

Consistency of the mean and the principal components of spatially distributed functional data

SIEGFRIED HÖRMANN¹ and PIOTR KOKOSZKA²

¹*Département de Mathématique, Université Libre de Bruxelles, CP 210, Bd du Triomphe, B-1050 Brussels - Belgium. E-mail: shormann@ulb.ac.be*

²*Department of Statistics, Colorado State University, Fort Collins, CO 80523-1877, USA. E-mail: Piotr.Kokoszka@colostate.edu*

This paper develops a framework for the estimation of the functional mean and the functional principal components when the functions form a random field. More specifically, the data we study consist of curves $X(\mathbf{s}_k; t)$, $t \in [0, T]$, observed at spatial points $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_N$. We establish conditions for the sample average (in space) of the $X(\mathbf{s}_k)$ to be a consistent estimator of the population mean function, and for the usual empirical covariance operator to be a consistent estimator of the population covariance operator. These conditions involve an interplay of the assumptions on an appropriately defined dependence between the functions $X(\mathbf{s}_k)$ and the assumptions on the spatial distribution of the points \mathbf{s}_k . The rates of convergence may be the same as for iid functional samples, but generally depend on the strength of dependence and appropriately quantified distances between the points \mathbf{s}_k . We also formulate conditions for the lack of consistency.

Keywords: consistency, estimation, functional data, functional principal components, spatial dependence, spatial sampling design.

1. Introduction

This paper develops aspects of theory for functional data observed at spatial locations. The data consist of curves $X(\mathbf{s}_k) = \{X(\mathbf{s}_k; t), t \in [0, T]\}$, observed at spatial points $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_N$. Such data structures are quite common, but often the spatial dependence and the spatial distribution of the points \mathbf{s}_k are not taken into account. A well-known example is the Canadian temperature and precipitation data used in Ramsay and Silverman (2005). The annual curves are available at 35 locations, some of which are quite close, and so the curves have very similar characteristics, others are very remote with notably different curves.

There has not been much research on fundamental properties of spatially distributed functional data. Delicado *et al.* (2010) review recent contributions to the methodology for spatially distributed functional data. For geostatistical functional data, several exploratory approaches to kriging have been proposed. Typically fixed basis expansions

are used, see Yamanishi and Tanaka (2003) and Bel *et al.* (2011). A general theoretical framework has to address several problems. The first issue is the dimensionality of the index space. While in time series analysis, the process is indexed by an equispaced scalar parameter, we need here a d -dimensional index space. For model building this makes a big difference since the dynamics and dependence of the process have to be described in all directions, and the typical recurrence equations used in time series cannot be employed. The model building is further complicated by the fact that the index space is often continuous (geostatistical data). Rather than defining a random field $\{\xi(\mathbf{s}); \mathbf{s} \in \mathbb{R}^d\}$ via a specific model equations, dependence conditions are imposed, in terms of the decay of the covariances or using mixing conditions. Another feature peculiar to random field theory is the design of the sampling points; the distances between them play a fundamental role. Different asymptotics hold in the presence of clusters and for sparsely distributed points. At least three types of point distributions have been considered (Cressie (1993)): When the region R_N where the points $\{\mathbf{s}_{i,N}; 1 \leq i \leq N\}$ are sampled remains bounded, then we are in the so-called *infill domain sampling* case. Classical asymptotic results, like the law of large numbers or the central limit theorem will usually fail, see Lahiri (1996). The other extreme situation is described by the *increasing domain sampling*. Here a minimum separation between the sampling points $\{\mathbf{s}_{i,N}\} \in R_N$ for all i and N is required. This is of course only possible if $\text{diam}(R_N) \rightarrow \infty$. We shall also explore the *nearly infill* situation studied by Lahiri (2003) and Park *et al.* (2009). In this case the domain of the sampling region becomes unbounded ($\text{diam}(R_N) \rightarrow \infty$), but at the same time the number of sites in any given subregion tends to infinity, i.e. the points become more dense. These issues are also studied by Zhang (2004), Loh (2005), Lahiri and Zhu (2006), Du *et al.* (2009). We formalize these concepts in Section 2. Finally, the interplay of the geostatistical spatial structure and the functional temporal structure must be cast into a workable framework.

The paper is organized as follows. Section 2 introduces the statistical setting. It also compares our conditions to those typically assumed for scalar spatial processes. In Sections 3 and 4 we establish consistency results, respectively, for the functional mean and the covariance operator. Section 5 explains, by means of general theorems and examples, when the sample principal components are not consistent. The proofs of the main results are collected in Section 6.

To make this presentation more streamlined, we have outsourced some proofs, further details and several examples. They are available as supplemental material: Hörmann and Kokoszka (2011).

2. Preliminaries and assumptions

We assume $\{X(\mathbf{s}), \mathbf{s} \in \mathbb{R}^d\}$ is a random field taking values in $L^2 = L^2([0, 1])$, i.e. each $X(\mathbf{s})$ is a square integrable function defined on $[0, 1]$. The value of this function at $t \in [0, 1]$ is denoted by $X(\mathbf{s}; t)$. With the usual inner product in L^2 , the norm of $X(\mathbf{s})$ is

$$\|X(\mathbf{s})\| = \left\{ \int X^2(\mathbf{s}; t) dt \right\}^{1/2}.$$

The mean function $\mu(\mathbf{s}) = \{EX(\mathbf{s}; t), t \in [0, 1]\}$ and the covariance operator is then defined for $x \in L^2$ by

$$C_{\mathbf{s}, \mathbf{s}}(x) = E[\langle X(\mathbf{s}) - \mu(\mathbf{s}), x \rangle (X(\mathbf{s}) - \mu(\mathbf{s}))].$$

More generally we define the cross-covariance operators

$$C_{\mathbf{s}_1, \mathbf{s}_2}(x) = E[\langle X(\mathbf{s}_1) - \mu(\mathbf{s}_1), x \rangle (X(\mathbf{s}_2) - \mu(\mathbf{s}_2))].$$

For the existence of $C_{\mathbf{s}_1, \mathbf{s}_2}$ a minimal assumption is that the variables have finite second moments in the sense that

$$E\|X(\mathbf{s})\|^2 < \infty, \quad \forall \mathbf{s}. \quad (1)$$

To think of our observations $X(\mathbf{s})$ as curves in L^2 is convenient and motivated this work, but our propositions in Section 3, 4 and 5 only require the general assumption that $\{X(\mathbf{s}), \mathbf{s} \in \mathbb{R}^d\}$ is a field taking values in some separable Hilbert space. So particularly our results hold true when L^2 is replaced by \mathbb{R}^p . To the best of our knowledge, our results are new as well in the vector case.

Our goal is to estimate the unknown mean curves and the principal components (FPC's). FPC's are intimately connected with covariance operators, as we will describe in some detail in the next section, and likewise estimation of FPC's is based on estimation of covariance operators. Thereby we are estimating *across space* and *not across time*. A minimal requirement for this to make sense is then that all locations share a common mean curve and that the covariance operator is the same for all locations, respectively:

$$\mu(\mathbf{s}) = \mu \quad \text{and} \quad C_{\mathbf{s}, \mathbf{s}} = C. \quad (2)$$

Although (2) is apparently necessary and would also be quite natural if we were in time series context, it may be not realistic in some spatial data situations. Let us briefly sketch how our methods can still be useful by employing a spatio-temporal framework. In this case we suppose to have for each location \mathbf{s} a functional time series $\{X_i(\mathbf{s}), i = 1, 2, \dots\}$. To avoid confusion between the time parameter i ($i = 1, 2, 3, \dots$) and the ‘‘intraday time’’ parameter t ($t \in [0, 1]$), we will employ for this paragraph the notation $X_i(\mathbf{s}; t)$. We then assume the following model

$$X_i(\mathbf{s}; t) = \alpha(\mathbf{s}; t) + \mu_i(t) + Y_i(\mathbf{s}; t), \quad i = 1, 2, 3, \dots,$$

where $\{Y_i(\mathbf{s}; t), t \in [0, 1]\}$ are curves at spatial locations which satisfy (2) with $\mu \equiv 0$ and where for each fixed \mathbf{s} the functional time series $\{X_i(\mathbf{s}; \cdot), i = 1, 2, 3, \dots\}$ is stationary and weakly dependent (e.g. as assumed in Hörmann and Kokoszka (2010)). The random curves $\{\mu_i(t), t \in [0, 1]\}$ are zero mean curves and form a certain ‘‘basis level’’ shared by all curves $\{X_i(\mathbf{s}; t), t \in [0, 1]\}$ across space at time i . In our setup the problem of interest is to determine for a given day i the curve $\{\mu_i(t), t \in [0, 1]\}$. We might think of daily temperature curves measured across some region. The curve $\{\mu_i(t), t \in [0, 1]\}$ amounts to a general deviation from normal on day i which persists over the whole region (e.g. on a hot day, it is not unlikely that all stations will show curves above average). Due to

geographical differences the mean curves $\alpha(\mathbf{s}; t)$ might be different at different locations, but they can be estimated individually by $\hat{\alpha}(\mathbf{s}; t) = \frac{1}{M} \sum_{i=1}^M X_i(\mathbf{s}; t)$, when we have a sample of M days. So we can assume that we can detrended our data and then work with new observations $X_i'(\mathbf{s}; t) = \mu_i(t) + Y_i(\mathbf{s}; t)$, to which our theory applies, when now i is fixed.

Besides (2), we don't impose any further stationarity assumption on the field $\{X(\mathbf{s})\}$.

