Development of archetypes to represent a residential building stock

Michel Bernier, Adam Neale, Michaël Kummert
Polytechnique Montréal, Montréal, Canada

Abstract

The main objective of this work is to develop a methodology to reduce the Québec single-family building stock into a limited number of archetypes that each represent a segment of the entire stock in terms of both the annual energy consumption and highest electrical demands. The selection of the archetype for each segment is based here on a technique similar to the one used to generate typical meteorological years using cumulative distribution functions (CDF). Using this technique, a set of seven archetypes are determined to represent 1.4 million single-family houses.

In the application section of the paper, the archetypes are used to simulate changes in space heating and domestic hot water systems and examine the impact on annual energy consumption, annual greenhouse gas emissions, and peak electrical demand.

Introduction

Building stock aggregation is referred to as a 'bottom up' approach where individual buildings are analysed and aggregated to evaluate the performance of a building stock (CMHC, 2004). For example, total energy use can be estimated by adding up the energy estimates for all the individual buildings within the stock. This may prove to be computationally intensive if the scale of the aggregation is at the national level with millions of buildings. To reduce the effort, it is possible to use a subset of statistically representative buildings or "archetypes" to estimate the characteristics of the entire stock (Reinhart & Cerezo Davila, 2016). Furthermore, archetypes can be segmented to represent buildings in a stock that have common features (e.g. all buildings equipped with gas furnaces and electric water heaters). In addition to archetypes for buildings, sub-archetypes for equipment and occupant behaviour can be created which makes it easier to model changes of a group of building archetypes.

This work is divided in two main parts. First, a methodology to reduce a building stock into a limited number of archetypes using cumulative distribution functions is presented. Then, the usefulness and computational efficiency of the archetypes is analyzed by simulating changes in space heating and domestic hot water (DHW) systems and examining the impact on: i) annual energy consumption; ii) annual greenhouse gas emissions (GHG) emissions; and iii) grid electrical demand.

Literature review

Methodologies to develop archetypes have been the subject of many investigations and Neale (2021) updated the review of Reinhart and Cerezo Davila (2016) on recent archetype work.

One of the earlier studies by Swan and Ugursal (Swan & Ugursal, 2009) reviewed bottom-up approaches including the generation of archetypes. De Jaeger et al. (De Jaeger et al., 2020) quantified the error caused by the simplification of using a limited number of archetype buildings instead of simulating all buildings. Different clustering techniques are used to identify buildings that are more similar to each other than to others of another group. Peak heat demand and annual heat demand for space heating are used as key performance index. Regarding the peak heat demand, the NRMSE is 13% if the buildings are grouped into ten clusters and 8% for 40 clusters compared to the full simulation.

The NREL approach for the development of ResStock follows five steps (Reyna et al., 2022):

1- Stock characterization based on building characteristics.
2- Sampling. ResStock uses 550,000 samples to represent 133 millions dwellings.
3- Physics-based simulations. Sample buildings are simulated using EnergyPlus.
4- Calibration and validation. Annual consumptions are validated against the data of the Residential Energy Consumption Survey (RECS) of the U.S. Energy Information Administration (EIA).
5- Model outputs. Annual and hourly energy use outputs.

Reyna et al. (2022) segmented ResStock into 165 subgroups based on five climate zones, two wall types, six housing types, and three vintage bins. Thermal energy use for each segment are presented. No attempt is made to use archetypes.

For work of a similar scale to the present study, Theodoridou et al. (Theodoridou et al., 2011) divided a stock of 2.5 millions buildings into 5 archetypes. Similarly, Dall'O' et al. (Dall'o' et al., 2012) divided 1320 buildings into 7 archetypes. Finally, for a good overview of clustering techniques, the reader is referred to the work of Ghiassi and Mahdavi (Ghiassi & Mahdavi, 2017) and Dahlström et al. (Dahlström et al., 2024).

Methodology

Single-family building stock in Québec

Data-based archetypes rely on building parameters (thermal resistance of the envelope, occupancy profile etc….) for their determination. Unfortunately, these parameters are typically known only for a few buildings. In the future, if electric smart meter data are made available it might be possible to extract these parameters with some confidence (Neale et al., 2022). In the mean time, one has to rely on statistical studies as well as general surveys to evaluate the building stock. One such study was performed for the single-family building stock in the province of Québec, Canada (Neale, 2021). This stock was developed using an approach similar to the development of ResStock (Five steps mentioned in the literature review).

Neale estimated that there were a total of 1 900 220 single-family (SF) buildings in the province of Québec in 2017 (Neale, 2021). A subset of 200 000 buildings out of this total, called the Québec Single Family Building Stock Energy Model (QSFBSEM), was developed by Neale (Neale, 2021) and covers 30 regions of the province. The QSFBSEM is composed of four types of single-family (SF) buildings that are shown in Figure 1. This stock will be used here to illustrate the methodology used to obtain archetypes.

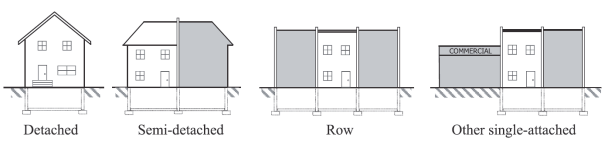

Figure 1:Single family (SF) dwelling types examined in the present study.

