Energy buildings management methodology

P.R.S. Jota
Centro Federal de Educação Tecnológica de Minas Gerais, CEFET-MG, BH, Minas Gerais, Brazil

V.R.B. Silva
CEFET-MG e Prefeitura Municipal de Belo Horizonte, BH, Minas Gerais, Brazil

ABSTRACT

In this paper, a new approach for energy consumption and peak demand predicting in buildings is shown. The method is based on a mathematical model of load curve and energy curve. This model is obtained after classification of typical curves using a multivariate technique called Cluster analysis. The performance of this predictor was evaluated using real data. The achieved results demonstrate the good precision reached with this system.

1. INTRODUCTION

A load-modelling task is very important to help a building energy analysis and operation. The energy manager has to use it in energy optimisation process. The load modelling quality strongly influences the energy management (Michalik and Mielczarski, 1998; Nazarko and Styczynski, 1999; Livik et al., 1993; Murata et al., 1991; Silva and Jota, 2004). The building load curve is a sum of all equipments demand. The building can have different daily load curves and its shape strongly influence the cost of the energy. To analyse the behaviour of daily load shape, the manager uses, in most of the cases, his experience. A load curve is a time series that represents the way the company uses energy. During a week, the company can or cannot have a very similar way to work and the load curve can show how similar it is. During the weekend, most of the companies have a different time schedule. The energy manager has to know how historically the load grows during the day and than compare with the load tendency to take decisions.

This paper presents a methodology that can predict the building energy and peak power demand. The dataset considered here is a set of several daily load curves corresponding to the electric power consumption of big building. The methodology analyses the historical load data and get information from them. First the data has to be analysed using cluster method to check how many curves are necessary to represent the building energy uses. After that, the typical load curves are built and models are obtained. Using the models, the accumulated energy in the end of the day and the peak power demand can be predicted with a good precision. The methodology had been used in a big hospital and the results were presented.

2. CLUSTER ANALYSIS

Cluster Analysis is the name of a group of multivariate techniques whose primary purpose is to identify similar entities from the characteristics they possess. The essence of clustering approaches is classification according to natural relationships. As such, the primary value of cluster analysis lies in the pre-classification of data, as suggested by natural grouping of the data itself (Jobson, 1992; Hair Junior, 1987). Cluster analysis has been used here to classify time series data in groups.

The nature of cluster analysis can be illustrated by graphic presentation of a bivariate example (Fig. 1). In problems with more than three dimensions it is impossible to illustrate by graphic presentation. The graphic presents two characteristics of the data. Each point presents one analysed object that has one value for characteristic X and one value for characteristic Y.
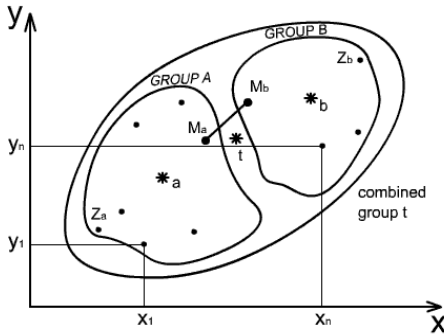
Figure 1: Similarities Measurements.

2.1 Similarity measures

There are many possibilities to measure similarity. One way is to look at the proximity or closeness between each pair of objects in order to determine their similarity. Another way is to look at difference or distance between the pairs of objects, as the distance is the complement of similarity. Distance measures are the most commonly used measures of similarity between objects. These distances can be based on a single dimension or multiple dimensions. The most straightforward way of computing distances between objects in a multi-dimensional space is to compute Euclidean distances. If we had a two- or three-dimensional space this measure is the actual geometric distance between objects in the space (i.e., as if measured with a ruler). In this work it is used a higher dimension (288-dimensional and four-dimensional space).

