Spatial Correlation Estimation for Sparsely Observed Functional Data

Workshop on High dimensional and dependent functional data, Bristol, Sep 2012

Surajit Ray
University of Glasgow

University of Glasgow

Contributors

Chong Liu

Doctoral Student(BU)

Mark Friedl

Dept. of Earth Sciences and Geography(BU)

Giles Hooker

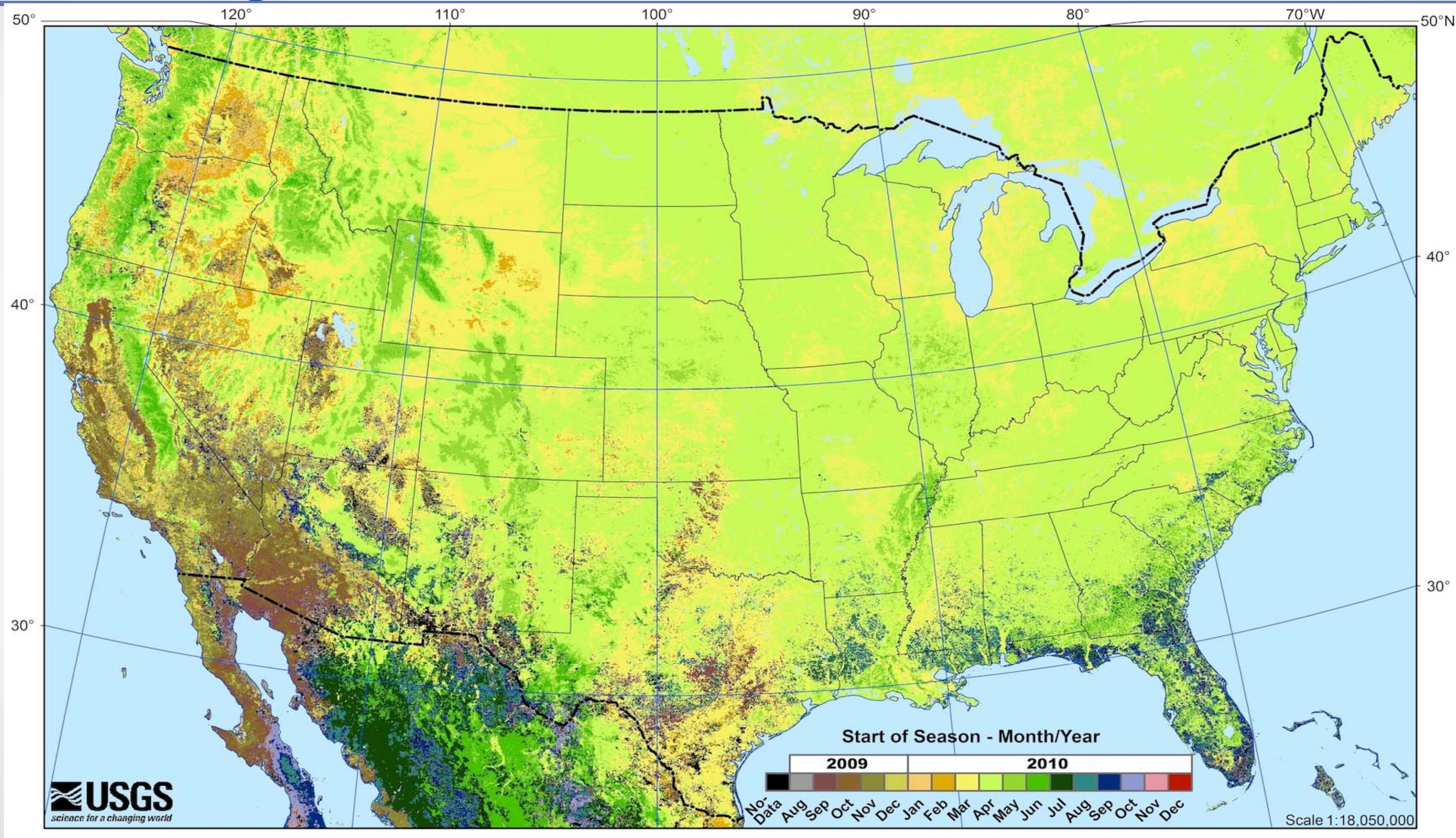
Dept. of Statistics (Cornell University)

