Biometrics , 1–26

Functional Principal Component Analysis for Multiple Variables on Different Riemannian Manifolds

Haixu Wang* and Jiguo Cao**

Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, B.C., Canada

**email:* haixuw@sfu.ca

***email:* jiguoc@sfu.ca

Summary:

Functional principal component analysis (FPCA) is a very important dimension reduction tool for functional data analysis. The conventional FPCA procedures can only model multiple functional variables on the same Riemannian manifold. We propose a new multivariate FPCA method to jointly analyze multiple variables on different Riemannian manifolds. We introduce a mapping procedure to project these variables to the same Euclidean space. The mapped variables are able to preserve temporal variations from the original ones, and their covariance functions can be defined and estimated as the same as those of Euclidean variables. A normalization method is also employed to mitigate the effects of different units of variables and scales of variation. The proposed method allows us to reconstruct curves for non-Euclidean variables with missing observations by using functional principal components and their scores. Several simulation studies are conducted to investigate the finite sample performance of the proposed method. We demonstrate our method by jointly analyzing the wind direction and speed data in three Canadian cities. It shows that the proposed multivariate FPCA method is a versatile tool and provides useful insights for variables on different manifolds.

Key words: Curve reconstruction; Dimension reduction; Multivariate analysis; Riemannian functional variables; Unsupervised learning.

This paper has been submitted for consideration for publication in *Biometrics*

1. Introduction

Functional data analysis (FDA) is a class of statistical methodology that is developed to analyze data recorded as curves, images, or other objects over a continuum, usually time, in scientific areas such as econometrics, ecology and medical science (Ramsay and Silverman, 2005; Vieu and Ferraty, 2006). These data can be viewed as functions of time, spatial locations or other continuum, hence these data are called functional data. One fundamental FDA tool is functional principal component analysis (FPCA). The main objective of FPCA is to explore the major modes of variation among functional data, which are represented by functional principal components (FPCs). Usually, a few FPCs can explain most variation among functional data. The functional data can then be expanded by a linear combination of FPCs. Then the infinite-dimensional functional data can be projected to the finite-dimensional space defined by FPCs. Therefore, FPCA is an important tool for the dimension reduction of functional data.

The asymptotic properties of FPCA have been investigated by Dauxois et al. (1982) through applying principal component analysis for infinite-dimensional objects. Later, Hall and Hosseini-Nasab (2005) and Hall et al. (2006) conduct thorough studies on the theories of FPCA. Following the groundwork, the studies on FPCA have expanded rapidly where extensions and applications are proposed. Lin et al. (2015) provides additional interpretations of FPCs with penalization techniques. Nie et al. (2018) combines information from supervised tasks for enhancing the prediction capabilities of FPCs. FPCA can also facilitate other statistical models. For example, Yao et al. (2005) used FPCA to estimate functional linear models, and Müller and Stadtmüller (2005) extended that to the generalized functional linear models. We can also use FPCA results to cluster and classify random trajectories as illustrated in Peng and Müller (2008) and Müller (2005), respectively. It is common to encounter the case where individual curves are not fully observed. To overcome the missing

data problem, Yao et al. (2005) proposed to pull available observations across all curves for estimating covariance functions and subsequently FPCs, and used conditional expectation to reconstruct individual curves based on the joint distribution of observations and FPC scores.

While univariate FPCA has been comprehensively studied and applied, multivariate FPCA becomes a pressing subject when several functional variables are simultaneously observed in real data applications. Existing methodologies for multivariate functional principal component analysis (MFPCA) include the approach discussed in Ramsay and Silverman (2005) that circumvents the multi-dimensionality by concatenating multivariate functions into a single one. Standard univariate FPCA can then be applied to the combined functional variables. Berrendero et al. (2011) proposed to do the regular PCA on the curve observations at the same time, and then connected the estimated principal components over time as the functional principal components. Both of the aforementioned methods assume that the functions are observed in the same domain. Happ and Greven (2018) investigates the case that functions are observed on different domains. Chiou et al. (2014) studied how to minimize the effects of the difference in units between functional variables through a normalization step.

A special challenge arises for the aforementioned MFPCA methods when functional variables are not uniformly defined in the same space. That is, each functional variable might reside in the space with a different metric system. One particular example of the challenge is the analysis of wind data. The most familiar variables of wind data consist of direction and speed. A demonstration of wind data is shown in Figure 1. The direction variable is defined on a 1-dimensional sphere, whereas the speed variable belongs to the usual Euclidean space. Naive joint analysis, treating speed and direction as two variables in the same Euclidean space, does not suffice to capture the covariances between these two variables nor acknowledge

the topological differences in their domains. Existing analyses of wind data depend on directional statistics that aim to use some specific distributions to describe the data (Mardia, 1972; Masseran et al., 2013; Belu and Korachin, 2013). There is little effort put into the joint analysis of these two variables. The existing MFPCA methods cannot be directly applied to wind data or variables in different topological spaces.

[Figure 1 about here.]

Before proposing an MFPCA method for variables on different topological spaces, it is interesting to formalize an appropriate FPCA method for variables on some smooth manifolds that are not limited to a Euclidean space. Bhattacharya and Patrangenaru (2003) provided large sample theories for variables defined on manifolds and examined the mapping of variables between manifolds and vector spaces. Later, Dai and Müller (2018) investigated how to perform FPCA for infinite-dimensional objects on Riemannian manifolds. These works provide instrumental help for developing a method to overcome the previous challenge, but Dai and Müller (2018) does not consider the case that functional variables are defined on different manifolds.

This article proposes an FPCA method for multivariate functional data where variables can be defined in their individual topological spaces. It serves as an unsupervised learning method for multivariate functional variables and identifies important features for doing dimension reduction and establishing statistical models. The features are obtained through the decomposition of covariance functions which are not directly available from the existing MFPCA methods. Our method overcomes the particular challenge in estimating covariance functions when variables are not uniformly defined in the same topological space. We utilize a mapping procedure that will allow variables to be analyzed in the same Euclidean space. Furthermore, we employ the normalization introduced in Chiou et al. (2014) to mitigate the effects of different units of variables and scales of variation. This ensures that the

estimation of cross-covariance functions will not prune to any particular variable. By doing the decomposition of the obtained covariance functions, we can capture some interpretable features of the data and pursue further analyses.

We examine the finite sample performance of the proposed method in two simulation studies. The first study is conducted under an ideal setting where individual curves are fully observed. The estimation of model components is consistent as expected over a finite sample. The second study investigates whether the method produces reliable results given various sparsity settings in the data. The study has demonstrated that estimations are not affected by the sparsity level. As aforementioned, wind data is an example that challenges the conventional MFPCA methods. We will present an application of the proposed method in extracting features of wind direction and speed for three cities in Canada. In the application, distinguishable features are captured by the obtained FPCs, and individual curves are well recovered even if they are not fully observed. The computing R codes for simulation studies and the application are available at `https://github.com/caojiguo/MFPCA`.

The article is organized as follows. Section 2 provides the detailed methodology of the proposed MFPCA for data on different manifolds. Section 3 includes simulation studies to investigate the finite-sample performance of the proposed method under various sparsity settings. An application on the wind data is presented in Section 4. The paper is concluded in Section 5.

2. FPCA for Multivariate Riemannian Functional Variables

Let $\boldsymbol{X}(t)$ be the set of variables defined in a Euclidean space and $\boldsymbol{W}(t)$ contain other variables on some Riemannian manifolds. The major obstacle in doing FPCA for $(\boldsymbol{X}(t), \boldsymbol{W}(t))$ is caused by the non-Euclidean topology of $\boldsymbol{W}(t)$, because the covariance function of $\boldsymbol{W}(t)$ cannot be conveniently defined in the same way as $\boldsymbol{X}(t)$. The obstacle also prevents us from defining the covariance function between $\boldsymbol{X}(t)$ and $\boldsymbol{W}(t)$ because they do not share the same

Euclidean metric system. Without a proper covariance function, the functional principal components cannot be obtained for analyzing the variation or facilitating the reconstruction of individual curves. Establishing a new metric system for the combined space is difficult and burdensome. Hence, we propose to use a mapping process to circumvent the problems raised by the difference in topology. The mapping process allows all variables to have a common space to analyze and compare variations. As a result, the covariance functions of $(\boldsymbol{X}(t), \boldsymbol{W}(t))$ can be defined and estimated as per usual practice.

Given that the Riemannian manifold $\mathcal{M}$ with metric g admits the Levi-Civita connection and is geodesically complete. Then, any two points $(p, q) \in \mathcal{M}$ can be connected with a length-minimizing geodesic based on the Hopf-Rinow theorem (Chavel, 2006). In the meantime, the manifold grants a mapping, called the exponential map, from its tangent space $T_p\mathcal{M}$ to itself around a point $p \in \mathcal{M}$. The exponential map $\exp_p : T_p\mathcal{M} \to \mathcal{M}$ allows the parameterization of geodesics based on the tangent vectors in $T_p\mathcal{M}$. Within the deformed tangent space on the manifold, there is an inverse mapping function, $\log_p$, for projecting geodesics as tangent vectors in $T_p\mathcal{M}$. If another point $q \in \mathcal{M}$ in the region inside the injectivity radius of $\exp_p$, then there exists a unique geodesic connecting p and q with the minimal length. As a result, the deviation from point q to p can be reparameterized by a tangent vector in $T_p\mathcal{M}$.

The local mapping between the manifold and its tangent spaces motivates the strategy of representing the variation of manifold-valued functions with real vectors in $\mathbb{E}$. It follows that the variation of $\boldsymbol{W}(t)$ can be examined as $\boldsymbol{X}(t)$ in a Euclidean space in which a covariance function is easily defined. Furthermore, the mapping easily concatenates $\boldsymbol{W}(t)$ with $\boldsymbol{X}(t)$ into a single vector for any joint analyses. In this paper, we will assume the existence and uniqueness of the mean function $\boldsymbol{\nu}(t)$ for the Riemannian variable $\boldsymbol{W}(t)$, and they can ensure the mapping procedure feasible for $\boldsymbol{W}(t)$. The mapping specifies a tangent space

rolling along the mean function $\boldsymbol{\nu}(t)$ on $\mathcal{M}$. At any given t, the geodesics connecting any function values to the mean function can be reparametrized by the corresponding tangent vector in $T_{\mathcal{M}}$. Consequently, taking the logarithm map of $\boldsymbol{W}(t)$ produces a functional variable $\boldsymbol{U}(t)$ in $\mathcal{H}$, i.e., $\boldsymbol{U}(t) = \log_{\boldsymbol{\nu}(t)}(\boldsymbol{W}(t))$. The resulting variable $\boldsymbol{U}(t)$ preserves the variation of $\boldsymbol{W}(t)$ and allows us to calculate its covariance function. Any results derived from the covariance function of $\boldsymbol{U}(t)$ can be conveniently mapped for the manifold-valued variable $\boldsymbol{W}(t)$ through the exponential map. For an example of the following Model (1), we can use a truncated approximation of $\boldsymbol{U}(t)$ to obtain a lower-dimensional representation of the infinite-dimensional object $\boldsymbol{W}(t)$.

