Multivariate data

ST733 – Spatial Statistics

Multivariate Data

- Say at each location there are q responses $Y_1(\mathbf{s}), ..., Y_q(\mathbf{s})$

- Examples:

Multivariate Data

- Advantages of a joint model of the q responses:

- Can you fit a joint model even if the responses are never measured at the same locations?

Multivariate Data

- Cross-covariogram
- Conditional models
- Separability
- Multivariate Matérn
- Linear model of coregionalization
- Spatially-varying coefficients

Cross-covariogram

- The covariogram is use to examine cross-dependence
- The cross-covarigram is defined as

$$\gamma_{jk}(\mathbf{h}) = \mathrm{E}\left\{[Y_j(\mathbf{s}+\mathbf{h}) - Y_j(\mathbf{s})][Y_k(\mathbf{s}+\mathbf{h}) - Y_k(\mathbf{s})]\right\}$$

- Empirical versions are defined in the obvious way

Conditional models

- Joint models can always be broken into a sequence of conditional models:

- If the data are measured at different locations this can be complicated

Separable model

- The separable model is

$$\text{Cov}(Y_j(\mathbf{s}), Y_k(\mathbf{s}')) = \Sigma_{jk} C(\mathbf{s}, \mathbf{s}')$$

  for $q \times q$ covariance matrix Σ and spatial correlation function C

- This implies:

- Kronecker products lead to fast computation

Separable model

The separable model implies:

- Variance: $\mathrm{V}(Y_j(\mathbf{s})) =$

- Cross covariance: $\mathrm{Cov}(Y_j(\mathbf{s},)Y_k(\mathbf{s})) =$

- Spatial covariance: $\mathrm{Cov}(Y_j(\mathbf{s},)Y_j(\mathbf{s}')) =$

Separable model

- Advantages:

- Disadvantages:

Multivariate Matérn

The multivariate Matérn covariance function[1] is

$$\text{Cov}(Y_j(\mathbf{s}), Y_k(\mathbf{s}')) = \sigma_j \sigma_k \rho_{jk} M(||\mathbf{s} - \mathbf{s}'||, \phi_{jk}, \nu_{jk})$$

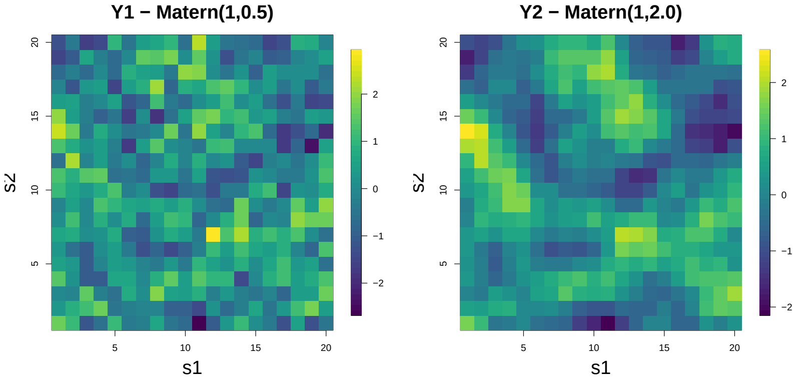
- σ_j^2 is a variance
- $\{\rho_{jk}\}$ is a $q \times q$ cross-correlation matrix
- ϕ_{jk} and ν_{jk} define the spatial correlation

[1] Gneiting et al, 2010

Multivariate Matérn

- Restrictions are needed on the covariance parameters to give a valid covariance function
- Examples/special cases:
  - Separable: $\phi_{jk} \equiv \phi$ and $\nu_{jk} \equiv \nu$
  - Parsimonious: $\phi_{jk} \equiv \phi$, $\nu_{jk} = (\nu_j + \nu_k)/2$, and ρ_{jk} is bounded
- Intuitively, why are restrictions needed?

