Electrical Power and Energy Systems 38 (2012) 33–45

Contents lists available at SciVerse ScienceDirect

Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes

PREDICT – Decision support system for load forecasting and inference: A new undertaking for Brazilian power suppliers

Ádamo L. Santana *, Guilherme B. Conde, Liviane P. Rego, Cláudio A. Rocha, Diego L. Cardoso, João C.W. Costa, Ubiratan H. Bezerra, Carlos R.L. Francês

Laboratory of High Performance Networks Planning, Federal University of Pará, R. Augusto Correa, 01, 66075-110 Belem, PA, Brazil

ARTICLE INFO

Article history:
Received 6 May 2009
Received in revised form 18 November 2011
Accepted 18 December 2011
Available online 24 January 2012

Keywords:
Decision support system
Power load forecast
Time series analysis
Wavelets
Neural networks
Bayesian networks

ABSTRACT

One of the most desired aspects for power suppliers is the acquisition/sale of energy for a future demand. However, power consumption forecast is characterized not only by the variables of the power system itself, but also related to social–economic and climatic factors. Hence, it is imperative for the power suppliers to project and correlate these parameters. This paper presents a study of power load forecast for power suppliers, considering the applicability of wavelets, time series analysis methods and artificial neural networks, for both mid and long term forecasts. Both the periods of forecast are of major importance for power suppliers to define the future power consumption of a given region. The paper also studies the establishment of correlations among the variables using Bayesian networks. The results obtained are much more effective when compared to those projected by the power suppliers based on specialist information. The research discussed here is implemented on a decision support system, contributing to the decision making for acquisition/sale of energy at a future demand; also providing them with new ways for inference and analyses with the correlation model presented here.

1. Introduction

Power load forecasting has always been the essential part of an efficient power system planning and operation [1–3]. With it, power suppliers can satisfactorily estimate the purchase of energy based on the future demand and prices, minimizing the difference between the amounts of energy bought and consumed.

The base input for the power load forecasting studies is the historical data of the consumption (in MW h), recorded at convenient intervals. Moreover, the data is influenced by many other random variables, such as temperature, humidity and social–economic factors [4,5].

The work here presented was developed together with the government of the State of Pará (State of the Brazilian Amazon area), the Brazilian Control Agency for Electrical Energy (ANEEL) and the Power Network Group (*Rede Energia*), a conglomerate with some of the main power suppliers in Brazil; aiming 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 situations of the power system. The project is entitled "PREDICT – Decision Support Tool for Load Forecast of Power Systems".

The decision support system is based on two approaches: (a) load forecasting, by means of time series models, artificial neural networks and wavelets and (b) correlation and analysis of dependences, using Bayesian networks.

The paper applies the models of time series with regression, neural networks and wavelets, as techniques for forecasting. The data used corresponds to the amount of required energy consumption, whose value is composed of several classes of energy consumption; these component classes are related to the industrial, residential, commercial and agricultural consumptions, as well as public illumination, among others.

There is a variety techniques developed to improve the load forecasting accuracy with many degrees of success. Theses methods can be divided into traditional forecasting and artificial intelligence; strategies range from the application of traditional regression analysis [6–8], to computational intelligence models; the latter usually consider the application of artificial neural network and bioinspired algorithms, to treat more complex, non-linear data [9–11]; evenly with pure and hybrid models [12–14]. These approaches consider load forecasting analysis in different application domains, usually with better accuracy, when compared with the strategy currently installed in the company under consideration. In this work, given the specificity of the application and data, the accuracy of the estimators and hybrid models developed

* Corresponding author. Tel.: +55 91 3201 8112.
E-mail addresses: adamo@ufpa.br (Á.L. Santana), conde@ufpa.br (G.B. Conde), liviane@ufpa.br (L.P. Rego), alex@ufpa.br (C.A. Rocha), diego@ufpa.br (D.L. Cardoso), jweyl@ufpa.br (J.C.W. Costa), bira@ufpa.br (U.H. Bezerra), rfrances@ufpa.br (C.R.L. Francês).

0142-0615/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ijepes.2011.12.018

for this particular domain will be compared based with the power supplier's models in effect for this task.

This work widens the previous study for load forecasting in [15], by incorporating more complex and sophisticated models, also considering the anomalies in the series, its intrinsic monthly and annual growth rate; as well as improving the estimator by separately treating the trends and fluctuations with wavelets based framework.

In order to provide a comprehensive decision support system, while considering demands by domain experts, together with more accurate models for mid and long term forecasting, tools were also implemented to measure the impact that other random variables have on the power consumption; such that new and alternate scenarios can be inferred considering different settings for electrical system operation (i.e. whether they are more economic, safe and/or reliable). Thus, the correlation of some exogenous variables to the electrical system, specifically related to climatic and social–economic factors, served as basis for the correlation and inference studies performed here.

