eSim 2018
«BUILDING SIMULATION TO SUPPORT BUILDING SUSTAINABILITY»
IBPSA Canada
The Canadian Chapter of the International Building Performance Simulation Association
Le chapitre canadien de International Building Performance Simulation Association

Generator: a stochastic virtual smart meter data generation model for residential building stock characterization

Adam Neale, Michaël Kummert, Michel Bernier[1]

[1] Polytechnique Montréal, Montréal, Canada.

Abstract: *Smart meters are becoming prevalent for the purpose of electricity consumption billing. In the province of Québec alone there are over 3.5 million smart meters collecting data at fifteen-minute intervals. In this paper a series of building energy simulations are used to produce virtual smart meter (VSM) data sets. This represents a first step towards using real smart meter data for building stock characterization, by developing an inverse model and training methodology in a controlled environment. The inputs for the VSM model are based on single family home building stock characteristics found in the province of Québec. Parameters for each simulation are randomly selected based on probability mass functions, which allow for realistic, statistically representative VSM data sets to be produced using building energy simulations. A data set of 100,000 VSM profiles is targeted, which will require big data analysis techniques for visualization and processing.*

Keywords: *residential, smart meter data, building stock modeling, building energy modeling*

INTRODUCTION

Residential buildings consume roughly 17% of all energy in Canada, almost double the energy consumption of commercial buildings (Natural Resources Canada, 2014). Understanding the energy consumption of buildings on a large scale is critical to many fields. For example, a government may wish to know what the impact would be if the minimum standard for insulation for homes were raised, or if the efficiency of heating equipment were increased in a certain number of homes. Alternatively, a new community may be planned on the outskirts of a city, and the local utility must plan the expansion of the electric distribution grid for the homes being built there. In those instances, the annual, monthly, daily or hourly energy consumption could be required. National and provincial average annual energy consumption for residential buildings is readily available, but beyond that a specific case-by-case analysis is typically required.

Depending on the information available, approximating the energy consumption of a building is not difficult at the individual building level. Building energy modeling is a field that has been widely researched and is used for many applications in industry (Higgins, 2012). For building stock energy modeling the difficulty lies in approximating the energy consumption in many buildings, such as at the community, city, or provincial levels. It becomes no longer practical to model each building individually and additional techniques must be used to reduce the dimensionality of the problem to a scale where simulation is possible.

The present study focuses on a new method for characterizing the single-family home (SFH) residential building stock in the province of Québec in terms of *building archetypes*. The literature review will illustrate briefly how current building archetype development techniques lack the capability to accurately estimate building stock energy consumption. A methodology to extract building stock parameters from smart meter data is presented.

This paper presents the first main component of the model used to generate virtual smart meter (VSM) data with some preliminary results, which represents the first step towards using real smart meter data for building stock characterization. Note that only buildings with electric systems are considered in this study.

BUILDING STOCK MODELING

A building stock is defined here as a group of buildings sharing some common root characteristic. This could be the same building type (i.e. residential) and/or the same geographic location (i.e. a community, city or province). A building stock does not need to represent all types of buildings in the same geographic location. For example, the energy consumption of all residential buildings in a province could be required for a demand-side management effort by a local energy distributor.

The method used to predict the energy consumption of a building stock depends on the information required. Swan and Ugursal (2009) performed a review of residential building stock energy modeling techniques. Two principal techniques for representing building stock energy use were illustrated: *top-down* and *bottom-up*.

Top-down models start at the figurative “top level” of the building stock, represented by the average energy consumption of all buildings of that type. The buildings could then be categorized in various ways until the individual building level is reached. Bottom-up models

ISBN 978-2-921145-88-6

start with individual buildings to characterize the energy use of a larger group of buildings (for example, a community).

The bottom-up category of stock modeling can be further divided into two subcategories: *statistical models*, which are based on the real energy use of buildings, and *engineering models*, which are based on physical processes in buildings. Building *archetypes* are one method falling under the engineering model category, which are used for stock modeling purposes (Swan and Ugursal, 2009).

Building archetypes

Building archetypes are templates used to simplify the energy consumption prediction for building stocks. In Europe the term *building typology* is used synonymously with building archetype (Loga, Stein and Diefenbach, 2016). Building archetypes are intended to contain characteristics that allow an energy modeler to represent the buildings within a stock without representing each individual case.

In order to represent buildings as archetypes, a building stock must go through a *segmentation* and *characterization* process (Sokol *et al.*, 2016). While the term *classification* has been coined for building archetypes by some authors, in this study the term *segmentation* will be used to avoid confusion with the machine learning process known as *classification*.

Segmentation is the process of categorizing buildings based on their physical parameters. For example, some archetypes segment stocks based on building age, HVAC systems, building shape, etc. (Reinhart and Cerezo Davila, 2016). The process of dividing the building stock depends on the parameters and the information available.

Characterization defines the approach used to assign values to the segmentation variables for a given building stock. Parameters are characterized either with information from the literature, such as the Canadian Survey of Household Energy Use, or taken from actual building data, which represents only a sample of the larger building stock (Sokol *et al.*, 2016).

The accuracy of a building archetype depends on the information that is available. Works such as those by Sokol et al. (2016) introduce additional techniques such as Bayesian calibration to further improve the accuracy of archetypes, because using the available deterministic building properties found in the literature was deemed insufficiently accurate.

