

# Nonparametric trend estimation in functional time series with application to annual mortality rates

Israel Martínez-Hernández  | Marc G. Genton 

King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

## Correspondence

Israel Martínez-Hernández, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia.  
 Email: [israel.martinezhernandez@kaust.edu.sa](mailto:israel.martinezhernandez@kaust.edu.sa)

## Abstract

We address the problem of trend estimation for functional time series. Existing contributions either deal with detecting a functional trend or assuming a simple model. They consider neither the estimation of a general functional trend nor the analysis of functional time series with a functional trend component. Similarly to univariate time series, we propose an alternative methodology to analyze functional time series, taking into account a functional trend component. We propose to estimate the functional trend by using a tensor product surface that is easy to implement, to interpret, and allows to control the smoothness properties of the estimator. Through a Monte Carlo study, we simulate different scenarios of functional processes to show that our estimator accurately identifies the functional trend component. We also show that the dependency structure of the estimated stationary time series component is not significantly affected by the error approximation of the functional trend component. We apply our methodology to annual mortality rates in France.

## KEYWORDS

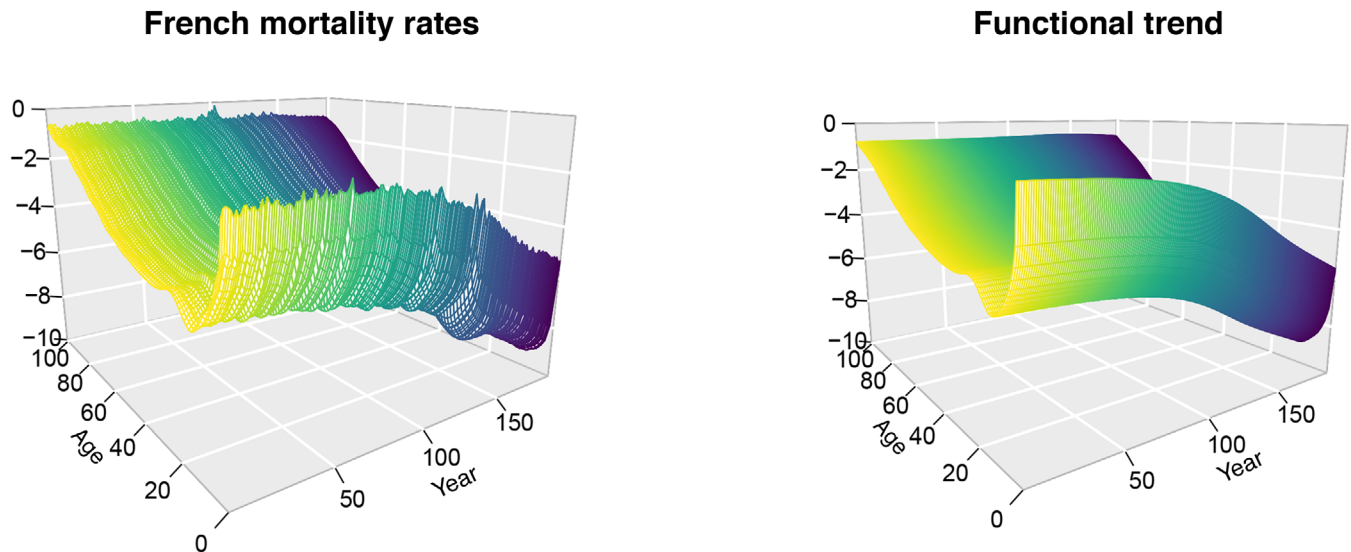
annual mortality rate, detrending functional time series, nonparametric estimator, nonstationary functional time series, penalized tensor product surface

## 1 | INTRODUCTION

In many phenomena, data are collected on a large scale, resulting in high-dimensional and high-frequency data. This is why there has been an increasing amount of interest in functional data analysis (FDA). FDA deals with data, called functional data, that are defined on an intrinsically infinite-dimensional space. When the functional data are time dependent, they are called functional time series. Some examples of data that can be considered as functional time series are the annual mortality rates and the annual temperature data. In practice, functional time series often tend to be nonstationary. This nonstationarity may be caused by structural breaks, functional random walk components, or deterministic trend components. Deterministic trends, or functional trends, can be observed

in different phenomena where functional data approaches have been used, for example, growth curves (Ramsay and Silverman, 2005), annual mortality rates (Hyndman and Ullah, 2007), gene networks (Telesca *et al.*, 2009), climate change (Fraiman *et al.*, 2014), electricity power systems (Horváth and Rice, 2015), and electroencephalography (EEG) data (Hasenstab *et al.*, 2017). The detection and estimation of the functional trend are crucial in data analysis, modeling, and forecasting.

The common method used to analyze functional time series involves projecting each curve on a finite dimensional space, for example, on the space generated by  $r$  eigenfunctions, and then modeling the projected values by using multivariate time series techniques (Hyndman and Ullah, 2007; Aue *et al.*, 2015). When the functional time series has a functional trend component, one can still



**FIGURE 1** Functional time series of log mortality rates in France from 1816 to 2006, for zero to 100 years of age (left), and the corresponding estimated functional trend (right). The estimated functional trend describes the smooth changes over time of the functional data

transform the curves into a vector and then model the trend component as in multivariate time series. However, using principal component analysis to reduce dimensionality may not be appropriate, since the estimation of the covariance operator is not consistent in this case. An alternative approach, similar to the univariate time series, is to estimate the functional trend directly from the functional data, then remove it, and analyze the remaining functional time series. In this paper, we adopt the latter approach.

Functional trends are challenging because of the complexity of the space where functional data are defined. In multivariate time series, trends have only one component, that is, they have the form  $h(t)$ , where  $t$  represents time, and  $h$  is a continuous function defined over time (see, eg, Wu and Zhao, 2007; Chen and Wu, 2019). Unlike in multivariate time series, functional trends have an additional component: the continuous parameter of each functional data. That is, functional trends can be written as a function with two variables  $T(s, t)$ , where  $s$  is the continuous parameter of each curve and  $t$  represents time.

A few attempts can be found in the literature on the study of functional trends. In Fraiman *et al.* (2014), a functional trend was defined by using the concept of records, where a record means the occurrence of new extreme observations, but nothing was mentioned about the estimation. In Kokoszka and Young (2017), a hypothesis test of trend stationarity of functional time series was proposed. In that paper, the functional trend was assumed to be separable and linear in time,  $T(s, t) = f(s)t$ , and a least squares estimator was used to estimate  $f(s)$ . Although this may cover a large number of cases, which depend linearly on time, it is still a very specific model. A functional trend can take very complex shapes, for example, Figure 1 shows log

annual mortality rates in France from 1816 to 2006, where each point of  $Y_n(s)$  represents the total mortality rate, in year  $n$ , at age  $s \in [0, 100]$ . Across the years  $n$ , the log mortality rate has been decreasing for almost all ages  $s$ . For ages between 0 and 60, it seems that the decrease behaves like a quadratic function, whereas for ages between 60 and 100 the values behave like a linear function. On the other hand, the  $s$  coordinate (age) is dominated by a U-shaped curve for each  $n$ . The right panel shows the resulting functional trend estimated by applying our proposed methodology. Here we analyze these data as a functional time series considering the functional trend  $T(s, t)$  (Section 5). Due to the complexity of functional trends, we propose describing  $T(s, t)$  using a nonparametric approach.

The functional time series approach has several advantages over the multivariate time series methods. Multivariate methods ignore information about the underlying continuity behavior of the data. For example, the bivariate time series of the annual mortality rates at ages  $s = 40$  and  $s = 41$ ,  $\{Y_n(40), Y_n(41)\}^T$ , is permutable in the multivariate setting. This leads to a rough surface for a functional trend estimator. In contrast, smoothness is an important property of functional data. Thus, FDA extracts additional information contained in a continuous function or in its derivative (Kokoszka (2012); Ullah and Finch (2013)).

There is still a gap in knowledge on functional trends in functional time series. To the best of our knowledge, previous research either involved detecting functional trends or assuming a simple model, but none involved estimating a general functional trend nor the analysis of functional time series with a functional trend component. Here, we describe a methodology to estimate the functional trend, and we show the analysis of the functional time series

when the trend is taken into account. We propose a method for estimating a functional trend that is easy to implement and to interpret and allows to control the smoothness properties of the estimator, which is useful in practice.

For instance, assume that  $t$  is fixed in  $T(s, t)$ ; thus  $T(\cdot, t)$  can be interpreted as the “common” curve that persists in different ways over time, weighted with the  $t$  component. For example, if the weight function is additive, that is,  $T(s, t) = f(s) + g(t)$ , then  $f(s)$  can be considered as the mean curve and consequently the functional trend is simply  $g(t)$ . Now, if we fix  $s \in D$ , where  $D$  represents the domain of the functional data,  $T(s, \cdot)$  is the trend over time, and it can take different forms for each  $s \in D$ . Therefore, for each coordinate,  $T(s, t)$  can take different shapes, and a nonparametric estimation for each coordinate seems reasonable. We propose using a B-spline to describe the different forms for each coordinate. When the sample size tends to infinity,  $T$  can be assumed to be continuous in  $s$  and  $t$ , and resulting in a tensor product surface. To obtain the smoothness property of the tensor product B-spline, similar ideas from the univariate case (Eilers and Marx, 1996) can be applied. One can opt to use one penalty parameter for both directions, or one for each direction, or a combination of both (see Wood, 2003; Xiao *et al.*, 2013). Here, we consider marginal penalizations as described in Wood (2006). This allows us to study the trend over time and a possible trend within the domain  $D$  separately. Also, this way of penalizing is easy to interpret and to control for each smoothness parameter.