2.1. Functional principal components

Functional principal components play a fundamental role in functional data analysis, much greater than the usual multivariate principal components. This is mostly due to the fact that the Karhunen-Loève expansion allows to represent functional data in a concise way. This property has been extensively used and studied in various settings. To name only a few illustrative references, we cite Hall and Hosseini-Nasab (2006), Reiss and Ogden (2007), Gabrys and Kokoszka (2007), Benko *et al.* (2009), Paul and Peng (2009), Jiang and Wang (2010) and Gabrys *et al.* (2010). Depending on the structure of the data, theoretical analyses emphasize various aspects of the estimation process, with smoothing in iid samples having being particularly carefully studied. This paper focuses on the spatial dependence and distribution of the curves, which has received no attention so far.

Suppose X_1, X_2, \dots, X_N are mean zero identically distributed elements of L^2 such that $E\|X\|^4 < \infty$. The eigenfunctions of the covariance operator of X_1 are the functional principal components, denoted v_k . Up to a sign, they are estimated by the empirical FPC's (EFPC's), denoted \hat{v}_k and defined as the eigenfunctions of the empirical covariance operator

$$\hat{C}_N(x) = \frac{1}{N} \sum_{n=1}^N \langle X_n, x \rangle X_n, \quad x \in L^2.$$

The distance between v_k and \hat{v}_k is determined by the distance between C and \hat{C}_N . This follows from Lemma 1. Due to its importance in the exposition that follows we provide a precise statement. An analog of Lemma 1 has often been used for iid functions under more restrictive assumptions. To state Lemma 1, consider two compact operators C and K with singular value decompositions

$$C(x) = \sum_{j=1}^{\infty} \lambda_j \langle x, v_j \rangle f_j, \quad K(x) = \sum_{j=1}^{\infty} \gamma_j \langle x, u_j \rangle g_j. \quad (3)$$

Recall that a linear operator K in a separable Hilbert space H is said to be Hilbert-Schmidt, if for some orthonormal basis $\{e_i\}$ of H

$$\|K\|_S^2 := \sum_{i \geq 1} \|K(e_i)\|_H^2 < \infty.$$

Then $\|\cdot\|_{\mathcal{S}}$ defines a norm on the space of all operators satisfying this condition. The norm is independent of the choice of the basis. This space is again a Hilbert space with the inner product

$$\langle K_1, K_2 \rangle_{\mathcal{S}} = \sum_{i \geq 1} \langle K_1(e_i), K_2(e_i) \rangle.$$

Set

$$v'_j = \hat{c}_j v_j, \quad \hat{c}_j = \text{sign}(\langle u_j, v_j \rangle).$$

Lemma 1. *Suppose $C, K \in \mathcal{L}$ are two compact operators with singular value decompositions (3). If C is symmetric, $f_j = v_j$ in (3), and its eigenvalues satisfy*

$$\lambda_1 > \lambda_2 > \dots > \lambda_d > \lambda_{d+1}, \quad (4)$$

then

$$\|u_j - v'_j\| \leq \frac{2\sqrt{2}}{\alpha_j} \|K - C\|_{\mathcal{L}}, \quad 1 \leq j \leq p,$$

where $\alpha_1 = \lambda_1 - \lambda_2$ and $\alpha_j = \min(\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1})$, $2 \leq j \leq p$.

Lemma 1 can be proven using Corollary 1.6 on p. 99 of Gohberg *et al.* (1990) and following the lines of the proof of Lemma 4.3 of Bosq (2000).

If the functional observations $X_k, k = 1, 2, \dots$, are independent, then

$$\limsup_{N \rightarrow \infty} NE \|\widehat{C}_N - C\|_{\mathcal{S}}^2 < \infty. \quad (5)$$

Consequently, for such functional observations, under (4),

$$\max_{1 \leq k \leq d} E \|\hat{c}_k \hat{v}_k - v_k\|^2 = O(N^{-1}).$$

Hörmann and Kokoszka (2010) showed that (5) continues to hold for weakly dependent time series, in particular for m -dependent X_k . It is not difficult to show that if spatially distributed functions are such that $X(\mathbf{s})$ is independent of $X(\mathbf{s}')$ if the distance between \mathbf{s} and \mathbf{s}' is greater than m , then (5) need not hold. It is even possible that EFPC's \hat{v}_k do not converge at all. See Section 5.

2.2. Dependence assumptions

To develop an estimation framework, we impose conditions on the decay of the cross-covariances $E[\langle X(\mathbf{s}_1) - \mu, X(\mathbf{s}_2) - \mu \rangle]$, as the distance between \mathbf{s}_1 and \mathbf{s}_2 increases. We shall use the distance function defined by the Euclidian norm in \mathbb{R}^d , denoted $\|\mathbf{s}_1 - \mathbf{s}_2\|_2$, but other distance functions can be used as well.

Assumption 1. *The spatial process $\{X(\mathbf{s}), \mathbf{s} \in \mathbb{R}^d\}$ satisfies (1) and (2). In addition,*

$$|E\langle X(\mathbf{s}_1) - \mu, X(\mathbf{s}_2) - \mu \rangle| \leq h(\|\mathbf{s}_1 - \mathbf{s}_2\|_2), \quad (6)$$

where $h : [0, \infty) \rightarrow [0, \infty)$ with $h(x) \searrow 0$, as $x \rightarrow \infty$.

If $\{e_j\}$ is an orthonormal basis (ONB) of L^2 , then it can be easily seen that (6) is equivalent to

$$\left| \sum_{j \geq 1} \langle C_{\mathbf{s}_1, \mathbf{s}_2}(e_j), e_j \rangle \right| \leq h(\|\mathbf{s}_1 - \mathbf{s}_2\|_2). \quad (7)$$

For any such ONB a Fourier expansion of $X(\mathbf{s})$ yields

$$X(\mathbf{s}; t) = \mu + \sum_{j=1}^{\infty} \xi_j(\mathbf{s}) e_j(t), \quad \mathbf{s} \in \mathbb{R}^d, t \in [0, 1], \quad (8)$$

where $\xi_j(\mathbf{s}) = \langle X(\mathbf{s}) - \mu, e_j \rangle$. Using the relation

$$\langle C_{\mathbf{s}_1, \mathbf{s}_2}(e_j), e_j \rangle = E[\xi_j(\mathbf{s}_1)\xi_j(\mathbf{s}_2)],$$

the more specific assumption

$$|E[\xi_j(\mathbf{s}_1)\xi_j(\mathbf{s}_2)]| \leq \phi_j(\|\mathbf{s}_1 - \mathbf{s}_2\|_2), \quad (9)$$

on the scalar fields, gives (6), if

$$\sum_{j \geq 1} \phi_j(\|\mathbf{s}_1 - \mathbf{s}_2\|_2) \leq h(\|\mathbf{s}_1 - \mathbf{s}_2\|_2). \quad (10)$$

Examples 1 and 2 consider typical spatial covariance functions, and show when condition (10) holds with a function h as in Assumption 1.

EXAMPLE 1. Suppose that the fields $\{\xi_j(\mathbf{s}), \mathbf{s} \in \mathbb{R}^d\}$, $j \geq 1$, are zero mean, strictly stationary and α -mixing. That is

$$\sup_{(A, B) \in \sigma(\xi_j(\mathbf{s})) \times \sigma(\xi_j(\mathbf{s} + \mathbf{h}))} |P(A)P(B) - P(A \cap B)| \leq \alpha_j(\mathbf{h}),$$

with $\alpha_j(\mathbf{h}) \rightarrow 0$ if $\|\mathbf{h}\|_2 \rightarrow \infty$. Let $\alpha'_j(h) = \sup\{\alpha_j(\mathbf{h}) : \|\mathbf{h}\|_2 = h\}$. Then $\alpha_j^*(h) = \sup\{\alpha'_j(x) : x \geq h\} \searrow 0$ as $h \rightarrow \infty$. Using stationarity and the main result in Rio (1993) it follows that

$$\begin{aligned} |E[\xi_j(\mathbf{s}_1)\xi_j(\mathbf{s}_2)]| &= |E[\xi_j(\mathbf{0})\xi_j(\mathbf{s}_2 - \mathbf{s}_1)]| \\ &\leq 2 \int_0^{2\alpha_j(\|\mathbf{s}_2 - \mathbf{s}_1\|_2)} Q_j^2(u) du \\ &\leq 2 \int_0^{2\alpha_j^*(\|\mathbf{s}_2 - \mathbf{s}_1\|_2)} Q_j^2(u) du \\ &=: \phi_j(\|\mathbf{s}_2 - \mathbf{s}_1\|_2), \end{aligned}$$

where $Q_j(u) = \inf\{t : P(|\xi_j(\mathbf{0})| > t) \leq u\}$ is the quantile function of $|\xi_j(\mathbf{0})|$. Note that $\alpha_h(\mathbf{h}) \leq 1/4$ for any \mathbf{h} , and thus $\phi_j(x) \leq 2 \int_0^1 Q_j^2(u) du = 2E[\xi_j^2(\mathbf{0})]$. If $\sum_{j \geq 1} E\xi_j^2(\mathbf{0}) <$

∞ , then (6) holds with $h(x) = \sum_{j \geq 1} \phi_j(x)$. (Note that $|h(x)| \searrow 0$ follows from $\alpha_j^*(x) \searrow 0$ and the monotone convergence theorem.)

We note that α -mixing is one of the classical assumptions in random field literature to establish limit theorems. It is in fact a much stronger assumption than ours and it is suitable if one needs more delicate results, like a central limit theorem (see e.g. Bolthausen (1982)) or uniform laws of large numbers, see Jenish and Prucha (2009). Besides the restriction to scalar observations, many papers restrict to the so-called “purely increasing domain sampling”, an assumption that we are going to further relax in the following.

