Wavelet estimation of the dimensionality of curve time series

Rodney V. Fonseca[1] · Aluísio Pinheiro[1]

Abstract

Functional data analysis is ubiquitous in most areas of sciences and engineering. Several paradigms are proposed to deal with the dimensionality problem which is inherent to this type of data. Sparseness, penalization, thresholding, among other principles, have been used to tackle this issue. We discuss here a solution based on a finite-dimensional functional subspace. We employ wavelet representation of random functions to estimate this finite dimension and successfully model a time series of curves. The proposed method is shown to have nice asymptotic properties. Moreover, the wavelet representation permits the use of several bootstrap procedures, and it results in faster computing algorithms. Besides the theoretical and computational properties, some simulation studies and an application to real data are provided.

Keywords Aggregate data · Bootstrap testing · Finite dimension · Functional data analysis

1 Introduction

Many phenomena, natural or anthropogenic, can be appropriately modeled by a function on a suitable domain. The underlying stochastic structure of these high-dimensional data can be understood as a technical tool toward reproducibil-

---

✉ Rodney V. Fonseca
  ra192588@ime.unicamp.br

[1] Department of Statistics, University of Campinas, Campinas, SP 13083-970, Brazil

ity/repeatability or inherent to the problem under study. Either way, a precise apportionment of deterministic and random components is paramount. Examples of relevant data sets and areas as well as paradigms for the statistical analysis of functional data can be found in Ramsay and Silverman (2005) and Morettin et al. (2017).

Some features are found in specific problems and should be dealt with accordingly. For instance, intraday and/or inter-day dependences are common in financial functional series (Aue et al. 2017; Abadir et al. 2013; Pakoš 2011). Aggregate data may be useful for energy (Dias et al. 2013, 2015), market shares (Berry and Haile 2014), demand/supply studies (Canale and Ruggiero 2016), and many others (Shang 2016; Amighini et al. 2014; Cholaquidis et al. 2014). Any model in this framework must deal with a very common and basic property: the dimension of the functional space and its impact on the proposed solution. Functional data analysis poses serious hindrances to parametric models. Some Bayesian proposals that deal with these dimensionality issues can be found in Schillings and Schwab (2016), Suarez and Ghosal (2017), and Canale and Ruggiero (2016).

Successful nonparametric solutions are found through sparseness (Yan et al. 2018; Aneiros and Vieu 2016; Li et al. 2016; Yao et al. 2016; Devijver 2017; Qu et al. 2018; Sienkiewicz et al. 2017; Voronin and Daubechies 2017), principal component analysis (Mousavi and Sørensen 2018; Li et al. 2016; Imaizumi and Kato 2018; Mirafzal 2018; Lakraj and Ruymgaart 2017; Hooker and Roberts 2016; Shang 2016), thresholding (Mousavi and Sørensen 2018; Breunig and Johannes 2016; Amato et al. 2017; Yang et al. 2017; Ivanescu 2017; Røislien and Winje 2013; Salvatore et al. 2016; Johnstone and Lu 2009), penalizing procedures (Mousavi and Sørensen 2018; Comte et al. 2017; Amato et al. 2017; Fan et al. 2015; Sienkiewicz et al. 2017; Lorenz and Resmerita 2017), sufficiency (Zhang et al. 2018; Li and Song 2017), and others (Zhang et al. 2017; Belloni et al. 2017).

Here we follow the setup studied by Bathia et al. (2010). Their idea is to model observable curves that are driven by a finite-dimensional process plus a noise term. This can be applied to a time series composed of curves. The methodology proposed by Bathia et al. (2010) does not need the assumption of negligible noise as sample size increases (Hall and Vial 2006), exploring instead the dynamic structure of the observed curves. Eigenfunctions are used to represent the curves, and bootstrap resampling is proposed to sequentially estimate the finite functional dimension. The methods of the latter were used, for instance, by Horta and Ziegelmann (2018) to compute the dimension of time series of density functions of stock indexes for prediction purposes. Additionally, the problem addressed by Bathia et al. (2010) was generalized by Horta and Ziegelmann (2016) for time series of Hilbertian random elements.

In this work we consider models in which the curves lie on a finite-dimensional subspace, and we propose to use wavelets to estimate this unknown dimension. This novelty on the basis allows us to propose a series of bootstrapping procedures besides the original one introduced by Bathia et al. (2010). Similar asymptotic properties are attained. Moreover, computational and mathematical advantages are discussed. We also prove that the estimation procedure may be used for aggregate data.

The text is organized as follows. In Sect. 2 we discuss the idea of finite functional dimension. In Sect. 3 we present the proposed wavelet solution for the estimation of the functional dimension. Two cases of particular interest are discussed in Sect. 4. We

then present the theoretical results for the proposed algorithms in Sect. 5. Simulation studies and an application to real data are presented in Sects. 6 and 7, respectively. A discussion and final remarks can be found in Sect. 8.

2 Functional dimension estimation

In what follows, we shall describe the problem of estimating the finite dimension of curve time series (Bathia et al. 2010). Consider random functions $Y_1, Y_2, \ldots$ in a Hilbert space $L^2 = L^2(I)$ of square integrable functions defined in a compact $I \subset \mathbb{R}$, with inner product $\langle Y, X \rangle = \int_I Y(x)X(x)\mathrm{d}x, \forall Y, X \in L^2$. These curves usually are not perfectly observed, being subject to errors of numerical or experimental nature, for example, or due to its nature, as is the case of conditional density functions. This means that in practice we do not know the curves of interest $X_t, t = 1, \ldots, n$, but we might have a sample of estimates $Y_1, \ldots, Y_n$ obtained after applying some smoothing method to the data at hand. The observed curves Y_t are taken as satisfying

$$Y_t(x) = X_t(x) + \varepsilon_t(x), \quad x \in I, t = 1, \ldots, n, \tag{1}$$

where X_t and ε_t are not observed and ε_t is supposed to be a noise, in the sense that

1. $\mathbb{E}[\varepsilon_t(x)] = 0, \forall t$ and $\forall x \in I$,
2. $\mathrm{Cov}(\varepsilon_t(x), \varepsilon_{t+k}(y)) = 0, \forall x, y \in I$ when $k \neq 0$,
3. $\mathrm{Cov}(X_t(x), \varepsilon_s(y)) = 0, \forall x, y \in I$ and $\forall t, s$.

With these conditions, the error in estimating X_t is intrinsic to time t and exogenous with respect to X_t. We assume that $X_1, X_2, \ldots$ are stationary, such that

$$\mu(x) = \mathbb{E}[X_t(x)] \quad \text{and} \quad M_k(x, y) = \mathrm{Cov}(X_t(x), X_{t+k}(y))$$

do not depend on t. Using the same notation of Bathia et al. (2010), we shall denote an operator with the kernel $M_k(x, y)$ as M_k, such that for any $f \in L^2$ we have $M_k f(x) = \int_{\mathcal{I}} M_k(x, y) f(y)\mathrm{d}y$. Since the process X_t lies in L^2, it admits the Karhunen–Loève expansion (Bosq 2000), i.e., it can be represented on the basis $\{\varphi_1, \varphi_2, \ldots\}$ formed by eigenfunctions of the zero-lag covariance operator M_0 (cf. Corollary 1 of Horta and Ziegelmann (2016)). Our interest is to identify the cardinality of this basis, which we assume to have a finite number d of elements. In that case, the curve X_t can be represented as

$$X_t(x) = \mu(x) + \sum_{j=1}^{d} \xi_{tj}\varphi_j(x), \quad \forall t \tag{2}$$

where $\xi_{tj} = \langle X_t - \mu, \varphi_j \rangle$ is a zero-mean random variable. However, estimation of M_0 from the observed curves $\{Y_t\}$ is not straightforward because $\mathrm{Cov}(Y_t(x), Y_t(y)) = M_0(x, y) + \mathrm{Cov}(\varepsilon_t(x), \varepsilon_t(y))$, unless further assumptions over the errors are made (Hall and Vial 2006). Nevertheless, Bathia et al. (2010) proposed to tackle this problem considering lagged operators $M_k, k \geq 1$, since $\mathrm{Cov}(Y_t(x), Y_t(y)) = M_k(x, y)$. Hence, we consider to identify d using the positive kernel

$$K(x, y) = \sum_{k=1}^{p} \int M_k(x, z) M_k(y, z) \mathrm{d}z,$$

for some $p \geq 1$. The range of the operator K is contained in the range of M_0. However, we need also to consider that an equality holds for these ranges (cf. assumption (A1) of Horta and Ziegelmann (2016)). Therefore, we shall consider throughout the paper that the following assumption is true:

(Assumption K) The range of the zero-lag covariance operator M_0 is finite-dimensional and coincides with the range of the operator K.

This condition allows us to use the idea of Bathia et al. (2010) to estimate d when the data are linear functional time series, but the methodology is not appropriate when the data are composed of independent data and might not be adequate when they are composed of curve times series with nonlinear dynamics.

The operator K belongs to $\mathcal{S}$, the space of Hilbert–Schmidt operators (Bosq 2000), whose norm we shall denote by $\|\cdot\|_{\mathcal{S}}$. Under the representation (2) and assumptions 1, 2 and 3 above, the linear part of the dynamics of $Y_1, Y_2, \ldots$ is captured by the d-dimensional time series $\boldsymbol{\xi}_t = (\xi_{t1}, \ldots, \xi_{td})^{\top}$. We consider the observed kernel

$$\hat{K}(x, y) = \sum_{k=1}^{p} \int \hat{M}_k(x, z) \hat{M}_k(y, z) \mathrm{d}z,$$

where p is fixed and

$$\hat{M}_k(x, y) = \frac{1}{n-p} \sum_{t=1}^{n-p} (Y_t(x) - \bar{Y}(x))(Y_{t+k}(y) - \bar{Y}(y)),$$

with $\bar{Y}(x) = \sum_{t=1}^{n} Y_t(x)/n$. These kernels also belong to $\mathcal{S}$.

The maximum lag p in practice can be taken as a small positive integer value (Bathia et al. 2010). The authors' idea in identifying d is to obtain eigenfunctions of $\hat{K}(x, y)$ through eigenvectors and eigenvalues of a matrix of dimension $(n - p) \times (n - p)$ whose elements are computed from inner products involving Y_t and $\bar{Y}$. We employ in this paper a wavelet representation to perform the eigenanalysis of $\hat{K}(x, y)$. Although in our methodology below we are expanding K using an alternative basis, we are still interested in identifying the dimension d through the number of nonzero eigenvalues of K, which does not depend on the chosen basis. The alternative basis is not considered in the previous definitions, but we shall make use of it in order to improve empirical aspects of the methodology. In Sect. 3 we briefly introduce wavelet methods and present the proposed wavelet procedure for dimension estimation.

3 Wavelet-based functional dimension

We discuss here wavelets as forming orthonormal bases for $L^2(\mathbb{R})$. This is done for notational simplification only, with no lack of generality. Wavelet orthonormal bases

can be constructed for $L^2(I)$ for compact I as well. See, for instance, (Mallat 1998, p. 289) for a detailed presentation on Daubechies basis for $L^2([0, 1])$. An orthonormal wavelet basis can be constructed by means of a multiresolution analysis (MRA), which is a tool presented by Mallat (1989) that consists of a nested sequence of closed subspaces $\{V_n, n \in \mathbb{Z}\}$ in $L^2(\mathbb{R})$ satisfying:

1. $\cdots \subset V_{-2} \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \cdots$;
2. $\cap_n V_n = \{0\}$ and $\overline{\cup_n V_n} = L^2(\mathbb{R})$;
3. the subspaces V_n are self-similar, in the sense that the map $x \mapsto f(2^j x)$ lies in V_j if and only if the map $x \mapsto f(x)$ lies in V_0;
4. V_0 is the closed span of the set formed by integer translations of ϕ.

In the literature, ϕ is known as a scale function and the subspaces V_j can be seen as resolution levels of approximation for a $L^2(\mathbb{R})$ function. Mallat (1989) shows that any function $f \in L^2(\mathbb{R})$ can be approximated in V_j by

$$P_j f(x) = \sum_{k \in \mathbb{Z}} \langle f, \phi_{j,k} \rangle \phi_{j,k}(x),$$

where $P_j f$ here denotes the orthogonal projection of f in V_j. Based on properties 1 and 2 of the MRA, we have $\lim_{j \to \infty} P_j f = f$ and $\lim_{j \to -\infty} P_j f = 0$, i.e., higher resolutions provide better approximations to $f(x)$ whereas, the lower the resolution is, the closer to zero is the approximation. The rate of this approximation can be precisely evaluated when f belongs to certain functional spaces, like Sobolev and Besov spaces (Härdle et al. 1998).

