Lecture 6 Multivariate Spatial Modeling

Shiwei Lan[1]

[1]School of Mathematical and Statistical Sciences
Arizona State University

STP598 Spatiotemporal Analysis
Fall 2021

Multivariate spatial modeling

- *Multivariate*: multiple (i.e. more than one) outcomes are measured at each spatial unit.
- Multivariate point-referenced data:
  - Levels of pollutants including ozone, nitric oxide, carbon monoxide, $PM_{2.5}$ etc. are measured at monitoring station
  - Surface temperature, precipitation, and wind speed in atmospheric modeling.
  - In examining real estate markets, both selling price and total rental income observed for individual property...
- Multivariate areal data:
  - In public health, supplies counts or rates for a number of diseases for each county or administrative unit.

Table of Contents

Multivariate spatial modeling for point-referenced data

- We model multivariate point-referenced data by either a *conditioning approach (kriging with external drift)* or a *joint approach (co-kriging)*.
- Inference focuses upon three major aspects:
  1. estimate associations among the processes
  2. estimate the strength of spatial association for each process
  3. predict the processes at arbitrary locations
- Let $\mathbf{Y}(\mathbf{s}) = (Y_1(\mathbf{s}), \cdots, Y_p(\mathbf{s}))^T$ be a $p \times 1$ vector of process referenced at $\mathbf{s} \in \mathcal{D}$.
- We seek to capture the association both within components of $\mathbf{Y}(\mathbf{s})$ and across $\mathbf{s}$.

Cross- variograms and covariances

- Assume $\mathrm{E}(Y(\mathbf{s}+\mathbf{h}) - Y(\mathbf{s})) = 0$. The joint second order (weak) stationarity hypothesis defines the *cross-variogram* as

$$\gamma_{ij}(\mathbf{h}) = \frac{1}{2}\mathrm{E}(Y_i(\mathbf{s}+\mathbf{h}) - Y_i(\mathbf{s}))(Y_j(\mathbf{s}+\mathbf{h}) - Y_j(\mathbf{s})) \tag{1}$$

  - $\gamma_{ij}(\mathbf{h}) = \gamma_{ij}(-\mathbf{h})$.
  - $|\gamma_{ij}(\mathbf{h})|^2 \leq \gamma_{ii}(\mathbf{h})\gamma_{jj}(\mathbf{h})$.
- The *cross-covariance* function is defined as

$$C_{ij}(\mathbf{h}) = \mathrm{E}(Y_i(\mathbf{s}+\mathbf{h}) - \mu_i)(Y_j(\mathbf{s}) - \mu_j) \tag{2}$$

  - $C_{ij}(\mathbf{h}) \neq C_{ji}(\mathbf{h})$.
  - $|C_{ij}(\mathbf{h})|^2 \leq C_{ii}(0)C_{jj}(0)$. $\quad |C_{ij}(\mathbf{h})|^2 \leq C_{ii}(\mathbf{h})C_{jj}(\mathbf{h})$?
- Eg: spatial delay models (Wackernagel, 2003): $Y_2(\mathbf{s}) = aY_1(\mathbf{s}+\mathbf{h}_0) + \epsilon(\mathbf{s})$.

Cross- variograms and covariances

- How to express $\gamma_{ij}(\mathbf{h})$ in terms of C_{ij}?

Cross- variograms and covariances

- How to express $\gamma_{ij}(\mathbf{h})$ in terms of C_{ij}?

$$\gamma_{ij}(\mathbf{h}) = C_{ij}(\mathbf{0}) - \frac{1}{2}(C_{ij}(\mathbf{h}) + C_{ij}(-\mathbf{h})) \tag{3}$$

- Cross-variogram only captures the even term of the cross-covariance function!
- *Pseudo* cross-variogram:
  - Clark et al. (1989) proposed $\pi_{ij}^c(\mathbf{h}) = \mathrm{E}(Y_i(\mathbf{s}+\mathbf{h}) - Y_j(\mathbf{s}))^2$
  - Myers (1991) defined $\pi_{ij}^m(\mathbf{h}) = \mathrm{Var}(Y_i(\mathbf{s}+\mathbf{h}) - Y_j(\mathbf{s}))$
  - $\pi_{ij}^c(\mathbf{h}) = \pi_{ij}^m(\mathbf{h}) + (\mu_i - \mu_j)^2$
- Positive, may not be even. Co-kriging uses $\pi_{ij}^m(\mathbf{h})$.

Co-kriging

- Given $\mathbf{Y} = (\mathbf{Y}(\mathbf{s}_1), \cdots, \mathbf{Y}(\mathbf{s}_n))^T$, we want to know $\mathbf{Y}(\mathbf{s}_0)$.
- Different from multi-output kriging for a univariate spatial process at multiple locations!
- In the regression framework, we could require the predicted value $\hat{\mathbf{Y}}(\mathbf{s}_0)$

$$\hat{\mathbf{Y}}(\mathbf{s}_0) = \sum_{i=1}^n \Lambda_i \mathbf{Y}(\mathbf{s}_i), \quad \sum_{i=1}^n \Lambda_i = I \tag{4}$$

$$\min_{\Lambda} \; \mathrm{tr}\mathrm{E}(\hat{\mathbf{Y}}(\mathbf{s}_0) - \mathbf{Y}(\mathbf{s}_0))(\hat{\mathbf{Y}}(\mathbf{s}_0) - \mathbf{Y}(\mathbf{s}_0))^T \tag{5}$$

- $\mathrm{tr}\mathrm{E}(\hat{\mathbf{Y}}(\mathbf{s}_0) - \mathbf{Y}(\mathbf{s}_0))(\hat{\mathbf{Y}}(\mathbf{s}_0) - \mathbf{Y}(\mathbf{s}_0))^T = \mathrm{E}(\hat{\mathbf{Y}}(\mathbf{s}_0) - \mathbf{Y}(\mathbf{s}_0))^T(\hat{\mathbf{Y}}(\mathbf{s}_0) - \mathbf{Y}(\mathbf{s}_0))$.

