Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Principal component analysis of spatially indexed functions

Thomas Kuenzer, Siegfried Hörmann & Piotr Kokoszka

To cite this article: Thomas Kuenzer, Siegfried Hörmann & Piotr Kokoszka (2020): Principal component analysis of spatially indexed functions, Journal of the American Statistical Association, DOI: 10.1080/01621459.2020.1732395

To link to this article: https://doi.org/10.1080/01621459.2020.1732395

Principal component analysis of spatially indexed functions

Thomas Kuenzer

Technische Universität Graz

Siegfried Hörmann

Technische Universität Graz

Piotr Kokoszka

Colorado State University

Corresponding author Piotr Kokoszka Piotr.Kokoszka@colostate.edu

Abstract

We develop an expansion, similar in some respects to the Karhunen–Loève expansion, but which is more suitable for functional data indexed by spatial locations on a grid. Unlike the traditional Karhunen–Loève expansion, it takes into account the spatial dependence between the functions. By doing so, it provides a more efficient dimension reduction tool, both theoretically and in finite samples, for functional data with moderate spatial dependence. For such data, it also possesses other theoretical and practical advantages over the currently used approach. The paper develops complete asymptotic theory and estimation methodology. The performance of the method is examined by a simulation study and data analysis. The new tools are implemented in an R package.

Keywords: Functional data, Principal Components, Spatial data, Spectral Analysis.

1 Introduction

Functional principal components (FPCs) have played a prominent role in functional data analysis and continue to do so, as explained in the monographs of Bosq (2000), Ramsay and Silverman (2005), Horváth and Kokoszka (2012), Hsing and Eubank (2015), among many other contributions. Many functional data sets have the form $X_{\mathbf{s}}(u), \mathbf{s} \in \mathcal{S},\ u \in \mathcal{U}$, where **s** is a spatial location, and $\mathcal{U}$ is a time interval. The curves $X_{\mathbf{s}}$ can be derived from

original observations, or are model residuals, many references are given in Chapter 9 of Kokoszka and Reimherr (2017). For example, $X_{\mathbf{s}}(u)$ can be maximum temperature at location **s** on day u in a time period $\mathcal{U}$, which represents a year or a longer time period. In inferential procedures, it is generally assumed that the functions $X_{\mathbf{s}}$ arise as a realization of a spatially stationary random field $\{X_{\mathbf{s}}, \mathbf{s} \in \mathbb{Z}^r\}$. To define the FPCs, it is enough to assume that the distribution of $X_{\mathbf{s}}$ does not depend on **s**. In that case, the FPCs $v_m, m \geq 1,$ are the eigenfunctions of the covariance operator of any and each $X_{\mathbf{s}}$. The Karhunen–Loève expansion is then (assuming zero mean function),

$$X_{\mathbf{s}}(u) = \sum_{m=1}^{\infty} \xi_{m,\mathbf{s}} v_m(u), \; u \in \mathcal{U}. \quad (1.1)$$

The FPCs v_m in (1.1) are not in any way related to the spatial dependence structure of the random field $\{X_{\mathbf{s}}, \mathbf{s} \in \mathbb{Z}^r\}$; if $\{X_{\mathbf{s}}', \mathbf{s} \in \mathbb{Z}^r\}$ is a field of iid functions such that for each **s**, $X_{\mathbf{s}} \overset{d}{=} X_{\mathbf{s}}'$, then $v_m = v_m'$. The objective of this paper is to develop a representation of functional data on a spatial domain which reflects spatial dependence, which is a chief characteristic of such data. At the population level, it will take the form

$$X_{\mathbf{s}}(u) = \sum_{m=1}^{\infty} \sum_{\mathbf{k}} Y_m(\mathbf{s}+\mathbf{k}) \phi_m(\mathbf{k})(u). \quad (1.2)$$

The index m has a similar interpretation as in (1.1), components with small m describe large scale behavior of the data. However, the single function v_m in (1.1) is replaced by the family of functions $\phi_m(\mathbf{k})$ indexed by spatial locations **k**. For each m, $\phi_m(\cdot)$ can be viewed as a deterministic field on $\mathbb{Z}^r$ with values in a function space. The form of these fields reflects spatial dependence. If there is no spatial dependence, they will vanish at all locations **k** except for the origin $\mathbf{0}$.

Expansion (1.2) should be contrasted with the Karhunen–Loève expansion of a *scalar* spatial field, $\eta(\mathbf{s}) = \sum_{j=1}^{\infty} a_j \alpha_j(\mathbf{s})$. The functions α_j are often called the empirical orthogonal functions (EOFs). This EOF analysis does not involve any functional objects. The EOFs can be estimated assuming iid replications η_t in discrete time t are available, in which case it does not require spatial stationarity, see Section 5.3 in Cressie and Wikle (2011). Research on EOFs, dates back almost seventy years, with many authors citing the work of Lorenz (1956) as initiating this methodology. A thorough review of these methods is given in Chapter 12 of Jolliffe (2002), which includes the functional PCA (Section 12.3) and the frequency domain PCA of Brillinger (1975) (Section 12.4), and several other more specialized techniques in earlier sections.

Expansion (1.2) is derived for a weakly stationary functional random field on $\mathbb{Z}^r$. It is based on spectral methods, which had to be purpose developed. We build on the theory of Brillinger (1975), Panaretos and Tavakoli (2013a), Panaretos and Tavakoli (2013b), Hörmann et al. (2015), and several other contributions which are introduced in the following. Panaretos and Tavakoli (2013b) emphasize connections to multivariate stationary random fields and Bochner's theorem and point out that ideas of this type can be found already in the monograph of Grenander and Rosenblatt (1957). We will show that expansion (1.2) leads to an optimal dimension reduction, and in this sense is superior to the ordinary FPC expansion (1.1). The scores $Y_m(\mathbf{s})$ and $Y_{m'}(\mathbf{s}')$ are uncorrelated whenever $m' \neq m$, leading to a convenient multilevel analysis of variance. Beyond the population level theory, we propose methodology for the estimation of all components in (1.2) and investigate its large and finite sample properties. We thus introduce a new framework, which we will refer to as *Spatial Functional Principal Components Analysis* (SFPCA).

Data sets whose analysis may benefit from the new methodology include outputs of computer climate models, e.g. Mearns et al. (2009), functional magnetic resonance imaging (fMRI) or positron emission tomography (PET),

e.g. Petersen et al. (2010), and remote sensing data, e.g. Campbell and Wynne (2011). All these data can be viewed as functions indexed by grid coordinates.

We illustrate the benefits of the new methodology using an example from the last field. Figure 1 shows a hyperspectral image together with a function $X_{\mathbf{s}}(u)$, for a single selected location **s** from a 614 × 512 grid. The argument u is the frequency. Details of this data and the application of the new methodology are presented in Section III of the Supplementary Materials. Table 1 shows approximation errors of dimension reduction based on the first p estimated components in (1.2) (Spatial) and (1.1) (Ordinary). With the same number of components, the new expansion provides a better approximation.

The body of research on spatial functional data is growing, and it is not possible to note numerous useful contributions. We list only a few papers, which we are most familiar with, and which motivated the research presented in this paper. Delicado et al. (2010) provide a useful introduction to selected methods for geostatistical functional data; the collection Mateu and Giraldo (2019) provides a comprehensive resource. Zhang et al. (2011) propose an early application of functional data analysis to climate studies. Gromenko *et al* (2012, 2017) use a framework of spatially indexed functional data to solve space physics problems. Aston et al. (2017) and Constantinou et al. (2017) develop tests of spatio–temporal separability for functional data. Liu et al. (2017) consider tests of anisotropy using expansion (1.1), but with a semiparametric estimation procedure custom-developed for their objectives. Happ and Greven (2018) extend representation (1.1) to the case of vector–valued observations, $\mathbf{X}_{\mathbf{s}}$, whose components are defined on possibly different domains, including spatial domains. The work of Bernardi et al. (2017) is an example of a substantive application of spatial functional modeling to a problem of practical importance.

The paper is organized as follows. Section 2 develops a spectral domain framework, which is needed to derive expansion (1.2). The properties and estimation of this expansion are addressed in Section 3. Section 4 focuses on

the finite sample behavior and real data applications. The contribution of the paper is summarized in Section 5, which also contains a discussion of further applications of our methodology.

Proofs of mathematical results and some details of data analysis are presented in Supplementary Materials. All computations presented in this paper are implemented in a purpose developed `R` package `fsd`, which internally calls selected functions of the `fda` package. Some plots were drawn using the `autoimage` package.

2 Spectral analysis of functional random fields

In this section we present a number of results, which are fundamental for the development of any frequency domain methodology and theory for functional random fields on a grid. While many of these results have analogs either in the frequency analysis of time series, e.g. Brockwell and Davis (1991), Shumway and Stoffer (2017), or in spatial statistics, e.g. Stein (1999), Chilès and Delfiner (1999), Gelfand et al. (2010), no spectral domain results are at present available for Hilbert space valued random fields. It is hoped that by filling this gap, we present in this section a useful reference which other researchers may draw upon. Section 2.4 is dedicated to spatial functional ARMA processes. All proofs of the results of this section are collected in Section I of the Supplementary Materials.

2.1 Notation and assumptions

Vectors in $\mathbb{Z}^r$ are denoted with bold font, with the null vector written as a bold zero, $\mathbf{0}$. The components of some vectors **s**, **h**, etc. are denoted by s_i, h_i, etc. We use $\|\mathbf{h}\|_\infty$ to denote the maximum norm of $\mathbf{h} \in \mathbb{Z}^r$, while $\|\mathbf{h}\|$ is its Euclidean norm.

We consider a functional random field $(X_\mathbf{s})_{\mathbf{s}\in\mathbb{Z}^r}$, with a fixed dimension $r \in \mathbb{N}$. Each $X_\mathbf{s}$ is a random function in the separable Hilbert space $H = L^2([0,1])$ of complex square-integrable functions on $[0,1]$. The inner product on H is defined by $\langle x, y\rangle = \int_0^1 x(u)\bar{y}(u)du$; it is linear in the first and antilinear in the

second argument. The norm is denoted by $\|x\| = \sqrt{\langle x, x \rangle}$. We define the tensor product as an operator on H by

$$x \otimes y(\cdot) := \langle \cdot, y \rangle x.$$

We will work within three classes of linear operators $A: H \to H'$, where H' is another separable Hilbert space. These are *trace class*, *Hilbert–Schmidt* and *bounded* operators. They are Schatten spaces $\mathcal{S}_p$, and it is useful to keep the following inclusions in mind

$$\text{Trace class } = \mathcal{S}_1 \subset \text{Hilbert} - \text{Schmidt } = \mathcal{S}_2 \subset \text{Bounded} = \mathcal{S}_\infty.$$

The corresponding norms are defined as

$$\|A\|_{\mathcal{S}_1} = \sum_{l \in \mathbb{N}} \left\langle (A^* A)^{1/2} \varphi_l, \varphi_l \right\rangle, \ \|A\|_{\mathcal{S}_2}^2 = \sum_{l \in \mathbb{N}} \|A\varphi_l\|^2, \ \|A\|_{\mathcal{S}_\infty} = \sup_{\|x\| \le 1} \|Ax\|,$$

where $\{\varphi_l, l \ge 1\}$ is any complete orthonormal system (CONS). These norms satisfy $\|A\|_{\mathcal{S}_\infty} \le \|A\|_{\mathcal{S}_2} \le \|A\|_{\mathcal{S}_1}$. Further details can be found e.g. in Chapter 4 of Hsing and Eubank (2015).

For a random element X of H, we write $X \in L_H^p$, for a $p > 0$, if $\mathbb{E} \| X \|^p < \infty$. If $X \in L_H^1$, we can define the mean curve μ by $\mu(u) = \mathbb{E}X(u)$. Any $X \in L_H^2$ possesses a covariance operator $C := \mathbb{E}\left[(X - \mu) \otimes (X - \mu)\right]$ with kernel $c(u, v) := \text{Cov}(X(u), X(v))$.

Further background on random elements and operators in Hilbert spaces is provided e.g. in Bosq (2000), Horváth and Kokoszka (2012) and Hsing and Eubank (2015).

The theory and methodology developed in this paper applies to stationary random fields defined as follows.