Archetypes have been found to accurately estimate building stock annual energy consumption, with errors ranging between 4% to 21% when compared to real building energy consumption (Reinhart and Cerezo Davila, 2016). Peak loads of building stocks are not well estimated by building archetypes however, with errors on the order of 40% (Heiple and Sailor, 2008).

Archetypes tend to be only as good as the information available and are generally only accurate for annual energy estimation.

PROPOSED METHODOLOGY

A coupled building energy modeling approach is proposed for the purpose of extracting building stock characteristics from high-resolution smart meter data. As shown in Figure 1, this is achieved using three branches referred to as *Generator*, *Classifier*, and *Extractor*. Each component of the model is described in further detail in the following sections for overall context. Note that in the following sections, number/letter combinations in curled braces {XX} refer to the corresponding process in Figure 1.

Generation of VSM data (*Generator*)

The first step of the approach, described in detail in this paper, is to produce a set of stochastically-generated virtual smart meter (VSM) data profiles. A VSM data profile is defined here as a set of electricity consumption data generated with a model at 15-minute time steps using realistic building characteristics as inputs, which are selected at random. A sufficient number of profiles will be generated to be statistically representative of the stochasticity of the buildings in the selected building stock without being redundant. The exact number of profiles produced will be determined as a function of the diversity in building characteristics for the studied building stock and on the computational resources available. The preliminary targeted data set will consist of 100,000 distinct VSM profiles, though only a subset of these profiles are presented in this paper.

The model inputs for each profile are randomly generated {1A} based on variables with probability distributions that represent the single family home (SFH) residential building stock of the province of Québec. For example, the building's location is determined at random based on the statistical distribution of SFH in the major regions of the province and the type of building is selected at random based on the region. Building loads, such as occupancy, lighting and domestic hot water, are also varied stochastically on a sub-hourly basis.

Once randomly selected, the inputs are used in a building energy modeling tool {1B} to determine the corresponding virtual electricity consumption profile, which is saved to a text file {1C}. This process continues until the total number of desired virtual profiles is attained {1D}. The full set of

ISBN 978-2-921145-88-6

profiles with their inputs are combined into a single data set for further analysis {1E}.

Classification of the VSM data set (*Classifier*)

The generated VSM data profiles are analysed using supervised machine learning algorithms {2A} to establish inverse models between the quarter hour data and the known inputs, such as the approximate building surface area, the window type, location, etc. The inverse models developed based on the VSM data are used to classify the inputs for the Extractor model {2B} and develop initial guesses for each parameter. This greatly reduces the number of simulations required for each set of real smart meter data by narrowing the range of possibilities for the input parameters in the *Extractor* component. The model does not need to iterate over the full range of possibilities for every parameter. The possibility of using artificial neural networks to develop a more complete inverse model for the building simulation inputs will also be investigated.

The full details of this component will be presented in a future paper.

Extract building parameters from smart meter data (*Extractor*)

The *Extractor* component is still in the concept phase of development and is presented here for context. The modeling approach used to generate VSM data will be applied to real smart meter (RSM) data {3A}. Model inputs will be initially based on the inverse models developed in the previous component, which will be used to generate educated guesses for the simulation parameters {3B}.

As with the *Generator* function, a building simulation is performed to generate a VSM profile using the best estimates for the inputs {3C}. The resulting electricity consumption profile is then compared to the RSM data based on sub-hourly (15-minute time steps), hourly, daily, monthly and annual metrics {3D}. Success criteria such as those presented in ASHRAE Guideline 14 will be evaluated, such as mean bias error (MBE) less than 10% and CV(RMSE) less than 30% for hourly data (ASHRAE, 2014). Differences in the profiles are evaluated based on past experience of the model (improvement or worsening from previous iterations) and are interpreted automatically {3E}. New model inputs are generated and the building simulation is repeated.

The process of generating new inputs, generating VSM data and comparing to the RSM data continues until the relative error between the VSM and RSM data fall below a certain threshold or a certain number of iterations is exceeded. The final set of input parameters extracted from the RSM data are then saved {3F} and further data can be

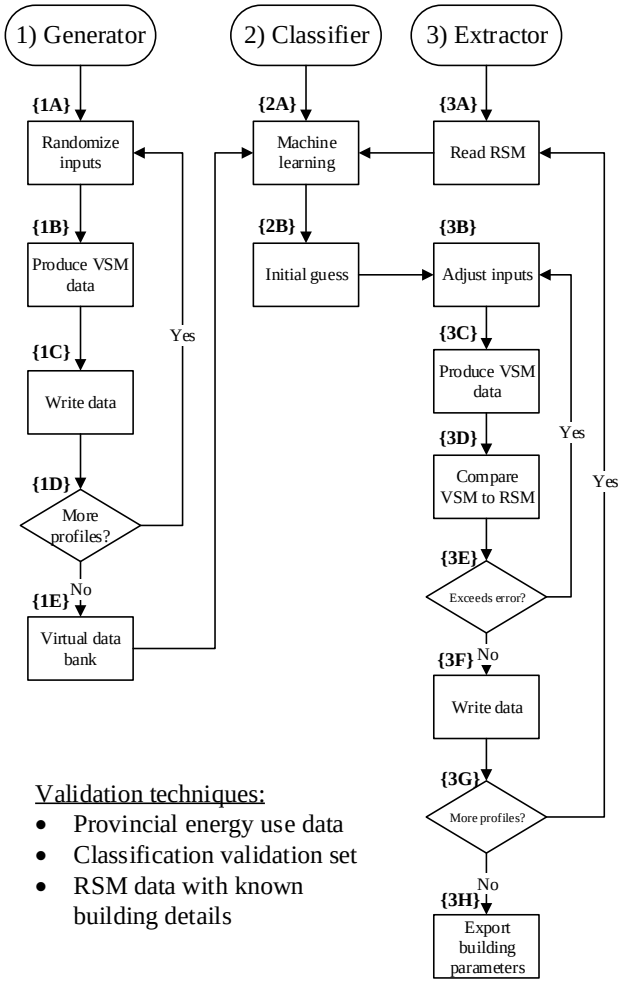

Figure 1: Overview of the model components

analyzed if required {3G}. The complete set of building parameters are combined into a single data set for further analysis {3H}.

