The 6th IBPSA Canada Conference
La 6e conférence d'IBPSA – Canada
Winnipeg, May 19 – 20

SIMULATION OF CONSUMPTION PROFILES BASED ON MONTHLY BILLING

François Laurencelle, Alain Moreau, Alain Poulin
Laboratoire des Technologies de l'Énergie, Institut de Recherche d'Hydro-Québec,
600, avenue de la Montagne, Shawinigan, Québec, Canada, G9N 7N5.

ABSTRACT

This paper presents an algorithm to construct a building's power consumption profile by 15-minute intervals from its monthly electricity consumption and monthly peak demand. These data can be retrieved through the building's billing history. The algorithm also requires a one-year record of the outdoor temperature as well as the building's operation schedule. The generated profile takes into account the heating and air conditioning loads, the building's operation load and the random power fluctuations. These loads are weighted through a descriptive model established from a collection of building electricity profiles.

The algorithm was validated by comparing generated profiles with metered profiles. Then, we analyzed the bias resulting from using synthetic profiles for sizing thermal storage equipments in different buildings, for a power demand management application.

The simulation algorithm is able to reproduce the main characteristics of the metered profiles. It can address the needs of engineering firms that support energy efficiency efforts or the needs of power utilities wishing to increase their targeted clientele through diagnostic tools, simulations, and marketing of energy efficient products.

1. INTRODUCTION

Hydro-Quebec is now installing communicating meters for its *Commercial and Institutional* (CI) customers. These devices can read the electricity consumption by 15-minute intervals (unlike the old meters providing only monthly summaries). The data recorded through the communicating meters open the way to a detailed analysis of electrical loads in buildings and to a better understanding of the impact of load management technologies, for which the benefits greatly depend on the temporal distribution of the power consumption.

However, as the market penetration of the new meters is still very limited and the metered data are not always available to the engineering firms, simulation is still the most accessible tool for estimating building power profiles.

Several works have been done to disaggregate the hourly load profiles of buildings to a variety of end-uses such as heating, cooling, lighting, water heating, plug loads, etc. and then perform load simulations [1, 2]. Generally, the behaviour of each end-use (load magnitude and hourly schedule) is retrieved from building simulations, survey statistics or energy audits. On the other hand, metered time series can be analysed for feeding a predictive model, without any assumption about the end-uses, for the purpose of load forecasting. Hahn *et al.* [3] present a review of the principal load forecasting methods.

In this work, we have developed an algorithm to simulate a power consumption profile by 15-minute intervals. It requires outdoor temperature time series and data available to all CI customers: i.e. monthly billing data and a simple schedule. For distributing the load with respect to the mean power and the peak demand, we use a *Load Duration Curve* (LDC) model calibrated with a previous statistical study.

We validated the present algorithm by comparing simulated profiles with metered ones having the same characteristics. Subsequently, by importing metered and simulated profiles into a computer program used for the sizing and economic analysis of *Electrical Thermal Storage* (ETS) systems, we verified the consistency of the calculated results based on metered and simulated data.

2. SIMULATION PROCESS

A realistic power consumption profile is generated from the monthly billing data of a CI building and its operation or occupancy schedule. The magnitudes of the loads are weighted through modeling LDCs that give the distribution of the power demand with respect to time. The profile is correlated with the outdoor temperature using a customized version of the *PRInceton Scorekeeping Method* (PRISM).

2.1. User's data

One should give a few input data to the program including the electricity billing data of a complete year consisting in the monthly electricity consumption in kWh and the monthly peak demand P_{max} in kW. The monthly electricity consumption are divided by the month durations, in hours, giving the monthly mean power $\overline{P}$. The monthly *Load Factors* (*LF*) are calculated as follow:

$$LF = \overline{P} / P_{\max} \tag{1}$$

The user should also specify a schedule that reflects the main electricity uses in the building. Usually, it corresponds to the HVAC system or the occupancy schedule. The building's operation profile $\vec{U}_{op}$ is set to "1" (ON) during the high demand hours, and to "0" (OFF) while the building systems are idle or inoperative (like during nights or weekends). The ratio of the time when the systems are active over the whole period, according to that schedule, is τ_{ON}.

The outdoor temperature profile $\vec{T}$, sampled at time intervals of up to 24 hours, should be given for the whole simulation period. Using the weather databases, we can retrieve the monthly mean temperature at the building's location $\overline{T}$ for the same time periods as the billing data.