EXAMPLE 2. Suppose (9) holds, and set $h(x) = \sum_{j \geq 1} \phi_j(x)$. If each ϕ_j is a powered exponential covariance function defined by

$$\phi_j(x) = \sigma_j^2 \exp \left\{ - \left(\frac{x}{\rho_j} \right)^p \right\}.$$

Then h satisfies the conditions of Assumption 1 if

$$\sum_{j \geq 1} \sigma_j^2 < \infty \quad \text{and} \quad \sup_{j \geq 1} \rho_j < \infty. \quad (11)$$

Condition (11) is also sufficient if all ϕ_j are in the Matérn class, see Stein (1999), with the same ν , i.e.

$$\phi_j(x) = \sigma_j^2 x^\nu K_\nu(x/\rho_j),$$

because the modified Bessel function K_ν decays monotonically and approximately exponentially fast; numerical calculations show that $K_\nu(s)$ practically vanishes if $s > \nu$. Condition (11) is clearly sufficient for spherical ϕ_j defined (for $d = 3$) by

$$\phi_j(x) = \begin{cases} \sigma_j^2 \left(1 - \frac{3x}{2\rho_j} + \frac{x^3}{2\rho_j^3} \right), & x \leq \rho_j \\ 0, & x > \rho_j \end{cases}$$

because ϕ_j is decreasing on $[0, \rho_j]$.

Assumption 1 is appropriate when studying estimation of the mean function. For the estimation of the covariance operator, we need to impose a different assumption. Recall that if z and y are elements in some Hilbert space H with norm $\|\cdot\|_H$, the operator $z \otimes y$, is defined by $z \otimes y(x) = \langle z, x \rangle y$. In the following assumption, we suppose that the mean of the functional field is zero. This is justified by notational convenience and because we deal with the consistent estimation of the mean function separately.

Assumption 2. *The spatial process $\{X(\mathbf{s}), \mathbf{s} \in \mathbb{R}^d\}$ satisfies (2) with $\mu \equiv 0$ and has 4 moments, i.e. $E\langle X(\mathbf{s}), x \rangle = 0, \forall x \in L^2$, and $E\|X(\mathbf{s})\|^4 < \infty$. In addition,*

$$|E\langle X(\mathbf{s}_1) \otimes X(\mathbf{s}_1) - C, X(\mathbf{s}_2) \otimes X(\mathbf{s}_2) - C \rangle_S| \leq H(\|\mathbf{s}_1 - \mathbf{s}_2\|_2), \quad (12)$$

where $H : [0, \infty) \rightarrow [0, \infty)$ with $H(x) \searrow 0$, as $x \rightarrow \infty$.

Assumption 2 cannot be verified using only conditions on the covariances of the scalar fields ξ_j in (8) because these covariances do not specify the 4th order structure of the model. This can be done if the random field is Gaussian or if additional structure is imposed. We discuss this Assumption 2 in more detail in supplemental material (Hörmann and Kokoszka, 2011, Section S.2).

2.3. Spatial distribution of the sampling points

As already noted, for spatial processes assumptions on the distribution of the sampling points are as important as those on the covariance structure. To formalize the different sampling schemes introduced in Section 1, we propose the following measure of “minimal dispersion” of some point cloud \mathfrak{S} :

$$I_\rho(\mathbf{s}, \mathfrak{S}) = |\{\mathbf{y} \in \mathfrak{S} : \|\mathbf{s} - \mathbf{y}\|_2 \leq \rho\}|/|\mathfrak{S}| \quad \text{and} \quad I_\rho(\mathfrak{S}) = \sup \{I_\rho(\mathbf{s}, \mathfrak{S}), \mathbf{s} \in \mathfrak{S}\},$$

where $|\mathfrak{S}|$ denotes the number of elements of \mathfrak{S} . The quantity $I_\rho(\mathfrak{S})$ is the maximal fraction of \mathfrak{S} -points in a ball of radius ρ centered at an element of \mathfrak{S} . Notice that $1/|\mathfrak{S}| \leq I_\rho(\mathfrak{S}) \leq 1$. We call $\rho \mapsto I_\rho(\mathfrak{S})$ the *intensity function* of \mathfrak{S} .

Definition 1. For a sampling scheme $\mathfrak{S}_N = \{\mathbf{s}_{i,N}; 1 \leq i \leq S_N\}$, $S_N \rightarrow \infty$, we consider the following conditions:

- (i) there is a $\rho > 0$ such that $\limsup_{N \rightarrow \infty} I_\rho(\mathfrak{S}_N) > 0$;
- (ii) for some sequence $\rho_N \rightarrow \infty$ we have $I_{\rho_N}(\mathfrak{S}_N) \rightarrow 0$;
- (iii) for any fixed $\rho > 0$ we have $S_N I_\rho(\mathfrak{S}_N) \rightarrow \infty$.

We call a deterministic sampling scheme $\mathfrak{S}_N = \{\mathbf{s}_{i,N}; 1 \leq i \leq S_N\}$

Type A if (i) holds;

Type B if (ii) and (iii) hold;

Type C if (ii) holds, but there is a $\rho > 0$ such that $\limsup_{N \rightarrow \infty} S_N I_\rho(\mathfrak{S}_N) < \infty$.

If the sampling scheme is stochastic we call it Type A, B or C if relations (i), (ii) and (iii) hold with $I_\rho(\mathfrak{S}_N)$ replaced by $E I_\rho(\mathfrak{S}_N)$.

Type A sampling is related to purely infill domain sampling which corresponds to $I_\rho(\mathfrak{S}_N) = 1$ for all $N \geq 1$, provided ρ is large enough. However, in contrast to the purely infill domain sampling, it still allows for a non-degenerate asymptotic theory for sparse enough subsamples (in the sense of Type B or C).

A brief reflection shows that assumptions (i) and (ii) are mutually exclusive. Combining (ii) and (iii) implies that the points intensify (at least at certain spots) excluding the purely increasing domain sampling. Hence the Type B sampling corresponds to the nearly infill domain sampling. If only (ii) holds, but (iii) does not (Type C sampling) then the sampling scheme corresponds to purely increasing domain sampling.

Our conditions are more general than those proposed so far. We treat below two special cases which are closely related to those considered by Lahiri (2003). The points

are assumed to be on a grid of an increasing size, or to have a density. We show how our more general assumptions look in these special cases, and provide additional intuition behind the sampling designs formulated in Definition 1. They also set a framework for some results of Sections 3 and 4.

2.4. Non-random regular design

Let $\mathcal{Z}(\boldsymbol{\delta})$ be a lattice in \mathbb{R}^d with increments δ_i in the i -th direction. Let $\delta_0 = \min\{\delta_1, \dots, \delta_d\}$, $\Delta^d = \prod_{i=1}^d \delta_i$ and let $R_N = \alpha_N R_0$, where R_0 is some bounded Riemann measurable Borel-set in \mathbb{R}^d containing the origin. A set is Riemann measurable if its indicator function is Riemann integrable. This condition excludes highly irregular sets R_0 . The scaling parameters $\alpha_N > 0$ are assumed to be non-decreasing and will be specified below in Lemma 2. We assume without loss of generality that $\text{Vol}(R_0) = 1$, hence $\text{Vol}(R_N) = \alpha_N^d$. Typical examples are $R_0 = \{x \in \mathbb{R}^d : \|x\| \leq z_{1,d}\}$, with $z_{1,d}$ equal to the radius of the d -dimensional sphere with volume 1, or $R_0 = [-1/2, 1/2]^d$. The sampling points \mathfrak{S}_N are defined as $\{\mathbf{s}_{k,N}, 1 \leq k \leq S_N\} = \mathcal{Z}(\eta_N \boldsymbol{\delta}) \cap R_N$, where η_N is chosen such that the sample size $S_N \sim N$. We remark that we only introduce S_N as it is generally not possible by the just described construction to define η_N such, that we would get exactly N points on the grid to fall in R_N . It is intuitively clear that $\text{Vol}(R_N) \approx \eta_N^d \Delta^d S_N$, suggesting

$$\eta_N = \frac{\alpha_N}{\Delta N^{1/d}}. \quad (13)$$

A formal proof that η_N in (13) assures $S_N \sim N$ is given in Section S.3 of the supplemental material.

The following Lemma, whose proof is also given in the supplemental material relates the non-random regular design to Definition 1. We write $a_N \gg b_N$ if $\limsup b_N/a_N < \infty$.

Lemma 2. *In the above described design the following pairs of statements are equivalent:*

- (i) α_N remains bounded \Leftrightarrow Type A sampling;
- (ii) $\alpha_N \rightarrow \infty$ and $\alpha_N = o(N^{1/d}) \Leftrightarrow$ Type B sampling;
- (iii) $\alpha_N \gg N^{1/d} \Leftrightarrow$ Type C sampling.

2.5. Randomized design

Let $\{\mathbf{s}_k, 1 \leq k \leq N\}$ be iid random vectors with a density $f(\mathbf{s})$ which has support on a Borel set $R_0 \subset \mathbb{R}^d$ containing the origin and satisfying $\text{Vol}(R_0) = 1$. Again we assume Riemann measurability for R_0 to exclude highly irregular sets. For the sake of simplicity we shall assume that on R_0 the density is bounded away from zero, so that we have $0 < f_L \leq \inf_{x \in R_0} f(x)$. The point set $\{\mathbf{s}_{k,N}, 1 \leq k \leq N\}$ is defined by $\mathbf{s}_{k,N} = \alpha_N \mathbf{s}_k$ for $k = 1, \dots, N$. For fixed N , this is equivalent to: $\{\mathbf{s}_{k,N}, 1 \leq k \leq N\}$ is an iid sequence on $R_N = \alpha_N R_0$ with density $\alpha_N^{-d} f(\alpha_N^{-1} \mathbf{s})$.

We cannot expect to obtain a full analogue of Lemma 2 in the randomized setup. For Type C sampling, the problem is much more delicate, and a closer study shows that it is related to the oscillation behavior of multivariate empirical processes. While Stute (1984) gives almost sure upper bounds, we would need here sharp results on the moments of the modulus of continuity of multivariate empirical process. Such results exist, see Einmahl and Ruymgaart (1987), but are connected to technical assumptions on the bandwidth for the modulus (here determined by α_N) which are not satisfied in our setup. Hence a detailed treatment would go beyond the scope of this paper. We thus state here the following lemma whose proof is given in the supplemental material.

