Soft Computing Techniques for Short Term Load Forecasting

Carolina Tranchita, Member IEEE , Álvaro Torres, Señor Member IEEE
Department of Electrical Engineering, Universidad de los Andes – Bogotá, Colombia

Abstract-- **Short term load forecasting is a recurrent topic in the operative planning activities of companies dedicated to the distribution and trade of energy around the world due to the competitive electricity market, in which an advantage in the previous knowledge of demand could mean the difference between obtaining big benefits or incur in economic losses.**
In this paper a novel method for short term load forecasting is proposed based on the similar day approach and the use of Soft Computing techniques. This approach is founded on the search for the most similar day in history, to the forecasted day, based on the explanatory meteorological variables forecast for the load of this day. Once the "similar" day is found, the load forecast will be the same of that day with an adjustment for load growth. LAMDA-fuzzy-clustering techniques, regression trees, CART classification and fuzzy inference for the peak power, daily energy and load curve forecast are used.
The validation of the proposed method is made with meteorological and load data from a Colombian city.

Keywords—**Short term load forecast, Peak Power and Daily Energy Forecast, Similar Day Approach, LAMDA, CART.**

I. INTRODUCTION

The art of load forecasting, particularly short term, became one of the top growing areas of study and significance since the emergence of competition in the energy markets. After the privatization and deregulation tendency of power systems in many countries, including Colombia among them, the purchase and sale of energy in this new market constitute the base of business for the distribution and electricity trade companies.

For these companies it is vital to have good knowledge, or to obtain a good forecast of the short term load with the purpose of fixing prices in advance, according to their service supply costs. This contributes to these companies savings, for example in Great Britain, the savings were in the order of US $30 millions for a 4% improvement in load forecasting [1].

Diverse methods for short term load forecasting have been tested with different degrees of success. There are two main tendencies of models for short term forecast: the causal or conventional models, which outline the relationships among load and factors influencing it, mainly social and weather variables; and the time series models which outline the relationships with values observed in the past.

The model presented in this paper is based in the similar day approach and a forecast for the load curve, the peak power and the daily energy consumed is achieved.

The "similar day" approach has been used in forecasting applications where the dependent variable is highly correlated with the meteorological variables and the calendar location. The method proposes to search in history a very similar day based in some known explaining independent variables, generally of climate and time (day of the week, month, etc.) and to predict the dependent variable, in this case the electric energy demand, in accordance to what had happened to this variable in the similar day found. It is important to note that in this method there are no mathematical functions found for the explaining variables.

The hourly electric load in a Colombian coast located company and the hourly temperature of the corresponding city are used for the research development. The data used for methodology validation go from January 1998 to June 2000.

II. STATE OF THE ART

Conventional methods in short-term load forecasting are among others: regression models, ARIMA, ARIMAX or SARIMAX models [2], Box and Jenkins transfer functions [3], optimization techniques [4], non-parametric regression, structural models, curve adjustment procedures and there are some more recent methods whose results are compared with the ones mentioned such as expert systems [5], artificial neural networks [6], fuzzy inference [7] and Neuro-fuzzy models [8].

The load to forecast in the system is a random non-stationary process composed by thousands of individual components. Therefore, the range of possible approaches to the forecast is wide. Usually the only possibility is to take a macroscopic view of the problem, and try to model the future load as a reflection of previous behavior. This still leaves open space to very different solutions. Due to the nature of load, the only objective method to evaluate the approaches is experimental evidence [9].

Expert systems have been also used in load forecasting. These systems are heuristic models, which usually can include both quantitative and qualitative factors. A typical approach is trying to imitate the reasoning of a human operator. The idea then is to reduce the analogical thought behind the intuitive forecasting to formal logic steps. A popular approach has been to develop rules based on *fuzzy logic* [9].

Another important technique in short term load forecasting is the neural network. In these models, the network inputs generally are present and past load values and the outputs are future load values. The network is trained using real load data from the past [10].

There have also been suggested models with no-supervised learning for load forecasting. The purpose of these models can be the classification of days in different types of day, or to choose the most appropriate days in history to be used as base for the real load forecasting.

III. MODEL PROPOSED

In the method proposed in this paper peak power and daily energy forecasts are made in order to predict the load curve, looking for the most similar day in history based on the meteorological forecasts for tomorrow's day. It is achieved using classification and identification with LAMDA and CART methodologies, obtaining and implementing a rule based Fuzzy Inference System and the similar day approach.