2.2. *Load Duration Curves* (LDC)

Poulin *et al.* [4] use an equation that fits the distribution of the power levels measured during a given period, which is the definition of a LDC. The dependent variable is the ratio *R* of the instant power over the maximum power measured during the period. It is sorted in descending order, so the independent variable τ, which is defined within the interval [0, 1], is the ratio of time where the instant power is greater than *R*. Figure 1 illustrates a typical LDC. The origin of a LDC is "1" and the curve decreases according to:

$$R(\tau) = \frac{P}{P_{\max}} = 1 - a\tau - b\tau^{c} + \frac{d}{1+e^{f(\tau-g)}} - \frac{d}{1+e^{-fg}} \tag{2}$$

The parameters (*a*, *b*, *c*, *d*, *f*, *g*) can be obtained form a non-linear regression of sorted and normalized power data. The annual consumption profiles of 332 buildings were used to establish the bounds of each of these parameters using sorted annual data series, as given in Table 1.

Fortunately, the integral of equation 2 is an analytic function which is equal to the *LF* when evaluated on the interval [0, 1]. Poulin *et al.* propose a method to obtain a random LDC matching a given *LF*. It consists in picking a synthetic value for each parameter between the bounds found in Table 1, computing parameter *d* at last to match the required *LF* with the integral of the LDC. The curve coefficients are considered valid for $[0.02 \le d \le 0.5]$ and $R(1) \ge 0$.

Table 1. Bounds[+] of *Load Duration Curve* parameters [4]

	Bounds
LF	$0.15 \le LF \le 0.85$
a	$0 \le a \le \min(2.00LF - 0.20\ ; -1.75LF + 1.58)$
b	$-1.82LF + 1.00 \le b \le -1.75LF + 1.58$
c	$0 \le c \le \min(1.40LF\ ; -1.75LF + 1.58)$
d	$0.02 \le d \le 0.50$
f	$25 \le f \le \min(-444.44LF + 400.00\ ; 444.44LF)$
g	$2.25LF - 1.03 \le g \le 3.00LF - 0.20$

[+]Established from annual data series

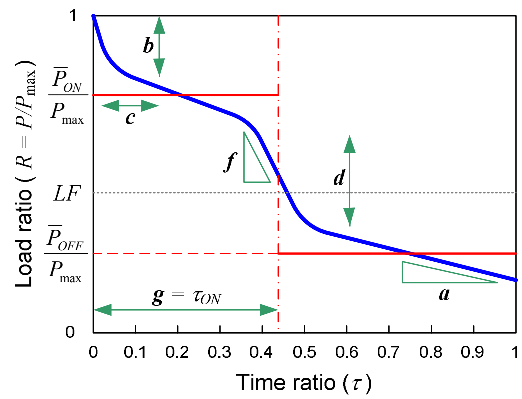

Figure 1. Typical *Load Duration Curve*

The former specification of the building's schedule allows us to set $g \equiv \tau_{ON}$ before retrieving the other LDC parameters. The authors of paper [4] are confident in the accuracy of their method to predict LDC parameters on a monthly basis, instead of the annual period used for the LDC model development. So, from each monthly *LF* value, we compute the average of several randomly generated LDCs to obtain a likely load distribution.

2.3. Load magnitudes

We assume that the building’s operation and idle loads are significantly different. So, the left side of the LDC (Figure 1) will keep together nearly all the scheduled operation load within the interval [0, τ_{ON}]. Oppositely, the right side will collect the idle load data. Therefore, the average power during building’s operation and idle hours should be respectively:

$$\overline{P}_{ON} = \frac{P_{\max}}{\tau_{ON}} \int_{\tau=0}^{\tau_{ON}} R(\tau)\, d\tau \tag{3}$$

$$\overline{P}_{OFF} = \frac{P_{\max}}{1-\tau_{ON}} \int_{\tau=\tau_{ON}}^{1} R(\tau) d\tau \tag{4}$$

The slope of a LDC, related to parameters *a* and *b*, is due to the dispersion of the load distribution. Indeed, the power demand reacts slowly to outdoor temperature variations but the slope also reflects the fast random-like fluctuations that occur due to load switching.

