COMILLAS
UNIVERSIDAD PONTIFICIA
ICAI

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

PHD THESIS

PROBABILISTIC FORECASTING OF FUNCTIONAL TIME SERIES: APPLICATION TO SCENARIO-GENERATION OF RESIDUAL DEMAND CURVES IN ELECTRICITY MARKETS

Author: Guillermo Mestre Marcos

Supervisors:
Antonio Muñoz San Roque
Estrella Alonso Pérez

Madrid
July 2021

This dissertation was typeset with LaTeX and compiled in TeXmaker using the MacTeX-2016 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®, R® and MATLAB®.

A mis padres

Agradecimientos

A mis directores de tesis, Antonio Muñoz y Estrella Alonso. Gracias por guiarme en este camino y apoyarme en los momentos difíciles. A Antonio, por la confianza que depositaste en mi para realizar una tesis bajo tu supervisión, por tu entera disposición y por compartir conmigo tu experiencia. A Estrella, por tu dedicación y apoyo constante, que ha sido fundamental para superar momentos difíciles. Sin vuestro apoyo y supervisión este trabajo no habría sido posible. Gracias a los dos.

Durante mi estancia en el Instituto de Investigación Tecnológica (IIT) he tenido el privilegio de coincidir con compañeros brillantes, tanto profesional como personalmente. Mención especial merece José Portela, con quien tantas horas he compartido en estos últimos cuatro años. Gracias por tantos buenos momentos y por tu ayuda a lo largo de esta etapa. A Eugenio Sánchez, de quien tanto he aprendido. Gracias por haber compartido conmigo tu afán investigador, echaré de menos esas reuniones en tu despacho que acababan con la pizarra llena de nuevas ideas.

Agradecer a los miembros del Departamento de Matemática Aplicada su cercanía y constante apoyo a lo largo de esta tesis. A Santiago Cano, por todo el interés que has mostrado en mi trabajo y por haber confiado en mí.

A mis compañeros IeFs, por tantos buenos ratos juntos. Gracias a Santi, Fran, Miguel, Fernando, Rafa, Otaola, Teresa, Jaime, Pedro, Salva y David por vuestros ánimos y apoyo durante este periodo.

A mi familia y amigos, por haberme acompañado en esta aventura. A mi hermano Jose Manuel, por su capacidad para ayudarme a desconectar y saber siempre qué decir para animarme. Vuestra ayuda y cercanía ha sido esencial en esta etapa.

A Cristina, por todo tu apoyo y la infinita paciencia que has tenido conmigo. Gracias por haber estado a mi lado en todo momento. Sin tu ayuda y apoyo constante estoy seguro de que no habría podido terminar esta tesis. Gracias de corazón.

A mis padres, por vuestro apoyo incondicional y la confianza que siempre habéis tenido en mí. Gracias por haberme enseñado a perseverar, a enfrentarme a las adversidades y a esforzarme para conseguir mis objetivos. Sin vosotros nunca habría llegado hasta aquí.

Contents

List of Figures

List of Tables

Abstract

Due to recent technological advances in science and computing, the collection and analysis of high-dimensional data has become a relevant topic in the field of applied statistics. Many of these observations arise from continuous processes observed over time, which can be naturally expressed as functions. Thus, the name functional data is commonly given to such observations. Statistical methodologies that deal with the analysis of functional data belong to the Functional Data Analysis (FDA) framework, a field of research devoted to the exploration and modelization of these data.

If these functional data are collected sequentially over time, it is natural to expect a time dependence between functional observations. The term Functional Time Series (FTS) is often used to refer to these processes. These series of functional observations appear frequently in numerous fields, such as medical research, finance and ecology. Nevertheless, this thesis is concerned with the applications in electricity markets that involve FTS. Particularly, the bidding information of market auctions can be used to define Residual Demand Curves (RDC), functions that model the market-clearing price as a function of the energy that agents are willing to trade. Having accurate estimations of the hourly series of RDCs is of utmost importance for market agents, as it would provide them with an accurate description of the bidding strategy of its competitors. RDCs often exhibit complex dynamics, as the bidding behaviour of the market agents is often influenced by external factors and often exhibits a strong seasonal dependence. The SARMAHX model is a generalization of the scalar ARMAX model to the functional data framework, which has proven to be an adequate model to obtain accurate short-term forecasts of FTS that appear in electricity markets.

The work described in this thesis is aimed at developing a novel functional probabilistic model that is able to generate coherent scenarios of functional time series. Probabilistic forecasts of the series are obtained by quantifying the uncertainty associated with the forecasts of a fitted SARMAHX model. Hence, a forecasting methodology for functional time series has been developed, which involves the identification and diagnosis of the SARMAHX model fitted to the series. Once deterministic forecasts of the series have been obtained, the conditional distribution of the residuals is characterized by the proposed probabilistic model, which then is used to generate future scenarios of the curves.

The proposed probabilistic forecasting methodology has been successfully applied to the short-term estimation of the series of hourly Residual Demand Curves in different electricity markets, outperforming other reference models commonly found in the functional data literature.

Resumen

En las últimas dos décadas el registro y análisis de datos de alta frecuencia ha cobrado una gran importancia dentro del ámbito de la Estadística. Muchas de estas observaciones provienen de procesos continuos que pueden ser tratados como funciones, por lo que comúnmente se les conoce como datos funcionales. Los métodos estadísticos desarrollados para analizar este tipo de datos se engloban dentro del ámbito del Análisis de Datos Funcionales (FDA, por sus siglas en inglés), un campo de la Estadística que busca desarrollar técnicas que permitan analizar y modelar conjuntos de observaciones funcionales.

Un tipo de datos funcionales muy importante por sus numerosas aplicaciones son las series temporales funcionales (STF), secuencias temporales de observaciones donde cada elemento de la serie es una función continua definida en un intervalo cerrado. Las series temporales funcionales aparecen con frecuencia en el ámbito de la investigación médica, las finanzas, la ecología y la energía, entre otras. El objetivo principal de esta tesis es el estudio de las series temporales funcionales que aparecen en distintas aplicaciones del sector de los mercados eléctricos, como pueden ser las Curvas de Demanda Residual. Estas curvas son funciones que determinan el precio de casación de la subasta en función de la cantidad de energía que los agentes están dispuestos a comprar y vender. Un agente que sea capaz de tener estimaciones precisas de estas curvas podrá optimizar su propia estrategia de oferta, ya que las curvas contienen información relevante sobre las estrategias de sus competidores. Sin embargo, existe una gran dificultad en obtener estimaciones robustas de las Curvas de Demanda Residual, debido a que suelen presentar una estructura dinámica compleja influenciada por numerosos factores externos. Uno de los modelos que han obtenido mejores resultados al modelar este tipo de series funcionales ha sido el modelo SARMAHX, un modelo que generaliza el modelo ARMAX clásico al ámbito funcional.

El objetivo principal de esta tesis es el desarrollo de un modelo probabilista funcional que permita generar escenarios funcionales coherentes de una serie temporal funcional. Estas predicciones probabilistas se obtienen caracterizando la distribución de los residuos de un modelo SARMAHX ajustado a la serie funcional. Para ello, se ha desarrollado una metodología de predicción probabilista para series temporales funcionales que requiere de un nuevo método de identificación y diagnóstico del modelo SARMAHX. Una vez que se ha ajustado el modelo SARMAHX se usan técnicas no-paramétricas para caracterizar la distribución condicional de los residuos del modelo, lo que permite generar escenarios realistas de la serie temporal funcional.

La metodología de predicción probabilista propuesta se ha aplicado con éxito a la predicción a corto plazo de Curvas de Demanda Residual en varios mercados de electricidad, obteniendo mejores resultados que el resto de modelos probabilistas funcionales de referencia.

1
Introduction

This first chapter introduces some justifications of this thesis as well as its main objectives. In addition, it provides the reader with a general overview of the organization and the outline of the dissertation in order to make it easier to follow.

1.1. Motivation

Due to recent technological advances in science and industry, the collection and analysis of high-dimensional data has become an important topic in the field of applied statistics. Many of these observations arise from continuous-valued processes over time, space or frequency domains, among others. These observations can be naturally expressed as functions; thus the name functional data is commonly given to such observations. The need of statistical methodologies to explore and analyze the structure of these high-dimensional data has led to the development of the Functional Data Analysis (FDA) framework, an active field of research that deals with the theory and methods required to extract information from functional observations.

Functional Data Analysis techniques have been successfully applied in numerous fields where the observations arise from a continuous process. These new statistical tools have been used in the study of demographic curves (Hyndman and Shang, 2009), the analysis of pollution curves (Aue, Norinho, et al., 2015) and finance, where FDA techniques have been used to model the dynamic evolution of yield curves (Sen and Klüppelberg, 2019). The applicability of functional methods to medical sciences is illustrated in Ullah and Finch (2013), where the authors perform a systematic review of the different applications of FDA to biomedical and public health problems. The ability of extracting information from a set of curves or finding relationship between two functional datasets was crucial to many of the case studies analyzed by the authors, highlighting the importance of these statistical tools across different research fields.

If these functional data are collected sequentially over time, it is natural to expect a time dependence between functional observations. This motivates the consideration of functional time series (FTS), a relevant type of functional data due to its wide range of applications in the industry. Due to the dependence structure and complex dynamics often exhibited by FTS, the

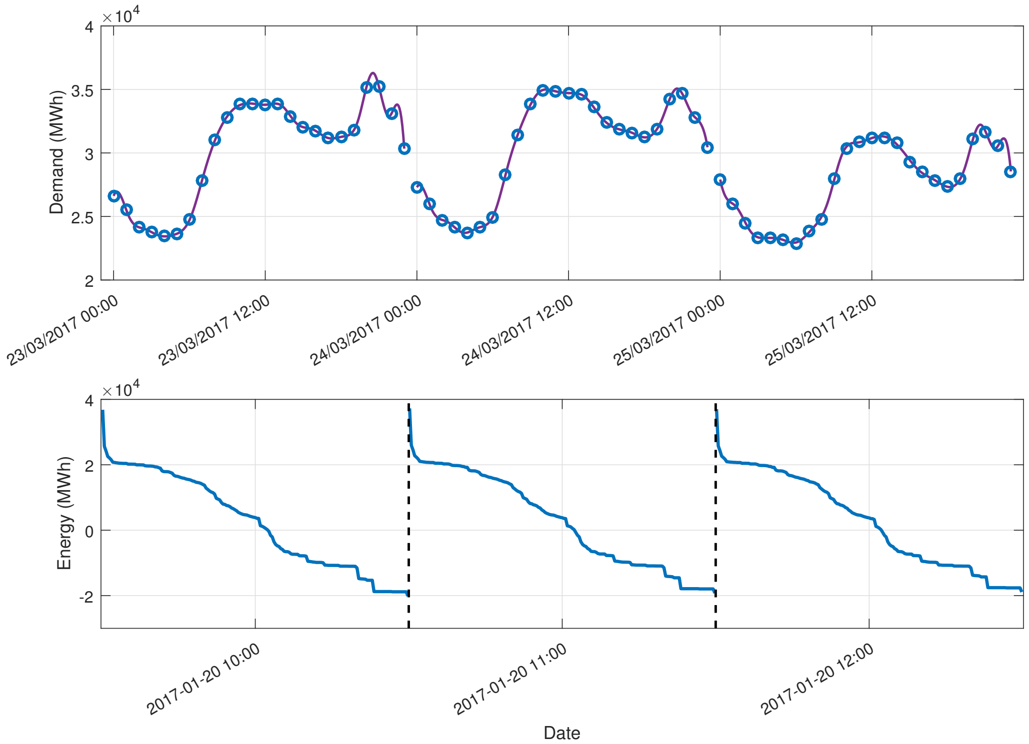

Figure 1.1. Top: discrete observations of the electricity load in the Spanish electricity market for three consecutive days together with the daily load profiles obtained using FDA smoothing techniques. Bottom: Residual Demand Curves for the Spanish Day-Ahead electricity market for three different hours.

analysis of these series requires the development of new statistical techniques that are able to take into account the infinite-dimensionality nature of these data.

Functional time series appear in numerous applications in electricity markets. Figure 1.1 shows two different examples of these series: the top panel shows the daily demand profile of three consecutive days in the Spanish system. Electricity load values have been registered each hour. However, the demand is a continuous process, having a different value for each time instant, which makes FDA techniques suitable for analyzing this dataset. Functional smoothing techniques have been used to obtain functional observations from the discrete consumption values, which exhibit continuous variations of demand throughout the day. Once the functional time series of daily demand profiles has been obtained, functional forecasting methods can be used to obtain short-term estimations of these curves, which is crucial for market agents and System Operators. Traditional demand forecasting methods aimed at estimating the 24 hourly values of the demand using non-functional models. However, the use of smart meters is increasing the sample frequency of electricity consumption from hourly values to a few minutes, highlighting the necessity of modelling the demand as a continuous-time process.

The previous series was obtained by slicing a continuous-time process into segments of equal length. However, there are certain types of functional time series where each observation is a high-dimensional process. An illustrative example of these series are Supply and Demand curves, which are a relevant part of electricity markets. Many Day-Ahead electricity markets are based on sealed-bid auctions where agents submit offers as price-quantity pairs (p, q), which refers to the amount of energy q the agent is willing to sell (buy) at price p. If these selling (buying) offers are sorted in increasing (decreasing) prices, the aggregated supply (demand) curve of the agent is built, which represents the bidding behaviour of the agent.

The analysis of Supply and Demand curves are of utmost importance to electricity companies, as having accurate estimations of the bidding strategy of its competitors allows them to optimize its own strategy in the market. Residual Demand Curves (RDC) are often used by market agents in this situation. RDCs expresses the clearing price of energy auctions as a function of the amount of energy the agent is willing to buy or sell. Hence, each auction defines one different Residual Demand Curve, which then can be used to market agents to analyze the bidding and offering behaviour of its competitors. The bottom panel of Figure 1.1 shows a set of three hourly Residual Demand Curves (RDC) for the Spanish Day-Ahead electricity market.

When modelling functional time series it is necessary to quantify the serial dependency exhibited by each element of the series, as well as identifying any trend or seasonal behaviour of the series. Classical identification methods for scalar time series, such as the Box-Jenkins methodology (Box et al., 2008), use the sample autocorrelation and partial autocorrelation function to identify the underlying serial correlation structure of the series. The generalization of these techniques to the functional data framework would be of great interest, as it would allow practitioners to include this information into functional forecasting models, improving its performance and obtaining more robust forecasts of the series.

Once a forecasting model has been identified and fitted to a given functional time series, future values of the series can be estimated. Although point-forecasts may be adequate in some cases, the vast majority of real-world applications require having probabilistic forecasts of future events. For example, market agents are often interested in having probabilistic forecasts of Residual Demand Curves, as it provides them with a measure of the future volatility of the market prices. This can be achieved by quantifying the uncertainty associated with the model's point-forecasts, in order to capture changes in the distribution of the residuals. Although there is a growing literature about deterministic functional forecasting models, the same cannot be said for the development of probabilistic models for functional time series. The infinite-dimensionality of the data hinders the definition of a proper density function, which makes measuring the probabilistic uncertainty of forecasts not as straightforward for functional datasets than for scalar series.

This dissertation contributes to the current state of the art by proposing a novel forecasting methodology that can be used to obtain sharp probabilistic forecasts of functional time series observed in electricity markets. The first original contribution is the definition of a functional version of the autocorrelation and partial autocorrelation functions for functional time series based on the L^2 norm of its lagged autocorrelation operators. These functions can then be used for selecting the order and assessing the adequacy of functional forecasting models, obtaining more robust estimations of the curves. The second contribution of this thesis is the proposal of a novel conditional probabilistic model for functional time series that it is able to capture the variations in the distribution of a set of functional residuals using a set of exogenous covariates. The proposed probabilistic model is able to outperform other reference models found in the functional data literature in several real-world applications involving Residual Demand Curves, obtaining coherent scenarios of the curves as well as sharper prediction bands for the forecasts. Additionally, scenario generation methods are of utmost importance in stochastic optimization (Kaut and Wallace, 2007), as they provide a method to represent the distribution of the input data of the problem. Henceforth, the scenario-generation model proposed in this thesis could be used to improve certain stochastic optimization models proposed in the literature, such as the offering optimization model proposed in Baillo et al. (2004).

1.2. Thesis objectives

The main objective of this thesis is to develop a novel functional probabilistic model that is able to generate coherent scenarios of the functional time series being modelled as well as quantifying the uncertainty associated with the forecasts of a given deterministic model. The scenarios obtained must reflect this forecasting uncertainty as well as exhibiting similar shape features as the original functional time series. This requires the definition of a new forecasting methodology for functional time series involving the identification and diagnosis of functional parametric models as well as the development of a probabilistic model that is able to capture changes in the distribution of the residuals of the aforementioned model using exogenous variables. Hence, the specific objectives pursued in this dissertation are summarized below:

- Development of an identification and diagnosis methodology for functional forecasting models, which can be used to identify relevant features of a functional time series and to asses the adequacy of a forecasting model fitted to that series.
- Development of a probabilistic model for functional time series that fills an existing gap in the functional probabilistic literature. The proposed model must obtain sharper probabilistic estimations about the series. Additionally, the new model must be able to generate functional scenarios that are coherent with both the actual functional time series and the point-forecasts of the deterministic model used to estimate the series, exhibiting similar shape features as the actual curves.
- Validate the proposed forecasting methodology by applying it to different real-world datasets obtained from electricity markets. The performance of the proposed models must be compared against other reference models found in the literature, highlighting the improvements made by the techniques developed in this thesis when applied to real-world problems.

1.3. Dissertation outline

This dissertation consists of six chapters including this first introductory one. The aim of Chapter 2 is twofold: on the one hand it provides a taxonomy of the different FDA techniques proposed in the literature, introducing the main tools used when analyzing functional data. The representation of curves using basis functions is addressed, together with the use of summary statistics as a tool for describing a functional dataset. On the other hand, functional time series are introduced, together with a review of the different forecasting models proposed in the literature for these data. Special emphasis is given to the SARMAHX model, a generalization of the scalar ARMAX model to the functional framework using linear operators. This model is able to model complex relationships between curves, proving itself to be a competitive model against other well-known methods for price forecasting. This makes the SARMAHX model a suitable model for forecasting functional time series from electricity markets, such as Supply Curves and Residual Demand Curves. Finally, a review of probabilistic models for functional time series is performed, analyzing how an analogue for the density function of a functional dataset can be defined, as well as describing the different probabilistic models proposed in the literature.

Chapter 3 presents a functional version of the autocorrelation and partial autocorrelation functions for functional time series based on the L^2 norm of the lagged autocovariance operators of the series. The proposed tools can be used to select the order and assessing the adequacy of functional SARMAHX models. The applicability of this identification and diagnosis methodology is illustrated with an application to several real-world datasets.

Chapter 4 is devoted to the development of the conditional probabilistic model for functional time series proposed in this dissertation. The model estimates the conditional density of the residuals of a fitted deterministic model using non-parametric techniques. By characterizing the distribution of the residuals, prediction regions and scenarios of a functional time series can be obtained. The applicability of the proposed model is validated in both a simulation study using synthethic data and a real-world problem using Complex Residual Demand Curves from the Iberian Day-Ahead market, outperforming the probabilistic forecasts obtained by other reference probabilistic models found in the functional data literature.

The applicability of the forecasting methodology to electricity markets is illustrated in Chapter 5, where two different real-world case studies are analyzed. Firstly, an application to the short-term forecasting of Supply Curves in the Italian Day-Ahead electricity market is presented. The identification methodology is used to select the order of a SARMAHX model, which is then fitted to the data to obtain one-step and 24-step ahead forecasts of the series. Additionally, probabilistic forecasts are obtained using the proposed conditional probabilistic model, which uses exogenous variables to capture changes in the distribution of the curves. Secondly, the short-term forecasting of Residual Demand Curves in the Spanish Day-Ahead electricity market is addressed. Similarly to the previous study, a SARMAHX model is identified and fitted to the hourly series. Once deterministic forecasts of the series have been obtained, the conditional probabilistic model is used to generate future scenarios of the Residual Demand Curves. The model is compared against other reference probabilistic models, obtaining sharper estimations of the series with the proposed model. These scenarios are used to obtain density forecasts of the market clearing price, illustrating the applicability of the probabilistic forecasts obtained.

Finally, Chapter 6 provides the concluding remarks of the dissertation, summarizing the contributions and future developments.

2

Functional time series analysis models and methods

This chapter serves the dual purpose of introducing the necessary concepts that will be used throughout this thesis and to provide an extensive review of the main functional forecasting methods found in the literature, with a discussion about probabilistic techniques proposed in the literature. The information is organized from the general to the more specific, starting with the definition of the main concepts and terms used in functional data analysis, and gradually progressing to the specific forecasting models and techniques proposed in the literature. Finally, probabilistic models for functional time series are introduced, including a review of scenario generation methods and measures to evaluate the performance of such probabilistic methods.

2.1. Introduction

The main purpose of statistics is to obtain information from a set of observations $x_1, x_2, \ldots, x_N$. In classical statistics, it is typically assumed that the observations are elements of $\mathbb{R}^d$, where each dimension of the feature space corresponds to a measured feature of the data. However, due to recent technological advances the collection and analysis of high frequency data has become an important research topic in the field of applied statistics. Many of these data X arise from continuous-valued processes over time, space or frequency domains, among others. The observation of each continuous-valued process over a given value $v \in \mathbb{R}$ of its domain will be denoted by $X(v)$.

Due to the high frequency of these processes, the values of each function $X_1, X_2, \ldots, X_N$ are not often observed, only the values of the curves measured on a reduced set of d discretization points $v_1, v_2, \ldots, v_d$. As such, classical multivariate techniques would only capture the information of the values of the curves in those discretization points $X_i(v_1), X_i(v_2), \ldots, X_i(v_d)$, without accounting for relevant information of the curves such as the shape of the curves. In addition, treating the observations as curves provides a more flexible framework for such data, as it does not require that the observations are measured at equally spaced intervals, nor that there are the same number of measurements for all curves. This makes multidimensional techniques obsolete for modeling these continuous-valued processes, highlighting the need for developing new statistical techniques that are able to model the behavior of these high-frequency processes without losing information about the infinite-dimensionality of the original curves. As these observations can be naturally expressed as functions; the name functional data is used to refer to this kind of data. This curve-based treatment of these functional data makes it possible to take into account the underlying structure of the curves, providing a more accurate description of the data.

This dissertation is concerned with these functional data, which are viewed as functions defined over some bounded interval $V \subset \mathbb{R}$. Therefore, a random function is defined as a function $\mathcal{X}$ defined over a probability space $(\Omega, \mathcal{H}, P)$, where Ω is a set, $\mathcal{H}$ denotes a $\sigma-$algebra on $\mathcal{H}$ and P denotes a probability measure on $(\Omega, \mathcal{H})$. A probability space is a mathematical model of a random experiment whose exact outcome cannot be told in advance (Çinlar, 2011). The set Ω stands for the collection of all possible outcomes of the experiment, which means that for each element $w \in \Omega$, $\mathcal{X}(w)$ is a deterministic function:

$$\begin{aligned} \mathcal{X} : \Omega &\to E \\ w &\mapsto \mathcal{X}(w) \end{aligned} \tag{2.1}$$

Throughout this dissertation, it will be assumed that all realizations $\mathcal{X}(w)$ are elements of the functional space $L^2(V)$ of square integrable functions defined over the bounded interval V. Following the standard definition, a functional stochastic process $\{\mathcal{X}_i\}_{i=1}^N$ is defined as a collection of random functions ordered in time.

The need of statistical methodologies to explore and analyze the structure of these high-dimensional data has led to the development of the Functional Data Analysis (FDA) framework, an active field of research that deals with the theory and methods required to extract information from functional observations. The field of FDA has seen rapid development over the last two decades, with many of its methods being applied quite broadly in several diverse fields. FDA has been applied in medical research, such as Erbas et al. (2007), where functional data models are used to forecast age-specific breast cancer mortality; or Hermanussen (2010), where human growth curves are analyzed. Functional time series models were used in Besse, Cardot, et al. (2000) to study the annual cycle of sea surface temperatures, illustrating the application of functional data to meteorology. Regarding the application of functional data in finance, the process of price formation in online auctions was studied in Bapna et al. (2008) using a functional regression model. The growing interest in functional data highlights the need for new methodologies that improve the treatment of this data. The work by Ramsay and Silverman (2005) provides a solid introduction to functional data, covering topics such as smoothing techniques and functional principal component analysis and laying the foundations of FDA as a formal sub-branch of statistics. The book Ramsay, Hooker, et al. (2009) explains how to implement those techniques using the R package `fda`, developed by the authors as a companion package to the book. Nonparametric regression models for functional data and their mathematical foundation are presented in Ferraty and Vieu (2006), while the monograph Bosq (2000) presents the mathematical theory behind functional linear time series models, establishing properties behind the functional autoregresive model within an even more general Banach space framework. Finally, the monographs Horváth and Kokoszka (2012) and Kokoszka and Reimherr (2017) present a general introduction to functional time series analysis and statistical inference methods.

This thesis is concerned with the study of functional time series (FTS), defined as the realization of a functional stochastic process collected over time. The temporal dependence exhibited by these curves must be taken into account when analyzing them, hence techniques for FDA have to be extended to include this information. Similarly to its scalar counterpart, functional time series appear in numerous real-world problems, such as electricity load forecasting (Antoniadis, Brossat, et al., 2016; Vilar, Aneiros, et al., 2018), the analysis of yield curves (Sen and Klüppelberg, 2019; Horváth, Liu, et al., 2020) and the prediction of daily pollution curves (Aue, Norinho, et al., 2015; Álvarez-Liébana and Ruiz-Medina, 2019), among others. As such, the analysis and modelization of such high-dimensional data is one of the most relevant topics in applied functional data analysis, and a vast portion of the functional data analysis literature has been devoted to this area.

The purpose of the following sections is to present the basic concepts in FDA that will be used throughout this document and the current state of the art in forecasting models for functional time series. The outline of this chapter is as follows: Section 2.2 introduces the mathematical background for operating with functional data, summarizing the most commonly used techniques when working with functional datasets. Section 2.3 is devoted to functional time series, introducing this particular case of functional data and performing a literature review of the the forecasting models proposed in the literature for these high-dimensional data. Finally, Section 2.4 details the background for functional probabilistic models, analyzing the definition of a density function for functional data and reviewing the models proposed in the literature to obtain probabilistic forecasts of a functional dataset.

2.2. Tools for functional data analysis

FDA is concerned with obtaining information from a sample of curves. Due to the infinite nature of the curves, new statistical tools have been developed to facilitate the analysis of the data. The main purpose of this section is to introduce the mathematical framework used in FDA, along with providing a description of the essential concepts and tools commonly used when analyzing a functional dataset, such as representation of functional data using basis expansions, summary statistics for functional data and a brief introduction to linear operators in Hilbert spaces. This section will follow the monographs Ramsay and Silverman (2005), Horváth and Kokoszka (2012) and Kokoszka and Reimherr (2017).

2.2.1. Basis expansion

A functional stochastic process $\{\mathcal{X}_i\}_{i=1}^{N}$ has been defined as a sequence of random functions ordered in time. However, in practice the value of each curve X_i is measured on a discrete grid $v_1, v_2, \ldots, v_{d_i}$, obtaining a (possibly high-dimensional) vector observation:

$$X_i(v_1), X_i(v_2), \ldots X_i(v_{d_i}). \tag{2.2}$$

Although it might be tempting to analyze these data as multivariate, there are several disadvantages to this approach: firstly, the curves may not have been observed on the same discretization points, so an initial pre-processing stage will be needed before analyzing the data. Secondly, the high-dimensionality of the data could lead to more complex multivariate models with a high number of parameters to estimate. As such, a curve-based treatment of the data is preferred, as it will be able to take advantage of the underlying structure of the functional data and it will reduce the number of parameters to estimate in a functional model.

Several techniques have been proposed in order to convert these observed discrete values to a functional observation computable for any desired argument value $X_i(v)$. While interpolation and kernel smoothing techniques can be applied to estimate the underlying curves, the most widely used tool for representing functional data are basis functions procedures (Ramsay and Silverman, 2005). By assuming that the functional observations X_i belong to the functional space $L^2(V)$, it is possible to select a basis $\{\phi_n\}_{n=1}^{\infty}$ of that space, obtaining the following representation of X_i as a linear combination of basis functions:

$$X_i(v) = \sum_{n=1}^{\infty} c_{n,i}\phi_n(v), \quad i = 1, \ldots, N, \tag{2.3}$$

where $c_{n,i}$ are the coefficients of the basis representation. Equation (2.3) shows that an infinite number of basis functions are required to represent the functional data with zero error. In practice, however, it is not computationally feasible to work with an infinite number of basis functions, so Equation 2.3 must be truncated in such a way that the reduced basis set is able to approximate the functions accurately:

$$X_i(v) \approx \sum_{n=1}^{M} c_{n,i}\phi_n(v), \quad i = 1, \ldots, N. \tag{2.4}$$

Therefore, the infinite-dimensional functional observation X_i is represented by the finite-dimensional coefficient vector $\boldsymbol{c}_i = (c_{1,i}, \ldots, c_{M,i})^\top$. These coefficients can be estimated using a wide variety of techniques, ranging from least squares estimation to smoothing techniques.

When analyzing functional data, the selection of the basis is of utmost importance, as the smoothed data will share some shape properties with the chosen basis functions. The nature of the data often motivates the selection of these functions: if the shape of the original data exhibits some kind of periodicity, the use of the Fourier basis is often recommended. Fourier series generally yield expansions which are uniformly smooth, and are especially useful for representing functions where there are no strong local features and the curvature tends to be of the same order everywhere. This basis is commonly used in real world applications such as fetal heart rate monitoring in medicine (Ratcliffe et al., 2002) and when analyzing environmental NOx and ozone cycles (Gao, 2007; Gao and Niemeier, 2008), among others. On the other hand, a popular system to represent non-periodic curves is the B-spline basis. The usefulness of splines comes from combining the fast computation properties of polynomials with greater flexibility in the representations of functions. If there is no prior information about the data generating process that is being analyzed, the B-spline basis is a suitable family to begin any FDA analysis. Some examples of this technique are Hyndman and Ullah (2007), where the authors use a spline basis to represent demographic curves; and Newell et al. (2006), where the authors analyze the structure of lactate curves obtained after using spline smoothing to the collected data. Nevertheless, the most common basis used in practical applications is the functional principal component basis, which will be detailed in Section 2.2.5. One of the main advantages of this representation is its ability to explain a great percentage of the original variance of the functional dataset using only a reduced number of basis, obtaining a low-dimensional representation of the curves. This technique has been widely used in medical research, as shown in Ullah and Finch (2013). However, in some cases the exact shape of the curves being represented is known and there is no need for smoothing the rough data observed, as it is the case when analyzing supply and demands curves from the electricity market (Aneiros, Vilar,

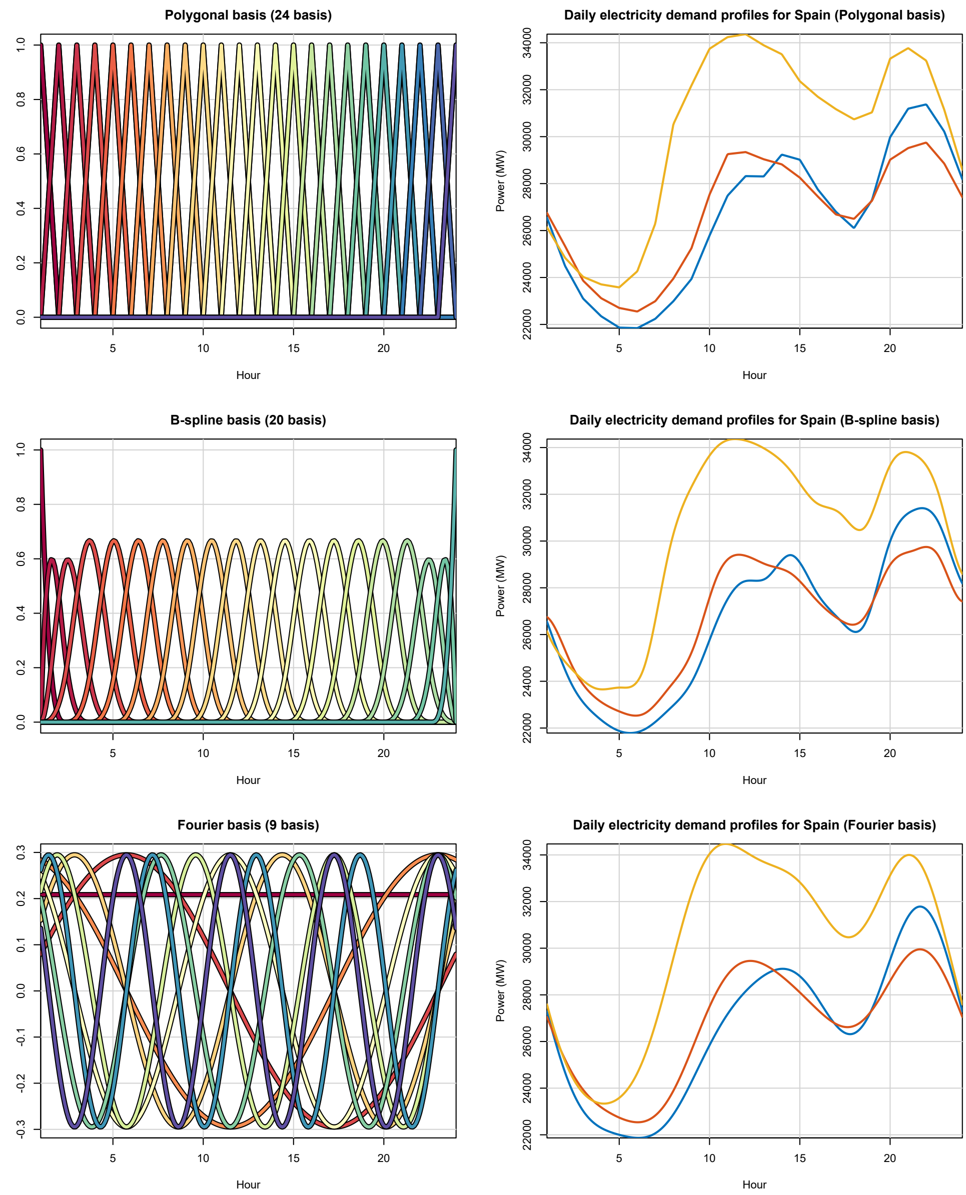

Figure 2.1. Functional data obtained using different basis expansion methods. Each curve is a daily load profile from the Spanish electricity market.

Cao, et al., 2013; Shah and Lisi, 2020). In this case, the polygonal basis can be used to represent the noiseless data.

Although it may seem that functional data analysis can be reduced to multivariate analysis, this is far from true. There is a trade-off between the accuracy of the functional representation and the number M of basis functions chosen, hence in practice selecting two values of M can result in two different analysis that yield non-identical results. If $M = d$ (the number of basis is equal to the number of discretization points of the curves) interpolation of the curves is achieved. Hence, the number M of basis functions should be seen as a measure of the degree to which the data X_i are smoothed, and chosen accordingly to the characteristics of the data. In addition, the representation of the curves is strongly related to the chosen basis: the shape of the functional representation relies on the characteristics of the basis.

Figure 2.1 illustrates the differences between representing the same set of discrete functional observations using different sets of basis functions. Daily load profiles from hourly data obtained for the Spanish electricity market have been obtained for three different days of February, 2019 employing different basis expansion methods. Each functional observation can be used to analyze the real-time electricity consumption in the market. According to the functional basis used to perform this transformation, the resulting curves will exhibit different properties. The top panel shows the load profiles obtained after representing the data using the polygonal basis. As can be seen, this is equivalent to performing linear interpolation between the discrete observations, and as such the resulting functional observations are rough. As the temporal evolution of the electricity load does not follow this linear pattern, it may be concluded that this basis is not adequate for representing this data. As the data does not show any kind of periodic behaviour, a more appropriate basis could be the B-spline basis, illustrated in the middle panel. This basis representation is able to smooth the discrete observations, obtaining a much more accurate image of the temporal evolution of the demand. This representation is compared with the one obtained after using a Fourier basis, as shown in the bottom panel. Due to the nature of the Fourier basis, a high level of smoothing is applied to the original data, resulting in a less accurate functional dataset as the one obtained with the B-spline basis. Load profiles obtained with the Fourier basis are over-smoothed, as can be seen in both the blue and yellow load profiles, which are not able to capture changes in the level of the demand between hours 12-14 and 16-18 with the same precision as the profiles obtained by using the B-spline basis. This example can be used to illustrate the importance of selecting an adequate set of basis functions that it is able to accurately represent the data, as the conclusions drawn by any subsequent analysis rely heavily on this pre-processing step.

These basis expansion methods serve a twofold purpose. On the one hand, they allow a computational feasible treatment of the high-dimensional data, providing a framework for working with the functional data at different evaluation points that may not be equidistant. On the other hand, these methods can be used for smoothing noisy functional data, including a penalty term in the basis expansion procedure that accounts for the roughness present in the original curves. This makes basis representation a crucial step when analyzing and modelling functional data, as choosing a suitable basis for the curves will enable the practitioner to obtain insight about the functional dataset being explored.

2.2.2. Norms and distances for functional data

Once functional data have been obtained from the discrete measurements of the curves, it is necessary to establish the mathematical framework required to analyze the properties of the functional dataset. Throughout this thesis, it will be assumed that the functional observations belong to the Hilbert space $L^2(V)$ of real valued square integrable functions defined on the closed interval V. This space is sufficient to handle most of the functional processes that are found in real-world applications, hence the vast majority of literature regarding functional data analysis makes similar assumptions.

A real-valued function of real variable X belongs to the space $L^2(V)$ if it satisfies $\int_V X^2(u)du < \infty$. From now on it will be assumed that an integral sign without the limits of integration is meant to denote the integral over the whole domain V: $\int X(u)du = \int_V X(u)du$. In addition, as long as there is no risk of confusion, the interval V will be omitted when referring to the space L^2. The space L^2 is a Hilbert space with the inner product

$$\langle X, Y \rangle = \int X(u)Y(u)du. \tag{2.5}$$

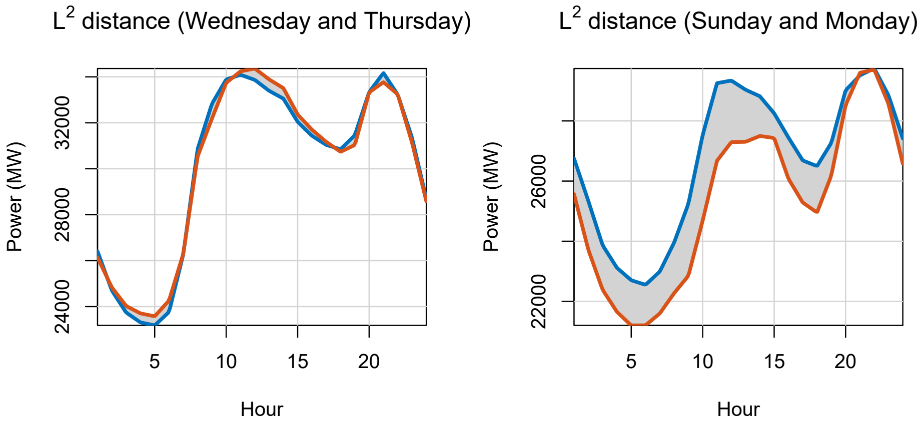

Figure 2.2. Visual representation of the L^2 distance between two pairs of electricity load curves: Wednesday-Thursday and Sunday-Monday.

The inner product induces a notion of magnitude over the functional observations, as the inner product generates the inner product norm:

$$\|X\| = \sqrt{\langle X, X\rangle} = \sqrt{\int X(u)^2 du}, \tag{2.6}$$

which also induces a distance function for the elements of the Hilbert space L^2:

$$\mathrm{d}(X,Y) = \|X - Y\| = \sqrt{\langle X-Y, X-Y\rangle} = \sqrt{\int (X(u) - Y(u))^2 du}. \tag{2.7}$$

Intuitively, a distance function provides a notion of closeness or similarity between two elements: if the distance between two objects is high, it can be concluded that these two objects exhibit different properties. This can be observed in Figure 2.2, where the L^2 distance is used to highlight the differences between two pairs of load curves. When comparing two curves that exhibit different behaviour (different mean level and shape), the L^2 distance is very high, effectively detecting the dissimilarity that exists between those functions. On the other hand, when two curves are near in the L^2 sense it can be concluded that both functions exhibit similar properties, as can be seen when comparing the Wednesday and Thursday curves. Hence, distance functions are commonly used in classification and clustering of functional datasets.

However, it should be noted that the L^2 distance defined in Equation 2.7 is not the only distance function that can be defined over L^2: a distance function $\mathrm{d}_*(.,.)$ on L^2 can be defined as any mapping

$$\begin{aligned} \mathrm{d}_* : L^2 \times L^2 &\longrightarrow \mathbb{R}^+ \\ (X,Y) &\longmapsto \mathrm{d}_*(X,Y) \end{aligned} \tag{2.8}$$

that satisfies the following properties:

P.1 $\mathrm{d}_*(X,Y) = \mathrm{d}_*(Y,X)$ for all $X, Y \in L^2$.

P.2 $\mathrm{d}_*(X,Y) = 0$ if and only if $X = Y$.

P.3 $\mathrm{d}_*(X,Y) \leq \mathrm{d}_*(X,Z) + \mathrm{d}_*(Z,Y)$ for all $X, Y, Z \in L^2$.

As such, other functional distances found in the literature are:

- **Hausdorff distance**: defined as the maximum distance from one point of the curve to the closest point of another curve.

$$\mathrm{d}_H(X,Y) = \max\{\sup_{x\in G(X)} \inf_{y\in G(Y)} \mathrm{d}_2(x,y), \sup_{y\in G(Y)} \inf_{x\in G(X)} \mathrm{d}_2(x,y)\}, \tag{2.9}$$

  where $\mathrm{d}_2(.,.)$ stands for the euclidean distance in $\mathbb{R}$ and $G(X) = \{(v, X(v)) \in \mathbb{R}^2\}$ is the graph of the curve X.

- **Weighted L^2 distance**: given a weight function $w(v)$ such that $\int w(u)du < \infty$, the weighted L^2 distance is defined as

$$\mathrm{d}_w(X,Y) = \sqrt{\left(\frac{1}{\int w(u)du}\int (X(u) - Y(u))^2 w(u)du\right)} \tag{2.10}$$

As can be seen, by defining new distance functions, different properties of the curves can be highlighted, enabling a different view of the relations between the curves of any given functional dataset. However, condition (P.2) is very restrictive when defining new distance functions, so most authors choose to remove this condition from the definition of distance and work with semi-metrics (Ferraty and Vieu, 2006; Febrero-Bande and Oviedo de la Fuente, 2012). The most common family of semi-metrics found in the literature are:

- **Derivative-based**: given two smooth functions, a new family of semi-metrics can be obtained by measuring the distance between the derivatives of the curves. More precisely, given two observed curves X and Y, the following semi-metric can be defined:

$$\mathrm{d}_m^S(X,Y) = \sqrt{\left(\int \left(X^{(m)}(u) - Y^{(m)}(u)\right)^2 du\right)}, \tag{2.11}$$

  where $X^{(m)}$ denotes the m-th derivative of X. This metric was used in Ferraty, Keilegom, et al. (2009) when analyzing a set of spectrometric curves from a food industry quality control problem.

- **Projection-based**: in many applications, curves are represented as a linear combination of basis functions $\{\phi_n\}_{n=1}^M$, effectively reducing the dimension of each functional observation X, as explained on Section 2.2.1. This decomposition can be used to measure distances between functional observations:

$$\mathrm{d}_M^P(X,Y) = \sqrt{\left(\sum_{i=1}^M \left(\int \{X(u) - Y(u)\}\phi_i(u)du\right)^2\right)}. \tag{2.12}$$

  As pointed out in Ferraty, Van Keilegom, et al. (2012), this semi-metric is especially well adapted when the functional observations are rough, as it can smooth out the data and work with the important features of the functional dataset.

- **Horizontal shift**: this measure provides a notion of proximity between two curves taking into account horizontal shift effects:

$$\mathrm{d}_{shift}\left(X, Y\right) = \min_{h \in \{-mh, mh\}} \{\mathrm{d}_2\left(X(v), Y(v+h)\right)\}, \tag{2.13}$$

  where mh stands for the maximum horizontal shift allowed and $\mathrm{d}_2\left(.,.\right)$ is the euclidean distance in $\mathbb{R}$. As this metric is invariant by translation, it is useful when working with functional data that exhibit some kind of horizontal displacement, such as the altometric curves collected by the satellite Topex/Poseidon (Dabo-Niang et al., 2004).

In addition, it can be proven that the sum of semi-metrics is itself a semi-metric, providing a tool for constructing new metric functions.

The use of semi-metrics plays a crucial role in nonparametric functional data analysis, as choosing a suitable metric function will highlight relevant features of the functional data. Nonparametric techniques have been successfully applied in many problems, such as functional classification (Ferraty and Vieu, 2003; Chang et al., 2014), clustering of functional data (Dabo-Niang et al., 2007; Delaigle, Hall, and Bathia, 2012), functional regression (Ferraty, Van Keilegom, et al., 2012; Raña, Aneiros, et al., 2016) and forecasting (Aneiros, Cao, et al., 2011; Vilar, Cao, et al., 2012), illustrating the flexible capabilities of nonparametric models.

The inner product defined in L^2 also defines the angle θ between two functions X and Y as

$$\cos(\theta) = \frac{\langle X, Y\rangle}{\|X\|\|Y\|}, \tag{2.14}$$

which makes the geometric structure of the space L^2 similar to that of the usual finite dimensional Euclidean space, at least at an intuitive level. Following the previous definition, the concept of orthogonality can be defined as follows: two functions X and Y are orthogonal (or perpendicular) if $\langle X, Y\rangle = 0$. Similarly to the usual finite dimensional Euclidean space, orthogonal functions will play a major role when working with basis expansions. When basis functions were introduced in Equation 2.3, it was not required that the elements of the basis were orthogonal. If the basis $\{\phi_n\}_{n=1}^{\infty}$ satisfies $\langle \phi_i, \phi_j\rangle = 0$ whenever $i \neq j$ and $\|\phi_i\| = 1$, it is said that $\{\phi_n\}_{n=1}^{\infty}$ is an orthonormal basis. Working with orthonormal basis simplifies Equation 2.3, as it can be seen that it becomes

$$X(v) = \sum_{n=1}^{\infty} c_n \phi_n(v) = \sum_{n=1}^{\infty} \langle X, \phi_n\rangle \phi_n(v), \tag{2.15}$$

where the linearity of the inner product can be used to obtain the value of the coefficients $c_n = \langle X, \phi_n\rangle$. As $\|\phi_i\| = 1$, the value $\langle X, \phi_i\rangle$ can be interpreted as the projection of the curve X over the direction specified by ϕ_i. In addition, for an orthonormal basis, Parseval's equality holds:

$$\int X^2(u)du = \|X\|^2 = \sum_{i=1}^{\infty} \langle X, \phi_i\rangle^2 = \sum_{i=1}^{\infty} \left(\int X(u)\phi_i(u)du\right)^2. \tag{2.16}$$

2.2.3. Summary statistics

In many practical applications, it is assumed that each curve of the observed sample $X_1, X_2, \ldots, X_N$ is a realization of a random function $\mathcal{X}$. As such, the main statistical properties of $\mathcal{X}$ must be analyzed to obtain insight about the observed data. The simplest summary statistics are the functional mean, variance and correlation.

The mean function $\mu_{\mathcal{X}}$ of the random function $\mathcal{X}$ is the function such that $\mu_{\mathcal{X}}(v) = E\left[\mathcal{X}(v)\right]$, where $E\left[.\right]$ denotes the expectation. The mean function can be estimated from the observed data by the sample mean function, defined as

$$\widehat{\mu}_{\mathcal{X}}(v) = N^{-1} \sum_{t=1}^{T} X_t(v), \tag{2.17}$$

which is the point-wise average of the observed curves.

Similarly to the finite-dimensional case, the variance function measures the deviations from the mean, and can be defined as the function such that $\text{var}_{\mathcal{X}}(v) = E\left[\left(\mathcal{X}(v) - \mu_{\mathcal{X}}(v)\right)^2\right]$. Its sample counterpart is defined as

$$\widehat{\text{var}}_{\mathcal{X}}(v) = (N-1)^{-1} \sum_{t=1}^{T} \left(X_t(v) - \widehat{\mu}_{\mathcal{X}}(v)\right)^2. \tag{2.18}$$

As in the scalar case, the square root of the variance function defines the standard deviation function.

Up to this point, the summary statistics presented are pointwise, meaning that the value of the statistic at a given discretization point u only takes into account the values of the functional observations for that same discretization point u. As such, the point-wise variance function can be used to measure the typical variability of the functional dataset for each discretization value. However, these values do not account for how the values of the curves at point u relate to those values at point v. In order to measure this, the covariance function of the functional variable $\mathcal{X}$, defined as $\text{c}_{\mathcal{X}}(v,u) = E\left[\left(\mathcal{X}(v) - \widehat{\mu}_{\mathcal{X}}(v)\right)\left(\mathcal{X}(u) - \widehat{\mu}_{\mathcal{X}}(u)\right)\right]$, is introduced. The values of this function quantify the joint deviations from the mean between two different discretization points v and u, effectively measuring the joint variability of the functional random variable $\mathcal{X}$ at those points. This function is estimated by the sample covariance function $\widehat{\text{c}}_{\mathcal{X}}$, defined as

$$\widehat{\text{c}}_{\mathcal{X}}(v,u) = (N-1)^{-1} \sum_{t=1}^{T} \left(X_t(v) - \widehat{\mu}_{\mathcal{X}}(v)\right)\left(X_t(u) - \widehat{\mu}_{\mathcal{X}}(u)\right). \tag{2.19}$$

This function can be seen as a counterpart for the covariance matrix in the multidimensional setting: large values of $\widehat{\text{c}}_{\mathcal{X}}(v,u)$ indicate that $X(v)$ and $X(u)$ tend to be simultaneously below or above the mean values at these two discretization points.

Similarly to the multidimensional case, the functional correlation for a functional variable $\mathcal{X}$ can be defined by normalizing the covariance function

$$\widehat{\rho}_{\mathcal{X}}(v,u) = \frac{\widehat{\text{c}}_{\mathcal{X}}(v,u)}{\sqrt{\widehat{\text{var}}_{\mathcal{X}}(v)\widehat{\text{var}}_{\mathcal{X}}(u)}}. \tag{2.20}$$

This function takes values between -1 and 1, allowing for a better interpretation of the covariance structure present in the functional data. In situations where there is no risk of confusion, the sub-index indicating the functional random variable which these summary

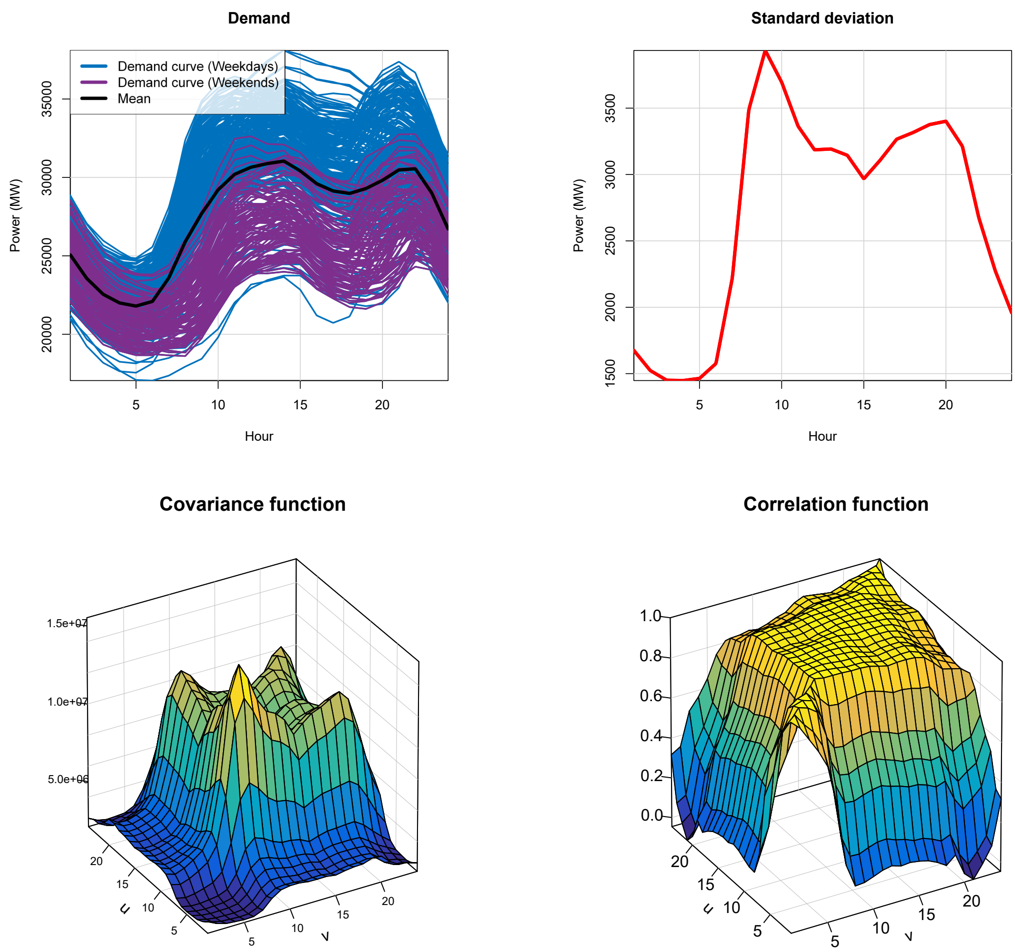

Figure 2.3. Point-wise summary statistics for the daily electricity load profiles of Spain. The top figures contain the sample mean and standard deviation functions, whereas the bottom images contain the sample covariance and correlation functions of the functional data.

statistics refer to will be omitted (e.g. $\widehat{\rho} = \widehat{\rho}_{\mathcal{X}}$), in order to ease the notation and provide a more comprehensive exposition of the techniques presented in this dissertation.

Figure 2.3 illustrates these concepts on the electricity load curves of the Spanish market in 2019. The point-wise sample mean and standard deviation curves are included, showing that the variation of the load is lower between hours 0 and 6, and its point of highest variation is found at hour 9. The covariance sample function contains information about the linear dependence structure across different discretization points of the curves, although it can be difficult to interpret due to the scale of the function. Hence, the correlation sample function is included, which is bounded between -1 and 1. This function provides a more straight-forward interpretation of the linear dependence structure exhibited by the data: the electricity load between hours 13 and 18 shows a high value of correlation between those same hours. A similar behaviour, albeit slightly lower, can be observed between the load of hours 7 to 9 and 21 to 23: the correlation between these hours has a value higher than 0.8, showing a strong linear dependence between these sections of the load curves.

To this point, all the summary statistics presented in this section only involve one random function. However, the covariance and correlation functions can be generalized to measure the

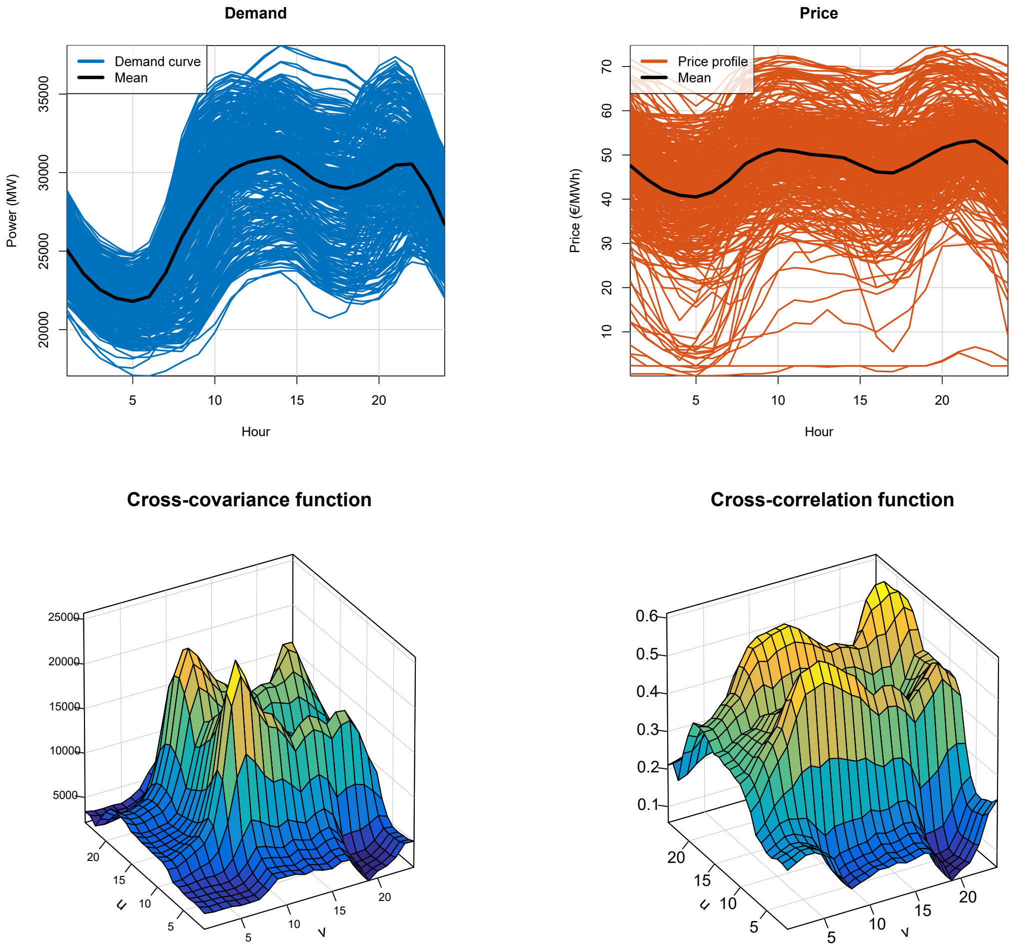

Figure 2.4. Top: electricity load and price profiles for the Spanish electricity market. Bottom: sample cross-covariance and sample cross-correlation functions of the two functional datasets.

variability between two random functions $\mathcal{X}$ and $\mathcal{Y}$. The cross-covariance function between $\mathcal{X}$ and $\mathcal{Y}$ is defined as $\mathrm{c}_{\mathcal{X},\mathcal{Y}}(v,u) = E\left[(\mathcal{X}(v) - \mu_{\mathcal{X}}(v))(\mathcal{Y}(u) - \mu_{\mathcal{Y}}(u))\right]$, and can be estimated via its sample counterpart

$$\widehat{\mathrm{c}}_{\mathcal{X},\mathcal{Y}}(v,u) = (N-1)^{-1} \sum_{t=1}^{T} \left(X_t(v) - \widehat{\mu}_{\mathcal{X}}(v)\right)\left(Y_t(u) - \widehat{\mu}_{\mathcal{Y}}(u)\right), \tag{2.21}$$

where (X_t, Y_t) are pairs of functional observations. This function can be used to quantify the linear dependence between those sets of functional data. The cross-correlation function provides a scale-free version of the cross-covariance function, as it is bounded between -1 and 1:

$$\widehat{\rho}_{\mathcal{X},\mathcal{Y}}(v,u) = \frac{\widehat{\mathrm{c}}_{\mathcal{X},\mathcal{Y}}(v,u)}{\sqrt{\widehat{\mathrm{var}}_{\mathcal{X}}(v)\widehat{\mathrm{var}}_{\mathcal{Y}}(u)}}. \tag{2.22}$$

The usefulness of this tool is highlighted in Figure 2.4, where both the cross-covariance and cross-correlation surfaces are obtained for the series of electricity load curves and electricity price profiles for the Spanish electricity spot market. The correlation between the two functional datasets for hours 7 to 10 and 19 to 21 is high, which can be interpreted as the price at those hours being directly affected by the demand at the same hours. In addition, there seems to be a high correlation between the electricity price at hours 17 to 20 and the electricity load between hours 7 to 10, meaning that the prices in that range of hours will tend to be directly related to the load values of the early hours. This information can be used by a functional regression model to obtain robust forecasts of the price profiles using the whole load curves as exogenous variables.

2.2.4. Operators in Hilbert spaces

Most functional data analysis techniques require the use of tools that are able to operate with the observed curves. Linear operators offer a mathematical tool to perform transformations on functional datasets, and are one of the main elements of the theory of Hilbert spaces. As such, a vast portion of the literature on functional data models has been devoted to the study of these operators. This section will introduce the notion of functional operators, with a special focus on the most common classes of operators found in practice.

A linear operator defined on the Hilbert space L^2 is any function $\Psi : L^2 \to L^2$ such that, for any scalars $a, b \in \mathbb{R}$ and any $X, Y \in L^2$, the following relation holds:

$$\Psi\left(aX + bY\right) = a\Psi\left(X\right) + b\Psi\left(Y\right). \tag{2.23}$$

An operator Ψ is said to be bounded if there exists a constant C such that for any $X \in L^2$, $\|\Psi(X)\| \leq C\|X\|$. It can be seen that an operator is bounded if and only if it is continuous, hence they are usually referred to as continuous operators.

A bounded operator Ψ is said to be compact if there exist two orthonormal basis $\{\phi_n\}_{n=1}^{\infty}$ and $\{f_n\}_{n=1}^{\infty}$ and a sequence of real positive values $\{\lambda_n\}_{n=1}^{\infty}$ converging to zero such that

$$\Psi(X)(v) = \sum_{j=1}^{\infty} \lambda_j \langle X, \phi_j \rangle f_j, \quad \forall\, X \in L^2. \tag{2.24}$$

This is called the singular value decomposition of the operator. One of the properties of compact operators is that they map every bounded set into a compact set, providing a stronger notion of continuity. If, in addition, condition $\sum_{j=1}^{\infty} \lambda_j^2 < \infty$ holds, the operator Ψ is said to be a Hilbert–Schmidt operator.

When modeling functional data, an important class of operators are integral operators, defined as

$$\Psi(X)(v) = \int \psi(u,v)X(u)du, \quad \forall\, X \in L^2. \tag{2.25}$$

The bivariate function $\psi(u, v)$ is called the kernel of the operator Ψ. It can be proven that an integral operator is a Hilbert–Schmidt operator if

$$\iint \psi^2(u,v)dudv < \infty. \tag{2.26}$$

If in addition the kernel of the Hilbert–Schmidt operator satisfy that $\psi(u, v) = \psi(v, u)$ and $\iint \psi(u, v)X(u)X(v)dudv \geq 0$ for all $X \in L^2$, the operator Ψ becomes a symmetric and positive-definite operator, admitting the following spectral decomposition:

$$\Psi(X) = \sum_{j=1}^{\infty} \lambda_j \langle X, \phi_j \rangle \phi_j, \quad \forall X \in L^2, \tag{2.27}$$

where $\{\lambda_n\}_{n=1}^{\infty}$ are the eigenvalues of the operator Ψ and $\{\phi_n\}_{n=1}^{\infty}$ are the corresponding orthonormal eigenfunctions, satisfying $\Psi(\phi_j) = \lambda_j \phi_j$. The spectral decomposition allows the following representation of the kernel function $\psi(u, v)$:

$$\psi(u, v) = \sum_{j=1}^{\infty} \lambda_j \phi_j(u) \phi_j(v). \tag{2.28}$$

This section has summarized the theory about functional operators that will be used throughout this document. Integral operators play a key role when defining parametric functional models, as will be seen later in this chapter. In addition, the spectral decomposition of the covariance operator is the basis for functional principal components analysis, which will be introduced in the next section. For an in-depth look at the theory of operators in Hilbert spaces, the reader is referred to Riesz and Sz.-Nagy (1990).

2.2.5. Functional principal components

Principal component analysis (PCA) is one of the most popular techniques for analyzing multidimensional data (Jolliffe, 2010). This dimensionality reduction technique is able to obtain a low-dimensional representation of the original data in such a way that most of the variation of the original data is captured. Due to its usefulness this technique has been extended to the functional data framework (Besse and Ramsay, 1986; Rice and Silverman, 1991), becoming the most prevalent tool when analyzing a functional dataset. Functional Principal Component Analysis (FPCA) can be used to reduce a set of infinite-dimensional functional data to a set of multivariate scores that account for most of the variance of the data. As a consequence, classical multidimensional techniques can be applied to the resulting scores, allowing to use known tools to model the observed functional data.

Similarly to its multidimensional counterpart, FPCA is based on the spectral decomposition of the covariance operator of the observed data. Given a random function $\mathcal{X}$ such that $E[\mathcal{X}] = 0$ and $E\left[\|\mathcal{X}\|^2\right] < \infty$, the associated covariance operator is defined as

$$\mathrm{C}_{\mathcal{X}}(Y) = E\left[\langle \mathcal{X}, Y \rangle \mathcal{X}\right], \quad \forall Y \in L^2. \tag{2.29}$$

As the functional data are assumed to be elements of L^2, the covariance operator can be represented as an integral operator whose kernel is the covariance function $\mathrm{c}_{\mathcal{X}}(u, v)$:

$$\mathrm{C}_{\mathcal{X}}(Y)(v) = \int \mathrm{c}_{\mathcal{X}}(u, v) Y(u) du, \quad \forall Y \in L^2. \tag{2.30}$$

It can be seen (Conway, 1990) that covariance operators are symmetric and non-negative Hilbert–Schmidt operators, hence Equation 2.28 holds and the covariance kernel $\mathrm{c}_{\mathcal{X}}(u, v)$ can be expressed as

$$\mathrm{c}_{\mathcal{X}}(u, v) = \sum_{j=1}^{\infty} \lambda_j \phi_j(u) \phi_j(v), \tag{2.31}$$

where the eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots$ represent the variability explained by the associated eigenfunctions $\{\phi_n\}_{n=1}^{\infty}$, called the (functional) principal components. The empirical counterparts of the eigenvalues and eigenfunctions of the sample covariance operator are denoted as $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \ldots$ and $\{\widehat{\phi_n}\}_{n=1}^{\infty}$, respectively.

As a consequence, each functional observation $\{X_i\}_{i=1}^{N}$ can be represented as a linear combination of its principal components. This expansion is known as the Karhunen-Loéve representation:

$$X_i = \widehat{\mu}_{\mathcal{X}} + \sum_{j=1}^{\infty} \widehat{\xi}_{i,j} \widehat{\phi}_j, \tag{2.32}$$

where $\widehat{\xi}_{i,j} = \int (X_i(u) - \widehat{\mu}_{\mathcal{X}}(u)) \widehat{\phi}_j(u) du$ and $\{\widehat{\phi}_n\}_{n=1}^{\infty}$ are the eigenfunctions of the sample covariance operator, which form a basis in L^2. The values $\widehat{\xi}_{i,j}$ are the coordinates of the functional observations when projected in the direction of $\widehat{\phi}_j$ and are usually called the scores of the curves. Methods for estimating these values can be found in Ramsay and Silverman (2005).

As already mentioned, one of the main purposes of FPCA is to reduce the dimension of the functional data to a finite number of M components. Once the value of M is selected, Equation 2.32 can be truncated in such a way that the selected principal components are able to represent the functions accurately:

$$X_i \approx \widehat{\mu}_{\mathcal{X}} + \sum_{j=1}^{M} \widehat{\xi}_{i,j} \widehat{\phi}_j. \tag{2.33}$$

Figure 2.5 shows the evolution of the scores of the first three FPC of the electricity load dataset, together with the estimated mean $\widehat{\mu}_{\mathcal{X}}$ and eigenfunctions $\widehat{\phi}_j$ of the FPC decomposition. As can be seen, the first principal component is always positive and retains a characteristic shape of the electricity demand curves, hence it can be interpreted as a level function. By analyzing the weekly seasonal pattern exhibited by its associated scores the difference between weekdays and weekends can be seen, due to weekends having a lower demand than the rest of days of the week. The remaining FPC can be seen as shape curves, with the second eigenfunction modeling the difference between valley hours and peak hours. The scores associated to the second FPC have negative values in the winter months, thus this difference is lower than in the summer months, where the scores are positive and there is a greater distance between valley and peak hours. As such, these curves modulate the shape of the load profile, and are associated with scores with lower values than the first FPC.

In most applications, it is crucial to determine a value of M such that the actual curves can be accurately replaced by the approximation shown in Equation 2.33. Usually this decision entails a compromise between the number of components and the total percentage of variance that the selected eigenfunctions are able to explain. This approach is commonly known as the Cumulative Percentage of total Variance (CPV) method (Kokoszka and Reimherr, 2017), and uses the magnitude of the estimated covariance eigenfunctions $\widehat{\lambda}_j$ to choose an appropriate number of components. The cumulative percentage of total variance explained by the first k sample eigenfunctions in Equation 2.33 is

$$\text{CPV}(k) = \frac{\sum_{j=1}^{k} \widehat{\lambda}_j}{\sum_{j=1}^{M} \widehat{\lambda}_j}. \tag{2.34}$$

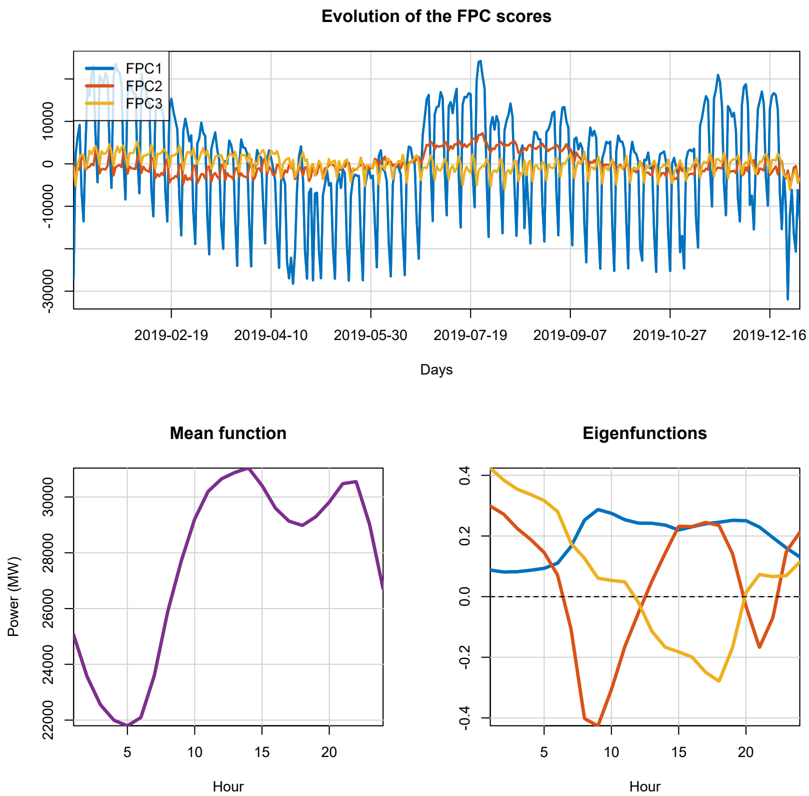

Figure 2.5. Top: evolution of the scores of the first 3 FPC of the daily electricity load profiles of Spain. Bottom: sample mean function and estimated eigenfunctions of the FPC decomposition.

