7th WSEAS Int. Conf. on MATHEMATICAL METHODS and COMPUTATIONAL TECHNIQUES IN ELECTRICAL ENGINEERING, Sofia, 27-29/10/05 (pp287-293)

Decision Support System for Power Systems Applying Load Forecasting and the Learning of Causal Relationships

CLÁUDIO A. ROCHA[2], ÁDAMO L. DE SANTANA[1], CARLOS RENATO L. FRANCÊS[1], ELOI FAVERO[1], VALQUIRIA MACEDO[1], UBIRATAN H. BEZERRA[1], ARMANDO TUPIASSÚ[3], VANJA GATO[3], LIVIANE REGO[1], RAFAEL COSTA[2], CIBELLE NASCIMENTO[1]
[1]Department of Electric Engineering and Computing
Federal University of Pará
Campus Universitário do Guamá, Rua Augusto Côrrea, 01 CEP 66075-110 - Belém - Pará
[2]Center of Natural and Technological Sciences
University of the Amazon
Av. Alcindo Cacela, 287 - 66060-902 - Belém – Pará
[3]Rede Celpa
Av. Gov. Magalhães Barata, 209 - 66040-170 - Belém - Pará
BRAZIL

Abstract: - This paper presents a decision support system for power load forecast and learning of causal relationships based on mathematical and computational intelligenge methods, with the purpose of defining the future power consumption of a given region, as well as to provide a mean for the analysis of correlations between the power consumption and the socio-economic and climatic data of the region.

Key-Words: Power systems, Load forecast, Power consumption, Regression, Bayesian networks.

1 Introduction

Load forecasting has always been the essential part of an efficient power system planning and operation [1]. Moreover, with the power estimations the power suppliers can estimate satisfactorily the purchase of power based on the future demand and in the relations of prices presented by the Brazilian suppliers, leading to a reduction of the difference between the amount of energy bought and consumed.

Load forecasting must manipulate historical data of power loads (in MW) recorded. Then, as basic input for the studies we have the historical data, obtained in convenient intervals. These data are influenced by many other random variables, such as temperature, humidity; seasonalities, such as vacation times, etc. All these factors are then part of the input data of the models, given the existing correlations with the consumption.

Since the methods used for load forecast only use the consumption data, it became necessary to offer a mean to analyze these correlations. Hence the use of Bayesian networks to codify the probabilistic relations of the variables.

The work here presented was originated from the studies proposed for the research project "PREDICT - Decision Support Tool for Load Forecast of Power Systems, approved by the National Agency of Power System of Brazil (ANEEL), in course since september of 2004. This project, developed together with the government of the State of Pará, the Brazilian Amazon and the power supplier of the State of Pará, aims at the implementation of a decision support system using mathematical and computational intelligence models to estimate the purchase of energy needed in the future and to make inferences on the situation of power system from the historical consumption and its correlation with the climatic and socio-economic data.

This paper is organized as follow: regression methods for load forecast is subject of section 2. In the section 3 is presented the Bayesian networks for measuring the correlations between consumption, climatic and socio-economic conditions. Results are presented and analyzed in the section 4. Section 5 presents the final remarks of the paper.

2 Demand prediction with regression methods

In this section the linear model of regression used for the data analysis is presented. The model was used to verify the trend of the data, examining the past behavior in order to produce a forecast model for it.

The data available for the analysis correspond to the power consumption of a certain period, more specifically from january of 1991 to march of 2005, in the State of Pará, as is presented on Figure 1.

