



Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

On projection methods for functional time series forecasting

Antonio Elías^{a,*}, Raúl Jiménez^b, Han Lin Shang^c^a OASYS group, Department of Applied Mathematics, Universidad de Málaga, Málaga, Spain^b Department of Statistics, Universidad Carlos III de Madrid, Madrid, Spain^c Department of Actuarial Studies and Business Analytics, Macquarie University, Sydney, Australia

ARTICLE INFO

Article history:

Received 4 February 2021

Received in revised form 18 May 2021

Accepted 21 October 2021

Available online 3 November 2021

AMS 2020 subject classifications:

62R10

62G30

Keywords:

Forecasting

Functional depth

Functional nonparametric

Functional time series

Nearest neighbors

ABSTRACT

Two nonparametric methods are presented for forecasting functional time series (FTS). The FTS we observe is a curve at a discrete-time point. We address both one-step-ahead forecasting and dynamic updating. Dynamic updating is a forward prediction of the unobserved segment of the most recent curve. Among the two proposed methods, the first one is a straightforward adaptation to FTS of the k -nearest neighbors methods for univariate time series forecasting. The second one is based on a selection of curves, termed the curve envelope, that aims to be representative in shape and magnitude of the most recent functional observation, either a whole curve or the observed part of a partially observed curve. In a similar fashion to k -nearest neighbors and other projection methods successfully used for time series forecasting, we “project” the k -nearest neighbors and the curves in the envelope for forecasting. In doing so, we keep track of the next period evolution of the curves. The methods are applied to simulated data, daily electricity demand, and NO_x emissions and provide competitive results with and often superior to several benchmark predictions. The approach offers a model-free alternative to statistical methods based on FTS modeling to study the cyclic or seasonal behavior of many FTS.

© 2021 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In the last few decades, technological advancements have simplified and decreased the cost of data collecting and storing processes. This new paradigm has made data scientific to confront not only big data sets but also complex data structures. Functional data analysis (FDA) arises naturally in this context to exploit the information recorded over a continuum such as time or space. The literature has experienced a prominent development of techniques and tools to take advantage of this kind of data. While the monographs of [23,38,59] provide complete picture of the methodological and theoretical contributions, recent advances in the field can be found in survey papers [1,18,27,78].

This article addresses the problem of forecasting functional time series (FTS). These are time series $\{y_k, k \in \mathbb{N}\}$ where each y_k is a random function $t \rightarrow y_k(t)$, $t \in [a, b]$. We refer the reader to Hörmann and Kokoszka [28] for theory and examples of FTS. One-step-ahead forecasting is one of the most important problems in FTS analysis, and it has been addressed by diverse prominent literature [see, e.g., 2,3,7,9,11,30,31,34,60]. Also, forward prediction of a curve which is already being observed, termed dynamic updating by Shang and Hyndman [66], a topic of current interest [64,67]. As many classical methods for time series forecasting, for example, those based on the ARIMA family models, the benchmark

* Correspondence to: OASYS group, Department of Applied Mathematics, University of Málaga, Ada Byron Research Building, C/ Arquitecto Francisco Peñalosa, 18, 29010, Málaga, Spain.

E-mail address: aelias@uma.es (A. Elías).

methods for FTS forecasting are based on fitting some statistical model of some approximated data generation. The most common is to fit ARIMA or VAR models to the Functional Principal Components Scores (FPC) of the functions, see FAR model [9,34]. Klepsch and Klüppelberg [37] considered Functional Moving Average (FMA) and FARMA. Other authors [41,42] consider functional autoregressive fractionally integrated moving average (ARFIMA). Here we discuss an alternative model-free approach that addresses both one-step-ahead forecasting and dynamic updating in a unified way.

The literature of artificial intelligence and machine learning for time series forecasting has gained considerable prominence over the last decade [see, e.g.,51]. Besides the lack of an underlying model, many of these methods exhibit interesting features compared with classical statistical methods, including their ability to capture nonlinearities [53]. Among the most powerful machine learning technics used for time series forecasting, we must mention artificial neural networks [82] and k -nearest neighbors (KNN). The simplicity of implementing the algorithm and its high accuracy have popularized KNN. It has demonstrated to be a strong contender in forecasting competitions [52]. The method can be schematized as follows:

1. Consider T observations, x_1, \dots, x_T , from a time series and a prediction horizon h . Our goal is to forecast $p = (x_{T+1}, \dots, x_{T+h})$, i.e. the time series h periods ahead.
2. Consider all the possible observed $(E + h)$ -tuples $(x_{i-(E-1)}, \dots, x_i, x_{i+1}, \dots, x_{i+h})$. This means $E \leq i \leq T - h$.
3. Denote $\mathbf{x}_i = (x_{i-(E-1)}, \dots, x_i)$ and its projection $P(\mathbf{x}_i) = (x_{i+1}, \dots, x_{i+h})$. Note that p is the projection of \mathbf{x}_T .
4. Find the k -nearest neighbors to \mathbf{x}_T from $\{\mathbf{x}_i, E \leq i \leq T - h\}$.
5. Predict p by a convex combination of the projections of the k -nearest neighbors.

In addition to the distance function used to find the nearest neighbors and the convex combination used for predicting, KNN involves setting the sometimes called embedding dimension E and the number of neighbors denoted by k . The simplex projection [71] and the S-map projection [70] are methods related to KNN that provide skillful forecasts, free of k .

Moreover, these model-free methods have outperformed any forecast based on models [58]. Surprisingly, functional versions have not been developed for FTS. In contrast, in the context of FDA, the KNN method has been explored and theoretically supported towards regression problems [10,12,40,43] or functional classification problems [13,29]. Biau et al. [10] introduce asymptotic results when the response variable has a finite dimension. Kara et al. [36] provide asymptotic results that, exceptionally, allow for an automatic choice data-driven of the number of neighbors for several operators, among them functional regression with a scalar response. Lian [43] proves the consistency of the KNN regression estimate when both dependent and independent variables are functions. Recently, Ferraty and Nagy [22] use the nearest neighbors method for bandwidth selection on a problem of scalar-on-function local linear regression. Remarkably, Zhang et al. [83] address the problem of forecasting final prices of auctions via functional KNN. Roughly, their forecasting is based on the k -nearest price curves to the item being predicted and their corresponding final prices. We consider dynamic updating in line with these authors. For example, if we predict afternoon electricity demand, we use afternoon data from days with morning data near to the morning of the day being predicted. For one-step-ahead forecasting, we mimic the projection procedure inherent to KNN. For example, for tomorrow's forecasting electricity demand, we consider next-day demands of daily curves near to today. Moreover, we take ideas of Sugihara and May [71], and we do not only consider the nearest daily curves but those that surround the most recent data from above and below, when this is possible, to select a neighbor that is representative of its shape and its magnitude. This is the key to produce coherent band predictions, in which the observed part of the day being predicted is inside the band delimited by the curves that we are using for predicting. Whereas the curve to predict may be entirely outside of the band delimited by its k -nearest curves. In this vein, we combine two concepts, nearness and centrality, commonly termed depth in the functional context.

Functional depth [25,56] has received a great deal of attention since it was introduced by Fraiman and Muniz [24]. It has been used for several applications; including classification [16,17,19,29,46], outlier detection [8,21,35,54,55], populations comparison [47,48,74], and clustering [68,75]. Functional versions of boxplots and other graphical tools based on different depths have also been proposed for visualizing curves to discover features from a sample that might not be apparent using other methods [20,32,62,72]. However, up to our knowledge, the concept of depth has not been used for forecasting.

Here we introduce a depth-based projection method for forecasting that constitutes a contribution to the literature of nonparametric FDA, an area popularized by Ferraty and Vieu [23] and that it is actively growing, see the review by Ling and Vieu [44]. Also, we consider an adaptation to FTS of the KNN briefly schematized above and we follow ideas from Kara et al. [36] to make it automatic. We also consider several benchmark methods in the field for comparisons. For that, we use daily functions of electricity demand and nitrogen monoxide (NOx) emissions. Both are FTS that have already been considered by the literature. Also, both exhibit a strong calendar-effect [79], making them appropriate to illustrate the excellent performance of the method discussed here. As we discuss below, the projection methods have been intended to forecast FTS with seasonal or cyclical patterns. Through a simulation study, we investigate the impact on forecast accuracy from the departure of stationarity. We remark that, although we do not assume any stationarity of first or second order, we do not consider applications to FTS with a significant trend. Specifically, we suppose the curves of the FTS fluctuate in a finite band.

This paper is organized as follows. Section 2 describes the projection methods considered here. Section 3 provides results of comparing the proposed methods with benchmark methods. Section 4 concludes with a general discussion.