The detail obtained after passing from a resolution j to $j + 1$ can be analyzed considering the orthogonal complement of V_j in V_{j+1}, which is denoted by W_j. Hence, $V_{j+1} = V_j \oplus W_j$, so that $V_j = \bigoplus_{k < j} W_k$ and $L^2(\mathbb{R}) = \bigoplus_{k \in \mathbb{Z}} W_k$. Mallat (1989) shows that $\{\psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k), k \in \mathbb{Z}\}$ is an orthonormal basis of W_j and $\{\psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k); k \in \mathbb{Z}, j \in \mathbb{Z}\}$ is a basis of $L^2(\mathbb{R})$. From the MRA, we have that f can be represented as

$$f(x) = \sum_{k \in \mathbb{Z}} \langle f, \phi_{j_0,k} \rangle \phi_{j_0,k}(x) + \sum_{j \geq j_0} \sum_{k \in \mathbb{Z}} \langle f, \psi_{j,k} \rangle \psi_{j,k}(x),$$

where the first series is the projection of f in resolution j_0, $\langle f, \phi_{j_0,k} \rangle$ being called as an approximation coefficient, and the second series contains the details corresponding to resolutions greater or equal to j_0, with $\langle f, \psi_{j,k} \rangle$ being called a detail coefficient. A widely used system consists in the Daubechies wavelets, which have compact support and nice properties regarding function regularity. We denote DAUBN as a Daubechies wavelet with N null moments. The case $N = 1$ corresponds to the famous Haar wavelet (Vidakovic 1999). Since $V_0 \subset V_1$, we have

$$\phi(x) = \sum_{k \in \mathbb{Z}} h_k \sqrt{2} \phi(2x - k), \tag{3}$$

where $\mathbf{h} = \{h_k, k \in \mathbb{Z}\}$ is square summable and is known as the wavelet filter. Equation (3) is known as scaling equation (Vidakovic 1999). In addition, since $W_0 \subset V_1$, we have that

$$\psi(x) = \sum_{k \in \mathbb{Z}} g_k \sqrt{2}\phi(2x - k),$$

for some $\mathbf{g} = \{g_k, k \in \mathbb{Z}\}$. It is possible to show that $g_n = (-1)^n h_{1-n}$, which is called the quadrature mirror relation. These filters play an important role on the computation of wavelet and scale functions. For instance, they are used to compute the discrete wavelet transformation with a cascade algorithm that uses $\mathbf{h}$ and $\mathbf{g}$ as filters of convolution operators.

Our idea is to employ wavelet decompositions of the observed functions Y_t to estimate the dimension of the process that generates these curves. For convenience, we follow the notation used by Pinheiro and Vidakovic (1997), considering

$$Y_t(x) = \sum_j a_j^t \phi_j(x),$$

where a_j^t represents both wavelet and approximation coefficients and ϕ_j represents both scale and wavelet functions. Hence,

$$Y_t(x) - \bar{Y}(x) = \sum_j (a_j^t - \bar{a}_j)\phi_j(x) = \sum_j c_j^t \phi_j(x),$$

where $\bar{a}_j = n^{-1} \sum_{t=1}^n a_j^t$ and c_j^t is given by $\langle Y_t - \bar{Y}, \phi_j \rangle$. Therefore, we obtain

$$\hat{M}_k(x, y) = \frac{1}{n - p} \sum_{t=1}^{n-p} \sum_j \sum_{j'} c_j^t \phi_j(x)\phi_{j'}(y)c_{j'}^{t+k}$$

and using that the functions $\phi_j(x)$ form an orthonormal system, we get

$$\hat{K}(x, y) = \frac{1}{(n - p)^2} \sum_{k=1}^p \sum_{t=1}^{n-p} \sum_{s=1}^{n-p} \sum_{j,j',l} c_j^t c_{j'}^s c_l^{t+k} c_l^{s+k} \phi_j(x)\phi_{j'}(y).$$

Our objective is to find eigenfunctions of $\hat{K}(x, y)$. Even though we are dealing with an infinite-dimensional problem, in practice we work with a vector of Y_t evaluated on a grid of points selected appropriately. Applying the discrete wavelet transform to those points, we obtain a finite number J of coefficients representing the discretized Y_t on the wavelet domain. This approach is called by Bathia et al. (2010) as approximation via discretization and allows us to employ the wavelet methods as well as other basis functions (splines, Hermite polynomials, etc). We opt for wavelet basis due to its sparse representation of thresholded functions and because it allows to perform different types of bootstraps on dimensionality tests. Notice also that this representation is loss free

and can be inverted (Vidakovic 1999). Therefore, in what follows we consider that Y_t can be represented on a basis with cardinality J. For the infinite-dimensional case, see Sect. 4 or Bathia et al. (2010). Considering now a candidate eigenfunction h_m, its wavelet representation is given by

$$h_m(x) = \sum_{j''} b_{j''}^m \phi_{j''}(x) = \Phi(x)^\top \mathbf{b}^m, \tag{4}$$

where $\Phi(x) = (\phi_1(x), \ldots, \phi_J(x))^\top$ and $\mathbf{b}^m = (b_1^m, \ldots, b_J^m)^\top$. Thus, considering that the same basis is being used in all decompositions, the indexes j, j' and j'' also vary in $\{1, \ldots, J\}$. Then

$$\int \hat{K}(x, y) h_m(y) \mathrm{d}y = \sum_j \left(\frac{1}{(n-p)^2} \sum_{k=1}^p \sum_{t=1}^{n-p} \sum_{s=1}^{n-p} \sum_{j',l} c_j^t c_{j'}^s c_l^{t+k} c_l^{s+k} b_{j'}^m \right) \phi_j(x)$$

$$= \sum_j (D_j \mathbf{b}^m) \phi_j(x) = \Phi(x)^\top (D\mathbf{b}^m), \tag{5}$$

where D_j represents a $1 \times J$ vector which is the j-th row of the $J \times J$ matrix D, whose (j, j') element is

$$D_{j,j'} = \frac{1}{(n-p)^2} \sum_{k=1}^p \sum_{t=1}^{n-p} \sum_{s=1}^{n-p} \sum_l c_j^t c_{j'}^s c_l^{t+k} c_l^{s+k}.$$

This matrix can also be obtained in the following way. Consider the $J \times n$ matrix $\mathcal{C}$ whose t-th column contains J coefficients c_j^t, then letting $\mathcal{C}_{J \times (k_1:k_2)}$ be a submatrix obtained selecting from the k_1-th until the k_2-th column of $\mathcal{C}$, $k_1 < k_2$, we have

$$D = \frac{1}{(n-p)^2} \mathcal{C}_{J \times (1:n-p)} \left(\sum_{k=1}^p (\mathcal{C}_{J \times (k+1:n-p+k)})^\top \mathcal{C}_{J \times (k+1:n-p+k)} \right) (\mathcal{C}_{J \times (1:n-p)})^\top.$$

Therefore, from (4) and (5), our goal is to find $\mathbf{b}^m$ such that $\Phi(x)^\top (D\mathbf{b}^m) = \lambda_m \Phi(x)^\top \mathbf{b}^m$ for some constant λ_m and $\forall x \in I$, i.e., we wish to solve for $\mathbf{b}^m$ the system

$$(D\mathbf{b}^m) = \lambda_m \mathbf{b}^m,$$

i.e., taking $\mathbf{b}^m$ as an eigenvector of D, with λ_m being its associated eigenvalue. Thus, letting $\mathbf{b}^1, \ldots, \mathbf{b}^{\hat{d}}$ be the eigenvectors of D associated with its $\hat{d}$ largest eigenvalues, we have that $h_1, \ldots, h_{\hat{d}}$ as in Eq. (4) are eigenfunctions of the operator $\hat{K}$. It is worth mentioning that this procedure resembles the functional principal component analysis (PCA) (Ramsay and Silverman 2005, p. 162), with the difference that instead of the matrix D we would consider for the latter the $J \times J$ matrix $n^{-1}\mathcal{C}\mathcal{C}^\top$.

Since the h_m's are orthonormal, we have that $\{h_1(\cdot), \ldots, h_{\hat{d}}(\cdot)\}$ forms an orthonormal system in L^2, the estimate of the functional of interest being thus

$$\hat{Y}_t(x) = \bar{Y}(x) + \sum_{l=1}^{\hat{d}} \hat{\eta}_{tl} h_l(x), \tag{6}$$

$\hat{\eta}_{tl} = \langle Y_t - \bar{Y}, h_l \rangle$. Hence, it follows from (6) that $\hat{\eta}_{tl} = \sum_j c_j^t b_j^l$, with the dynamics of Y_t being modeled through the multivariate time series $\hat{\boldsymbol{\eta}}_t = (\hat{\eta}_{t1}, \ldots, \hat{\eta}_{t\hat{d}})'$.

3.1 Bootstrap tests of dimensionality

We now present how the functional dimension can be estimated by means of bootstrap tests. We compare four bootstrap procedures for the estimation of d. The first is described by Bathia et al. (2010). Given the eigenvalues as above $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$, tests of the null hypothesis $\lambda_{d_0+1} = 0$ are performed sequentially until the first λ_{d_0+1} which significantly equals zero is found. In this case, the estimated dimension is taken as d_0. For instance, suppose that we want to test $H_0 : \lambda_{d_0+1} = 0$ against $H_1 : \lambda_{d_0+1} > 0$ for some positive integer d_0. We construct a function imposing the restriction of H_0:

$$\tilde{Y}_t(x) = \bar{Y}(x) + \sum_{l=1}^{d_0} \hat{\eta}_{tl} h_l(x) = \sum_j \left\{ \bar{a}_j + \sum_{l=1}^{d_0} \hat{\eta}_{tl} b_j^l \right\} \phi_j(x),$$

for which we already have the wavelet decomposition of $\tilde{Y}_t(x)$. Then we obtain the residuals $\hat{\epsilon}_t(x) = Y_t(x) - \tilde{Y}_t(x)$ and perform the following steps:

1. for each $t = 1, \ldots, n$, randomly select (with replacement) a residual $\epsilon_t^b(x)$ from $\{\hat{\epsilon}_1(x), \ldots, \hat{\epsilon}_n(x)\}$ and take $Y_t^b(x) = \tilde{Y}_t(x) + \epsilon_t^b(x)$;
2. obtain for the bootstrap sample $Y_1^b(x), \ldots, Y_n^b(x)$ the matrix D, and compute its $(d_0 + 1)$-th largest eigenvalue $\lambda_{d_0+1}^b$;
3. repeat steps 1 and 2 a large number of times, say B, then compute the bootstrap p value $p_{\text{boot}} = \#\{\hat{\lambda}_{d_0+1} < \lambda_{d_0+1}^b\}/(B + 1)$, where $\hat{\lambda}_{d_0+1}$ is the $(d_0 + 1)$-th largest eigenvalue obtained for $\hat{Y}$. Reject H_0 if p_{boot} is lower than some previously specified significance value.

From the wavelet decompositions of $\bar{Y}$ and h_l, $l = 1, \ldots, d_0$, we have

$$Y_t^b(x) = \tilde{Y}_t(x) + (Y_t(x) - \tilde{Y}_t(x))^b = \sum_j \left\{ a_j^{tb} + \sum_{l=1}^{d_0} (\hat{\eta}_{tl} - \hat{\eta}_{tl}^b) b_j^l \right\} \phi_j(x),$$

where the superscript b indicates the bootstrapped terms. Hence, the resampling of step 1 can be performed directly on the coefficients a_j^t and $\hat{\eta}_{tl}$, which reduces the computation time of the bootstrap procedure.

Three other bootstrap resampling schemes follow the same algorithm with some modifications that provide alternative methods to the previous procedure by exploring

the wavelet decomposition. Since wavelet thresholding leads to optimal minimax rates (Vidakovic 1999), it is natural to pursue bootstrapping based on thresholding rules. We propose three different bootstraps which make use of thresholding. The first one applies thresholding to the curves' estimation. The second one adds to the former residual thresholding as well. Finally, a third bootstrap proposed employs the idea of wavestrapping (Percival et al. 2000), which resamples on the wavelet domain.