University of Glasgow

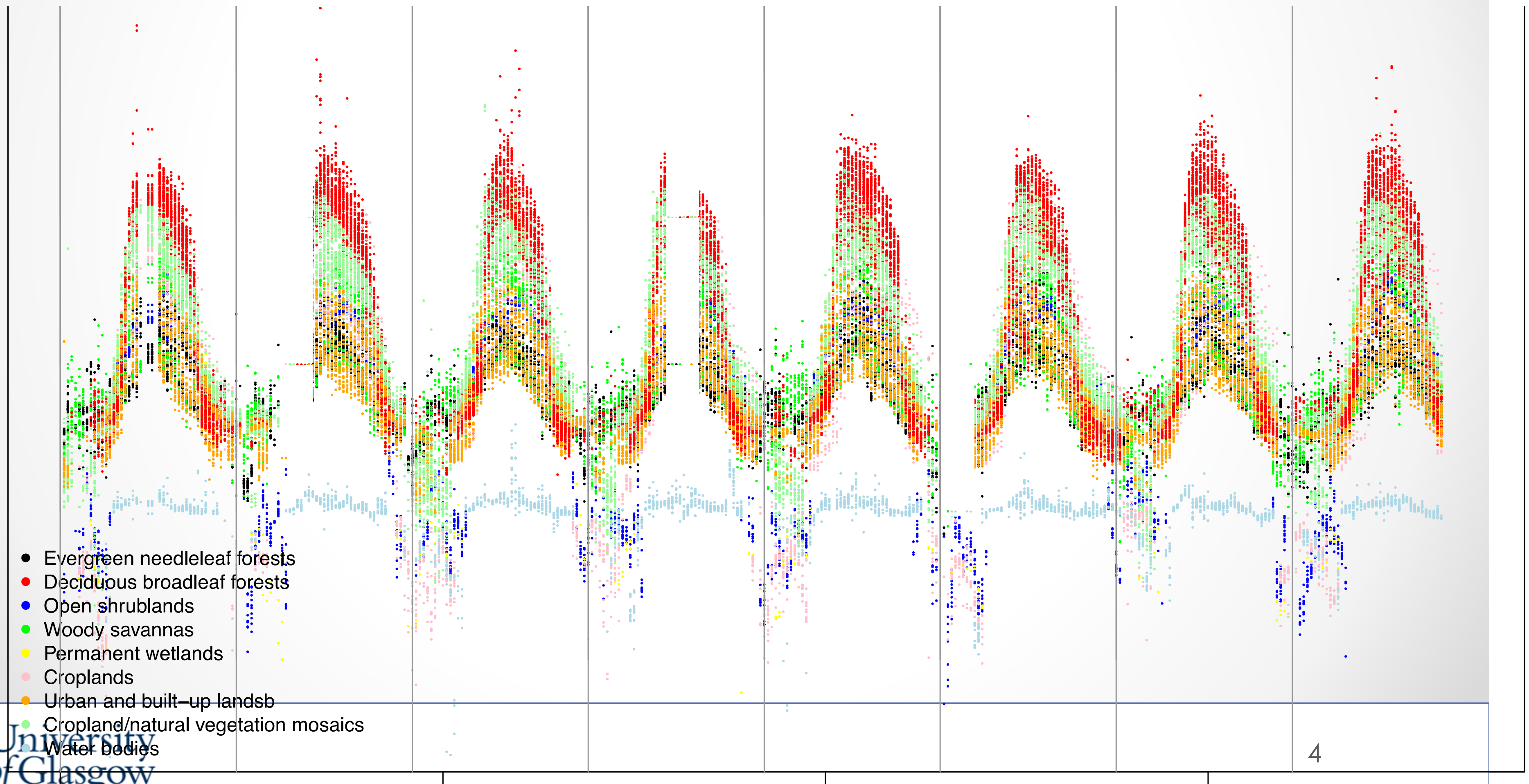
Outline

- Context and Background
- Moment Based Method
- Likelihood Based Method Method
- Curve Reconstruction Results
- Future Work

``Gap Filling" for missing remote sensing observation



Multi year remote sensing data

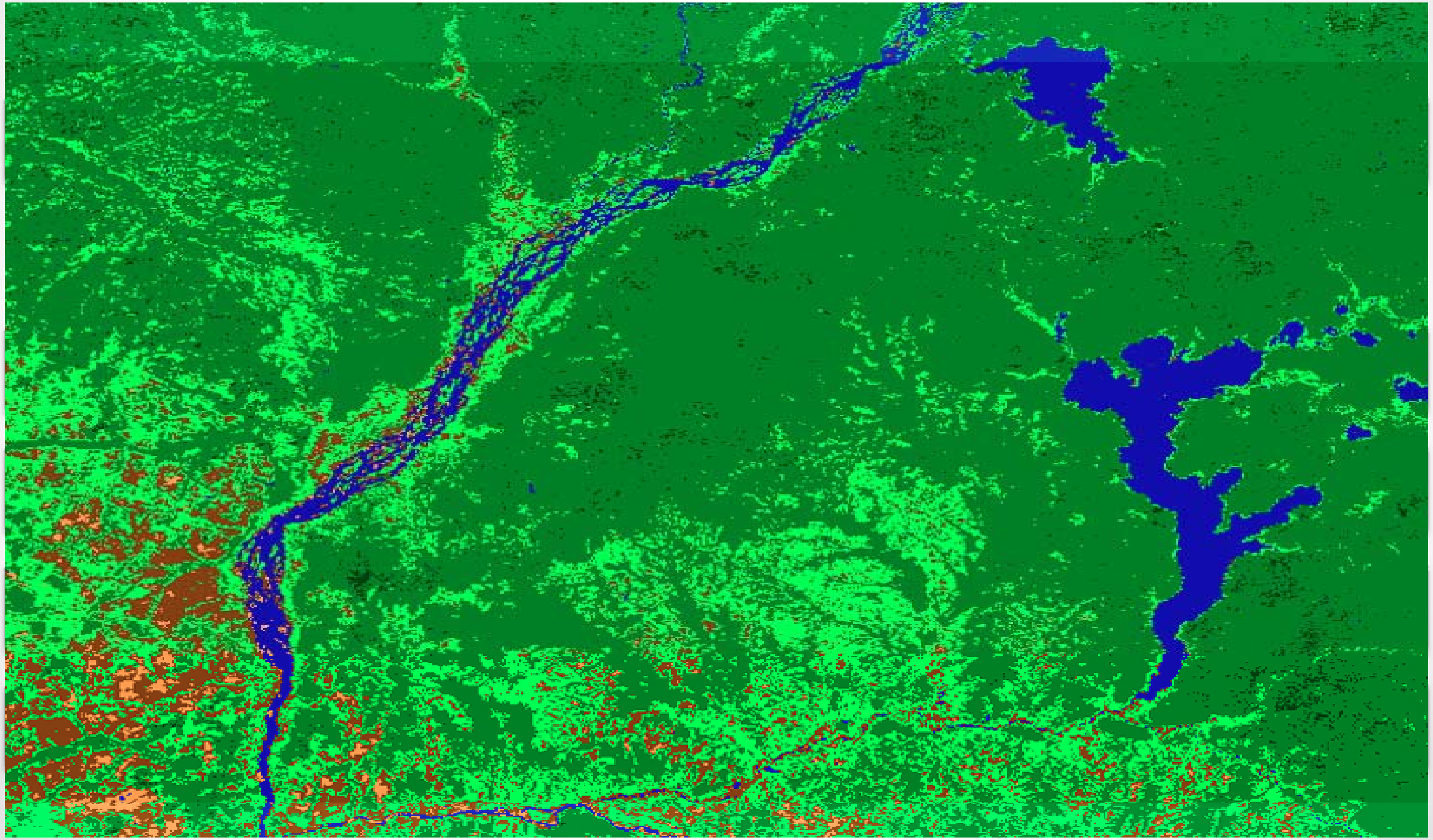


``Gap Filling" for missing remote sensing observation

- Longitudinal remote sensing data are collected for studying various climate ecosystem phenomenon
- Often a lot of observations are missing due to various reasons
    - Cloud Cover
    - Aerosol Content
    - Change of instrument
    - Fire
    - Snow

- Geoscientists would love to fill those gaps for applying "standard statistical techniques."

Properties of remotely sensed ecosystem data



Background and Context

Model for independent curves

$$Y_i(t) = \sum_{k=1}^{K} \gamma_{ik} \phi_k(t) + \varepsilon_i(t)$$

where eigen-scores γ_{ik} are independent across i.

- However, in real world, many applications have correlated $Y_i(t)$ across i.
  - ❖ example: spatial-temporal data, online auction data, time course gene expression data.
- Most existing work treat them as i.i.d.
- Asymptotic property holds for mild correlation.

University of Glasgow

Previous Research

- Yao,Muller and Wang 2004 outlined the **moment based** methods to estimate covariance surface and eigenstructure of the random process assuming i.i.d. curves.

