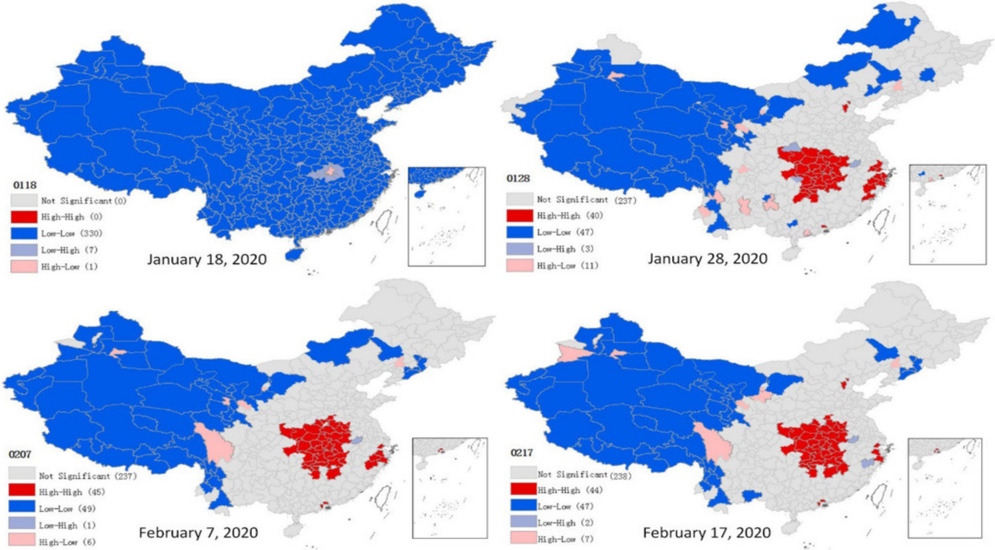
Lecture 10 Spatiotemporal Modeling

Shiwei Lan[1]

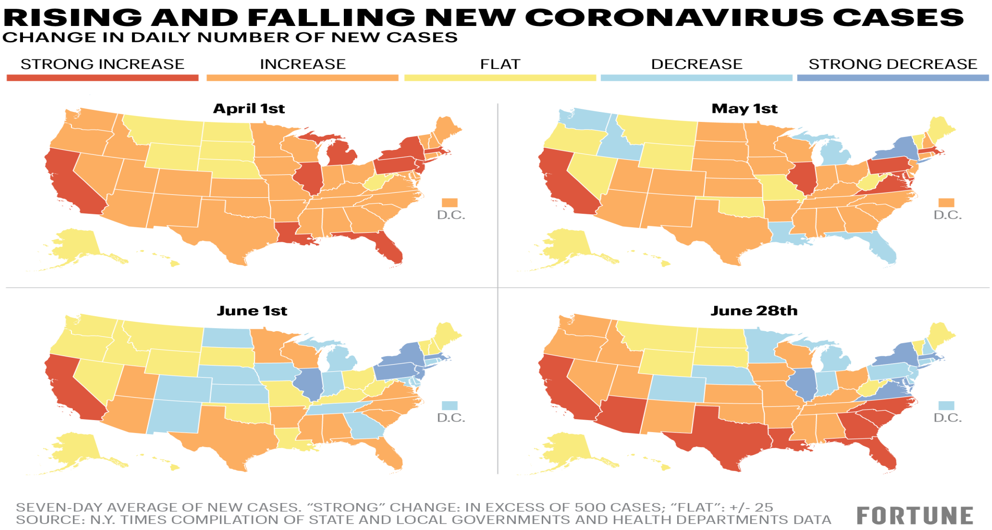
[1]School of Mathematical and Statistical Sciences
Arizona State University

STP598 Spatiotemporal Analysis
Fall 2021

Spatiotemporal Problems



Spatiotemporal Problems



Spatiotemporal modeling: challenges

When analyzing spatiotemporal data, researchers in various areas are faced with challenges:

- highly multivariate, with many important predictors and response variables,
- often having single history, or missing data, and
- intricate relationship between space and time.