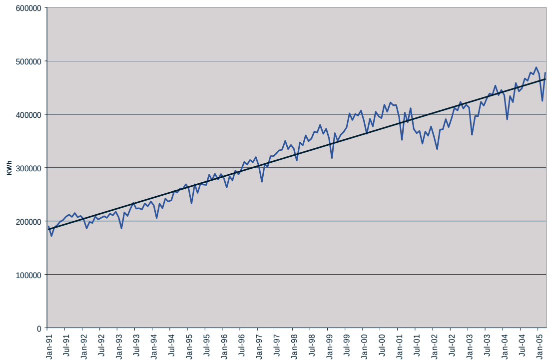
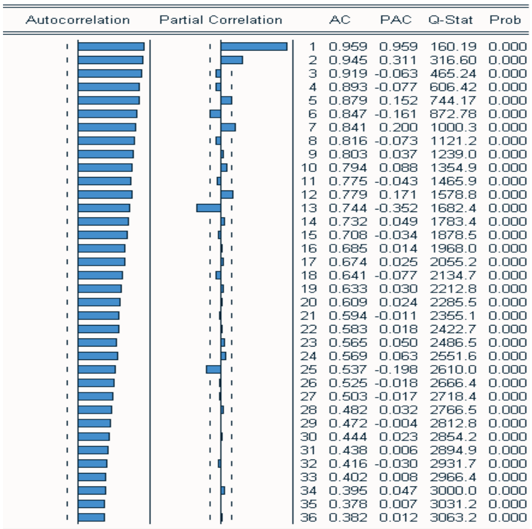


Figure 1. Power consumption from Jan/91 to Mar/05.

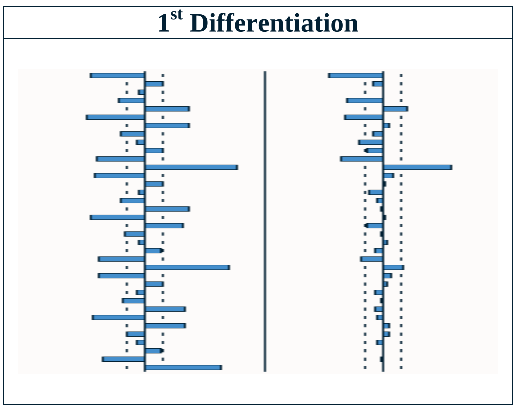
By observing the graphic we can see that neither its mean nor variance is constant, characterizing it as a non stationary series, as we can better see from its correlogram on Figure 2.

From the correlogram we can see the non stationarity of the series, not only in level, but also that it does not achieve stationarity on further differentiations (Figure 3).



	AC	PAC	Q-Stat	Prob
1	0.959	0.959	160.19	0.000
2	0.945	0.311	316.60	0.000
3	0.919	-0.063	465.24	0.000
4	0.893	-0.077	606.42	0.000
5	0.879	0.152	744.17	0.000
6	0.847	-0.161	872.78	0.000
7	0.841	0.200	1000.3	0.000
8	0.816	-0.073	1121.2	0.000
9	0.803	0.037	1239.0	0.000
10	0.794	0.088	1354.9	0.000
11	0.775	-0.043	1465.9	0.000
12	0.779	0.171	1578.8	0.000
13	0.744	-0.352	1682.4	0.000
14	0.732	0.049	1783.4	0.000
15	0.708	-0.034	1878.5	0.000
16	0.685	0.014	1968.0	0.000
17	0.674	0.025	2055.2	0.000
18	0.641	-0.077	2134.7	0.000
19	0.633	0.030	2212.8	0.000
20	0.609	0.024	2285.5	0.000
21	0.594	-0.011	2355.1	0.000
22	0.583	0.018	2422.7	0.000
23	0.565	0.050	2486.5	0.000
24	0.569	0.063	2551.6	0.000
25	0.537	-0.198	2610.0	0.000
26	0.525	-0.018	2666.4	0.000
27	0.503	-0.017	2718.4	0.000
28	0.482	0.032	2766.5	0.000
29	0.472	-0.004	2812.8	0.000
30	0.444	0.023	2854.2	0.000
31	0.438	0.006	2894.9	0.000
32	0.416	-0.030	2931.7	0.000
33	0.402	0.008	2966.4	0.000
34	0.395	0.047	3000.0	0.000
35	0.378	0.007	3031.2	0.000
36	0.382	0.012	3063.2	0.000

Figure 2. Correlogram in level of the series.