The QSFBSEM is based on 20 building parameters which are presented in Table 1. They cover a wide range of values (or categories) that are based on probability distribution developed using a segmentation and characterization process (Neale et al., 2020). For example, there are 24 categories of windows. There are also 15 different load profiles and up to 5 occupants per building giving a total of 75 different profiles for domestic hot water consumption, occupant heat gains, and electrical load (and associated heat gains) from appliances and lighting. The data set of the QSFBSEM is publicly available (Neale et al., 2020).

The QSFBSEM can be used to determine the energy consumption of a particular segment of this subset. For example, it is possible to determine the energy consumption of houses heated with natural gas in the Montréal region. The database consists of energy consumption profiles created by simulating the 200 000 buildings in TRNSYS with a 15 minute time step using actual weather for 2017 for each of the 30 regions. These calculations are computationally intensive and take about 36 hours on a recent computer. Some of the output values, which are evaluated at each time step for every building, are shown in Table 2.

Table 1 Building characteristics used in the QSFBSEM

Parameters	Number of Categories
Location	30
Building type	4
Load profiles	15
Window types	24
Heated floor area	5
Window-to-wall ratio	3
Building rotation	4
Occupants	1 to 5
Adjacent buildings	4
Number of floors	1 or 2
Wall thermal resistance	4
Roof thermal resistance	6
Foundation thermal resistance	4
Leakage area	5
Air conditioning	Yes or no
Heating type	8
DHW energy source	3
Aspect ratio	5
Pool	yes or no
Spa	yes or no

Table 2 Output values obtained from the QSFBSEM

Variables (all in kWh for 15 min.)	Definition
rawHC	Heating/cooling loads including contributions from internal gains and occupants
rawElec	Total electrical consumption
Internal gains	Sum of the internal gains (lighting and appliances)
Occupant load	Internal gains from occupants
DHWload	Energy requirement for domestic water heating
NatGas	Natural gas energy consumption (space and DHW)
HeatingOil	Heating oil energy consumption (space and DHW)

House archetype selection process

The first important step in the evaluation of archetypes is to determine the desired level of granularity in the segmentation process. For the present study, the objective is to examine the province-wide impact of changes made to the heating (space and DHW) systems. Seven segments are adequate to represent the various combinations of heating systems as shown in Table 3. They represent 73 % (1.4 million houses) of the entire SF building stock. The remaining 27% are houses where wood is an important contributor for space heating and will not be examined here. One archetype is to be determined for each segment.

Table 3 Segments of the SF building stock

Segment #	Energy source for space heating system	Energy source for DHW	Number of houses in the segment
1	Natural gas	Electric	31354
2	Natural gas	Natural gas	61558
3	Oil	Electric	74926
4	Oil	Oil	21824
5	Electric	Electric	877978
6	Heat pump	Electric	203172
7	Hybrid (oil/electric)	Electric	113092
		Total	1383904

The objective is to find an archetype that: i) has a heating/cooling load profile similar to the average of all houses in that segment and ii) reproduces adequately the highest electrical loads experienced by all houses in that segment.

The selection of the archetype house for each segment is based here on a technique similar to the one used to generate typical meteorological years (Wilcox & Marion, 2008) where cumulative distribution functions (CDF) for weather parameters (maximum dry bulb temperature, global solar radiation etc..) are combined to determine typical weather months. In the present case, two CDFs are evaluated. The first one calculates the CDF for the heating and cooling (HC) loads and the other evaluates the CDF for the highest electrical loads during the year. Cumulative differences between individual CDF profiles and the average CDF profile are calculated for both the HC loads and the highest electrical loads. The house that has the minimum weighted cumulative difference value is the archetype house.

In order to reduce load fluctuations and improve the archetype accuracy, it was found necessary to first modify the raw heating and cooling loads ($rawHC_{i,j}$) and electrical consumption data ($rawElec_{i,j}$) according to Equation 1:

$$\begin{aligned} HCLoads_{i,j} &= rawHC_{i,j} + Internal\ Gains_{i,j} + Occupant\ load_{i,j} \\ ElecLoads_{i,j} &= rawElec_{i,j} - eDHW\ load_{i,j} - Internal\ Gains_{i,j} \end{aligned} \tag{1}$$

where i is the time and j is the house number. Thus, $HCLoads_{i,j}$ are the net heating/cooling loads, which include envelope heat transfer (and infiltration) and solar gains but exclude internal gains and occupant loads. Furthermore, $ElecLoads_{i,j}$ excludes the electrical DHW consumption and internal gains from the raw electrical data.