Spatiotemporal modeling

- Previously in spatial modeling, we had point level (GP) versus areal unit level (CAR) modeling.
- Is time viewed as continuous (over $\mathbb{R}^+$ or some subinterval) or discrete (hourly, daily, ect.)?
- We also have continuous time model (GP) and discrete time (block average) model (time series).
- And we have different combinations when modeling spatiotemporal data:
  1. Point-level continuous time model
  2. Point-level discrete time model (multivariate spatial models)
  3. Area-level continuous time model
  4. Area-level discrete time model (multivariate time series)

- How do we characterize spatiotemporal correlation?
- How do we deal with sparse/missing data?
- How do we explain response changes with covariates?
- ...?

Table of Contents

Spatiotemporal modeling

- Consider point-referenced locations and continuous time.
- Let $Y(\mathbf{s}, t)$ denote the measurement at location $\mathbf{s}$ at time t. We adopt the following general form

$$Y(\mathbf{s}, t) = \mu(\mathbf{s}, t) + e(\mathbf{s}, t) \tag{1}$$

where $\mu(\mathbf{s}, t)$ denotes the mean structure and $e(\mathbf{s}, t)$ denotes the residual.

- If $\mathbf{x}(\mathbf{s}, t)$ is a vector of covariates associated with $Y(\mathbf{s}, t)$, then we could model $\mu(\mathbf{s}, t) = \mathbf{x}(\mathbf{s}, t)^T \boldsymbol{\beta}(\mathbf{s}, t)$, where it could be that $\boldsymbol{\beta}(\mathbf{s}, t) = \boldsymbol{\beta}$, $\boldsymbol{\beta}(\mathbf{s}, t) = \boldsymbol{\beta}_t$ or $\boldsymbol{\beta}(\mathbf{s}, t) = \boldsymbol{\beta}(\mathbf{s})$, etc..
- The residual $e(\mathbf{s}, t)$ can further be written as

$$e(\mathbf{s}, t) = w(\mathbf{s}, t) + \epsilon(\mathbf{s}, t) \tag{2}$$

where $\epsilon(\mathbf{s}, t)$ a white noise process and $w(\mathbf{s}, t)$ is a mean-zero process.

Spatiotemporal modeling

- One could model $Y(\mathbf{s},t) \sim f(y(\mathbf{s},t)|\mu(\mathbf{s},t), w(\mathbf{s},t))$ in the exponential family

$$f(y(\mathbf{s},t)|\mu(\mathbf{s},t), w(\mathbf{s},t)) = h(y(\mathbf{s},t))\exp\{\gamma[\eta(\mathbf{s},t)y(\mathbf{s},t) - \chi(\eta(\mathbf{s},t))]\} \quad (3)$$

  where γ is a positive dispersion parameter, and $g(\eta(\mathbf{s},t)) = \mu(\mathbf{s},t) + w(\mathbf{s},t)$ for some link function g.

- The spatiotemporal random process $w(\mathbf{s},t)$ is usually modeled with further details (Gelfand, Ecker, Knight, and Sirmans, 2004):

$$w(\mathbf{s},t) = \alpha(t) + w(\mathbf{s}) \quad \text{or} \quad \alpha(t)w(\mathbf{s}) \quad (4)$$

$$w(\mathbf{s},t) = \alpha_{\mathbf{s}}(t) \quad (5)$$

$$w(\mathbf{s},t) = w_t(\mathbf{s}) \quad (6)$$

  where $\epsilon(\mathbf{s},t) \overset{iid}{\sim} N(0,\sigma_\epsilon^2)$.

