Statistica Sinica **24** (2014), 1571-1596
doi:http://dx.doi.org/10.5705/ss.2013.305

MULTIVARIATE FUNCTIONAL PRINCIPAL COMPONENT ANALYSIS: A NORMALIZATION APPROACH

Jeng-Min Chiou, Yu-Ting Chen and Ya-Fang Yang

Academia Sinica

Abstract: We propose an extended version of the classical Karhunen-Loève expansion of a multivariate random process, termed a normalized multivariate functional principal component ($m\mathrm{FPC}_n$) representation. This takes variations between the components of the process into account and takes advantage of component dependencies through the pairwise cross-covariance functions. This approach leads to a single set of multivariate functional principal component scores, which serve well as a proxy for multivariate functional data. We derive the consistency properties for the estimates of the $m\mathrm{FPC}_n$, and the asymptotic distributions for statistical inferences. We illustrate the finite sample performance of this approach through the analysis of a traffic flow data set, including an application to clustering and a simulation study. The $m\mathrm{FPC}_n$ approach serves as a basic and useful statistical tool for multivariate functional data analysis.

Key words and phrases: Karhunen-Loève expansion, Mercer's theorem, multivariate functional data, normalization, traffic flow.

1. Introduction

Multivariate functional data typically comprise several simultaneously recorded time course measurements for a sample of subjects or experimental units. They are realizations sampled from multivariate random functions. Statistical methods for analysis of multivariate functional data that consider simultaneous variations of more than one random function are important, but have received relatively less attention than univariate functional data methods.

Functional principal component analysis (FPCA) has been widely used and serves as a fundamental tool for developing advanced methods for functional data analysis (FDA). Recent works that discussed the functional principal component (FPC) method and/or its asymptotic properties mostly focused on *univariate* functional data. These include, but are not limited to, approaches of the smoothed FPCA based on a roughness penalty (e.g., Rice and Silverman (1991) and Silverman (1996)), the FPC methods for sparsely sampled functional data (e.g., James, Hastie, and Sugar (2001), Yao, Müller, and Wang (2005)), and the asymptotic properties of the classical FPCA (e.g., Boente and Fraiman (2000), Hall and Hosseini-Nasab (2006), and Li and Hsing (2010)).

Methods for analyzing multivariate functional data that utilize univariate FPCA can be found sporadically. These include that the dynamical correlation analysis for multivariate longitudinal observations (Dubin and Müller (2005)), the modeling of the relationship of paired longitudinal observations (Zhou, Huang, and Carrol (2008)), regularized FPCA for multidimensional functional data (Kayano and Konishi (2009)), and linear manifold modeling to quantify functional dependence between the components of multivariate random processes (Chiou and Müller (2014)). As well, a method that repeatedly applies the classical PCA pointwise was investigated by Berrendero, Justel, and Svarc (2011).

While the development of FPCA is based on the Karhunen-Loève (K-L) representation of a random function, Balakrishnan (1960) and Kelly and Root (1960) independently derived a vector-valued version of the K-L representation of multivariate random processes that hinges on a vector-valued version of the Mercer's Theorem. Deville (1974) discussed statistical and computational methods for functional data based on the K-L representation, including extensions of the methods to vector-valued processes. Dauxois, Pousse, and Romain (1982) discussed asymptotic theory for FPCA of a vector random function, treating the random process as an operator.

When the components of multivariate functional data are measured in the same units and have similar variation, classical multivariate FPCA that concatenates the multiple functions into one to perform univariate FPCA can work well. A successful example is the bivariate FPCA of children's gait cycles (Ramsay and Silverman (2005, p.166)). However, when a particular component in the multivariate random functions has relatively large variability, the classical multivariate functional principal components (FPCs) may not do so well, as will be illustrated in Section 4.

We propose a *normalized multivariate functional principal component* ($m\mathrm{FPC}_n$) method. It accounts for differences in degrees of variability and units of measurements among the components of the multivariate random functions when defining the mFPCs. We propose to normalize each random function as a preliminary step, and discuss the normalization effects on the estimation of the multivariate FPC (mFPC) model. The multivariate approach leads to a *single* set of mFPC scores for each subject, which serves well as the proxy of *multivariate* functional data. This feature leads to natural extensions from univariate to multivariate FDA in many statistical methods. Weighted least squares estimates of the $m\mathrm{FPC}_n$ scores are suitable for dense, regular cases and for sparse, irregular designs. We derive the asymptotic properties for the $m\mathrm{FPC}_n$ scores the relevant asymptotic distributions of the predicted functions for inference purposes.

The article is organized as follows. Section 2 discusses the proposed $m\mathrm{FPC}_n$ approach. Section 3 presents the estimation of the $m\mathrm{FPC}_n$ model components. Section 4 gives a data application to traffic flow analysis, including an application to clustering and a simulation study, with additional numerical results compiled in Supplements S1-S2. Section 5 presents the asymptotic results relevant to the $m\mathrm{FPC}_n$ analysis, and Section 6 provides the technical proofs, with additional details in Supplement S3. Concluding remarks are in Section 7.

2. Multivariate Functional Principal Component Analysis

Let $\{\mathrm{X}_k\}_{k=1,\ldots,p}$ be a set of random functions with each X_k in $L_2(\mathcal{T})$, a Hilbert space of square integrable functions with respect to Lebesgue measure dt on an interval $\mathcal{T} = [a, b]$, $a < b$. Write $\mathbf{X} = (\mathrm{X}_1, \mathrm{X}_2, \ldots, \mathrm{X}_p)^\top$ as a vector in a Hilbert space of p-dimensional vectors of functions in $L_2(\mathcal{T})$, denoted by $\mathbb{H}$. We assume that $\mathbf{X}(t)$ has a continuous mean function $\boldsymbol{\mu}(t) = (\mu_1(t), \ldots, \mu_p(t))^\top$, $\mu_k(t) = \mathrm{E}(\mathrm{X}_k(t))$, and covariance function $\mathbf{G}(s,t) = \{\mathrm{G}_{kl}(s,t)\}_{1\le k,l\le p}$, $\mathrm{G}_{kl}(s,t) = \mathrm{cov}(\mathrm{X}_k(s), \mathrm{X}_l(t))$. Here $\mathbf{G}$ is symmetric in the sense that $\mathbf{G}(s,t) = \mathbf{G}(t,s)^\top$. The inner product of f and g in $L_2(\mathcal{T})$ is $\langle f, g\rangle = \int f(t)g(t)dt$ with the norm $\|\cdot\| = \langle\cdot,\cdot\rangle^{1/2}$. For $\boldsymbol{f} = (f_1, f_2, \ldots, f_p)^\top$ and $\boldsymbol{g} = (g_1, g_2, \ldots, g_p)^\top$ in $\mathbb{H}$, the inner product is $\langle \boldsymbol{f}, \boldsymbol{g}\rangle_{\mathbb{H}} = \sum_{k=1}^p \langle f_k, g_k\rangle$, and the norm $\|\cdot\|_{\mathbb{H}} = \langle\cdot,\cdot\rangle_{\mathbb{H}}^{1/2}$. More generally, the inner product is $\langle \boldsymbol{f}, \boldsymbol{g}\rangle_{\mathbb{H}} = \sum_{k=1}^p w_k\langle f_k, g_k\rangle$, where w_k is a weight associated with the kth function that needs to be pre-specified or estimated. A possible approach to this is to take the inverse of w_k as the square root of the sample functional variance (Suyundykov, Puechmorel, and Ferre (2010)). Here we adopt the inner product without weighting, and take into account heteroscedasticity between components of multivariate random functions through a modeling approach.

2.1. The $m\mathbf{FPC}_n$ method

Let $\boldsymbol{D}(t) = \mathrm{diag}(v_1(t)^{1/2}, \ldots, v_p(t)^{1/2})$, where $v_k(t) = \mathrm{G}_{kk}(t,t)$ for $k = 1, \ldots, p$, and $t \in \mathcal{T}$. We consider a stochastic representation for multivariate random functions,

$$\mathbf{X}(t) = \boldsymbol{\mu}(t) + \sum_{r=1}^{\infty} \xi_r \left(\boldsymbol{D}\boldsymbol{\phi}_r\right)(t). \tag{2.1}$$

Here the fixed components μ_k and v_k are the mean and the variance functions of X_k, and $\{\boldsymbol{\phi}_r\}_{r=1,2,\ldots}$ is a set of orthonormal basis functions in $\mathbb{H}$ satisfying

$$\langle \boldsymbol{\phi}_r, \boldsymbol{\phi}_q\rangle_{\mathbb{H}} = \sum_{l=1}^{p} \langle \phi_{lr}, \phi_{lq}\rangle = \delta_{rq}, \tag{2.2}$$

where $\boldsymbol{\phi}_r = (\phi_{1r}, \ldots, \phi_{pr})^\top$, and $\delta_{rq} = 1$ if $r = q$ and 0 otherwise. The set of random coefficients $\{\xi_r\}$ is independent of k. Observing that the random coefficients $\{\xi_r\}$ are representative of $\mathbf{X}$, the weight $\boldsymbol{D}$ in (2.1) takes the uneven extent of variations among $\{X_k\}$ into account.

If $\mathrm{Z}_k(t) = v_k(t)^{-1/2}\,\{\mathrm{X}_k(t) - \mu_k(t)\}$, then

$$\mathbf{Z}(t) = \sum_{r=1}^{\infty} \xi_r\, \boldsymbol{\phi}_r(t), \tag{2.3}$$

where $\mathbf{Z}(t) = (\mathrm{Z}_1(t), \ldots, \mathrm{Z}_p(t))^\top$. Here

$$\xi_r = \langle \mathbf{Z}, \boldsymbol{\phi}_r \rangle_{\mathbb{H}} = \sum_{k=1}^{p} \langle \mathrm{Z}_k, \phi_{kr} \rangle. \tag{2.4}$$

The expression of $\mathbf{Z}(t)$ in (2.3) is a multivariate version of the classical Karhunen-Loève representation based on the multivariate version of Mercer's Theorem (e.g., Balakrishnan (1960), Kelly and Root (1960)) that relies on a properly defined integral operator on the covariance kernel associated with the basis functions $\{\boldsymbol{\phi}_r\}$.

Let $\mathrm{C}_{kl}(s,t) = \{v_k(s)v_l(t)\}^{-1/2}\mathrm{G}_{kl}(s,t)$, where $1 \le k, l \le p$, and let $\mathbf{C}_k = (\mathrm{C}_{k1}, \ldots, \mathrm{C}_{kp})^\top$ and $\mathbf{C}(s,t) = \{\mathrm{C}_{kl}(s,t)\}$. We define an integral operator $\mathcal{A} : \mathbb{H} \to \mathbb{H}$ with the covariance kernel $\mathbf{C}(s,t)$, for any given $\boldsymbol{f} \in \mathbb{H}$, such that

$$(\mathcal{A}\boldsymbol{f})(s) = \int \mathbf{C}(s,t)\boldsymbol{f}(t)dt = \begin{pmatrix} \langle \mathbf{C}_1(s,\cdot), \boldsymbol{f} \rangle_{\mathbb{H}} \\ \vdots \\ \langle \mathbf{C}_p(s,\cdot), \boldsymbol{f} \rangle_{\mathbb{H}} \end{pmatrix}, \tag{2.5}$$

where $\langle \mathbf{C}_k(s,\cdot), \boldsymbol{f} \rangle_{\mathbb{H}} = \sum_{l=1}^{p} \langle \mathrm{C}_{kl}(s,\cdot), f_l \rangle$. Since $\mathcal{A}$ is linear, an eigenvalue λ and an eigenfunction $\boldsymbol{f}$ in $\mathbb{H}$ satisfy $(\mathcal{A}\boldsymbol{f})(s) = \lambda \boldsymbol{f}(s)$.

Since the covariance kernel $\mathbf{C}$ is continuous, symmetric, and nonnegative-definite ($\langle \mathcal{A}\boldsymbol{f}, \boldsymbol{f} \rangle_{\mathbb{H}} \ge 0$ for any $\boldsymbol{f} \in \mathbb{H}$), by the multivariate version of Mercer's theorem (cf., Withers (1974)), there exists a set of orthonormal basis functions $\boldsymbol{\phi}_r$ in $\mathbb{H}$ such that

$$\langle \mathbf{C}_k(s,\cdot), \boldsymbol{\phi}_r \rangle_{\mathbb{H}} = \sum_{l=1}^{p} \langle \mathrm{C}_{kl}(s,\cdot), \phi_{lr} \rangle = \lambda_r \phi_{kr}(s) \tag{2.6}$$

for $k = 1, \ldots, p$ and all r. Here, λ_r is the r-th eigenvalue, in non-increasing order, with the corresponding eigenfunction $\boldsymbol{\phi}_r$. With the eigen-equations (2.6) and under the constraint (2.2), we have $\mathrm{E}(\xi_r) = 0$, and $\mathrm{E}(\xi_r\xi_q) = \lambda_r\delta_{rq}$. The multivariate covariance function $\mathbf{C}$ has the representation $\mathbf{C}(s,t) = \sum_{r=1}^{\infty} \lambda_r \boldsymbol{\phi}_r(s)\boldsymbol{\phi}_r(t)^\top$, with the (k,l) element $\mathrm{C}_{kl}(s,t)$ of $\mathbf{C}(s,t)$

$$\mathrm{C}_{kl}(s,t) = \sum_{r=1}^{\infty} \lambda_r \phi_{kr}(s)\phi_{lr}(t). \tag{2.7}$$

The representation of $\mathbf{C}(s,t)$ in (2.7) converges absolutely and uniformly in both s and t. We note that the existence of $\{\boldsymbol{\phi}_r\}$ in (2.6) requires $\mathcal{A}$ in (2.5) to be a compact self-adjoint operator, based on Hilbert-Schmidt theorem (e.g., Section 7.5 in Hutson and Pym (1980)). It is easy to show that the symmetry property of $\mathbf{C}$ is a sufficient condition for the integral operator $\mathcal{A}$ in (2.5) to be self-adjoint.

