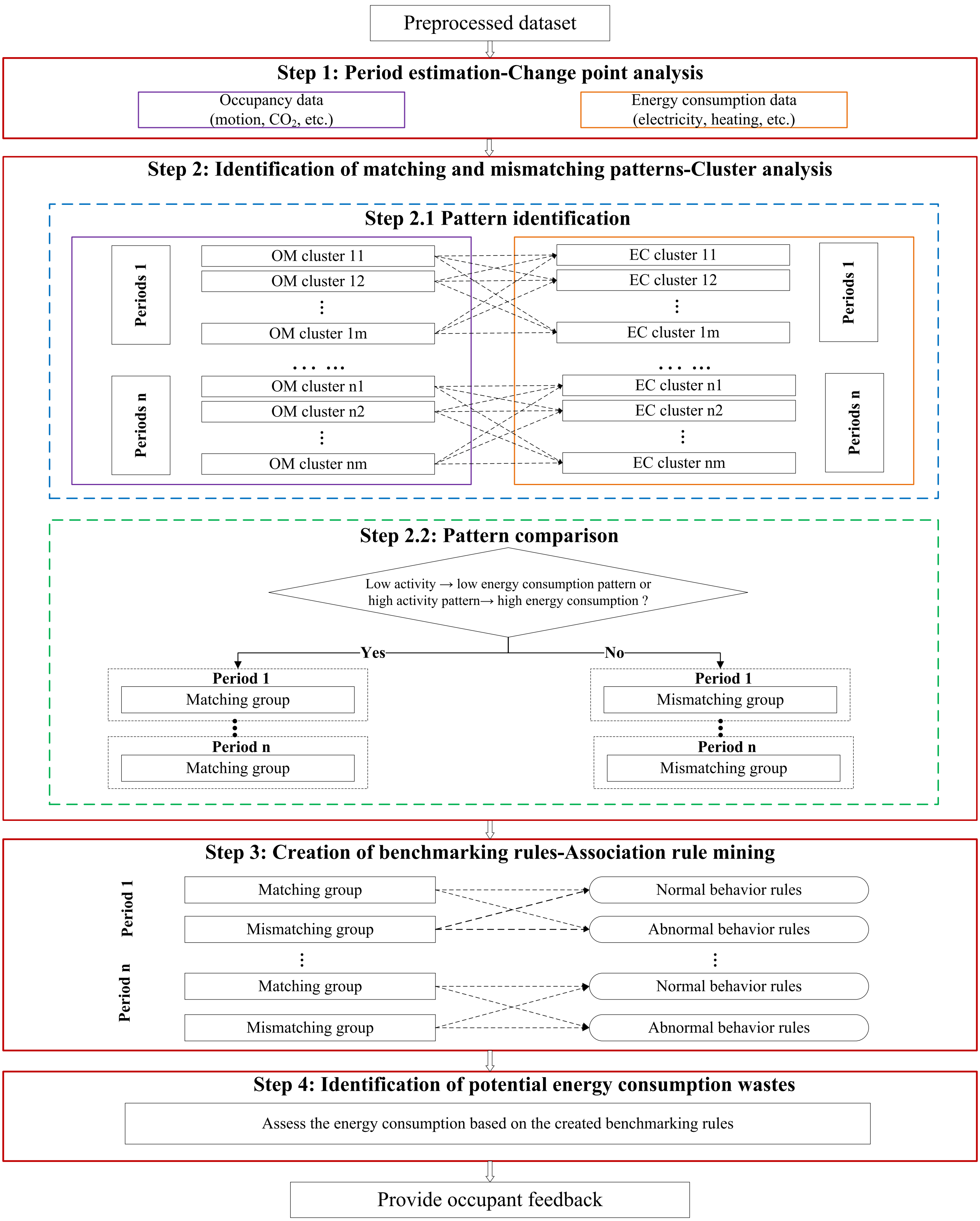


Fig. 1. Framework of the proposed methodology to identify energy waste patterns in residential buildings (EC and OM are the abbreviations of energy consumption and occupant movement).

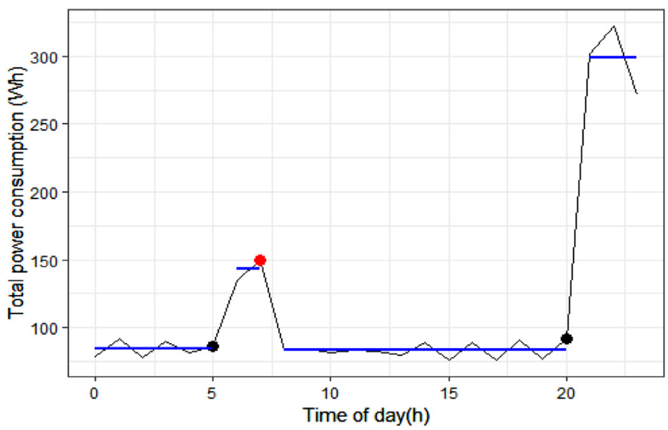


Fig. 2. Illustration of change point analysis (black dot and red dot represents an 'increase' and 'decrease' change point respectively). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

a day are mainly split based on general office timings (e.g. 9.00–12:00 as morning period, 2:00–5:00 as afternoon period, etc.,), in residential buildings, it is significant to determine the periods based on the daily profile of the occupants, which shows the distinct variations in occupancy and energy consumption of the respective household. In prior studies, due to the increasing availability of high-frequency occupancy and energy related data, change point analysis (CPA) [15], a statistical method to detect the changes in time series data, has been widely applied in building data analysis applications such as fault detection and building energy performance analysis [16,17]. These studies have proved its effectiveness in identifying multiple events (especially non-routine events) occurred in building system operation [17]. Consequently, CPA was used in this study to find the appropriate periods in a day that represent occupants' daily routines. In principle, a change point can be detected when the statistical distribution (e.g. mean or variance) after a point is significantly different from the distribution of current and its previous distribution. In this study, CPA was performed using '*changepoint*' package available in 'R' programming software [18]. Prior to performing CPA, the maximum number of change points ('Q') and the method to detect the change points should be assigned. Among the available methods (AMOC, PELT, SegNeigh, BinSeg), BinSeg method was used to detect the change points [18]. For the maximum number of change points, several values for 'Q' was tried and eventually 'Q' was set to three and four for apartment 1 and 2 (details about the apartments are given in Section 3.1), respectively. Fig. 2 is an illustration for defining the number and duration of periods in a given day using CPA by considering the changes detected in the energy consumption pattern. In the considered example, changes in energy consumption on a specific day occurred at 5:00, 7:00 and 20:00. Accordingly, periods for that day can be divided into Period 1 (0:00 h–5:00 h), Period 2 (5:00 h–7:00 h), Period 3 (7:00 h–20:00 h), and Period 4 (21:00 h–23:00 h).

2.2. Step 2: Identification of matching and mismatching patterns – Cluster analysis

As one of the most commonly used data mining techniques, cluster analysis shows superior performance in identifying unknown patterns/groups and has been extensively applied in multiple domains for building performance improvement. In the literature, different clustering methods (e.g. k-means and hierarchical clustering) were used to serve different application purposes, e.g. feature extraction for building energy prediction [19], load profiling [20], nearly zero energy buildings (nZEBs) grouping [21] and occupant behavior pattern recognition [22,23]. Among the available cluster analysis methods, *k-means* cluster analysis has gained the greatest popularity due to its simplicity and high effectiveness. In present study, once the periods in a day are distinguished based on CPA, *k-means* clustering analysis [24] is performed on the energy consumption and occupancy data of each period separately to split them into several subsets (Step 2.1). Note that, the occupancy/energy consumption dataset in these subsets are homogeneous internally and heterogeneous between clusters. To perform *k*-means clustering, prior knowledge regarding the optimal number of clusters is required and accordingly, the elbow method is adopted to find the optimal number of the clusters [24]. After clustering the occupancy and energy consumption data separately, comparisons are made between the respective occupancy and energy consumption clusters to identify the matching and mismatching patterns (i.e. OB-EC sub-clusters) in each period (refer to Step 2.2 in Fig. 1). The matching and mismatching days are identified by checking whether energy consumption is in accord with the occupancy level. If the occupancy level in a period aligns with the energy assumption, i.e. low (high) occupancy linked with low (high) energy consumption, it is considered as matching patterns, vice versa. In this way, the general relationship between occupancy and energy consumption are explicitly found in this study and they are interpreted accordingly. It should be noted that the matching and mismatching mechanisms do not guarantee an accurate description of the relationship between occupancy and energy consumption, however, by doing so, it is easier to further identify the actual reason behind the abnormal energy consumption pattern (occupancy variation or wasteful behavior).

2.3. *Step 3: Development of benchmarking rules* – Association rule mining

In general, energy wastes caused by occupant behavior most probably occur when the occupants are asleep or away from the home. A typical example is that occupants might forget to turn off the unneeded appliances (e.g. lights, computer, HVAC) during the sleeping/away period. Previous studies reported that 70% of lighting consumption could be saved in an institutional building if such wasteful behavior is avoided [25]. Furthermore, not completely turning OFF some of the household appliances (especially those with a relatively high standby power) can also be considered as potential energy waste. As there is increasing use of multiple ICT devices (e.g. set-top box and game console), there exists a noteworthy consumption of standby power which needs particular attention [26,27]. Such waste is prone to be often neglected by occupants due to the lack of energy-saving awareness or limited knowledge on equipment standby mode energy consumption. Considering the above said energy waste occurrences and diversity in the usage of household appliances, a reasonable baseline method that can discover different types of wasteful behavior is a key issue for assessing occupants' energy saving potentials. To do so, knowledge about occupants' general behavioral patterns and associated energy consumption needs to be extracted first before the development of benchmarking rules. Moreover, specific knowledge about which energy consumption level is strongly correlated with occupants' activities (sleeping or away) would be very helpful in identifying the actual energy wastes. Association rule mining (ARM) is another popular DM technique applied to building sector which can uncover interesting and recurring relationship from tremendous amounts of data. Many recent research works explored the effectiveness of ARM technique in analyzing building automation data, for instance, interactive associations among building subsystem operations and faulty condition detection [23,28–32]. Results of these studies indicated that building performance could be remarkably improved by using the extracted useful knowledge through ARM.