Co-kriging

- Assume a multivariate Gaussian spatial process $\mathbf{Y}(\mathbf{s})$ with zero mean.
- Suppose we have a finite cross-covariance function (*permissible* cross-variogram).
- Denote $\mathbf{Y} = (\mathbf{Y}(\mathbf{s}_1)^T, \cdots, \mathbf{Y}(\mathbf{s}_n)^T)^T$. Then we have $np \times np$ covariance matrix $\Sigma_{\mathbf{Y}}$.
- Denote $np \times 1$ vector $\mathbf{c}_0$ with jl-th element $c_{0j,l} = \mathrm{Cov}(Y_1(\mathbf{s}_0), Y_l(\mathbf{s}_j))$. Then

$$\mathrm{E}(Y_1(\mathbf{s}_0)|\mathbf{Y}) = \mathbf{c}_0^T \Sigma_{\mathbf{Y}}^{-1} \mathbf{Y} \tag{6}$$

$$\mathrm{Var}(Y_1(\mathbf{s}_0)|\mathbf{Y}) = C_{11}(\mathbf{0}) - \mathbf{c}_0^T \Sigma_{\mathbf{Y}}^{-1} \mathbf{c}_0 \tag{7}$$

Co-kriging

- *Intrinsic* co-kriging assumes $C(\mathbf{h}) = \rho(\mathbf{h})T$ with a valid correlation function $\rho(\cdot)$ and a positive definite covariance matrix T.
- Therefore $\Sigma_{\mathbf{Y}} = R \otimes T$, and

$$\mathrm{E}(Y_1(\mathbf{s}_0)|\mathbf{Y}) = \mathbf{c}_0^T \Sigma_{\mathbf{Y}}^{-1} \mathbf{Y} = t_{11} \mathbf{r}_0^T R^{-1} \tilde{\mathbf{Y}}_1 \tag{8}$$

  where $\mathbf{r}_0 = (\rho(\mathbf{s}_0 - \mathbf{s}_j))$ and $\tilde{\mathbf{Y}}_1$ is formed by the first components of $\mathbf{Y}(\mathbf{s}_j)$'s.
- Data availability (missing data):
  - *isotopy*: data is available for each variable at all sampling points
  - partial *heterotopy*: some variables share some sample locations
  - entirely *heterotopic*: the variables have no sample locations in common
- *Collocated* co-kriging makes use of $Y_l(\mathbf{s}_j)$ to help predict $Y_1(\mathbf{s}_0)$.

Cross-covariance function

- Consider a vector-valued spatial process $\{\mathbf{w}(\mathbf{s}) \in \mathbb{R}^p : \mathbf{s} \in \mathcal{D}\}$. Assume $\mathrm{E}[\mathbf{w}(\mathbf{s})] = \mathbf{0}$.
- The *cross-covariance function* is a matrix-valued function $\mathbf{C}(\mathbf{s}, \mathbf{s}')$ with (i, j)-th entry

$$C_{ij}(\mathbf{s}, \mathbf{s}') = \mathrm{Cov}(w_i(\mathbf{s}), w_j(\mathbf{s}')) = \mathrm{E}[w_i(\mathbf{s}) w_j(\mathbf{s}')] \tag{9}$$

- Let $w_i(\mathbf{s}) = Y_i(\mathbf{s}) - \mu_i$. Then $C(\mathbf{s}, \mathbf{s}') = \mathrm{Cov}(\mathbf{w}(\mathbf{s}), \mathbf{w}(\mathbf{s}')) = \mathrm{E}[\mathbf{w}(\mathbf{s})\mathbf{w}(\mathbf{s}')^T]$.
- We require $C(\mathbf{s}, \mathbf{s}') = C(\mathbf{s}', \mathbf{s})^T$.
- $\mathbf{w}(\mathbf{s})$ is *stationary* if $C(\mathbf{s}, \mathbf{s}') = C(\mathbf{h})$ is a function of $\mathbf{h} = \mathbf{s} - \mathbf{s}'$. Symmetric cross-covariance implies $C(-\mathbf{h}) = C(\mathbf{h})$.
- $\mathbf{w}(\mathbf{s})$ is *isotropic* if further $C(\mathbf{s}, \mathbf{s}') = C(\|\mathbf{h}\|)$, which directly implies symmetry in cross-covariance function.

Separable models

- Separable models for p-dimensional $\mathbf{Y}(\mathbf{s})$ assume the following cross-covariance function

$$C(\mathbf{s},\mathbf{s}') = \rho(\mathbf{s},\mathbf{s}') \cdot T \tag{10}$$

- The covariance matrix for $\mathbf{Y}$ has the following Kronecker product structure

$$\Sigma_{\mathbf{Y}} = H \otimes T \tag{11}$$

  where $H_{ij} = \rho(\mathbf{s}_i, \mathbf{s}_j)$.
- Pros: $|\Sigma_{\mathbf{Y}}| = |H|^p \cdot |T|^n$, $\Sigma_{\mathbf{Y}}^{-1} = H^{-1} \otimes T^{-1}$.
- Cons: *coherence* $\frac{\mathrm{Cov}(Y_\ell(\mathbf{s}), Y_{\ell'}(\mathbf{s}+\mathbf{h}))}{\sqrt{\mathrm{Cov}(Y_\ell(\mathbf{s}), Y_\ell(\mathbf{s}+\mathbf{h}))\mathrm{Cov}(Y_{\ell'}(\mathbf{s}), Y_{\ell'}(\mathbf{s}+\mathbf{h}))}} = \frac{T_{\ell\ell'}}{T_{\ell\ell}T_{\ell'\ell'}}$ regardless of $\mathbf{s}$ and $\mathbf{h}$: no spatial variation of dependences among components of $\mathbf{Y}(\mathbf{s})$!