Euclidean distance

This is probably the most commonly chosen type to calculate distance. It is the geometric distance in the multidimensional space. It is computed as:

$$d_{AB}^2 = \sum_{j=1}^{p}(x_{aj} - x_{bj})^2 \qquad (1)$$

Note that Euclidean distances are usually computed from raw data, and not from standardized data. This method has certain advantages (e.g., the distance between any two objects is not affected by the addition of new objects to the analysis, which may be outliers). However, the distances can be greatly affected by differences in scale among the dimensions from
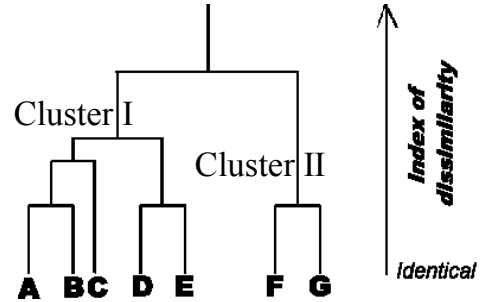

Figure 2: Dendrogram.

which the distances are computed.

2.2 Clustering algorithms

There are many different algorithms to implement in the portioning part of the cluster analysis. Most of the commonly used clustering algorithms can be classified into two general categories: hierarchical and non-hierarchical. Hierarchical procedures involve the construction of a hierarchical or three-like structure. There are basically two types of hierarchical procedures: agglomerative and divisive. In this work it is used the hierarchical agglomerative. In the agglomerative methods each observation starts out as its own cluster. In subsequent steps, the two closest clusters are combined into a new aggregate cluster, thus reducing the number of cluster in one in each step. Dendrogram graphic (Fig. 2) can illustrate this procedure.

2.3 Linkage Rules

At the first step, when each object represents its own cluster, the distances between those objects are defined by the chosen distance measure. However, once several objects have been linked together, a linkage or amalgamation rule is used to determine when two clusters are sufficiently similar to be linked together. It is used here the method called single linkage. It links two clusters together when any two objects in the two clusters are closer together than the respective linkage distance.

2.4 Hierarchical Tree

In a Hierarchical Tree Plot, each object begins in a class by itself (Fig. 2). Note that the clusters are joined (fused) at increasing levels of 'dissimilarity'. In small steps, the threshold regard-

ing the decision when to declare two or more objects to be members of the same cluster is adjusting smaller and new cluster has been creating.

It appears, from this Dendrogram, that two clusters can represent the data (I & II). However, as the number of cases increases it may not be so obvious. Indeed, one of the biggest problems with this Cluster Analysis is identifying the optimum number of clusters. As the fusion process continues increasingly dissimilar clusters must be fused, i.e. the classification becomes increasingly artificial. Deciding upon the optimum number of clusters is largely subjective, although looking at a graph of the level of similarity at fusion versus number of clusters may help. There will be sudden jumps in the level of similarity as dissimilar groups are fused.

3. METHODOLOGY APPLYIED IN A REAL CASE

The methodology proposed here is divided in two parts. The first one is responsible to predict the peak load. The other is responsible to predict the accumulated energy. Both philosophies are the same. Figure 3 shows the structure of the methodology.
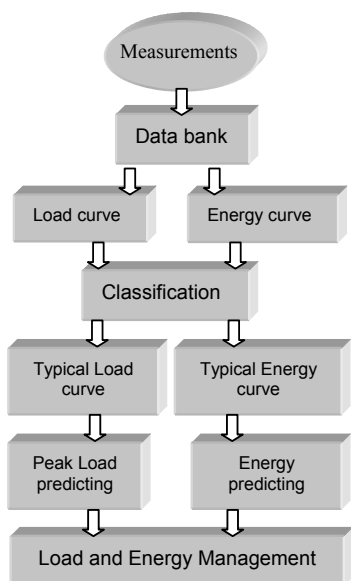

Figure 3: Methodology to predict.

The analysed building is a hospital complex. There is a second biggest hospital in Brazil (Fig. 4).

3.1 Peak Load Predicting

The methodology to predict the peak load is divided in three parts (Silva and Jota, 2004).

- Classification to obtain how many typical load curves is necessary.
- Mathematical model to represent the typical load curve.
- Predict the peak load.

3.1.1 Classification

To obtain the typical load curve of a consumer, it is necessary to analyse its historical load curves. The load curves contain all the historical of consumer energy use. It can be seen how the load grows during the day and how it is used during the week and weekend.

Because of the load curve is a time series, to classify the load curves in shape it is necessary to compare the shape during all day. This comparison can easily be done by a cluster analysis as presented before.