2. Projection methods for forecasting

Let $\{y_i, i \in \mathbb{N}\}$ be a FTS [28]. Here, we assume that each y_i is a univariate random function and, without loss of generality, we suppose that its domain is $[0, 1]$.

2.1. Basic notions

We consider two scenarios of prediction:

- One-step-ahead forecasting; when we have observed y_1, \dots, y_n on $[0, 1]$, and we want to predict $\{y_{n+1}(t) : t \in [0, 1]\}$.
- Dynamic updating; when we have observed y_1, \dots, y_{n-1} on $[0, 1]$ and $y_n(t)$ on $[0, q]$, for a given $0 \ll q < 1$, and we want to predict $\{y_n(t) : t \in (q, 1]\}$.

In both scenarios, we refer to the most recently observed functional datum, either $\{y_n(t) : t \in [0, 1]\}$ or $\{y_n(t) : t \in [0, q]\}$, as the focal curve. The latter will be denoted by f and its observation domain by D_f , regardless of whether it is $[0, 1]$ or $[0, q]$. For each $i, 1 \leq i \leq n - 1$, we consider the restriction of y_i to D_f , this is the curve with graph equal to $\{(t, y_i(t)) : t \in D_f\}$. We denote this restriction by $y_{i|f}$ and we refer to $\mathcal{Y}_n = \{y_{1|f}, \dots, y_{n-1|f}\}$ as the past-focal-curves sample. Note that in one-step-ahead forecasting $y_{i|f} \equiv y_i$.

Denote by p the curve or curve segment to predict, this is $\{y_{n+1}(t) : t \in [0, 1]\}$ or $\{y_n(t) : t \in (q, 1]\}$. The domain of p , either $[0, 1]$ or $(q, 1]$, is denoted by D_p . Then, for each $y_{i|f} \in \mathcal{Y}_n$, we consider the projection

$$\mathcal{P}y_{i|f} = \begin{cases} \{y_{i+1}(t), t \in [0, 1]\}, & \text{in one-step-ahead forecasting.} \\ \{y_i(t), t \in (q, 1]\}, & \text{in dynamic updating.} \end{cases}$$

Note that $\mathcal{P}f = p$ and that, although not explored it in this article, h -step-ahead projection could be achieved by $\mathcal{P}y_{i|f} = \{y_{i+h}(t), t \in [0, 1]\}$.

Many FTS regularly repeat dependency patterns between past focal curves and their projections. The change in shape and magnitude of the curves over time of some FTS can be visualized from the rainbow plots of Hyndman and Shang [32] and other related plots [65].

In dynamic updating, we observe that $y_{i|f}$ and $\mathcal{P}y_{i|f}$ are a partition of y_i in two curve segments. The dependence patterns between past focal curves and their projections can be observed from groups of curves with similar colors in the rainbow plots already mentioned. Similar curves may be distant in time due to seasons and cycles, making specific patterns be repeated over time. It is not necessarily an easy task to visualize dependency patterns between focal curves and their projections; it depends on the complexity of the FTS and the curves' shapes, and the number of curves involved. Nonetheless, the cyclic nature of many FTS suggests that there may be statistical regularity between the morphology of focal curves and the morphology of their projections. The core idea of the methods discussed here is based on the existence of this kind of regularity. Thus, if f is surrounded by past focal curves that capture both its shape and magnitude, we could predict p from these past focal curves' projections. Of course, when the FTS has a strong global trend, the future values are always out of the historical data magnitude. Therefore, the projection methods discussed here are not suitable without detrending in such cases.

2.2. The focal-curve envelope

The straightforward approach for surrounding f with past focal curves is by using KNN. However, as we already commented, f does not have to be necessarily centered in the band delimited by its k -nearest curves. Even f may be completely outside of this band. To produce bands where f is central, which is natural for band estimation, we use the concept of functional depth. In particular, we use the modified band depth of López-Pintado and Romo [47] with bands formed by two curves. This is a widely used depth measure that allows a million curves to be ranked in only tenths of seconds [73] making our algorithm efficient even for large data sets. Strictly speaking, for any $\mathcal{J} \subset \mathcal{Y}_n$ with $|\mathcal{J}| \geq 2$ and $y \in \mathcal{J} \cup \{f\}$, we consider

$$D(y, \mathcal{J}) = \frac{1}{2} \binom{|\mathcal{J}| + 1}{2}^{-1} \sum_{x, z \in \mathcal{J} \cup \{f\}} \lambda(\{t \in D_f : \min(x(t), z(t)) \leq y(t) \leq \max(x(t), z(t))\}),$$

$\lambda(\{t \in D_f : A(t)\})$ being the proportion of time that $A(t)$ is true on D_f . The larger $D(y, \mathcal{J})$, the deeper will be y in $\mathcal{J} \cup \{f\}$. Denote by \mathcal{J}_k the k deepest curves of \mathcal{J} . We remark that $f \notin \mathcal{J}_k$ although f may be the deepest curve of $\mathcal{J} \cup \{f\}$.

In line with Sun and Genton [72], who consider central regions based on the modified band depth for visualizing the spread of the deepest curves, we consider the Focal Central Region (FCR) delimited by \mathcal{J}_k . This is the region

$$R(\mathcal{J}_k) = \left\{ (t, y(t)) : t \in D_f, \min_{x \in \mathcal{J}_k} x(t) \leq y(t) \leq \max_{x \in \mathcal{J}_k} x(t) \right\}.$$

So that $R(\mathcal{J}_k)$ can capture the shape and magnitude of f , $R(\mathcal{J}_k)$ must be roughly centered at f and its mean width should be small. Having these objectives in mind, we look for a set \mathcal{J} of past focal curves, as large as possible in order of not restricting the range of k , such that:

- (1) f is deep in $\mathcal{J} \cup \{f\}$, the deepest if possible;
- (2) f is enveloped by \mathcal{J} as much as possible. Here, we say f is more enveloped by \mathcal{J} than by \mathcal{J}' if and only if

$$\lambda \left(\left\{ t : \min_{y \in \mathcal{J}} y(t) \leq f(t) \leq \max_{y \in \mathcal{J}} y(t) \right\} \right) > \lambda \left(\left\{ t : \min_{y \in \mathcal{J}'} y(t) \leq f(t) \leq \max_{y \in \mathcal{J}'} y(t) \right\} \right);$$

- (3) f is surrounded by near curves of \mathcal{J} , as many as possible.

Algorithm 1 provides a set of past focal curves with the three features above that we call the focal-curve envelope and denote by \mathcal{J} from now on. In a nutshell, the algorithm iteratively selects as many past curves as possible. From the nearest to the farthest to f , we obtain an envelope of f with an increasing depth.

Algorithm 1: Input: \mathcal{Y}_n, f . Output: \mathcal{J}

```

Initialize  $\mathcal{Y} = \mathcal{Y}_n, \mathcal{J} = \emptyset, D(f, \mathcal{J}) = 0$ 
while size of  $\mathcal{Y} \geq 2$  do
  Let  $y'$  be the nearest curve to  $f$  from  $\mathcal{Y}$  and  $\mathcal{N} = \{y'\}$ 
  for  $y \in \mathcal{Y} \setminus \{y'\}$ , from the nearest curve to the farthest from  $f$ , do
    if  $f$  is more enveloped by  $\mathcal{N} \cup \{y\}$  than by  $\mathcal{N}$  then
       $\mathcal{N} = \mathcal{N} \cup \{y\}$ 
    end if
  end for
  if  $D(f, \mathcal{J} \cup \mathcal{N}) \geq D(f, \mathcal{J})$  then
     $\mathcal{J} = \mathcal{J} \cup \mathcal{N}$ 
  end if
   $\mathcal{Y} = \mathcal{Y} \setminus \mathcal{N}$ 
end while

```

As an illustration of how Algorithm 1 works, see top panels of Fig. 1, where 100 curves from a simulated FTS are considered. The left panels show the twelve curves selected on the first iteration of Algorithm 1 (top panel) and the 12-nearest neighbors (bottom panel). For measuring nearness, hereafter, we use L_2 distance. Besides this is the most widely used distance in the KNN context, we base our choice by looking for a small mean squared error on the prediction method discussed in Section 2.3. We emphasize the intervals where the nearest curves do not envelope the focal curve. Right panels show the band delimited by the focal-curve envelope obtained after five iterations and formed by forty-six curves and the corresponding band from the 46-nearest curves. As a general rule, Algorithm 1 selects near curves that are enveloping the focal until it is the deepest one. In contrast, nearest neighbors neither necessarily envelope the focal curve nor make it the deepest one.

