Forecasting Functional Time Series with a New Hilbertian ARMAX Model: Application to Electricity Price Forecasting

José Portela González, Antonio Muñoz San Roque, and Estrella Alonso Pérez

***Abstract*—A functional time series is the realization of a stochastic process where each observation is a continuous function defined on a finite interval. These processes are commonly found in electricity markets and are gaining more importance as more market data become available and markets head toward continuous-time marginal pricing approaches. Forecasting these time series requires models that operate with continuous functions. This paper proposes a new functional forecasting method that attempts to generalize the standard seasonal ARMAX time series model to the L^2 Hilbert space. The structure of the proposed model is a linear regression where functional parameters operate on functional variables. The variables can be lagged values of the series (autoregressive terms), past observed innovations (moving average terms), or exogenous variables. In this approach, the functional parameters used are integral operators whose kernels are modeled as linear combinations of sigmoid functions. The parameters of each sigmoid are optimized using a Quasi-Newton algorithm that minimizes the sum of squared errors. This novel approach allows us to estimate the moving average terms in functional time series models. The new model is tested by forecasting the daily price profile of the Spanish and German electricity markets and it is compared to other functional reference models.**

***Index Terms*—Electricity price forecasting, functional ARMAX model, functional data analysis, functional time series.**

I. Introduction

FUNCTIONAL time series are defined as time sequences of functional observations. These observations, called functional data, can be one dimensional or multidimensional functions (curves or hypersurfaces respectively) taking values over a continuum which can be time, space location, wavelength, etc. Functional data can be found in numerous applications such as the study of demographic curves (mortality and fertility rates for each age observed every year [1]), or the study of ozone concentration over a certain period of time in different locations [2]. New mathematical tools are required for studying these processes and extracting useful information. The Functional Data Analysis framework (see [3]–[5] for comprehensive references) provides the necessary statistical background for analyzing functional variables, where each observation is a continuous function. It is a relatively recent research field which has experienced a substantial growth as new models are being developed and as these particular type of processes are becoming more popular. This paper is aimed at improving forecasting models for functional time series and showing its potential use in electricity markets.

There are two types of functional time series. On the one hand, the functional time series can be originated by a univariate time process of scalar values which is recorded at a finite number of equidistant time points, i.e., every hour. Then, by dividing the series into segments of equal length, we obtain a functional time series. (i.e., a time sequence of segments). Fig. 1 illustrates this transformation from the scalar process (Part a) into the functional process (Part b). The other type of functional time series is given when the observations are functions per se whose domain does not necessarily have to be time.

Functional time series processes can be found in numerous applications in electricity markets. Performing long horizon forecasts of the electricity demand or the day-ahead clearing price can be approached from a functional perspective. If the series is divided into segments of length equal to the horizon, it turns the longer forecast horizon problem into a one-step-ahead functional forecast. On the other hand, Residual Demand Curves or intraday continuous prices are directly observed as sequences of curves which stimulate the use of functional models. Hereafter, the above examples are detailed.

1) Electricity demand: Short term load forecasting is crucial for market agents and System Operators. As the electricity cannot be stored, the demand has to be satisfied instantly and overproduction incurs in cost penalties. In addition, load forecasting allows for an efficient management of resources, optimal scheduling and production planning for minimizing generation costs. In the case of retailers, the demand forecast is of utmost importance for buying the necessary energy quantity at the cheapest price.

Manuscript received September 2, 2016; revised December 22, 2016 and March 7, 2017; accepted April 16, 2017. Date of publication May 2, 2017; date of current version December 20, 2017. This work was supported in part by Ministerio de Economía y Competitividad, Spain under Grant MTM2013-48462-C2-2-R. Paper no. TPWRS-01330-2016. *(Corresponding author: José Portela González.)*

J. P. González and A. M. San Roque are with the Institute for Research in Technology (IIT), Technical School of Engineering (ICAI), Comillas Pontifical University, Madrid 28015, Spain (e-mail: Jose.Portela@iit.comillas.edu; Antonio.Munoz@iit.comillas.edu).

E. A. Pérez is with the Applied Mathematics Department, Technical School of Engineering (ICAI), Comillas Pontifical University, Madrid 28015, Spain (e-mail: Estrella.Alonso@comillas.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TPWRS.2017.2700287

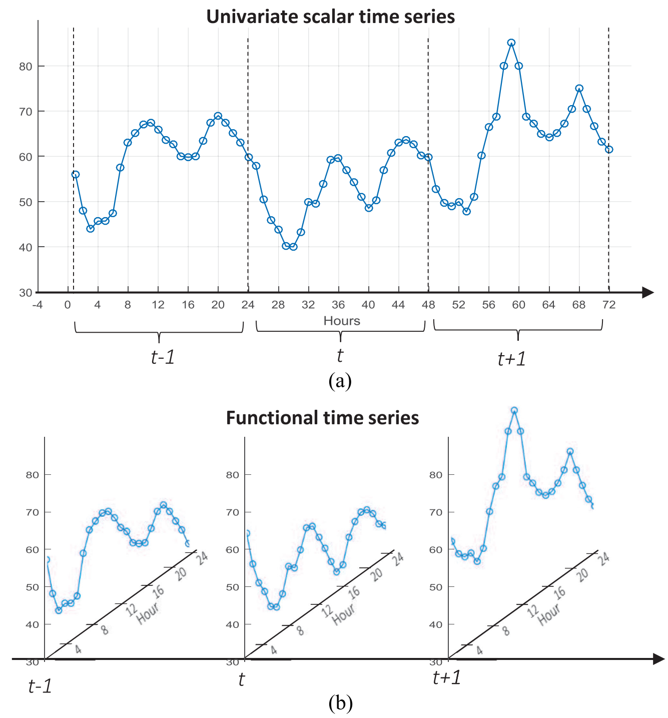

Fig. 1. Scalar time series vs. functional time series. (Part a) An example of a univariate scalar time series with hourly measurements. Dividing it into segments of equal length will produce a functional time series. (Part b) Functional time series. For each time t, a continuous function is observed.

Therefore, one of the interests of electricity companies is to forecast the demand for the 24 hours of the next day. However, smart metering is increasing the sample frequency of electricity consumption from hourly values to data every half-hour or every few minutes. Thus, in the near future, forecasting the demand for the next day would entail forecasting far more than 24 demand values, evidencing the need for new approaches such as the use of functional data methods (see [6]–[8], all based on functional nonparametric techniques).

2) Electricity prices: Forecasting electricity prices is an important matter when energy is traded in the market [9]. Usually, a clearing market price is obtained for each auction. Consequently, having an estimate of future market prices allows for an optimum bidding that maximizes profit. The recent research is leading to the possibility of creating continuous-time marginal pricing markets that can avoid the inefficiencies of having discrete time events for energy trading [10], [11]. This is a case where a functional approach can provide competitive models. Some examples of price forecasting with functional methods can be found in [12] and [13].
3) Residual demand curves: When the need of optimizing the bidding strategy of an agent arises, the concept of residual demand curve appears as a way of modeling the competitor's bidding behavior [14], [15]. This residual demand curve expresses the clearing price of the auction as a function of the amount of energy the agent is willing to buy or sell. Thus, a continuous function (curve) is obtained for each auction and retailers can use forecasts of the curves to optimize their bidding strategy [16]. Forecasting Residual Demand Curves from a functional approach using nonparametric methods is addressed in [17].
4) Intraday continuous price: EPEX spot market[1] has introduced a continuous intraday market in all the countries it comprises [18]. For example, in the case of Germany, starting at 3.00 pm on the current day, all hours of the following day can be traded until 30 minutes before delivery begins. This means that bids and offers are placed at any moment during that time span and are matched as soon as they are entered into the order book as opposed to waiting for gate closure. The benefit of local continuous trading is that market participants will be able to adjust their position as close as possible to real time [19]. Consequently, there is not an intraday price for each hour, but a price profile which starts from 3.00 pm of the former day up to 30 minutes before the hour. Functional models can be also used in this context to forecast these continuous-time price profiles.

