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

DAILY LOAD PROFILE CLUSTERING: A TOOL FOR SIMULATION CALIBRATION

Michaël Fournier, Eng., MASc and Karine Lavigne, Jr. Eng., MASc
Institut de recherche d'Hydro-Québec, Laboratoire des technologies de l'énergie
Shawinigan, Québec, Canada

ABSTRACT

This paper presents a clustering algorithm of building's electrical load profiles and its custom made visualization tool. The algorithm groups daily load profiles based on their shape in order to produce a reduced set of typical profiles. The visualization tool shows these building's typical profiles and their distribution throughout the year. Sample cases are given to demonstrate the diagnostic capability of the clustering approach for the calibration of building energy simulations. The first case shows how the observed typical daily load profile for a specific time period can be used as an aggregated operation schedule indicator. The approach is also used to validate the transition from heating to cooling dominated thermal response of the building. Another example exposes higher than predicted night consumption which suggests an improper temperature setback. This analysis tool, allowing the comparison of simulated to measured data, has been implemented in a DOE2.1-based simulation software called SIMEB.

INTRODUCTION

The impact of an energy-wise retrofit on a building performance can be determined by comparing the simulated consumption of the actual and retrofitted buildings. The use of energy simulations has proven to give meaningful insight on the energy saving potentials of the considered scenarios. Well-calibrated building simulations are particularly suitable to account for multiple energy end-uses, especially where interactions may occur between energy saving measures [ASHRAE Guideline 14-2002]. At the owner and operator levels, proper knowledge of the building's actual behavior is not only of interest approaching a retrofit; it can also be used for benchmarking and the optimization of its operation through the detection of malfunctions or unexpected behavior [Salsbury 2000].

Commonly used building energy simulation software is now hourly or sub-hourly based. Parameter's time dependencies (i.e. schedules) therefore become crucial inputs to the simulation. Moreover, the use of hourly simulation requires not only a detailed description of the building and its operation parameters but also a thorough calibration process. A review of the published concepts, issues and proposed procedures for simulation calibration can be found in [Reddy 2006]. The calibration of such detailed simulations requires a comprehensive corroboration, not only of the as-built building properties, but also of its systems operating mode in response to the prevailing weather and occupancy. Acknowledging that this detailed information is hardly available and expensive to collect, it is proposed to complement it by exposing the electrical load profile patterns of the building.

Automated meter reading (AMR) is becoming widely used in non-residential buildings. The time series of measured power loads reflect the building response to weather, operation mode and occupancy. The proportion of the response that impacts the electrical consumption is however not straightforward and changes from building to building. Nevertheless, the growing availability of the AMR data makes them attractive to decode for gaining building operation insight and simulation calibration.

A one year long hourly sampled load profile contains 8760 readings. Using a scatter plot is an ineffective way to graph them. Fluctuations occur on hourly, daily, weekly and seasonal scales, all of which cannot be grasped from a single traditional graph even for an experienced analyst. Several types of display have been proposed to ease the interpretation of hourly profiles [Bou-Saada 1995, McCray 1995]. Nevertheless, inexperienced users may still feel overwhelmed by the amount of information presented. To broaden the exploitation of load profiles, tools are required to extract and synthesize the meaningful information. Clear and user friendly ways to render them are also essential.

This paper proposes the use of clustering as an unsupervised data mining technique to identify typical daily load profiles of single buildings. Load profile clustering has been proposed to compound the load profiles of several customers at the utility level [ex. Chicco 2006], principally for cost allocation and definition of rate structures. To the authors' knowledge, the use of this feature extraction method for grouping individual daily load profiles of a single building is original. A single related application has been found [Omnisolve 2010] but no formal description of the process was exposed.

The paper describes AMR data preparation and clustering procedure. It also presents the display of the clustering solution as included in the SIMEB software. The tool is then used as a data mining and presentation technique to analyse measured and simulated load profiles for two different buildings.

METHODOLOGY

While most of the energy sources may be compiled on a monthly basis, only electricity commonly has a monitoring frequency allowing the observation of both weather dependency and operation schedule. Such behavior can be seen from 15 minutes and 1 hour interval time series. While 15 minutes intervals grasp more information, hourly data offers good balance between resolution and averaging random fluctuations.

Clustering is a technique of data segmentation in which individual components are matched based on their level of similarity. A variety of algorithms are used to create classes (or clusters) within the population of components such that the similarities between members of the same class are high while the similarities between members of different classes are low. The unique distribution of components to different classes as a result of the clustering algorithm with a determined set of parameters is called a solution.