2.3. Prediction

We make predictions by projecting the k -nearest curves to the focal and projecting the curves that belong to the focal-curve envelope. We compute weighted means of these projections for punctual prediction, assigning more weights to the curves closer to the focal. Following the literature on KNN [36,52], and simplex projections or S-map [58], we consider two types of weights: On one hand, weights inversely proportional to the square distance to the focal, which we identify in the sequel by $1/d$ weights. On the other hand, exponential weights. Specifically, let f_1, \dots, f_k be the k -nearest curves to f , $d_i = \|f_i - f\|^2$, and more general $d_y = \|y - f\|^2$, let \mathcal{J} be the focal-curve envelope of f and $\theta > 0$, what we do is to consider the following estimators of $p = \mathcal{P}f$:

- functional KNN projections (fKNN) with $1/d$ and exponential weights

$$\hat{p}_k = \frac{\sum_{i=1}^k \frac{1}{d_i} \mathcal{P}f_i}{\sum_{i=1}^k \frac{1}{d_i}}, \quad \text{and} \quad \hat{p}_{k,\theta} = \frac{\sum_{i=1}^k e^{-\theta d_i/d_1} \mathcal{P}y_i}{\sum_{i=1}^k e^{-\theta d_i/d_1}}; \tag{1}$$

- Envelope projections (EP) with $1/d$ and exponential weights

$$\hat{p} = \frac{\sum_{y \in \mathcal{J}} \frac{1}{d_i} \mathcal{P}f_i}{\sum_{y \in \mathcal{J}} \frac{1}{d_i}}, \quad \text{and} \quad \hat{p}_\theta = \frac{\sum_{y \in \mathcal{J}} e^{-\theta d_y/d_1} \mathcal{P}y}{\sum_{y \in \mathcal{J}} e^{-\theta d_y/d_1}}. \tag{2}$$

For fixed θ , k in (1) will be chosen by minimizing the prediction Mean Square Error (MSE) on a rolling forecasting origin procedure, as it is customary in the literature of KNN for time series. We denote by k^* the value that minimizes the MSE. Similarly, we can select θ in (2). Let θ^* be the resulting value of minimizing MSE by rolling origin. We discard selection of optimal pairs (θ, k) for $\hat{p}_{k,\theta}$ in (1) by a high computational cost involved in this two-dimensional optimization problem. To compare the methods, on one hand we compute \hat{p}_{k^*} and \hat{p} . On the other hand, we fix $\theta = 1$ and compute $\hat{p}_{k^*,\theta=1}$ and $\hat{p}_{\theta=1}$. In addition, we compute \hat{p}_{θ^*} .

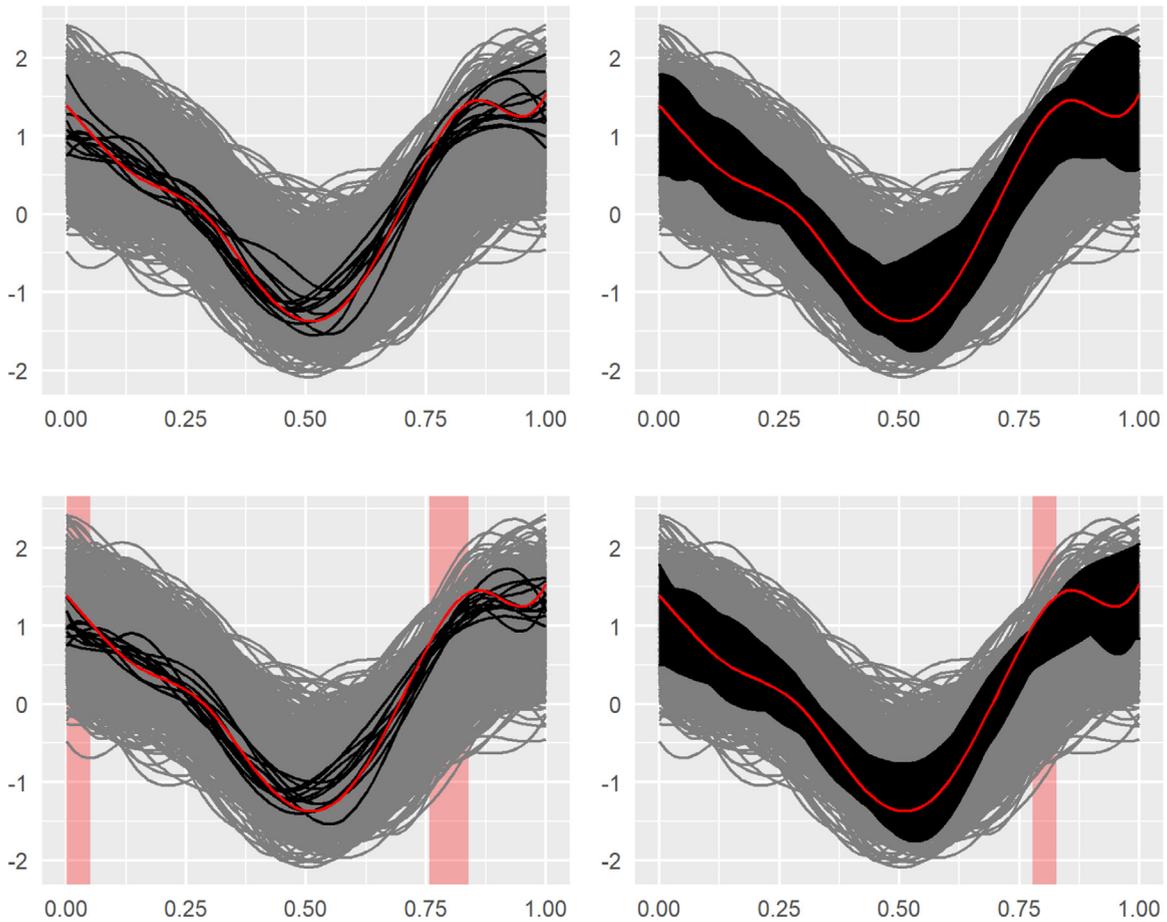


Fig. 1. Top panels: focal curves (in red), the twelve curves selected on the first iteration of Algorithm 1 (left panel) and band delimited by the focal-curve envelope from the 46 curves selected by the algorithm (right panel). Bottoms panels: The 12-nearest curves and the band delimited by the 46-nearest curves. The intervals where the nearest curves do not envelope the focal are shaded in red. Algorithm 1 selects near curves that envelope the focal until making it been the deepest one. The nearest curves do not necessarily envelope the focal nor making it been the deepest one. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

For band prediction, we consider the region delimited by the projections of the k deepest curves of \mathcal{J} . This is the band

$$\mathcal{PR}(\mathcal{J}_k) = \left\{ (t, y(t)) : t \in D_p, \min_{x \in \mathcal{J}_k} \mathcal{P}x(t) \leq y(t) \leq \max_{x \in \mathcal{J}_k} \mathcal{P}x(t) \right\}.$$

This band prediction corresponds to a central functional region such as those defined by [72]. As with the fKNN projections, k is chosen by minimizing a rolling forecasting origin procedure's cost function. For that, we consider the Winkler score. This is a measure that penalizes both bandwidth and non-coverage for a given $\alpha \in (0, 1)$, penalizing even more heavily when observations are remarkably outside the band the confidence levels are higher [26,80]. Thus, for a given $(1 - \alpha)$ mean coverage, i.e., the proportion of time that p is in the band, k is chosen by minimizing the score. Mimicking this procedure, we also consider the region delimited by the projection of the k nearest curves for band prediction based on fKNN.

Fig. 2 shows an example of point and band prediction based on 1000 curves from a simulated FTS corresponding to the model discussed in Section 3.1 with $\mu = 0$. Both one-step-ahead forecasting and dynamic updating provided by EP with \hat{p}_θ are illustrated.

