Graphical Principal Component Analysis of Multivariate Functional Time Series

Jianbin Tan
Joint work with Decai Liang, Yongtao Guan, and Hui Huang.

Department of Biostatistics & Bioinformatics
Duke University

July 14, 2024

Table of Contents

Multivariate Functional Time Series

Consider the random objects

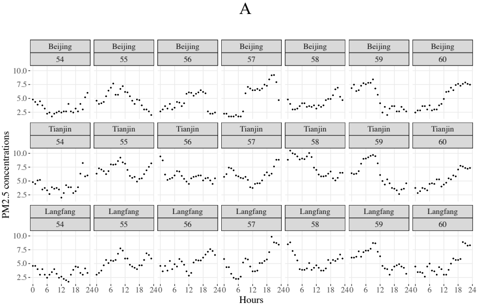
$$\varepsilon_j(t) \in \mathbb{R}^p, \qquad \text{for} \qquad j = 1, \ldots, J \text{ and } t \in \mathcal{T},$$

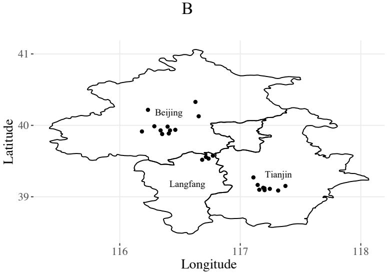
where

- j is the index of the discrete-time unit.
- $\mathcal{T}$ is the domain of the multivariate functions.
- Ignore the functional variable $t \rightarrow$ multivariate time series.
- Consider the functional variable $t \rightarrow$ **multivariate functional times series (MFTS)**.

Data Illustration

- Daily curves, monthly curves, or yearly curves from different subjects.





Two-way Dependency

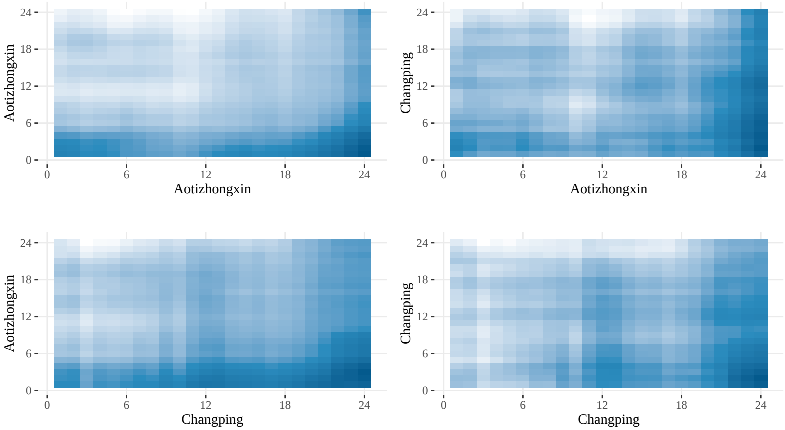


Figure – One-day-lag covariance functions between stations.

Two-way Dependency

The MFTS has both

- **Serial dependencies** over time
- **Multivariate dependence** structure among the different subjects

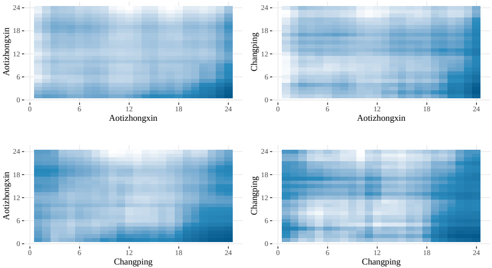


Figure – Two-day-lag covariance functions between stations.

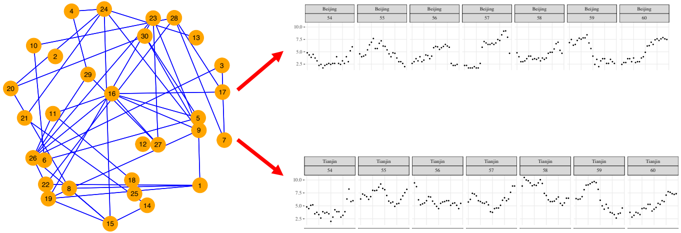
Dimension reduction on MFTS

We want to perform a dimension reduction on MFTS :

- Encode the **serial and multivariate dependencies** among MFTS.
- Transform the possibly **infinite-dimensional MFTS** into a multivariate time series.
- Provide a **data-adaptive model** for MFTS.