This paper proposes an innovative forecasting model for functional time series which can be used in power systems' applications. The original contribution is the extension of the standard seasonal ARMA model to the functional framework by means of functional operators. The proposed seasonal functional time series model (the ARMAHX model) accounts for autorregressive and moving average effects which can model complex time dependencies of a time series of curves. This is of great importance for modeling variables in electricity markets as the effect of business and everyday activities lead to weekly and daily seasonalities as well as peak and low demand hours. In addition, the model allows the inclusion of explanatory variables, which are important drivers of some time series such as the electricity prices. For example, weather (wind speed, precipitation, etc.) affects the production of renewal technologies with lower generation costs influencing the offering behavior of the agents.

The paper is structured as follows: Section II introduces and reviews some background on functional forecasting models. Section III is devoted to the proposed seasonal ARMAHX model, the structure and mathematical operators are described as well as the method to estimate the different parameters. Section IV evaluates the performance of the model with a real case study comparing this method to the reference functional methods found in the literature. In addition, it performs a comparison against other well known forecasting models for electricity prices. Finally, Section VI provides the concluding remarks.

II. Functional Time Series Forecasting

A functional time series Y is defined as a sequence of functional observations $\{Y_t(v), t \in \{1, 2, \ldots, T\}, v \in V\}$ where each observation at time t is a continuous function taking

[1]EPEX being the European Power Exange Market.

values on the interval V. These functions are considered as elements belonging to the L^2 Hilbert space of real square integrable functions defined on an interval V. As was exemplified in Section I, the functional time series can be originated by different continuous-time stochastic processes:

1) From a discrete univariate time series, as is the case of hourly demand or electricity spot prices. The set of hourly prices $\{y_{1,1}, \ldots, y_{1,24}, y_{2,1}, \ldots, y_{2,24}, \ldots, y_{T,1}, \ldots, y_{T,24}\}$ can be interpreted as discrete samples of a smooth price function. By dividing the time series into intervals of length 24, the scalar time series is transformed into a new time series of observed functions $\{Y_t(v), t \in \{1, 2, \ldots, T\}, v \in V\}$ where v takes values in the interval $V = [1, 24]$ and t is measured in days.
2) From measurements over some continuous domain v at different times $t = 1, \ldots, T$. As an example, demographic functional times series are studied in [20], where $Y_t(v)$ is the mortality rate at age v measured at year t. This is also the case of Residual Demand Curves where for each hourly auction t, a curve $R_t(q)$ is observed which represents the clearing market price as a function of the energy q sold in the market.

Forecasting models of different natures have been developed for functional time series. The reference functional linear model is the Autoregressive Hilbertian model (ARH) of order 1 [21], which is defined as follows:

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

where Y is a centered functional time series, ε_t is a sequence of i.i.d.[2] random functional values and $\Psi(\cdot)$ is an autoregressive L^2 operator. This operator is unknown and needs to be estimated from the data. The approaches employed to estimate the operator gave birth to the different methods. According to [4], they can be classified in parametric and nonparametric techniques. In addition, other forecasting models follow a different approach based on dimensionality reduction.

A. Parametric Approach

Parametric models are the generalization of multivariate regression to the functional domain. An overview of parametric functional regression can be found in [3] and [22]. In parametric regression, the operator $\Psi(\cdot)$ has the form of an integral operator in L^2, which is expressed as:

$$\Psi(x)(v) = \int \psi(u, v)\, x(u)\, du, \ \ x \in L^2, \tag{2}$$

where $\psi(\cdot, \cdot)$ is the kernel of the operator and x is a curve in the L^2 space. This kernel is, in fact, a bivariate real function, or surface, that establishes the relationship between the input and the output functions. Hence, estimating the functional operator Ψ implies estimating the associated kernel surface $\psi(u, v)$. This is denoted as a parametric approach (According to [4]) because the kernel is considered as the functional parameter. The standard approach, as seen in [3], is to project the kernel into a subspace spanned by a functional basis. Then, the coordinates of the kernel in that subspace are optimized so that the forecasting error is minimized. The most common functional basis are the Functional Principal Components (FPCs) as in [23] and [24], although other options such as spline estimators can be seen in [25]. Other methods look for different basis such as [26] which develops a technique that finds the functional basis that is most relevant to the prediction. Moreover, [20] identifies uncorrelated latent components by maximizing the covariance function between functional predictors and functional responses.

Another parametric approach is followed by [21]. It solves (1) using the equivalent Yule-Walker equations for Functional Data and estimating the operator Ψ as a function of the covariance operator of the time series (see [3] and [5] for more details). In [2], [21], and [27], this approach is extended by including extra lagged observations of the series as well as functional exogenous variables.

B. Nonparametric Approach

A nonparametric linear estimation estimates the regression operator without defining a fixed structure for the operator. An in-depth description of functional nonparametric time series models is provided in [4]. A review of different nonparametric methods can be found in [28]. The most common procedure is to use the Nadaraya-Watson kernel estimator, as in [29] while other methodologies are based on wavelets [30] applied to electricity load forecasting. In addition, a semi-parametric model was developed in [17] which allows the model to include scalar exogenous variables.

C. Dimensionality Reduction

Models based on dimensionality reduction techniques pursue the goal of transforming the functional time series into a reduced set of scalar variables so that multivariate techniques can be applied. Functional Principal Component Analysis is the most common technique as extracting a few FPC is usually enough to model a high percentage of the variance of the data [31]. Each time series of FPC scores can be estimated using standard multivariate models such as ARIMA time series [1], [31], [32]. Despite the fact that FPC scores are uncorrelated, the absence of serial correlation between them is not assured. [33]–[35] propose using vector time series models to estimate all the scores jointly. These methodologies also allow the inclusion of both scalar and functional covariates into the models. One drawback of this approach arises when the number of FPC needed to accurately represent the functional time series is large. Therefore, it would be necessary to adjust several ARMA models or a large VARMA which could become a tough estimation procedure. On the other hand, they can account for seasonal effects in the series as well as for moving average terms.

In summary, parametric and nonparametric models are suitable for accounting cross-effects on the whole domain of the curves. Nonetheless, they are more limited for modeling complex time dependencies which include moving average and seasonal effects. On the other hand, reduction techniques make the most of applying standard ARMA methods at the cost of reducing the dimensionality and losing information of the process. This paper develops a functional model with operators that

[2] independent and identially distributed

account for these complex time dependencies. This model does not require turning the curves to a limited number of components, thus avoiding the corresponding information loss.

III. ARMAHX Model

The novel methodology proposed in this paper combines Functional Data theory with standard time series models and neural networks. The core idea is to estimate a parametric functional operator (2) by a finite sum of bivariate sigmoid functions. Sigmoid functions are universal function approximates [36] which are commonly used in neural networks because of their properties to model non-linear relations.

Consequently, the Autoregressive Moving Average Hilbertian model with exogenous variables (ARMAHX model) is defined following the standard time series modeling proposed in [37] but extended to functional data using parametric Hilbert operators whose kernel surfaces $\psi(u,v)$ will be modeled as a sum of bivariate sigmoid functions. The following subsections present the definition and estimation procedure of the model.