2.2. Comparison with the classical approach

For comparisons, we give the classical multivariate functional principal component model ($m\mathrm{FPC}_u$). We define a linear operator $\mathcal{B}$ similar to $\mathcal{A}$, but with the covariance kernel $\mathbf{G}$ rather than $\mathbf{C}$. For $\mathbf{G}_k = (\mathrm{G}_{k1}, \ldots, \mathrm{G}_{kp})$, there exists a set of orthonormal basis functions $\boldsymbol{\psi}_r = (\psi_{1r}, \ldots, \psi_{pr})^\top$ in $\mathbb{H}$ such that $\langle \mathbf{G}_k(s,\cdot), \boldsymbol{\psi}_r \rangle_{\mathbb{H}} = \sum_{l=1}^{p} \langle \mathrm{G}_{kl}(s,\cdot), \psi_{lr} \rangle = \theta_r \psi_{kr}(s)$, for $1 \le k \le p$ and all $r \ge 1$. Here, θ_r is the r-th eigenvalue in non-increasing order with the corresponding eigenfunction $\boldsymbol{\psi}_r(s)$ satisfying $\langle \boldsymbol{\psi}_r, \boldsymbol{\psi}_q \rangle_{\mathbb{H}} = \sum_{l=1}^{p} \langle \psi_{lr}, \psi_{lq} \rangle = \delta_{rq}$. Then the random vector of functions $\mathbf{X}$ has the representation, based on the covariance kernel $\mathbf{G}$,

$$\mathbf{X}(t) = \boldsymbol{\mu}(t) + \sum_{r=1}^{\infty} \zeta_r \boldsymbol{\psi}_r(t), \tag{2.8}$$

where the random coefficients are $\zeta_r = \langle \mathbf{X} - \boldsymbol{\mu}, \boldsymbol{\psi}_r \rangle_{\mathbb{H}} = \sum_{k=1}^{p} \langle \mathrm{X}_k - \mu_k, \psi_{kr} \rangle$, with $E(\zeta_r) = 0$ and $E(\zeta_r \zeta_q) = \theta_r \delta_{rq}$.

The random coefficients $\{\xi_r\}$ in (2.1) minimize $\|\mathbf{Z} - \sum_{r=1}^{\infty} \xi_r \boldsymbol{\phi}_r\|_{\mathbb{H}}^2$ with respect to ξ_r, whereas $\{\zeta_r\}$ in (2.8) minimize of $\|\mathbf{X} - \boldsymbol{\mu} - \sum_{r=1}^{\infty} \zeta_r \boldsymbol{\psi}_r\|_{\mathbb{H}}^2$ with respect to ζ_r. Here ζ_r could be influenced by a particular term $\langle \mathrm{X}_l - \mu_l, \psi_{lr} \rangle$ for some l, which dominates the other terms. Each random function may require a different number of FPCs, due to different degrees of variabilities among the multivariate random functions. The random coefficients $\{\xi_r\}$ for the truncated expansion of (2.1) and (2.3), taking into account the discrepant amount of variabilities among the random functions $\{X_k\}$ through $\{v_k\}$, well represents the multivariate random function $\mathbf{Z}$ in practice.

We demonstrate in Section 4 that the proposed $m\mathrm{FPC}_n$ approach (2.1) performs better than the classical $m\mathrm{FPC}_u$ approach (2.8) in approximating $\mathbf{X}$ in terms of prediction errors, even though the variance functions are unknown and need to be estimated.

3. Estimation and prediction of multivariate trajectories

Suppose $\{\mathbf{X}_i\}_{i=1,\ldots,n}$ are sampled from a stochastic process $\mathbf{X}$ in $\mathbb{H}$, $\mathbf{X}_i(t) = (\mathrm{X}_{1i}(t), \ldots, \mathrm{X}_{pi}(t))^\top$. Each $\mathbf{X}_i(t)$ is associated with $\mathbf{Z}_i(t) = (\mathrm{Z}_{1i}(t), \ldots, \mathrm{Z}_{pi}(t))^\top$, where $\mathrm{Z}_{ki}(t) = v_k(t)^{-1/2} \{\mathrm{X}_{ki}(t) - \mu_k(t)\}$. Let $\mathbf{Y}_{ij} = (\mathrm{Y}_{1ij}, \ldots, \mathrm{Y}_{pij})^\top$ be the jth observation of the ith subject observed at T_{ij}, contaminated with measurement errors $\boldsymbol{\epsilon}_{ij}$, such that

$$\mathbf{Y}_{ij} = \mathbf{X}_i(T_{ij}) + \boldsymbol{\epsilon}_{ij} = \boldsymbol{\mu}(T_{ij}) + \sum_{r=1}^{\infty} \xi_{ri} \left\{(\boldsymbol{D}\boldsymbol{\phi}_r)(T_{ij})\right\} + \boldsymbol{\epsilon}_{ij}, \tag{3.1}$$

where $\boldsymbol{\mu}(\cdot)$, $(\boldsymbol{D}\boldsymbol{\phi}_r)(\cdot)$, and $\{\xi_{ri}\}$ are as before, and $\boldsymbol{\epsilon}_{ij} = (\epsilon_{1ij}, \ldots, \epsilon_{pij})^\top$ are mutually independent with mean 0 and variance $\boldsymbol{\sigma}^2 = \left(\sigma_1^2, \ldots, \sigma_p^2\right)^\top$. Here, the recording times T_{ij} are observed and treated as fixed scalars, sampled from a density function $f_T(\cdot)$. Similarly, each $\mathbf{Y}_{ij}$ is associated with an $\mathbf{U}_{ij} = (\mathrm{U}_{1ij}, \ldots, \mathrm{U}_{pij})^\top$ with $\mathrm{U}_{kij} = v_k(T_{ij})^{-1/2}\left(\mathrm{Y}_{kij} - \mu_k(T_{ij})\right)$ and $\mathbf{U}_{ij} = \mathbf{Z}_i(T_{ij}) + \boldsymbol{\varepsilon}_{ij} = \sum_{r=1}^{\infty} \xi_{ri}\boldsymbol{\phi}_r(T_{ij}) + \boldsymbol{\varepsilon}_{ij}$, where $\boldsymbol{\varepsilon}_{ij} = (\varepsilon_{1ij}, \ldots, \varepsilon_{pij})^\top$ are mutually independent with mean 0 and variance $\boldsymbol{\varsigma}_{ij}^2 = \left(\varsigma_{1ij}^2, \ldots, \varsigma_{pij}^2\right)^\top$, $\varsigma_{kij}^2 = \sigma_k^2 / v_k(T_{ij})$.

3.1. Estimation of the fixed model components

Let $K : \mathbb{R} \to \mathbb{R}$ be a symmetric kernel density functions with support $[-1, 1]$ for the smoothing procedures. In estimation of μ_k, we apply local linear regression (see, e.g., Fan and Gijbels (1996)) with the bandwidth b_{μ_k} to the pooled data $\{(T_{ij}, \mathrm{Y}_{kij}); i = 1, \ldots, n, j = 1, \ldots, m_i\}$ for each k, such that $\hat{\mu}_k(t) = \hat{\alpha}_0$, where

$$(\hat{\alpha}_0, \hat{\alpha}_1) = \underset{(\alpha_0, \alpha_1)}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{m_i} \sum_{j=1}^{m_i} \{\mathrm{Y}_{kij} - \alpha_0 - \alpha_1(t - T_{ij})\}^2 K\left(\frac{T_{ij} - t}{b_{\mu_k}}\right). \tag{3.2}$$

Here the weighting adjustment $(1/m_i)$ follows the approach of Li and Hsing (2010).

To estimate the variance function $v_k(t)$, let $\widetilde{\mathrm{G}}_{kk}(T_{ij}, T_{ij'}) = \{\mathrm{Y}_{kij} - \hat{\mu}_k(T_{ij})\}\{\mathrm{Y}_{kij'} - \hat{\mu}_k(T_{ij'})\}$ be a raw covariance estimate for $1 \le k \le p$. We apply the pooled data $\{(T_{ij}, T_{ij'}, \widetilde{\mathrm{G}}_{kk}(T_{ij}, T_{ij'}))\}$ to fit a local linear plane with the bandwidth (b_{G_k}, b_{G_k}) such that $\hat{\mathrm{G}}_{kk}(s, t) = \hat{\beta}_0$, where $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$ is the minimizer of

$$\frac{1}{n} \sum_{i=1}^{n} \frac{1}{M_i} \sum_{1 \le j \ne j' \le m_i} \left\{\widetilde{\mathrm{G}}_{kk}(T_{ij}, T_{ij'}) - \beta_0 - \beta_1(s - T_{ij}) - \beta_2(t - T_{ij'})\right\}^2 K\left(\frac{T_{ij} - s}{b_{\mathrm{G}_k}}\right) K\left(\frac{T_{ij'} - t}{b_{\mathrm{G}_k}}\right), \tag{3.3}$$

where $M_i = m_i(m_i - 1)$ is the number of pairs (j, j') in the summation for $1 \le j \ne j' \le m_i$, assuming $m_i \ge 2$. We then obtain the variance function estimates $\hat{v}_k(t) = \hat{\mathrm{G}}_{kk}(t, t)$ and the covariance function estimates $\hat{\mathrm{C}}_{kk}(s, t) = \{\hat{v}_k(s)\hat{v}_k(t)\}^{-1/2}\, \hat{\mathrm{G}}_{kk}(s, t)$.

To estimate the cross-covariance functions $\mathrm{C}_{kl}(s, t)$ for $k \ne l$, let

$$\widetilde{\mathrm{U}}_{kij} = \frac{\{\mathrm{Y}_{kij} - \hat{\mu}_k(T_{ij})\}}{\hat{v}_k(T_{ij})^{1/2}}, \tag{3.4}$$

and $\widetilde{\mathrm{C}}_{kl}(T_{ij}, T_{ij'}) = \widetilde{\mathrm{U}}_{kij}\widetilde{\mathrm{U}}_{lij'}$. We apply the pooled data $\{(T_{ij}, T_{ij'}, \widetilde{\mathrm{C}}_{kl}(T_{ij}, T_{ij'}))\}$ to fit a local linear plane with bandwidth (h_k, h_l) such that $\hat{\mathrm{C}}_{kl}(s, t) = \hat{\gamma}_0$, where

$(\hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_2)$ is the minimizer of

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{M_i}\sum_{1\le j\ne j'\le m_i}\left\{\widetilde{\mathrm{C}}_{kl}(T_{ij}, T_{ij'})-\gamma_0-\gamma_1(s-T_{ij})-\gamma_2(t-T_{ij'})\right\}^2 K\left(\frac{T_{ij}-s}{h_k}\right)K\left(\frac{T_{ij'}-t}{h_l}\right). \tag{3.5}$$

Given $\hat{\mathbf{C}}(s,t) = \{\hat{\mathrm{C}}_{kl}(s,t)\}_{1\le k,l\le p}$ and $\hat{\mathbf{C}}_k(s,t) = (\hat{\mathrm{C}}_{k1}(s,t),\ldots,\hat{\mathrm{C}}_{kp}(s,t))^\top$ obtained from (3.3) and (3.5), the estimates of the eigenvalues and eigenfunctions correspond to the discrete approximations of the solutions to the eigenequations

$$\langle \hat{\mathbf{C}}_k(s,\cdot), \hat{\boldsymbol{\phi}}_r\rangle_{\mathbb{H}} = \hat{\lambda}_r\hat{\phi}_{kr}(s), \tag{3.6}$$

subject to $\langle \hat{\boldsymbol{\phi}}_r, \hat{\boldsymbol{\phi}}_q\rangle_{\mathbb{H}} = \sum_{l=1}^{p}\langle\hat{\phi}_{lr}, \hat{\phi}_{lq}\rangle = \delta_{rq}$. The eigenfunction estimates are approximated by discretizing the smooth variance-covariance functions.

For the estimates of the measurement error variances σ_k^2, the diagonal values of the covariance $\mathrm{G}_{kk}(t,t)$ plus the constant measurement error variance σ_k^2, $\mathrm{W}_k(t) = \mathrm{G}_{kk}(t,t) + \sigma_k^2$, can be estimated by the smooth estimate of the data $\left\{(T_{ij}, \widetilde{\mathrm{G}}_{kk}(T_{ij}, T_{ij}))\right\}$ through a local linear regression with bandwidth h_{W_k} such that $\hat{\mathrm{W}}_k(t) = \hat{\eta}_0$, where $(\hat{\eta}_0, \hat{\eta}_1)$ is the minimizer of

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}\sum_{1\le j\le m_i}\left\{\widetilde{\mathrm{G}}_{kk}(T_{ij}, T_{ij})-\eta_0-\eta_1(t-T_{ij})\right\}^2 K\left(\frac{T_{ij}-t}{h_{\mathrm{W}_k}}\right). \tag{3.7}$$

Following Yao, Müller, and Wang (2005), the estimate of σ_k^2, where $k = 1,\ldots,p$, can be obtained as

$$\hat{\sigma}_k^2 = \frac{2}{|\mathcal{T}|}\int_{\mathcal{T}_1}\left\{\hat{\mathrm{W}}_k(t) - \hat{\mathrm{G}}_{kk}(t,t)\right\}dt, \tag{3.8}$$

where $|\mathcal{T}|$ denote the length of $\mathcal{T}$ and $\mathcal{T}_1$ is the interval $\mathcal{T}_1 = [\inf\{x : x \in \mathcal{T}\} + |\mathcal{T}|/4, \sup\{x : x\in\mathcal{T}\} - |\mathcal{T}|/4]$. For the estimate of ς_{kij}^2, we simply plug in the estimates of σ_k^2 and $v_k(T_{ij})$ to obtain

$$\hat{\varsigma}_{kij}^2 = \frac{\hat{\sigma}_k^2}{\hat{v}_k(T_{ij})}. \tag{3.9}$$

When applying one- and two-dimensional local polynomial regression methods to obtain the estimates, we need to choose the bandwidths. One way to choose the bandwidths data-adaptively is to apply the leave-one-subject-out cross-validation method (Rice and Silverman (1991)). In numerical studies, we use the generalized cross-validation method to choose the bandwidths.

