An Overview of Univariate and Multivariate Karhunen Loève Expansions in Statistics

Ranadeep Daw[1] · Matthew Simpson[2] · Christopher K. Wikle[1] ·
Scott H. Holan[1,3] · Jonathan R. Bradley[4]

Abstract

Dependent data are ubiquitous in statistics and across various subject matter domains, with dependencies across space, time, and variables. Basis expansions have proven quite effective in modeling such processes, particularly in the context of functional data and high-dimensional spatial, temporal, and spatio-temporal data. One of the most useful basis function representations is given by the Karhunen-Loève expansion (KLE), which is derived from the covariance kernel that controls the dependence of a random process, and can be expressed in terms of reproducing kernel Hilbert spaces. The KLE has been used in a wide variety of disciplines to solve many different types of problems, including dimension reduction, covariance estimation, and optimal spatial regionalization. Despite its utility in the univariate context, the multivariate KLE has been used much less frequently in statistics. This manuscript provides an overview of the KLE, with the goal of illustrating the utility of the univariate KLE and bringing the multivariate version to the attention of a wider audience of statisticians and data scientists. After deriving the KLE from a univariate perspective, we derive the multivariate version and illustrate the implementation of both via simulation and data examples.

✉ Scott H. Holan
   holans@missouri.edu

1   Department of Statistics, University of Missouri, Columbia, MO, US

2   SAS Inc., Cary, NC, US

3   Office of the Associate Director for Research and Methodology, U.S. Census Bureau,
    Washington, DC, US

4   Florida State University, Tallahassee, FL, US

1 Introduction

Observed data are often generated by random processes that induce dependence. Common examples of such data are time series, spatial data, longitudinal data, and functional data. Often, we may only have one realization from such a process over the domain of interest. This can make it challenging to consider realistic dependence structure in statistical models. Arguably, the most common tool for studying dependence structure is the covariance (or correlation) between any pairs of observations or hypothetical observations in the domain. In practice, this requires construction and computation with a covariance or correlation matrix, which can be challenging computationally when the number of observations or inference locations becomes large. An alternate way of analyzing such a random process is by representing it as a series expansion of some deterministic basis functions with associated random coefficients. In practice, a finite number of such basis functions are used to expand the process and the random coefficients are then estimated. This is sometimes referred to as "random field discretization" and it reduces the computational burden since the process is expressed as a low dimensional (reduced rank) series (Wikle 2010). Importantly, this expansion implies a covariance function.

There are many examples of random field discretization methods. The most common examples are spectral expansions (e.g., Shinozuka and Deodatis 1991; Grigoriu 1993). In addition, polynomial based expansion methods (typically called "polynomial chaos") are also quite common in the uncertainty quantification literature (see Li and Der Kiureghian 1993); Zhang and Ellingwood 1994); Xiu 2009) for a review of these methods). Another popular approach of direct relevance to this manuscript is the expansion method proposed by Karhunen (Karhunen 1946) and Loève (Lóeve 1955), which is now widely known as the Karhunen-Loève expansion (KLE).

As we show here, the KLE is analogous to the diagonalization of a given matrix and hence, has close ties with many commonly known statistical techniques such as principal component analysis (PCA) (Hotelling 1933), singular value decomposition (Golub and Van Loan 1996), empirical eigenfunction decomposition (Sirovich 1987), factor analysis (Mulaik 2009), proper orthogonal decomposition (Lumley 1967), and empirical orthogonal function (EOF) decomposition for spatio-temporal data (Cressie and Wikle 2011). One important aspect of the KLE is its bi-orthogonality; i.e., the use of orthogonal basis functions for the expansion and associated uncorrelated random coefficients. As discussed herein, the KLE is the optimal procedure among all reduced-rank expansion methods in the sense of reducing global mean square error.

The literature concerned with the KLE is vast and it is beyond the scope of this manuscript to provide a comprehensive review. Fortunately, many review articles are available that provide historical developments of the KLE and its application in various field. For example, the KLE is tied closely to matrix factorization and one can find a review of matrix decomposition methods in Stewart Gilbert (1993). Additional details of the history of the KLE are provided in Van Trees (2004). Application of the KLE in statistics can be divided into a few distinct categories. For

example, KLE methods are used to simulate random fields as described in Zheng and Dai (2017) and Gutiérrez et al. (1992). In addition, there is a very large literature on using KLE approaches in numerical solution to various types of problems. Many of these approaches are closely related to Galerkin and finite element methods as used in the computational solution of partial differential equations. Some examples can be found in Boente and Fraiman (2000); Phoon et al. (2004, 2002); Hu and Zhang (2015); Spanos et al. (2007); Ghanem and Spanos (2003), and Spanos and Ghanem (1989). Furthermore, Huang et al. (2001) studied the convergence of the KLE when used as a numerical solution approach. In the area of data analysis and applications, KLE methods have been used extensively to analyze spatio-temporal data in the geophysical sciences and statistical analysis of such data. Overviews can be found in Jolliffe and Cadima (2016); Cressie and Wikle (2011); Wikle et al. (2019), and examples of applications include Obled and Creutin (1986); Monahan et al. (2009); Hannachi et al. (2007); Fontanella and Ippoliti (2012); Bradley et al. (2017); Wikle et al. (2019); Hu (2013), and Dong et al. (2006). The KLE has also been applied to the study of the Gaussian Processes (GPs) (Levy 2008; Jin 2014; Greengard and O'Neil 2021) and non-Gaussian processes (Poirion 2016; Dai et al. 2019; Li et al. 2007; Li and Zhang 2013). Further application areas of the KLE are in pattern recognition (Zhao et al. 2020; Barat and Roy 1998; Yamashita et al. 1998; Kirby and Sirovich 1990), machine learning (Yeung et al. 2021; Rasmussen 2003), functional data analysis (Ramsay and Silverman 2002; Castrillon-Candas et al. 2021; Jacques and Preda 2014, 2014; Zapata et al. 2019), and discretization of random fields (Betz et al. 2014; Rahman 2018; Huang et al. 2014; Li et al. 2008) (among many others).

It is apparent that the KLE has become a common approach for modeling dependent processes across many disciplines. That said, by far the majority of these studies have only considered a univariate KLE. Much less attention has been devoted to the multivariate KLE and thus, our primary goal in this manuscript is to revisit the multivariate KLE and describe its utility when modeling in statistics, primarily in the context of spatial and spatio-temporal statistics. As we will show, the multivariate KLE can be cast similar to the univariate version when one considers them from an underlying reproducing kernel Hilbert space (RKHS) perspective. Thus, we also provide an overview of the univariate approach and show through simulation and an optimal spatial data aggregation example that these methods can be effective even when one does not know the underlying dependence structure corresponding to the data generating process.

The remainder of the manuscript is structured as follows. For background, we first present a brief overview of RKHS theory and its connection to the KLE in Sect. 2. This is followed by the multivariate version of the RKHS and KLE in Sect. 3. We then show how the KLE can be used in statistical modeling through simulation and application in Sect. 4, from both the univariate and multivariate perspectives. We provide a brief conclusion in Sect. 5.

2 Univariate Karhunen-Loève Expansion

To set the stage for discussion of the multivariate KLE, we first present theoretical motivation for the univariate KLE, which is based on the theory of reproducing kernel Hilbert spaces and associated representation theorems.

2.1 Reproducing Kernel Hilbert Space (RKHS)

The analysis of a stochastic process involves the study of functional data. That is, we consider the observed data to be a realization from an unknown function with the indexing variable being an argument that belongs to an infinite continuous set. The building blocks of such fields are functions that span the process over the entire domain of interest (e.g., a spatial domain). Thus, we shall begin the mathematical formulation of the KLE by studying a particular vector space formed by functions, namely the Hilbert space, which was first formulated in the beginning of the 20th century with David Hilbert and Erhard Schmidt's study of integral equations. We define the Hilbert space in Definition 2.1. Note, Table 1 provides the notation used in this subsection.

Definition 2.1 (*Hilbert Space*) A Hilbert space $\mathcal{H}$ is a vector space over its domain (e.g., $\mathbb{R}$, $\mathbb{R}^n$, $\mathbb{C}$, etc.) equipped with a norm $\| \cdot \|_{\mathcal{H}}$ such that the space is complete with respect to the norm.

Although this general definition of a Hilbert space may take any space as its argument, we shall specifically focus on spaces defined over $\mathbb{R}$ (univariate) and $\mathbb{R}^n$ (multivariate). We begin reviewing the univariate case and then extend it over $\mathbb{R}^n$ in the multivariate case, noting that any extension to a general metric space is similar to the study over $\mathbb{R}^n$.

The particular Hilbert space of interest in this manuscript is the one formed by all functions defined over a fixed domain and range set. In the remainder of this section we motivate the construction of such a Hilbert space, commonly known as a reproducing kernel Hilbert space (RKHS). More details about the RKHS can be found in Aronszajn (1950); Wahba (1990); Kailath (1971), and Wang (2008), among others.

Table 1 Notation and definitions for RKHS and KLE

Notation	Definitions
$\mathbb{R}^d$	Real space of dimension d
$\mathcal{H}$	A Hilbert space
$\mathcal{D}$	Domain of Hilbert Space, $\mathcal{D} \subseteq \mathbb{R}^M$
$\mathcal{Y}$	Range of Hilbert Space, $\mathcal{Y} \subseteq \mathbb{R}^N$
$\| \cdot \|_{\mathcal{Z}}$	Norm associated with a space $\mathcal{Z}$
$\langle \cdot, \cdot \rangle_{\mathcal{Z}}$	Inner product associated with a space $\mathcal{Z}$
$\mathcal{K}_{\mathcal{H}}$ or $\mathcal{K}$	Kernel associated to RKHS $\mathcal{H}$

We first briefly review the univariate RKHS and KLE. In that context, assume that $\mathcal{Y} \subseteq \mathbb{R}$, although the domain set can be from an arbitrary subspace of $\mathbb{R}^d$. For example, a spatio-temporal univariate random variable $X_{s_1,s_2,t}$ is a 1-dimensional stochastic process, but the domain set is 3-dimensional with, for example, latitude, longitude, and time serving as the three input arguments. Formally, the univariate RKHS is defined as:

Definition 2.2 (*Univariate Reproducing Kernel Hilbert Space*) A Hilbert space $\mathcal{H}$ of functions f from $\mathcal{D}$ to $\mathcal{Y}$ is called a RKHS if there exists a function $\mathcal{K}_{\mathcal{H}} : \mathcal{D} \times \mathcal{D} \to \mathcal{Y}$ such that the following holds:

i   For all $t \in \mathcal{D}$, the function $\mathcal{K}(.,t)$ is an element from the Hilbert space $\mathcal{H}$.
ii   For all $t \in \mathcal{D}$ and for every function $\phi \in \mathcal{H}$, the following relation is satisfied

$$\langle \phi, \mathcal{K}(.,t) \rangle_{\mathcal{H}} = \phi(t)$$

The second property given in (*ii*) above is called the "reproducing property" of the Hilbert space and hence, the function $\mathcal{K}_{\mathcal{H}}$ associated with the space $\mathcal{H}$ is called a reproducing kernel associated with the space $\mathcal{H}$.

To study the properties of a RKHS, we first consider its associated reproducing kernel. Consider the following lemmas (for proofs, see Appendix A):

Lemma 2.1 *The reproducing kernel $\mathcal{K}$ (or $\mathcal{K}_{\mathcal{H}}$) associated with $\mathcal{H}$ is symmetric in its argument.*

Lemma 2.2 *The reproducing kernel $\mathcal{K}_{\mathcal{H}}$ of a Hilbert space $\mathcal{H}$ is unique.*

We now drop the suffix from a reproducing kernel $\mathcal{K}_{\mathcal{H}}$ and simply denote it as $\mathcal{K}$. A reproducing kernel also has another very important property that will be utilized throughout the manuscript.

Lemma 2.3 *A reproducing kernel is always positive semi-definite in the sense of E. H. Moore. That is, given any set of scalars $\{c_i\}_{i=1}^n$ and any set of elements $\{u_i\}_{i=1}^n$ from $\mathcal{D}$ for $i = 1, \ldots, n$, $\sum_{i=1}^n \sum_{j=1}^n c_i c_j \mathcal{K}(u_i, u_j) \geq 0$ for any n.*

The above properties of symmetry and positive semi-definiteness are used to characterize kernels, which are functions of two arguments. Kernels are used extensively in many areas of statistics (non-parametric statistics, Gaussian processes, basis function expansions, etc.). Note, the terms "kernel" and "reproducing kernel" can be used interchangeably since every reproducing kernel satisfies both symmetry and positive semi-definiteness, which are the conditions that kernels must satisfy.

Now, associated with every such kernel $\mathcal{K}$ over a domain $\mathcal{D}$, there exists a Hilbert space $\mathcal{H}$ of functions from $\mathcal{D} \to \mathcal{Y}$ such that $\mathcal{K}$ reproduces the Hilbert Space $\mathcal{H}$. This is known as Moore-Aronszajn Theorem (Aronszajn 1950) and is stated as follows:

Theorem 2.1 (*Moore-Aronszajn Theorem*) *Let $\mathcal{K} : \mathcal{D} \to \mathcal{Y}$ be a kernel. For every $t \in \mathcal{D}$, define the function $\mathcal{K}_t$: $\mathcal{K}_t(x) = \mathcal{K}(t,x)$. Then there exists an unique RKHS $\mathcal{H}_{\mathcal{K}}$ such that $\mathcal{K}$ reproduces $\mathcal{H}_{\mathcal{K}}$. The space is given as:*

$$\mathcal{H}_{\mathcal{K}} = Closure\Big\{ \sum_i c_i \mathcal{K}_{t_i} \Big\}, \qquad c_i \in \mathcal{Y}.$$

The Moore-Aronszajn Theorem completes the connection between kernels and the associated RKHS in both directions – i.e., associated with every kernel there exists a RKHS and vice versa. The RKHS spanned by a kernel is formally called the "native space" associated with the kernel (Schaback 1998).

2.1.1 Examples of Kernels

There are an infinite number of kernels that can be used. In practice, some specific ones have proven useful. We present a few examples here for illustration (for more, see Small and McLeish 2011).

Example 2.1 Let $\{e_1, \ldots, e_n\}$ be an orthonormal basis of $\mathcal{H}$ and define:

$$\mathcal{K}(x, y) = \sum_{i=1}^{n} e_i(x) e_j(y).$$

Then, for any x,

$$\mathcal{K}_y(\cdot) = \mathcal{K}(\cdot, y) = \sum_{i=1}^{n} e_j(y) e_i(\cdot)$$

belongs to $\mathcal{H}$. Now, take any function ψ from $\mathcal{H}$ where

$$\psi(\cdot) = \sum_{i=1}^{n} \lambda_i e_i(\cdot).$$

Then, we get the following

$$\langle \psi, \mathcal{K}_y \rangle \rangle_{\mathcal{H}} = \Big\langle \sum_{i=1}^{n} \lambda_i e_i, \sum_{j=1}^{n} e_j(y) e_i \Big\rangle_{\mathcal{H}}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} \lambda_i e_j(y) \langle e_j, e_i \rangle_{\mathcal{H}}$$

$$= \sum_{i=1}^{n} \lambda_i e_i(y) = \psi(y).$$

Example 2.2 Let $\mathcal{D} = \mathbb{R}^+$. The function $\mathcal{K}(x, y) = min(x, y)$ is a kernel. To show this, we note that

$$\mathcal{K}(x,y) = \int_{\mathbb{R}^+} \int_{\mathbb{R}^+} \mathbf{1}_{[0,x]}(t)\mathbf{1}_{[0,y]}(t)dt,$$

which means there exists a map $\phi(x) = \mathbf{1}_{[0,x]}(t)$ such that $\mathcal{K}(x,y) = \langle \phi(x), \phi(y) \rangle_{\mathcal{H}}$ and hence, it is a kernel.

