ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA (ICAI)
Instituto de Investigación Tecnológica (IIT)

PHD THESIS

FUNCTIONAL TIME SERIES FORECASTING IN ELECTRICITY MARKETS: A NOVEL PARAMETRIC APPROACH

Author: José Portela González
Supervisors:
Prof. Dr. Antonio Muñoz San Roque
Prof. Dr. Estrella Alonso Pérez

MADRID
April 2017

This dissertation was typeset with LaTeX and compiled in TeXmaker using the MacTeX-2013 distribution. The font families used are Bitstream Charter, Utopia, Bookman, and Computer Modern. Unless otherwise noted, all figures were created by the author using Microsoft Visio®, Adobe Illustrator®, and MATLAB®.

Contents

List of Figures

List of Tables

1 Introduction

This first chapter introduces the motivation behind this thesis as well as its main objectives. In addition, a general overview of the organization can be found as well as the outline of the dissertation.

1.1. Motivation

Statistics is a branch of mathematics dealing with the collection, analysis, interpretation, presentation, and organization of observed data. These observations can be in the form of scalars, vectors or other objects. In particular, this dissertation is motivated by data consisting of functional observations, called functional data, where each observation takes infinite values over a continuum, which can be time, space location, wavelength, etc. It can be a function dependent on one or several variables, however, we will restrict the scope of this work to data consisting of curves. The observation of processes of continuous functions appear in nature, social sciences or industrial systems, evidencing the diversity of functional data that can be encountered.

Figure 1.1 shows two examples of functional data. On the left, female growth curves are shown[1]. Each curve is a functional observation which represents the evolution of a girl's height in the range from 0 to 18 years. This is a continuous curve because the height is defined as a function of time. Thus the set of observed curves could be expressed as the set $\{h_n(t)\}$ where $n = 1, 2, 3, 4$ and $t \in [0, 18]$ years. On the right, spectrometric curves are shown [2]. Four different pieces of raw meat are analyzed with a spectrometer, which gives the absorbance (as the $-log_{10}$ of the transmittance measured by the spectrometer) measured for each wavelength in the range 850 to 1050 nm. Therefore, an spectrometric curve is obtained for each piece of meat. This is again a continuous function, as the absorbance is defined for each possible wavelength in that

[1]Data obtained from http://www.functionaldata.org/
[2]Data obtained from [FV06]

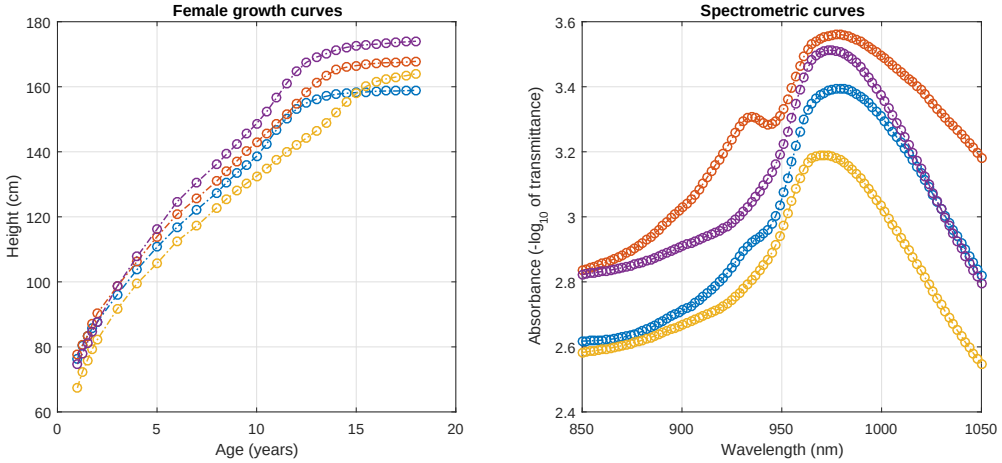


Figure 1.1. On the left, growth curves for four girls measured from the age 0 to 18. On the right, spectrometric curves, showing the absorbance measured at different wavelength for four pieces of raw meat.

range. Thus the set of observed curves could be expressed as the set $\{a_n(\lambda)\}$ where $n = 1, 2, 3, 4$ and $\lambda \in [850, 1050]$nm. It is worth mentioning that functional data are not always observed in its entirety. In some cases the curve is measured on discrete points over the continuum. For example, the spectrometric curves are recorded every 2nm of wavelength, or the height in the growth curves is measured at non-equidistant ages. Regardless the discretization procedure, the underlying function is continuous, and therefore, it is reasonable to approach the analysis from a functional point of view, considering observed functions as single entities, rather than just a sequence of point observations. It is considered that the data has an intrinsic structure that links each pair of adjacent points. The analysis of these processes of continuous functions require new statistical tools for extracting useful information.

Functional Data Analysis (FDA) is the statistical framework that provides the necessary tools for analyzing functional variables, where each observation is a continuous function. It is a relatively recent research field which has experienced a substantial growth in the past years. Among the different branches of FDA, we focus on functional time series (FTS), which are time sequences of functional observations. Two types of FTS can be considered. On the one hand, the FTS can be originated from a continuous process in time which is divided into segments of equal length, obtaining a time sequence of segments, i.e. a functional time series. An example of this type of process would be the evolution of the temperature at some location as a function of time, which is divided in daily segments, thus, obtaining a sequence of daily temperature profiles. Figure 1.2 illustrates this transformation from the continuous process (Part A) into the functional process (Part B). The other type of FTS is given when the observations are functions *per se* whose domain does not necessary have to be time. The former growth and spectrometric curves are examples of this type.

Functional time series can also be found in numerous forecasting applications in electricity markets. Hereafter, some examples are detailed, addressing the importance to market agents, System Operators or Market Operators:

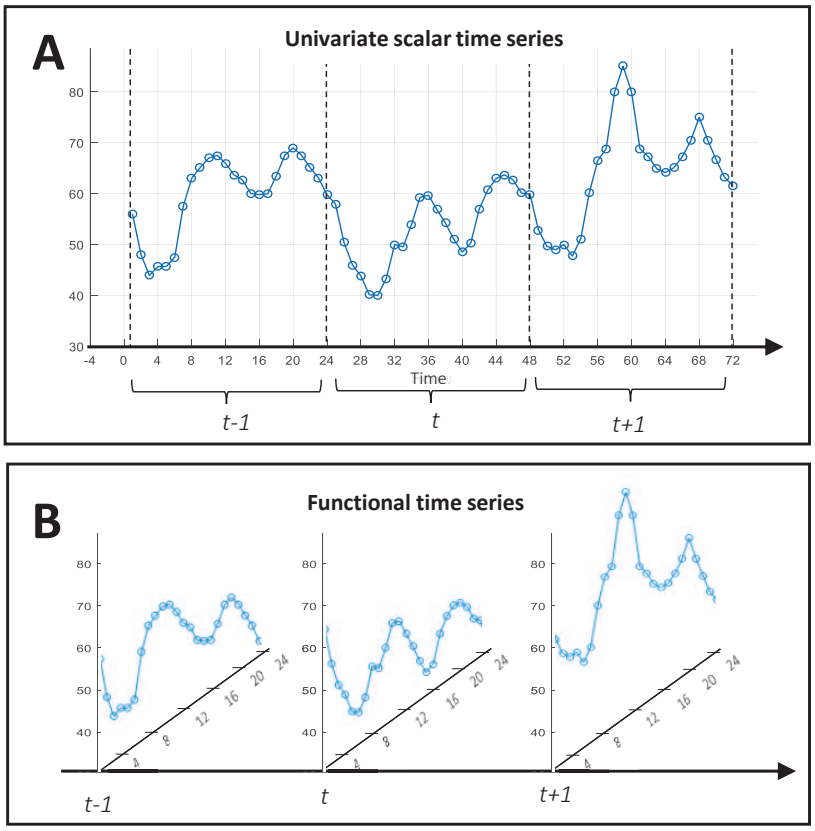


Figure 1.2. Scalar time series vs. functional time series. Part A. An example of a univariate scalar time series. When divided into segments of equal length, a functional time series is obtained. Part B. Functional time series. For each time t, a continuous function is observed.

- Electricity demand: The power consumption is a continuous-time process that can be analyzed with FDA methods. Short term load forecasting is crucial for market agents and System Operators. As the electricity cannot be stored, the demand has to be satisfied instantly and overproduction incurs in cost penalties. In addition, load forecasting allows for an efficient management of resources, optimal scheduling and production planning for minimizing generation costs. Moreover, distribution networks need continuous data for managing and balancing the electric network in real time. In the case of retailers, the demand forecast is of utmost importance for buying the necesary energy quantity at the lowest cost. Therefore, one of the interests of electricity companies is in forecasting the demand for the 24 hours of the next day. However, smart metering is increasing the sample frequency of electricity consumption from hourly values to data every half hour or every few minutes, as represented in Figure 1.3. Therefore, forecasting the demand for the next day is evolving to a continuous forecast. Some applications of FDA to electriciy load forecasting can be seen in [SL15] [Cha14], [PS13].

- Electricity prices: Everyday, the energy is traded in auction-like markets which yield a cleared energy and a cleared market price for each auction. Although the resulting time series of hourly prices is discrete, the sequence of daily price profiles can be analyzed as a functional time series. Forecasting electricity prices is an important matter when trading energy in the market as it allows for an optimum bidding strategy that maximizes

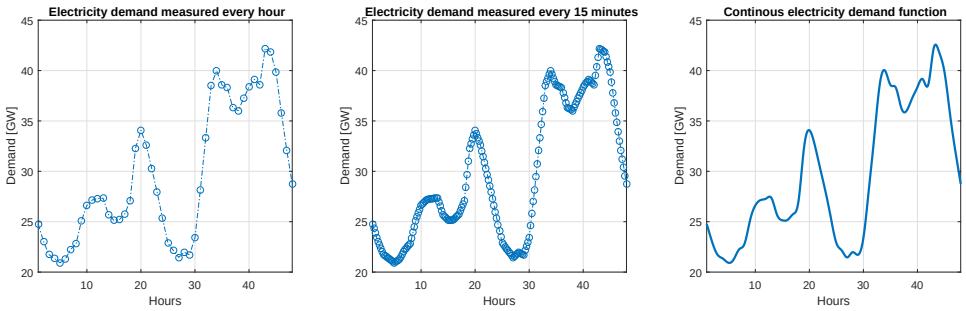


Figure 1.3. The effect of smart metering in the electricity demand. Traditionally, hourly measurements were available. With smart meters, the sampling rate approaches the continuous function.

profit. Some examples of price forecasting with a functional approach can be found in [Vil+12] and [Lie13]. In addition, the recent research and technological advances are leading to the possibility of creating continuous-time marginal pricing markets that can avoid the inefficiencies of having discrete time events for energy trading (see [PS16] and [PK16]). This effect is seen in the intraday continuous price mechanism introduced in EPEX[3]. Starting at 3.00pm on the current day, all hours of the following day can be traded until 30 minutes before delivery begins. This means that bids and offers are placed at any moment during that time period and are matched as soon as they are entered into the order book as opposed to waiting for gate closure. Consequently, there is not an intraday price for each hour, but a price profile which starts from 3.00pm of the former day up to 30 minutes before the clearing hour.

- Ancillary services: They are services and operations necessary to facilitate and support the continuous flow of electricity so that supply will always meet the demand. They generally include frequency control and reserves that provide additional energy when needed. This continuous management of resources and planning provides an application where FDA can be a meaningful tool.

- Offer Curves: In order to sell energy, the market agents have to submit offers to the market as price-quantity pairs, i.e. (p, q). It refers to the amount of energy q the agent is willing to sell at that price p. By sorting the selling offers in increasing prices, the aggregated offer curve for the agent is built, which represents the bidding behavior of the agent. Therefore, an offer curve is observed for each hour, i. e. an hourly functional time series is obtained. Figure 1.4 shows a visual example. It is worth noting that these functional data are observed in its entirety, as the entire offer function is fully defined for all the price range. Knowing in advance the shape of the offer curves of competitors allows to optimize the offering strategy, which motivates the developments of forecasting methods for supply functions as in [Pel13].

- Residual Demand Curves: When the need of optimizing the bidding strategy of an agent arises, the concept of residual demand curve appears as a mean of modeling the competitor's bidding behavior [Nie+14], [Por+16b]. A residual demand curve expresses the clearing price of the auction as a function of the amount of energy the agent is willing

[3]European Power Exange https://www.epexspot.com

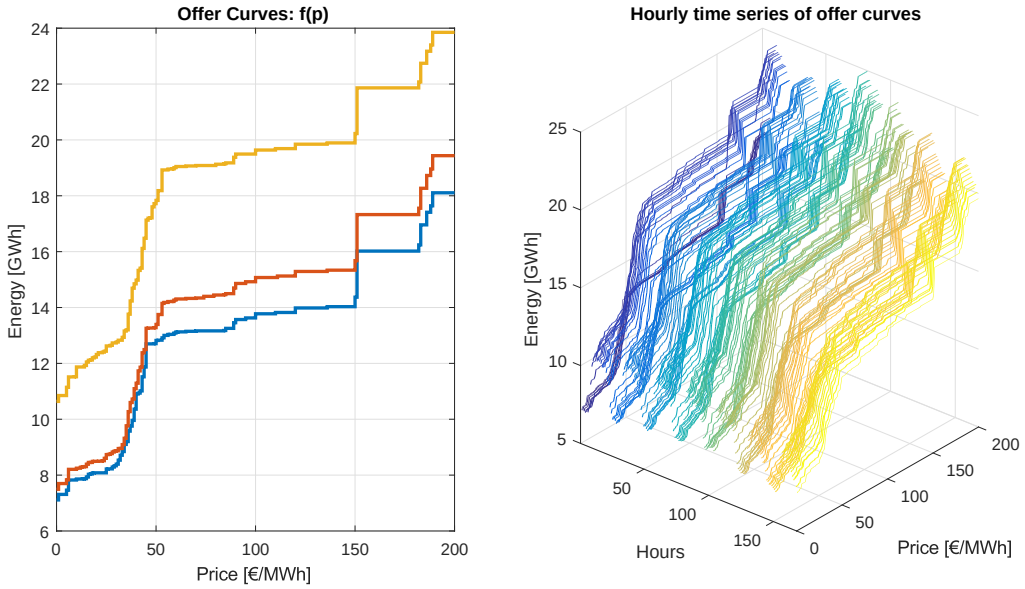


Figure 1.4. Offer curves in electricity markets. On the left, three offer curves defined in the price range 0 to 200 €/MWh are shown . On the right, the hourly sequence of offer curves is plotted. For each hour, a function defined in the price range 0 to 200 €/MWh is observed.

to buy or sell. Thus, similarly to the offer curves, a continuous function (curve) is obtained for each auction and retailers can use forecasts of the curves for optimizing their bidding strategy [Cam+16]. Some studies approach this forecasting problem using clustering techniques [Uge+03]. However, the functional approach fits in this application, as shown in [Ane+13], which makes use of functional nonparametric methods.

Being able to forecast these different time series can provide significant benefits to market agents, System Operators and Market Operators, thus, motivating the research of this dissertation. Any company involved in these markets will be interested in developing decision-making tools for optimal business and operation management. Therefore, forecasting the electricity demand, the price, offer curves or residual demand curves can provide a significant advantage over other competitors.

Many time series in electricity markets share some stylized facts. The effect of business and everyday activities lead to weekly and daily seasonalities as well as peak and low demand hours. In addition, explanatory variables are important drivers of some time series such as the electricity prices. For example, weather (wind speed, precipitation, etc.) affects the production of renewal technologies with lower generation costs influencing the offering behavior of the agents. Consequently, the development of functional forecasting models for electricity markets should take into account these properties by considering the effect of exogenous variables and seasonality among others.

The literature on functional time series forecasting provides some reference models which could be applied in electricity markets. The different approaches can be classified into parametric, dimensionality reduction and nonparametric models. Functional parametric

models make use of linear functional operators to operate with the curves. These are powerful tools as the operators allow to produce forecasts using the entirety of the input curve. However, estimating these operators is challenging and causes that most of the available models have a simple structure, i.e. seasonality is not taken into account. The dimensionality reduction approach represents the FTS in a lower dimensional space to further apply classical multivariate statistics to the reduced components. The advantage is that powerful multivariate models could be applied. However, reducing the dimensionality might incur in loss of precision. Finally, nonparametric models use nonlinear functional operators, which are usually focused on averaging historical curves from similar situations as the moment that it is being forecasted.

This dissertation contributes to the current state of the art by proposing a novel parametric functional time series model that can be used successfully in electricity markets' applications.The first original contribution is the definition of a new procedure for estimating functional linear operators. Each operator is defined by a bivariate kernel which, in our approach, is modeled as a finite sum of sigmoid functions. This allows for the extension of the standard seasonal ARIMAX model to the functional framework. Hence, the SARIMAHX model is developed, which significantly improves parametric models by accounting for seasonal autoregressive and moving average effects as well as accounting for the effects of scalar and functional explanatory variables.

1.2. Thesis objectives

The main objective of this dissertation is to develop a new parametric functional forecasting method that can provide competent functional forecasts when applied to functional time series in electricity markets. The particular features of these series require that the proposed functional model ought to consider the effects of exogenous variables and seasonality.

The specific objectives that are to be developed in the dissertation and that will become the main contributions are summarized:

- Provide a taxonomy of forecasting models for functional time series. The aim is to discuss the theoretical basis behind each formulation and evaluate its applicability to electricity markets.

- Propose a new functional forecasting method suited for forecasting functional time series in electricity markets. The new model should improve those features missing in literature models and obtain better estimates. It should be able to take into account explanatory variables, seasonality and complex time dependencies.

- Validate the approach by applying the proposed model to datasets from different electricity markets and compare the results against other reference models. These case studies should test the models with data of different applications to validate the versatility of the proposed model.

1.3. Dissertation outline

The objectives previously described are unfold in this document in the form of five chapters. After this introductory chapter, Chapter 2 provides the essential background of FDA. Its aim is twofold. Firstly, it introduces the main concepts and tools in Functional Data Analysis. This includes a formal definition of functional data, its representation using basis functions, the definition of norms, the extension of summary statistics (mean, variance, correlation...) to the functional framework and an introduction of operators in Hilbert spaces for operating with functions. Moreover, the Functional Linear Model is presented, which is the natural extension of multivariate regression to the functional setting. The aforementioned tools are the essential elements needed for analyzing functional data and provide the statistical context for deriving complex functional models. The second part of Chapter 2 is centered in functional time series, providing a formal definition as well as an extensive review of the main functional forecasting methods found in the literature.

Chapter 3 is devoted to the development of the forecasting models proposed in this dissertation. Firstly, a functional regression model is presented, which introduces a novel estimation method for functional linear operators. Then, the functional seasonal ARIMAHX model is presented. Two versions of the SARIMAHX model are proposed, both using a different type of linear operator. Simulation analyses are carried out for each model to test the estimation procedure in a controlled environment.

Chapter 4 presents the different empirical case studies where the proposed model is validated. Specifically, four different electricity market applications are selected where the SARIMAHX model is tested and compared with other functional reference models:

- Firstly, FDA is applied to the Ancillary Services market. This case study forecasts the hourly total amount of energy assigned in the tertiary and deviation management markets in the Spanish electricity market. It is set up as a pure functional linear regression problem, which allows to compare the proposed functional regression model with the standard method in the literature.
- Secondly, the problem of forecasting the electricity market price is addressed. In this case study, the Spanish and German day-ahead daily electricity price profiles are forecasted. The goal is to compute one-step horizon forecasts using explanatory variables as functional covariates and accounting for the weekly seasonality of the series. The proposed model is compared with the reference parametric and dimensionality reduction models. This chapter also includes a small test comparing the proposed functional model with some other non-functional reference forecasting models for electricity prices.
- Thirdly, forecasting offer curves is addressed. The aggregated supply function of the competitors of Enel in the Italian electricity market is forecasted. This is an hourly FTS, therefore, forecasting models will include double seasonal (daily and weekly) terms and will account for scalar explanatory variables. One-step as well as 24-step forecasts are considered, illustrating the differences encountered.
- Fourthly, the case of Residual Demand Curves is addressed. The goal in this case study is to forecast the hourly RDC of the market for the following auctions. However, the

concept of RDC entails an important discussion. When RDCs are used for optimizing the bidding strategy of an agent, it is of utmost importance that this curve represents the best as possible what would be the outcome of the market for different bids of the agent. The traditional way of computing RDCs does not take into account the non-linear effects that complex offering conditions have in the RDC. Therefore, the chapter is divided in two main sections. The first section covers the extensive research performed in order to achieve a better representation of the RDC. The second section is devoted to the forecasting problem where the newly defined RDCs are forecasted with functional models including explanatory variables as well as double seasonality. Moreover, one-step as well as 24-step forecasts are considered.

Finally, Chapter 5 provides the concluding remarks of the dissertation, summarizing the contributions and future developments.

2

Forecasting Functional Time Series

This chapter provides the necessary background for supporting the main contributions of this thesis. Firstly, functional data are introduced, illustrating the practical application of this new statistical framework. Then, the mathematical fundamentals of Functional Data Analysis is covered, including a description of the essential tools needed for analyzing functional variables. Finally, functional time series are elucidated including a summary description of different functional methods developed for forecasting functional time series.

2.1. Introduction

Statistics is concerned with the analysis of data obtained from observations of random variables. In classical multivariate analysis, these random variables x are real-valued elements, i.e. $x \in \mathbb{R}$. A random variable x has a probability distribution which specifies the probability that its value falls in any given interval. Moreover, that probability distribution can be modeled by its cumulative distribution function, i.e. $F_x(a) = P(x \leq a)$. Different moments can be computed to characterize the probability distribution, e.g. mean, variance, etc.

The data that motivate this dissertation are observed in the form of curves, i.e. each observation is a real-valued function that takes values in an infinite set. Therefore, a functional random variable X is defined as the function $X = \{X(v); v \in V\}$. In this dissertation, we restrict the analysis to functional random variables associated with an interval $V \subset \mathbb{R}$, nevertheless, the notion of functional variable covers a larger area, e.g. random surfaces when $V \subset \mathbb{R}^2$. Finally, a functional stochastic process is defined as a collection of functional random variables ordered in time.

Functional data can be found in numerous real life applications from different fields of applied sciences. The study of demographic curves (mortality and fertility rates for each age observed every year) can be found in [HF10], the analysis of the evolution of ozone concentration at each location is seen in [DG02], the study of sulfur dioxide levels is addressed in [DC+05] and traffic flow prediction in [Chi12]. In addition, the application to the medical field is wide as shown by the review in [UF13]. .The advances of modern technology has improved the data collection, processing and storage capabilities leading to the possibility of recording these continuous processes at increasingly high frequency. Consequently, it has increased the need for accounting for these new high dimensional data.

Functional data can be classified according to the nature of the underlying continuous process [HE15].

- The random element perspective assumes the functional data are recordings of random functional variables. Some examples are absortion spectrometric data for each object as seen in figure 1.1 or demographic functional times series [HS09], where $Y_t(v)$ is the mortality rate at age v measured at year t. This is also the case of Residual Demand Curves where for each hourly auction t, a curve $R_t(q)$ is observed which represents the clearing market price as a function of the energy q sold in the market.
- The stochastic process perspective assumes that functional time series can be originated from a univariate scalar continuous-time process. Then, by dividing the series into segments of equal length, a functional time series is obtained, i.e. a time sequence of segments. This is the case of temperature measurements in a specific location, the study of ozone concentration as a function of time [DG02] or the electricity demand.

In order to analyze this new type of data, Functional Data Analysis (FDA) is a new field of statistics that has arisen in the last two decades in order to encompass the statistical methodology for such data. Broadly interpreted, FDA deals with the analysis and theory of functional random variables. Some comprehensive references are found in (Ramsay and Silverman, 2005) which explores functional data covering the basic techniques to analyze functional variables: smoothing, basis function expansion, principal component analysis, canonical correlation, regression analysis and differential analysis. On the other hand, [FV06] approaches the statistical analysis using nonparametric techniques. [HK12] summarizes different advances in statistical inference for functional data. In addition, [HE15] gives deep theoretical foundations of FDA. The reference book for linear processes in functional data analysis is [Bos00], establishing properties of functional time series, as well as numerous developments for the functional autoregressive process.

Specifically, this thesis will cover the study of a particular type of data in FDA which are functional time series (FTS). A FTS can be defined as a time sequence of functional observations i.e. the realization of a functional stochastic process in time. Consequently, most of the concepts from FDA are also valid for analyzing FTS.

The following sections aim at describing the necessary background on FDA and the current FTS forecasting state of the art. Therefore, the rest of the chapter is outlined as follows: section 2.2 detail the theory of the mathematical approach for operating with functional data and summarizes the main tools that are available for analyzing functional data. Section 2.3 is

devoted to the particular case in FDA which are functional time series. Firstly, some definitions and assumptions regarding FTS are introduced which then lead to the introduction of the different approaches for modeling FTS.

2.2. Tools for functional data

The goals of FDA are to provide tools for statically inferring useful information on functional variables. This includes representing the data in ways that help further analyses, the study of the sources of variation among the data, comparing the similarities or differences of two or more functional variables and forecasting time series of functional data. This section covers the main tools of FDA that help to achieve those goals. The tools commonly used in multivariate statistics are not suitable for the study of infinitely dimensional data; therefore, they have to be redefined to account for the particularities of this framework. This chapter follows [Ram06], [FV06], [HK12], [HE15], [Bos00] and the survey [Wan+15]

Let us start by introducing the mathematical framework of FDA. A functional random variable X is a real-valued function $X(v)$ considered as an element of the Hilbert space L^2 of real square integrable functions defined on a closed interval $v \in [a, b]$. Therefore, X belongs to an infinite dimensional space $L^2_{[a,b]}$, satisfying $\int_a^b X_t^2(v)\, dv < \infty$. From now on, it is considered that an integral sign without the limits of integration denotes the integral over the whole interval of the domain where the functions are defined.

2.2.1. Representing functional data with basis functions

Each functional variable can be represented as a linear combination of known basis functions. A functional variable X belongs to an infinite dimensional space L^2 and, as a consequence, an infinite number of basis functions are needed for representing the functional variable with zero error. Hence, all the realizations $\{X_n\}_{n\in\mathbb{Z}}$ of a functional random variable X can be expressed in relation to the same functional basis as:

$$X_n(v) = \sum_{k=1}^{\infty} c_{n,k} e_k(v),$$

where $\{e_k(v)\}_{k=1}^{\infty}$ are the basis functions and $\{c_{n,k}\}$ are the coefficients or coordinates that represent each observation in the selected basis. In practice, working with an infinite basis is not feasible, so the usual approach is to truncate the number of basis functions to a finite number K, hence, the following approximation holds

$$X_n(v) \approx \sum_{k=1}^{K} c_{n,k} e_k(v).$$

Therefore, each observation $X_n(v)$ is represented by a vector of finite scores $c_{n,1}, \ldots, c_{n,K}$. Thus, the dimensionality of the curves has been reduced from infinite to K.

The choice of the basis system is critical. The number of basis functions K will impact the performance of some procedures. The smaller K, the less computation is required but the accuracy for representing the functions can be reduced and thus it may incur in a significant

loss of information. Usually, the selected basis depends on the application that is being tackled. [Ram06] dedicates several sections on this topic, and a variety of functional basis are used in the literature. Some examples are Fourier, B-splines [AM+13] and wavelets [Ant+07]. Nevertheless, a reference basis commonly used is the Functional Principal Component (FPC) basis, which deserves a more detailed explanation in section 2.2.5. The advantages of the FPC basis is that the coordinates are good descriptors of the data, thus it allows selecting a low number of basis functions while accounting for much of the variability of the original data.

Basis expansion goes beyond reducing to a multivariate setting. A representation on a functional basis is usually used for purposes such as smoothing the source data, or when the measurements of the observed functions are sampled at uneven values of the domain.

2.2.2. Norm and distances

Being able to measure distances between functional data is of utmost importance. Consequently, the definition of a norm that measures tge distance between two elements is an essential tool in FDA. In the L^2 Hilbert space, the following inner product $\langle \cdot, \cdot \rangle$ is defined for the functional variables X, Y

$$\langle X, Y \rangle = \int X(v)Y(v)dv.$$

The inner product generates the following L_2 norm $\| \cdot \|$:

$$\|X\| = \sqrt{\langle X, X \rangle}.$$

It is worth noting that a function e with $\|e\| = 1$ can be considered as a direction in the functional space and thus, $\langle X, e \rangle$ can be viewed as projecting the curve X over the direction given by e. Then, the Parseval's equality holds, which ensures that for a countable orthonormal functional basis $\{e_k\}_{k \in \mathbb{N}}$,

$$\langle X, Y \rangle = \sum_{k=1}^{\infty} \langle X, e_k \rangle \langle Y, e_k \rangle = \sum_{k=1}^{\infty} c_{X,k} c_{Y,k}$$

being $c_{X,k} = \langle X, e_k \rangle$ and $c_{Y,k} = \langle Y, e_k \rangle$ the scores of X and Y with respect to the basis function e_k.

A measure of distance between two curves $d(\cdot, \cdot)$ is derived from the inner product and norm defined:

$$d(X_1, X_2) = \|X_1 - X_2\| = \sqrt{\langle X_1 - X_2, X_1 - X_2 \rangle} = \sqrt{\int (X_1(v) - X_2(v))^2 dv}.$$

The equality $X_1 = X_2$ implies $d(X_1, X_2) = \|X_1 - X_2\| = 0$. In addition, it is worth noting that if a finite basis system e_k is chosen to represent the functions, the distance becomes a semi-metric, as it is a truncated norm:

$$d(X_1, X_2) = \sqrt{\int \left(\sum_{k=1}^{K} (c_{1,k} - c_{2,k})\, e_k(v) \right)^2 dv}$$

where $c_{1,k}$ and $c_{2,k}$ are the coordinates of X_1 and X_2 in the e_k basis.

This definition of distance between functional data will play a crucial role in non-parametric models [FV06], and allow developments on outlier detection techniques as seen in [AGR14], [HF10], [Rañ+15], [Tim+11] as well as depth measurements for functional data.

The inner product and the corresponding norm also allows to define the angle θ between two elements X_1 and X_2 given by

$$\cos(\theta) = \frac{\langle X_1, X_2 \rangle}{\|X_1\| \|X_2\|}$$

which can be useful for defining properties such as orthogonality i.e. elements X_1 and X_2 are said to be orthogonal if $\langle X_1, X_2 \rangle = 0$.

2.2.3. Summary statistics

Given a functional variable X, it is necessary to define its main statistical properties. These include the definition of the mean value of the functional variable, the variance and the correlation.

The mean function $\mu_X(v)$ of the functional variable X is the function such that $\mu_X(v) = E[X(v)]$ where $E[\cdot]$ is the expectation, which can be estimated through the observed data $\{X_n\}_{n\in\mathbb{N}}$. Hence, the sample mean is

$$\hat{\mu}_X(v) = N^{-1} \sum_{n=1}^{N} X_n(v)$$

which is the point-wise average of the functional observations. Similarly, the variance is a measure of the deviations from the mean, i.e. $var_X(v) = E[(X(v) - \mu_X(v))^2]$. Its sample counterpart is computed as

$$\widehat{var}_X(v) = N^{-1} \sum_{n=1}^{N} \left(X_n(v) - \hat{\mu}_X(v)\right)^2 .$$

Naturally the standard deviation function is the square root of the variance function.

Some examples of the sample mean and variance functions are shown to illustrate these concepts. Figure 2.1 shows spectrometric data, which are absorbance measurements at each wavelength for different pieces of raw meat. Figure 2.2 shows the yearly precipitation profiles for different weather stations in New Zealand, and figure 2.3 shows the yearly temperature profile for the same New Zealand weather stations.

The former variance only provides information about the point-wise variability of the curves for equal values of v. However, variability can also be measured across values of each curve. Then, the covariance function of the functional variable X is defined as $\Sigma_X(v,u) = E[(X(v) - \mu_X(v))(X(u) - \mu_X(u))]$. Each value $\Sigma_X(v,u)$ measures the joint variability of the functional variable at different points of the curve v and u. That is to say, it accounts for the

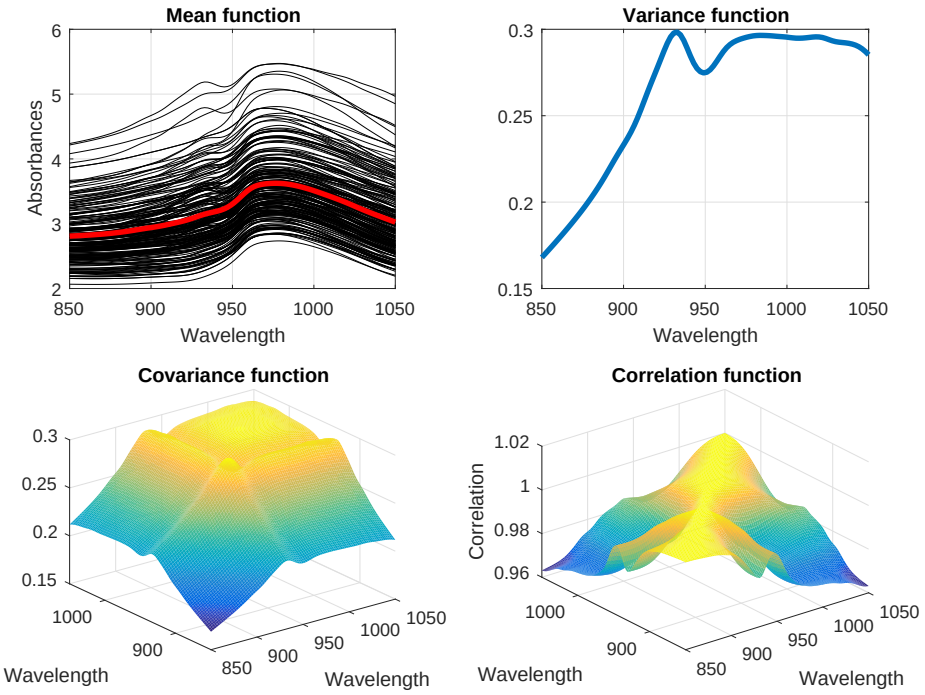


Figure 2.1. Summary statistics for Spectrometric data (absorbance measurements at each wavelength for different pieces of raw meat). The sample mean and variance functions as well as the sample covariance and correlation surfaces are shown

joint deviations from the mean of v and u. The sample autocovariance function is obtained by

$$\hat{\Sigma}_X(v,u) = N^{-1}\sum_{n=1}^{N}\left(X_n(v) - \hat{\mu}_X(v)\right)\left(X_n(u) - \hat{\mu}_X(u)\right).$$

The covariance function is one of the key tools for functional data analysis as it is the essential basis for many methods such as the extraction of Functional Principal Components. Several studies can be seen in [Yao+05a] [LH+10] which also proposes variations for computing smooths estimates of $\hat{\Sigma}(v,u)$.

Correlation in FDA is not a straight-away extension of scalar methods as for the variance and covariance counterparts. See [He+00] for a study on this issue. Several approaches to correlations can be found. Firstly, an autocorrelation measurement can be defined for a functional variable X as

$$\widehat{corr}_X(v,u) = \frac{\hat{\Sigma}_X(v,u)}{\sqrt{\widehat{var}_X(v)\widehat{var}_X(u)}}.$$

This correlation accounts for the point-wise correlation existing between v and u domain values of the observations in X. For example, Figure 2.1 shows the correlation surface of spectrometric data. As can be seen, the absortion for very different wavelengths are less correlated than the absortion values for similar wavelengths. Figure 2.2 shows that Winter months are highly correlated with themselves (higher correlation in the corners) while summer months are much less correlated with the rest of the months.

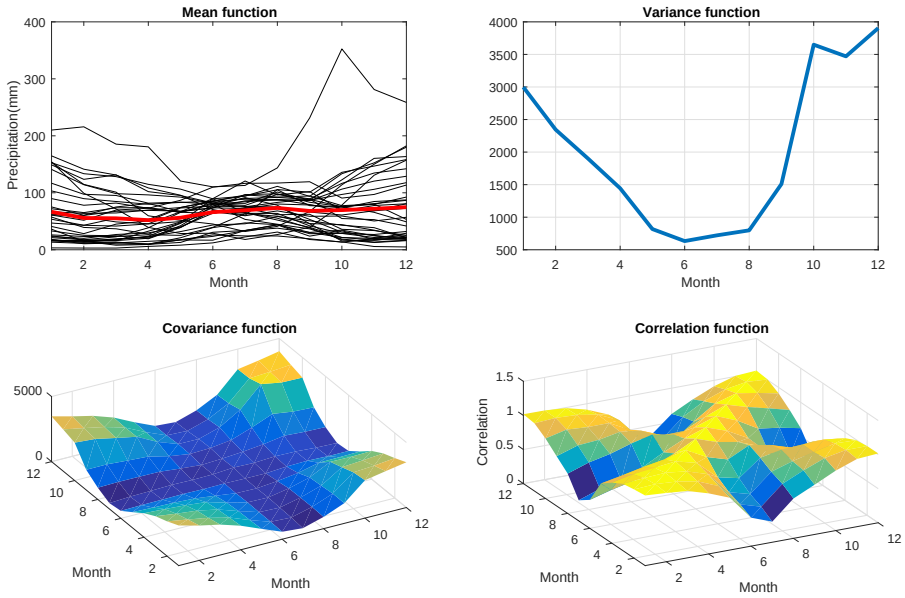


Figure 2.2. Summary statistics for the yearly precipitation profiles for different weather stations in New Zealand. The sample mean and variance functions as well as the sample covariance and correlation surfaces are shown

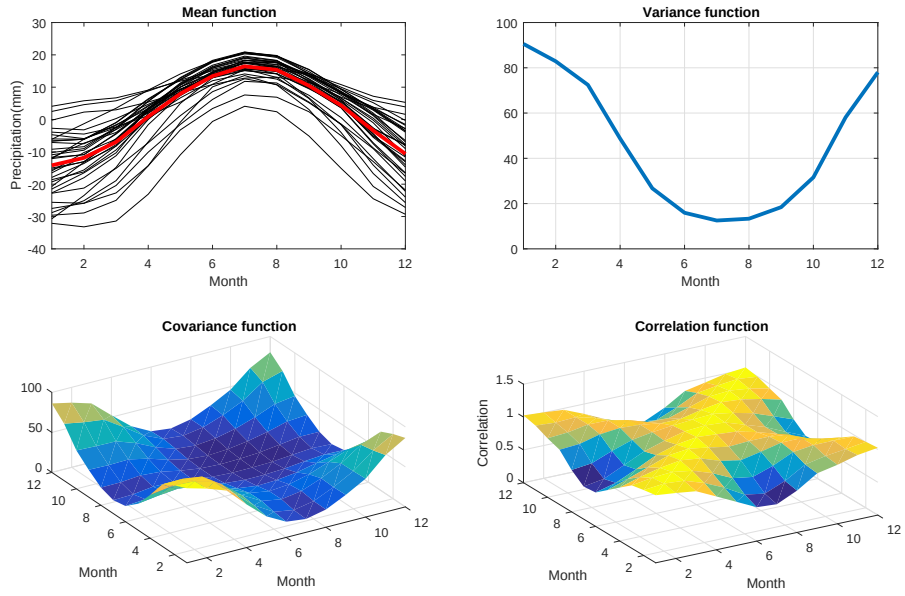


Figure 2.3. Summary statistics for the yearly temperature profiles for different weather stations in New Zealand. The sample mean and variance functions as well as the sample covariance and correlation surfaces are shown

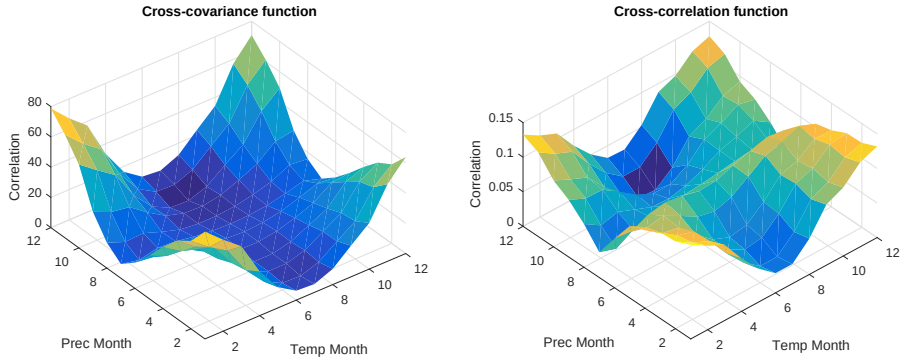


Figure 2.4. Cross-covariance and cross-correlation for precipitation data based on temperature data

In addition, covariance and correlation functions can also be defined between two functional variables. The covariance function between X and Y is defined as $\Sigma_{XY}(v,u) = E[(X(v) - \mu_X(v))(Y(u) - \mu_Y(u))]$. Therefore, the sample cross-covariance between two functional variables is obtained as

$$\hat{\Sigma}_{XY}(v,u) = N^{-1} \sum_{n=1}^{N} \left(X_n(v) - \hat{\mu}_X(v)\right) \left(Y_n(u) - \hat{\mu}_Y(u)\right).$$

This shows the joint variability between points of the two functional variables. It is worth noting that this covariance functions still holds when the domain ranges of X and Y are different. Finally, the cross-correlation function

$$\widehat{corr}_{XY}(v,u) = \frac{\hat{\Sigma}_{XY}(v,u)}{\sqrt{\widehat{var}_X(v)\widehat{var}_Y(u)}}.$$

Figure 2.4 show the cross-covariance and cross-correlation functions between the yearly temperature profile and the yearly precipitation profile datasets. As can be seen, correlation is higher in the "corners" of the surface. This means that the precipitation on the first months of the year are correlated with the temperature of the last months of the year, and vice versa. During summer months, however precipitation and temperatures are less correlated.

The cross-correlation function accounts for the point-wise correlation but other methods to assess functional correlation are found. Canonical Correlation Analysis [Ram06; HK12; He+00] finds the linear transformation such that the original functional variables X and Y are mapped into two subspaces such that the correlation between the transformed series is maximized. [DM05] defines some dynamical correlation to account for derivatives of he curves and a unified correlation for the different domain values. [GM12] attempt to simplify the correlation surface by defining an stickiness coefficient and [HR16] define a maximal autocorrelation approach.

It is worth noting that for uneven sampled functional data observations, the mean, the variance, the covariance and correlation estimates can be calculated in all the domain straight-away by previously expanding the data into a functional basis and working with the reconstructed curves. Otherwise, non-parametric estimators should be used. See [Wan+15] and references therein.

2.2.4. Operators in Hilbert spaces

Operators are the basic mathematical tool to operate with functional data. This section is devoted to linear transformations in the L^2 space. Different operators can be defined, but we will focus on Hilbert-Smidth linear operators and a brief description of the main theoretical properties is presented.

A continuous and bounded linear operator Ψ is said to be compact if there exists two orthonormal basis $\{e_k\}$ and $\{f_k\}$ and a sequence of real λ_k converging to zero such that

$$\Psi(X_n)(v) = \sum_{k=1}^{\infty} \lambda_k \langle X_n, e_k \rangle f_k \qquad X_n \in L^2$$

This representation is called the singular value decomposition. If $\sum_{k=1}^{\infty} \lambda_k^2 < \infty$, the operator is said to be Hilbert-Schmidt. Considering the space S of Hilbert-Schmidt operators the following scalar product is defined

$$\langle \Psi_1, \Psi_2 \rangle_S = \sum_{i=1}^{\infty} \langle \Psi_1(e_i), \Psi_2(e_i) \rangle$$

where $\{e_i\}$ is an arbitrary orthonormal basis. Hence, the norm of Ψ is $\|\Psi\|_S = \sqrt{\langle \Psi, \Psi \rangle_S}$, which can be shown that $\|\Psi\|_S^2 = \sum_{k=1}^{\infty} \lambda_k^2$.

An important class of linear operators in L^2 are the integral operators defined by:

$$\Psi(X_n)(v) = \int \psi(v,u) X_n(u) du \qquad X_n \in L^2$$

where $\psi(v,u)$ is the kernel of the operator. Such operators are Hilbert-Shmidt if

$$\iint \psi^2(v,u) dv du < \infty.$$

In that case, the norm of the operator becomes

$$\|\Psi\|_S = \sqrt{\iint \psi^2(v,u) dv du}.$$

If $\psi(u,v) = \psi(v,u)$ and $\iint \psi(v,u)x(v)x(u)dudv \geq 0$ becomes a symmetric and positive-definite operator so it admits the decomposition:

$$\Psi(X_n) = \sum_{i=1}^{\infty} \lambda_i \langle x, e_i \rangle e_i \qquad X_n \in L^2 \tag{2.1}$$

where e_i are orthonormal eigenfunctions and λ_i are eigenvalues which satisfy $\Psi(e_i) = \lambda_i e_i$. This is known as the Mercer's theorem [RS90]. This allows the kernel function $\psi(s,u)$ to be expressed as:

$$\psi(v,u) = \sum_{i=1}^{\infty} \lambda_i e_i(u) e_i(v)$$

Figure 2.5 shows a visual representation of a functional integral operator.