2.4. Climatic effect on power consumption

The PRISM [5] allows estimating the effect of the outdoor temperature on energy consumption. The heating load is expected to be proportional to the difference between a threshold temperature and the outdoor temperature: $T_H - T$, while $T_H > T$. In the same way, the cooling load is proportional to $T - T_C$ while $T > T_C$. The general equation for this model is:

$$P(T) = \underbrace{K}_{\text{Base load}} + \underbrace{H(T_H - T)^{+}}_{\text{Heating load}} + \underbrace{C(T - T_C)^{+}}_{\text{Cooling load}} \tag{5}$$

where the exponents "+" indicate that the term vanishes when the content of the brackets is negative. This simplistic method is used for convenience although it is not much accurate, especially to predict the cooling energy. Indeed, it could be improved by adding parameters to account for the solar heat gain, wind speed or humidity [6, 7], but this would exceed the purpose of this work.

A non-linear least squares regression method is used to fit $\overline{P}$ against $\overline{T}$ through equation (5). To insure relevant fitting results, we run several regressions where the heating and cooling terms are either included or excluded. The fitting solutions are rejected if any of the coefficients (K, H or C) are < 0 or if the heating or the cooling segment is represented by less than three months or if $T_C < T_H$. Among the satisfactory solutions, we adopt the one that maximizes the significance of each coefficient according to the statistical test of *Fisher*.

This fit allows to set the thresholds temperatures T_H and T_C. We assume that these thresholds reflect correctly the behaviours of $\overline{P}_{ON}$ and $\overline{P}_{OFF}$, which are then fitted to equation 5, after imposing T_H and T_C, and thus , we determine K_{ON}, H_{ON}, C_{ON}, K_{OFF}, H_{OFF} and C_{OFF}.

We are now ready to approximate the power consumption profile of the building. Since the building reacts slowly to outdoor temperature changes, we filter the hourly temperature vector $\vec{T}$ with a 24-hour mobile average window. Then we evaluate the PRISM equations for $P_{ON}(\vec{T})$ and $P_{OFF}(\vec{T})$, implemented with their respective sets of coefficients, and for every filtered outdoor temperature. The convolution with $\vec{U}_{op}$ sets the magnitude of the operating load of the building:

$$\vec{P}_{op} = \left(P_{ON}(\vec{T}) - P_{OFF}(\vec{T})\right) \cdot \vec{U}_{op} + P_{OFF}(\vec{T}) \tag{6}$$

2.5. Random load variations

The power consumption profile of an actual building is a noisy signal comprising the random switching of several automated and human controlled loads. This high frequency noise has a so called "pink spectrum" with amplitude coefficients proportional to the inverse of the frequency. We generate a noise signal $\vec{N}$ independently for each day. This noise is the *Inverse Fast Fourier Transform* (IFFT) of a spectrum with complex amplitude coefficients F_n computed as :

$$F_n = \frac{\alpha_n f_1}{f_n} e^{i(2\pi \varphi_n)} \tag{7}$$

where α_n and φ_n are random numbers picked in the interval [0, 1[. The fundamental frequency being f_1 = 1/day, the other frequencies are $f_n = n f_1$. To avoid breaking the continuity of the signal, it is shifted so that it passes through 0 at midnight.

The noise signal $\vec{N}$ is normalised as follow:

$$\vec{U}_N = \frac{1}{2} \cdot \frac{\vec{N}}{\sqrt{3}\,\sigma(\vec{N})} \tag{8}$$

where $\sigma(\vec{N})$ is the standard deviation of $\vec{N}$.

Approximately 92% of that normalized noise signal is within the range ± ½. The LDC do not allow to quantify the magnitude of the noise. However, disregarding the energy consumption related to climate, we could estimate the maximum magnitude of the noise during the building's operation hours as $\pm(\overline{P}_{\max} - \overline{P}_{ON})$. The noise is expected to follow the load, therefore its relative amplitude R_N must be in the order of:

$$R_N = \frac{P_{\max} - \overline{P}_{ON}}{\overline{P}_{ON}}. \tag{9}$$

R_N can be computed separately for each month. The noise contribution can be introduced within the profile using:

$$\vec{P} = \vec{P}_{op} + R_N \cdot \vec{P}_{op} \cdot \vec{U}_N \tag{10}$$

2.6. Final correction

Finally, the profile is adjusted to match exactly the monthly billing. We trim the peaks exceeding $P_{\max}$ and we apply a gamma correction to bring back the average power to $\overline{P}$, as in image processing [8]:

