Modeling multivariate spatial data

Alan Gelfand[1] and Andrew O. Finley[2]

[1] Department of Statistical Science, Duke University, Durham, North Carolina.

[2] Department of Forestry, Michigan State University, Lansing, Michigan.

September 9, 2014

- Point-referenced spatial data often come as multivariate measurements at each location.
- Examples:
  - Environmental monitoring: stations yield measurements on ozone, NO, CO, and $PM_{2.5}$.
  - Community ecology: assemblages of plant species due to water availibility, temperature, and light requirements.
  - Forestry: measurements of stand characteristics age, total biomass, and average tree diameter.
  - Atmospheric modeling: at a given site we observe surface temperature, precipitation and wind speed
- We anticipate dependence between measurements
  - at a particular location
  - across locations

Bivariate Linear Spatial Regression

- A single covariate $X(\mathbf{s})$ and a univariate response $Y(\mathbf{s})$
- At any arbitrary point in the domain, we conceive a linear spatial relationship:

$$\mathsf{E}[Y(\mathbf{s}) \,|\, X(\mathbf{s})] = \beta_0 + \beta_1 X(\mathbf{s});$$

  where $X(\mathbf{s})$ and $Y(\mathbf{s})$ are spatial processes.

- Regression on uncountable sets:

$$\text{Regress} \quad \{Y(\mathbf{s}) : \mathbf{s} \in \mathscr{D}\} \quad \text{on} \quad \{X(\mathbf{s}) : \mathbf{s} \in \mathscr{D}\}\ .$$

- Inference:
  - Estimate β_0 and β_1.
  - Estimate spatial surface $\{X(\mathbf{s}) : \mathbf{s} \in \mathscr{D}\}$.
  - Estimate spatial surface $\{Y(\mathbf{s}) : \mathbf{s} \in \mathscr{D}\}$.

Bivariate spatial process

- A bivariate distribution $[Y, X]$ will yield regression $[Y \mid X]$.
- So why not start with a bivariate process?

$$\mathbf{Z}(\mathbf{s}) = \begin{bmatrix} X(\mathbf{s}) \\ Y(\mathbf{s}) \end{bmatrix} \sim GP_2\left(\begin{bmatrix} \mu_X(\mathbf{s}) \\ \mu_Y(\mathbf{s}) \end{bmatrix}, \begin{bmatrix} C_{XX}(\cdot;\boldsymbol{\theta}_Z) & C_{XY}(\cdot;\boldsymbol{\theta}_Z) \\ C_{YX}(\cdot;\boldsymbol{\theta}_Z) & C_{YY}(\cdot;\boldsymbol{\theta}_Z) \end{bmatrix} \right)$$

- The cross-covariance function:

$$\mathbf{C}_Z(\mathbf{s},\mathbf{t};\boldsymbol{\theta}_Z) = \begin{bmatrix} C_{XX}(\mathbf{s},\mathbf{t};\boldsymbol{\theta}_Z) & C_{XY}(\mathbf{s},\mathbf{t};\boldsymbol{\theta}_Z) \\ C_{YX}(\mathbf{s},\mathbf{t};\boldsymbol{\theta}_Z) & C_{YY}(\mathbf{s},\mathbf{t};\boldsymbol{\theta}_Z) \end{bmatrix} ,$$

where $C_{XY}(\mathbf{s},\mathbf{t}) = \text{cov}(X(\mathbf{s}), Y(\mathbf{t}))$ and so on.

- Cross-covariance functions satisfy certain properties:

  $C_{XY}(\mathbf{s},\mathbf{t}) = \text{cov}(X(\mathbf{s}), Y(\mathbf{t})) = \text{cov}(Y(\mathbf{t}), X(\mathbf{s})) = C_{YX}(\mathbf{t},\mathbf{s}).$

  Caution: $C_{XY}(\mathbf{s},\mathbf{t}) \neq C_{XY}(\mathbf{t},\mathbf{s})$ and $C_{XY}(\mathbf{s},\mathbf{t}) \neq C_{YX}(\mathbf{s},\mathbf{t})$

  .

