Using robust FPCA to identify outliers in functional time series, with applications to the electricity market

Juan M.Vilar, Paula Raña and Germán Aneiros

Abstract

This study proposes two methods for detecting outliers in functional time series. Both methods take dependence in the data into account and are based on robust functional principal component analysis. One method seeks outliers in the series of projections on the first principal component. The other obtains uncontaminated forecasts for each data set and determines that those observations whose residuals have an unusually high norm are considered outliers. A simulation study shows the performance of these proposed procedures and the need to take dependence in the time series into account. Finally, the usefulness of our methodology is illustrated in two real datasets from the electricity market: daily curves of electricity demand and price in mainland Spain, for the year 2012.

1. Introduction

Functional data analysis (FDA) is a branch of Statistics that analyses data providing information about curves, surfaces or any other mathematical object varying over a continuum. The continuum is often time, but it may also be spatial location, wavelength, etc. These curves are defined by a functional form and are called functional data.

Over the last two decades there has been growing research on FDA and most statistical techniques have been generalized to the functional context. This includes linear regression models (Cardot, Ferraty, and Sarda, 1999; Li and Hsing, 2007; García-Portugués, González-Manteiga, and Febrero-Bande, 2014), nonparametric smoothing

Departamento de Matemáticas, Universidade da Coruña, Spain.
juan.vilar@udc.es, paula.rana@udc.es, ganeiros@udc.es

methods (Ferraty and Vieu, 2002; Delsol, Ferraty, and Vieu, 2011; Shang, 2014), classification (Cuevas, Febrero, and Fraiman, 2007; Baíllo, Cuesta-Albertos, and Cuevas, 2011; Sguera, Galeano, and Lillo, 2014), dimension reduction (Boente and Fraiman, 2000; Hall, Müller, and Wang, 2006) and bootstrap methods (González-Manteiga and Martínez-Calvo, 2011; Ferraty, van Keilegom, and Vieu, 2012). In addition, FDA has been successfully applied in a wide range of fields such as climatology (Besse, Cardot, and Stephenson, 2000), chemometrics (Ferraty and Vieu, 2002), environmetrics (Aneiros-Pérez et al., 2004), demography (Hyndman and Ullah, 2007), social sciences (Ocaña, Aguilera, and Escabias, 2007) and the electricity market (Aneiros et al., 2013 and 2016). Of course, the above references form a non-exhaustive list of recent methodological and practical presentations related to FDA. See the monographs by Ramsay and Silverman (2005) and Ferraty and Vieu (2006) for parametric and nonparametric methods, respectively. For a recent state of the art on FDA, see Ferraty and Romain (2011), Horváth and Kokoszka (2012) and Cuevas (2014).

Procedures for detecting functional outliers have also been proposed over recent years despite the fact that the functional nature of the data makes outliers hard to detect. As a matter of fact, a rigorous definition of functional outlier remains to be given. Throughout this paper, we define a functional outlier as an observation (functional datum) that has been generated by a stochastic process with a distribution different from the vast majority of the remaining observations, which are assumed to be identically distributed (note that this is the definition given in Febrero, Galeano, and González-Manteiga, 2008; Hyndman and Shang, 2010). The first papers that have addressed outlier identification in the context of functional data are Hyndman and Ullah (2007) and Febrero, Galeano, and González-Manteiga (2007, 2008). Hyndman and Ullah (2007) proposed a method for robust estimation of functional principal components, which is the basis of their methodology for forecasting functional time series. As a by-product, they constructed a method for detecting outliers based on the integrated squared error between each functional datum and its projection into a given number of robust principal components. The procedure in Febrero, Galeano, and González-Manteiga (2007) (Febrero, Galeano, and González-Manteiga, 2008) performs a distance-based (depth-based) test statistic for each curve, where the critical value is obtained with a bootstrap method. Several procedures for detecting outliers in functional data have been proposed from these works. They are generally based on functional principal components analysis (Hyndman and Shang, 2010; Sawant, Billor, and Shin, 2012; Yu, Zou, and Wang, 2012), functional depths (Sun and Genton, 2011; Gervini, 2012; Arribas-Gil and Romo, 2014) or random projections (Fraiman and Svarc, 2013). All of these papers deal with independent functional data.

This paper addresses the problem of outlier detection in functional time series coming from a real-valued continuous time stochastic process. Specifically, to define the functional time series, $\{\chi_i\}_{i=1}^{n}$, which are going to be used along this paper, we consider a real-valued continuous time stochastic process $\{\chi(t)\}_{t \in R}$. Then, we assume that such process is seasonal with seasonal length τ and we regard that it is observed on the in-

terval $(a,b]$ with $b = a + n\tau$. We define the functional time series $\{\chi_i\}_{i=1}^n$ in terms of $\{\chi(t)\}_{t \in R}$ as:

$$\chi_i(t) = \chi(a + (i-1)\tau + t) \text{ with } t \in [0, \tau].$$

As in the case of finite-dimensional data, dependence affects functional outlier detection (see Raña, Aneiros, and Vilar, 2015). This is clearly illustrated in Figure 1, which shows the sequential graph of a simulated functional time series contaminated with four outliers (left panel) and the corresponding curves (right panel). Looking closely at the left panel in Figure 1, one may suspect the possible presence of such four outliers; however, the same may not be said when observing the right panel in Figure 1. Local trends induced from the dependence structure could mask the presence of outliers; so, in functional time series, an observation could be an outlier despite being inside the range of the vast majority of the data. It therefore seems reasonable to believe that this kind of outlier cannot be detected by applying outlier detection procedures designed for independent data. To the best of our knowledge, the only paper that has addressed outlier detection in functional time series is Raña, Aneiros, and Vilar (2015). These authors suggested adapting the procedure in Febrero, Galeano, and González-Manteiga (2008) to the functional time series setting by considering bootstrap techniques that take into account the dependence between functional data (instead of standard bootstrap).
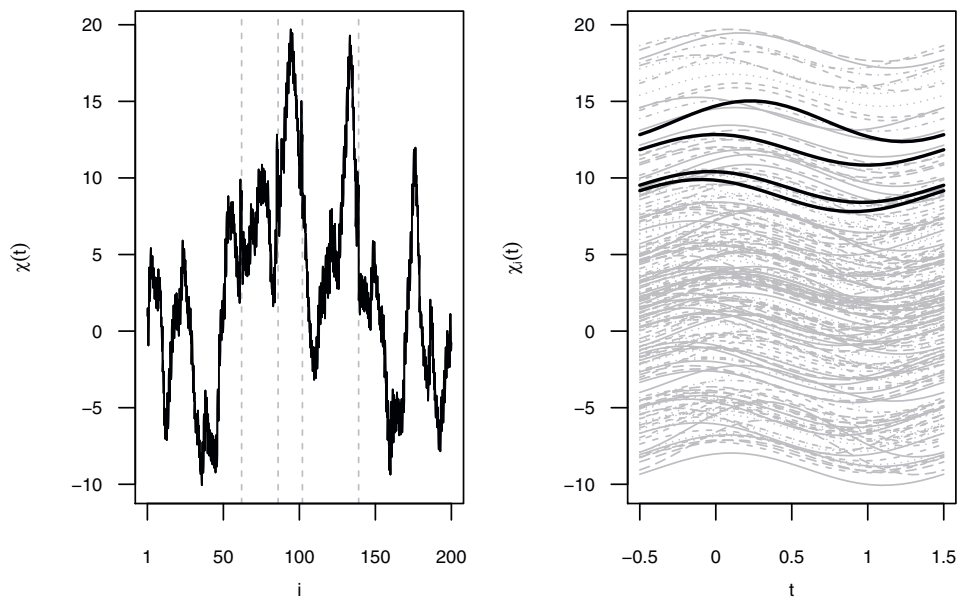

Figure 1: *Left panel: functional time series (i denotes the temporal index) contaminated with four outliers; the vertical dashed lines indicate the positions where the outliers emerged. Right panel: the corresponding curves $\chi_i(t)$ (the black curves are the outliers).*

This article proposes two procedures for detecting outliers in functional time series. Both methods are based on robust functional principal component analysis and use ideas developed by Hyndman and Ullah (2007) to forecast functional time series.

The remainder of this paper is organized as follows. Section 2 presents basic ideas on principal component analysis. The proposed methodology is described in Section 3. Its behavior is evaluated and compared with other approaches through a simulation study in Section 4. In Section 5 our methods are applied to electricity demand and price curves. Finally, Section 6 concludes with a discussion.

2. Functional principal component analysis

Since our proposed procedures for detecting functional outliers will be constructed based on functional principal component analysis (FPCA), this section presents a brief review on this topic. The interested reader can find a more complete review in Hall (2011). Without loss of generality, we assume that the considered functional random variable has zero mean.

