Regularized Estimation of Spatial Patterns

Wen-Ting Wang

June 18, 2017

Joint work with **Hsin-Cheng Huang** @ Academia Sinica

Outline

Climate Change

Climate Change

increases the odds of **extreme weather events** occuring,

Flood

Drought

Climate Change

affects **human health** and **quality of life**

Drought in the East Africa

2011

10,000

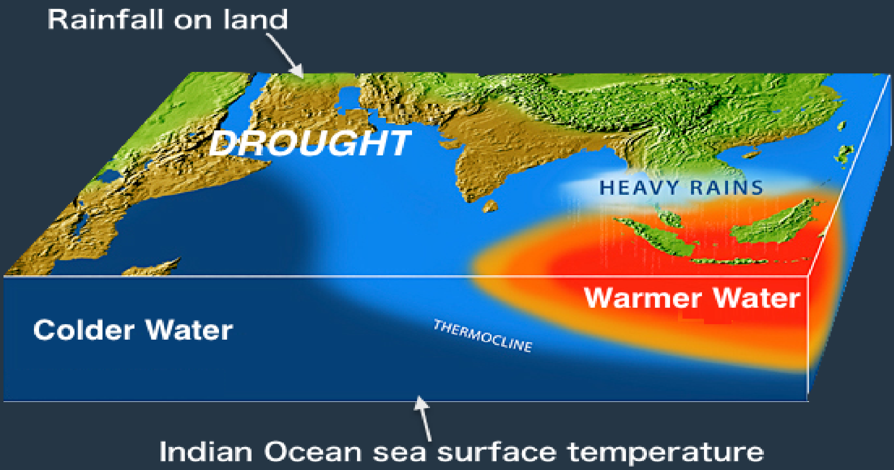
People have been killed
by the worst drought in 60 years.

Climate Change

are associated with atmospheric dynamics.

Atmospheric dynamics

can be studied through spatial patterns.


Rainfall on land
DROUGHT
HEAVY RAINS
Colder Water
Warmer Water
THERMOCLINE
Indian Ocean sea surface temperature

Outline

Background

- Spatial processes of interest:

$$\{\eta_i(\boldsymbol{s}); \boldsymbol{s} \in D\}\,; i = 1, \ldots, n$$

  - $D \subset \mathbb{R}^d$
  - mean zero
  - common covariance function: $C_\eta(\boldsymbol{s}^*, \boldsymbol{s}) = \text{cov}(\eta_i(\boldsymbol{s}^*), \eta_i(\boldsymbol{s}))$
  - $\eta_1(\cdot), \ldots, \eta_n(\cdot)$: uncorrelated

Background

- Spatial processes of interest:

$$\{\eta_i(\boldsymbol{s}); \boldsymbol{s} \in D\}\,; i = 1, \ldots, n$$

  - $D \subset \mathbb{R}^d$
  - mean zero
  - common covariance function: $C_\eta(\boldsymbol{s}^*, \boldsymbol{s}) = \text{cov}(\eta_i(\boldsymbol{s}^*), \eta_i(\boldsymbol{s}))$
  - $\eta_1(\cdot), \ldots, \eta_n(\cdot)$: uncorrelated

- Data at locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_p \in D$,

$$Y_i(\boldsymbol{s}_j) = \eta_i(\boldsymbol{s}_j) + \epsilon_{ij}; i = 1, \ldots, n,\ j = 1, \ldots, p$$

  - $\epsilon_{ij} \sim (0, \sigma^2)$: white noise
  - ϵ_{ij} and $\eta_i(\boldsymbol{s}_j)$ are uncorrelated for any i, j

Targets

1. Detect the dominant spatial patterns (modes) of $\eta_1(\cdot), \ldots, \eta_n(\cdot)$
   - interpret the variability of spatial data physically
2. Estimate spatial covariance function $C_\eta(\cdot, \cdot)$
   - no specific assumption (*e.g.*, parametric form or stationarity)
   - spatial prediction (kriging) of $\{\eta_i(\boldsymbol{s}); \boldsymbol{s} \in D\}$

Rank-K Spatial Model

- Data:

$$Y_i(\boldsymbol{s}_j) = \eta_i(\boldsymbol{s}_j) \qquad\qquad + \epsilon_{ij}; \quad i = 1, \ldots, n, \quad j = 1 \ldots, p$$

Rank-K Spatial Model

- Data:

$$Y_i(\boldsymbol{s}_j) = \sum_{k=1}^{K} \xi_{ik}\varphi_k(\boldsymbol{s}_j) + \epsilon_{ij}; \quad i = 1, \ldots, n, \quad j = 1 \ldots, p$$

  - $(\xi_{i1}, \ldots, \xi_{iK})' \sim (\mathbf{0}, \mathbf{\Lambda})$; $\mathbf{\Lambda}_{K\times K}$ is positive-definite
  - $\varphi_1(\cdot) \ldots, \varphi_K(\cdot)$: K unknown orthonormal functions
  - ξ_{ik} uncorrelated with ϵ_{ij}

Rank-K Spatial Model

- Data:

$$Y_i(\boldsymbol{s}_j) = \sum_{k=1}^{K} \xi_{ik}\varphi_k(\boldsymbol{s}_j) + \epsilon_{ij}; \quad i = 1, \ldots, n, \quad j = 1 \ldots, p$$

  - $(\xi_{i1}, \ldots, \xi_{iK})' \sim (\mathbf{0}, \boldsymbol{\Lambda})$; $\boldsymbol{\Lambda}_{K\times K}$ is positive-definite
  - $\varphi_1(\cdot) \ldots, \varphi_K(\cdot)$: K unknown orthonormal functions
  - ξ_{ik} uncorrelated with ϵ_{ij}

- Spatial covariance function:

$$C_\eta(\boldsymbol{s}^*, \boldsymbol{s}) = \sum_{k=1}^{K}\sum_{k'=1}^{K} \lambda_{kk'}\varphi_k(\boldsymbol{s}^*)\varphi_{k'}(\boldsymbol{s})$$

  - $\lambda_{kk'}$: (k, k') entry of $\boldsymbol{\Lambda}$

Goal

- Find $\varphi_1(\cdot), \ldots, \varphi_K(\cdot)$ to represent the dominant patterns.

Goal

- Find $\varphi_1(\cdot), \ldots, \varphi_K(\cdot)$ to represent the dominant patterns.
- Standard approach: principal component analysis (PCA)

Principal Component Analysis

- p-dimensional data vector:

$$\boldsymbol{Y}_i = (Y_i(\boldsymbol{s}_1), \ldots, Y_i(\boldsymbol{s}_p))' \sim (\boldsymbol{0}, \boldsymbol{\Sigma})$$

- Idea: find $\boldsymbol{\phi} \in \mathbb{R}^p$ with $\boldsymbol{\phi}'\boldsymbol{\phi} = 1$, which maximizes Var$(\boldsymbol{\phi}'\boldsymbol{Y}_i)$

Principal Component Analysis

- p-dimensional data vector:

$$\boldsymbol{Y}_i = (Y_i(\boldsymbol{s}_1), \ldots, Y_i(\boldsymbol{s}_p))' \sim (\boldsymbol{0}, \boldsymbol{\Sigma})$$

- Idea: find $\boldsymbol{\phi} \in \mathbb{R}^p$ with $\boldsymbol{\phi}'\boldsymbol{\phi} = 1$, which maximizes Var$(\boldsymbol{\phi}'\boldsymbol{Y_i})$
- Spectral decomposition: $\boldsymbol{\Sigma}$
  - eigenvalues: $\lambda_1 \geq \cdots \geq \lambda_p$
  - eigenvectors: $\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_p$

Principal Component Analysis

- p-dimensional data vector:

$$\boldsymbol{Y}_i = (Y_i(\boldsymbol{s}_1), \ldots, Y_i(\boldsymbol{s}_p))' \sim (\mathbf{0}, \boldsymbol{\Sigma})$$

- Idea: find $\boldsymbol{\phi} \in \mathbb{R}^p$ with $\boldsymbol{\phi}'\boldsymbol{\phi} = 1$, which maximizes Var$(\boldsymbol{\phi}'\boldsymbol{Y_i})$
- Spectral decomposition: $\boldsymbol{\Sigma}$
  - eigenvalues: $\lambda_1 \geq \cdots \geq \lambda_p$
  - eigenvectors: $\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_p$
- Dominant patterns: $\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_K$ (with $\lambda_1, \ldots, \lambda_K$ large)

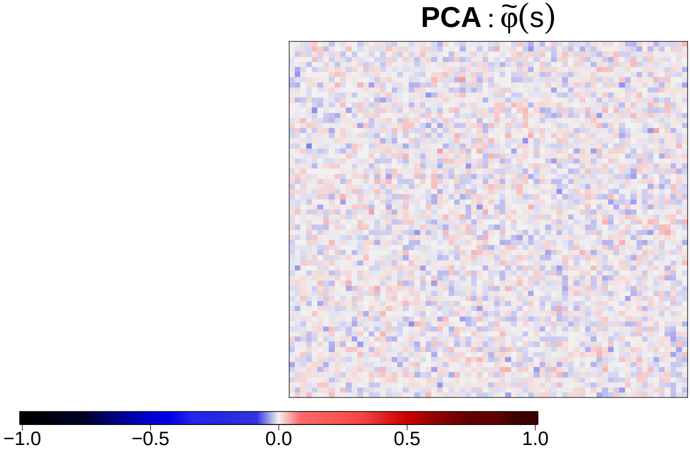
Sample Principal Component Analysis

- Data matrix: $\boldsymbol{Y}_{n\times p} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n)'$
- Sample covariance matrix: $\boldsymbol{S} = \boldsymbol{Y}'\boldsymbol{Y}/n$
- Spectral decomposition: $\boldsymbol{S}$
  - sample eigenvalues: $\tilde{\lambda}_1 \geq \cdots \geq \tilde{\lambda}_p$
  - sample eigenvectors: $\tilde{\boldsymbol{\phi}}_1, \ldots, \tilde{\boldsymbol{\phi}}_p$
- $\tilde{\boldsymbol{\phi}}_1, \ldots, \tilde{\boldsymbol{\phi}}_K$ are estimates of $\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_K$

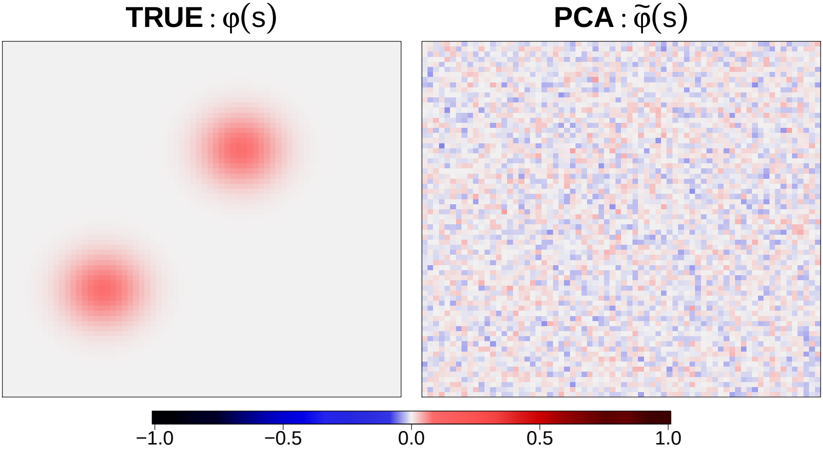
Sample Principal Component Analysis

- Data matrix: $\boldsymbol{Y}_{n\times p} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n)'$
- Sample covariance matrix: $\boldsymbol{S} = \boldsymbol{Y}'\boldsymbol{Y}/n$
- Spectral decomposition: $\boldsymbol{S}$
  - sample eigenvalues: $\tilde{\lambda}_1 \geq \cdots \geq \tilde{\lambda}_p$
  - sample eigenvectors: $\tilde{\boldsymbol{\phi}}_1, \ldots, \tilde{\boldsymbol{\phi}}_p$
- $\tilde{\boldsymbol{\phi}}_1, \ldots, \tilde{\boldsymbol{\phi}}_K$ are estimates of $\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_K$
- Problems:
  - high estimation variability: n is small or p is large
    $\rightarrow$ unstable and noisy patterns
    $\rightarrow$ weak physical interpretation
  - without spatial structure of $\boldsymbol{\phi}$

Example:



Example:



Motivation

To enhance the interpretability, we implement PCA by incorporating

Motivation

To enhance the interpretability, we implement PCA by incorporating

1. spatial structure of $\varphi_k(\cdot)$

Motivation

To enhance the interpretability, we implement PCA by incorporating

1. spatial structure of $\varphi_k(\cdot)$
2. spatial localized pattern of $\varphi_k(\cdot)$