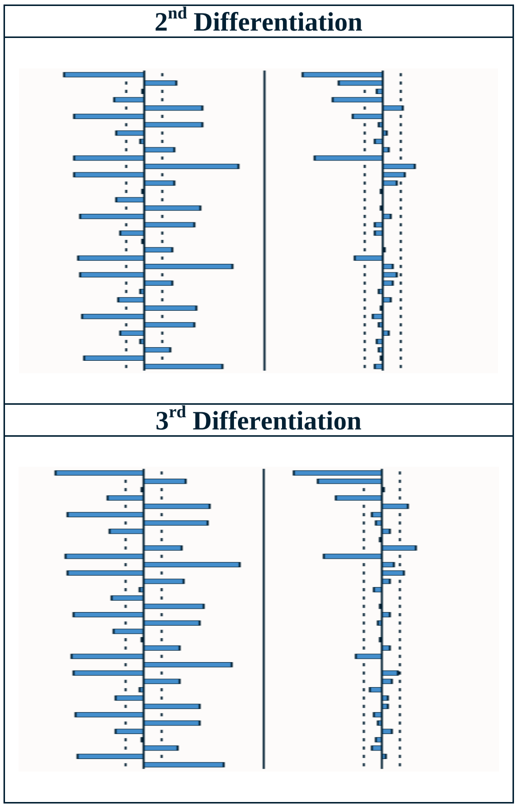
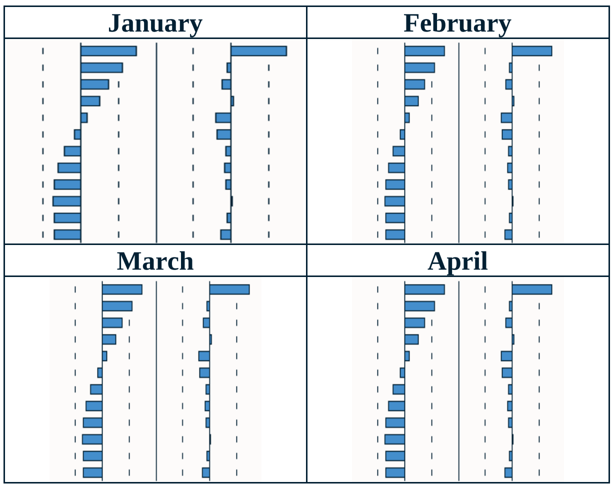


Figure 3. Autocorrelation and partial correlation of the 1st, 2nd e 3rd differentiations of the series.

Once verified from its behavior that the data represents an “explosive” series, and that it does not achieve stationarity when working with the series as a whole, a new approach was used, partitioning the once monthly series of data, in 12 annual series corresponding to the months from January to December.

From this approach, the series were then analyzed, presenting now a stationarity, as it can be seen by the correlograms on Figure 4.



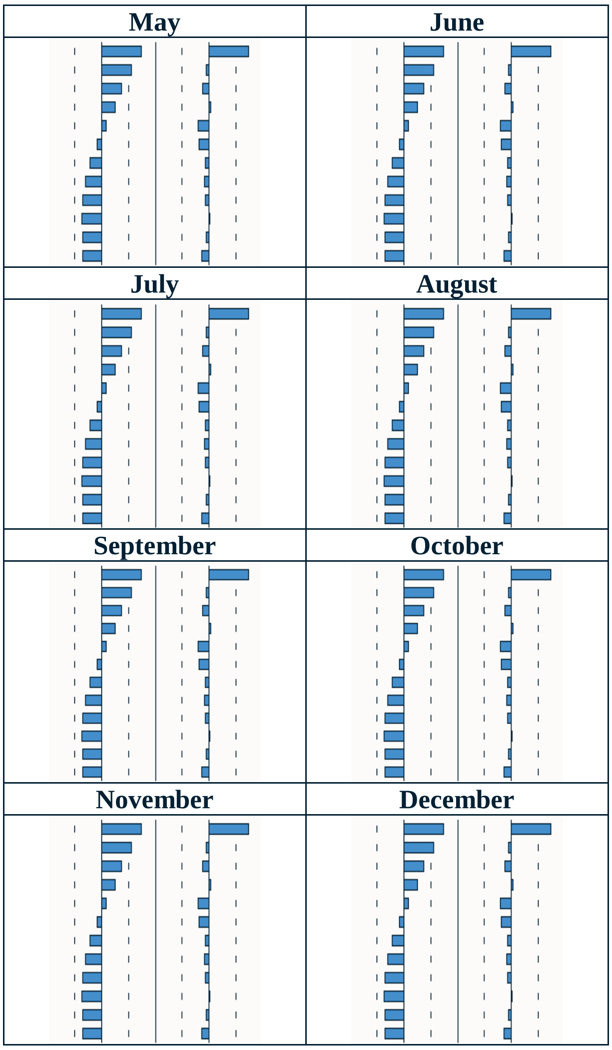


Figure 4. Autocorrelation and partial correlation of the data in the months from January to December.

