DEMAND AND ENERGY FORECAST: INVESTIGATION OF DRIVING FACTORS

Jorge YASUOKA[1], Reynaldo A. SOUZA Jr. [1], José A. JARDINI[1], Roberto CASTRO[2], Liliana M. V. BRITTES[2], André L. P. CRUZ[2], Hernán P. SCHMIDT[1]
[1]Escola Politécnica da Universidade de São Paulo and [2]CPFL – Companhia Paulista de Força e Luz - Brazil
jyasuoka@pea.usp.br

INTRODUCTION

Traditional methods for estimating future values of demand and energy do not normally take into account the effect of the so-called explaining variables, which include load geographical location, seasonal and climatic variations, availability restrictions of energy and Daylight Saving Time. This work proposes a methodology for assessing the impact of these driving factors on the estimation of future values of demand and energy, with a view to improving current practices which dictate demand and energy purchases (bid-offer area) in both the short term and the medium term.

THE LOAD FORECASTING PROBLEM

In general, the short and medium term forecast problem has been approached according to two traditional methodologies: temporary series and regression analysis. These conventional computational techniques may not present results sufficiently precise. In a similar way, computational methods based on complex algorithms demand a heavy processing task, and thus, they can converge slowly and in certain cases even diverge.

The approach through temporary series normally does not take into account the climatic information (ambient temperature, humidity, insolation, etc.) that constitutes an important parameter, especially when referring to residential loads. Consequently, problems resulting from little precision and numeric instability of the results are some serious problems observed in this technique. The objective of this model is to recognize a pattern inside the values of historical series and to extrapolate this in the future. On the other hand, the regression analysis tries to associate the climatic variables with the system load. In this case, a fixed functional relationship among these variables has to be established beforehand, and usually a linear model is adopted. The objective of this model is to discover the form of this relationship and to use it in the forecast of the dependent variables utilizing the future evolution of the independent variables (e.g.: simple and multiple regression). However, as the functional relationship itself varies through time, this approach frequently produces unsatisfactory results.

To bypass such obstacles, this article presents a work whose objective is to study the influence of the driving factors in the aspect of the demand and energy curve, and the use of Artificial Neural Networks (ANNs) technique to improve the forecast results

ANALYSIS OF DEMAND CURVES

For the studies, demand data extracted from CPFL (Companhia Paulista de Força e Luz, a Brazilian distribution utility) historical data base were used, including a period from 1997 to mid 2003. Special days, such as holidays, were not considered. Some of the studies made and the obtained results are described as follows.

Clusterization

To improve the results of the demand and energy forecast, analyses were made grouping the substations in families according to the shape of the demand curve. Because the system of this utility is comprised of approximately 300 substations, typical demand curves from each substation were compared with the intention of obtaining some typical curves that would be used as a basis for the load forecast for all other substations. By seeking a leveling in the amplitude of the demand curve, curves in 'pu' were used in the analyses in relation to the daily average energy; in other words, the ratio of demand data (kW) to the average daily power (daily energy/24h).

Firstly, 75 substations were tested, whose load data were inserted in a clusterization software (SAS), that groups the load curves in different families according to statistical models. The results demonstrate that it is possible to identify typical curves for each family (that resemble to the average curve of each group). Ten different clusters were obtained and one of them is shown in figure 1.

In the analysis of the clusters, the discrimination of substations with profiles predominantly residential, commercial and industrial was verified. It was noticed that the substations of the same cluster were not necessarily located geographically close to each other.

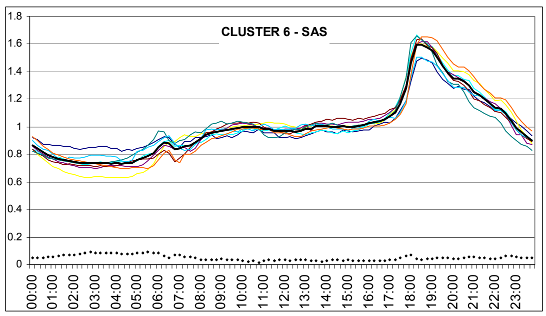

Fig. 1. The enhanced curve is the cluster curves average and the dotted curve consists of the mean deviation of the same.

Normalization (pu)

The load shape of each day of the week is different, however it was noticed that when the curves are normalized (in relation to the average daily energy), the curves of Tuesday, Wednesday and Thursday have the same shape and can be grouped. Thus, clustering and normalizing the curves are crucial to reduce the number of ANNs necessary to realize the forecasting. This process is essential because there are more than 300 substations in the CPFL's network.