Initially, note that in the previous bootstrap test the wavelet decomposition of the observed curves is obtained without thresholding and then we apply the bootstrap procedure to test the eigenvalues of the corresponding matrices D. Taking this into account, in the second bootstrap procedure we perform the same steps as above with the only difference that a hard thresholding is applied for the observed curves' coefficients before computing D. In this case, coefficients greater than a certain threshold value are kept unchanged, whereas the remaining are shrunk to zero. For different ways of choosing the threshold value, the reader is referred to Vidakovic (1999). In the third procedure, we apply a hard thresholding (indicated by the index thr) to $\bar{Y}$ and h_l in Eq. (6), such that

$$Y_t(x) = \bar{Y}(x) + \sum_{l=1}^{\hat{d}} \hat{\eta}_{tl} h_l(x) + \hat{\epsilon}_t(x) = \bar{Y}^{\text{thr}}(x) + \sum_{l=1}^{\hat{d}} \hat{\eta}_{tl} h_l^{\text{thr}}(x) + \hat{\epsilon}_t^{\text{thr}}(x),$$

where

$$\hat{\epsilon}_t^{\text{thr}}(x) = \hat{\epsilon}_t(x) + (\bar{Y}(x) - \bar{Y}^{\text{thr}}(x)) + \sum_{l=1}^{\hat{d}} \hat{\eta}_{tl}(h_l(x) - h_l^{\text{thr}}(x)).$$

Then, we apply the bootstrap procedure with $\hat{\epsilon}_t^{\text{thr}}(x)$ instead of $\hat{\epsilon}_t(x)$ and $\tilde{Y}_t(x)$ formed of $\bar{Y}^{\text{thr}}$ and h_l^{thr}. Hence, we have that the bootstrap curve in this case is

$$Y_t^{\text{b}}(x) = \sum_j \left\{ a_j^{t\text{b}} + \sum_{l=1}^{d_0} [\hat{\eta}_{tl} - \hat{\eta}_{tl}^{\text{b}}](b_j^l)^{\text{thr}} \right\} \phi_j(x),$$

which is similar to the previous bootstrap method, using the thresholded wavelet coefficient $(b_j^l)^{\text{thr}}$ and $a_j^{t\text{b}}$ instead of b_j^l and a_j^t. The last procedure we consider is based on the wavestrapping technique proposed by Percival et al. (2000), where for each t, a residual $\epsilon_t^{\text{b}}(x)$ is randomly selected from $\{\hat{\epsilon}_1(x), \ldots, \hat{\epsilon}_n(x)\}$, then its wavelet coefficients are resampled (with replacement) inside each detail level to obtain the coefficients of a new bootstrap residual, which is used to form the bootstrap curve $\tilde{Y}_t(x)$. An advantage of the wavestrapping over the first three bootstrap methods is that it has a much larger number of possible residuals, since it is based not only on random selection from n elements, but also considers resampling from their wavelet coefficients to obtain random samples.

4 Two cases of practical interest

The proposed wavelet method of functional dimension estimation can be applied on a variety of time series curves. In this section we highlight two such cases.

4.1 Functional data aggregation

There is a considerable attention on the literature to investigate situations where analyzing curves individually is not possible or very costly, requiring an analysis based on aggregate curves, as described in the references from Sect. 1. Moreover, aggregate time series are also well described in the literature, as can be seen in (Wei 2006, Chapter 20), since this kind of data is often found, as happens with economic data (Abraham 1982), for example. In this paper we consider aggregate data to identify the dimension of a functional time series.

Suppose that the observed curves as defined by (1) are not originally aggregates, but the number of observations for each time is not large. We then take linear combinations of $Y_t(\cdot)$ for a fixed number of successive t's as observed functions before applying the methods of dimensionality identification. Each $Y_t(\cdot)$ is multiplied by a weight that controls its contribution to time t. Hence, a model where a number δ of functions like model (1) are aggregated can be represented as

$$\mathcal{Y}_t(x) = \mathcal{X}_t(x) + \mathcal{E}_t(x), \quad x \in I,$$

where $\mathcal{Y}_t(x) = \sum_{s=t-\delta+1}^{t} \omega_{t-s} Y_s(x)$, $\mathcal{X}_t(x) = \sum_{s=t-\delta+1}^{t} \omega_{t-s} X_s(x)$ and $\mathcal{E}_t(x) = \sum_{s=t-\delta+1}^{t} \omega_{t-s} \varepsilon_s(x)$, for a positive integer δ and some coefficients $\omega_{\delta-1}, \ldots, \omega_0$ so that the problem is similar to a moving average process. Denoting $\sigma_\varepsilon^2(x, y) = \mathrm{Cov}(\varepsilon_t(x), \varepsilon_t(y))$, we have for $k \in \mathbb{N}$

$$\mathrm{Cov}(\mathcal{E}_t(x), \mathcal{E}_{t+k}(y)) = \begin{cases} 0 & \text{if } \delta - 1 < k, \\ \sigma_\varepsilon^2(x, y) \sum_{s=t-\delta+1}^{t} \sum_{v=t+k-\delta+1}^{t+k} \omega_{t-s} \omega_{t+k-v} & \text{if } \delta - 1 \geq k. \end{cases} \quad (7)$$

Considering assumption **K** on Sect. 2, using the decomposition of $X_t(\cdot)$ given by (2), we have

$$Y_t(x) = \mu(x) + \sum_{j=1}^{d} \xi_{tj} \varphi_j(x) + \varepsilon_t(x). \quad (8)$$

The aggregate observed function has the following decomposition:

$$\mathcal{Y}_t(x) = \mu(x) \sum_{s=t-\delta+1}^{t} \omega_{t-s} + \sum_{j=1}^{d} \varphi_j(x) G_{tj}(x) + \mathcal{E}_t(x),$$

where $G_{tj}(x) = \sum_{s=t-\delta+1}^{t} \omega_{t-s}\xi_{sj}$. It follows that

$$
\begin{aligned}
&\mathrm{Cov}\{\mathcal{Y}_t(u), \mathcal{Y}_{t+k}(v)\} \\
&= \sum_{s=t-\delta+1}^{t} \sum_{l=t+k-\delta+1}^{t+k} \omega_{t-s}\omega_{t+k-l}M_{|l-s|}(u,v) + \mathrm{Cov}\{\mathcal{E}_t(u), \mathcal{E}_{t+k}(v)\}.
\end{aligned}
$$

If $k > \delta - 1$, using Eq. (7) we have

$$
\mathrm{Cov}\{\mathcal{Y}_t(u), \mathcal{Y}_{t+k}(v)\} = \sum_{s=t-\delta+1}^{t} \sum_{l=t+k-\delta+1}^{t+k} \omega_{t-s}\omega_{t+k-l}M_{l-s}(u,v) = \mathcal{M}_k(u,v).
$$

Let $\lambda_1 \geq \cdots \geq \lambda_d$ be eigenvalues of the operator M_0, with corresponding eigenfunctions $\varphi_1, \ldots, \varphi_d$. The serial dependence of $Y_t(\cdot)$ is determined by $\boldsymbol{\xi}_t = (\xi_{t1}, \ldots, \xi_{td})^\top$, with $\mathbb{E}(\boldsymbol{\xi}_t) = \mathbf{0}$ and $\mathrm{Var}(\boldsymbol{\xi}_t) = \mathrm{diag}\{\lambda_1, \ldots, \lambda_d\}$. Additionally, under assumption **K**, we can use the approach of Bathia et al. (2010) to estimate the number of nonzero eigenvalues of the operator M_0 through lagged covariance operators. We define

$$
\begin{aligned}
\mathcal{N}_k(u,v) &= \int_I \mathcal{M}_k(u,z)\mathcal{M}_k(v,z)\mathrm{d}z \\
&= \sum_s \sum_l \sum_{s'} \sum_{l'} \omega_{t-s}\omega_{t+k-l}\omega_{t-s'}\omega_{t+k-l'} \int_I M_{l-s}(u,z)M_{l'-s'}(v,z)\mathrm{d}z,
\end{aligned}
$$

where $t - \delta + 1 \leq s, s' \leq t$ and $t + k - \delta + 1 \leq l, l' \leq t + k$. Here and throughout this section, we shall write such summation this way, the summation limits being implicit.

We have that $M_k(u,v) = \sum_{i,j=1}^{d} \sigma_{ij}^{(k)}\varphi_i(u)\varphi_j(v)$, where $\Sigma_k = \mathbb{E}(\boldsymbol{\xi}_t\boldsymbol{\xi}_{t+k}^\top) = \{\sigma_{ij}^{(k)}\}$. Therefore,

$$
\mathcal{M}_k(u,v) = \sum_{i,j=1}^{d} \alpha_{ij}^{k}\varphi_i(u)\varphi_j(v),
$$

where

$$
\alpha_{ij}^{k} = \sum_{s=t-\delta+1}^{t} \sum_{l=t+k-\delta+1}^{t+k} \omega_{t-s}\omega_{t+k-l}\sigma_{ij}^{(l-s)} = \sum_{s=0}^{\delta-1}\sum_{l=0}^{\delta-1} \omega_s\omega_l\sigma_{ij}^{(l-s+k)}. \tag{9}
$$

Since $1 \leq l - s + k \leq p$, we have that $1 \leq -\delta + 1 + k$ and $\delta - 1 + k \leq p$, and $\delta \leq k \leq p - \delta + 1$. Moreover,

$$
\mathcal{N}_k(u, v) = \sum_{i,j=1}^{d} \left(\sum_{l=1}^{d} \alpha_{il}^k \alpha_{jl}^k \right) \varphi_i(u) \varphi_j(v).
$$

Then, we shall consider $\mathcal{K}(u, v) = \sum_{k=\delta}^{p-\delta+1} \mathcal{N}_k(u, v)$ to estimate the process' dimension, with fixed integers δ and p, $p \geq 2\delta - 1$. For the aggregate data case, we consider as estimator of the covariance kernel

$$
\hat{M}_k(u, v) = \frac{1}{n - \delta - p + 1} \sum_{j=1}^{n-p-\delta+1} \{Y_j(u) - \bar{Y}(u)\}\{Y_{j+k}(v) - \bar{Y}(v)\},
$$

its aggregate version being given by

$$
\hat{\mathcal{M}}_k(u, v) = \sum_{s=t-\delta+1}^{t} \sum_{l=t+k-\delta+1}^{t+k} \omega_{t-s} \omega_{t+k-l} \hat{M}_{l-s}(u, v).
$$

Therefore,

$$
\hat{\mathcal{K}}(u, v) = \sum_{k=\delta}^{p-\delta+1} \int_I \hat{\mathcal{M}}_k(u, z) \hat{\mathcal{M}}_k(v, z) \mathrm{d}z
$$

$$
= \sum_{k=\delta}^{p-\delta+1} \sum_{s,l,s',l'} \frac{\omega_{t-s} \omega_{t+k-l} \omega_{t-s'} \omega_{t+k-l'}}{(n - p - \delta + 1)^2} \sum_{i,j=1}^{n-p-\delta+1} \{Y_i(u) - \bar{Y}(u)\}
$$

$$
\times \{Y_j(v) - \bar{Y}(v)\} \langle Y_{i+l-s} - \bar{Y}, Y_{j+l'-s'} - \bar{Y} \rangle.
$$

In Sect. 5, it is shown in Proposition 2 that $\hat{\mathcal{K}}$ shares the same nonzero eigenvalues as a $(n - p - \delta + 1) \times (n - p - \delta + 1)$ matrix, say $\mathbf{K}^*$. Moreover, letting $\boldsymbol{\gamma}_j = (\gamma_{1j}, \ldots, \gamma_{n-p-\delta+1,j})^\top$, $j = 1, \ldots, \hat{d}$, be eigenvectors of $\mathbf{K}^*$ corresponding to the $\hat{d}$ largest eigenvalues, we have that

$$
\sum_{i=1}^{n-p-\delta+1} \gamma_{ij}\{Y_i(\cdot) - \bar{Y}(\cdot)\}, \quad j = 1, \ldots, \hat{d}
$$

are eigenfunctions of $\hat{\mathcal{K}}$. These eigenfunctions can be transformed into an orthonormal system $\hat{\psi}_1(\cdot), \ldots, \hat{\psi}_{\hat{d}}(\cdot)$ using a Gram–Schmidt algorithm.