- In matrix terms, $\mathbf{C}_Z(\mathbf{s},\mathbf{t};\boldsymbol{\theta}_Z)^\top = \mathbf{C}_Z(\mathbf{t},\mathbf{s};\boldsymbol{\theta}_Z)$

- Positive-definiteness for any finite collection of points:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{a}_i^\top \mathbf{C}_Z(\mathbf{s}_i,\mathbf{t}_j;\boldsymbol{\theta}_Z)\mathbf{a}_j > 0 \ \text{ for all } \ \mathbf{a}_i \in \Re^2 \setminus \{\mathbf{0}\}.$$

Bivariare Spatial Regression from a Separable Process

- To ensure $\mathrm{E}[Y(\mathbf{s}) \mid X(\mathbf{s})] = \beta_0 + \beta_1 X(\mathbf{s})$, we assume

$$\mathbf{Z}(\mathbf{s}) = \begin{bmatrix} X(\mathbf{s}) \\ Y(\mathbf{s}) \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} T_{11} & T_{12} \\ T_{12} & T_{22} \end{bmatrix}\right) \text{ for every } \mathbf{s} \in \mathscr{D}$$

- Simplifying assumption :

$$\mathbf{C}_Z(\mathbf{s}, \mathbf{t}) = \rho(\mathbf{s}, \mathbf{t})\mathbf{T} \Longrightarrow \mathbf{\Sigma}_Z = \{\rho(\mathbf{s}_i, \mathbf{s}_j)\mathbf{T}\} = \mathbf{R}(\phi) \otimes \mathbf{T} \; .$$

- Then, $p(Y(\mathbf{s}) \,|\, X(\mathbf{s})) = N(Y(\mathbf{s}) \,|\, \beta_0 + \beta_1 X(\mathbf{s}), \sigma^2)$, where

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

- Regression coefficients are functions of process parameters.

- Estimate $\{\mu_1, \mu_2, T_{11}, T_{12}, T_{22}\}$ by sampling from

$$p(\phi) \times N(\boldsymbol{\mu} \,|\, \boldsymbol{\delta}, \mathbf{V}_\mu) \times IW(\mathbf{T} \,|\, r, \mathbf{S}) \times N(\mathbf{Z} \,|\, \boldsymbol{\mu}, \mathbf{R}(\phi) \otimes \mathbf{T})$$

- Immediately obtain posterior samples of $\{\beta_0, \beta_1, \sigma^2\}$.

Bivariate Spatial Regression with Misalignment

- Rearrange the components of $\mathbf{Z}$ to $\widetilde{\mathbf{Z}} = (X(\mathbf{s}_1), X(\mathbf{s}_2), \ldots, X(\mathbf{s}_n), Y(\mathbf{s}_1), Y(\mathbf{s}_2), \ldots, Y(\mathbf{s}_n))^\top$ yields
$$\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \mathbf{1} \\ \mu_2 \mathbf{1} \end{bmatrix} \ , \ \mathbf{T} \otimes \mathbf{R}\left(\phi\right) \right) \ .$$

- Priors: Wishart for T^{-1}, normal (perhaps flat) for (μ_1, μ_2), discrete prior for ϕ or perhaps a uniform on $(0, .5$max dist).

- Estimation: Markov chain Monte Carlo (Gibbs, Metropolis, Slice, HMC/NUTS); Integrated Nested Laplace Approximation (INLA).

Hierarchical approach (Royle and Berliner, 1999; Cressie and Wikle, 2011)

- $Y(\mathbf{s})$ and $X(\mathbf{s})$ observed over a finite set of locations $\mathscr{S} = \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n\}$.
- **Y** and **X** are $n \times 1$ vectors of observed $Y(\mathbf{s}_i)$'s and $X(\mathbf{s}_i)$'s, respectively.
- How do we model $\mathbf{Y} \mid \mathbf{X}$?
- No "conditional process"—meaningless to talk about the joint distribution of $Y(\mathbf{s}_i) \mid X(\mathbf{s}_i)$ and $Y(\mathbf{s}_j) \mid X(\mathbf{s}_j)$ for two distinct locations $\mathbf{s}_i$ and $\mathbf{s}_j$.
- Can model using $[\mathbf{X}] \times [\mathbf{Y} \mid \mathbf{X}]$ but can we interpolate/predict at arbitrary locations?

Hierarchical approach (contd.)