As such, the number of components k is chosen as the first value for which CPV(k) exceeds a desired level, typically between 85% and 95%. Figure 2.6 illustrates this concept: the left figure contains a visual representation of the total variance explained by each FPC. As can be seen the first 3 components already explain more than 95% of the variance of the original dataset. The right panel shows the difference between a real load curve and the reconstructed curves using 2,4,6 and 8 FPC. The representation of the curves improves with the inclusion of additional FPC to the decomposition, but even with more than 99% of the total variance explained by the first 4 FPC there are areas of the curves which are not well modeled by this approximation.

Similar dimension reduction of a functional dataset can be achieved by expanding the functional data into other function basis, such as the ones reviewed in Section 2.2.1. The main difference between the FPC basis and the other basis proposed in the literature is that among all basis expansions that use M elements, the FPC expansion explains most of the variance of the dataset in the L^2 sense. As such, the FPC basis is commonly used in practical applications such as exploratory data analysis (Meiring, 2007; Baladandayuthapani et al., 2008), regression (Maslova et al., 2010) and forecasting (Hyndman and Shang, 2009; Shang et al., 2019; Sen and Klüppelberg, 2019), among others.

2.2.6. Functional linear regression

Linear regression is perhaps the most widely used statistical model, and one of the most fundamental tools in statistics. When working with multidimensional data, the usual linear model is given by:

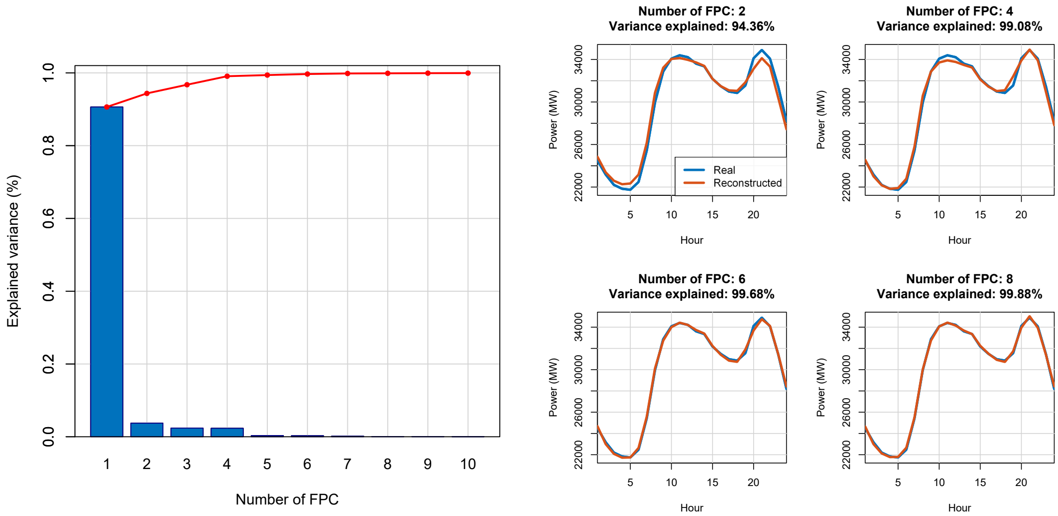

Figure 2.6. Left: Percentage of variance explained by each FPC. The red line indicates the Cumulative Percentage of total Variance explained by the first FPC. Right: Real and reconstructed electricity load profiles using 2,4,6 and 8 FPC.

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \varepsilon, \tag{2.35}$$

where y is usually called the response of the model, x_k are known scalar values known as covariates or regressors, β_k are called the coefficients of the regression model and ε are the innovations of the model, usually assumed to be i.i.d. errors. When generalizing this model to the functional data framework, some of these quantities will be curves, hence the model formulation and the coefficients of the model must be adapted accordingly.

Depending on whether the responses, the covariates, or both are functional values, functional linear models may be divided into three wide groups, described below. In all three cases, it will be assumed that the functional variables have been already centered.

- **Scalar covariates and functional response**: This model assumes that the functional variable Y is related to a set of scalar responses $\{x_k\}_{k=1}^{p}$. As such, the model can be formulated as

$$Y(v) = \sum_{k=1}^{p} \beta_k(v) x_k + \varepsilon(v), \tag{2.36}$$

where the parameters of the model are the functions β_k, which need to be estimated from the data. As the response of the model is functional, the residuals of the model are the zero mean functional i.i.d. values ε. The monograph Ramsay and Silverman (2005) describes the procedure to estimate the functional coefficients β_k using a least squares criterion, and illustrates the applicability of the model in a real world problem that involves predicting temperature curves from different climate zones. As this model allows to explain the behavior of a functional variable conditioned to several scalar covariates, it has been successfully applied in several medical applications, as illustrated in Yang et al. (2007) and Hutchinson et al. (2004).

- **Functional covariates and scalar response**: This model is used when the scalar variable y exhibits some dependence on a set of functional variables $\{X_k\}_{k=1}^p$, each defined in its own domain V_k. The model can be expressed as

$$y = \sum_{k=1}^{p} \int_{V_k} \beta_k(u) X_k(u) du + \varepsilon. \tag{2.37}$$

  Here, ε is a zero mean i.i.d. scalar error and the parameters of the model are the coefficient functions $\beta_k(v)$ which need to be estimated from the observed data. The classical approach for estimating these coefficient functions is to project the functional covariates X_k and its associated coefficients β_k in the space spanned by its first functional principal components in order to optimize its scores by least squares (Cardot et al., 1999). Due to its wide range of applications, the properties of this model have been studied thoroughly (Hall and Horowitz, 2007; Hilgert et al., 2013), and has been used extensively in several real-world applications (Ratcliffe et al., 2002; Wang, Su, et al., 2016).

- **Functional covariates and functional response**: In this model both the response variable Y and the covariates $\{X_k\}_{k=1}^p$ are functions, and the model assumes that there exists some kind of linear dependence between them. According to the type of relationship exhibited by the curves, two different function-on-function regression models have been proposed in the literature. The most general regression model is referred to as the Functional Linear Model (FLM):

$$Y(v) = \sum_{k=1}^{p} \int_{V_k} \beta_k(u, v) X_k(u) du + \varepsilon(v), \tag{2.38}$$

  where ε denotes the zero-mean i.i.d. residuals of the model, V_k is the domain of each covariate and the bivariate functions $\beta_k(u, v)$ are the coefficients of the model that need to be estimated. This model assumes that at any given point v, the value of the response $Y(v)$ can be modeled taking into account the entire trajectory of the covariates $\{X_k\}_{k=1}^p$. As such, this is a very flexible model that it is able to capture complex relations between functional data which may not share the same domain. However, such flexibility comes at a high price: in order to fit the FLM, the bivariate functions $\beta_k(u, v)$ must be estimated. In order to do this, a common approach (Ramsay and Dalzell, 1991) is to define some functional basis $\{\phi_l\}_{l=1}^L$, $\{\nu_m\}_{m=1}^M$ (usually, the first functional principal components) where the parameters $\beta_k(u, v)$ will be projected:

$$\beta_k(u, v) = \sum_{l=1}^{L} \sum_{m=1}^{M} b_{l,m}^{(k)} \phi_l(u) \nu_m(v), \quad k = 1, \dots, p. \tag{2.39}$$

  Therefore, the estimation of the bivariate coefficients $\beta_k(u, v)$ reduces to estimating their coordinates $b_{l,m}^{(k)}$ in this basis which reduce the forecasting error of the model.

  However, in many applications where the domain of the response and covariates is the same, the dependence exhibited between the curves at different discretization points $v \neq u$ is negligible. As such, given a discretization point v, the values of the response $Y(v)$ and the regressors $\{X_k(v)\}_{k=1}^p$ follow the classic regression model shown in Equation 2.35, where the regression coefficients may vary for different discretization points. The formulation of this second model, called concurrent functional linear model (Ramsay and Silverman, 2005), is as follows:

$$Y(v) = \sum_{k=1}^{p} \beta_k(v) X_k(v) + \varepsilon(v), \tag{2.40}$$

where ε denotes the zero-mean i.i.d. residuals of the model and $\beta_k(v)$ denotes the coefficient functions that have to be estimated. The reader is referred to Fan and Zhang (2008) for a review on estimation methods for this model, which involve kernel, polynomial and smoothing spline methods.

2.3. Functional time series

Up to this point, the functional data $\{X_i\}_{i=1}^N$ were assumed to be independent observations drawn from a functional random variable $\mathcal{X}$, hence the properties exhibited by each functional observation did not change over time. However, this is quite a strong assumption: usually, values X_t observed at a given time t exhibit some kind of dependence with previous observed values. This is the case with daily electricity prices profiles: the prices for a given day are usually influenced by the prices of previous days, so the assumption of independence between the curves must be dropped. In addition, Figure 2.3 shows that there exists a different pattern for the price profiles of weekends and holidays when compared with working days, so the distribution of the curves is not the same for every day. A better approach would be to model the functional observations $\{X_t\}_{t=1}^T$ as realizations of a functional stochastic process $\{\mathcal{X}_t\}_{t=1}^T$, defined as the collection of functional random variables ordered in time. Hence, the series $\{X_t\}_{t=1}^T$ is referred to as a functional time series (FTS).

Time series analysis is an active field of research which involves the study and modelization of these serially correlated data (Box et al., 2008). It has a wide range of applications in practically all fields of science and engineering, ranging from econometrics (Harvey, 1990) to environmental (Zhang, 2018) and medical sciences (Fried et al., 2017). In the last few decades there has been a growing research interest in generalizing these concepts to functional time series, as illustrated in Bosq (2000) and (Mas and Pumo, 2010). The ability to model dependent high-dimensional data such as electricity load curves (Antoniadis, Brossat, et al., 2016) and air pollution curves (Álvarez-Liébana and Ruiz-Medina, 2019) makes functional time series methods a key tool in a vast majority of real world applications. The monographs Horváth and Kokoszka (2012) and Kokoszka and Reimherr (2017) provide a panoramic view of the main theoretical and practical results for analyzing functional time series.

The purpose of this section is to review the fundamental concepts of functional time series analysis, focusing on forecasting models and estimation techniques. Firstly, Section 2.3.1 will introduce the notion of functional white noise, an important type of process which is the core of many identification tools for functional time series models. Then, a review of forecasting models for functional time series will be performed in Section 2.3.2, and finally the SARIMAHX model will be introduced in Section 2.3.3. As one of the objectives of this thesis is to propose an identification and diagnosis method for this model, details about the formulation and estimation of the SARIMAHX model have to be detailed.

2.3.1. Functional white noise

An important property of time series is stationarity, which refers to stochastic processes whose statistical properties, such as the mean and variance, do not vary over time. Most of the parametric forecasting models proposed in the literature assume that the time series is stationary, but as has already mentioned, most of the real world time series are not stationary. Hence, an initial step when modelling these non-stationary time series is to transform the data into a stationary process which then can be modelled using classical models found in the literature (Box et al., 2008). This highlights the importance of developing statistical tools that can be used to decide whether a time series is stationary.

The concept of stationarity can be generalized to functional time series (Hörmann, Kidziński, et al., 2015), where several authors have developed statistical procedures for testing the stationarity of functional time series (Horváth, Kokoszka, and Rice, 2014; Kokoszka and Young, 2016; Aue and Delft, 2020). Similarly to the scalar case, functional forecasting models assume that the functional time series is stationary, hence having methods for validating this assumption is crucial when modelling these high-dimensional series.

The most fundamental example of a stationary functional process is the functional white noise process (Bosq, 2000). A given functional stochastic process $\{\varepsilon_t(v)\}_{t=1}^{T}$ is a functional white noise process if the following assumptions are satisfied:

A1: $\mathrm{E}\left[\varepsilon_t\right] = 0$ for all $t = 1, \ldots, T$.

A2: The variables ε_i and ε_j are orthogonal for $i \neq j$, this is:

$$E\left[\langle \varepsilon_i, \mathcal{X}\rangle \langle \varepsilon_j, \mathcal{Y}\rangle\right] = 0, \quad \forall \mathcal{X}, \mathcal{Y} \in L^2. \tag{2.41}$$

If in addition to **A1** the process satisfies that

A3: The functional random variables ε_t are independent and identically distributed,

the process is called strong functional white noise (WN) process. Throughout this dissertation the term functional white noise will refer to a strong functional white noise process. Assumption **A3** implies that such a white noise series does not exhibit any degree of serial dependence, differentiating such a process from others that exhibit dependence between its terms. One example of functional white noise is the Brownian bridge process, defined by $\varepsilon_t(v) = W_t(v) - vW_t(1)$, where W_t denotes a Wiener process.

Tests for white noise in functional data are frequently used to check the adequacy of a fitted parametric model to a functional dataset, so they are of utmost importance in any practical forecasting application. For example, Gabrys, Horváth, et al. (2010) project the residual curves of a functional linear model on the space spanned by its first principal components to validate the white noise assumption, whereas in Horváth, Hušková, et al. (2013) the authors develop a Portmanteau test for testing the independence assumption of a functional time series. An in-depth review of these methods and how these can be applied to model identification and diagnosis is performed in Chapter 3.

2.3.2. Forecasting models for functional time series

When working with time series, it is often desired to model the temporal dependence exhibited by the data. As such, a vast majority of the literature regarding time series has been devoted to these types of forecasting models (the reader is to Chatfield (2001) for a survey on univariate forecasting methods). Due to the flexibility and generalization capabilities, these techniques have applications in diverse fields such as weather forecasting, production planning and load forecasting, among others.

The last two decades have seen seen a growing interest in forecasting models for functional time series, with authors developing novel techniques to model the complex dynamics often exhibited by these series. This section is devoted to reviewing the different modelling techniques found in the functional data literature, classifying them into different groups according to the statistical tools that they use: dimensionality reduction, parametric and nonparametric estimation methods.

2.3.2.1. Dimensionality reduction models

A common approach to forecasting functional time series is to project the functional observations into a suitable finite-dimensional space in order to apply multivariate models to estimate future values of the series $Y_1, Y_2, \ldots, Y_T$. In most applications the curves are projected into their first functional principal components, effectively reducing the dimension of the functional data and allowing the use of standard scalar time series models to forecast the future values of the scores $\widehat{\xi}_{t+1,j}$ of the decomposition. Then, using the Karhunen-Loéve representation Equation 2.33, estimated future curves can be reconstructed as

$$\widehat{Y}_{t+1}(v) = \sum_{j=1}^{K} \widehat{\xi}_{t+1,j} \widehat{\phi}_j(v). \tag{2.42}$$

As such, functional forecasting models based on dimensionality reduction can be classified according to the methods used to estimate the series of scores For example,Valderrama et al. (2002), Hyndman and Ullah (2007), Erbas et al. (2007), Hyndman and Shang (2009), and Wagner-Muns et al. (2018) project the curves into the space spanned by the first functional principal components of the functional time series and then fit univariate ARIMA models to each series of scores. An application of these methods can be found in Shang et al. (2019), where robust functional principal components are used to fit a projection-based functional regression model. The work by Shang (2012) summarizes different functional forecasting techniques that involve functional principal component decomposition methods.

Although the eigenfunctions of the FPC decomposition are uncorrelated, there may exist some kind of serial correlation between them. In order to account for these cross-correlations between scores, Aue, Norinho, et al. (2015), Klepsch et al. (2017), and Sen and Klüppelberg (2019) propose to fit a vector autoregressive model to the FPC scores, which is able to obtain a joint estimation of the multidimensional time series of the scores. In addition, the model proposed by Aue, Norinho, et al. (2015) is able to include functional covariates to its formulation, by projecting the exogenous functional observations into the space spanned by its first functional principal components. Once the dimensionality of the functional covariates has been reduced, the series of scores can be used in their multidimensional model.

2.3.2.2. Nonparametric models

Unlike the previous methodology, nonparametric and parametric models do not rely on explicitly reducing the dimension of the original functional dataset to fit a forecasting model to the data. These techniques assume that the functional time series follows some kind of relationship where functional linear operators are used to model this serial dependence. The reference model for forecasting functional time series is the autoregressive Hilbertian process of order 1 (Bosq, 2000), denoted by ARH(1), which is defined as

$$Y_t = \Psi(Y_{t-1}) + \varepsilon_t. \tag{2.43}$$

Here, Ψ is the autoregressive operator that needs to be estimated and ε_t denotes a functional white noise process. According to the techniques used to estimate these linear operators forecasting models for functional time series can be classified as parametric or nonparametric models.

The nonparametric framework does not assume a fixed structure for the linear operators, instead relying on a local weighting of the input variables to estimate the operators of the model (Ferraty and Vieu, 2006). The Nadaraya-Watson kernel estimator (Nadaraya, 1964; Watson, 1964) is often used to perform this local weighting of the variables, selecting the most similar historical values in the available dataset to predict future situations. As such, the ARH(1) operator can be estimated as

$$\widehat{\Psi}(Y_{t-1}) = \sum_{j=1}^{t-2} \omega_h\left(Y_{t-1}, Y_j\right) Y_{j+1}. \tag{2.44}$$

Here, $\omega_h(.,.)$ denotes the weight function that is given by

$$\omega_h\left(Y_{t-1}, Y_j\right) = \frac{K\left(h^{-1}\mathrm{d}\left(Y_{t-1}, Y_j\right)\right)}{\sum_{i=1}^{t-2} K\left(h^{-1}\mathrm{d}\left(Y_{t-1}, Y_i\right)\right)}, \tag{2.45}$$

where K denotes a kernel function (such as the standard Gaussian density function) and $h > 0$ is called the bandwidth parameter of the kernel, which accounts for the level of smoothing that will be applied to the data and needs to be estimated from the data. Equation 2.44 can be interpreted as a weighted average of past functional observations, where the similarity between the observations is quantified by the (semi-)metric function $\mathrm{d}\left(.,.\right)$. As such, the choice of the distance function is critical, as it will be used to highlight the relevant features of the curves.

Some early results about nonparametric functional models can be found in Besse, Cardot, et al. (2000), where the authors analyze the applicability of these nonparametric forecasting models against other conventional non-functional models, and Masry (2005) and Delsol (2009), where asymptotic properties of these estimators is established. The monograph (Ferraty and Vieu, 2006) provides an excellent introduction to nonparametric techniques, which helped to popularize these nonparametric statistical methods. Recent examples of the nonparametric autoregressive model of order p can be found in Zhu and Politis (2017).

In real world applications the inclusion of exogenous covariates is of utmost importance, as they are often the main drivers of the functional time series. As such, Aneiros and Vieu (2008) introduced the semi-parametric model, which allows the inclusion of exogenous variables to the nonparametric autoregressive model. This model has been successfully applied in electricity load and price forecasting (Vilar, Cao, et al., 2012; Aneiros, Vilar, and Raña, 2016; Vilar, Aneiros, et al., 2018) and nonparametric forecasting of residual demand curves (Aneiros, Vilar, Cao, et al., 2013).

A different approach to these techniques can be found in Antoniadis, Paparoditis, et al. (2006) and Antoniadis, Paparoditis, et al. (2009), where the authors introduced a combination of kernel and wavelet techniques to define a suitable semi-metric that can be used to forecasts functional time series using nonparametric techniques. This kernel-wavelet approach is able to measure to quantify the similarity between segments of the functional observations, allowing for a robust forecasts of non-smooth curves. This approach has been illustrated with a real world application in electricity load forecasting (Antoniadis, Brossat, et al., 2016).

2.3.2.3. Parametric models

Following the discussion found in Chapter 1 of Ferraty and Vieu (2006), parametric models for functional time series forecasting can be defined as models that assume that the operators that appear in the formulation of the forecasting model belong to some known family, such as integral operators. This approach was formalized in the monograph Bosq (2000), where estimation techniques and stability conditions for the ARH(1) model were derived.

When estimating the operator Ψ of the ARH(1) process, it is necessary to obtain the inverse of the covariance operator of the process. Due to the infinite-dimensional nature of the data, this operator does not have a bounded inverse, hence the authors propose to truncate the spectral decomposition of the operator to obtain a pseudo-inverse that can be used in the estimation procedure. This can be seen as estimating the covariance matrix for the projections of the original functional data when projected onto the space spanned by its first M functional principal components. Other works such as Mas and Pumo (2009) and Horváth and Kokoszka (2012) follow the same approach, which has become a standard in the functional parametric literature. It can be seen that the estimated operator $\widehat{\Psi}$ obtained by this method is an integral operator whose kernel function takes the form

$$\widehat{\psi}(u, v) = \frac{1}{T-1} \sum_{t=1}^{T-1} \sum_{j=1}^{M} \sum_{i=1}^{M} \widehat{\lambda}^{-1} \langle Y_t, \widehat{\phi}_j \rangle \langle Y_{t+1}, \widehat{\phi}_i \rangle \widehat{\phi}_j(u) \widehat{\phi}_i(v), \tag{2.46}$$

where $\{\widehat{\lambda}_j\}_{j=1}^{M}$ denotes the estimated eigenvalues of the covariance operator of the data and $\{\widehat{\phi}_j\}_{j=1}^{M}$ are the first functional principal components of the curves. This can be seen as finding the optimal coordinates of the bivariate kernel function $\widehat{\psi}(u, v)$ when projected in the functional basis defined by the first principal components of the functional dataset.

This model has been widely studied in the literature. Mas (2007) and Álvarez-Liébana, Bosq, et al. (2017) analyze the asymptotic properties of the ARH(1) model. Other estimation methods for the ARH(1) model are illustrated in Didericksen et al. (2012), where the authors compare the finite-sample estimation properties of the different techniques. Additionally, Kargin and Onatski (2008) introduces a different estimation procedure for the ARH(1) operator: instead of projecting the operator in the first functional principal componets of the series, the define a a novel dimensionality technique that is derived directly from the minimization of the prediction error. These new directions, called predictive factors, can improve the forecasting capabilities of the model when compared with other estimators. The work by Horváth, Liu, et al. (2020) compares the predictive capabilities of several of these estimation techniques in a real-world financial application.

However, in many applications the ARH(1) model may not be enough, as the serial dependence exhibited by the curves may be influenced by several lagged values of the functional time series. The works by Bosq (2000) and Mourid (2002) provide an in-depth study of functional autoregressive processes of order $p > 1$, denoted as ARH(p) . The inclusion of exogenous covariates in the ARH(p) model was analyzed in Damon and Guillas (2002) and Damon and Guillas (2005), where the authors analyze the theoretical and empirical properties of this model. This model, denoted as ARHX(p), can be formulated as

$$Y_t = \sum_{j=1}^{p} \Psi_j \left(Y_{t-j}\right) + \sum_{z=1}^{Z} \Gamma_z \left(X_t^z\right) + \varepsilon_t, \tag{2.47}$$

where Ψ_j are the autoregressive operators, Γ_z denotes the regression operator associated with the functional covariate X_t^z and ε_t are the i.i.d. innovations of the model. The functional moving average model MAH(1) was introduced in Turbillon et al. (2007), where the authors develop an estimator for the moving average operator in Hilbert spaces. In addition, Klepsch et al. (2017) study the functional ARMA model, denoted by ARMAH(p, q). The authors approximate the model

$$Y_t = \sum_{j=1}^{p} \Psi_j \left(Y_{t-j}\right) + \sum_{i=1}^{q} \Theta_i \left(\varepsilon_{t-i}\right) + \varepsilon_t \tag{2.48}$$

by a vector ARMA model, obtained after projecting both the curves, innovations and operators in the space spanned by the FPC of the original functional time series.

This section was devoted to reviewing the different forecasting models found in the functional data literature, highlighting the different assumptions that each model makes and its real-world applicability. Nevertheless, there is still a relevant functional forecasting model that, due to its relevance to the contents of this thesis, it deserves its own section. This is the SARMAHX model, a parametric model that is a generalization of the scalar seasonal ARMA model with exogenous covariates. This model, proposed in Portela (2017), has proven to be a competitive model when compared against other reference forecasting models, due to its capability to capture complex dynamics that are often present in real-world functional time series such as electricity prices (Portela, Muñoz San Roque, et al., 2018). The following section will introduce this model together with its main properties and assumptions.

2.3.3. The SARMAHX model

This section introduces the functional SARMAHX model, which is defined following the standard time series modeling approach proposed in Box et al. (2008) but extended to functional time series using integral operators. The proposed SARMAHX model generalizes the scalar ARMAX model by extending the scalar model to functional time series using linear operators defined on the Hilbert space L^2 as the parameters of the model. While the parameters of the classical ARMAX model are scalar values, the parameters of the SARMAHX model are functional operators that model the relation between input and output curves. In order to estimate these parameters, a new methodology was proposed in Portela (2017), which differs substantially from the parametric estimation methods described in Section 2.3.2.3. As such, the modelization and estimation of these functional operators is also detailed in this section. This section follows Portela (2017) and Mestre, Portela, Muñoz San Roque, et al. (2020).

2.3.3.1. Model formulation

The SARMAHX$(P_0, Q_0) \times (P_1, Q_1)_{s_1} \times (P_2, Q_2)_{s_2}$ 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:

$$\begin{aligned}\prod_{j=0}^{2}\left(I-\sum_{i=1}^{P_j}\Psi_{j,i}B^{i\cdot s_j}\right)(Y_t) = \prod_{k=0}^{2}\left(I-\sum_{l=1}^{Q_k}\Theta_{k,l}B^{l\cdot s_k}\right)(\varepsilon_t) \\ +\sum_{z=1}^{Z_f}\Gamma_z^f(X_t^z)+\sum_{z=1}^{Z_c}\Gamma_z^c(x_t^z),\end{aligned} \tag{2.49}$$

where:

- $\{Y_t\,(v)\,; t=1,2,...,T; v\in V\}$ a stationary and zero-mean functional time series.
- $\{X_t^z(v_z); z=1,2,\cdots,Z_f; t=1,2,\cdots,T; v_z\in V_z\}$ a set of Z_f functional covariates.
- $\{x_t^z; z=1,2,\cdots,Z_c; t=1,2,\cdots,T\}$ a set of Z_c scalar covariates.
- ε_t is a functional white noise process.
- I is the identity operator.
- Parameters P_0, P_1 and P_2 are the regular and the two seasonal autoregressive orders respectively.
- Parameters Q_0, 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. Parameter s_0 is equal to 0.
- $\Psi_{0,i}$, $\Psi_{1,i}$ and $\Psi_{2,i}$ are the regular and seasonal autoregressive operators.
- $\Theta_{0,l}$, $\Theta_{1,l}$ and $\Theta_{2,l}$ are the regular and seasonal moving average operators.
- Γ_z^f are the operators related to the Z_f explanatory variables.
- Γ_z^c are 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 $n\in\mathbb{N}$.

Two versions of the SARMAHX model are considered. If integral operators defined in the L^2 space are used for the ARMA terms, the fully functional SARMAHX model is obtained. This version of the model, illustrated in Portela, Muñoz San Roque, et al., 2018 in a electricity price forecasting problem, is able to capture the cross-correlation dependence between different discretization points of the curves. 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:

$$\begin{aligned}\widehat{Y}_t(v') &= \sum_{i=0}^{p}\sum_{j=0}^{P_1}\sum_{k=0}^{P_2}\iiint A(v,v',u,u')Y_{t-\tilde{t}}(u)\,dudu'dv \\ &-\sum_{i=0}^{q}\sum_{j=0}^{Q_1}\sum_{k=0}^{Q_2}\iiint B(v,v',u,u')\widehat{\varepsilon}_{t-\tilde{t}}(u)\,dudu'dv \\ &+\sum_{z=1}^{Z_f}\int \rho_z\left(v_z,v'\right)X_t^z\left(v_z\right)\,dv_z+\sum_{z=1}^{Z_c}\beta_z\left(v'\right)x_t^z,\end{aligned} \tag{2.50}$$

where $u,v,v' \in V$, $v_z \in V_z$, $\tilde{t}$ is defined as

$$\tilde{t} = i + j\cdot s_1 + k\cdot s_2,$$

$\widehat{\varepsilon}_t$ is the estimation of past innovations, which is given by:

$$\widehat{\varepsilon}_t\left(u\right) = Y_t\left(u\right) - \widehat{Y}_t\left(u\right), \tag{2.51}$$

and

$$A(v,v',u,u') = \psi_{0,i}\left(v,v'\right)\psi_{1,j}\left(u',v\right)\psi_{2,k}\left(u,u'\right), \tag{2.52}$$

$$B(v,v',u,u') = \theta_{0,i}\left(v,v'\right)\theta_{1,j}\left(u',v\right)\theta_{2,k}\left(u,u'\right). \tag{2.53}$$

In order to ease the comprehension of the SARMAHX model, the following example is detailed. The SARMAHX$(1,1)$ $\times(1,0)_{24}$ model would be defined following Equation 2.49:

$$Y_t = \Psi_{0,1}(Y_{t-1}) + \Psi_{1,1}(Y_{t-24}) - \Psi_{0,1}\Psi_{1,1}(Y_{t-25}) - \Theta_{0,1}(\varepsilon_{t-1}) + \varepsilon_t, \tag{2.54}$$

where the terms are, respectively, the regular autoregressive, the seasonal autoregressive, the autoregressive seasonal interaction and the moving average term. It should be noted that the term $\Psi_{0,1}\Psi_{1,1}(Y_{t-25})$ denotes the composition, i.e. $\Psi_{0,1}(\Psi_{1,1}(Y_{t-25}))$. Then, substituting each operator by its integral expression, the forecasting equation becomes:

$$\begin{aligned}\widehat{Y}_t(v') &= \int \psi_{0,1}(u,v')Y_{t-1}(u)\,du + \int \psi_{1,1}(u,v')Y_{t-24}(u)\,du \\ &- \iint \psi_{0,1}(v,v')\psi_{1,1}(u,v)Y_{t-25}(u)\,du\,dv \\ &- \int \theta_{0,1}(u,v')\widehat{\varepsilon}_{t-1}(u)\,du\end{aligned} \tag{2.55}$$

As mentioned in Section 2.2.6, if the dependence exhibited between the curves at different discretization points is negligible, concurrent operators may be used instead of integral operators. These operators are defined as

$$\Psi_i(f)(v) = \psi_i(v)f(v) \qquad f \in L^2, \tag{2.56}$$

and if concurrent operators are used in the ARMA terms of equation Equation 2.49 the concurrent version of the SARMAHX model is obtained. Due to its simpler formulation, the representation capabilities of these operators is more limited than integral operators, as they are not able to model cross correlations between curves (i.e. the first part of the input curve affecting the last part of the output curve). However, the formulation of the model is simplified, which can ease the optimization of the parameters. For example, Equation 2.55 using concurrent operators would have the following expression:

$$
\begin{aligned}
\widehat{Y}_t(u) = {} & \psi_{0,1}(u)Y_{t-1}(u) + \psi_{1,1}(u)Y_{t-24}(u) \\
& - \psi_{0,1}(u)\psi_{1,1}(u)Y_{t-25}(u) - \theta_{0,1}(u)\widehat{\varepsilon}_{t-1}(u).
\end{aligned}
\tag{2.57}
$$

As a conclusion, the resulting model (either the fully functional or the concurrent version) admits 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. Therefore, this model is suitable for most functional time series found in many real-world applications, such as price and load forecasting in electricity markets.

2.3.3.2. Estimation procedure

In order to fit the SARMAHX model, each integral operator must be estimated from the observed data. As already mentioned in Section 2.2.6, estimating an integral operator Ψ implies estimating the associated kernel surface $\psi(u, v)$. The standard approach for the estimation of these operators is to project the kernel into a subspace spanned by a functional basis, as explained in Section 2.3.2.3. Then, the coordinates of the kernel in that subspace are optimized so that the forecasting error is minimized. Thus, the selection of the functional basis is critical when applying these estimation methods.

In order to avoid choosing one fixed functional basis, the SARMAHX model follows a novel approach to functional parameter estimation: each kernel function is modeled as a finite sum of bivariate sigmoid functions (Portela, 2017; Portela, Muñoz San Roque, et al., 2018). Sigmoid functions are universal function approximates (Cybenko, 1989) which are commonly used in neural networks due to their properties to model non-linear relations. As such, each bivariate kernel $\psi(u, v)$ can be modeled as:

$$
\psi(u, v) = \alpha_0 + \sum_{g=1}^{G_\psi} \alpha_g \tanh\left(w_{g0} + w_{g1}u + w_{g2}v\right), \tag{2.58}
$$

where $w_{g0}, w_{g1}, w_{g2}, \alpha_g$ and α_0 are the parameters that define each sigmoid. The variables u and v take real values in the intervals in which the functional variables are defined (i.e. V or V_z). In the case of a concurrent operator (2.56), this expression is simplified, modeling each kernel as a finite sum of univariate sigmoids:

$$
\psi(v) = \alpha_0 + \sum_{g=1}^{G_\psi} \alpha_g \tanh\left(w_{g0} + w_{g1}v\right). \tag{2.59}
$$

This approach can be viewed as a Multi-Layer Perceptron (MLP) neural network with a particular configuration: an input layer with two input variables and a bias; one hidden layer with a number G_ψ of nonlinear hidden units with hyperbolic tangent as the activation function and w_{g0}, w_{g1}, w_{g2} as the weights for each input. Finally, one output layer with one linear output

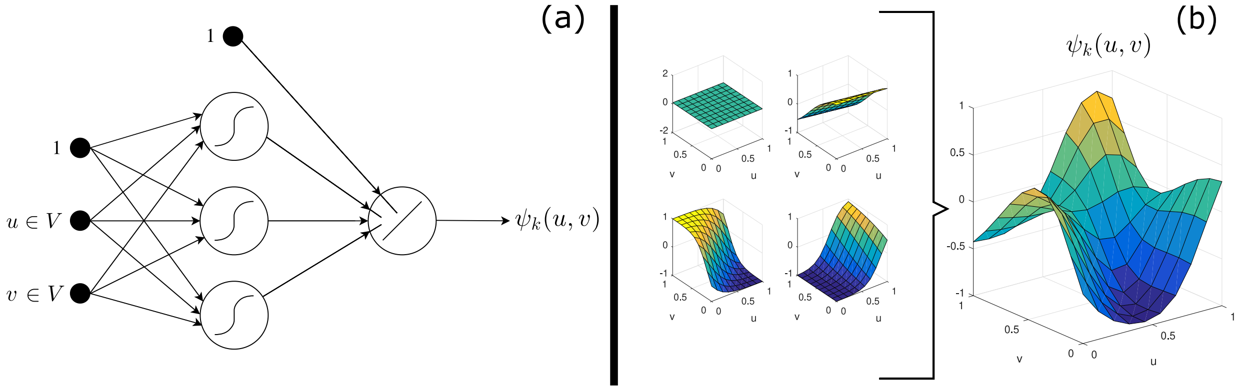

Figure 2.7. (Part a) Architecture diagram of the neural network used for optimizing the functional parameters of the SARMAHX model. (Part b) Kernel function estimated with 3 bivariate sigmoid functions and a constant surface.

unit having α_g as the weights for the activation of the hidden units. Figure 2.7 (Part a) shows the architecture diagram of the aforementioned network with $G_\psi = 3$ hidden layers. The property of sigmoidal surfaces being universal approximators is illustrated in Figure 2.7 (Part b), where a rather complex surface is modeled as the sum of 3 sigmoidal functions and a constant surface that models the level of the final surface. It has been observed that using 5 or 6 sigmoid functions when fitting the SARMAHX model is enough in the vast majority of practical applications, due to the flexibility of sigmoid functions.

The parameters $(w_{g0}, w_{g1}, w_{g2}, \alpha_g, \alpha_0)$ completely define each bivariate kernel in Equation 2.58. Therefore, the proposed SARMAHX model is estimated when values for all these parameters are estimated. In order to achieve this, a low-memory Quasi Newton method with random initial weights has been implemented to optimize these real-valued parameters so as to minimize a certain cost function. In Portela (2017), the cost function C for estimating the SARMAHX model is defined as the sum of the L^2 square errors,

$$C = \sum_{t=1}^{T} e_t, \tag{2.60}$$

where

$$e_t = \left\| Y_t - \widehat{Y}_t \right\|_{L^2}^2 = \int \left(Y_t(u) - \widehat{Y}_t(u) \right)^2 du. \tag{2.61}$$

The implemented Quasi Newton method is a gradient descent algorithm, so the derivatives of the error term shown in Equation 2.60 with respect to the sigmoid parameters are needed. The derivative of the error function with respect to a general parameter W is given by:

$$\frac{\partial C}{\partial W} = \sum_{t=1}^{T} \int 2 \left(Y_t(u) - \widehat{Y}_t(u) \right) \left(-\frac{\partial \widehat{Y}_t(u)}{\partial W} \right) du, \tag{2.62}$$

where $\frac{\partial \widehat{Y}_t(u)}{\partial W}$ is the partial derivative of the estimation with respect to the generic parameter W. In order to reduce computational times, analytical expressions for these derivatives have been obtained. The reader is referred to Portela, Muñoz San Roque, et al. (2018) for the formulation of the derivatives of the cost function with respect to the weight parameters.

However, neural network procedures are known to have a tendency towards overfitting. If no regularization techniques are used in the learning process, the resulting network will be so

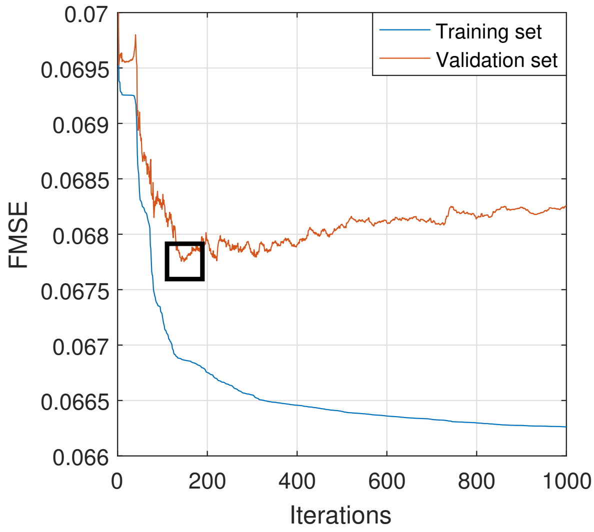

Figure 2.8. Evolution of both training and validation errors in the optimization process for the SARMAHX model. The training error always decreases in each iteration of the algorithm, but the same is not true for the validation set. The early-stopping method is used to select the parameter configuration that yields a lower validation error (marked in the figure with a square).

closely fitted to the training data that it causes the model to not generalize well in the presence of new data. In order to prevent this effect, early stopping methodology is used: the observed data is split into a training and a validation set and executing the optimization algorithm with the training set, thus minimizing the cost function over these data. At the same time, the adjusted model in each iteration is being evaluated with the validation set. Figure 2.8 illustrates this methodology: the training error of the model decreases in each iteration while the error among the validation set is oscillating. As the best model is the one with better generalization performance among an unknown dataset, the final model will be the one that produces a lower error in the validation data.

The training and evaluation of the SARMAHX 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, so both the time series $\{Y_t\}_{t=1}^T$ and the functional covariates $\{X_t\}_{t=1}^T$ are centered and normalized as

$$Y_t^*(v) = \left(Y_t(v) - \mu_{\mathcal{Y}}(v)\right)/\sigma_Y \quad , \quad X_t^*(v) = \left(X_t(v) - \mu_{\mathcal{X}}(v)\right)/\sigma_X, \tag{2.63}$$

where σ_Y and σ_X are normalization values for which $Y_t^*(v) \in [-1,1]$ and $X_t^*(v) \in [-1,1]$. 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 domain of the time series $\{Y_t^*\}_{t=1}^T$ and the functional covariates $\{X_t^*\}_{t=1}^T$ is moved to the range $[-1,1]$. By default, the sigmoid parameters are initialized with small values, causing the sigmoids to be centered around 0. As such, the position of the sigmoids wont need to move far from that initial position.

This section summarizes the main properties of the SARMAHX model, highlighting the features that make this model a suitable model to forecast functional time series that appear on electricity markets. The inclusion of exogenous covariates (either scalar or functional variables) and the ability of including up to two seasonal terms makes it possible to capture the complex dynamics that hourly functional time series, such as the series of residual demand curves, exhibit.

However, before fitting the SARMAHX model, the user needs to select both the seasonality present in the data and the order of the autoregressive and moving average terms. As such, it is necessary to develop some kind of identification and diagnosis technique for assessing the adequacy of these functional SARMAX models. Chapter 3 is devoted to this purpose, developing an analogue of the Box-Jenkins identification methodology for functional time series.

2.4. Probabilistic models for functional time series

The main goal of forecasting is to obtain estimations about the distribution of some future value of interest. The vast majority of research in forecasting models is aimed at producing and evaluating point forecasts (estimates of the mean of the distribution of interest). Point forecasts are of utmost importance, providing the forecaster with an easy to understand description of the future evolution of the series. However, while point forecasts are adequate in some cases, the vast majority of applications require having estimations of the uncertainty associated with the forecasts in order to optimize their decision-making process. As such, the last decade as seen a growing interest in the development of probabilistic forecasting models, see Gneiting and Katzfuss (2014) and Hong and Fan (2016) for an in-depth review of probabilistic electricity load forecasting models.

The aim of probabilistic forecasting models for scalar time series is to estimate the distribution of a certain quantity of interest. Probabilistic forecasts can be expressed in a variety of ways (Zhang, Wang, et al., 2014): through a probability measure, such as the variance or skewness of the future event; using quantiles and intervals of the estimated future distribution (Anastasiades and McSharry, 2013); and by providing the user with an estimation of the probability density function of the data (Hyndman and Fan, 2010). In several real-world applications, probabilistic forecasts can be used for risk-based decisions by quantifying the uncertainty associated with the forecasts. In addition, probabilistic models may overcome the limitations of point-forecasting models by allowing the generation of future scenarios that are consistent with the estimated density for future events.

Probabilistic models for time series can be classified in two large groups, based on how the model estimates the future distribution of the series of interest. The first set of probabilistic models aim at estimating the whole distribution of the output variables, often using non-parametric methods such as kernel density estimators, as shown in Monteiro et al. (2018); probabilistic radial basis function networks (Muñoz San Roque and Sanz-Bobi, 1998) or by characterizing the distribution of interest via its quantiles, using techniques such as quantile regression models (Bello et al., 2017). Although these models have been applied successfully in several real-world scenarios, they are often not as interpretable as other time-series based methods, which is a quality that is often desired in many applications. The second type of probabilistic models found in the literature are based on quantifying the uncertainty associated with a deterministic forecasting model, which can be expressed as

$$y_t = D(\boldsymbol{y}_{t-1}) + \varepsilon_t, \tag{2.64}$$

where $D(\boldsymbol{y}_{t-1})$ denotes the deterministic part of the model and ε_t is the stochastic part of the model. After the deterministic model is fitted, the residuals of the model are analyzed in order to obtain an estimation of the distribution of the errors of the model. Once the distribution of the residuals has been characterized, probabilistic statements can be obtained for the forecasts of the deterministic model. The monograph Chatfield (2001) contains an in-depth review of the most common techniques used to quantify these distributions.

Although there are many functional models for point-forecast (Hyndman and Shang (2009), Antoch et al. (2010) and Aneiros, Vilar, Cao, et al. (2013), among others), methods for obtaining prediction intervals for functional time series have not been studied extensively up to date. Some remarkable works in this direction are Antoniadis, Brossat, et al. (2016), where the authors use a nonparametric functional model to obtain pointwise prediction intervals for the electricity load in the, and Vilar, Aneiros, et al. (2018), where prediction bands are obtained for a semi-parametric functional model. One of the main objectives of this thesis is to develop a functional probabilistic model that is able to model the distribution of a given functional time series, providing an estimation of the conditional probability density function of any future event. As such, a model which is able to characterize the conditional distribution of the residuals of a functional deterministic model is preferred, as it allows us to benefit from the deterministic forecasts of functional time series models that are often used in practice, obtaining much sharper distributions. The advantages of this approach are twofold: firstly, it allows us for using the predictive capabilities of the SARMAHX model (Portela, Muñoz San Roque, et al., 2018) introduced in Section 2.3.3, which has proven to be a competitive model in functional time series forecasting. Secondly, by quantifying the uncertainty of the SARMAHX forecasts, realistic scenarios that are consistent with the deterministic forecasts can be generated.

This section is devoted to performing a review of probabilistic models for functional data proposed in the literature. Firstly, the definition of an analogue of the density function for functional datasets is presented in Section 2.4.1, analyzing the extension of the notion of density of a finite-dimensional dataset to the functional framework. Secondly, Section 2.4.2 will review the different probabilistic models found in the literature for functional data, grouping them according to the methodology employed to generate probabilistic forecasts. The advantages and disadvantages of each method will be analyzed, which will provide a framework for the contributions of this thesis.

2.4.1. On the lack of a density function for functional data

The major drawback for the development of probabilistic models for functional data is the lack of a propper probability density function (Delaigle and Hall, 2010; Cuevas, 2014). The non-existence of a density function in the L^2 space with respect to the Lebesge measure was established in Li and Linde (1999). The authors focused on the concept of "small ball probabilities", this is, the probability of a random function $\mathcal{X}$ falling into a small ball $\mathrm{B}\left(\mathcal{X}',\delta\right)$ of center $\mathcal{X}'$ and radius δ for the topology of its respective function space given the distance function $\mathrm{d}\left(.,.\right)$:

$$\mathrm{B}\left(\mathcal{X}',\delta\right) = \left\{\mathcal{X}^* \in L^2 \,|\, \mathrm{d}\left(\mathcal{X}',\mathcal{X}^*\right) < \delta\right\}. \tag{2.65}$$

The authors studied the probability $P\left(\mathcal{X} \in \mathrm{B}\left(\mathcal{X}',\delta\right)\right)$ and found that the expected number of functions falling into balls with a small radius is so low that densities do not even exist in most cases (Dai et al., 2017). This makes a substantial difference with scalar probabilistic models, as the vast majority of those models rely heavily on having estimates of the probability density function of the data. As such, the literature on functional probabilistic models is scarce, as authors must employ non-standard techniques for quantifying the uncertainty associated with functional models.

In order to overcome this, several alternatives have been proposed in the literature. In Delaigle and Hall (2010) the authors propose approximating the probability density function for functional data as the joint density function of the first M FPC scores of the curves. Given $X_1, X_2, \ldots, X_N$ a set of N functional observations obtained from the same distribution as the functional random variable $\mathcal{X}$, the projection-based density $p_{\mathcal{X}}^{(M)}$ of the curves can be estimated using a kernel density model:

$$\hat{p}_{\mathcal{X}}^{(M)}\left(\boldsymbol{u}, H\right) = \frac{1}{N|H|^{1/2}} \sum_{i=1}^{N} \mathbf{K}\left(H^{-1/2}\left(\boldsymbol{u} - \boldsymbol{x}_i\right)\right), \tag{2.66}$$

where H is the bandwidth matrix of the model, $|H|$ denotes the determinant of the matrix, and $\boldsymbol{x}_i$ are the scores of the first M FPC associated with the functional observation X_i. In this formulation, $\mathbf{K}$ denotes a multidimensional kernel function, such as the standard Gaussian density. The choice of the bandwidth matrix H is done using standard kernel density estimation methods (Silverman, 1986). The authors prove that their approximation captures the distribution of the curves up to a certain order, related to the number of principal components used in the decomposition. This technique has been used in both classification (Delaigle and Hall, 2012; Dai et al., 2017) and model-based clustering (Bouveyron and Jacques, 2011; Jacques and Preda, 2013; Jacques and Preda, 2014) of functional datasets.

This technique is illustrated in Figure 2.9, where the density of the electricity load profiles is approximated by the joint density of its first 2 FPC. As can be seen, the obtained density of the functional data presents a multimodal behavior, with four distinct density clusters that are associated with different shapes of the curves. The interpretation of the functional principal components that was carried out in Section 2.2.5 can be useful to analyze the estimated density function of our data: recall that the first principal component models the level of the curves, hence the leftmost cluster seen in the figure can be associated with days with a lower demand (typically weekends and holidays). In addition, the second component accounts for the difference between summer and winter, as extreme temperature values greatly affect the electricity consumption. The effect on the shape of the curves can be seen in the estimated distribution, where the highest density cluster is located in the negative range of the FPC2 and only a small number of points fall in the rightmost positive density cluster.

Other authors follow the definition of Ferraty, Kudraszow, et al. (2012) to estimate an equivalent version of the probability density for functional data based on some (semi-)metric function d^*. Let $\{\mathcal{X}_i\}_{i=1}^{T}$ be a set of functional random variables independent and identically distributed as $\mathcal{X}$ and valued in L^2. The authors analyze the small ball probabilities associated with the open balls in the topology associated with the semi-metric d^*, and assume that there exist functions $\varphi(.)$ and $p_{\mathcal{X}}^{\mathrm{d}^*}(.)$ such that for all $X \in L^2$,

$$P\left(\mathcal{X} \in \mathrm{B}\left(\mathcal{X}', \delta\right)\right) \sim p_{\mathcal{X}}^{\mathrm{d}^*}(X)\varphi(\delta) \quad \text{as } \delta \to 0, \tag{2.67}$$

which implies that $P\left(\mathcal{X} \in \mathrm{B}\left(\mathcal{X}', \delta\right)\right) = p_{\mathcal{X}}^{\mathrm{d}^*}(X)\varphi(\delta) + o(\varphi(\delta))$ as $\delta \to 0$. The function $p_{\mathcal{X}}^{\mathrm{d}^*}(X)$ can be considered as an infinite-dimensional analogue of the concept of probability density, while $\varphi(\delta)$ may be interpreted as a volume parameter. In order to obtain approximations to this density $p_{\mathcal{X}}^{\mathrm{d}^*}(X)$, the authors propose the following kernel estimator:

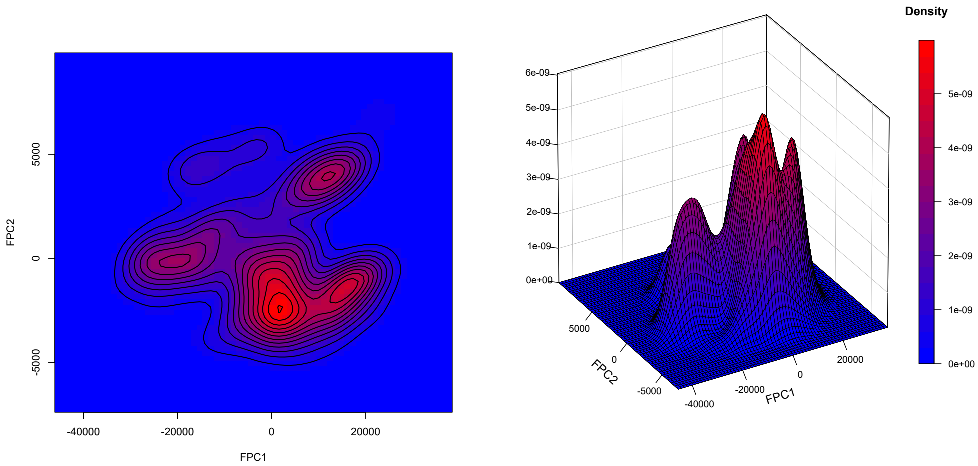

Figure 2.9. Contour plot (left) and perspective plot (right) of the projection-based density for the series of electricity load profiles using the first 2 FPC of the data.

$$\hat{p}_{\mathcal{X}}^{\mathrm{d}^*}(X,h) = \frac{1}{T-1}\frac{\sum_{i=1}^{T} K(h^{-1}\mathrm{d}^*\left(X, X_i\right))}{\sum_{i=1}^{T}\sum_{j\neq i} K(h^{-1}\mathrm{d}^*\left(X_j, X_i\right))}, \tag{2.68}$$

where K is a kernel function and $h > 0$ is the bandwidth parameter of the kernel. By employing a distance function as a notion of similarity between the curves, their model is able to define a notion of central tendency in a population of random functional variables. This approach has been used both in classification of functional datasets (Bongiorno and Goia, 2016) and (Ciollaro et al., 2014) for clustering functional data.

This distance-based density function has been implemented in order to compare this technique with the projection-based density described before and analyze the benefits of these two approaches. Figure 2.10 illustrates this method using the same electricity load profile dataset: the model described in Equation 2.68 has been fitted using the historical curves, and the fitted density has been evaluated over the 7 new load curves shown in the left panel. The metric function that has been used in this example is the L^2 distance, and the bandwidth parameter h has been selected using the method proposed in Ferraty, Kudraszow, et al. (2012). The right panel shows the estimated density for each new curve, using the color code represented in the legend. As can be seen, the density of curves 2, 3, 4 and 7 is higher than the remaining curves, due to these curves exhibiting a similar level and shape to the majority of the historical curves. The level of curve 5 is similar to curves 2, 3 and 7; however the odd shape of this new curve lowers the density given by the model. Finally, curves 1 and 6 are electricity load profiles for weekends, which explains their low probability.

Nevertheless, while these techniques have been applied in both classification and clustering for functional data, the same cannot be said for functional forecasting models. When dealing with a time series model, the practitioner is usually interested in measuring the uncertainty associated with its forecasts. As mentioned earlier, this often involves estimating the distribution of the residuals of the model. A common technique for approximating the distribution of the residuals of a time series model is the bootstrap (Efron and Tibshirani, 1994; Chatfield, 2001). This technique involves sampling from the empirical distribution of past fitted errors. The procedure effectively approximates the theoretical distribution of the innovations of the model

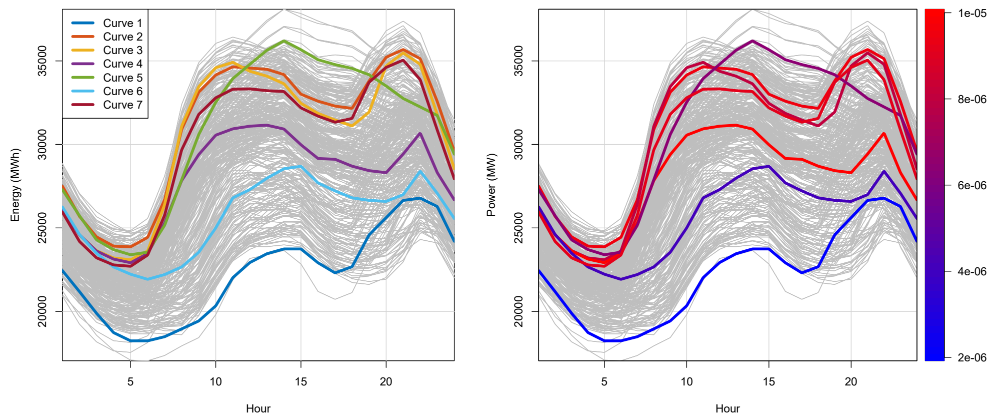

Figure 2.10. Left: set of historical electricity load curves used to fit the distance-based kernel density model. A set of new curves is included, in order to test the probability assigned by the model. Right: probability density given by the model to each new curve.

by the empirical distribution of the observed residuals. The literature in boostrap methods applied to functional time series models is growing rapidly: in Raña, Aneiros, et al. (2016) the authors perform a review of bootstrap methods in the functional data framework, together with its asymptotic properties. In Vilar, Aneiros, et al. (2018) the authors develop a bootstrap algorithm that accounts for heteroscedasticity in the residuals of a fitted functional forecasting model, whereas Rossini and Canale (2019) propose a block bootstrap method to quantify the uncertainty associated with a functional autoregressive model.

2.4.2. Probabilistic models for functional time series

The last section was devoted to analyzing the concept of a density function for functional data and to proposing different approaches to quantify the distribution of a functional random variable. This section is aimed at analyzing the models that have been proposed in the literature to obtain probabilistic forecasts of a functional time series. Most of the probabilistic models found in the literature are aimed at estimating the distribution of these functional variables through the generation of scenarios of the original curves, so the available scenario generation models will be classified according to the technique that is employed to obtain said functional scenarios.

2.4.2.1. Dimensionality-reduction models

Methods listed in this section are based on dimensionality-reduction models, which have already been introduced in Section 2.3.2.1. In these models, the functional time series is projected on a fixed functional basis (tipically FPCA) and then multivariate models are used to model the finite-dimensional time series of the scores. Once the future scores have been estimated, future curves can be reconstructed by applying the inverse transformation of the decomposition.

As these models effectively reduce the dimension of the data, classical techniques can be used to obtain prediction intervals for the multivariate series of scores. The prediction intervals obtained for the scores can then be combined to obtain prediction bands for the functional time series that is being modeled. This approach was used in Hyndman and Ullah (2007), Hyndman and Shang (2009), and Shang (2013), where the functional time series is decomposed in their Functional Principal Components to project the data into a low-dimensional space. Then, ARIMA models are fitted to forecast the series of the FPC scores, and the prediction intervals of the ARIMA models are used to define the prediction intervals for the functional observations. The authors illustrate this technique with applications to forecasting demographic curves and monthly sea temperatures. The R package `ftsa` contains functions to replicate the methods proposed in those works. In Liebl (2013), a functional factor model is proposed to obtain estimations of the electricity price using the electricity demand as an explanatory variable. The authors use FPCA to decompose the factor function and then uses a SARIMA model to estimate the scores of the decomposition. Then, the 95% prediction intervals of the fitted SARIMA models are used to generate prediction bands of the estimated price by reconstructing the upper and lower bands using the Karhunen-Loéve representation.

Although the models reviewed in this section are devoted to estimating prediction bands, functional scenarios can be obtained using these techniques. Standard resampling techniques can be used to simulate values of the FPC scores, which then can be used to obtain future functional scenarios by reconstructing the curves. One example of this technique can be found in Hyndman and Booth (2008), where the authors obtain bootstrap replications of the residuals of the fitted multivariate models to generate sample paths of the estimated demographic curves.

2.4.2.2. Assumptions on the distribution of the residuals

A common assumption when fitting a forecasting model is that the residuals of the model follow a normal distribution. This allows the practitioner to simulate future innovations of the process by sampling from the theoretical Gaussian distribution. This approach has been extended to the functional data framework, often used together with dimensionality-reduction techniques. In Mestekemper et al. (2010), the authors obtain estimations for the functional time series of water temperature by applying FPCA and estimating each component by using a Vector Autoregressive (VAR) model. Then, in order to obtain prediction bands for its forecasts, it is assumed that the residuals of the fitted VAR model follow a multivariate Gaussian distribution, hence defining 95% prediction intervals as

$$\widehat{\boldsymbol{y}}_{t+1} \pm 1.96\sqrt{\mathrm{diag}(\widehat{\Sigma})}, \tag{2.69}$$

where $\widehat{\Sigma}$ is the estimated covariance matrix of the residuals of the fitted VAR model.

A similar approach can be found in Shang (2012) and Erbas et al. (2007), where the authors use FPCA to reduce the dimensionality of the input curves to obtain estimations of age-specific fertility rates and breast cancer mortality, respectively. In both models, the residuals of the fitted multivariate are assumed to be normal, obtaining pointwise prediction intervals for its models.

2.4.2.3. Bootstrap replication

As mentioned in the previous section, bootstrap techniques are one of the most commonly used method to approximate the unknown distribution of the innovations of a time series model by sampling from the empirical distribution of the fitted residuals. Due to their ease of implementation and their good asymptotic properties, bootstrap methods have been widely used in the functional data literature. For example, Rossini and Canale (2019) proposes a moving block bootstrap to generate future scenarios of the daily supply and demand curves of the Italian natural gas market. From these functional scenarios, point-wise prediction bands can be obtained. In Ferraty, Keilegom, et al. (2009) and Zhu and Politis (2017) a bootstrap procedure is applied to the residuals of a functional non-parametric model to obtain confidence intervals for the estimation of the regression operator of a non-parametric functional model. The naive bootstrap method proposed by the authors can be summarized as follows: given $\{Y_t\}_{t=1}^T$ a functional time series, the authors fit a non-parametric autoregressive model (see Equation 2.44) to the series. The estimations of the model are denoted as $\{\widehat{Y}_t\}_{t=1}^T$. Then, the authors propose a naive bootstrap procedure for obtaining functional scenarios of the series:

1. For all $t = 1, \ldots, T$ define the residuals of the model as $\widehat{\varepsilon}_t = Y_t - \widehat{Y}_t$.

2. Let $\overline{\widehat{\varepsilon}} = T^{-1} \sum_{j=1}^T \widehat{\varepsilon}_j$ be the sample mean function of the model's residuals. Define new i.i.d. elements $\varepsilon_1^{\text{boot}}, \ldots, \varepsilon_T^{\text{boot}}$ such that

$$P\left(\varepsilon_i^{\text{boot}} = \widehat{\varepsilon}_j - \overline{\widehat{\varepsilon}}\right) = T^{-1} \tag{2.70}$$

for $i, j = 1, \ldots, T$.

3. For $t = 1, \ldots, T$, let

$$\widehat{Y}_t^{\text{boot}} = \widehat{Y}_t + \varepsilon_t^{\text{boot}}. \tag{2.71}$$

Repeat this process B times to obtain B sets of functional scenarios.

The authors then use these functional scenarios to obtain confidence intervals for the estimation of the non-parametric autoregressive operator. This work is improved in Ferraty, Van Keilegom, et al. (2012), where the residuals of the model are projected in their first principal components and the bootstrap procedure is employed to estimate the density of each component of the projected error. Hence, they are able to obtain confidence areas for the estimation of the regression operator of their model.

An in-depth comparison of bootstrap methods for non-parametric functional regression models can be found in Raña, Aneiros, et al. (2016). The methods proposed by the authors are generalized in Vilar, Aneiros, et al. (2018), where bootstrap algorithms are proposed to account for heteroscedasticity in the residuals of a fitted semi-parametric data. Point-wise prediction bands are obtained using the bootstrapped residuals, which are then used in a case study using data from the Spanish electricity market. The prediction intervals obtained are able to account for both the variability on the estimation of the autoregressive operator and the uncertainty of the model's errors. The model proposed by Antoniadis, Brossat, et al. (2016) obtains a wavelet decomposition of the original curves, and then uses a non-parametric model to estimate the curves based on the wavelet distance. In order to obtain prediction intervals, a weighted bootstrap algorithm is proposed, which uses the distribution induced by the weights of the fitted non-parametric model to sample historical residuals. This method is illustrated with a case-study to forecast French daily electricity load profiles.

As seen, most of the functional probabilistic models found in the literature are concerned with modeling the unconditional distribution of the residuals of a deterministic forecasting model. However, when modeling functional time series that appear in electricity markets, such as the daily price profiles or residual demand curves, the distribution of the data is often influenced by changes in other exogenous variables. One example of this behavior can be seen in Figure 2.9, where the density of the electricity load curves was shown. The estimated density illustrates how extreme temperatures and the level of industrial activity affect the distribution of the curves. As such, in order to develop a competitive functional probabilistic model, the conditional distribution of the residuals has to be modeled. This will allow the model to provide more accurate estimations of the uncertainty of the forecasts and to generate future scenarios that follow the estimated conditional distribution exhibited by the original data.

2.5. Conclusions

This chapter has been devoted to introducing the theoretical background and the main techniques used for Functional Data Analysis (FDA), a novel research field that aims to analyze and model the behavior of functional random variables. As these variables are real-valued functions that belong to an infinite dimensional space, classical multivariate techniques cannot be applied directly to functional datasets. This has motivated the development of new methodologies that are able to extend the standard statistical concepts to infinite-dimensional observations. A brief summary of these statistical tools has been included in this chapter, such as functional principal components and a review of functional regression models.

This thesis is concerned with modeling and forecasting functional time series, which are realizations of a functional stochastic process. Functional time series appear in many practical applications, so it is of utmost importance to develop models that are able to capture the dynamics exhibited by the series. A review of the different forecasting models for functional time series has been performed, classifying each model in three large groups (dimensionality reduction, nonparametric and parametric models) according to the estimation techniques used to obtain estimations of the series. In addition, this chapter includes an introduction to the SARMAHX model, a parametric model that generalizes the classical seasonal ARMA models to the functional framework. Due to its flexibility, the SARMAHX model is able to model the complex behavior often exhibited by functional time series found in electricity markets, making it a suitable model in many real-world applications. However, when fitting the SARMAHX model to a functional dataset, it is critical to define a methodology that allows the practitioner to select the optimal orders of the model. One of the main objectives of this thesis is to propose an identification and diagnosis procedure that can be used to quantify the serial correlation of a given functional time series, providing a useful tool when selecting the parameters of the model.

Although the literature on deterministic forecasting models for functional time series is quite extensive, the same cannot be said about functional probabilistic models. The development of a model that is able to account for the uncertainty associated with point forecasts would be of major interest, as this will allow users to make risk-based decisions. The literature on functional probabilistic models is scarce, often relying in normality assumptions or resampling techniques to quantify the uncertainty of the models. This lack of functional probabilistic models can be explained by the infinite-dimensionality of the observations, making it impossible to define a proper density function for functional datasets. As such, new approaches to the concept of density for functional data have been developed, which provide an approximation to the density function of a functional dataset. These methods have been successfully applied to the classification and clustering of functional data, but up to our knowledge, they have not yet been applied to probabilistic forecasting of functional time series. This thesis is concerned with developing a probabilistic model for functional time series that is able to model the distribution of the forecasts and to generate future scenarios that are consistent with that distribution. Therefore, the statistical concepts reviewed in this chapter will be of utmost importance when formulating this new model.

3

Identification and diagnosis of the SARMAHX model

Quantifying the serial correlation across time lags is a crucial step in the identification and diagnosis of a time series model. Simple and partial autocorrelation functions of the time series are the most widely used tools for this purpose with scalar time series. Nevertheless, there is a lack of an established method for the identification of functional time series (FTS) models. Functional versions of the autocorrelation and partial autocorrelation functions for FTS based on the L^2 norm of the lagged autocovariance operators of the series are proposed. Diagnostic plots of these functions coupled with prediction bounds derived from large sample results for the autocorrelation and partial autocorrelation functions estimated from a strong functional white noise series are proposed as fast and efficient tools for selecting the order and assessing the adequacy of functional SARMAX models. These methods are studied in numerical simulations with both white noise and serially correlated functional processes, which show that the structure of the processes can be well identified using the proposed techniques. The applicability of the method is illustrated with two real-world datasets: Eurodollar futures contracts and electricity price profiles. The content of this Chapter is a version of the journal article Mestre, Portela, Rice, et al. (2021).

3.1. Introduction

Due to recent technological advances in science and industry, the collection and analysis of high-dimensional data has become an important topic in the field of applied statistics. Many of these observations arise from continuous processes observed, for example, over time, space or frequency domains. These observations can be naturally expressed as functions, and thus the name functional data is commonly used to such observations. The need for statistical methodologies to explore and analyze the structure of these high-dimensional data has led to the development of the Functional Data Analysis (FDA) framework. The monographs Ramsay and Silverman (2005) and Ramsay, Hooker, et al. (2009) and Kokoszka and Reimherr (2017) provide excellent introductions to the key topics of functional data and its applications.

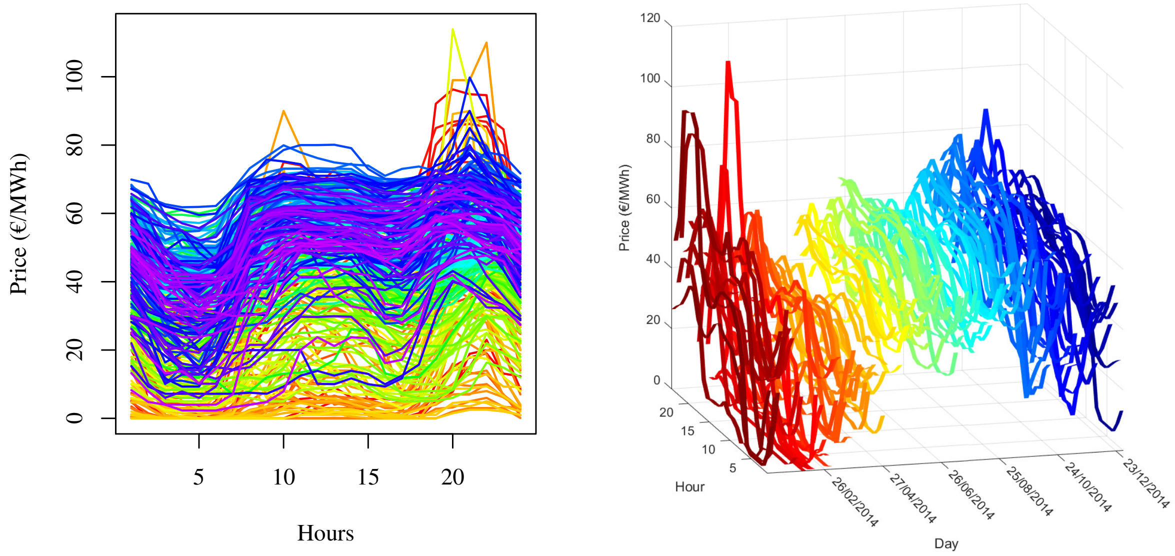

Figure 3.1. Left: Rainbow plot of the FTS of daily electricity price profiles from the Spanish electricity market for the year 2014. Color denotes time, with the oldest curves in red and the most recent data in violet. Right: Rainbow plot with time in a third axis.

If these functional data are collected sequentially over time, it is natural to expect a time dependence between functional observations. This motivates the consideration of functional time series (FTS); see e.g. Bosq (2000) and Mas and Pumo (2010). Due to its wide range of real world applications, there has been growing research interest on the analysis and identification of such functional processes, as presented in Horváth and Kokoszka (2012) where several tests are introduced to measure the serial dependence structure and other properties of FTS.

To present the main ideas, let $\{Y_t(u)\,;\, t = 1, \ldots, T;\; u \in [0, 1]\}$ denote an observed stretch of length T of a functional time series. Here we assume that each observation $Y_t(u)$ is a functional observation defined on $[0, 1]$ whose sample paths are in $L^2([0, 1])$, which is the space of square integrable functions defined on the unit interval. Since typically functional data are only observed at some discrete collection of points, we are assuming that the elements of the FTS under consideration have been obtained after a pre-processing step, such as linear interpolation or basis smoothing; see e.g. Ramsay and Silverman (2005). We could consider instead FTS taking values in more general function spaces, but proceed with $L^2([0, 1])$ in this presentation given the data applications we present below.

