Decision support in power systems based on load forecasting models and influence analysis of climatic and socio-economic factors

Cláudio A. Rocha*[a], Ádamo L. de Santana[b], Carlos R. Francês[b],
Ubiratan Bezerra[b], Armando Tupiassú[c], Vanja Gato[c], Liviane Rego[b], João Costa[b]
[a]University of the Amazon, Av. Alcindo Cacela, 287, 66060-902, Belém, PA, Brazil;
[b]Federal University of Pará, R. Augusto Côrrea, 01, 66075-110, Belém, PA, Brazil;
[c]Power Supplier of the State of Pará - CELPA, Rodovia Augusto Montenegro, Km 8,5
66823-010, Belém, PA, Brazil.

ABSTRACT

This paper presents a decision support system for power load forecast and the learning of influence patterns of the socio-economic and climatic factors on the power consumption 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 these factors. Here we use a linear modelo of regression for the forecasting, also presenting a comparative analysis with neural networks, to prove its efectiveness; and also Bayesian networks for the learning of causal relationships from the data.

Keywords: 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 [2].

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 Brazilian Control Agency of Power System (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, a comparative application to the regresion analysis using neural networks is also presented in this section. Section 5 presents the final remarks of the paper.

* alex@interconect.com.br; phone +55 91 8131-9925; lprad.ufpa.br

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.

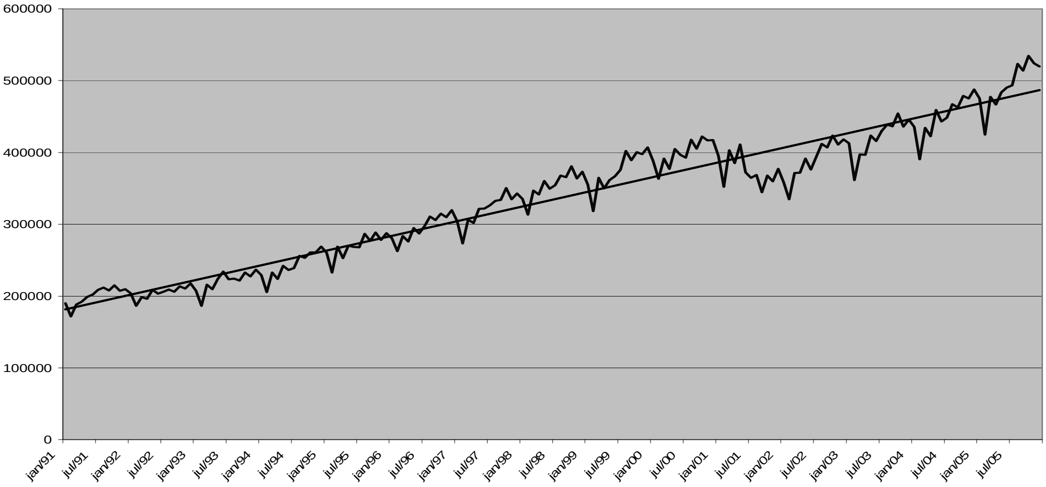
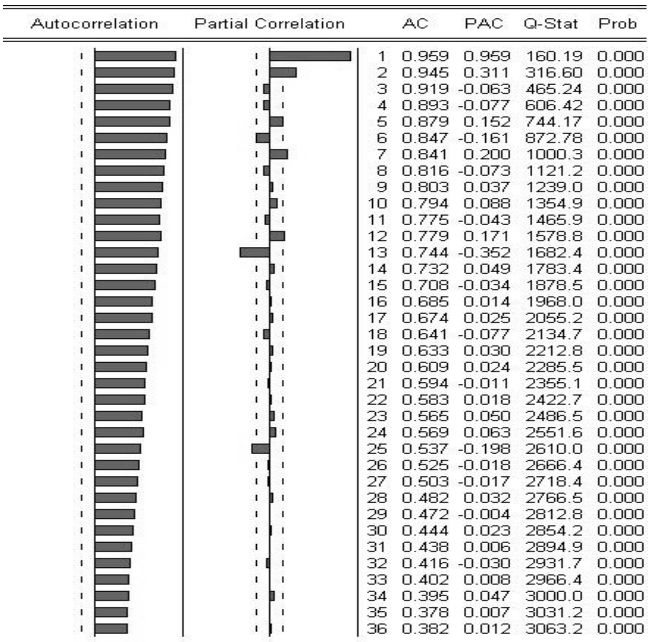


Fig. 1. Power consumption from Jan/91 to Dec/05.