Lemma 3. *In the above described sampling scheme the following statements hold:*

- (i) α_N remains bounded \Rightarrow Type A sampling;
- (ii) $\alpha_N \rightarrow \infty$ and $\alpha_N = o(N^{1/d}) \Rightarrow$ Type B sampling;

3. Consistency of the sample mean function

We start with a general setup, and show that the rates can be improved in special cases. The proofs of the main results, Propositions 1, 2, 3, are collected in Section 6.

For independent or weakly dependent functional observations X_k ,

$$E \left\| \frac{1}{N} \sum_{k=1}^N X_k - \mu \right\|^2 = O(N^{-1}). \quad (14)$$

Proposition 1 shows that for general functional spatial processes, the rate of consistency may be much slower than $O(N^{-1})$; it is the maximum of $h(\rho_N)$ and $I_{\rho_N}(\mathfrak{S}_N)$ with ρ_N from (ii) of Definition 1. Intuitively, the sample mean is consistent if there is a sequence of increasing balls which contain a fraction of points which tends to zero, and the decay of the correlations compensates for the increasing radius of these balls.

Proposition 1. *Let Assumption 1 hold, and assume that \mathfrak{S}_N defines a non-random design of Type A, B or C. Then for any $\rho_N > 0$,*

$$E \left\| \frac{1}{N} \sum_{k=1}^N X(\mathbf{s}_{k,N}) - \mu \right\|^2 \leq h(\rho_N) + h(0)I_{\rho_N}(\mathfrak{S}_N). \quad (15)$$

Hence, under the Type B or Type C non-random sampling, with ρ_N as in (ii) of Definition 1, the sample mean is consistent.

EXAMPLE 3. Assume that N points $\{\mathbf{s}_{k,N}, 1 \leq k \leq N\}$ are on a regular grid in $\alpha_N[-1/2, 1/2]^d$. Then, as we have seen in Section 2.4, $I_\rho(\mathfrak{S}_N)$ is proportional to $(\rho/\alpha_N)^d$.

For example, if $h(x) = 1/(1+x)^2$, then choosing $\rho_N = \alpha_N^{d/(d+2)}$ we obtain that

$$h(\rho_N) + h(0)I_{\rho_N}(\mathfrak{S}_N) \ll \alpha_N^{-2d/(d+2)} \vee N^{-1}.$$

(Recall that $I_{\rho_N}(\mathfrak{S}_N) \geq N^{-1}$.)

In Hörmann and Kokoszka (2011), Section S.4, we show that bound (15) is optimal, in the sense that under the assumptions of Proposition 1 it is always possible to find an example where the rate in (15) is precise and cannot be improved. This of course doesn't mean that the obtained rate is uniformly optimal. If we impose a more regular sampling design, we can get better rates.

Proposition 2. *Assume the sampling design of Section 2.4. Let Assumption 1 hold with h such that $x^{d-1}h(x)$ is monotone on $[b, \infty)$, $b > 0$. Then under Type B sampling*

$$\begin{aligned} E \left\| \frac{1}{S_N} \sum_{k=1}^{S_N} X(\mathbf{s}_{k,N}) - \mu \right\|^2 \\ \leq \frac{1}{\alpha_N^d} \left\{ d(3\Delta)^d \int_0^{K\alpha_N} x^{d-1}h(x)dx + o(1) \sup_{x \in [0, K\alpha_N]} x^{d-1}h(x) \right\}, \end{aligned} \quad (16)$$

for some large enough constant K which is independent of N . Under Type C sampling $1/\alpha_N^d$ in (16) is replaced by $O(N^{-1})$.

The technical assumptions on h pose no practical problem, they are satisfied for all important examples, see Example 2. A common situation is that $x^{d-1}h(x)$ is increasing on $[0, b]$ and decreasing thereafter. We recall again that $S_N \sim N$ and that $S_N = N$ usually cannot be achieved due to the construction of the sampling design.

Our next example shows that under nearly infill domain sampling the rate of consistency may be much slower than for the iid case, if the size of the domain does not increase fast enough.

EXAMPLE 4. Suppose the functional spatial process has representation (8), and (9) holds with the covariance functions ϕ_j as in Example 2 (powered exponential, Matérn or spherical). Define $h(x) = \sum_{j \geq 1} \phi_j(x)$, and assume that condition (11) holds. Assumption 1 is then satisfied and

$$\int_0^\infty x^{d-1}h(x)dx < \infty \quad \text{and} \quad \sup_{x \in \mathbb{R}} x^{d-1}h(x) < \infty. \quad (17)$$

Therefore, for the sampling design of Section 2.4,

$$E \left\| \frac{1}{S_N} \sum_{k=1}^{S_N} X(\mathbf{s}_{k,N}) - \mu \right\|^2 = \begin{cases} O(\alpha_N^{-d} \vee N^{-1}), & \text{under Type B sampling} \\ O(N^{-1}), & \text{under Type C sampling} \end{cases} \quad (18)$$

The next example shows that formula (18) is far from universal, and that the rate of consistency may be even slower if the covariances decay slower than exponential.

EXAMPLE 5. Consider the general setting of Example 4, but assume that each covariance function ϕ_j has the quadratic rational form

$$\phi_j(x) = \sigma_j^2 \left\{ 1 + \left(\frac{x}{\rho_j} \right)^2 \right\}^{-1}.$$

Condition (11) implies that $h(x) = \sum_{j \geq 1} \phi_j(x)$ satisfies Assumption 1, but now $h(x) \sim x^{-2}$, as $x \rightarrow \infty$. Because of this rate, condition (17) holds only for $d = 1$ (and so for this dimension (18) also holds). If $d \geq 2$, (17) fails, and to find the rate of the consistency, we must use (16) directly. We focus only on Type B sampling, and assume implicitly that the rate is slower than N^{-1} . We assume (11) throughout this example.

If $d = 2$,

$$\begin{aligned} \int_0^{K\alpha_N} x^{d-1} h(x) dx &= \sum_j \sigma_j^2 \int_0^{K\alpha_N} x \left\{ 1 + \left(\frac{x}{\rho_j} \right)^2 \right\}^{-1} dx \\ &= \sum_j \sigma_j^2 \rho_j^2 O \left(\int_1^{K\alpha_N} x^{-1} dx \right) = O(\ln \alpha_N) \end{aligned}$$

and similarly $\sup_{x \in [0, K\alpha_N]} x^{d-1} h(x) = O(1)$.

If $d \geq 3$, the leading term is

$$\int_0^{K\alpha_N} x^{d-1} h(x) dx = O(\alpha_N^{d-3}).$$

We summarize these calculations as

$$E \left\| \frac{1}{S_N} \sum_{k=1}^{S_N} X(\mathbf{s}_{k,N}) - \mu \right\|^2 = \begin{cases} O(\alpha_N^{-1}), & \text{if } d = 1 \\ O(\alpha_N^{-2} \ln(\alpha_N)), & \text{if } d = 2 \\ O(\alpha_N^{-2}), & \text{if } d \geq 3, \end{cases}$$

for Type B sampling scheme (provided the rate is slower than N^{-1}).

The last example shows that for very persistent spatial dependence, the rate of consistency can be essentially arbitrarily slow.

EXAMPLE 6. Assume that $h(x)$ decays only at a logarithmic rate, $h(x) = \{\log(x \vee e)\}^{-1}$. Then, for any $d \geq 1$, the left hand side in (16) is $\ll (\log \alpha_N)^{-1}$.

We now turn to the case of the random design.

Proposition 3. *Assume the random sampling design of Section 2.5. If the sequence $\{\mathbf{s}_{k,N}\}$ is independent of the process X , and if Assumption 1 holds, then we have for any $\varepsilon_N > 0$*

$$E \left\| \frac{1}{N} \sum_{k=1}^N X(\mathbf{s}_{k,N}) - \mu \right\|^2 \leq 6 h(0) \sup_{\mathbf{s} \in R_0} f^2(\mathbf{s}) \varepsilon_N^d + h(\alpha_N \varepsilon_N) + \frac{h(0)}{N}.$$

Choosing ε_N such that $\varepsilon_N \rightarrow 0$ and $\alpha_N \varepsilon_N \rightarrow \infty$, it follows that under Type B or Type C sampling, the sample mean is consistent.

The bound in Proposition 3 can be easily applied to any specific random sampling design and any model for the functions ϕ_j in (8). It nicely shows that what matters for the rate of consistency is the interplay between the rate of growth of the sampling domain and the rate of decay of dependence.

Let us explain in slightly more detail a Type C sampling situation. Here typically we have $\alpha_N = N^{1/d}$. Then taking $\varepsilon_N = aN^{-1/d} \log N$, $a > 0$, we see that the rate of consistency is $h(a \log N) \vee N^{-1}$. For typical covariance functions ϕ_j , like powered exponential, Matérn or spherical, $h(a \log N)$ decays faster than N^{-1} . In such cases, the rate of consistency is, up to some logarithmic factor, the same as for an iid sample. For ease of reference, we formulate the following corollary, which can be used in practical applications.

Corollary 1. *Assume the random sampling design of Section 2.5 with the sequence $\{\mathbf{s}_{k,N}\}$ independent the process X . Suppose that $X(\mathbf{s})$ has representation (8) and that (9) holds with the ϕ_j in one of the families specified in Example 2. If Condition (11) holds, and $\alpha_N \geq N^{1/d}$ then (14) holds up to some multiplicative logarithmic factor.*

4. Consistency of the empirical covariance operator

In Section 3 we found the rates of consistency for the functional sample mean. We now turn to the rates for the sample covariance operator. Assuming the functional observations have mean zero, the natural estimator of the covariance operator C is the sample covariance operator given by

$$\hat{C}_N = \frac{1}{N} \sum_{k=1}^N X(\mathbf{s}_k) \otimes X(\mathbf{s}_k).$$

In general, the sample covariance operator is defined by

$$\hat{\Gamma}_N = \frac{1}{N} \sum_{k=1}^N (X(\mathbf{s}_k) - \bar{X}_N) \otimes (X(\mathbf{s}_k) - \bar{X}_N),$$

where

$$\bar{X}_N = \frac{1}{N} \sum_{k=1}^N X(\mathbf{s}_k).$$

Both operators are implemented in statistical software packages, for example in the popular R package `FDA` and in a similar MATLAB package, see Ramsay *et al.* (2009). The operator $\hat{\Gamma}_N$ is used to compute the EFPC's for centered data, while \hat{C}_N for data without centering.

