energies

MDPI

Article

Statistical Analysis of Baseline Load Models for Residential Buildings in the Context of Winter Demand Response

Alain Poulin *, Marie-Andrée Leduc and Michaël Fournier

Institut de Recherche d'Hydro-Québec (IREQ), 600 av de la Montagne, Shawinigan, QC G9N 7N5, Canada; leduc.marie-andree@hydroquebec.com (M.-A.L.); fournier.michael@hydroquebec.com (M.F.)
* Correspondence: poulin.alain@hydroquebec.com

Abstract: By reducing electricity consumption during peak times, peak shaving could reduce the need for carbon intensive resources and defer capacity related investments. Households, where they use electricity for space or water heating, are major contributors to the winter peak demand and promising candidates for related demand response (DR) initiatives. The impact of such initiatives is determined by comparing the actual consumption during a DR event to a baseline, i.e., the estimated consumption that would have occurred in the absence of an event. This paper explores the challenges associated with modeling a baseline in the context of residential winter DR programs with individual performance-based incentives. A sample of more than a thousand residential load profiles was used in this study to provide a statistical comparison of performance metrics for different baseline load models. Arithmetic, regression based, and matching-day models were considered. Results show that adjusted arithmetic models achieve similar performances to the more complex regression model without the need for weather data. These simpler models were also found to be less sensitive to the number of events called during the season. Performing individual adjustments for each of the two daily peak periods also provides better accuracy.

Keywords: demand response; measurement and verification; baseline model; cold climate

check for updates

Citation: Poulin, A.; Leduc, M.-A.; Fournier, M. Statistical Analysis of Baseline Load Models for Residential Buildings in the Context of Winter Demand Response. *Energies* **2022**, *15*, 4441. https://doi.org/10.3390/en15124441

Academic Editors: Raman B. Paranjape and Zhanle Wang

Received: 14 March 2022
Accepted: 13 June 2022
Published: 18 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In the province of Quebec, eastern Canada, cold climate (ASHRAE climate zone 6A-7 [1]), low residential electricity tariffs [2], and intensive electrification programs in the 1960s and 1970s [3] have led to a widespread use of electricity for domestic water heating and for space heating. A significant portion of the residential building stock is thus all-electric as it uses solely electricity as an energy source. It is estimated that 68% of households have electric heating equipment in their home, e.g., electric baseboards and heat pumps [4]. This reality makes electricity consumption strongly influenced by the outside air temperature in winter and thus induces stress on the electric grid during extreme cold conditions. Households, through their routinized practices, contribute strongly to the two winter daily peaks (typically in the morning from 6 to 9 A.M. and in the evening from 4 to 8 P.M.) experienced by the grid [5]. Demand response (DR) resulting in peak shaving could contribute to relieving this stress and hence defer capacity-related investments. Hydro-Quebec, the largest utility in Quebec, forecasts the contribution of such DR resources at 992 MW in 2028–2029 for dynamic pricing and aggregated new DR resources for residential and SMEs (small and medium enterprises) [6] including smart communicating thermostats' control to shift and shed heating loads [7].

Users become DR resources if they change their electricity consumption behavior during peak periods. The Federal Energy Regulatory Commission (FERC) gives a comprehensive definition of DR which consists of "changes in electric usage by end-use customers from their normal consumption patterns in response to changes in the price of electricity over time, or to incentive payments designed to induce lower electricity use at times of

high wholesale market prices or when system reliability is jeopardized" [8]. Incentive payments are either based on participation (fixed incentive) or on performance (variable incentive). Most utility programs targeting residential customers compensate on a participation basis (e.g., fixed monthly credit as NV Energy's Cool Share program [9]), or installation rebate and yearly credit [10]. This approach is based on a global program evaluation, using experimental design with treatment and control groups, to calibrate the appropriate uniform compensation to be given to all participants to the program. As for the performance-based incentive approach, it uses a baseline, sometimes referred to as customer baseline (CBL), in order to evaluate the shedding performance during a DR event (real load minus the CBL). This corresponds to the load that would have been if no DR event had been effective at a given time. Whereas there exists measurement and verification (M&V) standards for energy efficiency evaluations ([11,12]), it is not the case for DR impacts evaluations. The standard [13] from NAESB (North American Energy Standard Board) is intended to facilitate DR in wholesale electricity markets by providing a common framework for transparency, accountability, and consistency but does not, however, make recommendations on M&V for settlement and program evaluation. More recently, the CalTRACK initiative was put in place in an attempt to develop methods "in an open and transparent stakeholder process that uses empirical testing to define replicable methods for calculating normalized metered energy consumption using either monthly or interval data from an existing conditions baseline" [14].

The question of baseline accuracy is central to the feasibility of performance-based incentives. Systematic underestimating of customer performances could lead to participation erosion but systematic overestimating could lead to a non-cost-effective program. Few utilities prefer to compensate residential DR participants based on individual performance. For example, Commonwealth Edison Company (ComEd), which provides electric services across northern Illinois, has a performance-based incentive program which "uses a formula based on your typical usage history to predict how much energy you would normally have used during Peak-Time Savings Hours. After this number is normalized for weather and other variables, ComEd then subtracts the amount of energy that you actually used to determine how much energy you saved" [15].

Table 1 provides a non-exhaustive listing of DR programs using performance incentives to compensate residential participants.

Table 1. Utility DR programs using performance incentives.

Utility	DR Program	Summer/Winter	Baseline Method
Pepco (Maryland and District of Columbia)	Pepco Peak Energy Savings Credit [16]	summer	Arithmetic High 3 of 30
Connexus Energy (Minnesota)	Connexus Energy Peak-Time Rebate [17]	summer	Arithmetic High 3 of 10
BGE (Maryland)	BGE's Energy Savings Day [18]	summer	Matching day with similar weather
Portland General (Oregon)	Peak-Time Rebates [19]	summer/winter	Arithmetic Mean of 10
ComEd (Illinois)	Peak-Time Savings Hours [15]	winter	Regression using weather

There is a perception that providing performance-based incentives is more challenging given the fact that residential load profiles tend to be highly variable [20], hence making difficult the task of accurately evaluating the individual compensations. In an extensive study of M&V for demand response, Goldberg and Agnew [21] found that weather sensitivity and variability are key determinants of baseline accuracy. They argue that since air conditioning is a major contributor to the summer household load profile and that this load is highly variable, baseline would then be less accurate for residential load profiles than, for example, large industrial load profiles.