Principal component analysis (PCA) is a standard approach to explore variability in multivariate data, $\mathbf{X} \in \mathbb{R}^d$. This approach specifies the d directions, $\{\mathbf{v}_k\}_{k=1}^d \in \mathbb{R}^d$, that maximize the variance along each component, subject to the orthonormal condition. Reducing the dimension is especially important when data belong to infinite dimensional spaces, this being the case of functional data. In this article we focus on curves observed in $[a,b]$ ($-\infty < a < b < \infty$) and square integrable. Then, if χ denotes a functional random variable, the aim of FPCA is to find the functions $\phi_k : [a,b] \to \mathbb{R}$ such that the variance of

$$\beta_k = \int_a^b \phi_k(t)\chi(t)dt \tag{1}$$

is maximized subject to the constraints

$$\int_a^b \phi_k^2(t)dt = 1 \text{ and } \int_a^b \phi_k(t)\phi_j(t)dt = 0 \ (k \neq j). \tag{2}$$

The functional principal components $\phi_k(\cdot)$ can also be defined as the orthonormal functions verifying

$$\int_a^b \mathbf{C}(t,s)\phi_k(s)ds = \lambda_k\phi_k(t) \ (t \in [a,b], \ k = 1,2,\ldots), \tag{3}$$

where $\mathbf{C}(t,s)$ denotes the covariance between $\chi(t)$ and $\chi(s)$. Finally, dimension reduction is performed by considering the approximation

$$\chi(\cdot) \approx \sum_{k=1}^{K} \beta_k \phi_k(\cdot), \tag{4}$$

where $K < \infty$ and $\sum_{k=1}^{K} \lambda_k$ is close to $\sum_{k=1}^{\infty} \lambda_k$ (we have assumed that $\lambda_k > \lambda_{k+1}$, $k = 1, 2, \ldots$). For details, see e.g. Ramsay and Silverman (2005).

Functional principal components, $\phi_k(\cdot)$, depend on the unknown covariance operator $\mathbf{C}(\cdot, \cdot)$. Assuming that one has observations $\{\chi_i\}_{i=1}^{n}$ identically distributed from the functional random variable χ, estimates for $\phi_k(\cdot)$ can be obtained by using

$$\widehat{\mathbf{C}}(t,s) = \frac{1}{n} \sum_{i=1}^{n} (\chi_i(t) - \overline{\chi}(t))(\chi_i(s) - \overline{\chi}(s)), \text{ where } \overline{\chi}(t) = \frac{1}{n} \sum_{i=1}^{n} \chi_i(t),$$

instead of $\mathbf{C}(t,s)$ in (3). See Horváth and Kokoszka (2012) for the consistency of $\widehat{\mathbf{C}}$ and of the corresponding eigenfunctions and eigenvalues, under either independent curves or weakly dependent functional time series.

It is worth noting that, apart being used for dimension reduction, FPCA can also be used as a tool for outlier detection. Nevertheless, as noted in the previous paragraph, the estimation of functional principal components is based on the estimated covariance operator $\widehat{\mathbf{C}}(\cdot, \cdot)$, which is known to be sensitive to outliers. Thus, if the goal is to construct an approach based on principal components to identify functional outliers, robust FPCA should be considered. In this way, Hyndman and Ullah (2007) propose estimating the functional principal components by means of the functions $\widehat{\phi}_k(\cdot)$ that maximize the variance of the scores

$$z_{i,k} = w_i \int_a^b \phi_k(t) \chi_i(t) dt \tag{5}$$

subject to the constraints (2). The weights w_i are computed as

$$w_i = \begin{cases} 1 & \text{if} \quad v_i < S + \lambda\sqrt{S} \\ 0 & \text{otherwise} \end{cases}$$

where

$$v_i = \int_a^b (\chi_i(t) - \sum_{k=1}^{K} \widetilde{\beta}_{i,k} \widetilde{\phi}_k(t))^2 dt \tag{6}$$

with $\widetilde{\phi}_k(\cdot)$ being initial (highly robust) projection-pursuit estimates of $\phi_k(\cdot)$ obtained from the RAPCA algorithm (see Hubert, Rousseeuw, and Verboven, 2002) considering equal weights w_i in (5), while $\widetilde{\beta}_{i,k} = \int_a^b \widetilde{\phi}_k(t) \chi_i(t) dt$. In addition, S is the median of

$\{v_1, \ldots, v_n\}$ and $\lambda > 0$ is a tuning parameter to control the degree of robustness. Once the robust estimates $\widehat{\phi}_k(\cdot)$ are obtained, the coefficients corresponding to the curve χ_i are constructed as

$$\widehat{\beta}_{i,k} = \int_a^b \widehat{\phi}_k(t)\chi_i(t)dt. \tag{7}$$

As a by-product, Hyndman and Ullah (2007) proposed an outlier detection method (the ISE method): the curve χ_i is detected as outlier if $w_i = 0$. For other FPCA-based procedures to identify outliers, see e.g. Hyndman and Shang (2010) and Sawant, Billor, and Shin (2012).

3. Outlier detection in functional time series

As noted in Section 1, the dynamics in the data should be taken into account to detect outliers in functional time series. In other words, methods based only on the set of curves and not on the dependence structure among them, cannot detect the outliers that remain hidden among all of the curves (note that these outliers make sense in time series).

We propose two procedures to detect outliers in functional time series. Both proposals are based on the suggestions of Hyndman and Ullah (2007) for obtaining robust forecasting in functional time series. We establish our methods in the following subsections.

3.1. Method based on projections

Our first proposal detects outliers on the first K robust principal component scores and then map the detected outliers into the functional space.

Specifically, the method based on projections proposes to detect outliers in functional time series with the following algorithm:

- Step 1. Perform robust FPCA and construct the corresponding time series of coefficients $\{(\widehat{\beta}_{i,1}, \ldots, \widehat{\beta}_{i,K})\}_{i=1}^n$.

- Step 2. Identify outliers in the series constructed in Step 1 by means of a time-series outlier detection method.

- Step 3. Establish the set of outliers as $\mathscr{O} = \{\chi_i : i \in \mathscr{I}\}$, where $\mathscr{I} = \{i : (\widehat{\beta}_{i,1}, \ldots, \widehat{\beta}_{i,K})$ was detected as outlier in Step 2$\}$.

The key points in this method are the use of robust FPCA together with procedures to detect outliers in time series. Given that the estimated functional principal components $\widehat{\phi}_k$ are not affected by the outliers, the corresponding projections $\widehat{\beta}_{i,k}$ reflect the main

features of the datum χ_i. Thus, we may expect that if a curve is an outlier, its projection on the directions of maximum variance (the first principal components) will also be an outlier.

In practice, both a robust FPCA and a time-series outlier detection method must be fixed to implement our proposal. On the one hand, the robust FPCA proposed in Hyndman and Ullah (2007) could be considered (for a brief exposition, see last paragraph in Section 2). On the other hand, it is worth being noted that the principal component scores $\widehat{\beta}_{i,k}$ and $\widehat{\beta}_{i,l}$ are uncorrelated for $k \neq l$. Thus, as suggested in Hyndman and Ullah (2007), each univariate time series $\{\widehat{\beta}_{i,k}\}_{i=1}^n$, $k = 1, \ldots, K$, can be studied independently. In this way, we propose to use some univariate time-series outlier detection method to identify outliers in each of such scalar time series, and, in Step 2, consider that $(\widehat{\beta}_{i,1}, \ldots, \widehat{\beta}_{i,K})$ is an outlier if some of its components was detected as outlier in the univariate study (for the univariate time-series outlier detection method based on ARIMA models used in this paper, see Section 11.2 in Cryer and Chan, 2008). Another alternative would be to use a multivariate time-series outlier detection method (see, for instance, Tsay, Peña, and Pankratz, 2000).

3.2. Method based on errors

Unlike the previous method, our second procedure takes the whole of each curve into account. Using techniques for robust forecasting in functional time series, it constructs a non-contaminated version for each curve, which is compared with the corresponding original curve. A curve is considered an outlier if it is substantially different from its uncontaminated version.

Specifically, this method proposes to detect outliers in functional time series with the following algorithm:

- Step 1. Perform robust FPCA and construct the corresponding time series of coefficients $\{(\widehat{\beta}_{i,1}, \ldots, \widehat{\beta}_{i,K})\}_{i=1}^n$.

- Step 2. Fit a robust model to the time series constructed in Step 1.

- Step 3. Obtain the fitted values $\{(\widehat{\beta}_{i,1}^*, \ldots, \widehat{\beta}_{i,K}^*)\}_{i=1}^n$ from the model constructed in Step 2.

- Step 4. Construct the residual curves $\{\chi_i - \widehat{\chi}_i\}_{i=1}^n$ and compute some kind of norm $\{u_i\}_{i=1}^n$ for such curves. We have denoted

$$\widehat{\chi}_i = \sum_{k=1}^K \widehat{\beta}_{i,k}^* \widehat{\phi}_k.$$

- Step 5. Identify "abnormally high values" in $\{u_i\}_{i=1}^n$, and set the functional outliers as $\mathscr{O} = \{\chi_i : i \in \mathscr{J}\}$, where $\mathscr{J} = \{i : u_i \text{ was identified as abnormally high}\}$.

As in the method based on projections, robust FPCA plays a main role (together with robust modelling of nonfunctional time series). Note that, because the fitted values obtained in Step 3 are not contaminated by the outliers, $\widehat{\chi}_i$ can be seen as the "expected value" of the functional time series at time i when no contamination is present. Thus, an abnormally high value for u_i suggests that χ_i is an outlier.

Note that our proposal can be seen as an extension in different ways of the ISE method proposed in Hyndman and Ullah (2007) (for a brief exposition, see last paragraph in Section 2). Clearly, our main contribution is related to the dependence in the functional time series: our procedure takes the dependence among the sample into account (see Step 3) to construct the coefficients associated to each functional data χ_i ($\widehat{\beta}_{i,k}^*$ in Step 4 above), while the method in Hyndman and Ullah (2007) does not do so (see $\widetilde{\beta}_{i,k}$ in (6)). As it will be clearly shown in the simulation study to be presented in the next Section 4, this seemingly minor modification will greatly improve the performance of the method when applied on functional time series.