The mapping strategy not only provides a way to perform FPCA for $\boldsymbol{W}(t)$ but also facilitates the FPCA for $\boldsymbol{X}(t)$ and $\boldsymbol{W}(t)$ jointly. Projecting Riemannian functional variables into $\mathcal{H}$ allows traditional functional analyses to be employed without perplexed by the additional topology. It is also efficient to concatenate $\boldsymbol{U}(t)$ with the functional variable $\boldsymbol{X}(t)$, inherently defined in $\mathcal{H}$, for any joint analyses of the two. In this way, we can simply create a new variable $\boldsymbol{V}(t) = (\boldsymbol{X}(t), \boldsymbol{U}(t))$ and apply multivariate analyses as it is defined in a Euclidean space. Furthermore, it is possible to include more sets of variables defined on various topological spaces as long as all of them possess an intrinsic mapping similar to the exponential-log mapping on Riemannian manifolds.

While the mapping strategy diminishes the interference caused by the topological difference between $\boldsymbol{X}(t)$ and $\boldsymbol{W}(t)$, the differences can exist between $\boldsymbol{X}(t)$ and $\boldsymbol{W}(t)$ and within each set of variables in units of measurements and scales of variation. These differences can lead to the FPCA results overshadowed by any particular variable. The units in $\boldsymbol{W}(t)$ are discarded after undergoing the mapping procedure, and there are no units in $\boldsymbol{U}(t)$ for comparisons with those in $\boldsymbol{X}(t)$ when they are concatenated into the variable $\boldsymbol{V}(t)$. The difference in scales of variations is usually more serious in interfering with the FPCA result because it could lead to

the covariance function being dominated by some variables with greater variability. The same problem also occurs to both the functional principal components and scores of individual curves. This situation is often encountered in multivariate analyses. To ensure that the covariance function of $\boldsymbol{V}(t)$ will not be swayed into variables with greater variability, we can employ a normalization step to balance the discrepancy in variabilities among variables. This normalization step is an extension of the method proposed in Chiou et al. (2014). Note that the method proposed in Chiou et al. (2014) was only applied to traditional functions, while our method is employed on functional variables in different topological spaces for facilitating the subsequent multivariate FPCA. In the rest of the section, we will specify the model that addresses the aforementioned difficulties along with the model estimation procedure for doing FPCA for multiple variables on different Riemannian manifolds.

2.1 *Model Specification*

Let $\boldsymbol{X}(t) = (X_1(t), ..., X_L(t))$ and $\boldsymbol{W}(t) = (W_1(t), ..., W_R(t))$ be two sets of functional variables where each is located in a different topological space. Considering a simple but general scenario, $\boldsymbol{X}(t)$ is defined in a Hilbert space in which each $X_l(t)$ is assumed to be a square-integrable function within $L^2(\tau)$ over an observation interval $\tau = [a, b]$. On the other hand, $\boldsymbol{W}(t)$ consists of functional variables defined on some Riemannian manifolds which can be universally or individually specified. That is, each $W_r(t)$ can be located on its individual manifold $\mathcal{M}_r$ with metric g_r or a single manifold $\mathcal{M}$ with a metric system g for all. In the following sections, we will assume that $W_r(t)$'s are on the same manifold $(\mathcal{M}, g)$, or equivalently, $\boldsymbol{W}(t)$ is a Riemannian functional variable on $(\mathcal{M}, g)$.

The primary objective of FPCA on $(\boldsymbol{X}(t), \boldsymbol{W}(t))$ is to identify the functional principal components that can explain the variation of both sets of functional variables. The objective is also embedded in finding the covariance functions because the functional principal components are obtained by the eigen-decomposition of covariance functions. Subsequent to

identifying the functional principal components, the functional variables can be represented as

$$\begin{aligned} \boldsymbol{X}(t) &= \boldsymbol{\mu}(t) + \sum_{k=1}^{\infty} \xi_k \cdot \{\boldsymbol{\Gamma}_{\boldsymbol{X}} \cdot \boldsymbol{\psi}_k(t)\}, \\ \boldsymbol{W}(t) &= \exp_{\boldsymbol{\nu}(t)}\Big[\sum_{k=1}^{\infty} \xi_k \cdot \{\boldsymbol{\Gamma}_{\boldsymbol{W}} \cdot \boldsymbol{\omega}_k(t)\}\Big], \end{aligned} \tag{1}$$

where $\boldsymbol{\psi}_k(t) = (\psi_{k1}, ..., \psi_{kL})$, $\boldsymbol{\omega_k}(t) = (\omega_{k1}, ..., \omega_{kR})$, and $\{\boldsymbol{\phi}_k(t)\} = \{(\boldsymbol{\psi}_k(t), \boldsymbol{\omega}_k(t))\}$ denotes a set of orthonormal basis functions in a D-dimensional Hilbert space $\mathcal{H}$ where $D = L + R$. The $\exp_{\boldsymbol{\nu}(t)}$ is the aforementioned exponential map for projecting the variation of $\boldsymbol{W}(t)$ into a Hilbert space. The two weighting matrices $\boldsymbol{\Gamma}_{\boldsymbol{X}}$ and $\boldsymbol{\Gamma}_{\boldsymbol{W}}$ balance the heteroscedasticity among different functional variables and are used to normalize their variations to the same scale. For example in our numerical studies, we define $\boldsymbol{\Gamma}_{\boldsymbol{X}}$ to be a $L \times L$ diagonal matrix with the diagonal element to be $\sqrt{\text{Cov}(X_l(t), X_l(t)}$. Also, $\boldsymbol{\Gamma}_{\boldsymbol{W}}$ corresponds to a $R \times R$ diagonal matrix with elements of $\sqrt{\text{Cov}(U_r(t), U_r(t)}$ where we assume the variation of $\boldsymbol{W}(t)$ can be captured by that of $\boldsymbol{U}(t)$. The representation (1) emphasizes on the functional principal components $\boldsymbol{\phi}_k(t)$'s that are extending over both topological spaces. Moreover, they will be derived from a joint covariance function rather than individually for each dimension. This will ensure that variations between dimensions are also analyzed and used for the multivariate analysis.

For obtaining the FPCs, we will start from examining the concatenated variable $\boldsymbol{V}(t)$. Let $\boldsymbol{\eta}(t) = (\boldsymbol{\mu}(t), \boldsymbol{\nu}(t))$ and $\boldsymbol{G}(s,t) = \{G_{dd'}(s,t)\}_{d,d'=1,...,D}$ to be the mean and covariance function of $\boldsymbol{V}(t)$ respectively. The normalization for $\boldsymbol{V}(t)$ can be carried out as

$$Z_d(t) = \frac{V_d(t) - \eta_d(t)}{\sqrt{G_{dd}(t,t)}} \tag{2}$$

in each dimension $d = 1, ..., D$ and produces a noramlized functional variable $\boldsymbol{Z}(t) = (Z_1(t), ..., Z_D(t))$. Subsequently, we can apply the multivariate FPCA for the normalized functional variable $\boldsymbol{Z}(t)$ as normal functiona variables in a Hilbert space. Define $\boldsymbol{C}(s,t)$ to

be the covariance function of $\boldsymbol{Z}(t)$, and it is specified in the usual way as

$$\boldsymbol{C}(s,t) = \mathbb{E}[\boldsymbol{Z}(s) - \mathbb{E}\boldsymbol{Z}(s)][\boldsymbol{Z}(t) - \mathbb{E}\boldsymbol{Z}(t)]^T$$

where each element $C_{dd'}(s,t) = \mathbb{E}[Z_d(s) - \mathbb{E}Z_d(s)][Z_{d'}(t) - \mathbb{E}Z_{d'}(t)]$ for $d, d' = 1, ..., D$. Based on the covariance function $\boldsymbol{C}(s,t)$, the integral operator $L : \mathcal{H} \to \mathcal{H}$ performs a transformation as follows

$$L\boldsymbol{Z}(s) = \int_\tau \boldsymbol{C}(s,t)\boldsymbol{Z}(t)dt = \langle \boldsymbol{C}, \boldsymbol{Z}\rangle_{\mathcal{H}} = \begin{pmatrix} \langle \boldsymbol{C}_{1\cdot}, \boldsymbol{Z}\rangle_{\mathcal{H}} \\ \langle \boldsymbol{C}_{2\cdot}, \boldsymbol{Z}\rangle_{\mathcal{H}} \\ \vdots \\ \langle \boldsymbol{C}_{D\cdot}, \boldsymbol{Z}\rangle_{\mathcal{H}} \end{pmatrix}$$

where $\boldsymbol{C}_{d\cdot}$ corresponds to the row elements of $\boldsymbol{C}(s,t)$, $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ is the inner product operator indulged in $\mathcal{H}$, and $\langle \boldsymbol{C}_{1\cdot}, \boldsymbol{v}\rangle_{\mathcal{H}} = \sum_{d=1}^{D} \int_\tau C_{1d}(s,t)v_d(t)dt$. The kernel of the integral operator L is a covariance function which is symmetric by its definition, then the operator is self-adjoint because

$$\langle L\boldsymbol{u}, \boldsymbol{v}\rangle_{\mathcal{H}} = \langle \boldsymbol{u}, L\boldsymbol{v}\rangle_{\mathcal{H}} \ , \ \forall \boldsymbol{u}, \boldsymbol{v} \in \mathcal{H}.$$