3.2. Estimation of $m\mathbf{FPC}_n$ scores

In practice, the infinite series in the expansions on the right-hand-side of (2.1) and (2.3) are truncated at L, chosen data-adaptively. In this study, we choose

L as the minimal number of components such that the leading L FPCs explains $100\delta_0\%$ of the total variability. With L determined, we assume that the remaining term $\sup_{t\in\mathcal{T}}\sum_{r=L+1}^{\infty}\xi_{ri}\boldsymbol{\phi}_r(t)$ is negligible or is confounded with the measurement error $\boldsymbol{\varepsilon}_{ij}$. Let $\boldsymbol{\xi}_{i,L}=(\xi_{1i},\ldots,\xi_{Li})^\top$, where $\xi_{ri}=\langle\mathbf{Z}_i,\boldsymbol{\phi}_r\rangle_{\mathbb{H}}$. While $\boldsymbol{\phi}_r$ can be estimated, $\mathbf{Z}_i$ is observed through $\{\widetilde{\mathrm{U}}_{kij}\}$ and contaminated with measurement errors. To estimate $\boldsymbol{\xi}_{i,L}$, let $\widetilde{\mathbf{U}}_i=(\widetilde{\mathbf{U}}_{1i}^\top,\ldots,\widetilde{\mathbf{U}}_{pi}^\top)^\top$, where $\widetilde{\mathbf{U}}_{ki}=(\widetilde{\mathrm{U}}_{ki1},\ldots,\widetilde{\mathrm{U}}_{kim_i})^\top$, and let $\widetilde{\boldsymbol{\varphi}}_{i,L}=(\widetilde{\boldsymbol{\phi}}_{1i},\widetilde{\boldsymbol{\phi}}_{2i},\ldots,\widetilde{\boldsymbol{\phi}}_{Li})^\top$, where $\widetilde{\boldsymbol{\phi}}_{ri}=(\widetilde{\phi}_{1ri}^\top,\ldots,\widetilde{\phi}_{pri}^\top)^\top$ and $\widetilde{\phi}_{kri}=(\phi_{kr}(T_{i1}),\ldots,\phi_{kr}(T_{im_i}))^\top$. Further, let $\widetilde{\boldsymbol{\varepsilon}}_i=(\widetilde{\boldsymbol{\varepsilon}}_{1i}^\top,\ldots,\widetilde{\boldsymbol{\varepsilon}}_{pi}^\top)^\top$ with $\widetilde{\boldsymbol{\varepsilon}}_{ki}=(\varepsilon_{ki1},\ldots,\varepsilon_{kim_i})^\top$, and $\boldsymbol{\Gamma}_i=\mathrm{cov}(\widetilde{\boldsymbol{\varepsilon}}_i)=\mathrm{diag}\left((\boldsymbol{\varsigma}_{1i}^2)^\top,\ldots,(\boldsymbol{\varsigma}_{pi}^2)^\top\right)$ with $\boldsymbol{\varsigma}_{ki}^2=(\varsigma_{ki1}^2,\ldots,\varsigma_{kim_i}^2)^\top$. Given $\widetilde{\mathbf{U}}_i$, $\widetilde{\boldsymbol{\varphi}}_{i,L}$ and $\boldsymbol{\Gamma}_i$, we consider the weighted least squares (WLS) estimate of $\boldsymbol{\xi}_{i,L}$, minimizing

$$(\widetilde{\mathbf{U}}_i-\widetilde{\boldsymbol{\varphi}}_{i,L}^\top\,\boldsymbol{\xi}_{i,L})^\top\boldsymbol{\Gamma}_i^{-1}(\widetilde{\mathbf{U}}_i-\widetilde{\boldsymbol{\varphi}}_{i,L}^\top\,\boldsymbol{\xi}_{i,L}),$$

such that $\boldsymbol{\xi}_{i,L}^{WLS}=(\xi_{1i}^{WLS},\xi_{2i}^{WLS},\ldots,\xi_{Li}^{WLS})^\top=\boldsymbol{\Psi}_i^{-1}\widetilde{\boldsymbol{\varphi}}_{i,L}\boldsymbol{\Gamma}_i^{-1}\widetilde{\mathbf{U}}_i$, where $\boldsymbol{\Psi}_i=\widetilde{\boldsymbol{\varphi}}_{i,L}\boldsymbol{\Gamma}_i^{-1}\widetilde{\boldsymbol{\varphi}}_{i,L}^\top$.

Let $\hat{\boldsymbol{\phi}}_{ri}=(\hat{\phi}_{1ri}^\top,\ldots,\hat{\phi}_{pri}^\top)^\top$ with $\hat{\phi}_{kri}=(\hat{\phi}_{kr}(T_{i1}),\ldots,\hat{\phi}_{kr}(T_{im_i}))^\top$. By substituting $\widetilde{\boldsymbol{\phi}}_{ri}$ in $\widetilde{\boldsymbol{\varphi}}_{i,L}$ with $\hat{\boldsymbol{\phi}}_{ri}$ and ς_{kij}^2 in $\boldsymbol{\Gamma}_i$ with $\hat{\varsigma}_{kij}^2$, we obtain $\hat{\boldsymbol{\Gamma}}_i=\mathrm{diag}((\hat{\boldsymbol{\varsigma}}_{1i}^2)^\top,\ldots,(\hat{\boldsymbol{\varsigma}}_{pi}^2)^\top)$ with $\hat{\boldsymbol{\varsigma}}_{ki}^2=(\hat{\varsigma}_{ki1}^2,\ldots,\hat{\varsigma}_{kim_i}^2)^\top$, and $\hat{\boldsymbol{\Psi}}_i=\hat{\boldsymbol{\varphi}}_{i,L}\hat{\boldsymbol{\Gamma}}_i^{-1}\hat{\boldsymbol{\varphi}}_{i,L}^\top$. The estimate of $\boldsymbol{\xi}_{i,L}^{WLS}$ is

$$\hat{\boldsymbol{\xi}}_{i,L}^{WLS}=\hat{\boldsymbol{\Psi}}_i^{-1}\hat{\boldsymbol{\varphi}}_{i,L}\hat{\boldsymbol{\Gamma}}_i^{-1}\widetilde{\mathbf{U}}_i. \tag{3.10}$$

Alternatively, we can extend the conditional expectation (CE) approach of Yao, Müller, and Wang (2005) from the univariate to the proposed $m\mathrm{FPC}_n$ model, which is especially useful when data are very sparse. By assuming that the scores ξ_{ri} and the measurement errors $\boldsymbol{\varepsilon}_{ij}$ are jointly Gaussian, the $m\mathrm{FPC}_n$ scores obtained by the conditional expectation $\mathrm{E}(\boldsymbol{\xi}_{i,L}|\widetilde{\mathbf{U}}_i)$ are $\boldsymbol{\xi}_{i,L}^{CE}=(\xi_{1i}^{CE},\xi_{2i}^{CE},\ldots,\xi_{Li}^{CE})^\top=\mathbf{H}_i\boldsymbol{\Sigma}_{\mathbf{U}_i}^{-1}\widetilde{\mathbf{U}}_i$, where $\mathbf{H}_i=(\lambda_1\widetilde{\boldsymbol{\phi}}_{1i},\lambda_2\widetilde{\boldsymbol{\phi}}_{2i},\ldots,\lambda_L\widetilde{\boldsymbol{\phi}}_{Li})^\top$ and the (k,l)th $m_i\times m_i$ block element of $\boldsymbol{\Sigma}_{\mathbf{U}_i}$ is $\{\mathrm{C}_{kl}(T_{ij},T_{ir})+\varsigma_{kij}^2\delta_{kljr}\}_{1\le j,r\le m_i}$, with δ_{kljr} being 1 for $k=l$ and $j=r$ and 0 otherwise. Substituting the unknown quantities λ_r and $\tilde{\boldsymbol{\phi}}_{ri}$, C_{kl}, and ς_{kij}^2 with $\hat{\lambda}_r$ and $\hat{\boldsymbol{\phi}}_{ri}$, $\hat{\mathrm{C}}_{kl}$, and $\hat{\varsigma}_{kij}^2$ in (3.3), (3.5), and (3.9) leads to the estimate

$$\hat{\boldsymbol{\xi}}_{i,L}^{CE}=\hat{\mathbf{H}}_i\hat{\boldsymbol{\Sigma}}_{\mathbf{U}_i}^{-1}\widetilde{\mathbf{U}}_i. \tag{3.11}$$

3.3. Prediction of multivariate random trajectories

Let ϑ be a superscript indicating either the WLS or CE approach used to estimate the $m\mathrm{FPC}_n$ scores. The predicted random functions $\mathbf{Z}_i$ and $\mathbf{X}_i$ for subject i with the leading L random components are

$$\hat{\mathbf{Z}}_i^{L,\vartheta}(t) = \sum_{r=1}^{L} \hat{\xi}_{ri}^{\vartheta} \hat{\boldsymbol{\phi}}_r(t), \tag{3.12}$$

$$\hat{\mathbf{X}}_i^{L,\vartheta}(t) = \hat{\boldsymbol{\mu}}(t) + \sum_{r=1}^{L} \hat{\xi}_{ri}^{\vartheta} \, (\hat{\boldsymbol{D}} \hat{\boldsymbol{\phi}}_r)(t), \tag{3.13}$$

where $\hat{\boldsymbol{D}}(t) = \mathrm{diag}\left(\hat{v}_1(t)^{1/2}, \ldots, \hat{v}_p(t)^{1/2}\right)^\top$.

Let $\boldsymbol{\xi}_{i,L}^{\vartheta} = (\xi_{1i}^{\vartheta}, \ldots, \xi_{Li}^{\vartheta})^\top$ be the vector of the $m\mathrm{FPC}_n$ scores for the L leading components based on ϑ= WLS or CE. For the WLS estimates, the covariance of $(\boldsymbol{\xi}_{i,L}^{WLS} - \boldsymbol{\xi}_{i,L})$ is $\boldsymbol{\Omega}_{i,L}^{WLS} = \boldsymbol{\Psi}_i^{-1}$. By extending the CE approach from the univariate to the multivariate setting, we obtain the covariance of $(\boldsymbol{\xi}_{i,L}^{CE} - \boldsymbol{\xi}_{i,L})$, $\boldsymbol{\Omega}_{i,L}^{CE} = \boldsymbol{\Lambda} - \mathbf{H}_i \boldsymbol{\Sigma}_{\mathbf{U}_i}^{-1} \mathbf{H}_i^\top$, where $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_L)$.

Substituting the unknown quantities λ_r, $\tilde{\boldsymbol{\phi}}_{ri}$, C_{kl}, and ς_{kij}^2 with $\hat{\lambda}_r$, $\hat{\boldsymbol{\phi}}_{ri}$, $\hat{\mathrm{C}}_{kl}$, and $\hat{\varsigma}_{kij}^2$ leads to the estimates $\hat{\boldsymbol{\Omega}}_{i,L}^{WLS} = \hat{\boldsymbol{\Psi}}_i^{-1}$ and $\hat{\boldsymbol{\Omega}}_{i,L}^{CE} = \hat{\boldsymbol{\Lambda}} - \hat{\mathbf{H}}_i \hat{\boldsymbol{\Sigma}}_{\mathbf{U}_i}^{-1} \hat{\mathbf{H}}_i^\top$. Let $\boldsymbol{\phi}_{L,t} = (\boldsymbol{\phi}_1(t), \ldots, \boldsymbol{\phi}_L(t))^\top$, $\hat{\boldsymbol{\phi}}_{L,t} = (\hat{\boldsymbol{\phi}}_1(t), \ldots, \hat{\boldsymbol{\phi}}_L(t))^\top$, and $\mathbf{Z}_i^L(t) = \boldsymbol{\phi}_{L,t}^\top \boldsymbol{\xi}_{i,L}$, $\hat{\mathbf{Z}}_i^{L,\vartheta}(t) = \hat{\boldsymbol{\phi}}_{L,t}^\top \hat{\boldsymbol{\xi}}_{i,L}^{\vartheta}$. The distribution of $\left\{\hat{\mathbf{Z}}_i^{L,\vartheta}(t) - \mathbf{Z}_i^L(t)\right\}$ is approximately normal $N(\mathbf{0}, \hat{\boldsymbol{\phi}}_{L,t}^\top \hat{\boldsymbol{\Omega}}_{i,L}^{\vartheta} \hat{\boldsymbol{\phi}}_{L,t})$ for large n. More details will be given in Section 5. Given a scalar p-vector $\boldsymbol{a} = (a_1, \ldots, a_p)^\top$, the $(1-\alpha)$ asymptotic pointwise confidence intervals and simultaneous confidence bands (with respect to time t) for $\boldsymbol{a}^\top \hat{\mathbf{Z}}_i^{L,\vartheta}(t)$ can be obtained by the arguments for the univariate case of Yao, Müller, and Wang (2005, Sec. 2.4),

$$[\underline{\boldsymbol{z}}^{\vartheta}, \overline{\boldsymbol{z}}^{\vartheta}] = \boldsymbol{a}^\top \hat{\mathbf{Z}}_i^{L,\vartheta}(t) \pm \psi^{-1}\left(1 - \frac{\alpha}{2}\right) \left\{\boldsymbol{a}^\top \ \hat{\boldsymbol{\phi}}_{L,t}^\top \ \hat{\boldsymbol{\Omega}}_L^{\vartheta} \ \hat{\boldsymbol{\phi}}_{L,t} \ \boldsymbol{a}\right\}^{1/2}, \tag{3.14}$$

$$[\underline{\boldsymbol{z}}_B^{\vartheta}, \overline{\boldsymbol{z}}_B^{\vartheta}] = \boldsymbol{a}^\top \hat{\mathbf{Z}}_i^{L,\vartheta}(t) \pm \left\{\mathcal{X}_{L,1-\alpha}^2 \boldsymbol{a}^\top \ \hat{\boldsymbol{\phi}}_{L,t}^\top \ \hat{\boldsymbol{\Omega}}_L^{\vartheta} \ \hat{\boldsymbol{\phi}}_{L,t} \ \boldsymbol{a}\right\}^{1/2}, \tag{3.15}$$

where $\psi^{-1}(1-\alpha/2)$ is the $(1-\alpha/2)$th percentile of the standard Gaussian distribution and $\mathcal{X}_{L,1-\alpha}^2$ is the $(1-\alpha)$th percentile of the chi-squared distribution with L degrees of freedom. In particular, when taking $\boldsymbol{a} = \mathbf{e}_k$ as a unit vector whose k-th component is one, (3.14) and (3.15) reduce to the standard univariate case for component Z_k. The $(1-\alpha)$ asymptotic point wise confidence intervals and simultaneous confidence bands for X_k can be approximated by $[\hat{\mu}_k(t) + \underline{z}^{\vartheta} \hat{v}_k^{1/2}(t), \hat{\mu}_k(t) + \overline{z}^{\vartheta} \hat{v}_k^{1/2}(t)]$ and $[\hat{\mu}_k(t) + \underline{z}_B^{\vartheta} \hat{v}_k^{1/2}(t), \hat{\mu}_k(t) + \overline{z}_B^{\vartheta} \hat{v}_k^{1/2}(t)])$, respectively.

