Stationary spatial process models
Generalized linear spatial process modeling
Areal data models
Generalized linear areal data modeling

Lecture 5 Hierarchical modeling of spatial data

Shiwei Lan[1]

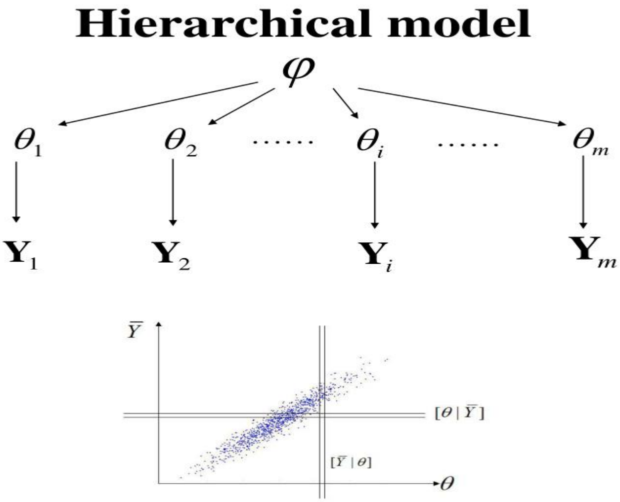
[1]School of Mathematical and Statistical Sciences
Arizona State University

STP598 Spatiotemporal Analysis
Fall 2020

Table of Contents

1. Stationary spatial process models
2. Generalized linear spatial process modeling
3. Areal data models
4. Generalized linear areal data modeling

Bayesian hierarchical modeling



Stationary spatial process models

- The basic model is
$$Y(\mathrm{s}) = \mu(\mathrm{s}) + w(\mathrm{s}) + \epsilon(\mathrm{s}) \tag{1}$$
- The mean structure $\mu(\mathrm{s}) = \mathrm{x}^T \boldsymbol{\beta}$.
- The residual is partitioned into two parts: the spatial part $w(\mathrm{s})$ and the non-spatial errors $\epsilon(\mathrm{s})$.
- Recall that the $w(\mathrm{s})$ introduces the partial sill (σ^2) and the range ϕ; while the $\epsilon(\mathrm{s})$ adds the nugget τ^2.
- Assume stationarity. Pure error $\epsilon(\mathrm{s})$ vs spatial error $w(\mathrm{s})$. We further assume
  - $w(\mathrm{s}+\mathrm{h}) - w(\mathrm{s}) \to 0$ as $\mathrm{h} \to 0$
  - $[w(\mathrm{s}+\mathrm{h}) + \epsilon(\mathrm{s}+\mathrm{h})] - [w(\mathrm{s}) + \epsilon(\mathrm{s})] \not\to 0$ as $\mathrm{h} \to 0$
- *Microscale* view: $eps(\mathrm{s})$ is a spatial process with very rapid decay in association, and only matters at high resolution.

Isotropic models

- Suppose we have data $Y(\mathrm{s}_i)$, $i = 1, \cdots, n$, and let $\mathrm{Y} = (Y(\mathrm{s}_1), \cdots, Y(\mathrm{s}_n))^T$. In the basic Gaussian isotropic kriging model, we assume

$$\Sigma = \sigma^2 H(\phi) + \tau^2 I \tag{2}$$

where H is correlation matrix with $H_{ij} = \rho(\mathrm{s}_i - \mathrm{s}_j; \phi)$ and ρ is a valid isotropic correlation function on $\mathbb{R}^2$, e.g. $\rho(\mathrm{s}_i - \mathrm{s}_j; \phi) = \exp(-\phi \|\mathrm{s}_i - \mathrm{s}_j\|)$

- Collecting all model parameters into a vector $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma^2, \tau^2, \phi)$, we have

$$p(\boldsymbol{\theta}|\mathrm{y}) \propto f(\mathrm{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta}) \tag{3}$$

where

$$\mathrm{Y}|\boldsymbol{\theta} \sim N(X\boldsymbol{\beta}, \sigma^2 H(\phi) + \tau^2 I) \tag{4}$$

Prior specification

- Typically, independent priors are chosen for the different parameters

$$p(\boldsymbol{\theta}) = p(\boldsymbol{\beta})p(\sigma^2)p(\tau^2)p(\phi) \tag{5}$$

- We usually adopt the following priors

$$\begin{aligned} \beta &\sim N_{p+1}(\mu_0, \Lambda_0) && (6) \\ \sigma^2 &\sim \Gamma^{-1}(a_1, b_1) && (7) \\ \tau^2 &\sim \Gamma^{-1}(a_2, b_2) && (8) \\ \phi &\sim \Gamma(a_3, b_3) && (9) \end{aligned}$$

- In Matérn class, the product $\sigma^2\phi^{2\nu}$ can be identified but not the individuals (Zhang, 2004). Often, a very informative prior is imposed for ϕ (e.g. uniform over interval) and a relatively vague prior is used for σ^2.