Example 2.3 Let $\mathcal{L}_2$ denote the set of second order functions from $\mathcal{D} \to \mathbb{R}$. Then, $\mathcal{L}_2$ is a Hilbert space with inner product

$$\langle f, g \rangle_{\mathcal{L}_2} = \int \int fg d\mathbb{P}.$$

Example 2.4 Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and let $\mathcal{H}$ be a Hilbert space with inner product $\langle ., . \rangle_{\mathcal{H}}$. Let $\mathcal{L}_2(\Omega, \mathcal{A}, \mathbb{P})$ be the set of random variables X with values in $\mathcal{H}$. Then, $\mathcal{L}_2(\Omega, \mathcal{A}, \mathbb{P})$ is a Hilbert space with inner product given by

$$\langle f, g \rangle_{\mathcal{L}^2} = \mathbb{E}_{\mathbb{P}}(\langle X, Y \rangle_{\mathcal{H}}).$$

2.2 Representation Theorems

We now shift our focus to representation theorems that facilitate the Karhunen-Loève representation. Representation of any Hilbert space corresponds to finding a set of vectors in another space such that the inner product of the "representation vector" remains the same as the original inner product in the range space. We seek to represent a stochastic process as an infinite series through the Mercer representation theorem, which in turn leads to the Karhunen-Loéve representation. First, we formally define representation.

Definition 2.3 (*Representation*) We consider a stochastic process $X_t, t \in \mathcal{D}$ with the associated kernel being the covariance function $\boldsymbol{C}(s,t) = Cov(x_s, x_t)$. A family of vectors $\{\phi(t) : \mathcal{D} \to \mathcal{Y}\}$ is called a representation of X_t if for all $s, t \in \mathcal{D}$, we have $\langle \phi(s), \phi(t) \rangle = \boldsymbol{C}(s,t)$.

Next, we discuss the Mercer representation of a second order stochastic process $X_t : t \in \mathcal{D}$ with covariance function R (De Vito et al. 2013; Carmeli et al. 2006). The assumptions are the same as before, but we make some additional assumptions concerning the existence of two quantities, v and L_v, as defined by

$$\begin{aligned} v(A) &= \int_A \frac{1}{1 + \|\boldsymbol{C}(x,x)\|_{\mathcal{Y}}} d\mu(x); \ A \in \mathcal{D}, \\ (L_v f)(\cdot) &= \int_A \boldsymbol{C}(\cdot, t) f(t) dv(t); \forall f \in \mathcal{H}. \end{aligned} \tag{1}$$

If the measure v is a bounded, positive measure, the integral $\int \boldsymbol{C}(x,x)dv(x) < \infty$. When this is satisfied, we get the Mercer representation theorem (Theorem 2.2).

Theorem 2.2 (*Mercer Representation Theorem*) *Assume that the conditions in* (1) *hold. Then there exists a countable set* I *and corresponding countable sequence of eigenfunctions* $\{\phi_i : \mathcal{D} \to \mathcal{Y}, i \in I\}$*, orthonormal with respect to* $\langle \cdot, \cdot \rangle_C$*, along with corresponding non-increasing non-negative eigenvalues* $\{\lambda_i\}_{i \in I}$ *such that*

  i $C(u, v) = \sum_{i \in I} \lambda_i \phi_i(u) \phi_i(v)$
  ii   *For all* $i \in I$, $\int_{\mathcal{D}} C(u, v) \phi_i(v) dv = \lambda_i \phi_i(u)$.

The second equation is known as *Fredholm Integral Equation* (Freiberger and Grenander 1965; Holmstrom 1977; Davenport et al. 1958). The proof of Mercer's theorem is beyond the scope of this manuscript (see Riesz and Nagy 1955 for the proof). Given Mercer's theorem, we note the following important lemma.

Lemma 2.4 *Below are the consequences of the Mercer's theorem.*

   i   The family $\{\phi_i\}_{i \in I}$ is an orthogonal basis of L_v
   ii $\int \phi_i(s) \phi_j(s) ds = \delta_{ij}$
  iii The family $\{\sqrt{\lambda_i} \phi_i\}_{i \in I}$ is orthonormal in $\mathcal{H}_C$.

2.2.1 Karhunen-Loève Representation Theorem

Mercer's theorem and the associated lemma lead directly to the Karhunen-Loéve (KL) representation of a stochastic process. Without loss of generality, assume we have mean zero univariate random variables from a stochastic process. Recall, we seek a KLE that represents such a process through a series with orthogonal basis functions and uncorrelated random coefficients. The formal KL representation theorem is stated in Theorem 2.3.

Theorem 2.3 (*Karhuen-Loéve Representation Theorem*) *Let* $\{X_t, t \in \mathcal{D}\}$ *be a zero mean second order stochastic process with continuous covariance function* $C(s, t)$*. Then, there exists a sequence of mean zero, uncorrelated random variables* $\{\alpha_i : i \in I\}$*, with corresponding non-increasing non-negative eigenvalues* $\lambda_i : i \in I$ *with* $\sum_{i \in I} \lambda_i^2 < \infty$ *and orthonormal eigenfunctions* $\phi_i : i \in I$ *such that the following holds*:

$$\int C(u, v) \phi_i(v) dv = \lambda_i \phi_i(u)$$

$$C(u, v) = \sum_{i \in I} \lambda_i \phi_i(u) \phi_i(v)$$

$$\alpha_i = \int \phi_i(t) X_t dt$$

$$X_t = \sum_{i \in I} \alpha_i \phi_i(t).$$

Additionally, if X_t is a Gaussian process, $\alpha_i \sim \mathbb{N}(0, \lambda_i)$.

To recap, the KLE is a direct consequence of the Mercer representation theorem. Since Mercer's theorem finds an orthonormal basis to represent a positive semi-definite matrix, the basis is a representation of a covariance kernel in an infinite series. Because this representation deals with a linear transformation, one can find the associated eigenvalue-eigenvector pairs corresponding to the linear operator. Thus, one obtains an orthonormal basis function representation of the process where the basis expansion coefficients are uncorrelated. This in turn shows the bi-orthogonal property of the KLE.

Without loss of generality, let the index set I be $\{1, 2, \ldots\}$. Then, the convergence of the KLE is in mean square, i.e., $\mathbb{E}[X_t - \sum_{i=1}^{K} \alpha_i \phi_i(t)]^2 \to 0$ as K increases. This is very useful in practice since it is impossible to work directly with an infinite sum of representative features. Hence, in practice, one truncates the summation. Fixing the number of terms, the following theorem provides the rationale as to why the KLE is the best representation in terms of squared-error loss.

Theorem 2.4 *Consider the following representation of a stochastic process* X_t

$$X_p(t) = \sum_{i=1}^{p} \alpha_i \phi_i(t),$$

where $\phi_i(t)$s are mutually orthonormal. Define the error of this truncated representation as $e_p(t) = X(t) - X_p(t)$. Then among all such expansions, the KLE minimises the integrated mean squared error. That is, $\int_t \mathbb{E}[e_p^2(t)]dt$ is minimized when the expansion is the KLE. In this case, we have

$$e_p(t) = \sum_{i=p+1}^{\infty} \alpha_i \phi_i(t).$$

As with principal component analysis, one may decide to choose p terms that explain a pre-specified amount of the variation (e.g., 90% or 95%) using the first few eigenvalues, or consider scree plots, etc. (see Wikle et al. 2019 for examples in spatio-temporal statistics). This also provides the link between the KLE and many statistical techniques close to PCA (e.g., EOFs as mentioned in Sect. 1).

2.3 Implementation of the KLE

For numerical computation, the main idea behind the KLE is to use the Fredholm integral equation. For example, assume that we are given a $n \times n$ covariance matrix C, for which we need to find the KL decomposition numerically. For any such matrix, the eigenfunctions and eigenvalues can be obtained by performing an eigendecomposition using any numerical integration method (e.g., Riemann integration, Gaussian quadrature, trapezoid integration, etc.); i.e., by solving $CW\Psi = \Lambda\Psi$, where Ψ is the collection of eigenvectors and Λ is the diagonal matrix of eigenvectors. The weight matrix W can vary depending on the integral method,

e.g., $W = \frac{1}{n}I$ for Riemann integral, or $W = \text{diag}(\frac{1}{2(n-1)}, \frac{1}{n-1}, \ldots, \frac{1}{n-1}, \frac{1}{2(n-1)})$ for Trapezoid integral etc.

A more common approach in this setting is to consider an expansion method using any choice of basis functions and then project them linearly onto the KL eigenfunctions. This is useful when one has many data points and cannot take the eigendecomposition of a large covariance matrix. The initially chosen basis functions are called "generating basis functions (GBF)" and they can be chosen from any family of basis functions, e.g., the Fourier basis, Legendre polynomials, Haar basis, radial basis, bisquare basis, spline basis, etc.

Consider a one-dimensional index set (e.g., time). Denote the j-th GBF at time t as $\theta_j(t)$ and the i-th KL eigenfunction as $\psi_j(t)$. In matrix form, the GBF matrix is written by $\boldsymbol{\Theta} = ((\theta_{ij})) = \theta_j(i)$. Then the linear expansion of GBFs is accomplished by assuming $\boldsymbol{\Psi} = \boldsymbol{\Theta}F$, where the matrix F contains the expansion coefficients. The Fredholm equation is then used to solve for the unknown coefficients in F, which then are plugged in the above equation to get $\boldsymbol{\Psi}$. The detailed computation for the above method of solving F is explained in Appendix B. It should be noted that when one uses K many GBFs, the expansion method only needs the eigendecomposition of a $K \times K$ positive definite matrix, which is more computationally feasible than doing the same to a $n \times n$ covariance matrix for large n. Hence, this expansion method using GBFs is usually chosen for practical purposes, and will be of our main interest in this manuscript.

We now discuss the case where we assume that the true covariance model is not known and instead we are given one or more realizations from a stochastic process that has an unknown covariance function. When more than one realization is present (for example, in the cases of functional data), an empirical covariance matrix can be computed from the data, which is then used to get the KL decomposition of the unknown covariance matrix. A more common scenario (specifically with spatial or spatio-temporal data) is the presence of a single realization, where an empirical covariance matrix can not be computed without further model assumptions. Even in this case, the KLE can be used, implicitly providing a parameterization that allows for estimation in a manner that corresponds to Mercer's theorem. We discuss one algorithm to implement the KLE in this context as presented in Bradley et al. (2017) and Bradley et al. (2021) for the case of one realization, where the covariance kernel is estimated under a Bayesian framework, and thereby posterior estimates of the KLE are obtained from this estimated covariance. We call this the 'Obled-Creutin' (OC) basis model following a similar early development in Obled and Creutin (1986).

2.4 Obled-Creutin Basis Model

Assume that $Y = (Y_1, \ldots, Y_n)$ are observations following an unknown covariance matrix C with kernel $C(s, t)$ and we want to find the KLE of C based on Y. This is modeled as:

$$Y = \Psi\eta + \epsilon, \tag{2}$$

where Ψ contains the KL eigenfunctions and ϵ are truncation and measurement errors. Like before, starts with a set of GBFs $\theta_j(t)$ and then linearly projects them to get the KL eigenfunctions as $\Psi = \Theta F$. However, without the knowledge of the covariance kernel $C(s,t)$, we cannot estimate the coefficients F and hence the KL eigenfunctions Ψ can not be computed from Θ. So, the algorithm is modified as below. First, define W as the matrix with (i, j)th element as $W_{ij} = \int_t \theta_i(t)\theta_j(t)dt$. For the KL eigenfunctions to satisfy the Fredholm equation, we need the following condition to hold

$$F^T W F = I. \tag{3}$$

Assuming that W is positive definite and let Q be the Cholesky decomposition of W^{-1}. Then, $F = QG$ satisfies Condition (3), where G is any orthonormal matrix of proper size. Hence, given G, and a solution of the KL eigenfunction can be obtained as $\Psi = \Theta QG$. Plugging this into the Model 2, one gets

$$Y = \Theta QG\eta + \epsilon = \Phi v + \epsilon.$$

where $\Phi = \Theta Q$ and $v = G\eta$. Note that Φ is an orthonormal basis function and is computed in an unsupervised manner with only the knowledge of the GBFs, but v needs to be estimated here. If $\lambda_1 \geq \lambda_2 \geq \cdots$ are the eigenvalues of the unknown covariance matrix C, $\text{Cov}(\eta) = \Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots)$ and hence $\text{Cov}(v) = G\Lambda G^T$, where ϵ corresponds to the measurement and truncation error. Now, to perform Bayesian inference, we can specify the following prior distributions:

$$
\begin{aligned}
Y &= \Phi v + \epsilon \\
\epsilon &\sim \mathcal{N}_n(0, \tau^2 I) \\
v &\sim \mathcal{N}_K(0, \Sigma) \\
\Sigma &= \frac{1}{\sigma^2}\left[R^{-1}\mathcal{A}^{(+)}Q'(I - A)QR^{-T} \right]^{-1} = \frac{1}{\sigma^2}M^{-1} \\
\sigma^2 &\sim IG(a_\sigma, b_\sigma) \\
\sigma^2 &\sim IG(a_\tau, b_\tau).
\end{aligned} \tag{4}
$$

This prior for Σ has been used in Bradley et al. (2017) and Bradley et al. (2015) and is called a the Moran's I (MI) prior due to it's similarity with Moran's I statistic in spatial statistics. Here A is the adjacency matrix of the locations, (Q, R) are the QR decomposition of the basis matrix Φ and $\mathcal{A}^{(+)}(\cdot)$ is the best positive definite approximation of a matrix. With model and prior distributions, we can perform MCMC estimation, generating Gibbs samples of the posteriors. Details of the Gibbs sampler for this model ar given in Appendix C. The posterior estimates of Σ are then used to get the eigendecomposition as $\widehat{\Sigma} = \widehat{G}\widehat{\Lambda}\widehat{G}^T$, which is used to get the KL eigenfunctions as $\widehat{\Psi} = \Phi\widehat{G}$ and eigenvectors as diagonal elements of $\widehat{\Lambda}$.

Note that in general one does not need to build the full Bayesian structure as given above to obtain the KLE of the unknown process if one has regularly space observations and enough replicates to estimate the required $n \times n$ covariance matrix. Then, the symmetric decomposition of the estimated covariance matrix can give the appropriate KL eigenfunctions. However, as originally demonstrated in Obled and Creutin (1986), in cases where one has irregularly spaced observations (say, in a spatial case), and seeks to have eigenfunctions corresponding to any location, then the KLE approach presented here is optimal, as it takes into account the "area of influence" of a particular data point through the integration of the GBFs.

3 Multivariate Karhunen-Loéve Expansion

Increasingly there is a need to model multivariate dependent processes. The multivariate case is more complicated than the univariate setting because one needs to account for the dependence structure between the random variables as well as the different indexing points. Although the multivariate KLE (MKLE) is analogous to the univariate case presented in Sect. 2, it is more complicated in the sense that there is no unique MKLE.

Before presenting the mathematical details, we shall first discuss the scenario at hand. The response variable in this section will be treated as a multivariate random variable, X_t, indexed by t from an arbitrary continuous index set $\mathcal{D}$. We assume $X_t = (X_{t1}, \ldots, X_{tK})'$. Additional assumptions are the same as in Sect. 2; that is, $t \in \mathcal{D} \in \mathbb{R}^m$ and $\mathcal{H}$ is the Hilbert space of functions from $\mathcal{D}$ to $\mathcal{Y}$, where $\mathcal{Y}$ is now a subset of $\mathbb{R}^K$ for some $K > 1$ with a (multivariate) norm $\| \cdot \|_{\mathcal{H}}$ and inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$. In the remainder of this section we briefly discuss multivariate (reproducing) kernels, the multivariate extension of Mercer's theorem, and the associated multivariate Karhunen-Lóeve representation theorem (for details, see Carmeli et al. 2006; Berlinet and Thomas-Agnan 2004; Chiou et al. 2014).

3.1 Multivariate RKHS

In this subsection we describe the multivariate RKHS. First, we define a multivariate reproducing kernel.

