Overview

Simultaneous confidence bands for the mean of functional data

David Degras*

The mean function is a central object of inquiry in the analysis of functional data. Typical questions related to the mean function include quantifying estimation uncertainty, testing parametric models, and making comparisons between populations. To make probabilistic statements about the mean function over its entire domain, rather than at a single location, it is necessary to infer all of its values simultaneously. Pointwise inference is not appropriate for this task and indeed produces anticonservative results, i.e., the coverage level of confidence regions is too low and the significance level of hypothesis tests too high. In contrast, simultaneous confidence bands (SCB) provide a flexible framework for conducting simultaneous inference on the mean function and other functional parameters. They also offer powerful visualization tools for communicating analytic results to interdisciplinary audiences. The construction of SCB in the context of functional data requires specific theory and methods. In particular, it is not addressed by the nonparametric regression literature. Although software is available to perform individual steps of an SCB procedure, resources that provide end-to-end computations are scarce. Applications of SCB to one- and two-sample inferences are illustrated here with the R package SCBmeanfd. © 2017 The Authors. *WIREs Computational Statistics* published by Wiley Periodicals, Inc.

How to cite this article:
WIREs Comput Stat 2017, 9:e1397. doi: 10.1002/wics.1397

Keywords: Functional data, simultaneous confidence bands, mean function, nonparametric smoothing, bootstrap, hypothesis test, R package

INTRODUCTION

Functional Data

Technological advances in science and industry have enabled the routine collection of high-resolution data over time, space, frequency domains, and so on. Such data can be thought of as discretely observed functions (of time, space, etc.) and are thus called *functional data*. Functional data are ubiquitous in fields as diverse as climate science,[1] finance,[2] medical imaging,[3] bioinformatics,[4] and more. Although they may appear as simple instances of say, multivariate, time series, or image data, functional data present a unique combination of features: high dimensionality, structure, and smoothness that require new theory and methods.[5] This has motivated the development of *functional data analysis* (FDA), a recent and highly active area of statistical research.[6] Important FDA tasks include mean function estimation, functional principal component analysis (FPCA), functional regression and classification, differential analysis, and curve registration. Mathematically, functional data are represented as discrete observations of a latent stochastic process or random function defined over a continuous domain, say D. This domain, e.g., a time interval or frequency range, displays structure such as ordering or metricity. (Typically, D is a compact subset of $\mathbb{R}^d$ for some $d \geq 1$ such as $[0, 1]^d$ or a sphere.) The latent

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: ddegrasv@gmail.com

Department of Mathematics, University of Massachusetts Boston, Boston, MA, USA

Conflict of interest: The author has declared no conflicts of interest for this article.

stochastic process, say X, is assumed to have finite variance. Its mean function μ and covariance function Γ are supposed to be smooth, but no parametric shape is imposed to these functions. Multiple realizations of X are observed, either on a fine grid (dense functional data) or at a few random locations in D (sparse functional data). Figure 1 shows an example of (dense) functional data arising in the study of multiple sclerosis (MS) by brain tractography. These data will be examined in the application section of the article.

Pointwise versus Simultaneous Inference

When conducting statistical inference with functional data, one most often wants to assess functional parameters over their entire domain D (or a subregion thereof) rather than at a single point $t_0 \in D$. For this purpose, it is necessary to *simultaneously* assess all the values of the functional parameter, say θ, over D. If one *separately* infers $\theta(t)$ at the confidence level $1 - \alpha \in (0, 1)$ for each $t \in D$, in general, these inferences will not *collectively* attain the confidence level $1 - \alpha$. Consider for example the construction of simultaneous confidence bands (SCB) for the mean function μ of the random process X. Given discrete and possibly noisy observations of X, the objective is to determine functional statistics L_s and U_s such that the band $\{[L_s(t), U_s(t)] : t \in D\}$ fully contains the mean function μ with probability $1 - \alpha$:

$$P(\mu(t) \in [L_s(t), U_s(t)],\ \forall t \in D) = 1 - \alpha. \qquad (1)$$

Now, a method that builds pointwise confidence intervals $[L_p(t), U_p(t)]$ of level $1 - \alpha$ for each $\mu(t), t \in D$, and then aggregates them into a pointwise confidence band (PCB) $\{[L_p(t), U_p(t)] : t \in D\}$ will fail to satisfy (1). (Subscripts s and p have been added to the lower bounds $L(t)$ and upper bounds $U(t)$ to distinguish SCB from PCB.) Indeed, separately building multiple confidence intervals for each $\mu(t), t \in D$, increases the probability of obtaining *at least one* invalid inference beyond the level α. This implies that the simultaneous coverage level of a PCB is usually (far) less than $1 - \alpha$. In other words, PCB may give the unaware practitioner a largely false sense of confidence in making global inferences about μ and thus lead to erroneous conclusions. Figure 2 illustrates the fact that *pointwise methods are not appropriate to simultaneously infer functional parameters*. Of course PCB are perfectly valid inferential tools as long as they are used for their original purpose which is to infer μ at a single point $t_0 \in D$.

A simple idea for correcting the deficiency of PCB with respect to simultaneous coverage is to increase their pointwise coverage. In principle, there exists a level $1 - \alpha' > 1 - \alpha$ such that a PCB of pointwise coverage level $1 - \alpha'$ is also an SCB of simultaneous coverage level $1 - \alpha$ or more. However, the determination of the corrected level $1 - \alpha'$ is not analytically tractable. In particular, multiple comparison methods for multivariate data (e.g., Bonferroni correction, Scheffé procedure) are of no use in the functional data setup because the infinite cardinality of D

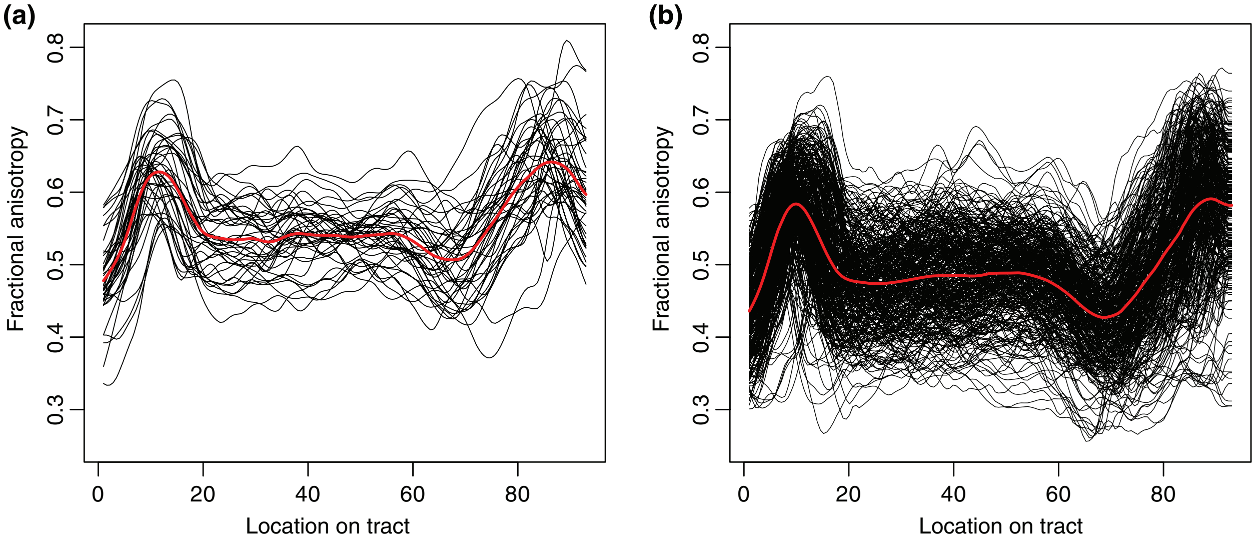

FIGURE 1 | Diffusion tensor imaging (DTI) data of 42 control subjects (left) and 334 multiple sclerosis patients (right). The x-axis indicates location along a brain tract in the corpus callosum (arbitrary units). The y-axis displays fractional anisotropy (FA), which measures water diffusivity and ranges in [0, 1]. Low FA values indicate possible neurological damage. The average profile (red line) is consistently lower in the patient group than in the control group. The same observation holds for the standard deviation function.

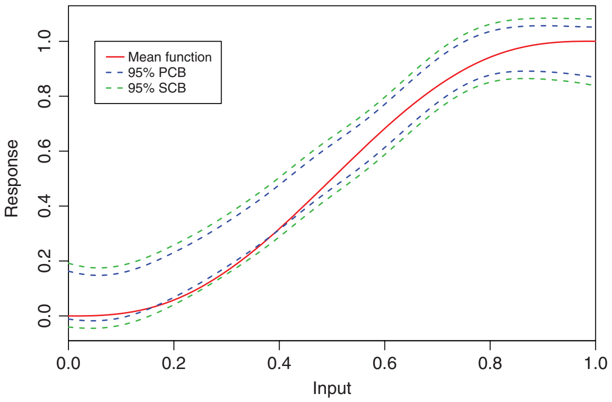

FIGURE 2 | Comparison between pointwise confidence bands (PCBs) and simultaneous confidence bands (SCB) using artificial data. Here, X is a Gaussian process with mean function $\mu(t) = 10t^3 - 15t^4 + 6t^5$ and covariance $\Gamma(s, t) = (0.1)^2\exp(-2|s - t|)$; see the Supporting information for simulation details. The SCB covers μ whereas the PCB is too narrow and fails to cover it. For a nominal coverage level of 95% (in the pointwise sense for PCB, simultaneous sense for SCB), the observed simultaneous coverage level is 74.1 % for PCB and 92.8 % for SCB; the observed pointwise coverage level (averaged over D) is 94.2 % for PCB and 98.7 % for SCB.

would lead to unbounded confidence regions.[7] It is also noteworthy that despite the strong connection between mean function estimation with functional data and nonparametric regression, the SCB methods of the latter do not apply to the former because of their entirely different frameworks. (Nonparametric regression assumes a discrete-time stochastic structure with little or no temporal dependence whereas FDA posits a continuous-time process with strong temporal dependence.) This motivates the development of specific SCB methods for functional data.