Motivation

To enhance the interpretability, we implement PCA by incorporating

1. spatial structure of $\varphi_k(\cdot)$
2. spatial localized pattern of $\varphi_k(\cdot)$

retain the orthogonal constraint of ϕ_k

Quick Recap

- Data $\boldsymbol{Y}_i = (Y_i(\boldsymbol{s}_1), \ldots, Y_i(\boldsymbol{s}_p))'$; $i = 1, \ldots, n$
  - $Y_i(\boldsymbol{s}_j) = \sum_{k=1}^{K} \xi_{ik}\varphi_k(\boldsymbol{s}_j) + \epsilon_{ij}; \quad j = 1 \ldots, p$
    - $\varphi_1(\cdot) \ldots, \varphi_K(\cdot)$: K unknown orthonormal functions
    - $(\xi_{i1}, \ldots, \xi_{iK})' \sim (\mathbf{0}, \boldsymbol{\Lambda})$; $\boldsymbol{\Lambda}_{K\times K} \succ \mathbf{0}$
    - $\epsilon_{ij} \sim (0, \sigma^2)$; ϵ_{ij}: uncorrelated with ξ_{ik}

Quick Recap

- Data $\boldsymbol{Y}_i = (Y_i(\boldsymbol{s}_1), \ldots, Y_i(\boldsymbol{s}_p))'$; $i = 1, \ldots, n$
  - $Y_i(\boldsymbol{s}_j) = \sum_{k=1}^{K} \xi_{ik} \varphi_k(\boldsymbol{s}_j) + \epsilon_{ij}; \quad j = 1 \ldots, p$
    - $\varphi_1(\cdot) \ldots, \varphi_K(\cdot)$: K unknown orthonormal functions
    - $(\xi_{i1}, \ldots, \xi_{iK})' \sim (\mathbf{0}, \boldsymbol{\Lambda})$; $\boldsymbol{\Lambda}_{K\times K} \succ \mathbf{0}$
    - $\epsilon_{ij} \sim (0, \sigma^2)$; ϵ_{ij}: uncorrelated with ξ_{ik}
- Spatial covariance function:

$$C_\eta(\boldsymbol{s}^*, \boldsymbol{s}) = \sum_{k=1}^{K} \sum_{k'=1}^{K} \lambda_{kk'} \varphi_k(\boldsymbol{s}^*) \varphi_{k'}(\boldsymbol{s})$$

  - $\lambda_{kk'}$: (k, k') entry of $\boldsymbol{\Lambda}$

Quick Recap

- Data $\boldsymbol{Y}_i = (Y_i(\boldsymbol{s}_1), \ldots, Y_i(\boldsymbol{s}_p))'$; $i = 1, \ldots, n$
  - $Y_i(\boldsymbol{s}_j) = \sum_{k=1}^{K} \xi_{ik}\varphi_k(\boldsymbol{s}_j) + \epsilon_{ij}; \quad j = 1 \ldots, p$
    - $\varphi_1(\cdot) \ldots, \varphi_K(\cdot)$: K unknown orthonormal functions
    - $(\xi_{i1}, \ldots, \xi_{iK})' \sim (\boldsymbol{0}, \boldsymbol{\Lambda})$; $\boldsymbol{\Lambda}_{K\times K} \succ \boldsymbol{0}$
    - $\epsilon_{ij} \sim (0, \sigma^2)$; ϵ_{ij}: uncorrelated with ξ_{ik}
- Spatial covariance function:

$$C_\eta(\boldsymbol{s}^*, \boldsymbol{s}) = \sum_{k=1}^{K}\sum_{k'=1}^{K} \lambda_{kk'}\varphi_k(\boldsymbol{s}^*)\varphi_{k'}(\boldsymbol{s})$$

  - $\lambda_{kk'}$: (k, k') entry of $\boldsymbol{\Lambda}$
- Unknown parameters: $\varphi_1(\cdot), \ldots, \varphi_K(\cdot)$, $\boldsymbol{\Lambda}$, σ^2

PCA (alternative version)

- Data matrix: $\boldsymbol{Y}_{n\times p} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n)'$
- PCA :

$$\tilde{\boldsymbol{\Phi}} = \underset{\boldsymbol{\Phi}:\boldsymbol{\Phi}'\boldsymbol{\Phi}=\boldsymbol{I_K}}{\arg\min} \|\boldsymbol{Y} - \boldsymbol{Y}\boldsymbol{\Phi}\boldsymbol{\Phi}'\|_F^2$$

  - $\boldsymbol{\Phi}_{p\times K} = (\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_K)$ with $\phi_{jk} = \varphi_j(\boldsymbol{s}_k)$
  - $\|M\|_F^2 = \sum_{i=1}^{n}\sum_{j=1}^{p} m_{ij}^2$

Regularized PCA

- Data matrix: $\boldsymbol{Y}_{n\times p} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n)'$
- $\boldsymbol{\Phi}_{p\times K} = (\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_K)$ with $\phi_{jk} = \varphi_j(\boldsymbol{s}_k)$
- Objective function

$$\|\boldsymbol{Y} - \boldsymbol{Y}\boldsymbol{\Phi}\boldsymbol{\Phi}'\|_F^2$$

subject to $\boldsymbol{\Phi}'\boldsymbol{\Phi} = \boldsymbol{I}_K$

Regularized PCA

- Data matrix: $\boldsymbol{Y}_{n\times p} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n)'$
- $\boldsymbol{\Phi}_{p\times K} = (\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_K)$
- Objective function

$$\|\boldsymbol{Y} - \boldsymbol{Y}\boldsymbol{\Phi}\boldsymbol{\Phi}'\|_F^2 + \tau_1 \sum_{k=1}^{K} J(\varphi_k) + \tau_2 \sum_{k=1}^{K}\sum_{j=1}^{p} |\varphi_k(\boldsymbol{s}_j)|$$

subject to $\boldsymbol{\Phi}'\boldsymbol{\Phi} = \boldsymbol{I}_K$

- $J(\varphi_k) = \displaystyle\sum_{z_1+\cdots+z_d=2} \int_{R^d} \left(\frac{\partial^2 \varphi_k(\boldsymbol{s})}{\partial x_1^{z_1} \ldots \partial x_d^{z_d}}\right)^2 d\boldsymbol{s}$
  - $\boldsymbol{s} = (x_1, \ldots, x_d)'$
- τ_1: smoothness parameter
- τ_2: sparseness parameter

Regularized PCA

- Data matrix: $\boldsymbol{Y}_{n\times p} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n)'$
- $\boldsymbol{\Phi}_{p\times K} = (\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_K)$
- Objective function

$$\|\boldsymbol{Y} - \boldsymbol{Y}\boldsymbol{\Phi}\boldsymbol{\Phi}'\|_F^2 + \tau_1 \sum_{k=1}^{K} J(\varphi_k) + \tau_2 \sum_{k=1}^{K}\sum_{j=1}^{p} |\varphi_k(\boldsymbol{s}_j)|$$

subject to $\boldsymbol{\Phi}'\boldsymbol{\Phi} = \boldsymbol{I}_K$

- $J(\varphi_k) = \displaystyle\sum_{z_1+\cdots+z_d=2} \int_{R^d} \left(\frac{\partial^2 \varphi_k(\boldsymbol{s})}{\partial x_1^{z_1} \ldots \partial x_d^{z_d}} \right)^2 d\boldsymbol{s}$
  - $\boldsymbol{s} = (x_1, \ldots, x_d)'$
- τ_1: smoothness parameter
- τ_2: sparseness parameter

Spatial PCA (SpatPCA)

- $J(\varphi_k) = \boldsymbol{\phi}_k' \boldsymbol{\Omega} \boldsymbol{\phi}_k$
  - $\boldsymbol{\Omega}_{p \times p}$: determined only by $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_p$
  - Ref: Green and Silverman (1994)

Spatial PCA (SpatPCA)

- $J(\varphi_k) = \boldsymbol{\phi}_k' \boldsymbol{\Omega} \boldsymbol{\phi}_k$
  - $\boldsymbol{\Omega}_{p\times p}$: determined only by $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_p$
  - Ref: Green and Silverman (1994)
- Proposal: SpatPCA

$$\hat{\boldsymbol{\Phi}} = \underset{\boldsymbol{\Phi}:\boldsymbol{\Phi}'\boldsymbol{\Phi}=\boldsymbol{I}_K}{\arg\min} \left\{ \|\boldsymbol{Y} - \boldsymbol{Y}\boldsymbol{\Phi}\boldsymbol{\Phi}'\|_F^2 + \tau_1 \sum_{k=1}^{K} \boldsymbol{\phi}_k' \boldsymbol{\Omega} \boldsymbol{\phi}_k + \tau_2 \sum_{k=1}^{K} \sum_{j=1}^{p} |\phi_{jk}| \right\}$$

Spatial PCA (SpatPCA)

- $J(\varphi_k) = \boldsymbol{\phi}_k' \boldsymbol{\Omega} \boldsymbol{\phi}_k$
  - $\boldsymbol{\Omega}_{p \times p}$: determined only by $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_p$
  - Ref: Green and Silverman (1994)

- Proposal: SpatPCA

$$\hat{\boldsymbol{\Phi}} = \operatorname*{arg\,min}_{\boldsymbol{\Phi}: \boldsymbol{\Phi}'\boldsymbol{\Phi} = \boldsymbol{I}_K} \left\{ \|\boldsymbol{Y} - \boldsymbol{Y}\boldsymbol{\Phi}\boldsymbol{\Phi}'\|_F^2 + \tau_1 \sum_{k=1}^{K} \boldsymbol{\phi}_k' \boldsymbol{\Omega} \boldsymbol{\phi}_k + \tau_2 \sum_{k=1}^{K} \sum_{j=1}^{p} |\phi_{jk}| \right\}$$

- As $\tau_1 = \tau_2 = 0$, $\hat{\boldsymbol{\phi}}_k$ is the k-th eigenvector of $\boldsymbol{S}$.

SpatPCA: $\hat{\varphi}_1(\cdot), \ldots, \hat{\varphi}_K(\cdot)$

- $(\hat{\varphi}_1(\cdot), \ldots, \hat{\varphi}_K(\cdot))$ minimizes

$$\|\boldsymbol{Y} - \boldsymbol{Y}\boldsymbol{\Phi}\boldsymbol{\Phi}'\|_F^2 + \tau_1 \sum_{k=1}^{K} J(\varphi_k) + \tau_2 \sum_{k=1}^{K}\sum_{j=1}^{p} |\varphi_k(\boldsymbol{s}_j)|,$$

subject to $\boldsymbol{\Phi}'\boldsymbol{\Phi} = \boldsymbol{I}_K$

SpatPCA: $\hat{\varphi}_1(\cdot), \ldots, \hat{\varphi}_K(\cdot)$

- $(\hat{\varphi}_1(\cdot), \ldots, \hat{\varphi}_K(\cdot))$ minimizes

$$\|\boldsymbol{Y} - \boldsymbol{Y}\boldsymbol{\Phi}\boldsymbol{\Phi}'\|_F^2 + \tau_1 \sum_{k=1}^{K} J(\varphi_k) + \tau_2 \sum_{k=1}^{K}\sum_{j=1}^{p} |\varphi_k(\boldsymbol{s}_j)|,$$

  subject to $\boldsymbol{\Phi}'\boldsymbol{\Phi} = \boldsymbol{I}_K$

- $\hat{\varphi}_k(\cdot)$: smoothing spline based on $\hat{\boldsymbol{\phi}}_k$

$$\hat{\varphi}_k(\boldsymbol{s}) = \sum_{i=1}^{p} a_i g(\|\boldsymbol{s} - \boldsymbol{s}_i\|) + b_0 + \sum_{j=1}^{d} b_j x_j$$

SpatPCA: $\hat{\varphi}_1(\cdot), \ldots, \hat{\varphi}_K(\cdot)$

- $(\hat{\varphi}_1(\cdot), \ldots, \hat{\varphi}_K(\cdot))$ minimizes

$$\|\boldsymbol{Y} - \boldsymbol{Y}\boldsymbol{\Phi}\boldsymbol{\Phi}'\|_F^2 + \tau_1 \sum_{k=1}^{K} J(\varphi_k) + \tau_2 \sum_{k=1}^{K} \sum_{j=1}^{p} |\varphi_k(\boldsymbol{s}_j)|,$$

  subject to $\boldsymbol{\Phi}'\boldsymbol{\Phi} = \boldsymbol{I}_K$

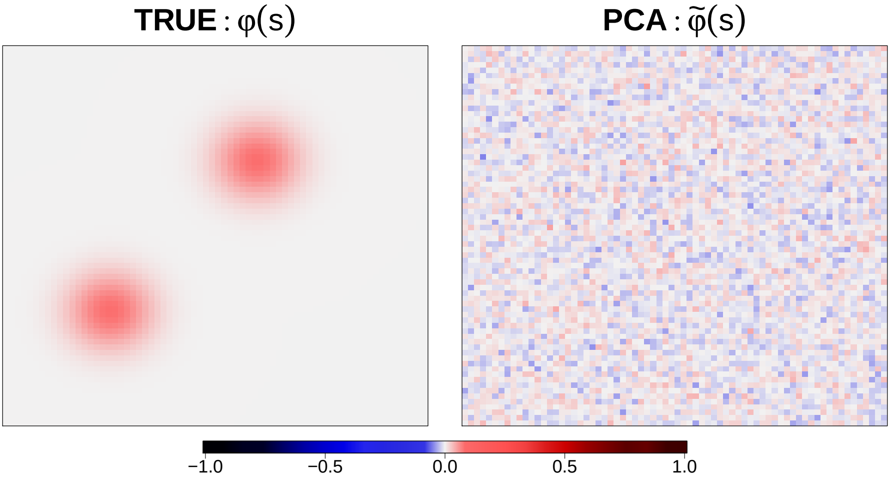
- $\hat{\varphi}_k(\cdot)$: smoothing spline based on $\hat{\boldsymbol{\phi}}_k$

$$\hat{\varphi}_k(\boldsymbol{s}) = \sum_{i=1}^{p} a_i g(\|\boldsymbol{s} - \boldsymbol{s}_i\|) + b_0 + \sum_{j=1}^{d} b_j x_j$$