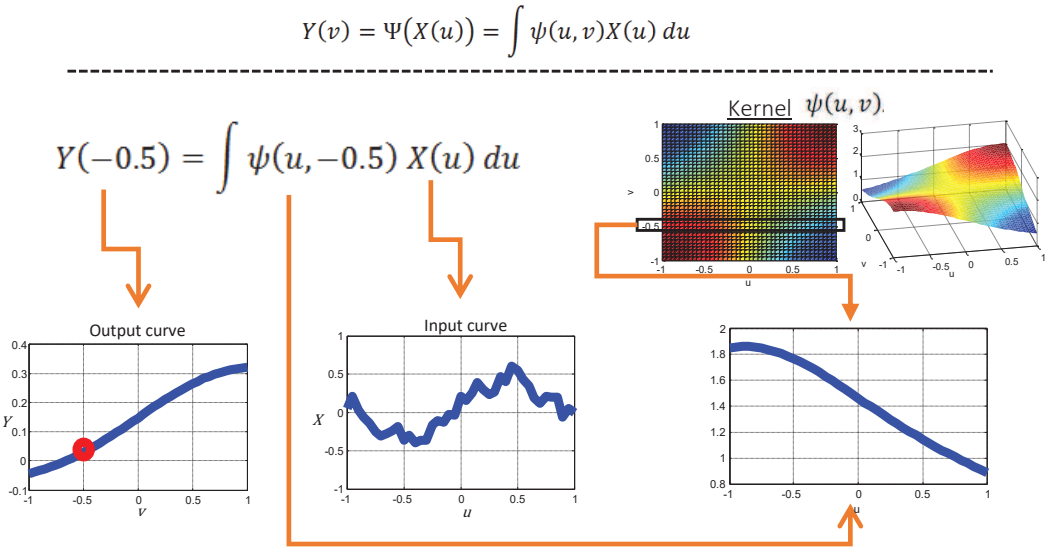


Figure 2.5. Visual representation of a functional integral operator. In the example, the output curve Y at point (-0.5) is the result of the integral of the multiplication of the input X and the slice of the operator at -0.5.

2.2.5. Covariance operator and functional Principal Components

Principal Component Analysis [Jol02] has proven to be a key dimensionality reduction tool in multivariate analysis. This technique has been extended to functional data in numerous works such as [Yao+05b; Hal+06], which have approached the definition, calculation and applications of the functional principal components (FPC). Functional Principal Component Analysis (FPCA) has become the most prevalent tool in FDA because it facilitates the conversion of the infinite-dimensional functional data to a reduced set of scores accounting for the great majority of the variability of the data. The conversion is achieved by projecting the original data into a functional basis spanned by the functional principal components, which are the eigenfunctions of the autocovariance operator of the functional variable X.

The autocovariance operator C is defined as $C(\cdot) = E\left[\langle X, \cdot \rangle X\right]$. More precisely, it is an integral operator whose kernel is the autocovariance function $\Sigma_X(u,v)$:

$$C(Y)(v) = \int \Sigma_X(u,v)Y(u)du \qquad Y \in L^2$$

Because of the covariance function properties, the covariance operator falls into the class of symmetric and positive definite Hilbert-Shmidt operators, thus admits the decomposition in eigenfunctions shown in equation 2.1. Hence, the kernel of the covariance operator (i.e. $\Sigma_X(u,v)$) can be expressed as:

$$\Sigma_X(u,v) = \sum_{j=1}^{\infty} \lambda_i e_i(u) e_i(v)$$

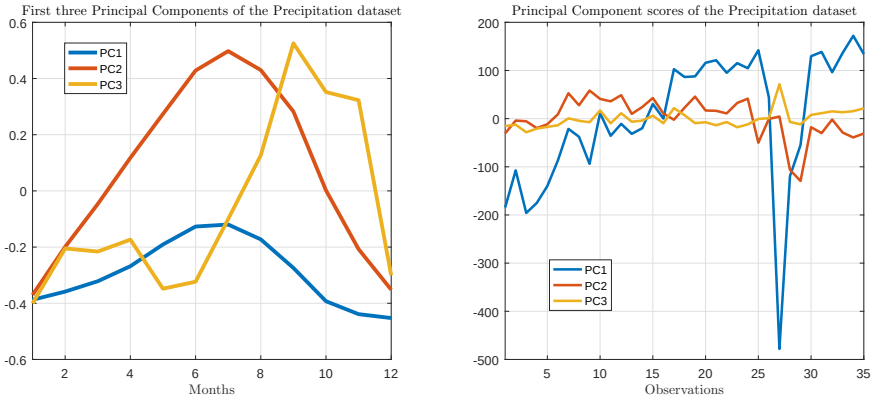


Figure 2.6. Principal Components and scores obtained for the Precipitation dataset.

where the orthogonal eigenfunctions e_i are called the principal components and the eigenvalues λ_i represent the variability explained by the corresponding principal component. We will assume that the λ_i are ordered decreasingly, providing a meaningful interpretation of the principal components. Each e_i represent the direction in that subspace that accounts for λ_i variance of X. Therefore, the first principal component is the direction which explains the most variability of the data.

Every observation $\{X_n\}$ can be represented as a linear combination of the eigenfunctions obtained, which is known as the Karhunen-Loéve representation.

$$X_n(v) = \sum_{k=1}^{\infty} \langle X_n, e_k \rangle \, e_k$$

The scalar products $\langle X_n, e_k \rangle$ are called the component scores (or coordinates of the curves in the FPC basis) and have zero mean, are uncorrelated, and have variance equal to λ_k.

Figure 2.6 shows the first three Principal Components obtained for the Precipitation dataset. These components were extracted from the covariance function seen in Figure 2.2. All the values of the first PC are always negative, meaning that it accounts for the level of the precipitation, being the winter months the one that have a greater variability than summer months. In addition, the sequence of PC scores for each component is plotted. These represent the projected coordinates of each curve.

As commented, one of the main purposes of FPCA is to reduce the dimensionality of the series from infinite to a finite number K of components. Then, an essential discussion is the choice of K. It depends on the compromise between the number of components and the percent of variance explained by the selected eigenfunctions. This is known as the Cumulative Percentage of total Variance (CPV) method which is defined as:

$$CPV(k) = \sum_{i=1}^{k} \frac{\lambda_i}{\sum_{j=1}^{\infty} \lambda_j}$$

assuming the λ_i are ordered in decreasing values, the K is chosen so that $CPV(k)$ explains most of the variability of the data without needing a high number of components.

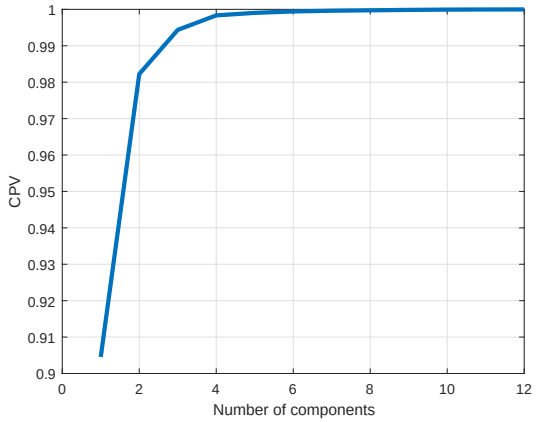


Figure 2.7. Cumulative Variance Explained for the different number of PC extracted in the Precipitation dataset.

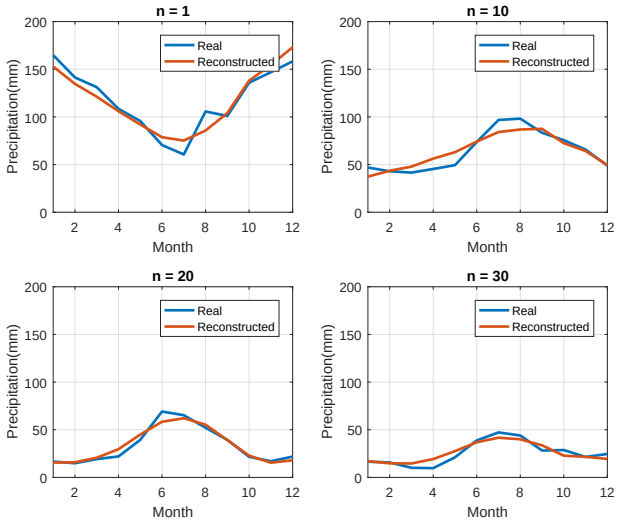


Figure 2.8. Real and reconstructed precipitation profiles with three PC. Four different observations are shown $n = 1, 10, 20, 30$

Figure 2.7 shows the CPV curve. For the different number of components extracted, the cumulative variance explained is shown. Extracting three components is seems a reasonable trade off between complexity and accuracy.

Figure 2.8 shows a comparison of four real precipitation profiles and the corresponding reconstructed curves with three PC. As can be seen, even though these three components account for the 99.44% of the variability of the data, there are some areas which are not modeled by this approximation.

2.2.6. Functional linear regression

Functional regression is the extension of the standard multivariate linear regression to functional data. We will summarize the main approaches for this problem although it is still an active area of research. Without loss of generality, we consider the functional variables X and Y with zero

mean. Different methods can be found depending on the nature of the predictor and response variables:

- Functional covariates and scalar response: In this model, a scalar variable y_n is assumed to be dependent on a functional variable X_n. Then, the model has the following representation:

$$y_n = \langle X_n, \beta \rangle + \epsilon_n = \int X_n(u)\beta(u)du + \epsilon_n$$

  where ϵ_n is a zero mean finite variance random scalar error and β is the coefficient function that has to be estimated from the data. This model has been studied extensively [Car+99; Car+03; HH+07]. The standard approach is to expand the covariate X and the coefficient function β in the same functional basis, such as the B-spline basis or an eigenbasis, and then optimizing the coordinates in that subspace by least squares. It should be noted that if an eigenbasis is used, the model becomes a standard multivariate regression problem.

- Functional covariates and functional response: This model relates the functional variable Y defined on some interval V_Y depending of functional variable X defined on some interval V_X, two major models have been considered. On the one hand, we have the concurrent functional linear model

$$Y_n(v) = \beta(v)X_n(v) + \varepsilon_n(v)$$

  where ε_n is a zero mean finite variance random functional error and β is the coefficient function that has to be estimated. This model assumes that the value of Y_n at v depends only on the current value of $X_n(v)$ and not on the cross dependencies. A naïve approach to estimate β is to fit the model locally in a neighborhood of v using theb ordinary least square method and then smooth the estimate $\hat{\beta}(v)$ across the whole domain. Nevertheless, other estimation methods can be found [WY02; FZ08].

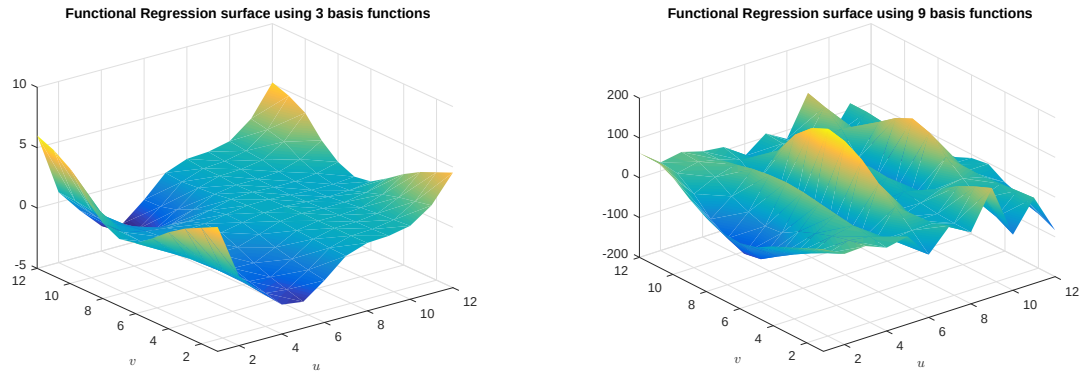
  The second approach is referred to as the Functional Linear Model:

$$Y_n(v) = \int \alpha(u, v)X_n(u)du + \varepsilon_n(v)$$

  where $\alpha(u, v)$ can be considered as the functional parameter. At any given point v, the value of $Y_n(v)$ depends on the entire trajectory of X_n. It is a direct extension of traditional linear models with multivariate response and vector covariates by changing the inner product from the Euclidean vector space to L^2. It is worth noting that the domain of X and Y does not need to be in the same range. The common approach to estimate the functional parameter is to project the regression operator into some functional basis:

$$\alpha(u, v) = \sum_{l=1}^{L} \sum_{m=1}^{M} b_{l,m} e_l(u) f_m(v),$$

  where $b_{l,m}$ are scalar coefficients and e_l and f_m some basis functions which need not be orthonormal. The basis functions e_l and f_m are usually chosen by the user (i.e., fourier, splines...), and the parameters $b_{l,m}$ are optimized so as to minimize prediction error. The



(a) Kernel using 3 basis functions

(b) Kernel using 9 basis functions

Figure 2.9. Kernel for the regression operator for estimating the precipitation data with the temperature data.

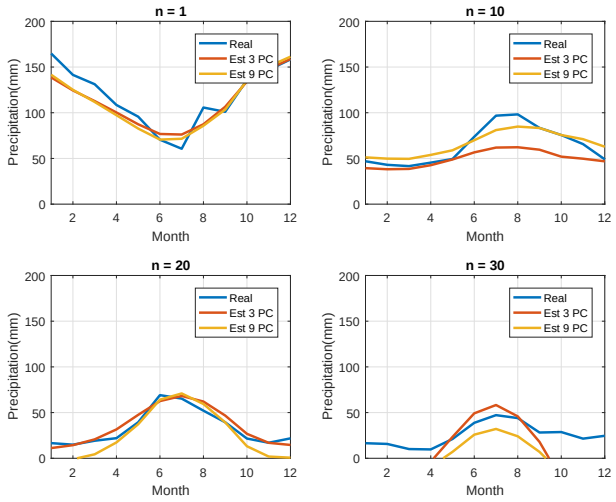


Figure 2.10. Comparison of estimations from the functional linear model for four different observations: $n = 1, 10, 20, 30$.

most common basis functions used are the functional principal components, i.e, [Far97], although other options are available as seen in [Ant+10] which uses a B-spline estimator. Other approaches are found in [He+10] based on canonical correlation analysis.

Figure 2.9 shows the resulting regression kernels when the yearly profiles of precipitation data are explained with the yearly profiles of temperature data. In this case, the basis functions are chosen to be the Principal Components of both input and output. Figure 2.9 shows the resulting kernels when $K = 3$ PC or $K = 9$ PC are used. As can be seen, the use of a higher number of components result in a noisier kernel. Figure 2.10 shows some examples of the real curves and forecasts with each method.

These models can be directly extended to account for multiple predictor variables. A functional coefficient of determination R^2 for the functional linear models has also been defined in [Yao+05a], [Ram06], [HK12].

2.2.7. Error measures

This section briefly describes accuracy measures that can be calculated for the residuals of a prediction model in FDA. In general, there is not a standard criterion for functional errors and it usually depends on the application. Nevertheless, some functional accuracy measures can be computed based on the L^p norm:

The Functional Mean Average Error (FMAE), also called the Mean Integrated Average Error (MIAE), is defined as

$$\text{FMAE} = N^{-1}\sum_{t=1}^{T}\|Y_t - \hat{Y}_t\|_{L_1} = N^{-1}\sum_{t=1}^{T}\int |Y_t(u) - \hat{Y}_t(u)|du.$$

The Functional Mean Square Error (FMSE), also called the Mean Integrated Square Error (MISE), is defined as

$$\text{FMSE} = N^{-1}\sum_{t=1}^{T}\|Y_t - \hat{Y}_t\|^2_{L_2} = N^{-1}\sum_{t=1}^{T}\int \left(Y_t(u) - \hat{Y}_t(u)\right)^2 du$$

A root version of FMSE can be computed as

$$\text{FRMSE} = \sqrt{N^{-1}\sum_{t=1}^{T}\|Y_t - \hat{Y}_t\|^2_{L_2}} = \sqrt{N^{-1}\sum_{t=1}^{T}\int \left(Y_t(u) - \hat{Y}_t(u)\right)^2 du}$$

These error measurements can provide some results that are hard to interpret because the integration is done in the domain range of the functions $u \in [a, b]$. A way to normalize the results is by dividing each measure by $b - a$ as shown by [Ane+13].

Finally, a functional version of MAPE could be defined as the norm of the residual normalized by the norm of the real functional observation

$$\text{FMAPE} = N^{-1}\sum_{t=1}^{T}\frac{\|Y_t - \hat{Y}_t\|_{L^2}}{\|Y_t\|_{L^2}} = N^{-1}\sum_{t=1}^{T}\frac{\int |Y_t(u) - \hat{Y}_t(u)|du}{\int |Y_t(u)|du}$$

2.3. Forecasting functional time series

This section focuses on Functional Time Series (FTS), which is the motivation of the dissertation. Firstly, FTS are defined and classified according to the nature of the underlying continuous process that generates the FTS. Then the notion of a white noise process in functional data is endorsed, setting the basis for the development of forecasting models. The main functional forecasting models proposed int the literature for FTS are revised. Pros and cons of each method are highlighted, which provide the framework for the contributions of the thesis.

A FTS $\{Y_t\}_{t\in T}$ is the realization of functional stochastic process, i.e. a time sequence of functional random variables $Y_t = \{Y_t(v), v \in V\}$, where t is some measurement of time. Therefore, a FTS is a dependent functional process, where each random variable in the series

could depend on previous realizations of the same series. The following sections show the particular methods developed for these dependent functional data

2.3.1. Functional white noise

An important concept in time series forecasting is the notion of white noise process. Hereafter, the extension of this process to FDA is addressed following [Bos00]. Let $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ be a functional stochastic process of random functions $\varepsilon_t = \{\varepsilon_t(v), v \in V\}$. Then, $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a functional white noise (WN) process if:

- It has zero mean $E\varepsilon_t = 0$
- It has a finite variance $0 < E\|\varepsilon_t\|^2 = \sigma_\varepsilon < \infty$ and thus, an associated autocovariance operator $C_{\varepsilon_t} = C_\varepsilon$.
- The cross-covariance operator between elements of the process is zero, i.e. $C_{\varepsilon_t,\varepsilon_s}(\cdot) = E\left[\langle\varepsilon_s, \cdot\rangle\, \varepsilon_t\right] = 0$ for all $t \neq s$.

In addition, if $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is i.i.d., it is denoted as a strong functional white noise process.

Let us highlight the implications of the properties of a WN process. As the cross-covariance function is zero, it means that there is no linear relation between an element of the process and any of its past realizations. Therefore, the process cannot be modeled based on linear time dependent correlations. Surprisingly, there are no assumption on the shape of the autocovariance function thus, the functions $\{\varepsilon_t\}$ can have some internal structure in the variance. C_{ε_t} admits the spectral decomposition and FPC can be extracted. Thus, the WN process $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ can be expressed as its projection on the basis spanned by the FPC $\{e_k\}$:

$$\varepsilon_t(v) = \sum_{k=1}^{\infty} c_{t,k} e_k(v).$$

Therefore, a functional Gaussian WN functional process does not only correspond to the case when each point in each curve is gaussian i.e. $\varepsilon_t(v) \sim N(0, \sigma_v)$, which would be the case in a scalar univariate or multivariate white noise. In the functional setting, a gaussian white noise implies that the coordinates with respect to the functional basis are gaussian distributed i.e. $c_{t,i} \sim N(0, \sigma_i)$. From a practical perspective, this representation is useful for generating synthetic WN functional processes by choosing an orthonormal functional basis and then, generating random coordinates $c_{t,i} \sim N(0, \sigma_i)$ for the observations.

2.3.2. Forecasting methods

Several approaches have been developed for forecasting functional time series in the past two decades. The novelty of FTS has caused a high diversity of approaches combining theoretical developments as well as some practical implementations. We have grouped the different methods into three main categories, which we will refer to as parametric models, dimensionality reduction models and nonparametric models.

Following [FV06], parametric models will be considered as linear functional models where the functional operators used to model the dependency in the series will generally fall into the

category of integral operators in L^2. Therefore, the kernel of the operator can be considered as the functional parameter that has to be estimated. Dimensionality reduction accounts for models that expand the FTS into a functional basis and then, applying standard multivariate techniques to the reduced coordinates. These two approaches are separated by a very thin line. Parametric methods usually end up expanding the kernel surface into a functional basis and then, optimizing the coordinates in that subspace. Finally, nonparametric models differ significantly from the other two as they do not assume any linear structure for the functional operators and can thus be considered as non-linear models.

2.3.2.1. Parametric models

The reference parametric model for forecasting FTS is the Autoregressive Hilbertian Process of order one ARH(1) proposed in [Bos00] which extensively covers theoretical properties of functional linear processes as well a some estimation procedures. It has also been studied in [Mou02], [MP11] and [HK12]. Assuming a zero-mean functional time series Y, the ARH(1) model is defined as:

$$Y_t = \Psi(Y_{t-1}) + \varepsilon_t \tag{2.2}$$

where ε_t is a white noise process in L^2 and Ψ is the functional autoregressive operator that needs to be estimated. Analogously to the univariate scalar setting [Bos00] proves that a stationary and casual solution for the ARH(1) can be found under the condition that $\|\Psi\| < 1$.

[Bos00] proposes estimating Ψ using an equivalent set of Yule-Walker equations for functional data. The equivalent formulation can be derived from equation 2.2:

$$E\left[\langle Y_t, \cdot\rangle\, Y_{t-1}\right] = E\left[\langle \Psi(Y_{t-1}), \cdot\rangle\, Y_{t-1}\right] + E\left[\langle \varepsilon_t, \cdot\rangle\, Y_{t-1}\right].$$

Since ε_t has zero mean and is not correlated with Y_{t-1}, $E\left[\langle \varepsilon_t, \cdot\rangle\, Y_{t-1}\right]$ cancels. By defining C as the covariance operator and D as the lag-1 autocovariance operator $D(\cdot) = E\left[\langle\, Y_t, \cdot\rangle\, Y_{t+1}\right]$, the expression is simplified as $D = \Psi C$. Hence, Ψ can be calculated as:

$$\Psi = DC^{-1}$$

However, the operator C lies on an infinite dimensional space, thus, it does not have a bounded inverse. The solution is to represent the operator C by its spectral decomposition, but limiting the number of eigenfunctions e_j and eigenvalues λ_j to a finite value K. Then, the inverse of C is approximated as

$$C^{-1}(\cdot) \approx \sum_{j=1}^{K} \lambda_j^{-1} \langle \cdot, e_j\rangle\, e_j.$$

This methodology can be seen as if the covariance operator C had been calculated for the data projected onto a finite subspace formed by the K functional principal components of C. Now, given the empirical estimation the lag-1 covariance operator $\hat{D}(\cdot) = (T-1)^{-1} \sum_{t=1}^{T-1} \langle\, Y_t, \cdot\rangle\, Y_{t+1}$, the estimation of Ψ is derived:

$$\hat{\Psi}(\cdot) = \hat{D}\hat{C}^{-1} = \frac{1}{T-1} \sum_{t=1}^{T-1} \sum_{j=1}^{K} \lambda_j^{-1} \langle \cdot, e_j\rangle \langle Y_t, e_j\rangle\, Y_{t+1}.$$

[Did+11] emphasizes that typically, an addition smoothing step is introduced by the approximation $Y_{t+1} \approx \sum_{i=1}^{K} \langle Y_{t+1}, e_i \rangle e_i$ which projects Y_{t+1} into the subspace generated by the FPC basis. Then, it leads to the estimator

$$\hat{\Psi}(\cdot) = \frac{1}{T-1} \sum_{t=1}^{T-1} \sum_{j=1}^{K} \sum_{i=1}^{K} \lambda_j^{-1} \langle \cdot, e_j \rangle \langle Y_t, e_j \rangle \langle Y_{t+1}, e_i \rangle e_i.$$

It can be shown how $\hat{\Psi}(\cdot)$ is an integral operator with the following kernel function $\hat{\psi}(u, v)$:

$$\hat{\psi}(u, v) = \frac{1}{T-1} \sum_{t=1}^{T-1} \sum_{j=1}^{K} \sum_{i=1}^{K} \lambda_j^{-1} \langle Y_t, e_j \rangle \langle Y_{t+1}, e_i \rangle e_j(u) e_i(v).$$

Other approaches are developed in the literature that estimate the linear Ψ in the ARH(1) model. Reviews and comparison of different methods can be found in [Bes+00], which applies them to forecast climatic variations. A summary of these approaches follows:

- Regression estimation: This estimation procedure models Ψ as in the Functional Linear Model, where Ψ is an integral operator in L^2. Then, it builds the corresponding regression model where the kernel $\psi(u, v)$ is estimated. The usual approach is to expand the kernel into some functional basis and then, optimized the coordinates in that basis so that forecasting error is minimized. [Ant+10] uses a B-spline estimator for forecasting electricity consumption. [Chi12] applies the functional linear model combined with a functional clustering to account for different situations.

- Partial Least Squares, developed by [HS09], computes a basis decomposition method that extracts uncorrelated latent component scores by maximizing the covariance function between functional predictors and functional responses. The model uses and integral operator with a kernel of the form:

  $$\psi(u, v) = \sum_{m=1}^{M} f_m(u) g_m(v)$$

  where g_m are the latent components of the responses, and f_m are functions such that $\langle f_m, Y_t \rangle = \zeta_{m,t}$ which are the common latent component scores.

- Predictive factors: [KO08] develop a technique that finds the most relevant directions for the prediction. It focuses on the optimal expansion of Ψ, which is, theoretically, the best predictor of Y_t, rather than the optimal expansion of Y_{t-1}. The estimation procedure involves complex formulations and the selection of two parameters.

Despite the numerous advances in the estimation of the ARH(1) model, this method may not be enough for modeling more complex functional time series. The extension to the ARH(p) model, in which more than one lag is taken into account for the prediction, can be found in [Bos00]. Additionally, the ARHX(1) model which includes exogenous variables was introduced

by [DG02] and [DG05] defines the ARHX(p) model which combines both improvements. The ARHX(p) model is therefore, given by:

$$Y_t = \Psi_1(Y_{t-1}) + \cdots + \Psi_p(Y_{t-p}) + a_1(X_{t,1}) + \cdots + a_z(X_{t,z}) + \varepsilon_t,$$

where $X_{t,1}, \cdots, X_{t,z}$ are the z autoregressive exogenous variables, which are associated respectively with operators $u_1, \cdots, u_z$ and strong white noises $\mu_1, \cdots, \mu_z$, i.e. $X_{t,i} = u_i(X_{t-1,i}) + \mu_{t,i}$. By grouping the different terms in matrix form and denoting:

$$T_{t-1} = \begin{pmatrix} Y_{t-1} \\ \vdots \\ Y_{t-p} \\ X_{t,1} \\ \vdots \\ X_{t,z} \end{pmatrix} \qquad \varepsilon'_t = \begin{pmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \\ \mu_{t,1} \\ \vdots \\ \mu_{t,z} \end{pmatrix} \qquad \Psi' = \begin{pmatrix} \Psi_1 & \cdots & \Psi_p & a_1 & \cdots & a_z \\ I & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & I & 0 & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & u_1 & \cdots & 0 \\ \vdots & \cdots & \cdots & 0 & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & \cdots & u_z \end{pmatrix}$$

the following ARH(1) process is obtained, which can be estimated using the different techniques presented earlier

$$T_t = \Psi'(T_{t-1}) + \varepsilon'_t.$$

Regarding more complex linear models, some theoretical developments for functional moving average processes are carried by [Tur+07] and [Bos00] while the functional ARMA(p, q) process is theoretically introduced in [Bos00] as

$$\sum_{j=0}^{p} \Psi_j(Y_{t-j}) = \sum_{j=0}^{q} \Theta_j(\varepsilon_{t-j})$$

where $\Psi_0 = \Theta_0 = I$. However, not much work is found on estimating the operators of this model for real applications. [Sen10], [Kle+16] thoroughly analyze the ARMA(p, q) process deriving conditions such that the model has a stationary solution. In addition, they showed that if Y_t, Ψ_j, ε_t and Θ_j are projected to the FPC of Y, then, the functional ARMA process becomes a Vector ARMA(p, q) process for the time series of scores. This leads directly to the next category of models, based on dimensionality reduction.

2.3.2.2. Dimensionality reduction

This model pursues the goal of reducing the functional forecasting problem to a multivariate version so that standard multivariate techniques can be applied to the reduced set of variables. This is achieved by dimensionality reduction transformation.

The common approach is to project the data Y_t into the space spanned by the K most relevant functional principal components $\{e_k\}_{k=1}^{K}$. Recall that in practices, a few components are enough to obtain a reasonable result. Therefore, each function is approximated by

$$Y_t(v) \approx \sum_{k=1}^{K} c_{t,k} e_k(v)$$

where $c_{t,k}$ are the principal component scores, which form the reduced dimensional multivariate dataset. Then, by computing estimates of $c_{t,k}$, forecast of functions are obtained:

$$\hat{Y}_{t+1}(v) = \sum_{k=1}^{K} \hat{c}_{t+1,k} e_k(v).$$

The time seres of scores $\hat{c}_{t+1,k}$ can be forecasted with standard ARIMA models or neural networks. Once the estimates are calculated, the estimated curve is easily obtained. This method is used in [AM+13], [Val+02], [HSU07] [HS09]. These methodologies also allow the inclusion of both scalar and functional covariates to the models. For including the functional covariates,[Aue+15] proposes extracting the FPC scores from the functional covariates and adding them as scalar covariates in the multivariate models.

Despite the fact that FPC scores are uncorrelated, the absence of serial correlation between them is not assured. [Sen10], [Aue+15] and [Kle+16] propose using vector time series models so as to estimate all the scores jointly.

Other dimensionality reduction approaches are based on Functional Factor models. [Lie13] and [Hay+12], propose using a basis expansion of the time series into a number K of unobserved functional factors. Both the basis functions and scores are unobserved and have to be determined form the data so that the reconstruction error is minimized. Once the basis functions and scores are obtained, standard time series models are used to forecast the time series of scores.

[Sal+15] transforms the FAR model into a functional additive autoregressive model (FAAR) which can take into account nonlinear effects by k-Nearest Neighbors estimation on the principal component scores.

2.3.2.3. Nonparametric models

Nonparametric regression attempts to estimate the functional operator without previously defining a linear structure. It can also be considered as a nonlinear method. The first advances in functional nonparametric time series modeling were provided in [FV06] and [Mas05]. A review of different nonparametric methods can be found in [AP+11]. The most common procedure is to use the Nadaraya-Watson kernel estimator, with the form:

$$\Psi(Y_{t-1}) = \sum_{i=2}^{t-1} \omega_h(Y_{t-1}, Y_i - 1) Y_i$$

where $\omega_h(\cdot, \cdot)$ is the weight function given by

$$\omega_h(Y_{t-1}, Y_i) = \frac{K(\frac{d(Y_{t-1}, Y_i)}{h})}{\sum_{j=1}^{N-1} K(\frac{d(Y_{t-1}, Y_j)}{h})}$$

where K denotes a kernel function and $h > 0$ is the bandwidth parameter which needs to be estimated. In essence, the estimation is a weighted average of past observations of Y where the weights are given by ω_h.

Some examples of applications of this methodology can be found in [DC+05] for Surfur Dioxide levels prediction, and [Ane+13], which apply these technique to forecasting Residual demand curves. [Vil+12] use this non-parametric estimation for the next-day electricity demand and price. The last two also develop a method so as to include exogenous variables with the Semi-functional Partial linear model. [SL15] as well as [PS13] propose decomposing the functional time series into the trend and a stochastic part, then, nonparametric is used for the stochastic component.

As previously mentioned, the semimetric chosen is critical, as computing distances depend on that. [Tim+11] tackle this problem by making a wide comparison of semimetrics. [Ant+06] make use of wavelets to build semimetrics and apply the nonparmetric functional regression to forecasting electrical load. [AP+11] propose a local-linear approach for the nonparametric estimation of $\Psi(\cdot)$.

2.4. Conclusion

This chapter has introduced the theoretical background for Functional Data Analysis (FDA) which is the framework designed for analyzing functional random variables. These variables are real-valued functions that belong to an infinite dimensional space. Therefore, multivariate analysis tools cannot be directly applied to functional data. Throughout the main developments of FDA, the key goal has been to extend the standard tools from the multivariate to the functional framework and to introduce new necessary concepts for dealing with infinitely dimensional observations.

One of the most important concepts in FDA is the reduction of functional data into a finite functional basis system. Reducing the dimensionality from infinite to the number K of selected basis functions provides a way to approach functional data from a multivariate perspective. In this sense, the FPC basis has been proven to be very useful as it can represent the source data with a high fidelity using a small number of components. Additionally, integral operators in L^2 have been substantially used for modeling functional relations between functional variables.

Functional time series are the realization of functional stochastic processes, i.e. time sequences of functional random variables. Three main groups of models arise when forecasting FTS is addressed: Parametric models, Dimensionality Reduction and Nonparametric models. Parametric models make use of integral operators for modeling the time dependency between past observations. The second group reduces the dimensionality of the functional data to a finite number K of components which allow using multivatiate tools to forecast the reduced time series. These forecasts can be reconstructed to obtain the estimated curve. Finally, nonparametric models use nonlinear functional operators that find historical curves that happened in situations similar to the curve that it is being forecasted.

This dissertation is aimed at forecasting functional time series in the context of electricity markets. Usually, stochastic processes found in electricity markets are characterized by the effect of business and everyday activities which lead to weekly and daily seasonalities as well as peak and low demand hours. In addition, explanatory variables are important drivers of some time series such as the electricity prices. As an example, the weather (wind speed, precipitation,

etc.) affects the production of renewal technologies with lower generation costs influencing the offering behavior of the agents.

We are interested in methods that can model complex time dependencies as well as including other covariates. Most research in parametric and nonparametric models is focused on estimating the autoregressive model of order one, in which the following curve depends only on the last observation. This might not seem enough for fully modeling the functional time dependencies in electricity markets. Dimensionality reduction, on the other hand, can apply VARMA or ARIMA models to the series so that seasonality can be accounted for in the model as well as moving average terms. One drawback to this approach arises when the number of FPC needed to accurately represent the functional time series is large. In that case, it would be necessary to adjust several ARMA models or a large VARMA which could become a tough estimation procedure.

In summary, parametric and nonparametric models are well designed for accounting cross-effects on the whole domain of the curves however, they are more limited for modeling complex time dependencies such as accounting for seasonality as well as moving average effects. On the other hand, dimensionality reduction techniques take advantage of applying standard ARMA methods at the cost of reducing the dimensionality and losing information of the process. The model proposed in this thesis develops a method with functional operators that can take into account these complex time dependencies while avoiding the projection of the curves to a limited number of components, thus avoiding the corresponding information loss.

3

The SARIMAHX model

This chapter is devoted to the functional forecasting model developed in this thesis. The model is a seasonal autoregressive Moving Average Hilbertian model which includes exogenous covariates (SARIMAHX). The aim is to extend the standard ARIMA modeling for scalar time series to the functional framework using integral operators in the L^2 Hilbert space. As opposed to other approaches in the literature, the proposed estimation method for the operators does not require to reduce the dimensionality of the data.

3.1. Introduction

A functional time series (FTS) $\{Y_t\}_{t\in\mathbb{Z}}$ is a time-ordered sequence of random functions. These real-valued functions $Y_t(v)$ are elements of the Hilbert space L^2 of real square integrable functions defined on a closed interval $[a, b]$. Therefore, each observation is a curve. What we are mainly concerned in this thesis is to forecast future values of the time series based on past realizations of the same. The methods developed in the literature for forecasting functional time series have progressed significantly in the past two decades. As was detailed in chapter 2, functional data require new statistical tools for analyzing and studying these infinite dimensional processes. Regarding functional linear models for forecasting FTS, the reference method is the autoregressive model of order 1 (ARH(1)) which has been the center of numerous developments. This model assumes that the curve at time t depends on the curve observed at time $t-1$. The relation between Y_t and Y_{t-1} is modeled by means of an integral operator in L^2 which is defined by a bivariate function called the kernel of the operator. Hence, estimating the ARH(1) model implies estimating that bivariate kernel. Despite the fact that this model has been widely studied, it only accounts for the last observation in order to forecast the following curve, which might be too simple for cases where the FTS exibits complex time dependencies. Therefore, more advanced approaches can be seen in [Sen10], [Kle+16] where the functional ARMA(p, q) model is addressed. These authors propose estimating the functional parameters by projecting the time series into a functional subspace spanned by the Functional Principal

Component (FPC) basis. Then, the functional ARMA model is approximated by a VARMA model of the principal component scores. This approach has the advantage that with only a few FPC, a great variability of the model is accounted and thus, the VARMA only has to account for a few dimensions. On the contrary, by reducing the dimensionality, some information is being lost.

As opposed to the available methods in the literature, this thesis proposes a fully functional method that allows modeling complex time dependencies as well as provinding a novel estimation method for integral operators. Therefore, the objectives are twofold. On the one hand, it extends the functional time series model far beyond the ARMA(p,q) by including double seasonality with multiplicative interaction (which is required for hourly time series) and differentiation as well as scalar and functional covariates. Similiar to the reference functional linear model, the operators for each term are considered as integral operators in the Hilbert space L^2 which are defined by a bivariate function or kernel. The difficulty lies in the estimation of those kernels without reducing the dimensionality and thus, avoiding the loss of information. Therefore, this leads to the second contribution: an innovative methodology is proposed to estimate the bivariate kernel based on neural networks. The core idea is to model the operators' kernel by a finite sum of bivariate sigmoid functions. The sum of sigmoid functions are universal function approximates [Cyb89] which are commonly used in neural networks because of their properties to model non-linear relations. Artificial neural networks have found increasing consideration due to their function approximation capabilities, being the Multilayer Perceptron (MLP) one of the most popular neural network models.

Similar to the two versions of the functional linear model seen in chapter 2, we also contribute with the concurrent SARIMAHX model. This is a simpler model that still accounts for all the complex time interactions but the operators used do not model cross dependencies between curves. However, the model greatly improves the training speed.

The rest of the chapter is organized as follows. Firstly, functional regression is addressed, which introduces a novel methodology for estimating the kernel surfaces of an integral operator in L^2. Then, the functional SARIMAX model is introduced. The equation of the model is explained and the representation of the operators are also given. Then, the estimation algorithm is explained and a simulation study is performed to test the model adjustment. Finally, the concurrent SARMAHX model is detailed.

3.2. A new approach for functional regression

This section presents a new estimation procedure for the Functional Linear Model described in Section 2. Consider a set $\{Y_t(v)\}_{v\in V}$ of T observations from a zero-mean functional random variable Y a set of observations of functional covariates $\{X_t^z(v_z)\}$ with $z = 1, \ldots, Z_f$, $v_z \in V_z$ and a set of scalar random covariates $\{x_t^z\}$. where $z = 1, \ldots, Z_c$. The extended Functional Linear Model that accounts for both functional and scalar covariates can be defined as follows:

$$Y_t = \sum_{z=1}^{Z_f} \Gamma_z^f X_t^z + \sum_{z=1}^{Z_c} \Gamma_z^c x_t^z + \varepsilon_t \tag{3.1}$$

where ε_t is a functional white noise, Γ_z^f are the operators related to the Z_f functional explanatory variables and Γ_z^c are the operators related to the Z_c scalar explanatory variables. The usual representation for the functional operators are integral operators in L^2. Therefore, equation 3.1 can be expressed as follows for computing the estimation of Y_n:

$$\widehat{Y}_t(v') = \sum_{z=1}^{Z_f} \int \rho_z\left(v_z, v'\right) X_t^z\left(v_z\right) dv_z + \sum_{z=1}^{Z_c} \beta_z\left(v'\right) x_t^z, \tag{3.2}$$

being $\rho_z\left(v_z, v'\right)$ the kernels of the integral operators related to the functional covariates and $\beta_z\left(v'\right)$ the functional parameters for the scalar covariates.

As was explained in Section 2, the usual method employed for estimating ρ_z and β_z is to select some fixed functional basis for expanding the operators and then, optimizing the scores. The following subsections explain the new approach followed in this dissertation. The objective is not to depend on extracting a fixed functional basis from the data.

3.2.1. Parametric operator

Each bivariate kernel is modeled as a weighted sum of bivariate sigmoid functions. Therefore:

$$\rho_z\left(u, v\right) = \alpha_0^{\rho_z} + \sum_{g=1}^{G_{\rho_z}} \alpha_g^{\rho_z} \tanh(w_{g0}^{\rho_z} + w_{g1}^{\rho_z} u + w_{g2}^{\rho_z} v)$$

where $w_{g0}^{\rho_z}$, $w_{g1}^{\rho_z}$, $w_{g2}^{\rho_z}$ and $\alpha_g^{\rho_z}$ are the parameters that define the linear combination of sigmoids. The variables u and v take real values in the intervals in which the functional variables are defined (i.e. V or V_z). This approach can be viewed as a MLP neural network with a particular configuration: an input layer with two input variables u and v. One hidden layer with a number G_{ρ_z} of nonlinear hidden units with hyperbolic tangents as activation functions and $w_{g0}^{\rho_z}$,$w_{g1}^{\rho_z}$,$w_{g2}^{\rho_z}$ as the weights for each input. Finally, one output layer with one linear output unit having $\alpha_g^{\rho_z}$ as the weights for the activation of the hidden units.

Let us define $\bar{\boldsymbol{\rho}}_{\boldsymbol{z}}$ as the vector of parameters defining the sigmoids of operator ρ_z, hence $\bar{\boldsymbol{\rho}}_{\boldsymbol{z}} = (\alpha_g^{\rho_z}, w_{g0}^{\rho_z}, w_{g1}^{\rho_z}, w_{g2}^{\rho_z})$.

Figure 3.1 illustrates the modeling of the kernel by means of a combination of sigmoids.

For the scalar covariates, the functional parameter is modeled following the same approach, although univariate sigmoids are used as activation functions

$$\beta_z(v) = \alpha_0^{\beta_z} + \sum_{g=0}^{G_{\beta_z}} \alpha_g^{\beta_z} \tanh(w_{g0}^{\beta_z} + w_{g1}^{\beta_z} v)$$

where the set $\bar{\boldsymbol{\beta}}_{\boldsymbol{z}} = \left(\alpha_0^{\beta_z}, \alpha_g^{\beta_z}, w_{g0}^{\beta_z}, w_{g1}^{\beta_z}\right)$ correspond to the group of parameters defining the kernels of the scalar covariates' operators.

Therefore, the model is estimated when values for the parameters $\bar{\boldsymbol{\beta}}_{\boldsymbol{z}}$ and $\bar{\boldsymbol{\rho}}_{\boldsymbol{z}}$ are estimated.

$$\rho_z(u,v) = \alpha_0^{\rho_z} + \sum_{g=1}^{G_{\rho_z}} \alpha_g^{\rho_z} \tanh(w_{g0}^{\rho_z} + w_{g1}^{\rho_z} u + w_{g2}^{\rho_z} v)$$

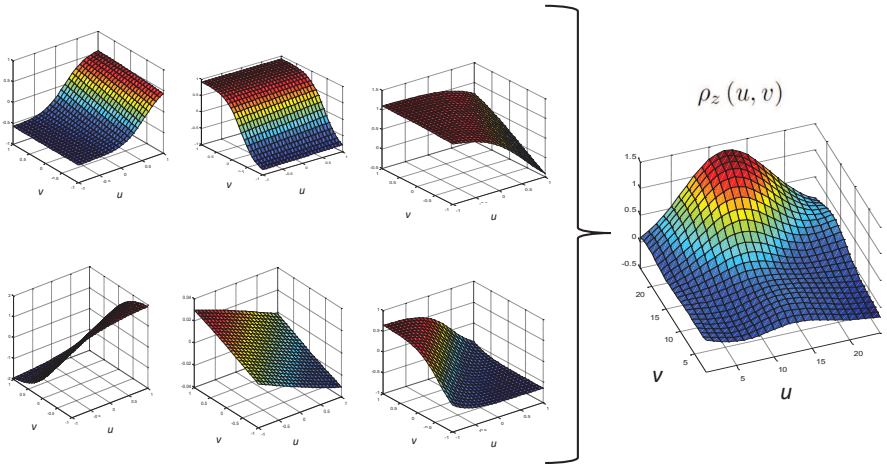


Figure 3.1. Illustration of the representation of a bivariate kernel as a function of bivariate sigmoids.