Pre and post-rationing period comparison (2001)

As it is known, the rationing of energy which occurred in Brazil, in 2001, promoted several changes in the energy sector, mainly in the value and in the form of the demand curve. The inherent alterations of the rationing were studied, com-paring previous and subsequent data to 2001, based on the load data of the main substations of the CPFL electrical sys-tem. The results show that the tendency was that the load shape maintained in most of the analyzed places, signaling that the data previous to 2001 can be used, provided that only the shape of the curve is considered; in other words, analyzing the curves in pu, it is noticed that the shape of the curve was not altered even after the rationing period. One of the results is shown in figure 2. From April/1997 to July/2003 the Mondays average of each month was made.

Daylight Saving Time (DST)

The demand and energy data in periods in which the Daylight Saving Time (DST) is in force are different in relation to the other periods of the year. To verify this, analyses of the curves of main CPFL's substations were made, aiming at this comparison, it was verified that it is necessary to make the forecast for the days of DST period separately. Hence it was verified that there is great similarity in the aspect of the demand curves of the substations out of the peak time, and that the differences lie basically in the peak time, mainly in substations with predominance of industrial/commercial profile. Figure 3 shows the case of a substation located in the state of São Paulo.

Climatic Variables

The relationship between the electric charge and its driving factors is complex and non-linear, hindering its modelling through linear mappings. Not only does it lack the necessary precision, but also the traditional techniques are not sufficiently robust. ANNs can represent patterns that are not detectable by traditional statistical methods that utilize auto-correlation coefficients or crossed correlation.

Besides the data of the load themselves, the inclusion of ambient temperature, relative air humidity and insolation (incident radiation) was analyzed as causal variables for the load forecasts.

It is noticed intuitively that these variables influence the consumption of energy, although the mechanisms of such impact are unknown. A pre-analysis was made only to detect possible correlations between the load (kWh) and those driving factors. If, for a variable, this qualitative analysis had revealed some level of correlation (linear or not) with the load, it would mean that this variable could be considered to aid the forecast. The climatic data, available every 10 minutes, were supplied by an automatic collection station of climatic data located in the city of Campinas, São Paulo.

The first variable analyzed was the average temperature T_{AVG} defined as:

$$T_{AVG} = \frac{T_{MAX} + T_{MIN}}{2} \tag{1}$$

being T_{MAX} and T_{MIN} the maximum and the minimum daily temperature, respectively. Through an analysis of simple linear regression a high correlation level was verified between this variable and the daily energy of a certain place, especially in the working days of the week. Thus, it was defined that the average temperature would be taken into account in the load forecast.

Figure 4 shows the dispersion diagram and the calculated regression line, along with the determination coefficient for a Friday.

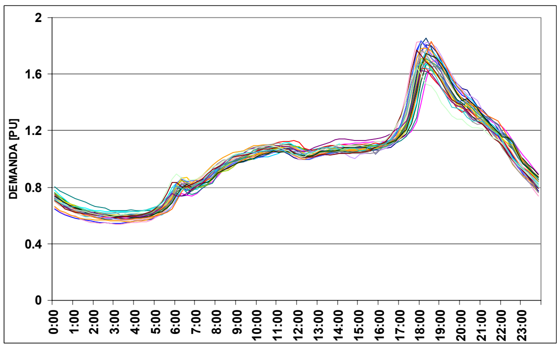

Fig. 2. Demand curve in pu of a substation located in the central region of the state of São Paulo.

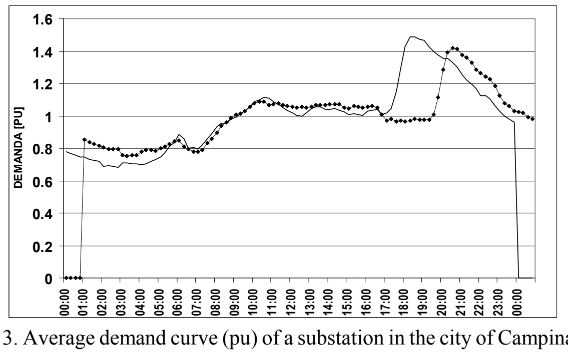

Fig. 3. Average demand curve (pu) of a substation in the city of Campinas in the years of 2002 and 2003, highlighting the differentiation in the consumption in the Daylight Saving Time (DST) (dotted curve). A predominant residential profile is noticed.