A. Model Definition

Consider a stationary and centered functional time series $\{Y_t(v), t \in \{1,2,\ldots,T\}, v \in V\}$ and a set $\{X_t^z(u_z), z \in \{1,\ldots,Z\}, t \in \{1,2,\ldots,T\}, u_z \in V_z\}$ of Z functional covariates. The ARMAHX $(p,q) \times (P,Q)_s$ model is a functional seasonal stochastic process defined by the following parameters: Parameters p and P are the regular and seasonal autoregressive orders respectively. Parameters q and Q are the regular and seasonal moving average orders respectively. The s parameter is the seasonal period. The expression for the model is defined as:

$$\begin{aligned}&(I - \Psi_1 B - \cdots - \Psi_p B^p)\left(I - \Phi_1 B^s - \cdots - \Phi_P B^{P\cdot s}\right) Y_t \\ &= (I - \Theta_1 B - \cdots - \Theta_q B^q)\left(I - \Upsilon_1 B^s - \cdots - \Upsilon_Q B^{Q\cdot s}\right)\varepsilon_t \\ &\quad + \Gamma_1 X_t^1 + \cdots + \Gamma_Z X_t^Z, \end{aligned} \tag{3}$$

where Ψ_i and Φ_i are the regular and seasonal autoregressive operators, Θ_i and Υ_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 $n \in \{1,\ldots,N\}$. Finally, I is the identity operator.

It should be noted that when $q = 0$ and $Q = 0$ it becomes a pure autoregressive model. On the other hand, it becomes a pure moving average model when $p = 0$ and $P = 0$. In addition, if the p, P, q and Q terms are zero but there are exogenous variables, the model becomes a pure functional regression model. Moreover, the multiplicative effect between the regular and seasonal terms allows us to include extra lag dependencies by combining the regular and seasonal operators, as in the scalar case.

Reordering terms, a simplified version of (3) is obtained as:

$$\begin{aligned}Y_t =& -\sum_{i=0}^{p}\sum_{j=0}^{P} \Psi_i \Phi_j B^{i+j\cdot s} Y_t + \sum_{i=0}^{q}\sum_{j=0}^{Q} \Theta_i \Upsilon_j B^{i+j\cdot s} \varepsilon_t \\ &+ \sum_{z=1}^{Z} \Gamma_z X_t^z + \varepsilon_t,\end{aligned}$$

where Ψ_0, Φ_0, Θ_0 and Υ_0 are negative identity operators $-I$, and B^0 is the null operator.

This paper follows a parametric approach and therefore, each term in the ARMAHX model is considered as a functional integral operator, as seen in (2). In addition, when this model is used for forecasting, all the error terms are unobserved. Thus, the empirical forecast equation can be expressed as follows (note that the operators have been expanded to their integral form):

$$\begin{aligned}\widehat{Y}_t(v') =& -\sum_{i=0}^{p}\sum_{j=0}^{P} \iint \psi_i(v,v')\,\phi_j(u,v)\, B^{i+j\cdot s} Y_t(u)\, du dv \\ &+ \sum_{i=0}^{q}\sum_{j=0}^{Q} \iint \theta_i(v,v')\,\vartheta_j(u,v)\, B^{i+j\cdot s} \widehat{\varepsilon}_t(u)\, du dv \\ &+ \sum_{z=1}^{Z} \int \rho_z(u_z,v')\, X_t^z(u_z)\, du_z,\end{aligned} \tag{4}$$

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

$$\widehat{\varepsilon}_t(u) = Y_t(u) - \widehat{Y}_t(u) \tag{5}$$

Up to this point, the forecasting method has been extended to functional time series using integral operators but a methodology should be defined to estimate each kernel surface.

B. Parametric Operator

Each bivariate kernel (for simplicity, only the regular autoregressive kernel $\psi_k(u,v)$ is detailed) is modeled as a weighted sum of bivariate sigmoid functions:

$$\psi_k(u,v) = \alpha_{k0}^r + \sum_{g=1}^{G_{\psi_k}} \alpha_{kg}^r \tanh(w_{kg0}^r + w_{kg1}^r u + w_{kg2}^r v)$$

where $w_{kg0}^r, w_{kg1}^r, w_{kg2}^r, \alpha_{kg}^r$ are the parameters defining 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).

This approach can be viewed as a MLP neural network with a particular configuration: an input layer with two input variables u and v. One hidden layer with a number G_{ψ_k} of nonlinear hidden units with hyperbolic tangent as the activation function and $w_{kg0}^r, w_{kg1}^r, w_{kg2}^r$ as the weights for each input. Finally, one output layer with one linear output unit having α_{kg}^r as the weights for the activation of the hidden units.

The same configuration can be extended for the rest of the terms in the model. The weight parameters defining each kernel can be grouped into several sets:

1) Regular autoregressive terms:

$$\mathbf{r}_k = \left(\alpha_{k0}^r, \alpha_{kg}^r, w_{kg0}^r, w_{kg1}^r, w_{kg2}^r\right)$$

2) Seasonal autoregressive:

$$\mathbf{R}_k = \left(\alpha_{k0}^R, \alpha_{kg}^R, w_{kg0}^R, w_{kg1}^R, w_{kg2}^R\right)$$

3) Regular moving average terms:

$$\mathbf{m}_k = \left(\alpha_{k0}^m, \alpha_{kg}^m, w_{kg0}^m, w_{kg1}^m, w_{kg2}^m\right)$$

4) Seasonal moving average terms:

$$\mathbf{M}_k = (\alpha^M_{k0}, \alpha^M_{kg}, w^M_{kg0}, w^M_{kg1}, w^M_{kg2})$$

5) Explanatory variables' terms:

$$\mathbf{b}_k = (\alpha^b_{k0}, \alpha^b_{kg}, w^b_{kg0}, w^b_{kg1}, w^b_{kg2})$$

Therefore, the model is estimated when values for the $\mathbf{r}_k$, $\mathbf{R}_k$, $\mathbf{m}_k$, $\mathbf{M}_k$ and $\mathbf{b}_k$ parameters are given.

C. Learning Algorithm

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

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

The derivatives of the error function with respect to a general parameter W is given by,

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

where $\frac{\partial \widehat{Y}_t(v)}{\partial W}$ is the derivative of the estimation with respect to the parameter W. The full mathematical expressions for the derivatives with respect to each parameter can be found in the Appendix.

Neural network procedures are known to have a tendency towards overfitting. This event occurs when the error reaches a very small value, but when new data are presented to the network the error is large. The network memorized the data, but it did not learn to generalize to new situations. In order to prevent this effect, the data are divided into two different sets: A training set and a test set (usually, 80% of the data for training and 20% for test). The optimization algorithm is executed with the training set and the error is minimized. At the same time, the adjusted model in each iteration is evaluated with the test set. The neural network with the best performance is the one that generalizes best to the unknown part of the dataset. Therefore, the selected parameter values are the ones that produce a lower error in the test set. An example of the evolution of the training and test errors is shown in Fig. 2.

IV. Application to Electricity Price Forecasting

This section evaluates the performance of the proposed model in different case studies. Two empirical comparisons are analyzed: Firstly, functional approaches are compared. Secondly, the ARMAHX model is compared with other well known existing methods for price forecasting using the same case studies that were published originally.

