Systematic Approach to Provide Building Occupants with Feedback to Reduce Energy Consumption

Milad Ashouri[a], Benjamin C. M. Fung[b], Fariborz Haghighat[a*], Hiroshi Yoshino[c]
[a] Department of Building, Civil and Environmental Engineering, Concordia University, Montreal, Quebec, Canada H3G 1M8
[b] School of Information Studies, McGill University, Montreal, Quebec, Canada H3A 1X1
[c] Department of Architecture and Building Science, Tohoku University, Japan
Corresponding author: haghi@bcee.concordia.ca

Official publication: https://doi.org/10.1016/j.energy.2019.116813

Abstract

Many technical solutions have been developed to reduce buildings' energy consumption, but limited efforts have been made to adequately address the role or action of building occupants in this process. Our earlier investigations have shown that occupants play a significant role in buildings' energy consumption: It was shown that savings of up to 20% could be achieved by modifying occupant behavior thorough direct feedback and recommendations. Studying the role of occupants in building energy consumption requires an understanding of the interrelationships between climatic conditions; building characteristics; and building services and operation. This paper describes the development of a systematic procedure to provide building occupants with direct feedback and recommendations to help them take appropriate action to reduce building energy consumption. The procedure is geared toward developing a *Reference Building* (*RB*) (an energy-efficient building) for a specific given building. The RB is then compared against its given building to inform the occupants of the given building how they are using end-use loads and how they can improve them. The *RB* is generated using a data-mining approach, which involves clustering analysis and neural networks. The framework is based on clustering similar buildings by effects unrelated to occupant behavior. The buildings are then grouped based on their energy consumption, and those with lower consumption are combined to generate the *RB*. Performance evaluation is determined by comparison of a given building with an *RB*. This comparison provides feedback that can lead occupants to take appropriate measures (e.g., turning off unnecessary lights or heating, ventilation, and air conditioning (HVAC), etc.) to improve building energy performance. More accurate, scalable, and realistic results are achiveable through current methodology which is shown through comparison with existing literature.

Keywords: energy use evaluation; building energy management; data mining; occupant behavior

1. Introduction

1.1 Background and Motivation

Occupants' contributions to the energy consumption of buildings could be significant. By monitoring occupants' behavior in this regard, we can find opportunities to save energy. This is important in the sense that modifying occupant behavior is an inexpensive way to reduce building energy consumption, especially if the occupants could benefit financially (reducing energy bills) and without any additional costs. Occupant behavior means occupants' presence, activities, and operations. One of the best ways to assess this behavior is through comparison. Climate conditions, building envelope, and building systems can influence occupant behavior. However, if two buildings with the same types of factors are studied, we may notice, for example, that one of the buildings has lower energy expenditure toward a certain end-use load. This shows that some occupants are aware of their activities or behavior, which could help reduce energy consumption. For example, the occupants of one building may use little energy for kitchen appliances, the operation of heating, ventilation, and air conditioning (HVAC), or both. This can be due to several reasons, such as culture or lifestyle. On the other hand, occupants may overuse some appliances, which makes them high-energy consumers. However, simply choosing low consumer buildings regarding one specific end-use load (by methods such as k-nearest neighbors or similar clustering methods) and comparing it with a given building may not give us accurate results knowing that two buildings may not be exactly the same regarding climate conditions, physical parameters, and number of occupants. Based on this, it is possible to create a digital twin of the building that has all the positive characteristics of low-consumption buildings, so we can alert the occupants of the building in question of how much energy they can save by copying the *Reference Building* (*RB*). The idea is shown in Figure 1. As expected, buildings have different occupants and characteristics, such as the number of people, age, the level of activities, building thermal characteristics such as wall area, insulations, etc., and weather conditions such as cold and humid, or warm and dry climates. This is shown in building 1 to building n (Figure 1). Every building has its own occupants, climatic conditions, and building characteristics and consequently its own appliances and end-use loads. This is shown in bar charts in red and blue in Figure 1 (red means the occupants show a higher energy use regarding those specific appliances and blue means less energy use). Therefore, one cannot simply compare the high-energy-consuming buildings with those with the lowest energy consumption. The main objective of the current project is to generate an *RB* so that any other building can be compared with it. The *RB* should be very similar to the building in question in terms of number of occupants, building characteristics, and weather conditions. In the previous studies, occupant behavior was analyzed using data mining of several existing buildings to reveal opportunities to save energy [1-3]. In this study, the development of a novel methodology to generate an *RB* from data on existing buildings for performance evaluation is described. The result is a generic, accurate, and scalable tool to evaluate the energy consumption of a given building. In the following sections, an overview of previous works is presented before the current work methodology is introduced.