Wavelets can be applied to aggregate data analogously to what was done in Sect. 3. Taking the wavelet decomposition of $Y_t(\cdot)$ on the expression for $\hat{\mathcal{K}}(\cdot, \cdot)$, we have

$$
\begin{aligned}
\hat{\mathcal{K}}(u, v) = \sum_{k=\delta}^{p-\delta+1} \sum_{s,l,s',l'} & \frac{\omega_{t-s}\omega_{t+k-l}\omega_{t-s'}\omega_{t+k-l'}}{(n-p-\delta+1)^2} \\
\times & \sum_{i,j=1}^{n-p-\delta+1} \sum_{r',r'',r'''} c_{r'}^i c_{r''}^j c_{r'''}^{i+l-s} c_{r'''}^{j+l'-s'} \phi_{r'}(u)\phi_{r''}(v),
\end{aligned}
$$

and considering $h_m(y) = \sum_q b_q^m \phi_q(y)$ the wavelet decomposition of an eigenfunction of $\hat{\mathcal{K}}(\cdot, \cdot)$, we have that

$$
\int \hat{\mathcal{K}}(x, y) h_m(y) \mathrm{d}y = \Phi(x)^\top \left(D\mathbf{b}^m \right)
$$

where in this case, the element (r', r'') of D is given by

$$
D_{r',r''} = \sum_{k=\delta}^{p-\delta+1} \sum_{s,l,s',l'} \frac{\omega_{t-s}\omega_{t+k-l}\omega_{t-s'}\omega_{t+k-l'}}{(n-p-\delta+1)^2} \sum_{i,j=1}^{n-p-\delta+1} \sum_{r'''} c_{r'}^i c_{r''}^j c_{r'''}^{i+l-s} c_{r'''}^{j+l'-s'}.
$$

Hence, we can estimate the eigenvalues of $\mathcal{K}(\cdot, \cdot)$ computing the eigenvalues of this matrix D, and the corresponding eigenvectors $\mathbf{b}_1, \mathbf{b}_2, \ldots$ contain the wavelet coefficients of the eigenfunctions of that operator.

4.2 Density time series

A common random curve analyzed in applications is the density function of some random variable of interest. The problem of estimating the dimension of density functions was investigated by Horta and Ziegelmann (2018), which apply the method of Bathia et al. (2010) to financial data. The former considered as curves of interest density functions f_t taking values on $L^2(I)$, $I \subset \mathbb{R}$. The observed densities can be taken as curves g_t obtained after applying some density estimation method to the data at hand. Therefore, the assumption on the model is like Eq. (1), say

$$
g_t(x) = f_t(x) + \epsilon_t(x), \quad x \in I,
$$

with ϵ_t being a noise satisfying the same assumptions made for model (1), but with the additional condition that $\int \epsilon_t(x)\mathrm{d}x = 0$, since both f_t and g_t must integrate one. Nevertheless, according to Horta and Ziegelmann (2018), this last assumption is not easily satisfied since estimated density functions are usually biased, but they do discuss a workaround.

Wavelet-based estimators for dependent time series density functions have some results established in the literature regarding its performance and consistency, like the contributions of Masry (1994, 1997) and Chacón and Rodríguez-Casal (2005).

Another approach to analyze the dimension of the density functions follows from the idea of Pinheiro and Vidakovic (1997), where, instead of estimating the density directly, we estimate its square root, with a wavelet estimator $\check{g}_t$ say. This change has two main advantages. First, the density can be estimated taking the square of $\check{g}_t$, which ensures that we obtain only nonnegative values for the estimated density; second, letting $\check{g}_t(x) = \sum_j \check{a}_j \phi_j(x), x \in I$, be the estimator's wavelet decomposition, by normalizing these coefficients such that $\sum_j \check{a}_j^2 = 1$, it follows from Parseval's identity that

$$\|\check{g}_t\|_{L^2(I)}^2 = \int_I \check{g}_t(x)^2 \mathrm{d}x = 1,$$

which guarantees that $\check{g}_t^2$ is a *bona fide* estimator of the density function and that $\check{g}_t$ belongs to $L^2(I)$. Hence, by shifting attention to $\sqrt{f_t}$ and applying the method of Pinheiro and Vidakovic (1997), one can evaluate the dimension of $\sqrt{f_t}$, with the benefits of having automatically integral equal to one and nonnegative estimates for the observed densities.

5 Theoretical results

We prove in this section that the dimension estimators for both aggregate and non-aggregate data have the same asymptotic properties proven by Bathia et al. (2010) for non-aggregate data. Proposition 1 states that the eigenfunctions of $\mathcal{K}$ span the space that generates the curve time series. Proposition 2 shows how to obtain eigenvalues of $\hat{\mathcal{K}}(\cdot, \cdot)$ as presented with the previous heuristic. The last result is a theorem showing convergence for the covariance operators and eigenvalues.

Proposition 1 *Let $\mathcal{M}$ be the space of dimension d that generates the curve time series. Also, suppose $\Sigma_k = \left\{\sigma_{ij}^{(k)}\right\}$ has full rank for some $k \in \mathbb{N}$. Then, $\mathcal{N}_k$ and $\mathcal{K}$ (for $p \geq k$) have exactly d nonzero eigenvalues and $\mathcal{M}$ is spanned by the corresponding functions.*

Proof Let us denote the adjoint operator of $\mathcal{M}_k$ by $\mathcal{M}_k^*$. Since for any $f \in L^2(I)$ we have

$$(\mathcal{N}_k f)(u) = \int_I \mathcal{N}_k(u, v) f(v) \mathrm{d}v = (\mathcal{M}_k \mathcal{M}_k^* f)(u),$$

it follows that $\mathcal{N}_k = \mathcal{M}_k \mathcal{M}_k^*$. In this case, $\hat{\mathcal{K}} = \sum_{k=\delta}^{p-\delta+1} \mathcal{M}_k \mathcal{M}_k^*$. We also have range$(\mathcal{N}_k)$ = range$(\mathcal{M}_k \mathcal{M}_k^*)$ = range$(\mathcal{M}_k)$, where range$(\cdot)$ is the operator's range.

We can also represent $\mathcal{M}_k$ as $\sum_{i,j=1}^d \alpha_{ik}^{(k)} \varphi_i \otimes \varphi_j$, where the symbol $\otimes$ indicates that $\mathcal{M}_k f = \sum_{i,j=1}^d \alpha_{ik}^{(k)} \langle \varphi_i, f \rangle \varphi_j$ for $f \in L^2$, and $\alpha_{ik}^{(k)}$ is given in Equation 9. Then

$$(\mathcal{M}_k f)(u) = \sum_{i=1}^{d} \lambda_i^{(k)} \langle \varphi_i, f \rangle \rho_i^{(k)}(u),$$

where

$$\rho_i^{(k)}(u) = \sum_{j=1}^{d} \frac{\alpha_{ij}^{(k)}}{\lambda_i^{(k)}} \varphi_j(u) \quad \text{and} \quad \lambda_i^{(k)} = \left\| \sum_{j=1}^{d} \alpha_{ij}^{(k)} \varphi_j \right\|.$$

Let now $\boldsymbol{\beta}$ be an arbitrary vector in $\mathbb{R}^d$, $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_d)^\top$, $\boldsymbol{\rho}_k = (\rho_1^{(k)}, \ldots, \rho_d^{(k)})^\top$ and $\mathbf{A}_k = \{\alpha_{ij}^{(k)}\}$. Since $\{\varphi_j, 1 \le j \le d\}$ is an orthonormal system, the equation

$$\boldsymbol{\beta}^\top \boldsymbol{\rho}_k = \boldsymbol{\beta}^\top \mathbf{A}_k \boldsymbol{\varphi} = 0$$

has a nontrivial solution iff $\boldsymbol{\beta}^\top \mathbf{A}_k = \mathbf{0}^\top$, i.e., for all j it holds that

$$\sum_{r=1}^{d} \beta_r \alpha_{rj}^{(k)} = \sum_{r=1}^{d} \beta_r \left(\sum_{s,l} \omega_{t-s} \omega_{t+k-l} \sigma_{rj}^{(l-s)} \right) = \sum_{s,l} \omega_{t-s} \omega_{t+k-l} \left(\sum_{r=1}^{d} \beta_r \sigma_{rj}^{(l-s)} \right) = 0,$$

or in matrix form

$$\begin{bmatrix} \omega_{\delta-1} \\ \vdots \\ \omega_0 \end{bmatrix}^\top \begin{bmatrix} \sum_{r=1}^{d} \beta_r \sigma_{rj}^{(k)} & \sum_{r=1}^{d} \beta_r \sigma_{rj}^{(k+1)} & \cdots & \sum_{r=1}^{d} \beta_r \sigma_{rj}^{(k+\delta-1)} \\ \sum_{r=1}^{d} \beta_r \sigma_{rj}^{(k-1)} & \sum_{r=1}^{d} \beta_r \sigma_{rj}^{(k)} & \cdots & \sum_{r=1}^{d} \beta_r \sigma_{rj}^{(k+\delta-2)} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{r=1}^{d} \beta_r \sigma_{rj}^{(k-\delta+1)} & \sum_{r=1}^{d} \beta_r \sigma_{rj}^{(k-\delta+2)} & \cdots & \sum_{r=1}^{d} \beta_r \sigma_{rj}^{(k)} \end{bmatrix} \begin{bmatrix} \omega_{\delta-1} \\ \vdots \\ \omega_0 \end{bmatrix} = 0.$$

$$(10)$$

It is known that some quadratic form satisfies $x^\top A x = 0$ for all x iff A is skew-symmetric, i.e., $A + A^\top$ exists and equals the zero matrix. Since the coefficients $\boldsymbol{\omega} = (\omega_{\delta-1}, \ldots, \omega_0)$ are arbitrary, this holds for all $\boldsymbol{\omega} \in \mathbb{R}^d$ and for all $j \in \{1, \ldots, d\}$. Therefore, the matrix in (10) must be skew-symmetric, which leads to

$$\sum_{r=1}^{d} \beta_r \sigma_{rj}^{(k)} = 0, \quad \forall j = 1, \ldots, d.$$

Since $\Sigma_k = \{\sigma_{ij}^{(k)}\}$ is of full rank by assumption, the only solution is $\boldsymbol{\beta} = \mathbf{0}$, which implies that $\{\rho_j^{(k)}\}$ is linearly independent. Applying the same arguments of Lemma 2 of Bathia et al. (2010), the result follows. □

Proposition 2 *The operator $\hat{\mathcal{K}}(\cdot, \cdot)$ has the same nonzero eigenvalues of the finite matrix $\mathbf{K}^*$ of dimension $(n - p - \delta + 1) \times (n - p - \delta + 1)$, whose (m, i)-th element is*

$$
\sum_{r=1}^{n-p-\delta+1} \sum_{k=\delta}^{p-\delta+1} \sum_{s,l,s',l'} \frac{\omega_{t-s}\omega_{t+k-l}\omega_{t-s'}\omega_{t+k-l'}}{(n - p - \delta + 1)^2} \langle Y_{m+l-s} - \bar{Y}, Y_{r+l'-s'} - \bar{Y} \rangle
$$
$$
\times \langle Y_r - \bar{Y}, Y_i - \bar{Y} \rangle.
$$

Proof Let $\hat{\theta}_j > 0$ be an eigenvalue of $\mathbf{K}^*$ and $\boldsymbol{\gamma}_j = (\gamma_{1j}, \ldots, \gamma_{n-p-\delta+1,j})^\top$ its corresponding eigenvector. The m-th element of $\mathbf{K}^* \boldsymbol{\gamma}_j = \hat{\theta}_j \boldsymbol{\gamma}_j$ is

$$
\frac{1}{(n - p - \delta + 1)^2} \sum_{i,r=1}^{n-p-\delta+1} \sum_{k=\delta}^{p-\delta+1} \sum_{s,l,s',l'} \omega_{t-s}\omega_{t+k-l}\omega_{t-s'}\omega_{t+k-l'}
$$
$$
\times \langle Y_{m+l-s} - \bar{Y}, Y_{r+l'-s'} - \bar{Y} \rangle \langle Y_r - \bar{Y}, Y_i - \bar{Y} \rangle \gamma_{ij} = \hat{\theta}_j \gamma_{mj}.
$$