Definition 3.1 (*Multivariate (Reproducing) Kernel*) A multivariate kernel is a function $\mathcal{K}$ from $\mathcal{D} \times \mathcal{D}$ to $\mathcal{Y}$ such that the following holds.

i $\mathcal{K}$ is symmetric, i.e., $\mathcal{K}(u, v) = \mathcal{K}(v, u)^T$.
ii $\mathcal{K}$ is positive definite in the sense of E.H. Moore; i.e., for any collection of vectors $\{c_i\}_{i=1}^n$ and any set of elements $\{u_i\}_{i=1}^n$ from $\mathcal{D}$, the following holds:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i^T \mathcal{K}(u_i, u_j)c_j = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i^T \langle \mathcal{K}(u_i), \mathcal{K}(u_j)\rangle_{\mathcal{H}} c_j$$

$$= \left\langle \sum_{i=1}^{n} c_i^T \mathcal{K}(u_i), \sum_{j=1}^{n} c_j^T \mathcal{K}(u_j)\right\rangle_{\mathcal{H}} \geq 0.$$

Thus, multivariate kernels are analogous to the univariate kernels discussed in Sect. 2, with the extension that they are now matrices rather than scalars. As in the univariate case, any multivariate kernel $\mathcal{K}$ has an associated native Hilbert space $\mathcal{H}_{\mathcal{K}}$. This native space can be constructed as

$$\mathcal{H}_{\mathcal{K}} = Closure\left\{ \sum_i c_i \mathcal{K}_{t_i} \right\} \qquad c_i \in \mathcal{Y},$$

where $\mathcal{K}_t$ are similarly defined. Interested readers can also see De Vito et al. (2013); Carmeli et al. (2006), which discuss extensively the properties of such kernels.

Multivariate kernels still hold the reproducing property, i.e., for any ϕ from the Hilbert space $\mathcal{H}$ associated with the kernel $\mathcal{K}$, $\langle \phi, \mathcal{K}_t\rangle = \phi(t)$ is still satisfied. We can define $\mathcal{K}_t^j$ as the j-th coordinate of the function $\mathcal{K}_t(\cdot)$, which simplifies notation:

$$\mathcal{K}_t^j \in \mathcal{H}j = 1, \ldots, N$$
$$\phi(t) = (\langle \phi, \mathcal{K}_t^1\rangle_{\mathcal{H}}, \ldots, \langle \phi, \mathcal{K}_t^N\rangle_{\mathcal{H}}).$$

3.2 Multivariate Representation Theorems

Similar to the univariate case, Mercer's theorem allows the multivariate kernel to be represented as an infinite series

Theorem 3.1 (*Multivariate Mercer Theorem*) *Assume a kernel $\mathcal{K} : \mathcal{D} \times \mathcal{D} \to \mathcal{Y} \subseteq \mathbb{R}^N$, $\mathcal{H}_{\mathcal{K}}$ is separable, and $\int_{\mathcal{D}} \mathcal{K}(t,t)dt < \infty$. Then, there exists a countable sequence of continuous orthonormal eigenfunctions $\{\phi_i : \mathcal{D} \to \mathcal{Y} : i \in I\}$ and a sequence of non-negative decreasing eigenvalues $\{\lambda_i : i \in I\}$ with $\sum_{i \in I} \lambda_i^2 < \infty$ such that*

$$\lambda_i \phi_i(u) = \int_{\mathcal{D}} \mathcal{K}(u, v)\phi_i(v)dv$$
$$\mathcal{K}(u, v) = \sum_{i \in I} \lambda_i \phi_i(u)\phi_i^T(u).$$

Following from the Mercer representation theorem, the multivariate Karhunen-Loève representation theorem for the multivariate stochastic process X_t is given in Theorem 3.2.

Theorem 3.2 (*Multivariate KL Representation Theorem*) *Assume* $X : \mathcal{D} \to \mathcal{Y}$ *denotes a mean-zero square integrable stochastic process with covariance matrix* $C(t, s) = Cov(X_t, X_s)$. *Then, there exists a countable sequence of random variables* $\{\alpha_i : i \in I\}$ *and associated eigenvectors* $\{\lambda_i : i \in I\}$ *and orthonormal eigenfunctions* $\{\phi_i : \mathcal{D} \to \mathcal{Y} : i \in I\}$, *such that*

$$\lambda_i \phi_i(u) = \int_{\mathcal{D}} \mathcal{K}(u, v) \phi_i(v) dv$$

$$C(u, v) = \sum_{i \in I} \lambda_i \phi_i(u) \phi_i^T(u)$$

$$\alpha_k = \int_{\mathcal{D}} \phi_i(u)^T X(u) du$$

$$X_t = \sum_{i \in I} \alpha_i \phi_i(t),$$

and if X_t *is Gaussian Process,* $\alpha_i \sim \mathcal{N}(0, \lambda_k)$.

Again, this is the direct consequence of the multivariate Mercer theorem.

3.3 Relation between Multivariate and Univariate KLEs

Consider the MKLE as in Sect. 3.2 of the form $X_t = \sum_i \alpha_i \phi_i(t)$, where the multivariate eigenfunctions are vector valued (N-dimensional) functions. An immediate question is how is the multivariate KLE related to the univariate KLEs of each of the corresponding processes. This can be found in Proposition 5 from Happ and Greven (2018), which is stated in the theorem below. Here, we denote the full covariance matrix of X is denoted as the block matrix:

$$C = \begin{bmatrix} C_{11} & \dots & C_{1N} \\ \vdots & \vdots & \vdots \\ C_{N1} & \dots & C_{NN} \end{bmatrix},$$

where the elements of j, k-th blocks are $C_{st}^{jk} = C^{jk}(s, t) = Cov(X_s^{(j)}, X_t^{(k)})$. Let the eigenvalues and eigenvectors of C be $\{\lambda_1 \geq \lambda_2 \geq \dots\}$ and $\{\phi_1, \phi_2, \dots\}$, and let the j-th element of ϕ_i be denoted as $[\phi_i]_j$. The j-th univariate process with covariance matrix C_{jj} has the KLE $X_t^{(j)} = \sum_i \alpha_i^{(j)} \psi_i^{(j)}(t)$, with eigenvalues $\lambda_1^{(j)} \geq \lambda_2^{(j)} \geq \dots$ and eigenfunctions $\{\psi_1^{(j)}, \psi_2^{(j)}, \dots\}$. Then, we have the following theorem to link the multivariate KLE to the univariate KLEs.

Theorem 3.3 (*Relation between Multivariate KLE and process specific Univariate KLEs*) *The multivariate vector* $X_t = (X_t^{(1)}, \dots, X_t^{(N)}) : t \in \mathcal{D}$ *has a KL expansion as* $X_t = \sum_i \alpha_i \phi_i(t)$ *if and only if each of the univariate processes* $X_t^{(1)}, \dots, X_t^{(N)} : t \in \mathcal{D}$ *has an univariate KL expansion. Then the following two conditions hols.*

1. Given the multivariate KLE, the eigenvalues of the covariance kernel corresponding to the j-th univariate process correspond to the eigenvalues of the matrix $\boldsymbol{R}^{(j)}$ with the (m, l)-th element as

$$\boldsymbol{R}^{(j)}_{ml} = \sqrt{\lambda_m \lambda_l} \langle [\boldsymbol{\phi}_m]_j [\boldsymbol{\phi}_l]_j \rangle.$$

Consider the k-th orthonormal eigenvector of $\boldsymbol{R}^{(j)}$ as $\boldsymbol{u}^{(j)}_k$ and denote the l-th entry as $u^{(j)}_{kl}$. Then, the eigenfunctions of $\boldsymbol{C}_{jj}$ are given by

$$\psi^{(j)}_k(t) = \frac{1}{\sqrt{\lambda^{(j)}_k}} \sum_l \sqrt{\lambda_l} u^{(j)}_{kl} [\boldsymbol{\phi}_k]_j(t).$$

2. Given the univariate KLEs of each individual process, the eigenvalues of the large covariance matrix $\boldsymbol{C}$ correspond to the positive eigenvalues of the block matrix $\boldsymbol{K}$ with blocks $\boldsymbol{K}^{(jk)}$, where the (m, l)-th element of $\boldsymbol{K}^{(jk)}$ is $\boldsymbol{K}^{(jk)}_{ml} = \mathrm{Cov}\big(\alpha^{(j)}_m, \alpha^{(k)}_l\big)$. The j-th element of the k-th multivariate eigenfunction of $\boldsymbol{C}$ is constructed as

$$[\boldsymbol{\phi}_k]_j(t) = \sum_i \boldsymbol{v}^{(j)}_{ki} \psi^{(j)}_k(t),$$

where $\boldsymbol{v}^{(j)}_{ki}$ denotes the i-th element of the j-th block of an orthonormal eigenvector $\boldsymbol{v}_k$ of $\boldsymbol{K}$ associated with eigenvalue λ_k.

Theorem 3.3 shows that one does not need to directly deal with the full covariance matrix and instead can work with the process specific KLEs for each of the univariate process. We discuss this alternative approach for constructing the MKLE starting from the univariate processes in the next section.

3.4 Alternative MKLE Construction

We note that it is more difficult to work with the multivariate KL representation for developing expansions as compared to univariate expansions because of the matrix-valued formulation. Hence, it is often of interest to consider whether a reasonable multivariate expansion can be obtained from the individual univariate KLEs. As with any multivariate process, there are multiple ways to achieve such an expansion. We next summarize two such approaches as presented in Cho et al. (2013).

3.4.1 Multiple Uncorrelated KLEs (muKL)

Perhaps the simplest approach is to treat the K-dimensional multivariate process as K observations from one univariate stochastic process by tweaking the indexing variables. To demonstrate this in simple terms, assume that we have a K-variate random stochastic process $\boldsymbol{X}_t = (X^{(1)}_t, \ldots, X^{(K)}_t)$ and assume that we observe n many such vector observations, namely $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, observed at some index points

$t_1, \ldots, t_n \in \mathcal{D}$. The covariance structure is given by a block matrix C with the (j, k)-th block $C^{(jk)}$ with elements $C_{st}^{(jk)} = C^{(jk)}(s, t) = \mathrm{Cov}(X_s^{(j)}, X_t^{(k)})$. The goal here is to create an assembled (i.e., augmented or stacked) univariate process from the vectorized data, such that

$$
\begin{aligned}
\widetilde{X}_t &= X_t^{(1)} \text{ if } t \le n, \\
&= X_{(t-n)}^{(2)} \text{ if } n < t \le 2n, \\
&\vdots \\
&= X_{(t-(K-1)n)}^{(K)} \text{ if } (K-1)n < t \le Kn.
\end{aligned}
$$

The assembled process is still a stochastic process, with a new covariance matrix of size $Kn \times Kn$ with elements

$$
\begin{aligned}
\widetilde{C}(s, t) &= \mathrm{Cov}(\widetilde{X}_s, \widetilde{X}_t) \\
&= \mathrm{Cov}(X_{(s-(i-1)n)}^{(i)}, X_{(t-(j-1)n)}^{(j)}) \\
&= C^{(ij)}(s - (i-1)n, t - (j-1)n),
\end{aligned}
$$

where $(i-1)n < s \le in, (j-1)n < t \le jnt$.

The above representation is simple and easily connected to the univariate setup from Sect. 2. The covariance matrix of the assembled process $\widetilde{X}, \widetilde{C}$, is still the same as the original "unassembled" process, but is indexed using different sets of parameters when compared to the original process. The vector process is now expanded using this assembled covariance matrix $\widetilde{C}$, which is assumed to have an eigen-decomposition with eigenvalues $\lambda_1 \ge \lambda_2 \ge \cdots$ and eigenfunctions $\{\phi_1, \phi_2, \ldots\}$. Hence, using the univariate formulations, $\widetilde{X}$ has a KLE as

$$
\widetilde{X}_t = \sum_{i=1}^{\infty} \alpha_i \phi_i(t), \text{ where } \alpha_i = \int_{\mathcal{D}} \widetilde{X}_t \phi_i(t) dt.
$$

Note that in our vector representation all the observations from the first variable are grouped together, followed by those for the second variable, and so on. Following the same ordering, we can now back-calculate to determine the KLE of any particular variable X_j as follows. First, define $\psi_j^i(t) = \phi_j(t + (i-1)n)$ where $(i-1)n < t \le in$. Then, define the vector $\boldsymbol{\psi}_j(t) = (\psi_j^{(1)}(t), \ldots, \psi_j^{(N)}(t))'$. Now, the multivariate process has the following expansion

$$
X_t = \sum_{j=1}^{\infty} \alpha_j \boldsymbol{\psi}_j(t).
$$

Importantly, the random coefficients (α_is) are shared among different variables in this expansion and hence, induces dependence. This suggests that this approach is best used when the individual processes are somewhat alike. Importantly, one must be careful because the covariance matrix constructed from a collection of small

matrices is not guaranteed to be positive definite, even though individual blocks represent valid covariance matrices Cho et al. (2013).

3.5 Multiple Correlated KLEs (mcKL)

An alternative approach to implement the multivariate KLE is given here. This is a more generally applicable scenario where one has an arbitrary set of correlated stochastic processes. We use the same notation for the available data and covariance matrices as discussed previously. In this approach, one uses the KLE of each individual variable separately first then computes the matrix of correlation coefficients that give the joint covariance matrix (C from above).

Mathematically, consider the j-th process X_j to have a KLE as

$$X_t^{(j)} = X_j(t) = \sum_{m=1}^{\infty} \alpha_m^{(j)} \phi_m^{(j)}(t).$$

For each j, this can be derived from univariate KLE of C_{jj}. For a fixed j, the coefficients $\alpha_m^{(j)}$ are uncorrelated for different m and the variances of $\alpha_m^{(j)}$ is given by the m-th eigenvalue of C_{jj}. From these univariate KLEs, the next step is to get the full covariance matrix by estimating the terms of the cross covariance functions between the variables. For that, we define

$$\mathcal{K}_{ij}^{ml} = \mathrm{Cov}\left[\alpha_m^{(i)}, \alpha_l^{(j)}\right] = \mathbb{E}\left[\alpha_m^{(i)}, \alpha_l^{(j)}\right].$$

Then, the cross-covariances are obtained as the following weighted linear combination of the univariate eigenfunctions

$$
\begin{aligned}
\mathrm{Cov}[X_{si}, X_{tj}] &= \mathrm{Cov}\left[\sum_{m=1}^{\infty} \alpha_m^{(i)} \phi_m^{(i)}(t), \sum_{l=1}^{\infty} \alpha_l^{(j)} \phi_l^{(j)}(t)\right] \\
&= \sum_{m=1}^{\infty} \sum_{l=1}^{\infty} \phi_m^{(i)}(s) \phi_l^{(j)}(t) Cov[\alpha_m^{(i)}, \alpha_l^{(j)}] \\
&= \sum_{m=1}^{\infty} \sum_{l=1}^{\infty} \mathcal{K}_{ij}^{ml} \phi_m^{(i)}(s) \phi_l^{(j)}(t).
\end{aligned}
$$

Hence, to get a multivariate KLE, the remaining task here is to solve for $\mathcal{K}_{ij}^{ml}$. One can show that this can be solved using the following relationship

$$\mathcal{K}_{ij}^{ml} = \int_{\mathcal{D}} \int_{\mathcal{D}} C_{ij}(s,t) \phi_m^{(i)}(s) \phi_l^{(j)}(t)\, ds\, dt. \tag{5}$$

Let K_{ij} be the matrix with (m, l)-th element $\mathcal{K}_{ij}^{ml}$ and define K to be the block matrix with the (i, j)-th block K_{ij}, where the diagonal blocks K_{ii}s are defined to be an identity matrix. Also collect all the variable-wise expansion coefficients into a large vector as in $\mathrm{A} = \left(\alpha_{11}, \alpha_{21}, \ldots, \alpha_{1K}, \alpha_{2K}, \ldots\right)'$. To get the multivariate KLE, consider a Cholesky decomposition of $K = RR^T$ and then define $\widetilde{\mathrm{A}} = R^{-1}\mathrm{A}$. So, $\widetilde{\mathrm{A}}$ are

uncorrelated because $\mathbb{E}\left[\widetilde{A}\widetilde{A}^T\right] = \boldsymbol{R}^{-1}\mathbb{E}\left[AA^T\right]\boldsymbol{R}^{-T} = \boldsymbol{R}^{-1}\boldsymbol{K}\boldsymbol{R}^{-T} = \mathbf{I}$. Similarly the univariate eigenfunctions $\phi_m^{(i)}(\cdot)$ can be collected in a vector to define a multivariate eigenfunction $\Phi_m^{(i)}(\cdot)$ and then they can be made orthonormal by a similar transformation as in $\widetilde{\Phi}(\cdot) = \Phi(\cdot)\boldsymbol{R}$. Thus, we obtain the orthonormal eigenfunctions and uncorrelated coefficients that gives the MKLE from the univariate KLEs of each process.

