Functional clustering and linear regression for peak load forecasting

Aldo Goia[*], Caterina May[1], Gianluca Fusai[2]

Dipartimento di Scienze Economiche e Metodi Quantitativi (SEMeQ), Università del Piemonte Orientale "A. Avogadro", Via Perrone, 18, 28100 Novara, Italy

Abstract

In this paper we consider the problem of short-term peak load forecasting using past heating demand data in a district-heating system. Our data-set consists of four separate periods, with 198 days in each period and 24 hourly observations in each day. We can detect both an intra-daily seasonality and a seasonality effect within each period. We take advantage of the functional nature of the data-set and propose a forecasting methodology based on functional statistics. In particular, we use a functional clustering procedure to classify the daily load curves. Then, on the basis of the groups obtained, we define a family of functional linear regression models. To make forecasts we assign new load curves to clusters, applying a functional discriminant analysis. Finally, we evaluate the performance of the proposed approach in comparison with some classical models.
© 2009 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

Keywords: Short-term forecasting; Out-of-sample; Load curve; Seasonality; Functional regression; Functional clustering; Functional linear discriminant analysis

1. Introduction

Load demand forecasting is becoming more and more important as power generation costs increase and market competition intensifies: accurate forecasts are relevant to energy systems for scheduling generator maintenance and choosing an optimal mix of on-line capacity. The literature on load forecasting considers three main problems: long-term forecasts for system planning, medium-term forecasts for maintenance programs, and short-term prediction for the day-to-day operation, scheduling and load-shedding plans of power utilities. A central role is played by the intra-daily pattern of the load demand, known as the *load curve*, which describes the amount of energy consumed to satisfy the load demand of customers over the course of the day.

---

[*] Corresponding author. Tel.: +39 0321 375 319; fax: +39 0321 375 305.

E-mail addresses: aldo.goia@eco.unipmn.it (A. Goia), cmay@mat.unimi.it (C. May), gianluca.fusai@eco.unipmn.it (G. Fusai).

URL: http://semeq.unipmn.it/ (A. Goia, C. May, G. Fusai).

[1] Tel.: +39 0321 375 334; fax: +39 0321 375 305.
[2] Tel.: +39 0321 375 312; fax: +39 0321 375 305.

The focus of the present work is on short-term (i.e. around 24 h) forecasting of the daily peak load in a district-heating (or "teleheating") system, which is the maximum of the daily demand for heating. A district-heating system involves distributing the heat for residential and commercial requirements via a network of insulated pipes. We analyze here data on heat consumption in a major Italian centre, Turin, where the district heating is produced through a co-generation system. This technology allows for energy saving and reduces emissions compared to old technologies; it is therefore treated as a renewable energy source and delivers substantial economic and environmental benefits.

In the recent literature concerning prediction in district-heating systems (see, for example, Dotzauer, 2002, and Nielsen & Madsen, 2006, for some applications and references), the algorithms employed are usually similar to those used in the prediction of electrical-power loads. A review of statistical methods for electrical load forecasting has been given by Weron (2006), for instance. These methods are mainly based on ARIMA models, regression models, exponential smoothing, and generalizations of these. Typically, weather variables are used for the prediction of electricity loads, and a great range of modeling approaches are presented in the literature. Developments in forecasting methodologies are also reflected by the contributions in the special issue of the *International Journal of Forecasting* on energy forecasting that appeared in 2008. For instance, Dordonnat, Koopman, Ooms, Dessertaine, and Collet (2008) develop a multi-equation model with time-varying parameters, while Alves da Silva, Ferreira, and Velasquez (2008) propose automated input selection procedures for forecasting models based on neural networks. Whereas electric and wind power, gas consumption and electricity price forecasting are discussed in the cited issue, the present paper deals with heat consumption in a residential area. In this case, the strong correlation between heating and external temperature has led us to a parsimonious model without exogenous variables, which forecasts satisfactorily without requiring the inclusion of weather variables.

In this paper we propose a method based on a functional statistics approach. The functional statistics approach has become the object of an increasing amount of attention on the part of many researchers and practitioners in recent years, since it can be applied when data are collections of discrete observations effected on curves, images or shapes. A survey of these techniques can be found in the monographs of Ramsay and Silverman (2005) and Ferraty and Vieu (2006), for instance. Our data-set consists of hourly observations of the heat consumption over four separated periods, with 198 days in each period, over the years 2001-2005. The load data are registered in megawatts, and have been rescaled to be between 0 and 1. The goal of the present work is to build and estimate models on the basis of the first three periods, and then to evaluate the *out-of-sample* performance of different forecasts of the peak load on the whole fourth period.

First, we define a sample formed by the daily curves of heat consumptions (we will refer to them as "load curves"), and consider a functional linear regression model where the peak demand on a given day is the scalar response and the load curve of the previous day is the functional regressor. Noting that intra-daily effects change with the season within each period, we then propose a methodology for improving the forecasting ability of the functional regression model: we partition the observational curves into homogeneous groups using the functional clustering technique presented by Abraham, Cornillon, Matzner-Lober, and Molinari (2003). The aim of our clustering procedure is to find some characteristic patterns in the data-set that may determine changes in the heat demand peaks. The classification of the load curves into groups is a solution to the problem of modeling the seasonality effect during each period. Then, we estimate a specific functional linear coefficient for each group, obtaining a family of functional regression models. Note that various clustering techniques for classifying similar load patterns have been considered in the literature; for instance, Chicco, Napoli, and Piglione (2001) and Amin-Naseri and Soroush (2006) refer to neural networks clustering; here we use clustering in a functional context instead. In order to assign the new curves to clusters in the forecasting procedure, we use a functional linear discriminant analysis, as was done by James and Hastie (2001). Finally, we compare the out-of-sample performance of our models with classical regression approaches in which the functional nature of the data is not taken