The proposed methodology will be applied to RSM data sets with known building characteristics to validate the approach.

GENERATOR COMPONENT DESCRIPTION

This section describes the approach used to develop a model to simulate a residential building. Under normal circumstances a user would vary building parameters directly within a building simulation tool and that would be sufficient for the purposes of modeling energy consumption of the building. However, in this case the number of buildings studied (up to 100,000) and the stochastic nature of the inputs requires different approaches for generating building input files and post processing. For this reason, a coupled modeling approach using Matlab as the control program and TRNSYS (Klein *et al.*, 2017) as the building simulation software was selected.

ISBN 978-2-921145-88-6

Table 1: Model inputs (PMF: probability mass function, TBD: to be determined, UPD: uniform probability distribution, NYI: not yet implemented, SHEU: survey of household energy use).

Property	Distribution	Dependencies	Data source	Description (type of data contained)
1. Location	PMF	N/A	Census (StatCan, 2016), SHEU (Natural Resources Canada, 2011)	Region in Québec
2. Weather	N/A	▪ Location	SIMEB Database (Hydro-Québec, no date)	Based on the location selected, CWEC weather file.
3. Building type	PMF	▪ Location	Census (StatCan, 2016), SHEU (Natural Resources Canada, 2011)	Single detached, semi-detached, row, etc.
4. Building age	PMF	▪ Building type ▪ Location	SHEU Table 4.1 (Natural Resources Canada, 2011)	Age brackets, probability. Data seems to be for Canada only.
5. Shape	UPD	▪ Building type ▪ Location ▪ Age?	TBD	Aspect ratio, e.g. 0.8, 1.0, 1.3
6. Rotation	UPD	N/A	N/A	-90° to +90°
7. # Floors	PMF	▪ Building type	SHEU Table 4.1 (Natural Resources Canada, 2011)	One or two floors
8. Renovated	NYI	▪ Building type ▪ Building age	SHEU Table 5.2 (Natural Resources Canada, 2011)	Building renovated yes or no? To what effective age?
9. Wall construction	TBD	▪ Building age ▪ Location ▪ Renovated	TBD	Brick, stucco, insulation level, 2x4, 2x6, etc.
10. Roof construction	TBD	▪ Building age ▪ Location ▪ Renovated	TBD	Attic, cathedral, flat, etc
11. Window type	PMF	▪ Building age ▪ Renovated	SHEU Table 4.5 (Natural Resources Canada, 2011)	Single, double, low-e, etc
12. WWR	UPD	▪ Building type	TBD	e.g. 0.1, 0.2, etc.
13. Basement	NYI	▪ Building type ▪ Location ▪ Building age	SHEU Table 4 (Natural Resources Canada, 2011)	Yes/no/crawl space/…
14. HVAC	NYI	▪ Building age ▪ Building type	SHEU Table 6 (Natural Resources Canada, 2011) SHEU Table 7 (Natural Resources Canada, 2011)	Electric baseboards, central air, heat recovery ventilator, air conditioning
15. Setpoints	NYI	N/A	SHEU Table 8.2 (Natural Resources Canada, 2011)	Thermostat setpoints for heating and cooling
16. Infiltration	TBD	▪ Building age ▪ Location ▪ Weather ▪ HVAC	TBD	Rate of infiltration
17. Appliances	Stochastic	▪ Building type ▪ Occupants	CREST demand model (McKenna and Thomson, 2016)	Number and type of appliances
18. Lighting	Stochastic	▪ Building type	CREST demand model (McKenna and Thomson, 2016)	Type of lighting, density
19. Appliance usage	Stochastic	▪ Appliances ▪ Occupancy	CREST demand model (McKenna and Thomson, 2016)	Equipment electricity consumption
20. Lighting usage	Stochastic	▪ Lighting ▪ Occupancy	CREST demand model (McKenna and Thomson, 2016)	Lighting electricity consumption
21. Occupants	Stochastic	▪ Building type ▪ Location	CREST demand model (McKenna and Thomson, 2016), Census (StatCan, 2016), SHEU (Natural Resources Canada, 2011)	Number of occupants
22. Occupancy	Stochastic	▪ Occupants ▪ Location ▪ Building type ▪ Time ▪ Day	CREST demand model (McKenna and Thomson, 2016)	Occupant schedule and activity schedule.
23. DHW consumption	Stochastic	▪ Occupants ▪ Occupancy ▪ Time ▪ Day	CREST demand model (McKenna and Thomson, 2016)	Volume of hot water