3. Results

For one-period-ahead forecasting, we compare fKNN and EP estimators with benchmark predictions for stationary FTS. In particular, we implement a variant of the method proposed by Hyndman and Ullah [34] and Hyndman and Shang [31], who suggest functional prediction based on modeling the functional principal component scores as independent time series. Following Aue et al. [9], we will consider the first d functional scores as a d -variate time series since the score

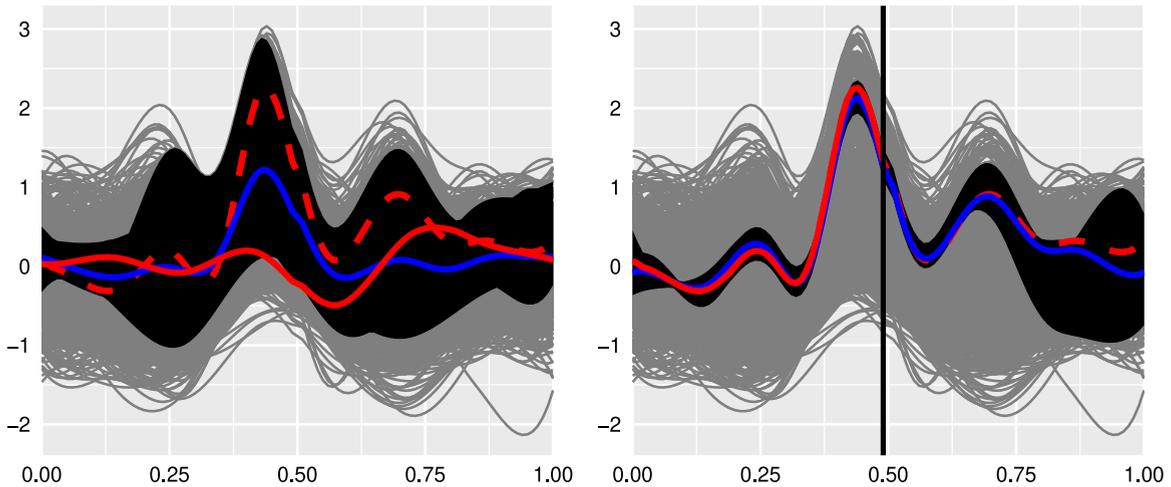


Fig. 2. 1000 curves from a simulated FTS (gray). Focal curve f in solid red, curve to predict p in dashed red, point prediction \hat{p}_1 in blue and prediction band $\mathcal{PR}(\mathcal{J}_{10})$ in black. Left panel: One-step-ahead prediction. Right panel: Dynamic updating from 0.5 to 1. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

vectors could have non-diagonal autocorrelations. The number of scores d is chosen as the smallest integer such that the first d principal components explain at least 80% of the variance of the data. Then, we fit a d -dimensional autoregressive model [50] to the scores' time series. This method is termed FPCF. For validating FPCF as a benchmark method, we repeated the simulation study conducted by Aue et al. [9] and obtained competitive results with the best of their methods (see Appendix A.1).

For dynamic updating, we compare fKNN, and EP with the method of [39] for predicting partially observed functional data. We also tested the well-known method of Yao et al. [81] but Kraus [39] achieved better results in our experiments. When the FTS is constructed by slicing a periodic univariate time series, we also consider the Block-Moving (BM) approach by Shang and Hyndman [66]. The BM approach is designed to re-arrange the time series so that all the FTS are ultimately observed and, then, to apply FPCF is possible. The literature devoted to dynamic updating has successfully proposed methodologies based on ordinary least squares (OLS) regression, ridge regression (RR), penalized least squares (PLS) regression, or functional linear regression [64,66,67]. However, we do not compare with them because we deal with large data sets where these methods might be computationally inefficient. In other circumstances, they certainly are the benchmark of the field.

Finally, to understand the scale of how superior a method is with respect to the other, we also report about the two most basic estimators. They are the historical mean and the naive prediction. This is just the most recent observed datum corresponding to the prediction range. In our notation, the naive predictor of $\mathcal{P}y_n$ is $\mathcal{P}y_{n-1}$.

The implementation of the Envelope Projection method is available at <https://github.com/aefd/nnFTS>. The benchmark methods for Functional Time Series are available in the `fts` R package by Hyndman and Shang [33]. Finally, the method for estimating partially observed data was implemented by [39] with the code available from https://www.davidkraus.net/?page_id=349.

3.1. Stationary FTS with random shocks

To simulate FTS with the features discussed at the end of Section 2.1, we consider slices of a scalar process with complex periodic rhythm that artificially corrupt with shocks. This process is denoted Y_μ , being $\mu \in [0, 1)$ the expected proportion of affected periods. The process is the sum of three independent components corresponding to:

1. A noise $\{X(t) : t \geq 0\}$, modeled by a zero-mean stationary Gaussian process;
2. A random periodic function $f : [0, +\infty) \rightarrow \mathbb{R}$;
3. A random shock process $\{S_\mu(t) : t \geq 0\}$ that occasionally affects two contiguous periods of f . For $\mu = 0$, $S_\mu = 0$.

See Appendix A.2 for more details about these processes. Thus $Y_\mu = X + f + S_\mu$ and the considered FTS is $y_i(t) = Y_\mu(t + i - 1)$, $0 \leq t \leq 1$, $i \in \mathbb{N}$. If $\mu = 0$, there are no shocks and the corresponding FTS is stationary. With increasing μ , we explore different non-stationary regimes. Fig. 3 shows two random trajectories of $Y_{0.2}$ and their respective FTS.

We performed 300 realizations of Y_μ , 100 for each level of contamination considered. They were $\mu = 0.0, 0.2$ and 0.4 . For each realization, we considered 1000 periods and, from the corresponding 1000 slices, the exercise of one-step-ahead prediction and dynamic updating on the last 100 slices. For dynamic updating, we chose $q = 0.5$, corresponding to half slice. The resulting MSEs are shown in Table 1. In summary, we found the proposed methods were superior to the benchmark method. They were even better at higher contamination levels, and fKNN is often superior to EP.

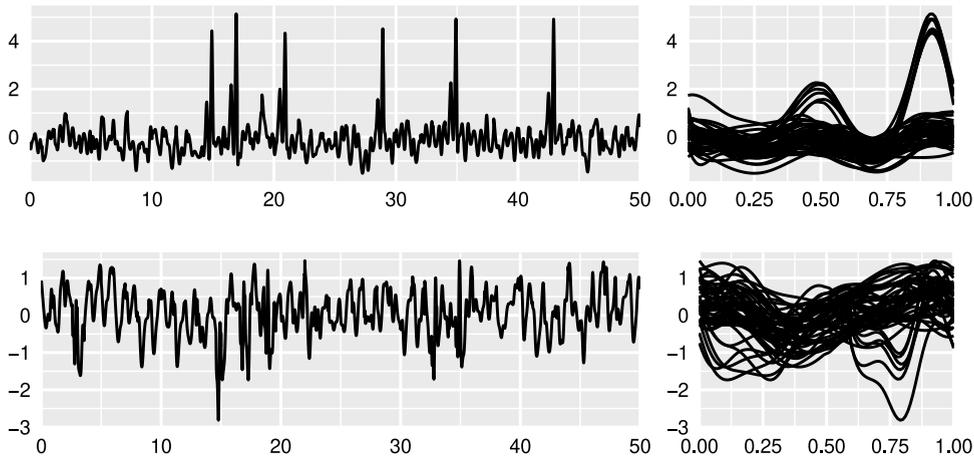


Fig. 3. Two random trajectories of $Y_{0,2}$ (left panel) and their corresponding FTS (right panel).

Table 1

Mean Squared Errors for one-step-ahead forecasting and dynamic updating from a stationary FTS ($\mu = 0$) contaminated with random shocks at two levels of contamination ($\mu = 0.2$ and $\mu = 0.4$). The observed minimum MSE among the considered methods is remarked in bold.

One-step-ahead prediction									
	Average	Naive	fKNN		EP			FPCF	
			\hat{p}_{k^*}	$\hat{p}_{k^*, \theta=1}$	\hat{p}	$\hat{p}_{\theta=1}$	\hat{p}_{θ^*}		
$\mu = 0.0$	0.192	0.385	0.1880	0.1879	0.192	0.199	0.192	0.190	
$\mu = 0.2$	2.465	5.380	1.5288	1.5153	1.646	1.679	1.594	1.662	
$\mu = 0.4$	4.513	11.35	2.5893	2.5974	2.912	2.832	2.744	2.767	
Dynamic updating									
	Average	Naive	fKNN		EPU			BM	Kraus
			\hat{p}_{k^*}	$\hat{p}_{k^*, \theta=1}$	\hat{p}	$\hat{p}_{\theta=1}$	\hat{p}_{θ^*}		
$\mu = 0.0$	0.192	0.382	0.1582	0.1582	0.160	0.172	0.157	0.182	0.166
$\mu = 0.2$	2.527	5.520	1.2109	1.1920	1.669	1.421	1.239	2.173	1.599
$\mu = 0.4$	4.783	12.126	1.4078	1.3805	2.278	1.640	1.418	3.756	2.750

3.2. Spanish electricity demand

Functional methods for electricity demand forecasting have been widely used [4–6,14,57,60,63,76,77]. These methods often consider particular features of the data, such as weekly seasonality, the effect of holidays, and outliers' presence. Thus, researchers process weekdays separately from Saturdays and Sundays and apply methods for replacing outliers. In addition to this, Raña et al. [60] use additional data from functional exogenous covariates in their Functional Additive Models (FAM) for next-day forecasting. Certainly, they provide the benchmark predictions in functional Spanish electricity demand forecasting.

