Significance Tests for Functional Data with Complex Dependence Structure

Ana-Maria Staicu and Soumen N. Lahiri and Raymond J. Carroll

Department of Statistics, North Carolina State University;
Department of Statistics, North Carolina State University;
Department of Statistics, Texas A&M University

Abstract: We propose an L^2-norm based global testing procedure for the null hypothesis that multiple group mean functions are equal, for functional data with complex dependence structure. Specifically, we consider the setting of functional data with a multilevel structure of the form groups-clusters or subjects-units, where the unit-level profiles are spatially correlated within the cluster, and the cluster-level data are independent. Orthogonal series expansions are used to approximate the group mean functions and the test statistic is estimated using the basis coefficients. The asymptotic null distribution of the test statistic is developed, under mild regularity conditions. To our knowledge this is the first work that studies hypothesis testing, when data have such complex multilevel functional and spatial structure. Two small-sample alternatives, including a novel block bootstrap for functional data, are proposed, and their performance is examined in simulation studies. The paper concludes with an illustration of a motivating experiment.

Key words and phrases: Block bootstrap; Functional data; Group mean testing; Hierarchical modeling; Significance tests; Spatially correlated curves.

1 Introduction

Advancements in technology and computation have led to a rapidly increasing number of applications where repeated functional data are observed per subject, for many subjects. These developments have been accompanied and, in some cases, anticipated by intense methodological development in functional data analysis (Ramsay and Silverman, 2005; Ferraty and Vieu, 2006). Although much work has been done on estimation in various models for multilevel functional data (Morris and Carroll, 2006; Di et al., 2009; Crainiceanu et al., 2009; Staicu, et al., 2010), there is only limited work on inference for the fixed effects in these more complex models. This paper focuses on closing this gap for functional data that have a natural multilevel structure, group - cluster - unit with functional-type measurements at unit level, such that conditional on the subject, the unit-level measurements are spatially correlated. Accounting for the complex dependence among the curves when carrying hypothesis testing about the group means is very important, since the common testing procedures, applied by ignoring the curve-dependence, yield misleading results.

When there is a single curve per cluster, thus all the curves are independent, testing the significance of the group mean functions has been studied extensively (Fan and Lin, 1998; Shen and Faraway, 2004; Cuevas et al., 2004; Staicu et al., 2013). For example, to assess the equality of the group means when the random curves have a stationary time series covariance, Fan and Lin (1998) proposed a powerful overall test based on the decomposition of the original functional data into Fourier or wavelet series expansions. Cuevas et al. (2004) developed an ANOVA-like test statistic for the hypothesis testing about group mean functions, when the curves come from independent samples of independent curves; their setting assumes the curves are fully observed and without noise. Zhang and Chen (2007) considered a similar model setup and discussed an L^2-norm test in the context where the curves are observed on dense grid of points and are corrupted with measurement error. For two samples of curves, Zhang et al. (2010) discussed both a pointwise t-test and a global L^2-norm based test statistic and employed bootstrap procedure to approximate the null distribution. However, all these methods rely on the assumption that the curves in the samples are independent, and extending them to the setting where curves have complex correlation structures is far from straightforward. For paired functional data, where the data consist of independent pairs of functions and each pair of functions exhibit complex dependence, Crainiceanu et al. (2012) discussed bootstrap-based inferential methods for the difference in the mean profiles, and Staicu et al. (2013) proposed likelihood-ratio type statistics. These results, while are in the direction of our research, are not applicable to our setting where we have ‘clusters’ of dependent curves contained in multiple independent samples.

The objective of this paper is to develop inferential methods for testing hypotheses about group mean and group mean differences for hierarchical functional data of the type group - cluster - unit, when the unit-level functions exhibit spatial correlation. We take a nonparametric approach and study a test statistic based on the L^2 distance among the group mean functions. To the best of our knowledge, no asymptotic distribution results are currently available for hypothesis testing about the group mean functions, when data are curves with complex multilevel functional and spatial structure. Small-sample alternatives based on bootstrap procedure are proposed and are examined in simulation studies. The main contributions of this paper are (i) the proposal and development of the asymptotic null distribution of the testing procedure for group mean functions for functional data with such complex dependence; and (ii) the proposal of a novel block bootstrap approach for functional data, inspired from spatial statistics.

The remaining of the paper is structured as follows. Section 2 introduces the notation, describes the hypothesis testing problem that we consider, introduces the testing procedure and outlines the working assumptions. Section 3 presents the asymptotic study of the null distribution of the testing procedure. Bootstrap approximations to the asymptotic null distribution are detailed in Section 4. Illustration of the proposed approach in finite sample is via a simulation study in Section 5 and application to a long range infrared light detection and ranging (LIDAR) study in Section 6. Section 7 concludes with a discussion.

2 Statistical framework

2.1 Preliminaries

We first introduce the notation, the model assumptions, and the testing procedure. Broadly, the structure of the data is groups - clusters or subjects - units, where the unit-level data consist of a sequence of repeated measurements, the unit-level data are spatially correlated within the cluster or subject, and the cluster or subject-level data are assumed independent of each other. For exposition simplicity we use 'subjects' throughout the paper. Let i index the subjects, j index the units, and denote by Y_{ijl} the lth repeated measurement which corresponds to the time point t_{ijl}. Moreover it is assumed that the units are 'ordered' and denote by s_{ij} the location of the jth unit within the ith subject. Let N_{ij} be the number of repeated measurements for unit j within subject i, let M_i be the number of units within subject i, and n be the total number of subjects. The subjects are separated into D groups, and let $G(i)$ denote the group membership of the ith subject, $G(i) \in \{1, \ldots, D\}$, and let n_d be the number of subjects in group d.

It is assumed that observed data are realizations of a random process on discrete grids, which are further contaminated by noise, as follows:

$$Y_{ijl} = \mu_{G(i)}(t_{ijl}) + V_i(t_{ijl}, s_{ij}) + \epsilon_{ijl}, \tag{1}$$

for $l = 1, \ldots, N_{ij}$, $j = 1, \ldots, M_i$ and $i = 1, \ldots, n$. Here $\mu_d(\cdot)$ is the unknown mean function in group d and the main object of inference, $V_i(\cdot, \cdot)$ a mean-zero bi-variate random process defined on $\mathcal{T} \times \mathcal{D}$, where $\mathcal{T} \in R$ and $\mathcal{D} \in R^2$, and ϵ_{ijl} is random error. We assume that $V_i(\cdot, \cdot)$'s are independent and identically distributed over i, and ϵ_{ijl} are independent and identically distributed with mean zero and variance σ_ϵ^2, and furthermore are independent of $V_i(\cdot, \cdot)$.

Our objective is to test the hypothesis that the group mean functions $\mu_d(\cdot)$ are equal

$$H_0 : \mu_1(\cdot) = \ldots = \mu_D(\cdot) \qquad \text{versus} \qquad H_a : \mu_{d_1}(\cdot) \neq \mu_{d_2}(\cdot), \quad \text{for some} \quad d_1 \neq d_2. \tag{2}$$

In the case when $\{Y_{ijl} : l = 1, \ldots, N_{ij}\}$ are random curves observed on the entire domain, say $\{Y_{ij}(t) : t \in [0, 1]\}$, and without measurement error such that $Y_{ij}(t) = \mu_{G(i)}(t) + W_{ij}(t)$, for independent and identically distributed zero-mean processes $W_{ij}(\cdot)$, then the testing hypothesis (2) has also been considered by Cuevas et al. (2004). Our model framework (1) reduces to the framework considered by Cuevas et al. (2004), when $V_{ij}(t, s_{ij}) = W_{ij}(t)$ and $\epsilon_{ij\ell} = 0$ for all i, j, ℓ. The authors proposed a test statistic that quantifies the "between" groups variability in this functional framework: specifically, when $M_i = 1$ for all i, using our notation, their test is $\sum_{d<d'} n_d \|\widehat{\mu}_d(\cdot) - \widehat{\mu}_{d'}(\cdot)\|_{L^2}^2$, where n_d is the number of subjects in group d, $\widehat{\mu}_d$ is the sample mean estimator of the group mean function μ_d and $\| \cdot \|_{L^2}^2$ refers to the L^2 norm induced by the inner product $< f, g >_{L^2} = \int_0^1 f(t)g(t)\, dt$. Cuevas et al. developed the asymptotic distribution of this test statistic, when the null hypothesis (2) is true. Nevertheless, testing the null hypothesis (2) under a more realistic and general framework that does not restrict the curves to be observed entirely (or on a regular dense grid) and without noise, nor to be all mutually independent, has not been considered yet. A naïve application of the methods described in Cuevas et al. (2004), by ignoring the complex dependence among the curves within the same cluster (subject), or the measurement error may result in considerably increased size.

To the best of the authors' knowledge this research is the first to propose a testing procedure for (2) in the case when 1) the curves are contaminated with measurement error, 2) the curves are not fully observed, as in Cuevas et al. (2004), and 3) the random deviations, as described by $V_i(t, s_{ij})$ in (1) are not independent and identically distributed over both indices i and j. We consider that the random deviations $V_i(t, s_{ij})$ are bi-variate stochastic processes and have a complex covariance structure that combines covariance components commonly encountered in functional as well as spatial data analysis. Specifically, assume that $V_i(t, s_{ij})$ have the the following decomposition

$$V_i(t, s_{ij}) = Z_i(t) + W_{ij}(t) + U_i(s_{ij}), \tag{3}$$

where $Z_i(t), W_{ij}(t)$ and $U_i(s)$ are independent random components, $Z_i(t)$ is the subject-specific random effect and the term $\{W_{ij}(t) + U_i(s_{ij})\}$ represents the unit-specific random deviation from the subject mean. The latter term consists of two components: $U_i(s_{ij})$, which varies with the spatial location, s_{ij}, and $W_{ij}(t)$, which varies with the time index, t. This modeling assumption has also been considered by Staicu et al. (2010) in the context of modeling multilevel functional data that are spatially correlated. It is assumed that $Z_i(\cdot), W_{ij}(\cdot)$ are square integrable random processes on the closed and bounded set $\mathcal{T}$, that $U_i(\cdot)$ is second order stationary on some domain $\mathcal{D}$, and they all have mean zero and continuous covariance functions. For simplicity set $\mathcal{T} = [0, 1]$ and assume that the sampling region $\mathcal{D}$ is a bounded subset of $\mathcal{R}^2$. Furthermore, it is assumed that $s_{ij} = M_i^{1/2} S_{ij}$, where S_{ij} are independent and identically distributed random variables defined on a bounded domain, see Lahiri (2003, Chapter 12).