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).



Autocorrelation	Partial Correlation		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

Fig. 2. Correlogram in level of the series.

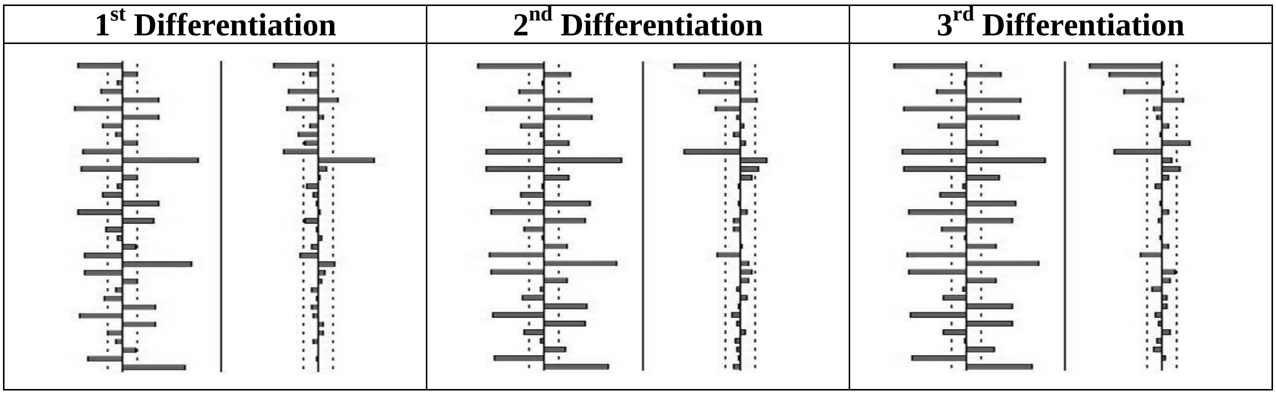


Fig. 3. Autocorrelation and partial correlation of the 1st, 2nd and 3rd differentiations of the series.

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.