Extensive studies, both by academics and practicians, have been conducted to evaluate the performance of the different baseline load models based on historical data for commercial and industrial customers enrolled in DR programs. Granderson and Price [22] assessed the accuracy of five baseline models applied to 29 commercial buildings. Goldberg and Agnew [23] used 646 commercial accounts to assess the performances of 18 baseline models used in the industry. Bode et al. [24] tested both accuracy and settlement errors for 48 baseline models on data from 95 small to large commercial customer accounts for the Ontario Power Authority (OPA). A study for PJM used 4565 accounts to estimate summer baseline accuracy for 11 models [25]. A more recent study by Granderson et al. [26] assesses the accuracy of advanced M&V regression approaches compared to simple baseline on 453 commercial meters. They found that these hourly regressions did not provide significant accuracy improvements over simpler averaging methods. Furthermore, the industry published several review papers on the subject including Enernoc's white paper [27] summarizing differences and challenges characterizing different baseline models such as the issues with adjustments (scalar versus additive, window timing and duration, and load capping) for arithmetic models (ex. High X of Y).

Few studies focus on the residential baseline models but Goldberg and Agnew [21] and George et al. [28] provide some insight into their performance. Goldberg and Agnew [21] made a broad review of measurement and verification issues that are settlement and impact estimation in the context of DR programs. They identified relevant studies and pilots. They state some of the challenges associated with baseline load models for residential load profiles which include weather sensitivity typical of these loads, variability, and the importance of availability of comparable days. George et al. [28] studied a sample of 2000 residential load profiles, testing for 21 baseline models' accuracy and incentive payments impacts. Their key findings on baseline accuracy include poor individual load profile accuracy (20% of sample has $\pm$24% error) and regression models and weather-matching models, although providing better individual accuracy only marginally increased overall accuracy.

Most studies involve DR programs in summer peaking utilities and few have evaluated the performances of baseline load models for event days outside the summer season. Goldberg and Agnew [23] discussed non-summer conditions and KEMA [25] provided additional insight into the winter morning events for non-residential load profiles.

The aim of this study is to characterize the performance of common baseline load models when used for individual load profiles, in the context of winter DR programs for residential customers with two daily peaks and performance-based incentives. We were interested in methods that are either in-use or considered for real life operationalization of DR programs, thus excluding machine-learning-based methods for the moment. This quantitative analysis is based on the statistical evaluation of performance metrics for the studied data sample and should provide answers to the following research tasks:

- Task 1: Are regression-type baseline load models significantly more accurate than other models?
- Task 2: For adjusted arithmetic baseline load models, do individual adjustments for each peak period provide additional overall precision to the baseline?
- Task 3: How do the performances of the baseline load models change with the number of events called during the DR season?

2. Materials and Methods

In this study, common baseline load models were used to predict load profiles for a sample of households. Load profile predictions were made for ten proxy event days (PEDs) with two events per day (morning and evening) and thus 20 events per winter. That compares to Hydro-Québec Winter Credit Option which has between 25 and 33 events for a maximum of 100 h [29]. Performance metrics were then calculated using the predicted and measured load profiles for all PEDs. A statistical analysis was then performed on

these metrics; statistical distributions of performance metrics were obtained and analyzed. Details of this methodology will be discussed in the following paragraphs.

2.1. Data Sample and Processing

As mentioned earlier, electricity is widely used in Québec for both space and domestic water heating purposes in households. Within the residential market, customers with all-electric detached single-family housing are the preferred target for a DR program. They have greater peak reduction potential than customers living in apartments, or non-electric-heated houses.

In Québec, a typical all-electric detached single-family household (without air conditioning, pool or spa) consumes annually about 22 000 kWh, 55–60% of which is consumed for space heating. The typical building overall heat transfer coefficient (*UA*) of this market segment is evaluated around 120–180 W/°C using a linear change-point model [30,31]. The overall heat transfer coefficient characterizes the effect of the outside air temperature (*OAT*) on the electricity consumption. Average demand reductions during peak periods for this market segment using typical space heating DR strategies are expected to range from 1 to 2 kW [7], which corresponds to about 15–20% of the normal demand without DR.

The sample used in this study consists of nearly two thousand residential households located in the greater Montreal area. One year of 15 min interval meter data was available (1 November 2012 to 31 December 2013) for each household. No DR events were called on these households during the studied period. Since most baseline load models use hourly load data, interval meter data were converted to hourly time series by integrating all interval data for each hour.

As no information was available on the type of building (apartments, single detached, etc.) and whether these households use electricity as their sole source of energy, the following inclusion criteria were used to select the households most likely belonging to the targeted market of all-electric detached single-family household:

- Annual electricity consumption (>5000 kWh);
- Weather sensitivity (*UA* > 75 W/°C);
- Amount of missing data (<20%).

After applying the inclusion criteria, the sample contained the data of 1178 households.

Hourly *OAT* data for the nearest weather station for each household were obtained from an online meteorological dataset [32].

2.2. Sample Characterization

Retained households' load profiles were characterized by their annual electricity consumption (kWh), variability (*COV*, average value of the coefficient of variability of the electric consumption during both peak periods of the ten coldest winter days) and weather sensitivity (*UA*). Table 2 and Figure 1 show the characteristics of the sample.

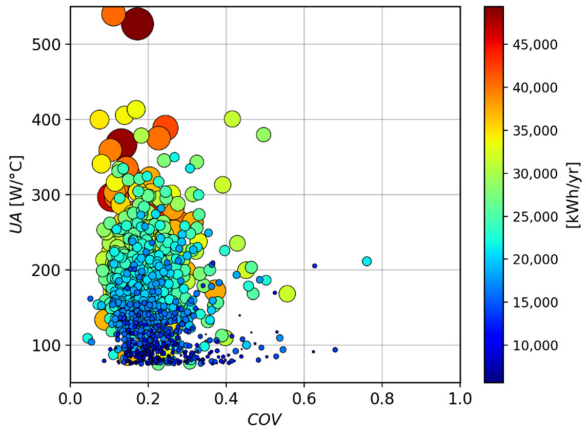


Figure 1. Sample characteristics (marker sizes are proportional to the annual electricity consumption).

Table 2. Statistics of the sample characteristics.