The remainder of our paper is organized as follows. In Section 2, we introduce the model that is assumed in this paper, and we develop the proposed estimator for the functional trend. In Section 3, we study the theoretical properties of the proposed estimator as well as the selection of the smoothing parameters. In Section 4, we conduct a simulation study to evaluate the performance of the proposed estimator under different simulation settings. In Section 5, we analyze a dataset of annual mortality rates assuming a functional trend component. Section 6 presents some discussion. Proofs and additional material are provided in the Web Appendix.

## 2 | TREND IN FUNCTIONAL TIME SERIES

### 2.1 | Preliminaries

Assume that we observe a functional time series with a sample size  $N$ ,  $\{Y_1, \dots, Y_N\}$ , taking values on a separable Hilbert space  $\mathcal{H}$  that will be defined in Section 3.2, that is,  $Y_n(s) : D \rightarrow \mathbb{R}$  is a continuous function for  $n = 1, \dots, N$ .

Now, assume that  $\{Y_n\}$  follows the model

$$Y_n(s) = T(s, n/N) + X_n(s), \quad (1)$$

where  $T(s, t) : D \times [0, 1] \rightarrow \mathbb{R}$  is a deterministic function and  $\{X_n\}$  is a stationary functional time series with  $\mathbb{E}(X_n) = 0$ . Thus,  $\mathbb{E}(Y_n) = T(s, n/N)$  and  $\{Y_n\}$  is not weakly stationary. The function  $T(s, t)$  is the trend component.

A technique that is widely used in time series to obtain the stationarity property is considering the first difference of  $\{Y_n, n \geq 1\}$ , that is,  $\Delta Y_n := Y_n - Y_{n-1}$ . If the functional time series has a random walk component or if it is a  $I(1)$  functional process,  $\{\Delta Y_n\}$  is stationary (Beare *et al.*, 2017). However, if the nonstationary component is a deterministic function, as in model (1), the transformation  $\{\Delta Y_n\}$  does not guarantee to remove the trend component  $T(s, t)$ . Moreover,  $\Delta X_n$  might be nonstationary even though  $\{X_n\}$  is stationary, and as a consequence  $\{\Delta Y_n\}$  might be nonstationary. To clarify the above idea, assume for instance that  $T(s, t) = \sin(2\pi t + s)$  in model (1). Thus,  $T(s, \frac{n}{N}) - T(s, \frac{n-1}{N})$  depends on  $n$ , and then  $\Delta Y_n$  depends on  $n$  as well. Therefore, the estimation of the functional trend  $T(s, t)$  is necessary.

### 2.2 | Nonparametric functional trend estimator

We observe that, for  $n_0$  fixed in model (1),  $Y_{n_0}(\cdot) = T(\cdot, n_0/N) + X_{n_0}(\cdot)$ . Thus  $T(\cdot, n_0/N)$  represents the mean curve of the functional data  $Y_{n_0}$  at time  $n_0$ . If  $s_0 \in D$  is fixed, then  $\{Y_n(s_0), n = 1, \dots, N\}$  is a univariate time series and  $T(s_0, \cdot)$  represents the deterministic trend at  $s_0$ . In the latter case,  $T(s_0, \cdot)$  can be obtained via nonparametric estimation, such as Nadaraya-Watson, local polynomial, wavelet, or spline methods. Here we use the spline method, that is, we assume that  $T(s_0, \cdot) = \sum_{i=1}^{k_2} b_i \eta_i(\cdot) = \mathbf{b}^\top \boldsymbol{\eta}(\cdot)$ , where  $\boldsymbol{\eta}^\top = (\eta_1, \dots, \eta_{k_2})$  is a B-spline basis function defined on  $[0, 1]$ .

Similarly, one could repeat this procedure for a finite set of  $s$  values and apply a multivariate time series technique. However, since  $Y_n$  is assumed to be a continuous function in  $s$ , multivariate methods cannot be extended to functional data. Multivariate methods ignore the continuity (smoothness) property of  $Y_n$ , that is,  $Y_n(s_0)$  and  $Y_n(s_0 + \epsilon)$  are considered permutable for any  $\epsilon > 0$ . In addition, these would involve estimating infinite parametric or nonparametric tendencies. Instead, we allow each coefficient  $b_i$  to be a smooth continuous function of  $s$ , that is,  $T(s, \cdot) = \mathbf{b}^\top(s) \boldsymbol{\eta}(\cdot)$ , and  $b_i(s)$  can be modeled nonparametrically as well. Let  $\boldsymbol{\nu}^\top = (\nu_1, \dots, \nu_{k_1})$  be another B-spline basis function defined on  $D$ , such that  $b_i(s) = \sum_{j=1}^{k_1} \theta_{ji} \nu_j(s)$  for

$i = 1, \dots, k_2$ . Then,  $T(s, t)$  can be written as

$$T(s, t) = \sum_{j=1}^{k_1} \sum_{i=1}^{k_2} \theta_{ji} \nu_j(s) \eta_i(t) = \boldsymbol{\nu}^\top(s) \boldsymbol{\Theta} \boldsymbol{\eta}(t). \quad (2)$$

We propose estimating the functional trend by using a tensor product of the two spaces  $\text{span}\{\nu_1, \dots, \nu_{k_1}\}$  and  $\text{span}\{\eta_1, \dots, \eta_{k_2}\}$ . To obtain smoothness properties of  $T(s, t)$ , we consider penalty terms associated with each coordinate (Wood, 2006). That is,

$$P(T) = \lambda_1 \int_{[0,1]} (P_1 T)(t) dt + \lambda_2 \int_D (P_2 T)(s) ds, \quad (3)$$

where  $P_1 T = \int \{ \frac{\partial^2}{\partial s^2} T(s, t) \}^2 ds$  and  $P_2 T = \int \{ \frac{\partial^2}{\partial t^2} T(s, t) \}^2 dt$ . Other quadratic penalties can be considered, such as  $\int \int \{ (LT)(t, s) \}^2 ds dt$ , with  $L$  a linear operator (eg, the Laplacian). Here, we adopt the marginal penalty (3), where  $\lambda_1$  and  $\lambda_2$  control the smoothness of  $T(s, t)$  in the first component and the second component, respectively. This penalty is invariant to a linear rescaling of the functional data, which is useful since, in practice the domain  $D$  of the functions is rescaled to the interval  $[0,1]$ . Also,  $P(T)$  is easily interpretable and allows us to control the smoothness in the direction of the domain  $D$  and in the direction of the time domain, separately, which is desirable for the estimation of the functional trend.

We observe that if  $\lambda_1 \gg 0$ , then  $T(\cdot, n/N)$  is a linear function on  $D$  for each  $n = 1, \dots, N$ , and if  $\lambda_1 = 0$ , then  $T(\cdot, n/N)$  is close to the shape of the functional data  $Y_n$ , that is,  $T(\cdot, n/N) \approx Y_n$ . Thus, to only capture the trend over time and without removing the inherent shape of the functional data, a  $\lambda_1$  different from zero should be considered. Similarly, if  $\lambda_2 \gg 0$ , then  $T(s, \cdot)$  represents a linear trend for each  $s$ , whereas when  $\lambda_2 = 0$ , then  $T(s, \cdot)$  represents interpolation of  $Y_1(s), \dots, Y_n(s)$  for each  $s$ , and so  $T(s, t)$  results in a rough surface. In Section 3.3, we describe how to select these parameters taking into account the dependency structure of  $\{X_n\}$ . In practice, users are free to choose the values of  $\lambda_1$  and  $\lambda_2$  as well as the number of basis functions in each coordinate,  $k_1$  and  $k_2$ .

Given  $P(T)$  we obtain the estimator of  $T(s, t)$  by using a penalized least square estimator, that is, we obtain  $\hat{\Theta}$  minimizing the mean integrated squared error

$$\hat{\Theta} = \arg \min_{\Theta} \left[ \sum_{n=1}^N \int_D \{ Y_n(s) - \boldsymbol{\nu}^\top(s) \boldsymbol{\Theta} \boldsymbol{\eta}(n/N) \}^2 ds + P(T) \right]. \quad (4)$$

Consequently, we define  $\hat{T}(s, t) = \boldsymbol{\nu}^\top(s) \hat{\Theta} \boldsymbol{\eta}(t)$ .

In summary, we propose describing the deterministic trend in functional time series by using a smooth tensor

product surface. A tensor product surface is very flexible in the sense that it can represent complex structures in functional data. Because of the penalization term, a few numbers of basis functions (or knots) are required, and it is computationally feasible. In Section 4, we show the performance of our proposed estimator under different scenarios.

### 2.3 | Modeling with estimated functional trend

Once the functional trend has been estimated, we make an  $h$ -step ahead forecast for the functional time series  $\{Y_n\}$  by forecasting each component of the model (1), that is,  $\hat{Y}_{N+h} = \hat{T}_{N+h} + \hat{X}_{N+h}$ . The  $h$ -step ahead forecast for each component is computed as follows: For the stationary functional time series component, we obtain  $\hat{X}_{N+h}$  by modeling the functional time series  $\{\tilde{X}_n := Y_n(s) - \hat{T}(s, n/N)\}_{n=1}^N$ . For example, one can use the methodology described in Aue *et al.* (2015) (see Section 5). To obtain the  $h$ -step ahead forecast for the functional trend component, we use a Taylor expansion in the time direction. Specifically, we define the 1-step ahead forecast as

$$\hat{T}_{N+1}(s) := \hat{T}(s, 1) + \frac{1}{N+1} \frac{\partial}{\partial t} T(s, t) \Big|_{t=1}, \quad (5)$$

where  $\hat{T}(s, 1)$  corresponds to the trend estimated at time  $N$ . This 1-step ahead forecast is iterated  $h$  times, with  $\hat{T}(s, 1)$  being the last trend observed or forecasted in each iteration. After the iterations, we obtain the  $h$ -step ahead forecast  $\hat{T}_{N+h}$ . In general,  $T(s, t)$  can be assumed to be a function with slow variation over time, as evidenced in Figure 1. Thus, in this paper we use the linear approximation (5).