Definition 2.1. A functional random field $(X_{\mathbf{s}})_{\mathbf{s} \in \mathbb{Z}^r}$ is called weakly stationary, if for all $\mathbf{s} \in \mathbb{Z}^r$, $X_{\mathbf{s}} \in L_H^2$ and

(i) $\mathbb{E}X_{\mathbf{s}} = \mathbb{E}X_{\mathbf{0}}$,

(ii) for all $\mathbf{s},\mathbf{h} \in \mathbb{Z}^r$ and $u,v \in [0,1]$

$$c_{\mathbf{h}}(u,v) := \mathrm{Cov}(X_{\mathbf{h}}(u), X_{\mathbf{0}}(v)) = \mathrm{Cov}(X_{\mathbf{s}+\mathbf{h}}(u), X_{\mathbf{s}}(v)).$$

The integral operator defined by the autocovariance kernel $c_{\mathbf{h}}$ is then denoted by $C_{\mathbf{h}}$.

Throughout the paper, we work under the following assumption.

Assumption 2.1. *The field $(X_{\mathbf{s}})_{\mathbf{s}\in\mathbb{Z}^r}$ is weakly stationary according to Definition 2.1 and has mean zero.*

Another assumption, which will often be made, is that the autocovariance operators of $(X_{\mathbf{s}})_{\mathbf{s}\in\mathbb{Z}^r}$ are absolutely summable:

$$\sum_{\mathbf{h}\in\mathbb{Z}^r} \|C_{\mathbf{h}}\|_{\mathcal{S}_2} < \infty. \qquad \text{(A.1)}$$

Condition (A.1) is a functional analog of the summability of spatial covariances of a stationary scalar field. It is thus a very mild assumption, as the most widely used parametric covariance models exhibit at least exponential decay, or even a finite range, see Gelfand et al. (2010), Section 3.5, or Chilès and Delfiner (1999), Section 2.5. This is also the case for functional data, e.g. the algorithm for principal component analysis proposed by Liu et al. (2017) uses a Matérn model.

2.2 Spectral density operator and inversion formula

The following definition generalizes the concept of the spectral density of a functional time series, which was introduced by Panaretos and Tavakoli (2013a).

Definition 2.2. Let $(X_{\mathbf{s}})_{\mathbf{s}\in\mathbb{Z}^r}$ be a weakly stationary functional random field. Set

$$f_\theta^X(u,v) := \frac{1}{(2\pi)^r} \sum_{h \in \mathbb{Z}^r} c_{\mathbf{h}}(u,v) e^{-\mathrm{i}\mathbf{h}^\top \theta}, \qquad \boldsymbol{\theta} \in [-\pi,\pi]^r,$$

where $\mathrm{i} = \sqrt{-1}$ denotes the imaginary unit. The integral operator with the above kernel is called the spectral density operator of $(X_{\mathbf{s}})$ at the spatial frequency $\boldsymbol{\theta}$. We denote it by $\mathcal{F}_\theta^X$.

The operators $\mathcal{F}_\theta^X$ are understood as elements of the space $L^2_{\mathcal{S}(H,H)}([-\pi,\pi]^r)$, which needs some explanation. Here $\mathcal{S}(H,H)$ is the Hilbert space of all Hilbert–Schmidt operators from *H* to *H* (which defines a separable Hilbert space if *H* is) and for any separable Hilbert space $\mathcal{H}$ we define $L^2_{\mathcal{H}}([-\pi,\pi]^r)$ as the space of measurable mappings $x : [-\pi,\pi]^r \to \mathcal{H}$ satisfying $\int_{[-\pi,\pi]^r} \|x(\boldsymbol{\theta})\|^2 d\boldsymbol{\theta} < \infty$. Using the inner product

$$(x,y)_2 := \frac{1}{(2\pi)^r} \int_{[-\pi,\pi]^r} \langle x(\boldsymbol{\theta}), y(\boldsymbol{\theta}) \rangle d\boldsymbol{\theta}$$

with corresponding norm $\|\cdot\|_2$, $L^2_{\mathcal{H}}([-\pi,\pi]^r)$ becomes yet another Hilbert space.

Proposition 2.1 lists fundamental properties of the operator $\mathcal{F}_\theta^X$. Its proof is similar to the proof of Proposition 7 in Hörmann et al. (2015), and so is omitted. It states, in particular, that the $\mathcal{F}_\theta^X$ have versions such that $\boldsymbol{\theta} \mapsto \mathcal{F}_\theta^X$ is continuous. We use these continuous versions in some proofs.

Proposition 2.1. *Let Assumption 2.1 and condition (A.1) hold. Then, for every spatial frequency* $\boldsymbol{\theta}$*, the operator* $\mathcal{F}_\theta^X$ *is a non-negative, self-adjoint Hilbert–Schmidt operator, so it admits the decomposition*

$$\mathcal{F}_\theta^X = \sum_{m \geq 1} \lambda_m(\boldsymbol{\theta})\, \varphi_m(\boldsymbol{\theta}) \otimes \varphi_m(\boldsymbol{\theta}), \qquad (2.1)$$

using its eigenvalues $\lambda_1(\boldsymbol{\theta}) \geq \lambda_2(\boldsymbol{\theta}) \geq \ldots \geq 0$ *and standardized eigenfunctions* $\varphi_m(\boldsymbol{\theta}) \in H$*. Furthermore,* $\boldsymbol{\theta} \mapsto \mathcal{F}_\theta^X$ *has a continuous version and it holds that*

(a) the eigenvalues $\lambda_m(\boldsymbol{\theta})$ *are continuous functions of* $\boldsymbol{\theta}$ *,*

(b) if (A.1) *is strengthened to*

$$\sum_{\mathbf{h}\in\mathbb{Z}^r} \|\mathbf{h}\|_\infty \|C_{\mathbf{h}}\|_{S_2} < \infty, \quad \text{(A.2)}$$

then the eigenvalues $\lambda_m(\boldsymbol{\theta})$ *are Lipschitz-continuous functions of* $\boldsymbol{\theta}$ *,*

(c) *if* $X_{\mathbf{s}}$ *is real-valued, then for any* $m \geq 1$ *and any spatial frequency* $\boldsymbol{\theta}$ *, the eigenvalues* $\lambda_m(\boldsymbol{\theta})$ *are even functions and the eigenfunctions* $\varphi_m(\boldsymbol{\theta}) \in H$ *are Hermitian in* $\boldsymbol{\theta}$ *, i.e.* $\lambda_m(\boldsymbol{\theta}) = \lambda_m(-\boldsymbol{\theta})$ *and* $\varphi_m(\boldsymbol{\theta}) = \overline{\varphi_m(-\boldsymbol{\theta})}$ *.*

Remark 2.1. The norm $\|\cdot\|_\infty$ in (A.2) can be replaced by any norm $\|\cdot\|$ on $\mathbb{Z}^r$ because of the equivalence of norms in finite-dimensional spaces.

We now consider an inversion formula, which plays a central role in our mathematical development. While the result is analogous to a finite-dimensional inversion formula, we were unable to find a Hilbert–space formulation we need. We will denote by $e_{\mathbf{k}} \in L^2([-\pi,\pi]^r)$ the function $e_{\mathbf{k}}(\boldsymbol{\theta}) = e^{\mathrm{i}\mathbf{k}^\top\boldsymbol{\theta}}$.

Proposition 2.2. *Suppose that* $x \in L_H^2([-\pi,\pi]^r)$ *and define the* ***k****-th Fourier coefficient* $f_{\mathbf{k}} \in H$ *by*

$$f_{\mathbf{k}} := \frac{1}{(2\pi)^r} \int_{[-\pi,\pi]^r} x(\boldsymbol{\theta}) e^{-\mathrm{i}\mathbf{k}^\top\boldsymbol{\theta}} d\boldsymbol{\theta}.$$

The sequence of partial sums

$$S_n(\boldsymbol{\theta}) := \sum_{\|\mathbf{k}\|_\infty \leq n} f_{\mathbf{k}} e_{\mathbf{k}} = \sum_{\|\mathbf{k}\|_\infty \leq n} f_{\mathbf{k}} e^{\mathrm{i}\mathbf{k}^\top\boldsymbol{\theta}}$$

has a mean square limit $S \in L_H^2([-\pi,\pi]^r)$ *with* $x(\boldsymbol{\theta}) = S(\boldsymbol{\theta})$ *for almost all* $\boldsymbol{\theta}$ *, i.e.*

$$\int_{[-\pi,\pi]^r} \|S_n(\boldsymbol{\theta}) - x(\boldsymbol{\theta})\|^2 d\boldsymbol{\theta} \xrightarrow{n\to\infty} 0.$$

2.3 Spatial dependence and linear filters

To develop large sample justifications for the methodology we propose, spatial dependence conditions, which go beyond stationarity, must be imposed. We develop in this section, in the setting of functional random fields, the concept of L^p-m-approximability introduced in the context of functional time series by Hörmann and Kokoszka (2010). We note that similar concepts were used earlier, e.g. by Shao and Wu (2007). The L^p-m-approximability has emerged as a useful assumption in the asymptotic analysis of functional time series, see Horváth et al. (2013), Horváth et al. (2014) and Zhang (2016), among many other contributions. We expect that our theory would remain valid under different dependence assumptions, most notably various forms of mixing, like those used in Prause and Steland (2018), or analogs of cumulant conditions used by Panaretos and Tavakoli (2013a). Care would have to be taken because not all mixing concepts are meaningful for random fields, see Bradley (2007), p. 205.

Definition 2.3. Let $(\varepsilon_{\mathbf{s}})_{\mathbf{s}\in\mathbb{Z}^r} \subseteq S$ be an i.i.d. random field taking values in some measurable space S and let $f : S^{\mathbb{Z}^r} \to H$ be a measurable function. Suppose the functional random field $(X_{\mathbf{s}})_{\mathbf{s}\in\mathbb{Z}^r} \subseteq H$ has the form $X_{\mathbf{s}} = f((\varepsilon_{\mathbf{s}-\mathbf{k}})_{\mathbf{k}\in\mathbb{Z}^r})$. Taking an independent copy $(\varepsilon'_{\mathbf{s}})_{\mathbf{s}\in\mathbb{Z}^r}$ of $(\varepsilon_{\mathbf{s}})_{\mathbf{s}\in\mathbb{Z}^r}$ and $m \in \mathbb{N}$, we define $(X^{(m)}_{\mathbf{s}})_{\mathbf{s}\in\mathbb{Z}^r}$ by

$$X^{(m)}_{\mathbf{s}} = f((\, 1_{\{\|\mathbf{k}\|_\infty < m\}}\, \varepsilon_{\mathbf{s}-\mathbf{k}} + 1_{\{\|\mathbf{k}\|_\infty \geq m\}}\, \varepsilon'_{\mathbf{s}-\mathbf{k}})_{\mathbf{k}\in\mathbb{Z}^r}). \qquad (2.2)$$

For $1 \leq p < \infty$, we say that $(X_{\mathbf{s}})_{\mathbf{s}\in\mathbb{Z}^r}$ is *L^p-m-approximable* if

$$\sum_{\mathbf{h}\in\mathbb{Z}^r} \left(\mathbb{E} \left\| X_{\mathbf{0}} - X^{(\|\mathbf{h}\|_\infty)}_{\mathbf{0}} \right\|^p \right)^{1/p} < \infty. \qquad (2.3)$$

The essence of Definition 2.3 is that the impact of innovations $\varepsilon_{\mathbf{s}'}$ for $\mathbf{s}'$ far away from the location $\mathbf{s}$ becomes negligible; they can be replaced by independent copies. The effect of this replacement is quantified by (2.3); the

convergence of the infinite series implies that the mean L^p distance tends to zero, as the spatial separation increases.

To lighten the notation, we henceforth set

$$\nu_p(Y) := \left(\mathbb{E}\left[\|Y\|^p\right]\right)^{1/p}.$$

Then the definition of L^p-m-approximability in (2.3) can be modified to the equivalent condition

$$\sum_{m\in\mathbb{N}} m^{r-1} \nu_p\left(X_{\mathbf{0}} - X_{\mathbf{0}}^{(m)}\right) < \infty. \quad (2.4)$$

This follows easily from the fact that $(X_{\mathbf{s}})$ is stationary and that $\left|\left\{\mathbf{h}\in\mathbb{Z}^r : \|\mathbf{h}\|_\infty = m\right\}\right|$ has order of magnitude m^{r-1}. Setting $r = 1$, Definition 2.3 almost coincides with the definition in the context of time series proposed by Hörmann and Kokoszka (2010). The difference is that the defining function f is naturally only one-sided for time series, while in our setting the function f always uses $\varepsilon_{\mathbf{s}}$ on the entire grid.

Our first proposition shows that L^2-m-approximability is sufficient for most results established in this work.