Based on the identified matching/mismatching groups through cluster analysis, in this study ARM is performed on each sub clusters, to extract a set of rules which can be categorized into several categories: (1) rules that explain the positive relationship between occupancy, energy consumption are considered as normal behavior rules (e.g. high occupancy level with high energy consumption; (2) rules describing low occupancy level with relatively high-energy consumption are considered as abnormal behavior rules, as the energy consumption in these rules lies in a relatively high range, which is different from the normal behavior rules. This is because if the energy consumption is higher than the energy consumption value specified in the benchmarking rules when occupancy level is in the lowest range for a defined duration, it hints a high chance for energy wastage; (3) rules that indicates the relationship between the lowest occupancy level in each period with the lowest energy consumption are considered as the benchmarking rules. Though these rules also can be categorized as 'normal rules', it is categorized as 'benchmarking rules' since such rules serve as the baseline to identify the actual and more reasonable energy waste patterns. As the rules obtained by ARM are time-independent, a threshold for the duration of identified instances needs to be added in the benchmarking rules to identify the abnormal behavior. This means that whenever the total energy consumption pattern above its defined threshold value prolonged continuously for the defined duration (e.g. 1 h) and if absence/sleeping event is recorded, such a pattern is specified as energy waste. Hence, the framed benchmarking rules must include the following: (i) indicators/rules to recognize households' unoccupied and sleeping period, and (ii) energy and time-based benchmarks (i.e. setting the thresholds, above which the energy consumption pattern in a given period can be specified as energy waste instance). Several algorithms exist to mine association rules from a large dataset. The frequent pattern growth (FP-growth) algorithm is adopted in this work due to its high effectiveness and wide application. The open-source data mining software *"RapidMiner"* [33] is used to perform ARM.

2.4. Step 4: Identification of potential energy wastes

On the basis of the identified benchmarking rules, an algorithm is developed to identify the potential energy wastes and send feedback to the occupants, as shown in Fig. 3. The collected occupancy and energy consumption data serve as input to the algorithm. The instances that do not satisfy the benchmarking rules are collected to check their time continuity. If the potential waste instances are continuous and the duration exceeds a specific threshold, it is considered as one instance of energy waste patterns.

3. Data collection and preliminary analysis

3.1. Data description

The dataset collected from a high-performance residential building of a district located in Lyon, France is used in this study. To monitor the indoor environment and energy performance, each apartment in the building is equipped with a home energy management system (HEMS) and the data on both occupancy movement and power consumption (i.e. plug and lighting consumption) were recorded for every minute. In this study, these two categories of data collected in the year 2016 were used. It should be noted that, though individual power sensors were installed in these apartments, the information regarding their connected appliances and zones are unknown due to the privacy issue.

To show the applicability of the proposed methodology to different buildings, two apartments with different footprint, occupancy and energy profiles were selected. Table 1 shows the major characteristics of the two apartments. In these two apartments, motion sensors were installed in the living room (LR), bedroom (BR), kitchen (KIT), corridor (COR), bathroom (BTR), and toilet (TOL). The details pretaining to the type of sensors installed in apartment 1 and 2, their accuracy, measurement resolution are given in Table 2. Fig. 4 shows the floor plan and motion sensors locations in different rooms in apartment 1 and 2, respectively.

3.2. Data preprocessing and preparation

3.2.1. Data preprocessing

Data preprocessing is an important step before applying DM techniques, as the raw dataset obtained from the BAS/BEMS may contain both not available (NA)/missing values and outliers. Accordingly, the quantile method is used to detect outliers. If the dataset is missing or filled with (NA) continuously for several hours in a day, then the particular day is omitted from the dataset. If the data is not continuously missing or filled with NA's, then the missing values/ values with NA's are replaced by the average of previous two values in the dataset. After processing the outliers and missing values, the dataset for 310 days (446,400 observations) and 315 days (453,600 observations) were considered for apartment 1 and 2, respectively. Note that the collected data was in the one-minute time interval and data aggregation was carried out to aggregate one-minute data into hourly data.

3.2.2. Data preparation for clustering analysis

To find the distinct groups of occupancy and energy consumption patterns, respective features for occupancy and energy consumption dataset were determined for the cluster analysis. For occupancy, the total number of occupant movements during a given period was used as the feature, as it can represent the occupancy and activity level. For energy consumption, four statistical measures, i.e. minimum, maximum, average and standard deviation, were calculated for each period and used as features for the cluster analysis. The advantage of using statistical measures rather than energy consumption data at each time step is that it would significantly reduce the dimensionality of clustering but can still describe the temporal variation. Note that, all the features used for cluster analysis were scaled into the range 0–1 by using the min–max normalization method [7].

3.2.3. Data preparation for associate rule mining

The important step before performing ARM through the frequent pattern (FP)-growth algorithm is the discretization of all the numerical attributes into categorical values. Accordingly, equal frequency method was used for data discretization. Note that, 4 levels of categorization (range 1, 2, 3 and 4) were considered for energy consumption in order to find a reasonable threshold for energy wastes identification. Regarding the occupancy data, quartile method (1st, 3rd quartile, median and maximum) was used to check the distribution of occupant's movement data and accordingly, the dataset was discretized into two to three categories based on the corresponding data distribution in each cluster. For instance, in high activity cluster where the total number of movements is high, the motion data is discretized into three ranges ('no', 'less', and 'frequent' movement). For the low and very low cluster, the data is categorized into either 2 (i.e. no movement or movement) or 3 (no movement, one movement and frequent movement) categories. Note that, the value defined for 'frequent movement' differs for each cluster based on the data distribution.

1. Define a $m \times n$ matrix $M = (a_1, \quad a_2, \quad \cdots, \quad a_n)$ where $a_1, a_2, \ldots a_n$ are vectors representing the benchmarking values for occupancy and energy consumption patterns in Period1, Period2, … , Period n;

2. **Input:** Occupancy data* $X=\{x_1, x_2, \ldots, x_t\}$, energy consumption data $Y=\{y_1, y_2, \ldots, y_t\}$, time vector $Z=\{z_1, z_2, \ldots, z_t\}$, where t is the total number of data points

3. **For** i =1 to *t*
   - (a) If x_i matches with the benchmarking value of occupancy
   - (b) Check the time continuity with the previous instances
   - (c) If duration greater than the defined threshold (e.g. 1h)
   - (d) Identify occupancy patterns $S=\{S1, \ldots, S_k\}$

4. **For** j=1 to *k*
   - (a) Compute the average energy consumption in the identified intervals
   - (b) If the average energy consumption greater than the benchmarking value of energy consumption
   - (c) Compute the total energy wastes

5. **Output:** Energy waste patterns

* Based on the availability, the number of inputs for the occupancy ($x_1, x_2, \ldots$.) and energy consumption data ($y_1, y_2, \ldots$.) varies.

Fig. 3. The pseudo code of the algorithm to identify potential energy wastes instances.

Table 1
Description of major characteristics of apartment 1 and 2.

Name	Floor area (m^2)	No. of bedrooms	No. of motion sensors	No. of power sensors	No. of lighting sensors
Apartment 1	51.65	1	10	13	8
Apartment 2	97.6	3	14	18	14

Table 2
Description of sensors installed in apartment 1 and 2.

Sensor	Company name	Type	Accuracy	Measurement resolution
Presence detector	Theben	PlanoCentro A-KNX	- (detection area 64 m^2 if seated)	Event[a]
Power	ABB	KNX Energy Module: EM/S 3.16.1	±2/3/6%	1 min

[a] Event-based sensors can be triggered at any time. The motion data used was transformed into the structured data of which all the sensors are organized at 1-min resolution. In other words, if one or more movements are detected within one minute, it is recognized as one.

4. Results

4.1. CPA results

Fig. 5 (a) & (b) show the frequency of detected change points for total energy consumption and occupancy profiles of apartment 1, respectively. Since there are several change points detected for both occupancy and energy consumption dataset, a threshold for frequency (i.e. 0.25) was set to determine the reasonable number of change points or periods in a day. From Fig. 5 (a) & (b), it is inferred that in apartment 1, the frequent change points (i.e. change points with frequency ≥0.25) occurred at 5:00 h, 6:00 h, 7:00 h, 8:00 h, 19:00 h, 20:00 h in terms of both the energy consumption and occupancy profiles. Thus, after considering both frequency and the statistical features, four periods were identified in apartment 1 and corresponding statistical data in terms of energy and occupancy are summarized in Table 3. This table shows that the average hourly energy consumption and mean occupancy in *Period 2* & *Period 4* are relatively higher, implying higher energy consumption patterns and higher occupancy in these two periods. As motion levels and energy consumption in nighttime (i.e. *Period 1*) are the lowest, it can be construed that the sleeping activity dominates this period. A relative lower occupancy level but the high variance in the day time (i.e. Period 3) denotes that for most of the days, there is a higher chance that occupants are absent from the home.

Fig. 6 depicts the frequency of the discovered change points for apartment 2 and Table 4 summarizes the identified periods and respective statistical data in terms of energy consumption and occupancy. In apartment 2, seven frequent change points for energy consumption profiles were obtained at 1:00 h, 7:00 h, 10:00 h, 11:00 h, 12:00 h, 18:00 h, 19:00 h while only two were found for occupancy profiles at 6:00 h and 19:00. This indicates that there is a noticeable difference between energy consumption and occupancy profiles in terms of the discovered frequent points and respective frequency value. For instance, an interesting change point (i.e. 1:00 h) was only observed for energy consumption at nighttime while occupancy pattern does not significantly change at the same hour. The possible reason behind this is that some

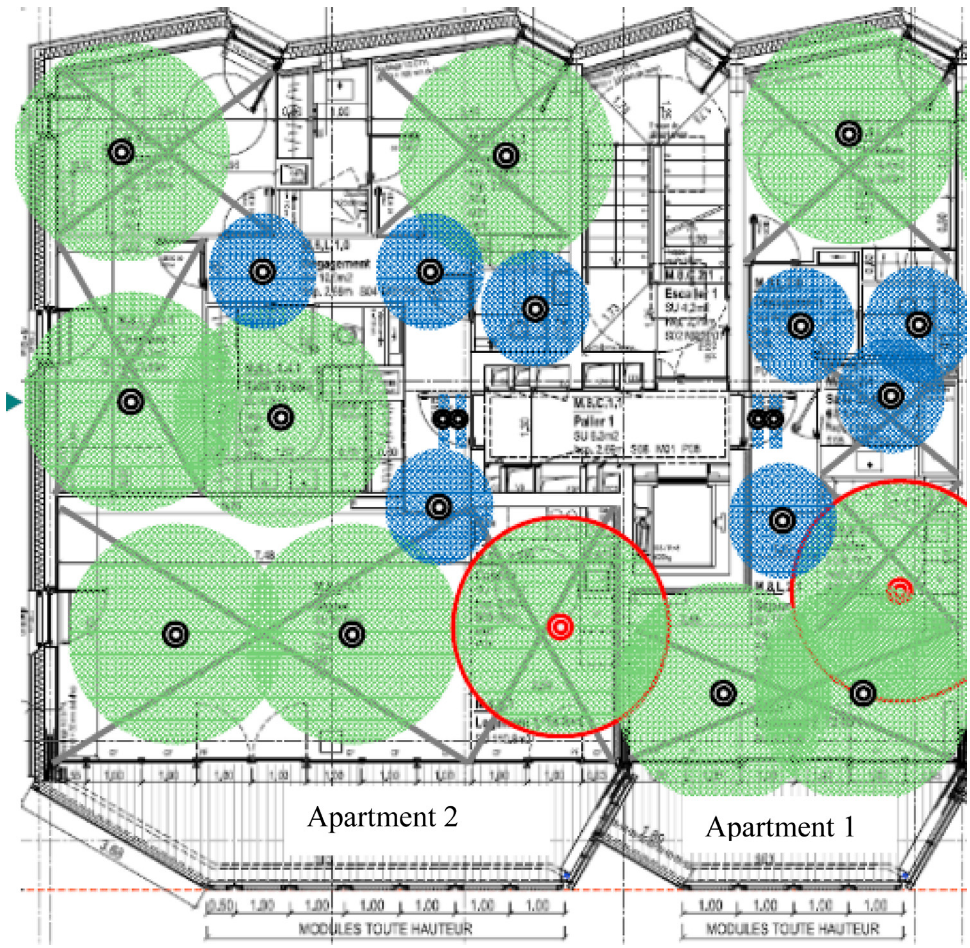


Fig. 4. Floor plan and motion sensors deployed in apartment 1 and apartment 2.