Consider the function $\tilde{\psi}_j(\cdot) = \sum_{i=1}^{n-p-\delta+1} \gamma_{ij}\{Y_i(\cdot) - \bar{Y}(\cdot)\}$. Then

$$
(\hat{\mathcal{K}}\tilde{\psi}_j)(u) = \int_I \hat{\mathcal{K}}(u, v)\tilde{\psi}_j(v)\mathrm{d}v
$$
$$
= \sum_{k=\delta}^{p-\delta+1} \sum_{s,l,s',l'} \frac{\omega_{t-s}\omega_{t+k-l}\omega_{t-s'}\omega_{t+k-l'}}{(n - p - \delta + 1)^2} \sum_{m,r=1}^{n-p-\delta+1} \{Y_m(u) - \bar{Y}(u)\}
$$
$$
\times \langle Y_r - \bar{Y}, \hat{\psi}_j \rangle \langle Y_{m+l-s} - \bar{Y}, Y_{r+l'-s'} - \bar{Y} \rangle
$$
$$
= \sum_{m=1}^{n-p-\delta+1} \{Y_m(u) - \bar{Y}(u)\}\gamma_{mj}\hat{\theta}_j = \hat{\theta}_j \tilde{\psi}_j(u).
$$

Therefore, $\hat{\psi}_j$ is an eigenfunction of $\hat{\mathcal{K}}$, with corresponding eigenvalue $\hat{\theta}_j$.   □

Theorem 1 *Suppose the following conditions are satisfied:*

C1. $\{Y_t(\cdot)\}$ *is strictly stationary and ψ-mixing with the mixing coefficient defined as*

$$
\psi(l) = \sup_{A \in \mathcal{F}_{-\infty}^0, B \in \mathcal{F}_l^\infty, P(A)P(B)>0} |1 - P(B|A)/P(B)|,
$$

where $\mathcal{F}_i^j$ is the σ-algebra generated by $Y_i(\cdot), \ldots, Y_j(\cdot)$ for any $j \geq i$. In addition, it holds that $\sum_{l=1}^\infty l\psi^{1/2}(l) < \infty$.
C2. $\mathbb{E}\{\int_I Y_t(u)^2\mathrm{d}u\}^2 < \infty$.
C3. $\theta_1 > \cdots > \theta_d > 0 = \theta_{d+1} = \cdots$, *i.e., all the nonzero eigenvalues of $\mathcal{K}$ are different.*
C4. $\mathrm{Cov}\{X_s(u), \epsilon_t(v)\} = 0$ *for all s, t and $u, v \in I$.*

Then it holds that

(i) $\left\| \hat{\mathcal{K}} - \mathcal{K} \right\|_{\mathcal{S}} = O_p(n^{-1/2}).$

(ii) *For* $j = 1, \ldots, d, |\hat{\theta}_j - \theta_j| = O_p(n^{-1/2})$ *and*

$$\left(\int_I \{\hat{\psi}_j(u) - \psi(u)\}^2 du \right)^{1/2} = O_p(n^{-1/2}).$$

(iii) *For* $j \geq d + 1, \hat{\theta}_j = O_p(n^{-1}).$

(iv) *Let* $\{\psi_j : j \geq d + 1\}$ *be a complete orthonormal basis of* $\mathcal{M}^\perp$, *and put*

$$f_j(\cdot) = \sum_{i=d+1}^{\infty} \langle \psi_i, \hat{\psi}_j \rangle \psi_i(\cdot).$$

Then for any $j \geq d + 1$,

$$\left(\int_I \left\{ \sum_{i=1}^{d} \langle \psi_i, \hat{\psi}_j \rangle \psi_i(u) \right\}^2 du \right)^{1/2} = \left(\int_I \left\{ \hat{\psi}_j(u) - f_j(u) \right\}^2 du \right)^{1/2} = O_p(n^{-1/2}).$$

Proof Initially, we want to show that Theorem 1(i) of Bathia et al. (2010) holds also for the aggregate case. Let $\mathcal{S}$ denote the space of operators with a finite Hilbert–Schmidt norm. Since p and δ are fixed and finite, we set $n \equiv n - p - \delta + 1$. Let now $Z_{tk} = (Y_t - \mu) \otimes (Y_{t+k} - \mu) \in \mathcal{S}$ and consider the kernel $\rho : \mathcal{S} \times \mathcal{S} \to \mathcal{S}$ given by $\rho(A, B) = AB^*$, with $A, B \in \mathcal{S}$. We have that

$$\hat{M}_k \hat{M}_k^* = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \rho(Z_{ik}, Z_{ik}) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} Z_{ik} Z_{ik}^*.$$

Therefore,

$$\hat{\mathcal{M}}_k \hat{\mathcal{M}}_k^* = \left(\sum_{s,l} \omega_{t-s} \omega_{t+k-l} \hat{M}_{l-s} \right) \left(\sum_{s',l'} \omega_{t-s'} \omega_{t+k-l'} \hat{M}_{l'-s'} \right)^*$$

$$= \sum_{s,l} \sum_{s',l'} \omega_{t-s} \omega_{t+k-l} \omega_{t-s'} \omega_{t+k-l'} \hat{M}_{l-s} \hat{M}_{l'-s'}^*,$$

and similarly

$$\mathcal{M}_k \mathcal{M}_k^* = \sum_{s,l} \sum_{s',l'} \omega_{t-s} \omega_{t+k-l} \omega_{t-s'} \omega_{t+k-l'} M_{l-s} M_{l'-s'}^*.$$

Hence,

$$\left\|\hat{\mathcal{M}}_k\hat{\mathcal{M}}_k^* - \mathcal{M}_k\mathcal{M}_k^*\right\|_{\mathcal{S}} \leq \sum_{s,l}\sum_{s',l'}\omega_{t-s}\omega_{t+k-l}\omega_{t-s'}\omega_{t+k-l'}\left\|\hat{M}_{l-s}\hat{M}_{l'-s'}^* - M_{l-s}M_{l'-s'}^*\right\|_{\mathcal{S}}.$$

On the other hand, we note that

$$\hat{M}_{l-s}\hat{M}_{l'-s'}^* = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\rho(Z_{i,l-s}, Z_{i,l'-s'})$$

is a $\mathcal{S}-$valued von Mises functional, just as $\hat{M}_k\hat{M}_k^*$, which enables us to use Lemma 3 of Bathia et al. (2010) to get

$$\mathbb{E}\left\|\hat{M}_{l-s}\hat{M}_{l'-s'}^* - M_{l-s}M_{l'-s'}^*\right\|_{\mathcal{S}}^2 = O(n^{-1}).$$

Moreover, for some distinct indexes a, b, s and l, from Schwarz inequality we obtain

$$\mathbb{E}\left(\left\|\hat{M}_a\hat{M}_b^* - M_aM_b^*\right\|_{\mathcal{S}}\left\|\hat{M}_s\hat{M}_l^* - M_sM_l^*\right\|_{\mathcal{S}}\right)$$
$$\leq \left\{\mathbb{E}\left\|\hat{M}_a\hat{M}_b^* - M_aM_b^*\right\|_{\mathcal{S}}^2\mathbb{E}\left\|\hat{M}_s\hat{M}_l^* - M_sM_l^*\right\|_{\mathcal{S}}^2\right\}^{1/2} = O(n^{-1}).$$

Then, since the ω's, δ, p and k are fixed, we have that

$$\mathbb{E}\left\|\hat{\mathcal{M}}_k\hat{\mathcal{M}}_k^* - \mathcal{M}_k\mathcal{M}_k^*\right\|_{\mathcal{S}}^2$$
$$\leq \mathbb{E}\left(\sum_{s,l}\sum_{s',l'}\omega_{t-s}\omega_{t+k-l}\omega_{t-s'}\omega_{t+k-l'}\left\|\hat{M}_{l-s}\hat{M}_{l'-s'}^* - M_{l-s}M_{l'-s'}^*\right\|_{\mathcal{S}}\right)^2 = O(n^{-1}).$$

Thus, there exist n_0, n_1 such that $n\mathbb{E}\left\|\hat{\mathcal{M}}_k\hat{\mathcal{M}}_k^* - \mathcal{M}_k\mathcal{M}_k^*\right\|_{\mathcal{S}}^2 \leq n_1, \forall n > n_0$. Then, for some $k \geq 1$ and Chebyshev inequality, it follows that for all $\epsilon > 0$,

$$P\left(n^{1/2}\left\|\hat{\mathcal{M}}_k\hat{\mathcal{M}}_k^* - \mathcal{M}_k\mathcal{M}_k^*\right\|_{\mathcal{S}} > n_1^k\right) \leq \frac{\mathbb{E}\left\{n\left\|\hat{\mathcal{M}}_k\hat{\mathcal{M}}_k^* - \mathcal{M}_k\mathcal{M}_k^*\right\|_{\mathcal{S}}^2\right\}}{n_1^{2k}} < \frac{1}{n_1^{2k-1}} < \epsilon,$$

by choosing k sufficiently large. This means that $\left\|\hat{\mathcal{M}}_k\hat{\mathcal{M}}_k^* - \mathcal{M}_k\mathcal{M}_k^*\right\|_{\mathcal{S}} = O_p(n^{-1/2})$, and as consequence

$$\left\|\hat{\mathcal{K}} - \mathcal{K}\right\|_{\mathcal{S}} \leq \sum_{k=\delta}^{p-\delta+1}\left\|\hat{\mathcal{M}}_k\hat{\mathcal{M}}_k^* - \mathcal{M}_k\mathcal{M}_k^*\right\|_{\mathcal{S}} = O_p(n^{-1/2}).$$

Applying the same arguments of Theorem 1(ii) of Bathia et al. (2010), we observe that it also holds for the aggregate case

$$|\hat{\theta}_j - \theta_j| = O_p(n^{-1/2}) \quad \text{and} \quad \left\| \hat{\psi}_j - \psi_j \right\| = O_p(n^{-1/2}), \quad j = 1, \ldots, d.$$

Additionally, we have that

$$\mathbb{E} \left\| \hat{M}_{l-s} \hat{M}^*_{l'-s'} - \hat{M}_{l-s} M^*_{l'-s'} \right\|^2_{\mathcal{S}} = O(n^{-2}),$$

which gives

$$\mathbb{E} \left\| \hat{\mathcal{M}}_k \hat{\mathcal{M}}^*_k - \mathcal{M}_k \mathcal{M}^*_k \right\|^2_{\mathcal{S}} = O(n^{-2}).$$

Therefore, using the same arguments of Bathia et al. (2010) for $\hat{\mathcal{M}}_k$ and $\mathcal{M}_k$ instead of $\hat{M}_k$ and M_k, we can conclude that Theorem 1(iii) and 1(iv) also hold for the aggregate case. □

6 Simulations

In this section we present a simulation study to evaluate the performance of the method described in Sect. 3, to estimate the dimension of a functional generating subspace via wavelets. We used the same settings employed by Bathia et al. (2010), considering as true function and noise, respectively

$$X_t(x) = \sum_{l=1}^{d} \xi_{tl} \varphi_l(x) \quad \text{and} \quad \epsilon_t(x) = \sum_{i=1}^{10} \frac{Z_{ti}}{2^{i-1}} \zeta_i(x), \quad x \in [0, 1],$$

where $\{\xi_{tl}, t \geq 1\}$ are mutually independent AR(1) processes with coefficients $(-1)^l(0.9 - 0.5l/d)$, the coefficients Z_{ti} are independent random variables, independent from the ξ_{tl}'s, following standard normal distribution and the eigenfunctions used are

$$\varphi_l(x) = \sqrt{2} \cos(\pi l x) \quad \text{and} \quad \zeta_i(x) = \sqrt{2} \sin(\pi i x).$$

The sample sizes considered are $n \in \{100, 300, 600\}$ and the dimensions are $d \in \{2, 4, 6\}$, while the maximum lag used is $p = 5$. The wavelet basis is the Daubechies with four null moments. We perform a discretization of the problem, evaluating $X_t(x)$ and $\epsilon_t(x)$ on a grid of 256 equally spaced points $x \in [0, 1]$, and then, we obtain a vector of the observed curve $Y_t(x)$ evaluated at these points. For each $t = 1, \ldots, n$ the decomposition for the 2^8 points is performed using a minimum resolution level of 5 and maximum resolution level 7. The simulation described is based on 1000 replications, and we consider the four bootstrap procedures described in Sect. 3. The threshold

Table 1 Proportion (%) that each value $\hat{d}$ is selected as dimension of the process when the true dimension is d for each sample size n

d	2			4			6		
n	$\hat{d}$								
	1	2	3	3	4	5	5	6	7
Ordinary bootstrap									
100	46.1	51.8	2	41.7	27.9	1.7	23.6	8.9	0.4
300	0.3	93.5	5.9	2.3	92.9	4.6	9.2	84.3	6
600	0	93.3	6.5	0	93.2	6.5	0	95	4.9
Applying thresholding before bootstrap									
100	45.4	51.6	3	41.6	28	1.5	24.3	11	0.4
300	1.2	92.8	5.5	2.6	91.7	5.5	8	86.7	4.6
600	0	94.2	5.8	0	94.2	5.4	0	95.2	4.7
Bootstrap with the residual $\hat{\epsilon}_t^{\text{thr}}$									
100	43.1	53.3	3.6	42.2	27.7	1.9	24.9	8.7	0.1
300	0.7	93.2	5.8	2.4	91	6.4	7.2	86.7	5.6
600	0	94.6	5.2	0	94.3	5.6	0	94.9	5
Wavestrap									
100	45.3	51.8	2.8	43.4	26.9	1.4	24.9	10.5	0.3
300	0.9	93.2	5.6	2.3	91.4	6.1	7.5	88.8	3.4
600	0	93.6	6.1	0	93.5	6.2	0	93.8	6.1

value used in the second and third bootstrap procedures presented in this section is chosen according to the universal threshold of Donoho and Johnstone (1994).