Provided that L is a compact and self-joint operator, the Hilbert-Schmidt theorem (Renardy and Rogers (2004)) states the existence of a set of eigenfunctions $\boldsymbol{\phi}_k(t)$ with the corresponding eigenvalues λ_k, in monotone non-increasing order, for $k = 1, ..., \infty$. The covariance function $\boldsymbol{C}(s,t)$ is non-negative, and so is the integral operator L. By the Mercer's theorem (Minh et al. (2006)), there exists a orthonormal basis in $\mathcal{H}$ as the set of eigenfunctions by which the continuous kernel $\boldsymbol{C}(s,t)$ can be written as the summation

$$\boldsymbol{C}(s,t) = \sum_{k=1}^{\infty} \lambda_k \boldsymbol{\phi}_k(s)\boldsymbol{\phi}_k(t)^T,$$

which converges absolutely for $(s,t) \in (\tau \times \tau)$ and uniformly for $s \in \tau$. The eigenfunctions

are orthonormal, hence satisfy

$$\langle \boldsymbol{\phi}_k(t), \boldsymbol{\phi}_{k'}(t) \rangle_{\mathcal{H}} = \sum_{d=1}^{D} \langle \phi_{kd}(t), \phi_{k'd}(t) \rangle_{\mathcal{H}} = \delta_{kk'} = \begin{cases} 1 \text{ , if } k = k', \\ 0 \text{ , if } k \neq k'. \end{cases}$$

One merit of the proposed method is to allow the functional principal components to be obtained from a single covariance function and span over all dimensions. Once the functional principal components become available, we can arrange $\boldsymbol{\phi}_k(t)$'s into $(\boldsymbol{\psi}_k(t), \boldsymbol{\omega}_k(t))$'s for assessing variations with respect to each dimension in both $\boldsymbol{X}(t)$ and $\boldsymbol{W}(t)$. Furthermore, we can use a fixed number of functional principal components to construct a lower-dimensional space for reconstructing individual curves. With the addition of the mapping and normalization step, the modified FPCA offers a better-suited method for a combination of variables on different topological spaces than the traditional technique. As result, the functional principal components are equally useful with respect to both sets of variables because they would be estimated and examined simultaneously. Given that the additional steps do not disturb the traditional FPCA methodology, the modification allows the estimation to be carried out in a similar way as that of the traditional method. Also, any benefits of the traditional FPCA are retained in the proposed multivariate FPCA method.

One application of FPCA is to reconstruct individual curves from their discrete observations. Furthermore, the reconstruction can be accomplished even when functions are partially observed, which will be illustrated in the next section. Subsequent to obtaining the functional principal components $\boldsymbol{\phi}_k(t)$'s, individual functions can be interpreted through their FPC scores as projection in the aforementioned space spanned by the functional principal components. A normalized random function $\boldsymbol{z}_i(t)$ accepts the approximation by using a selection of K functional principal components as the following

$$\boldsymbol{z}_i(t) = \sum_{k=1}^{K} \xi_{ik} \boldsymbol{\phi}_k(t), \tag{3}$$

where $\{\xi_{ik}\}_{k=1}^{K}$ are the FPC scores obtaining from the projection onto each functional

principal component:

$$\xi_{ik} = \int_{\tau} \boldsymbol{z}_i(t)\boldsymbol{\phi}_k(t)dt$$

for $k = 1, ..., K$. The representation (3) allows us to use an array of scores to reconstruct any random curves. Through the exponential map defined on $\mathcal{M}$, manifold-valued functions are also easily recovered based on the FPCA result as illustrated in Model (1).

2.2 *Model Estimation*

The first step in the model estimation is to obtain the Fréchet mean function $\boldsymbol{\nu}(t)$ required in the mapping procedure for the manifold-valued variable $\boldsymbol{W}(t)$. We know that the Fréchet mean $\boldsymbol{\nu}(t)$ is a general minimizer of a squared distance function. For any time $t \in [a, b]$, $\boldsymbol{\nu}(t)$ should minimize the following distance function as

$$\boldsymbol{\nu}(t) = \arg\min_{p} \frac{1}{N} \sum_{i=1}^{N} d_g(p, \boldsymbol{w}_i(t))^2$$

given any set of random trajectories $\boldsymbol{w}_i(t)$'s for $i = 1, .., N$. Let $y_{ij}^{\boldsymbol{W}} = \boldsymbol{w}_i(t_j)$ be the observation of i-th manifold-valued function at time t_j for $j = 1, ..., M$. Then, the empirical estimate of the mean function $\boldsymbol{\nu}(t)$ at any observation point t_j is obtained by the following minimization:

$$\hat{\boldsymbol{\nu}}(t_j) = \arg\min_{p} \frac{1}{N} \sum_{i=1}^{N} d_g(p, \boldsymbol{y}_{ij}^{\boldsymbol{W}})^2 \tag{4}$$

One simplification of the estimation process is to reduce the observation point indices from t_{ij} to t_j, and this will align the observations from all random curves and discard any missing observations in the minimization process. Furthermore, we focus on acquiring estimates of the mean function at t_j's instead of an estimate of the entire function, since the discrete estimates are assumed to be sufficient for mapping observations $\boldsymbol{y}_{ij}^{\boldsymbol{W}}$. Consequently, the mapped observations $\boldsymbol{y}_{ij}^{\boldsymbol{U}} = \log_{\hat{\boldsymbol{\nu}}(t_{ij})} \boldsymbol{y}_{ij}^{\boldsymbol{W}}$ are treated as the observations of $\boldsymbol{U}(t)$.

The normalization procedure in (2) requires the mean and covariance function of $\boldsymbol{V}(t)$.

The estimation of the mean function $\boldsymbol{\eta}(t) = (\mathrm{E}(\boldsymbol{X}(t)), \mathrm{E}(\boldsymbol{U}(t)))$ is separated into two pieces with respect to $\boldsymbol{X}(t)$ and $\boldsymbol{U}(t)$. However, the estimation of $\mathrm{E}(\boldsymbol{U}(t))$ is ignored because the log-mapping dictates that the mean function of the log-mapped variables are inherently zero. Hence, only $\boldsymbol{\mu}(t) = \mathrm{E}(\boldsymbol{X}(t))$ needs to be estimated for doing the normalization.

The estimation of mean functions of $\boldsymbol{X}(t)$ is a standard problem in functional data analysis. One obstacle frequently undermines the estimation and are caused by the missing observations of individual curves. The problem can be solved by pulling together available observations from all functions to estimate the underlying mean function. Based on the collected observations, the local linear regression is employed for obtaining a smooth estimate of the mean function. The local linear regression technique was originally utilized for estimating functions based on observations of a single random function. Later, Li and Hsing (2010) and Yao et al. (2005) advance the local regression approach for a collection of observations.

The mean function $\mu_l(t)$, for $l = 1, ..., L$, is estimated by the intercept of a local linear regression line from all the observations included within a sliding kernel density function $k_l(t)$. Adjusting for the range and density of the observation interval, the kernel needs a transformation $k_l(t) = \frac{1}{h_l} k(\frac{t}{h_l})$ with a bandwidth h_l. Then, a regression line $\alpha_0 + \alpha_1(t_{ij} - t)$ is estimated as the minimizer of

$$(\hat{\alpha_0}, \hat{\alpha_1}) = \underset{\alpha_0, \alpha_1}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^{N} \frac{1}{m_i} \sum_{j=1}^{m_i} (y_{ijl}^{\boldsymbol{X}} - \alpha_0 - \alpha_1(t_{ij} - t))^2 k_l(t_{ij} - t). \tag{5}$$

The estimate of the mean function $\hat{\mu}_l(t) = \hat{\alpha_0}$ is obtained. After estimating $\boldsymbol{\mu}(t)$, the mean function of $\boldsymbol{V}(t)$ has the estimate as $\hat{\boldsymbol{\eta}}(t) = (\hat{\boldsymbol{\mu}}(t), \mathbf{0})$, and individual indices for $\boldsymbol{X}(t)$ and $\boldsymbol{U}(t)$ can be ignored for further estimations. Given the concatenated observation $\boldsymbol{y}_{ij}^{\boldsymbol{V}} = (\boldsymbol{y}_{ij}^{\boldsymbol{X}}, \boldsymbol{y}_{ij}^{\boldsymbol{U}})$ of i-th function $\boldsymbol{v}_i(t)$ at time t_{ij} for $j = 1, ..., m_i$, we can proceed to estimate covariance functions of $\boldsymbol{V}(t)$ for the subsequent normalization.

To complete the normalization, each variable in $\boldsymbol{V}(t)$ needs to be scaled by its covariance function as shown in (2). We need to first estimate individual covariance functions, i.e.,

$G_{dd}(s,t) = \mathrm{Cov}(V_d(s), V_d(t))$ for $d = 1, ..., D$. The covariance functions are assumed to be smooth functions, hence the sample covariance matrix $\hat{\boldsymbol{G}}_{dd}(t_j, t_{j'}) = \sum_{i=1}^{N} y_d^{\boldsymbol{V}}(t_{ij}) y_d^{\boldsymbol{V}}(t_{ij'})$ can be a good approximation if the number of observation points m_i is sufficiently large. We can again take the advantage of local linear regression by fitting a linear plane to the observations in the neighborhood of the paired time points (s,t) to estimate $G_{dd}(s,t)$. The local plane is in the form of $\alpha_0 - \alpha_1(t_{ij} - s) - \alpha_2(t_{ij'} - t)$. Moreover, an estimate $\hat{G}_{dd}(s,t)$ is obtained as the intercept estimate $\hat{\alpha}_0$ through the minimization of

$$\frac{1}{N}\sum_{i=1}^{N} \frac{1}{m_i(m_i-1)} \sum_{j=1}^{m_i} \sum_{\substack{j'=1 \\ j' \neq j}}^{m_i} (y_{ijd}^{\boldsymbol{V}} y_{ij'd}^{\boldsymbol{V}} - \alpha_0 - \alpha_1(t_{ij} - s) - \alpha_2(t_{ij'} - t))^2 k_d(t_{ij} - s) k_d(t_{ij'} - t) \quad (6)$$

where $k_d(\cdot)$ is a kernel function with the bandwidth h_d.

Once the estimated covariance functions become available, the observations of centered functional variables can be scaled by the diagonal elements of the covariance functions. We obtain the normalized observations $\boldsymbol{y}_{ij}^{\boldsymbol{Z}}$'s of $\boldsymbol{Z}(t)$ as follows

$$y_{ijd}^{\boldsymbol{Z}} = \frac{y_{ijd}^{\boldsymbol{V}} - \eta_d(t_{ij})}{\sqrt{\hat{G}_{dd}(t_{ij}, t_{ij})}}.$$

Furthermore, it follows that the individual covariance function of normalized variables $C_{dd}(s,t) = \mathrm{Cov}(Z_d(s), Z_d(t))$ can be derived as

$$C_{dd}(s,t) = \frac{G_{dd}(s,t)}{\sqrt{G_{dd}(s,s) G_{dd}(t,t)}}. \quad (7)$$

The estimated covariance functions $\hat{C}_{dd}(s,t)$ can be obtained by plugging the estimated $\hat{G}_{dd}(s,t)$ into (7).