The following averaged values for the whole segment are then calculated for each time step:

$$\begin{aligned} \overline{Internal\ Gains}_{i,arch} &= \textstyle\sum_{k=1}^{nhouses} Internal\ Gains_{i,k}/nhouses \\ \overline{Occupant\ load}_{i,arch} &= \textstyle\sum_{k=1}^{nhouses} Occupant\ load_{i,k}/nhouses \\ \overline{DHWload}_{i,arch} &= \textstyle\sum_{k=1}^{nhouses} DHWload_{i,k}/nhouses \\ \overline{NatGas}_{i,arch} &= \textstyle\sum_{k=1}^{nhouses} NatGas_{i,k} \\ \overline{HeatingOil}_{i,arch} &= \textstyle\sum_{k=1}^{nhouses} HeatingOil_{i,k}/nhouses \end{aligned} \tag{2}$$

where $nhouses$ is the number of houses in that segment. Values in Equation 2 can be considered to be sub-archetypes. Each archetype has its own set of sub-archetypes.

In addition, averaged heating/cooling loads and electrical loads for all houses in that segment are calculated at each time step:

$$\begin{aligned} \overline{HCLoads}_i &= \sum_{j=1}^{nTimeStep} HCLoads_{i,j}/nTimeStep \\ \overline{ElecLoads}_i &= \sum_{j=1}^{nTimeStep} ElecLoads_{i,j}/nTimeStep \end{aligned} \tag{3}$$

where $nTimestep$ is the number of time steps (=35040 in the present case, i.e. 15 minute time steps for a full year).

Then, the 200 highest values (out of 35040) of $\overline{ElecLoads}_i$ and the time i at which they occur are determined. The corresponding values of $ElecLoads_{i,j}$ for each individual house are also determined. The value of 200 is a compromise and is based on preliminary simulations. If only one value is selected (i.e. the overall peak value) then the match between the electrical peak of all houses and the archetype will be very good. However, the timing of this peak might not correspond to the maximum for the electrical grid. By increasing the number of high values to 200, the highest value of the archetype might not be as accurate but there are better chances of predicting an accurate electrical load for houses when the electrical grid is at its maximum usage.

Cumulative distribution function (CDF)

Cumulative distribution functions (CDF) are calculated for $HCLoads$ (CDF_{HC}) and the 200 highest $ElecLoads$ (CDF_{elec}) for each house and for the average of all houses. Recall that the CDF gives the proportion of values that are less or equal to specific values. For example, if a CDF value of 0.7 is obtained for a specific $HCLoads$ value of 1 kWh, then 70% of $HCLoads$ are below 1 kWh.

Individual CDF curves for $HCLoads$ ($CDF_{HC,j}$) and the CDF curve for the averaged $HCLoads$ ($\overline{CDF_{HC}}$) are evaluated. Similarly, a CDF curve for the 200 highest values of all houses ($\overline{CDF_{elec}}$) and the corresponding values for individual houses ($CDF_{elec,j}$) are determined. Representative CDF curves are presented in Figure 2 for segment #6 (see Table 3). This figure presents CDF curves for a sample (150) of individual houses, and the average of all houses. Then, the absolute value of the difference between $CDF_{HC_{j,z}}$ and $\overline{CDF_{HC,z}}$ and between $CDF_{elec_{j,z}}$ and $\overline{CDF_{elec,z}}$ are calculated at different intervals (or bins) on the x-axis:

$$\begin{aligned} \delta_{HC,j,z} &= \left| CDF_{HC_{j,z}} - \overline{CDF_{HC,z}} \right| \\ \delta_{elec,j,z} &= \left| CDF_{elec_{j,z}} - \overline{CDF_{elec,z}} \right| \end{aligned} \tag{4}$$

where z is the bin number. The distances between CDF curves are determined based on the Finkelstein-Schafer (FS) formula:

$$FS_{HC,j} = \sum_{z=1}^{n} \delta_{HC,j,z} \text{ , } FS_{elec,j} = \sum_{z=1}^{n} \delta_{elec,j,z} \qquad (5)$$

where $n = 100$ (FS values do not change for $n > 100$). Finally, the archetype house is determined as the minimum value of WS_j :

$$WS_j = WS_{HC} \times FS_{HC,j} + WS_{elec} \times FS_{elec,j} \qquad (6)$$

where WS_{HC} and WS_{elec} are weighting factors (from 0 to 1 with $WS_{HC} + WS_{elec} = 1$).

After some exploratory runs, it was found that values of $WS_{HC} = 0.5$ and $WS_{elec} = 0.5$ minimized the value of WS_j.

As shown in Figure 2, there is very good agreement between the archetype and the average of all houses for the $HCLoads$. The agreement for the 200 highest values of the electrical loads is also very good especially for the largest values (> 2.6 kWh).