Construction of the MKLE from the univariate KLEs is the most interesting feature due to it's practical implementation and simplicity of the expansion. The truncated univariate expansion of each process is easy to obtain and a simple cartesian product of the eigenfunctions is all that is needed to construct the multivariate expansion. Discussions and theorems supporting this approach can be found in Steinwart and Christmann (2008); Flaxman et al. (2017); De Vito et al. (2013). Specifically, the Proposition 3.5 from De Vito et al. (2013) provides the necessary intuition that when we deal with a bivariate process, the m-th bivariate eigenfunction can be obtained by multiplying the m-th univariate eigenfunctions from the individual univariate KLEs. Hence, when one knows the univariate expansions, the dependence structure of the off-diagonal elements can be easily derived. For example, if one starts with two random variables, each with a Gaussian covariance kernel : $\mathcal{K}_1(x, y) = \exp(-(x-y)^2/2r_1^2)$ and $\mathcal{K}_2(x, y) = \exp(-(x-y)^2/2r_2^2)$, one can set the following eigenexpansion of the j-th variable as

$$\psi_i^{(j)}(t) = \frac{t^i}{\sqrt{i!}r_j^i}\exp(-x^2/2r_j^2),$$

since the following is satisfied:

$$\sum_i \psi_i^{(j)}(x)\psi_i^{(j)}(y) = \sum_i \frac{x^i}{\sqrt{i!}r_j^i}\exp(-x^2/2r_j^2)\frac{y^i}{\sqrt{i!}r_j^i}\exp(-y^2/2r_j^2)$$

$$= \sum_i \frac{(xy)^i}{i!r_j^{2i}}\exp(-(x^2+y^2)/2r^2)$$

$$= \exp(-(x-y)^2/2r_j^2) = \mathcal{K}_j(x, y).$$

The off-diagonal elements of the full covariance matrix can now be given by

$$\mathcal{K}_{1,2}(x, y) = \sum_i \psi_i^{(1)}(x)\psi_i^{(2)}(y) = \exp\left(-\frac{1}{2}(x/\sigma_1 - y/\sigma_2)^2\right).$$

Note that the off-diagonal elements are not Gaussian kernels.

As another example, a univariate Brownian bridge kernel between [0, 1] is given by

$$\mathcal{K}(x, y) = \min(xy) - xy,$$

and has a KLE with the following eigenfunctions and eigenvalues:

$$e_i(x) = \sqrt{2}sin(i\pi x)$$

$$\lambda_i = \frac{1}{\pi^2 i^2}$$

$$\mathcal{K}(x,y) = \sum_i \frac{\sqrt{2}}{\pi^2 i^2} sin(i\pi x).$$

Given two such correlated variables, both individually following the same kernel, the off-diagonals of the covariance matrix of the bivariate process can then be constructed from the expansion

$$\mathcal{K}_{1,2}(x,y) = \sum_i \frac{2}{\pi^4 i^4} \sin(i\pi x) \sin(i\pi y).$$

3.6 Implementation of the MKLE from GBFs

The implementation of the MKLE is similar to the univariate case. The muKL method essentially transforms the multivariate problem into an univariate problem and hence the analysis is analogous to starting with univariate GBFs to get an univariate KLE of the vectorized process and then re-indexing the eigenfunctions. That said, the mcKL approach is truly multivariate and is of our interest for this discussion. Noting again that mcKL requires individual KLEs from each stochastic process, one needs to start with a set of GBFs for each individual random variable and perform analogous operations as in the univariate case to get the individual KLEs. After that, the computation of the multivariate KLE is done by estimating the K matrix in (5) from Sect. 3.5. The mathematical details are given in Appendix D.

Alternatively, one may also start with multivariate GBFs to get multivariate KLEs. For a K-variate random process, one starts with any multivariate basis function $\phi_k : \mathcal{D} \to \mathbb{R}^K$ for $k = 1,\ldots,M$. Letting $\Phi(t)$ be the $T \times M$ matrix with the k-th column as ϕ_k, the KL eigenfunctions are constructed as

$$\psi_i(\cdot) = \sum_{j=1} f_{ij}\phi_j(\cdot).$$

Similar to the univariate case, one must satisfy the condition $\boldsymbol{F}^T\boldsymbol{W}\boldsymbol{F} = \boldsymbol{I}$ where the i,j-th element of $\boldsymbol{W}$ is $W_{ij} = \int \phi_i(t)^T \phi_j(t)\,dt$. Hence, if $\boldsymbol{Q}$ is a Cholesky decomposition of $\boldsymbol{W}^{-1}$ and if $\boldsymbol{G}$ is some orthonormal matrix, we get $\boldsymbol{F} = \boldsymbol{Q}\boldsymbol{G}$ and thus the KL eigenfunctions are given by $\boldsymbol{\Psi} = \boldsymbol{Q}\boldsymbol{G}\boldsymbol{\Phi}$. Although this is more in line with the univariate GBF expansion, we note that it can be problematic to choose multivariate GBFs from individual processes. Also, this involves much larger matrix operations compared to the univariate KLEs.

We conclude this section by re-emphasizing that when the true (or any target) covariance matrix $\boldsymbol{K}$ is known, the coefficient matrix $\boldsymbol{F}$ can be exactly solved. However, the analogous MKLE case where the multivariate covariance matrix is unknown is still unsolved and is the subject of on-going research.

4 Simulations and Applications

To illustrate the KLE we provide a simple univariate construction example with simulated data in Sect. 4.1. This is followed in Sect. 4.2 by an important application that uses the univeraite KLE as part of a criterion to determine the optimal regionalization of spatial data. This is then followed in Sects. 4.3 and 4.4 with simulated examples illustrating the MKLE. Finally, Sect. 4.5 presents a data example considering multivariate spatio-temporal data (maximum and minimum temperature for weather stations over the USA from 1990 to 1993).

4.1 Univariate: Expansion of a Univariate Exponential Covariance Kernel

In this example, we demonstrate the univariate KLE construction approach using GBFs. The covariance matrix chosen here is based on the covariance function of the form $C(s,t) = \exp\{-\frac{\|s-t\|}{\lambda}\}$. We choose $\lambda = 1$ and $n = 200$ locations in 1-d space to evaluate the covariance matrix, which gives us a 200×200 covariance matrix. The evaluation locations are chosen on a regular grid $t = \{t_1, \ldots, t_n\} = \{-1, -1+\delta, -1+2\delta, \ldots, 1-2\delta, 1-\delta, 1\}$, where $\delta = \frac{2}{(n-1)} = \frac{2}{199}$. As explained in Sect. 2, our goal is to construct the KL eigenfunctions $\psi_j(\cdot)$, which will be obtained as a linear projection of some chosen family of GBFs. Here, we use Legendre polynomial basis functions for GBFs (see Appendix B for details of Legendre polynomials and the computation of the KL eigenfunctions).

Following the methodology described in Sect. 2.3, the first two eigenfunctions from the KLE and the reconstructed autocovariance functions are given in Figs. 1 and 2, respectively. We have used different numbers of GBFs for illustration, but find little difference between the KLE eigenvalues and eigenfunctions for the different values (e.g., see Fig. 1 for a plot of the first two eigenfunctions based on the 10 basis function representation). In addition, the reconstructed autocovariance plots in Fig. 2 show that 10 or more basis functions do a reasonable job of representing the true correlation structure.

4.2 CAGE: Criterion for Spatial Aggregation Error

Spatial change of support is a long-standing problem in geography and spatial statistics in which one seeks to do spatial inference at different spatial scales without inducing effects from aggregation (the so-called "ecological fallacy"). Bradley et al. (2017) showed how the KLE could be used to mitigate this through a criterion for spatial aggregation error (CAGE). We repeat the essence of this approach here to illustrate the value of the univariate KLE in generating optimal regionalizations for spatial statistical inference.

Consider a spatial stochastic process $\mathcal{Y}$ that has been observed across several regions at a particular spatial scale. For example, one might consider the unemployment rate for counties across different states in the United States, yet we might wonder how we could aggregate the county level estimates at different
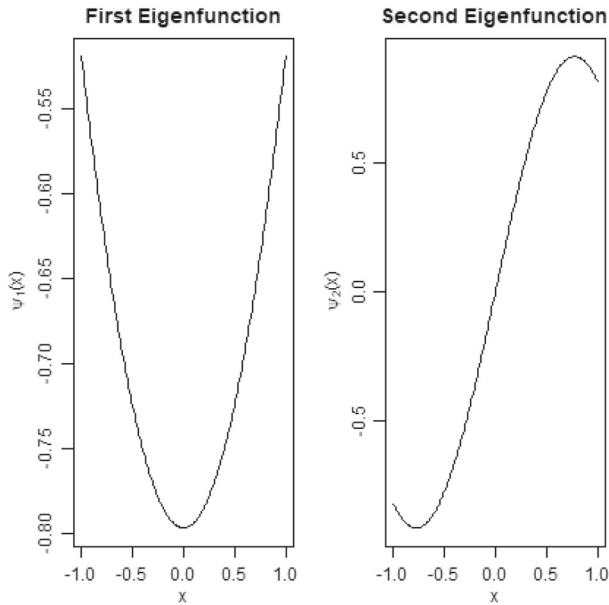
Fig. 1 The first two eigenfunctions $\psi_1(\cdot)$ and $\psi_2(\cdot)$ obtained from the univariate KL representations of an exponential covariance function using 10 generating basis functions
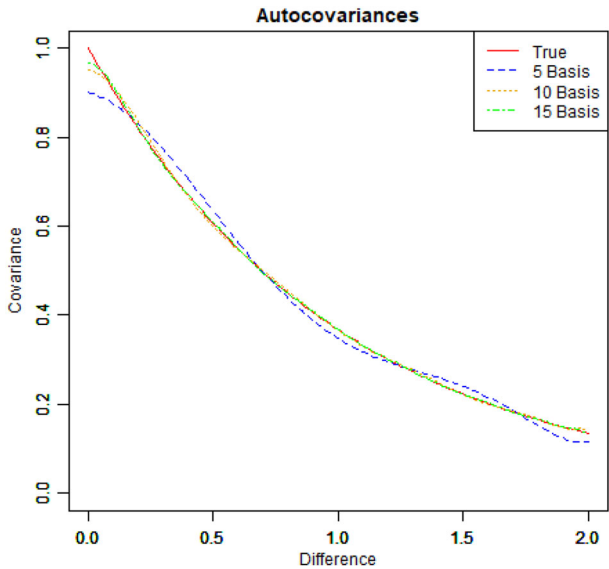

Fig. 2 Reconstructed autocovariance function for the exponential covariance function as a function of univariate lag differences. The true autocorrelation function is given by the red line, and the univariate KL expansions with 5, 10 and 15 basis functions are shown by the blue, golden and green lines, respectively

geographies than those given by the state boundaries. The CAGE procedure assumes there exists some point (or very small spatial) level process Z (denoted $Z_s(\cdot)$) and an associated larger-scale areal level process (denoted $Z_A(\cdot)$). Consider the stochastic process $Z_s(\cdot)$ to have a covariance function $\text{Cov}(Z_s(t), Z_s(u)) = C(t, u)$ and note the KLE of this covariance function is given by

$$C(t, u) = \sum_{j=1}^{\infty} \phi_{s,j}(t) \phi_{s,j}(u) \lambda_j$$

$$Z_s(t) = \sum_{j=1}^{\infty} \phi_{s,j}(t) \alpha_j,$$

where $\text{var}(\alpha_j) = \lambda_j$ and $\alpha_j = \int_{\mathcal{D}} \phi_{s,j}(t) Z_s(t) \, dt$. Then, the areal random variable is obtained from

$$Z_A(U) = \int_{t \in U} Z_s(t) \, dt.$$

The KLE of the areal level process can then be obtained from the point level KLE as

$$\phi_{A,j}(U) = \int_{t \in U} \phi_{s,j}(t) \, dt$$

$$Z_A(U) = \sum_{j=1}^{\infty} \phi_{A,j}(U) \alpha_j$$

$$\text{Cov}(Z_A(U), Z_A(V)) = \sum_{j=1}^{\infty} \phi_{A,j}(U) \phi_{A,j}(V) \lambda_j.$$

The proof of this result can be found in Bradley et al. (2017). This representation in turn helps one to provide an optimal discretization of the point level process that can be used to construct an optimal areal level process from the point level data. The CAGE statistic provided in Bradley et al. (2017) can be used for this purpose and is given by:

$$CAGE(U) = \frac{1}{|U|} \int_{t \in U} \left[\sum_{j=1}^{\infty} \lambda_j [\phi_{s,j}(t) - \phi_{A,j}(U)]^2 \right] dt. \tag{6}$$

The CAGE statistic can be implemented in the `rcage` (Bradley et al. 2021) R package. Here, we apply it to an ocean color dataset (Leeds et al. 2014; Wikle et al. 2013) over the coastal Gulf of Alaska. The dataset contains SeaWiFS ocean color satellite observations – ocean color is a proxy for phytoplankton at the near surface of the ocean. We then use ROMS ocean model output to predict the ocean color because the SeaWiFS data have areas of missing data due to cloud cover. Specifically, we use the ROMS ocean model output variables chlorophyll, sea surface temperature, and sea surface height as covariates to predict the response, SeaWiFs satellite ocean color, using the model given in (4). W consider data for May 12, 2000, which contains 4718 observations. The approach we consider is described in

Sect. 2.4, with a slight change given the existence of the covariates. Specifically, we consider

$$\begin{aligned}
\boldsymbol{Y} &= \boldsymbol{\mu} + \boldsymbol{\Phi}\boldsymbol{v} + \boldsymbol{\epsilon} \\
\boldsymbol{\mu} &= \mu\boldsymbol{1} \\
\mu &\sim \mathcal{N}(0, \sigma_\mu^2) \\
\boldsymbol{\epsilon}|\tau^2 &\sim \mathcal{N}_n(0, \tau^2\boldsymbol{I}) \\
\boldsymbol{v}|\sigma^2 &\sim \mathcal{N}_K(0, \boldsymbol{\Sigma}) \\
\boldsymbol{\Sigma} &= \sigma^2\left[\boldsymbol{R}^{-1}\mathcal{A}^{(+)}\boldsymbol{Q}'(\boldsymbol{I} - \boldsymbol{A})\boldsymbol{Q}\boldsymbol{R}^{-T}\right]^{-1} = \sigma^2\boldsymbol{M}^{-1} \\
\sigma^2 &\sim IG(a_\sigma, b_\sigma) \\
\tau^2 &\sim IG(a_\tau, b_\tau).
\end{aligned}$$

We use bisquare basis functions as the GBFs here, which are defined as

$$\begin{aligned}
\theta_i^{(j)}(t) &= \left(1 - \frac{(t - \boldsymbol{c}_j)^2}{\omega^2}\right)^2 \quad \text{if } \|t - \boldsymbol{c}_j\| \leq \omega \\
&= 0 \qquad\qquad\qquad\quad \text{otherwise,}
\end{aligned}$$

where ω is chosen as 1.5 times the minimum distance among the locations. The knots are carefully chosen following the space-filling "coverage designs" using a swapping algorithm (Johnson et al. 1990). As shown in the model description, the "MI prior" is the choice for the prior for the unknown covariance matrix $\boldsymbol{\Sigma}$. In addition, σ_μ^2 is chosen to be large, $a_\tau = b_\tau = 1$ and a_σ and b_σ are chosen based on the suggestions in Sect. 3.2 of Sørbye and Rue (2014). Posteriors are obtained using Gibbs sampling, and this is then used to estimate the CAGE Criterion (6).