4. Application to Traffic Flow Analysis and Simulation

In road traffic monitoring, vehicle speed (average speed in kilometers per hour), flow rate (vehicle count per unit time), and occupancy (time percentage

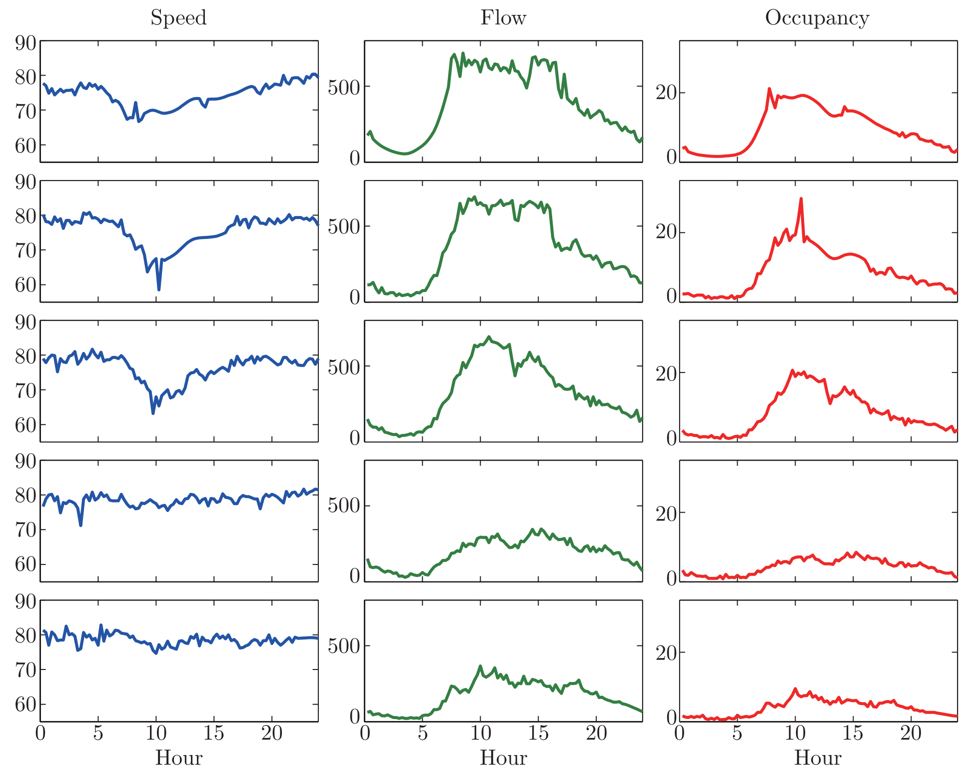

Figure 1. Samples of observed daily speed-flow-occupancy trajectories.

of a unit length of roadway occupied by a vehicle) are the basic quantities that provide input to intelligent transportation systems. These quantities are continuously recorded by vehicle detectors that are installed in the pavement of selected roads at regular intervals, the most common type of vehicle detector being an inductive loop. Vehicle speed, flow rate, and occupancy then form a triplet of multivariate random functions that play a central role for describing the traffic stream.

Large sets of such multivariate functional data are constantly generated for road networks. We specifically analyze data that were recorded by a single dual-loop detector on Highway 5 in Taiwan, for 92 days (3 months) in 2009. We take X_1 as speed, X_2 as flow and X_3 as occupancy. Figure 1 illustrates 5 randomly selected samples of daily traffic monitoring profiles, illustrating the three components corresponding to continuously observed traffic characteristics for one 24-hour day. There is substantial random variation across these trajectories. The notion of taking daily traffic flow rates as realizations of a stochastic process was adopted for traffic flow prediction by Chiou (2012). We notice that the profiles of flow and occupancy give rise to similar shapes, while that of speed stands apart.

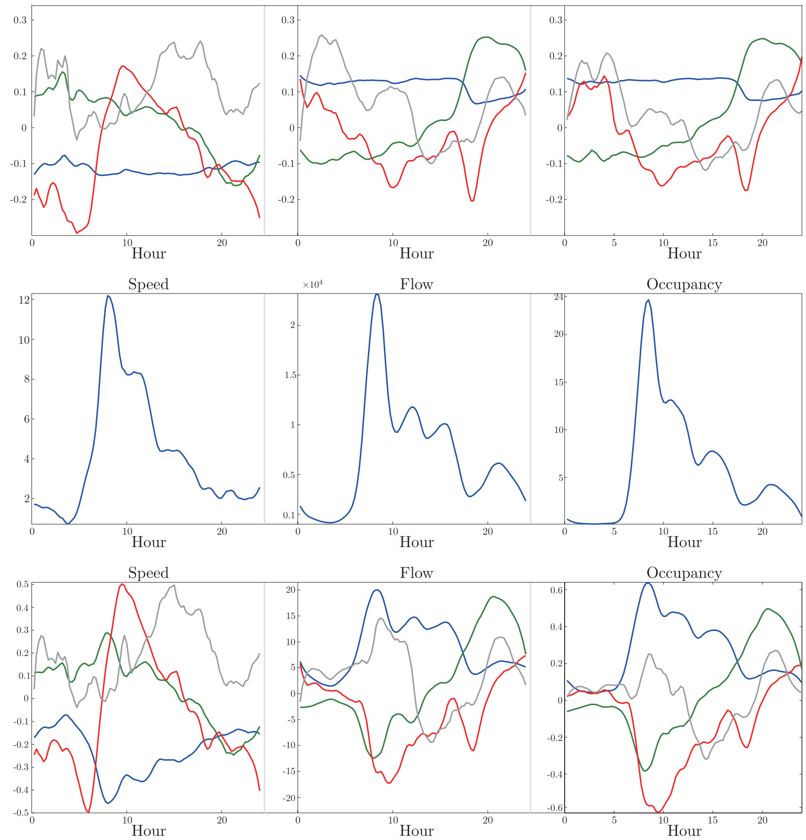

Figure 2. Upper: Estimates of the eigenfunctions $\{\hat{\phi}_{kr}\}$ in (3.6) for $r = 1$ (blue), $r = 2$ (green), $r = 3$ (red) and $r = 4$ (gray). Middle: Estimates of the variance functions $\hat{v}_k$. Bottom: Direction of variation adjusted by variance function, $\hat{v}_k^{1/2}\hat{\phi}_{kr}$.

4.1. The $m\mathbf{FPC}_n$ analysis and application to clustering multivariate functional data

The five leading FPCs based on $m\text{FPC}_n$, accounting for 67%, 15%, 5%, 2% and 1%, explain just above 90% of the total variance. Figure 2 displays the first four eigenfunctions in (3.6) and the estimated variance functions for each of the variables. While the directions of variation in flow and occupancy behave similarly, variation in speed is in the opposite direction. The first and the second eigenfunctions show a contrast to each other between day and night hour traffic for the three parameters. The overall shapes of the variance function estimates look similar, with the peak of variation occurring around 8 a.m. (middle panels). The bottom panels display the relative directions of variations $\hat{v}_k^{1/2}\hat{\phi}_{kr}$, the direction of variation adjusted by the effect of the variance function. Figure 3

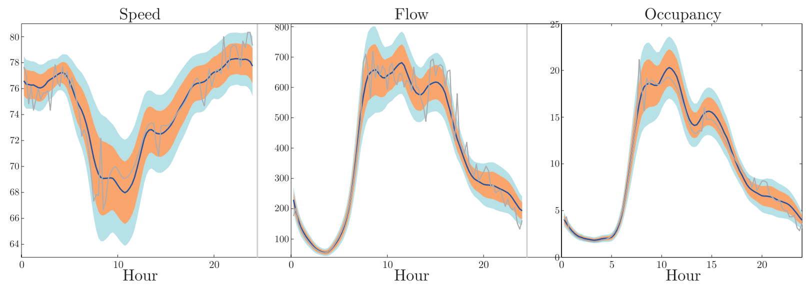

Figure 3. The predicted curves, superimposed on the observed trajectories, with the asymptotic 95% pointwise confidence intervals (shaded in orange) and simultaneous confidence bands (shaded in cyan) for a random sample of multivariate traffic trajectories, based on $m\mathrm{FPC}_n$.

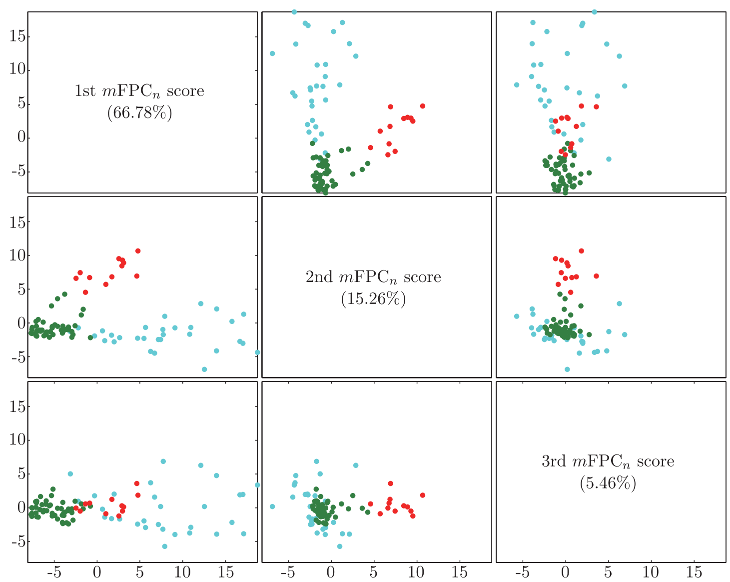

Figure 4. Pairwise scatterplot of the first three $m\mathrm{FPC}_n$ scores marked for holiday (blue), weekday (green) and weekday-before-holiday (red).

shows the 95% pointwise confidence intervals and the simultaneous confidence bands as obtained in (3.14) and (3.15) for a randomly selected observation. The pairwise scatterplots of the three leading $m\mathrm{FPC}_n$ scores are displayed in Figure 4. We mark the $m\mathrm{FPC}_n$ scores for holidays (blue), weekdays (green) and weekday-before-holidays (red), indicating three typical daily traffic patterns.

The information of distinct daily traffic flow patterns is useful in traffic control and prediction. For univariate functional data clustering, a simple method clusters the set of the univariate FPC (uFPC) scores by the classical K-means algorithm for multivariate data. A more advanced method applies K-centers

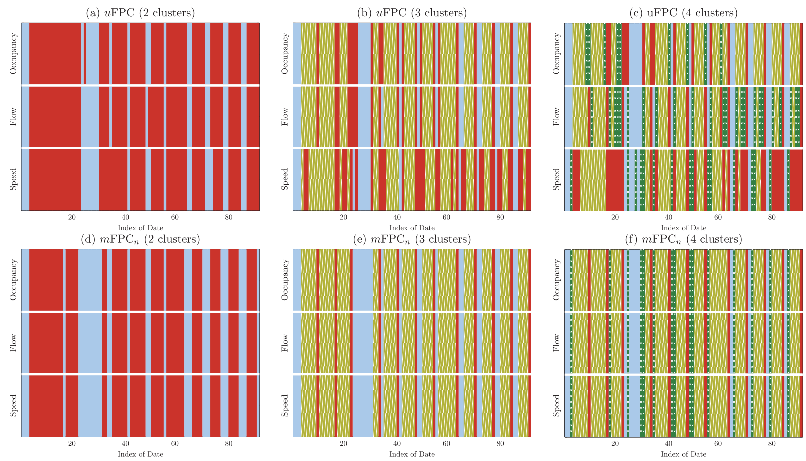

Figure 5. Cluster memberships of uFPC [(a)–(c)] versus $m\mathrm{FPC}_n$ [(d)–(f)] for 2, 3 and 4 clusters.

functional clustering (Chiou and Li (2007)) for considering distinct patterns in the mean and covariance functions. However, using the univariate approach to multivariate functional data can lead to inconsistent results of cluster memberships for different variables within the same unit or subject. This issue can be resolved using the *single* set of $m\mathrm{FPC}_n$ scores for the multivariate trajectories, in contrast to the *multiple* sets of uFPC scores with each corresponding to a random function. We compare the clustering results for the traffic flow data based on clustering the $m\mathrm{FPC}_n$ and uFPC scores as displayed in Figure 5 with 2–4 clusters. The cluster memberships of the days based on uFPC approach (upper panels) are not consistent for the same day across speed, flow, and occupancy.

4.2. Comparison in prediction

We define the cluster-specific average squared errors (cASE) between the observed $Y_{ki}(t_{ij})$ and the predicted $\hat{X}^{(c)}_{ki}(t_{ij})$ trajectories, for the kth variable, by

$$\mathrm{cASE}_k = N_c^{-1} \sum_{c=1}^{N_c} n_c^{-1} \sum_{i=1}^{n_c} m_i{}^{-1} \sum_{j=1}^{m_i} \left\{ \hat{X}^{(c)}_{ki}(t_{ij}) - Y_{ki}(t_{ij}) \right\}^2, \tag{4.1}$$

where N_c is the number of clusters. The numerical results based on cASE for comparing the univariate (uFPC), the classical multivariate ($m\mathrm{FPC}_u$) and the proposed multivariate $m\mathrm{FPC}_n$ approaches are given in Table 1.