Spatiotemporal modeling

- Consider areal unit data with discrete time.
- Let Y_{it} denote the measurement for unit i at time period t. Similarly we have

$$Y_{it} = \mu_{it} + e_{it} \tag{7}$$

- Now $\mu_{it} = \mathbf{x}_{it}^T \boldsymbol{\beta}_t$, and $e_{it} = w_{it} + \epsilon_{it}$ where the ϵ_{it} are unstructured heterogeneity terms and the w_{it} are spatiotemporal random effects.
- We again model $Y_{it} \sim f(y_{it}|\mu_{it} w_{it})$ in the exponential family

$$f(y_{it}|\mu_{it} w_{it}) = h(y_{it}) \exp\{\gamma[\eta_{it} y_{it} - \chi(\eta_{it})]\} \tag{8}$$

where γ is a positive dispersion parameter, and $g(\eta_{it}) = \mu_{it} + w_{it}$ for some link function g.

Spatiotemporal modeling

- Consider Gaussian model in the case of point-referenced locations and continuous time.
- Denote $\mathbf{Y} = (\mathbf{Y}_1', \cdots, \mathbf{Y}_T')'$ and each $\mathbf{Y}_t = (Y(\mathbf{s}_1, t), \cdots, Y(\mathbf{s}_n, t))'$. Let $\boldsymbol{\mu} = \mathbf{X}(\mathbf{s}, t)\boldsymbol{\beta}$ and $\boldsymbol{\epsilon}$ be defined similarly.
- Under additive spatiotemporal random effect model (4), denote $\boldsymbol{\alpha} = (\alpha(1), \cdots, \alpha(T))'$ and $\mathbf{w} = (w(\mathbf{s}_1), \cdots, w(\mathbf{s}_n))'$. We can write

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\alpha} \otimes \mathbf{1}_n + \mathbf{1}_T \otimes \mathbf{w} + \boldsymbol{\epsilon} \tag{9}$$

- Assuming $\mathbf{w} \sim N(\mathbf{0}, \sigma_w^2 H(\delta))$, and $\alpha(t) \sim AR(1)$ with autocovariance $\sigma_\alpha^2 A(\rho)$. Then we have the following model

$$\mathbf{Y}|\boldsymbol{\beta}, \sigma_\epsilon^2, \sigma_\alpha^2, \rho, \sigma_w^2, \delta \sim N(\boldsymbol{\mu}, \sigma_\alpha^2 A(\rho) \otimes \mathbf{1}_n \mathbf{1}_n' + \sigma_s^2 \mathbf{1}_T \mathbf{1}_T' \otimes H(\delta) + \sigma_\epsilon^2 I_{Tn}) \tag{10}$$

Spatiotemporal modeling

- Under temporally evolving spatial effect model (5), denote $\boldsymbol{\alpha} = (\boldsymbol{\alpha}(1), \cdots, \boldsymbol{\alpha}(T))'$ and $\boldsymbol{\alpha}(t) = (\alpha_{\mathbf{s}_1}(t), \cdots, \alpha_{\mathbf{s}_n}(t))'$. We can write

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\alpha} + \boldsymbol{\epsilon} \tag{11}$$

- If $\alpha_{\mathbf{s}_i}(t) \sim AR(1)$ independently across i, then marginalizing over $\boldsymbol{\alpha}$

$$\mathbf{Y}|\boldsymbol{\beta}, \sigma_\epsilon^2, \sigma_\alpha^2, \rho \sim N(\boldsymbol{\mu}, \sigma_\alpha^2 A(\rho) \otimes I_n + \sigma_\epsilon^2 I_{Tn}) \tag{12}$$