Note that in the CAGE procedure, the clustering component is done in two stages. Consider that S many samples have been drawn from the posterior distribution. Then, based on each sample there is a "prediction" for each element of $\boldsymbol{Y}$, i.e., $\boldsymbol{Y}_s | \cdot : s = 1, \ldots, S$. Now, using any naive clustering algorithm, one obtains clusters for each Gibbs sample. Note that, in most applications, we recommend that one consider a clustering procedure that explicitly accounts for spatial location and can keep elements of the cluster to be spatially contiguous (e.g., in the rcage package one can select structural hierarchical clustering to accomplish this). The cluster centroids are used as the center of the areal units. Then in the second step, the CAGE criterion is calculated based on these areal units to determine the optimal number of clusters. For this example, we considered a range for the number of clusters from $g_L = 200$ to $g_U = 250$, where 217 clusters were retained as optimal. Figure 3 shows the optimal regionalizations that we obtained from this procedure. Note that this aggregation represents an order of magnitude reduction in the number of spatial units compared to the point level prediction. Thus, the KLE, which is fundamental to the CAGE methodology, has demonstrated value in providing dimension reduction for spatial data.
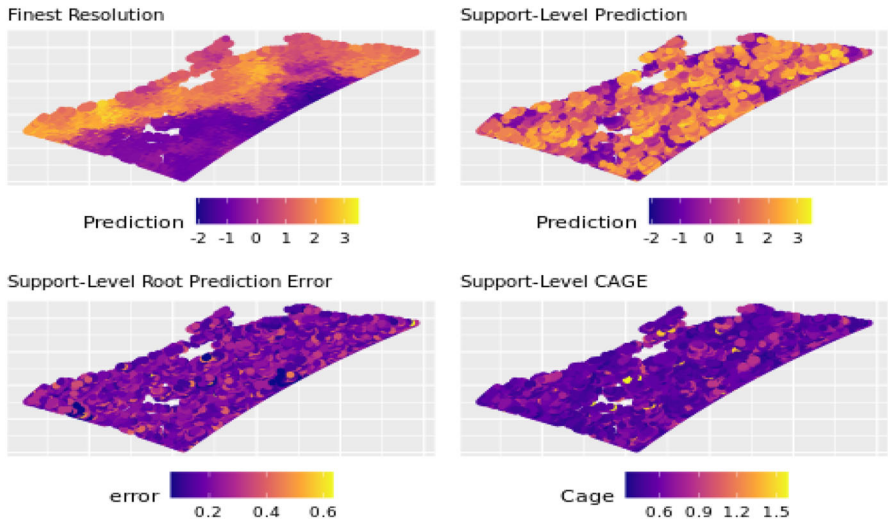
Fig. 3 Regionalization of ocean color by the CAGE approach. The top left panel shows the prediction at the point level. The top right panel shows the prediction for the optimal areal units obtained from the CAGE approach. The bottom left panel shows the areal level prediction errors and the bottom right panel shows the evaluation of the CAGE criterion over the areal units

4.3 Multivariate: Expansion of a Bivariate Exponential Covariance Function

In this example, we demonstrate the MKLE estimation procedure with a bivariate stochastic process $X_t = (X_{t1}, X_{t2})$. The j-th processes is assumed to follow an exponential covariance function C_j given by

$$\text{Cov}(X_{sj}, X_{tj}) = \exp\left(-\frac{\|X_{sj} - X_{tj}\|}{\lambda_j}\right),$$

and the cross-covariance function C_{12} is specified as

$$\text{Cov}(X_{s1}, X_{t2}) = \exp\left(-\frac{\|X_{s1} - X_{t2}\|}{\lambda_{12}}\right).$$

This representation gives a valid covariance function if $\lambda_1 \lambda_2 \le \lambda_{12}^2$. Hence, we consider $\lambda_1 = 0.5$, $\lambda_2 = 0.8$, $\lambda_{12} = 0.6$.

For both processes, orthonormal Legendre polynomials are used as the GBFs. Figure 4 shows the first two eigenfunctions as constructed from the individual KLE of each of the two univariate process. Figure 5 shows the reconstructed autocovariance function for both the univariate processes and also for the cross-covariance. The true and the estimated autocovariance functions align with one another, which is expected from the KL theory.
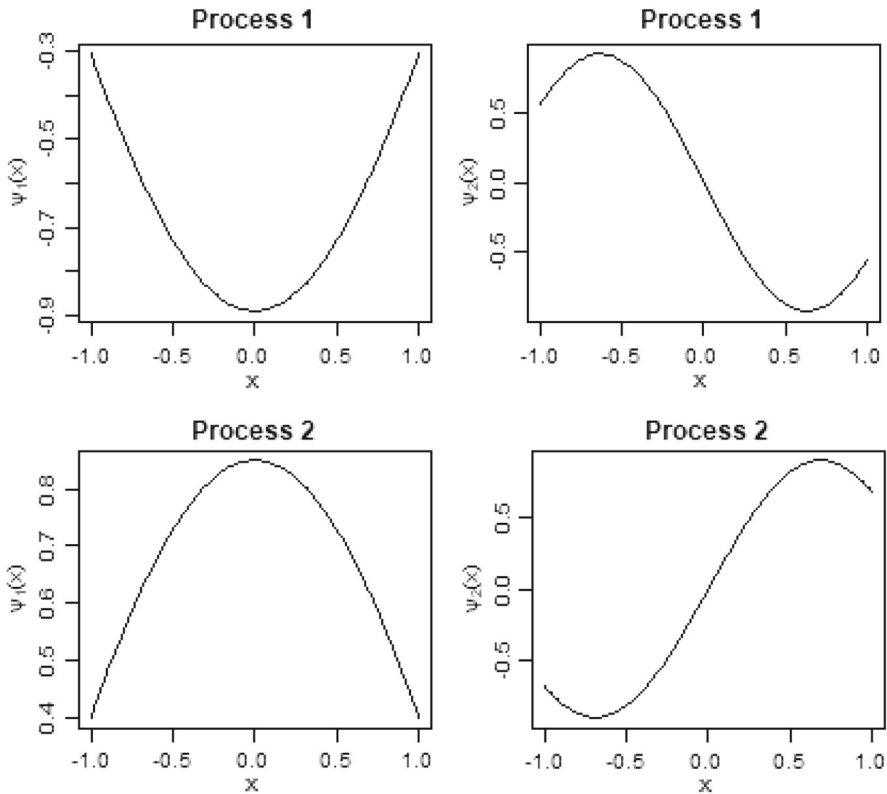
Fig. 4 KL eigenfunctions from two individual univariate processes where the KLE for each univariate process is first obtained using 10 Legendre polynomials as GBFs. The top row contains the first two eigenfunctions $(\psi_1(\cdot)$ and $\psi_2(\cdot))$ from the first process and bottom row contains the same for the second process

4.4 Multivariate: Estimating KLE from Functional Data

Consider the full covariance matrix

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix},$$

where C_{jj} is a covariance matrix based on exponential covariance functions

$$C_{jj}(s,t) = \exp\left(-\frac{\|X_{sj} - X_{tj}\|}{\lambda_j}\right)$$

$$C_{12}(X_{s1}, X_{t2}) = \exp\left(-\frac{\|X_{s1} - X_{t2}\|}{\lambda_{12}}\right).$$

Consider the index points $t_1, \ldots, t_T$ with $X_1, \ldots, X_n (n = 1000)$ simulated from a multivariate Gaussian distribution with the covariance matrix C and mean 0, where
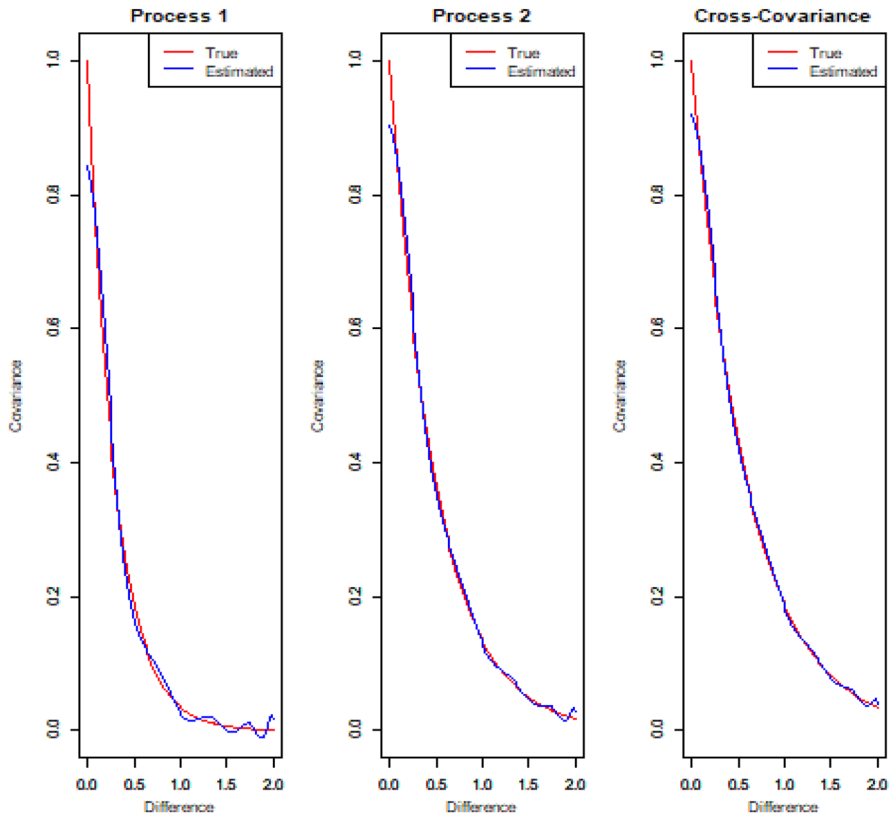
Fig. 5 Left panel: the estimated autocovariance function based on the truncated KLE for process 1 compared to the truth; Middle panel: the estimated autocovariance function based on the truncated KLE for process 2 compared to the truth; Right panel: the estimated cross-covariance function from the MKLE approach compared to the truth

$X_i(t) = (X_{i1}(t), X_{i2}(t))$ with $X_{ij}(t)$ being the i-th observation of the j-th variable at time t. We denote $X = [X_1; X_2]$, where X_j is a $n \times K$ matrix, with each row as an observation and the k-th column of X_j being the evaluations of the j-th variable at the k-th time. Our goal is to get a KLE of the covariance matrix and thereby obtain the expansion of the cross-covariance matrix from individual KLEs. We first get the estimated covariance matrix

$$\widehat{C} = \begin{bmatrix} \widehat{C}_{11} & \widehat{C}_{12} \\ \widehat{C}_{21} & \widehat{C}_{22} \end{bmatrix},$$

where $\widehat{C}_{ij} = \frac{1}{N-1} X_i^T X_j$. From the simulated data, we will use the expansion of the estimated covariance matrix, $\widehat{C}$.

Now, we begin with the individual covariance matrices C_{11} and C_{22} and compute KLEs for each of them following the previously mentioned technique, i.e., start with

K Legendre polynomial basis functions $\boldsymbol{\Theta}^{(j)}$ and then compute the KL eigenfunctions $\boldsymbol{\Psi}^{(j)}$ and eigenvalues $\lambda_i^{(j)}$. Then the multivariate KLE can be found from $\widehat{\boldsymbol{C}}_{12}$ and by orthonormalizing the process-specific random coefficients and eigenfunctions (see details in Appendix D).

The individual eigenfunctions from the two processes reconstructed by the individual KLEs are shown in Fig. 6. The eigenfunctions are similar to the ones in the univariate examples above. We then perform the multivariate KLE by estimating the $\boldsymbol{K}$ matrix and reconstruct the full covariance matrix by estimating the off-diagonals. We compare the true versus the target versus the reconstructed covariances for the 3 different blocks $(\boldsymbol{C}_{11}, \boldsymbol{C}_{22}, \boldsymbol{C}_{12})$ in Fig. 8. Here 'true' is the true covariance model, 'target' is the estimated covariance matrix from the data $\boldsymbol{X}$ (this is what we are attempting to reconstruct from the individual KLEs) and 'estimated' is the reconstructed covariance function from the KLEs.

With the functional data, one can compute the scores (i.e., expansion coefficients) for the data $\boldsymbol{X}$. Given the matrix $\boldsymbol{\Psi}$, the scores are computed via a linear model as in $\boldsymbol{X} = \boldsymbol{\Psi}\boldsymbol{\eta} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ is the prediction error. We plot the score
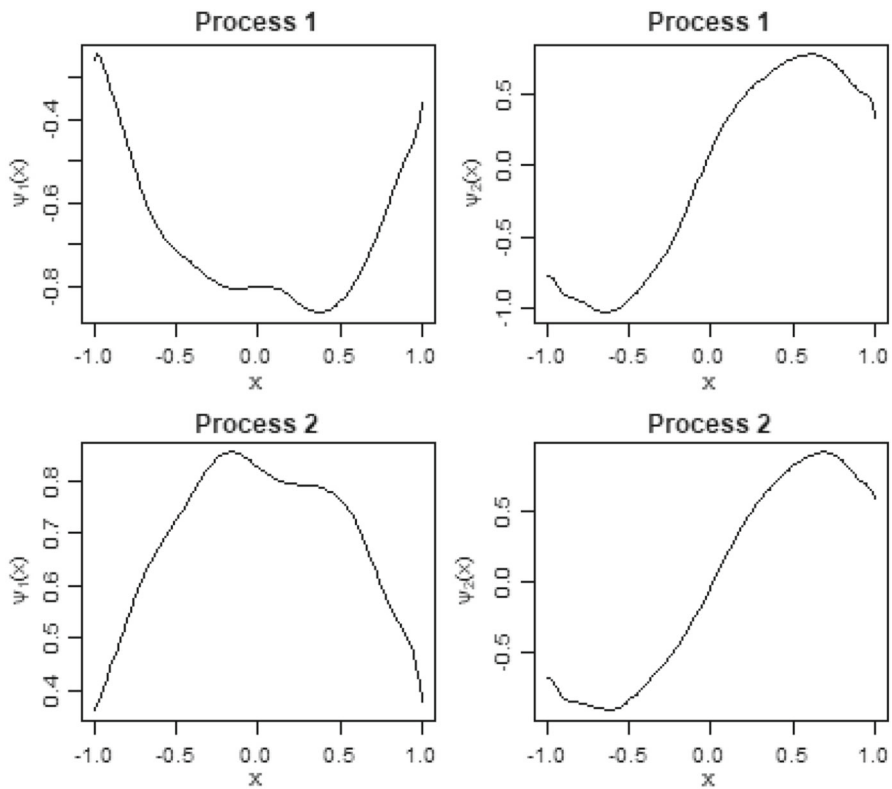

Fig. 6 Univariate process eigenfunctions obtained from the individual process specific KLEs where 10 Legendre polynomials are used for the expansion in each case. The top row contains the first two eigenfunctions $(\psi_1(\cdot)$ and $\psi_2(\cdot))$ from Process 1 and bottom row contains the same for Process 2

variances corresponding to the eigenvalues in Fig. 7 and examine the 95% credible intervals in Table 2 to show that these are consistent with what is expected from the theory.

4.5 Multivariate Data Application: Maximum and Minimum Temperatures Over Space and Time

We consider the KLE of a bivariate process that generates daily maximum and minimum temperature observations obtained from the US National Oceanic and Atmospheric Administration (NOAA) National Climatic Data Center. The spatial locations considered are 138 weather stations in USA (between $32°N - 46°N$ and $80°W - 100°W$) and the observations are recorded daily between 1990 and 1993. The two variables of importance are denoted as *Tmax* (maximum daily temperature) and *Tmin*. For the purposes of demonstration, we consider the observations as replications over the spatial domain and the temporal process at each location is our bivariate stochastic process of interest. So, these data are treated as observations from a bivariate functional (temporal) process with spatial replicates and we demonstrate how an expansion of the bivariate autocorrelation function is obtained from the individual KL expansions of each variable.