    - Also suggested reconstructing curve trajectories ("gap filling") using expected principal component scores.

Question: How to estimate eigenstructure and reconstructing trajectories assuming correlated curves?

Concurrent Research

- "Reduced Rank Mixed Effects Models for Spatially Correlated Hierarchical Functional Data" JASA 2010, Zhou et al.
  - Mentioned by Maurice Berk during his multilevel talk.

- Principal components analysis for sparsely observed correlated functional data using a kernel smoothing approach, Electron. J. Statist. Volume 5 (2011). Paul and Peng

- Some that I have missed …

Our Model for correlate data

2. Our Model:

$$Y_i(t) = \sum_{k=1}^{K} \gamma_{ik}\phi_k(t) + \varepsilon_i(t)$$

where

$$\operatorname{cov}(\gamma_{ip}, \gamma_{jq}) = \begin{cases} 0, & p \neq q \\ \rho^{|i-j|}\lambda_k, & p = q = k \end{cases}$$

more general model:

$$\operatorname{cov}(\gamma_{ip}, \gamma_{jq}) = \begin{cases} 0, & p \neq q \\ matern(d(i,j), \alpha_k, \beta_k)\lambda_k, & p = q = k \end{cases}$$

Goal: estimating ρ or more generally α_k and β_k

University of Glasgow

Estimation of Parameters

Moment Based Method

- local linear smoothing of the covariance surface $\sum_0$ and
- lag-d cross-covariance surface $\sum_d$ where d = 1,2,...D,
- take ratios of eigenvalues as covariance estimates and fit matern or other parametric models.

Likelihood based method

- Marginal likelihood is hard to optimize.
- Turn to joint likelihood of Y and and treat as random effects.
- Use EM(expectation maximization) to solve the optimal parameters ρ, α_k and β_k

University of Glasgow

Moment based method

- Covariance Model

$$\text{cov}(Y_i, Y_j) = \Phi \, \text{cov}(\gamma_i, \gamma_j)\Phi' + \sigma^2 \text{I}$$

$$= \Phi \, diag(\text{cov}(\gamma_{i1}, \gamma_{j1}), \text{cov}(\gamma_{i2}, \gamma_{j2}), \ldots, \text{cov}(\gamma_{iK}, \gamma_{jK}))\Phi' + \sigma^2 \text{I}$$

- Smooth covariance to get estimate of $\sum_0 = \text{cov}(Y_i, Y_i)$
  - raw covariance $G_i(T_{is}, T_{it}) = (Y_{is} - \mu(T_{is}))(Y_{it} - \mu(T_{it}))$
  - Smooth $G_{ij}(T_{is}, T_{jt})$ using local linear smoother and get

University of Glasgow

Moment based method

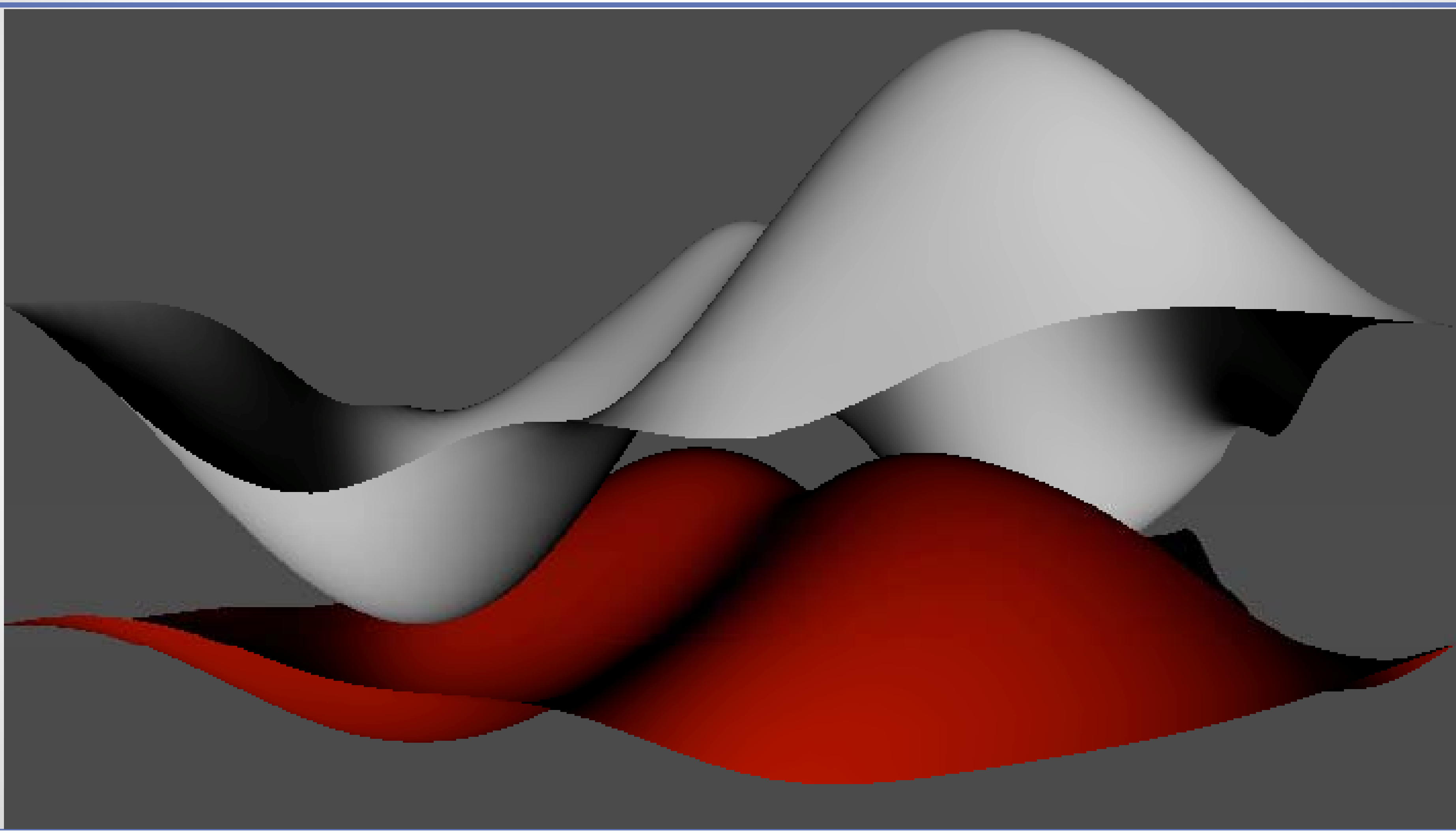
- Smooth lag-d covariance to get estimate of $\Sigma_{d(i,j)} = \text{cov}(Y_i, Y_j)$

  where $d(i,j) = d$

  - raw covariance $G_{ij}(T_{is}, T_{jt}) = (Y_{is} - \mu(T_{is}))(Y_{jt} - \mu(T_{jt}))$