The covariance function $\boldsymbol{C}(s,t)$ of $\boldsymbol{Z}(t)$ is also composed of cross-covariance functions $C_{dd'}(s,t)$ for $d, d' = 1, ..., D$ and $d \neq d'$. These cross-covariance functions are obtained from the normalized observations $\boldsymbol{y}_{ij}^{\boldsymbol{Z}}$'s. The estimation of $C_{dd'}(s,t)$'s uses the same methodology as that of $G_{dd}(s,t)$. Let a local linear plane to be defined in the same way as in (6), an estimate $\hat{C}_{dd'}(s,t)$ is obtained as the intercept term of the plane after fitting the pooled

observation located within the neighborhood around (s,t). That is, $\hat{C}_{dd'}(s,t)$ corresponds to $\hat{\alpha}_0$ as the result of minimizing

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{m_i(m_i-1)}\sum_{j=1}^{m_i}\sum_{\substack{j'=1\\ j'\neq j}}^{m_i}(y_{ijd}^{\boldsymbol{Z}}y_{ijd'}^{\boldsymbol{Z}}-\alpha_0-\alpha_1(t_{ij}-s)-\alpha_2(t_{ij'}-t))^2k_d(t_{ij}-s)k_{d'}(t_{ij}-t), \quad (8)$$

where $k_d(\cdot)$ and $k_{d'}(\cdot)$ in (6) are kernel functions with the bandwidth h_d and $h_{d'}$.

Let $\epsilon_{ijd}\sim \mathrm{N}(0,\sigma_d^2)$ be the i.i.d. random errors associated within the observations of the i-th random function of $\boldsymbol{V}$ as $y_{ijd}^{\boldsymbol{V}}=V_d(t_{ij})+\epsilon_{ijd}$. The variance of observations $y_{ijd}^{\boldsymbol{V}}$ is composed of two parts as

$$\mathrm{Var}(y_{ijd}^{\boldsymbol{V}})=G_{dd}(t_{ij},t_{ij})+\sigma_d^2,$$

for $d=1,...,D$. If we use all observations to estimate $G_{dd}(s,t)$'s, then it would be overestimated with the addition of the error variance σ_d^2 when $s=t$. For that matter, the estimation $G_{dd}(s,t)$ described in (6) does not use any information that would be contaminated with random errors when $j=j'$. In order to estimate σ_d, we can consider a new function $\Omega_d(t)=G_{dd}(t,t)+\sigma_d^2$ for isolating σ_d from $G_{dd}(t,t)$. The estimation of $\Omega_d(t)$ relies on the diagonal elements of the sample covariance matrix. Define a local line $\alpha_0-\alpha_1(t_{ij}-t)$ in the neighborhood of t_{ij}, an estimate of $\Omega_d(t)$ corresponds to $\hat{\alpha}_0$ as the result of minimizing

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}\sum_{j=1}^{m_i}((y_{ijd}^{\boldsymbol{V}})^2-\alpha_0-\alpha_1(t_{ij}-t))^2k_d(t_{ij}-t). \quad (9)$$

The estimation of the error variance σ_d^2 for each functional variable V_d follows Yao et al. (2005) and Chiou et al. (2014), i.e.,

$$\hat{\sigma}_d^2=\frac{1}{T}\int_0^T[\hat{\Gamma}_d(t)-\hat{G}_{dd}(t,t)]dt \quad (10)$$

where $\hat{G}_{dd}(t,t)$ is the estimated covariance function based on (6).

The last estimation concerns with FPC scores which are used for the reconstruction of individual curves. The most straightforward way to calculate FPC scores ξ_{ik}'s are to calculate the inner products for projecting the i-th function in the space spanned by the FPCs, i.e.,

$\xi_{ik} = \langle \boldsymbol{y}_i^{\boldsymbol{Z}}, \hat{\boldsymbol{\phi}}_k \rangle$, where $\hat{\boldsymbol{\phi}}_k, k = 1, ..., K$, are the estimated functional principal components by taking eigen-decomposition of $\hat{\boldsymbol{C}}(s,t)$. The inner products involve integrations which cannot be well calculated when a random function has too few of observations. One alternative to the direct integration is to consider the calculation as if it is a regression problem. That is, the FPC scores of a random function is the equivalent to a K-dimension regression coefficients which come from

$$\tilde{\boldsymbol{y}}_i^{\boldsymbol{Z}} = \tilde{\boldsymbol{\phi}}_i \boldsymbol{\xi}_i + \tilde{\boldsymbol{\zeta}}_i \tag{11}$$

where $\tilde{\boldsymbol{y}}_i^{\boldsymbol{Z}} = (\boldsymbol{y}_{i\cdot 1}^{\boldsymbol{Z}}, ..., \boldsymbol{y}_{i\cdot D}^{\boldsymbol{Z}})$ with $\boldsymbol{y}_{i\cdot d}^{\boldsymbol{Z}} = (y_{i1d}^{\boldsymbol{Z}}, ..., y_{im_id}^{\boldsymbol{Z}})$, $\tilde{\boldsymbol{\phi}}_i = (\tilde{\boldsymbol{\phi}}_{i1}, ..., \tilde{\boldsymbol{\phi}}_{iK})$ with $\tilde{\boldsymbol{\phi}}_{ik} = (\boldsymbol{\phi}_{ik1}, ..., \boldsymbol{\phi}_{ikD})$ and $\boldsymbol{\phi}_{ikd} = (\boldsymbol{\phi}_{kd}(t_{i1}), ..., \boldsymbol{\phi}_{kd}(t_{im_i}))^T$, $\tilde{\boldsymbol{\zeta}}_i = (\boldsymbol{\zeta}_{i\cdot 1}, ..., \boldsymbol{\zeta}_{i\cdot D})$ with $\boldsymbol{\zeta}_{i\cdot d} = (\zeta_{i1d}, ..., \zeta_{im_id})$, and ζ_{ijd}'s are errors of normalized variables with variances $\sigma_d^2 / G_{dd}(t_{ij}, t_{ij})$. The estimation of FPC scores $\boldsymbol{\xi}_i$ follows the routine of regression problems which is the minimizing of the least square objective function $||\tilde{\boldsymbol{y}}_i^{\boldsymbol{Z}} - \tilde{\boldsymbol{\phi}}_i \boldsymbol{\xi}_i||^2$, where $\tilde{\boldsymbol{\phi}}_i$ is obtained by the estimated functional principal components $\hat{\boldsymbol{\phi}}_k$'s.

The regression framework is easy to employ but is not reliable given functions are not fully observed. As a remedy, we can use the conditional expectation method which was initially introduced in Yao et al. (2005) and extended by Chiou et al. (2014) for multivariate case. First, we can consider that both $\boldsymbol{\xi}_i$ and $\tilde{\boldsymbol{\eta}}_i$ are random quantities, hence (11) is equivalent to a random effect model. Then the conditional expectation method provides the best linear unbiased predictions of $\boldsymbol{\xi}_i$. Let $\tilde{\boldsymbol{y}}_i^{\boldsymbol{Z}}$ and $\boldsymbol{\xi}_i$ be jointly Gaussian:

$$\begin{bmatrix} \tilde{\boldsymbol{y}}_i^{\boldsymbol{Z}} \\ \boldsymbol{\xi}_i \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{\tilde{\boldsymbol{y}}_i^{\boldsymbol{Z}}} & \boldsymbol{\Sigma}_{\tilde{\boldsymbol{y}}_i^{\boldsymbol{Z}} \boldsymbol{\xi}_i} \\ \boldsymbol{\Sigma}_{\boldsymbol{\xi}_i \tilde{\boldsymbol{y}}_i^{\boldsymbol{Z}}} & \boldsymbol{\Sigma}_{\boldsymbol{\xi}_i} \end{bmatrix} \right),$$

where $\boldsymbol{\Sigma}_{\tilde{\boldsymbol{y}}_i^{\boldsymbol{Z}}}$ is filled with block matrices $\{C_{dd'}(t_{ij}, t_{ij'}) + \delta_{jj'}\delta_{dd'}\frac{\sigma_d^2}{G_{dd}(t_{ij},t_{ij})}\}_{d,d'=1,...,D}$ for $j = 1, ..., m_i$ where $\delta_{jj'} = 1$ and $\delta_{dd'} = 1$ when $j = j'$ and $d = d'$, respectively. The off-diagonal

block matrix $\boldsymbol{\Sigma}_{\boldsymbol{\xi}_i\tilde{\boldsymbol{y}}_i^{\boldsymbol{Z}}}$ corresponds to the covariance matrix of $\boldsymbol{\xi}_i$ and $\tilde{\boldsymbol{y}}_i^{\boldsymbol{Z}}$:

$$\mathrm{cov}(\boldsymbol{\xi}_i, \tilde{\boldsymbol{y}}_i^{\boldsymbol{Z}}) = \begin{pmatrix} \mathrm{cov}(\xi_{i1}, \tilde{\boldsymbol{y}}_i^{\boldsymbol{Z}}) \\ \cdots \\ \mathrm{cov}(\xi_{iK}, \tilde{\boldsymbol{y}}_i^{\boldsymbol{Z}}) \end{pmatrix},$$

where $\mathrm{cov}(\xi_{ik}, \tilde{\boldsymbol{y}}_i^{\boldsymbol{Z}}) = \mathrm{cov}(\xi_{ik}, \xi_{ik}\tilde{\boldsymbol{\phi}}_{ik}) = \lambda_k\tilde{\boldsymbol{\phi}}_{ik}$. The conditional expectation of $\boldsymbol{\xi}_i$ given observations $\tilde{\boldsymbol{y}}_i^{\boldsymbol{Z}}$ is derived as

$$\hat{\boldsymbol{\xi}}_i = \boldsymbol{\Sigma}_{\boldsymbol{\xi}_i\tilde{\boldsymbol{y}}_i^{\boldsymbol{Z}}}\boldsymbol{\Sigma}_{\tilde{\boldsymbol{y}}_i^{\boldsymbol{Z}}}^{-1}\tilde{\boldsymbol{y}}_i^{\boldsymbol{Z}}. \tag{12}$$

Substituting the estimated quantities into (12), we can obtain the estimated FPC scores $\hat{\boldsymbol{\xi}}_i$ for all individual functions. The entire algorithm is summarized as follows:

Algorithm 1: FPCA for multivariate variables on different manifolds

---

Data: $\{\boldsymbol{y}_{ij}^{\boldsymbol{X}}\}$, $\{\boldsymbol{y}_{ij}^{\boldsymbol{W}}\}$