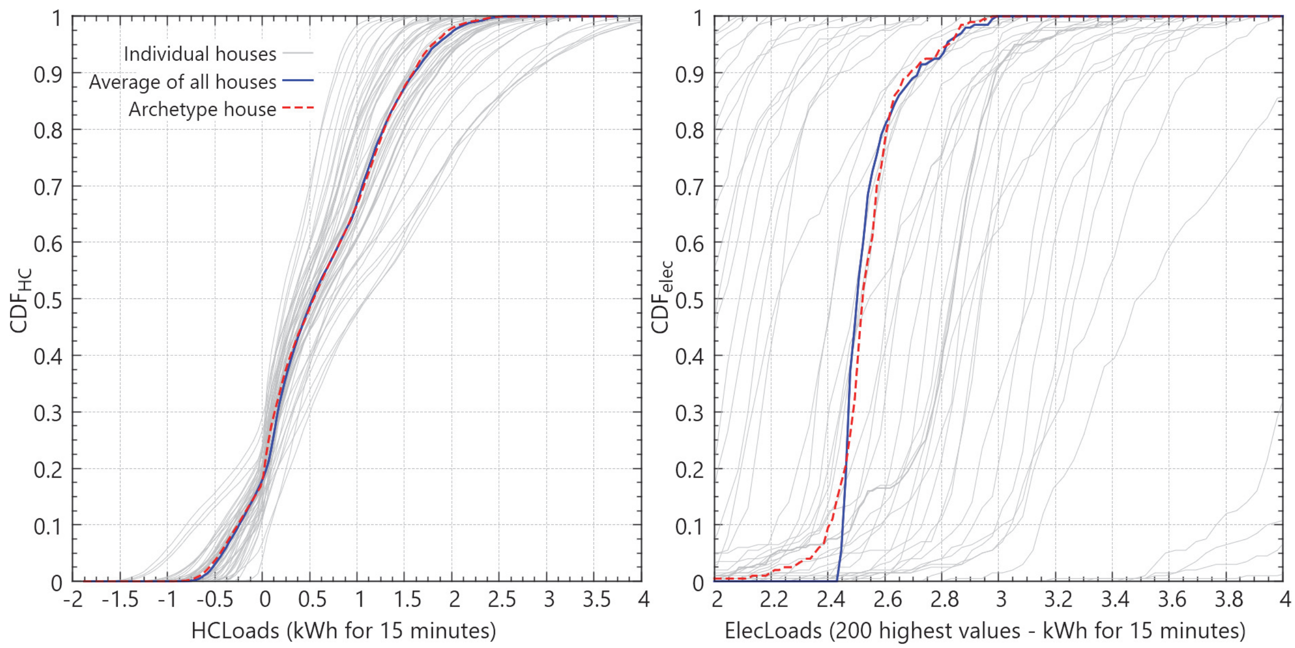

Figure 2: CDF values for the heating/cooling loads and for the 200 highest values of electrical loads.

At this stage, values of the heating/cooling loads ($HCLoads_{i,arch}$) and the electrical load($ElecLoads_{i,arch}$) profile for the archetype are known. It should be recalled that these values do not include equipment and occupant loads. The sub-archetypes for equipment and occupant loads, which have been evaluated through Equation 2, are used to obtain the final HC and electrical load profiles:

$$\begin{aligned} FinalHCloads_{i,arch} &= HCLoads_{i,arch} - \overline{Internal\ Gains}_{i,arch} \\ &\quad - \overline{Occupant\ load}_{i,arch} \\ FinalElecLoads_{i,arch} &= ElecLoads_{i,arch} + \overline{DHW\ load}_{i,arch} \\ &\quad + \overline{Internal\ Gains}_{i,arch} \end{aligned} \qquad (7)$$

House archetype profile

Figure 3 gives an example of the annual electrical load profile for houses in segment #6. Figure 3a gives the total electrical load (as calculated with equation 7) for the archetype house as well as the average of all houses in that segment while Figure 3b provides the difference (in kW) between the two curves in Figure 3a. Finally, Figure 3c is a zoomed portion of Figure 3a concentrating on the last week of 2017 with t = 8592 h corresponding to midnight on Sunday, December 24th. These figures show that the archetype adequately reproduces the behaviour of the houses for that particular segment. The RMSE (in Figure 3b) is 0.37 kW with a maximum difference of 3.08 kW occurring at t = 1476 h. As shown in Figure 3c, at the time of the Hydro-Québec grid peak in 2017 (at t = 8682 h), the total electrical load for the archetype is 9.79 kW and it is 9.69 kW for the average of all houses, a 1% difference.

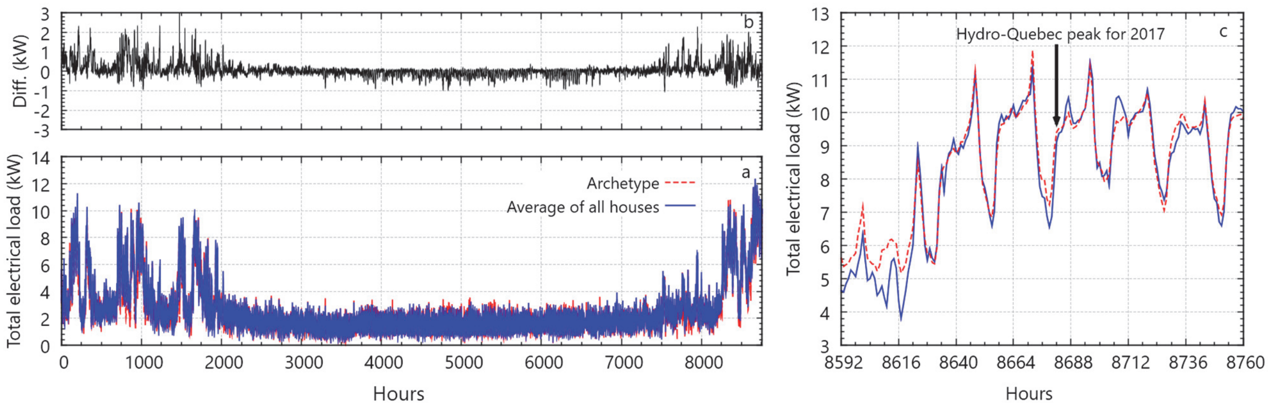

Figure 3: Total electrical loads for the archetype house and the average of all houses in segment #6.