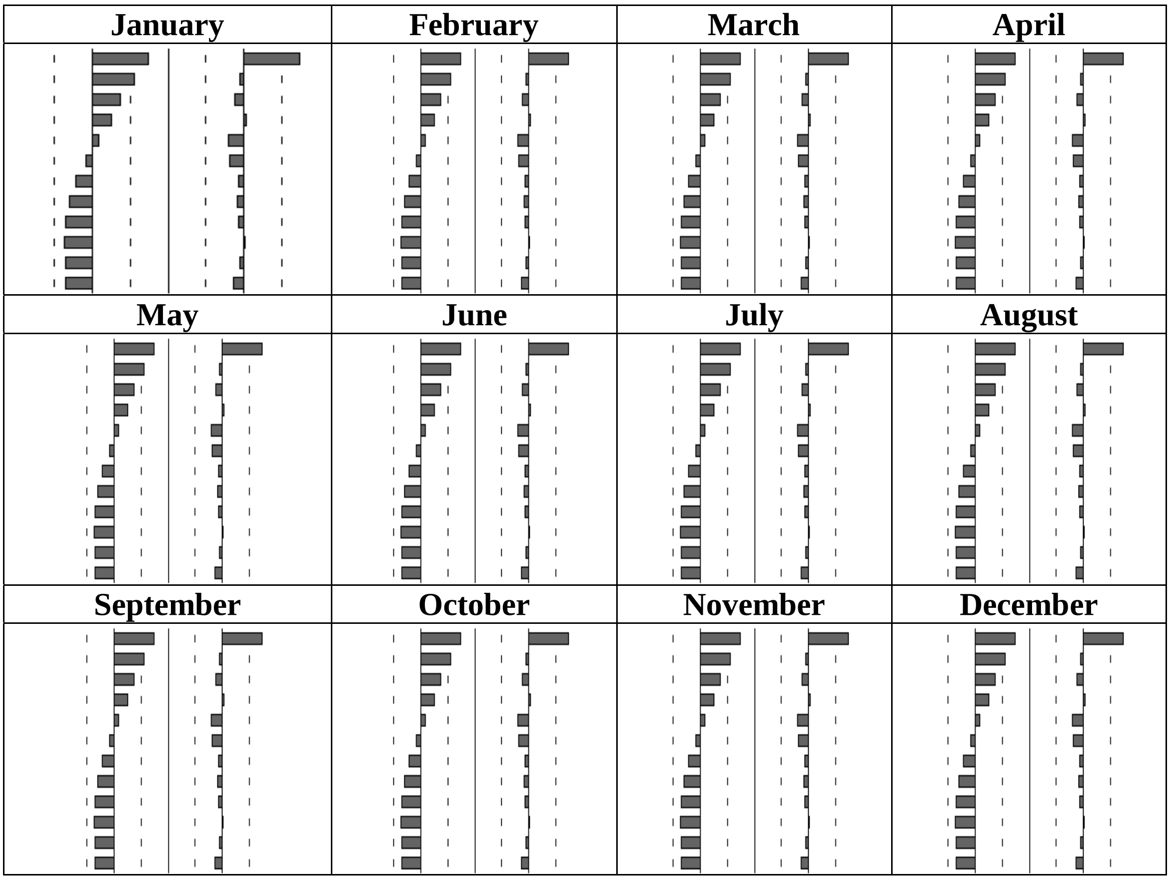


Fig. 4. Autocorrelation and partial correlation of the data in the months from January to December.

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 [3], [4], [5]); converging the regression model to a Best Linear Unbiased Estimator (BLUE).

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

$$Y_t = \alpha + \beta T_t + \varepsilon_t \tag{1}$$

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.

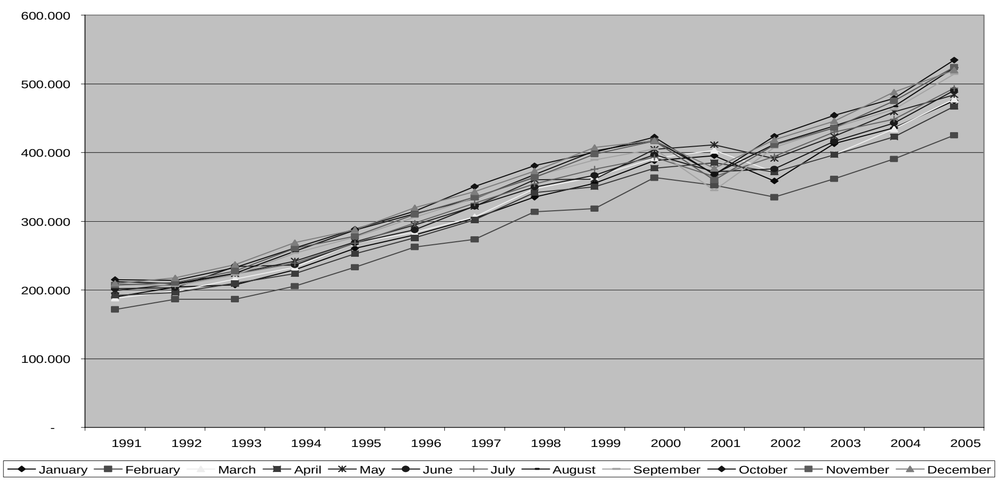


Fig. 5. Power consumption from 1991 to 2005 separated monthly.

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 [6].

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.

$$Y_t = Y_0 (1 + r)^t \tag{2}$$

Where Y_0 is the initial value of the variable and r is the growth rate.

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

$$\ln Y_t = \ln Y_0 + t \ln(1 + r) \tag{3}$$

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

$$\ln Y_t = \alpha + \beta T + \varepsilon_t \tag{4}$$

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

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

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 2004 (Jan/91 to Dec/04), estimating from them the sequence of values for the year of 2005; in order to verify the trustworthiness of the estimator, and only then follow for a projection of its behavior for the year of 2006.

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 [7] [8]. 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 [9]. Together, these components offer an efficient representation of the joint probability distribution of the set of variables X for a domain [10]. The joint distribution is given by the following equation:

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

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 pluviometric, temperature and humidity data;
- **Socio-economic**, including the data of collection, evolution of the formal employment, inflationary indices, number of constitutions and extinction of companies in some economic sectors. Based on the Bayesian network, it was verified, and confirmed by the domain experts, the correlation of the power consumption with some of the remaining variables, especially with the following: the number of employments in the sectors of the transformation industries and agropecuary, and the values of the total turnover and of the dollar; which were shown to be more representative than the others.

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 as it follows.

Table 1. Datasets created for the extration of climatic and socio-economic influence paterns over the power consumption.