Simultaneous Inference with Functional Data

The simultaneous inference of the mean function is a well studied topic in FDA. Several SCB methods and statistical tests are available for this task, especially in the prevalent case of dense functional data which we focus on in this article. The practical construction of SCB involves estimating the mean function μ and the covariance function Γ of the data-generating process X. The mean function estimator $\hat{\mu}$ can be derived with any suitable nonparametric smoothing technique, e.g., spline smoothing,[8] local polynomial smoothing,[9] or basis function approaches.[6,10] Data-driven methods such as cross validation[11,12] and plug-in estimation[13,14] are useful to determine how much to smooth the data in this task. Standard covariance estimation methods include regularizing the sample covariance function, performing its functional PCA (FPCA), or using multivariate smoothing methods.[15] A large sample approximation to the distribution of the (scaled, centered) estimator $\sqrt{n}(\hat{\mu}-\mu)$ is then used to build the SCB, where n is the sample size. This approximation can be obtained numerically with resampling techniques such as parametric or nonparametric bootstrap[9,16] or analytically using e.g., Gaussian random field theory[17] or Karhunen–Loève expansions.[7] Alongside dense functional data, SCB have been studied in the context of sparse functional data[18,19] and survey sampling.[20,21]

In addition to their usefulness for assessing estimation uncertainty or building parametric models for μ, SCB find rich applications to hypothesis testing, e.g., in specification tests, two-sample equality tests, and goodness-of-fit tests.[8,9,22] Bayesian methods for mean function estimation,[23] functional analysis of variance,[1] and functional mixed models[24,25] can also be used to produce simultaneous credible bands[26,27] for mean functions. Other methods based e.g., on covariance regularization,[28] generalizations of multivariate tests,[29,30] or FPCA[31] have been employed to simultaneously infer mean functions. In comparison to the latter, statistical tests based on SCB present several advantages: (1) they offer powerful visualization tools, (2) they can easily be interpreted even by nonstatistical experts, and (3) they have localization power in the sense that they indicate where and to what extent a statistical hypothesis holds, or is violated.

ESTIMATING THE MEAN AND COVARIANCE OF FUNCTIONAL DATA

In the analysis of functional data, the latent stochastic process X is conveniently decomposed as the sum of its mean function $\mu = EX$ and a zero mean process $Z = X - \mu$:

$$X(t) = \mu(t) + Z(t), \quad t \in D. \tag{2}$$

The collected data consist in discrete and possibly noisy observations of realizations $X_1, \ldots, X_n$ of X. A general model for these data is

$$Y_{ij} = X_i(t_{ij}) + \varepsilon_{ij}, \quad i = 1, \ldots, n, \ j = 1, \ldots, p_i, \tag{3}$$

where Y_{ij} is the observation of the ith subject/statistical unit at location t_{ij} and ε_{ij} is a measurement error. The X_i are typically assumed to be mutually independent and independent of the ε_{ij}; the ε_{ij} are usually assumed to be independent across units (i) but not necessarily within. In the frequent case of dense functional data, the X_i are all observed on a fine grid of

TABLE 1 | Types of Designs for Functional Data

Design	Data	Details
Dense	$Y_i(t_j)$, $i = 1, \ldots, n$, $j = 1, \ldots, p$	p large; t_j, p fixed
Sparse	$Y_i(t_{ij})$, $i = 1, \ldots, n$, $j = 1, \ldots, p_i$	p_i small; t_{ij}, p_i random

deterministic points in D, say $t_1, \ldots, t_p$. In the less frequent case of sparse functional data, the observation points t_{ij} are random and each X_i is only observed a few times. Table 1 summarizes the differences between these observational designs. *For reasons of space, we focus on the study of dense functional data in this article and only devote a brief section to sparse functional data.*

Mean Function Estimation

In an ideal situation, the best estimator of the mean function μ would be the sample mean of $X_1, \ldots, X_n$, i.e., $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. This oracle estimator is however not feasible due to the finite size of the observation grid $\{t_j : j = 1, \ldots, p\}$ and to the measurement errors (noisy observations Y instead of X). A reasonable approach is to compute the sample mean $\overline{Y}_j = (1/n)\sum_{i=1}^{n} Y_{ij}$ at each grid point t_j and smooth these average values in a nonparametric fashion. The benefits of smoothing the data are twofold: first, it yields a well-defined estimator $\hat{\mu}$ over the entire domain D; second, it preserves the features of the smooth function μ while reducing the effect of the noise ε.

Nonparametric Smoothing

The key idea of nonparametric smoothing is to locally average data. In the estimation of μ, the usual nonparametric smoothers can be written as

$$\hat{\mu}(t) = \sum_{j=1}^{p} w_j(t)\,\overline{Y}_j \qquad (4)$$

where the w_j are weight functions that depend on the data and the smoothing method.

A classic example of nonparametric smoother is the Nadaraya–Watson estimator[32]

$$\hat{\mu}(t) = \frac{\sum_{j=1}^{p} K\left(\frac{t_j - t}{h}\right)\overline{Y}_j}{\sum_{j=1}^{p} K\left(\frac{t_j - t}{h}\right)},$$

where K is a kernel function and $h > 0$ is a bandwidth (K and h must be chosen by the user). Here, the weight functions are $w_j(t) = K\left(\frac{t_j - t}{h}\right) / \sum_{k=1}^{p} K\left(\frac{t_k - t}{h}\right)$; they satisfy $\sum_{j=1}^{p} w_j(t) = 1$ for all $t \in D$. The kernel function K is typically a unimodal function with mode at zero, e.g., the normal density function $K(t) = (2\pi)^{-1/2}\exp(-t^2/2)$ or the triangular function $K(t) = \max(0, 1 - |t|)$. The bandwidth h determines the effective size of the local averaging window. For a given location t, small values of h give large weights $w_j(t)$ to observations $\overline{Y}_j$ for which t_j is close to t. Conversely, larger values of h give more relative importance to distant observations. For fixed h and t_j, the weight $w_j(t)$ decreases as t gets farther from t_j.

Nonparametric smoothing techniques can be broadly divided in three categories: penalization methods, e.g., smoothing splines[33] and P-splines[34]; basis function methods, e.g., B-splines,[35] Fourier basis,[6] polynomials, and wavelets[36]; and local smoothing methods, e.g., kernel smoothing[32] and local polynomial modeling.[37] Table 2 shows examples of objective functions used to define popular estimators. The choice of a nonparametric smoothing technique for a given application should be guided by the characteristics of the problem at hand but also by user preference. It is widely agreed that the choice of a smoothing technique has far less influence on statistical results than the selection of smoothing parameters (e.g., h for the Nadaraya–Watson estimator), which directly determines how much smoothing to apply to the data.

Smoothing Parameter Selection

For a given $t \in D$, a sensible measure of the accuracy of $\hat{\mu}(t)$ is the mean squared error

$$\mathrm{MSE} = E\{\hat{\mu}(t) - \mu(t)\}^2.$$

As a reminder, the MSE can be decomposed as squared bias plus variance:

$$\mathrm{MSE} = \underbrace{\{E\hat{\mu}(t) - \mu(t)\}^2}_{\mathrm{Bias}^2} + \underbrace{E\{\hat{\mu}(t) - E\hat{\mu}(t)\}^2}_{\mathrm{Variance}}.$$

Hence, an estimator with small MSE has both small bias and small variance, two desirable statistical properties. To fix notations, let's call h the smoothing parameter used to define $\hat{\mu}(t)$. Typically, the same value h is used for each $t \in D$ although it can be made location-dependent: $h = h(t)$. The smoothing parameter h expresses a trade-off between bias and variance in the estimation, or between fidelity to the

TABLE 2 | Examples of Nonparametric Smoothing Techniques When $D \subset \mathbb{R}$