A. *STAGE 0: Data Analysis*

The data used for methodology validation go from January of 1998 to June of 2000. For a more complete proof, load data of a longer period of time would be preferred.

Some cyclical effects for the load data, see Figure 1, can be observed starting from the working day; the weekend rhythm is followed by most people. In labor days social activities are in a higher level than in Saturdays and Sundays, therefore the load is also higher. The series begins with five quite similar patterns that are the Monday to Friday load curves. Next two different patterns continue for Saturday and Sunday. Then this same weekly pattern is repeated [10].

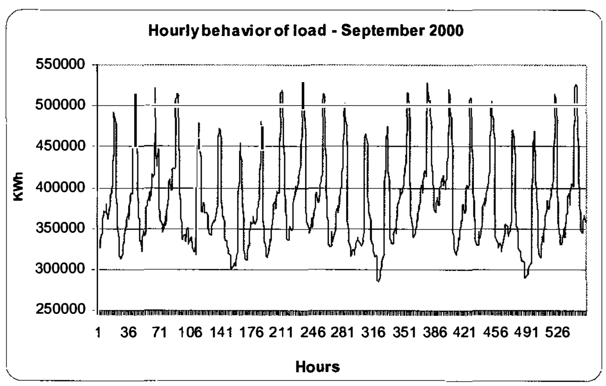

Fig.1. Monthly demand of energy during September 2000.

On the other hand the daily rhythm takes place because of the synchronous behavior of people during the day. Most of people sleep at night; consequently the demand is low during night hours. Also during the day, many activities tend to be simultaneous for most of people (labor schedule, lunch hour, etc.). The daily rhythm changes throughout the year [10].

There are also differences between days of the same month or meteorological condition. Therefore, in load forecasting days are often divided in numerous *types of day*, each has its own characteristic load patterns. Saturdays and Sundays have different load curves than other days. Often Mondays and/or Fridays also separate from other labor days, because the proximity of weekend can have a slight effect in load [10].

In the proposed model the explanatory variables used for the forecast were: the day of the week, the type of day (classified as labor, weekends and holidays) and the meteorological conditions of the day to which the load is to be predicted.

B. *STAGE I: LAMDA to determine the similar day*

The similar day technique is founded on the search for the most similar day in history to tomorrow, or to the day to predict, based on the explanatory meteorological variables for the electric power consumption.

The objective is that with the meteorological forecasting for tomorrow, or for the day to predict, the most similar day in history can be found. Some software programs, available in the market, look for the most similar day using clustering methods with neural networks.

In the work developed the LAMDA (*Learning Algorithm Multivariate Data Analysis*) classification method is proposed to be used with this aim. LAMDA was chosen because it shows some advantages that other classification systems don't have. One of these is determining the number of clusters, without being the user who decides the input space partition, and this is not excessive. Another interesting feature is that it can be chosen whether the supervised or the not supervised classification is used [11].

LAMDA is a fuzzy classification method that is rooted in the adequacy degree analysis of each object to a class, concept related with the membership degree to a fuzzy set. The contribution of each one of the attributes or individual's describers, to the membership degree to a class is called marginal adequacy (MAD, Marginal Adequacy Degree). These marginal adequacies are combined with the use of "Fuzzy Mixed" operators to finally find the adequacy degree of the individual to each one of the classes (GAD, Global Adequacy Degree) [11].

As no pattern was known to determine similar days it was necessary to carry out first not supervised classification, so that later supervised classification could be done in such a way that when entering a new day (individual), the algorithm was able to identify at least one (1) most similar day.

The "learning" consists on finding parameters defining each one of the classes' describers. The type of parameters and the way to find them depends on whether the work is done with quantitative or qualitative describers, and the means to determine them depends on whether a supervised learning is being carried out or not [11].

A non information class exists (NIC=Non Informative Class), which determines the minimum threshold of global adequacy that an individual should have in order to belong to some class. In the case of supervised learning, the individuals that don't overcome the GAD of the NIC class are said not to belong to any of the classes. For not supervised learning, starting from the NIC parameters and the individual that was not classified, a new class is generated. So starting from only one group, all the individuals are finally classified locating them in groups that are generated depending on the GAD´s.