As with the majority of scalar and vector-valued time series models, most FTS models can be expressed as

$$Y_t = f(Y_{t-1}, \ldots, Y_{t-p}, \varepsilon_{t-1}, \ldots, \varepsilon_{t-q}) + \varepsilon_t, \tag{3.1}$$

where $\{\varepsilon_t\,;\, t \in \mathbb{N}\}$ denotes a sequence of independent and identically distributed random functional processes with zero mean, and f is an operator of the p past values of the time series and the q past values of the i.i.d innovation processes. Functional models such as the functional autoregressive model (Bosq, 2000; Aue, Norinho, et al., 2015; Martínez-Hernández et al., 2019), the functional ARHX model (Damon and Guillas, 2002; Damon and Guillas, 2005), the functional moving average model (Turbillon et al., 2007), the functional ARMA model (Klepsch et al., 2017), the SARMAHX model (Portela, Muñoz San Roque, et al., 2018; Mestre, Portela,

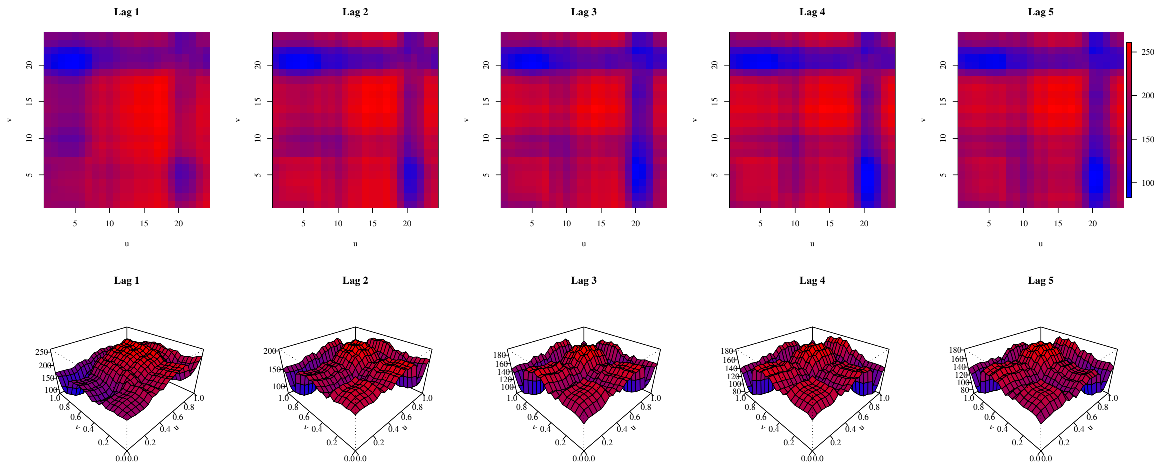

Figure 3.2. First 5 lagged autocovariance functions for the FTS of daily electricity price profiles (top) and a three-dimensional representation of the surfaces (bottom). These functions provide detailed, though perhaps difficult to interpret, information on covariance structure between curves Y_t and Y_{t-h}.

Muñoz San Roque, et al., 2020) and the functional ARCH model (Hörmann, Horváth, et al., 2013) all follow the general equation (3.1).

Identifying relevant features of the time series such as serial correlation, trend or seasonality is of utmost importance when identifying a time series model. Several methods have been proposed for revealing these attributes. Graphical tools can be found in Hyndman and Shang (2010), where the rainbow plot was proposed for visualizing large time indexed functional data sets and performing outlier detection. Curves that exhibit an odd shape or magnitude can be identified with this method. In addition, it provides some information about the trend of the series. Other tools can be found in Canale and Vantini (2016), where the authors use visual representations of the autocorrelation functions of the curves to obtain insights about the dependence across functions observed at different times.

For example, Figures 3.1 and 3.2 illustrate these tools for a FTS composed of daily electricity price profiles from the Spanish electricity market in 2014 (Portela, Muñoz San Roque, et al., 2018). In particular, Figure 3.1 shows the rainbow plot for the price profiles time series. Increases or changes in trend of the daily price profiles are clearly visible in the summer months, as well as in January (red curves). It is however difficult to decipher from this plot the nature of the serial dependence between the curves. Figure 3.2 illustrates the autocovariance surfaces, or kernels for different time lags; see Equation 3.5 for a formal definition. It can be seen that the electricity prices for hours 12 to 18 are correlated with past curves, and that hours 19 to 22 are less influenced by the former price curves. These graphical tools provide some insights on the trend and outliers of the FTS, but they are not able to describe the serial dependence between the curves.

Moreover, once a model of the form (3.1) has been fitted to a given FTS, the plausibility of the main assumptions made by the model should be investigated. In particular, the residuals of the model should behave as independent, identically distributed curves. Several tests have been proposed in order to check the validity of the serial independence assumption for functional data. A Functional Principal Component (FPCA)-based Portmanteau test is proposed in Gabrys and Kokoszka (2007). In Gabrys, Horváth, et al. (2010), the authors project the residuals of the functional linear model on an "optimal" finite dimensional subspace to test for model goodness-of-fit. The white noise test proposed in Horváth, Hušková, et al. (2013) is based on the sum of the L^2 norms of the autocovariance operators of the FTS, rather than on summaries of the FPCA projections. Kokoszka, Rice, et al. (2017) also develop portmanteau tests and inferential procedures for the autocovariance operators of a FTS that are valid in the presence of conditional heteroscedasticity. In addition, frequency-domain tests have been proposed recently in Zhang (2016) and Bagchi et al. (2018) and Characiejus and Rice (2020), where the authors use the distance between the spectral density operator of the series and its best L^2-approximation by a spectral density operator corresponding to a white noise process as a measure of the deviation of the time series from a white noise process.

The host of aforementioned tests would all benefit from the subsequent use of methods that provide insightful information about the underlying dependence structure of the FTS. In the event that these tests suggest that the model errors do not plausibly form a white noise, one often hopes to identify specific lags at which substantial correlation exists in order to fit an improved model to the data. Classical identification methods for scalar time series, such as the Box–Jenkins methodology (Box et al., 2008), use the sample autocorrelation and partial autocorrelation functions of the time series in order to simultaneously test the white noise assumption and to identify the underlying correlation structure of the time series. Identifying the rate of decay of the autocorrelation and specific lags with strong autocorrelation can be used to select the appropriate autoregressive and moving average orders in an ARIMA time series model.

While these kinds of identification methods have been well addressed in the literature of scalar time series, the same cannot be said for FTS. An extension of the autocorrelation function for FTS was proposed in Kokoszka, Rice, et al. (2017) for the purpose of quantifying conditional heteroscedasticity in functional data, but its use in FTS model selection has not been explored. Further, to our knowledge, the notion of partial autocorrelation with functional data has not been explored to date, nor how such summary information could be used in FTS model building.

This chapter proposes methods based on the L^2 norms of lagged covariance operators in order to quantify the autocorrelation structure of FTS. A functional version of the autocorrelation function is studied in the context of time series model identification. In addition, a generalization of the partial autocorrelation function is proposed, which allows the practitioner to identify functional autoregressive and moving average processes of higher order. We show how these methods can be used in order to identify autocorrelation and seasonality in FTS, which are useful in model selection within the functional SARMAHX family of models introduced in Portela, Muñoz San Roque, et al. (2018) and Mestre, Portela, Muñoz San Roque, et al. (2020). These models generalize the aforementioned functional linear models, allowing the inclusion of both autoregressive and moving average effects as well as incorporating the seasonal behavior present in the data. As such, identifying the serial correlation structure of the data is of utmost importance to fitting the SARMAHX model.

This chapter is organized as follows: Section 3.2 presents the main contributions and the practical implementation of the proposed identification procedure. Section 3.3 contains the results of a Monte-Carlo simulation study of the proposed methodology. Demonstrations and applications of these model identification procedures are presented in Section 3.4, where two real-world datasets are explored. Section 3.5 gives some concluding remarks, summarizes the results of the methodology developed in the previous sections, and points to some directions for future research. Proofs for the main results presented in this chapters can be found in the appendix of the journal article Mestre, Portela, Rice, et al. (2021).

3.2. Functional autocorrelation measures

We begin by introducing some notation that will be used throughout this chapter and state the main assumptions of the proposed methodology. Throughout this dissertation, $L^2([0,1]^d)$ denotes the Hilbert space of real valued square integrable functions defined on $[0,1]^d$ with the usual inner product $\langle .,. \rangle$ and the norm $\|.\|$ it generates. The dimension is made clear based on the domain of the input function.

3.2.1. Autocorrelation function for FTS

Here an FTS $\{Y_t\,;\, t = 1, \ldots, T\}$ is assumed to be a realization of length T of a given functional stochastic process $\{\mathcal{Y}_t\,;\, t \in \mathbb{N}\}$, where each random variable $\mathcal{Y}_t$ is a square integrable function $\{\mathcal{Y}_t(v)\,;\, v \in [0,1]\}$. For the rest of the chapter, it will be assumed that all the FTS are stationary. The main interest of this section is to develop a statistical tool that can be used to identify the underlying serial dependence structure of a given time series. In order to achieve this objective, the first step is to test the white noise assumption on a given sequence of functional observations. The definition of functional white noise was presented in Section 2.3.1, as an functional stochastic process with the following properties:

Definition 3.2.1. The functional stochastic process $\{\varepsilon_t(v)\,;\, t \in \mathbb{N}; v \in [0,1]\}$ is a (strong) functional white noise process if the following assumptions are satisfied:

- **A1**: The functional random variables ε_t are independent and identically distributed.
- **A2**: $E\left[\varepsilon_i(v)\right] = 0$ for all $i = 1, \ldots, T$.

The first assumption implies that such a white noise series does not exhibit any degree of serial dependence, differentiating such a process from others that exhibit serial dependence, such as the functional linear process (Bosq, 2000).

Throughout this dissertation, it will be assumed that all the functional random variables have finite fourth order moments, which is to say that $E\left[\|\mathcal{Y}_t\|^4\right] < \infty$. This implies that the (lag zero) autocovariance kernel $\mathrm{c}_0(u,v)$, defined as

$$\mathrm{c}_0(u,v) = E\left[\left(\mathcal{Y}_1(u) - E\left[\mathcal{Y}_1(u)\right]\right)\left(\mathcal{Y}_1(v) - E\left[\mathcal{Y}_1(v)\right]\right)\right], \tag{3.2}$$

is square integrable. Hence, there are real positive eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots$ and orthonormal eigenfunctions $\phi_1, \phi_2, \ldots$ satisfying

$$\lambda_i \phi_i(u) = \int \mathrm{c}_0(u,s)\phi_i(s)ds, \qquad i = 1, 2, \ldots \tag{3.3}$$

The dependence structure of a given stationary FTS can be analyzed via its lagged autocovariance kernels, defined as

$$\mathrm{c}_h(u,v) = E\left[\left(\mathcal{Y}_1(u) - E\left[\mathcal{Y}_1(u)\right]\right)\left(\mathcal{Y}_{1+h}(v) - E\left[\mathcal{Y}_{1+h}(v)\right]\right)\right], \tag{3.4}$$

where $h = 1, \ldots, T-1$ is a given lag parameter. Indeed fourth order moments are more than what is required here to make the above quantities well defined, two would suffice, although we require fourth order moments in order to establish large sample Gaussian approximations for autocovariance operator estimates presented below.

Given $Y_1, Y_2, \ldots, Y_T(v)$ a realization of the FTS, the autocovariance kernel $\mathrm{c}_h(u, v)$ can be estimated via the empirical lagged autocovariance kernels:

$$\widehat{\mathrm{c}}_h(u, v) = \frac{1}{T}\sum_{i=1}^{T-h}(Y_i(u) - \widehat{\mu}(u))(Y_{i+h}(v) - \widehat{\mu}(v)), \tag{3.5}$$

where

$$\widehat{\mu}(u) = \frac{1}{T}\sum_{i=1}^{T} Y_i(t) \tag{3.6}$$

denotes the sample mean function. Using right integration as in Equation 3.3, these kernels may be used to define autocovariance and empirical autocovariance operators, and so the terms kernels and operators will be used interchangeably in this setting.

Due to the fact that a functional white noise process exhibits no serial correlation between its terms, the norm of its lagged autocovariance operators should be close to zero for each positive lag h. Horváth, Hušková, et al. (2013) for example propose to use the sums of the L^2-norms of the lagged autocovariance operators up to a user specified maximum lag H in order to obtain a white noise test for functional data.

However, such a test does not provide the practitioner with specific information about the lagged dependence structure of the FTS. This information may be summarized simply by considering the magnitudes and plots of the empirical covariance operators as a function of the lag. Evaluating whether the magnitudes $\|\widehat{\mathrm{c}}_h\|$ are consistent with the series following a white noise process, and further identifying specific lags where strong correlation is present, is facilitated by comparing $\|\widehat{\mathrm{c}}_h\|$ to an estimate of its distribution assuming the data are indeed a white noise. The following result describes the large sample behavior of $\|\widehat{\mathrm{c}}_h\|$ under assuming the data follow a white noise; see Kokoszka, Rice, et al. (2017) for a proof.

Theorem 3.2.1. *If the FTS $\{Y_t\}_{t=1}^{T}$ is independent and identically distributed with $E\left[\|\mathcal{Y}_t\|^4\right] < \infty$, then*

$$\widehat{Q}_{T,h} = \iint T\,\widehat{\mathrm{c}}_h^2(u, v)dudv \xrightarrow{D} Q = \sum_{j=1}^{\infty}\sum_{l=1}^{\infty}\lambda_j\lambda_l\chi_{j,l}^2(1), \qquad h = 1, 2, \ldots \tag{3.7}$$

as $T \to \infty$. The λ_i's are the eigenvalues of the covariance operator of Y_t and $\{\chi_{j,l}^2(1)\ ;\ j, l \in \mathbb{N}\}$ are independent random variables following a chi-squared distribution $\chi^2(1)$.

Following Kokoszka, Rice, et al. (2017), the functional autocorrelation coefficient at lag h can be defined as

$$\rho_h = \frac{\|\mathrm{c}_h\|}{\int \mathrm{c}_0(u, u)du}\ ,\ \|\mathrm{c}_h\|^2 = \iint \mathrm{c}_h^2(u, v)dudv. \tag{3.8}$$

It follows from the Cauchy–Schwarz inequality applied pointwise to $\mathrm{c}_h(u, v)$ that $0 \leq \rho_h \leq 1$, and gives a scale free measure of the serial correlation in the FTS at lag h. One could also define ρ_h as $\|\mathrm{c}_h\|/\|\mathrm{c}_0\|$, which would also give a scale free measure of the autocorrelation structure of a given FTS. Both definitions are equally useful in identifying the lagged dependence structure encoded by $\|\mathrm{c}_h\|$. Henceforth, in this paper the definition in Equation 3.8 is used, as this definition is more easily adapted to a functional version of a partial autocorrelation function, which we pursue below.

The functional autocorrelation at lag h ρ_h admits the plug in estimate

$$\widehat{\rho}_h = \frac{\|\widehat{\mathrm{c}}_h\|}{\int \widehat{\mathrm{c}}_0(u,u)du} = \frac{\sqrt{Q_{T,h}}}{\sqrt{T}\int \widehat{\mathrm{c}}_0(u,u)du}. \tag{3.9}$$

Based on Theorem 3.2.1, a procedure to produce a prediction bound for $\widehat{\rho}_h$ under the assumption that the data are a white noise can be summarized as:

- **Step 1**: Estimate $\widehat{\lambda}_1, \dots, \widehat{\lambda}_d$ from

$$\widehat{\mathrm{c}}_0(u,v) = \frac{1}{T}\sum_{i=1}^{T}(Y_i(u) - \widehat{\mu}(u))(Y_i(v) - \widehat{\mu}(v)) \tag{3.10}$$

  for a fixed value of d selected by the practitioner, so that

$$\widehat{\lambda}_i\widehat{\phi}_i(u) = \int \widehat{\mathrm{c}}_0(u,s)\widehat{\phi}_i(s)ds\,. \tag{3.11}$$

- **Step 2**: Under the assumption of functional white noise, Theorem 3.2.1 provides an approximate distribution for the sequence $\{\iint T\,\widehat{\mathrm{c}}_h^2(t,s)dtds\,;\, h = 1,\dots,H\}$. This distribution can be estimated as

$$\widehat{Q}(d) = \sum_{j=1}^{d}\sum_{l=1}^{d}\widehat{\lambda}_j\widehat{\lambda}_l\chi^2_{j,l}(1), \tag{3.12}$$

  where d is a large number (below we select d as the number of eigenvalues $\widehat{\lambda}_i$ of $\widehat{\mathrm{c}}_0(u,v)$ such that $\widehat{\lambda}_i/\widehat{\lambda}_1 > 0.0001$. This threshold was observed to give good results in the different empirical tests performed). This distribution is a linear combination of independent chi-squared random variables, and in practice it must be estimated. The method proposed in Imhof (1961) provides a suitable method to do so, providing good approximations in the tails of the distribution.

- **Step 3**: For a given confidence level $\alpha \in (0,1)$, and under the white noise assumption, an upper prediction bound for the sequence $\{\widehat{\rho}_h\,;\, h = 1,\dots,H\}$ can be set at

$$\frac{\sqrt{\widehat{Q}(d)_{(1-\alpha)}}}{\sqrt{T}\int \widehat{\mathrm{c}}_0(u,u)du}, \tag{3.13}$$

  where $\widehat{Q}(d)_{(1-\alpha)}$ denotes the $(1-\alpha)$ critical value of the distribution of $\widehat{Q}(d)$.

For small α, significant deviations of $\{\widehat{\rho}_h\,;\, h = 1,\dots,H\}$ beyond the boundary given by Equation 3.13 can be used to identify structural dependence in the FTS, and also for rejecting the white noise hypothesis at a given lag.

3.2.2. Partial autocorrelation function for FTS

In classical time series analysis, the autocorrelation function is often accompanied by the partial autocorrelation function, which measures the correlation between two lags of the time series after removing the influence of the intermediate terms. By plotting these two functions together, autoregressive and moving average processes of any order can be identified by comparing them to the theoretical behavior of these plots when the order of the process is known. This section is devoted to developing the notion of a partial autocorrelation function for functional data, which can be used in conjunction with the proposed FACF to identify the correlation structure of a given FTS. There are a number of challenges in extending the notion of partial autocorrelation to a FTS $\{Y_t \,:\, t = 1, \ldots, T\}$, although here we propose and establish some of the large sample properties of such a quantity that is calculated analogously to the functional autocorrelation function defined in Section 3.2.1. It is similarly simple to interpret and calculate, and is useful in FTS model selection and diagnostic checking.

In plain language, the partial autocovariance between Y_t and Y_{t+h} is the autocovariance of the residuals of Y_t and Y_{t+h} produced by "regressing" them linearly on the intermediate variables $Y_{t+1}, \ldots, Y_{t+h-1}$. Towards a more formal definition of such a regression operation, first let $\boldsymbol{Y}_{t,h} = (Y_{t+1}, \ldots, Y_{t+h-1})$ denote the collection of these intermediary variables. Linear regression in this context can be carried out by selecting or estimating appropriate linear operators that relate $\boldsymbol{Y}_{t,h}$ to both Y_t and Y_{t+h}. Let $\mathcal{K}_h$ denote the collection of Hilbert–Schmidt kernel integral (linear) operators mapping $(L^2([0,1]))^{\otimes h}$ to $L^2([0,1])$. More precisely, if $\boldsymbol{\Psi}_h \in \mathcal{K}_h$, and if $\boldsymbol{f}_h \in (L^2([0,1]))^{\otimes h}$ with $\boldsymbol{f}_h$ having component functions $\boldsymbol{f}_h = (f_{1,h}, \ldots, f_{h,h})$, then for almost every $s \in [0,1]$

$$\boldsymbol{\Psi}_h(\boldsymbol{f}_h)(s) = \sum_{j=1}^{h} \int \psi_h^{(j)}(v,s) f_{j,h}(v) dv, \tag{3.14}$$

for some component kernel functions $\psi_h^{(j)},\ j = 1, \ldots, h-1$ satisfying $\|\psi_h^{(j)}\| < \infty$. The collection of such operators may be made into its own Hilbert space endowed with the Hilbert–Schmidt norm

$$\|\boldsymbol{\Psi}_h\|_{HS} = \left(\sum_{j=1}^{h} \|\psi_h^{(j)}\|^2 \right)^{1/2}. \tag{3.15}$$

For example, $\{Y_t \,:\, t \in \mathbb{N}\}$ follows a functional autoregressive model of order p if there exists an operator $\boldsymbol{\Psi}_p \in \mathcal{K}_p$ and an innovation sequence $\{\varepsilon_t \,:\, t \in \mathbb{N}\}$ so that

$$Y_t(u) = \boldsymbol{\Psi}_p(\boldsymbol{Y}_{t,p+1})(u) + \varepsilon_t(u) = \sum_{j=1}^{p} \int \psi^{(j)}(v,u) Y_{t-j}(v) dv + \varepsilon_t(u). \tag{3.16}$$

In this case the kernels $\psi^{(j)}\ j = 1, \ldots, p$ are referred to as the autoregressive kernels. Now let

$$Proj(\mathcal{X}, \boldsymbol{Y}_{t,h}) = \boldsymbol{\Psi}(\boldsymbol{Y}_{t,h}), \tag{3.17}$$

where $\boldsymbol{\Psi}$ satisfies

$$E\left[\|\mathcal{X} - \boldsymbol{\Psi}(\boldsymbol{Y}_{t,h})\|^2\right] = \inf_{\boldsymbol{\Psi}' \in \mathcal{K}_{h-1}} E\left[\|\mathcal{X} - \Psi'(\boldsymbol{Y}_{t,h})\|^2\right]. \tag{3.18}$$

Let $\mathcal{A}_t = \mathcal{Y}_t - Proj(\mathcal{Y}_t, \boldsymbol{Y}_{t,h})$ and $\mathcal{B}_t = \mathcal{Y}_{t+h} - Proj(\mathcal{Y}_t, \boldsymbol{Y}_{t,h})$. With this defined an intuitive definition of the partial autocovariance kernel at lag h of the series is

$$\mathsf{c}_{(h,h)}(u,v) = E\left[\left(\mathcal{A}_t(u) - E\left[\mathcal{A}_t(u)\right]\right)\left(\mathcal{B}_t(v) - E\left[\mathcal{B}_t(v)\right]\right)\right]. \tag{3.19}$$

The functional partial autocorrelation function (FPACF) may then be defined as

$$\rho_{(h,h)} = \frac{\|\mathrm{c}_{(h,h)}\|}{\gamma_{1,h}\gamma_{2,h}}, \tag{3.20}$$

where

$$\gamma_{1,h} = \left(\int E\left[\mathcal{A}_t(u)\right]^2 du\right)^{1/2}, \tag{3.21}$$

and

$$\gamma_{2,h} = \left(\int E\left[\mathcal{B}_t(u)\right]^2 du\right)^{1/2}. \tag{3.22}$$

It follows by the Cauchy–Schwarz inequality that $0 \leq \rho_{(h,h)} \leq 1$.

Although this may serve as a logical counterpart to the scalar PACF, a number of theoretical and practical challenges are faced here stemming from the fact that estimating the projections of Y_t and Y_{t+h} onto $\boldsymbol{Y}_{t,h}$, or equivalently estimating the operators $\boldsymbol{\Psi}_1$ and $\boldsymbol{\Psi}_2$ so that

$$E\left[\|\mathcal{Y}_t - \boldsymbol{\Psi}_{1,h}(\boldsymbol{Y}_{t,h})\|^2\right] = \inf_{\boldsymbol{\Psi}'\in\mathcal{K}_{h-1}} E\left[\|\mathcal{Y}_t - \Psi'(\boldsymbol{Y}_{t,h})\|^2\right], \tag{3.23}$$

and

$$E\left[\|\mathcal{Y}_{t+h} - \boldsymbol{\Psi}_{2,h}(\boldsymbol{Y}_{t,h})\|^2\right] = \inf_{\boldsymbol{\Psi}'\in\mathcal{K}_{h-1}} E\left[\|\mathcal{Y}_{t+h} - \Psi'(\boldsymbol{Y}_{t,h})\|^2\right], \tag{3.24}$$

in general presents an ill-posed inverse problem. Nonetheless, these operators may still be estimated in practice using standard regularization techniques, see e.g. Chapter 9 of Bosq (2000). For any fixed h, let $\widehat{\boldsymbol{\Psi}}_{1,h}$ denote the operator obtained by performing a regularized functional linear regression of Y_t on $\boldsymbol{Y}_{t,h}$, and let $\widehat{\boldsymbol{\Psi}}_{2,h}$ denote the operator obtained by performing a regularized functional linear regression of Y_{t+h} on $\boldsymbol{Y}_{t,h}$. We note that in the second case this may be viewed as fitting a functional autoregressive model of order $h-1$ to the series. Define

$$\widehat{\mathrm{c}}_{(h,h)}(u,v) = \frac{1}{T-h}\sum_{t=1}^{T-h}[Y_t(u) - \widehat{\Psi}_{h,1}(\boldsymbol{Y}_{t,h})(u)][Y_{t+h}(v) - \widehat{\Psi}_{h,2}(\boldsymbol{Y}_{t,h})(v)], \tag{3.25}$$

$$\widehat{\gamma}_{1,h} = \left(\int \frac{1}{T-h}\sum_{t=1}^{T-h}[Y_t(u) - \widehat{\Psi}_{h,1}(\boldsymbol{Y}_{t,h})(u)]^2 du\right)^{1/2}, \tag{3.26}$$

and

$$\widehat{\gamma}_{2,h} = \left(\int \frac{1}{T-h}\sum_{t=1}^{T-h}[Y_{t+h}(u) - \widehat{\Psi}_{h,2}(\boldsymbol{Y}_{t,h})(u)]^2 du\right)^{1/2}. \tag{3.27}$$

Then, the empirical partial autocorrelation may then be defined as

$$\widehat{\rho}_{(h,h)} = \frac{\|\widehat{\mathrm{c}}_{(h,h)}\|}{\widehat{\gamma}_{1,h}\widehat{\gamma}_{2,h}}. \tag{3.28}$$

It may be shown that under mild conditions on the estimators $\widehat{\boldsymbol{\Psi}}_{1,h}$ and $\widehat{\boldsymbol{\Psi}}_{2,h}$ that $\widehat{\rho}_{(h,h)}$ has the same large sample properties as $\widehat{\rho}_h$ when the data are a white noise, as described in the following theorem.

Theorem 3.2.2. *If $\{Y_t \ : \ t \in \mathbb{N}\}$ is independent and identically distributed with $E\left[\|\mathcal{Y}_t\|^4\right] < \infty$, $\|\widehat{\Psi}_{h,1}\|_{HS} = o_P(T^{-1/4})$, and $\|\widehat{\Psi}_{h,2}\|_{HS} = o_P(T^{-1/4})$, then*

$$\widehat{Q}_{h,h} = T\|\widehat{c}_{(h,h)}\|^2 \xrightarrow{D} Q = \sum_{j=1}^{\infty}\sum_{l=1}^{\infty} \lambda_j \lambda_l \chi^2_{j,l}(1), \qquad h = 1, 2, \ldots \tag{3.29}$$

as $T \to \infty$. The $\lambda_i's$ are the eigenvalues of the covariance operator of Y_i. Additionally

$$\widehat{\gamma}_{1,h} \xrightarrow{P} \left[\int \widehat{c}_0(u,u)du\right]^{1/2}, \text{ and } \widehat{\gamma}_{2,h} \xrightarrow{P} \left[\int \widehat{c}_0(u,u)du\right]^{1/2}. \tag{3.30}$$

Remark 3.2.1. Most standard regularized estimators in this setting would satisfy $\|\widehat{\mathbf{\Psi}}_{h,1}\|_{HS} = o_P(T^{-1/4})$, and $\|\widehat{\mathbf{\Psi}}_{h,2}\|_{HS} = o_P(T^{-1/4})$ under mild conditions, including those derived from principal component regression and Tychonoff regularization. We elaborate on the PCA case, see e.g. Didericksen et al. (2012) and Bosq (2000) and Zhang (2016). In case of Tychonoff regularization we refer to Kargin and Onatski (2008). In the case of PCA regression $\widehat{\mathbf{\Psi}}_{h,2}$, and similarly $\widehat{\mathbf{\Psi}}_{h,1}$, is estimated so that

$$\widehat{\mathbf{\Psi}}_{h,2}(\boldsymbol{f}_h)(s) = \sum_{j=1}^{h-1} \int \widehat{\psi}^{(j)}_{h,2}(v,s) f_{j,h}(v) dv. \tag{3.31}$$

The kernels $\widehat{\psi}^{(j)}_h$ are constructed so that

$$\widehat{\psi}^{(j)}_{h,2}(v,s) = \sum_{r,\ell=1}^{\kappa} \widehat{\psi}^{(j)}_{h,2}[r,\ell]\widehat{\phi}_r(v)\widehat{\phi}_\ell(s), \tag{3.32}$$

and the coefficient matrix $\widehat{\psi}^{(j)}_{h,2}[\cdot,\cdot]$ is estimated to minimize the least squares criterion (3.24). For example, when $h = 2$, so that regression is performed on a single intermediary variable, this amounts to taking

$$\widehat{\psi}^{(j)}_{h,2}[r,\ell] = \frac{\widehat{\lambda}^{-1}_\ell}{T-j} \sum_{t=1}^{T-j} \langle Y_t, \widehat{\phi}_\ell\rangle\langle Y_{t+j}, \widehat{\phi}_r\rangle, \tag{3.33}$$

where the $\widehat{\lambda}'_\ell s$ and $\widehat{\phi}'_\ell s$ are the empirical eigenvalues and eigenvectors of the covariance operator defined in Equation 3.5. κ in this case serves as a truncation level for the number of principal components employed. When κ is a fixed integer, then under the assumption that the data follow a strong white noise and standard eigenvalue spacing conditions $\|\widehat{\mathbf{\Psi}}_{h,2}\|_{HS} = O_P(T^{-1/2})$, and even satisfies the Central Limit Theorem when normalized by multiplying by $T^{1/2}$. In order that the regression operator would be consistently estimated assuming that the underlying subspace on which the functional data take their values in $L^2([0,1])$ is infinite dimensional, it is often allowed that $\kappa = \kappa_T$ is a function of the sample size that is allowed to increase slowly as T increases. In this case under sufficient decay conditions on the eigenvalues of the covariance operator defined in Equation 3.5, it still holds that $\|\widehat{\mathbf{\Psi}}_{h,2}\|_{HS} = o_P(T^{-1/4})$, see Chapter 9 of Bosq (2000) and Zhang (2016). Below all such regressions are carried out using FPCA where κ is chosen to explain more than 95% of the total variation of the data.

Theorem 3.2.2 shows that under the hypothesis that the series of interest is a strong white noise having fourth order moments, then the exact same prediction intervals for $\{\widehat{\rho}_h, \quad h =$

$1, \ldots, H\}$ are asymptotically valid for $\{\widehat{\rho}_{(h,h)}, \quad h = 1, \ldots, H\}$. Hence comparing $\{\widehat{\rho}_{(h,h)}, \quad h = 1, \ldots, H\}$ to the threshold

$$\frac{\sqrt{\hat{Q}(d)_{(1-\alpha)}}}{\sqrt{T}\int \widehat{\mathfrak{c}}_{(h,h)}(u,u)du} \tag{3.34}$$

for a user specified confidence level $(1-\alpha)$ provides a further way of evaluating the whiteness of the series or residuals under study.

Similarly to the standard scalar PACF, the functional PACF (FPACF) will tend to zero at lags larger than the autoregressive order when applied to consistently estimated ARH(p) models. In order to make this statement precise, we assume that $\{Y_t \ : \ t \in \mathbb{N}\}$ follows an ARH(p) model

$$Y_t(u) = \mathbf{\Psi}_p^{ARH}(\mathbf{Y}_{t,p+1})(u) + \varepsilon_t(u) = \sum_{j=1}^{p} \int \psi_p^{ARH,j}(v,u)Y_{t-j}(v)dv + \varepsilon_t(u), \tag{3.35}$$

for some $\mathbf{\Psi}_p^{ARH} \in \mathcal{K}_p$ with components $\psi_p^{ARH,j}$. So that we may discuss the approximation of this operator as performed in evaluating the FPACF coefficient at lags larger than p, let $\mathbf{\Psi}_{h,p}^{ARH}$ denote the operator in $\mathcal{K}_h$ with component kernels $\psi_h^{ARH,j}$ satisfying

$$\psi_h^{ARH,j} = \begin{cases} \psi_p^{ARH,j} & j \leq p, \\ 0, & j > p. \end{cases} \tag{3.36}$$

Theorem 3.2.3. *Suppose that* $\{Y_t \ : \ t \in \mathbb{N}\}$ *satisfies the ARH(p) model in* (3.36) *with* $0 < E\left[\|\varepsilon_0\|^4\right] < \infty$, *and further suppose that the model admits a stationary and causal solution. If* $h > p$, $\|\widehat{\mathbf{\Psi}}_{2,h} - \mathbf{\Psi}_{h,p}^{ARH}\|_{HS} = o_P(1)$, *and there exists an operator* $\mathbf{\Psi}_{1,h} \in \mathcal{K}_{h-1}$ *so that* $\|\widehat{\mathbf{\Psi}}_{1,h} - \mathbf{\Psi}_{1,h}\|_{HS} = o_P(1)$, *then* $\widehat{\rho}_{(h,h)} \xrightarrow{P} 0$ *as* $T \to \infty$.

Theorem 3.2.3 implies that asymptotically the FPACF coefficients cutoff after lag p when estimated based on data following an ARH(p) model, which aligns with the traditional properties of the PACF. Necessary and sufficient conditions for the existence of a stationary and causal solution to a ARH(p) model are discussed in Chapter 5 of Bosq (2000).

3.3. Simulation study

Methods for estimating the proposed autocorrelation and partial autocorrelation functions for functional data have been implemented in the R package `fdaACF`, available in CRAN (`http://cran.r-project.org/package=fdaACF`), which will be used in the following simulation study.

3.3.1. Application to strong white noise processes

The FACF and FPACF introduced above can be used to evaluate the whiteness of a given FTS. In order to validate the proposed method, we consider several synthetic functional white noise processes. This section investigates the finite-sample properties of the method when applied to such processes.

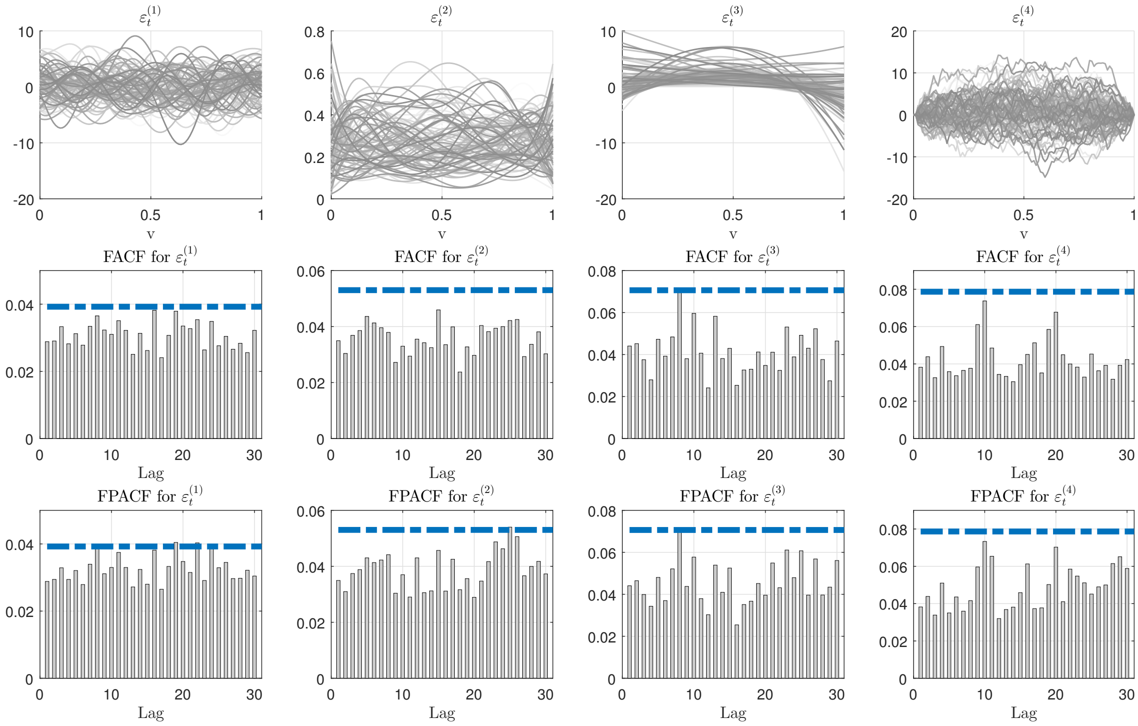

Figure 3.3. Simulated white noise processes and FACF/FPACF of each series. The dotted line denotes the 99% critical value of the distribution obtained by the proposed method.

Each functional observation has been discretized in a grid of 100 equi-spaced points in the interval $[0, 1]$. Four different white noise processes have been simulated, and the methods employed to generate each process are described below.

The first 3 series were generated by selecting a functional basis $\phi_k(v)$ and i.i.d. coefficients $b_{t,k}$, thus obtaining the elements of the process as

$$\varepsilon_t(v) = \sum_{k=1}^{K} b_{t,k}\phi_k(v) \,, \quad t = 1, \dots, T. \tag{3.37}$$

As the coefficients of the series are i.i.d with respect to t, the FTS should not exhibit any kind of serial correlation. By varying the functional basis ϕ_k and the distributions of the coefficients $b_{t,k}$, different processes have been generated:

- $\varepsilon_t^{(1)}$ is obtained using the first 7 elements of the Fourier basis, due to the fact that this basis is often used when dealing with periodic functional data. The coefficients $b_{t,k}$ follow a standard normal distribution $N(0, 1)$.

- $\varepsilon_t^{(2)}$ is obtained using the first 7 elements of the B-spline basis, due to the fact that this basis is often used when dealing with non-periodic functional data. The coefficients are drawn from a Beta$(2, 5)$ distribution.

- $\varepsilon_t^{(3)}$ is obtained selecting 3 functions and applying the Gram–Schmidt method to obtain 3 orthonormal functions that will be used as a basis in Equation 3.37. In this case, the functions selected were $\theta_1(v) = \sin(v)$, $\theta_2(v) = \exp v$ and $\theta_3(v) = \cos(v)$. In this case, the coefficients of Equation 3.37 follow a Exp(1) distribution.

The last white noise process has been obtained simulating a brownian bridge random process, where each innovation function is defined as

$$\varepsilon_t(v) = W_t(v) - vW_t(1), \tag{3.38}$$

where $W_t(v)$ is a Wiener processes defined on the domain $[0, 1]$.

- $\varepsilon_t^{(4)}$ is generated by a brownian bridge process where the variance of the Wiener process is 1.

Figure 3.3 shows a representative example of the first 50 values of the FACF of each white noise process, including the 99% prediction bound provided by Theorem 3.2.1 that has been obtained using the `obtain_FACF` function of the `R` package `fdaACF`. The value of d was selected as the number of eigenvalues $\widehat{\lambda}_i$ of $\widehat{c}_0(u, v)$ such that $\hat{\lambda}_i/\hat{\lambda}_1 > 0.0001$. Evidently most of the coefficients of the FACF fall below this limit, which is consistent with the white noise hypothesis.

In order to test the empirical properties of the proposed statistic for different confidence values, all 4 white noise series described in this section have been simulated, using different configurations of the length of the time series $(T = 100, 250, 500, 1000, 2000)$ and using different lags when obtaining the limit value for the lagged autocorrelations. 50 repetitions of each configuration has been obtained, and Table 3.1 shows the percentage of out-of-bounds autocorrelations obtained in the previous analysis for the confidence levels $\alpha = 0.1, 0.05, 0.01$. Although it has been observed that the empirical out-of-bound rate seems to be quite conservative for significance level 10% when increasing the number of lags considered, the overall performance of the test agrees with the selected confidence level α. Similar results are obtained when using the FPACF instead of the FACF, and so we omit those results due to space considerations.

3.3.2. Application to serially correlated processes

This section present several simulations of functional dependent processes, each one exhibiting some kind of serial correlation. The proposed identification method based on the FACF and the FPACF will be used in order to identify the underlying structure of the simulated FTS; fit a model using that information and check that the residuals of the fitted model plausibly correspond to a white noise process, validating the model.

A crucial step in the modeling of scalar time series is the identification of the structure of the underlying stochastic process. The Box-Jenkins methodology (Box et al., 2008) proposes a standard model identification procedure for scalar time series. Once the stationarity and seasonality of the series has been addressed, the next step is to identify the seasonal and regular order of the AR and MA component of the transformed time series. The main tools used in this identification procedure are the autocorrelation and partial autocorrelation functions, whose plots can be used to obtain useful information about the underlying structure of the stochastic process that has generated the series. These tools can also be used in the diagnosis of the model in order to check the white noise assumption of the residuals.

The purpose of this section is to apply the proposed method in order to obtain a similar tool for FTS. Regarding the identification of FTS models, there has been several attempts to obtain the optimum order of some functional processes. In Kokoszka and Reimherr (2013), a test statistic is derived in order to check if a ARH(p) model fits the data better than a ARH($p-1$)

	T	H	Significance level 10%	5%	1%		T	H	Significance level 10%	5%	1%
$\varepsilon_t^{(1)}$	100	10	7.11%	3.65%	0.58%	$\varepsilon_t^{(3)}$	100	10	5.77%	3.46%	1.15%
		20	3.84%	1.73%	0.38%			20	6.92%	2.98%	0.57%
		30	3.01%	1.08%	0.12%			30	5.7%	2.88%	0.51%
	250	10	8.46%	3.07%	0.57%		250	10	9.03%	4.23%	1.15%
		20	6.53%	3.26%	0.48%			20	6.82%	3.65%	0.48%
		30	5.7%	2.75%	0.32%			30	6.98%	3.14%	0.32%
	500	25	7.36%	3.28%	0.88%		500	25	10.48%	5.04%	1.12%
		50	6.88%	2.92%	0.56%			50	8.4%	4.12%	0.88%
		100	3.88%	1.92%	0.12%			100	6.96%	3.36%	0.68%
	1000	25	10.08%	4.88%	0.88%		1000	25	9.76%	4.16%	0.56%
		50	7.84%	3.72%	0.96%			50	8.96%	4.44%	0.8%
		100	6.56%	2.98%	0.56%			100	7.66%	3.8%	0.74%
	2000	25	7.76%	3.84%	0.56%		2000	25	11.52%	5.12%	0.8%
		50	9.28%	4.56%	0.88%			50	9.08%	4.64%	1.04%
		100	8.26%	4.1%	0.68%			100	8.74%	4.24%	0.92%
$\varepsilon_t^{(2)}$	100	10	7.69%	3.27%	0.57%	$\varepsilon_t^{(4)}$	100	10	8.46%	4.42%	0.38%
		20	6.53%	3.56%	0.77%			20	6.82%	2.98%	0.46%
		30	5.06%	2.43%	0.44%			30	6.21%	2.75%	0.51%
	250	10	9.23%	4.42%	0.96%		250	10	10.77%	4.42%	0.76%
		20	9.13%	4.32%	0.86%			20	8.46%	3.94%	0.29%
		30	8.78%	4.23%	0.89%			30	8.84%	4.42%	0.06%
	500	25	9.28%	4.32%	0.4%		500	25	9.2%	5.12%	1.28%
		50	7.72%	3.72%	0.64%			50	7.6%	3.88%	0.72%
		100	6.42%	2.9%	0.4%			100	7.58%	3.38%	0.62%
	1000	25	9.28%	4.56%	1.2%		1000	25	12.08%	5.92%	1.36%
		50	9.2%	4.72%	1.04%			50	9.68%	4.6%	0.92%
		100	8.12%	3.84%	0.6%			100	8.38%	3.64%	0.58%
	2000	25	9.04%	4.32%	1.04%		2000	25	10.4%	6.24%	1.44%
		50	10.08%	4.88%	1.16%			50	9.6%	4.36%	0.78%
		100	8.66%	4.1%	0.8%			100	9.26%	4.58%	0.98%

Table 3.1. Empirical rate of out-of bounds FACF coefficients obtained for several simulated functional white noise processes.

model, by checking if the operator ϕ_p of the formulation of the ARH(p) is significantly different from the null operator. However, up to the authors' knowledge, no general identification procedure analogous to the Box-Jenkins methodology for scalar time series has been proposed for FTS. This section is aimed at showing that the FACF and FPACF proposed in Section 3.2 can be considered as a first step in this direction, offering a useful tool for the identification of certain patterns in the structure of the residuals of a fitted functional model that should be included in the model in order to increase its accuracy.

In order to illustrate the applicability of this method as a diagnosis tool, several functional series have been simulated including some specific patterns of temporal dependence. A functional SARMAHX model (Portela, Muñoz San Roque, et al., 2018; Mestre, Portela, Muñoz San Roque, et al., 2020) will be fitted to each time series, and the method proposed in this paper will be used to test the white noise assumption on the residuals of the fitted model. The SARMAHX model is an extension of the standard seasonal ARMA model to the functional framework by means of functional operators, accounting for autoregressive and moving average effects as well as allowing the inclusion of exogenous variables to the model. Given a stationary and centered FTS $\{Y_t(v)\,;\,t=1,2,...,T\,;\,v\in V\}$ and a set

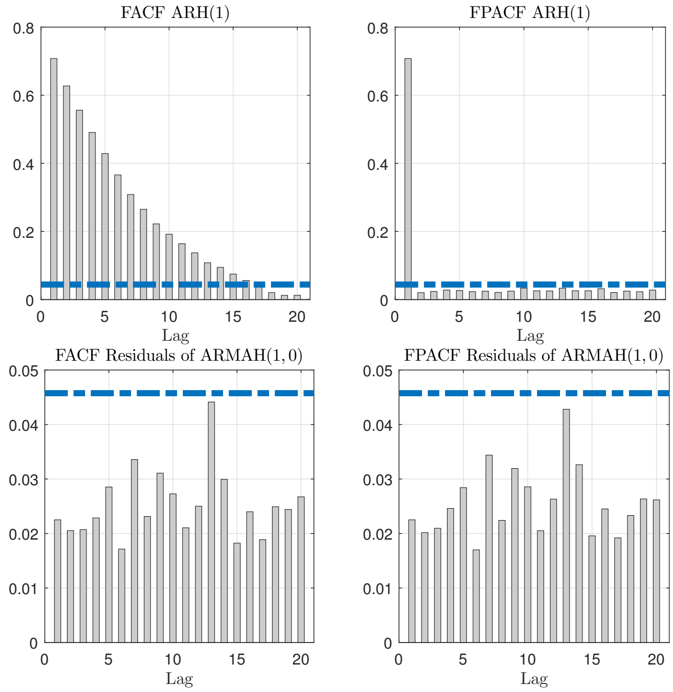

Figure 3.4. Top: FACF and FPACF for a simulated ARH(1) process. Bottom: FACF and FPACF of the residuals of a fitted ARMAH$(1,0)$ model. The dotted line represents the 99% white noise prediction bounds for the FACF/FPACF coefficients under the assumption of white noise.

$\{X_t^z(u_z)\,;\, z \in \mathbb{Z}\,;\, t = 1, 2, \ldots, T\,;\, u_z \in V_z\}$ of Z, potentially functional covariates, the expression for the SARMAHX $(p, q) \times (P, Q)_s$ model is defined as:

$$\begin{aligned}
&(I - \Psi_{0,1}B - \cdots - \Psi_{0,p}B^p)\left(I - \Psi_{1,1}B^s - \cdots - \Psi_{1,P}B^{P\cdot s}\right) Y_t = \\
&(I - \Theta_{0,1}B - \cdots - \Theta_{0,q}B^q)\left(I - \Theta_{1,1}B^s - \cdots - \Theta_{1,Q}B^{Q\cdot s}\right) \varepsilon_t \\
&+ \Gamma_1(X_t^1) + \cdots + \Gamma_Z(X_t^Z),
\end{aligned} \tag{3.39}$$

where $\Psi_{0,i}$ and $\Psi_{1,i}$ are the regular and seasonal autoregressive operators, $\Theta_{0,i}$ and $\Theta_{1,i}$ the regular and seasonal moving average operators, and Γ_i the operators related to the Z explanatory variables. B^n is the lag operator which is defined as $B^n Y_t = Y_{t-n}$ where $n \in \mathbb{N}$. Finally, I is the identity operator.

Each function of the series will be discretized in a grid of 100 equi-spaced points in the interval $[0, 1]$. The length of the simulated processes will be 1500 functional observations, after leaving 500 observations as a burn-in period in order to stabilize the series.

The first simulated process $Y_t^{(1)}$ is an ARH(1) model, given by the equation

$$Y_t^{(1)}(v) = \Psi(Y_{t-1}^{(1)}(v)) + \varepsilon_t(v) = \int \psi(u, v) Y_{t-1}^{(1)}(u)du + \varepsilon_t(v), \tag{3.40}$$

where $\Psi(.)$ is an integral operator that defines the process, $\psi(u, v)$ denotes the kernel of said integral operator and $\varepsilon_t(v)$ denotes the white noise innovations of the process. This process, studied in Bosq (2000), is stationary if $\|\Psi\| < 1$, so the simulated series will use the Gaussian kernel

$$\psi(u, v) = 0.6 \exp\left(\frac{u^2 + v^2}{2}\right), \ (u, v) \in [0, 1] \times [0, 1], \tag{3.41}$$

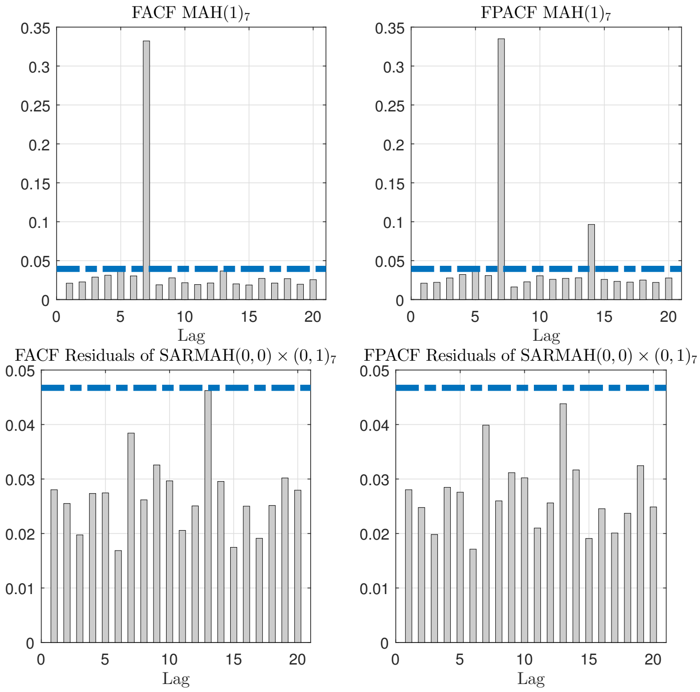

Figure 3.5. Top: FACF and FPACF for the simulated MAH$(1)_7$ process. Bottom: FACF and FPACF and the residuals of a fitted SARMAHX$(0,0) \times (0,1)_7$ model. The dotted line represents the 99% white noise prediction bounds for the FACF/FPACF coefficients under the assumption of white noise.

so $\|\Psi\| \approx 0.7$, ensuring that the resulting series will be stationary. The brownian bridges simulated in the last section will be used as the innovations of the process. The top panels of Figure 3.4 show the FACF and FPACF of the process $Y_t^{(1)}$ for $H = 20$ lags, together with the 99% i.i.d. bounds provided by Theorem 3.2.1. The coefficients of the FACF exhibit an exponential decay that agrees with the expected behavior of the ARH(1) process.

However, the FACF alone is not enough to identify the order of the autoregressive process– an ARH process of higher order could produce a similar FACF structure. Similarly to the scalar case, the FPACF can be used to better identify the order of autoregressive process, as it aims to remove the influence of the intermediate terms of the series at a given lag. This is illustrated in the top FPACF of Figure 3.4, where only the first coefficient is significant, which is consistent with the dependence structure of an ARH(1) process.

As most coefficients of the FACF fall above the critical value, the white noise assumption is rejected. The test proposed in Horváth, Hušková, et al. (2013) (referred to as the HHR hereafter) also rejects the white noise hypothesis at levels $\alpha = 0.1, 0.05$ and 0.01. Because of this, a ARMAHX$(1,0)$ model has been fitted to this FTS, and both FACF and FPACF of the residuals of the model are plotted on the bottom panels of Figure 3.4. As shown in the image, no pattern can be found in the residuals of the model, and all the coefficients fall below the critical value, suggesting the residuals of the model can be treated as a functional white noise. This is corroborated by the HHR test, which does not reject the white noise hypothesis for confidence level $\alpha = 0.1, 0.05$ and 0.01.

The next simulated process $Y_t^{(2)}$ will have the structure of a seasonal MAH$(1)_7$ (Aue, Norinho, et al., 2015), where each observation depends on the 7th lagged term of the series $Y_{t-7}^{(2)}$, following the equation

$$Y_t^{(2)}(v) = \Theta(\varepsilon_{t-7}(v)) + \varepsilon_t(v) = \int \theta(u,v)\varepsilon_{t-7}(u)du + \varepsilon_t(v), \tag{3.42}$$

where $\Theta(.)$ denotes an integral operator with kernel $\theta(u,v)$, and $\varepsilon_t(v)$ denotes white noise variables. As in the previous series, the innovations will be independent brownian bridges, and the kernel used to define the integral operator will be the Gaussian kernel defined in Equation 3.41, so $\|\Theta\| \approx 0.7$, resulting in a stationary series. The FACF and FPACF of the simulated series are shown in the top panels of Figure 3.5. Due to the innovations being uncorrelated, the FACF plot only shows dependence of the lag 7, whereas the FPACF plot shows a certain, approximately geometric, decay for the seasonal lags 7 and 14. These plots exhibit the shape of a moving average process with seasonality 7. When applied to this data, the HHR test rejects the null hypothesis of functional white noise at levels $\alpha = 0.1, 0.05$ and 0.01.

Once the process has been identified, a ARMAHX$(0,0) \times (0,1)_7$ model is fitted to the data. In order to validate the model, the main hypotheses of the model have to be checked. The values of the lagged autocorrelations of the residuals are obtained as well as the 99% limit distribution of the statistic to test the white noise assumption on the residuals. The HHR test does not reject the null hypothesis of functional white noise at level $\alpha = 0.1$, and when using the method proposed in this paper, similar results were obtained: as shown in the bottom panels of Figure 3.5; none of the autocorrelation values surpasses the 99% limit, hence the assumption of white noise can not be rejected, and the residuals of the fitted model can be regarded as functional white noise.

These results show the applicability of the proposed method in the diagnosis of FTS models, providing the practitioner with a graphical tool to visualize the structure of the FTS and identify possible deterministic components in the residuals of a model that have not been explained. Figure 3.6 shows the FACF and FPACF of several functional autoregressive and moving average processes that can be used to identify seasonality, the dependence structure of FTS and the order of the process: an exponential decay on the FACF is associated with an autoregressive process, and significant values of the FPACF indicate the order of the process. On the other hand, moving average processes are associated with a FPACF that exhibits an exponential decay and a FACF that has only a reduced number of values above the i.i.d. upper bound, where the number of significant coefficients of the FACF indicate the order of the moving average process. In the simulations of this section, both the proposed method and the HHR test have proven to provide similar results and rejection rates for FTS that exhibited some kind of time dependency between its terms. Next section will apply this identification and diagnosis method to several real-world datasets, in order to test the usefulness when dealing with real data.

3.4. Application to real-world datasets

To illustrate the applicability of the proposed method, two widely studied real-world datasets will be analyzed. Firstly, the FACF and FPACF will be obtained to quantify the correlation structure of the FTS in order to check the white-noise assumption on the original time series and to identify AR and MA trends. Secondly, a functional linear model will be fitted to each series and the test will be applied to the residuals of the models, in order to diagnose the models. If the white noise assumption is not rejected on the residuals, it will indicate that the model has

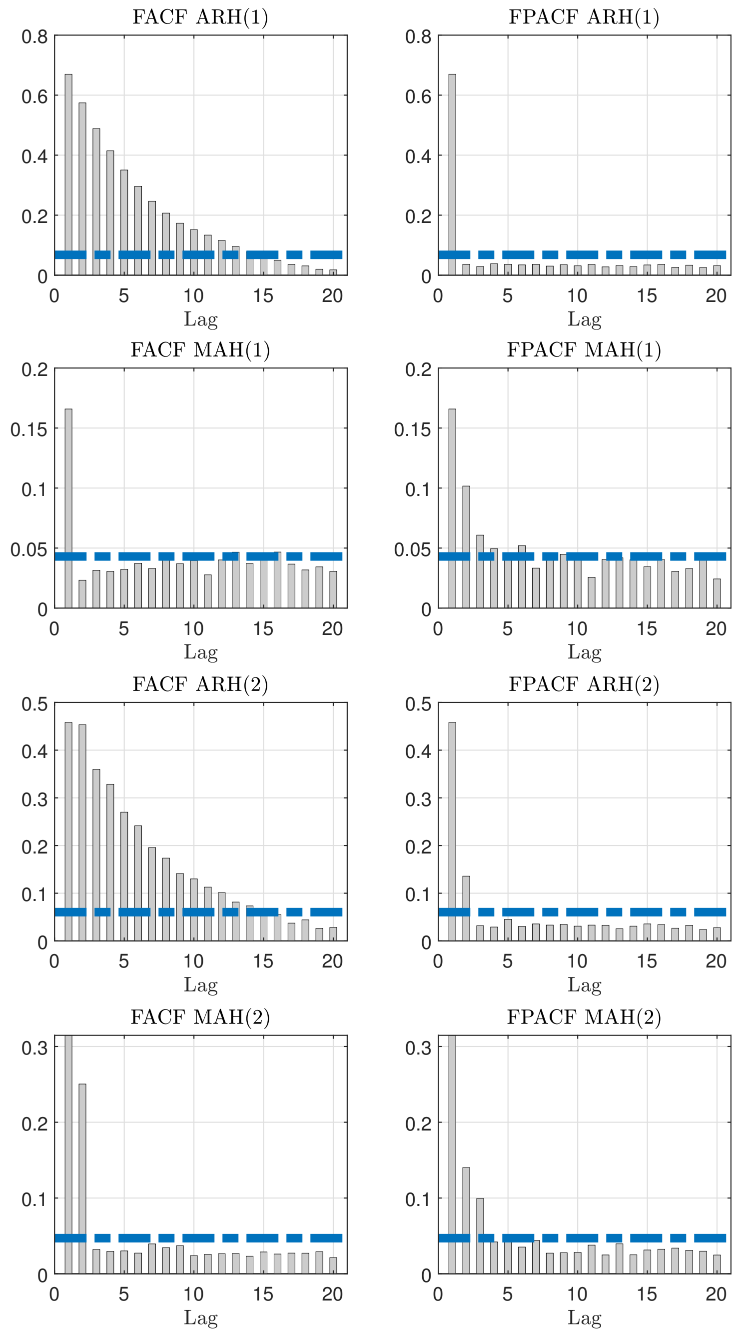

Figure 3.6. FACF and FPACF of four different simulated functional autoregressive and moving average processes. The shape exhibited by the autocorrelation function of each process can be used to identify each process. The dotted line represents the 99% white noise prediction bounds for the FACF/FPACF coefficients under the assumption of white noise.

extracted all the information of the original time series. In order to validate the proposed test, the results obtained will be compared with the HHR test.

3.4.1. Eurodollar futures contracts

The first real-world dataset to be analyzed consists of Eurodollar future rates (Kargin and Onatski, 2008; Horváth, Hušková, et al., 2013). As stated in those works, each Eurodollar futures contract is an obligation to deliver a 3-month deposit of $1,000,000 to a bank account outside of the United States at a specified time. Given v a scalar value, if one defines $Y_t^E(v)$ as the price of a contract with closing date v months from day t, data from this dataset can be

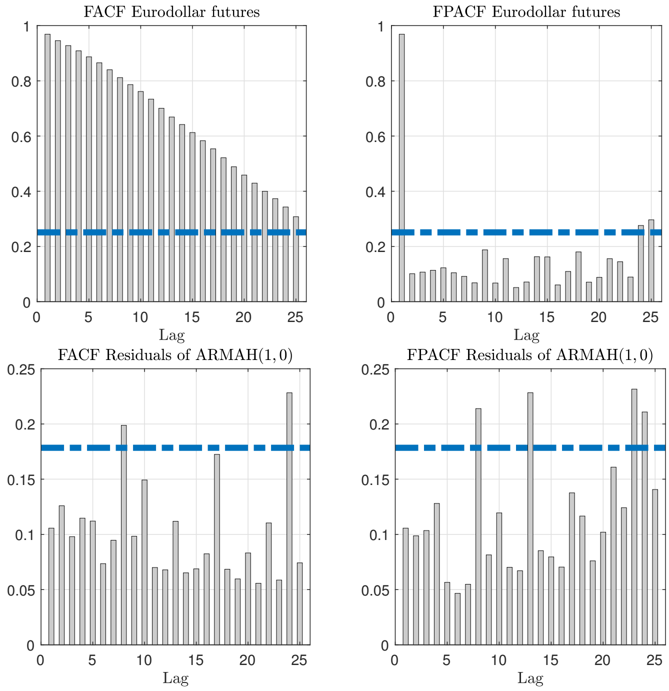

Figure 3.7. Top: FACF and FPACF of the Eurodollar dataset. Bottom: FACF and FPACF of residuals of a fitted ARMAHX$(1, 0)$ model. The dotted line represents the 95% white noise prediction bounds for the FACF/FPACF coefficients under the assumption of white noise.

regarded as a FTS. The Eurodollar dataset construct those curves using 114 points for each day, and in order to replicate the results of the HHR test (Horváth, Hušková, et al., 2013), the sample considered in this analysis consists of 100 days of data taken from January 24 to June 17, 1997.

An ARH(1) model was identified in Kargin and Onatski (2008), where the authors develop a technique to estimate the autoregression operator of the model, and illustrate this technique to the Eurodollar dataset. However, no identification procedure was used to check if the data can be modeled as an ARH(1). The top panels of Figure 3.7 show the FACF and FPACF of the Eurodollar series. The shape of the FACF plot shows an exponential decay, whereas only the first coefficient of the FPACF is significant. The shape exhibited by the correlogram is similar to the simulated ARH(1) process shown in Figure 3.4; hence validating the selection of an autoregressive model to model the data. The presence of structural dependence between the data is also confirmed by the 95% critical line displayed in the plot: all of the coefficients of the FACF fall above of the critical bound, hence the assumption of functional white noise must be dropped. This was also observed in Horváth, Hušková, et al. (2013), where the authors applied the HHR test to the same data and rejected the white noise hypothesis at 99% confidence level.

Once the structure of the process has been identified, an ARMAHX$(1, 0)$ model has been fitted to the data, and the FACF and FPACF of the model's residuals are displayed on the bottom panels of Figure 3.7. The HHR test does not reject the white noise hypothesis at confidence level $\alpha = 0.1$. This result is consistent with the outcome of the diagnosis methodology developed in this paper: most of the coefficients of the FACF and FPACF obtained for the residuals fall below the 95% critical bound and no structure can be identified in the plots, hence assuming that the residuals of the model are white noise, validating the selected model.

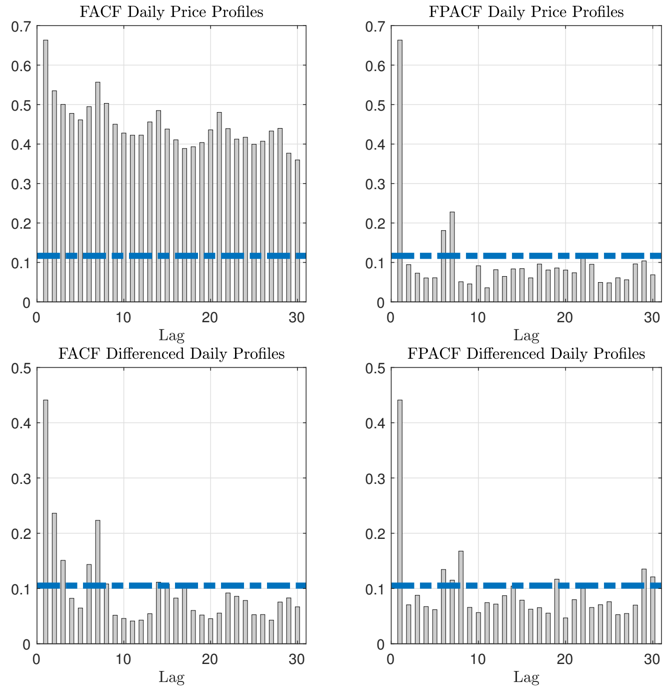

Figure 3.8. Top: FACF and FPACF of the daily price profiles series Y_t^P. Bottom: FACF and FPACF of the seasonally differenced series Z_t^P. The dotted line represents the 99% white noise prediction bounds for the FACF/FPACF coefficients under the assumption of white noise.

3.4.2. Spanish electricity daily price profiles

The second dataset was studied in Portela, Muñoz San Roque, et al. (2018), where the authors analyze the series of hourly Spanish electricity spot prices provided by the Spanish electricity Market Operator (`www.omie.es`). As explained in the paper, the spot price series $\{y_{t,h}\}$ is transformed into a FTS $\{Y_t^P(v)\,;\, t = 1, \ldots, T\,;\, v \in [1, 24]\}$ where each function Y_t is observed at discrete hours $v_i \in \{1, \ldots, 24\}$, thus $Y_t^P(v_i) = y_{t,v_i}$. Hence each observation is a daily price profile. In order to replicate the results of that work, the sample considered in this analysis consists of 1 year of data taken from January 1, 2014 to December 31, 2014. The daily price profiles are shown in Figure 3.1. We here note that all below functional regressions models fitted to obtain the FPACF are estimated using FPCA regularization with level chosen by so that 95% of the total variance is explained by the selected number of FPC's.

In order to model the electricity price series, Portela, Muñoz San Roque, et al. (2018) proposed a SARMAHX$(1,0) \times (0,1)_7$ model, using the daily demand X_t^D and wind profiles X_t^W as exogenous functional variables. However, the authors relied on FPCA reduction techniques and testing different configurations in order to identify the structure of the model. Besides, no white noise identification test was applied on the residuals. The purpose of this section is to illustrate the application of the autocorrelation functions in the identification and diagnosis of a model, by testing the white noise hypothesis on the residuals.

The top panels of Figure 3.8 show the FACF and FPACF for the first 30 lags of the price functional time series. This diagnostic tool provides useful information about the correlation structure of the series: firstly, all the coefficients of the FACF lie above the 99% critical bound, so the white noise assumption is strongly rejected. The HHR test provides the same conclusion. Secondly, the diagnostic plot exhibits a seasonal dependency with period 7, which can be explained by the daily nature of the series. Moreover, a slow decrease in the values of the FACF is observed, which might imply that the series should be differenced. The bottom panels of

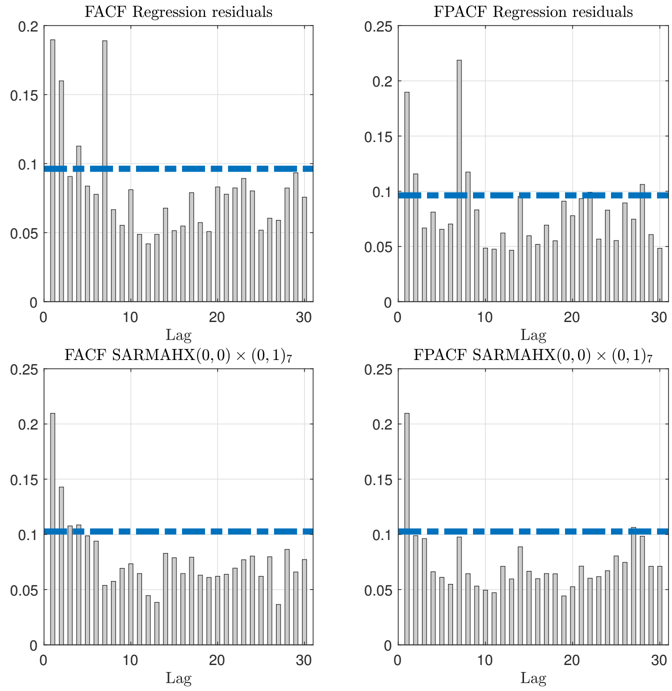

Figure 3.9. Top: FACF and FPACF of the residuals of the regression model. Bottom: FACF and FPACF of the residuals of the fitted SARMAHX$(0,0) \times (0,1)_7$ model. The dotted line represents the 99% white noise prediction bounds for the FACF/FPACF coefficients under the assumption of white noise.

Figure 3.8 show the FACF and FPACF of the price series after being seasonally differenced at lag 7. As can be seen the slow decrease of the coefficients in the FACF is no longer observed. Hence, the differenced series are used, which are denoted as

$$Z_t^P = Y_t^P - Y_{t-7}^P \tag{3.43}$$

$$U_t^D = X_t^D - X_{t-7}^D \tag{3.44}$$

$$U_t^W = X_t^W - X_{t-7}^W, \tag{3.45}$$

where Z_t^P denotes the series of seasonally differenced daily electricity price profiles and U_t^D and U_t^W denote the demand and wind production series after the seasonal differentiation.

In order to model the price time series, an initial SARMAHX regression model is estimated using the differenced demand and wind production series as inputs. The expression of this regression model is as follows:

$$Z_t^P = \Gamma_W(U_t^W) + \Gamma_D(U_t^D) + \varepsilon_t, \tag{3.46}$$

where Γ_W and Γ_D are the integral operators related to the differenced demand and wind exogenous variables. The top panels of Figure 3.9 show the FACF and FPACF of the regression model residuals, which can be used to extract information about the correlation structure of the underlying process that generated the series. Most of the values of the FACF fall below the critical limit, however there seems to be a strong correlation on the first lags, and the white noise assumption is rejected at the 99% confidence level using the HHR test. Consequently, using the exogenous variables is not enough for modeling the series. When compared with Figures 3.4 and 3.5, the FACF and FPACF plots seem to have two components: a regular autoregressive behavior (identified by the first coefficient of the FPACF) and a moving average effect on lag 7.