Method	Definition of estimator
Smoothing spline	$\min_{g \in W^{2,m}(D)} \frac{1}{p} \sum_{j=1}^{p} \left\{ \overline{Y}_j - g(t_j) \right\}^2 + \lambda \int_D \{ g^{(m)}(t) \}^2 dt$
Basis functions	$\min_{\theta_1, \ldots, \theta_L} \frac{1}{p} \sum_{j=1}^{p} \left\{ \overline{Y}_j - \sum_{l=1}^{L} \theta_l B_l(t) \right\}^2$
P-spline	$\min_{\theta_1, \ldots, \theta_L} \frac{1}{p} \sum_{j=1}^{p} \left\{ \overline{Y}_j - \sum_{l=1}^{L} \theta_l B_l(t) \right\}^2 + \lambda \int_D \left\{ \sum_{l=1}^{L} B_l^{(m)}(t) \right\}^2 dt$
Local polynomial	$\min_{\beta_0, \ldots, \beta_m} \frac{1}{ph} \sum_{j=1}^{p} \left\{ \overline{Y}_j - \sum_{k=0}^{m} \beta_k (t_j - t)^k \right\}^2 K\left(\frac{t_j - t}{h} \right)$

Notations: $W^{2,m}(D) = \{g : g^{(0)}, \ldots, g^{(m-1)} \text{abs.continuous}; g^{(m)} \in L_2(D)\}$: Sobolev space; $\lambda > 0$: smoothing parameter; $(B_1, \ldots, B_\kappa)$ arbitrary basis set in general or B-spline basis for P-spline smoothing; K: kernel function; $h > 0$: bandwidth.

data and stability/interpretability of the estimator. More smoothing (larger values of h) reduces the variance but increases the bias and conversely, less smoothing (smaller values of h) reduces the bias but increases the variance. One may average the MSE across the observation points t_j, $j = 1, \ldots, p$, or integrate it over D to obtain a global accuracy measure for $\hat{\mu}$. For simplicity, let's consider the average $\text{AMSE}(h) = \frac{1}{p} \sum_{j=1}^{p} E\{\hat{\mu}_h(t_j) - \mu(t_j)\}^2$, where dependence on h is now explicitly indicated. The goal of smoothing parameter selection methods is to determine from the data values of h that approximately minimize the AMSE.

Cross validation is a celebrated method for this task. In its standard version, cross validation consists in finding the smoothing parameter h for which the estimator based on all data points but one best predicts the remaining value. Applied to model (3), this type of cross validation ('leave-one-point-out' cross validation) would lead to minimize the score $\frac{1}{np} \sum_{i=1}^{n} \sum_{j=1}^{p} \left(\hat{\mu}_h^{-j}(t_j) - Y_{ij} \right)^2$ with respect to h, where $\hat{\mu}_h^{-j}$ is the estimator based on all data points but the $(t_j, Y_{ij}), i = 1, \ldots, n$. In the functional data setup, however, statistical units are entire *curves* or functions rather than single observation points. For this reason, it is more natural to minimize the leave-one-curve-out cross-validation score[11]

$$\text{CV}(h) = \frac{1}{np} \sum_{i=1}^{n} \sum_{j=1}^{p} \left(\hat{\mu}_h^{-i}(t_j) - Y_{ij} \right)^2, \qquad (5)$$

where $\hat{\mu}_h^{-i}$ is the estimator based on all data curves but the ith. It is known that $E\{CV(h)\} \approx \text{AMSE}(h) + c$, where c is a constant that does not depend on h. That is, the expected cross-validation score and the AMSE are minimized by the same value h. In addition, the cross-validation smoothing parameter $\hat{h}_{cv}$ is asymptotically optimal as $n, p \to \infty$.[12]

Another classical procedure for selecting smoothing parameters is *plug-in estimation*. The principle of this method is to: (1) analytically derive an asymptotic expression for the error measure, say AMSE(h), as $n, p \to \infty$ and deduce the value h_{opt} that minimizes this expression, (2) estimate all unknown quantities that h_{opt} may depend on (typically, derivatives of μ and Γ) from the data, and (3) replace these quantities by their estimates to obtain the plug-in smoothing parameter $\hat{h}_{\text{plug}}$. Plug-in estimation has been studied in the local polynomial estimation of μ and its derivatives based on functional data,[14] where its numerical performances were comparable or slightly inferior to cross validation.

In the case of penalized estimation methods such as smoothing splines or P-splines (see Table 2 for a definition of these estimators), it is often possible to reformulate the estimation problem in terms of a mixed linear model.[38] The smoothing parameter can then be estimated as a variance component by (restricted) maximum likelihood or the Expectation-Maximization (EM) algorithm. Note, however, that this approach requires a good pilot estimator of the data covariance.

Covariance Estimation

After obtaining a pointwise estimator $\hat{\mu}$ of the mean function μ, the next step in constructing SCB is to assess the covariance structure of $\hat{\mu}$. The estimator covariance is closely related to the covariance Γ of the data-generating process X. Indeed, for all $s, t \in D$ and all usual nonparametric estimators $\hat{\mu}$, it holds that[13,39]

$$\text{Cov}(\hat{\mu}(s), \hat{\mu}(t)) \approx \frac{\Gamma(s,t)}{n} \qquad (6)$$

as n, $p \to \infty$. Hence, evaluating the estimator covariance reduces to estimating Γ.

A common way to estimate Γ is to first apply to each (discretized) individual curve $(Y_{ij})_{j=1,\ldots,p}$ (for $i = 1, \ldots, n$) the same smoothing technique that was applied to the average curve $(\overline{Y}_j)_{j=1,\ldots,p}$ to obtain the estimator $\hat{\mu}$. This operation is called *presmoothing* and can be interpreted as an attempt to recover the curves X_i underlying the data. The presmoothed data express as

$$\hat{X}_i(t) = \sum_{j=1}^{p} w_j(t) Y_{ij}, \qquad (7)$$

where the w_j are the weight functions of (4). The estimator $\hat{\mu}$ can thus be interpreted as the average of the presmoothed data: $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} \hat{X}_i$. The sample covariance of the $\hat{X}_i$ is then employed to estimate the covariance Γ:

$$\hat{\Gamma}(s,t) = \frac{1}{n-1}\sum_{i=1}^{n}\left\{\hat{X}_i(s) - \hat{\mu}(s)\right\}\left\{\hat{X}_i(t) - \hat{\mu}(t)\right\}. \qquad (8)$$

One may alternatively utilize FPCA[6] to obtain a low-rank estimator $\hat{\Gamma}$. In a nutshell, FPCA calculates the eigenvalues $\hat{\lambda}_1 \ge \cdots \ge \hat{\lambda}_n \ge 0$ and associated eigenfunctions $\hat{\varphi}_1, \ldots, \hat{\varphi}_n$ of the sample covariance (8). The $\hat{\varphi}_k$ are orthonormal ($\int_D \hat{\varphi}_k \hat{\varphi}_l = \delta_{kl}$ with δ_{kl} the Kronecker symbol) and represent the successive directions of maximal variation among the $\hat{X}_i$. By Mercer's theorem (see e.g., Ref 40, Chap. 4) the sample covariance (8) expands as $\sum_{k=1}^{n} \hat{\lambda}_k \hat{\varphi}_k(s) \hat{\varphi}_k(t)$. With a suitable truncation order κ, the low-rank approximation $\hat{\Gamma}(s,t) = \sum_{k=1}^{\kappa} \hat{\lambda}_k \hat{\varphi}_k(s) \hat{\varphi}_k(t)$ preserves most of the variations of (8) while saving substantial storage space ($O(\kappa d)$ versus $O(d^2)$ with usually $\kappa \ll min$ (n, d)). More importantly, this covariance estimator provides an efficient way to simulate the limiting Gaussian distribution of $\sqrt{n}(\hat{\mu} - \mu)$ in view of SCB.[9,22,41]

A third way to estimate Γ is via the decomposition $\Gamma(s, t) = E(X(s)X(t)) - \mu(s)\mu(t)$. Having already estimated μ, it suffices to estimate $\phi(s, t) = E(X(s)X(t))$ by smoothing the data $(t_{ijj'}, C_{ijj'})$ for $1 \le i \le n$, $1 \le j \ne j' \le p_i$, where $t_{ijj'} = (t_{ij}, t_{ij'})$ and $C_{ijj'} = Y_{ij} Y_{ij'}$.[15] Although the resulting estimator $\hat{\Gamma}(s,t) = \hat{\phi}(s,t) - \hat{\mu}(s)\hat{\mu}(t)$ is less biased than the first two (sample covariance and FPCA), it can become computationally expensive for large datasets and is in general *not* positive semidefinite. This may prove problematic in the construction of SCB because a proper, i.e., positive semidefinite covariance function is required to simulate the limiting Gaussian process associated to $\hat{\mu}$.

Figure 3 shows covariance and correlation estimates for the brain tractography data introduced in Figure 1 and analyzed in the application section. For these data, there is virtually no difference between the sample covariance function (8) and the low-rank covariance estimate based on FPCA.