- $X(\mathbf{s}) \sim GP(\mu_X(\mathbf{s}), C_X(\cdot; \boldsymbol{\theta}_X))$. Therefore, $\mathbf{X} \sim N(\boldsymbol{\mu}_X, \mathbf{C}_X(\boldsymbol{\theta}_X))$.
- $\mathbf{C}_X(\boldsymbol{\theta}_X)$ is $n \times n$ with entries $C_X(\mathbf{s}_i, \mathbf{s}_j; \boldsymbol{\theta}_X)$.
- $e(\mathbf{s}) \sim GP(0, C_e(\cdot; \boldsymbol{\theta}_e))$; $\mathbf{C}_e$ is analogous to $\mathbf{C}_X$.

$$Y(\mathbf{s}_i) = \beta_0 + \beta_1 X(\mathbf{s}_i) + e(\mathbf{s}_i)\ , \quad \text{for} \quad i = 1, 2, \ldots, n\ .$$

- Joint distribution of **Y** and **X**:

$$\left(\begin{array}{c} \mathbf{X} \\ \mathbf{Y} \end{array} \right) \sim N\left(\left[\begin{array}{c} \boldsymbol{\mu}_X \\ \boldsymbol{\mu}_Y \end{array} \right], \left[\begin{array}{cc} \mathbf{C}_X(\boldsymbol{\theta}_X) & \beta_1 \mathbf{C}_X(\boldsymbol{\theta}_X) \\ \beta_1 \mathbf{C}_X(\boldsymbol{\theta}_X) & \mathbf{C}_e(\boldsymbol{\theta}_e) + \beta_1^2 \mathbf{C}_X(\boldsymbol{\theta}_X) \end{array} \right] \right)\ ,$$

where $\boldsymbol{\mu}_Y = \beta_0 \mathbf{1} + \beta_1 \boldsymbol{\mu}_X$.

This joint distribution arises from a bivariate spatial process:

$$\mathbf{W}(\mathbf{s}) = \begin{bmatrix} X(\mathbf{s}) \\ Y(\mathbf{s}) \end{bmatrix} \quad \text{and} \quad \mathrm{E}[\mathbf{W}(\mathbf{s})] = \boldsymbol{\mu}_W(\mathbf{s}) = \begin{bmatrix} \mu_X(\mathbf{s}) \\ \beta_0 + \beta_1 \mu_X(\mathbf{s}) \end{bmatrix} .$$

and cross-covariance

$$C_W(\mathbf{s}, \mathbf{s}') = \left[\begin{array}{cc} C_X(\mathbf{s}, \mathbf{s}') & \beta_1 C_X(\mathbf{s}, \mathbf{s}') \\ \beta_1 C_X(\mathbf{s}, \mathbf{s}') & \beta_1^2 C_X(\mathbf{s}, \mathbf{s}') + C_e(\mathbf{s}, \mathbf{s}') \end{array} \right] ,$$

where we have suppressed the dependence of $C_X(\mathbf{s}, \mathbf{s}')$ and $C_e(\mathbf{s}, \mathbf{s}')$ on $\boldsymbol{\theta}_X$ and $\boldsymbol{\theta}_e$ respectively. This implies that $\mathrm{E}[Y(\mathbf{s}) \mid X(\mathbf{s})] = \beta_0 + \beta_1 X(\mathbf{s})$ for any arbitrary location **s**, thereby specifying a well-defined spatial regression model for an arbitrary **s**.

Coregionalization (Wackernagel)

- Separable models assume one spatial range for both $X(\mathbf{s})$ and $Y(\mathbf{s})$.
- Coregionalization helps to introduce a second "range parameter."
- Introduce two "latent" independent GP's, each having its own parameters:

$$v_1(\mathbf{s}) \sim GP(0, \rho_1(\cdot; \phi_1)) \quad \text{and} \quad v_2(\mathbf{s}) \sim GP(0, \rho_2(\cdot; \phi_2))$$

- Construct a bivariate process as the linear transformation:

$$\begin{aligned} w_1(\mathbf{s}) &= a_{11} v_1(\mathbf{s}) \\ w_2(\mathbf{s}) &= a_{21} v_1(\mathbf{s}) + a_{22} v_2(\mathbf{s}) \end{aligned}$$

Coregionalization

- Short form:

$$\mathbf{w}(\mathbf{s}) = \begin{bmatrix} a_{11} & 0 \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} v_1(\mathbf{s}) \\ v_2(\mathbf{s}) \end{bmatrix} = \mathbf{A}\mathbf{v}(\mathbf{s})$$

- Cross-covariance of $\mathbf{v}(\mathbf{s})$:

$$\mathbf{C}_v(\mathbf{s},\mathbf{t}) = \begin{bmatrix} \rho_1(\mathbf{s},\mathbf{t};\phi_1) & 0 \\ 0 & \rho_2(\mathbf{s},\mathbf{t};\phi_2) \end{bmatrix}$$

- Cross-covariance of $\mathbf{w}(\mathbf{s})$:

$$\mathbf{C}_w(\mathbf{s},\mathbf{t}) = \mathbf{A}\mathbf{C}_v(\mathbf{s},\mathbf{t})\mathbf{A}^\top \ .$$

  It is a valid cross-covariance function (by construction).