To select the proper set of clustering parameters and algorithm, several of their combinations have been applied to a database of load profiles. That database contains the load profiles of more than 200 commercial and institutional customers with different building main activity: health care, banking, groceries, general retailing and education. The Cluto clustering toolkit [Karypis 2009] was used to generate the solutions. The iteratively produced clustering solutions were visually compared allowing the selection of the combination leading to the overall best balance between details' representation and interpretability.

Data formatting

The input format of the clustering algorithm can be formulated as vectors, each vector representing the load profile of a single day. Assuming hourly sampled data, the input space has 24 dimensions. In order for an event occurring at 1:00 AM local time to always be represented by the first dimension of that space, retrieved load profiles often have to be adjusted for daylight saving time. This can simply be done by *duplicating* the last value occurring before the change to daylight saving time and *deleting* the last one occurring before the return to normal time. This way, the nth daily load profile (D) since the beginning of the time series can be expressed as:

$$D_n = [H_{n,1} \quad H_{n,2} \quad \dots \quad H_{n,24}], \qquad (1)$$

where H are the hourly loads and the second index is the local time at which the data has been recorded. The so described input vector makes use of no formatting. Each hourly load is the total electrical power load measured over the hour (kWh/h or kW) preceding the timestamp. The time series can be of any number of days. About a year long dataset is however required to monitor weather related behavior.

Similarity function

Similarity functions are applied to the vector input space and are used to define de level of likeness between two vectors. The similarity function selected for this application is the cosine, stated as:

$$\cos(D_j, D_k) = \frac{D_j^{\,t} \cdot D_k}{\|D_j\| \, \|D_k\|}. \qquad (2)$$

This function gives the angle formed between the two input vectors. It is thus insensitive to their respective magnitude; only their direction matters. The cosine function responds to the time of the day at which changes in consumption occur. In this way, it leads to more valuable solutions than the Euclidean distance. The latter is more responsive to the average load of the day.

Algorithm and objective function

The Cluto toolkit allows using different allocating algorithms. Based on the relatively limited dimensionality, number of elements and number of desired classes, the "direct" algorithm [Karypis 2009] was expected to give good results. A two step process however gave slightly better results. The first step uses a "direct" method in which 10 clusters are resolved simultaneously. That solution globally maximizes the selected objective function:

$$\sum_{i=1}^{k} \sqrt{\sum_{D_j, D_k \in S_i} sim\left(D_j, D_k\right)}, \qquad (3)$$

where k is the number of desired classes (10), sim is the selected similarity function, D_j, D_k are a pair of daily load profiles vectors belonging to the $S_{i^{th}}$ class. This function maximizes the sum, over all classes, of the sums of similarities for the pairs of vectors belonging to a class. The second step involves the sequentially merger of pairs of classes until the desired number of classes is left. The pair of classes to merge at each step is selected so as to maximize the same objective function used for the first step, i.e. equation 3. For example, the solution having 4 classes is obtained by merging two classes of the solution having 5 classes.

Number of classes

The number of classes created is generally an input to the clustering algorithms. Based on our observations, the optimal number of classes when clustering daily load profiles of a single building is 6 or less. However, the optimal number of classes is expected to vary with building occupation schedules and operating mode and thus from building to building. Solutions having 2, 3, 4, 5 and 6 classes are hence produced. While having 5 solutions to work with allows flexibility, it is desirable to propose the most appropriate one to the user. For such recommendation, a criterion based on the distribution of the classes throughout the year has been developed.

From our observations, when the number of classes is too low, the classes form large homogenous groups in a calendar. Groups are defined as contiguous days in the calendar that belong to the same class. See the left part of figure 1 for an example of calendar. On the other hand, when the number of classes is too high, the days belonging to different classes are more or less randomly intertwined in a calendar, resulting in a high number of groups for each class. We observed that the best