BUILDING SIMULTANEOUS CONFIDENCE BANDS

Under regularity conditions on the process X, the observation points t_j, and the noise ε in (3), and under joint conditions on the rates of the sample size n, grid p, and smoothing parameter h, the estimator $\hat{\mu}$ satisfies a functional central limit theorem

$$\sqrt{n}(\hat{\mu} - \mu) \to \mathcal{G}(0, \Gamma) \qquad (9)$$

as n, $p \to \infty$, where $\mathcal{G}(0, \Gamma)$ denotes a Gaussian process with mean zero and covariance Γ and where convergence takes place in the space of continuous functions on D equipped with the supremum norm.[8,9]

The functional limit theorem (9), along with results of uniform convergence for $\hat{\Gamma}$, forms the cornerstone upon which SCB are built for μ. For example, let $\sigma(t) = \Gamma(t, t)^{1/2}$ and $\rho(s,t) = \frac{\Gamma(s,t)}{\sigma(s)\sigma(t)}$ be the standard deviation and correlation functions of X, respectively, and let $Z \sim \mathcal{G}(0, \rho)$ be a Gaussian process. Assume that for a given confidence level $1 - \alpha$, we can determine the quantile $z_{\alpha,\rho}$ such that $P\left(\sup_{t \in D} |Z(t)| \le z_{\alpha,\rho}\right) = 1 - \alpha$. As for pointwise confidence intervals based on the normal distribution, the confidence region

$$\left\{\left[\hat{\mu}(t) - z_{\alpha,\rho}\frac{\sigma(t)}{\sqrt{n}}, \hat{\mu}(t) + z_{\alpha,\rho}\frac{\sigma(t)}{\sqrt{n}}\right] : t \in D\right\} \qquad (10)$$

is an approximate SCB of level $1 - \alpha$ for μ. That is, for large n and p,

$$P\left(\mu(t) \in \left[\hat{\mu}(t) \pm z_{\alpha,\rho}\frac{\sigma(t)}{\sqrt{n}}\right], \forall t \in D\right) \approx 1 - \alpha. \qquad (11)$$

Note that one could also build SCB based on the limit distribution $\mathcal{G}(0, \Gamma)$ rather than its standardized version $\mathcal{G}(0, \rho)$. Letting $z_{\alpha,\Gamma}$ be such that

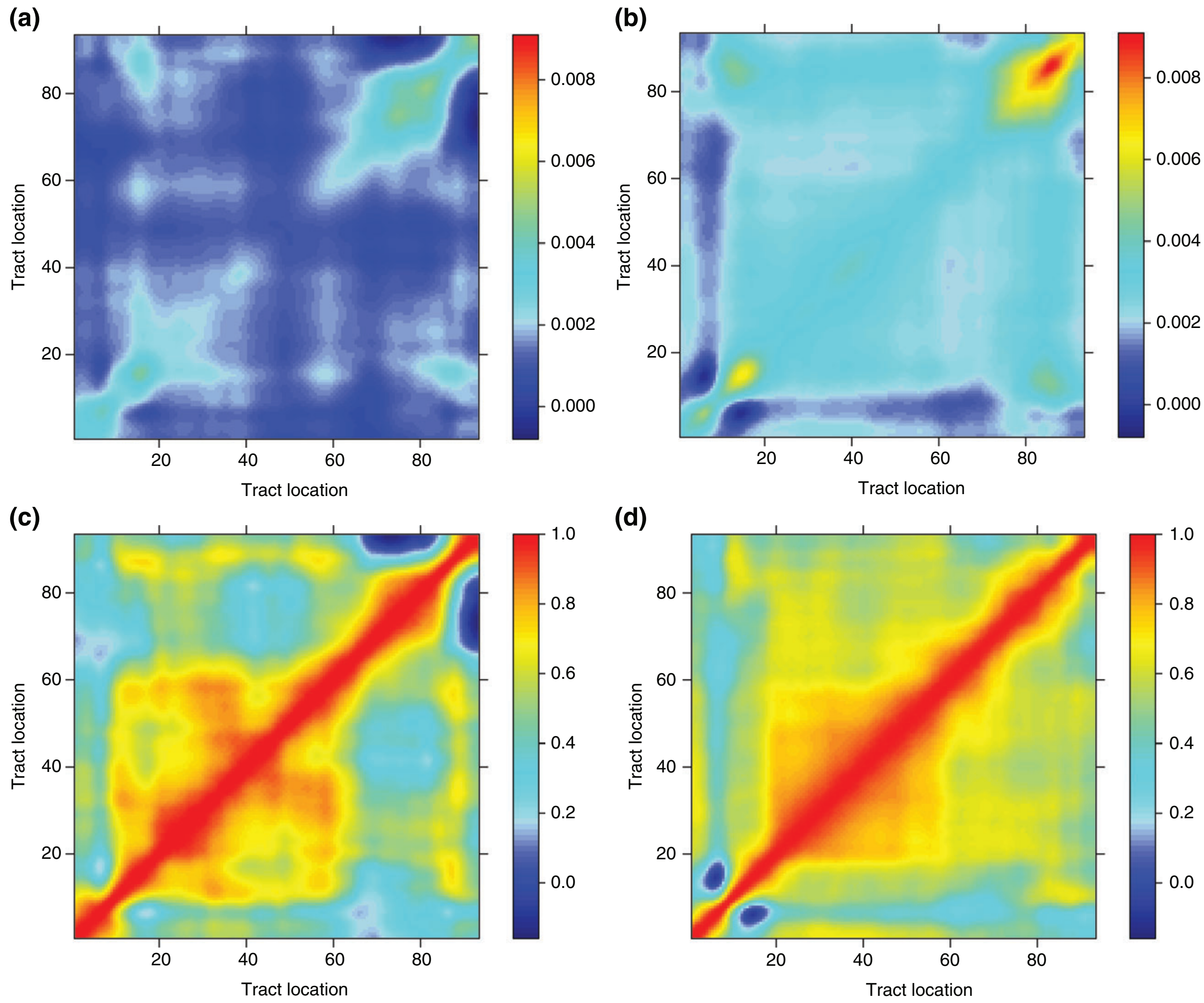

FIGURE 3 | Estimated covariance (top) and correlation (bottom) functions for the brain tractography data of 42 control subjects (left) and 334 multiple sclerosis patients (right). After presmoothing the data by local linear smoothing (bandwidth h = 0.98 for the controls and h = 0.52 for the patients selected by leave-one-curve-out cross validation), the population covariance (resp. correlation) function of each group was estimated by the corresponding sample covariance (resp. correlation) function. The estimated covariance functions display similar patterns in both groups although the scale (variance) is much smaller in the control group. The estimated correlation functions have similar global features in both groups but in the control group, it is rougher and closer to zero away from the diagonal.

$P\left(\sup_{t\in D}|\sigma(t)Z(t)|\le z_{\alpha,\Gamma}\right)=1-\alpha$ where $\sigma Z\sim\mathcal{G}(0,\Gamma)$, the corresponding SCB of approximate level $1-\alpha$ would be $\left\{\left[\hat{\mu}(t)\pm n^{-1/2}z_{\alpha,\Gamma}\right]:t\in D\right\}$. The latter (covariance-based) SCB has uniform width over D whereas (10) tracks more adaptively the estimation uncertainty around $\mu(t), t\in D$.

In practice, the unknown standard deviation $\sigma(t)$ is replaced by its estimator $\hat{\sigma}(t)=\hat{\Gamma}(t,t)^{1/2}$ in (10). The quantile $z_{\alpha,\rho}$, which cannot be obtained analytically, is approximated by numerical simulation of $\mathcal{G}(0,\hat{\rho})$ conditionally on $\hat{\rho}$. The resulting SCB retains approximate coverage level $1-\alpha$:

$$P\left(\mu(t)\in\left[\hat{\mu}(t)\pm z_{\alpha,\hat{\rho}}\frac{\hat{\sigma}(t)}{\sqrt{n}}\right],\forall t\in D\right)\approx 1-\alpha. \qquad (12)$$

Note that the stronger the correlation structure ρ, the smaller the quantile $z_{\alpha,\rho}$. In the limit case of a perfectly correlated limit process ($\rho(s,t)=1$ for all $s,t\in D$), $z_{\alpha,\rho}$ is exactly the quantile of level $1-\alpha/2$ of the standard normal distribution, e.g., 1.96 for $1-\alpha=0.95$. At the other end of the spectrum, if $\rho(s,t)\longrightarrow\delta(s,t)$ where $\delta(s,t)=1$ if $s=t$ and $\delta(s,t)=0$ otherwise (Gaussian white noise), then $z_{\alpha,\rho}\longrightarrow\infty$.