Table 1 contains the proportions of the dimension estimates for different true dimensions and sample sizes when we apply the bootstrap tests with 5% significance level. All bootstrap procedures tend to select the true value d as the sample sizes increases. When the sample size is 300 or larger, the four methods perform well, but have poor performance when $n = 100$, especially when the true dimension has a large value. The results of the four bootstrap procedures are very close, but slight advantages can be noted for the procedures where thresholding is applied or when the residual $\hat{\epsilon}_t^{\text{thr}}$ is used. For instance, when $d = 2$ these two procedures have lower rates of dimension overestimation for sample sizes 300 and 600. Figure 1 displays boxplots of the bootstrap p values corresponding to the first method. Results for the other three methods were similar, and corresponding figures are included in a supplementary material. We can observe that for samples of size $n = 100$, the tests select lower dimensions with considerable frequency (not rejecting that the d-th largest eigenvalue is zero), but for larger sample sizes the tests present a better performance, selecting the correct dimension more often (rejecting that the d-th largest eigenvalue is null and not rejecting that the $(d + 1)$-th largest eigenvalue is zero).
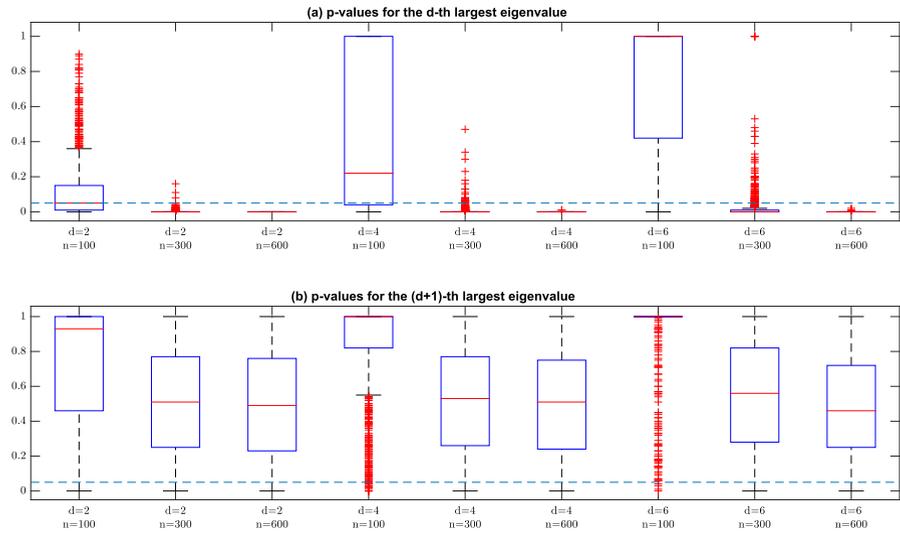
Fig. 1 Boxplots of the p values of the tests for the d-th and $(d+1)$-th largest eigenvalues, for each sample size and true value of d using the ordinary bootstrap, which corresponds to the bootstrap proposed by Bathia et al. (2010) applied in the wavelet coefficients. The segmented line represents the significance level used (5%)

6.1 Numerical evaluation of aggregate data

In this subsection we report a simulation study performed to evaluate the method of data aggregation considering the same functions used in the previous numerical evaluation. A comparison with the eigenvalues obtained without data aggregation is also presented. For both methods we analyze the estimated eigenvalues and the testing performances in correctly selecting the functional dimension.

We begin with the analysis of the estimated eigenvalues obtained using the covariance operator. The functions $\{\varphi_i, i = 1, \ldots, d\}$ of the previous simulation study form an orthonormal system in $L^2([0,1])$, and the AR(1) processes $\{\xi_{tl}, t \geq 1\}$ have coefficients $\vartheta_l = (-1)^l(0.9 - 0.5l/d)$ and are independent for different l's. The white noise in the AR processes is random variables with distribution $N(0, \sigma_w^2)$, where $\sigma_w^2 = 1.5$ was used during simulations. Therefore, the covariance kernel can be written as

$$M_k(u, v) = \mathrm{Cov}\{X_t(u), X_{t+k}(v)\} = \mathrm{Cov}\left\{\sum_{j=1}^{d} \xi_{t,j}\varphi_j(u), \sum_{i=1}^{d} \xi_{t+k,i}\varphi_i(v)\right\}$$

$$= \sum_{j=1}^{d} \varphi_j(u)\varphi_j(v)\sigma_{jj}^{(k)},$$

where $\sigma_{jj}^{(k)} = \sigma_w^2 \vartheta_j^k/(1 - \vartheta_j^2)$, $j = 1, \ldots, d$, are eigenvalues of M_k. Knowing that $M_k = \sum_{i,j=1}^{d} \sigma_{ij}^{(k)}\varphi_i\varphi_j$ with $\Sigma_k = \{\sigma_{ij}^{(k)}\} = \mathrm{diag}\{\sigma_{11}^{(k)}, \ldots, \sigma_{dd}^{(k)}\}$, let us denote

$N_k = \sum_{i,j=1}^{d} w_{ij}^{(k)} \varphi_i \varphi_j$ and $W_k = \{w_{ij}^{(k)}\} = \Sigma_k \Sigma_k^\top = \text{diag}\{(\sigma_{11}^{(k)})^2, \ldots, (\sigma_{dd}^{(k)})^2\}$
$= \text{diag}\{w_{11}^{(k)}, \ldots, w_{dd}^{(k)}\}$. Thus,

$$\int_0^1 K(u,v)\varphi_j(v)\mathrm{d}v = \int_0^1 \left(\sum_{k=1}^{p} N_k(u,v)\right)\varphi_j(v)\mathrm{d}v = \left(\sum_{k=1}^{p} (\sigma_{jj}^{(k)})^2\right)\varphi_j(u),$$

giving the eigenvalues of the operator K for the non-aggregate case. Applying data aggregation, the term in Eq. (9) is $\alpha_{ij}^{(k)} = 0$ if $i \neq j$. Then $\mathcal{N}_k(u,v) = \sum_{i=1}^{d}(\alpha_{ii}^{(k)})^2 \varphi_i(u)\varphi_i(v)$, and

$$\int_0^1 \mathcal{K}(u,v)\varphi_j(v)\mathrm{d}v = \left(\sum_{k=\delta}^{p-\delta+1} (\alpha_{jj}^{(k)})^2\right)\varphi_j(u),$$

which gives the eigenvalues of $\mathcal{K}$. With these results, we can compare the eigenvalues obtained with and without data aggregation with their respective true eigenvalues of the random function X_t.

The sample sizes considered in this simulation are $n \in \{100, 300, 600\}$, and the dimensions are $d \in \{2, 4, 6\}$, while the maximum lag used is $p = 5$. The numerical study is based on 1000 replications, and Table 2 presents the average of the largest eigenvalues obtained for each n and d, as well as the true eigenvalues for each dimension. For each replication we used $\delta = 3$, with weights $\omega_2 = 0.1$, $\omega_1 = 0.3$ and $\omega_0 = 0.5$. For means of comparison, Table 2 also presents the analogous results applying the method of Bathia et al. (2010) directly, without data aggregation, which corresponds to using a single weight $\omega_0 = 1$. Results of Table 2 are summarized in Figure 2, which shows the estimate's averages of nonzero eigenvalues obtained for each sample size. To make it easier to discriminate between different curves, we considered the logarithm of these averages in Figure 2. Overall, we note that estimates tend to get closer to their corresponding true values as the sample size increases.

We now present a simulation study to compare the method's performance under data aggregation. Different weights are used as follows: case 1, with weights $\omega_2 = 0.1$, $\omega_1 = 0.1$ and $\omega_0 = 0.8$; case 2, with weights $\omega_2 = 1/3$, $\omega_1 = 1/3$ and $\omega_0 = 1/3$; case 3, with weights $\omega_2 = 0.1$, $\omega_1 = 0.3$ and $\omega_0 = 0.5$; and case 4, with only $\omega_0 = 1$. Table 3 summarizes the results. The proportions of dimension selection for $d = 2, 4$ and 6 for each of the four cases are displayed. As expected, case 4 (no aggregation) presents the best overall performances notwithstanding the method. Moreover, case 1 (which is the aggregate closest to case 4) has the second best overall performances. Figure 3 presents the boxplots for the fourth and fifth largest eigenvalues when $d = 4$. As n increases, the proposed procedures correctly lead to a dimension equal to 4 with high empirical frequency.

Two important observations are due. The estimates for the eigenvalues are close to their true theoretical counterparts for the aggregate case. This means that the apparent poorer performance for aggregate data versus non-aggregate data can be interpreted as a good performance for a harder problem. The second issue is that we should only

Table 2 Simulation study for aggregate and non-aggregate data

d	n	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$	$\hat{\theta}_5$	$\hat{\theta}_6$	$\hat{\theta}_7$	$\hat{\theta}_8$	$\hat{\theta}_9$	$\hat{\theta}_{10}$
Using method with data aggregation											
2	100	0.0591	0.0093	0.0013	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	300	0.033	0.0055	0.0005	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	600	0.026	0.005	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	True	0.0186	0.0046	0	0	0	0	0	0	0	0
4	100	0.5383	0.0672	0.022	0.0056	0.0007	0.0000	0.0000	0.0000	0.0000	0.0000
	300	0.4354	0.0434	0.0156	0.0024	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000
	600	0.4229	0.0374	0.0161	0.0018	0.0002	0.0000	0.0000	0.0000	0.0000	0.0000
	True	0.3971	0.0256	0.0186	0.0010	0	0	0	0	0	0
6	100	1.5702	0.2265	0.0767	0.0306	0.0114	0.0037	0.0005	0.0000	0.0000	0.0000
	300	1.317	0.1789	0.0609	0.0217	0.0063	0.0017	0.0002	0.0000	0.0000	0.0000
	600	1.2134	0.1549	0.0544	0.02	0.0053	0.0012	0.0002	0.0000	0.0000	0.0000
	True	1.1759	0.1433	0.0497	0.0186	0.0046	0.0007	0	0	0	0
Using method without data aggregation											
2	100	6.6003	0.9931	0.1515	0.0143	0.0032	0.0011	0.0003	0.0001	0.0000	0.0000
	300	5.3412	0.7516	0.0553	0.0046	0.001	0.0004	0.0001	0.0000	0.0000	0.0000
	600	5.1887	0.6861	0.028	0.0024	0.0006	0.0002	0.0000	0.0000	0.0000	0.0000
	True	4.8693	0.6073	0	0	0	0	0	0	0	0
4	100	26.122	5.8327	2.1884	0.943	0.2496	0.0034	0.0008	0.0002	0.0001	0.0000
	300	21.212	5.364	1.8662	0.7305	0.0882	0.0011	0.0003	0.0001	0.0000	0.0000
	600	20.921	5.0998	1.7707	0.6855	0.0444	0.0006	0.0001	0.0000	0.0000	0.0000
	True	19.5567	4.8693	1.6290	0.6073	0	0	0	0	0	0
6	100	52.188	14.454	5.9776	2.9913	1.6645	0.9401	0.334	0.001	0.0003	0.0000
	300	40.075	12.757	5.2998	2.543	1.3797	0.7322	0.1235	0.0003	0.0001	0.0000
	600	37.857	12.41	5.177	2.4472	1.2812	0.6805	0.0626	0.0002	0.0000	0.0000
	True	35.2583	11.7016	4.8693	2.3012	1.1669	0.6073	0	0	0	0