We take a nonparametric approach, similar to Cuevas et al. (2004), and consider a testing procedure that measures the L^2 distance between the group mean functions. Let $\widehat{\mu}_d(t)$ be a smooth estimator of the group mean function $\mu_d(t)$, and let $\widehat{\mu}.(t) = \sum_{d=1}^{D}(m_d/m)\widehat{\mu}_d(t)$ be a weighted estimator of the overall mean function $\mu(t)$, where $m_d = \sum_{i:G(i)=d} M_i$ is the number of curves in group d and $m = \sum_{i=1}^{n} M_i$ is the total number of curves. It is assumed that $m_d, m \to \infty$ for all d such that the limit $q_d = \lim m_d/m$ exists and is in $(0, 1)$. We propose to test the null hypothesis (2) using the global test:

$$T_n = \sum_{d=1}^{D} \int_0^1 n_d\{\widehat{\mu}_d(t) - \widehat{\mu}.(t)\}^2 \, dt. \tag{4}$$

When $M_i = M_1$ for all i, we have $\widehat{\mu}.(t) = \sum_{d=1}^{D}(n_d/n)\widehat{\mu}_d(t)$; thus when the group sample mean functions are used to estimate $\mu_d(t)$, this test is proportional to the one discussed in Cuevas et al. (2004). Nevertheless its null asymptotic distribution will be different from the one developed by Cuevas et al. (2004), due to the complex structure that is assumed for the covariance of the random component $V_i(t, s_{ij})$. Also, when $D = 2$ and $n_d = 1$ the testing procedure (4) is similar to Horváth et al. (2013), who considered the problem of testing the equality of means of two functional samples which exhibit temporal dependence. Here we develop the asymptotic null distribution for (4) when the observed data Y_{ijl} are discrete realizations from a bi-variate stochastic process having a functional/spatial dependence in a hierarchical setting, as in (1).

2.2 The testing procedure

If the sampling design is regular, $t_{ijl} = t_l$, then the group mean functions, $\mu_d(\cdot)$ can be estimated as common group sample means (see Cuevas et al., 2004). To bypass the restriction on the design regularity, other techniques use local or global smoothing techniques (e.g. Yao et al. 2005, Crainiceanu et al. 2012, etc.), under a working independence assumption. We take a similar viewpoint and consider orthogonal basis expansions for the group mean functions. Specifically, let $\{\psi_\ell(\cdot)\}_{\ell\geq 1}$ be an orthogonal pre-determined basis in $L^2[0,1]$, and write $\mu_d(t) = \sum_{\ell\geq 1}\psi_\ell(t)\beta_{d,\ell}$, where $\beta_{d,\ell}$ are uniquely determined by $\beta_{d,\ell} = \int_0^1 \mu_d(t)\psi_\ell(t)\,dt$. For fixed truncation value L, the group mean can be approximated by $\mu_d^L(t) = \sum_{\ell=1}^L \psi_\ell(t)\beta_{d,\ell}$. Estimation of the basis coefficients $\{\beta_{d,\ell} : \ell = 1,\ldots,L\}_d$ can proceed via a sum of squares criterion using L^2 norm. Specifically, the estimators $\widehat{\beta}_{d,\ell}$ and furthermore $\widehat{\beta}_{\cdot,\ell}$ and are calculated by

$$\widehat{\beta}_{d,\ell} = m_d^{-1}\sum_{\{i:G(i)=d\}}\sum_{j=1}^{M_i}\sum_{l=1}^{N_{ij}} Y_{ijl}\int_{A_{ijl}}\psi_\ell(t)\,dt, \tag{5}$$

$$\widehat{\beta}_{\cdot,\ell} = m^{-1}\sum_{i=1}^{n}\sum_{j=1}^{M_i}\sum_{l=1}^{N_{ij}} Y_{ijl}\int_{A_{ijl}}\psi_\ell(t)\,dt, \tag{6}$$

where $A_{ijl} = [t_{ijl}, t_{ij(l+1)})$, for $l = 1,\ldots,N_{ij}$. The basis coefficients $\widehat{\beta}_{d,\ell}$ are estimated by using integrals of the basis functions over smaller intervals, which is different from the common approach that uses basis functions evaluated at single time points (see for example, Fan & Lin, 1998). Our preference for this approach is based mainly on the simplicity of the expressions of the estimators; our practical experience is that the estimation/testing results obtained with the two approaches are very close. The consistency of the estimators $\widehat{\beta}_{d,\ell}$ and $\widehat{\beta}_{\cdot,\ell}$ is proved in Appendix A.1 and is based on regularity assumptions of the sampling design, group mean function and covariance function, $K^Z(\cdot,\cdot)$, of the process $Z_i(\cdot)$. It follows that the group mean functions can be estimated by $\widehat{\mu}_d^L(t) = \sum_{\ell=1}^L \widehat{\beta}_{d,\ell}\psi_\ell(t)$, and the overall mean function by $\widehat{\mu}^L{\cdot},(t) = \sum_{\ell=1}^L \widehat{\beta}_\ell\psi_\ell(t)$.

These mean estimators are consistent, and to avoid digression from the main point of the paper we defer the discussion of their asymptotic properties to Appendix A1.

Using the group estimates above, the test statistic T_n is approximated by

$$T_n^L = \sum_{d=1}^{D} \sum_{\ell=1}^{L} n_d (\widehat{\beta}_{d,\ell} - \widehat{\beta}_{\cdot,\ell})^2 \tag{7}$$

since $\{\psi_\ell(\cdot)\}_{\ell \geq 1}$ is an orthogonal basis on $[0, 1]$ and thus $\int \psi_\ell(t)\psi_{\ell'}(t)\,dt = 1$ if $\ell = \ell'$ and 0 otherwise. Here the superscript L emphasizes the truncation used in the basis representation of the group mean functions $\mu(t)$. The asymptotic distribution of the test, when the null hypothesis, that the group mean functions are the same is true is developed next. We present first the regularity assumptions on which we base our results.

Assumption 1 (A1): The group mean functions $\mu_d(\cdot)$'s have the following properties:
(a) there exists $\alpha > 0$ such that $\mu_d(\cdot)$ is α-Hölder on $[0, 1]$; μ_d is differentiable;
(b) $\mu_d(\cdot) \in L^2[0, 1]$ and $\int_0^1 |\mu'_d(t)|\,dt < \infty$, where $\mu'_d(t) = \partial\mu_d(t)/\partial t$.

Assumption 2 (A2): The bivariate process $V_i(t, s_{ij})$ admits the decomposition $V_i(t, s_{ij}) = Z_i(t) + W_{ij}(t) + U_i(s_{ij})$, where the independent components $Z_i(\cdot), W_{ij}(\cdot)$ and $U_i(\cdot)$ satisfy the conditions:
(a) The random processes $Z_i(\cdot)$, $W_{ij}(\cdot)$ are square integrable on $[0, 1]$ and have zero-mean functions and covariance functions $K^Z(\cdot, \cdot)$ and $K^W(\cdot, \cdot)$, respectively that are both uniformly bounded in $L^2[0, 1]$. Furthermore the covariance function $K^Z(\cdot, \cdot)$ is assumed twice continuously differentiable and $E(\|Z_i(\cdot)\|_{L^2}^4) < \infty$.
(b) The random process $U_i(\cdot)$ is second order stationary on $\mathcal{D}$, with zero-mean and continuous covariance function. The unit locations $\{s_{ij} : j = 1, \ldots, M_i\}$ are generated by a spatial stochastic design through the relation $s_{ij} = M_i^{1/2} X_{ij}$ for independent and identically distributed random vectors X_{ij}, independent of the other random variables, with density f on some prototype set R_0. Furthermore f is assumed continuous and positive on R_0.

Assumption 3 (A3): We require the following assumptions about the sampling design:
(a) $n_d \to \infty$ and $M_i \to \infty$ for all $i = 1, \ldots, n$. For every $d = 1, \ldots, D$ we have $n_d/n \to p_d > 0$, and $m_d/m \to q_d > 0$, where $m_d = \sum_{\{i:G(i)=d\}} M_i$ and $m = \sum_{i=1}^n M_i$.
(b) There exists $0 < c_1 < c_2 < \infty$ such that $c_1 < M_i/M_{i'} < c_2$ for all i, i' such that $G(i) = G(i')$.
(c) For every $d = 1, \ldots, D$ we have $\min\{N_{ij} : G(i) = d\} > n_d^\theta$, where $\theta > 1/(2\alpha)$,

where α is given in condition A1.

Generally, the selection of the orthonormal basis is important, in the sense that some orthonormal bases may be more appropriate than others under a given situation. However, the theoretical properties of the estimators are independent of the particular basis, as long as it is a pre-determined orthonormal basis (Fourier, orthonormal wavelets, orthonormal B-splines and so on). As a result the choice of basis is expected to have little effect on the testing procedure; the number of basis functions L that does not change considerably the results (size/power) would vary with the choice of the basis. In particular, a smaller value L would suffice if the mean and error process are approximated well by the first few basis functions, than otherwise. We recommend to select L carefully in any particular application.

In our simulation study and data application we used the Fourier basis $\{\psi_1(t) = 1, \psi_{2\ell+1}(t) = \sqrt{2}\cos(2\ell\pi t), \psi_{2\ell}(t) = \sqrt{2}\sin(2\ell\pi t)$, for $\ell \geq 1\}$, which is flexible for smooth functions (see also Ramsay and Silverman, 2005). This choice is mainly motivated by the Fourier computational advantage and by their rigorous theoretical study in the literature. For differentiable functions μ_d, with bounded derivative in absolute value, the basis coefficients $\beta_{d,\ell}$ decay at the rate ℓ^{-1} (see Efromovich, 1999). Typically, the smoother a function is, the faster its Fourier coefficients decay to zero.

3 Main result

To derive the asymptotic distribution of the test statistic T_n^L we assume also that the covariance K^Z has finite trace, that is $\text{tr}(K^Z) = \int K^Z(t,t)\,dt = \sum_{k\geq 1}\lambda_k < \infty$ where λ_k's are the eigenvalues of K^Z; this assumption is common in the functional data literature (see Zhang and Chen, 2007; Horváth and Kokoszka, 2012). Denote by $\kappa > 0$ the number of positive eigenvalues λ_k; $\kappa = \infty$ if all the eigenvalues are positive.