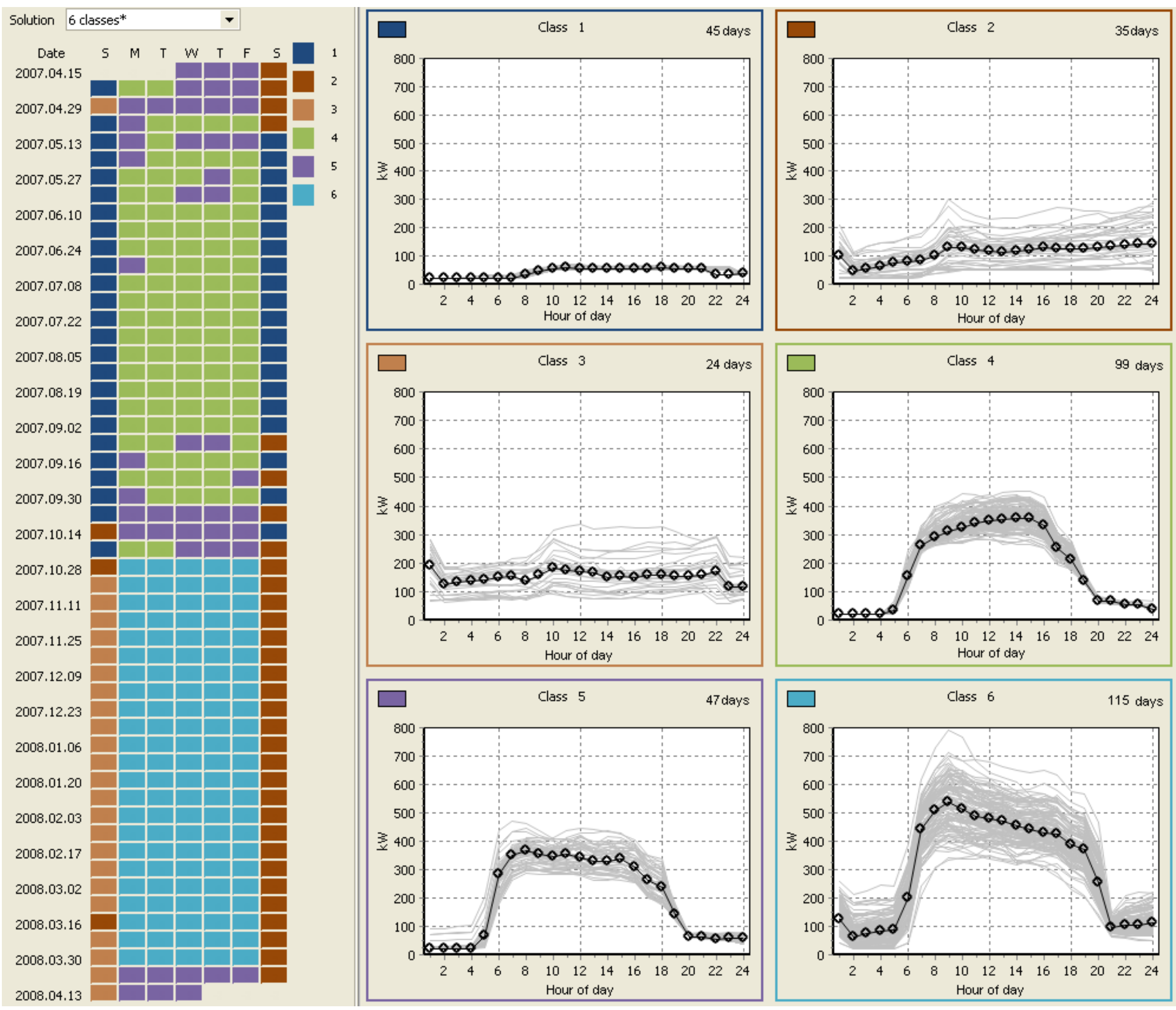

Figure 1: Building #1, simulated data

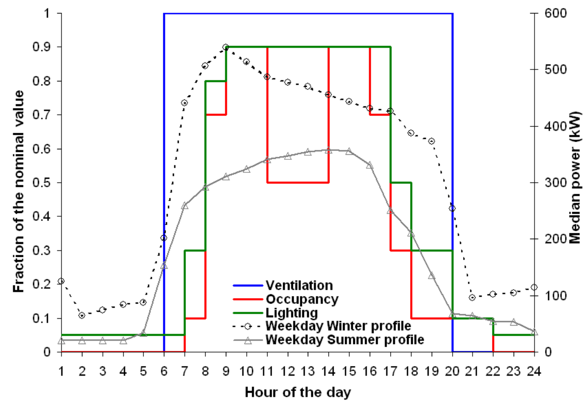

Figure 2 : Schedules used for weekdays and observed typical load profiles

solutions often have a ratio of the total number of groups over the number of classes being near three but not exceeding it. How the solutions are displayed is discussed in the next section.

SOLUTION DISPLAY

An appropriate display of the solution should be easy to understand, highlight the most important trends and keep enough details to allow proper interpretation. Figure 1 gives an example of the proposed display. It is composed of two parts: the first one shows the temporal distribution of the classes while the second presents the typical load profile of each one.

