Test of Partial Separability for Multivariate Functional Data

Fangzhi Luo, Wei Zhang and Decai Liang

Sun Yat-sen University, Peking University and Nankai University

Supplementary Material

S1 Proof of Theorem 1

S1.1 Preliminaries

Following standard definitions in Hilbert spaces, we denote the space of Hilbert-Schmidt operators on $L^2(\mathcal{T})$ or $(L^2(\mathcal{T}))^p$ as $\mathcal{B}_{HS}\{L^2(\mathcal{T})\}$ or $\mathcal{B}_{HS}\{(L^2(\mathcal{T}))^p\}$. Let $(f_1 \otimes f_2)$ be the operator from $L^2(\mathcal{T})$ to $L^2(\mathcal{T})$ defined by $(f_1 \otimes f_2)(g) = \langle f_1, g\rangle f_2$ for any $f_1, f_2, g \in L^2(\mathcal{T})$, and $(\boldsymbol{f}_1 \otimes_p \boldsymbol{f}_2)$ be the operator from $(L^2(\mathcal{T}))^p$ to $(L^2(\mathcal{T}))^p$ defined by $(\boldsymbol{f}_1 \otimes_p \boldsymbol{f}_2)(\boldsymbol{g}) = \langle \boldsymbol{f}_1, \boldsymbol{g}\rangle_p \boldsymbol{f}_2$ for any $\boldsymbol{f}_1, \boldsymbol{f}_2, \boldsymbol{g} \in (L^2(\mathcal{T}))^p$. The space of $\mathcal{B}_{HS}\{L^2(\mathcal{T})\}$ forms a Hilbert space, equipped with the inner product $\langle \cdot, \cdot \rangle_{HS}$ s.t. $\langle \mathcal{K}_1, \mathcal{K}_2\rangle_{HS} = \sum_{l=1}^{\infty} \langle \mathcal{K}_1 e_l, \mathcal{K}_2 e_l\rangle$ for any orthonormal basis $\{e_l\}$ in $L^2(\mathcal{T})$, and the induced norm $\|\cdot\|_{HS}$. Similarly, $\mathcal{B}_{HS}\{(L^2(\mathcal{T}))^p\}$ is a Hilbert space with the inner product $\langle \cdot, \cdot \rangle_{HS_p}$ s.t. $\langle \boldsymbol{\mathcal{K}}_1, \boldsymbol{\mathcal{K}}_2\rangle_{HS_p} = \sum_{j=1}^{\infty} \langle \boldsymbol{\mathcal{K}}_1 \boldsymbol{e}_j, \boldsymbol{\mathcal{K}}_2 \boldsymbol{e}_j\rangle_p$ for any orthonormal $\{\boldsymbol{e}_j\}$ in $(L^2(\mathcal{T}))^p$, and the induced norm $\|\cdot\|_p$. For an operator $\mathcal{K} \in \mathcal{B}_{HS}\{L^2(\mathcal{T})\}$ and a matrix $\mathbf{B} \in \mathbb{R}^{p\times p}$, the operator $\mathbf{B}\mathcal{K} \in \mathcal{B}_{HS}\{(L^2(\mathcal{T})^p\}$ is defined by $(\mathbf{B}\mathcal{K})\boldsymbol{f} = \mathcal{K}(\mathbf{B}\boldsymbol{f})$ for any $\boldsymbol{f} = (f_1, \ldots, f_p)^{\mathrm{T}} \in$

$(L^2(\mathcal{T}))^p$, where $\mathbf{B}\boldsymbol{f} \in (L^2(\mathcal{T})^p$ takes values $\mathbf{B}\{\boldsymbol{f}(t)\} \in \mathbb{R}^p$, $\mathcal{K}(\boldsymbol{f}) = \{\mathcal{K}(f_1), \ldots, \mathcal{K}(f_p)\}^{\mathrm{T}} \in (L^2(\mathcal{T}))^p$.

For a multivariate functional process $\boldsymbol{X} = (X_1, \ldots, X_p)^{\mathrm{T}} \in (L^2(\mathcal{T}))^p$ with $\mathbb{E}(\boldsymbol{X}) = \boldsymbol{\mu} = (\mu_1, \ldots, \mu_p)^{\mathrm{T}}$, the covariance operator $\boldsymbol{\mathcal{G}}$ is defined as

$$\boldsymbol{\mathcal{G}} = \mathbb{E}\{(\boldsymbol{X} - \boldsymbol{\mu}) \otimes_p (\boldsymbol{X} - \boldsymbol{\mu})\},$$

and for $j, k = 1, \ldots, p$, the cross-covariance operator $\mathcal{G}_{jk}$ for each $\{X_j, X_k\}$ is defined as

$$\mathcal{G}_{jk} = \mathbb{E}\left\{(X_j - \mu_j) \otimes (X_k - \mu_k)\right\}.$$

We refer to Chapter 7 of Hsing and Eubank (2015) for more details on the (cross) covariance operators in Hilbert space.

Following the definition of partial separability, we say that a covariance operator $\boldsymbol{\mathcal{G}}$ is partially separable if there exist an orthonormal basis $\{\varphi_l\}_{l=1}^{\infty}$ of $L^2(\mathcal{T})$ and a sequence of $p \times p$ positive-definite matrices $\{\boldsymbol{\Sigma}_l = (\sigma_l(j,k))_{j,k=1,\ldots p}\}$, s.t.

$$\boldsymbol{\mathcal{G}} = \sum_{l=1}^{\infty} \boldsymbol{\Sigma}_l \, \varphi_l \otimes \varphi_l, \tag{S1.1}$$

or equivalently, for any $j, k = 1, \ldots, p$,

$$\mathcal{G}_{jk} = \sum_{l=1}^{\infty} \sigma_l(j,k) \, \varphi_l \otimes \varphi_l. \tag{S1.2}$$

Suppose we observe i.i.d. samples $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ from $\boldsymbol{X}$. The sample mean function is obtained by $\hat{\boldsymbol{\mu}} = n^{-1} \sum_{i=1}^{n} \boldsymbol{X}_i$. We estimate $\boldsymbol{\mathcal{G}}$ by the sample covariance operator

$$\widehat{\boldsymbol{\mathcal{G}}} = \frac{1}{n} \sum_{i=1}^{n} (\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}) \otimes_p (\boldsymbol{X}_i - \hat{\boldsymbol{\mu}}),$$

and $\mathcal{G}_{jk}$ by

$$\widehat{\mathcal{G}}_{jk} = \frac{1}{n}\sum_{i=1}^{n}(X_{ij} - \hat{\mu}_j) \otimes (X_{ik} - \hat{\mu}_k).$$

If $\boldsymbol{\mathcal{G}}$ is partially separable, the marginal covariance operator $\mathcal{H} := \sum_{j=1}^{p} \mathcal{G}_{jj}$ has the decomposition

$$\mathcal{H} = \sum_{l=1}^{\infty} \lambda_l \, \varphi_l \otimes \varphi_l,$$

where $\lambda_l = \mathrm{tr}(\boldsymbol{\Sigma}_l)$. And $\mathcal{H}$ is estimated by $\widehat{\mathcal{H}} = \sum_{j=1}^{p} \widehat{\mathcal{G}}_{jj}$.