Parsimonious Matérn with $\rho_{12} = 0.8$



Multivariate Matérn

- Advantages:

- Disadvantages:

Linear model of coregionalization (LMC

- The LCM is basically spatial factor analysis
- Example: the responses are q pollutants,
  - $Y_1(\mathbf{s})$ is ozone
  - ...
  - $Y_q(\mathbf{s})$ is particulate matter
- The q pollutants as driven by L latent sources,
  - $f_1(\mathbf{s})$ is emission from cars
  - ...
  - $f_L(\mathbf{s})$ is emission from power plants

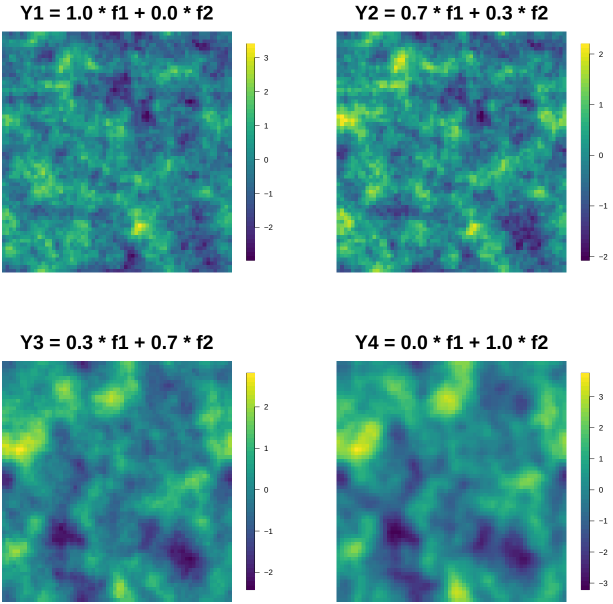
Linear model of coregionalization

- The factor analysis model is:

- Taking these together gives the model:

- Identification:

LMC with $q = 4$ and $L = 2$



Linear model of coregionalization

The sample cross-covariance is

	Y_1	Y_2	Y_3	Y_4
Y_1	1.00	0.86	0.19	-0.14
Y_2	0.86	1.00	0.66	0.39
Y_3	0.19	0.66	1.00	0.95
Y_4	-0.14	0.39	0.95	1.00

Linear model of coregionalization

- Separable model: If the latent processes f_j are iid GPs with mean zero, variance 1 and correlation $C(||\mathbf{s} - \mathbf{s}'||; \Theta)$, then the covariance of Y is:

- Non-separable models: If the latent processes f_j are independent with mean zero and variance one, but latent process j has $C(||\mathbf{s} - \mathbf{s}'||; \Theta_j)$, then the covariance of Y is:

Linear model of coregionalization

- Advantages:

- Disadvantages:

Spectral methods

- In the spectral domain,

$$Y_j(\mathbf{s}) = \int \exp(i\mathbf{s}^T\omega)Z_j(\omega)d\omega$$

  where Z_j is WN with $\text{Var}(Z_j(\omega)) = f_j(\omega)$

- A MV model for $\mathbf{Z}(\omega) = [Z_1(\omega), ..., Z_q(\omega)]^T$ implies a MV model for Y

- Example: $\text{Cov}[\mathbf{Z}(\omega)] = \Sigma(\omega)$ measures dependence a frequency ω

Spectral methods

- Coherence:
- Separable model:
- Non-separable model

Spatially-varying coefficients (SVCs)

- MV spatial processes are also used for SVCs for a single outcome variable
- The model is

$$Y(\mathbf{s}) = \beta_0(\mathbf{s}) + \sum_{j=1}^{p} X_j(\mathbf{s})\beta_j(\mathbf{s}) + \varepsilon(\mathbf{s})$$

- A MV spatial prior can be used for $\boldsymbol{\beta}(\mathbf{s}) = [\beta_0(\mathbf{s}), ..., \beta_p(\mathbf{s})]$
- Why might the covariate effects change with space?