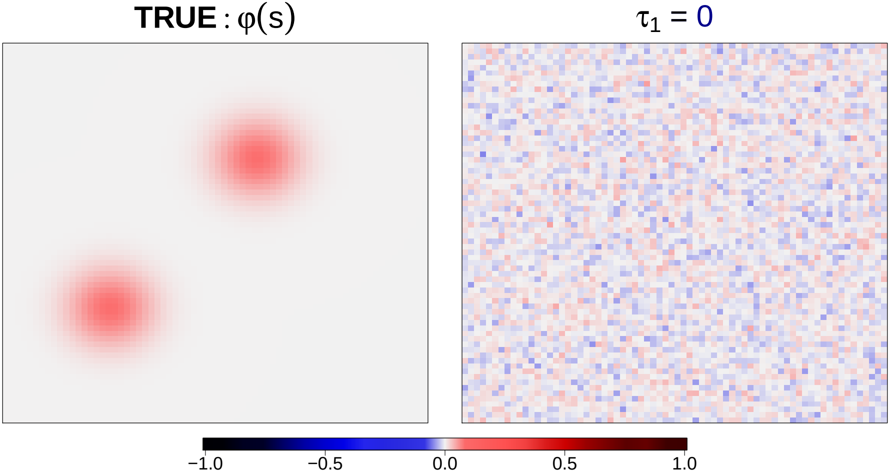
  - $\boldsymbol{s} = (x_1, \ldots, x_d)'$
  - $g(r) = \begin{cases} \frac{1}{16\pi} r^2 \log r; & \text{if } d = 2, \\ \frac{\Gamma(d/2-2)}{16\pi^{d/2}} r^{4-d}; & \text{if } d = 1, 3, \end{cases}$
  - $\boldsymbol{a} = (a_1, \ldots, a_p)'$ and $\boldsymbol{b} = (b_0, b_1, \ldots, b_d)'$ depend on $\hat{\boldsymbol{\phi}}_k$

Why considering two penalties?

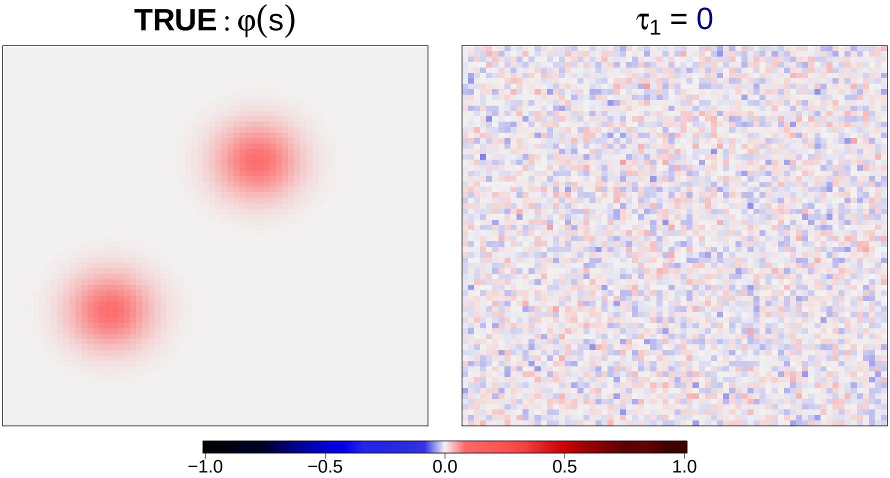
2D Example



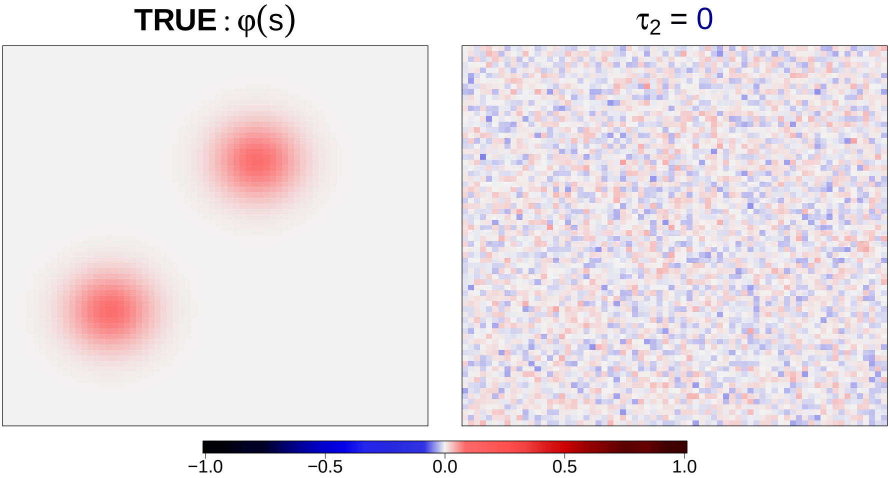
Case 1: $\hat{\varphi}(\cdot)$ as $\tau_2 = 0$ (only smoothness)



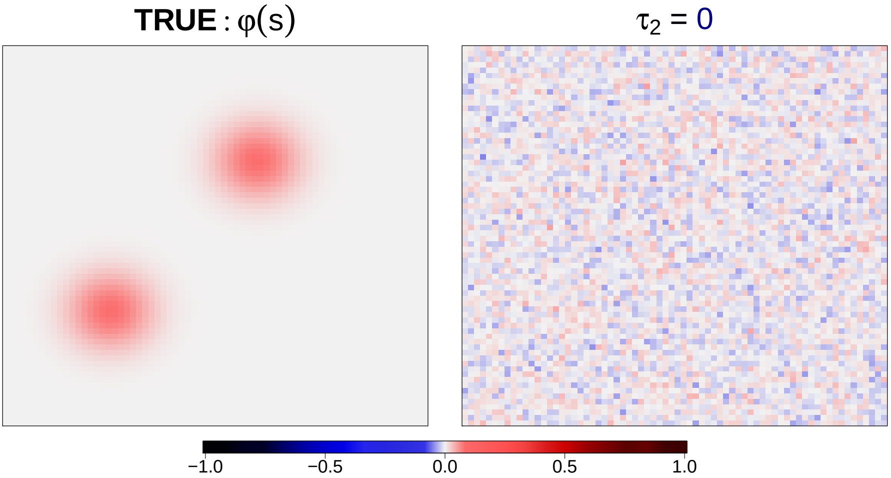
Case 1: $\hat{\varphi}(\cdot)$ as $\tau_2 = 0$ (only smoothness)



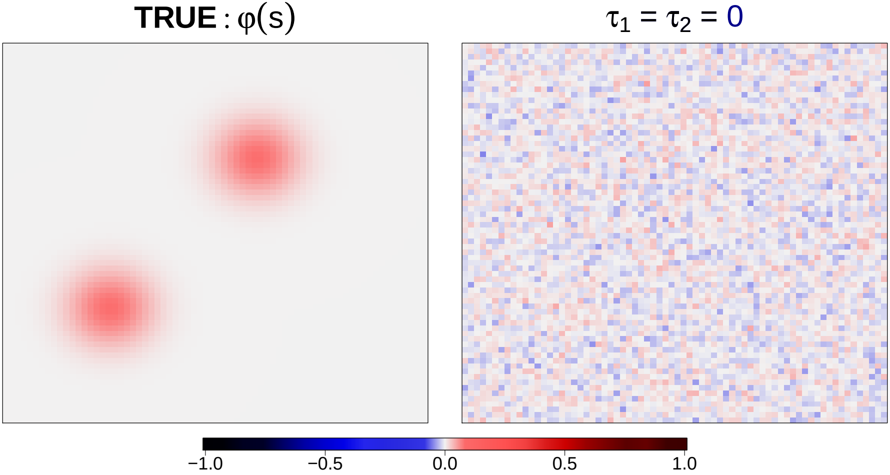
Case 2: $\hat{\varphi}(\cdot)$ as $\tau_1 = 0$ (only sparseness)



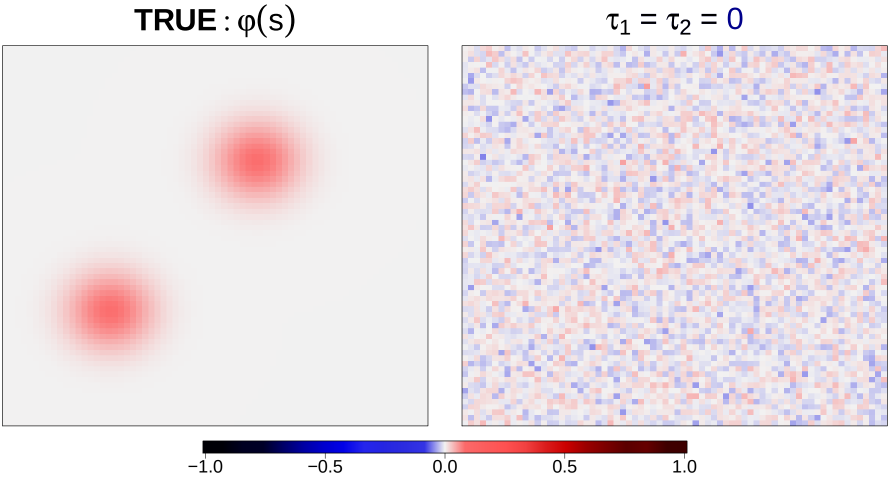
Case 2: $\hat{\varphi}(\cdot)$ as $\tau_1 = 0$ (only sparseness)



Case 3: $\hat{\varphi}(\cdot)$ as $\tau_1 = \tau_2$



Case 3: $\hat{\varphi}(\cdot)$ as $\tau_1 = \tau_2$



Selection of (τ_1, τ_2)

- M-fold cross-validation:

$$\text{CV}(\tau_1, \tau_2) = \frac{1}{M} \sum_{m=1}^{M} \|\boldsymbol{Y}^{(m)} - \boldsymbol{Y}^{(m)} \hat{\boldsymbol{\Phi}}_{\tau_1,\tau_2}^{(-m)} (\hat{\boldsymbol{\Phi}}_{\tau_1,\tau_2}^{(-m)})' \|_F^2$$

Selection of (τ_1, τ_2)

- M-fold cross-validation:

$$\text{CV}(\tau_1, \tau_2) = \frac{1}{M} \sum_{m=1}^{M} \| \boldsymbol{Y}^{(m)} - \boldsymbol{Y}^{(m)} \hat{\boldsymbol{\Phi}}_{\tau_1,\tau_2}^{(-m)} (\hat{\boldsymbol{\Phi}}_{\tau_1,\tau_2}^{(-m)})' \|_F^2$$

  - Partition $\{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n\}$ into M parts with equal (or roughly) size
  - $\boldsymbol{Y}^{(m)}$: the sub-matrix of $\boldsymbol{Y}$ corresponding to the m-th part
  - $\hat{\boldsymbol{\Phi}}_{\tau_1,\tau_2}^{(-m)}$: the estimate of $\boldsymbol{\Phi}$ for (τ_1, τ_2) based on $\boldsymbol{Y}^{(-m)}$
    - $\boldsymbol{Y}^{(-m)}$: remaining data, i.e., $\boldsymbol{Y}$ excluding $\boldsymbol{Y}^{(m)}$

Selection of (τ_1, τ_2)