  - Smooth $G_{ij}(T_{is}, T_{jt})$ using local linear smoother and get $\hat{\Sigma}_{d(i,j)}$

Moment based method



University of Glasgow

Moment based method

How to use the lagged covariance surfaces

Calculate Eigenvalue ratio

eigenvalues of $\hat{\Sigma}_0$: $\pi_{0,k}, k = 1, 2 \ldots K$

eigenvalues of $\hat{\Sigma}_d$: $\pi_{d,k}, k = 1, 2 \ldots K$

For given k, we have $\hat{cor}_d(\gamma_{ik}, \gamma_{jk}) = \dfrac{\pi_{d,k}}{\pi_{0,k}}$ for d = 1,2,…D

These $\dfrac{\pi_{d,k}}{\pi_{0,k}}$ can be used to fit model parameters

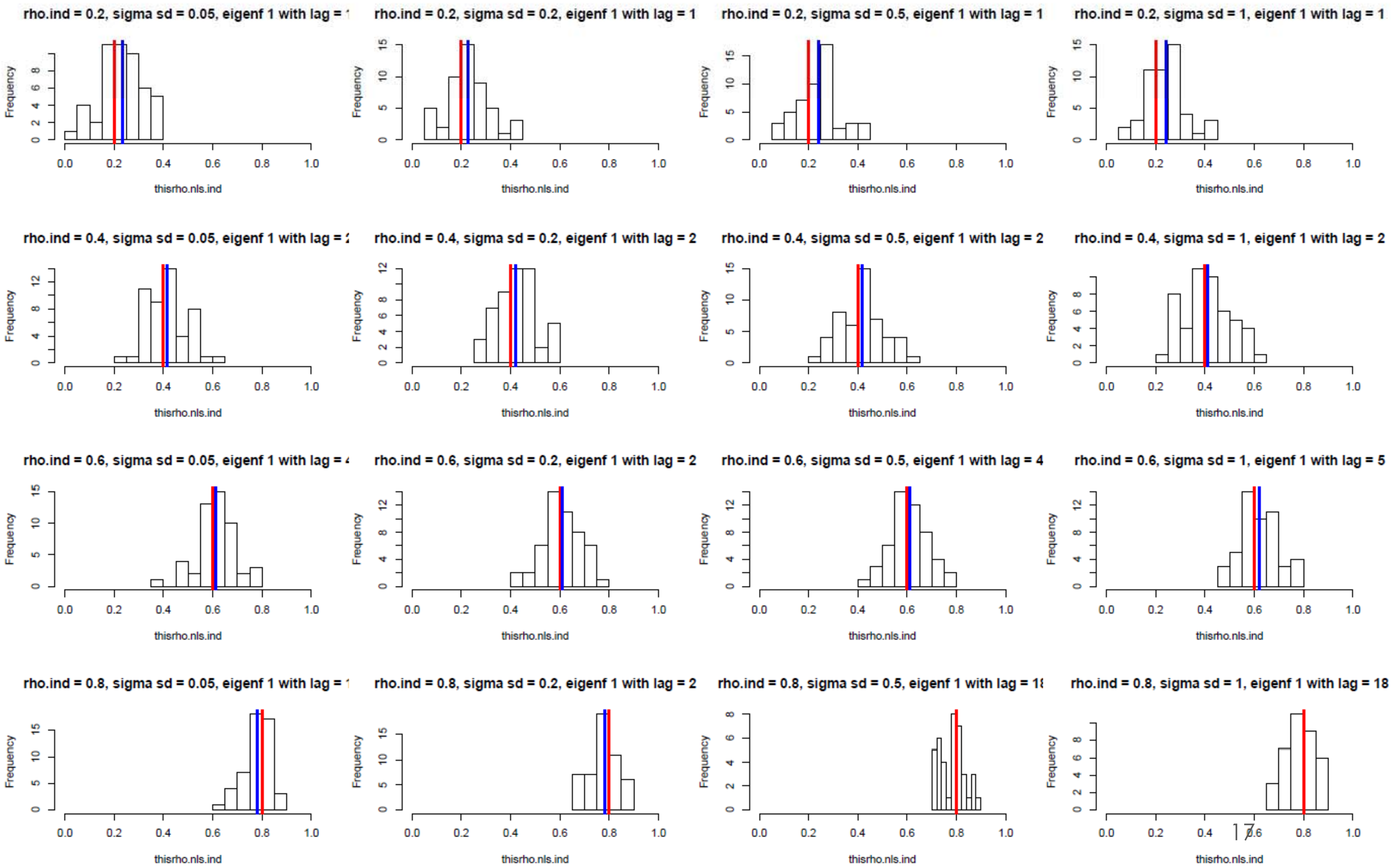
Simulation Results for moment based method