MFTS on a Graph

We treat an **univariate functional time series** as a node from a graph.



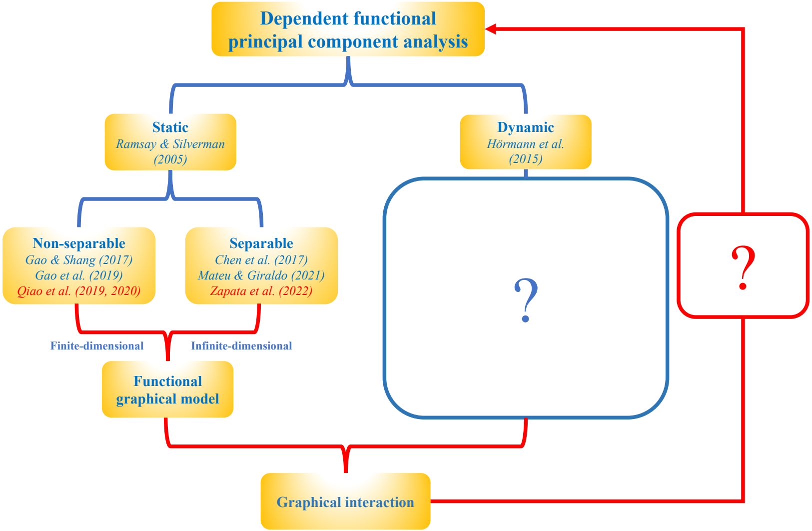
- Provide a powerful and illustrative framework to describe multivariate dependencies.
- Defined for multivariate variables (Friedman et al., 2001) and multivariate time series (Dahlhaus, 2000 ; Dahlhaus and Eichler, 2003).
- **Do not define for MFTS.**

Existing Methods

We focus on the dimension reduction of MFTS on a graph.

- **Conventional static Karhunen-Loève (KL) expansions** for each univariate functional time series (Gao and Shang, 2017 ; Gao et al., 2019).
- Dynamic functional principal component analysis (DFPCA ; Hörmann et al., 2015) ; **More general and efficient than the static KL expansion**.
- Both these methods ignore graphical interactions in the MFTS – a loss of statistical efficiency.

Dependent Functional Principal Component Analysis



Contributions

- Define a graph model for **infinite-dimensional MFTS**.
- Extend the DFPCA to the multivariate case.
  - Embed a graph structure into the DFPCA and give an **optimal representation for MFTS, preserving all information on the graph structure**.
  - The resulting representation is **more general** than the models proposed in the literature (Mateu and Giraldo, 2021 ; Zapata et al., 2022).
- Develop a novel procedure for dimension reduction of contaminated MFTS, considering

  graphical interactions and serial dependencies.

Spectral Density Kernel

Define the **auto-covariance function** of the MFTS $\{\varepsilon_j; j \in \mathbb{Z}\}$ by

$$\mathbb{E}\varepsilon_{j_1}(t)\varepsilon_{j_2}(s)^T = \boldsymbol{C}_g(t,s) \in \mathbb{R}^{p \times p}, \tag{1}$$

$\forall j_1, j_2 \in \mathbb{Z}$ and $t, s \in [0,1]$, where $g = j_1 - j_2$.

- **Two-way dependence** : serial and multivariate dependencies on random functions.
- Define the **spectral density kernel** by the Fourier transformation

$$\boldsymbol{f}(t,s|\theta) := \frac{1}{2\pi}\sum_{g \in \mathbb{Z}} \boldsymbol{C}_g(t,s)\exp(\mathrm{i}\, g\theta),\ \ \theta \in [-\pi,\pi],\ \ t,s \in [0,1]. \tag{2}$$

Weak separability on frequency domain

For each $\forall\theta \in [-\pi, \pi]$, $\boldsymbol{f}(t, s|\theta)$ can be decomposed as

$$\boldsymbol{f}(t, s|\theta) = \sum_{k=1}^{\infty} \omega_k(\theta)\boldsymbol{\delta}_k(s|\theta)\left\{\boldsymbol{\delta}_k(t|\theta)\right\}^*. \tag{3}$$