ISBN 978-2-921145-88-6

Table 2: Household probability mass functions by region and by building type

City	Number of households						Probability distributions					
							STEP1:	STEP2: Determine the building type by region				
	SFH	DET1	DET2	Row	SDH	OSA	Determine the location	SFH	DET1	DET2	Row	SDH	OSA
1. Rimouski	16,505	10,650	2,940	620	2,150	145		1.4%	64.5%	17.8%	3.8%	13.0%	0.9%
2. Saguenay	44,200	28,911	7,979	1,585	5,460	265		3.7%	65.4%	18.1%	3.6%	12.4%	0.6%
3. Quebec City	180,375	116,759	32,226	8,920	21,130	1,340		14.9%	64.7%	17.9%	4.9%	11.7%	0.7%
4. Sherbrooke	49,770	33,405	9,220	2,450	4,325	370		4.1%	67.1%	18.5%	4.9%	8.7%	0.7%
5. Victoriaville	13,495	8,664	2,391	190	2,205	45		1.1%	64.2%	17.7%	1.4%	16.3%	0.3%
6. Trois-Rivières	40,715	26,426	7,294	1,845	4,785	365		3.4%	64.9%	17.9%	4.5%	11.8%	0.9%
7. Shawinigan	15,015	10,792	2,978	210	710	325		1.2%	71.9%	19.8%	1.4%	4.7%	2.2%
8. Drummondville	25,235	17,382	4,798	800	2,050	205		2.1%	68.9%	19.0%	3.2%	8.1%	0.8%
9. Granby	20,765	13,570	3,745	315	2,635	500		1.7%	65.3%	18.0%	1.5%	12.7%	2.4%
10. Montreal	713,720	442,191	122,044	56,765	86,455	6,265		59.0%	62.0%	17.1%	8.0%	12.1%	0.9%
11. Gatineau	90,850	50,713	13,997	7,950	17,895	295		7.5%	55.8%	15.4%	8.8%	19.7%	0.3%
Total	**1,210,645**	**759,464**	**209,611**	**81,650**	**149,800**	**10,120**							

SFH: Single-family home, DET1: 1-storey detached home, DET2: 2-storey detached home, Row: row house, SDH: semi-detached home, OSA: other single-attached home.

Building simulation inputs

The complete list of input parameters for the building simulation are described in Table 1. It is important to note that all of the illustrated parameters are implemented in the building simulation but some are fixed values that do not yet represent the Quebec residential building stock. This is often due to the fact that an appropriate source of data for the parameter has not yet been identified (items marked *TBD (to be determined)* or *NYI (not yet implemented)* in Table 1). Otherwise, parameters with specified distributions such as *PMF (probability mass function)* or *Stochastic* will be described in further detail later in the paper. Some dependencies are also illustrated in the table, such as how the distribution of building types depends on the location of the building. In the absence of data in the literature, best available values will be used based on best practice information, codes and standards or other sources.

Stochastic building stock parameters

For a given simulation, i.e. the generation of one profile, a series of parameters must be set for the model to produce an output. These parameters must be selected in a specific order as there are some interdependencies. For example, depending on the region the ratio between the different types of dwellings (detached homes, semi-detached, and row houses) can vary. The same holds true for the weather, which varies based on the location.

As a general rule, each input for the model can be described as a discrete random variable based on the studied building stock. It should be noted that all of the inputs will eventually be varied stochastically for the *Generator* and that the inputs above were selected simply as a preliminary proof-of-concept.

As an example of the approach used for stochastic building parameters, the location and building type probability mass functions are illustrated in Table 2. The first step is to determine the building location, followed by the type of the randomly generated building (detached, semi-detached, row, etc.).

Stochastic internal loads

Building internal loads are largely occupancy driven – lights, appliances, and domestic hot water all depend on when the occupants are home and how many occupants are in the house. The objective of producing stochastic VSM profiles led to an investigation into existing models that could represent these loads stochastically so that they may be used as inputs in a building simulation. The CREST demand model was selected as an appropriate tool for that purpose (McKenna and Thomson, 2016).

The CREST model is an open-source Excel-based spreadsheet model of a residential building. The model includes a complete thermal model of a building and all buildings loads. There are several limitations to the model, including the fact that it can only calculate the loads of the building for a single day of the year that the user specifies. The stochastic loads are based on a detailed time-of-use data survey sent to building occupants in the United Kingdom. Based on the collected data, probabilities for a number of different parameters were produced:

ISBN 978-2-921145-88-6

- Occupancy profile: number of occupants at home;
- Activity profile: when each occupant is active;
- Lighting profile: lighting power;
- Appliance profile: appliance power, based on the probability of appliances being used at a given time of day;
- Domestic hot water profile: domestic hot water volume used based on occupancy.

In order to produce inputs for the building simulation, the CREST model had to be modified to produce a full year's worth of loads. Each day in the annual profile is distinct and all load profiles are unique. Two separate load profile documents were produced by the modified CREST model: 1) occupancy, activity, lighting and appliance data at fifteen-minute intervals, and 2) domestic hot water consumption at one-minute intervals.

Each load profile produced represents one household with 1 to 5 occupants (the CREST model is limited to a maximum of 5). The number of occupants was selected at random based on a probability mass function specified by the user. The distribution of occupants in residential buildings in the province of Québec was extracted from Statistics Canada data and inserted into the CREST model (StatCan, 2011). The occupancy data is summarized in Table 3.

The load profiles are generated when the modified CREST model is executed. Note that the thermal model component of CREST was not used for this project. An example of the load profiles produced with the modified CREST model, the occupancy, activity and DHW profiles for two consecutive days are illustrated in Figure 2.