The relative performance of $m\mathrm{FPC}_n$ to $m\mathrm{FPC}_u$ is summarized in the column R_1, the ratio of cASE of $m\mathrm{FPC}_n$ to $m\mathrm{FPC}_u$. Most of the values are significantly smaller than 1, indicating that $m\mathrm{FPC}_n$ performs better. For the single cluster

Table 1. Prediction performance for the traffic data based on cASE for $m\text{FPC}_n$, $m\text{FPC}_u$ and $u\text{FPC}$, with R_1, cASE ratio of $m\text{FPC}_n$ to $m\text{FPC}_u$, and R_2, cASE ratio of $m\text{FPC}_n$ to $u\text{FPC}$.

	WLS					CE				
	$m\text{FPC}_u$	$m\text{FPC}_n$	$u\text{FPC}$	R_1	R_2	$m\text{FPC}_u$	$m\text{FPC}_n$	$u\text{FPC}$	R_1	R_2
(1 cluster)										
Speed	0.686	0.507	0.467	0.739	1.086	0.630	0.514	0.467	0.816	1.101
Flow($\times 10^2$)	2.622	1.890	2.547	0.721	0.742	2.738	2.084	2.548	0.761	0.818
Occupancy	0.233	0.154	0.202	0.661	0.762	0.208	0.175	0.202	0.841	0.866
(2 clusters)										
Speed	0.642	0.404	0.401	0.629	1.008	0.645	0.423	0.402	0.656	1.052
Flow($\times 10^2$)	2.183	1.542	1.760	0.706	0.876	2.226	1.599	1.762	0.718	0.908
Occupancy	0.219	0.127	0.155	0.580	0.819	0.224	0.132	0.155	0.589	0.852
(3 clusters)										
Speed	0.670	0.386	0.362	0.576	1.066	0.603	0.397	0.365	0.658	1.088
Flow($\times 10^2$)	1.864	1.380	1.501	0.740	0.919	1.889	1.392	1.506	0.737	0.925
Occupancy	0.205	0.113	0.131	0.551	0.863	0.198	0.117	0.131	0.591	0.893
(4 clusters)										
Speed	0.450	0.361	0.355	0.802	1.017	0.435	0.369	0.355	0.848	1.039
Flow($\times 10^2$)	1.123	1.198	1.427	1.067	0.840	1.130	1.220	1.431	1.080	0.853
Occupancy	0.133	0.109	0.124	0.820	0.879	0.135	0.103	0.125	0.763	0.824

case the cASEs based on $m\text{FPC}_n$ coupled with WLS decrease by about 26%, 28%, and 34% for speed, flow and occupancy as compared to $m\text{FPC}_u$, and the reductions are 18%, 24%, and 16% when using the CE method. The cASE results based on 2–4 clusters are similar.

For the comparison of $m\text{FPC}_n$ with $u\text{FPC}$, the cASE ratios of $m\text{FPC}_n$ to $u\text{FPC}$ indicate that $m\text{FPC}_n$ greatly reduces cASEs for flow and occupancy, while cASEs for speed based on $m\text{FPC}_n$ are slightly larger. For the single cluster case the selected number of components is 5 (with FVE 91.4%) in $m\text{FPC}_n$, 2 (with FVE 93.2%) in $m\text{FPC}_u$ and 3, 2, and 2 (with FVE 90.9%, 93.3% and 94.5%) for the variables, speed, flow, and occupancy, in $u\text{FPC}$. In general, $m\text{FPC}_n$ coupled with WLS performs slightly better than CE for the traffic data analysis.

We conclude that clustering of multivariate functional data based on the $m\text{FPC}_n$ approach not only renders consistent cluster memberships across the variables but also leads to better performance in terms of prediction errors.

4.3. Simulation

We examined the finite sample performance of the $m\text{FPC}_n$, $m\text{FPC}_u$, and $u\text{FPC}$ methods. We took $p = 3$, and generated curves according to the truncated version of (3.1) up to $L = 60$ components. We considered equally spaced

Table 2. Relative performance in terms of cASE ratios of $m\mathrm{FPC}_n$ to $m\mathrm{FPC}_u$ (R_1) and cASE ratios of $m\mathrm{FPC}_n$ to $u\mathrm{FPC}$ (R_2) based on 200 simulation replicates for Settings I and II.

		Setting I				Setting II			
		WLS		CE		WLS		CE	
n	Variable	R_1	R_2	R_1	R_2	R_1	R_2	R_1	R_2
100	X_1	0.887	0.959	0.878	0.958	0.918	0.959	0.921	0.960
	X_2	0.909	0.987	0.909	0.986	0.808	1.022	0.810	1.021
	X_3	0.855	0.950	0.856	0.951	0.987	0.967	0.989	0.967
	Avg.	0.880	0.965	0.881	0.965	0.904	0.983	0.907	0.983
500	X_1	0.875	0.934	0.876	0.934	0.930	1.008	0.932	1.008
	X_2	0.902	1.006	0.902	1.006	0.813	1.048	0.814	1.047
	X_3	0.860	0.968	0.862	0.969	0.992	0.958	0.993	0.959
	Avg.	0.879	0.969	0.880	0.970	0.912	1.005	0.913	1.005

recording time points on [0, 5] with $m_i = 51$ for all i. We constructed $n = 100$ and $n = 500$ random trajectories with 200 simulated replicates. We set $\boldsymbol{\mu(t)} = \mathbf{0}$, and $\boldsymbol{v}(t) = (t + 0.5, \sin(t) + \cos(t) + 10, 0.5t^2 - t + 1)^\top$ for Setting I and $\boldsymbol{v}(t) = (t + 0.5, 0.04, (t-1)^2 + 1)^\top$ for Setting II. The multivariate FPC scores $\{\xi_{ri}\}$ were generated from $N(0, \lambda_r)$ and the measurement errors $\{\epsilon_{ki}\}$ from $N(0, \sigma_k^2)$ for $k = 1, \ldots, p$, where λ_r and σ_k^2 were taken from the estimates of the traffic flow analysis. We set the measurement error variance $\boldsymbol{\sigma}^2 = (1, 2.25, 0.64)^\top$ for Setting I and $\boldsymbol{\sigma}^2 = (2.25, 0.04, 5)^\top$ for Setting II. The settings of $\{\boldsymbol{\phi}_r\}$ and $\{\lambda_r\}$ were more complicated; more detail regarding these settings is provided in Supplement S1, and an additional simulation study that mimics our traffic flow analysis is compiled in Supplement S2.

Table 2 summarizes the simulation results via relative performance of cASE ($N_c = 1$) ratios of $m\mathrm{FPC}_n$ to $m\mathrm{FPC}_u$, R_1, and to $u\mathrm{FPC}$, R_2. Setting I was designed to have similar scales of the variance functions with higher correlations between the pairwise random functions, while Setting II was designed to have uneven magnitudes of the variance functions with lower between-variable correlations. For Setting I, the averaged R_1 ratios for the cases of WLS, CE, $n = 100$ and $n = 500$ are about 0.88, showing that using $m\mathrm{FPC}_n$ has substantial gains over $m\mathrm{FPC}_u$. The averaged R_2 ratios are about 0.97, indicating that that the prediction errors of $m\mathrm{FPC}_n$ can be smaller than $u\mathrm{FPC}$ by reducing 3% of cASE. For Setting II, the averaged R_1 ratios are around 0.91, and particularly the ratios for X_2 are as low as 0.81. In general, the performances of WLS and CE are quite close.

Figure 6 displays the boxplots for the number of components and fraction of total variance explained (FVE) based on 200 replicates of Setting II. Under the selection criterion of achieving 90% of total variance, $m\mathrm{FPC}_n$ selects three

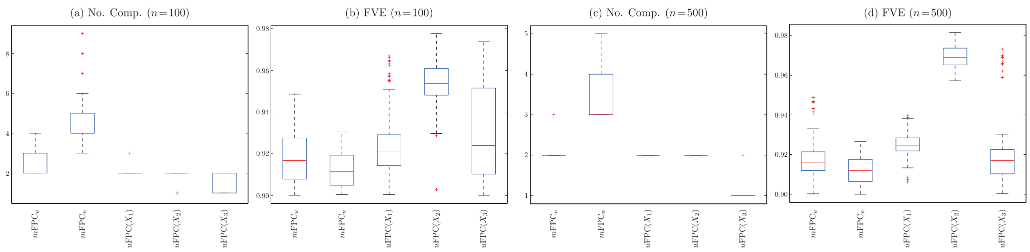

Figure 6. Boxplots for the number of components and fraction of total variance explained (FVE) based on 200 simulation replicates for the mFPC$_u$, mFPC$_n$ and uFPC methods in Setting II.

components mostly with FVE interquartile ranges from 90.5% to 91.9% for $n = 100$ and with FVE interquartile ranges from 90.6% to 91.8% for $n = 500$, while mFPC$_u$ selects two components with FVE interquartile ranges from 90.8% to 92.8% for $n = 100$ and from 91.2% to 92.1% for $n = 500$. For uFPC, the median number of components for X_1 and X_2 is 2, while the selected number is 1 for X_3, and FVEs are generally higher for X_2.

5. Asymptotic Properties

We investigate the asymptotic properties of our mFPC$_n$ approach, borrowing the theoretical results for the uFPC method in Li and Hsing (2010) and Yao, Müller, and Wang (2005), where the convergence rates in mean and covariance estimations are established. The asymptotic results consider the normalization effects on estimation for the representation in (2.3)and (2.1) under the mFPC framework.

For a vector $\boldsymbol{u} = (u_1, \ldots, u_p)^\top$ and a matrix $\mathbf{A} = \{\mathrm{A}_{kl}\}_{1 \le k,l \le p}$, we take $\|\boldsymbol{u}\|_2 = \left(\sum_{k=1}^p u_k^2\right)^{1/2}$ and $\|\boldsymbol{A}\|_2 = \left(\sum_{k,l=1}^p \mathrm{A}_{kl}^2\right)^{1/2}$. Further, let $\gamma_{nk} = \left(n^{-1} \sum_{i=1}^n m_i^{-k}\right)^{-1}$, where $m_i \ge 2$ and $k \in \mathbb{N}$. We need the following conditions.

(C1) The density function, f_T, of T_{ij} is bounded and differentiable with a bounded derivative.

(C2) The kernel $K(\cdot)$ is a symmetric probability density function with support [-1,1] and is differentiable with a bounded derivative.

(C3) The mean function $\mu_k(\cdot)$ is twice continuously differentiable with the second derivative bounded on $\mathcal{T}$ for all k.

(C4) The covariance function $\mathrm{G}_{kk}(t,t)$ is bounded above by M_{v_k} and bounded below by m_{v_k} for some $M_{v_k} \ge m_{v_k} > 0$. The second order derivative of $\mathrm{G}_{kl}(s,t)$ exists and is bounded on $\mathcal{T}^2$ for all k and l.

(C5) $\mathrm{E}\left(\sup_{t\in\mathcal{T}}|\mathrm{X}_k(t)|^{\lambda_{h_1}}\right) < \infty$ and $\mathrm{E}\left(|\epsilon_{kij}|^{\lambda_{h_1}}\right) < \infty$ for each $\lambda_{h_1} > 2$; $h_1 \to 0$ and $\left(h_1^2 + h_1/\gamma_{n1}\right)^{-1}(\log n/n)^{1-2/\lambda_{h_1}} \to 0$ as $n \to \infty$.

(C6) $\mathrm{E}\left(\sup_{t\in\mathcal{T}}|\mathrm{X}_k(t)|^{2\lambda_{h_2}}\right) < \infty$ and $\mathrm{E}\left(|\epsilon_{kij}|^{2\lambda_{h_2}}\right) < \infty$ for each $\lambda_{h_2} > 2$; $h_2 \to 0$ and $\left(h_2^4 + h_2^3/\gamma_{n1} + h_2^2/\gamma_{n2}\right)^{-1}(\log n/n)^{1-2/\lambda_{h_2}} \to 0$ as $n \to \infty$.

(C1)$-$(C4) hold as general assumptions for functional data analysis. The condition in (C4) that G_{kk} is bounded away from zero is required for the normalization approach and, with $\mathrm{C}_{kl}(s,t) = \{v_k(s)v_l(t)\}^{-1/2}\mathrm{G}_{kl}(s,t)$ and $v_k(t) = \mathrm{G}_{kk}(t,t)$, (C4) also holds for C_{kl}. The h_1 in (C5) corresponds to the bandwidth for a one-dimensional local linear smoother, such as b_{μ_k} in (3.2), h_{W_k} in (3.7) and h_{V_k} used to estimate V_k, while h_2 in (C6) corresponds to the bandwidth used in a two-dimensional local linear smoother, such as b_{G_k} in (3.3) and the h_k and h_l used in (3.5). The moment conditions (C5) and (C6) are required in establishing the results with the uniform rates of convergence. When the scores ξ_{ri} and the measurement errors $\boldsymbol{\varepsilon}_{ij}$ are jointly Gaussian, (C5) and (C6) hold. These conditions were used in Hall, Müller, and Wang (2006). Here (C5) and (C6) are required for *each* $\lambda_{h_1} > 2$ and for *each* $\lambda_{h_2} > 2$, unlike the conditions in Li and Hsing (2010).

For any bandwidth h_1 and h_2, let

$$\tau_{n1}(h_1) = h_1^2 + \left[\left\{1 + (h_1\gamma_{n1})^{-1}\right\}\left(\frac{\log n}{n}\right)\right]^{1/2},$$
$$\tau_{n2}(h_2) = h_2^2 + \left[\left\{1 + (h_2\gamma_{n1})^{-1} + \left(h_2^2\gamma_{n2}\right)^{-1}\right\}\left(\frac{\log n}{n}\right)\right]^{1/2}.$$

The bandwidths h_1 and h_2 satisfy (C5) and (C6), respectively. Take $\tau_\mu = \max_{1\le k\le p}\{\tau_{n1}(b_{\mu_k})\}$, $\tau_{\mathrm{G}} = \max_{1\le k\le p}\{\tau_{n2}(b_{\mathrm{G}_k})\}$, and $\tau_{kl} = \max_{k\neq l}\{\tau_{n2}(h_{kl})\}$, where $h_{kl} = \max\{h_k, h_l\}$, $\tau_{\mathrm{W}} = \max_{1\le k\le p}\{\tau_{n1}(h_{\mathrm{W}_k})\}$.