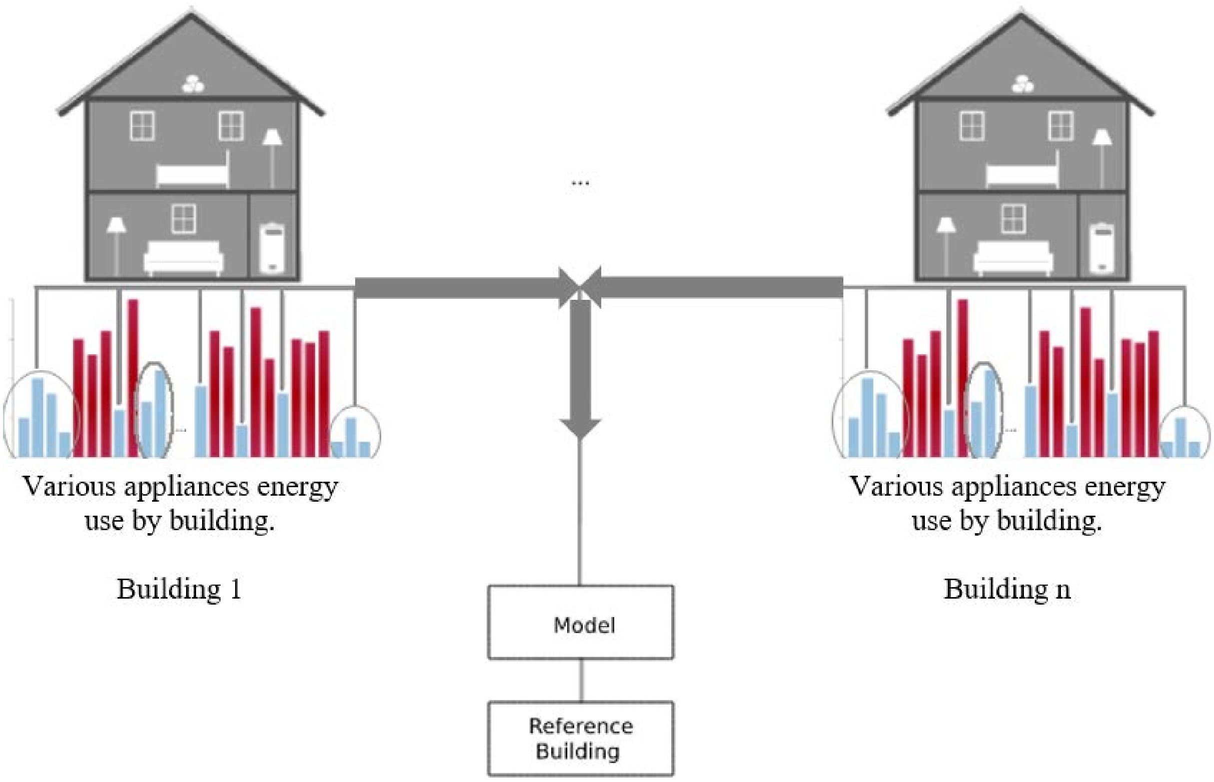

Figure 1 Creation of reference building through analysis of appliances energy consumption. Each building has its own end-use loads energy usage, some are high, and some are low (shown in red and blue bars). The idea is to make a model which incorporates all low energy use patterns of the buildings.

1.2 State of the art

Occupant behaviour and its great role in energy consumption has been widely accepted by researches such that it has been identified as the key to bridge the gap between simulated and actual building energy consumption [1]. Together with recent technological improvements in building industry such as design and operation of building systems [2], a vast number of studies are seen in the literature regarding occupant behaviour. Two factors are common among the works. The first factor is related to stochastic and complex nature of occupant behaviour and the second factor lies in interrelation between other influencing parameters in building energy consumption such as climate, building materials and characteristics, and economics. These two factors make it difficult to model the exact occupant behaviour. With abundance of data in building through sensors (to measure human interactions such as movement. CO2 detection, windows, blinds), Wi-Fi signals, control system (for HVAC and lighting), in-depth analysis of occupant behaviour has been enabled [3].

As a solution, researchers moved to data analytics to explore occupant related actions. Cluster analysis is a well-known tool for dividing data into subsets with similar characteristics regarding variables. Fan et al. [4] determined the typical operation patterns of a building cooling system by clustering the data on energy consumption over a year. The results suggested two power consumption patterns: one for weekends and one for weekdays. In a similar project Xiao and Fan [5] suggested three patterns using the same technique. Using clustering, association rule mining and regression analysis, D'Oca and Hong [6] investigated the patterns of window opening and closing as an important factor in energy consumption in sixteen office buildings. For example, patterns of window opening duration during the day (long, medium, or short intervals) affect infiltration and, consequently, energy consumption. The mentioned unsupervised data analytics were sufficient to highlight the behavioral patterns seen in the buildings.

In order to isolate the effect of influential factors in energy consumption, do Carmo and Christensen [7] clustered hourly heating load data of 139 single-family detached houses into three separate groups. The goal was to eliminate the effect of weather conditions to isolate the effect of household and building characteristics on the thermal load demand. It also helped to find daily routines for each household. These data can be compared to best, worst, or "normal" consumption data to study occupants' impact. Abreu et al. [8] developed a framework to extract the daily routines of households and patterns of energy consumption over a year. They identified recurrent behaviors during the day (daily routines of households) by applying the PCA to the daily electricity energy consumption data. Moreover, Abreu et al. [8] extracted patterns of energy consumption using clustering as the unoccupied period "baseline," cold weekend days, cold working days, and hot and temperate working days.

In order to isolate the effects of factors unrelated to occupant behavior, Yu et al. [9] applied k-means clustering based on four factors unrelated to occupant behavior, i.e., climate, building physical parameters, number of occupants, and building systems. The difference between energy consumption of buildings in the same cluster were referred to different occupant behavior and actions. Other examples of pattern discovery using clustering are occupancy schedule [10], and indoor air quality [11].

Pattern discovery also helps to optimize the operation of HVAC and reduce the unnecessary loads. Capozzoli et al. [12] used occupancy pattern analysis in order to reduce the HVAC energy consumption. By grouping similar occupancy patterns in the same thermal zone, they changed the HVAC control strategy to reduce the load while keeping the thermal comfort. Wi-Fi signals gives us a lot of information about the occupancy patterns, and how occupants interact with the building. Wang and Shao [13] used Wi-Fi signals to find occupancy patterns and used association rule mining to find energy wastage instances. It was claimed that up to 26% of lighting energy could be saved through this method. Kastner et al. [14] proposed a web-based intervention plan to encourage occupants to more energy efficient behavior through giving them recommendations (either habit-based intervention or knowledge-based intervention) and recording their actions. Field studies show that cultural differences of different sites influence implementation levels. In another study, meinke et al. [15] used feed forward information to give the occupants information about their choice of cooling strategy (removing shirt, turning on ceiling fan, turning on air conditioner or tilting the windows) and how they impact energy use and their level of comfort. The result showed that most of the participants changed their action after getting the related information. This shows the effectiveness of giving occupants information and recommendation on their behavior. Knowing the effectivity of feedback-based systems on household electricity use, Fischer [16] developed a psychological model to evaluate which features of feedback works best for the occupants. The feedback features are frequency, duration, content, breakdown, medium and way of presentation, comparisons, and combinations with other instruments. The results indicate that an effective feedback should be frequent, over a long time, be appliance specific breakdown, and be presented in an appealing way using computerized and interactive tools.

Association rule mining, as an unsupervised learning technique, was used by Yu et al. [17] to find the relationship between the power consumption of different appliances (dishwasher, microwave, TV, etc.) and climatic parameters (temperature, relative humidity, etc.) to provide occupants with recommendations on how to reduce the appliance usage and to find energy inefficient behaviors.