Interpolation, (spatial) prediction and regression

- Consider response process $Z(\mathbf{s})$ and a vector of covariates $\mathbf{x}(\mathbf{s})$.
- Partition our set of sites into three mutually disjoint groups
  1. S_Z: the sites where only the response $Z(\mathbf{s})$ has been observed
  2. S_X: the the set of sites where only the covariates have been observed
  3. S_{ZX}: the set where both $Z(\mathbf{s})$ and the covariates have been observed
  4. S_U: the set of sites where no observations have been taken.
- Formalize three types of inference questions:
  1. *interpolation*: concerns $Y(\mathbf{s})$ when $\mathbf{s} \in S_X$
  2. *prediction*: concerns $Y(\mathbf{s})$ when $\mathbf{s} \in S_U$
  3. *spatial regression*: concerns the functional relationship between $X(\mathbf{s})$ and $Y(\mathbf{s})$ at an arbitrary site $\mathbf{s}$, along with other covariate information $\mathbf{U}(\mathbf{s})$, $\mathrm{E}[Y(\mathbf{s})|X(\mathbf{s}), \mathbf{U}(\mathbf{s})]$.

Interpolation, (spatial) prediction and regression

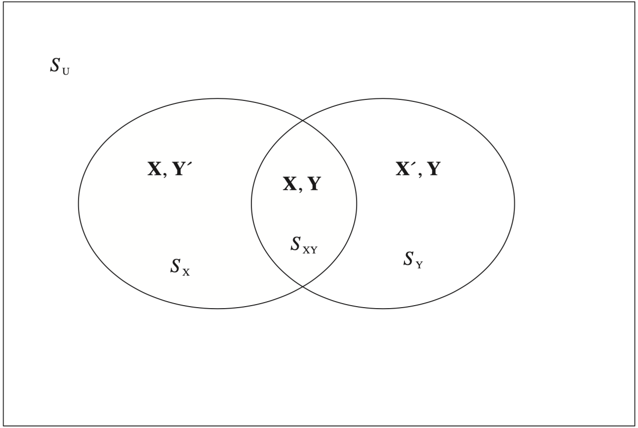

Figure 9.1 *A graphical representation of the S sets. Interpolation applies to locations in* S_X*, prediction applies to locations in* S_U*, and regression applies to all locations.* $\mathbf{X}_{aug} = (\mathbf{X}, \mathbf{X}')$, $\mathbf{Y}_{aug} = (\mathbf{Y}, \mathbf{Y}')$.

Regression in the Gaussian case

- Learn about the conditional distribution for $Y(\mathbf{s}_0)|X(\mathbf{s}_0)$.
- Considering a bivariate Gaussian spatial process $\mathbf{W}(\mathbf{s}) = (X(\mathbf{s}), Y(\mathbf{s}))^T$ with mean $\boldsymbol{\mu}(\mathbf{s}) = (\mu_X(\mathbf{s}), \mu_Y(\mathbf{s}))^T$ and a separable cross-covariance function, we have

$$\mathbf{W}(\mathbf{s}) = (X(\mathbf{s}), Y(\mathbf{s}))^T \sim N(\boldsymbol{\mu}(\mathbf{s}), T) \tag{12}$$

- For simplicity, suppose $\boldsymbol{\mu}(\mathbf{s}) = (\mu_1, \mu_2)^T$. We have the conditional

$$p(y(\mathbf{s})|x(\mathbf{s}), \beta_0, \beta_1, \sigma^2) = N(\beta_0 + \beta_1 x(\mathbf{s}), \sigma^2) \tag{13}$$

$$\beta_0 = \mu_2 - \frac{T_{12}}{T_{11}}\mu_1, \quad \beta_1 = \frac{T_{12}}{T_{11}}, \quad \sigma^2 = T_{22} - \frac{T_{12}^2}{T_{11}} \tag{14}$$

- Therefore, *regression*: $\mathrm{E}[Y(\mathbf{s})|x(\mathbf{s})] = \beta_0 + \beta_1 x(\mathbf{s})$.

Interpolation/Prediction in the Gaussian case

- Now let $\mathbf{s}_0$ be a new site where we want to make prediction.
- We have

$$\mathbf{W}^* = (\mathbf{W}(\mathbf{s}_0), \cdots, \mathbf{W}(\mathbf{s}_n))^T \sim N\left(\mathbf{1}_{n+1} \otimes \boldsymbol{\mu}, H^*(\phi) \otimes T\right) \tag{15}$$

where $H^*(\phi) = \begin{bmatrix} H(\phi) & \mathbf{h}(\phi) \\ \mathbf{h}(\phi)^T & \rho(0;\phi) \end{bmatrix}$, and $\mathbf{h}(\phi) = (\rho(\mathbf{s}_0 - \mathbf{s}_j; \phi))$.