We first derive the rates of consistency for \hat{C}_N assuming $EX(\mathbf{s}) = 0$. Then we turn to the operator $\hat{\Gamma}_N$. The proofs are obtained by applying the technique developed for the estimation of the functional mean. It is a general approach based on the estimation of the second moments of an appropriate norm (between estimator and estimand) so that the conditions in Definition 1 can come into play. It is broadly applicable to all statistics obtained by simple averaging of quantities defined at single spatial location. The proofs are thus similar to those presented in the simplest case in Section 6, but the notation becomes more cumbersome because of the increased complexity of the objects to be averaged. To conserve space these proofs are not included.

We begin by observing that

$$\begin{aligned} E\|\hat{C}_N - C\|_{\mathcal{S}}^2 &= E\langle \hat{C}_N - C, \hat{C}_N - C \rangle_{\mathcal{S}} \\ &= \frac{1}{N^2} \sum_{k=1}^N \sum_{\ell=1}^N E\langle X(\mathbf{s}_k) \otimes X(\mathbf{s}_k) - C, X(\mathbf{s}_\ell) \otimes X(\mathbf{s}_\ell) - C \rangle_{\mathcal{S}}. \end{aligned}$$

It follows that under Assumption 2

$$E\|\hat{C}_N - C\|_{\mathcal{S}}^2 \leq \frac{1}{N^2} \sum_{k=1}^N \sum_{\ell=1}^N H(\|\mathbf{s}_k - \mathbf{s}_\ell\|_2). \quad (19)$$

Relation (19) is used as the starting point of all proofs, cf. the proof of Proposition 1 in Section 3. Modifying the proofs of Section 3, we arrive at the following results.

Proposition 4. *Let Assumption 2 hold, and assume that \mathfrak{S}_N defines a non-random design of Type A, B or C. Then for any $\rho_N > 0$*

$$E\|\hat{C}_N - C\|_{\mathcal{S}}^2 \leq H(\rho_N) + H(0)I_{\rho_N}(\mathfrak{S}_N).$$

Hence under the Type B or Type C non-random sampling, with ρ_N as in (ii) of Definition 1, the empirical covariance operator is consistent.

Proposition 5. *Assume the sampling design of Section 2.4. Let Assumption 2 hold, with some function H such that $x^{d-1}H(x)$ is monotone on $[b, \infty)$, $b > 0$. Then under Type B sampling*

$$E\|\hat{C}_N - C\|_{\mathcal{S}}^2 \leq \frac{1}{\alpha_N^d} \left\{ d(3\Delta)^d \int_0^{K\alpha_N} x^{d-1} H(x) dx + o(1) \sup_{x \in [0, K\alpha_N]} x^{d-1} H(x) \right\},$$

for some large enough constant K which is independent of N . Under Type C sampling, the factor $1/\alpha_N^d$ is replaced by $O(N^{-1})$.

Proposition 6. *Assume the random sampling design of Section 2.5. If the sequence $\{\mathbf{s}_{k,N}\}$ is independent of the process X and if Assumption 2 holds, then we have for any $\varepsilon_N > 0$,*

$$E\|\widehat{C}_N - C\|^2 \leq 6H(0) \sup_{\mathbf{s} \in \mathcal{R}_0} f^2(\mathbf{s}) \varepsilon_N^d + H(\alpha_N \varepsilon_N) + \frac{H(0)}{N}.$$

It follows that under Type B or Type C sampling the sample covariance operator is consistent.

Introducing the (unobservable) operator

$$\tilde{\Gamma}_N = \frac{1}{N} \sum_{k=1}^N (X(\mathbf{s}_k) - \mu) \otimes (X(\mathbf{s}_k) - \mu),$$

we see that

$$\tilde{\Gamma}_N - \hat{\Gamma}_N = (\bar{X}_N - \mu) \otimes (\bar{X}_N - \mu).$$

Therefore

$$E\|\hat{\Gamma}_N - C\|_{\mathcal{S}}^2 \leq 2E\|\tilde{\Gamma}_N - C\|_{\mathcal{S}}^2 + 2E\|(\bar{X}_N - \mu) \otimes (\bar{X}_N - \mu)\|_{\mathcal{S}}^2.$$

The bounds in Propositions 4, 5 and 6 apply to $E\|\tilde{\Gamma}_N - C\|_{\mathcal{S}}^2$. Observe that

$$E\|(\bar{X}_N - \mu) \otimes (\bar{X}_N - \mu)\|_{\mathcal{S}}^2 = E\|\bar{X}_N - \mu\|^4.$$

If $X(\mathbf{s})$ are bounded variables, i.e. $\sup_{t \in [0,1]} |X(\mathbf{s}; t)| \leq B < \infty$ a.s., then $\|\bar{X}_N - \mu\|^4 \leq 4B^2 \|\bar{X}_N - \mu\|^2$. It follows that under Assumption 1 we obtain the same order of magnitude for the bounds of $E\|\bar{X}_N - \mu\|^4$ as we have obtained in Propositions 1, 2 and 3 for $E\|\bar{X}_N - \mu\|^2$. In general $E\|\bar{X}_N - \mu\|^4$ can neither be bounded in terms of $E\|\bar{X}_N - \mu\|^2$ nor with $E\|\hat{C}_N - C\|_{\mathcal{S}}^2$. To bound fourth order moments, conditions on the covariance between the variables $Z_{k,\ell} := \langle X(\mathbf{s}_{k,N}) - \mu, X(\mathbf{s}_{\ell,N}) - \mu \rangle$ and $Z_{i,j}$ for all $1 \leq i, j, k, \ell \leq N$ are unavoidable. However, a simpler general approach is to require higher order moments of $\|X(\mathbf{s})\|$. More precisely, we notice that for any $p > 1$, by the Hölder inequality,

$$E\|\bar{X}_N - \mu\|^4 \leq (E\|\bar{X}_N - \mu\|^2)^{1/p} \left(E\|\bar{X}_N - \mu\|^{\frac{4p-2}{p-1}} \right)^{(p-1)/p}.$$

Thus as long as $E\|X(\mathbf{s})\|^{\frac{4p-2}{p-1}} < \infty$, we conclude that, by stationarity,

$$E\|\bar{X}_N - \mu\|^4 \leq M(p) (E\|\bar{X}_N - \mu\|^2)^{1/p},$$

where $M(p)$ depends on the distribution of $X(\mathbf{s})$ and on p , but not on N . It is now evident how the results of Section 3 can be used to obtain bounds for $E\|\hat{\Gamma}_N - C\|_{\mathcal{S}}^2$. We state in Proposition 7 the version for the general non-random design. The special cases follow, and the random designs are treated analogously. It follows that if Assumptions 1 and 2 hold, then $E\|\hat{\Gamma}_N - C\|_{\mathcal{S}}^2 \rightarrow 0$, under Type B or C sampling, provided $E\|X(\mathbf{s})\|^{4+\delta} < \infty$.

Proposition 7. *Let Assumptions 1 and 2 hold and assume that for some $\delta > 0$ we have $E\|X(\mathbf{s})\|^{4+\delta} < \infty$. Assume further that \mathfrak{S}_N defines a non-random design of Type A, B or C. Then for any $\rho_N > 0$ we have*

$$E\|\hat{\Gamma}_N - C\|_{\mathfrak{S}}^2 \leq 2\{H(\rho_N) + H(0)I_{\rho_N}(\mathfrak{S}_N)\} + 2K(\delta)\{h(\rho_N) + h(0)I_{\rho_N}(\mathfrak{S}_N)\}^{\frac{\delta}{2+\delta}}, \quad (20)$$

where $K(\delta)$ depends only on δ .

If $X(\mathbf{s}_1)$ is a.s. bounded by some finite constant B , then we can formally let δ in (20) go to ∞ , with $K(\infty) = 4B^2$.

5. Inconsistent empirical functional principal components

We define $X^* = \langle X(\mathbf{0}), \cdot \rangle X(\mathbf{0})$. Observe that $X^*(X(\mathbf{0})) = \|X(\mathbf{0})\|^2 X(\mathbf{0})$. Thus $\|X(\mathbf{0})\|^2 = \sum_{j=1}^{\infty} \xi_j^2(\mathbf{0})$ is an eigenvalue of X^* . Note also that for $x \in L^2$,

$$X^*(x)(t) = \left(\int X(\mathbf{0}; u)x(u)du \right) X(\mathbf{0}; t) = \int c^*(t, u)x(u)du,$$

where

$$c^*(t, u) = X(\mathbf{0}; t)X(\mathbf{0}; u).$$

Since

$$E \iint (c^*(t, u))^2 dt du = E\|X(\mathbf{0})\|^4 < \infty,$$

the operator X^* is Hilbert–Schmidt almost surely.