Finally, as indicated in Table 4, the annual energy consumption is 21 679 kWh for the archetype house and 22 309 kWh for the average of all houses, a 2.8% difference.

The results for the other segments are also presented in Table 4. First, column 1 shows that the RMSE between the hourly values of the electrical load for the archetype and all houses is small; it reaches a value of 0.47 kW for segment #7. However, the maximum difference observed during a year (column 2) is much greater and can reach a value of 3.25 kW (for segment #5). This is likely due to weather conditions (solar gains for example) that are different for the archetype and the average of all houses at the time of the maximum difference. However, the time of the maximum difference does not coincide with any the fifty highest hourly grid peaks.

Two peak conditions are examined in Table 4: i) time at which the electrical loads of all houses are at their peak; and ii) time when the Hydro-Québec (HQ) grid peaked in 2017. The first observation is that not all segments peak at the same time (column 4). A comparison of columns 5 and 6 shows that the peak load predictions of the archetype house are in good agreement with the average results obtained for all houses. For example, for segment #6, the peak is at t = 8672 h with an archetype prediction of 11.34 kW while the average house has a peak of 11.87 kW, a difference of 4.6%. As shown by comparing columns 7 and 8, the predictions are much better at the time of Hydro-Québec's peak (t = 8782 h in 2017). For instance, as indicated earlier, the load predicted by the archetype is 9.79 kW while it is 9.69 kW for the average of all houses for segment #6.

Finally, the predictions of the annual energy consumption (last 6 columns in Table 4) are also in good agreement except perhaps for the oil consumption in segment #7, which shows a larger difference.

As a final point of comparison, the electrical load during Hydro-Québec's peak and the combined energy consumption (electricity+ natural gas + oil) for the entire single-family building stock (1.4 million homes) is compared to those calculated using the archetypes in Table 5.

Table 4 Comparison between the archetypes and the average of all houses for each segment

	Electrical Load								Annual energy consumption					
									Electrical		Natural gas		Oil	
Segment	RMSE (kW)	Maximum difference (kW)	Time of maximum difference (h)	Time of peak for all houses (h)	All houses at time of peak (kW)	Archetype at time of peak (kW)	All houses at time of HQ peak (kW)	Archetype at time of HQ peak (kW)	Archetype house (kWh)	Average houses (kWh)	Archetype house (kWh)	Average houses (kWh)	Archetype house (kWh)	Average houses (kWh)
1	0.08	0.75	6158	5120	2.82	2.80	1.95	1.94	11398	11306	21512	21068	0	0
2	0.06	0.67	5515	3882	2.20	2.18	1.21	1.23	7570	7599	26313	25714	0	0
3	0.08	0.81	6161	6944	2.81	2.82	1.94	1.96	11344	11193	0	0	27518	25298
4	0.08	0.94	4236	3882	2.02	1.88	1.27	1.32	7575	7526	0	0	31856	31138
5	0.20	3.25	7840	8672	11.62	10.95	9.47	9.51	29807	29521	0	0	0	0
6	0.37	3.08	1476	8672	11.87	11.34	9.69	9.79	21679	22309	0	0	0	0
7	0.47	2.66	1540	8312	7.34	7.36	1.86	1.73	23539	24304	0	0	4093	5629
column->	1	2	3	4	5	6	7	8	9	10	11	12	13	14

Table 5 Load at HQ's peak and annual energy consumption for the entire single-family building stock

Segment	Elec. load of all houses at HQ peak (MW)	Elec. load predicted by archetypes at HQ peak (MW)	Total energy consumption of all houses (TWh)	Total energy consumption predicted by archetypes (TWh)
1	61.1	60.8	1.0	1.0
2	74.5	75.7	2.1	2.1
3	145.4	146.9	2.7	2.9
4	27.7	28.8	0.8	0.9
5	8314.5	8349.6	25.9	26.2
6	1968.7	1989.1	4.5	4.4
7	210.4	195.6	3.4	3.1
Total	10802.2	10846.5	40.5	40.6
Difference		-0.41%		-0.27%

As shown in Table 5, the prediction of the electrical load during Hydro-Québec's 2017 peak using the archetypes is within 0.41% of the value using the actual building stock for the seven segments. As for the total annual energy consumption, the prediction is within 0.27% of the building stock value.

All these results indicate that the seven archetypes can be used with confidence to predict the electrical peak and the total annual energy consumption of the entire stock. The main characteristics of the seven archetype houses are given in Table A-1 in the Appendix.