Proposition 2.3. *If* $(X_{\mathbf{s}})_{\mathbf{s}\in\mathbb{Z}^r}$ *is* L^2*-m-approximable, then* (A.1) *holds and the operators* $\mathcal{F}_{\boldsymbol{\theta}}^X$, $\boldsymbol{\theta}\in[-\pi,\pi]^r$, *are trace-class.*

The following proposition establishes a sufficient condition for a linear functional random field to be L^p-m-approximable. Scalar linear fields have been extensively studied, see e.g. El Machkouri et al. (2013) and Meyer et al. (2017), and references in those papers.

Proposition 2.4. *Let* $(X_{\mathbf{s}})_{\mathbf{s}\in\mathbb{Z}^r}$ *be a linear process, i.e.*

$$X_{\mathbf{s}} = \sum_{\mathbf{k}\in\mathbb{Z}^r} \Psi_{\mathbf{k}} \varepsilon_{\mathbf{s}-\mathbf{k}}, \quad (2.5)$$

where $(\varepsilon_{\mathbf{s}})_{\mathbf{s}\in\mathbb{Z}^r}$ *is an i.i.d. functional random field with* $\nu_p(\varepsilon_{\mathbf{0}})<\infty$ *for some* $p\geq 1$*, and* $\Psi_{\mathbf{k}}$ *are operators on H. If the operators* $\Psi_{\mathbf{k}}$ *satisfy*

$$\sum_{m\in\mathbb{N}} m^{r-1}\sum_{\|\mathbf{k}\|_\infty\geq m}\|\Psi_{\mathbf{k}}\|<\infty,$$

then $(X_{\mathbf{s}})_{\mathbf{s}\in\mathbb{Z}^r}$ *is L^p-m-approximable.*

Linear filters — (2.5) is an example — and their spectral domain representation (the frequency response function) play a fundamental role in time series analysis. We need a corresponding theory for functional random fields. We develop its main elements in the remainder of this section. We define the spatial lag operator by $B^{\mathbf{k}}X_{\mathbf{s}}:=X_{\mathbf{s}-\mathbf{k}},\mathbf{k}\in\mathbb{Z}^r$. The *spatial functional filter* (SFF) is a sequence $\mathbf{\Psi}=(\Psi_{\mathbf{l}}:\mathbf{l}\in\mathbb{Z}^r)$ of linear operators mapping *H* into another separable Hilbert space H'. The filtered functional random field is defined by

$$Y_{\mathbf{s}}=\Psi(B)X_{\mathbf{s}}=\sum_{\mathbf{k}\in\mathbb{Z}^r}\Psi_{\mathbf{k}}(X_{\mathbf{s}-\mathbf{k}})=\lim_{n\to\infty}\sum_{\|\mathbf{k}\|_\infty\leq n}\Psi_{\mathbf{k}}(X_{\mathbf{s}-\mathbf{k}}),\quad(2.6)$$

provided the limit exists in a suitable sense. The frequency response function of the SFF is defined by

$$\Psi_\theta=\Psi(e^{-\mathrm{i}\theta})=\sum_{\mathbf{k}\in\mathbb{Z}^r}\Psi_{\mathbf{k}}e^{-\mathrm{i}\mathbf{k}^\top\theta}.\quad(2.7)$$

A sufficient condition for (2.7) to converge in operator norm is

$$\sum_{\mathbf{k}\in\mathbb{Z}^r}\|\Psi_{\mathbf{k}}\|<\infty,\quad(2.8)$$

which also implies continuity of Ψ_θ with respect to θ. However, we generally do not want to rely on this condition. In fact, pointwise convergence of (2.7) is not needed for our framework.

In the development of spatial functional principal components, we will need convergence results stated in the following two propositions. We denote by $\mathcal{S}(H, H')$ the space of Hilbert–Schmidt operators from H into H'.

Proposition 2.5. *Let* $\mathbf{\Psi} = (\Psi_{\mathbf{k}} : \mathbf{k} \in \mathbb{Z}^r)$ *be a spatial functional filter. Suppose that each* Ψ_k *is a Hilbert–Schmidt operator and that*

$$\int_{[-\pi,\pi]^r} \left\| \sum_{\|\mathbf{k}\|_\infty \le n} \Psi_{\mathbf{k}} e^{-i\mathbf{k}^\top \boldsymbol{\theta}} - \Psi_{\boldsymbol{\theta}} \right\|^2_{\mathcal{S}(H,H')} d\boldsymbol{\theta} \to 0 \qquad (2.9)$$

and

$$\operatorname*{ess\,sup}_{\boldsymbol{\theta} \in [-\pi,\pi]^r} \left\| \Psi_{\boldsymbol{\theta}} \right\|_{\mathcal{S}_2} < \infty. \qquad (2.10)$$

Let $(Y_{\mathbf{s}})$ be defined by (2.6). If $(X_{\mathbf{s}})_{\mathbf{s} \in \mathbb{Z}^r}$ satisfies Assumption 2.1 and

$$\operatorname*{ess\,sup}_{\boldsymbol{\theta} \in [-\pi,\pi]^r} \left\| \mathcal{F}^X_{\boldsymbol{\theta}} \right\|_{\mathcal{S}_2} < \infty, \qquad (2.11)$$

then the following statements hold.

(a) The series (2.6) defining $Y_{\mathbf{s}}$ *converges in* $L^2_{H'}$.

(b) $(Y_{\mathbf{s}})$ *is weakly stationary with autocovariance operator*

$$C^Y_{\mathbf{h}} = \sum_{\mathbf{k} \in \mathbb{Z}^r} \sum_{\mathbf{l} \in \mathbb{Z}^r} \Psi_{\mathbf{k}} C^X_{\mathbf{l}-\mathbf{k}+\mathbf{h}} \Psi^*_{\mathbf{l}}$$

and spectral density operator

$$\mathcal{F}^Y_{\boldsymbol{\theta}} = \Psi_{\boldsymbol{\theta}} \mathcal{F}^X_{\boldsymbol{\theta}} \Psi^*_{\boldsymbol{\theta}},$$

where equality holds for almost all $\boldsymbol{\theta} \in [-\pi, \pi]^r$.

(c) $\operatorname*{ess\,sup}_{\boldsymbol{\theta} \in [-\pi,\pi]^r} \left\| \mathcal{F}^Y_{\boldsymbol{\theta}} \right\|_{\mathcal{S}_1} < \infty$.

Proposition 2.5 is formulated under fairly general assumptions, in particular we don't require (2.8). This has a good reason: in the framework of SFPCs, we will encounter filter functions (3.1) which are square-summable but not necessarily absolutely summable. A simpler and stronger result can be derived assuming (2.8).

Proposition 2.6. *If* (2.8) *holds and* $(X_{\mathbf{s}})_{\mathbf{s}\in\mathbb{Z}^r}$ *satisfies Assumption 2.1, then the statements (a) and (b) of Proposition 2.5 still hold (without assuming* (2.9)*,* (2.10), (2.11)*). If furthermore the autocovariances of* $(X_{\mathbf{s}})$ *are summable in the sense of* (A.1)*, then* (A.1) *also holds for the filtered random field* $(Y_{\mathbf{s}})$*, and* $\mathcal{F}_{\theta}^{Y} = \Psi_{\theta}\mathcal{F}_{\theta}^{X}\Psi_{\theta}^{*}$ *holds pointwise for all* $\boldsymbol{\theta}$*.*

We will specifically consider the case where $H' = \mathbb{R}^p$, for some $p \in \mathbb{N}$. By the Riesz representation theorem, there are $\psi_{m,\mathbf{l}} \in H, 1 \le m \le p, \mathbf{l} \in \mathbb{Z}^r$, such that each $\Psi_{\mathbf{l}}$ takes the form $\Psi_{\mathbf{l}}(x) = (\langle x, \psi_{1,\mathbf{l}}\rangle, \ldots, \langle x, \psi_{p,\mathbf{l}}\rangle)^{\top}$.

Corollary 2.1. *Assume that* $\Psi_{\mathbf{l}}(x) = (\langle x, \psi_{1,\mathbf{l}}\rangle, \ldots, \langle x, \psi_{p,\mathbf{l}}\rangle)^{\top}$ *with* $\psi_{m,\mathbf{l}} \in H$ *for all* $1 \le m \le p, \mathbf{l} \in \mathbb{Z}^r$*. If the filter* Ψ *and the functional random field* $(X_{\mathbf{s}})_{\mathbf{s}\in\mathbb{Z}^r}$ *fulfil the conditions of Proposition 2.5, then (2.6) converges in mean square to a limit* $Y_{\mathbf{s}}$*. This vector process* $(Y_{\mathbf{s}})$ *is stationary and has the spectral density matrix*

$$\mathcal{F}_{\theta}^{Y} = \begin{pmatrix} \langle \mathcal{F}_{\theta}^{X}(\psi_{1,\theta}^{\star}), \psi_{1,\theta}^{\star}\rangle & \cdots & \langle \mathcal{F}_{\theta}^{X}(\psi_{p,\theta}^{\star}), \psi_{1,\theta}^{\star}\rangle \\ \vdots & \ddots & \vdots \\ \langle \mathcal{F}_{\theta}^{X}(\psi_{1,\theta}^{\star}), \psi_{p,\theta}^{\star}\rangle & \cdots & \langle \mathcal{F}_{\theta}^{X}(\psi_{p,\theta}^{\star}), \psi_{p,\theta}^{\star}\rangle \end{pmatrix},$$

where $\psi_{m,\boldsymbol{\theta}}^{\star} := \sum_{\mathbf{l}\in\mathbb{Z}^r} \psi_{m,\mathbf{l}} e^{\mathrm{i}\mathbf{l}^{\top}\boldsymbol{\theta}}$ *and equality holds for almost all* $\boldsymbol{\theta} \in [-\pi, \pi]^r$*.*

2.4 Spatial functional ARMA processes

We will, in this section, lay the groundwork for spatial functional ARMA processes and their simulation. Results on stationarity and prediction of functional ARMA time series can be found in Klepsch et al. (2017). Ruiz-Medina (2012) introduced spatial functional AR processes and establishes

conditions for stationarity in certain settings. Spatial functional ARMA processes are a subject of independent interest and illustrate some of the theory developed in the previous sections. These models will be used as data generating processes in Section 4.1.

Definition 2.4. (SFARMA(*P*, *Q*)) Let $P,Q \subset \mathbb{Z}^r$ be two finite index sets with $\mathbf{0} \notin P$ and $(A_\mathbf{k})_{\mathbf{k}\in P}, (B_\mathbf{k})_{\mathbf{k}\in Q}$ bounded linear operators mapping from *H* to *H*. Let $(\varepsilon_\mathbf{s})_{\mathbf{s}\in\mathbb{Z}^r}$ be a functional random field of white noise (i.e. weakly stationary and uncorrelated at all spatial lags) with mean zero. Let $(X_\mathbf{s})_{\mathbf{s}\in\mathbb{Z}^r}$ be a functional random field satisfying the recursion

$$X_\mathbf{s} := \sum_{\mathbf{k}\in P} A_\mathbf{k}\ X_{\mathbf{s}-\mathbf{k}} + \sum_{\mathbf{k}\in Q} B_\mathbf{k}\ \varepsilon_{\mathbf{s}-\mathbf{k}}. \qquad (2.12)$$

If $(X_\mathbf{s})_{\mathbf{s}\in\mathbb{Z}^r}$ is weakly stationary, we call it a *spatial functional ARMA(P, Q) process*.

Remark 2.2. Because there are no restrictions on the autoregressive lags *P*, Definition 2.4 defines a multi–sided recursion, which is not as easy to handle as a univariate causal recursion. Therefore, some authors consider the restriction of *P* to the non-symmetric half-space

$$S_r := \bigcup_{i=0}^{r-1} \left(\mathbb{Z}^i \times \mathbb{N} \times \{0\}^{r-i-1} \right)$$

and call such a model *unilateral*. In practice, this can arise naturally from directed processes, where constant sea currents, winds or sloped terrain play a predominant role and establish a spatial order. But in the absence of such a directional influence, a unilateral definition can also be a choice of convenience, as it ensures *causality* of $(X_\mathbf{s})$, which simplifies most, if not all, calculations, see Ruiz-Medina (2012) and Drapatz (2016). For *r* = 1, unilaterality means that $P \subset \{1, 2, \ldots\}$. This yields the common time series formulation of ARMA processes. Whittle (1954) showed that in the case of *r* = 2 with scalar observations, the unilateral formulation is not very restrictive compared to the two-sided definition. However, finding the unilateral

representation of a two-sided ARMA process can be cumbersome and yield results that are not very elegant.