The two factors mentioned in the beginning of the section 1.2 show the necessity of in-depth occupant behavior analysis to reveal new opportunities toward energy efficient buildings while maintaining the comfort. In this study, a new approach for evaluation of energy consumption in a single dwelling is presented, using clustering analysis and ANN models. The following is a summary of the contributions of the current work:

1. The current paper is the first featuring a proposal to create a nonexistent *RB* to assess the energy consumption of a given building. Knowing that each building may show a low energy consumption pattern regarding one specific end-use load such as HVAC or lighting, it is possible to make an *RB* that contains all energy-saving characteristics of different buildings. Therefore, for a given building, its *RB* is created by the proposed methodology (Section 2). Potential savings are revealed through comparison of the given building with the *RB*.
2. The creation of an *RB* through this methodology is an accurate approach to assess the real performance of a building. (A comparison of existing work with another authors' publication is described in section 3.4.)
3. The methodology introduced here is a generic and scalable approach, meaning that by increasing the number of buildings under investigation, we can achieve even better and more robust models as *RB*s to assess the performance of a given building.

2. Methodology

Figure 2 shows the overview of the framework. The tasks are outlined in the framework: 1) data selection; 2), data aggregation; outlier detection and diagnosis; and normalization; 3) database creation; 4) grey relational analysis (GRA); 5) level 1-1 clustering; 6) level 1-2 clustering; 7) level 2 clustering; 8) cluster ranking and combination; and 9) model development. Each task is meant for a purpose and is applied to a portion of the dataset. The designed process is as follows:

1) Data on energy consumption from each building are selected, summarized, cleaned, and integrated together. The data consist of weather parameters, building characteristics, and energy consumption of all end-use loads. Energy consumption data are averaged to annual values and normalized and stored in the database. This methodology uses the annual building energy consumption, but it can be further improved to show seasonal energy consumption.
2) GRA is performed on the database to give weight to input parameters. The contribution of each parameter on a specific end-use load may be different. GRA is performed for each specific load separately for use in generating the *RB*.
3) A two-level clustering is performed on the dataset to group buildings with similar characteristics in terms of climatic and physical and occupant information, as shown in Figure 2. This step is called level one clustering.
4) A second clustering is performed on each resultant dataset to place them into groups of low- and high-energy consumers.
5) Clusters containing buildings with low-energy consumer tags are combined to develop the dataset of low-energy consumers. Note that steps 3–5 are performed for each end-use load separately.
6) A model is developed using the low-energy consumers. The inputs to the model are eleven variables which consist of building characteristics, climatic characteristics, and occupant information. The output is the energy use intensity of the end-use load.

During the process, clustering is used for different tasks. The first two put the building in similar climatic and physical characteristics, while the third is clustering based on energy consumptions to distinguish the high energy consumers from low energy consumers enabling us to build our

model based on low consumer buildings. For each specific load, this process is repeated meaning that for one end-use load (for example kitchen appliances) the low energy consumers are different than another end-use load (for example HVAC energy consumption).
The following sections describe each step in more detail.

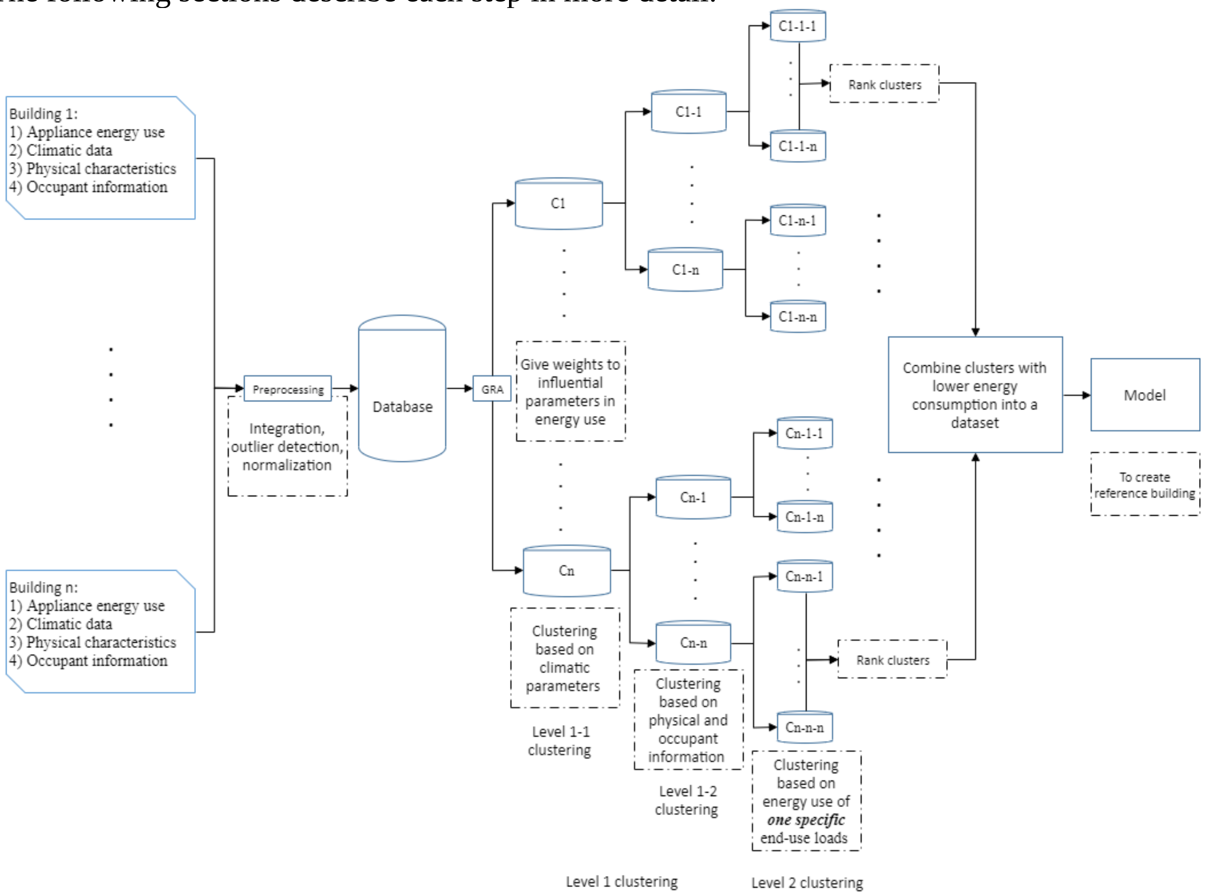

Figure 2. The proposed data mining framework. The inputs are home appliance energy use, weather, physical data, and occupant information. The output is the model to create an *RB*.