- Under spatially varying temporal effect model (6), denote $\mathbf{w} = (\mathbf{w}_1, \cdots, \mathbf{w}_T)'$ and $\mathbf{w}_t = (w_t(\mathbf{s}_1), \cdots, w_t(\mathbf{s}_n))'$. We can write

$$\mathbf{Y} = \boldsymbol{\mu} + \mathbf{w} + \boldsymbol{\epsilon} \tag{13}$$

- If $\mathbf{w} \sim N(\mathbf{0}, \sigma_w^{2(t)} H(\delta^{(t)}))$, independently for t, then marginalizing over $\mathbf{w}$

$$\mathbf{Y}|\boldsymbol{\beta}, \sigma_\epsilon^2, \sigma_w^2, \boldsymbol{\delta} \sim N(\boldsymbol{\mu}, D(\sigma_w^2, \boldsymbol{\delta}) + \sigma_\epsilon^2 I_{Tn}) \tag{14}$$

where $D(\sigma_w^2, \boldsymbol{\delta})$ is block diagonal with the t-th block $\sigma_w^{2(t)} H(\delta^{(t)})$.

Prediction and forecasting

- Forecasting involves prediction at location $\mathbf{s}_0$ and time t_0, i.e. of $Y(\mathbf{s}_0, t_0)$. Typically $t_0 > T$ if of interest.
- Within the Bayesian framework, we have

$$f(Y(\mathbf{s}_0, t_0)|\mathbf{Y}) = \int f(Y(\mathbf{s}_0, t_0)|\boldsymbol{\beta}, \sigma_\epsilon^2, \alpha(t_0), w(\mathbf{s}_0)) \\ \times dF(\boldsymbol{\beta}, \boldsymbol{\alpha}, \mathbf{w}, \sigma_\epsilon^2, \sigma_\alpha^2, \rho, \sigma_w^2, \delta, \alpha(t_0), w(\mathbf{s}_0)|\mathbf{Y}) \tag{15}$$

- Given a random draw $(\boldsymbol{\beta}^*, \sigma_\epsilon^{2*}, \alpha(t_0)^*, w(\mathbf{s}_0)^*)$ from the posterior $f(\boldsymbol{\beta}, \sigma_\epsilon^2, \alpha(t_0), w(\mathbf{s}_0)|\mathbf{Y})$, if we draw $Y^*(\mathbf{s}_0, t_0)$ from $N(X'(\mathbf{s}_0, t_0)\boldsymbol{\beta}^* + \alpha(t_0)^* + w(\mathbf{s}_t)^*, \sigma_e^{2*})$, marginally $Y^*(\mathbf{s}_0, t_0) \sim f(Y(\mathbf{s}_0, t_0)|\mathbf{Y})$.

Table of Contents

Point-level modeling with continuous time

- Suppose $\mathbf{s} \in \mathbb{R}^2$ and $t \in \mathbb{R}^+$. We define a spatiotemporal process $Y(\mathbf{s}, t)$ using Gaussian process.
- We need to specify a covariance function, which frequently adopts a *separable* form

$$\mathrm{Cov}(Y(\mathbf{s}, t), Y(\mathbf{s}', t')) = \sigma^2 \rho^{(1)}(\mathbf{s} - \mathbf{s}'; \phi)\rho^{(2)}(t - t'; \psi) \tag{16}$$

where $\rho^{(1)}$ is the spatial correlation function and $\rho^{(2)}$ is the temporal correlation function.

- For given I locations and J time points, the covariance matrix of $\mathbf{Y}$ is

$$\Sigma_{\mathbf{Y}}(\sigma^2, \phi, \psi) = \sigma^2 H_{\mathbf{s}}(\phi) \otimes H_t(\psi) \tag{17}$$

- Pros? Cons?