The use of diagnostics modelling analyse the impact that certain factors (endogenous or exogenous) have on the operation, distribution and planning of electrical systems has attracted much attention in the literature; examples of this are: fault diagnosis of equipment; detection of frauds and other non-technical losses; reliability assessment of electrical systems; and scenario analysis of power consumption [16–18].

Hence the use of Bayesian networks to codify the probabilistic relations of the variables and to make inferences on the conditions of the power system from the historical consumption and its correlation with the climatic and social–economic data.

This paper is organized as follows: the regression method for load forecast is subject of Section 2. Section 3 presents the neural networks model used and the definition of its parameters. Section 4 presents the wavelets model applied for the forecasting. In Section 5, the Bayesian networks for measuring the correlations from the consumption, climatic and social–economic conditions are presented. Section 6 presents the final remarks of the paper.

2. Time series and regression model for load forecasting

In this section the regression model used for the data analysis is presented. The model was used to verify the trend of the data, examining the past behaviour to develop a forecast model for it. The data available for the analysis correspond to the total power consumption class. The study used the historical power consumption data available in the period from January 1991 to December 2006 (Fig. 1).

As discussed in previous studies [15], the consumption time series is tendentious and non-stationary. The series, by studying its correlograms, does not achieve stationarity either on successive differentiations (Fig. 2).

Once verified from its behaviour that the data represents an "explosive" series, and that it does not achieve stationarity when working with the entire series, a new approach was used; partitioning the series, in twelve annual series corresponding to the months from January to December (Fig. 3). These partitioned series when studied showed the stationarity feature.

An almost linear growth could then be observed in the series throughout time, apart from the period ranging from 2001 to 2002, characterized by the occurrence of a national measure for energy rationing which drastically reduced the power consumption [19]. This rationing was due to the fact that during that period water in the reservoirs was below the acceptable level. It is important to stress that 85% of Brazil's energy is originated from hydroelectric facilities.

The estimator for the consumption prediction uses a multiple regression analysis (see [20–22]), based on the value of the consumption at a previous time and two additional terms. The general formula of the multiple regression model can be specified as follows:

$$Y_i = A_0 + A_1X_{1i} + A_2X_{2i} + \cdots + A_kX_{ki} + u_i \tag{1}$$

Thus, the general system of the multiple regressions can be seen in a matrix notation

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1k} \\ 1 & X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{nk} \end{bmatrix} \times \begin{bmatrix} A_0 \\ A_1 \\ \vdots \\ A_k \end{bmatrix} + \begin{bmatrix} u_0 \\ u_1 \\ \vdots \\ u_k \end{bmatrix} \tag{2}$$

where Y is a column vector, with dimension $n \times 1$; X is a matrix of size $n \times k$, that is, with n observations and k variables; with the first column representing the intercept A_0; A is a vector with $k \times 1$ unknown parameters; u is a vector with $n \times 1$ disturbances.

Although the non-stationary issue is somewhat solved by partitioning the series, a loss of knowledge must also be considered, due to events exogenous to the standard behaviour of the system. That is, as example, new governmental programs, or contracts acquired by the suppliers, would cause a shift in the series; which might not be fully accounted for in a direct forecasting of each month, separately.

With this in mind, in order to consider the impact of such events and thus obtain a more adjustable value for the prediction, a variable quantifying the annual trend of the consumption was added.

The variable included was obtained from a factorial analysis (for a more complete view on factorial analysis see [23]), in order to condense the information and trends occurred in the year. The factorial analysis denotes the reduction of a set of variables from a domain to a model with fewer factors, maintaining the representativity and relations among the original variables; by applying it over the twelve annual series, which presented a near linear growing behaviour, the correlated growth varying in the series was modelled as a single variable (factor) that best represented the series (around 99.6% of representativity was obtained) and, consequently, the annual behaviour.

A second term was also added to the model, for containing the impact from anomalies in the historical power consumption data. Here we approach the anomalous period of June 2001–February 2002 by adding a binary artificial variable to the monthly series. This variable indicates the presence or absence of anomalies in the historical data, attributing 1 or 0, respectively. Not only the period when the rationing measure was installed is treated, but also the months that followed, persisting in a decrease in the power consumption, until the series returned to normality.

The model is firstly applied to verify the reliability of the estimator. Thus, from the existing data (January/1991–December/2006), some tests were conducted over the regression model. Only then, the estimator was applied to project the behaviour for 2007 and 2008. The results achieved by the application of the regression model, as well as its significance will be presented next. The performance of the model will be evaluated according to mean absolute percentage error (MAPE), calculated according to

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{|y_i - \hat{y}_i|}{y_i}\right) \times 100\% \tag{3}$$

where N is the number of existing samples, y is the real historical value and $\hat{y}$ is the estimated value. The results from the neural networks and the wavelets methods, presented in Sections 3 and 4, respectively, are also based on MAPE.