Prior specification

- When $\tau^2 \equiv 0$, with the exponential covariance, Berger et al. (2001) impose the *objective* priors of the form

$$p(\boldsymbol{\beta}, \sigma^2, \phi) \propto \frac{p(\phi)}{(\sigma^2)^\alpha} \tag{10}$$

- This implies a flat prior for $\boldsymbol{\beta}$. With a uniform prior for ϕ, it can be shown that an improper posterior arises for $\alpha < 2$.
- If $\sigma^2 \sim \Gamma^{-1}(\epsilon, \epsilon)$, then $\alpha = 1 + \epsilon$ getting an improper posterior for small ϵ.
- If $\sigma^2 \sim \Gamma^{-1}(a_1, b_1)$, $a_1 \geq 1$ is recommended $\alpha \geq 2$.
- Closed form of *marginal* posterior of $\boldsymbol{\beta}$ may be available (e.g. with Gaussian likelihood).

Hierarchical modeling

- The hierarchical modeling shares the following generic format

$$[\text{data}|\text{process}, \text{parameters}]\ [\text{process}|\text{parameters}]\ [\text{parameters}] \tag{11}$$

- The previous spatial process model (4) can be recast as a hierarchical model with two stages

$$\mathsf{Y}|\boldsymbol{\theta}, \mathsf{W} \sim N(X\boldsymbol{\beta} + \mathsf{W}, \tau^2 I) \tag{12}$$
$$\mathsf{W}|\sigma^2, \phi \sim N(0, \sigma^2 H(\phi) + \tau^2 I) \tag{13}$$

where $\mathsf{W} = (w(\mathsf{s}_1), \cdots, w(\mathsf{s}_n))^T$ is the spatial random effect that captures spatial dependence.

- The parameter space is now augmented from $\boldsymbol{\theta}$ to $(\boldsymbol{\theta}, \mathsf{W})$, with the dimension increased by n.

Hierarchical modeling

- The resulting $p(\boldsymbol{\theta}|\mathrm{y})$ is the same, but we have the choice of using MCMC to fit either $f(\mathrm{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$ or $f(\mathrm{y}|\boldsymbol{\theta},\mathrm{W})p(\mathrm{W}|\boldsymbol{\theta})p(\boldsymbol{\theta})$.
- Interest is often in the spatial surface $\mathrm{W}|\mathrm{y}$ as well as prediction for $W(\mathrm{s}_0)|\mathrm{y}$ for various choices of s_0.
- Note $p(\mathrm{W}|\mathrm{y})$ can be recovered from $p(\boldsymbol{\theta}|\mathrm{y})$:

$$p(\mathrm{W}|\mathrm{y}) = \int p(\mathrm{W}|\boldsymbol{\theta},\mathrm{y})p(\boldsymbol{\theta}|\mathrm{y})d\boldsymbol{\theta} \tag{14}$$

which can be sampled by one for one composition of posterior sampling: $\mathrm{W}^{(g)} \sim p(\mathrm{W}|\boldsymbol{\theta}^{(g)},\mathrm{y})$ with $\boldsymbol{\theta}^{(g)} \sim p(\boldsymbol{\theta}|\mathrm{y})$.

Hierarchical modeling

- Bayesian kriging involves prediction of $Y_0 \equiv Y(\mathrm{s}_0)$ at a new location s_0 with associated covariate $\mathrm{x}(\mathrm{s}_0)$:

$$p(y_0|\mathrm{y}, X, \mathrm{x}_0) = \int p(y_0, \boldsymbol{\theta}|\mathrm{y}, X, \mathrm{x}_0) d\boldsymbol{\theta} = \int p(y_0|\mathrm{y}, \boldsymbol{\theta}, \mathrm{x}_0) p(\boldsymbol{\theta}|\mathrm{y}, X) d\boldsymbol{\theta} \quad (15)$$