The solution used for the linear model, in order to estimate future moments of the data series, is attained according to the Method of Ordinary Least Squares, assuming the restrictions of the Gauss-Markov Theorem (see [2], [3], [4]); converging the regression model to a Best Linear Unbiased Estimator (BLUE).

The linear model can be expressed according to equation (1).

$$Y_t = \alpha + \beta T_t + \varepsilon_t \qquad (1)$$

where:

Y_t is the value of the variable on period t;

α is the intercept of the regression;

β is the slope coefficient of the regression;

T_t is the value of the time variable in a period t;

ε_t characterizes the random error.

Figure 5 presents the graphic for the series, now seraparated monthly.

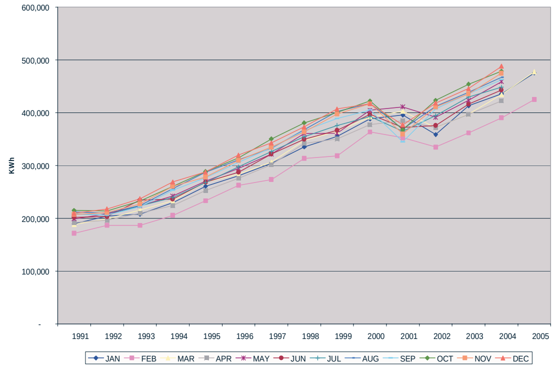


Figura 5. Power consumption from 1991 to 2005 separated monthly.

A linear growth can be observed in the series throughout time, apart from the period that goes from 2001 to 2002, characterized by the occurrence of a national measure for energy rationing. This rationing was due to the fact that around 85% of the country's energy is generated from hydroelectric facilities, and that during such period, the water level in the dams were way below the acceptable level; hence the Federal Government established a rationing, which drastically reduced the power consumption [5].

In order to calculate the growth rate of the series, so that we could verify its future behavior, a geometric growth rate model (2) was calculated.

$$Y_t = Y_0(1+r)^t \qquad (2)$$

where:

Y_0 is the initial value of the variable;

r is the growth rate.

Applying the natural logarithm on both sides of the equation we have (3):

$$\ln Y_t = \ln Y_0 + t\ln(1+r) \qquad (3)$$

which we can rewrite as (4), assuming $\alpha = \ln Y_0$ and $\beta = \ln(1+r)$.

$$\ln Y_t = \alpha + \beta T + \varepsilon_t \qquad (4)$$

The growth rate becomes then related to the variable β, which can be calculated according to (5):

$$\beta = \frac{n.\sum(Y.T) - \sum Y.\sum T}{n.\sum T^2 - \left(\sum T\right)^2} \qquad (5)$$

From the existing data (Jan/91 to Mar/05), a prior analysis was made, using only the data corresponding to the interval until the year of 2003 (Jan/91 to Dec/03), estimating from them the sequence of values for the year of 2004; in order to verify the trustworthiness of the estimator, and only then follow for a projection of its behavior for the year of 2005

Thus the growth rates were calculated for each one of the 12 series, allowing a posterior estimation for each.

The results achieved by the application of the regression model, as well as its significance will be further explained on section 4.

3 Bayesian Networks to measure the correlations among consumption, climatic, socio-economic conditions

Bayesian networks can be seen as models that codify the probabilistic relationships between the variables that represent a certain domain [6]. These models are formed by a qualitative structure, representing the dependencies between the nodes, and a quantitative one (conditional probability tables of these nodes), quantifying these dependencies in probabilistic terms [7], [8]. Together, these components offer an efficient representation of the joint probability distribution of the set of variables X for a domain [9]. The joint distribution is given by the following equation:

$$P(X) = \prod_{i=n}^{n} P(X_i \mid Pa_i) \qquad (6)$$

in which Pa_i are the father-nodes of X_i. This representation makes it possible to substantially reduce the number of probabilities that are handled.