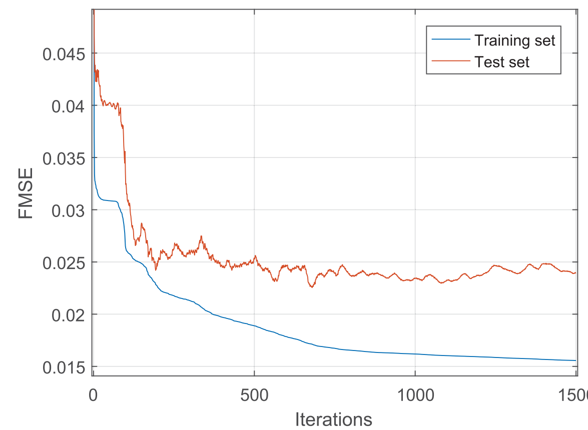

Fig. 2. Training and test errors in the optimization process for the ARMAHX model. The training error always decreases in each iteration while the test error shows oscillations. The model with minimum test error is selected.

A. Empirical Comparison With Functional Approaches

The performance of the proposed functional model is validated with two electricity price time series. The first case is the hourly Spanish electricity spot prices provided by the Spanish electricity Market Operator (www.omie.es). The second case is the hourly German electricity spot prices provided by the Open Power System Data[3]. The time range of both datasets goes from January 1, 2014 to December 31, 2015. The case study is presented as follows: Firstly, the Spanish market data is used to compare different model settings to validate the need to include complex time dependencies in the models, e.g moving average terms. Then, the German case compares the two best competing models using the same settings as the Spanish case.

The data set is divided into two different periods:

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

As explained in the Section II, the spot price series $\{y_{t,h}\}$ is transformed into a functional time series $\{Y_t(v), t \in \{1, 2, \ldots, T\}, v \in [1, 24]\}$ where each function Y_t is observed at discrete hours $v_i \in \{1, \cdots, 24\}$, thus $Y_t(v_i) = y_{t,v_i}$. Hence, each observation is a daily price profile. In addition, electricity demand and wind power production are relevant variables to forecast prices as shown in [38] and [39]. Consequently, the daily demand and wind profiles are added to the models as exogenous functional variables $\{D_t(v), t \in \{1, 2, \ldots, T\}, v \in [1, 24]\}$ and $\{W_t(v), t \in \{1, \ldots, T\}, v \in [1, 24]\}$ respectively.

Prior to the training of the functional forecasting models, it is convenient to verify that the time series are stationary. As mentioned in [40], if the sample autocorrelation function of the series goes to zero slowly, it means that the series are not stationary in the mean and both the output and the input series

[3] Available at http://open-power-system-data.org/, Data Package Time series, version 2016-10-28

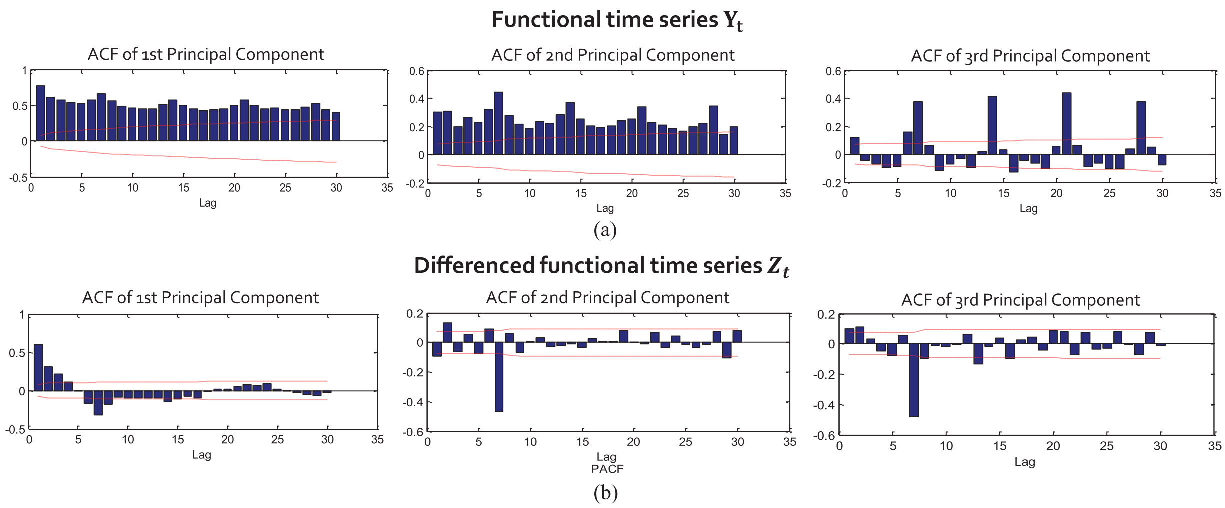

Fig. 3. Autocorrelation Functions (ACF) of principal component scores. (Part a) ACF for the time series of the three principal component scores for the price curves data Y_t. The slow decay in the correlation terms of the ACF plot evidence the need for differencing to make the series stationary. (Part b) ACF of the scores of the differenced time series Z_t. The slow decay is no longer seen and a moving average effect can be clearly identified from these plots.

should be differenced. Thus stationary is verified by extracting functional principal components and validating that each component's scores time series is stationary. Fig. 3 shows the Autocorrelation functions (ACF) for the three principal component scores time series. A slow decay in the correlation terms is observed within the ACF plot on a regular basis as well as on the seasonal component. By seasonal differencing the price time series at lag 7, the new ACF no longer shows the slow decay, evidencing that the mean in the transformed time series is stationary. Demand and wind power functional time series should also be differentiated at lag 7 in accordance to the transformation of the price data [40]. Hence, this provides the following sets:

$$Z_t(v) = Y_t(v) - Y_{t-7}(v)$$
$$U_t(v) = D_t(v) - D_{t-7}(v)$$
$$V_t(v) = W_t(v) - W_{t-7}(v)$$

The different methods that are compared in the analysis are described hereafter. When possible, the models will account for the seasonal component at lag 7, due to the weekly dependency in the series shown in Fig. 3.

1) *Naïve:* The Naïve method consists in using past observed data as the forecast. This simple method provides an initial benchmark in the comparison. Due to the fact that there is a strong weekly correlation, the forecast is obtained as
$$\widehat{Y}_t^{\mathrm{Naive}}(v) = Y_{t-7}(v)$$
2) *Functional reference method:* Among the functional linear methods, we have chosen the method proposed in [21] and [27] as the reference functional model. This method has the ability to include functional covariates which are needed for this case study. Three models have been adjusted for this study, a pure regular autoregressive (ARH_DG) model, a regular autoregressive with exogenous variables (ARHX_DG) and an autoregressive model (ARHs_DG) that includes a seasonal component at lag 7 as if it were an extra explanatory variable. The models have been adjusted using the R package *far*, which provides the code for training and evaluating of the model. The details can be found in [27]. Therefore, each model can be expressed as:
$$\widehat{Z}_t^{\mathrm{ARH\_DG}} = \Psi_1(Z_{t-1})$$
$$\widehat{Z}_t^{\mathrm{ARHs\_DG}} = \Psi_1(Z_{t-1}) + \Psi_7(Z_{t-7})$$
$$\widehat{Z}_t^{\mathrm{ARHX\_DG}} = \Psi_1(Z_{t-1}) + \Gamma_1(U_t) + \Gamma_2(V_t)$$
3) *FPC dimension reduction:* The functional dimension reduction approach is also compared. FPCA is applied to the functional time series Y using the PACE Matlab toolbox for functional data as described in [41]. Three components are extracted which model the 98% of the variane. The scores are forecasted using univariate ARMA models. Different models are adjusted varying the regular autoregressive order and the seasonal moving average order. The model named AR_PC uses an $\mathrm{ARMA}(1,0)$ to forecast each score time series. ARMA_PC includes the seasonal moving average component, hence, it uses an ARMA $(1,0)\times(0,1)_7$. Models ARX_PC and ARMAX_PC have the same configuration as the previous two models but include the hourly wind production and electricity demand as exogenous variables. This is done as suggested by [34]: Functional Principal Components are extracted from the wind and demand functional time series and the resulting scores are used as input variables to the ARMAX models.
4) *Optimized Functional ARMAHX:* This is the model proposed in this paper. Four configurations have been adjusted with different structures:

$$\hat{Z}^{\mathrm{ARH\_NN}} = \mathrm{ARMAH}(1,0)\times(0,0)_7 = \Psi(Z_{t-1})$$
$$\hat{Z}^{\mathrm{ARMAH\_NN}} = \mathrm{ARMAH}(1,0)\times(0,1)_7$$