DATASETS	ATTRIBUTES
Power Consumed/Climatic	Power Consumed monthly in the State of Pará and its respective climatic attributes.
Power Supplied/Climatic	Power Supplied monthly in the State of Pará and its respective climatic attributes.
Consumption by class/ Climatic	Power Consumed by class in the State of Pará and its respective climatic attributes.
City Consumption by class/ Climatic	Power Consumed in the cities by class in the State of Pará and its respective climatic attributes.
Power Consumed/Socio-Economic	Power Consumed monthly in the State of Pará and its respective socio-economic attributes.
Power Supplied/Socio-Economic	Power Supplied monthly in the State of Pará and its respective socio-economic attributes.
Consumption by class/Socio-Economic	Power Consumed by class in the State of Pará and its respective socio-economic attributes.

3.2. Pattern extraction

After the preparation and construction, the datasets were submitted to the PredictBayes, software developed to extract the parameters (probabilities) from the database into a Bayesian network, created together with the domain specialist. The PredictBayes implements a propagation algorithm based on the junction trees method proposed by Jensen [11].

3.3. Generated Bayesian networks

For the learning of the BN graphical model, that is, for the learning if existing correlations between the variables, it was used the search and scoring algorithm K2 [11], which allows us to find the most probable belief network structure B_S given a dataset D. The K2 algorithm applies a Bayesian scoring method, according (7).

$$P(B_S \mid D) = \prod_{i=1}^{n} \prod_{j=1}^{q_i} \frac{\Gamma(r_i)}{\Gamma(r_i + N_{ij})} \prod_{k=1}^{r_i} \Gamma(N_{ijk} + 1) \quad (7)$$

Where:

n is the number of nodes;

q_i is the number of configurations of the parents of the variable X_i;

r_i is the number of possible values of X_i;

N_{ijk} is the number of cases in D where the attribute X_i is evidenced with its value k, and the configuration of the parents of X_i is evidenced with value j;

N_{ij} is the number of observations in which the configuration of the parents of X_i is evidenced with the value j, being $N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$.

This way, a Bayesian network was generated from each dataset. After the construction of the Bayesian networks, the PredictBayes propagation algorithm was used to make inferences on these networks. To exemplify the Bayesian networks created and its inferences, consider the *Power Consumption/Socio-Economic* and *City Consumption* (in this example, the city of Oriximiná) *by class/Climatic* datasets (Figures 6 and 7).

From Figure 6, it is possible to visualize, besides the correlations between the climatic attributes and the power consumed in the state of Pará, the marginal probabilities which, for simplification matters, are presented only for some attributes of the network.

commercial	
[126,900 to 145,500)	0.1764
[145,500 to 164,000)	0.2731
[164,000 to 182,600)	0.2151
[182,600 to 201,100)	0.1181
[201,100 to 219,600]	0.2172

pluviometric_rate	
[1.497 to 64.260)	0.5215
[64.260 to 127.000)	0.0599
[127.000 to 189.800)	0.1740
[189.800 to 252.000)	0.1716
[252.000 to 219.300]	0.0730

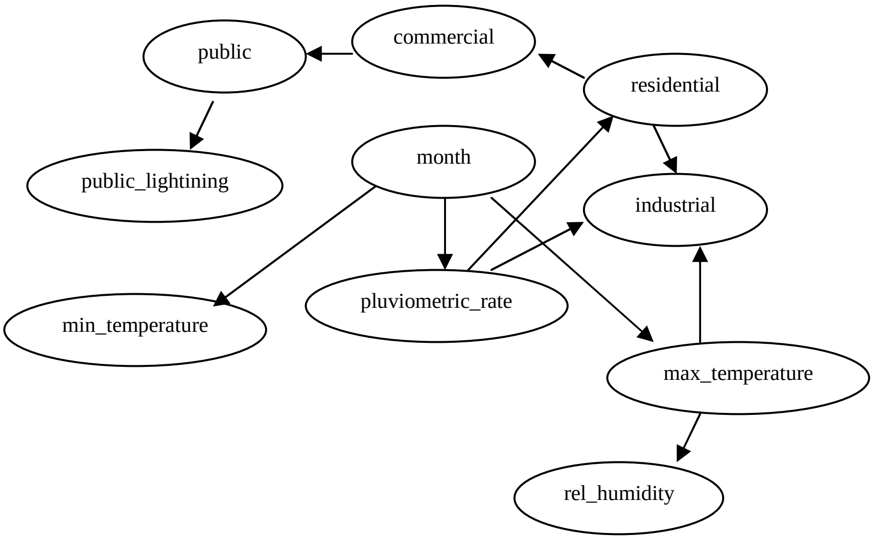


Fig. 6. Bayesian network created from the City Consumption by class/climatic dataset.