Proposition 8. *Suppose representation (8) holds with stationary mean zero Gaussian processes ξ_j such that*

$$E[\xi_j(\mathbf{s})\xi_j(\mathbf{s} + \mathbf{h})] = \lambda_j \rho_j(h), \quad h = \|\mathbf{h}\|,$$

where each ρ_j is a continuous correlation function, and $\sum_j \lambda_j < \infty$. Assume the processes ξ_j and ξ_i are independent if $i \neq j$. If $\mathfrak{S}_N = \{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_N\} \subset \mathbb{R}^d$ with $\mathbf{s}_N \rightarrow \mathbf{0}$, then

$$\lim_{N \rightarrow \infty} E\|\hat{C}_N - X^*\|_{\mathfrak{S}}^2 = 0. \quad (21)$$

Proposition 8 is proven in Section 6. Since the eigenvalues of X^* are random they cannot be close to any of the λ_j . The eigenfunctions of \hat{C}_N are also close to random functions in L^2 , and do not converge to the FPC's e_j .

We now present a very specific example that illustrates Proposition 8.

EXAMPLE 7. Suppose

$$X(s; t) = \zeta_1(s)e_1(t) + \sqrt{\lambda}\zeta_2(s)e_2(t), \quad (22)$$

where the ζ_1 and ζ_2 are iid processes on the line, and $0 < \lambda < 1$. Assume that the processes ζ_1 and ζ_2 are Gaussian with mean zero and covariances $E[\zeta_j(s)\zeta_j(s+h)] = \exp\{-h^2\}$, $j = 1, 2$. Thus, each $Z_j := \zeta_j(0)$ is standard normal. Rearranging the terms, we obtain

$$X^*(x) = \left(Z_1^2 \langle x, e_1 \rangle + \sqrt{\lambda} Z_1 Z_2 \langle x, e_2 \rangle \right) e_1 + \left(\sqrt{\lambda} Z_1 Z_2 \langle x, e_1 \rangle + \lambda Z_2^2 \langle x, e_2 \rangle \right) e_2.$$

The matrix

$$\begin{bmatrix} Z_1^2 & \sqrt{\lambda} Z_1 Z_2 \\ \sqrt{\lambda} Z_1 Z_2 & \lambda Z_2^2 \end{bmatrix}$$

has only one positive eigenvalue $Z_1^2 + \lambda Z_2^2 = \|X(0)\|^2$. A normalized eigenfunction associated with it is

$$f := \frac{X(0)}{\|X(0)\|} = [Z_1^2 + \lambda Z_2^2]^{-1/2} \left(Z_1 e_1 + \sqrt{\lambda} Z_2 e_2 \right). \quad (23)$$

Denote by \hat{v}_1 a normalized eigenfunction corresponding to the largest eigenvalue of \hat{C}_N . By Lemma 1, \hat{v}_1 is close in probability to $\text{sign}(\langle \hat{v}_1, f \rangle) f$. It is thus not close to $\text{sign}(\langle \hat{v}_1, e_1 \rangle) e_1$.

Ten simulated \hat{v}_1 are shown in Figure 1. To compute each \hat{v}_1 , we simulated curves $X(s_n)$ given by (22) with $e_1(t) = \sqrt{2} \sin(2\pi t)$, $e_2(t) = \sqrt{2} \cos(2\pi t)$, $\lambda = 0.5$. We set $s_n = n^{-1}$, $n = 1, 2, \dots, N$, $N = 100$. The random vectors $[\zeta_j(s_1), \zeta_j(s_2), \dots, \zeta_j(s_N)]^T$ were generated using the R function `rmvnorm` which uses the singular value decomposition to simulate Gaussian vectors with predetermined covariances (in our case, $\text{Cov}(\zeta_j(s_k), \zeta_j(s_\ell)) = \exp\{-|k^{-1} - \ell^{-1}|\}$).

The EFPC \hat{v}_1 is a linear combination of e_1 and e_2 with random weights. As formula (23) suggests, the function e_1 is likely to receive a larger weight. The weights, and so the simulated \hat{v}_1 , cluster because both Z_1 and Z_2 are standard normal.

We now state a general result showing that Type A sampling generally leads to inconsistent estimators if the spatial dependence does not vanish.

Proposition 9. *Assume that $E\langle X(\mathbf{s}_1) - \mu, X(\mathbf{s}_2) - \mu \rangle \geq b(\|\mathbf{s}_1 - \mathbf{s}_2\|_2) > 0$, where $b(x)$ is non-increasing. Then under Type A sampling the sample mean \bar{X}_N is not a consistent estimator of μ . Similarly, if $EX(\mathbf{s}) = 0$ and*

$$E\langle X(\mathbf{s}_1) \otimes X(\mathbf{s}_1) - C, X(\mathbf{s}_2) \otimes X(\mathbf{s}_2) - C \rangle_{\mathcal{S}} \geq B(\|\mathbf{s}_1 - \mathbf{s}_2\|_2) > 0, \quad (24)$$

where $B(x)$ is non-increasing, then under Type A sampling the sample covariance \hat{C}_N is not a consistent estimator of C .

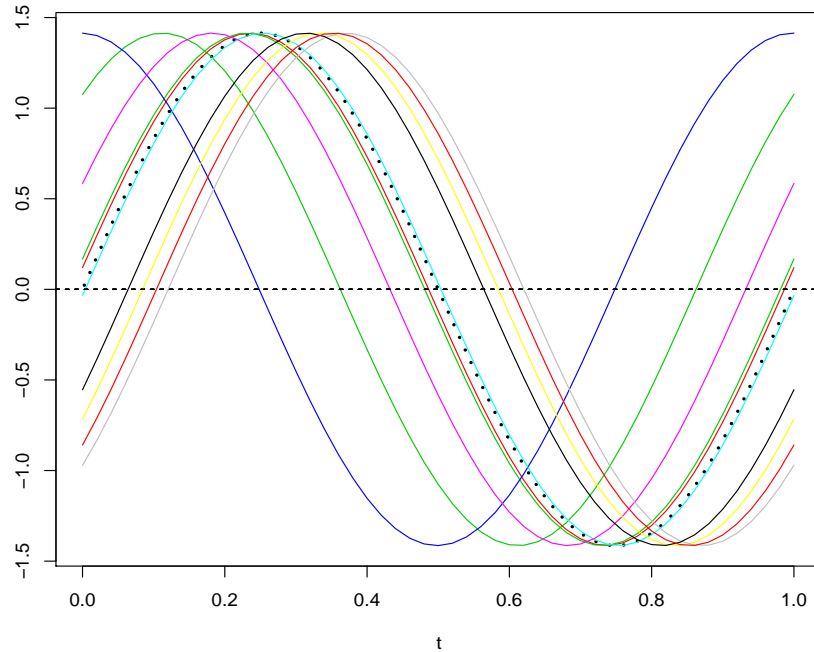


Figure 1. Ten simulated EFPC's \hat{v}_1 for process (22) with $\lambda = 0.5$ and $e_1(t) = \sqrt{2}\sin(2\pi t)$, $e_2(t) = \sqrt{2}\cos(2\pi t)$ ($N = 100$). The dotted line is the population eigenfunction.

We illustrate Proposition 9 with a further example that complements Proposition 8 in a sense that in Proposition 8 the functional model was complex, but the spatial distribution of the \mathbf{s}_k simple. In Example 8, we allow a general Type A distribution, but consider the simple model (22).

EXAMPLE 8. We focus on condition (24) for the FPC's. For the general model (8), the left-hand side of (24) is equal to

$$\kappa(\mathbf{s}_1, \mathbf{s}_2) = \sum_{i,j \geq 1} \text{Cov}(\xi_i(\mathbf{s}_1)\xi_j(\mathbf{s}_1), \xi_i(\mathbf{s}_2)\xi_j(\mathbf{s}_2)).$$

If the processes ξ_j satisfy the assumptions of Proposition 8, then, by Lemma 4,

$$\text{Cov}(\xi_i(\mathbf{s}_1)\xi_j(\mathbf{s}_1), \xi_i(\mathbf{s}_2)\xi_j(\mathbf{s}_2)) = \lambda_i^2 r_i + \lambda_j^2 r_j + \lambda_i \lambda_j \frac{r_i + r_j}{2} - \left(\lambda_i^{3/2} r_i + \lambda_j^{3/2} r_j \right) \sqrt{\lambda_i + \lambda_j},$$

where $r_i = \rho_i(\|\mathbf{s}_1 - \mathbf{s}_2\|)$.

To calculate $\kappa(\mathbf{s}_1, \mathbf{s}_2)$ in a simple case, corresponding to (22), suppose

$$\lambda_1 = 1, \lambda_2 = \lambda, 0 < \lambda < 1, \lambda_i = 0, i > 2, \quad \text{and} \quad \rho_1 = \rho_2 = \rho. \quad (25)$$

Then,

$$\kappa(\mathbf{s}_1, \mathbf{s}_2) = f(\lambda)\rho(\|\mathbf{s}_1 - \mathbf{s}_2\|),$$

where

$$f(\lambda) = (3 - 2\sqrt{2})(1 + \lambda^2) + 2 \left[1 + \lambda + \lambda^2 - (1 + \lambda^{3/2})(1 + \lambda)^{1/2} \right].$$

The function f increases from about 0.17 at $\lambda = 0$ to about 0.69 at $\lambda = 1$.

We have verified that if the functional random field (8) satisfies the assumptions of Proposition 8 and (25), then \widehat{C}_N is an inconsistent estimator of C under Type A sampling, whenever $\rho(h)$ is a nonincreasing function of h .

6. Proofs of the results of Sections 3, 4 and 5

We will use the following well-known Lemma.