3.2.2. Learning algorithm

In order to optimize the vector of real parameters so as to minimize a certain cost function, a low-memory Quasi Newton method with random initial weights is used. This is a gradient descent optimization algorithm; therefore, the derivatives of the error with respect to the network parameters are needed. These derivatives can be effectively calculated by applying a backpropagation procedure. In this dissertation, the cost function is defined as the sum of the L^2 square errors $E = \sum_{t=1}^{T} e_t$, where:

$$e_t = \left\| Y_t - \widehat{Y}_t \right\|^2 = \int \left(Y_t(v) - \widehat{Y}_t(v) \right)^2 dv.$$

It should be noted that different cost functions could be used depending on the application. The derivatives of the error function with respect to a general parameter W is given by,

$$\frac{\partial E}{\partial W} = \sum_{t=1}^{T} \int 2 \left(Y_t(v) - \widehat{Y}_t(v) \right) \left(-\frac{\partial \widehat{Y}_t(v)}{\partial W} \right) dv,$$

where $\frac{\partial \widehat{Y}_t(v)}{\partial W}$ is the derivative of the estimation with respect to the parameter W. If the parameter W is $(\bar{\boldsymbol{\rho}}_{\boldsymbol{n}})_l$, that is to say, the l-th component of $\bar{\boldsymbol{\rho}}_{\boldsymbol{n}}$ of kernel parameters, this derivative can be obtained deriving equation (3.2)

$$\frac{\partial \widehat{Y}_t(v')}{\partial (\bar{\boldsymbol{\rho}}_{\boldsymbol{n}})_l} = \int \frac{\partial \rho_n(v, v')}{\partial (\bar{\boldsymbol{\rho}}_{\boldsymbol{n}})_l} X_t^n(u)\, du$$

Finally, the derivatives of the kernels with respect to the each of its sigmoid parameters can be calculated. Its derivative with respect to each particular weight $(\bar{\boldsymbol{\rho}}_{\boldsymbol{n}})_l$ are obtained as:

$$\begin{aligned}
\frac{\partial \psi_n(u,v)}{\partial \alpha_0^{\rho_n}} &= 1 \\
\frac{\partial \psi_n(u,v)}{\partial \alpha_g^{\rho_n}} &= \tanh(w_{g0}^{\rho_n} + w_{g1}^{\rho_n} u + w_{g2}^{\rho_n} v) \\
\frac{\partial \psi_n(u,v)}{\partial w_{g0}^{\rho_n}} &= \alpha_{ng}^{\rho_n} \cdot \tanh'(w_{g0}^{\rho_n} + w_{g1}^{\rho_n} u + w_{g2}^{\rho_n} v) \\
\frac{\partial \psi_n(u,v)}{\partial w_{g1}^{\rho_n}} &= \alpha_g^{\rho_n} \cdot \tanh'(w_{g0}^{\rho_n} + w_{g1}^{\rho_n} u + w_{ng2}^{r} v) \cdot u \\
\frac{\partial \psi_n(u,v)}{\partial w_{g2}^{\rho_n}} &= \alpha_g^{\rho_n} \cdot \tanh'(w_{g0}^{\rho_n} + w_{g1}^{\rho_n} u + w_{g2}^{\rho_n} v) \cdot v
\end{aligned}$$

For the case of the scalar covariates, the derivative of the estimation with respect to the parameter $(\bar{\boldsymbol{\beta}}_{\boldsymbol{n}})_l$, that is to say, the l-th component of $\bar{\boldsymbol{\beta}}_{\boldsymbol{n}}$ of kernel parameters, is

$$\frac{\partial \widehat{Y}_t(v')}{\partial (\bar{\boldsymbol{\beta}}_{\boldsymbol{n}})_l} = \frac{\partial \beta_n(v)}{\partial (\bar{\boldsymbol{\beta}}_{\boldsymbol{n}})_l} x_t^n$$

and the derivatives of it's kernel with respect to each parameter is given by

$$\begin{aligned}
\frac{\partial \beta_n(v)}{\partial \alpha_0^{\beta_n}} &= 1 \\
\frac{\partial \beta_n(v)}{\partial \alpha_g^{\beta_n}} &= \tanh(w_{g0}^{\beta n} + w_{g1}^{\beta_n} v) \\
\frac{\partial \beta_n(v)}{\partial w_{g0}^{\beta_n}} &= \alpha_g \cdot \tanh'(w_{g0}^{\beta_n} + w_{g1}^{\beta_n} v) \\
\frac{\partial \beta_n(v)}{\partial w_{g1}^{\beta_n}} &= \alpha_g \cdot \tanh'(w_{g0}^{\beta_n} + w_{g1}^{\beta_n} v) \cdot v
\end{aligned}$$

3.2.3. Practical implementation

The training and evaluation of the model has been implemented in C programming language. This includes the quasi-Newton algorithm and the derivatives which feed the quasi-Newton. It is a common procedure in neural network implementations to normalize or standardize the input and output variables. Similarly, In order to help the algorithm to converge faster and as a way to minimize the risk of falling into a local minimum, the following normalization steps are done to the inputs:

- The domain of the time series $\{Y_t\}$ as well as for the functional covariates $\{X_t\}$ is moved to the range $[-1, 1]$. By default, the sigmoid parameters are initialized with small values, causing the sigmoids to be centered in the surroundings of 0. With this transformation, the position of the sigmoids wont need to move far from that initial position.

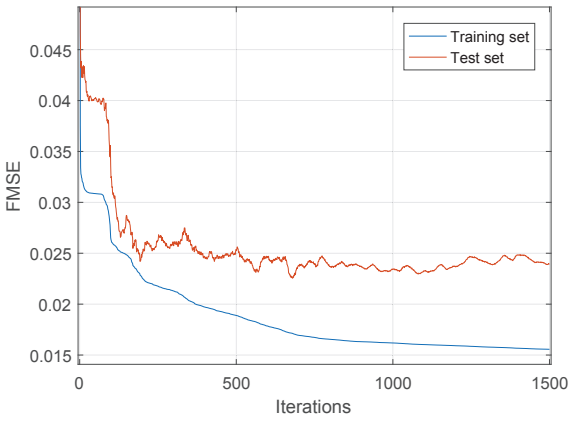


Figure 3.2. Training and test errors for each iteration of the training process.

- The time series $\{Y_t\}$ as well as for the functional covariates $\{X_t\}$ are normalized as $Y_t(v)/\sigma_Y$ and $Y_t(v)/\sigma_Y$, where σ_Y and σ_X are some constants.

In addition to the normalization step, a stopping criterion for the optimization has to be set. Neural network procedures are known to have a tendency towards overfitting. This event occurs when the training error is driven to a very small value, but when new data is presented to the network the error is large. The network has memorized the data, but it has not learned to generalize to new situations. In order to prevent this effect, early stopping methodology is used. The data is divided into a training set and a test set. The optimization algorithm is executed with the training set and the error is minimized. At the same time, the adjusted model in each iteration is evaluated with the test set. The neural network with the best performance is the one that generalized best to the unknown part of the dataset. Therefore, the selected parameter values are the ones that produce the lowest error in the test set. An example of the evolution of the training and test errors is shown in Figure 3.2.

3.3. Hilbertian SARIMAHX model

This section considers the functional SARIMAHX model. It is defined following the standard time series modeling proposed in [BJ70] but extended to functional time series using Hilbert operators. The following subsections present the definition and estimation procedure of the model.

3.3.1. Model definition

Consider a stationary and zero-mean functional time series $\{Y_t(v), t = 1, 2, ..., T, v \in V\}$, a set $\{X_t^z(v_z), z = 1, 2, \cdots, Z_f, t = 1, 2, \cdots, T, v_z \in V_z\}$ of Z_f functional covariates and a set $\{x_t^z, z = 1, 2, \cdots, Z_c, t = 1, 2, \cdots, T\}$ of Z_c scalar covariates. The SARIMAHX$(p, d, q) \times (P_1, D_1, Q_1)_{s_1} \times (P_2, D_2, Q_2)_{s_2}$ model is a functional Seasonal Autoregressive Moving Average Hilbertian model with two seasonalities (although it could be generalized to any number

of seasonalities) which includes both functional and scalar explanatory variables. The full expression for the model is defined as follows:

$$\left(I-\sum_{i=1}^{p}\Psi_i B^i\right)\left(I-\sum_{j=1}^{P_1}\Phi_j B^{j\cdot s_1}\right)\left(I-\sum_{k=1}^{P_2}\Xi_k B^{k\cdot s_2}\right)(I-B)^d\,(I-B^{s_1})^{D_1}\,(I-B^{s_2})^{D_2}\,Y_t$$
$$=\left(I-\sum_{i=1}^{q}\Theta_i B^i\right)\left(I-\sum_{j=1}^{Q_1}\Upsilon_j B^{j\cdot s_1}\right)\left(I-\sum_{k=1}^{Q_2}\Lambda_k B^{k\cdot s_2}\right)\varepsilon_t+\sum_{z=1}^{Z_f}\Gamma_z^f X_t^z+\sum_{z=1}^{Z_c}\Gamma_z^c x_t^z \tag{3.3}$$

where:

- ε_t is a functional white noise process.
- I is the identity operator.
- Parameters p, P_1 and P_2 are the regular and the two seasonal autoregressive orders respectively.
- Parameters q, Q_1 and Q_2 are the regular and the two seasonal moving average orders respectively.
- Parameters s_1 and s_2 are the seasonal time span for repeated patterns.
- Parameters d, D_1 and D_2 are the regular and seasonal integration orders respectively. For simplicity, we will restrict the integration orders to be 0 or 1, although it can be easily generalized for greater orders.
- Ψ_i, Φ_j and Ξ_k are the regular and seasonal autoregressive operators
- Θ_i, Υ_j and $\Lambda_k$the regular and seasonal moving average operators
- Γ_z^f the operators related to the Z_f explanatory variables.
- Γ_z^c the operators related to the Z_c explanatory variables.
- B^n is the lag operator which is defined as $B^nY_t = Y_{t-n}$ where $n \in \mathbb{N}$.

Let us briefly highlight the main features of this model. By choosing the order of the autoregressive and moving average terms, a great number of time behaviors can be modeled. Some naïve configurations would be a pure autorregresive model if $q = 0$, $Q_1 = 0$ and $Q_2 = 0$, a pure moving average model when $p = 0$, $P_1 = 0$ and $P_2 = 0$ or a functional regression model when all configuration parameters are zero. Moreover, the multiplicative effect between the regular and seasonal terms allows including extra lag dependencies by combining the regular and seasonal operators, as in the scalar case.

Equation (3.3) can be simplified by expanding the products and reordering terms, thus

$$Y_t=-\sum_{i=0}^{p}\sum_{j=0}^{P_1}\sum_{k=0}^{P_2}\Psi_i\Phi_j\Xi_k B^{(t^*+d^*)}Y_t+\sum_{i=0}^{q}\sum_{j=0}^{Q_1}\sum_{k=0}^{Q_2}\Theta_i\Upsilon_j\Lambda_k B^{t^*}\varepsilon_t+\sum_{z=1}^{Z_f}\Gamma_z^f X_t^z+\sum_{z=1}^{Z_c}\Gamma_z^c x_t^z+\varepsilon_t, \tag{3.4}$$

where

$$\begin{aligned} t^* &= i + j \cdot s_1 + k \cdot s_2 \\ d^* &= (1-d)(1-D_1)(1-D_2) - 1 \end{aligned}$$

where Ψ_0, Φ_0, Ξ_0, Θ_0, Υ_0 and Λ_0 are negative identity operators $-I$, and B^0 is the null operator. As mentioned earlier, this functional model makes use of integral operators in L^2 i.e, the regular autorregresive operator is defined as:

$$\Psi_i(f)(v) = \int \psi_i(u,v) f(u) du \qquad f \in L^2$$

The operators for the rest of the terms are defined alike except for the scalar covariates' operator, which is defined as $\Gamma_z(x)(v) = \beta_z(v)\, x$. In addition, when this model is used for forecasting, all the error terms are unobserved. Thus, expanding the functional operators to its integral form and simplifying the lag operators, the empirical forecast equation can be expressed as follows:

$$\begin{aligned} \widehat{Y}_t(v') = &\sum_{i=0}^{p}\sum_{j=0}^{P_1}\sum_{k=0}^{P_2} \iiint \psi_i\left(v,v'\right)\phi_j\left(u,v\right)\xi_k\left(u,v\right) Y_{t-(t^*+d^*)}\left(u\right)\, du\, du'\, dv \\ &- \sum_{i=0}^{q}\sum_{j=0}^{Q_1}\sum_{k=0}^{Q_2} \iiint \theta_i\left(v,v'\right)\vartheta_j\left(u,v\right)\lambda_k\left(u,v\right)\widehat{\varepsilon}_{t-t^*}\left(u\right)\, du\, du'\, dv \\ &+ \sum_{z=1}^{Z_f}\int \rho_z\left(v_z,v'\right) X_t^z\left(v_z\right)\, dv_z + \sum_{z=1}^{Z_c}\beta_z\left(v'\right) x_t^z, \end{aligned} \tag{3.5}$$

where

$$\begin{aligned} t^* &= i + j \cdot s_1 + k \cdot s_2 \\ d^* &= (1-d)(1-D_1)(1-D_2) - 1 \end{aligned}$$

and $u, v \in V$, $v_z \in V_z$, and $\widehat{\varepsilon}_t$ is the estimation of past innovations, which is given by:

$$\widehat{\varepsilon}_t\left(u\right) = Y_t\left(u\right) - \widehat{Y}_t\left(u\right) \tag{3.6}$$

Up to this point, the seasonal ARIMA forecasting method has been extended to functional time series using integral operators. In the following paragraphs, the estimation procedure is defined based on the operators' kernel structure described in the previous section.

Therefore, each of the operators' kernel is modeled by the finite sum of bivariate sigmoids The same configuration can be extended for the rest of the terms in the model. The weight parameters defining each kernel can been grouped into several sets:

- Regular autoregressive terms: $\bar{\boldsymbol{\psi}}_{\boldsymbol{i}} = \left(\alpha_0^{\psi_i}, \alpha_g^{\psi_i}, w_{g0}^{\psi_i}, w_{g1}^{\psi_i}, w_{g2}^{\psi_i}\right)$
- Seasonal autoregressive terms for first seasonality: $\bar{\boldsymbol{\phi}}_{\boldsymbol{j}} = \left(\alpha_0^{\phi_j}, \alpha_g^{\phi_j}, w_{g0}^{\phi_j}, w_{g1}^{\phi_j}, w_{g2}^{\phi_j}\right)$

- Seasonal autoregressive terms for second seasonality: $\bar{\boldsymbol{\xi}}_{\boldsymbol{k}} = \left(\alpha_0^{\xi_k}, \alpha_g^{\xi_k}, w_{g0}^{\xi_k}, w_{g1}^{\xi_k}, w_{g2}^{\xi_k}\right)$
- Regular moving average terms: $\bar{\boldsymbol{\theta}}_{\boldsymbol{i}} = \left(\alpha_0^{\theta_i}, \alpha_g^{\theta_i}, w_{g0}^{\theta_i}, w_{g1}^{\theta_i}, w_{g2}^{\theta_i}\right)$
- Seasonal moving average terms for second seasonality: $\bar{\boldsymbol{\lambda}}_{\boldsymbol{k}} = \left(\alpha_0^{\lambda_k}, \alpha_g^{\lambda_k}, w_{g0}^{\lambda_k}, w_{g1}^{\lambda_k}, w_{g2}^{\lambda_k}\right)$
- Seasonal moving average terms for first seasonality: $\bar{\boldsymbol{\vartheta}}_{\boldsymbol{j}} = \left(\alpha_0^{\vartheta_j}, \alpha_g^{\vartheta_j}, w_{g0}^{\vartheta_j}, w_{g1}^{\vartheta_j}, w_{g2}^{\vartheta_j}\right)$
- Functional Explanatory variables' terms: The sets $\bar{\boldsymbol{\rho}}_{\boldsymbol{z}}$ and $\bar{\boldsymbol{\beta}}_{\boldsymbol{z}}$ already defined before.

In order to optimize the vector of real parameters so as to minimize a certain cost function, a low-memory Quasi Newton method with random initial weights is used. This is a gradient descent optimization algorithm; therefore, the derivatives of the error with respect to the network parameters are needed. These derivatives can be effectively calculated by applying a backpropagation procedure. In this paper, the cost function is defined as the sum of the L^2 square errors $E = \sum_{t=1}^{T} e_t$, where:

$$e_t = \left\| Y_t - \widehat{Y}_t \right\|^2 = \int \left(Y_t(v) - \widehat{Y}_t(v) \right)^2 dv.$$

It should be noted that different cost functions could be used depending on the application. The derivatives of the error function with respect to a general parameter W is given by,

$$\frac{\partial E}{\partial W} = \sum_{t=1}^{T} \int 2 \left(Y_t(v) - \widehat{Y}_t(v) \right) \left(-\frac{\partial \widehat{Y}_t(v)}{\partial W} \right) dv,$$

where $\frac{\partial \widehat{Y}_t(v)}{\partial W}$ is the derivative of the estimation with respect to the parameter W. If the parameter W is $(\bar{\boldsymbol{\psi}}_{\boldsymbol{n}})_l$, that is to say, the l-th component of $\bar{\boldsymbol{\psi}}_{\boldsymbol{n}}$ of autoregressive parameters, this derivative can be obtained deriving equation (3.5)

$$\begin{aligned} \frac{\partial \widehat{Y}_t(v')}{\partial (\bar{\boldsymbol{\psi}}_{\boldsymbol{n}})_l} = & \sum_{j=0}^{P_1} \sum_{k=0}^{P_2} \iint \frac{\partial \psi_n(v, v')}{\partial (\bar{\boldsymbol{\psi}}_{\boldsymbol{n}})_l} \phi_j \left(u', v\right) \xi_k \left(u, u'\right) Y_{t-(t^{n*}+d^*)} \left(u\right) \, du \, dv \\ & + \sum_{i=0}^{q} \sum_{j=0}^{Q_1} \sum_{k=0}^{Q_2} \iiint \theta_i \left(v, v'\right) \vartheta_j \left(u', v\right) \lambda_k \left(u, u'\right) \frac{\partial \widehat{\varepsilon}_{t-t^*} \left(u\right)}{\partial (\bar{\boldsymbol{\psi}}_{\boldsymbol{n}})_l} \, du \, du' \, dv. \end{aligned}$$

As can be seen, due to the fact that $\widehat{\varepsilon}_{t-t^*}$ depends on past estimations of Y, it is also dependent on $r_{n,l}$ and, therefore, can have a nonzero derivative. The derivatives of the error terms with respect to a certain parameter W can be expressed as a function of former calculated derivatives. Hence, deriving equation (3.6) we obtain:

$$\frac{\partial \widehat{\varepsilon}_t \left(v\right)}{\partial W} = \frac{\partial Y_t \left(v\right) - \partial \widehat{Y}_t \left(v\right)}{\partial W} = -\frac{\partial \widehat{Y}_t(v)}{\partial W}$$

Note that $Y_t(v)$ is the real observation of the functional time series, that is to say, a constant value. Thus, its derivative is null. As a result, the derivatives can be calculated recursively by substituting the error derivative terms.

$$\frac{\partial \widehat{Y}_t(v')}{\partial(\bar{\boldsymbol{\psi}}_{\boldsymbol{n}})_l} = \sum_{j=0}^{P_1}\sum_{k=0}^{P_2}\iint \frac{\partial \psi_n(v,v')}{\partial(\bar{\boldsymbol{\psi}}_{\boldsymbol{n}})_l}\phi_j\left(u',v\right)\xi_k\left(u,u'\right)Y_{t-(t^{n*}+d^*)}\left(u\right)\,du\,dv$$
$$+\sum_{i=0}^{q}\sum_{j=0}^{Q_1}\sum_{k=0}^{Q_2}\iiint \theta_i\left(v,v'\right)\vartheta_j\left(u',v\right)\lambda_k\left(u,u'\right)\frac{-\partial \widehat{Y}_{t-t^*}(u)}{\partial(\bar{\boldsymbol{\psi}}_{\boldsymbol{n}})_l}\,du\,du'\,dv.$$

where $t^{n*} = n + j\cdot s_1 + k\cdot s_2$. Similarly, the derivatives for the rest of the autoregressive and moving average terms can be calculated. The full expressions can be found ahead:

- AR from first seasonality $\phi_n(u,v)$

$$\frac{\partial \widehat{Y}_t(v')}{\partial(\bar{\boldsymbol{\phi}}_{\boldsymbol{n}})_l} = \sum_{i=0}^{p}\sum_{k=0}^{P_2}\iint \psi_i\left(v,v'\right)\frac{\partial \phi_n\left(u',v\right)}{\partial(\bar{\boldsymbol{\phi}}_{\boldsymbol{n}})_l}\xi_k\left(u,u'\right)Y_{t-(t^{n*}+d^*)}\left(u\right)\,du\,dv$$
$$+\sum_{i=0}^{q}\sum_{j=0}^{Q_1}\sum_{k=0}^{Q_2}\iiint \theta_i\left(v,v'\right)\vartheta_j\left(u',v\right)\lambda_k\left(u,u'\right)\frac{-\partial \widehat{Y}_{t-t^*}(u)}{\partial(\bar{\boldsymbol{\phi}}_{\boldsymbol{n}})_l}\,du\,du'\,dv.$$

- AR from second seasonality $\xi_n(u,u')$

$$\frac{\partial \widehat{Y}_t(v')}{\partial(\bar{\boldsymbol{\xi}}_{\boldsymbol{n}})_l} = \sum_{i=0}^{p}\sum_{j=0}^{P_1}\iint \psi_i\left(v,v'\right)\phi_j\left(u',v\right)\frac{\partial \xi_n(u,u')}{\partial(\bar{\boldsymbol{\xi}}_{\boldsymbol{n}})_l}Y_{t-(t^{n*}+d^*)}\left(u\right)\,du\,dv$$
$$+\sum_{i=0}^{q}\sum_{j=0}^{Q_1}\sum_{k=0}^{Q_2}\iiint \theta_i\left(v,v'\right)\vartheta_j\left(u',v\right)\lambda_k\left(u,u'\right)\frac{-\partial \widehat{Y}_{t-t^*}(u)}{\partial(\bar{\boldsymbol{\xi}}_{\boldsymbol{n}})_l}\,du\,du'\,dv.$$

- Regular MA $\theta_n\left(v,v'\right)$

$$\frac{\partial \widehat{Y}_t(v')}{\partial(\bar{\boldsymbol{\theta}}_{\boldsymbol{n}})_l} = -\sum_{j=0}^{Q_1}\sum_{k=0}^{Q_2}\iint \frac{\partial \theta_n\left(v,v'\right)}{\partial(\bar{\boldsymbol{\theta}}_{\boldsymbol{n}})_l}\vartheta_j\left(u',v\right)\lambda_k\left(u,u'\right)\widehat{\varepsilon}_{t-t^{n*}}\left(u\right)\,du\,dv$$
$$+\sum_{i=0}^{q}\sum_{j=0}^{Q_1}\sum_{k=0}^{Q_2}\iiint \theta_i\left(v,v'\right)\vartheta_j\left(u',v\right)\lambda_k\left(u,u'\right)\frac{-\partial \widehat{Y}_{t-t^*}(u)}{\partial(\bar{\boldsymbol{\theta}}_{\boldsymbol{n}})_l}\,du\,du'\,dv.$$

- MA from first seasonality $\vartheta_n\left(u,v\right)$

$$\frac{\partial \widehat{Y}_t(v')}{\partial(\bar{\boldsymbol{\vartheta}}_{\boldsymbol{n}})_l} = -\sum_{i=0}^{q}\sum_{k=0}^{Q_2}\iint \theta_n\left(v,v'\right)\frac{\partial \vartheta_j\left(u',v\right)}{\partial(\bar{\boldsymbol{\vartheta}}_{\boldsymbol{n}})_l}\lambda_k\left(u,u'\right)\widehat{\varepsilon}_{t-t^{n*}}\left(u\right)\,du\,dv$$
$$+\sum_{i=0}^{q}\sum_{j=0}^{Q_1}\sum_{k=0}^{Q_2}\iiint \theta_i\left(v,v'\right)\vartheta_j\left(u',v\right)\lambda_k\left(u,u'\right)\frac{-\partial \widehat{Y}_{t-t^*}(u)}{\partial(\bar{\boldsymbol{\vartheta}}_{\boldsymbol{n}})_l}\,du\,du'\,dv.$$

- MA from second seasonality $\lambda_n(u,v)$

$$\frac{\partial \widehat{Y}_t(v')}{\partial(\bar{\boldsymbol{\lambda}}_{\boldsymbol{n}})_l} = -\sum_{i=0}^{q}\sum_{j=0}^{Q_1}\iint \theta_n(v,v')\,\vartheta_j(u',v)\,\frac{\partial \lambda_k(u,u')}{\partial(\bar{\boldsymbol{\lambda}}_{\boldsymbol{n}})_l}\widehat{\varepsilon}_{t-t^{n*}}(u)\,du\,dv$$
$$+\sum_{i=0}^{q}\sum_{j=0}^{Q_1}\sum_{k=0}^{Q_2}\iiint \theta_i(v,v')\,\vartheta_j(u',v)\,\lambda_k(u,u')\,\frac{-\partial \widehat{Y}_{t-t^*}(u)}{\partial(\bar{\boldsymbol{\lambda}}_{\boldsymbol{n}})_l}\,du\,du'\,dv.$$

3.3.2. Simulation study

This section is intended to validate the model for synthetic case studies. The finite sum of bivariate sigmoid functions can approximate any kernel (some restrictions apply, [Cyb89]), and the analytic derivatives assure that the quasi-Newton method always decreases the error function to be minimized. Nevertheless, to test the performance of the model in a controlled environment, some simulation analysis are carried out.

Table 3.1. Processes simulated in the synthetic case study

SARIMAH$(1,0,0)$	$Y_t = \tau_1(Y_{t-1}) + \varepsilon_t$
SARIMAH$(1,0,0)_5$	$Y_t = \tau_2(Y_{t-5}) + \varepsilon_t$
SARIMAH$(1,0,0)_{20}$	$Y_t = \tau_3(Y_{t-20}) + \varepsilon_t$
SARIMAH$(0,0,1)$	$Y_t = \tau_1(\varepsilon_{t-1}) + \varepsilon_t$
SARIMAH$(0,0,1)_5$	$Y_t = \tau_2(\varepsilon_{t-5}) + \varepsilon_t$
SARIMAH$(0,0,1)_{20}$	$Y_t = \tau_3(\varepsilon_{t-5}) + \varepsilon_t$
SARIMAH$(1,0,0)(1,0,0)_5$	$Y_t = \tau_1(Y_{t-1}) + \tau_2(Y_{t-5}) - \tau_1\tau_2(Y_{t-6}) + \varepsilon_t$
SARIMAH$(1,0,0)(0,0,1)_5$	$Y_t = \tau_1(Y_{t-1}) + \tau_2(\varepsilon_{t-5}) + \varepsilon_t$
SARIMAH$(0,0,1)(0,0,1)_5$	$Y_t = \varepsilon_t - \tau_1(Y_{t-1}) - \tau_2(Y_{t-1}) + \tau_1\tau_2(Y_{t-1})$
SARIMAH$(1,0,0)(1,0,0)_5(1,0,0)_{20}$	$Y_t = \tau_1(Y_{t-1}) + \tau_2(Y_{t-5}) + \tau_3(Y_{t-6}) - \tau_1\tau_2(Y_{t-6})$ $- \tau_1\tau_3(Y_{t-21}) - \tau_2\tau_3(Y_{t-26}) + \tau_1\tau_2\tau_3(Y_{t-27}) + \varepsilon_t$
SARIMAH$(0,0,1)(0,0,1)_5(0,0,1)_{20}$	$Y_t = \varepsilon_t - \tau_1(\varepsilon_{t-1}) - \tau_2(\varepsilon_{t-5}) - \tau_3(\varepsilon_{t-6})$ $+ \tau_1\tau_2(\varepsilon_{t-6}) + \tau_1\tau_3(\varepsilon_{t-21}) + \tau_2\tau_3(\varepsilon_{t-26}) - \tau_1\tau_2\tau_3(\varepsilon_{t-27})$

Table 3.1 shows the different SARIMAHX configurations that will be simulated and forecasted. where τ_i are integral functional operators and ε_t are some white noise processes. Both τ_i and ε_t are synthetically defined to explore different cases. Three operators are be chosen and three different WN time series are simulated. The setup for the simulations has been based on simulation studies found in the literature (see [Did+11; Hor+16] as some examples)

Firstly, the generation of the WN processes is addressed. A functional WN process is a sequence of zero-mean and finite variance functional random variables. Three different WN processes have been created in the range $[-1,1]$. Figure3.3 shows five samples for each one. The formulation of each WN process is described ahead:

- The first WN process is simply a white noise process for each element of the curve and time.

$$\varepsilon_t^{(1)}(v) \sim N(0,1)$$

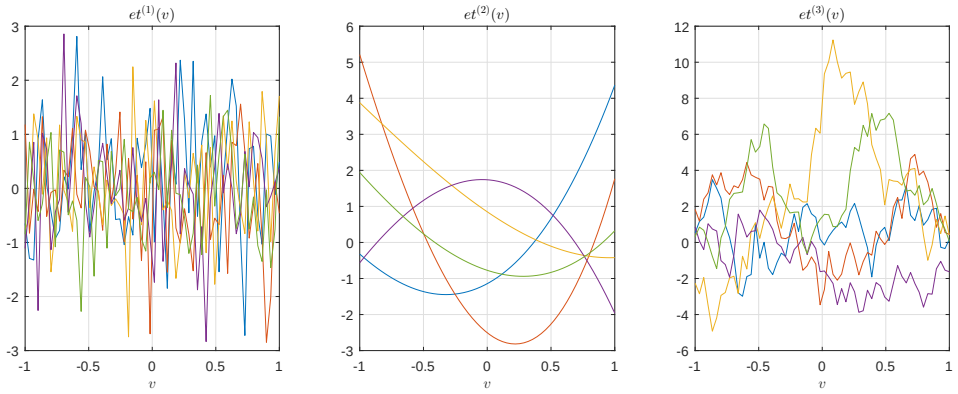


Figure 3.3. Representation of five samples of each white noise process used in the simulation study.

- The second approach is based on the idea that a Gaussian functional time series ε_t can be generated using a truncated functional basis $\{e_k(v)\}$ and random coefficients c_{ti} (see section 2.3.1). Then,

$$\varepsilon_t^{(2)}(v) = \sum_{k=1}^{K} c_{t,k} e_k(v)$$

  If the coefficients $c_{t,k}$ are randomly distributed following a normal distribution N(0, σ_k), $\varepsilon_t^{(2)}$ should be a gaussian white noise process. An initial truncated functional basis out of three functions $\sin(v^*)$, $\exp(v^*)$ and $\cos(v^*)$ is selected where $v^* = v/2 + 0.5$. Then, they have been orthonormalized using Gram-Smith method, thus, obtaining three orthonormal functions e_k. Finally, the WN process $\varepsilon_t^{(2)}$ is generated giving random values to $c_{t,k}$ from the normal distribution $N(0, 1)$

- The third approach is to generate a Brownian Bridge random process, which is a common WN seen in [Did+11]. Each innovation function is created as:

$$\varepsilon_t^{(3)}(v) = W(v^*) - v^* \cdot (W(1))$$

  where $v^* = v/2 + 0.5$ and $W(v^*)$ is a Wiener process in the domain v^*.

Secondly, the operators τ_i are defined. They are considered integral operators $\tau_i(f)(v) = \int \kappa_i(u, v) f(u) du$. Hence, the following synthetic kernels are used for each operator:

$$\kappa_1(u, v) = 0.6 \cdot e^{-(u^{*2} + v^{*2})}$$
$$\kappa_2(u, v) = \min(u^*, v^*)$$
$$\kappa_3(u, v) = -0.5 \min(u^*, v^*)$$

where $v^* = v/2 + 0.5$ and $u^* = u/2 + 0.5$. The resulting norms of the operators are $\|\tau_1\| = 0.89$, $\|\tau_2\| = 0.81$, $\|\tau_3\| = 0.4$. A visual representation of the kernels are shown in figure 3.4.

Therefore, three different versions of each SARIMAH configuration of table 3.1 are generated, one for each of the previous WN processes used. Each process is generated with 60 discrete points in the range $v \in [-1, 1]$ with a number T of 1000 observations, and ten replications will

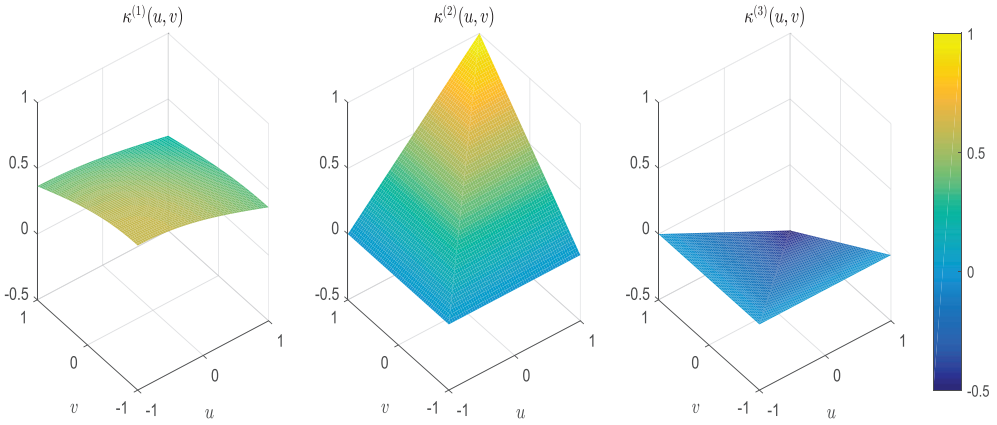


Figure 3.4. Visual representation of the three synthetic kernels used in the simulation study.

be generated for each method. Models are trained using the first 500 observations, while the last 500 is left for testing.

Several error measurements are accounted in order to asses the training. We will follow measurements proposed in [Did+11] for validating the adjustment. On the one hand, the distance between the real synthetic operator and the estimated one will be accounted by three different measurements: the Square Distance (SD), the Absolute Distance (AD) and the Relative Absolute Distance (RAD) which are defined as:

$$\text{SD} = \|\tau_i - \hat{\tau}_i\|^2 = \iint (\kappa_i(u,v) - \hat{\kappa_i}(u,v))^2 \, du \, dv$$

$$\text{AD} = \iint |\kappa_i(u,v) - \hat{\kappa_i}(u,v)| \, du \, dv$$

$$\text{RAD} = \frac{\iint |\kappa_i(u,v) - \hat{\kappa_i}(u,v)| \, du \, dv}{\iint |\kappa_i(u,v)| \, du \, dv}$$

In addition, residuals are also compared by measuring the Functional Root Mean Square Error (FRMSE) and the Functional Mean Average Error (FMAE):

$$\text{FRMSE} = \sqrt{T^{-1} \sum_{t=500}^{T} \int (Y_t(v) - \hat{Y}_t(v))^2 dv}$$

$$\text{FMAE} = T^{-1} \sum_{t=500}^{T} \int |Y_t(v) - \hat{Y}_t(v)| dv$$

The generated series Y_t are compared with two forecasts: The forecast produced by forecasting the series with the real operators, i.e. ideal prediction, and the estimation given by the models adjusted with the SARIMAHX. These results help addressing if the errors observed in the estimation of the operators have great impact on the residuals.

Average values across replications are calculated for each error measurement previously described. Tables 3.2, 3.3 and 3.4 show the errors in the estimation of the shape of the operators' kernel for each simulated model in table 3.1 and each WN process. These results

are analyzed. On the one hand, the estimation of the operator's kernel yielded errors in the following ranges:

- For kernel $\kappa^{(1)}$: SD $\in [0.06, 0.17]$, AD $\in [0.10, 0.29]$ and RAD $\in [0.05, 0.17]$
- For kernel $\kappa^{(2)}$: SD $\in [0.11, 0.25]$, AD $\in [0.12, 0.41]$ and RAD $\in [0.09, 0.7]$
- For kernel $\kappa^{(3)}$: SD $\in [0.08, 0.31]$, AD $\in [0.14, 0.52]$ and RAD $\in [0.21, 0.78]$

The different distances measured depend on the shape of the kernel that is estimated and on the type of coefficient it models. It could be said that kernel $\kappa^{(1)}$ is, in general, easier to estimate than kernel $\kappa^{(2)}$ or $\kappa^{(3)}$ as the errors observed for the latter two are higher. Indeed, the first kernel is somewhat flat, which helps the adjustment of the sigmoids. The second and third kernel present a higher non-linear behavior, such as the sharp corner along the diagonal which challenges the optimization. Moreover, kernel 3 has a lower norm, thus the normalized SAD distances for this operator tend to be higher. It is worth noting that the estimation of the operators is not very sensitive to the WN process of the series, however, errors vary between configurations of the SARIMAH models. As the number of operators to be estimated increases, the distance measured slightly increase because the optimization problem becomes more complex. When there is only one operator to be estimated, the results are similar for both AR and MA terms. However, when more terms are included, MA operators seem to be harder to estimate.

These distance values are not easily interpreted. In order to grasp and visualize the implications of these results, Figure 3.5 shows an estimation for each of the three synthetic operators defined. The distances measured for each estimation are the following: For the first kernel: SD $= 0.107$, AD $= 0.165$, RAD $= 0.094$. For the second kernel: SD $= 0.197$, AD $= 0.340$, RAD $= 0.194$. For the third kernel: SD $= 0.188$, AD $= 0.264$, RAD $= 0.396$. Comparing the shape of the estimated operators with the synthetic ones shown in figure 3.4, it could be said that the estimated kernels approximate the shape of the real surfaces.

In addition, the effects in the forecasting errors of underestimating the kernels are shown in tables 3.5 and 3.6. The residuals of the ideal prediction for each WN process simulated is compared to the residuals obtained with the trained models. As shown, all trained models provide good estimation errors, close to the ideal prediction.

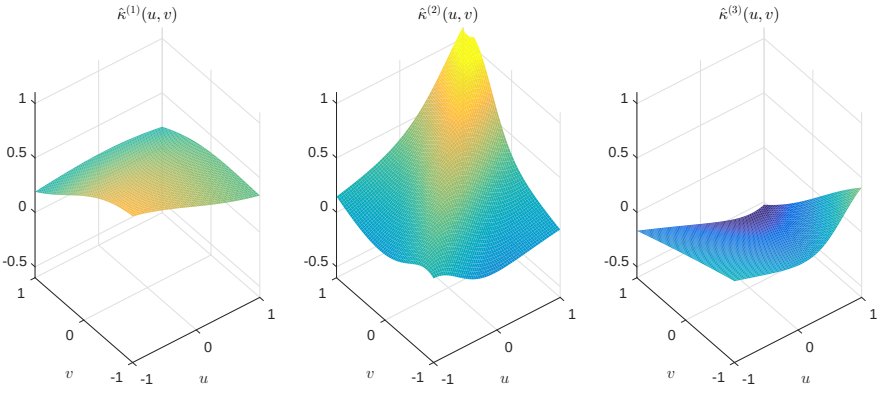


Figure 3.5. Visual representation of the estimation of three synthetic kernels, each measuring the following distances $SD = 0.122$, $AD = 0.230$, $RAD = 0.134$

Table 3.2. Average SD distances across replications in the estimation of the shape of the τ_i kernels

	τ_1			τ_2			τ_3		
	ε_t^A	ε_t^B	ε_t^C	ε_t^A	ε_t^B	ε_t^C	ε_t^A	ε_t^B	ε_t^C
ARIMAH$(1,0,0)$	0.064	0.068	0.098						
ARIMAH$(1,0,0)_5$				0.116	0.121	0.137			
ARIMAH$(1,0,0)_{20}$							0.135	0.097	0.087
ARIMAH$(0,0,1)$	0.072	0.074	0.067						
ARIMAH$(0,0,1)_5$				0.145	0.164	0.147			
ARIMAH$(0,0,1)_{20}$							0.136	0.128	0.154
ARIMAH$(1,0,0)(1,0,0)_5$	0.072	0.063	0.084	0.078	0.080	0.114			
ARIMAH$(1,0,0)(0,0,1)_5$	0.068	0.072	0.080	0.183	0.178	0.214			
ARIMAH$(0,0,1)(0,0,1)_5$	0.101	0.110	0.119	0.186	0.182	0.276			
ARIMAH$(1,0,0)(1,0,0)_5(1,0,0)_{20}$	0.172	0.105	0.176	0.242	0.201	0.259	0.293	0.205	0.310
ARIMAH$(0,0,1)(0,0,1)_5(0,0,1)_{20}$	0.172	0.105	0.176	0.242	0.201	0.259	0.293	0.205	0.310

Table 3.3. Average AD distances across replications in the estimation of the shape of the τ_i kernel

	τ_1			τ_2			τ_3		
	ε_t^A	ε_t^B	ε_t^C	ε_t^A	ε_t^B	ε_t^C	ε_t^A	ε_t^B	ε_t^C
ARIMAH$(1,0,0)$	0.103	0.108	0.163						
ARIMAH$(1,0,0)_5$				0.175	0.198	0.223			
ARIMAH$(1,0,0)_{20}$							0.206	0.156	0.140
ARIMAH$(0,0,1)$	0.120	0.121	0.108						
ARIMAH$(0,0,1)_5$				0.236	0.271	0.226			
ARIMAH$(0,0,1)_{20}$							0.224	0.205	0.252
ARIMAH$(1,0,0)(1,0,0)_5$	0.115	0.102	0.137	0.127	0.130	0.186			
ARIMAH$(1,0,0)(0,0,1)_5$	0.108	0.114	0.131	0.302	0.292	0.330			
ARIMAH$(0,0,1)(0,0,1)_5$	0.169	0.182	0.185	0.301	0.299	0.425			
ARIMAH$(1,0,0)(1,0,0)_5(1,0,0)_{20}$	0.297	0.170	0.293	0.419	0.311	1.016	0.522	0.330	0.465
ARIMAH$(0,0,1)(0,0,1)_5(0,0,1)_{20}$	0.297	0.170	0.293	0.419	0.311	1.016	0.522	0.330	0.465

Table 3.4. Average RAD distances across replications in the estimation of the shape of the τ_i kernel

	ρ_1			ρ_2			ρ_3		
	ε_t^A	ε_t^B	ε_t^C	ε_t^A	ε_t^B	ε_t^C	ε_t^A	ε_t^B	ε_t^C
ARIMAH$(1,0,0)$	0.058	0.061	0.092						
ARIMAH$(1,0,0)_5$				0.100	0.113	0.127			
ARIMAH$(1,0,0)_{20}$							0.303	0.234	0.210
ARIMAH$(0,0,1)$	0.068	0.069	0.061						
ARIMAH$(0,0,1)_5$				0.134	0.154	0.129			
ARIMAH$(0,0,1)_{20}$							0.337	0.309	0.378
ARIMAH$(1,0,0)(1,0,0)_5$	0.065	0.058	0.078	0.095	0.097	0.139			
ARIMAH$(1,0,0)(0,0,1)_5$	0.061	0.065	0.074	0.172	0.166	0.188			
ARIMAH$(0,0,1)(0,0,1)_5$	0.096	0.103	0.1058	0.226	0.224	0.318			
ARIMAH$(1,0,0)(1,0,0)_5(1,0,0)_{20}$	0.169	0.097	0.167	0.314	0.233	0.762	0.784	0.495	0.697
ARIMAH$(0,0,1)(0,0,1)_5(0,0,1)_{20}$	0.169	0.097	0.167	0.314	0.233	0.762	0.784	0.495	0.697

Table 3.5. Average FRMSE forecast error across replications

	ε_t^A	ε_t^B	ε_t^C
Ideal Prediction	2.0070	5.9293	21.1831
ARIMAH$(1,0,0)$	2.0071	5.9428	21.2218
ARIMAH$(1,0,0)_5$	2.0073	5.9514	21.1990
ARIMAH$(1,0,0)_{20}$	2.0206	5.9919	21.0531
ARIMAH$(0,0,1)$	2.0114	5.9351	21.2668
ARIMAH$(0,0,1)_5$	2.0075	5.9584	21.1920
ARIMAH$(0,0,1)_{20}$	2.0075	5.9735	21.2509
ARIMAH$(1,0,0)(1,0,0)_5$	2.007	5.967	21.265
ARIMAH$(1,0,0)(0,0,1)_5$	2.008	6.020	21.445
ARIMAH$(0,0,1)(0,0,1)_5$	2.0091	6.0419	21.7815
ARIMAH$(1,0,0)(1,0,0)_5(1,0,0)_{20}$	2.054	7.658	27.769
ARIMAH$(0,0,1)(0,0,1)_5(0,0,1)_{20}$	2.054	7.658	27.769

Table 3.6. Average FMAE forecast error across replications

	ε_t^A	ε_t^B	ε_t^C
Ideal Prediction	1.5984	2.6552	5.0481
ARIMAH$(1,0,0)$	1.5984	2.6596	5.0535
ARIMAH$(1,0,0)_5$	1.5985	2.6589	5.0515
ARIMAH$(1,0,0)_{20}$	1.6038	2.6596	5.0558
ARIMAH$(0,0,1)$	1.6001	2.6571	5.0640
ARIMAH$(0,0,1)_5$	1.5985	2.6618	5.0516
ARIMAH$(0,0,1)_{20}$	1.5986	2.6649	5.0576
ARIMAH$(1,0,0)(1,0,0)_5$	1.598	2.666	5.063
ARIMAH$(1,0,0)(0,0,1)_5$	1.598	2.676	5.086
ARIMAH$(0,0,1)(0,0,1)_5$	1.599	2.682	5.134
ARIMAH$(1,0,0)(1,0,0)_5(1,0,0)_{20}$	1.616	3.047	5.884
ARIMAH$(0,0,1)(0,0,1)_5(0,0,1)_{20}$	1.616	3.047	5.884

3.3.3. Multi-step-ahead forecasts

The training of the model is aimed at minimizing the forecasting error for forecasting horizon equal to one. However, once the values of the different sigmoid parameters are obtained, the model can be used to forecast a multiple-step horizon following a simple procedure:

- For the first curve to predict, the model is evaluated as normal for step one and the forecast is obtained.
- For the second curve to predict, the real curve at previous time is not available. In order to solve the issue, the previous forecast obtained for that time is considered as the real value as it is the best estimation available. Then, the second curve is forecasted by evaluating the one-step forecast.
- This procedure is done iteratively for every horizon needed. Unobserved future curves are substituted with the corresponding forecasts and assuming they are real values.