Proposition 2.7 establishes sufficient conditions for the existence of a stationary solution to the SFARMA(*P*, *Q*) equations (2.12). Theorem 3.1 in Bosq (2000) on the stationarity of a FAR(1) model is a special case of Proposition 2.7 (*r* = 1, $P = \{1\}, Q = \{1\}, B_1 = I$). Our result is also more general because we allow the innovations $Z_{\mathbf{s}}$ to form a weakly stationary process rather than using independent errors $\varepsilon_{\mathbf{s}}$, as in Definition 2.3.

Proposition 2.7. *Let* $(Z_{\mathbf{s}})_{\mathbf{s}\in\mathbb{Z}^r}$ *be a centered, weakly stationary functional random field with* $\mathbb{E}\|Z_{\mathbf{s}}\|^2 < \infty$, $P \subset \mathbb{Z}^r \setminus \{\mathbf{0}\}$ *a finite index set and* $(A_{\mathbf{k}})_{\mathbf{k}\in P}$ *bounded linear operators mapping from H to H. Define* $X_{\mathbf{s}}$ *by the recursion*

$$X_{\mathbf{s}} := \sum_{\mathbf{k}\in P} A_{\mathbf{k}}\, X_{\mathbf{s}-\mathbf{k}} + Z_{\mathbf{s}}. \qquad (2.13)$$

Suppose that

$$\exists\, N \in \mathbb{N}: \quad \sum_{\mathbf{k}_1,\ldots,\mathbf{k}_N \in P} \left\| A_{\mathbf{k}_1} \cdots A_{\mathbf{k}_N} \right\|_{\mathcal{S}_\infty} < 1. \qquad (2.14)$$

Then there exists a unique weakly stationary solution to (2.13) *given by the functional moving average representation*

$$X_{\mathbf{s}} = \sum_{l=1}^{\infty} \sum_{\mathbf{k}_1,\ldots,\mathbf{k}_l \in P} A_{\mathbf{k}_l} \cdots A_{\mathbf{k}_1} Z_{\mathbf{s}-\sum \mathbf{k}_j} + Z_{\mathbf{s}}. \qquad (2.15)$$

This series converges in mean square and almost surely. Furthermore, if $(Z_{\mathbf{s}})_{\mathbf{s}\in\mathbb{Z}^r}$ *is L^p-m-approximable, then so is* $X_{\mathbf{s}}$.

Corollary 2.2. *Let* $(X_{\mathbf{s}})_{\mathbf{s}\in\mathbb{Z}^r}$ *be a SFARMA(P, Q) process as defined in Definition 2.4 and satisfying condition (2.14). Then* $(X_{\mathbf{s}})_{\mathbf{s}\in\mathbb{Z}^r}$ *is weakly stationary and uniquely defined. If furthermore* $(\varepsilon_{\mathbf{s}})_{\mathbf{s}\in\mathbb{Z}^r}$ *is an i.i.d. functional*

random field satisfying $\mathbb{E}\|\varepsilon_{\mathbf{s}}\|^p < \infty$ *for some* $p \geq 2$*, then* $(X_{\mathbf{s}})_{\mathbf{s}\in\mathbb{Z}^r}$ *is L^p-m-approximable.*

The simulation of a general SFARMA process can be done by reapplying m times the recursion formula (2.13) on the transformed noise $Z_{\mathbf{s}} = \sum_{\mathbf{k}\in Q} B_{\mathbf{k}}\varepsilon_{\mathbf{s}-\mathbf{k}}$, which after m steps effectively yields a truncated version of (2.15). Convergence is typically fast, depending on the structure of the autoregressive lags and operators, and a bound on the approximation error can be seen directly from the proof of Proposition 2.7. For unilateral formulations of the autoregressive part, simulation is straight-forward, and one does not have to resort to this approximation.

3 Spatial functional principal components

In this section, we present the theory underlying the spectral decomposition of functional data on a spatial domain. Fundamental definitions and properties are presented in Section 3.1. This is followed by the estimation theory in Sections 3.2 and 3.3. We establish consistency rates and display their dependence on the spatial dimension r and the strength of spatial dependence. The proofs of all results presented in this section are given in Section II of Supplementary Materials.

3.1 Existence and fundamental properties

Recall Definition 2.2 and Proposition 2.1. Denote by $\varphi_m(u\,|\,\boldsymbol{\theta})$ the value of the eigenfunction $\varphi_m(\boldsymbol{\theta})$ at $u \in [0,1]$. Because $\|\varphi_m(\boldsymbol{\theta})\| = 1$, by Fubini's theorem,

$$(2\pi)^r = \int_{[-\pi,\pi]^r}\int_0^1 |\varphi_m(u\,|\,\boldsymbol{\theta})|^2\, du\, d\boldsymbol{\theta} = \int_0^1\int_{[-\pi,\pi]^r} |\varphi_m(u\,|\,\boldsymbol{\theta})|^2\, d\boldsymbol{\theta}\, du.$$

We see that for almost all $u \in [0,1]$, $\varphi_m(u\,|\,\boldsymbol{\theta}) \in L^2([-\pi,\pi]^r) \subseteq L^1([-\pi,\pi]^r)$. This means that for almost all $u \in [0,1]$, we can define the Fourier coefficients

$$\phi_{m,\mathbf{l}}(u) := \frac{1}{(2\pi)^r}\int_{[-\pi,\pi]^r} \varphi_m(u\,|\,\boldsymbol{\theta}) e^{-\mathrm{i}\mathbf{l}^\top\boldsymbol{\theta}}\, d\boldsymbol{\theta}, \qquad (3.1)$$

and the expansion

$$\varphi_m(u \mid \boldsymbol{\theta}) = \sum_{\mathbf{l} \in \mathbb{Z}^r} \phi_{m,\mathbf{l}}(u)\, e^{\mathrm{i}\mathbf{l}^\top \boldsymbol{\theta}}.$$

The following definition must be considered in conjunction with Proposition 3.1, which ensures the existence of the scores $Y_{m,\mathbf{s}}$ below.

Definition 3.1. Suppose $(X_\mathbf{s})_{\mathbf{s} \in \mathbb{Z}^r}$ satisfies Assumption 2.1 and (A.1) holds. Then the *m*-th *spatial functional principal component* (SFPC) score is defined by

$$Y_{m,\mathbf{s}} := \sum_{\mathbf{l} \in \mathbb{Z}^r} \left\langle X_{\mathbf{s}-\mathbf{l}}, \phi_{m,\mathbf{l}} \right\rangle, \quad (3.2)$$

where $\phi_{m,\mathbf{l}}$ is defined by (3.1). The corresponding SFPC filter is denoted $\boldsymbol{\Phi}_m := (\phi_{m,\mathbf{l}})_{\mathbf{l} \in \mathbb{Z}^r}$.

If $r = 1$, the SFPC scores in Definition 3.1 coincide with the dynamic FPCs introduced in Hörmann et al. (2015).

Proposition 3.1. *In the setting of Definition 3.1, i.e. under Assumption 2.1 and* (A.1):

(a) The series defining $Y_{m,\mathbf{s}}$ *in* (3.2) *converges in mean square, with*

$$\mathbb{E} Y_{m,\mathbf{s}} = 0, \qquad \mathbb{E} Y_{m,\mathbf{s}}^2 = \sum_{\mathbf{l} \in \mathbb{Z}^r} \sum_{\mathbf{k} \in \mathbb{Z}^r} \left\langle C^X_{\mathbf{l}-\mathbf{k}}(\phi_{m,\mathbf{l}}), \phi_{m,\mathbf{k}} \right\rangle.$$

(b) If $(X_\mathbf{s})$ *is real-valued, then both the Fourier coefficients* $\phi_{m,\mathbf{l}}(u)$ *in* (3.1) *and the* $Y_{m,\mathbf{s}}$ *are real.*

(c) If $C^X_\mathbf{h} = 0$ *for all* $\mathbf{h} \neq \mathbf{0}$*, then* $Y_{m,\mathbf{s}}$ *coincide with ordinary FPC scores computed from* $C^X_\mathbf{0}$*.*

Moreover, if there exists an index set $\{i_1, \ldots, i_k\} \subset \{1, \ldots, r\}$ *with* $C^X_\mathbf{h} = 0$ *if* $h_{i_j} \neq 0$ *for* $1 \leq j \leq k$*, then the r-dimensional SFPC scores coincide with*

the $(r-k)$ *-dimensional SFPC scores that are obtained by considering only the* $(r-k)$ *-dimensional random field* $(X_{\mathbf{s}} : s_{i_j} = 0 \; \forall \, 1 \le j \le k)$.

(d) For $m \neq m'$ *and any* $\mathbf{s}, \mathbf{s}' \in \mathbb{Z}^r$, *the SFPC scores* $Y_{m,\mathbf{s}}$ *and* $Y_{m',\mathbf{s}'}$ *are uncorrelated. In general, the covariance between score vectors can be calculated as in Proposition 2.5(b).*

(e) It holds that

$$\lim_{\mathbf{n}\to\infty} \frac{1}{|R_{\mathbf{n}}|} \mathrm{Var}\Big(\sum_{\mathbf{k}\in R_{\mathbf{n}}} Y_{m,\mathbf{k}}\Big) = (2\pi)^r \lambda_m(\mathbf{0}), \qquad (3.3)$$

where the limit $\mathbf{n}\to\infty$ *means that all components of* $\mathbf{n}$ *tend to infinity, and* $R_{\mathbf{n}}$ *is the hyperrectangle of size* $\mathbf{n} \in \mathbb{Z}^r$, *that is* $R_{\mathbf{n}} = \{\mathbf{s} \in \mathbb{Z}^r : 1 \le s_i \le n_i \; \forall 1 \le i \le r\}$.

Theorem 3.1 below is one of the main results of this paper. Expansion (3.4) is a decomposition of the spatial functional field into uncorrelated component fields of decreasing importance and smoothness.

Theorem 3.1 (SPATIAL KARHUNEN–LOÈVE EXPANSION). *Assume* $(X_{\mathbf{s}})_{\mathbf{s}\in\mathbb{Z}^r}$ *is a functional random field satisfying Assumption 2.1 and* (A.1). *Let* $\phi_{m,\mathbf{s}}$ *and* $Y_{m,\mathbf{s}}$ *be defined by* (3.1) *and* (3.2). *Then*

$$X_{\mathbf{s}}(u) = \sum_{m\ge 1} X_{m,\mathbf{s}}(u), \quad \text{with} \quad X_{m,\mathbf{s}}(u) := \sum_{\mathbf{l}\in\mathbb{Z}^r} Y_{m,\mathbf{s}+\mathbf{l}}\, \phi_{m,\mathbf{l}}(u), \qquad (3.4)$$

where both series converge in mean square. The curves $X_{m,\mathbf{s}}$ *are uniquely determined and uncorrelated across m, i.e.*

$$\mathrm{Cov}(X_{m,\mathbf{s}}(u), X_{n,\mathbf{s}'}(u')) = 0, \text{ if } m \neq n, \; \forall \, \mathbf{s}, \mathbf{s}' \in \mathbb{Z}^r, u, u' \in [0,1]. \qquad (3.5)$$

The next theorem states that for any fixed *p*, the approximation of each function $X_{\mathbf{s}}$ by $\sum_{m=1}^{p} X_{m,\mathbf{s}}$ is optimal in the following sense. Suppose the scores $Y_{m,s}$ in (3.2) are computed using a different set of functions, say $\psi_{m,\mathbf{l}}$, rather

than the functions $\phi_{m,\mathbf{l}}$ defined by (3.1). This gives us a new set of "scores" $\tilde{Y}_{m,\mathbf{s}} = \sum_{\mathbf{l}\in\mathbb{Z}^r} \langle X_{\mathbf{s}-\mathbf{l}}, \psi_{m,\mathbf{l}} \rangle$. Next, instead of computing the functions $X_{m,\mathbf{s}}$ as in the second equation in (3.4), compute them using a different set of functions, say $\nu_{m,\mathbf{l}}$, instead of the $\phi_{m,\mathbf{l}}$, and the $\tilde{Y}_{m,\mathbf{s}}$ instead of the $Y_{m,s}$ in (3.2). Denote these functions by $\tilde{X}_{m,\mathbf{s}}$. Theorem 3.2 states that the $\phi_{m,\mathbf{l}}$ and the $\nu_{m,\mathbf{l}}$ cannot be chosen in such a way that the approximation by $\sum_{m=1}^{p} \tilde{X}_{m,\mathbf{s}}$ is better than the approximation by $\sum_{m=1}^{p} X_{m,\mathbf{s}}$.