Theorem 3.1. *Assume that A1-A3 hold. Then, under the null hypothesis H_0 we have:*

$$T_n^L \to_d \sum_{k=1}^{\kappa} \lambda_k \xi_k^T A \xi_k \tag{8}$$

where $\to_d$ denotes convergence is in distribution as $n \to \infty$ and $L \to \infty$ such that $n_d^{1/2} L^{-1} = o(1)$ for all d. Here $\xi_k \sim$ Normal$(0, I_{D-1})$ for $k \geq 1$, $A = I_{D-1} + R_B^T(q_{-D} - p_{-D})(q_{-D} - p_{-D})^T R_B$, $q_{-D} = (q_1, \ldots, q_{D-1})^T$ and $p_{-D} =$

$(p_1, \ldots, p_{D-1})^T$, *I_D is the $D \times D$ identity matrix and R_B is the Cholesky factor of B, i.e. $B = R_B R_B^T$, where $B = diag(p_1^{-1}, \ldots, p_{D-1}^{-1}) + p_D^{-1} 1_{D-1} 1_{D-1}^T$.*

The proof is in Appendix A.2. When $p_d = q_d$ for all d, which yields $A = I_{D-1}$, Theorem 3.1 implies that the distribution of T_n^L is asymptotically the same as that of a χ^2-type mixture,. Specifically, in this situation, the null asymptotic distribution of T_n^L simplifies to $\sum_{k=1}^{\kappa} \lambda_k \Xi_k$, where $\Xi_k \sim \chi^2_{D-1}$. An example of setting $p_d = q_d$ is when $M_i = M$ for all $i = 1, \ldots, n$. In the particular case $M_i = 1$, Theorem 3.1 is in agreement with the results of Zhang and Chen (2007) for the testing hypothesis that the group mean functions are equal.

The test statistic T_n^L depends on the number of basis components used for the representation of the group mean functions, L. Intuitively, L needs to be sufficiently large in order to approximate well the group mean functions; on the other hand, a large value L accumulates large stochastic noise. In practice we recommend to select L using a hard truncation approach of the Fourier coefficients; see Donoho and Johnstone (1994). Specifically, estimate L by $\widehat{L} = \text{argmin}_\ell \{\ell : |\widehat{\beta}_{d,\ell}| \leq \lambda\}$, where λ is a tuning parameter, in our application in Section 6 the choice $\lambda = 0.03 n^{-1/2}$ was used. However, this threshold should be carefully tuned in any other particular application using simulations.

Hypothesis testing (2) can be tested more generally via contrasts: Zhang and Chen (2007) discussed this problem for $M_i = 1$. For example, consider the hypothesis testing of interest

$$H_0 : C\boldsymbol{\mu}(t) \equiv \boldsymbol{\mu}_0(t), \quad \forall t \qquad \text{versus} \qquad H_a : C\boldsymbol{\mu}(t) \neq \boldsymbol{\mu}_0(t), \quad \text{for some t;} \tag{9}$$

where C is a $r \times D$ matrix of contrasts, $\boldsymbol{\mu}(t)$ and $\boldsymbol{\mu}_0(t)$ are D-dimensional vectors of mean functions, with $\boldsymbol{\mu}_0(t)$ known. Remark that as $L \to \infty$ and $n_d \to \infty$ such that $n_d^{1/2} L^{-1} = o(1)$, the limit of the asymptotic distribution of $n^{1/2}(CP_n^{-1}C^T)^{-1/2}\{C\widehat{\boldsymbol{\mu}}_L(t) - \boldsymbol{\mu}_0(t)\}$ is $AGP(0, I_r K^Z)$, where $AGP(\eta, \gamma)$ denotes an asymptotic Gaussian process with mean function $\eta(t)$ and covariance function $\gamma(t, t')$, and $P_n = \text{diag}\{n_1/n, \ldots, n_D/n\}$. Then a test statistic of the form

$$T_{n,C} = n \int_0^1 \|(CP_n^{-1}C^T)^{-1/2}\{C\widehat{\boldsymbol{\mu}}(t) - \boldsymbol{\mu}_0(t)\}\|^2 \, dt \tag{10}$$

can be used to test (9); here $\|\cdot\|$ denotes the usual Euclidean vector norm and $\widehat{\boldsymbol{\mu}}(t)$ is the D dimensional vector with group mean estimates $\widehat{\mu}_d(t)$. When the group mean functions are estimated using truncated basis function expansion, as described in Section

2.2, then $T_{n,C}$ is approximated by $T_{n,C}^{L}$; the superscript emphasizes the dependence on the truncation L. One can show that, under the regularity assumptions A1-A3 stated above and when the null hypothesis (9) holds true, then

$$T_{n,C}^{L} \to_d \sum_{k=1}^{\kappa} \lambda_k \Xi_k, \tag{11}$$

as $L \to \infty$ and $n_d \to \infty$ such that $n_d^{1/2} L^{-1} = o(1)$, where $\Xi_k \sim \chi_r^2$.

An important characteristic of both T_n^L and $T_{n,C}^L$ is that the asymptotic sampling distributions are typically unknown, because they are based on unknown quantities, such as the covariance function of $K^Z(\cdot,\cdot)$, p_d's and q_d's. In practice, one can use consistent estimators of these quantities, and substitute their value into the expression used by the asymptotic distribution. For example, Staicu et al. (2010) propose ways to obtain a consistent estimator of $K^Z(\cdot,\cdot)$ in the case of balanced design for the grid points at which the unit profiles are sampled. In such situations, we can use the estimators of the eigenvalues, $\widehat{\lambda}_k$'s and the eigenfunctions $\widehat{\Phi}_k(\cdot)$'s corresponding to $\widehat{K}^Z(\cdot,\cdot)$.

The main downside of using the asymptotic distribution of the test statistic is the poor performance for small sample sizes n_d. When the asymptotic distribution with the plug-in estimates for the parameters involved is used, the test T_n^L shows an increased Type I error rate for small/moderate sample sizes; similar performance is expected for $T_{n,C}^L$. This is due to the finite sample bias collected by terms such as $n_d^{1/2}\{\mu_d^L(t) - \mu_d(t)\}$, on which the test is based. To address this limitation, in the following, we discuss two bootstrapping procedures that allow approximation of the sampling distribution of the tests. While the description of the procedures will be tailored on the first test, T_n^L it can be easily adapted to be used for the more general test, $T_{n,C}^L$.

4 Bootstrap approximations

Bootstrap methodology has attracted recent interest in the context of functional data (see for example Cuevas et al., 2006, Hall and Keilegom, 2007, Cuevas, 2014). In this section we propose two practical bootstrap-based alternatives for the approximation of the null sampling distribution of T_n. The first method, the *single-level bootstrap*, involves resampling the subject-level data, under the assumption that the group means are equal. The second method, the *nested bootstrap*, involves two steps: first resampling the subject-level data, and second resampling the unit-profile data within the resampled subjects. The sampling strategy for the resampling at the unit level is based on the spatial block bootstrap (Lahiri, 2003, Chapter 12), and uses the spatial location of the units within the subject. The latter method may seem somewhat counter-intuitive, since the spatial covariance component does not have any effect on the asymptotic distribution of the test T_n^L. However, our simulation studies in Section 5 show that by accounting for the spatial dependence, the Type I error rate of the test, using the *nested bootstrap*, is considerably improved for small samples.

Both bootstrap approaches use the so called 'bootstrap of the residuals'. Fix $L > 0$ and let $\widehat{\mu}_d^L(t) = \sum_{\ell=1}^L \psi_\ell(t)\widehat{\beta}_{d,\ell}$ and $\widehat{\mu}_\cdot^L(t) = \sum_{\ell=1}^L \psi_\ell(t)\widehat{\beta}_{\cdot,\ell}$ be the estimate of the d^{th} group mean function and the overall mean function respectively, where $\widehat{\beta}_{d,\ell}$ and $\widehat{\beta}_{\cdot,\ell}$ are the estimated Fourier coefficients determined by (5) and (6) respectively. The selection of L will be discussed later. Denote by $\widetilde{Y}_{ijl}$ the de-trended data, which is obtained by $\widetilde{Y}_{ijl} = Y_{ijl} - \widehat{\mu}_{G(i)}^L(t_{ijl})$, for all i's and j's. It follows that the "curves" $\{\widetilde{Y}_{ijl} : 1 \leq l \leq N_{ij}\}$ have mean zero, irrespective of the i^{th} subject group membership, $G(i)$. Let $\widetilde{Y}_{ij}$ be the N_{ij}-dimensional vector with the lth element equal to $\widetilde{Y}_{ijl}$, and by $\widetilde{Y}_i$ the vector obtained by stacking $\widetilde{Y}_{ij}$ over $j = 1, \ldots, M_i$.

Single-level bootstrap. The single-level bootstrap is simply an extension of the common bootstrap for independently and identically distributed scalar random variables to independently and identically distributed random processes. We define the bootstrap set as $\mathcal{B} = \{\widetilde{Y}_i : i = 1, \ldots, n\}$. For each group $d = 1, \ldots, D$, obtain $\{\widetilde{Y}_i^{(b)} : G(i) = d\}$ by sampling with replacement n_d vectors from $\mathcal{B}$. The corresponding bootstrap sample is $Y_{ijl}^{*(b)} = \widetilde{Y}_{ijl}^{(b)} + \widehat{\mu}_\cdot^L(t_{ijl})$. The estimators of $\widehat{\beta}_{d,\ell}^{(b)}$, and $\widehat{\beta}_{\cdot,\ell}^{(b)}$ are obtained as detailed in Section 2.2 corresponding to the resample of subjects and the resampled data $Y^{*(b)}$. The test statistic is then calculated using the expressions given earlier: e.g. $T_n^{L,(b)} = \sum_{d=1}^D \sum_{\ell=1}^L \{\widehat{\beta}_{d,\ell}^{(b)} - \widehat{\beta}_{\cdot,\ell}^{(b)}\}^2$. Because the *single-level bootstrap* is based on resampling independent objects, it is not hard to check that, for fixed L, the distributions of $T_n^{L,(b)}$ and T_n^L are asymptotically the same. The null distribution of the test $T_n^{L,(b)}$ is always available, and furthermore it requires little computational cost.

The *nested bootstrap* is a more complex bootstrap approach that accounts for the spatial dependence of the random curves within a subject. It encompasses resampling at the subject level and resampling at the unit level. At the first step, a bootstrap sample is obtained using the *single-level bootstrap* technique (i.e. bootstrapping the subjects). Let $\{\widetilde{Y}_i^{(b_1)} : G(i) = d\}$, for $d = 1, \ldots, D$ be such a sample, where the superscript

(b_1) emphasizes the use of the *single-level bootstrap*. At the second step, we propose to further resample the subject-level data for each selected subject by using the spatial locations of the unit-level profiles; we do this by employing a method inspired by block bootstrapping, a standard technique for dependent data such as time series or spatial data (Lahiri, 2003, Chapter 12). We describe this approach next.