S1.2 Proof of Theorem 1

To alleviate the notation, we assume $\mathbb{E}(\boldsymbol{X}) = 0$ w.l.o.g. To begin with, we state the weak convergence result for the covariance estimators $\widehat{\boldsymbol{\mathcal{G}}}$, $\widehat{\mathcal{G}}_{jk}$ and $\widehat{\mathcal{H}}$, which is based on the central limit theorem for random elements of a Hilbert space (see. e.g. Theorem 7.7.6 of Hsing and Eubank, 2015). Define

$$\boldsymbol{\mathcal{Z}}_n = \sqrt{n}(\widehat{\boldsymbol{\mathcal{G}}} - \boldsymbol{\mathcal{G}}), \mathcal{Z}_n(j,k) = \sqrt{n}(\widehat{\mathcal{G}}_{jk} - \mathcal{G}_{jk}) \text{ and } \mathcal{Z}_{n,p} = \sqrt{n}(\widehat{\mathcal{H}} - \mathcal{H}) = \sum_{j=1}^{p} \mathcal{Z}_n(j,j).$$

By Condition 1 and Theorem 8.1.2 of Hsing and Eubank (2015), $\boldsymbol{\mathcal{Z}}_n$ converges weakly to a mean-zero Gaussian random element, say $\boldsymbol{\mathcal{Z}}$, in $B_{HS}\{(L^2(\mathcal{T}))^p\}$, which also implies the asymptotic normality of $\mathcal{Z}_n(j,k)$ and $\mathcal{Z}_{n,p}$.

Proof of Theorem 1(a). According to the definition of $T_n(l,j,l',k)$ and the

notations in Section S1.1, we have

$$\begin{aligned}
T_n(l,j,l',k) =& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\theta}_{i,lj}\hat{\theta}_{i,l'k} = \frac{1}{\sqrt{n}}\langle X_{ij},\hat{\varphi}_l\rangle\langle X_{ik},\hat{\varphi}_{l'}\rangle \\
=& \frac{1}{\sqrt{n}}\left\langle \sum_{i=1}^{n} X_{ij}\otimes X_{ik}, \hat{\varphi}_l\otimes\hat{\varphi}_{l'}\right\rangle_{HS} \\
=& \sqrt{n}\langle \widehat{\mathcal{G}}_{jk}, \hat{\varphi}_l\otimes\hat{\varphi}_{l'}\rangle_{HS} \\
=& \sqrt{n}\langle \widehat{\mathcal{G}}_{jk}-\mathcal{G}_{jk}, \varphi_l\otimes\varphi_{l'}\rangle_{HS} + \sqrt{n}\langle \mathcal{G}_{jk}, (\hat{\varphi}_l-\varphi_l)\otimes\varphi_{l'}\rangle_{HS} \\
&+ \sqrt{n}\langle \mathcal{G}_{jk}, \varphi_l\otimes(\hat{\varphi}_{l'}-\varphi_{l'})\rangle_{HS}
\end{aligned}$$

for$1 \le l \ne l' \le L_n, 1 \le j \le k \le p$. Recall that $\{\hat{\varphi}_l\}$ are the eigenfunctions of $\widehat{\mathcal{H}}$. By Condition 2 and Theorem 5.1.8 of Hsing and Eubank (2015), we have

$$(\hat{\varphi}_l-\varphi_l) = \mathcal{M}_l(\widehat{\mathcal{H}}-\mathcal{H})\varphi_l + o_p(\hat{\varphi}_l-\varphi_l),$$

where $\mathcal{M}_l = \sum_{m\ne l}(\lambda_l-\lambda_m)^{-1}\varphi_m\otimes\varphi_m \in \mathcal{B}_{HS}\{L^2(\mathcal{T})\}$. It follows that

$$\begin{aligned}
T_n(l,j,l',k) =& \sqrt{n}\langle \widehat{\mathcal{G}}_{jk}-\mathcal{G}_{jk}, \varphi_l\otimes\varphi_{l'}\rangle_{HS} + \sqrt{n}\langle \mathcal{G}_{jk}, \mathcal{M}_l(\widehat{H}-H)\varphi_l\otimes\varphi_{l'}\rangle_{HS} \\
&+ \sqrt{n}\langle \mathcal{G}_{jk}, \varphi_l\otimes\mathcal{M}_{l'}(\widehat{H}-H)\varphi_{l'}\rangle_{HS} + o_p(1). \qquad \text{(S1.3)}
\end{aligned}$$

Denote the first three terms on the right hand side of (S1.3) as I_1, I_2 and I_3, respectively. Recall the definitions of $\widehat{\mathcal{G}}_{jk}$ and $\widehat{\mathcal{H}}$ in Section S1.1, and notice that the partial separability assumption implies $\langle \mathcal{G}_{jk}, \varphi_l\otimes\varphi_{l'}\rangle = \mathbb{E}(\langle X_j,\varphi_l\rangle\langle X_k,\varphi_{l'}\rangle) = \mathbb{E}(\theta_{lj}\theta_{l'k}) = 0$ and similarly

$\langle H, \varphi_l \otimes \varphi_{l'} \rangle = 0$, it follows that

$$\begin{aligned} I_1 &= \sqrt{n}\langle \hat{\mathcal{G}}_{jk}, \varphi_l \otimes \varphi_{l'} \rangle_{HS} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \langle X_{ij} \otimes X_{ik}, \varphi_l \otimes \varphi_{l'} \rangle_{HS} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \langle X_{ij}, \varphi_l \rangle \langle X_{ik}, \varphi_{l'} \rangle = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \theta_{i,lj} \theta_{i,l'k}, \end{aligned} \tag{S1.4}$$

and then, noting (S1.2),

$$\begin{aligned} I_2 &= \left\langle \sum_{q=1}^{\infty} \sigma_q(j,k)\, \varphi_q \otimes \varphi_q, \mathcal{M}_l(\widehat{\mathcal{H}} - \mathcal{H}) \varphi_l \otimes \varphi_{l'} \right\rangle_{HS} \\ &= \sum_{q=1}^{\infty} \sigma_q(j,k) \langle \varphi_q, \mathcal{M}_l\, \widehat{\mathcal{H}}\, \varphi_l \rangle \langle \varphi_q, \varphi_{l'} \rangle, \end{aligned}$$

where the infinite sum and inner product are exchangeable since $\sum_{q=1}^{\infty} \sigma_q(j,k) < \infty$. Due to the orthogonality of eigenfunctions, i.e., $\langle \varphi_l, \varphi_{l'} \rangle = I(l = l')$, we then have