The determination of the quantile $z_{\alpha,\hat{\rho}}$ by Gaussian process simulation can be viewed as a parametric bootstrap of the standardized estimator $\frac{\sqrt{n}}{\hat{\sigma}(t)}(\hat{\mu}(t)-\mu(t)), t \in D$. It proceeds as follows: first discretize the process $Z \sim \mathcal{G}(0,\hat{\rho})$ over a fine grid $\tau = \{\tau_1,\dots,\tau_m\} \subset D$ and call Z_m the resulting random vector. In principle, the grid size m should be taken as large as possible to render the effect of discretization negligible. However, larger values of m incur higher computational costs, especially given that Z_m must be simulated many times. In practice, sensible values of m should increase with the number p of observation points, the size of D, and the roughness of $\hat{\rho}$. When $D = [0,1]$, a choice of m between 100 and 500 is suitable in most cases. The next step is to simulate $Z_m \sim N\left(0, M_{\hat{\rho}}\right)$where $M_{\hat{\rho}} = (\hat{\rho}(\tau_j,\tau_k))_{1 \le j,k \le m}$. The direct simulation of Z_m has a memory cost $O(m^2)$ for storing $M_{\hat{\rho}}$ and computation cost $O(m^3)$ (the eigenvalue or Cholesky decomposition of $M_{\hat{\rho}}$ is required), which is impractical if m is large. However, if $\hat{\Gamma}$ (or equivalently $\hat{\rho}$) can be written as a low-rank tensor of basis functions $\varphi_1, \dots, \varphi_\kappa$: $\hat{\Gamma}(s,t) = \sum_{k=1}^{\kappa} \sum_{l=1}^{\kappa} \hat{\lambda}_{kl} \varphi_k(s) \varphi_l(t)$ with $\kappa \ll m$, the memory cost and computation cost both dramatically decrease to $O(\kappa m)$. This low-rank decomposition is obtained, e.g., when $\hat{\Gamma}$ is based on the FPCA of the smoothed data $\hat{X}_i$ (here the φ_k are the first few eigenfunctions of the sample covariance function (8) with associated eigenvalues $\hat{\lambda}_{kk}$ and $\hat{\lambda}_{kl} = 0$ for $k \neq l$) or when the data Y_{ij} have been expanded in a basis of functions $\varphi_1, \dots, \varphi_\kappa$, say, B-spline or wavelets: $\hat{X}_i(t) = \sum_{k=1}^{\kappa} \hat{x}_{ik} \varphi_k(t)$, and $\hat{\Gamma}$ is the sample covariance (8) (here the $\hat{\lambda}_{kl}$ are the sample covariances of the basis coefficients $\hat{x}_{1k},\dots,\hat{x}_{nk}$, and $\hat{x}_{1l},\dots,\hat{x}_{nl}$). To simulate $Z_m \sim N\left(0, M_{\hat{\rho}}\right)$ in this low-dimensional setting, it suffices to simulate a random vector $Z_\kappa \sim N\left(0, M_{\hat{\lambda}}\right)$, where $M_{\hat{\lambda}} = \left(\hat{\lambda}_{kl} / \left(\hat{\lambda}_{kk} \hat{\lambda}_{ll}\right)^{1/2}\right)$ has low-dimension $\kappa \times \kappa$, and premultiply Z_κ by the $m \times \kappa$ matrix $\Phi = (\varphi_k(\tau_j))$. After simulating Z_m, one computes its ℓ_∞ norm, i.e., the maximum of the absolute values of its entries. The same process is repeated many times, say $N = 10^4$ times, and $z_{\alpha,\hat{\rho}}$ is taken as the quantile of level $1 - \alpha$ of the N simulated realizations of $\|Z_m\|_\infty$. Note that the above parametric bootstrap ignores the uncertainty associated with smoothing parameter selection and covariance estimation. In small to moderate samples, this may cause the actual coverage level of the SCB procedure to be lower than its target $1 - \alpha$.

Alternatively one can approximate the distribution of the (supremum norm of the) standardized estimator $sup_{t \in D} \sqrt{n} |(\hat{\mu}(t) - \mu(t))| / \hat{\sigma}(t)$ by nonparametric bootstrap. This is done by generating a nonparametric bootstrap sample $\left(X_1^*,\dots,X_n^*\right)$ of the presmoothed curves $\hat{X}_1,\dots,\hat{X}_n$ and computing the bootstrap estimator $\mu^* = \frac{1}{n} \sum_{i=1}^{n} X_i^*$ many times in order to determine the distribution of $\sup_{t \in D} \left| \frac{\sqrt{n}}{\hat{\sigma}(t)} (\mu^*(t) - \hat{\mu}(t)) \right|$. (A number $N = 2500$ of bootstrap replications seem sufficient in practice.) It then suffices to substitute the $(1 - \alpha)$-quantile of this bootstrap distribution to $z_{\alpha,\hat{\rho}}$ in the SCB. This nonparametric bootstrap typically requires much more computational effort than the parametric bootstrap but is also potentially more accurate because it accounts for the uncertainty in covariance estimation and does not rely on large sample normal approximations. Note however that this procedure uses the same smoothing parameter in all bootstrap replications (the one used to estimate μ with the original data). A strict implementation of the nonparametric bootstrap would require bootstrapping the original data (not $\hat{X}_1,\dots,\hat{X}_n$) and performing smoothing parameter selection for each bootstrap replication,[42] in addition to all other computations. At great computational expense, this could bring the coverage level of the SCB even closer to $1 - \alpha$.

For the sake of completeness, we mention an entirely different approach that utilizes FPCA and the geometry of Hilbert spaces to build confidence regions for μ shaped as hyperellipses or hyperrectangles.[7] The confidence regions can be derived in analytic form, which greatly speeds up computations. Although they have zero coverage, they are very close to regions with proper coverage for large n, p. In addition, the confidence ellipses can be expanded and transformed into conservative SCB.

The Case of Sparse Functional Data

The construction of SCB with sparse functional data is markedly different from the case of dense functional data. This is because here, the sparsity of observations within curves (see Table 1) renders longitudinal dependence asymptotically negligible as $n \to \infty$.[43]

The nonparametric methods used to estimate μ in the dense case can be identically applied in the sparse case:

$$\hat{\mu}(t) = \sum_{i=1}^{n} \sum_{j=1}^{p_i} w_{ij}(t) Y_{ij}. \qquad (13)$$

In model (3) with a sparse design, it is often assumed that (1) the observation points t_{ij} are independent and identically distributed (i.i.d.), (2) the numbers p_i of points per curve are i.i.d., (3) the errors ε_{ij} are i.i.d. with variance function $\sigma_\varepsilon^2(t) = V(\varepsilon(t))$, and (4) all these random variables are mutually independent.

Two important functional parameters for the sparse design are the variance $\sigma_Y^2(t) = \Gamma(t,t) + \sigma_\varepsilon^2(t)$ of $Y(t) = X(t) + \varepsilon(t)$ and the common density function f of the t_{ij}. The first can be estimated by smoothing the points $\left(t_{ij}, \hat{\varepsilon}_{ij}^2\right)$, where $\hat{\varepsilon}_{ij} = Y_{ij} - \hat{\mu}\left(t_{ij}\right)$ are the residuals of the estimation of μ. The second can be estimated with any suitable nonparametric density estimator, e.g., $\hat{f}(t) = \frac{1}{N}\sum_{i,j}\frac{1}{h}K\left(\frac{t_{ij}-t}{h}\right)$ where K is a kernel function, h is a bandwidth, and $N = \sum_i^n p_i$ is the total number of data points.

Considering the estimation of μ over $D = [0,1]$ by linear spline smoothing[18] or local linear smoothing[19] and assuming the above i.i.d. conditions on the t_{ij}, p_i, and ε_{ij}, moment conditions on the p_i and ε_{ij}, regularity conditions on μ, Γ, σ_ε^2, f, and additional conditions on the smoothing parameters, the simple SCB

$$\left\{\left[\hat{\mu}(t) \pm \frac{C\hat{\sigma}_Y(t)}{\left(Nh\hat{f}(t)\right)^{1/2}} Q_h(\alpha)\right] : t \in D\right\} \qquad (14)$$

has approximate coverage level 1 – α for μ, where C is a constant that depends on the estimation method, h is a smoothing parameter akin to a bandwidth, and $Q_h(\alpha)$ is a quantile function that only depends on the smoothing method.[18,19] That is, the simultaneous coverage level of (14) tends to 1 – α as $n \to \infty$.

Software

Many packages are available in the R programming environment for building SCB for the mean of functional data. Table 3 presents popular R packages that can be harnessed to perform individual steps of the SCB construction, e.g., mean function estimation, covariance estimation, and resampling. To this date, however, only the package SCBmeanfd[44] provides end-to-end computations. There seems to be no dedicated MATLAB toolboxes for SCB with functional data.

All packages listed in Table 3 are available on https://cran.r-project.org except for fregion which can be obtained on https://github.com.

TABLE 3 | R Packages for Mean Function Estimation, Covariance Estimation, and Simultaneous Confidence Bands with Functional Data

Statistical Method	Package
Smoothing splines and B-splines	Splines (univariate); gss (multivariate)
Penalized splines	Face, fda, uniReg
Local kernel smoothing	KernSmooth, locfit
Wavelets	Wavethresh, wavelets
Functional PCA	fda, refund (dense); face, fdapace (sparse)
Simultaneous confidence bands	Fregion, SCBmeanfd

APPLICATIONS TO HYPOTHESIS TESTING

SCB constitute a useful tool for exploratory data analysis and model building. They also find important applications to hypothesis testing, e.g., specification tests, equality tests, and goodness-of-fit tests.[8,9,22] Below are a few examples implemented with the R package SCBmeanfd[44] (R code provided as Supporting information).