We consider the challenge of predicting the daily curves of a whole year (2018) without considering any specific peculiarity of the electricity demand by using raw data. For dynamic updating, we considered the practical case of half-day ahead prediction. Our study is based on electricity demand at every 10-minute interval, from January 1, 2014, to December 31, 2018. The data set is available at <http://www.ree.es/es/>. Following Raña et al. [60], we compute the mean absolute percentage error (MAPE) in addition to the MSE. Viewing the weekly seasonality of electricity demand and the naive prediction corresponding to yesterday for predicting today, we also consider the naive prediction corresponding to the same day one week ago, identified below as seasonal naive (Naive' in the tables). The results are shown in Table 2.

From this table, we remark that FPCF is slightly inferior to the naive method and significantly inferior to seasonal naive. However, the projection methods were superior to the benchmark competitors for both one-step-ahead forecasting and dynamic updating, and EP slightly superior to fKNN.

Although they are different studies, MAPE of Raña et al. [60] (in the last four rows of the last column of their Table 1) might reference the expected magnitude of the forecasting performance. The naive method in both studies produces the same MAPE (6.383 vs. 6.39). However, we remark that MAPE computed by Raña et al. [60] corresponds to 2012 prediction instead of 2018, using demand each hour, instead of every 10 min, and therefore conclusions might be taken with caution. From this extrapolation, FPCF in Table 2 seems to be by far inferior to FAM. This points out the convenience of removing outliers and taking into account the data's weekly seasonality to improve predictions as suggested by Raña et al. [60].

Table 2

Forecasting accuracy for daily electricity demand during 2018. The performance is measured by MSE, MSE/MSE_{min}, being the MSE divided by the minimum MSE among the different methods and MAPE. The minimum value of each row is remarked in bold.

One-step-ahead prediction									
				fKNN		EP			
	Average	Naive	Naive'	\hat{p}_{k^*}	$\hat{p}_{k^*, \theta=1}$	\hat{p}	$\hat{p}_{\theta=1}$	\hat{p}_{θ^*}	FPCF
MSE	10139732	8011267	4448292	1537719	1530080	2144232	1440256	1416428	5771944
$\frac{MSE}{MSE_{min}}$	7.158	5.655	3.140	1.086	1.080	1.514	1.017	1.000	4.075
MAPE	8.857	6.383	4.647	2.57953	2.59538	3.495	2.666	2.542	6.583
Dynamic updating									
				fKNN		EP			
	Average	Naive	Naive'	\hat{p}_{k^*}	$\hat{p}_{k^*, \theta=1}$	\hat{p}	$\hat{p}_{\theta=1}$	\hat{p}_{θ^*}	BM KRAUS
MSE	11545143	7805649	5245784	433620	426642	534009	414877	413508	2823846 512749.9
$\frac{MSE}{MSE_{min}}$	27.92	18.877	12.687	1.049	1.032	1.291	1.003	1.000	6.829 1.240
MAPE	9.061	6.183	5.035	1.55566	1.55052	1.79564	1.54113	1.53637	4.43493 1.7558

Table 3

Prediction bands for daily electricity demand during 2018 and NOx emissions during 2017. The performance is measured by mean coverage and Winkler score and results are reported for different values of α .

α	Electricity Demand					NoX Emission			
	One-day-ahead		Dynamic updating			One-day-ahead		Dynamic updating	
	EPF	fKNN	EPU	fKNN		EPF	fKNN	EPU	fKNN
0.05	Coverage	0.96	0.96	0.96	0.93	0.94	0.94	0.96	0.98
	(width)	6340.22	5581.59	3909.38	6871.60	88.64	81.07	99.56	95.59
	Winkler	7245.76	6200.67	4391.11	10326.45	103.94	94.95	108.16	97.85
0.10	Coverage	0.93	0.92	0.92	0.87	0.89	0.89	0.89	0.95
	(width)	5272.801	4477.16	3126.253	5881.41	71.16	68.91	60.41	79.47
	Winkler	6005.07	5097.22	3672.03	9969.236	87.23	84.90	71.70	83.81
0.20	Coverage	0.83	0.84	0.82	0.80	0.80	0.79	0.80	0.87
	(width)	4228.828	3355.45	2457.981	4794.38	56.93	54.54	47.43	64.04
	Winkler	5390.77	4185.80	3162.14	8344.702	75.13	75.06	60.34	71.79

Finally, MAPE of fKNN and EP turned out to be by far smaller than those of FAM, despite being based on data without preprocessed, without splitting the data into Weekdays, Saturdays, and Sundays, and without covariates.

Notably, Raña et al. [60] also report coverage and mean Winkler scores of prediction bands, providing an approximate order of magnitude of the results to compare them with the prediction bands based on projection methods introduced at the end of Section 2.3. Table 3 displays the scores resulting from projection methods for $\alpha = 0.05, 0.10$ and 0.20 . The first noteworthy result is that the mean coverages obtained are often above $(1 - \alpha)$. Second, the mean Winkler scores are often smaller than that those previously reported.

Also, we want to remark that the projection methods take into account the weekly and yearly periodicity of the data without any prior information. For illustrating this, we plot the frequency distribution of the number of days between predicted days and days used in the EP prediction in Fig. 4. The exponential weights properly averaged and scaled are also plotted in the same figure. This figure reveals the most relevant past information to forecast one-period-ahead and the electricity demand's seasonal structure. In particular, we observe that the most recent day contributes the most to the point forecast, followed by the same weekday but one week before (highest bars at the bottom panel). This weekly contribution vanishes until we approach the values of one year ago (highest bars at the top panel). For example, if we aim to forecast the demand for a given Wednesday, then the envelope would consider data from the last Tuesday, past Wednesdays, and same calendar days of other years in the same season, in order of importance.

3.3. NOx emissions in madrid city center

Air quality is nowadays in the spotlight due to its impact on human health [45]. Many cities gather almost continuous data of different air quality measures, such as nitrogen monoxide (NOx). The Cityhall of Madrid has recorded data since 2001 at different city points and has made them publicly available at <https://datos.madrid.es/portal/site/egob/>. In this section, we analyze the data from a station located in Plaza de España, Madrid downtown, because it has been functioning uninterrupted since the very beginning. By discarding days with fails on record, we end up with 6187 daily functions from 2001 to 2017 observed each hour.

Following [59], we evaluate these functions every 15 min by applying a positive smoothing with $K = 11$ Fourier basics functions (periodic data) and a λ obtained from minimizing the generalized cross-validation coefficient. As we did with

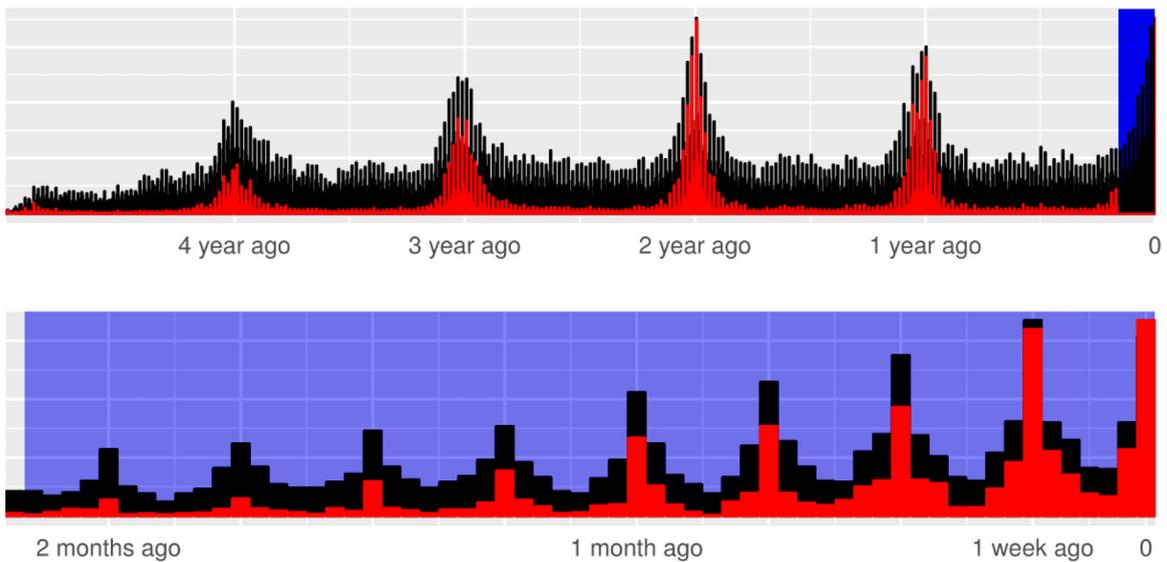


Fig. 4. Black bars: frequency distribution of the number of days between predicted day and the ones used in the EP prediction. Red bars: the normalized average of exponential weights used in the EP prediction. Bottom panel: zoom of the blue shaded area in the top panel. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4

Forecast accuracy for NOx emissions during 2017. The performance is measured by MSE/MSE_{min} , being the MSE divided by the minimum MSE among the different methods. The minimum value of each row is remarked in bold.