Input: K,$\{h_d\}$,$\{h_l\}$

Output: $\{\lambda_k\}$,$\{\boldsymbol{\phi}_k\}$,$\{\xi_{ik}\}$

Step 1: Estimate mean functions $\hat{\boldsymbol{\nu}}(t)$ and $\hat{\boldsymbol{\mu}}(t)$ from the minimization of (4) and (5)

Step 2: Perform log-mapping on $\boldsymbol{W}(t)$ to define $\boldsymbol{U}(t)$ through $\log_{\boldsymbol{\nu}(t)}\boldsymbol{W}(t)$, and obtain observations $\{\boldsymbol{y}_{ij}^{\boldsymbol{U}}\}$ of $\boldsymbol{U}(t)$ by $\log_{\hat{\boldsymbol{\nu}}(t_{ij})}\boldsymbol{y}_{ij}^{\boldsymbol{W}}$

Step 3: Estimate weighting functions $\boldsymbol{\Gamma}_{\boldsymbol{X}}(\cdot)$ and $\boldsymbol{\Gamma}_{\boldsymbol{W}}(\cdot)$ which correspond to $\{\sqrt{\mathrm{Cov}(X_l(t), X_l(t))}\}$ and $\{\sqrt{\mathrm{Cov}(U_r(t), U_r(t))}\}$

Step 4: Concatenate $\boldsymbol{V}(t) = (\boldsymbol{X}(t), \boldsymbol{U}(t))$ and normalize to $\boldsymbol{Z}(t) = (\frac{\boldsymbol{X}(t)-\boldsymbol{\mu}(t)}{\boldsymbol{\Gamma}_{\boldsymbol{X}}}, \frac{\boldsymbol{U}(t)}{\boldsymbol{\Gamma}_{\boldsymbol{W}}})$ which provide observations $\boldsymbol{y}_{ij}^{\boldsymbol{Z}} = (\frac{\boldsymbol{y}_{ij}^{\boldsymbol{X}}-\hat{\boldsymbol{\mu}}(t_{ij})}{\hat{\boldsymbol{\Gamma}}_{\boldsymbol{X}}(t_{ij})}, \frac{\boldsymbol{y}_{ij}^{\boldsymbol{U}}}{\hat{\boldsymbol{\Gamma}}_{\boldsymbol{W}}(t_{ij})})$

Step 5: Perform multivariate FPCA on $\boldsymbol{Z}(t)$ to obtain $\{\lambda_k\}$, $\{\boldsymbol{\phi}_k\}$ and $\{\xi_{ik}\}$ with $\{\boldsymbol{y}_{ij}^{\boldsymbol{Z}}\}$

---

3. Simulations

In this section, we will examine the proposed multivariate FPCA method with two simulation studies. We assess the estimation accuracy of model components under an ideal setting. Subsequently, we explore the consistency of estimations when the sparsity of data varies at different degrees.

3.1 *Simulation I*

The proposed multivariate FPCA method is first examined under a simulation setting where all functional variables are fully observed. Let $\boldsymbol{X}(t) = (X_1(t), X_2(t))$ be real-valued functions and $W(t)$ locate on 1-sphere, $\mathbb{S}^1$. On $\mathbb{S}^1$, the metric for calculating the geodesic distance (great-circle distance) between any two points p, q is

$$d_g(p, q) = \arccos(\vec{n}_p, \vec{n}_q)$$

where $\vec{n}$ is the normal vector of a point on the sphere. Using the ambient space $\mathbb{E}^2$ of $\mathbb{S}^1$, the tangent space of any point p corresponds to all the orthogonal vectors $\vec{v}$ such that $T_p\mathcal{M} = \{\vec{v}|\vec{v}^T \cdot \vec{n}_p = 0\}$. The logarithm map from $\mathcal{M}$ to $T_p\mathcal{M}$ is as follows

$$\log_p(q) = \frac{\vec{u}}{||\vec{u}||} d_g(p, q)$$

acting as a linear mapping parameterized by the distance and the orthogonal vector $\vec{u} = \vec{n}_q - \vec{n}_p^T \vec{n}_q \vec{n}_p$ at the point p. The exponential map is defined as

$$\exp_p(\vec{v}) = \cos(\vec{v})\vec{n}_p + \sin(\vec{v})\frac{\vec{v}}{||\vec{v}||}$$

for any $v \in T_p\mathcal{M}$.

The simulated data are generated as follows. First, using a set of orthogonal functions , trigonometric and constant functions, we can generate kernel functions $\tilde{G}_{dd}(s, t)$ for each dimension. The following method can produce the diagonal elements of a covariance function

for multivariate functional variables

$$C_{dd}(s,t) = \frac{\tilde{G}_{dd}(s,t)}{\sqrt{\tilde{G}_{dd}(s,s)}\sqrt{\tilde{G}_{dd}(t,t)}}$$

The eigen-decomposition of $\tilde{G}_{dd}(s,t)$ defines a basis with eigenvalues γ_{kd} and $\psi_{kd}(t)$. Then, the off-diagonal elements $C_{dd'}(s,t)$'s are filled by the following

$$C_{dd'}(s,t) = \sum_k (\frac{1}{3}\sum_d^3 \gamma_{kd})\psi_{kd}(s)\psi'_{kd}(t)^T$$

Subsequently, the decomposition of $\boldsymbol{C}(s,t)$ can define $\boldsymbol{\phi}_k(t)$ and λ_k accordingly. The weighting functions are for adding heteroscedasticity into each variable as indicated in (1), and they are set to be $\boldsymbol{\Gamma_X}(t) = (\sqrt{0.1+(1-t)^2}, \sqrt{0.1+0.5(t-1)^4})$ and $\Gamma_W(t) = \sqrt{0.2+0.2\cos(2\pi t)}$. The mean function $\boldsymbol{\mu}(t)$ is set to be zero for simplicity, and $\boldsymbol{\nu}(t)$ is defined to be $90 + 100\sin^2(2\pi t)$ for mapping between the manifold and its tangent space. Choosing the first three eigen-functions $\{\boldsymbol{\phi}_k\}_{k=1}^3$ and setting the eigen-values to be $(0.65, 0.3, 0.05)$, we can now simulate random curves by the random generation of FPC scores $\xi_{ik} \sim \mathrm{N}(0, \lambda_k)$ through the following

$$\begin{aligned} \boldsymbol{y}_{ij}^X &= \sum_{k=1}^3 \xi_{ik}(\boldsymbol{\Gamma_X} \circ \boldsymbol{\psi}_k(t)) + \boldsymbol{\epsilon}_{ij}^{\boldsymbol{X}} \\ \boldsymbol{y}_{ij}^W &= \exp(\sum_{k=1}^3 \xi_{ik}(\boldsymbol{\Gamma}_W \circ \boldsymbol{\omega}_k(t)) + \boldsymbol{\epsilon}_{ij}^W) \end{aligned}$$

where the random noises $\boldsymbol{\epsilon}_{ij}^{\boldsymbol{X}}$ and $\boldsymbol{\epsilon}_{ij}^W$ have distributions $\epsilon_{ij1}^{\boldsymbol{X}} \sim \mathrm{N}(0, 0.09)$, $\epsilon_{ij2}^{\boldsymbol{X}} \sim \mathrm{N}(0, 0.4)$ and $\epsilon_{ij}^W \sim \mathrm{N}(0, 0.01)$ respectively.

We then apply the proposed multivariate FPCA method on the simulated data. This simulation study is repeated for 100 times. First, we assess the estimation of $\boldsymbol{G}(s,t)$ which is used for normalizing functional variables. Figure 2 displays the point-wise average of the estimated covariance functions $G_{dd}(t,t)$ for $d = 1, 2, 3$. They are all very close to the true covariance functions used for generating the simulated data. Such result allows us to confidently normalize the functional variables with the estimated covariance functions, and

the variations among normalized variables can be compared and calculated regardless of individual units or scales of variability.

[Figure 2 about here.]

Following the normalization, the covariance function $\boldsymbol{C}(s,t)$ of the normalized variable $\boldsymbol{Z}(t)$ can be estimated. The functional principal components is then estimated as the eigenfunctions of $\boldsymbol{C}(s,t)$. Figure 3 displays the average of the estimated functional principal components (FPCs), which are all close the true FPCs. The multivariate FPCA method involves concatenating all individual functional variable together into one long function. Hence, the estimated FPCs at connection points, equivalently at the start and end of functions, slightly deviate from the true FPCs.

[Figure 3 about here.]

3.2 *Simulation II*

In some real data applications, individual functions are not fully observed due to the nature of variables or measuring methods. The second simulation study is to investigate the accuracy of the proposed method given different degrees of sparsity in the data. The full data are generated in the same settings as the first simulation study, but we randomly sample 10%, $20\%, \cdots, 100\%$ data from the full as the simulation data. Let the case where functions are fully observed be a benchmark, then we can compare the estimation results across various sparsity settings. First, the estimated covariance functions $G_{dd}(s,t), d = 1,2,3$, are compared based on the integrated squared errors (ISEs), which is defined as

$$ISE(G_{dd}) = \int_{\tau}\int_{\tau} |G_{dd}(s,t) - \hat{G}_{dd}(s,t)|^2 dsdt. \tag{13}$$

The simulation is replicated for 100 times. Table 1 summarizes the ISEs for estimated covariance function $G_{dd}(s,t)$'s in various sparsity settings. It shows that the estimation is

considerably consistent for all levels of sparsity. However, a slight increase in the variability in ISE is observed as the portion of missing data increases.

[Table 1 about here.]

We also compare the integrated squared errors of the estimated FPCs in each dimension, which is defined as

$$ISE(\phi_{kd}) = \int_{\tau} |\phi_{kd}(t) - \hat{\phi}_{kd}(t)|^2 dt. \tag{14}$$

Table 2 provides the estimation result of the first FPC under various sparsity settings. It shows that the estimation of $\boldsymbol{\phi}_1(t)$ is consistent for all dimensions in different topological spaces. There is no significant difference in terms of ISE across different levels of sparsity in the data. The ISE only jumps a little bit from the complete data case to those with missing observations. The local polynomial regression technique used in our proposed method provides reliable estimates because it can pull time-neighboring observations together for mitigating problems caused by missing observations. The mapping procedure also ensures the success of joint analyses between variables on different manifolds.

[Table 2 about here.]