Testing the Specification of a Mean Function

SCB have a straightforward application to specification tests of the form $H_0 : \mu = \mu_0$ versus the alternative $H_a : \mu \neq \mu_0$, where μ_0 is a prespecified function. A simple way to test the null hypothesis H_0 in model (3) is to build an approximate SCB of level 1 – α for μ as in (12). One then retains H_0 if the SCB contains μ_0 and rejects H_0 otherwise. By construction this statistical test has (approximate) significance level α. In addition the test is consistent against H_a, i.e., its statistical power tends to 1 as $n, p \to \infty$. In other words, the test will detect any misspecification H_a of μ with probability close to 1 for large sample size n and grid size p. Under regularity conditions, the test is also consistent against local alternatives approaching H_0 at a rate slower than $n^{-1/2}$, i.e., alternatives of the form $H_n : \mu = \mu_n$ such that $n^{1/2}\sup_{t\in D}|\mu_0(t) - \mu_n(t)| \to \infty$ as $n, p \to \infty$.[9] One may also exploit the SCB (12) to localize the region $R = \{t \in D : \mu(t) \neq \mu_0(t)\}$ where H_0 does not hold. Indeed the region $\hat{R} = \{t \in D : \mu_0(t) \notin [\hat{\mu}(t) \pm z_{\alpha,\hat{\rho}}\hat{\sigma}(t)n^{-1/2}]\}$ where the SCB does not contain μ_0 is close to R (in the sense that the symmetric difference $R \Delta \hat{R} = \left(R \setminus \hat{R}\right) \cup \left(\hat{R} \setminus R\right)$ tends to a negligible set) with probability at least 1 – α for large n, p.

We now illustrate the above specification test and localization procedure with an example from cytogenetics. Array comparative genomic hybridization (aCGH) is a cytogenetic technique for detecting changes in DNA copy numbers along the genome. Applications of aCGH are mainly directed at detecting genomic abnormalities in cancer. Although it is progressively replaced by high throughput sequencing techniques, aCGH can aid in the identification and localization of cancer causing genes. It also finds clinical use in diagnosis, cancer classification, and prognosis. Given a collection of aCGH profiles of cancerous tumors, a common research problem is to detect chromosomal regions where copy numbers are either abnormally high (gain) or abnormally low (loss). This task can be carried out with descriptive methods based on the frequency of samples with copy numbers above (for gains) or below (for losses) a given threshold at each location. It can also be tackled with statistical methods for data segmentation and change point detection.[45]

Here, we cast the gain/loss detection problem in terms of the population average aCGH profile μ (of cancer patients) expressed as the log ratio of the observed copy number versus a baseline reference. The task at hand is to test the null hypothesis $H_0 : \mu \equiv 0$ versus $H_a : \overline{H_0}$ (i.e., $\mu(t_0) \neq 0$ for some $t_0 \in D$) and to detect regions where the mean function μ is nonzero. As proposed above, we accomplish this task by building an SCB for μ over its domain D, here the genome indexed by its base pairs, and identifying the regions of D where the SCB does not contain zero. The mean function μ is then deemed to be nonzero in these regions with the prescribed confidence level, say 95%. This inferential procedure was applied to aCGH profiles of primary colorectal tumors (n = 124) measured at approximately 2000 'clones,' i.e., locations along the genome (p = 2031, resolution ~1.4 Mb per clone).[46] Hereafter, we show to conduct the analysis with the R package SCBmeanfd which uses local linear smoothing for mean function estimation and FPCA for covariance estimation.

```
> require(aCGH)
> data(colorectal, package="aCGH")
> y = t(log2.ratios(colorectal))
> ## Locations of clones
> start = cumsum(c(0, human.chrom.info.Jul03$length[1:23]))
> x = start[clones.info(colorectal)$Chrom] + clones.info(colorectal)$kb
> ## Missing data imputation
> na.ind = which(is.na(y),T)
> ybar = colMeans(y,na.rm=T)
> y[na.ind] = ybar[na.ind[,2]]
> ## SCB
> scb = scb.mean(x, y, bandwidth=2.5e4, scbtype="both", gridsize=5e3)
> plot(scb) # basic plot
> ## Hypothesis test
> scb.model(x, y, model=0, bandwidth=2.5e4, scbtype="both", gridsize=5e3)

Goodness-of-fit test
Model for the mean function: zero
Bandwidth: 25000
SCB type: normal and bootstrap
Significance level: 0.05
Test statistic and p value
  stat   normal p  bootstrap p
  7.983  <1e-16    0.0014
```

The hypothesis test produced a test statistic $T = n^{1/2}\max_{t \in D}|\hat{\mu}(t)/\hat{\sigma}(t)| = 7.983$ with P values $< 10^{-16}$ (normal-based) and 0.0014 (bootstrap-based), giving overwhelming statistical evidence that the average aCGH profile of cancer patients is not uniformly zero. In other words, there is considerable empirical evidence for significant gains/losses in DNA copy numbers among the cancer population. Figure 3 shows normal and bootstrap SCB of level 95% for the population average aCGH profile. Regions with relatively frequent gains/losses in Ref 46 (chromosomes 8, 17, 18, 20) are also found here to have log-ratios significantly different from zero. In addition, Figure 4 indicates a transition from loss to gain on chromosome 1 and relatively large losses on chromosomes 5 and 15.

Rather than performing a single hypothesis test with only one value of the smoothing parameter(s), it is often more informative to conduct the test for a range of smoothing parameter values (here the bandwidth h) and examine the variations of the P value. This approach bears strong connections to scale space methods, e.g., SiZer.[47] It is particularly relevant in the context of hypothesis testing where no theory is available to guide the selection of smoothing parameters. Indeed methods like cross validation are in a certain sense optimal for estimation but not necessarily for inference. The plot of the P value in function of the smoothing parameter is called a *significance trace*.[48,49] An example illustrating the previous specification test ($H_0 : \mu \equiv 0$ versus $\overline{H_0}$) is provided in Figure 5.

Comparing Mean Functions in Two Populations

The SCB procedure (12) presented in the case of one sample can be extended to two samples in a straightforward manner. It suffices to first estimate the mean function μ_i and covariance function Γ_i of the process

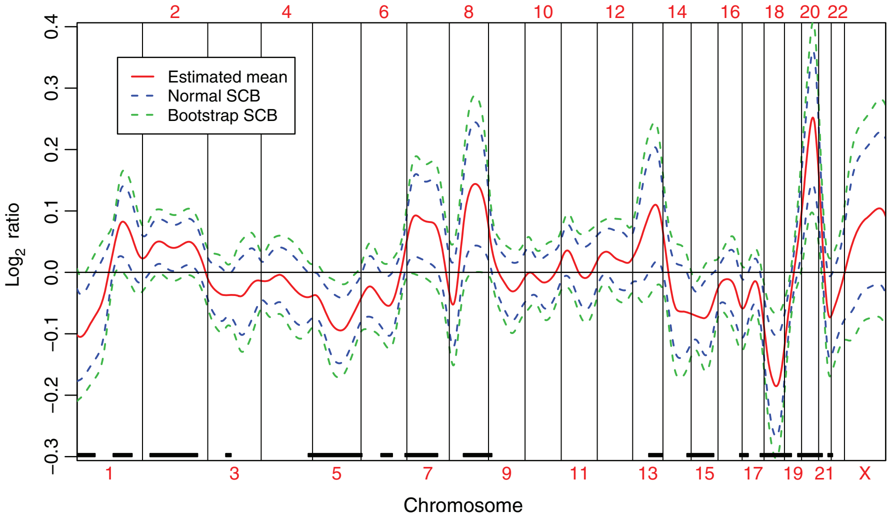

FIGURE 4 | Simultaneous confidence bands of level 95% for the population-level average aCGH profile of colorectal tumors. Chromosomal regions with significant gain/loss in DNA copy number are marked with black rectangles above the x-axis. These regions may contain cancer causing genes.

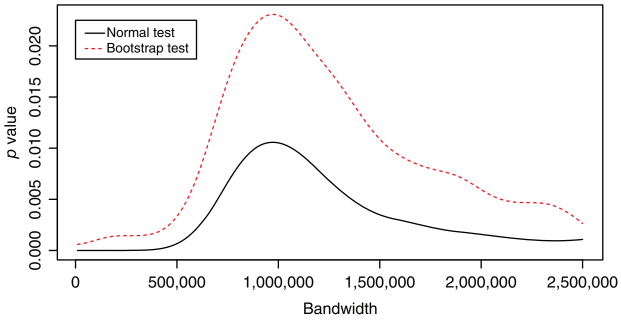

FIGURE 5 | Significance trace of a specification test for the average aCGH profile μ of the cancer population. The null hypothesis $H_0 : \mu \equiv 0$ (no gain/loss in DNA copy numbers) can be rejected at the 5% significance level for all bandwidths in both the normal and bootstrap tests. H_0 can be rejected at the 1% level for virtually all bandwidth values in the normal test and a relatively wide range of bandwidths in the bootstrap test.

