GINA CODY
SCHOOL OF ENGINEERING
AND COMPUTER SCIENCE

Department of Building, Civil
& Environmental Engineering

Occupancy data analysis in BC dwellings using ecobee smart thermostat data

Final report of research services submitted to

Jessica Webster, MCIP RPP

Energy Planning Analyst
Buildings and Renewables, CanmetENERGY Ottawa
National Research Council of Canada

Karthik Panchabikesan, Ph.D.,
Post-doctoral Researcher

Mohamed Ouf, Ph.D., P.Eng
Assistant Professor
Ursula Eicker, Ph.D.,
Professor and Canada Excellence Research Chair (CERC) in Smart, Sustainable and Resilient Communities and Cities.

March 31, 2021

1455 De Maisonneuve Blvd. W., EV-6.139, Montreal, Quebec, Canada H3G 1M8
Tel. 514-848-2424 ext. 3200 concordia.ca/ginacody

Executive summary

This report presents the research findings by Concordia University (under the contract with the CanmetENERGY-Ottawa division of Natural Resources Canada on the Canadian Energy End-use Mapping project) to support the input required related to occupancy and occupant schedules in energy simulations for the City of Kelowna. Research activities focused on analyzing occupant movement detection data recorded by the smart thermostats and remote sensors. To perform the analysis, data collected from ecobee thermostats installed in 461 dwellings across British Columbia between 2015 and 2019 was used. Since the data available for the houses located in Kelowna is limited (32 dwellings), the analysis was extended to include households located in British Columbia (429 dwellings).

The report is organized into six main sections: 1) Introduction, 2) Objectives of the study, 3) Data description and understanding, 4) Research method *(data processing and analysis)*, 5) Results and discussion, and 6) Conclusion.

Results provide an overview of occupancy in residential dwellings and show the variations in hourly occupancy schedules across dwellings located in British Columbia. The data processing steps required to convert raw sensor-recorded occupancy data into useful results are explained in detail. The difference between current assumptions made in the Housing Technology Assessment Platform (HTAP) housing energy simulations and the study's results is highlighted. The k-means clustering is used to extract distinct occupancy profiles from the dataset. The average 24-h occupancy profile for weekdays and weekends in each cluster was presented. The characteristics of the dwellings grouped in each cluster were analyzed and reported. Finally, the procedure adopted in the study, the main outcomes of the report, the applicability of the results, and the scope for future works are consolidated in conclusion.

1455 De Maisonneuve Blvd. W., EV-6.139, Montreal, Quebec, Canada H3G 1M8
Tel. 514-848-2424 ext. 3200 concordia.ca/ginacody

1. Introduction

Occupant's schedules and their energy-related behavior are substantial inputs for building energy simulations. In residential buildings, electrical appliances and thermostat usage depend on the occupant's energy use behavior. Hence, assigning harmonized schedules for occupancy, electrical appliances, and thermostat setpoint temperature is important. In most energy simulation software, the heating, ventilation, and air conditioning loads are estimated based on the temperature setpoint schedules and internal gains. Thus, it is highly recommended to define realistic schedules related to occupancy, setpoints, and equipment usage. Knowing the importance of occupant behavior in evaluating the building energy performance, International Energy Agency (IEA) have dedicated Annexes 53 [1], 66 [2], and 79 [3] focused on studying the influence of occupants in building energy use, simulation methods of occupant behavior, and addressing the research issues related to occupant-centric building design and operation, respectively. These IEA annexes emphasize the research potential and need for understanding, modeling the occupancy in building performance simulation.

In general, the occupant schedule is highly uncertain [4]. Thus, in most building energy simulations, default/fixed schedules (based on the building codes, national averages) are considered [5], and the temporal variations associated with occupancy are often missed out, leading to uncertainties in building performance simulations. The main reason for considering the default/standardized schedules is the data non-availability, and in most cases, the actual occupant behavior is unknown [6]. Recently, there is a wide implementation of smart energy meters in buildings, resulting in massive data collection related to occupancy, energy, and indoor environmental condition. One example of such a comprehensive data source is the 'Donate Your Data' initiated by the thermostat manufacturer ecobee Inc [7]. In this context, it is interesting to extract occupancy schedules from the data collected from the dwellings.

2. Objectives

The study's main objectives are 1) to calculate the percentage of the time the occupants are at home and 2) to extract distinct occupancy profiles from the real time occupant movement detection data collected from the dwellings located in British Columbia. The first objective helps define

1455 De Maisonneuve Blvd. W., EV-6.139, Montreal, Quebec, Canada H3G 1M8
Tel. 514-848-2424 ext. 3200 concordia.ca/ginacody

appropriate occupancy percentage in HTAP simulations using the HOT 2000 tool to evaluate building energy performance. The second objective provides detailed hourly occupancy profiles (for weekdays and weekends, respectively) obtained through k-means clustering analysis.

3. Data description and understanding

3.1. Data overview

The data used in this report was obtained from the users in British Columbia who agreed to share their thermostat usage, and occupancy data anonymously under the 'Donate Your Data (DYD)' program from the year 2015 to 2019 administrated by ecobee Inc., [7]. The data from the DYD program is partially user-reported (e.g., data pertaining to dwelling characteristics such as the number of floors, dwelling age, and floor area) and partially collected from ecobee thermostats (e.g., occupancy movement detected in the house by the PIR motion sensor embedded within the thermostat and the remote sensors). The occupancy-related data available in the DYD dataset is presented in Table 1.

Table 1: Ecobee data description

Attribute	Description
House ID	Anonymous unique ID of each user
Thermostat motion sensor	Indicates the occupant movement detected as 0 or 1 by the motion detection sensor installed in the ecobee thermostat
Remote sensor	Indicates the occupant movement detected as 0 or 1 by the remote sensors
Province	User inputted value
City	User inputted value
Dwelling type	User inputted value (Detached, Apartment, Condominium, and others)
Floor area (ft^2)	User inputted value
Number of Floors	User inputted value
Age of the house	User inputted value
Number of occupants	User inputted value
Auxiliary heat fuel type	Fuel type for the auxiliary heat system

The occupant movement data in each dwelling is recorded every five minutes. In total, the research leverages the data collected from 461 dwellings (32 dwellings from Kelowna and 429 dwellings from British Columbia). Since the data available for houses located in Kelowna is limited (32 houses), the analysis was extended to households located in British Columbia (BC) to ensure the results' representativeness. The distribution of house types in Kelowna and BC is shown in Figure 1. As depicted, most dwellings from which ecobee data was available in Kelowna and the rest of BC are single-family dwellings (SFDs). With respect to the number of floors, floor area, and building age, most dwellings have one or two floors, are between 2000 and 2500 ft^{2,} and are between one and ten years old for both Kelowna and BC (as shown in Figure 2 and Figure 3, respectively).

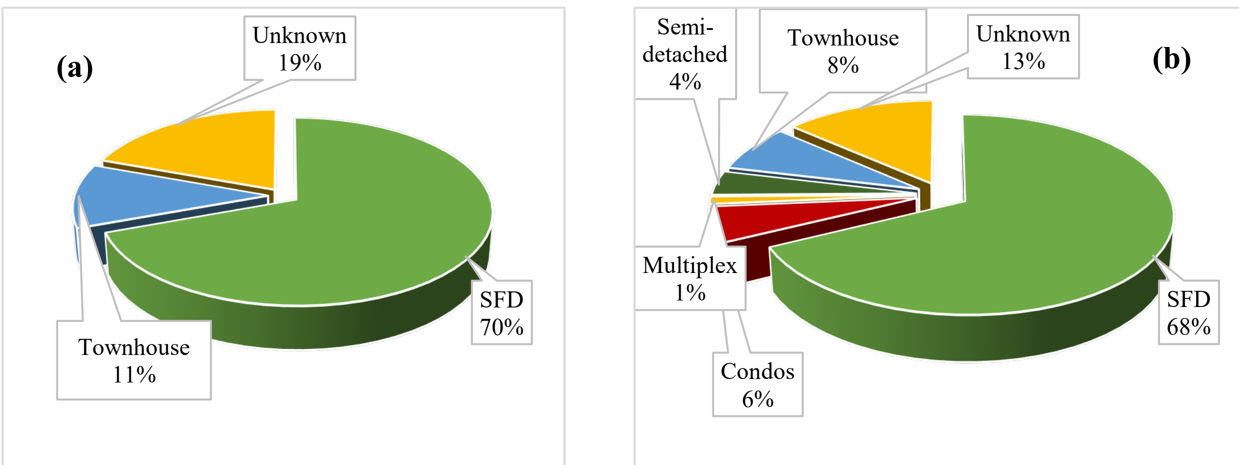


Figure 1: Distribution of dwelling types in ecobee dataset in (a) Kelowna, (b) British Columbia