For the similar day identification, the IDEAM (Colombian institution for meteorology) made possible the access to their meteorological database. Basically data of years 1998, 1999 and 2000 were obtained. However there were problems with the database as it was not fully completed and the variables measures are not hourly taken. Therefore the variables used for obtaining the similar day are the following: average temperature and average humidity in the morning, afternoon, and night of the next and current day; solar shine in the morning and afternoon of the current day; average speed of the winds in the morning, afternoon, and night of the current day; maximum and minimum speeds of the winds in the current day and maximum and minimum temperature in the next and current day.

Once the variables were chosen, tests on the not supervised classification were made with the software elaborated by the control group of the Universidad de Los Andes. Several simulations with different demand values and distributions were made using 900 of the base of 1088 data for training, because in this type of problem there is one seasonal pattern per year and it was important to try to embrace all of the possible cases.

After the tests it was determined that the best distribution for classifying data was the normal when the demand parameter is 0.39 since the number of classes is not excessive and the classification is more homogeneous that in other cases. The result of not supervised classification is shown in Figure 2.

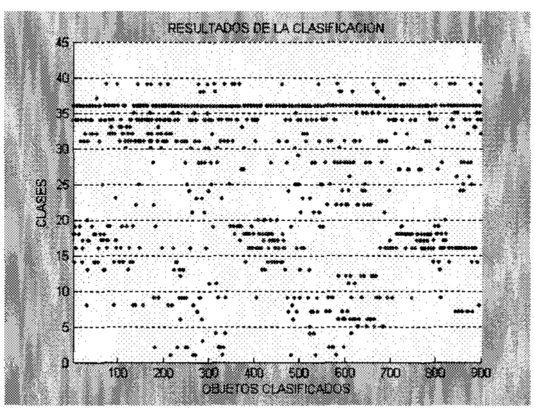

Fig. 2. Classification of the days according to meteorological conditions.

C. *STAGE II: Forecast for peak power and daily energy*

1) Part a: Search of relevant variables and obtaining of if-then Rules using CART

Once made the daily load curve analysis it was decided to forecast the peak power and the daily energy, other important data such as the load factor and the minimum power were not forecasted since they are quite constant with the test data.

CART is a classification and regression trees algorithm. It makes supervised classification and it can find the decision rules of the cluster and in the application for load forecasting it was used for this purpose only after the classification with LAMDA was completed.

As the clusters had already been determined, it was necessary to determine whether all the variables were relevant, or if it was better to add other non meteorological variables in order to obtain better functions of Takagi Sugeno type for the defuzzyfication. This way, several models were suggested depending on meteorological variables, the day of the week, the type of day and the variable to predict in previous times.

With CART it was found that the meteorological variables, the type of day and the day of the week were enough to make the peak power and energy forecasts. When autoregressive models were used, that is, the output in previous times, the error obtained was similar to the model that only uses meteorological data but the more autoregressive the output was the less necessary the meteorological variables were.

Once defined the clusters and variables, if-then rules were also found for the fuzzy inference system using CART. However CART only provides Takagi – Sugeno rules of zero order, which was not the optimum for the problem. This was solved using least minimum squared on each cluster as later explained.

As example, in the decision tree obtained for peak power one of the rules obtained is:

*If the day type is weekend or holiday **and** the average speed of winds of the current day in the afternoon is more than 4.45 **and** the humidity of the current day is less than or equal to 94.5 **and** the temperature of the next day in the night is less than or equal to 26.95 **then** the peak power of the next day is of 436873.*

2) Part b: Takagi-Sugeno (TS) Rule Adjustment using least minimum squared regressions

With the CART algorithm the classification rules for each cluster were found but of zero order (a constant value), therefore it was necessary to make an adjustment to that regression. We proceeded to make the regressions using least minimum squared.

There were found excellent adjustments in several clusters that exceeded all the statistical tests, this means, the t test that indicates the adjustment tolerance of each variable, the F test that indicates the global adjustment tolerance of the model and the value of R squared that indicates how much can the model forecast. However, there were also found clusters where the value of R squared did not exceed 0.6 (fairly good) and the value of F was barely more than the critical F, but in spite of this in the great majority of cases the typical error of adjustment was less than in the case where the rule was left with a constant value. In the cases where the defuzzyfication of the rule is constant, it is because the quantity of cluster data was too small as to use least minimum squared.