- For *interpolation*: $x(\mathbf{s}_0)$ is observed, we obtain

$$p(y(\mathbf{s}_0)|x(\mathbf{s}_0), \mathbf{y}, \mathbf{x}) = \int p(y(\mathbf{s}_0)|x(\mathbf{s}_0), \mathbf{y}, \mathbf{x}, \boldsymbol{\mu}, \phi, T) p(\boldsymbol{\mu}, \phi, T|\mathbf{y}, \mathbf{x}) \tag{16}$$

- For *prediction*: $x(\mathbf{s}_0)$ is not observed, we still have

$$p(y(\mathbf{s}_0)|\mathbf{y}, \mathbf{x}) = \int p(y(\mathbf{s}_0)|x(\mathbf{s}_0), \mathbf{y}, \mathbf{x}, \boldsymbol{\mu}, \phi, T) p(\boldsymbol{\mu}, \phi, T, x(\mathbf{s}_0)|\mathbf{y}, \mathbf{x}) \tag{17}$$

Regression in a probit model

- Now suppose we have binary response $Z(\mathbf{s})$ in a point-source spatial dataset.
- Let $Y(\mathbf{s})$ be a latent spatial process such that $Z(\mathbf{s}) = 1$ only if $Y(\mathbf{s}) > 0$. Let $X(\mathbf{s})$ be a process that generate values of a covariate.
- Again we consider a bivariate Gaussian spatial process $\mathbf{W}(\mathbf{s}) = (X(\mathbf{s}), Y(\mathbf{s}))^T$, but where now $\boldsymbol{\mu}(\mathbf{s}) = (\mu_1, \mu_2 + \boldsymbol{\alpha}^T\mathbf{U}(\mathbf{s}))^T$ with $\mathbf{U}(\mathbf{s})$ regarded as a $p \times 1$ vector of fixed covariates.
- We can set $T_{22} = 1$ due to non-identifiability. Thus we formulate a probit regression model

$$P(Z(\mathbf{s}) = 1 | x(\mathbf{s}), \mathbf{U}(\mathbf{s}), \boldsymbol{\alpha}, \boldsymbol{\mu}, T_{11}, T_{12}) =$$
$$\Phi\left([\beta_0 + \beta_1 X(\mathbf{s}) + \boldsymbol{\alpha}^T\mathbf{U}(\mathbf{s})] / \sqrt{1 - T_{12}^2 / T_{11}} \right)$$

where $\beta_0 = \mu_2 - (T_{12}/T_{11})\mu_1$, and $\beta_1 = T_{12}/T_{11}$.

Regression in a probit model

- Now we observe $\mathbf{z} = (z(\mathbf{s}_1), \cdots, z(\mathbf{s}_n))^T$ and $\mathbf{X} = (X(\mathbf{s}_1), \cdots, X(\mathbf{s}_n))^T$, but not $\mathbf{Y} = (Y(\mathbf{s}_1), \cdots, Y(\mathbf{s}_n))^T$. Again we have

$$\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} = N\left(\begin{bmatrix} \mu_1 \mathbf{1} \\ \mu_2 \mathbf{1} + \mathbf{U}\beta \end{bmatrix}, T \otimes H(\phi) \right) \tag{18}$$

- Assuming appropriate hyper-priors, we can obtain posterior samples from $p(\boldsymbol{\mu}, \boldsymbol{\alpha}, T_{11}, T_{12}, \phi | \mathbf{x}, \mathbf{z})$.
- Given $x(\mathbf{s}_0)$, we could obtain posterior estimates of the "success probability" $P(Z(\mathbf{s}_0) = 1 | x(\mathbf{s}_0), \mathbf{U}(\mathbf{s}_0), \boldsymbol{\alpha}, \boldsymbol{\mu}, T_{11}, T12)$.
- Without $x(\mathbf{s}_0)$, we could still obtain $P(Z(\mathbf{s}_0) = 1 | \mathbf{U}(\mathbf{s}_0), \boldsymbol{\alpha}, \boldsymbol{\mu}, T_{11}, T12)$ from

$$\int P(Z(\mathbf{s}_0) = 1 | x(\mathbf{s}_0), \mathbf{U}(\mathbf{s}_0), \boldsymbol{\alpha}, \boldsymbol{\mu}, T_{11}, T12) p(x(\mathbf{s}_0), \mu_1, T_{11}) dx(\mathbf{s}_0) \tag{19}$$

Conditional modeling

- Previously, we consider a bivariate Gaussian process to model $Y(\mathbf{s})$ and $X(\mathbf{s})$ *jointly*. Alternatively, we could directly consider a *conditional* approach.
- Can we model $\mathbf{Y}|\mathbf{X}$ using a condition *process* $Y(\mathbf{s})|X(\mathbf{s})$? What is the joint distribution of $Y(\mathbf{s}_i)|X(\mathbf{s}_i)$ and $Y(\mathbf{s}_j)|X(\mathbf{s}_j)$?
- Assume $X(\mathbf{s})$ is a univariate Gaussian spatial process with mean $\mu_X(\mathbf{s})$ and covariance function $C_X(\cdot; \boldsymbol{\theta}_X)$. Then we can model for any finite collection of n locations

$$Y(\mathbf{s}_i) = \beta_0 + \beta_1 X(\mathbf{s}_i) + e(\mathbf{s}_i), \quad i = 1, \cdots, n \tag{20}$$

where $e(\mathbf{s})$ is another GP with zero mean and covariance function $C_e(\cdot; \boldsymbol{\theta}_e)$ independent of $X(\mathbf{s})$.

Conditional modeling

- Therefore we have the joint distribution of **X** and **Y**

$$\begin{bmatrix}\mathbf{X}\\ \mathbf{Y}\end{bmatrix} \sim N\left(\begin{bmatrix}\boldsymbol{\mu}_X\\ \beta_0\mathbf{1}+\beta_1\boldsymbol{\mu}_X\end{bmatrix}, \begin{bmatrix}\Sigma_X(\boldsymbol{\theta}_X) & \beta_1\Sigma_X(\boldsymbol{\theta}_X)\\ \beta_1\Sigma_X(\boldsymbol{\theta}_X) & \Sigma_e(\boldsymbol{\theta}_e)+\beta_1^2\Sigma_X(\boldsymbol{\theta}_X)\end{bmatrix}\right) \tag{21}$$

- It arises from a legitimate bivariate process $\mathbf{W}(\mathbf{s}) = (X(\mathbf{s}), Y(\mathbf{s}))^T$ with mean $\boldsymbol{\mu}_{\mathbf{W}}(\mathbf{s}) = (\mu_X(\mathbf{s}), \beta_0 + \beta_1\mu_X(\mathbf{s}))$ and cross covariance

$$C_{\mathbf{W}}(\mathbf{s},\mathbf{s}') = \begin{bmatrix}C_X(\mathbf{s},\mathbf{s}') & \beta_1 C_X(\mathbf{s},\mathbf{s}')\\ \beta_1 C_X(\mathbf{s},\mathbf{s}') & C_e(\mathbf{s},\mathbf{s}')+\beta_1^2 C_X(\mathbf{s},\mathbf{s}')\end{bmatrix} \tag{22}$$

- We can define spatial regression model $E[Y(\mathbf{s})|X(\mathbf{s})] = \beta_0 + \beta_1 X(\mathbf{s})$.