In practice, several choices must be done to implement our algorithm. As in the method based on projections, we suggest to consider the robust FPCA proposed in Hyndman and Ullah (2007) and construct univariate models instead of multivariate ones (see Section 3.1). Specifically, we suggest to fit, for each series $\{\widehat{\beta}_{1,k}, \ldots, \widehat{\beta}_{n,k}\}$, $k = 1, \ldots, K$, the univariate robust ARIMA models studied in Cryer and Chan (2008) (for details, see Section 11.2 in the cited reference). As for the norm to be used to construct the set $\{u_i\}_{i=1}^n$ in Step 4, one might consider, for instance, the L_1-norm or the L_2-norm (or even the squared of the L_2-norm, as in Hyndman and Ullah, 2007). Finally, we suggest to consider that u_i is high enough to be considered as abnormally high if $u_i > q_{0.75} + 1.32(q_{0.75} - q_{0.25})\}$ (q_p denotes the quantile of order p of $\{u_1, \ldots, u_n\}$). Actually, this is the rule given by the classical boxplot; that is, under normality, the probability of detecting no outliers is 0.993, when no outliers are actually present (note that the usual constant factor 1.5 was changed to 1.32 because low values are not considered outliers).

3.3. Tuning parameter

As common to all FDA procedures using FPCA, the proposed methods depend on the number of principal components considered, K. In practice, the value of K must be specified. Hyndman and Ullah (2007) suggest choosing K to minimize the integrated squared forecast error (ISFE), while Hyndman and Booth (2008) find that the forecasts are insensitive to the choice of K, provided K is large enough. Then, Hyndman and Booth (2008) recommend using a value K that is apparently larger than actually required by the components. In the cited works of Hyndman and Ullah (2007) and Hyndman and Booth (2008), and also in Liebl (2013), in the study of different applications using this technique they use a value of K which explains, at least, 98% of the variability.

We have carried out sensitivity studies for the values of K in our methods, using the dependent simulated data considered in the next section (Models 1, 2 and 3). On the

one hand, our findings agree with the general suggestion given in Hyndman and Booth (2008): to consider a larger than necessary value K (for instance, a value explaining at least 98% of the variability). On the other hand, to detect "shape outliers" (that arise when they are within the range of the rest of the data but differ from them in shape; see Hyndman and Shang, 2010) by means of the method based on projections (PB), the recommendation is to select a value K even higher (for instance, explaining at least the 99.9% of variability). To justify this very high value we argue that (i) the PB method only uses scores (and not the whole of the curve), (ii) the first scores inform about the possible presence of "magnitude outliers" (that arise when they lie outside the range of the majority of the data; see Hyndman and Shang, 2010) and (iii) the scores of higher order inform about the possible presence of shape outliers.

4. Simulation study

A simulation study was conducted to compare the performance of our methods with other methods available in the statistical literature.

On the one hand, three main models were constructed to generate functional time series. They are the superposition of a deterministic signal and random noise. Noise in main Models 1, 2, and 3 was the superposition of a scalar AR(1) process and functional AR(1)-, MA(1)- and ARMA(1,1)-type processes, respectively. On the other hand, another main model (Main Model 0) was constructed in the same way, but considering independent noise instead of dependent one. Note that main Models 1, 2 and 3 are favourable to methods that take dependence in the sample into account, while Main Model 0 is favourable to methods designed for independent data. From each main model, two contaminated models were constructed by randomly adding either three magnitude outliers or three shape outliers.

More specifically, we considered the following main models:

- Main Model 0:

$$\zeta_i(t) = \cos(\pi t)(1 - c) + a_i(t) \quad \text{if} \quad -n + 1 \leq i \leq n.$$

- Main Model 1:

$$\zeta_i(t) = \begin{cases} \cos(\pi t) & \text{if} \quad i = -n + 1 \\ \cos(\pi t)(1 - c) + \rho \zeta_{i-1}(t) + a_i(t) + b_i & \text{if} \quad -n + 1 < i \leq n. \end{cases}$$

- Main Model 2:

$$\zeta_i(t) = \cos(\pi t)(1 - c) + \theta a_{i-1}(t) + a_i(t) + b_i \quad \text{if} \quad -n + 1 \leq i \leq n.$$

- Main Model 3:

$$\zeta_i(t) = \begin{cases} \cos(\pi t) & \text{if} \quad i = -n+1 \\ \cos(\pi t)(1-c) + \rho\zeta_{i-1}(t) + \theta a_{i-1}(t) + a_i(t) + b_i & \text{if} \quad -n+1 < i \leq n. \end{cases}$$

In the processes above we have denoted $a_i(t) = X_i \sin(\pi t)$ with X_i being i.i.d. Gaussian variables with mean 0 and standard deviation 0.3, while $\{b_i\}$ is a scalar Gaussian AR(1) process with correlation coefficient $d = 0.8$ and standard deviation $(1-d^2)^{-1/2}$. $c = 0.8$ and $t \in [-0.5, 1.5]$ were considered.

Values ρ and θ manage the dependence strength in the functional time series. We consider two options, one with low dependence (LD, $\rho = 0.5$ and $\theta = -0.5$) and other with high dependence (HD, $\rho = 0.8$ and $\theta = 0.8$).

Then, given each main model, ζ_i, methods were applied on the following contaminated models to detect outliers:

- Contaminated model with magnitude outliers:

$$\chi_i(t) = \zeta_i(t) + k 1_{\{i \in \{I_j\}\}}, \ 1 \leq i \leq n.$$

- Contaminated model with shape outliers:

$$\chi_i(t) = \zeta_i(t) + k \cos(3\pi t) 1_{\{i \in \{I_j\}\}}, \ 1 \leq i \leq n.$$

Note that k is a contamination size while $1_{\{.\}}$ and I_j denote the indicator function and i.i.d. random variables with discrete uniform distribution on $\{1, \ldots, n\}$, respectively. The curves χ_i were discretized on a grid $\{t_j\}$ of 30 equispaced points in $[-0.5, 1.5]$. Note also that in the simulation process we generate curves corresponding to the double of the sample size n. That is, we simulate the curves $\{\zeta_i(t)\}$, where $-n+1 \leq i \leq n$, but we use only the last half of the curves, $\{\zeta_i(t) : 1 \leq i \leq n\}$, for the contaminated models. The first n realizations are not used in order to avoid the impact of the initial values. The number of outliers introduced in the models was $j = 0.02n$ (that is, 2% of the curves). Value of k was 0.75 for Contaminated Model 0, in which dependence does not affect, and 5 for contaminated Models 1, 2 and 3. It is worth noting here that the contamination size, k, considered in this study is low compared with other existing simulation studies (see, e.g. Sun and Genton, 2011).

Figure 2 shows curves simulated from these four contaminated models. First row corresponds to the Model 0 (no dependence), and the other three rows to the Models 1, 2 and 3 (functional time series), respectively. Last three models are shown for the HD case. We can see in the figure the difference between the data simulated from Model 0 and from Models 1, 2 and 3: in the case of functional time series, outliers are almost always hidden within the rest of the curves.
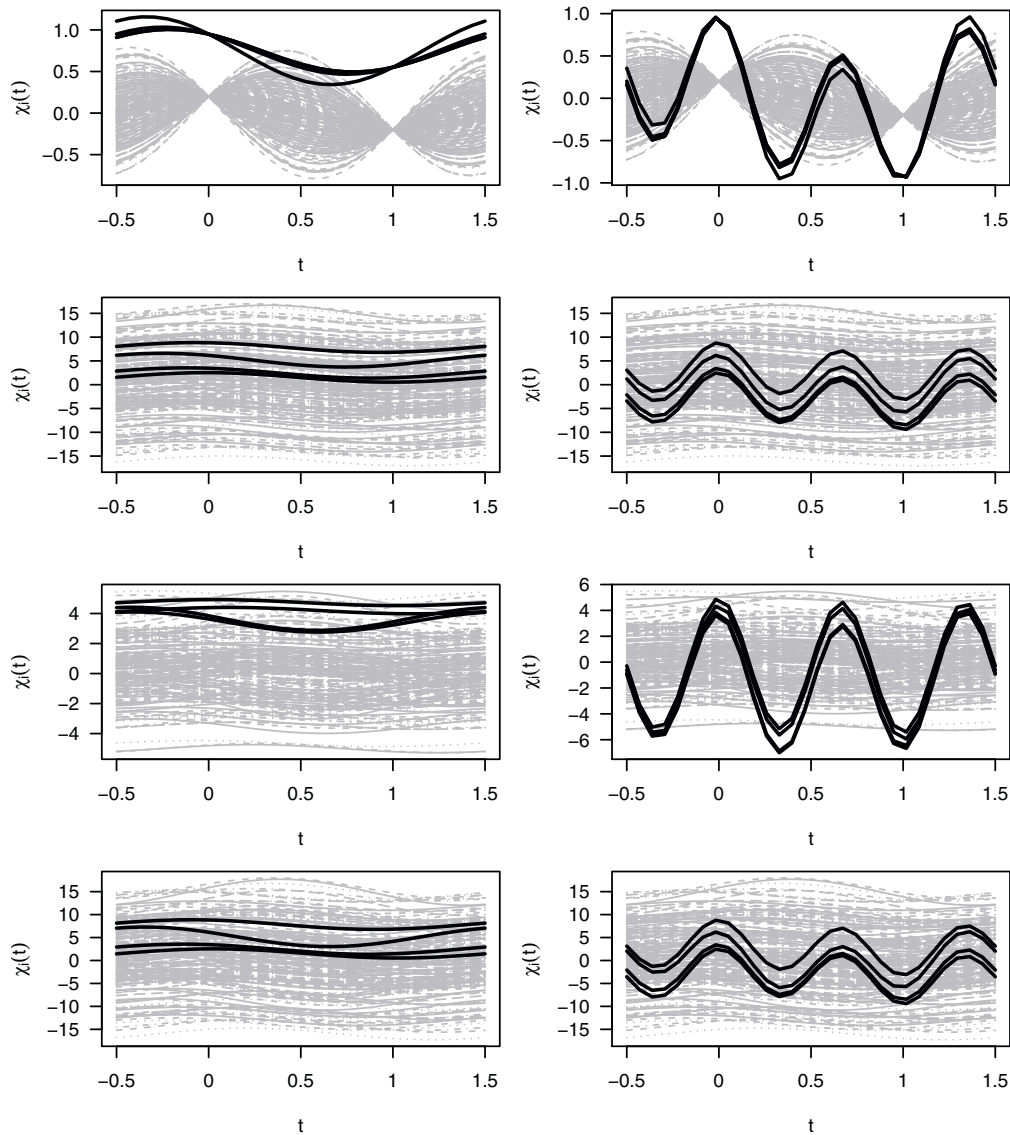
Figure 2: *Left panels: from top to bottom, curves* $(\chi_i(t))$ *generated from contaminated Models 0, 1, 2 and 3, respectively (the black curves are magnitude outliers). Right panels: from top to bottom, curves* $(\chi_i(t))$ *generated from contaminated Models 0, 1, 2 and 3, respectively (the black curves are shape outliers).*