Bayesian networks were chosen once that they ease the model understanding and the decision-making process due to the fact that the relation among the domain variables could be graphically visualized as well as their probabilistic measures.

3.1 Data selection and preparation

The database used to generate the Bayesian networks was provided by the power supplier of the State of Pará, the climatic data by the National Institute of Spatial Researches and the socio-economic by the government of the State of Pará.

The tables used to create the database were obtained considering the components of consumption, climatic and socio-economic conditions, specified as follow:

- **Consumption**, containing the data regarding the monthly consumption and its class (residential, industrial, commercial, etc);
- **Climatic**, containing the data of temperature and humidity;
- **Economic**, including the data of collection, evolution of the formal employment, inflationary indices, number of constitutions and extinction of companies in some economic sectors.

To reach the objectives specified, it was necessary the construction of datasets from original tables, considering the analyses of the dependences between the consumption and the components climatic and economic. The datasets created are presented below: Economic_Consumption, regarding the values of these two components (recorded from January of 1991 to march of 2005), in order to extract their correlations. Table 1 shows the dataset scheme.

Table 1. Schematical representation of the dataset for extration of the influence patterns of the economic factors on the power consumption.

Power consumption	**Inflationary indices**	**Number of company constitutionss**	**Collection (ICMS)**	**...**
Monthly records of the state of Pará regarding the values of the attributes of the components Consumption and Economic				

Moreover, to carry out the influence analyses of the Climatic Component on the Consumption, datasets were created for the analysis of this influence by city. The measures used were: monthly power consumption averages, maximum and minimum temperatures, relative air humidity of available and the pluviometric index of the cities. These datasets were created following the same principle adopted for the dataset Economic_Consumption.

Lastly, it is worth mentioning that the data preparation was extremely long due to the many sources used for the construction of the datasets used; being need to check and relate the information provided (power supplier, SEPOF and INPE).

3.2 Pattern extraction

After the preparation and construction of the datasets, they were submitted to the PredictBayes, software developed to extract the parameters (probabilities) from the database to a Bayesian network structure, created together with the domain specialist. The PredictBayes implements a propagation algorithm based on the junction trees method proposed by Jensen [10].

3.3 Generated Bayesian Networks

A Bayesian network was generated for the dataset Economic_Consumption and for each of the Climatic_Consumption, so that it would be possible to make a vast analysis of the impact of the climatic conditions in the consumption of energy in the State of Pará. The data used was provenient from six cities, one from each of the state region (Marajó, Southeast, Lower Amazon, Metropolitan, Southwest and Northeast).

After the creation of the Bayesian network, the propagation algorithm of the PredictBayes was used to make the inferences. In order to exemplify the resulting Bayesian networks, as well as the inferences made, consider the networks created from the dataset Economic_Consumption (Figure 6) and Climatic_Consumption of a certain city (Figure 7).

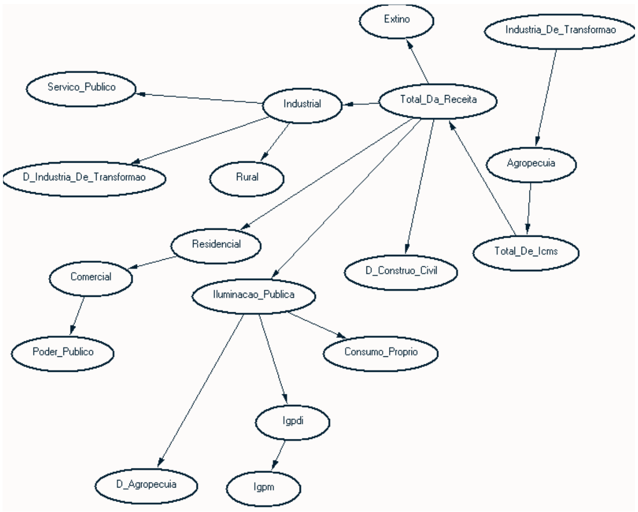


Figure 6. Bayesian network generated by the PredictBayes from the dataset Economic_Consumption.