Coregionalization models

- Consider a constructive modeling strategy to add flexibility to separable models while retaining interpretability and computational tractability.
- The approach is through the *linear model of coregionalization* (LMC).
- The most basic coregionalization model, a.k,a. *intrinsic specification* (Matheron, 1982): $\mathbf{Y}(\mathbf{s}) = A\mathbf{w}(\mathbf{s})$, where $w_j(\mathbf{s}) \overset{iid}{\sim} (0, \rho(h))$. Therefore

$$\mathrm{E}[\mathbf{Y}(\mathbf{s})] = \mathbf{0}, \quad \Sigma_{\mathbf{Y}(\mathbf{s}),\mathbf{Y}(\mathbf{s}')} = C(\mathbf{s}-\mathbf{s}') = \rho(\mathbf{s}-\mathbf{s}')AA^T \tag{23}$$

- *Intrinsic*: specification only requires the first and second moments of differences in measurement vectors and

$$\mathrm{E}[\mathbf{Y}(\mathbf{s}) - \mathbf{Y}(\mathbf{s}')] = \mathbf{0}, \quad \Sigma_{\mathbf{Y}(\mathbf{s})-\mathbf{Y}(\mathbf{s}')} = G(\mathbf{s}-\mathbf{s}') \tag{24}$$

- We denote $T = AA^T$ and assume A full rank and lower triangular.

Coregionalization models

- A more general LMC: $\mathbf{Y}(\mathbf{s}) = A\mathbf{w}(\mathbf{s})$, where $w_j(\mathbf{s}) \overset{inid}{\sim} (\mu_j, \rho_j(h))$. Therefore

$$\mathrm{E}[\mathbf{Y}(\mathbf{s})] = A\boldsymbol{\mu}, \quad \Sigma_{\mathbf{Y}(\mathbf{s}),\mathbf{Y}(\mathbf{s}')} = C(\mathbf{s}-\mathbf{s}') = \sum_{j=1}^{p} \rho_j(\mathbf{s}-\mathbf{s}')T_j \tag{25}$$

  where $T_j = \mathbf{a}_j\mathbf{a}_j^T$ with $\mathbf{a}_j$ the j-th column of A. Note $\sum_j T_j = T$.
- Alternatively, we can have a general *nested covariance model* (Wackernagel, 1998)

$$\mathbf{Y}(\mathbf{s}) = \sum \mathbf{Y}^{(u)}(\mathbf{s}) = \sum_{u=1}^{r} A^{(u)}\mathbf{w}^{(u)}(\mathbf{s}) \tag{26}$$

  where the $\mathbf{Y}^{(u)}$ are independent intrinsic LMC specifications with the components of $\mathbf{w}^{(u)}$ having correlation function ρ_u. Then the cross-covariance function is ($T^{(u)} = A^{(u)}(A^{(u)})^T$ *coregionalization* matrices.)

$$C(\mathbf{s}-\mathbf{s}') = \sum_{u=1}^{r} \rho_u(\mathbf{s}-\mathbf{s}')T^{(u)} \tag{27}$$

Coregionalization models

- In a general multivariate spatial model

$$\mathbf{Y}(\mathbf{s}) = \boldsymbol{\mu}(\mathbf{s}) + \mathbf{v}(\mathbf{s}) + \boldsymbol{\epsilon}(\mathbf{s}) \tag{28}$$

where $\boldsymbol{\epsilon}(\mathbf{s}) \sim N(\mathbf{0}, D)$, $D = \mathrm{diag}(\tau_j^2)$, $\mathbf{v}(\mathbf{s}) = A\mathbf{w}(\mathbf{s})$, and $\mu_j(\mathbf{s}) = \mathbf{X}_j^T(\mathbf{s})\boldsymbol{\beta}_j$.

- This can be cast into a hierarchical model

$$\mathbf{Y}(\mathbf{s}_i)|\mu(\mathbf{s}_i), \mathbf{v}(\mathbf{s}_i) \overset{ind}{\sim} N(\mu(\mathbf{s}_i) + \mathbf{v}(\mathbf{s}_i), D) \tag{29}$$

$$\mathbf{v} \sim N(\mathbf{0}, \sum_{j=1}^{p} H_j \otimes T_j) \tag{30}$$

- Concatenating $\mathbf{Y}(\mathbf{s}_i)$ into $\mathbf{Y}$ and marginalizing over $\mathbf{v}$ yields

$$p(\mathbf{Y}|\{\boldsymbol{\beta}_j\}, D, \{\rho_j\}, T) = N\left(\boldsymbol{\mu}, \sum_{j=1}^{p} H_j \otimes T_j + I_{n\times n} \otimes D\right) \tag{31}$$

Spatially varying coefficient models

- Recall the usual Gaussian stationary spatial process model

$$Y(\mathbf{s}) = \mu(\mathbf{s}) + w(\mathbf{s}) + \epsilon(\mathbf{s}) \tag{32}$$

where $\mu(\mathbf{s}) = \mathbf{x}(\mathbf{s})^T\beta$, $\epsilon(\mathbf{s})$ is a white noise process $(0, \tau^2\delta(\mathbf{s}, \mathbf{s}'))$ and $w(\mathbf{s})$ is a second-order stationary process with 0 mean and covariance function $\sigma^2\rho(\mathbf{s}, \mathbf{s}'; \phi)$.