One-step-ahead									
				fKNN		EP			FPCF
MSE	Average	Naive	Naive'	\hat{p}_{k^*}	$\hat{p}_{k^*, \theta=1}$	\hat{p}	$\hat{p}_{\theta=1}$	\hat{p}_{θ^*}	
	661.5623	582.1691	945.5869	386.1636	386.2248	383.0795	387.2456	381.9375	502.6298
$\frac{MSE}{MSE_{min}}$	1.732	1.524	2.476	1.011	1.011	1.003	1.014	1.000	1.316
Dynamic updating									
				fKNN		EP			KRAUS
MSE	Average	Naive	Naive'	\hat{p}_{k^*}	$\hat{p}_{k^*, \theta=1}$	\hat{p}	$\hat{p}_{\theta=1}$	\hat{p}_{θ^*}	
	733.5805	512.6575	963.5414	257.2160	256.0309	273.2082	257.6278	257.7905	272.6422
$\frac{MSE}{MSE_{min}}$	2.8652	2.0023	3.7633	1.0046	1.0000	1.0671	1.0062	1.0069	1.0648

the Spanish electricity demand, we forecast the last recorded year (2017), predicting half a day for dynamic updating exercise. The results are also shown in Table 4.

Results for BM are not reported for the high computational cost involved. Neither MAPE is reported for avoiding divisions by zero. Once again, projection methods were superior to benchmark competitors in real case studies, as shown in Table 4.

4. Conclusion

The functional projection methods offer an empirical insight into forecasting functional time series of cyclic nature that repeat curve patterns and dependency patterns between contiguous curves. Both one-step-ahead prediction and dynamic updating are addressed in a unified way, without attending to any statistical model that could explain data generation's mechanism.

The projection methods were superior to benchmark methods from the functional time series literature for simulated and real data. The Spanish electricity demand curves illustrated how projection methods take into account structural features of the underlying functional time series, such as weekly and yearly periodicity, crucial in the literature of electricity demand forecasting. Particular characteristics of the electricity demand related to the calendar effect and outliers' presence were anticipated by the method for making accurate predictions. Although these peculiarities are known and often used to model and forecast electricity demand, they could be unknown for other data sets. In these cases, the projection methods may be particularly attractive.

According to the data set, the two projection methods were competitive with each other, being on average one better than the other. By construction, they are similar if the k -nearest curves surround the most recent functional datum (termed

here focal curve), making it deep. However, they differ if this curve is outlying with respect to their nearest curves. The method based on curve envelope projections performs better in the simulation and empirical data analyses in such cases. The superiority is because both punctual and band predictions are computed from curves surrounding the focal from below and above. However, for focal curves inlying with respect to their nearest curves, the method based on k -nearest neighbors projections can be even superior, making both methods similar in average, as we observed with the data sets considered in the previous section. This result reveals the common statistical tradeoff between efficiency and robustness that merits a separate study between both methods. Nevertheless, our main goal is to propose a parameter-free forecasting method for functional time series. We remark that the envelope method with $\theta = 1$ presents a competitive and entirely data-driven method. Based on our implementations and executions of the algorithms, the projection methods discussed here lead to substantive improvements in the computational time with respect to the benchmark. Additionally, the method could easily accommodate explanatory variables by-products of a suitable distance between multivariate functions and a multivariate functional depth, already available in the literature [15,29,35,49]. Finally, we note that the methodology is intuitive and data-driven, making the valuable approach to a broad audience, even without any prior knowledge of functional time series.

CRedit authorship contribution statement

Antonio Elías: Writing – original draft, Writing – review & editing, Conceptualization, Methodology, Software, Data curation, Visualization. **Raúl Jiménez:** Writing – original draft, Writing – review & editing, Conceptualization, Methodology, Supervision. **Han Lin Shang:** Writing – review & editing, Conceptualization, Methodology, Supervision.

Acknowledgments

The authors acknowledge insightful comments and suggestions from two reviewers and the Associate Editor. Antonio Elías is supported by the Spanish Ministerio de Educación, Cultura y Deporte under grant FPU15/00625 and the research stay grant EST17/00841. Antonio Elías and Raúl Jiménez are partially supported by the Spanish Ministerio de Economía y Competitividad under grant ECO2015-66593-P and PID2019-109196GB-I00/AEI/10.13039/501100011033. Funding for open access charge: Universidad de Málaga / CBUA. Part of this article was conducted during a stay at Australian National University. Antonio Elías is grateful to Han Lin Shang for his hospitality and insightful and constructive discussions.

Appendix

A.1. Functional autoregressive processes

With the primary aim of validating FPCF as a benchmark method, we repeated the simulation study conducted by Aue et al. [9]. This study was based on 100 realizations of eight classes of functional autoregressive (FAR) processes, 800 realizations in total. Each class is characterized by a pair of values (κ_1, κ_2) and a vector of standard deviations. Two vectors were chosen by the authors, denoted by (σ_1) and (σ_2) . See Section 6.3 of Aue et al. [9] for full description of these models. For generating FAR processes data, we used the `fChange` package [69]. For each realization, the study considered small n (from 180 to 199) and large n (from 900 to 999). This involves 2000 predictions for small n on each model and 10,000 for large n . We chose $q = 0.25$. Both for EPF and $EPU_{0.25}$ we considered $m = 100$, for small n , and $m = 500$, for large n . Smaller values of m (20 for small n and 100 for large n) were also considered and yield similar results. MSEs produced by the methods mentioned above are shown in Table 1.

From this table and Table 1¹ of Aue et al. [9], we conclude:

- FPCF is competitive with the best method reported by Aue et al. [9].
- FPCF has advantages over EPF.
- MSE based on $DB-U_{0.25}$ was by far the smallest.

We emphasize that FAR processes are not necessarily the kind of FTS that we have discussed at the end of Section 2.1, and consequently, neither EPF nor EPU should work well, although their performance is reasonably good. For the pair (0.8, 0.0), EPF was competitive. For these values, the future realization is generated only with the focal curve and innovations. Unlike (0.0, 0.8), the focal curve is not involved, making EPF a little competitive. For (0.4, 0.4), the difference of MSE based on FPCF and EPF was roughly 0.1, below to 7% of relative difference. For (0.2, 0.0), the FTS is mainly noise, and the average is the best prediction (see Table 5).

¹ There is a misprint in Table 1 of Aue et al. [9]: the results for (σ_1) and (σ_2) were switched (personal communication with the authors).

Table 5

MSE based on functional average, FPCF, EPF and EPU_{0.25}, for the two vector standard deviations, (σ_1) and (σ_2), chosen by Aue et al. [9]. The first row of each setting (κ_1, κ_2) corresponds to small n and the second row to large n .

κ_1	κ_2	σ_1				σ_2			
		Average	FPCF	EPF	EPU _{0.25}	Average	FPCF	EPF	EPU _{0.25}
0.2	0.0	1.63	1.65	1.69	0.92	2.30	2.33	2.42	2.02
		1.62	1.61	1.65	0.88	2.30	2.30	2.36	1.91
0.8	0.0	2.04	1.59	1.62	1.11	2.63	2.43	2.43	2.20
		2.14	1.65	1.67	1.05	2.41	2.32	2.37	2.09
0.4	0.4	1.92	1.74	1.84	0.70	2.47	2.45	2.54	2.13
		1.83	1.63	1.74	0.97	2.41	2.32	2.43	2.00
0.0	0.8	2.15	1.70	2.23	1.13	2.64	2.45	2.75	2.27
		2.19	1.66	2.26	1.03	2.57	2.36	2.67	2.11

A.2. Stationary FTS with random shocks

The noise X is generated with a stationary Gaussian process with zero mean and squared exponential autocovariance function

$$\text{Cov}(X(t), X(s)) = \exp\left(-\frac{|t - s|^2}{2l_x^2}\right).$$

This is the default autocovariance function in Gaussian processes simulation [61]. The parameter l_x determines the length of the ‘wiggles’ in the trajectories.

The random function f is a trajectory of a stationary Gaussian process with zero mean and periodic autocovariance function

$$\text{Cov}(f(t), f(s)) = \exp\left(-\frac{2 \sin^2(\pi|t - s|)}{l_f^2}\right).$$

The parameter l_f determines lengthscale in the same way as in the squared exponential autocovariance function. We emphasize that the trajectories of f are periodic functions of period 1.