Implementation in TRNSYS

Once the archetypes have been selected, the corresponding TRNSYS assembly is extracted from the data base of houses of the QSFBSEM. Each archetype is also associated with its own sub-archetype for $\overline{Internal\ Gains}_{i,arch}$, $\overline{Occupant\ load}_{i,arch}$, and $\overline{DHW\ load}_{i,arch}$ calculated at 15 minute time intervals.

Figure 4 shows a simplified version of the resulting TRNSYS assembly for segment#6. The heating/cooling loads are calculated in the building model (Type 56 in TRNSYS) based on the Montréal weather and $\overline{Internal\ Gains}_{i,arch}$ and $\overline{Occupant\ load}_{i,arch}$ from a user file generated during the CDF calculations (Equation 2). The $\overline{DHW\ load}_{i,arch}$ data is transferred directly to the "Energy calculations" calculator where the energy calculations related to the heating/cooling systems are performed separately. It is then relatively easy to modify these energy calculations to study other heating systems.

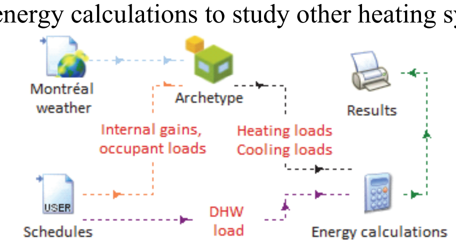

Figure 4: Typical TRNSYS assembly

Simulating the seven archetypes with a 15 minute time step takes about 1 minute of computer time. This is about three orders of magnitude faster than calculating the 200 000 houses of the QSFBSEM. The results obtained for one archetype are then multiplied by the number of houses in that segment to obtain the energy consumption of that segment.

Application

The use of archetypes facilitates the analysis of various changes to the building stock. In this section, a change often discussed in the province of Québec is presented. It concerns the replacement of natural gas for space heating and water heating (see segments #1 and #2 in Table 3). Five replacement scenarios are examined for space heating:

a) electric heating (with a 100% efficiency)
b) regular heat pumps
c) cold-climate heat pumps
d) ground-source heat pumps (GSHP)
e) hybrid heat pump system

For water heating, natural gas is replaced with an electric hot water tank.

Scenarios b, c, and d have electric backup when capacity is insufficient at cold external temperatures. In the hybrid scenario, a regular heat pump is used (including electric backup if required) down to an external temperature of - 12 ℃ at which point a gas furnace is employed (with an efficiency of 90%) for space heating.

The performance of the heat pumps used for this analysis are shown in Figure A-1 in the Appendix. Figure A-1a presents the coefficient of performance (COP) while Figure A-1b gives the fraction of the heating load met by the heat pump both in terms of the external temperature. Data for the regular heat pump are those used by Neale (Neale, 2021). As shown in Figure A-1b, it is assumed that the capacity of a regular heat pump is zero below -12 ℃. The performance of the GSHP is based on a commercial model. Both the capacity and COP values of the GSHP, which are typically dependent on the return temperature from boreholes, were converted to values dependent on the external temperature based on a recent study (Viviescas & Bernier, 2023). This approximation implies that simulation of boreholes is not required when simulating GSHP. Finally, the cold-climate heat pump performance has been adapted from data of a commercial unit.

In addition to energy consumption and electrical load prediction, the results of the next section include the decrease in greenhouse gas (GHG) emissions resulting from the five replacement scenarios. These calculations are based on emissions factors given in Table 6. It is further assumed that the emission factors for electricity would not be affected by an increase (or decrease) in the total grid load.

Table 6 : Emissions factors for GHG emissions

	Value	Units	Reference
Electricity	0.002 to 0.026 [1]	kgCO2eq/kWh	TEQ (2023) and Neale (2021)
Natural gas	49.86	kgCO2eq/GJ	TEQ (2023)
Heating Oil	71.032	kgCO2eq/GJ	TEQ (2023)
Refrigerant leaks	315 [2]	kgCo2eq/year	Pistochini et. al. (2022)

[1] The higher value occured at the end of 2017 when HQ imported electricty with a relatively high CO2 content
[2] An annual 5% leak is assumed for a typical 3 kg charge of R-410A (GWP=2100)

Results

Simulation results using the archetypes for segment #1 and #2 are shown in Figures 5 and 6 for the five replacement scenarios. Figure 5 shows the change in the grid load caused by each of the scenarios for the last week of 2017, which recorded very cold temperatures (see top portion of Figure 5). It is to be noted that the curves for the regular heat pump and hybrid heat pump are identical for t < 8623 h (when T_{ext} > -12 °C) and that the electric and regular heat pump curves are the same for t > 8623 h (when T_{ext} < - 12 °C). The maximum grid load increase occurs when electric systems or regular heat pumps are used. It is 951.5 MW at t = 8671.5 h. During Hydro-Québec's peak (t = 8682 h), the increase is 742.2 MW. The hybrid heat pump scenario increases the grid peak by only 43 MW during Hydro-Québec's peak.