- In practice, we use MCMC to obtain posterior samples $\{\boldsymbol{\theta}^{(g)}\}_{g=1}^G$ with $\boldsymbol{\theta}^{(g)} \sim p(\boldsymbol{\theta}|\mathrm{y}, X)$, and approximate the above predictive distribution with

$$\hat{p}(y_0|\mathrm{y}, X, \mathrm{x}_0) = \frac{1}{G} \sum_{g=1}^G p(y_0|\mathrm{y}, \boldsymbol{\theta}^{(g)}, \mathrm{x}_0) \quad (16)$$

- Replicates of prediction $y_0^{(g)}$ can be obtained using the composition sampling.
- Multi-output prediction for $\mathrm{Y}_0 = (Y(\mathrm{s}_{01}), \cdots, Y(\mathrm{s}_{0m}))^T$ is also available

$$p(\mathrm{y}_0|\mathrm{y}, X, \mathrm{x}_0) = \int p(\mathrm{y}_0|\mathrm{y}, \boldsymbol{\theta}, \mathrm{x}_0) p(\boldsymbol{\theta}|\mathrm{y}, X) d\boldsymbol{\theta} \approx \frac{1}{G} \sum_{g=1}^G p(\mathrm{y}_0|\mathrm{y}, \boldsymbol{\theta}^{(g)}, \mathrm{x}_0) \quad (17)$$

Table of Contents

1. Stationary spatial process models
2. Generalized linear spatial process modeling
3. Areal data models
4. Generalized linear areal data modeling

Hierarchical Model

S.Lan

Stationary spatial process models

Generalized linear spatial process modeling

Areal data models

Generalized linear areal data modeling

Generalized linear spatial process modeling

- In some point-referenced data sets, measurements $Y(\mathbf{s})$ are not naturally modeled as a normal distribution. e.g. binary data, and counting data.
- The observations $Y(\mathbf{s}_i)$ are modeled independent conditioned on $\boldsymbol{\beta}$ and $w(\mathbf{s}_i)$ within the class of exponential family:

$$p(y(\mathbf{s}_i)|\boldsymbol{\beta}, w(\mathbf{s}_i), \gamma) = h(y(\mathbf{s}_i), \gamma) \exp\{\gamma[y(\mathbf{s}_i)\eta(\mathbf{s}_i) - \psi(\eta(\mathbf{s}_i))]\} \tag{18}$$

where $g(\eta(\mathbf{s}_i)) = \mathbf{x}^T(\mathbf{s}_i)\boldsymbol{\beta} + w(\mathbf{s}_i)$ for some link function g, and γ is a dispersion parameter.

- We presume the $w(\mathbf{s}_i)$ to be the spatial random effect coming from a Gaussian process, i.e. $\mathbf{W} \sim N(0, \sigma^2 H(\phi) + \tau^2 I)$.

Generalized linear spatial process modeling

- Using conditional independence, we have the joint distribution

$$f(y(\mathrm{s}_1), \cdots, y(\mathrm{s}_n)|\boldsymbol{\beta}, \sigma^2, \phi, \gamma) = \int \left(\prod_{i=1}^{n} f(y(\mathrm{s}_i)|\boldsymbol{\beta}, w(\mathrm{s}_i), \gamma) \right) p(\mathrm{W}|\sigma^2, \phi) d\mathrm{W} \tag{19}$$

- We presume the $w(\mathrm{s}_i)$ to be the spatial random effect coming from a Gaussian process, i.e. $\mathrm{W} \sim N(0, \sigma^2 H(\phi) + \tau^2 I)$. Note W may not be marginalized out in general.
- Binary data. We model $Y(\mathrm{s})$ through the latent process $Z(\mathrm{s}) = \mathrm{x}(\mathrm{s})^T \boldsymbol{\beta} + w(\mathrm{s}) + \epsilon(\mathrm{s})$:

$$\Pr(Y(\mathrm{s}) = 1) = \Pr(Z(\mathrm{s}) \geq 0) = g^{-1}(\mathrm{x}(\mathrm{s})^T \boldsymbol{\beta} + w(\mathrm{s})) \tag{20}$$

for some link function $g(\cdot)$ such that g^{-1} takes $p(\mathrm{s}) := \Pr(Y(\mathrm{s}) = 1)$ to $\mathbb{R}^1$.

- Possible choices: the logit $g(x) = \log \frac{x}{1-x}$ and the probit $g(x) = \Phi^{-1}(x)$.