In Figure 7, for example, the following inference could be made: which is probability for the consumption to achieve the highest value (be in the 5th band of values of the attribute *Power consumed*), given the evidence of an increase in the number of employment in the agropecuary sector was within the band from 2,113 to 2,725 (4th band of values of the *Agropecuary* atribute)? The answer to this question is observed in the probability presented for the band with the highest power consumption (40.74%).

trans_industry	
[1,266 to 1,990)	0.0017
[1,990 to 2,714)	0.0037
[2,714 to 3,439)	0.4968
[3,439 to 4,163)	0.4968
[4,163 to 4,887]	0.0010

power_consumed	
[2.689E8 to 2.871E8)	0.0018
[2.871E8 to 3.053E8)	0.0018
[3.053E8 to 3.235E8)	0.1567
[3.235E8 to 3.417E8)	0.4324
[3.417E8 to 3.598E8]	0.4074

agropecuary	
[275 to 888)	0.0000
[888 to 1,500)	0.0000
[1,500 to 2,113)	0.0000
[2,113 to 2,725)	**1.0000**
[2,728 to 3,338]	0.0000

turnover_total	
[9.290E7 to 1.171E8)	0.0010
[1.171E8 to 1.429E8)	0.0010
[1.429E8 to 1.687E8)	0.0010
[1.687E8 to 1.946E8)	0.4988
[1.946E8 to 2.204E8]	0.4983

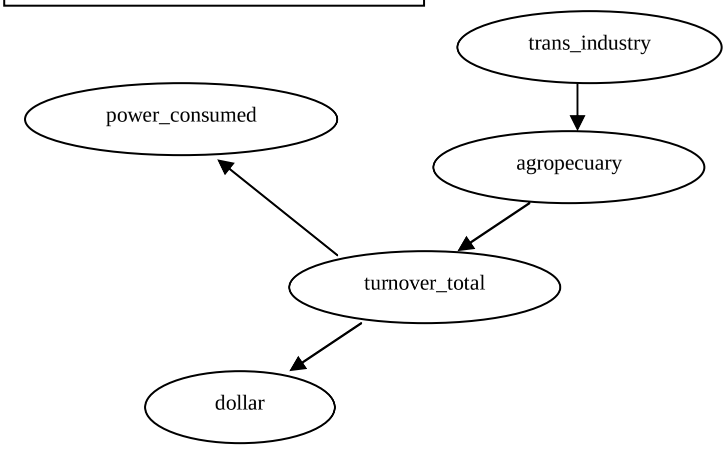


Fig. 7. Bayesian network created from the power consumed/socio-economic dataset.

It is possible, from the prospections (inferences) over the Bayesian networks, to present the users with many scenarios which can promote variations in the power consumption, given the climatic and socio-economic conditions of the state of Pará. This analysis can considerably aid in the analysis process of the power demands, what can lead to an anticipated decision making and, consequentially, a reduction in the operation costs of the system.

4. EVALUATION OF THE OBTAINED RESULTS

The evaluation of the results obtained by the application of the mathematical and computational methods was made considering two aspects: load forecast with regression methods and the analysis and visualization of the dependencies.

4.1. Load forecast with regression methods

Here the prospection studies were 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 2005 based on the data form Jan/91 to Dec/04. 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%. This reduction represents, evidently, a considerable economy for energy purchase in a future market.

As a comparative study, a load forecasting analysis using neural networks [13] and Kalman filters [6] were made. The neural network used in this study was a feed-forward multi-layer network with two hidden layers and using as learning algorithm the adaptative backpropagation.

The same data of the annual consumption series supplied as input for the regression analysis, as described in section 2, were also submitted to the neural network and the Kalman filter. The results obtained for the regression methods, neural network and Kalman filter were of 1.47% 4.08% and 5.08% respectively.

Once verified the effectiveness of the regression estimation model for the data series, a projection of its behavior for the year of 2006 (Figure 8).