The random shocks process S_μ is based on two independent processes. First, an occurrence process

$$\chi_\rho(t) = \sum_{i \geq 1} x_i \mathbb{1}_{[i-1, i)}(t),$$

$\{x_i, i \geq 0\}$ being a two states Markov chain, with $x_0 = 0$ and transition probabilities

$$\mathbb{P}(x_{i+1} = 1|x_i) = \rho(1 - x_i), \quad \mathbb{P}(x_{i+1} = 0|x_i) = 1 - \mathbb{P}(x_{i+1} = 1|x_i).$$

Second, a random periodic function for modeling irregular shock patterns. Specifically, we use $Z(t) = \sigma_g^2(g^2(t) - g^2(0))$, g being an independent copy of f . The parameter σ_g determines the average distance of g away from zero. The squares are taking for generating a non-zero mean pattern, and $g^2(0)$ is subtracted for beginning to shock from zero. The random shock at time t is then

$$S_\mu(t) = \begin{cases} 0, & \text{for } t < u \\ Z(t - u)\chi_\rho(t - u), & \text{for } t \geq u \end{cases}$$

u being the minimum time that can elapse before the first shock and $\rho = \mu/(2 - \mu)$. In our simulations, u is uniformly distributed on $[0, 1]$. Since the stationary distribution of $\{x_i\}$ is $\mathbb{P}(x_i = 1) = \rho/(1 + \rho)$ and each shock affects two contiguous periods with probability equal to 1, the proportion of periods affected by shocks, when $t \rightarrow \infty$, is μ .

The parameters considered for examples and simulations are $l_f = 1$, $l_x = 0.2$, and $\sigma_g^2 = 10$.

References

- [1] G. Aneiros, R. Cao, R. Fraiman, C. Genest, P. Vieu, Recent advances in functional data analysis and high-dimensional statistics, *J. Multivariate Anal.* 170 (2019) 3–9, Special Issue on Functional Data Analysis and Related Topics.
- [2] G. Aneiros, R. Cao, J.M. Vilar, Functional methods for time series prediction: a nonparametric approach, *J. Forecast.* 30 (4) (2011) 377–392.
- [3] G. Aneiros, P. Vieu, Nonparametric time series prediction: A semi-functional partial linear modeling, *J. Multivariate Anal.* 99 (5) (2008) 834–857.
- [4] G. Aneiros, J.M. Vilar, R. Cao, A. Muñoz, Functional prediction for the residual demand in electricity spot markets, *IEEE Trans. Power Syst.* 28 (4) (2013) 4201–4208.
- [5] G. Aneiros, J.M. Vilar, P. Raña, Short-term forecast of daily curves of electricity demand and price, *Int. J. Electr. Power Energy Syst.* 80 (2016) 96–108.
- [6] J. Antoch, L. Prchal, M.R. De Rosa, P. Sarda, Electricity consumption prediction with functional linear regression using spline estimators, *J. Appl. Stat.* 37 (12) (2010) 2027–2041.

- [7] A. Antoniadis, E. Paparoditis, T. Sapatinas, A functional wavelet-kernel approach for time series prediction, *J. R. Stat. Soc. Ser. B Stat. Methodol.* 68 (5) (2006) 837–857.
- [8] A. Arribas-Gil, J. Romo, Shape outlier detection and visualization for functional data: the outliergram, *Biostatistics* 15 (4) (2014) 603–619.
- [9] A. Aue, D.D. Norinho, S. Hörmann, On the prediction of stationary functional time series, *J. Am. Stat. Assoc.: Theory Methods* 110 (509) (2015) 378–392.
- [10] G. Biau, F. Cerou, A. Guyader, Rates of convergence of the functional k -nearest neighbor estimate, *IEEE Trans. Inform. Theory* 56 (4) (2010) 2034–2040.
- [11] D. Bosq, *Linear Processes in Function Spaces: Theory and Applications*, Springer, New York, 2000.
- [12] F. Burba, F. Ferraty, P. Vieu, K -nearest neighbour method in functional nonparametric regression, *J. Nonparametr. Stat.* 21 (4) (2009) 453–469.
- [13] F. Cérou, A. Guyader, Nearest neighbor classification in infinite dimension, *ESAIM: PS* 10 (2006) 340–355.
- [14] H. Cho, Y. Goude, X. Brossat, Q. Yao, Modeling and forecasting daily electricity load curves: A hybrid approach, *J. Am. Stat. Assoc.: Appl. Case Stud.* 108 (501) (2013) 7–21.
- [15] G. Claeskens, M. Hubert, L. Slaets, K. Vakili, Multivariate functional halfspace depth, *J. Amer. Statist. Assoc.* 109 (505) (2014) 411–423.
- [16] J.A. Cuesta-Albertos, M. Febrero-Bande, M. Oviedo de la Fuente, The DD^c -classifier in the functional setting, *TEST* 26 (1) (2017) 119–142.
- [17] J. Cuesta-Albertos, A. Nieto-Reyes, The random Tukey depth, *Comput. Statist. Data Anal.* 52 (11) (2008) 4979–4988.
- [18] A. Cuevas, A partial overview of the theory of statistics with functional data, *J. Statist. Plann. Inference* 147 (2014) 1–23.
- [19] A. Cuevas, M. Febrero-Bande, R. Fraiman, Robust estimation and classification for functional data via projection-based depth notions, *Comput. Statist.* 22 (3) (2007) 481–496.
- [20] W. Dai, M.G. Genton, Multivariate functional data visualization and outlier detection, *J. Comput. Graph. Statist.* 27 (4) (2018) 923–934.
- [21] M. Febrero-Bande, P. Galeano, W. González-Manteiga, Outlier detection in functional data by depth measures, with application to identify abnormal NOx levels, *Environmetrics* 19 (4) (2008) 331–345.
- [22] F. Ferraty, S. Nagy, Scalar-on-function local linear regression and beyond, *Biometrika* (2021).
- [23] F. Ferraty, P. Vieu, *Nonparametric Functional Data Analysis: Theory and Practice*, Springer-Verlag New York, 2006.
- [24] R. Fraiman, G. Muniz, Trimmed means for functional data, *Test* 10 (2) (2001) 419–440.
- [25] I. Gijbels, S. Nagy, On a general definition of depth for functional data, *Statist. Sci.* 32 (4) (2017) 630–639.
- [26] T. Gneiting, A. Raftery, Strictly proper scoring rules, prediction, and estimation, *J. Am. Stat. Assoc.: Rev. Article* 102 (477) (2007) 359–378.
- [27] A. Goia, P. Vieu, An introduction to recent advances in high/infinite dimensional statistics, *J. Multivariate Anal.* 146 (2016) 1–6, Special Issue on Statistical Models and Methods for High or Infinite Dimensional Spaces.
- [28] S. Hörmann, P.P. Kokoszka, Functional time series, in: *Time Series Analysis*, in: *Handbook of Statistics*, 30, Elsevier B.V., Netherlands, 2012, pp. 157–186.
- [29] M. Hubert, P. Rousseeuw, P. Segaeert, Multivariate and functional classification using depth and distance, *Adv. Data Anal. Classif.* 11 (3) (2017) 445–466.
- [30] R.J. Hyndman, H. Booth, Stochastic population forecasts using functional data models for mortality, fertility and migration, *Int. J. Forecast.* 24 (3) (2008) 323–342.
- [31] R.J. Hyndman, H.L. Shang, Forecasting functional time series, *J. Korean Stat. Soc.* 38 (3) (2009) 199–211.
- [32] R.J. Hyndman, H.L. Shang, Rainbow plots, bagplots, and boxplots for functional data, *J. Comput. Graph. Stat.* 19 (1) (2010) 29–45.
- [33] R.J. Hyndman, H.L. Shang, *ftsa: Functional time series analysis*, 2020, URL <https://CRAN.R-project.org/package=ftsa> R package version 6.0.
- [34] R.J. Hyndman, M. Ullah, Robust forecasting of mortality and fertility rates: A functional data approach, *J. Comput. Graph. Stat.* 51 (10) (2007) 4942–4956.
- [35] F. Ieva, A.M. Paganoni, Depth measures for multivariate functional data, *Comm. Statist. Theory Methods* 42 (7) (2013) 1265–1276.
- [36] L.-Z. Kara, A. Laksaci, M. Rachdi, P. Vieu, Data-driven kNN estimation in nonparametric functional data analysis, *J. Multivariate Anal.* 153 (2017) 176–188.
- [37] J. Klepsch, C. Klüppelberg, An innovations algorithm for the prediction of functional linear processes, *J. Multivariate Anal.* 155 (2017) 252–271.
- [38] P. Kokoszka, M. Reimherr, *Introduction To Functional Data Analysis*, first ed., Chapman and Hall/CRC, 2017.
- [39] D. Kraus, Components and completion of partially observed functional data, *J. R. Stat. Soc. B* 77 (4) (2015) 777–801.
- [40] N.L. Kudraszow, P. Vieu, Uniform consistency of kNN regressors for functional variables, *Statist. Probab. Lett.* 83 (8) (2013) 1863–1870.
- [41] D. Li, P.M. Robinson, H.L. Shang, Local whittle estimation of long-range dependence for functional time series, *J. Time Series Anal.* (2020a).
- [42] D. Li, P.M. Robinson, H.L. Shang, Long-range dependent curve time series, *J. Am. Stat. Assoc.: Theory Methods* 115 (530) (2020b) 957–971.
- [43] H. Lian, Convergence of functional k -nearest neighbor regression estimate with functional responses, *Electron. J. Stat.* 5 (2011) 31–40.
- [44] N. Ling, P. Vieu, Nonparametric modelling for functional data: selected survey and tracks for future, *Statistics* 52 (4) (2018) 934–949.
- [45] Liu, et al., Ambient particulate air pollution and daily mortality in 652 cities, *N. Engl. J. Med.* 381 (8) (2019) 705–715.
- [46] S. López-Pintado, J. Romo, Depth-based classification for functional data, in: R.S. Regina Y. Liu, D.L. Souvaine (Eds.), *Data Depth: Robust Multivariate Analysis, Computational Geometry and Applications*, Vol. 72, AMS and DIMACS, 2006, p. 103.
- [47] S. López-Pintado, J. Romo, On the concept of depth for functional data, *J. Am. Stat. Assoc.: Theory Methods* 104 (486) (2009) 718–734.
- [48] S. López-Pintado, J. Romo, A. Torrente, Robust depth-based tools for the analysis of gene expression data, *Biostatistics* 11 (2) (2010) 254–264.
- [49] S. López-Pintado, Y. Sun, J.K. Lin, M.G. Genton, Simplicial band depth for multivariate functional data, *Adv. Data Anal. Classif.* 8 (3) (2014) 321–338.
- [50] H. Lütkepohl, *New Introduction to Multiple Time Series Analysis*, Springer-Verlag, New York, 2006.
- [51] S. Makridakis, E. Spiliotis, V. Assimakopoulos, Statistical and machine learning forecasting methods: Concerns and ways forward, *PLOS ONE* 13 (3) (2018) 1–26.
- [52] F. Martínez, M.P. Frías, F. Charte, A.J. Rivera, A methodology for applying k -nearest neighbor to time series forecasting, *Artif. Intell. Rev.* 52 (3) (2019a) 2019–2037.
- [53] F. Martínez, M.P. Frías, F. Charte, A.J. Rivera, Time series forecasting with KNN in r: the tsfkn package, *R J.* 11 (2) (2019b) 229–242.
- [54] S. Nagy, I. Gijbels, D. Hlubinka, Depth-based recognition of shape outlying functions, *J. Comput. Graph. Statist.* 26 (4) (2017) 883–893.
- [55] N.N. Narisetty, V.N. Nair, Extremal depth for functional data and applications, *J. Am. Stat. Assoc.: Theory Methods* 111 (516) (2015) 1705–1714.
- [56] A. Nieto-Reyes, H. Bettey, A topologically valid definition of depth for functional data, *Statist. Sci.* 31 (1) (2016) 61–79.
- [57] E. Paparoditis, T. Sapatinas, Short-term load forecasting: The similar shape functional time-series predictor, *IEEE Trans. Power Syst.* 28 (4) (2013) 3818–3825.
- [58] C.T. Perretti, S.B. Munch, G. Sugihara, Model-free forecasting outperforms the correct mechanistic model for simulated and experimental data, *Proc. Natl. Acad. Sci.* 110 (13) (2013) 5253–5257.
- [59] J. Ramsay, B. Silverman, *Functional Data Analysis*, second ed., Springer, New York, 2005.
- [60] P. Raña, J. Vilar, G. Aneiros, On the use of functional additive models for electricity demand and price prediction, *IEEE Access* 6 (2018) 9603–9613.
- [61] C.E. Rasmussen, C.K.I. Williams, *Gaussian Processes for Machine Learning*, The MIT Press, Cambridge, 2005.