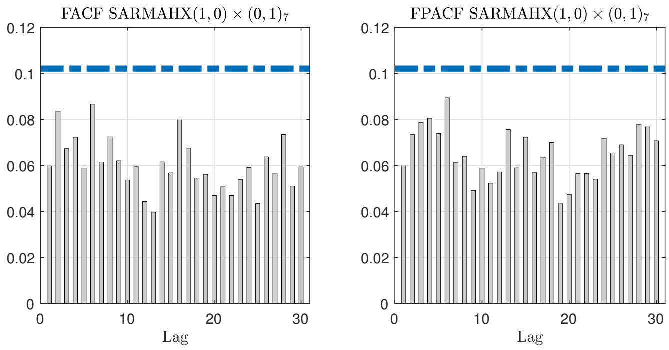

Figure 3.10. FACF and FPACF of the residuals of the SARMAHX$(1,0) \times (0,1)_7$ model. The dotted line represents the 99% white noise prediction bounds for the FACF/FPACF coefficients under the assumption of white noise.

The next step in this analysis is to fit a SARMAHX$(0,0) \times (0,1)_7$ model to the differenced series, in order to capture this seasonal moving average effect. The expression for this models is:

$$Z_t^P = \Gamma_W(U_t^W) + \Gamma_D(U_t^D) - \Theta(\varepsilon_{t-7}) + \varepsilon_t. \tag{3.47}$$

The correlation structure of the residuals of that model is shown in the bottom panels of Figure 3.9. As can be seen, the moving average effect has been removed. However, the first lags of the FACF function lie above the critical limit and exhibit a certain exponential decay pattern in the FACF, whereas the FPACF still has one significant value at lag 1. This causes both the proposed test and the HHR test to reject the white noise hypothesis at 99% confidence level.

Finally, a SARMAHX$(1,0) \times (0,1)_7$ model has been fitted to the differenced series of daily price profiles. The expression for this model is:

$$Z_t^P = \Gamma_W(U_t^W) + \Gamma_D(U_t^D) + \Psi(Z_{t-1}^P) - \Theta(\varepsilon_{t-7}) + \varepsilon_t. \tag{3.48}$$

The autocorrelation plot of the residuals of the model is shown in Figure 3.10. All the values of both the FACF and FPACF fall below the critical limit, hence the white noise assumption cannot be rejected. This is validated by the HHR test.

This process illustrates the proposed identification and diagnosis procedure that has been developed for FTS linear models. For models within the SARMAHX family of models, this tool is useful to select the optimum configuration of its parameters. This enables the practitioner to capture the dynamics of the underlying process, obtaining more accurate forecasts of the FTS.

3.5. Conclusions

This section has presented a set of statistical tools that can be used to identify the serial dependence structure of a given FTS, identify suitable FTS models and diagnose fitted FTS models. The motivation for these tools was to define versions of the autocorrelation and partial autocorrelation functions for FTS, with similar properties as their scalar counterpart. The statistical tool proposed in this section differs from other white noise testing procedures found in the literature in several aspects: where most testing methods rely on FPCA to reduce the dimension of the curves of the series, the proposed methods use the correlation structure of the FTS to capture the structure of the data. Hence, no dimensionality reduction is applied, avoiding the need of choosing functional principal components and approximating the functional series with a finite number of terms, thus losing some information in this process. In addition, while existing functional white noise tests can be used to evaluate whether a given series or sequence of residuals is plausibly a white noise, the use of FACF and FPACF plots is helpful in identifying the specific serial correlation structure, providing some insight when fitting a linear model to that series. This is particularly useful in identifying the order of functional autoregressive and moving average processes.

The performance of the method has been validated both numerically and empirically, identifying the correlation structure of both Eurodollar future contracts and Spanish electricity price profiles FTS and characterizing the residuals of the fitted models as functional white noise. It has been shown that the white noise identification has a similar performance to other white noise tests. The proposed methods have been implemented and are currently available in the `R` package `fdaACF`.

Although the proposed FACF and FPACF can be used in a similar way as their scalar counterparts to identify and diagnose FTS models, the fact that these functions are based on the L^2 norm of the lagged covariance operators makes it difficult to identify whether the correlation exhibited by the series is "positive" or "negative" in nature; only the magnitude is measured. The development of a signed version of the proposed autocorrelation functions could overcome these problems, providing the practitioner with a useful tool to model FTS. This topic is left as an area of future research.

4

Probabilistic forecasting of functional time series

This chapter is devoted to the functional probabilistic model developed in this thesis. Once a functional deterministic model has been identified and fitted to the data, it is necessary to characterize the distribution of its residuals to obtain probabilistic forecasts of the series being analyzed. The proposed model aims at estimating the conditional projection-based density of the residuals of a forecasting model using non-parametric techniques. The relationship between the distribution of the residuals and a set of covariates is quantified using kernel functions, and a leave-one-out cross-validation is implemented to select the bandwidth parameters of the model. A review of different quality measures for functional probabilistic forecasts is performed, which highlights the benefits and drawbacks of the different metrics proposed in the literature. In addition, a generalization of the log-score for functional data is developed as a novel quality measure that it is suitable to measure the adequacy of a set of functional scenarios. These techniques are validated in both a simulation study and with an case study using Complex Residual Demand Curves obtained from the Iberian Electricity Market.

4.1. Introduction

The main goal of forecasting is to obtain accurate estimations about the conditional distribution of some quantity of interest. Usually, forecasting models aim at estimating a certain parameter of that distribution, commonly the conditional mean of the process. These deterministic models are able to capture the dynamics of the series, providing the user with a meaningful description of the evolution of the data. As such, deterministic forecasting models are of utmost importance across a huge number of real-world problems that involve understanding the temporal evolution of certain values. However, in numerous real-world applications it is necessary to quantify the risk associated with each possible outcome. In these situations deterministic models may not be enough, as the user is usually interested in characterizing the whole future distribution of the series. Henceforth, deterministic forecasts must be accompanied by some measure of the uncertainty of each future estimation, which motivates the development of probabilistic forecasting models.

One of the main objectives of this thesis is to develop a probabilistic model for functional time series based on the SARMAHX model presented in Section 2.3.3. In order to present the main ideas, recall than $\{Y_t\}_{t=1}^T$ denotes a functional time series (FTS). Suppose that a forecasting model has been fitted to the series, and denote by $\{\widehat{Y}_{t,h}\}_{t=1}^T$ the h-step ahead estimations of the original series made by the model. Then, the h-step ahead residuals of the model can be defined as $\widehat{\varepsilon}_{t,h} = Y_t - \widehat{Y}_{t,h}$. As mentioned in the review performed in Section 2.4, the vast majority of probabilistic models for time series are aimed at characterizing the distribution of the residuals of the model. The most common technique used to obtain probabilistic forecasts of a functional time series is bootstrap resampling of the residuals of a functional deterministic model, as evidenced in Raña, Aneiros, et al. (2016), Vilar, Aneiros, et al. (2018), and Rossini and Canale (2019), among others. These models assume that the residuals do not exhibit any kind of dependence, so the bootstrap replications of the residuals approximate the true underlying distribution of the non-deterministic part of the model. A similar technique that has been commonly used to define a probabilistic model for functional time series is the projection-based density proposed in Delaigle and Hall (2010). This technique can be used to approximate the unconditional distribution of the residuals of a functional forecasting model, which can then be used to generate functional scenarios of the curves.

However, in many real-world applications these probabilistic models may not be adequate, as the assumption of independence in the residuals is quite restrictive. Usually, reference probabilistic models proposed in the functional data literature do not perform any kind of statistical test to validate the i.i.d. assumption of the residuals. Hence, the bootstrapped residuals may exhibit some kind of linear dependence between its terms, so the obtained functional scenarios may not capture the dynamics present in the original data. In this regard the identification and diagnosis methodology proposed in the previous chapter can be seen as a useful tool for these models, as the FACF/FPACF proposed can be used to test whether the residuals of a functional deterministic forecasting model present serial correlation. However, even after validating the i.i.d. hypothesis on the model's residuals, there may be some kind of dynamics on some higher order moments of the residuals, such as the variance of the residuals being influenced by some extreme values of some exogenous variable. This is a common occurrence in industrial applications such as electricity price forecasting: the uncertainty associated with the price forecasts is higher when there are some extreme values of external drives such as the temperature or the electricity load. As such, any competitive probabilistic model must be able to take into account these exogenous effects that modify the distribution of the residuals of the model.

As opposed to the methods found in the literature, this thesis proposes a probabilistic model that is able to model the conditional distribution of the residuals of a forecasting model for functional time series. This model is based on the projection-based density given by Delaigle and Hall (2010), which approximates the underlying distribution that generates the functional time series. In addition, techniques commonly used in conditional density estimation (Rosenblatt, 1969), such as kernel weighting of the input variables, are introduced to define this new model. This provides a suitable framework for quantifying the variations of the residuals' distribution. The estimated conditional density can then be used to define probabilistic regions at a user specified level, which provide a useful tool for assessing the quality of the predictions of the forecasting model. Additionally, functional scenarios will be obtained from the fitted density, which can be used to analyze the set of possible outcomes observed in a future event; or as inputs for stochastic programming models (Kaut and Wallace, 2007). In addition to the proposed probabilistic model, a comparison between different measures for assessing the quality of functional scenarios proposed in the literature will be performed, highlighting the advantages and drawbacks of each metric. A new quality measure for functional scenarios will be proposed, which will be able to capture the likelihood of functional scenarios when compared with the original functional series.

This chapter is organized as follows: firstly, the formulation of the functional probabilistic model proposed in this thesis is addressed in Section 4.2, introducing a leave-one-out cross-validation methodology to select the bandwidth parameters of the model. The model is able to estimate the conditional projection-based density of a set of functional data, which then is used to generate functional scenarios and to define prediction regions. Then, Section 4.3 performs a review of the different quality measures proposed in the functional data literature to compare probabilistic forecasts and proposes a generalization of the log-score metric to the functional data framework. These novel techniques are validated in a simulation study performed in Section 4.4. Finally, the applicability of the proposed method is illustrated in Section 4.5, which presents a real case study using Complex Residual Demand Curves obtained from the Iberian Electricity Market.

4.2. An estimator for the conditional projection-based density for functional data

This section presents a probabilistic model for functional time series based on a non-parametric estimator for the conditional projection-based density of the residuals $\{\widehat{\varepsilon}_t\}_{t=1}^T$ of a forecasting model. The proposed model aims at modelling the conditional distribution of the projections of the residuals on a truncated functional basis using as inputs a set of exogenous covariates that are able to explain the variations of the projection-based density of the residuals. Once the distribution has been estimated, the functional basis used for decomposing the curves can be used to reconstruct the curves, which can be used to define predictive bands for the forecasts of the fitted deterministic model and to generate functional scenarios that are consistent with the estimated conditional density. As such, this section is structured as follows: firstly, the formulation of the model will be detailed in Section 4.2.1, providing an estimator for the conditional density of a set of functional residuals. A leave-one-out cross-validation procedure is proposed, which enables the automatic selection for the bandwidth parameters of the proposed model. Then, Section 4.2.2 illustrates how the projection-based density estimated by the model can be employed to generate coherent functional scenarios and to obtain prediction regions that capture the expected volatility of the data.

4.2.1. Model formulation

As mentioned earlier, the proposed probabilistic model aims at estimating the conditional density of the residuals of a functional forecasting model such as the SARMAHX model presented in Section 2.3.3. This model will provide a useful tool to characterize the residuals of the forecasting model, which can then be used both to diagnose the adequacy of the forecasting model and to obtain probabilistic forecasts of the modelled time series. Due to the lack of a proper density function for infinite-dimensional data, the projection-based density $p_\varepsilon^{(M)}$ will be used instead, which provides an approximation to the underlying density of the functional data. In many applications it would be enough to estimate the unconditional distribution of the residuals of the model. However, due to the complex relationships that are often exhibited by the functional time series obtained from electricity markets, a more in-depth analysis of the residuals of a fitted forecasting model is needed. As such, modelling the conditional density of the residuals is preferred, as it can provide information about sudden changes in the dynamic of the series that can be used to improve the initial forecasting model.

As mentioned earlier, the proposed model aims at estimating the projection-based density $p_\varepsilon^{(M)}$ of the residuals of a functional forecasting model. Hence, the first step to estimate the conditional density of the data is to project the observed residuals $\{\widehat{\varepsilon}_t\}_{t=1}^{T}$ into the finite-dimensional subspace defined by the truncation of the first FPC of the residuals (Delaigle and Hall, 2010):

$$\widehat{\varepsilon}_t \approx \widehat{\mu}_\varepsilon + \sum_{j=1}^{M} \widehat{e}_{t,j}\widehat{\phi}_j = \widehat{\mu}_\varepsilon + \widehat{\boldsymbol{e}}_t\widehat{\boldsymbol{\phi}}, \tag{4.1}$$

where $\widehat{\boldsymbol{e}}_t = (\widehat{e}_{t,1}, \dots, \widehat{e}_{t,M})^\top$ and $\widehat{\boldsymbol{\phi}} = \left(\widehat{\phi}_1, \dots, \widehat{\phi}_M\right)$ are the scores of the decomposition and the eigenfunctions, respectively. The number M of FPC is chosen via the CPV method described in Section 2.2.5. The conditional joint density of the scores $\widehat{e}_{t,j}$ of the decomposition is estimated using a conditional kernel density estimator (CKDE), a non-parametric model that is able to model changes in the distribution of the data given a set of covariates. This model was first introduced in Rosenblatt (1969) as a generalization of the standard Parzen-Rosenblat kernel density estimator (Rosenblatt, 1956; Parzen, 1962), which aims at estimating the unconditional density of a set of scalar data. The authors proposed the following model to estimate the density p of a set of scalar values $\{y_i\}_{i=1}^{N}$ conditioned to the values of a given set of scalar covariates $\{\boldsymbol{x}_i\}_{i=1}^{N}$:

$$\hat{p}_{\mathscr{H}}(y|\boldsymbol{x}) = \frac{\hat{p}_{\mathscr{H}}(y,\boldsymbol{x})}{\hat{p}_{\mathscr{H}}(\boldsymbol{x})} = \frac{\sum_{i=1}^{N} K_{h_1}(y-y_i)\, K_{h_2}(\|\boldsymbol{x}-\boldsymbol{x}_i\|)}{\sum_{i=1}^{N} K_{h_2}(\|\boldsymbol{x}-\boldsymbol{x}_i\|)}, \tag{4.2}$$

where $\mathscr{H} = \{h_1, h_2\}$ is the set of bandwidth parameters of the fitted model, $\hat{p}_{\mathscr{H}}(y,\boldsymbol{x})$ is the joint density estimation of $(y,\boldsymbol{x})$, $\hat{p}_{\mathscr{H}}(\boldsymbol{x})$ stands for the density estimator for the covariate $\boldsymbol{x}$ and $K_h(u) = h^{-1}K\left(h^{-1}u\right)$ denotes a scaled kernel function. An interesting application of this model in a forecasting problem is illustrated in Dudek (2010), where the CKDE model is used to obtain short-term forecasts of the probability density of the hourly demand of the Polish electricity market; or in Arora and Taylor (2016), where density estimations of the electricity consumption recorded by smart meters are obtained using the conditional kernel density estimation model. In addition, conditional kernel density estimation models have been successfully applied in wind power forecasting, as evidenced in Bessa et al. (2012) and Jeon and Taylor (2012).

The proposed model for estimating the conditional joint density of the FPCA scores is based on the classical CKDE model, extending the formulation of that model to a multidimensional framework and allowing the inclusion of functional covariates using a semi-metric function. Hence, this new model is able to capture the variation in the distribution of the series of residuals caused by sudden changes in the set of exogenous covariates included in the model. Given $\{\boldsymbol{e}_t\}_{t=1}^{T}$ the series of scores and $\{X_t^z\}_{t=1}^{T}, z = 1, \dots, Z$ a set of exogenous covariates (either scalar or functional), the proposed CKDE model is formulated as follows:

$$\hat{p}_{\mathscr{H}}^{(M)}(\boldsymbol{e}|\boldsymbol{X}) = \frac{\hat{p}_{\mathscr{H}}^{(M)}(\boldsymbol{e},\boldsymbol{X})}{\hat{p}_{\mathscr{H}}^{(M)}(\boldsymbol{X})} = \frac{\sum_{t=1}^{T} \mathbf{K}_{\boldsymbol{H}}(\boldsymbol{e}-\boldsymbol{e}_t) \prod_{z=1}^{Z} K_{h_z}\left(\mathrm{d}^z\left(X^z, X_t^z\right)\right)}{\sum_{t=1}^{T} \prod_{z=1}^{Z} K_{h_z}\left(\mathrm{d}^z\left(X^z, X_t^z\right)\right)}, \tag{4.3}$$

where $\mathscr{H} = \{H, h_1, \dots, h_Z\}$ denotes the set of bandwidth parameters of the model, $\boldsymbol{X} = \left(X^1, \dots, X^Z\right)$ denotes the set of new exogenous covariates used as inputs for the model, $\mathrm{d}^z\left(.,.\right)$ are suitable semi-metric functions defined for each covariate X^z included in the model and $\mathbf{K}_{\boldsymbol{H}}$ denotes a multidimensional scaled kernel, defined as $\mathbf{K}_{\boldsymbol{H}}(\boldsymbol{u}) = |H|^{-1/2}\mathbf{K}\left(H^{-1/2}\boldsymbol{u}\right)$. The model is able to quantify the likelihood of a new observation $\left(\boldsymbol{e}, X^1, \dots, X^Z\right)$ with respect to the historical data used to train the model, giving more importance to past situations that are similar to the new data. In addition, the model allows the inclusion of functional variables via the use of a suitable semi-metric function that quantifies the similarity between the historical covariates and the new curve being evaluated. As such, the selection of a proper distance function is of utmost importance when fitting the proposed model, as it needs to capture the relevant properties of the input curves.

A more compact expression of Equation 4.3 can be obtained by rearranging its terms as follows:

$$\hat{p}^{(M)}_{\mathscr{H}}\left(\boldsymbol{e}|\boldsymbol{X}\right) = \sum_{t=1}^{T} \omega_t\left(\boldsymbol{X}\right)\mathbf{K}_{\boldsymbol{H}}\left(\boldsymbol{e} - \boldsymbol{e}_t\right), \tag{4.4}$$

where

$$\omega_t\left(\boldsymbol{X}\right) = \frac{\prod_{z=1}^{Z} K_{h_z}\left(\mathrm{d}^z\left(X^z, X_t^z\right)\right)}{\sum_{t=1}^{T}\prod_{z=1}^{Z} K_{h_z}\left(\mathrm{d}^z\left(X^z, X_t^z\right)\right)}. \tag{4.5}$$

Due to its similarity to the Nadaraya-Watson kernel regression estimator, this form is known as the Nadaraya-Watson form of the CKDE model (De Gooijer and Zerom, 2003). As such, the model proposed in Equation 4.3 can be regarded as a weighted version of the standard kernel density estimator, where the contribution of each past observation is pondered by the similarity to the new observations. This allows for a straight-forward interpretation of the output of the model: the regions of the FPC-defined subspace that exhibit a positive density will be associated with a positive weight $\omega_i\left(\boldsymbol{X}\right)$, whereas regions associated with unrelated events will have a negligible weight.

In order to fit the proposed CKDE model, both the family of the kernel functions and the bandwidth parameters used by the model must be selected by the user. As mentioned in Wand and Jones (1995) the choice of the kernel functions is of secondary importance, as the performance of the models does not vary significantly when selecting different kernel functions. Popular choices of kernel functions are often defined in terms of univariate and unimodal probability density functions. The current implementation of the probabilistic model uses standard Gaussian density functions as kernels, as these functions are widely used in the kernel density estimation literature due to its convenient mathematical properties. In Hyndman and Yao (2002) the authors compared the effectiveness of the CKDE model using both standard Gaussian kernels and truncated Gaussian kernels, and found that the performance of the two models was identical, which is consistent with our experience. However, the same cannot be said for the selection of the bandwidth parameters, which control the degree of smoothing applied to the density estimate. As with all kernel estimators, the performance of the model is critically dependent on a suitable choice for the bandwidth parameters. Therefore, a suitable criterion for selecting the optimal set of bandwidth parameters must be developed in order to fit the conditional density estimator to a set of data.

In order to derive an efficient way to estimate the bandwidth parameters of the CKDE model, some authors have studied the asymptotic properties of the estimator, which leads to plug-in estimators and reference rules (Fan, 1996; De Gooijer and Zerom, 2003). However, these methods are known to exhibit an unsatisfactory performance in the finite sample setting, hence the use of a data-driven criterion is preferred. As mentioned in Holmes et al. (2010), the most common data-driven bandwidth score used in the CKDE literature is the mean integrated squared error (MISE), which can be expressed in the following form for the model proposed in Equation 4.3:

$$\text{MISE}\left(\mathscr{H}\right) = \iint \left(p^{(M)}(\boldsymbol{e}|\boldsymbol{X}) - \hat{p}^{(M)}_{\mathscr{H}}\left(\boldsymbol{e}|\boldsymbol{X}\right)\right)^2 d\boldsymbol{e}\left(p^{(M)}(\boldsymbol{X})\right) d\boldsymbol{X}, \tag{4.6}$$

where $p^{(M)}(\boldsymbol{e}|\boldsymbol{X})$ denotes the true projection-based conditional density of the data. As showed in Fan and Yim (2004), the minimization of the MISE is equivalent to minimizing the quantity

$$\iint \left(\hat{p}^{(M)}_{\mathscr{H}}\left(\boldsymbol{e}|\boldsymbol{X}\right)\right)^2 d\boldsymbol{e}\left(p^{(M)}(\boldsymbol{X})\right) d\boldsymbol{X} - 2\iint \left(\hat{p}^{(M)}_{\mathscr{H}}\left(\boldsymbol{e}|\boldsymbol{X}\right) p^{(M)}(\boldsymbol{e},\boldsymbol{X})\right) d\boldsymbol{e}d\boldsymbol{X}. \tag{4.7}$$

Hence, the authors proposed a consistent estimator of the MISE using cross-validation techniques, which can be expressed as:

$$\widehat{\text{MISE}}\left(\mathscr{H}\right) = \frac{1}{T}\sum_{t=1}^{T}\int\left(\hat{p}^{(M)}_{-t,\mathscr{H}}\left(\boldsymbol{e}|\boldsymbol{X}_t\right)\right)^2 d\boldsymbol{e} - \frac{2}{T}\sum_{t=1}^{T}\hat{p}^{(M)}_{-t,\mathscr{H}}\left(\boldsymbol{e_t}|\boldsymbol{X}_t\right), \tag{4.8}$$

where $\hat{p}^{(M)}_{-t,\mathscr{H}}$ denotes $\hat{p}^{(M)}_{\mathscr{H}}$ evaluated with the pair $(\boldsymbol{e}_t, \boldsymbol{X}_t)$ left out.

However, this cross-validated estimator is computationally intensive, as it involves several nested summation terms, which can make it computationally unfeasible when working with high-dimensional data. As such, the estimator that will be used to select the parameters of the proposed CKDE model for the projection-based density of the functional residuals will be a version of the cross-validated log-likelihood estimator proposed in Holmes et al. (2010). Given $\mathscr{H}$ a set of bandwidth parameters for the proposed CKDE model, the cross-validation log-likelihood of the model is defined as follows:

$$\text{L}\left(\mathscr{H}\right) = \frac{1}{T}\sum_{t=1}^{T}\log\left(\hat{p}^{(M)}_{-t,\mathscr{H}}\left(\boldsymbol{e_t}|\boldsymbol{X}_t\right)\hat{p}^{(M)}_{-t,\mathscr{H}}\left(\boldsymbol{X}_t\right)\right), \tag{4.9}$$

where $\hat{p}^{(M)}_{-t,\mathscr{H}}\left(\boldsymbol{X}_t\right)$ is the denominator of Equation 4.3 fitted without the value $\boldsymbol{X}_t$, evaluated on $\boldsymbol{X}_t$. As mentioned in Holmes et al. (2010), this criterion aims at optimizing the conditional density of the data $\hat{p}^{(M)}_{\mathscr{H}}\left(\boldsymbol{e}|\boldsymbol{X}\right)$, with each contribution being weighted by $\hat{p}^{(M)}_{\mathscr{H}}(\boldsymbol{X})$ so that the most frequent conditional distributions have more influence on the choice of bandwidth.

The following example illustrates how the proposed log-likelihood function can be used to estimate the bandwidth parameters of the model. A multivariate set of functional data $\{(Y_t, X_t)\}_{t=1}^{300}$ has been simulated, where the variance of the first 2 FPC of the output $\{Y_t\}_{t=1}^{300}$ is related to the shape of the functional covariates $\{X_t\}_{t=1}^{300}$: if the L^2-norm of the curve X_t is high, the scores of Y_t will exhibit a high variance. On the other hand, the dispersion of the scores will be low if its associated covariate X_t has a low L^2-norm. The model introduced in this section seems appropriate to capture this conditional distribution of the data, but first the set of bandwidth parameters must be selected.

A leave-one-out cross-validation method has been implemented, which allows for automatic selection of the bandwidth parameters of the model. Figure 4.1 illustrates this procedure: the top image shows the average log-likelihood obtained by the cross-validation procedure for each set of bandwidth parameters tested. As it can be seen, when extreme values are selected for the bandwidth parameters of the model, the mean log-likelihood of the model is quite low, as the model is either being overfitted or underfitted. In order to illustrate the log-likelihood measure being optimized by the cross-validation criterion, three sets of bandwidth parameters are selected: an optimal set (represented in blue), an overfitted set (represented in orange) and an underfitted set, which is represented by the yellow lines. Then two different volatility scenarios are presented: the first case corresponds to a curve with a low L^2 norm, which is associated with low variance in the FPC scores. The second case presents a situation of high volatility, as the curve X_t presents a high L^2 norm. The underfitted model is not able to distinguish between these two situations, hence estimating a similar density in both cases. Conversely, the overfitted model is not able to generalize, as it cannot capture the correct level of variance given by the exogenous variables. Finally, the model that uses the optimal parameters selected by the cross-validation procedure is able to model these changes in the distribution of the scores, as evidenced by the conditional densities that estimates.

This section was aimed at formalizing the proposed functional conditional density estimator model. Given a set of functional curves whose distribution exhibits some kind of relation with other exogenous variables, the proposed model is able to capture this dynamic and to obtain estimations of the projection-based conditional density of the curves. By characterizing the distribution of the residuals of a functional forecasting model, probabilistic forecasts of a functional time series can be obtained. The following section will illustrate how the estimations of the model can be used to obtain functional scenarios that capture the uncertainty about a future event and to define prediction regions that quantify the uncertainty associated with the forecasts of a deterministic model.

4.2.2. Application to scenario generation and prediction regions

The previous section was devoted to the development of a probabilistic model for functional time series aimed at estimating the conditional projection-based density of a set of curves. This model can be used for characterizing the conditional distribution of the residuals of a functional deterministic model, which enables the generation of probabilistic statements for the forecasts of the model. Once the model has been fitted it is necessary to convey the probabilistic information captured by the model to the user. The most widely used tools found in the forecasting literature for this purpose are prediction intervals, which represent the uncertainty about the forecasts of the model (Patel, 1989). Prediction intervals estimate the possible range of values in which a future observation may fall at a given confidence level α, typically 80% or 95% (Hyndman and Athanasopoulos, 2018). As such, prediction intervals are used in numerous practical applications as a tool for quantifying the risk associated with the forecasts of a given model.

Closely related to the estimation of prediction intervals are scenario generation methods, which aim at generating a set of equiprobable future values of the variable that is being modelled in such a way that they capture the variability associated with the forecasting model. The set of future scenarios represent the set of possible outcomes of a variable after taking into account the relative impact of all possible factors or drivers that explain the dynamic of the data and the interactions between them. Scenario forecasting methods can be used to generate a wide range of possible forecasts and to identify extremes situations, which can be extremely useful for agents that need to make risk-based decisions.

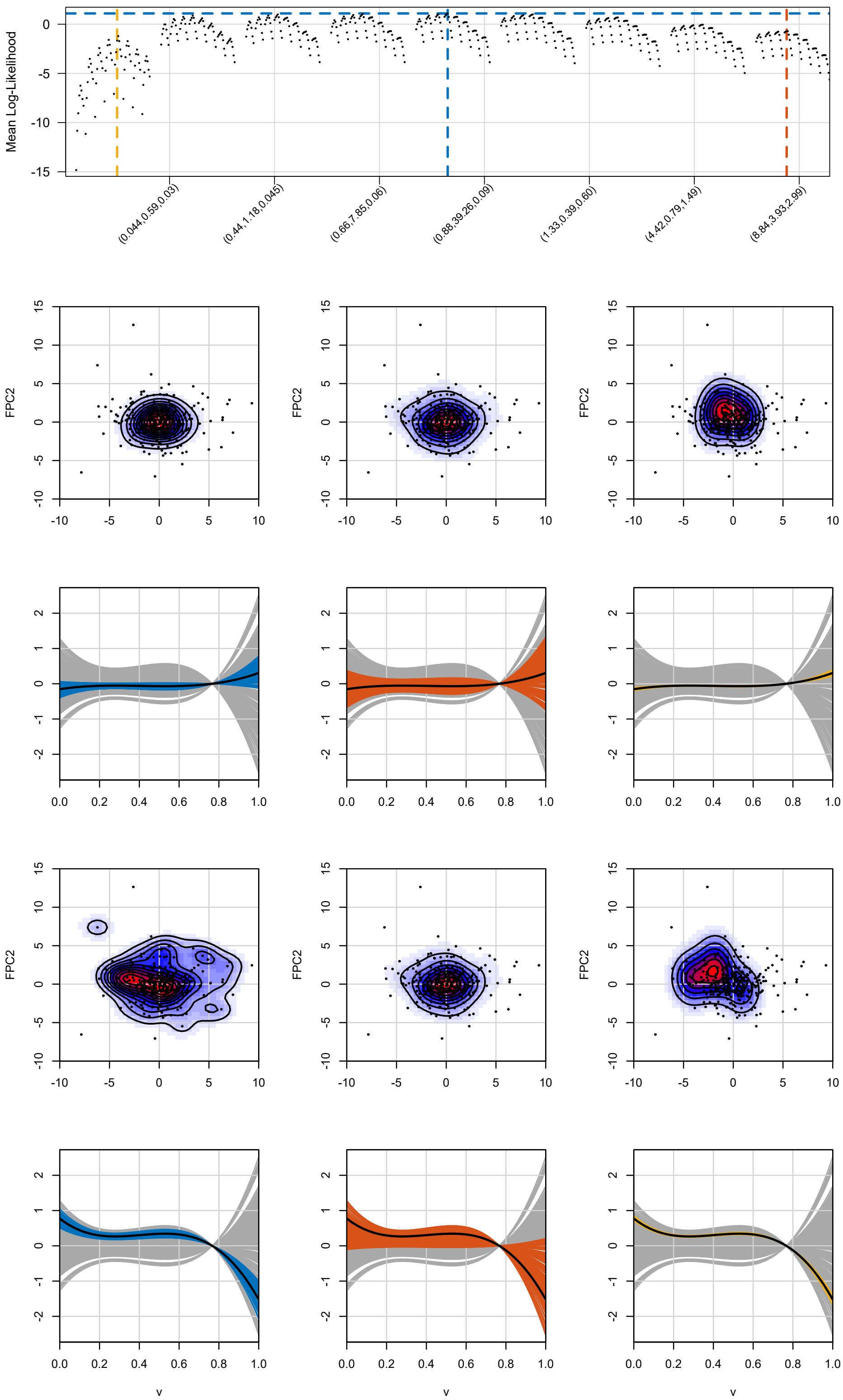

Figure 4.1. Top figure: average log-likelihood obtained for each combination of bandwidth parameters used in the cross-validation procedure. Second and third figures: conditional density estimated for the 2-dimensional projection of a simulated functional dataset and functional covariates selected by the model using different sets of bandwidth parameters for a low-volatility period. Fourth and fifth figures: conditional density estimated for the FPC scores of a simulated functional dataset in a high-volatility period for three different sets of bandwidth parameters and functional covariates selected by the model.

4.2.2.1. Prediction regions for functional data

In the functional data framework the natural extension of prediction intervals are prediction bands (Antoniadis, Brossat, et al., 2016; Vilar, Aneiros, et al., 2018), which consist of an upper and lower limit functions, denoted by $\mathrm{U}_t^{1-\alpha}$ and $\mathrm{L}_t^{1-\alpha}$ respectively, defined in such a way that the functional observations Y_t satisfy

$$P\left(Y_t \in \mathrm{R}_t^{1-\alpha}\right) = 1 - \alpha, \tag{4.10}$$

where $\mathrm{R}_t^{1-\alpha}$ denotes the prediction region defined by the upper and lower limits $\mathrm{U}_t^{1-\alpha}$ and $\mathrm{L}_t^{1-\alpha}$ (Lei et al., 2015). Hence, the area between the upper and lower limits of the prediction bands will cover curve Y_t with probability $1 - \alpha$. This region can be used as a tool to quantify the variability associated with the forecasts of a given model, hence it can be used to compare two forecasting models based on the uncertainty associated with each model.

Due to the infinite dimension of the functional data, requiring Equation 4.10 to hold is very ambitious. In many practical applications the user is only interested in the main structural features of the curves, so most methods found in the literature aim at estimating point-wise prediction bands, which are defined in such a way that, for every discretization point v of the curves, the functional observations satisfy

$$P\left(\mathrm{L}_t^{1-\alpha}(v) \leq Y_t(v) \leq \mathrm{U}_t^{1-\alpha}(v)\right) = 1 - \alpha, \quad v \in V. \tag{4.11}$$

The vast majority of probabilistic models for functional data proposed in the literature follow this approach, as can be seen in Antoniadis, Brossat, et al. (2016), Cabrera and Schulz (2017), Vilar, Aneiros, et al. (2018), and Raña, Vilar, et al. (2018), among others.

Even though point-wise prediction bands are very informative and can be more than enough in many practical applications, this approach may not be appropriate in some cases, as it assumes that the measurements of the curves at different discretization points are uncorrelated. Ignoring the relation exhibited by different sections of the curves may lead to prediction bands that are not realistic and do not reflect the complete uncertainty of the forecasts. A different approach for obtaining prediction regions was propposed in Lei et al. (2015), where instead of obtaining point-wise prediction bands the authors proposed a method to define regions that satisfy

$$P\left(\Pi\left(Y_t\right) \in \Pi\left(\mathrm{R}_t^{1-\alpha}\right)\right) = 1 - \alpha, \tag{4.12}$$

where $\Pi\left(.\right)$ denotes the projection operator of a set of curves into a finite dimensional space, such as the subspace spanned by the first Functional Principal Components or any orthonormal basis of L^2. The authors found that prediction bands obtained using this method can capture shape features of the curves with much more detail than their point-wise counterparts. This definition is consistent with the projection-based density that is being used in this thesis as an approximation to the density of a functional dataset, as they both aim at estimating the distribution of the curves when projected on a finite-dimensional subspace. Hence, this method will be used in order to define the prediction regions associated with the conditional density estimated by the probabilistic model proposed in Section 4.2.1.

Once the proposed probabilistic model has been fitted, it can be used to obtain estimations of the conditional density of the projections of the curves $\hat{p}^{(M)}_{\mathscr{H}}(\boldsymbol{e}|\boldsymbol{X})$ given a new set of exogenous covariates $\boldsymbol{X}$. The estimated density satisfies

$$\int_{\mathbb{R}^M} \hat{p}^{(M)}_{\mathscr{H}}(\boldsymbol{z}|\boldsymbol{X})\, d\boldsymbol{z} = 1. \tag{4.13}$$

Given a confidence level α, the prediction region at level α can be obtained by finding a subset $\mathrm{C}_t^{1-\alpha} \subset \mathbb{R}^M$ such that

$$P\left(\Pi\left(Y_t\right) \in \mathrm{C}_t^{1-\alpha}\right) = \int_{\mathrm{C}_t^{1-\alpha}} \hat{p}^{(M)}_{\mathscr{H}}(\boldsymbol{z}|\boldsymbol{X})\, d\boldsymbol{z} = 1-\alpha. \tag{4.14}$$

The estimation of this region can be achieved using a method similar to the one proposed in Hyndman and Shang (2010) to define functional boxplots. Once the region $\mathrm{C}_t^{1-\alpha}$ has been estimated the functional prediction region $\mathrm{R}_t^{1-\alpha}$ can be obtained as

$$\mathrm{R}_t^{1-\alpha} = \left\{ Z = \widehat{\mu} + \sum_{j=1}^{M} \widehat{\xi}_j \widehat{\phi}_j \in L^2 \,\middle|\, \left(\widehat{\xi}_1, \ldots, \widehat{\xi}_M\right) \in \mathrm{C}_t^{1-\alpha} \right\}, \tag{4.15}$$

where $\widehat{\mu}$ and $\widehat{\phi}_j$ are the sample mean function and the functional principal components of the FPC decomposition of the curves, respectively. Equation 4.15 can be interpreted as the set of curves Z obtained by using the points of $\mathrm{C}_t^{1-\alpha}$ as coefficients of the basis expansion that define the curve Z. Therefore, prediction bands can be directly obtained as the envelope of the prediction region $\mathrm{R}_t^{1-\alpha}$:

$$\mathrm{L}_t^{1-\alpha}(v) = \min\left\{Z(v) \,|\, Z \in \mathrm{R}_t^{1-\alpha}\right\}, \quad \mathrm{U}_t^{1-\alpha}(v) = \max\left\{Z(v) \,|\, Z \in \mathrm{R}_t^{1-\alpha}\right\}. \tag{4.16}$$

The proposed methodology can be used to define prediction regions from the estimated conditional density of the residuals of a given functional forecasting model. As these predictive regions quantify the uncertainty of the model, the deterministic forecasts can be accompanied by the prediction regions (or the associated prediction bands) to provide a measure of the expected variability of a future estimation. In addition, the prediction region obtained is based on the projection-based of the curves, so it is able to capture relevant shape features of the curves. This is a desirable property that may not be satisfied by point-wise prediction regions proposed in the literature.

4.2.2.2. Scenario generation for functional data

As mentioned at the beginning of this section, scenario-based forecasts are of utmost importance in many practical applications, as they are able to represent the range of possible future values of the data being modelled, which can be used to quantify the risk associated with the forecasts. Unlike prediction bands, which only provide information about the boundaries of the expected value, scenarios represent the full spectrum of the possible outcomes of the future value being estimated by the model. Having a scenario-generation model is of utmost importance for stochastic programming models (Kaut and Wallace, 2007), which use these scenarios to model the distribution of the input data, as illustrated by the offering optimization model proposed in Baillo et al. (2004). The quality of these scenarios directly influences the quality of the solution of these stochastic programming models. In addition, scenarios of functional time series are commonly used as input for other optimization models, such as the market equilibrium model proposed in Campos et al. (2016). This model use scenarios of residual demand curves as inputs to include information about the offering behaviour of the different market agents into the model, taking into account the market uncertainty. As such, there is a growing interest in developing new scenario generation methodologies that are able to produce accurate scenarios that represent the variability in the evolution of a quantity of interest .

In the functional data literature scenarios have been usually obtained through bootstrap simulation, as illustrated in Hyndman and Shang (2009) where the authors obtain scenarios of demographic curves using bootstrap replication methods on the residuals of a fitted dimensionality-reduction model. A similar approach is found in Wagner-Muns et al. (2018), where recent historical residuals of a SARIMA model fitted to the scores of the first FPC are sampled with replacement to generate functional scenarios. In Castro et al. (2005) the authors decompose the residuals of a fitted ARH(1) model into their first FPC to then obtain bootstrap samples of the residuals, which are then used to generate scenarios of the functional time series. Recently there have been some improvements over these methods, as seen in Antoniadis, Brossat, et al. (2016) where the authors propose a weighted bootstrap approach to obtain a more accurate representation of the volatility of the forecasts, which allows them to generate scenarios of daily electricity load curves that represent the uncertainty associated with the forecasts. In addition, Vilar, Aneiros, et al. (2018) proposes residual-based bootstrap methods that are able to account for heteroscedasticity in the residuals of a functional forecasting model, which allows them to generate accurate scenarios of the series of electricity demand and price curves in the spanish electricity market.

The proposed probabilistic model takes a different approach for scenario generation. Instead of relying on residual-based bootstrap techniques, the estimated conditional density of the projections of the residuals of the model is used to generate new residual curves that follow the density estimated by the probabilistic model. Equation 4.4 can be interpreted as a weighted sum of multidimensional kernel functions, where each kernel is centered at the coordinates of the residuals of the forecasting model and the weight terms are given by the historical values of the exogenous variables used as inputs for the probabilistic model. As such, given a new set of covariates $\boldsymbol{X}_{T+1} = \left(X^1_{T+1}, \ldots, X^Z_{T+1}\right)$, functional scenarios that follow the estimated conditional density can be obtained as follows:

1. Following Equation 4.5, compute the weights $\omega_t\left(\boldsymbol{X}_{T+1}\right)$ for $t = 1, \ldots, T$.

2. The weights $\omega_t\left(\boldsymbol{X}_{T+1}\right)$ are non-negative values that satisfy that $\sum_{t=1}^{T} \omega_t\left(\boldsymbol{X}_{T+1}\right) = 1$, with higher values of $\omega_i\left(\boldsymbol{X}_{T+1}\right)$ indicating that the future situation is similar to the historical situation given by the pair $(\boldsymbol{e}_i, \boldsymbol{X}_i)$. Hence, a center $\boldsymbol{e}^*$ of the kernels of Equation 4.4 can be chosen according to the distribution induced by the the weights:

$$P\left(\boldsymbol{e}^* = \boldsymbol{e}_i \mid \boldsymbol{X}_{T+1}\right) = \omega_i\left(\boldsymbol{X}_{T+1}\right). \tag{4.17}$$

3. Once the center of the multidimensional kernel has been selected, a random value $\boldsymbol{\xi}^*$ can be drawn from the distribution of the kernel centered at $\boldsymbol{e}^*$ and with a bandwidth parameter H, which has been selected previously using the proposed cross-validation method. This is the standard sampling method from kernel density estimation models (Silverman, 1986).

4. The functional residual curve $\widehat{\varepsilon}^*_{T+1}$ can be obtained from the new set of coefficients $\widehat{\boldsymbol{\xi}}^*_{T+1}$ using the truncated Karhunen-Loéve expression given by Equation 2.33:

$$\widehat{\varepsilon}^*_{T+1} = \widehat{\mu} + \widehat{\boldsymbol{\xi}}^*_{T+1}\widehat{\boldsymbol{\phi}}, \tag{4.18}$$

   where $\widehat{\phi}$ denotes the first FPC of the residuals.

5. The functional scenario associated with this particular residual can be defined as follows:

$$\widehat{Y}^*_{T+1} = \widehat{Y}_{T+1} + \widehat{\varepsilon}^*_{T+1}. \tag{4.19}$$

6. This process is repeated B times to obtain a set of functional scenarios.

The set of curves obtained by this procedure are drawn according to the conditional density estimated for the residuals of the forecasting model, providing a measure of the uncertainty associated with its forecasts. This method differs significantly from the vast majority of scenario generation methods proposed in the functional data literature, as this methodology is able to generate new curves that have not been previously observed. The residual-based bootstrap methods proposed in the literature rely on resampling the observed error curves, so they are not able to extrapolate this information to generate new curves. In addition, the information about the exogenous covariates is taken into account in the scenario generation process, so the functional scenarios will follow the volatility estimated by probabilistic model given the new set of explanatory variables.

4.3. Quality measures for functional probabilistic models

The last sections have been devoted to introducing a new probabilistic model for functional time series and to illustrate how the estimated density of the model can be used to obtain prediction regions and functional scenarios of the series being modelled. However, once the probabilistic forecasts have been obtained, it is of utmost importance to quantify the quality of those estimations. Hence, it is necessary to define metrics that can be used to measure the predictive capabilities of the fitted forecasting model. When the forecasts are given as a set of functional scenarios one should expect that they exhibit similar properties as the actual observations of the series, so the practitioner is usually interested in measuring the likelihood of this set of scenarios.

Suppose that two different scenario sets are obtained for the same functional time series. In order to choose one of the scenario sets, some kind of performance measure is needed to select the most suitable set of scenarios. When dealing with non-functional time series, this is usually done by means of a probabilistic scoring rule: a metric that provides summary measures for the evaluation of probabilistic forecasts by assigning a numerical score based on the predictive distribution and on the actual outcome of the series (Nowotarski and Weron, 2018), which minimized the scoring rule. The most common scoring rules found in the literature are the Continuous Ranked Probability Score or CRPS (Matheson and Winkler, 1976) which computes the mean square error of a predicted cumulative density function and the empirical cumulative density function of the data, and it is often seen as an extension of the mean absolute error to the case of probabilistic forecasts; the Pinball Score (Hong and Fan, 2016) which is often used to evaluate the accuracy of quantile forecasting models; or the Winkler Score (Winkler, 1972)

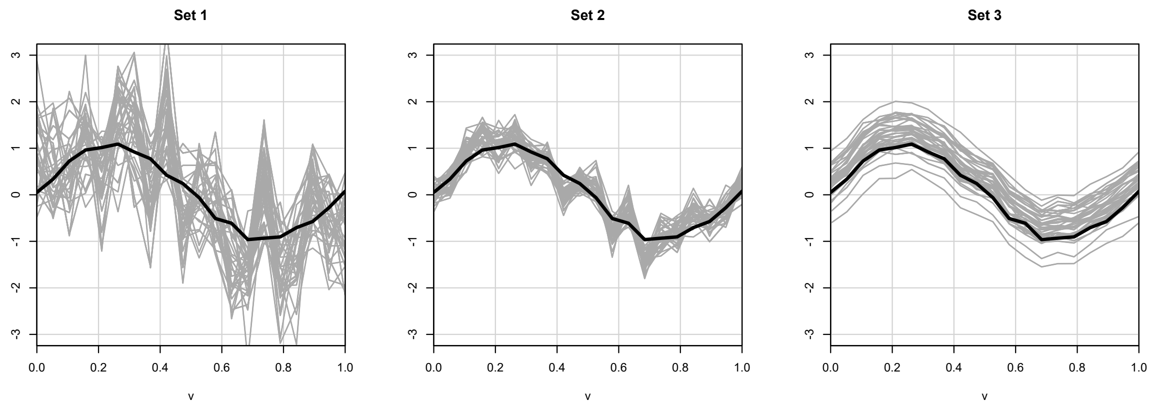

Figure 4.2. Three sets of synthetic scenarios that will be used to illustrate different probabilistic performance measures. The black curve denotes the actual functional observation, whereas each grey curve represents each functional scenario.

which allows for a joint assessment of the unconditional coverage of a given prediction interval and the width of said interval. The work Nowotarski and Weron (2018) reviews probabilistic models for estimating electricity prices and includes an in-depth analysis of the most common probability scores used in probabilistic electricity price forecasting.

These metrics provide a measure of the sharpness of the distribution, a measure of the concentration of the predictive distribution. This refers to how tightly the prediction distribution covers the actual one. For example, the Winkler Score favours narrow prediction intervals, but it also penalizes intervals that do not capture the real observation. As such, according to this scoring rule, the best probabilistic model will cover the real series most of the times with narrow prediction intervals. Nevertheless, in the functional data literature, concepts such as the likelihood of a set of curves are not well defined. As such, there is not a defined set of probabilistic scoring rules to compare different scenario sets. Authors often rely on point-wise metrics to evaluate the probabilistic capabilities of their functional scenarios, without taking into account the functional nature of the curves.

This section will provide a review of the probabilistic scores found in the functional data literature, including examples that illustrate each metric. In addition, a new metric to evaluate the quality of a set of functional scenarios is proposed, that is able to generalize the concept of maximum likelihood to the functional data framework. Three sets of synthetic scenarios will be used to illustrate this section: the first and second set of scenarios are obtained by adding a high (low) variance Gaussian white noise process to each discretization point of the real curves, respectively. The third set of scenarios are obtained by shifting the mean of the real curve and adding a small noise process to each simulated curve. As seen in Figure 4.2, only the third set of scenarios retain the original shape of the real curves. This is a desirable property of the scenarios: when simulating future curves, it is desirable that the shape of the curves is preserved.

4.3.1. Point-wise Pinball Loss and CRPS

In Narajewski and Ziel (2019) the authors proposed a methodology to generate scenarios of transaction counting curves in the German Intraday Market. In order to select the optimal

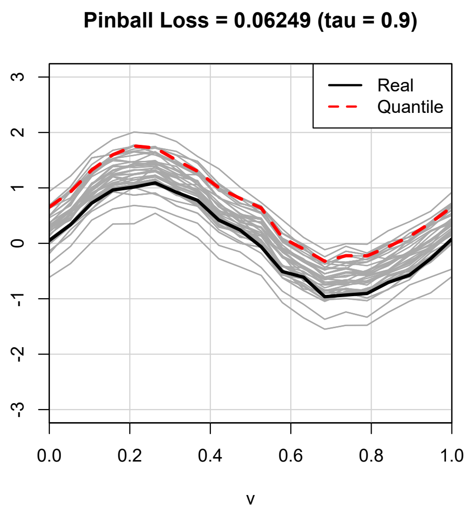

Figure 4.3. In red, point-wise 90% quantile curve obtained from the set of functional scenarios.

distribution of the scenarios, a functional version of the Pinball Loss function was proposed. Given a confidence level α, the functional Pinball Loss is defined as

$$\mathrm{fPB}_{\alpha,t} = \int \mathrm{pb}_\alpha \left(Y_t(v), \mathrm{Q}_t^\alpha(v)\right) dv, \tag{4.20}$$

where Y_t is the real functional observation at time t, pb_α is the Pinball Loss function at level α for a given quantile forecast at level α of exceedance q_α:

$$\mathrm{pb}_\alpha(y, q_\alpha) = (1-\alpha)(q_\alpha - y)\mathbb{1}_{\{y<q_\alpha\}} + \alpha(y - q_\alpha)\mathbb{1}_{\{y \geq q_\alpha\}}, \tag{4.21}$$

and Q_t^α is the point-wise α percentile curve of the functional scenarios.

As such, the procedure for obtaining the functional Pinball Loss can be summarized as:

1. For each discretization point, obtain the α percentile of the scenarios.
2. A point-wise α percentile curve is obtained.
3. The Pinball loss function $\mathrm{pb}_\alpha(.)$ is applied to each discretization point of the difference between the real curve and the percentile curve.
4. The resulting curve is integrated in its domain to obtain the functional Pinball Loss.

The Functional Pinball Loss for the whole functional time series can be obtained as the mean value of the Functional Pinball Loss across all the curves:

$$\mathrm{fPB}_\alpha = \frac{1}{T}\sum_{t=1}^{T} \mathrm{fPB}_{\alpha,t}. \tag{4.22}$$

In addition, the authors propose to approximate the CRPS metric to functional data using the Functional Pinball Loss:

$$\mathrm{fCRPS} = \frac{1}{R}\sum_{\alpha \in r} \mathrm{fPB}_\alpha, \tag{4.23}$$

for an equidistant grid of probabilities r between 0 and 1 of size R.

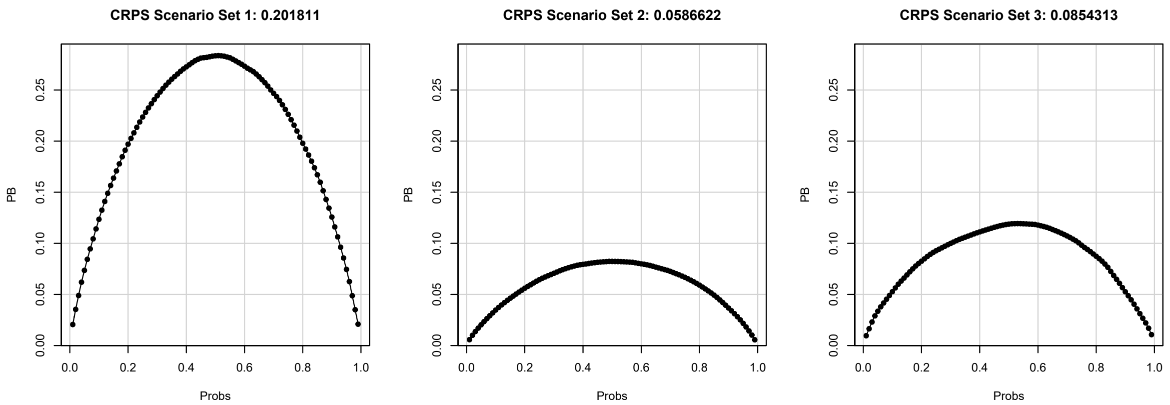

Figure 4.4. Mean functional Pinball Loss for each of the 3 sets of scenarios across an equidistant grid of probabilities. The Functional CRPS is obtained as the mean value of the Pinball Loss.

By looking at their definition, it can be concluded that these metrics offer a point-wise view of the curves, as they only look at the point-wise percentile curve obtained from the scenarios without taking into account the shape of each scenario curve. Hence, this metric will not discriminate scenarios that exhibit a much different shape than the real curve, as long as the point-wise coverage of the scenarios is optimal. This is illustrated in Figure 4.4, where the functional Pinball Loss fPB_α across the grid of probabilities $\alpha \in \{0.01, 0.02, \ldots, 0.99\}$. The lowest fCRPS is obtained by the second set of scenarios, showing that the shape of the scenario curves is not as relevant as the coverage of the scenarios.

4.3.2. Point-wise Winkler Score

When forecasting functional time series, the authors are often interested in obtaining prediction bands as a way of quantifying the uncertainty associated with their forecasts. As such, they tend to use evaluation metrics that measure the quality of the prediction bands. For example, Antoniadis, Paparoditis, et al. (2006) and Antoniadis, Brossat, et al. (2016) measure the empirical coverage of the prediction bands to ensure that the bands achieve the desired point-wise coverage probability. In order to enhance the interpretation of the prediction bands, in Vilar, Aneiros, et al. (2018) the authors use the point-wise Winkler Score, that not only measures the coverage of the bands but also takes into account the width of the prediction bands. Unlike the functional Pinball Loss and the Continuous Ranked Probability Score, the Winkler Score measures the area covered between the prediction bands. The authors define point-wise prediction bands from the simulated scenarios in an straightforward manner: for each discretization point, the upper and lower limit of the band are obtained as the percentiles that capture $\alpha\%$ of the sample. Figure 4.5 illustrates this procedure, showing the bands that each set of scenarios generate.

Given a real curve and a symmetric prediction band, the point-wise functional Winkler Score is obtained as the integral of the function obtained after applying the scalar Winkler Score function to each discretization point. More formally, let Y_t be a functional observation and let $\mathrm{U}_t, \mathrm{L}_t$ be the upper and lower prediction bands, respectively. The functional Winkler Score is defined as:

$$\mathrm{fWS}_{\alpha,t} = \int \mathrm{ws}_\alpha \left(Y_t(v), \mathrm{U}_t(v), \mathrm{L}_t(v) \right) dv, \tag{4.24}$$

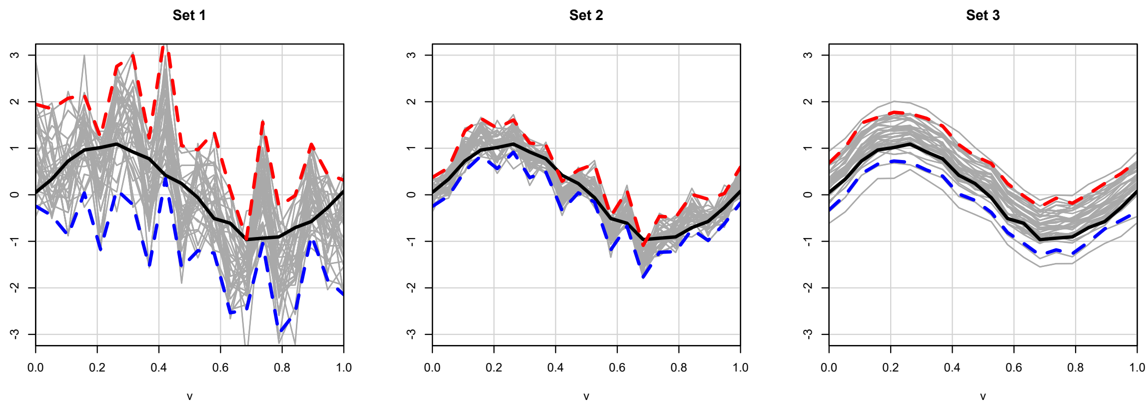

Figure 4.5. Point-wise symetric 90% prediction bands obtained from the 3 sets of functional scenarios. The red and blue dashed lines indicate the upper and lower bounds of the prediction bands, respectively.

where ws(.) denotes the scalar Winkler Score loss function:

$$\mathrm{ws}_\alpha(y,u,l) = (u-l) + \frac{2}{\alpha}(l-y)\,\mathbb{1}_{\{y<l\}} + \frac{2}{\alpha}(y-u)\,\mathbb{1}_{\{y>u\}}. \tag{4.25}$$

The functional Winkler Score for the whole functional time series can be obtained as the mean value of the functional Winkler Score across all the curves:

$$\mathrm{fWS}_\alpha = \frac{1}{T}\sum_{t=1}^{T} \mathrm{fWS}_{\alpha,t}. \tag{4.26}$$

Similarly to the fCRPS case, the functional Winkler Score does not measure differences in the shape of the scenario curves with respect to the real curves. Figure 4.6 shows the Winkler Score for the 3 synthetic set of functional scenarios: the best set according to this criterion is again Set 2, despite having scenarios that do not follow the much smoother dynamics of the real curves. As such, it can be concluded that the fWS_α only measures the coverage of the pointwise prediction bands and its width, without taking into account the shape of the modelled functional time series.

4.3.3. Functional Log-score

As seen in the previous sections, the vast majority of the performance measures for functional scenarios proposed in the literature are focused on the width and coverage of the prediction bands, neglecting the shape of the curves being modelled. The simulation carried out illustrates this problem, as if one had to choose the best set of functional scenarios by minimizing the existing probabilistic scores, a set of scenarios that do not capture the shape of the curves would have been selected. This highlights the need for a new probabilistic metric that measures both the coverage and the shape of the scenarios.

As such, a new probabilistic score is proposed, based on the projection-based surrogate density introduced in Delaigle and Hall (2010). The main idea is to project the functional scenarios on the subspace spanned by the first functional principal components of the modelled functional time series. Once projected, the density of the scenarios is estimated via a non-parametric density estimator, and the fitted probability density function is used to estimate the likelihood of the projection of the real curve given by the set of scenarios. This is an

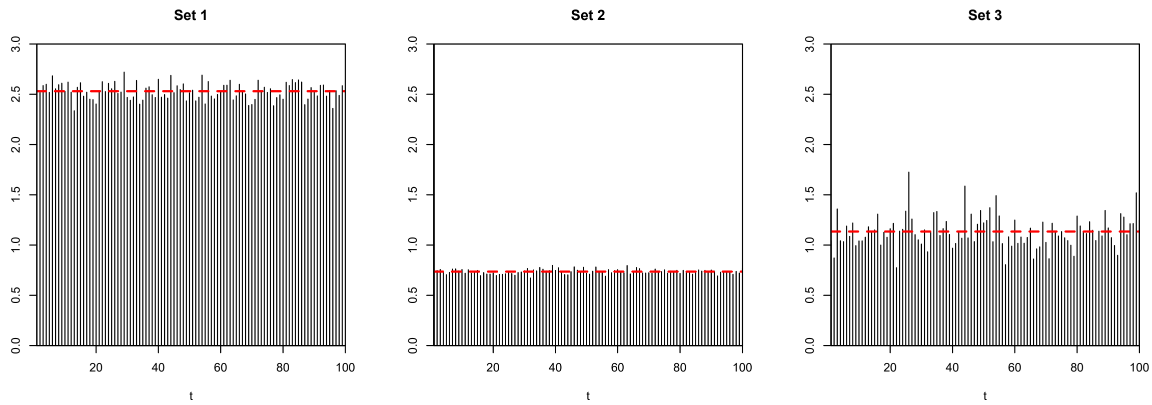

Figure 4.6. Functional Winkler Score for each observation of the functional time series. The mean value is shown in red.

approximation to the concept of maximum likelihood for functional data: the best set of scenarios will be the one that gives most likelihood to the observation of the real curve.

The definition of the proposed metric is as follows: given Y_t a functional observation and $\{\widehat{Y}_t^p\}_{t=1}^P$ a set of functional scenarios for that functional time series, the score can be defined as

$$\text{fLS}_t = -\log\left(\widehat{f}\left(\boldsymbol{y}_t|\widehat{\boldsymbol{y}}_t^1, \widehat{\boldsymbol{y}}_t^2, \ldots, \widehat{\boldsymbol{y}}_t^P\right)\right), \tag{4.27}$$

where $\widehat{f}(.|\widehat{\boldsymbol{y}}_t^1, \widehat{\boldsymbol{y}}_t^2, \ldots, \widehat{\boldsymbol{y}}_t^P)$ is the estimated density of the projection of the functional scenarios $\widehat{\boldsymbol{y}}_t^1, \widehat{\boldsymbol{y}}_t^2, \ldots, \widehat{\boldsymbol{y}}_t^P$ in the subspace spanned by the first FPC of the real functional time series, and $\boldsymbol{y}_t$ stands for the coefficients of the functional observation Y_t in its FPC-decomposition:

$$Y_t \approx \widehat{\mu}_{\mathcal{Y}} + \boldsymbol{y}_t^{\intercal}\Phi, \qquad \widehat{Y}_t^P \approx \widehat{\mu}_{\mathcal{Y}} + \widehat{\boldsymbol{y}}_t^{p\intercal}\Phi, \quad \forall\, p = 1, \ldots, P \tag{4.28}$$

Following a similar approach to the other metrics, the functional Log-Score for the whole functional time series can be obtained as the mean value of the Log-Score across all the curves:

$$\text{fLS} = \frac{1}{T}\sum_{t=1}^{T} \text{fLS}_t. \tag{4.29}$$

This formulation is similar to the classical logarithmic score (Gneiting and Raftery, 2007), providing a way of measuring the pseudo-likelihood that a set of functional scenarios give to the real curve. Scenarios with a low fLS will be preferred, since the likelihood that they give to the real curves is higher.

This approach is illustrated with an application to the synthetic datasets that have been used in the previous sections. The estimated densities for the 3 sets of scenarios for a given time instant are shown in Figure 4.7 As can be seen, the estimated density function for Set 1 has a higher variance than the other 2 sets, hence giving a log-score of -2.04 to the real curve, much lower than the other two densities. Due to the non-smooth nature of the scenarios of Set 2, the distance of the projections of these curves in the FPC-spanned subspace is higher than those of Set 3, hence the log-score associated with Set 2 is -5.417, a higher value than the log-score of Set 3: -7.43. Hence, this criterion will select Set 3 as the scenarios that better approximate the distribution of the real functional time series.

	fCRPS	fWS_{90}	fLS
Set 1	0.202	2.53	-1.868
Set 2	**0.059**	**0.74**	-4.992
Set 3	0.085	1.13	**-6.499**

Table 4.1. Mean values of the three functional probabilistic scores for the 3 synthetic scenario sets.

Table 4.1 shows the mean average values of the three functional probabilistic scores for the 3 synthetic scenario sets obtained for 100 observations of the simulated functional time series. As can be seen, the non-smooth curves of Set 2 achieve the best results in both the functional CRPS and the Winkler score, which illustrates how these two metrics can be inadequate for quantifying the quality of the functional scenarios. Hence, the proposed log-score metric seems more adequate for this task, as it is able to account for shape discrepancies between the real curves and the scenarios.

This concludes the review of quality measures for functional probabilistic models. The different metrics proposed in the literature have been analyzed, showing the benefits and drawbacks of each method. A functional version of the log-score metric has been proposed, which is able to quantify the likelihood of a set of functional scenarios given a real observation. This new tool can be used to asses the quality of different scenario generation methods for functional time series, which is of utmost importance in numerous real-world applications.

4.4. Simulation study

The previous sections have introduced a novel probabilistic model for functional data and several probabilistic scores that can be used to compare the adequacy of a set of functional probabilistic models. This section is devoted to illustrating the applicability of the proposed model using several synthetic datasets. Hence, the usefulness of the model can be validated in a controlled environment, providing some insight about the characteristics of the probabilistic methods developed in this chapter. Two different probabilistic reference models commonly used in the functional data literature, the unconditional kernel density model (Delaigle and Hall, 2010) and the residual-based bootstrap model (Rossini and Canale, 2019), will also be included in this simulation study to quantify the improvements made by the proposed model.

The first synthetic dataset included in this study is comprised of a set of real functional observations $\{Y_t\}_{t=1}^{300}$ and a set of functional residuals $\{\varepsilon_t^{(1)}\}_{t=1}^{300}$ whose distribution exhibits a direct relationship with the levels of an scalar covariate $\{x_t^1\}_{t=1}^{300}$: higher values of x_t^1 are associated with high-volatility periods of variable $\varepsilon_t^{(1)}$, as illustrated in Figure 4.8. The estimated series is defined as $\widehat{Y}_t = Y_t - \varepsilon_t^{(1)}$. Although the residuals do not exhibit any kind of serial correlation, as evidenced by the FACF/FPACF shown in Figure 4.9, the moments of higher order of the functional variables display a strong relationship with an exogenous variable. As such, a model that it is able to identify and capture this relation will be able to obtain more accurate probabilistic statements about the series being modelled, as it will be able to detect low-volatility periods of time and define narrower prediction regions for those periods.

Once the dependency between the distribution of the residuals and the covariates has been identified, the conditional probabilistic model can be fitted. The projection-based density of the residuals $\varepsilon_t^{(1)}$ will be obtained using their first two FPC, as they are able to explain more

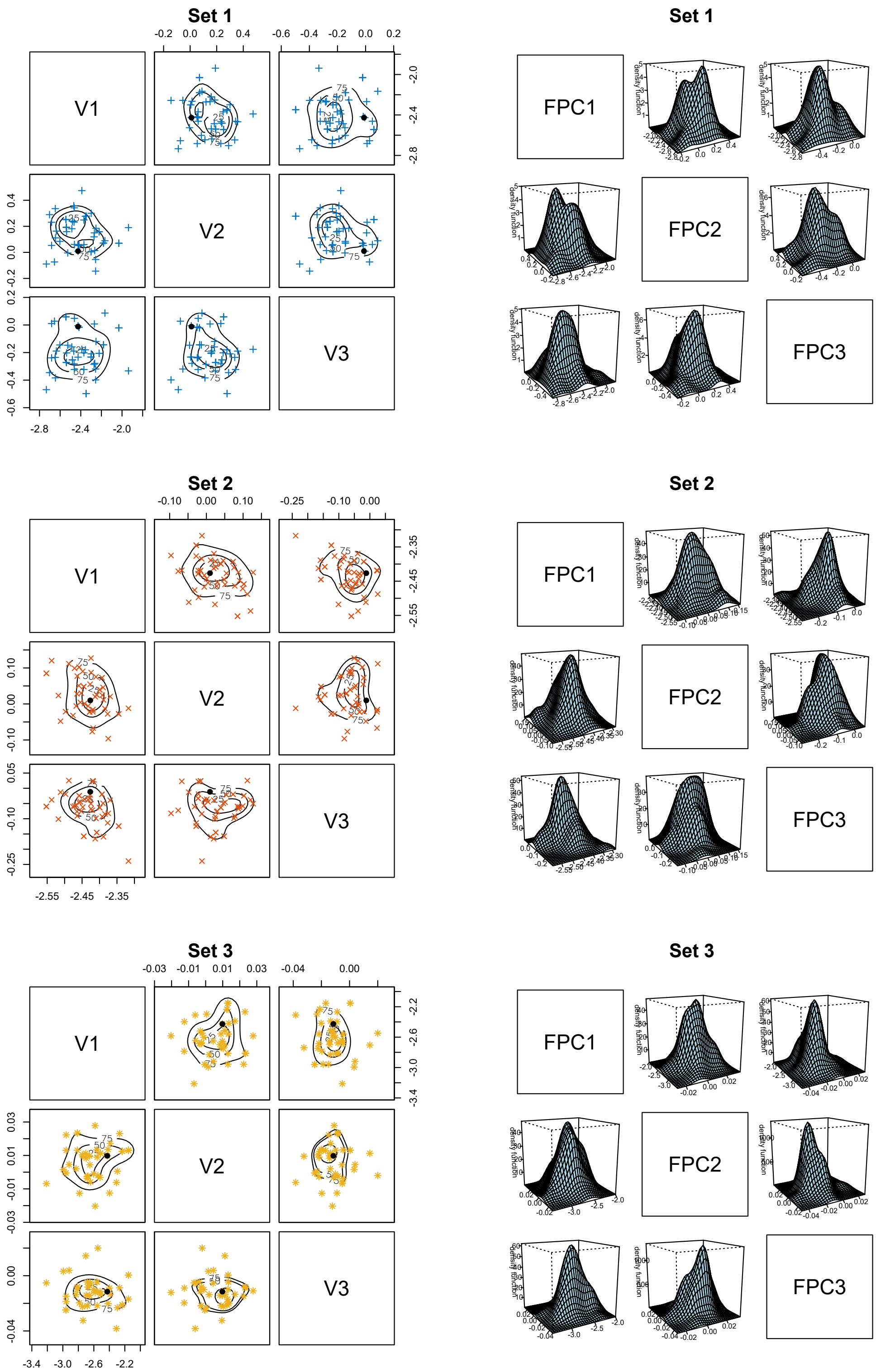

Figure 4.7. Left: Estimated density of the projections of the functional scenarios for the 3 sets. The coloured dots represent the projection of the scenarios, the black dot is the FPC-projection of the real functional observation. Right: Pairwise visualization of the estimated density surfaces

than 95% of the variance of the data. The formulation of the conditional density model for this dataset is as follows:

$$\hat{p}^{(2)}_{\mathscr{H}}\left(\boldsymbol{e^{(1)}}|x\right) = \sum_{t=1}^{300} \omega_t\left(x\right) \mathbf{K}_{\boldsymbol{H}}\left(\boldsymbol{e}^{(1)} - \boldsymbol{e}^{(1)}_t\right), \tag{4.30}$$

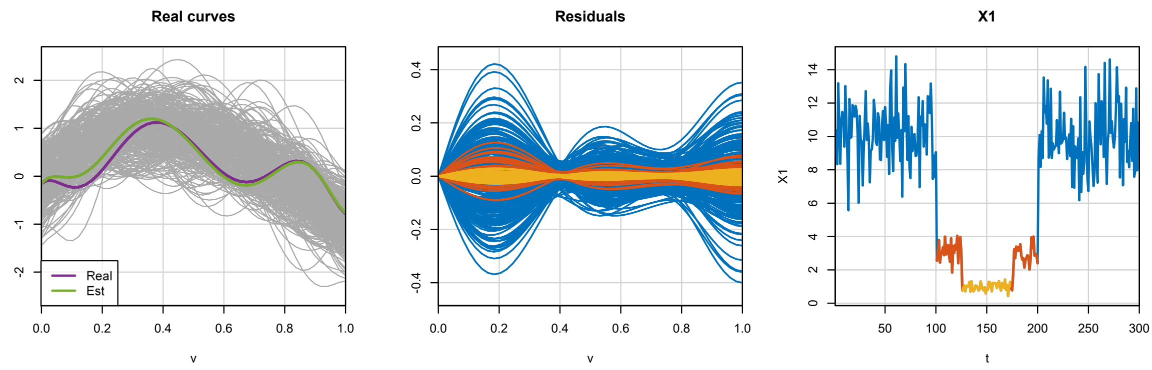

Figure 4.8. Left: synthetic functional dataset used in the first simulation. Grey curves represent the real observations, whereas the coloured curves show examples of the real and estimated curves. Middle: set of simulated functional residuals. The curves have been coloured following the different levels of the scalar covariate x_t included in the simulation. Right: exogenous variable. Regions with a different mean level are displayed using different colors.