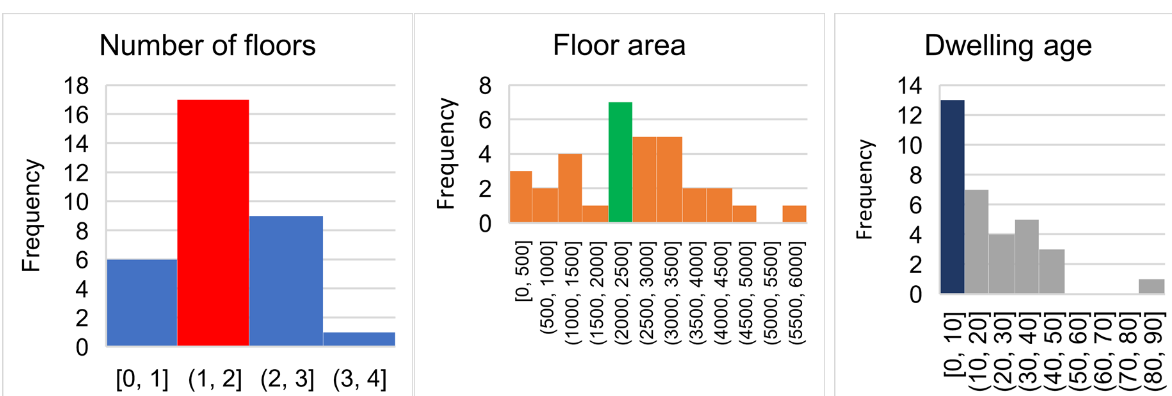


Figure 2: Dwelling characteristics in Kelowna

1455 De Maisonneuve Blvd. W., EV-6.139, Montreal, Quebec, Canada H3G 1M8
Tel. 514-848-2424 ext. 3200 concordia.ca/ginacody

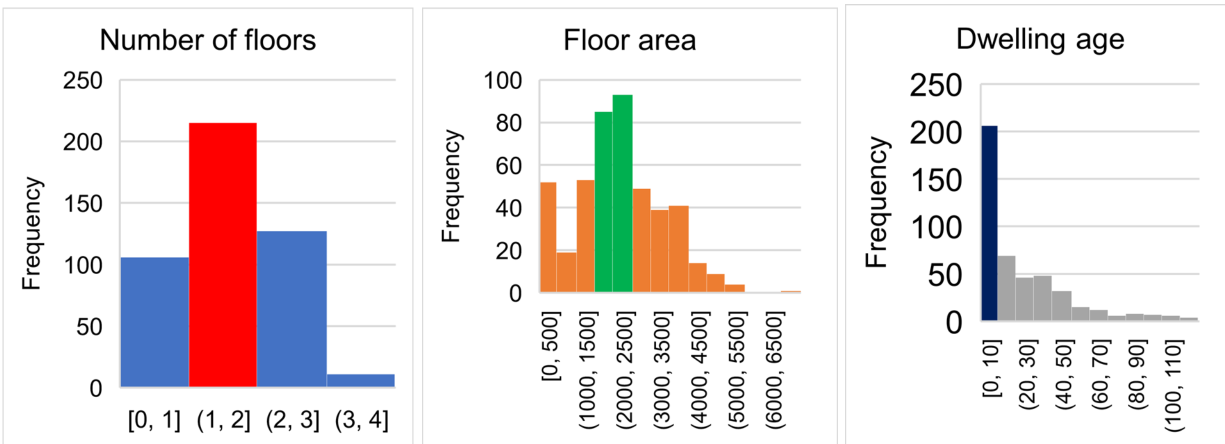


Figure 3: Dwelling characteristics in British Columbia

3.2. Data availability

The data availability for each dwelling was based on the user's enrolment in the DYD program. The data collection's start and end date for the ecobee dataset used in this study were from September 2015 to September 2019. Therefore, data available for each dwelling represented a subset of the above-mentioned period. For example, some dwellings have the data measured from September 2015 till September 2019, whereas some dwellings have only available for 2019. In this report, all available data for each dwelling was used for the analysis irrespective of duration. Figures 4 and 5 show the data available in months for each dwelling in Kelowna and BC, respectively. It can be seen from Figure 4 that 50% of the dwellings in Kelowna have the data available between 17 to 27 months. For BC, 35%, 31%, 17%, and 17% of the dwellings have the data for 1 to 12, 13 to 24, 25 to 36, and 37 to 47 months. It is to be mentioned that the start and end date of the data availability might be different for each dwelling.

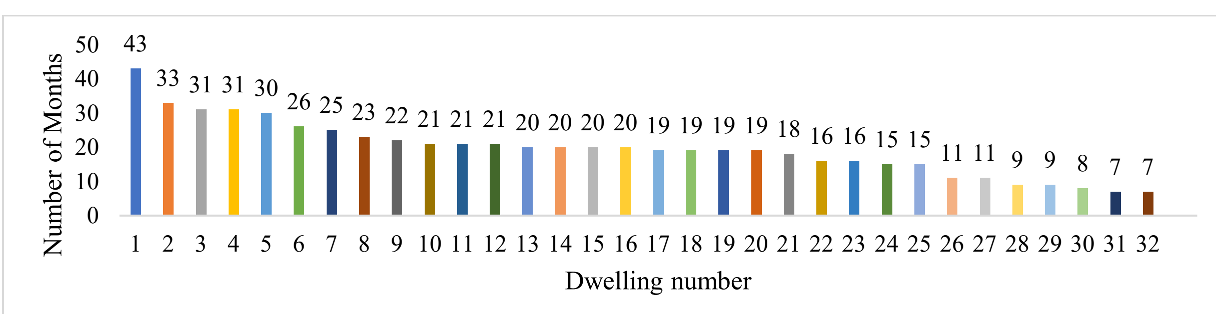


Figure 4: Data availability in months for each dwelling in Kelowna

1455 De Maisonneuve Blvd. W., EV-6.139, Montreal, Quebec, Canada H3G 1M8
Tel. 514-848-2424 ext. 3200 concordia.ca/ginacody

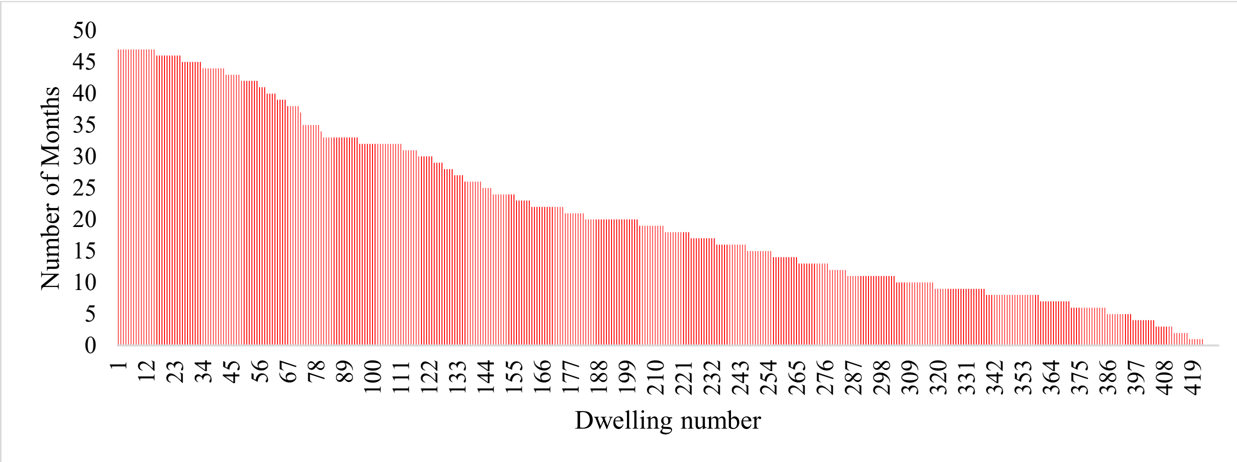


Figure 5: Data availability in Months for each dwelling in BC

3.3. Number of sensors installed in each dwelling

In this report, the occupant movements detected by both the PIR motion sensor (installed within the thermostat) and the remote sensors (installed in different rooms of the dwelling) were used for the analysis. Note that the number of remote sensors installed differs randomly for each dwelling. Further, the location of the thermostat is unknown. Therefore, occupant movements depend on the sensor's location, time of the day, and occupant activity throughout the 24-hour period.

During the data analysis, it was found that 73 dwellings have no data recorded by both the PIR and remote sensors (related to occupancy). The possible reason could be that the users might turn OFF the option to record the occupant movement detected inside the house. Hence, in the study, the data collected from 388 dwellings (28 and 360 dwellings for Kelowna and BC, respectively) were used hereafter for the analysis. Figures 6(a) and 6(b) shows the number of sensors installed in dwellings in Kelowna and BC, respectively. Note that the number of sensors mentioned in the figures includes the PIR thermostat and remote sensors. As seen from Figure 6, 64% of the dwellings in Kelowna are equipped with two sensors. For BC, 48%, 24%, and 15% (covering 89%) of dwellings have 2, 4, and 1 sensor, respectively.