Simulation scheme

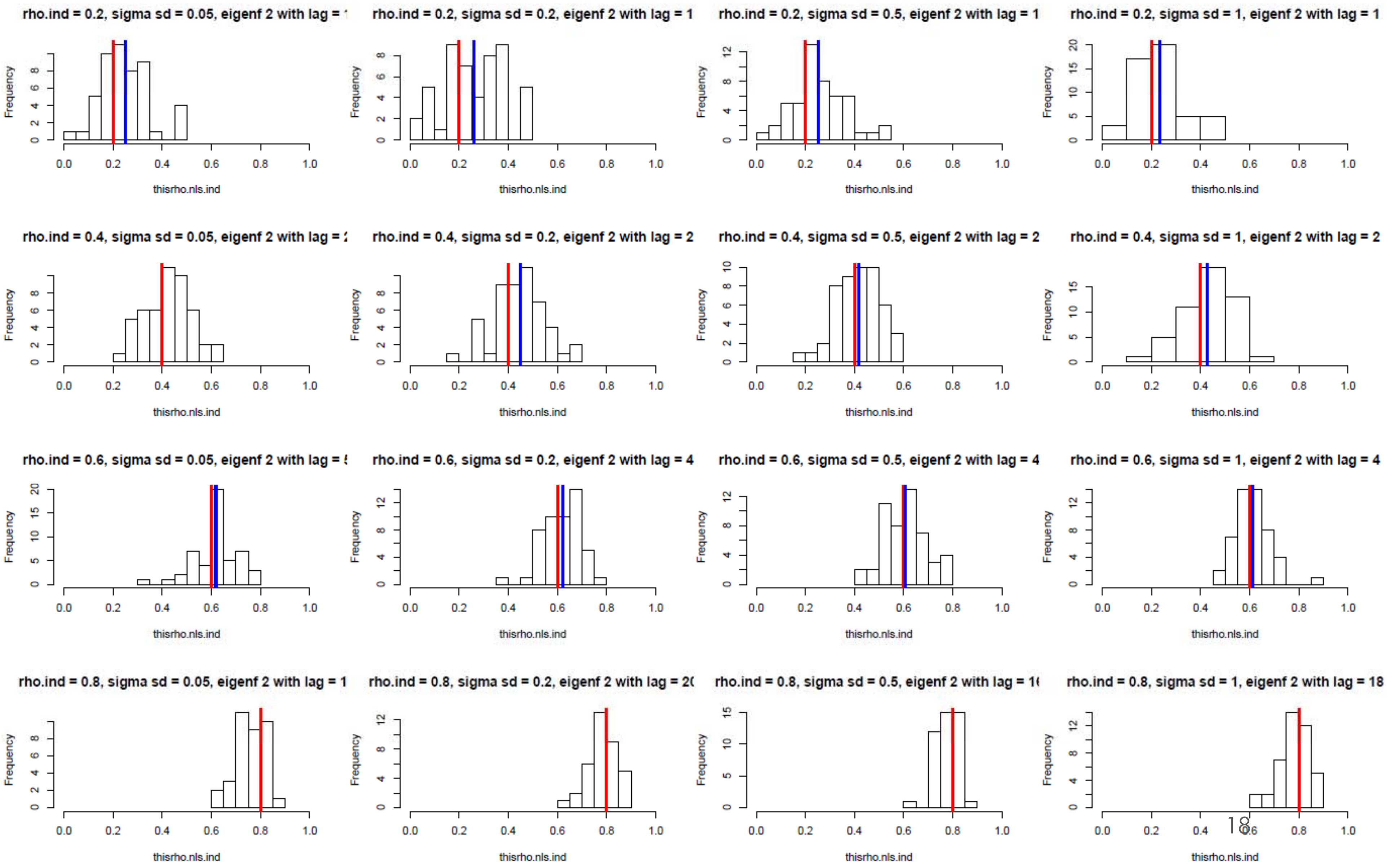
- 3 eigenfunctions, 100 curves, 10 time points.
- AR(1) type spatial correlation(parameter is ρ) =0.2, 0.4, 0.6, 0.8.
- noise standard deviation σ = 0.05, 0.2, 0.5, 1.
- Results for estimation of ρ eigenfunction 1 to 3 with moment based method

University of Glasgow

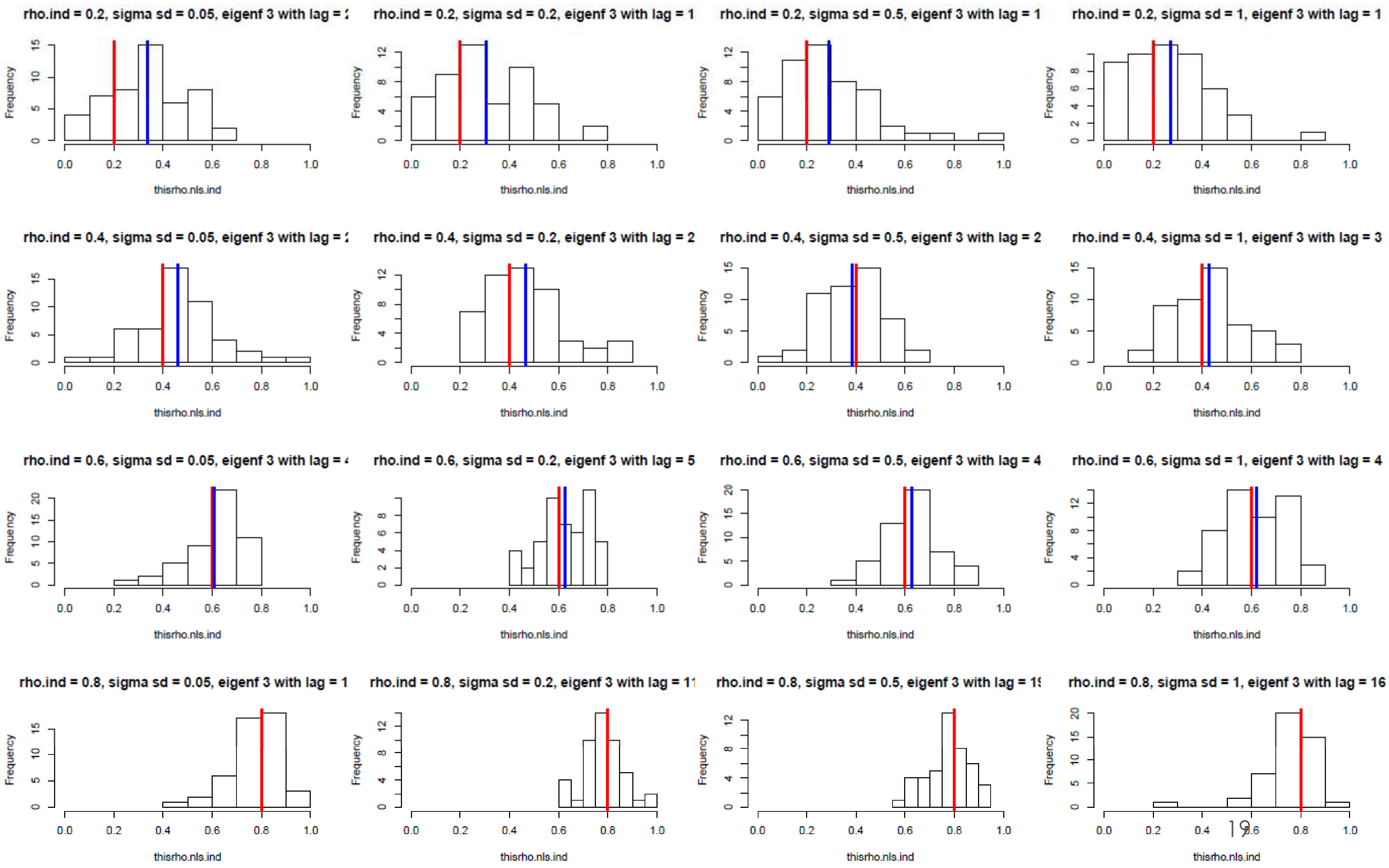
Simulation Results: Ratio of first eigen values