It should be noted that when forecasts of the model are used as real values, the moving average term does not affect the following forecasts as the observed error is 0. Consequently, in the same way as an ARIMA model, the future forecasts can be slightly degraded if the moving average term is very significant.

In this proposed model, there is a parameter that is chosen by the practitioner which is the number of neurons for each term's operator. The tests have suggested that usually, there is no need for selecting a high number of neurons. In practice, with five or six sigmoids the representation capabilities are quite good, as a great diversity of shapes can be modeled.

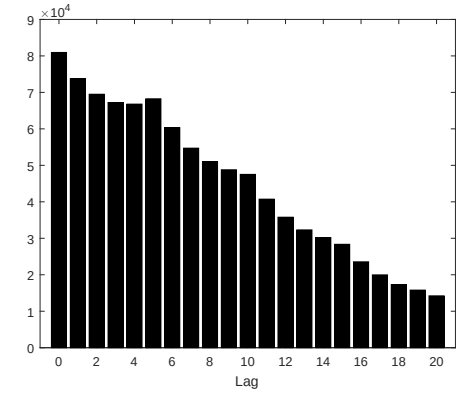
For computational purposes, the average training time is shown in table 3.7. It shows the average training time of 100 iterations of the optimization process for the different configurations. As can be seen, the algorithm is computational intensive, which increases the training time as the model becomes more complex. The reason is found in the computation of the integrals, causing the majority of time to be dedicated to calculations that could be easily paralelized. Consequently, the training times can be significantly reduced if the code is further optimized.

Table 3.7. Average training time for 100 iteration of the functional SARIMAHX model

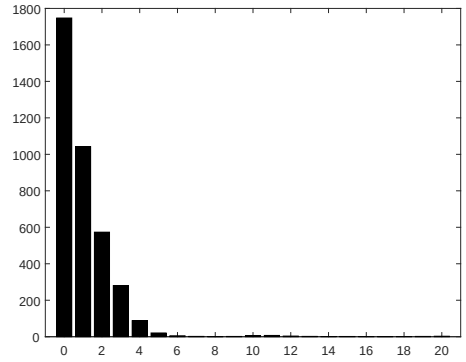
Model	Time (min)
ARIMAH$(1,0,0)$	5.92
ARIMAH$(0,0,1)$	8.51
ARIMAH$(1,0,0)(1,0,0)_5$	23.34
ARIMAH$(1,0,0)(0,0,1)_5$	19.02
ARIMAH$(0,0,1)(0,0,1)_5$	42.12
ARIMAH$(1,0,0)(1,0,0)_5(1,0,0)_{20}$	602.23
ARIMAH$(0,0,1)(0,0,1)_5(0,0,1)_{20}$	692.10

3.3.4. Model identification and analysis of residuals

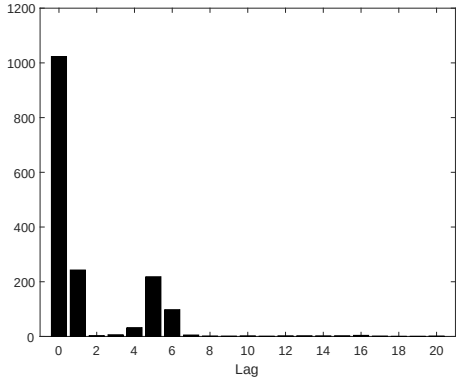
It is a common approach in standard ARIMA time series to identify relevant lags of the time series that are correlated with the value to be forecasted. This is materialized into the Autocorrelation



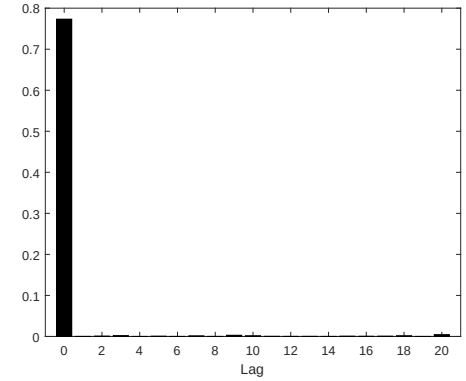
(a) Functional ACF for SARIMAH(1,0)x(1,0)



(b) Functional ACF for SARIMAH(1,0)x(0,1)



(c) Functional ACF for SARIMAH(0,1)x(0,1)



(d) Functional ACF for residuals

Figure 3.6. Functional ACF plots for different simulated processes

Function and Partial Autocorrelation Function (ACF and PACF respectively). Their purpose is to identify the structure of the model (AR and MA orders). The ACF and PACF are also used for testing the residuals [BJ70].

Finding an equivalent approximation to the functional framework would be very useful. One of the properties of white noise processes is that the covariance function between lag i and lag j i.e. $C_{i,j}$, is zero (see section 2.3.1). Then, a lagged index could be defined as the norm of the lagged covariance as:

$$c_{i,j} = \iint C_{i,j}(u,v)dudv$$

If any of these indexes take high values, it would mean the covariance at that lag is not zero, and thus, the series is not a white noise process. In addition, if these lag indexes are plotted for different structures of the time series, various behaviors are observed. This is shown in figure 3.6, where the evolution of the lagged covariance index is plotted for different configurations. Depending on the AR or MA terms of the model, the behavior of the index is different. It is worth noting that the plot for the residuals the covariance index at lag 0 (autocovariance) is high, but for the rest of the lags, the index is close to zero.

3.4. The concurrent SARIMAHX model

Based on the previous functional SARIMAHX model, a concurent version can be defined. The functional concurrent model is specified in section 2 as an alternative to the functional linear model, where integral operators are substituted by the following linear operators

$$\Psi(f)(u) = \psi(u) \cdot f(u). \qquad f \in L^2$$

These are simpler operators, which do not model cross-relations between the input and the response functional variable.

3.4.1. Model definition

The concurrent SARIMAHX model equation is the same as equation 3.3, however, the operators used are concurrent linear operators with a univariate function as the functional parameter. Hence, the concurrent empirical forecast equation is given by:

$$\begin{aligned}\widehat{Y}_t(v) = &\sum_{i=0}^{p}\sum_{j=0}^{P_1}\sum_{k=0}^{P_2} \psi_i(v)\,\phi_j(v)\,\xi_k(v)\,Y_{t-(t^*+d^*)}(v) - \sum_{i=0}^{q}\sum_{j=0}^{Q_1}\sum_{k=0}^{Q_2} \theta_i(v)\,\vartheta_j(v)\,\lambda_k(v)\,\widehat{\varepsilon}_{t-t^*}(v) \\ &+ \sum_{z=1}^{Z_f} \rho_z(v)\,X_t^z(v_z) + \sum_{z=1}^{Z_c} \beta_z(v)\,x_t^z,\end{aligned} \tag{3.7}$$

$$\begin{aligned} t^* &= i + j \cdot s_1 + k \cdot s_2 \\ d^* &= (1-d)(1-D_1)(1-D_2) - 1 \end{aligned}$$

In order to estimate the model, a similar approach to the functional SARIMAHX model is followed. The coefficient function is modeled as a sum of sigmoid functions. For simplicity, only the regular autorregression term is shown:

$$\rho_z(u,v) = \alpha_0^{\rho_z} + \sum_{g=1}^{G_{\rho_z}} \alpha_g^{\rho_z} \tanh(w_{g0}^{\rho_z} + w_{g1}^{\rho_z} v)$$

where $w_{g0}^{\rho_z}$, $w_{g1}^{\rho_z}$, $w_{g2}^{\rho_z}$ and $\alpha_g^{\rho_z}$ are the parameters that define the linear combination of sigmoids. The estimation of the parameters for all the terms in the model can be done in a similar approach to the fully functional model. A quasi-Newton algorithm minimizes the cost function defined as the sum of the L^2 square errors $E = \sum_{t=1}^{T} e_t$. The derivatives of the error function with respect to a general parameter W is given by,

$$\frac{\partial E}{\partial W} = \sum_{t=1}^{T} \int 2\left(Y_t(v) - \widehat{Y}_t(v)\right)\left(-\frac{\partial \widehat{Y}_t(v)}{\partial W}\right) dv,$$

where $\frac{\partial \widehat{Y}_t(v)}{\partial W}$ is the derivative of the estimation with respect to the parameter W. Hereafter are whown the expressions that result from deriving equation (3.7) with respect to each of the parameter to be optimized:

- Regular AR $\psi_n(v)$

$$\frac{\partial \widehat{Y}_t(v)}{\partial(\bar{\boldsymbol{\psi}}_{\boldsymbol{n}})_l} = \sum_{j=0}^{P_1}\sum_{k=0}^{P_2} \frac{\partial \psi_n(v)}{\partial(\bar{\boldsymbol{\psi}}_{\boldsymbol{n}})_l} \phi_j(v)\, \xi_k(v)\, Y_{t-(t^{n*}+d^*)}(v)\, dudv + \sum_{i=0}^{q}\sum_{j=0}^{Q_1}\sum_{k=0}^{Q_2} \theta_i(v)\, \vartheta_j(v)\, \lambda_k(v)\, \frac{-\partial \widehat{Y}_{t-t^*}(v)}{\partial(\bar{\boldsymbol{\psi}}_{\boldsymbol{n}})_l}.$$

- AR from first seasonality $\phi_n(v)$

$$\frac{\partial \widehat{Y}_t(v)}{\partial(\bar{\boldsymbol{\phi}}_{\boldsymbol{n}})_l} = \sum_{i=0}^{p}\sum_{k=0}^{P_2} \psi_i(v) \frac{\partial \phi_n(v)}{\partial(\bar{\boldsymbol{\phi}}_{\boldsymbol{n}})_l} \xi_k(v)\, Y_{t-(t^{n*}+d^*)}(v) + \sum_{i=0}^{q}\sum_{j=0}^{Q_1}\sum_{k=0}^{Q_2} \theta_i(v)\, \vartheta_j(v)\, \lambda_k(v)\, \frac{-\partial \widehat{Y}_{t-t^*}(v)}{\partial(\bar{\boldsymbol{\phi}}_{\boldsymbol{n}})_l}.$$

- AR from second seasonality $\xi_n(v)$

$$\frac{\partial \widehat{Y}_t(v)}{\partial(\bar{\boldsymbol{\xi}}_{\boldsymbol{n}})_l} = \sum_{i=0}^{p}\sum_{j=0}^{P_1} \psi_i(v)\, \phi_j(v)\, \frac{\partial \xi_n(v)}{\partial(\bar{\boldsymbol{\xi}}_{\boldsymbol{n}})_l} Y_{t-(t^{n*}+d^*)}(v) + \sum_{i=0}^{q}\sum_{j=0}^{Q_1}\sum_{k=0}^{Q_2} \theta_i(v)\, \vartheta_j(v)\, \lambda_k(v)\, \frac{-\partial \widehat{Y}_{t-t^*}(v)}{\partial(\bar{\boldsymbol{\xi}}_{\boldsymbol{n}})_l}.$$

- Regular MA $\theta_n(v)$

$$\frac{\partial \widehat{Y}_t(v)}{\partial(\bar{\boldsymbol{\theta}}_{\boldsymbol{n}})_l} = -\sum_{j=0}^{Q_1}\sum_{k=0}^{Q_2} \frac{\partial \theta_n(v)}{\partial(\bar{\boldsymbol{\theta}}_{\boldsymbol{n}})_l} \vartheta_j(v)\, \lambda_k(v)\, \widehat{\varepsilon}_{t-t^{n*}}(v) + \sum_{i=0}^{q}\sum_{j=0}^{Q_1}\sum_{k=0}^{Q_2} \theta_i(v)\, \vartheta_j(v)\, \lambda_k(v)\, \frac{-\partial \widehat{Y}_{t-t^*}(v)}{\partial(\bar{\boldsymbol{\theta}}_{\boldsymbol{n}})_l}.$$

- MA from first seasonality $\vartheta_n(v)$

$$\frac{\partial \widehat{Y}_t(v)}{\partial(\bar{\boldsymbol{\vartheta}}_{\boldsymbol{n}})_l} = -\sum_{i=0}^{q}\sum_{k=0}^{Q_2} \theta_n(v)\, \frac{\partial \vartheta_j(v)}{\partial(\bar{\boldsymbol{\vartheta}}_{\boldsymbol{n}})_l}(v)\, \lambda_k(v)\, \widehat{\varepsilon}_{t-t^{n*}}(v) + \sum_{i=0}^{q}\sum_{j=0}^{Q_1}\sum_{k=0}^{Q_2} \theta_i(v)\, \vartheta_j(v)\, \lambda_k(v)\, \frac{-\partial \widehat{Y}_{t-t^*}(v)}{\partial(\bar{\boldsymbol{\vartheta}}_{\boldsymbol{n}})_l}.$$

- MA from second seasonality $\lambda_n(v)$

$$\frac{\partial \widehat{Y}_t(v)}{\partial(\bar{\boldsymbol{\lambda}}_{\boldsymbol{n}})_l} = -\sum_{i=0}^{q}\sum_{j=0}^{Q_1} \theta_i(v)\vartheta_j(v)\, \frac{\partial \lambda_k(v)}{\partial(\bar{\boldsymbol{\lambda}}_{\boldsymbol{n}})_l}(v)\, \widehat{\varepsilon}_{t-t^{n*}}(v) + \sum_{i=0}^{q}\sum_{j=0}^{Q_1}\sum_{k=0}^{Q_2} \theta_i(v)\, \vartheta_j(v)\, \lambda_k(v)\, \frac{-\partial \widehat{Y}_{t-t^*}(v)}{\partial(\bar{\boldsymbol{\lambda}}_{\boldsymbol{n}})_l}.$$

Finally, the derivatives of the kernels with respect to their parameters are given as:

$$\begin{aligned}
\frac{\partial \psi_n(v)}{\partial \alpha^r_{n0}} &= 1 \\
\frac{\partial \psi_n(v)}{\partial \alpha^r_{ng}} &= \tanh(w^r_{ng0} + w^r_{ng1} v) \\
\frac{\partial \psi_n(v)}{\partial w^r_{ng0}} &= \alpha_{ng} \cdot \tanh'(w^r_{ng0} + w^r_{ng1} v) \\
\frac{\partial \psi_n(v)}{\partial w^r_{ng1}} &= \alpha_{ng} \cdot \tanh'(w^r_{ng0} + w^r_{ng1} v) \cdot v
\end{aligned}$$

The same concepts described for the functional SARIMAHX model also holds for the concurrent model. The normalization steps as well as the training and test division of the data is also performed so as to achieve a robust model.

3.4.2. Simulation study

As it was done with the functional model, a simulation study is also performed to validate this new approach. A similar configuration of the simulations is followed. The processes to be simulated are those in table 3.1, although the operators τ_i are concurrent operators. The WN processes $\varepsilon_t^{(j)}$ are the same as those used in the functional SARIMAHX simulation. Regarding the operators, they are defined as:

$$\begin{aligned}
\kappa_1(v) &= v^2 \\
\kappa_2(v) &= exp(v) \\
\kappa_3(v) &= -0.6 exp(-4v)
\end{aligned}$$

Figure 3.7 shows a visualization of these kernels. The error terms for comparing the residuals are equally calculated as the functional case. However, the distances for measuring the accuracy in the operators' estimation are redefined: the Square Distance (SD), the Absolute Distance (AD) and the Relative Absolute Distance (RAD) which are defined as:

$$\begin{aligned}
\text{SD} &= \|\tau_i - \hat{\tau}_i\|^2 = \int (\kappa_i(v) - \hat{\kappa_i}(v))^2 dv \\
\text{AD} &= \int |\kappa_i(v) - \hat{\kappa_i}(v)| dv \\
\text{RAD} &= \frac{\int |\kappa_i(v) - \hat{\kappa_i}(v)| du dv}{\int |\kappa_i(v)| dv}
\end{aligned}$$

Regarding the results in tables 3.8, 3.9, 3.10, 3.12, 3.11, similar results to the functional case are obtained. The operators are estimated with fairly low error and the results show that the residuals from the estimated model approximates the true ones. This model however, is much less computationally intensive than the former model. Table 3.13 shows the training times taken by each model configuration.

Table 3.8. Average SD errors across replications in the estimation of the shape of the concurrent operators' kernel

	ρ_1			ρ_2			ρ_3		
	ε_t^A	ε_t^B	ε_t^C	ε_t^A	ε_t^B	ε_t^C	ε_t^A	ε_t^B	ε_t^C
ARIMAH$(1,0,0)$	0.035	0.060	0.063						
ARIMAH$(1,0,0)_5$				0.081	0.134	0.130			
ARIMAH$(1,0,0)_{20}$							0.015	0.040	0.048
ARIMAH$(0,0,1)$	0.051	0.063	0.062						
ARIMAH$(0,0,1)_5$				0.014	0.059	0.058			
ARIMAH$(0,0,1)_{20}$							0.206	0.156	0.140
ARIMAH$(1,0,0)(1,0,0)_5$	0.056	0.057	0.078	0.062	0.075	0.103			
ARIMAH$(1,0,0)(0,0,1)_5$	0.068	0.072	0.080	0.183	0.178	0.214			
ARIMAH$(0,0,1)(0,0,1)_5$	0.101	0.110	0.119	0.186	0.182	0.276			
ARIMAH$(1,0,0)(1,0,0)_5(1,0,0)_{20}$	0.130	0.083	0.146	0.204	0.191	0.482	0.470	0.353	0.496
ARIMAH$(0,0,1)(0,0,1)_5(0,0,1)_{20}$	0.172	0.105	0.176	0.242	0.201	0.259	0.293	0.205	0.310

Table 3.9. Average AD errors across replications in the estimation of the shape of the concurrent operators' kernel

	ρ_1			ρ_2			ρ_3		
	ε_t^A	ε_t^B	ε_t^C	ε_t^A	ε_t^B	ε_t^C	ε_t^A	ε_t^B	ε_t^C
ARIMAH$(1,0,0)$	0.051	0.082	0.093						
ARIMAH$(1,0,0)_5$				0.097	0.160	0.154			
ARIMAH$(1,0,0)_{20}$							0.016	0.047	0.055
ARIMAH$(0,0,1)$	0.083	0.093	0.092						
ARIMAH$(0,0,1)_5$				0.016	0.071	0.067			
ARIMAH$(0,0,1)_{20}$							0.175	0.198	0.223
ARIMAH$(1,0,0)(1,0,0)_5$	0.088	0.091	0.123	0.097	0.117	0.165			
ARIMAH$(1,0,0)(0,0,1)_5$	0.108	0.114	0.131	0.302	0.292	0.330			
ARIMAH$(0,0,1)(0,0,1)_5$	0.169	0.182	0.185	0.301	0.299	0.425			
ARIMAH$(1,0,0)(1,0,0)_5(1,0,0)_{20}$	0.195	0.113	0.202	0.264	0.223	0.617	0.304	0.219	0.308
ARIMAH$(0,0,1)(0,0,1)_5(0,0,1)_{20}$	0.297	0.170	0.293	0.419	0.311	1.016	0.522	0.330	0.465

Table 3.10. Average RAD errors across replications in the estimation of the shape of the concurrent operators' kernel

	ρ_1			ρ_2			ρ_3		
	ε_t^A	ε_t^B	ε_t^C	ε_t^A	ε_t^B	ε_t^C	ε_t^A	ε_t^B	ε_t^C
ARIMAH$(1,0,0)$	0.042	0.092	0.087						
ARIMAH$(1,0,0)_5$				0.138	0.227	0.218			
ARIMAH$(1,0,0)_{20}$							0.0382	0.108	0.126
ARIMAH$(0,0,1)$	0.053	0.080	0.078						
ARIMAH$(0,0,1)_5$				0.023	0.101	0.095			
ARIMAH$(0,0,1)_{20}$							0.309	0.234	0.210
ARIMAH$(1,0,0)(1,0,0)_5$	0.054	0.061	0.086	0.075	0.097	0.129			
ARIMAH$(1,0,0)(0,0,1)_5$	0.061	0.065	0.074	0.172	0.166	0.188			
ARIMAH$(0,0,1)(0,0,1)_5$	0.096	0.103	0.1058	0.226	0.224	0.318			
ARIMAH$(1,0,0)(1,0,0)_5(1,0,0)_{20}$	0.130	0.083	0.146	0.205	0.191	0.482	0.470	0.354	0.496
ARIMAH$(0,0,1)(0,0,1)_5(0,0,1)_{20}$	0.169	0.097	0.167	0.314	0.233	0.762	0.784	0.495	0.697

Table 3.11. Average FRMSE concurrent forecast error across replications

	ε_t^A	ε_t^B	ε_t^C
Ideal Prediction	2.0070	5.9293	21.1831
ARIMAH$(1,0,0)$	2.0073	5.9374	21.217
ARIMAH$(1,0,0)_5$	2.0366	6.078	21.4751
ARIMAH$(1,0,0)_{20}$	2.0072	5.9337	21.2067
ARIMAH$(0,0,1)$	2.0099	5.9298	21.2403
ARIMAH$(0,0,1)_5$	2.0072	5.9336	21.1961
ARIMAH$(0,0,1)_{20}$	2.0206	5.9919	21.0531
ARIMAH$(1,0,0)(1,0,0)_5$	2.0075	5.9675	21.274
ARIMAH$(1,0,0)(0,0,1)_5$	2.008	6.020	21.445
ARIMAH$(0,0,1)(0,0,1)_5$	2.0091	6.0419	21.7815
ARIMAH$(1,0,0)(1,0,0)_5(1,0,0)_{20}$	2.041	6.895	24.910
ARIMAH$(0,0,1)(0,0,1)_5(0,0,1)_{20}$	2.054	7.658	27.769

Table 3.12. Average FMAE concurrent forecast error across replications

	ε_t^A	ε_t^B	ε_t^C
Ideal Prediction	1.5984	2.6552	5.0481
ARIMAH$(1,0,0)$	1.5985	2.6582	5.0533
ARIMAH$(1,0,0)_5$	1.6096	2.6815	5.0863
ARIMAH$(1,0,0)_{20}$	1.5985	2.6562	5.0515
ARIMAH$(0,0,1)$	1.5995	2.6559	5.059
ARIMAH$(0,0,1)_5$	1.5985	2.6557	5.0492
ARIMAH$(0,0,1)_{20}$	1.6038	2.6596	5.0558
ARIMAH$(1,0,0)(1,0,0)_5$	1.5986	2.666	5.063
ARIMAH$(1,0,0)(0,0,1)_5$	1.598	2.676	5.086
ARIMAH$(0,0,1)(0,0,1)_5$	1.599	2.682	5.134
ARIMAH$(1,0,0)(1,0,0)_5(1,0,0)_{20}$	1.612	2.878	5.53
ARIMAH$(0,0,1)(0,0,1)_5(0,0,1)_{20}$	1.616	3.047	5.884

Table 3.13. Average training time for 100 iterations of the concurrent model

Model	Time (min)
ARIMAH$(1,0,0)$	0.65
ARIMAH$(0,0,1)$	0.82
ARIMAH$(1,0,0)(1,0,0)_5$	2.41
ARIMAH$(1,0,0)(0,0,1)_5$	2.36
ARIMAH$(0,0,1)(0,0,1)_5$	2.75
ARIMAH$(1,0,0)(1,0,0)_5(1,0,0)_{20}$	5.39
ARIMAH$(0,0,1)(0,0,1)_5(0,0,1)_{20}$	7.45

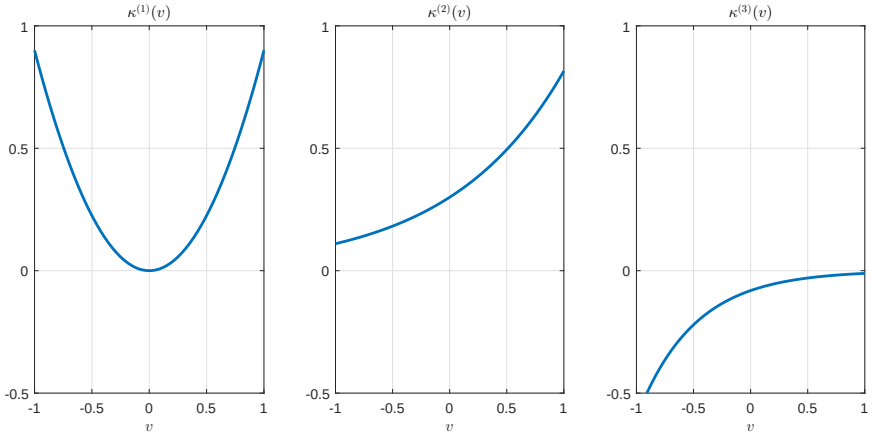


Figure 3.7. Visual representation of the three concurrent kernels.

3.5. Conclusion

This chapter has been devoted to the description of a new method to extend the seasonal ARIMA to the functional framework. Therefore, this model accounts for autoregressive and moving average terms, both with a multiplicative seasonal interaction. As opposed to other forecasting methods in the literature, the estimation procedure does not require to reduce the dimensionality of the functional time series by projecting the data into a reduced number of functional principal components. Thus, the loss of information caused by the dimensionality reduction is avoided.

The SARIMAHX model extends the standard ARIMA modeling to the functional framework by using functional linear operators in L^2. Two methods have been proposed, a fully functional linear method and a concurrent method. The first makes use of integral operators, which are defined by a bivariate kernel to model the relation between the input and the output curve. The latter is a concurrent version as the functional parameter is a univariate function acting as a coefficient for each element in the curve. The difference between the two is that the bivariate kernel makes each point in the response curve dependent of all the points of the input curve as opposed to the concurrent version, where each point in the response

Each of the functional parameters (bivariate and univariate functions of each operator) are modeled as a finite sum of sigmoid functions whose parameters are optimized for minimizing the average square forecasting error. Therefore, the shape of the sigmoids are adjusted by the algorithm to produce the best forecasts. Normalization is applied so as to help the convergence of the algorithm and an early stopping criterion is used for avoiding overfitting.

As it has been shown in the simulation case studies, this model can be used to successfully estimate functional SARIMAHX processes. The model is able to approximate the synthetic operators for the different simulated SARIMAH processes and the residuals of the estimation are similar to the real white noise processes used for the simulations.

4

Forecasting applications in electricity markets

This chapter is devoted to the empirical application of Functional Data Analysis to the electricity market field. The proposed modeling approach is evaluated on four different case studies. In each case study, the proposed SARIMAHX model of this dissertation is compared with other reference models, providing a numeric validation of the approach. The different settings of each case study allows to test the model in different environments. Firstly, an application to the Spanish ancillary services market is set up as a regression problem. Secondly, the case of forecasting electricity prices is addressed and compared with functional as well as non-functional methods. Thirdly, offer curves from the Italian electricity market are forecasted, addressing multiple-step ahead functional forecasts. Finally, Residual Demand Curves in electricity markets are forecasted. In this case, a new modeling approach is proposed that improves the definition of RDCs and then, functional forecasting is performed.

4.1. Introduction

The electric power industry in different countries has experienced a deregulation process in the last decades which has given rise to liberalized markets. They allow companies to trade energy in organized auctions. Generally, day-ahead electricity markets are based on sealed-bid auctions where companies submit their selling offers and buying bids to the Market Operator who then determines the market-clearing price and the set of accepted bids and offers for each time period. In this chapter, we will consider markets with an energy auction for each hour h of the day.

In a simple-bid market, each offer (or bid) is defined by a price p and a quantity q, which refers to the amount of energy q the agent is willing to sell (or buy) at that price p. By sorting

the selling (buying) offers in increasing (decreasing) prices, the aggregated supply (demand) function for the agent is built. Once all the agents have submitted their offer curves, the sum of the supply functions results in the system supply function $S_h(p)$ and the sum of the demand functions of each firm results in the system demand function $D_h(p)$. Market-clearing price p_h^* is computed for each hour as the intersection of the system aggregated supply and demand curves, hence $D_h(p_h^*) - S_h(p_h^*) = 0$. Then, two possible outcomes take effect depending on the auction type. In pay-as-bid auctions, prices paid to winning suppliers are based on their actual bids. On the other hand, in marginal pricing auctions each supplier is paid the market-clearing price.

Usually, each day-ahead market has subsequent smaller auctions to deal with real-time delivery of energy, such as the intraday market. Moreover, ancillary services markets are built as a way to support the transmission of electric power from seller to purchaser. The volume of energy traded in these auctions are smaller than the day-ahead market, but are not negligible when power systems are analyzed.

For a given generation company i participating in a simple-bid market, it is of utmost importance to plan ahead and manage the available resources as efficiently as possible. The forecasts of different significant market indicators and variables can provide useful estimations for the agents. For example, load forecasting allows for an efficient management of resources, optimal scheduling and production planning for minimizing generation costs [BF85]. Estimates of the price allow to plan ahead and cover their operation costs and hedge against price movements [Wer14]. Furthermore, the market agent is usually interested in optimizing its bidding strategy in the market. This can be done using Residual Demand Curves (RDCs)[Bai+04]. They can be defined for each hour as the function that models the offering and bidding behavior of all the competitors and can be calculated as

$$q = R_h(p) = D_h(p) - S_h^{-i}(p) \tag{4.1}$$

where $S_h^{-i}(p)$ is the system supply function minus the firm's supply function i.e. $S_h^{-i}(p) = S_h(p) - S_h^i(p)$. For a given price value p, R_h gives the maximum energy quantity q that the company can sell in the market at hour h.

Consequently, an electricity company interested in participating in the different markets would be inclined to used forecasting techniques to obtain estimates of load, price, supply curves or Residual Demand Curves that would allow a better management of its resources. Forecasting time series in electricity markets is challenging. The effect of business and everyday activities lead to weekly and daily seasonalities as well as peak and low demand hours. In addition, explanatory variables are important drivers of some time series such as the electricity prices. For example, weather affects the production of renewal technologies with lower generation costs influencing the offering behavior of the agents. Therefore, functional forecasting models for electricity markets should take into account these properties by considering the effect of exogenous variables and seasonality among others.

The forecasting model developed in this dissertation is very versatile, admitting a wide variety of configurations: autoregressive and moving average terms up to two seasonalities as well as the inclusion of scalar and functional explanatory variables. This allows the model to adapt

to the different necessities of each application. In order to test the versatility of the proposed models, four different real case studies are extracted from the electricity market field. Each one is briefly described ahead, detailing the source data, its application, the seasonal configuration of the SARIMAHX and the type of explanatory variables used. Table 4.1 summarizes the setup of each case study and the particular configurations of each setting.

- The first application is devoted to ancillary services. In particular, to the hourly tertiary reserve and deviation management markets. These markets account for the amount of available energy which is left unassigned before the delivery. The decisions of the agents will depend on the different values of the explanatory variables along the day, therefore, it is appealing to apply a functional approach that can forecast each hour accounting for all the daily profile of the explanatory variables. This case study is set as a long term forecasting problem. Therefore, functional regression is applied and no temporal relation is assumed for the time series. The different hourly time series (output and explanatory variables) are transformed into daily profiles, which are then used for training and evaluation of the models.

- The second application is the forecasting of the electricity price. It is an important matter for an agent that acts as a price taker. Electricity prices can be analyzed as a functional time series transforming the hourly series into a sequence of daily price profiles. This case study is set up as a forecasting problem that estimates the daily price profile for the next day. The comparison is aimed at testing the SARIMAHX model with the competing dimensionality reduction model based on Functional Principal Components as well as with the reference functional parametric model. This is a one-step ahead forecast application where the output functional time series exhibits a single weekly seasonality. Moreover, functional explanatory variables are used, which are the daily profiles of the demand and the wind power production. Forecasting electricity prices have a long history of models proposed in the literature. Therefore, a small comparison with other reference non-functional models is provided.

- The third application is on forecasting offer curves in the Italian day-ahead market. Forecasting offer curves allow for a very detailed estimation of the competitors' bidding strategy. It can be useful for optimizing the bidding strategy in markets when the demand is inelastic. The setup of the case study is based on the work [Pel13] which forecasts the supply function of the competitors of the most important agent in the Italian electricity market. As opposed to the former two case studies, the offer curves time series are not composed of an univariate process that is segmented to form a functional time series. The data are functional in itself as there exists a supply curve $S_h(p)$ for each hour which is defined over a fixed range of prices. Therefore, the hourly functional time series, which will exhibit two seasonalities (daily and weekly). Furthermore, in the real application, the 24 hourly offer curves should be forecasted at once, thus, a 24-step forecast is analyzed. For each hour, the available explanatory variables are scalar values of demand, wind, etc.

- The fourth application is dedicated to Residual Demand Curves. As was mentioned, RDCs are used by utilities in the decision making process in the electricity market. Given the RDC for an hour, the optimum amount of energy to be traded can be calculated so as to obtain the maximum profit. This methodology can be used for short-term applications

Table 4.1. Summary of the case studies performed in Chapter 4.

Market	Variable to be forecasted	Type of variable	Type of forecast	Seasonality	Explanatory variable
Ancillary Services	Tertiary and deviation management energy	Daily profile	Regression	None	Functional
Day-ahead	Electricity prices	Daily profile	1-step forecast	weekly	Functional
Day-ahead	Offer Curves	Hourly functions	24-step forecast	Daily and weekly	Scalar
Day-ahead	Residual Demand Curves	Hourly functions	24-step forecast	Daily and weekly	Scalar

such as unit commitment [GB00] or for long term planning, as can be seen in [BB11]. However, the standard definition of RDCs do not account for the effects of complex bids in the market-clearing process. Therefore, the first step is to redefine RDCs and to develop an algorithm to build RDCs which takes into account complex conditions in the market as well as international exchanges. Then, the newly defined RDCs are forecasted with functional methods. The forecasting problem shares many similarities with the offer curves case study. The data are already functional in itself and the hourly RDC time series also exhibits a double seasonal effect. Again, a 24-step horizon forecast is needed for the real application and scalar explanatory variables are available.

4.2. Application to ancillary services

The first application is devoted to the ancillary services market. A functional regression model is proposed for performing long term forecasts of the energy assigned to the Spanish tertiary and deviation management markets. The ancillary services market is an adjustment service whose aim is to resolve the technical constraints of the system by limiting and modifying the production schedules of the power stations so as to allow technical constraints to be resolved with the lowest cost for the system. In the Spanish market, the following ancillary services can be found:

- Additional Upward Reserve Power: Ancillary service whose purpose is to provide the electricity system with the necessary level of upward power reserve taking into account the available reserve in the estimated schedule of the day-ahead horizon.
- Secondary Control Band: Automatic control whose purpose is to maintain the system frequency and to correct power deviations with respect to the anticipated power exchanges. Its temporary action horizon ranges from 20 seconds to 15 minutes.
- Tertiary regulation: Its purpose is to resolve the deviations between generation and consumption and the restoration of the secondary control band reserve used. The tertiary control band reserve is defined as the maximum variation of power generation that

a generation unit can perform within a maximum of 15 minutes, and which can be maintained for at least 2 hours.

- Deviation management: Its purpose is to resolve the possible deviations between generation and consumption that could appear after the closing of each intraday session up to the start of the following session.

The total amount of energy assigned in ancillary services during 2009 was 22.501 GWh, that means the 8,95% of the total energy supplied. Day-Ahead Technical Constraints solution process (39%) and Tertiary Regulation (25%) were the ancillary services that manage more amount of energy.

The integration of large shares of wind generation requires an increase in the amount of reserves that are needed to balance generation and load. Studies described in [Ban+08] showed that large scale integration of wind generation does not create problems in terms of primary reserve levels. So, the analysis should only be considered in terms of the operating reserve management.

The primary scope of reserve predictions is to reduce balancing costs via dynamically allocation of reserve and if possible with the help of non-fossil fuel capacity. If a system operator or balance responsible party can schedule reserve more dynamically, the costs for imbalances become lower and the energy system more efficient. For lowest operational costs reserve should be shared rather than bound to specific units or pools. Reserve forecasting is therefore best applied for the full system in question with explicit forecasts of load, wind and solar, while other imbalance sources are simulated as a kind of noise term.

The modeling and forecasting of prices from the ancillary services markets is rather rare in the literature, but there are some exceptions. For instance, [Ma+04] develop neural network models for forecasting real-time LMP before and after the day-ahead market is cleared, and test them using data from the PJM and New England markets; [OS08] build a model for balancing prices at Nord Pool using combined seasonal ARIMA and discrete Markov processes.

Ancillary services are usually hard to forecast, specially, in the case of tertiary reserve and deviations. They account for the last energy that is traded to cover for imbalances, which usually shows a noisy behavior. However, a functional approach would allow to tackle the problem with a different perspective. Instead of analyzing hourly data as independent measurements, daily profiles are accounted for. Moreover, the traders usually are restricted to the units that are available and that have not been committed before. Hence, the functional approach is also justified.

In this application, the goal is to study the energy volume that is assigned in the tertiary and deviation management markets with FDA models. Generally, the outcome of both markets have the same upward or downward direction. Thus, the energy assigned will be, for each hour, of positive value if it is a selling quantity, or of negative value if it is considered a buying quantity. Let us consider as the variable of study, the hourly T&D quantity as the sum of the energies that have been assigned in the Tertiary and Deviation management markets for the same hour. The goal is to explain the T&D quantity based on some market-related explanatory variables, therefore, the problem is set as a regression problem. For each hour, the available

explanatory variables are the production classified by its technology (nuclear, Combined Cycle (CC), hydraulic, coal, wind and solar), the total electricity demand as well as the day-ahead market price.

The empirical data of the study corresponds to the added upward and downward tertiary reserve and deviation management energies in the Spanish electricity market from years 2015 to 2016. The in-sample period is set from 01/01/2015 up to 30/06/2016. The rest of 2016 is left for the Out-Of-Sample period. All variables (T&D and explanatory variables) are converted to functional variables whose domain is in the range $[1, 24]$ hours.

The models that are compared in this case study are the following:

- Naive: The naive model is set up as a simple linear regression between the explanatory variables and the T&D quantity.
- Functional Linear Model with Principal Component estimation (FLM-PC). This model is the functional linear regression described in chapter 2. Specifically, the response series $Y_t(v)$ is modeled as:

$$Y_t(v) = \sum_{z=1}^{Z} \Gamma_z(X_t^z(v))$$

  where each operator Γ_z is an integral operator in L^2, X_t^z are the explanatory variables and Y_t is the T&D series in its functional form. The estimation method consists of decomposing each operator into its spectral decomposition. Therefore, FPCA is applied to each series and components are extracted. The program used for the extraction is PACE, available in [He+10], and letting the program the automatic selection of the number of components to favor the estimation of the operator. In order to visualize this expansion, figure 4.1 shows the mean function of the response functional series, the five principal components extracted as well as the time series of scores. The same can be obtained for the functional explanatory variables, which are seen in 4.2, 4.3 and 4.4. It is worth noting that the randomness inherent to the T&D variable can be clearly seen in the principal components. While explanatory variables need 1 or 2 variables for explaining the majority of the data, the response variable needs up to five. The resulting kernel operators obtained for each explanatory variable can bee seen in figure 4.5.
- Optimized Functional Linear Model with neural networks (FLM-NN). This method is also a functional linear regression with integral operators, however, the operators' kernel are optimized with the proposed regression approach in Chapter 3. The parameters of the bivariate sigmoids are optimized so that the error is minimized on a fraction of the In-Sample data. Finally, the model that best performs with the rest of the In-Sample data is used. The resulting optimized kernels for each explanatory variable are shown in figure 4.6.

Functional errors (FMAE and FRMSE described in chapter 2.2.7) are obtained for both periods and the results are shown in table 4.2, where the three methods are compared. Diebold-Mariano test statistic [DM95] for comparing predictive accuracy on the Out-Of-Sample period is shown in table 4.3. Each cell shows the result of the test when the model of the cell's column is compared against the model of the cell's row.

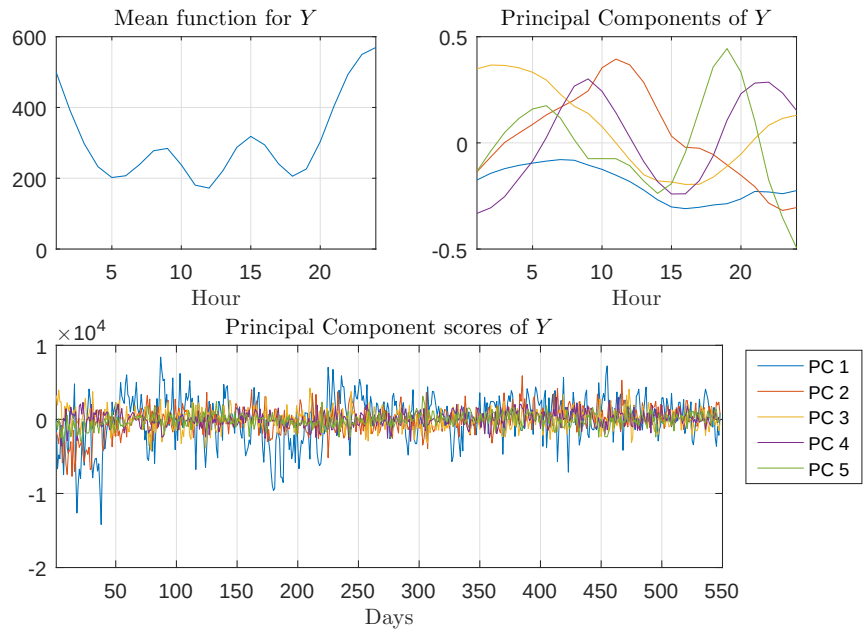


Figure 4.1. Mean function, principal components and scores extracted for the dependent variable in the Ancillary Services case study.

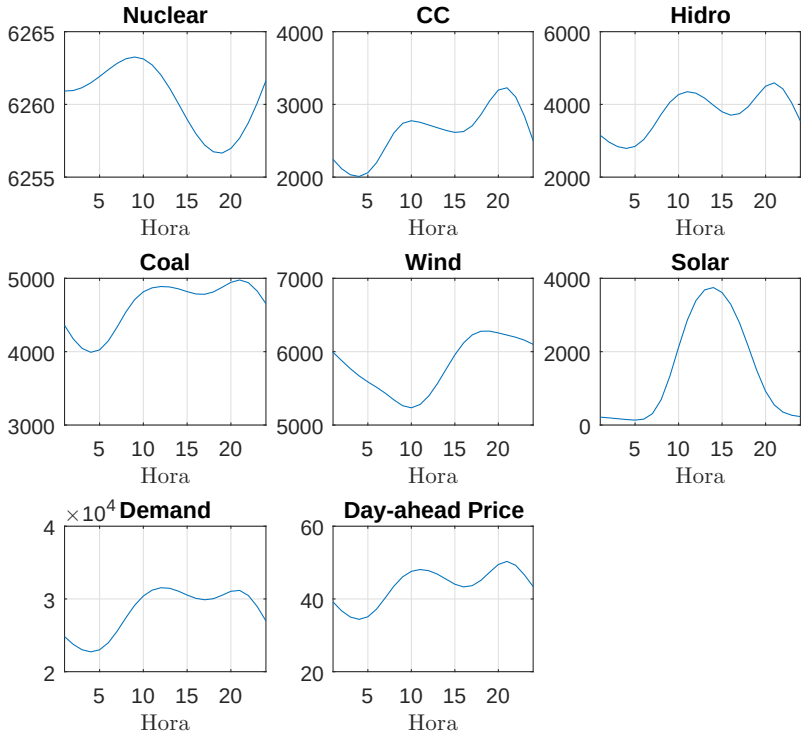


Figure 4.2. Mean functions extracted for the explanatory variables of the Ancillary Services case study.

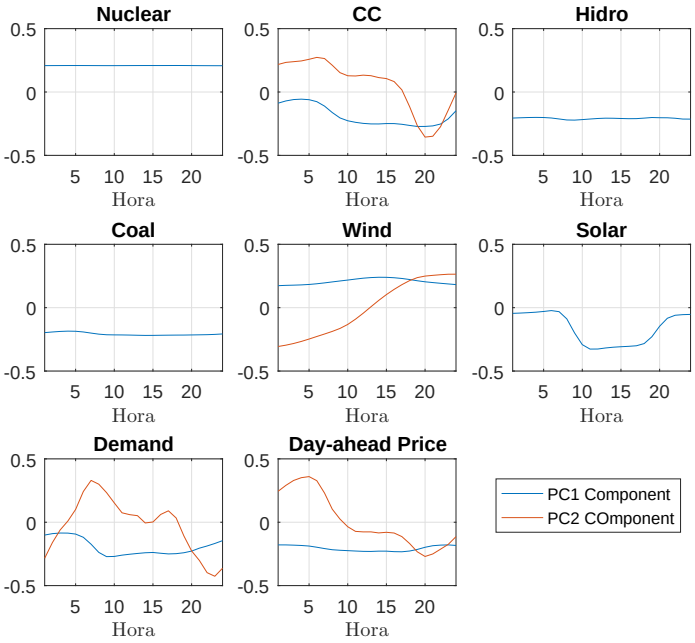


Figure 4.3. Principal Components extracted for the explanatory variables in the Ancillary Services case study.