TABLE I
AVERAGE ERRORS FOR EACH METHOD IN THE SPANISH MARKET

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

$$= \Psi(Z_{t-1}) + \Upsilon(\hat{e}_{t-7})$$

$$\hat{Z}^{\text{ARHX\_NN}} = \text{ARMAHX}(1,0) \times (0,0)_7$$
$$= \Psi(Z_{t-1}) + \Gamma_1(U_t) + \Gamma_2(V_t)$$

$$\hat{Z}^{\text{ARMAHX\_NN}} = \text{ARMAHX}(1,0) \times (0,1)_7$$
$$= \Psi(Z_{t-1}) + \Upsilon(\hat{e}_{t-7}) + \Gamma_1(U_t) + \Gamma_2(V_t)$$

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

The parameters of all the models are adjusted with the In-Sample data and then evaluated in the Out-Of-Sample period. Even though the forecast of the functional models is a continuous function, we are interested in comparing the prediction against the real hourly prices. Therefore, error measurements usually found in electriciy price forecasting literature (see [9]) are calculated in order to compare the models:

1) Mean Absolute Error (MAE)

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

2) Root Mean Squared Error (RMSE)

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

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

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

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

The Diebold-Mariano (DM) test [42] is used to compare the predictive accuracy of the different forecasts. The DM test is a head-to-head method which compares the forecast error of two different models[4]. Table II shows the test results for the Out-Of-Sample comparison. Each cell contains the p-value that results from comparing the model of the cell's row against the model of the cell's column. A low p-value means that the null hypothesis has to be rejected and therefore the average errors shown in Table I can be considered statistically different. The DM test requires that the loss differential be covariance stationary. In order to verify that assumption, a KPSS test is performed for each loss differential. Underlined values in Table I correspond to those comparisons which did not pass the stationarity test and whose results should be therefore analyzed with caution.

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

The different methods in each group are compared. All the pure regular autorregressive models have similar performances as seen by the DM test in Table II, however, the seasonal methods without explanatory variables are significantly different and we can see how the moving average component is justified, being the method presented in this paper the one that provides the best results. When considering the group of autorregressive

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

TABLE II
HEAD-TO-HEAD DIEBOLD-MARIANO TEST'S P-VALUES FOR OUT-OF-SAMPLE PERIOD OF THE SPANISH MARKET CASE

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

TABLE III
AVERAGE ERRORS FOR THE TWO COMPETING METHODS FOR THE GERMAN MARKET

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

TABLE IV
OUT-OF-SAMPLE ERRORS FOR THE TWO COMPETING METHODS FOR EACH DAY OF THE WEEK IN THE SPANISH AND THE GERMAN MARKET

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

models with the exogenous variables, Table II shows that all three methods perform alike. Finally, the last group only consists of two models. It can be seen how the ARMAHX method is further improved by adding the moving average term and the improvement is statistically validated with the p-values obtained in Table II.

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

In order to understand the results from the proposed method, Fig. 4 shows the optimized kernels for the trained ARMAHX$(1, 0) \times (0, 1)_7$ model in each country. These models were defined by four functional operators: two for the explanatory variables, one for the regular autoregressive term and one for the seasonal moving average term. Each of these surfaces is formed by summing up 8 bivariate hyperbolic tangent functions, which are the output of the optimization process. They show the influence that each input variable has on the output variable. The u axis represents the input variable and the v axis, the output function. The autoregressive kernel shows that all the hours of the price profile are very much dependent on the value at hour 24 of the price profile of the previous day. The moving average kernel shows high values in the diagonal, which means that each hour of the output is influenced by the error at that hour 7 days ago. The wind explanatory variable kernel shows that there is a

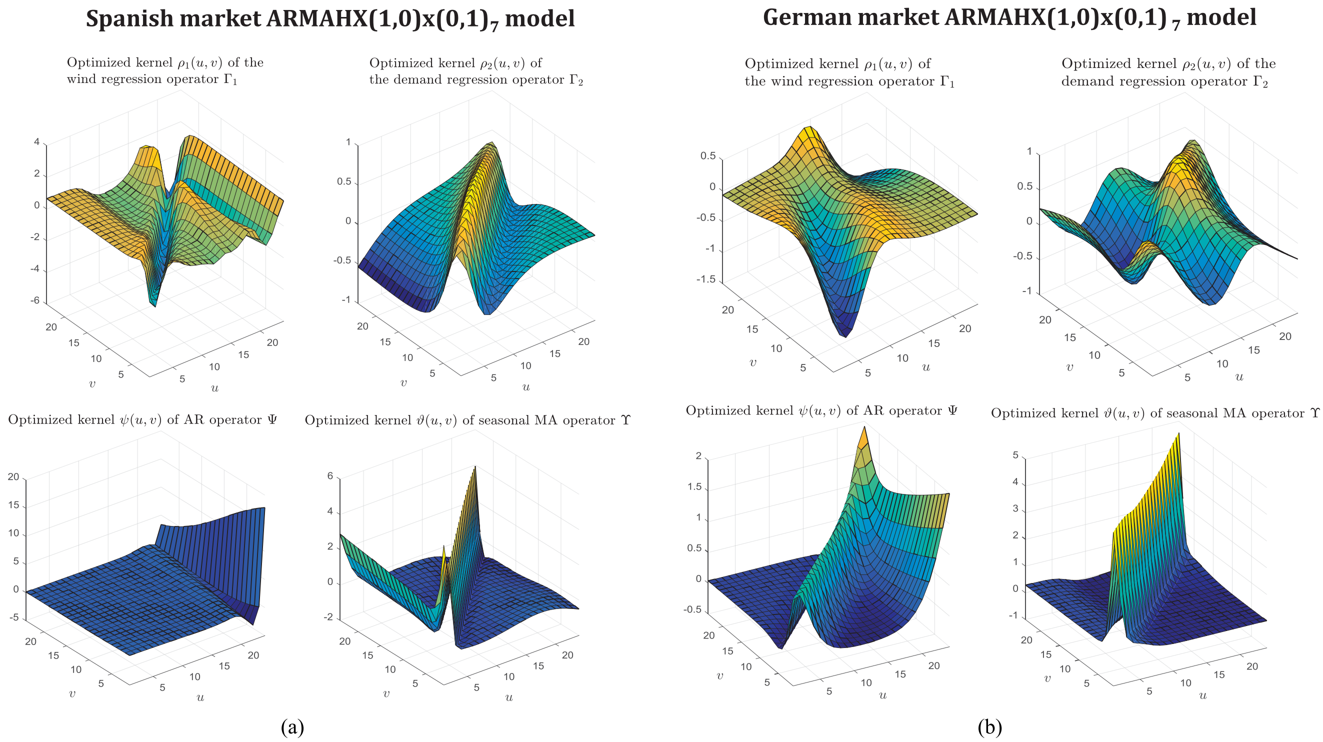

Fig. 4. Estimated kernel functions of the operators of the best ARMAHX models. (Part a) Kernel functions for the model trained with the Spanish case. (Part b) Kernel functions for the model trained with the German case.