Nonseparable spatiotemporal models

- The separable form for the spatiotemporal covariance function is convenient for computation and offers attractive interpretation.
- However, its form limits the nature of space-time interaction.
- A simple way to extend it is through *mixing*. Suppose $w(\mathbf{s},t) = w_1(\mathbf{s},t) + w_2(\mathbf{s},t)$, with w_1 and w_2 independent processes, each with a separable spatiotemporal covariance function

$$c_\ell(\mathbf{s}-\mathbf{s}', t-t') = \sigma_\ell^2 \rho_\ell^{(1)}(\mathbf{s}-\mathbf{s}')\rho_\ell^{(2)}(t-t'), \quad \ell = 1,2 \tag{18}$$

  Then the covariance function for $w(\mathbf{s},t)$ is not separable.
- Alternatively, De Iaco et al. (2002) considered

$$c_\ell(\mathbf{s}-\mathbf{s}', t-t') = \sigma^2 \int \rho^{(1)}(\mathbf{s}-\mathbf{s}', \phi)\rho^{(2)}(t-t', \psi) G_\gamma(d\phi, d\psi) \tag{19}$$

- Let's consider some nonseparable models using (infinite) spectral decomposition.

Spatiotemporal Gaussian Process*

- The spatiotemporal data $\{y_{ij} : i = 1, \cdots, I; j = 1, \cdots, J\}$ are usually modeled using the standard (separable) STGP model:

$$\begin{aligned} y_{ij} &= f(\mathbf{x}_i, t_j) + \epsilon_{ij}, \quad \epsilon_{ij} \overset{iid}{\sim} \mathcal{N}(0, \sigma^2_\epsilon) \\ f(\mathbf{z}) &\sim \mathcal{GP}(0, \mathcal{C}_\mathbf{z}), \quad \mathbf{z} := (\mathbf{x}, t) \end{aligned} \tag{20}$$

where the (joint) kernel $\mathcal{C}_\mathbf{z}$ has the following separability assumption

$$\text{model 0}: \quad \mathcal{C}_\mathbf{z} = \mathcal{C}_\mathbf{x} \otimes \mathcal{C}_t, \quad \mathcal{C}_\mathbf{x} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}, \mathcal{C}_t : \mathcal{T} \times \mathcal{T} \to \mathbb{R} \tag{21}$$

- However, conditioned on any $t \in \mathcal{T}$, the covariance of $f(\mathbf{x}, t)$ is static:

$$\mathrm{Cov}[f(\mathbf{x}, t), f(\mathbf{x}', t)] \propto \mathcal{C}_\mathbf{x}(\mathbf{x}, \mathbf{x}'), \quad \forall t \in \mathcal{T} \tag{22}$$

- Separable kernel fails to characterize TESD!

Two Generalizations*

- Generalization I: replace $\mathcal{C}_{\mathbf{x}}$ with $\mathcal{C}_{\mathbf{x}|t}$ in separable STGP

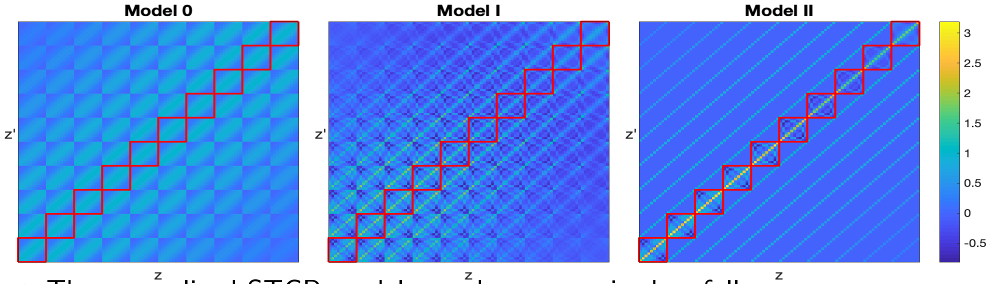
$$
\begin{aligned}
y(\mathbf{z}) = f(\mathbf{z}) + \epsilon, &\quad \epsilon \sim \mathcal{GP}(0, \sigma_\epsilon^2 \mathcal{I}_{\mathbf{z}}) \\
f(\mathbf{z}) \sim \mathcal{GP}(0, \mathcal{C}_{\mathbf{z}}), &\quad \mathcal{C}_{\mathbf{z}} = \mathcal{C}_{\mathbf{x}|t} \otimes \mathcal{C}_t
\end{aligned} \tag{23}
$$