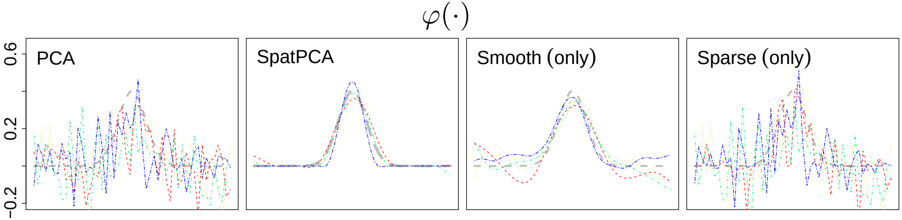
- M-fold cross-validation:

$$\text{CV}(\tau_1, \tau_2) = \frac{1}{M} \sum_{m=1}^{M} \|\boldsymbol{Y}^{(m)} - \boldsymbol{Y}^{(m)} \hat{\boldsymbol{\Phi}}_{\tau_1,\tau_2}^{(-m)} (\hat{\boldsymbol{\Phi}}_{\tau_1,\tau_2}^{(-m)})' \|_F^2$$

  - Partition $\{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n\}$ into M parts with equal (or roughly) size
  - $\boldsymbol{Y}^{(m)}$: the sub-matrix of $\boldsymbol{Y}$ corresponding to the m-th part
  - $\hat{\boldsymbol{\Phi}}_{\tau_1,\tau_2}^{(-m)}$: the estimate of $\boldsymbol{\Phi}$ for (τ_1, τ_2) based on $\boldsymbol{Y}^{(-m)}$
    - $\boldsymbol{Y}^{(-m)}$: remaining data, i.e., $\boldsymbol{Y}$ excluding $\boldsymbol{Y}^{(m)}$

- Find τ_1 and τ_2 which minimize $\text{CV}(\tau_1, \tau_2)$

Example (1D): Estimation of $\varphi(\cdot)$



- τ_1, τ_2 selected by 5-fold CV

Reconstruction of $\eta(\cdot)$

- What if we want to reconstruct the spatial process $\eta(\cdot)$
- Observe only at p locations

Reconstruction of $\eta(\cdot)$

- What if we want to reconstruct the spatial process $\eta(\cdot)$
- Observe only at p locations
- Spatial best linear unbiased prediction of $\eta_i(\boldsymbol{s}_0)$:

$$\hat{\eta}_i(\boldsymbol{s}_0) = \boldsymbol{c}'(\boldsymbol{s}_0)\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}_i$$

  - $\boldsymbol{c}(\boldsymbol{s}_0) = (C_\eta(\boldsymbol{s}_0, \boldsymbol{s}_1), \ldots, C_\eta(\boldsymbol{s}_0, \boldsymbol{s}_p))'$
  - $\boldsymbol{\Sigma} = \text{var}(\boldsymbol{Y}_i) = \boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}' + \sigma^2\boldsymbol{I}$
  - $C_\eta(\boldsymbol{s}^*, \boldsymbol{s}) = \sum_{k=1}^{K}\sum_{k'=1}^{K} \lambda_{kk'}\varphi_k(\boldsymbol{s}^*)\varphi_{k'}(\boldsymbol{s})$
  - Ref: Cressie and Johannesson (2008)

Reconstruction of $\eta(\cdot)$

- What if we want to reconstruct the spatial process $\eta(\cdot)$
- Observe only at p locations
- Spatial best linear unbiased prediction of $\eta_i(\boldsymbol{s}_0)$:

$$\hat{\eta}_i(\boldsymbol{s}_0) = \boldsymbol{c}'(\boldsymbol{s}_0)\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}_i$$

  - $\boldsymbol{c}(\boldsymbol{s}_0) = (C_\eta(\boldsymbol{s}_0, \boldsymbol{s}_1), \ldots, C_\eta(\boldsymbol{s}_0, \boldsymbol{s}_p))'$
  - $\boldsymbol{\Sigma} = \text{var}(\boldsymbol{Y}_i) = \boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}' + \sigma^2\boldsymbol{I}$
  - $C_\eta(\boldsymbol{s}^*, \boldsymbol{s}) = \sum_{k=1}^{K}\sum_{k'=1}^{K} \lambda_{kk'}\varphi_k(\boldsymbol{s}^*)\varphi_{k'}(\boldsymbol{s})$
  - Ref: Cressie and Johannesson (2008)
- Till now, σ^2 and $\boldsymbol{\Lambda}$ are unknown

Estimation of σ^2 and Λ

- $\mathbf{\Lambda}$ has $K(K+1)/2$ unknown elements

Estimation of σ^2 and Λ

- $\mathbf{\Lambda}$ has $K(K+1)/2$ unknown elements
- Apply the regularized method (Tzeng and Huang, 2015):

$$(\hat{\sigma}^2, \hat{\mathbf{\Lambda}}) = \operatorname*{arg\,min}_{(\sigma^2, \mathbf{\Lambda}): \sigma^2 \geq 0,\, \mathbf{\Lambda} \succeq \mathbf{0}} \left\{ \frac{1}{2} \| \boldsymbol{S} - (\hat{\mathbf{\Phi}} \mathbf{\Lambda} \hat{\mathbf{\Phi}}' + \sigma^2 \boldsymbol{I}) \|_F^2 + \gamma \| \hat{\mathbf{\Phi}} \mathbf{\Lambda} \hat{\mathbf{\Phi}}' \|_* \right\}$$

  - 1st term : goodness of fit based on $\mathrm{var}(\boldsymbol{Y}_i) = \mathbf{\Phi} \mathbf{\Lambda} \mathbf{\Phi}' + \sigma^2 \boldsymbol{I}$
  - $\hat{\mathbf{\Phi}}$: given SpatPCA estimate
  - $\gamma \geq 0$ (selected by M-fold CV)
  - $\|\boldsymbol{M}\|_* = \mathrm{tr}((\boldsymbol{M}'\boldsymbol{M})^{1/2})$

Estimation of σ^2 and $\boldsymbol{\Lambda}$

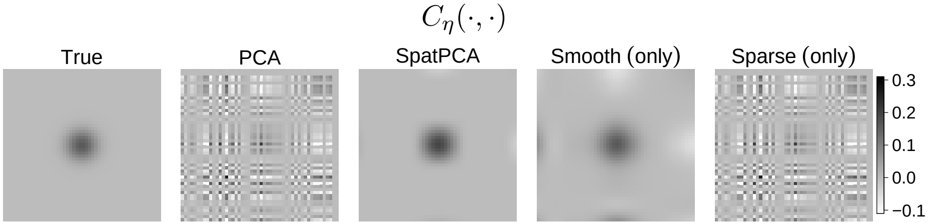
- $\boldsymbol{\Lambda}$ has $K(K+1)/2$ unknown elements
- Apply the regularized method (Tzeng and Huang, 2015):

$$(\hat{\sigma}^2, \hat{\boldsymbol{\Lambda}}) = \underset{(\sigma^2, \boldsymbol{\Lambda}): \sigma^2 \geq 0,\, \boldsymbol{\Lambda} \succeq \mathbf{0}}{\arg\min} \left\{ \frac{1}{2} \left\| \boldsymbol{S} - (\hat{\boldsymbol{\Phi}} \boldsymbol{\Lambda} \hat{\boldsymbol{\Phi}}' + \sigma^2 \boldsymbol{I}) \right\|_F^2 + \gamma \| \hat{\boldsymbol{\Phi}} \boldsymbol{\Lambda} \hat{\boldsymbol{\Phi}}' \|_* \right\}$$

  - 1st term : goodness of fit based on $\text{var}(\boldsymbol{Y}_i) = \boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}' + \sigma^2 \boldsymbol{I}$
  - $\hat{\boldsymbol{\Phi}}$: given SpatPCA estimate
  - $\gamma \geq 0$ (selected by M-fold CV)
  - $\|\boldsymbol{M}\|_* = \text{tr}((\boldsymbol{M}'\boldsymbol{M})^{1/2})$

- Spatial prediction of $\eta_i(\boldsymbol{s}_0)$: $\hat{\eta}_i(\boldsymbol{s}_0) = \hat{\boldsymbol{c}}'(\boldsymbol{s}_0)(\hat{\boldsymbol{\Phi}} \hat{\boldsymbol{\Lambda}} \hat{\boldsymbol{\Phi}}' + \hat{\sigma}^2 \boldsymbol{I})^{-1} \boldsymbol{Y}_i$
  - $\hat{\boldsymbol{c}}(\boldsymbol{s}_0) = (\hat{C}_\eta(\boldsymbol{s}_0, \boldsymbol{s}_1), \ldots, \hat{C}_\eta(\boldsymbol{s}_0, \boldsymbol{s}_p))'$
  - $\hat{C}_\eta(\boldsymbol{s}^*, \boldsymbol{s}) = \sum_{k=1}^K \sum_{k'=1}^K \hat{\lambda}_{kk'} \hat{\varphi}_k(\boldsymbol{s}^*) \hat{\varphi}_{k'}(\boldsymbol{s})$

Example (1D): Covariance Function Estimation



- γ selected by 5-fold CV

Outline

1. **Principal Component Analysis**
   - Background
   - Proposed Method: Spatial PCA
2. **Maximum Covariance Analysis**
   - Background
   - Proposed Method: Spatial MCA
3. **Numerical Example**
4. **Summary**

How rainfalls in East Africa are affected by sea surface temperature in the Indian Ocean?

How rainfalls in East Africa are affected by sea surface temperature in the Indian Ocean?

- Analyze this problem via their coupled spatial patterns
- Ref: Omondi et al.,2013

Background

- Bivariate spatial processes:

$$\{(\eta_{1i}(\boldsymbol{s}_1), \eta_{2i}(\boldsymbol{s}_2)) : \boldsymbol{s}_1 \in D_1, \boldsymbol{s}_2 \in D_2\}; \quad i = 1, \dots, n$$

  - $D_1, D_2 \subset \mathbb{R}^d$
  - $\eta_{11}(\boldsymbol{s}_1), \dots, \eta_{1n}(\boldsymbol{s}_1)$: uncorrelated and mean zero
  - $\eta_{21}(\boldsymbol{s}_2), \dots, \eta_{2n}(\boldsymbol{s}_2)$: uncorrelated and mean zero
  - common spatial covariance function:
    - $C_{11}(\boldsymbol{s}_1, \boldsymbol{s}_1^*) = \text{cov}(\eta_{1i}(\boldsymbol{s}_1), \eta_{1i}(\boldsymbol{s}_1^*))$
    - $C_{22}(\boldsymbol{s}_2, \boldsymbol{s}_2^*) = \text{cov}(\eta_{2i}(\boldsymbol{s}_2), \eta_{2i}(\boldsymbol{s}_2^*))$
    - $C_{12}(\boldsymbol{s}_1, \boldsymbol{s}_2) = \text{cov}(\eta_{1i}(\boldsymbol{s}_1), \eta_{2i}(\boldsymbol{s}_2))$

Background

- Data at locations $\boldsymbol{s}_{11}, \ldots, \boldsymbol{s}_{1p_1} \in D_1$ and $\boldsymbol{s}_{21}, \ldots, \boldsymbol{s}_{2p_2} \in D_2$
  - $Y_{1i}(\boldsymbol{s}_{1j}) = \eta_{1i}(\boldsymbol{s}_{1j}) + \epsilon_{1ij}$; $j = 1, \ldots, p_1$
    - $\epsilon_{1ij} \sim (0, \sigma_1^2)$
    - ϵ_{1ij}: uncorrelated with $\eta_1(\cdot)$
  - $Y_{2i}(\boldsymbol{s}_{2j}) = \eta_{2i}(\boldsymbol{s}_{2j}) + \epsilon_{2ij}$; $j = 1, \ldots, p_2$
    - $\epsilon_{2ij} \sim (0, \sigma_1^2)$
    - ϵ_{2ij}: uncorrelated with $\eta_2(\cdot)$

  - $i = 1, \ldots, n$

Targets

- Find dominant coupled patterns between $\eta_{1i}(\cdot)$ and $\eta_{2i}(\cdot)$
  - to study how variations of $\eta_{1i}(\cdot)$ affect $\eta_{2i}(\cdot)$

Rank-K Cross-Covariance Model

- $D_1, D_2 \subset \mathbb{R}^d$: continuous domain
- Azaïez and Belgacem (2015),

$$C_{12}(\boldsymbol{s}_1, \boldsymbol{s}_2) = \text{cov}(\eta_{1i}(\boldsymbol{s}_1), \eta_{2i}(\boldsymbol{s}_2)) = \sum_{k=1}^{\infty} d_k u_k(\boldsymbol{s}_1) v_k(\boldsymbol{s}_2)$$

  - nonnegative singular values: $d_1 \geq d_2 \geq \cdots$
  - $\{u_k(\cdot)\}$ and $\{v_k(\cdot)\}$: sets of orthonormal basis functions
  - similar to the Karhunen-Loéve expansion

Rank-K Cross-Covariance Model

- $D_1, D_2 \subset \mathbb{R}^d$: continuous domain
- Azaïez and Belgacem (2015),

$$C_{12}(\boldsymbol{s}_1, \boldsymbol{s}_2) = \text{cov}(\eta_{1i}(\boldsymbol{s}_1), \eta_{2i}(\boldsymbol{s}_2)) = \sum_{k=1}^{\infty} d_k u_k(\boldsymbol{s}_1) v_k(\boldsymbol{s}_2)$$