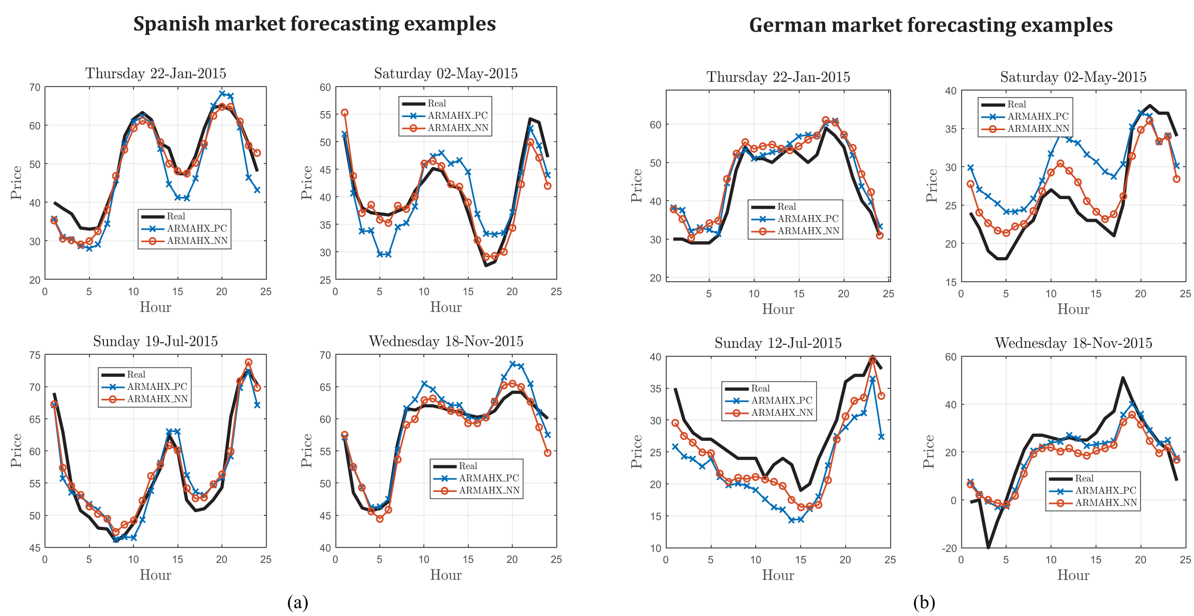

Fig. 5. Real price profiles and corresponding forecasts of the two competing models. (Part a) Four days from the Spanish market case. (Part b) Four days from the German market case.

strong negative dependence on the diagonal, which means that when the wind power is high for one hour, prices will drop at that hour and viceversa. In addition, the demand operator shows a positive dependency on the diagonal.

Fig. 5 shows forecasting examples for different weekdays for each country. The real price profiles are compared with the estimation provided by the best competing models.

B. *Empirical Comparison With Non-Functional Approaches*

This paper proposes a functional forecasting model which can be applied to diverse problems in electricity markets, as it was introduced in Section I. Nevertheless, in the case of forecasting the hourly electricity price, non-functional approaches have been traditionally used in the literature. It is out of the scope of the paper to perform a detailed comparison of the proposed

TABLE V
OUT-OF-SAMPLE MAPE FOR THE CASE STUDY AND MODELS PROPOSED IN [38]

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

Note: Functional ARMAHX results are included.

model with the latest price forecasting methods; however, we have included an empirical comparison of the performance of the proposed model with other well known existing methods using the same case studies that were published originally.

The price forecasting setup in [38] has been selected for the empirical comparison. There, the Spanish day-ahead electricity price of years 2007 and 2008 is forecasted by a wide variety of models, ranging from MLP Neural Network to dynamic regression and periodic models. The latter methods have been used extensively in price forecasting (see [43] and [44] respectively). Seasonality as well as load and wind explanatory variables are included in the models.

A competing functional ARMAHX model is adjusted with the same in-sample period used in [38] (01/01/2007 to 10/08/2007) and validated with the same Out-Of-Sample period (11/05/2007 to 31/07/2008). The hourly electricity spot price, load and wind energy are transformed to functional time series as explained in Section IV-A. Firstly, both explanatory variables and the price functional time series are differenced at lag 1. Then, an $\text{ARMAHX}(0,1) \times (1,0)_7$ is adjusted for the differenced time series with 5 sigmoids for each operator.

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

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

One of such improvements could be the extension of intervention analysis [45] to the functional framework, which would allow to model the effect of a sudden change in the time series to be forecasted. Dummy intervention variables of different nature can account for the occurrence of an event that affects the response time series, e.g. a change in the regulatory policy would be modeled as a step function, or a strike by means of a pulse function. These intervention variables could be added as explanatory variables to the functional model.

V. SUMMARY

This paper fills an existing gap in forecasting functional time series which is the lack of functional methods that model complex time dependencies such as accounting for seasonality as well as moving average effects. The methods found in the literature address this issue by applying standard ARMA methods at the cost of reducing the dimensionality and losing information of the process.

This manuscript contributes significantly by defining a functional version of the seasonal ARMAHX model based on Hilbert operators. These operators are able to account for cross-effects on the entire curves, and do not require to turn the curves to a limited number of components, thus avoiding the corresponding loss of information. This paper models the kernel of the Hilbert integral operator as a sum of bivariate sigmoid functions, which is a novel approach in the literature. The parameters of the sigmoids are optimized by minimizing the forecasting error and a recursive algorithm allows for the estimation of the moving average terms of the ARMAHX model.

Functional time series appear in electricity markets when analyzing load and price time series, Residual Demand Curves or continuous time intraday markets. All of them are sensitive to weather conditions and to the effects of business and everyday activities that lead to weekly and daily seasonalities. Therefore, a suitable forecasting model has to account for all these events.

The performance of the model has been validated with the Spanish and German day-ahead hourly electricity market price. It has been shown how the proposed model outperforms the other reference functional models. Moreover, it has proven to be a competitive model against other well known existing methods for price forecasting.

This method opens a wide range of future developments such as the aforementioned inclusion of dummy intervention variables, the usage of a population-based algorithm (particle-swarm, differential evolution...) in the optimization process or the possibility of generating forecast intervals and scenarios.

VI. APPENDIX

This appendix follows the derivation of the expressions for the derivatives of the model parameters from Section III-C. If the parameter W is $r_{k,n}$, i.e., a parameter that belongs to the set $\mathbf{r}_k$ of regular autoregressive parameters, this derivative can be directly obtained deriving (4):

$$\frac{\partial \widehat{Y}_t(v')}{\partial r_{k,n}} = -\sum_{j=0}^{P} \iint \frac{\partial \psi_k(v,v')}{\partial r_{k,n}} \phi_j(u,v)\, Y_{t-(k+j\cdot s)}(u)\, du dv$$

$$+ \sum_{i=0}^{q}\sum_{j=0}^{Q} \iint \theta_i(v,v')\, \vartheta_j(u,v)\, \frac{\partial \widehat{\varepsilon}_{t-(i+j\cdot s)}(u)}{\partial r_{k,n}} du dv.$$

As can be seen, due to the fact that $\widehat{\varepsilon}_{t-(i+j\cdot s)}$ depends on past estimations of Y, it is also dependent on r_i and, therefore, can have a nonzero derivative. Similar expressions can be derived for each set of parameters.