$$\vec{P} \rightarrow P_{\max} \cdot (\vec{P} / P_{\max})^{\gamma} \tag{11}$$

In summary, the algorithm uses monthly energy and peak power data to generate a realistic load distribution, through the LDC model. A part of the load distribution is considered as being synchronized with a pre-defined schedule. Then, the monthly load parameters are fitted to the temperature data through the PRISM model. A synthetic profile consistent with the daily average temperature and the building's operation schedule is generated. A random noise signal is added to the synthetic profile and finally, a correction is applied to match exactly the monthly billing data.

3. SIMULATION EXAMPLE

We present an example of a simulated profile. The monthly data $P_{\max}$ and $\overline{P}$ have been computed from a metered profile (15-minute time step) of a building located in Montreal (to enable the comparison between metered and simulated profiles). These data metered during year 2005 were extracted from the database of Hydro-Quebec's *Visilec* program. We specified the building's operating schedule $\vec{U}_{op}$ as follows: the loads are activated between 6:00 and 18:00 from Monday to Friday, but not during the weekend. With this schedule, $\tau_{ON} = 0.357$.

Figure 2 and Table 2 show the main parameters of the simulation for each month. $\overline{P}_{ON}$ and $\overline{P}_{OFF}$ are calculated by integrating the LDCs computed from $LF = \overline{P} / P_{\max}$. The values of γ exponents are generally close to 1, indicating a minor final correction.

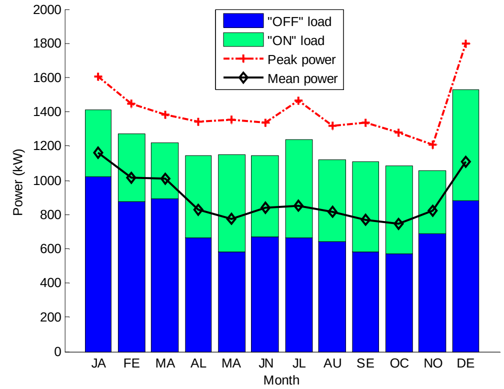

Figure 2. Monthly power data

Table 2. Monthly parameters

Month	$\overline{T}$ (°C)	$\overline{P}$ (W)	$P_{\max}$ (W)	$\overline{P}_{ON}$ (W)	$\overline{P}_{OFF}$ (W)	γ –
Jan.	-10.5	1159	1605	1420	1028	0.99
Feb.	-6.0	1015	1445	1266	880	1.12
Mar.	-2.8	1009	1382	1220	885	0.88
Apr.	7.9	827	1341	1145	660	1.01
May	11.9	777	1356	1135	590	1.08
June	21.7	841	1339	1151	669	1.00
July	22.9	852	1464	1248	652	1.01
Aug.	22.1	815	1318	1122	645	1.09
Sep.	18.1	767	1336	1121	574	1.11
Oct.	10.3	743	1280	1078	576	1.12
Nov.	3.1	821	1206	1054	691	1.17
Dec.	-7.0	1111	1801	1526	879	0.95

The monthly data $\overline{P}$, $\overline{P}_{ON}$ and $\overline{P}_{OFF}$ were correlated with the outdoor temperature through curve fits with the *PRISM* equation. The coefficients obtained from these regressions are given in Table 3 and the curves are displayed in Figure 3.

Table 3. Curve fitting of monthly power data using *PRISM* model[+]

Parameter	$\overline{P}_{ON}$	$\overline{P}_{OFF}$	$\overline{P}$
Constant *K* (kW)	611.3	1090.8	777.8
Heating coef. *H* (kW/°C)	25.4	22.2	24.6
Cooling coef. *C* (kW/°C)	11.3	15.8	12.8
Std. dev. of error (kW)	33.7	67.7	26.4

[+] T_H = 5.3°C and T_C = 17.7°C (fitted to $\overline{P}$ *vs.* $\overline{T}$).