Denote by $\mathcal{B}_i = \{\widetilde{Y}_{ij}^{(b)} : j = 1, \ldots, M_i\}$ the set of unit-level profiles, and by $\mathcal{S}_i = \{s_{i1}, \ldots, s_{iM_i}\}$ the set of unit locations corresponding to subject i of the single-level bootstrap sample. The basic idea is first to resample the unit locations by using block-bootstrapping, and second to form the bootstrap samples of the subject-level data by collecting the unit-level profiles that correspond to the selected sample of unit locations. For simplicity, consider the case when the spatial domain is $\mathcal{D} = [0, S) \subset \mathcal{R}$ and we refer the reader to Lahiri (2003, Chapter 12) for general sampling regions. Let $b_u > 0$ be some constant, commonly known as 'block length', and construct M_i' overlapping blocks, of length b_u, $B_i(j) = [s_{ij}, s_{ij} + b_u)$, and such that $s_{ij} + b_u \leq S$. Corresponding to each block $B_i(j)$, define the set of spatial locations included in this block as $J_i(j) = \{s_{ij'} \in \mathcal{S}_i : s_{ij} < s_{ij'} < s_{ij} + b_u\}$. To account for possible sparsity in the sampling spatial sampling design we consider only the sets $J_i(j)$ for which their cardinality is at least 6, that is $|J_i(j)| \geq 6$. With little abuse of notation assume there are M_i' such pairs. Let $n_{S,b_u} = \lfloor S/b_u \rfloor$. Then to construct the bootstrap sample for subject i, a number of n_{S,b_u} blocks $B_i(j^*)$ are drawn with replacement from $\{B_i(j) : j \in M_i'\}$. Denote the blocks obtained by $B_i(j_1^{(b_2)}), \ldots, B_i(j_{n_{S,b_u}}^{(b_2)})$; the blocks of unit locations will be aligned in the order they were picked and such that the ℓ^{th} selected sample to start at $(\ell - 1)b_u$ for $1 \leq \ell \leq n_{S,b_u}$. Corresponding to the selection of blocks, let $S_i^{(b_2)} = \{s_{ij}^{(b_2)} : 1 \leq j \leq M_i^{(b_2)}\}$ be the bootstrap of sample of unit locations.

The corresponding resample of the i^{th} subject data is obtained by collecting the trajectories $\{Y_{ijl}^{(b_1)} : 1 \leq l \leq N_{ij}\}$ according to the sample locations s_{ij} that are included in the selected bootstrap samples $J_i(j_\ell^{(b_2)})$, for all ℓ's; denote by $[\{Y_{ijl}^{(b)} : 1 \leq l \leq N_{ij}\} : j = 1, \ldots, M_i^{(b)}]$ the bootstrap de-trended data, using the two-step procedure. The nested bootstrap sample is then $\widetilde{Y}_{ijl}^{*(b)} = \widetilde{Y}_{ijl}^{(b)} + \widehat{\mu}_{\cdot}^{L}(t_{ijl})$.

One may argue that standard block bootstrap produces replicates which are non-smooth near the joint-points, and thus the use of such approach in our context may be debatable. However, even with non-smooth replicates, the block bootstrap is known to perform better than the independently and identically distributed bootstrap, and this is what motivated us to apply it at the unit level data. The selection of the block length, b_u, is another issue one has to consider. The most prominent selection rule for the optimal block length in standard block bootstrap is the method by Hall et al. (1995); for a more complete list of methods see Lahiri (2003, Chapter 7). However, no results exist for irregularly spaced spatial data, to the best of our knowledge. In the simulation study and data application we consider an ad-hoc criterion and determine b_u by requiring that there are at least 6 blocks per subject of cardinality 6 or more. The performance of the two bootstrap approaches is investigated numerically in several simulation scenarios, as illustrated in the next section.

5 Simulation studies

We conducted a simulation study to investigate the finite sample performance of the test. In this section we summarize the main findings based on data sets, each consisting of $D = 3$ groups of $n_d = 10, 15, 30, 50$ subjects per group, and $M_i = 20$ units per subject. The unit-level profiles correspond to a grid of equidistant $N_{ij} = N_{i1}$ points in $[0, 1]$, and N_{i1} are generated uniformly between 27 and 36. Each data set is generated from the model $Y_{ijl} = \mu_{G(i)}(t_{ijl}) + Z_i(t_{ijl}) + W_{ij}(t_{ijl}) + U_i(s_{ij}) + \epsilon_{ijl}$, and $G(i) = \lfloor (i-1)/n_d \rfloor$ under all the possible combinations from the following scenarios:
Scenario A: (i) $\mu_d(t) = 4.2 + cos(2t\pi)$ for $d = 1, 2, 3$; (ii) $\mu_d(t) = 4.2 + cos(2t\pi) + 1(d = 2)0.5t$ for $d = 1, 2, 3$; (iii) $\mu_d(t) = 4.2 + cos(2t\pi) + 1(d = 2)0.5$ for $d = 1, 2, 3$. Case (Ai) corresponds to a situation where the null hypothesis H_0 is true; cases (Aii) and (Aiii) describe situations where the null hypothesis is false, and the departure from the null is moderate and stronger respectively. In particular, the mean functions in (Aii) are separated by at most a linear trend, while they are separated by a constant trend in case (Aiii).

Scenario B: $Z_i(t) = \sum_{k\geq 1} \lambda_{Z,k}^{1/2} \xi_{i,k} \phi_{Z,k}(t)$ where $\lambda_{Z,1} = 0.5$, $\lambda_{Z,2} = 0.125$, and $\lambda_{Z,k} = 0$ otherwise and (i) $\phi_{Z,1}(t) = 1$, $\phi_{Z,2}(t) = \sqrt{2}\sin(2\pi t)$; (ii) $\phi_{Z,1}(t) = \sqrt{3}(2t-1)$, $\phi_{Z,2}(t) = \sqrt{5}(6t^2 - 6t + 1)$. We take $W_{ij}(t) = \sum_{k\geq 1} \lambda_{W,k}^{1/2} \zeta_{ij,k} \phi_{W,k}(t)$, where $\lambda_{W,1} = 0.33$, and $\lambda_{W,2} = 0.11$, and $\lambda_{W,k} = 0$ otherwise, and $\phi_{W,1}(t) = 1$, $\phi_{W,2}(t) = \sqrt{7}(20t^3 - 30t^2 + 12t - 1)$. The random coefficients $\{\xi_{i,k}\}_{i,k}$ and $\{\zeta_{ij,k}\}_{i,j,k}$ are assumed mutually uncorrelated, and identically distributed as standard normal variables. In addition $\epsilon_{ijl} \sim \text{Normal}(0, 0.1)$.

Scenario C: U_i is stationary Gaussian process with mean 0, variance 0.5 and autocorrelation function specified by Matérn function defined by $\rho(\Delta; \phi, \nu) = 2^{1-\nu}\{\Gamma(\nu)\}^{-1} (\Delta/\phi)^{\nu} K_{\nu} (\Delta/\phi)$ where ϕ and ν are the unknown parameters and K_ν is the modified Bessel function of order ν (see Stein, 1999). We consider $\nu = 1.5$, $\phi = 70$; this corresponds to a setting where the correlation is negligible (i.e. with values smaller than 0.003) for $\Delta > 640$. It is assumed a uniform sampling design for the units, that is the unit locations are independent and identically distributed as Uniform $[0, S]$, where $S = 15,000$.

For each setting, obtained by combining the above scenarios, we test the null hypothesis that the group mean functions $\mu_d(\cdot)$ are equal. Cuevas et al. (2004) cannot be applied directly to test this hypothesis, because of three reasons: 1) the curves are observed at discrete points, 2) the number of grids per curve is varying, and 3) the observations include measurement error. The testing procedure proposed by Zhang and Chen (2007) via contrasts, and assuming a working independence among all the curves, is highly misleading and results in very large Type I errors, because the variability of the test statistic under the null assumption is under estimated. Due to these considerations we do not pursue these two approaches in our study. We carry this hypothesis testing using our proposed test statistic T_n^L. The distribution of the test statistic T_n^L is approximated using the single-level and nested bootstrap with $B = 1000$ bootstrap samples. We examine the Type I error rate corresponding to the significance levels $\alpha = 0.01, 0.05, 0.10, 0.15, 0.20$ when the group mean functions are equal, and investigate the power at these levels, when the group mean functions are different, under different covariance structures and for various sample sizes. For each data set, the size and power probabilities are based on estimated tail probabilities $P(T_n^L > t_n^{L,0})$, where P is the null distribution of T_n^L as approximated by single-level or nested bootstrap, and $t_n^{0,L}$ is the observed value of the test, corresponding to the particular data set. The size of the test corresponding to a nominal level α is then estimated by $\sum_{k=1}^{N_{sim}} 1\{P(T_n^L > t_n^{L,0}) \leq \alpha\}/N_{sim}$, presuming the data are generated under the null hypothesis, where N_{sim} is the number of simulated data sets.

In all the simulations the number of Fourier basis function is set to $L = 9$ - for our setting this choice corresponds to undersmoothing the group mean functions. For com-

Table 1: Estimated Type I error rate of T_n^L (and $L = 9$) using single-level (SB) / nested bootstrap (NB), for various group sizes n_d and significance levels. The data sets are generated using mean functions specified by (Ai) and covariance functions described by (Bi) and (C).

$100\alpha\%$	1%	5%	10%	15%	20%
n_d	SB / NB	SB / NB	SB / NB	SB / NB	SB / NB
10	2.47 / 1.80	9.27 / 7.40	14.97 / 12.57	19.67 / 16.90	24.80 / 21.40
15	2.23 / 1.50	7.20 / 5.67	12.50 / 10.47	18.33 / 15.30	23.20 / 20.30
30	1.33 / 0.93	5.67 / 4.33	10.87 / 8.37	16.23 / 13.47	21.33 / 18.47
50	1.27 / 0.83	5.20 / 4.00	10.40 / 8.27	15.50 / 13.03	20.37 / 16.97

Table 2: Estimated power of T_n^L (and $L = 9$) using single-level (SB) / nested bootstrap (NB), for various group sizes n_d and significance levels. The data sets are generated using mean functions specified by (Aii), block column M_1, and (Aiii), block column M_2, and covariance functions described by (Bi) and (C).