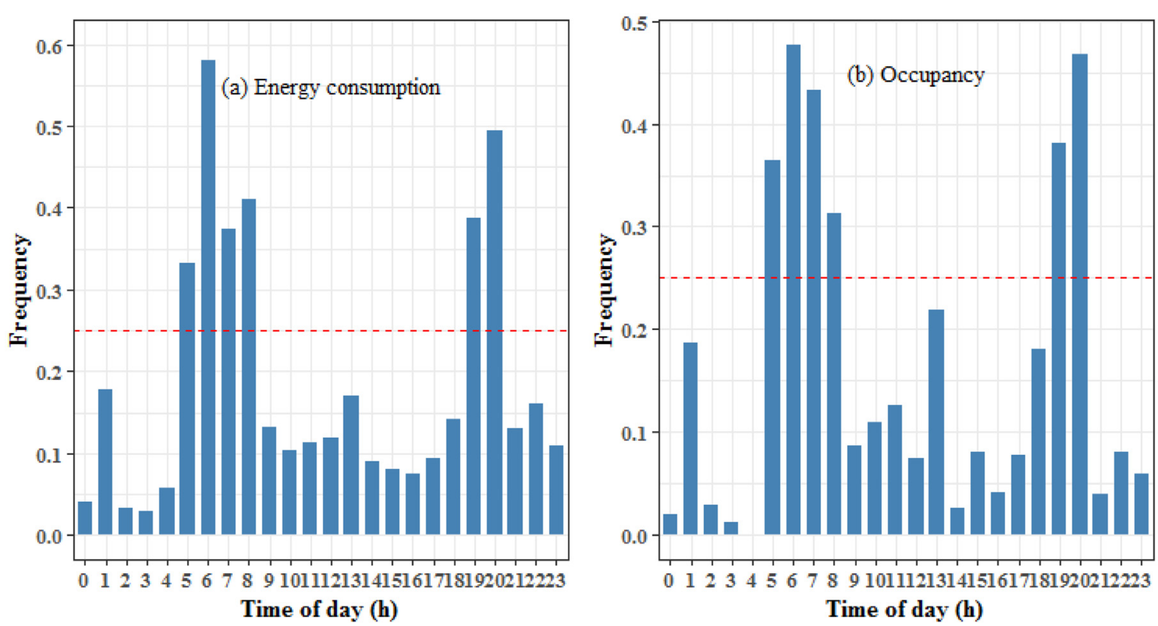


Fig. 5. Results of CPA for energy consumption and occupancy (apartment 1).

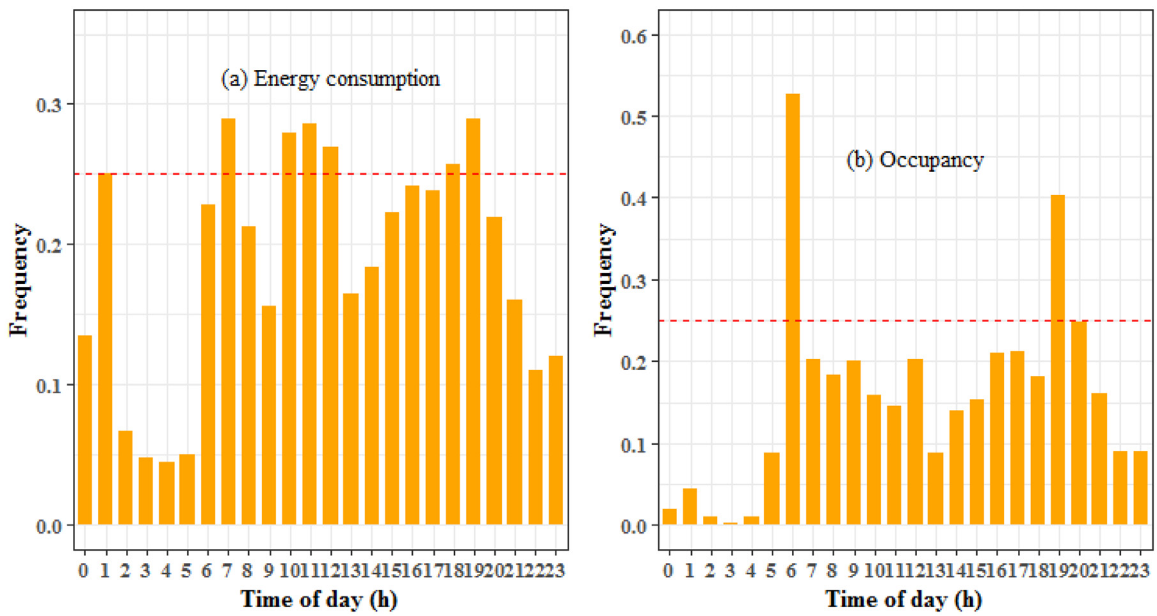


Fig. 6. Results of CPA for energy consumption and occupancy (apartment 2).

Table 3
Energy consumption and occupancy characteristics of four estimated periods in apartment 1.

Period	Time	Energy consumption (Wh)		Occupant movement (no. of counts)	
		Mean	Std.deviation	Mean	Std.deviation
Period 1	00:00–05:00	104	47.9	4	12.5
Period 2	06:00–08:00	216	192.9	22	26.9
Period 3	09:00–19:00	147	166.4	12	27.4
Period 4	20:00–23:00	255	200.2	37	35.6

Table 4
Energy consumption and occupancy characteristics of five estimated periods in apartment 2.

Period	Time	Energy consumption (Wh)		Occupant movement (no. of counts)	
		Mean	Std. deviation	Mean	Std.deviation
Period 1	02:00–07:00	143	83.3	11	22.1
Period 2	08:00–10:00	376	293.8	52	41.2
Period 3	11:00–12:00	463	410.7	50	43
Period 4	13:00–18:00	354	336.2	43	41.9
Period 5	19:00–01:00	342	267.2	22	29.2

particular appliances (such as dishwashers) might be preprogrammed by the occupants before they go to sleep. Such an energy use pattern is significantly different from the results found in apartment 1. Furthermore, a relatively lower frequency value in energy consumption profile indicates that the energy use is highly variable and diverse in apartment 2 implying more different patterns could possibly exist. For occupancy profiles, a high frequency of the identified two frequent change points (i.e. 6:00 and 19:00) was observed which denotes that the occupants may perform similar activities. For example, occupant might regularly wake up between 6:00 and 6:59 and after 19:00, occupant usually perform some activities with low motion level (such as watching TV). In addition, according to the statistical data shown in Table 4, the number of movements and energy consumption is relatively high in *Period 2* to *Period 5* indicating a high probability of occupants' presence at home with a wide range of different occupant activities. Especially in *Period 3*, the occupants might involve in cooking activities, implying higher energy consumption in this period.

Comparing the frequent change points found in apartment 1 and 2, it indicates the different lifestyle and daily routines of occupant activities between these two households. Such difference clearly emphasizes the necessity of finding unique divisions (in terms of periods in a day) rather than considering the same periods for all the apartments/residential buildings. The results obtained from CPA denote that it is critical to take into consideration the distinct energy use behavior in these periods (especially during nighttime) to assess the energy waste instances associated with occupant behavior.

4.2. Cluster analysis results

Considering the interpretation of the clustering results and further comparison, identical cluster numbers were set for both occupancy and energy consumption dataset. Based on the results of the elbow method, for apartment 1 four clusters were identified for *Period 1, Period* 2, *Period* 4; and three clusters for *Period 3*. Fig. 7 shows the heatmap of cluster centroids for each dataset belonging to the identified periods in apartment 1. The characteristics of each cluster can be described by the respective cluster centroid values, which represent the mean value of the extracted features. In terms of energy consumption clusters, for *Period 1* in apartment 1, cluster 2 has the highest value in terms of statistical measures, which