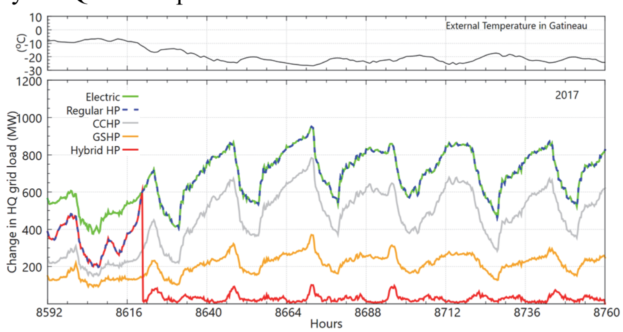

Figure 5: Change in the HQ grid load for all five scenarios during the last week of 2017.

Figure 6 summarises the annual results for all five scenarios in terms of: i) increase in grid load at Hydro-Québec's peak; ii) increase in annual electrical consumption, iii) and decrease in GHG emissions. The GSHP scenario leads to the lowest increase in electrical energy consumption at 0.62 TWh/yr. Scenarios b, c, and d experience almost the same GHG reduction at approximately 370 kt CO2/yr. This is slightly less than the electric heating scenario because of GHG emissions associated with refrigerant leakage for scenarios b, c, and d. The hybrid heat pump scenario is not far behind the others with a reduction of 288 kT CO2/yr.

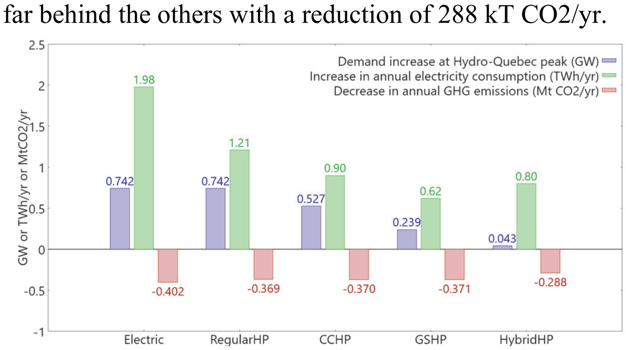

Figure 6: Change in grid load, annual electrical energy consumption and GHG emissions for all five scenarios.

Results presented in this section are only applicable to 2017. Weather conditions and grid demand from other years may lead to different conclusions.

Conclusion

A methodology to determine province-wide archetype houses in Québec has been presented. The database of the Québec single-family building stock energy model (QSFBSEM) is first divided into seven segments (see Table 3). The segmentation is based here on the type of heating system, but the methodology is adaptable to other types of segmentation. Each segment has its own archetype (which are summarised in Table A-1) and sub-archetypes for internal gains, occupant load, and DHW load.

The selection of the archetype house for each segment is based on the use of cumulative distribution functions (CDF) for the heating and cooling (HC) loads and for the 200 highest electrical loads during the year. Cumulative differences between individual CDF profiles and the average CDF profile are calculated for both the HC loads and the highest electrical loads (see Figure 2). The house that has the minimum weighted cumulative difference value is the archetype house.

It is shown that the seven archetypes can be used to accurately predict the performance of 1.4 million homes both in terms of the annual electrical energy consumption (within 0.27%) and peak electrical demand (within 0.41%). The use of archetypes reduces the calculation time by three orders of magnitude when compared to a full simulation of the entire QSFBSEM.

In the final part of the paper, the usefulness of the archetypes is shown by examining five different scenarios to replace gas heating systems in Québec. As shown in Figure 6, the use of hybrid heat pumps would lead to the lowest increase in electrical demand (43 MW at Hydro-Québec's peak) while the use of ground-source heat pumps would lead to the lowest increase in annual energy consumption (0.62 TWh/yr).

Future work

While the archetype methodology developed here is accurate, there is room for possible improvements. For example, the weighting factors presented in Equation 2 might not be optimum in all situations.

In addition, more work is needed to determine if the 200 highest electrical load values used for the cumulative distribution function captures enough of the peaks in the load profile.

Finally, the approach used here should also be tested in conditions where cooling loads are dominant.

Acknowledgement

The financial aid provided by IVADO (Institut de VAlorisation des DOnnéees) in the initial stages of this work is gratefully acknowledged.

References

CMHC. (2004). Stock Aggregation - Methods for Evaluating the Environmental Performance of Building Stocks. Annex 31. https://www.iisbe.org/annex31/pdf/L_stock_aggregation.pdf.

Dahlström, L., Johari, F., Broström, T., & Widén, J. (2024). Identification of representative building archetypes: A novel approach using multi-parameter cluster analysis applied to the Swedish residential building stock. *Energy & Buildings*, 303, 113823.

Dall'o', G., Galante, A., & Torri, M. (2012). A methodology for the energy performance classification of residential building stock on an urban scale. *Energy and Buildings*, 48, 211–219.

De Jaeger, I., Reynders, G., Callebaut, C., & Saelens, D. (2020). A building clustering approach for urban energy simulations. *Energy and Buildings*, 208. 109671.

Ghiassi, N., & Mahdavi, A. (2017). Reductive bottom-up urban energy computing supported by multivariate cluster analysis. *Energy and Buildings*, 144, 372–386.

Neale, A. (2021). Development of a bottom-up white-box residential building stock energy model. Thèse de doctorat, Polytechnique Montréal, Génie mécanique.