	$100\alpha\%$	1%	5%	10%	15%	20%
Model	n_d	SB / NB	SB / NB	SB / NB	SB / NB	SB / NB
M_1	10	6.2 / 5.0	16.8 / 13.9	27.2 / 22.8	34.4 / 30.2	39.9 / 36.6
	15	8.3 / 6.2	18.8 / 16.0	29.4 / 25.5	38.9 / 34.5	45.8 / 41.7
	30	14.7 / 11.5	33.3 / 29.9	46.3 / 41.0	54.0 / 50.2	60.5 / 58.1
	50	24.7 / 18.9	46.9 / 43.1	60.6 / 55.4	69.5 / 65.4	76.3 / 72.9
M_2	10	20.1 / 18.1	37.4 / 33.6	48.6 / 45.4	55.7 / 52.2	63.4 / 59.4
	15	29.9 / 24.4	49.8 / 45.9	61.1 / 56.6	68.0 / 64.4	73.8 / 71.0
	30	57.8 / 53.3	76.8 / 72.5	85.6 / 83.2	89.3 / 87.4	91.9 / 90.1
	50	84.1 / 80.1	94.5 / 92.9	97.2 / 96.7	98.4 / 97.9	98.9 / 98.4

parison we also examined the results corresponding to $L = 3$ and $L = 15$ and observed that they barely change; for example in the case $n_d = 10$, the estimated type I error rate for 10% nominal level is 15.00% and 14.97% respectively for single-level bootstrap and 12.60% and 12.53% respectively for the nested bootstrap; the results remain unchanged when the nominal level equals 1% or 5%. Generally, the number of basis functions L is a tuning parameter and its selection can be compared to the selection of a smoothing parameter in the context of penalized splines.

Table 1, presents the estimated Type I error of the test using the two bootstrap approaches, based on $N_{sim} = 3000$ generated data sets with mean functions specified by (Ai) and covariance functions specified by (Bi) and (C); results corresponding to a covariance function described (Bii) are similar and are omitted out of brevity. Several nominal sizes α and group sample sizes n_d are investigated. The results emphasize that single-level bootstrap performs well for moderate and large sample sizes, $n_d = 30$ or $n_d = 50$ confirming the theoretical expectations. However, it gives an inflated Type I error when the group sample sizes are smaller like $n_d = 10, 15$. On the other hand, the nested bootstrap has an excellent performance, particularly for smaller group sample sizes, in having a type I error close to the nominal level. The estimated Type I error rate with the nested bootstrap is much improved over the single-level bootstrap: compare the results for $n_d = 10$ and $n_d = 15$ obtained with both types of bootstrap. For moderate or larger group sample sizes, both bootstrap procedures work well in terms of accurately estimating a Type I error rate of the test, with the single-level bootstrap tending to be more liberal, while the nested bootstrap more conservative. The block size for the nested bootstrap was fixed to 4192 - a value determined by requiring that there are at least 6 blocks per subject of cardinality 6 or more; all the simulation results are based on this value.

Table 2, in the blocks labeled M_1 and M_2, gives the estimated power of the test using the two bootstrap approaches, based on $N_{sim} = 1000$ generated data sets with mean functions specified by (Aii) and (Aiii) respectively, and the covariance functions of the random components specified by (Bi) and (C). As expected, from the analysis of the Type I error, the power of the test with single-level bootstrap is larger comparative to when nested bootstrap is used. The difference decreases as the sample size increases or the departure from the null hypothesis is stronger.

6 Data Analysis

The proposed testing procedure was applied to a long range infrared light detection and ranging (LIDAR) study, with the objective to test whether the backscatter efficiency is affected by the type of aerosol clouds. The study comprises measurements of the spectral backscatter taken at different time periods and corresponding to various CO_2 laser wavelengths for two types of clouds: control clouds that were non-biological in nature and treatment clouds that were biological. This is an example where the biological clouds may be a threat (perhaps from a terrorist) while the non-biological ones are benign. So there is a lot of interest in knowing whether the two types of clouds are different. The data have been previously described in Carroll et al. (2012) and discussed recently in Serban et al. (2013) and Xun et al. (2013).

In the experiment, 30 aerosol clouds are investigated: control clouds that were non-biological in nature and treatment clouds that were biological. For each cloud $i = 1, \dots, 30$ at 50 time periods (called bursts), which are sampled at one second apart, and various CO_2 laser wavelengths, the background corrected received signal is observed at 250 equally spaced range values. Here we concentrate on the (range invariant) backscatter efficiency of the true signal as estimated using the algorithm of Warren et al. (2008, 2009), but applied to the observed data rather than the deconvolved data.

Because of physical properties, the backscatter efficiency can be viewed as a function of the wavelength for each burst (see Serban et al., 2013). Define the response Y_{ijl} as the backscatter efficiency for CO_2 laser wavelength t_{ijl} for the jth burst sampled at s_{ij} within cloud i. The burst level profiles are sampled at regular wavelengths $t_{ijl} \in \{1, \ldots, 19\}$ and the measurements are likely contaminated with measurement error. Furthermore because of the nature of the bursts, the dependence among responses for the same cloud depends on the relative location of the bursts, not through the mean, but rather through the covariance, as a function of the distance between the burst locations. It is reasonable to assume that the spectral backscatter can be modeled using (1), where it is assumed that the cloud-type specific mean trend and covariance structure are of the form (3). Figure 1 shows the spectral backscatter profiles for all the clouds in the control group and for all the clouds in the treatment group.

We are interested to test the null hypothesis that the mean backscatter efficiency for the control and treatment group are equal, i.e. $\mu_{ctrl}(t) \equiv \mu_{trt}(t)$ for all wavelengths t. Hitherto, there are no available approaches to test this null hypothesis, when data exhibit this complex correlation structure. The proposed testing approach was applied and the number of Fourier coefficients was allowed to vary between L = 3 and L = 19. In our case $D = 2$ and $M_i = M_1$ and thus the two testing procedures T_n and $T_{n,C}$ for $C = (1, -1)$ agree with one another and, not surprisingly, their null distribution is the same. The value of the test statistic ranges from 0.0046 when $L = 3$, to 0.0071 when $L = 6$ and to 0.0081 when $L = 19$. The p-value is estimated using the three approximations of the null distribution of the test: via single bootstrap and nested bootstrap with B = 10,000 replications, and by using the null distribution with the estimated model components.

Figure 2 shows that the p-value varies from 0.023 to 0.056 for the single-level and from 0.016 to 0.044 for different truncation levels $L = 3, \ldots, 19$. Using the hard truncation criterion described earlier, we obtain L = 17, the test statistic value is 0.008 and the corresponding p-values equal to 0.028 and 0.021 for the singlelevel and nested bootstrap, respectively. For the nested bootstrap, the size of the block bootstrap was fixed at 0.2; further investigation of the analysis for varying block size between 0.15 to 0.4 indicates that the overall results remain roughly the same.

Finally, we considered the approximation of the null distribution of the T_n^L given by (8) with the eigenvalues of K^Z, λ_k, replaced by their estimated values $\widehat{\lambda}_k$. In our case $D = 2$ and $M_i = M_1$, thus the null distribution of T_n^L is $\sum_{k\geq 1} \lambda_k \chi_1^2$. We use the estimation algorithm proposed by Staicu et al. (2010) to compute the estimated covariance $\widehat{K}^Z$, and thus the estimated eigenvalues $\widehat{\lambda}_k$. Using a percentage of explained variance equal to 0.95 we obtain 3 positive eigenvalues of $\widehat{K}^Z$, $\widehat{\lambda}_1 = 0.0013, \widehat{\lambda}_2 = 0.0003, \widehat{\lambda}_3 = 0.0001$. The p-value with this approach is 0.016. All the results indicate strong evidence of significant differences between the backscatter efficiency mean trend corresponding to the two types of clouds.

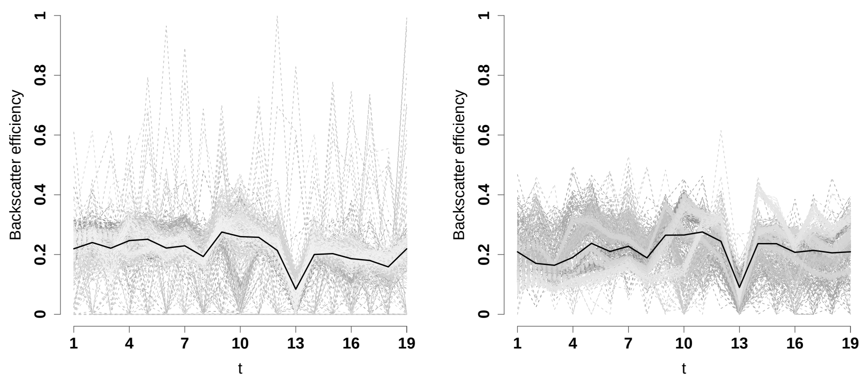

Figure 1: Backscatter efficiency profiles for the 16 types of clouds in the control group (left panel) and for the 14 clouds in the treatment group (right panel) and al their bursts; same color is used for the responses measured on the same cloud. Overlayed in solid line is the group mean profile obtained using $L = 17$.

7 Discussion and Extensions

The present paper develops testing procedures for assessing the equality of the group mean functions of several groups of curves, when the data have a multilevel structure of the form groups-subjects or clusters-units with the unit-level profiles being spatially correlated. We show that the asymptotic distribution of the significance tests depends solely on the subject level covariance, provided that an analysis of variance-like decomposition of the functional processes according to the levels of hierarchy, subject and unit, and the spatial correlation holds. The lack of dependence of the asymptotic distribution of the tests statistics on the unit-level profiles may seems surprising. Intuitively, it appears

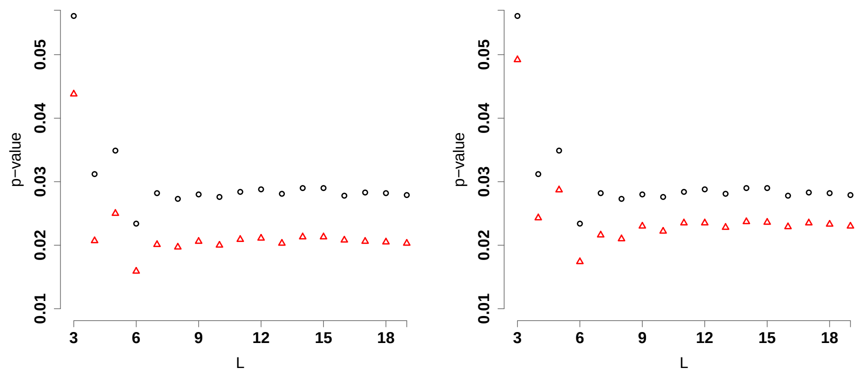

Figure 2: P-values for various truncations L for testing the null hypothesis that all the group means are equal. Displayed are results with single-level bootstrap (circle symbol) and nested bootstrap (triangle). Size of the block bootstrap is fixed at 0.2 (left panel) and 0.4 (right).

to be the result of the combination between the model assumption for the covariance structure and the increasing domain asymptotics used to handle the spatial dependence. Such assumptions work well for settings similar to our data application. However, the asymptotic distribution of the tests would most likely change structurally, by using infill asymptotics, which intuitively means that as number of the unit-level curves increases, the correlation between them also increases, under a stationary spatial dependence assumption.