Remarks. (C5) implies $\tau_{n1}(h_1) \to 0$ as $n \to \infty$. To see this, observe that $m_i \ge 2$ for $i = 1, \ldots, n$, and thus $\gamma_{n1} \ge 1$,

$$\begin{aligned}(h_1\gamma_{n1})^{-1}\Big(\frac{\log n}{n}\Big) \le h_1^{-1}\Big(\frac{\log n}{n}\Big) &\le \left\{\frac{(h_1^2 + h_1)}{2}\right\}^{-1}\Big(\frac{\log n}{n}\Big)\\ &\le 2\left(h_1^2 + \frac{h_1}{\gamma_{n1}}\right)^{-1}\left(\frac{\log n}{n}\right)^{1-2/\lambda_{h_1}}\end{aligned}$$

converges to zero by (C5), where $h_1 > (h_1^2 + h_1)/2$ for $n \gg 0$. When $\max_{1\le i\le n}\{m_i\} < M$ for some fixed $M < \infty$, it follows that $\gamma_{n1} \le M$ and $1 + (h_1\gamma_{n1})^{-1} = O(1/h_1)$, and thus $\tau_{n1}(h_1) = O\left(h_1^2 + \{\log n/(nh_1)\}^{1/2}\right)$. On the other hand, if

$\min_{1\le i\le n}\{m_i\}\ge M_n$ such that $M_n = O(1/h_1)$, then $\tau_{n1}(h_1) = O\left(h_1^2 + \{\log n/n\}^{1/2}\right)$. Similarly, it can be shown that (C6) implies $\tau_{n2}(h) \to 0$ as $n \to \infty$.

Lemma 1. *Let* $\hat{\boldsymbol{\mu}}(t) = (\hat\mu_1(t), \ldots, \hat\mu_p(t))^\top$, *where* $\hat\mu_k(t)$ *is obtained by* (3.2), *and* $\hat{\boldsymbol{v}}(t) = (\hat v_1(t), \ldots, \hat v_p(t))^\top$, *where* $\hat v_k(t)$ *is obtained by* (3.3), *for* $k = 1, \ldots, p$. *Under* (C1)−(C6), *and if* $b_{\mu_k} \asymp h_1$ *and* $b_{G_k} \asymp h_2$, *then*

(a) $\|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|_{\mathbb{H}} = O(\tau_\mu)$ *a.s.*.

(b) $\|\hat{\boldsymbol{v}} - \boldsymbol{v}\|_{\mathbb{H}} = O(\tau_\mu + \tau_{\mathrm{G}})$ *a.s.*.

Let $\hat{\boldsymbol{\sigma}}^2 = (\hat\sigma_1^2, \ldots, \hat\sigma_p^2)^\top$, *where* $\hat\sigma_k^2$ *is obtained by* (3.8). *If* $h_{\mathrm{W}_k} \asymp h_1$, *then*

(c) $\|\hat{\boldsymbol{\sigma}}^2 - \boldsymbol{\sigma}^2\|_2 = O(\tau_{\mathrm{W}} + \tau_\mu + \tau_{\mathrm{G}})$ *a.s.*.

To prove (a) and (b) of Lemma 1. we follow the proofs of Theorems 3.1 and 3.3 of Li and Hsing (2010) although the estimates in (b) are slightly different. Details are in Section 6. Lemma 1 still holds if the the moment conditions in (C5) and (C6) are relaxed to hold for *some* $\lambda_{h_q} \in (2, \infty)$, rather than for *each* $\lambda_{h_q} > 0$ where $q = 1, 2$.

Lemma 2. *Under* (C1)−(C6), *if* $b_{\mu_k} \asymp h_1$ *and* $b_{G_k} \asymp h_2$, *then for all* $k = 1, \ldots, p$, *and* $i = 1, \ldots, n$,

$$\sup_{1\le j\le m_i} |\widetilde{\mathrm{U}}_{kij} - \mathrm{U}_{kij}| = O\left(\tau_{n2}(b_{\mathrm{G}_k}) + \tau_{n1}(b_{\mu_k})\right) \ a.s..$$

The same rate holds for convergence in probability if the moment conditions in (C5) and (C6) are satisfied for *some* $\lambda_{h_q} \in (2, \infty)$, where $q = 1, 2$.

Theorem 1. *Let* $\hat{\mathbf{C}}(s,t) = \left\{\hat{\mathrm{C}}_{kl}(s,t); 1 \le k, l \le p\right\}$, *where* $\hat{\mathrm{C}}_{kl}(s,t)$ *is obtained by* (3.5). *Under* (C1)−(C6), *if* $b_{\mu_k} \asymp h_1$ *and* $b_{\mathrm{G}_k} \asymp h_{kl} \asymp h_2$, *then it holds that*

(a) $\sup_{s,t\in\mathcal{T}} \|\hat{\mathbf{C}}(s,t) - \mathbf{C}(s,t)\|_2 = O(\tau_{kl} + \tau_{\mathrm{G}} + \tau_\mu)$ *a.s.*.

If in addition $h_{W_k} \asymp h_1$, *then for* $i = 1, \ldots, n$,

(b) $\sup\limits_{1\le j\le m_i} |\hat\varsigma_{ij}^2 - \varsigma_{ij}^2| = O(\tau_W + \tau_\mu + \tau_{\mathrm{G}})$ *a.s.*.

Theorem 2. *Let* $\hat\lambda_r$ *and* $\hat{\boldsymbol{\phi}}_r(t) = \left(\hat\phi_{1r}(t), \ldots, \hat\phi_{pr}(t)\right)^\top$ *be the estimates of* λ_r *and* $\boldsymbol{\phi}_r(t)$ *corresponding to the solutions of the eigenequations* (3.6). *Under* (C1)−(C6) *with* $b_{\mu_k} \asymp h_1$ *and* $b_{\mathrm{G}_k} \asymp h_{kl} \asymp h_2$, *if* λ_r *is simple, then*

(a) $|\hat\lambda_r - \lambda_r| = O(\tau_{kl} + \tau_{\mathrm{G}} + \tau_\mu)$ *a.s.*.
(b) $\|\hat{\boldsymbol{\phi}}_r - \boldsymbol{\phi}_r\|_{\mathbb{H}} = O(\tau_{kl} + \tau_{\mathrm{G}} + \tau_\mu)$ *a.s.*.

(c) $\sup_{t\in\mathcal{T}} \|\hat{\boldsymbol{\phi}}_r(t) - \boldsymbol{\phi}_r(t)\|_2 = O\left(\tau_{kl} + \tau_{\mathrm{G}} + \tau_\mu\right)$ *a.s.*.

Theorem 1 establishes the uniform consistency result for the multivariate covariance and cross-covariance functions. The rates of convergence in Theorem 2 are those of Theorem 1 (a). This is consistent with the finding in Yao, Müller, and Wang (2005). The truncated number of components $L = L(n, \delta_0)$ depends on the threshold δ_0 and the sample size n, and for a fixed $0 < \delta_0 < 1$, a corresponding L_0 can be determined as $L_0 = L_0(\delta_0) = \mathrm{argmin}_M\{\sum_{r=1}^M \lambda_r / \sum_{r=1}^\infty \lambda_r > \delta_0\}$. Moreover, by the consistency property in Theorem 2 along with Slusky's theorem, it can be shown that for each $0 < \delta_0 < 1$, $\lim_{n\to\infty} L(n, \delta_0) = L_0(\delta_0)$ (cf., Sec. 3.4 in Dauxois, Pousse, and Romain (1982)). Further, $\lim_{\delta_0\to 1}\lim_{n\to\infty} L(n, \delta_0) = \infty$.

Theorem 3. *Given a fixed* $0 < \delta_0 < 1$*, set* $L = L_0(\delta_0)$*. Assume* (C1)−(C6) *hold with* $b_{\mu_k} \asymp h_{\mathrm{W}_k} \asymp h_1$ *and* $b_{\mathrm{G}_k} \asymp h_{kl} \asymp h_2$*, and assume* $m_i \to \infty$ *as* $n \to \infty$*. Then*

(i) *there exist* $0 < \delta_1 < \infty$ *and* $0 < \eta < \infty$ *such that* $E\left(|\varepsilon_{kij}^2|^{1+\delta_1}\right) < \eta$ *for all* $1 \le k \le p$ *and* $1 \le j \le m_i$*;*

(ii) *there exist* $0 < \delta_2 < \infty$ *such that* $(pm_i)^{-1}\boldsymbol{\Psi}_i$ *is nonsingular for* m_i *sufficiently large with det*$\left\{(pm_i)^{-1}\boldsymbol{\Psi}_i\right\} > \delta_2 > 0$.

Under (i) *and* (ii), $\hat{\boldsymbol{\xi}}_{i,L}^{WLS} \xrightarrow{a.s.} \boldsymbol{\xi}_{i,L}$ *as* $n \to \infty$*, and*

$$\hat{\boldsymbol{\Psi}}_i^{-1/2}\left(\hat{\boldsymbol{\xi}}_{i,L}^{WLS} - \boldsymbol{\xi}_{i,L}\right) \xrightarrow{d} N(0, \boldsymbol{I}_L), \quad \textit{as } n \to \infty.$$

The theorem indicates that the asymptotic normality property of $(\hat{\boldsymbol{\xi}}_{i,L}^{WLS} - \boldsymbol{\xi}_{i,L})$ holds for densely collected functional data. With $\boldsymbol{\omega}_{i,L}^{\vartheta}(s,t) = \boldsymbol{\phi}_{L,s}^\top \boldsymbol{\Omega}_{i,L}^{\vartheta}\boldsymbol{\phi}_{L,t}$ for $s, t \in \mathcal{T}$, where ϑ denotes WLS or CE, $\left\{\boldsymbol{\omega}_{i,L}^{\vartheta}(s,t)\right\}$ is a sequence of continuous positive definite matrices of the functions in s and t.

(C7) There exists a continuous positive definite matrix of function $\boldsymbol{\omega}_i^{\vartheta}(s,t)$ such that $\boldsymbol{\omega}_{i,L}^{\vartheta}(s,t) \to \boldsymbol{\omega}_i^{\vartheta}(s,t)$ element wise as $L \to \infty$ for all $s, t \in \mathcal{T}$.

Further, let $\hat{\boldsymbol{\omega}}_{i,L}^{\vartheta}(s,t) = \hat{\boldsymbol{\phi}}_{L,s}^\top \hat{\boldsymbol{\Omega}}_{i,L}^{\vartheta}\hat{\boldsymbol{\phi}}_{L,t}$. By the consistency of the estimates λ_r, $\boldsymbol{\phi}_r(s)$ and $\mu_k(t)$, it holds under (C7) that $\lim_{L\to\infty}\lim_{n\to\infty}\hat{\boldsymbol{\omega}}_{i,L}^{\vartheta}(s,t) = \boldsymbol{\omega}_i^{\vartheta}(s,t)$ *a.s.*.

Corollary 1. *Assume the conditions of Theorem* 3 *hold.*

(a) *If* (C7) *holds, then for any* $t \in \mathcal{T}$,

$$\lim_{L\to\infty}\lim_{n\to\infty}\left(\hat{\boldsymbol{\omega}}_{i,L}^{WLS}(t,t)\right)^{-1/2}\left(\hat{\mathbf{Z}}_i^{L,WLS}(t) - \mathbf{Z}_i(t)\right) = \mathbf{D} \sim N(0, \mathbf{I}_p).$$

(b) *For a fixed L,*

$$\lim_{n\to\infty} P\left\{\sup_{t\in\mathcal{T}} \frac{\left|\boldsymbol{a}^\top\left\{\hat{\mathbf{Z}}_i^{L,WLS}(t)-\mathbf{Z}_i^L(t)\right\}\right|}{\left\{\boldsymbol{a}^\top\hat{\boldsymbol{\omega}}_{i,L}^{WLS}(t,t)\boldsymbol{a}\right\}^{1/2}} \le \sqrt{\mathcal{X}^2_{L,1-\alpha}}\right\} \ge 1-\alpha,$$

where $\mathbf{Z}_i^L(t) = \sum_{r=1}^L \xi_{ri}\boldsymbol{\phi}_r(t)$, $\boldsymbol{a} = (a_1,\ldots,a_p)^\top \in \mathbb{R}^p$ *is a p-vector and* $\mathcal{X}^2_{L,1-\alpha}$ *is the* $(1-\alpha)$*th percentile of the chi-square distribution with L degrees of freedom.*

Corollary 1 (a) gives the asymptotic behavior of $\{\hat{\mathbf{Z}}_i^{L,WLS}(t) - \mathbf{Z}_i(t)\}$ when δ_0 tends to one, where only the pointwise distribution can be attained. For the case of $0 < \delta_0 < 1$, the distribution uniformly in $\mathcal{T}$ can be further attained as in (b). Parallel to the WLS estimates $\boldsymbol{\xi}_{i,L}^{WLS}$ (3.10), the asymptotic results similar to Theorem 3 and Corollary 1 can be obtained for the conditional expectation estimate $\boldsymbol{\xi}_{i,L}^{CE}$ (3.11) under the additional assumption that the scores ξ_{ri} and the measurement errors $\boldsymbol{\varepsilon}_{ij}$ are jointly Gaussian.

6. Proofs

We provide proofs of the asymptotic properties discussed in Section 5. The proofs of Lemmas 1 and 2, and Corollary 1 are compiled in the Supplementary Material.

To prove Theorem 1, we consider the following conditions.

(C5.1) $\mathrm{E}(\sup_{t\in\mathcal{T}}|\mathrm{Z}_k(t)|^{\lambda_{h_1}}) < \infty$ and $\mathrm{E}(|\varepsilon_{kij}|^{\lambda_{h_1}}) < \infty$ for each $\lambda_{h_1} > 2$; $h_1 \to 0$, and $\left(h_1^2 + h_1/\gamma_{n1}\right)^{-1}(\log n/n)^{1-2/\lambda_{h_1}} \to 0$ as $n\to\infty$.