Average of the ten largest eigenvalues estimated for different sample sizes n and true dimensions d and true values of the eigenvalues for each dimension, presenting results with and without data aggregation
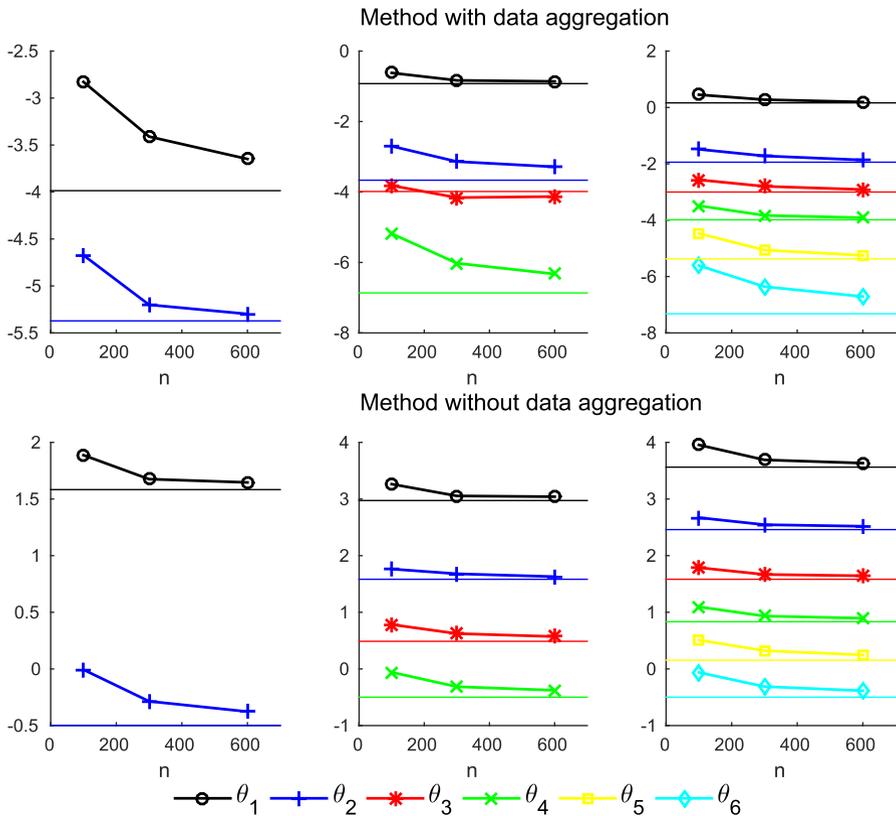
Fig. 2 Simulation study for aggregate and non-aggregate data. Logarithm of averages of the estimated nonzero eigenvalues (vertical axis) for different sample sizes n (horizontal axis) for the methods with and without data aggregation. The number of curves in each plot corresponds to the true dimensions, and horizontal solid lines correspond to the true values of the estimated eigenvalues of same color

proceed with the aggregate data when small sample sizes hinder good estimates for some curves, while the non-aggregate data are employed. Moreover, since the weights may be chosen by the analyst, we can do it in such a way as to minimize the aggregation burden.

7 Application

In this section we present an application of the proposed wavelet methods in dimension estimation of functional time series. We analyze the data set of Australian fertility rates since the year 1921 until 2010. The data are in the Australian Bureau of Statistics available at http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/3105.0.65.0012014?OpenDocument and consist of the numbers of births per 1000 women during each year according to the age-group of the mother (15–19, 20–24, 25–29, 30–34, 35–39, 40–44, 45–49). This data set was analyzed by Hyndman and

Table 3 Proportion (%) that each value $\hat{d}$ is selected as dimension of the process when the true dimension is d for each sample size n when we use aggregation with different weights

d	2			4			6		
n	$\hat{d}$								
	1	2	3	3	4	5	5	6	7
Weights $\omega_2 = 0.1, \omega_1 = 0.1, \omega_0 = 0.8$									
100	77.5	19.0	3.5	18.5	12.5	4.0	14.5	6.5	1.5
300	41.5	55.0	3.5	31.0	47.0	6.5	38.0	38.5	5.0
600	17.5	76.0	6.0	14.0	70.0	14.0	14.0	73.5	11.5
Weights $\omega_2 = 1/3, \omega_1 = 1/3, \omega_0 = 1/3$									
100	94.5	3.5	2.0	6.0	0.5	0	0	0	0
300	57.5	37.0	5.5	37.0	16.5	2.0	22.5	10.0	1.5
600	9.0	80.5	10.5	22.0	60.0	11.0	31.0	51.0	9.0
Weights $\omega_2 = 0.1, \omega_1 = 0.3, \omega_0 = 0.5$									
100	98.5	1.5	0	0.5	0	0	0	0	0
300	83.0	15.5	1.5	28.0	2.0	1.0	3.0	1.0	0
600	56.0	39.0	5.0	59.0	13.0	1.0	32.5	8.0	0.5
Weight $\omega_0 = 1$ (method without aggregation)									
100	38.0	57.0	5.0	33.0	39.5	1.5	28.0	24.0	2.5
300	1.5	91.0	7.5	1.5	94.0	4.5	2.0	95.5	2.5
600	0	95.0	5.0	0	92.5	7.5	0	98.0	2.0

Ullah (2007) in the context of functional time series for the years of 1921 until 2000, where the authors considered the center of each age-group as the age for which the corresponding fertility rate was observed and also assigned the value 0.005 for the ages 13 and 52 for all years. Following this procedure, we fitted a curve for the logarithm of the fertility rate of each year using smoothing splines and considered these functions as our observed curves. Figure 4 displays the observed curves for some years. It is noteworthy from this figure that the lower log rate for women under 25 is observed in the year 2010, which might be associated with the tendency of women in developed countries to bear less children and later than in previous years. We also performed a stationarity test (Horváth et al. 2014) and obtained a p value of 2.4%. Although we have an indication that the process is not stationary, we shall still employ the methods discussed in paper to estimate the process' dimension and compare the conclusion with the results obtained by Hyndman and Ullah (2007).

To estimate the dimension of the process that generates these random functions using the wavelet-based method, we evaluated the observed curves in an equally spaced grid of 64 points between the ages of 13 and 52, which gives $J = 84$ wavelet coefficients for each curve. Next, we performed the procedures of Sect. 3 using the Daubechies wavelet basis with four null moments with minimum and maximum resolution levels of 3 and 5, respectively. The value of maximum lag used was the same as in the simulation experiments, i.e., $p = 5$. The application of hard thresholding on the wavelet coefficients for the Australian fertility curves results in the shrinkage of

(a) p-values for the 4-th largest eigenvalue

(b) p-values for the 5-th largest eigenvalue
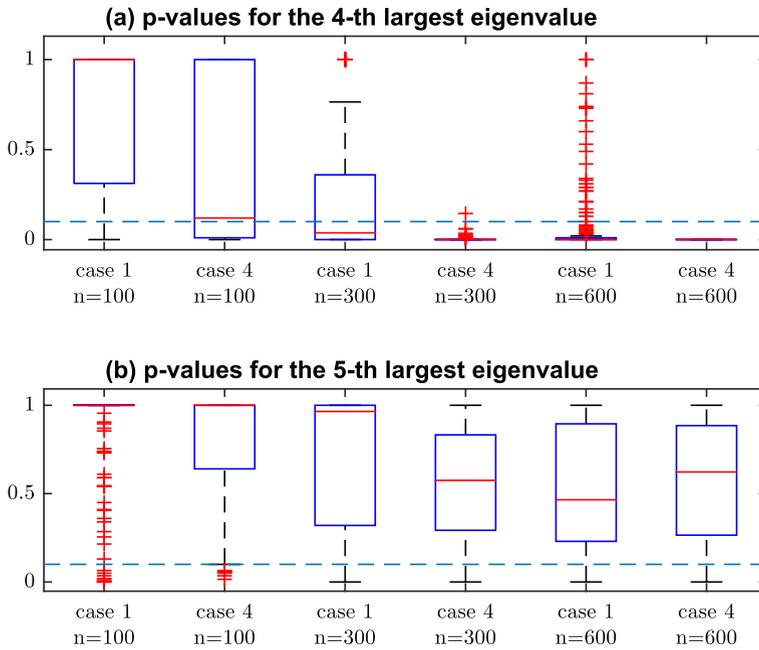
Fig. 3 Boxplots of the p values of the tests for the fourth and fifth largest eigenvalues of cases 1 ($\omega_2 = 0.1$, $\omega_1 = 0.1$, $\omega_0 = 0.8$) and 4 ($\omega_0 = 1$) of weights used in the aggregated method, for each sample size and true dimension $d = 4$ fixed. The segmented line represents the significance level used (5%)
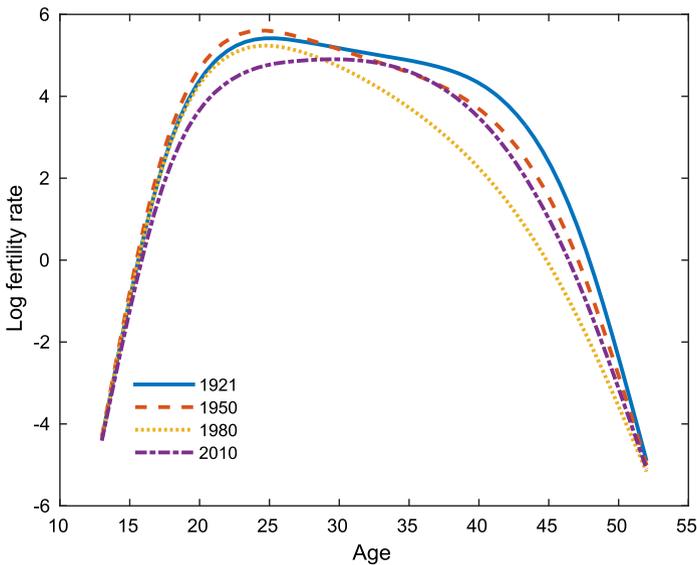
Fig. 4 Australian fertility rates on 1921–2010 (Hyndman and Ullah 2007). The curves are logarithms of the Australian fertility rate: 1921; 1950; 1980; 2010

Table 4 Australian fertility rates on 1921–2010 (Hyndman and Ullah 2007)

Method	Eigenvalues				
	λ_1	λ_2	λ_3	λ_4	λ_5
Wavelet-based	99.8677	5.1413	0.0603	0.0021	< 0.0001
Aggregate	99.9678	2.5363	0.0522	0.0003	< 0.0001
Non-aggregate	99.9531	3.0614	0.0573	0.0006	< 0.0001

Five largest eigenvalues divided by the norm of all eigenvalues when applied the wavelet-based method and the methods with and without data aggregation for the fertility data

56% of the estimated coefficients to zero. This means that a sparse representation of the functions is attained, and that prediction variance and bias will diminish as well. Table 4 contains the five largest eigenvalues computed from the matrix D obtained after the wavelet decomposition of the log rate curves. Table 4 also contains eigenvalues obtained with the dimension estimation methods with and without data aggregation. For ease of comparison, all values were divided by the norm of all eigenvalues obtained with the same method. The values in Table 4 are close for the three methods and indicate that the time series might be generated from a four- or five-dimensional process. Performing the bootstrap test for dimensionality in the wavelet-based method (using the residual $\hat{\epsilon}_t^{\text{thr}}$) with 301 replications and significance level of 5%, the result also indicates that the process has dimension 5. (We reject that $\lambda_5 = 0$ and fail to reject that $\lambda_6 = 0$.)
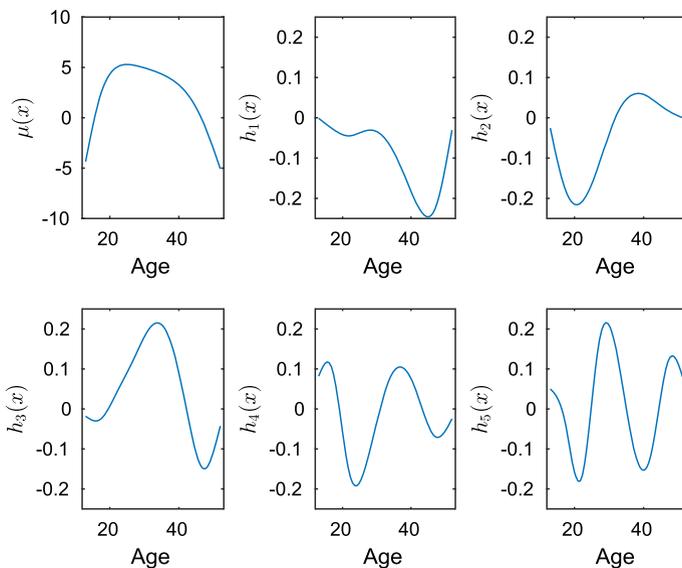
Fig. 5 Australian fertility rates on 1921–2010 (Hyndman and Ullah 2007). Mean function and selected eigenfunctions after estimating the dimension of the process ($\hat{d} = 5$)

Figure 5 presents the mean function and the eigenfunctions corresponding to the five largest eigenvalues. These eigenfunctions are similar (apart from sign) to the eigenfunctions presented by Hyndman and Ullah (2007) for the years of 1921 until 2000. Additionally, similar functions are obtained if we apply the method of Bathia et al. (2010). The first eigenfunction has a marked impact on ages greater than 30 years; this impact is negative whenever the latent coefficient is positive, and vice versa. For ages below 30, h_1 seems to have little influence, fluctuating around some value close to zero. The second eigenfunction seems to oscillate around zero at the age of 30 years; for instance, for positive latent coefficients, it has a negative influence on ages below 30 and a positive influence on ages greater than 30. With respect to the third eigenfunction, it has a direct impact on ages around 30, with a positive impact for positive latent coefficients and vice versa. The remaining eigenfunctions can be analyzed similarly.

