



HAL
open science

Should we correct the bias in Confidence Bands for Repeated Functional Data?

Emilie Devijver, Adeline Samson

► **To cite this version:**

Emilie Devijver, Adeline Samson. Should we correct the bias in Confidence Bands for Repeated Functional Data?. 2024. hal-04599968

HAL Id: hal-04599968

<https://hal.science/hal-04599968>

Preprint submitted on 4 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Emilie Devijver¹ CNRS, Univ. Grenoble Alpes, Grenoble INP, LIG, 38000 Grenoble, France
 Adeline Samson Univ. Grenoble Alpes, CNRS, Grenoble INP, LJK, 38000 Grenoble, France

Date published: 2024-06-03 Last modified: 2024-06-03

Abstract

While confidence intervals for finite quantities are well-established, constructing confidence bands for objects of infinite dimension, such as functions, poses challenges. In this paper, we explore the concept of parametric confidence bands for functional data with an orthonormal basis. Specifically, we revisit the method proposed by Sun and Loader, which yields confidence bands for the projection of the regression function in a fixed-dimensional space. This approach can introduce bias in the confidence bands when the dimension of the basis is misspecified. Leveraging this insight, we introduce a corrected, unbiased confidence band. Surprisingly, our corrected band tends to be wider than what a naive approach would suggest. To address this, we propose a model selection criterion that allows for data-driven estimation of the basis dimension, balancing the trade-off between bias and variance. The bias is then automatically corrected after dimension selection. Throughout the paper, we illustrate these strategies using an extensive simulation study. We conclude with an application to real data.

Keywords: functional data, repeated data, confidence band, Kac-Rice formulae, bias, dimension selection

1 Contents

2	1 Introduction	2
3	2 Statistical Model	4
4	2.1 Functional regression model	4
5	2.2 Approximation of the model on a finite family	6
6	2.3 Estimator	8
7	2.3.1 Estimation of the regression function	8
8	2.3.2 Statistics	9
9	2.3.3 Bias	10
10	3 Confidence Bands of \underline{f}^{L,L^*} and f^{L,L^*} for a fixed L	10
11	3.1 Confidence band for \underline{f}^{L,L^*}	10
12	3.2 Asymptotic confidence band for f^{L,L^*}	12
13	4 Confidence Band of f^{L^*} by correcting the bias	13
14	4.1 Construction of the band of f^{L^*} for a given L	14
15	4.2 Influence of L	15
16	4.3 Comparison with the confidence bands of Section 3	17

¹Corresponding author: emilie.devijver@univ-grenoble-alpes.fr

17	5 Selection of the best confidence band with a criteria taking into account the bias	18
18	6 Real data analysis	21
19	7 Conclusion	22
20	References	22
21	8 Appendix: proofs	24
22	8.1 Proof of Proposition 2.3	24
23	8.2 Proof of Theorem 3.2	24
24	8.3 Proof of Proposition 4.1	24

25 **1 Introduction**

26 Functional data analysis is widely used for handling complex data with smooth shapes, finding
27 applications in diverse fields such as neuroscience (e.g., EEG data, Zhang (2020)), psychology (e.g.,
28 mouse-tracking data, Quinton et al. (2017)), and sensor data from daily-life activities (Jacques and
29 Samardžić (2022)).

30 We consider multiple independent observations of the same function, yielding noisy functional data.
31 To analyze such data, a common approach, typically in the parametric setting, involves projecting the
32 data onto a functional space defined by a family of functions (Li, Qiu, and Xu (2022), Kokoszka and
33 Reimherr (2017)). When the family serves as an orthonormal basis, e.g., Legendre (with the standard
34 scalar product) or Fourier (with another scalar product), the projection is clearly understood, but
35 widely used families such as splines are not orthonormal for the standard scalar product. Leveraging
36 an approximate functional space offers a key advantage: it simplifies the inference problem to
37 estimating coefficients, for example through methods like least squares or maximum likelihood
38 estimation. Subsequently, the function is estimated as the mean of the functional data following
39 projection onto the functional basis.

40 Measuring the uncertainty of an estimator is usually done using confidence intervals. In this paper,
41 our focus lies specifically on providing a simultaneous confidence band for the function means, rather
42 than point-wise confidence intervals. This task presents several challenges: the confidence band
43 must effectively control the simultaneous functional type-I error rate, as opposed to point-wise rates;
44 it must strike a balance between being sufficiently conservative to maintain a confidence level while
45 not being overly so as to render it meaningless; and the method used to construct this confidence
46 band should be computationally feasible for practical application.

47 Several developments have already been proposed to answer these questions. First, consider the
48 case with only one individual (no repetition) but with many time points. Some methods study
49 the asymptotic distribution of the infinity norm between the true function and its estimator. The
50 asymptotic in the number of time points is studied in Hall (1991), Claeskens and Van Keilegom (2003).
51 This approach works only for large datasets in time and is likely to be too conservative otherwise. For
52 small samples, bootstrap methods have been developed to compute the confidence band (Neumann
53 and Polzehl (1998), Claeskens and Van Keilegom (2003)), but with a high computational cost. Another
54 approach is to construct confidence bands based on the volume of the tube formula. Sun and Loader
55 (1994) studied the tail probabilities of suprema of Gaussian random processes. This approach is
56 based on an unbiased linear estimator of the regression function. Zhou, Shen, and Wolfe (1998)
57 used the volume-of-tube formula for estimation by regression splines. Krivobokova, Kneib, and
58 Claeskens (2010) used this method for the construction of confidence bands by penalized spline
59 estimators. They proposed to mix Bayesian and frequentist approaches, to get the good properties

60 from the Bayesian world but reducing the variability to be less conservative using the frequentist
61 approach. The bias is considered through spline modeling, assuming sufficient knots are considered.
62 Wang (2022) proposed Kolmogorov-Smirnov simultaneous confidence band by modeling the error
63 distribution, avoiding the estimation of the covariance structure of the underlying stochastic process.
64 They rely on B-splines for the estimation of the mean curve. Liebl and Reimherr (2019) have proposed
65 a method based on random field theory and the volume-of-tube formula. They provide a band with
66 locally varying widths using an unbiased estimator. Their method does not require the estimation of
67 the full covariance function of the estimator, but only its diagonal. This reduces the computational
68 time. From a practical viewpoint, Sachs, Brand, and Gabriel (2022) introduce a package to popularize
69 simultaneous confidence bands, in the context of survival analysis.

70 Some papers, like ours, rely on several observations of the same function. Bunea, Ivanescu, and
71 Wegkamp (2011) propose a threshold-type estimator and derive error bounds and simultaneous
72 confidence bands, having an unbiased estimator. Telschow and Schwartzman (2022) propose a
73 simultaneous confidence band based on the Gaussian kinematic formula. Again, it assumes access to
74 an unbiased estimator of the function of interest. Note that recent extensions have been proposed, to
75 nonstationary random field in Telschow et al. (2023), based on conformal prediction in Diquigiovanni,
76 Fontana, and Vantini (2022), or having a prediction goal in mind in Hernández, Cugliari, and Jacques
77 (2023) by considering functional time series data set. These extensions are out of the scope of this
78 paper, focusing on the simple functional case.

79 One limitation of all those approaches is that they do not generally take into account the bias of
80 the functional estimator. Sun and Loader (1994) proposed a bias correction for a particular class of
81 functions but left the smoothing parameter choice open, leading to an unusable estimator. In the
82 nonparametric framework, the bias is approximated using the estimator of the second derivative of
83 the underlying mean function (Xia (1998)). But in general, there is a lack of discussion on how to
84 handle the bias of the functional estimator, even in the simple case of a functional space of finite
85 dimension.

86 The objective of this paper is to address the bias problem in confidence band construction for a
87 general function, utilizing a finite functional orthonormal family. Our contributions are as follows:

- 88 • we disentangle the bias issue by explicitly defining the parameter of interest within the approach
89 of Sun and Loader (1994);
- 90 • we propose a bias correction method in a new confidence band for the function of interest;
- 91 • we illustrate this confidence band, concluding on the conservatism of the procedure;
- 92 • finally, we propose a method for selecting the dimension of the approximation space, treating
93 it as a model selection problem, with a trade-off between conservatism and confidence level
94 assurance.

95 Note that while the model selection paradigm has been extensively studied in the literature, in
96 multivariate statistics or functional data analysis (e.g., Goepp, Bouaziz, and Nuel (Submitted), Aneiros,
97 Novo, and Vieu (2022), Basna, Nassar, and Podgórski (2022)), it has not been explored in the context
98 of confidence band construction.

99 The paper is organized as follows: Section 2 introduces the functional regression model, the considered
100 functional family and the corresponding approximate regression models, as well as an estimator
101 defined in the finite space, along with descriptions of the error terms. In Section 3, we propose a
102 confidence band for the approximate regression function in the space of finite dimension, where the
103 dimension is fixed. Section 4 proposes a strategy to construct a confidence band for the true function.
104 This last confidence band being too conservative, Section 5 introduces a model selection criterion
105 to select the best confidence band, doing a trade-off between conservatism and confidence level
106 assurance. Section 7 ends the paper by a conclusion and discussion of perspectives. The different

107 estimation procedures are illustrated throughout the sections.

108 2 Statistical Model

109 In this paper, we consider time series as discrete measurements of functional curves. We first present
 110 the general functional regression model (Section 2.1) where the regression function belongs to a
 111 finite functional family of dimension L^* . In practice, this dimension L^* is unknown and we will
 112 work on functional space of dimension L . The regression model on the finite family of functions is
 113 presented in Section 2.2, and an estimator is proposed in Section 2.3, with a description of the error
 114 terms.

115 2.1 Functional regression model

116 Let y_{ij} be the measure at fixed time $t_j \in [a, b]$ for individual $i = 1, \dots, N$, with $j = 1, \dots, n$. We restrict
 117 ourselves to $[a, b] = [0, 1]$, without loss of generality. We assume these observations are discrete-time
 118 measurements of individual curves, which are independent and noisy realisations of a common
 119 function f that belongs to a functional space. Thus for each individual i , we consider the following
 120 functional regression model

$$y_{ij} = f(t_j) + \varepsilon_{ij},$$

121 where $\varepsilon_i = (\varepsilon_{i1}, \dots, \varepsilon_{in})$ is the noise representing the individual functional variation around f . We
 122 assume that the ε_i are independent. Their distribution is detailed below.

123 For each individual $i = 1, \dots, N$, we denote $y_i = (y_{i1}, \dots, y_{in})$ the $n \times 1$ vector of observations, $t_i =$
 124 (t_1, \dots, t_n) the $n \times 1$ vector of observation times and $f(t) = (f(t_1), \dots, f(t_n))$ the $n \times 1$ vector of the
 125 function f evaluated in t . We also denote $\mathbf{y} = (y_1, \dots, y_N)$ the whole matrix of observations.

126 Let us introduce the functional space $\mathcal{S}^{L^*} = \text{Vect}((t \mapsto B_\ell^{L^*}(t))_{1 \leq \ell \leq L^*})$ with L^* functions $(B_\ell^{L^*})_{1 \leq \ell \leq L^*}$
 127 assumed to be linearly independent. Then, for any $f \in \mathcal{S}^{L^*}$, there exists a unique vector of coefficients
 128 $(\mu_\ell^{L^*})_{1 \leq \ell \leq L^*}$ such that, for all t , $f(t) = \sum_{\ell=1}^{L^*} \mu_\ell^{L^*} B_\ell^{L^*}(t)$. The regression function f verifies the following
 129 assumption:

130 **Definition 2.1.** The function f belongs to the space \mathcal{S}^{L^*} of dimension L^* . It is denoted f^{L^*} and
 131 defined as:

$$f(t) = f^{L^*}(t) = \sum_{\ell=1}^{L^*} \mu_\ell^{L^*} B_\ell^{L^*}(t).$$

132 Many functional spaces are available in the literature, as Splines, Fourier or Legendre families. Let
 133 us consider the space $L^2([0, 1])$ with its standard scalar product $\langle f_1, f_2 \rangle = \int_0^1 |f_1(t)f_2(t)|dt$, for
 134 $f_1, f_2 \in L^2([0, 1])$. We introduce the following assumption:

135 **Definition 2.2.** The functional family $(t \mapsto B_\ell^{L^*}(t))_{1 \leq \ell \leq L^*}$ is orthonormal with respect to the standard
 136 scalar product $\langle \cdot, \cdot \rangle$.