- [62] R. Serfling, U. Wijesuriya, Depth-based nonparametric description of functional data, with emphasis on use of spatial depth, *Comput. Statist. Data Anal.* 105 (2017) 24–45.
- [63] H.L. Shang, Functional time series approach for forecasting very short-term electricity demand, *J. Appl. Stat.* 40 (1) (2013) 152–168.
- [64] H.L. Shang, Functional time series forecasting with dynamic updating: An application to intraday particulate matter concentration, *Econometrics Stat.* 1 (2017) 184–200.
- [65] H.L. Shang, Visualizing rate of change: an application to age-specific fertility rates, *J. R. Stat. Soc. Ser. A* 182 (1) (2019) 249–262.
- [66] H.L. Shang, R.J. Hyndman, Nonparametric time series forecasting with dynamic updating, *Math. Comput. Simulation* 81 (7) (2011) 1310–1324.
- [67] H.L. Shang, Y. Yang, F. Kearney, Intraday forecasts of a volatility index: functional time series methods with dynamic updating, *Ann. Oper. Res.* 282 (2018) 331–354.
- [68] S.K. Singh, H. McMillan, A. Bádosy, C. Fateh, Nonparametric catchment clustering using the data depth function, *Hydrol. Sci. J.* 61 (15) (2016) 2649–2667.
- [69] O. Sonmez, A. Aue, G. Rice, fChange: Change point analysis in functional data, 2019, URL <https://cran.r-project.org/package=fChange> R package version 0.2.1, removed from CRAN.
- [70] G. Sugihara, B.T. Grenfell, R.M. May, H. Tong, Nonlinear forecasting for the classification of natural time series, *Philos. Trans. R. Soc. Lond. Ser. A Phys. Eng. Sci.* 348 (1688) (1994) 477–495.
- [71] G. Sugihara, R.M. May, Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series, *Nature* 344 (6268) (1990) 734–741.
- [72] Y. Sun, M.G. Genton, Functional boxplots, *J. Comput. Graph. Stat.* 20 (2) (2011) 316–334.
- [73] Y. Sun, M.G. Genton, D.C. Nychka, Exact fast computation of band depth for large functional datasets: How quickly can one million curves be ranked? *Stat* 1 (2012) 68–74.
- [74] N. Tarabelloni, F. Ieva, R. Biasi, A.M. Paganoni, Use of depth measure for multivariate functional data in disease prediction: An application to electrocardiograph signals, *Int. J. Biostat.* 11 (2015) 189–201.
- [75] L.L. Tupper, D.S. Matteson, C.L. Anderson, L. Zephyr, Band depth clustering for nonstationary time series and wind speed behavior, *Technometrics* 60 (2) (2017) 245–254.
- [76] J. Vilar, G. Aneiros, P. Raña, Prediction intervals for electricity demand and price using functional data, *Int. J. Electr. Power Energy Syst.* 96 (2018) 457–472.
- [77] J.M. Vilar, R. Cao, G. Aneiros, Forecasting next-day electricity demand and price using nonparametric functional methods, *Int. J. Electr. Power Energy Syst.* 39 (1) (2012) 48–55.
- [78] J.-L. Wang, J.-M. Chiou, H.-G. Müller, Functional data analysis, *Annu. Rev. Stat. Appl.* 3 (1) (2016) 257–295.
- [79] E. Wang, D. Cook, R.J. Hyndman, Calendar-based graphics for visualizing people's daily schedules, *J. Comput. Graph. Statist.* 29 (3) (2020) 1–13.
- [80] R.L. Winkler, A decision-theoretic approach to interval estimation, *J. Am. Stat. Assoc.: Theory Methods* 67 (337) (1972) 187–191.
- [81] F. Yao, H.G. Müller, J.L. Wang, Functional data analysis for sparse longitudinal data, *J. Am. Stat. Assoc.: Theory Methods* 100 (470) (2005) 577–590.
- [82] G. Zhang, B. Eddy Patuwo, M. Y. Hu, Forecasting with artificial neural networks: The state of the art, *Int. J. Forecast.* 14 (1) (1998) 35–62.
- [83] S. Zhang, W. Jank, G. Shmueli, Real-time forecasting of online auctions via functional K-nearest neighbors, *Int. J. Forecast.* 26 (4) (2010) 666–683.