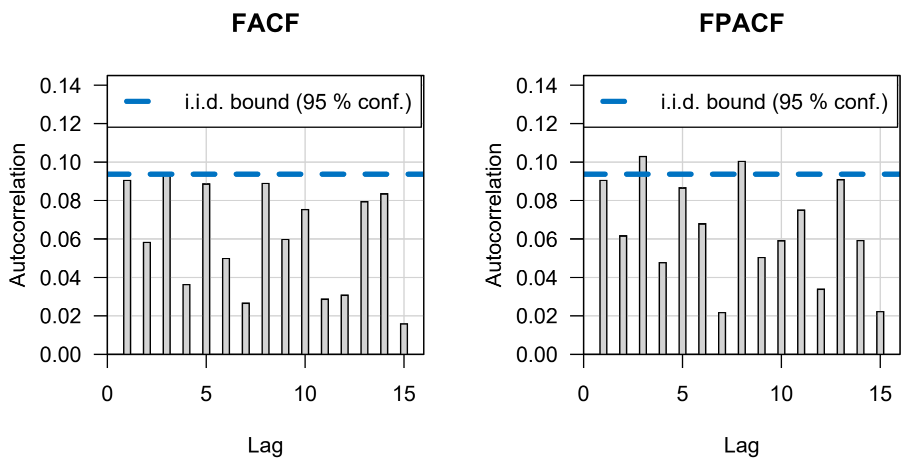

Figure 4.9. FACF/FPACF of the series of functional residuals $\varepsilon_t^{(1)}$ included in the first simulation. The dotted line represents the 99% white noise prediction bounds for the FACF/FPACF coefficients under the assumption of white noise

where the weight function of the scalar covariate is given by

$$\omega_t\left(x\right) = \frac{K_{h_1}\left(\|x^1 - x_t^1\|\right)}{\sum_{j=1}^{300} K_{h_1}\left(\|x^1 - x_j^1\|\right)}. \tag{4.31}$$

The scores of the FPC decomposition of $\varepsilon_t^{(1)}$ are denoted as $\boldsymbol{e}_t^{(1)}$, and the family of bandwidth parameters is denoted as $\mathscr{H}$. In order to obtain estimations with this model, it is necessary to select an optimal set of bandwidth parameters that maximize the likelihood of the model. Hence, the proposed cross-validation method is used to optimize these parameters. A grid of different bandwidth values are used in this procedure to select the ones that give a better likelihood to the observed data, as illustrated in Figure 4.10. The values selected by the cross-validation procedure ensure that the model is neither overfitted or underfitted, as the extreme bandwidth values are associated with a lower log-likelihood.

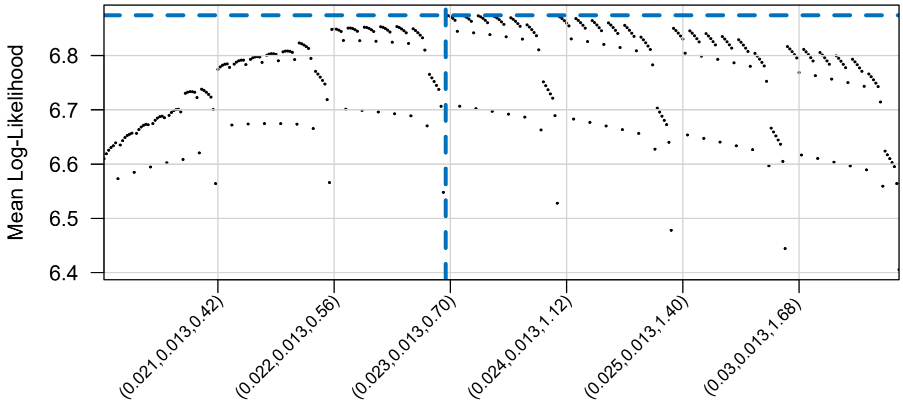

Figure 4.10. Average log-likelihood obtained for each combination of bandwidth parameters used in the cross-validation procedure for the first simulation.

Once the training of the conditional probabilistic model has been completed, it can be used to estimate the prediction regions given by the estimated conditional density of the residuals. Due to how the data has been generated, periods with low values of the exogenous covariate are associated with a lower dispersion of the residuals. Hence, if the proposed model has been able to capture this behaviour, the prediction regions obtained will be narrower than the regions defined by the other reference models included in this study, as they are not able to include information about exogenous variables in their formulation. This is validated by the prediction regions shown in Figure 4.11: the top images correspond to time periods where the value of the exogenous variable is high. Therefore, the variance of the residuals is quite high, so the three models included in this study obtain similar prediction regions. As it can be seen, the coverage of the unconditional models is not enough in this case, as even the 95% prediction bands do not cover the real curve in the range $[0.4, 0.6]$. This can be interpreted as the result of the reference models not being able to distinguish between the different volatility states given by the scalar covariate, which results in prediction regions with an unsatisfactory performance. Unlike these models, the coverage of the conditional model seems better, as it is able to capture the real curve inside the 95% region obtained from the estimated conditional density of the data. A different situation is illustrated in the bottom images of Figure 4.11, where a low volatility period of the residuals is shown. As the reference models assume that the residuals are i.i.d. the prediction regions that they obtain have a width similar to the ones obtained in the top images, which are too wide for this period. As the dispersion of the residuals is closely related with the values of the scalar covariate included in this study, the conditional model detailed in this chapter is able to model this relationship and to define narrower prediction regions with an appropriate coverage level for the real curve. This illustrates the benefits of employing the proposed model when modelling situations where the dispersion of the residuals can be explained using some external variable, generalizing the reference models proposed in the functional data literature.

The second simulated dataset uses the same real curves as the previous case, but instead of using a scalar exogenous variable that conditions the distribution of the synthetic residuals a functional variable will be used instead. However, unlike the previous case, the variance of the residuals will not be correlated with the level of the functional covariate, but will be related with the slope of the exogenous curves. As such, residuals $\varepsilon_t^{(2)}$ associated with exogenous curves X_t^1

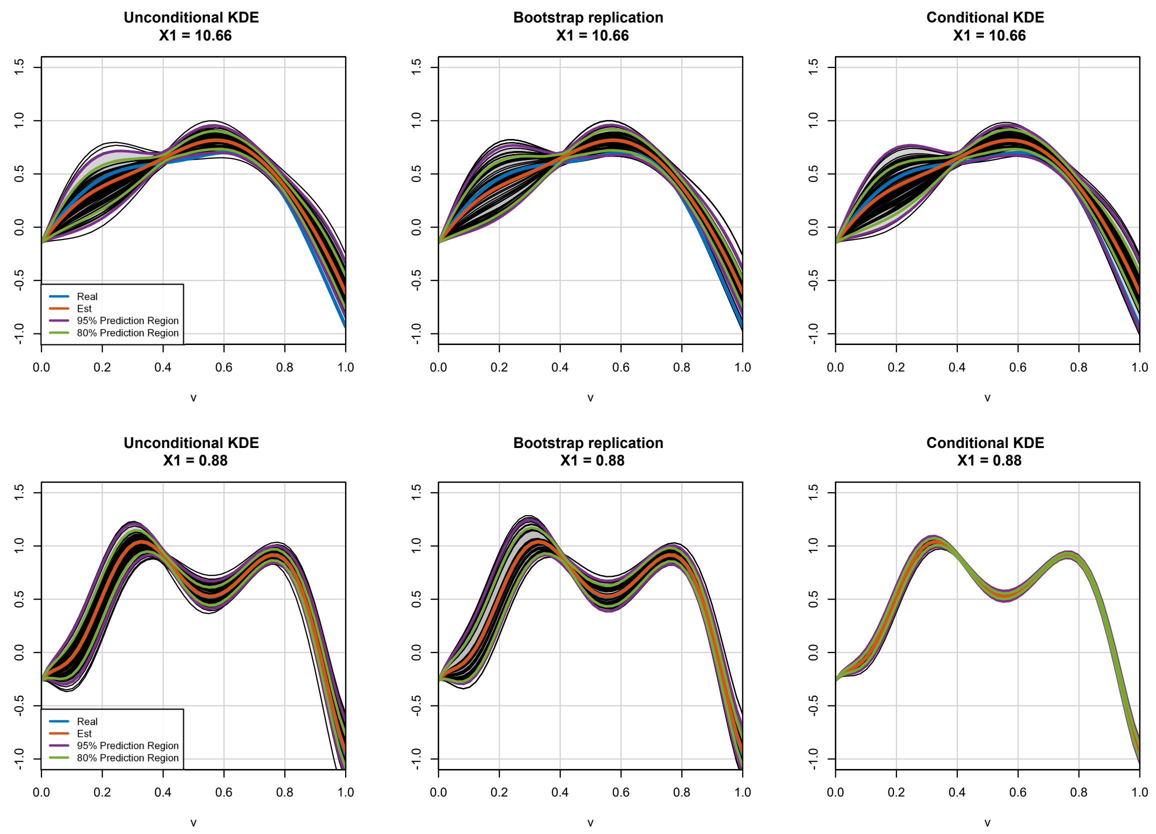

Figure 4.11. Prediction regions defined by the residual-based bootstrap, unconditional and conditional models fitted to the synthetic series for the first simulation. The black curves represent the different scenarios obtained by each method. The top row corresponds to a high-volatility period, whereas the bottom figures display the curves for a low-volatility period.

with a positive slope will exhibit a greater dispersion than the remaining residuals. Figure 4.12 illustrates the setup for this simulation, colouring each family of residuals and its associated functional covariates accordingly.

This simulation is devoted to showing how the proposed model is flexible enough to use functional data as inputs for estimating the conditional density of a set of functional residuals by selecting a suitable distance function that is able to detect the relevant information given by the functional covariates. As the dispersion of the residuals is related with the slope of variable X_t^1, the derivative-based semi-metric introduced in Equation 2.11 will be used in the conditional model for this simulation:

$$\hat{p}_{\mathscr{H}}^{(2)}\left(\boldsymbol{e}^{(\mathbf{2})}|x\right) = \sum_{t=1}^{300} \omega_t\left(X\right)\mathbf{K}_{\boldsymbol{H}}\left(\boldsymbol{e}^{(2)} - \boldsymbol{e}_t^{(2)}\right), \tag{4.32}$$

where the weight function of the functional covariate is given by

$$\omega_t\left(X\right) = \frac{K_{h_1}\left(\mathrm{d}_1^S\left(X^1, X_t^1\right)\right)}{\sum_{j=1}^{300} K_{h_1}\left(\mathrm{d}_1^S\left(X^1, X_j^1\right)\right)}. \tag{4.33}$$

Similarly to the previous model, $\boldsymbol{e}_t^{(2)}$ denotes the scores of the FPC decomposition of the residuals $\varepsilon_t^{(2)}$. As the slope of the functional covariate X_t^1 influences the dispersion of the

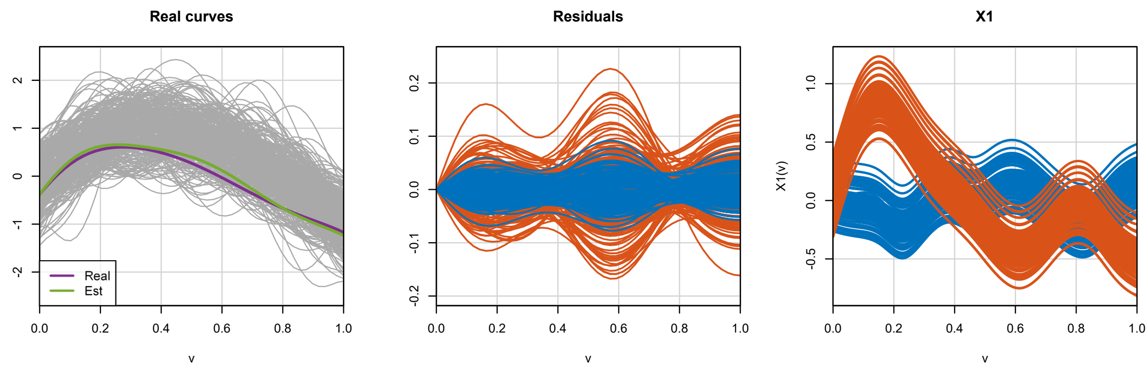

Figure 4.12. Left: synthetic functional dataset used in the second simulation. Grey curves represent the real observations, whereas the coloured curves show examples of the real and estimated curves. Middle: set of simulated functional residuals. The curves have been coloured following the different volatility levels given by the functional covariate X_t^1 included in the simulation. Right: exogenous functional variable. Curves with a different slope are displayed using different colors.

residuals $\varepsilon_t^{(2)}$, the proposed model uses the semi-metric $\mathrm{d}_1^S(.,.)$, which is the L^2 norm of the first derivative of X_t^1. The bandwidth parameters of the model are selected by employing the cross-validation method used in the previous simulation.

Once the model has been identified, it can be applied to obtaining estimations of the conditional density of the projections of the synthetic residuals, which then can be employed to define both prediction regions and to generate functional scenarios that follow the estimated density. In order to obtain a good coverage of the curves, the model must have captured the relation between the slope of the functional covariates and the dispersion of the residuals. The images shown in Figure 4.13 provide some insight about the improvements made by the proposed model over the two reference unconditional models included in this study: if the slope of the curves X_t^1 is high the proposed model obtains a similar estimation than the one given by the reference models, as can be seen in the top images of Figure 4.13. However, when the L^2 norm of the first derivative of variable X_t^1 is low the model detects that this is a situation where the variance of the residuals is low, so the estimated prediction region is narrower than the ones obtained by the reference models. This illustrates the capability of the proposed model to include functional covariates into its formulation, and also highlights the necessity of selecting a suitable semi-metric function that it is able to highlight the relevant features of the exogenous curves.

In order to provide a global comparison of the different models included in this section, a comparative table including the average probabilistic scores obtained for each model and dataset is included. Table 4.2 contains the values of the functional CRPS, Winkler Score and Log-Score detailed in Section 4.3 for the two reference models included in this study as well as the scores obtained by the proposed conditional probabilistic model. The mean scores validate the improvement made by the conditional model over the reference unconditional models, as both the coverage of its prediction bands and the likelihood of the functional scenarios generated by the model is superior to the scores obtained by the remaining models. The applicability of the proposed model when including functional covariates has been validated in the second simulation, where the model defined in Equation 4.32 is able to capture the variations in the dispersion of the residuals given by the slopes of the functional covariates, obtaining a better probabilistic score when compared with the reference models.

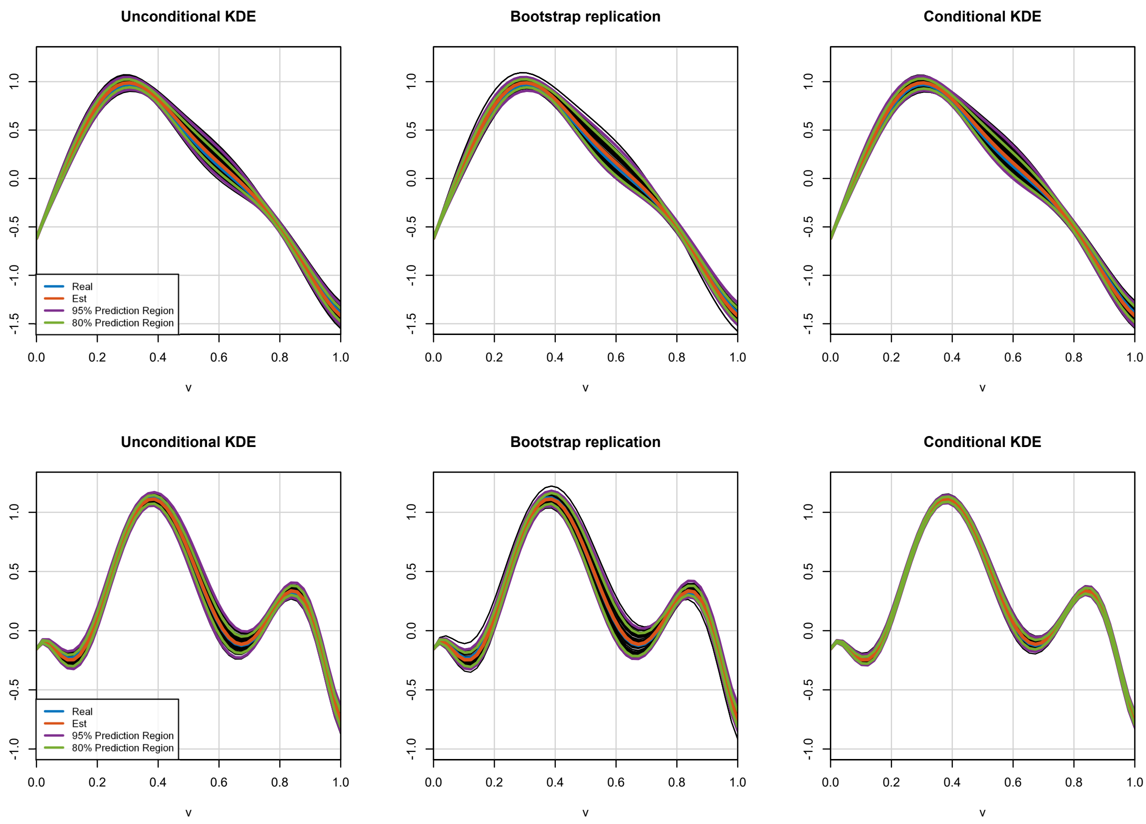

Figure 4.13. Prediction regions defined by the residual-based bootstrap, unconditional and conditional models fitted to the synthetic series for the second simulation. The black curves represent the different scenarios obtained by each method. The top row corresponds to a high-volatility period, whereas the bottom figures display the curves for a low-volatility period.

	Simulation 1			Simulation 2		
Model	fCRPS	fWS_{95}	fLS	fCRPS	fWS_{95}	fLS
Bootstrap	0.022	0.3196	-14.02	0.01187	0.1879	-14.38
Unconditional	0.019	0.3101	-15.28	0.01036	0.1699	-16.41
Conditional	**0.017**	**0.2681**	**-18.34**	**0.0099**	**0.1434**	**-18.04**

Table 4.2. Average scores for each probabilistic model in each simulation study. The lower scores obtained for each metric are marked in bold.

This concludes the simulation study performed in this section, which was devoted to illustrating how the proposed model can be used to model conditional distributions of the residuals of a functional deterministic model. The estimations of the model are compared with two reference functional probabilistic models often used in the literature, showing the improvements made by the proposed model in situations where the distribution of the residuals exhibits a certain relation with a set of exogenous variables, either scalar or functional. This is a common situation in real-world forecasting applications, as the volatility of the residuals of a fitted deterministic model is often related with sudden changes in the market. As such, next section will illustrate the applicability of the conditional probabilistic model in a real-world application, in order to test the usefulness of the model when dealing with real data.

4.5. Application to real-world datasets

This section is devoted to illustrating the real-world applicability of the proposed functional probabilistic model in a case study that analyses the Complex Residual Demand Curves (CRDC) obtained from the Spanish Electricity Market. The definition of CRDC was introduced in Portela, Muñoz, et al. (2017) as a generalization of standard Residual Demand Curves(RDC) that include complex conditions of energy markets into its formulation, such as market splitting occurrence. Due to its ability to explain the strategic bidding behaviour of the market agents, there has been a growing interest in forecasting these residual demand curves. This section will follow the analysis performed in Portela (2017), where the authors proposed a SARMAHX$(2,0) \times (1,1)_{24} \times (1,1)_{168}$ model to obtain 1-step and 24-step ahead forecasts of the series of CRDC for the year 2013.

The bidding behaviour of the agents is conditioned by weather conditions that affect the demand, wind and solar generation; and several external factors that might influence the decisions of the traders. As such, the SARMAHX model proposed by the authors uses the demand of the system and the wind and hydraulic production as scalar covariates to the fitted model. The resulting model is compared against other well-known forecasting models commonly used in the RDC forecasting literature, proving that the SARMAHX model is a competitive model for forecasting RDC as it outperforms the reference deterministic models. Henceforth the analysis presented below will focus on the results obtained by the authors using the SARMAHX model.

Having accurate estimations of the future residual demand curves is of utmost importance to any market agent, as this information can be used to optimize its offering strategy. However, as pointed out in Berzal et al. (2001), deterministic forecasts of the curves may not be enough as the market uncertainty must be taken into account when obtaining an optimal bidding strategy for a market agent (Baillo et al., 2004; Campos et al., 2016). As such this case study is devoted to analyse the capability of the proposed functional probabilistic model to quantify the uncertainty associated with the forecasts of the SARMAHX model and to measure the quality of the functional scenarios obtained. The proposed model will be compared against two well-known scenario generation models found in the functional data literature: the unconditional kernel density estimator based on the techniques proposed in Delaigle and Hall (2010) and the residual-based bootstrap method used in Vilar, Aneiros, et al. (2018) and Rossini and Canale (2019), among others. Once the models are evaluated, the quality of the obtained probabilistic forecasts will be measured using the 3 scores described in Section 4.3, in order to find which model is able to obtain a more coherent set of scenarios for the series of CRDC.

The setup for this study is as follows: the series of Complex Residual Demand Curves between days $01/02/2013$ and $10/05/2013$ is obtained, as well as the set of scalar covariates used in Portela (2017) to fit a SARMAHX model to the series. The first 2000 hours of the dataset are used to fit the parameters of the SARMAHX model, leaving the remaining curves as an Out-of-sample set to measure the generalization capabilities of the model. Once fitted, the SARMAHX model is evaluated to obtain 1-step ahead forecasts of the CRDC, following the case study presented by the authors. Figure 4.14 shows both the data used to fit the SARMAHX model and the 1-step ahead estimations obtained by the model, together with the residuals of the model. As mentioned in previous sections, it is necessary to characterize the distribution of these residuals to obtain probabilistic forecasts of the series. Due to the infinite-dimensionality of the functional time series being modelled, the projection-based density of the curves must be used as an approximation for the density function of the data. Hence, in order to fit the

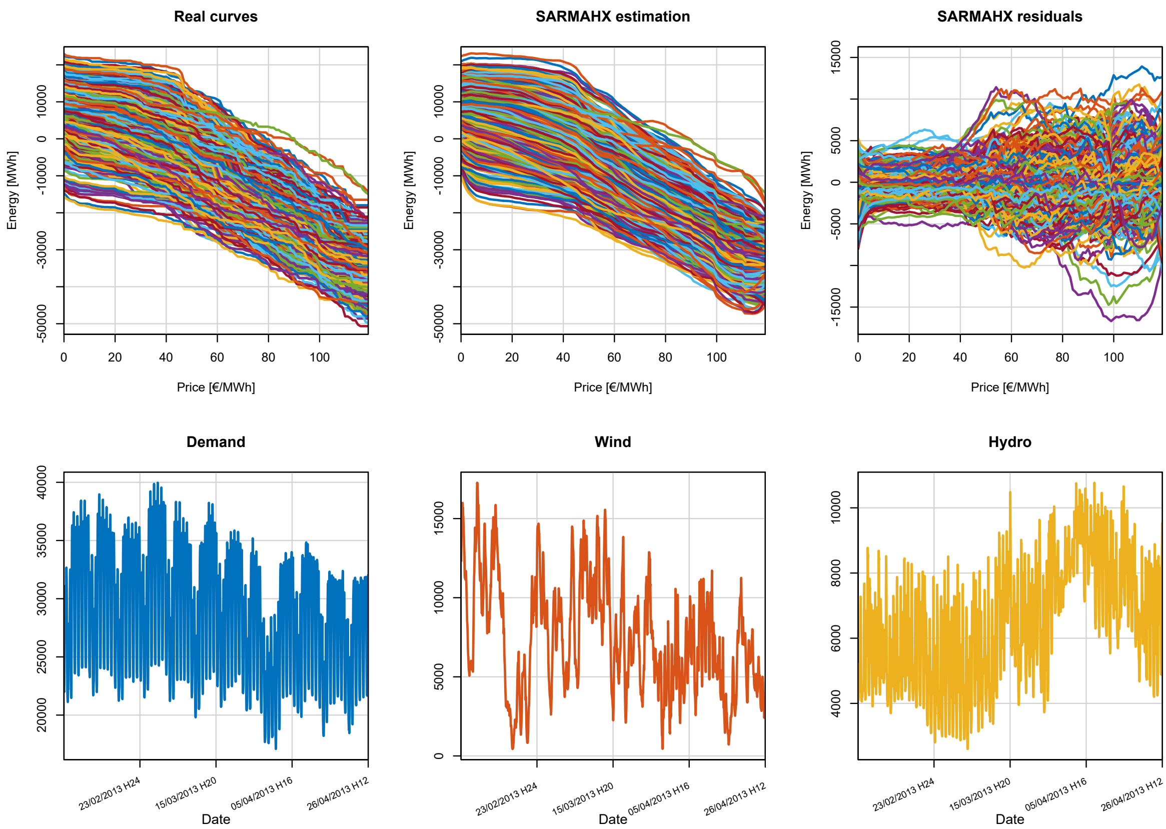

Figure 4.14. Top: actual hourly Complex Residual Demand curves together with the 1-step ahead forecasts obtained by the fitted SARMAHX model and the residuals of the model. Bottom: Hourly values of the three exogenous variables included in the regression part of the SARMAHX model.

proposed probabilistic model the 1-step ahead residuals of the SARMAHX model must be decomposed into their first FPC. The number of components used in the decomposition is selected using the CPV criterion introduced in Section 2.2.5, which shows that using 4 FPC in the decomposition allows us to explain more than 90% of the total variance of the residual curves of the model.

The main benefit of the conditional probabilistic model proposed in this dissertation is its ability to model the variations in the distribution of the curves based on a set of exogenous variables. Hence, an exploratory analysis is carried out to identify possible relations between the covariates and the distribution of the scores of the FPC decomposition of the residuals. Figure 4.15 illustrates the distribution of each FPC's scores given a certain level of the exogenous covariates used to fit the deterministic SARMAHX model. As this is a hourly forecasting problem, the distribution of the residuals grouped by hours is also included in order to check if there are any set of hours where the uncertainty of the model is lower. This analysis shows that the distribution of the scores exhibits a different behaviour for different hours of the day: the dispersion of the scores for hours 1-2 and 19-24 is very high when compared with the low dispersion that can be found between hours 3-6. In addition, the distribution of the scores for some groups of hours have a non-zero mean, as evidenced in the distribution of the scores of the first principal component for hours 7-8 and 11-12. This fact is quite surprising: although the regression residuals of the SARMAHX model exhibit a zero-mean level, differences can be seen when grouping the data by hour, showing that the functional regression model may benefit from the inclusion of some intervention variable that is able to model the aforementioned

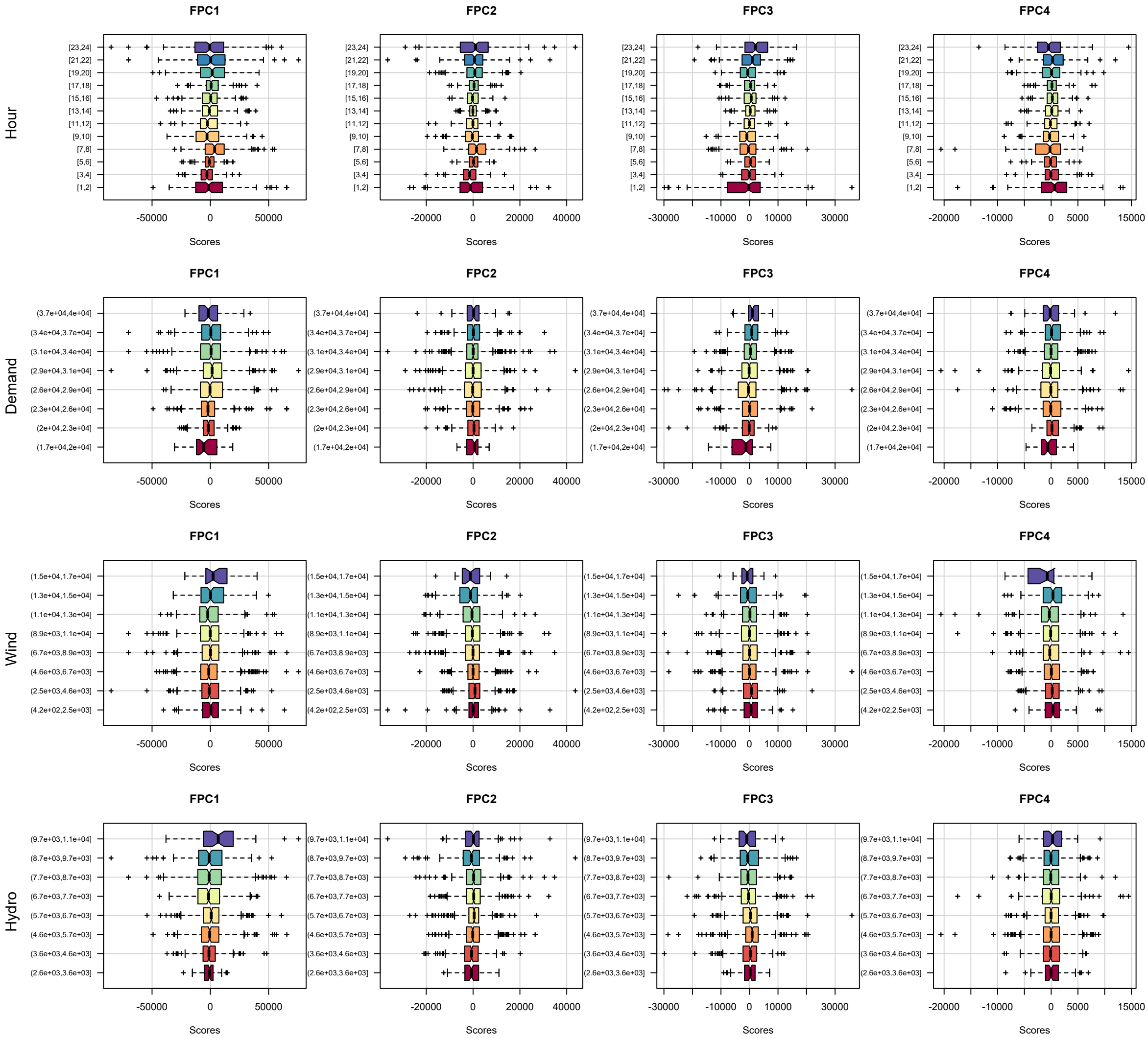

Figure 4.15. Distribution of the FPC scores of the SARMAHX 1-step ahead residuals grouped by levels of the exogenous variables. Each column is associated with the scores of a certain FPC, whereas each row is associated with a different covariate.

differences. The remaining covariates do not exhibit such a clear relationship between the levels of the variables and the dispersion of the scores, where only the variable Hydro can be seen as a variable that influences the variance of the scores. In order to whether the levels of the covariates influence the distribution of the residuals' scores, the two-sample Kolmogorov-Smirnov (KS) test (Hollander and Wolfe, 1973) is used. This is a nonparametric test that assesses whether two samples have been drawn from the same distribution. If the p-value of the statistic is lower than 0.05, the null hypothesis of both samples being drawn from the same distribution can be rejected at a 95% confidence level. Table 4.3 contains the minimum p-value obtained for the grouped distributions of the exogenous variables shown in Figure 4.15 for each FPC score. In order to provide a robust estimations of the differences in the distribution, only grouping of the variables that contained a representative sample of the data have been included in this study. Hence, variables Hour (hour of the day) and Hydro (hydroelectric power generation) will be used as inputs for the probabilistic model, as these are the variables selected by the KS test, showing that they are able to explain most of the variation in the distribution of the residuals' scores. Figure 4.16 illustrates the differences in the distribution of the residuals: the variance exhibited by the model's residuals is low for hour 6 and low levels of hydraulic

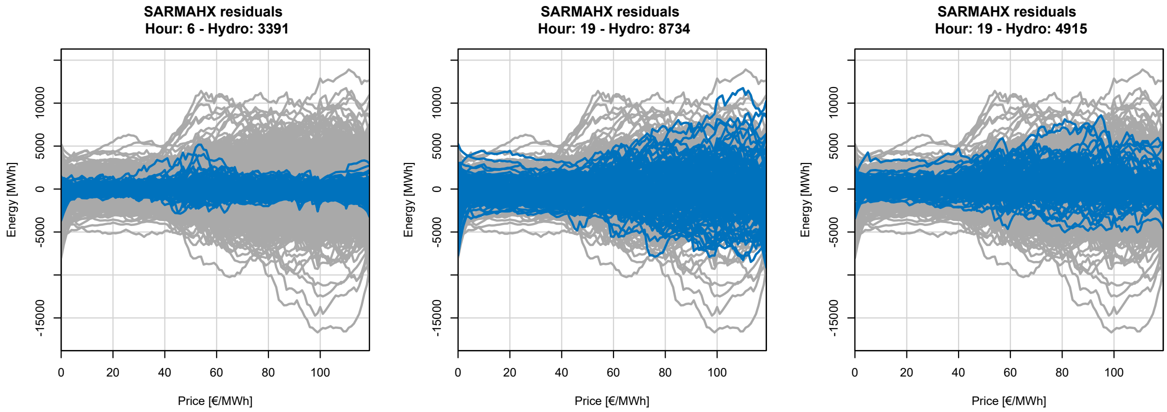

Figure 4.16. Distribution of the residuals of the model for three different values of the exogenous variables Hour and Hydro for the 1-step ahead forecasts. The whole set of residuals is displayed using grey color, whereas the curves with a positive weight in the probabilistic model are shown in blue.

	FPC1	FPC2	FPC3	FPC4
Hour	**0.00013**	**0.0061**	**0.00081**	**0.0028**
Demand	**0.0249**	0.146	0.0671	0.21047
Wind	0.06207	**0.01135**	0.0514	0.135
Hydro	**0.00031**	**0.00073**	**0.00011**	0.08059

Table 4.3. p-values of the Kolmogorov-Smirnov test obtained for the distributions of the one-step ahead residuals' scores grouped by levels of the exogenous variables. p-values lower than 0.05 are marked in bold.

production, whereas high values of hydraulic production and hours in the range 19-24 usually entails greater dispersion in the residuals. As illustrated by the third panel, using only variable Hour is not enough to model the conditional distribution of the data, as the residuals selected by the model for hour 19 but a lower level of Hydro than the second panel exhibit a lower dispersion.

The probabilistic model for the 1-step ahead forecasts is formulated as follows

$$\hat{p}^{(4)}_{\mathscr{H}}\left(\boldsymbol{e}^{(1)}|\boldsymbol{x}\right) = \sum_{t=1}^{T} \omega_t\left(\boldsymbol{x}\right) \mathbf{K}_{\boldsymbol{H}}\left(\boldsymbol{e}^{(1)} - \boldsymbol{e}^{(1)}_t\right), \tag{4.34}$$

with the weight function of the covariates given by

$$\omega_t\left(\boldsymbol{x}\right) = \frac{K_{h_1}\left(\|x^1 - x^1_t\|_{S^1}\right) K_{h_2}\left(\|x^2 - x^2_t\|\right)}{\sum_{j=1}^{T} K_{h_1}\left(\|x^1 - x^1_j\|_{S^1}\right) K_{h_2}\left(\|x^2 - x^2_j\|\right)}. \tag{4.35}$$

In this expression, $\boldsymbol{e}^{(1)}$ denotes the scores of the FPC decomposition of the 1-step ahead residuals, $\boldsymbol{x} = (x^1, x^2)$ stands for the values of variables Hour and Hydro respectively. As the Hydro variable included in the model is scalar, the euclidean distance $\|.\|$ in $\mathbb{R}$ is used by the model to weight the contribution of these values. Nevertheless, the same distance function cannot be used with variable Hour, as the distance between hour 24 and 1 should be equal to 1. In order to capture this relationship, the following distance function is used:

$$\|x - y\|_{S^1} = \min\{|x - y|, 24 - |x - y|\}. \tag{4.36}$$

This function is able to model the notion of proximity between hours, which is relevant for the proposed model. The leave-one-out cross-validation procedure proposed in Section 4.2.1 is used to select the optimal bandwidth parameters for this model. Once fitted, 80% and 95% prediction regions are obtained from the conditional density estimated by the model in both the In-Sample and Out-Of-Sample periods. Figure 4.17 illustrates the obtained predictive regions together with a set of 100 scenarios of Complex Residual Demand Curves. The estimations of the unconditional kernel density model and the residual-based bootstrap model are also included, in order to highlight the performance of the proposed conditional model. As evidenced by this comparison, the proposed model is able to capture the differences in the uncertainty of the forecasts, which results in narrower predictive regions for low-volatility periods of time and coherent scenarios of the functional time series being modelled. In addition, the upper and lower limits of the predictive region defined by the conditional density model better captures the shape of the real curves, as these regions are derived from a proper density function unlike the prediction regions obtained by the bootstrap method. This enables an easier quantification of the uncertainty of each forecasted curve, allowing for a better interpretation of the results. As shown in Figure 4.17 both the 95% and 80% prediction regions obtained by the proposed model outperform the regions estimated by the other unconditional models: in the bottom row, the real curve is inside the 80% region obtained by the conditional density estimation method, whereas it lies outside the boundaries defined by the other two models. This illustrates the sharpness of the probabilistic statements produced by the proposed model, proving itself to be a competitive model when compared with the other probabilistic methods considered in this study.

Up to this point the analysis has been focused on the 1-step ahead forecasts obtained by the SARMAHX model, as the training process for this model aims at minimizing the 1-step ahead forecasting error. Hence, analyzing this case can be useful to obtain insight about the formulation of the model as well as providing an example for illustrating the probabilistic capabilities of the proposed model. However, in real operation the auctions for the next 24 hours are cleared at the same time, so market agents will need to have forecasts of these 24 CRDC before the auctions take place. Hence, the SARMAHX model will be evaluated to obtain 24-step ahead forecasts of the series, better illustrating the applicability of the model to a real world problem. Once the 24-step ahead estimations have been obtained the proposed model will be used to obtain probabilistic forecasts for each following day. This has a twofold purpose: on the one hand, it will show how the proposed model can be adapted to a variable horizon forecasting problem. On the other hand, it will validate the probabilistic forecasts obtained in a real case application.

Similarly to the 1-step ahead case, the residuals obtained after the evaluation of the SARMAHX model with a forecasting horizon of 24 hours must be decomposed into their first FPC in order to estimate the projection-based density of the curves. In this particular case the first 4 FPC are able to explain more than 95% of the variance of the original set of curves. Once the curves have been projected into the subspace spanned by these first 4 principal components, it is necessary to analyze whether the distribution of the scores exhibits any kind of dependency with the exogenous variables available for this forecasting problem. Figure 4.18 shows the differences in the distribution of the scores when aggregated by different levels of the covariates used to fit the SARMAHX model. Unlike the 1-step ahead case, the distribution of the 24-step ahead residuals do not exhibit a significant difference for different hours of the day. However, as illustrated by the figure, the distribution of the first FPC exhibits a strong relation with the demand: low demand levels are typically associated with negative scores,

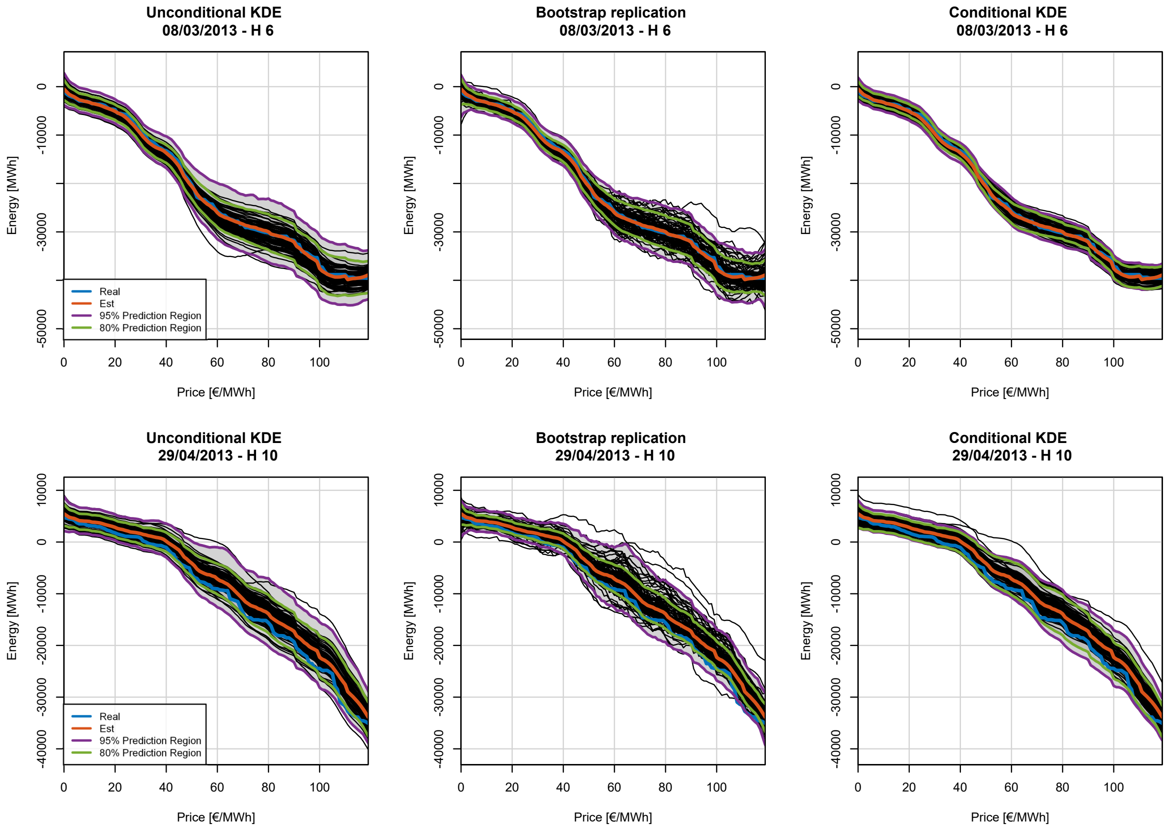

Figure 4.17. Prediction regions defined by the residual-based bootstrap, unconditional and conditional models fitted to the CRDC series for the 1-step ahead forecasting problem. The black curves represent the different scenarios obtained by each method. The top row corresponds to the In-Sample period, whereas the bottom figures display curves for the out-of-sample period.

whereas the variance of the scores decreases with the level of the demand. The hydraulic production also influences the distribution of the scores, as the dispersion of the scores increases with the levels of this variable. In addition, high levels of hydraulic production are associated with a distribution of the scores that is not centered at 0. The Kolmogorov-Smirnov (KS) test is consistent with this selection, as the variables selected exhibit a lower p-value across all 4 FPC scores, as shown in Table 4.4. Extreme values of wind production are hardly present in the original dataset, so the KS test does not find enough evidence to find differences in the distribution of the scores after removing those outlying values from the sample. This is further exemplified in Figure 4.19, where the 24-step ahead residuals of hours with similar values of demand and hydraulic production are highlighted, showing that the distribution of the SARMAHX residuals is influenced by these two variables. Hence, variables Demand and Hydro will be used as inputs for the 24-step ahead probabilistic model as they are able to explain most of the variation in the distribution of the scores.

Once the input variables are selected, the bandwidth parameters of the conditional probabilistic model are selected by the leave-one-out cross-validation procedure proposed in this chapter. Once fitted, the model is evaluated in both the In-sample and Out-Of-Sample periods to obtain both 80% and 95% predictive regions defined by the conditional density estimated by the model. In addition, a set of 100 functional scenarios are obtained for each hour which follow the estimated conditional distribution of the residuals of the SARMAHX model. Similarly to the 1-step ahead case the residual-based bootstrap and unconditional kernel density estimator models have also been fitted to this dataset to verify if the proposed probabilistic

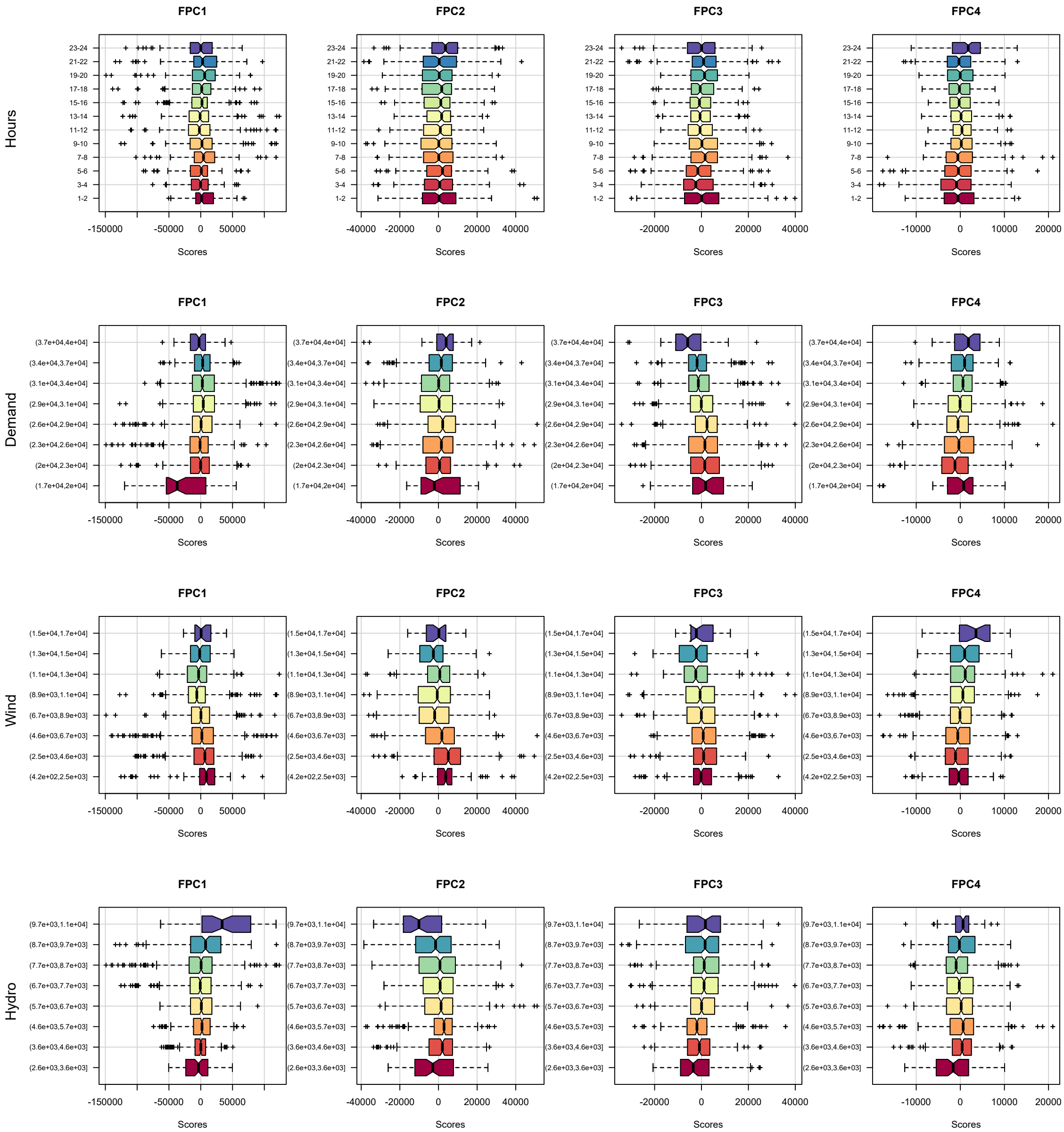

Figure 4.18. Distribution of the FPC scores of the SARMAHX 24-step ahead residuals grouped by levels of the exogenous variables. Each column is associated with the scores of a certain FPC, whereas each row is associated with a different covariate.

model is able to improve the probabilistic statements obtained by these two models. Figure 4.20 contains examples of both the predictive regions and functional scenarios obtained by each of the 3 methods included in this comparison. As it can be seen, both the unconditional and the wild bootstrap models are unable to capture the variations in the uncertainty associated with the 24-step ahead forecasts obtained by the SARMAHX model, as indicated by the wide prediction regions obtained in both the In-Sample and the Out-Of-Sample periods. This analysis also illustrates the improvements made by the proposed probabilistic model over these two widely used functional models, as the conditional density model is able to better capture the variations in the distribution of the SARMAHX model's residuals, obtaining narrower prediction regions in low volatility periods of the series and wider regions where the uncertainty of the model increases.

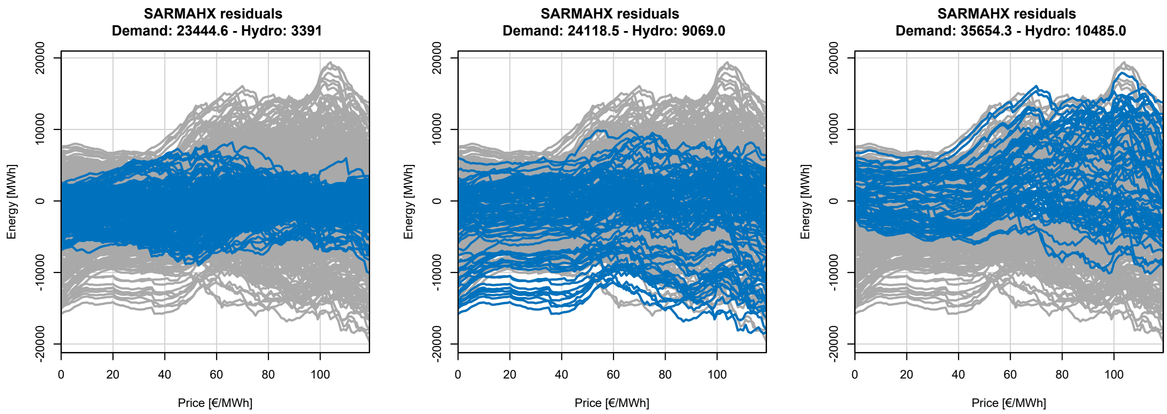

Figure 4.19. Distribution of the residuals of the model for three different values of the exogenous variables Demand and Hydro for the 24-step ahead forecasts. The whole set of residuals is displayed using grey color, whereas the curves with a positive weight in the probabilistic model are shown in blue.

	FPC1	FPC2	FPC3	FPC4
Hour	**0.00445**	0.06224	**0.0463**	0.0781
Demand	**0**	**0.00062**	**0.00022**	**0.000581**
Wind	**0.00028**	**0.00685**	0.063	0.0829
Hydro	**0**	**0**	**0.000014**	**0.000393**

Table 4.4. p-values of the Kolmogorov-Smirnov test obtained for the distributions of the 24-step ahead residuals' scores grouped by levels of the exogenous variables. p-values lower than 0.05 are marked in bold.

In order to test the quality of the three functional probabilistic models included in this section, the probabilistic scores introduced in Section 4.3 have been obtained for both the In-Sample and Out-Of-Sample periods and the 1-step and 24-step ahead forecasts of the series. Table 4.5 shows the average scores achieved by each model, which illustrate the advantage of the proposed conditional model over the reference models: the prediction region obtained by the proposed model achieves an average Winkler Score in the Out-Of-Sample period of 848817.7 for the 1-step ahead study, a clear improvement over the Winkler Score of 1013025.2 obtained by the residual-based bootstrap model for the same period of data. Similar values are obtained for the 24-step ahead forecasts, with the conditional model obtaining an average Winkler Score of 1541468, outperforming the best reference model that only achieves a Winkler Score of 2179392. As the Winkler Score quantifies the conditional coverage and mean width of the prediction regions, it can be concluded that the sharpness of the probabilistic bounds obtained by the proposed model is higher than the other reference models found in the literature. When analyzing the functional CRPS metric, the conditional model obtains a better score in the In-Sample period, but both the proposed model and the bootstrap models are outperformed by the unconditional model in the Out-Of-Sample period. However, as seen in the comparative analysis performed in Section 4.3, the functional CRPS measure may not be adequate for quantifying the quality of the functional scenarios, as it is unable to distinguish between the conditional coverage of the scenarios nor to capture the shape similarities of the curves. Nevertheless, as stated before, many real-world applications seek to obtain realistic scenarios of the actual series being modelled, as these scenarios are then used to optimize the strategy of the company. Hence, it is desirable to maximize the likelihood of the obtained scenarios to ensure the consistency of scenarios with observations. As such, the proposed functional Log-Score metric is used to quantify this property, and in both the In-Sample and Out-Of-Sample periods the scenarios

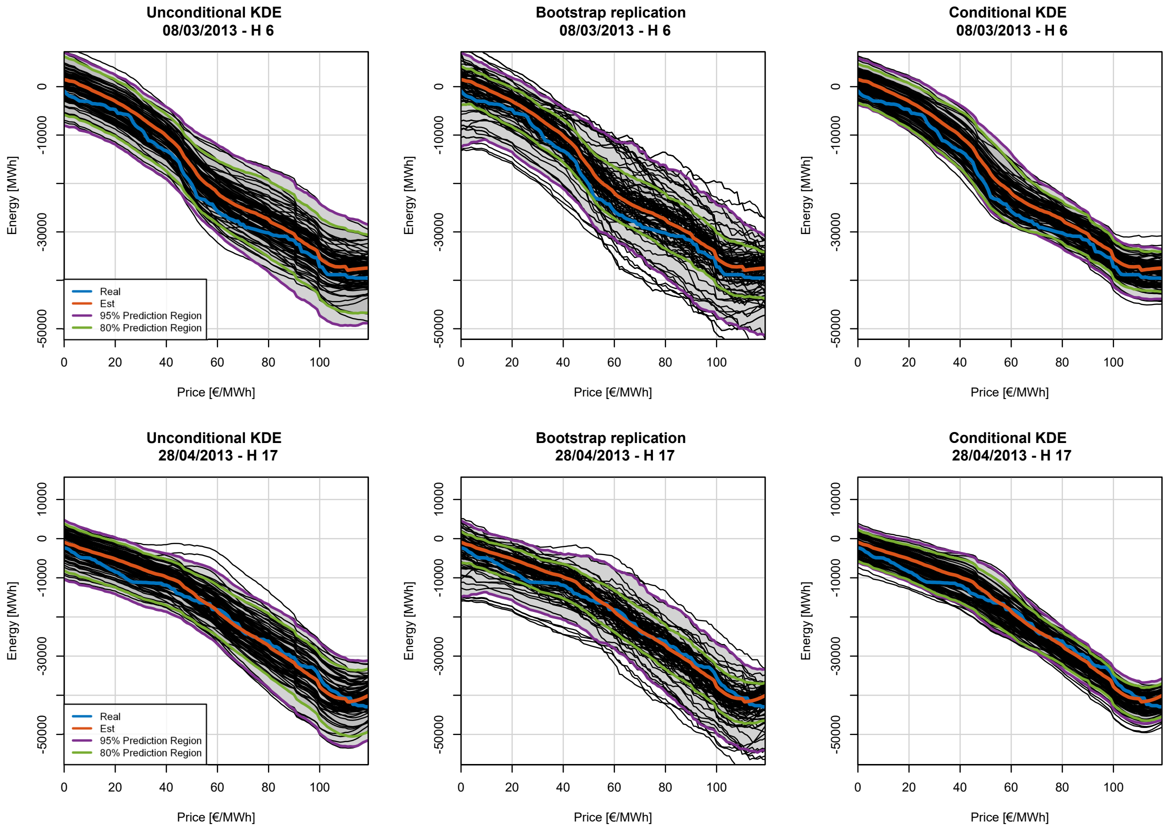

Figure 4.20. Prediction regions defined by the residual-based bootstrap, unconditional and conditional models fitted to the CRDC series for the 24-step ahead forecasting problem. The black curves represent the different scenarios obtained by each method. The top row corresponds to the In-Sample period, whereas the bottom figures display curves for the out-of-sample period.

obtained by the conditional model attain a better score than the scenarios generated by the unconditional reference models. This metric is consistent with the graphical representations included in Figures 4.17 and 4.20, as the scenarios obtained with the conditional model exhibit a similar shape as the real Complex Residual Demand Curves in addition to having a lower dispersion than the remaining scenarios obtained by the reference probabilistic models. The metrics obtained for the 1-step ahead study are lower than the scores obtained for the 24-step ahead forecasts, as the uncertainty associated with the 1-step ahead estimations is lower than the other forecasting study.

The analysis carried out in this section has been devoted to illustrating the applicability of the proposed functional probabilistic model in a real-world problem, using a dataset of Complex Residual Demand Curves obtained from the Spanish Electricity Market. In order to validate the usefulness of the conditional model developed in this section, 1-step ahead and 24-step ahead probabilistic forecasts of the series have been obtained, and then compared with the two most widely used probabilistic models in the functional data analysis framework. An exploratory analysis has been carried out to identify variations in the distribution of the SARMAHX residuals that are related to the levels of certain covariates. Hence, periods of high uncertainty in the bidding behaviour of the market agents may be characterized with extreme values in the variables considered in this study. This allows for modelling the conditional distribution of the residuals of the SARMAHX model, which results in more accurate probabilistic forecasts of the original series. In addition, the proposed model is able to generate new functional scenarios that have not been previously observed, unlike residual-based bootstrap methods. This is a desirable

Forecasting horizon	Model	In-Sample fCRPS	In-Sample fWS_{95}	In-Sample fLS	Out-Of-Sample fCRPS	Out-Of-Sample fWS_{95}	Out-Of-Sample fLS
1-Step Ahead	Bootstrap	54119.54	965100.9	30.75	42773.03	1013025.2	30.24
	Unconditional	54017.84	1115016.1	30.35	**41642.51**	1110242.6	30.10
	Conditional	**51614.39**	**887061.3**	**30.15**	41807.88	**848817.7**	**29.98**
24-Step Ahead	Bootstrap	117977.6	2378571	32.86	88172.79	2372640	32.47
	Unconditional	113540.1	2187799	32.55	**85461.91**	2179392	32.39
	Conditional	**104614.6**	**1648718**	**32.06**	86735.09	**1541468**	**32.17**

Table 4.5. Average scores for each probabilistic model in the 1-step and 24-step ahead Complex Residual Demand Curve scenario generation study. The lower scores obtained for each metric are marked in bold.

property of scenario generation methods as it allows for the generation of future scenarios that follow the conditional density estimated by the model, resulting in more realistic and coherent scenarios. A set of quality metrics for evaluating the probabilistic performance of the models have been obtained, showing that the proposed conditional density model is a competitive model against other well known existing methods for functional scenario generation for both 1-step and 24-step ahead forecasting.

4.6. Conclusions

This chapter has been devoted to introducing a probabilistic model for functional time series models that it is able to obtain sharp estimations of the conditional distribution of the projection-based density of the residuals of a given deterministic model using information given by a set of exogenous variables that influence the distribution of the residuals. This is achieved by combining classical kernel density estimation techniques with nonparametric techniques that allows the model to weight the contribution of each set of covariates. A leave-one-out cross-validation method has been proposed to select the optimal bandwidth parameters for the model, which seeks to maximize the likelihood of the conditional densities. Once fitted, the proposed model can be used to define predictive regions at a given confidence level α, which are the natural generalization of prediction intervals to the functional data framework. The upper and lower boundaries of these regions provide a useful tool to quantify the uncertainty associated with the forecasts of a given functional time series model. In addition, functional scenarios that follow the estimated conditional density can be obtained, which are able to represent the range of possible future values of the functional time series being modelled.

The applicability of the probabilistic model has been validated with a simulation study using synthetic datasets and a case study using Complex Residual Demand Curves from the Spanish Electricity Market. Deterministic 1-step and 24-step ahead forecasts of the series have been obtained using the SARMAHX model, and the conditional density of the model's residuals has been characterized. A set of exogenous variables that are able to explain the variations in the dispersion of the data has been identified and then used as input for the proposed model, obtaining sharp estimations of the volatility of the series. In order to measure the probabilistic performance of the proposed model two reference probabilistic models for functional time series have also been fitted to the data as a way to compare the efficiency of the proposed model. A set of quality measures for functional probabilistic models has been obtained for all three models, illustrating the improvements made by the proposed model over the reference models, proving itself to be a competitive model in real-world applications. The logarithmic score proposed in this section has been used to validate the consistency of the functional scenarios obtained by each model. Functional scenarios that are drawn from the estimated conditional density distribution are able to outperform the other scenarios obtained by the reference models, validating the performance of the proposed functional probabilistic model.

5

Functional probabilistic forecasting applications in electricity markets

This chapter is devoted to the empirical application of the proposed probabilistic forecasting methodology to the electricity market field. The forecasting methodology developed in this dissertation is evaluated on two different case studies. In each case study, the identification methodology based on the FACF/FPACF is used to select the optimal configuration of a SARMAHX model for the modelled series. Once a suitable deterministic model has been fitted to the data, probabilistic forecasts of the series will be obtained using the proposed conditional probabilistic model and then compared with other reference models often found in the probabilistic FDA literature. The first study is aimed at obtaining short-term probabilistic forecasts of the hourly Supply Curves for the Italian Day-Ahead electricity market. The second case study is devoted to scenario generation of Residual Demand Curves for the Spanish Day-Ahead electricity market, which illustrates how the proposed methodology is able to obtain accurate short-term estimations of the bidding behaviour of market agents.

5.1. Introduction

One of the main appeals of Functional Data Analysis is its wide range of applications across different scientific disciplines. The applicability of these techniques to biomedical sciences was illustrated in the extensive review performed by Ullah and Finch (2013), where the authors analyzed the relevance of FDA techniques in data registration and modelization employed by medical researchers. Recently, Matabuena et al. (2019) proposed a functional model to predict the maximum heart rate of an individual which reduced the predictive error more than 50 % when compared with current models found in the medical literature. In addition, functional data techniques have been applied successfully to financial markets: due to the high-frequency and granularity that financial time series often exhibit it is customary to analyze these data as

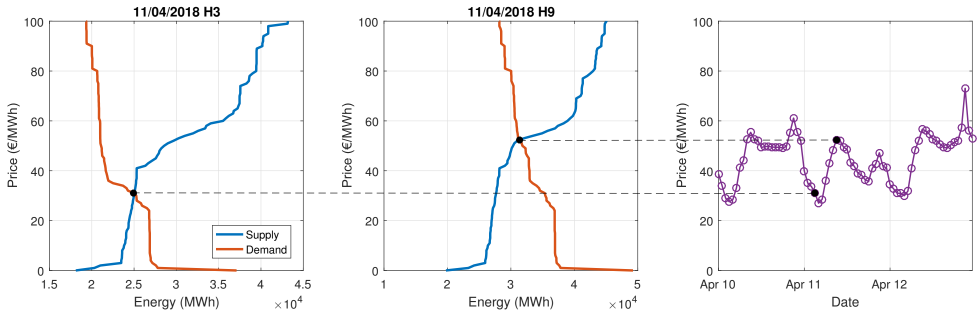

Figure 5.1. Supply and Demand Curves for the Spanish Day-Ahead electricity market. The intersection of both curves determines the market clearing price for each hour, as illustrated by the series of hourly prices shown in the right panel.

continuous curves (Shang et al., 2019). As such, functional forecasting models have been used to obtain accurate forecasts of forward curves derived from commodities futures (Horváth, Liu, et al., 2020), USA yield curves (Sen and Klüppelberg, 2019) and even estimating the Value-at-Risk associated with intra-day return curves (Rice, Wirjanto, et al., 2020). Nevertheless, recently there has been a growing interest in the application of FDA techniques to electricity markets, with authors developing new models to obtain accurate estimations of the daily load curves (Antoniadis, Brossat, et al., 2016; Vilar, Cao, et al., 2012) and electricity price curves (Raña, Vilar, et al., 2018; Portela, Muñoz San Roque, et al., 2018), among others. The FDA techniques proposed by these authors are able to outperform other well-known reference models often found in the literature, which illustrate the real-world applicability of the FDA framework to industry.

Although functional data arises naturally in many problems related with electricity markets, perhaps one of the most relevant applications of FDA in this field is the forecasting of supply and offer curves (Shah and Lisi, 2020). Many different countries have undergone a deregulation process in the last decades that has given rise to liberalized electricity markets that allow companies to trade energy in organized auctions (Joskow, 2008). As such, many day-ahead electricity markets are based on sealed-bid auctions where agents submit their selling and buying offers to the Market Operator, who then determines the market-clearing prices and the set of accepted bids for each time period. From here on we will only consider markets with an energy auction for each hour h of the day. This deregulation process has introduced new challenges to market agents regarding electricity price forecasting, highlighting the necessity of new modelling techniques that are able to model the underlying price-generating process.

Consider, for the sake of simplicity and without loss of generality, a simple-bid market. In this market each offer (or bid) is defined by a price p and a quantity q, which refers to the amount of energy that 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) curve for that agent is built. Once all the agents have submitted their offers, the sum of all the supply curves results in the system supply curve $S_h(p)$, and the sum of the demand curves of each firm results in the system demand curve $D_h(p)$. Market-clearing price p_h^* is computed for each hour as the intersection of the system aggregated supply and demand curves, hence $D_h(p_h^*) - S_h(p_h^*) = 0$. According to the auction type, two possible outcomes generally take effect: in discriminatory auctions, prices paid to winning suppliers are based on their actual bids, whereas in uniform-price auctions all suppliers are paid the market-clearing price. This process is illustrated on Figure 5.1, where the supply and demand curves of the Spanish Day-Ahead market for different hours are shown.

The clearing price of the market is determined by the intersection of these two curves, but having accurate estimations of these curves is not only useful to determine future market prices: as these curves contain the bidding strategy of all market agents, a company can use this information to optimize its own market strategy. In addition, this information can be used to quantify the sensibility of the electricity price when faced with changes in renewable production, as this usually entails a shift in the bidding curves.

For a given generation company i participating in a 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 (Bunn and Farmer, 1985). Estimates of the price allow to plan ahead and cover their operation costs and hedge against price movements (Weron, 2014). As a consequence, load (Chen et al., 2019), price (Zhang, Zhang, et al., 2019; Monteiro et al., 2018) or wind forecasting (Wang, Hu, et al., 2019) are widely studied and continuously being improved.

Furthermore, the market agent is usually interested in optimizing its bidding strategy in the market (Liu et al., 2012). This can be done using Residual Demand Curves (RDCs) (Baillo et al., 2004; Campos et al., 2016; Xu and Baldick, 2007; Prete and Hobbs, 2015). RDCs can be defined for each hour as the function that models the offering and bidding behaviour of all the competitors and can be calculated as

$$q_i = R_h(p) = D_h(p) - S_h^{-i}(p), \tag{5.1}$$

where $S_h^{-i}(p)$ is the supply function of the competitors of firm i, which is given by the system supply function $S_h(p)$ minus the firm's supply function $S_h^i(p)$ 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_i that the generation company i can sell in the market at hour h.

Consequently, it would be desirable for electricity companies to obtain short-term forecasts of supply curves and RDCs, as it would provide them with an accurate description of both the strategy of its competitors as well as a robust estimation of the price generation process of the market. Time series observed in electricity markets often exhibit complex dynamics, such as weekly and daily seasonal patterns as well as calendar effects or influence of exogenous variables such as the renewable production. In order to obtain a competitive model to forecast these complex data, all these features should be taken into account to produce robust forecasts of the series. The SARMAHX model detailed in Section 2.3.3 is able to include this information into its formulation, becoming a suitable model to obtain accurate estimations of both supply curves and RDCs for different electricity markets. The studies carried out in this chapter will illustrate the real-world applicability of the SARMAHX model to model functional time series that arise from electricity markets, validating its usefulness with a comparison against other reference models often found in the functional data literature. The identification and diagnosis methodology proposed in Chapter 3 will be used to select the optimal configuration of the SARMAHX model, providing the practitioner with much sharper forecasts of the series.

Although deterministic forecasts of supply curves and RDCs obtained by the SARMAHX model may seem adequate in many applications, market agents are often interested in having probabilistic forecasts of these curves, as it will provide them with a measure of the future volatility of the market. Having sharp estimations of the uncertainty associated with its forecasts

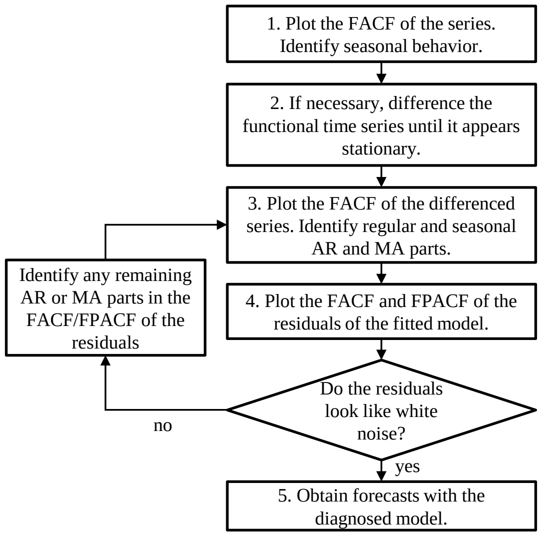

Figure 5.2. General procedure for identifying and diagnosing a SARMAHX model using the proposed FACF and FPACF.

allows them to hedge against future risk by optimizing its bidding strategy including future scenarios of the curves into their decision-making process. As mentioned earlier, it is safe to assume that exogenous conditions may influence the distribution of the residuals of the SARMAHX model. As such, the conditional probabilistic model proposed in Chapter 4 will be used to obtain probabilistic forecasts of the functional time series being analyzed. The model's capability to capture variations in the conditional distribution of the residuals determined by a set of exogenous covariates will be useful when estimating the density of the residual curves. A comparison between this model and other reference models often used in the literature will be carried out, in order to highlight how the proposed model is able to improve upon standard forecasting techniques for functional time series.