$$\begin{aligned} I_2 =& \sigma_{l'}(j,k) \left\langle \varphi_{l'}, \mathcal{M}_l\, \widehat{H}\, \varphi_l \right\rangle \\ =& \frac{1}{\sqrt{n}} \sigma_{l'}(j,k) \cdot \sum_{m \neq l} \sum_{j=1}^{p} \sum_{i=1}^{n} \left\langle \varphi_{l'}, \left\{ \left((\lambda_l - \lambda_m)^{-1} \varphi_m \otimes \varphi_m \right) (X_{ij} \otimes X_{ij}) \right\} \varphi_l \right\rangle \\ =& \frac{1}{\sqrt{n}} \sigma_{l'}(j,k) \cdot \sum_{m \neq l} \sum_{j=1}^{p} \sum_{i=1}^{n} \sum_{q,q'=1}^{\infty} \left\langle \varphi_{l'}, \left\{ (\lambda_l - \lambda_m)^{-1} (\theta_{i,qj} \theta_{i,q'j}) \langle \varphi_m, \varphi_{q'} \rangle \langle \varphi_q, \varphi_l \rangle \right\} \varphi_m \right\rangle \\ =& \frac{1}{\sqrt{n}} \sigma_{l'}(j,k) \cdot \sum_{m \neq l} \sum_{j=1}^{p} \sum_{i=1}^{n} \left\langle \varphi_{l'}, \left\{ (\lambda_l - \lambda_m)^{-1} \theta_{i,lj} \theta_{i,mj} \right\} \varphi_m \right\rangle \\ =& \frac{1}{\sqrt{n}} \sigma_{l'}(j,k) \cdot \sum_{m \neq l} \sum_{j=1}^{p} \sum_{i=1}^{n} \left\{ (\lambda_l - \lambda_m)^{-1} \theta_{i,lj} \theta_{i,mj} \right\} \langle \varphi_{l'}, \varphi_m \rangle \\ =& \frac{1}{\sqrt{n}} \frac{\sigma_{l'}(j,k)}{\lambda_l - \lambda_{l'}} \sum_{i=1}^{n} \sum_{j=1}^{p} \theta_{i,lj} \theta_{i,l'j}. \end{aligned} \tag{S1.5}$$

Similarly to (S1.5), it can be derived that

$$I_3 = \left\langle \mathcal{G}_{jk}, \varphi_l \otimes \left(\mathcal{M}_{l'}\, \widehat{H}\, \varphi_{l'} \right) \right\rangle_{HS} = \frac{1}{\sqrt{n}} \frac{\sigma_l(j,k)}{\lambda_{l'} - \lambda_l} \sum_{i=1}^{n} \sum_{j=1}^{p} \theta_{i,lj} \theta_{i,l'j}. \tag{S1.6}$$

Combining (S1.4) to (S1.6), we have

$$\begin{aligned} T_n(l,j,l',k) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\theta_{i,lj}\theta_{i,l'k} + \frac{1}{\sqrt{n}}\frac{\sigma_{l'}(j,k)}{\lambda_l-\lambda_{l'}}\sum_{i=1}^{n}\sum_{j=1}^{p}\theta_{i,lj}\theta_{i,l'j} \\ &\quad + \frac{1}{\sqrt{n}}\frac{\sigma_{l}(j,k)}{\lambda_{l'}-\lambda_{l}}\sum_{i=1}^{n}\sum_{j=1}^{p}\theta_{i,lj}\theta_{i,l'j} + o_p(1) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\theta_{i,lj}\theta_{i,l'k} + \frac{\eta_{l,l'}(j,k)}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{p}\theta_{i,lj}\theta_{i,l'j} + o_p(1), \end{aligned}$$

where $\eta_{l,l'}(j,k) = (\lambda_l - \lambda_{l'})^{-1}(\sigma_{l'}(j,k) - \sigma_l(j,k))$. Define

$$Z_n(l,j,l',k) := \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\theta_{i,lj}\theta_{i,l'k} + \frac{\eta_{l,l'}(j,k)}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{p}\theta_{i,lj}\theta_{i,l'j}, \tag{S1.7}$$

it follows that $T_n(l,j,l',k) = Z_n(l,j,l',k)+o_p(1)$. The asymptotic normality of $Z_n(l,j,l',k)$ follows directly from the standard central limit theorem for random variables, observing that $\mathbb{E}\left\|\eta_{l,l'}(j,k)\sum_{j=1}^{p}\theta_{lj}\theta_{l'j} + \theta_{lj}\theta_{l'k}\right\|^2 < \infty$ implied by Condition 1. This completes the proof of Theorem 1(a).

Proof of Theorem 1(b). Based on the arguments for deriving I_1 to I_3 in the proof of Theorem 1(a), we actually have $Z_n(l,j,l',k) = I_1 + I_2 + I_3$. According to (S1.3) and recall the definitions of $\mathcal{Z}_n(j,k)$ and $\mathcal{Z}_{n,p}$, we can rewrite $Z_n(l,j,l',k)$ by

$$\begin{aligned} Z_n(l,j,l',k) &= \langle \mathcal{Z}_n(j,k), \varphi_l \otimes \varphi_{l'}\rangle_{HS} + \sigma_{l'}(j,k)\langle \varphi_{l'}, \mathcal{M}_l\mathcal{Z}_{n,p}\,\varphi_l\rangle + \sigma_l(j,k)\langle \varphi_l, \mathcal{M}_{l'}\mathcal{Z}_{n,p}\,\varphi_{l'}\rangle \\ &= \langle \mathcal{Z}_n(j,k), \varphi_l \otimes \varphi_{l'}\rangle_{HS} + \langle \sigma_{l'}(j,k)\mathcal{M}_l\mathcal{Z}_{n,p} + \sigma(j,k)\mathcal{M}_{l'}\mathcal{Z}_{n,p}, \varphi_l \otimes \varphi_{l'}\rangle_{HS} \\ &= \langle \mathcal{Z}_n(j,k) + \sigma_{l'}(j,k)\mathcal{M}_l\mathcal{Z}_{n,p} + \sigma(j,k)\mathcal{M}_{l'}\mathcal{Z}_{n,p}, \varphi_l \otimes \varphi_{l'}\rangle_{HS}, \end{aligned} \tag{S1.8}$$

which actually defines a continuous linear mapping from $\boldsymbol{\mathcal{Z}}_n$ to $Z_n(l,j,l',k)$. To be explicit, define $\widetilde{\boldsymbol{\varphi}}_{lj} \in (L^2(\mathcal{T}))^p$ such that the jth element of $\widetilde{\boldsymbol{\varphi}}_{lj}$ is φ_l and other el-