Bootstrap alternatives are discussed and in particular a novel block bootstrap procedure for functional data is proposed, which accounts for the spatial dependence between the curves. The block bootstrap, referred to as nested bootstrap, provides a very accurate approximation of the null distribution of the test, and in particular for small sample sizes. For such sample sizes, the regular bootstrap, referred to as single-level bootstrap has poor performance and yields inflated Type I error rates. For larger sample sizes the nested bootstrap has a good performance and tends to be more conservative. The challenge with using the block bootstrap approach is the selection of the block length, a challenge inherited from the classical block bootstrap for spatial statistics (see Lahiri 2003, Chapter 7). While three has been considerable research in optimal selection of the size of the dependent bootstrap (see for example Lahiri, 1999, Patton, et al., 2009) on optimal selection of the block bootstrap, based on data, remains an open problem.

Acknowledgement

Staicu's research was supported by U.S. National Science Foundation grant number DMS 1007466. Lahiri's research was partially supported by National Foundation grants number DMS 0707139 and DMS 1007703. Carroll's research was supported by a grant from the National Cancer Institute (R37-CA057030). This publication is based in part on work supported by Award Number KUS-CI-016-04, made by King Abdullah University of Science and Technology (KAUST).

Appendix

This Appendix contains three sections. Section A.1 discusses asymptotic properties of the basis coefficients estimators in terms of consistency and rate of convergence. Section A.2 gives the asymptotic distribution of the testing procedure, including the proof of Theorem 3.1.

A.1 Asymptotic properties of the basis functions estimators

Recall the model assumption is $Y_{ijl} = \mu_{G(i)}(t_{ijl}) + V(t_{ijl}, s_{ij}) + \epsilon_{ijl}$, for $G(i) \in \{1, \ldots, D\}$, where $t_{ijl} \in \mathcal{T}$ and $s_{ij} \in \mathcal{D}$, $l = 1, \ldots, N_{ij}$, $j = 1, \ldots, M_i$, $i = 1, \ldots, n$. For simplicity we set $\mathcal{T} = [0, 1]$ and $\mathcal{D} = R^2$. It is assumed that ϵ_{ijl} is white noise with mean zero and constant variance σ_ϵ^2, for all l, j, and i.. Without log of generality we assume that there are $N_{ij} + 1$ observations per curve instead of N_{ij}; this assumption is made to simplify notations.

Consistency of the estimators

We develop the consistency of the estimators of the basis functions coefficients.

Lemma A.1. *Assumptions A1-A3 hold. Then as $n_d \to \infty$, we have $|\widehat{\beta}_d - \beta_d| = o_p(1)$.*

Proof of Lemma A.1: We show that $\widehat{\beta}_{d,\ell} - \beta_{d,\ell} = o_p(1)$ for all $\ell \geq 1$. Because $E\{|\widehat{\beta}_{d,\ell} - \beta_{d,\ell}|^2\} = \text{var}(\widehat{\beta}_{d,\ell}) + \{E(\widehat{\beta}_{d,\ell}) - \beta_{d,\ell}\}^2$ it is suffices to show that $\text{bias}(\widehat{\beta}_{d,\ell}) \to 0$ and $\text{var}(\widehat{\beta}_{d,\ell}) \to 0$, as $n_d \to \infty$,

Let $\mu_{d,ij}(t) = \sum_{l=1}^{N_{ij}} \mu_d(t_{ijl}) 1\{t \in A_{ijl}\}$, for all i, j, d. Recall that $\{t_{ijl} : l = 1, \ldots, N_{ij}+1\}$ are considered equally spaced in $[0, 1]$ with $t_{ijl} = (l-1)/N_{ij}$ and $A_{ijl} =$

$[t_{ijl}, t_{ij,l+1})$. We write $\widehat{\beta}_{d,\ell} = \widetilde{\beta}_{d,\ell} + I^V_{d,\ell} + I^\epsilon_{d,\ell}$, where $\widetilde{\beta}_{d,\ell} = m_d^{-1}\sum_{\{i:G(i)=d\}}\sum_{j=1}^{M_i}\int_0^1 \mu_{d,ij}(t)\psi_\ell(t)\,dt$, $I^V_{d,\ell} = m_d^{-1}\sum_{\{i:G(i)=d\}}\sum_{j=1}^{M_i}\sum_{l=1}^{N_{ij}} V_{ij}(t_{ijl}, s_{ij})a_{ijl,\ell}$, $I^\epsilon_{d,\ell} = m_d^{-1}\sum_{\{i:G(i)=d\}}\sum_{j=1}^{M_i}\sum_{l=1}^{N_{ij}}\epsilon_{ij}(t_{ij,\ell})a_{ijl,\ell}$ and $a_{ijl,\ell} = \int_{A_{ijl}}\psi_\ell(t)\,dt$. It is sufficient to show that $|\widetilde{\beta}_{d,\ell} - \beta_{d,\ell}| = o(n_d^{-1/2})$ and $\text{var}\{\widehat{\beta}_{d,\ell}\} = O(n_d^{-1})$.

To simplify notation, in what follows we assume that all integrals without a range of integration are over A_{ijl}. We show first that $|\widetilde{\beta}_{d,\ell} - \beta_{d,\ell}| = o(n_d^{-1/2})$ as $n_d \to \infty$, using the following inequalities:

$$|\widetilde{\beta}_{d,\ell} - \beta_{d,\ell}| = \left| m_d^{-1}\sum_{\{i:G(i)=d\}}\sum_{j=1}^{M_i}\sum_{l=1}^{N_{ij}}\int \{\mu_d(t) - \mu_d(t_{ijl})\}\psi_\ell(t)\,dt \right| \le \frac{C}{\sqrt{2\alpha+1}} n_d^{-\alpha\theta},$$

which is of order $o(n_d^{-1/2})$ as $\lim n_d = \infty$, since $\theta\alpha > 1/2$ and A3(c) holds. This sequence of equalities and inequalities uses assumption A1(a) as well as Hölder's and Cauchy's inequalities.

Next, we show that $\text{var}\{\widehat{\beta}_{d,\ell}\} = O(n_d^{-1})$. For simplicity we discuss first the error term, $I^\epsilon_{d,\ell}$, and then the 'random' term $I^V_{d,\ell}$.

Error term, $I^\epsilon_{d,\ell}$: We show that $I^\epsilon_{d,\ell} = o_p(n_d^{-1/2})$.

$$I^\epsilon_{d,\ell} = m_d^{-1}\sum_{\{i:G(i)=d\}}\sum_{j=1}^{M_i}\sum_{l=1}^{N_{ij}}\epsilon_{ijl}\int_{A_{ijl}}\psi_\ell(t)\,dt = O_p(m_d^{-1/2}n_d^{-1/2}),$$

which is of order $o_p(n_d^{-1/2})$ as $\lim n_d = \infty$. Thus $\text{var}(I^\epsilon_{d,\ell}) = o_p(n_d^{-1})$.

Covariance term, $I^V_{d,\ell}$: Using the ANOVA-like decomposition of the bivariate process $V_{ij}(t,s)$, we decompose the term $I^V_{d,\ell}$ into the components $I^V_{d,\ell} = I^Z_{d,\ell} + I^W_{d,\ell} + I^U_{d,\ell}$, using the modeling assumption (3). We will show that $\text{var}(I^Z_d) = O(n_d^{-1})$, $\text{var}(I^W_d) = O(m_d^{-1})$, and $\text{var}(I^U_d) = O(m_d^{-1})$.

Claim: var$(I^Z_{d,\ell}) = O(n_d^{-1})$ for all ℓ.

We write $\text{var}(I^Z_{d,\ell}) = m_d^{-2}\sum_{\{i:G(i)=d\}}E(F^2_{i\ell})$, where $F_{i\ell} = \sum_{j=1}^{M_i}\sum_{l=1}^{N_{ij}} Z_i(t_{ijl})a_{ijl,\ell}$. Let $K^Z(\cdot,\cdot)$ be the covariance function of $Z(\cdot)$ defined by $K^Z(t,t') = \text{cov}\{Z(t), Z(t')\}$. Then

$$\text{cov}(I^Z_{d,\ell}, I^Z_{d,\ell'}) \le m_d^{-2}\sum_{\{i:G(i)=d\}}\sum_{j=1}^{M_i}\sum_{j'=1}^{M_i}\sum_{l=1}^{N_{ij}}\sum_{l'=1}^{N_{ij'}}\left|a_{ijl,\ell}a_{ij'l',\ell'}K^Z(t_{ijl}, t_{ij',l'})\right|$$
$$= O(n_d^{-1}),$$

since $K^Z(\cdot,\cdot)$ is uniformly bounded, from A2(a). The last equality is also based on $\min_{\{i:G(i)=d\}} M_i \le M_i \le c_2 \min_{\{i:G(i)=d\}} M_i$, which follows from assumption A3(b).

Claim: var$(I^W_{d,\ell}) = O(n_d^{-1})$ for all ℓ.

In fact we can show a much stronger result, namely that var$(I^W_{d,\ell}) = O(m_d^{-1})$. The reasoning is similar to the above, with few modifications. First we write var$(I^W_{d,\ell}) = m_d^{-2}\sum_{i:G(i)=d}\sum_{j=1}^{M_i} E(F^2_{ij\ell})$, where $F_{ij\ell} = \sum_{l=1}^{N_{ij}} W_{ij}(t_{ijl})a_{ijl,\ell}$. It is easy to show that cov$(I^W_{d,\ell}, I^W_{d,\ell'}) = O(m_d^{-1})$, again using assumption A2(a). Here $\|K^W\| = \sup_{t,t'\in[0,1]} K^W(t,t') < \infty$, following the assumption that $\int_0^1 K^W(t,t)\,dt < \infty$.
Claim: var$(I^U_{d,\ell}) = O(n_d^{-1})$ for all ℓ.