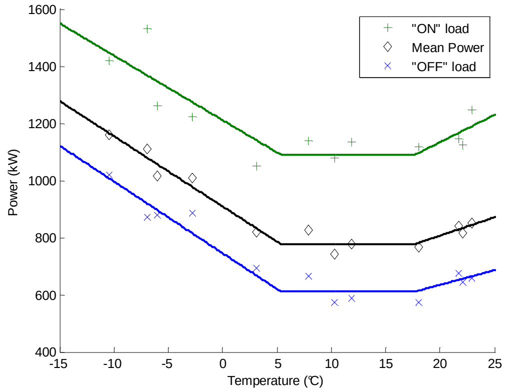

Figure 3. Curve fitting of monthly power data using the PRISM model

The metered and simulated data of two weeks in winter and two weeks in spring are shown respectively in Figures 4 and 5. The January data (Figure 4) have a wavy appearance due to a significant heating load. On the other hand, in May (Figure 5), the average temperature is outside the heating and cooling ranges bounded by T_H and T_C, so the climatic load is negligible. In the metered profiles shown in these two Figures, we note that a load is occasionally added in the weekends and during nights. For simulation purpose, the schedule could have been defined in a different way to reflect that behaviour. The average monthly LDC is shown in Figure 6. The distribution of the power demand is much smoother in metered than in simulated data because of some partial loads that do not respect the proposed schedule.

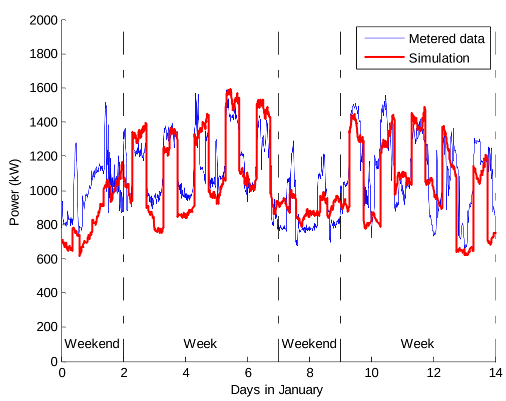

Figure 4. First weeks of January (metered and simulated data)

Figures 4 and 5 allow viewing the random noise, which is added to the profile at the end of the calculation. In addition, Figure 7 compares the average of the frequency spectra of every days calculated using a *Fast Fourier Transform* (FFT), for metered and simulated data. The slope of the dotted line represents the inverse of the frequency, which is the envelope of the daily power spectrum. The ripples in the spectrum of simulated data are a consequence of the lack of variability in the schedule.

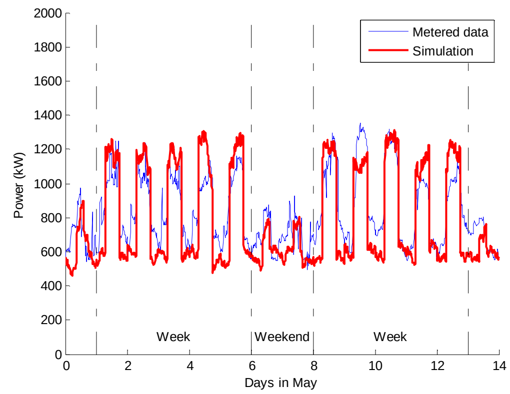

Figure 5. First weeks of May (metered and simulated data)

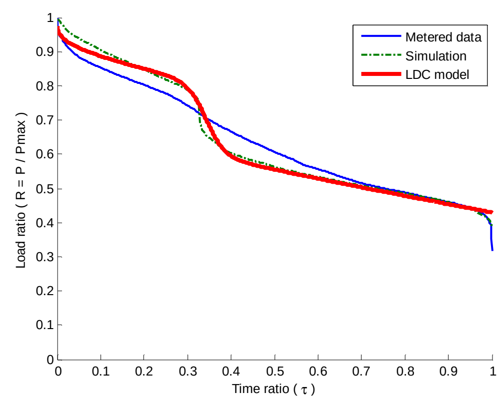

Figure 6. Load duration curve for an average month.

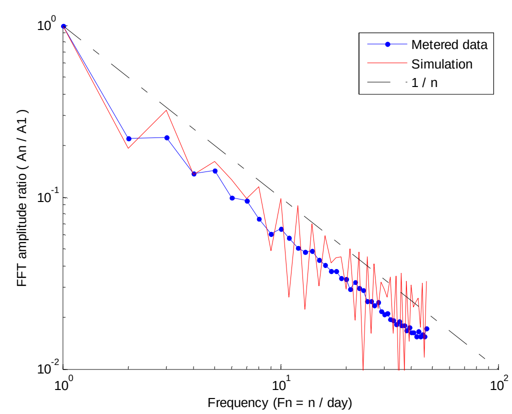

Figure 7. Amplitude of daily frequency spectrum (average amplitude through the days).