We applied the following four procedures on each generated sample in addition to the proposed projections-based (PB) and errors-based (EB) methods.

- Functional highest density region boxplot (HDR). This graphical method is based on the bivariate HDR boxplot (Hyndman, 1996) applied to the first two robust principal component scores (for details, see Hyndman and Shang 2010). The coverage probability $(1 - \alpha_0)$ of the outer region must be prespecified.

- Integrated squared error (ISE). This is the method proposed in Hyndman and Ullah (2007) (for a brief exposition, see last paragraph in Section 2). Both the parameter that controls the degree of robustness (λ) and the quantity of principal components (K) need to be prespecified.

- Depth-based trimming (DBT). Performs a test statistic. A curve is considered an outlier if its depth is lower than a cutoff. The cutoff is determined by a bootstrap method based on trimming the sample (for details, see Febrero, Galeano, and González-Manteiga, 2008). The functional depth, the signification level (α_1), the proportion of potential outliers (α_2), the parameter used to obtain smoothed bootstrap samples (γ) and the number of bootstrap samples (B) must be prespecified.

- DBT for dependent data. Adapts the DBT procedure to the functional time series setting. For that, to determine the cutoff, bootstrap techniques for dependent data are used instead of standard bootstrap. In addition, the functional boxplot of Sun and Genton (2011) is used to trim the sample in the first stage (for details, see Raña, Aneiros, and Vilar, 2015). The functional depth, the signification level (α_1), the bootstrap technique and the number of bootstrap samples (B) need to be prespecified.

Note that the methods HDR, ISE and DBT are designed to detect outliers in samples of independent curves, even if they were also applied to functional time series. Meanwhile, DBT-MBB, PB and EB are specifically address to deal with the problem of outlier detection in the context of functional time series. Along this simulation study, we will compare the performance of the cited methods in situations of both independent and dependent data.

$M = 500$ simulations were run for each model. The percentage of correctly identified outliers p_c (100 times the number of correctly identified outliers over the number of outliers in the sample, or sensitivity) and the percentage of false positives p_f (100 times the number of wrongly identified outliers over the number of non-outlying curves in the sample, or false detection percentage) were computed for each simulation and for each method considered.

Routines fboxplot and foutliers, available in the R package rainbow, were used to detect outliers from the HDR and ISE procedures, respectively. $\alpha_0 = 0.01$ was considered in the HDR method while values for λ and K in the ISE approach were chosen following the suggestion given in Hyndman and Ullah (2007): $\lambda = 3.29$ and K being the value minimizing the ISFE. The method DBT is implemented in the routine Outliers.fdata, available in the R package fda.usc. We considered $\alpha_1 = \alpha_2 = 0.01$, $\gamma = 0.05$ and $B = 200$, while the functional depth was the h-modal depth (Cuevas, Febrero, and Fraiman, 2006), as recommended in Febrero, Galeano, and González-Manteiga (2008). Moving blocks bootstrap (Künsch, 1989) was used in the DBT-for-dependent-data procedure (DBT-MBB) while α_1 and B took the same values as in the DBT method. The functional h-modal depth was also considered in this case.

The choices related to the procedures PB and EB were done following the recommendations given in Section 3.1 and 3.2, respectively. The tuning parameter K was chosen as suggested in Section 3.3. Specifically, we chose $K = 1$ for magnitude outliers (for both methods PB and EB). In the case of shape outliers, we chose $K = 3$ for the PB method and $K = 1$ for the EB. This election explains more than 98% of the variability (in some cases, even with only the first component, it explains around 99.5%), increasing until 99.9% when we use PB method to detect shape outliers. This choice agrees with the guidelines given in Section 3.3 about the requirement of more components when dealing with shape outlier detection and the PB method. In the case of Model 0, due to the simplest performance of the data, it is enough to take $K = 1$ for the EB and $K = 2$ for the PB method and both kind of outliers. The signification level used to detect scalar outliers in the PB method (Step 2) was $\alpha_3 = 0.01$. In the case of the norm to be used in Step 4 of the EB procedure, both the L_1-norm and the L_2-norm were considered. Because similar results were obtained, we only show the corresponding ones to the L_2-norm.

Our first simulation study employs $n = 200$ and the results are reported in tables 1, 2 and 3. These tables show the mean and standard deviation of the values of both p_c and p_f obtained from the two proposed procedures (PB and EB) and the other four considered methods (HDR, ISE, DBT and DBT-MBB) when they are applied to the different contaminated models. In Table 1 the Model 0 is considered (independent data), which is contaminated with magnitude or shape outliers. In Tables 2 and 3, the Models 1, 2 and 3 (dependent data) and the two cases of dependence (low and high dependence) are considered (see Table 2 for contamination with magnitude outliers and Table 3 for the case of shape outliers).

Table 1: *Mean and standard deviation (in parentheses) of the percentage of correctly and falsely identified outliers in Model 0 contaminated with magnitude or shape outliers.*

Model 0				
	Magnitude outliers		Shape outliers	
Method	$\widehat{p}_c$	$\widehat{p}_f$	$\widehat{p}_c$	$\widehat{p}_f$
HDR	40.60 (14.21)	0.19 (0.29)	40.50 (14.24)	0.19 (0.29)
ISE	100.00 (0.00)	0.00 (0.02)	100.00 (0.00)	0.00 (0.00)
DBT	87.00 (26.12)	0.64 (0.48)	84.15 (27.46)	0.61 (0.47)
DBT-MBB	99.80 (4.47)	2.83 (1.44)	99.80 (4.47)	2.83 (1.44)
PB	95.40 (10.44)	0.02 (0.10)	95.15 (10.98)	0.02 (0.10)
EB	100.00 (0.00)	2.14 (1.23)	95.75 (15.61)	2.10 (1.22)

Table 2: *Mean and standard deviation (in parentheses) of the percentage of correctly and falsely identified outliers in Models 1, 2 and 3 (with low or high dependence) contaminated with magnitude outliers.*

Model 1

Method	Low dependence		High dependence	
	$\widehat{p}_c$	$\widehat{p}_f$	$\widehat{p}_c$	$\widehat{p}_f$
HDR	16.50 (15.91)	0.68 (0.32)	9.85 (14.04)	0.82 (0.29)
ISE	22.00 (22.01)	15.33 (2.69)	25.25 (22.74)	21.07 (3.65)
DBT	26.30 (23.28)	1.19 (0.87)	10.55 (15.67)	1.14 (1.17)
DBT-MBB	30.45 (24.04)	2.07 (1.78)	13.10 (16.93)	2.38 (2.24)
PB	70.65 (35.95)	0.31 (0.45)	62.05 (38.65)	0.56 (0.59)
EB	88.55 (17.12)	3.71 (1.24)	84.10 (20.81)	4.07 (1.26)

Model 2

Method	Low dependence		High dependence	
	$\widehat{p}_c$	$\widehat{p}_f$	$\widehat{p}_c$	$\widehat{p}_f$
HDR	28.25 (16.92)	0.44 (0.35)	30.25 (17.08)	0.40 (0.35)
ISE	24.30 (23.73)	14.60 (2.51)	27.00 (23.39)	16.64 (2.79)
DBT	66.75 (25.49)	0.81 (0.69)	67.60 (24.83)	0.76 (0.69)
DBT-MBB	73.15 (24.58)	1.60 (1.17)	73.40 (24.00)	1.55 (1.24)
PB	67.60 (37.26)	0.07 (0.20)	68.40 (37.24)	0.07 (0.18)
EB	91.60 (14.83)	3.17 (1.29)	91.60 (15.08)	3.23 (1.28)

Model 3

Method	Low dependence		High dependence	
	$\widehat{p}_c$	$\widehat{p}_f$	$\widehat{p}_c$	$\widehat{p}_f$
HDR	14.70 (15.31)	0.72 (0.31)	10.75 (14.27)	0.80 (0.29)
ISE	19.60 (21.78)	13.31 (2.72)	30.65 (24.04)	28.96 (3.40)
DBT	26.30 (23.55)	1.20 (0.86)	10.65 (15.44)	1.13 (1.15)
DBT-MBB	30.30 (24.35)	2.08 (1.67)	12.85 (16.78)	2.35 (2.03)
PB	69.30 (36.30)	0.33 (0.47)	60.95 (38.36)	0.61 (0.64)
EB	88.45 (16.76)	3.63 (1.24)	84.20 (20.28)	3.81 (1.17)