In fact we will prove a stronger result that var$(I^U_{d,\ell}) = O(m_d^{-1})$.
Let var$(I^U_{d,\ell}) = m_d^{-2}\sum_{i:G(i)=d} E\{Var(H^2_{i\ell}|\mathbf{S}_i)\}$, where $\mathbf{S}_i$ is the M_i-dimensional vector of S_{ij}, $H_{i\ell} = \sum_{j=1}^{M_i} U_i(S_{ij})a_\ell$, $a_\ell = \int_0^1 \psi_\ell(t)\,dt$; notice $a_\ell = 1$ for $k = 1$ and 0 otherwise. We have that $E\{Var(H^2_{i\ell}|\mathbf{S}_i)\} = \sum_{j=1}^{M_i}\sigma_U(0) + E\{\sum_{j=1}^{M_i}\sum_{j'=1,j'\neq j}^{M_i}\sigma_U(\|S_{ij} - S_{ij'}\|)\} = M_i\sigma_U(0) + (M_i-1)\int\sigma_U(\Delta)f(s)f(s+\Delta)\,ds\,d\Delta$. Hence var$(I^U_{d,\ell})$ equals

$$m_d^{-1}\sigma_U(0) + m_d^{-2}\sum_{\{i:G(i)=d\}}(M_i-1)\int\int\sigma_U(\Delta)f(s)f(s+\Delta)\,ds\,d\Delta = O(m_d^{-1}), \quad (A.1)$$

since $|\int\sigma_U(\Delta)f(s)f(s+\Delta)\,ds\,d\Delta| < \infty$ following the assumption that the density $f(\cdot)$ and the covariance function $\sigma_U(\cdot)$ are non-zero on a finite interval. Here we used that for each i, the spatial locations S_{ij} are independent and identically distributed with density function $M_i^{-1/2}f(M_i^{-1/2}s)$ and the result that if S_1 and S_2 are independent and identically distributed with density function $f(\cdot)$, then $g(\Delta) = \int f(s)f(s+\Delta)\,ds$ is the density function $S_1 - S_2$.

Asymptotic normality of the estimators

Next we show the asymptotic normality of $n_d^{1/2}(\widehat{\beta}_d - \beta_d)$ for fixed but arbitrary truncation L.

Lemma A.2. *Suppose the assumptions A1-A3 hold. Let L be fixed truncation, and denote by $\beta_d = (\beta_{d,1}, \ldots, \beta_{d,L})^T$, and by $\widehat{\beta}_d$ its analogous estimator, by suppressing the dependence on L. Then, for $n_d = |\{i : G(i) = d\}|$, we have that as $n_d \to \infty$, $n_d^{1/2}(\widehat{\beta}_d - \beta_d) \to$ Normal$(0, \Sigma)$, where Σ is $L \times L$ matrix with (ℓ, ℓ') element equal to*

$$\int_0^1\int_0^1 \psi_\ell(t)\psi_{\ell'}(t')K^Z(t,t')\,dt\,dt'. \quad (A.2)$$

The proof is based on the remark that $\widehat{\beta}_d - \beta_d = (\widetilde{\beta}_d - \beta_d) + I^Z_d + (I^W_d + I^U_d + I^\epsilon_d)$, and that 1) $(\widetilde{\beta}_{d,\ell} - \beta_d) = o(n_d^{-1/2})$, and 2) each of I^W_d, I^U_d, I^ϵ_d are $o_p(n_d^{-1/2})$. It follows that the asymptotic distribution of $n_d^{1/2}(\widehat{\beta}_d - \beta_d)$ is the same as that of $n_d^{1/2}I^Z_d$. Lemma A.3 shows that $n_d^{1/2}I^Z_d \to$ Normal$(0, \Sigma)$.

Using this result, one can derive the asymptotic distribution of the group mean function estimator of $\mu_d(t)$, $\widehat{\mu}_d^L(t) = \sum_{\ell=1}^{L} \widehat{\beta}_{d,\ell}\psi_\ell(t)$. In particular, as $L \to \infty$ and provided that $\max_{\ell \geq L+1} |\beta_{d,\ell}| = o(n_d^{-1/2})$, it follows that the limiting distribution of $n_d^{1/2}\{\widehat{\mu}_d^L(t) - \mu_d(t)\}$ is $AGP\{0, K^Z\}$, where $AGP(\eta, \gamma)$ denotes an asymptotic Gaussian process with mean function $\eta(t)$ and covariance function $\gamma(t, t')$. The assumption $\max_{\ell \geq L+1} |\beta_{d,\ell}| = o(n_d^{-1/2})$ is related to the rate of decay of the Fourier coefficients $\beta_{d,\ell}$; this assumption ensures that the group mean estimators $\widehat{\mu}_d^L(t)$ are unbiased asymptotically.

Lemma A.3. *Suppose that $Z_i(\cdot)$ are independent and identically distributed as the stochastic process $Z(\cdot)$ for which assumptions A2(a) and A3(b, c) hold. Then, as $n_d \to \infty$ we have*

$$n_d^{1/2}(I_{d,1}^Z, \ldots, I_{d,L}^Z)^T \to Normal(0, \Sigma), \tag{A.3}$$

where the convergence is in distribution and Σ is $L \times L$ matrix defined above.

Proof of Lemma A.3: We will show this is two steps. In Step 1 we prove that $n_d^{1/2}(I_{d,\ell}^Z - \widetilde{I}_{d,\ell}^Z)$ converges in probability to zero, where $\widetilde{I}_{d,\ell}^Z = m_d^{-1}\sum_{\{i:G(i)=d\}} M_i \int_0^1 Z_i(t)\psi_\ell(t)$. In Step 2 we show that $n_d^{1/2}\widetilde{I}_d^Z \to \text{Normal}(0, \Sigma)$. The result then follows by an application of the Slutsky's theorem.

We begin with proving Step 1. Our proof relies on the assumption that the covariance function of Z_i is twice continuously differentiable. Then by a Taylor expansion

$$K^Z(t, t') = K^Z(t_1, t_2) - K^Z(t, t_2) - K^Z(t_1, t') + K_{t,t'}^Z(t_1, t_2)(t - t_1)(t' - t_2) + o\{N_{min}^{-2}\},$$

for $|t - t_1| < N_{min}^{-1}$ and $|t' - t_2| < N_{min}^{-1}$. Here $K_{t,t'}^Z(t_1, t_2) = \partial^2 K^Z(t_1, t_2)/\partial t \partial t'$.

Let $\ell \geq 1$ be arbitrary. It is sufficient to show that $E[n_d(I_{d,\ell}^Z - \widetilde{I}_{d,\ell}^Z)^2] \to 0$. Let $V_{ij}(t) = \sum_{l=1}^{N_{ij}} Z_i(t_{ijl}) 1(t \in A_{ijl})$. Simple algebra gives that $E[n_d(I_{d,\ell}^Z - \widetilde{I}_{d,\ell}^Z)^2] = n_d m_d^{-2}\sum_{\{i:G(i)=d\}} \sum_{j=1}^{M_i} \sum_{j'=1}^{M_i} S_{jj'}$, where

$$\begin{aligned} |S_{jj'}| &\leq \|K_{t_1,t_2}^Z\| \sum_{l=1}^{N_{ij}} \sum_{l'=1}^{N_{ij'}} \int_{A_{ijl}} \int_{A_{ij'l'}} |(t - t_{ijl})(t' - t_{ij',l'})||\psi_\ell(t)||\psi_\ell(t')|\, dt\, dt' + o(N_{d,min}^{-2}) \\ &= \||K_{t_1,t_2}^Z\| N_{ij}^{-1} N_{ij'}^{-1}/3, \end{aligned} \tag{A.4}$$

for $N_{d,min} = \min\{N_{ij} : G(i) = d\}$, and $\sup\{|K_{t_1,t_2}^Z(t, t')| : 0 \leq t, t' \leq 1\} = \|K_{t_1,t_2}^Z\| < \infty$. We obtain that $|E[n_d(I_{d,\ell}^Z - \widetilde{I}_{d,\ell}^Z)^2]| \leq C N_{d,min}^{-2} n_d \sum_{\{i:G(i)=d\}} M_i^2/m_d^2$ which converges to zero as $n_d \to \infty$, from assumption A3(c).

Next we prove Step 2. Let $\mathbf{b} = (b_1, \ldots, b_L)^T$ be a vector and denote by $\Psi_{\mathbf{b}}(t) = \sum_{k=1}^{L} \psi_\ell(t) b_\ell$; furthermore let $Z_{i,\psi} = \int_0^1 Z_i(t)\Psi_{\mathbf{b}}(t)\,dt$. Then $Z_{i,\psi}$ are independent and identically distributed with mean zero and finite variance. The variance is finite because

$$E[Z_{i,\psi}^2] \;=\; \sum_{k=1}^{L}\sum_{k'=1}^{L} b_\ell b_{k'} \int_0^1\int_0^1 K^Z(t,t')\psi_\ell(t)\psi_{\ell'}(t')\,dt\,dt'.$$

Thus $|\int_0^1\int_0^1 K^Z(t,t')\psi_\ell(t)\psi_{\ell'}(t')\,dt\,dt'| \le \{\int_0^1\int_0^1 |K^Z(t,t')|^2\,dt\,dt'\}^{1/2} \times \{\int_0^1 \psi_\ell^2(t)\,dt\} < \infty$ using Hölder's inequality, the last step being a consequence of $\int_0^1\int_0^1 |K^Z(t,t')|^2\,dt\,dt' < \infty$. Note that this constraint of the covariance operator of Z_i's is met when $\|K^Z\| < \infty$, but also it can be met by assuming the less restrictive assumption $E(\|Z_i\|_{L^2}^4) < \infty$.

Thus the convergence of $n_d^{1/2}\mathbf{b}^T\widetilde{I}_d^Z$ is obtained by applying the Central Limit Theorem for independent random variables. More specifically, let $X_i = n_d M_i/m_d Z_{i,\psi}$, $R_d = \sum_{\{i:G(i)=d\}} X_i$ and note that R_d has zero-mean and variance equal to $n_d\tau_d^2$, where $\tau_d^2 = n_d \sum_{\{i:G(i)=d\}} M_i^2/m_d^2 E[Z_{i,\psi}^2]$; of course $\tau_d = O(1)$ since $Z_{i,\psi}$ has finite variance, say $\sigma_{Z,\psi}^2$, and $n_d m_d^{-2}\sum_{i:G(i)=d} M_i^2 = O(1)$. Then the Lindeberg's condition is satisfied by using the dominated convergence theorem along with Chebyshev's inequality and it follows that $n_d^{-1/2}\sum_{\{i:G(i)=d\}} X_i \to \text{Normal}(0, \sigma_{Z,\psi}^2)$. The result $n_d^{1/2}\widetilde{I}_{d,\ell}^Z \to \text{Normal}(0,\Sigma)$ follows from Cramér-Wold device, where Σ is given by expression (A.2) above.