## 3 | THEORETICAL PROPERTIES

The theoretical properties of penalized splines have been studied when errors are uncorrelated. For the one-dimensional setting, see, for example, Li and Ruppert (2008), and Claeskens *et al.* (2009). Some papers that have studied the two-dimensional setting are Lai and Wang (2013) and Xiao (2019). Xiao (2019) studied the asymptotic behavior of bivariate penalized tensor-product splines, extending the idea from the one-dimensional setting. Here, we adopt the same approach as in Xiao (2019) to study the consistency of the functional trend estimator  $\hat{T}(s, t)$ .

Let  $\mathcal{P}_1$  and  $\mathcal{P}_2$  be the fixed marginal penalty matrices, for the first component and the second component of  $T(s, t)$ , respectively. Thus, the first component of the penalty term in (3) can be written as  $\int (P_1 T)(t) dt = \int \{ \boldsymbol{\Theta} \boldsymbol{\eta}(t) \}^\top$

$\mathcal{P}_1 \Theta \eta(t) dt = \{\text{vec}(\Theta)\}^\top \mathbf{J}_\eta \otimes \mathcal{P}_1 \text{vec}(\Theta)$ , and the second component as  $\int (P_2 T)(s) ds = \int \nu^\top(s) \Theta \mathcal{P}_2 \Theta^\top \nu(s) ds = \{\text{vec}(\Theta)\}^\top \mathcal{P}_2 \otimes \mathbf{J}_\nu \text{vec}(\Theta)$ , where  $\mathbf{J}_\nu = \int \nu(s) \nu^\top(s) ds$ , and  $\mathbf{J}_\eta = \int \eta(t) \eta^\top(t) dt$ . Therefore,

$$P(T) = \{\text{vec}(\Theta)\}^\top \{\lambda_1 \mathbf{J}_\eta \otimes \mathcal{P}_1 + \lambda_2 \mathcal{P}_2 \otimes \mathbf{J}_\nu\} \text{vec}(\Theta), \quad (6)$$

where  $\lambda_1$  and  $\lambda_2$  are the smoothing parameters and they need to be estimated.

### 3.1 | Functional representation

We assume that the functional time series  $Y_n(s)$  are given in the functional form. To establish the consistency of the functional trend estimator  $\hat{T}(s, t)$ , we introduce some concepts for functional time series. Let  $\mathcal{H}$  be a Hilbert space of square integrable functions defined on a compact interval  $D$ , with inner product  $\langle f, g \rangle = \int_D f(s)g(s) ds$ . Let  $\{X_n(s), s \in D\}$  be a sequence of random variables in  $\mathcal{H}$  with finite moments of order 2, that is, for each  $n$ ,  $\mathbb{E}(\|X_n\|_{\mathcal{H}}^2) < \infty$ , where  $\|\cdot\|_{\mathcal{H}}$  is the norm induced by the inner product in  $\mathcal{H}$ . Similarly to the univariate case, where the  $\alpha$ -mixing concept is required in the smoothing spline models with correlated random errors (Wang, 1998), one can assume short-range dependency in the functional time series  $\{X_n\}$ . We use the  $L^p - m$ -approximable concept (see the Supporting Information for more details of this concept).

Also, we will use the following assumptions that are concerned with the number of basis functions and the smoothing parameters.

**Assumption 1.**  $k_1 k_2 = o(N^r)$  for some  $r \in (0, 1)$  and  $\lim_{N \rightarrow \infty} k_1/k_2 = k_0$ , for some constant  $k_0$ .

**Assumption 2.**  $\lambda_1 k_1^4 = O(N)$  and  $\lambda_2 k_2^4 = O(N)$ .

**Assumption 3.** The knots for the spline bases  $\eta$  and  $\nu$  are equidistantly distributed.

**Proposition 1.** Let  $\{Y_n(s), s \in D\}$ , for  $n = 1, \dots, N$ , be the functional time series observed, and following model (1). Suppose that  $\{X_n\}$  is an  $L^4 - m$ -approximable sequence, and that the functional trend has a tensor product representation  $T(s, t) = \nu^\top(s) \Theta \eta(t)$  with fourth-order derivatives. Then, under Assumptions 1-3,

$$\mathbb{E} \left\{ \|\hat{T}(s, t) - T(s, t)\|_{L_2}^2 \right\} = o(1).$$

In a nonparametric regression estimation, the long-run covariance of the time series plays an important role when errors are correlated. The assumption of  $\{X_n\}$  being an

$L^4 - m$ -approximable sequence implies that the corresponding long-run covariance operator is convergent.

*Remark 1.* The weak dependence condition on  $\{X_n\}$  is across time,  $n$ . Thus, for each  $n$ ,  $\{X(s) := X_n(s), s \in D\}$  can be a nonstationary process. This is another advantage of FDA over the multivariate methods.

### 3.2 | Matrix representation

In practice, we do not observe continuous curves. Instead, each functional data  $Y_n(s)$  is observed on a grid of points  $\mathbf{s}_n = \{s_{n1}, \dots, s_{nm}\}$ . Without loss of generality, let us assume identical grids  $\mathbf{s}_n \equiv \mathbf{s} = \{s_1, \dots, s_m\}$  for  $n = 1, \dots, N$ . Let  $\mathbf{V} = \{\nu(\mathbf{s})\}^\top$  be the  $m \times k_1$ -matrix of the evaluation of  $k_1$  basis functions on  $m$  locations  $\mathbf{s}$ , let  $\mathbf{Z} = \{\eta(\mathbf{t})\}^\top$  be the  $N \times k_2$ -matrix of the evaluation of  $k_2$  basis functions on  $N$  times  $\mathbf{t} = \{1/N, 2/N, \dots, 1\}$ , and let  $\mathbf{Y} = \{Y_1(\mathbf{s}), Y_1(\mathbf{s}), \dots, Y_N(\mathbf{s})\}^\top$  be the  $m \times N$ -matrix of the observed functional time series, where each column represents observations of each continuous curve. Then, considering (2), model (1) can be written as

$$\mathbf{Y} = \mathbf{V} \Theta \mathbf{Z}^\top + \mathbf{X}, \quad (7)$$

where  $\mathbf{X}$  denotes the  $m \times N$ -matrix representing the evaluation of the functional time series  $X_n(s)$  at  $\mathbf{s}$ , for  $n = 1, \dots, N$ .

Thus, by using (6), the optimization problem (4) is equivalent to  $\arg \min_{\Theta} \|\mathbf{Y} - \mathbf{V} \Theta \mathbf{Z}^\top\|^2 + \{\text{vec}(\Theta)\}^\top \{\lambda_1 \mathbf{J}_\eta \otimes \mathcal{P}_1 + \lambda_2 \mathcal{P}_2 \otimes \mathbf{J}_\nu\} \text{vec}(\Theta)$ , where  $\|\cdot\|$  is the Frobenius norm, that is,  $\|\mathbf{E}\| = (\sum \sum |e_{ij}|^2)^{1/2}$  if  $\mathbf{E} = (e_{ij})$ . Thus, the solution  $\hat{\Theta}$  for  $\Theta$  satisfies,

$$\begin{aligned} & [(\mathbf{Z} \otimes \mathbf{V})^\top (\mathbf{Z} \otimes \mathbf{V}) + \lambda_1 \mathbf{J}_\eta \otimes \mathcal{P}_1 + \lambda_2 \mathcal{P}_2 \otimes \mathbf{J}_\nu] \text{vec}(\hat{\Theta}) \\ & = (\mathbf{Z} \otimes \mathbf{V})^\top \text{vec}(\mathbf{Y}). \end{aligned} \quad (8)$$

Then, given  $\lambda_1$  and  $\lambda_2$ , Equation (8) can be solved with the `smooth.bibasis` function in the `fda` R package.

**Proposition 2.** Assume that the functional time series is observed in a matrix form  $\{Y_n(s_i)\}$ ,  $n = 1, \dots, N$ ,  $i = 1, \dots, m$ , on a regular grid  $\mathbf{s} = \{s_1, \dots, s_m\}$ , and follows model (1). Suppose that  $\{X_n\}$  is an  $L^4 - m$ -approximable sequence, and that the functional trend has a tensor product representation  $T(s, t) = \nu^\top(s) \Theta \eta(t)$  with fourth-order derivatives. Then, under Assumptions 1-3, with the sample size  $Nm$ ,

$$\mathbb{E} \left\{ \|\hat{T}(s, t) - T(s, t)\|_{L_2}^2 \right\} = o(1),$$

where  $\hat{T}(s, t) = \nu^\top(s) \hat{\Theta} \eta(t)$ , and  $\hat{\Theta}$  is the solution of Equation (8).

*Remark 2.* If each curve of the functional time series is observed on an irregular or sparse grid, we can always write model (1) in a matrix form as in (7), with  $\mathbf{V}$  and  $\mathbf{Z}$  matrices evaluated on the corresponding grids.

### 3.3 | Smoothing parameters selection

When considering penalized regression splines, the number of basis functions  $k_1$  and  $k_2$  (or knots) do not have a significant influence on the resulting penalized fit (Ruppert, 2002). Usually, the number of basis functions grows with the sample size, but at a slower rate. Thus, the selection of  $\lambda_1$  and  $\lambda_2$  is more crucial, since these parameters control the flexibility of the tensor product. One of the advantages of tensor product surfaces is that all methods for curves are generalized easily. In particular, the methods to estimate the smoothing parameter can be extended to surfaces, such as cross-validation (CV), generalized cross-validation (GCV), or Akaike information criterion (AIC). In Wood (2006), the GCV method is used to estimate the smoothing parameters  $\lambda_1$  and  $\lambda_2$ . While these methods perform well for uncorrelated errors, they perform poorly with correlated errors, tending to underestimate (or overestimate) the smoothing parameters. In general, nonparametric estimators are sensitive to the presence of correlation in the errors, and several methods have been proposed. In Opsomer *et al.* (2001), one can find a general review of the literature in kernel regression, smoothing splines, and wavelet regression under correlated errors.