2.1 Data Selection

As part of energy efficiency program in Japanese residential buildings, a project entitled "Investigation on energy consumption of residents all over Japan" was performed by the Architecture Institute of Japan from 2002 to 2004. Field surveys regarding energy related data and environmental data were carried out on 80 buildings (76 building were selected for this work) located in six diverse regions in Japan to cover different parts of the country with all kinds of climatic conditions: Hokkaido, Tohoku, Hokuriko, Kanto, Kansai, and Kyushu. Table 1 and Figure 3 show the survey items and method along with the measuring tools to monitor temperature, gas, electricity, and kerosene. The available data were daily, so we developed the methodology on a daily basis. It is important to mention that all energy consumption data were converted to joule unit for comparison purpose.

Table 1. Data gathering methods and items [18].

Method	Items	Measurement interval
Field Measurements	• Energy data (Electricity, Gas, and Kerosene) • Climatic data (e.g. indoor air temperature, humidity, wind speed, etc. 1.1m above ground)	• Daily • Hourly (original data resolution: 15 minutes)
Questionnaire	Number of occupants, equipment uses.	—
Inquiring survey	Building characteristics (building types, area, heat loss coefficient, equivalent leakage area, type of heating and hot water and kitchen equipment)	—

Figure 3. Measurement tools [19].

The data containing the information about home appliance energy use, weather data, building characteristics, and occupants' information were input. Home appliances' energy use were summed up in eight categories as follows:

1) HVAC
2) Hot water supply (HWS)
3) Lighting (LIGHT)
4) Kitchen (KITCH)
5) Refrigerator (FRIDGE)
6) Entertainment and information (E&I)
7) Housework and sanitary (H&S)
8) Other end-use loads (OTHER)

The energy consumption of each of these end-use loads was monitored and measured every day for each building. There are 76 buildings in total, and the study was performed between 2002 and 2004. Therefore, approximately $3 \times 365 \times 76 = 83220$ datapoints were collected. Climate data included:

1) Annual average outside air temperature (T)
2) Annual average relative humidity (RH)
3) Annual mean wind speed (WS)
4) Annual mean global solar radiation (IR)

In some cases, where the values were missing, the nearest weather station's data were used. Building physical parameters were:

5) House types (HT) (detached or apartment)
6) Building area (A) (m^2)
7) Equivalent leakage area (ELA) (cm^2/m^2)
8) Type of space heating and cooling (HC) (electric or nonelectric)
9) Type of hot water supply (hws) (electric or non-electric)
10) Kitchen equipment (KE) (electric or nonelectric)

11) Heat loss coefficient (HLC) (W/m^3K)

Occupant information included only:

12) Number of occupants (NO)

2.2 Data Aggregation, Outlier Detection, and Normalization

The data from all buildings were aggregated, and outlier detection was performed using the lower and upper quantile method. This means that data lower than $\mathrm{Q1} - 1.5(\mathrm{Q3} - \mathrm{Q1})$ or greater than $\mathrm{Q3} + 1.5(\mathrm{Q3} - \mathrm{Q1})$ were considered as outliers, where Q1 and Q3 were the first and third quantiles, respectively. After outlier removal, they were filled using a regression model based on other available attributes. Data were then normalized using min–max normalization to obtain the values between 0 and 1. Furthermore, the categorical values were converted to 0, 1, or in between.

2.3 Database Development

Once preprocessing was completed, the database was developed for the DM tasks. For each task, a subset of the attributes was selected. For example, in level 1-1 clustering, climate data were used to cluster the buildings into similar groups, and other attributes were kept, while in level 1-2 clustering, the building's physical information was used. Finally, in level 2 clustering, the energy consumption data of a specific end-use load were used.

2.4 Grey Relational Analysis

GRA identifies the causative factors of an objective (energy use, here) and weighs them according to their contribution[20]. The contribution of each of these factors in the overall energy consumption of buildings varies greatly. For example, a variation in the number of occupants may have larger effects on the KITCH energy consumption than would variation in the house type. In this study, GRA was applied to the following variables: temperature, relative humidity, solar radiation, wind speed, house type, equivalent leakage area, heat loss coefficient, number of occupants, heating/cooling type, hot water supply, and kitchen equipment. The GRA results were then multiplied by their corresponding variable. The process was repeated for each end-use load. Results were stored in the database for the next steps.

2.5 Level 1-1 Clustering

This step filters out the effects of weather parameters by clustering buildings with similar weather conditions together. The attributes used for this clustering are annual mean air temperature, annual mean relative humidity, annual mean wind speed, and annual mean global solar radiation. The output of this clustering is groups of buildings with similar weather parameters. It is worth mentioning that it is possible to prioritize the attributes used in this clustering, for example, making the outside temperature the main attribute by giving it a weight. However, in this study, all four climatic attributes were given the same importance (weight) and used in clustering.

2.6 Level 1-2 Clustering

At this stage, the buildings are grouped by occupant information (number of occupants) and building physical characteristics (house type, equivalent leakage area, heat loss coefficient, heating/cooling type, hot water supply, and kitchen equipment). The result is the grouping of buildings with similar occupants and physical parameters.

2.7 Level 2 Clustering

The differences between energy consumption of buildings in level 2 clustering are due to different occupant behaviour. At this level, the buildings are clustered in terms of energy use intensity (EUI, energy use per unit of area), which is an indicator of the energy consumption of the occupants. The steps in sections 2.4–2.7 are repeated for all eight end-use loads (mentioned in section 2.1) separately and stored in the database.