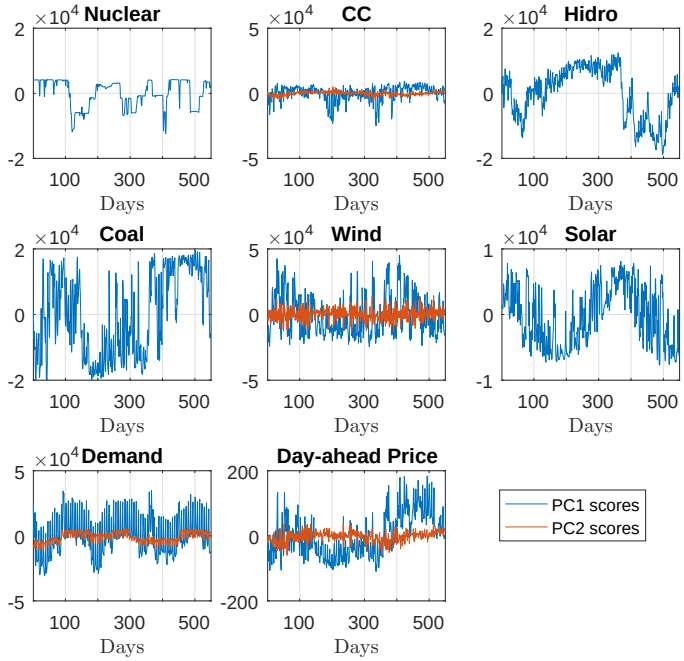


Figure 4.4. Principal component scores extracted for the explanatory variables in the Ancillary Services case study.

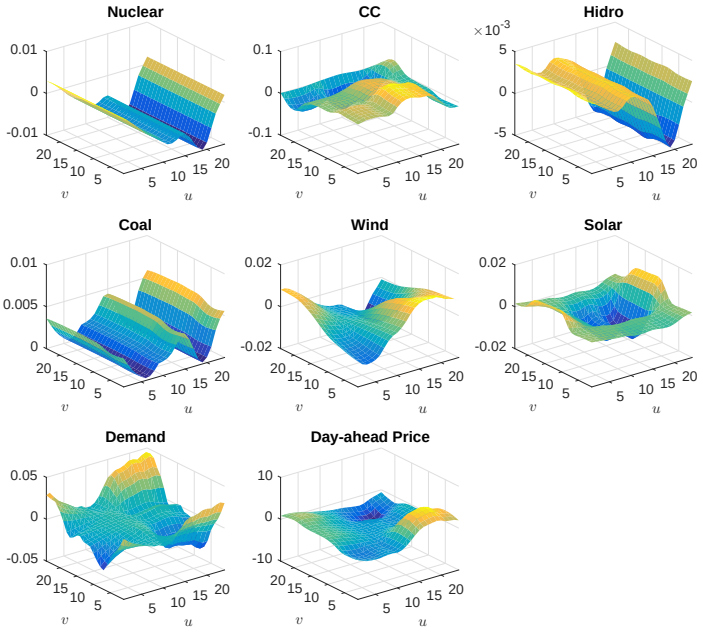


Figure 4.5. Operators' kernels for the FLM-PC

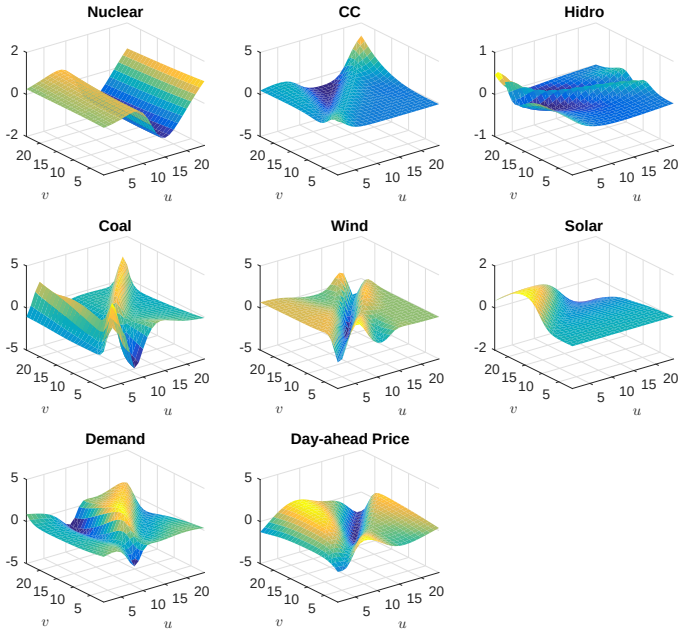


Figure 4.6. Operators' kernels for the FLM-NN

Table 4.2. Average errors for each method in the one-step ahead prediction in the Ancillary Services case study

	In-Sample		Out-Of-Sample	
Model	FMAE [MWh]	FRMSE [MWh]	FMAE [MWh]	FRMSE [MWh]
Naïve	673.80	885.77	615.05	787.56
FLM-PC	695.51	915.33	592.4	770.80
FLM-NN	**561.55**	**727.19**	**558.10**	**716.87**

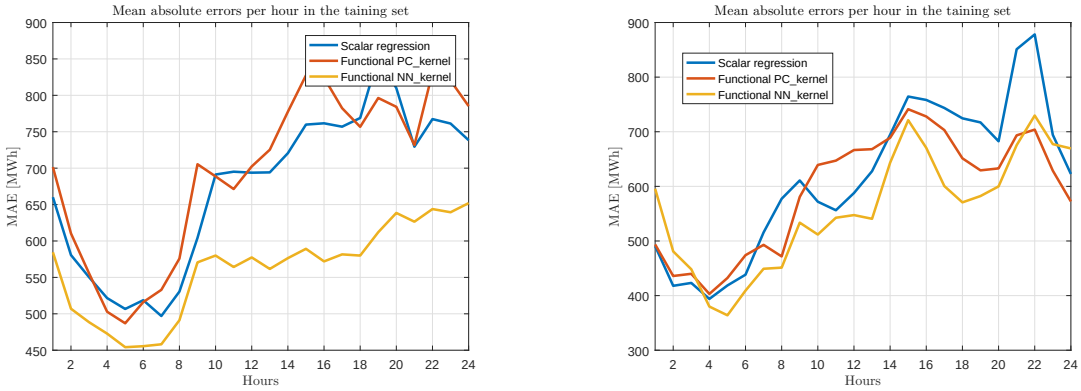
Table 4.3. Diebold-Mariano statistics pvalues for Out-Of-Sample period in the regression forecasts in the Ancillary Services case study

Models	Naïve	FLM-PC	FLM-NN
Naïve	-		
FLM-PC	0.028	-	
FLM-NN	**0**	**0**	-

Analyzing the results, it can be seen how the proposed functional model outperforms the other two competing models in both In-Sample and Out-Of-Sample periods and DM test validate the significant improvement. The FLM-PC also improves the Naïve for the Out-Of-Sample, but does not provide better results in the In-Sample. Figure 4.7 shows MAE errors per hour for both periods. The first hours of the day appear to be the easiest ones to forecast, as opposed to the later hours in the day. The proposed approach performs better in most of the hours, being in the In-Sample period where the improvement is prominent.

Figure 4.8 exhibit a comparison of the T&D time series for both periods with the corresponding forecasts. It should be remarked how the Naïve and FLM-PC models are more restrained regarding the variance of the estimation. Indeed, they appear to follow the tendency, but they behave as if they were some moving average of the series. On the contrary, the forecasts given by the proposed approach not only seem to follow the big changes, but also the variance of the estimation is more similar to the real series variance. Finally, figure 4.9 show some real daily profiles of the T&D values as well as the results of the competing models.

It is worth discussing the differences between the reference regression model and the proposed approach. The structure of both models are the same, both they differ in the estimation method. The representation capabilities of the FLM-PC is limited by the functional principal components extracted from each explanatory variable and from the output series. On the contrary, the proposed FLM-NN does not depend on some fixed functions, but the sigmoid functions that are optimized allow the operator's kernel to acquire the shape more suitable for the problem. In some cases there is not much difference between the FLM-PC kernel and the optimized kernel in FLM-NN e.g. the Nuclear Production kernel in this case study, however, usually, the difference is more accused, e.g the Wind Production kernel from both FLM models.



(a) MAE In-Sample **(b)** MAE Out-Of-Sample

Figure 4.7. Mean absolute errors for each hour and method in the In-Sample and Out-Of-Sample periods in the Ancillary Services case study

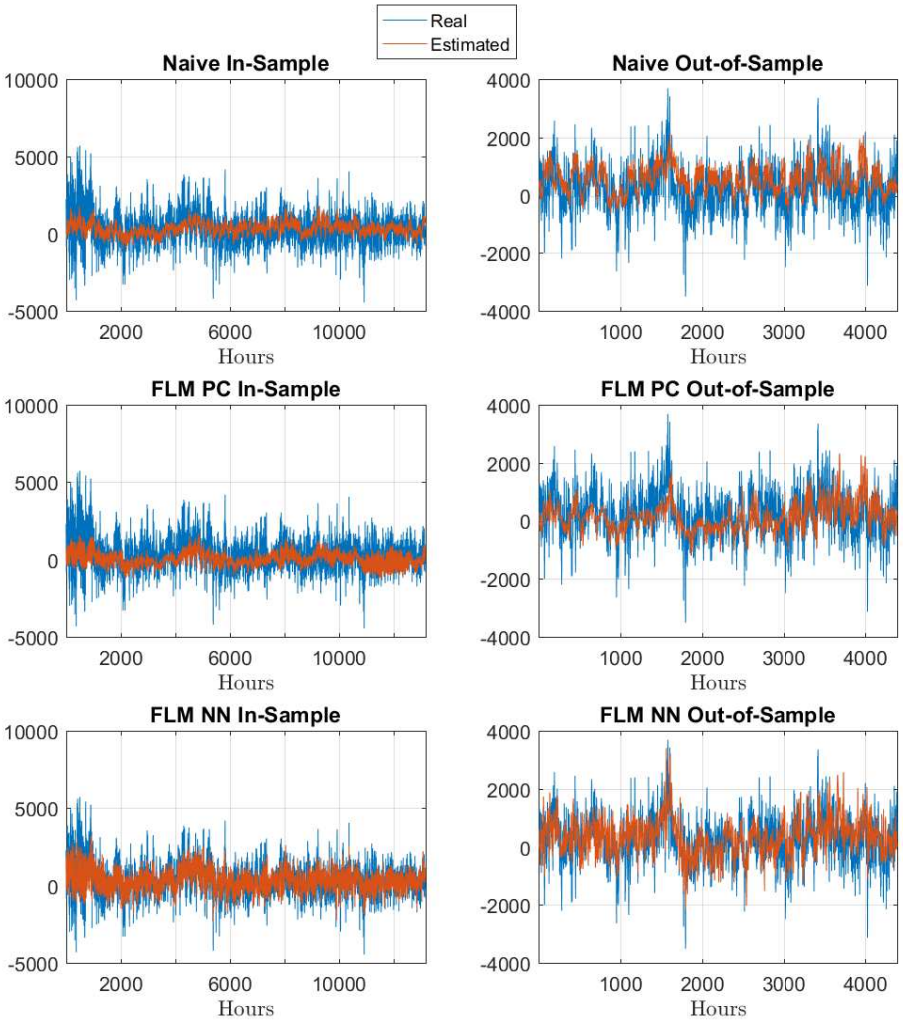


Figure 4.8. Real and estimated series in the In-Sample and Out-of-Sample for the different models trained in the Ancillary Services case study

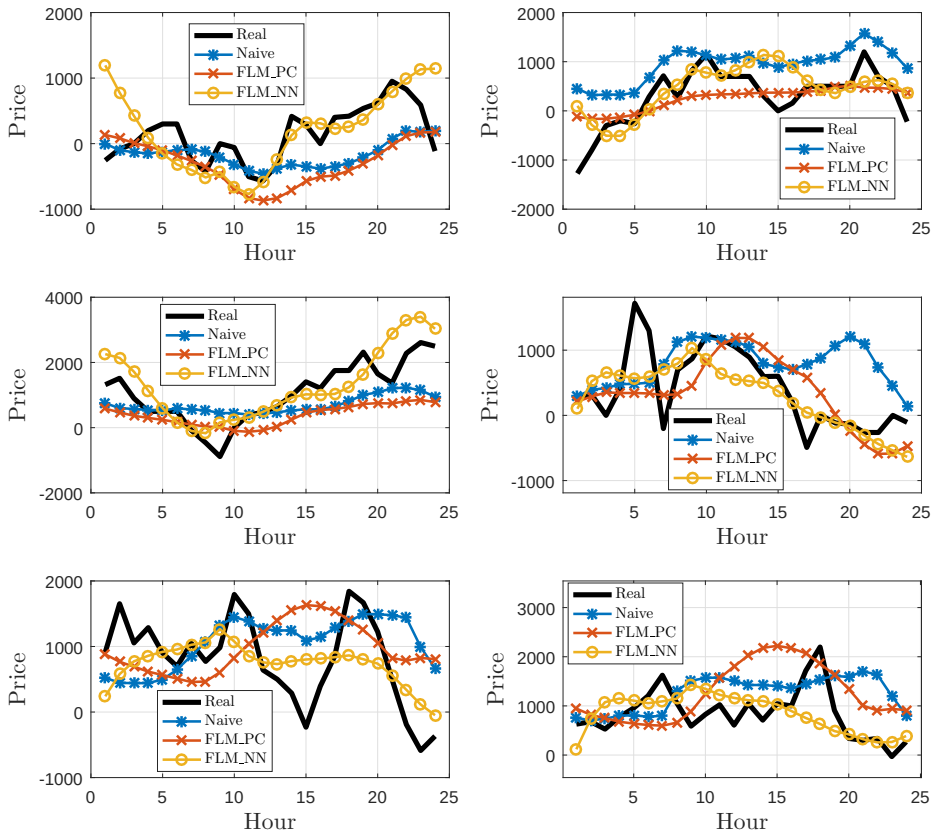


Figure 4.9. Forecast examples of the Out-of-sample period in the Ancillary Services case study

4.3. Application to electricity price forecasting

This section is devoted to the case study of forecasting electricity prices in day-ahead markets. An agent bidding in the market as a price taker would be highly interested in estimating future market prices. It allows for deciding the optimum bidding strategy and its own production or consumption schedule so that the expected profit can be maximized and the risk of trading in the market is reduced.

A wide variety of models have been developed for electricity price forecasting. An extensive review can be found in [Wer14]. Among the differnt alternatives, forecasting prices with functional methods can be found in [Vil+12] and [Lie13].

The electricity price time series can be analyzed as a functional time series by transforming the hourly series into a sequence of daily price profiles. This allows the application of functional models. Thus, the proposed SARIMAHX model is used for forecasting the the daily price profile for the next day.

Two empirical comparisons are analyzed: Firstly, functional approaches are compared. Secondly, the SARIMAHX model is compared with with other well known existing methods for

price forecasting using the same case studies that were published originally. A simplified version of the results of this chapter can be found in [Por+16a]

4.3.1. Empirical comparison with functional approaches

The performance of the proposed functional model is validated with two electricity price time series. The first case is the hourly Spanish electricity spot prices provided by the Spanish electricity Market Operator (www.omie.es). The second case is the hourly German electricity spot prices provided by the Open Power System Data [1] The time range of both datasets goes from January 1, 2014 to December 31, 2015. The case study is presented as follows: Firstly, the Spanish market data is used to compare different model settings to validate the need to include complex time dependencies in the models, e.g. moving average terms. Then, the German case compares the two best competing models using the same settings as the Spanish case.

The data set is divided into two different periods:

- In-Sample: From January 1, 2014 to December 31, 2014. This period is used to optimize the parameters of the models.
- Out-Of-Sample: From January 1, 2015 to December 31, 2015. This period is used to evaluate the generalization capabilities or forecasting performance of the model.

The spot price series $\{y_{t,h}\}$ is transformed into a functional time series $\{Y_t(v), t = 1, 2, ..., T, v \in [1, 24]\}$ where each function Y_t is observed at discrete hours $v_i \in \{1, \cdots, 24\}$, thus $Y_t(v_i) = y_{t,v_i}$. The rest of the values in the interval have been interpolated. Hence, each observation is a daily price profile. In addition, electricity demand and wind power production are relevant variables to forecast prices [Cru+11] [Jon+13]. Consequently, the daily demand and wind profiles are added to the models as exogenous functional variables $\{D_t(v), t = 1, 2, ..., T, v \in [1, 24]\}$ and $\{W_t(v), t = 1, 2, ..., T, v \in [1, 24]\}$.

Prior to the training of the functional forecasting models, it is convenient to verify that the time series are stationary. As mentioned in [Pan91], if the sample autocorrelation function of the series goes to zero slowly, it means that the series are not stationary in the mean and both the output and the input series should be differenced. Thus stationary is verified by extracting functional principal components and validating that each component's scores time series is stationary. Figure 4.10 shows the Autocorrelation functions (ACF) for the three principal component scores time series. A slow decay in the correlation terms is observed within the ACF plot on a regular basis as well as on the seasonal component. By seasonal differencing the price time series at lag 7, the new ACF no longer shows the slow decay, evidencing that the mean in the transformed time series is stationary. Demand and wind power functional time series should also be differentiated at lag 7 in accordance to the transformation of the price data [Pan91]. Hence, this provides the following sets:

$$
\begin{aligned}
Z_t(v) &= Y_t(v) - Y_{t-7}(v) \\
U_t(v) &= D_t(v) - D_{t-7}(v) \\
V_t(v) &= W_t(v) - W_{t-7}(v)
\end{aligned}
$$

[1] Accessible in http://open-power-system-data.org/, version 2016-10-28.

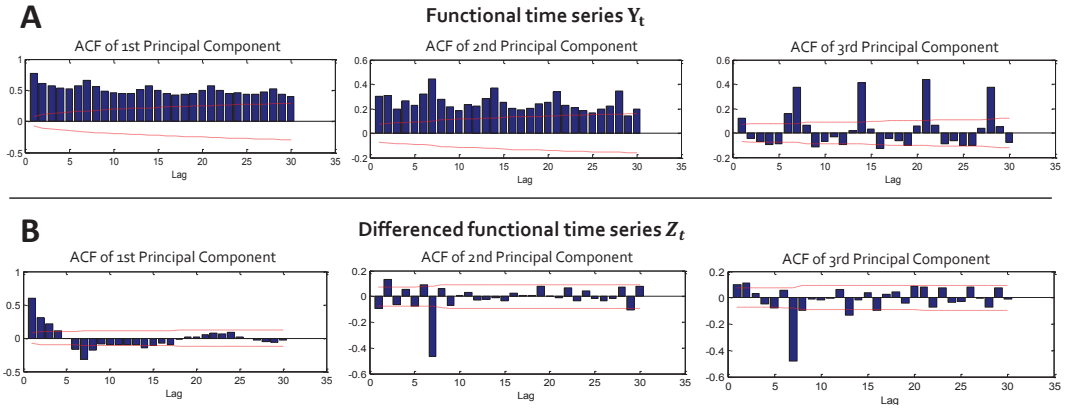


Figure 4.10. Autocorrelation Functions (ACF) of principal component scores. Part A. ACF for the time series of the three principal component scores for the price curves data Y_t. The slow decay in the correlation terms of the ACF plot evidence the need for differencing to make the series stationary. Part B. ACF of the scores of the differenced time series Z_t. The slow decay is no longer seen and a moving average effect can be clearly identified from these plots.

The different methods that are compared in the analysis are described hereafter. When possible, the models will account for the seasonal component at lag 7, due to the weekly dependency in the series shown in figure 4.10.

- Naïve: The Naïve method consists in using past observed data as the forecast. This simple method provides an initial benchmark in the comparison. Due to the fact that there is a strong weekly correlation, the forecast is obtained as

$$\widehat{Y}_t^{Naive}(v) = Y_{t-7}(v)$$

- Functional reference method: Among the functional linear methods, we have chosen the method proposed in [Bos00] and [DG05] as the reference functional model. This method has the ability to include functional covariates which are needed for this case study. Three models have been adjusted for this study, a pure regular autoregressive (ARH_DG) model, a regular autoregressive with exogenous variables (ARHX_DG) and an autoregressive model (ARHs_DG) that includes a seasonal component at lag 7 as if it were an extra explanatory variable. The models have been adjusted using the R package *far*, which provides the code for training and evaluating of the model. The details can be found in [DG05]. Therefore, each model can be expressed as:

$$\begin{aligned}
\widehat{Z}_t^{\text{ARH_DG}} &= \Psi_1(Z_{t-1}) \\
\widehat{Z}_t^{\text{ARHs_DG}} &= \Psi_1(Z_{t-1}) + \Psi_7(Z_{t-7}) \\
\widehat{Z}_t^{\text{ARHX_DG}} &= \Psi_1(Z_{t-1}) + \Gamma_1(U_t) + \Gamma_2(V_t)
\end{aligned}$$

- FPC dimension reduction: The functional dimension reduction approach is also compared. FPCA is applied to the functional time series Y using the PACE Matlab® toolbox for functional data as described in [MÖ8]. Three components are extracted which model the 98% of the variane. The scores are forecasted using univariate ARMA models. Different

models are adjusted varying the regular autoregressive order and the seasonal moving average order. The model named AR_PC uses an ARMA$(1,0)$ to forecast each score time series. ARMA_PC includes the seasonal moving average component, hence, it uses an ARMA $(1,0)\times(0,1)_7$. Models ARX_PC and ARMAX_PC have the same configuration as the previous two models but include the hourly wind production and electricity demand as exogenous variables. This is done as suggested by [Aue+15], Functional Principal Components are extracted from the wind and demand functional time series and the scores are input variables to the ARMAX models.

- Optimized Functional SARIMAHX: This is the model proposed in this paper. Four configurations have been adjusted with different structures:

$$\begin{aligned}
\hat{Z}^{\text{ARH_NN}} &= \text{ARMAH}(1,0,0) = \Psi(Z_{t-1}) \\
\hat{Z}^{\text{ARMAH_NN}} &= \text{ARMAH}(1,0,0)\times(0,0,1)_7 \\
&= \Psi(Z_{t-1}) + \Upsilon(\hat{e}_{t-7}) \\
\hat{Z}^{\text{ARHX_NN}} &= \text{ARMAHX}(1,0,0) \\
&= \Psi(Z_{t-1}) + \Gamma_1(U_t) + \Gamma_2(V_t) \\
\hat{Z}^{\text{ARMAHX_NN}} &= \text{SARIMAHX}(1,0,0)\times(0,0,1)_7 \\
&= \Psi(Z_{t-1}) + \Upsilon(\hat{e}_{t-7}) + \Gamma_1(U_t) + \Gamma_2(V_t)
\end{aligned}$$

  The number of neurons were selected by trial and error, choosing 8 neurons for each functional parameter. Each model is trained by running the optimization algorithm for 2000 cycles.

The parameters of all the models are adjusted with the In-Sample data and then evaluated in the Out-Of-Sample period. Even though the forecast of the functional models is a continuous function, we are interested in comparing the prediction against the real hourly prices. Therefore, error measurements usually found in electriciy price forecasting literature (see [Wer14]) are calculated in order to compare the models:

- Mean Absolute Error (MAE)

$$\frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{24}\sum_{v=1}^{24}\left|Y_t(v)-\widehat{Y}_t(v)\right|\right]$$

- Root Mean Squared Error (RMSE)

$$\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left[\frac{1}{24}\sum_{v=1}^{24}\left(Y_t(v)-\widehat{Y}_t(v)\right)^2\right]}$$

- Daily-weighted Mean Absolute Error (DMAE), which is used as a substitute of the Mean Absolute Percentage Error (MAPE) to avoid very high values caused by prices equal to zero:

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\sum_{v=1}^{24}\left|Y_t(v)-\widehat{Y}_t(v)\right|}{\sum_{v=1}^{24}|Y_t(v)|}$$

Table 4.4. Average errors for each method in the Spanish market.

	Model	In-Sample			Out-Of-Sample		
		MAE [€/MWh]	RMSE [€/MWh]	DMAE [%]	MAE [€/MWh]	RMSE [€/MWh]	DMAE [%]
	Naïve	8.76	11.78	23.93	8.03	10.87	18.01
Without expl. variable	AR_PC	7.62	10.15	20.51	7.04	9.29	15.55
	ARH_DG	6.59	8.9	17.44	6.11	8.13	13.46
	ARH_NN	6.53	8.81	17.37	6.09	8.14	13.47
	ARMA_PC	6.59	8.73	17.53	5.93	7.78	13.24
	ARHs_DG	6.59	8.89	17.47	6.07	8.09	13.4
	ARMAH_NN	**5.65**	**7.55**	**14.86**	**5.44**	**7.16**	**12.07**
With expl. variable	ARX_PC	5.03	6.58	13.13	4.76	6.21	10.4
	ARHX_DG	4.74	6.16	12.53	4.71	6.11	10.23
	ARHX_NN	4.77	6.21	12.65	4.64	6.02	10.11
	ARMAX_PC	4.45	5.79	11.60	4.13	5.37	9.04
	ARMAHX_NN	**4.36**	**5.68**	**11.45**	**3.92**	**5.09**	**8.64**

Table 4.4 shows the In-Sample and Out-Of-Sample errors for each model. The tested models can be classified into four groups according to their structure: Pure regular autoregressive, regular autoregessive with seasonal component, regular autoregressive with exogenous variables and regular autoregessive with seasonal component and exogenous variables.

The Diebold-Mariano (DM) test [DM95] is used to compare the predictive accuracy of the different forecasts. The DM test is a head-to-head method which compares the forecast error of two different models[2]. Table 4.5 shows the test results for the Out-Of-Sample comparison. Each cell contains the p-value that results from comparing the model of the cell's row against the model of the cell's column. A low p-value means that the null hypothesis has to be rejected and therefore the average errors shown in table 4.4 can be considered statistically different. The DM test requires that the loss differential be covariance stationary. In order to verify that assumption, a KPSS test is performed for each loss differential. Underlined values in table 4.4 correspond to those comparisons which did not pass the stationarity test and whose results should be analyzed with caution.

The main results are commented hereafter. On the one hand, the naïve method is outperformed in all cases, meaning that relevant information can be extracted from the time series. Then, looking at the average performance of the models in each group, it can be verified that when the models become more complex, the forecasting error is reduced. Models with a seasonal component outperform those that only take into account the regular autoregressive component. In addition, models that include the explanatory variable perform better than models that do not use it. These results are coherent to what is expected. Firstly, due to the seasonal differentiation, a strong moving average component appears in the series. Furthermore, the error reduction from the inclusion of the explanatory variables was expected as the electricity demand and wind power production are known to have a significant effect on the price.

[2]Diebold-Mariano tests the null hypothesis that the mean of the loss differential i.e. $d_t = L(\epsilon_{1,t}) - L(\epsilon_{2,t})$ is zero. In this study, $L(\epsilon_{i,t})$ is considered as the daily absolute error of model i. The test statistic is calculated as $DM = \bar{d}/\hat{\sigma}$, where $\bar{d}$ is the mean of the loss differential and $\hat{\sigma}$ is a consistent estimate of its standard deviation.

Table 4.5. Head-to-head Diebold-Mariano test's p-values for Out-Of-Sample period of the Spanish market case.

Models	Naïve	AR_PC	ARH_DG	ARH_NN	ARMA_PC	ARHs_DG	ARMAH_NN	ARX_PC	ARHX_DG	ARHX_NN	ARMAX_PC	ARMAHX_NN
Naïve	-											
AR_PC	0	-										
ARH_DG	0	0	-									
ARH_NN	0	0	0.713	-								
ARMA_PC	0	0	0.285	0.342	-							
ARHs_DG	0	0	0.023	0.733	0.392	-						
ARMAH_NN	**0**	**0**	**0**	**0**	**0**	**0**	**-**					
ARX_PC								-				
ARHX_DG								0.581	-			
ARHX_NN								0.174	0.205	-		
ARMAX_PC								0	0	0	-	
ARMAHX_NN								**0**	**0**	**0**	**0**	**-**

Table 4.6. Average errors for the two competing methods for the german market.

Model	In-Sample MAE [€/MWh]	In-Sample RMSE [€/MWh]	In-Sample DMAE [%]	Out-Of-Sample MAE [€/MWh]	Out-Of-Sample RMSE [€/MWh]	Out-Of-Sample DMAE [%]
Naïve	7.45	10.76	26.94	8.65	11.9	32.51
ARMAX_PC	4.02	5.68	13.68	4.50	6.21	16.01
ARMAHX_NN	**3.65**	**5.18**	**12.49**	**4.15**	**5.64**	**14.59**

The different methods in each group are compared. All the pure regular autoregressive models have similar performances as seen by the Diebold-Mariano test in Table 4.5, however, the seasonal methods without explanatory variables are significantly different and we can see how the moving average component is justified, being the method presented in this paper the one that provides the best results. When considering the group of autoregressive models with the exogenous variables, Table 4.5 shows that all three methods perform alike. Finally, the last group only consists of two models. It can be seen how the ARMAHX method is further improved by adding the moving average term and the improvement is statistically validated with the p-values obtained in Table 4.5.

The two most competing models in the Spanish case (ARMAX_PC and ARMAHX_NN) have also been applied to forecast the German electricity price. The results are shown in table 4.6. The proposed method shows better results for both In-Sample and Out-Of-Sample periods and the Diebold-Mariano test yielded a p-value equal to zero in the comparison of both forecasts. In addition, table 4.7 shows the average errors of the two competing models for each day of the week for both market cases. The proposed method outperforms the reference model for each day.

Table 4.7. Out-Of-Sample Errors for the two competing methods for each day of the week in the Spanish and the German market.

Market	Model	Error	Mon	Tues	Wed	Thu	Fri	Sat	Sun
Spanish	ARMAX_PC	MAE [€/MWh]	4.43	3.8	4.4	3.82	3.47	4.61	4.33
		RMSE [€/MWh]	5.71	4.88	5.68	4.84	4.51	6.07	5.73
		DMAE [%]	8.89	7.58	8.94	7.74	6.86	10.96	12.33
	ARMAHX_NN	MAE [€/MWh]	**4.14**	**3.68**	**4.16**	**3.73**	**3.44**	**4.14**	**4.18**
		RMSE [€/MWh]	**5.18**	**4.7**	**5.38**	**4.81**	**4.45**	**5.43**	**5.62**
		DMAE [%]	**8.23**	**7.39**	**8.47**	**7.67**	**6.85**	**9.9**	**12.04**
German	ARMAX_PC	MAE [€/MWh]	5.17	4.65	3.95	4.38	4.17	4.07	5.07
		RMSE [€/MWh]	7.09	6.08	5.2	5.71	5.58	5.22	8.06
		DMAE [%]	15.97	13.89	11.79	13.63	13.27	16.00	27.15
	ARMAHX_NN	MAE [€/MWh]	**4.81**	**4.17**	**3.86**	**3.95**	**3.92**	**3.82**	**4.52**
		RMSE [€/MWh]	**6.42**	**5.39**	**5.02**	**5.05**	**5.17**	**4.89**	**7.16**
		DMAE [%]	**14.7**	**12.52**	**11.66**	**11.98**	**12.27**	**14.99**	**24.01**

In order to understand the results from the proposed method, Figures 4.11 shows the optimized kernels for the trained ARMAHX$(1,0) \times (0,1)_7$ model in each country. These models were defined by four functional operators: two for the explanatory variables, one for the regular autoregressive term and one for the seasonal moving average term. Each of these surfaces is formed by summing up 8 bivariate hyperbolic tangent functions, which are the output of the optimization process. They show the influence that each input variable has on the output variable. The u axis represents the input variable and the v axis, the output function. The autoregressive kernel shows that all the hours of the price profile are very much dependent on the value at hour 24 of the price profile of the previous day. The moving average kernel shows high values in the diagonal, which means that each hour of the output is influenced by the error at that hour 7 days ago. The wind explanatory variable kernel shows that there is a strong negative dependence on the diagonal, which means that when the wind power is high for one hour, prices will drop at that hour and viceversa. In addition, the demand operator shows a positive dependency on the diagonal.

Figures 4.12 shows forecasting examples for different weekdays for each country. The real price profiles are compared with the estimation provided by the best competing models.

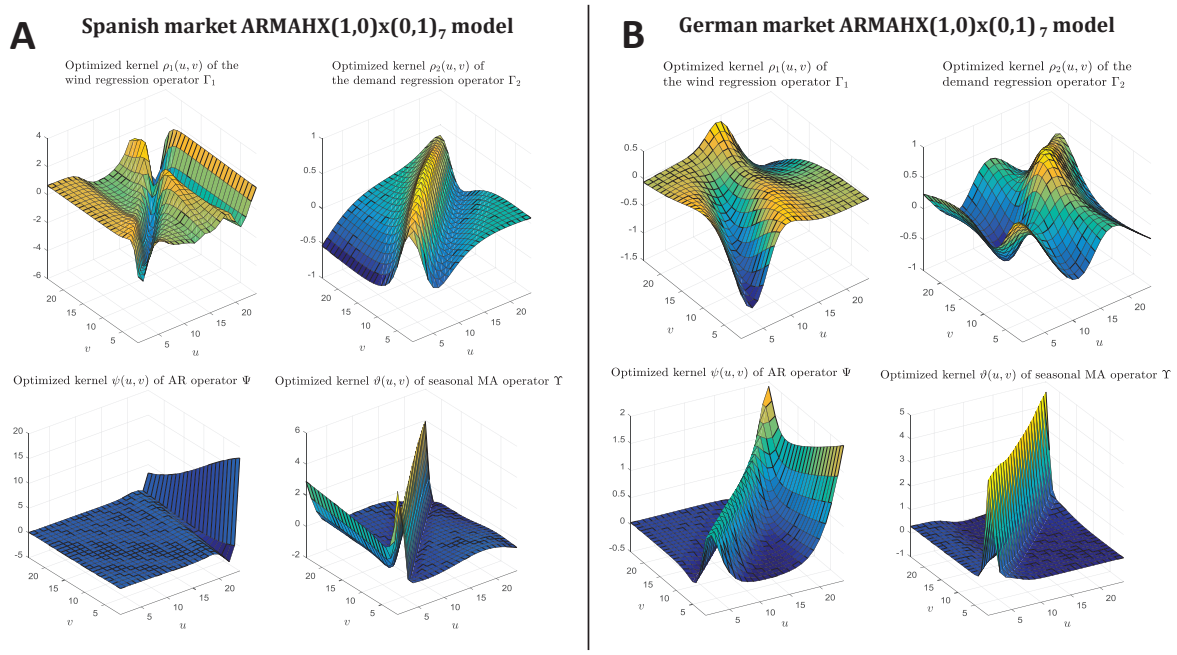


Figure 4.11. Estimated kernel functions of the operators of the best ARMAHX models. Part A. Kernel functions for the model trained with the Spanish case. Part B. Kernel functions for the model trained with the German case.

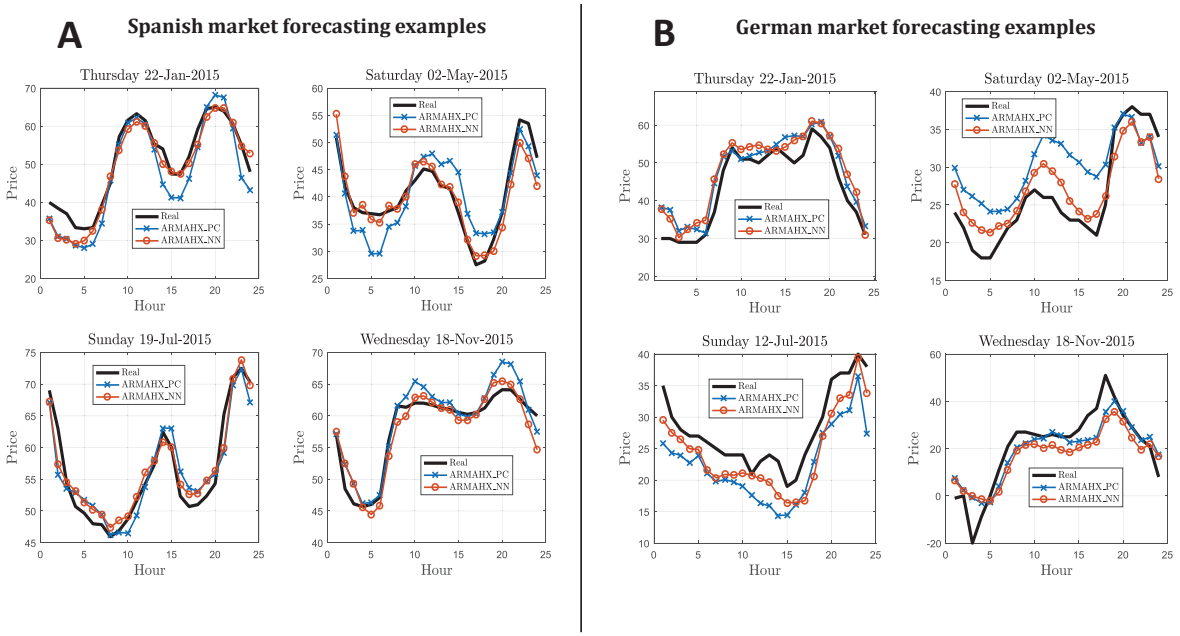


Figure 4.12. Real price profiles and corresponding forecasts of the two competing models. Part A. Four days from the Spanish market case. Part B. Four days from the German market case.

4.3.2. Empirical comparison with non-functional approaches

This dissertation proposes a functional forecasting model which can be applied to diverse problems in electricity markets. Nevertheless, in the case of forecasting the hourly electricity price, non-functional approaches have been traditionally used in the literature. It is out of the scope of this case study to perform a detailed comparison of the proposed model with the latest price forecasting methods; however, we have included an empirical comparison of the performance of the proposed model with other well known existing methods using the same case studies that were published originally.

The price forecasting setup in [Cru+11] has been selected for the empirical comparison. There, the Spanish day-ahead electricity price of years 2007 and 2008 is forecasted by a wide variety of models, ranging from MLP Neural Network to dynamic regression and periodic models. The latter methods have been used extensively in price forecasting (see [NC06] and [Koo+07] respectively). Seasonality as well as load and wind explanatory variables are included in the models.

A competing functional ARMAHX model is adjusted with the same in-sample period used in [Cru+11] (01/01/2007 to 10/08/2007) and validated with the same Out-Of-Sample period (11/05/2007 to 31/07/2008). The hourly electricity spot price, load and wind energy are transformed to functional time series as explained in section 4.3.1. Firstly, both explanatory variables and the price functional time series are differenced at lag 1. Then, a SARIMAHX$(0,0,1) \times (1,0,0)_7$ is adjusted for the differenced time series with 5 sigmoids for each operator.

Table 4.8 shows monthly MAPE errors for the MLP, the Dynamic Regression model (DR), the periodic model (PER) and the proposed functional model. The results exhibit a very good performance of the ARMAHX model for the Out-of-sample average error.

Consequently, the proposed functional approach can provide competent forecasts against other existing methods in the literature. Moreover, it should be remarked that functional methods are a recent field of research which will be further developed with new features that will provide better competing models.

One of such improvements could be the extension of intervention analysis [BT75] to the functional framework, which would allow to model the effect of a sudden change in the time series to be forecasted. Dummy intervention variables of different nature can account for the occurrence of an event that affects the response time series, e.g. a change in the regulatory policy would be modeled as a step function, or a strike by means of a pulse function. These intervention variables could be added as explanatory variables to the functional model.

Table 4.8. Out-Of-Sample MAPE for the case study and models proposed in [Cru+11]. Functional ARMAHX results are included.

MONTH	MAPE [%]			
	MLP	DR	PER	ARMAHX
2007/08	9.61	8.70	8.30	**7.83**
2007/09	7.04	6.61	**6.35**	6.76
2007/10	8.33	8.09	8.16	**7.67**
2007/11	9.07	8.43	8.19	**7.40**
2007/12	11.54	**9.10**	9.18	9.20
2008/01	9.13	7.87	**7.35**	7.65
2008/02	8.38	6.75	**6.30**	6.98
2008/03	7.36	7.69	7.63	**5.68**
2008/04	6.29	6.33	6.02	**5.85**
2008/05	6.35	6.11	6.21	**5.47**
2008/06	5.87	5.27	5.32	**5.17**
2008/07	4.83	4.88	5.09	**4.40**
TOTAL	7.77	7.12	6.98	**6.65**

4.4. Forecasting Offer curves

This section is devoted to the case study of forecasting offer curves in electricity markets. This is very useful for electricity companies that want to foresee the competitors' behavior and optimize their bidding strategy. A supply function $S_h(p)$ represents for each price p the total amount of energy offered by an agent in the auction of hour h . If the demand is inelastic at , the residual demand of equation (4.1) becomes

$$q = D_h - S^{-i}(p)$$

Therefore, the problem of forecasting the residual demand could be separated into the forecast of the hourly demand and the forecast of the offer curves. Demand forecasting is widely studied in the literature, hence, this chapter addresses the forecasting of supply functions with a functional approach.

[Pel13] analyzes the forecasting of supply functions of competitors in the Italian electricity market using functional models. The set up of this case study is influenced by that published work. The Italian electricity market is a complex system which belongs to the European market. There exists 6 interconnected zones where buying bids and selling offers can be placed in the market. If the exchange limits between zones are not saturated, the supply and demand functions of each zone are aggregated and a single clearing price is obtained. However, if congestion in the network occur, the market can split into different areas, each with a different clearing-market quantity and price.

A company with generation units in all the zones might want to optimize the bidding strategy for the whole portfolio, and thus, predictions of each zonal competitors' supply curve would be necessary. However, for simplicity, and following [Pel13], the Italian market is analyzed as if there were no saturation constraints between zones and the 6 areas are considered as a single

Italian market. The detailed research of each zone would be of much interest, but is out of the scope of this dissertation.

Consequently, offer curves for the competitors of Enel, a mayor electricity company in Italy, are obtained by summing all the zonal competitors' offer curves. Offer curves are limited to the price range $[0, 200]$ €/MWh. It should be remarked that the functional time series to be forecasted does not correspond to a time series which has been obtained from a univariate scalar time series, as was the case of the electricity prices. Each functional observation is the hourly aggregated offer curve submitted to the market. As a consequence, the offer curves time series will exhibit two significant seasonalities: daily and weekly.

The setup of the real case study is as follows. Similar to [Pel13], Enel competitors' are considered and their offer curves are forecasted. The time range of the data is from 01/03/2015 up to 28/02/2016 thus, consisting of 8784 curves. The data is divided into two sets: The In-Sample period is considered from 01/03/2015 to 31/08/2015, which will be used for training the models. The rest is left for the Out-Of-Sample period.

The offering behavior of the agents is conditioned to the weather and the particular circumstances of each day. For that reason, explanatory variables are used as a way to account for the external factors that might influence the traders decision. The explanatory variables used in this case are the following:

- Total demand of Italy. The demand is of utmost importance to account for the consumption of energy in the country.
- Total wind power production. The south of Italy hosts a great number of wind farms which have a significant impact on the offer curves in windy days.
- Total solar production. The north of Italy is the region where most solar power capacity installed. Therefore, the solar production should be significant.
- Thermal availability. It is the sum of all the energy offered in the market by thermal units.
- Energy exchanged from Italy to to the adjacent European countries: France, Switzerland, Austria, Slovenia, Greece and Malta. These exchanges play a very important role in the energy trading of the country.

As the output time series are hourly curves and these explanatory variables are hourly values, the model will consider them as scalar covariates and not functional covariates.

It should be remarked that in most of the markets, the auctions of the 24 hours for the following day are cleared at the same time. Therefore, two different analysis are analyzed. On the one hand, a one-step case study is analyzed. The proposed model is trained to minimize the one-step ahead forecast error, then, it is reasonable to study this theoretical comparison. On the other hand, a 24-step ahead forecast is also evaluated. This serves two main purposes. It shows the adaptation of the model to a variable horizon forecast as well as validating the use in a real case application.

The models to be compared are described and analyzed:

- Naïve. This is a simple benchmark model that is used to obtain a simple reference forecast and provides a reference to compare the rest of the models. Two versions are used, depending whether the simulation is one-step or 24-step forecast. In the first case, the forecast is simply the last curve observed in the data, i.e. the curve from the previous hour. The second case, the forecast takes into account that the curves of the same day are not observed. Then, the forecast will depend on the type of weekday. For Saturdays, Sundays and Mondays, the forecast will be the hourly curve of previous week, while for Tuesdays, Wednesdays, Thursdays and Fridays, the forecast will be the hourly curve of the previous day.