Theorem 3.2 (OPTIMALITY). *For arbitrary SFFs* $(\psi_{m,\mathbf{k}} : \mathbf{k} \in \mathbb{Z}^r)$ *and* $(\nu_{m,\mathbf{k}} : \mathbf{k} \in \mathbb{Z}^r)$ *in H that fulfill the conditions of Proposition 2.5 set*

$$\tilde{X}_{m,\mathbf{s}}(u) = \sum_{\mathbf{l}\in\mathbb{Z}^r} \tilde{Y}_{m,\mathbf{s}+\mathbf{l}} \nu_{m,\mathbf{l}}(u), \quad \tilde{Y}_{m,\mathbf{s}} = \sum_{\mathbf{l}\in\mathbb{Z}^r} \langle X_{\mathbf{s}-\mathbf{l}}, \psi_{m,\mathbf{l}} \rangle. \qquad (3.6)$$

Then, for all $p \geq 1$,

$$\mathbb{E}\left\| X_{\mathbf{s}} - \sum_{m=1}^{p} X_{m,\mathbf{s}} \right\|^2 = \sum_{m>p} \int_{[-\pi,\pi]^r} \lambda_m(\boldsymbol{\theta})\, d\boldsymbol{\theta} \leq \mathbb{E}\left\| X_{\mathbf{s}} - \sum_{m=1}^{p} \tilde{X}_{m,\mathbf{s}} \right\|^2. \quad (3.7)$$

We now explain the main advantages of expansion (3.4) over the ordinary Karhunen–Loève expansion (1.1). For this purpose, it is convenient to rewrite formula (1.1) as

$$X_{\mathbf{s}}(u) = \sum_{m\geq 1} \tilde{X}_{m,\mathbf{s}}(u), \quad \tilde{X}_{m,\mathbf{s}}(u) = \xi_{m,\mathbf{s}} \nu_m(u). \qquad (3.8)$$

For each $\mathbf{s}$, the scores $\xi_{m,\mathbf{s}}$ are uncorrelated, i.e. $\mathrm{Cov}(\xi_{m,\mathbf{s}}, \xi_{m',\mathbf{s}}) = 0$ if $m' \neq m$. It follows that for any u, u', $\mathrm{Cov}(\tilde{X}_{m,\mathbf{s}}(u), \tilde{X}_{m',\mathbf{s}}(u')) = 0$. However, if $\mathbf{s}' \neq \mathbf{s}$, in general, $\mathrm{Cov}(\xi_{m,\mathbf{s}}, \xi_{m',\mathbf{s}'}) \neq 0$. Property (3.5) thus does not hold for the component fields $\tilde{X}_{m,\mathbf{s}}(u)$ based on the usual Karhunen–Loève expansion; these fields are not uncorrelated. The zero correlation of the whole fields is a very useful property, as will be discussed in Section 5. Relation (3.7) shows

that the first p functions $X_{m,\mathbf{s}}(u)$ capture more variance than any other decomposition, including the standard FPCs. They thus provide the most efficient tool for dimension reduction. Observe that decomposition (3.8) coincides with (3.6) if $\nu_{m,\mathbf{l}} = \psi_{m,\mathbf{l}} = 0$ for all $\mathbf{l} \neq \mathbf{0}$ and $\nu_{m,\mathbf{0}} = \psi_{m,\mathbf{0}} = \nu_m$.

We will call decomposition (3.4) the *spatial Karhunen–Loève expansion* of the field $(X_{\mathbf{s}})$. We next consider estimation of its elements.

3.2 Estimation of the spectral density operator

We assume that the functions $X_{\mathbf{s}}$ are observed at $\mathbf{s} = (s_1, s_2, \ldots, s_r), 1 \le s_i \le n_i$. The total number of the spatial locations on the grid is thus $N = \prod_{i=1}^{r} n_i$. We use the notation $\mathbf{n} = (n_1, n_2, \ldots, n_r)$.

Based on Definition 2.2, we estimate the operator $\mathcal{F}_\theta^X$ by

$$\mathcal{F}_\theta^X := \frac{1}{(2\pi)^r} \sum_{|\mathbf{h}| \le \mathbf{q}} w(\mathbf{h}/\mathbf{q})\, \hat{C}_{\mathbf{h}}\, e^{-\mathrm{i}\mathbf{h}^\top \theta}, \qquad (3.9)$$

where w is a weight function and the vector $\mathbf{q} = (q_1, q_2, \ldots, q_r)$ has positive coordinates. The sample autocovariance operators are defined by

$$\hat{C}_{\mathbf{h}} := \frac{1}{N} \sum_{\mathbf{s} \in M_{\mathbf{h},\mathbf{n}}} (X_{\mathbf{s}+\mathbf{h}} - \bar{X}) \otimes (X_{\mathbf{s}} - \bar{X}), \quad (3.10)$$

with $M_{\mathbf{h},\mathbf{n}} = \{\mathbf{s} \in \mathbb{Z}^r : 1 \le s_i, s_i + h_i \le n_i \ \forall 1 \le i \le r\}$. If the set $M_{\mathbf{h},\mathbf{n}}$ is empty, we set $\hat{C}_{\mathbf{h}} = 0$.

If r = 1, the weight function is often taken to be the Bartlett kernel $w(z) = (1 - |z|)_+$. A straightforward way of generalizing this kernel into higher dimensions is using some norm $\|\cdot\|$ on $\mathbb{Z}^r$ and setting

$$w(\mathbf{z}) = (1 - \|\mathbf{z}\|)_+ . \qquad (3.11)$$

Instead of using (3.11), one can use many weight functions of the form $w(\mathbf{z}) = w_1(\|\mathbf{z}\|)$, where w_1 is some univariate kernel. Since the study of the impact of the form of kernel is not central to the methodology we develop, we focus on the kernel (3.11).

The following proposition establishes the consistency of the estimators $\mathcal{F}_\theta^X$ in a sense that is suitable for the subsequent applications.

Proposition 3.2. *Assume that* $(X_\mathbf{s})_{\mathbf{s}\in\mathbb{Z}^r}$ *is L⁴-m-approximable and consider the estimator* $\mathcal{F}_\theta^X$ *defined by (3.9) with weights (3.11). Assume that for all dimensions* $1 \le i \le r$*, as* $N \to \infty$*,*

$$n_i \to \infty, \quad q_i \to \infty, \quad q_i = \mathcal{O}(n_i^\delta) \ \text{ for some } 0 < \delta < \frac{1}{2}.$$

Then $\mathcal{F}_\theta^X$ *is consistent in integrated mean square. i.e.*

$$\int_{[-\pi,\pi]^r} \mathbb{E}\left\| \mathcal{F}_\theta^X - \mathcal{F}_\theta^X \right\|_{\mathcal{S}_2}^2 d\boldsymbol{\theta} \xrightarrow{N\to\infty} 0. \quad (3.12)$$

We emphasize that in the spatial case, care must be taken to ensure the asymptotics for n_i, q_i (for all $1 \le i \le r$) and the shape of the multivariate kernel interact properly. As noted above, and as is apparent from the proof of Proposition 3.2, a variety of other kernel functions will yield the same asymptotic result, provided the resulting weights w satisfy suitable conditions and the n_i and q_i are suitably chosen. We also explain in the proof of Proposition 3.2 that the assumption that all dimensions n_i tend to infinity is needed to ensure that the estimator $\hat{C}_\mathbf{h}$ is asymptotically unbiased. Any norm on $\mathbb{Z}^r$ can be used to define the weights $w(\cdot)$ in (3.11), without influence on the asymptotic result.

3.3 Sample filters and scores

To compute a sample version of the spatial Karhunen–Loève expansion given by (3.4) in Theorem 3.1, we must be able to compute sample versions of the filter functions $\phi_{m,\mathbf{l}}$ and the scores $Y_{m,\mathbf{s}}$. In light of (3.1), we set

$$\hat{\phi}_{m,\mathbf{l}}(u) := \frac{1}{(2\pi)^r} \int_{[-\pi,\pi]^r} \hat{\varphi}_m(u \mid \boldsymbol{\theta}) e^{-\mathrm{i}\mathbf{l}^\top \boldsymbol{\theta}} d\boldsymbol{\theta}, \qquad (3.13)$$

where the functions $\hat{\varphi}_m(\boldsymbol{\theta})$ are the eigenfunctions of an estimator $\mathcal{F}_{\boldsymbol{\theta}}^X$ of the spectral density operator $\mathcal{F}_{\boldsymbol{\theta}}^X$. The population scores $Y_{m,\mathbf{s}}$ are then approximated by

$$\hat{Y}_{m,\mathbf{s}} := \sum_{\|\mathbf{l}\|_\infty \le L} \left\langle X_{\mathbf{s}-\mathbf{l}}, \hat{\phi}_{m,\mathbf{l}} \right\rangle, \quad 1 \le m \le p, \qquad (3.14)$$

assuming that $1 + L \le s_i \le n_i - L$ for all $1 \le i \le r$, where *L* is an integer-valued truncation parameter. The sample analog of (3.4) is then

$$X_{\mathbf{s}}(u) \approx \sum_{m=1}^{p} \hat{X}_{m,\mathbf{s}}(u), \quad \text{with} \quad \hat{X}_{m,\mathbf{s}}(u) := \sum_{\|\mathbf{l}\|_\infty \le L} \hat{Y}_{m,\mathbf{s}+\mathbf{l}} \hat{\phi}_{m,\mathbf{l}}(u), \qquad (3.15)$$

assuming $1 + 2L \le s_i \le n_i - 2L$ for $1 \le i \le r$.

We now establish consistency results for the estimators in (3.15). We begin by stating the required assumptions.

Assumption 3.1. *The estimator $\mathcal{F}_{\boldsymbol{\theta}}^X$ for the spectral density operator is consistent in integrated mean operator norm, i.e.* $\int_{[-\pi,\pi]^r} \mathbb{E} \left\| \mathcal{F}_{\boldsymbol{\theta}}^X - \mathcal{F}_{\boldsymbol{\theta}}^X \right\|_{\mathcal{S}_\infty} d\boldsymbol{\theta} \to 0$.

Assumption 3.1 is implied by (3.12) in Proposition 3.2 because

$$\left(\mathbb{E} \int \left\| \mathcal{F}_{\boldsymbol{\theta}}^X - \mathcal{F}_{\boldsymbol{\theta}}^X \right\|_{\mathcal{S}_\infty} d\boldsymbol{\theta} \right)^2 \le \left(\mathbb{E} \int \left\| \mathcal{F}_{\boldsymbol{\theta}}^X - \mathcal{F}_{\boldsymbol{\theta}}^X \right\|_{\mathcal{S}_2} d\boldsymbol{\theta} \right)^2 \le C_r \, \mathbb{E} \int \left\| \mathcal{F}_{\boldsymbol{\theta}}^X - \mathcal{F}_{\boldsymbol{\theta}}^X \right\|_{\mathcal{S}_2}^2 d\boldsymbol{\theta}.$$

Proposition 3.2 shows that Assumption 3.1 holds under general, weak conditions on the spatial domain and spatial dependence, but in what follows we do not need these conditions, only Assumption 3.1 is required.

We next define

$$\alpha_1(\boldsymbol{\theta}) = \lambda_1(\boldsymbol{\theta}) - \lambda_2(\boldsymbol{\theta}),$$
$$\alpha_m(\boldsymbol{\theta}) = \min(\lambda_{m-1}(\boldsymbol{\theta}) - \lambda_m(\boldsymbol{\theta}), \lambda_m(\boldsymbol{\theta}) - \lambda_{m+1}(\boldsymbol{\theta})),$$

where $\lambda_l(\boldsymbol{\theta})$ is the l-th largest eigenvalue of $\mathcal{F}_\theta^X$. The functions α_m are, by definition, nonnegative. Under the assumptions of Proposition 2.1, all α_m are continuous functions.

Assumption 3.2. *For all* $1 \le m \le p$, $\inf_{\theta \in [-\pi,\pi]^r} \alpha_m(\boldsymbol{\theta}) =: \beta_m > 0.$

Assumption 3.2 is analogous to requiring that the eigenvalues λ_m of the standard functional principal components v_m are separated, which means that these components can be uniquely identified, up to the sign of v_m.

The signs of $\hat{v}_m$ and v_m are arbitrary, and these functions are close if they have the same orientation, i.e. if $\langle v_m, \hat{v}_m \rangle > 0.$ In the context of this paper, the eigenfunction $\varphi_m(\boldsymbol{\theta})$ are uniquely determined only up to multiplication by a number on the complex unit circle. So henceforth we tacitly only consider pairs $\hat{\varphi}_m(\boldsymbol{\theta})$ and $\varphi_m(\boldsymbol{\theta})$ for which $\langle \varphi_m(\boldsymbol{\theta}), \hat{\varphi}_m(\boldsymbol{\theta}) \rangle > 0, \boldsymbol{\theta} \in [-\pi, \pi]^r$. With this convention, we now state the consistency results for the filters, the scores and the SFPC decomposition.