ements are zero, i.e., $\widetilde{\boldsymbol{\varphi}}_{lj} = (0, \cdots, \varphi_l, \cdots, 0)^{\mathrm{T}}$. Consider the continuous mapping $\mathcal{A}_{jkll'}$ from $\mathcal{B}_{HS}\{(L^2(\mathcal{T}))^p\}$ to $\mathbb{R}$ that $\mathcal{A}_{jkll'}(\boldsymbol{\mathcal{F}}) = \langle \boldsymbol{\mathcal{F}}, \widetilde{\boldsymbol{\varphi}}_{lj} \otimes_p \widetilde{\boldsymbol{\varphi}}_{l'k} \rangle_{HS}$ for any $\boldsymbol{\mathcal{F}} \in \mathcal{B}_{HS}\{(L^2(\mathcal{T}))^p\}$, then $\langle \mathcal{Z}_n(j,k), \varphi_l \otimes \varphi_{l'} \rangle_{HS} = \mathcal{A}_{jkll'}(\boldsymbol{\mathcal{Z}}_n)$ and $\langle \mathcal{M}_l \mathcal{Z}_{n,p}, \varphi_l \otimes \varphi_{l'} \rangle_{HS} = \sum_{q=1}^{p} \mathcal{A}_{qqll'}(\boldsymbol{I}_p \mathcal{M}_l \boldsymbol{\mathcal{Z}}_n)$, where $\boldsymbol{I}_p$ is the p-dimensional identity matrix. Denote $\boldsymbol{Z}_n = (\mathcal{Z}_n(l,j,l',k) : 1 \le l \ne l' \le L_n, 1 \le j \le k \le p)^{\mathrm{T}}$, which is the counterpart of $\boldsymbol{T}_n$ with length q. Let $\Psi_{l,j,l',k} = \mathcal{A}_{jkll'} + \sigma_{l'}(j,k) \sum_{q=1}^{p} \mathcal{A}_{qqll'} \boldsymbol{I}_p \mathcal{M}_l + \sigma_l(j,k) \sum_{q=1}^{p} \mathcal{A}_{qqll'} \boldsymbol{I}_p \mathcal{M}_l'$, and Ψ be the mapping from $\mathcal{B}_{HS}\{(L^2(\mathcal{T}))^p\}$ to $\mathbb{R}^q$ with $\Psi(\boldsymbol{\mathcal{F}})_{l,j,l',k} = \Psi_{l,j,l',k}(\boldsymbol{\mathcal{F}})$, i.e. for any $\boldsymbol{\mathcal{F}} \in \mathcal{B}_{HS}\{(L^2(\mathcal{T}))^p\}$, $\Psi(\boldsymbol{\mathcal{F}})$ is a q-dimensional vector with the (l,j,l',k)th element being $\Psi_{l,j,l',k}(\boldsymbol{X})$. Based on (S1.8) and the definition of $\mathcal{A}_{jkll'}$ and $\Psi_{l,j,l',k}$, we actually have

$$\boldsymbol{Z}_n = \Psi(\boldsymbol{\mathcal{Z}}_n),$$

and it follows from the asymptotic normality of $\boldsymbol{\mathcal{Z}}_n$ and the continuous mapping theorem in metric spaces (e.g. Aston et al., 2017) that $\boldsymbol{Z}_n$ converges to a Gaussian random vector in $\mathbb{R}^q$, which consequently implies that $\boldsymbol{T}_n$ is asymptotically jointly Gaussian.

We then derive the asymptotic covariance structure $\boldsymbol{\Theta}$ based on the form (S1.7) of $Z_n(l,j,l',k)$. Define $\kappa_{l,l',j_1,j_2,k_1,k_2} = \mathbb{E}\left(\theta_{lj_1}\theta_{l'k_1}\theta_{lj_2}\theta_{l'k_2}\right)$ and note that partial separability implies $\mathbb{E}\left(\theta_{l_1 j_1}\theta_{l_1' k_1}\theta_{l_2 j_2}\theta_{l_2' k_2}\right) = I(l_1 = l_2) I(l_1' = l_2') \kappa_{l_1,l_1',j_1,j_2,k_1,k_2}$ and

$$\frac{1}{n}\mathbb{E}\left\{\left(\sum_{i=1}^{n}\sum_{j=1}^{p}\theta_{i,l_1 j}\theta_{i,l_1' j}\right)\left(\sum_{i'=1}^{n}\sum_{k=1}^{p}\theta_{i',l_2 k}\theta_{i',l_2' k}\right)\right\} = I(l_1 = l_2) I(l_1' = l_2') \sum_{j=1}^{p}\sum_{k=1}^{p} \kappa_{l_1,l_1',j,j,k,k},$$

$$\frac{1}{n}\mathbb{E}\left\{\left(\sum_{i=1}^{n}\sum_{j=1}^{p}\theta_{i,l_1 j}\theta_{i,l_1' j}\right)\left(\sum_{i'=1}^{n}\theta_{i',l_2 j_2}\theta_{i',l_2' k_2}\right)\right\} = I(l_1 = l_2) I(l_1' = l_2') \sum_{j=1}^{p} \kappa_{l_1,l_1',j,j_2,j,k_2}.$$

It then follows that

$$\mathbb{E}\left\{Z_n(l_1,j_1,l_1',k_1)Z_n(l_2,j_2,l_2',k_2)\right\} = I(l_1=l_2)I(l_1'=l_2')\left\{\sum_{j=1}^{p}\kappa_{l_1,l_1',j,j_2,j,k_2}\eta_{l_1l_1'}(j_1,k_1)\right.$$
$$\left.+\sum_{j=1}^{p}\kappa_{l_1,l_1',j,j_1,j,k_1}\eta_{l_1l_1'}(j_2,k_2)+\kappa_{l_1,l_1',j_1,j_2,k_1,k_2}+\sum_{j=1}^{p}\sum_{k=1}^{p}\kappa_{l_1,l_1',j,j,k,k}\eta_{l_1l_1'}(j_1,k_1)\eta_{l_1l_1'}(j_2,k_2)\right\},$$

which gives the expression of $\boldsymbol{\Theta}\left(l_1,j_1,l_1',k_1,l_2,j_2,l_2',k_2\right)$. Practically, a plug-in estimator of $\boldsymbol{\Theta}$ can be used with $\sigma_l(j,k)$ and $\kappa_{l,l',j_1,j_2,k_1,k_2}$ being estimated by $n^{-1}\sum_{i=1}^{n}(\theta_{i,lj}\theta_{i,lk})$ and $n^{-1}\sum_{i=1}^{n}(\theta_{i,lj_1}\theta_{i,l'k_1}\theta_{i,lj_2}\theta_{i,l'k_2})$.

S1.3 Proof of Corollary 1

Recall the definition of S_n and let the spectral decomposition of the matrix $\boldsymbol{\Theta}$ be $\boldsymbol{\Theta} = U\boldsymbol{\Gamma}U^T$, where U is orthonormal and $\boldsymbol{\Gamma}$ is diagonal with diagonal entries $\gamma_1,\ldots,\gamma_q$ being the eigenvalues of $\boldsymbol{\Theta}$. Based on the null distribution of $\boldsymbol{T}_n$ in Theorem 1, it then follows that

$$S_n = \|\boldsymbol{T}_n\|^2 = \|U^{\mathrm{T}}\boldsymbol{T}_n\|^2 = \Sigma_{i=1}^{q}\gamma_i Z_i^2,$$

where $\{Z_i\}$ are i.i.d standard normal random variables, which indicates that the null distribution of S_n is a χ^2 type mixture distribution.

On the other hand, if $\boldsymbol{X}(t)$ is not partially separable, then by definition the covariance of $\theta_{i,lj}$ and $\theta_{i,l'k}$ would not be zero. As a result, the term I_1 in the proof of Theorem 1(a) is of order $n^{1/2}$ for at least one set of (l,l'), which leads to $S_n\to\infty$ in probability under alternative hypothesis.