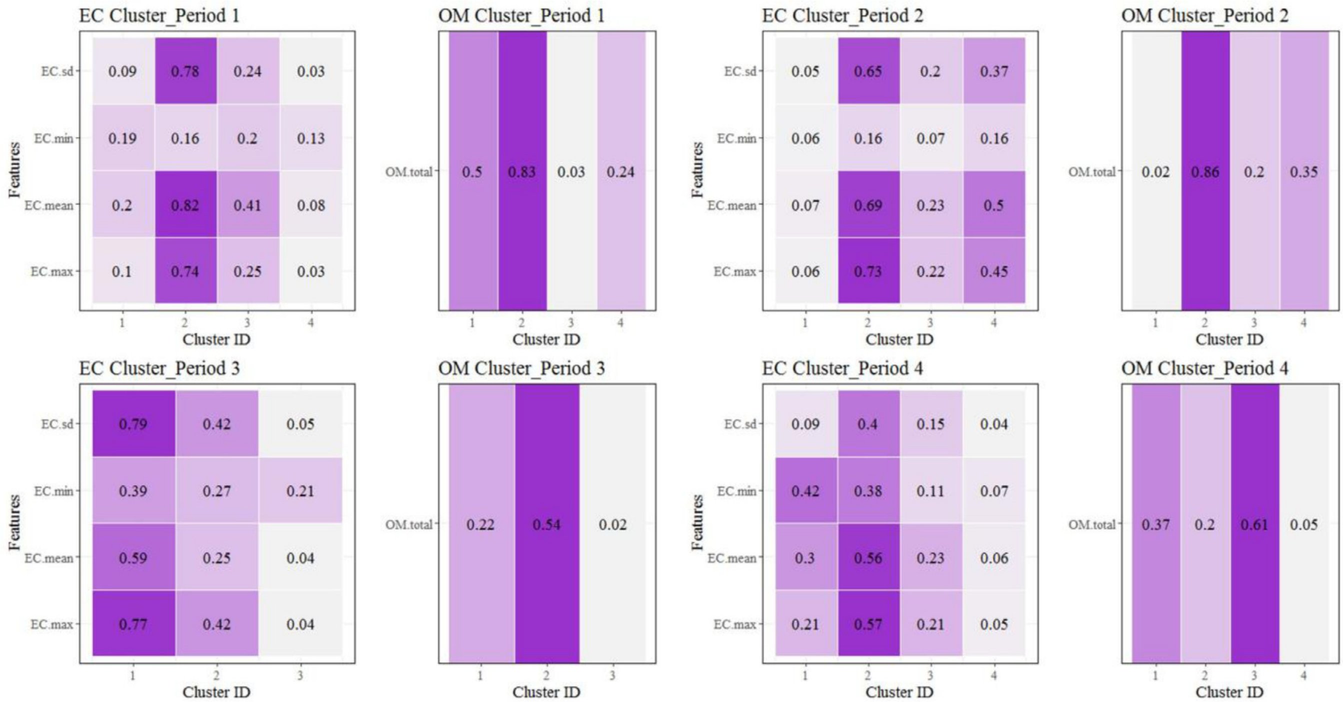


Fig. 7. Visualization of cluster centroids of discovered clusters for all the *Periods* in apartment 1.

indicates that total energy consumption is relatively higher than other clusters. Therefore, cluster 1, 2, 3 and 4 are represented as 'Low (L.EC)', 'High' (H.EC), 'Medium' (M.EC) and 'Very low' (V.EC) energy consumption, respectively. Similarly, for occupancy clusters of *Period 1,* the centroid value for the total number of movements in cluster 2 is higher, implying the higher occupant's involvement in more household activities than the other three clusters. Hence, this cluster is described as 'High occupant movement' (H.OM); correspondingly, the centroid value for cluster 3 is the lowest and is described as 'Very low occupant movement' (VL.OM); cluster 1 and cluster 4 is described as 'Medium occupant movement' (M.OM) and 'Low occupant movement' (L.OM), respectively. The interpretations were made in a similar way for other energy consumption and occupancy clusters of different periods. For apartment 2, the number of clusters was identified as three for *Period* 1 and four for the remaining period. Fig. 8 depicts the heatmap of clusters centroids for each period in apartment 2. The same definitions for occupancy and energy consumption clusters can be adapted to apartment 2.

To further identify the matching and mismatching days (refer Section 2.2) in these two apartments, the distribution of OM clusters, EC clusters are compared with each other and subsequently, OM-EC sub-clusters are formed. For the illustration purposes, the comparison of OM with EC clusters and the respective results obtained for *Period 1* in apartment 1 is illustrated in Fig. 9. It can be seen from the figure that for *Period 1*, 65.4 % (203 days) of the days in the whole year belongs to 'VL.OM' cluster of which occupant activity is at a lower level, indicating that 'VL.OM' is the regular/most frequently occurred patterns in *Period 1*; whereas in terms of EC clusters, 199 days belong to 'VL.EC' clusters which are the most frequently occurred energy consumption patterns. Regarding the OM-EC sub-clusters, in 'VL.OM' cluster, 80.8 % of days relates to 'VL.EC' pattern, which implies the variation of energy use is largely consistent (matching) with occupancy behavior. However, there are 38 (18.7%) and 1 (0.5%) days belonged to 'L.EC' and 'H.EC' clusters, respectively which imply a mismatch between the occupancy and energy consumption. Due to the high consistency between occupancy and energy consumption in apartment 1, the mismatching instances found in this apartment are interesting which hints a high chance for unusual energy use caused by occupant energy inefficient behavior. In addition, Fig. 9 shows that when irregular occupant activities ('M.OM' and 'H.OM' in *Period 1*) occur, energy consumption level does not always follow a similar change, indicating that the extent of energy use variation in this period is not closely correlated with occupancy variation.

Similarly, Fig. 10 depicts the distribution of OM, EC clusters and OM-EC sub clusters for *Period* 1 in apartment 2. This figure shows that there are two clusters (L.OM and M.OM) where the number of days in both occupancy and energy consumption is greater than 100 (exceeding one-third of a year). This indicates that, compared the patterns discovered in apartment 1, more diverse and complex patterns of both energy consumption and occupancy are explored in apartment 2, which further verifies the results obtained from CPA. In terms of the data distribution of OM-EC sub-clusters, Fig. 10 shows that in 'L.OM' cluster of *Period 1*, about 95.5% of the days were observed in 'L.OM-L.EC' sub-groups. The above-mentioned results verify the hypothetical relationship between occupancy and energy consumption between these apartments.

4.3. ARM results

Tables 5 and 6 summarize the association rules obtained for different periods in apartment 1 and 2, respectively. In these two tables, association rules involving low energy consumption range, and lowest occupancy level (i.e. sleeping/absent) and some interesting rules are presented. The rules shown in the table are used for defining benchmarking rules and identifying the abnormal energy use patterns.

According to the association rules obtained from different OM-EC sub-clusters, it can be construed that for apartment 1 and 2, when the occupants are sleeping or away from home, energy consumption falls into the lowest level, e.g. Rule 1–4 in apartment 1 (Table 5) and Rule 1–5 in apartment 2 (Table 6). The high support and high confidence value verify that such benchmarking

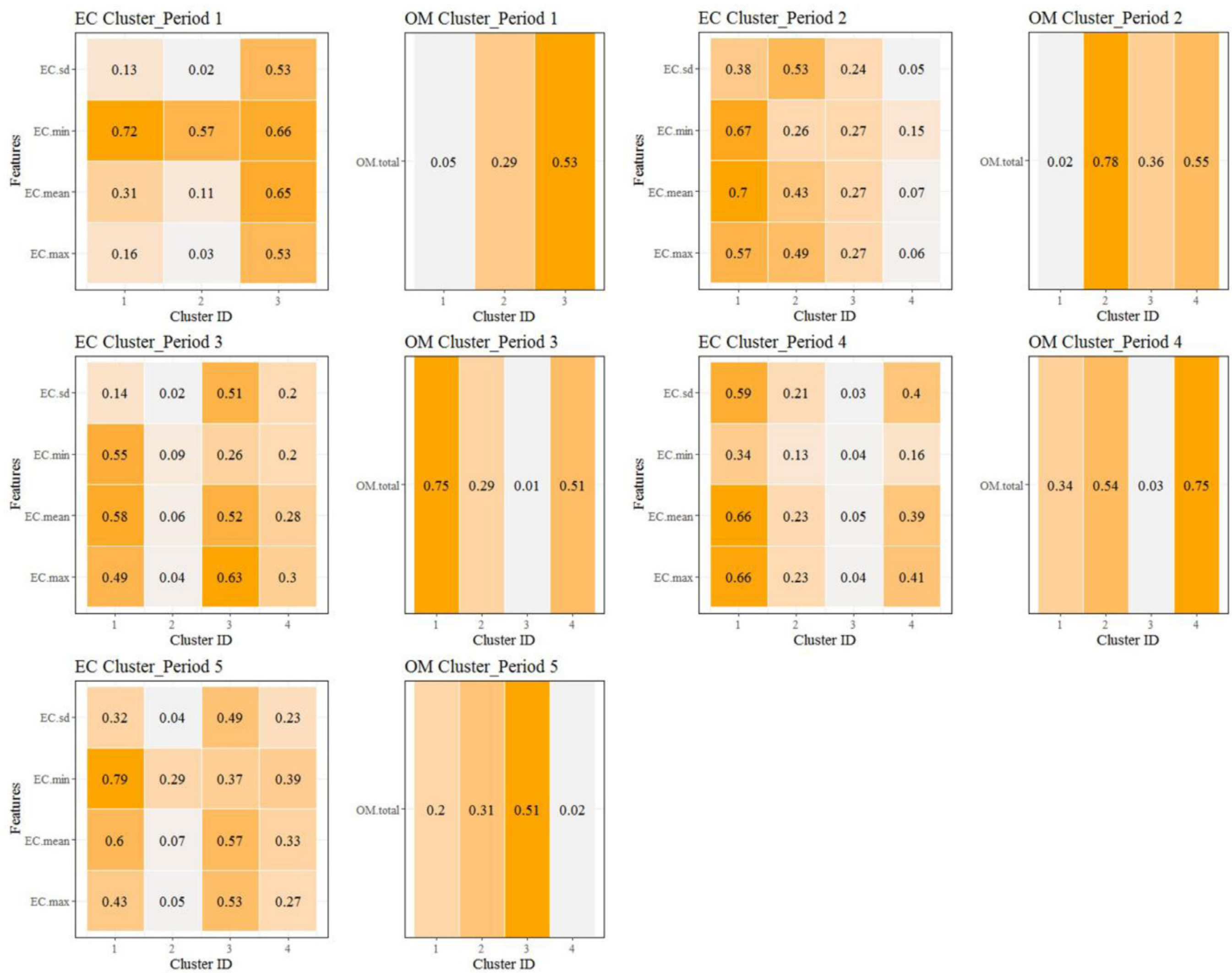


Fig. 8. Visualization of cluster centroids of discovered clusters for all the periods in apartment 2.