137 Note that if Definition 2.2 holds, one get $\mu_\ell^{L^*} = \langle f^{L^*}, B_\ell^{L^*} \rangle$ for $\ell = 1, \dots, L^*$. The Legendre family is
 138 orthonormal, the Fourier family is orthogonal for the standard scalar product (but not orthonormal),
 139 and the B-splines family is not orthogonal.

140 We also consider a functional noise through the following assumption.

141 **Definition 2.3.** The sequence ε_i is functional and belongs to the functional space $\mathcal{S}^{L^\varepsilon}$ of dimension
 142 L^ε . Then, there exists a sequence of coefficients $(c_{i\ell})_{1 \leq \ell \leq L^\varepsilon}$ such that

$$\varepsilon_{ij} = \sum_{\ell=1}^{L^\varepsilon} c_{i\ell} B_\ell^{L^\varepsilon}(t_j).$$

143 We also assume that the coefficients are Gaussian: for all $i = 1, \dots, N$ and $\ell = 1, \dots, L^\varepsilon$,

$$c_{i\ell} \sim_{iid} \mathcal{N}(0, \sigma^2).$$

144 Definition 2.1 and Definition 2.3 imply that each curve y_i belongs to a finite family: for $j = 1, \dots, n$,

$$y_{ij} = \sum_{\ell=1}^{L^*} \mu_\ell^{L^*} B_\ell^{L^*}(t_j) + \sum_{\ell=1}^{L^\varepsilon} c_{i\ell} B_\ell^{L^\varepsilon}(t_j).$$

145 As the observations are recorded at discrete time points $(t_j)_{1 \leq j \leq n}$, we introduce the family of functions
 146 evaluated at the discrete times of observations. For $L \in \mathbb{N}$, let us denote \mathbf{B}^L the matrix of $n \times L$ with
 147 coefficient in row j and column ℓ equal to $B_\ell^L(t_j)$.

148 Let us introduce $c_i = (c_{i1}, \dots, c_{iL^\varepsilon})$ the $L^\varepsilon \times 1$ vector. Then $\varepsilon_i = \mathbf{B}^{L^\varepsilon} c_i$. The vectors $y_i \in \mathbb{R}^n$ are thus
 149 independent and $y_i \sim \mathcal{N}_n(f(t), \sigma^2 \Sigma^{L^\varepsilon})$ with $\Sigma^{L^\varepsilon} = \mathbf{B}^{L^\varepsilon} (\mathbf{B}^{L^\varepsilon})^T$.

150 To illustrate the model, we simulate a regression functional model with $n = 50$ points per individual
 151 and $N = 40$ individuals. In Figure 1, the function f (red curve) belongs to the Fourier (resp. Legendre
 152 and Spline) family with $L^* = 10$ and the noisy observations y_{ij} (black curves) have a functional noise
 153 in dimension $L^\varepsilon = 20$, also in the Fourier (resp. Legendre and Spline) family on the left plot (resp.
 154 middle and right).

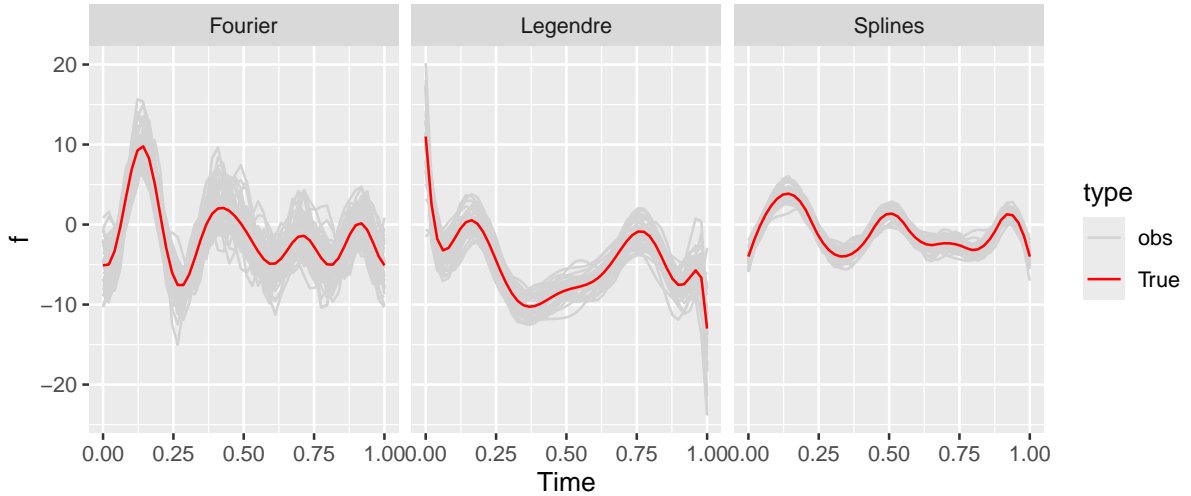


Figure 1: Illustrative example. We generate a regression functional model in the Fourier (left), Legendre (middle) and Splines (right) families. The red curve corresponds to the true function, and the gray curves correspond to noisy observations.

155 The objective of this paper is to construct a tight confidence bound for f^{L^*} using data $(y_{ij})_{ij}$. The
 156 main challenge is that the true dimension L^* is unknown. In the rest of the paper, we will work with
 157 a collection of models defined on a finite family of dimension L with $L \in \{L_{\min}, \dots, L_{\max}\}$, L_{\max} being
 158 chosen to be sufficiently large by the user, expecting that $L^* \leq L_{\max}$. Then we will propose different
 159 strategies to choose the best bandwidths among the different collections.

160 First, in Section 2.2 and Section 2.3, we define for a fixed L the corresponding regression model and
 161 its estimator. Then Section 3, Section 4 and Section 5 will introduce the different bandwidths.

2.2 Approximation of the model on a finite family

Let $f^{L^*} \in \mathcal{S}^{L^*}$ with L^* unknown, and consider the space \mathcal{S}^L for $L \in \{L_{\min}, \dots, L_{\max}\}$ fixed. As \mathcal{S}^L is a family of linearly independent functions, there always exists a unique vector μ^{L, L^*} of coefficients defining $f^{L, L^*}(t) = \sum_{\ell=1}^L \mu_{\ell}^{L, L^*} B_{\ell}^L(t) = B^L(t) \mu^{L, L^*}$ such that

$$f^{L, L^*} = \arg \min_{f \in \mathcal{S}^L} \{\|f^{L^*} - f\|_2^2\},$$

and if the family is orthonormal (Definition 2.2), it corresponds to the projected coefficients μ_{ℓ}^{L, L^*} :

$$\mu_{\ell}^{L, L^*} := \langle f^{L^*}, B_{\ell}^L \rangle.$$

We can prove the following property:

Proposition 2.1. *Under Definition 2.1,*

$$f^{L^*, L^*} = f^{L^*}.$$

Moreover, if Definition 2.2 also holds, the projection coefficients verify

$$\mu_{\ell}^{L, L^*} = \mu_{\ell}^{L^*} \quad \text{for } \ell = 1, \dots, \min(L, L^*).$$

In practice, data are observed at discrete time, we consider the operator \mathbf{P}^L defined as the matrix $\mathbf{P}^L = ((\mathbf{B}^L)^T \mathbf{B}^L)^{-1} (\mathbf{B}^L)^T$ of size $L \times n$ (this operator is a bit more complex when the functional family is not orthonormal wrt the standard scalar product). Then we define the coefficients $\underline{\mu}^{L, L^*}$ which are the coefficients of μ^{L, L^*} approximated on the vector space, denoted \mathbf{S}^L , defined by the matrix \mathbf{B}^L .

$$\underline{\mu}^{L, L^*} := \mathbf{P}^L \mathbf{B}^{L^*} \mu^{L^*}.$$

The corresponding finite approximated regression function is denoted \underline{f}^{L, L^*} and is defined, for all $t \in [0, 1]$, as

$$\underline{f}^{L, L^*}(t) = B^L(t) \underline{\mu}^{L, L^*}.$$

We can prove the following properties linking L, L^* and the number of timepoints n :

Proposition 2.2. *Under Definition 2.1 and Definition 2.2,*

- When $L \geq L^*$, $\mathbf{P}^L \mathbf{B}^{L^*}$ has L^* diagonal elements equal to 1 and other non-diagonal elements close to 0. The first L^* elements of $\underline{\mu}^{L, L^*}$ are equal to μ^{L^*} when $n > L$.
- When $L < L^*$, $\mathbf{P}^L \mathbf{B}^{L^*}$ has L diagonal elements equal to 1. The first L elements of $\underline{\mu}^{L, L^*}$ are different to $\mu_{\ell}^{L^*}$. When $n \rightarrow \infty$, $\underline{\mu}_{\ell}^{L, L^*} \rightarrow \mu_{\ell}^{L^*}$ for $\ell = 1, \dots, \min(L, L^*)$.
- If $n > L^*$, then $f^{L^*} = \underline{f}^{L^*, L^*} = \underline{f}^{L^*, L^*}$.

These properties are illustrated in Figure 2. The true dimension is $L^* = 11$. Three families are considered, Fourier, Legendre and Splines. The plots display the absolute difference between the coefficients $\mu_{\ell}^{L^*}$ and the projected coefficients $\underline{\mu}_{\ell}^{L, L^*}$, for different ℓ in x-axis and for different values of L and n in the y-axis, namely a case with $L < L^*$ and two values of n : $L = 7, n = 20$ and $L = 7, n = 100$; and a case with $L > L^*$ and two values of n : $L = 15, n = 20$ and $L = 15, n = 100$. The absolute difference is represented as a gradient of color, this gradient being adapted to each functional family. We can see that as Legendre (resp. Fourier) are orthonormal (resp. orthogonal) families, the differences are close to 0 when $L = 15$, whatever the values of n . When $L < L^*$, the difference is close to 0 when n is large. This property does not hold for the spline family, which is not orthogonal.