1455 De Maisonneuve Blvd. W., EV-6.139, Montreal, Quebec, Canada H3G 1M8
Tel. 514-848-2424 ext. 3200 concordia.ca/ginacody

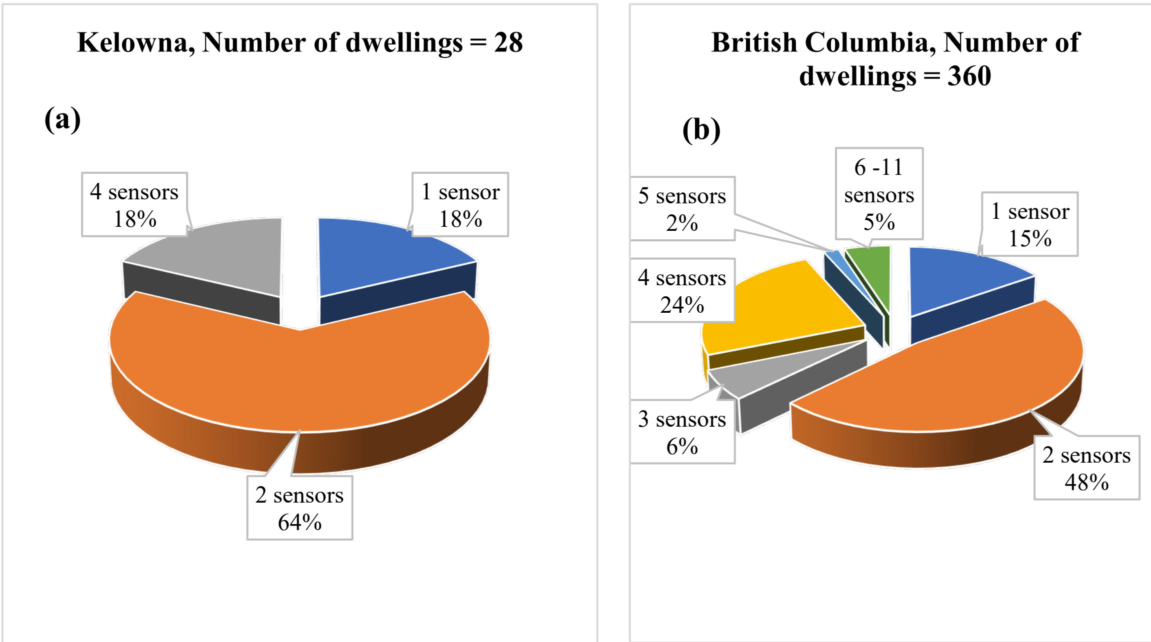


Figure 6: Total number of sensors (inclusive of both thermostat motion sensor and remote sensors) installed in Kelowna dwellings

4. Research method

4.1. Data processing and transformation

Occupant movements were detected by passive infrared (PIR) and remote sensors every 5-minutes as binary values ('1' means movement detected and '0' no movements detected). To derive the hourly average occupancy profile for each dwelling, at first, the movement detected for every 5 minutes by each sensor was averaged to hourly values, respectively. As the next step, the hourly average occupancy value obtained for all the sensors was averaged to obtain the overall average occupancy value for each hour of the day, resulting in occupancy values between 0 to 1. Figure 7(a) and 7(b) shows the weekdays and weekends average hourly occupancy profile for each of 388 dwellings.

1455 De Maisonneuve Blvd. W., EV-6.139, Montreal, Quebec, Canada H3G 1M8
Tel. 514-848-2424 ext. 3200 concordia.ca/ginacody

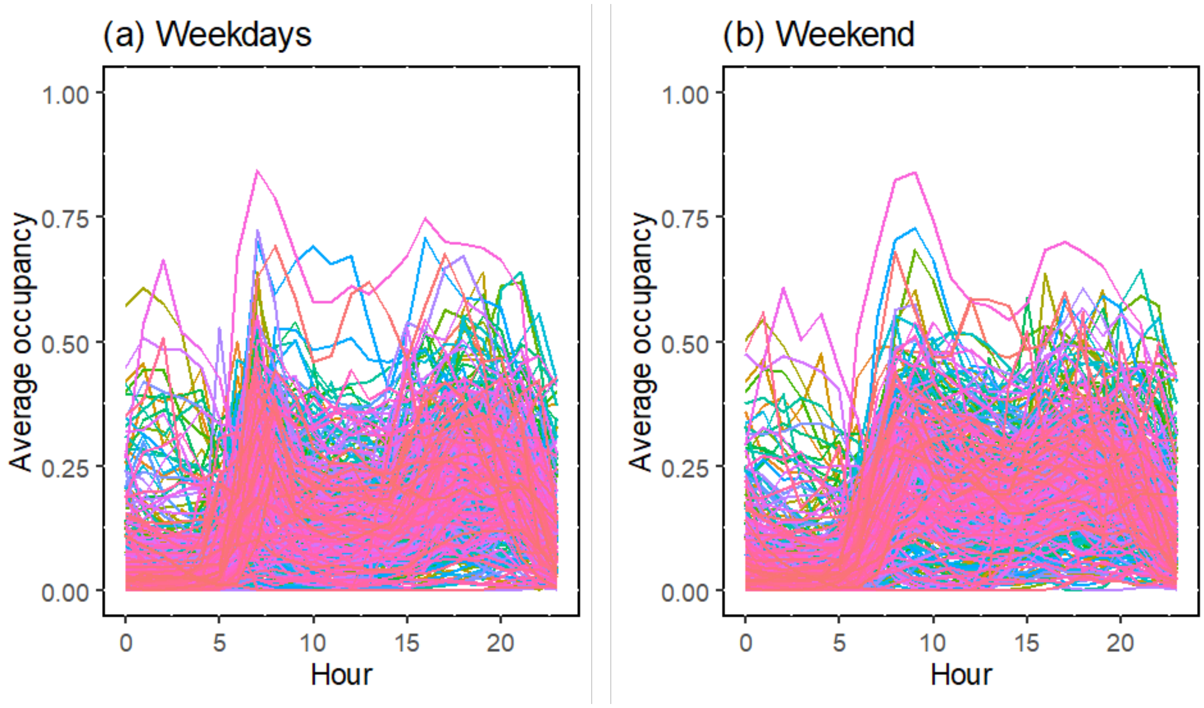


Figure 7: Average occupancy profile for 388 dwellings (a) Weekdays profile, (b) Weekend profile.

The hourly average occupancy profile shown in Figure 8 indicates that the average occupancy value during the night hours (22:00h to 07:00h) is less than the day hours. This is due to the sleeping activity of the occupants. To calculate the percentage of the occupants' time at home, it is meaningful to convert the hourly average occupancy values to either 0 or 1 for each dwelling, respectively. Because currently, the average occupancy value represents the occupants' activity level in the dwelling and does not explicitly indicate whether the occupants are present at home for a specific hour of the day. In this context, different scenarios were considered to convert the average occupancy profile (which is currently a value between 0 and 1) to either 0 or 1 and to modify the occupancy values during the nighttime. To do so, in each scenario, different threshold values (median, 1st quartile) were considered for daytime and nighttime to modify the occupancy values as 0 or 1. It is to be mentioned that the median and 1st quartile value that was considered as threshold values varies for each hour and each dwelling, respectively. If the average value for the specific hour is greater than the same hour's threshold, then the occupants were considered to be at home ('1'), else not ('0'). Table 2

1455 De Maisonneuve Blvd. W., EV-6.139, Montreal, Quebec, Canada H3G 1M8
Tel. 514-848-2424 ext. 3200 concordia.ca/ginacody

describes the conditions used in different scenarios to modify the average occupancy values. It is to be mentioned that 3rd quartile was not considered as the threshold because it underestimated the occupancy. In this study, hereafter, the average occupancy converted based on the defined threshold is referred to as 'modified occupancy'.

Table 2: Different scenarios and description

Scenario	**Description**
1	*For both nighttime* (22:00h to 07:00h) and daytime* (08:00h to 21:00h):* Converting the average occupancy values based on each hour's median and for each dwelling, respectively. In other words, if the average value is less than the median, then the modified occupancy value is 0, else 1
2	*For nighttime:* The average occupancy value is changed to 1 for the entire night time, irrespective of the average occupancy values. *For the daytime:* The median is considered as the threshold as Scenario 1.
3	*For nighttime*: The average occupancy value is modified to 1 based on the median. Additionally, a condition was applied such that if the modified occupancy was found to be '1' in any of the night hours of the day, the entire night is changed to '1' or vice versa. *For daytime:* The median is considered as the threshold as Scenario 1.
4	*For nighttime:* The average value is modified to '1' if the average value itself is found to be greater than 0. Additionally, the condition was applied such that if the modified occupancy was found to be '1' during any of the night hours of the day, the entire night is changed to '1' and vice versa. *For daytime:* The median is considered as the threshold as Scenario 1.
5	*For nighttime:* The average value is modified to '1' if the average value itself is found to be greater than 0. Additionally, the condition was applied such that if the modified occupancy was found to be '1' during any of the night hours of the day, the entire night is changed to '1' and vice versa. *For the daytime:* The average value during the daytime is also modified to '1' if the average value itself is found to be greater than 0.
6	*For nighttime:* The average occupancy value is modified to 1 based on the 1st quartile. Additionally, a condition was applied such that if the modified occupancy was found to be '1' in any of the night hours of the day, the entire night is changed to '1' or vice versa. *For daytime:* The 1st quartile is considered as the threshold, accordingly, the average occupancy value is modified to either 0 or 1.

* For all the scenarios, the same daytime and nighttime are considered.

1455 De Maisonneuve Blvd. W., EV-6.139, Montreal, Quebec, Canada H3G 1M8
Tel. 514-848-2424 ext. 3200 concordia.ca/ginacody

4.2.Data analysis

After data processing, the first step entailed was calculating the percentage of time occupants are at home considering the modified occupancy values obtained based on each scenario. Successively, an appropriate scenario was determined. The outcome of this analysis (home occupied percentage) shall be straightaway used in HTAP simulations by replacing the current assumptions related to the occupancy with the outcomes of this study.