4. Application

We will apply the proposed multivariate FPCA on the motivating wind data. Under a simple setting, the wind data include the direction and speed of the wind over time. The wind direction and speed variable do not share the same topology, because the wind direction variable is usually defined on a circle while the wind speed variable resides in a Euclidean space. The topological differences between these two functional variables are often ignored in joint analyses as well as the different units. In addition, the wind data usually have a

large amount of missing data. We will show how the proposed multivariate FPCA method can address these difficulties of joint analyses and the missing data problem.

The application is carried out on a hourly wind direction and speed from Vancouver, Toronto and Halifax in the years from 1998 to 2018. The data is provided by Environment and Climate Change Canada (2019) and are available on `https://climate.weather.gc.ca/`. We focus on the data in Decembers of 1998 to 2018, because we assume that the days in Decembers have the same covariance structure. Figure 4 displays a few curves randomly selected. It shows that the wind direction variable has more missing observations than the speed variable.

[Figure 4 about here.]

First, the covariance functions of both wind direction and speed are estimated in order to apply the normalization step. Figure 5 illustrates the covariance surfaces of Vancouver's wind data. It shows that the variability of wind direction becomes the strongest as time is approaching midnight and diminishes around noon. On the contrary, the degree of variation for wind speed is almost constant throughout the day. We also observe some small correlation between mornings and evenings in wind direction, whereas the wind speed demonstrates strong correlation only among adjacent hours. Figure 5 also shows that there is a great difference in the scale of covariance functions of these two variables. The subsequent normalization helps the joint analyses and avoids excessive influences from a particular variable.

[Figure 5 about here.]

[Figure 6 about here.]

After estimating the individual covariance functions, we can apply the normalization to both wind direction (after log-mapping) and speed. Given the normalized functional variables, the proposed multivariate FPCA method can be performed for obtaining the

functional principal components (FPCs) in order to examine the major modes of variations among the wind data. Figure 6 displays the estimated top four FPCs for three cities across Canada including Vancouver, Toronto, and Halifax. The top four FPCs explain 86.97%, 92.70% and 92.60% of the total variation for Vancouver, Toronto and Halifax respectively. We can observe both similarities and distinctions in the FPCs across three cities. In the first FPC, the variation in the wind direction is more potent for Vancouver compared to other cities where the variation in the wind speed is consistent for all cities. In the fourth FPC within the wind direction dimension, Vancouver possesses more oscillatory features compared to the other two cities. In general, We can see that the FPCs of Vancouver in the wind direction have more fluctuations compared to the other two, whereas variations in the wind speed is more agreeable across cities.

[Figure 7 about here.]

After obtaining the FPCs, we can construct individual curves. By using the nonparametric method in estimating covariance functions and subsequent FPCs, random curves can be conveniently recovered by (1) without individual smoothing. Curves of the wind speed variable are reconstructed as expected for all three cities. Figure 7 displays some estimated individual curves for the wind direction and speed variables for the three cities. The trajectories from the wind direction variable show more interesting results. The wind direction variable is defined on 1-sphere, hence the distance and variation should be derived by the great-circle distance rather than the direct difference between any two points. As a result, the temporal variation would not be overestimated, and using the mapped variable ensures that variations are more precise and reflecting the true features of the data. From the reconstruction in Halifax, the recovered curve is not chasing the observation (labeled as point "A") that seems far away from the rest. In fact, that observation is similar to the adjacent points since we need to use the distance metric on the manifold. The same insight can be observed from the other two

cities as well. In the data, the wind direction variable has more missing observations but the proposed multivariate FPCA can still recover individual trajectories with satisfactions.

5. Conclusions

In this work, we have proposed a functional principal component analysis method for multiple variables on different Riemannian manifolds. The diversity of variables overwhelms the conventional FPCA methods, and the proposed one addresses a particular situation where variables have different topological features. The proposed MFPCA introduces a mapping procedure that allows the joint analysis to mitigate the difference in topologies. It also includes a normalization step so that functional variables can be easily investigated as a unity regardless of their variation scales or topological differences. One merit of the method is the utilization of smooth estimations which allow any model components to be approximated without requiring all individual curves to be fully observed. The FPCA results can enable us to reconstruct individual curves based on sparse observations.

Subsequently, the proposed method is applied to a wind data set collected from three cities in Canada. The obtained functional principal components provide useful insights for distinguishing the nature of variation between cities. Furthermore, a lower-dimensional space spanned by the FPCs allows us to recover infinite-dimensional objects using FPC scores even if data curves have randomly missing observations. The consistency of estimations has been demonstrated under various sparsity settings in the simulation studies. In conclusion, the proposed multivariate FPCA is a versatile tool and can be applied to any set of functional variables, since it relaxes the requirement of a common topology.

References

Ramsay, Jim and Silverman, B.W. (2005). *Functional Data Analysis*, 2nd edition. New York: Springer.

Wang, Jane-Ling and Chiou, Jeng-Min and Müller, Hans-Georg (2016). Functional data analysis. *Annual Review of Statistics and Its Application* **3,** 257–295.

Lin, Zhenhua and Wang, Liangliang and Cao, Jiguo (2015). Interpretable functional princiap component analysis. *Biometrics* **72,** 846–854.

Chiou, J.M. and Chen, Y.T. and Yang, Y.F. (2014). Multivariate functional principal component analysis: a normalization approach. *Statistica Sinica* **24,** 1571–1597.

Berrendero, J.R. and Justel, A. and Svarc, M. (2011). Principal components for multivariate functional data. *Computational Statistics and Data Analysis* **55,** 2619–2634.

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

Mardia, K.v. (1972). *Statistics of directional data*. New York: Academic Press.

Belu, R. and Korachin, D. (2013). Statistical and spectral analysis of wind characteristics relevant to wind energy assessment using tower mesurements in complex terrain. *Journal of Wind Energy* **2013** 1–12.

Masseran, N. and Razali, A.M. and Ibrahim, K. and Latif, M.T. (2013). Fitting a mixture of von Mises distributions in order to model data on wind direction in Peninsular Malaysia. *Energy Conversion and Management* **72,** 94–102.

Dai, X. and Müller, H. (2018). Principal component analysis for functional data on Riemannian manifolds and spheres. *The Annals of Statistics* **46,** 3334–3361.

Bhattacharya, R. and Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. *The Annals of Statistics* **31,** 1–29.

Chavel, I. (2006). *Riemannian Geometry: A Modern Introduction*, 2nd edition. Cambridge: University Press.

Renardy, Michael and Rogers, Robert C. (2004). *An Introduction to Partial Differential*

Equations, 2nd edition. New York: Springer.

Minh, Ha Quang and Niyogi, Partha and Yao, Yuan. (2006). Mercer's Theorem, Feature Maps, and Smoothing. *Proceedings of the 19th Annual Conference on Learning Theory* 154–168.

Yao, Fang and Müller, Hans-Georg and Wang, Jane-Ling. (2005). Functional Data Analysis for Sparse Longitudinal Data. *Journal of the American Statistical Association* **100,** 577–590.

Li, Yehua and Hsing, Tailen. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics* **6,** 3321–3351.

Environment and Climate Change Canada. (2019). Historical Climate Data. Retrieved from `climate.weather.gc.ca`.

Hall, Peter and Hosseini-Nasab, Mohammad. (2005). On properties of functional principal components analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68,** 109–126.

Hall, Peter and Müller, Hans-Georg and Wang, Jane-Ling. (2006). Properties of principal component methods for functional and longitudinal data analysis. *The Annals of Statistics* **34,** 1493–1517.

Dauxois, J. and Pousse, A. and Romain, Y.. (1982). Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *Journal of Multivariate Analysis* **12,** 136–154.

Nie, Yunlong and Wang, Liangliang and Liu, Baisen and Cao, Jiguo. (2018). Supervised functional principal component analysis. *Statistics and Computing* **28,** 713–723.

Sang, Peijun and Wang, Liangliang and Cao, Jiguo. (2017). Parametric functional principal component analysis. *Biometrics* **73,** 802–810.

Vieu, Philippe and Ferraty, édéric. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. New York: Springer.

Yao, Fang and Müller, Hans-Georg and Wang, Jane-Line. (2005). Functional linear regression analysis for longitudinal data. *The Annals of Statistics* **33,** 2873–2903.

Müller, Hans-Georg and Stadtmüller, Ulrich. (2005). Generalized functional linear models. *The Annals of Statistics* **33,** 774–805.

Peng, Jie and Müller, Hans-Georg. (2008). Distance-based clustering of sparsely observed stochastic processes, with applications to online auctions. *The Annals of Applied Statistics* **2,** 1056–1077.

Müller, Hans-Georg. (2005). Functional modelling and classification of longitudinal data. *Scandinavian Journal of Statistics* **32,** 223–240.

