Available online at www.sciencedirect.com
ScienceDirect
International Journal of Forecasting 24 (2008) 630–644
international journal of forecasting
www.elsevier.com/locate/ijforecast

Modeling and forecasting short-term electricity load: A comparison of methods with an application to Brazilian data

Lacir J. Soares [a], Marcelo C. Medeiros [b,*]

[a] UENF, Campos dos Goytacazes, Brazil
[b] Department of Economics, Pontifical Catholic University of Rio de Janeiro, Brazil

Abstract

The goal of this paper is to describe a forecasting model for the hourly electricity load in the area covered by an electric utility located in the southeast of Brazil. A different model is constructed for each hour of the day. Each model is based on a decomposition of the daily series of each hour in two components. The first component is purely deterministic and is related to trends, seasonality, and the special days effect. The second is stochastic, and follows a linear autoregressive model. Nonlinear alternatives are also considered in the second step. The multi-step forecasting performance of the proposed methodology is compared with that of a benchmark model, and the results indicate that our proposal is useful for electricity load forecasting in tropical environments.

Keywords: Short-term load forecasting; Time series; Seasonality; Linear models; SARIMA; Time-series decomposition

1. Introduction

One kind of time series which is of major interest, both academic and practical, is the hourly electricity load. From an academic point of view, the series are remarkable because they have a number of interesting features, such as trends, annual and daily seasonal patterns, an influence of external variables, and possible nonlinearities. In addition, load series have been used over the years as a benchmark data set for different forecasting models.

From the applied point of view, short-term load forecasting is a very important task for electric utilities in order to manage the production, transmission, and distribution of electricity in a more efficient and secure way. As an example of the importance of accurate forecasts, it was estimated that an increase of only 1% in the forecast error (in 1984) caused an increase of 10 million pounds in operating costs per year for one electric utility in the United Kingdom (Bunn & Farmer, 1985b).

Over the years, different forecasting techniques have been developed to model the electricity load, both in the classical time series literature (Al-Hamadi & Soliman, 2004; Amjady, 2001; Bunn & Farmer, 1985a;

* Corresponding author. Departament of Economics. Pontifical Catholic University of Rio de Janeiro (PUC-Rio) Rua Marquês de São Vicente, 225 - Gávea 22453-900 Rio de Janeiro, RJ Brazil. Tel.: +55 21 3527 1078; fax: +55 21 3527 1084.

E-mail addresses: lacir@pesquisador.cnpq.br (L.J. Soares), mcm@econ.puc-rio.br (M.C. Medeiros).

0169-2070/$ - see front matter © 2008 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.ijforecast.2008.08.003

Cancelo, Espasa, & Grafe, 2008-this issue; Dordonnat, Koopman, Ooms, Dessertaine, & Collet, 2008-this issue; Huang, 2003; Huang, Huang, & Wang, 2005; Nowicka-Zagrajek & Weron, 2002; Taylor, 2008-this issue; Taylor, de Menezes, & McSharry, 2006; and Weron, 2006), and in the machine intelligence framework (da Silva, Ferreira, & Velasquez, 2008-this issue; Hippert, Bunn, & Souza, 2005; Hippert, Pedreira, & Souza, 2001; Khotanzad, Afkhami-Rohani, & Maratukulam, 1998; and Metaxiotis, Kagiannas, Askounis, & Psarras, 2003). Feinberg and Genethliou (2005) provide an updated review of different methods.

In this paper, we consider a methodology based solely on rigorous statistical arguments for modelling and forecasting the hourly electricity load of part of the southeast of Brazil. The area covered by the electric utility represents 25% of the state of Rio de Janeiro, totalling 11,132 km^2, and with a population of more than ten million people. The energy consumption corresponds to 75% of the total consumption in the Rio de Janeiro state. It is worth mentioning that this is one of the most important regions for tourism in Latin America. We adopt the same strategy as Fiebig, Bartels, and Aigner (1991), Peirson and Henley (1994), Ramanathan, Engle, Granger, Vahid-Arahi, and Brace (1997), Cottet and Smith (2003), and Soares and Souza (2006), treating each hour as a separate time series, such that 24 different models are estimated, one for each hour of the day. The approach considered in this paper is based on a two-step decomposition of the load series. In the first step, a component based on Fourier series, dummy variables, and a linear trend, is estimated to describe the long-run trend, the annual seasonality, the effects of the days of the week, and any other special days effects such as public holidays. In the second step, different linear autoregressive (AR) models are estimated. Neural network models are also considered in the second step. The type of decomposition considered here is not new. Similar proposals have been discussed in the literature during the last two decades; see, for example, Harvey and Koopman (1993), Temraz, Salama, and Quintana (1996), and Nowicka-Zagrajek and Weron (2002). However, we contribute to the literature in several different ways. First, to the best of our knowledge, the way in which we specify the models in each component is not common in the load forecasting literature, and relies only on rigorous classical statistical arguments. Recently, Cottet and Smith (2003) proposed a similar approach, but their methodology is fully based on Bayesian statistics and is computationally very demanding. Our methodology is simpler and can be used efficiently for real-time online load forecasting. Second, although very simple, this methodology is very flexible, allowing for different specifications in the second step. For example, neural networks and other nonlinear models may be estimated instead of a simple AR model. However, we show that the nonlinear effects are mainly related to the time-varying conditional variance and are not present in the conditional mean. Thus, the linear model is adequate to describe the dataset considered here. Furthermore, based on the bootstrap resampling technique, confidence intervals may easily be constructed under mild assumptions on the errors of the model. Finally, exogenous variables, when available, may easily be incorporated into the model.

The plan of the paper is as follows. Section 2 describes the dataset used in the paper. Section 3 presents the model and the modeling strategy. The benchmark model is discussed in Section 4. Section 5.1 shows the modeling results, and Section 5.2 presents the forecasting results. Final remarks are found in Section 6.

2. The Data

We consider a dataset containing hourly loads from January 1, 1990 to December 31, 2000. The period from January 1, 1990 to December 31, 1998 is used for estimation purposes (in-sample), and the data from the years 1999 and 2000 are left for forecast evaluation (out-of-sample). The data were obtained from an utility company from Rio de Janeiro, Brazil, and are shown in Fig. 1. This is the same dataset considered by Soares and Souza (2006). Fig. 1 shows the daily loads for each hour of the day during the in-sample period.

3. The Model

3.1. Mathematical Definition

The hourly load is modeled as the sum of two components. The first component is deterministic, representing the trend, the annual cycle, and the effects