$\boldsymbol{f}(t, s|\theta)$ is **weakly separable** if and only if

$$\boldsymbol{f}(t, s|\theta) = \sum_{k=1}^{\infty} \boldsymbol{\eta}_k(\theta)\overline{\psi_k(t|\theta)}\psi_k(s|\theta), \tag{4}$$

where the **vector-valued** $\boldsymbol{\delta}_k(t|\theta)$ degenerates to the **scalar-valued** $\psi_k(t|\theta)$. As such, the scalar-valued $\omega_k(\theta)$ becomes as the **eigen-matrix** $\boldsymbol{\eta}_k(\theta)$.

- The above condition is called the **dynamic weak separability.**

Degeneration of separability

- The concept of separability is firstly proposed for **covariance functions**. In our case, we may assume

$$\boldsymbol{C}_g(t,s) = \boldsymbol{C}^{(1)} \cdot C_g^{(2)}(t,s) \text{ or } \boldsymbol{C}_g(t,s) = \boldsymbol{C}_g^{(3)} \cdot C^{(4)}(t,s).$$

- $\boldsymbol{C}_g(t,s) = \boldsymbol{C}_g^{(3)} \cdot C^{(4)}(t,s)$ can be generalized as

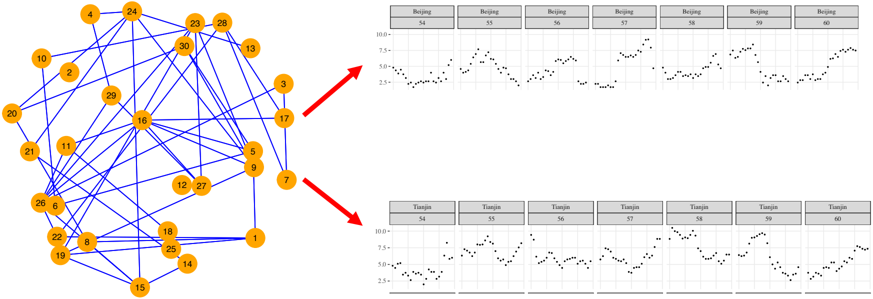
$$\boldsymbol{C}_g(t,s) = \sum_{k=1}^{\infty} \boldsymbol{\vartheta}_{k,g}\ \varphi_k(t)\varphi_k(s). \tag{5}$$

- If $p=1$, (5) is called **weak separability** in Liang et al. (2023), and if $\boldsymbol{C}_g(t,s) = 0$, $\forall t,s \in [0,1]$ and $g \neq 0$, (5) is called **partial separability** in Zapata et al. (2022).
- Our separability condition is **more general** than the above conditions.

Graphical models of MFTS

Define $V = \{1, \cdots, p\}$:

- Given $i_1, i_2 \in V$, define a **partial correlation relation** between $\{\varepsilon_{i_1 j}(\cdot); j \in \mathbb{Z}\}$ and $\{\varepsilon_{i_2 j}(\cdot); j \in \mathbb{Z}\}$ **conditioning on other functional time series**.
- Let $E \subset V^2$ be an edge set, where $(i_1, i_2) \in E$ if and only if $\{\varepsilon_{i_1 j}(\cdot); j \in \mathbb{Z}\}$ and $\{\varepsilon_{i_2 j}(\cdot); j \in \mathbb{Z}\}$ are partially correlated.



Graphical models of MFTS

Theorem

Under the dynamic weak separability (4), define

$$\mathbf{\Phi}_k(\theta) = \{\boldsymbol{\eta}_k(\theta)\}^{-1}$$

and $[\cdot]_{V_1,V_2}$ *is the operation to extract subsets* V_1, V_2 *of rows and columns of a matrix. Then,*

$$(i_1, i_2) \notin E \text{ iff } [\mathbf{\Phi}_k(\theta)]_{i_1,i_2} = 0, \ \forall\theta \in [-\pi, \pi] \text{ and } k \geq 1.$$

- The existence of graphical models for **infinite-dimensional MFTS**.
- **What are the roles of $\eta_k(\theta)$ and $\mathbf{\Phi}_k(\theta)$ for MFTS ?**