The forecasting methodology proposed in this thesis provides a very flexible framework to obtain estimations of functional time series that exhibit a complex dynamical behaviour. The FACF and FPACF introduced in Chapter 3 can be used to identify the optimal configuration of the autoregressive and moving average orders of the SARMAHX model for a given FTS, obtaining more accurate forecasts of the curves. This identification and diagnosis procedure is illustrated in Figure 5.2. Once the deterministic SARMAHX model has been fitted to the data it can be evaluated to obtain estimations of the functional time series. In order to obtain probabilistic forecasts of the series the uncertainty associated with the forecasts has to be modelled. This can be achieved by fitting the proposed conditional probabilistic model proposed in Chapter 4 to the series of SARMAHX residuals, obtaining a estimation of the projection-based density of the functional residuals conditioned to the values of a set of exogenous variables that influence the distribution of the residuals. This allows the model to adapt to different market regimes or changes in the bidding behaviour of the market agents, obtaining sharper probabilistic forecasts of the series being modelled.

In order to test the versatility of the techniques proposed in this dissertation, two real case studies involving data from electricity markets have been analyzed. Each case study will be briefly described, detailing the application, the nature and source of the data and the forecasting strategy that will be used.

- The first application is devoted to the short term forecasting of the hourly aggregated supply curves in the Italian Day-Ahead market. Having accurate forecasts of these supply curves is of utmost importance for an electricity company, as it provides a detailed estimation of the bidding strategy of its competitors, allowing them to optimize its bidding strategy in markets where the demand is considered inelastic. The application of functional forecasting models is appropriate for this setup, as the series of supply curves naturally define an hourly functional time series $S_h(p)$, with each curve defined over a fixed range of prices. In addition to obtaining accurate deterministic forecasts of the series, having coherent future scenarios of the supply curves of the next 24 hours is very useful for market agents, as it will provide them with information about the different strategies that they may face in the market. As such, the proposed probabilistic model will be used to estimate the distribution of the one-step and 24-step ahead residuals of the SARMAHX model, conditioned to the values of several scalar covariates that influence the volatility of the residuals. Then, prediction bands together with functional scenarios of the supply curves will be obtained. This case study is an updated version of the journal article Mestre, Portela, Muñoz San Roque, et al. (2020).

- The second application is aimed at forecasting Residual Demand Curves obtained for the Spanish Day-Ahead electricity market. As mentioned earlier, RDCs are used by utilities and other institutions to help the analysis of market agents' behaviour and in the decision making process in electricity markets. One of such applications is the estimation of the slope of RDCs of a given market participant, as it provides useful information regarding its capability to influence market prices. This allows market regulators to compare the potential market power of the different companies. In addition, short-term forecasting of RDCs has been used in both unit commitment (Garcia-Gonzalez and Barquin, 2000) and long term planning (Baslis and Bakirtzis, 2011). Similarly to the previous case, the series of RDCs define an hourly functional time series, so functional forecasting models will be used to estimate future values of RDCs. In order to improve the estimation of the curves, information about the system demand and the wind and hydraulic production will be used to account for the external factors that might influence the behaviour of the agents. Once the curves have been forecasted, functional scenarios of the RDCs will be obtained using the conditional density model proposed in this thesis. These forecasts open the door to many different applications, such as the evaluation of estimating the effect of renewable energy on the market clearing price, as an increase of the renewable production will shift the supply curve of the system, varying the electricity price for that hour. Similarly, the probabilistic forecasts of the residual demand curves can be used to obtain accurate estimations of the density of market clearing prices taking into account possible extreme events that may influence the distribution of the prices.

5.2. Forecasting hourly supply curves

This section is devoted to the short-term forecasting of aggregated hourly supply curves in electricity markets. As mentioned earlier, a supply curve $S_h(p)$ represents the total amount of energy offered by an agent for each price p in the auction of hour h. If the demand is inelastic to price, the residual demand of Equation 5.1 becomes

$$q = D_h - S^{-i}(p). \tag{5.2}$$

Therefore, the problem of forecasting the residual demand can be splitted into the forecasts of the hourly inelastic demand and the forecast of the supply curves. As demand forecasting has been widely studied in the literature (Aneiros, Vilar, and Raña, 2016; Huang and Shih, 2003; Fan and Hyndman, 2012), this section will focus on forecasting supply curves. The setup for this study is based on Pelagatti (2012), where the author analyzed the forecasting of supply curves of the competitors of the major Italian electricity company using functional models. Despite the fact that the Italian market consists in 6 interconnected zones, for simplicity and following Pelagatti (2012) the Italian market is considered as a single market, as if there were no saturation constraints between zones.

This first section is aimed at illustrating the applicability of the SARMAHX model to obtain short-term forecasts of the aggregated supply curves in the Italian Day-Ahead Market. The Day-Ahead Market (MGP) is the venue for the trading of electricity supply offers and demand bids for each hour of the next day. The MGP opens at 8:00 A.M. of the ninth day before the day of delivery and closes at 12:00 P.M. of the day before the day of delivery. Participants submit their bids where they specify the amount and the maximum and minimum price at which they are willing to buy and sell electricity for each hourly auction. Once the MGP closes, bids and offers are accepted based on the economic merit-order criterion and taking into account transmission capacity limits between zones. Where these constraints are binding, the market effectively splits and distinct zonal clearing prices are determined. While generators are paid the relevant zonal price, the accepted demand bids pertaining to consuming units are evaluated at the single national price (PUN) which is computed as the average of the zonal prices weighted by zonal consumption (Gianfreda and Grossi, 2012).

The results of the MGP are made known within 12.55 P.M. of the day before the day of delivery. As pointed out in (Shah and Lisi, 2019), due to confidentiality policies, the Italian Market Operator (GME) does not publish the detailed information relative to demand bids and supply offers for a confidentiality period of seven days, beginning on the day following the close of the market sitting to which they refer. However, for market participants, anonymized confidential results (including the acceptance/rejection of the bids and the aggregated supply curves without information about the geographical zone of each bid) are available at `https://www.IPEX.it`. As this study is concerned with forecasting the aggregated supply curves (instead of the zonal supply curves), all the data from hourly supply curves up to day D is available to forecast the hourly supply curves for day $D+1$ before the closing of MGP auction for day $D+1$.

Consequently, the setup for the real case study is as follows. Supply curves for the competitors of Enel, a major electricity company in Italy, are obtained by aggregating all the zonal competitors' supply curves. For this study, supply curves are limited to the price range $[0, 200]$ €/MWh. Therefore, each functional observation is the aggregated supply curve submitted to the market in each hour. The time range of the data is from 01/03/2015 up to 29/02/2016, thus consisting of 8784 curves. The data has been obtained from the Italian Electricity Market Operator (`www.mercatoelettrico.org`) and 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 remaining data is left for the Out-Of-Sample period.

The offering behaviour of the agents is conditioned to the weather conditions that affect the demand, wind and solar generation; as well as the particular circumstances of each day. Consequently, 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 study are the following:

- Total electricity 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 supply curves in windy days.
- Total solar production. The north of Italy is the region with the greatest 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.

It is worth noting that real values, instead of forecasts, of the explanatory variables are used for both training and validation of all the models considered in this study. In real operation the actual values of these variables are unknown, so future scenarios of the exogenous variables have to be used instead. The Italian Market Operator publishes hourly forecasts of the total electricity demand and international exchanges for day $D+1$ at day D, so the proposed models can use them to obtain estimations of the hourly supply curves before the auction for day $D+1$ closes. Other variables, such as the the thermal availability and solar and wind production should be forecasted in order to use them as inputs for the forecasting models. These models provide useful indicators for any market agent, hence in this study it is assumed that market participants already have forecasts of these variables. Historical information about the aforementioned variables can be obtained from the Transparency Platform of ENTSO-E, the European Network of Transmission System Operators (`https://transparency.entsoe.eu/`). 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.

The remainder of the section is divided into three parts. Firstly, the fitting of the SARMAHX model is detailed, showing the identification process proposed in this dissertation. Secondly, a comparison with other reference deterministic models is presented. Finally, probabilistic forecasts are obtained for the one-step and 24-step ahead forecasts using the proposed conditional probabilistic model for functional time series.

5.2.1. SARMAHX model identification

The identification and diagnosis technique proposed in this thesis is used to identify the functional SARMAHX model. Once the regular and seasonal autoregressive and moving average orders of the model have been selected, the model is trained with the In-Sample data of this case study. Both integral and concurrent operators are used. Firstly, the autocorrelation structure of the series of supply curves in the Italian electricity market is identified using the FACF and FPACF of the series together with the proposed identification procedure shown in Figure 5.2. Figure 5.3 show the seasonal behavior of the series: a high autocorrelation on lags 24 and 168 indicates that the series has strong daily and weekly seasonalities. The decay pattern of the FACF indicates the presence of an autoregressive part in the series. The serial autocorrelation is still present in Figure 5.4, which shows the autocorrelation structure of the residuals of a fitted functional regression model including the aforementioned exogenous variables. These

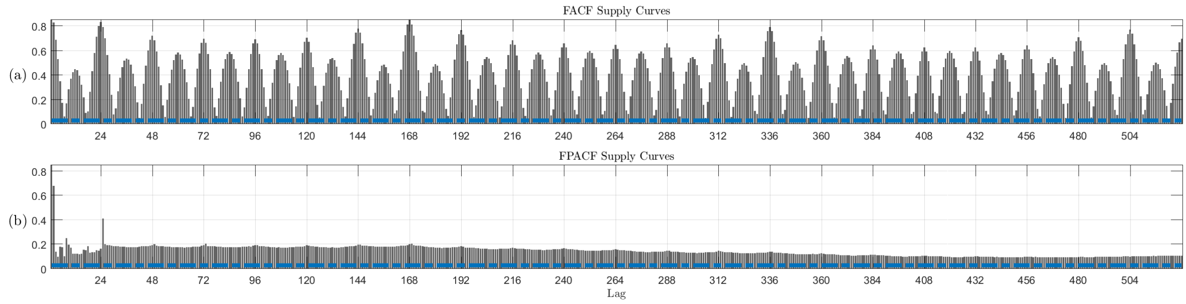

Figure 5.3. First 528 lags of the functional auto- and partial autocorrelation function (FACF / FPACF) of the functional time series of Italian supply curves. The horizontal line represents the 95% upper limit of the distribution of the statistic under the white noise hypothesis.

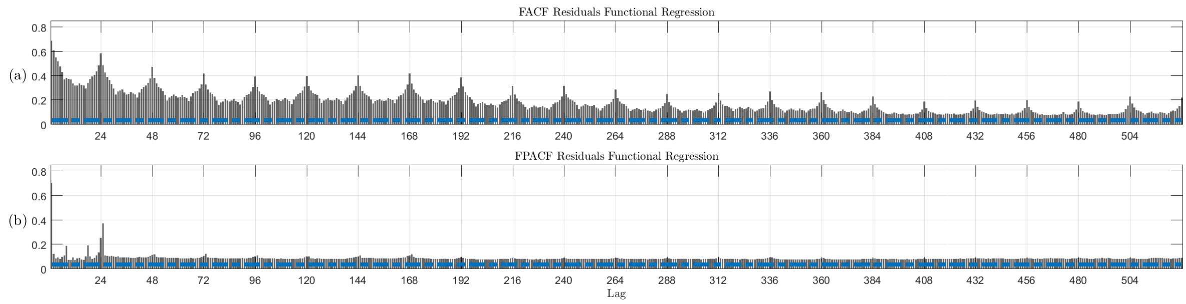

Figure 5.4. First 528 lags of the functional auto- and partial autocorrelation function (FACF / FPACF) of the regression residuals of a fitted functional regression model including the aforementioned exogenous variables. The horizontal line represents the 95% upper limit of the distribution of the statistic under the white noise hypothesis.

explanatory variables are included to capture the effect of these exogenous terms in the supply curves. As the output time series are hourly curves and the explanatory variables are hourly values, the model will consider them as scalar covariates. The strong serial correlation in the regression residuals of the SARMAHX model indicates that the dynamic of the series has to be modelled in order to obtain accurate forecasts of our data.

A SARMAHX$(1,0)\ \times(1,0)_{24}\ \times(1,0)_{168}$ model is fitted to model the autoregressive part observed in the residuals of the regression model. Although the slow decay pattern exhibited by the FACF of Figure 5.4 may be modelled by regular and seasonal differentiation of the series, it has been decided to include an autoregressive term in the model, as empirical tests have shown that the performance of the model increases with the inclusion of this autoregressive term instead of differencing the series. Figure 5.5 shows the functional autocorrelation of the residuals of this SARMAHX$(1,0)\ \times(1,0)_{24}\ \times(1,0)_{168}$ model. As can be seen, significant correlated lags are found at multiples of 24 and 168 hours of the FACF, whereas the FPACF of the series shows a decaying pattern on the seasonal lags of the series. Hence, there are daily and weekly effects that have not been captured by the explanatory variables and the AR terms that should be modelled. For example, the FACF of the daily lags shows two significative values at lags 24 and 48, whereas the FPACF of the same lags shows an exponential decrease. This behaviour corresponds to a seasonal moving average effect of order two, which should be added in our final SARMAHX model in order to capture the dynamic of the functional time series being modeled.

In order to select the optimal autoregressive and moving average orders of the SARMAHX model, the identification methodology presented in Chapter 3 has been used. Regular and

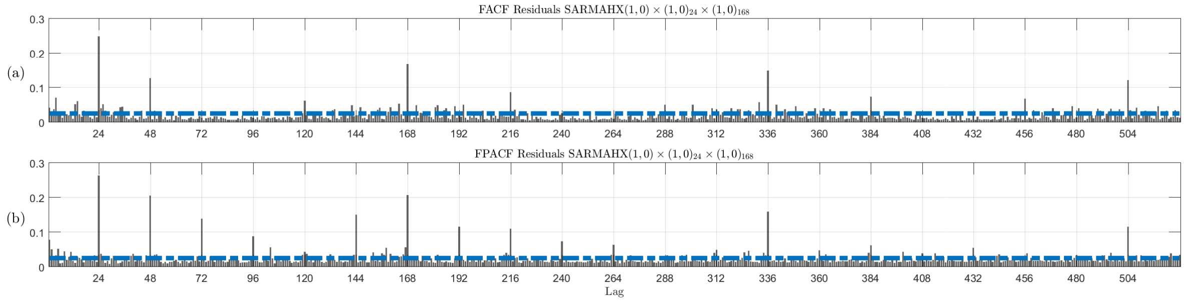

Figure 5.5. First 528 lags of the functional auto- and partial autocorrelation function (FACF / FPACF) of the residuals of the fitted SARMAHX$(1,0) \times (1,0)_{24} \times (1,0)_{168}$ concurrent model including exogenous variables. A high autocorrelation value is found at multiples of 24 and 168 hours. The horizontal line represents the 95% upper limit of the distribution of the statistic under the white noise hypothesis.

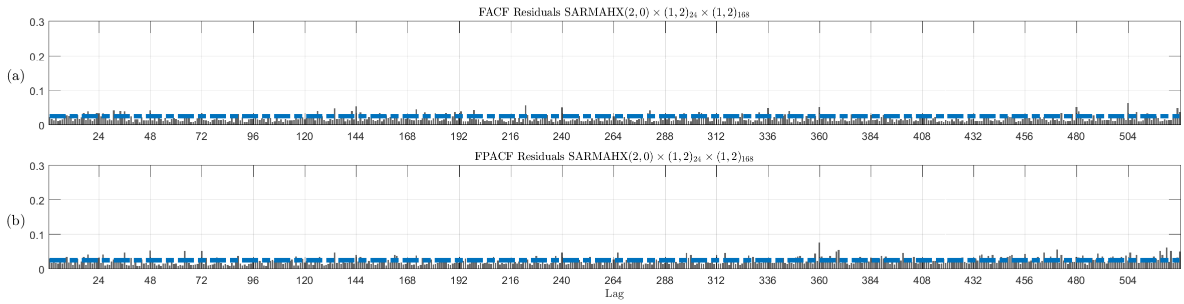

Figure 5.6. First 528 lags of the functional auto- and partial autocorrelation function (FACF / FPACF) of the residuals of the fitted SARMAHX$(2,0) \times (1,2)_{24} \times (1,2)_{168}$ concurrent model including exogenous variables. As there are no significative autocorrelation values for any lags, the residuals can be regarded as a functional white noise process. The horizontal line represents the 95% upper limit of the distribution of the statistic under the white noise hypothesis.

seasonal moving average terms have been added into the model to account for the significant correlated lags found in the previous FACF/FPACF residual plot of the previous model. An iterative process has been carried out until the residuals of the fitted model are functional white noise. The final adjusted model for both concurrent and functional cases is a SARMAHX$(2,0)$ $\times(1,2)_{24}$ $\times(1,2)_{168}$ model. Throughout this section, these models will be denoted by SARMAHX (functional case) and SARMAHXconc (concurrent case). Figure 5.6 show the autocorrelation and partial autocorrelation functions of the residuals of the fitted concurrent SARMAHX model with the 95% upper bound of the statistics under the white noise hypothesis. As most values fall below this upper bound, the hypothesis of the residuals being white noise cannot be rejected, hence validating the model from a linear point of view.

It is worth analyzing the resulting shapes of the fitted functional parameters. Figure 5.7 shows the concurrent regressor operators. 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 energy offered at all prices increases. In fact, the offer increases more at higher prices than at lower prices, meaning 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 more energy is imported, the supply curve has less energy being offered.

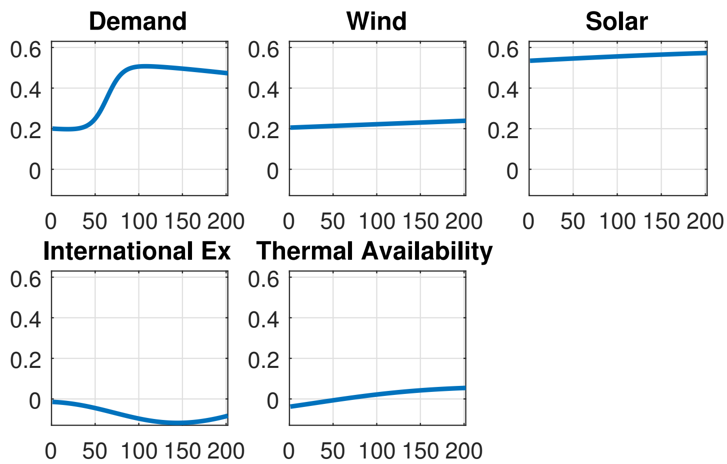

Figure 5.7. Operators' kernels for the regressors of the SARMAHX model in Supply Curve forecasting study.

5.2.1.1. Empirical comparison

This section compares the fitted SARMAHX models with some reference models. Two different analysis are presented. On the one hand, as the proposed model is trained to minimize the one-step ahead forecast error, a one-hour ahead forecast is analyzed. However, in the Italian electricity market, the auctions for the 24 hours of the day are cleared at the same time. Therefore, a 24-hour ahead forecast is also considered, validating the use in a real case application.

All the models included in this study were fitted using data from the In-Sample period. Once their parameters have been estimated, estimations are obtained for both the In-Sample and the Out-Of-Sample period, without re-calibrating the parameters of the models. For the 24-hour ahead forecast, a rolling window approach is used: when forecasting the first hour $h = 1$ of day $D + 1$, the estimation $\widehat{S}_{D+h}$ uses the real curves up to day D. However, when forecasting hours $h > 1$, the real curves have not been observed, so past estimations of the model are used as inputs for the model.

Following (Ferraty and Vieu, 2006) and (Portela, Muñoz San Roque, et al., 2018), forecasting models for functional time series can be classified in 3 main groups, according to the techniques that are used to obtain estimations of the parameters of the model: parametric, nonparametric and dimensionality reduction models. Parametric models, such as the proposed SARMAHX model, assume that the parameters can be expressed as integral operators. Conversely, non-parametric models do not define a fixed structure for the operators of the model, instead relying on kernel estimators or other non-parametric techniques to fit the parameters of the models. Finally, dimensionality reduction models pursue the goal of transforming the functional time series into a reduced set of scalar variables so that multivariate techniques can be applied to the estimation of the final model. In order to provide a fair comparison, one model for each group will be included in this study, to test the forecasting capabilities of the proposed model against the best performing functional models found in the literature. The models to be compared are described ahead:

- *Naïve*. Simple benchmark model that provides a reference to compare the tested 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. In the second case, for Tuesdays, Wednesdays, Thursdays and

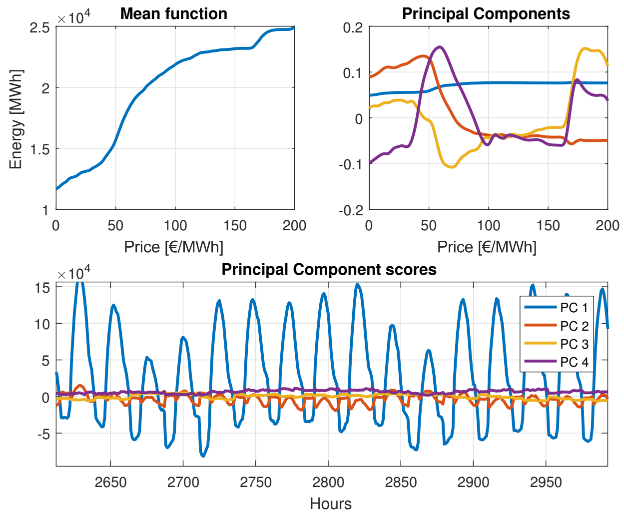

Figure 5.8. Mean function, Principal Components and scores of the supply curves.

Fridays, the forecast will be the hourly curve of the previous day, while for Saturdays, Sundays and Mondays, the forecast will be the hourly curve of the previous week.

- *Principal Components approach*. This method is used in Pelagatti (2012) for supply curves forecasting. It extracts the first Functional Principal Components of the curves and the corresponding time series of scores. Then, the scores are forecasted by means of Transfer Function (TF) models (Pankratz, 1991), which include explanatory variables. The final estimation of the curves is done by reconstructing the estimated scores. The In-Sample period is used for training all models. 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.

  Figure 5.8 represents 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. It is worth noting that each forecasted curve from this method is a linear combination of only the 3 or 4 principal component extracted. These models will be denoted as PC_FT3 and PC_FT4, respectively.

- *Functional nonparametric approach*. When forecasting functional time series, one of the most popular models found in the literature is the functional nonparametric model (the reader is referred to (Ferraty and Vieu, 2006) for a comprehensive review of nonparametric techniques applied to functional data). The functional nonparametric model (Ferraty, Van Keilegom, et al., 2012) assumes that the functional response Y_t can be modelled as a non-linear function of functional covariates X_t.

$$Y_t(v) = m(X_t)(v) + \varepsilon_t(v), \tag{5.3}$$

  where $m(.)$ is an unknown nonlinear operator and ε_t is a functional white noise process. Nonparametric models do not define a fixed structure for the operator $m(.)$, instead they employ a functional version of the Nadaraya-Watson kernel estimator to estimate the regression operator:

$$\widehat{m}_h(X_t) = \frac{\sum_{i=1}^{T} Y_i K\left(h^{-1} d(X_t, X_i)\right)}{\sum_{i=1}^{T} K\left(h^{-1} d(X_t, X_i)\right)}, \tag{5.4}$$

where K is a kernel function, $h > 0$ is a bandwidth parameter and d is a semimetric. A nonparametric version of the functional autoregressive model of order 1 can be defined with this model, using past realizations of the process as the functional covariate:

$$Y_t(v) = m(Y_{t-1})(v) + \varepsilon_t(v). \tag{5.5}$$

This model was used in (Shah and Lisi, 2019) to obtain forecasts of supply and purchase curves in the Italian Day-Ahead market in order to obtain accurate estimations of the market price. However, as all the models that will be compared against the proposed SARMAHX model use exogenous variables, a variation of this model will be included in the comparison study instead. The semi-functional partial lineal model (Aneiros, Vilar, Cao, et al., 2013; Vilar, Aneiros, et al., 2018) allows the inclusion of exogenous scalar variables, generalizing the functional nonparametric model (Equation 5.3). The expression of the model is as follows:

$$Y_t(v) = \boldsymbol{x}_t^{\intercal}\boldsymbol{\beta}(v) + m(Y_{t-1})(v) + \varepsilon_t(v) \tag{5.6}$$

where $\boldsymbol{x}_t^{\intercal} = (x_{t,1}, \ldots, x_{t,p})$ is a vector of p exogenous scalar covariates and $\boldsymbol{\beta}(v) = (\beta_1(v), \ldots, \beta_p(v))$ is a vector of unknown functional parameters to be estimated. The authors obtain estimations for these parameters using kernel smoothing and ordinary least squares ideas. In order to fit this model to the supply curve dataset, the parameters have been selected as follows: the kernel function K used was the Epanechnikov, defined as $K(u) = 3/4(1 - u^2)$; the bandwidth parameter h has been selected using the k-nearest neighbours method proposed in Antoniadis, Paparoditis, et al. (2009); and the semimetric d selected is based on the L^2 norm of the curves. Throughout this section, this model will be denoted as NPARHX.

All the models are trained with the In-Sample period and using the same exogenous variables. 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. In the 24-step ahead forecast study, hour 1 will always be estimated using a forecasting horizon of 1; hour 2 will be estimated using a forecasting horizon of 2; and finally the forecasting horizon used to estimate hour 24 will be 24. Henceforth, 1-step and 24-step ahead forecasts are produced for the whole data range. Functional errors FMAE and FRMSE are calculated, which are defined as

$$\text{FMAE} = T^{-1}\sum_{t=1}^{T}\int |Y_t(u) - \hat{Y}_t(u)|du \tag{5.7}$$

$$\text{FRMSE} = \sqrt{T^{-1}\sum_{t=1}^{T}\int \left(Y_t(u) - \hat{Y}_t(u)\right)^2 du} \tag{5.8}$$

When forecasting certain types of data, some post-processing method has to be implemented to improve the raw output of the forecasting model being used. One such case is the one that concerns us: forecasting aggregated supply curves in the Day-Ahead electricity market. These curves are always monotonically increasing curves, due to the sheer definition of the supply functions. The formulation of the SARMAHX model does not ensure that the estimations of the model are non-decreasing functions, so in order to obtain faithful estimations of these curves

Forecasting horizon	Model	In-Sample FMAE [MWh]	In-Sample FRMSE [MWh]	Out-Of-Sample FMAE [MWh]	Out-Of-Sample FRMSE [MWh]
1-Step Ahead	Naïve	909.39	1258.46	818.59	1179.16
	PC_FT3	353.85	470.51	480.54	694.15
	PC_FT4	276.65	378.83	416.05	569.36
	NPARHX	389.30	553.53	641.92	884.50
	SARMAHX	264.62	363.10	313.11	429.66
	SARMAHXconc	**248.58**	**354.97**	**288.66**	**407.81**
24-Step Ahead	Naïve	1075.3	1433.5	1382.3	1812.50
	PC_FT3	672.06	871.33	856.62	1148.1
	PC_FT4	639.46	834.46	811.69	1085.07
	NPARHX	743.47	1010.01	987.41	1311.51
	SARMAHX	590.08	787.34	709.70	963.49
	SARMAHXconc	**557.37**	**764.73**	**655.19**	**885.86**

Table 5.1. Average errors for each method in the one-step and 24-step ahead prediction in Supply Curve forecasting study. The lower errors obtained for each forecasting horizon are marked in bold.

with the proposed model, output curves of the model $\widehat{Y}_t$ are transformed into monotonically increasing curves $\widehat{Y}_t^+$ obtained as the solution of the following optimization problem:

$$\underset{\widehat{Y}_t^+}{\text{minimize}} \quad \frac{1}{2}\left\|\widehat{Y}_t^+ - \widehat{Y}_t\right\|^2 \tag{5.9a}$$

$$\text{subject to} \quad 0 \leq \widehat{Y}_t^+(v_i) \leq \widehat{Y}_t^+(v_{i+1}),\ i = 1, \ldots, N-1, \tag{5.9b}$$

where $\{v_1, \ldots, v_N\}$ denotes the discretization points of the functional observations. The objective function to be minimized (5.9a) is the distance between the estimated curves and their monotonic transformation, while constraint (5.9b) ensures that the new curve is non-decreasing. The optimization problem is solved by linear least squares.

This transformation guarantees that the estimated curves are monotonically increasing, while retaining the shape that has been estimated by the SARMAHX model. It should be noted that the problem of estimating curves that can be non-monotonic curves is a known issue that happens to the vast majority of forecasting models, for example if the parameter estimates for one regressor is negative and a new value of this variable is positive. Henceforth, in order to provide a fair comparison between the SARMAHX model and other forecasting models found in the literature, the proposed transformation will be applied to the estimates of all models considered in this section.

Table 5.1 shows the functional errors for the 1-step and the 24-step ahead forecasts for both the In-Sample and Out-Of-Sample periods. The proposed SARMAHX model shows a clear advantage over the reference models, obtaining a FMAE of 288.66 MWh that is much lower than the FMAE of the PC_FT4 (the best reference model), which achieves a FMAE of 416.05 MWh in the Out-Of-Sample period of the 1-step ahead forecasting study. For the 24-step ahead forecasting case, the results are similar: while the PC_FT4 model obtains a FMAE of 811.69 MWh, it is not able to improve the results of the concurrent SARMAHX model, that achieves a

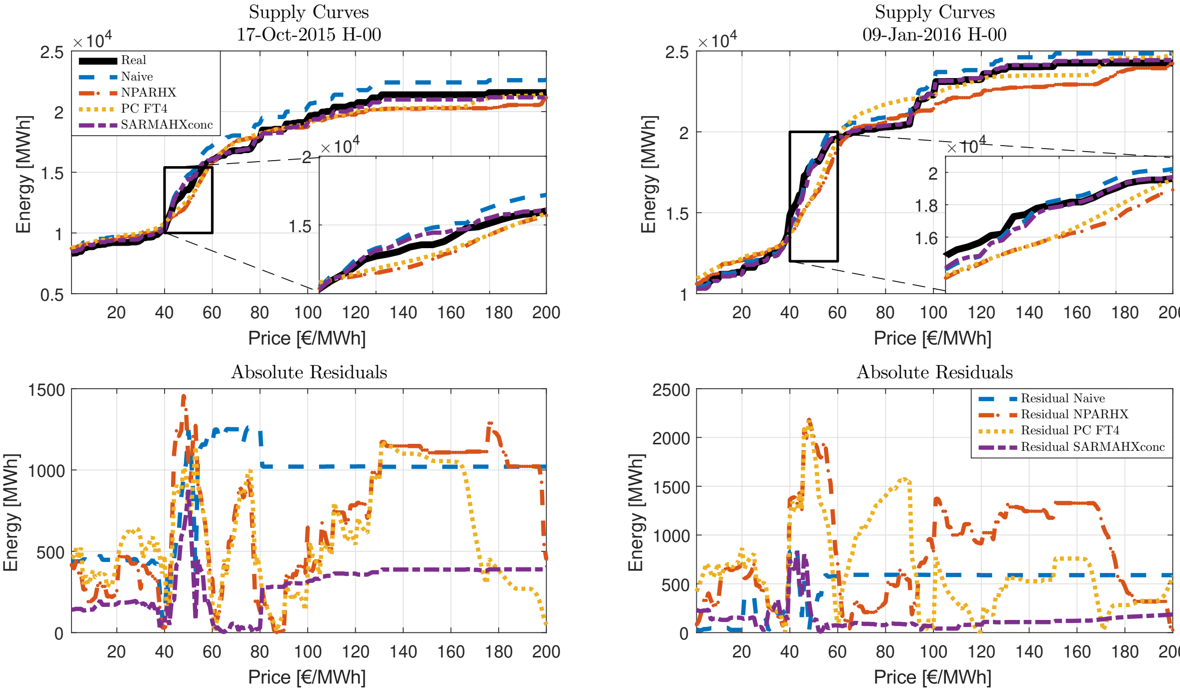

Figure 5.9. Top: Forecast examples for 1-step ahead estimations in the Out-Of-Sample periods in Supply Curve forecasting study. Curves in the price range $[40, 60]$ €/MWh are shown in detail, highlighting the differences between the different forecasting models. Bottom: Absolute value of the residual curves of the models.

FMAE of 655.19 MWh in the Out-Of-Sample period. The estimation results are analyzed ahead in detail.

Firstly, 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, with a mean absolute error of 289 MWh in the Out-of-Sample period, whereas the FMAE of the PC_FT4 model is 416 MWh. This difference validates the usefulness of the SARMAHX model as a competitive model to forecast supply curves. Several conclusions can be drawn from these results. Firstly, even though FPCA approach also uses a time series model, the fact of reducing the dimensionality significantly affects the performance. As the Principal Components are kept untouched, the forecasts cannot adapt to changes in the functional time series and the reconstruction of the curves 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 account for changes in the series. In addition, as the nonparametric approach does not take into account the seasonality structure of the data, it is not able to achieve the same results as the other reference models. Finally, the use of concurrent operators in the SARMAHX model returns better results than integral operators. This means that the cross correlations between supply curves might not be very important, hence, the adjustment of the parameters might be easier for the concurrent model. Figure 5.9 shows some forecasting examples for the one step-ahead case, where the capability of the concurrent SARMAHX model of estimating the shape of the supply curves is highlighted. Whereas the proposed model is able to accurately estimate the supply curves, the reference models are not capable of capturing the complex bidding behaviour exhibited by the supply curves, providing only a smooth approximation to the curves that does not take into account the step-wise nature of the curves.

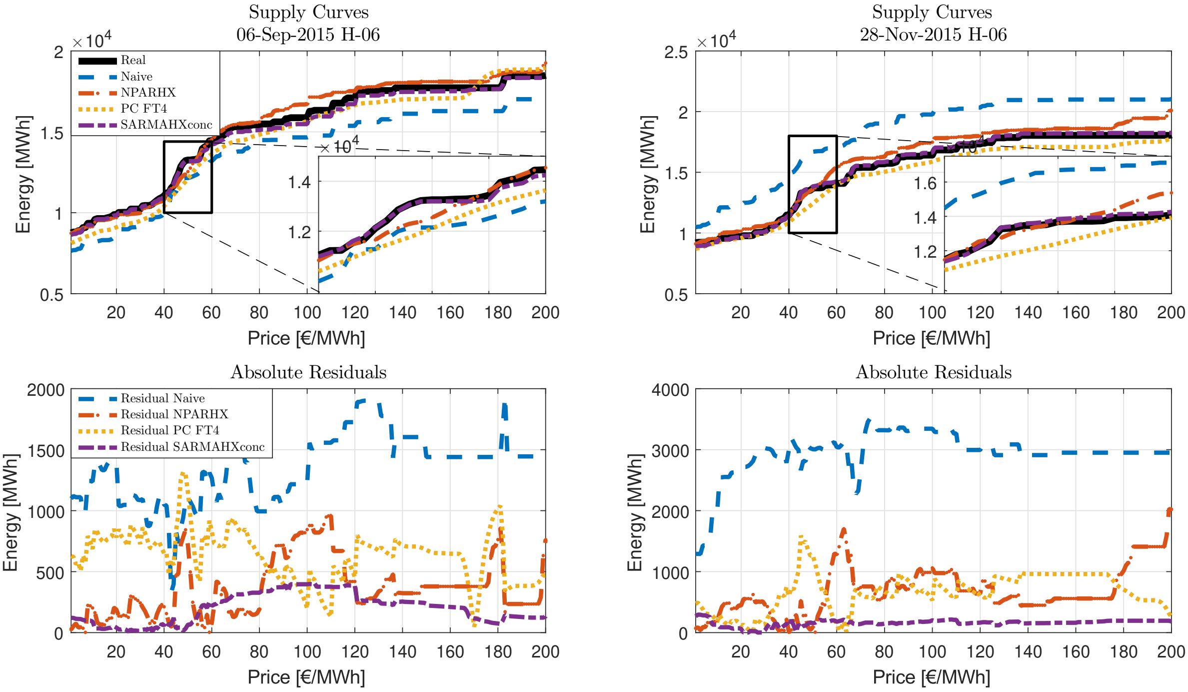

Figure 5.10. Top: Forecast examples for 24-step ahead estimations in the Out-Of-Sample periods in Supply Curve forecasting study. Curves in the price range $[40, 60]$ €/MWh are shown in detail, highlighting the differences between the different forecasting models. Bottom: Absolute value of the residual curves of the models.

Secondly, the 24-step forecasts show similar results to the 1-step ahead case. The concurrent SARMAHX model provides better average errors with respect to the other models. While the PC_FT4 model has a mean absolute error of 811 MWh for the Out-Of-Sample period, the concurrent SARMAHX model has a much lower mean absolute error of 655 MWh, proving itself to be a competitive forecasting 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. Supply 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 behaviour for the whole day. This is the case if there is some technical constraint 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 5.11. Forecasting the first hour is very important to the rest of the prediction, specially for the first hours of the day. Figure 5.10 shows some forecasting examples for the 24-step case. As can be seen, the concurrent SARMAHX model provides better estimations of the supply curves than the other models, obtaining a much lower error than the reference models.

Additionally, Tables 5.2 and 5.3 show the Diebold-Mariano (DM) (Diebold and Mariano, 1995) test applied to each pair of forecasts. This test is a head-to-head method which compares the forecast error of two different models. Each cell of Tables 5.2 and 5.3 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 forecasting errors can be considered statistically different. The DM test validates the Out-Sample results, yielding very low p-values between model comparisons. Then, the equal forecast null hypothesis can be rejected. This analysis emphasizes the improvement between the proposed model and the other models considered in this study.

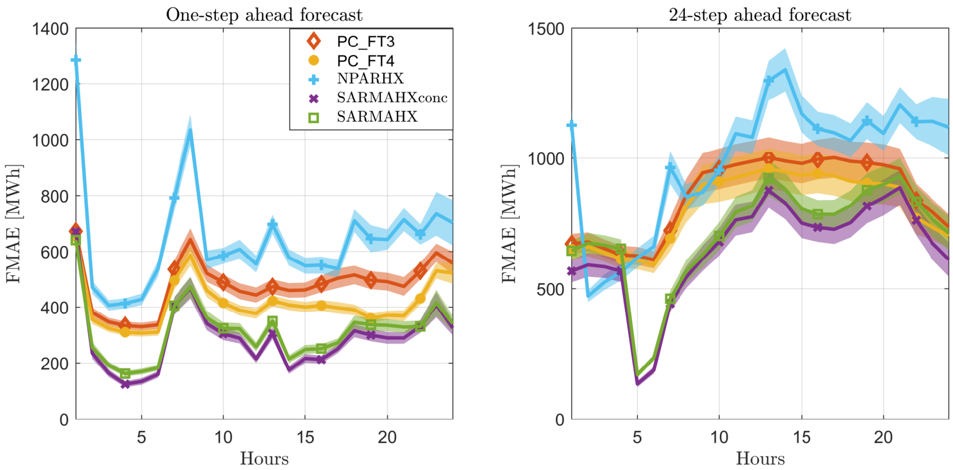

Figure 5.11. FMAE for each hour in the 1-step and 24-step ahead forecasts in the Out-Of-Sample period in Supply Curve forecasting study. As the FMAE error of the Naive model is significantly higher than the other models, it is not included. The shaded regions are 90% confidence bands for the FMAE.

Models	Naïve	PC_FT3	PC_FT4	NPARHX	SARMAHX	SARMAHXconc
Naïve	-					
PC_FT3	**0**	-				
PC_FT4	**0**	**0**	-			
NPARHX	**0**	**0**	**0**	-		
SARMAHX	**0**	**0**	**0**	**0**	-	
SARMAHXconc	**0**	**0**	**0**	**0**	**0**	-

Table 5.2. Diebold-Mariano test's p-values for Out-Of-Sample period in the one-step ahead forecasts in Supply Curve forecasting study. p-values that are lower than 0.05 are marked in bold.

5.2.1.2. In-depth analysis of the results

In the Italian Day-Ahead market, the market-clearing price is usually located in the price range $[30, 70]$ €/MWh (Shah and Lisi, 2019), hence the volume of the agents' bids will be significantly higher in that portion of the aggregated supply curve. Thus, obtaining accurate forecast of the supply curve in that region is of utmost importance for any company. As can be seen in both Figure 5.9 and Figure 5.10, the concurrent SARMAHX model is able to capture the shape of the supply curve in that price range, providing an accurate description of the agents' bidding behaviour in the zone of interest of the supply curve.

The shape of the estimated supply curves is essential when analyzing the strategic offering of the competitors of an electricity company, because it contains valuable information about the offering behaviour of the agents. Forecasted supply curves, together with estimations of the demand, can be used to define RDCs, which are often used to calibrate market equilibrium models (Díaz et al., 2010) as well as offering optimization models (Baillo et al., 2004). In Portela, Muñoz, et al. (2017), the authors analyze the importance of obtaining faithful estimations of the slopes of RDCs, since it indicates an agent's capability to influence market prices. This

Models	Naïve	PC_FT3	PC_FT4	NPARHX	SARMAHX	SARMAHXconc
Naïve	-					
PC_FT3	**0**	-				
PC_FT4	**0**	**0**	-			
NPARHX	**0**	**0.001**	**0**	-		
SARMAHX	**0**	**0.0001**	0.1162	-		
SARMAHXconc	**0**	**0**	**0**	**0**	**0**	-

Table 5.3. Diebold-Mariano test's p-values for Out-Of-Sample period in the 24-step ahead forecasts in Supply Curve forecasting study. p-values that are lower than 0.05 are marked in bold

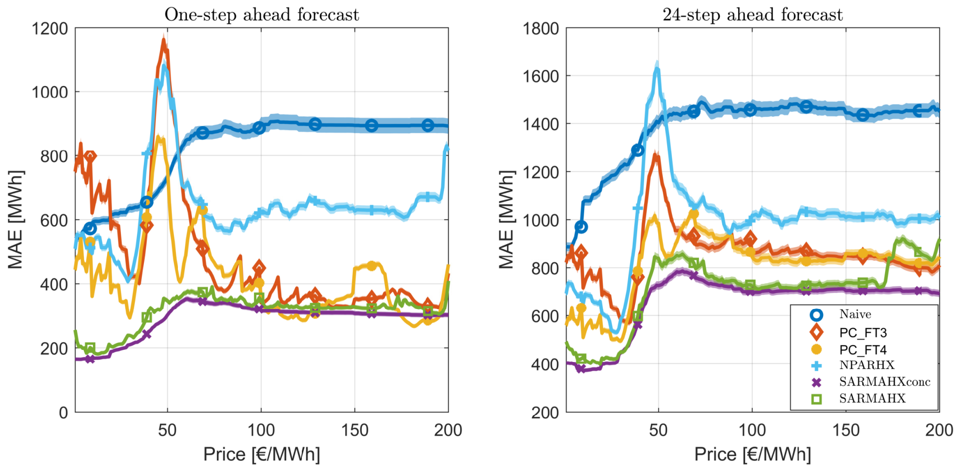

Figure 5.12. Mean absolute error for each price in the 1-step and 24-step ahead forecasts in the Out-Of-Sample period in Supply Curve forecasting study. The shaded regions are 90% confidence bands for the MAE.

highlights the need for a forecasting model that not only captures the general level of the supply curves, but also provides an accurate estimation of the shape of the curve.

As can be seen in Figure 5.12, MAE errors are presented for each price for the Out-Of-Sample period and for the 1-step and 24-step ahead forecasts, which are obtained as the mean value of the absolute error for each price of the curves. The figures show the performance of each method for different ranges of prices. It can be seen how the the SARMAHX 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. The most significant improvement is seen in the aforementioned price range of interest: for example, whereas the mean absolute error of the PC_FT4 model at price 50 €/MWh is 1000 MWh, the concurrent SARMAHX model achieves a mean absolute error of 730 MWh in the 24-step ahead forecast. This difference is more pronounced in the one-step ahead study, where the mean absolute errors of the PC_FT4 model in the price range of interest are close to 800 MWh as opposed to the errors of the SARMAHX model, which are near 350 MWh. In addition, confidence bands for the MAE errors are included to ease the interpretation of the results. As the confidence bands do not overlap, it can be

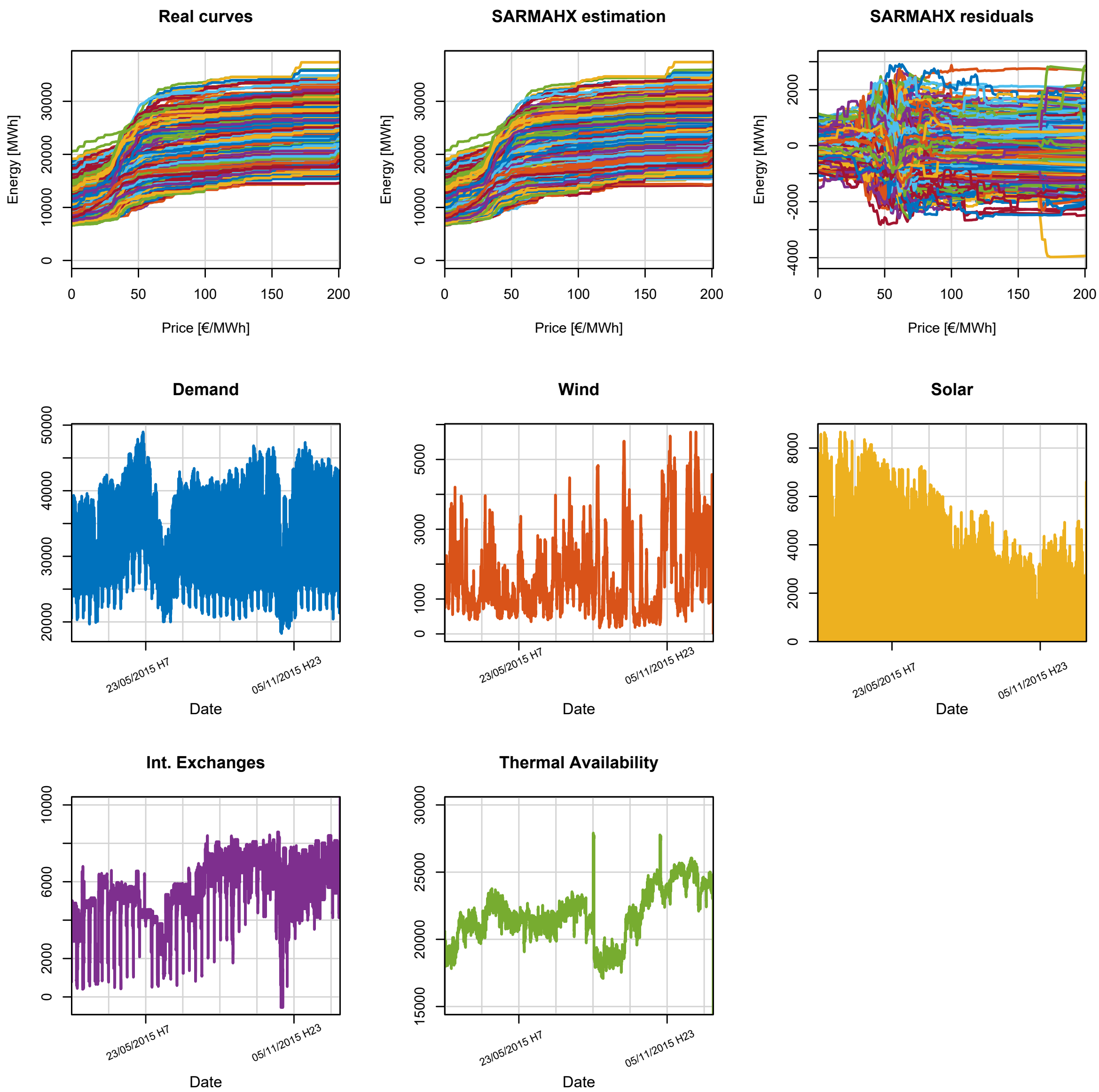

Figure 5.13. Top: actual hourly aggregated Supply Curves together with the 1-step ahead forecasts obtained by the fitted SARMAHX model and the residuals of the model. Bottom: Hourly values of the five exogenous variables included in the regression part of the SARMAHX model.

concluded that the errors of the SARMAHX models are significantly lower than the other models considered in this study.

5.2.2. Probabilistic forecasting

The previous section was devoted to obtaining short-term deterministic forecasts of the hourly aggregated supply curves in the Italian Day-Ahead market. Figure 5.13 shows the real supply curves and the one-step ahead forecasts of the concurrent SARMAHX model identified in the previous section, together with the set of exogenous variables that have been used in this case study. Once the deterministic forecasts have been obtained the uncertainty associated with those estimations must be quantified, in order to provide the practitioner with a measure of the confidence given to the forecasts. This information can then be used to obtain coherent scenarios of future residual demand curves, which can then be used by market agents to optimize its bidding strategy, hedging its market position against the possible strategies of its competitors.

As mentioned earlier, in the Italian Day-Ahead market the auctions for the 24 hours of the day are cleared at the same time. Hence, in order to simulate the real-world application of the proposed forecasting methodology, functional scenarios of the aggregated supply curves will be obtained for the one-step and 24-step ahead SARMAHX estimations.

The setup for this study is as follows: all the probabilistic models included in this comparison were fitted using data from the In-Sample period defined in the previous section. Once the internal parameters of the models have been selected, prediction bands for both the In-Sample and the Out-Of-Sample period are obtained, without re-calibrating the models. Functional scenarios of the supply curves are also obtained following the methodology detailed in Section 4.2.2, and finally the probabilistic performance metrics reviewed in Section 4.3 are used to compare the applicability of the proposed conditional density model against the remaining reference models considered in this study, which are described ahead:

- *Unconditional density estimator*. This model is based on the concept of projection-based density for functional data proposed in Delaigle and Hall (2010). The authors use a standard kernel density estimator to approximate the distribution of the FPC scores of the curves, which can then be used to generate new functional scenarios by sampling from the fitted distribution. The bandwidth parameters of the kernel density estimator is selected using the scores of the FPC decomposition of the SARMAHX's residuals for the In-Sample period. Then, the model assumes that the distribution of the residuals for the Out-Of-Sample period is the same as the In-Sample period, so prediction bands and functional scenarios can be obtained for the whole range of data. This model aims at estimating the unconditional density of a set of functional data, so it is not able to include information about exogenous variables in its formulation.

- *Bootstrap resampling*. The most common technique used in the literature to obtain probabilistic forecasts of functional time series is using bootstrap resampling of the residuals of a fitted deterministic model to quantify the uncertainty of the forecasts (Raña, Aneiros, et al., 2016; Vilar, Aneiros, et al., 2018; Rossini and Canale, 2019). This model assumes that the residuals do not exhibit any kind of dependence, so the bootstrap pseudo-replications of the curves are able to approximate the true underlying distribution of the non-deterministic part of the forecasting model. Similarly to the model presented above, the bootstrap model only takes into account the set of functional residuals, without including information about the possible influence of external factors to the distribution of the curves.

Unlike the previous reference models, the conditional density model proposed in this thesis assumes that the distribution of the SARMAHX residuals is influenced by a set of covariates. Hence, once the In-Sample residuals of the SARMAHX model have been decomposed into their first 4 FPC (as they are able to account for more than 95% of the variance of the original data), an exploratory analysis is carried out to identify possible relations between the set of exogenous variables and the projection-based distribution of the residuals. Figure 5.14 illustrates the differences in the distribution of each FPC's score for different levels of the covariates being used in this case study. In addition, the hour of the day is included in this analysis, as it can be a driving factor to identify periods of the day where the uncertainty of the model is lower. As can be seen, for the one-step ahead forecasts the hour of the day is a good indicative of the dispersion of the residuals: for hours in the 1-2 hour range the variance of the residuals is higher than the remaining hours, with hours in the range 3-6 showing narrower distributions. This may be caused by competitors defining their bidding strategy for whole days, which makes

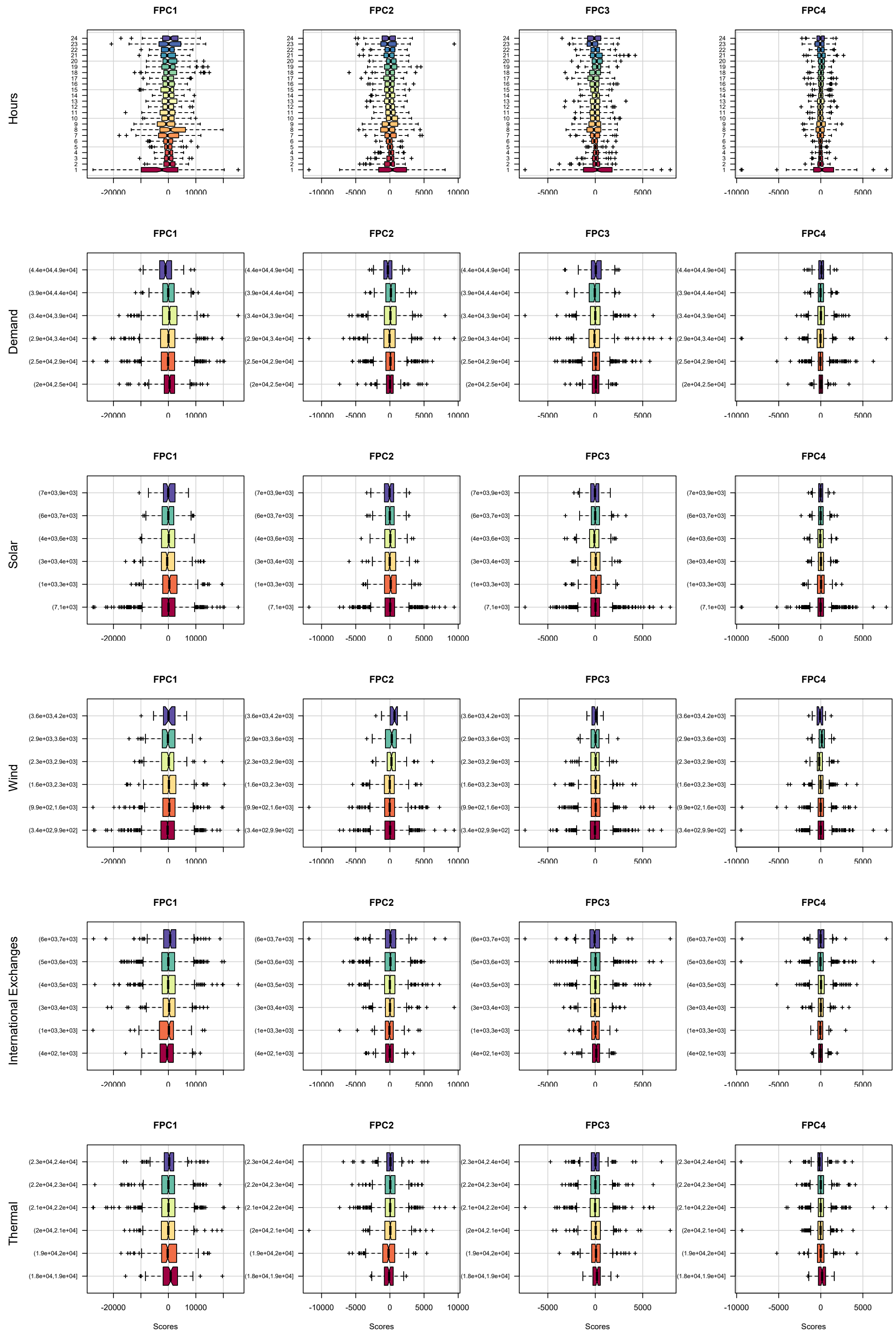

Figure 5.14. Distribution of the FPC scores of the SARMAHX 1-step ahead residuals for the Supply Curve forecasting study, grouped by levels of the exogenous variables. Each column is associated with the scores of a certain FPC, whereas each row is associated with a different covariate.

	FPC1	FPC2	FPC3	FPC4
Hours	**0**	**0**	**0.00092**	**0**
Demand	**0.02533**	**0.0361**	0.1091	0.09668
Solar	**0.01839**	0.1196	0.06406	0.1027
Wind	**0.00058**	**0.00396**	**0.01904**	**0.00029**
International Exchanges	**0.0054**	0.06945	0.0517	**0.0276**
Thermal	**0.00681**	**0.00182**	**0.0422**	**0.00026**

Table 5.4. p-values of the Kolmogorov-Smirnov test obtained for the distributions of the one-step ahead residuals' scores grouped by levels of the exogenous variables used in the Supply Curve forecasting study. p-values lower than 0.05 are marked in bold.

the first hour of the day the most difficult to estimate given the historical data. The effect of the remaining covariates is not as clear as the effect of the hour, so the the two-sample Kolmogorov-Smirnov (KS) test (Hollander and Wolfe, 1973) is used to test the statistical difference between the distribution of the scores for different values of the remaining covariates. This is a nonparametric test that assesses whether two samples have been drawn from the same distribution. If the p-value of the statistic is lower than 0.05, the null hypothesis of both samples being drawn from the same distribution can be rejected at a 95% confidence level. Table 5.4 contains the minimum p-value obtained for the grouped distributions of the exogenous variables shown in Figure 5.14 for each FPC score. This highlights the wind generation and thermal generation as the next 2 variables that most influence the distributions of the one-step ahead residuals' scores, as they are the ones with a significant p-value across all 4 FPC. Specifically, higher wind generation is associated with low dispersion of the residuals, whereas differences in the thermal generation are related with changes in the shape of the residuals, as the differences in the distribution are seen for the second and third FPC's scores. The system demand and solar generation do not generate any significant differences in the distribution of the residuals for different values of those variables.

As a result, variables Wind and Thermal will be used as inputs for the conditional probabilistic model for the one-step ahead forecasts, in addition to the hour of the day previously mentioned. Figure 5.15 illustrates the differences in the distribution of the residuals' scores for different values of the input variables. The dispersion exhibited by the model's residuals is very high for hour 1 and a combination of low wind generation and high thermal availability. This is a very different scenario than the one shown in the second panel, where the historical residuals selected by the model for hour 4 and high wind and thermal production exhibit a lower variance than the previous case. This highlights the capability of the proposed model to capture differences in the uncertainty of the model's forecasts using a set of exogenous covariates.

Once relevant variables for estimating the conditional distribution of the residuals have been identified, the parameters of the model must be estimated. The bandwidth parameters are obtained using the leave-one-out cross-validation procedure described in Section 4.2.1 with data from the In-Sample period, and using standard Gaussian kernels in the model's formulation. As one of the inputs for the model is variable Hour, the circular metric $\|.\|_{S^1}$ introduced in Equation 4.36 is used to capture the notion of proximity between hours in the model. Once fitted, 80% and 95% prediction regions are obtained from the estimated conditional densities in both the In-Sample and the Out-Of-Sample periods. This is illustrated in Figure 5.16, where the prediction regions obtained for all probabilistic models included in this study are shown, together with a set of 100 functional scenarios of the one-step ahead predictions of the supply curves.

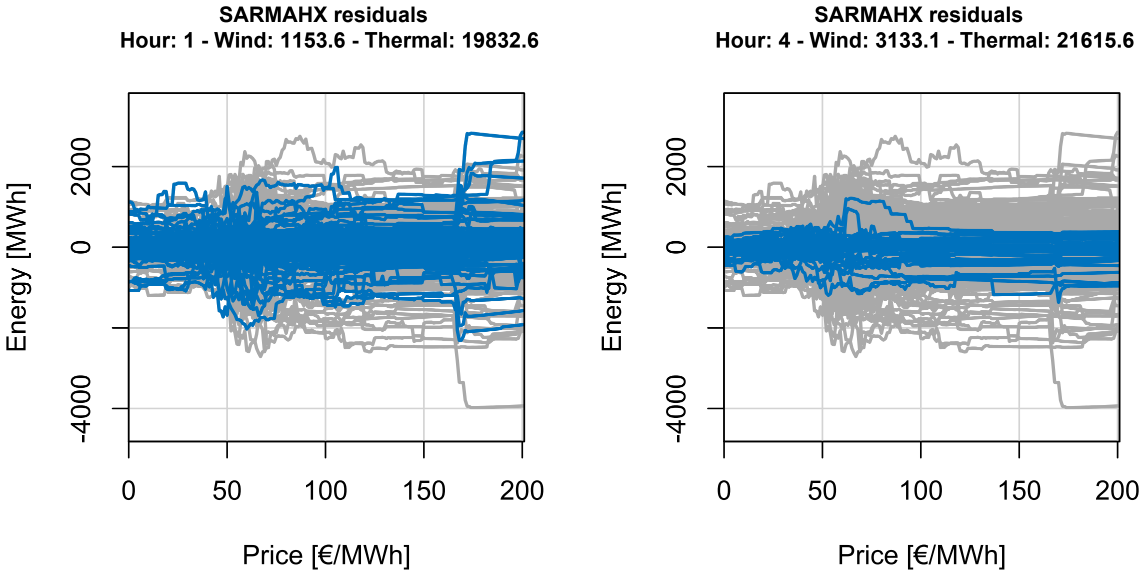

Figure 5.15. Distribution of the residuals of the model for two different values of the exogenous variables Hour, Wind and Thermal for the 1-step ahead forecasts of the Supply Curves. The whole set of residuals is displayed using grey color, whereas the curves with a positive weight in the probabilistic model are shown in blue.

As evidenced by this comparison, the proposed model is able to capture the differences in the uncertainty of the forecasts conditioned to the input variables, which results in narrower predictive regions for low-volatility periods and coherent scenarios of the supply curves being modelled. As the reference models included in this comparison assume that the residuals are distributed independently, the prediction bands obtained by those models often overestimate the uncertainty of the forecasts, which results in wider predictive regions and scenarios of the aggregated supply curves that are not as coherent with the observed curves as the ones obtained by the proposed model for both the In-Sample and Out-Of-Sample periods.

Up to this point, the analysis has been focused on the one-step ahead forecasts of the series of supply curves, as the identification and diagnosis of the fitted concurrent SARMAHX model uses the 1-step ahead estimations of the model. The FACF and FPACF of the one-step ahead residuals of the final SARMAHX model did not show any autocorrelation between different lags of the series, which motivated the use of unconditional probabilistic models as there was no dependence in the distribution of the residuals. However, as illustrated by the previous analysis, although the distribution of the residuals did not show any dependence with its lagged values, there were higher-order moments of the residuals' distribution that were influenced by the values of some exogenous variables. This dependence made the proposed conditional probabilistic model more adequate for generating one-step ahead scenarios of the aggregated supply curves. Nevertheless, in a real application, the auctions for the 24 hours of the following day are cleared at the same time, so market agents would need 24-step ahead scenarios of the curves to optimize its bidding strategy. Henceforth, a similar analysis will be carried out using the 24-step ahead forecasts of the series of supply curves, with the objective of obtaining a set of functional scenarios that capture the uncertainty associated with the estimations of the SARMAHX model evaluated at this forecasting horizon.

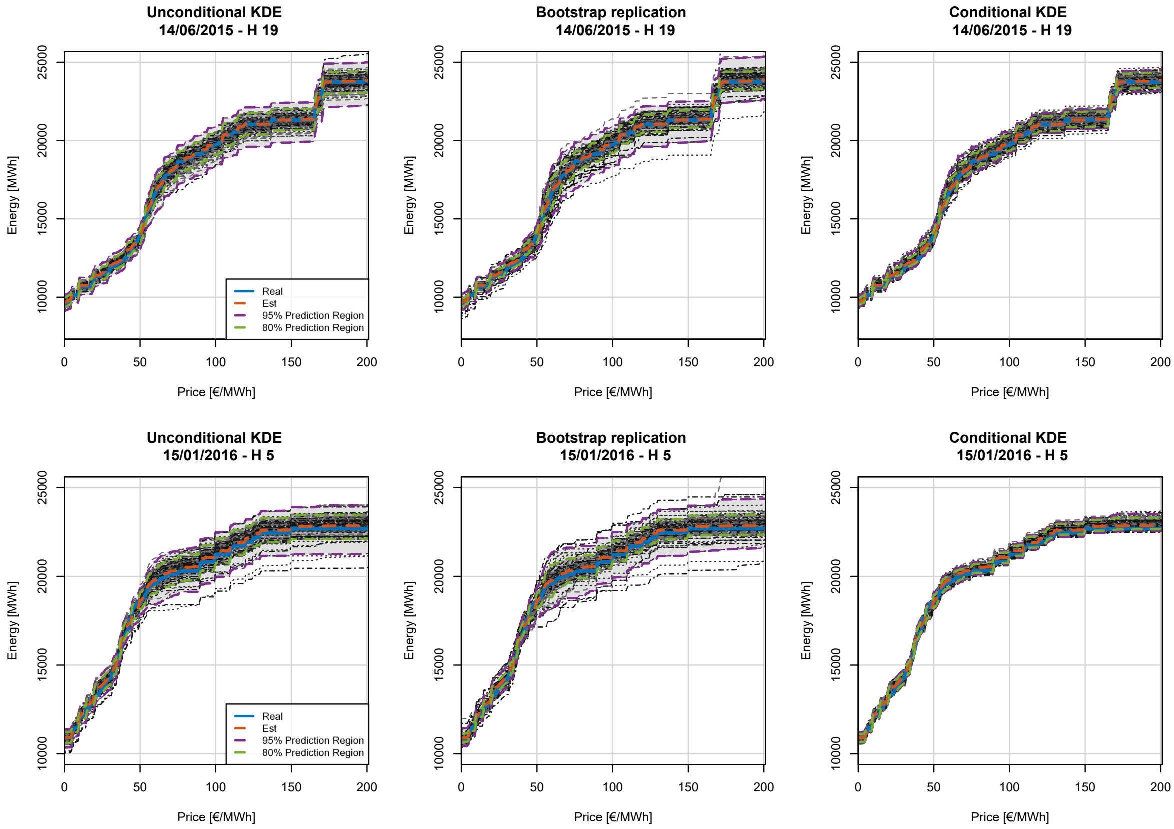

Figure 5.16. Prediction regions defined by the residual-based bootstrap, unconditional and conditional models fitted to the series of Supply Curves for the 1-step ahead forecasting problem. The black curves represent the different scenarios obtained by each method. The top row corresponds to the In-Sample period, whereas the bottom figures display curves for the out-of-sample period.

Following a similar procedure to the one used in the one-step ahead case, the 24-step ahead SARMAHX residuals must be decomposed into the space spanned by its first FPC in order to estimate the projection-based density of the curves. In this setup, the first 4 FPC of the In-Sample residuals account for more than 97% of the original functional dataset's variance. Once this decomposition has been performed, an exploratory analysis has been carried out to identify any dependence between the distribution of the FPC scores and the values of the exogenous variables used in this study. Figure 5.17 shows the distribution of the residuals' scores grouped by certain levels of the exogenous variables. Unlike the one-step ahead case, the distribution of the residuals does not seem to be strongly influenced by the hour of the day, as all hours minus 5-8 present similar distributions. The differences in the scores between hours 5-8 may be explained by changes in other exogenous variables, such as the demand. Nevertheless, the distribution of the 24-step ahead residuals is influenced by variables such as the demand, wind and thermal generation and international exchanges, as seen in the boxplots included in Figure 5.17. Particularly, high demand and wind generation levels influence the mean of the first FPC, whereas different values of thermal generation and international exchanges account for changes in the shape of the residuals (third and fourth FPC). Similarly to the one-step ahead study, the Kolmogorov-Smirnov test is used to select those variables that influence the distribution of the 24-step ahead residuals' scores the most. Table 5.5 contains the p-values of the KS test, which reject the null hypothesis for variables Demand, Wind, International Exchanges and Thermal production having a similar distribution across all FPC.

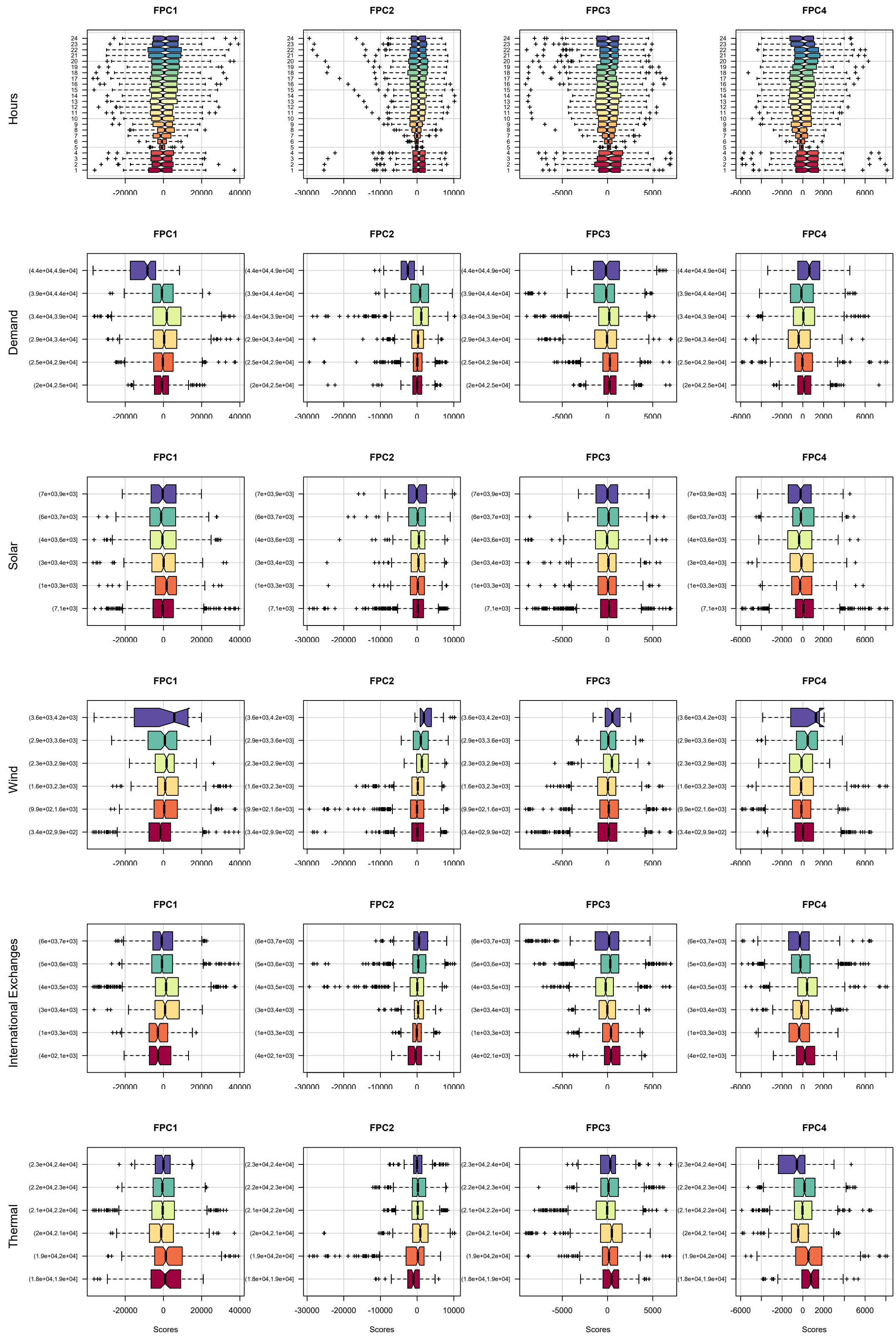

Figure 5.17. Distribution of the FPC scores of the SARMAHX 24-step ahead residuals for the Supply Curve forecasting study, grouped by levels of the exogenous variables. Each column is associated with the scores of a certain FPC, whereas each row is associated with a different covariate.