As shown with the simulation example of an multivariate functional data in Sect. 4.4, we label the time indices as $t_1, \ldots, t_T$. Removing all the missing observations, the full data with $T = 87$ days are treated as bivariate time series with $n = 1461$ spatial locations serving as replicates. As before, the data are denoted as $X_1, \ldots, X_n$,
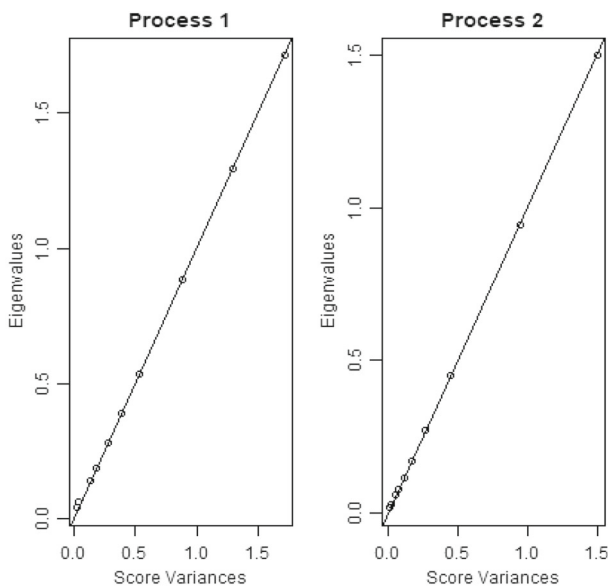

Fig. 7 Score variances vs. eigenvalues for the simulated bivariate data. In this case, close agreement between the score variances and the eigenvalues is observed

Table 2 Summary of the scores (Coefficients of linear expansion using the KL eigenfunctions) : Individual scores are expected to have mean 0. The table presents the 95% credible intervals and the true and observed variances for these scores

Process 1					Process 2				
Mean	Lower 95%	Upper 95%	Observed variance	True variance	Mean	Lower 95%	Upper 95%	Observed variance	True variance
0.00	−0.08	0.08	1.71	1.71	0.05	−0.03	0.12	1.50	1.50
−0.08	−0.15	−0.00	1.29	1.30	−0.04	−0.10	0.02	0.94	0.94
−0.02	−0.08	0.04	0.88	0.88	−0.02	−0.06	0.02	0.45	0.45
0.03	−0.02	0.07	0.53	0.53	0.00	−0.03	0.03	0.27	0.27
0.03	−0.01	0.07	0.39	0.39	−0.02	−0.05	0.00	0.17	0.17
0.02	−0.02	0.05	0.28	0.28	−0.02	−0.04	0.00	0.11	0.11
−0.01	−0.04	0.01	0.19	0.19	0.01	−0.01	0.02	0.08	0.08
0.00	−0.02	0.03	0.14	0.14	0.02	0.00	0.03	0.06	0.06
0.00	−0.01	0.02	0.05	0.06	0.00	−0.01	0.01	0.02	0.02
0.01	−0.00	0.02	0.03	0.04	−0.00	−0.01	0.01	0.01	0.02

where $X_i(t) = (X_{i1}(t), X_{i2}(t))'$ with $X_{ij}(t)$ being the i-th observation of the j-th variable at time t. Write $X = [X_1; X_2]$, where X_j is a $n \times T$ matrix, with each row as an observation and the k-th column of X_j being the value of the j-th variable at the k-th time. Our goal is to get a KLE of the covariance matrix and thereby obtain the expansion of the cross-covariance matrix from individual KLEs.

As was the case for the multivariate functional data example, we estimate the covariance matrix

$$\widehat{C} = \begin{bmatrix} \widehat{C}_{11} & \widehat{C}_{12} \\ \widehat{C}_{21} & \widehat{C}_{22} \end{bmatrix},$$

where $\widehat{C}_{ij} = \frac{1}{N-1} X_i^T X_j$. We will use the expansion of $\widehat{C}$ matrix for the KLE. The individual covariance matrices C_{11} and C_{22} are first expanded using individual KLEs, i.e., using GBFs $\Theta^{(j)}$ and then projecting them linearly to compute the KL eigenfunctions $\Psi^{(j)}$ and eigenvalues $\lambda_i^{(j)}$. As stated in Theorem 3.3, univariate KLEs for the two processes are used to compute the the bivariate KLE of the given bivariate process. This is done by estimating the terms of the K using the estimated cross-covarince matrix $\widehat{C}_{12}$ and then by comuputing K matrix (see details in Appendix D). After computing K, the eigenvalues and eigenvectors of K are computed to get the bivariate orthonormal eigenfunctions and eigenvalues.

Here we first show the estimated eigenfunctions from the two individual univariate stochastic processes in Fig. 9. The multivariate KLE is then used to compute the scores or the coefficients (i.e., the expansion coefficients αs in the KLE $X_t = \sum_i \alpha_i \phi_i(t)$). Note that the k-th estimated coefficient from the variables in X_j are random variables with mean 0 and variance $\lambda_k^{(j)}$. We show this in Table 3.
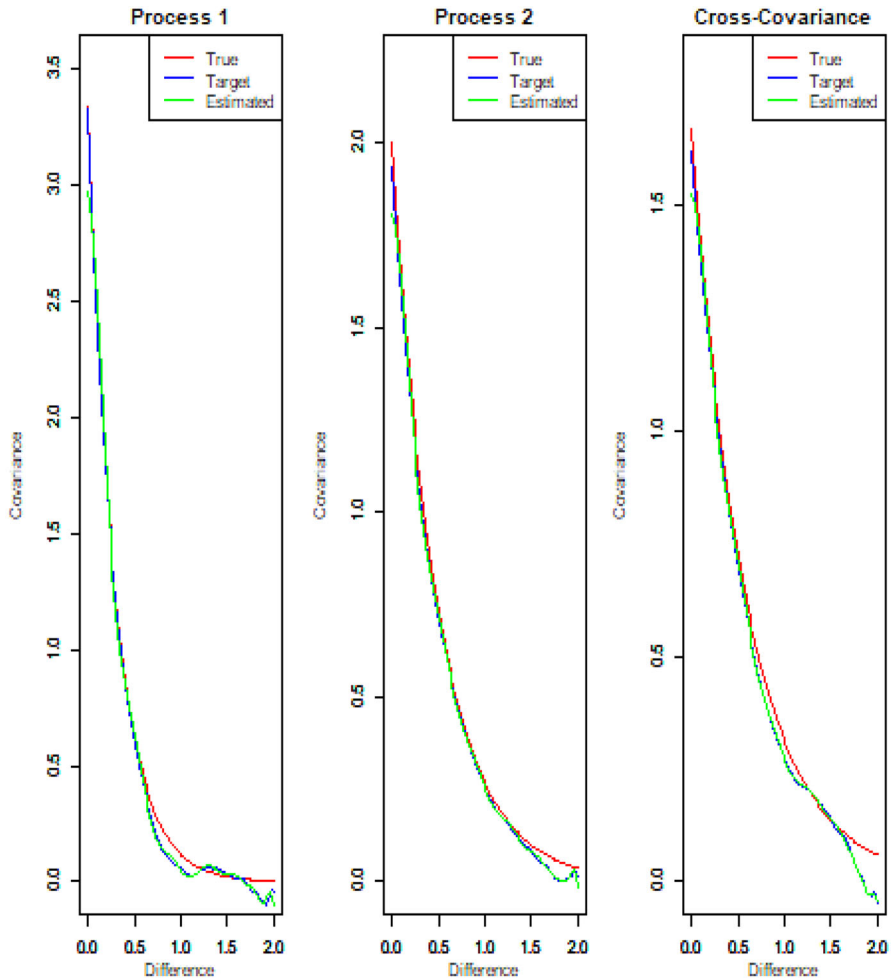
Fig. 8 Left panel: the estimated autocovariance function based on the truncated KLE for Process 1 compared to the truth; Middle panel: the estimated autocovariance function based on the truncated KLE for Process 2 compared to the truth; Right panel: the estimated cross-covariance function from the MKLE approach compared to the truth

Finally, we show in Fig. 10 how the covariance reconstruction performs compared to the original data. Similarly, the cross-covariances are shown in Fig. 11. In both cases, 'target' is the estimated function that we are targeting to reconstruct, and 'estimated' is the estimated function from the KLEs.
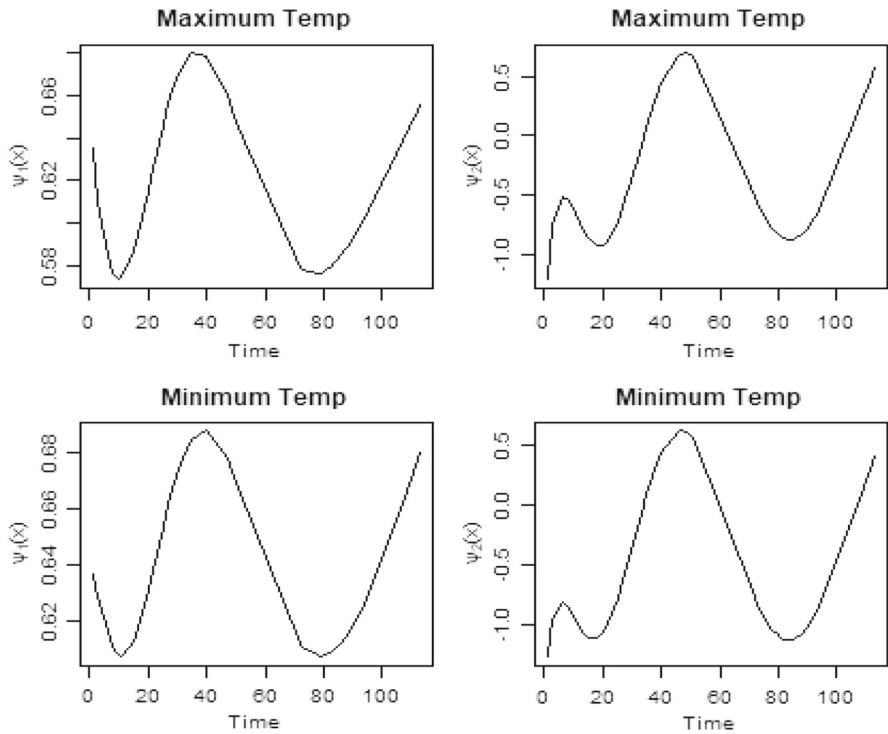
Fig. 9 The first $(\psi_1(\cdot))$ and second $(\psi_2(\cdot))$ estimated eigenfunctions for *Tmax* and *Tmin* plotted against Time

Table 3 Summary of coefficient scores associated with the estimated KLE eigenfunctions

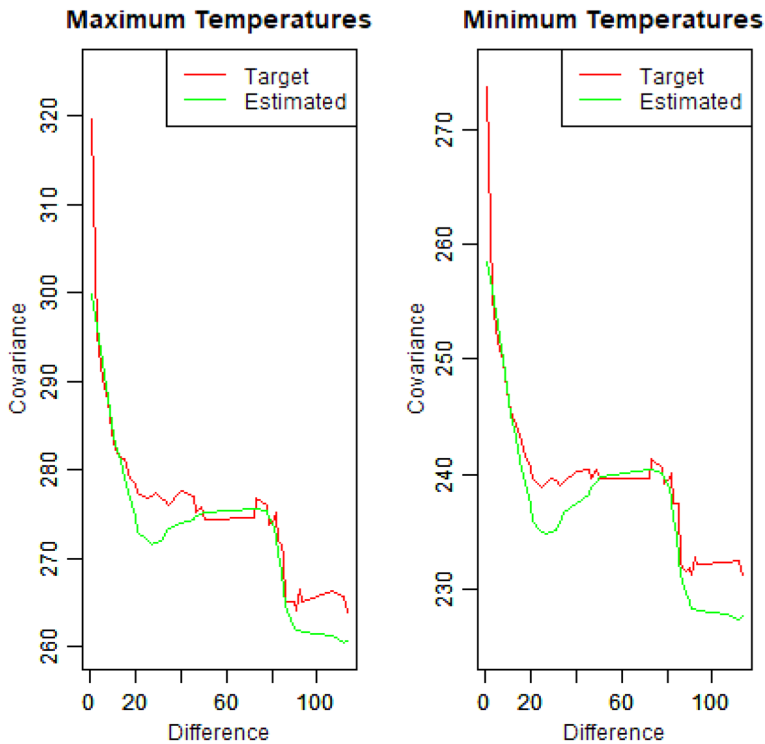| Process 1					Process 2				
Mean	Lower 95%	Upper 95%	Observed variance	True variance	Mean	Lower 95%	Upper 95%	Observed variance	True variance
−0.00	−1.23	1.23	575.18	555.91	−0.00	−1.14	1.14	488.93	482.10
0.00	−0.42	0.42	67.27	20.17	−0.00	−0.38	0.38	55.65	18.63
−0.00	−0.21	0.21	16.31	13.90	−0.00	−0.21	0.21	16.16	9.47
0.00	−0.11	0.11	4.75	4.11	−0.00	−0.10	0.10	3.78	3.13
−0.00	−0.06	0.06	1.33	1.24	0.00	−0.05	0.05	1.03	0.95
0.00	−0.05	0.05	0.79	0.80	0.00	−0.05	0.05	1.10	0.62
−0.00	−0.05	0.05	1.02	0.56	0.00	−0.04	0.04	0.61	0.37
−0.00	−0.05	0.05	1.12	0.41	0.00	−0.05	0.05	0.86	0.28
0.00	−0.06	0.06	1.22	0.30	0.00	−0.05	0.05	0.88	0.22
−0.00	−0.02	0.02	0.19	0.14	−0.00	−0.02	0.02	0.16	0.13

Fig. 10 Individual Process Autocovariances: The target autocovariance function vs the estimated autocovariance function from the KLE for the maximum and minimum temperature example

5 Conclusion

The univariate KLE is well-known in statistics and has many practical uses, including dimension reduction and determining optimal regionalizations in space. In addition, considering the increasing interest in recent years in data from multivariate correlated processes, the multivariate KLE is also a highly relevant topic, but has not seen a great deal of research or application in the statistics literature. The multivariate KLE is applicable over any general multivariate stochastic process with stationary or non-stationary covariance functions. Hence, it is very useful in applications because it allows one to account for dependence structure, both between different processes and within each process.

The multivariate KLE is more complex than the univariate KLE due to the matrix-valued structure of the kernels. But, the univariate KLE can be derived in terms of a RKHS. Similarly, the multivariate case is based on multivariate version of RKHS and corresponding matrix valued kernels. We discussed how one can utilize the univariate KLE of each process to get the multivariate expansion by estimating the terms of the cross covariance matrix. The same can also be done by re-indexing different processes and thereby converting it to a vectorized univariate process, but this latter method is only suitable for a multivariate process where
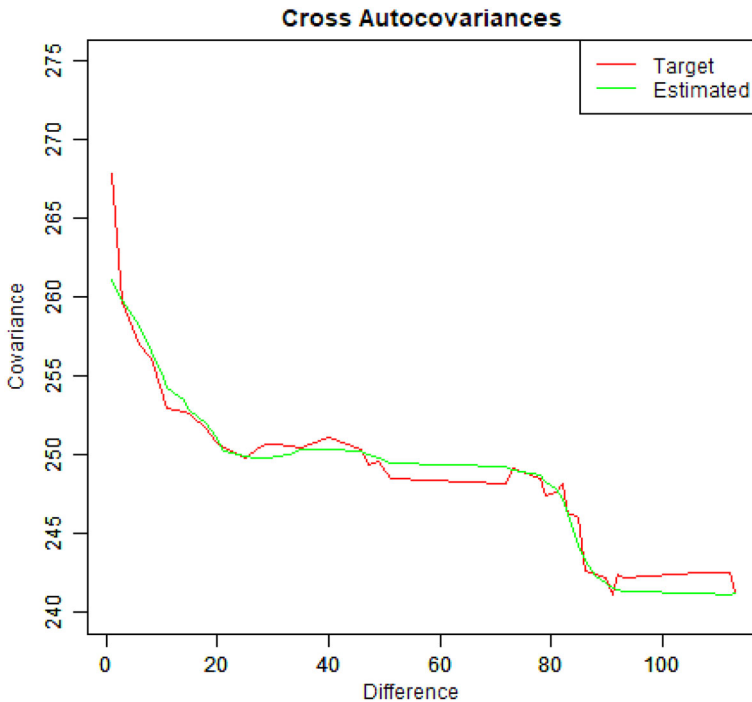
Fig. 11 Target and estimated autocovariance functions from the KLE for the maximum and minimum temperature example

individual processes are somewhat alike. The takeaway here is that there can be more than one way to approach the problem of multivariate KLE and this is a topic of interest for future study.

From the numerical perspective, the use of generalized basis functions (GBFs) has been described here. One can further focus on how the choice different basis functions affects the multivariate KLE. Other similar expansion methods (e.g., polynomial expansions, Fourier series, etc.) are applied as the methods described here. Other approaches to induce dependence between processes (e.g., copula models) could be considered here and are a subject of future research.