Dynamic FPCA

The **dynamic KL expansion** of $\{\varepsilon_{ij}(\cdot); j \in \mathbb{Z}\}$ (Hörmann et al., 2015) is

$$\varepsilon_{ij}(t) = \sum_{k=1}^{\infty} \sum_{l \in \mathbb{Z}} \phi_{ikl}(t) \xi_{i(j+l)k}. \tag{6}$$

- $\{\phi_{ikl}; l \in \mathbb{Z}\}$, is called the **functional filters**, and ξ_{ijk} is the score variable.
- The expansion (6) is more general than the static KL expansion

$$\varepsilon_{ij} = \sum_{k=1}^{\infty} \varphi_{ik} \xi_{ijk}.$$

- Although (6) can be used on each of the univariate component in MFTS, it ignores **graphical interactions**.

Weakly-separable KL expansion

Under the dynamic weak separability (4), we can prove that

$$\varepsilon_j(t) := \sum_{k=1}^{\infty} \sum_{l \in \mathcal{Z}} \phi_{kl}(t) \boldsymbol{\xi}_{\cdot,(j+l)k} \tag{7}$$

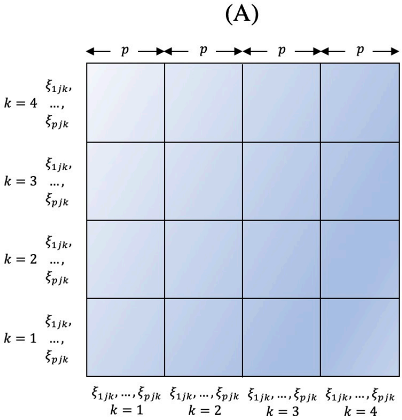
is an **optimal representation** for MFTS, where

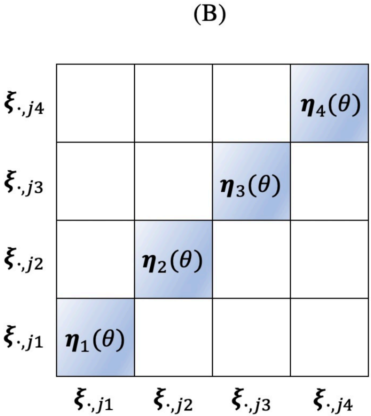
$$\phi_{kl}(t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \psi_k(t|\theta) \exp(-\,\mathrm{i}\, l\theta)\, \mathrm{d}\theta \tag{8}$$

is **free of subject label** i, and

- $\{\boldsymbol{\xi}_{\cdot,jk}; j \in \mathbb{Z}\}$ is a **multivariate time series** with the **spectral density matrix** $\boldsymbol{\eta}_k(\theta)$ and **precision matrix** $\boldsymbol{\Phi}_k(\theta)$.
- $\{\boldsymbol{\xi}_{\cdot,jk}; j \in \mathbb{Z}\}$, $k = 1, 2, \cdots$, are **uncorrelated** for different k.

An Illustration of Dynamic Weak Separability





Union of Graph Information

Define the partial correlation graph of $\{\boldsymbol{\xi}_{\cdot,jk}; j \in \mathbb{Z}\}$ (Dahlhaus, 2000).

Theorem

Under the dynamic weak separability (4), let (V, E) and (V, E_k) be the graphs of $\{\varepsilon_j; j \in \mathbb{Z}\}$ and $\{\boldsymbol{\xi}_{\cdot,jk}; j \in \mathbb{Z}\}$, respectively, then

$$E = \cup_{k=1}^{\infty} E_k. \tag{9}$$

- Scores preserve all information on the graph structure of the original MFTS.
- Define **partial mutual information** (Brillinger, 1996) defined as

$$\mathcal{I}_{i_1,i_2} = -\frac{1}{2\pi} \sum_{k=1}^{K} \int_{-\pi}^{\pi} \log \left\{ 1 - \frac{\left|[\boldsymbol{\Phi}_k(\theta)]_{i_1,i_2}\right|^2}{[\boldsymbol{\Phi}_k(\theta)]_{i_1,i_1}[\boldsymbol{\Phi}_k(\theta)]_{i_2,i_2}} \right\} \mathrm{d}\theta.$$

Graphical FPCA

We refer to the representation

$$\boldsymbol{\varepsilon}_j(t) := \sum_{k=1}^{\infty} \sum_{l \in \mathcal{Z}} \phi_{kl}(t) \boldsymbol{\xi}_{\cdot,(j+l)k}$$

as the **graphical DFPCA (GDFPCA)** for MFTS.