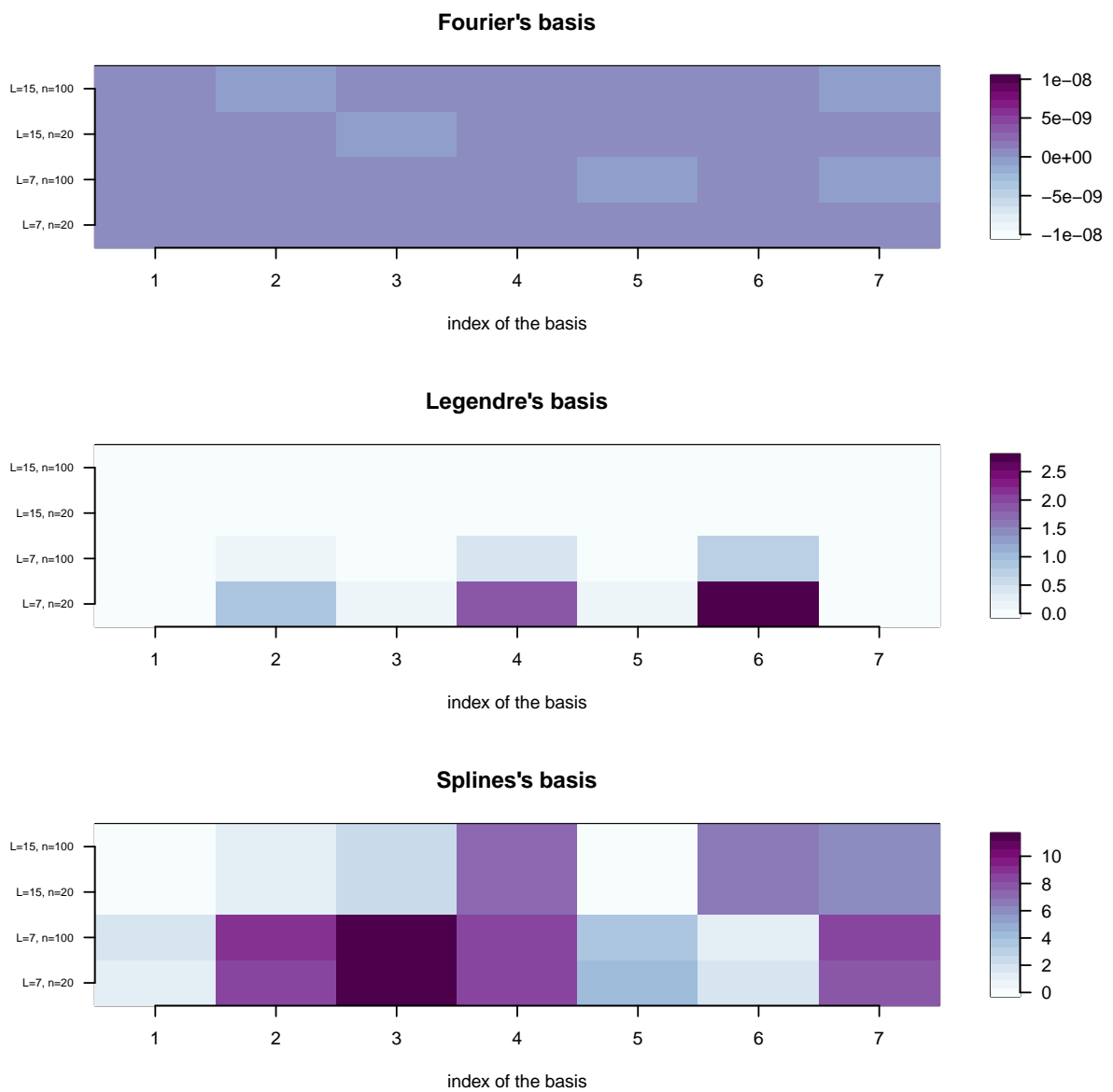


Figure 2: Illustrative example. The true dimension is 11, we generate the coefficients with three families, Fourier (which is orthogonal), Legendre (which is orthonormal) and the splines (which are not orthogonal wrt the standard scalar product). In the y-axis, two dimensions of the family (7 or 15) and two numbers of timepoints (20 or 100) are compared. We plot in x-axis the value of the absolute difference between the true coefficients and their approximations for the first 7 coefficients of the basis. The color scale is adapted to each functional basis.

2.3 Estimator

Let $L \in \{L_{\min}, \dots, L_{\max}\}$. This section presents the least square estimator of the regression function on the space of dimension L defined by the family \mathbf{B}^L and discusses its error.

2.3.1 Estimation of the regression function

When considering the estimation of the regression function f^{L^*} on the space of dimension L defined by the family \mathbf{B}^L , we do not directly estimate f^{L^*} but its projection on this finite space, which corresponds to the projected function $\underline{f}^{L,L^*}(t)$ and its associated coefficients $(\underline{\mu}_\ell^{L,L^*})_{1 \leq \ell \leq L}$.

Definition 2.4. The vector of coefficients $(\underline{\mu}_\ell^{L,L^*})_{1 \leq \ell \leq L}$ is estimated by the least square estimator $\hat{\underline{\mu}}^{L,L^*}$ defined as:

$$\hat{\underline{\mu}}^{L,L^*} := \frac{1}{N} \sum_{i=1}^N \mathbf{P}^L y_i.$$

For a fixed $t \in [0, 1]$, the estimator of the function $\underline{f}^{L,L^*}(t)$ is defined by:

$$\underline{f}^{L,L^*}(t) = \sum_{\ell=1}^L \hat{\underline{\mu}}_\ell^{L,L^*} B_\ell^L(t) = B^L(t) \hat{\underline{\mu}}^{L,L^*}. \quad (1)$$

Equation 1 directly implies that the estimator is thus the empirical mean of the functional approximation of each individual vector of observations. Because we work with least squares estimators, we can easily study the error of estimation of $\hat{\underline{\mu}}^{L,L^*}$ and \underline{f}^{L,L^*} .

Proposition 2.3. Under Definition 2.1 and Definition 2.3, we have

$$\hat{\underline{\mu}}^{L,L^*} \sim \mathcal{N}_L \left(\underline{\mu}^{L,L^*}, \frac{\sigma^2}{N} \Sigma_B^{L,L^*} \right),$$

where the $L \times L$ covariance matrix Σ_B^{L,L^*} is defined as $\Sigma_B^{L,L^*} := \mathbf{P}^L \Sigma^{L^*} (\mathbf{P}^L)^T$ with $\Sigma^{L^*} = \mathbf{B}^{L^*} (\mathbf{B}^{L^*})^T$.

Moreover, $B^L(\cdot) \mathbf{P}^L y_i$ is a Gaussian process with mean $\underline{f}^{L,L^*}(\cdot)$ and covariance function $(s, t) \mapsto \sigma^2 B^L(s) \Sigma_B^{L,L^*} (B^L(t))^T$, and $(\underline{f}^{L,L^*} - \hat{\underline{f}}^{L,L^*})(\cdot)$ is a centered Gaussian process with covariance function $C^{L,L^*} : (s, t) \mapsto \frac{\sigma^2}{N} B^L(s) \Sigma_B^{L,L^*} B^L(t)^T$.

The proof is given in Appendix.

Figure 3 displays estimators calculated with different dimensions L . Data are generated with $L^* = 11$, $L^\varepsilon = 20$, $n = 50$ and $N = 40$. The true function and its projection \underline{f}^{L,L^*} are in cyan, and the estimator $\hat{\underline{f}}^{L,L^*}$ is in red. We compute it for the three families, Legendre, Fourier and splines. In all cases, the estimators are very precise when considering the relevant dimension, but estimating a function of dimension L^* with a function of dimension $L < L^*$ is not consistent. Note that the performance of the estimator for the spline family is also good, even if the family is not orthonormal, because we work here at the level of the function (and not at the level of the coefficients).

Even if the estimator $\hat{\underline{f}}^{L,L^*}$ is defined on the functional space associated to \mathbf{S}^L , it can also be seen as an estimator of the function f^{L^*} which lies in the space \mathcal{S}^{L^*} . In that case, the error includes a functional approximation term due to the approximation of f^{L^*} on the space \mathcal{S}^L , which will be nonzero if

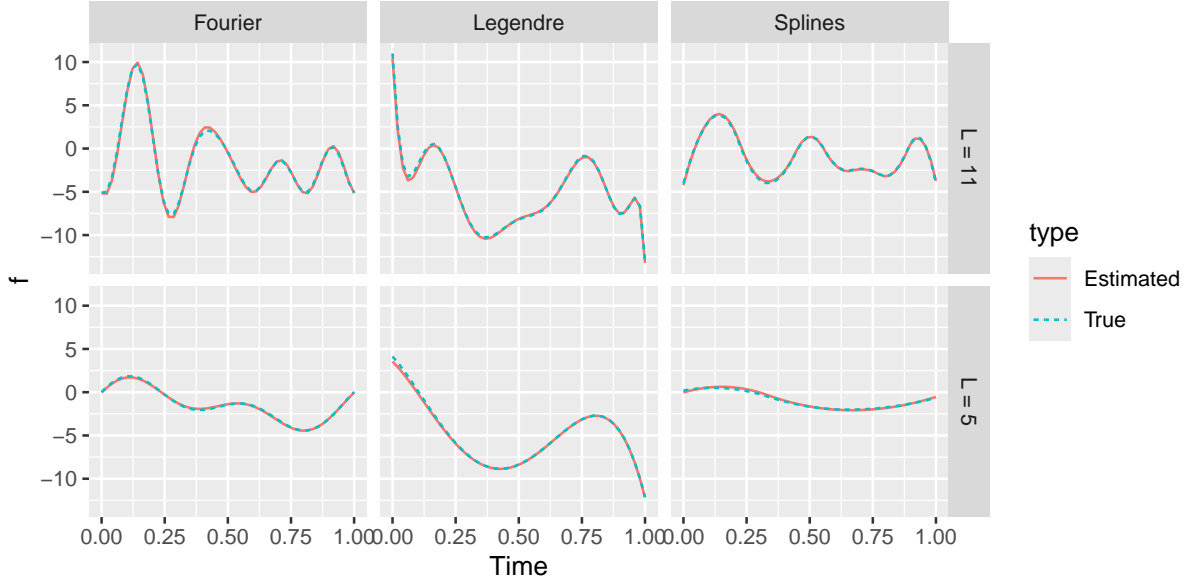


Figure 3: Illustrative example. For each family (Fourier which is orthogonal, Legendre which is orthonormal and the splines which are not orthogonal wrt the standard scalar product), we consider a function with true dimension 11 (top), and its projection on the space of dimension 5 (bottom), displayed in cyan. The estimators in dimensions 11 and 5 are displayed in red.

222 $L \neq L^*$. It corresponds to the bias of the estimator $\underline{f}^{\mathcal{J},L^*}$, i.e. the difference between its expectation
 223 and the true f^{L^*} . Indeed, recalling that $f^{L^*} = \underline{f}^{L^*,L^*}$, the error of estimation can be decomposed into

$$\underline{f}^{\mathcal{J},L^*}(t) - f^{L^*}(t) = \underline{f}^{\mathcal{J},L^*}(t) - \underline{f}^{L^*,L^*}(t) + \underline{f}^{L^*,L^*}(t) - f^{L^*}(t) =: \text{Stat}_{L,L^*}(t) + \text{Bias}_{L,L^*}(t), \quad (2)$$

224 The first term $\text{Stat}_{L,L^*}(t) = \underline{f}^{\mathcal{J},L^*}(t) - \underline{f}^{L^*,L^*}(t)$ is the (unrescaled) statistics of the model. The second
 225 term $\text{Bias}_{L,L^*}(t) = \mathbb{E}(\underline{f}^{\mathcal{J},L^*}(t)) - \underline{f}^{L^*,L^*}(t)$ is the bias of the estimator $\underline{f}^{\mathcal{J},L^*}(t)$ when estimating the true
 226 function $\underline{f}^{L^*,L^*}(t)$.

227 Let us remark that this bias is different than the bias of the estimator $\underline{f}^{\mathcal{J},L^*}(t)$ when estimating the
 228 projected function $\underline{f}^{L^*,L^*} = f^{L^*}$, which is 0.

229 The two terms defined in Equation 2 are more detailed in the two next subsections.

230 2.3.2 Statistics

231 The statistics of the model, $t \mapsto \text{Stat}_{L,L^*}(t) = \underline{f}^{\mathcal{J},L^*}(t) - \underline{f}^{L^*,L^*}(t)$, is a random functional quantity which
 232 depends on the estimator $\underline{f}^{\mathcal{J},L^*}$. From Proposition 2.3, for any $t \in [0, 1]$, we define the centered and
 233 rescale statistics $Z_L(t)$ such that:

$$Z_L(t) := \frac{\text{Stat}_{L,L^*}(t)}{\sqrt{\text{Var}(\text{Stat}_{L,L^*}(t))}} = \frac{\underline{f}^{\mathcal{J},L^*}(t) - \underline{f}^{L^*,L^*}(t)}{\sqrt{C^{L,L^*}(t,t)}} \sim \mathcal{N}(0, 1).$$

234 The covariance function can be estimated using the observations y_i as

$$\hat{C}^{L,L^*}(s,t) = \frac{1}{N-1} \sum_{i=1}^N (B^L(s)\mathbf{P}^L y_i - \underline{f}^{\mathcal{J},L^*}(s))(B^L(t)\mathbf{P}^L y_i - \underline{f}^{\mathcal{J},L^*}(t)).$$

235 **2.3.3 Bias**

236 The bias is due to the fact that the estimation is potentially performed in a different (finite) space
 237 than the space where the true function \underline{f}^{L^*,L^*} lives. This is a functional bias, which is not random. It
 238 corresponds to the approximation (orthogonal projection if Definition 2.2 holds) of f^{L^*} from \mathcal{S}^{L^*}
 239 to the space \mathcal{S}^L . It can be written as follows:

$$\text{Bias}_{L,L^*}(t) = B^L(t)\underline{\mu}^{L,L^*} - B^{L^*}(t)\mu^{L^*}.$$

240 Thus, we can deduce that when $L < L^*$ and if the family is orthonormal (Definition 2.2 holds),

$$\text{Bias}_{L,L^*}(t) = \sum_{\ell=1}^L B_{\ell}^L(t)\underline{\mu}_{\ell}^{L,L^*} - \sum_{\ell=1}^{L^*} B_{\ell}^{L^*}(t)\underline{\mu}_{\ell}^{L^*} = \sum_{\ell=L+1}^{L^*} B_{\ell}^{L^*}(t)\underline{\mu}_{\ell}^{L^*}.$$

241 From Proposition 2.3, we can directly deduce the following proposition:

242 **Proposition 2.4.** *Under Definition 2.1 and Definition 2.3, the mean is, for all $t \in [0, 1]$,*

- 243 • for $L < L^*$, $\text{Bias}_{L,L^*}(t) \neq 0$,
- 244 • for $L \geq L^*$, $\text{Bias}_{L,L^*}(t) = 0$.