Theorem 3.3. *Let* $p \ge 1$ *be fixed. Under Assumptions 3.1, and 3.2, for* $1 \le m \le p$,

$$\max_{\mathbf{l} \in \mathbb{Z}^r} \left\| \phi_{m,\mathbf{l}} - \hat{\phi}_{m,\mathbf{l}} \right\| = \mathcal{O}(G(N)), \quad (3.16)$$

where

$$G(N) := \int_{[-\pi,\pi]^r} \left\| \mathcal{F}_{\theta}^{X} - \mathcal{F}_{\theta}^{X} \right\|_{\mathcal{S}_\infty} d\boldsymbol{\theta} \xrightarrow{P} 0.$$

In particular, it follows that $\max_{\mathbf{l}\in\mathbb{Z}^r} \| \phi_{m,\mathbf{l}} - \hat{\phi}_{m,\mathbf{l}} \| \xrightarrow{P} 0$ *, as* $N \to \infty$ *.*

Theorem 3.4. *Suppose that Assumptions 3.1 and 3.2 hold. Then, for* $1 \le m \le p$ *,*

$$\left| \hat{Y}_{m,\mathbf{s}} - Y_{m,\mathbf{s}} \right| = \mathcal{O}_P(L^r G(N) + H_m(L)),$$

where G(N) is defined in Theorem 3.3, and

$$H_m(L) := \Big(\sum_{\|\mathbf{l}\|_\infty > L} \left\| \phi_{m,\mathbf{l}} \right\|^2 \Big)^{1/4}.$$

In particular, if $L^r G(N) \xrightarrow{P} 0$ *, then* $| \hat{Y}_{m,\mathbf{s}} - Y_{m,\mathbf{s}} | \xrightarrow{P} 0$ *.*

Theorem 3.5. *Let* $1 \le m \le p$ *and suppose that Assumptions 3.1 and 3.2 hold. If* $L^r H_m(L) \to 0$ *, and for an increasing sequence* $L = L(N)$ *it holds that* $L^{2r} \mathbb{E}[G(N)] \to 0$ *, then for each* ***s****,* $\mathbb{E} \left\| \hat{X}_{m,\mathbf{s}} - X_{m,\mathbf{s}} \right\| \to 0$ *, as* $N \to \infty$ *.*

The condition $L^r H_m(L) \to 0$ is basically a smoothness assumption on $\varphi_m(\boldsymbol{\theta})$ and ultimately on $\mathcal{F}_{\theta}^{X}$. What we mean here by smoothness is that the Fourier coefficients $\phi_{m,\mathbf{l}}$ decay in a way such that the summation tail $\sum_{\|\mathbf{l}\|_\infty > L} \left\| \phi_{m,\mathbf{l}} \right\|^2$ converges to zero at a suitably fast rate. For example, if *r* = 2, then a sufficient condition for this is $\left\| \phi_{m,\mathbf{l}} \right\| = \mathcal{O}(\|\mathbf{l}\|^{-\gamma})$ with $\gamma > 5$.

The criterion $L^{2r} \mathbb{E}[G(N)] \to 0$ is used for the choice of *L*. Suppose the process $(X_{\mathbf{s}})$ is *L*4-*m*-approximable with exponentially decaying autocovariances $C_{\mathbf{h}}^{X}$, and the sample size **n** grows with the same order in all dimensions. Then it follows from the details in the proof of Proposition 3.2 that (3.12) is of asymptotic order $N^{-1/2} \prod_{i=1}^{r} q_i + \sum_{i=1}^{r} q_i^{-1}$. If we set $q_i = \lfloor c_q n_i^{\delta} \rfloor$ for some $c_q > 0$ and

a fixed $0 < \delta < \frac{1}{2}$, then we have a polynomial convergence rate for $\mathbb{E}[G(N)] \to 0$. Therefore we can choose a constant $\rho > 0$ only depending on δ, such that $L(N) = N^{\rho}$ satisfies $L^{2r}\mathbb{E}[G(N)] \to 0$.

In practice, one will typically choose L large enough so that $H_m(L)$ (or more practically, the sample version $\hat{H}_m(L)$) is small.

4 Finite sample properties and an application to climate data

In this section, we examine finite sample behavior of the SFPCA. We do it by a simulation study followed by an application to sea surface temperature. An additional application to hyperspectral images is given in Section III of Supplementary Materials. We consider the behavior of the filter functions $\hat{\phi}_{m,\mathbf{l}}$ defined by (3.13), which we will also call the *spatial functional principal components* (SFPCs). If the data are defined on a 2D grid, the components of $\mathbf{l} = (l_1, l_2)$ can be thought of as longitudinal and latitudinal lags. The ordinary estimated functional PCs, $\hat{v}_m$, do not depend on any spatial parameters. It is convenient to think that the lag zero filters $\hat{\phi}_{m,\mathbf{0}}$ provide a rough approximation, analogous to that provided by the $\hat{v}_m$, and the remaining $\hat{\phi}_{m,\mathbf{l}}$ improve it by reflecting the spatial dependence. We will also consider the sample scores $\hat{Y}_{m,\mathbf{s}}$ defined by (3.14). These scores are to some degree analogous to the ordinary functional PC scores $\hat{\xi}_{m,\mathbf{s}}$, but the whole array must be used to obtain the functions $\hat{X}_{m,\mathbf{s}}$ in (3.15). As a main criterion for the usefulness of the new methodology, we consider the error of the approximation of the observed functions $X_{\mathbf{s}}$ by the low dimensional expansions $\sum_{m=1}^{p} \hat{X}_{m,\mathbf{s}}$.

Calculation of the $\hat{X}_{m,\mathbf{s}}$ as defined in (3.15) is not directly possible at locations $\mathbf{s}$ which are less than $2L$ grid points away from the boundary of the grid. We executed all calculations in this section by extending the formula (3.15) to all points $\mathbf{s}$ on the sample grid with the scores $\hat{Y}_{m,\mathbf{s}'}$ set to zero for locations $\mathbf{s}'$

outside of the grid. This introduces a certain bias near the boundary, but we found that the impact on the overall quality of the approximation $\sum_m \hat{X}_{m,\mathbf{s}}$ is small.

In all computations, we use the Euclidean norm for the estimator $\mathcal{F}_\theta^X$. As for the tuning parameter **q**, $q_i = \lceil n_i^{0.4} \rceil$ yields good results and is an asymptotically valid choice. In small samples or with highly correlated data, this rule-of-thumb is not always able to capture the covariance structure sufficiently. We recommend using the data-adaptive method discussed in Section IV of Supplementary Materials.

The choice of *L* takes into account both the sample size and the mass of the SFPC filters captured, measured by $\hat{W}_m(L) := \sum_{\|l\|_\infty \leq L} || \hat{\phi}_{m,\mathbf{l}} ||^2$. We want to capture in all cases at least 95 percent of the mass of the first SFPC, i.e. $\hat{W}_1(L) > 0.95$. In the examples we considered, this yields an *L* that is small in comparison to the sample size, which is a very desirable property. However, $\hat{W}_m(L)$ will usually be a decreasing function in *m* and with this choice, 95 percent will not be reached for the principal components *m* > 1. Therefore, we suggest to use this criterion merely to find a sensible lower bound for the parameter. Only larger samples allow for *L* to be further increased while it stays proportionately small in comparison to the sample size **n**. This way, $\hat{W}_2(L)$ and $\hat{W}_3(L)$ can also be increased to an acceptable level.

Before presenting the numerical results, we note that in the case of spatial data, *r* = 2, run–times needed to obtain our SFPCA are between 10^2 and 10^3 longer than those needed for the ordinary FPCA. The ratio depends on **q**, *L* and the number of grid points for the integration in the spectral domain; for smaller values of these tuning parameters it is smaller. This is not an impediment to the practical application of our methodology as the run-time on a typical laptop was generally in the range of a few seconds.

The package implementing the SFPCA is available on Github and can be installed directly from the R console using the appropriate function from the `devtools` package: `install_github("kuenzer/fsd")`.

4.1 A simulation study

We simulate realizations

$$X_{s,t}(u), \quad s,t = 1,2,\ldots 50, \; u \in [0,1]$$

of a SFARMA process of Section 2.4 defined by the equations

$$X_{s,t} = A_{10}X_{s-1,t} + A_{01}X_{s,t-1} + \varepsilon_{s,t} + B_{10}\varepsilon_{s-1,t} + B_{01}\varepsilon_{s,t-1} + B_{11}\varepsilon_{s-1,t-1}. \qquad (4.1)$$

The operators $A_{k\ell}$ and $B_{k\ell}$ are Hilbert–Schmidt operators defined in Section IV of Supplementary Materials. The coefficients of their finite dimensional representations change from replication to replication so as not to favor a specific data generating process; a similar approach was used by Aue et al. (2015) and Paparoditis (2018) in the context of functional time series. It is ensured that each replication is a realization of a random field, which is stationary in the sense of Definition 2.1 and for which condition (A.1) holds. The error functions $\varepsilon_{t,s}$ are i.i.d. with covariance functions specified in Section IV of Supplementary Materials, and with marginal normal or t_5 distributions.

To each realization we apply both the new SFPCA and ordinary functional PCA. The quality of dimension reduction is evaluated by the normalized mean squared error (NMSE) defined by

$$\mathrm{NMSE}(p) = \frac{\sum_{\mathbf{s}\in R_{\mathbf{n}}} \left\| X_{\mathbf{s}} - \sum_{m=1}^{p} \hat{X}_{m,\mathbf{s}} \right\|^2}{\sum_{\mathbf{s}\in R_{\mathbf{n}}} \left\| X_{\mathbf{s}} \right\|^2}. \qquad (4.2)$$

The average is calculated over the full region $R_{\mathbf{n}} = \{\mathbf{s} \in \mathbb{Z}^r : 1 \le s_i \le n_i \;\forall 1 \le i \le r\}$. As already noted, the approximation $\sum \hat{X}_{m,\mathbf{s}}$ is less accurate at the boundary,

but to give a fair comparison to ordinary FPCA, we do not exclude the values close to the boundary. The estimation was performed as described in Sections 3.2 and 3.3, with the Bartlett kernel using **q** chosen by our data-adaptive method, which in this setting mostly yields $q_1 = q_2 = 3$ or 4. For the filter functions, we used $L = 3$. In case of ordinary functional PCA, $\hat{X}_{m,\mathbf{s}} = \hat{\xi}_{m,\mathbf{s}} \hat{v}_m$ are the usual estimators of the components in (1.1).

Note that by (3.7), the error (4.2) can be approximated without actually calculating the scores and the reconstructions $\hat{X}_{m,\mathbf{s}}$. In fact, we only need the eigenvalues of the spectral density operator. Thus, one could alternatively use

$$\mathrm{NMSE}^*_{\mathrm{spat}}(p) = 1 - \frac{\sum_{m \le p} \int_{[-\pi,\pi]^r} \hat{\lambda}_m(\boldsymbol{\theta})\, d\boldsymbol{\theta}}{\sum_{m \ge 1} \int_{[-\pi,\pi]^r} \hat{\lambda}_m(\boldsymbol{\theta})\, d\boldsymbol{\theta}}. \qquad (4.3)$$

We observed the values of $\mathrm{NMSE}^*_{\mathrm{spat}}(p)$ to be usually quite close to those computed in (4.2). $\mathrm{NMSE}^*_{\mathrm{spat}}(p)$ and $\mathrm{NMSE}(p)$ for the climate data are shown in Figure 4. The measure NMSE^* describes the quality of the approximation with no boundary effects and $L = \infty$.

Table 2 summarizes the results of our simulations. The new SFPCs provide a more effective dimension reduction in the function space; depending on the setting, the advantage ranges from 3.6 to 26.7 percentage points. If we exclude the boundary of the region and take the NMSE only over the region $\{\mathbf{s} \in \mathbb{Z}^r : 1 + 2L \le s_i \le n_i - 2L \ \forall 1 \le i \le r\}$, then the NMSE drops again by about 0.5 percentage points across all settings in Table 2. (The **q** chosen by the automatic routine is not optimal, but we did not use a better value obtained by experimentation.)