Appendix A : Proofs of Lemmas

Proof of Lemma 2.1 For each $t \in \mathcal{D}$, we define the functions $\mathcal{K}_t$ induced by the reproducing kernel $\mathcal{K}_{\mathcal{H}}$ as $\mathcal{K}_t(\cdot) = \mathcal{K}(\cdot, t)$. From the reproducing property, taking $\phi = \mathcal{K}_x$, we get the following immediate consequence

$$\langle \mathcal{K}_x, \mathcal{K}_t \rangle_{\mathcal{H}} = \mathcal{K}_x(t) = \mathcal{K}(x, t).$$

Since inner product is symmetric, this leads to the following

$$\langle \mathcal{K}_x, \mathcal{K}_t \rangle_{\mathcal{H}} = \langle \mathcal{K}_t, \mathcal{K}_x \rangle_{\mathcal{H}}$$
$$\Rightarrow \mathcal{K}(x, t) = \mathcal{K}_x(t) = \mathcal{K}_t(x) = \mathcal{K}(t, x).$$

$\square$

Proof of Lemma 2.2 Suppose there exists two reproducing kernels $\mathcal{K}$ and $\mathcal{K}'$ for $\mathcal{H}$ defined over $\mathcal{D}$. Then for each $t \in \mathcal{D}$, we get the following

$$\begin{aligned}
\|\mathcal{K}_t - \mathcal{K}'_t\|^2_{\mathcal{Y}} &= \langle \mathcal{K} \rangle_t - \mathcal{K}'_x, \mathcal{K}_x - \mathcal{K}'_x \rangle \\
&= \langle \mathcal{K}_x - \mathcal{K}'_x, \mathcal{K}_x \rangle - \langle \mathcal{K}_x - \mathcal{K}'_x, \mathcal{K}'_x \rangle \\
&= (\mathcal{K}_x - \mathcal{K}'_x)(x) - (\mathcal{K}_x - \mathcal{K}'_x)(x) \\
&= 0.
\end{aligned}$$

Hence, $\mathcal{K}_t = \mathcal{K}'_t$, i.e for every $x \in \mathcal{D}$, $\mathcal{K}_t(x) = \mathcal{K}'_t(x)$. But since $\mathcal{K}_t(x) = \mathcal{K}(x, t)$, this would mean $\mathcal{K}(x, t) = \mathcal{K}'(t, x)$, i.e $\mathcal{K} \equiv \mathcal{K}'$. This provides the proof of the lemma. $\square$

Proof of Lemma 2.3

$$\begin{aligned}
\sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j \mathcal{K}(u_i, u_j) &= \sum_{i=1}^{n} \sum_{j=1}^{n} c_i c_j \langle \mathcal{K}_{u_i}, \mathcal{K}_{u_j} \rangle_{\mathcal{H}} \\
&= \langle \sum_{i=1}^{n} c_i K_{u_i}, \sum_{j=1}^{n} c_j K_{u_j} \rangle_{\mathcal{H}} \\
&= \langle \sum_{i=1}^{n} c_i K_{u_i}, \sum_{i=1}^{n} c_i K_{u_i} \rangle_{\mathcal{H}} \geq 0.
\end{aligned}$$

$\square$

Appendix B : Univariate KLE from a GBF

Here we are interested in estimating a covariance kernel $C(s, t) : s, t \in \mathcal{D}$ using the expansion method from a GBF. Assume that the covariance kernel is evaluated at n points $\{t_1, \ldots, t_n\}$ from $\mathcal{D}$. We denote the i-th KL eigenfunction as $\psi_i(\cdot)$ and the i-th eigenvalue as λ_i. These eigenfunctions and eigenvalues will be estimated from GBFs. We denote the j-th GBF as $\theta_j(\cdot)$. The KL eigenfunction is obtained by projecting the GBFs linearly as below

$$\psi_i(t) = \sum_{j=1}^{M} F_{ij}\theta_j(t)$$

$$\Rightarrow \mathbf{\Psi} = \mathbf{\Theta F}. \tag{7}$$

Let $\mathbf{\Psi}$ be the $n \times K$ matrix of KL eigenfunctions, where the $i - j$th element is $\psi_j(i)$. Similarly, the $n \times K$ GBF matrix $\mathbf{\Theta}$ is similarly defined from $\theta_j(i)$s. Denote the j-th row of $\mathbf{\Psi}$ as the row vector $\boldsymbol{\psi}_{j.}$ and similarly define $\boldsymbol{\theta}_{j.}$. Using the Fredholm equation, we hae

$$\int_s \mathbf{C}(s,t)\psi_i(s)ds = \lambda_i\psi_i(t)$$

$$\Rightarrow \int_s \mathbf{C}(s,t){\boldsymbol{\psi}_{s.}}^T ds = \mathbf{\Lambda}{\boldsymbol{\psi}_{t.}}^T$$

$$\Rightarrow \int_s \mathbf{C}(s,t)\mathbf{F}^T{\boldsymbol{\Theta}_{s.}}^T ds = \mathbf{\Lambda F}^T{\boldsymbol{\Theta}_{t.}}^T$$

$$\Rightarrow \mathbf{F}^T \int_s \mathbf{C}(s,t){\boldsymbol{\Theta}_{s.}}^T ds = \mathbf{\Lambda F}^T{\boldsymbol{\Theta}_{t.}}^T$$

$$\Rightarrow \mathbf{F}^T \int_s \mathbf{C}(s,t){\boldsymbol{\Theta}_{s.}}^T\boldsymbol{\Theta}_{t.}ds\mathbf{F} = \mathbf{\Lambda F}^T{\boldsymbol{\Theta}_{t.}}^T\boldsymbol{\Theta}_{t.}$$

$$\Rightarrow \mathbf{F}^T \int_s \int_t \mathbf{C}(s,t){\boldsymbol{\Theta}_{s.}}^T\boldsymbol{\Theta}_{t.}ds\,dt = \mathbf{\Lambda F}^T \int_t {\boldsymbol{\Theta}_{t.}}^T\boldsymbol{\Theta}_{t.}dt.$$

Denote $\mathbf{A}^T = \int_s \int_t \mathbf{C}(s,t){\boldsymbol{\Theta}_{s.}}^T\boldsymbol{\Theta}_{t.}ds\,dt$, $\mathbf{B}^T = \int_t {\boldsymbol{\Theta}_{t.}}^T\boldsymbol{\Theta}_{t.}dt$. Then, we get the following

$$\mathbf{F}^T\mathbf{A}^T = \mathbf{\Lambda B}^T\mathbf{F}^T$$

$$\Rightarrow \mathbf{AF} = \mathbf{BF\Lambda}.$$

The above problem is the well-known generalized eigenvalue problem, which is equivalent to finding the eigenvalues of $\mathbf{B}^{-1}\mathbf{A}$. This gives the solution of $\mathbf{F}$ and $\mathbf{\Lambda}$, which is now plugged in (7) to get the KL eigenfunctions. The eigenvalues are the diagonal elements of $\mathbf{\Lambda}$. This completes the KL expansion of the covariance matrix $\mathbf{C}$.

As the choice of GBFs, the `rcage` package considers Gaussian, Bisquare or Wendland basis functions. The R package `funData` (Happ-Kurz 2020), `mfpca` (Happ and Greven 2018) uses similar basis functions, but also includes the spline basis, cosine basis, radial basis, Fourier basis, and others. In our simulated example, we have tried different basis functions including Fourier, Haar, Legendre polynomials, radial basis functions, and decided that the Legendre polynomials worked well for the choice of GBFs in the examples presented here. Legendre polynomials are orthonormal families of basis functions defined over $[-1, 1]$ and given by

$$\mathbb{P}_1(t) = \frac{1}{\sqrt{2}}$$

$$\mathbb{P}_2(t) = \frac{3}{\sqrt{2}}t$$

$$\mathbb{P}_{k+1}(t) = \sqrt{\frac{2k-1}{2}}\left[\frac{2k-1}{k}t\mathbb{P}_k(t) - \frac{k-1}{k}\mathbb{P}_{k-1}(t)\right].$$

$$\vdots$$

Using recursion, it is straight-forward to show that $\int_{-1}^{1}\int_{-1}^{1}\mathbb{P}_i(s)\mathbb{P}_j(t)\,ds\,dt = \delta_{ij}$. In particular, when K many GBFs are chosen, the GBF matrix $\boldsymbol{\Theta}$ becomes an $n \times K$ matrix with the (i,j)-th element $\theta_{i,j} = \mathbb{P}_j(t_i)$.

Appendix C : Univariate Bayesian KLE

Suppose that we have n observations $\boldsymbol{Y} = (Y_1, \ldots, Y_n)$ at indices $t_1, \ldots, t_n$ from a stochastic process $\mathcal{Y}_t : t \in \mathcal{D}$ with covariance kernel $\boldsymbol{C}(s,t)$, which is unknown. In this case we use the following hierarchical model

$$\begin{aligned}
\boldsymbol{Y} &= \boldsymbol{\Psi}\boldsymbol{\eta} + \boldsymbol{\epsilon} \\
\boldsymbol{\Psi} &= \boldsymbol{\Theta}\boldsymbol{F} \\
\text{Cov}(\boldsymbol{\eta}) &= \boldsymbol{\Lambda} = \text{diag}(\lambda_1 \geq \lambda_2 \geq \cdots) \\
\boldsymbol{\epsilon} &\sim \mathcal{N}_n(0, \tau^2\boldsymbol{I}).
\end{aligned} \tag{8}$$

One can start with any choice of GBF $\boldsymbol{\Theta}$ here (as stated in Appendix B), where the (i,j)-th element of $\boldsymbol{\Theta}$ is the j-th GBF corresponding to the i-th observation, i.e., $\theta_j(t_i)$. To compute $\boldsymbol{\Psi}$ from $\boldsymbol{\Theta}$, we need first define $\boldsymbol{W}$ as the matrix with i,j th element as $W_{ij} = \int_t \theta_i(t)\theta_j(t)dt$. Then, the Fredholm integral equation $\boldsymbol{F}^T\boldsymbol{W}\boldsymbol{F} = \boldsymbol{I}$ must be satisfied for $\boldsymbol{\Psi}$ to be orthonormal. Let $\boldsymbol{Q}$ be the Cholesky decomposition of $\boldsymbol{W}^{-1}$ and let $\boldsymbol{G}$ be any orthonormal matrix of proper dimension. Then, $\boldsymbol{F} = \boldsymbol{Q}\boldsymbol{G}$ satisfies the Fredholm integral equation since

$$\boldsymbol{F}^T\boldsymbol{W}\boldsymbol{F} = \boldsymbol{G}^T\boldsymbol{Q}^T\boldsymbol{Q}^{-T}\boldsymbol{Q}^{-1}\boldsymbol{Q}\boldsymbol{G} = \boldsymbol{G}^T\boldsymbol{G} = \boldsymbol{I}.$$

The first stage of the model in (8) now becomes

$$\begin{aligned}
\boldsymbol{Y} &= \boldsymbol{\Psi}\boldsymbol{\eta} + \boldsymbol{\epsilon} \\
&= \boldsymbol{\Theta}\boldsymbol{F}\boldsymbol{\eta} + \boldsymbol{\epsilon} \\
&= \boldsymbol{\Theta}\boldsymbol{Q}\boldsymbol{G}\boldsymbol{\eta} + \boldsymbol{\epsilon} \\
&= \boldsymbol{\Phi}\boldsymbol{v} + \boldsymbol{\epsilon} \quad \left(\text{where } \boldsymbol{\Phi} = \boldsymbol{\Theta}\boldsymbol{Q}, \boldsymbol{v} = \boldsymbol{G}\boldsymbol{\eta}\right).
\end{aligned}$$

Note that $\boldsymbol{\Phi}$ can be pre-calculated from the GBFs $\boldsymbol{\Theta}$ and $\text{Cov}(\boldsymbol{v}) = \boldsymbol{G}\boldsymbol{\Lambda}\boldsymbol{G}^T = \boldsymbol{\Sigma}$). We need to specify a prior on $\boldsymbol{\Sigma}$ to complete the full Bayesian approach. The prior used on $\boldsymbol{\Sigma}$ is the so-called Moran's I (MI) prior developed in (Bradley et al. 2015, 2017), which is computed as

$$\Sigma = \sigma^2 \left[R^{-1} \mathcal{A}^{(+)} \left(Q'(I - A)Q \right) R^{-T} \right]^{-1} = \sigma^2 M^{-1},$$

where A is the adjacency matrix of the locations, (Q, R) is the QR decomposition of the basis matrix Φ and $\mathcal{A}^{(+)}(\cdot)$ is the best positive definite approximation of a matrix as explained in Higham (1988).

In addition, σ^2 and τ^2 are given inverse gamma priors with $a_\sigma = 10, b_\sigma = 10$, $a_\tau = 10, b_\tau = 1$. The remaining part of the above expression is known once we compute Φ and hence the prior specification on σ^2 completes the prior specifications. We now compute the posteriors of the unknown parameters, specifically the posterior predictive distribution of v. The complete hierarchical model specification is given by

$$Y = \Phi v + \epsilon$$
$$\epsilon | \tau^2 \sim \mathcal{N}_n(0, \tau^2 I)$$
$$v | \sigma^2 \sim \mathcal{N}_K \left(0, \sigma^2 \left[R^{-1} \mathcal{A}^{(+)} Q'(I - A)QR^{-T} \right]^{-1} \right)$$
$$\sigma^2 \sim IG(a_\sigma, b_\sigma)$$
$$\tau^2 \sim IG(a_\tau, b_\tau).$$

The full-conditional distributions are then given by:

$$\pi(v, \tau^2, \sigma^2 \mid Y) \propto (\tau^2)^{-(1+a_\tau+\frac{n}{2})} (\sigma^2)^{-(1+a_\sigma+\frac{K}{2})}$$
$$\exp - \frac{(Y - \Phi v)^T (Y - \Phi v) + 2b_\tau}{2\tau^2} \exp - \frac{v^T M v + 2b_\sigma}{2\sigma^2}$$
$$\pi(\tau^2 \mid \cdot) \sim IG \left(\frac{n}{2} + a_\tau, \frac{(Y - \Phi v)^T (Y - \Phi v) + 2b_\tau}{2} \right)$$
$$\pi(\sigma^2 \mid \cdot) \sim IG \left(\frac{K}{2} + a_\sigma, \frac{v^T M v + 2b_\sigma}{2} \right)$$
$$\pi(v \mid \cdot) \sim \mathcal{N}_K \left((\frac{1}{\tau^2} \Phi^T \Phi + \frac{1}{\sigma^2} M)^{-1} \frac{1}{\tau^2} \Phi^T Y, (\frac{1}{\tau^2} \Phi^T \Phi + \frac{1}{\sigma^2} M)^{-1} \right).$$

Posterior samples of the parameters are obtained after running the MCMC for 5000 iterations with a burn-in of 1000 and then thinned the chains, keeping every 5-th observation. Convergence was assessed through visual inspection of the sample chains, with no lack of convergence detected. After getting the posterior samples, one can finally construct the KL eigenfunctions from the posterior samples of vs and compute the covariance matrix $\widehat{\Sigma} = \widehat{G} \widehat{\Lambda} \widehat{G}^T$. Now, we can get the KL eigenfunctions under this prior by computing $\widehat{\Psi} = \Phi \widehat{G}$.

Note that the reconstructed eigenfunctions and eigenvalues may depend heavily on the priors chosen. In our examples, we found that the performance of the algorithm is typically worse than the method when the true covariance function is known, as expected. However, we note that the true covariance function is never known in practice and this provides an elegant way to perform the KLE for the data

when one has no knowledge apart from only one realization of the stochastic process. We also note that in cases where more than one replicate is available, one can estimate the covariance matrix from the data and then fall back to solving for the unknown coefficients through Fredholm equation directly.

Appendix D : Multiple Separate GBFs

For the multiple separate KLE problem, one needs to proceed as in the univariate case for each of the individual process. Consider a multivariate stochastic process $X_t = (X_{t1}, \ldots, X_{tK})$ with a block covariance matrix

$$C = \begin{bmatrix} C_{1,1} & \ldots & C_{1,K} \\ \vdots & \vdots & \vdots \\ C_{K,1} & \ldots & C_{K,K} \end{bmatrix},$$

where each block matrix C_{jk} is a $n \times n$ covariance matrix with elements $C_{jk}(s, t) = Cov(X_{js}, X_{kt})$. We need to construct individual process specific KLEs first.