S2 Analysis for Spatial Correlated Functional Data

S2.1 Motivation

Let $X(\boldsymbol{s}, t)$ be a mean-zero spatio-temporal process on $\mathcal{S} \times \mathcal{T}$, where $\mathcal{S} \subset \mathbb{R}^2$ is the spatial domain and $\mathcal{T}$ is the time domain. Following Liang et al. (2022), a weakly separable spatial functional field is represented by

$$X(\boldsymbol{s}, t) = \sum_{l=1}^{\infty} \xi_l(\boldsymbol{s}) \varphi_l(t), \tag{S2.9}$$

where $\{\varphi_l(\cdot)\}$ is an orthonormal basis in $L^2(\mathcal{T})$ and $\{\xi_l(\cdot)\}$ are uncorrelated spatial random fields. Consequently, the cross covariance of $X(\boldsymbol{s}, t)$ yields

$$C(\boldsymbol{s}_1, t_1; \boldsymbol{s}_2, t_2) = \sum_{r=1}^{\infty} c_l(\boldsymbol{s}_1, \boldsymbol{s}_2) \varphi_l(t_1) \varphi_l(t_2),$$

where $c_l(\cdot, \cdot)$ is the spatial covariance function of $\xi_l(\cdot)$. As a common practice, the temporal component $\varphi_l(\cdot)$ can be modelled using nonparametric methods in FDA, while the spatial covariance $c_l(\cdot, \cdot)$ is often restricted to some parametric class such as the well-known Matérn family. This semiparametric modeling approach is widely employed in pieces of literatures on spatiotemporal data or spatial functional data (e.g. Liu et al., 2017; **?**; Zhang and Li, 2021).

Suppose we observe spatially correlated functional data $X_i(\boldsymbol{s}_j, t)$ over spatial locations $\{\boldsymbol{s}_j\}_{j=1}^{p}$ for each subject $i = 1, \ldots, n$. Following the definition of partial separability, one can assume that

$$X_i(\boldsymbol{s}_j, t) = \sum_{l=1}^{\infty} \theta_{il}(\boldsymbol{s}_j) \varphi_l(t), \tag{S2.10}$$

with p-variate random vector $\boldsymbol{\theta}_{il} = (\theta_l(\boldsymbol{s}_1), \ldots, \theta_l(\boldsymbol{s}_p))^{\mathrm{T}}$ and $\mathrm{cov}\{\theta_{il}(\boldsymbol{s}_j), \theta_{l'}(\boldsymbol{s}_k)\} = 0$ for any $l \neq l'$. This actually coincides with the definition of weak separability (S2.9) or a similar notion called coregionalization (e.g. Zhang and Li, 2021). Typically the spatial correlation appears to be smaller as the distance between two locations decreases, which is different from the typical multivariate functional data (e.g. the multichannel profile data) whose multivariate dependence is in general free of correlation structure. As a result, the performance of the proposed test for such type of spatially correlated functional data should be specially studied.

Note that for classic spatiotemporal data, it is often the case that only one realization of $\boldsymbol{X}(s,t)$ is available, which gives rise to challenges for the estimation of $C(\boldsymbol{s}_1, t_1; \boldsymbol{s}_2, t_2)$. As a result, the spatial covariance is often assumed to be stationary, isotropic or separable (e.g. Liu et al., 2017; **?**; Li and Guan, 2014). This also distinguishes the proposed test that is suitable to spatially correlated functional data with replicates, from the test in Liang et al. (2022) which aims at non-replicated spatial functional field and requires the assumption of spatial stationarity.

S2.2 A simulation study

To generate spatially correlated functional data, we modify the model (4.17) as

$$X_i(\boldsymbol{s}_j; t) = \sum_{l=1}^{L=10} \theta_{il}(\boldsymbol{s}_j)\varphi_l(t),$$

where the spatial locations $\{\boldsymbol{s}_j\}_{j=1}^p$ consist of regular grids $\{\frac{2}{\sqrt{p}},\ldots,2\}\otimes\{\frac{2}{\sqrt{p}},\ldots,2\}$ on $\mathcal{S}=[0,2]\times[0,2]$, and $\varphi_l(t)=\sqrt{2}\cos(l\pi t)$ when l is odd and $\varphi_l(t)=\sqrt{2}\sin\{(l-1)\pi t\}$ when l is even, under the same setup as Liang et al. (2022). To reflect the spatial correlation, the covariance matrix in Section 4.1 is modified by

(i*) $\boldsymbol{\Sigma}_{ps}=\mathrm{diag}(\boldsymbol{\Sigma}_1,\ldots,\boldsymbol{\Sigma}_L)$ with $\sigma_l(j,k)=\omega_l\,\mathrm{M}(|\boldsymbol{s}_j-\boldsymbol{s}_k|;\nu_l,\phi_l)$, where $\mathrm{M}(d;\nu,\phi)$ denotes the isotropic Matérn covariance function;

(ii*) $\boldsymbol{\Sigma}_{nps}$ with the diagonal block $\boldsymbol{\Sigma}_l$ in (i*) and the off-diagonal blocks $\boldsymbol{\Sigma}_{1,2}=\boldsymbol{\Sigma}_{2,1}=\{\sigma_{12}(j,k)\}$, where $\sigma_{12}(j,k)=\rho_{12}\sqrt{\omega_1\omega_2}\mathrm{M}(|\boldsymbol{s}_j-\boldsymbol{s}_k|;\nu_{12},\phi_{12})$.

To ensure the validity of bivariate Matérn cross-covariance $\sigma_{12}(j,k)$, the values of (ν_1,ν_{12},ν_2) are determined following the condition in Theorem 3 of **?**. The specific values of parameters $\{\omega_l,\phi_l,\nu_l\}_{l=1}^L$ are set the same as that in Liang et al. (2022). Similarly to the setup in Section 4.1, the parameter ρ_{12} controls the departure from partial separability, and $\boldsymbol{\Sigma}_{nps}$ degenerates to $\boldsymbol{\Sigma}_{ps}$ when $\rho_{12}=0$. In this scenario, we adopt a spatially stationary covariance structure for $X(s_j,t)$, but our tests are in general free of structural covariance assumptions.

We evaluate the size or power of the proposed testing procedures through 1000 or 200 runs. From the rejection rate results in Table 1, we can see both χ^2 type mixture and high-dimensional tests control the type I error well, except that the high-dimensional test is slightly undersized when $p=25$. As ρ increases, both tests become more powerful, and it is clear that χ^2 test behaves uniformly more powerful than high-dimensional test. This

Table 1: Empirical rejection rates(%) of χ^2 type mixture test (χ^2) and high-dimensional test (HD) for spatially correlated functional data.

		ρ=0 ($\mathbb{H}_0$)		ρ=0.2		ρ=0.4		ρ=0.6	
p	FVE	χ^2	HD	χ^2	HD	χ^2	HD	χ^2	HD
25	80%	4.3	3.1	30.0	9.5	99.5	44.0	100	95.0
	90%	4.8	3.0	28.0	9.5	92.5	44.0	100	95.0
64	80%	4.4	3.8	87.5	17.5	100	94.5	100	100
	90%	5.2	3.8	79.5	17.5	100	94.5	100	100
100	80%	3.6	4.2	96.5	14.5	100	97.0	100	100
	90%	4.4	4.2	92.5	14.5	100	97.0	100	100

indicates that for spatially correlated multivariate functional data, the quadratic form of statistics may be more efficient to detect the violation of partial separability compared with the l^∞-norm statistic. It can also be observed that the powers appear to be larger as p increases, since a larger size of spatial samples that are more correlated can be included and it is common in spatial statistics that the departure from the null hypothesis is usually stronger when the test statistic is constructed based on more adjacent spatial points (Liang et al., 2022). In summary, we conclude that the χ^2 type mixture test is preferable for spatially correlated functional data especially when p is not large.