Characteristics	Value
Size	1178
Sampling dates	1 November 2012 to 31 December 2013
Annual consumption [kWh]	
Mean	21,911
Median	21,818
Std	8109
COV	
Mean	0.21
Median	0.19
Std	0.10
UA [W/°C]	
Mean	165
Median	158
Std	64

2.3. Proxy Event Days (PED)

Baseline load models are used to generate an hourly profile for several days, referred to as PED. PEDs were chosen as the ten coldest admissible days (normal working days, excluding weekends and holidays, e.g., Christmas break) in the historical data period. Montreal *OAT* was used for PED selection. On those days, the average daily *OAT* varied between −13 °C and −22 °C. Two events were considered during each PED. Event's time corresponded to the peaking periods of Hydro-Québec's network, i.e., in the morning between 6 and 9 A.M. and in the evening between 4 and 8 P.M.

Two scenarios were studied using the PEDs. The first, referred to as *S1*, evaluates baseline load models for every PED by including other PEDs as admissible days (like they were regular days) while the second, *S2*, excludes other PEDs from the admissible days (like they were event days) to calculate the baseline. *S2* represents a more realistic scenario where the DR program is called repeatedly leaving no very cold days as admissible days (also referred to as comparable days). Unavailability of comparable days (especially for weather-sensitive load profiles) [18] typically has a detrimental effect on the performance of baseline load models, as does having too few recent non-event days (this effect is referred to as a static baseline).

2.4. Baseline Load Models

This study uses Type-1 baseline models as defined by NAESB, i.e., "baseline performance evaluation methodology baseline model based on historical meter data for a Demand Resource that may also include other parameters such as weather and calendar data" [13]. Several Type-1 baseline models have been proposed and evaluated in the literature and used in practice to predict non-residential baseline load for DR events ([25,26,33]). Among these models, thirteen were included in this study. A baseline load model developed by the authors in a previous study provided good results when applied to commercial load profiles and therefore was included in this study. It is referred to as *SGE* [34].

The tested models can be grouped into three calculation types:

1. Arithmetic (11 models); models using the hourly loads (L_h) of a certain number of admissible days (X) within a given baseline window (Y) to predict the baseline load ($\hat{L}_h^{PED}$); Tested models include:
    - *High X of Y*: with Y = 5 or 10 and X = 3, 4, 5 or 7, 8, 9, 10. These models consider only the X maximal L_h of the last Y admissible days [L_h^Y];

$$\hat{L}_h^{PED} = \frac{1}{X} \sum_{j \in High(X,Y)} L_h^j \text{ with } h \in [1,\ 24] \text{ and } \begin{cases} X = 3,\ 4,\ 5 \ \ if\ Y = 5 \\ \text{or} \\ X = 7,\ 8,\ 9,\ 10 \ \ if\ \ Y = 10 \end{cases}, \quad (1)$$

- *Mid X of Y*: with Y = 5 or 10 and X = 3, 6, or 8. These models discard the maximal and minimal L_h to keep only the X center ones;

$$\hat{L}_h^{PED} = \frac{1}{X} \sum_{j \in Mid(X,Y)} L_h^j \text{ with } h \in [1,\ 24] \text{ and } \begin{cases} X = 3 \ \ if\ Y = 5 \\ \text{or} \\ X = 6,\ 8 \ \ if\ \ Y = 10 \end{cases}, \quad (2)$$

- *Monthly High 5*: just like a High X of Y model with X = 5 and Y = duration of the billing period, in days (approximately a calendar month).

2. Regression (2 models): Models using a correlation between the load and other variables (such as weather data, occupancy, weekday, production, etc.) to predict the load during a DR event. *OAT* is the only variable considered in this study as it is recognized to have the strongest effect on residential load profiles during winter in cold climate. The regression model is based on this equation:

$$\hat{L}_h^{PED} = aOAT_h^{PED} + b, \quad (3)$$

Tested models include:

- *Seasonal weather*: Historical data are used to calculate coefficients of a linear regression of hourly load to hourly *OAT*. The coefficients are calculated using data from all the heating season admissible days (1 December to 31 March).
- *10-day weather*: Model using a similar linear regression as the seasonal weather model but the coefficients are calculated using data from only the ten previous admissible days.

3. Matching-day: *SGE* model matches the conditions of the PED to the historic admissible day closest in terms of weather and operating conditions. Days with similar operating conditions are found using fuzzy c-mean clustering (unsupervised machine learning) of daily load consumption profile. Clustering of daily load profile is used to find days with similar operating conditions [35].

2.5. Baseline Adjustments

Baseline load profiles obtained from *Arithmetic* models can be adjusted to align the baseline with the observed conditions prevailing during DR event days (*Adjusted arithmetic*). The additive adjustment factor studied here is based on pre-event measured load. The adjustment factor is not capped either positively or negatively.

As mentioned by Coughlin [36], deciding on the time adjustment window to use for the adjustment factor calculation can be a challenge in the context where DR events are called hours prior to the peak periods. Intentional modulation of the measured load (e.g., gaming) can be used by DR resources as a means to increase compensation. Unintentional modulation can also take place such as when preconditioning of the building is carried out during the adjustment window. Preheating has indeed been proven as an effective strategy to mitigate residential occupants' discomfort in a winter DR context [37].

The adjustment window used in this study corresponds to the third and fourth hours prior to the peak periods, i.e., between 2 and 4 A.M. for morning events and between 12 and 2 P.M. for evening events. This leaves a two-hour window for preconditioning of the building before the peak period, a realistic assumption. For the evening events, an adjustment based on the morning adjustment window was also considered. Results for the two adjustment windows for the evening events will be referred to as early morning (*EM*) and midday (*MD*) window.

Figure 2 presents examples of CBL profiles predicted using the different models for 2 households on a cold day. Figure 2a shows the unadjusted profiles while Figure 2b shows the adjusted ones (dotted lines) using the *EM* window. Results demonstrate the positive

effect of the adjustment on the predicted profiles: they are much closer to the measured one over the event hours (7 to 9).