245 In the next section, we explain how to use this property to derive confidence bands of \underline{f}^{L,L^*} and f^{L,L^*} .

246 **3 Confidence Bands of \underline{f}^{L,L^*} and f^{L,L^*} for a fixed L**

247 The objective is to construct a confidence band for the two functions \underline{f}^{L,L^*} and f^{L,L^*} , based on the
 248 observations \mathbf{y} , for a given value $L \in \{L_{\min}, \dots, L_{\max}\}$. The band for \underline{f}^{L,L^*} enters the framework
 249 proposed by Sun and Loader (1994) which relies on an unbiased and linear estimator of the function.
 250 This is the case for the estimator \underline{f}^{L,L^*} which is an unbiased estimator of \underline{f}^{L,L^*} . We recall in Section 3.1
 251 the construction of this confidence band which attains a given confidence level in a non-asymptotic
 252 setting, that is for a finite number of observations n for each individual. Then in Section 3.2, we prove
 253 that the confidence band proposed by Sun and Loader (1994) can be viewed as a confidence band for
 254 f^{L,L^*} with an asymptotic confidence level, the asymptotic framework being considered when $n \rightarrow \infty$.

255 **3.1 Confidence band for \underline{f}^{L,L^*}**

256 Let $L \in \{L_{\min}, \dots, L_{\max}\}$. Consider $1 - \alpha$ as a fixed confidence level. Our aim is to find a function $d^L()$
 257 such that

$$\mathbb{P}\left(\forall t \in [0, 1], \underline{f}^{L,L^*}(t) - d^L(t) \leq \underline{f}^{L,L^*}(t) \leq \underline{f}^{L,L^*}(t) + d^L(t)\right) = 1 - \alpha.$$

258 Consider the normalized statistics $Z_L(t)$ which is a centered and reduced Gaussian process. We want
 259 to find the quantile q^L satisfying

$$q^L = \arg \min_q \left\{ \mathbb{P}\left(\max_{t \in [0,1]} |Z_L(t)| \leq q\right) = 1 - \alpha \right\}. \quad (3)$$

260 Then we can take $d^L(t) = q^L \sqrt{C^{L,L^*}(t, t)}$. The covariance function $C^{L,L^*}(t, t)$ can be replaced by its
 261 estimator $\hat{C}^{L,L^*}(t, t)$, making the distribution a Student's distribution with $N - 1$ degrees of freedom.
 262 Thus, it only requires to be able to compute the critical value q^L .

263 This can be done following Sun and Loader (1994) who propose a confidence band for a centered
 264 Gaussian process. Their procedure is based on an unbiased linear estimator of the function of interest,
 265 which is the case for $\underline{f}^{\underline{L}, L^*}$ when we consider a band for $\underline{f}^{\underline{L}, L^*}$. We recall their result in the following
 266 proposition, the computation of the value q^L is detailed thereafter.

267 **Theorem 3.1** (Sun and Loader (1994)). *Set Definition 2.1 and Definition 2.3 and a probability $\alpha \in [0, 1]$.
 268 Then, we have*

$$\mathbb{P}\left(\forall t \in [0, 1], \left| \underline{f}^{\underline{L}, L^*}(t) - \underline{f}^{\underline{L}, L^*}(t) \right| \leq \hat{d}^L(t)\right) = 1 - \alpha$$

269 with

$$\hat{d}^L(t) = \hat{q}^L \sqrt{\hat{C}^{\underline{L}, L^*}(t, t)/N};$$

270 and \hat{q}^L defined as the solution of the following equation, seen as a function of q^L :

$$\alpha = \mathbb{P}(|t_{N-1}| > q^L) + \frac{\|\tau^L\|_1}{\pi} \left(1 + \frac{(q^L)^2}{N-1}\right)^{-(N-1)/2}, \quad (4)$$

271 with $(\tau^L)^2(t) = \partial_{12}c(t, t) = \text{Var}(Z_L(t))'$ where we denote $\partial_{12}c(t, t)$ the partial derivatives of a function
 272 $c(t, s)$ in the first and second coordinates and then evaluated at $t = s$.

273 We can thus deduce a confidence band of level $1 - \alpha$ for $\underline{f}^{\underline{L}, L^*}$:

$$CB_1(\underline{f}^{\underline{L}, L^*}) = \{\forall t \in [0, 1], [\underline{f}^{\underline{L}, L^*}(t) - \hat{d}^L(t); \underline{f}^{\underline{L}, L^*}(t) + \hat{d}^L(t)]\}.$$

274 The value \hat{q}^L is defined implicitly in Equation 4 which involves the quantity $t \mapsto \tau^L(t)$. Liebl and
 275 Reimherr (2019) propose to estimate $\tau^L(t)$, for all t , by

$$\begin{aligned} \hat{\tau}^L(t) &= \left(\widehat{\text{Var}}((U^L)'_1(t), \dots, (U^L)'_N(t))\right)^{1/2} \\ &= \left(\frac{1}{N-1} \sum_{i=1}^N \left((U^L)'_i(t) - \frac{1}{N} \sum_{j=1}^N (U^L)'_j(t)\right)^2\right)^{1/2}, \end{aligned}$$

276 where $U_i^L(t) = (P^L y_i(t) - \underline{f}^{\underline{L}, L^*}(t))/(\hat{C}^{\underline{L}, L^*}(t))^{1/2}$ and $(U^L)'_i$ is a smooth version of the differentiated
 277 function U_i^L . Then we take the L_1 -norm of $\hat{\tau}^L$.

278 Let us describe the behavior of \hat{d}^L :

- 279 • $\|\hat{d}^L\|_\infty$ increases with L
- 280 • When the functions $(B_t^L)_{1 \leq t \leq L}$ consists in an orthonormal family, $\|\hat{d}^L\|_\infty$ increases with L until
 281 $L = L^*$ and then $\|\hat{d}^L\|_\infty$ is constant with L .

282 This band is illustrated on Figure 4. It displays on the top row several functional data generated
 283 under either the Fourier family (left), Legendre (middle) or Spline (right), on the middle row the
 284 confidence bands of $\underline{f}^{\underline{L}, L^*}$ for different values of $L = 3, 5$ and 11 , and on the bottom row the bound \hat{d}^L .
 285 The true functions $\underline{f}^{\underline{L}, L^*}$ are displayed in cyan and the confidence bands in purple. The bands are
 286 very precise for each L . The behavior of \hat{d}^L increases with L . As d^L can be seen as a variance, $\hat{d}^L(t)$ is
 287 larger on the boundary of the time domain, as there are less observations near 0 and 1.

288 We also evaluate numerically the levels of the obtained confidence bands. For this, 1000 datasets are
 289 simulated, the confidence band is estimated for each of them. The empirical confidence level is then

290 evaluated as the proportion of confidence bands that contain the true function. Table 1 presents the
 291 empirical confidence levels for different values of L and two sample sizes $n = 50$ and $n = 150$, with
 292 $N = 40$. The level is the expected one whatever the value of L , especially when $L < L^*$ and $L > L^*$
 293 but also when $L > L^e$. We will see in the next sections that this will not be the case for the debiased
 294 confidence band.

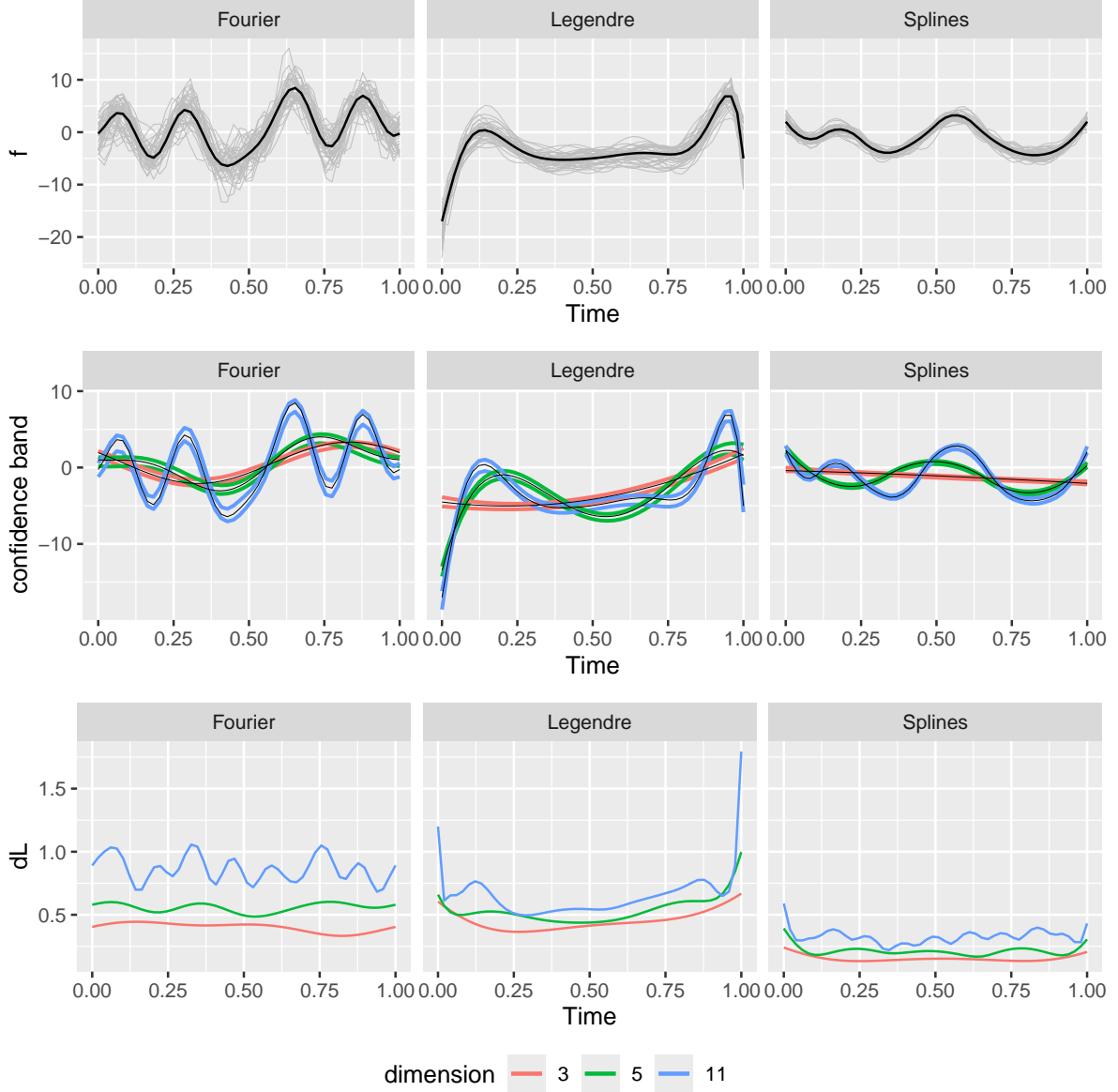


Figure 4: Illustrative example. For the three families, resp. Fourier, Legendre and the splines, we display on the top row the observed functional data, on the middle row the confidence bands for different values of L (3, 5 and 11), and on the bottom row the bound dL .

295 3.2 Asymptotic confidence band for f^{L,L^*}

296 Note that if one works in the asymptotic framework $n \rightarrow \infty$, the previous definition of \hat{d}^L induces a
 297 natural asymptotic confidence band for the function f^{L,L^*} . Indeed, we can prove that

298 **Theorem 3.2.** Set Definition 2.1 and Definition 2.3 and a probability $\alpha \in [0, 1]$. Then, we have,

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(\forall t \in [0, 1], |\underline{f}^{L,L^*}(t) - f^{L,L^*}(t)| \leq \hat{d}^L(t) \right) = 1 - \alpha,$$

Table 1: Illustrative example. The confidence level of the confidence band is evaluated from 1000 repetitions. Confidence bands are calculated with the Legendre family, for several L in rows and several n in columns.