- Principal component dimension reduction approach. This method extracts the Functional Principal Components of the curves and the corresponding scores time series. Then, the scores are forecasted by means of Transfer Function (TF) models, which include explanatory variables. The final estimation of the curves is done by reconstructing the estimated scores. This method is also used in [Pel13], although it uses a simpler time series model.

  The In-Sample period is used for all the training. FPCs are extracted for that range and the parameters of the TF are estimated. In the Out-Of-Sample period, the component scores are obtained, by projecting the curves into the basis spanned by the FPCs previously extracted.

  For illustration purposes, in the one-step ahead case the ideal prediction results are shown. Instead of resconstruncting the forecasted scores, the ideal forecast is the reconstruction of the real scores. In [Bos00] and [Kle+16], they make reference to this model as the best prediction model. Indeed, once the functional time series is projected on to a functional basis, the best possible forecast is to reconstruct the curves according to the real scores. This model could never be applied in real life as the real scores of future hours or days are unknown, but it is a meaningful comparison that illustrates the amount of information that is lost in the dimension reduction procedure.

  Figure 4.13 represent the mean function, the principal components and the scores obtained for the curves time series. Three and four principal components are extracted, which explain the 98% and 99% of the variance of the data respectively. By analyzing the FPCs it can be observed how the first FPC represents the level of the curve, as it is mostly flat. This is the component that best represent the data, and that is why the first scores time series is significantly greater that the rest. The interpretation of the rest of the scores is less intuitive. The second FPC, for example, shows that when energy in the range up to price 50 €/MWh increases, the values for prices higher than 70 €/MWh decrease.

- Functional SARIMAHX. The proposed functional and concurrent time series models are trained with the In-Sample data of this case study. In both cases, the explanatory variables are scalar hourly values thus, the univariate sigmoid kernel is used. The resulting shapes of those functional parameters are shown in figure 4.14. Analyzing these shapes, the effect of each variable on the offering curve can be seen. The demand has positive values for each price, meaning that for higher demand, the offer curve has more volume of energy offered. In fact, the offer increases more at higher prices than at lower prices, meaning

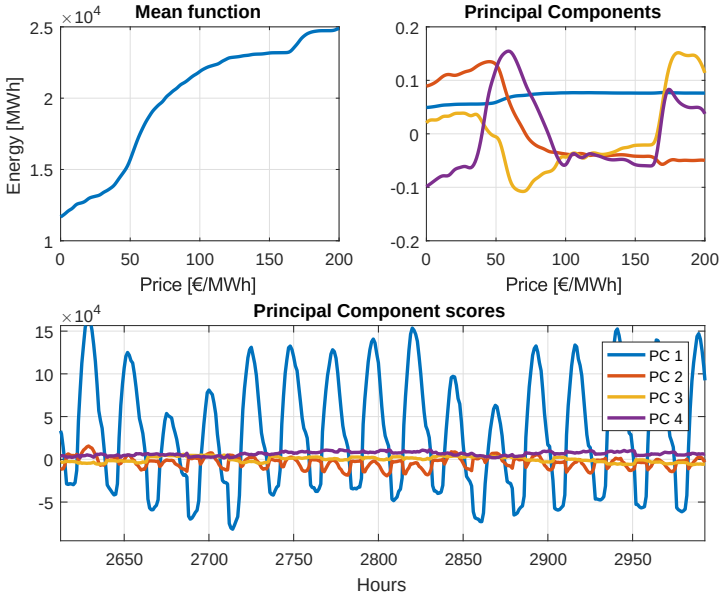


Figure 4.13. Mean function, Principal Components and scores of the offer curves

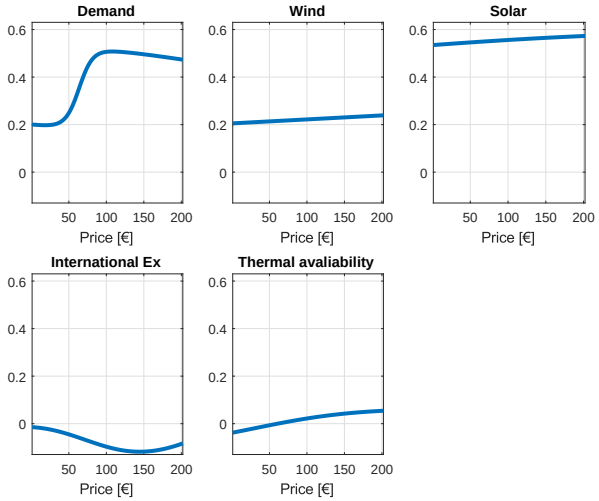


Figure 4.14. Operators' kernels for the regressors of the SARIMAHX model in Offer Curves forecasting study

that the increase in demand is usually covered with more expensive generation. Wind and solar production have somewhat flat coefficients. Usually, these renewal production is offered at price 0, thus, simply displacing the curve up or down. The international exchange, on the other hand, has negative values. This means that when there are more imports, the offer curve has less energy being offered. Once the regressive terms are adjusted, the dynamic terms for the concurrent and functional versions are adjusted. Several models were trained and the ones with less training error are described ahead:

– Concurrent SARIMAHX: This model is the concurrent version proposed in section 3. The final adjusted model was a concurrent SARIMAHX$(2,0,0) \times (1,0,1) \times (1,0,1)$.

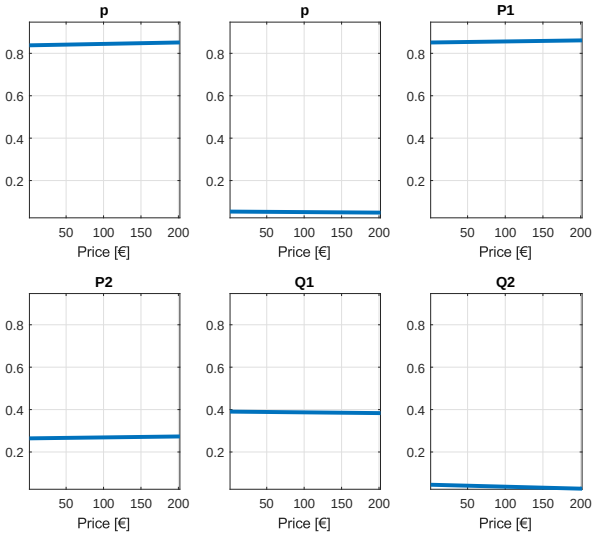


Figure 4.15. Operators' kernels for the concurrent SARIMAHX model in Offer Curves forecasting study

Figure 4.15 shows the kernels trained for the functional models. In this case, these kernels are somewhat linear, indicating that the temporal behavior does not change much between prices. It can be seen how the first regular and the first seasonal autoregressive terms take high values with respect to the other functional parameters, meaning that they are the most relevant coefficients in the model.

– Functional SARIMAHX: This is the fully functional forecast model described in section 3. Again, the final adjusted model was a SARIMAHX$(2,0,0) \times (1,0,1) \times (1,0,1)$. Figure 4.16 shows the kernels trained for the functional model. Analyzing the shape of the bivariate kernels, the first regular and first seasonal functional coefficients are the most significant as they show high values on the diagonal. This means that not much relation is found between different parts of the curve and the behavior of the offer curve at some price is mostly dependent on previous curves at that same price. If the shapes of all functional coefficients had all high values in the diagonal, the model would be equivalent to the concurrent functional model.

All the models are trained with the In-Sample period. Then, each one produces a 1-step ahead forecast, assuming that the curve of last hour is known, and a 24-step ahead forecast, where the 24 hours of the following day are forecasted being the last observed curve the hour 24 of the current day. Functional errors FMAE, FRMSE and FMAPE from section 2 are calculated. Table 4.9 and table 4.11 show the functional errors for the 1-step and the 24-step ahead forecasts. Each table shows In-Sample as well as Out-Of-Sample errors. In addition, tables 4.9 and 4.13 show the Diebold-Mariano tests applied to each pair of forecasts. Table 4.10 show the results with the ideal Principal Component models.

The estimation results are analyzed. Firstly, by looking at the 1-step ahead forecasts, it can be seen how the functional approach outperforms the other methods in both the In-Sample and Out-Of-Sample periods. Moreover, the concurrent functional approach is the one that provides the best results. The most relevant functional coefficients seem to give less importance to cross

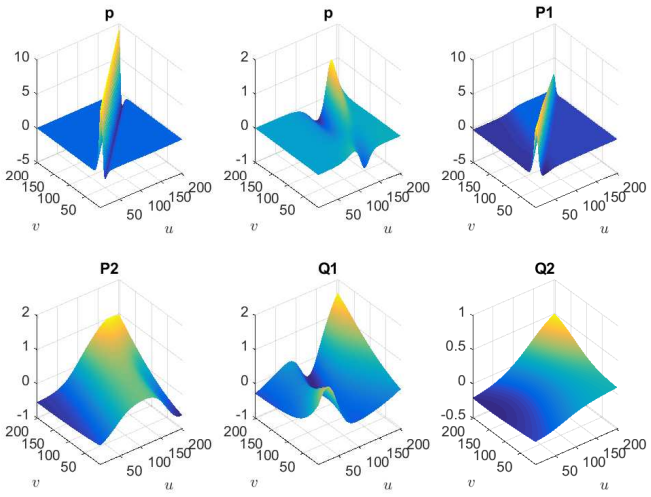


Figure 4.16. Operators' kernels for the SARIMAHX model in Offer Curves forecasting study

correlations among the price axis. Therefore, the adjustment of the parameters might be easier for the concurrent model than for the fully functional model and in this case that simpler model adjusts better to the functional series. DM test validate the Out-Sample results, yielding very low p-values between model comparisons. Then, the equal forecast null hypothesis can be rejected.

In addition, it is worth to compare the results against the ideal principal component prediction. As can be seen, on the In-Sample period, the ideal method significantly outperforms the functional concurrent model, as it is expected. However, for the Out-Of-Sample period, the errors are quite similar. In fact, the SARIMAHX functional model outperforms the 3PC and 4PC ideal models. This implies that the error incurred only by reducing the dimensionality of the series is already higher than the SARIMAHX forecast. Then, the estimation of the scores can only yield worse results in that period. The reason is that the Principal Components have been extracted with the In-Sample period. The In-Sample reconstruction error is lower, however, on the Out-Of-Sample, the Principal Components are kept untouched, thus they cannot adapt to a significant change in the functional time series and the perfect reconstruction looses precision. On the contrary, the proposed functional model does not rely on any basis expansion of the series and it takes into account the whole curve values from the recent past. Therefore, it can better adapt to changes in the series.

Secondly, the 24-step forecasts show similar results as the 1-step ahead case. The concurrent functional SARIMAHX model provides better average estimations with respect to the other model. Globally, 24-step ahead errors are much higher than 1-step ahead errors, i.e. around 4% error for the 24-step and around 2% error for the 1-step. Offer curves are the result of aggregating the offers of competitors in the market. These offers respond to the different strategies of the agents which, in many cases, have a common behavior for the hole day e.g. There is some restriction on a power plant for that day. Therefore, strategies are planed for the whole day, and consequently, once the curve for the first hour is known, it helps to improve significantly the forecast of the following hours. This can be seen in figure 4.18. Forecasting

Table 4.9. Average errors for each method in the one-step ahead prediction in Offer Curves forecasting study

Model	In-Sample			Out-Of-Sample		
	FMAE [MWh]	FRMSE [MWh]	FMAPE [%]	FMAE [MWh]	FRMSE [MWh]	FMAPE [%]
Naïve	909.39	1258.46	4.50	818.59	1180.16	4.18
ARIMA 3 PC	353.85	470.51	1.83	479.51	690.59	2.48
ARIMA 4 PC	276.65	378.83	1.46	415.06	565.38	2.19
Con SARIMAHX	**255.19**	**365.73**	**1.34**	**292.40**	**409.30**	**1.56**
Fun SARIMAHX	268.57	375.24	1.41	309.34	424.81	1.64

Table 4.10. Average errors for the ideal principal component forecast in Offer Curves forecasting study

Model	In-Sample			Out-Of-Sample		
	FMAE [MWh]	FRMSE [MWh]	FMAPE [%]	FMAE [MWh]	FRMSE [MWh]	FMAPE [%]
ARIMA 3 PC ideal	276.10	369.41	1.40	390.94	607.09	1.99
ARIMA 4 PC ideal	175.67	236.32	0.91	319.11	454.27	1.66

the first hour is very important to the rest of the prediction, specially for the first hours of the day.

To complement the results and provide more understanding on the different forecasts, figure 4.17 presents MAE errors for each price for the In-Sample and Out-Of-Sample periods for the 1-step and 24-step ahead forecasts. The figures show how each method estimates in the different range of prices. It can be seen how the the SARIMAHX models yield smoother errors across prices. Again, as the proposed models do not depend on the FPCs, it can adapt better to the bidding steps in the curve. In addition, the SARIMAHX yields lower errors for prices under 100 €/MWh, which is the zone of interest. It is worth stopping at figure 4.19, which shows MAE results for each day of the week. To end up, Figures 4.20 and 4.21 show some forecasting examples that illustrate how are the curves affected by each model.

Table 4.11. Average errors for each method in the 24-step ahead prediction in Offer Curves forecasting study

Model	In-Sample			Out-Of-Sample		
	FMAE [MWh]	FRMSE [MWh]	FMAPE [%]	FMAE [MWh]	FRMSE [MWh]	FMAPE [%]
Naïve	1254.62	1845.61	7.71	1241.69	1826.60	7.83
ARIMA 3 PC	672.06	871.33	3.41	855.77	1146.97	4.43
ARIMA 4 PC	**639.47**	**834.46**	**3.26**	810.98	1085.07	4.22
Con SARIMAHX	666.43	879.32	3.43	**761.79**	**990.0**	**3.97**
Fun SARIMAHX	687.45	897.05	3.51	798.68	1045.43	4.12

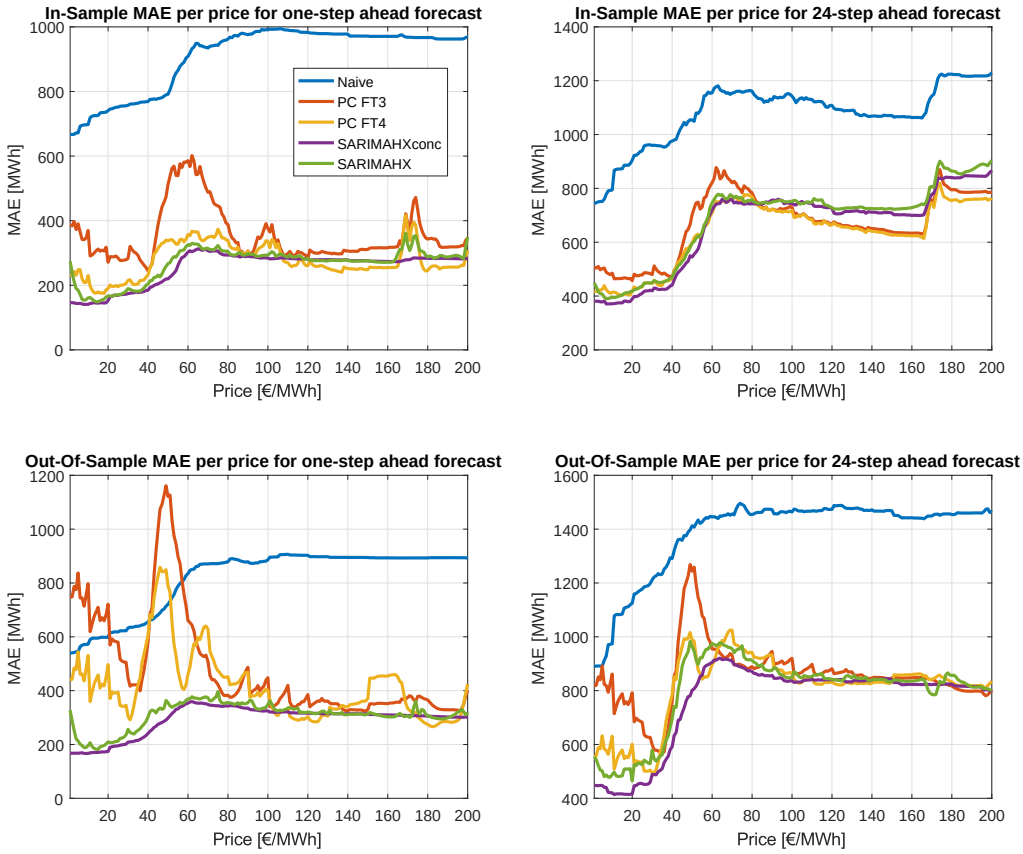


Figure 4.17. MAE for each price in the 1-step and 24-step ahead forecasts in the In-Sample and Out-Of-Sample periods in Offer Curves forecasting study

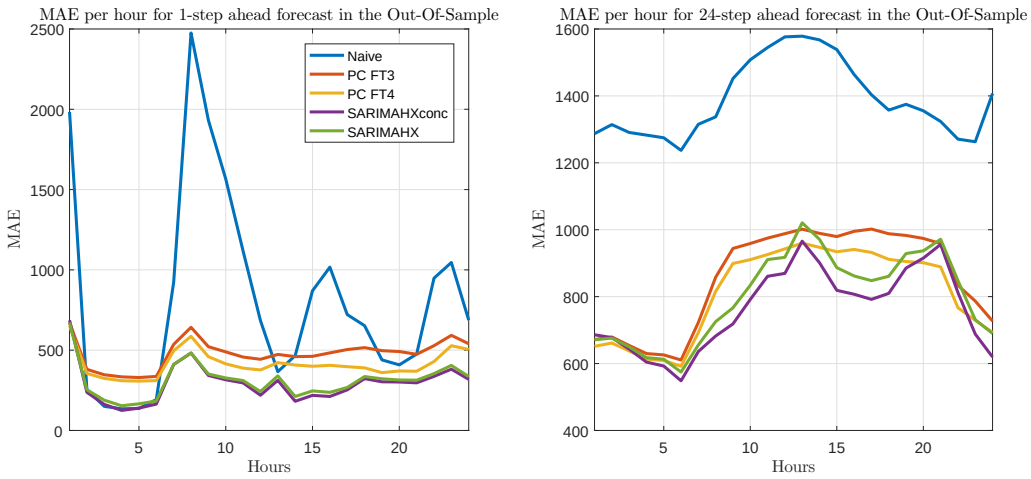


Figure 4.18. MAE for each hour in the 1-step and 24-step ahead forecasts in the In-Sample and Out-Of-Sample periods in Offer Curves forecasting study

Table 4.12. Diebold-Mariano test's pvalues for Out-Of-Sample period in the one-step ahead forecasts in Offer Curves forecasting study

Models	Naïve	ARIMA 3 PC	ARIMA 4 PC	Con SARIMAHX	Fun SARIMAHX
Naïve	-				
ARIMA 3 PC	0	-			
ARIMA 4 PC	0	0	-		
Con SARIMAHX	**0**	**0**	**0**	-	
Fun SARIMAHX	0	0	0	**0**	-

Table 4.13. Diebold-Mariano test's pvalues for Out-Of-Sample period in the 24-step ahead forecasts in Offer Curves forecasting study

Models	Naïve	ARIMA 3 PC	ARIMA 4 PC	Con SARIMAHX	Fun SARIMAHX
Naïve	-				
ARIMA 3 PC	0	-			
ARIMA 4 PC	0	0	-		
Con SARIMAHX	0	0.0005	**0.053**	-	
Fun SARIMAHX	0	0 .026	0.620	0	-

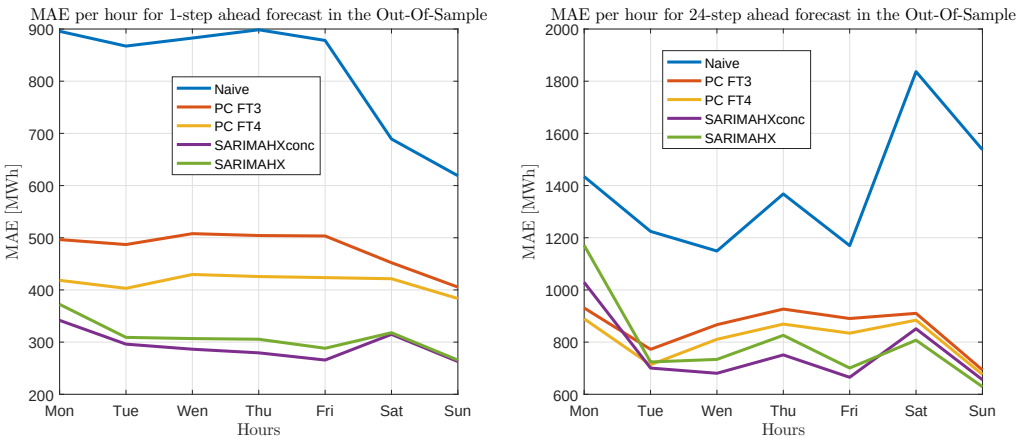


Figure 4.19. MAE for each weekday in the 1-step and 24-step ahead forecasts in the In-Sample and Out-Of-Sample periods in Offer Curves forecasting study

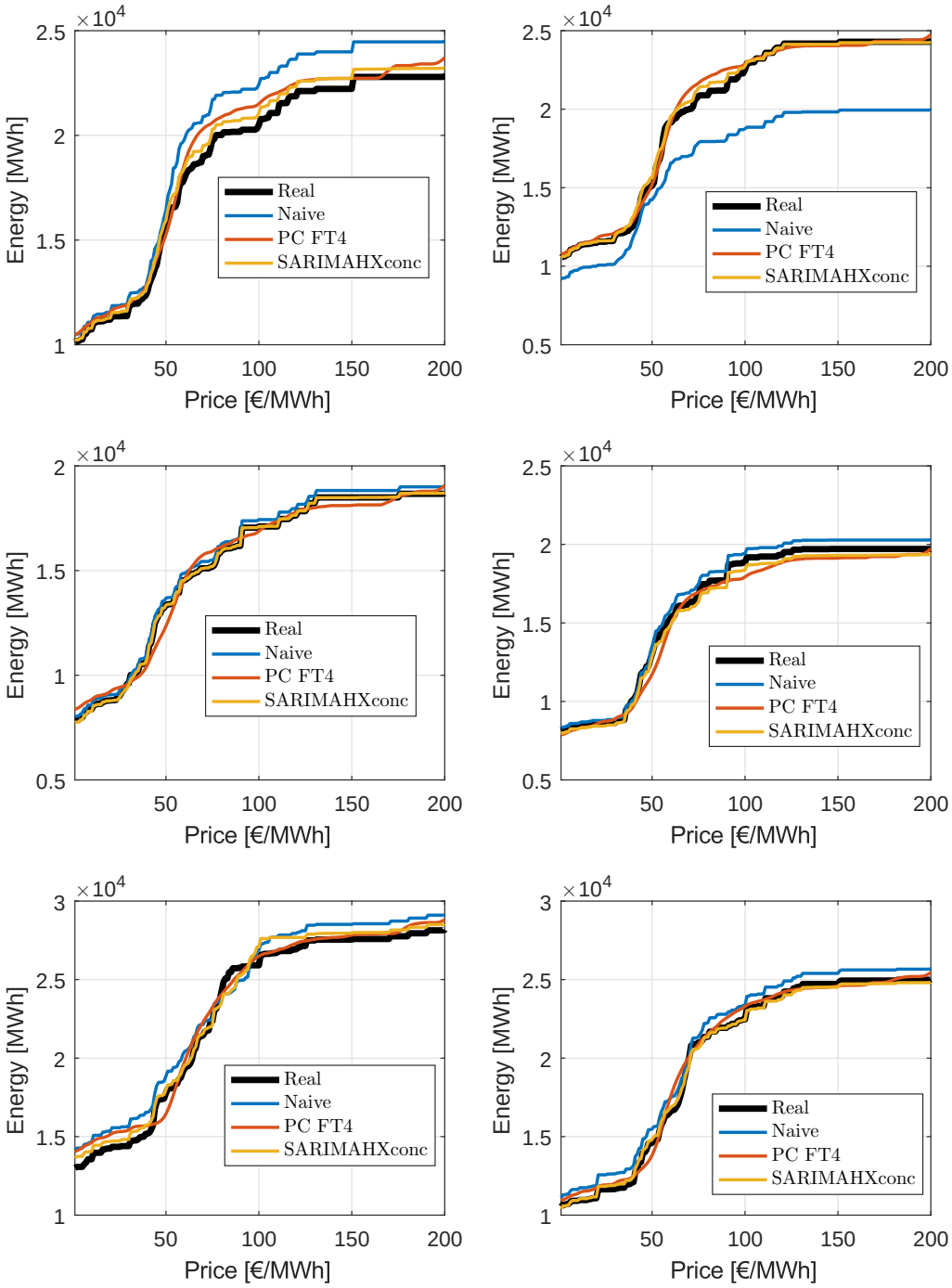


Figure 4.20. Forecast examples for 1-step ahead estimations in the Out-Of-Sample periods in Offer Curves forecasting study

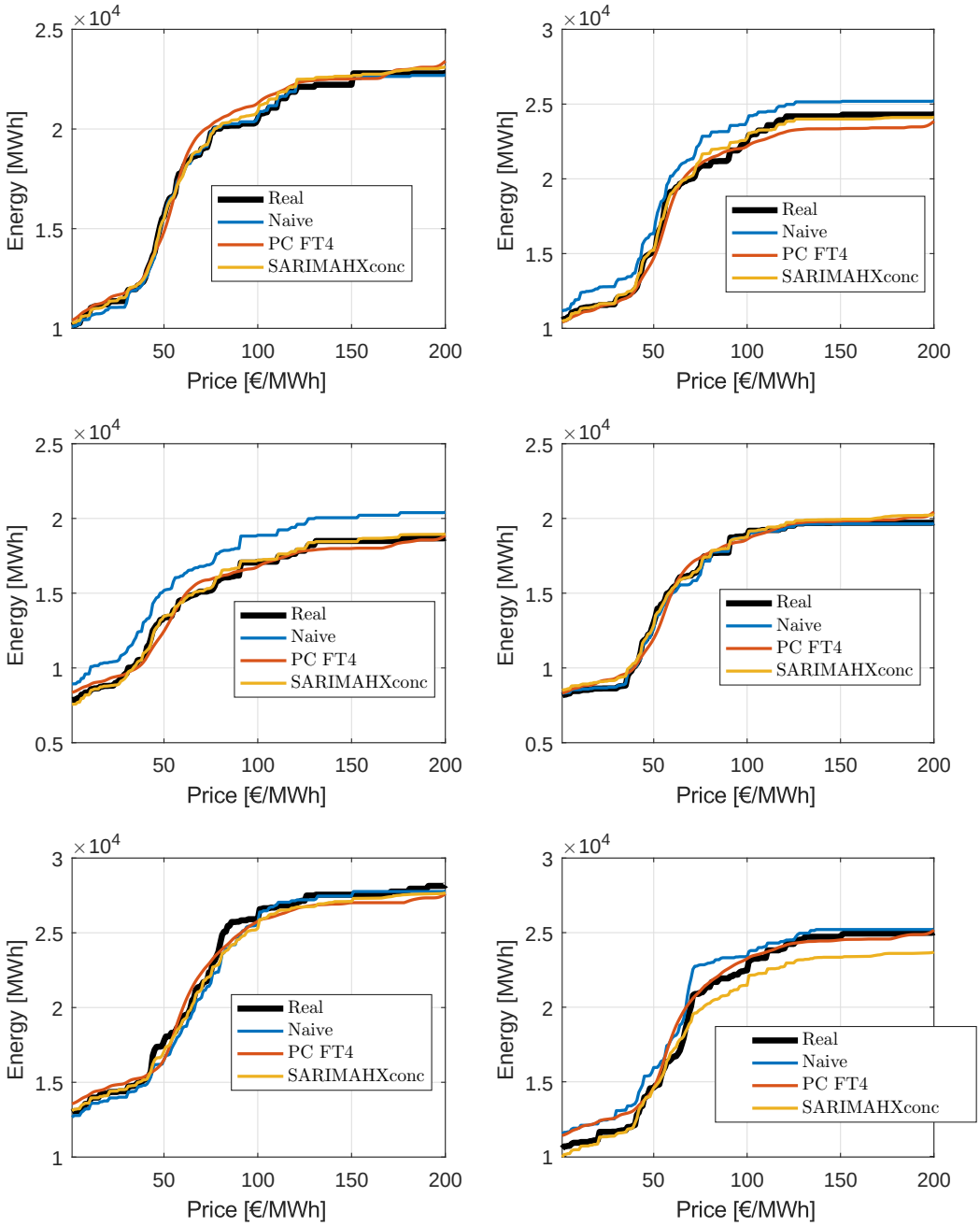


Figure 4.21. Forecast examples for 24-step ahead estimations in the Out-Of-Sample periods in Offer Curves forecasting study

4.5. Forecasting Residual Demand Curves

This section is devoted to the application of forecasting Residual Demand Curves in electricity markets. This problem is of great interest to electricity companies as it is the foundation for optimization programs that allow optimizing the bidding strategy of the agent in the different electricity markets.

In a simple-bid market, each offer (or bid) is defined by a price p and a quantity q, which refers to the amount of energy q the agent is willing to sell (or buy) at that price p. By sorting the selling (buying) offers in increasing (decreasing) prices, the aggregated supply (demand) function for the agent is built. Once all the agents have submitted their offer curves, the sum of the supply functions results in the system supply function $S_h(p)$ and the sum of the demand functions of each firm results in the system demand function $D_h(p)$. Market-clearing price p_h^* is computed for each hour as the intersection of the system aggregated supply and demand curves, hence $D_h(p_h^*) - S_h(p_h^*) = 0$.

For a given generation company i participating in this simple market, a Residual Demand Curve (RDC) can be defined for each hour as the function $R_h : \mathbb{R} \to \mathbb{R}$ that models the offering and bidding behavior of all the competitors. Given the hourly system supply function $S_h(p)$, the system demand function $D_h(p)$ and the firm's supply function $S_h^i(p)$, the supply function submitted by the firm's competitors is easily calculated as $S_h^{-i}(p) = S_h(p) - S_h^i(p)$. Hence, assuming that the firm i is a pure generation company the residual demand function $R_h(p)$ that the firm faces see, e.g., [Bal+04] can be calculated as

$$q = R_h(p) = D_h(p) - S_h^{-i}(p)$$

For a given price value p, R_h gives the maximum energy quantity q that the company can sell in the market at hour h. The inverse of the residual demand, expressed as $p = R_h^{-1}(q)$, is interpreted from the point of view of the company. If the slope of the curve R_h^{-1} is steep, it means that the company is a price maker, as small changes in the energy offer produces great changes on the clearing price. On the other hand, a flat curve means that the generation company acts as a price taker as the clearing price would remain constant and independent on the quantity sold. RDCs are used by utilities and other institutions in order to help the analysis of participants' behavior and the decision making process in the electricity market. The two main applications where RDCs are applied are described hereafter.

The first application is related to the regulator's concern about the possible exercise of market power by the electric companies. As it has been mentioned above, the slope of the RDCs provides useful information regarding the capability to influence market prices of the company for which the RDCs have been computed. Therefore, the analysis of historical data allows the regulator to compare the potential market power of the different firms. Some previous works that follow this approach are [Lee+11; XB07; Wol03; Mar+08].

The second main application of RDCs is the design of optimal offering strategies, which is very related to the subject of interest in this work. A generation company submitting an energy offer for hour h, would obtain the following profits as a function of the offering quantity q:

$$\beta_h(q) = p_h^*(q) \cdot q - c_h(q),$$

where $c_h(q)$ is the cost function, and $p_h^*(q)$ is the hourly market-clearing price which depends on the submitted offer. As it was previously defined, this dependency can be modeled with the RDC. Therefore, in a marginal pricing auction, the profit function can be redefined as:

$$\beta_h(q) = R_h^{-1}(q) \cdot q - c_h(q)$$

Hence, given the RDC for that hour, the optimum amount of energy to be traded can be calculated so as to obtain the maximum profit. This optimization problem has been widely studied as can be seen in [del+02; Ber+01; Bai+04; Cam+16]. Notice that assuming a cost function $c_h(q)$ is a very simplified approach, as the cost structure of real generator links the decisions made in different hours due to the existence of well-known technical constraints: minimum stable load, start-up and shut-down trajectories, ramps, water reservoirs for hydro generators, etc. Therefore, Mixed Integer Linear Programming models are developed that allow optimizing the offering strategy for generation units of different technologies given a set of RDCs, generation costs and the technical constrains that link the operation of unit throughout the day. This methodology can be used for short-term applications such as unit commitment [GB00] or for long term planning, as can be seen in [BB11].

In order to estimate future RDCs, two steps are needed:

- Building the RDCs: The first step is to calculate the RDC for every past hour for the company of interest. The market operator makes publicly available all the bidding information for the auctions of past days. Therefore, supply and demand functions can be constructed and RDCs calculated applying equation (4.1). This method for calculating RDC assumes that the market is a simple clearing market, where no complex conditions affect the resulting clearing price. However, complex bids and exchange capacity limitations can change drastically the outcome of the clearing process as the supply and demand functions are altered. These effects are not modeled by equation (4.1), causing the resulting RDC to be inaccurate.
- Forecasting the RDCs: once the past hourly curves have been built, a time series of curves is obtained, i.e a functional time series. Then, this is the pure forecasting problem that can be tackled with FDA. The proposed approach in this dissertation is compared against other functional approaches.

As a consequence, the following sections cover these two steps needed in the forecasting of RDC.

4.5.1. Modeling RDC in complex markets

This section explains in detail the problem of modeling RDC. This chapter is based on the published work [Por+16b], which can be consulted for more in-depth analysis. The market based on simple bids is the building block of electricity markets. However, in order to encourage competition between agents, different market designs have been proposed in the literature [PA+13]. In complex markets, offers and bids are submitted along with complex conditions so that, in case they are not fulfilled, the offer or bid cannot be cleared. Although each market has its own complex conditions, many of them involve hour-coupling constraints. In addition, electricity markets are usually interconnected by cross-border transmission lines and the market

Table 4.14. List of PXs within the PCR initiative. The countries belonging to each PX are shown as well as the number of bidding areas the PX operates.

Power Exange	Countries	Bidding areas
APX	Great Britain, Netherlands	2
Belpex	Belgium	1
EPEX	SPOT France, Germany, Austria, Switzerland	3
GME	Italy	6
Nord Pool Spot	Norway, Sweden, Finland, Denmark, Estonia, Latvia, Lithuania	15
OMIE	Spain, Portugal	2
OTE	Czech Republic, Slovakia, Hungary, Romania	4

solution must fulfill the technical constraints of these interconnections. Transmission congestion can be managed in two different ways: nodal pricing or zonal pricing [HL12]. Each of these approaches has a different clearing procedure:

- Nodal pricing: This market design uses a very detailed representation of the network. The spot price for each node of the network is obtained by solving a social welfare maximization problem subject to some constraints. All accepted offers are paid according to the associated nodal price. For each line, the difference between the prices at extreme nodes is used to calculate the congestion rents which are used to remunerate the transmission costs. This approach is used, for example, in Argentina, Chile, and several US states such as Texas and California.

- Zonal pricing: This design bundles nodes together into bigger regions called bidding areas. A bidding area is a network area in which market participants can offer energy without having to acquire transmission capacity to conclude their trades. Inter-area congestion is considered, but each region has its own uniform market-clearing price. This design is sometimes called market splitting, and it is used in many European countries. As a consequence, the nodal transmission constraints of the network are reduced to a single constraint between the bidding areas. This simplification is carried out by defining an Available Transfer Capacity (ATC) of the interconnection which establishes the maximum energy flow that can circulate between bidding areas.

We have focused on the EU markets and the recently developed Price Coupling of Regions (PCR) project, where a single algorithm [Bis+14a; Bis+14b] is used to clear simultaneously the day-ahead markets of several Power Exchanges (PXs) in Europe. Some PXs operate several bidding areas that are subject to transmission congestion. This algorithm solves an optimal pricing problem giving a solution for the prices, quantities and power flows between each bidding area. Table 4.14 shows the different PXs that are involved in the PCR project.

In the case of the Iberian market (Spain and Portugal), both countries belong to the same PX, but each one of them is considered as a single bidding area. Therefore, PCR determines the prices for each country in the Iberian PX, as well as the congestion state and power flow of the transmission line. If the interconnection reaches its physical limits exceeding the ATC, market splitting occurs resulting in different prices for Spain and Portugal. On the contrary, if the transmission constraint is not active, the market prices in both bidding areas are equal.

Inter-temporal as well as spatial constraints can cause a clearing price significantly different from what would be obtained in an isolated and pure simple-bid market. The standard definition of the RDC given by (4.1) does not consider the effect of these constraints, and, as a consequence, it can be an unreliable method for modeling complex and constrained markets like the day-ahead electricity market. Some examples that address this inaccuracy issue are [Lee+11; XB07; Xu+11b; Xu+11a], where a transmission-constrained residual demand derivative is calculated so that the effect of transmission congestion is considered in the price responsiveness. However, this approach does not take into account inter-temporal constraints such as complex offers or bids of agents and, in addition, the market configuration they are studying is nodal pricing, which is different from the zonal pricing used in Europe. On the other hand, [Váz+14] attempts to modify the RDC accounting for the most relevant complex condition in the Spanish market (known as the Minimum Income Condition). However, it does not take into account transmission constraints (Market Splitting).

A recent article which also shows the importance of modeling properly the RDC is [Reg14]. In this paper, the objective is to estimate marginal costs, startup costs and forward contracts of a firm by analyzing the offering and bidding behavior of such firm on a time interval. The estimation method includes a detailed modeling of the Spanish market complex conditions. However, its analysis is focused on obtaining marginal information, and therefore, it does not study how complex offers or bids affect at price ranges far from the clearing point. Thus, its approach could be classified as a local approach focused on obtaining the slopes of the residual demand curves around the clearing point. In addition, it develops a simulation experiment for a generation company. It defines an optimization model that finds the optimum offering strategy for the company. This optimization model makes use of the standard RDC in order to compute the revenue function. Therefore, while complex conditions are used for estimating the firm's parameters, they are not taken into account in the RDC, which can be biased in price ranges far from the clearing point.

This section of the dissertation is aimed at quantifying the effects of inter-temporal complex offers in RDCs as well as market splitting effects and defining a more reliable modeling strategy. In addition, this RDC is required to be calculated for the full range of prices (not only in the vicinity of the clearing point) so that it can be used in applications such as optimizing the offering strategy of generators.

Specifically, a model is developed for the Iberian electricity market (from the perspective of a Spanish agent), taking into account its current complex conditions. While the model proposed by [Váz+14] consists in shifting a constant quantity the RDC along the energy axis, the proposed methodology improves the modeling by considering different energy variations for each point in the curve. In addition, the coupling effect of Minimum Income Condition and Market Splitting are included in the model.

Firstly, the Iberian day-ahead electricity market is briefly explained along with its clearing process, complex conditions and transmission congestion management. Secondly, the consequences of using standard RDCs in this market are stated. Then, the main contributions are presented. The concept of RDC is redefined and a new methodology of building RDCs is explained. Finally, a case study is analyzed comparing the new approach against the traditional one.

4.5.1.1. Bidding and offering in the Iberian Day-ahead electricity market

The Spanish market and the Portuguese market operate under the regulation of the Iberian electricity market. Before 2014, the market was controlled by the Iberian Market Operator OMIE (OMIE stands for Operador del Mercado Ibérico Español). However, since 2014, each county in the Iberian market is cleared within the mentioned European PCR (Price Coupling of Regions) project. The clearing algorithm is called Euphemia. The different offers that a selling agent can submit as well as the clearing processes are explained in the following sections.

4.5.1.1.1. Simple and Complex offers

As it was mentioned before, a simple sale offer that is placed in the market is defined by a quantity q and a price p, which indicates the amount of energy q that an agent is willing to sell at that minimum price p. In addition, generation agents are allowed to submit different complex conditions to selling offers, which are now explained:

- Energy Gradient: This condition is applied to the offering unit. It forces that the difference of the energy cleared by a generating unit between two consecutive hours has to limit respect a positive and a negative value.
- Indivisible offers: This condition can be imposed to single offers. It ensures the offer to be cleared for its full energy value, and not for a fraction of it. One individual condition is allowed per hour for each offering unit.
- Minimum Income Condition (MIC): This condition affects all the offers of an offering unit u. It is defined by a fixed term (MIC^u_{fix}) in Euros and a variable term (MIC^u_{var}) in €/MWh. It forces that, unless the unit receives a minimum income in the day (β^u), it does not enter the clearing process. Mathematically, it can be expressed as:

$$\beta^u > MIC^u_{fix} + MIC^u_{var} \cdot \sum_h e^u_h,$$

  where e^u_h is the cleared energy of the unit u for hour h. This is very restrictive, because leaving out an offering unit implies that all the unit's offers in the day have to be extracted from the clearing process, producing a significant change in the supply functions.
- Shutdown scheduling: This condition can be imposed to single offers. For the first three hours of the day, in case the unit is dropped out due to MIC condition, the offer can still be cleared, so as to allow the unit to follow the shut-down trajectory.

4.5.1.1.2. Clearing Process

As it has been previously stated, the clearing market algorithm has changed recently. In order to highlight that this change does not affect the proposed method for RDC calculation, both clearing procedures are summarized, giving special attention to the handling of MIC units. All the details can be found in the aforementioned documentation.

Former Iberian Clearing Procedure (Before 2014)

The day-ahead market clearing algorithm carried out by OMIE can be summarized in three steps:

- Initial solution: Firstly, Spanish and Portuguese buying and selling agents submit bids and offers on the market for every hour of the following day. Hourly aggregated supply and demand curves are built and, for every hour, the initial clearing price is given by the point in which the two curves intersect. Then, the Energy Gradient and Indivisible Offers conditions are checked, and the clearing price is slightly modified so as to adapt any offer that did not match the condition until those conditions are verified for every unit.

- MIC: Secondly, MIC complex conditions are checked and an iterative process is followed in case there are some units that do not verify it. The iterative process finishes when all of the offering units which have cleared offers during the day receive at least the required incomes. In particular, for every unit that does not match MIC requirements, the difference between an estimated market price derived from MIC and the price that would result because of clearing the unit is obtained. The unit for which the difference is bigger is removed, and all the offers submitted by that unit are extracted from the clearing process of the whole day. Then, a new price is calculated for each hour, and MIC is again checked. The process is repeated until a first initial solution is obtained.

- Solution refinement: This step is devoted to optimize the solution by checking if there is a better combination of MIC units for the market. The algorithm search for new combination of active MIC units that minimizes the income margin, defined by

$$\sum_{u} \left(\beta^{u} - \left(MIC^{u}_{fix} + MIC^{u}_{var} \cdot \sum_{h} e^{u}_{h} \right) \right) \qquad u \in MICunits$$

  If a combination with lower income margin than the initial solution is found, it becomes the final solution. Figure 4.22 shows the effect produced by the elimination of offers which have not fulfilled complex conditions. It can be seen that the cleared supply curve is shifted to the left, showing that there has been energy withdrawn from the clearing process.

- Market Splitting: Once the final solution is obtained, the exchange capacity between Spain and Portugal is analyzed by OMIE (Iberian Market Operator). For every hour, the cleared energy that is exchanged between Spain and Portugal is calculated. If this amount is under the maximum limit (imposed by the technical capacity of the interconnection), the process is finished. However, for those hours in which this condition is not fulfilled, the markets split and they are cleared separately. Buying offers and selling bids that had been cleared are separated according to the country of origin and a new bid with the energy value of the interconnection capacity is added to both Spanish and Portuguese polls. For the Spanish poll, this bid is a buying bid when the traded energy is being exported to Portugal or a selling offer when the energy is being imported. The analogous criteria are followed for the Portuguese poll. The prices of these bids are set administratively to ensure their acceptance (i.e., 0 e/MWh for selling and 180 e/MWh for buying). New aggregated supply and demand curves for each country are obtained and thus, two different prices are calculated. Figure 4.23 shows the supply and demand curves for each country and the effect of Market Splitting.

Euphemia clearing procedure

Euphemia is in charge of giving a clearing price for each European bidding areas within this project as well as the optimal transmission flows between them. It has two main steps:

- Welfare maximization problem: The different method of this algorithm with respect to the former procedure lies in the fact that Euphemia runs an optimization process that clears all the markets at the same time, and gives an initial solution of feasible interchanges between bidding areas maximizing global welfare. This solution also considers the entire complex conditions explained as well as other conditions related to other countries. The algorithm performs branch and cut methodology so as to find the initial feasible solution. For the areas in the Iberian Market, this initial solution provides the set of accepted MIC units and the set of MIC units that have been left out and the congestion state of the transmission line.
- Price determination subproblem: This step is very similar to the solution refinement in the former algorithm. It is in charge of verifying that all MIC units that have been left out indeed cannot be accepted in the final solution. Then, the algorithm follows an iterative procedure so as to find a new combination of accepted MIC units that has a higher global welfare than the initial solution.
- OMIE processing: The Iberian PX collects the data of the Euphemia' solution and publishes the information for the agents bidding in this market. If Market Splitting has occurred, the bidding information is separated by countries, whereas if the transmission lines are not congested, the information is aggregated as if it were a single market.