Table 3: *Mean and standard deviation (in parentheses) of the percentage of correctly and falsely identified outliers in Models 1, 2 and 3 (with low or high dependence) contaminated with shape outliers.*

Model 1

Method	Low dependence		High dependence	
	$\widehat{p}_c$	$\widehat{p}_f$	$\widehat{p}_c$	$\widehat{p}_f$
HDR	16.15 (19.20)	0.69 (0.39)	13.90 (18.77)	0.74 (0.38)
ISE	100.00 (0.00)	14.58 (2.74)	100.00 (0.00)	20.15 (3.59)
DBT	95.75 (16.47)	0.21 (0.41)	64.25 (38.55)	0.61 (0.91)
DBT-MBB	99.40 (6.69)	0.70 (1.16)	56.05 (37.45)	0.58 (1.06)
PB	95.20 (10.71)	0.04 (0.17)	95.00 (11.19)	0.04 (0.24)
EB	100.00 (0.00)	2.58 (1.26)	100.00 (0.00)	2.59 (1.29)

Model 2

Method	Low dependence		High dependence	
	$\widehat{p}_c$	$\widehat{p}_f$	$\widehat{p}_c$	$\widehat{p}_f$
HDR	10.75 (15.20)	0.80 (0.31)	9.10 (14.23)	0.83 (0.23)
ISE	100.00 (0.00)	13.85 (2.51)	100.00 (0.00)	15.95 (2.67)
DBT	96.40 (15.57)	0.38 (0.45)	98.30 (11.06)	0.33 (0.42)
DBT-MBB	100.00 (0.00)	1.81 (1.36)	100.00 (0.00)	1.67 (1.35)
PB	95.20 (10.71)	0.03 (0.16)	95.15 (10.63)	0.04 (0.16)
EB	100.00 (0.00)	2.61 (1.38)	100.00 (0.00)	2.64 (1.34)

Model 3

Method	Low dependence		High dependence	
	$\widehat{p}_c$	$\widehat{p}_f$	$\widehat{p}_c$	$\widehat{p}_f$
HDR	17.35 (19.26)	0.67 (0.39)	4.65 (11.40)	0.93 (0.23)
ISE	100.00 (0.00)	12.52 (2.62)	100.00 (0.00)	27.86 (3.23)
DBT	94.40 (19.15)	0.22 (0.42)	49.60 (38.54)	0.79 (1.04)
DBT-MBB	99.40 (6.69)	0.47 (1.23)	41.80 (34.50)	0.94 (1.32)
PB	95.05 (10.82)	0.05 (0.20)	94.95 (11.45)	0.14 (0.31)
EB	100.00 (0.00)	2.51 (1.26)	100.00 (0.00)	2.06 (1.19)

Several conclusions can be drawn from these results. First of all, we look at Contaminated Model 0 in Table 1, which considers independent data. Under that situation, ISE method gets the best result for both kind of outliers. On the contrary, HDR presents poor results with the lowest sensitivity, but also its false detection rate is low. Looking at the pair of DBT and DBT-MBB method, we can see an improvement with the second option, even if dependence is not affecting this data. p_c is much better for the DBT-MBB method, compared to the DBT, but also the p_f is higher. Note that DBT-MBB not only adapts DBT to work with functional time series (by taking dependence into account), but also improves the method itself by changing some other aspects. This is why we can see different results even when they are applied to independent data. Our both proposals, PB and EB, are very competitive in this situation, even compared to methods designed to work with independent data. They maintain high and low values for p_c and p_f, respectively. Their sensitivity is greater than 95% and there is no big difference between magnitude and shape outliers. We can see that PB detects less outliers than EB but also its false detection rate is lower.

Now, we focus on the simulated models that include dependence structure; that is, contaminated Models 1, 2 and 3. The role of this analysis is two-fold: to illustrate the performance of the two proposed procedures and to show the need to take into account the dependence in the functional time series. We restrict first to the magnitude outliers under both situations of low dependence (LD) and high dependence (HD), which results are given in Table 2. In general we do not observe major differences in the behaviour of the proposed methods (PB and EB) when the dependence scenario changes (LD or HD), and we can see that the best results are achieved by the methods that take into account dependence (DBT-MBB, PB and EB). Results are analysed below in a deeper way. HDR and ISE methods lose their effectiveness in detecting outliers when dealing with dependent data. We may highlight the large p_f (around 20%) of the ISE method, indicating a high volatility in its behaviour. We look now at the pair DBT and DBT-MBB methods (remember that DBT-MBB adapts DBT to work with functional time series). It is true that DBT-MBB gets always higher p_c, which clearly indicates that taking dependence in the data into account is outstanding. Both methods are also better than HDR and ISE in most of the cases. Despite of getting worse p_c than ISE when dealing with Models 1 and 3 under high dependence, they get significantly lower p_f. Both methods (DBT and DBT-MBB) also show a sharp difference between dependence scenarios for the Models 1 and 3, in which the outlier detection rate decreases as the dependence structure becomes more relevant.

All the methods analysed above are overcome by our two proposals PB and EB. Both options achieve high sensitivity, greater than DBT-MBB (excepting Model 2 in which DBT-MBB overcomes PB) and far away from the other considered methods that not take into account dependence. PB method holds lower sensitivity than EB, but also lower false detection rate. To obtain a trade-off between high sensitivity and low false detection rate, in general, the proposed EB seems to be a good choice for magnitude outlier detection under the considered dependence scenarios.

Table 3 shows the results when the models are contaminated with shape outliers. HDR still performance very similar to the magnitude outliers case, however ISE methods shows an improvement by detecting all the shape outliers (at the expense of a large false detection rate). DBT and DBT-MBB behaves also similarly to the magnitude outliers case, with a remarkable difference in the levels of p_c. They achieve now very high sensitivity with low dependence (around $95 - 100\%$) but under high dependence they provide low values, around $40 - 60\%$ for Models 1 and 3.

Proposed methods PB and EB show high sensitivity (95% and 100%, respectively) and low false detection rate (0.05% and 2.5%), being very stable for the three simulated models. As in Table 2 for magnitude outliers, also with shape outliers there is no major differences between both dependence scenarios (LD and HD). In summary, even if both methods obtain very good results for shape outlier detection under dependence, EB seems to be a better choice due its great success detecting all the outliers.

A second simulation study is developed in order to study the influence of the sample size (n) over the analysed methods for outlier detection. In this case we restrict to Models 1, 2 and 3 (simulated functional time series) contaminated with magnitude outliers. Table 4 shows the mean of the percentage of correctly and falsely identified outliers (p_c and p_f, respectively) when the sample size varies within the values $n = 100, 200, 300$ and 400. These results are obtained under the scenario of high dependence (HD) and the number of outliers introduced in each sample follows the same rule as the previous results (including $j = 0.02n$ outliers; that is, 2% of the curves).

Results given by the two proposed methods (PB and EB) in Table 4 clearly overcome the rest of the methods included in the comparison (HDR, ISE, DBT and DBT-MBB) in almost all the situations (except when one considers $n = 100$ in Model 2). That is, for the three contaminated models and the different values of the sample size n (except the combination Model 2, $n = 100$), PB and EB get the best performing. On the one hand, HDR, ISE, DBT and DBT-MBB show poor results with very low p_c and also, in the case of ISE method, very high false detection rate. DBT-MBB gets always better results than DBT, showing again the importance of taking dependence in the data into account. Actually, both DBT and DBT-MBB are very competitive for the Model 2, specially with the lowest sample size $n = 100$, but they are overcome by PB and EB as n increases.

On the other hand, also HDR, ISE, DBT and DBT-MBB remains stable when the sample size varies. Indeed, we can see a slight decrease in the p_c and increase in p_f as long as the sample size n increases. On the opposite, for the three contaminated models, proposed methods PB and EB clearly improve the sensitivity (p_c) meanwhile the false detection rate (p_f) decreases slightly as n increases. The reason for this is that PB and EB methods are based on fitting univariate time series (of the coefficients given by FPCA) as a previous step to the outlier detection. Therefore, by increasing the sample size n the fit of the univariate time series is improved and, accordingly, also the outlier detection with PB and EB methods improves.