2.8 Cluster Ranking and Combination

The result of level 2 clustering is the grouping of buildings with similar energy use regarding one specific end-use load (for example, HVAC energy consumption). The centroids of clusters (average energy use of each cluster) are then sorted from the highest to the lowest: the cluster with the lowest centroid represents building occupants who are cautious regarding energy use of the end-use load under investigation. By extracting all low consumer buildings and their corresponding information (listed in level 1-1 and 1-2 clustering), we were able to make a database, which was a subset of the original database but contained only low-consumption buildings, as shown in Figure 2. This process was performed for all eight end-use loads listed in section 2.1. The resulting data lay the foundation to create the *RB*.

2.9 Model Development

The aim of this step is to use the developed databases as the base to generate the *RB*. As mentioned before, all buildings in the created datasets were low-consumption buildings. The goal was to develop a model that could estimate the energy consumption of an end-use load. The model inputs were number of occupants, climatic conditions, and building physical information. The model took the inputs and estimated how much energy the building would approximately consume if it belonged to the set of low-energy-consumption buildings. The model used an artificial neural network to map inputs (characteristics) to outputs (energy consumption). This process was performed for all eight end-use loads, and a model was created for each end-use load. Therefore, the *RB* consisted of eight ANN models. The outputs of the *RB* were in the range of low-consumption buildings because the databases were created using such buildings. After comparing the model outputs and real values from a given building, we could evaluate said building's thermal performance. This is shown in Figure 4. It is worth mentioning that the *RB* was customized to each given building, considering that the inputs of the models came from the given building and those developing the *RB* estimated the desired values from the low-consumption buildings using an ANN. This means that the reference building is unique to each given building and changes dynamically with respect to any given building. In other words, for two different given buildings two different reference buildings are created using the methodology described in section 2.

The ANN model used in this study was a multilayer perception model with regularization parameters. The Scikit-learn package of Python 3.7.3 was used for the analysis, and the optimal parameters were chosen according to the grid-search method available with the package scikit-learn [21]. A three-layer neural network was used in all cases, and the activation function, regularization parameter, and learning rate method were chosen by the grid search.

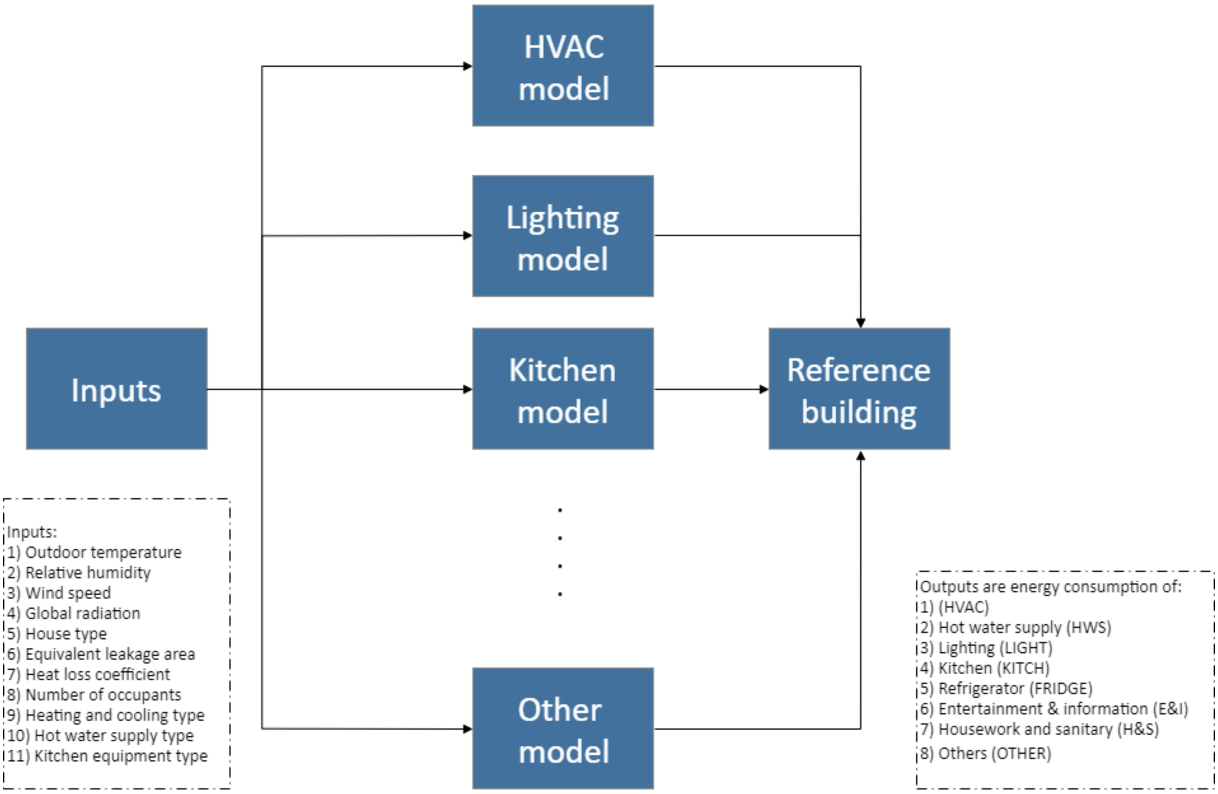

Figure 4. Creation of reference building using eight ANN models.

3. Results and Discussion

3.1 Grey Relational Analysis

Table 2 shows the result of GRA for every end-use load. As can be seen, for every load, the contributions of characteristics are different. The colors in each column sort the values from highest to lowest. Pure red shows the maximum, pure green shows the minimum, and others are in between. For example, in HVAC energy consumption, NO, followed by T, are the key variables, while HT is the least important characteristic. This shows that the type of dwelling (detached or apartment) has the least effect on HVAC energy consumption by occupants. Regarding HWS, RH and NO are the dominant factors, which seems reasonable, while KE is the least important characteristic. In general, the important characteristics are NO, HLC, T, and IR, while HT and KE are considered less important characteristics. Weather factors and NO play a greater role than building characteristics. The contributions of each characteristic were multiplied by the normalized energy consumption data for clustering purposes (next section).

Table 2 Results of GRA for Each End-Use Load.