Also in some of the clusters it was found that a fairly good model with the explanatory variables did not exist, so that a division of the cluster had to be made depending on the day type or the day of the week, for example, the cluster 23 had the following rule:

*If the day of the week is Monday or Tuesday or Wednesday or Thursday or Friday **and** the average speed of the winds of the current day is more than to 3.75 and the temperature in the night of the next day is less than or equal to 27.15 **and.........** then the daily energy of the next day is 8.46487E+006.*

As can be seen, there was no difference whether the day was Monday, Tuesday, Wednesday, Thursday or Friday. When a least minimum squared was intended, none of the statistical tests was satisfactory; therefore more explanatory variables were used to improve the model adjustment. For this purpose binary variables or dummies were manipulated to represent the day of the week because it was known that this variable is relevant for the load forecasting. In this case there was better adjustment making these distinctions although it incurred in a greater division of the input space.

For example node 23 changes from 1 to 4 rules, as follows:

- ***Node 23 a:*** *If the day of the week is Monday and the average speed of the winds of the current day is more than 3.75 and.......*
- ***Node 23 b:*** *If the day of the week is Tuesday and.....*
- ***Node 23 c:*** *If the day of the week is Wednesday and....*
- ***Node 23 d:*** *If the day of the week is Friday and the average speed of the winds of the current day is more than 3.75 and......*

In total there were about 350 tests made to find 87 functions for the consequents of the TS rules.

3) Part c: Implementation of the Fuzzy Inference System

Once the Takagi – Sugeno rules and the number of membership functions were obtained the following step was the implementation of the FIS. To create the fuzzy inference system the "fuzzy" function of the Fuzzy Logic toolbox in MATLAB 6.1 was used.

The system has in total 25 inputs, for the 7 days of the week, the three day types (labour, holiday or weekend), and the 15 meteorological variables. Since the day of the week and the type of day are concrete values the membership function that better fits is the singleton, not yet implemented in the toolbox, therefore a very narrow triangular function was used.

Bell functions were used to implement meteorological variables and function parameters were estimated using least minimum squared estimators. With the two FIS, for peak power and energy, the forecast of these variables was made.

D. STAGE III: Integration of the fuzzy clustering and FIS

In Stage I, with the use of LAMDA meteorologically similar days are found and in Stage II forecasts for peak power and daily energy peak are made. Now the user will have to choose which the most similar day in history is and make the growth adjustments to be able to forecast the load curve of the day to forecast following the flux diagram in Figure 3.

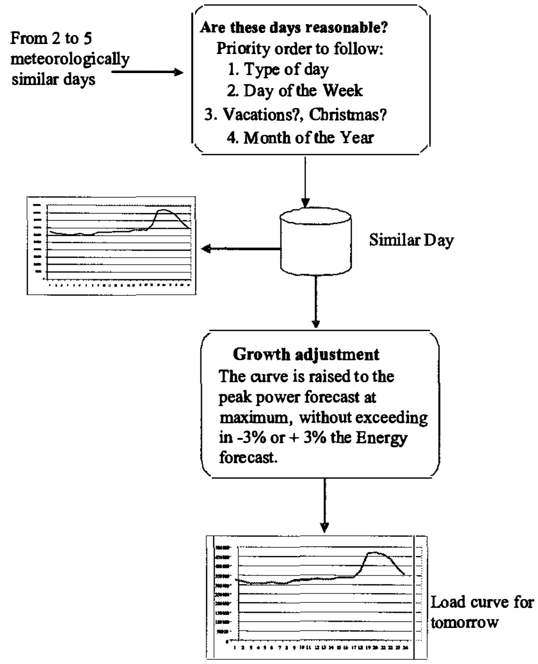

Fig.3. Decision diagram for load curve forecast.