L	n	
	50	150
3	0.945	0.949
5	0.950	0.949
11	0.871	0.944
15	0.904	0.953
21	0.914	0.946
25	0.914	0.946

with $\hat{d}^L(t) = \hat{q}^L \sqrt{\hat{C}^{L,L^*}(t,t)/N}$ and \hat{q}^L defined as the solution of Equation 4.

The proof is given in Appendix.

Then a confidence band for f^{L,L^*} at the asymptotic confidence level $1 - \alpha$ for a large number of observations n is given by

$$CB(f^{L,L^*}) = \{\forall t \in [0, 1], [\underline{f}^{L,L^*}(t) - \hat{d}^L(t); \underline{f}^{L,L^*}(t) + \hat{d}^L(t)]\}.$$

We do not provide any illustration of this property, as it would be similar than the previous ones. Indeed, we notice that the asymptotic is achieved even when n is small on our examples.

4 Confidence Band of f^{L^*} by correcting the bias

The function of interest is $f^{L^*} = \underline{f}^{L^*,L^*}$, rather than \underline{f}^{L,L^*} . Therefore, our aim is to construct a confidence bound for f^{L^*} . However, an unbiased estimator of f^{L^*} is unavailable by definition, since the true dimension L^* is unknown. Instead, we propose to work with the estimator \underline{f}^{L,L^*} and to debias the corresponding confidence band.

To achieve this, we use the decomposition outlined in Equation 2 between the bias term and the statistical term. The idea is to bound the infinity norm of these two terms. A first strategy consists in bounding each term separately and then summing the two bounds to construct the confidence band. However, this approach tends to produce a band that is too large and conservative. The reason is that applying the infinite norms on each term before bounding them does not take into account the functional nature of the two terms.

A second strategy consists in keeping the functional aspect by bounding the infinity norm of the sum of the functional two terms. This approach is detailed in this section.

In Section 4.1, we first rewrite the band as a band around $\underline{f}^{L,L^*}(t)$. We then use a first subsample \mathbf{y}^1 to estimate the bound as defined in Section 3. A second subsample \mathbf{y}^2 is used to estimate the bias term (without the infinite norm). This yields a pointwise correction of the bias, and the final confidence band is centered around $\underline{f}^{L_{\max},L^*}$. This procedure provides a collection of confidence bands, for $L \in \{L_{\min}, \dots, L_{\max}\}$ with varying width. Then, in Section 4.2, we propose a criterion to select the “best” band by minimizing its width.

4.1 Construction of the band of f^{L^*} for a given L

We introduce two independent sub-samples \mathbf{y}^1 and \mathbf{y}^2 of \mathbf{y} of length N_1 and N_2 such that $N_1 + N_2 = N$.

We use the first sub-sample \mathbf{y}^1 to calculate $\hat{f}_1^{L,L^*}(t)$, an estimator of $\underline{f}^{L,L^*}(t)$ and a functional bound denoted \hat{d}_1^L that controls the bias term $\underline{f}^{L,L^*}(t) - \hat{f}_1^{L,L^*}(t)$. This bound is defined in Section 3 applied on \mathbf{y}^1 , for a given level α , such that:

$$\mathbb{P}\left(\forall t \in [0, 1], -\hat{d}_1^L(t) \leq \underline{f}^{L,L^*}(t) - \hat{f}_1^{L,L^*}(t) \leq \hat{d}_1^L(t)\right) = 1 - \alpha. \quad (5)$$

Then, we need to control the bias $Bias_{L,L^*}(t) = \underline{f}^{L,L^*}(t) - f^{L^*}(t)$. Recall that when L_{\max} is large enough and $n > L_{\max}$, $f^{L^*} = \underline{f}^{L_{\max},L^*}$. Therefore, we want to control the $Bias_{L,L^*}(t) = \underline{f}^{L,L^*}(t) - \underline{f}^{L_{\max},L^*}(t)$. It would be tempting to replace $Bias_{L,L^*}(t)$ by its estimation based on the second sample \mathbf{y}^2 . But this would introduce an estimation error that we also need to control, in the same spirit than what is done in Lacour, Massart, and Rivoirard (2017). We can use again Section 3 to compute the function $\hat{d}_2^{L,L_{\max}}(t)$ on the sample \mathbf{y}^2 , and the functional estimators $\hat{f}_2^{L,L^*}(t)$ and $\hat{f}_2^{L_{\max},L^*}(t)$ of $\underline{f}^{L,L^*}(t)$ and $\underline{f}^{L_{\max},L^*}(t)$, respectively. This allows to construct the following band for $\underline{f}^{L,L^*}(t) - \underline{f}^{L_{\max},L^*}(t)$ for a confidence level $1 - \beta$,

$$\mathbb{P}\left(\forall t \in [0, 1], -\hat{d}_2^{L,L_{\max}}(t) \leq \underline{f}^{L_{\max},L^*}(t) - \underline{f}^{L,L^*}(t) - (\hat{f}_2^{L_{\max},L^*}(t) - \hat{f}_2^{L,L^*}(t)) \leq \hat{d}_2^{L,L_{\max}}(t)\right) = 1 - \beta. \quad (6)$$

Combining Equation 5 and Equation 6, we can provide a debiased confidence band of $f^{L^*}(t)$.

Proposition 4.1. *Let us define*

$$\begin{aligned} \hat{\theta}_1^L(t) &:= -\hat{d}_1^L(t) - \hat{d}_2^{L,L_{\max}}(t) + \hat{f}_2^{L_{\max},L^*}(t) - \hat{f}_2^{L,L^*}(t) \\ \hat{\theta}_2^L(t) &:= \hat{d}_1^L(t) + \hat{d}_2^{L,L_{\max}}(t) + \hat{f}_2^{L_{\max},L^*}(t) - \hat{f}_2^{L,L^*}(t), \end{aligned}$$

where $\hat{d}_1^L(t)$ is defined on sample \mathbf{y}^1 by Equation 5 for a level α and $\hat{d}_2^{L,L_{\max}}(t)$ is defined on sample \mathbf{y}^2 by Equation 6 for a level β . Then we have

$$\mathbb{P}\left(\forall t \in [0, 1], \hat{\theta}_1^L(t) \leq f^{L^*}(t) - \hat{f}_1^{L,L^*}(t) \leq \hat{\theta}_2^L(t)\right) \geq 1 - \alpha\beta.$$

The proof is given in Appendix.

This defines a confidence band which can be defined either around \hat{f}_1^{L,L^*} :

$$CB_2(\underline{f}^{L^*}) = \{\forall t \in [0, 1], [\hat{f}_1^{L,L^*}(t) + \hat{\theta}_1^L(t); \hat{f}_1^{L,L^*}(t) + \hat{\theta}_2^L(t)]\}$$

or around $\hat{f}_2^{L_{\max},L^*}$:

$$CB_2(\underline{f}^{L^*}) = \{\forall t \in [0, 1], [\hat{f}_2^{L_{\max},L^*}(t) + \hat{f}_1^{L,L^*}(t) - \hat{f}_2^{L,L^*}(t) - \hat{d}_1^L(t) - \hat{d}_2^{L,L_{\max}}(t); \hat{f}_2^{L_{\max},L^*}(t) + \hat{f}_1^{L,L^*}(t) - \hat{f}_2^{L,L^*}(t) + \hat{d}_1^L(t) + \hat{d}_2^{L,L_{\max}}(t)]\}$$

344 The two functions $\hat{d}_1^L(t)$ and $\hat{d}_2^{L,L_{\max}}(t)$ are of the same order as they are built with the same approach.
 345 They depend on the length of the samples. To obtain the thinnest band, the best strategy is to split
 346 the sample in two sub-samples of equal length $n_1 = n_2 = n/2$.

347 The behavior of \hat{d}_1^L has been described in Section 3. Let us describe the behavior of $\hat{d}_2^{L,L_{\max}}$:

- 348 • $\|\hat{d}_2^{L,L_{\max}}\|_{\infty}$ decreases with L .
- 349 • When $L > L^{\varepsilon}$, $\|\hat{d}_2^{L,L_{\max}}\|_{\infty}$ is constant with L and the probability in Equation 6 is equal to 1.
- 350 • When $L^* < L < L^{\varepsilon}$, $\|\hat{d}_2^{L,L_{\max}}\|_{\infty}$ is constant with L when the functions B_{ξ}^L consists in an
 351 orthonormal family. Otherwise, the behavior is erratic.

352 It means that when the band defined in Proposition 4.1 is calculated for $L > L^{\varepsilon}$, the confidence level
 353 is $1 - \alpha$ instead of $1 - \alpha\beta$.

354 The advantages of this approach is that the bias of the band is corrected and the level for the true
 355 function f^{L^*} is guaranteed when L^{ε} is large. This was the main objective of the paper.

356 We illustrate numerically those advantages. In Figure 5, top row, we plot the confidence bands
 357 obtained for different dimensions $L \in \{3, 5, 11, 15, 21\}$ with Fourier, Legendre and Splines families
 358 and $\alpha = \beta = \sqrt{0.05} \approx 0.22$. We can see that all the confidence bands are alike. Especially, they
 359 are unbiased, even for $L = 3$. A larger dimension L provides a smoother band. On the middle and
 360 bottom rows of Figure 5, we illustrate the two terms that enter the confidence band, $t \mapsto \hat{d}_1^L(t)$ and
 361 $t \mapsto \hat{d}_2^{L,L_{\max}}(t)$. Their behavior is the same along time. The function $\hat{d}_1^L(t)$ can be seen as a variance,
 362 this is why it is larger near 0 and 1 where there are less observations. The function $\hat{d}_2^{L,L_{\max}}(t)$ is
 363 smaller than $\hat{d}_1^L(t)$ because it controls the remaining rest after the projection. Note that as expected
 364 when $L > L^{\varepsilon}$, $\hat{d}_2^{L,L_{\max}}(t)$ is close to 0. As explained before, the influence of L is not the same for the
 365 two functions. When L increases, $\hat{d}_1^L(t)$ increases while $\hat{d}_2^{L,L_{\max}}(t)$ decreases.

366 In Table 2, we simulate 1000 repeated datasets with two sample sizes $n = 50$ and $n = 150$. For each
 367 dataset, we compute the confidence band defined in Proposition 4.1 with a theoretical confidence
 368 level of $1 - \alpha\beta = 0.95$ and for different values of L . Then the confidence level is approximated as the
 369 proportion of confidence bands containing the true function f . Remark that when $L < L^{\varepsilon}$, the level
 370 is the expected one, that is 0.95. When $L > L^{\varepsilon}$, the level is not more ensured, as explained before.
 371 Indeed the term $\hat{d}_2^{L,L_{\max}}$ is mainly equal to 0, and the level is close to $1 - \alpha$ instead of $1 - \alpha\beta$. This is
 372 not the case for the band in Section 3, as this is due to the correction of the bias.