$X^{(i)}$ under study for each population (i = 1,2). For large sample sizes n_1, n_2, and under regularity conditions, the difference estimator $(\hat{\mu}_1 - \hat{\mu}_2)$ is approximately distributed as a Gaussian process $\mathcal{G}\left(\mu_1 - \mu_2, \frac{\Gamma_1}{n_1} + \frac{\Gamma_2}{n_2}\right)$. Then, given a confidence level 1 − α, build the SCB

$$\left\{ \left[\hat{\mu}_1(t) - \hat{\mu}_2(t) \pm z_{\alpha,\hat{\rho}_{12}} \left(\frac{\hat{\sigma}_1^2(t)}{n_1} + \frac{\hat{\sigma}_2^2(t)}{n_2} \right)^{1/2} \right] : t \in D \right\}, \tag{15}$$

where $\hat{\sigma}_i^2(t) = \hat{\Gamma}_i(t,t)$ is the variance estimator for $X^{(i)}(t)$, $\hat{\rho}_{12}$ is the estimated correlation function of $(\hat{\mu}_1 - \hat{\mu}_2)$, and the quantile $z_{\alpha,\hat{\rho}_{12}}$ is obtained as before by Gaussian-process simulation or by stratified bootstrap of the two data samples. The SCB (15) has approximate coverage 1 − α for the difference $(\mu_1 - \mu_2)$. The hypothesis $H_0 : \mu_1 = \mu_2$ can be very simply tested against $H_a : \mu_1 \neq \mu_2$ as follows. If the zero function is entirely contained in the SCB, retain H_0; otherwise, reject H_0. Like in the one-sample case, this test has approximate significance level α and is consistent against fixed and local alternatives. In comparison to related statistical tests that only return a *P* value, the SCB test bears the important advantage that it not only answers the question *whether* two mean functions are different but also the questions '*where?*' and '*by how much?*' (are they different). The SCB test measures distance between functions with the L^∞ norm and is therefore best suited to detect local differences of relatively large amplitude. In contrast, tests based on the L^2 distance are more effective to detect differences of smaller amplitude but greater spatial extent. For visualization purposes, the SCB (15) can be centered around one of the mean function estimates, say $\hat{\mu}_1$, instead of $\hat{\mu}_1 - \hat{\mu}_2$ (see Figure 5). In this case H_0 is retained or rejected according to whether the other estimator, $\hat{\mu}_2$, lies within the SCB centered on $\hat{\mu}_1$.

We now illustrate the above procedure with a brain tractography study of MS. The MS is a neurological disorder that affects an estimated 2.3 million people in the world. Its most common symptoms are overwhelming fatigue, visual disturbances, altered sensation, and difficulties with mobility. Although there is currently no cure for this disease, neuroimaging research holds the

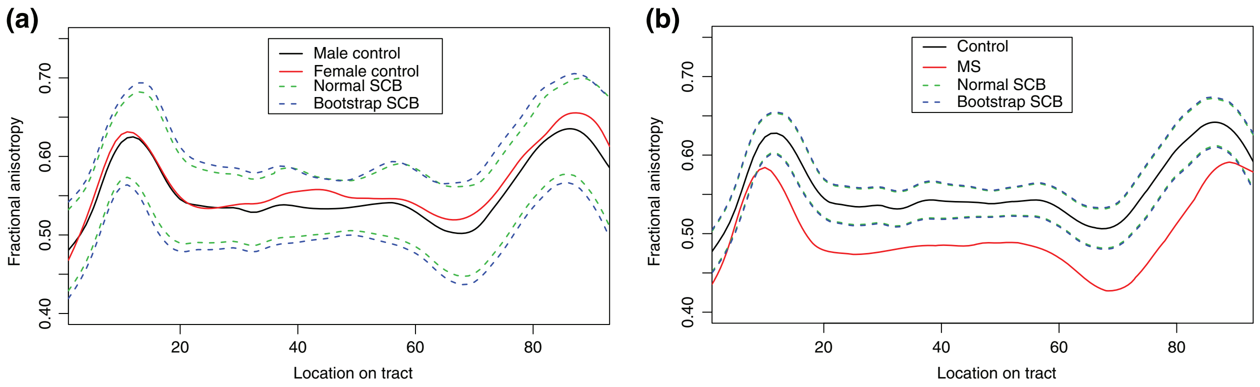

FIGURE 6 | Comparisons of mean tract profiles in diffusion tensor imaging fractional anisotropy data. Left: male control group versus female control group. Right: control group versus multiple sclerosis group (all genders).

potential to improve its diagnosis and prognosis predictions. Diffusion tensor imaging (DTI) is a magnetic resonance imaging technique that maps white matter tractography in the brain based on water diffusivity. White matter tracts are known to be altered by MS and as such, characterizing differential patterns in their orientation and water diffusivity may shed light on the workings of MS. The goal here is to detect significant differences in the average tract profiles of MS patients and healthy controls, as measured by fractional anisotropy (FA) which is a ratio between 0 (perfect spherical diffusion) and 1 (ideal linear diffusion). Low FA indicates potential neurological damage and pathology. The DTI data were collected at Johns Hopkins University and the Kennedy Krieger Institute and obtained from the R package refund.[50] They consist in FA measurements of 382 subjects (340 MS patients, 42 controls) at 93 tract locations along the corpus callosum.

Computations were conducted with the R package SCBmeanfd.[44] Hereafter is example code for comparing the male control and female control groups. (This comparison is for exploratory purposes. The more important comparison of MS patients and controls was also carried out but is not shown here for reasons of space.)

```
> library(SCBmeanfd)
> library(refund)
> y1 = with(DTI,cca[case==0 & sex=="male",]) # male control group
> y2 = with(DTI,cca[case==0 & sex=="female",]) # female control group
> x = 1:ncol(y1) # tract locations
> h1 = cv.select(x,y1) # bandwidth selection
> h2 = cv.select(x,y2)
> (test = scb.equal(x, list(y1,y2), c(h1,h2), scbtype="both")) # equality test

Equality test for mean functions

Bandwidths: 1.1668 1.658
SCB type: normal and bootstrap
Significance level: 0.05
Test statistic and p value
  stat   normal p  bootstrap p
  2.097  0.2482    0.3488

> plot(test) # basic plot
```

The results of the SCB procedure are displayed in Figure 6. The left panel shows a comparison of males and females in the control group. Both estimated mean functions are within the 95% SCB, which indicates no significant difference (P value 0.25 for normal SCB and 0.35 for bootstrap SCB). The right panel compares the average profiles of the MS and control group (all genders confounded). The massive differences between groups (P value less than 10^{-16} for both normal and bootstrap SCB) are visually indicated by the fact that the SCB do not contain both mean function estimates. In addition, a significant difference was found between males and women in the MS group (P value 0.001 for normal SCB, 0.0008 for bootstrap SCB). This difference can however be explained by the considerably larger size of this group. These results agree with previous statistical studies of the same data.[51,52]

CONCLUSIONS

This article gave a brief introduction to the theory, methods, and implementation of SCB with functional data. Although the discussion has focused on mean functions, many of the ideas presented here also apply to functional parameters such as quantile functions, survival functions, and covariance functions. The main takeaways are as follows. First, in the analysis of functional data, simultaneous inference is the correct way to *globally* assess the mean function. Pointwise methods are inadequate for this purpose because they fail to account for the large (infinite) number of inferences conducted and thus produce excessively optimistic inferences. Second, SCB provide a versatile platform for the exploration and simultaneous inference of the mean function. In addition to providing guidance in model building, they can also be employed in specification tests, goodness-of-fit tests for parametric models, comparisons between two populations, inference of extrema, and more. Third, the powerful visualization capabilities of SCB facilitate the communication of inferential results to all publics, including nonstatistical experts. Fourth, the construction of SCB with functional data is both computationally and conceptually straightforward. It parallels the construction of PCBs in that it requires three main steps: first estimate the mean function, then the covariance function, and finally obtain numerically or theoretically (the quantile that determines) the amplitude of the SCB required to achieve the target confidence level. In other words, the construction of SCB is as simple as that of (often incorrectly used) PCB.

In recent years, several new research directions have emerged in FDA. In particular, there has been increasing attention to spatial and temporal dependence in functional data, which has opened a broad range of new applications in earth sciences, finance, energy, and more.[53] Consistency results have been established for functional means and principal components[54] and inferential methods have been proposed[55,56] but as a whole, research on dependent functional data is still in its early days and presents a wealth of theoretical, methodological, and computational challenges. Functional regression is another new area of intense research.[57] There too, the construction of SCB is a significant challenge and requires more in-depth scrutiny.

Two related open problems in the simultaneous inference of functional data concern the selection of smoothing parameters and the optimality of confidence regions. More precisely, smoothing parameter selection methods such as cross validation and plugin estimation, that are optimal for pointwise estimation, are in general suboptimal for inference. Selection methods that can: (1) guarantee inferences at close-to-nominal level even for small samples and (2) produce confidence regions of optimal size and/or statistical power, are yet to be developed and would constitute a major statistical contribution.

As noted earlier in the article, software resources for computing and visualizing SCB with functional data are scarce. SCB packages in R, Matlab, Python, and so on are strongly needed and will be instrumental in promoting the use of SCB methods in applied research.

FURTHER READING

The reader interested in convergence rates for mean function estimation with dense or sparse functional data may usefully refer to:

Cai TT, Yuan M. Optimal estimation of the mean function based on discretely sampled functional data: phase transition. *Ann Stat* 2011, 39:2330–2355.

Zhang X, Wang J-L. From sparse to dense functional data and beyond. *Ann Stat* 2016, 44:2281–2321.

The construction of SCB for derivatives of the mean function has been addressed in:Cao G, Wang J, Wang L, Todem D. Spline confidence bands for functional derivatives. *J Stat Plann Infer* 2012, 142:1557–1570.

Cao G. Simultaneous confidence bands for derivatives of dependent functional data. *Elec J Stat* 2014, 8:2639–2663.