(C6.1) $\mathrm{E}(\sup_{t\in\mathcal{T}}|\mathrm{Z}_k(t)|^{2\lambda_{h_2}}) < \infty$ and $\mathrm{E}(|\varepsilon_{kij}|^{2\lambda_{h_2}}) < \infty$ for each $\lambda_{h_2} > 2$; $h_2 \to 0$, and $\left(h_2^4 + h_2^3/\gamma_{n1} + h_2^2/\gamma_{n2}\right)^{-1}(\log n/n)^{1-2/\lambda_{h_2}} \to 0$ as $n\to\infty$.

Lemma 3.
(a) (C5) *implies* (C5.1).
(b) (C6) *implies* (C6.1).

The proof of Lemma 3 is in Supplement S3.

6.1. Proof of Theorem 1

(a) For the case of $k = l$, it follows by Lemma 1 (b) that

$$\sup_{s,t\in\mathcal{T}}\left|\hat{\mathrm{C}}_{kk}(s,t) - \mathrm{C}_{kk}(s,t)\right| \le \frac{1}{m_{v_k}(m_{v_k}-\delta_0)}\left\{M_{v_k}\sup_{s,t\in\mathcal{T}}\left|\hat{\mathrm{G}}_{kk}(s,t) - \mathrm{G}_{kk}(s,t)\right|\right.$$

$$+(M_{v_k}+\delta_0)\sup_{s,t\in\mathcal{T}}\left|\hat{\mathrm{G}}_{kk}(s,t)-\mathrm{G}_{kk}(s,t)\right|\Bigg\}\ a.s..$$
$$=O(\tau_{n2}(b_{\mathrm{G}_k})+\tau_{n1}(b_{\mu_k}))\ a.s..$$

For the case of $k\neq l$, since (C6.1) is satisfied, Lemma 3 (b), the proof is similar to that of Lemma 1 (b), with the following notation. Let

$$Q_{pq}=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{M_i}\sum_{j\neq j'}\widetilde{U}_{kij}\widetilde{U}_{lij'}\left(\frac{T_{ij}-s}{h_k}\right)^p\left(\frac{T_{ij'}-t}{k_l}\right)^q K\left(\frac{T_{ij}-s}{h_k}\right)K\left(\frac{T_{ij'}-t}{h_l}\right),$$

$$Q^*_{pq}=Q_{pq}-C_{kl}(s,t)L_{pq}-h_k\frac{\partial}{\partial s}C_{kl}(s,t)L_{p+1,q}-h_l\frac{\partial}{\partial t}C_{kl}(s,t)L_{p,q+1},$$

where

$$L_{pq}=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{M_i}\sum_{j\neq j'}\left(\frac{T_{ij}-s}{b_{G_k}}\right)^p\left(\frac{T_{ij'}-t}{b_{G_k}}\right)^q K\left(\frac{T_{ij}-s}{h_l}\right)K\left(\frac{T_{ij'}-t}{h_k}\right).$$

Then, it can be shown that

$$\left(\hat{C}_{kl}-C_{kl}\right)(s,t)=(\mathcal{A}_1Q^*_{00}-\mathcal{A}_2Q^*_{10}-\mathcal{A}_3Q^*_{01})\mathcal{B}_0^{-1},$$

where $\mathcal{A}_r$, $r=1,2,3$, and $\mathcal{B}_0$ are defined as in the proof of Lemma 1 (b) and have the same asymptotic orders. Hence, it remains to investigate Q^*_{00} since the other terms are of lower order. Let $\eta^*_{klijj'}=\widetilde{U}_{kij}\widetilde{U}_{lij'}-C_{kl}(T_{ij},T_{ij'})$. By Taylor's expansion,

$$Q^*_{00}=\frac{1}{n}\sum_{i=1}^{n}\frac{1}{M_i}\sum_{i\neq j}\eta^*_{klijj'}K\left(\frac{T_{ij}-s}{h_k}\right)K\left(\frac{T_{ij'}-t}{h_l}\right)+O(h_{kl}^2).$$

Here $\mathrm{E}(U_{kij}U_{lij'}|T_{ij},T_{ij'})=C_{kl}(T_{ij},T_{ij'})$ and, by Lemma 2, $|\widetilde{\mathrm{U}}_{kij}-\mathrm{U}_{kij}|=O\left(\tau_{n2}(b_{\mathrm{G}_k})+\tau_{n1}(b_{\mu_k})\right)$ *a.s.* for each $1\leq j\leq m_i$. Thus, we have $\mathrm{E}(Q^*_{00})=O\left(\tau_{n2}(b_{\mathrm{G}_k})+\tau_{n1}(b_{\mu_k})\right)$ and $Q^*_{00}=O(\tau_{n2}(h_{kl})+\tau_{n2}(b_{\mathrm{G}_k})+\tau_{n1}(b_{\mu_k})+\tau_{n2}(b_{\mathrm{G}_l})$ $+\tau_{n1}(b_{\mu_l}))$ *a.s.*. Therefore, $\left|\hat{\mathrm{C}}_{kl}(s,t)-\mathrm{C}_{kl}(s,t)\right|=O(\tau_{n2}(h_{kl})+\tau_{n2}(b_{\mathrm{G}_k})+\tau_{n1}(b_{\mu_k})+\tau_{n2}(b_{\mathrm{G}_l})+\tau_{n1}(b_{\mu_l}))$ *a.s.* uniformly in $\mathcal{T}^2$. Then $\sup_{s,t\in\mathcal{T}}\|\hat{\mathbf{C}}(s,t)-\mathbf{C}(s,t)\|_2=O\left(\tau_{kl}+\tau_{\mathrm{G}}+\tau_{\mu}\right)$ *a.s.*.

(b) By the consistency properties of $v_k(t)$ and σ_k obtained in Lemma 1 (b) and (c), together with the Slutsky's theorem, the result follows.

6.2. Proof of Theorem 2

We take $\|\cdot\|_{\mathbb{B}}$ as an induced norm for any linear operator in $\mathbb{H}$, such that, for $\mathcal{B}\in\mathbb{B}=\boldsymbol{B}(\mathbb{H})$, a set of all linear operators whose domain and range are both in $\mathbb{H}$, $\|\mathcal{B}\|_{\mathbb{B}}=\sup_{\|\boldsymbol{u}\|_{\mathbb{H}}\leq 1}\|\mathcal{B}\boldsymbol{u}\|_{\mathbb{H}}=\|\mathcal{B}\boldsymbol{u}^*\|_{\mathbb{H}}$ for some $\boldsymbol{u}^*$ with $\|\boldsymbol{u}^*\|_{\mathbb{H}}=1$.

We define $\hat{\mathcal{A}}$ similarly to $\mathcal{A}$ (2.5) by replacing $\mathbf{C}(s,t)$ with $\hat{\mathbf{C}}(s,t)$. Recall from (2.6) that the eigen-pair $(\boldsymbol{\phi}_r, \lambda_r)$ of $\mathcal{A}$ satisfies the eigen-equation $\mathcal{A}\boldsymbol{\phi}_r = \lambda_r\boldsymbol{\phi}_r$ and the set $\{\boldsymbol{\phi}_r\}$ forms an orthonormal basis in $\mathbb{H}$. The estimated eigen-pair $(\hat{\boldsymbol{\phi}}_r, \hat{\lambda}_r)$ of $\hat{\mathcal{A}}$ is analogously defined in (3.6). Then

$$\begin{aligned}\|\hat{\mathcal{A}} - \mathcal{A}\|_{\mathbb{B}}^2 &\leq \sum_{k=1}^{p} \int \|\hat{\mathbf{C}}_k(s,\cdot) - \mathbf{C}_k(s,\cdot)\|_{\mathbb{H}}^2\, ds \\ &\leq \sum_{1\geq k,l\leq p} |\mathcal{T}|^2 \sup_{s,t\in\mathcal{T}} |\hat{\mathrm{C}}_{kl}(s,t) - \mathrm{C}_{kl}(s,t)|^2 \\ &= O\left\{(\tau_{kl} + \tau_{\mathrm{G}} + \tau_{\mu})^2\right\}\ a.s.. \qquad (6.1)\end{aligned}$$

For any $\boldsymbol{u}$ in $\mathbb{H}$, we define a projector $\mathcal{P}_r$ (resp. $\hat{\mathcal{P}}_r$) corresponding to λ_r (resp. $\hat{\lambda}_r$), for a fixed r, as

$$\mathcal{P}_r(\boldsymbol{u}) = u_{E_r} \in E_r,$$

where E_r is the eigen-space of λ_r, $E_r = \{\mathrm{span}\,(\boldsymbol{\phi}_{r'}) : \lambda_{r'} = \lambda_r\}$ and $\boldsymbol{u} = u_{E_r} + u_{E_r}^{\perp}$, where $u_{E_r}^{\perp}$ is the complement of u_{E_r}. Thus, $\mathcal{P}_r(\boldsymbol{u}) = \sum_{\{r':\lambda_{r'}=\lambda_r\}} \langle\boldsymbol{\phi}_{r'}, \boldsymbol{u}\rangle\boldsymbol{\phi}_{r'}$. The projector $\hat{\mathcal{P}}_r$ is defined analogously.

In particular, $\mathcal{P}_r(\boldsymbol{u}) = \langle\boldsymbol{\phi}_r, \boldsymbol{u}\rangle\boldsymbol{\phi}_r$ if λ_r is simple. Hence, given that λ_r is simple,

$$\begin{aligned}\|\hat{\mathcal{P}}_r - \mathcal{P}_r\|_{\mathbb{B}}^2 = \sup_{\|\boldsymbol{u}\|\leq 1} \|(\hat{\mathcal{P}}_r - \mathcal{P}_r)(\boldsymbol{u})\|_{\mathbb{H}}^2 &\geq \|(\hat{\mathcal{P}}_r - \mathcal{P}_r)(\hat{\boldsymbol{\phi}}_r)\|_{\mathbb{H}}^2 \\ &= \langle\hat{\boldsymbol{\phi}}_r - \langle\boldsymbol{\phi}_r, \hat{\boldsymbol{\phi}}_r\rangle_{\mathbb{H}}\ \boldsymbol{\phi}_r, \hat{\boldsymbol{\phi}}_r - \langle\boldsymbol{\phi}_r, \hat{\boldsymbol{\phi}}_r\rangle_{\mathbb{H}}\ \boldsymbol{\phi}_r\rangle_{\mathbb{H}} \\ &= 1 - \langle\boldsymbol{\phi}_r, \hat{\boldsymbol{\phi}}_r\rangle_{\mathbb{H}}^2 \\ &\geq \frac{1}{2}\|\hat{\boldsymbol{\phi}}_r - \boldsymbol{\phi}_r\|_{\mathbb{H}}^2. \qquad (6.2)\end{aligned}$$

That is, $\|\hat{\boldsymbol{\phi}}_r - \boldsymbol{\phi}_r\|_{\mathbb{H}} \leq \sqrt{2}\|\hat{\mathcal{P}}_r - \mathcal{P}_r\|_{\mathbb{B}}$.

To investigate the order of $\|\hat{\mathcal{P}}_r - \mathcal{P}_r\|_{\mathbb{B}}$, we introduce the resolvent of $\mathcal{A}$ (resp. $\hat{\mathcal{A}}$), $\mathcal{R} : \mathbb{C}\backslash\{\lambda_r\text{'s}\} \to \mathbb{B}$, where $\mathbb{C}$ denotes the set of all complex numbers, defined as $\mathcal{R}(z) = (\mathcal{A} - z\boldsymbol{I})^{-1}$ (resp. $\hat{\mathcal{R}}(z) = (\hat{\mathcal{A}} - z\boldsymbol{I})^{-1}$) .

By decomposition of the spectrum (c.f., Kato (1984, III-6)), $\mathcal{P}_r = -(1/2\pi i) \int_{\Lambda_{\rho,r}} \mathcal{R}(z)dz$, where $\Lambda_{\rho,r} = \{z \in \mathbb{C} : |z - \lambda_r| = \rho\}$ and ρ is taken so that $2\rho \leq \min_{r\neq r'} |\lambda_r - \lambda_{r'}|$. A similar argument applies to $\hat{P}_r$. By direct calculation, one has

$$\mathcal{R}(z) - \hat{\mathcal{R}}(z) = \hat{\mathcal{R}}(z)(\mathcal{A} - \hat{\mathcal{A}})\mathcal{R}(z). \qquad (6.3)$$

Since λ_r is simple and $\hat{\mathcal{A}} \to \mathcal{A}$ in the $\mathbb{B}$-norm sense as $n \to \infty$, by Lemma 2.1 of Gil' (1999), there exists a ρ^* such that among all eigeinvalues of $\mathcal{A}$ and $\hat{\mathcal{A}}$, only λ_r and $\hat{\lambda}_r$ lie in $\Lambda_{\rho^*,r}$ when n is large enough. Note that $\mathcal{R}(z)$ is a continuous linear operator and, thus, is bounded. Let $M_1 = \sup_{z\in\Lambda_{\rho^*,r}} \{\|\mathcal{R}(z)\|_{\mathbb{B}}\}$. By (6.1) and the definition of $\hat{\mathcal{R}}$, we have $\hat{\mathcal{R}} \overset{a.s.}{\to} \mathcal{R}$ in the $\|\cdot\|_{\mathbb{B}}$ sense. Thus, $\sup_{z\in\Lambda_{\rho^*,r}} \left\{\|\hat{\mathcal{R}}(z)\|_{\mathbb{B}}\right\} \leq M_1 + \delta_1$ for some fixed $\delta_1 > 0$ when $n \gg 0$. By (6.3),

$$\|\hat{\mathcal{P}}_r - \mathcal{P}_r\|_{\mathbb{B}} \leq \frac{1}{2\pi}\int_{\Lambda_{\rho*,r}} \|\hat{\mathcal{R}}(z)\|_{\mathbb{B}}\|\mathcal{A}-\hat{\mathcal{A}}\|_{\mathbb{B}}\|\mathcal{R}(z)\|_{\mathbb{B}}dz \leq \rho^* M_1\left(M_1+\delta_1\right)\|\mathcal{A}-\hat{\mathcal{A}}\|_{\mathbb{B}}. \tag{6.4}$$

By the inequality (6.2) together with the result of (6.1), we have $\|\hat{\boldsymbol{\phi}}_r - \boldsymbol{\phi}_r\|_{\mathbb{H}} = O\left(\tau_{kl} + \tau_{\mathrm{G}} + \tau_\mu\right)$ *a.s.*.