373 [1] 3
 374 [1] 5
 375 [1] 11
 376 [1] 15
 377 [1] 21
 378 [1] 25
 379 [1] 3
 380 [1] 5
 381 [1] 11
 382 [1] 15
 383 [1] 21
 384 [1] 25

385 4.2 Influence of L

386 This approach gives a collection of confidence bands for different values of L . The confidence bands
 387 have different widths for a same confidence level $1 - \alpha\beta$. It is thus natural to want to select one

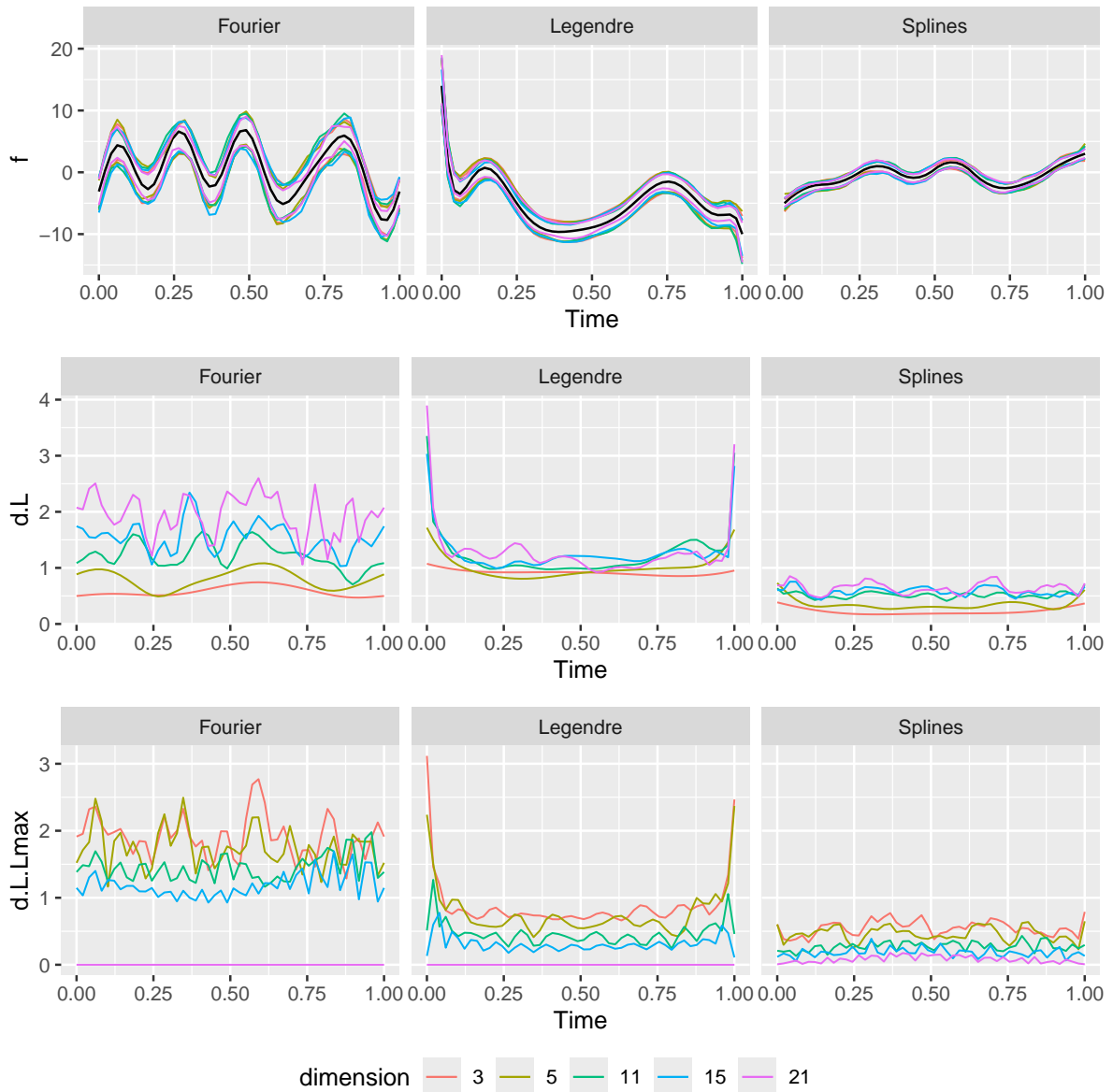


Figure 5: Illustrative example. For a given dataset, we plot several confidence bands (top row), functions dL (middle row) and $dLLmax$ (bottom row). Bands and functions are estimated with Fourier (left column), Legendre (middle column) and Spline (right column) basis and several dimensions L (3, 5, 11, 15, 21).

Table 2: Illustrative example. We display the level of confidence for the proposed confidence band, for several L in rows and several n in columns.

L	n	
	50	150
3	0.946	0.941
5	0.961	0.958
11	0.961	0.970
15	0.944	0.957
21	0.787	0.790
25	0.787	0.790

388 of them. This means we want to select the best dimension L among the collection $\{L_{\min}, \dots, L_{\max}\}$.
 389 We need to define what “best” means. It is quite intuitive to focus on the band that is the thinnest.
 390 Thinnest could be thought in different norms. Here we consider the infinity norm of the width of the
 391 confidence band, which gives a preference to smooth bands. We thus define the following criteria

$$\hat{L} = \arg \min_L \left\{ \sup_t |\hat{\theta}_2^L(t) - \hat{\theta}_1^L(t)| \right\} = \arg \min_L \left\{ \sup_t |\hat{d}^L(t) + \hat{d}^{L, L_{\max}}(t)| \right\}. \quad (7)$$

392 We illustrate the different terms involved in Equation 7. In Figure 6, we plot for a given dataset,
 393 the infinity norm of the width of the band $\hat{d}^L(t) + \hat{d}^{L, L_{\max}}(t)$ (top), of $\hat{d}^L(t)$ (middle) and $\hat{d}^{L, L_{\max}}(t)$
 394 (bottom) functions obtained with the Fourier (left column), Legendre (middle column) and Spline
 395 (right column) basis. As already said, $\|\hat{d}^L\|_{\infty}$ increases with L while $\|\hat{d}^{L, L_{\max}}\|_{\infty}$ decreases (and is zero
 396 when $L > L^{\epsilon}$). The width of the band wrt L does not have a U -shape, as expected. It is thus difficult
 397 to minimize this criterion.

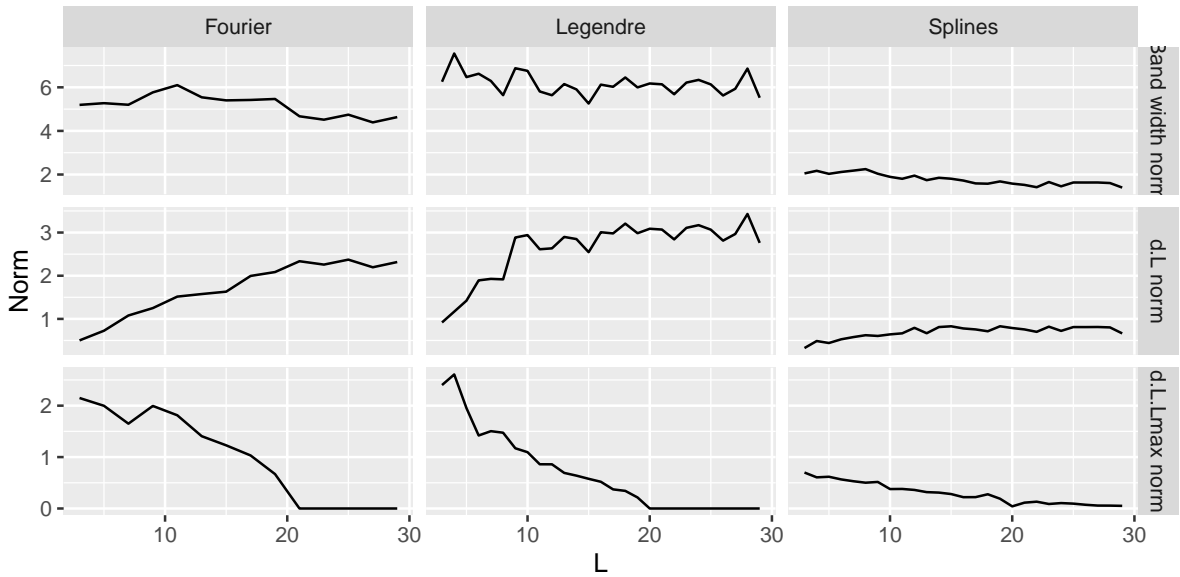


Figure 6: Illustrative example. For a given dataset, we calculate the norm of the width of the confidence band (top), of the dL function (middle) and the dLLmax function (bottom), for several dimensions L and for Fourier (left column), Legendre (middle column) and Splines (right column) basis.

398 We then evaluate the performance of the selection criteria. To do that, we simulate 100 repeated
 399 datasets. Confidence bands and the norm of their widths are computed for several L . We apply the
 400 selection criteria and plot the distribution of the estimated dimension \hat{L} in Figure 7, for the three
 401 basis families. The estimated dimension is almost always larger than the true $L^* = 11$. Being larger
 402 is not a problem and the selected band is unbiased and has the correct level as soon as L^{ϵ} is large,
 403 which was the objective. However, the criteria has the tendency to select a (too) smooth band. In
 404 Section 4.3, we also illustrate that this band is too conservative.

405 4.3 Comparison with the confidence bands of Section 3

406 The reformulation of the band around $\underline{f}_2^{I_{\max}, L^*}$ is close to the band presented in Section 3 for $L = L_{\max}$,
 407 that is a band centered around $\underline{f}^{I_{\max}, L^*}$. A natural question is to understand what is the gain by

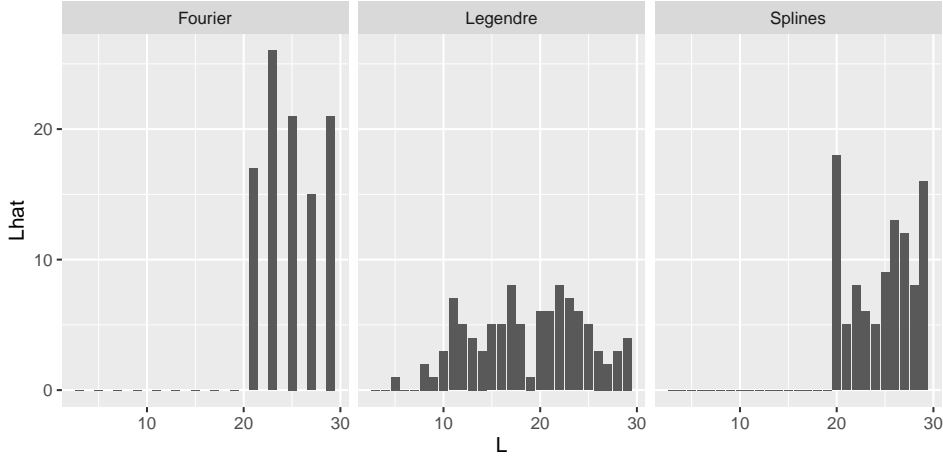


Figure 7: Illustrative example. From 100 datasets, we calculate the distribution of the estimated dimension L . The true dimension is $L^* = 11$.

408 doing so instead of using the band from Section 3 with $L = L_{\max}$, namely the band $[\underline{f}^{L_{\max}, L^*}(t) -$
 409 $\hat{d}^{L_{\max}}(t); \underline{f}^{L_{\max}, L^*}(t) + \hat{d}^{L_{\max}}(t)]$. To do that, we have to understand the behavior of the different terms.

410 It is difficult to compare theoretically the width of the two bands. We compare them with simulations.
 411 For 100 repeated datasets, we compute three different confidence bands: the band CB_2 defined in
 412 Proposition 4.1 with \hat{L} defined in Equation 7, the band CB_1 constructed in Section 3 with L_{\max} and
 413 the ideal (and not accessible) band constructed in Section 3 with the true L^* . In Figure 8, we present
 414 the boxplots of the norms of the width of the band with \hat{L} (left), with L_{\max} (middle) and L^* (right).
 415 The width of the confidence band with the true L^* is smaller, which is expected but unfortunately not
 416 achievable. What was not expected, but sad, is that the width of the confidence band CB_1 constructed
 417 in Section 3 with some L_{\max} is smaller than our band CB_2 with a correction of the bias and the model
 418 selection criteria. This may be understood because we estimate two different quantities, on smaller
 419 dataset, for more conservative level ($1 - \alpha$ and $1 - \beta$ respectively) to achieve at the end the confidence
 420 level of $1 - \alpha\beta$. The use of the two independent subsamples is mandatory to control the probability
 421 in the proof of Theorem 3.1. Therefore it is not possible to correct this problem. In the next section,
 422 we come back to the confidence bands proposed in Section 3 and propose a model selection criterion
 423 to take into account the bias.

424 5 Selection of the best confidence band with a criteria taking into 425 account the bias

426 In this section, we want to use the collection of confidence bands defined in Section 3 without
 427 correcting their bias but instead by proposing a criteria which is a trade-off between this bias and the
 428 dimension of the basis. To do that, we propose a new heuristic criteria going back to the definition
 429 of the band itself seen as the estimation of a quantile of a certain empirical process. The criteria is
 430 inspired by model selection tools to select the best dimension L . In the following, we assume that
 431 L_{\max} is large enough such that $\underline{f}^{L_{\max}, L^*} = f^{L^*}$.