Simulation Results: Ratio of 2nd eigen values



Simulation Results: Ratio of 3rd eigen values



Likelihood Method

- Spatial correlation introduced through random effects
- Random effects are zero mean and satisfy the following

$$\mathrm{cov}(\gamma_{ip}, \gamma_{jq}) = \begin{cases} 0, & p \neq q \\ \rho^{|i-j|}\lambda_k, & p = q = k \end{cases}$$

Then

$$\mathrm{cov}(\tilde{\gamma}) = \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{pmatrix} \otimes \Lambda$$

where

$$\Lambda = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_K \end{pmatrix}$$

Likelihood Method

We express the eigenfunctions $\{\phi_k(t)\}_{k=1}^K$ by basis expansion. Let $\{\mathbf{B}_i\}_{i=1}^n$ be the evaluation matrix of basis and $\mathbf{\Theta}$ be the coefficient matrix of $\{\phi_k(t)\}_{k=1}^K$ on basis functions. Then $\mathbf{\Phi}_i = \mathbf{B}_i\mathbf{\Theta}$ Hence, the full model is expressed as

$$\begin{pmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_n \end{pmatrix} = \begin{pmatrix} \mathbf{B}_1\mathbf{\Theta} & & \\ & \ddots & \\ & & \mathbf{B}_n\mathbf{\Theta} \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{B}_1 & & \\ & \ddots & \\ & & \mathbf{B}_n \end{pmatrix} \begin{pmatrix} \mathbf{\Theta} & & \\ & \ddots & \\ & & \mathbf{\Theta} \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

where

- $\mathbf{Y}_i \to n_i \times 1$ vector of observed values for curve i.
- $\boldsymbol{\epsilon}_i \to n_i \times 1$ vector of measurement errors for curve i.
- $\mathbf{\Phi}_i \to n_i \times K$ matrix of eigenfunction evaluation on curve i.
- $\gamma_i \to K \times 1$ vector of random effects of curve i
- $\mathbf{B}_i \to n_i \times P$ matrix of P basis functions evaluation.
- $\mathbf{\Theta} \to P \times K$ matrix of coefficient matrix of K eigenfunctions on P basis functions.

The compact format becomes,

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{B}}\tilde{\mathbf{\Theta}}\tilde{\gamma} + \tilde{\epsilon}$$

Likelihood Steps

Model:

$$\tilde{\mathbf{Y}} \sim \mathcal{N}(\mathbf{0}, \tilde{\mathbf{B}}\tilde{\mathbf{\Theta}}\mathbf{\Gamma}\tilde{\mathbf{\Theta}}'\tilde{\mathbf{B}}' + \sigma^2\mathbf{I})$$

Marginal Likelihood

$$L(\tilde{\mathbf{Y}}) = (2\pi)^{-N/2}|\tilde{\mathbf{B}}\tilde{\mathbf{\Theta}}\mathbf{\Gamma}\tilde{\mathbf{\Theta}}'\tilde{\mathbf{B}}' + \sigma^2\mathbf{I}|^{-1/2}\exp(-\frac{1}{2}\tilde{\mathbf{Y}}'(\tilde{\mathbf{B}}\tilde{\mathbf{\Theta}}\mathbf{\Gamma}\tilde{\mathbf{\Theta}}'\tilde{\mathbf{B}}' + \sigma^2\mathbf{I})^{-1}\tilde{\mathbf{Y}})$$

- Direct maximizing of this likelihood over $\varrho, \sigma^2, \Lambda$ and Θ is a difficult non-convex optimization problem.
- Solution: EM procedure by treating random effects as missing data and focus on the joint log likelihood

Joint log likelihood

$$\begin{aligned}\mathcal{L}(\tilde{\mathbf{Y}}, \tilde{\gamma}) &= C - N\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}(\mathbf{Y}_i - \mathbf{B}_i\mathbf{\Theta}\boldsymbol{\gamma}_i)'(\mathbf{Y}_i - \mathbf{B}_i\mathbf{\Theta}\boldsymbol{\gamma}_i) \\ &\quad -\frac{1}{2}\left(K\log D(\rho) + n\sum_{k=1}^{K}\log\lambda_k + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{K} g_{ij}(\rho)\lambda_k^{-1}\gamma_{ik}\gamma_{jk}\right)\end{aligned}$$

University of Glasgow

EM Step

E-Step:

It is easy to show that $\mathbb{E}(\mathcal{L}(\mathbf{Y}, \tilde{\gamma})|\mathbf{Y}, \boldsymbol{\Delta})$ depends on $\boldsymbol{\gamma}_i$ only through $\{\widehat{\gamma}_i = \mathbb{E}(\boldsymbol{\gamma}_i|\mathbf{Y}, \boldsymbol{\Delta})\}_{i=1}^n$, $\{\widehat{\gamma_{ik}\gamma_{jk}} = \mathbb{E}(\gamma_{ik}\gamma_{jk}|\tilde{\mathbf{Y}}, \boldsymbol{\Delta})\}_{k=1,i\neq j=1}^{K,n}$ and $\{\widehat{\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i'} = \mathbb{E}(\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i'|\tilde{\mathbf{Y}}, \boldsymbol{\Delta})\}_{i=1}^n$. Note that for each i, $\widehat{\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i'} = \widehat{\gamma}_i\widehat{\gamma}_i{}' + \mathbb{V}(\boldsymbol{\gamma}_i|\tilde{\mathbf{Y}}, \boldsymbol{\Delta})$

M-Step

M-step is to maximize $\mathbb{E}(\mathcal{L}(\tilde{\mathbf{Y}}, \tilde{\gamma})|\tilde{\mathbf{Y}}, \boldsymbol{\Delta})$ over $\boldsymbol{\Lambda}$, $\boldsymbol{\Theta}$, σ^2 and ρ.

- Note that Λ and ϱ are separated from Θ and σ^2
- optimization over these parameters one at a time and do it iteratively.