- Generalization II: replace $\mathbf{\Sigma}_t$ with $\mathcal{C}_{\mathbf{x}|t}$ in dynamic covariance model

$$
\begin{aligned}
y_t(\mathbf{x}) | m_{\mathbf{x}|t}, \mathcal{C}_{\mathbf{x}|t} &\sim \mathcal{GP}_{\mathbf{x}}(m_{\mathbf{x}|t}, \mathcal{C}_{\mathbf{x}|t} \otimes \mathcal{I}_t) \\
m_{\mathbf{x}|t} &\sim \mathcal{GP}_t(0, \mathcal{I}_{\mathbf{x}} \otimes \mathcal{C}_t)
\end{aligned} \tag{24}
$$

- TESD could be characterized by $\mathcal{C}_{\mathbf{x}|t}$!

Two Kernels*



- The generalized STGP models can be summarized as follows

$$
\begin{aligned}
y(\mathbf{z})|m, \mathcal{C}_{y|m} &\sim \mathcal{GP}(m, \mathcal{C}_{y|m}) \\
m(\mathbf{z}) &\sim \mathcal{GP}(0, \mathcal{C}_m) \\
\text{model I}: \quad \mathcal{C}_{y|m} &= \underbrace{\sigma_\epsilon^2 \mathcal{I}_{\mathbf{x}} \otimes \mathcal{I}_t}_{likelihood}, \quad \mathcal{C}_m = \underbrace{\mathcal{C}_{\mathbf{x}|t} \otimes \mathcal{C}_t}_{prior} \\
\text{model II}: \quad \mathcal{C}_{y|m} &= \underbrace{\mathcal{C}_{\mathbf{x}|t} \otimes \mathcal{I}_t}_{likelihood}, \quad \mathcal{C}_m = \underbrace{\mathcal{I}_{\mathbf{x}} \otimes \mathcal{C}_t}_{prior}
\end{aligned}
\tag{25}
$$

Time-Dependent Spatial Kernel*

- By Mercer's theorem, we have the following representation of $\mathcal{C}_{\mathbf{x}}$

$$\mathcal{C}_{\mathbf{x}}(\mathbf{x}, \mathbf{x}') = \sum_{\ell=1}^{\infty} \lambda_\ell^2 \phi_\ell(\mathbf{x}) \phi_\ell(\mathbf{x}') \tag{26}$$

- Let $\{\lambda_\ell\}$ change with time, thus denoted as $\lambda(t) := \{\lambda_\ell(t)\}$. Assume

Assumption

$$\lambda \in \ell^2(L^2(\mathcal{T})), \quad i.e. \quad \|\lambda\|_{2,2}^2 := \sum_{\ell=1}^{\infty} \|\lambda_\ell(\cdot)\|_2^2 < +\infty \tag{27}$$

- We model λ_ℓ as random draws from independent GP's:

$$\lambda_\ell(\cdot) \sim \mathcal{GP}(0, \mathcal{C}_{\lambda,\ell}), \qquad \mathcal{C}_{\lambda,\ell} = \gamma_\ell^2 \mathcal{C}_u, \qquad \sum_{\ell=1}^{\infty} \gamma_\ell^2 < \infty \tag{28}$$

$$\lambda_\ell(t) = \gamma_\ell u_\ell(t), \qquad u_\ell(\cdot) \overset{iid}{\sim} \mathcal{GP}(0, \mathcal{C}_u) \qquad \text{for } \ell \in \mathbb{N} \tag{29}$$

Time-Dependent Spatial Kernel*

Theorem (Wellposedness of Mercer's Kernels)

Under Assumption 1, both of the following are well defined non-negative definite kernels on $\mathcal{Z} = \mathcal{X} \times \mathcal{T}$.