The left part of the figure is a running calendar in which each row corresponds to a week and each line corresponds to a day of the week, starting on a Sunday. In this calendar, days having a daily load profile belonging to a common class are represented by a unique color. The number of classes of the solution to

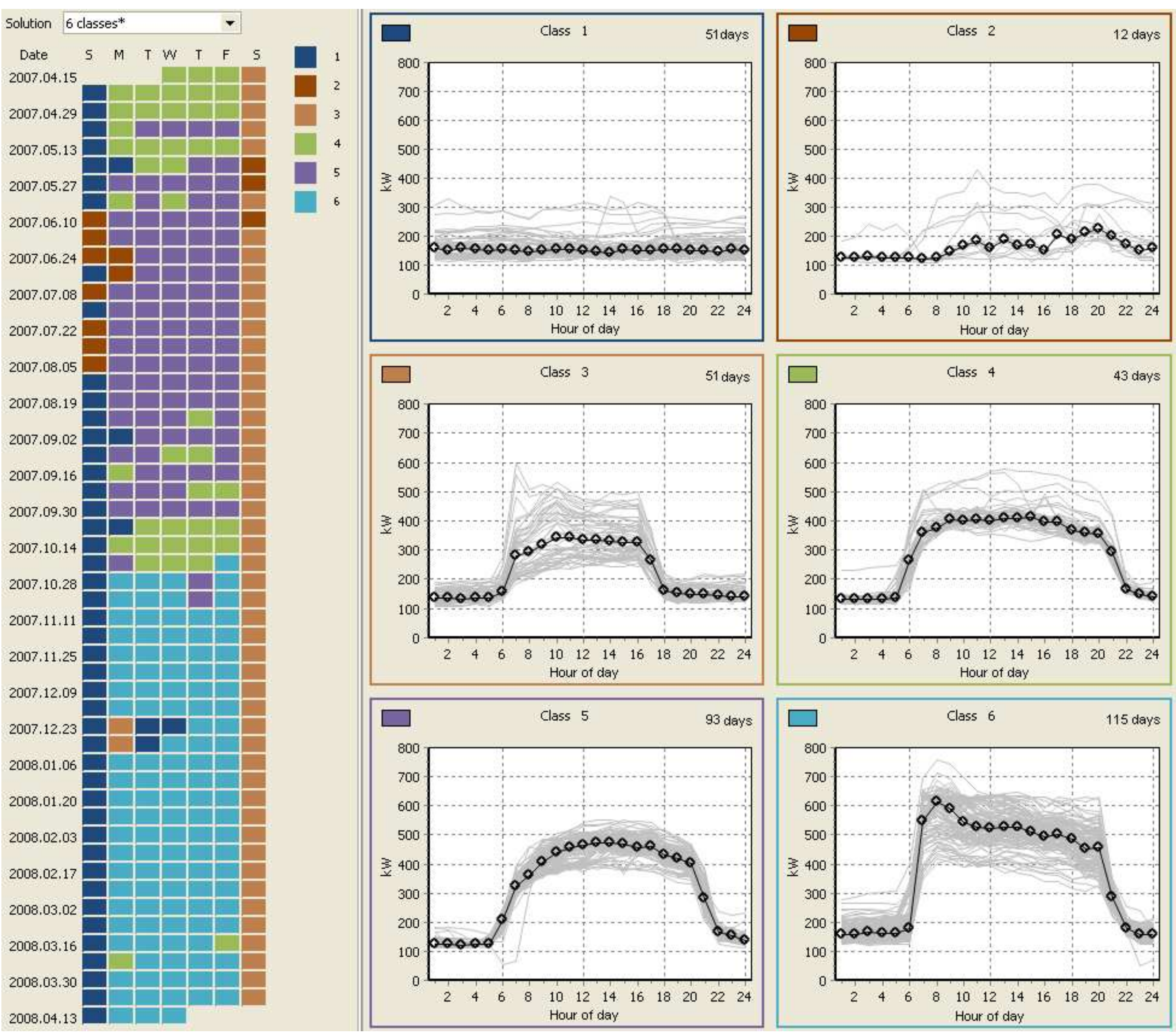

Figure 3 : Building #1, AMR data

be displayed can be selected, at the calendar's top, from a drop-down list. Therein, the recommended solution is indicated with an asterisk.

On the right part, there is one graph for each of the identified class. Each graph shows each and every load profile associated with this class as well as the typical load profile of that particular class shown in color. The same color is used to represent that class in the calendar. The typical load profile is taken as the median load profile of the members of the class. That lowers the influence of extreme or misclassified profiles. The number of members of each class is shown above the graph.

ANALYSIS

This section presents the results obtained with the clustering tool for two different building simulations and their respective measured data. Those examples show how the visualization tool can highlight differences between the measured and simulated data and offer a diagnosis of the building's operation.