- Let $\mu(\mathbf{s}) = \beta_0 + \beta_1 x(\mathbf{s})$, and $w(\mathbf{s}) = \beta_0(\mathbf{s}) + \beta_1(\mathbf{s})x(\mathbf{s})$. Then we can denote $\tilde{\beta}_0(\mathbf{s}) = \beta_0 + \beta_0(\mathbf{s})$ and $\tilde{\beta}_1(\mathbf{s}) = \beta_1 + \beta_1(\mathbf{s})$.
- The model can be written as

$$Y(\mathbf{s}) = \tilde{\beta}_0(\mathbf{s}) + \tilde{\beta}_1(\mathbf{s})x(\mathbf{s}) + \epsilon(\mathbf{s}) \tag{33}$$

where we have $\mathrm{Cov}(Y(\mathbf{s}), Y(\mathbf{s}')|\beta_0, \beta_1, \tau^2, \sigma_0^2, \sigma_1^2) = \sigma_0^2\rho_0(\mathbf{s} - \mathbf{s}'; \phi_0) + \sigma_1^2 x(\mathbf{s})x(\mathbf{s}')\rho_1(\mathbf{s} - \mathbf{s}'; \phi_1) + \tau^2\delta(\mathbf{s}, \mathbf{s}')$ nonstationary.

Spatially varying coefficient models

- For the $p \times 1$ covariate vector $\mathbf{X}(\mathbf{s})$ including 1, we consider

$$Y(\mathbf{s}) = \mathbf{X}^T(\mathbf{s})\tilde{\beta}(\mathbf{s}) + \epsilon(\mathbf{s}) \tag{34}$$

  where $\tilde{\beta}(\mathbf{s})$ is assumed to follow a p-variate spatial process model.
- Denote $\mathbf{X}$ as $n \times np$ block diagonal having as block for the i-th row $\mathbf{X}^T(\mathbf{s}_i)$. Then we can write $\mathbf{Y} = \mathbf{X}^T\tilde{\mathbf{B}} + \boldsymbol{\epsilon}$, where $\tilde{\mathbf{B}}$ is $np \times 1$ the concatenated vector of $\tilde{\beta}(\mathbf{s})$, and $\boldsymbol{\epsilon} \sim N(0, \tau^2 I)$.
- Denote $\boldsymbol{\mu}_\beta = (\beta_1, \cdots, \beta_p)^T$. We assume separable models for $\tilde{\mathbf{B}}$

$$\tilde{\mathbf{B}} \sim N(\mathbf{1}_{n\times 1} \otimes \boldsymbol{\mu}_\beta, H(\phi) \otimes T) \tag{35}$$

- If we write $\tilde{\mathbf{B}} = \mathbf{B} = \mathbf{1}_{n\times 1} \otimes \boldsymbol{\mu}_\beta$, then we have

$$Y(\mathbf{s}) = \mathbf{X}^T(\mathbf{s})\boldsymbol{\mu}_\beta + \mathbf{X}^T(\mathbf{s})\beta(\mathbf{s}) + \epsilon(\mathbf{s}) \tag{36}$$

Spatially varying coregionalization models

- A possible extension of the LMC would replace A by $A(\mathbf{s})$ to get the following *spatially varying LMC*:

$$\mathbf{Y}(\mathbf{s}) = A(\mathbf{s})\mathbf{w}(\mathbf{s}) \tag{37}$$

  Therefore the covariance function becomes non-stationary

$$C(\mathbf{s}, \mathbf{s}') = \sum_{j=1}^{p} \rho_j(\mathbf{s} - \mathbf{s}')\mathbf{a}_j(\mathbf{s})\mathbf{a}_j(\mathbf{s}')^T \tag{38}$$

- $T_j(\mathbf{s}) = \mathbf{a}_j(\mathbf{s})\mathbf{a}_j^T(\mathbf{s})$ with $\mathbf{a}_j(\mathbf{s})$ the j-th column of $A(\mathbf{s})$. $\sum_j T_j(\mathbf{s}) = \mathbf{T}(\mathbf{s})$.
- Extending the intrinsic specification for $\mathbf{Y}(\mathbf{s})$ yields the following covariance function

$$C(\mathbf{s}, \mathbf{s}') = \rho(\mathbf{s} - \mathbf{s}')\mathbf{T}(\mathbf{s}) \tag{39}$$

  which is a multivariate version of the case of a spatial process with a spatially varying variance.

Table of Contents

Multivariate models for areal data

- We explore the extension of univariate CAR methodology to the multivariate setting.
- Multivariate CAR (MCAR) models can also provide coefficients in a multiple regression setting that are dependent and spatially varying at the areal unit level.
- Gamerman et al. (2002) investigate a Gaussian Markov random field (GMRF) model (a multivariate generalization of the pairwise difference IAR model).
- Assuncão et al. (2002) consider *space-varying coefficient models*.
- Zhang, Hodges and Banerjee (2009) develop an alternative approach building upon the techniques of smoothed ANOVA (SANOVA).

The multivariate CAR (MCAR) distribution

- Let $\phi^T = (\phi_1, \cdots, \phi_n)$ where each $\phi_i = (\phi_{i1}, \cdots, \phi_{ip})^T$. Under the MRF assumption, we specify the full conditionals

$$p(\phi_i|\phi_{j\neq i}, \Gamma_i) = N\left(\sum_{j\sim i} B_{ij}\phi_j, \Gamma_i\right), \quad , i,j = 1, \cdots, n \tag{40}$$

where Γ_i and B_{ij} are $p \times p$ matrices.