TABLE I
TEST CASE: SIMILAR DAYS

	TOMORROW'S FORECAST	SIMILAR DAYS FOUND					
DAY	18	3	11	23	14		
MONTH	10	8	5	11	10		
YEAR	2000	2000	2000	1998	1998		
DiaSem	3.0	5	5	2	4		
TipDia	1.0	1	1	1	1	Desv Est	Media
Temasig	25.4	25.8	24.7	24.4	25.9	0.76	25.20
Tetasig	29.8	29.8	30.2	30.6	30.2	0.33	30.20
Tenosig	28.0	27.6	27.4	28.6	28.2	0.55	27.95
Temaact	25.8	25.9	26.3	25.2	25.1	0.57	25.63
Tetaact	29.8	32.4	31.8	26.5	27.0	3.11	29.43
Tenoact	26.6	27.5	27.1	25.9	26.0	0.80	26.63
Humasig	95.0	94.0	94.0	99.0	97.0	2.45	96.00
Hutasig	83.0	83.0	82.0	80.0	86.0	2.50	82.75
Hunosig	93.0	89.0	93.0	92.0	98.0	3.74	93.00
Humact	96.0	94.0	94.0	99.0	98.0	2.63	96.25
Hutaact	88.0	88.0	84.0	99.0	90.0	6.34	90.25
Hunoact	94.0	93.0	93.0	98.0	94.0	2.38	94.50
Bractman	6.3	3.5	4.9	5.0	4.2	0.70	4.40
Bracttar	1.8	1.0	3.4	3.3	2.5	1.11	2.55
Veactmax	8.0	8.4	8.1	6.0	6.9	1.11	7.35
Veactmin	0.0	0.3	0.0	0.6	0.6	0.29	0.38
Vepactman	2.0	3.7	1.3	1.4	1.5	1.14	1.96
Vepacttar	5.1	6.6	5.3	2.5	2.4	2.08	4.19
Vepactnoc	4.0	3.8	4.0	4.2	2.9	0.56	3.69
Temaxsig	31.5	31.8	31.4	31.2	31.1	0.31	31.38
Teminsig	25.2	25.0	24.7	25.2	24.7	0.24	24.90
Temaxact	33.4	32.9	34.4	29.1	28.3	2.94	31.18
Teminact	24.1	25.1	23.3	23.2	24.0	0.88	23.90
CLASE	5	5.0	5.0	5.0	5.0		
GADMAX	0.82342	0.7138	0.7015	0.6530	0.6526		

Table I shows the result of the meteorologically most similar days to a test day, on this case Wednesday October 18 of 2000. It was found according to the GAD of the data in the class that the most similar days are: Friday August 3 of 2000, Friday May 5 of 2000, Tuesday November 23 of 1998 and Thursday October 14 of 1998. We observe that the standard deviation in most of the variables for the similar days is low.

Figure 4 shows one variable of those with which the search of the similar day was made. As it can be seen these variables are quantitatively very near. Figure 5 shows the load curves of the similar days obtained based only on the meteorological variables. These load curves are similar evidencing that the meteorological variables have explanation power for the electric power consumption.

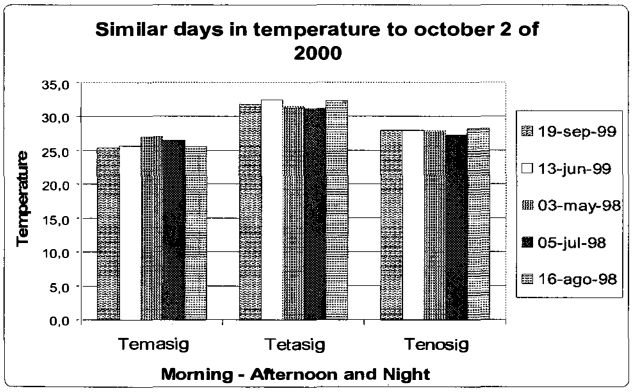

Fig.4. Temperatures of similar days to October 2 according to the GAD.

Once the similar days were identified the meteorological data, the day type and the day of the week were entered to the fuzzy inference system for the short term load forecast. Table II shows the result when the Takagi – Sugeno rules consequent of first order were used.

However there is still error due to the small database, since it has not been possible to characterize all the different types of days. One can observe that the days December 25 are non-typical days that could not be well characterized as consequence of few learning data.

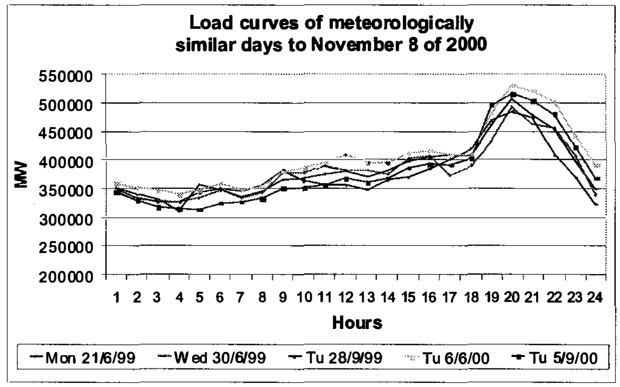

Fig.5. Load curves of the meteorologically similar days to Nov/ 8/2000.