S2.3 Real data example

Since January 2013, a large monitoring network for air quality assessment has been established in China. We consider the hourly recordings of PM2.5, the fine particulate matter with aerodynamic diameters less than $2.5\mu m$, from 11 Guokong (state controlled) monitoring stations in Beijing (**?**). The hourly PM2.5 concentration observed at each station is regarded as a realization of a random function with $T = 24$, and the multiple curves across $p = 11$ stations constitute the spatially correlated functional data for use.

Liang et al. (2022) studied the weak separability of a non-replicated spatio-temporal PM2.5 field in North China Plain, a broader topographical region including Beijing, within a time interval from December 1 to 30, 2016. By contrast, we work with PM2.5 recordings from March 1, 2015 to February 29, 2016, consisting of spatially correlated functional data with $n = 366$ replicates. Provided with multiple realizations across different days, the partial separability test we considered is free of stationary or isotropic assumption.

We apply the χ^2 type mixture test due to its superior performance in Section S2.2. As shown by Table 2, for the PM2.5 data in the whole year, the p-value is larger than 0.05 if FVE$< 95\%$ or $L_n < 7$, indicating the rationality of a partial separability model. To give a more rounded analysis, we then divide the data into four seasons, i.e., spring (March to May), summer (June to August), autumn (September to November) and winter (December to February), and conduct the tests separately. As presented by

Table 2: The p-values of χ^2 type mixture (χ^2) and high-dimensional (HD) test for Beijing's PM2.5 data.

	Whole Year	Spring	Summer	Autumn	Winter
FVE=80%	L_n=2	L_n=2	L_n=2	L_n=2	L_n=2
χ^2	0.434	0.425	0.141	0.534	0.642
HD	0.531	0.800	0.671	0.307	0.363
FVE=90%	L_n=4	L_n=5	L_n=5	L_n=3	L_n=3
χ^2	0.064	0.010	0.003	0.435	0.602
HD	0.112	0.497	0.293	0.336	0.399
FVE=95%	L_n=7	L_n=9	L_n=10	L_n=5	L_n=7
χ^2	0.047	0.003	< 0.001	0.360	0.600
HD	0.129	0.441	0.293	0.353	0.353

the remaining four columns of Table 2, both tests do not reject partial separability in autumn or winter, while the χ^2 type mixture test yields p-values less than 0.01 in spring and summer when FVE$\geq$ 90%. This indicates that the violation of partial separability for Beijing's PM2.5 data may arise in spring and summer, and a partially separable model with more than three components is more appropriate to be assumed in autumn and winter. In addition, the high-dimensional test tends to have a larger p-value than χ^2 test in spring and summer, which is possibly due to its low power for spatially correlated functional data as suggested in Section S2.2.

S3 Additional Simulation Results

S3.1 A more complex case for high-dimensional functional data

We consider a more realistic simulation setting of graphical functional data to resemble real-world data. We modify setting (ii†) in Section 4.2 and generate the covariance $\mathbf{\Sigma}_{nps}$ of $\boldsymbol{\theta}_l$'s as follows.

- First, generate $\{\Omega_l\}_{l=1,\ldots,L=11}$ following the same procedure in Section 4.2.
- Second, generate the off-block precision matrices $\Omega_{l,l'} = 0.5\{\Omega_l - \text{diag}(\Omega_l) + \Omega_{l'} - \text{diag}(\Omega_{l'})\}$ for $1 \leq l \neq l' \leq L$.
- Next, for each $l \neq l'$, randomly set a proportion of $(1 - 1/|l - l'|)$ elements in $\Omega_{l,l'} \in \mathbb{R}^{p\times p}$ as zero.
- Finally, compute $\mathbf{\Sigma}_{nps}$ by $\mathbf{\Sigma}_{nps} = \text{diag}(\mathbf{\Sigma}_{ps})^{1/2}\mathbf{\Omega}_{nps}^{-1}\text{diag}(\mathbf{\Sigma}_{ps})^{1/2}$, where $\mathbf{\Omega}_{nps} \in \mathbb{R}^{Lp\times Lp}$ is formed by stacking the block matrices $\{\Omega_{l,l'}\}_{l,l'=1,\ldots,L}$, and $\mathbf{\Sigma}_{ps}$ is defined in setting (i†) of Section 4.2.

The third step of the above approach is utilized to avoid the singularity of $\mathbf{\Omega}_{nps}$ and attenuate the signals when the indices l and l' become larger. Additionally, by generating the covariance structure using precision matrices, this approach ensures that each $\Omega_{l,l'}$'s contains nonzero elements, reflecting a more complex graphical structure for $\boldsymbol{\theta}_l$'s. Therefore we refer to it as the "dense setting", which differs from the "sparse setting" described in setting (ii†) of Section 4.2. To illustrate the difference between them, Figure 1

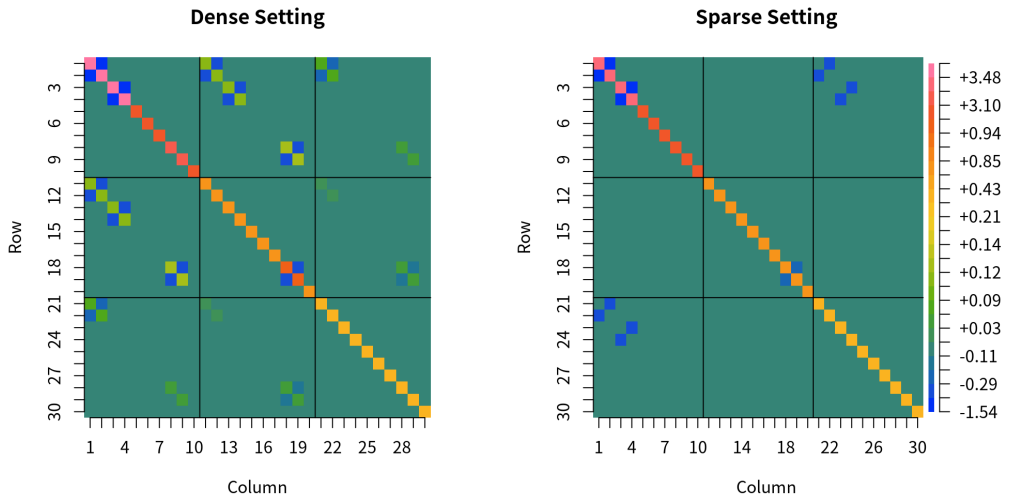

Figure 1: Precision matrices $\mathbf{\Omega}_{nps}$ of the first three principal scores $(\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T, \boldsymbol{\theta}_3^T) \in \mathbb{R}^{3p}$ for the dense setting and sparse setting in one simulation. Each matrix is divided into blocks by the black lines, with the (l, l')th block representing $\Omega_{l,l'}$. The matrix elements are visualized using colors, with the color bar on the right side indicating the corresponding values.

visualizes the covariance structure of the leading three random coefficients $(\boldsymbol{\theta}_1^{\mathrm{T}}, \boldsymbol{\theta}_2^{\mathrm{T}}, \boldsymbol{\theta}_3^{\mathrm{T}})^{\mathrm{T}}$ in one simulation with $p = 10$ and $\pi = 0.3$. It can be observed that the covariance structure for the "dense setting" exhibits nonzero elements in all off-diagonal blocks, while the covariance for the "sparse setting" only has nonzero elements in $\Sigma_{1,3}$ and $\Sigma_{3,1}$.