$$\mathcal{C}^{I}_{m}(\mathbf{z}, \mathbf{z}') = \mathcal{C}^{\frac{1}{2}}_{\mathbf{x}|t}\mathcal{C}^{\frac{1}{2}}_{\mathbf{x}|t'} \otimes \mathcal{C}_t(\mathbf{z}, \mathbf{z}') = \sum_{\ell=1}^{\infty} \lambda_\ell(t)\mathcal{C}_t(t, t')\lambda_\ell(t')\phi_\ell(\mathbf{x})\phi_\ell(\mathbf{x}')$$

$$\mathcal{C}^{II}_{y|m}(\mathbf{z}, \mathbf{z}') = \mathcal{C}_{\mathbf{x}|t} \otimes \mathcal{I}_t(\mathbf{z}, \mathbf{z}') = \sum_{\ell=1}^{\infty} \lambda^2_\ell(t)\delta(t = t')\phi_\ell(\mathbf{x})\phi_\ell(\mathbf{x}')$$

- Similar to "coregionalization" model Banerjee (2015) for the spatial process $\mathbf{Y}(\mathbf{x}) = \mathbf{A}\mathbf{w}(\mathbf{x})$ with $\mathbf{T} = \mathbf{A}\mathbf{A}^\mathsf{T}$ and $w_j(\cdot) \sim \mathcal{GP}(0, \rho_j)$ independent:

$$\mathrm{Cov}(\mathbf{Y}(\mathbf{x}), \mathbf{Y}(\mathbf{x}')) = \sum_{j=1}^{p} \rho_j(\mathbf{x} - \mathbf{x}')\mathbf{T}_j$$

where $\mathbf{T}_j = \mathbf{a}_j\mathbf{a}_j^\mathsf{T}$ with $\mathbf{a}_j$ being the j-th column of $\mathbf{A}$.

Table of Contents

Dynamic spatiotemporal models

- Now consider spatiotemporal data continuous in space and discrete in time.
- They can be viewed as a time series of spatial processes and allow straightforward computation using Kalman filtering.
- Dynamic linear models, often referred to as state-space models in the time-series literature.
- Let $\mathbf{Y}_t$ be an $m \times 1$ vector of observables at time t, $\boldsymbol{\theta}_t$ be a $p \times 1$ state vector. Consider the Gaussian-linear state space model

$$\mathbf{Y}_t = F_t \boldsymbol{\theta}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(0, \Sigma_t^{\epsilon}) \tag{30}$$

$$\boldsymbol{\theta}_t = G_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \Sigma_t^{\eta}) \tag{31}$$

where F_t and G_t are $m \times p$ and $p \times p$ matrices, respectively.

- Note the covariance matrices

$$\mathrm{Cov}(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t-1}) = G_t \mathrm{Var}(\boldsymbol{\theta}_{t-1}), \quad \mathrm{Cov}(\mathbf{Y}_t, \mathbf{Y}_{t-1}) = F_t G_t \mathrm{Var}(\boldsymbol{\theta}_{t-1}) F_t^T \tag{32}$$

Table of Contents

Areal unit space-time modeling

- Let's return to spatiotemporal modeling for areal unit data with discrete time.
- Here, we focus on the spatiotemporal disease mapping setting.
- Denote $Y_{i\ell t}$ and $E_{i\ell t}$ as the observed and expected disease counts in county i and demographic subgroup ℓ (race, gender, etc.) during time period t. Denote $n_{i\ell t}$ as the number of persons at risk in county i during year t. With internally standardization, $E_{i\ell t} = n_{i\ell t}(\sum_{i\ell t} Y_{i\ell t} / \sum_{i\ell t} n_{i\ell t})$.
- Consider the Poisson regression model

$$Y_{i\ell t}|\mu_{i\ell t} \overset{ind}{\sim} \mathrm{Pois}(E_{i\ell t} e^{\mu_{i\ell t}}) \tag{33}$$

  where $\mu_{i\ell t}$ is the log-relative risk of disease for region i, subgroup ℓ, and year t.
- We need to specify the main effect and interaction components of $\mu_{i\ell t}$.