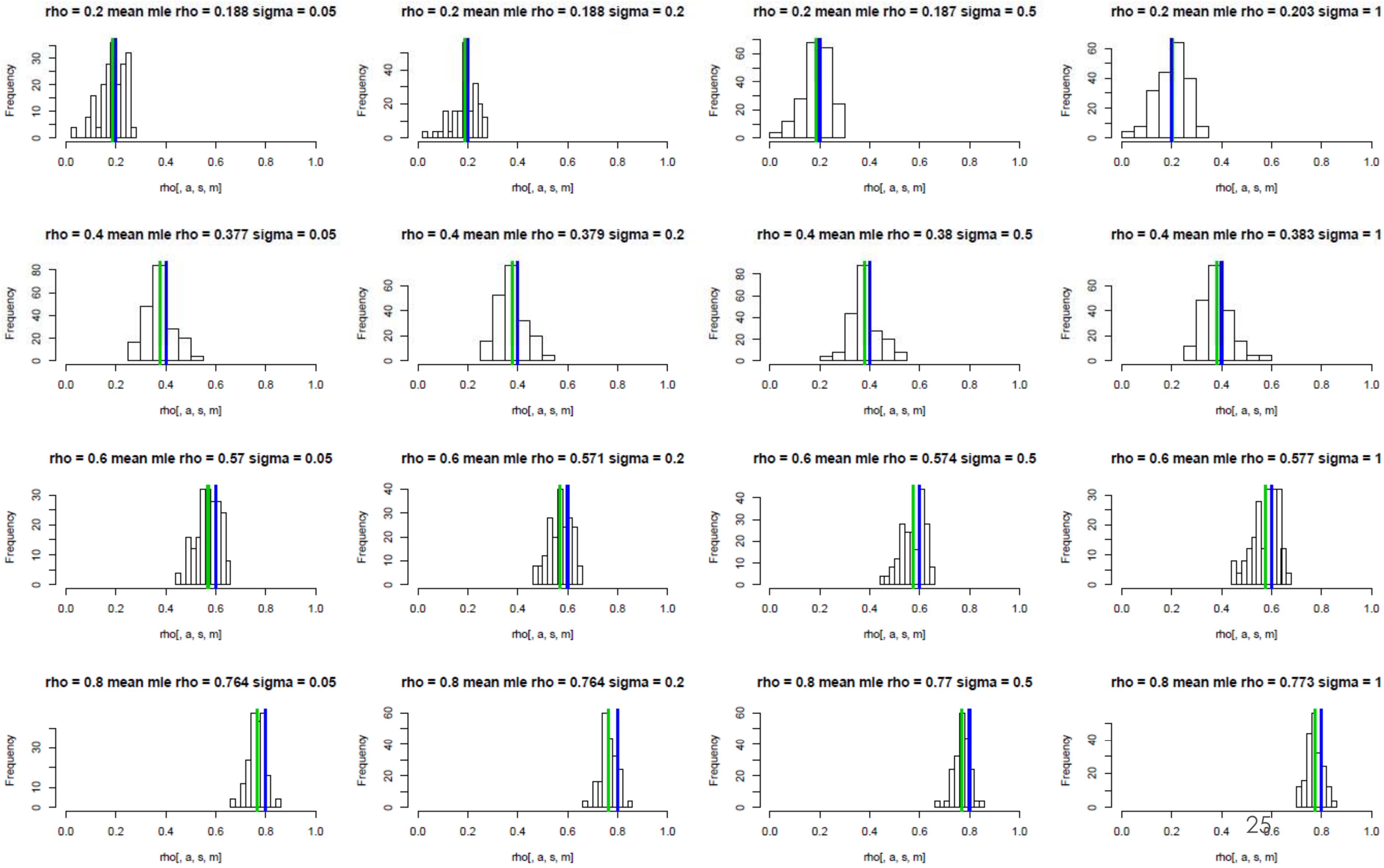
University of Glasgow

Simulation Results for moment based method

Simulation scheme

- 3 eigenfunctions, 100 curves, 10 time points.
- AR(1) type spatial correlation(parameter is ϱ) =0.2, 0.4, 0.6, 0.8.
- noise standard deviation σ = 0.05, 0.2, 0.5, 1.
- Results for estimation of ϱ eigenfunction 1 to 3 with moment based method

EM estimates of correlation parameter



Reconstruction Results

- Given the estimate of spatial correlation structure, we can compute the expected principal component scores γ_{ik}

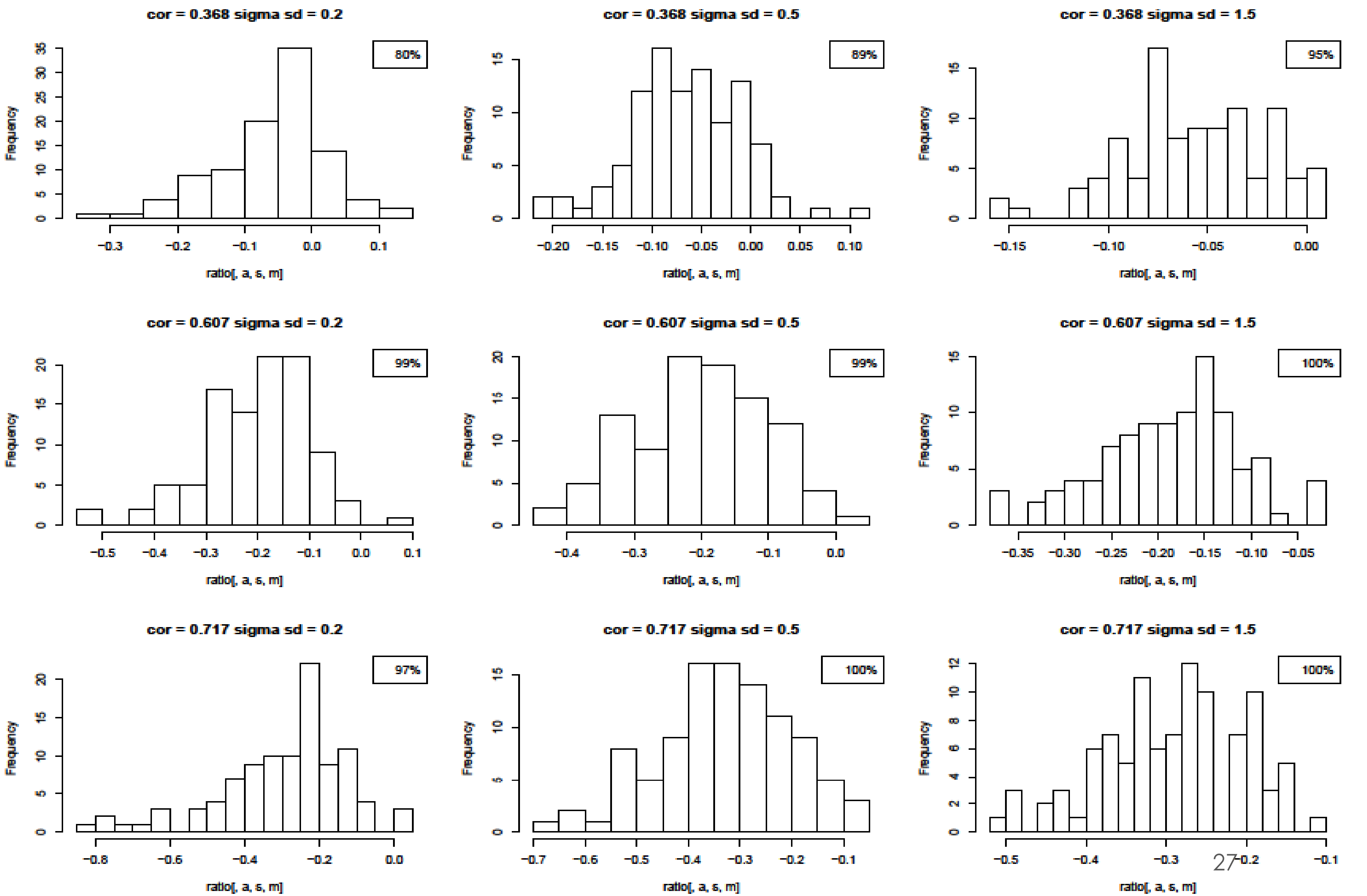
$$\hat{\hat{\gamma}}_{ik} = E(\gamma_{ik} \mid Y_1, Y_2, \ldots, Y_N)$$