432 We work on the quantile q^L introduced in Equation 3, its oracle version q^{L^*} for the level L^* and the
 433 estimation \hat{q}^L . All of them are scalar, in a collection of scalars, with $L = L_{\min}, \dots, L_{\max}$. A natural
 434 criteria to choose the best L is such that the estimator \hat{q}^L minimizes the quadratic error $\mathbb{E}(\|q^{L^*} - \hat{q}^L\|^2)$.
 435 However, this quadratic error is unknown as q^{L^*} is unknown. We can not directly use it.

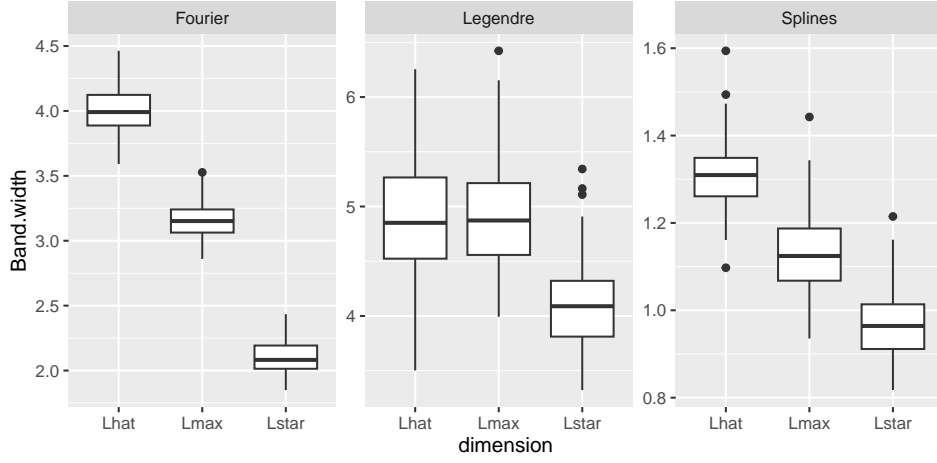


Figure 8: Illustrative example. We display within boxplots the confidence band's width over 100 repetitions for the band CB2 (Lhat), some fixed Lmax and the true (unknown) L^* for Fourier (left), Legendre (middle) and Splines (right) basis.

436 Instead, we study $\|\hat{q}^{L_{\max}} - \hat{q}^L\|^2$. While the theoretical quadratic error $E(\|q^{L^*} - \hat{q}^L\|^2)$ decreases when
 437 $L < L^*$ and increases when $L > L^*$, the approximation $\|\hat{q}^{L_{\max}} - \hat{q}^L\|^2$ of this error is still decreasing
 438 when $L > L^*$, as illustrated in Figure 9.

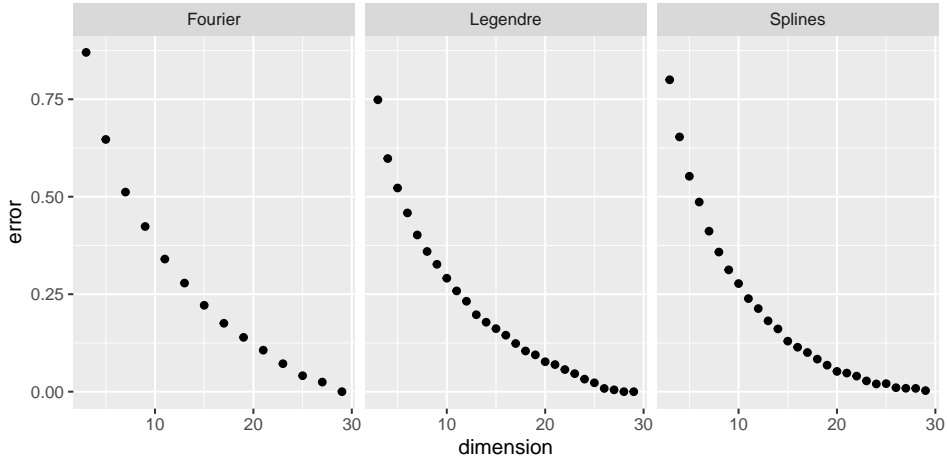


Figure 9: Illustrative example. For a given simulated dataset, we show the behavior of the approximation of the quadratic error of the quantile, as a function of dimension L , for Fourier (left), Legendre (middle) and Splines (right) basis.

439 We recognize a behavior similar to a bias, high when dimension is small, and small when dimension
 440 is large. Selecting a dimension using this criterion will always overfit the data. Thus, we propose to
 441 penalize this quantity by the dimension L divided by the sample size N , as usual in model selection
 442 criterion. For that, we introduce a regularisation parameter $\lambda > 0$ which balances the two terms. A
 443 natural criteria to select the best L is then

444

$$\widetilde{crit}(L) = \|\hat{q}^{L_{\max}} - \hat{q}^L\|^2 + \lambda \frac{L}{N}.$$

445 Then we define

$$\tilde{L} = \arg \min_L \widetilde{crit}(L),$$

446 and take the band centered around $\underline{f}^{\tilde{L}, L^*}$:

$$CB_3(\underline{f}^{L^*}) = CB_1(\underline{f}^{\tilde{L}, L^*})$$

447 In Figure 10, we illustrate the behavior of this selection criterion on simulated data, with $\lambda = 1$ for
 448 the three basis. We can see that \tilde{L} is overestimated. As we work with nested spaces, it ensures that \tilde{L}
 449 has the tendency to be larger than L^* and thus the confidence band is automatically unbiased.

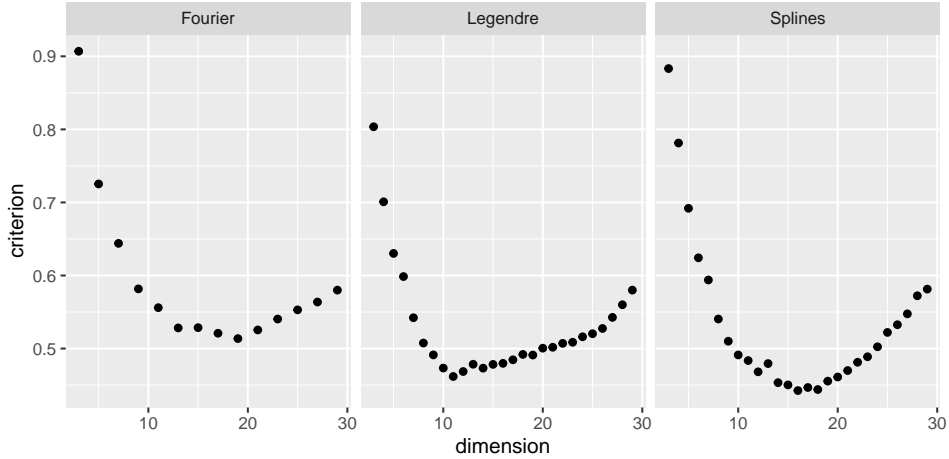


Figure 10: Illustrative example. For a given simulated dataset, we show the behavior of the criteria as a function of dimension L , for Fourier (left), Legendre (middle) and Splines (right) basis.

450 In Figure 11, we test which model is selected over 100 repetitions for the three basis. The estimated
 451 dimension is equal or larger than the true $L^* = 11$. As in Figure 7, being larger is not a problem.
 452 However, we can see that \widetilde{crit} performs better than $crit$. Indeed, the selected dimension \tilde{L} is smaller in
 453 distribution (Figure 11), and closer to the true value than \hat{L} (Figure 7). Moreover, as we then use the
 454 confidence band of Section 3, the confidence level is ensured to be the expected one.

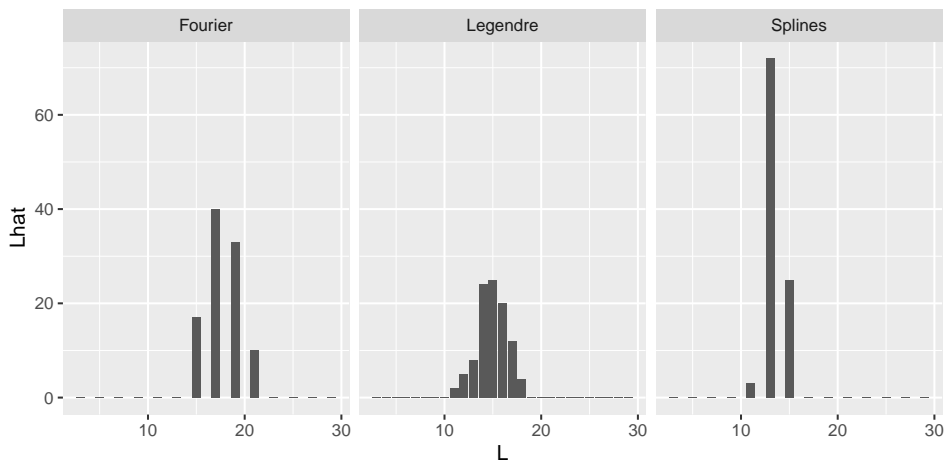


Figure 11: Illustrative example. We show the distribution of the selected model, over 100 repetitions, with the new criteria used to select a model for different basis.

455 We then show in Figure 12 that the width of the selected model is better than the width of the
 456 confidence band with a large level L_{\max} , which one should have used to avoid model selection.

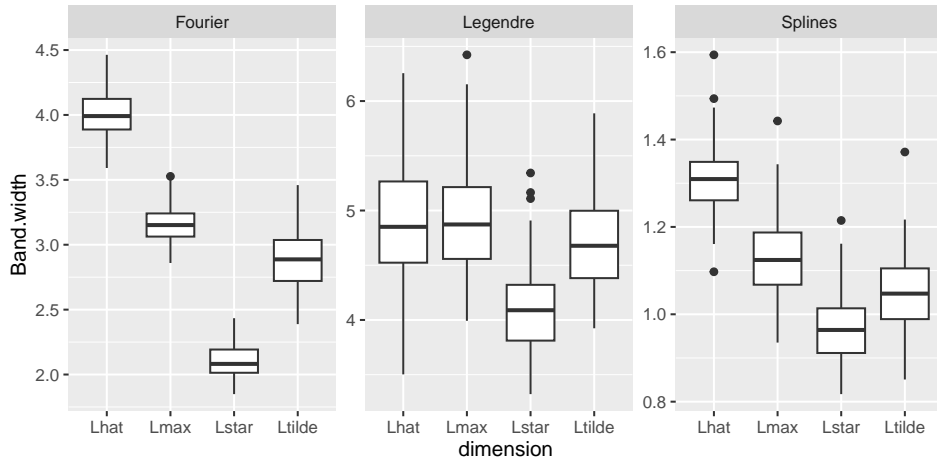


Figure 12: Illustrative example. We display within boxplots the confidence band's width over 100 repetitions for the dimension selected by the several criterion introduced in this paper, for Fourier (left), Legendre (middle) and Splines (right) basis.

6 Real data analysis

In this section, we illustrate the proposed method on the Berkeley Growth Study data. It consists of the heights in centimeters of 39 boys at 31 ages from 1 to 18. We approximate those curves by our 3 basis, namely Legendre, Splines and Fourier. We select the level of each basis using the method introduced in Section 5.

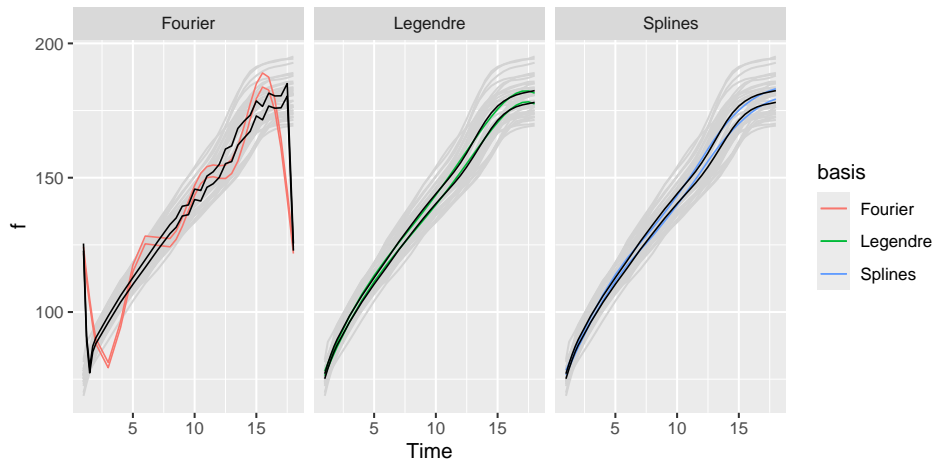


Figure 13: Real data analysis example. We display the confidence bands for Fourier (left), Legendre (middle) and Splines (right) basis on the Berkeley Growth Study data. Black curves correspond to the confidence bands with L_{max} , while colored one are the confidence bands CBE.