The next step focused on obtaining hourly average modified occupancy values for each dwelling. This average hourly profile shall be considered as the occupants' presence percentage in the dwelling. Once the 24-h occupancy profiles were obtained for each dwelling, clustering analysis was performed. As one of the most used data mining techniques, clustering shows superior performance in identifying unknown patterns/groups and has been extensively applied in multiple domains for building performance improvement. Different clustering methods were used in the literature to serve different application purposes, e.g., feature extraction for building energy prediction, load profiling, nearly zero energy buildings (nZEBs) grouping, and occupant behavior pattern recognition. Among the available cluster analysis methods, k-means cluster analysis has gained the greatest popularity due to its simplicity and high effectiveness. Subsequently, in this study, k-means clustering is performed to extract the distinct patterns of occupancy profiles. Since data is available for only 28 houses in Kelowna, clustering was not performed separately for Kelowna's dwellings. Instead, the clustering is performed considering all the 388 dwellings altogether. To perform the 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 clusters. After identifying the clusters, the weekday and weekend occupancy profile for each cluster is estimated and analyzed.

5. Results and discussion

5.1. Home occupied percentage

Currently, the occupancy in the Housing Technology Assessment Platform (HTAP) simulations is defined by assuming that the occupants are present at home for 50%. Further, the information on which period of the day the occupants are present at home is not mentioned. In this study, following

1455 De Maisonneuve Blvd. W., EV-6.139, Montreal, Quebec, Canada H3G 1M8
Tel. 514-848-2424 ext. 3200 concordia.ca/ginacody

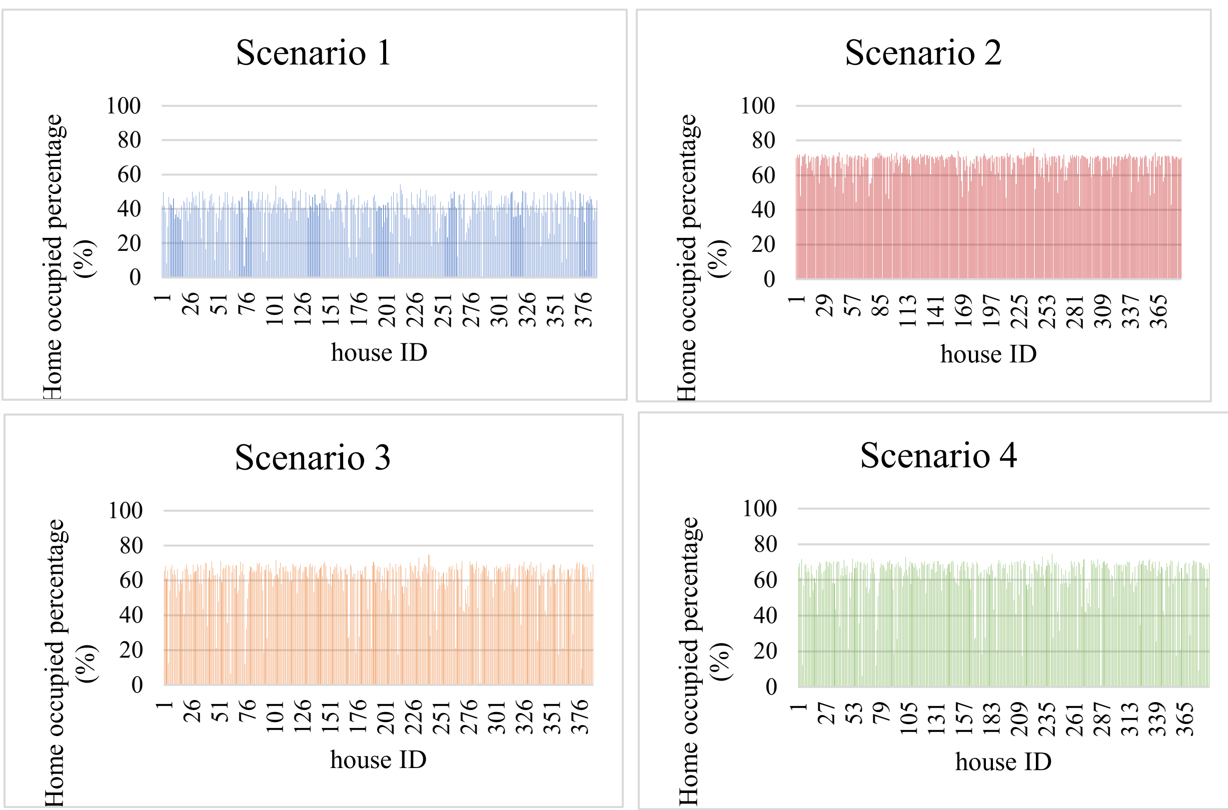
the data processing steps explained in section 4, the percentage of time each dwelling was occupied considering different scenarios was calculated as shown in Figure 8. Figure 9 depicts the number of dwellings in each percentage bin (histogram chart). As depicted in Figure 8, scenario 1 has the lowest home occupied percentage than the other scenarios, where 96% of the dwellings were occupied for <50%. The low occupied percentage was because of the condition applied to process the nighttime occupancy in Scenario 1. As the median was considered as the threshold for both daytime and nighttime, and no additional conditions were applied for the night hours (refer to Table 1), the modified occupancy for most of the night hours in scenario 1 was 0 affecting the overall home occupied percentage. While in scenario 2, a simple condition was assumed, such as the dwelling was occupied for the entire nighttime irrespective of the average value derived from the occupant movements detected by the sensors. Though this simple condition improves the occupancy during the nighttime, in some cases, it might overestimate the occupancy during the night hours. As a result, it can be seen from Figure 9 that 40% of the dwellings have an occupied percentage between 70% to 80%.

Scenarios 3 and 4 have a resemblance in the estimated home occupied percentage. In both scenarios, it is estimated that ~66% of the dwellings have the home occupied percentage between 60% to 70%. In Scenarios 3 and 4, a systematic approach was adopted compared to Scenario 2. After converting the average occupancy to either 0 or 1 based on the threshold, an additional condition was applied (refer to Table 2), which addresses the shortcoming of Scenario 2. The only difference in Scenarios 3 and 4 is the threshold, wherein the former, the median was considered, and in the latter, the average value itself was considered as the threshold. In Scenario 5, the average occupancy value was considered as the threshold for both day and nighttime. However, this assumption exaggerated the home occupied percentage where 43% of the dwellings have the home occupied percentage greater than 80%. To try a different threshold, in scenario 6, 1st quartile was used to calculate the home occupied percentage. This assumption increased the occurrences of 1 in the modified occupancy (especially in daytime compared to scenarios 1 to 4). Subsequently, ~30% of the dwellings have the home occupied percentage between 80% to 90%.

Overall, comparing the results of different scenarios, the home occupied percentage obtained for scenarios 3 and 4 seems to be realistic (as scenario 1 underestimates the time the occupants are at

1455 De Maisonneuve Blvd. W., EV-6.139, Montreal, Quebec, Canada H3G 1M8
Tel. 514-848-2424 ext. 3200 concordia.ca/ginacody

home, whereas scenarios 2, 5, and 6 overestimated the home occupied percentage). Among scenarios 3 and 4, the modified occupancy values (0 or 1) obtained based on scenario 4 were considered in the study's next steps to derive the 24-h average occupancy profile for each dwelling. Since the average values were calculated based on the occupant movement detected by the sensors, inherently, the occupancy during the nighttime was low, and hence it is reasonable to consider the average value as the threshold rather than the median. It is observed from Figure 9 (scenario 4) that ~66% of the dwellings were occupied for 60 to 70% of the time, which is deviating from the assumption made for occupancy in the HTAP housing simulations. Since the home occupied percentage influences the electrical appliances and domestic hot water usage, it is suggested to update the current assumptions in HTAP housing energy simulations with the results obtained in this study.



1455 De Maisonneuve Blvd. W., EV-6.139, Montreal, Quebec, Canada H3G 1M8
Tel. 514-848-2424 ext. 3200 concordia.ca/ginacody