Table 3: Occupancy distribution (StatCan, 2011)

Category	Quebec	Fraction
Total households	3,395,340	-
1-person households	1,094,410	0.322
2-person households	1,181,240	0.348
3-person households	496,140	0.146
4-person households	421,080	0.124
5-person or more households	202,475	0.060
Total persons in households	7,732,830	-
Average occupants in households	2.3	-

BUILDING ENERGY MODEL

The *Generator* model requires two tools: Matlab and TRNSYS. In brief, Matlab prepares the model inputs, writes the necessary files for the building simulation, and then TRNSYS performs specific building-related simulations to write outputs that will subsequently be used within the Matlab environment. This process greatly simplifies the process of running batch simulations where many inputs must be varied.

TRNSYS is a component-based modeling tool that links a number of components (referred to as "*Types*") together with connectors. A *Type* is a system of equations that define some relationship between inputs and outputs, representing any kind of process or piece of equipment, such as a pump, a building, a heating element, a set of controls, etc.

Two types of simulations are performed: domestic hot water (DHW) and building energy use.

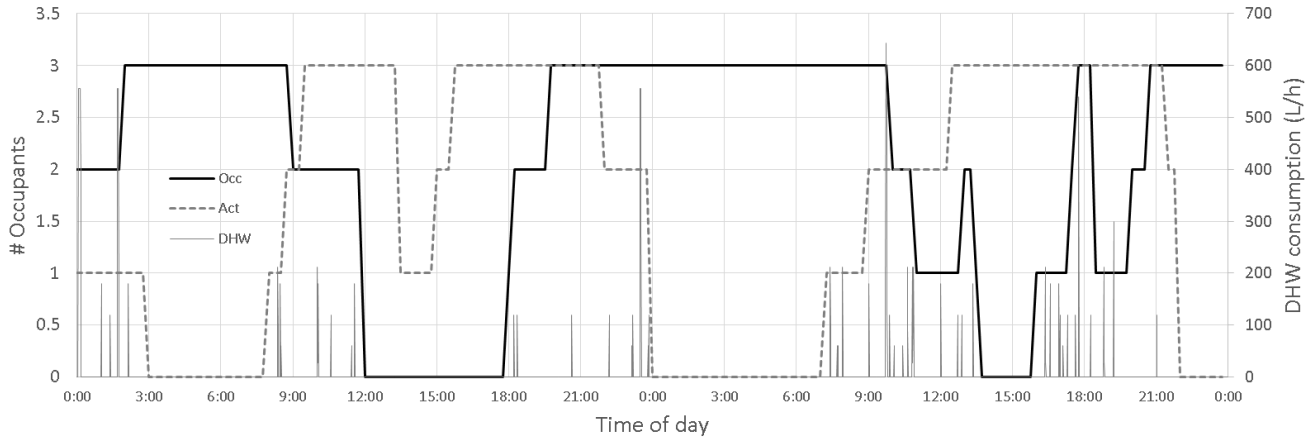

Figure 2. Example illustrating occupancy, activity and domestic hot water (DHW) profiles for two days of the year.

ISBN 978-2-921145-88-6

Domestic hot water

The DHW model included a hot water tank heated by two electric elements (typically 4.5 kW each) controlled by aquastats at different heights in the reservoir. The heating elements operated based on a master-slave control scheme that prevented both from activating at once, with priority to the top element.The water heater energy consumption was calculated based on the annual hot water demand generated using the CREST demand model. The weather file for the location of the building was also used for the city mains water temperature. The DHW model generated electricity consumption at 15-minute intervals based on the hot water demand.

Building model

The key element of the model is the "Building" component, which is *Type 56* within the TRNSYS environment. This *Type* allows a user to specify a building file to be used with customized inputs and outputs, based on the user's needs. In this case, several approaches were possible for how the building files were developed.

Without entering into great levels of detail, the TRNSYS model allows for a user to specify the input parameters presented previously in Table 1. For example, the weather file in the *Weather* type, the building geometry and thermal characteristics in the *Building* type, etc. Some are allocated directly in the TRNSYS environment, such as the building rotation in a user-defined "Rotation" equation editor, while others are specified in the building file using a TRNSYS editing tool called *TRNBuild*. Buildings are specified in *building files* (*.b18 extension) and the connections and simulation parameters for the model are specified in a *deck file* (*.dck extension).

The buildings modeled in TRNBuild are simple one-zone buildings that are allocated dimensions based on the randomly selected properties for each simulation. For example, the randomly-selected total surface area of the home is translated into a heated surface area per floor, depending on whether it is a one- or two-storey home. The building perimeter, height, window surfaces and other dimensions are then calculated based on the surface area, aspect ratio of the building footprint, window-to-wall ratio (WWR) and other relevant data. The wall external surfaces are then specified with the appropriate dimensions to properly represent the home in three dimensions.

VSM data generation

The virtual smart meter (VSM) profiles were stochastically generated based on the available building stock characteristics and are therefore statistically representative of a single family home (SFH) electricity consumption in the province of Québec. The purpose of the virtual smart meter profiles and the associated model is to be able to generate a large quantity of data for trend analysis and classification. Given that the input parameters for the model are known, unlike for most real smart meter data, direct links between the electricity consumption and the building characteristics can be established. These links can subsequently be used to extract building parameters from real measured smart meter data profiles.

Random number generation

The input parameters for the model were described previously in Table 1. The stochastic inputs were presented in terms of uniform probability distributions and probability mass functions. In either case, each of the stochastic parameters is assigned a distinct random number generated from 0 to 1. This random number is used differently depending on the type of probability distribution, but in both cases the end result must be a discrete integer representing the chosen input category.