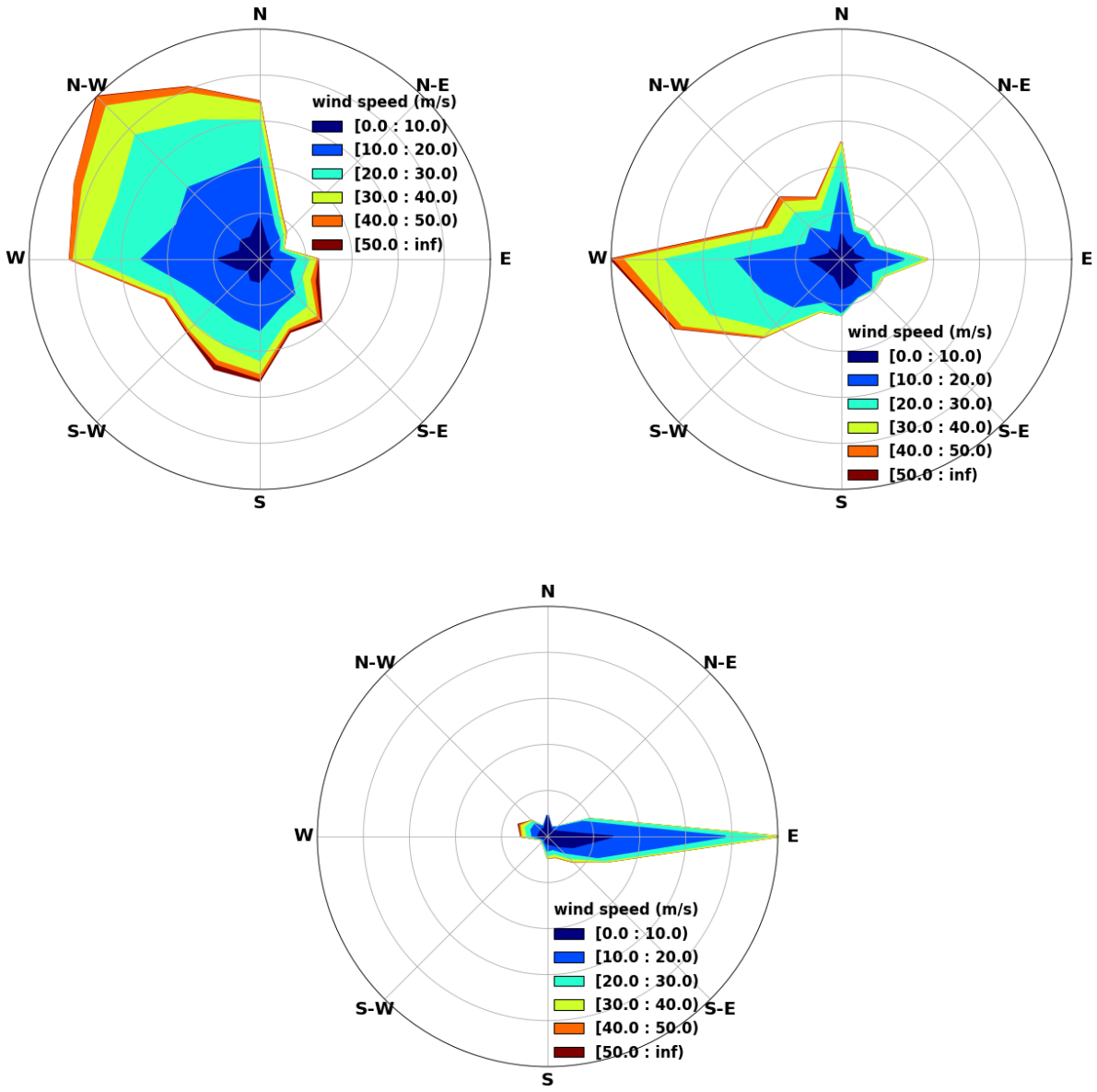

Figure 1. Circular histograms of hourly wind direction and speed at Halifax (top left), Toronto (top right) and Vancouver (bottom) in the month of December from 1998 to 2018. The spoke corresponds to the proportion of wind at a certain speed interval for a given direction.

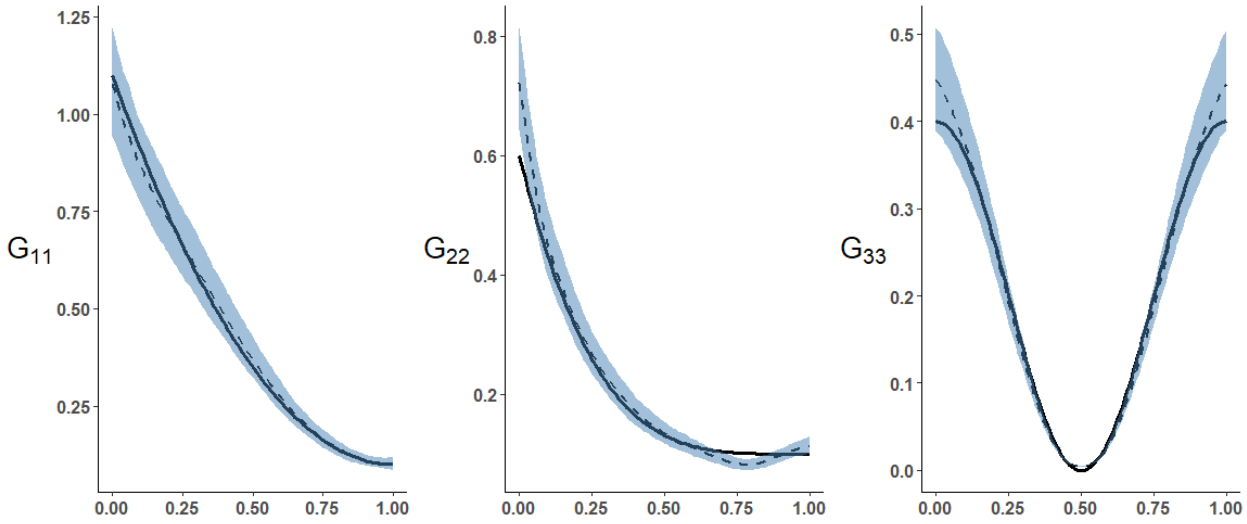

Figure 2. The point-wise average of the estimated covariance functions $\widehat{G}_{dd}(t,t)$ for $d = 1,2,3$ in the Simulation Study I. The solid lines represent the true covariance functions used for generating the simulated data. The shaded areas are 95% point-wise confidence intervals of the estimated covariance functions.

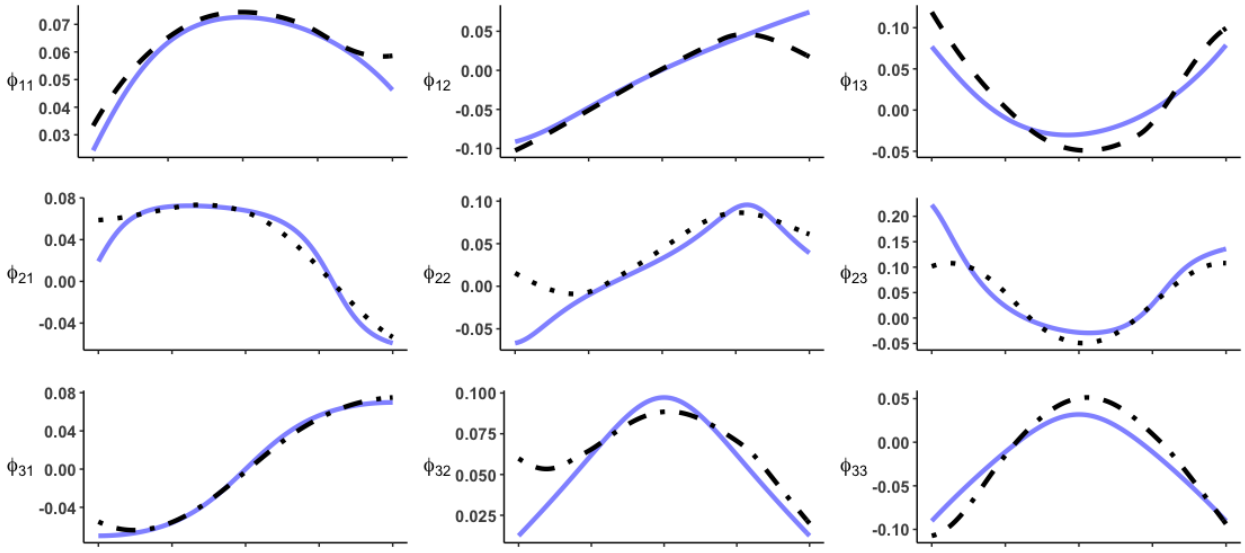

Figure 3. The point-wise average of the estimated functional principal components (FPCs) in the Simulation Study I (dashed lines). The solid lines are the true FPCs.

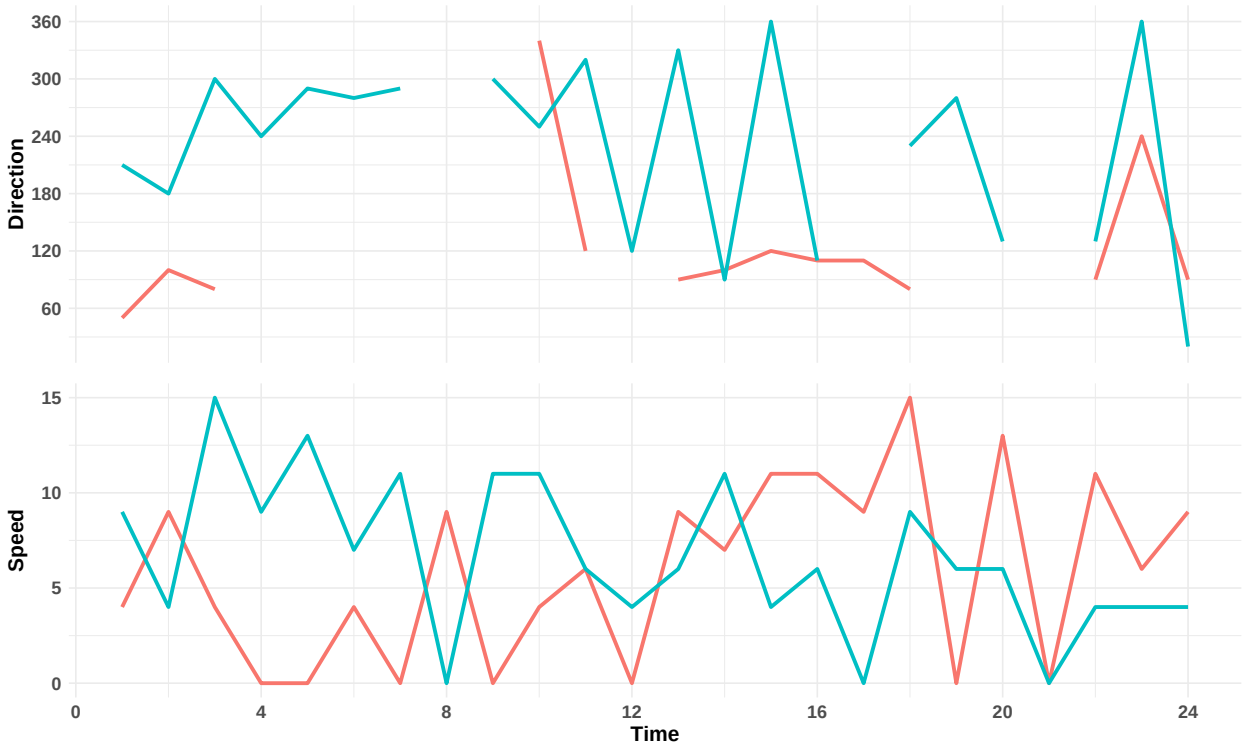

Figure 4. The two individual curves randomly selected from the hourly wind direction and speed of Vancouver in the month of December from 1998 to 2018. Each color indicates an individual curve. Missing observations are shown as the gaps in trajectories.