Consider the j-th process $\{X_{tj} : t \in \mathcal{D}\}$ with covariance matrix as the j-th diagonal block C_{jj}. Denote the corresponding GBF as the matrix $\Theta^{(j)}$ with the (t, i)-th element given by $\theta_i^{(j)}(t)$. Similarly, the i-th eigenvalue and eigenfunction for the j-th process are then given by $\psi_i^{(j)}(\cdot)$ and $\lambda_i^{(j)}$, respectively. Define $\Psi^{(j)}$ as the vector with the t, i-th element to be $\psi_i^{(j)}(\cdot)$. These KL eigenfunctions are obtained as linear projections from $\Theta^{(j)}$ using $\Psi^{(j)} = \Theta^{(j)} F^{(j)}$.

From the Fredholm integral equation, $F^{(j)}$ is estimated by using the following derivation

$$\int_s C_j(s, t) \psi_i^{(j)}(s) ds = \lambda_i^{(j)} \psi_i^{(j)}(t).$$

Proceeding similarly as in the univariate case, one gets the matrix $A^{(j)}$ and $B^{(j)}$ with (k, l)-th element as $\int_s \int_t C_{jj}(s, t) \theta_i^{(j)}(s) \theta_j^{(j)}(t) \, ds \, dt$ and $\int_t \theta_i^{(j)}(t) \theta_j^{(j)}(t) \, dt$ respectively. To get the KLE, the generalized eigenvalue problem $A^{(j)} F^{(j)} = B^{(j)} F^{(j)} \Lambda^{(j)}$ is solved, where $\Lambda^{(j)}$ and $F^{(j)}$ are the eigenvalues and eigenvectors of $B^{(j)-1} A^{(j)}$, the diagonals of $\Lambda^{(j)}$ are the KL eigenvalues, and $\Psi^{(j)} = \Theta^{(j)} F^{(j)}$ are the KL eigenfunctions.

After solving for each univariate process, we need to obtain the KLE which we get by computing the following

$$\mathcal{K}_{ij}^{ml} = \frac{1}{\sqrt{\lambda_m^{(1)} \lambda_l^{(2)}}} \int_s \int_t C_{ij}(s, t) \psi_m^{(i)}(s) \psi_l^{(j)}(t) \, ds \, dt.$$

Now as described in Sect. 3.6, we can define the block matrix K, where the (i, j)-th

block is the matrix $\boldsymbol{K}_{ij}$ with m, l-th element $\mathcal{K}_{ij}^{ml}$. The diagonal blocks $\boldsymbol{K}_{ii}$'s are defined to be identity matrices. We then collect the variable-wise expansion coefficients into a large vector as in $A = \left(\alpha_{11}, \alpha_{21}, \ldots, \alpha_{1K}, \alpha_{2K}, \ldots\right)'$. To get the multivariate KLE, we obtain the Cholesky decomposition $\boldsymbol{K} = \boldsymbol{R}\boldsymbol{R}^T$ and then define $\widetilde{A} = \boldsymbol{R}^{-1}A$. $\widetilde{A}$ thus becomes uncorrelated since $\mathbb{E}\left[\widetilde{A}\widetilde{A}^T\right] = \boldsymbol{R}^{-1}\mathbb{E}\left[AA^T\right]\boldsymbol{R}^{-T} = \boldsymbol{R}^{-1}\boldsymbol{K}\boldsymbol{R}^{-T} = \mathbf{I}$. Similarly the univariate eigenfunctions $\phi_m^{(i)}(.)$ can be collected in a vector to define a multivariate eigenfunction $\Phi_m^{(i)}(.)$ and then they can be made orthonormal by a similar transformation as in $\widetilde{\Phi}(\cdot) = \Phi(\cdot)\boldsymbol{R}$. Thus, we get the orthonormal eigenfunctions and uncorrelated coefficients and this gives the multivariate KLE from the univariate KLEs of each process.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Aronszajn N (1950) Theory of reproducing Kernels. Trans Am Math Soc 68(3):337–404

Barat P, Amitava R (1998) Modification of Karhunen-Loève transform for pattern recognition. Sadhana 23(4):341–350

Levy BC (2008) Karhunen-Loève expansion of Gaussian processes. Principles of Signal Detection and Parameter Estimation, Springer, Boston, MA

Betz W, Papaioannou I, Straub D (2014) Numerical methods for the discretization of random fields by means of the Karhunen-Loève expansion. Comput Methods Appl Mech Eng 271:109–129

Boente G, Fraiman R (2000) Kernel-based functional principal components. Stat Probab Lett 48(4):335–345

Bradley JR, Holan Scott H, Wikle Christopher K (2015) Multivariate Spatio-temporal models for high-dimensional areal data with application to Longitudinal Employer-Household Dynamics. Ann Appl Stat 9(4):1761–1791

Bradley JR, Wikle Christopher K, Holan Scott H (2017) Regionalization of multiscale spatial processes by using a criterion for spatial aggregation error. J Royal Stat Soc: Series B (Statistical Methodology) 79(3):815–832

Bradley Jonathan R, Wikle Christopher K, Holan Scott H, Holloway Shannon T (2021) Rcage: Regionalization of Multiscale Spatial Processes, R package version 1.1

Carmeli C, De Vito E, Toigo A (2006) Vector valued reproducing kernel Hilbert spaces of integrable functions and mercer theorem. Anal Appl 4(04):377–408

Castrillon-Candas JE, Liu D, Kon M (2021) Stochastic functional analysis with applications to robust machine learning. arXiv preprint arXiv:2110.01729

Chiou J-M, Chen Y-T, Yang Y-F (2014) Multivariate functional principal component analysis: a normalization approach. Statistica Sinica 1:1571–1596

Heyrim C, Venturi D, Karniadakis GE (2013) Karhunen-Loève expansion for multi-correlated stochastic processes. Probab Eng Mech 34:157–167

Cressie N, Wikle CK (2011) Statistics for spatio-temporal data. John Wiley Sons, UK

Dai H, Zheng Z, Ma H (2019) An explicit method for simulating non-Gaussian and non-stationary stochastic processes by Karhunen-Loève and polynomial chaos expansion. Mech Syst Signal Process 115:1–13

Davenport WB, Root WL et al (1958) An introduction to the theory of random signals and noise, vol 159. McGraw-Hill, New York

De Vito E, Umanità V, Villa S (2013) An extension of mercer theorem to matrix-valued measurable kernels. Appl Comput Harmon Anal 34(3):339–351

Dong D, Fang P, Bock Y, Webb F, Prawirodirdjo L, Kedar S, Jamason P (2006) Spatiotemporal filtering using principal component analysis and Karhunen-Loève expansion approaches for regional GPS network analysis. J Geophys Res: Solid Earth. https://doi.org/10.1029/2005JB003806

Flaxman S, Teh YW, Sejdinovic D (2017) Poisson intensity estimation with reproducing Kernels. In: Artificial Intelligence and Statistics, PMLR, pp 270–279

Fontanella Lara, Ippoliti Luigi (2012) Karhunen-Loève expansion of temporal and spatio-temporal processes. In: Handbook of Statistics, Elsevier, vol. 30, pp 497–520

Freiberger W, Grenander U (1965) On the formulation of statistical meteorology. Revue de l'Institut International de Statistique, pp 59–86

Ghanem Roger G, Spanos Pol D (2003) Stochastic finite elements: a spectral approach. Courier Corporation

Golub GH, Van Loan CF (1996) Matrix Computations. Johns Hopkins University Press; 3rd edition

Greengard P, O'Neil M (2021) Efficient reduced-rank methods for Gaussian processes with eigenfunction expansions. arXiv preprint arXiv:2108.05924

Grigoriu M (1993) On the spectral representation method in simulation. Probab Eng Mech 8(2):75–90

Ramón G, Carlos RJ, Valderrama MJ (1992) On the numerical expansion of a second order stochastic process. Appl Stochastic Models Data Anal 8(2):67–77

Abdel H, Jolliffe IT, Stephenson DB (2007) Empirical orthogonal functions and related techniques in atmospheric science: a review. Int J Climatol: A J Royal Meteorol Soc 27(9):1119–1152

Happ C, Greven S (2018) Multivariate functional principal component analysis for data observed on different (dimensional) domains. J Am Stat Assoc 113(522):649–659

Happ-Kurz C (2020) Object-oriented software for functional data. J Stat Softw 93(5):1–38

Higham NJ (1988) Computing a nearest symmetric positive semidefinite matrix. Linear Algebra Appl 103:103–118

Holmstrom I (1977) On empirical orthogonal functions and variational methods. In: Proceedings of a workshop on the use of empirical orthogonal functions in meteorology, European Center for Medium Range Forecast Reading, UK, pp 8–20

Hotelling H (1933) Analysis of a complex of statistical variables into principal components. J Educ Psychol 24(6):417

Hu J (2013) Optimal low rank model for multivariate spatial data. PhD thesis, Purdue University

Juan H, Zhang H (2015) Numerical methods of Karhunen-Loève expansion for spatial data. Econom Quality Contr 30(1):49–58

Huang J, Griffiths DV, Lyamin AV, Krabbenhoft K, Sloan SW (2014) Discretization errors of random fields in finite element analysis. Appl Mech Mater 553: 405–409

Huang SP, Quek ST, Phoon KK (2001) Convergence study of the truncated karhunen-loeve expansion for simulation of stochastic processes. Int J Numer Meth Eng 52(9):1029–1043

Jacques J, Preda C (2014) Model-based clustering for multivariate functional data. Comput Stat Data Anal 71:92–106

Jin S (2014) Gaussian processes: Karhunen-Loève expansion, small ball estimates and applications in time series models

Johnson ME, Moore LM, Donald Y (1990) Minimax and maximin distance designs. J Stat Plan Infer 26(2):131–148

Jolliffe IT, Cadim J (2016) Principal component analysis: a review and recent developments. Philos Trans Royal Soc A: Math, Phys Eng Sci 374(2065):20150202

Kailath T (1971) RKHS approach to detection and estimation problems-i: deterministic signals in Gaussian noise. IEEE Trans Inf Theory 17(5):530–549

Karhunen K (1946) Zur spektraltheorie stochastischer prozesse. Ann Acad Sci Fennicae, AI, 34

Kirby M, Sirovich L (1990) Application of the Karhunen-Loeve procedure for the characterization of human faces. IEEE Trans Pattern Anal Mach Intell 12(1):103–108

Leeds William B, Wikle Christopher K, Fiechter Jerome F (2014) Emulator-assisted reduced-rank ecological data assimilation for nonlinear multivariate dynamical Spatio-temporal processes. Stat Methodol 17:126–138

Li CF, Feng YT, Owen DRJ, Li DF, Davis IM (2008) A Fourier-Karhunen-Loève discretization scheme for stationary random material properties in sfem. Int J Numer Meth Eng 73(13):1942–1965

Li CC, Der Kiureghian A (1993) Optimal discretization of random fields. J Eng Mech 119(6):1136–1154

Li Heng, Zhang Dongxiao (2013) Stochastic representation and dimension reduction for non-Gaussian random fields: review and reflection. Stoch Env Res Risk Assess 27(7):1621–1635

Li LB, Phoon KK, Quek ST (2007) Comparison between Karhunen-Loève expansion and translation-based simulation of non-Gaussian processes. Comput Struct 85(5–6):264–276

Lumley JL (1967) The structure of inhomogeneous turbulent flows. Atmos Turbul Radio Wave Prop

Lóeve MM (1955) Probability Theory. Van Nostrand, Princeton, N.J

Monahan AH, Fyfe JC, Ambaum MH, Stephenson DB, North GR (2009) Empirical orthogonal functions: the medium is the message. J Clim 22(24):6501–6514

Mulaik SA (2009) Foundations of factor analysis. CRC press

Obled Ch, Creutin JD (1986) Some developments in the use of empirical orthogonal functions for mapping meteorological fields. J Appl Meteorol Climatol 25(9):1189–1204

Phoon KK, Huang HW, Quek ST (2004) Comparison between Karhunen-Loève and wavelet expansions for simulation of Gaussian processes. Comput Struct 82(13–14):985–991

Phoon KK, Huang SP, Quek ST (2002) Implementation of Karhunen-Loève for simulation using a wavelet-Galerkin scheme. Probab Eng Mech 17(3):293–303

Poirion F (2016) Karhunen-Loève expansion and distribution of non-Gaussian process maximum. Probab Eng Mech 43:85–90

Rahman S(2018) A Galerkin isogeometric method for Karhunen-Loève approximation of random fields. Comput Methods Appl Mech Eng 338:533–561

Ramsay JO, Silverman BW (2002) Applied functional data analysis: methods and case studies, vol. 77. Springer

Rasmussen CE (2003) Gaussian processes in Machine Learning. In: Summer school on machine learning, Springer, pp 63–71

Riesz F, Nagy BS (1990) Functional analysis. Dover Publications, Inc., NewYork. First published 1176, 3(6):35, 1955

Schaback Robert (1999) Native Hilbert spaces for radial basis functions i. New developments in approximation theory

Masanobu Shinozuka, George Deodatis (1991) Simulation of stochastic processes by spectral representation. Appl Mech Rev 44(4):191–204

Sirovich L (1987) Turbulence and the dynamics of coherent structure. part i, ii, iii. Quart Appl Math, 3:583

Small CG, McLeish DL (2011) Hilbert space methods in probability and statistical inference, 920. John Wiley Sons, UK

Sørbye SH, Rue H (2014) Scaling intrinsic Gaussian markov random field priors in spatial modelling. Spatial Statistics 8:39–51

Spanos PD, Beer M, Red-Horse J (2007) Karhunen-Loève expansion of stochastic processes with a modified exponential covariance kernel. J Eng Mech 133(7):773–779

Spanos PD, Roger G (1989) Stochastic finite element expansion for random media. J Eng Mech 115(5): 1035–1053

Steinwart I, Christmann A (2008) Support vector machines. Springer Science & Business Media

Stewart Gilbert W (1993) On the early history of the singular value decomposition. SIAM Rev 35(4):551–566

Berlinet C, Thomas-Agnan A (2004) Reproducing kernel hilbert spaces in probability and statistics. Springer, Boston, MA

Van Trees, Harry L (2004) Detection, estimation, and modulation theory, part I: detection, estimation, and linear modulation theory. John Wiley and Sons

Wahba G (1990) Spline models for observational data. SIAM

Wang L (2008) Karhunen-Loeve expansions and their applications. London School of Economics and Political Science, UK

Wikle CK (2010) Low-rank representations for spatial processes. In: M. Fuentes A. E. Gelfand, P. J. Diggle and P. Guttorp, (Eds.) Handbook of Spatial Statistics, Handbook of Spatial Statistics, vol 30, pp 107–118. Chapman and Hall-CRC, UK

Wikle CK, Milliff RF, Herbei R, Leeds WB (2013) Modern statistical methods in oceanography: a hierarchical perspective. Stat Sci 1:466–486

Wikle CK, Zammit-Mangion A, Cressie N (2019) Spatio-temporal Statistics with R. Chapman and Hall/CRC

Xiu D (2009) Fast numerical methods for stochastic computations: a review. Commun Comput Phys 5(2–4): 242–272

Yamashita Y, Ikeno Y, Ogawa H (1998) Relative Karhunen-Loève transform method for pattern recognition. In: Proceedings 14th international conference on pattern recognition (Cat. No. 98EX170), vol. 2, IEEE, pp 1031–1033

Yeung Y-H, Barajas-Solano DA, Tartakovsky AM (2021) Physics-informed machine learning method for large-scale data assimilation problems. arXiv preprint arXiv:2108.00037

Zapata J, Oh S-Y, Petersen A (2019) Partial separability and functional graphical models for multivariate Gaussian processes. arXiv preprint arXiv:1910.03134

Zhang J, Ellingwood B (1994) Orthogonal series expansions of random fields in reliability analysis. J Eng Mech 120(12):2660–2677

Zhao Z, Li B, Liu W, Yuan H, Song C (2020) A qualitative recognition method based on Karhunen-Loève decomposition for near-infrared (NIR) and mid infrared (MIR) spectroscopy analysis. Anal Methods 12(1):112–121

Zheng Z, Dai H (2017) Simulation of multi-dimensional random fields by Karhunen-Loève expansion. Comput Methods Appl Mech Eng 324:221–247