Building #1

The first example is a 8097 square-meter office building with electrical heating and cooling and a specific consumption of 321 kWh/m²/year.

The building was simulated using a DOE2.1-based simulation software called SIMEB (SIMEB 2009). The simulation was run from April 18th 2007 to April 16th 2008 since the measured data was available for this period. The weather file used was constructed using data from the closest weather station for the period of simulation. No standard weather file was used. The schedules used were the software's default schedules for an office building. Those were originally defined in the Canadian Building Incentive Program. Occupancy, ventilation and lighting weekday schedules are shown on figure 2. During weekends and holidays, occupancy and ventilation schedules are set to 0 while lighting schedule was maintained at 5% of the maximal lighting capacity.

The simulation predicts a specific consumption of 216 kWh/m²/year, which is about 33% lower than the actual consumption.

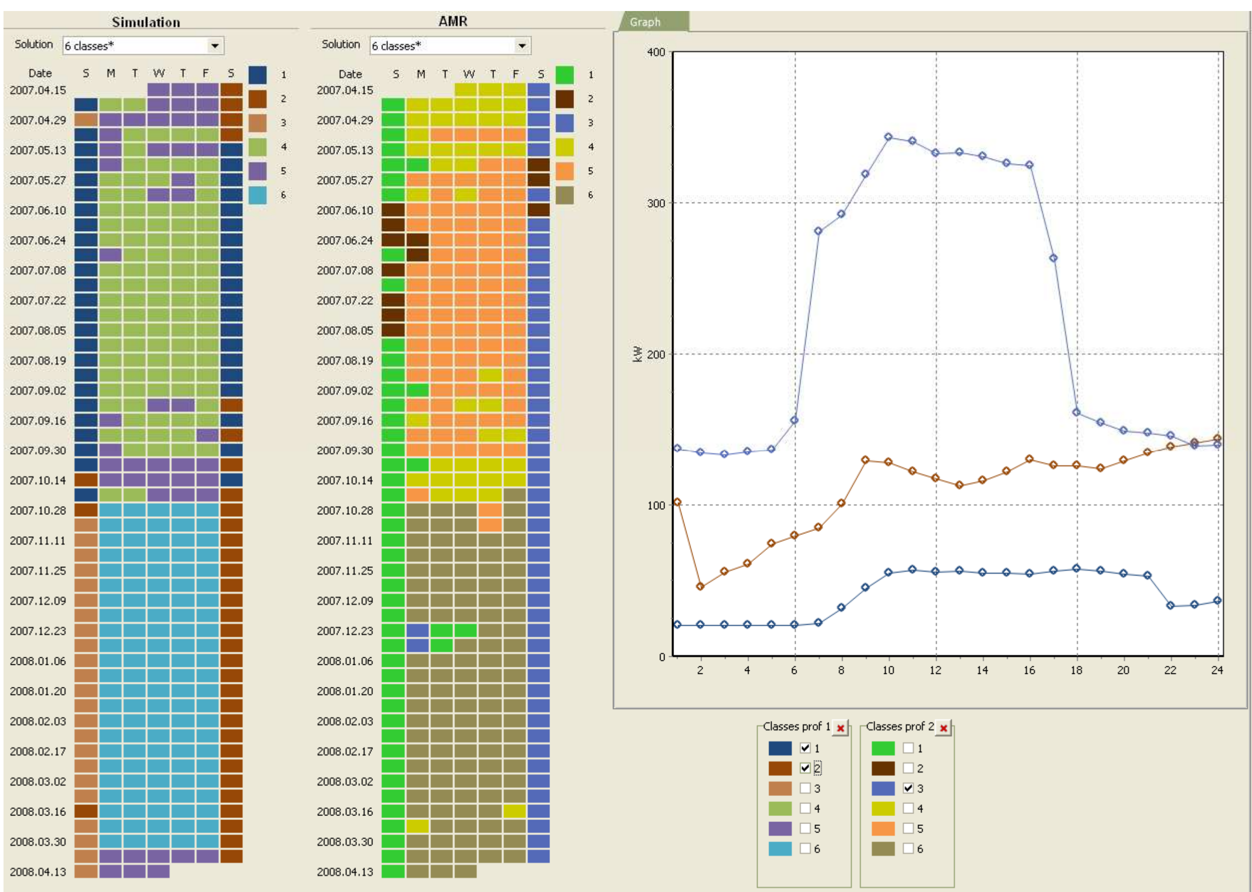

Figure 4 : Building #1, comparison of Saturdays' median profiles

The results obtained with the clustering tool are shown on figures 1 and 3, respectively for the simulated and AMR data. The simulated winter and summer median weekday profiles were also plotted on figure 2. The graph clearly indicates that the profiles are mainly driven by the ventilation schedule. In figures 1 and 3, the summer and winter weekday load profiles are easily distinguished. In winter, profiles show a sharp peak at the systems' start-up while the summer week profiles usually reach a maximum around midday. Transition periods, between summer and winter modes, occur in April and October for both sets of data. During those periods, the HVAC systems oscillate between heating and cooling. The resulting daily load profiles thus do not exhibit the characteristics of either heating or cooling profiles. As can be seen on the figure 1, class 5, and on the figure 3, class 4, graphs, the median load profile is rather flat during daytime.