- Mardia (1988) proved, using a multivariate analogue of Brook's Lemma, the joint distribution

$$p(\phi|\{\Gamma_i\}) \propto \exp\left\{-\frac{1}{2}\phi^T\Gamma^{-1}(I - \tilde{B})\phi\right\} \tag{41}$$

where Γ is block-diagonal with blocks Γ_i, and $\tilde{B}$ is $np \times np$ with (i,j)-th block B_{ij}.

The multivariate CAR (MCAR) distribution

- A convenient special case to guarantee the symmetry of $\Gamma^{-1}(I - \tilde{B})$ is $B_{ij} = b_{ij} I_p$ and $b_{ij}\Gamma_j = b_{ji}\Gamma_i$.
- Analogous to the univariate case, we set $b_{ij} = w_{ij}/w_{i+}$ and $\Sigma_i = w_{i+}^{-1}\Sigma$.
- Note, in this case, we have $\tilde{B} = B \otimes I$ and $\Gamma = D^{-1} \otimes \Sigma$. Therefore,

$$\Gamma^{-1}(I - \tilde{B}) = (D \otimes \Sigma^{-1})(I - B \otimes I) = (D - W) \otimes \Sigma^{-1} \tag{42}$$

- Note the singularity of $D - W$ implies that $\Gamma^{-1}(I - \tilde{B})$ is singular. We denote this distribution as $\mathrm{MCAR}(1, \Sigma)$.
- To consider remedies to the impropriety, Mardia (1988) proposed rewriting $N\left(R_i \sum_{j\sim i} B_{ij}\phi_j, \Gamma_i\right)$. The symmetry condition becomes $\Gamma_j B_{ij}^T R_i^T = R_j B_{ji} \Gamma_i$.
- Setting $R_{ij} \equiv R = \rho I$ yields a *separable* model $\phi \sim N(0, (D - \rho W)^{-1} \otimes \Sigma)$, denoted as $\mathrm{MCAR}(\rho, \Sigma)$.

Proper non-separable MCAR distribution

- However, the assumption of a common ρ for $j=1,\cdots,p$ may well be too strong.
- We may use ρ_j for each component:

$$\phi \sim N_{np}\left(\mathbf{0}, [\mathrm{diag}(U_1^T,\cdots,U_p^T)(\Lambda\otimes I_n)\mathrm{diag}(U_1,\cdots,U_p)]^{-1}\right) \tag{43}$$

where $U_j^T U_j = D - \rho_j W$, $j=1,\cdots,p$ and $\Lambda = \Sigma^{-1}$.

- This leads to to a *non-separable* model. We denote this distribution as $\mathrm{MCAR}(\rho_1,\cdots,\rho_p,\Sigma)$, or simply $\mathrm{MCAR}(\boldsymbol{\rho},\Sigma)$.

Modeling with a proper, non-separable MCAR distribution

- Previous MCAR is specified for spatial random effects in a hierarchical model.
- Suppose we have a linear model with continuous data $\mathbf{Y}_{ik}$, $i = 1, \cdots, n$, $k = 1, \cdots, m_i$, where $\mathbf{Y}_{ik}$ is a $p \times 1$ vector denoting the k-th response at the i-th areal unit.
- The first stage models the mean $\boldsymbol{\mu}_{ik}$ with $\mu_{ikj} = (\mathbf{X}_{ik})_j \beta^{(j)} + \phi_{ij}$, $j = 1, \cdots, p$. Here $\mathbf{X}_{ik}$ is a $p \times s$ matrix with covariates associated with $\mathbf{Y}_{ik}$ having j-th row $(\mathbf{X}_{ik})_j$, $\beta^{(j)}$ is an $s \times 1$ coefficient vector associated with the j-th component of $\mathbf{Y}_{ik}$'s, and ϕ_{ij} is the j-th component of the $p \times 1$ vector $\boldsymbol{\phi}_i$. We have
$$\mathbf{Y}_{ik} | \{\beta^{(j)}\}, \boldsymbol{\phi}_i, V \sim N(\boldsymbol{\mu}_{ik}, V) \tag{44}$$
- The second stage specifies priors for $\{\beta^{(j)}\}$ and V, and an MCAR model for $\boldsymbol{\phi}_i$.
- Finally, a hyperprior on the MCAR parameters completes the model.

Conditionally specified generalized MCAR (GMCAR) distributions

- Jin, Carlin and Banerjee (2005) expand upon this idea by building the joint distribution for a multivariate Markov random field (MRF) through specifications of simpler conditional and marginal models.
- For simplicity, we consider bivariate case. Assume the joint distribution of ϕ_1 and ϕ_2 is

$$\begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22} \end{bmatrix} \right) \tag{45}$$

- Then we have $\phi_1|\phi_2 \sim N(A\phi_2, \Sigma_{11\cdot 2})$, where $A = \Sigma_{12}\Sigma_{22}^{-1}$, and $\Sigma_{11\cdot 2} = \Sigma_{11} - \Sigma 12 \Sigma_{22}^{-1}\Sigma_{12}^T$. If $\phi_2 \sim N(0, \Sigma_{22})$, then $p(\phi) = p(\phi_1|\phi_2)p(\phi_2)$.
- Jin et al. (2005) propose

$$\phi_1|\phi_2 \sim N(A\phi_2, [(D - \rho_1 W)\tau_1]^{-1}) \tag{46}$$

$$\phi_2 \sim N(\mathbf{0}, [(D - \rho_2 W)\tau_2]^{-1}) \tag{47}$$

where $A = \eta_0 I + \eta_1 W$.