  - nonnegative singular values: $d_1 \geq d_2 \geq \cdots$
  - $\{u_k(\cdot)\}$ and $\{v_k(\cdot)\}$: sets of orthonormal basis functions
  - similar to the Karhunen-Loéve expansion

- Assume $d_{K+1} = 0$.
- $(u_1(\cdot), v_1(\cdot)), \ldots, (u_K(\cdot), v_K(\cdot))$: K dominant coupled patterns

Goal

- Find $(u_1(\cdot), v_1(\cdot)), \ldots, (u_K(\cdot), v_K(\cdot))$ as K coupled patterns

Goal

- Find $(u_1(\cdot), v_1(\cdot)), \ldots, (u_K(\cdot), v_K(\cdot))$ as K coupled patterns
- Common approach: maximum covariance analysis (MCA)

Bivariate Data Vector

- $\begin{pmatrix} \boldsymbol{Y}_{1i} \\ \boldsymbol{Y}_{2i} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\eta}_{1i} \\ \boldsymbol{\eta}_{2i} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\epsilon}_{1i} \\ \boldsymbol{\epsilon}_{2i} \end{pmatrix}; i = 1, \ldots, n$
  - $\begin{pmatrix} \boldsymbol{Y}_{1i} \\ \boldsymbol{Y}_{2i} \end{pmatrix} = \begin{pmatrix} (Y_{1i}(\boldsymbol{s}_{11}), \ldots, Y_{1i}(\boldsymbol{s}_{1p_1}))' \\ (Y_{2i}(\boldsymbol{s}_{21}), \ldots, Y_{2i}(\boldsymbol{s}_{2p_2}))' \end{pmatrix}$
  - $\begin{pmatrix} \boldsymbol{\eta}_{1i} \\ \boldsymbol{\eta}_{2i} \end{pmatrix} = \begin{pmatrix} (\eta_{1i}(\boldsymbol{s}_{11}), \ldots, \eta_{1i}(\boldsymbol{s}_{1p_1}))' \\ (\eta_{2i}(\boldsymbol{s}_{21}), \ldots, \eta_{2i}(\boldsymbol{s}_{2p_2}))' \end{pmatrix}$
  - $\begin{pmatrix} \boldsymbol{\epsilon}_{1i} \\ \boldsymbol{\epsilon}_{2i} \end{pmatrix} = \begin{pmatrix} (\epsilon_{1i1}, \ldots, \epsilon_{1ip_1})' \\ (\epsilon_{2i1}, \ldots, \epsilon_{2ip_2})' \end{pmatrix}$
- Assume $p_1 \geq p_2$

Maximum Covariance Analysis (MCA)

- Bivariate data vector

$$\begin{pmatrix} \boldsymbol{Y}_{1i} \\ \boldsymbol{Y}_{2i} \end{pmatrix} \overset{i.i.d.}{\sim} \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right)$$

  - $\boldsymbol{\Sigma}_{12} = \text{cov}(\boldsymbol{Y}_{1i}, \boldsymbol{Y}_{2i}) = \text{cov}(\boldsymbol{\eta}_{1i}, \boldsymbol{\eta}_{2i})$

- Idea: find $\boldsymbol{u} \in \mathcal{R}^{p_1}$ and $\boldsymbol{v} \in \mathcal{R}^{p_2}$, with $\|\boldsymbol{u}\|_2 = \|\boldsymbol{v}\|_2 = 1$, which maximize

$$d = \text{cov}(\boldsymbol{u}'\boldsymbol{Y}_{1i}, \boldsymbol{v}'\boldsymbol{Y}_{2i}) = \boldsymbol{u}'\boldsymbol{\Sigma}_{12}\boldsymbol{v}$$

Maximum Covariance Analysis (MCA)

- Bivariate data vector

$$\begin{pmatrix} \boldsymbol{Y}_{1i} \\ \boldsymbol{Y}_{2i} \end{pmatrix} \overset{i.i.d.}{\sim} \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right)$$

  - $\boldsymbol{\Sigma}_{12} = \text{cov}(\boldsymbol{Y}_{1i}, \boldsymbol{Y}_{2i}) = \text{cov}(\boldsymbol{\eta}_{1i}, \boldsymbol{\eta}_{2i})$

- Idea: find $\boldsymbol{u} \in \mathcal{R}^{p_1}$ and $\boldsymbol{v} \in \mathcal{R}^{p_2}$, with $\|\boldsymbol{u}\|_2 = \|\boldsymbol{v}\|_2 = 1$, which maximize

$$d = \text{cov}(\boldsymbol{u}'\boldsymbol{Y}_{1i}, \boldsymbol{v}'\boldsymbol{Y}_{2i}) = \boldsymbol{u}'\boldsymbol{\Sigma}_{12}\boldsymbol{v}$$

- Singular value decomposition (SVD): $\boldsymbol{\Sigma}_{12} = \boldsymbol{UDV}'$
  - Singular values: $\boldsymbol{D}_{K\times K} = \text{diag}(d_1, \ldots, d_K); d_1 \geq \cdots \geq d_K > 0$
  - Left singular vectors: $\boldsymbol{U}_{p_1\times K} = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_K\}$
  - Right singular vectors:$\boldsymbol{V}_{p_2\times K} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K\}$

Maximum Covariance Analysis (MCA)

- Bivariate data vector

$$\begin{pmatrix} \boldsymbol{Y}_{1i} \\ \boldsymbol{Y}_{2i} \end{pmatrix} \overset{i.i.d.}{\sim} \left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \right)$$

  - $\boldsymbol{\Sigma}_{12} = \text{cov}(\boldsymbol{Y}_{1i}, \boldsymbol{Y}_{2i}) = \text{cov}(\boldsymbol{\eta}_{1i}, \boldsymbol{\eta}_{2i})$

- Idea: find $\boldsymbol{u} \in \mathcal{R}^{p_1}$ and $\boldsymbol{v} \in \mathcal{R}^{p_2}$, with $\|\boldsymbol{u}\|_2 = \|\boldsymbol{v}\|_2 = 1$, which maximize

$$d = \text{cov}(\boldsymbol{u}'\boldsymbol{Y}_{1i}, \boldsymbol{v}'\boldsymbol{Y}_{2i}) = \boldsymbol{u}'\boldsymbol{\Sigma}_{12}\boldsymbol{v}$$

- Singular value decomposition (SVD): $\boldsymbol{\Sigma}_{12} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}'$
  - Singular values: $\boldsymbol{D}_{K\times K} = \text{diag}(d_1, \ldots, d_K); d_1 \geq \cdots \geq d_K > 0$
  - Left singular vectors: $\boldsymbol{U}_{p_1\times K} = \{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_K\}$
  - Right singular vectors: $\boldsymbol{V}_{p_2\times K} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K\}$

- Coupled pattern: $(\boldsymbol{u}_1, \boldsymbol{v}_1), \ldots, (\boldsymbol{u}_K, \boldsymbol{v}_K)$

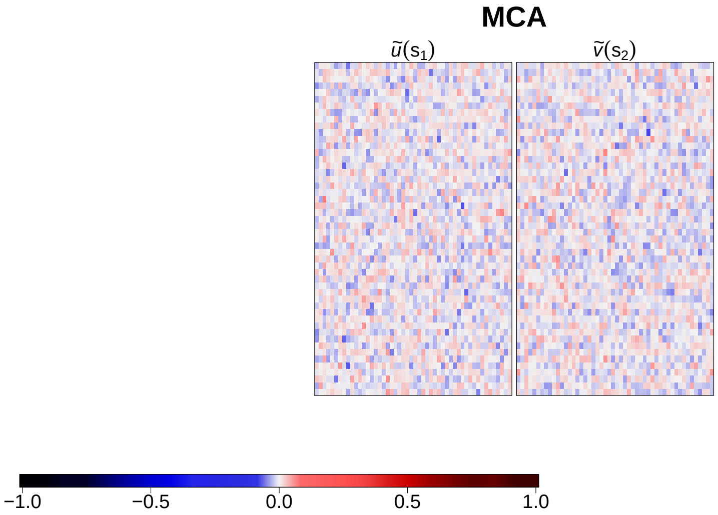
Sample Maximum Covariance Analysis

- Bivariate data matrix: $(\boldsymbol{Y}_1, \boldsymbol{Y}_2)$
  - $\boldsymbol{Y}_1 = (\boldsymbol{Y}_{11}, \ldots, \boldsymbol{Y}_{1n})'$; $\boldsymbol{Y}_2 = (\boldsymbol{Y}_{21}, \ldots, \boldsymbol{Y}_{2n})'$
- Sample cross-covariance matrix: $\boldsymbol{S}_{12} = \boldsymbol{Y}_1'\boldsymbol{Y}_2/n$
- Singular value decomposition (SVD): $\boldsymbol{S}_{12} = \tilde{\boldsymbol{U}}\tilde{\boldsymbol{D}}\tilde{\boldsymbol{V}}'$
  - $\tilde{\boldsymbol{D}}_{p_2 \times p_2} = \text{diag}(\tilde{d}_1 \ldots \tilde{d}_{p_2})$; $\tilde{d}_1 \geq \cdots \geq \tilde{d}_{p_2} \geq 0$
  - $\tilde{\boldsymbol{U}}_{p_1 \times p_2} = \{\tilde{\boldsymbol{u}}_1, \ldots, \tilde{\boldsymbol{u}}_{p_2}\}$
  - $\tilde{\boldsymbol{V}}_{p_2 \times p_2} = \{\tilde{\boldsymbol{v}}_1, \ldots, \tilde{\boldsymbol{v}}_{p_2}\}$
- $(\tilde{\boldsymbol{u}}_1, \tilde{\boldsymbol{v}}_1), \ldots, (\tilde{\boldsymbol{u}}_K, \tilde{\boldsymbol{v}}_K)$: estimates of $(\boldsymbol{u}_1, \boldsymbol{v}_1), \ldots, (\boldsymbol{u}_K, \boldsymbol{v}_K)$

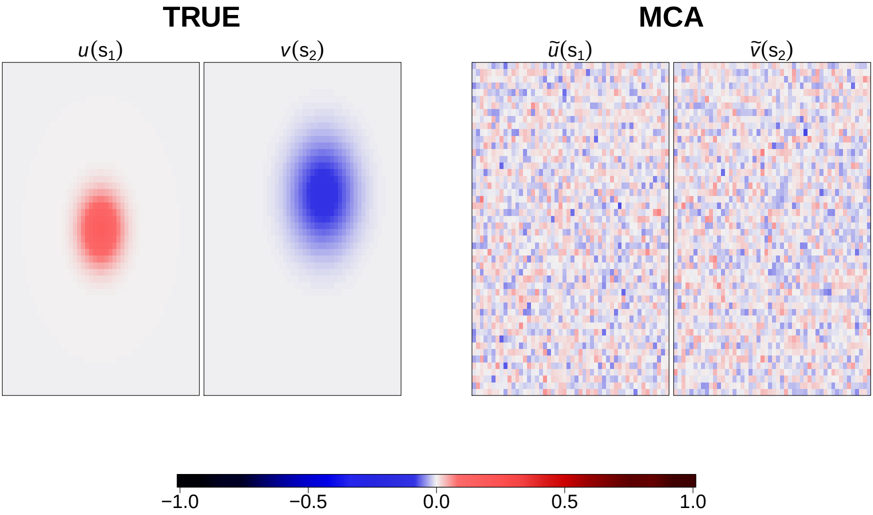
Sample Maximum Covariance Analysis

- Bivariate data matrix: $(\boldsymbol{Y}_1, \boldsymbol{Y}_2)$
  - $\boldsymbol{Y}_1 = (\boldsymbol{Y}_{11}, \ldots, \boldsymbol{Y}_{1n})'$; $\boldsymbol{Y}_2 = (\boldsymbol{Y}_{21}, \ldots, \boldsymbol{Y}_{2n})'$
- Sample cross-covariance matrix: $\boldsymbol{S}_{12} = \boldsymbol{Y}_1'\boldsymbol{Y}_2/n$
- Singular value decomposition (SVD): $\boldsymbol{S}_{12} = \tilde{\boldsymbol{U}}\tilde{\boldsymbol{D}}\tilde{\boldsymbol{V}}'$
  - $\tilde{\boldsymbol{D}}_{p_2 \times p_2} = \text{diag}(\tilde{d}_1 \ldots \tilde{d}_{p_2})$; $\tilde{d}_1 \geq \cdots \geq \tilde{d}_{p_2} \geq 0$
  - $\tilde{\boldsymbol{U}}_{p_1 \times p_2} = \{\tilde{\boldsymbol{u}}_1, \ldots, \tilde{\boldsymbol{u}}_{p_2}\}$
  - $\tilde{\boldsymbol{V}}_{p_2 \times p_2} = \{\tilde{\boldsymbol{v}}_1, \ldots, \tilde{\boldsymbol{v}}_{p_2}\}$
- $(\tilde{\boldsymbol{u}}_1, \tilde{\boldsymbol{v}}_1), \ldots, (\tilde{\boldsymbol{u}}_K, \tilde{\boldsymbol{v}}_K)$: estimates of $(\boldsymbol{u}_1, \boldsymbol{v}_1), \ldots, (\boldsymbol{u}_K, \boldsymbol{v}_K)$
- Problem:
  - high estimation variability: n small, p_1 or p_2 large
    - noisy patterns $\rightarrow$ low interpretation
  - without a spatial structure of $(\boldsymbol{u}, \boldsymbol{v})$

Example:



Example:



Motivation

To enhance the interpretability, we implement MCA by incorporating

Motivation

To enhance the interpretability, we implement MCA by incorporating

1. spatial structure of $(u_k(\cdot), v_k(\cdot))$

Motivation

To enhance the interpretability, we implement MCA by incorporating

1. spatial structure of $(u_k(\cdot), v_k(\cdot))$
2. spatial localized patterns of $(u_k(\cdot), v_k(\cdot))$

Motivation

To enhance the interpretability, we implement MCA by incorporating

1. spatial structure of $(u_k(\cdot), v_k(\cdot))$
2. spatial localized patterns of $(u_k(\cdot), v_k(\cdot))$

retain that orthogonal constraints for $(\boldsymbol{u}, \boldsymbol{v})$.