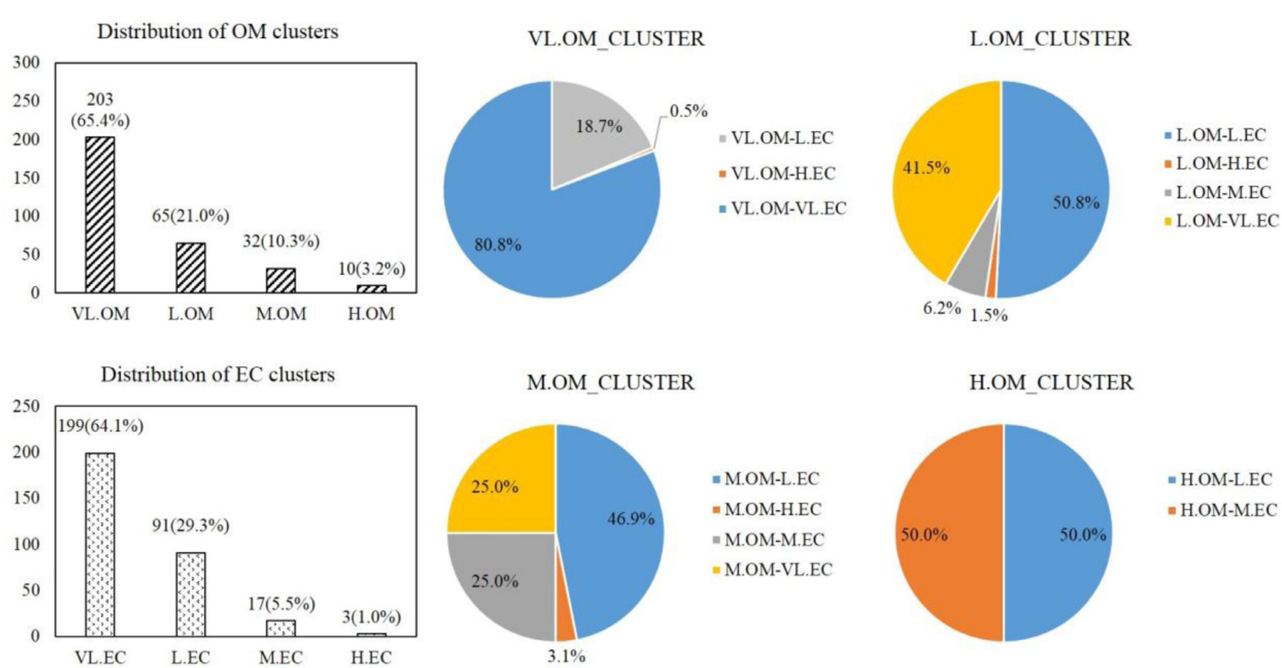


Fig. 9. Comparison of occupancy and energy consumption clusters for *Period 1* (0:00–5:00), apartment 1.

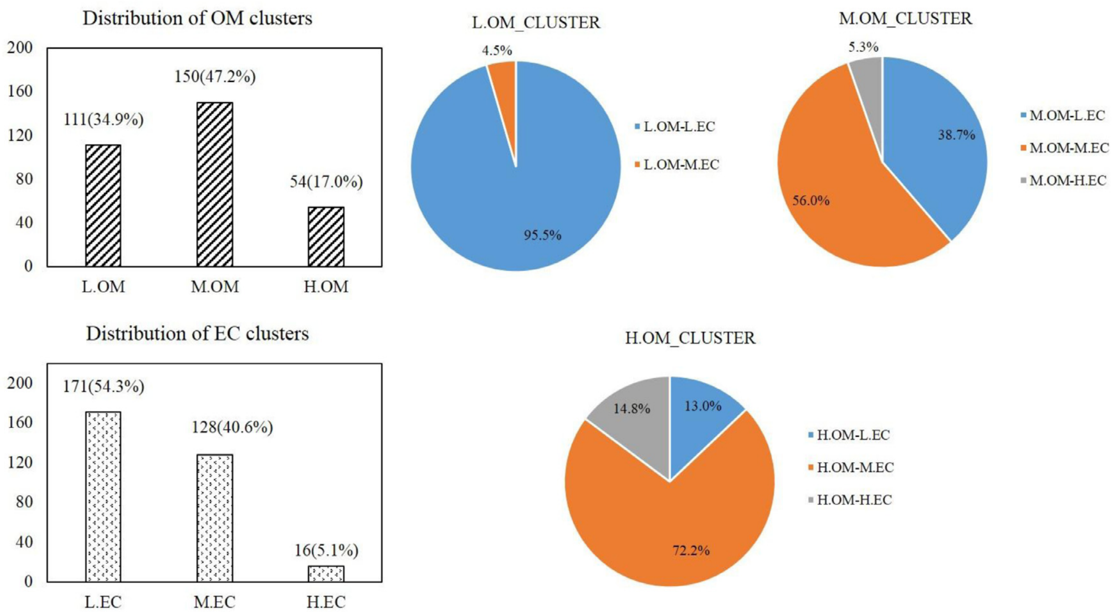


Fig. 10. Comparison of occupancy and energy consumption clusters for *Period 1* (2:00–7:00), apartment 2.

Table 5
Rules used to define benchmarking rules in apartment 1.

Rule No.	Period	Cluster	Premises	Conclusion	Sup.	Conf.	Lift	Count
1	1	VL.OM-VL.EC	Motion = [0, 3) [a]	Energy = [0,104)	0.74	0.87	1.03	725
2	2	VL.OM-VL.EC	Motion = [0, 1) Day of week = Sunday	Energy = [0,114)	0.34	0.89	1.18	56
3	3	L.OM-L.EC	Motion = [0, 1)	Energy = [0,114)	0.77	0.86	1.02	1693
4	4	VL.OM-V.EC	Motion = [0, 1)	Energy = [0,111)	0.50	0.80	1.14	134
5	1	VL.OM-L.EC	Energy = [>110)	Motion = [0,3)	0.46	0.77	0.98	105

[a] The total number of movements is not always zero during *Period 1* (sleeping period) since the motion sensors can be triggered in the bedroom when occupants are asleep. Hence, the occupancy counts up to 3 is considered for *Period* 1. The same interpretation for *Period 1* is applicable for apartment 2.

Table 6
Rules used to define benchmarking rules in apartment 2.

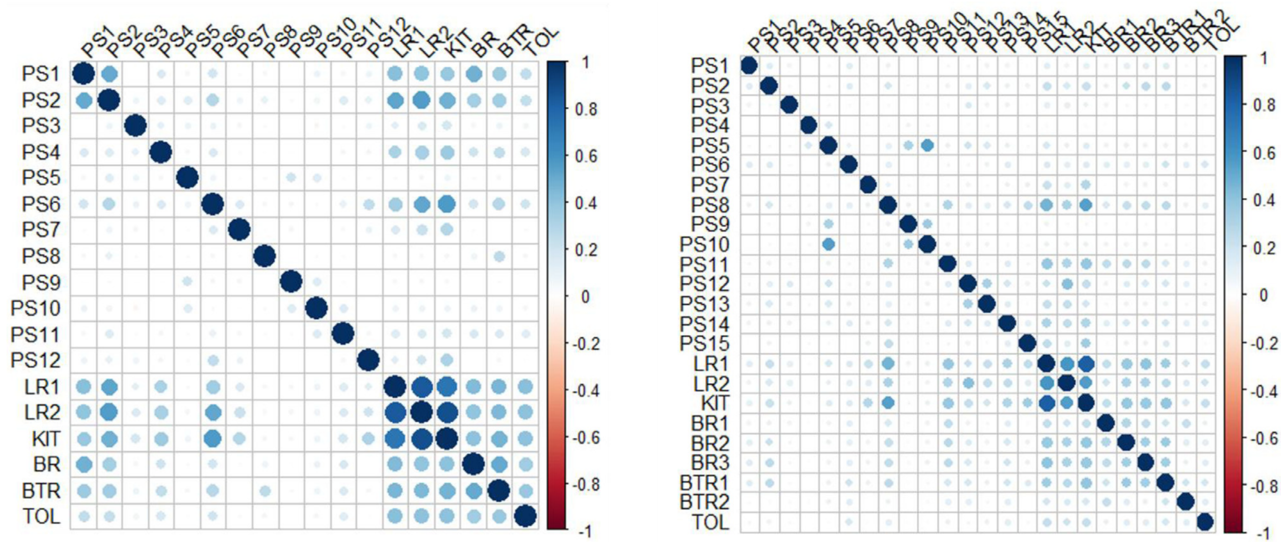
Rule No.	Period	Cluster	Premises	Conclusion	Sup.	Conf.	Lift	Count
1	1	L.OM-L.EC	Motion = [0, 5)	Energy = [0,125)	0.73	0.85	1.06	467
2	2	VL.OM-VL.EC	Motion = [0, 1)	Energy = [0,137)	0.90	0.99	1.07	184
3	3	VL.OM-VL.EC	Motion = [0, 1)	Energy = [0,125)	0.85	0.89	1.04	144
4	4	VL.OM-VL.EC	Motion = [0, 1)	Energy = [0,118)	0.77	0.85	1.09	395
5	5	VL.OM-VL.EC	Motion = [0, 1)	Energy = [0,125)	0.87	0.96	1.08	403
6	5	L.OM-L.EC	Hour = 01:00	Energy = [125, 286)	0.06	0.70	2.12	37

behavior rules appeared frequently which means occupants usually turned off unneeded appliances. Note that Rule 2 in Table 5 is different from other benchmarking rules which shows the low energy consumption has a strong positive relationship with no movement on Sunday. This is because in 'VL.OM-VL.EC' clusters, the majority of instances belongs to Sunday. Similarly, Rule 6 in Table 6 shows a strong positive relationship between time of day (i.e. Hour = 1:00h) and second-range energy consumption (i.e. Energy = [125, 286)), which indicates that energy consumption belongs to a relatively high level at 1:00 h. Such rule also indicates that the energy consumption pattern is not necessarily correlated with occupancy level which is consistent with CPA results (a frequent change point exists when the occupancy status keeps unchanged). According to the further analysis, there is a high chance that occupants in apartment 2 might use some appliances (e.g. dishwashers) before going to sleep. Thus, when assessing the energy saving potentials for apartment 2, such a rule is considered as occupants' normal behavior and needs to be differentiated from energy-inefficient behavior rules. This result indicates that the proposed framework is able to identify the difference in distinct energy-consuming behavior and habits among the selected apartments and differentiate them from abnormal behavior. Based on the abovementioned association rules, duration of 2 h was considered in this study as a threshold to affirm the minimum length of occupants' sleeping or away instances.

4.4. Identification of energy waste instances

4.4.1. Categorization of total energy consumption data

To facilitate further analysis of the energy waste patterns, individual power sensor (PS) data and motion sensor data collected in these two apartments were used to categorize the wastage instances discovered through the proposed framework. Considering that the specific appliance data is not available (see Section 3.1), it is necessary and reasonable to use detailed and actual occupancy data in each room to classify the unknown power sensors as the energy consumption is highly associated with occupant activities in the home. Accordingly, correlation analysis was carried out be-



(a) Apartment 1 (b) Apartment 2

Fig. 11. Visualization of the correlation matrix of between hourly energy consumption and hourly total no. of movements (Note: 'PS' in the figure is the abbreviation of power sensor).

Table 7
Summary of categories of all power sensors (PS) in the considered apartments.