A comprehensive reference on asymptotic theory for stochastic processes, including functional central limit theorems, is:van der Vaart AW, Wellner JA. *Weak Convergence and Empirical Processes. With Applications to Statistics*. New York: Springer-Verlag; 1996.

REFERENCES

1. Kaufman CG, Sain SR. Bayesian functional ANOVA modeling using gaussian process prior distributions. *Bayesian Anal* 2010, 5:123–149.
2. Hall P, Müller H-G, Yao F. Estimation of functional derivatives. *Ann Stat* 2009, 37(6A):3307–3329.
3. Usset J, Staicu A-M, Maity A. Interaction models for functional regression. *Comput Stat Data Anal* 2016, 94:317–329.
4. Cao J, Zhao H. Estimating dynamic models for gene regulation networks. *Bioinformatics* 2008, 24:1619–1624.

5. Wang J-L, Chiou J-M, Müller H-G. Functional data analysis. *Annu Rev Stat Appl* 2016, 3:257–295.
6. Ramsay JO, Silverman BW. *Functional Data Analysis*. Springer Series in Statistics. 2nd ed. New York: Springer; 2005.
7. Choi H , Reimherr M. A geometric approach to confidence regions and bands for functional parameters. arXiv:1607.07771v2 [stat.ME], 2016.
8. Cao G, Yang L, Todem D. Simultaneous inference for the mean function based on dense functional data. *J Nonparametr Stat* 2012, 24:359–377.
9. Degras D. Simultaneous confidence bands for nonparametric regression with functional data. *Stat Sin* 2011, 21:1735–1765.
10. Bunea F, Ivanescu AE, Wegkamp MH. Adaptive inference for the mean of a Gaussian process in functional data. *J R Stat Soc Series B Stat Methodol* 2011, 73:531–558.
11. Rice JA, Silverman BW. Estimating the mean and covariance structure nonparametrically when the data are curves. *J R Stat Soc Series B* 1991, 53:233–243.
12. Hart JD, Wehrly TE. Consistency of cross-validation when the data are curves. *Stoch Process Their Appl* 1993, 45:351–361.
13. Hart JD, Wehrly TE. Kernel regression estimation using repeated measurements data. *J Am Stat Assoc* 1986, 81:1080–1088.
14. Benhenni K, Degras D. Local polynomial estimation of the mean function and its derivatives based on functional data and regular designs. *ESAIM Probab Stat* 2014, 18:881–899.
15. Yao F, Müller H-G, Wang J-L. Functional data analysis for sparse longitudinal data. *J Am Stat Assoc* 2005, 100:577–590.
16. Cuevas A, Febrero M, Fraiman R. On the use of the bootstrap for estimating functions with functional data. *Comput Stat Data Anal* 2006, 51:1063–1074.
17. Degras D. Nonparametric estimation of a trend based upon sampled continuous processes. *C R Math Acad Sci Paris* 2009, 347:191–194.
18. Ma S, Yang L, Carroll RJ. A simultaneous confidence band for sparse longitudinal regression. *Stat Sin* 2012, 22:95–122.
19. Zheng S, Yang L, Härdle WK. A smooth simultaneous confidence corridor for the mean of sparse functional data. *J Am Stat Assoc* 2014, 109:661–673.
20. Cardot H, Degras D, Josserand E. Confidence bands for Horvitz-Thompson estimators using sampled noisy functional data. *Bernoulli* 2013, 19(5A):2067–2097.
21. Cardot H, Goga C, Lardin P. Uniform convergence and asymptotic confidence bands for model-assisted estimators of the mean of sampled functional data. *Electron J Stat* 2013, 7:562–596.
22. Crainiceanu CM, Staicu A-M, Ray S, Punjabi N. Bootstrap-based inference on the difference in the means of two correlated functional processes. *Stat Med* 2012, 31:3223–3240.
23. Yang J, Zhu H, Choi T, Cox DD. Smoothing and mean-covariance estimation of functional data with a Bayesian hierarchical model. *Bayesian Anal* 2016, 11:649–670.
24. Morris JS, Carroll RJ. Wavelet-based functional mixed models. *J R Stat Soc Series B Stat Methodol* 2006, 68:179–199.
25. Liu Z, Guo W. Functional mixed effects models. *Wiley Interdiscip Rev Comput Stat* 2012, 4:527–534.
26. Krivobokova T, Kneib T, Claeskens G. Simultaneous confidence bands for penalized spline estimators. *J Am Stat Assoc* 2010, 105:852–863.
27. Sørbye SH, Rue H. Simultaneous credible bands for latent Gaussian models. *Scand J Stat* 2011, 38:712–725.
28. Mas A. Testing for the mean of random curves: a penalization approach. *Stat Infer Stoch Process* 2007, 10:147–163.
29. Antoniadis A, Sapatinas T. Estimation and inference in functional mixed-effects models. *Comput Stat Data Anal* 2007, 51:4793–4813.
30. Ghiglietti A, Ieva F, Paganoni AM. Statistical inference for stochastic processes: two-sample hypothesis tests. *J Stat Plann Infer* 2017, 180:49–68.
31. Benko M, Härdle W, Kneip A. Common functional principal components. *Ann Stat* 2009, 37:1–34.
32. Wand MP, Jones MC. *Kernel Smoothing*. London: Chapman and Hall; 1995.
33. Green PJ, Silverman BW. *Nonparametric Regression and Generalized Linear Models. A Roughness Penalty Approach*. London: Chapman & Hall; 1994.
34. Eilers P, Marx B. Flexible smoothing with B-splines and penalties. *Stat Sci* 1996, 89:89–121.
35. De Boor C. *A Practical Guide to Splines. Applied Mathematical Sciences*. New York: Springer-Verlag; 1978.
36. Nason G. *Wavelet Methods in Statistics with R*. New York: Springer; 2008.
37. Fan J, Gijbels I. *Local Polynomial Modelling and Its Applications*. London: Chapman & Hall; 1996.
38. Ruppert D, Wand MP, Carroll RJ. *Semiparametric Regression*. Cambridge: Cambridge University Press; 2003.
39. Degras D. Asymptotics for the nonparametric estimation of the mean function of a random process. *Stat Probab Lett* 2008, 78:2976–2980.
40. Indritz J. *Methods in Analysis*. New York and London: The Macmillan Co. and Collier-Macmillan Ltd.; 1963.
41. Goldsmith J, Greven S, Crainiceanu C. Corrected confidence bands for functional data using principal components. *Biometrics* 2013, 69:41–51.

42. Efron B. Estimation and accuracy after model selection. *J Am Stat Assoc* 2014, 109:991–1007.

43. Yao F. Asymptotic distributions of nonparametric regression estimators for longitudinal or functional data. *J Multivariate Anal* 2007, 98:40–56.

44. Degras D. *SCBmeanfd: Simultaneous Confidence Bands for the Mean of Functional Data*. R package version 1.2.2, 2016.

45. Hocking TD, Schleiermacher G, Janoueix-Lerosey I, Boeva V, Cappo J, Delattre O, Bach F, Vert J-P. Learning smoothing models of copy number profiles using breakpoint annotations. *BMC Bioinformatics* 2013, 14:1–15.

46. Nakao K, Mehta KR, Fridlyand J, Moore DH, Jain AN, Lafuente A, Wiencke JW, Terdiman JP, Waldman FM. High-resolution analysis of DNA copy number alterations in colorectal cancer by array-based comparative genomic hybridization. *Carcinogenesis* 2004, 25:1345–1357.

47. Chaudhuri P, Marron JS. SiZer for exploration of structures in curves. *J Am Stat Assoc* 1999, 94:807–823.

48. King E, Hart JD, Wehrly TE. Testing the equality of two regression curves using linear smoothers. *Stat Probab Lett* 1991, 12:239–247.

49. Bowman AW, Azzalini A. *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations*. Press, New York: Oxford Univ; 1997.

50. Goldsmith J, Scheipl F, Huang L, Wrobel J, Gellar J, Harezlak J, McLean MW, Swihart B, Xiao L, Crainiceanu C, et al. *refund: Regression with Functional Data*. R package version 0.1-16, 2016.

51. Goldsmith J, Bobb J, Crainiceanu CM, Caffo B, Reich D. Penalized functional regression. *J Comput Graph Stat* 2011, 20:830–851.

52. Pomann G-M, Staicu A-M, Ghosh S. A two-sample distribution-free test for functional data with application to a diffusion tensor imaging study of multiple sclerosis. *J R Stat Soc Ser C Appl Stat* 2016, 65:395–414.

53. Horváth L, Kokoszka P. *Inference for Functional Data with Applications*. New York, NY: Springer New York; 2012.

54. Hörmann S, Kokoszka P. Consistency of the mean and the principal components of spatially distributed functional data. *Bernoulli* 2013, 19(5A):1535–1558.

55. Rakêt LL, Markussen B. Approximate inference for spatial functional data on massively parallel processors. *Comput Stat Data Anal* 2014, 72:227–240.

56. Horváth L, Rice G. Testing equality of means when the observations are from functional time series. *J Time Ser Anal* 2015, 36:84–108.

57. Morris JS. Functional regression. *Annu Rev Stat Appl* 2015, 2:321–359.