	HVAC[l]	HWS[m]	LIGHT[n]	KITCH[o]	FRIDGE[p]	E&I[q]	H&S[r]	OTHER[s]	
HT[a]	0.625	0.653	0.637	0.562	0.614	0.621	0.588	0.707	
NO[b]	0.778	0.726	0.732	0.765	0.812	0.777	0.763	0.698	
HLC[c]	0.77	0.708	0.697	0.739	0.802	0.751	0.779	0.679	green: max
ELA[d]	0.734	0.696	0.694	0.706	0.723	0.696	0.74	0.702	red: min
HC[e]	0.688	0.683	0.645	0.641	0.699	0.704	0.673	0.73	
hws[f]	0.744	0.639	0.707	0.649	0.751	0.712	0.725	0.696	
KE[g]	0.667	0.605	0.645	0.546	0.651	0.658	0.642	0.691	
T[h]	0.773	0.688	0.721	0.709	0.783	0.747	0.753	0.671	
RH[i]	0.722	0.734	0.668	0.66	0.693	0.703	0.704	0.767	
WS[j]	0.753	0.668	0.707	0.702	0.766	0.699	0.768	0.677	
IR[k]	0.768	0.684	0.707	0.738	0.806	0.766	0.784	0.66	
max	0.778	0.734	0.732	0.765	0.812	0.777	0.784	0.767	
min	0.625	0.605	0.637	0.546	0.614	0.621	0.588	0.66	

[a] house type (detached or apartment)
[b] number of occupants
[c] heat loss coefficient (W/m^{3}K)
[d] equivalent leakage area (cm^2/m^2)
[e] heating/cooling type (electric vs non-electric)
[f] hot water supply type (electric vs non-electric)
[g] kitchen equipment (electric vs non-electric)
[h] annual average outside air temperature (C)
[i] annual average relative humidity (%)
[j] annual average wind speed (m/s)
[k] annual mean global solar radiation
[l] HVAC energy consumption
[m] hot water supply energy consumption
[n] lighting energy consumption
[o] kitchen appliances energy consumption
[p] refrigerator energy consumption
[q] entertainment and information energy consumption
[r] household and sanitary related energy consumption
[s] other energy consumption

3.2 Clustering Results

Figure 5 shows the result of data clustering regarding HVAC energy consumption; Df represents the original dataset after applying GRA and Df1, Df2, etc. represent the sub-datasets after performing clustering in different levels. The number of clusters was determined using the silhouette index [22] which was two in this case. Df1 and Df2 are the output of level 1-1 clustering regarding climatic conditions of the buildings. This means that all buildings in Df1 and Df2 have similar weather parameters. Cluster Df1 contains buildings in warmer regions considering that in their centroid, temperature and solar irradiation have higher average values. Clustering level 1-2 divides each cluster further into groups of buildings with similar physical characteristics and NO (Df1-1 to Df1-2), as shown in Figure 5. Df1-1 contains buildings with a nonelectric HWS and KE, whereas Df1-2 is all electric. The details of clustering are shown in Figure 5. The last clustering step is level 2, which partitions data based on target end-use load (HVAC energy consumption here). This last clustering step is performed based on the end-use load. As seen, the high-energy-consumption buildings have normalized HVAC consumption of more than 0.60 in all clusters, while their low-consumption counterparts use less than 0.30 in all clusters. Through comparing cluster centroids of level 2 clustering, the clusters with lowest energy consumption were selected and combined. This dataset made the “low-consumption buildings,” which are shown in Figure 5 in the case of HVAC energy consumption. It is important to note that clustering level 2 may end up with more than two clusters (here, we ended up with two clusters, so we named them low- and high-energy-consumption buildings). The buildings belonging to lower consumption groups are chosen and combined regardless of the number of clusters.