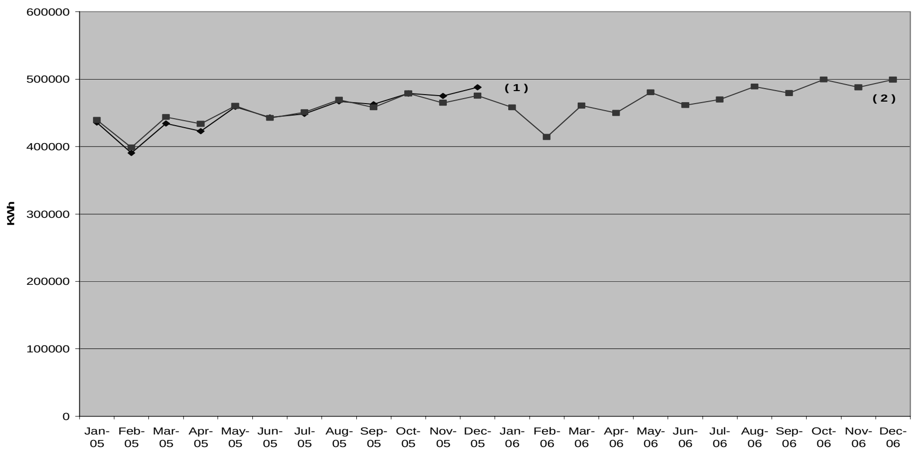


Fig. 8. (1) Real values of the power consumption (Jan/05 to Dec/05); (2) Estimated values (Jan/05 a Dec/06).

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.

4.2. Analysis and visualization of the dependencies

The correlation among 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.

The application of the propagation algorithm, provided by the PredictBayes, made possible a series of inferences suggested by the specialist. Thus, the exploratory character of the Bayesian networks propitiated, through the inferences made in the variable of the domain, the interactive analysis of the many situations of interest. Some examples of the analyses made are presented as follow:

- The diagnosis of the influence of the economic and climatic factors on the power consumption. The inferences made allowed the domain specialists to visualize, in quantitative terms, trivial influences such as that the temperature increase inflicts a proportional increase in the consumption; as well as those that are not so trivial, such as the influence that the increase of employments in the agropecuary sector inflicts on this consumption;
- The residential consumption class is the one which suffer most influence with the action of rains. When compared with the industrial and commercial classes, for example, the impact of the pluviometric index increase on the residential consumption is, usually, 12% higher than on the consumption of these two classes.
- Once that the economy of Pará is strongly directed toward the extrativist and agropecuary sectors, it was possible to verify that the increase of employment in the agropecuary sector influences more on the industrial consumption than the employment in the transformation industry.

Moreover, it is possible to perceive the many diagnostics of the consumption and its relations with the economic and climatic aspects of the state of Pará; what, allied to the module of prediction, allows to observe and measure the impacts from changes in the climatic scenarios and in certain sectors of the power consumption economy.

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.

ACKNOWLEDGEMENTS

The authors would like to thank the institutions CELPA/ANEEL, CAPES, and CNPq for their financial support.

REFERENCES

1. T. Senjyu, H. Takara, K. Uezato, T. Funabashi, "One-hour-ahead Load Forecasting Using Neural Network", IEEE Transactions on Power Systems, Volume 17, 1, (2002).
2. A. P. Douglas, A. M. Breipohl, F. N. Lee, R. Adapa. "The impacts of temperature forecast uncertainty on Bayesian load forecasting", IEEE Transactions on Power Systems, Volume 13, (1998).
3. J. Gill, *Bayesian Methods: A Social and Behavioral Sciences Approach*, Chapman & Hall/CRC, 2002.
4. R. S. Pindick, D. L. Rubinfeld, *Econometric Models and Economic Forecasts,* Irwin/McGraw-Hill, 2004.
5. D. N. Guajarati, *Basic Econometrics*, McGraw-Hill, 1995.
6. ANEEL, *Brazilian power systems atlas*, Agência Nacional de Energia Elétrica, (in Portuguese), 2003.
7. S. Russel, P. Norvig, *Artificial Intelligence - A Modern Approach*, Prentice Hall, 2003.
8. K. B. Korb, A. E. Nicholson, *Bayesian Artificial Intelligence*, CRC PRESS, 2003.
9. Z. Chen, *Data Mining and Uncertain Reasoning - an Integrated Approach*, John Wiley Professional, 2001.
10. J. Pearl, *Probabilistic Reasoning in Intelligent System*, Morgan Kaufmann Publishers, 1988.
11. F. V. Jensen, F. Jensen, "Optimal junction trees", In Proceedings of the 10th Conference on Uncertainty in Artificial Intelligence, 360-366, (1994).
12. G. Cooper, E. Herskovitz, "A Bayesian Method for the Induction of Probabilistic Networks from Data", Machine Learning, 9, 309-347, (1992).
13. S. Haykin, *Neural networks: a comprehensive foundation*, Prentice Hall, 1998.