Lemma 4. *Suppose X and Y are jointly normal mean zero random variables such that $EX^2 = \sigma^2$, $EY^2 = \nu^2$, $E[XY] = \rho\sigma\nu$. Then*

$$\text{Cov}(X^2, Y^2) = 2\rho^2\sigma^2\nu^2.$$

Proof of Proposition 1. By Assumption 1 we have

$$\begin{aligned} & E \left\| \frac{1}{N} \sum_{k=1}^N X(\mathbf{s}_{k,N}) - \mu \right\|^2 \\ &= \frac{1}{N^2} \sum_{k=1}^N \sum_{\ell=1}^N E \langle X(\mathbf{s}_{k,N}) - \mu, X(\mathbf{s}_{\ell,N}) - \mu \rangle \\ &\leq \frac{1}{N^2} \sum_{k=1}^N \sum_{\ell=1}^N h(\|\mathbf{s}_{k,N} - \mathbf{s}_{\ell,N}\|_2) \\ &\leq \frac{1}{N^2} \sum_{k=1}^N \sum_{\ell=1}^N (h(\rho_N) I\{\|\mathbf{s}_{k,N} - \mathbf{s}_{\ell,N}\|_2 \geq \rho_N\} + h(0) I\{\|\mathbf{s}_{k,N} - \mathbf{s}_{\ell,N}\|_2 \leq \rho_N\}) \\ &\leq h(\rho_N) + h(0) I_{\rho_N}(\mathfrak{S}_N). \end{aligned}$$

□

The following Lemma is a simple calculus problem and will be used in the proof of Proposition 2.

Lemma 5. *Assume that f is a non-negative function which is monotone on $[0, b]$ and on $[b, \infty)$. Then*

$$\sum_{k=0}^L f\left(\frac{k}{N}\right) \frac{1}{N} \leq \int_0^{L/N} f(x) dx + \frac{2}{N} \sup_{x \in [0, L/N]} |f(x)|.$$

Proof of Proposition 2. By Assumption 1,

$$E \left\| \frac{1}{S_N} \sum_{k=1}^{S_N} X(\mathbf{s}_{k,N}) - \mu \right\|^2 \leq \frac{1}{S_N^2} \sum_{k=1}^{S_N} \sum_{\ell=1}^{S_N} h(\|\mathbf{s}_{k,N} - \mathbf{s}_{\ell,N}\|_2).$$

Let $\mathbf{a} = (a_1, \dots, a_d)$ and $\mathbf{b} = (b_1, \dots, b_d)$ be two elements on $\mathcal{Z}(\delta)$. We define $d(\mathbf{a}, \mathbf{b}) = \min_{1 \leq i \leq d} v_i(\mathbf{a}, \mathbf{b})$, where $v_i(\mathbf{a}, \mathbf{b})$ is the number of edges between a_i and b_i . For any two points $\mathbf{s}_{k,N}$ and $\mathbf{s}_{\ell,N}$ we have

$$d(\mathbf{s}_{k,N}, \mathbf{s}_{\ell,N}) = m \quad \text{from some } m \in \{0, \dots, KN^{1/d}\}, \quad (26)$$

where K depends on $\text{diam}(R_0)$. It is easy to see that the number of points on the grid having distance m from a given point is less than $2d(2m+1)^d$, $m \geq 0$. Hence the number of pairs for which (26) holds is $< 2d(2m+1)^{d-1}N$. On the other hand, if $d(\mathbf{s}_{k,N}, \mathbf{s}_{\ell,N}) = m$, then $\|\mathbf{s}_{k,N} - \mathbf{s}_{\ell,N}\|_2 \geq m\delta_0\eta_N$. Let us assume without loss of generality that $\delta_0 = 1$. Noting that there is no loss of generality if we assume that $x^{\delta-1}h(x)$ is also monotone on $[0, b]$, we obtain by Lemma 5 for large enough N and $K < K' < K''$

$$\begin{aligned} & \frac{1}{S_N^2} \sum_{k=1}^{S_N} \sum_{\ell=1}^{S_N} h(\|\mathbf{s}_{k,N} - \mathbf{s}_{\ell,n}\|_2) \\ & \leq 2d \sum_{m=1}^{K'N^{1/d}} \frac{(2m+1)^{d-1}}{N} h(m\eta_N) + \frac{2h(0)}{N} \\ & \leq 2d \left(\frac{3}{\eta_N}\right)^{d-1} \sum_{m=0}^{K'N^{1/d}+1} \left(\frac{m}{N} N\eta_N\right)^{d-1} h\left(\frac{m}{N} N\eta_N\right) \frac{1}{N} + \frac{2h(0)}{N} \\ & \leq 2d \left(\frac{3}{\eta_N}\right)^{d-1} \left(\int_0^{K''N^{1/d-1}} (N\eta_N x)^{d-1} h(N\eta_N x) dx \right. \\ & \quad \left. + \frac{2}{N} \sup_{x \in [0, K''\alpha_N/\Delta]} x^{d-1} h(x) \right) + \frac{2h(0)}{N} \\ & = \frac{(3\Delta)^{dd}}{\alpha_N^d} \int_0^{K''\alpha_N/\Delta} x^{d-1} h(x) dx \\ & \quad + \frac{4d(3\Delta)^{d-1}}{\alpha_N^{d-1} N^{1/d}} \sup_{x \in [0, K''\alpha_N/\Delta]} x^{d-1} h(x) + \frac{2h(0)}{N}. \end{aligned}$$

By Lemma 2, Type B sampling implies $\alpha_N \rightarrow \infty$ and $\alpha_N = o(N^{1/d})$. This shows (16). Under Type C sampling $1/\alpha_N^d \ll 1/N$. The proof is finished. \square

Proof of Proposition 3. This time we have

$$\begin{aligned} E \left\| \frac{1}{N} \sum_{k=1}^N X(\mathbf{s}_{k,N}) - \mu \right\|^2 &\leq \frac{1}{N^2} \sum_{k=1}^N \sum_{\ell=1}^N E h(\|\mathbf{s}_{k,N} - \mathbf{s}_{\ell,N}\|_2) \\ &\leq \alpha_N^{-2d} \int_{R_N} \int_{R_N} h(\|\mathbf{s} - \mathbf{r}\|_2) f(\alpha_N^{-1} \mathbf{s}) f(\alpha_N^{-1} \mathbf{r}) d\mathbf{s} d\mathbf{r} + \frac{h(0)}{N} \\ &= \int_{R_0} \int_{R_0} h(\alpha_N \|\mathbf{s} - \mathbf{r}\|_2) f(\mathbf{s}) f(\mathbf{r}) d\mathbf{s} d\mathbf{r} + \frac{h(0)}{N}. \end{aligned}$$

Furthermore, for any $\varepsilon_N > 0$,

$$\begin{aligned} &\int_{R_0} \int_{R_0} h(\alpha_N \|\mathbf{s} - \mathbf{r}\|_2) f(\mathbf{s}) f(\mathbf{r}) d\mathbf{s} d\mathbf{r} \\ &\leq h(0) \int_{R_0} \int_{R_0} f(\mathbf{s}) f(\mathbf{r}) I\{\|\mathbf{s} - \mathbf{r}\|_2 \leq \varepsilon_N\} d\mathbf{s} d\mathbf{r} + h(\alpha_N \varepsilon_N) \\ &\leq h(0) \sup_{\mathbf{s} \in R_0} f^2(\mathbf{s}) \times \int_{R_0} \int_{R_0} I\{\|\mathbf{s} - \mathbf{r}\|_2 \leq \varepsilon_N\} d\mathbf{s} d\mathbf{r} + h(\alpha_N \varepsilon_N). \end{aligned}$$

Now for fixed \mathbf{r} it is not difficult to show that $\int_{R_0} I\{\|\mathbf{s} - \mathbf{r}\|_2 \leq \varepsilon_N\} d\mathbf{s} \leq 6\varepsilon_N^d$. (The constant 6 could be replaced with $\pi^{d/2}/\Gamma(d/2 + 1)$). \square

Proof of Proposition 8. Observe that

$$\|\widehat{C}_N - X^*\|_{\mathcal{S}}^2 = \iint \left\{ \frac{1}{N} \sum_{n=1}^N [X(\mathbf{s}_n; t)X(\mathbf{s}_n; u) - X(\mathbf{0}; t)X(\mathbf{0}; u)] \right\}^2 dt du.$$

Therefore,

$$\|\widehat{C}_N - X^*\|_{\mathcal{S}}^2 \leq 2I_1(N) + 2I_2(N),$$

where

$$I_1(N) = \iint \left\{ \frac{1}{N} \sum_{n=1}^N X(\mathbf{s}_n; t)(X(\mathbf{s}_n; u) - X(\mathbf{0}; u)) \right\}^2 dt du$$

and

$$I_2(N) = \iint \left\{ \frac{1}{N} \sum_{n=1}^N X(\mathbf{0}; u)(X(\mathbf{s}_n; t) - X(\mathbf{0}; t)) \right\}^2 dt du.$$

We will show that $EI_1(N) \rightarrow 0$. The argument for $I_2(N)$ is the same. Observe that

$$\begin{aligned} I_1(N) &= \frac{1}{N^2} \sum_{k,\ell=1}^N \iint X(\mathbf{s}_k; t)(X(\mathbf{s}_k; u) - X(\mathbf{0}; u))X(\mathbf{s}_\ell; t)(X(\mathbf{s}_\ell; u) - X(\mathbf{0}; u))dtdu \\ &= \frac{1}{N^2} \sum_{k,\ell=1}^N \int X(\mathbf{s}_k; t)X(\mathbf{s}_\ell; t)dt \int (X(\mathbf{s}_k; u) - X(\mathbf{0}; u))(X(\mathbf{s}_\ell; u) - X(\mathbf{0}; u))du. \end{aligned}$$

Thus,

$$EI_1(N) \leq \frac{1}{N^2} \sum_{k,\ell=1}^N \left\{ E \left(\int X(\mathbf{s}_k; t)X(\mathbf{s}_\ell; t)dt \right)^2 \right\}^{1/2} \left\{ E \left(\int Y_k(u)Y_\ell(u)du \right)^2 \right\}^{1/2},$$

where

$$Y_k(u) = X(\mathbf{s}_k; u) - X(\mathbf{0}; u).$$

We first deal with the integration over t :

$$\begin{aligned} E \left(\int X(\mathbf{s}_k; t)X(\mathbf{s}_\ell; t)dt \right)^2 &\leq E \int X^2(\mathbf{s}_k; t)dt \int X^2(\mathbf{s}_\ell; t)dt \\ &= E [\|X(\mathbf{s}_k)\|^2 \|X(\mathbf{s}_\ell)\|^2] \leq \{E\|X(\mathbf{s}_k)\|^4\}^{1/2} \{E\|X(\mathbf{s}_\ell)\|^4\}^{1/2} = E\|X(\mathbf{0})\|^4. \end{aligned}$$