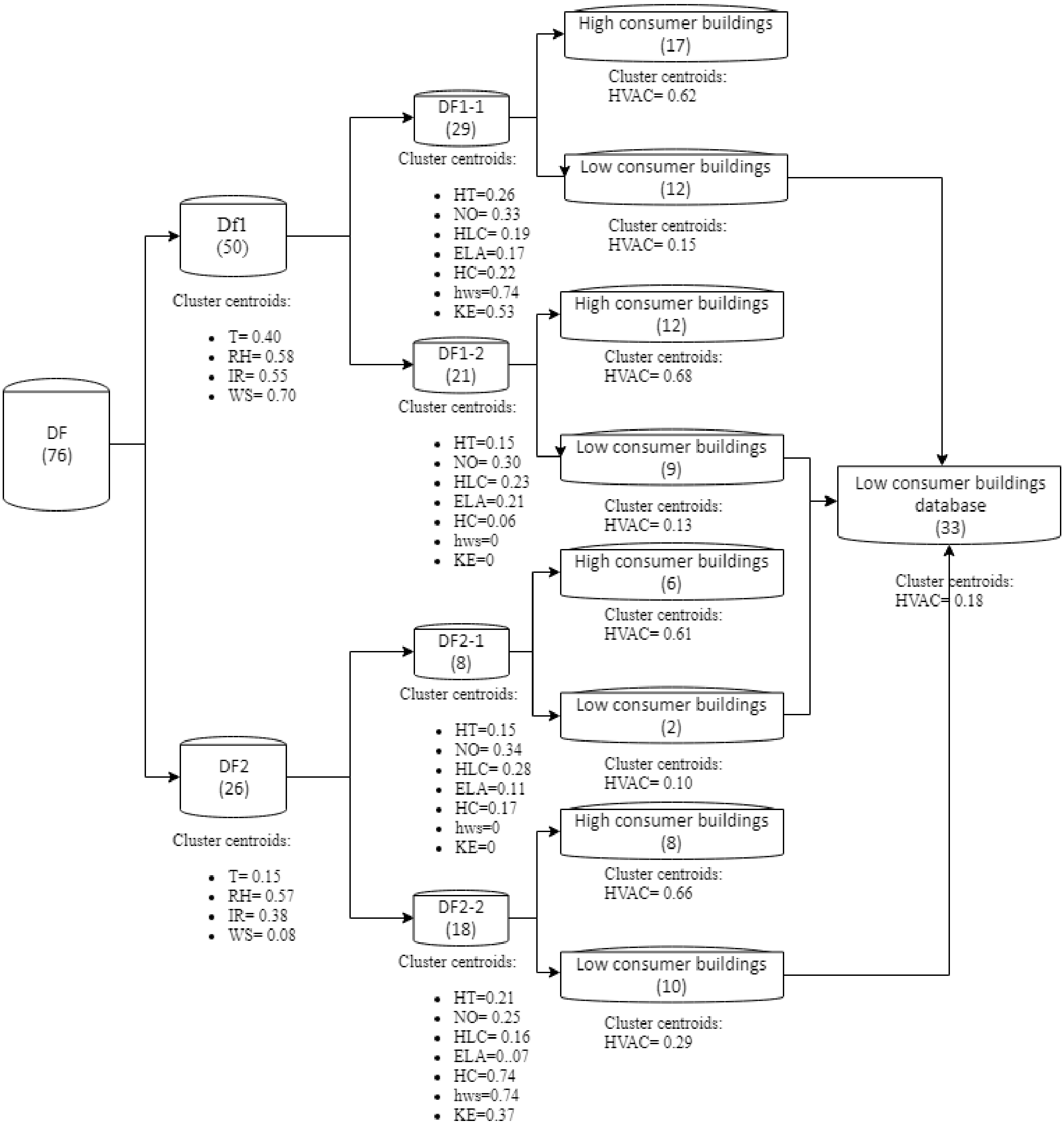

Figure 5. Clustering schematic considering HVAC energy consumption as the target.

3.3 Reference Building Generation and Building Performance Evaluation

Figure 6 shows the result of clustering regarding HVAC energy consumption. There are two sets of buildings: one with higher energy consumption (shown in orange) and the other with lower energy consumption (shown in blue). The high consumer buildings and low consumer buildings are shown in the figure. The low consumer buildings were used as the training set to create the reference building using an ANN model. Therefore, by plugging in the characteristics of any given building (shown in red in Figure 6), the model output is an estimation of how much the HVAC energy consumption of the given building would have been had it belonged to low-consumer buildings (*RB shown in red in* Figure 6). The difference between the actual and estimated value of the given building and the *RB* shows the possible energy consumption savings. It is important to mention that because the training dataset comes from a combination of clusters, the input of the model is sufficiently diverse, so the model is a reliable estimator for buildings from different types, characteristics and climate.

Figure 6. Distribution of buildings regarding HVAC energy consumption, showing clusters of high- and low-consumption buildings.

Table 3 shows the results of the eight models, along with the real consumption of eight end-use loads. The difference between them is the possible energy savings. For example, HVAC energy consumption is 0.634 (normalized values) in the given building and 0.256 in the *RB*; the difference (0.378) is the possible energy-saving potential. The same logic applies regarding HWS, FRIDGE, E&I, H&S, and OTHER. However, it is evident that LIGHT and KITCH have negative energy savings. This shows that the given building is performing better than the *RB*. In other words, the occupants of this building are cautious regarding LIGHT and KITCH appliances and may be focusing on other end-use loads to reduce their bills even further.

Table 3 Comparison of Given and *Reference Building* for Calculation of Possible Energy Savings. The values are annual energy consumption of various end-use loads normalized by the 76-building dataset.

	HVAC	**HWS**	**LIGHT**	**KITCH**	**FRIDGE**	**E&I**	**H&S**	**OTHER**
Given building	0.634	0.189	0.102	0.092	0.639	0.754	0.366	0.605
Reference Building	*0.256*	0.062	0.144	0.601	0.527	0.507	0.342	0.033
Energy savings	0.378	0.127	-0.042	-0.509	0.112	0.247	0.024	0.572

3.4 Discussion

Table 4 shows the result of the given building assessment through the developed methodology and the previous work [23] (a different approach, based on mere comparison of similar buildings with each other). It is observed that for all end-use loads except KITCH, the *RB* implies that there are more energy-saving opportunities, while according to the previous authors' work, there are limited or no such opportunities. This is because the *RB* assumes the advantage over all low-consumption buildings, while the previous work took similar buildings with lower average energy consumption than the given building.

Table 4 Comparison of Given Building through *Reference Building* and Comparison with Existing Buildings [23]. The values are annual energy consumption of various end-use loads normalized by the 76-building dataset.

	HVAC	**HWS**	**LIGHT**	**KITCH**	**FRIDGE**	**E&I**	**H&S**	**OTHER**
Given Building	0.634	0.189	0.102	0.092	0.639	0.754	0.366	0.605
Energy Savings According to *Reference Building*	0.378	0.127	-0.042	-0.509	0.112	0.247	0.024	0.572
Energy Savings According to Comparison with Existing Buildings [23]	0.161	-0.02	-0.135	-0.245	0.067	0.182	-0.033	0.284

It is seen that reference building can fully take advantage of low consumer buildings and make a baseline to assess any given building energy consumption. It adapts to the given building by taking all the characteristics as the input (listed in Figure 4) except the occupant behavior. Therefore, the difference between the given building and the reference building can be attributed to occupant behavior. The advantage of current work in comparison to other existing works in literature is that the recommendations coming from the reference building are realistic mainly because they originate form real buildings not simulations or buildings with different characteristics.

4. Limitations and Future Insights

The energy consumption data in this study were collected annually. However, a seasonal analysis would enhance the accuracy of detection of low-consumption buildings. For example, a building that has a high HVAC energy consumption in winter and a very low consumption in summer could be categorized among low-consumption buildings when considering yearly energy consumption. However, the same building may be a high energy consumer if one considers only the winter season or a low consumer if one considers only summer. In addition, we have used annual weather data in clustering level 1-1, which does not capture underlying fluctuations in weather parameters. One solution would be to include standard deviations of the weather parameters or to use 24-hour temperature profiles of each building throughout the year. More information about individual buildings would enhance the process. For instance, the number of occupants is present in the dataset; however, more information—age; level of activity; number of adults and children; time of arrival and departure; and so on—could be used in clustering analysis when putting similar buildings together.