Categories	PS in apartment 1	PS in apartment 2
LR-related	PS2	PS12, PS13
KIT-related	PS3, PS6, PS7, PS12	PS4, PS7, PS8, PS15
BR-related	PS1	PS1, PS2
BTR-related	PS8	PS6
Others	PS4, PS5, PS9, PS10, PS11	PS3, PS5, PS9, PS10, PS14, PS11

tween all power and occupancy sensors. Fig. 11 visualizes the correlation analysis results obtained by using hourly occupancy and energy consumption data. In Fig. 11, some power sensors show a strong correlation with one specific motion sensor, e.g. PS2 in apartment 1 is correlated with motion sensor in the living room (i.e. LR1 & LR2) and PS7 is correlated with motion sensor in the kitchen (i.e. KIT). Such relationships were further validated by their specific load features (e.g. working time, power range and use frequency). It can be also seen that some power sensors (such as PS9, PS10 in apartment 1 and PS4, PS9, PS10 in apartment 2) show no correlation with any occupancy motion sensor. This is because appliances connected to the sensors constantly consume energy. In addition, some show weak correlation with multiple motion sensors, which implies the use of appliances connected these sensors are less dependent on occupancy activities in one specific room. The possible reason is that power sensors are connected with different rooms to measure some particular appliances (usually high-intensity appliances) and have a relatively long working cycle (e.g. washing machine). All the categories based on the correlation analysis are summarized in Table 7.

4.4.2. Illustration of identified energy waste instances

The algorithm developed based on the benchmarking rules was applied to identify the energy wastes and assess the corresponding energy saving potentials in apartment 1 and 2. Based on the labeled PS data (as shown in Table 7) and separate lighting consumption data, the identified energy waste instances were grouped into six categories (i.e. lighting related, living room (LR) related, kitchen (KIT) related, bedroom (BR) related, bathroom (BTR) and others (OTH)). Fig. 12 presents the distribution of different categories of energy waste instances and Fig. 13 illustrates the different

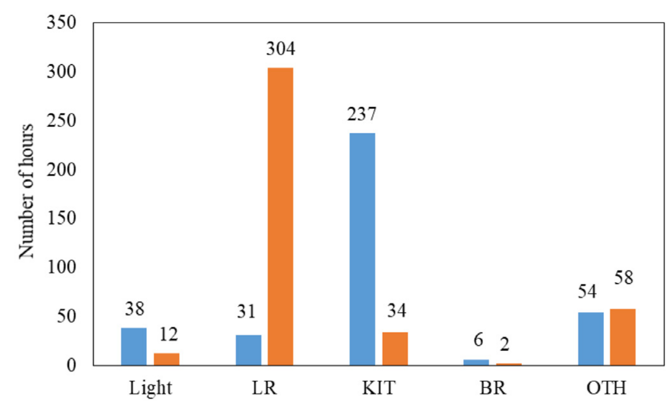


Fig. 12. Comparison of different categories of abnormal EC patterns in apartment 1& 2.

types of energy waste categories identified in apartment 1 and 2. For apartment 1, the inference from Fig. 12 is that approximately 70% of the time the energy wastes were found in the kitchen. Fig. 13(b) illustrates one example of energy waste instances associated with the refrigerator during a whole-day absence period in apartment 1. As seen in Fig. 13(b), the energy consumption pattern of the refrigerator (constantly keeps at a high level during the whole day) is abnormal implying that refrigerator did not work in a normal condition. One possible reason is that the fridge door was not closed properly and resulting in changes in the operation of the refrigerator from a cyclic operation to a constantly working condition which is further verified by follow-up interviews. As there is a high frequency to found wastes in kitchen appliances, it hints that the use of kitchen appliances deserves more attention for occupants in apartment 1 in practice. The further inference from Fig. 12 is that in apartment 1, the fraction of lighting waste exceeding 10% implies that occupants' wasteful behavior would result in an unexpected increase in lighting consumption even though the building was designed as energy-efficient (one goal is to reduce lighting consumption and make the best of natural lighting). One example of lighting waste in apartment 1 is shown in Fig. 13(a). As depicted in the figure, the lights consumed energy

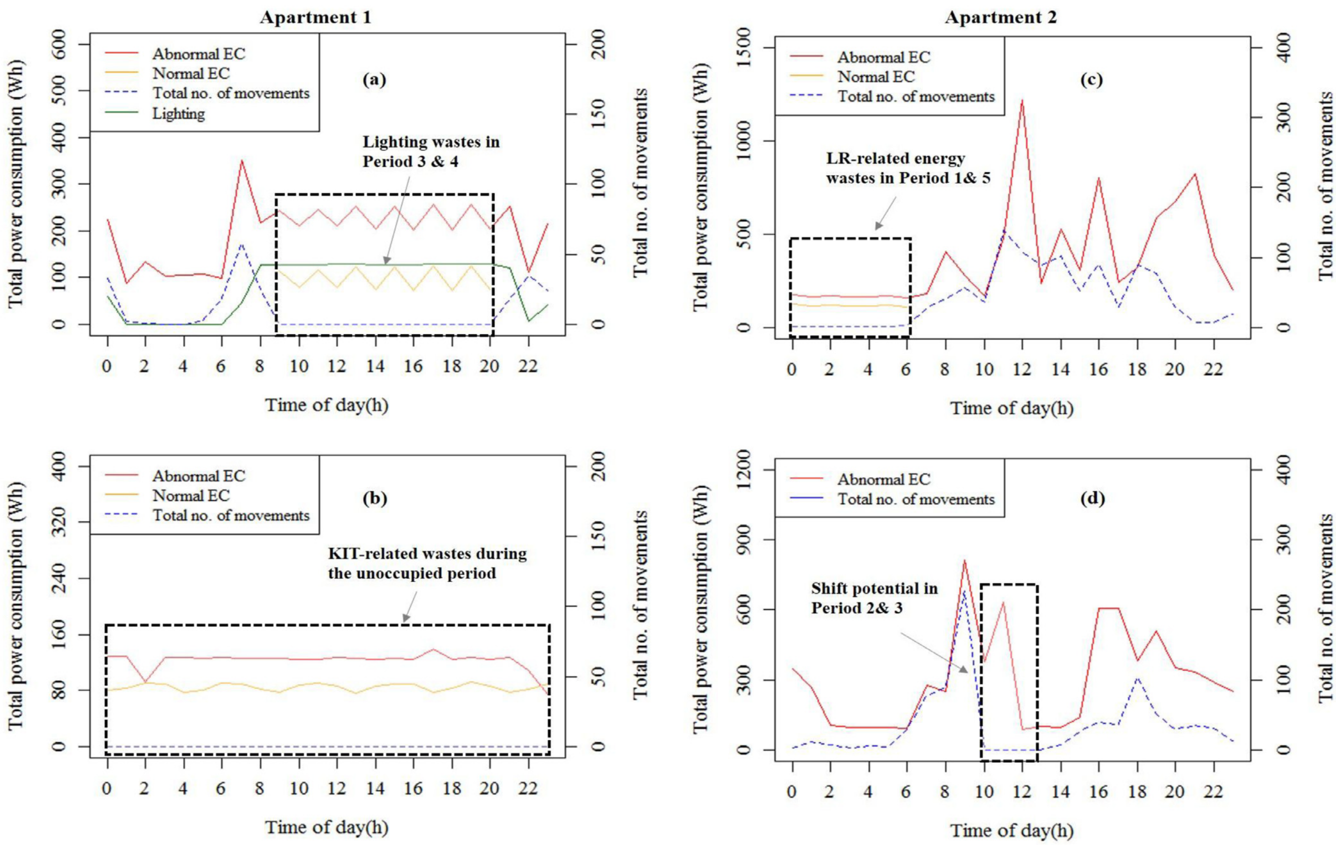


Fig. 13. Illustration of different types of energy waste patterns during a sleeping/away pattern event in apartment 1 & 2.

when the occupants were not at home for more than 10h. This means that occupants in apartment 1 wasted 1782 Wh energy during this period, which is 1.6 times higher compared with the benchmarking scenario of *Period 3 & 4*.

For apartment 2, a large proportion of energy wastes was associated with appliances in the living room (as shown in Fig. 12), which implies such waste occurs frequently. Fig. 13 (c) shows one example of energy waste instances in the living room in *Period 1* (when occupants were asleep). This energy waste is majorly caused by the standby mode power consumption associated with ICT appliances (e.g. TV, set-top box, Audio Hi-Fi) as the appliances in the living room might consume the active standby power (i.e. appliances were still on but did not perform its main function). The feedback about standby wastes is expected to help residents in apartment 2 to increase their energy-saving awareness and reduce standby energy consumption.

Additionally, these two apartments show a noteworthy proportion of energy saving potentials in the other (OTH) category. The abnormal EC patterns discovered in this category might be majorly due to the energy-intensity appliances (such as washing machines) left on when the occupant leaves the home (as illustrated in Fig. 13(d)). Although such instance might not be an actual energy waste, it can still be given as useful feedback information to the suggest occupants to shift such operation of such appliances from peak to off-peak hours (e.g. 22:00 h) for the purpose of cost savings when low electricity rate is applicable.

4.4.3. Summary of identified energy waste instances

Table 8 shows the statistical data of the duration of all identified abnormal EC patterns (i.e. energy waste instances) and corresponding energy saving potentials. As shown in Table 8, in both the apartments, there are more than 20% of days (65 days in apartment 1 and 109 days in apartment 2) that have energy waste instances which imply huge saving potentials associated with improving occupant behavior in both the apartments. However, substantial differences were found in terms of the duration of these instances and corresponding energy consumption. For example, the average energy saving potential in apartment 1 was about 389 Wh per day while only 194 Wh for apartment 2. Note that, though the number of days with energy saving potentials in apartment 1 is less than that of apartment 2, the total energy saving potential in apartment 1 is higher than that of apartment 2. A possible explanation behind this is that the occupants in apartment 1 have higher saving potentials with high-energy-consuming appliances while for apartment 2 the majority of energy waste was possibly correlated with the usage of small-appliances or standby mode power consumption. In addition, as discussed in previous sections, the occupancy and energy consumption patterns in each period show distinct characteristics. Therefore, energy wastes in different periods were examined and plotted as Fig. 14. It can be seen from the figure that in apartment 1, Period 3 accounts for the largest proportion (greater than 70%) and in Apartment 2, over 80% was found in Period 1.

5. Discussion

While specifying the energy waste instances, it is important to differentiate the actual energy wastes from the irregular energy consumption caused by different occupant activities. For instance, the occupants might use several appliances (such as dishwashers, washing machine, cooking appliances) simultaneously and the occupancy activity in the apartment might still be at a low level. In this case, such a type of energy use behavior was not frequently occurred in the considered apartments (i.e. irregular energy consumption patterns) and such energy consumption

Table 8
Summary of statistics with regard to the duration of abnormal EC patterns and associated saving potentials found in Apartment 1 and Apartment 2.