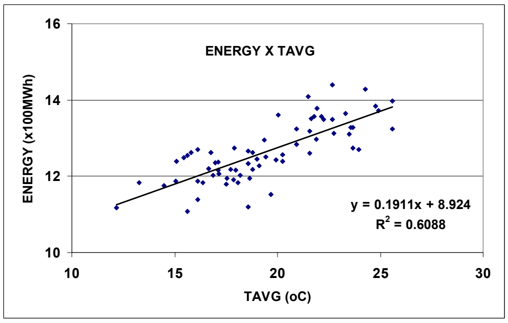

Fig. 4. Adjusted straight line for energy in relation to the temperature. By the determination coefficient calculated in the linear modeling, it is noticed that approximately 60% of the energy variance is explained by the variation of the average temperature.

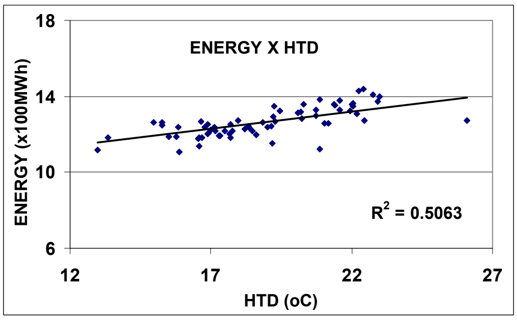

Fig. 5. Line adjusted for energy in function of daily HTD for Fridays. The determination coefficient calculated for this case was smaller. The HTD was calculated in function of the minimum daily relative humidity and T_{AVG}.

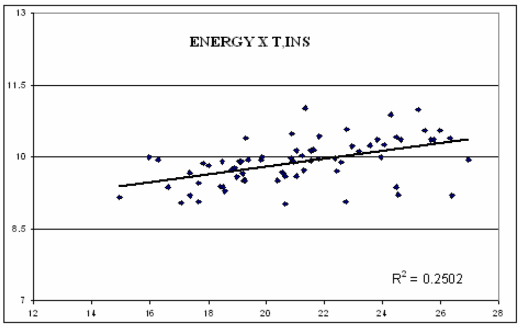

Fig. 6. Line adjusted for Saturdays of the daily Energy in relation to the daily T_{AVG} and daily average of incident radiation. This variable has slightly higher influence on the weekends.

The other studied variable was the relative air humidity. Only this item presents very low correlation with the consumed load, as revealed by several types of regression curves adjustments. However, when associated to the temperature, there is the effect of thermal comfort, which occurs because of the air saturation through humidity, that associated to thermal effect, hinders the exchange of heat between the human body and the atmosphere. This effect occurs through the Index of Humidity Temperature Discomfort (HTD)[1], that relates ambient temperature to the relative air humidity according to the equation:

$$HTD = T - 0{,}55.(1-H).(T-14) \qquad (2)$$

where T is the ambient temperature in Celsius (°C) and H is the relative air humidity in percentage (%).
Therefore, it is possible to check the effect of the temperature and relative humidity to, if proved its influence on the consumption of energy, consider it as input in the load forecast. The same data that served for analysis of the temperature were also used for analysis of the influence of HTD on the load behavior, as shown in figure 5.

In general, there was a good linear correlation between HTD and the daily energy, although smaller in the case of T_{AVG}. For this variable, the average of the determination coefficients obtained for the working days was 0.5554, and in the weekends it was 0.2053. Now in the case of HTD, the average of the working days was 0.5093 and in the weekends it was 0.1936. From these results, it was concluded that the relative air humidity does not significantly add information to the temperature influence. Therefore, the relative air humidity will not be taken into account in the load forecast.

Finally, an evaluation of the influence of direct insolation in the consumption of energy was made. The variable INS, which consists of the daily average of the incident radiation measured in a certain place, was defined. Preliminary tests showed that this variable has irrelevant impact on the consumption of energy. Statistical analyses based on multiple regressions were carried out, considering not only the INS variable, but also the humidity temperature index. One of the dispersion diagrams obtained can be seen in figure 6.

Comparing with the previous case, it was verified that there is practically no improvement with the insertion of the INS variable, either on the weekends or the working days. Therefore, it was decided that this last item will not be considered as a causal variable in the load forecast. Consequently, the only climatic variable to be considered for forecast was the ambient temperature. However, other analyses showed that in hotter and not rainy days the insolation and relative humidity, added to the temperature, presented good correlation with the load, even higher than the cases obtained with T_{AVG} only. This shows that a more complex relation can exist between these two variables.