Neale, A., Kummert, M., & Bernier, M. (2020). Development of a stochastic virtual smart meter data set for a residential building stock–methodology and sample data. *Journal of Building Performance Simulation*, 13(5), 583–605.

Neale, A., Kummert, M., & Bernier, M. (2022). Discriminant analysis classification of residential electricity smart meter data. *Energy and Buildings*, 258, 111823.

Pistochini, T., Silveira, R., Ross, D., Mcmurry, R., & Umarani, A. (2022). Analysis of Refrigerant Use for AC Equipment Maintenance. *ASHRAE Journal*, 64(10), 44–54.

Reinhart, C. F., & Cerezo Davila, C. (2016). Urban building energy modeling – A review of a nascent field. *Building and Environment*, 97, 196–202.

Reyna, J., Wilson, E., Parker, A., Satre-Meloy, A., Egerter, A., Bianchi, C., Praprost, M., Speake, A., Liu, L., Horsey, R., Dahlhausen, M., Caradonna, C., & Rothgeb, S. (2022). U.S. Building Stock Characterization Study: A National Typology for Decarbonizing U.S. Buildings. https://www.nrel.gov/docs/fy22osti/83063.pdf.

Swan, L. G., & Ugursal, V. I. (2009). Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and Sustainable Energy Reviews*, 13(8), 1819–1835.

TEQ-Transition Énergétique Québec. (2023). Facteurs d'émissions et de conversion. https://transitionenergetique.gouv.qc.ca/fileadmin/medias/pdf/FacteursEmission.pdf

Theodoridou, I., Papadopoulos, A. M., & Hegger, M. (2011). Statistical analysis of the Greek residential building stock. *Energy and Buildings*, 43, 2422–2428.

Viviescas, G., & Bernier, M. (2023). A variable speed water-to-water heat pump model used for ground-source applications. *Science and Technology for the Built Environment*, DOI: 10.1080/23744731.2023.2295822

Wilcox, S., & Marion, W. (2008). Users Manual for TMY3 Data sets. Technical Report NREL/TP-581-43156.

Appendix A

Table A-1 Caracteristics of the seven archetypes

Parameters	Unit	1	2	3	4	5	6	7
Location	-	Gatineau	Gatineau	Montréal	Montréal	Montréal	Montréal	Montréal
Building type	-	Single-detached	Single-detached	Single-attached	Row	Single-detached	Single-detached	Single-detached
Load profile[1]	-	average of all houses	average of all houses	average of all houses	average of all houses	average of all houses	average of all houses	average of all houses
Window type[2]	-	200002	200002	200012	202012	213204	300002	200002
Heated floor area	m^2	220.4	158.5	160.2	206.2	214.5	203.2	164.8
Window-to-wall ratio	-	0.20	0.15	0.10	0.15	0.20	0.20	0.15
Building rotation	(°)	0	270	90	0	180	270	270
# occupants[1]	-	average of all houses	average of all houses	average of all houses	average of all houses	average of all houses	average of all houses	average of all houses
Adjacent buildings	-	No buildings adj.	No buildings adj.	East side adjacent	Both sides adjacent	No buildings adj.	No buildings adj.	No buildings adj.
# floors	-	1-storey	1-storey	2-storeys	2-storeys	1-storey	1-storey	1-storey
Wall thermal res.	m^2K/W	2	2	1	1	2	1	2
Roof thermal res.	m^2K/W	5	2	8	2	2	8	2
Found. thermal res.	m^2K/W	1	1	1	1	1	1	1
Leakage area	cm^2	406	775	556	775	556	556	406
Air conditioning	-	yes	yes	yes	yes	yes	yes	no
Heating type	-	Natural Gas	Natural Gas	Heating Oil	Heating Oil	Electric	Heat Pump	Heating Oil/Electric
DHW energy source	-	Electric	Natural gas	Electric	Oil	Electric	Electric	Electric
Aspect ratio	-	1.1	1.1	1	0.8	1	0.9	1
Pool	-	No pool	Pool	No pool	No pool	No pool	No pool	No pool
Spa	-	No spa	No spa	No spa	No spa	No spa	No spa	No spa

[1] Each archetype has its own sub-archetypes for Internal Gains, Occupant load, and DHW load

[2] See Table 6 in Neale et al. (2020)

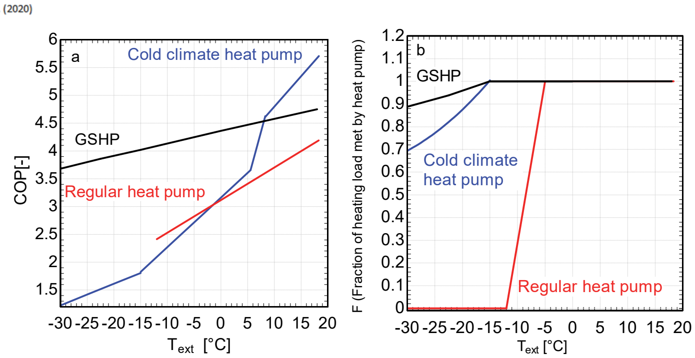

Figure A-1 COP and capacity of heat pumps used in the present study