- Characterize graphical dependencies in MFTS naturally.
- More general then the **static form** :

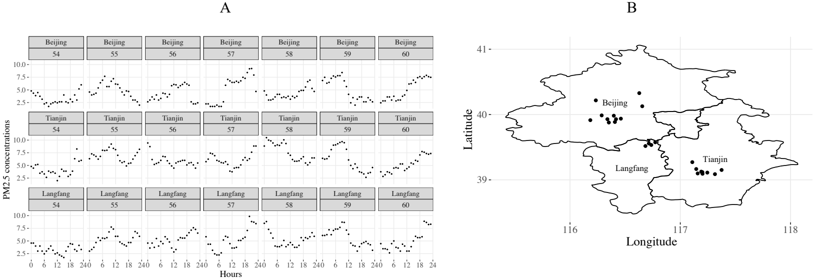
$$\varepsilon_{ij}(t) = \sum_{k \in \mathbb{Z}} \varphi_k(t) \xi_{ijk} \text{ with } \xi_{ijk} = \langle \varepsilon_{ij}, \varphi_k \rangle,$$

  which is used to model spatial functional data (Mateu and Giraldo, 2021 ; Liang et al., 2021) or graphical functional data (Zapata et al., 2022).

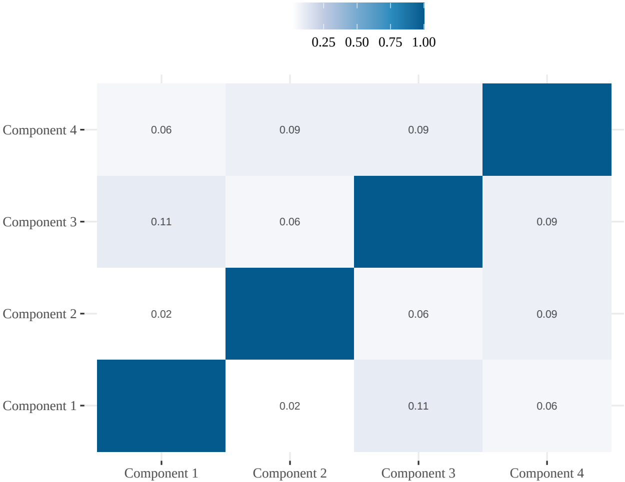
1. Background
2. Theoretical Framework
3. Data Illustration

PM2.5 concentration

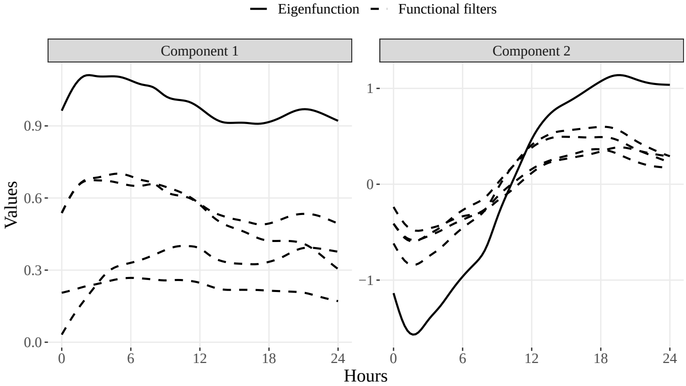
- Consider hourly readings of PM2.5 concentration (measured in $\mu g/m^3$) that were collected from 24 monitoring stations (in three cities in China) in the winter of 2016, with a total length of 60 days.
- We observe a discrete MFTS with $p = 24$.