Quick Recap

- Data: $\boldsymbol{Y}_{1i} = (Y_{1i}(\boldsymbol{s}_{11}), \ldots, Y_{1i}(\boldsymbol{s}_{1p_1}))'$, $\boldsymbol{Y}_{2i} = (Y_{2i}(\boldsymbol{s}_{21}), \ldots, Y_{2i}(\boldsymbol{s}_{p_2}))'$; $i = 1, \ldots, n$
  - $Y_{1i}(\boldsymbol{s}_{1j}) = \eta_{1i}(\boldsymbol{s}_{1j}) + \epsilon_{1ij}$; $j = 1, \ldots, p_1$
    - $\epsilon_{1ij} \sim (0, \sigma_1^2)$
    - ϵ_{1ij}: uncorrelated with $\eta_1(\cdot)$
  - $Y_{2i}(\boldsymbol{s}_{2j}) = \eta_{2i}(\boldsymbol{s}_{2j}) + \epsilon_{2ij}$; $j = 1, \ldots, p_2$
    - $\epsilon_{2ij} \sim (0, \sigma_1^2)$
    - ϵ_{2ij}: uncorrelated with $\eta_2(\cdot)$

Quick Recap

- Data: $\boldsymbol{Y}_{1i} = (Y_{1i}(\boldsymbol{s}_{11}), \ldots, Y_{1i}(\boldsymbol{s}_{1p_1}))'$, $\boldsymbol{Y}_{2i} = (Y_{2i}(\boldsymbol{s}_{21}), \ldots, Y_{2i}(\boldsymbol{s}_{p_2}))'$; $i = 1, \ldots, n$
  - $Y_{1i}(\boldsymbol{s}_{1j}) = \eta_{1i}(\boldsymbol{s}_{1j}) + \epsilon_{1ij}$; $j = 1, \ldots, p_1$
    - $\epsilon_{1ij} \sim (0, \sigma_1^2)$
    - ϵ_{1ij}: uncorrelated with $\eta_1(\cdot)$
  - $Y_{2i}(\boldsymbol{s}_{2j}) = \eta_{2i}(\boldsymbol{s}_{2j}) + \epsilon_{2ij}$; $j = 1, \ldots, p_2$
    - $\epsilon_{2ij} \sim (0, \sigma_1^2)$
    - ϵ_{2ij}: uncorrelated with $\eta_2(\cdot)$

- Spatial cross-covariance function:

$$C_{12}(\boldsymbol{s}_1, \boldsymbol{s}_2) = \mathrm{cov}(\eta_{1i}(\boldsymbol{s}_1), \eta_{2i}(\boldsymbol{s}_2)) = \sum_{k=1}^{K} d_k u_k(\boldsymbol{s}_1) v_k(\boldsymbol{s}_2)$$

  - $d_1 \geq \cdots \geq d_K \geq 0$
  - $u_1(\cdot), \ldots, u_K(\cdot)$: K unknown orthonormal functions
  - $v_1(\cdot), \ldots, v_K(\cdot)$: K unknown orthonormal functions

MCA (alternative version)

- Sample cross-covariance matrix: $\boldsymbol{S}_{12} = \boldsymbol{Y}_1'\boldsymbol{Y}_2/n$
- MCA: perform SVD of $\boldsymbol{S}_{12}$

- Alternative method:

$$(\tilde{\boldsymbol{U}}, \tilde{\boldsymbol{V}}) = \underset{\boldsymbol{U},\boldsymbol{V}}{\arg\max}\, \mathrm{tr}(\boldsymbol{U}'\boldsymbol{S}_{12}\boldsymbol{V}),$$

  subject to $\boldsymbol{U}'\boldsymbol{U} = \boldsymbol{V}'\boldsymbol{V} = \boldsymbol{I}_K$
  - $\boldsymbol{U}_{p_1 \times K} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_K)$ with $u_{jk} = u_k(\boldsymbol{s}_{1j})$
  - $\boldsymbol{V}_{p_2 \times K} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K)$ with $v_{jk} = v_k(\boldsymbol{s}_{2j})$

Regularized MCA

- Sample cross-covariance matrix: $\boldsymbol{S}_{12} = \boldsymbol{X}'\boldsymbol{Y}/n$
- $\boldsymbol{U}_{p_1 \times K} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_K)$ with $u_{jk} = u_k(\boldsymbol{s}_{1j})$
- $\boldsymbol{V}_{p_2 \times K} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K)$ with $v_{jk} = v_k(\boldsymbol{s}_{2j})$

- Objective function:

  tr$(\boldsymbol{U}'\boldsymbol{S}_{12}\boldsymbol{V})$

  subject to $\boldsymbol{U}'\boldsymbol{U} = \boldsymbol{V}'\boldsymbol{V} = \boldsymbol{I}_K$

Regularized MCA

- Sample cross-covariance matrix: $\boldsymbol{S}_{12} = \boldsymbol{X}'\boldsymbol{Y}/n$
- $\boldsymbol{U}_{p_1 \times K} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_K)$ with $u_{jk} = u_k(\boldsymbol{s}_{1j})$
- $\boldsymbol{V}_{p_2 \times K} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K)$ with $v_{jk} = v_k(\boldsymbol{s}_{2j})$

- Objective function:
  $$\mathrm{tr}(\boldsymbol{U}'\boldsymbol{S}_{12}\boldsymbol{V}) - \sum_{k=1}^{K} \left\{ \tau_{1u} J(u_k) + \tau_{2u} \sum_{j=1}^{p_1} |u_k(\boldsymbol{s}_{1j})| + \tau_{1v} J(v_k) + \tau_{2v} \sum_{j=1}^{p_2} |v_k(\boldsymbol{s}_{2j})| \right\}$$
  subject to $\boldsymbol{U}'\boldsymbol{U} = \boldsymbol{V}'\boldsymbol{V} = \boldsymbol{I}_K$

Regularized MCA

- Sample cross-covariance matrix: $\boldsymbol{S}_{12} = \boldsymbol{X}'\boldsymbol{Y}/n$
- $\boldsymbol{U}_{p_1 \times K} = (\boldsymbol{u}_1, \ldots, \boldsymbol{u}_K)$ with $u_{jk} = u_k(\boldsymbol{s}_{1j})$
- $\boldsymbol{V}_{p_2 \times K} = (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_K)$ with $v_{jk} = v_k(\boldsymbol{s}_{2j})$

- Objective function:

$$\mathrm{tr}(\boldsymbol{U}'\boldsymbol{S}_{12}\boldsymbol{V}) - \sum_{k=1}^{K} \left\{ \tau_{1u} J(u_k) + \tau_{2u} \sum_{j=1}^{p_1} |u_k(\boldsymbol{s}_{1j})| + \tau_{1v} J(v_k) + \tau_{2v} \sum_{j=1}^{p_2} |v_k(\boldsymbol{s}_{2j})| \right\}$$

subject to $\boldsymbol{U}'\boldsymbol{U} = \boldsymbol{V}'\boldsymbol{V} = \boldsymbol{I}_K$

  - $J(u_k(\cdot)) = \displaystyle\sum_{z_1+\cdots+z_d=2} \int_{R^d} \left(\frac{\partial^2 u_k(\boldsymbol{s})}{\partial x_1^{z_1} \ldots \partial x_d^{z_d}} \right)^2 d\boldsymbol{s}$
    - $\boldsymbol{s} = (x_1, \ldots, x_d)'$
  - τ_{1u}, τ_{1v}: smoothness parameter
  - τ_{2u}, τ_{2v}: sparseness parameter

Spatial MCA (SpatMCA)

- $J(u_k(\cdot)) = \boldsymbol{u}_k' \boldsymbol{\Omega}_1 \boldsymbol{u}_k, J(v_k(\cdot)) = \boldsymbol{v}_k' \boldsymbol{\Omega}_2 \boldsymbol{v}_k$
  - $\boldsymbol{\Omega}_1 \succ \mathbf{0}$: determined only by $\boldsymbol{s}_{11}, \ldots, \boldsymbol{s}_{1p_1}$
  - $\boldsymbol{\Omega}_2 \succ \mathbf{0}$: determined only by $\boldsymbol{s}_{21}, \ldots, \boldsymbol{s}_{2p_2}$
  - Green and Silverman (1994)

Spatial MCA (SpatMCA)

- $J(u_k(\cdot)) = \boldsymbol{u}_k'\boldsymbol{\Omega}_1\boldsymbol{u}_k$, $J(v_k(\cdot)) = \boldsymbol{v}_k'\boldsymbol{\Omega}_2\boldsymbol{v}_k$
  - $\boldsymbol{\Omega}_1 \succ \boldsymbol{0}$: determined only by $\boldsymbol{s}_{11}, \ldots, \boldsymbol{s}_{1p_1}$
  - $\boldsymbol{\Omega}_2 \succ \boldsymbol{0}$: determined only by $\boldsymbol{s}_{21}, \ldots, \boldsymbol{s}_{2p_2}$
  - Green and Silverman (1994)

- SpatMCA: $(\hat{\boldsymbol{U}}, \hat{\boldsymbol{V}})$ maximizes:

$$\mathrm{tr}(\boldsymbol{U}'\boldsymbol{S}_{12}\boldsymbol{V}) - \sum_{k=1}^{K}\left\{\tau_{1u}\boldsymbol{u}_k'\boldsymbol{\Omega}_1\boldsymbol{u}_k + \tau_{2u}\sum_{j=1}^{p_1}|u_{jk}| + \tau_{1v}\boldsymbol{v}_k'\boldsymbol{\Omega}_2\boldsymbol{v}_k + \tau_{2v}\sum_{j=1}^{p_2}|v_{jk}|\right\}$$

subject to $\boldsymbol{U}'\boldsymbol{U} = \boldsymbol{V}'\boldsymbol{V} = \boldsymbol{I}_K$

Spatial MCA (SpatMCA)

- $J(u_k(\cdot)) = \boldsymbol{u}_k'\boldsymbol{\Omega}_1\boldsymbol{u}_k$, $J(v_k(\cdot)) = \boldsymbol{v}_k'\boldsymbol{\Omega}_2\boldsymbol{v}_k$
  - $\boldsymbol{\Omega}_1 \succ \boldsymbol{0}$: determined only by $\boldsymbol{s}_{11}, \dots, \boldsymbol{s}_{1p_1}$
  - $\boldsymbol{\Omega}_2 \succ \boldsymbol{0}$: determined only by $\boldsymbol{s}_{21}, \dots, \boldsymbol{s}_{2p_2}$
  - Green and Silverman (1994)

- SpatMCA: $(\hat{\boldsymbol{U}}, \hat{\boldsymbol{V}})$ maximizes:

$$\operatorname{tr}(\boldsymbol{U}'\boldsymbol{S}_{12}\boldsymbol{V}) - \sum_{k=1}^{K}\left\{\tau_{1u}\boldsymbol{u}_k'\boldsymbol{\Omega}_1\boldsymbol{u}_k + \tau_{2u}\sum_{j=1}^{p_1}|u_{jk}| + \tau_{1v}\boldsymbol{v}_k'\boldsymbol{\Omega}_2\boldsymbol{v}_k + \tau_{2v}\sum_{j=1}^{p_2}|v_{jk}|\right\}$$

subject to $\boldsymbol{U}'\boldsymbol{U} = \boldsymbol{V}'\boldsymbol{V} = \boldsymbol{I}_K$