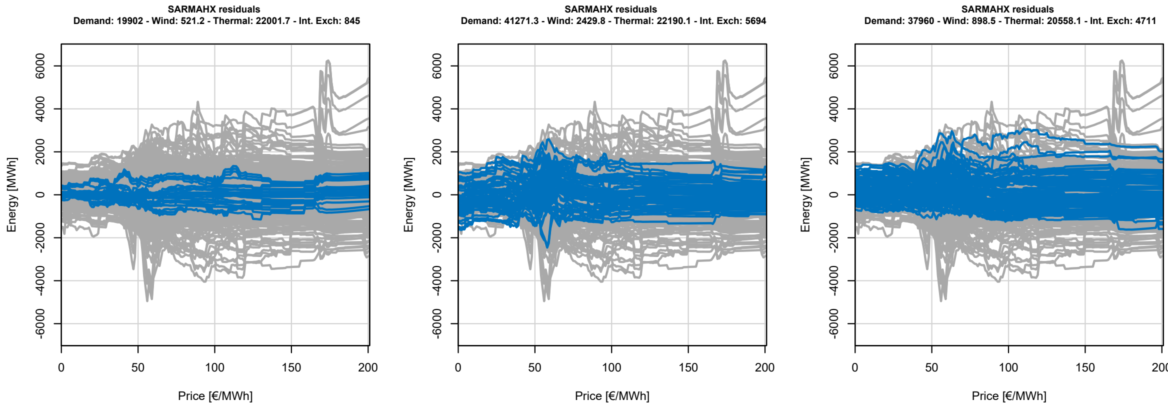

Figure 5.18. Distribution of the residuals of the model for three different values of the exogenous variables Demand, Wind, Thermal and Hydro for the 24-step ahead forecasts of the Supply Curves. The whole set of residuals is displayed using grey color, whereas the curves with a positive weight in the probabilistic model are shown in blue.

	FPC1	FPC2	FPC3	FPC4
Hours*	**0.0037**	0.0862	0.0614	**0.0039**
Demand	**0.00034**	**0**	**0.00051**	**0**
Solar	**0.00135**	**0.0399**	0.1408	0.0724
Wind	**0.00895**	**0.00715**	**0**	**0.01709**
International Exchanges	**0**	**0.00139**	0.01245	**0**
Thermal	**0**	**0**	**0.0226**	**0**

Table 5.5. p-values of the Kolmogorov-Smirnov test obtained for the distributions of the 24-step ahead residuals' scores grouped by levels of the exogenous variables used in the Supply Curve forecasting study. p-values lower than 0.05 are marked in bold. The distribution of variable Hours has been obtained after removing hours 5 and 6, as those effects were explained by other exogenous covariates included in the final probabilistic model.

These differences are illustrated in Figure 5.18, where the set of historical 24-step ahead functional residuals selected by the proposed probabilistic model are shown for three different combinations of the selected exogenous variables. The first image shows a configuration of the variables related with a set of residuals that exhibit low volatility, associated with low values of the demand and wind generation together with high values of thermal generation. The variance of the second set of residuals is higher than the previous set, as the values of both the wind generation and demand are very high. Finally, the third set of residuals exhibits a positive trend, which, according to Figure 5.17, is likely caused by the values of thermal production and international exchanges.

The previous analysis illustrates the influence that exogenous variables have over the distribution of the 24-step ahead residuals of the SARMAHX model. A probabilistic model that is able to capture this dependence will be able to obtain future estimations that exhibit similar shape features as the actual curves. Hence, the proposed conditional probabilistic model is fitted using the selected input variables for the In-Sample period. Once the bandwidth parameters of the model have been selected by the proposed leave-one-out cross-validation procedure, the model is evaluated for both the In-Sample and Out-Of-Sample periods, defining 80% and 95% prediction regions, together with a set of 100 24-step ahead functional scenarios of the supply curves. Figure 5.18 compares the results obtained by the reference models and the conditional probabilistic model for the In-Sample and Out-Of-Sample periods of the data. Although the

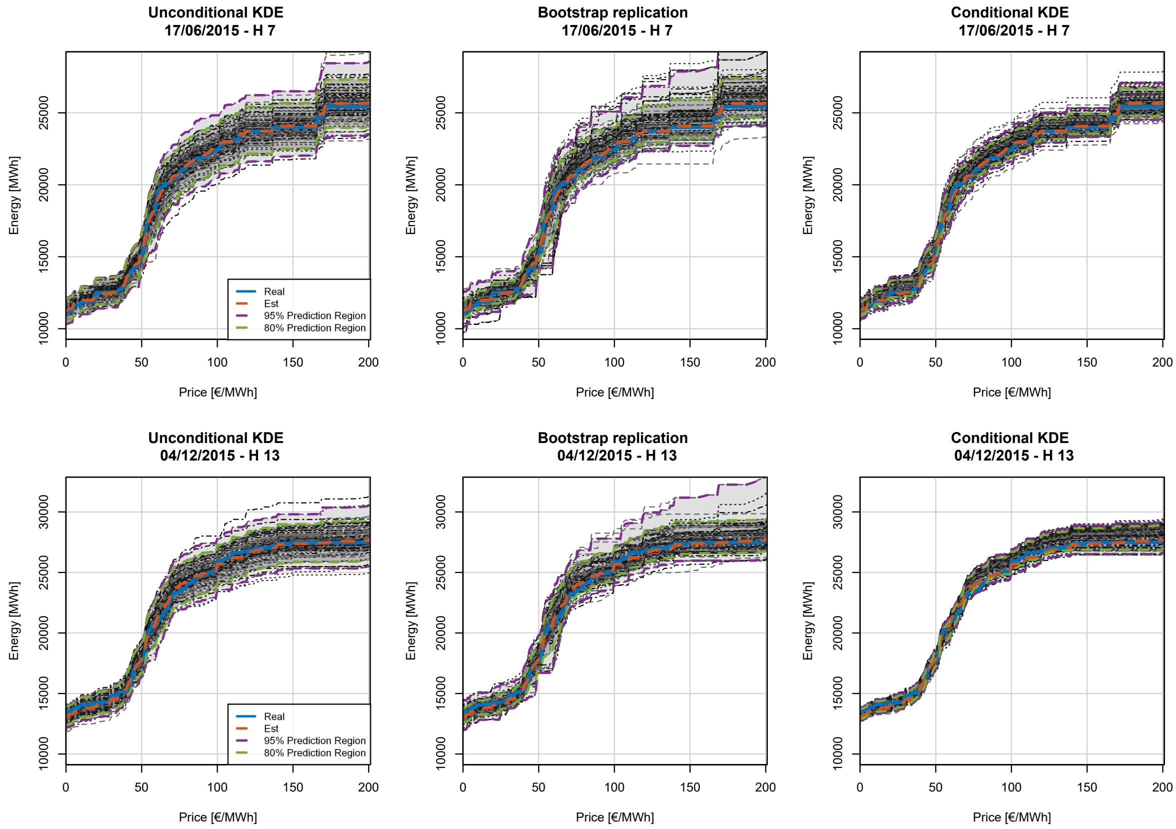

Figure 5.19. Prediction regions defined by the residual-based bootstrap, unconditional and conditional models fitted to the series of Supply Curves for the 24-step ahead forecasting problem. The black curves represent the different scenarios obtained by each method. The top row corresponds to the In-Sample period, whereas the bottom figures display curves for the Out-Of-Sample period.

Forecasting horizon	Model	In-Sample			Out-Of-Sample		
		fCRPS	fWS_{95}	fLS	fCRPS	fWS_{95}	fLS
1-Step Ahead	Bootstrap	20477.17	461073.1	26.17	**23411.44**	462085.2	26.08
	Unconditional	20258.42	445599.3	25.84	23424.07	446792.3	26.03
	Conditional	**19359.63**	**266334.2**	**25.15**	24935.15	**244645.2**	**25.82**
24-Step Ahead	Bootstrap	46642.82	900791.5	28.93	52663.89	906108.8	28.68
	Unconditional	44804.68	810456.7	28.53	**52477.56**	815007.2	28.48
	Conditional	**21502.56**	**381369.6**	**26.19**	67737.62	**405488.4**	**27.76**

Table 5.6. Average scores for each probabilistic model in the 1-step and 24-step ahead Supply Curve scenario generation study. The lower scores obtained for each metric are marked in bold.

uncertainty of these forecasts if higher than the one-step ahead case (as expected, due to the greater forecasting horizon), the proposed model is able to produce sharper probabilistic forecasts of the curves, obtaining narrower prediction regions than the reference models.

In order to provide a fair comparison of the performance of the models, Table 5.6 shows the values of the probabilistic scores defined in Section 4.3 for the one-step and 24-step ahead forecasts in both the In-Sample and Out-Of-Sample periods. The scenarios obtained by the proposed conditional probabilistic model show a clear advantage over the unconditional reference models, obtaining a functional logarithmic score of 25.82 in the Out-Of-Sample period of the one-step ahead case, a lower value than the best reference model which obtains a fLS of 26.03. In addition, the conditional predictive regions provide a better coverage of the real curves while being narrower than the unconditional regions: the Out-Of-Sample functional Winkler Score obtained by the conditional model is 244645.2, whereas the best reference model has a value of 446792.3. The results are similar for the 24-step ahead case: the logarithmic score of the proposed model is 27.76, whereas the best reference model only achieves a fLS of 28.48. This can be interpreted as conditional scenarios being more coherent with the actual observations of the series. In addition, the proposed model also outperforms the reference models in terms of Winkler Score, obtaining a value of 405488.4 against the unconditional probabilistic model which obtains a fWS_{95} of 815007.2. These values are consistent with Figures 5.16 and 5.19, where unconditional scenarios showed a greater dispersion over the real curves.

Henceforth, it can be concluded that better probabilistic forecasts of the series of aggregated supply curves can be obtained by including relevant information about the electricity market in the formulation of the probabilistic model being used. This information can improve the forecasting capabilities of the SARMAHX model by providing market agents with sharper estimations of the uncertainty of the SARMAHX estimations, providing them with an useful tool to optimize its own short-term bidding strategy against its competitors.

5.3. Forecasting hourly residual demand curves

This section is devoted to the application of short-term forecasts of Residual Demand Curves (RDC) in Day-Ahead electricity markets. As mentioned before, the analysis of RDC is of utmost importance for both market agents, as it allows them to optimize its bidding strategy (Baillo et al., 2004; Campos et al., 2016); and market regulators, as this analysis provides an useful tool to analyze the market power of different generation companies.

This study is aimed at modelling the Residual Demand Curves of the Spanish Day-Ahead electricity market, which provides market agents with a venue to submit their selling and takeover bids for electrical energy for next day's 24 hours. Each day at 12:00 CET an auction takes places which sets the prices and volume of energy being traded for the following 24 hours. The price and energy volume for each hour h is obtained as the intersection of the supply S_h and demand D_h curves, as illustrated in Figure 5.1. This clearing procedure is determined by the European clearing algoritm Euphemia (EUPHEMIA, 2016). Hence, this analysis pursues obtaining accurate short-term forecasts of the Residual Demand Curves (RDC) of the system, defined as $R_h(p) = D_h(p) - S_h(p)$. This is an hourly series where each observation is a whole curve, which can be modelled as an hourly functional time series.

The setup for this case study is as follows. Hourly supply and demand curves for the Spanish Day-Ahead market are obtained by aggregating all the offers that have been submitted by market agents for each hour. The criterion used to select those offers that will be used to define the supply and demand curves is the following: only cleared offers will be used for prices lower than the clearing price of that specific hour, whereas all submitted offers for higher prices are considered. This methodology ensures that the intersection of the supply and demand curves equals the real market price for each hour. Once the supply and demand curves have been obtained, the RDCs can be obtained as the difference between those curves.

Similarly to the previous forecast study, each RDC is modeled as a function $R_t(p)$ which denotes the residual demand of the system at hour t for price p in the range [0,180] €/MWh. The data has been obtained between dates 01/03/2018 and 28/02/2019 from the Iberian

Market Operator OMIE (`www.omie.es`), and has been divided into two sets: the In-Sample period is considered from 01/03/2018 to 31/08/2018, and will be used to identify and fit all the forecasting models included in this study. The remaining data is used to the Out-Of-Sample period, and will be used to compare the generalization capabilities of the models.

The offering behaviour of the agents is often conditioned to the particular market conditions of each day, so explanatory variables will be used to account for those externals factors that influence the trader's decision. The set of covariates using in this study are:

- System demand: The demand is usually an important factor, as it defines the amount of purchased energy that is expected, which usually drives agents to increase its bidding prices.
- Wind production: This renewable energy has a very low generation cost, and it has a relevant presence in the Spanish system: in Spain, wind production accounts for the 20% of the total production.
- Hydraulic production: Hydraulic energy has the ability to regulate and accounts for an important share of the generation in Spain.

As mentioned earlier, the auction for the 24 hours of the next day are cleared simultaneously. Hence, in a real application the agents must have forecasts of the next 24 curves, so two different analysis will be carried out: on the one hand, one-step ahead forecasts will be obtained, as the parameters of the forecasting models are obtained by minimizing the one-step ahead error. This analysis can be useful to obtain insight about the forecasting capabilities of the different models, and to detect any weaknesses that the model may present. On the other hand, a 24-step ahead forecasting study will be included, with the purpose of illustrating the applicability of the proposed techniques to a realistic situation.

The remainder of this section has been divided into three parts. The first part will detail the identification of the deterministic forecasting models, focusing on the application of the FACF and FPACF proposed in this thesis to identify the optimal structure of the SARMAHX model. Secondly, a comparison of the different forecasting models will be performed, in order to evaluate the usefulness of these techniques on a real application. Finally, probabilistic forecasts of the series of RDC will be obtained for the one-step and 24-step ahead forecasts. The proposed conditional probabilistic model will be compared against other reference models found in the functional data literature to analyze its applicability in real scenarios. Finally, probabilistic forecasts are obtained for the one-step and 24-step ahead forecasts using the proposed conditional probabilistic model for functional time series.

5.3.1. SARMAHX model identification

The identification and diagnosis technique proposed in this thesis is used to identify the functional SARMAHX model. Once the regular and seasonal autoregressive and moving average orders of the model have been selected, the model is trained with the In-Sample data of this case study. Both integral and concurrent operators are used in this study. Firstly, the autocorrelation structure of the series of Residual Demand Curves in the Spanish Day-Ahead electricity market is identified using the FACF and FPACF of the series together with the proposed identification procedure shown in Figure 5.2. Figure 5.20 shows the seasonal behaviour of the series: a high autocorrelation on lags 24 and 168 indicates that the series has strong daily and weekly seasonalities. The serial autocorrelation is still present in Figure 5.21, which shows the

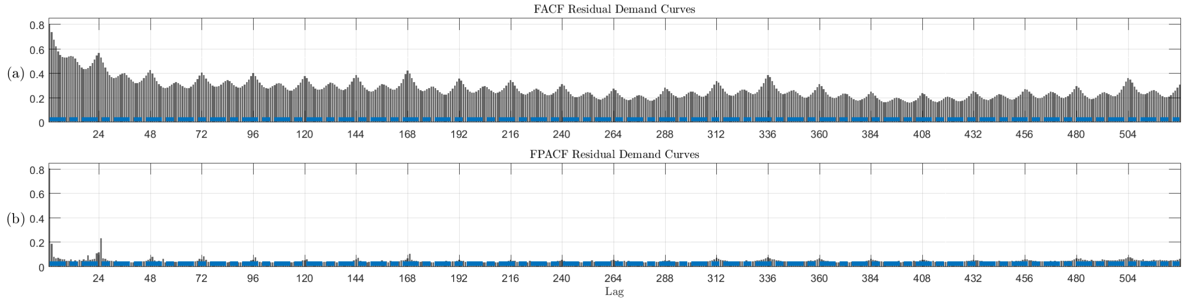

Figure 5.20. First 528 lags of the functional auto- and partial autocorrelation function (FACF / FPACF) of the functional time series of Residual Demand Curves from the Spanish Day-Ahead electricity market.The horizontal line represents the 95% upper limit of the distribution of the statistic under the white noise hypothesis.

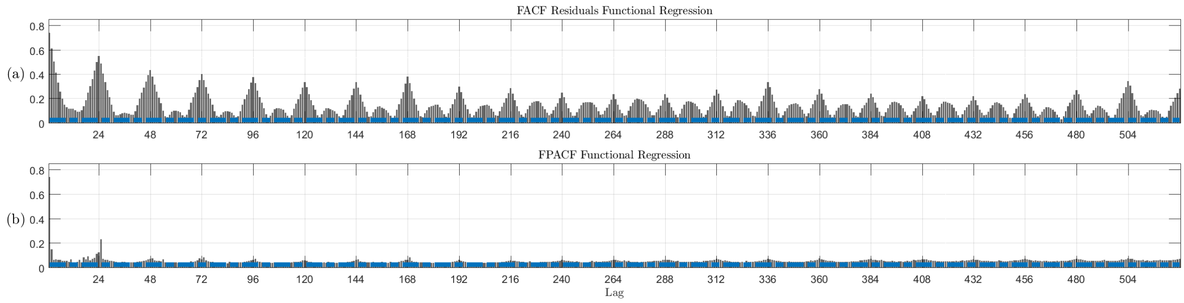

Figure 5.21. First 528 lags of the functional auto- and partial autocorrelation function (FACF / FPACF) of the regression residuals of a fitted functional regression model including the aforementioned exogenous variables. The horizontal line represents the 95% upper limit of the distribution of the statistic under the white noise hypothesis.

autocorrelation structure of the residuals of a fitted functional regression model including the exogenous variables described in the previous section. As the output time series are hourly curves and the explanatory variables are hourly values, the model will consider them as scalar covariates. The strong serial correlation in the regression residuals of the SARMAHX model indicates that the dynamic of the series has to be modelled in order to obtain accurate forecasts of our data.

A SARMAHX$(1,0)$ $\times(1,0)_{24}$ $\times(1,0)_{168}$ model is fitted to model the autoregressive part observed in the residuals of the regression model. Similarly to the previous case study, it has been decided to include autoregressive terms in the model instead of differencing the functional time series, as it has been observed that the performance of the model improved after including these autoregressive terms. Figure 5.22 shows the functional autocorrelation of the residuals of this SARMAHX$(1,0)$ $\times(1,0)_{24}$ $\times(1,0)_{168}$ model. As can be seen, most of the autocorrelation values observed in the previous figures have been captured by the pure autoregressive model. However, significant correlated lags are still found at multiples of 24 and 168 hours of the FACF, whereas the FPACF of the series shows a decaying pattern on the seasonal lags of the series. Hence, there are daily and weekly effects that have not been captured by the explanatory variables and the AR terms that should be modelled. The first lags of both the FACF and FPACF of the series still exhibit a decaying pattern, which will be modeled as a combination of regular autoregressive and moving average terms. Additionally, the FACF of the daily lags shows two significative values at lags 24 and 48, whereas the FPACF of the same lags shows an exponential decrease. This behaviour corresponds to a seasonal moving average effect of order two, which should be added in our final SARMAHX model in order to capture the dynamic

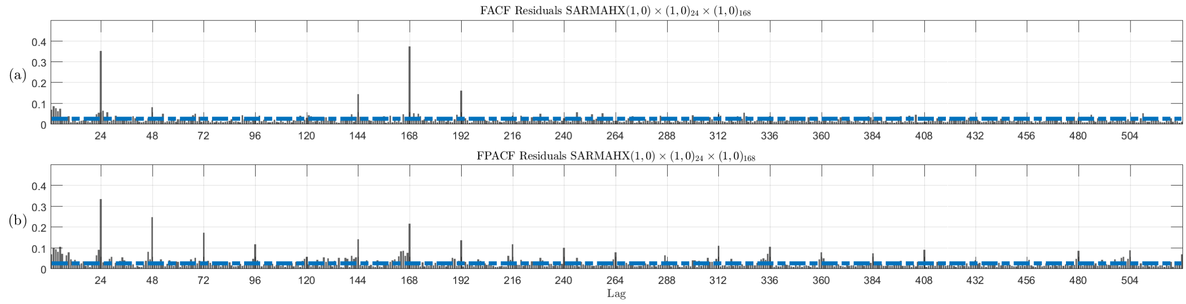

Figure 5.22. First 528 lags of the functional auto- and partial autocorrelation function (FACF / FPACF) of the residuals of the fitted SARMAHX$(1, 0) \times (1, 0)_{24} \times (1, 0)_{168}$ concurrent model including exogenous variables. A high autocorrelation value is found at multiples of 24 and 168 hours. The horizontal line represents the 95% upper limit of the distribution of the statistic under the white noise hypothesis.

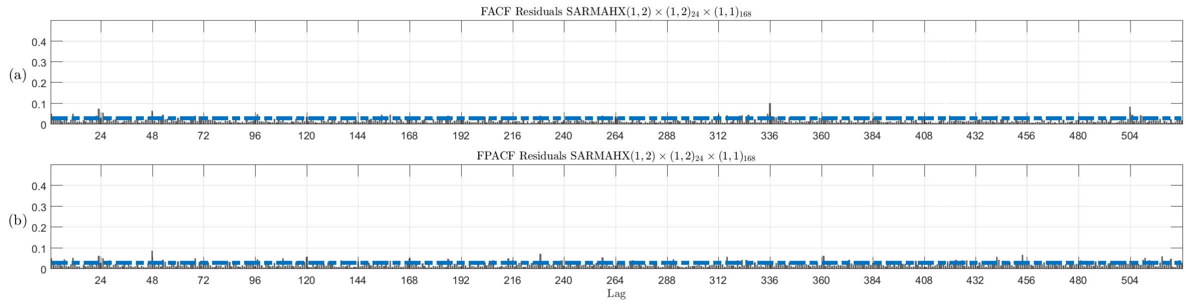

Figure 5.23. First 528 lags of the functional auto- and partial autocorrelation function (FACF / FPACF) of the residuals of the fitted SARMAHX$(1, 2) \times (1, 2)_{24} \times (1, 1)_{168}$ concurrent model including exogenous variables. As there are no significative autocorrelation values for any lags, the residuals can be regarded as a functional white noise process. The horizontal line represents the 95% upper limit of the distribution of the statistic under the white noise hypothesis.

of the functional time series being modelled. A similar pattern can be observed in the weekly lags of the autocorrelation, with only a significant value is observed in the FACF. The decaying pattern in the weekly values of the FPACF indicates that this can be modelled with a seasonal moving average term.

In order to select the optimal autoregressive and moving average orders of the SARMAHX model, the identification methodology presented in Chapter 3 has been used. Regular and seasonal moving average terms have been added into the model to account for the significant correlated lags found in the previous FACF/FPACF residual plot of the previous model. An iterative process has been carried out until the residuals of the fitted model are functional white noise. The final adjusted model for both concurrent and functional cases is a SARMAHX$(1, 2)$ $\times(1, 2)_{24}$ $\times(1, 1)_{168}$ model. Throughout this section, these models will be denoted by SARMAHX (functional case) and SARMAHXconc (concurrent case). Figure 5.6 shows the autocorrelation and partial autocorrelation functions of the residuals of the fitted concurrent SARMAHX model with the 95% upper bound of the statistics under the white noise hypothesis. As most values fall below this upper bound, the hypothesis of the residuals being white noise cannot be rejected, hence validating the model from a linear point of view.

It is worth analyzing the resulting shapes of the fitted functional parameters. Figure 5.24 shows the concurrent regressor operators for the selected SARMAHX model. Analyzing these shapes, the effect of each variable on the residual demand curve can be seen. The demand coefficient only takes positive values, which means that higher demand values increases the average level of the RDCs, as expected due to being more demand that can be covered in

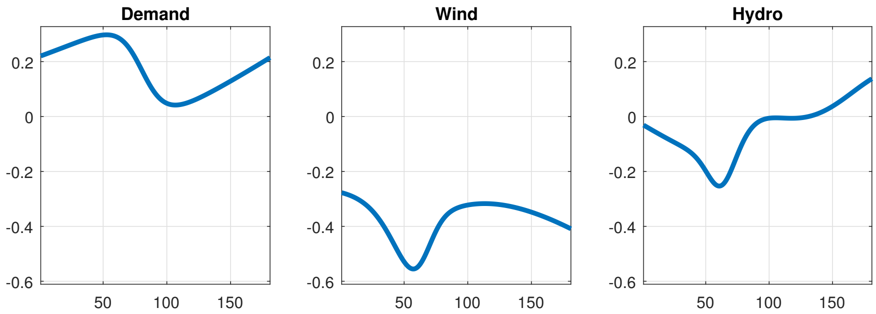

Figure 5.24. Operators' kernels for the regressors of the SARMAHX model in the Residual Demand Curve forecasting study.

the market. The concurrent operator has lower values between prices 50 and 100 €/MWh, meaning that the effect of the demand is weaker on that price range. The wind coefficient takes negative values for all prices, which can be interpreted as RDCs having lower average values with high wind generation. This was expected, as wind is usually offered at prices closer to 0 €/MWh, shifting the values of the curves. Finally, the hydro production coefficient exhibits negative values up to price 100 €/MWh, where the values present a flat shape. Hence, this can be seen as hydro production modifying the shape of the RDCs, moving the first section of the RDC down.

5.3.1.1. Empirical comparison

This section compares the fitted SARMAHX models against other reference functional forecasting models. As mentioned earlier, two different analysis will be performed. On the one hand, a one-step ahead forecast is analyzed, as the models are trained to minimize the one-step ahead forecasting error. However, as the auctions for the 24 hours of the next day are cleared at the same time in the Spanish Day-Ahead market, a 24-step ahead forecast is also included, validating the use of the SARMAHX model in a real case application forecasting Residual Demand Curves.

All the models included in this study have been fitted using data from the In-Sample period. Once their parameters have been estimated, estimations are obtained for both the In-Sample and the Out-Of-Sample period, without re-calibrating the parameters of the models. For the 24-hour ahead forecast, a rolling window approach is used: when forecasting the first hour $h = 1$ of day $D + 1$, the estimation $\widehat{R}_{D+h}$ uses the real curves up to day D. However, when forecasting hours $h > 1$, the real curves have not been observed, so past estimations of the model are used as inputs for the model.

This study will compare the SARMAHX models identified in the previous section against the same set of functional forecasting models considered in the previous case study. This ensures that the performance of the SARMAHX model, a parametric model, is compared with both dimensionality reduction models and non-parametric models. The models included in this comparison are listed below:

- *Naïve*. Simple benchmark model that provides a reference to compare the tested 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. In the second case, for Tuesdays, Wednesdays, Thursdays and

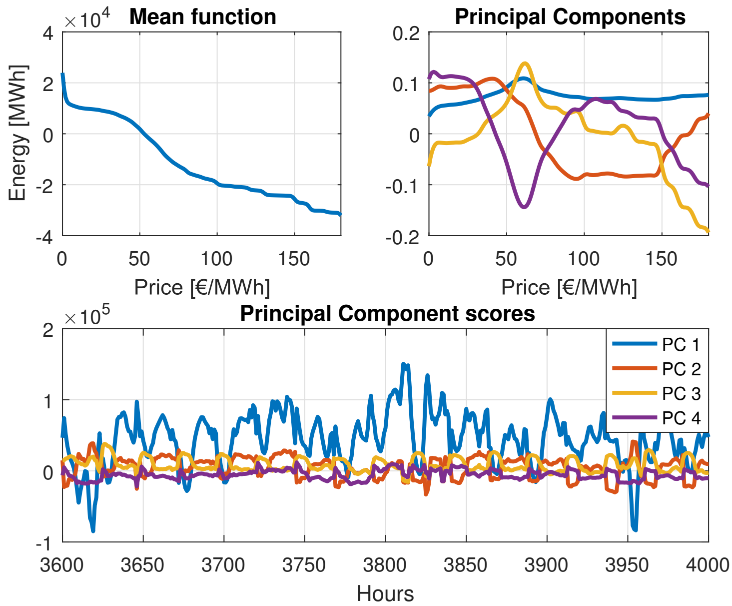

Figure 5.25. Mean function, Principal Components and scores of the Residual Demand Curves.

Fridays, the forecast will be the hourly curve of the previous day, while for Saturdays, Sundays and Mondays, the forecast will be the hourly curve of the previous week.

- *Principal Components approach*. This model aims at decomposing the series of RDC into its first functional Principal Components and the time series of scores, to then forecast each score using a Transfer Function (TF) model (Pankratz, 1991), which are able to include explanatory variables into its formulation. Once the series of scores has been estimated, future forecasted curves can be reconstructed. As with all the models included in this section, the In-Sample period is used to extract the FPC of the RDC and to fit the parameters of the TF models of the scores. Once the models have been fitted, they are evaluated in both the In-Sample and Out-Of-Sample period.

  Figure 5.25 represents 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 96% and 98% of the variance of the data respectively. It is worth noting that each forecasted curve from this method is a linear combination of only the 3 or 4 principal component extracted. These models will be denoted as PC_FT3 and PC_FT4, respectively.

- *Functional nonparametric approach*. This method is used in Aneiros, Vilar, Cao, et al. (2013) for Residual Demand Curve forecasting. The model proposed by the authors is able to include exogenous scalar variables to the classical functional nonparametric autoregressive model (Ferraty, Van Keilegom, et al., 2012) described in Equation 5.5, which leads to more accurate forecasts of the series of RDC. Throughout this section, this model will be denoted as NPARHX.

Similarly to the previous case study, the In-Sample period of the data is used to train all the models, which use the same set of exogenous variables. 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. In the 24-step ahead forecast study, hour 1 will always be estimated

Forecasting horizon	Model	In-Sample FMAE [MWh]	In-Sample FRMSE [MWh]	Out-Of-Sample FMAE [MWh]	Out-Of-Sample FRMSE [MWh]
1-Step Ahead	Naïve	1196.4	1867.9	1439.3	2037.7
	PC_FT3	1085.9	1621.3	1288.7	1868.2
	PC_FT4	1007.2	1591.1	1221.5	1808.1
	NPARHX	1125.1	1798.2	1369.9	1994.1
	SARMAHX	908.7	1374.3	1147.6	1643.7
	SARMAHXconc	**821.8**	**1333.9**	**1028.3**	**1551.2**
24-Step Ahead	Naïve	3029.7	4213.3	2752.2	3585.3
	PC_FT3	2406.8	3301.6	2422.2	3101.2
	PC_FT4	2087.6	2922.1	2371.6	3018.5
	NPARHX	2801.2	4099.1	2692.3	3491.8
	SARMAHX	1972.9	2627.4	2452.1	3151.3
	SARMAHXconc	**1271.9**	**1827.2**	**1922.4**	**2550.3**

Table 5.7. Average errors for each method in the one-step and 24-step ahead prediction in the Residual Demand Curve forecasting study. The lower errors obtained for each forecasting horizon are marked in bold.

using a forecasting horizon of 1; hour 2 will be estimated using a forecasting horizon of 2; and finally the forecasting horizon used to estimate hour 24 will be 24, so the curves will be estimated using only information available up to the previous day. Henceforth, 1-step and 24-step ahead forecasts are produced for the whole data range. Functional errors FMAE and FRMSE, defined in Equations (5.7) and (5.8) respectively, are obtained for both the In-Sample and Out-Of-Sample periods. As the RDCs are monotonically decreasing functions, a variation of the optimization problem presented in Equation 5.9 is used to transform the estimated Residual Demand Curves $\widehat{R}_t$ into monotonically decreasing curves $\widehat{R}_t^-$. This transformation guarantees that the estimated RDCs are monotonically decreasing, while retaining the shape that has been estimated by the functional forecasting models.

Table 5.7 shows the functional errors for the 1-step and the 24-step ahead forecasts for both the In-Sample and Out-Of-Sample periods. The proposed SARMAHX concurrent model shows a clear advantage over the reference models, obtaining a FMAE of 1028.3 MWh, much lower than the FMAE of the PC_FT4 (the best reference model) which only achieves a FMAE of 1221.5 MWh in the Out-Of-Sample period of the 1-step ahead forecasting study. For the 24-step ahead forecasting case, the results are similar: while the PC_FT4 model obtains a FMAE of 2371.6 MWh, it is not able to improve the results of the concurrent SARMAHX model, that achieves a FMAE of 1922.4 MWh in the Out-Of-Sample period. The estimation results are analyzed ahead in detail.

The 1-step ahead average errors of the models show that the proposed concurrent SARMAHX model outperforms the reference models in both the In-Sample and Out-Of-Sample periods. Due to the strong hourly and weekly seasonality of the series of Residual Demand Curves, the nonparametric autoregressive approach is not able to achieve similar results than the other models included in this study. The dimensionality reduction models are able to include up to two seasonal terms, allowing them to model this complex seasonal pattern exhibited by the curves. However, its performance is hindered by the decomposition in Functional Principal Components, as the estimations are not able to fully capture the shape of the RDCs being

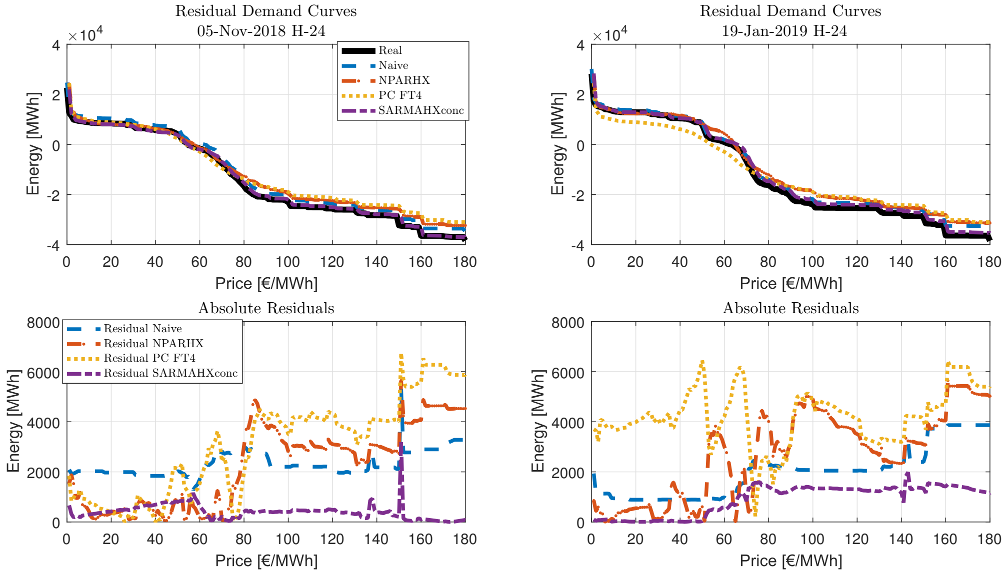

Figure 5.26. Top: Forecast examples for one-step ahead estimations in the Out-Of-Sample periods in the Residual Demand Curve forecasting study. Bottom: Absolute value of the residual curves of the fitted models.

modelled. Interestingly, the concurrent SARMAHX model obtains more accurate estimations than the fully functional version. This usually happens when the series presents low cross-correlation across different curves, which results in diagonal integral operators. As such, due to the lower complexity of the concurrent operators, the adjustment of the parameters is easier for the concurrent model, obtaining more robust forecasts of the series. Figure 5.26 shows some examples of the one-step ahead forecasts of the RDCs, highlighting the accuracy of the proposed SARMAHX concurrent model when compared with the reference models.

The results of the 24-step ahead forecasts are similar to the one-step ahead case. As expected, the global average error for this case is higher than the one-step ahead analysis, due to the higher forecasting horizon. Nevertheless, the SARMAHX estimations are able to obtain more precise estimations than the reference models: while the best reference model has a FMAE of 2371.6 MWh in the Out-Of-Sample period, the concurrent SARMAHX model achieves a FMAE of 1922.4 MWh. It should be noted that the average error of the PC_FT4 model is lower than the fully functional SARMAHX model in the Out-Of-Sample period. This may be caused by a lack of cross-correlation between different sections of the Residual Demand Curves, which hinders the optimization of the integral operators of the model. Figure 5.27 shows some forecasting examples for the 24-step ahead case, which again validates the usefulness of the concurrent SARMAHX model when forecasting RDCs.

Additionally, Tables 5.8 and 5.9 show the Diebold-Mariano (DM) test applied to each pair of forecasts. As mentioned in the previous section, this is a head-to-head method which compares the forecast error of two different models. Each cell of Tables 5.8 and 5.9 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 forecasting errors can be considered statistically different. The DM test validates the Out-Sample results, yielding very low p-values between model comparisons. Then, the equal forecast null hypothesis can be rejected. This analysis emphasizes the statistical difference between the proposed model and the other models considered in this study.

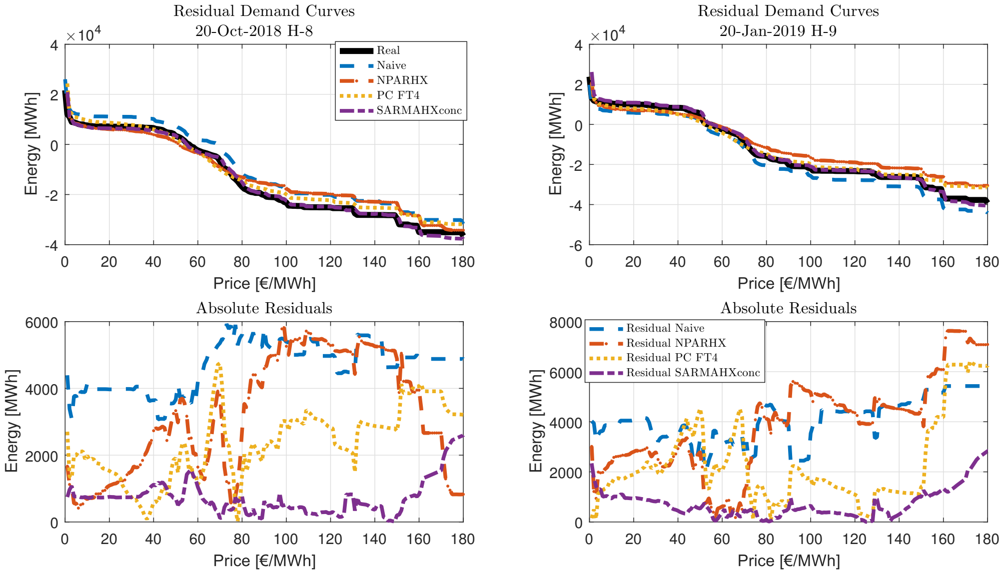

Figure 5.27. Top: Forecast examples for 24-step ahead estimations in the Out-Of-Sample periods in the Residual Demand Curve forecasting study. Bottom: Absolute value of the residual curves of the fitted models.

Models	Naïve	PC_FT3	PC_FT4	NPARHX	SARMAHX	SARMAHXconc
Naïve	-					
PC_FT3	**0**	-				
PC_FT4	**0**	0.1029	-			
NPARHX	**0**	**0**	**0**	-		
SARMAHX	**0**	**0**	**0**	**0**	-	
SARMAHXconc	**0**	**0**	**0**	**0**	**0**	-

Table 5.8. Diebold-Mariano test's p-values for Out-Of-Sample period in the one-step ahead forecasts in the Residual Demand Curve forecasting study. p-values that are lower than 0.05 are marked in bold.

5.3.1.2. In-depth analysis of the results

In order to provide a comprehensive analysis of the forecasting errors of the models, the average absolute error of the different models have been obtained for each hour of the day. As seen in Figure 5.28, the hours with a higher average error for the one-step ahead forecast are the first hour of the day, 8 and 9 and the range of hours between 19 and 21. Market agents usually use similar strategies for the whole day, so when forecasting the first curve of the next day there is usually no past information than can be used to improve the estimation of the RDC. This explains the high error observed on hour 0. The effect of the rolling window approach for estimating the 24-step ahead RDCs can be seen on the right panel of Figure 5.28. The hourly evolution of the absolute errors of the reference models shows a positive trend, which can be explained by the higher forecasting horizons. Interestingly, the trend of the concurrent SARMAHX's errors is not as steep as the other models, which provides an additional measure of the robustness of its predictions.

Models	Naïve	PC_FT3	PC_FT4	NPARHX	SARMAHX	SARMAHXconc
Naïve	-					
PC_FT3	**0**	-				
PC_FT4	**0**	**0.0022**	-			
NPARHX	**0**	**0.0001**	**0.001**	-		
SARMAHX	**0.0008**	**0**	**0**	-		
SARMAHXconc	**0**	**0**	**0**	**0**	**0**	-

Table 5.9. Diebold-Mariano test's p-values for Out-Of-Sample period in the 24-step ahead forecasts in the Residual Demand Curve forecasting study. p-values that are lower than 0.05 are marked in bold

Figure 5.29 shows the mean absolute errors for the one-step and 24-step ahead studies in the Out-Of-Sample period, grouping them by prices. These graphs can be used to identify those price ranges that are more difficult to estimate when forecasting RDCs: as can be seen on both panels the reference models are not able to obtain accurate forecasts of the Residual Demand Curves between prices 60 and 80 €/MWh in both the one-step and 24-step ahead cases. This effect is smoother in the errors of the concurrent SARMAHX model, which presents a flatter level of error across all prices of the curves. These facts make the concurrent model a more reliable forecasting model than its competitors, as it is less frequent to observe sudden spikes in the residuals of the model.

5.3.2. Probabilistic forecasting

The previous section was aimed at obtaining accurate short-term deterministic forecasts of the hourly series of RDCs in the Spanish Day-Ahead electricity market. Figure 5.30 shows the real RDCs together with the 1-step ahead estimated curves, the residuals of the concurrent SARMAHX model and the set of exogenous variables used by the model. However, as mentioned earlier, market agents are usually interested in having probabilistic forecasts of these curves in order to optimize its bidding strategy, hedging against possible risks that the may face in the market. In order to obtain such probabilistic forecasts the uncertainty associated with the deterministic forecasts must be quantified to provide agents with a confidence measure about the deterministic forecasts. Being able to model this uncertainty will allow them to make probabilistic statements about the future evolution of the curves. As the auctions for the 24 hours of the next day are cleared simultaneously in the Spanish Day-Ahead market, a 24-step ahead study will be analyzed in addition to the one-step ahead case, in order to illustrate the real-world applicability of the proposed methodology.

The setup for this study is as follows: all the probabilistic models included in this comparison were fitted using data from the In-Sample period defined in the previous section. Once the internal parameters of the models have been selected, prediction bands for both the In-Sample and the Out-Of-Sample period are obtained, without re-calibrating the models. Functional scenarios of the Residual Demand Curves are also obtained following the methodology detailed in Section 4.2.2, and finally the probabilistic performance metrics reviewed in Section 4.3 are used to compare the applicability of the proposed conditional density model against the remaining reference models considered in this study, which are described ahead:

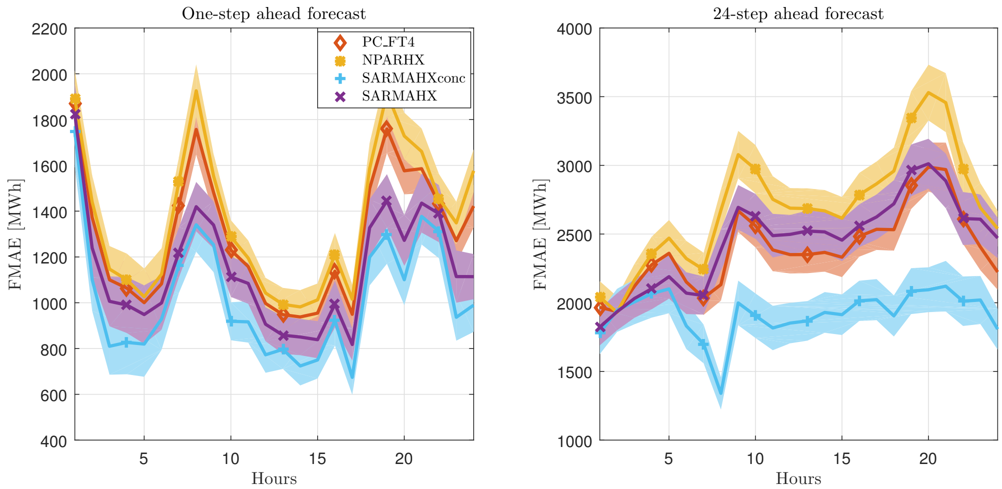

Figure 5.28. FMAE for each hour in the 1-step and 24-step ahead forecasts in the Out-Of-Sample period in the Residual Demand Curve forecasting study. As the FMAE error of the Naive model is significantly higher than the other models, it is not included. The shaded regions are 90% confidence bands for the FMAE.

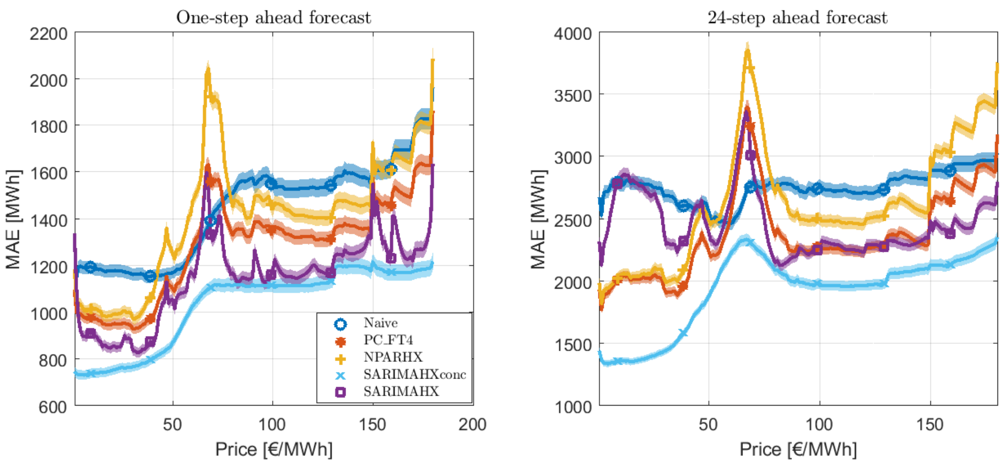

Figure 5.29. Mean absolute error for each price in the 1-step and 24-step ahead forecasts in the Out-Of-Sample period in the Residual Demand Curve forecasting study. The shaded regions are 90% confidence bands for the MAE.

- *Unconditional density estimator.* This model is based on the concept of projection-based density for functional data proposed in Delaigle and Hall (2010). The authors use a standard kernel density estimator to approximate the distribution of the FPC scores of the curves, which can then be used to generate new functional scenarios by sampling from the fitted distribution. The bandwidth parameters of the kernel density estimator is selected using the scores of the FPC decomposition of the SARMAHX's residuals for the In-Sample period. Then, the model assumes that the distribution of the residuals for the Out-Of-Sample period is the same as the In-Sample period, so prediction bands and functional scenarios can be obtained for the whole range of data. This model aims at estimating the unconditional density of a set of functional data, so it is not able to include information about exogenous variables in its formulation.

- *Bootstrap resampling*. The most common technique used in the literature to obtain probabilistic forecasts of functional time series is using bootstrap resampling of the residuals of a fitted deterministic model to quantify the uncertainty of the forecasts (Raña,

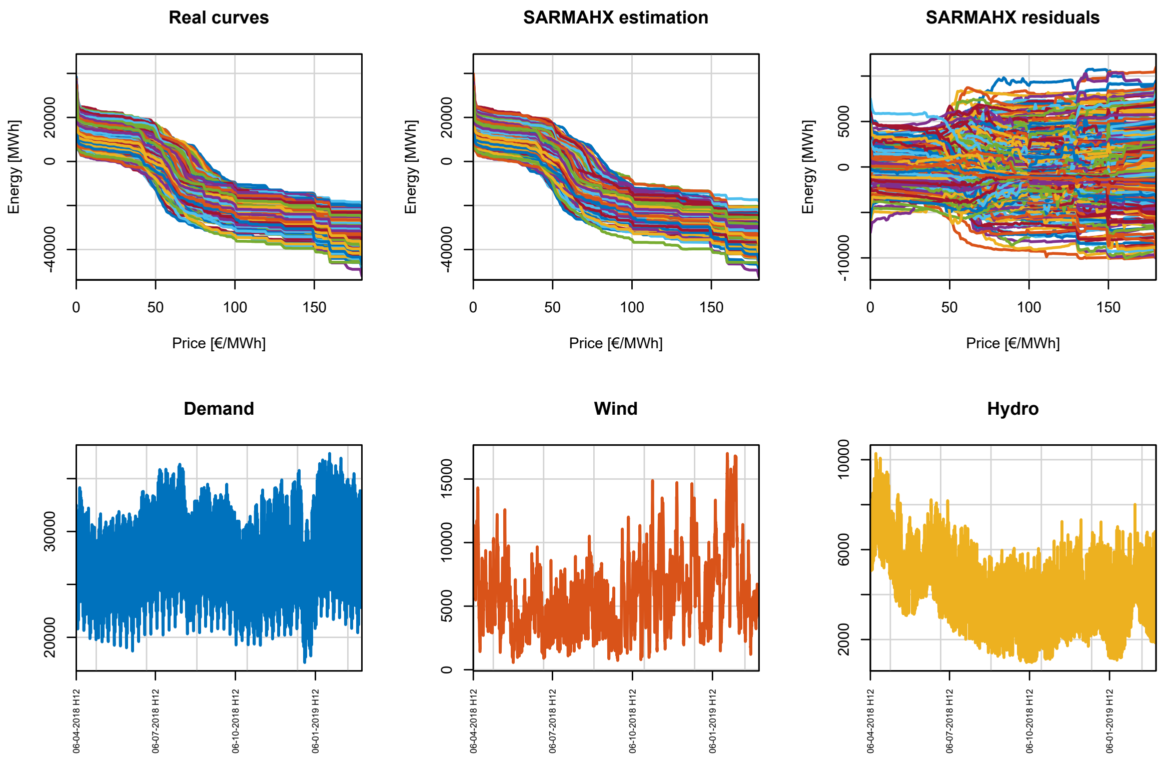

Figure 5.30. Top: actual hourly Complex Residual Demand Curves together with the 1-step ahead forecasts obtained by the fitted SARMAHX model and the residuals of the model. Bottom: Hourly values of the three exogenous variables included in the regression part of the SARMAHX model.

Aneiros, et al., 2016; Vilar, Aneiros, et al., 2018; Rossini and Canale, 2019). This model assumes that the residuals do not exhibit any kind of dependence, so the bootstrap pseudo-replications of the curves are able to approximate the true underlying distribution of the non-deterministic part of the forecasting model. Similarly to the model presented above, the bootstrap model only takes into account the set of functional residuals, without including information about the possible influence of external factors to the distribution of the curves.

Figure 5.28 showed the uneven distribution of the concurrent SARMAHX's errors, with certain hours having a systematically lower error than the remaining hours. This behaviour should be taken into account to obtain sharper probabilistic forecasts of the series. Hence, an exploratory analysis is performed in order to analyze which variables influence the distribution of the one-step ahead residuals. Decomposing the one-step ahead residuals of the concurrent SARMAHX model into their first 4 FPC accounts for 96% of the original dataset's variance, which is adequate to approximate the density of the curves (Delaigle and Hall, 2010). Figure 5.31 shows the distribution of the FPC scores grouping them by different levels of the exogenous variables. The distribution of the residuals is heavily influenced by the hour of the day, with hours 1, 7-10 and 19-24 showing a greater dispersion than the rest of hours. Interestingly, the distribution of the third and fourth FPC associated with hour 1 is not centered, which indicates that the shape of the hour 1 residuals is significantly different that the remaining curves. Variables Demand and Hydro also influence the distribution of the residuals, although not as strongly as the effect seen with the hour of the day. Unlike the reference probabilistic models found in the functional data literature, the conditional probabilistic model proposed in this thesis is able to include the information of the exogenous variables to obtain sharper

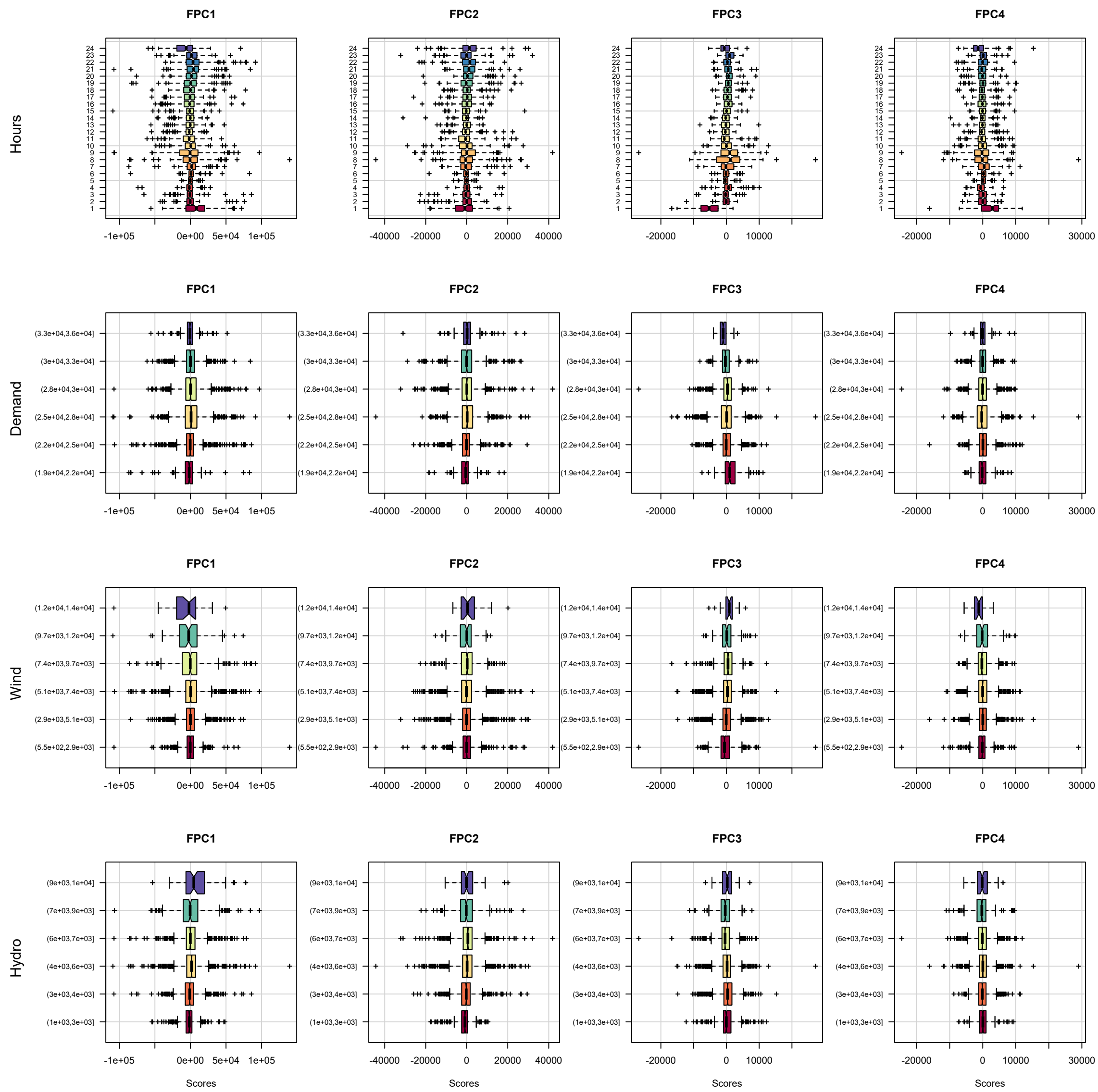

Figure 5.31. Distribution of the FPC scores of the SARMAHX 1-step ahead residuals for the Residual Demand Curve forecasting study, grouped by levels of the exogenous variables. Each column is associated with the scores of a certain FPC, whereas each row is associated with a different covariate.

estimations of the density of the residuals, which may improve the probabilistic forecasts of the RDCs.

In order to perform a robust variable selection for the model, the previous analysis has been compared with the results of performing a two-sample Kolmogorov-Smirnov (KS) test to the distributions of the covariates for each FPC score. The p-values of each variable are displayed in Table 5.10, highlighting those values where the KS test has found a significant difference between the distributions of the residuals' scores. As a result variables Hour, Demand and Hydro will be used as inputs for the one-step ahead conditional probabilistic model. Figure 5.32 illustrates the differences in the distribution of the one-step ahead functional residuals for different values of the input variables. The dispersion exhibited by the model's residuals is very high for hour 1 and a combination of high Demand and Hydro, as seen on the left panel. The right panel shows a very different situation, where the residuals selected by the conditional

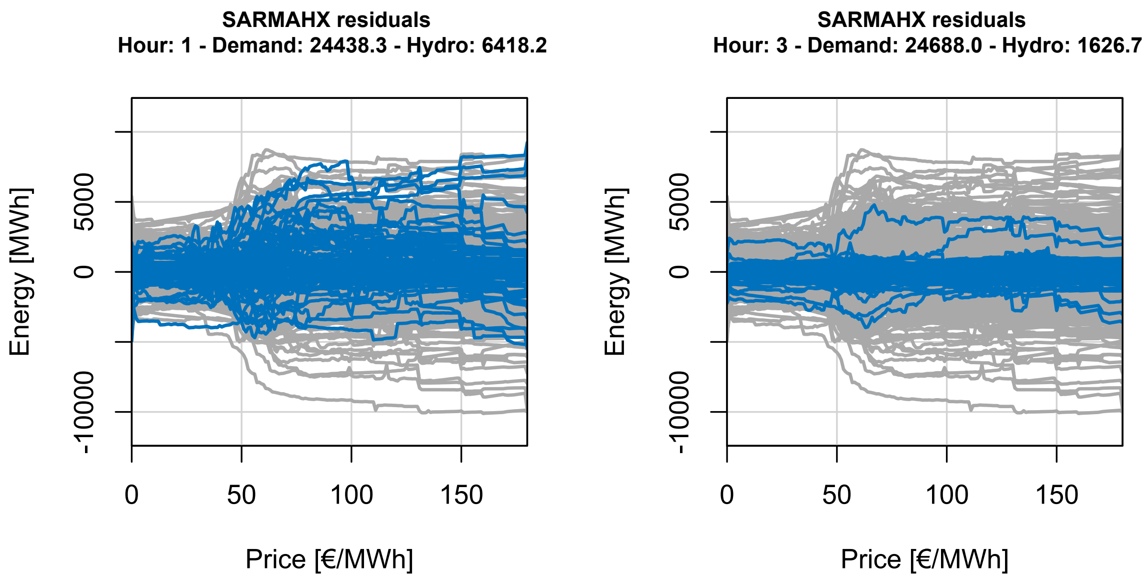

Figure 5.32. Distribution of the residuals of the model for two different values of the exogenous variables Hour, Demand and Hydro for the 1-step ahead forecasts of the Residual Demand Curves. The whole set of residuals is displayed using grey color, whereas the curves with a positive weight in the probabilistic model are shown in blue.

	FPC1	FPC2	FPC3	FPC4
Hour	**0**	**0**	**0**	**0.00002**
Demand	**0**	**0.00172**	**0.00065**	**0**
Wind	**0.00124**	0.01205	**0.00037**	0.0524
Hydro	**0**	**0**	**0.00261**	**0.0042**

Table 5.10. p-values of the Kolmogorov-Smirnov test obtained for the distributions of the one-step ahead residuals' scores grouped by levels of the exogenous variables used in the Residual Demand Curve forecasting study. p-values lower than 0.05 are marked in bold.

model for hour 3 and low levels of Hydro production are plotted. The dispersion of the residuals is very low, showing less variance than the previous case. This highlights the capability of the proposed model to capture differences in the distribution of the residuals given a set of exogenous covariates.

Once the relevant variables for estimating the conditional distribution of the residuals have been identified, the parameters of the models must be estimated from the observed data. The leave-one-out cross-validation procedure detailed in Section 4.2.1 is used to select the bandwidth parameters of the conditional model using the In-Sample period of the data. As one of the inputs for the model is variable Hour, the circular metric $\|.\|_{S^1}$ introduced in Equation 4.36 is used to capture the notion of proximity between hours in the model. Once the model has been fitted, 80% and 95% prediction regions are obtained from the estimated densities for both the In-Sample and Out-Of-Sample periods. Figure 5.33 compares the prediction regions obtained by the reference models against the conditional model are shown, together with a set of 100 functional scenarios of the RDCs. As evidenced by this comparison, the proposed conditional probabilistic model is able to outperform the reference models by capturing the variation in the uncertainty distribution of the forecasts conditioned to the input variables. This techniques obtains narrower prediction regions for those periods where the historical volatility of the

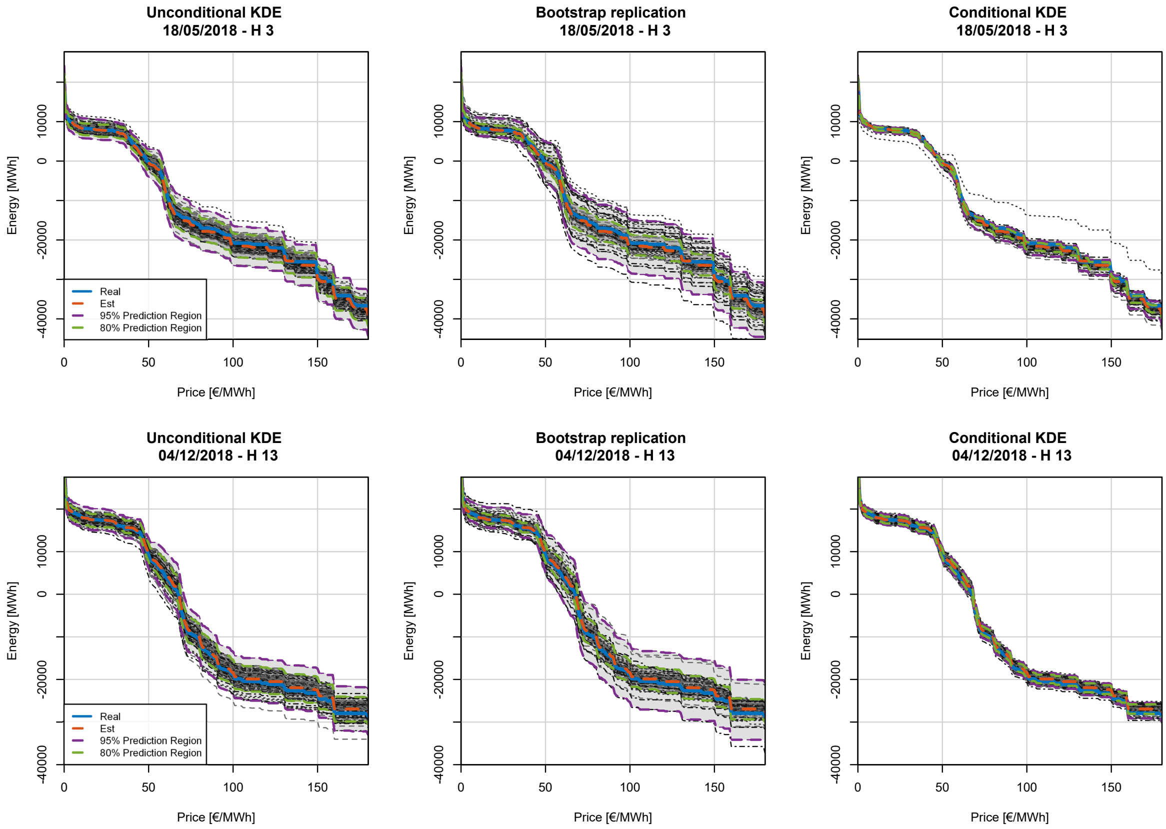

Figure 5.33. Prediction regions defined by the residual-based bootstrap, unconditional and conditional models fitted to the RDC series for the 1-step ahead forecasting problem. The black curves represent the different scenarios obtained by each method. The top row corresponds to the In-Sample period, whereas the bottom figures display curves for the out-of-sample period.

residuals has been low, obtaining coherent one-step ahead RDC scenarios. As the reference models assume that the residuals are distributed independently, they often overestimate the density of the residuals, which results in wider prediction regions.

As mentioned before, in a real application the RDCs for the next 24 hours must be forecasted at the same time, as the auctions for each hour of the next day are cleared at the same instant. Hence, market agents need to have probabilistic forecasts of the series for those 24 hours before the auction takes place in order to optimize its bidding strategy. As such, an additional study will be performed in order to analyze the applicability of the proposed model to this real situation. The first step in this study is to identify any possible relation between the projection-based density of the 24-step ahead residuals of the concurrent SARMAHX model and the set of explanatory variables that have been used throughout this section. This analysis is aimed at obtaining a set of coherent 24-step ahead functional scenarios of the RDCs.

Following a similar procedure to the one used in the one-step ahead case, the 24-step ahead SARMAHX residuals must be decomposed into the space spanned by its first FPC in order to estimate the projection-based density of the curves. In this setup, more than 98% of the variance of the 24-step ahead residuals is explained by the first 4 FPC. Once the residuals have been decomposed, an exploratory analysis is performed to identify those variables that influence the distribution of the residuals' scores the most. Figure 5.34 shows the distribution of the FPC scores grouped by different levels of the covariates. Unlike the one-step ahead case the distribution of the 24-step ahead residuals is not influenced by the hour of the day, with all

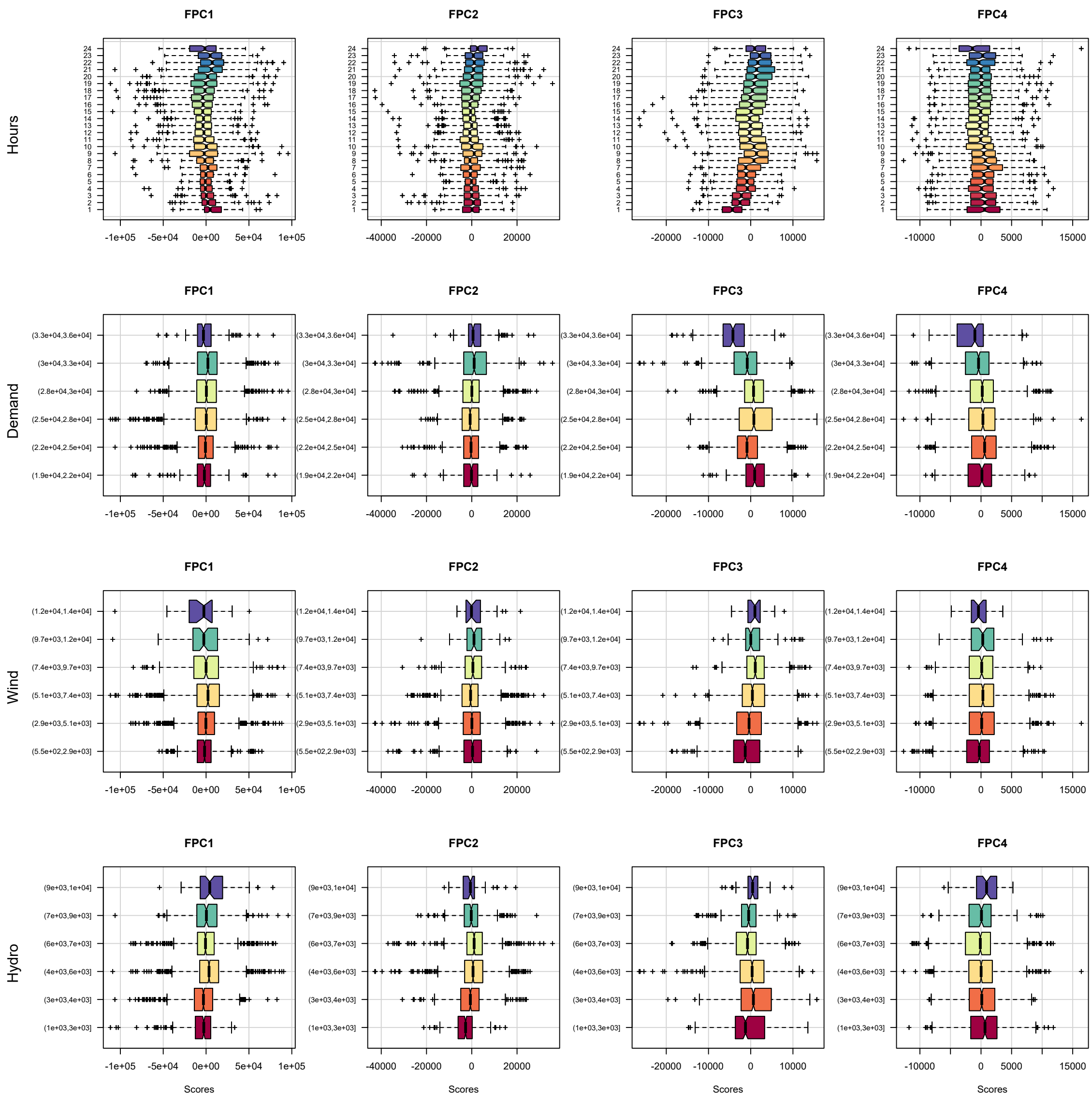

Figure 5.34. Distribution of the FPC scores of the SARMAHX 24-step ahead residuals of the Residual Demand Curves forecasting study, grouped by levels of the exogenous variables. Each column is associated with the scores of a certain FPC, whereas each row is associated with a different covariate.

hours having a similar distribution. This agrees with the results observed in the right panel of Figure 5.29. However, the boxplots shown in Figure 5.34 reveal that the distribution of the residuals is heavily influenced by the Demand, Wind and Hydro generation. As can be seen, extreme values of the demand exhibit a lower dispersion than the intermediate values. In addition, the distribution of the third FPC's scores are not centered, which implies that the Demand influences the shape of the residuals. Finally, it can be seen that the variance of the scores of the first FPC increases with the level of the Wind and Hydro generation. This analysis is coherent with the results of the Kolmogorov-Smirnov test, whose p-values are shown in Table 5.11. Most of the p-values obtained are 0, which means that the test has found strong evidences in the data to reject the null hypothesis of equal distributions across different levels of the exogenous variables.