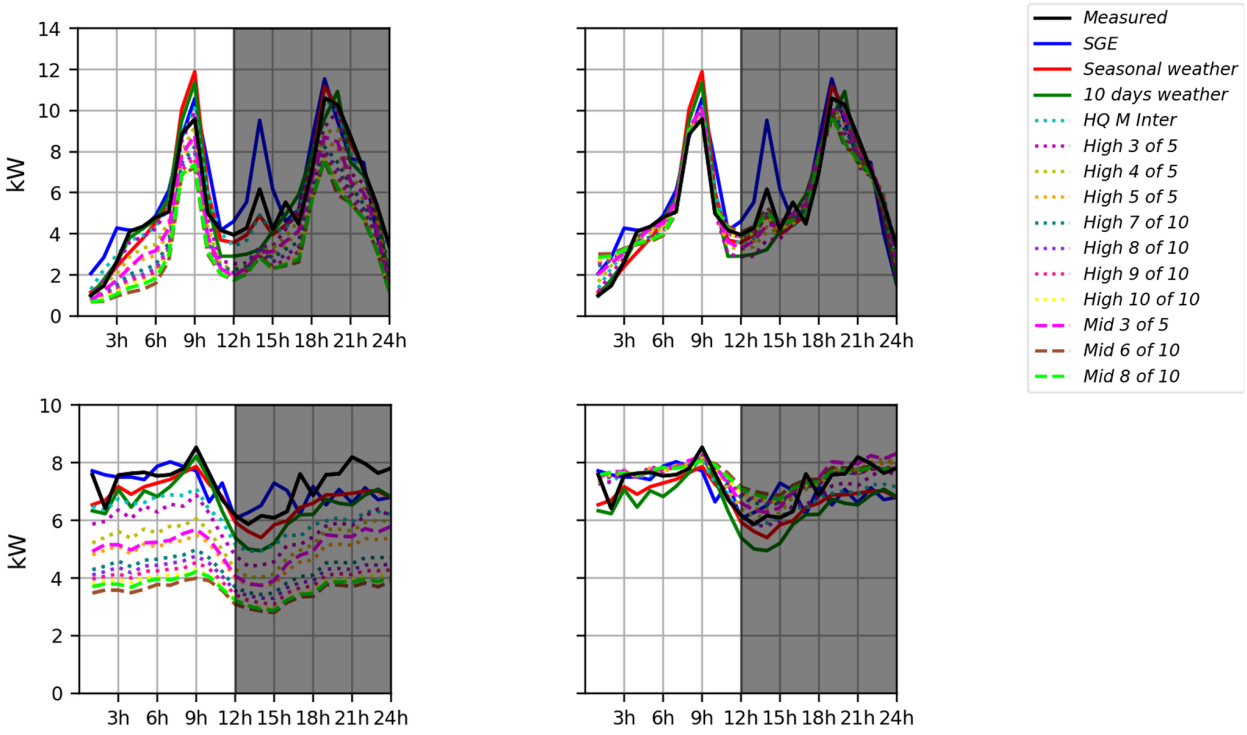


Figure 2. Examples of the resulting baseline profiles for 2 households, (**a**) without and (**b**) with adjustment.

2.6. Performance Metrics

Baselining and forecasting literature were reviewed to identify bias and accuracy metrics commonly used in those fields ([22,38]). Several metrics were tested but for the purpose of simplicity, only two are presented in this paper, i.e., mean bias error (*MBE*) and mean absolute percent error (*MAPE*). Indeed, it was found that models' ranking, as presented in the results section, does not significantly change when using different metrics for the considered sample.

MBE and *MAPE* are calculated using the measured and predicted load during the morning and evening peak periods for every PED. Distinct performance metrics distributions are obtained for each combination of baseline model, peak periods, and scenario.

2.6.1. Bias Metric: *MBE*

Bias indicates the tendency of a baseline load model to systematically under- or overestimate the measured load. *MBE* has been used in [11,22] and is given by:

$$MBE = \frac{\sum_{h=1}^{n}(L_h - \hat{L}_h)}{\sum_{h=1}^{n}(L_h)}, \quad (4)$$

where h = 7 to 9 and h = 17 to 20 for the morning and evening peak periods, respectively.

MBE varies between ±1 and indicates how much the predicted load ($\hat{L}_h$) deviates from the measured load (L_h). Underestimation is given by positive value while negative value indicates overestimation.

2.6.2. Accuracy Metric: *MAPE*

Accuracy is defined as the overall difference between predicted and measured values. Accuracy thus combines both bias and precision which is a measure of the variance.

MAPE is widely used in the literature as a measure of accuracy in the context of forecasting [39–41] as well as baselining [42]. *MAPE* is given by:

$$MAPE = \frac{\sum_{h=1}^{n}\left(\left|\left(L_h - \hat{L}_h\right)/L_h\right|\right)}{N}, \quad (5)$$

MAPE is a scale-independent metric which means that it can be used to compare predictions for a very heterogeneous population. It can be prone to unrealistic results under certain conditions such as when measured values are close to or equal to zero [43]. This is not the case in the present study as the original data were cleaned from outside range data and the considered peak periods typically have high load.

3. Results

In this section, we will provide results to help answer the four research questions initially formulated. Distributions of *MBE* and *MAPE* obtained from each combination of load profiles and PED (i.e., 11,780 points) for scenarios *S1* and *S2* were used to analyze the results. Figure 3 gives an example of the distributions obtained. Distributions are characterized by their median and range, defined here as the difference between the 95th and the 5th percentile. Baseline load models with the best performance should have median *MBE* close to zero, low median *MAPE*, and narrow range value for both *MBE* and *MAPE*.