This part of the analysis can be done using the measured power (each 5 minutes), for the historical days. In this case, the load curve is represented in a very good and precise way. The cluster algorithm had the shape and the information of time to analyse and compare (time series) the data.

Figure 5 shows a daily load curve. It is composed of power measured for each five minutes, in kW. Each day is then represented by 288 power values. Each power value has its correlation with a time. Then, the cluster algorithm

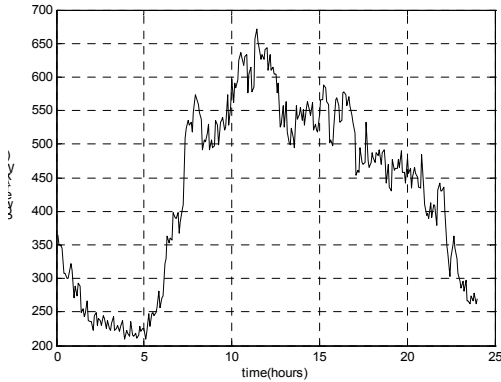Figure 4: Analysed Building.

Figure 4: Load curve.

compares the dayA with the dayB using the Euclidian distance in a 288-dimensional space, (2).

$$d_{AB}^2 = (x_{A1} - x_{B1})^2 \ldots (x_{A288} - x_{B288})^2 \qquad (2)$$

The result of cluster analysis can be represented in a Dendrogram graphic. Although, it is used more than 30 days to classify, the Dendrogram presents only the last classification where the maximum cluster number is 30. Analysing the output of the cluster analysis, it can be seen that the Saturdays are classified on group G1, the Sundays on G2 and the weekdays on G3, as showing in Figure 6. Analysing the error, it is decided that two cluster is sufficient to represent the load curves, than it is considered G12 and G3.

3.1.2 Mathematical load of the typical load curve
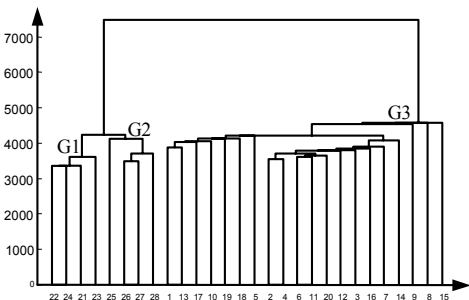
After the number of typical curves has been found, their shape has to be chosen. The power engineers normally build only one typical curve. They take the mean values of the load curves (weekdays curves) and the peak power demand is the maximum measured. In fact, they did not build a load curve but take the more common shape to be the typical one.

It has been proposed here to build a typical load curve in two manners: using the adjusted mean load curve and the adjusted upper quartile load curve. Equation 1 is the mathematical model of a normalised load curve of group G3. It is a polynomial of 9th order and it presents a correlation of 88,33% and a significance of 99,95% (t-test).

$$p = -0,09t^9 - 0,0618t^8 + 0,7269t^7$$
$$+ 0,3302t^6 - 2,0974t^5 - 0,387t^4 \qquad (3)$$
$$+ 2,3664t^3 - 0,3715t^2 - 0,585t + 0,9045$$

This curve is very important to make the contract with the energy company and to management the energy. In this contract the maximum value of the curve is established and the load curve cannot get more than that without a fine. To get the peak value normally it is used the maximum value of the measured curves. Using this curve the energy manager can allocate loads during the day without problems.

3.1.3 Predicting the peak power

Using the measured value of power of a time t_1, $D(t_1)$, it can predicted the peak power demand using Eq. 4

$$D_{max} = \frac{1}{D_n(t_1)}(D(t_1) + D_{min}(D_n(t_1) - 1)) \qquad (4)$$

where:
D_{max}: is the maximum demand (peak power),
D_{min}: is the minimum value of the demand in the day (it occurs between 3 to 6am),
$D_n(t_1)$: is the normalized demand in time $t = t_1$.

The predicting error using this method is approximately 10%. Considering that this consumer has a variation of maximum demand between 400-760kW this error is considered a reasonable.