Building	No. of days	Duration of abnormal patterns(hours/day)			Energy saving potentials (Wh/day)		
		Min.	Avg.	Max.	Min.	Avg.	Max.
Apartment 1	65	2	5.2	22	48	389.3	1756.0
Apartment 2	109	2	3.0	7.0	39	173.7	913

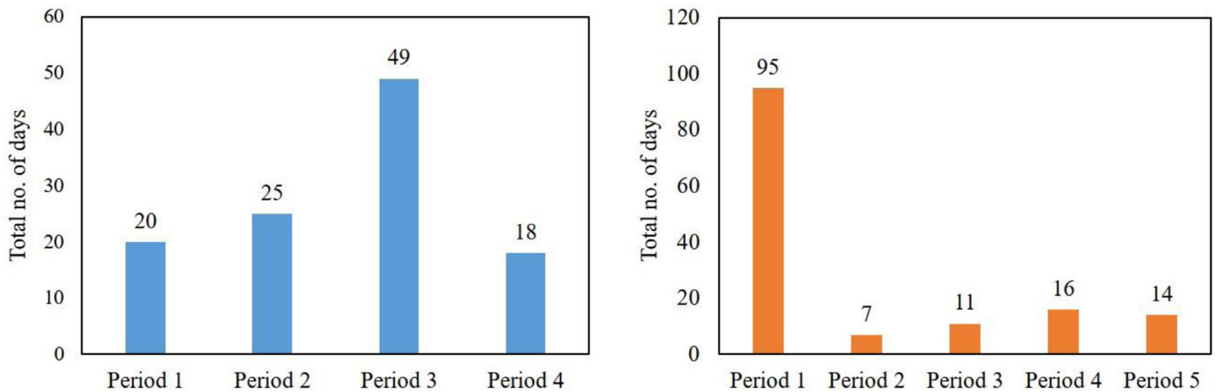


Fig. 14. Distribution of abnormal EC patterns found in each period in apartment 1 & 2.

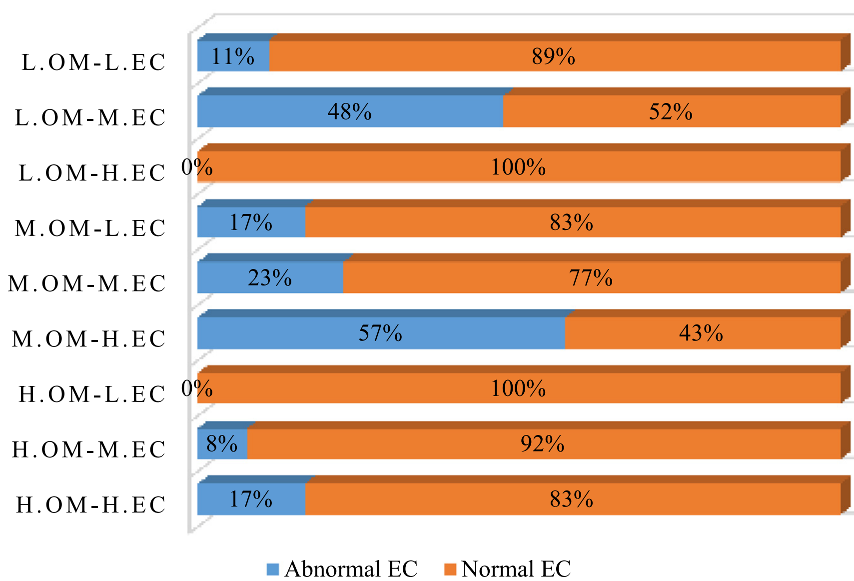


Fig. 15. Distribution of abnormal and normal EC patterns in all OB-EC sub-clusters for *Period 3*, apartment 1.

patterns cannot be considered as the energy waste. Hence, in this study, the differences in identifying abnormal EC patterns in matching and mismatching groups was further examined. Since higher percentages of energy waste were found in *Period 3* in apartment 1 and *Period 1* in apartment 2, these periods were used in this section for the analysis. Fig. 15 presents the distribution of normal and abnormal EC patterns in all OM-EC sub-clusters in *Period 3*, apartment 1. It can be seen from Fig. 15 that, abnormal EC patterns are identified in both matching and mismatching groups, however, there are significant differences in the possibility of discovering the abnormal EC patterns among them. Especially, the percentages of abnormal EC patterns found in the mismatchings groups i.e., 'L.OM-M.EC' and 'M.OM-H.EC' were 48% and 57%, respectively. The major cause for the energy wastages in these clusters are found to be lighting energy wastes and others (peak shift potential instances). These results indicate that there is a relatively higher probability of identifying the energy wastes in the mismatching patterns and the discovered wastes were associated with the occupants energy-inefficient behavior. As mentioned earlier in this section, it should be noted that the mismatching patterns do not always correspond with abnormal EC patterns and there are some cases that these mismatching patterns are normal patterns which are due to different occupant activities and use of high-intensity appliances (e.g. the instances in 'L.OM-H.EC'). For the matching groups, Fig. 15 shows that 11%, 17% and 23% were found in 'L.OM-L.EC', 'M.OM-M.EC', 'H.OM-H.EC' sub-clusters, respectively. This is because, in these sub-clusters, the majority of wastes were KIT-related wastes (especially the refrigerator). Note that some instances are also found in the living room and others in 'M.OM-M.EC' and 'H.OM-H.EC' sub-clusters. For example, occupants left the ICT appliances (e.g. TV) on when they were away from home in a relatively long time.

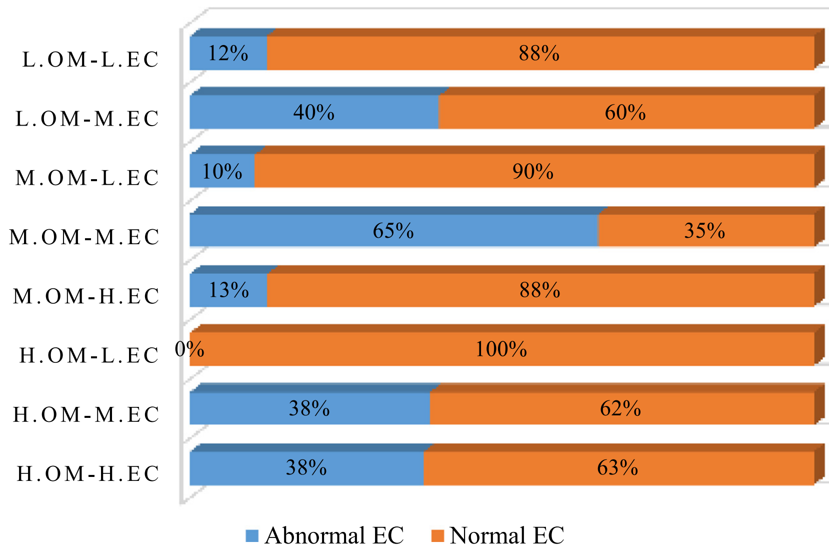


Fig. 16. Distribution of abnormal and normal EC patterns in all OM-EC sub-clusters for Period 1, apartment 2.

Similarly, Fig. 16 depicts the distribution of normal and abnormal EC patterns against all OM-EC sub-clusters in *Period 1*, apartment 2. It can be seen from Fig. 16, unlike the wastes found in apartment 1, a noticeable amount of wastes was found in the matching groups. For example, 65% and 12% of abnormal EC patterns were observed in 'M.OM-M.EC' and 'L.OM-L.EC' sub-clusters respectively. This is because most of the wastes found in this apartment were related to living room appliances and occupants in this room often left these appliances ON in the standby mode. In overall, it is construed that the energy wastage patterns can be identified in both matching and mismatching OM-EC sub-clusters. However, the probability of finding energy wastage instances are higher in the mismatching groups than the matching groups.

6. Practical considerations, limitations and future work

In this study, the authors explored the possibility of applying a DM-based framework to discover useful knowledge about occupancy behavior, energy consumption patterns and their relationship to assess household's energy use behavior and identifying associated energy-saving potentials. The results demonstrated the effectiveness and feasibility of the proposed framework in providing useful feedback information to the occupants about their energy waste behavior. However, there are some practical implementation considerations and limitations of this study.

First, high frequency and detailed occupancy (e.g. motion detection) and electricity data are used in this study, which might lead to privacy concerns as they can be used to infer occupants' activities and behavior. Hence, when implementing the framework in practice, there is a necessity to protect occupants' private information and thereby, security strategies such as data encryption and anonymization should be used. Also, the installation of motion sensors that cover each area in a home seems impractical and costly which may possibly hinder its real applications. As a result, effective methods to reasonably deploy motion sensors [34] or a combination with other low-cost detection techniques (e.g. Wi-Fi signals) can be used to address this issue. Considering the characteristics of occupancy-relevant data, investigations to residential buildings with other types of occupancy data are required to suggest more feasible recommendations for real-time applications (the combination of different sensor data). It should be noted that occupants would significantly benefit from the in-depth analysis of occupancy and energy consumption behavior in terms of not only electricity bill savings but also a better understanding of social and environmental impacts.

Second, the proposed framework was tested and validated in two different apartments of which the discovered energy saving potentials and benchmark values are useful for residents in the context of similar buildings and electricity consuming behavior. Nonetheless, the DM based framework proposed is systematic and hence the procedure adopted to identify the energy waste patterns could be also applied in wider building stocks, where the occupancy movement data are not available. For example, other sensors data such as CO_2 level indicator, thermostat data, Wi-Fi signal data, etc., can be used to discover the occupancy profile/routine and subsequently, the proposed DM framework can be adopted. However, further investigations with different sensor data and wider building stocks are required to explore the applicability and generalizability of the proposed framework.

Third, the energy wastes discovered in this study were obtained by analyzing the total energy consumption and some end-use data (which is labeled by motion data and domain knowledge). However, more specific end-use data (i.e. appliance-level) data would be beneficial to find more energy wastes. Thus, future studies would be carried out to use energy disaggregation techniques to obtain a more detailed and specific end-use load. Also, detailed occupancy related data such as the number of occupants, age, and profession of the households are unknown in this study. If such details are known, it is possible to provide more useful and tailor-made feedback to the occupants based on their lifestyles.

7. Conclusions

A systematic DM based framework was developed with the objective of discovering unusual energy use patterns in residential buildings by extracting the correlation between dynamic occupancy and building energy consumption data. The developed framework was applied to two datasets monitored through HEMS in two apartments from a high-performance building located in Lyon, France. Initially, CPA was performed to detect the distinct variations in occupancy movement, energy consumption patterns and accordingly, periods in a day for the specific household was determined. This step enabled the flexibility of determining unique number and duration of periods for specific household based on their daily routines rather than specifying a monotonous number and duration of periods for all the households. Then, occupancy, energy consumption data of respective periods are clustered separately and compared with each other to identify the matching and mismatching days. The clustering results explicitly demonstrated the association between occupancy patterns and energy consumption in each period. Also, it divided the dataset into different subclusters, which reduced the data size for ARM and subsequently the rule interpretation was made easy. Several normal behavior and interesting rules were obtained through ARM, which was used to define the benchmarking rules.