It is important to highlight that the market-clearing algorithm, which includes the existence of complex offers and a maximum interconnection capacity between neighboring countries, makes it impossible to clear each hour independently. Consequently, the 24 hours of the day are coupled.

4.5.1.2. Redefining Residual Demand Curves

The standard approach for building residual demand functions, e.g. [Uge+03; Ane+13], starts by assuming a simple-bid market and applying equation (4.1). Nevertheless, in order to partially capture the effects of complex conditions, the aggregated supply and demand curves are built using the cleared offers and bids. However, there are no cleared offers for prices higher than the clearing price (see Figure 4.22) then, submitted offers with higher prices are used to extend both supply and demand curves to the full range in price. Figure 4.25 illustrates this concept. In essence, this method builds residual demand curves which consider a fixed operating point, given by the MIC conditions that have been activated and the congestion state of the transmission line. Therefore, this approach is, in fact, only valid when price changes are small, and the activation of MIC as well as the congestion state do not change.

Without loss of generality, let us analyze the case of a virtual company entering the market. As this new company has not submitted any offer to the market in the day of study, every agent that has participated in the market is considered a competitor. Then, the supply function of the company's competitors $S_h^{-i}(p)$ is, in fact, the system supply function $S_h(p)$. This case study is a useful approach because it corresponds to the case of a new agent entering in the market

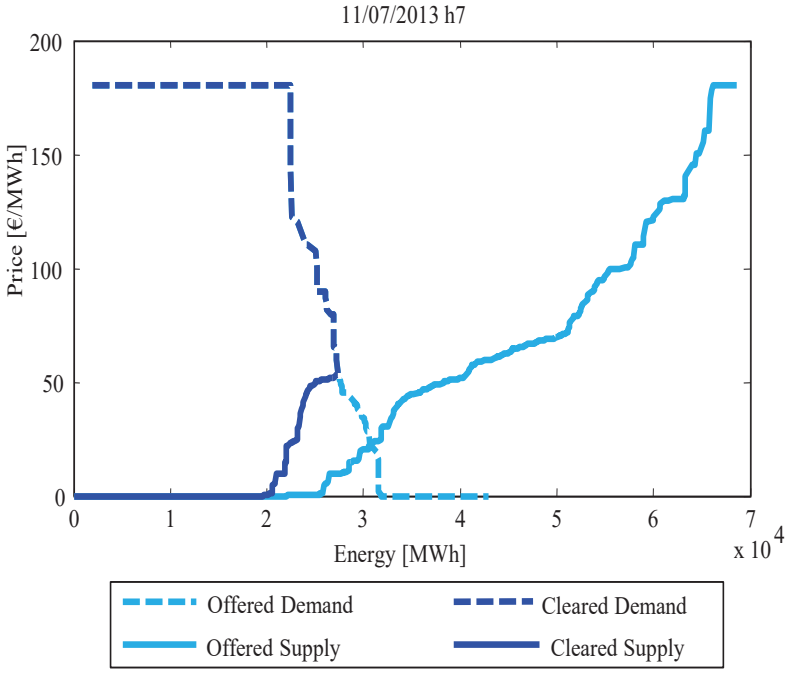


Figure 4.22. Aggregated supply and demand curves for hour 7 of day 11/07/2013. The offered supply curve is built with all the offers that were submitted to the market. The cleared supply curve is built with the final cleared offers. Because of the removed offers due to complex condition, the shapes of these curves are different. Source: www.omie.es

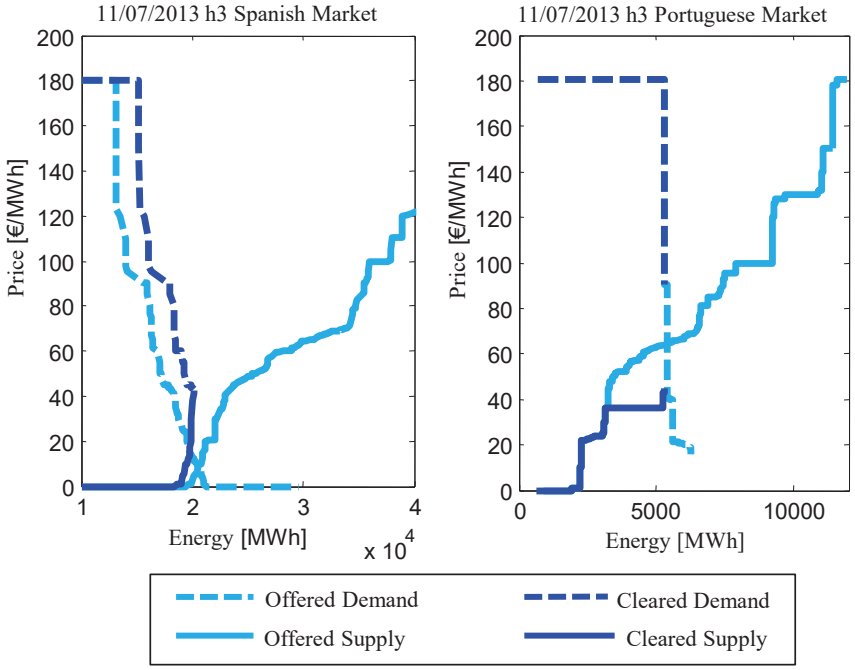


Figure 4.23. Aggregated supply and demand curves for Spain and Portugal when Market Splitting has occurred (11/07/2013, hour 3). In this case, the congestion of the transmission capacity has occurred when the energy is being exported from Spain to Portugal. Then, the Spanish offered demand is only showing the demand bids of Spanish agents while the Spanish cleared demand includes a demand bid with the energy value of the interconnection capacity to model the selling to the Portuguese market.

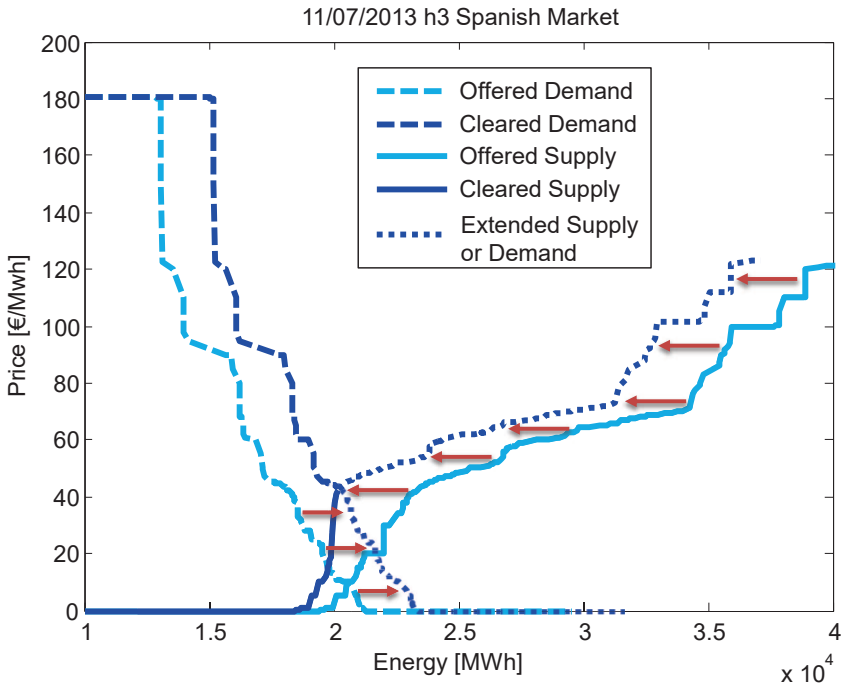


Figure 4.24. Extension of the Spanish cleared aggregated curves with submitted aggregated curves for day 11/July/2013, hour 3.

or to the case of a small generator or retailer. Then, the RDC for a new entering company is calculated for the hourly auction of Figure 4.24. Figure 4.25 shows the RDC obtained (in dark blue). The only true point in this curve is A, which is the real clearing price. The rest of the curve gives an estimation of the expected clearing price for different energy increments assuming that the activation of MIC is unchanged. Suppose that the energy is incremented in 4.2 GWh and MIC is constant. Then, the estimated market price would be Price B. However, if the change in price is significant enough so that MIC activation changes, different outcomes are possible. On the one hand, if more MIC is activated, more supply units are left out giving less cleared supply energy, and thus, the new price will be price C, which is higher than B. On the other hand, if the price change deactivates MIC, more supply units are cleared, giving a price, D. This analysis is also valid for negative values of energy (as can be seen in points B^*, C^* and D^*). While different outcomes could be obtained, there exist some limits. In cyan it is shown the residual demand curve built using only the submitted offers and bids into the market (SRRD curve, as will be described in the following subsection). Market prices can never be lower than this curve.

A similar effect can be found with market splitting constrains. Modifying the energy submitted to the market may arrive to a solution where the exchange capacity saturates, thus, producing a price change not modeled by the RDC.

As it was explained above, the 24 hours of the day are coupled by the clearing algorithm. As a consequence, obtaining hourly independent RDCs with equation 4.1 could lead to inappropriate results. In order to model the coupling effect, the residual demand should be a 24-dimensional vector function $\mathbf{R} : \mathbb{R}^{24} \rightarrow \mathbb{R}^{24}$ defined as $\mathbf{q} = \mathbf{R}(\mathbf{p})$, where $\mathbf{p} = (p_1, \ldots, p_{24})$ and $\mathbf{q} =$

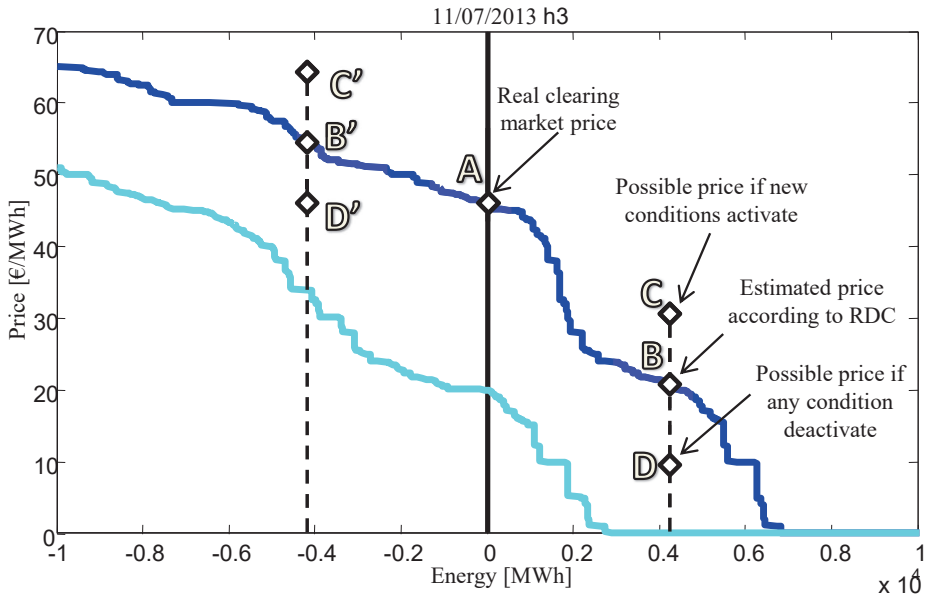


Figure 4.25. In dark blue, RDC for a new company offering in the hourly market described in Figure 3. In cyan, the RDC built using only submitted offers and bids.

$(q_1, \ldots, q_{24})$ are vectors of 24 hourly prices and energies respectively. Hence, the real residual demand is not a curve, but a multidimensional hypersurface. A generation company aiming to optimize its profits would have to optimize a problem in a 24 dimensional space of energy quantities. This method presents several aspects to deal with in practice:

- Computational issues: Evaluating the function $\mathbf{R}$ means replicating the clearing process which is a time consuming iterative algorithm. Optimizing in a 24 dimensional space implies a large number of directions to explore and a large number of evaluations of $\mathbf{R}$. This makes it an unreachable task to perform in a limited time.

- Forecasting issues: In practice, for optimizing production for future days, unknown RDCs are forecasted (as can be seen in [AP+11]). In order to forecast function $\mathbf{R}$ it would be necessary to forecast each of the competitors' offers and bids and MIC conditions, significantly increasing the complexity of the model.

- Interpretability issues: As it was previously mentioned, the slope of the residual demand curve gives information about competitors' influence of market price. With the 24 dimensional surface, it is harder to interpret. Summing up, real multidimensional residual demand function $\mathbf{R}$ would be the ideal modeling approach; however, it would be desirable in practice to have a set of equivalent 24 one-dimensional curves. The aim of this paper is to find reasonable assumptions and restrictions that can be applied to simplify the function $\mathbf{R}$ into a set of 24 equivalent curves.

A new approach for computing equivalent RDCs is now introduced. This approach is, without loss of generality, for a generation company offering in the Iberian market whose units are located in Spain. We will start by rewriting the RDC equation $R(p) = D(p) - S^{-i}(p)$. Firstly, the supply curve can be divided into offers from generation units of competitors with MIC conditions $S^{-i,MIC}$ and units without MIC conditions $S^{-i,noMIC}$, as shown in equation (4.2).

Taking one step further, demand and supply can be disaggregated into demand and supply from Spain (D_S, $S_S^{-i,noMIC}$, $S_S^{-i,MIC}$) and from Portugal (D_P, S_P^{noMIC}, S_P^{MIC}) (note that on Portugal, all the supply is competitors' supply, as the company that is analyzed has its units in Spain). This is shown in equation (4.4). Finally, the Portuguese terms can be grouped as a net exchange traded with Spain, as seen in equation (4.4). All the terms in equations (4.2), (4.3) and (4.4) are functions that depend on p. In order to simplify the terms, this dependency is omitted.

$$\begin{aligned} R &= D - S^{-i,noMIC} - S^{-i,MIC} = && (4.2)\\ &= D_S - S_S^{-i,noMIC} - S_S^{-i,MIC} + D_P - S_P^{noMIC} - S_P^{MIC} = && (4.3)\\ &= D_S - S_S^{-i,noMIC} - S_S^{-i,MIC} + N_P && (4.4) \end{aligned}$$

Taking advantage of the formulation in (4.4), we will define the following set of curves that can be built with the available historical data:

- Submitted Relaxed Residual Demand (SRRD): These curves are calculated with all the offers and bids that were submitted to the market. They result from equation (4.1) using the offered Supply and Demand curves as if all the blocks were simple offers.
- Relaxed Residual Demand (RRD): This is the standard RDC that is used in the existing literature. It is obtained from equation (4.1) using the cleared Supply and Demand curves. For energies greater than the cleared energy, no-cleared offers and bids are used to extend the curve to the full range.
- Complex Residual Demands (CRD): these are new residual demand curves which implicitly model complex conditions. They are built in two consecutive phases:
  - Phase 1 (CRDI): The first phase takes into account MIC market conditions.
  - Phase 2 (CRDII): The second phase starts from the solution obtained in phase 1 and applies market splitting conditions.

The process followed to build the proposed complex residual demand curves is described hereafter.

4.5.1.3. Complex Residual Demand Curve Phase 1: Accounting for MIC

The CRDI is the complex residual demand curve of phase I that implicitly models MIC conditions. The method is based on equation (4.2), taking into account that offers submitted by units without MIC are invariable regardless the clearing price, while offers with MIC may or may not be cleared depending on the daily income, which depends on the clearing price. Firstly, the 24-dimensional vector function $\mathbf{R} : \mathbb{R}^{24} \to \mathbb{R}^{24}$ is expressed as 24 scalar functions $r_h : \mathbb{R}^{24} \to \mathbb{R}$:

$$\mathbf{R}(\mathbf{p}) = (r_1(\mathbf{p}), \ldots, r_{24}(\mathbf{p}))$$

By fixing a price profile $p = (p_1, \ldots, p_{24})$ for the day, the market simulation simplifies considerably, as it is easy to calculate the amount of energy that would be cleared. The method presented is developed to account for MIC; however, if other complex conditions (such as ramp constraints or Indivisible Offers) wanted to be taken into account, the constraints

would have to be included at this stage of the algorithm, computing a feasible vector of cleared energies for the price profile defined. The process of obtaining r_h is described hereafter. The total cleared energy $e_h^u(x)$ for a given hour h and for a unit u, as a function of a clearing price x can be calculated as:

$$e_h^u(x) = \sum_b^{B_{u,h}} C\left((p_{u,h,b}, q_{u,h,b}), x\right),$$

where $(p_{u,h,b}, q_{u,h,b})$ is the block b of the offer submitted to the market by an offering unit u at hour h and $B_{u,h}$ is the number of blocks submitted by unit u at hour h. The function $C(\cdot,\cdot)$ gives the cleared quantity of an offer (p, q) in an auction with clearing price x:

$$C\left((p,q),x\right) = \begin{cases} q & \text{if } p \leq x \\ 0 & \text{if } p > x \end{cases}$$

Therefore, the total cleared supplied energy of the set $U^{-i,noMIC}$ of competitor units with no MIC for hourly price p_h will be:

$$S_h^{-i,noMIC}(p_h) = \sum_u e_h^u(p_h) \qquad u \in U^{-i,noMIC}$$

As for units with MIC, the cleared energy also depends on the total unit income $\beta^u = \sum_h p_h \cdot e_h^u(p_h)$. Then:

$$e^{*u}{}_h(\mathbf{p}) = \begin{cases} e_h^u(p_h) & \text{if } \beta^u > MIC_{fix}^u + MIC_{var}^u \cdot \sum_j e_j^u(p_j) \\ 0 & \text{otherwise} \end{cases}$$

$$S_h^{-i,MIC}(\mathbf{p}) = \sum_u e^{*u}{}_h(\mathbf{p}) \qquad u \in U^{-i,MIC}$$

Finally, the residual demand for hour h, given the price profile p, would be calculated according to equation (2)

$$q_h = r_h(\mathbf{p}) = D_h(p_h) - S_h^{-i,noMIC}(p_h) - S_h^{-i,MIC}(\mathbf{p})$$

The resulting vector $\mathbf{q} = (q_1, \cdots, q_{24})$ of energies provides only one point for each hourly residual demand curve. It is worth highlighting that this algorithm is free from oscillations. Oscillations might appear in a market-clearing algorithm when a removed unit not satisfying MIC, has to be added back due to the increase in the prices caused by its previous withdrawal. In our method, as our strategy is to fix the 24-dimensional hourly price profile, we do not allow for an oscillating solution. Therefore, the model can be applied in practice and convergence is always ensured. In order to build the rest of the curve, it is necessary to sample the 24-dimensional price space. Historically it can be observed that the daily price profiles are related, evidencing that the embedded dimension is smaller. The dimensionality can be reduced through Principal Component Analysis (PCA) [Jol02]. This decomposition technique finds the linear transformation that forms a subspace that maximizes the variance of the data. The first principal component extracted $\mathbf{v} = (v_1, \cdots, v_{24})$ (the eigenvector associated to the highest eigenvalue of the transformation matrix) is the direction that models the highest percentage of the variance of the data. Figure 4.26 plots the eigenvector $\mathbf{v}$ obtained when PCA is applied to the daily price

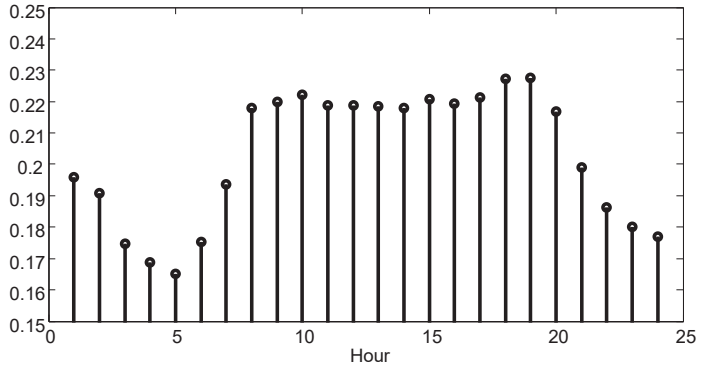


Figure 4.26. Values of the first principal component extracted for daily price of 2013. Higher values mean that prices change at a higher rate than hours with lower values.

profiles of year 2013 in the Spanish market, accounting for the 82.81 percent of the variance of the data. Consequently, we decide to reduce the price space to one dimension. It is important to highlight that the temporal link that exists in all the hours is being modeled thanks to this PCA reduction. Then, in order to create a full curve, the 24 dimensional space of price profiles is explored by searching along the embedded dimension given by $\mathbf{v}$.

The CRDI is built incrementally. Starting from an initial price profile, the search is performed by modifying the price profile along the search direction. In order to move through the search path, a weight parameter $w \in R$ is defined so that the hourly prices are modified proportional to the direction vector **v**. Therefore, the updated prices p_h^+ are calculated as $p_h^+ = p_h + w \cdot v_h$. The building algorithm is summarized thereafter:

1. Firstly the parameter w is set to 0, thus, the updated prices correspond to the initial point which is the day's real hourly prices $p^+ = p = (p_1, \cdots, p_{24})$.

2. A point in each hourly CRDI is defined evaluating the defined function r_h as:

$$CRDI_h(p_h^+) = r_h(p^+)$$

3. The weight parameter w is modified by adding constant increments $w = w + \delta w$ so as to explore new price profiles. The increment δw to be used depends on the resolution wanted for the resulting RDC. In this paper, δw has been set to 1. This means that, for example, the price of hour 18 will be increasing by 0.23 €/MWh in each iteration (see Figure 4.26).

4. New updated prices are obtained for the incremented w:

$$p_h^+ = p_h + w \cdot v_h$$

5. Then, go to step 2) to keep obtaining new CRDI points until a stopping criterion is reached. The stopping criterion is defined as the moment when any of the updated prices reach the maximum possible value, which is 180 €/MWh for the Iberian market.

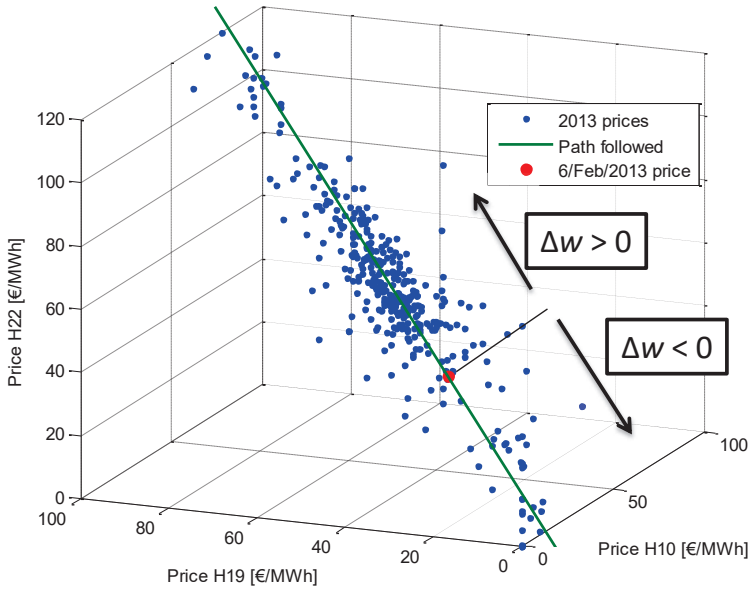


Figure 4.27. Three dimensional illustration of the price profile exploration. The red dot is the starting price profile. When the increment δw is positive, the prices are updated following the continuous path towards higher prices. When δw is negative, the path is followed towards lower prices.

6. If the stopping criterion is reached, then, repeat steps 1) to 5) but setting δw to be -1. Therefore, the path is explored by decrementing the prices. The stopping criterion would be when one of the updated hourly prices becomes 0.

In order to illustrate the proposed exploration of the price profile, Figure 4.27 shows a three dimensional equivalence. Instead of a 24 dimensional space, a three dimensional space composed of the prices of hours 10, 19 and 22 is represented. Each dot represents the prices in these 3 hours for a given day. For example, let's assume that the red dot is the starting profile, which corresponds to the 6th of February 2013. The green line is the direction given by the first principal component $\mathbf{v}$. When the weight parameter w is changed, the profile is updated following that path. As can be seen, the direction given by the first principal component, allows obtaining new price profiles which are coherent with the historically price profiles.

When the algorithm finishes, a set of 24 residual demand curves is obtained so that, for a given clearing price, returns the amount of energy the company would be able to clear considering the units that can be left out or included if the prices decrease or rise. Figure 4.28 shows a comparison between the SRRD, the RRD and the CRDI calculated for a new entering company. The difference observed between the SRRD and the RRD is due to the effect of MIC conditions. At the clearing market price, RRD and CRDI overlap in the same cleared energy. This is coherent, as the CRDI is built starting with the real price profile of that day. Then, as the price gets smaller, more units with MIC do not fulfill the condition and there are less supplied energy so the curve moves right. On the contrary, as the price is higher, more MIC units fulfill the condition and more energy is cleared, moving the residual demand to the left. When the price approaches the maximum possible, no unit is left out because of MIC, and the CRDI tends to the SRRD.

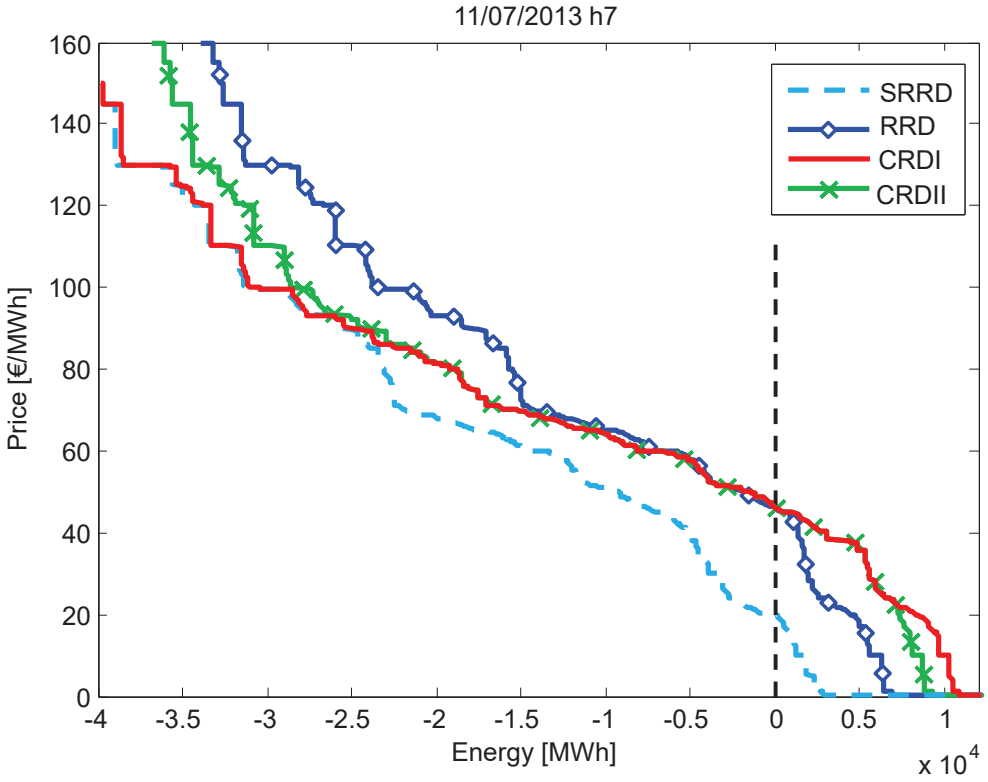


Figure 4.28. Comparison of the different RDCs that have been defined: standard approaches (SRRD & RRD) and the proposed definitions (CRDI & CRDII)

4.5.1.4. Complex Residual Demand Curve Phase 2: Accounting for market splitting

The CRDII is built over the solution obtained in the CRDI, as market splitting condition is applied to the cleared offers and bids. Recalling the formulation in equation (4.4), Portuguese supply and demand can be grouped as a net energy being traded with the Spanish market. It can be positive if Spanish market is exporting or negative if it is importing.

Because this exchange has some maximum limits, market splitting conditions can be modeled as a saturation of the Portuguese net energy. Market Splitting can be modeled by limiting the Portuguese net energy to the capacity limits. Mathematically, this is expressed as:

$$N_p = \begin{cases} L_{imp} & \text{if } N_p^{noLim} \leq L_{imp} \\ N_p^{noLim} & \text{if } L_{imp} < N_p^{noLim} < L_{exp} \\ L_{exp} & \text{if } N_p^{noLim} \geq L_{exp} \end{cases}$$

where L_{exp} and L_{imp} are the min and max limits imposed by the SO and $N_p^{noLim} = D_P - S_P^{noMIC} - S_P^{MIC}$ corresponds to the net energy. Figure 4.28 shows the differences between the CRDI and CRDII for the same hour. It can be seen that, in the range where the Net Exchange is between the two limits ($\pm$2000MWh for that hour), both curves are equal. However, when saturation activates, the curves diverge, as CRDII is only taking into account differences on the Spanish offers and bids.

4.5.2. RDC forecasting study

After the residual demand curves have been modeled in such a way that the clearing process is better represented, the following task is to forecast these RDC. The hourly sequence of RDC forms a functional time series and therefore, the proposed SARIMAHX model is applied and compared to other functional models.

Complex RDCs have been obtained for the year 2013 and are used as the data to test the models. The year 2013 is contaminated with several strategy changes from companies both in the first and last months of the year. Therefore, the dataset used is comprised of the months March to September i.e. 2016 curves.

This case study presents many similarities with the offer curves case study. The functional time series are composed of functional observations $Y_t(v)$ the hourly aggregated offer curve submitted to that zone for each price v in the range $0 - 120$ €/MWh. The data in this functional time series is observed on an hourly basis and, as a consequence, the RDC time series will exhibit two significant seasonalities: daily and weekly.

The offering and bidding behavior of the agents is conditioned to the weather and the particular conditions of each day. For that reason, explanatory variables are used as a way to account for the external factors that might influence the traders decision. These explanatory variables are the following:

- System demand: The demand is usually an important factor, as it defines the amount of purchase that is expected and can favor agents to rise prices.
- Wind production: This renewable energy has a very low generation cost. In Spain, wind production accounts for the % of the production.
- Hydraulic production: Hydraulic energy has the ability to regulate and accounts for an important share of the generation in Spain.

The explanatory variables are scalar values for each hour, therefore, they are considered as scalar covariates to the forecasting model.

It should be remarked that in most of the markets, the auctions of the 24 hours for the following day are cleared at the same time. Therefore, two different analysis are analyzed. On the one hand, a one-step case study is analyzed. The proposed model is trained to minimize the one-step ahead forecast error, then, it is reasonable to study this theoretical comparison. On the other hand, a 24-step ahead forecast is also evaluated. This serves two main purposes. It shows the adaptation of the model to a variable horizon forecast as well as validating the use in a real case application.

The models to be compared are described and analyzed.

- Naïve. This is a simple benchmark model that is used to obtain a simple reference forecast and provides a reference to compare the rest of the models. Two versions are used, depending whether the simulation is one-step or 24-step forecast. In the first case, the forecast is simply the last curve observed in the data, i.e. the curve from the previous hour. The second case, the forecast takes into account that the curves of the same day

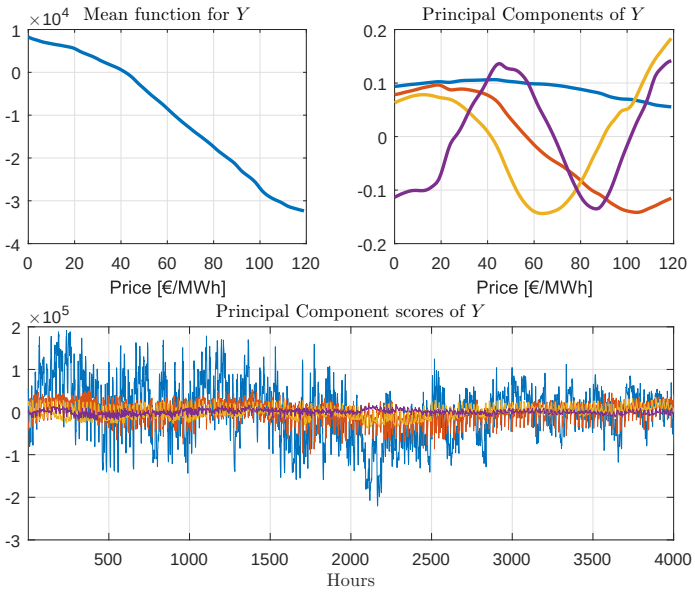


Figure 4.29. Mean function, Principal Components and scores of the offer curves

are not observed. Then, the forecast will depend on the type of weekday. For Saturdays, Sundays and Mondays, the forecast will be the hourly curve of previous week, while for Tuesdays, Wednesdays, Thursdays and Fridays, the forecast will be the hourly curve of the previous day.

- Principal component dimension reduction approach. This method extracts the Functional Principal Components of the curves and the corresponding scores time series. Then, the scores are forecasted by means of Transfer Function (TF) models, which include explanatory variables. The final estimation of the curves is done by reconstructing the estimated scores. The In-Sample period is used for all the training. FPCs are extracted for that range and the parameters of the TF are estimated. In the Out-Of-Sample period, the component scores are obtained, by projecting the curves into the basis spanned by the FPCs previously extracted.

  For illustration purposes, in the one-step ahead case the ideal prediction results are shown. Instead of resconstruncting the forecasted scores, the ideal forecast is the reconstruction of the real scores. In [Bos00] and [Kle+16], they make reference to this model as the best prediction model. Indeed, once the functional time series is projected on to a functional basis, the best possible forecast is to reconstruct the curves according to the real scores. This model could never be applied in real life as the real scores of future hours or days are unknown, but it is a meaningful comparison that illustrates the amount of information that is lost in the dimension reduction procedure.

  Figure 4.29 represent the mean function, the principal components and the scores obtained for the curves time series. Three and four principal components are extracted, which explain the 98% and 99% of the variance of the data respectively. Analyzing the FPCs it can be observed how the first FPC represents the level of the curve, as it is mostly flat. This is the component that best represent the data, and that is why the first scores time

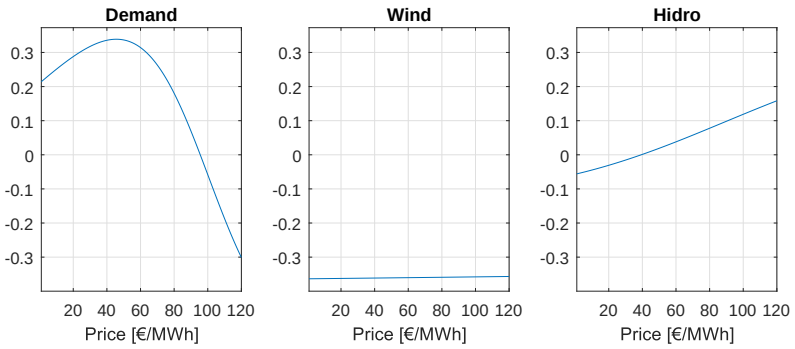


Figure 4.30. Operators' kernels for the regressors of the SARIMAHX model of the offer curves

series is significantly greater that the rest. The interpretation of the rest of the scores is less intuitive. In fact, in this case they look like sinusoidal functions.

- Functional SARIMAHX. The proposed functional and concurrent time series models are trained with the In-Sample data of this case study. In both cases, the explanatory variables are scalar hourly values thus, the univariate sigmoid kernel is used. The resulting shapes of those functional parameters are shown in figure 4.30. Analyzing these shapes, the effect of each variable on the residual demand curve can be seen. As was expected, a higher demand rises the RDC as there is more demand that can be covered in the market. The negative values in wind coefficients imply that the RDC decreases uniformly. Wind production offered at prices close to 0, imply that there are less demand to be covered.

  Once the regressive terms are adjusted, the dynamic terms for the concurrent and functional versions are adjusted. Several models were trained and the ones with less training error are described ahead:

  – Concurrent SARIMAHX: This model is the concurrent version proposed in section 3. The final adjusted model was a concurrent SARIMAHX$(2,0,0) \times (1,0,1) \times (1,0,1)$. Figure 4.32 shows the kernels trained for the functional models. In this case. It can be seen how the first regular and the first seasonal autoregressive terms take high values with respect to the other functional parameters, meaning that they are the most relevant coefficients in the model. It is worth noting how the functional coefficients vary for each price, meaning that the different parts of the curves present different temporal behaviors.

  – Functional SARIMAHX: This is the fully functional forecast model described in section 3. Again, the final adjusted model was a SARIMAHX$(2,0,0) \times (1,0,1) \times (1,0,1)$. Figure 4.31 shows the kernels trained for the functional model. Analyzing the shape of the bivariate kernels, the first regular and first seasonal functional coefficients are the most significant as they show high values on the diagonal. This means that not much relation is found between different parts of the curve and the behavior of the offer curve at some price is mostly dependent on previous curves at that same price. If the shapes of all functional coefficients had all high values in the diagonal, the model would be equivalent to the concurrent functional model.

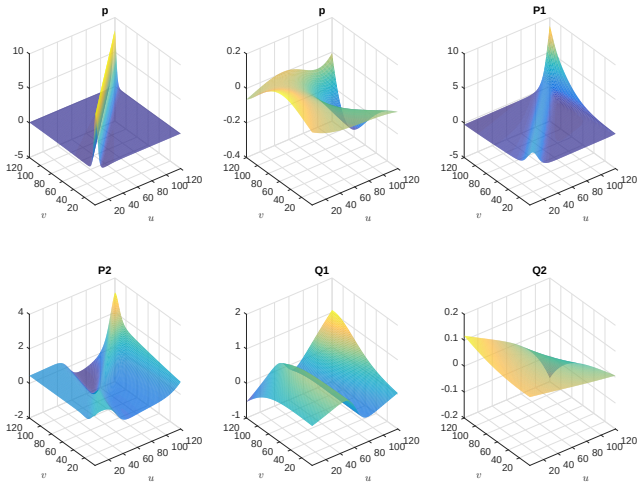


Figure 4.31. Operators' kernels for the SARIMAHX model of the offer curves

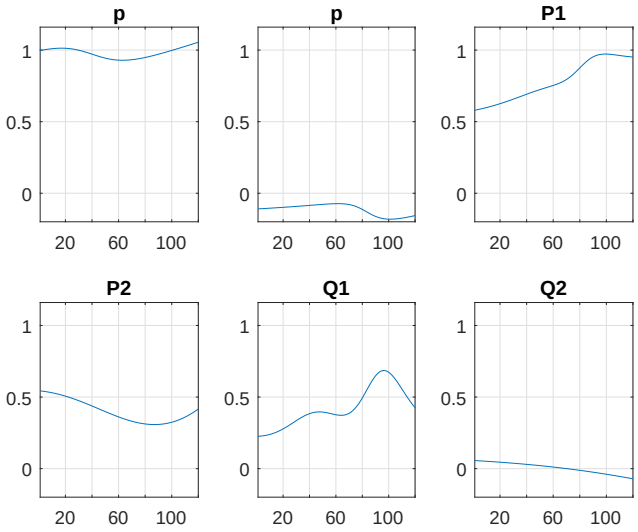


Figure 4.32. Operators' kernels for the concurrent SARIMAHX model

All the models are trained with the In-Sample period. Then, each produces a 1-step ahead forecast, assuming that the curve of last hour is known, and a 24-step ahead forecast, where the 24 hours of the following day are forecasted being the last observed curve the hour 24 of the current day. Functional errors FMAE, FRMSE and FMAPE from section 2 are calculated. Table 4.15 and table 4.17 show the functional errors for the 1-step and the 24-step ahead forecasts. Each table shows In-Sample as well as Out-Of-Sample errors. In addition, tables 4.18 and 4.19 show the Diebold-Mariano tests applied to each pair of forecasts. In addition, Table 4.16 show the results with the ideal Principal Component models.

Now, the estimation results are analyzed. Firstly, by looking at the 1-step ahead forecasts, it can be seen how the functional approach outperforms the other methods in both the In-Sample and Out-Of-Sample periods. Moreover, the fully functional SARIMAHX model is the one that

Table 4.15. Average errors for each method in the one-step ahead prediction in RDC forecasting study

Model	In-Sample			Out-Of-Sample		
	FMAE [MWh]	FRMSE [MWh]	FMAPE [%]	FMAE [MWh]	FRMSE [MWh]	FMAPE [%]
Naïve	1288.5	1982	9.504	1329.1	1929.1	9.928
ARIMA 3 PC	894.9	1175.7	6.392	871.1	1134	6.358
ARIMA 4 PC	776.1	1044.1	5.555	793.4	1045.1	5.801
Con SARIMAHX	694.4	1013.8	4.991	**659.7**	939.67	**4.846**
Fun SARIMAHX	**689.7**	**980.05**	**4.965**	663.8	**928.8**	4.877

provided the best results, although for some measurements, the concurrent versions was better. However, Diebol-Mariano statistical test for forecasting accuracy showed a high p-value between the SARIMAHX forecasts, meaning that models do not perform significantly different.

In addition, it is worth to compare the results against the ideal principal component prediction. As can be seen, on the In-Sample period, the ideal method significantly outperforms the functional concurrent model, as it is expected. However, for the Out-Of-Sample period, the errors are quite similar. In fact, the functional model outperforms the 3PC ideal model. This implies that the error incurred only by reducing the dimensionality of the series reaches levels equivalent to the forecasting. The reason is that the Principal Components have been extracted with the In-Sample period. The In-Sample reconstruction error is lower, however, on the Out-Of-Sample, the Principal Components are kept untouched, thus they cannot adapt to a significant change in the functional time series and the perfect reconstruction looses precision. On the contrary, the proposed functional model does not rely on any basis expansion of the series and it takes into account the whole curve values from the recent past. Therefore, it can better adapt to changes in the series.

Secondly, the 24-step forecasts show some differences in comparison to the results to the 1-step ahead case. The functional SARIMAHX model performs lower on average estimations with respect to the other model. Globally, 24-step ahead errors are much higher than 1-step ahead errors e.g. around 10% error for the 24-step and around 5% error for the 1-step. Similar to the offer curves, the RDC respond to the different strategies of the agents which, in many cases, have a common behavior for the hole day e.g. There is some restriction on a power plant for that day. Therefore, strategies are planed for the whole day, and consequently, once the first hour is observed, it provides a lot of information useful for the forecasing for the following hours . This can be seen in figure 4.34. One-step ahead errors on the firsts hours of the day are significantly reduced after observing hour 1.

To complement the results and provide more understanding on the different forecasts, figure 4.33 presents MAE errors for each price for the In-Sample and Out-Of-Sample periods for the 1-step and 24-step ahead forecasts. Looking at the figures, it can be seen how each method estimates in the different range of prices. The SARIMAHX model provides better estimations for lower prices in training.It is worth stopping at figure 4.35, which shows MAE results for each day of the week.To end up, Figures 4.36 and 4.37 show some forecasting examples that show how are the curves affected by each model.