- As $\tau_{1u} = \tau_{1v} = \tau_{2u} = \tau_{2v} = 0$, $(\hat{\boldsymbol{U}}, \hat{\boldsymbol{V}}) = (\tilde{\boldsymbol{U}}, \tilde{\boldsymbol{V}})$ (sample MCA)

SpatMCA: $(\hat{u}_1(\cdot), \hat{v}_1(\cdot)), \dots, (\hat{u}_K(\cdot), \hat{v}_K(\cdot))$

- $(\hat{u}_1(\cdot), \hat{v}_1(\cdot)), \dots, (\hat{u}_K(\cdot), \hat{v}_K(\cdot))$ maximizes
  $$\mathrm{tr}(\boldsymbol{U}'\boldsymbol{S}_{12}\boldsymbol{V}) - \sum_{k=1}^{K} \left\{ \tau_{1u} J(u_k) + \tau_{2u} \sum_{j=1}^{p_1} |u_k(\boldsymbol{s}_{1j})| + \tau_{1v} J(v_k) + \tau_{2v} \sum_{j=1}^{p_2} |v_k(\boldsymbol{s}_{2j})| \right\},$$
  subject to $\boldsymbol{U}'\boldsymbol{U} = \boldsymbol{V}'\boldsymbol{V} = \boldsymbol{I}_K$

- $$\hat{u}_k(\boldsymbol{s}_1) = \sum_{i=1}^{p_1} a_{1i} g(\|\boldsymbol{s}_1 - \boldsymbol{s}_{1i}\|) + b_{10} + \sum_{j=1}^{d} b_{1j} x_{1j}$$
  $$\hat{v}_k(\boldsymbol{s}_2) = \sum_{i=1}^{p_2} a_{2i} g(\|\boldsymbol{s}_2 - \boldsymbol{s}_{2i}\|) + b_{20} + \sum_{j=1}^{d} b_{2j} x_{2j}$$

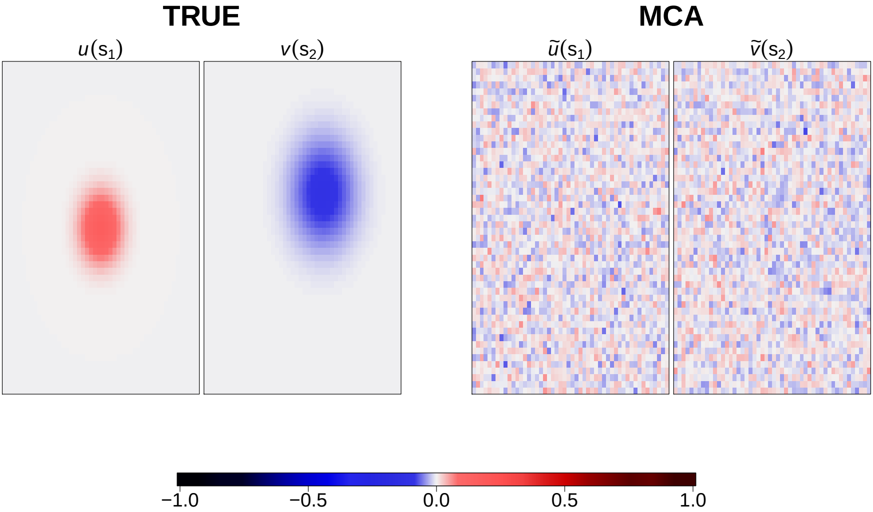
SpatMCA: $(\hat{u}_1(\cdot), \hat{v}_1(\cdot)), \ldots, (\hat{u}_K(\cdot), \hat{v}_K(\cdot))$

- $(\hat{u}_1(\cdot), \hat{v}_1(\cdot)), \ldots, (\hat{u}_K(\cdot), \hat{v}_K(\cdot))$ maximizes
  $$\mathrm{tr}(\boldsymbol{U}'\boldsymbol{S}_{12}\boldsymbol{V}) - \sum_{k=1}^{K}\left\{\tau_{1u}J(u_k)+\tau_{2u}\sum_{j=1}^{p_1}|u_k(\boldsymbol{s}_{1j})| + \tau_{1v}J(v_k)+\tau_{2v}\sum_{j=1}^{p_2}|v_k(\boldsymbol{s}_{2j})|\right\},$$
  subject to $\boldsymbol{U}'\boldsymbol{U} = \boldsymbol{V}'\boldsymbol{V} = \boldsymbol{I}_K$

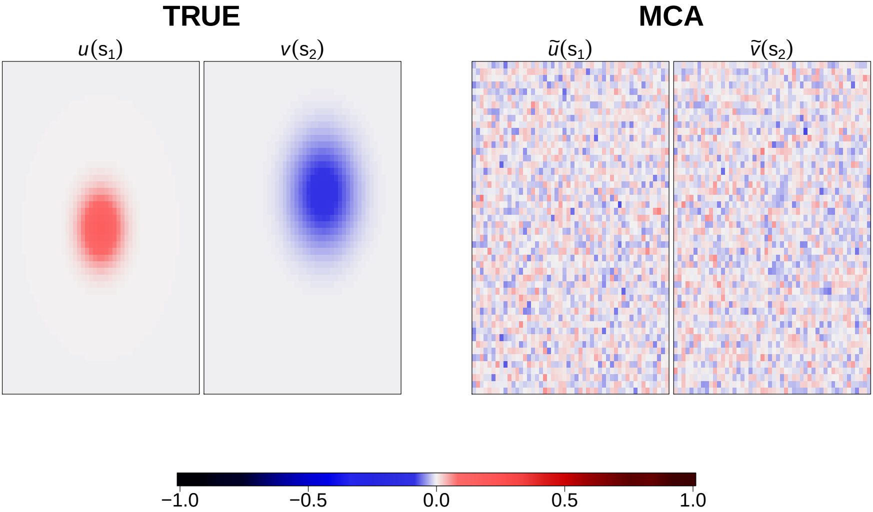
- $$\hat{u}_k(\boldsymbol{s}_1) = \sum_{i=1}^{p_1} a_{1i}g(\|\boldsymbol{s}_1 - \boldsymbol{s}_{1i}\|) + b_{10} + \sum_{j=1}^{d} b_{1j}x_{1j}$$
  $$\hat{v}_k(\boldsymbol{s}_2) = \sum_{i=1}^{p_2} a_{2i}g(\|\boldsymbol{s}_2 - \boldsymbol{s}_{2i}\|) + b_{20} + \sum_{j=1}^{d} b_{2j}x_{2j}$$
  - $\boldsymbol{s}_1 = (x_{11}, \ldots, x_{1d})'$; $\boldsymbol{s}_2 = (x_{21}, \ldots, x_{2d})'$
  - $$g(r) = \begin{cases} \frac{1}{16\pi} r^2 \log r; & \text{if } d = 2, \\ \frac{\Gamma(d/2-2)}{16\pi^{d/2}} r^{4-d}; & \text{if } d = 1, 3, \end{cases}$$
  - $\boldsymbol{a}_1 = (a_{11}, \ldots, a_{1p_1})'$ and $\boldsymbol{b}_1 = (b_{10}, b_{11}, \ldots, b_{1d})'$ based on $\hat{\boldsymbol{u}}_k$
  - $\boldsymbol{a}_2 = (a_{21}, \ldots, a_{2p_2})'$ and $\boldsymbol{b}_2 = (b_{20}, b_{21}, \ldots, b_{2d})'$ based on $\hat{\boldsymbol{v}}_k$

Why roughness and Lasso penalties?

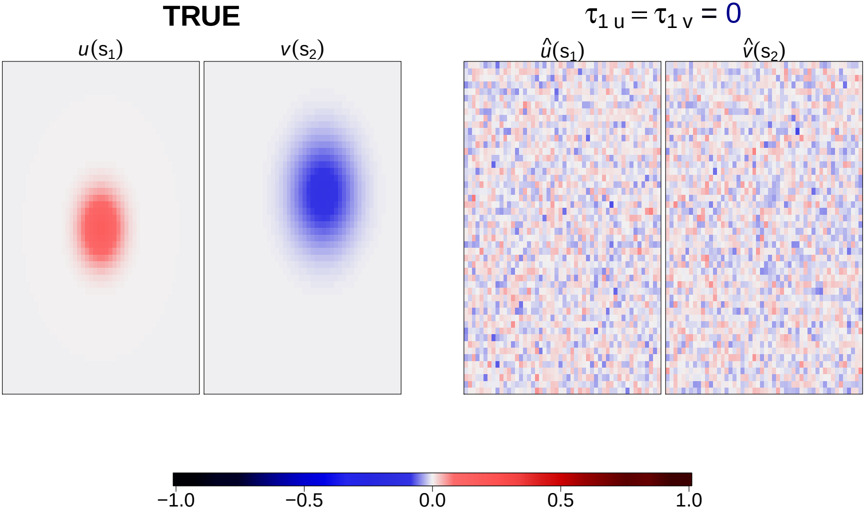
2D Example



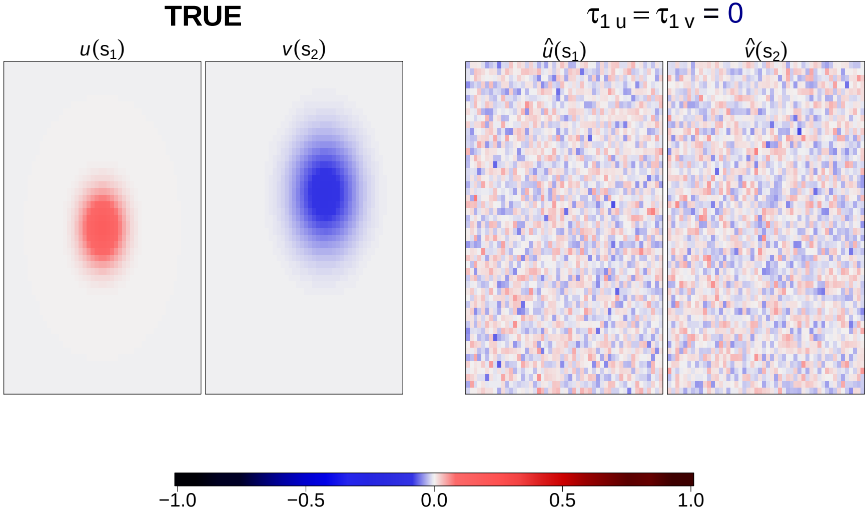
2D Example



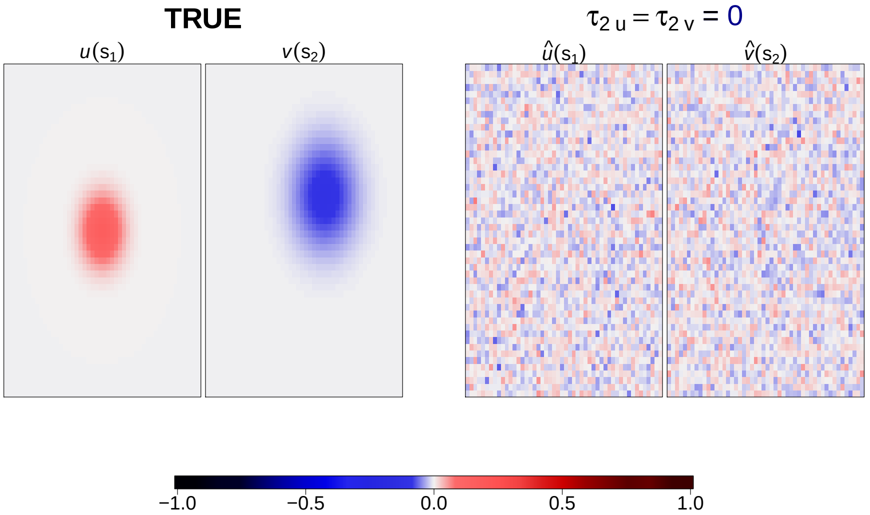
Case 1: $\tau_{2u} = \tau_{2v} = 0$ (only smoothness)



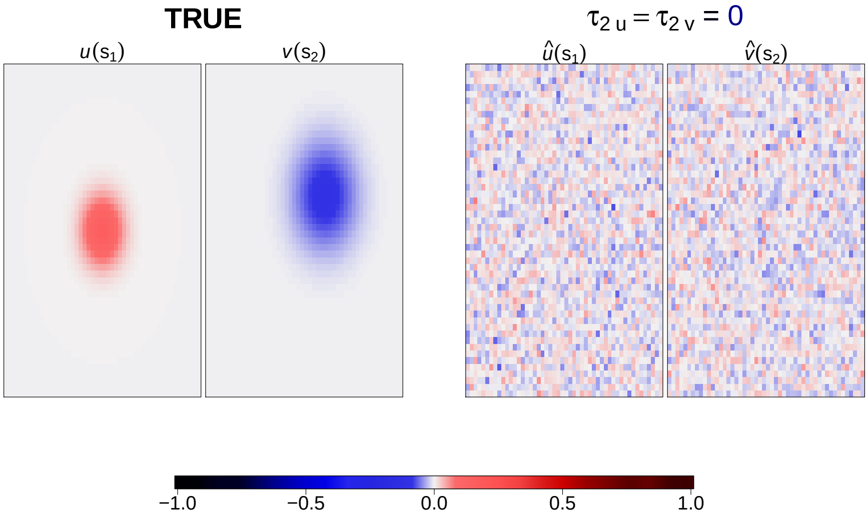
Case 1: $\tau_{2u} = \tau_{2v} = 0$ (only smoothness)