In Figure 13, we display the confidence bands associated to Section 3 in black and the one associated to Section 5, for the three basis. As the data is not periodic, the Fourier basis is meaningless, and so is the associated confidence band, whatever the dimension considered. Both splines and Legendre basis give similar confidence bands. When analyzing the width of the bands in Table 3, compared with the one obtained with L_{max} , we see that there are less smooth but also smaller, and from our empirical study we guess that it makes a trade-off between bias and variance.

Table 3: ?(caption)

(a)

	Basis		
	Legendre	Splines	Fourier
Width Lmax	2.124539	2.124536	2.204230
Width selected	1.989604	1.947361	2.078467
Model selected	6.000000	5.000000	7.000000

Real data analysis example, Berkeley Growth Study data. We display the width of the confidence bands for Fourier, Legendre and Splines basis for the confidence band of Section 3 with Lmax and the confidence band of Section 5. We also precise the dimension of the selected model.

7 Conclusion

This paper discusses the construction of confidence bands when considering a linear model over a functional family. Depending on the nature of the family (an orthogonal or orthonormal basis, or just a vector space), theoretical guarantees of the linear estimator are reminded and illustrated. Then, several confidence bands are proposed. Throughout the paper, extensive experimental study on Fourier, Legendre and Spline basis have illustrated the theoretical and methodological proposition, and a real data study is proposed to conclude the paper.

First, when considering a functional family with a fixed dimension, we discuss the confidence band derived from Sun and Loader (1994). It is biased if the dimension is not high enough to approximate well the true function. Then, a new confidence band is proposed that correct this bias. To do so, the bias is estimated and the additional randomness is controlled. A selection criterion is proposed to select the best dimension. Unfortunately, the two kinds of randomness are leading to a wider confidence band, and this result is then not more interesting than the naive one, which consists in taking the largest possible dimension L_{\max} . Finally, a heuristic selection criterion is proposed to select the dimension on the first confidence band, that did not correct the bias. It takes into account the bias as well as the variance, to select a moderate dimension.

The last selection criterion is heuristic, while each term is intuitive. An interesting next step, but out of the scope of this paper, consists of a theoretical study of this criterion. No result, to our knowledge, exist for confidence band with the supremum norm. The euclidean norm is well-studied in general, but is not of interest here, where we want to ensure that the tube is valid as a whole. The supremum norm, on its side, is difficult to study theoretically. A keypoint here also is the randomness of the criterion, that has also to be taken into account, through an oracle inequality for example.

References

- Aneiros, Germán, Silvia Novo, and Philippe Vieu. 2022. “Variable Selection in Functional Regression Models: A Review.” *Journal of Multivariate Analysis* 188: 104871. <https://doi.org/https://doi.org/10.1016/j.jmva.2021.104871>.
- Basna, Rani, Hiba Nassar, and Krzysztof Podgórski. 2022. “Data Driven Orthogonal Basis Selection for Functional Data Analysis.” *Journal of Multivariate Analysis* 189: 104868. <https://doi.org/https://doi.org/10.1016/j.jmva.2021.104868>.
- Bunea, Florentina, Andrada E. Ivanescu, and Marten H. Wegkamp. 2011. “Adaptive Inference for the Mean of a Gaussian Process in Functional Data.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73 (4): 531–58. <https://doi.org/https://doi.org/10.1111/j.1467-9868.2010.00768.x>.

- 501 Claeskens, G., and I. Van Keilegom. 2003. "Bootstrap Confidence Bands for Regression Curves and
502 Their Derivatives." *Ann. Stat.*
- 503 Diquigiovanni, Jacopo, Matteo Fontana, and Simone Vantini. 2022. "Conformal Prediction Bands for
504 Multivariate Functional Data." *Journal of Multivariate Analysis* 189: 104879. <https://doi.org/https://doi.org/10.1016/j.jmva.2021.104879>.
- 505
506 Goepf, V., O. Bouaziz, and G. Nuel. Submitted. "Spline Regression with Automatic Knot Selection,"
507 Submitted.
- 508 Hall, P. 1991. "On Convergence Rates of Suprema." *Probab Theory Related Fields*.
- 509 Hernández, Nicolás, Jairo Cugliari, and Julien Jacques. 2023. "Simultaneous Predictive Bands for
510 Functional Time Series Using Minimum Entropy Sets." <https://arxiv.org/abs/2105.13627>.
- 511 Jacques, Julien, and Sanja Samardžić. 2022. "Analysing Cycling Sensors Data Through Ordinal
512 Logistic Regression with Functional Covariates." *Journal of the Royal Statistical Society Series C:
513 Applied Statistics* 71 (4): 969–86. <https://doi.org/10.1111/rssc.12563>.
- 514 Kokoszka, P., and M. Reimherr. 2017. *Introduction to Functional Data Analysis*. Chapman & Hall /
515 CRC Numerical Analysis and Scientific Computing. CRC Press. [https://books.google.dk/books?
516 id=HlxIvgAACAAJ](https://books.google.dk/books?id=HlxIvgAACAAJ).
- 517 Krivobokova, Tatyana, Thomas Kneib, and Gerda Claeskens. 2010. "Simultaneous Confidence Bands
518 for Penalized Spline Estimators." *Journal of the American Statistical Association* 105 (490): 852–63.
- 519 Lacour, C., P. Massart, and V. Rivoirard. 2017. "Estimator Selection: A New Method with Applications
520 to Kernel Density Estimation." *Sankhya A*.
- 521 Li, Yehua, Yumou Qiu, and Yuhang Xu. 2022. "From Multivariate to Functional Data Analysis:
522 Fundamentals, Recent Developments, and Emerging Areas." *Journal of Multivariate Analysis* 188:
523 104806. <https://doi.org/https://doi.org/10.1016/j.jmva.2021.104806>.
- 524 Liebl, D, and M. Reimherr. 2019. "Fast and Fair Simultaneous Confidence Bands for Functional
525 Parameters."
- 526 Neumann, M., and J. Polzehl. 1998. "Simultaneous Bootstrap Confidence Bands in Nonparametric
527 Regression." *J Nonparametr Statist*.
- 528 Quinton, J-C., E. Devijver, A. Leclercq-Samson, and A. Smeding. 2017. "Functional Mixed Effect
529 Models for Mouse-Tracking Data in Social Cognition." In *ESCON Transfer of Knowledge Conference,
530 Gdansk, Poland*.
- 531 Sachs, Michael C., Adam Brand, and Erin E. Gabriel. 2022. "Confidence Bands in Survival Analysis."
532 *The British Journal of Cancer. Supplement* 127: 1636–41. [https://doi.org/10.1038/s41416-022-
533 01920-5](https://doi.org/10.1038/s41416-022-01920-5).
- 534 Sun, Jiayang, and Clive R. Loader. 1994. "Simultaneous Confidence Bands for Linear Regression and
535 Smoothing." *Ann. Statist.* 22 (3): 1328–45. <https://doi.org/10.1214/aos/1176325631>.
- 536 Telschow, Fabian J. E., Dan Cheng, Pratyush Pranav, and Armin Schwartzman. 2023. "Estimation
537 of expected Euler characteristic curves of nonstationary smooth random fields." *The Annals of
538 Statistics* 51 (5): 2272–97. <https://doi.org/10.1214/23-AOS2337>.
- 539 Telschow, Fabian J. E., and Armin Schwartzman. 2022. "Simultaneous Confidence Bands for Func-
540 tional Data Using the Gaussian Kinematic Formula." *Journal of Statistical Planning and Inference*
541 216: 70–94. <https://doi.org/https://doi.org/10.1016/j.jspi.2021.05.008>.
- 542 Wang, L. And Yang, J. And Gu. 2022. "Oracle-Efficient Estimation for Functional Data Error Dis-
543 tribution with Simultaneous Confidence Band." *Computational Statistics & Data Analysis* 167:
544 107363.
- 545 Xia, Y. 1998. "Bias-Corrected Confidence Bands in Nonparametric Regression." *J.R. Stat. Soc. Ser. B*.
- 546 Zhang, C. And Wu, Y. And Wang. 2020. "Prediction of Working Memory Ability Based on EEG by
547 Functional Data Analysis." *J Neur. Meth.* 333: 108552.
- 548 Zhou, S., X. Shen, and D. A. Wolfe. 1998. "Local Asymptotics for Regression Splines and Confidence
549 Regions." *Ann. Statist.*

550 8 Appendix: proofs

551 8.1 Proof of Proposition 2.3

552 Let us prove the first point. We have

$$\mathbb{E}(\hat{\underline{\mu}}^{L,L^*}) = (\mathbf{B}_L^T \mathbf{B}_L)^{-1} \mathbf{B}_L^T \mathbb{E}(\mathbf{y}) = (\mathbf{B}_L^T \mathbf{B}_L)^{-1} \mathbf{B}_L^T \mathbf{B}_{L^*} \underline{\mu}^{L^*} =: \underline{\mu}^{L,L^*}.$$

553 The theory of the linear model gives that the variance of $\hat{\underline{\mu}}^L$ is equal to $\sigma^2 (\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T \Sigma \mathbf{B} (\mathbf{B}^T \mathbf{B})^{-1}$ with
554 $\Sigma = \text{Diag}(\Sigma_1, \dots, \Sigma_N)$ the $nN \times nN$ covariance matrix of \mathbf{y} . So finally, we have

$$\hat{\underline{\mu}}^{L,L^*} \sim \mathcal{N}(\underline{\mu}^{L,L^*}, \sigma^2 \Sigma_B^{L,L^*}).$$

555 Now we can easily deduce the distribution of $\hat{\underline{f}}^{L,L^*}(t)$, for each $t \in [0, 1]$:

$$\hat{\underline{f}}^{L,L^*}(t) - \mathbf{f}^{L,L^*}(t) \sim \mathcal{N}(0, \sigma^2 \mathbf{B}(t) \Sigma_B^{L,L^*} \mathbf{B}(t)^T).$$

556 8.2 Proof of Theorem 3.2

557 We have

$$P(\forall t \in [0, 1], |\hat{\underline{f}}^{L,L^*}(t) - \mathbf{f}^{L,L^*}(t)| \leq \hat{d}^L(t)) = P(\forall t \in [0, 1], |\hat{\underline{f}}^{L,L^*}(t) - \underline{f}^{L,L^*}(t) + \underline{f}^{L,L^*}(t) - \mathbf{f}^{L,L^*}(t)| \leq \hat{d}^L(t))$$

558 Set Definition 2.1 and Definition 2.3 and a probability $\alpha \in [0, 1]$. Then, we have,

$$\lim_{n \rightarrow +\infty} P(\forall t \in [0, 1], |\hat{\underline{f}}^{L,L^*}(t) - \mathbf{f}^{L,L^*}(t)| \leq \hat{d}^L(t)) = 1 - \alpha$$

559 with $\hat{d}^L(t) = \hat{c}^L \sqrt{\hat{C}_L(t, t)/N}$ and \hat{c}^L defined as the solution of Equation 4.

560 8.3 Proof of Proposition 4.1

561 To simplify the notations, let us denote $a(t) = \underline{f}^{L,L^*}(t) - \underline{f}_1^{L,L^*}(t)$ and $b(t) = \underline{f}^{L_{\max}, L^*}(t) - \underline{f}^{L,L^*}(t) -$
562 $(\underline{f}_2^{L_{\max}, L^*}(t) - \underline{f}_2^{L,L^*}(t))$. We have

$$\begin{aligned} P(\exists t | a(t) + b(t) | \geq \hat{d}_1^L(t) + \hat{d}_2^{L, L_{\max}}(t)) &\leq P(\exists t | a(t) | + | b(t) | \geq \hat{d}_1^L(t) + \hat{d}_2^{L, L_{\max}}(t)) \\ &= P(\exists t | a(t) | \geq \hat{d}_1^L(t)) P(\exists t | b(t) | \geq \hat{d}_2^{L, L_{\max}}(t)) = \alpha \beta. \end{aligned}$$

563 The last equality holds thanks to the independence of the two sub-samples.