In clustering level 1, it is quite possible that buildings in the same cluster might have similar non-occupant-related characteristics but are dissimilar in others. To reduce the effect of this problem, each characteristic was multiplied by its contribution to energy consumption by means of GRA. Therefore, those characteristics with higher GRA values are dominant in determining the cluster to which the building belongs. In addition, clustering level 1 was divided into two subsections, levels 1-1 and 1-2, to prioritize the weather parameters in clustering. If one attempts to increase the accuracy of the clustering in terms of any characteristics, it is possible to increase the subsections of level 1 clustering to even more than two (for example, level 1-3 as number of occupants). However, this makes the cluster sizes smaller, and we were limited by the database size in this study. Increasing the size of the database can solve this issue, as discussed in the next point.

The number of buildings in this study was 76, which is not sufficient for such an analysis. Having more buildings would enhance the clustering analysis by breaking it into more levels (such as levels 1-1, 1-2, 1-3, etc.) in order of importance of variables. Furthermore, the neural network models would be better created if there were sufficient data points. In this case, a few thousands of buildings would work well. However, the purpose of this study was to demonstrate the methodology, which can be applied to similar datasets.

5. Conclusions

A general methodology was developed to generate an energy-efficient building—*RB*—using field-measured data. The generated *RB* could be used to evaluate the thermal energy performance of any given building and provide information to building occupants about their building energy performance with respect to any specific end use (such as HVAC, kitchen appliances, entertainment, etc.). A dataset of 76 buildings was studied to extract buildings with low energy consumption regarding the specific end use. Considering eight end-use loads, a dataset of eight buildings with low energy consumption was extracted from the original dataset and named as a set of low-consumption buildings. By building neural network models using these low-consumption buildings as the training dataset, mapping of building characteristics to an estimation of energy use was performed. Therefore, any given building could be compared with its *RB* (which was generated using ANN models). The results show that by comparing the two, it is possible to spot the end-use loads that are consuming more than the *RB*; therefore, the occupants may focus on those energy uses and take measures (i.e., turning off unnecessary lights or HVAC, etc.) to improve building energy performance. The methodology introduced here can be applied to larger building datasets in any climate.

Acknowledgment

The authors would like to express their gratitude to Concordia University for supporting this research.

REFERENCES

[1] M. Jia, R.S. Srinivasan, A.A. Raheem, From occupancy to occupant behavior: An analytical survey of data acquisition technologies, modeling methodologies and simulation coupling mechanisms for building energy efficiency, Renew. Sustain. Energy Rev. 68 (2017) 525–540. doi:10.1016/J.RSER.2016.10.011.

[2] M. Sameti, F. Haghighat, Optimization of 4th generation distributed district heating system: Design and planning of combined heat and power, Renew. Energy. 130 (2019) 371–387. doi:10.1016/J.RENENE.2018.06.068.

[3] T. Hong, D. Yan, S. D'Oca, C. Chen, Ten questions concerning occupant behavior in buildings: The big picture, Build. Environ. 114 (2017) 518–530. doi:10.1016/J.BUILDENV.2016.12.006.

[4] 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. http://www.sciencedirect.com/science/article/pii/S0926580514002507 (accessed September 14, 2017).

[5] F. Xiao, C. Fan, Data mining in building automation system for improving building operational performance, Energy Build. 75 (2014) 109–118.

[6] 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.

[7] C.M.R. do Carmo, T.H. Christensen, Cluster analysis of residential heat load profiles and the role of technical and household characteristics, Energy Build. 125 (2016) 171–180. doi:http://doi.org/10.1016/j.enbuild.2016.04.079.

[8] J.M. Abreu, F.C. Pereira, P. Ferrão, Using pattern recognition to identify habitual behavior in residential electricity consumption, Energy Build. 49 (2012) 479–487.

[9] 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:https://doi.org/10.1016/j.enbuild.2011.02.002.

[10] X. Liang, T. Hong, G.Q. Shen, Occupancy data analytics and prediction: a case study, Build. Environ. 102 (2016) 179–192.

[11] M. Saarikoski, A data mining approach to indoor environment quality assessment: A study on five detached houses in Finland, (2016).

[12] A. Capozzoli, M.S. Piscitelli, A. Gorrino, I. Ballarini, V. Corrado, Data analytics for occupancy pattern learning to reduce the energy consumption of HVAC systems in office buildings, Sustain. Cities Soc. 35 (2017) 191–208. doi:10.1016/J.SCS.2017.07.016.

[13] Y. Wang, L. Shao, Understanding occupancy pattern and improving building energy efficiency through Wi-Fi based indoor positioning, Build. Environ. 114 (2017) 106–117. doi:10.1016/J.BUILDENV.2016.12.015.

[14] I. Kastner, E. Matthies, Implementing web-based interventions to promote energy efficient behavior at organizations – a multi-level challenge, J. Clean. Prod. 62 (2014) 89–97. doi:10.1016/J.JCLEPRO.2013.05.030.

[15] A. Meinke, M. Hawighorst, A. Wagner, J. Trojan, M. Schweiker, Comfort-related feedforward information: occupants' choice of cooling strategy and perceived comfort, Build. Res. Inf. 45 (2017) 222–238. doi:10.1080/09613218.2017.1233774.

[16] C. Fischer, Feedback on household electricity consumption: a tool for saving energy?, Energy Effic. 1 (2008) 79–104. doi:10.1007/s12053-008-9009-7.

[17] Z. Yu, J. Li, H.Q. Li, J. Han, G.Q. Zhang, A Novel Methodology for Identifying Associations and Correlations Between Household Appliance Behaviour in Residential Buildings, Energy Procedia. 78 (2015) 591–596. doi:10.1016/J.EGYPRO.2015.11.024.

[18] 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).

doi:10.1016/j.enbuild.2018.04.052.

[19] Z. (Jerry) 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.

[20] C. Fu, J. Zheng, J. Zhao, W. Xu, Application of grey relational analysis for corrosion failure of oil tubes, Corros. Sci. 43 (2001) 881–889. doi:http://dx.doi.org/10.1016/S0010-938X(00)00089-5.

[21] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine Learning in Python, J. Mach. Learn. Res. 12 (2011) 2825–2830.

[22] L. Rokach, O. Maimon, The Data Mining and Knowledge Discovery Handbook: A Complete Guide for Researchers and Practitioners, (2005).

[23] 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.