Based on these and other numerical experiments we have performed, we conclude that the SFPCA works well for data that exhibit typical spatial dependence, with a roughly exponential decay. For independent spatially indexed functions, our method has no advantage as compared to ordinary

functional PCA. Since the scores of the ordinary principal components are already uncorrelated at all spatial lags, the SFPC filters concentrate at zero lags, and actually coincide with the ordinary FPC functions. The extra filters generate noise. The other extreme case where our method doesn't improve much over ordinary FPCA, is when the spatial dependence is very strong — approaching persistence. For very strongly spatially correlated functions, the SFPC filter functions again have most of the mass concentrated at the origin.

4.2 Application to sea surface temperature data

In this section, we consider sea surface temperature (SST) data, which is part of the NOAA Optimum Interpolation Sea Surface Temperature data obtained by aggregation of satellite observations and in-situ measurements by ships and buoys. This produces daily observations on a grid with resolution 0.25 degrees, covering the sea area of the whole globe. We use the temperature recorded daily for 33 years since 1982. A snapshot of this data set can be seen in Figure 3. Due to the annual quasi-periodicity, these data can be naturally treated as replications of a spatially indexed functional random field. From this large dataset, we extracted a portion of the Indian ocean ranging from approximately 60° to 93° E longitude and 15° to 44° S latitude. We chose this part because it looks relatively homogeneous and devoid of any major oceanic currents; we can expect that the field is close to covariance-stationary. Since the data are very strongly correlated (neighboring observations are often almost identical), we reduced the grid resolution to 0.75 degree. This eases the computational burden and increases our confidence that condition (A.1) holds; small differences between the observations at the 0.25 degree grid produce very slowly decaying spatial ACFs. The remaining daily data are projected onto a Fourier basis of 15 basis functions, thus obtaining functions $X_{s,t,i}(u), u \in [0,1]$, where the the indices s, t and i represent, respectively, the longitude, latitude and year, while the argument u represents the intra-year time scaled to the unit interval $[0,1]$. Since there is a clear latitudinal trend, we detrended the data by subtracting the mean over the 33 years available in the dataset. Hence, for each year $1982 \le i \le 2014$ we first compute

$$\hat{X}_{s,t}^{(i)} = X_{s,t,i} - \frac{1}{33}\sum_{j=1982}^{2014} X_{s,t,j}, \quad 1 \le s \le n_1 = 45, 1 \le t \le n_2 = 40.$$

This produces a field of SST anomalies on a 45×40 grid, similar in size to the grid used in Section 4.1 (with the dimensions longitude and latitude). For each fixed year *i*, the functional spatial data $\hat{X}_{s,t}^{(i)}$ are still not centered, because of a general temperature increase during this period. Thus, in a second step, we also subtract the spatial average from these anomalies, i.e. we consider

$$\tilde{X}_{s,t}^{(i)} = \hat{X}_{s,t}^{(i)} - \frac{1}{n_1 n_2}\sum_{\sigma,\tau} \hat{X}_{\sigma,\tau}^{(i)}.$$

We note that the physical distances between the nodes of the grid are not equal, but because the region lies at low latitudes, we regard this effect to be negligible for the purpose of our demonstration. A more careful grid construction may be needed for different regions and specific climate science applications.

For each year *i*, we computed the spectral density estimates $\mathcal{F}_{\theta}^{\tilde{X}^{(i)}}$ on a 200 × 200 grid of frequencies in $[-\pi, \pi]^2$. Because of the high correlation in the data, our method chose relatively high values of **q**. The average choice was $q_i = 15$. Again we use NMSE defined in (4.2) to compare the performance of the new SFPCA to ordinary functional PCA. For each year, and each $p = 1, 2, 3$, the NMSE of the reconstruction using SFPCA was considerably smaller than using ordinary FPCA, but the differences were generally smaller than in Table 2 which was based on simulated data. Figure 4 presents the comparison of the NMSEs for all years and $p = 1$. On average, the advantage of $\mathrm{NMSE}_{\mathrm{spat}}(1)$ is 18 percentage points. Representative graphs of the SFPCs and scores, along with details on the selection of **q**, are presented in Section IV of Supplementary Materials.

5 Summary and follow up research

We proposed a dimension reduction technique suitable for functional data indexed by spatial locations on a grid. As a prerequisite, and a contribution of

independent value, we developed mathematical foundations of the spectral analysis of such data, including the spectral theory for spatial linear filters. We introduced a general definition of weak dependence for spatial functional data and showed that spatial functional ARMA processes satisfy it. We established the existence and properties of the decomposition analogous to the Karhunen–Loève expansion, but designed for spatially dependent rather that i.i.d. functions. We proposed estimators of the elements of this decomposition and established their consistency, including convergence rates. Using simulated and real data, we demonstrated that the new technique is superior to the ordinary Karhunen–Loève expansion, if the functions exhibit moderate spatial dependence. The `R` package we developed implements all procedures, and can be used not only for functional, but also for multivariate spatial data.

Our paper has similar objectives as the work of Hörmann et al. (2015) who considered time series of functions. However, to our knowledge, there is no previous work that employs spectral analysis to derive a similar kind of principal component decomposition—even in the context of multivariate data—on a spatial grid. This requires appropriate models to deal with the spatial dependence, the latter being considerably more involved than serial dependence in Hörmann et al. (2015). For example, we cannot employ the concept of causality which is a natural and common assumption in time series analysis, so we developed the framework of spatial L^p-m-approximability. We advanced the theory spatial ARMA processes. We had to develop appropriate theory of spatial filtering and estimation of the spatial spectral density, which are novel results of independent interest. Our conditions for consistency of the individual SFPCs are more detailed than in Hörmann et al. (2015). In particular, we derive convergence rates for the estimators. In most of these results the dependence on the spatial dimension r is highlighted, and many involve the joint behavior of the data in various spatial directions; these issues are absent if $r = 1$. In addition to these theoretical advances, we have also created an R package, so that the complicated calculations and algebraic manipulations needed for the spatial design are conveniently implemented. The package accompanying Hörmann et al. (2015) cannot be used in the

spacial setting. In most general terms, while the concepts for $r = 1$ and $r \geq 2$ are analogous, there are many difficult issues which occur even for $r = 2$, which are not present in a time series context.

In the data analyses presented in Section 4, we focused on the optimal dimension reduction offered by our approach, see relation (3.7). Another key outcome of our construction is the zero-correlation property (3.5), which is not shared by the ordinary Karhunen–Loève expansion. The fact that a spatial functional random field can be decomposed into a few uncorrelated fields opens up possibilities of constructing inferential procedures which might otherwise be very complex and have poor finite sample properties. We illustrate this point by hinting how the SFPCA could be used to test normality, which is often assumed in spatial statistics. If the functions $X_{\mathbf{s}}$ are jointly Gaussian, then the SFPC scores $Y_{m,\mathbf{s}}$ are Gaussian as well. Furthermore we know that for all $m_1 \neq m_2$ and any $\mathbf{s}_1, \mathbf{s}_2 \in \mathbb{Z}^r$, the scores $Y_{m_1,\mathbf{s}_1}$ and $Y_{m_2,\mathbf{s}_2}$ are uncorrelated and therefore independent. A spatial analogue of the commonly used Jarque–Bera statistics would be

$$T = \sum_{m=1}^{p} \left(\frac{N\hat{\mu}_{m,3}^2}{6\hat{F}_m^{(3)}} + \frac{N(\hat{\mu}_{m,4} - 3\hat{\mu}_{m,2}^2)^2}{24\hat{F}_m^{(4)}} \right), \quad (5.1)$$

where $\hat{\mu}_{m,k}$ are the centered empirical moments,

$$\hat{\mu}_{m,k} = \frac{1}{N} \sum_{\mathbf{s} \in R_{\mathbf{n}}} (\hat{Y}_{m,\mathbf{s}} - \bar{Y}_m)^k,$$

$R_{\mathbf{n}} = \{\mathbf{s} \in \mathbb{Z}^r : 1 \leq s_i \leq n_i \ \forall 1 \leq i \leq r\}, |R_{\mathbf{n}}| = N$, whose variance can be estimated by

$$\hat{F}_m^{(k)} = \sum_{\|\mathbf{h}\| \leq L_{\max}} (\hat{\gamma}_{\mathbf{h}}^{Y_m})^k, \hat{\gamma}_{\mathbf{h}}^{Y_m} = \int_{[-\pi,\pi]^r} \hat{\lambda}_m(\mathbf{s}) e^{-\mathrm{i}\mathbf{h}^\top \mathbf{s}} d\mathbf{s}.$$

The key point is that under the null hypothesis of normality, the p terms in (5.1) can be expected to be approximately independent, so the asymptotic null distribution of T would be χ_{2p}^2. (The fact that each component is

asymptotically χ^2_2 should follow by extensions of the arguments in Lobato and Velasco (2004) and Górecki et al. (2018).)

Testing stationarity of a functional field by reducing the problem to testing the stationarity of p uncorrelated *scalar* fields of scores should also be possible. Another broad class of applications of the zero correlation property could be to bootstrap procedures. At least under normality, the independence of the sequences $(Y_{m_1,\cdot})$ and $(Y_{m_2,\cdot})$, if $m_1 \neq m_2$, can be expected to lead to simpler resampling schemes. Finally, our methodology requires that functions are completely observed and defined on a regular spatial grid. A task of practical importance would be to develop methods for transforming functional data which are sparsely observed and not defined on regular grids to such a form in some optimal way; French and Kokoszka (2020) make an effort in this direction. One can also conceive that methodology that allows to work more directly with irregularly spaced and incomplete functions might be developed. It is hoped that this work has laid a useful foundation for research in these directions, and will be received with some interest by the functional data and spatial statistics research communities.

6 Supplementary materials

Supplementary Materials contain proofs of all mathematical statements of Sections 2 and 3, an additional data example and elaborations on the numerical and graphical results presented in Section 4.

Acknowledgements We thank the Associate Editor and four anonymous referees for detailed and constructive advice, which contributed to the enhancement of the substance of this research and to its presentation. This research was partially supported by the NSF grant DMS–1914882.

References

Aston, J., Pigoli, D. and Tavakoli, S. (2017). Tests for separability in nonparametric covariance operators of random surfaces. *The Annals of Statistics*, **45,** 1431–1461.

Aue, A., Norinho, D. D. and Hörmann, S. (2015). On the prediction of stationary functional time series. *Journal of the American Statistical Association*, **110,** 378–392.

Bernardi, M. S., Sangalli, L. M., Mazza, G. and Ramsay, J. O. (2017). A penalized regression model for spatial functional data with application to the analysis of the production of waste in Venice province. *Stochastic Environmental Research and Risk Assessment*, **31,** 23–38.

Bosq, D. (2000). *Linear Processes in Function Spaces*. Springer.

Bradley, R. C. (2007). *Introduction to Strong Mixing Conditions*, volume 1. Kendrick Press.

Brillinger, D. R. (1975). *Time Series: Data Analysis and Theory*. Holt, New York.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer, New York.

Campbell, J. B. and Wynne, R. H. (2011). *Introduction to Remote Sensing*. Guilford Press, New York.

Chilès, J. P. and Delfiner, P. (1999). *Geostatistics: Modeling Spatial Uncertainty*. Wiley.

Constantinou, P., Kokoszka, P. and Reimherr, M. (2017). Testing separability of space-time functional processes. *Biometrika*, **104,** 425–437.

Cressie, N. and Wikle, C. K. (2011). *Statistics for Spatio–Temporal Data*. Wiley.

Delicado, P., Giraldo, R., Comas, C. and Mateu, J. (2010). Statistics for spatial functional data: some recent contributions. *Environmetrics*, **21,** 224–239.

Drapatz, M. (2016). Strictly stationary solutions of spatial ARMA equations. *Annals of the Institute of Statistical Mathematics*, **68,** 385–412.

El Machkouri, M., Volný, D. and Wu, W. B. (2013). A central limit theorem for stationary random fields. *Stochastic Processes and their Applications*, **123,** 1–14.

French, J. and Kokoszka, P. (2020). A sandwich smoother for spatio-temporal data. *Spatial Statistics*, **00,** 000–000; Under revision.

Gelfand, A. E., Diggle, P. J., Fuentes, M. and Guttorp, P. (2010) (eds). *Handbook of Spatial Statistics*. CRC Press.

Górecki, T., Hörmann, S., Horváth, L. and Kokoszka, P. (2018). Testing normality of functional time series. *Journal of Time Series Analysis*, **39,** 471–487.

Grenander, U. and Rosenblatt, M. (1957). *Statistical Analysis of Stationary Time Series*. Chelsea Publishing Company.