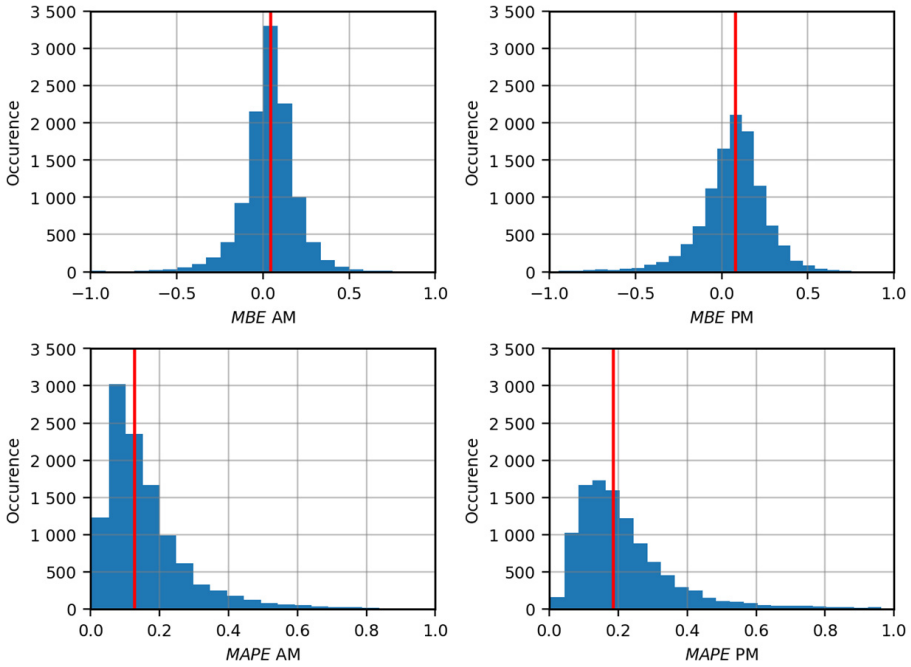
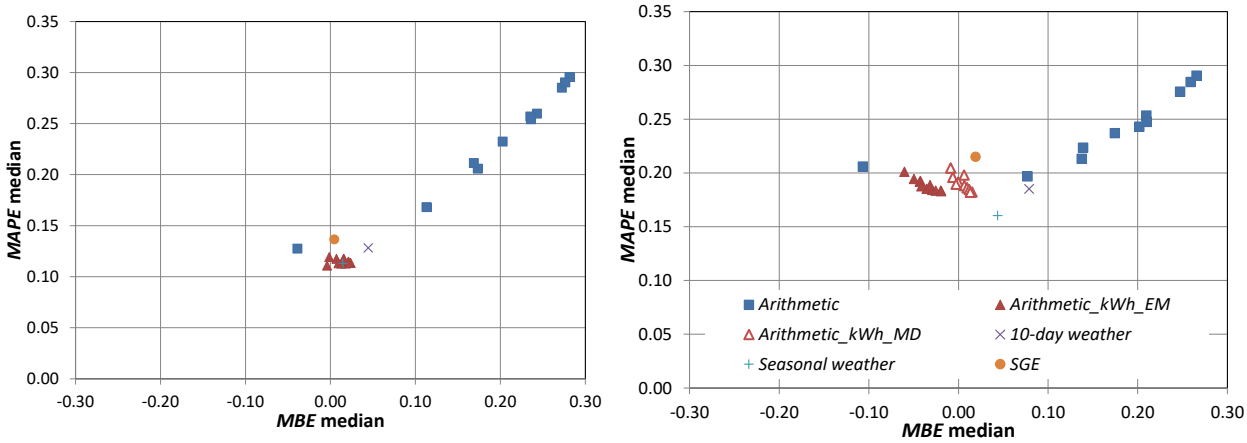


Figure 3. *MBE* and *MAPE* distributions for the *10-day weather* model, scenario *S1*.

3.1. Models Ranking

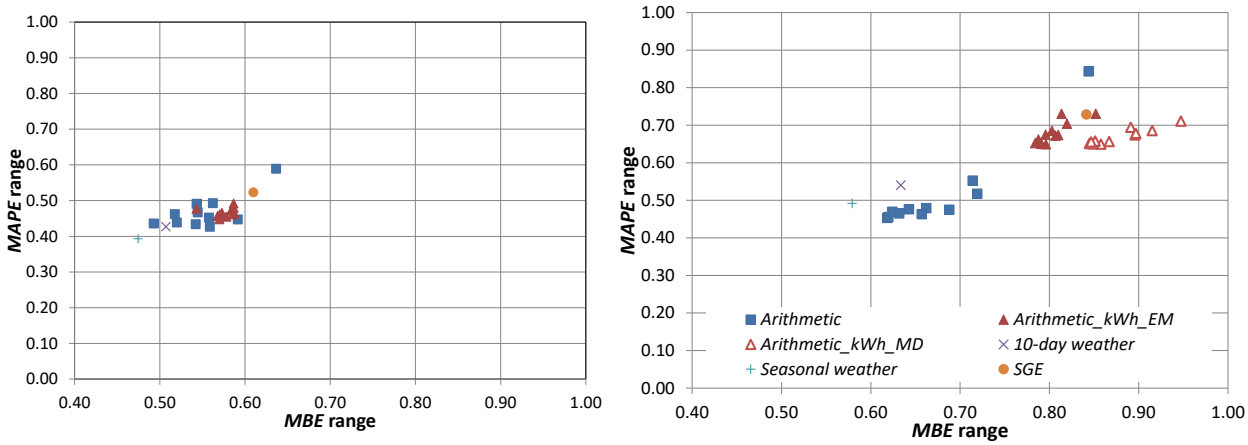
Performance analysis of all models will be discussed according to their relative ranking with respect to the metrics' distribution characteristics. The results for morning and evening events will be presented separately. Models' ranking will be first presented for scenario S1 only (independent events). The *Adjusted arithmetic* models presented in the figures are labelled as *Arithmetic_kWh_YY* where *YY* refers to the adjustment window (*EM*: early morning and *MD*: midday).

Figure 4 shows all studied models' median results for *MAPE* plotted against *MBE* for scenario *S1*. Figure 5 shows the range of distribution of both metrics. For clarity purposes, models were grouped by calculation type (marker shape), with each point representing a different model.



(**a**) (**b**)

Figure 4. MAPE versus MBE for median results for scenario *S1* (**a**) morning and (**b**) evening peak period.



(**a**) (**b**)

Figure 5. *MAPE* versus *MBE* for range results for scenario *S1* (**a**) morning and (**b**) evening peak period.

According to Figures 4 and 5, the following observations can be made:

- *Arithmetic* models have the largest median *MBE* combined with high *MAPE* values for both peak periods. Most of these models underestimate the load (median *MBE* > 0), except for *Monthly High 5* which overestimates. *Arithmetic* models lead to some of the smallest range value for both metrics and peak periods;
- Compared to the *Arithmetic* models, *Adjusted arithmetic* models present a significant reduction in the distributions' medians. Range value results show that baseline adjustments tend to widen the distributions, especially for the evening events. Performance's metrics across *Adjusted arithmetic* models are similar, meaning none of the models really stand out;
- The *SGE* model presents slightly worse results than the best ones. The *SGE* model does not perform well, likely because it is based on load pattern recognition which is not compatible with the high-variability characteristics of residential load profiles;
- Apart from the *Arithmetic* models, the *10-day weather* model has the biggest bias as shown by its *MBE* median values for both peak periods. However, its low range values indicate that it produced consistent results (narrow distributions);
- The *Seasonal weather* model is among the best performing models in terms of median and range value for both metrics and peak periods;
- For the morning period (Figures 4a and 5a), the best *Adjusted arithmetic* models compare favorably with the *Seasonal weather* model especially in terms of bias. For the

evening period, *Seasonal weather* has a slight advantage compared to the best *Adjusted arithmetic* models except for the MBE median, meaning that it is more biased.