One possible solution to the correlated error problem is using a linear mixed effect model to represent the spline model. For instance, assume that the functional time series  $\{Y_n\}$  follows a Gaussian process. Thus,  $\text{vec}(\mathbf{Y})$  is a vector with Gaussian distribution, and  $\text{vec}(\Theta)$  can be estimated from the penalized log-likelihood function. Let  $\hat{\Theta}_{\text{ML}}$  be the estimator obtained from the penalized log-likelihood function. If the vector  $\text{vec}(\mathbf{X})$  in model (7) has each entry being an independent random variable, then  $\hat{\Theta}_{\text{ML}}$  satisfies Equation (8). Since the penalized tensor product in (7) can be considered as a linear mixed effect model, the estimator  $\hat{\Theta}_{\text{ML}}$  results in the posterior Bayes estimate (or best linear unbiased predictor). The latter has the advantage that the smoothing parameters  $\lambda_1$  and  $\lambda_2$  can be selected by using restricted maximum likelihood (REML). Moreover, in Krivobokova and Kauermann (2007) it is shown that the selection of the smoothing parameters based on REML is robust under correlation structures. Based on these observations, we propose using the REML to select  $\lambda_1$  and  $\lambda_2$  under the assumption of independent residuals and a Gaussian distribution, that is,  $\text{vec}(\mathbf{X}) \sim N(\mathbf{0}, \sigma_X^2 \mathbf{I}_{mN})$  with  $\mathbf{I}_{mN}$  as the identity matrix. Although the extension is straightforward for surfaces, it is

computationally expensive. Since the penalty (3) of  $T(s, t)$  is for each coordinate separately, by taking into account the average on the other coordinate, we propose using the REML on the marginal mean data instead of using the whole dataset. With our proposal, the computational time is drastically reduced without losing the accuracy of the estimator.

Specifically, we estimate  $\lambda_1$  by using REML with the empirical mean  $\frac{1}{N} \sum_{n=1}^N Y_n$  of the observed functional time series at  $\mathbf{s}$ . Similarly, we estimate  $\lambda_2$  by using the univariate time series  $\{\int Y_1(s)ds, \dots, \int Y_N(s)ds\}$ . To gain an intuition about this estimation, see the Supporting Information. Once we have estimated  $\lambda_1$  and  $\lambda_2$  we solve (8) to obtain  $\hat{\Theta}$ .

*Remark 3.* The estimated smoothing parameter  $\hat{\lambda}_1$  controls the shape of the mean curve of the functional time series  $\{Y_n\}$ . On the other hand, the mean curve represents the common shape of the functional data over time. Thus,  $\hat{T}(s, t)$  is expected to represent the shape in the  $s$  coordinate. The estimated smoothing parameter  $\hat{\lambda}_2$  represents the shape of the trend of the average data in each period  $n = 1, \dots, N$ . That is,  $\hat{\lambda}_2$  controls the common trend of the functional time series, and so  $\hat{T}(s, t)$  represents the shape of the functional trend.

The methodology proposed here to estimate the functional trend in functional time series is easy to implement and computationally efficient. To obtain the estimators  $\hat{\lambda}_1$  and  $\hat{\lambda}_2$ , we can use the `gam` function in the `mgcv` package. Given  $\hat{\lambda}_1$  and  $\hat{\lambda}_2$ , we can obtain  $\hat{\Theta}$  using the `smooth.bibasis` function in the `fda` package. An R code example of this implementation is included in the Supporting Information.

## 4 | NUMERICAL PROPERTIES

### 4.1 | Preliminaries

We investigate the performance of our proposed method under different scenarios. We use the `gam` function combined with the `smooth.bibasis` function in the `mgcv` and `fda` (Ramsay *et al.*, 2018) packages, respectively. To the best of our knowledge, there is no paper addressing functional trend estimation when functional data are observed over time.

In the literature of nonparametric models, we can find methods related to the estimation of  $T(s, t)$ . However, these methods assume that the residuals  $\{X_n(s_i)\}$  are independent (or uncorrelated) for all  $n = 1, \dots, N$  and  $i = 1, \dots, m$ . Notice that our method does not require independence or stationarity of  $\{X_n(s_i)\}_{i=1}^m$ .

We denote our method as  $T_{\text{TPS}}$ , where TPS stand for tensor product surface. We compare it with the following estimators:

- 1) Finite element method:  $\hat{T}_{\text{FEM}}(s, t) := \sum_{j=1}^k \hat{a}_j \psi_j(s, t)$ , where  $\psi_j$  is a quadratic basis function associated with each node defined at points where data are observed, that is,  $(s_i, t_n)$ . The coefficients  $\hat{a}_j$  are obtained using a penalized least square method. Details of this method can be found in Azzimonti *et al.* (2015). For this method, we use the *fdaPDE* package (Lila *et al.*, 2019) to obtain  $\hat{T}_{\text{FEM}}(s, t)$ .
- 2) Kernel method:  $\hat{T}_{\text{Ker}}(s, t) := \frac{\sum_{n=1}^N Y_n(s) K\left\{\frac{(s,t)-(s,n/N)}{h}\right\}}{\sum_{n=1}^N K\left\{\frac{(s,t)-(s,n/N)}{h}\right\}}$ , where  $K(s, t) = \frac{1}{2\pi} \exp\left\{-\frac{s^2+t^2}{2}\right\}$ . The bandwidth  $h$  is selected via CV.
- 3) Linear trend:  $\hat{T}_{\text{Lin}}(s, t) := \hat{\mu}(s) + t\hat{f}(s)$ , where  $\hat{\mu}(s) = \bar{Y}_n(s) - \hat{f}(s) \frac{N+1}{2}$ , and  $\hat{f}(s) = \frac{1}{s_N} \sum_{n=1}^N (n - \frac{N+1}{2}) Y_n(s)$  with  $s_N = \sum_{n=1}^N (n - \frac{N+1}{2})^2$ .
- 4) Naive method: For each  $s_i \in \{s_1, \dots, s_m\}$ ,  $\hat{T}_{\text{Naiv}}(s_i, t) := \sum_{j=1}^{k_2} \hat{\theta}_j(s_i) \eta_j(t)$ , where  $(\eta_1, \dots, \eta_{k_2})$  is a B-spline basis function. The coefficients  $\hat{\theta}_j$  are obtained via a penalized least squares method, and using the *fda* package (Ramsay *et al.*, 2018). The penalty term is selected via GCV.
- 5) Sandwich smoother:  $\hat{T}_{\text{Sand}}(s, t) := \sum_{j=1}^{k_1} \sum_{i=1}^{k_2} \hat{a}_{j,i} \nu_j(s) \eta_i(t)$ . This estimator has the same form as (2), but the smoothing method is different (see Xiao *et al.*, 2013). The sandwich smoother is implemented in the *fbps* command in the *refund* package (Goldsmith *et al.*, 2018), and the corresponding smoothing parameters are selected via GCV.
- 6) Thin plate regression splines:  $\hat{T}_{\text{ThinP}}(s, t) := \sum_{j=1}^{k_1} \sum_{i=1}^{k_2} \hat{b}_{j,i} \nu_j(s) \eta_i(t)$ . This estimator is a tensor product as well with a thin plate energy penalty (see Wood, 2003). This method is implemented in the *gam* command in the *mgcv* package. The corresponding smoothing parameter is selected via GCV.

The estimator  $\hat{T}_{\text{FEM}}(s, t)$  is commonly used in cases where the domain is complex. In our case, the domain is a simple rectangle.  $\hat{T}_{\text{Lin}}(s, t)$  is a basic parametric linear trend model, and we use it as a baseline to measure the accuracy of our method in the simplest case (linear).  $\hat{T}_{\text{Naiv}}(s, t)$  is the most commonly used estimator on functional magnetic resonance imaging (fMRI) data to detrend time series at each voxel separately, see, for example, Tanabe *et al.* (2002) and Ombao *et al.* (2017).  $\hat{T}_{\text{Sand}}(s, t)$  and  $\hat{T}_{\text{ThinP}}(s, t)$  are smooth tensor product surfaces proposed in nonparametric regression.