It is also interesting to note that holidays are underlined in the measured data's calendar (May 21st, June 25th, July 2nd, September 3rd, October 8th, also December 25th-26th and January 1st).

When comparing the simulation with the measured data, a major difference can be observed. The simulation shows no activity for all weekends whereas the measurements clearly indicate that there is some activity on Saturdays throughout the year. As can been seen on figure 4, the visualization tool allows a graphical comparison between the two analysis. The median for Saturdays of the measured data is plotted against the median summer and winter simulated Saturday load profiles. This difference in the profiles can account for about 21 kWh/m²/year. If the profiles were corrected, the difference in the annual specific consumption would reduce to 26%. Furthermore, the median power demand during unoccupied hours is

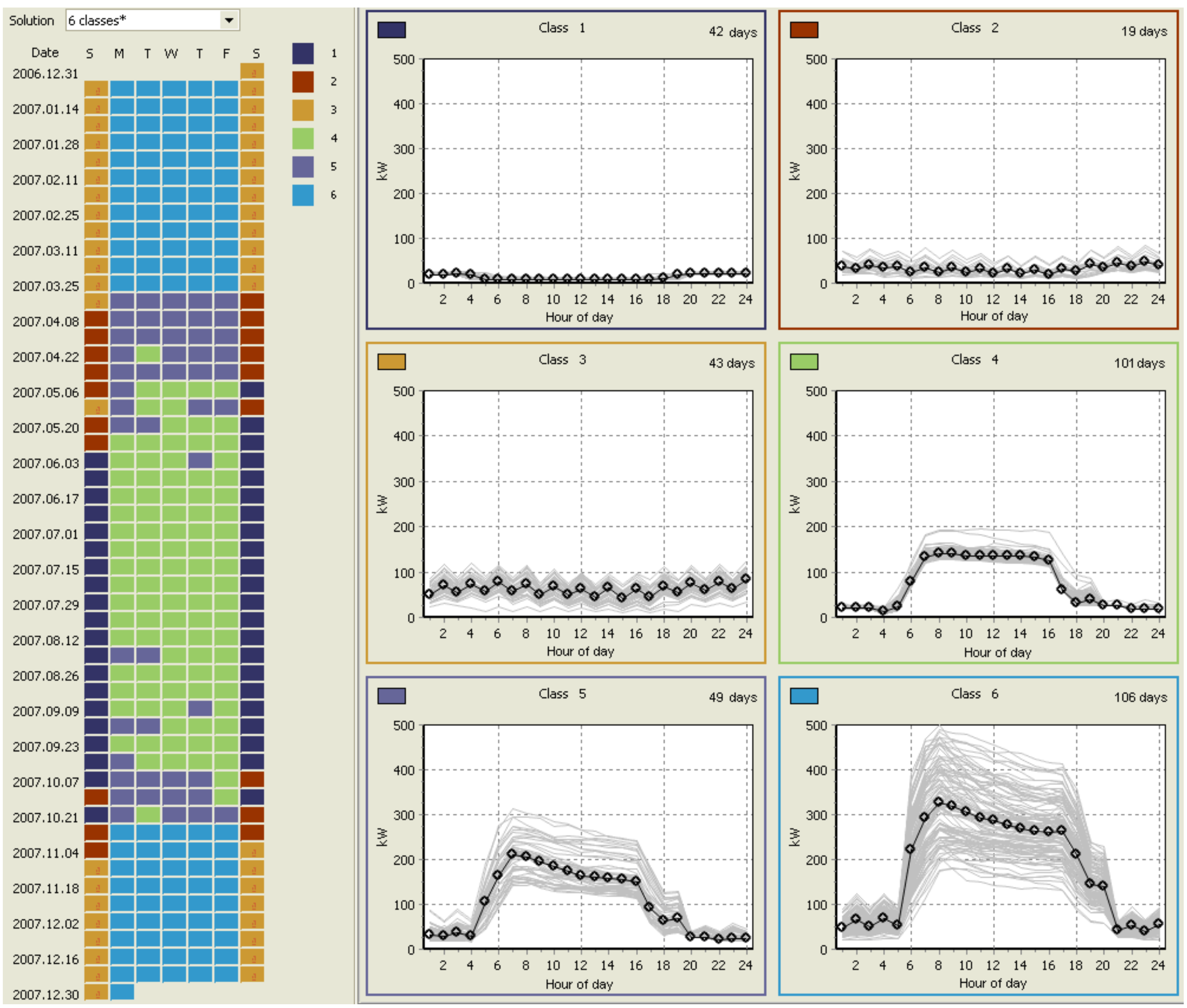

Figure 5 : Building #2, simulated data

underestimated in the simulation. The probable causes of this discrepancy are: improper temperature setbacks, higher than estimated night plug loads and erroneous HVAC schedules. The gap in power load during these hours causes a difference of about 41 kWh/m²/year between the simulation and the measurements. If this could be fixed, the error on the annual specific consumption would reduce to about 13%.

Building #2

The second building is one of Hydro-Quebec service center. It is a 4035 square-meter building including office, kitchen, workshop and garage spaces. There are nine rooftop units with electric cooling and heating. Garages and workshops are also equipped with electric unit heaters. The specific consumption of the building is 264 kWh/m²/year.

A detailed on-site survey was conducted in order to adjust the simulation parameters adequately. For this case study, following an interview with the maintenance personnel, the default schedules were modified in order to represent the real operation of the building.

The simulated building resulting specific consumption is 216 kWh/m²/year, 18% lower than the measured value. As for the first building, the simulation was performed with local weather data for the simulation period.

Figures 5 and 6 show respectively the results obtained with the clustering tool for the simulated and measured data. It should be noted that, for the measured data, although the tool recommended the two-class solution, the four-class solution was chosen for an easier comparison with the simulation. The occupied hours seem to be similar for both results. However, there are more fluctuations in the measured data classes. The distribution of classes is not as straightforward as in the