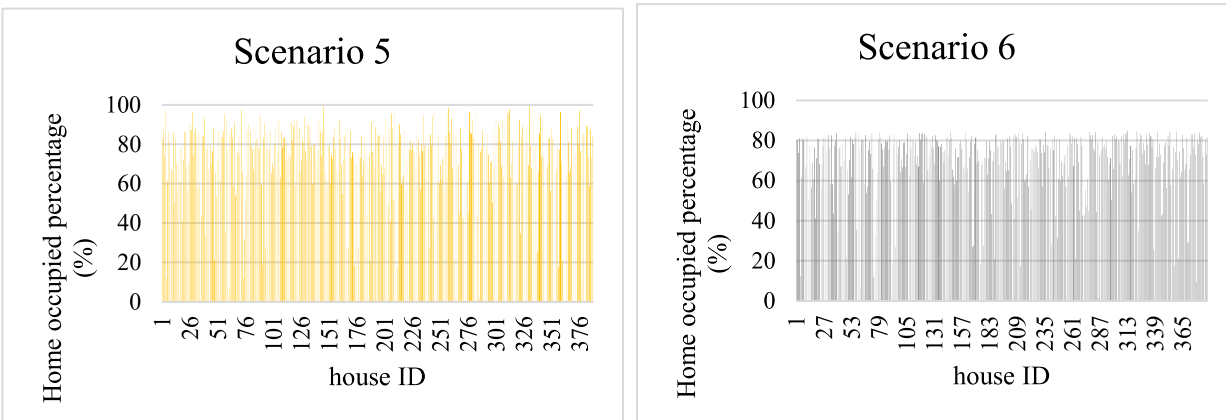
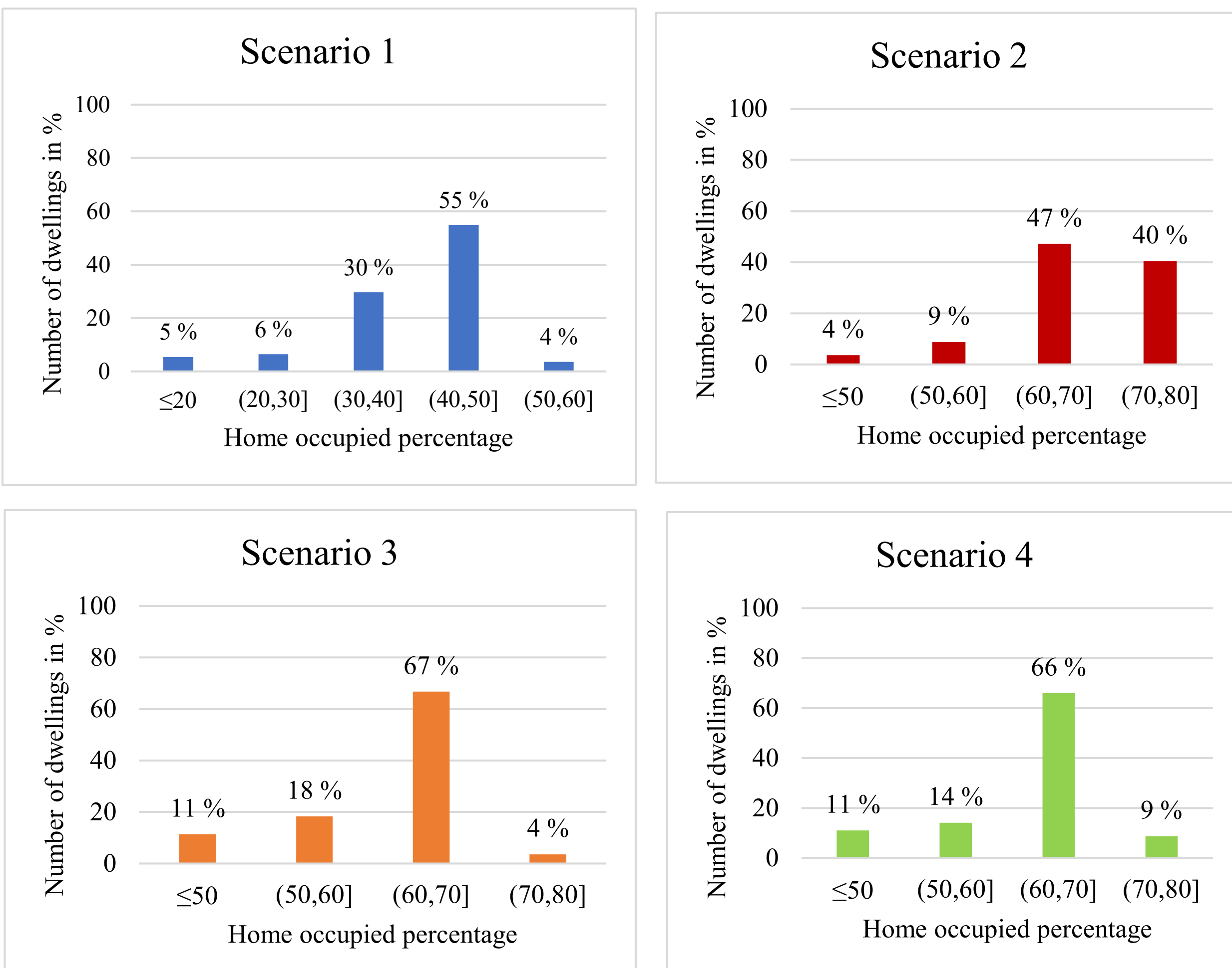


Figure 8: Home occupied percentage for different scenarios



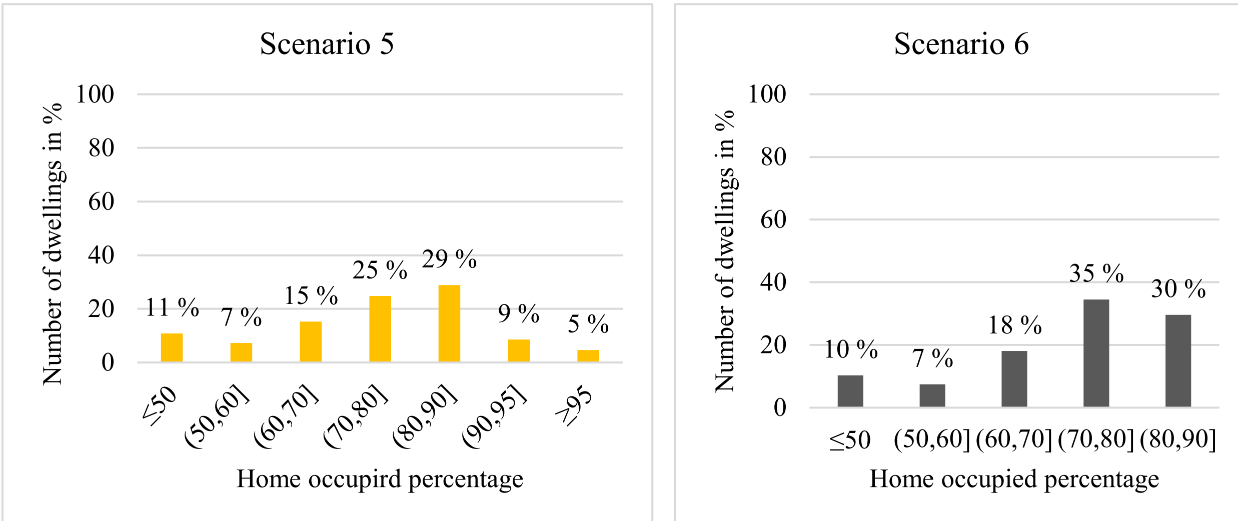


Figure 9: Frequency distribution of dwellings in % with respect to home occupied percentage for each scenario

5.2. Clustering results

The 24-h occupancy profile used for clustering analysis is shown in Figure 10. The occupancy profile shown in the figure was derived by taking the hourly average of the modified occupancy values (for each dwelling) that was obtained based on scenario 4. In other words, Figure 10 represents the occupant presence (in fraction) at home rather than the average occupancy (indicating the activity level), as shown in Figure 7. As seen from Figure 10, diversified occupancy profiles are observed for each dwelling (especially during the night hours). Subsequently, to extract different occupancy schedules, in this study, k-means clustering is performed. In k-means clustering, it is important to determine the appropriate number of clusters before performing the clustering. Subsequently, using the elbow method, the optimal number of clusters was set to '4', and the results are shown in Figure 11.

1455 De Maisonneuve Blvd. W., EV-6.139, Montreal, Quebec, Canada H3G 1M8
Tel. 514-848-2424 ext. 3200 concordia.ca/ginacody