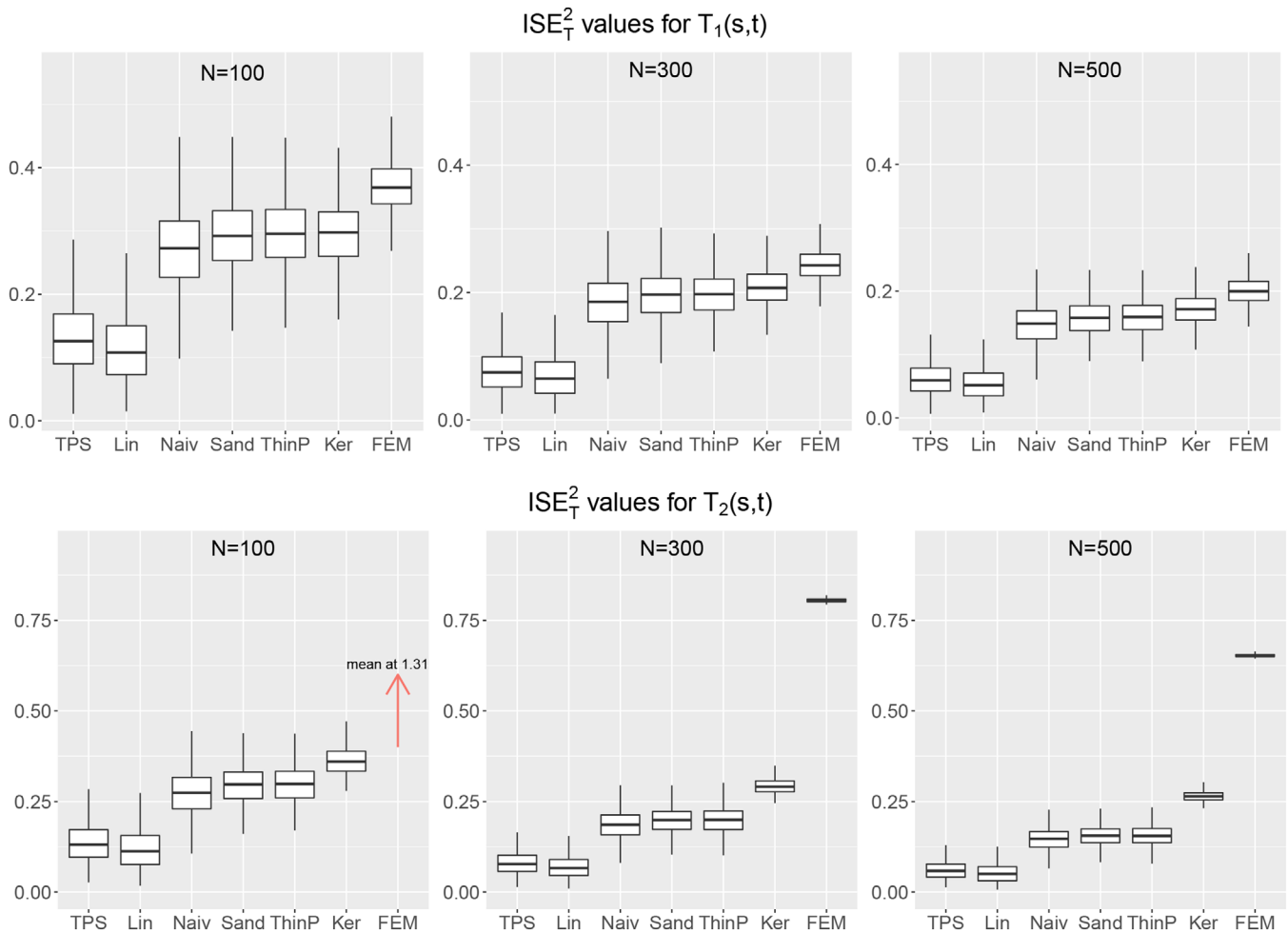
## 4.2 | Simulation setting

We simulate  $\{Y_n(s); s \in [0, 1], n = 1, \dots, N\}$  from model (1) with six different functional trends,  $T(s, t)$ , defined as follows: (a)  $T_1(s, t) = 2s + 30t$ , (b)  $T_2(s, t) = 25t \sin(2\pi s)$ , (c)  $T_3(s, t) = 20t^2 - 5t + 5$ , (d)  $T_4(s, t) = 2(0.5s + 4t)^2$ , (e)  $T_5(s, t) = 28 \sin(2\pi t + s)$ , and (f)  $T_6(s, t) = \frac{4.5}{\pi\sigma_s\sigma_t} \exp\left\{\frac{-(s-0.2)^2}{\sigma_s^2} - \frac{(t-0.3)^2}{\sigma_t^2}\right\} + \frac{2.7}{\pi\sigma_s\sigma_t} \exp\left\{\frac{-(s-0.7)^2}{\sigma_s^2} - \frac{(t-0.8)^2}{\sigma_t^2}\right\}$ , with  $\sigma_s = 0.3$  and  $\sigma_t = 0.4$ . The function  $T_6(s, t)$  was used in Wood (2003) and in Xiao *et al.* (2013) to study the performance of  $\hat{T}_{\text{ThinP}}(s, t)$  and  $\hat{T}_{\text{Sand}}(s, t)$ , respectively. The resulting surfaces for each of these models can be visualized in the Supporting Information.

The stationary functional time series component,  $\{X_n\}$ , is simulated from the functional autoregressive model of order one (FAR(1)), defined as  $X_n(s) = C_1 \int_{[0,1]} \beta(u, s) X_{n-1}(u) du + W_n(s)$ , with kernel  $\beta(u, v) = \exp\{-(u^2 + v^2)/2\}$ , and functional white noise  $\{W_n\}$  as independent Brownian motion defined in  $[0, 1]$ , where the scalar  $C_1$  is such that the norm of the corresponding coefficient operator is 0.7, that is,  $\{\int_{[0,1]} \int_{[0,1]} \beta^2(u, v) dudv\}^{1/2} = 0.7$ . We consider different sample sizes  $N = 100, 300$ , and 500. For each  $n = 1, \dots, N$ , we simulate  $Y_n(s)$  on an equispaced 50-point grid on  $[0, 1]$ . Each simulation set is replicated 1000 times.

For each simulation, we compute the functional trend. For our method  $T_{\text{TPS}}$  and for methods 4), 5), and 6), we fix  $k_1 = 10$  and  $k_2 = 15$  in all cases. To compare the performance of our estimator  $\hat{T}_{\text{TPS}}$  with the competitors, we consider two different criteria.

First, we evaluate the accuracy of the estimation of the functional trend component, computing the corresponding integrated squared error ( $\text{ISE}_T$ ) defined as  $\text{ISE}_T^2 = \int_{[0,1]} \int_{[0,1]} \{T(s, t) - \hat{T}(s, t)\}^2 ds dt$ . Second, we evaluate the accuracy of the estimation of the kernel  $\beta(u, v)$  after removing the estimated functional trend. To do this, we estimate the kernel  $\beta$  from the residual functional time series  $\{\tilde{X}_n(s)\} = \{Y_n(s) - \hat{T}(s, n/N)\}$ . We denote this estimator by  $\hat{\beta}_Y$ . Since our goal is not to have the best estimator of the kernel  $\beta$ , we assume that  $\hat{\beta}_X$  is the truth, where  $\hat{\beta}_X$  is the estimator obtained from the original simulated functional time series  $\{X_n\}$ . Thus, we compare the estimator  $\hat{\beta}_Y$  with the estimator  $\hat{\beta}_X$  by computing the corresponding integrated squared error ( $\text{ISE}_\beta$ ) defined as  $\text{ISE}_\beta^2 = \int_D \int_D \{\hat{\beta}_X(s, t) - \hat{\beta}_Y(s, t)\}^2 ds dt$ . The kernel estimators  $\hat{\beta}_Y$  and  $\hat{\beta}_X$  are obtained by using the *linmod* function with 15 B-spline basis functions for each coordinate  $u$  and  $v$ . Other parameters required in the *linmod* function are set to be equal in both cases,  $\hat{\beta}_Y$  and  $\hat{\beta}_X$ , to make them comparable.



**FIGURE 2** Boxplots of the  $ISE_T^2$  values for each simulation  $\{Y_n, n = 1, \dots, N\}$  with functional trends  $T_1$  and  $T_2$ , and different sample sizes  $N = 100, 300$ , and  $500$ . A red arrow indicates that the  $ISE_T^2$  values are out of visual range, and its mean is reported. Our proposed estimator  $\hat{T}_{TPS}$  and  $\hat{T}_{Lin}$  outperform the others

The value  $ISE_T$  represents the error approximation of the functional trend, while  $ISE_\beta$  indicates the difference between  $\{Y_n(s) - \hat{T}(s, n/N)\}$  and  $\{X_n\}$  in terms of dependency structure over time. Thus,  $ISE_\beta$  can be interpreted as the error dependency structure between  $\{Y_n(s) - \hat{T}(s, n/N)\}$  and  $\{X_n\}$  that is caused by the error approximation  $T(s, t) - \hat{T}(s, t)$  of the functional trend.

### 4.3 | Simulation results

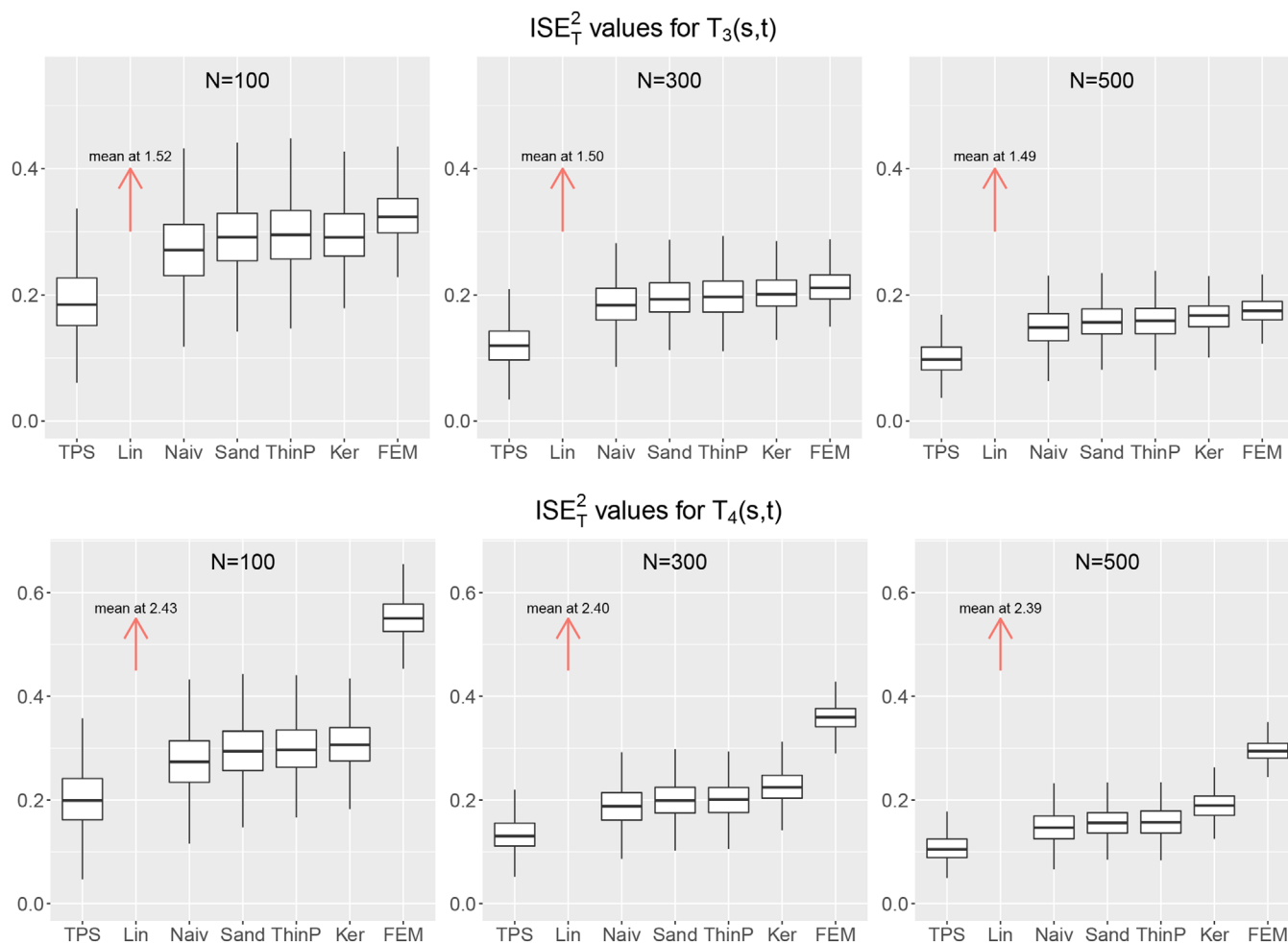
We present the results according to the shape of the functional trends over time: linear (Figure 2), quadratic (Figure 3), and complex (Figure 4).