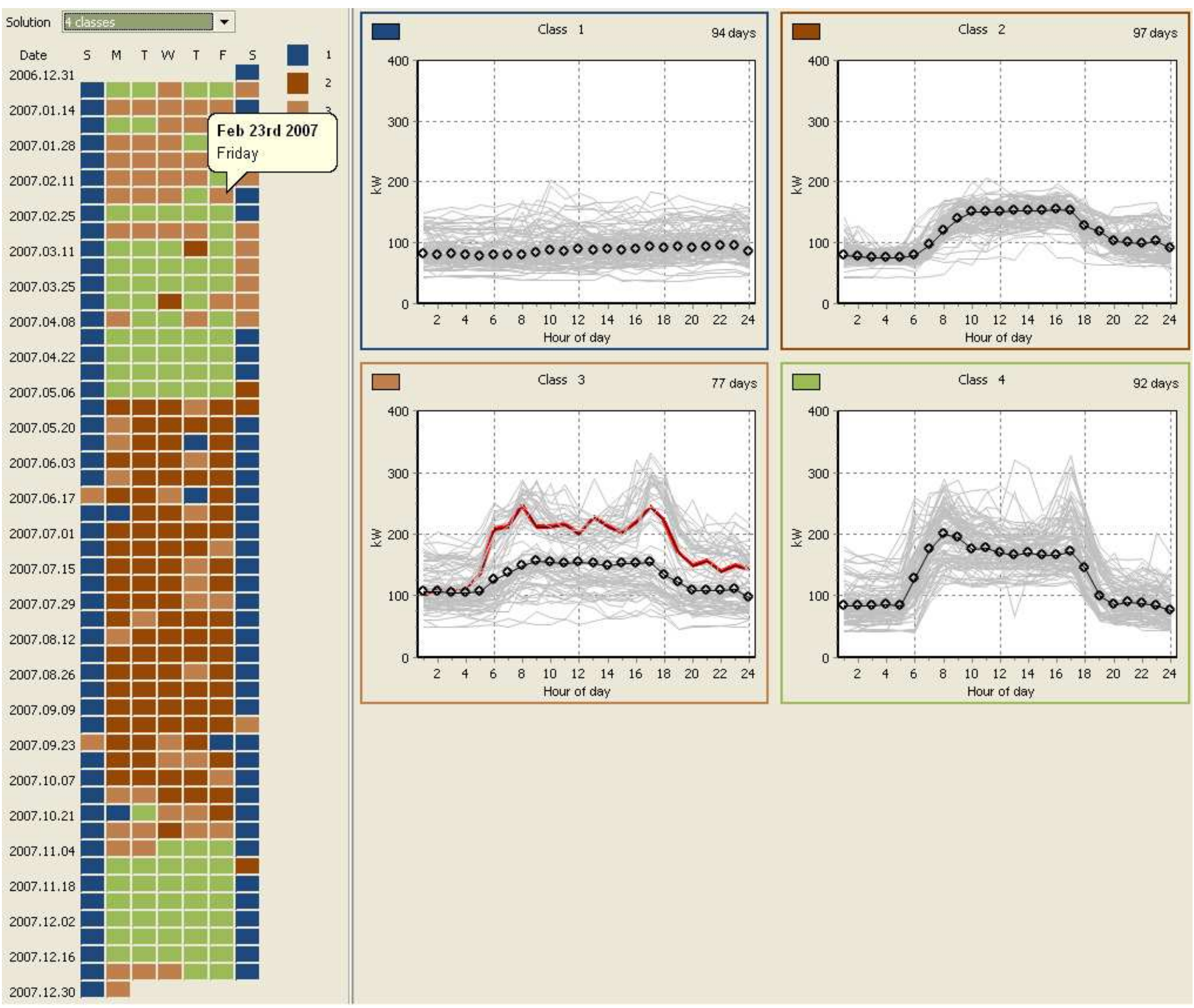

Figure 6 : Building #2, AMR data

first building. The algorithm attributed load profiles to the third class whereas some should have been included in other classes. For instance, February 23rd profile is highlighted in figure 6. From its shape and amplitude, it could have been part of class 4. However, from the criteria embedded in the clustering algorithm, it was attributed to another group.

Nevertheless, this example demonstrates that, even though a detailed on-site survey was done, there is still a large difference between simulated and measured median power demand during unoccupied hours. The measurements indicate that there is not an important night load reduction in the building, even in the winter. This could be due to equipments running continuously or to constant temperature set points in the different functional spaces. Sub-measurement should be performed on the building to acquire overlooked data and determine the cause of discrepancies.

CONCLUSIONS

The developed clustering tool makes AMR data a convenient source of information for energy simulation calibration. By grouping daily load profiles and identifying typical ones, it highlights the building's behaviour over the analysed period. Without any input from the user, it depicts some aspects of the building's response to the weather, as it changes with the time of the year, and to the operation schedules.

The exposed examples illustrate how the synthesized information can be interpreted for the validation of the simulation hypothesis. Example one showed how the tool can help the user identify heating and cooling periods as well as the transition between the two. It also emphasises the link between schedules and median power profiles and how this can be exploited to correct the simulated data in order to fit the measurements. The second example, by drawing attention to a load level unexpected from the simulation hypothesis, suggested improvements to the simulation.