Uniform probability distributions (UPD) have equal probability for each possible outcome. The desired output must be a whole number from 1 to *n*, where *n* is the number of distinct categories for that variable. For example, there are five different possible building surface area categories, so the result for this variable must be a whole number from 1 to 5, inclusively.

Probability mass functions (PMF) have unequal probabilities for each outcome and therefore require a different approach than for UPD. In the case of a PMF, such as the building location, a discrete probability distribution (DPD) is created from the individual category probabilities. An example DPD is presented in Figure 3 for the building location. The exact values used were presented previously in Table 2.

In practical terms, the DPD presented in Figure 3 would simply be programmed to verify in which location a building would lie given a specific random number. For example, a value of 0.1 on the x-axis would result in a category number of 3 for the location, which corresponds to Quebec City. A value of 0.8 yields category 10, which corresponds to Montréal.

As described previously, Matlab and TRNSYS are coupled to read an input set, process the inputs, perform domestic hot water and building simulations and ultimately write a VSM data profile. A building file and a deck file are required to run a TRNSYS simulation. These files were created dynamically for each simulation based on the randomly selected input parameters.

ISBN 978-2-921145-88-6

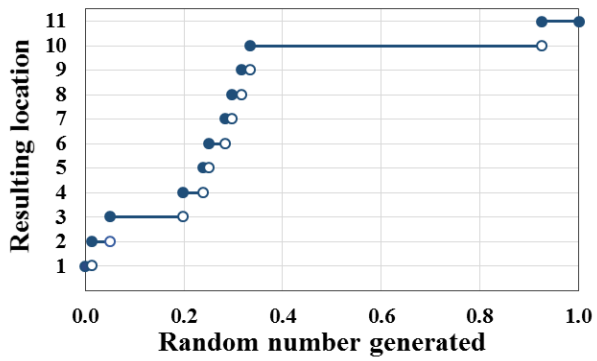

Figure 3: An example discrete probability distribution used to determine the location category number. Closed points represent inclusive values, open points represent exclusive values.

Once the model input files are written with the appropriate parameters, a simulation is executed. The output of the simulation is a VSM profile saved to a text file, which is subsequently read by an external program for post-processing. Each simulation takes approximately 200 seconds to run on a relatively slow computer (Intel Quad CPU Q9400@2.66GHz with 4 Gb RAM, SATA hard drive). For the full set of 100,000 profiles targeted in the project methodology the total simulation time would reach approximately six months. Subsequent to preliminary tests, with parallel processing and a more performant computer the total simulation time can be reduced to the order of 10 days.

Model validation

In order to provide some preliminary validation of the virtual smart meter model, the annual electricity consumption for a number of random cases was broken down into several segments: heating, air conditioning, lighting, appliances and domestic hot water. These values were compared with the provincial averages taken from Natural Resources Canada's National Energy Use Database (NRCan, 2017). The results of the comparison are illustrated in Figure 4.

The results in Figure 4 illustrate that the model produces annual energy consumption values consistent with provincial averages. First of all, the heating and cooling loads seem to be correctly represented by the model and the values scale proportionately based on the building surface area. Lighting and equipment also appear to be the correct order of magnitude. Domestic hot water electricity consumption seems generally slightly underestimated. A further in depth analysis of the CREST DHW consumption values and the TRNSYS DHW simulation will be required to understand the reason for the discrepancy, but given the proof of concept nature of the model at this stage, the