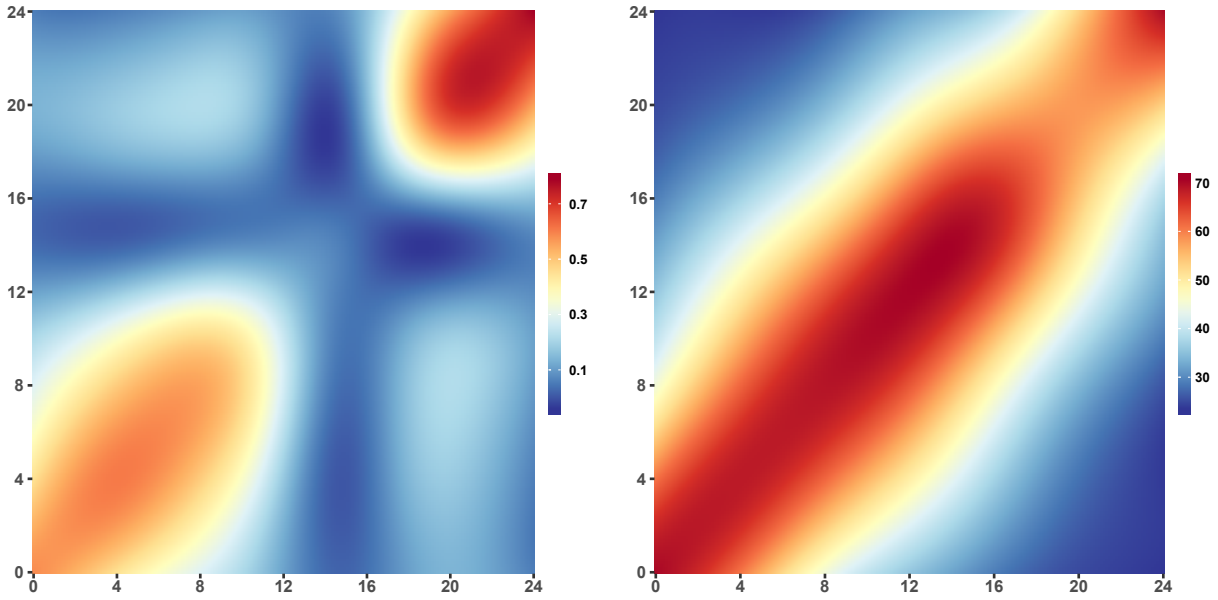

Figure 5. The covariance functions of the hourly wind direction (left panel) and speed (right panel) estimated from Vancouver's wind data in the month of December from 1998 to 2018.

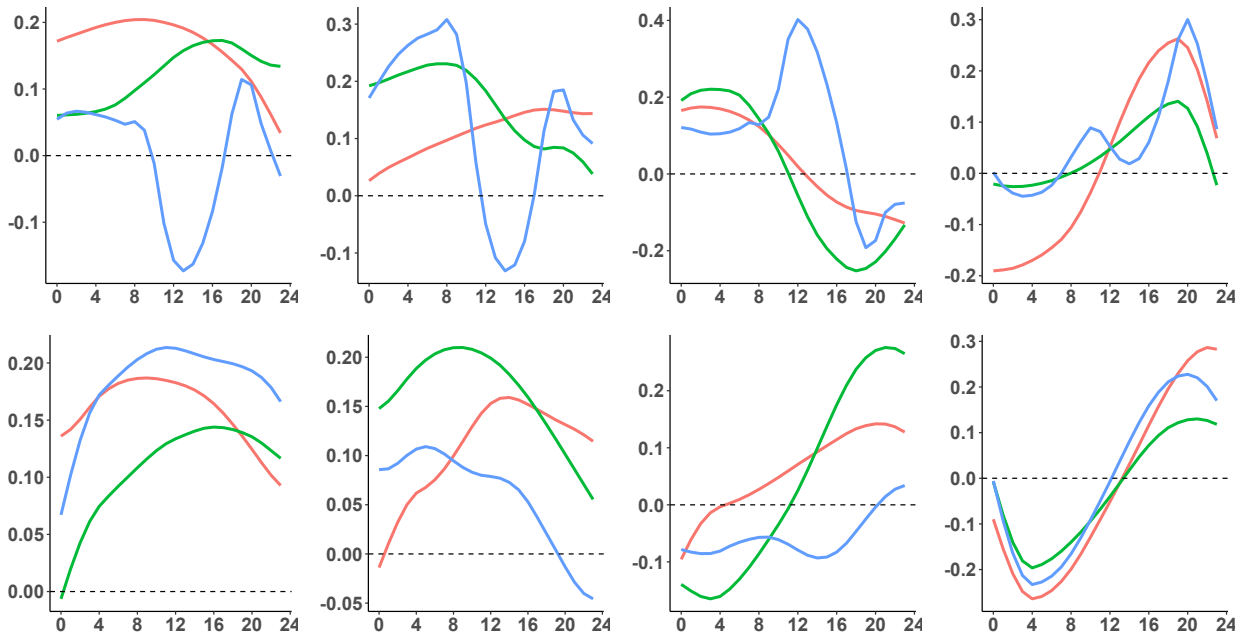

Figure 6. The top four functional principal components (FPCs) estimated by the proposed multivariate FPCA method from the wind data in Vancouver, Toronto, and Halifax. The blue, red, and green lines correspond to Vancouver, Toronto, and Halifax, respectively. The first, second, third and fourth FPCs are ordered from left to right. The top panel includes the top four FPCs for the wind direction variable, and the bottom panel presents the top four FPCs for the wind speed variable.

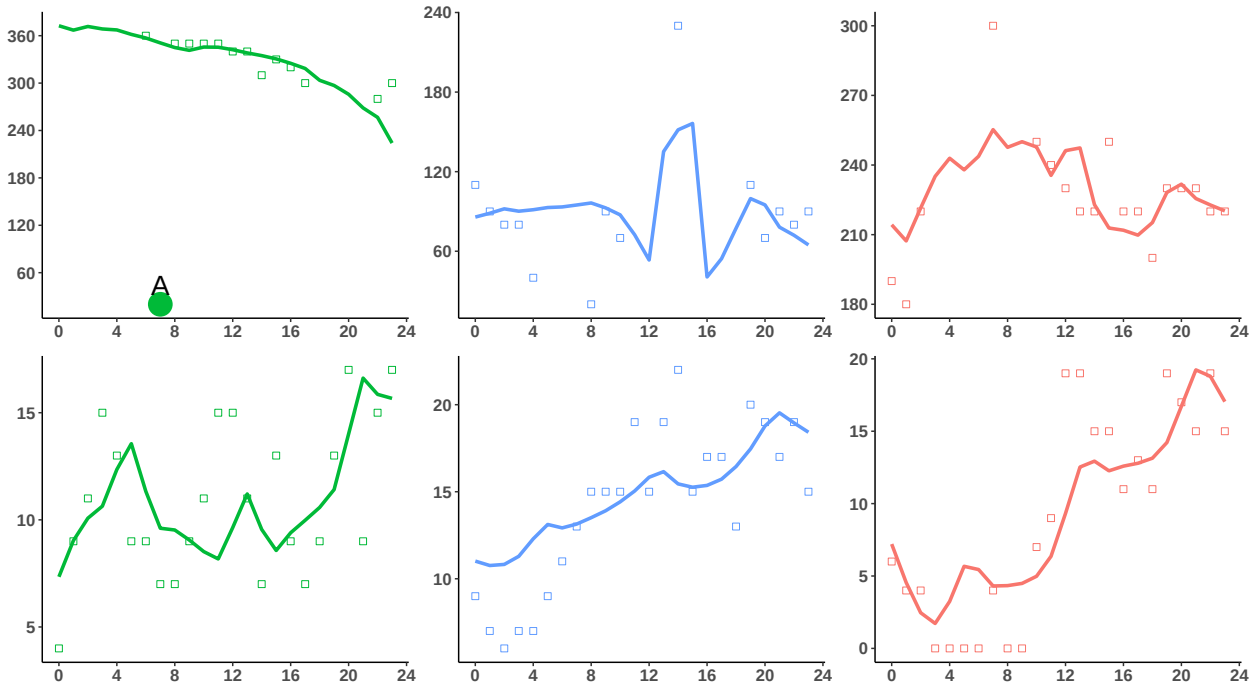

Figure 7. The estimated trajectories for the hourly wind direction and speed on December 7th, 2007 in Halifax, Vancouver and Toronto (ordered from left to right). The top panel is the wind direction, and the bottom panel corresponds to the wind speed.

Table 1
The means and standard deviations (in parentheses) of the integrated squared errors (ISEs) defined in (13) for the estimated covariance functions $G_{dd}(s,t), d = 1, 2, 3$, with various sparsity settings when the simulated data are generated by randomly sampling 10%, 20%, $\cdots$, 100% data from the full data in the Simulation Study II.

Sparsity	ISE$(G_{11}(s,t)) * 10^3$	ISE$(G_{22}(s,t)) * 10^2$	ISE$(G_{33}(s,t)) * 10$
100%	1.7(1.8)	8.2(7.3)	1.35(0.061)
90%	2.4(2.3)	8.7(8.5)	1.42(0.065)
80%	2.2(2.1)	8.6(6.8)	1.42(0.070)
70%	2.1(1.8)	8.8(8.1)	1.41(0.053)
60%	2.0(1.9)	8.7(7.6)	1.40(0.060)
50%	2.5(2.0)	8.8(7.5)	1.42(0.052)
40%	2.6(2.2)	8.8(9.7)	1.41(0.062)
30%	2.4(2.6)	8.7(9.6)	1.41(0.062)
20%	3.5(3.3)	8.7(1.1)	1.43(0.082)
10%	.3.7(1.8)	8.8(1.0)	1.41(0.120)

Table 2

The means and standard deviations (in parentheses) of the integrated squared errors (ISEs) defined in (14) for the estimated first functional principal component (FPC) for $(X_1(t), X_2(t), W(t))$ with various sparsity settings when the simulated data are generated by randomly sampling 10%, 20%, $\cdots$, 100% data from the full data in the Simulation Study II.

Sparsity	ISE(ϕ_{11})	ISE(ϕ_{12})	ISE(ϕ_{13})
100%	0.29(0.07)	1.36(0.12)	0.26(0.11)
90%	1.69(0.18)	4.62(0.10)	3.80(0.66)
80%	1.71(0.17)	4.63(0.11)	3.72(0.60)
70%	1.75(0.18)	4.65(0.10)	3.64(0.58)
60%	1.71(0.15)	4.62(0.07)	3.76(0.55)
50%	1.72(0.17)	4.64(0.11)	3.69(0.56)
40%	1.70(0.18)	4.64(0.11)	3.90(0.79)
30%	1.73(0.17)	4.64(0.11)	3.73(0.62)
20%	1.73(0.22)	4.65(0.16)	3.77(0.80)
10%	1.80(0.31)	4.70(0.53)	4.34(4.58)