We thus see that

$$\begin{aligned} EI_1(N) &\leq \{E\|X(\mathbf{0})\|^4\}^{1/2} \frac{1}{N^2} \sum_{k,\ell=1}^N \left\{ E \left(\int Y_k(u)Y_\ell(u)du \right)^2 \right\}^{1/2} \\ &\leq \{E\|X(\mathbf{0})\|^4\}^{1/2} \frac{1}{N^2} \sum_{k,\ell=1}^N \left\{ E \left(\int Y_k^2(u)du \right)^2 \right\}^{1/4} \left\{ E \left(\int Y_\ell^2(u)du \right)^2 \right\}^{1/4} \\ &= \{E\|X(\mathbf{0})\|^4\}^{1/2} \left[\frac{1}{N} \sum_{k=1}^N \left\{ E \left(\int Y_k^2(u)du \right)^2 \right\}^{1/4} \right]^2. \end{aligned}$$

Consequently, to complete the verification of (21), it suffices to show that

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N \left\{ E \left(\int Y_k^2(u)du \right)^2 \right\}^{1/4} = 0.$$

The above relation will follow from

$$\lim_{k \rightarrow \infty} E \left(\int Y_k^2(u)du \right)^2 = 0. \quad (27)$$

To verify (27), first notice that, by the orthonormality of the e_j ,

$$\int Y_k^2(u) du = \sum_{j=1}^{\infty} (\xi_j(\mathbf{s}_k) - \xi_j(\mathbf{0}))^2.$$

Therefore, by the independence of the processes ξ_j ,

$$\begin{aligned} E \left(\int Y_k^2(u) du \right)^2 &= \sum_{j=1}^{\infty} E (\xi_j(\mathbf{s}_k) - \xi_j(\mathbf{0}))^4 \\ &\quad + \sum_{i \neq j} E (\xi_i(\mathbf{s}_k) - \xi_i(\mathbf{0}))^2 E (\xi_j(\mathbf{s}_k) - \xi_j(\mathbf{0}))^2. \end{aligned}$$

The covariance structure was specified so that

$$E (\xi_j(\mathbf{s}_k) - \xi_j(\mathbf{0}))^2 = 2\lambda_j(1 - \rho_j(\|\mathbf{s}_k\|)),$$

so the normality yields

$$\begin{aligned} E \left(\int Y_k^2(u) du \right)^2 &\leq 12 \sum_{j=1}^{\infty} \lambda_j^2 (1 - \rho_j(\|\mathbf{s}_k\|))^2 \\ &\quad + 4 \left\{ \sum_{j=1}^{\infty} \lambda_j (1 - \rho_j(\|\mathbf{s}_k\|)) \right\}^2. \end{aligned}$$

The right hand side tends to zero by the Dominated Convergence Theorem. This establishes (27), and completes the proof of (21). \square

Proof of Proposition 9. We only check inconsistency of the sample mean. In view of the proof of Proposition 1 we have now the lower bound

$$\begin{aligned} E \left\| \frac{1}{N} \sum_{k=1}^N X(\mathbf{s}_{k,N}) - \mu \right\|^2 &\geq \frac{1}{N^2} \sum_{k=1}^N \sum_{\ell=1}^N b(\|\mathbf{s}_{k,N} - \mathbf{s}_{\ell,N}\|_2) \\ &\geq b(\rho) I_\rho^2(\mathfrak{S}_N), \end{aligned}$$

which is by assumption bounded away from zero for $N \rightarrow \infty$. \square

Acknowledgements.

The research was partially supported by NSF grants DMS-0804165 and DMS-0931948 at Utah State University, by the Banque National de Belgique and Communauté française de Belgique - Actions de Recherche Concertées (2010–2015). We thank O. Gromenko and X. Zhang for performing the numerical simulations reported in this paper. We also would like to thank two referees and the associated editor for giving clear guidelines to improve the presentation of this paper.

Supplementary Material

Supplement to "Consistency of the mean and the principal components of spatially distributed functional data"

(doi: COMPLETED BY THE TYPESETTER; .pdf). We provide additional examples, some remarks concerning Assumption 2 and the regular sampling design as well as a remark on the sharpness of our bounds.

References

- Bel, L., Bar-Hen, A., Cheddadi, R. and Petit, R. Bel *et al.* (2011). Spatio-temporal functional regression on paleo-ecological data. *Journal of Applied Statistics*, **38**, 695–704.
- Benko, M., Härdle, W. and Kneip, A. Benko *et al.* (2009). Common functional principal components. *The Annals of Statistics*, **37**, 1–34.
- Bolthausen, E. Bolthausen (1982). On the central limit theorem for stationary mixing random fields. *Annals of Probability*, **10**, 1047–1050.
- Bosq, D. Bosq (2000). *Linear Processes in Function Spaces*. Springer, New York.
- Cressie, N. A. C. Cressie (1993). *Statistics for Spatial Data*. Wiley.
- Delicado, P., Giraldo, R., Comas, C. and Mateu, J. Delicado *et al.* (2010). Statistics for spatial functional data: some recent contributions. *Environmetrics*, **21**, 224–239.
- Du, J., Zhang, H. and Mandrekar, V. S. Du *et al.* (2009). Fixed-domain asymptotic properties of tapered maximum likelihood estimators. *Annals of Statistics*, **37**, 3330–3361.
- Einmahl, J. H. J and Ruymgaart, F. G. Einmahl and Ruymgaart (1987). The order of magnitude of the moments of the modulus of continuity of the multiparameter Poisson and empirical processes. *Journal of Multivariate Analysis*, **21**, 263–273.
- Gabrys, R., Horváth, L. and Kokoszka, P. Gabrys *et al.* (2010). Tests for error correlation in the functional linear model. *Journal of the American Statistical Association*, **105**, 1113–1125.
- Gabrys, R. and Kokoszka, P. Gabrys and Kokoszka (2007). Portmanteau test of independence for functional observations. *Journal of the American Statistical Association*, **102**, 1338–1348.
- Gohberg, I., Golberg, S. and Kaashoek, M. A. Gohberg *et al.* (1990). *Classes of Linear Operators*. Operator Theory: Advances and Applications, volume 49. Birkhäuser.
- Hall, P. and Hosseini-Nasab, M. Hall and Hosseini-Nasab (2006). On properties of functional principal components. *J. R. Statist. Soc B*, **68**, 109–126.
- Hörmann, S. and Kokoszka, P. Hörmann and Kokoszka (2010). Weakly dependent functional data. *The Annals of Statistics*, **38**, 1845–1884.
- Hörmann, S. and Kokoszka, P. Hörmann and Kokoszka (2011). Supplement to "consistency of the mean and the principal components of spatially distributed functional data". Technical Report.

- Jenish, N. and Prucha, I.R. Jenish and Prucha (2009). Central limit theorems and uniform laws of large numbers for arrays of random fields. *Journal of Econometrics*, **150**, 86–98.
- Jiang, C-R. and Wang, J-L. Jiang and Wang (2010). Covariate adjusted functional principal components analysis for longitudinal data. *The Annals of Statistics*, **38**, 1194–1226.
- Lahiri, S. N. Lahiri (1996). On inconsistency of estimators based on spatial data under infill asymptotics. *Sankhya Series A*, **58**, 403–417.
- Lahiri, S. N. Lahiri (2003). Central limit theorems for weighted sums of a spatial process under a class of stochastic and fixed designs. *Sankhya, Series A*, **65**, 356–388.
- Lahiri, S. N. and Zhu, J. Lahiri and Zhu (2006). Resampling methods for spatial regression models under a class of stochastic designs. *Annals of Statistics*, **34**, 1774–1813.
- Loh, W.-L. Loh (2005). Fixed-domain asymptotics for a subclass of Matern-type Gaussian random fields. *Annals of Statistics*, **33**, 2344–2394.
- Park, B. U., Kim, T. Y., Park, J-S. and Hwang, S. Y. Park *et al.* (2009). Practically applicable central limit theorem for spatial statistics. *Mathematical Geosciences*, **41**, 555–569.
- Paul, D. and Peng, J. Paul and Peng (2009). Consistency of restricted maximum likelihood estimators of principal components. *The Annals of Statistics*, **37**, 1229–1271.
- Ramsay, J., Hooker, G. and Graves, S. Ramsay *et al.* (2009). *Functional Data Analysis with R and MATLAB*. Springer.
- Ramsay, J. O. and Silverman, B. W. Ramsay and Silverman (2005). *Functional Data Analysis*. Springer.
- Reiss, P. T. and Ogden, R. T. Reiss and Ogden (2007). Functional principal component regression and functional partial least squares. *Journal of the American Statistical Association*, **102**, 984–996.
- Rio, E. Rio (1993). Covariance inequalities for strongly mixing processes. *Ann. Inst. H. Poincaré Probab. Statist.*, **29**, 587–597.
- Stein, M. L. Stein (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer.
- Stute, W. Stute (1984). The oscillation behavior of empirical processes: the multivariate case. *The Annals of Probability*, **12**, 361–379.
- Yamanishi, Y. and Tanaka, Y. Yamanishi and Tanaka (2003). Geographically weighted functional multiple regression analysis: A numerical investigation. *Journal of the Japanese Society of Computational Statistics*, **15**, 307–317.
- Zhang, H. Zhang (2004). Inconsistent estimation and asymptotically equivalent interpolations in model-based geostatistics. *Journal of the American Statistical Association*, **99**, 250–261.