Further improvements of the tool are expected. Some of the produced classes are too heterogeneous. The suggested number of classes can also be perfected.

ACKNOWLEDGEMENTS

The authors would like to thank their teammates of the SIMEB and "Simulation et caractérisation de la consommation électrique" projects. The relevant and informative comments of the reviewers are also acknowledged.

REFERENCES

ASHRAE (2002), *ASHRAE Guideline 14-2002, Measurement of Energy and Demand Savings*.

Bou-Saada T.E. and J.S. Haberl (1995), 'An improved procedure for developing calibrated hourly simulation models, *IBSPA Conference*, Madison, WI, Aug. 14-16, 475-484.

Chicco G. et al. (2006), 'Comparisons among clustering techniques for electricity customer classification', *IEEE transactions on power systems*, 21, 2, 933-940.

Karypis G. (2009), Cluto – software for clustering high-dimensional datasets, software documentation, http://glaros.dtc.umn.edu/gkhome/cluto/cluto/overview, retrieved on 09/01/12.

McCray J.A. et al. (1995), 'Using data visualization tools for the calibration of hourly DOE-2.1 simulations, *IBSPA Conference*, Madison, WI, Aug. 14-16, 461-466.

Omnisolve (2010), Powervise, Energy engineering metering analysis, http://www.powervise.com/PowerVise/Tour/Base.htm, retrieved on 2010/02/22.

Reddy T.A. (2006), 'Literature review on calibration of building energy simulation programs: uses, problems, procedures, uncertainty and tools', *ASHRAE trans.*, 112, 226-240.

Reiter P.D (1986), 'Early results from commercial ELCAP buildings: Schedules as a primary determinant of load shapes in the commercial sector', *ASHRAE trans.*, 92, 297-309.

Salsbury T. and R. Diamond (2000), 'Performance validation and energy analysis of HVAC systems using simulation', *Energy and buildings*, 32, 5-17.

SIMEB (2009), www.simeb.ca.