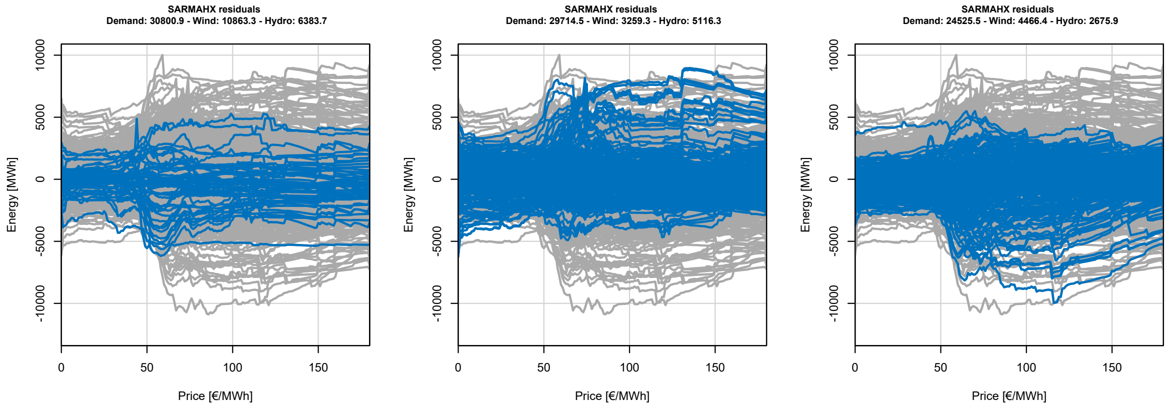

Figure 5.35. Distribution of the residuals of the model for three different set of values of the exogenous variables Demand, Wind and Hydro for the 24-step ahead forecasts of the Residual Demand Curves. The whole set of residuals is displayed using grey color, whereas the curves with a positive weight in the probabilistic model are shown in blue.

	FPC1	FPC2	FPC3	FPC4
Hour	**0**	0.05162	**0**	0.0658
Demand	**0.00004**	**0**	**0**	**0.0002**
Wind	**0.00084**	**0.00202**	**0**	**0.0072**
Hydro	**0**	**0**	**0**	**0.00279**

Table 5.11. p-values of the Kolmogorov-Smirnov test obtained for the distributions of the 24-step ahead residuals' scores grouped by levels of the exogenous variables used in the Residual Demand Curve forecasting study. p-values lower than 0.05 are marked in bold.

The differences in the distribution of the 24-step ahead residuals are illustrated in Figure 5.35, where the set of historical residuals selected by the fitted conditional probabilistic model for different sets of covariates is shown. The first panel shows the set of historic residual curves for high Demand and extreme values of Hydro and Wind generation. These curves exhibit a low level of dispersion, identifying periods of low uncertainty in the SARMAHX's forecasts. The second panel shows a set of historical residuals that present a high level of volatility with a skew towards positive values, associated with high demand and low values of Hydro and Wind production. Finally, the third set of residuals exhibits a negative skew which corresponds to a situation that combines low values of Demand, Hydro and Wind.

The previous analysis reinforces the hypothesis that exogenous variables are able to explain the variation in the distribution of the 24-step ahead residuals. Hence, once the proposed conditional model is fitted with the In-Sample data, the model is evaluated to obtain estimation of the 80% and 95% prediction regions of the curves for the whole period of dates, together with a set of 100 functional scenarios that follow the estimated conditional distribution. Figure 5.36 compares the results obtained by the reference models and the conditional probabilistic mode for the In-Sample and Out-Of-Sample periods of the data. As can be seen, both the unconditional and the wild bootstrap models are unable to capture the variations in the uncertainty associated with the 24-step ahead forecasts obtained by the concurrent SARMAHX model, as indicated by the wide prediction regions obtained in both the In-Sample and the Out-Of-Sample periods. This also illustrates the improvements made by the proposed model over the reference methods, as the conditional distribution obtained by the proposed model is able to capture better the differences in the distribution of the residuals, which results in narrower prediction regions in low volatility periods of the series, and wider prediction regions where the uncertainty of the model increases.

In order to provide a fair comparison of the performance of the models, Table 5.12 shows the values of the probabilistic scores defined in Section 4.3 for the one-step and 24-step ahead forecasts in both the In-Sample and Out-Of-Sample periods. The scenarios obtained by the proposed conditional probabilistic model show a clear advantage over the unconditional reference models, obtaining a functional logarithmic score of 29.21 in the Out-Of-Sample period of the one-step ahead case, a lower value than the best reference model which obtains a fLS of 30.17. In addition, the conditional predictive regions provide a better coverage of the real curves while being narrower than the unconditional regions: the Out-Of-Sample functional Winkler Score obtained by the conditional model is 804632.1, nearly half the value obtained by the best reference model that only achieved a Winkler Score of 1587586.8. The results are similar for the 24-step ahead case: the logarithmic score of the proposed model is 31.36, whereas the best reference model has a fLS of 31.58. This can be interpreted as conditional scenarios being more coherent with the actual observations of the series. In addition, the proposed model also outperforms the reference models in terms of Winkler Score, obtaining a value of 1265318 against the unconditional probabilistic model which obtains a fWS_{95} of 2010097. These values are consistent with Figures 5.33 and 5.36, where unconditional scenarios showed a greater dispersion over the real curves.

As mentioned when introducing the concept of RDCs there are a great number of applications of the short-term functional scenarios obtained across energy markets. RDCs defined in this section model the offering and bidding behavior of market agents in the Spanish Day-Ahead electricity market. Hence, a straight-forward application of these curves is to obtain short-term probabilistic estimations of the market clearing price, as it is defined as the value $R_t(0)$. Hence, the intersection of each RDC scenario $\widehat{R}_t^{(i)}(0)$ defines a plausible clearing price, thus the set of 100 scenarios can be used to define a probability density of the market clearing price. Figure 5.37 illustrates this concept, by comparing the density estimation made by the 3 functional probabilistic models considered in this study, for the Out-Of-Sample period and the 24-step ahead forecasts of the Residual Demand Curves. The first and second figures show how the conditional scenarios induce a sharper price distribution, which usually gives more likelihood to the actual clearing price for that hour. The third panel shows a rather interesting situation, where the conditional probabilistic model has identified a bimodal distribution of the prices using only past historical data. As can be seen, the actual clearing price falls inside one of the components of the bimodal distribution, whereas the unconditional models are unable to capture that relationship, thus estimating a unimodal distribution and not providing a good estimation of the price density.

In order to validate this probabilistic analysis of the clearing prices, the price density estimated by all the models in both the one-step ahead and 24-step ahead forecasting study has been obtained for all of the In-Sample and Out-Of-Sample periods. Then, the likelihood that each distribution gives to the actual clearing price has been quantified using a probabilistic score often used in density forecasting, the average logarithmic score (Gneiting and Raftery, 2007). This metric is defined as the negative value of the arithmetic mean of the log-likelihood of the real data given the estimated density:

$$\text{logScore} = -\frac{1}{N}\sum_{i=1}^{N} \log\left(\widehat{P}_i(y_i)\right), \tag{5.10}$$

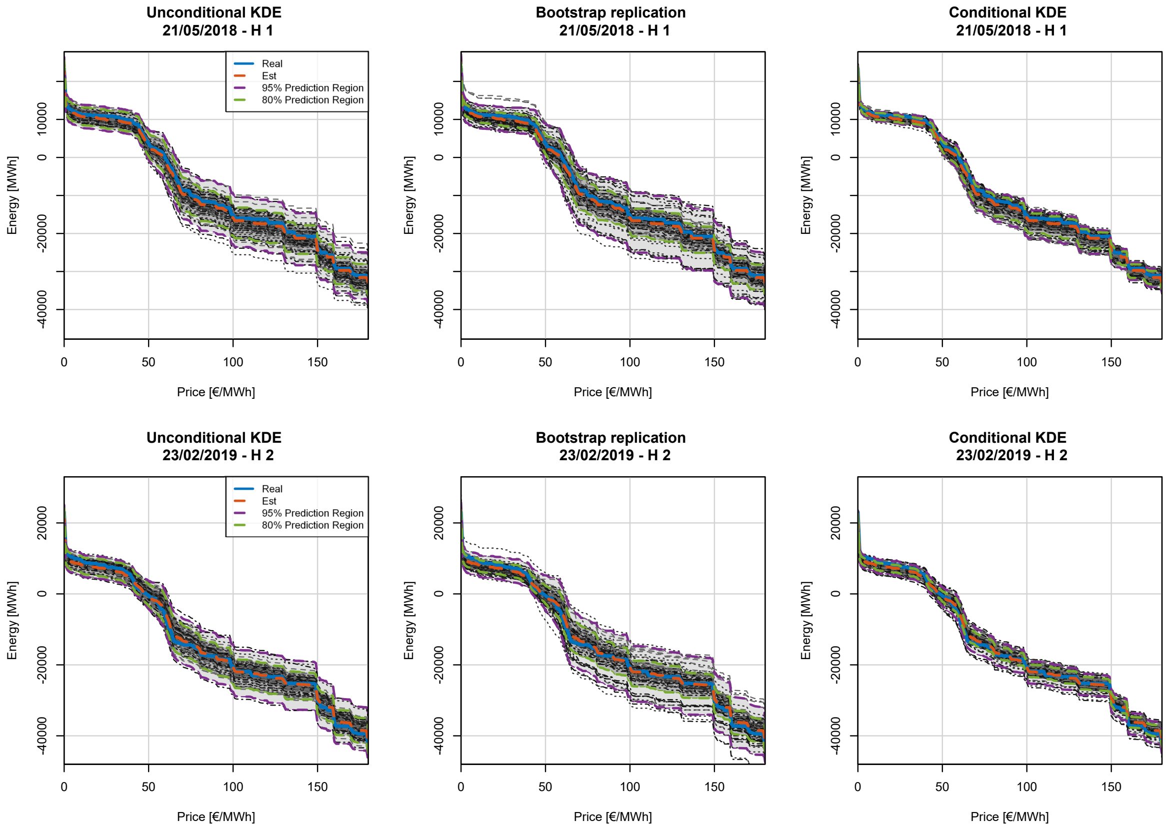

Figure 5.36. Prediction regions defined by the residual-based bootstrap, unconditional and conditional models fitted to the RDC series for the 24-step ahead forecasting problem. The black curves represent the different scenarios obtained by each method. The top row corresponds to the In-Sample period, whereas the bottom figures display curves for the Out-Of-Sample period.

Forecasting horizon	Model	In-Sample fCRPS	In-Sample fWS_{95}	In-Sample fLS	Out-Of-Sample fCRPS	Out-Of-Sample fWS_{95}	Out-Of-Sample fLS
1-Step Ahead	Bootstrap	64958.87	1581943.3	30.03	77746.03	1587586.8	30.33
	Unconditional	64917.66	2044942.1	30.01	77283.63	2049195.7	30.17
	Conditional	**56294.31**	**878175.5**	**28.75**	**77053.13**	**804632.1**	**29.21**
24-Step Ahead	Bootstrap	93621.98	1993667	30.94	141471.1	2010097	31.58
	Unconditional	96014.91	2301784	31.29	**139535.3**	2313833	31.77
	Conditional	**83001.28**	**1239106**	**30.28**	140459.5	**1265318**	**31.36**

Table 5.12. Average scores for each probabilistic model in the 1-step and 24-step ahead Residual Demand Curve scenario generation study. The lower scores obtained for each metric are marked in bold.

where $\widehat{P}_i$ denotes the estimated density and y_i denotes the actual observations. As with all proper probabilistic scores, a lower value of the logarithmic score implies that the fitted density is better than the other estimations included in the comparison. The results are shown in Table 5.13. As expected, the average log-scores are much lower in the one-step ahead case, as the uncertainty of the forecasts is lower when compared with the 24-step ahead forecasts. Additionally, the average log-score is higher in the Out-Of-Sample period, as it measures the estimations of the market clearing prices' density over six months, without any recalibration of the models whatsoever. Nevertheless, the density estimated by the conditional model consistently outperforms the reference models in both the In-Sample and Out-Of-Sample periods, obtaining sharper density estimations that the remaining models. In the 24-step ahead study, the conditional model has

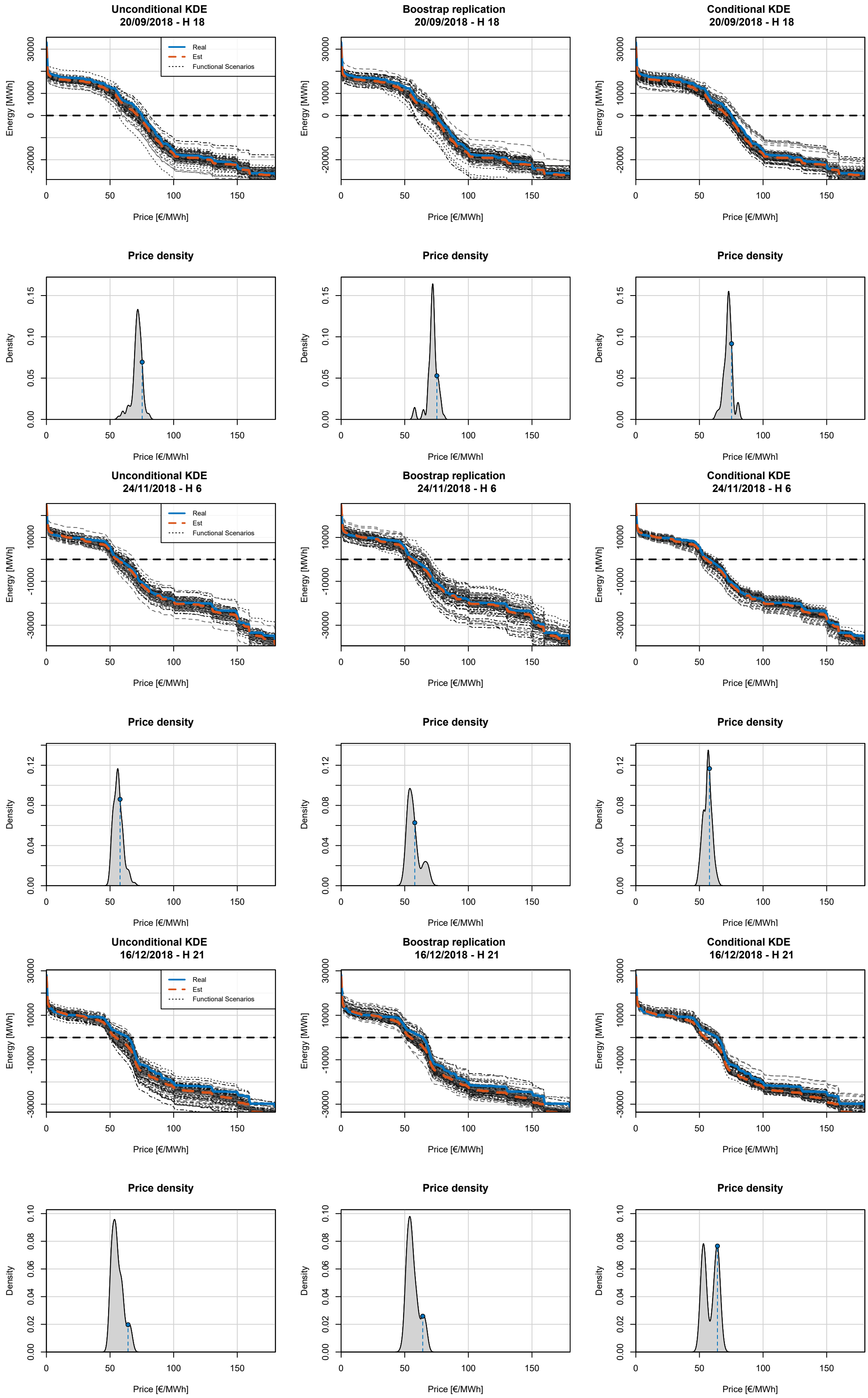

Figure 5.37. Price density induced by the 24-step ahead RDC scenarios obtained using the bootstrap, unconditional and conditional probabilistic models for three different hours. The estimated probability density is shown in grey, whereas the likelihood of the real clearing price is shown in blue.

Model	1-step ahead In-Sample	1-step ahead Out-Of-Sample	24-step ahead In-Sample	24-step ahead Out-Of-Sample
Unconditional	1.8323	2.1042	2.3438	2.7503
Bootstrap	1.8483	2.1618	2.4753	2.8392
Conditional	**1.5501**	**1.8507**	**2.1774**	**2.6449**

Table 5.13. Average log-scores for both the one-step and 24-step ahead density forecasts of the market clearing price in the In-Sample and Out-Of-Sample period of the study. Lower log-scores are marked in bold.

obtained a mean logarithmic score of 2.64 in the Out-Of-Sample period, against a value of 2.75 obtained by the unconditional model's density estimations and a value of 2.84 given by the bootstrap scenarios.

The analysis presented in this section has illustrated how the proposed methodology for functional time series is able to obtain short-term probabilistic forecasts of Residual Demand Curves which outperform the current reference models found in the functional data literature. The proposed conditional model has been able to obtain sharper probabilistic forecasts of the curves by including information about exogenous variables that represent different regimes in the electricity market. Additionally, an application to estimating the price density of the market clearing price has been presented, highlighting the real-world applicability of the forecasting methodology presented in this dissertation.

5.4. Conclusions

This chapter has been devoted to the empirical application of functional probabilistic models to the estimation of Residual Demand Curves (RDC) in electricity markets. If the demand is assumed to be inelastic, this problem reduces to the estimation of the series of Supply Curves, as the inelastic demand can be estimated using other non-functional methods. This situation has been illustrated with a case study using data from the Italian Day-Ahead market. An additional analysis has been performed without making any prior assumptions about the demand curve in the market, where the series of Residual Demand Curves from the system have been modelled. This study uses data from the Spanish Day-Ahead electricity market to define the series of RDCs and the sets of exogenous variables that will be used to obtain short-term estimations of the curves. The high dimensionality of the series of Supply Curves and Residual Demand Curves provide a suitable framework to compare different functional forecasting models.

Both of the functional time series analyzed in this section are hourly series, thus they will present both a daily and weekly seasonality. Hence, in order to obtain accurate estimations of these series, this information must be included in the forecasting models that will be used to obtain estimations of the curves. The current implementation of the SARMAHX model is able to include up to two seasonal terms in its formulation (but it could be extended to any number of seasonal orders), which has proven to be of utmost importance to obtain accurate estimation of the curves in both studies. This model has outperformed the other reference deterministic models in both the one-step and 24-step ahead cases. The identification and diagnosis methodology based on the FACF and FPACF proposed in this thesis has been used to select the optimal parameters of the SARMAHX model when training the model, enhancing the forecasting capabilities of the model.

Once the deterministic forecasting models have been evaluated, probabilistic forecasts of the functional time series have been obtained by quantifying the uncertainty of the concurrent SARMAHX model (as this was the best performing model for both the one-step and 24-step ahead forecasting study). An exploratory analysis has revealed that the SARMAHX residuals were not independently distributed, as there were periods that exhibited a high level of volatility, possible due to sudden changes in the bidding behaviour of market agents. The conditional probabilistic model proposed in this thesis has proven itself to be of much use in these applications, as it has been able to capture the relation between these high-volatility periods and the set of exogenous variables from the electricity market. This model was able to obtain sharper estimations of the density of the residuals, outperforming the reference functional probabilistic models in both the one-step and 24-step ahead forecasts of the series. An application of the probabilistic estimations of the series of RDCs has been included, which illustrates how the techniques developed in this dissertation can be used to obtain density estimations of the market clearing prices. This application highlights the applicability of the proposed forecasting methodology to electricity markets, providing market agents with an useful set of tools to model the bidding behaviour of its competitors.

6

Conclusions, contributions and future work

This last chapter, summarizes the contributions of this thesis. 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.

6.1. Summary and conclusions

Many applications in electricity markets involve the analysis and modelization of functional time series. An illustrative example of functional time series in electricity markets are Supply and Demand curves, obtained by aggregating the bidding information of the different market auctions. These curves contain relevant information related to the strategic bidding behaviour of market agents, making the analysis of Supply and Demand curves of utmost importance to electricity companies. The flexibility of functional forecasting models makes them suitable for this task, which makes FDA a competitive framework to model these high-dimensional data.

In many electricity markets there is one auction for each hour of the day, hence obtaining an hourly series of Supply and Demand curves. Consequently, these series exhibit both a daily and weekly seasonal dependency. Being able to identify the seasonality of a functional time series, as well as the dependence structure between its elements, is of utmost importance when fitting a functional time series model to the curves. Henceforth, the FACF and FPACF proposed in this dissertation are useful tools that allow the practitioner to capture this information for a given functional time series. This methodology improves the current state of the art in the selection and diagnosis of parametric functional forecasting models, as it is able to identify the presence of serial correlation in a functional time series, in addition to quantifying the type of dependence (auto-regressive or moving average). This information can then be used to select the order and assessing the adequacy of functional SARMAHX models by testing the white noise assumption on the residuals of the fitted model, improving the forecasting capability and robustness of these models.

Once a deterministic model has been selected and fitted to a given functional time series future values of the series can be estimated. Although point-forecasts may be adequate in some cases, the vast majority of real-world applications require having probabilistic forecasts of future events. In order to obtain probabilistic forecasts of the fitted deterministic model, such as the aforementioned SARMAHX model, it is necessary to quantify the uncertainty associated with its forecasts by analyzing the distribution of the models' residuals. The vast majority of probabilistic models proposed in the functional data literature assume that these residuals are independent and are identically distributed, which results in similar predictive regions for the whole range of curves. However, this is a strong assumption when dealing with data obtained from electricity markets, which is often influenced by the system's demand, business effects and other external factors. The bidding behaviour of market agents is also sensitive to these indicators, so in order to obtain sharp probabilistic forecasts of the series this information must be taken into account. The functional probabilistic model developed in this thesis aims at estimating the conditional projection-based density of the residuals of a given functional forecasting model, using a set of exogenous variables as inputs that influence the distribution of the residuals. The bandwidth parameters of the model are selected using a leave-one-out cross-validation procedure which ensures that the final model is not overfitted. The conditional distribution estimated by the model can then be used to define prediction regions that characterize the uncertainty of the residuals. Additionally, coherent scenarios of the functional time series can be obtained, which outperform the unconditional scenarios obtained with the reference models proposed in the literature. These scenarios could be used as inputs for several stochastic optimizations models, which would benefit from the scenario generation model proposed in this thesis.

The different developments presented in this dissertation can be used to define a forecasting methodology for functional time series, which involves the identification and diagnosis of a functional SARMAHX model and the generation of coherent scenarios obtained from the conditional density of the SARMAHX's residuals. The performance of this methodology has been validated with two case studies involving different electricity markets. Firstly, an application to the short-term forecasting of Supply Curves in the Italian Day-Ahead electricity market is presented. As this is an hourly functional time series the forecasting models considered in this study include two seasonal terms to model the daily and weekly dependency of the data. The concurrent SARMAHX model outperforms the reference models, obtaining a lower error in both the one-step and the 24-step ahead cases. The proposed conditional probabilistic model is fitted to the residuals of the one-step and 24-step ahead forecasts, using as inputs a set of exogenous variable that are able to model the conditional distribution of the residuals. The resulting prediction regions and functional scenarios provide better estimations than the ones obtained by the reference probabilistic models, which are not able to distinguish between periods of high and low volatility in the curves.

Secondly, the short-term forecasting of Residual Demand Curves of the Spanish Day-Ahead electricity market is addressed. These are functions that model the market-clearing price as a function of the energy that agents are willing to trade. Similarly to the previous case study, the identification and diagnosis methodology is used to select the best parameters for a SARMAHX model, whose estimations are compared against other forecasting models often used in the functional data literature. The best model for both the one-step and 24-step ahead study is again the concurrent SARMAHX model. Once evaluated, probabilistic forecasts of the series are obtained from the proposed conditional model, whose estimations outperform the reference probabilistic models included in the comparison. The 24-step ahead scenarios are used to obtain density estimations of the market clearing price, illustrating the applicability of the proposed forecasting methodology.

6.2. Original contributions

The techniques developed in this thesis have yielded several original contributions to the current state-of-the-art in terms of probabilistic forecasting of functional time series and its application to electricity markets. The main contributions of this dissertations are described below:

- Definition of a functional version of the autocorrelation and partial autocorrelation functions for functional time series based on the L^2 norms of the lagged autocovariance operators of the series. Prediction bounds for both the FACF and FPACF under the hypothesis of functional white noise have been obtained, which can be used as a fast and efficient tool to analyze the correlation structure of a given functional time series.
- Development of an identification and diagnosis methodology for the SARMAHX model. This methodology uses the FACF and FPACF proposed in this thesis and can be used for selecting the order and assessing the adequacy of the functional SARMAHX model. This tool is of utmost importance when fitting this model to a functional time series, as selecting the optimal configuration of parameters for the model allows the practitioner to obtain accurate forecasts of the curves.
- Development of a conditional probabilistic model for functional time series that aims at modelling the projection-based conditional distribution of the residuals of a fitted functional forecasting model. A leave-one-out cross-validation methodology has been developed for selecting the bandwidth parameters of the proposed model. Once fitted, the model can be used to obtain prediction regions of the series and to generate coherent functional scenarios of the curves.
- Proposal of a novel probabilistic metric for quantifying the likelihood of a real curve given a set of functional scenarios. This metric, based on the classical logarithmic score, can be used to asses the quality of different scenario generation methods for functional time series, providing an useful tool in model selection.
- Successful application of the proposed probabilistic forecasting methodology to modelling the hourly series of Residual Demand Curves in several electricity markets, obtaining one-step and 24-step ahead scenarios of the series. These case studies have validated the proposed functional forecasting methodology against other well-known reference models often found in the functional data literature.

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

- G. Mestre, J. Portela, A. Muñoz, E. Alonso. *Forecasting hourly supply curves in the Italian day-ahead electricity market with a double-seasonal SARMAHX model*. International Journal of Electrical Power & Energy Systems, 121, 106083. April 2020. DOI: `https://doi.org/10.1016/j.ijepes.2020.106083`

- G. Mestre, J. Portela, G. Rice, A. Muñoz, E. Alonso. *Functional time series model identification and diagnosis by means of auto- and partial autocorrelation analysis*. Computational Statistics & Data Analysis, 155, 107108. March 2021. DOI: `https://doi.org/10.1016/j.csda.2020.107108`

Additionally, at the time of writing, another two papers are under review:

- G. Mestre, J. Portela, A. Muñoz, E. Alonso. *Scenario-Generation of Residual Demand Curves using a conditional probabilistic model for functional time series*. (Submitted to Mathematics)

- G. Mestre, E. F. Sánchez-Úbeda, A. Muñoz, E. Alonso. *The Arithmetic of Stepwise Offer Curves*. (Submitted to Energy)

Conference papers

- G. Mestre, J. Portela, A. Muñoz, E. Alonso. *White noise testing for functional time series. Application to model identification and diagnosis*. 2nd Satellite CRoNoS Workshop on Functional Data Analysis - CRoNoS FDA 2018, Iasi, Romania. 31 August - 2 September 2018.

Invited conference lectures

- G. Mestre, J. Portela, G. Rice, A. Muñoz, E. Alonso. *Functional time series identification and diagnosis by means of autocorrelation analysis*. XXXVIII Congreso Nacional de Estadística e Investigación Operativa - SEIO 2019, Alcoy, Spain, 3-6 September 2019.

6.3. Future work

This dissertation aimed at pushing forward research on identification and diagnosis techniques for functional time series, and the development of a conditional probabilistic model for functional data. The forecasting framework proposed in this thesis has been applied successfully to real-world problems in electricity markets. These contributions lead to a number of future lines of research that could yield valuable contributions to the functional data field, which are summarized below:

- In many forecasting problems the actual value of the series is often influenced by lagged values of the exogenous variables. In order to capture this dynamic dependence the methodology used to define the FACF and FPACF could be adapted to obtain a functional version of the cross-covariance function. This tool could then be used to define a generalization of the impulse response function for classical Dynamic Regression models to the functional data framework.

- When fitting the SARMAHX model it would be useful to have a measure of the significance of the different parameters of the models. Developing a statistical test to quantify the importance of each term would be very useful to the practitioner, as it would provide them with a tool for selecting the most relevant variables when modelling a functional time series.

- When modelling functional time series it is customary to find that the volatility of the series presents serial correlation, which can be identified by using the proposed FACF/FPACF. The estimation procedure used by the SARMAHX model could be adapted to define a generalization of the standard GARCH model to the functional data framework. This model could capture the conditional volatility presented by the curves, which then can be used to obtain probabilistic short-term estimates of the series with a parametric model.

- The analysis performed in Chapter 5 revealed that the residuals of the SARMAHX's 24-step ahead forecasts where not independently distributed. This may be explained by the current estimation procedure for the model, which aims at minimizing the one-step ahead forecasting error. Hence, the fitted parameters may not be optimal for cases where multiple-step ahead forecasts are needed. The probabilistic model proposed in this dissertation can be used to mitigate this effect, as it is able to capture the conditional distribution of the 24-step ahead residuals. Nevertheless, the development of an optimization procedure that provides the best set of SARMAHX's parameters for a multiple step-ahead forecasts would be a relevant contribution to the field.
- The forecasting study of the aggregated Supply Curves presented in Section 5.2 has considered the Italian Day-Ahead electricity market as a single market. Although this simplification is a common approach found the literature, the Italian market is a much complex system which consists in 6 interconnected zones where saturation constraints may happen. Hence, in practice each zone has its own Supply Curve. An in-depth analysis of this market would be of great interest for market agents, as it would provide them with a detailed description of the bidding behaviour of its competitors for each market zone. The forecasting approach presented in this thesis could be adapted to this new framework to obtain short-term probabilistic forecasts of the zonal Supply Curves of this market.

Bibliography

Álvarez-Liébana, J., Bosq, D., and Ruiz-Medina, M. (2017). "Asymptotic properties of a component-wise ARH(1) plug-in predictor". In: *Journal of Multivariate Analysis* 155, pp. 12–34. DOI: 10.1016/j.jmva.2016.11.009.

Álvarez-Liébana, J. and Ruiz-Medina, M. D. (2019). "Prediction of air pollutants PM10 by ARBX(1) processes". In: *Stochastic Environmental Research and Risk Assessment* 33.10, pp. 1721–1736. DOI: 10.1007/s00477-019-01712-z.

Anastasiades, G. and McSharry, P. (2013). "Quantile Forecasting of Wind Power Using Variability Indices". In: *Energies* 6.2, pp. 662–695. DOI: 10.3390/en6020662.

Aneiros, G., Vilar, J. M., Cao, R., and Muñoz San Roque, A. (2013). "Functional Prediction for the Residual Demand in Electricity Spot Markets". In: *IEEE Transactions on Power Systems* 28.4, pp. 4201–4208. DOI: 10.1109/TPWRS.2013.2258690.

Aneiros, G., Cao, R., and Vilar, J. M. (2011). "Functional methods for time series prediction: a nonparametric approach". In: *Journal of Forecasting* 30.4, pp. 377–392. DOI: 10.1002/for.1169.

Aneiros, G. and Vieu, P. (2008). "Nonparametric time series prediction: A semi-functional partial linear modeling". In: *Journal of Multivariate Analysis* 99.5, pp. 834–857. DOI: 10.1016/j.jmva.2007.04.010.

Aneiros, G., Vilar, J., and Raña, P. (2016). "Short-term forecast of daily curves of electricity demand and price". In: *International Journal of Electrical Power & Energy Systems* 80, pp. 96–108. DOI: 10.1016/j.ijepes.2016.01.034.

Antoch, J., Prchal, L., Rosaria De Rosa, M., and Sarda, P. (2010). "Electricity consumption prediction with functional linear regression using spline estimators". In: *Journal of Applied Statistics* 37.12, pp. 2027–2041. DOI: 10.1080/02664760903214395.

Antoniadis, A., Brossat, X., Cugliari, J., and Poggi, J.-M. (2016). "A prediction interval for a function-valued forecast model: Application to load forecasting". In: *International Journal of Forecasting* 32.3, pp. 939–947. DOI: 10.1016/j.ijforecast.2015.09.001.

Antoniadis, A., Paparoditis, E., and Sapatinas, T. (2006). "A functional wavelet–kernel approach for time series prediction". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68.5, pp. 837–857.

Antoniadis, A., Paparoditis, E., and Sapatinas, T. (2009). "Bandwidth selection for functional time series prediction". In: *Statistics & Probability Letters* 79.6, pp. 733–740. DOI: 10.1016/j.spl.2008.10.028.

Arora, S. and Taylor, J. W. (2016). "Forecasting electricity smart meter data using conditional kernel density estimation". In: *Omega* 59, pp. 47–59. DOI: 10.1016/j.omega.2014.08.008.

Aue, A. and Delft, A. van (2020). "Testing for stationarity of functional time series in the frequency domain". In: *Annals of Statistics* 48.5, pp. 2505–2547. DOI: 10.1214/19-AOS1895.

Aue, A., Norinho, D. D., and Hörmann, S. (2015). "On the Prediction of Stationary Functional Time Series". In: *Journal of the American Statistical Association* 110.509, pp. 378–392. DOI: 10.1080/01621459.2014.909317.

Bagchi, P., Characiejus, V., and Dette, H. (2018). “A Simple Test for White Noise in Functional Time Series”. In: *Journal of Time Series Analysis* 39.1, pp. 54–74. DOI: 10.1111/jtsa.12264.

Baillo, A., Ventosa, M., Rivier, M., and Ramos, A. (2004). “Optimal Offering Strategies for Generation Companies Operating in Electricity Spot Markets”. In: *IEEE Transactions on Power Systems* 19.2, pp. 745–753. DOI: 10.1109/TPWRS.2003.821429.

Baladandayuthapani, V., Mallick, B. K., Young Hong, M., Lupton, J. R., Turner, N. D., and Carroll, R. J. (2008). “Bayesian Hierarchical Spatially Correlated Functional Data Analysis with Application to Colon Carcinogenesis”. In: *Biometrics* 64.1, pp. 64–73. DOI: 10.1111/j.1541-0420.2007.00846.x.

Bapna, R., Jank, W., and Shmueli, G. (2008). “Price formation and its dynamics in online auctions”. In: *Decision Support Systems* 44.3, pp. 641–656. DOI: 10.1016/j.dss.2007.09.004.

Baslis, C. G. and Bakirtzis, A. G. (2011). “Mid-Term Stochastic Scheduling of a Price-Maker Hydro Producer With Pumped Storage”. In: *IEEE Transactions on Power Systems* 26.4, pp. 1856–1865. DOI: 10.1109/TPWRS.2011.2119335.

Bello, A., Bunn, D. W., Reneses, J., and Muñoz San Roque, A. (2017). “Medium-Term Probabilistic Forecasting of Electricity Prices: A Hybrid Approach”. In: *IEEE Transactions on Power Systems* 32.1, pp. 334–343. DOI: 10.1109/TPWRS.2016.2552983.

Berzal, D., Fuente, J. de la, and Gomez, T. (2001). “Building generation supply curves under uncertainty in residual demand curves for the day-ahead electricity market”. In: *2001 IEEE Porto Power Tech Proceedings (Cat. No.01EX502)*. Porto, Portugal: IEEE, p. 6. DOI: 10.1109/PTC.2001.964623.

Bessa, R. J., Miranda, V., Botterud, A., Wang, J., and Constantinescu, E. M. (2012). “Time Adaptive Conditional Kernel Density Estimation for Wind Power Forecasting”. In: *IEEE Transactions on Sustainable Energy* 3.4, pp. 660–669. DOI: 10.1109/TSTE.2012.2200302.

Besse, P. and Ramsay, J. (1986). “Principal components analysis of sampled functions”. In: *Psychometrika* 51.2, pp. 285–311. DOI: 10.1007/BF02293986.

Besse, P. C., Cardot, H., and Stephenson, D. B. (2000). “Autoregressive Forecasting of Some Functional Climatic Variations”. In: *Scandinavian Journal of Statistics* 27.4, pp. 673–687.

Bongiorno, E. G. and Goia, A. (2016). “Classification methods for Hilbert data based on surrogate density”. In: *Computational Statistics & Data Analysis* 99, pp. 204–222. DOI: 10.1016/j.csda.2016.01.019.

Bosq, D. (2000). *Linear Processes in Function Spaces - Theory and Applications*. Lecture Notes in Statistics 149. Springer-Verlag New York.

Bouveyron, C. and Jacques, J. (2011). “Model-based clustering of time series in group-specific functional subspaces”. In: *Advances in Data Analysis and Classification* 5.4, pp. 281–300. DOI: 10.1007/s11634-011-0095-6.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (2008). *Time series analysis: forecasting and control*. 4th ed. Wiley series in probability and statistics. OCLC: ocn176895531. Hoboken, N.J: John Wiley.

Bunn, D. W. and Farmer, E. (1985). *Comparative Models for Electrical Load Forecasting*. New York, NY: John Wiley & Sons, Inc.

Cabrera, B. L. and Schulz, F. (2017). “Forecasting Generalized Quantiles of Electricity Demand: A Functional Data Approach”. In: *Journal of the American Statistical Association* 112.517, pp. 127–136. DOI: 10.1080/01621459.2016.1219259.

Campos, F. A., Muñoz, A., Sánchez-Úbeda, E. F., and Portela, J. (2016). “Strategic Bidding in Secondary Reserve Markets”. In: *IEEE Transactions on Power Systems* 31.4, pp. 2847–2856. DOI: 10.1109/TPWRS.2015.2453477.

Canale, A. and Vantini, S. (2016). “Constrained functional time series: Applications to the Italian gas market”. In: *International Journal of Forecasting* 32.4, pp. 1340–1351. DOI: 10.1016/j.ijforecast.2016.05.002.

Cardot, H., Ferraty, F., and Sarda, P. (1999). “Functional linear model”. In: *Statistics & Probability Letters* 45.1, pp. 11–22. DOI: 10.1016/S0167-7152(99)00036-X.

Castro, B. F. de, Guillas, S., and Manteiga, W. G. (2005). “Functional Samples and Bootstrap for Predicting Sulfur Dioxide Levels”. In: *Technometrics* 47.2, pp. 212–222. DOI: 10.1198/004017005000000067.

Chang, C., Chen, Y., and Ogden, R. T. (2014). “Functional data classification: a wavelet approach”. In: *Computational Statistics* 29.6, pp. 1497–1513. DOI: 10.1007/s00180-014-0503-4.

Characiejus, V. and Rice, G. (2020). “A general white noise test based on kernel lag-window estimates of the spectral density operator”. In: *Econometrics and Statistics* 13, pp. 175–196. DOI: 10.1016/j.ecosta.2019.01.003.

Chatfield, C. (2001). *Time-series forecasting*. Chapman & Hall/CRC.

Chen, K., Chen, K., Wang, Q., He, Z., Hu, J., and He, J. (2019). “Short-Term Load Forecasting With Deep Residual Networks”. In: *IEEE Transactions on Smart Grid* 10.4, pp. 3943–3952. DOI: 10.1109/TSG.2018.2844307.

Çinlar, E. (2011). *Probability and Stochastics*. Vol. 261. Graduate Texts in Mathematics. New York, NY: Springer New York. DOI: 10.1007/978-0-387-87859-1.

Ciollaro, M., Genovese, C., Lei, J., and Wasserman, L. (2014). “The functional mean-shift algorithm for mode hunting and clustering in infinite dimensions”. In: *arXiv:1408.1187 [stat]*. arXiv: 1408.1187.

Conway, J. B. (1990). *A course in Functional Analysis*. 2nd ed. Graduate texts in mathematics 96. Springer.

Cuevas, A. (2014). “A partial overview of the theory of statistics with functional data”. In: *Journal of Statistical Planning and Inference* 147, pp. 1–23. DOI: 10.1016/j.jspi.2013.04.002.

Cybenko, G. (1989). “Approximation by superpositions of a sigmoidal function”. In: *Mathematics of Control, Signals, and Systems* 2.4, pp. 303–314. DOI: 10.1007/BF02551274.

Dabo-Niang, S., Ferraty, F., and Vieu, P. (2004). “Nonparametric unsupervised classification of satellite wave altimeter forms”. In: *Proceedings in Computational Statistics*. Prague, pp. 879–886.

Dabo-Niang, S., Ferraty, F., and Vieu, P. (2007). “On the using of modal curves for radar waveforms classification”. In: *Computational Statistics & Data Analysis* 51.10, pp. 4878–4890. DOI: 10.1016/j.csda.2006.07.012.

Dai, X., Müller, H.-G., and Yao, F. (2017). “Optimal Bayes classifiers for functional data and density ratios”. In: *Biometrika* 104.3, pp. 545–560. DOI: 10.1093/biomet/asx024.

Damon, J. and Guillas, S. (2002). “The inclusion of exogenous variables in functional autoregressive ozone forecasting”. In: *Environmetrics* 13.7, pp. 759–774. DOI: 10.1002/env.527.

Damon, J. and Guillas, S. (2005). “Estimation and Simulation of Autoregressive Hilbertian Processes with Exogenous Variables”. In: *Statistical Inference for Stochastic Processes* 8.2, pp. 185–204. DOI: 10.1007/s11203-004-1031-6.

De Gooijer, J. G. and Zerom, D. (2003). “On Conditional Density Estimation”. In: *Statistica Neerlandica* 57.2, pp. 159–176. DOI: 10.1111/1467-9574.00226.

Delaigle, A. and Hall, P. (2010). “Defining probability density for a distribution of random functions”. In: *The Annals of Statistics* 38.2, pp. 1171–1193. DOI: 10.1214/09-AOS741.

Delaigle, A. and Hall, P. (2012). “Achieving near perfect classification for functional data”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 74.2, pp. 267–286. DOI: 10.1111/j.1467-9868.2011.01003.x.

Delaigle, A., Hall, P., and Bathia, N. (2012). “Componentwise classification and clustering of functional data”. In: *Biometrika* 99.2, pp. 299–313. DOI: 10.1093/biomet/ass003.

Delsol, L. (2009). “Advances on asymptotic normality in non-parametric functional time series analysis”. In: *Statistics* 43.1, pp. 13–33. DOI: 10.1080/02331880802184961.

Díaz, C. A., Villar, J., Campos, F. A., and Reneses, J. (2010). "Electricity market equilibrium based on conjectural variations". In: *Electric Power Systems Research* 80.12, pp. 1572–1579. DOI: 10.1016/j.epsr.2010.07.012.

Didericksen, D., Kokoszka, P., and Zhang, X. (2012). "Empirical properties of forecasts with the functional autoregressive model". In: *Computational Statistics* 27.2, pp. 285–298. DOI: 10.1007/s00180-011-0256-2.

Diebold, F. X. and Mariano, R. S. (1995). "Comparing Predictive Accuracy". In: *Journal of Business and Economic Statistics* 13.3. Journal Article, pp. 253–263.

Dudek, G. (2010). "Short-Term Load Forecasting Based on Kernel Conditional Density Estimation". In: *Przegląd Elektrotechniczny (Electrical Review)* 86.8, pp. 164–167.

Efron, B. and Tibshirani, R. J. (1994). *An Introduction to the Bootstrap*. Monographs on Statistics and Applied Probability. CRC press.

Erbas, B., Hyndman, R. J., and Gertig, D. M. (2007). "Forecasting age-specific breast cancer mortality using functional data models". In: *Statistics in Medicine* 26.2, pp. 458–470. DOI: 10.1002/sim.2306.

EUPHEMIA (2016). *Euphemia Public Description - PCR Market Coupling Algorithm*.

Fan, J. (1996). "Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems". In: *Biometrika* 83.1, pp. 189–206. DOI: 10.1093/biomet/83.1.189.

Fan, J. and Yim, T. H. (2004). "A crossvalidation method for estimating conditional densities". In: *Biometrika* 91.4, pp. 819–834. DOI: 10.1093/biomet/91.4.819.

Fan, J. and Zhang, W. (2008). "Statistical methods with varying coefficient models". In: *Statistics and Its Interface* 1.1, pp. 179–195. DOI: 10.4310/SII.2008.v1.n1.a15.

Fan, S. and Hyndman, R. J. (2012). "Short-Term Load Forecasting Based on a Semi-Parametric Additive Model". In: *IEEE Transactions on Power Systems* 27.1, pp. 134–141. DOI: 10.1109/TPWRS.2011.2162082.

Febrero-Bande, M. and Oviedo de la Fuente, M. (2012). "Statistical Computing in Functional Data Analysis: The R Package fda". In: *Journal of Statistical Software* 51.4, pp. 1–28. DOI: 10.18637/jss.v051.i04.

Ferraty, F. and Vieu, P. (2003). "Curves discrimination: a nonparametric functional approach". In: *Computational Statistics & Data Analysis* 44.1-2, pp. 161–173. DOI: 10.1016/S0167-9473(03)00032-X.

Ferraty, F., Keilegom, I. V., and Vieu, P. (2009). "On the Validity of the Bootstrap in Non-Parametric Functional Regression: Bootstrap in functional regression". In: *Scandinavian Journal of Statistics* 37.2, pp. 286–306. DOI: 10.1111/j.1467-9469.2009.00662.x.

Ferraty, F., Kudraszow, N., and Vieu, P. (2012). "Nonparametric estimation of a surrogate density function in infinite-dimensional spaces". In: *Journal of Nonparametric Statistics* 24.2, pp. 447–464. DOI: 10.1080/10485252.2012.671943.

Ferraty, F., Van Keilegom, I., and Vieu, P. (2012). "Regression when both response and predictor are functions". In: *Journal of Multivariate Analysis* 109, pp. 10–28. DOI: 10.1016/j.jmva.2012.02.008.

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis - Theory and Practice*. Springer Series in Statistics. Springer-Verlag New York.

Fried, R., Abbas, S., Borowski, M., and Imhoff, M. (2017). "Online Analysis of Medical Time Series". In: *Annual Review of Statistics and Its Application* 4.1, pp. 169–188. DOI: 10.1146/annurev-statistics-060116-054148.

Gabrys, R., Horváth, L., and Kokoszka, P. (2010). "Tests for Error Correlation in the Functional Linear Model". In: *Journal of the American Statistical Association* 105.491, pp. 1113–1125. DOI: 10.1198/jasa.2010.tm09794.

Gabrys, R. and Kokoszka, P. (2007). "Portmanteau Test of Independence for Functional Observations". In: *Journal of the American Statistical Association* 102.480, pp. 1338–1348. DOI: 10.1198/016214507000001111.

Gao, H. O. (2007). "Day of week effects on diurnal ozone/NOx cycles and transportation emissions in Southern California". In: *Transportation Research Part D: Transport and Environment* 12.4, pp. 292–305. DOI: 10.1016/j.trd.2007.03.004.

Gao, H. O. and Niemeier, D. A. (2008). "Using functional data analysis of diurnal ozone and NOx cycles to inform transportation emissions control". In: *Transportation Research Part D: Transport and Environment* 13.4, pp. 221–238. DOI: 10.1016/j.trd.2008.02.003.

Garcia-Gonzalez, J. and Barquin, J. (2000). "Self-unit commitment of thermal units in a competitive electricity market". In: *2000 Power Engineering Society Summer Meeting (Cat. No.00CH37134)*. Vol. 4. Seattle, WA, USA: IEEE, pp. 2278–2283. DOI: 10.1109/PESS.2000.867346.

Gianfreda, A. and Grossi, L. (2012). "Forecasting Italian electricity zonal prices with exogenous variables". In: *Energy Economics* 34.6, pp. 2228–2239. DOI: 10.1016/j.eneco.2012.06.024.

Gneiting, T. and Katzfuss, M. (2014). "Probabilistic Forecasting". In: *Annual Review of Statistics and Its Application* 1.1, pp. 125–151. DOI: 10.1146/annurev-statistics-062713-085831.

Gneiting, T. and Raftery, A. E. (2007). "Strictly Proper Scoring Rules, Prediction, and Estimation". In: *Journal of the American Statistical Association* 102.477, pp. 359–378. DOI: 10.1198/016214506000001437.

Hall, P. and Horowitz, J. L. (2007). "Methodology and convergence rates for functional linear regression". In: *The Annals of Statistics* 35.1, pp. 70–91. DOI: 10.1214/009053606000000957.

Harvey, A. C. (1990). *The Econometric Analysis of Time Series*. 2nd ed. London School of Economics handbooks in economics series. MIT Press.

Hermanussen, M. (2010). "Auxology: An Update". In: *Hormone Research in Paediatrics* 74.3, pp. 153–164. DOI: 10.1159/000317440.

Hilgert, N., Mas, A., and Verzelen, N. (2013). "Minimax adaptive tests for the functional linear model". In: *The Annals of Statistics* 41.2, pp. 838–869. DOI: 10.1214/13-AOS1093.

Hollander, M. and Wolfe, D. A. (1973). *Nonparametric Statistical Methods*. Wiley.

Holmes, M. P., Gray, A. G., and Isbell, C. L. (2010). "Fast kernel conditional density estimation: A dual-tree Monte Carlo approach". In: *Computational Statistics & Data Analysis* 54.7, pp. 1707–1718. DOI: 10.1016/j.csda.2010.01.011.

Hong, T. and Fan, S. (2016). "Probabilistic electric load forecasting: A tutorial review". In: *International Journal of Forecasting* 32.3, pp. 914–938. DOI: 10.1016/j.ijforecast.2015.11.011.

Hörmann, S., Horváth, L., and Reeder, R. (2013). "A functional version of the ARCH model". In: *Econometric Theory* 29.02, pp. 267–288. DOI: 10.1017/S0266466612000345.

Hörmann, S., Kidziński, Ł., and Hallin, M. (2015). "Dynamic functional principal components". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 77.2, pp. 319–348. DOI: 10.1111/rssb.12076.

Horváth, L., Hušková, M., and Rice, G. (2013). "Test of independence for functional data". In: *Journal of Multivariate Analysis* 117, pp. 100–119. DOI: 10.1016/j.jmva.2013.02.005.

Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. Springer Series in Statistics. Springer-Verlag New York.

Horváth, L., Kokoszka, P., and Rice, G. (2014). "Testing stationarity of functional time series". In: *Journal of Econometrics* 179.1, pp. 66–82. DOI: 10.1016/j.jeconom.2013.11.002.

Horváth, L., Liu, Z., Rice, G., and Wang, S. (2020). “A functional time series analysis of forward curves derived from commodity futures”. In: *International Journal of Forecasting* 36.2, pp. 646–665. DOI: 10.1016/j.ijforecast.2019.08.003.

Huang, S. J. and Shih, K. R. (2003). “Short-term load forecasting via ARMA model identification including non-Gaussian process considerations”. In: *IEEE Transactions on Power Systems* 18.2, pp. 673–679. DOI: 10.1109/TPWRS.2003.811010.

Hutchinson, R., McLellan, P., Ramsay, J., Sulieman, H., and Bacon, D. (2004). “Investigating the impact of operating parameters on molecular weight distributions using functional regression”. In: *Macromolecular Symposia* 206.1, pp. 495–508. DOI: 10.1002/masy.200450238.

Hyndman, R. J. and Fan, S. (2010). “Density Forecasting for Long-Term Peak Electricity Demand”. In: *IEEE Transactions on Power Systems* 25.2, pp. 1142–1153. DOI: 10.1109/TPWRS.2009.2036017.

Hyndman, R. J. and Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. 2nd ed. Melbourne, Australia: Otexts.

Hyndman, R. J. and Booth, H. (2008). “Stochastic population forecasts using functional data models for mortality, fertility and migration”. In: *International Journal of Forecasting* 24.3, pp. 323–342. DOI: 10.1016/j.ijforecast.2008.02.009.

Hyndman, R. J. and Shang, H. L. (2009). “Forecasting functional time series”. In: *Journal of the Korean Statistical Society* 38.3, pp. 199–211. DOI: 10.1016/j.jkss.2009.06.002.

Hyndman, R. J. and Shang, H. L. (2010). “Rainbow Plots, Bagplots, and Boxplots for Functional Data”. In: *Journal of Computational and Graphical Statistics* 19.1, pp. 29–45. DOI: 10.1198/jcgs.2009.08158.

Hyndman, R. J. and Ullah, M. S. (2007). “Robust forecasting of mortality and fertility rates: A functional data approach”. In: *Computational Statistics & Data Analysis* 51.10, pp. 4942–4956. DOI: 10.1016/j.csda.2006.07.028.

Hyndman, R. J. and Yao, Q. (2002). “Nonparametric Estimation and Symmetry Tests for Conditional Density Functions”. In: *Journal of Nonparametric Statistics* 14.3, pp. 259–278. DOI: 10.1080/10485250212374.

Imhof, J. P. (1961). “Computing the Distribution of Quadratic Forms in Normal Variables”. In: *Biometrika* 48.3/4, pp. 419–426. DOI: 10.2307/2332763.

Jacques, J. and Preda, C. (2013). “Funclust: A curves clustering method using functional random variables density approximation”. In: *Neurocomputing* 112, pp. 164–171. DOI: 10.1016/j.neucom.2012.11.042.

Jacques, J. and Preda, C. (2014). “Model-based clustering for multivariate functional data”. In: *Computational Statistics & Data Analysis* 71, pp. 92–106. DOI: 10.1016/j.csda.2012.12.004.

Jeon, J. and Taylor, J. W. (2012). “Using Conditional Kernel Density Estimation for Wind Power Density Forecasting”. In: *Journal of the American Statistical Association* 107.497, pp. 66–79. DOI: 10.1080/01621459.2011.643745.

Jolliffe, I. T. (2010). *Principal Component Analysis*. Springer Series in Statistics. Springer.

Joskow, P. L. (2008). “Lessons Learned from Electricity Market Liberalization”. In: *The Energy Journal* 29.SI2. DOI: 10.5547/ISSN0195-6574-EJ-Vol29-NoSI2-3.

Kargin, V. and Onatski, A. (2008). “Curve forecasting by functional autoregression”. In: *Journal of Multivariate Analysis* 99.10, pp. 2508–2526. DOI: 10.1016/j.jmva.2008.03.001.

Kaut, M. and Wallace, S. W. (2007). “Evaluation of scenario-generation methods for stochastic programming”. In: *Pacific Journal of Optimization* 3.2, pp. 257–271.

Klepsch, J., Klüppelberg, C., and Wei, T. (2017). “Prediction of functional ARMA processes with an application to traffic data”. In: *Econometrics and Statistics* 1, pp. 128–149. DOI: 10.1016/j.ecosta.2016.10.009.

Kokoszka, P. and Reimherr, M. (2013). “Determining the order of the functional autoregressive model”. In: *Journal of Time Series Analysis* 34.1, pp. 116–129. DOI: 10.1111/j.1467-9892.2012.00816.x.

Kokoszka, P. and Reimherr, M. (2017). *Introduction to Funcional Data Analysis*. Statistical Theory & Methods. CRC press.

Kokoszka, P., Rice, G., and Shang, H. L. (2017). “Inference for the autocovariance of a functional time series under conditional heteroscedasticity”. In: *Journal of Multivariate Analysis* 162, pp. 32–50. DOI: 10.1016/j.jmva.2017.08.004.

Kokoszka, P. and Young, G. (2016). “KPSS test for functional time series”. In: *Statistics* 50.5, pp. 957–973. DOI: 10.1080/02331888.2015.1128937.

Lei, J., Rinaldo, A., and Wasserman, L. (2015). “A conformal prediction approach to explore functional data”. In: *Annals of Mathematics and Artificial Intelligence* 74.1-2, pp. 29–43. DOI: 10.1007/s10472-013-9366-6.

Li, W. V. and Linde, W. (1999). “Approximation, Metric Entropy and Small Ball Estimates for Gaussian Measures”. In: *Annals of Probability* 27.3, pp. 1556–1578. DOI: 10.1214/aop/1022677459.

Liebl, D. (2013). “Modeling and forecasting electricity spot prices: A functional data perspective”. In: *The Annals of Applied Statistics* 7.3, pp. 1562–1592. DOI: 10.1214/13-AOAS652.

Liu, Z., Yan, J., Shi, Y., Zhu, K., and Pu, G. (2012). “Multi-agent based experimental analysis on bidding mechanism in electricity auction markets”. In: *International Journal of Electrical Power & Energy Systems* 43.1, pp. 696–702. DOI: 10.1016/j.ijepes.2012.05.056.

Martínez-Hernández, I., Genton, M. G., and González-Farías, G. (2019). “Robust depth-based estimation of the functional autoregressive model”. In: *Computational Statistics & Data Analysis* 131, pp. 66–79. DOI: 10.1016/j.csda.2018.06.003.

Mas, A. (2007). “Weak convergence in the functional autoregressive model”. In: *Journal of Multivariate Analysis* 98.6, pp. 1231–1261. DOI: 10.1016/j.jmva.2006.05.010.

Mas, A. and Pumo, B. (2009). “Linear Processes for Functional Data”. In: *arXiv:0901.2503*.

Mas, A. and Pumo, B. (2010). “Linear Processes for Functional Data”. In: *The Oxford Handbook of Functional Data*. Ed. by F. Ferraty and Y. Romain. Vol. 1. Oxford University Press, pp. 47–51. DOI: 10.1093/oxfordhb/9780199568444.013.3.

Maslova, I., Kokoszka, P., Sojka, J., and Zhu, L. (2010). “Statistical significance testing for the association of magnetometer records at high-, mid- and low latitudes during substorm days”. In: *Planetary and Space Science* 58.3, pp. 437–445. DOI: 10.1016/j.pss.2009.11.004.

Masry, E. (2005). “Nonparametric regression estimation for dependent functional data: asymptotic normality”. In: *Stochastic Processes and their Applications* 115.1, pp. 155–177. DOI: 10.1016/j.spa.2004.07.006.

Matabuena, M., Vidal, J. C., Hayes, P. R., Saavedra-Garcia, M., and Trillo, F. H. (2019). “Application of Functional Data Analysis for the Prediction of Maximum Heart Rate”. In: *IEEE Access* 7, pp. 121841–121852. DOI: 10.1109/ACCESS.2019.2938466.

Matheson, J. E. and Winkler, R. L. (1976). “Scoring Rules for Continuous Probability Distributions”. In: *Management Science* 22.10, pp. 1087–1096. DOI: 10.1287/mnsc.22.10.1087.

Meiring, W. (2007). “Oscillations and Time Trends in Stratospheric Ozone Levels: A Functional Data Analysis Approach”. In: *Journal of the American Statistical Association* 102.479, pp. 788–802. DOI: 10.1198/016214506000000825.

Mestekemper, T., Windmann, M., and Kauermann, G. (2010). “Functional hourly forecasting of water temperature”. In: *International Journal of Forecasting* 26.4, pp. 684–699. DOI: 10.1016/j.ijforecast.2009.10.010.

Mestre, G., Portela, J., Muñoz San Roque, A., and Alonso, E. (2020). "Forecasting hourly supply curves in the Italian Day-Ahead electricity market with a double-seasonal SARMAHX model". In: *International Journal of Electrical Power & Energy Systems* 121, p. 106083. DOI: `https://doi.org/10.1016/j.ijepes.2020.106083`.

Mestre, G., Portela, J., Rice, G., Muñoz San Roque, A., and Alonso, E. (2021). "Functional time series model identification and diagnosis by means of auto- and partial autocorrelation analysis". In: *Computational Statistics & Data Analysis* 155, p. 107108. DOI: `10.1016/j.csda.2020.107108`.

Monteiro, C., Ramirez-Rosado, I. J., Fernandez-Jimenez, L. A., and Ribeiro, M. (2018). "New probabilistic price forecasting models: Application to the Iberian electricity market". In: *International Journal of Electrical Power & Energy Systems* 103, pp. 483–496. DOI: `10.1016/j.ijepes.2018.06.005`.

Mourid, T. (2002). "Estimation and Prediction of Functional Autoregressive Processes". In: *Statistics* 36.2, pp. 125–138. DOI: `10.1080/02331880212048`.

Muñoz San Roque, A. and Sanz-Bobi, M. A. (1998). "An incipient fault detection system based on the probabilistic radial basis function network: Application to the diagnosis of the condenser of a coal power plant". In: *Neurocomputing* 23.1-3, pp. 177–194. DOI: `10.1016/S0925-2312(98)00082-4`.

Nadaraya, E. A. (1964). "On Estimating Regression". In: *Theory of Probability & Its Applications* 9.1, pp. 141–142. DOI: `10.1137/1109020`.

Narajewski, M. and Ziel, F. (2019). "Estimation and Simulation of the Transaction Arrival Process in Intraday Electricity Markets". In: *Energies* 12.23, p. 4518. DOI: `10.3390/en12234518`.

Newell, J., McMillan, K., Grant, S., and McCabe, G. (2006). "Using functional data analysis to summarise and interpret lactate curves". In: *Computers in Biology and Medicine* 36.3, pp. 262–275. DOI: `10.1016/j.compbiomed.2004.11.006`.

Nowotarski, J. and Weron, R. (2018). "Recent advances in electricity price forecasting: A review of probabilistic forecasting". In: *Renewable and Sustainable Energy Reviews* 81, pp. 1548–1568. DOI: `10.1016/j.rser.2017.05.234`.

Pankratz, A. (1991). *Forecasting with dynamic regression models*. Wiley series in probability and mathematical statistics. New York: Wiley.

Parzen, E. (1962). "On Estimation of a Probability Density Function and Mode". In: *The Annals of Mathematical Statistics* 33.3, pp. 1065–1076. DOI: `10.1214/aoms/1177704472`.

Patel, J. K. (1989). "Prediction intervals - a review". In: *Communications in Statistics - Theory and Methods* 18.7, pp. 2393–2465. DOI: `10.1080/03610928908830043`.

Pelagatti, M. (2012). *Supply Function Prediction in Electricity Auctions*. Tech. rep.

Portela, J. (2017). "Functional time series forecasting in electricity markets: a novel parametric approach". PhD thesis. Madrid: Universidad Pontificia Comillas.

Portela, J., Muñoz, A., Sánchez-Úbeda, E. F., García-González, J., and González, R. (2017). "Residual Demand Curves for Modeling the Effect of Complex Offering Conditions on Day-Ahead Electricity Markets". In: *IEEE Transactions on Power Systems* 32.1, pp. 50–61. DOI: `10.1109/TPWRS.2016.2552240`.

Portela, J., Muñoz San Roque, A., and Alonso, E. (2018). "Forecasting Functional Time Series with a New Hilbertian ARMAX Model: Application to Electricity Price Forecasting". In: *IEEE Transactions on Power Systems* 33.1, pp. 545–556. DOI: `10.1109/TPWRS.2017.2700287`.

Prete, C. L. and Hobbs, B. F. (2015). "Market power in power markets: an analysis of residual demand curves in California's day-ahead energy market (1998-2000)". In: *The Energy Journal* 0.2.

Ramsay, J. and Dalzell, C. J. (1991). "Some Tools for Functional Data Analysis". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 53.3, pp. 539–572. DOI: `10.1111/J.2517-6161.1991.TB01844.X`.

Ramsay, J., Hooker, G., and Graves, S. (2009). *Functional Data Analysis with R and MATLAB*. New York, NY: Springer New York. DOI: 10.1007/978-0-387-98185-7.

Ramsay, J. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer Series in Statistics. Springer-Verlag New York.

Raña, P., Aneiros, G., Vilar, J., and Vieu, P. (2016). "Bootstrap confidence intervals in functional nonparametric regression under dependence". In: *Electronic Journal of Statistics* 10.2, pp. 1973–1999. DOI: 10.1214/16-EJS1156.

Raña, P., Vilar, J., and Aneiros, G. (2018). "On the Use of Functional Additive Models for Electricity Demand and Price Prediction". In: *IEEE Access* 6, pp. 9603–9613. DOI: 10.1109/ACCESS.2018.2805819.

Ratcliffe, S. J., Heller, G. Z., and Leader, L. R. (2002). "Functional data analysis with application to periodically stimulated foetal heart rate data. II: Functional logistic regression". In: *Statistics in Medicine* 21.8, pp. 1115–1127. DOI: 10.1002/sim.1068.

Rice, G., Wirjanto, T., and Zhao, Y. (2020). "Forecasting value at risk with intra-day return curves". In: *International Journal of Forecasting*. DOI: 10.1016/j.ijforecast.2019.10.006.

Rice, J. A. and Silverman, B. W. (1991). "Estimating the Mean and Covariance Structure Nonparametrically When the Data are Curves". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 53.1, pp. 233–243.

Riesz, F. and Sz.-Nagy, B. (1990). *Functional Analysis*. Dover Books on Mathematics. Dover Publications.

Rosenblatt, M. (1956). "Remarks on Some Nonparametric Estimates of a Density Function". In: *Annals of Mathematical Statistics* 27.3, pp. 832–837. DOI: 10.1214/aoms/1177728190.

Rosenblatt, M. (1969). "Conditional probability density and regression estimators". In: *Multivariate Analysis II*. Ed. by P. R. Krishnaiah. New York: Academic Press.

Rossini, J. and Canale, A. (2019). "Quantifying prediction uncertainty for functional-and-scalar to functional autoregressive models under shape constraints". In: *Journal of Multivariate Analysis* 170, pp. 221–231. DOI: 10.1016/j.jmva.2018.10.007.

Sen, R. and Klüppelberg, C. (2019). "Time series of functional data with application to yield curves". In: *Applied Stochastic Models in Business and Industry* 35.4, pp. 1028–1043. DOI: 10.1002/asmb.2443.

Shah, I. and Lisi, F. (2019). "Forecasting of electricity price through a functional prediction of sale and purchase curves". In: *Journal of Forecasting*, pp. 1–18. DOI: 10.1002/for.2624.

Shah, I. and Lisi, F. (2020). "Forecasting of electricity price through a functional prediction of sale and purchase curves". In: *Journal of Forecasting* 39.2, pp. 242–259. DOI: 10.1002/for.2624.

Shang, H. L. (2012). "Point and interval forecasts of age-specific fertility rates: a comparison of functional principal component methods". In: *Journal of Population Research* 29.3, pp. 249–267. DOI: 10.1007/s12546-012-9087-4.

Shang, H. L. (2013). "ftsa: An R package for analyzing functional time series". In: *The R Journal* 5.1, pp. 64–72.

Shang, H. L., Yang, Y., and Kearney, F. (2019). "Intraday forecasts of a volatility index: functional time series methods with dynamic updating". In: *Annals of Operations Research* 282.1-2, pp. 331–354. DOI: 10.1007/s10479-018-3108-4.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Boston, MA: Springer US. DOI: 10.1007/978-1-4899-3324-9.

Turbillon, C., Marion, J.-M., and Pumo, B. (2007). "Estimation of the moving-average operator in a Hilbert space". In: *Recent Advances in Stochastic Modeling and Data Analysis*. Chania, Greece: World Scientific Publications, pp. 597–604. DOI: 10.1142/9789812709691_0070.

Ullah, S. and Finch, C. F. (2013). "Applications of functional data analysis: A systematic review". In: *BMC medical research methodology* 13.1, p. 43.

Valderrama, M. J., Ocana, F. A., and Aguilera, A. M. (2002). "Forecasting PC-ARIMA models for functional data". In: *Proceedings in Computational Statistics*. Ed. by W. Härdle and Rönz. Physica, Heidelberg, pp. 25–36. DOI: 10.1007/978-3-642-57489-4_3.

Vilar, J., Aneiros, G., and Raña, P. (2018). "Prediction intervals for electricity demand and price using functional data". In: *International Journal of Electrical Power & Energy Systems* 96, pp. 457–472. DOI: 10.1016/j.ijepes.2017.10.010.

Vilar, J. M., Cao, R., and Aneiros, G. (2012). "Forecasting next-day electricity demand and price using nonparametric functional methods". In: *International Journal of Electrical Power & Energy Systems* 39.1, pp. 48–55. DOI: 10.1016/j.ijepes.2012.01.004.

Wagner-Muns, I. M., Guardiola, I. G., Samaranayke, V. A., and Kayani, W. I. (2018). "A Functional Data Analysis Approach to Traffic Volume Forecasting". In: *IEEE Transactions on Intelligent Transportation Systems* 19.3, pp. 878–888. DOI: 10.1109/TITS.2017.2706143.

Wand, M. P. and Jones, M. C. (1995). *Kernel smoothing*. 1st ed. Monographs on statistics and applied probability 60. London ; New York: Chapman & Hall.

Wang, G., Su, Y., and Shu, L. (2016). "One-day-ahead daily power forecasting of photovoltaic systems based on partial functional linear regression models". In: *Renewable Energy* 96, pp. 469–478. DOI: 10.1016/j.renene.2016.04.089.

Wang, Y., Hu, Q., Srinivasan, D., and Wang, Z. (2019). "Wind Power Curve Modeling and Wind Power Forecasting With Inconsistent Data". In: *IEEE Transactions on Sustainable Energy* 10.1, pp. 16–25. DOI: 10.1109/TSTE.2018.2820198.

Watson, G. S. (1964). "Smooth Regression Analysis". In: *Sankhyā: The Indian Journal of Statistics* 26.4, pp. 359–372.

Weron, R. (2014). "Electricity price forecasting: A review of the state-of-the-art with a look into the future". In: *International Journal of Forecasting* 30.4, pp. 1030–1081. DOI: 10.1016/j.ijforecast.2014.08.008.

Winkler, R. L. (1972). "A Decision-Theoretic Approach to Interval Estimation". In: *Journal of the American Statistical Association* 67.337, p. 187. DOI: 10.2307/2284720.

Xu, L. and Baldick, R. (2007). "Transmission-Constrained Residual Demand Derivative in Electricity Markets". In: *IEEE Transactions on Power Systems* 22.4, pp. 1563–1573. DOI: 10.1109/TPWRS.2007.907511.

Yang, X., Shen, Q., Xu, H., and Shoptaw, S. (2007). "Functional regression analysis using an F test for longitudinal data with large numbers of repeated measures". In: *Statistics in Medicine* 26.7, pp. 1552–1566. DOI: 10.1002/sim.2609.

Zhang, J.-L., Zhang, Y.-J., Li, D.-Z., Tan, Z.-F., and Ji, J.-F. (2019). "Forecasting day-ahead electricity prices using a new integrated model". In: *International Journal of Electrical Power & Energy Systems* 105, pp. 541–548. DOI: 10.1016/j.ijepes.2018.08.025.

Zhang, X. (2016). "White noise testing and model diagnostic checking for functional time series". In: *Journal of Econometrics* 194.1, pp. 76–95. DOI: 10.1016/j.jeconom.2016.04.004.

Zhang, Y., Wang, J., and Wang, X. (2014). "Review on probabilistic forecasting of wind power generation". In: *Renewable and Sustainable Energy Reviews* 32, pp. 255–270. DOI: 10.1016/j.rser.2014.01.033.

Zhang, Z. (2018). *Multivariate Time Series Analysis in Climate and Environmental Research*. Cham: Springer International Publishing. DOI: 10.1007/978-3-319-67340-0.

Zhu, T. and Politis, D. N. (2017). "Kernel estimates of nonparametric functional autoregression models and their bootstrap approximation". In: *Electronic Journal of Statistics* 11.2, pp. 2876–2906. DOI: 10.1214/17-EJS1303.