To validate the simulation, we have compared the maximum, minimum and average power of each day of the year between the metered and simulated profiles. Figure 8 shows the correlation between these daily data. In Table 4, we observe that the correlation coefficient *r* is 0.94 between the daily mean power series, and over 0.8 for the maximum and minimum power data, in spite of the noise. Of course, the algorithm is not expected to reproduce exactly an actual profile but its main characteristics such as the daily minimal and maximal power. The load schedule and the random power variations are approximated with satisfactory results, given the small amount of inputs needed for the simulation program.

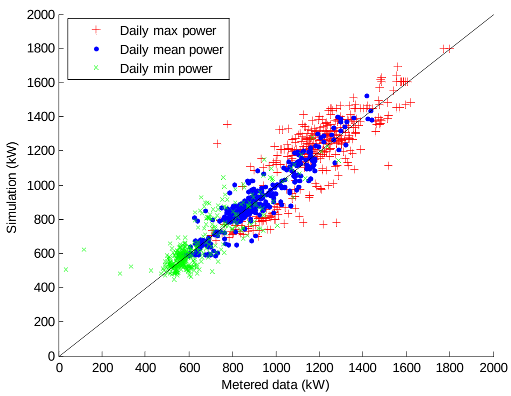

Figure 8. Maximum, mean and minimum power daily

Table 4. Daily data statistics

Type of daily power data	Average value (kW)	Average error (kW)	Std. dev. of error (kW)	*r*
Max. of each day	1181	6.8	127	0.84
Min. of each day	648	-2.5	93	0.83
Mean of each day	895	0.0	59	0.94

4. EXAMPLE OF APPLICATION: Sizing thermal storage systems

Several demand management technologies are used to reduce peak electricity demand, which is expensive for the consumers and power utilities. The billing of monthly peak demand is an incentive for customers to adopt best practices of power management. However, before implementing new technologies, the customers want to know the economic benefits of installing the proposed equipments. The calculation of potential savings requires knowledge of the consumption profiles on a finer scale than the monthly billing data. The simulated profiles can therefore meet the needs of customers without access to communicating meters.

Electrical Thermal Storage (ETS) systems can store heat into a mass of a thermal storage material during the night and reduce the power consumption for heating during day [9]. Metered and simulated profiles have been used as inputs to a computer program developed for sizing these systems. That software calculates the load that can be shifted to off-peak periods using the heat storage capacity. To do so, it analyzes the consumption profile (by 15-minute time steps) that is provided. Table 5 shows the results of sizing and savings estimated by the software for 5 buildings, as the consumption profiles used as input are either metered or simulated. We note that the proposed configurations are the same, whether obtained from the metered or the corresponding simulated profiles. Nevertheless, the financial savings differ slightly, with an average deviation of ±5% (maximum deviation is ±16%), which is acceptable at the level of a technico-economic analysis.

Table 5. Validation of the simulated profiles through the sizing of ETS systems.

Building	Data type[+]	Required Storage capacity (kWh)	Annual energy bill before install ($)	Annual energy bill after install ($)	Savings ($)	Error on savings (%)
Office building[++]	M	1 920	500 921	478 204	22 717	-11%
	S	1 920	500 926	480 688	20 238	
HQ building 1	M	1 920	1 057 581	1 023 465	34 116	0%
	S	1 920	1 057 592	1 023 604	33 988	
HQ building 2	M	1 440	409 401	381 471	27 929	13%
	S	1 440	409 403	377 951	31 452	
Primary school 1	M	480	67 938	58 525	9 413	-16%
	S	480	67 940	59 996	7 944	
Primary school 2	M	480	25 966	18 424	7 543	0%
	S	480	25 986	18 442	7 544	

[+]M = Metered data ; S = Simulated data.
[++]The metered and simulated data of the "Office building" are presented in Section 3.

5. DISCUSSION

Simulating a load profile based only on monthly data and a daily schedule involves the risk of incorrectly distributing the power consumption over time.