In their analysis, Hyndman and Ullah (2007) employ a PCA with three basis functions by applying procedures similar to the ones presented by Ramsay and Silverman (2005). The former authors mention that only 0.8% of the variation is left unexplained, and from our results, this amount could still account for part (maybe not essential for their practical purposes) of the process that generates the curves. These extra information may be quite nonlinear in nature.

8 Discussion

We study in this manuscript the problem of estimating the dimension of finite-dimensional random functions, which can be used for modeling time series of curves. This problem has been discussed by Hall and Vial (2006) and Bathia et al. (2010). The latter have used the underlying temporal stochastic structure to propose a statistical procedure which has nice asymptotic properties. We use wavelet representation in this setup, and have attained the same asymptotic results. Moreover, besides the original bootstrap procedures, wavelets allow us to employ three additional bootstrap schemes. The wavelet method is loss free, can be inverted and has some computational advantages as well. We also show that such method may be employed for aggregate data and that the resulting statistical methodology has similar theoretical properties. The proposed method is illustrated in simulation studies and on a real data set.

References

Abadir, K. M., Caggiano, G., Talmain, G. (2013). Nelson–Plosser revisited: The ACF approach. *Journal of Econometrics*, *175*(1), 22–34.

Abraham, B. (1982). Temporal aggregation and time series. *International Statistical Review/Revue Internationale de Statistique*, *50*(3), 285–291.

Amato, U., Antoniadis, A., De Feis, I., Goude, Y. (2017). Estimation and group variable selection for additive partial linear models with wavelets and splines. *South African Statistical Journal*, *51*(2), 235–272.

Amighini, A., Bongiorno, E. G., Goia, A. (2014). A clustering method for economic aggregates by using concentration curves. *Contributions in infinite-dimensional statistics and related topics*, pp. 25–30. Bologna: Esculapio.

Aneiros, G., Vieu, P. (2016). Comments on: Probability enhanced effective dimension reduction for classifying sparse functional data. *TEST*, *25*(1), 27–32.

Aue, A., Horváth, L., Pellatt, D. F. (2017). Functional generalized autoregressive conditional heteroskedasticity. *Journal of Time Series Analysis*, *38*(1), 3–21.

Bathia, N., Yao, Q., Ziegelmann, F. (2010). Identifying the finite dimensionality of curve time series. *The Annals of Statistics*, *38*(6), 3352–3386.

Belloni, A., Chernozhukov, V., Fernández-Val, I., Hansen, C. (2017). Program evaluation and causal inference with high-dimensional data. *Econometrica*, *85*(1), 233–298.

Berry, S. T., Haile, P. A. (2014). Identification in differentiated products markets using market level data. *Econometrica*, *82*(5), 1749–1797.

Bosq, D. (2000). *Linear processes in function spaces: Theory and applications*. New York: Springer.

Breunig, C., Johannes, J. (2016). Adaptive estimation of functionals in nonparametric instrumental regression. *Econometric Theory*, *32*(3), 612–654.

Canale, A., Ruggiero, M. (2016). Bayesian nonparametric forecasting of monotonic functional time series. *Electronic Journal of Statistics*, *10*(2), 3265–3286.

Chacón, J. E., Rodríguez-Casal, A. (2005). On the l1-consistency of wavelet density estimates. *Canadian Journal of Statistics*, *33*(4), 489–496.

Cholaquidis, A., Fraiman, R., Kalemkerian, J., Llop, P. (2014). An optimal aggregation type classifier. In *Contributions in infinite-dimensional statistics and related topics* (pp 85–90). Bologna: Esculapio.

Comte, F., Mabon, G., Samson, A. (2017). Spline regression for hazard rate estimation when data are censored and measured with error. *Statistica Neerlandica*, *71*(2), 115–140.

Devijver, E. (2017). Model-based regression clustering for high-dimensional data: Application to functional data. *Advances in Data Analysis and Classification*, *11*(2), 243–279.

Dias, R., Garcia, N. L., Ludwig, G., Saraiva, M. A. (2015). Aggregated functional data model for near-infrared spectroscopy calibration and prediction. *Journal of Applied Statistics*, *42*(1), 127–143.

Dias, R., Garcia, N. L., Schmidt, A. M. (2013). A hierarchical model for aggregated functional data. *Technometrics*, *55*(3), 321–334.

Donoho, D. L., Johnstone, J. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, *81*(3), 425–455.

Fan, Y., James, G. M., Radchenko, P. (2015). Functional additive regression. *The Annals of Statistics*, *43*(5), 2296–2325.

Hall, P., Vial, C. (2006). Assessing the finite dimensionality of functional data. *Journal of the Royal Statistical Society, Series B*, *68*(4), 689–705.

Härdle, W., Kerkyacharian, G., Picard, D., Tsybakov, A. (1998). *Wavelets, approximation, and statistical applications*. New York: Springer.

Hooker, G., Roberts, S. (2016). Maximal autocorrelation functions in functional data analysis. *Statistics and Computing*, *26*(5), 945–950.

Horta, E., Ziegelmann, F. (2016). Identifying the spectral representation of Hilbertian time series. *Statistics & Probability Letters*, *118*, 45–49.

Horta, E., Ziegelmann, F. (2018). Dynamics of financial returns densities: A functional approach applied to the Bovespa intraday index. *International Journal of Forecasting*, *34*(1), 75–88.

Horváth, L., Kokoszka, P., Rice, G. (2014). Testing stationarity of functional time series. *Journal of Econometrics*, *179*(1), 66–82.

Hyndman, R. J., Ullah, M. S. (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics & Data Analysis*, *51*(10), 4942–4956.

Imaizumi, M., Kato, K. (2018). PCA-based estimation for functional linear regression with functional responses. *Journal of Multivariate Analysis*, *163*, 15–36.

Ivanescu, A. E. (2017). Adaptive inference for the bivariate mean function in functional data. *Advances in Data Science and Adaptive Analysis*, *9*(3), 1750005.

Johnstone, I. M., Lu, A. Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. *Journal of the American Statistical Association*, *104*(486), 682–693.

Lakraj, G. P., Ruymgaart, F. (2017). Some asymptotic theory for Silverman's smoothed functional principal components in an abstract Hilbert space. *Journal of Multivariate Analysis*, *155*, 122–132.

Li, B., Song, J. (2017). Nonlinear sufficient dimension reduction for functional data. *The Annals of Statistics*, *45*(3), 1059–1095.

Li, G., Shen, H., Huang, J. Z. (2016). Supervised sparse and functional principal component analysis. *Journal of Computational and Graphical Statistics*, *25*(3), 859–878.

Lorenz, D. A., Resmerita, E. (2017). Flexible sparse regularization. *Inverse Problems*, *33*(1), 014002.

Mallat, S. G. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *11*(7), 674–693.

Mallat, S. G. (1998). *A wavelet tour of signal processing*. San Diego: Academic Press.

Masry, E. (1994). Probability density estimation from dependent observations using wavelets orthonormal bases. *Statistics & Probability Letters*, *21*(3), 181–194.

Masry, E. (1997). Multivariate probability density estimation by wavelet methods: Strong consistency and rates for stationary time series. *Stochastic Processes and Their Applications*, *67*(2), 177–193.

Mirafzal, S. M. (2018). More odd graph theory from another point of view. *Discrete Mathematics*, *341*(1), 217–220.

Morettin, P. A., Pinheiro, A., Vidakovic, B. (2017). *Wavelets in functional data analysis*. Cham: Springer.

Mousavi, S. N., Sørensen, H. (2018). Functional logistic regression: A comparison of three methods. *Journal of Statistical Computation and Simulation*, *88*(2), 250–268.

Pakoš, M. (2011). Estimating intertemporal and intratemporal substitutions when both income and substitution effects are present: The role of durable goods. *Journal of Business & Economic Statistics*, *29*(3), 439–454.

Percival, D., Sardy, S., Davison, A. (2000). *Nonlinear and nonstationary signal processing, chapter Wavestrapping time series: Adaptive wavelet-based bootstrapping*, pp. 442–471. Cambridge: Cambridge University Press.

Pinheiro, A., Vidakovic, B. (1997). Estimating the square root of a density via compactly supported wavelets. *Computational Statistics & Data Analysis*, *25*(4), 399–415.

Qu, L., Song, X., Sun, L. (2018). Identification of local sparsity and variable selection for varying coefficient additive hazards models. *Computational Statistics & Data Analysis*, *125*, 119–135.

Ramsay, J. O., Silverman, B. W. (2005). *Functional data analysis* 2nd ed. New York: Springer.

Røislien, J., Winje, B. (2013). Feature extraction across individual time series observations with spikes using wavelet principal component analysis. *Statistics in Medicine*, *32*(21), 3660–3669.

Salvatore, S., Bramness, J. G., Røislien, J. (2016). Exploring functional data analysis and wavelet principal component analysis on ecstasy (MDMA) wastewater data. *BMC Medical Research Methodology*, *16*, 81.

Schillings, C., Schwab, C. (2016). Scaling limits in computational Bayesian inversion. *ESAIM: Mathematical Modelling and Numerical Analysis*, *50*(6), 1825–1856.

Shang, H. L. (2016). Mortality and life expectancy forecasting for a group of populations in developed countries: A multilevel functional data method. *The Annals of Applied Statistics*, *10*(3), 1639–1672.

Sienkiewicz, E., Song, D., Breidt, F. J., Wang, H. (2017). Sparse functional dynamical models—A big data approach. *Journal of Computational and Graphical Statistics*, *26*(2), 319–329.

Suarez, A. J., Ghosal, S. (2017). Bayesian estimation of principal components for functional data. *Bayesian Analysis*, *12*(2), 311–333.

Vidakovic, B. (1999). *Statistical modeling by wavelets*. New York: Wiley.

Voronin, S., Daubechies, I. (2017). An iteratively reweighted least squares algorithm for sparse regularization. In *Functional analysis, harmonic analysis, and image processing: A collection of papers in honor of Björn Jawerth*, volume 693 of *Contemporary Mathematics* (pp. 391–411). Providence: American Mathematical Society.

Wei, W. (2006). *Time series analysis: Univariate and multivariate methods* 2nd ed. Boston: Pearson.

Yang, J., Stahl, D., Shen, Z. (2017). An analysis of wavelet frame based scattered data reconstruction. *Applied and Computational Harmonic Analysis*, *42*(3), 480–507.

Yan, H., Paynabar, K., Shi, J. (2018). Real-time monitoring of high-dimensional functional data streams via spatio-temporal smooth sparse decomposition. *Technometrics*, *60*(2), 181–197.

Yao, F., Wu, Y., Zou, J. (2016). Probability-enhanced effective dimension reduction for classifying sparse functional data. *TEST*, *25*(1), 1–22.

Zhang, J., Blum, R. S., Kaplan, L. M., Lu, X. (2017). Functional forms of optimum spoofing attacks for vector parameter estimation in quantized sensor networks. *IEEE Transactions on Signal Processing*, *65*(3), 705–720.

Zhang, X., Wang, C., Wu, Y. (2018). Functional envelope for model-free sufficient dimension reduction. *Journal of Multivariate Analysis*, *163*, 37–50.