The derivatives of the error terms with respect to a certain parameter W can be expressed as a function of former calculated derivatives. Therefore, by deriving (5) we obtain:

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

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

The complete derivatives are shown for all of the parameters:

$$\frac{\partial\widehat{Y}_t(v')}{\partial r_{k,n}} = -\sum_{j=0}^{P}\iint \frac{\partial\psi_k(v,v')}{\partial r_{k,n}}\phi_j(u,v)\,Y_{t-(k+j\cdot s)}(u)dudv$$

$$-\sum_{i=0}^{q}\sum_{j=0}^{Q}\iint \theta_i(v,v')\,\vartheta_j(u,v)\,\frac{\partial\widehat{Y}_{t-(i+j\cdot s)}(u)}{\partial r_{k,n}}dudv,$$

$$\frac{\partial\widehat{Y}_t(v')}{\partial R_{k,n}} = -\sum_{i=0}^{p}\iint \psi_i(v,v')\,\frac{\partial\phi_k(u,v)}{\partial R_{k,n}}Y_{t-(i+k\cdot s)}(u)dudv$$

$$-\sum_{i=0}^{q}\sum_{j=0}^{Q}\iint \theta_i(v,v')\,\vartheta_j(u,v)\,\frac{\partial\widehat{Y}_{t-(i+j\cdot s)}(u)}{\partial R_{k,n}}dudv,$$

$$\frac{\partial\widehat{Y}_t(v')}{\partial m_{k,n}} = \sum_{j=0}^{Q}\iint \frac{\partial\theta_k(v,v')}{\partial m_{k,n}}\vartheta_j(u,v)\,\widehat{\varepsilon}_{t-(k+j\cdot s)}(u)dudv$$

$$-\sum_{i=0}^{q}\sum_{j=0}^{Q}\iint \theta_i(v,v')\,\vartheta_j(u,v)\,\frac{\partial\widehat{Y}_{t-(i+j\cdot s)}(u)}{\partial m_{k,n}}dudv,$$

$$\frac{\partial\widehat{Y}_t(v')}{\partial M_{k,n}} = \sum_{i=0}^{q}\iint \theta_i(v,v')\,\frac{\partial\vartheta_k(u,v)}{\partial M_{k,n}}\widehat{\varepsilon}_{t-(i+k\cdot s)}(u)dudv$$

$$-\sum_{i=0}^{q}\sum_{j=0}^{Q}\iint \theta_i(v,v')\,\vartheta_j(u,v)\,\frac{\partial\widehat{Y}_{t-(i+j\cdot s)}(u)}{\partial M_{k,n}}dudv,$$

$$\frac{\partial\widehat{Y}_t(v')}{\partial b_{k,n}} = \int \frac{\partial\rho_k(u_k,v')}{\partial b_{k,n}}X_t^k(u_k)\,du_k$$

$$-\sum_{i=0}^{q}\sum_{j=0}^{Q}\iint \theta_i(v,v')\,\vartheta_j(u,v)\,\frac{\partial\widehat{Y}_{t-(i+j\cdot s)}(u)}{\partial b_{k,n}}dudv.$$

Finally, the derivatives of the kernels $\frac{\partial\psi_k(u,v)}{\partial r_{k,n}}$, $\frac{\partial\phi_k(u,v)}{\partial R_{k,n}}$, $\frac{\partial\theta_k(u,v)}{\partial m_{k,n}}$, $\frac{\partial\vartheta_k(u,v)}{\partial M_{k,n}}$ and $\frac{\partial\rho_k(u,v)}{\partial b_{k,n}}$ are calculated. At this point, these derivatives are calculated the same way because all of them are specified in the same manner. Hence, the derivatives of the kernel $\psi_k(u,v)$ (or $\phi_k(u,v), \theta_k(u,v), \vartheta_k(u,v), \rho_k(u,v)$) with respect to each particular weight $r_{k,n}$ (or $R_{k,n}, m_{k,n}, M_{k,n}, b_{k,n}$) are obtained as:

$$\frac{\partial\psi_k(u,v)}{\partial\alpha_{k0}^r} = 1$$

$$\frac{\partial\psi_k(u,v)}{\partial\alpha_{kg}^r} = \tanh(w_{kg0}^r + w_{kg1}^r u + w_{kg2}^r v)$$

$$\frac{\partial\psi_k(u,v)}{\partial w_{kg0}^r} = \alpha_{kg}\cdot\tanh'(w_{kg0}^r + w_{kg1}^r u + w_{kg2}^r v)$$

$$\frac{\partial\psi_k(u,v)}{\partial w_{kg1}^r} = \alpha_{kg}\cdot\tanh'(w_{kg0}^r + w_{kg1}^r u + w_{kg2}^r v)\cdot u$$

$$\frac{\partial\psi_k(u,v)}{\partial w_{kg2}^r} = \alpha_{kg}\cdot\tanh'(w_{kg0}^r + w_{kg1}^r u + w_{kg2}^r v)\cdot v$$

ACKNOWLEDGMENT

The authors would like to thank the reviewers for the valuable comments and suggestions that have helped improving the final version of this paper.

REFERENCES

[1] R. J. Hyndman and M. Shahid Ullah, “Robust forecasting of mortality and fertility rates: A functional data approach,” *Comput. Statist. Data Anal.*, vol. 51, no. 10, pp. 4942–4956, Jun. 2007.

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

[3] J. O. Ramsay and B. W. Silverman, *Functional Data Analysis* (ser. Springer Series in Statistics), 2nd ed. New York, NY, USA: Springer, Jun. 2005.

[4] F. Ferraty and P. Vieu, *Nonparametric Functional Data Analysis: Theory and Practice*. New York, NY, USA: Springer, Jun. 2006.

[5] L. Horváth and P. Kokoszka, *Inference for Functional Data With Applications*. New York, NY, USA: Springer, 2012, vol. 200.

[6] I. Shah and F. Lisi, “Day-ahead electricity demand forecasting with nonparametric functional models,” in *Proc. 12th Int. Conf. Eur. Energy Market*, May 2015, pp. 1–5.

[7] M. Chaouch, “Clustering-based improvement of nonparametric functional time series forecasting: Application to intra-day household-level load curves,” *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 411–419, Jan. 2014.

[8] E. Paparoditis and T. Sapatinas, “Short-term load forecasting: The similar shape functional time-series predictor,” *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 3818–3825, Nov. 2013.

[9] R. Weron, “Electricity price forecasting: A review of the state-of-the-art with a look into the future,” *Int. J. Forecasting*, vol. 30, no. 4, pp. 1030–1081, Oct. 2014.

[10] M. Parvania and A. Scaglione, “Unit commitment with continuous-time generation and ramping trajectory models,” *IEEE Trans. Power Syst.*, vol. 31, no. 4, pp. 3169–3178, Jul. 2016.

[11] M. Parvania and R. Khatami, “Continuous-time marginal pricing of electricity,” *IEEE Trans. Power Syst.*, vol. 32, no. 3, pp. 1960–1969, May 2017.

[12] J. M. Vilar, R. Cao, and G. Aneiros, “Forecasting next-day electricity demand and price using nonparametric functional methods,” *Int. J. Elect. Power Energy Syst.*, vol. 39, no. 1, pp. 48–55, 2012.

[13] D. Liebl, “Modeling and forecasting electricity spot prices: A functional data perspective,” *Ann. Appl. Statist.*, vol. 7, no. 3, pp. 1562–1592, 2013.

[14] A. A. S. de la Nieta, J. Contreras, J. I. Muñoz, and M. O’Malley, “Modeling the impact of a wind power producer as a price-maker,” *IEEE Trans. Power Syst.*, vol. 29, no. 6, pp. 2723–2732, Nov. 2014.

[15] J. Portela, A. Muñoz, E. F. Sánchez-Úbeda, J. García-González, and R. González, “Residual demand curves for modeling the effect of complex offering conditions on day-ahead electricity markets,” *IEEE Trans. Power Syst.*, vol. 32, no. 1, pp. 50–61, Jan. 2017.