Table 4.16. Average errors for the ideal principal component forecast in RDC forecasting study

Model	In-Sample			Out-Of-Sample		
	FMAE [MWh]	FRMSE [MWh]	FMAPE [%]	FMAE [MWh]	FRMSE [MWh]	FMAPE [%]
ARIMA 3 PC ideal	675.3	868.1	4.788	670.8	850.0	4.861
ARIMA 4 PC ideal	505.3	659.7	3.585	567.7	715.9	4.113

Table 4.17. Average errors for each method in the 24-step ahead prediction in RDC forecasting study

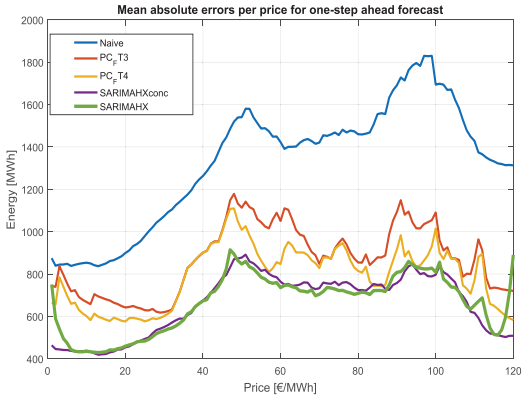
Model	In-Sample			Out-Of-Sample		
	FMAE [MWh]	FRMSE [MWh]	FMAPE [%]	FMAE [MWh]	FRMSE [MWh]	FMAPE [%]
Naïve	2700	3571.7	19.158	2502.3	3189.4	17.946
ARIMA 3 PC	1541.6	1960.5	10.864	1509.8	1940.6	10.945
ARIMA 4 PC	**1502.8**	**1910.4**	**10.588**	**1486.3**	**1910.2**	**10.777**
Con SARIMAHX	1543.5	2021.3	10.844	1594.3	2016.0	11.521
Fun SARIMAHX	1507.0	1950.1	10.618	1572.2	2008.1	11.388

Table 4.18. Diebold-Mariano test's pvalues for Out-Of-Sample period in the one-step ahead forecasts in RDC forecasting study

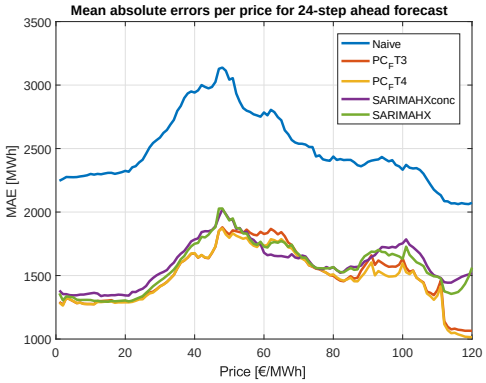
Models	Naïve	ARIMA 3 PC	ARIMA 4 PC	Con SARIMAHX	Fun SARIMAHX
Naïve	-				
ARIMA 3 PC	0	-			
ARIMA 4 PC	0	0	-		
Con SARIMAHX	0	0	0	-	
Fun SARIMAHX	**0**	**0**	**0**	**0.519**	-

Table 4.19. Diebold-Mariano test's pvalues for Out-Of-Sample period in the 24-step ahead forecasts in RDC forecasting study

Models	Naïve	ARIMA 3 PC	ARIMA 4 PC	Con SARIMAHX	Fun SARIMAHX
Naïve	-				
ARIMA 3 PC	0	-			
ARIMA 4 PC	**0**	**0**	-		
Con SARIMAHX	0	0.229	**0.158**	-	
Fun SARIMAHX	0	0.320	**0.228**	0.760	-

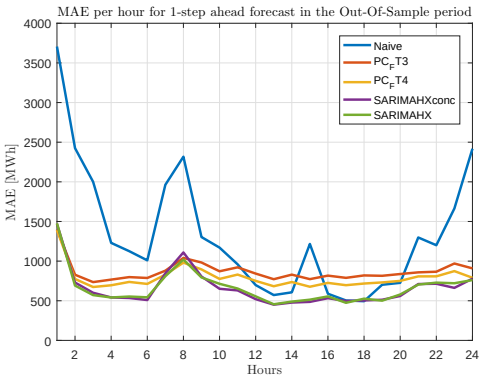


(a) Out-Sample MAE per price for one-step forecast

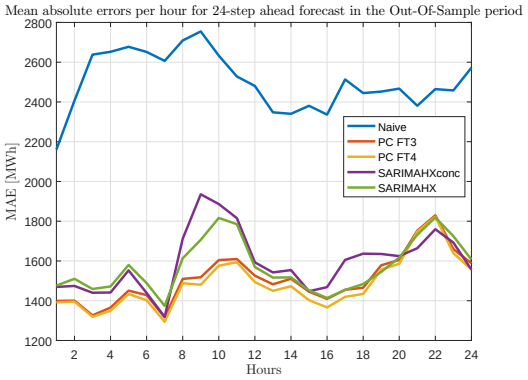


(b) Out-Sample MAE per price for 24-step forecast

Figure 4.33. Out-Sample MAE for each price in the 1-step and 24-step ahead forecast results in RDC forecasting study



(a) Out-Sample MAE per hour for one-step forecast



(b) Out-Sample MAE per hour for one-step forecast

Figure 4.34. Mean absolute errors for each hour in RDC forecasting study

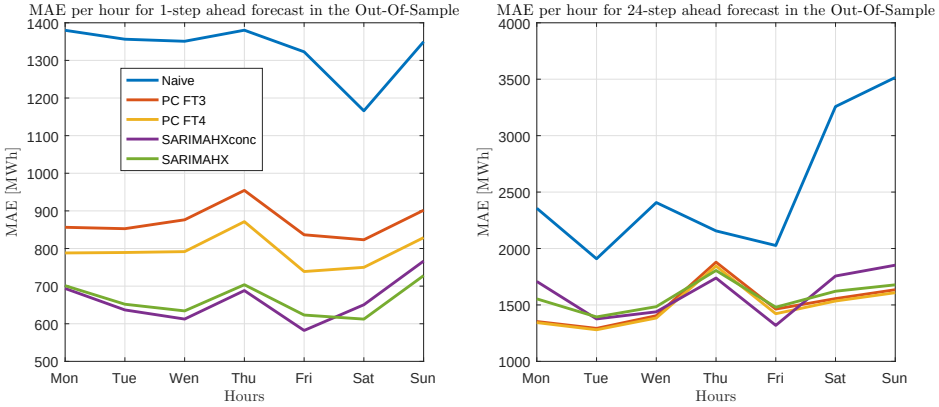


Figure 4.35. Mean absolute errors for each weekday in the 1-step and 24-step ahead forecasts in the In-Sample and Out-Of-Sample periods in RDC forecasting study

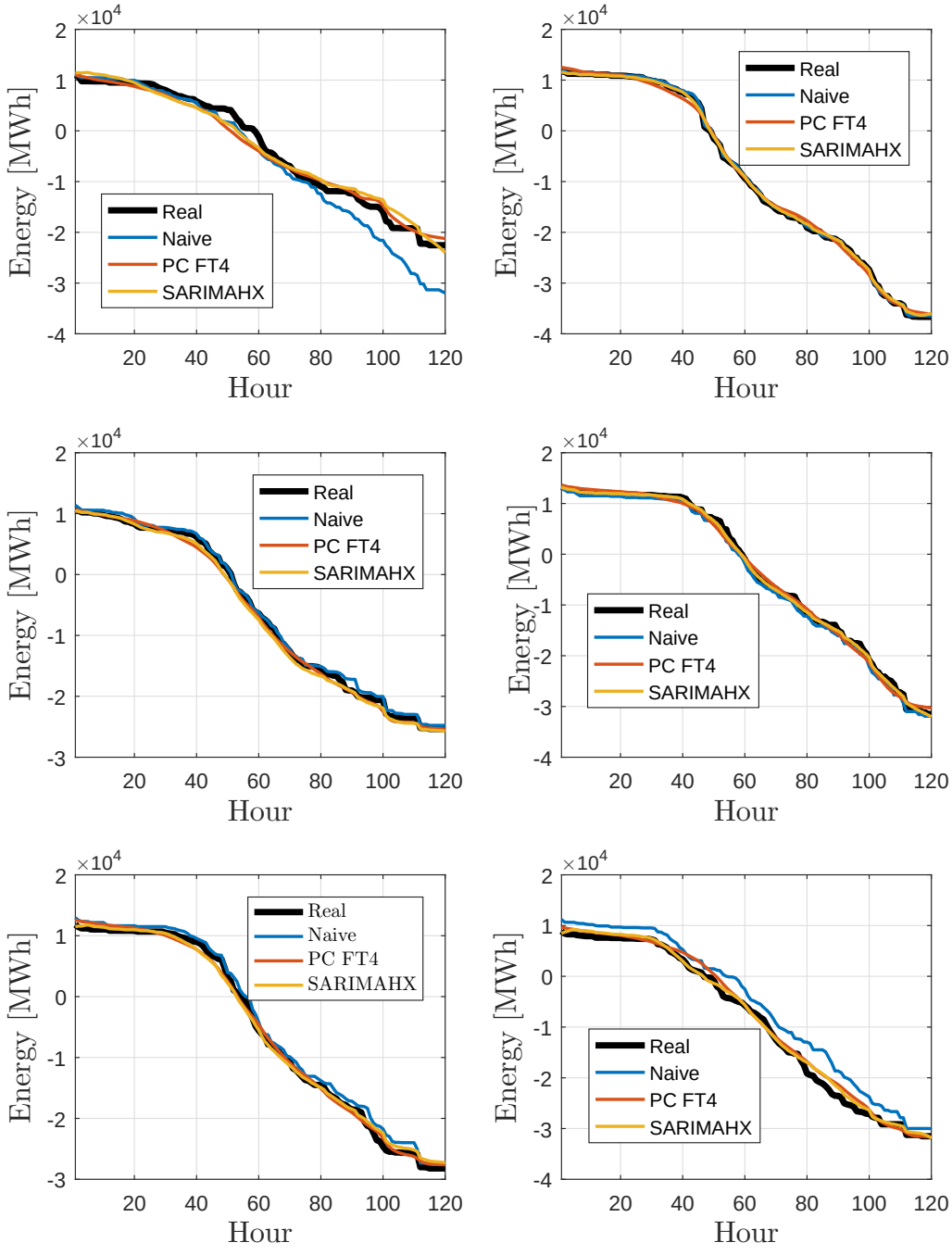


Figure 4.36. Forecast examples for 1-step ahead estimations in the Out-Of-Sample periods in RDC forecasting study

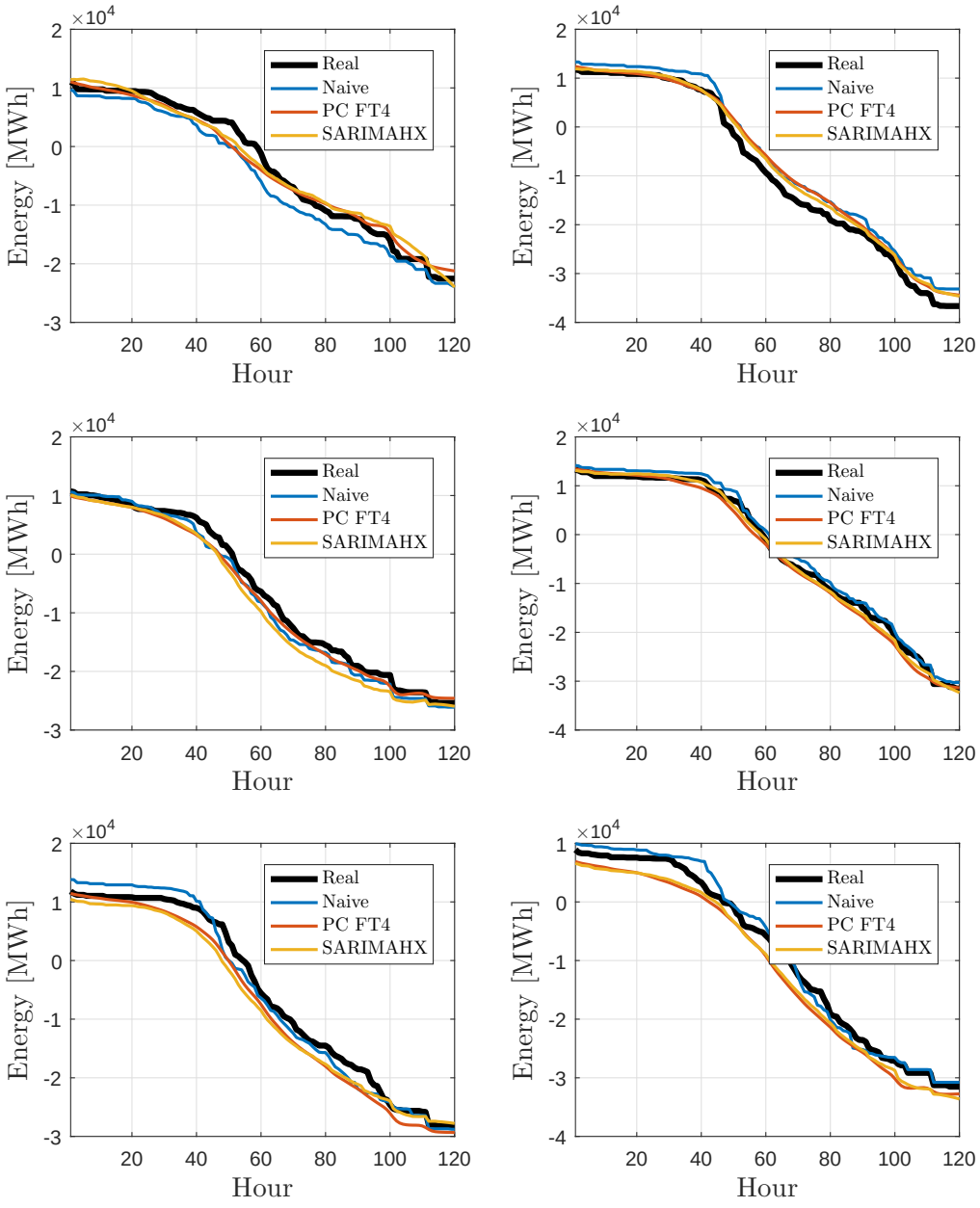


Figure 4.37. Forecast examples for 24-step ahead estimations in the Out-Of-Sample periods in RDC forecasting study

4.6. Conclusion

This chapter has been devoted to the empirical application of functional forecasting models to the electricity market. The different markets available for trading energy provide a perfect framework for the use of this new statistical approach.

The proposed functional model in this dissertation is proven to be of much use for these applications, outperforming the other reference model in most of the cases. The one-step horizon forecast is the one that produced better results, while the 24-step ahead prediction handed closer results to the reference model. This can be due to numerous factors. For example, the model is trained so as to minimize the one-step ahead error and that optimization might not be the best for multiple step ahead forecast. Hereafter, the different case studies are summarized.

Firstly, an application to the ancillary services market was analyzed. The case study was setup as a regression problem, which estimated the energy assigned to the Spanish tertiary and deviation management markets. The proposed functional regression proved to outperform the reference model thanks to the novel estimation algorithm proposed.

Secondly, the Spanish and German day-ahead hourly electricity market prices were studied. The daily price profile was forecasted as a functional time series. One-step ahead forecasts were produced with each functional model, where explanatory variables and weekly seasonality were included. The proposed model showed a better performance when accounting for moving average terms in the models. Moreover, an additional case study was provided which showed that the proposed model can compute competitive forecasts against other well known existing non-functional methods for price forecasting.

Thirdly, the forecasting of offer curves in the Italian electricity market was addressed. Offer curves are hourly FTS and thus, the functional models compared included two seasonalities (daily and weekly). In addition, scalar explanatory variables were used. One-step as well as 24-step ahead forecasts are compared, being the concurrent SARIMAHX model the winning method, as it provided better estimations for both horizons.

Finally, the case of forecasting Residual Demand Curves in the Spanish electricity market was addressed. RDCs are functions that give the market-clearing prices as a function of the energy an agent is willing to buy or sell. However, the standard definition of RDCs do not account for the effects of complex bids in the market-clearing process. Therefore, the first step was to redefine RDCs and to develop an algorithm to build RDCs which takes into account complex offering conditions in the market as well as international exchanges. On the other hand, the newly defined curves were forecasted in a similar setting as the offer curves, comparing the results of the SARIMAHX model against other reference models. The proposed model outperformed the dimensionality reduction model in one-step ahead forecasts and yielded similar errors in the 24-step ahead forecast.

5

Conclusions, Contributions and Future Work

This last chapter summarizes the developments of this dissertation. The main conclusions that can be drawn from the experiments carried out are set forth and the most original contributions are highlighted. Finally, the open issues that have not been tackled in the thesis, as well as possible future research lines, are discussed.

5.1. Summary and conclusions

Functional time series (FTS) can be found in numerous applications in electricity markets. For example, the electricity demand which is usually considered as a discrete process is in fact a continuous process as the power consumption is a continuous function of time. By dividing the univariate series into segments of equal length, i.e. daily or weekly segments, a FTS is obtained. Electricity prices can also be analyzed as a FTS considering daily and weekly price profiles as functional data. Moreover, FTS can also be obtained from the bidding information of market auctions. Aggregated offer curves and Residual Demand Curves can be calculated as functions that model the competitive behavior of agents in the market. Hence, time sequences of hourly curves are observed. Being able to forecast these functional time series is of utmost importance for market agents, which aim at optimizing their business, market operation and and bidding strategy in the market. Hence, the electricity market is an important field of application of Functional Data Analysis (FDA).

Functional time series that appear in electricity markets share some common features. They are sensitive to weather conditions and to the effects of business and everyday activities, which lead to weekly and daily seasonalities. Therefore, an appropriate forecasting model for these series has to account for all of these effects. FDA is a relatively recent research field and three main functional approaches are found in the literature: parametric, dimensionality reduction

and nonparametric models. Functional parametric models make use of linear functional operators to operate with the curves. These are powerful tools as the operators allow to produce forecasts using the entirety of the input curve. However, estimating these operators is challenging and causes that most of the available models have a simple structure, i.e. seasonality is not taken into account. The dimensionality reduction approach represents the FTS in a lower dimensional space to further apply classical multivariate statistics to the reduced components. The advantage is that powerful multivariate models can then be applied. However, reducing the dimensionality might incur in loss of precision. Finally, nonparametric models use nonlinear functional operators, which are usually focused on averaging historical curves from similar situations as the moment that it is being forecasted.

This dissertation fills an existing gap in forecasting functional time series, which is the development of functional linear forecasting models that do not require representing the functional time series by means of a functional basis. This manuscript contributes significantly by proposing a novel estimation method for the kernel of a Hilbert integral operator in L^2. Inspired on neural networks, it models the bivariate kernel as a sum of bivariate sigmoid functions, whose parameters are optimized by minimizing the forecasting error. Hence, this approach does not need to choose a fixed functional basis for estimating the kernel as the sigmoid functions are optimized in the training process. Furthermore, this estimation method allows to develop complex functional models that can tackle the known effects in time series from electricity markets. Three methods were developed. Firstly, a functional linear regression approach is presented that combines functional as well as scalar covariates. Secondly, a functional ARMAHX method with double multiplicative seasonality was developed which also accounts for explanatory variables. In addition, a concurrent version of the SARIMAHX was developed.

The performance of the proposed model has been validated with several case studies, where the SARIMAHX model was used in different electricity market forecasting applications. Firstly, an application to the ancillary services market was analyzed. The case study was setup as a regression problem, which estimated the energy assigned to the Spanish tertiary and deviation management markets. The proposed functional regression proved to outperform the reference model thanks to the novel estimation algorithm proposed. Secondly, the Spanish and German day-ahead hourly electricity market prices were studied. The daily price profile was forecasted as a functional time series. One-step ahead forecasts were produced with each functional model, where explanatory variables and weekly seasonality were included. The proposed model showed a better performance when accounting for moving average terms in the models. Moreover, an additional case study was provided which showed that the proposed model can compute competitive forecasts against other well known existing non-functional methods for price forecasting. Thirdly, the forecasting of offer curves in the Italian electricity market was addressed. Offer curves are hourly FTS and thus, the functional models compared included two seasonalities (daily and weekly). In addition, scalar explanatory variables were used. One-step as well as 24-step ahead forecasts are compared, being the concurrent SARIMAHX model the winning method, as it provided better estimations for both horizons.

Finally, the case of forecasting Residual Demand Curves in the Spanish electricity market was addressed. RDCs are functions that give the market-clearing prices as a function of the energy an

agent is willing to buy or sell. However, the standard definition of RDCs do not account for the effects of complex bids in the market-clearing process. Therefore, the first step was to redefine RDCs and to develop an algorithm to build RDCs which takes into account complex offering conditions in the market as well as international exchanges. On the other hand, the newly defined curves were forecasted in a similar setting as the offer curves, comparing the results of the SARIMAHX model against other reference models. The proposed model outperformed the dimensionality reduction model in one-step ahead forecasts and yielded similar errors in the 24-step ahead forecast.

5.2. Original contributions

The development of this thesis has yielded three main original contributions to the current state-of-the-art in terms of functional forecasting and its application to the electricity markets. Hereafter these contributions are detailed

Firstly, a novel method has been proposed to estimate an integral operator in the L^2 Hilbert space. Based on Machine Learning techniques, the bivariate kernel of the operator is modeled as a finite sum of bivariate sigmoid functions whose parameters are optimized by minimizing some custom cost function. This is a meaningful contribution in functional data analysis as it provides a greater approximation capability. Other approaches in the literature estimate the operator by expanding the kernel into a fixed functional basis and then optimizing the scores. On the contrary, the proposed method adapts the sigmoid functions to minimize the error. The benefit of the proposed approach with respect to the reference method (where Functional Principal Components are used to estimate the kernel) is showed in the ancillary services case study. The proposed method exhibited greater capabilities in the estimation of the regression kernels and yielded lower errors in both In-Sample and Out-Of-Sample. This operator estimation method provides the flexibility needed for developing more complex functional models.

The second main contribution is the extension of the Seasonal ARIMAX method to the functional framework using integral operators. The Seasonal Autoregressive and Moving Average Hilbertian model that includes explanatory variables (SARIMAHX) is proposed and estimated by means of the powerful estimation method previously described. This is a groundbreaking methodology for modeling complex functional time series, as it operates directly with functions and avoids the need of expanding to a functional basis. It opens a new framework for developing functional time series and extending the Box and Jenkins methodology [BJ70] to the functional setting. In addition, a concurrent version of the functional SARIMAHX model is also proposed where the integral operators are substituted with their concurrent alternative. The advantage is that the estimation procedure is less computationally intensive.

Finally, the third main contribution is the application of this forecasting method to different applications in electricity markets, showing the wide applicability of FDA in this field. Moreover, the case study of forecasting Residual Demand Curves (RDC) yielded a significant contribution. An algorithm that builds historical RDCs considering the effect of complex offering conditions and transmission constraints in the market clearing process was developed. This methodology allows the RDC to better estimate the outcome of the clearing market algorithm for different bidding quantities.

The research developed in the field of Functional Data Analysis and the study of Residual Demand Curves has been materialized in the following journal publications and conference papers.

Journal publications

- J. Portela, A. Muñoz, E. Alonso, Forecasting functional time series with a new Hilbertian ARMAX model: application to electricity price forecasting, (Under second review in IEEE Transactions on Power Systems, March 2017)

- J. Portela, A. Muñoz, E.F. Sánchez-Úbeda, J. García-González, R. González Hombrados. Residual demand curves for modeling the effect of complex offering conditions on day-ahead electricity markets. IEEE Transactions on Power Systems. vol. 32, no. 1, pp. 50-61, January 2017.

- F.A. Campos, A. Muñoz, E.F. Sánchez-Úbeda, J. Portela. Strategic bidding in secondary reserve markets. IEEE Transactions on Power Systems. vol. 31, no. 4, pp. 2847-2856, July 2016.

Conference papers

- J. Portela, A. Muñoz, E. Alonso, Forecasting residual demand curves of the day-ahead electricity market with a Hilbertian ARMAX model, 36th International Symposium on Forecasting - ISF 2016. Santander, Spain, 19-22 June 2016

- J. Portela, A. Muñoz, E. Alonso, Hilbertian ARMA model for forecasting functional time series, 8th International Conference of the ERCIM WG on Computational and Methodological Statistics - CMStatistics 2015. London, United Kingdom, 12-14 December 2015

- J. Portela, A. Muñoz, E. Alonso, Forecasting residual demand time series in electricity markets: a functional approach, 21st International Conference on Computational Statistics - COMPSTAT 2014. Geneva, Switzerland, 19-22 August 2014

- J. Portela, A. Muñoz, E. Alonso, Combination of forecasts in functional data: a study applied to day-ahead residual demand curve forecasting In electricity markets, 2nd Workshop on Industry and Practices for Forecasting - WIPFOR 13. Paris, France, 05-07 June 2013

- F.A. Campos, A. Muñoz, E.F. Sánchez-Úbeda, J. Portela, R. González Hombrados, J. Rodríguez Marcos, A. González Castrillón, Optimization of the bidding curve in reserve markets, IEEE 10th International conference on the european energy market - EEM2013. Stockholm, Sweden, 27-31 May 2013

- J. Portela, A. Muñoz, E. Alonso, Day-Ahead residual demand curve forecasting in electricity markets, 32nd Annual International Symposium on Forecasting. Boston, United States of America, 24-27 June 2012

Invited conference lectures

- J. Portela, A. Muñoz, E. Alonso, Forecasting functional time series in electricity markets with a seasonal ARIMAX Hilbertian model, Satellite CRoNoS Workshop on Functional Data Analysis - 2016 CRoNoS FDA. Oviedo, Spain, 26-28 August 2016

5.3. Future work

This dissertation has advanced in the development of parametric models for forecasting functional time series and its applications in electricity markets. These contributions lead to a number of future lines of research that could be explored. This section summarizes some of them:

- Development of the multiple step-ahead forecast optimization: The current optimization algorithm minimizes the 1-step ahead forecast and therefore, might not be optimal for cases where multiple step ahead forecasts are needed. Researching the best optimization strategy in those cases is a top priority.
- Identification of the model and diagnosis of residuals: Developing a functional methodology that would allow to identify the SARIMAHX model is found appealing. Moreover, it would be valuable to have a tool to diagnose the residuals of the model so as to determine if it is a white noise process (the model fitted is has extracted all temporal information of the series) or, on the contrary, there exists significant lags that are not being modeled. The functional autocorrelation measure defined in chapter 3 could be a starting point for the development of this methodology.
- Probabilistic approach: The model developed is designed to produce point-forecasts of functional time series. Some very valuable improvements would be the generation of confidence intervals and scenarios. In addition, the extensions to functional GARCH models would be a relevant contribution.
- Additional features: Taking advantage of the flexibility that gives the estimation procedure, some additional features could be added to the SARIMAHX model. Some proposals could be the inclusion of intervention analysis to model sudden changes in the time series, the use of warping functions for modeling side displacements of the functions, and the definition of a generalized version of the model for modeling nonlinear effects.
- Modeling the Residual Demand Curve in electricity markets is another potential research field. The extension of the new RDC proposed in section 4.5.1 to markets with multiple interconnecting zones is found appealing. Moreover, the inclusion of other complex conditions in the model can provide for a better generalization of the methodology.

Bibliography

[AGR14] A. Arribas-Gil and J. Romo, "Shape outlier detection and visualization for functional data: The outliergram", *Biostatistics*, vol. 15, no. 4, pp. 603–619, 2014.

[AM+13] M. C. Aguilera-Morillo, A. M. Aguilera, M. Escabias, and M. J. Valderrama, "Penalized spline approaches for functional logit regression", *Test*, vol. 22, no. 2, pp. 251–277, 2013.

[Ane+13] G. Aneiros, J. M. Vilar, R. Cao, and A. Muñoz San Roque, "Functional prediction for the residual demand in electricity spot markets", *IEEE Transactions on Power Systems*, vol. Early Access Online, 2013.

[Ant+06] A. Antoniadis, E. Paparoditis, and T. Sapatinas, "A functional wavelet kernel approach for time series prediction", *Journal of the Royal Statistical Society: Series B. Statistical Methodology*, vol. 68, no. 5, pp. 837–857, 2006. (visited on 09/03/2012).

[Ant+07] A. Antoniadis *et al.*, "Wavelet methods in statistics: Some recent developments and their applications", *Statistics Surveys*, vol. 1, pp. 16–55, 2007.

[Ant+10] J. Antoch, L. Prchal, M. Rosaria De Rosa, and P. Sarda, "Electricity consumption prediction with functional linear regression using spline estimators", *Journal of Applied Statistics*, vol. 37, no. 12, pp. 2027–2041, 2010. (visited on 09/04/2014).

[AP+11] G. Aneiros-Pérez, R. Cao, and J. M. Vilar-Fernández, "Functional methods for time series prediction: A nonparametric approach", *Journal of Forecasting*, vol. 30, no. 4, pp. 377–392, 2011. (visited on 08/31/2012).

[Bai+04] A. Baillo, M. Ventosa, M. Rivier, and A. Ramos, "Optimal offering strategies for generation companies operating in electricity spot markets", *IEEE Transactions on Power Systems*, vol. 19, no. 2, pp. 745–753, 2004.

[Bal+04] R. Baldick, R. Grant, and E. Kahn, "Theory and application of linear supply function equilibrium in electricity markets", *Journal of regulatory economics*, vol. 25, no. 2, pp. 143–167, 2004.

[Ban+08] H. Banakar, C. Luo, and B. T. Ooi, "Impacts of wind power minute-to-minute variations on power system operation", *IEEE Transactions on Power Systems*, vol. 23, no. 1, pp. 150–160, 2008.

[BB11] C. G. Baslis and A. G. Bakirtzis, "Mid term stochastic scheduling of a price maker hydro producer with pumped storage", *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 1856–1865, 2011.

[Ber+01] D. Berzal, J. I. delaFuente, and T. Gómez, "Building generation supply curves under uncertainty in residual demand curves for the dayahead electricity market", in *Power Tech Proceedings, 2001 IEEE Porto*, IEEE, vol. 1, 2001, 6–pp.

[Bes+00] P. C. Besse, H. Cardot, and D. B. Stephenson, "Autoregressive forecasting of some functional climatic variations", *Scandinavian Journal of Statistics*, vol. 27, no. 4, pp. 673–687, 2000.

[BF85] D Bunn and E. D. Farmer, "Comparative models for electrical load forecasting", 1985.

[Bis+14a] P. N. Biskas, D. I. Chatzigiannis, and A. G. Bakirtzis, "European electricity market integration with mixed market designs-part i: Formulation", *IEEE Transactions on Power Systems*, vol. 29, no. 1, pp. 458–465, 2014.

[Bis+14b] ——, "European electricity market integration with mixed market designs-part ii: Solution algorithm and case studies", *IEEE Transactions on Power Systems*, vol. 29, no. 1, pp. 466–475, 2014.

[BJ70] G. E. P. Box and G. Jenkins, *Time series analysis, forecasting and control*. Place of Publication: San Francisco, CA, USA. Country of Publication: USA.: Holden-Day, 1970, Book, Whole.

[Bos00] D. Bosq, *Linear processes in function spaces: Theory and applications*. Springer Verlag, 2000, vol. 149.

[BT75] G. E. Box and G. C. Tiao, "Intervention analysis with applications to economic and environmental problems", *Journal of the American Statistical association*, vol. 70, no. 349, pp. 70–79, 1975.

[Cam+16] F. A. Campos, A. M. n. S. Roque, E. F. Sánchez-Úbeda, and J. P. González, "Strategic Bidding in Secondary Reserve Markets", *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 2847–2856, Jul. 2016.

[Car+03] H. Cardot, F. Ferraty, and P. Sarda, "Spline estimators for the functional linear model", *Statistica Sinica*, pp. 571–591, 2003.

[Car+99] ——, "Functional linear model", *Statistics & Probability Letters*, vol. 45, no. 1, pp. 11–22, 1999.

[Cha14] M. Chaouch, "Clustering-Based Improvement of Nonparametric Functional Time Series Forecasting: Application to Intra-Day Household-Level Load Curves", *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 411–419, Jan. 2014.

[Chi12] J.-M. Chiou, "Dynamical functional prediction and classification, with application to traffic flow prediction", *The Annals of Applied Statistics*, pp. 1588–1614, 2012.

[Cru+11] A. Cruz, A. Muñoz, J. L. Zamora, and R. Espínola, "The effect of wind generation and weekday on spanish electricity spot price forecasting", *Electric Power Systems Research*, vol. 81, no. 10, pp. 1924–1935, 2011.

[Cyb89] G. Cybenko, "Approximation by superpositions of a sigmoidal function", *Mathematics of control, signals and systems*, vol. 2, no. 4, pp. 303–314, 1989.

[DC+05] B. F. De Castro, S Guillas, and W. G. Manteiga, "Functional samples and bootstrap for predicting sulfur dioxide levels", *Technometrics*, vol. 47, no. 2, pp. 212–222, 2005.

[del+02] S. delaTorre, J. M. Arroyo, A. J. Conejo, and J. Contreras, "Price maker selfscheduling in a poolbased electricity market: A mixed-integer lp approach", *IEEE Transactions on Power Systems*, vol. 17, no. 4, pp. 1037–1042, 2002.

[DG02] J. Damon and S. Guillas, "The inclusion of exogenous variables in functional autoregressive ozone forecasting", *Environmetrics*, vol. 13, no. 7, pp. 759–774, 2002.

[DG05] ——, "Estimation and simulation of autoregressive hilbertian processes with exogenous variables", *Statistical inference for stochastic processes*, vol. 8, no. 2, pp. 185–204, 2005.

[Did+11] D. Didericksen, P. Kokoszka, and X. Zhang, "Empirical properties of forecasts with the functional autoregressive model", *Computational Statistics*, pp. 1–14, 2011.

[DM05] J. A. Dubin and H.-G. Müller, "Dynamical correlation for multivariate longitudinal data", *Journal of the American Statistical Association*, vol. 100, no. 471, pp. 872–881, 2005.

[DM95] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy", *Journal of Business & Economic Statistics*, vol. 13, no. 3, pp. 253–263, 1995.

[Far97] J. J. Faraway, "Regression analysis for a functional response", *Technometrics*, pp. 254–261, 1997. (visited on 09/03/2012).

[FV06] F. Ferraty and P. Vieu, *Nonparametric Functional Data Analysis: THEORY And Practice*, en. Springer, Jun. 2006.

[FZ08] J. Fan and W. Zhang, "Statistical methods with varying coefficient models", *Statistics and its Interface*, vol. 1, no. 1, p. 179, 2008.

[GB00] J GarciaGonzalez and J Barquin, "Self unit commitment of thermal units in a competitive electricity market", in *Power Engineering Society Summer Meeting, 2000. IEEE*, IEEE, vol. 4, 2000, pp. 2278–2283.

[GM12] A. Gottlieb and H.-G. Müller, "A stickiness coefficient for longitudinal data", *Computational Statistics & Data Analysis*, vol. 56, no. 12, pp. 4000–4010, 2012.

[Hal+06] P. Hall, H.-G. Müller, and J.-L. Wang, "Properties of principal component methods for functional and longitudinal data analysis", *The annals of statistics*, pp. 1493–1517, 2006.

[Hay+12] S. Hays, H. Shen, and J. Z. Huang, "Functional dynamic factor models with application to yield curve forecasting", *The Annals of Applied Statistics*, pp. 870–894, 2012.

[He+00] G. He, H. G. Muller, and J. L. Wang, "Extending correlation and regression from multivariate to functional data", *Asymptotics in statistics and probability*, pp. 197–210, 2000. (visited on 07/27/2012).

[He+10] G. He, H.-G. Müller, J.-L. Wang, and W. Yang, "Functional linear regression via canonical analysis", *Bernoulli*, pp. 705–729, 2010.

[HE15] T. Hsing and R. Eubank, *Theoretical foundations of functional data analysis, with an introduction to linear operators*. John Wiley & Sons, 2015.

[HF10] R. J. Hyndman and S. Fan, "Density Forecasting for Long-Term Peak Electricity Demand", *IEEE Transactions on Power Systems*, vol. 25, no. 2, pp. 1142–1153, May 2010. (visited on 01/21/2016).

[HH+07] P. Hall, J. L. Horowitz, *et al.*, "Methodology and convergence rates for functional linear regression", *The Annals of Statistics*, vol. 35, no. 1, pp. 70–91, 2007.

[HK12] L. Horváth and P. Kokoszka, *Inference for functional data with applications*. Springer, 2012, vol. 200. (visited on 07/05/2013).

[HL12] P. Holmberg and E. Lazarczyk, "Congestion management in electricity networks", 2012.

[Hor+16] L. Horváth, G. Rice, and S. Whipple, "Adaptive bandwidth selection in the long run covariance estimator of functional time series", *Computational Statistics & Data Analysis*, vol. 100, pp. 676–693, 2016.

[HR16] G. Hooker and S. Roberts, "Maximal autocorrelation functions in functional data analysis", *Statistics and Computing*, vol. 26, no. 5, pp. 945–950, 2016.

[HS09] R. J. Hyndman and H. L. Shang, "Forecasting functional time series", *Journal of the Korean Statistical Society*, vol. 38, no. 3, pp. 199–211, 2009. (visited on 09/03/2012).

[HSU07] R. J. Hyndman and M. Shahid Ullah, "Robust forecasting of mortality and fertility rates: A functional data approach", *Computational Statistics & Data Analysis*, vol. 51, no. 10, pp. 4942–4956, Jun. 2007. (visited on 07/25/2012).

[Jol02] I. Jolliffe, *Principal component analysis*. Wiley Online Library, 2002.

[Jon+13] T. Jonsson, P. Pinson, H. A. Nielsen, H. Madsen, and T. S. Nielsen, "Forecasting Electricity Spot Prices Accounting for Wind Power Predictions", *IEEE Transactions on Sustainable Energy*, vol. 4, no. 1, pp. 210–218, Jan. 2013.

[Kle+16] J. Klepsch, C. Klüppelberg, and T. Wei, "Prediction of functional ARMA processes with an application to traffic data", *ArXiv:1603.02049*, Mar. 2016. (visited on 09/02/2016).

[KO08] V. Kargin and A. Onatski, "Curve forecasting by functional autoregression", *Journal of Multivariate Analysis*, vol. 99, no. 10, pp. 2508–2526, Nov. 2008. (visited on 01/31/2012).

[Koo+07] S. J. Koopman, M. Ooms, and M. A. Carnero, "Periodic seasonal reg-arfima–garch models for daily electricity spot prices", *Journal of the American Statistical Association*, vol. 102, no. 477, pp. 16–27, 2007.

[Lee+11] Y.-Y. Lee, J. Hur, R. Baldick, and S. Pineda, "New indices of market power in transmission-constrained electricity markets", *IEEE Transactions on Power Systems*, vol. 26, no. 2, pp. 681–689, 2011.

[LH+10] Y. Li, T. Hsing, *et al.*, "Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data", *The Annals of Statistics*, vol. 38, no. 6, pp. 3321–3351, 2010.

[Lie13] D. Liebl, "Modeling and forecasting electricity spot prices: A functional data perspective", *The Annals of Applied Statistics*, vol. 7, no. 3, pp. 1562–1592, 2013. (visited on 09/09/2014).

[MÖ8] H.-G. Müller, "Functional modeling of longitudinal data", *Longitudinal data analysis (Handbooks of modern statistical methods). CRC, New York*, pp. 223–252, 2008.

[Ma+04] Y. Ma, P. B. Luh, K. Kasiviswanathan, and E. Ni, "A neural network-based method for forecasting zonal locational marginal prices", in *Power Engineering Society General Meeting, 2004. IEEE*, IEEE, 2004, pp. 296–302.

[Mar+08] V. Marques, I. Soares, and A. Fortunato, "Measuring market power in the wholesale electricity iberian market through the residual demand curve elasticity", in *Electricity Market, 2008. EEM 2008. 5th International Conference on European*, IEEE, 2008, pp. 1–7.

[Mas05] E. Masry, "Nonparametric regression estimation for dependent functional data: Asymptotic normality", *Stochastic Processes and their Applications*, vol. 115, no. 1, pp. 155–177, Jan. 2005. (visited on 09/12/2013).

[Mou02] T. Mourid, "Estimation and prediction of functional autoregressive processes", *Statistics: A Journal of Theoretical and Applied Statistics*, vol. 36, no. 2, pp. 125–138, 2002.

[MP11] A. Mas and B. Pumo, "Linear Processes for Functional Data", *The Oxford Handbook of Functional Data*, pp. 47–71, 2011.

[NC06] F. Nogales and A. J. Conejo, "Electricity price forecasting through transfer function models", *Journal of the Operational Research Society*, vol. 57, no. 4, pp. 350–356, 2006.

[Nie+14] A. A. S. d. l. Nieta, J. Contreras, J. I. Muñoz, and M. O'Malley, "Modeling the Impact of a Wind Power Producer as a Price-Maker", *IEEE Transactions on Power Systems*, vol. 29, no. 6, pp. 2723–2732, Nov. 2014.

[OS08] M. Olsson and L. Soder, "Modeling real-time balancing power market prices using combined sarima and markov processes", *IEEE Transactions on Power Systems*, vol. 23, no. 2, pp. 443–450, 2008.

[PA+13] I. J. PérezArriaga *et al.*, *Regulation of the power sector*. Springer, 2013.

[Pan91] A. Pankratz, *A Primer on ARIMA Models*. Wiley Online Library, 1991, pp. 24–81.

[Pel13] M. Pelagatti, "Supply function prediction in electricity auctions", in *Complex Models and Computational Methods in Statistics*, Springer, 2013, pp. 203–213.

[PK16] M. Parvania and R. Khatami, "Continuous-Time Marginal Pricing of Electricity", *IEEE Transactions on Power Systems*, vol. PP, no. 99, pp. 1–1, 2016.

[Por+16a] J. Portela, A. Muñoz, and E. Alonso, "Forecasting functional time series with a new hilbertian armax model: Application to electricity price forecasting", Working paper, "2016.

[Por+16b] J. Portela, A. Muñoz, E. Sanchez-Ubeda, J. Garcia-Gonzalez, and R. Gonzalez, "Residual demand curves for modeling the effect of complex offering conditions on day-ahead electricity markets", *IEEE Transactions on Power Systems*, vol. PP, no. 99, pp. 1–1, 2016.

[PS13] E. Paparoditis and T. Sapatinas, "Short-Term Load Forecasting: The Similar Shape Functional Time-Series Predictor", *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 3818–3825, Nov. 2013.

[PS16] M. Parvania and A. Scaglione, "Unit Commitment With Continuous-Time Generation and Ramping Trajectory Models", *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 3169–3178, Jul. 2016.

[Ram06] J. O. Ramsay, *Functional data analysis*. Wiley Online Library, 2006.

[Rañ+15] P Raña, G Aneiros, and J. M. Vilar, "Detection of outliers in functional time series", *Environmetrics*, vol. 26, no. 3, pp. 178–191, 2015.

[Reg14] M. Reguant, "Complementary bidding mechanisms and startup costs in electricity markets", *The Review of Economic Studies*, vol. 81, no. 4, pp. 1708–1742, 2014.

[RS90] F. Riesz and B. Sz-Nagy, *Functional Analysis*, en. Courier Corporation, 1990.

[Sal+15] N. Salish, A. Gleim, and A. Statkraft, "Forecasting methods for functional time series", Working paper, University of Bonn, Tech. Rep., 2015.

[Sen10] R. Sen, "Time series of functional data", *Technical report*, 2010.

[SL15] I. Shah and F. Lisi, "Day-ahead electricity demand forecasting with nonparametric functional models", in *2015 12th International Conference on the European Energy Market (EEM)*, May 2015, pp. 1–5.

[Tim+11] C. Timmermans, L. Delsol, and R. VonSachs, "Bases giving distances. a new semimetric and its use for nonparametric functional data analysis", in *Recent Advances in Functional Data Analysis and Related Topics*, Springer, 2011, pp. 307–313.

[Tur+07] C. Turbillon, J.-M. Marion, B. Pumo, and R. Lenôtre, "Estimation of the moving-average operator in a hilbert space", *Recent Advances in Stochastic Modeling and Data Analysis, World Sci. Publ., Hackensack, NJ*, pp. 597–604, 2007.

[UF13] S. Ullah and C. F. Finch, "Applications of functional data analysis: A systematic review", *BMC medical research methodology*, vol. 13, no. 1, p. 43, 2013.

[Uge+03] A Ugedo, E Lobato, A Franco, L Rouco, J Fernandez-Caro, J de-Benito, J Chofre, and J De-la-Hoz, "Stochastic model of residual demand curves with decision trees", in *Power Engineering Society General Meeting, 2003, IEEE*, IEEE, vol. 2, 2003, pp. 979–984.

[Val+02] M. J. Valderrama, F. A. Ocana, and A. M. Aguilera, "Forecasting PC-ARIMA models for functional data", *Proceedings in computational statistics*, pp. 25–36, 2002.

[Vil+12] J. M. Vilar, R. Cao, and G. Aneiros, "Forecasting next-day electricity demand and price using nonparametric functional methods", *International Journal of Electrical Power & Energy Systems*, vol. 39, no. 1, pp. 48–55, 2012. (visited on 03/12/2015).

[Váz+14] S. Vázquez, P. Rodilla, and C. Batlle, "Residual demand models for strategic bidding in european power exchanges: Revisiting the methodology in the presence of a large penetration of renewables", *Electric power systems research*, vol. 108, pp. 178–184, 2014.

[Wan+15] J.-L. Wang, J.-M. Chiou, and H.-G. Müller, "Review of functional data analysis", *ArXiv preprint arXiv:1507.05135*, 2015.

[Wer14] R. Weron, "Electricity price forecasting: A review of the state-of-the-art with a look into the future", en, *International Journal of Forecasting*, vol. 30, no. 4, pp. 1030–1081, Oct. 2014. (visited on 03/10/2016).

[Wol03] F. A. Wolak, "Measuring unilateral market power in wholesale electricity markets: The california market, 19982000", *The American economic review*, vol. 93, no. 2, pp. 425–430, 2003.

[WY02] C. O. Wu and K. F. Yu, "Nonparametric varying-coefficient models for the analysis of longitudinal data", *International Statistical Review*, vol. 70, no. 3, pp. 373–393, 2002.

[XB07] L. Xu and R. Baldick, "Transmissionconstrained residual demand derivative in electricity markets", *IEEE Transactions on Power Systems*, vol. 22, no. 4, pp. 1563–1573, 2007.

[Xu+11a] L. Xu, R. Baldick, and Y. Sutjandra, "Bidding into electricity markets: A transmission-constrained residual demand derivative approach", *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1380–1388, 2011.

[Xu+11b] ——, "Transmission-constrained inverse residual demand jacobian matrix in electricity markets", *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2311–2318, 2011.

[Yao+05a] F. Yao, H. G. Müller, and J. L. Wang, "Functional linear regression analysis for longitudinal data", *The Annals of Statistics*, vol. 33, no. 6, pp. 2873–2903, 2005.

[Yao+05b] F. Yao, H.-G. Müller, and J.-L. Wang, "Functional data analysis for sparse longitudinal data", *Journal of the American Statistical Association*, vol. 100, no. 470, pp. 577–590, 2005.

[Aue+15] A. Aue, D. Norinho, and S. Hörmann, "On the prediction of stationary functional time series", *Journal of the American Statistical Association*, vol. 110, no. 509, pp. 378–392, 2015. (visited on 10/08/2015).