Figure 2 shows that our estimator  $\hat{T}_{TPS}$  and  $\hat{T}_{Lin}$  are highly accurate for the linear functional trends  $T_1$  and  $T_2$ . Both estimators have the lowest error values, and they decrease when the sample size increases. Thus, in these cases, our proposed estimator performs as well as the

parametric estimator  $\hat{T}_{Lin}$ , with the advantage that our estimator does not require the specification of the functional trend shape. The results are similar for the functional trends  $T_3$  and  $T_4$  (Figure 3). The  $ISE_T$  values for  $\hat{T}_{TPS}$  remain as accurate as in the linear trends, except that the  $ISE_T$  values for  $\hat{T}_{Lin}$  become significantly larger, which is expected since the functional trends are not linear anymore. Therefore, our proposed estimator outperforms the rest of the estimators on the quadratic functional trend. The latter conclusion extends to the  $T_5$  and  $T_6$  functional trends. Also, we observe that, in the case of nonlinear trends, the  $\hat{T}_{Naiv}$  estimator is the second best estimator after our method.

Next, we analyze the  $ISE_\beta$  values that represent the errors of the dependency structure caused by the error approximation of the functional trend estimator. We only present results corresponding to sample size  $N = 300$ . The results from  $N = 100$  and  $N = 500$  are similar. Boxplots of all cases can be found in the Supporting Information.





**FIGURE 3** Boxplots of the  $ISE_T^2$  values for each simulation  $\{Y_n, n = 1, \dots, N\}$  with functional trends  $T_3$  and  $T_4$  and different sample sizes  $N = 100, 300$ , and  $500$ . A red arrow indicates that the  $ISE_T^2$  values are out of visual range, and its mean is reported. Our proposed estimator  $\hat{T}_{TPS}$  outperforms the others

Table 1 shows the corresponding mean values and the standard deviations in parenthesis. We observe that the  $ISE_\beta$  values behave similarly to the  $ISE_T$  values in almost all cases of different functional trends, except for the trend  $T_6$ . The  $ISE_\beta$  values are similar for  $\hat{T}_{TPS}$  and  $\hat{T}_{Lin}$  when considering functional trends  $T_1$  and  $T_2$ . For  $T_3$  and  $T_4$ , the  $ISE_\beta$  values are significantly larger with the competitor estimators, whereas, for the  $\hat{T}_{TPS}$  estimator, the  $ISE_\beta$  values remain small. The conclusion is the same for the functional trend  $T_5$ . For  $T_6$ , surprisingly, the estimator  $\hat{T}_{FEM}$  presents the lowest mean value of  $ISE_\beta$ . However,  $\hat{T}_{FEM}$  performs poorly in all cases when approximating the functional trend, that is,  $\hat{T}_{FEM}$  presents the largest  $ISE_T$  values.

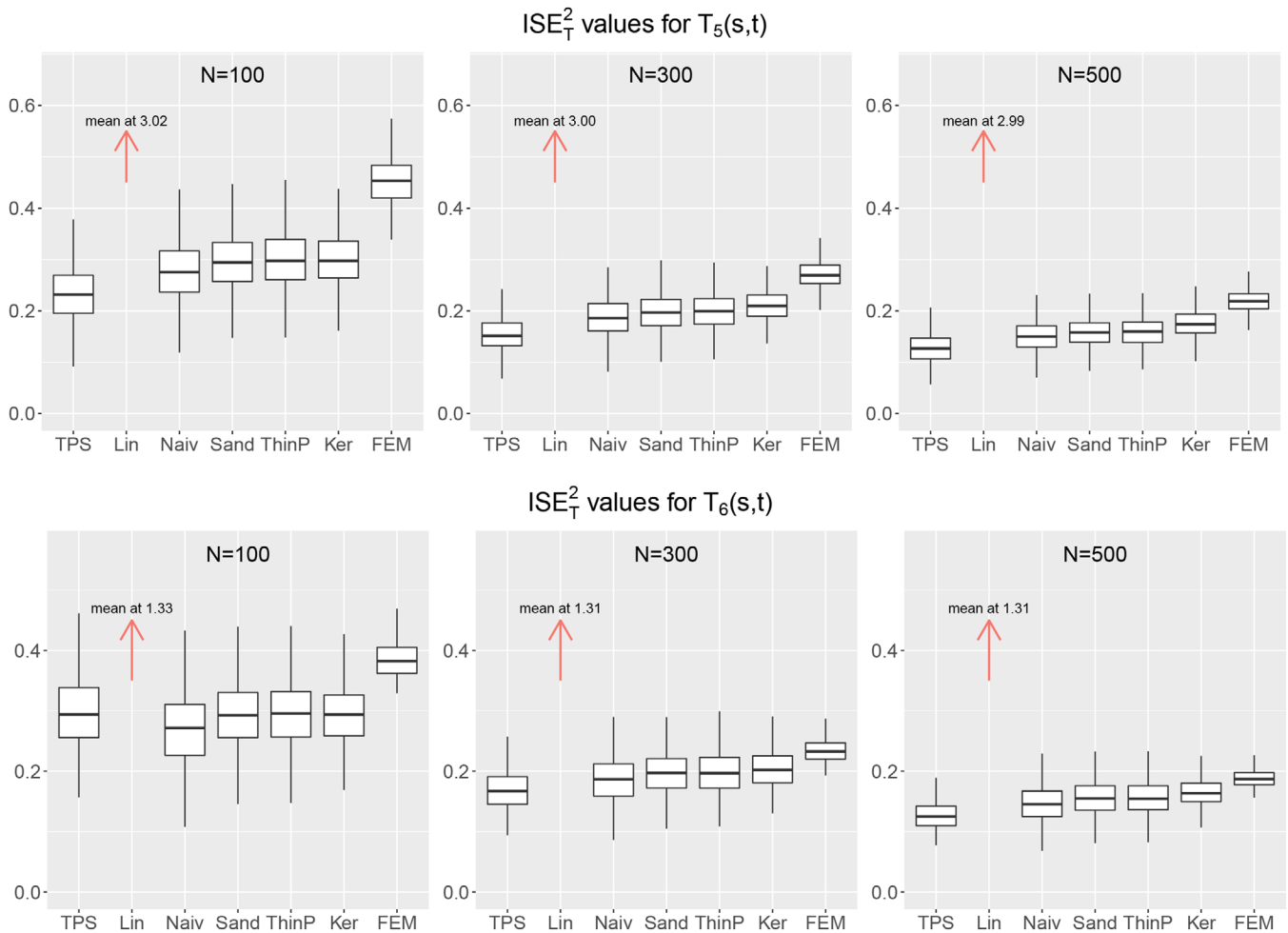
In general, we conclude that our proposed estimator performs well in all cases, even with simple models such as models  $T_1$  and  $T_2$  of the functional trend. It has the advantage of being applicable to a general class of functional trends with complex structures, and accurately describes the functional trends.

## 5 | DATA ANALYSIS

### 5.1 | Objectives

In this section, we apply our methodology on annual mortality rates in France. Our goal is to show that the consideration of a functional trend from a functional point of view improves data analysis, in particular data forecasting. We model the dataset considering the functional trend described in Section 2.2. Then, we compare the forecasts with the model without considering the functional trend.