Gromenko, O., Kokoszka, P. and Sojka, J. (2017). Evaluation of the cooling trend in the ionosphere using functional regression with incomplete curves. *The Annals of Applied Statistics*, **11,** 898–918.

Gromenko, O., Kokoszka, P., Zhu, L. and Sojka, J. (2012). Estimation and testing for spatially indexed curves with application to ionospheric and magnetic field trends. *The Annals of Applied Statistics*, **6,** 669–696.

Happ, C. and Greven, S. (2018). Multivariate functional principal component analysis for data observed on different (dimensional) domains. *Journal of the American Statistical Association*, **113,** 649–659.

Hörmann, S., Kidziński, L. and Hallin, M. (2015). Dynamic functional principal components. *Journal of the Royal Statistical Society(B)*, **77,** 319–348.

Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *The Annals of Statistics*, **38,** 1845–1884.

Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. Springer.

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

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

Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Wiley.

Jolliffe, I. T. (2002). *Principal Component Analysis*. Springer.

Klepsch, J., Klüppelberg, C. and Wei, T. (2017). Prediction of functional ARMA processes with an application to traffic data. *Econometrics and statistics*, **1,** 128–149.

Kokoszka, P. and Reimherr, M. (2017). *Introduction to Functional Data Analysis*. CRC Press.

Liu, C., Ray, S. and Hooker, G. (2017). Functional principal components analysis of spatially correlated data. *Statistics and Computing*, **27,** 1639–1654.

Lobato, I. and Velasco, C. (2004). A simple test of normality for time series. *Econometric Theory*, **20,** 671–689.

Lorenz, E. N. (1956). Empirical orthogonal functions and statistical weather prediction. Technical Report. Massachusetts Institute of Technology. Science Report 1, Department of Meteorology.

Mateu, J. and Giraldo, R. (2019) (eds). *Geostatistical Functional Data Analysis: Theory and Methods*. Wiley. Forthcoming.

Mearns, L., Gutowski, W., Jones, R., Leung, R., McGinnis, S., Nunes, A. and Qian, Y. (2009). A regional climate change assessment program for North America. *EOS, Transactions American Geophysical Union*, **90,** 311–311.

Meyer, M., Jentsch, C. and Kreiss, J-P. (2017). Baxter's inequality and sieve bootstrap for random fields. *Bernoulli*, **23,** 2988–3020.

Panaretos, V. M. and Tavakoli, S. (2013a). Fourier analysis of stationary time series in function space. *The Annals of Statistics*, **41,** 568–603.

Panaretos, V. M. and Tavakoli, S. (2013b). Cramér–Karhunen–Loéve representation and harmonic principal component analysis of functional time series. *Stochastic Processes and their Applications*, **123,** 2779–2807.

Paparoditis, E. (2018). Sieve bootstrap for functional time series. *The Annals of Statistics*, **46,** 3510–3538.

Petersen, R. C., Aisen, P. S., Beckett, L. A., Donohue, M. C., Gamst, A. C., Harvey, D. J., Jack, C. R., Jagust, W. J., Shaw, L. M., Toga, A. W., Trojanowski, J. Q. and Weiner, M. W. (2010). Alzheimer's disease neuroimaging initiative (ADNI). *Neurology*, **74,** 201–209.

Prause, A. and Steland, A. (2018). Estimation of the asymptotic variance of univariate and multivariate random fields and statistical inference. *Electronic Journal of Statistics*, **12,** 890–940.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer.

Ruiz-Medina, M. D. (2012). Spatial functional prediction from spatial autoregressive Hilbertian processes. *Environmetrics*, **23,** 119–128.

Shao, X. and Wu, W. B. (2007). Asymptotic spectral theory for nonlinear time series. *The Annals of Statistics*, **35,** 1773–1801.

Shumway, R. H. and Stoffer, D. S. (2017). *Time Series Analysis and Its Applications with R Examples*. Springer.

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

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, **41,** 434–449.

Zhang, X. (2016). White noise testing and model diagnostic checking for functional time series. *Journal of Econometrics*, **194,** 76–95.

Zhang, X., Shao, X., Hayhoe, K. and Wuebbles, D. (2011). Testing the structural stability of temporally dependent functional observations and application to climate projections. *Electronic Journal of Statistics*, **5,** 1765–1796.

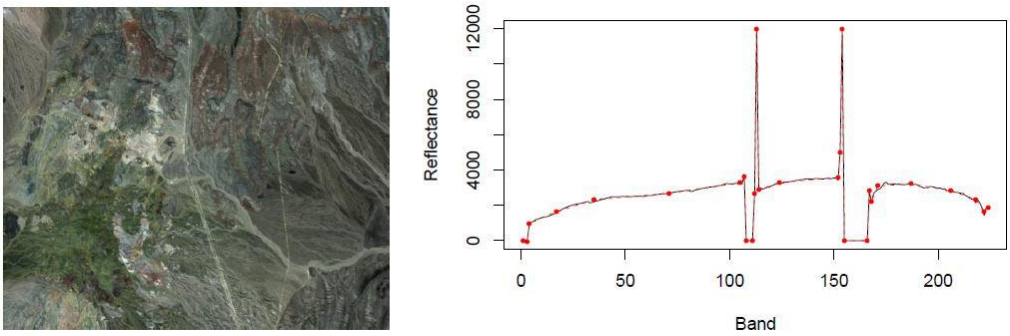

Fig. 1 Left: the hyperspectral image of the Cuprite mining area in Nevada in a pseudo true color composite. The used frequency bands for the red, green and blue channels are band 32, 21, and 13, which correspond to wavelengths of approximately 670, 560 and 480 nm. Right: The reflectance curve for the pixel with coordinates (150, 40) in the lower left part of the image. The raw data is the continuous black line, while the fitted B-Spline curve (with its knots marked) is the dashed line. The two lines almost overlap visually.

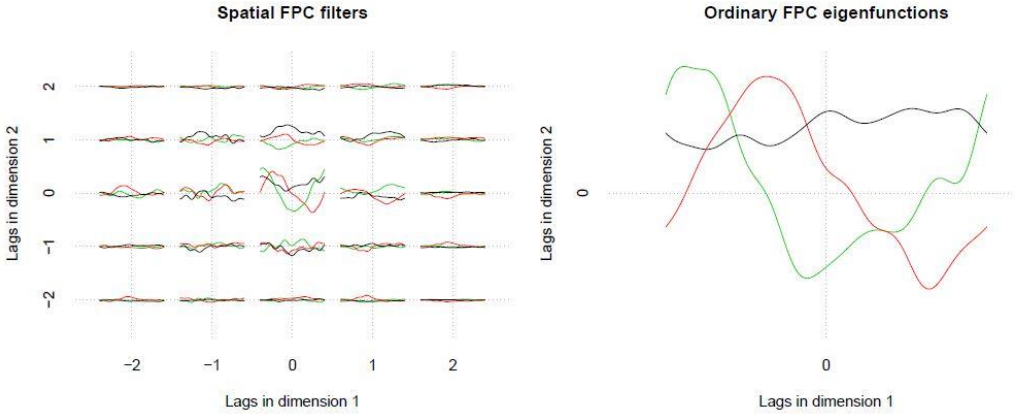

Fig. 2 The principal components for one simulated sample. Left: SFPC filter functions $\hat{\phi}_{m,\mathbf{l}}$ for the first three SFPCs. Right: ordinary functional PCs. For this particular replication, the NMSE is 0.4960 with the new SFPCs and 0.6346 with ordinary FPCs.

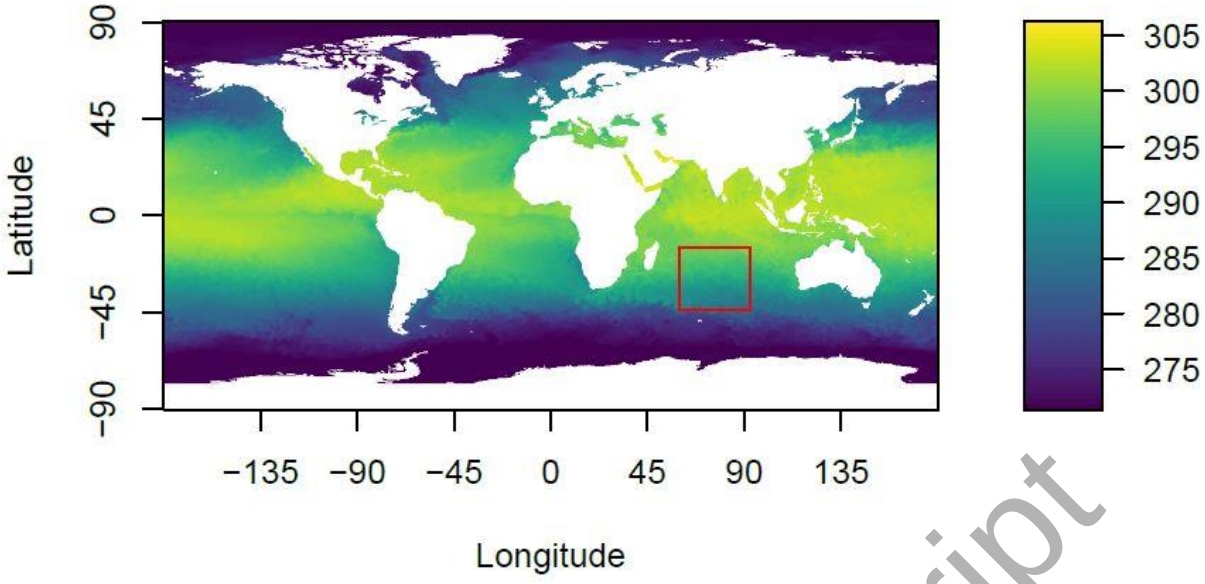

Fig. 3 The sea surface temperature (SST) for July 1, 2012, recorded on a grid with resolution 0.25° for both longitude and latitude. The scale is in Kelvin. The red rectangle marks the area we used for demonstrating the performance of the SFPCA.

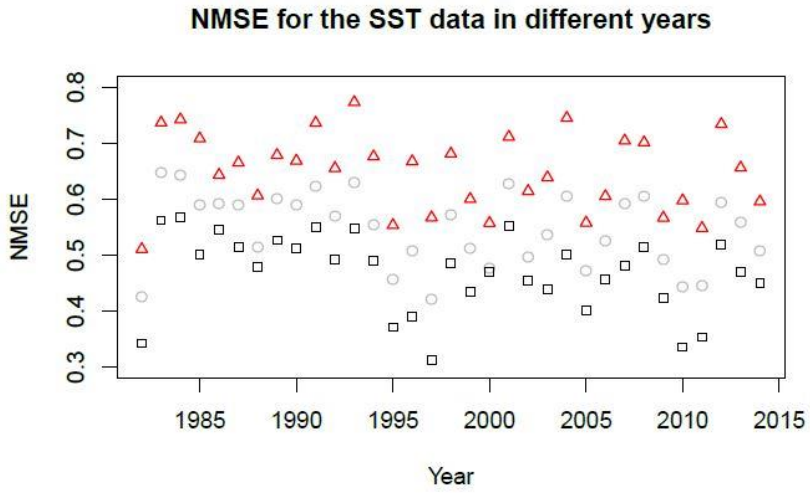

Fig. 4 The NMSE for the detrended SST data $(\tilde{X}_{s,t}^{(i)})_{s,t}$ for all years *i* and the first PC (p = 1). Red triangle: ordinary FPCA; Black square: SFPCA based on L = 12. Grey circle: $\mathrm{NMSE}_{\mathrm{spat}}^{*}$ (defined in (4.3)).

Table 1 Normalized mean squared errors of dimension reduction for the Cuprite hyperspectral image data with different numbers, p, of principal components. The efficiency gain is computed from the ratio of the error of the reconstruction using ordinary FPCs to the error based on the new SFPCs.

p	Spatial	Ordinary	Efficiency gain
1	0.5360	0.6885	28%
2	0.2967	0.4384	48%
3	0.1619	0.2047	26%

Table 2 The normalized mean squared errors (4.2). The average and the standard deviation (in brackets) are based on 1000 replications.

	$p = 1$		$p = 2$		$p = 3$	
Noise	Spatial	Ordinary	Spatial	Ordinary	Spatial	Ordinary
Gaussian	0.4738	0.6016	0.2968	0.3731	0.1970	0.2367
	(0.0400)	(0.0429)	(0.0229)	(0.0311)	(0.0170)	(0.0218)
t_5	0.4720	0.6020	0.2956	0.3730	0.1962	0.2368
	(0.0408)	(0.0438)	(0.0235)	(0.0319)	(0.0172)	(0.0223)