[16] F. A. Campos, A. Muñoz, E. F. Sánchez-Úbeda, and J. Portela, “Strategic bidding in secondary reserve markets,” *IEEE Trans. Power Syst.*, vol. 31, no. 4, pp. 2847–2856, Jul. 2016.

[17] G. Aneiros, J. Vilar, R. Cao, and A. Muñoz, "Functional prediction for the residual demand in electricity spot markets," *IEEE Trans. Power Syst.*, vol. 28, no. 4, pp. 4201–4208, Nov. 2013.
[18] EPEXSPOT Press Release. (Mar. 30, 2017). [Online]. Available: https://www.epexspot.com/en/press-media/press/details/press/_30-minute_ continuous_Intraday_trading_successfully _launched_in_France _Germany_and_Switzerland
[19] K. Neuhoff, N. Ritter, A. Salah-Abou-El-Enien, and P. Vassilopoulos, "Intraday markets for power: Discretizing the continuous trading," Cambridge Working Papers in Economics 1616, Faculty Econ., Univ. Cambridge, Cambridge, U.K., Mar. 2016. [Online]. Available: https://ideas.repec.org/p/cam/camdae/1616.html
[20] R. J. Hyndman and H. L. Shang, "Forecasting functional time series," *J. Korean Statist. Soc.*, vol. 38, no. 3, pp. 199–211, 2009.
[21] D. Bosq, *Linear Processes in Function Spaces: Theory and Applications*. New York, NY, USA: Springer-Verlag, 2000, vol. 149.
[22] G. He, H. G. Müller, and J. L. Wang, "Extending correlation and regression from multivariate to functional data," in *Asymptotics in Statistics and Probability*, M. L. Puri, ed. Leiden, Netherlands: VSP, pp. 197–210, 2000.
[23] F. Yao, H. G. Müller, and J. L. Wang, "Functional linear regression analysis for longitudinal data," *Ann. Statist.*, vol. 33, no. 6, pp. 2873–2903, 2005.
[24] J. J. Faraway, "Regression analysis for a functional response," *Technometrics*, vol. 39, pp. 254–261, 1997.
[25] J. Antoch, L. Prchal, M. Rosaria De Rosa, and P. Sarda, "Electricity consumption prediction with functional linear regression using spline estimators," *J. Appl. Statist.*, vol. 37, no. 12, pp. 2027–2041, 2010.
[26] V. Kargin and A. Onatski, "Curve forecasting by functional autoregression," *J. Multivariate Anal.*, vol. 99, no. 10, pp. 2508–2526, Nov. 2008.
[27] J. Damon and S. Guillas, "Estimation and simulation of autoregressive Hilbertian processes with exogenous variables," *Statist. Inference Stoch. Process.*, vol. 8, no. 2, pp. 185–204, 2005.
[28] G. Aneiros-Pérez, R. Cao, and J. M. Vilar-Fernández, "Functional methods for time series prediction: A nonparametric approach," *J. Forecasting*, vol. 30, no. 4, pp. 377–392, 2011.
[29] E. Masry, "Nonparametric regression estimation for dependent functional data: Asymptotic normality," *Stoch. Process. Appl.*, vol. 115, no. 1, pp. 155–177, Jan. 2005.
[30] A. Antoniadis, E. Paparoditis, and T. Sapatinas, "A functional waveletkernel approach for time series prediction," *J. Roy. Statist. Soc., B*, vol. 68, no. 5, pp. 837–857, 2006.
[31] M. J. Valderrama, F. A. Ocana, and A. M. Aguilera, "Forecasting PC-ARIMA models for functional data," in *Proc. Comput. Statist.*, 2002, pp. 25–36.
[32] R. J. Hyndman, H. Booth, and F. Yasmeen, "Coherent mortality forecasting: The product-ratio method with functional time series models," *Demography*, vol. 50, no. 1, pp. 261–283, 2013.
[33] R. Sen, "Time series of functional data," Tech. Rep., TU Munich, 2010. [Online]. Available: http://mediatum.ub.tum.de/doc/1079227/file.pdf
[34] A. Aue, D. Norinho, and S. Hörmann, "On the prediction of stationary functional time series," *J. Amer. Statist. Assoc.*, vol. 110, no. 509, pp. 378–392, 2015.
[35] J. Klepsch, C. Klüppelberg, and T. Wei, "Prediction of functional ARMA processes with an application to traffic data," *Econometr. Statist.*, vol. 1, pp. 128–149, Jan. 2017.
[36] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Math. Control, Signals Syst.*, vol. 2, no. 4, pp. 303–314, 1989.
[37] G. E. P. Box and G. Jenkins, *Time Series Analysis, Forecasting and Control*. San Francisco, CA, USA: Holden-Day, 1970.
[38] A. Cruz, A. Muñoz, J. L. Zamora, and R. Espínola, "The effect of wind generation and weekday on spanish electricity spot price forecasting," *Elect. Power Syst. Res.*, vol. 81, no. 10, pp. 1924–1935, 2011.
[39] T. Jonsson, P. Pinson, H. A. Nielsen, H. Madsen, and T. S. Nielsen, "Forecasting electricity spot prices accounting for wind power predictions," *IEEE Trans. Sustain. Energy*, vol. 4, no. 1, pp. 210–218, Jan. 2013.
[40] A. Pankratz, *A Primer on ARIMA Models*. Hoboken, NJ, USA: Wiley, 1991.
[41] H. G. Müller, "Functional modeling of longitudinal data," in *Longitudinal Data Analysis (Handbooks of Modern Statistical Methods)*. New York, NY, USA: CRC, 2008, pp. 223–252.
[42] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," *J. Bus. Econ. Statist.*, vol. 13, no. 3, pp. 253–263, 1995.
[43] F. Nogales and A. J. Conejo, "Electricity price forecasting through transfer function models," *J. Oper. Res. Soc.*, vol. 57, no. 4, pp. 350–356, 2006.
[44] S. J. Koopman, M. Ooms, and M. A. Carnero, "Periodic seasonal Reg-ARFIMA–GARCH models for daily electricity spot prices," *J. Amer. Stat. Assoc.*, vol. 102, no. 477, pp. 16–27, 2007.
[45] G. E. P. Box and G. C. Tiao, "Intervention analysis with applications to economic and environmental problems," *J. Amer. Statist. Assoc.*, vol. 70, no. 349, pp. 70–79, 1975.

José Portela González received the Industrial Engineer degree in 2011 from the Universidad Pontificia Comillas, Madrid, Spain. He is currently working toward the Ph.D. degree at the Instituto de Investigacin Tecnolgica (Technological Research Institute). His areas of interest include forecasting methods applied to electricity markets along with forecasting techniques in the functional data analysis field.

Antonio Muñoz San Roque received the Electrical Engineering degree in 1991 and the Ph.D. degree in 1996 from the Universidad Pontificia Comillas, Madrid, Spain. He is currently with the research staff at the Instituto de Investigacin Tecnolgica, Madrid, Spain, and a Full Professor at the ICAI School of Engineering, Comillas University. His areas of interest include the application of artificial intelligence techniques to the monitoring and diagnosis of industrial processes, nonlinear system identification and time series forecasting, analog electronics and digital signal processing.

Estrella Alonso Pérez received the mathematics degree from the Universidad Autnoma de Madrid in 1998, and the Ph.D. degree from the Universidad Complutense de Madrid, Spain, in 2007. She is currently a Lecturer at the ICAI School of Engineering, Comillas University. Her areas of interest include the auctions theory and functional data analysis.