To forecast functional time series, we adopt one of the most feasible and commonly used procedures. Let  $\{Z_n(s), n = 1, \dots, N\}$  be a functional time series with sample size  $N$ . For each  $n$ ,  $Z_n$  is transformed into a vector time series of dimension  $r$ ,  $\mathbf{Z}_n = (z_{n,1}, \dots, z_{n,r})^\top$ , by projecting  $Z_n$  into  $r$  functional principal components. Then, the multivariate time series  $\{\mathbf{Z}_n, n = 1, \dots, N\}$  is modeled



**FIGURE 4** Boxplots of the  $ISE_T^2$  values for each simulation  $\{Y_n, n = 1, \dots, N\}$  with functional trends  $T_5$  and  $T_6$ , and different sample sizes  $N = 100, 300$ , and  $500$ . A red arrow indicates that the  $ISE_T^2$  values are out of visual range, and its mean is reported. Our estimator  $\hat{T}_{TPS}$  has a good performance in all cases

**TABLE 1** Mean of the  $ISE_\beta^2$  values for each simulation  $\{Y_n, n = 1, \dots, N\}$  with different functional trends,  $T_i(s, t)$ , and sample size  $N = 300$

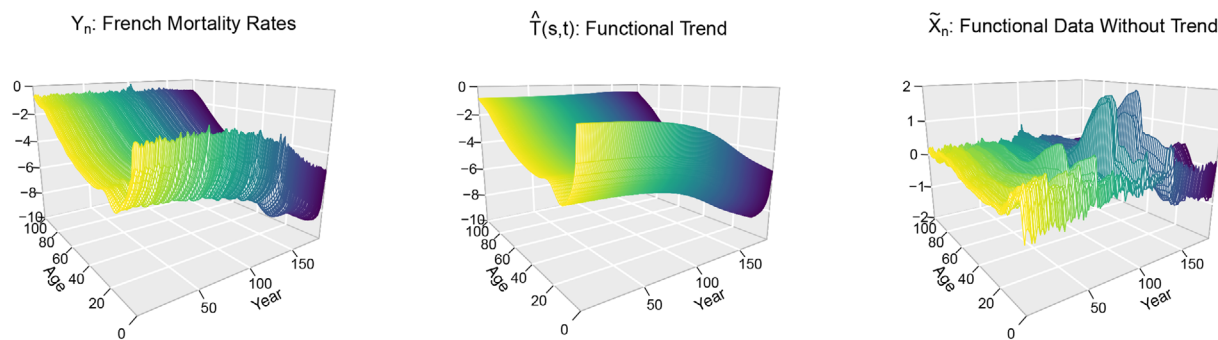
	$T_1(s, t)$	$T_2(s, t)$	$T_3(s, t)$	$T_4(s, t)$	$T_5(s, t)$	$T_6(s, t)$
TPS	0.028 (0.02)	0.030 (0.02)	<b>0.057</b> (0.03)	<b>0.076</b> (0.04)	<b>0.114</b> (0.05)	0.371 (0.09)
Lin	<b>0.014</b> (0.01)	<b>0.015</b> (0.02)	0.452 (0.07)	0.464 (0.07)	1.071 (0.07)	0.955 (0.05)
Naiv	0.151 (0.06)	0.154 (0.06)	0.154 (0.06)	0.158 (0.06)	0.153 (0.06)	0.153 (0.06)
Sand	0.157 (0.06)	0.159 (0.06)	0.159 (0.06)	0.162 (0.06)	0.157 (0.06)	0.158 (0.06)
ThinP	0.165 (0.06)	0.169 (0.06)	0.168 (0.06)	0.173 (0.06)	0.166 (0.06)	0.168 (0.06)
Ker	0.263 (0.06)	0.320 (0.07)	0.276 (0.06)	0.209 (0.06)	0.251 (0.06)	0.297 (0.07)
FEM	0.080 (0.04)	0.432 (0.04)	0.068 (0.03)	0.221 (0.06)	0.160 (0.04)	<b>0.145</b> (0.03)

Bold font is used to highlight the best performance. The corresponding standard deviations are indicated in parenthesis.

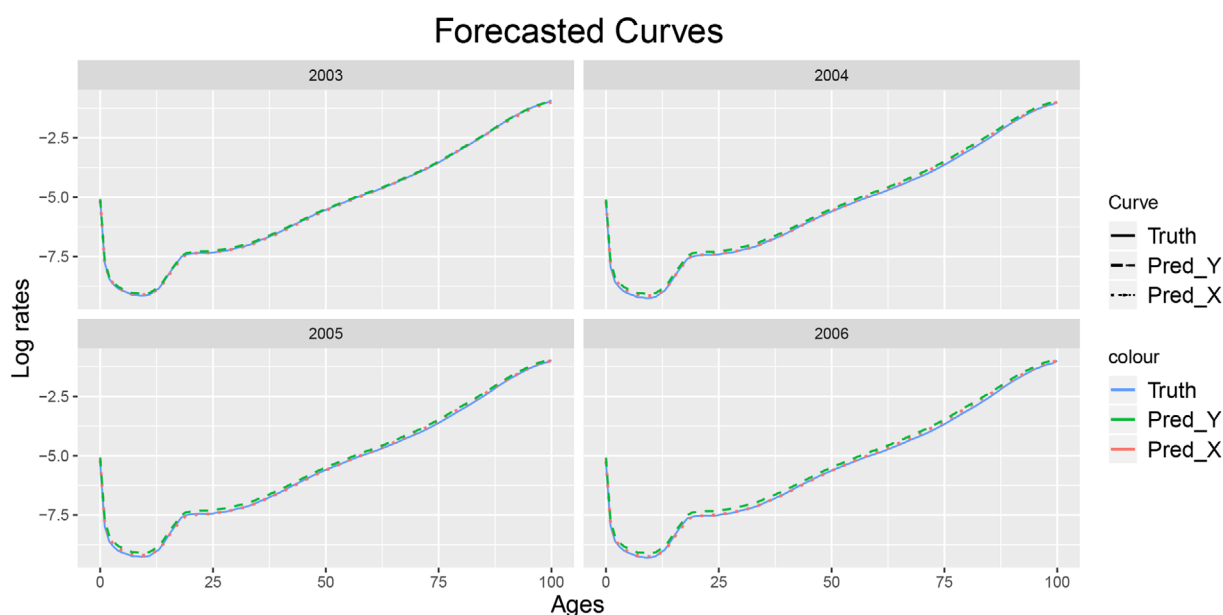
by using VAR( $p$ ) or autoregressive integrated moving average (ARIMA) models. Using the fitted time series model, and for  $h$  fixed, we obtain the  $h$ -step ahead forecast  $\hat{\mathbf{Z}}_{N+h} = (\hat{z}_{N+h,1}, \dots, \hat{z}_{N+h,r})^\top$ . Finally, we multiply the predicted vector  $\hat{\mathbf{Z}}_{N+h}$  by the  $r$  estimated principal components to obtain the  $h$ -step ahead forecast of functional time series

$\hat{Z}_{N+h}(s)$  (see Hyndman and Ullah, 2007; Aue *et al.*, 2015, for more details). Here, we model each component of  $\{\mathbf{Z}_n\}$  separately, similarly as in Hyndman and Ullah (2007).

Thus, to see the differences between considering and not considering the functional trend  $T(s, t)$ , we apply the latter methodology described in the functional time



(a) Functional data  $\{Y_n\}$  observed (left). Estimated functional trend (center), and functional data after removing the estimated functional trend (right).



(b) Forecasting when the functional trend is considered, and when the functional trend is not considered.

**FIGURE 5** Results of data analysis. (a) Estimated components of the model (1). (b) Four consecutive curves of log mortality rates with their corresponding forecasted curves. The solid curves (blue) represent the true curves  $Y_{2003}(s), \dots, Y_{2006}(s)$ ; the dotted curves (red) represent the forecasted curves considering the functional trend, using the time series  $\{\tilde{x}_{n,r}\}$ ; the dashed curves (green) represent the predicted curves without considering the functional trend, using the time series  $\{y_{n,r}\}$

series  $\{Y_n, n = 1, \dots, N\}$ , and in the functional time series  $\{\tilde{X}_n, n = 1, \dots, N\}$ , where  $\tilde{X}_n(s) := Y_n(s) - \hat{T}(s, n/N)$  and  $\hat{T}(s, n/N)$  is obtained as described in Section 2. The corresponding models for the univariate time series are selected with AIC.

## 5.2 | Mortality rates in France

This dataset consists of 191 curves of annual mortality rates in France, from 1816 to 2006, for individuals from zero to

100 years old. Each point of the curve  $Y_n(s)$  represents the log of the mortality rate, in year  $n$ , at age  $s$ . At first glance from Figure 5a (left), we can say that the functional time series  $\{Y_n\}$  is nonstationary, and also we can observe a decreasing trend over the years. After applying the stationarity test proposed by Horváth *et al.* (2014), we obtain a  $p$ -value equal to 0.003, and the smaller the  $p$ -value, the more evidence against the stationarity. Thus, we consider model (1).

To evaluate the performance of the forecast, we remove the last four curves of  $\{Y_n\}$ , that is, we only consider

curves from 1816 to 2002, with  $N = 187$ . Figure 5 shows the resulting functional time series  $Y_n$ , the estimated functional trend  $\hat{T}(s, t)$ , and the functional time series  $\{\tilde{X}_n\}$  after removing the trend (left to right). We fit autoregressive moving average (ARMA) models for the coefficients  $\{\tilde{x}_{n,r}, n = 1816, \dots, 2002\}$ ,  $r = 1, 2, 3, 4$ . Then, we forecast the four curves  $\hat{X}_{2003}, \hat{X}_{2004}, \hat{X}_{2005}$ , and  $\hat{X}_{2006}$ . The models fitted for  $\{\tilde{x}_{n,r}\}$  are: ARMA(1,0) with zero mean and coefficient 0.7506, ARMA(1,0) with zero mean and coefficient 0.9825, ARMA(1,1) with zero mean and coefficients (ar = 0.9212, ma = -0.5593), and ARMA(2,0) with zero mean and coefficient (ar1 = 0.4492, ar2 = 0.3601), for  $r = 1, 2, 3$ , and 4, respectively. Also, we forecast the four functional trends  $\hat{T}_{2003}, \hat{T}_{2004}, \hat{T}_{2005}$ , and  $\hat{T}_{2006}$  as described in (5). Finally, we obtain the forecast of the log mortality rate  $\hat{Y}_{2002+h}(s) = \hat{T}_{2002+h}(s) + \hat{X}_{2002+h}(s)$  for  $h = 1, 2, 3, 4$ .