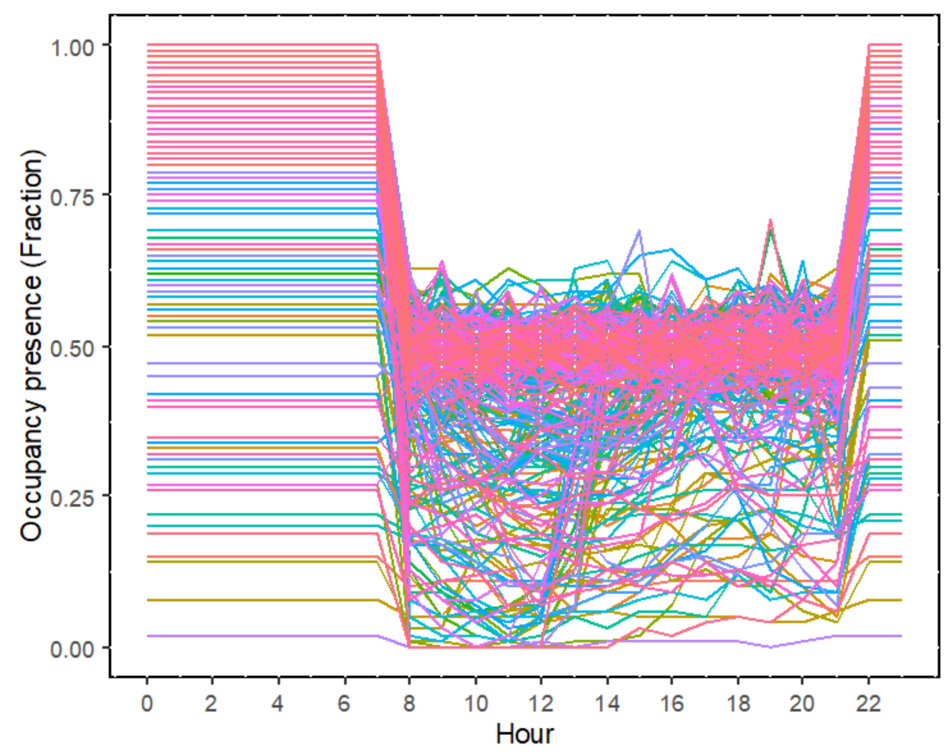


Figure 10: 24-h profile of modified occupancy for all 388 dwellings in BC

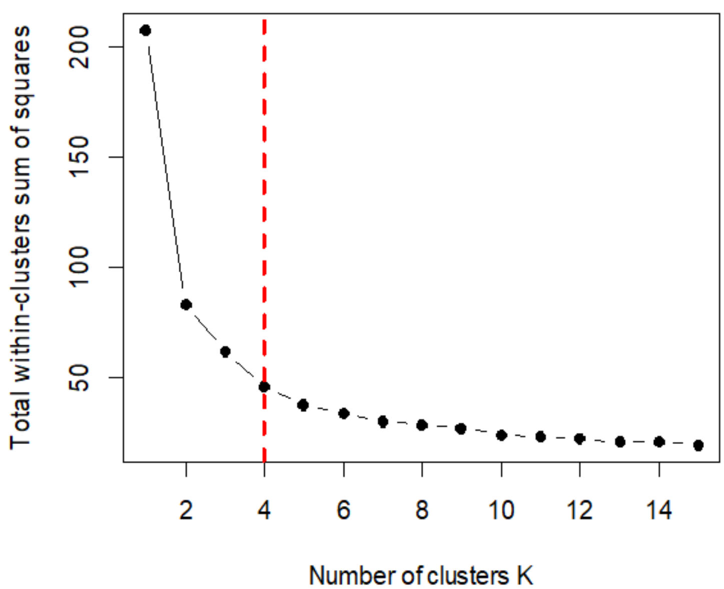


Figure 11: Elbow method showing the appropriate number of 'k'.

1455 De Maisonneuve Blvd. W., EV-6.139, Montreal, Quebec, Canada H3G 1M8
Tel. 514-848-2424 ext. 3200 concordia.ca/ginacody

Figure 12 shows the centroids of each cluster, and Figure 13 shows the 24-h occupancy profiles of the dwellings grouped in clusters 1 to 4. The overall distribution of dwellings in clusters 1, 2, 3, and 4 are 13%, 69%, 7%, and 11%, indicating that a plurality of the dwellings has the occupancy profile of cluster 2. The inference from Figure 12 is that during nighttime, clusters 2 and 4 followed similar patterns, whereas, in cluster 4, lower occupancy was observed during the daytime (specifically in the morning hours between 08:00h to 12:00h). After 12:00h, the occupancy in the dwellings grouped in cluster 4 increases gradually, and in the evening period (after 16:00h), the occupancy is similar to cluster 2. On the other hand, cluster 1 has a similar trend as cluster 2 during the daytime, whereas during the night hours, occupancy is lower for cluster 1 compared to cluster 2. Cluster 3 is distinct and has the lowest occupancy among all the clusters. The occupancy for the dwellings grouped in cluster 3 is low throughout the day, indicating that the dwellings might be the vacation homes.

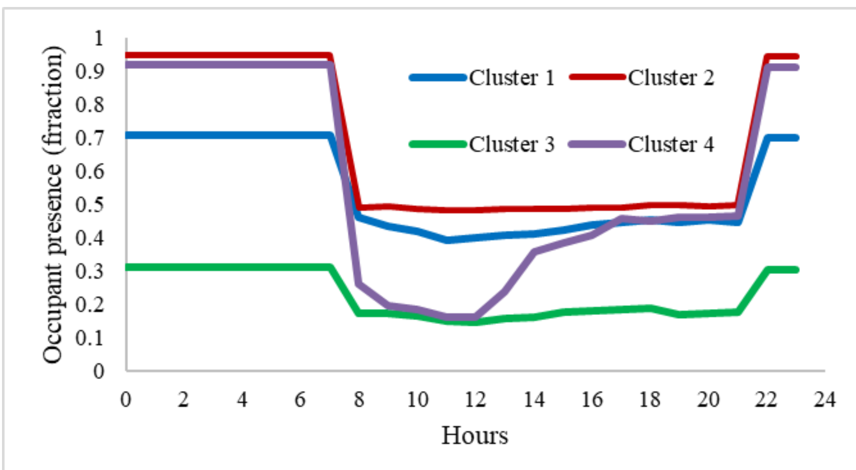


Figure 12: Centroids of each cluster

1455 De Maisonneuve Blvd. W., EV-6.139, Montreal, Quebec, Canada H3G 1M8
Tel. 514-848-2424 ext. 3200 concordia.ca/ginacody

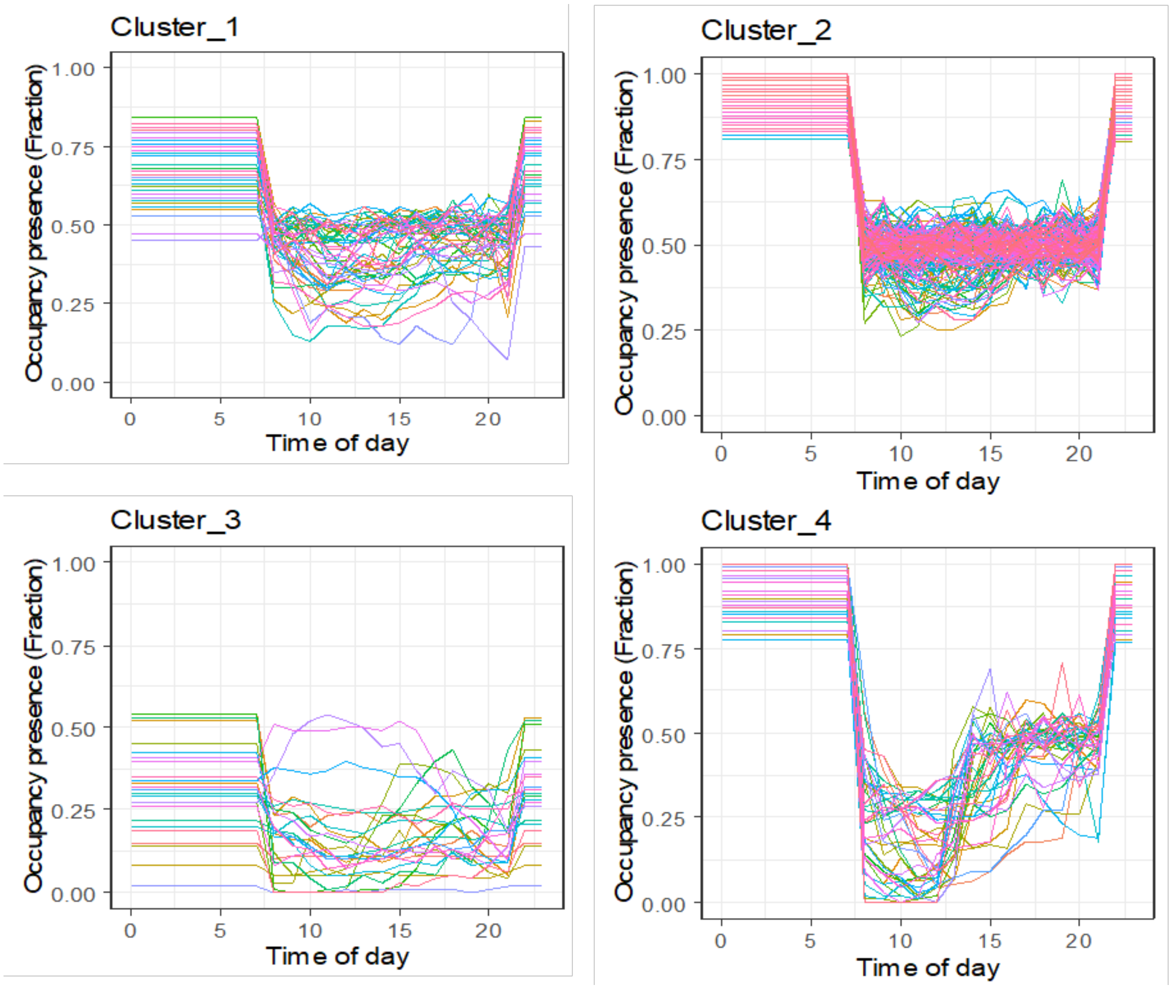


Figure 13: Occupancy profiles of the dwellings grouped in each cluster

The weekday and weekend profile for each cluster is presented in Figure 14. As seen from the figure, the occupancy for weekends during the daytime is relatively higher than the weekdays. Considering such variations in occupancy schedule might increase the accuracy of building performance simulations.