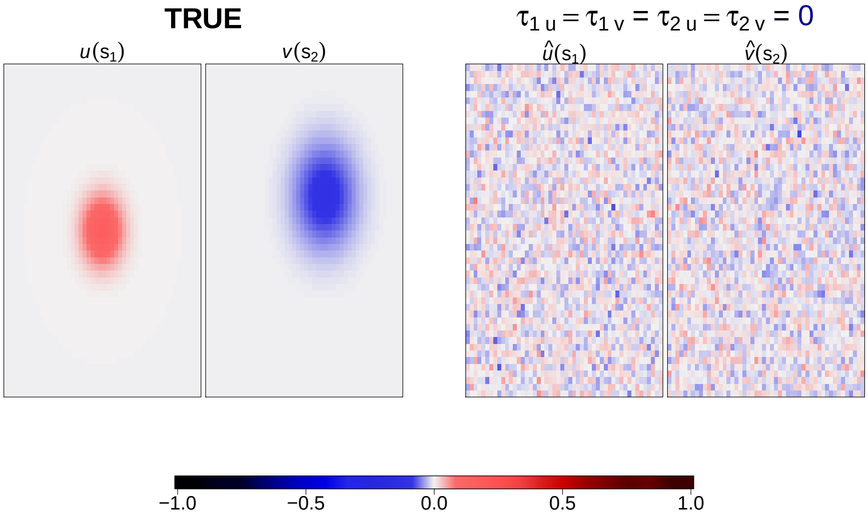
Case 2: $\tau_{1u} = \tau_{1v} = 0$ (only sparseness)



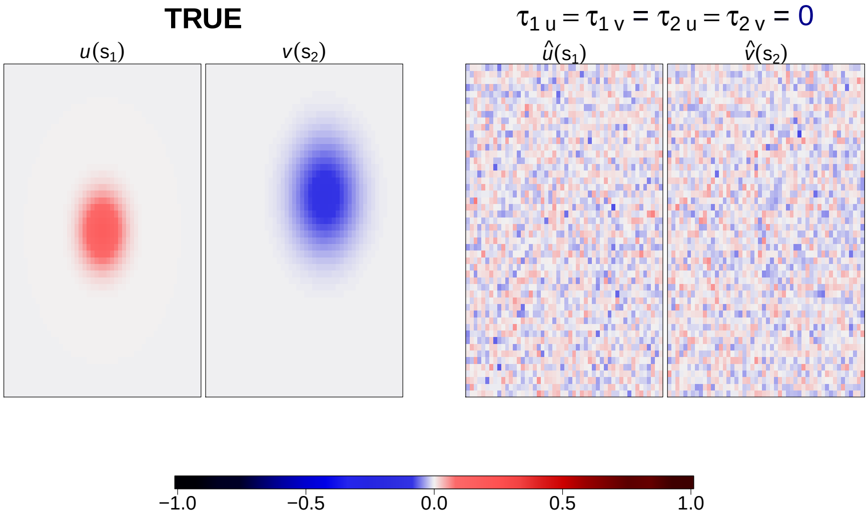
Case 2: $\tau_{1u} = \tau_{1v} = 0$ (only sparseness)



Case 3: $\tau_{1u} = \tau_{1v} = \tau_{2u} = \tau_{2v}$



Case 3: $\tau_{1u} = \tau_{1v} = \tau_{2u} = \tau_{2v}$



Estimation of D

Given the SpatMCA estimate $(\hat{\boldsymbol{U}}, \hat{\boldsymbol{V}})$,

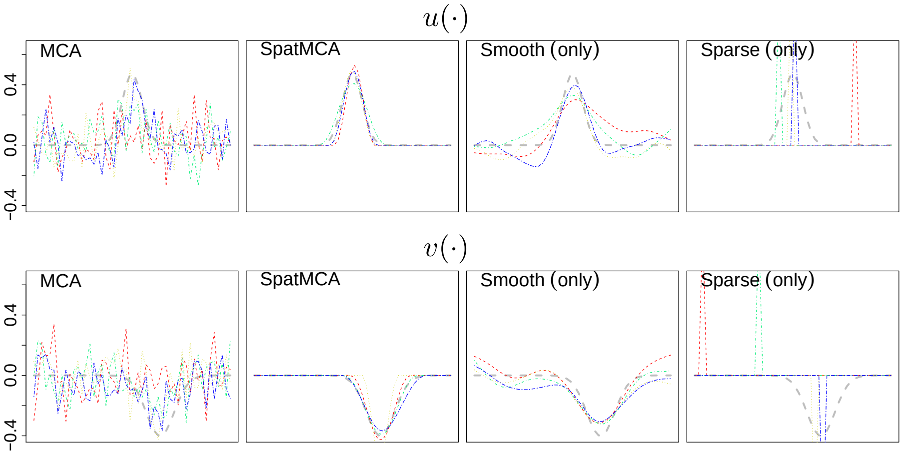
$$\hat{\boldsymbol{D}} = \operatorname*{arg\,min}_{d_1,\ldots,d_K \geq 0} \|\boldsymbol{S}_{12} - \hat{\boldsymbol{U}}\boldsymbol{D}\hat{\boldsymbol{V}}'\|_F^2 = \mathrm{diag}(\hat{d}_1, \ldots, \hat{d}_K)$$

- $\hat{d}_k = \min(\hat{u}_k' \boldsymbol{S}_{12} \hat{v}_k, 0)$

Selection of $(\tau_{1u}, \tau_{2u}, \tau_{1v}, \tau_{2v})$

- Similar to SpatPCA, $(\tau_{1u}, \tau_{2u}, \tau_{1v}, \tau_{2v})$ is selected by K-fold cross validation
  - CV criterion is based on rank-K approximation

Example (1D): 5-fold CV



Outline

Real data analysis

- Bivariate data:
  - Sea surface temperature (SST):
    - Region: Indian Ocean (20°N and 30°S; 20°E and 120°E)
    - Number of grids: $p_1 = 3,591$
  - Rainfall:
    - Region: Eastern African (6°N and 12°S; 20°E and 42°E)
    - Number of grids: $p_2 = 255$
  - Time period (monthly): Jan. 2011- Dec. 2015 $\rightarrow n = 60$
  - Remove monthly mean

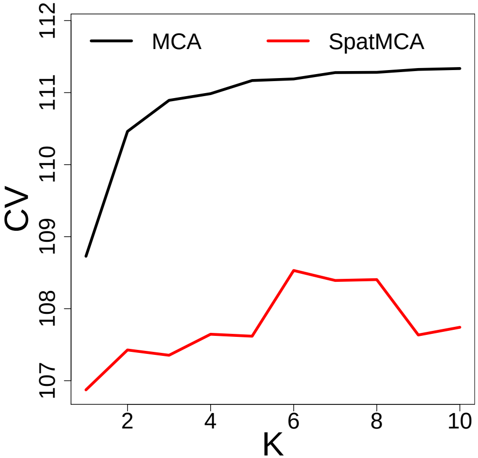
Real data analysis

- Bivariate data:
  - Sea surface temperature (SST):
    - Region: Indian Ocean (20°N and 30°S; 20°E and 120°E)
    - Number of grids: $p_1 = 3,591$
  - Rainfall:
    - Region: Eastern African (6°N and 12°S; 20°E and 42°E)
    - Number of grids: $p_2 = 255$
  - Time period (monthly): Jan. 2011- Dec. 2015 $\rightarrow n = 60$
  - Remove monthly mean
- Goal: find coupled patterns of the SST and rainfall data
- Reference: Omondi et al. (2013)

Real data analysis

- Randomly decompose the data into two parts with 30 time points
  - Training data
  - Validation data
- SpatMCA: based on 5-fold CV

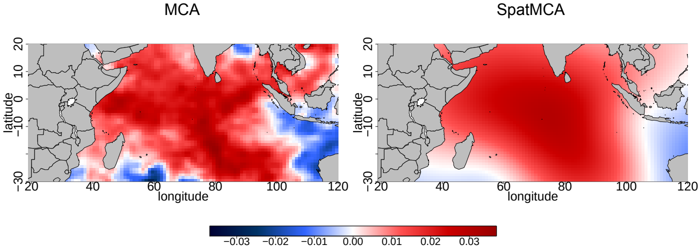
Result: CV vs. K



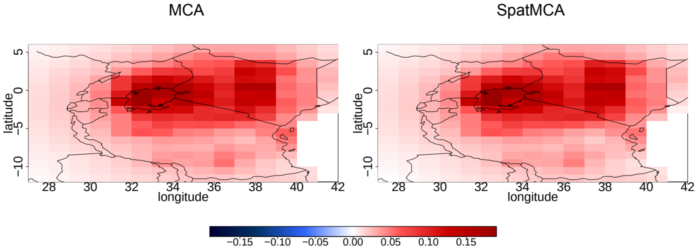
- $\hat{K} = 1$ for MCA and SpatMCA

Result: 1st coupled pattern

- SST

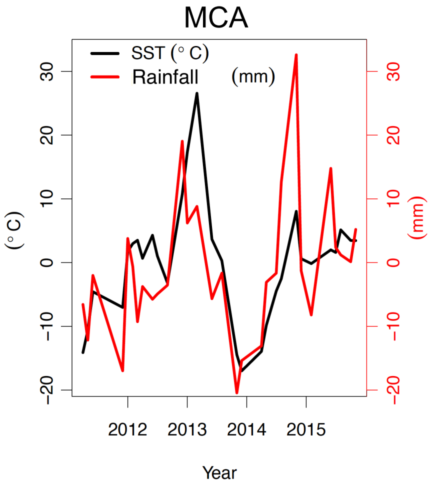


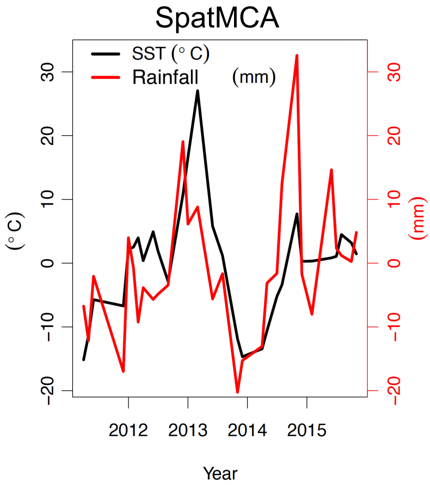
- Rainfall



Result: 1st Maximum covariance Variables

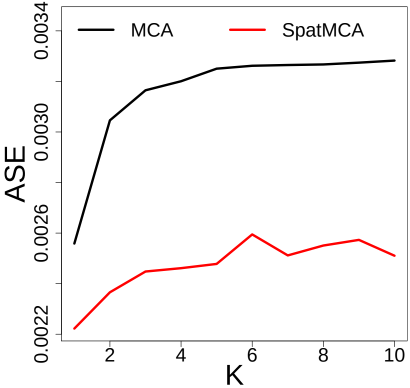
- 1st maximum covariance variables: $\{\hat{u}_1' \boldsymbol{Y}_{1i}\}$; $\{\hat{v}_1' \boldsymbol{Y}_{2i}\}$
- Pearson's correlation: 0.6 for MCA and SpatMCA





Result: Average Squared Error (ASE)

- ASE $= \frac{1}{p_1 p_2} \| \boldsymbol{S}^v_{12} - \hat{\boldsymbol{U}}_K \hat{\boldsymbol{D}}_K \hat{\boldsymbol{V}}'_K \|_F^2$
  - $\boldsymbol{S}^v_{12}$: sample cross-covariance matrix based on validation data
- Result:



Outline

Summary

SpatPCA/ SpatMCA:

- high-dimensional structure $\rightarrow$ low-dimensional structure
- with the roughness and Lasso penalties
- enhance physical interpretation, e.g. spatial localized patterns

Summary

SpatPCA/ SpatMCA:

- high-dimensional structure $\rightarrow$ low-dimensional structure
- with the roughness and Lasso penalties
- enhance physical interpretation, e.g. spatial localized patterns
- non-stationary spatial covariance function (SpatPCA)
- can cope with irregular spaced locations
- R packages on CRAN: *SpatPCA; SpatMCA*

Thanks for your attention!