DEMAND AND ENERGY FORECAST

The type and horizon of forecast covered in this study are:

- The daily demand curve forecast represented by 288 values (5 minutes demand values) and 96 values (15 minutes demand values) for 1 week ahead;
- the daily demand curve forecast represented by 96 values (15 minutes demand values) for 1 month ahead;
- the energy forecast of 1 week, 1 month and 1 year ahead.

This forecast is made at the distribution substation level, and the estimated global curve is obtained by aggregating curves from all substations (more than 300). The input and output variables considered in the ANN structure for the forecasts can be seen in figure 7.

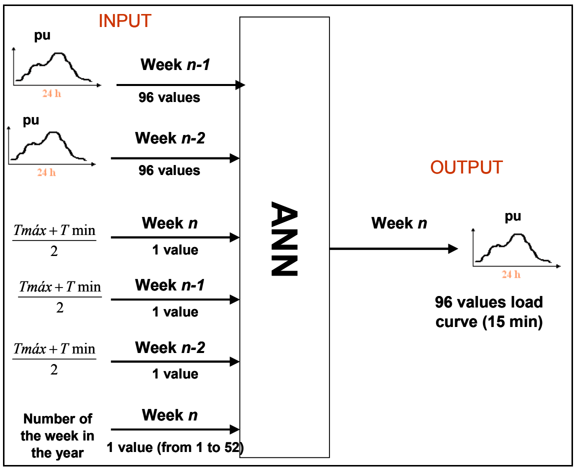

Fig. 7. Topology considered in the forecasting.

For example, to process the 1 week ahead forecast (load curve of week *n*), the input variables are:

- the load curve (96 values) of 1 week ago (week *n-1*);
- the load curve (96 values) of 2 weeks ago (week *n-2*);
- the forecasted average ambient temperature of week *n;*
- the average ambient temperature of week *n-1*;
- the average ambient temperature of week *n-2*;
- and optionally, the number of the week in the year, to represent the seasonal information.

The same ideia is used to process the 1 month ahead forecast. Instead of using the previous 2 weeks datas, the inputs for this case are the previous 2 months datas.

The general procedure for forecasting the future load is described as follows:

Step 1) Gather the normalized load curves according to its pattern similarity of substation load shape (clustering);
Step 2) process the demand forecast (pu) for the desired substation and time interval;
Step 3) process the energy forecast (MWh) to the same location and time interval;
Step 4) recompose the original load curve (MW) multiplying the load curve (pu) by its forecasted value of energy.

An example of forecasted load curve can be seen in figure 8. The forecasts results obtained are quite satisfactory, although there is room for further improvement. The Mean Absolute Percentage Error (MAPE) is about 3%.

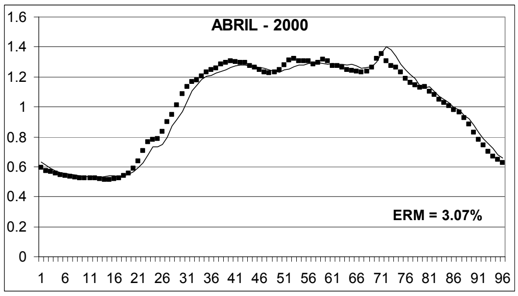

Fig. 8. Forecast accomplished for the last Monday of the month of April 2000 for one substation. The dotted curve is the one calculated and the continuous curve is the measured one. The MAPE is 3.07%.

CONCLUSION

The article presented a work that has the purpose of developing a methodology for demand and energy forecast for more accurate processes than the current models based on typical variables and forecast techniques, with application in the global forecast of the load in the outlook of short and medium terms, directed to the bid-offer operations of energy. The load forecast is of extreme importance in this area, and an improvement in the results implicates in expressive gains for the Electric Utility.
The analyses of the driving factors prior to the execution of the forecast, contribute greatly in the results of the demand and energy forecasts and the application of more robust techniques (such as ANNs), in substitution to the traditional techniques based on statistical methods allows to treat variables that these last ones do not, such as the climatic variables.

REFERENCES

[1] WALLACE, J.M. e P.V. HOBBS, Atmospheric Science: an introductory survey. Academic Press, 1977.
[2] LUTGENS, F.K. e E.J. TARBUCK, The Atmosphere: an introduction to Metorology. Prentice Hall, 1989.