We evaluate the test performance for both "dense" and "sparse" settings with $p = 64$, which is identical to that for the EEG data in Section 5.2. The corresponding empirical rejection rates with various sparsity parameters π are collected in Table 3.

Table 3: Empirical rejection rates(%) of χ^2 type mixture test (χ^2) and high-dimensional test (HD) for dense setting and sparse setting of graphical functional data.

	π=0.05		π=0.1		π=0.2		π=0.3	
FVE	χ^2	HD	χ^2	HD	χ^2	HD	χ^2	HD
Dense setting								
80%	56.0	97.0	77.5	99.5	97.5	100	100	100
90%	65.0	97.0	89.0	99.5	100	100	100	100
Sparse setting								
80%	18.5	79.0	23.5	76.5	33.5	74.5	46.0	75.5
90%	20.5	94.5	30.0	99.5	44.5	99.5	56.5	99.5

Similar to the results in Table 2 of our manuscript, it can be observed that the high-dimensional test behaves uniformly more powerful than the χ^2 type mixture test, and both tests become more powerful when π increases. In particular, both tests exhibit higher power under the "dense setting" compared to the "sparse" setting, which aligns with our expectation as the "dense" setting has a larger number of nonzero elements in the off-diagonal blocks of $\mathbf{\Omega}_{nps}$, indicating more deviation from partial separability. These findings are consistent with those presented in Section 4.2.

S3.2 Decay Rate of the Eigenvalues

To investigate the effect of FVE under different decay rates, we extend the settings in case (i) of Section 4.1 by

$$\sigma_l(j,k) = \mathrm{cov}\{\theta_{i,lj}, \theta_{i,lk}\} = l^{-a}\, r^{|j-k|} \tag{S3.11}$$

Table 4: Empirical rejection rates(%) of χ^2 type mixture test (χ^2) and high-dimensional test (HD) under different decay rates of eigenvalues.

		$\rho = 0$ ($\mathbb{H}_0$)		ρ=0.3		ρ=0.5		ρ=0.7	
r	FVE	χ^2	HD	χ^2	HD	χ^2	HD	χ^2	HD
					$a = 1.5$				
0.2	80%	5.4	3.4	15.0	11.0	41.5	34.0	77.5	66.0
	90%	4.9	3.4	13.5	11.0	39.0	34.0	72.0	65.5
0.4	80%	4.9	3.8	52.0	44.0	98.5	92.0	100	100
	90%	4.9	3.7	49.0	44.0	98.5	92	100	100
					$a = 2$				
0.2	80%	4.5	3.8	15.0	10.5	51.0	38.5	85.5	78.0
	90%	5.2	3.8	14.5	10.5	48.5	37.5	82.5	78.0
0.4	80%	4.8	3.7	70.0	52.5	100	99.0	100	100
	90%	4.6	3.7	67.5	52.5	100	99.0	100	100
					$a = 2.5$				
0.2	80%	5.2	3.8	18.0	12.0	58.0	43.0	94.0	82.5
	90%	5.2	3.8	18.0	12.0	58.0	43.0	94.0	82.5
0.4	80%	5.2	2.8	81.0	70.5	100	99.5	100	100
	90%	5.2	2.8	81.0	70.5	100	99.5	100	100

where $a = 1.5$, 2 or 2.5 reflects different decay rates of the eigenvalue, and a larger value of a yields a faster decay rate. We then conduct the simulation following the same procedure under the multivariate normal case in Section 4.1, and display the results in Table 4. Note that the result for $a = 2$ is actually the same as that in Table 1 of our

manuscript.

Table 4 shows that the proposed tests using FVE perform well in controlling the type I errors across various decay rates. Furthermore, the test power generally increases as the decay rate becomes faster. This is consistent with the expectation that the tests tend to be more powerful when a smaller value of L_n is chosen, as the signal of deviation from partial separability is concentrated in the off-diagonal block $\Sigma_{1,2}$. Specifically, we observe that as the decay rate becomes faster, the selected L_n at an identical FVE level becomes relatively smaller; e.g. as a increases from 1.5 to 2.5, the averaged value of L_n decreases from 2.2 to 2 when FVE= 80% , and from 3 to 2 when FVE= 90%. These numerical results indicate that the FVE criterion is adaptive to different decay rates and effectively determines the value of L_n.

S4 More Numerical Results on the Relationship between Partial Separability and Separability

To further investigate the connection between partial separability and separability, we compare the proposed tests for partial separability with those for separability through empirical studies. To test the assumption of separability, we employ the testing procedure proposed by Aston et al. (2017), of which the main idea is introduced as follows. Consider a random element $X \in H_1 \otimes H_2$, where H_1, H_2 are two separable Hilbert spaces. Let $\widehat{C}_N$ be the sample covariance operator of $X_1, \ldots, X_N \overset{\text{i.i.d.}}{\sim} X$ and $\widehat{C}_{1,N}$ or $\widehat{C}_{2,N}$ be the marginal covariance estimator on H_1 or H_2. The separability test is based

on the projection from $\widehat{C}_N - \widehat{C}_{1,N}\tilde{\otimes}\widehat{C}_{2,N}$ onto the tensor product of the first several eigenfunctions of $\widehat{C}_{1,N}$ and $\widehat{C}_{2,N}$, where $\widehat{C}_{1,N}\tilde{\otimes}\widehat{C}_{2,N}$ provides a separable approximation of $\widehat{C}_N$. Refer to Aston et al. (2017) for more technical details. To implement this separability test for the considered multivariate functional process $X \in (L^2(\mathcal{T}))^p$, we treat X as an element in $L^2(\{1,\ldots,p\}) \otimes L^2(\mathcal{T})$. We denote $\widehat{C}_{1,N}$ or $\widehat{C}_{2,N}$ as the marginal covariance operators of X on $L^2(1,\ldots,p)$ or $L^2(\mathcal{T})$ respectively. To preserve the multivariate dependence structure we choose p eigenfunctions of $\widehat{C}_{1,N}$ for the test statistics, whereas the first L_n eigenfunctions of $\widehat{C}_{2,N}$ are involved, the latter is consistent with the truncation number in the proposed partial separability tests.

S4.1 Simulation

We conduct the simulation study by extending the covariance setting in Section 4.1 to two cases: separable (SEP) and partially separable but not separable (PNS).

a. For the SEP case, we use the same covariance $\mathbf{\Sigma}_{ps}$ as case (i) of Section 4.1. According to Proposition 1(b) of our manuscript, $\mathbf{\Sigma}_{ps}$ is actually separable since $\sigma_l(j,k) = \omega_l\tilde{\sigma}(j,k)$, where $\omega_l = l^{-2}$ and $\tilde{\sigma}(j,k) = r^{|j-k|}$.

b. For the PNS case, we modify each $\sigma_l(j,k)$ in the SEP case as

$$\sigma_l^*(j,k) = \begin{cases} l^{-2}\, r^{|j-k|} & l = 1,3,5,\ldots \\ l^{-2} I(j=k) & l = 2,4,6,\ldots. \end{cases}$$

It can be seen from Proposition 1(b) that such a setup of $\sigma_l^*(j,k)$ yields a covariance

Table 5: Empirical rejection rates(%) of the partial separability test (PS-T) and the separability test (Sep-T) under SEP and PNS cases.