3.2. Effect of the Adjustment Window

For the evening events, two adjustment windows were considered for the *Adjusted arithmetic* models, one at the beginning of the day between 2 and 4 A.M. and one closer to the beginning of the event, i.e., between 12 and 2 P.M.

The results in Figures 4b and 5b present the performance metrics distribution characteristics for all the *Adjusted arithmetic* models for scenario *S1* for the evening peak periods. The results tend to show that the adjustment window selection has more impact on the *MBE* distributions than the *MAPE*. Using an adjustment window closer to the event period reduces the bias and *MBE* median, but increases the *MBE* range.

3.3. Effect of the Number of Events on the Performance Metrics

So far, the results presented assumed that each PED is considered the only event day occurring during the winter season, with all other PEDs being admissible days for CBL evaluation (scenario *S1*). However, considering every PED as an event day and removing them from the admissible days for CBL evaluation (scenario *S2*) brings some insights on how the repeated use of DR affects the performance of CBL evaluation. Figure 6 compares results for scenarios *S1* and *S2*. In this figure, the shadow areas indicate a decrease in performance due to a high number of PEDs. Results lying on the diagonal of these graphs indicate no effect of the number of PEDs.

Results indicate that globally, *Seasonal weather* and *Adjusted arithmetic* using the adjustment window closest to the DR event are the least sensitive models to the number of events during the winter season. The performance of the other models deteriorates when the number of PEDs increases.

Arithmetic, *10-day weather*, and *SGE* models all rely on the presence of cold-weather admissible days to establish the model and thus provide good baseline prediction otherwise they can fall in the extrapolation range. *Seasonal weather* is less sensitive to the absence of cold-weather days among the admissible days because it uses a larger number of days to establish the linear regression.

Regarding *Adjusted arithmetic* (*Arithmetic_kWh*) models, it seems that as long as the shape of the average load profile is good, the measured load-based adjustment will lead to similar results independently of the number of PEDs.

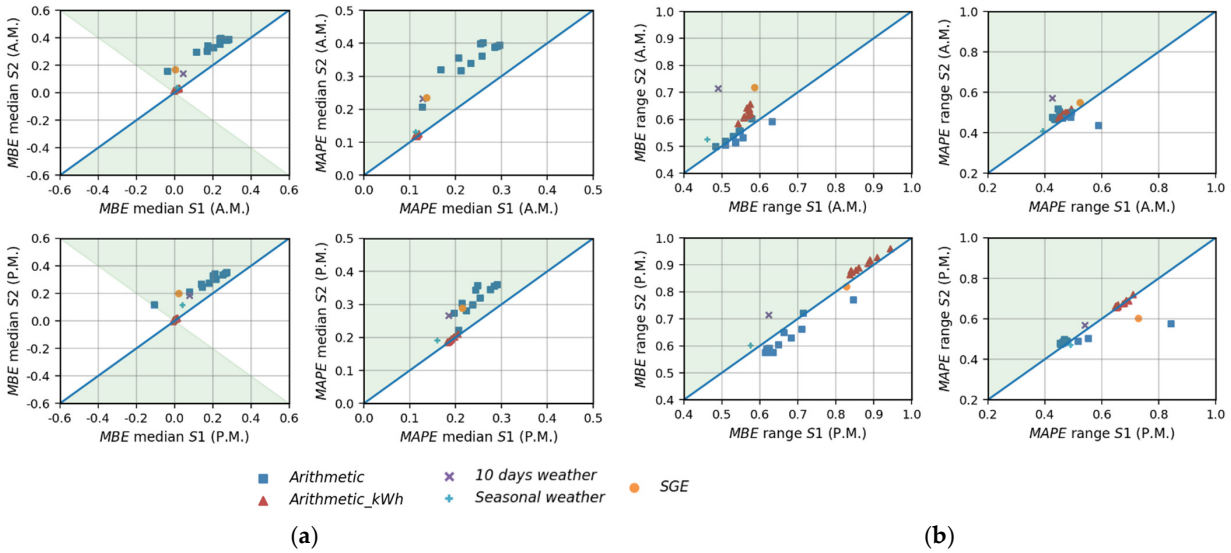


Figure 6. Comparison of (**a**) the median and (**b**) range of the *MBE* and *MAPE* results for scenarios *S1* and *S2* (**top** = A.M. peak period, **bottom** = P.M. peak period).

4. Discussion

The results in Section 3 provide answers to the tasks formulated in the introduction of this paper.

Task 1: Are regression-based baseline load models significantly more accurate than other models?

In terms of *MBE* and *MAPE*, the results have shown that the *Seasonal weather* model presents the best overall performance. This model was also proven to be fairly independent of the number of PEDs called during the winter season.

There is a clear advantage to the use of regression-based models such as *Seasonal weather* compared to simple *Arithmetic* models. However, this advantage is reduced when adjustments are applied to arithmetic models to produce *Adjusted arithmetic* models. It was shown that *Adjusted arithmetic* models based on measured load can achieve performances similar to the *Seasonal weather* model without the need for *OAT* data.

When comparing *Seasonal weather* to *Adjusted arithmetic* models, other considerations than performance could favor the use of the latter in a DR program. Compared to the *Seasonal weather* model for which the results are only available at the end of the winter season, *Adjusted arithmetic* models have the advantage of being able to give near-real-time feedback to DR participants.

Task 2: For adjusted arithmetic models, do individual adjustments for each peak period provide additional precision to the baseline?

The results have shown that the adjustment window should be taken as close as possible to the DR event while keeping in mind the gaming possibilities by DR resources. The adjustment window therefore has to be wisely chosen from the standpoint of the DR program coordinator.

In order to avoid the gaming problem, it has been suggested that the adjustment factor could be calculated using the *OAT* instead of the measured load. Some calculations, not presented in this paper, have been performed using an adjustment based on the *OAT*. Compared to load-based adjustment, *OAT*-based adjustment has shown similar results. However, one must bear in mind that baseline load models using data other than interval meter data or post-event data are considered more complex to implement.

Using *Adjusted arithmetic* models, one must also consider the risks of intentional modulation of the measured load by participants, as mentioned in Section 2.5.

Task 3: How does the repeated use of DR resources during the winter season affect the performances of the baseline load models?

The number of events during the season can have a slight to moderate effect on the performance of most models, as shown under the conditions prevailing in this study, i.e., ten days with two daily events. Globally, *Seasonal weather* and *Adjusted Arithmetic* using the adjustment window closest to the DR event are the least sensitive models to the number of events during the winter season. *Arithmetic*, *10-day weather* and *SGE* models are all largely dependent on the characteristics of a smaller number of admissible days matching those of the PED otherwise they fall in the extrapolation range.