Later, to find the demand curve, the most similar day is chosen between the similar days found by fuzzy clustering, for this it should be kept in mind the day of the week, the day type and the calendar date, for example, if it is Christmas day, it has priority a similar day of one earlier Christmas that other days of the year.

Making the forecast of load curve for example for Thursday November 24 of 2000, the most similar days found were: Tuesday June 2 of 1998, Friday June 5 of 1998, Thursday July 9 of 1998, Tuesday July 27 of 1999 and Wednesday November 10 of 1999. Having as a priority the similarity of the day of the week, it is chosen as similar day the day Thursday July 9 of 1998, load data can be seen in Table III.

Keeping in mind the criteria of moving up the curve to the peak power, without passing above 3% of the daily energy forecasted, the curve goes up in a delta calculated as follows:

$$\Delta = \frac{Ppeak\_forecasted - Ppeak\_of\_similar\_day}{2}$$

The result of the adjusted curve is shown in Figure 6. For this day it can be observed that the error in the peak power forecast was 0.55% and in the energy demand was 2.74%. When the similar day approach is used for the demand curve forecast it has an error of 2.96% on average, and the form of the curve is similar.

TABLE II

FORECAST FOR POWER PEAK USING THE TAKAGI – SUGENO RULES CONSEQUENT FUNCTIONS OF FIRST ORDER

				Real Value		Forecast Value			
				Ppeak	Energy	PPeakDay	Energy	Error Ppeak	Error Energy
Sunday	1	10	2000	489001	8652600	477568	8584621	2338%	0.786%
Wednesday	18	10	2000	509341	9711360	506233	9685412	0.610%	0.267%
Monday	30	10	2000	523222	9494236	525333	9547410	0.403%	0.560%
Sunday	5	11	2000	422995	7583608	433889	7935265	2575%	4.637%
Wednesday	8	11	2000	502626	9367717	501353	9404157	0.253%	0.389%
Wednesday	23	11	2000	501449	8910666	497966	9023493	0.694%	1.266%
Thursday	24	11	2000	520910	9331410	523782	9587424	0.551%	2.744%
Saturday	2	12	2000	462336	8559706	434077	8262351	6112%	3.447%
Sunday	3	12	2000	475611	8288801	487116	8447687	2419%	1.917%
Thursday	7	12	2000	533879	9564478	504708	9395214	5464%	1.77%
Tuesday	19	12	2000	536531	9330476	528959	9295416	1411%	0.376%
Monday	25	12	2000	447482	8143619	469725	8375258	4971%	2.844%

TABLE III

RESULTS

Hour	Similar Day	Forecasted Day	Real Curve	Error
1	342858.70	353373.10	339131.09	3.829%
2	331094.36	341608.76	331291.49	2.869%
3	330674.71	341189.11	324908.11	4.533%
4	327033.03	337547.43	321075.04	4.636%
5	335020.55	345534.95	328428.48	4.703%
6	347062.28	357576.68	359440.10	0.495%
7	333768.72	344283.12	345302.55	0.281%
8	343406.38	353920.77	350573.28	0.899%
9	378386.44	388900.84	390082.03	0.289%
10	377155.95	387670.35	375215.54	3.052%
11	389574.66	400089.06	390674.50	2.235%
12	382696.71	393211.11	397602.64	1.061%
13	382035.24	392549.64	376290.40	3.935%
14	376188.63	386703.03	386828.50	0.031%
15	402149.84	412664.24	392172.66	4.717%
16	407662.15	418176.55	392519.53	5.829%
17	374097.51	384611.91	393839.14	2.279%
18	390065.43	400579.83	405623.15	1.196%
19	432805.41	443319.81	453581.47	2.199%
20	492238.80	502753.20	520909.72	3.431%
21	463944.64	474459.04	503761.02	5.867%
22	455722.40	466236.80	481282.61	3.066%
23	407832.86	418347.26	426340.30	1.815%
24	340376.38	350890.78	379798.90	7.827%
Energy	9143851.80	9406711.77	9366672.24	2.96%

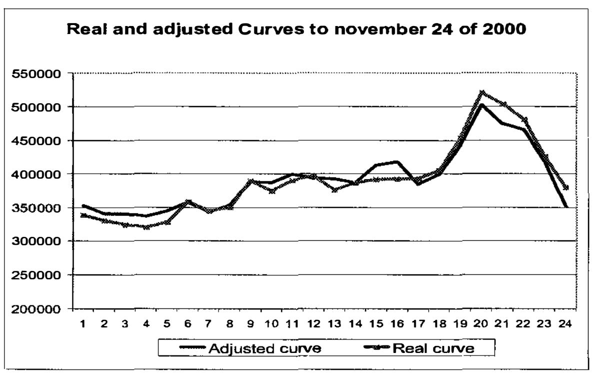

Fig.6. Real and adjusted (forecast) load curves.