It was possible to diagnosis that the rise of the volume of tax collection over the merchandise Circulation (Total_De_ICMS) considerably influences in the increase of the power consumption of in the industrial class. In probabilistic terms an increase of the Total_De_ICMS collection band from 1,60 to 1,85 billion to 1,85 to 2,1 billion, causes an increase in the probability of occurrence of a high power consumption of the industrial class in the State of Pará (around 760000MW) from 30% to 75%. This scenario could be used as an example for a better decision making process on the purchase of future energy, based on possible investments of the government and an increase of the collection of this important attribute.

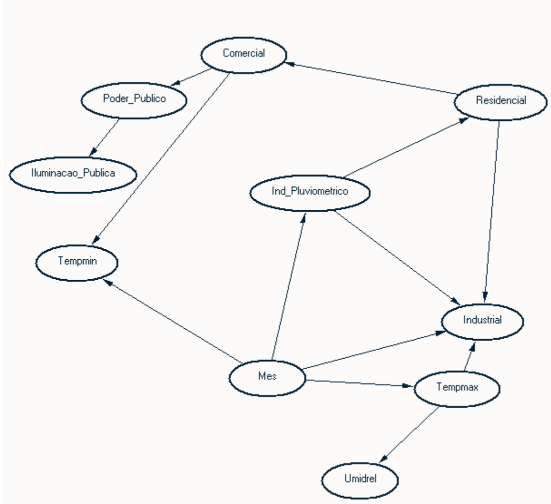


Figure 7. Bayesian network generated by the PredictBayes from the dataset Climatic_Consumo for a certain city

Following the example for the network generated from the Economic_Consumption dataset, the analyses for the consumption variation with the climatic conditions of a determined city could also be made. For example, using the Bayesian network presented in figure 7 and the following situation: we want to know the probability of the residential power consumption of a city to reach high values (between 660MW and 790MW) in the month of October, characterized by the occurrence of high temperatures (average of 33,5º to 37º) and a low pluviometric index for the Amazonian standards (less than 35mm); which would be of 43%.

This way, it was possible to, from the prospections on the Bayesian networks, estimate the many scenaries that would promote variations in the power consumption, considering the climatic and socio-economic conditions of the State of Pará. This analysis considerably diminishes the margin of error in the prediction of energy purchase based on the future demand and the relations of prices presented by the Brazilian suppliers, what would lead to a reduction of the operation costs of the system.

4 Evaluation of the obtained results

The results evaluation achieved by the application of the mathematical and computational methods was made considering two aspects:

Analysis and visualization of the dependencies, the correlation between the variables of study, which provides a graphical view, through the Bayesian networks, of the dependencies between the provided database parameters of consumption, weather and economy, as well as the quantification of these dependencies in terms of probabilities.

The evaluation of the results was made from the resources of the PredictBayes, together with the supervision of the domain specialist.

This way, the own graphical environment of the PredictBayes was used in order to verify the dependence among Bayesian network nodes. For example, from the Bayesian network showed in Figure 6 it is possible to analyze the dependences among the nodes, as well as quantifying these dependences, consulting the conditional and marginal probabilities, likelihood and variance.

Among the results obtained from the Bayesian networks we can mention:

- Due to the friendly interface and the easy understanding of the models (Bayesian networks) generated, it was possible to verify the efficiency of the learning method used;
- It was possible to diagnosis the many influences of the economic and climatic data on the power consumption. From inferences easily visualized by the domain specialist, such as that the increase of temperature causes a proportional increase the power consumption, to non trivial analyses such as that the increase of the number of employment in agriculture considerably increases the industrial power consumption;
- There is a strong trend that the power consumption of the residential class suffers influences of the pluviometric index much more than the industrial and commercial consumption classes, over 12% more;
- Once that the State of Pará possess an eminently extrativist and agricultural economy, it can be verified that the impact of the increase of the employment in the agricultural sector on the rise of the industrial power consumption is greater that the impact of employment in the transformation industries on this class of consumption.

The application of the propagation algorithm, provided by the PredictBayes made possible to accomplish the inferences suggested by the expert. Thus the exploratory character of the tool provided a sufficiently interactive way to verify the many situations of interest.

Load forecast with regression methods, from which prospection studies are made in order to estimate the power consumption values.

As it was previously specified on section 2, a prior study was made for the year of 2004 based on the data form Jan/91 to Dec/03. The result achieved by the estimation presented an error of approximately 1,47%, a value considered not only acceptable, but also inferior to all of the statistical methods used by the national power suppliers, which runs around 4%.