To increase the performance of the baseline, results suggest avoiding calling all DR participants on every DR event. Keeping cold-weather days free of DR events would extend the interpolation range of most baseline load models. This, of course, must be weighed against the cost of such measures such as reduced DR resource volume for the DR program coordinator.

5. Conclusions

This paper has provided insights into the performance of common baseline load models when used for individual residential winter load profile prediction, in the context of a performance-based DR program. It was found that *Adjusted arithmetic* models based on measured load achieve performances similar to the *Seasonal weather* model without the need to use weather data. *Adjusted arithmetic* models have also shown to not be very

sensitive to the number of events called during the season. Results have also shown that the adjustment window for these methods should ideally be taken as close as possible to the DR event.

Other considerations from overall baseline model performance should also be considered when choosing a model for a DR program. Simplicity and ease of implementation are two key factors to consider when making such a choice for a DR program coordinator.

The ease of implementation must also be considered when choosing a baseline load model since a more complex model might result in additional operating costs both during implementation and exploitation. Models involving regression calculations based on *OAT* (*Seasonal weather*, *10-day weather*) involve providing for mechanisms to allocate and extract weather data and sharing these to the participants and are considered more complex to implement, as is the *Matching-day* (*SGE*) method which involves additional algorithm development and implementation.

The simplicity of the baseline load model, or the capability of all involved in the DR program, for example customers or aggregators, to understand and calculate the baseline model is another aspect to consider. In the present case, the *Seasonal weather* method's relative advantage in terms of performance has to be weighed against its lack of simplicity and implications in terms of implementation.

Finally, the purpose of the DR impact evaluation also influences the choice of a baseline load model. Near-real-time individual compensation evaluation has different constraints compared to an overall program evaluation.

Author Contributions: Conceptualization, A.P. and M.-A.L.; methodology, A.P. and M.-A.L.; software, A.P.; validation, A.P., M.-A.L. and M.F.; formal analysis, A.P. and M.-A.L.; investigation, A.P., M.-A.L. and M.F.; resources, A.P.; data curation, A.P.; writing—original draft preparation, A.P. and M.-A.L.; writing—review and editing, A.P., M.-A.L. and M.F.; visualization, A.P.; supervision, M.F.; project administration, M.F.; funding acquisition, M.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Hydro-Québec.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study cannot be shared due to ethical, legal and privacy issues.

Conflicts of Interest: Authors are research scientists at Hydro-Quebec, therefore, the funding sponsor had a role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. *ANSI/ASHRAE/IES Standard 90.1-2016*; American Society of Heating, Refrigerating and Air-Conditioning Engineers. Energy Standard for Buildings Except Low-Rise Residential Buildings. ASHRAE: Atlanta, GA, USA, 2016.
2. Hydro-Québec. Comparaison des prix de L'électricité dans les Grandes Villes Nord-Américaines-Tarifs en Vigueur le 1er Avril 2021. 2021. Available online: https://www.hydroquebec.com/data/documents-donnees/pdf/comparaison-prix-electricite.pdf (accessed on 7 December 2021).
3. Sabourin, M. Les programmes de vente devraient produire leurs effets en 1970. In *Entre-nous*; Hydro-Québec Archives: Quebec, QC, Canada, 1970; p. 8.
4. Natural Resource Canada. Comprehensive Energy Use Database-Residential Sector-Quebec, Table 21. 2021. Available online: https://oee.nrcan.gc.ca/corporate/statistics/neud/dpa/showTable.cfm?type=CP&sector=res&juris=qc&rn=21&page=0 (accessed on 7 December 2021).
5. Hydro-Québec. Using Energy More Wisely in Cold Weather. 2021. Available online: https://www.hydroquebec.com/residential/customer-space/electricity-use/winter-electricity-consumption.html (accessed on 7 December 2021).
6. Hydro-Québec. État D'Avancement 2021 du Plan D'Approvisionnement 2020–2029. 2021. Available online: http://www.regie-energie.qc.ca/audiences/Suivis/Suivi%20HQD_PlanAppro2020-2029/%C3%89tat%20d\T1\textquoterightavancement%202021.pdf (accessed on 7 December 2021).