However, the dependence over the temperature seems to be correctly reconstructed in the cases of buildings with common HVAC systems. A more difficult problem stands in representing the day and night loads. We solved it using a schedule and LDCs. We could eventually refine the model to include partial loads in the user-defined schedule. The LDC generation algorithm has been calibrated with a large set of buildings data. For the purpose of this work, a large number of synthetic LDC were generated and the resulting average curve was used.

When the *LF* is low, the random noise occupies a large part of the signal magnitude. As it is modeled with very basic frequency dependence, it cannot reflect real building load fluctuations, which are not truly random. On the other hand, when the *LF* is very high, the weak distinction between the day and the night loads makes even more difficult to estimate their magnitude correctly. With medium *LFs*, the LDC parameters are allowed to take large range of values. Hence, giving the statistical distribution of the LDC parameters, instead of only their range, would increase the confidence in the computed parameter values.

Therefore it appears that the simulation procedure works best for buildings with common systems, schedules dominated by occupancy and climatic loads, and intermediate *LF* values, e.g. office buildings. In these conditions, we can rely on the simulated profiles both for forecast and for evaluating the profitability of new technologies like heat storage systems used for load management.

6. CONCLUSION

We have developed a method for simulating energy consumption profiles similar to those measured by the communicating meters in the Hydro-Quebec's program *Visilec* (with 15-minute time steps). It only requires a year of electricity billing on a monthly basis, and temperature data on hourly or daily basis. It weights the power demands during days and nights through the *Load Duration Curve* model. Simulated profiles can be used by engineering firms when metered data are not available to design systems and calculate the cost savings, when analysing energy efficiency projects or peak demand smoothing options.

In this perspective, we tested the quality of simulated profiles by comparing the outputs of an electrical thermal storage system design tool developed by Hydro-Quebec's *Laboratoire des Technologies de l'Énergie* (LTE). Metered and simulated profiles of 5 buildings were provided as input to that tool. These preliminary tests showed that the measured and simulated profiles lead to similar conclusions about the sizing of thermal storage systems, even if the estimation of savings differs up to ± 16%. This divergence is quite low, given the limited information involved in the implementation of the model.

We have identified some means for improving the simulation algorithm, such as creating a larger database of *Load Duration Curves* being classified by building types. In addition, sub-measurement and detailed simulations could provide relevant information about the loads and schedules, thereby validating the model assumptions. The simulation algorithm could also be modified to manage partial loads.

In conclusion, a new simulation tool was developed to assist the implementation of efficiency measures by consulting engineering. However, the performance of the algorithm has been verified for only few cases. An extensive testing procedure would be desirable before the deployment of the tool.

REFERENCES

[1] J.A. Dominguez-Navarro, J.L. Bernal-Agustín, R. Dufo-López, Data mining methodology for disaggregation of load demand, *Electric Power Systems Research 79* (2009) 1393-1399.

[2] H. Akbari, Validation of an algorithm to disaggregate whole-building hourly electrical load into end uses, *Energy 20* (1995) 1291-1301.

[3] H. Hahn, S. Meyer-Nieberg, S. Pickl, Electric load forecasting methods: Tools for decision making, *European Journal of Operational Research 199* (2009) 902–907.

[4] A. Poulin, M. Dostie, M. Fournier, S. Sansregret, Load duration curve: A tool for technico-economic analysis of energy solutions, *Energy and Buildings 40* (2006) 29-35.

[5] M. F. Fels, PRISM: an introduction, *Energy and Buildings 9* (1986) 5-18)

[6] D. Popescu, F. Ungureanu, A. Hernández-Guerrero, Simulation models for the analysis of space heat consumption of buildings, *Energy 34* (2009) 1447-1453.

[7] C.-L. Hor, S.J. Watson and S. Majithia, Analyzing the impact of weather variables on monthly electricity demands, *IEEE Transactions on Power Systems 20(4)* (2005), 2078–2085.

[8] G. Finlayson, R. Xu, Illuminant and gamma comprehensive normalisation in logRGB space, *Pattern Recognition Letters 24(11)* (2003), 1679-1690.

[9] A. Moreau, T. Steffes, Managing the Heating Demand in Institutional Buildings with an Innovative Thermal Storage System, *11th International Conference on Thermal Energy Storage - Effstock 2009 - Thermal Energy Storage for Energy Efficiency and Sustainability*, Stockholm (2009), July 14-17.