Assessment of Dynamic Weak Separability

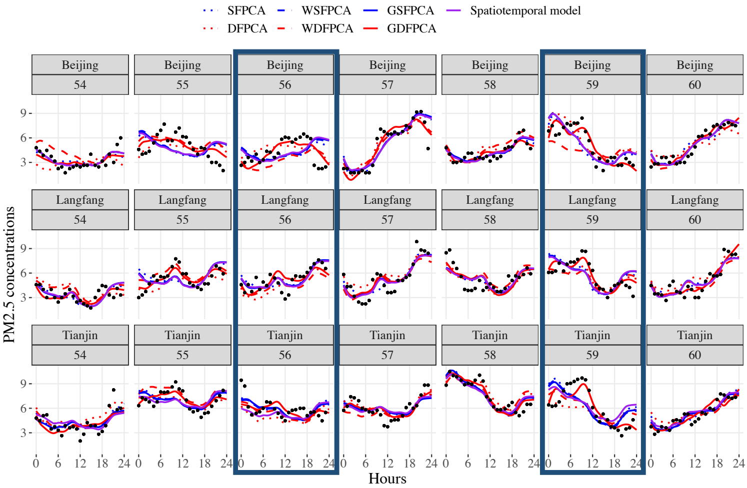


Dynamic vs Static



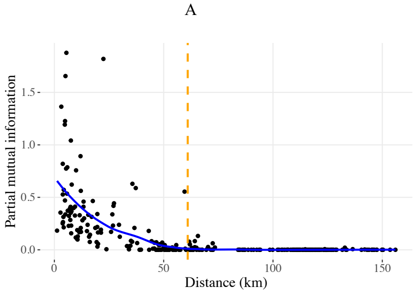
- Static forms of separability (Liang et al., 2021 ; Mateu and Giraldo, 2021 ; Zapata et al., 2022) may not be suitable for the spatial-temporal data.

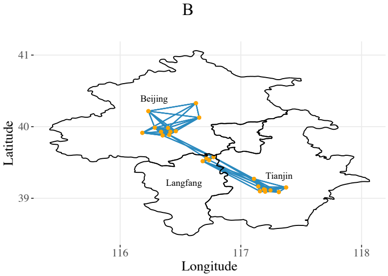
Reconstruction



Partial Mutual Information

Unlike conventional spatiotemporal models, we do not assume the **spatial stationarity condition** for the spatial-temporal data.





The PM2.5 concentration among the stations in Tianjin and Langfang is more partially correlated.

Related Literature

- Tan, J., Liang, D., Guan, Y., and Huang, H. (2024). Graphical principal component analysis of multivariate functional time series. *Journal of the American Statistical Association*, 1-24.

References I

Brillinger, D. R. (1996). Remarks concerning graphical models for time series and point processes. *Brazilian Review of Econometrics*, 16(1) :1–23.

Dahlhaus, R. (2000). Graphical interaction models for multivariate time series 1. *Metrika*, 51(2) :157–172.

Dahlhaus, R. and Eichler, M. (2003). Causality and graphical models in time series analysis. *Oxford Statistical Science Series*, pages 115–137.

Friedman, J., Hastie, T., Tibshirani, R., et al. (2001). *The elements of statistical learning*, volume 1. Springer series in statistics New York.

Gao, Y. and Shang, H. L. (2017). Multivariate functional time series forecasting : Application to age-specific mortality rates. *Risks*, 5(2) :21.

Gao, Y., Shang, H. L., and Yang, Y. (2019). High-dimensional functional time series forecasting : An application to age-specific mortality rates. *Journal of Multivariate Analysis*, 170 :232–243.

Hörmann, S., Kidziński, Ł., and Hallin, M. (2015). Dynamic functional principal components. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 77(2) :319–348.

Liang, D., Huang, H., Guan, Y., and Yao, F. (2023). Test of weak separability for spatially stationary functional field. *Journal of the American Statistical Association*, 118(543) :1606–1619.

Liang, D., Zhang, H., Chang, X., and Huang, H. (2021). Modeling and regionalization of chinas pm2. 5 using spatial-functional mixture models. *Journal of the American Statistical Association*, 116(533) :116–132.

Mateu, J. and Giraldo, R. (2021). *Geostatistical Functional Data Analysis*. John Wiley & Sons.

Zapata, J., Oh, S.-Y., and Petersen, A. (2022). Partial separability and functional graphical models for multivariate gaussian processes. *Biometrika*, 109(3) :665–681.