- If $\mathbf{s} = \mathbf{t}$, then $\mathbf{C}_w(\mathbf{s},\mathbf{s}) = \mathbf{A}\mathbf{A}^\top$. No loss of generality to specify $\mathbf{A}$ as (lower) triangular.

- If $v_1(\mathbf{s})$ and $v_2(\mathbf{s})$ have identical correlation functions, then $\rho_1(\mathbf{s},\mathbf{t}) = \rho_2(\mathbf{s},\mathbf{t})$ and

$$\mathbf{C}_w(\mathbf{s}) = \rho(\mathbf{s},\mathbf{t};\phi)\mathbf{A}\mathbf{A}^\top \Longrightarrow \text{separable model}$$

- Coregionalized Spatial Linear Model

$$\begin{bmatrix} X(\mathbf{s}) \\ Y(\mathbf{s}) \end{bmatrix} = \begin{bmatrix} \mu_X(\mathbf{s}) \\ \mu_Y(\mathbf{s}) \end{bmatrix} + \begin{bmatrix} w_1(\mathbf{s}) \\ w_2(\mathbf{s}) \end{bmatrix} + \begin{bmatrix} e_X(\mathbf{s}) \\ e_Y(\mathbf{s}) \end{bmatrix} ,$$

where $e_X(\mathbf{s})$ and $e_Y(\mathbf{s})$ are independent white-noise processes

$$\begin{bmatrix} e_X(\mathbf{s}) \\ e_Y(\mathbf{s}) \end{bmatrix} \sim N_2\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_X^2 & 0 \\ 0 & \tau_Y^2 \end{bmatrix} \right) \text{ for every } \mathbf{s} \in \mathscr{D} .$$

Generalizations

- Each location contains m spatial regressions

$$Y_k(\mathbf{s}) = \mu_k(\mathbf{s}) + w_k(\mathbf{s}) + \epsilon_k(\mathbf{s}),\ k = 1, \ldots, m.$$

- Let $v_k(\mathbf{s}) \sim GP(0, \rho_k(\mathbf{s}, \mathbf{s}'))$, for $k = 1, \ldots, m$ be m independent GP's with unit variance.
- Assume $\mathbf{w}(\mathbf{s}) = A(\mathbf{s})\mathbf{v}(\mathbf{s})$ arises as a *space-varying* linear transformation of $\mathbf{v}(\mathbf{s})$. Then:

$$\mathbf{C}_w(\mathbf{s}, \mathbf{t}) = A(\mathbf{s})\mathbf{C}_v(\mathbf{s}, \mathbf{t})A^\top(\mathbf{t})$$

  is a valid cross-covariance function.
- $A(\mathbf{s})$ is unknown!
  - Should we first model $A(\mathbf{s})$ to obtain $\mathbf{C}_w(\mathbf{s}, \mathbf{s})$?
  - Or should we model $\mathbf{C}_w(\mathbf{s}, \mathbf{t})$ first and derive $A(\mathbf{s})$?
  - $A(\mathbf{s})$ is completely determined from within-site associations.

Other approaches for cross-covariance models

- Convolutions of processes and covariance functions
  - Gaspari and Cohn (*Quart. J. Roy. Met. Soc.*, 1999).
  - Majumdar and Gelfand (*Math. Geo.*, 2007).
- Latent dimension approach:
  - Apanasovich and Genton (*Biometrika*, 2010).
  - Apanasovich et al. (*JASA*, 2012).
- Multivariate Matérn family
  - Gneiting et al. (*JASA*, 2010).
- Nonstationary variants of coregionalization
  - Space-varying: Gelfand et al. (*Test*, 2010).
  - Dimension-reducing (over space): Guhaniyogi et al. (*JABES*, 2012).
  - Dimension-reducing (over outcomes): Ren and Banerjee (*Biometrics*, 2013).
  - Variogram modeling: De Iaco et al. (*Math. Geo.*, 2003).