Conditionally specified generalized MCAR (GMCAR) distributions

- Under these assumptions, we can obtain

$$\Sigma_{11} = [\tau_1(D - \rho_1 W)]^{-1} + (\eta_0 I + \eta_1 W)[\tau_2(D - \rho_2 W)]^{-1}(\eta_0 I + \eta_1 W) \tag{48}$$

$$\Sigma_{12} = (\eta_0 I + \eta_1 W)[\tau_2(D - \rho_2 W)]^{-1} \tag{49}$$

$$\Sigma_{22} = [\tau_2(D - \rho_2 W)]^{-1} \tag{50}$$

- Jin et al. (2005) denote this new model by $\mathrm{GMCAR}(\rho_1, \rho_2, \eta_0, \eta_1, \tau_1, \tau_2)$.
- Setting $\rho_1 = \rho_2 = \rho$ and $\eta_1 = 0$ produces the separable model with $\Sigma^{-1} = \Lambda \otimes (D - \rho W)$, where $\tau_1 = \Lambda_{11}$, $\tau_2 = \Lambda_{11} - \frac{\Lambda_{12}^2}{\Lambda_{11}}$, and $\eta_0 = -\frac{\Lambda_{12}}{\Lambda_{11}}$.
- Further setting $\rho = 1$ produces an improper MIAR.
- If $\rho_1 \neq \rho_2$ and $\eta_0 = \eta_1 = 0$, this is equivalent to two separate univariate CAR models.
- If $\rho_1 = \rho_2 = 0$ and $\eta_0 \neq 0$, the model becomes iid bivariate normal.

Coregionalized MCAR distributions

- GMCAR depends on the order of conditioning. To obviate this issue, Jin, Banerjee and Carlin (2007) develop an order-free framework for multivariate areal modeling.
- This approach is based on an adaptation of the *linear model of coregionalization* (LMC) to areal data.
- The essential idea is to develop richer spatial association models using linear transformations of much simpler spatial distributions.
- Write the spatial effects in terms of latent processes: $\phi = (A \otimes I_n)\mathbf{u}$, where $\mathbf{u} = (\mathbf{u}_1^T, \cdots, \mathbf{u}_p^T)^T$ is $np \times 1$ with each $\mathbf{u}_j$ being an $n \times 1$ areal process.
- A proper distribution for $\mathbf{u}$ ensures a proper distribution for ϕ subject to the non-singularity of A.

Case 1: Independent and identical latent CAR variables

- First we assume the random spatial processes $\mathbf{u}_j$, $j = 1, \cdots, p$ are independent and identical. Give a CAR structure for each of them

$$\mathbf{u}_j \sim N_n(\mathbf{0}, (D - \alpha W)^{-1}), \quad j = 1, \cdots, p \tag{51}$$

- Therefore $\mathbf{u} \sim N_{np}(\mathbf{0}, I_p \otimes (D - \alpha W)^{-1})$.
- The joint distribution of $\phi = (A \otimes I_n)\mathbf{u}$ is

$$\phi \sim N_{np}(\mathbf{0}, \Sigma \otimes (D - \alpha W)^{-1}) \tag{52}$$

where $\Sigma = AA^T$. We denote this distribution as $\mathrm{MCAR}(\alpha, \Sigma)$.
- The model is independent of the choice of A. Without loss of generality, we specify A as the upper-triangular Cholesky factor of Σ.
- We need to require $\frac{1}{\xi_{\min}} < \alpha < \frac{1}{\xi_{\max}}$, where $\xi_{\min}$ and $\xi_{\max}$ are the minimum and maximum eigenvalues of $D^{-\frac{1}{2}} W D^{-\frac{1}{2}}$.

Case 2: Independent but not identical latent CAR variables

- Now assume $\mathbf{u}_j$ are independent, but relax them being identically distributed:

$$\mathbf{u}_j \sim N_n(\mathbf{0}, (D - \alpha_j W)^{-1}), \quad j = 1, \cdots, p \tag{53}$$

- The joint distribution of $\phi = (A \otimes I_n)\mathbf{u}$ becomes

$$\phi \sim N_{np}(\mathbf{0}, (A \otimes I_n)\Gamma^{-1}(A \otimes I_n)^T) \tag{54}$$

where $\Sigma = AA^T$ and Γ is an $np \times np$ block diagonal matrix with $n \times n$ diagonal entries $\Gamma_j = D - \alpha_j W$. We denote this distribution as $\mathrm{MCAR}(\boldsymbol{\alpha}, \Sigma)$.

- Again we may specify A as the upper-triangular Cholesky factor of Σ.
- Note there is no unique joint distribution for ϕ.

Case 3: Dependent and not identical latent CAR variables

- Finally, in this case we assume $\mathbf{u}_j$ are neither independent nor identically distributed.
- Assume that $u_{ij} \perp u_{i,l\neq j} | u_{k\neq i,j}, u_{k\neq i,l\neq j}$, where $l, j = 1, \cdots, p$ and $i, k = 1, \cdots, n$.
- Denote $b_{jl} = \mathrm{Cov}(\mathbf{u}_j, \mathbf{u}_l)$. Then we have

$$\mathbf{u} \sim N_{np}(\mathbf{0}, (I_p \otimes D - B \otimes W)^{-1}) \tag{55}$$

  where B is a $p \times p$ matrix with the elements b_{jl}, $j, l = 1, \cdots, p$.
- Note $I_p \otimes D - B \otimes W = (I_p \otimes D)^{\frac{1}{2}}(I_{np} - B \otimes D^{-\frac{1}{2}} W D^{-\frac{1}{2}})(I_p \otimes D)^{\frac{1}{2}}$. Denote the eigenvalues of B as ζ_j. We need to require $\frac{1}{\xi_{\min}} < \zeta_j < \frac{1}{\xi_{\max}}$.
- The joint distribution of $\phi = (A \otimes I_n)\mathbf{u}$ becomes

$$\phi \sim N_{np}(\mathbf{0}, (A \otimes I_n)(I_p \otimes D - B \otimes W)^{-1}(A \otimes I_n)^T) \tag{56}$$