Areal unit space-time modeling

- The main effect can consist of a demographic part $\epsilon_\ell = \mathbf{x}_\ell'\boldsymbol{\beta}$ (linear regression) and a temporal part δ_t (AR(1)).
- The spatiotemporal interactions $\psi_{it} = \mathbf{z}_i'\mathbf{w} + \theta_{it} + \phi_{it}$ can be modeled using the nested model

$$\theta_{it} \overset{iid}{\sim} N(0, 1/\tau_t), \quad \phi_{it} \sim \mathrm{CAR}(\lambda_t) \tag{34}$$

where the θ_{it} capture *heterogeneity* and the ϕ_{it} capture regional *clustering*; $\tau_t \overset{iid}{\sim} \Gamma(a, b)$ and $\lambda_t \overset{iid}{\sim} \Gamma(c, d)$.
- The most general model for $\mu_{i\ell t}$ is

$$\mu_{i\ell t} = \mathbf{x}_\ell'\boldsymbol{\beta} + \delta_t + \mathbf{z}_i'\mathbf{w} + \theta_{it} + \phi_{it} \tag{35}$$

with corresponding joint posterior distribution proportional to

$$L(\boldsymbol{\beta}, \delta, \mathbf{w}, \boldsymbol{\theta}, \phi; \mathbf{y}) p(\delta) p(\boldsymbol{\theta}|\tau) p(\tau) p(\lambda) \tag{36}$$

Table of Contents

Areal-level continuous time modeling

- Finally, we consider the less common setting where space is discrete and time is continuous.
- Quick et al. (2013) propose a class of Bayesian space-time models based upon a dynamic MRF that evolves continuously over time. Consider

$$Y_i(t) = \mu_i(t) + Z_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \overset{ind}{\sim} N(0, \tau_i^2), \; \textit{for } i = 1, 2, \cdots, N_s \tag{37}$$

  where $\mu_i(t)$ captures large scale variation or trends, $Z_i(t)$ is an underlying areally-referenced stochastic process over time that captures smaller-scale variations in the time scale while also accommodating spatial associations.
- A temporally evolving MRF for the areal units at any time is specified through the full conditional of $Z_i(t)$

$$p(Z_i(t)|Z_{j\neq i}(t)) \sim N\left(\sum_{j\sim i} \alpha \frac{w_{ij}}{w_{i+}} Z_j(t), \frac{\sigma^2}{w_{i+}}\right) \tag{38}$$

  where $w_{i+} = \sum_{j\sim i} w_{ij}$, $\sigma^2 > 0$.

Areal-level continuous time modeling

- Denote $\mathbf{Z}(t) = (Z_1(t), \cdots, Z_{N_s}(t))^T$. Then we have

$$\mathbf{Z}(t) \sim N(\mathbf{0}, \sigma_t^2 (D - \alpha_t W)^{-1}) \tag{39}$$

- Alternatively we could use a constructive approach similar to that used in linear models of coregionalization (LMC): $\mathbf{Z}(t) = A(t)\mathbf{v}(t)$. Then the covariance of $\mathbf{Z}(t)$, $K_Z(t, u)$ is

$$K_Z(t, u) = \rho(t, u; \phi) A(t) A(t)^T \tag{40}$$

- If $A(t) = A$ be some square-root (e.g. Cholesky) of the $N_s \times N_s$ dispersion matrix $\sigma^2 (D - \alpha W)^{-1}$, then

$$K_Z(t, u) = \sigma^2 \rho(t, u; \phi)(D - \alpha W)^{-1}, \quad \Sigma_Z = R(\phi) \otimes \sigma^2 (D - \alpha W)^{-1} \tag{41}$$