1455 De Maisonneuve Blvd. W., EV-6.139, Montreal, Quebec, Canada H3G 1M8
Tel. 514-848-2424 ext. 3200 concordia.ca/ginacody

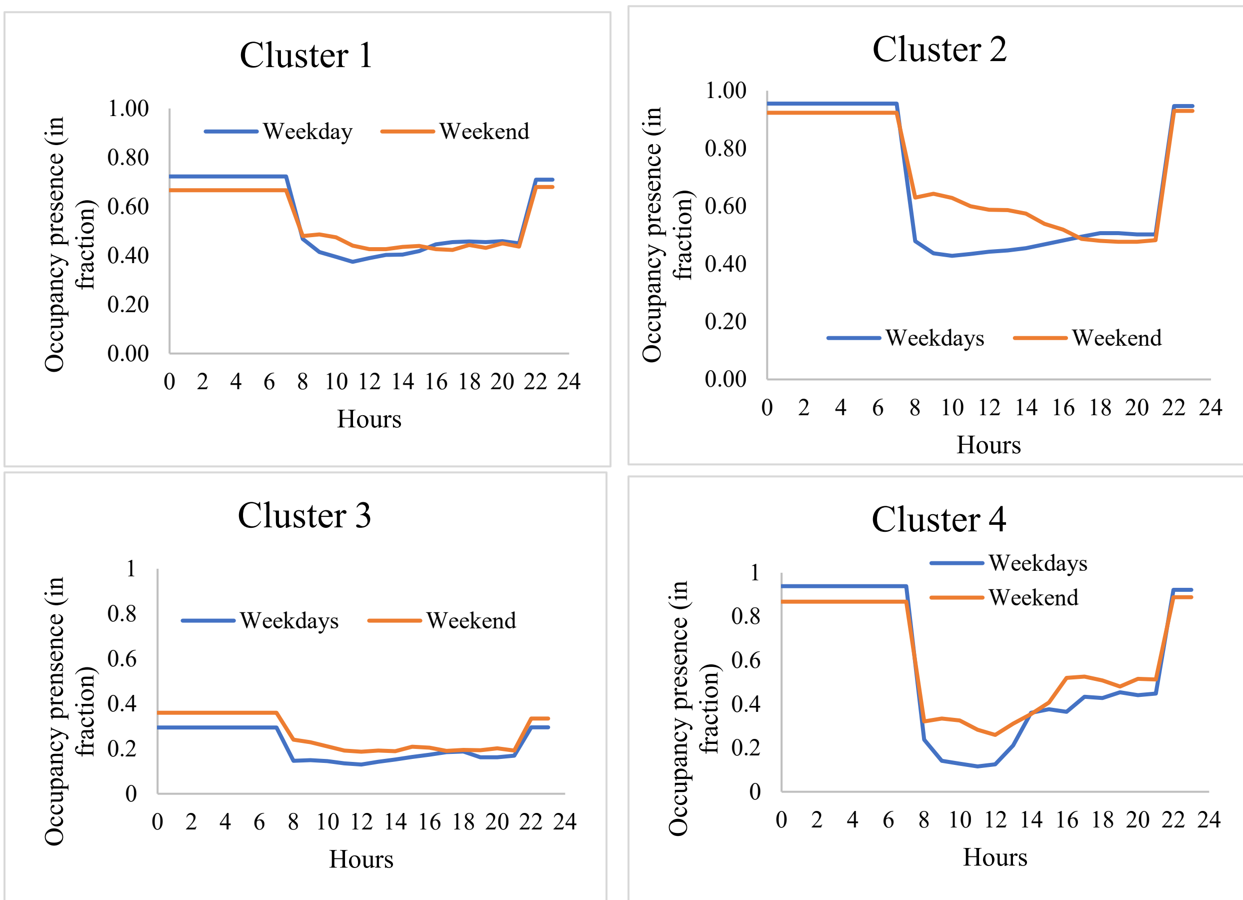


Figure 14: Weekdays and Weekend occupancy profile for each cluster

5.3. Analyzing the impact of the number of sensors

Since the number of sensors in each dwelling (refer to section 3.3) varies, it is interesting to study whether the number of sensors impacts the derived occupancy schedules and the clustering results. Such analysis will help to understand whether higher occupant presence was obtained for the dwellings with more sensors or vice-versa. Figure 15 displays the number of sensors in the dwellings concerning each cluster. As shown in the figure, the majority of the dwellings have two sensors in all the clusters. Even in cluster 3, where the lowest occupancy profile was observed, 61% of the dwellings are equipped with 2 sensors. Based on these observations, it was determined that no conclusive evidence could be drawn to identify the impact of the number of sensors over the derived occupancy profiles.

1455 De Maisonneuve Blvd. W., EV-6.139, Montreal, Quebec, Canada H3G 1M8
Tel. 514-848-2424 ext. 3200 concordia.ca/ginacody

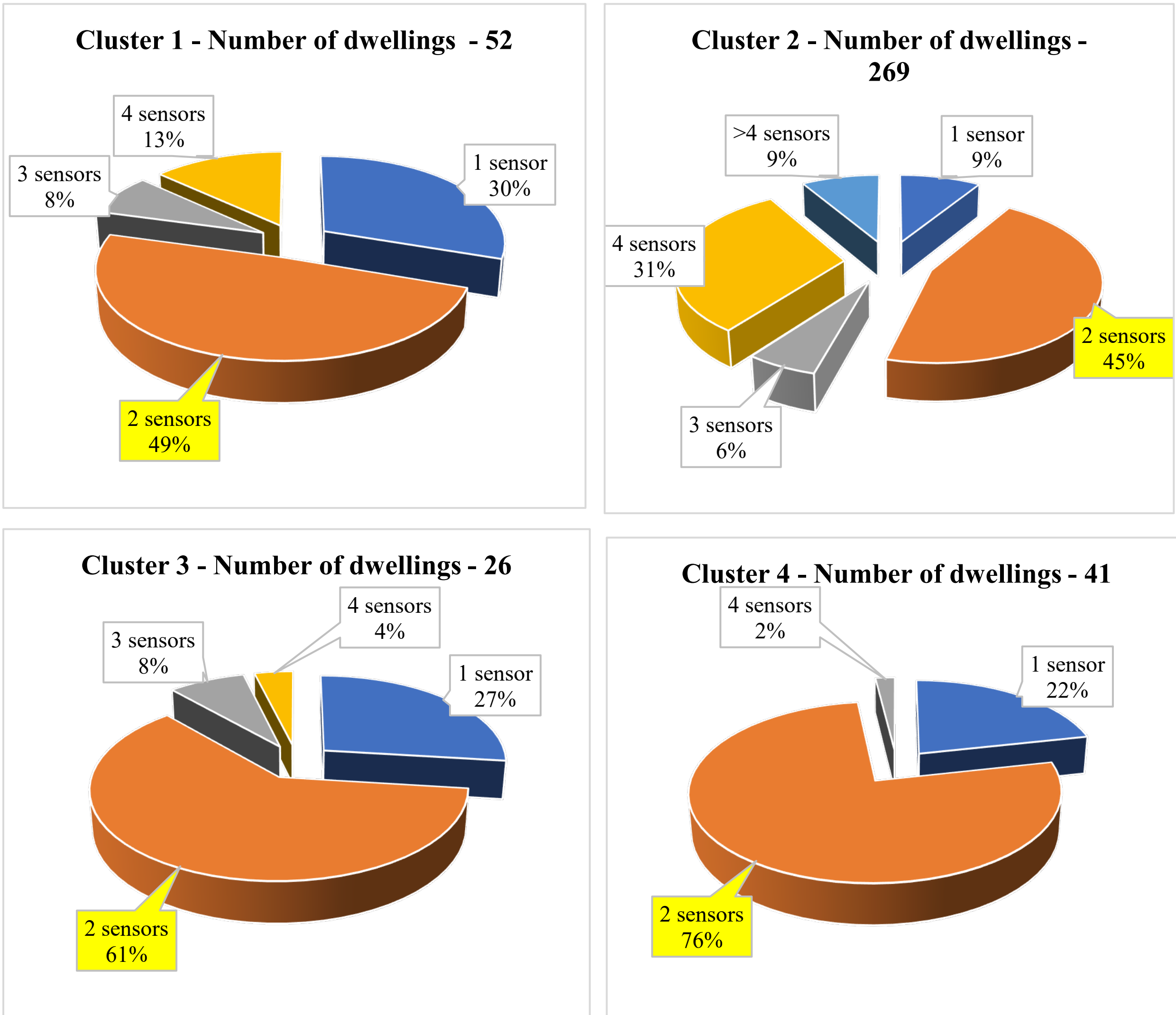


Figure 15: Number of sensors installed in the dwellings with respect to each cluster

Additionally, the characteristics (in terms of dwelling type, number of floors, age and number of floors) of the dwellings grouped under each cluster were studied. Regarding the dwelling type, most of the dwellings belonged to SFDs in all the clusters, being a monopoly. Similarly, the majority of the dwellings in all the clusters have one or two floors and are newly built with the building age between 0 to 10 years. However, a considerable difference was observed between each cluster for the floor area. Figure 16 shows the frequency distribution of the number of dwellings for different floor area bins. As depicted, for cluster 1, 40% of the dwellings have a floor area between 1000 to 2000 ft^2, whereas for cluster 2, 36% and 29% of the dwellings have a floor area between 2000 to 3000 ft^2 and 1000 to 2000 ft^2, respectively. For clusters 3 and 4, more buildings have a floor area lesser than 1000 ft^2.