7. Fournier, M.; Leduc, M.; Sansregret, S.; Poulin, A. Making the Connection: Testing Line-Voltage Communicating Thermostats for Baseboard Heaters in DR and EE Experiments. In Proceedings of the 2018 ACEEE Summer Study on Energy Efficiency in Buildings, Pacific Grove, CA, USA, 12–17 August 2018.
8. Federal Energy Regulatory Commission FERC. Reports on Demand Response and Advanced Metering. 2021. Available online: https://www.ferc.gov/industries-data/electric/power-sales-and-markets/demand-response/reports-demand-response-and (accessed on 7 December 2021).
9. NV Energy. Cool Share Support. 2021. Available online: https://www.nvenergy.com/help/coolshare-support (accessed on 7 December 2021).
10. PSEG Long Island. Smart Savers Thermostat Program. 2021. Available online: https://www.psegliny.com/saveenergyandmoney/energystarrebates/SmartSavers (accessed on 7 December 2021).
11. American Society of Heating, Refrigerating and Air-Conditioning Engineers. *ASHRAE Guideline 14-2014: Measurement of Energy, Demand and Water Savings*; ASHRAE: Atlanta, GA, USA, 2014.
12. IPMVP Committee. *International Performance Measurement and Verification Protocol: Concepts and Options for Determining Energy and Water Savings*; National Renewable Energy Lab.: Golden, CO, USA, 2001; Volume I, (No. DOE/GO-102001-1187; NREL/TP-810-29564).
13. North American Energy Standards Board. *WEQ-015 Measurement and Verification of Wholesale Electricity Demand Response*, NAESB WEQ Business Practice Standards version 003.1; Standards and Models Relating to Measurement and Verification of Wholesale Electricity Demand Response, NAESB: Houston, TX, USA, 2015.
14. CalTRACK. CalTRACK Technical Documentation: Modeling Hourly Methods. 2018. Available online: http://docs.caltrack.org/en/latest/index.html (accessed on 9 November 2021).
15. Commonwealth Edison Company (ComEd). Peak Time Savings. 2021. Available online: www.comed.com/SmartEnergy/InnovationTechnology/Pages/PeakTimeSavingsFAQ.aspx#:~{}:text=ComEd%20will%20credit%20your%20energy,open%20and%20factories%20are%20running (accessed on 7 December 2021).
16. PEPCO. About Peak Energy Savings Credit. 2021. Available online: www.pepco.com/WaysToSave/ForYourHome/Pages/MD/AboutPeakEnergySavingsCredit.aspx (accessed on 7 December 2021).
17. Connexus Energy. Peak-Time Rebate. 2021. Available online: www.connexusenergy.com/save-money-and-energy/programs-rebates/peak-time-rebate (accessed on 7 December 2021).
18. BGE. Energy Savings Days. 2021. Available online: www.bge.com/WaysToSave/ForYourHome/Pages/EnergySavingsDays.aspx (accessed on 7 December 2021).
19. Portland General. Peak Time Rebates. 2021. Available online: https://portlandgeneral.com/ptr (accessed on 7 December 2021).
20. Elliot, C.; Shelton, J.; Hanna, D. Challenges of estimating hourly baselines for residential customers. In Proceedings of the 2013 International Energy Program Evaluation Conference, Chicago, IL, USA, 13–15 August 2013.
21. Goldberg, M.L.; Agnew, G.K. *Measurement and Verification for Demand Response*; US Department of Energy: Washington, DC, USA, 2013.
22. Granderson, J.; Price, P.N. Development and application of a statistical methodology to evaluate the predictive accuracy of building energy baseline models. *Energy* **2014**, *66*, 981–990. [CrossRef]
23. Goldberg, M.L.; Agnew, G.K. Protocol Development for Demand Response Calculation-Findings and Recommendations. 2003. Available online: www.calmac.org/publications/2002-08-02_XENERGY_REPORT.pdf (accessed on 17 June 2022).
24. Bode, J.; Perry, M.; Morgan, S.; Freeman, S. *Assessment of Settlement Baseline Methods for Ontario Power Authority's Commercial & Industrial Event Based Demand Response Programs*; Prepared by Freement, Sullivan & Co. for Ontario Power Authority: Toronto, ON, Canada, 2010.
25. KEMA Inc. PJM Empirical Analysis of Demand Response Baseline Methods White Paper. PJM Load Management Task Force. 2011. Available online: www.pjm.com/-/media/markets-ops/demand-response/pjm-analysis-of-dr-baseline-methods-full-report.ashx (accessed on 17 June 2022).
26. Granderson, J.; Sharma, M.; Crowe, E.; Jump, D.; Fernandes, S.; Touzani, S.; Johnson, D. Assessment of Model-Based peak electric consumption prediction for commercial buildings. *Energy Build.* **2021**, *245*, 111031. [CrossRef]
27. EnerNoc. The Demand Response Baseline Whitepaper. 2009. Available online: www.naesb.org/pdf4/dsmee_group3_100809w3.pdf (accessed on 12 October 2021).
28. George, S.; Bode, J.; Berghman, D. 2012 San Diego Gas & Electric Peak Time Rebate Baseline Evaluation. Prepared for San Diego Gas & Electric. 2013. Available online: www.calmac.org/publications/SDGE_PTR_Baseline_Evaluation_Report_-_Final.pdf (accessed on 17 June 2022).
29. Hydro-Quebec. Winter Credit Option. 2021. Available online: https://www.hydroquebec.com/residential/customer-space/rates/winter-credit-option.html (accessed on 28 October 2021).
30. Poulin, A.; Fournier, M.; Dostie, M. Characterization and modelling of electricity consumption. *Eur. J. Electr. Eng.* **2010**, *13*, 717–740. [CrossRef]
31. American Society of Heating. Refrigerating and Air-Conditioning Engineers. In *ASHRAE Handbook: Fundamentals*, S-I ed.; ASHRAE: Atlanta, GA, USA, 2013.
32. Hydro-Quebec. SIMEB Weather Database. 2021. Available online: https://www.simeb.ca/ (accessed on 4 December 2021).

33. Coughlin, K.; Piette, M.A.; Goldman, C.; Kiliccote, S. *Estimating Demand Response Load Impacts: Evaluation of Baseline Load Models for Non-Residential Buildings in California (No. LBNL-63728)*; Ernest Orlando Lawrence Berkeley National Laboratory: Berkeley, CA, USA, 2008.
34. Leduc, M.A.; Lavigne, K.; Daoud, A.; Poulin, A. Demand response in commercial buildings: A cold climate field study. *ASHRAE Trans.* **2015**, *121*, 1U.
35. Richard, M.A.; Fortin, H.; Poulin, A.; Leduc, M.A.; Fournier, M. Daily load profiles clustering: A powerful tool for demand site management in medium-sized industries. In Proceedings of the ACEEE Summer Study on Energy Efficiency in Industry, Denver, CO, USA, 15–18 August 2017.
36. Coughlin, K.; Piette, M.A.; Goldman, C.; Kiliccote, S. Statistical analysis of baseline load models for non-residential buildings. *Energy Build.* **2009**, *41*, 374–381. [CrossRef]
37. Fournier, M.; Leduc, M.A.; Nadrault, G. Winter Demand Response Using Baseboard Heaters: Achieving Substantial Demand Reduction without Sacrificing Comfort. In Proceedings of the ACEEE Summer Study on Energy Efficiency in Buildings, Pacific Grove, CA, USA, 21–16 August 2016.
38. Armstrong, J.S. *Principles of Forecasting: A Handbook for Researchers and Practitioners*; Kluwer Academic Publishers: Norwell, MA, USA, 2001.
39. Wang, P.; Liu, B.; Hong, T. Electric load forecasting with recency effect: A big data approach. *Int. J. Forecast.* **2016**, *32*, 585–597. [CrossRef]
40. Bowerman, B.L.; O'Connell, R.T.; Koehler, A.B. *Forecasting, Time Series and Regression: An Applied Approach*; Thomson Brooks/Cole: Belmont, CA, USA, 2004.
41. Henke, J.E.; Reitsch, A.G. *Business Forecasting*, 5th ed.; Prentice-Hall: Englewood Cliffs, NJ, USA, 1995.
42. Granderson, J.; Price, P.N.; Addy NSohn, M. Automated measurement and verification: Performance of public domain whole-building electric baseline models. *Appl. Energy* **2015**, *144*, 106–113. [CrossRef]
43. Makridakis, S.; Wheelwright, S.; Hyndman, R.J. *Forecasting: Methods and Applications*, 3rd ed.; John Wiley and Sons: New York, NY, USA, 1998.