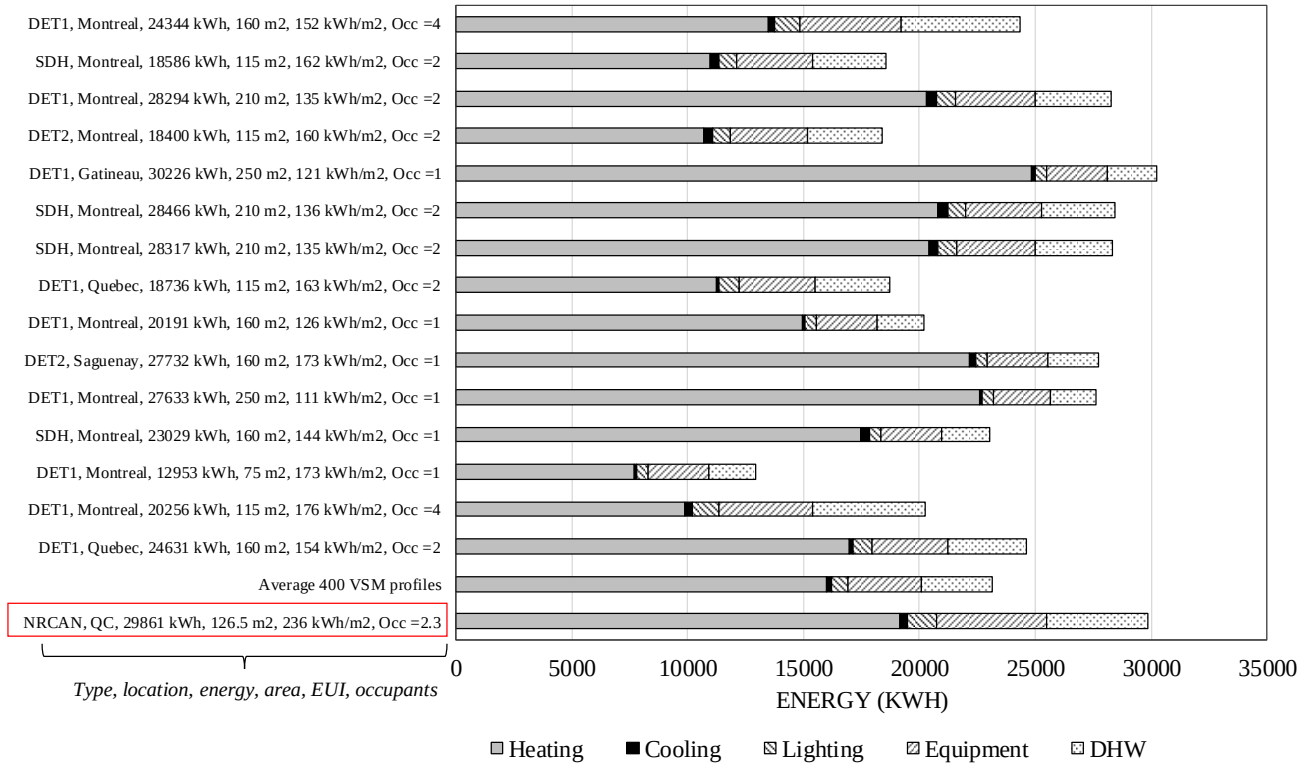

Figure 4: Comparison of annual energy consumption by usage for several VSM cases.

ISBN 978-2-921145-88-6

results were considered acceptable enough to proceed with the generation of a moderate number of VSM profiles for further analysis.

RESULTS

For the purpose of this paper 400 VSM profiles were generated. These profiles were used to examine trends between the input parameters and the resulting electricity consumption for each building and to develop classification models to be used for input category prediction (not presented in this paper).

VSM profiles

In order to visualize the stochasticity of the data set and to demonstrate some of the data variability, the range of possible daily electricity consumption values (from min to max) for the 400 profiles are plotted over the period of one year, as illustrated in Figure 5. The goal of such a graph is to illustrate the "average" profile in addition to highlighting the possible difference in order of magnitude from the lowest to highest energy consumption values. For example, in the winter there was as much as a factor of six difference in the possible daily energy consumption for the virtual buildings. This is quite feasible given the range of possible building parameters specified in the model, in particular due to the building surface area.

Another way of illustrating large quantities of data is through the use of carpet plots, which present the possibility of demonstrating the interaction between several

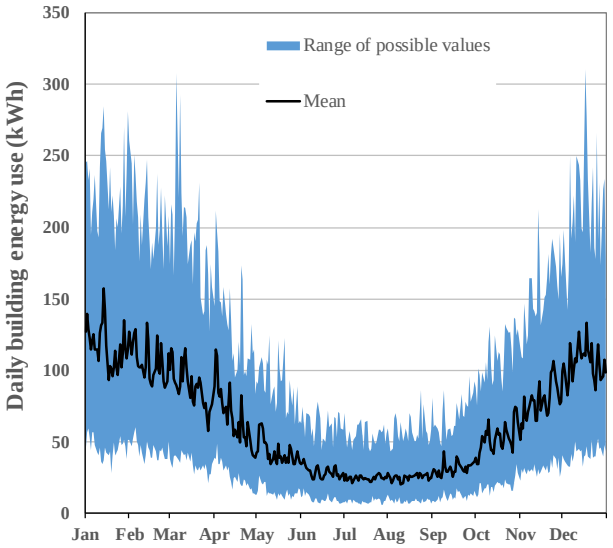

Figure 5: Illustration of the range of possible daily energy consumption values from 400 VSM profiles

variables. An example carpet plot is presented in Figure 6 for the VSM profiles based on daily energy consumption.

The carpet plots in Figure 6 demonstrate how visual tools can illustrate certain trends in large quantities of data, which can be useful for qualitative analysis of the data set. For example, consider the carpet plot in Figure 6(a), which presents the VSM profiles in the original randomized order. The difference in energy consumption magnitude can be observed for the winter (Jan.-April, Oct.-Dec, lighter shade) and summer (May-Sept., darker shade) periods.