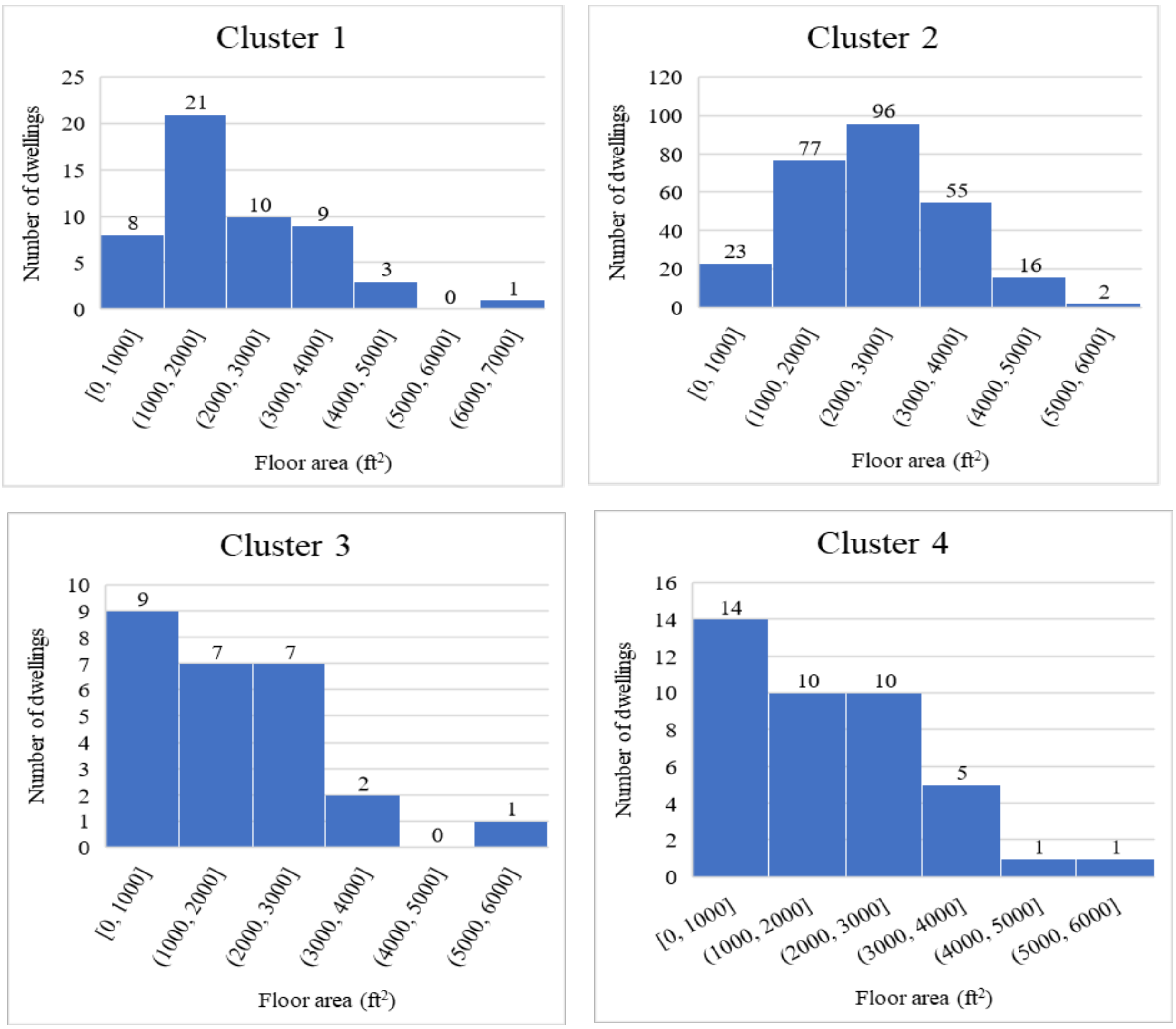


Figure 16: Frequency distribution of number of dwellings with respect to the floor area

6. Conclusion

The report used the occupancy data collected from the ecobee thermostats installed in 461 households across British Columbia. During the data analysis, it was found that no data related to occupant movement detections were available for 73 dwellings. Hence, for this analysis, occupancy data collected from 388 dwellings were used. The study's major objectives were to estimate the home occupied percentage and extract different average hourly occupancy profiles using clustering analysis. Appropriate data processing techniques were followed to process the occupant movement detected by multiple motion detection sensors (at 5 minutes time resolution) installed in each dwelling. The first step followed in the data processing procedure was aggregating the 5 minutes value to the hourly values, representing occupancy values between 0 to 1. The major challenge in the study was processing the average occupancy values obtained during the night hours (22:00h to 07:00h). Due to the sleeping activity, the aggregated occupancy values were found to be lower compared to the daytime. Using the average value as obtained in the next steps of the study would lead to results indicating lower occupancy during night time, which is not realistic for residential buildings. Therefore, different threshold values were considered for daytime and nighttime occupancy data transformation. Accordingly, the average values were converted to either 1 or 0, representing whether the occupants are at home for the day's specific hour. Six different scenarios were considered, and among the six scenarios, the results obtained for scenario 4 indicated the majority of dwellings were occupied for 60 to 70% of the time. The other scenarios (except scenario 3) either underestimated or overestimated the occupancy rate in dwellings. These results suggested that the current assumptions made in HTAP models for occupancy (which assumes occupants are at home for 50% of the time) contradict the ecobee data results. Since electrical appliance loads and domestic hot water usage have a strong correlation with the occupancy rate (especially in residential buildings), it is recommended to update the current HTAP model assumptions related to occupancy considering the results provided in this report.

After calculating the home occupied percentage, the next step entailed obtaining the average 24-h occupant presence profile for each dwelling. For this analysis, the modified occupancy values obtained considering the conditions mentioned in scenario 4 were averaged based on the hour of the day. Subsequently, an average 24-h occupancy profile was obtained for each dwelling. Afterward, k-means clustering was performed to extract distinct occupancy profiles. The occupancy profile of 388 houses was grouped in four clusters based on the elbow method results (an approach to determine the optimal number of clusters). Among the four clusters, 69% of the dwellings were grouped in cluster 2.

1455 De Maisonneuve Blvd. W., EV-6.139, Montreal, Quebec, Canada H3G 1M8
Tel. 514-848-2424 ext. 3200 concordia.ca/ginacody

Further, it was observed that the occupant schedule for weekends during the day hours was relatively higher compared to the weekdays. Accounting for such variations in occupancy shall be helpful in increasing the accuracy of the energy simulation models. Although at present, hourly schedules for occupancy were not considered in HTAP models, the presented results could be beneficial in the near future, especially if other more detailed simulation tools are considered. Since other energy simulation tools rely on dynamic occupancy profiles, it is recommended to use the hourly occupancy schedule in HTAP models. In addition, the dwelling characteristics of each cluster were analyzed. Results showed that the majority of the dwellings are SFDs in all clusters. However, variations in floor area were observed between each cluster. Additionally, in this report, the effect of the number of remote sensors over the derived occupancy profiles was studied to check whether a higher occupancy profile was obtained for dwellings with more sensors. Based on the analysis, it was determined that no conclusive evidence could be drawn to identify the impact of the number of sensors over the derived occupancy profiles.

Overall, the results presented in this report shall be helpful to update the current assumptions related to occupancy in HTAP models. The first applicability of the results shall be considering the house occupied percentage as 60 to 70% instead of 50%. In the future, the hourly occupancy schedules derived in this report can be used for energy simulations instead of defining only the occupancy rate. The scope for future studies includes extending the data analysis to other Canadian provinces to see if there is any difference in occupancy profiles between different provinces. Furthermore, increasing the dataset could provide more results for other dwelling types that can be useful to define distinct occupancy profiles for different building archetypes. The other potential scope for future works could be analyzing the correlation between the occupancy and thermostat setpoint settings and exploring options for enhancing energy efficiency by providing appropriate and useful energy-related feedback to users on the thermostat usage.

Acknowledgment

The Concordia team thank Ecobee Inc. for providing the data for the analysis through their 'donate your data' program.

1455 De Maisonneuve Blvd. W., EV-6.139, Montreal, Quebec, Canada H3G 1M8
Tel. 514-848-2424 ext. 3200 concordia.ca/ginacody

References

[1] H. Yoshino, T. Hong, N. Nord, IEA EBC annex 53: Total energy use in buildings—Analysis and evaluation methods, Energy and Buildings, 152 (2017) 124-136.
[2] D. Yan, T. Hong, B. Dong, A. Mahdavi, S. D'Oca, I. Gaetani, X. Feng, IEA EBC Annex 66: Definition and simulation of occupant behavior in buildings, Energy and Buildings, 156 (2017) 258-270.
[3] W. O'Brien, A. Wagner, M. Schweiker, A. Mahdavi, J. Day, M.B. Kjærgaard, S. Carlucci, B. Dong, F. Tahmasebi, D. Yan, T. Hong, H.B. Gunay, Z. Nagy, C. Miller, C. Berger, Introducing IEA EBC annex 79: Key challenges and opportunities in the field of occupant-centric building design and operation, Building and Environment, 178 (2020) 106738.
[4] J. Rouleau, L. Gosselin, P. Blanchet, Robustness of energy consumption and comfort in high-performance residential building with respect to occupant behavior, Energy, 188 (2019) 115978.
[5] J. Li, Z.J. Yu, F. Haghighat, G. Zhang, Development and improvement of occupant behavior models towards realistic building performance simulation: A review, Sustainable Cities and Society, (2019) 101685.
[6] G. Happle, J.A. Fonseca, A. Schlueter, A review on occupant behavior in urban building energy models, Energy and Buildings, 174 (2018) 276-292.
[7] Ecobee Inc. 2018. "Donate Your Data." https://www.ecobee.com/donateyourdata/

1455 De Maisonneuve Blvd. W., EV-6.139, Montreal, Quebec, Canada H3G 1M8
Tel. 514-848-2424 ext. 3200 concordia.ca/ginacody