Table 4: *Mean of the percentage of correctly and falsely identified outliers in Models 1, 2 and 3, with high dependence, contaminated with magnitude outliers and for n = 100, 200, 300 and 400.*

Model 1

	n=100		n=200		n=300		n=400	
Method	$\widehat{p}_c$	$\widehat{p}_f$	$\widehat{p}_c$	$\widehat{p}_f$	$\widehat{p}_c$	$\widehat{p}_f$	$\widehat{p}_c$	$\widehat{p}_f$
HDR	12.30	0.77	9.85	0.82	7.43	0.87	7.28	0.87
ISE	26.10	22.44	25.25	21.07	22.57	21.10	23.13	20.94
DBT	12.10	0.34	10.55	1.14	9.23	1.60	9.63	1.86
DBT-MBB	12.80	1.13	13.10	2.38	11.80	2.99	12.03	2.98
PB	29.20	0.56	62.05	0.56	80.23	0.42	87.90	0.28
EB	58.80	4.52	84.10	4.07	90.83	3.99	93.60	4.00

Model 2

	n=100		n=200		n=300		n=400	
Method	$\widehat{p}_c$	$\widehat{p}_f$	$\widehat{p}_c$	$\widehat{p}_f$	$\widehat{p}_c$	$\widehat{p}_f$	$\widehat{p}_c$	$\widehat{p}_f$
HDR	34.20	0.33	30.25	0.40	28.57	0.44	26.50	0.48
ISE	32.60	17.10	27.00	16.64	22.50	16.51	21.50	16.45
DBT	63.70	0.47	67.60	0.76	68.63	0.90	68.87	1.01
DBT-MBB	73.60	1.66	73.40	1.55	72.30	1.52	72.08	1.54
PB	34.50	0.10	68.40	0.07	84.93	0.04	89.43	0.04
EB	73.70	3.60	91.60	3.23	95.70	3.35	96.25	3.27

Model 3

	n=100		n=200		n=300		n=400	
Method	$\widehat{p}_c$	$\widehat{p}_f$	$\widehat{p}_c$	$\widehat{p}_f$	$\widehat{p}_c$	$\widehat{p}_f$	$\widehat{p}_c$	$\widehat{p}_f$
HDR	15.80	0.70	10.75	0.80	8.70	0.84	8.40	0.85
ISE	31.90	29.08	30.65	28.96	30.10	28.74	29.80	28.84
DBT	12.00	0.32	10.65	1.13	9.37	1.61	9.58	1.85
DBT-MBB	12.70	1.18	12.85	2.35	11.17	2.78	11.40	2.77
PB	29.80	0.60	60.95	0.61	81.10	0.40	87.00	0.29
EB	54.90	3.68	84.20	3.81	90.67	3.78	93.18	3.81

5. Applications in the electricity market

Nowadays, in many countries all over the world, the production and sale of electricity is traded under competitive rules in free markets. The agents involved in this market (namely, system operators, regulatory agencies, producers and consumers) are greatly interested in the study of electricity load and price. Since electricity cannot be stored, the demand must be satisfied instantaneously and producers need to anticipate future demands to avoid overproduction. So good forecasting of electricity demand is very important for the agents in the market. On the other hand, if reliable predictions of electricity price are available to producers and consumers, they can develop their bidding strategies and establish a pool-bidding technique to achieve a maximum benefit. Consequently, the prediction of electricity demand and price pose significant concerns to this sector. In recent years, these concerns have been addressed from a functional perspective. Regression models with functional covariates (and even functional response) have been used to forecast electricity demand and price. Some related papers are Antoniadis, Paparoditis, and Sapatinas (2006), Antoch et al. (2010), Vilar, Cao, and Aneiros (2012), Cho et al. (2013), Lielb (2013) and Aneiros et al. (2016). It is well known that the presence of outliers affects the accuracy of forecasts obtained from regression models. Thus, outlier detection represents a first step in any descriptive analysis of a dataset, prior to any type of modelling or prediction method. In that sense, depending on the objective of the study, one of the following strategies can be used once the outliers are identified: if outliers come from gross errors, they are subsequently removed from the sample. In another case, robust prediction methods or complex models that take into account the existence of outliers (for instance, introducing dummy variables) can be used.

In the next two sections, the proposed methods for detecting outliers in functional time series, projections-based (PB) and errors-based (EB) methods, are applied on daily curves of electricity demand and price. The corresponding tuning parameters were selected in a similar way as in the simulation study.

5.1. Case study: electricity demand

We are interested in outlier detection in time series of electricity demand curves. Data collect hourly electricity demand in the Spanish mainland electricity market on Mondays, ..., Fridays in the year 2012. They are available at http://www.omie.es, the official website of Operador del Mercado Ibérico de Energía. These hourly data present a trend. Thus, by subtracting the trend (estimated by means of a kernel regression) we obtained the corresponding detrended hourly series. The functional dataset under analysis is composed of the $n = 261$ daily demand curves obtained from this detrended hourly series, measured in Megawatt-hour (MWh). The quantity of functional principal components considered was $K = 9$. These K principal components explained, at least, 98% of the variance.
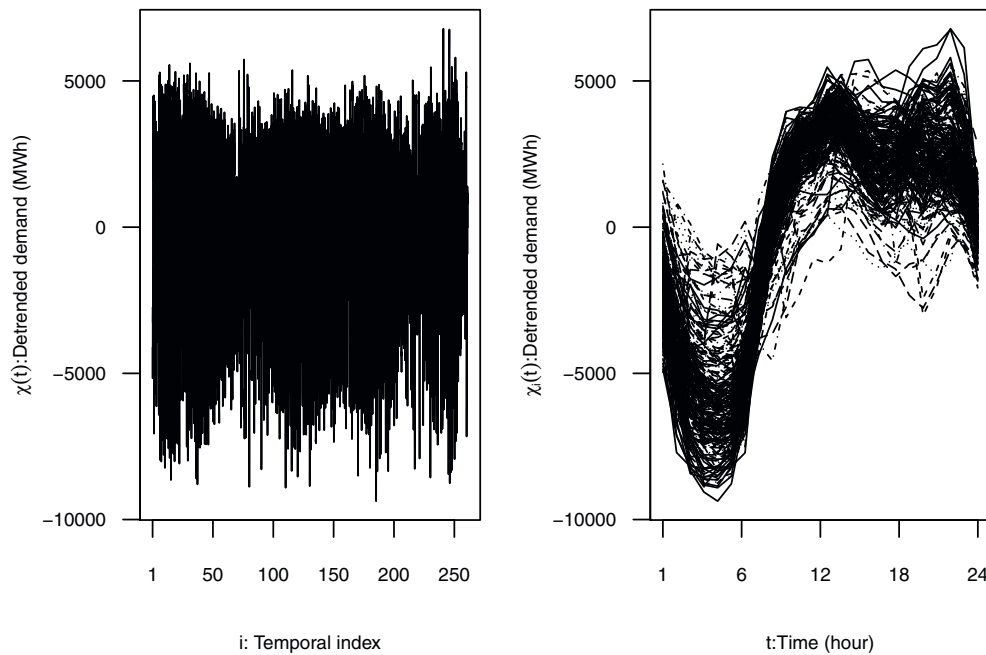
Figure 3: *Left panel: time series of electricity demand. Right panel: daily electricity demand curves.*

The functional time series and the corresponding daily curves are shown in Figure 3. Higher demands are observed in the interval 10:00h–22:00h while lower ones correspond to the interval 3:00h–5:00h.

The outliers identified from the proposed procedures PB and EB are listed in Table 5. Methods PB and EB detect 11 and 15 outliers, respectively, resulting in 20 different curves.

We can find possible causes for most of these abnormal curves. For example, April 16 and 19, November 1 and 2 and also December 24 correspond to days with zero price hours. During some hours in these days, the overproduction of wind power decreases the electricity price fixed by daily market. This have to do with the different taxations of this "green energies" because, as the wind power production increases, the electricity price decreases. As a result, if the wind power production covers and abnormally high percentage of the electricity demand, the price can drop even until zero during a period of time (this being the case of the cited days). We find also as outliers some previous or posterior days to these "zero price days", such as April 24 and 26, which are also affected by the disturbance in the price. Some of the outliers correspond to nonworking days in which the people usually behaves in a different way than the rest of the regular days (simply because most of the economical and industrial activities stop during these days), affecting the electrical consumption and, as a consequence, also the demand. This is the case of May 1 (Labour Day), August 15 (Assumption Day), October 12

Table 5: *Outliers detected in the demand data from proposed procedures PB and EB.*

Day	Method		Day	Method	
	PB	EB		PB	EB
February 14	X		November 2		X
April 16	X	X	November 7		X
April 19	X	X	November 14	X	X
April 24		X	December 3	X	
April 26	X		December 6	X	X
May 1	X	X	December 10		X
May 4	X		December 21		X
August 15	X	X	December 24	X	
October 12		X	December 25		X
November 1		X	December 28		X

(National holiday in Spain), November 1 (All Saints Day), December 6 (Constitution Day in Spain) and 25 (Christmas). December 24 (Christmas Eve) is also a special day, even if it is not officially a holiday. Friday, November 2, besides being a zero-price day, is situated also in the middle of a long weekend caused by All Saints Day, in which a lot of people take some holidays. Finally, November 14 was a strike day in Spain, which clearly affects electrical consumption as it can be considered in some sense as a holiday.

Finally, it is worth pointing out that electricity demand curves observed at days April 16, November 14 and December 6 are detected as outliers simultaneously with the two proposed methods, but no one of these curves is identified as an outlier from either the HDR or DBT procedures (remember that neither HDR nor DBT take dependence in the data into account). Actually, as can be seen in Figure 4, these three curves have features that can, to say the least, be considered suspicious: demand curve observed at April 16 takes high values throughout the first hours (possibly because the electricity price at 3:00h–6:00h was zero); demand curve corresponding to November 14 (strike day) maintains low values from 7:00h, this being the typical behaviour of demand curves corresponding to nonworking days; December 6 is a holiday.