3.2 Predicting the accumulated day energy

To predict the accumulated energy it is used the same methodology as in predicting peak load.
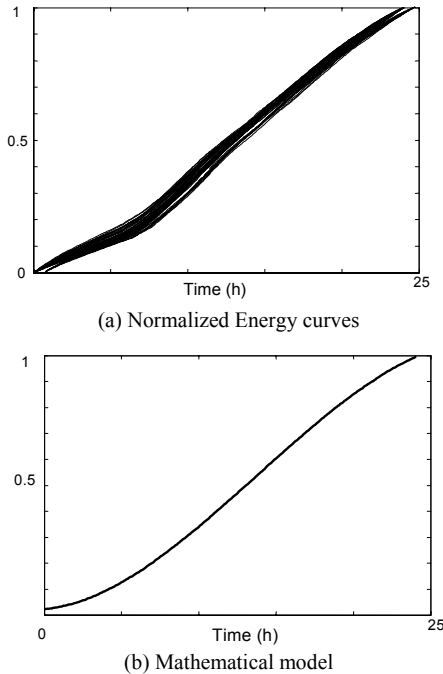

Figure 6: Dendrogram of data.

(a) Normalized Energy curves

(b) Mathematical model

Figure 7: Mathematical model of energy curve.

First of all the classification process of the curves has be done. The classification process indicates that two clusters have to be used. Cluster A is composed of curves of Fridays and Saturdays and cluster B with curves between Sundays to Thursdays (Fig. 7).

The mathematical model was obtained for both curves and is represented of Eq. 5. The coefficients of that equation are presented in Table 1.

$$E_n = a_3 t^3 + a_2 t^2 + a_1 t + a_o \qquad (5)$$

Having the number of typical curves and the mathematical model of each one, it is possible to predict the accumulated energy in the end of the day using a measured accumulated energy between 7 and 8am. This building has a routine that is established until 8am. The energy manager measured an accumulated energy until that time and getting the normalized value of the energy in Eq. 5 for the same time than, using Eq. 6 it can be calculate the total accumulated energy.

$$E_{t(predicted)} = \frac{E_1}{E_n} \qquad (6)$$

Table 1: Coefficients of Eq. 5

Coefficients	a_3	a_2	a_1	a_0
Friday to Saturday	-6,1e-5	2,5e-3	1,5e-2	1,6e-2
Sunday to Thursday	-9,4e-5	3,9e-3	1,2e-3	2,6e-2

4. CONCLUSIONS

The methodology proposed here can make a synthesis of the load curves and calculate the equation of the typical load curve. Using the historical data it is possible to predict the accumulated energy in the end of the day and the peak power load. This estimation can help the energy manager to identifying energy consumption anomalies, managing energy costs, and automating demand response strategies.

The proposed methodology has been applied in a real data of a big consumer where is installed a hospital. The results show that the methodology can predict energy and demand with a very small error. The energy use is not a static function but dynamic. Because of that, the methodology has to use a dynamic historical data that can incorporate the evolution of the load and its use.

REFERENCES

Jobson, J.D., 1992. Applied Multivariate Data Analysis: Categorical and Multivariate Methods. Vol. II. New York: Springer, 731p.

Hair, J.F., Junior, R.E. Anderson and R.L. Tatham, 1987. Multivariate Data Analysis. 2.ed. New York: Macmillan Publishers, 449p.

Michalik, G. and W. Mielczarski, 1998. Modelling Of Energy Use Patterns In The Residential Sector Using Linguistic Variable. IEEE, vol.1, May, pp. 217-220.

Nazarko, J. and Z.A. Styczynski, 1999. Application Of Statistical And Neural Approaches To The Daily Load Profile Modelling In Power Distribution Systems. presented at Transmission and Distribution Conference, vol.1, April, pp. 320-325.

Livik, K., N. Fielberg and J.A. Foosnaes, 1993. Estimation Of Annual Coincident Peak Demand And Load Curves Based On Statistical Analysis And Typical Load Data Electricity Distribution, v.6, May, pp.6.20/1-6.20/6.

Murata, J., K. Kimura, S. Sagara and T. Kamakura, 1991.

Daily load curve forecasting via automatic selection of models, Power System Monitoring and Control, June, p.270-272.

Silva, V.R.B. and P.R.S. Jota, 2004. Predicting Load using Cluster Analysis, WSEAS Transactions on Circuit and Systems, Issue 8, volume 3, October, pp. 1706-1711.