- In the i.i.d curve case, the information set only include curve i,

$$\hat{\gamma}_{ik} = E(\gamma_{ik} \mid Y_i)$$

- For AR(1) correlated curves, curves were reconstructed using using $\hat{\gamma}_{ik}$ and $\hat{\hat{\gamma}}_{ik}$
- Performance measured by sum of squared errors over all curves.
- Negative log ratio suggest better performance

Histogram of log ratio of squared error



Future Work

Consistency

- For moment based estimates

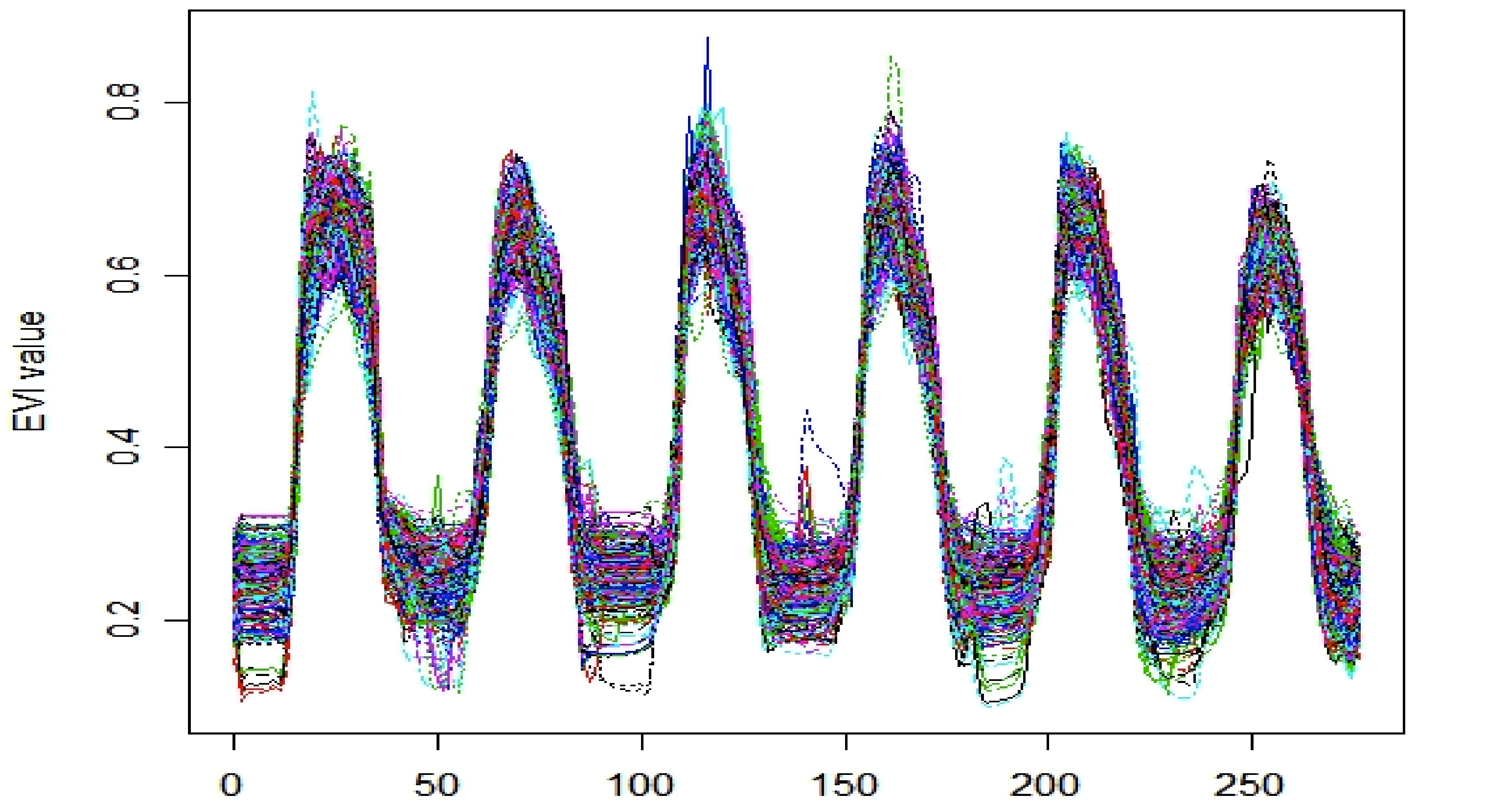
Strict geometric constraints for solving the EM steps.

- Hypothesize better convergence results from this approach.
- Following Peng and Paul's , 2009 paper on "A geometric approach to maximum likelihood estimation of the functional principal components from sparse longitudinal data"

Software:

- SPACE (Spatial Principal Analysis by Conditional Estimation)
- First version with limited spatial correlation choices.
- Follow up versions with more general spatial correlation.

Future Work



Acknowledgement

Collaborators:

- Giles Hooker, Cornell University
- Mark Friedl, Boston University
- Chong Liu, Boston University (Doctoral Students)

Grants:

- NSF Award No: #0934739: Functional Data Modeling of Climate-Ecosystem Dynamics 09/01/09- 08/31/13
- NSF Award No: #0947950: GLACIER-Global Change Initiative-Education & Research. 03/15/2010-03/14/2014

Related paper:

- Liu, C., Ray, S., Hooker, G., Friedl, M.F. (2012) Functional Factor Analysis For Periodic Remote Sensing Data. *Annals of Applied Statistics*, **6**:2, 601-624.