# A Functional Regression Approach for Prediction in a District-Heating System

Aldo Goia

# A Functional Regression Approach for Prediction in a District-Heating System

Aldo Goia

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

**Résumé**

Nous considérons le problème de la prédiction à court terme des pics de demande dans un système de chauffage urbain. Notre dataset consiste en quatre périodes séparées, avec 198 jours pour chaque période et 24 observations horaires dans chaque jour relatifs à la consommation de chaleur et le climat. Nous tenons en considération la nature fonctionnelle des données et proposons une méthodologie de prédiction basée sur la régression fonctionnelle. L'influence de variables explicatives exogènes est modelée d'une façon appropriée. Le résultats "out-of-sample" de l'approche proposée sont évalués.

**Abstract**

We consider the problem of short-term peak demand forecasting in a district heating system. Our dataset consists of four separated periods, with 198 days each period and 24 hourly observations within each day relative to heat consumption and climate. We take advantage of the functional nature of the data and we propose a forecasting methodology based on functional regression. The influence of exogenous explanatory variables is modelled in a suitable way. The out-of-sample performances of the proposed approach are evaluated.

**Mots clés**

Functional linear model, penalized splines estimation, peak load forecasting, district heating system

## Introduction

Among the activities of support in the coordination, maintenance and planning of an energy system, the prediction of the load demand is one of the most important. In particular, short-term forecasting, which is made within the 24 hours of the following day, and in special way the prevision of peaks of demand, plays a central role in guaranteeing an efficient generating capacity, maintaining the system stability.

In this work we consider the problem of modelling and predicting the peak of heat demand in a district heating (called also "teleheating") system. This consists in distributing the heat for residential and commercial requirements, via a network of insulated pipes. The dataset analyzed has been provided by AEM Turin Group, a municipal utility of the northern Italy city of Turin, which produces heat by means of cogeneration technology and distributes it, guaranteeing the heating to over a quarter of the town.

In the recent literature concerning load prediction in district-heating (see for example Dotzauer (2002) and Nielsen and Madsen (2006) for some applications and references), the algorithms employed are based on regression or time series models. They are often similar to the models used in the prediction of electrical-power loads (for a review see e.g. Weron (2006)). Weather factors are often used as major variables in predicting energy load and, among the others, the outdoor temperature is considered the most important factor in the short term forecasting.

These methodologies skip sometimes the fact that the data used are discretization points of curves: they are observed with high frequency and are very highly correlated, exhibiting some seasonality patterns. This fact has stimulated us to explore the functional approach (see e.g. Ferraty and Vieu (2006) and Ramsay and Silverman (2006) for a review). We propose here a linear model, combining real regressors with functional ones.

## Forecasting Peak Load

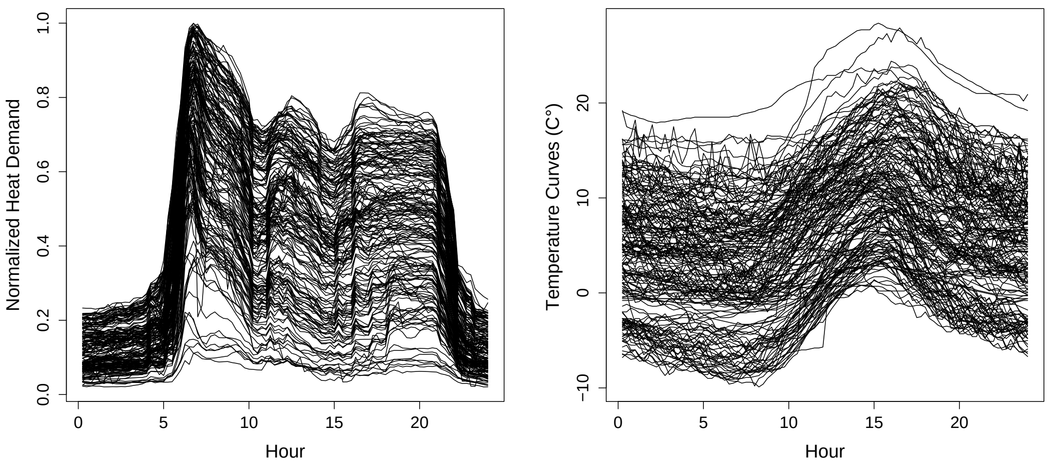
The dataset analyzed consists of measurements of heat consumption and temperature taken every hour, during the periods 15 October - 30 April in years 2001-02, 2002-03, 2003-04 and 2004-05. We take advantage of the functional nature of the data and we divide, in a natural way, the observed series of heat demand of each period in 198 *functional observations*, each one coincident with a specific daily load curve. We denote by $C_{y,d} = \{C_{y,d}(t),\ t \in [0, 24]\}$ the daily load curve of period $y$ and day $d$, with $y = 1, \ldots, 4$ and $d = 1, 2, \ldots, 198$. Each of these functional data is observed on a finite mesh of discrete times: $t_1 < t_2 < \ldots < t_{24}$ . Analogously we define the daily temperature curve. Figure 1 reports the observed loads and temperature curves of the first period.

Let us consider the forecasting problem of the daily peak load, defined as $P_{y,d} = \max_{j=1,\ldots,24} C_{y,d}(t_j)$. According to the literature (see e.g. Weron (2006)) we construct a linear model based on the decomposition of the load demand in a sum of two main components, namely the *load component* and the *weather-dependent component*, plus a stochastic residual. The first component includes:

- the seasonal effect, described by a suitable moving average of past daily means of consumptions;

- the intra-daily effect, modelled by a weighed sum of second derivative of the load curve of the previous day. A reason to consider second derivative rather than the

Figure 1: Normalized load and temperature daily curves in the period 15 October 2001 - 30 April 2002.




original curves is that data show an evident vertical shift and taking the second derivative annihilates this effect;

- calendar effects (week-days, weekend-days, holy-days).

About the weather-dependent part, we use the daily temperature curve, weighted by a suitable functional coefficient.

Combining in an additive way the components previously identified and described, we arrive to the specification of a linear model with scalar response (the peak of heat demand), two scalar regressors (the seasonal part and the dummy indicating the calendar effects), and two functional regressors (the second derivative of the past daily load curve and temperature curve).

The model is estimated on the base of the training-set corresponding to the data observed in the first three periods (2001-02, 2002-03 and 2003-04): we use here an estimation procedure proposed in Cardot *et al.* (2003) and based on B-splines. Then we carry out an out-of-sample forecasting study on the whole fourth period (2004-05), evaluating the results obtained. The estimated model fits well and the out-of-sample performances are good: we may compare them with the ones in Goia *et al.* (2010), where some functional and standard prediction methods are proposed to make forecasting on the same dataset.

# References

[1] Cardot, H., Ferraty, F. and Sarda, P. (2003). Spline estimators for the functional linear model, *Statistica Sinica*, 13, 571-591.

[2] Dotzauer, E. (2002). Simple model for prediction of loads in district-heating systems. *Applied Energy*. 73(3), 277-284.

[3] Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis*, Springer, New York.

[4] Goia, A., May, C. and Fusai, G. (2010). Functional clustering and linear regression for peak load forecasting. *International Journal of Forecasting*. Forthcoming.

[5] Nielsen, H.A., and Madsen, H. (2006). Modelling the heat consumption in district heating systems using a grey-box approach. *Energy and Buildings*, 38:63-71.

[6] Ramsay, J.O. and Silverman, B.W. (2006). *Functional data analysis*, Second Edition, Springer, New York.

[7] Weron, R. (2006). *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*, Wiley, Chichester.