5.2. Case study: electricity price

A similar study is conducted in this section for electricity price. Prices were available for the same period as demands, and they were obtained from the same source. The units were cents (euro) per kilowatt-hour (cents/kWh). Unlike the previous case, there was no
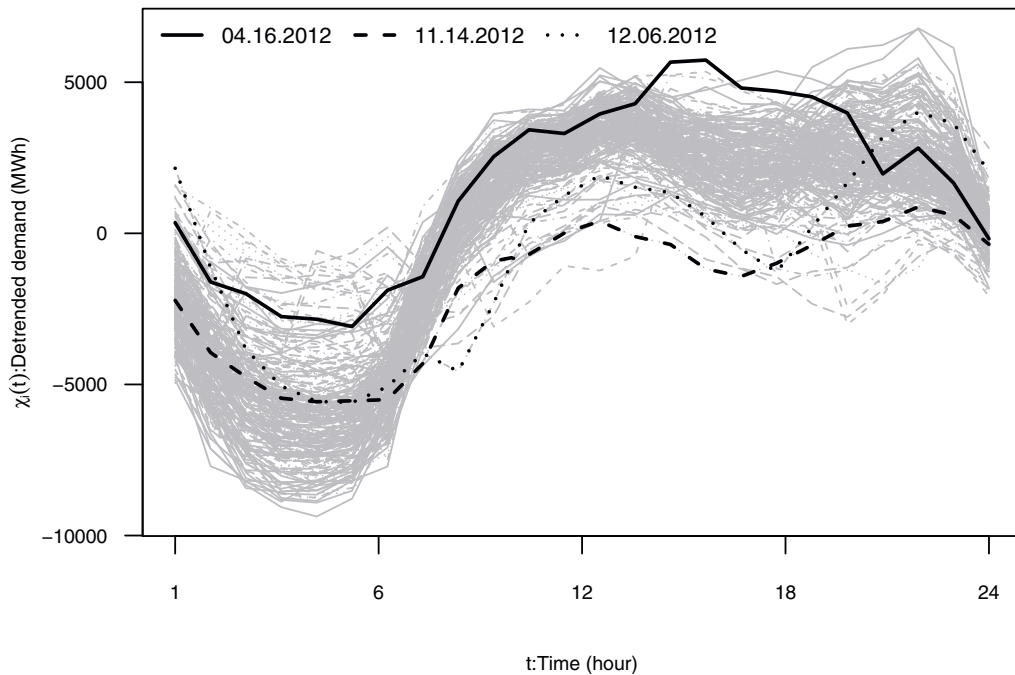
Figure 4: *Outliers simultaneously detected in the demand curves with both of the proposed procedures, but not detected when a method designed for independent data is applied.*

trend. The number of functional principal components considered was $K = 8$. These K principal components explained, at least, 98% of the variance.

Figure 5 displays the functional time series of electricity prices and the associated daily curves. Note that periods of low and high prices roughly correspond with periods of low and high demand, respectively. Greater variability is observed in the time series of prices, taking into account the different scale with respect to the electricity demand. It is easy to distinguish some of the zero-price days present in some points along the year, caused by the overproduction of wind power, and also in the daily curves (specially between 3:00h–6:00h).

The outliers identified by the proposed procedures PB and EB are listed in Table 5. Note that a total of 20 observations are detected as abnormal curves (13 from the PB method and 15 from the EB method). In addition, 7 of the 20 days corresponding to such outliers were days when demand curves were also identified as outliers (compare Tables 4 and 5). Following the classical rules of any kind of market, it is usual that demand and price are very interconnected, this being also the case of electricity markets and the reason why some of the outlying curves in demand are repeated as outliers in the electricity price. As in the previous application, one can argue causes for most of the abnormal curves of electricity price, being most of them already cited in the study of outliers in electricity demand. Some of the outliers correspond to zero-price days, as
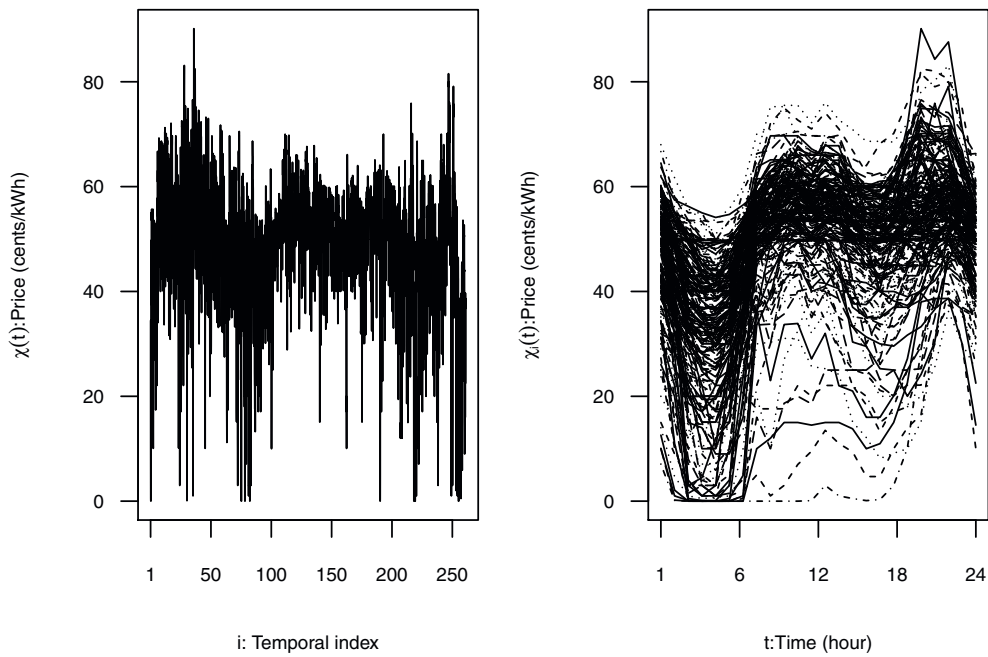
Figure 5: *Left panel: time series of electricity price. Right panel: daily electricity demand curves.*

Table 6: *Outliers detected in the price data from proposed procedures PB and EB.*

Day	Method PB	EB	Day	Method PB	EB
February 13	X		August 15	X	X
February 21		X	August 16		X
April 6	X		September 24		X
April 10	X	X	October 1		X
April 11	X	X	October 24		X
April 19	X		November 1	X	X
April 25	X	X	November 2	X	
May 1	X	X	December 14		X
May 8		X	December 24	X	X
June 11	X		December 25	X	X

April 19 and 25, September 24, November 1 and 2 or December 24 or days with a period close to zero price (February 13). Holidays have also some kind of influence over electricity prices, as April 6 (Good Friday) or May 1, August 15, November 1 and
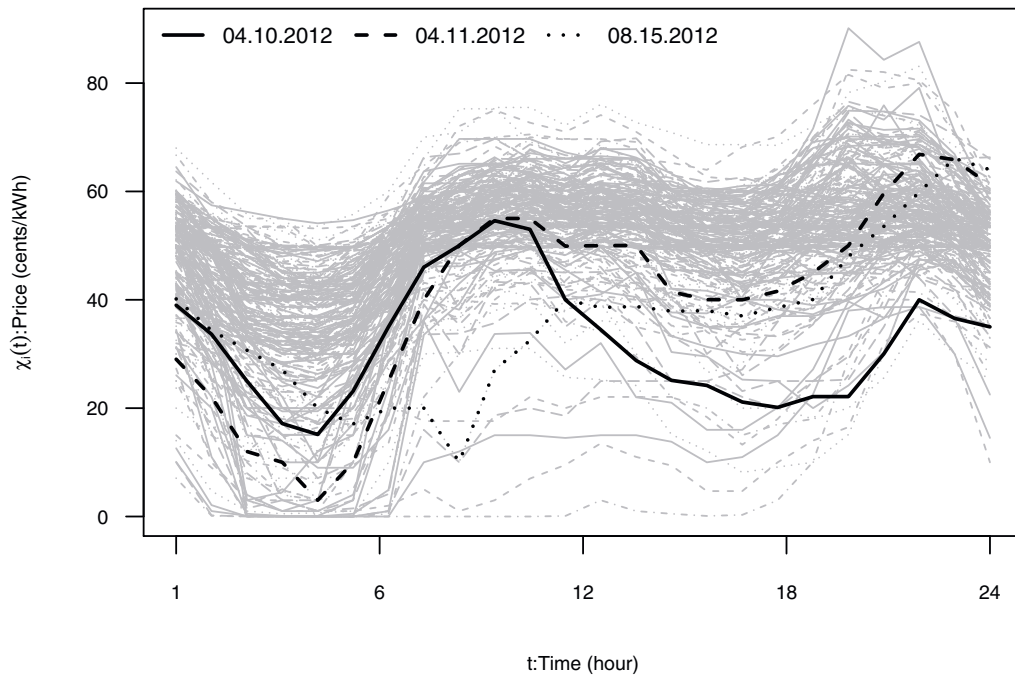
Figure 6: *Outliers in the price curves simultaneously detected using both proposed procedures, but not detected when applying a method designed for independent data.*

December 25. Finally, we find also some special days related to other holidays or linked to holidays, this being the case of February 21 (Carnival, holiday in part of Spain), August 16 (posterior to a nonworking day), November 2 (in the middle of a long weekend) or December 24 (Christmas Eve).

Finally, again as in the case of the demand, three price curves are detected as outliers simultaneously by the two proposed methods, but none of these curves is identified as an outlier with either the HDR or DBT procedure; we refer to the curves corresponding to April 10 and 11, and August 15, see Figure 6. It seems to make sense to consider them as outliers: the price was very low in the second half of the day on April 10 and the first half of the following day, April 11; August 15 is a holiday, and the pattern of the corresponding price curve of this day is different from the working days pattern.

6. Conclusions

This article proposes two methods to detect outliers in functional time series, the projections-based (PB) and the errors-based (EB) methods. These methods take dependence in the data into account and use robust functional principal component analysis (FPCA).

Our simulation studies have shown that the proposed methods present good performance when they are applied either on independent curves or dependent curves.