Through the proposed framework, different types of wastes were effectively identified for households with different lifestyles and everyday energy-consuming activities. For apartment 1, lighting usage and kitchen appliances deserve extra attention; for apartment 2, occupants need to improve their behavior with the usage of appliances in the living room (especially standby consumption). The results indicated that though the considered residential building was design to be highly energy-efficient, there still exists the considerable potential for energy savings by improving occupants' wasteful behavior. Meanwhile, it was found that a mismatching between occupancy and energy consumption hints a higher probability of finding energy wastes than that of matching groups which demonstrates the need of occupancy data while identifying energy wastes in residential buildings. Furthermore, the proposed framework can successfully differentiate high energy consumption associated with occupant's infrequently occurred activities from the abnormal EC pattern (i.e. wasteful behavior).

Though several DM based frameworks are available in the literature to interpret the energy consumption pattern in buildings, the framework proposed in this study is unique, since it systematically analyzes the building's energy use with actual and dynamic occupancy related data and explores specific knowledge about energy waste patterns in residential buildings. It can be provided to occupants through an online feedback system. Such an attempt will enable the occupants to understand their energy consumption patterns and will create awareness about their energy use and make them more responsible toward achieving energy savings, which is a tangential benefit for both energy suppliers and users. The developed framework is generic, flexible and it can be extended to a wide range of buildings irrespective of their household size, occupancy behavior and energy profiles. And it is also applicable to other types of buildings such as office buildings. Future studies will be carried out to improve the analysis framework by integrating the disaggregated individual energy consumption (e.g. appliance-specific) and by this mean; it is possible to identify more unusual energy use patterns explicitly.

Conflict of interest

The research work meets all applicable standards about the ethics of research integrity, and the following is being certified/declared true.

- The paper has been submitted with full responsibility, following due ethical procedure, and there is no duplicate publication, fraud or plagiarism.
- None of the authors of this paper has a financial or personal relationship with other people or organizations that could inappropriately influence or bias the content of the paper.
- It is to specifically state that "No Competing interests are at stake and there is No Conflict of Interest" with other people or organizations that could inappropriately influence or bias the content of the paper

Acknowledgments

The authors would like to express their gratitude to Concordia University for the support through the Concordia Research Chair – Energy & Environment and the Horizon Program.

References

[1] A.L. Pisello, F. Asdrubali, Human-based energy retrofits in residential buildings: a cost-effective alternative to traditional physical strategies, Appl. Energy 133 (2014) 224–235, doi:10.1016/J.APENERGY.2014.07.049.
[2] A. Francisco, J.E. Taylor, Designing community-scale energy feedback, Energy Procedia 158 (2019) 4178–4183, doi:10.1016/J.EGYPRO.2019.01.812.
[3] A. Nilsson, M. Wester, D. Lazarevic, N. Brandt, Smart homes, home energy management systems and real-time feedback: lessons for influencing household energy consumption from a Swedish field study, Energy Build. 179 (2018) 15–25, doi:10.1016/J.ENBUILD.2018.08.026.
[4] W. Abrahamse, L. Steg, C. Vlek, T. Rothengatter, The effect of tailored information, goal setting, and tailored feedback on household energy use, energy-related behaviors, and behavioral antecedents, J. Environ. Psychol. 27 (2007) 265–276, doi:10.1016/J.JENVP.2007.08.002.
[5] T. Ueno, F. Sano, O. Saeki, K. Tsuji, Effectiveness of an energy-consumption information system on energy savings in residential houses based on monitored data, Appl. Energy 83 (2006) 166–183, doi:10.1016/J.APENERGY.2005.02.002.
[6] G. Ma, J. Lin, N. Li, Longitudinal assessment of the behavior-changing effect of app-based eco-feedback in residential buildings, Energy Build. 159 (2018) 486–494. https://doi.org/10.1016/j.enbuild.2017.11.019.
[7] Z. Yu, B.C.M. Fung, F. Haghighat, H. Yoshino, E. Morofsky, A systematic procedure to study the influence of occupant behavior on building energy consumption, Energy Build. 43 (2011) 1409–1417, doi:10.1016/J.ENBUILD.2011.02.002.
[8] Zy). Yu, F. Haghighat, B.C.M. Fung, E. Morofsky, H. Yoshino, A methodology for identifying and improving occupant behavior in residential buildings, Energy 36 (2011) 6596–6608, doi:10.1016/J.ENERGY.2011.09.002.
[9] M. Ashouri, F. Haghighat, B.C.M. Fung, A. Lazrak, H. Yoshino, Development of building energy saving advisory: a data mining approach, Energy Build. 172 (2018) 139–151, doi:10.1016/J.ENBUILD.2018.04.052.
[10] M. Ashouri, F. Haghighat, B.C.M. Fung, H. Yoshino, Development of a ranking procedure for energy performance evaluation of buildings based on occupant behavior, Energy Build. 183 (2019) 659–671, doi:10.1016/J.ENBUILD.2018.11.050.
[11] J.M. Abreu, F. Câmara Pereira, P. Ferrão, Using pattern recognition to identify habitual behavior in residential electricity consumption, Energy Build. 49 (2012) 479–487, doi:10.1016/J.ENBUILD.2012.02.044.
[12] S. Ahmadi-Karvigh, B. Becerik-Gerber, L. Soibelman, A framework for allocating personalized appliance-level disaggregated electricity consumption to daily activities, Energy Build. 111 (2016) 337–350, doi:10.1016/J.ENBUILD.2015.11.029.
[13] S. Ahmadi-Karvigh, A. Ghahramani, B. Becerik-Gerber, L. Soibelman, Real-time activity recognition for energy efficiency in buildings, Appl. Energy 211 (2018) 146–160, doi:10.1016/J.APENERGY.2017.11.055.
[14] S. Naylor, M. Gillott, T. Lau, A review of occupant-centric building control strategies to reduce building energy use, Renew Sustain. Energy Rev. 96 (2018) 1–10, doi:10.1016/J.RSER.2018.07.019.
[15] P.F. Pereira, N.M.M. Ramos, Detection of occupant actions in buildings through change point analysis of in-situ measurements, Energy Build. 173 (2018) 365–377, doi:10.1016/J.ENBUILD.2018.05.050.
[16] C. Miller, F. Meggers, Mining electrical meter data to predict principal building use, performance class, and operations strategy for hundreds of non-residential buildings, Energy Build. 156 (2017) 360–373, doi:10.1016/J.ENBUILD.2017.09.056.
[17] S. Touzani, B. Ravache, E. Crowe, J. Granderson, Statistical change detection of building energy consumption: applications to savings estimation, Energy Build. 185 (2019) 123–136, doi:10.1016/J.ENBUILD.2018.12.020.
[18] R. Killick, I.A. Eckley, changepoint: an R package for changepoint analysis, J. Stat. Softw. 1 (Issue 3) (2014).
[19] G. Bode, T. Schreiber, M. Baranski, D. Müller, A time series clustering approach for building automation and control systems, Appl. Energy 238 (2019) 1337–1345, doi:10.1016/J.APENERGY.2019.01.196.
[20] J.Y. Park, X. Yang, C. Miller, P. Arjunan, Z. Nagy, Apples or oranges? Identification of fundamental load shape profiles for benchmarking buildings using a large and diverse dataset, Appl. Energy 236 (2019) 1280–1295, doi:10.1016/J.APENERGY.2018.12.025.
[21] P. Huang, Y. Sun, A clustering based grouping method of nearly zero energy buildings for performance improvements, Appl. Energy 235 (2019) 43–55, doi:10.1016/J.APENERGY.2018.10.116.
[22] C. Miller, Z. Nagy, A. Schlueter, A review of unsupervised statistical learning and visual analytics techniques applied to performance analysis of non-residential buildings, Renew Sustain. Energy Rev. 81 (2018) 1365–1377, doi:10.1016/J.RSER.2017.05.124.
[23] S. D'Oca, T. Hong, A data-mining approach to discover patterns of window opening and closing behavior in offices, Build. Environ. 82 (2014) 726–739, doi:10.1016/J.BUILDENV.2014.10.021.
[24] J. Han, J. Pei, M. Kamber, Data Mining: Concepts and Techniques, Elsevier, 2011.
[25] D.F. Motta Cabrera, H. Zareipour, Data association mining for identifying lighting energy waste patterns in educational institutes, Energy Build. 62 (2013) 210–216, doi:10.1016/J.ENBUILD.2013.02.049.

[26] M. Pothitou, R.F. Hanna, K.J. Chalvatzis, ICT entertainment appliances' impact on domestic electricity consumption, Renew Sustain. Energy Rev. 69 (2017) 843–853, doi:10.1016/J.RSER.2016.11.100.
[27] M. Coleman, N. Brown, A. Wright, S.K. Firth, Information, communication and entertainment appliance use—Insights from a UK household study, Energy Build. 54 (2012) 61–72, doi:10.1016/J.ENBUILD.2012.06.008.
[28] F. Xiao, C. Fan, Data mining in building automation system for improving building operational performance, Energy Build. 75 (2014) 109–118, doi:10.1016/J.ENBUILD.2014.02.005.
[29] Zy. Yu, F. Haghighat, B.C.M. Fung, L. Zhou, A novel methodology for knowledge discovery through mining associations between building operational data, Energy Build. 47 (2012) 430–440, doi:10.1016/J.ENBUILD.2011.12.018.
[30] C. Fan, F. Xiao, H. Madsen, D. Wang, Temporal knowledge discovery in big BAS data for building energy management, Energy Build. 109 (2015) 75–89, doi:10.1016/J.ENBUILD.2015.09.060.
[31] C. Fan, F. Xiao, C. Yan, A framework for knowledge discovery in massive building automation data and its application in building diagnostics, Autom. Constr. 50 (2015) 81–90, doi:10.1016/J.AUTCON.2014.12.006.
[32] Z. Yu, B.C.M. Fung, F. Haghighat, Extracting knowledge from building-related data – a data mining framework, Build. Simul. 6 (2013) 207–222, doi:10.1007/s12273-013-0117-8.
[33] Rapid-miner. Rapid 2012.
[34] E. Soltanaghaei, K. Whitehouse, Practical occupancy detection for programmable and smart thermostats, Appl. Energy 220 (2018) 842–855, doi:10.1016/J.APENERGY.2017.11.024.