Once verified the effectiveness of the estimation model for the data series, a projection of its behavior for the year of 2005 was made, as it can be seen on Figure 8.

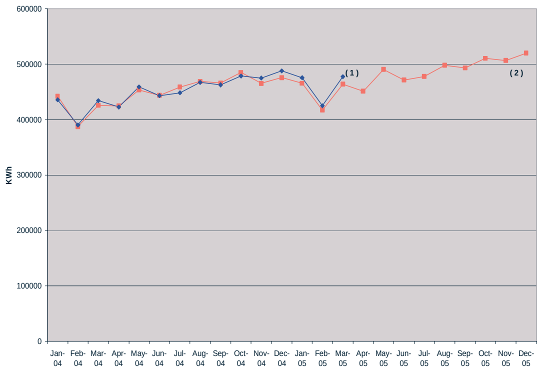


Figure 8. (1) Real values of the power consumption (Jan/04 to Mar/05); (2) Estimated values of the power consumption (Jan/04 to Dec/05).

We also point that these results were achieved in the presence of a long period of anomalous power consumption (between the years of 2001 and 2002), caused by the occurrence of a measure for energy rationing in all of the Brazilian territory.

5 Final Remarks

The objective of this study is to diagnose the relations among the consumption and economic and climatic aspects of the State of Pará, as well as to apply an estimate for the power consumption. Thus, it is possible to create a decision support system for the managers of the power suppliers, who can establish more advantageous energy contracts in the future market and analyze the favorable scenarios based on the climatic variations and the social and economic conditions of a certain region; and for the users of government agencies, so they would be able to establish policies and investments for the development of a certain region of the state.

From this point of view, the main contribution of this work was to apply the process of pattern extraction from the power consumption and estimate it, in order to establish more advantageous contracts of energy purchase in the future market for the power suppliers, as well as to provide ways to create governmental programs for social inclusion, specially since the expansion of the power supply in the

Amazon region is a predominant factor of social development.

Another relevant aspect presented here was to consider the historic of the climatic conditions and the great variety of existing socio-economic scenarios of the Amazonic Region, represented in this study by the State of Pará.

At the moment an optimized search using genetic algorithms for mining optimal inference sets based on bayesian networks is also on development, so that we could obtain the best possible set of events of the available attributes to be inferenced so that a particular attribute would be maximized.

Further studies must also be made, regarding factors such as the reliability of the system and the quality of energy supplied, which are also part of the next stage of the Predict project.

References:

[1] T. Senjyu, H. Takara, K. Uezato, and T.Funabashi, One-hour-ahead Load Forecasting Using Neural Network, *IEEE Transactions on Power Systems*, vol. 17, no. 1, february 2002.

[2] J. Gill, *Bayesian Methods: A Social and Behavioral Sciences Approach*, Chapman & Hall/CRC, 2002.

[3] R. S. Pindick, D. L. Rubinfeld, *Econometrc Models and Economic Forecasts*, Irwin/McGraw-Hill, 2004.

[4] D. N. Guajarati, *Basic Econometrics*, McGraw-Hill, 1995.

[5] ANEEL, *Atlas de energia elétrica do Brasil, Agência Nacional de Energia Elétrica*, Brasília, DF, 2003.

[6] S, Russel, P. Norvig, *Artificial Intelligence – A Modern Approach*. Prentice Hall, 2003.

[7] Z. CHEN, *Data Mining and Uncertain Reasoning - an Integrated Approach*, John Wiley Professional, 2001.

[8] S. O. Rezende, C. A. J. Rocha, Bayesian Networks for Knowledge Discovery in a Database from the Program for Genetic Improvement of the Nelore Breed. In: *2nd International Conference on Data Mining*, 2000, London, In *Proceedings of 2nd International Conference on Data Mining*, London: WIT Press, p.15 – 24, 2000.

[9] J. Pearl, *Probabilistic Reasoning in Intelligent System*, Morgan Kaufmann Publishers, 1988.

[10]F. V. Jensen e F. Jensen. Optimal junction trees, In *Proceedings of the 10th Conference on Uncertainty in Artificial Intelligence*, p. 360–366. 1994.