IV. CONCLUSIONS

A novel method is shown to make short term demand forecast using the bases of fuzzy logic and statistical methods. This method allows making the forecast of the whole demand curve contrary to others in which the curve is forecasted by means of an hourly forecast.

CART classification and regression trees allow identifying the relevant variables for STLF without using autoregressive models in the load. In the test case, the variables used were temperature, relative humidity and speed of the winds.

Based on data analysis support rules can be generated for the expert that show the reasoning line followed by the model. Is a tool for peak power and daily energy forecast integrated to the similar day approach through the fuzzy inference system.

The model developed can be applied to any dynamic mime system in which forecast is necessary to be made. However, a limitation of the proposed method is the propagation of error because a high degree of meteorological forecast is needed.

V. BIBLIOGRAPHY

[1] Ku Anne, "The Art of Forecasting Demand", Global Energy Business March/April 2002, pp. 20-23

[2] Karanta I. y Ruusunen J., 1991, "Short term load forecasting in comunal electric utilities", Research report A40, Systems Analysis Laboratory, Helsinki University of Technology

[3] Hagan M. y Klein R., "Identification Techniques of Box and Jenkins Applied to the Problem of Short Term Load Forecasting", IEEE PES SM, Paper A 77 168-2, Mexico City, Mexico, Jul. 1977

[4] Toyoda, J., M.-C. Chen, y Y. Inoue, 1970, "An application of state estimation to short-term load forecasting, Part I: Forecasting modeling, Part II: implementation", IEEE Transactions on Power Apparatus and Systems, VOL. PAS-89, No. 7, September/October 1970, pp. 1678-1688

[5] Rahman, S., Bhatnagar, R., 1988, "An expert system based algorithm for short term load forecast", IEEE Transactions on Power Systems, Vol.3, No.2, May 1988, pp.392-399

[6] Park, D. C., M. A. El-Sharkawi, R. J. Marks II, L. E. Atlas, M. J. Damborg, 1991a, "Electric load forecasting using an artificial neural network", IEEE Transactions on Power Systems, Vol. 6, No. 2, May 1991, pp. 442-449

[7] H. Mori y H. Kobayashi, "Optimal Fuzzy Inference for Short Term Load Forecasting", IEEE Trans. on Power Systems, Vol. 11, No. 1, pp. 390-396, Feb. 1996

[8] Ajith, A. y Balkunth, N., 2001, "A Neuro Fuzzy Approach for modelling electricity demand in Victoria". Monash University. Churchill, Australia 3842

[9] Murto P, 1998, "Neural network models for short-term load forecasting" Helsinki University of Technology, Department of Engineering Physics and Mathematics, Finland, January, 1998

[10] Khotanzad, A., Abaye, A. y Maratukulam, D., 1997, "A Neural Network Based Electric Load Forecasting Systems", IEEE Transactions on Neural Networks, July 1997

[11] Gauthier, A., Isaza, C. V., Martinez, H. "Desarrollo de una herramienta computacional para clasificación e identificación de sistemas complejos con técnicas difusas". CIFI, Universidad de los Andes, Bogota, Junio 2003

VI. BIOGRAPHY

Álvaro Torres Macías: Electrical Engineer, UIS. Master in Electrical Power Eng. Master in Systems and Computers Eng. Ph.D in Electrical Engineering, Rensselaer University, U.S. Specialist in Energy Transport, ENSEM, France. Professor and Investigator of the Electric and Electronic Department at the Universidad de los Andes. Technical Manager of Consultoría Colombiana. Work areas: planning, modeling, simulating, optimization, line and distribution systems design, software and information systems.

Carolina Tranchita Rátiva: Electrical Engineer, Master in Electrical Engineering and Doctorate student in Electrical Engineering at Universidad de los Andes. Professor at the Universidad de los Andes. Work Areas: planning, reliability and expert systems application in power systems.