For the convergence rate of $(\hat{\lambda}_r - \lambda_r)$ in (a), we note that $\langle\boldsymbol{\phi}_r, \mathcal{A}(\boldsymbol{\phi})\rangle_{\mathbb{H}} = \lambda_r$, and $\langle\hat{\boldsymbol{\phi}}_r, \hat{\mathcal{A}}(\hat{\boldsymbol{\phi}})\rangle_{\mathbb{H}} = \hat{\lambda}_r$,

$$\begin{aligned}|\lambda_r - \hat{\lambda}_r| &\leq \left|\langle\boldsymbol{\phi}_r, (\mathcal{A}-\hat{\mathcal{A}})(\boldsymbol{\phi}_r)\rangle_{\mathbb{H}}\right| + \left|\langle\boldsymbol{\phi}_r - \hat{\boldsymbol{\phi}}_r, \hat{\mathcal{A}}(\boldsymbol{\phi}_r)\rangle_{\mathbb{H}}\right| + \left|\langle\hat{\boldsymbol{\phi}}_r, \hat{\mathcal{A}}(\boldsymbol{\phi}_r - \hat{\boldsymbol{\phi}}_r)\rangle_{\mathbb{H}}\right| \\ &\leq \|\mathcal{A}-\hat{\mathcal{A}}\|_{\mathbb{B}} + 2\|\hat{\mathcal{A}}\|_{\mathbb{B}}\,\|\boldsymbol{\phi}_r - \hat{\boldsymbol{\phi}}_r\|_{\mathbb{H}} = O\left(\tau_{kl}+\tau_{\mathrm{G}}+\tau_\mu\right)\ a.s..\end{aligned}$$

To obtain (c), use

$$\begin{aligned}\|\hat{\lambda}_r\hat{\boldsymbol{\phi}}_r(s) - \lambda_r\boldsymbol{\phi}_r(s)\|_2 \leq &\int_{\mathcal{T}} \|\hat{\mathbf{C}}(s,t) - \mathbf{C}(s,t)\|_2\ \|\hat{\boldsymbol{\phi}}_r(t)\|_2 dt \\ &+ \int_{\mathcal{T}} \|\mathbf{C}(s,t)\|_2\ \|\hat{\boldsymbol{\phi}}_r(t) - \boldsymbol{\phi}_r(t)\|_2 dt.\end{aligned}$$

By Hölder's inequality and Theorem 1,

$$\begin{aligned}\|\hat{\lambda}_r\hat{\boldsymbol{\phi}}_r(s) - \lambda_r\boldsymbol{\phi}_r(s)\|_2 &\leq \left\{\int_{\mathcal{T}} \|\hat{\mathbf{C}}(s,t) - \mathbf{C}(s,t)\|_2^2 dt\right\}^{1/2} \left\{\int_{\mathcal{T}} \|\hat{\boldsymbol{\phi}}_r(t)\|_2^2 dt\right\}^{1/2} \\ &\quad + \left\{\int_{\mathcal{T}} \|\mathbf{C}(s,t)\|_2^2 dt\right\}^{1/2} \left\{\int_{\mathcal{T}} \|\hat{\boldsymbol{\phi}}_r(t) - \boldsymbol{\phi}_r(t)\|_2^2 dt\right\}^{1/2} \\ &\leq |\mathcal{T}|^{1/2} \sup_{s,t\in\mathcal{T}} \|\hat{\mathbf{C}}(s,t) - \mathbf{C}(s,t)\|_2 + M\|\hat{\boldsymbol{\phi}}_r - \boldsymbol{\phi}_r\|_{\mathbb{H}} \\ &= O\left(\tau_{kl}+\tau_{\mathrm{G}}+\tau_\mu\right)\ a.s.,\end{aligned}$$

where M is a fixed constant as the bound of $\left\{\int_{\mathcal{T}} \|\mathbf{C}(s,t)\|_2^2 dt\right\}^{1/2}$ whose existence is assured by (C4). Assuming at this point that one of λ_r or $\hat{\lambda}_r$ is not 0, say, $\lambda_r > 0$, then $\|(\hat{\lambda}_r/\lambda_r)\hat{\boldsymbol{\phi}}_r(s) - \boldsymbol{\phi}_r(s)\|_2 = O\left(\tau_{kl}+\tau_{\mathrm{G}}+\tau_\mu\right)$ *a.s.* uniformly in $s \in \mathcal{T}$. Since $|\hat{\lambda}_r - \lambda_r| = O\left(\tau_{kl}+\tau_{\mathrm{G}}+\tau_\mu\right)$ *a.s.*, it follows that $\sup_{t\in\mathcal{T}} \|\hat{\boldsymbol{\phi}}_r(t) - \boldsymbol{\phi}_r(t)\|_2 = O\left(\tau_{kl}+\tau_{\mathrm{G}}+\tau_\mu\right)$ *a.s.*. For the remaining case $\lambda_r = \hat{\lambda}_r = 0$, we have $\boldsymbol{\phi}_r(s) = \hat{\boldsymbol{\phi}}_r(s) = 0$ since we assume λ_r is simple and (c) is satisfied.

Remarks. The technique of applying a resolvent was used in the proofs of Proposition 3 by Dauxois, Pousse, and Romain (1982) and Theorem 2 by

Yao, Müller, and Wang (2005) for univariate FPCA. Major differences are summarized as follows.

We adopt the induced norm ($\|\cdot\|_{\mathbb{B}}$) in the proof, rather than the Hilbert-Schmidt norm ($\|\cdot\|_{HS}$) used in Yao, Müller, and Wang (2005). This was due to the concern that the norm of the resolvent $\mathcal{R}(z) = (\mathcal{A} - z\boldsymbol{I})^{-1}$ needs to be finite. Boundedness of $\mathcal{R}(z)$ based on the induced norm is assured by definition and is used in (6.4) to obtain the consistency of $\hat{\boldsymbol{\phi}}_r$. In contrast, since the Hilbert space $L_2(\mathcal{T}) \equiv \mathbb{H}$ is of infinite dimension, the Hilbert-Schmidt norm for the identity is not finite. In addition, we introduce Lemma 2.1 of Gil' (1999) to restrict the integration of $\mathcal{R}(z)$ and $\hat{\mathcal{R}}(z)$ to the same domain, which is crucial when estimating $\|\hat{\mathcal{P}}_r - \mathcal{P}_r\|_{\mathbb{B}}$.

6.3. Proof of Theorem 3

We note that $\|\hat{\boldsymbol{\xi}}_{i,L}^{WLS} - \boldsymbol{\xi}_{i,L}\|_2 \leq \|\hat{\boldsymbol{\xi}}_{i,L}^{WLS} - \boldsymbol{\xi}_{i,L}^{WLS}\|_2 + \|\boldsymbol{\xi}_{i,L}^{WLS} - \boldsymbol{\xi}_{i,L}\|_2$. By Theorems 1 and 2, the consistency results of $\hat{\varsigma}_{ij}^2$ and $\hat{\boldsymbol{\phi}}_r$ imply $\hat{\boldsymbol{\xi}}_{i,L}^{WLS} \overset{a.s.}{\to} \boldsymbol{\xi}_{i,L}^{WLS}$ and $\hat{\boldsymbol{\Omega}}_{i,L}^{WLS} \overset{a.s.}{\to} \boldsymbol{\Omega}_{i,L}^{WLS}$. It remains to show that $\boldsymbol{\xi}_{i,L}^{WLS} \overset{a.s.}{\to} \boldsymbol{\xi}_{i,L}$ and $\boldsymbol{\Psi}_i^{-1/2}\left(\boldsymbol{\xi}_{i,L}^{WLS} - \boldsymbol{\xi}_{i,L}\right) \overset{d}{\to} N(0, \boldsymbol{I}_L)$, as $n \to \infty$.

Let $\widetilde{\mathbf{U}}_i^* = \boldsymbol{\Gamma}_i^{-1/2}\widetilde{\mathbf{U}}_i$, $\widetilde{\boldsymbol{\varrho}}_{i,L}^{\top} = \boldsymbol{\Gamma}_i^{-1/2}\widetilde{\boldsymbol{\varphi}}_{i,L}^{\top}$ and $\widetilde{\boldsymbol{\varkappa}}_i = \boldsymbol{\Gamma}_i^{-1/2}\widetilde{\boldsymbol{\varepsilon}}_i$, where $\widetilde{\boldsymbol{\varkappa}}_i = (\widetilde{\boldsymbol{\varkappa}}_{1i}^{\top}, \ldots, \widetilde{\boldsymbol{\varkappa}}_{pi}^{\top})^{\top}$ with $\widetilde{\boldsymbol{\varkappa}}_{ki} = (\varkappa_{ki1}, \ldots, \varkappa_{kim_i})^{\top}$. It follows that $\{\varkappa_{kij}\}$ are mutually independent with $E(\varkappa_{kij}^2) = 1$. By considering the transformed linear regression model $\widetilde{\mathbf{U}}_i^* = \widetilde{\boldsymbol{\varrho}}_{i,L}^{\top}\boldsymbol{\xi}_{i,L} + \widetilde{\boldsymbol{\varkappa}}_i$ and noting that $\widetilde{\boldsymbol{\varrho}}_{i,L}\widetilde{\boldsymbol{\varrho}}_{i,L}^{\top} = \boldsymbol{\Psi}_i$, the results follow under (i) and (ii), using Lemmas 1 and 2 of White (1980).

7. Concluding Remarks

The new $m\mathrm{FPC}_n$ model in (2.1) provides a general stochastic representation for a vector of multivariate random functions, which takes the varying extent of variations into account and takes advantage of dependency through pairwise correlations among the random functions. It serves as a basic and useful tool for multivariate FDA. In a data application, we treated the daily traffic trajectories as independent observations, even though the data may not fit the independence assumption. Further investigation of the dependency structures between daily traffic trajectories, and its effect on traffic flow prediction is warranted.

Supplementary Material

The supplement provides more detail on generating multivariate functional data and the simulation settings (Supplement S1), an additional simulation study (Supplement S2), and the proofs of Lemmas 1−2 and Corollary 1 (Supplement S3).

Acknowledgement

The authors thank Professor N. Wang and anonymous reviewers for helpful suggestions. This study was supported by NSC Grant 101-2118-M-001-011-MY3 and Academia Sinica Grant 99-ASIA-02.

References

Balakrishnan, A. (1960). Estimation and detection theory for multiple stochastic processes. *J. Math. Anal. Appl.* **1**, 386-410.

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

Boente, G. and Fraiman, R. (2000). Kernel-based functional principal components. *Statist. Probab. Lett.* **48**, 335-345.

Chiou, J.-M. (2012). Dynamical functional prediction and classification, with application to traffic flow prediction. *Ann. Appl. Statist.* **6**, 1588-1614.

Chiou, J.-M. and Li, P.-L. (2007). Functional clustering and identifying substructures of longitudinal data. *J. Roy. Statist. Soc. B* **69**, 679–699.

Chiou, J.-M. and Müller, H. G. (2014). Linear manifold modelling of multivariate functional data. *J. Roy. Statist. Soc. B* **76**, 605-626.

Dauxois, J., Pousse, A. and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *J. Multivariate Anal.* **12**, 136-156.

Deville, J. C. (1974). Méthodes statistiques et numériques de l'analyse harmonique. *Annale de l'INSEE* **15**, 3-101.

Dubin, J. and Müller, H. G. (2005). Dynamical correlation for multivariate longitudinal data. *J. Amer. Statist. Assoc.* **100**, 872-881.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications.* Chapman and Hall, London.

Gil', M. I. (1999). Perturbations of simple eigenvectors of linear operators. *Manuscripta Math.* **100**, 213-219.

Hall, P. and Hosseini-Nasab, M. (2006). On properties of functional principal components analysis. *J. Roy. Statist. Soc. B* **68**, 109-126.

Hall, P., Müller, H. G. and Wang, J. L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.* **34**, 1493-1517.

Hutson, V. and Pym, J. S. (1980). *Applications of Functional Analysis and Operator Theory.* Academic Press, London.

James, G., Hastie, T. G. and Sugar, C. A. (2001). Principal component models for sparse functional data. *Biometrika* **87**, 587-602.

Kato, T. (1984). *Perturbation Theory for Linear Operators.* Springer-Verlag, New York.

Kayano, M. and Konishi, S. (2009). Functional principal component analysis via regularized Gaussian basis expansions and its application to unbalanced data. *J. Statist. Plann. Inference* **139**, 2388-2398.

Kelly, E. and Root, W. (1960). A representation of vector-valued random processes. *J. Math. Phys.* **39**, 211-216.

Li, Y. and Hsing, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Ann. Statist.* **38**, 3321-3351.

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

Rice, J. and Silverman, B. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Roy. Statist. Soc. B* **53**, 233-243.

Silverman, B. W. (1996). Smoothed functional principal component analysis by choice of norm. *Ann. Statist.* **24**, 1-24.

Suyundykov, R., Puechmorel, S. and Ferre, L. (2010). Multivariate functional data clusterization by PCA in Sobolev space using wavelets. *42èmes Journées de Statistique.*

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* **48**, 817-838.

Withers, C.S. (1974). Mercer's theorem and Fredholm resolvents. *Bull. Austral. Math. Soc.* **11**, 373-380.

Yao, F., Müller, H. G. and Wang, J. L. (2005). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100**, 577-590.

Zhou, L., Huang, J. Z. and Carrol, R. (2008). Joint modelling of paired sparse functional data using principal components. *Biometrika* **95**, 601–619.

Institute of Statistical Science, Academia Sinica, 128 Academia Road, Section 2, Taipei 11529, Taiwan.

E-mail: jmchiou@stat.sinica.edu.tw

Institute of Statistical Science, Academia Sinica, 128 Academia Road, Section 2, Taipei 11529, Taiwan.

E-mail: bboy0302@gmail.com

Institute of Statistical Science, Academia Sinica, 128 Academia Road, Section 2, Taipei 11529, Taiwan.

E-mail: yfyang@stat.sinica.edu.tw

(Received October 2013; accepted November 2013)