A similar result has been discussed by Zhang and Chen (2007) in the context of groups of independent noisy curves. Their rate of convergence, of order $m^{-1/2}$, is faster than ours, of order $n_d^{-1/2}$, due to the more restrictive assumption used by Zhang and Chen (2007) - that all the curves are independent - which is not met in our setting. Lemma A.2 can be used to derive the asymptotic distribution of a linear combination $\sum_{d=1}^{D} c_d\widehat{\mu}_d^L(t)$. Such a result is particularly useful for null hypotheses testing of the type $\sum_{d=1}^{D} c_d\mu_d = 0$ which are discussed in more detail in Section 3.

Corollary A.1. *Assume conditions A1-A3 hold and consider that* $\mu_0(t) = \sum_{d=1}^{D} c_d\mu_d(t)$. *Then as* $L \to \infty$ *and provided that* $L^{-1}n_d^{1/2} = o(1)$ *we have:*

$$n^{1/2}\{\sum_{d=1}^{D} c_d\widehat{\mu}_d^L(t) - \mu_0(t)\} \to AGP(0, c^T P^{-1} c K^Z) \qquad \text{(A.5)}$$

where the convergence is in distribution as $n \to \infty$, $c^T = (c_1, \ldots, c_D)^T$, $P = diag\{p_1, \ldots, p_D\}$, *and* $p_d = \lim n_d/n$ *as* $n_d, n \to \infty$ *for each* d *with* $p_d \in (0,1)$.

A.2 Asymptotic distribution of the test

We present now the proofs on which the asymptotic distribution of the test is based.

Lemma A.4. *Assume assumptions A1, A2, and A3 and let* $\widehat{\mu}^L(t) = \sum_{\ell=1}^{L} \psi_\ell(t)\widehat{\beta}_\ell$. *Then, under the null hypothesis* H_0, *as* $L \to \infty$, $n \to \infty$, *and* $L^{-1}n_d^{1/2} = o(1)$.

$$n^{1/2}\{\widehat{\boldsymbol{\mu}}(t) - \mathbf{1}_D\widehat{\mu}^L(t)\} \to AGP[0,\ (QBQ^T K^Z)] \tag{A.6}$$

where $\widehat{\boldsymbol{\mu}}$ *stands for the vector of* $\widehat{\mu}_d^L$*'s,* $\mathbf{1}_D$ *is the* D *column vector of ones, and* Q *is* $D \times (D-1)$ *dimensional matrix with the* (d, d') *element equal to* $Q_{dd} = 1 - q_d$ *for* $d = d'$ *and* $Q_{dd'} = -q_d'$ *for* $d \neq d'$. *Also* $B = diag(p_1^{-1}, \dots, p_{D-1}^{-1}) + p_D^{-1}\mathbf{1}_{D-1}\mathbf{1}_{D-1}^T$. *Here the* p_d*'s and* q_d*'s are defined by assumption A3(a).*

Proof of Lemma A.4

Let $P_n = \text{diag}\{n_1/n, \dots, n_D/n\}$ and Q_N be the $D \times (D-1)$ dimensional matrix that is the analogue of Q with q_d replaced by m_d/m. Then the left hand side of (A.6) can he represented under the null hypothesis H_0 as $F_n\mathbf{V}_n$, where $F_n = Q_N \times [I_{D-1}| - \mathbf{1}_{D-1}] \times P_n^{-1/2}$, and $\mathbf{V}_n$ is the D dimensional vector with the d^{th} component equal to $n_d^{1/2}\{\widehat{\mu}_d^L(t) - \mu_d(t)\}$. Here $[I_{D-1}| - \mathbf{1}_{D-1}]$ denotes $(D-1) \times D$ matrix with the left $(D-1)\times(D-1)$ block matrix equal to the identity matrix I_{D-1}, and the D^{th} remaining column equal to $-\mathbf{1}_{D-1}$.

When $n_d \to \infty$ and in addition $L \to \infty$ such that $L^{-1}n_d^{1/2} = o(1)$ it follows that $\|\beta_{d,\ell}\| = o(n_d^{-1/2})$ for all $\ell \geq L+1$ and all d; thus we have that limiting distribution of $\mathbf{V}_n$ is $AGP(0, I_D K^Z)$. It follows that the limiting distribution of $n^{1/2}\{\widehat{\boldsymbol{\mu}}(t) - \mathbf{1}_D\widehat{\mu}^L(t)\}$ is $AGP(0, QBQ^T K^Z)$, since as $n_d \to \infty$ we have $F_nF_n^T \to QBQ^T$.

Proof of Theorem 3.1

Simple algebra shows that if $B = R_BR_B^T$ is the Cholesky decomposition of B, then $R_B^TQ^TPQR_B = I_{D-1} + R_B^T(q_{-D} - p_{-D})(q_{-D} - p_{-D})^TR_B$, since $B^{-1} = Q^TPQ - (q_{-D} - p_{-D})(q_{-D} - p_{-D})^T$, using Woodbury formula (Woodbury, 1950). Using Lemma A.4, and the continuity theorem one can show that when $n_d \to \infty$ and in addition $L \to \infty$ the null distribution of T_n^L is $\sum_{k=1}^{\kappa}\lambda_k\xi_k^TA\xi_k$, provided that $L^{-1}n_d^{1/2} = o(1)$.

References

Carroll, R. J., Delaigle, A. and Hall, P. (2012). Deconvolution when classifying noisy data involving transformations. *Journal of the American Statistical Association*, 106, 1166-1177.

Crainiceanu, C. M.,Staicu, A.-M., Ray, S., and Punjabi, N. M. (2012). Bootstrap-based inference on the difference in the means of two correlated functional processes. *Statist. Med.* **31**, 3223–3240.

Crainiceanu C.M., Staicu, A.M. and Di, C. (2009). Generalized multilevel functional regression. *J. Am. Statist. Assoc.*, **104**, 1550–1561.

Cuevas, A., Febrero, M. and Fraiman, R. (2004). An ANOVA test for functional data. *Computat. Statist. Data Anal.* **47**, 111–122.

Cuevas, A., Febrero, M. and Fraiman, R. (2006). On the use of the bootstrap for estimating functions with functional data. *Computat. Statist. Data Anal.* **51**, 1063–1074.

Cuevas, A. (2014).A partial overview of the theory of statistics with functional data. *J. of Statistical Planning and Inference* **147**, 1–23.

Di, C., Craininceanu, C. M., Caffo, B. S. and Punjabi, N. M. (2009). Multilevel functional principal component analysis. *Ann. Appl. Statist.* **3**, 458–488.

Donoho, D. L. and Johnstone, I. M. (1994). Ideal Spatial Adaptation by Wavelet Shrinkage. *Biometrika*, **81**, 425–55.

Efromovich, S. (1999), *Nonparametric Curve Estimation: Methods, Theory and Applications.* Springer: New York.

Fan, J. Q. and Lin, S. K. (1998). Test of significance when data are curves. *J. Am. Statist. Assoc.* **93**, 1007–1021.

Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis: theory and practice.* Springer.

Hall, P. and Van Keilegom, I. (2007). Two-sample tests in functional data analysis starting from discrete data. *Statist. Sinica* **17**, 1511–1531.

Hall, P., Horowitz, J. L. and Jing, B-Y. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika* **82**, 561–574.

Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*, New York: Springer-Verlag.

Horváth, L., Kokoszka, P., and Reeder, R. (2013). Estimation of the mean of functional time series and a two sample problem. *J. of the Royal Statist. Society: Series B*, **75**, 103–122.

Lahiri, S.N. (2003). *Resampling Methods for Dependent Data*, New-York: Springer-Verlag.

Lahiri, S. N.(1999). Theoretical comparisons of block bootstrap methods. *Ann. Statist.* **27**, 386–404.

Morris, J. S., and Carroll, R. J. (2006). Wavelet-based functional mixed models. Journal of the Royal Statistical Society: Series B (Statistical Methodology), **68**, 179–199.

Patton, A., Politis, D. N., and White, H. (2009). 'CORRECTION TO "Automatic Block-Length Selection for the Dependent Bootstrap" by D. Politis and H. White', *Econometric Reviews* **28**, 372–375.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. New York: Springer-Verlag.

Shen, Q. and Faraway, J. (2004). An F test for linear models with functional responses. *Statist. Sinica* **14**, 1239–1257.

Staicu, A.M., Crainiceanu, C. M., and Carroll R. J. (2010). Fast Methods for Spatially Correlated Multilevel Functional Data. *Biostatistics*, **11**, 177–194.

Staicu, A.M., Li, Y., Crainiceanu C.M. and Ruppert, D. (2013) Likelihood ratio tests for dependent data with applications to longitudinal and functional data analysis. *Scandinavian Journal of Statistics*, to appear

Stein, M., L. (1999). *Interpolation of Spatial Data. Some Theory for Kriging*. Series: Springer Series in Statistics.

Warren, R. E., Vanderbeek, R. G., Ben-David, A., and Ahl, J. L. (2008). Simultaneous estimation of aerosol cloud concentration and spectral backscatter from multiple-wavelength lidar data. *Applied Optics*, 47, 4309-4320.

Warren, R. E., Vanderbeek, R. G., and Ahl, J. L. (2009). Detection and classification of atmospheric aerosols using multi-wavelength LWIR lidar. *Proceedings of SPIE*, 7304, 73040E.

Woodbury, M. A. (1950). *Inverting modified matrices.* Memorandum Rept. 42, Statistical Research Group, Princeton University, Princeton, NJ.

Xun, X., Cao, J., Mallick, B. K., Maity, A. and Carroll, R. J. (2013). Parameter estimation of partial differential equation models. *Journal of the American Statistical Association*, 108, 1009-1020.

Yao, F., Müller, H.-G. and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* **100**, 577–590.

Zhang, C., Peng, H. and Zhang, J.-T. (2010). Two samples tests for functional data. *Comm. Statist. - Theory and Methods* **39**, 559–578.

Zhang, J.-T. and Chen, J. (2007). Statistical inferences for functional data. *The Ann. of Statist.* **35**, 1052–1079.

Ana-Maria Staicu
Department of Statistics, North Carolina State University,
E-mail: `staicu@stat.ncsu.edu`

Soumen N. Lahiri
Department of Statistics, North Carolina State University,
E-mail: `slahiri@ncsu.edu`

Raymond J. Carroll
Department of Statistics, Texas A&M University,
E-mail: `carroll@stat.tamu.edu`