For the case in which the functional trend is not considered, we fit ARIMA models for the coefficients of the projected functional time series,  $\{y_{n,r}\}$ . In this case, the models fitted are ARIMA(1,1,1) with coefficients (ar = 0.6562, ma = -0.8259, drift = -0.1213), ARIMA(1,1,1) with coefficients (ar = 0.7606, ma = -0.9668), ARIMA(1,0,1) with coefficients (ar = 0.8853, ma = -0.5156), ARIMA(3,1,1) with coefficients (ar1 = 0.2569, ar2 = 0.2362, ar3 = -0.1590, ma1 = -0.6719), for  $r = 1, 2, 3$ , and 4, respectively. We observe that, when the functional trend is not removed, the time series corresponding to the first principal component  $\{y_{n,1}\}$  seems to absorb the trend component. The corresponding time series plots can be found in the Supporting Information (Figure 6).

Figure 5b shows the four forecasted curves. We use different line types and colors to indicate the true curves and forecasted curves (these figures appear in color in the electronic version of this article, and any mention of color refers to that version). The solid curves (blue) represent the true curves  $Y_{2002+h}(s)$ , the dotted curves (red) represent the forecasted curves considering the functional trend, that is, using the time series  $\{\tilde{x}_{n,r}\}$  and forecasting the functional trend, and the dashed curves (green) represent the forecasted curves without considering the functional trend, that is, using the time series  $\{y_{n,r}\}$ . Although both methods seem to perform well, the forecasted curves obtained when considering functional trend are more accurate. Namely, the sum of the  $L_1$  distance between the truth curves and the predicted curves for each method is 0.449 and 0.164, without/with considering functional trend, respectively.

We observe that the forecasted curves obtained when considering a functional trend are more accurate, that is, they are closer to the true curves, whereas the forecasted curves obtained when a functional trend is not taken into

account are farther away from the true curves. Thus, the consideration of estimating the functional trend improves data analysis. Based on this, we conclude that the statistical analysis is more accurate when the functional trend is taken into account from the functional point of view. We recommend estimating such a functional trend before modeling the stochastic component  $\{X_n\}$  in model (1), either using dimension reduction techniques such as functional principal component, or using a functional time series model such as the functional autoregressive models, FAR( $p$ ).

## 6 | DISCUSSION

In our study, we assumed a functional time series with a trend component (functional trend). We proposed estimating the functional trend by using a tensor product surface, and taking into account the dependency of the data. To obtain smoothness properties of the estimator, we used marginal penalties. The smoothing parameters were selected based on restricted maximum likelihood, which is robust under correlation structures. We showed that the proposed estimator of the functional trend is consistent when the sample sizes go to infinity. One of the advantages of our proposal is that it is easy to implement by using existing R packages, and it can handle large data. In the Monte Carlo simulation, we showed that our functional trend estimator performs well for simple and complex structures of the functional trend. With the annual mortality rates data, we showed that when the functional trend is estimated, it improves the inference and the forecasting.

With this work, we want to encourage taking into account the deterministic component and estimate it from a functional point of view for a functional time series. So, we believe this work will be of interest for data applications. Also, this work leads to a future project that is the extension to functional time series with domain in  $\mathbb{R}^2$ , called surface time series (Martínez-Hernández and Genton, 2020). Such an extension could benefit, for example, fMRI data and spatio-temporal data in general.

### DATA AVAILABILITY STATEMENT

The data that support the findings in this paper are available from the *demography* R package [https://cran.r-project.org/src/contrib/demography\\_1.22.tar.gz](https://cran.r-project.org/src/contrib/demography_1.22.tar.gz) (Hyndman *et al.*, 2019).

### ORCID

Israel Martínez-Hernández  <https://orcid.org/0000-0002-4122-2529>

Marc G. Genton  <https://orcid.org/0000-0001-6467-2998>

## REFERENCES

- Aue, A., Norinho, D.D. and Hörmann, S. (2015) On the prediction of stationary functional time series. *Journal of the American Statistical Association*, 110, 378–392.
- Azzimonti, L., Sangalli, L.M., Secchi, P., Domanin, M. and Nobile, F. (2015) Blood flow velocity field estimation via spatial regression with PDE penalization. *Journal of the American Statistical Association*, 110, 1057–1071.
- Beare, B.K., Seo, J. and Seo, W.-K. (2017) Cointegrated linear processes in Hilbert space. *Journal of Time Series Analysis*, 38, 1010–1027.
- Chen, L. and Wu, W.B. (2019) Testing for trends in high-dimensional time series. *Journal of the American Statistical Association*, 114, 869–881.
- Claeskens, G., Krivobokova, T. and Opsomer, J.D. (2009) Asymptotic properties of penalized spline estimators. *Biometrika*, 96, 529–544.
- Eilers, P.H.C. and Marx, B.D. (1996) Flexible smoothing with B-splines and penalties. *Statistical Science*, 11, 89–121. With comments and a rejoinder by the authors.
- Fraiman, R., Justel, A., Liu, R. and Llop, P. (2014) Detecting trends in time series of functional data: a study of Antarctic climate change. *The Canadian Journal of Statistics*, 42, 597–609.
- Goldsmith, J., Scheipl, F., Huang, L., Wrobel, J., Gellar, J., Harezlak, J., McLean, M.W., Swihart, B., Xiao, L., Crainiceanu, C. and Reiss, P.T. (2018). *refund: Regression with Functional Data*. R package version 0.1-17.
- Hasenstab, K., Scheffler, A., Telesca, D., Sugar, C.A., Jeste, S., DiStefano, C. and Şentürk, D. (2017) A multi-dimensional functional principal components analysis of EEG data. *Biometrics*, 73, 999–1009.
- Horváth, L., Kokoszka, P. and Rice, G. (2014) Testing stationarity of functional time series. *Journal of Econometrics*, 179, 66–82.
- Horváth, L. and Rice, G. (2015) Testing equality of means when the observations are from functional time series. *Journal of Time Series Analysis*, 36, 84–108.
- Hyndman, R.J., Booth, H., Tickle, L. and Maindonald, J. (2019). *demography: Forecasting Mortality, Fertility, Migration and Population Data*. R package version 1.22.
- Hyndman, R.J. and Ullah, M.S. (2007) Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics & Data Analysis*, 51, 4942–4956.
- Kokoszka, P. (2012) Dependent functional data. *ISRN Probability and Statistics*, 2012, 1–30. Article ID 958254.
- Kokoszka, P. and Young, G. (2017) Testing trend stationarity of functional time series with application to yield and daily price curves. *Statistics and Its Interface*, 10, 81–92.
- Krivobokova, T. and Kauermann, G. (2007) A note on penalized spline smoothing with correlated errors. *Journal of the American Statistical Association*, 102, 1328–1337.
- Lai, M.-J. and Wang, L. (2013) Bivariate penalized splines for regression. *Statistica Sinica*, 23, 1399–1417.
- Li, Y. and Ruppert, D. (2008) On the asymptotics of penalized splines. *Biometrika*, 95, 415–436.
- Lila, E., Sangalli, L.M., Ramsay, J. and Formaggia, L. (2019). *fdaPDE: Functional Data Analysis and Partial Differential Equations; Statistical Analysis of Functional and Spatial Data, Based on Regression with Partial Differential Regularizations*. R package version 0.1-6.
- Martínez-Hernández, I. and Genton, M.G. (2020) Recent developments in complex and spatially correlated functional data. *Brazilian Journal of Probability and Statistics*, 34, 204–229.
- Ombao, H., Lindquist, M., Thompson, W. and Aston, J. (Eds.). (2017) *Handbook of Neuroimaging Data Analysis*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. Boca Raton, FL: CRC Press.
- Opsomer, J., Wang, Y. and Yang, Y. (2001) Nonparametric regression with correlated errors. *Statistical Science*, 16, 134–153.
- Ramsay, J.O. and Silverman, B.W. (2005). *Functional Data Analysis*, 2nd edition. Springer Series in Statistics. New York: Springer.
- Ramsay, J.O., Wickham, H., Graves, S. and Hooker, G. (2018). *fda: Functional Data Analysis*. R package version 2.4.8.
- Ruppert, D. (2002) Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics*, 11, 735–757.
- Tanabe, J., Miller, D., Tregellas, J., Freedman, R. and Meyer, F.G. (2002) Comparison of detrending methods for optimal fmri preprocessing. *NeuroImage*, 15, 902–907.
- Telesca, D., Inoue, L.Y., Neira, M., Etzioni, R., Gleave, M. and Nelson, C. (2009) Differential expression and network inferences through functional data modeling. *Biometrics*, 65, 793–804.
- Ullah, S. and Finch, C.F. (2013) Applications of functional data analysis: A systematic review. *BMC Medical Research Methodology*, 13, 1–12.
- Wang, Y. (1998) Smoothing spline models with correlated random errors. *Journal of the American Statistical Association*, 93, 341–348.
- Wood, S.N. (2003) Thin plate regression splines. *Journal of the Royal Statistical Society. Series B*, 65, 95–114.
- Wood, S.N. (2006) Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics*, 62, 1025–1036.
- Wu, W.B. and Zhao, Z. (2007) Inference of trends in time series. *Journal of the Royal Statistical Society: Series B Statistical Methodology*, 69, 391–410.
- Xiao, L. (2019) Asymptotics of bivariate penalised splines. *Journal of Nonparametric Statistics*, 31, 289–314.
- Xiao, L., Li, Y. and Ruppert, D. (2013) Fast bivariate P-splines: the sandwich smoother. *Journal of the Royal Statistical Society. Series B*, 75, 577–599.

## SUPPORTING INFORMATION

Web Appendix referenced in Section 3, and Figures referenced in Sections 4, and 5 are available with this paper at the Biometrics website on Wiley Online Library. The R code referenced in Section 3 is also available with this paper at the Biometrics website on Wiley Online Library or in GitHub repositories <https://github.com/Martinez-Hernandez/Nonparametric-Trend-Estimation-in-Functional-TS>.

**How to cite this article:** Martínez-Hernández I, Genton MG. Nonparametric trend estimation in functional time series with application to annual mortality rates. *Biometrics*. 2021;77:866–878. <https://doi.org/10.1111/biom.13353>