Generalized linear spatial process modeling

- Consider real estate data at 50 locations in Baton Rouge, LA.
- $Y(s) = 1$ indicates that the price of the property at location s is "high" (above the median for the region); $Y(s) = 0$ indicates that the price is "low".
- Covariates: the house's age, total living area, and other area in the property.
- We fit the model with the logit link, assuming vague priors for $\boldsymbol{\beta}$, a $\mathrm{unif}(0, 10)$ prior for ϕ and a $\Gamma^{-1}(0.1, 0.1)$ prior for σ^2.

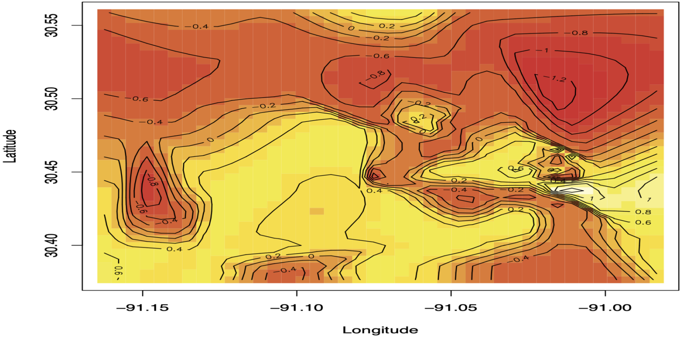


Figure 6.7 Image plot of the posterior median surface of the latent spatial process w(s), binary

Generalized linear spatial process modeling

- Counting data. We model $Y(\mathrm{s})$ through the latent process $Z(\mathrm{s}) = \mathrm{x}(\mathrm{s})^T\boldsymbol{\beta} + w(\mathrm{s}) + \epsilon(\mathrm{s})$:

$$Y(\mathrm{s}) \sim \mathrm{pois}(g^{-1}(Z(\mathrm{s}))) \tag{21}$$

for some link function $g(\cdot)$, e.g. the canonical link $g(x) = \log(x)$.

- This is related to the log-Gaussian Cox Process (LGCP) model.

Table of Contents

1. Stationary spatial process models
2. Generalized linear spatial process modeling
3. Areal data models
4. Generalized linear areal data modeling

Disease mapping

- An area of strong biostatistical and epidemiological interest is that of *disease mapping*.
- We typically have count data of the following:

$$\begin{aligned} Y_i &= \text{observed number of cases of disease in county } i, i = 1, \cdots, I \\ E_i &= \text{expected number of cases of disease in county } i, i = 1, \cdots, I \end{aligned}$$

  where E_i are thought of as fixed and known functions of n_i, the number of persons at risk for the disease in county i.
- One can simply assume

$$E_i = n_i \bar{r} = n_i \frac{\sum_i y_i}{\sum_i n_i} = \sum_i y_i \frac{n_i}{\sum_i n_i} \tag{22}$$

  i.e. $\bar{r}$ is the overall disease rate in the entire study region.

Disease mapping

- However, such *internal standardization* is “cheating”. And one might modify it by age-adjusted rates for the disease

$$E_i = n_{ij} r_j \tag{23}$$

where n_{ij} is the person-years at risk in area i for age group j, and $\bar{r}_j$ is the disease rate in age group j. This process is called *external standardization*.

- The usual model for Y_i is the Poisson model

$$Y_i | \eta_i \sim \mathrm{Pois}(E_i \eta_i) \tag{24}$$

where η_i is the true *relative risk* of disease in region i.

- The maximum likelihood estimate (MLE) of η_i is

$$\hat{\eta}_i = \mathrm{SMR}_i = \frac{Y_i}{E_i} \tag{25}$$

- $\mathrm{Var}(\mathrm{SMR}_i) = \mathrm{Var}(Y_i)/E_i^2 = \eta_i/E_i$. $\widehat{\mathrm{Var}}(\mathrm{SMR}_i) = \hat{\eta}_i/E_i = Y_i/E_i^2$.

Hierarchical Bayesian methods

- For detecting extra-Poisson variability (overdispersion) in the observed rates, we seek *random effects* model for η_i through hierarchical Bayesian modeling.
- Poisson-gamma model

$$Y_i|\eta_i \overset{ind}{\sim} \mathrm{pois}(E_i\eta_i), \quad i = 1, \cdots, I \tag{26}$$

$$\eta_i \overset{iid}{\sim} \Gamma(a, b) \tag{27}$$

Table of Contents

1. Stationary spatial process models
2. Generalized linear spatial process modeling
3. Areal data models
4. Generalized linear areal data modeling