However, procedures designed for independent data, such as the functional HDR box-plot (Hyndman and Shang, 2010), the depth-based trimming (Febrero, Galeano, and González-Manteiga, 2008) or the integrated squared error (Hyndman and Ullah, 2007) methods, fail to detect outliers in functional time series. Thus, it has also been shown the need to take dependence in the time series into account. The PB method has very low false detection rate (p_f) while the sensitivity (p_c) of the EB approach is very high. Although in our simulation study small contamination sizes have been considered, both methods show acceptable trade-off between p_c and p_f. In fact, they improve the trade-off corresponding to the DBT for dependent data (Raña, Aneiros, and Vilar, 2015), this being (to the best of our knowledge) the only method in the statistical literature that includes the effect of dependence in the detection of outliers. Both PB and EB have also shown good performance in different situations, considering the kind of outlier (magnitude or shape outliers) and also de dependence scenario (low or high dependence). Their output is generally better than the other methods included in the comparison. Regarding sensitivity (p_c), EB seems to be the best option for both magnitude or shape outliers. Furthermore, PB is also very accurate and, although its p_c is lower than the EB, its false detection rate is the lowest of all the methods. We have also shown that both proposals improve their results as long as the sample size increases. The practical usefulness of our methodology has been illustrated on the daily curves of electricity demand and price.

Finally, it is worth pointing out that, as in all procedures based on FPCA, the proposed methods depend on the quantity of principal components considered, K. In this article, K was selected by imposing a lower bound to the cumulative percentage of variance explained from the first K principal components (cumulative percentage variance (CVA) approach). As in Hyndman and Booth (2008), we find that a general recommendation is to use a larger than necessary K (for instance, a K explaining at least 98% or, even, 99.9% of the variability). Alternatives to the CVA approach are, for instance, methods based on the cross-validation score (Yao, Müller, and Wang, 2005a) or the Akaike information criterion (Yao, Müller, and Wang, 2005b).

Acknowledgments

References

Aneiros-Pérez, G., Cardot, H, Estévez-Perez, G., and Vieu, P. (2004). Maximum ozone concentration fore-casting by functional non-parametric approaches. *Environmetrics*, 15, 675–685.

Aneiros, G., Vilar, J.M., Cao, R., and Muñoz-San-Roque, A. (2013). Functional prediction for the residual demand in electricity spot markets. *IEEE Transactions on Power Systems*, 28, 4201–4208.

Aneiros, G., Vilar, J., and Raña, P. (2016). Short-term forecast of daily curves of electricity demand and price. *Electrical Power and Energy Systems*, 80, 96–108.

Antoch, J., Prchal, L., De Rosa, M.R., and Sarda, P. (2010). Electricity consumption prediction with functional linear regression using spline estimators. *Journal of Applied Statistics*, 37, 2027–2041.

Antoniadis, A., Paparoditis, E., and Sapatinas, T. (2006). A functional wavelet kernel approach for time series prediction. *Journal of the Royal Statistics Society B*, 68, 837–857.

Arribas-Gil, A., and Romo, J. (2014). Shape outlier detection and visualization for functional data: the outliergram. *Biostatistics*, 15, 603–619.

Baíllo, A., Cuesta-Albertos, J.A., and Cuevas, A. (2011). Supervised classification for a family of Gaussian functional models. *Scandinavian Journal of Statistics*, 38, 480–498.

Besse, P.C., Cardot, H., and Stephenson, D. (2000). Autoregressive forecasting of some functional climatic variations. *Scandinavian Journal of Statistics*, 27, 673–688.

Boente, G., and Fraiman, R. (2000). Kernel-based functional principal components. *Statistics and Probability Letters*, 48, 335–345.

Cardot, H., Ferraty, F., and Sarda, P. (1999). Functional linear model. *Statistics and Probability Letters*, 45, 11–22.

Cho, H., Goude, Y., Brossat, X., and Yao, Q. (2013). Modeling and forecasting daily electricity load curves: a hybrid approach. *Journal of the American Statistical Association*, 108, 7–21.

Cryer, J.D., and Chan, K.S. (2008). *Time Series Analysis*, New York: Springer.

Cuevas, A. (2014). A partial overview of the theory of statistics with functional data. *Journal of Statistical Planning and Inference*, 147, 1–23.

Cuevas, A., Febrero, M., and Fraiman, R. (2006). On the use of the bootstrap for estimating functions with functional data. *Computational Statistics and Data Analysis*, 51, 1063–1074.

Cuevas, A., Febrero, M., and Fraiman, R. (2007). Robust estimation and classification for functional data via projection-based depth notions. *Computational Statistics*, 22, 481–496.

Delsol, L., Ferraty, F., and Vieu, P. (2011). Structural test in regression on functional variables. *Journal of Multivariate Analysis*, 102, 422–447.

Febrero, M., Galeano, P., and González-Manteiga, W. (2007). Functional analysis of NOx levels: location and scale estimation and outlier detection. *Computational Statistics*, 22, 411–427.

Febrero, M., Galeano, P., and González-Manteiga, W. (2008). Outlier detection in functional data by depth measures, with application to identify abnormal NOx levels. *Environmetrics*, 19, 331–345.

Ferraty, F., and Romain, Y. (Eds.) (2011). *The Oxford Handbook of Functional Data Analysis*, Oxford: Oxford University Press.

Ferraty, F., van Keilegom, I., and Vieu, P. (2012). Regression when both response and predictor are functions. *Journal of Multivariate Analysis*, 109, 10–28.

Ferraty, F., and Vieu, P. (2002). The functional nonparametric model and application to spectrometric data. *Computational Statistics*, 17, 545–564.

Ferraty, F., and Vieu, P. (2006). *Nonparametric Functional Data Analysis*, New York: Springer-Verlag.

Fraiman, R., and Svarc, M. (2013). Resistant estimates for high dimensional and functional data based on random projections. *Computational Statistics and Data Analysis*, 58, 326–338.

García-Portugués, E., González-Manteiga, W., and Febrero-Bande, M. (2014). A goodness-of-fit test for the functional linear model with scalar response. *Journal of Computational and Graphical Statistics*, 23, 761–778.

Gervini, D. (2012). Outlier detection and trimmed estimation for general functional data. *Statistica Sinica*, 22, 1639–1660.

González-Manteiga, W., and Martínez-Calvo, A. (2011). Bootstrap in functional linear regression. *Journal of Statistical Planning and Inference*, 141, 453–461.

Hall, P. (2011). Principal component analysis for functional data: methodology, theory, and discussion. *The Oxford Handbook of Functional Data Analysis*, F. Ferraty and Y. Romain, Eds., Oxford: Oxford University Press, 210–234.

Hall, P., Müller, H.G., and Wang, J.L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Annals of Statistics*, 34, 1493–1517.

Horváth, L., and Kokoszka, P. (2012). *Inference for Functional Data with Applications*, New York: Springer.

Hubert, M., Rousseeuw, P.J., and Verboven, S. (2002). A fast method of robust principal components with applications to chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 60, 101–111.

Hyndman, R.J. (1996). Computing and graphing highest density regions. *The American Statistician*, 50, 120–126.

Hyndman, R.J., and Ullah, M.S. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics and Data Analysis*, 51, 4942–4956.

Hyndman, R.J., and Booth, H. (2008). Stochastic population forecasts using functional data models for mortality, fertility and migration. *International Journal of Forecasting*, 24, 323–342.

Hyndman, R.J., and Shang, H.L. (2010). Rainbow plots, bagplots, and boxplots for functional data. *Journal of Computational and Graphical Statistics*, 19, 29–45.

Künsch, H.R. (1989). The jacknife and the bootstrap for general stationary observations. *Annals of Statistics*, 17, 1217–1241.

Li, Y., and Hsing, T. (2007). On rates of convergence in functional linear regression. *Journal of Multivariate Analysis*, 98, 1782–1804.

Liebl, D. (2013). Modeling and forecasting electricity spot prices: a functional data perspective. *Annals of Applied Statistics*, 7, 1562–1592.

Ocaña, F.A., Aguilera, A.M., and Escabias, M. (2007). Computational considerations in functional principal component analysis. *Computational Statistics*, 22, 449–465.

Ramsay, J.O., and Silverman, B.W. (2005). *Functional Data Analysis*. New York: Springer-Verlag.

Raña, P., Aneiros, G., and Vilar, J.M. (2015). Detection of outliers in functional time series. *Environmetrics*, 26, 178–191.

Sawant, P., Billor, N., and Shin, H. (2012). Functional outlier detection with robust functional principal component analysis. *Computational Statistics*, 27, 83–102.

Sguera, C., Galeano, P., and Lillo, R. (2014). Spatial depth-based classification for functional data. *Test*, 23, 725–750.

Shang, H.L. (2014). Bayesian bandwidth estimation for a functional nonparametric regression model with mixed types of regressors and unknown error density. *Journal of Nonparametric Statistics*, 26, 599–615.

Sun, Y., and Genton, M.G. (2011). Functional boxplots. *Journal of Computational and Graphical Statistics*, 20, 316–334.

Tsay, R.S., Peña, D., and Pankratz, A.E. (2000). Outliers in multivariate time series. *Biometrika*, 87, 789–804.

Vilar, J.M., Cao, R., and Aneiros, G. (2012). Forecasting next-day electricity demand and price using nonparametric functional methods. *International Journal of Electrical Power and Energy Systems*, 39, 48–55.

Yao, F., Müller, H.G., and Wang, J.L. (2005a). Functional linear regression analysis for longitudinal data. *Annals of Statistics*, 33, 2873–2903.

Yao, F., Müller, H.G., and Wang, J.L. (2005b). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100, 577–590.

Yu, G., Zou, C., and Wang, Z. (2012). Outlier detection in functional observations with applications to profile monitoring. *Technometrics*, 54, 308–318.