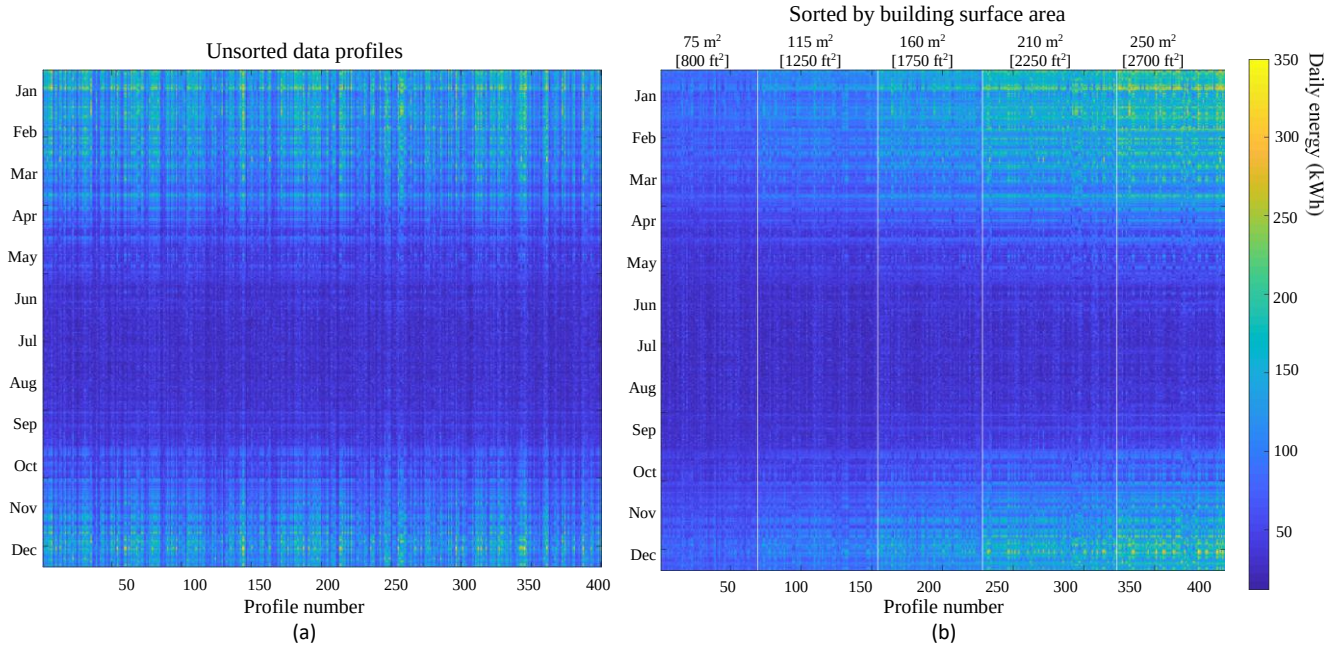

Figure 6: Carpet plots illustrating daily energy consumption for each annual VSM profile with (a) unfiltered data and (b) data organized by increasing building surface area category. Note: white lines illustrate the category boundaries in (b).

ISBN 978-2-921145-88-6

If the profiles are organized based on the input parameters some additional observations can be made. Figure 6(b) presents a carpet plot of the daily energy consumption when the profiles are organized by building surface areas. The white lines in the plot indicate the category boundaries. The categories clearly indicate a trend of increasing energy consumption as the building surface area increases, though still with a high stochasticity level due to the other parameters that are varied. The building surface area clearly has a large influence on the overall building energy consumption, as expected.

CONCLUSION

This paper presented the first step of an overall methodology (shown in Figure 1) to extract relevant building parameters from anonymous smart meter data. The first component of the model, called *Generator*, was shown to stochastically generate virtual smart meter (VSM) data based on known building stock information. The approach presented will be applied to generate a large virtual data set that will be used to link smart meter energy consumption with known building parameters.

A subset of 400 VSM profiles was presented. The energy consumption of the associated buildings was consistent with provincial energy use for similar buildings. The domestic hot water model will require some minor adjustments to more closely match residential hot water consumption in the province of Québec. Otherwise, the energy consumption of the heating, cooling, lighting and appliances was consistent with the chosen building stock.

Ultimately the *Generator* component will be used to produce a large-scale stochastic virtual smart meter data set, which will be used for the archetype development methodology presented in this paper. The VSM data will allow the study of the relationships between energy consumption data at smart meter intervals with known building parameters. There is often a disconnect between smart meter data and the building's characteristics, which prevents supervised machine learning algorithms from being effective. The proposed approach will assist in extracting useful information from anonymous smart meter data.

The authors intend to release the VSM profiles to the public along with their corresponding building parameters once the 100,000 profile data set is complete.

ACKNOWLEDGEMENTS

The financial contribution from the Institute of Data Valorization (IVADO) under its fundamental research program is gratefully acknowledged.

REFERENCES

ASHRAE (2014) *Guideline 14-2004 - Measurement of energy, demand and water savings*.

Heiple, S. and Sailor, D. J. (2008) 'Using building energy simulation and geospatial modeling techniques to determine high resolution building sector energy consumption profiles', Energy and Buildings, 40(8), pp. 1426–1436.

Higgins, J. A. (2012) 'Energy Modeling Basics', ASHRAE Journal. American Society of Refrigerating Engineers, 54(12), p. 26.

Hydro-Québec (no date) SIMEB - Simulation énergétique des bâtiments. Available at: https://www.simeb.ca/.

Klein, S. A. et al. (2017) TRNSYS 18: A Transient System Simulation Program, Solar Energy Laboratory, University of Wisconsin, Madison, USA. Available at: http://sel.me.wisc.edu/trnsys.

Loga, T., Stein, B. and Diefenbach, N. (2016) 'TABULA building typologies in 20 European countries—Making energy-related features of residential building stocks comparable', Energy and Buildings, 132, pp. 4–12.

McKenna, E. and Thomson, M. (2016) 'High-resolution stochastic integrated thermal–electrical domestic demand model', Applied Energy, 165, pp. 445–461.

Natural Resources Canada (2011) Survey of Household Energy Use.

Natural Resources Canada (2014) Canada's Secondary Energy Use by Sector, End Use and Subsector, Energy Use Data Handbook Tables.

NRCan (2017) National Energy Use Database, Natural Resources Canada.

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

Sokol, J. et al. (2016) 'Validation of a Bayesian-Based Method for Defining Residential Archetypes in Urban Building Energy Models', Energy and Buildings. Elsevier B.V., 134, pp. 11–24.

StatCan (2011) Household size, by province and territory, Census of Population and Statistics Canada catalogue no. 98-313-XCB.

StatCan (2016) 2016 Census, Statistics Canada. Available at: http://www12.statcan.gc.ca/census-recensement/2016/dp-pd/index-eng.cfm.

Swan, L. G. and 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), pp. 1819–1835.