	$r = 0.2$		r=0.3		r=0.4	
L_n	PS-T	Sep-T	PS-T	Sep-T	PS-T	Sep-T
			SEP			
2	5.4	3.8	4.7	5.1	5.5	4.7
3	5.8	3.8	5.1	6.1	5.7	4.8
			PNS			
2	5.8	22.5	5.5	78.6	4.4	98.2
3	5.3	33.2	5.9	92.0	4.2	99.0

structure that is partially separable but not separable. Besides, the deviation from separability becomes larger as r increases. Table 5 displays the empirical rejection rates of the separability test in Aston et al. (2017) and proposed partial separability test for the SEP and PNS cases. We utilize the χ^2 type mixture test for the partial separability test due to its proven effectiveness in typical multivariate functional data, as demonstrated in Section 4.1. It can be observed that both tests control type I errors well under different truncation levels for the SEP case, considering the covariance is also partially separable. On the other hand, for the PNS case, the partial separability test maintains good control of type I error, whereas the separability test exhibits the ability in rejecting the null hypothesis. Besides, the power of the separability test increases significantly with higher values of r. These findings align with our expectations and clearly illustrate

Table 6: The p-values of partial separability tests (PS-T) and separability tests (Sep-T) by Aston et al. (2017) with different L_n for multichannel tonnage data.

L_n	2	5	8	11	14
PS-T	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Sep-T	0.004	0.004	0.008	0.003	0.007

the effectiveness of the separability and partial separability tests.

S4.2 Real data application

We apply the separability test to the real data examples in Section 5. For the multichannel tonnage data, Table 6 presents the test results of the separability test proposed by Aston et al. (2017) as well as the partial separability test based on χ^2 type mixture. The truncation levels L_n are selected in alignment with those in Section 5.1. It can be observed that the separability test also rejects the null hypothesis under different truncation levels. This outcome is in line with expectations as partial separability is a weaker concept than separability.

For the EEG data, Table 7 displays the p-values of the separability tests under different L_n, all of which are below 0.001. In contrast, the partial separability is not rejected by the proposed tests. These findings provide support for justifying the assumption of partial separability for the EEG data instead of separability. Moreover, they demonstrate the flexibility of partial separability in accommodating the cross-covariance structure of the data.

Table 7: The p-values of partial separability tests (PS-T) and separability tests (Sep-T) by Aston et al. (2017) with different L_n for EEG data.

L_n	3	6	9
PS-T	0.135	0.121	0.110
Sep-T	< 0.001	< 0.001	< 0.001

S5 Subsequent analysis of patial separability tests on real data application

We study the eventual effect of the proposed tests on the multichannel change-point problem and the functional graphs for EEG data in Section 5. For the multichannel tonnage data, our analysis shows that the null hypothesis is rejected for the entire dataset comprising all four channels. However, the results also indicate that the assumption of partial separability can be justified for a specific subgroup of channels 3 and 4. To further explore the relationship between partial separability and change-point detection, we evaluate the Wald statistic Q_τ (defined in Section 2.3) from $\tau = 1$ to 200 across different subgroups, as displayed in Figure 2. It can be observed that the peak point for channels $(3, 4)$ occurs at $\hat{\tau} = 151$, while for channels $(1, 2)$ the maximum value is found at $\hat{\tau} = 80$. This finding demonstrates that misusing the partial separability assumption can potentially result in incorrect change-point detection. It is interesting to note that the estimated change point $\hat{\tau} = 151$ for channel $(3, 4)$ aligns with the findings of Paynabar et al. (2016), who assumed partial separability for the entire four-variate process and

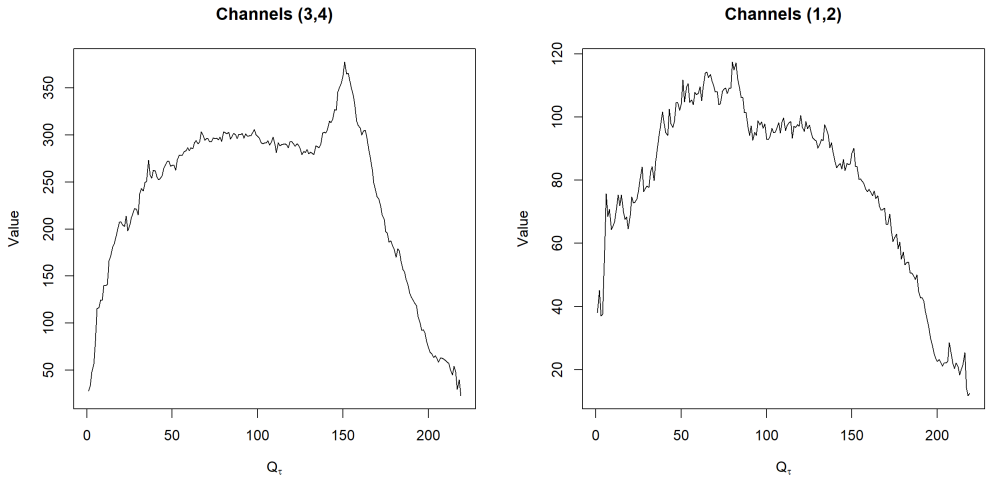

Figure 2: The value of the change-detection statistic Q_τ across different channels.

drew this conclusion for the entire dataset. This suggests a certain level of robustness of the change-point method to the partial separability assumption, provided that the deviation from partial separability in subgroups is not substantial.

For the EEG data in Section 5.2, our test of partial separability does not reject the null hypothesis, which provides justification for assuming partial separability in the subsequent application of the functional graphical model discussed in Section 2.3. It is also noteworthy that the separability assumption is demonstrated to be violated, as stated in Section S4.2.

References

Aston, J. A., D. Pigoli, S. Tavakoli, et al. (2017). Tests for separability in nonparametric covariance operators of random surfaces. *The Annals of Statistics 45*(4), 1431–1461.

Chu, T., J. Zhu, and H. Wang (2019). Semiparametric modeling with nonseparable and non-stationary spatio-temporal covariance functions and its inference. *Statistica Sinica 29*(3), 1233–1252.

Gneiting, T., W. Kleiber, and M. Schlather (2010). Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association 105*(491), 1167–1177.

Hsing, T. and R. Eubank (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators*. John Wiley & Sons.

Li, Y. and Y. Guan (2014). Functional principal component analysis of spatio-temporal point processes with applications in disease surveillance. *Journal of the American Statistical Association 109*(507), 1205–1215.

Liang, D., H. Huang, Y. Guan, and F. Yao (2022). Test of weak separability for spatially stationary functional field. *Journal of the American Statistical Association*, 1–14.

Liu, C., S. Ray, and G. Hooker (2017). Functional principal component analysis of spatially correlated data. *Statistics and Computing 27*(6), 1639–1654.

Paynabar, K., C. Zou, and P. Qiu (2016). A change-point approach for phase-i analysis in

multivariate profile monitoring and diagnosis. *Technometrics 58*(2), 191–204.

Zhang, H. and Y. Li (2021). Unified principal component analysis for sparse and dense functional data under spatial dependency. *Journal of Business & Economic Statistics*, 1–15.

Zhang, S., B. Guo, A. Dong, J. He, Z. Xu, and S. X. Chen (2017). Cautionary tales on air-quality improvement in beijing. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 473*(2205), 20170457.