Lecture 3 Areal Data Models

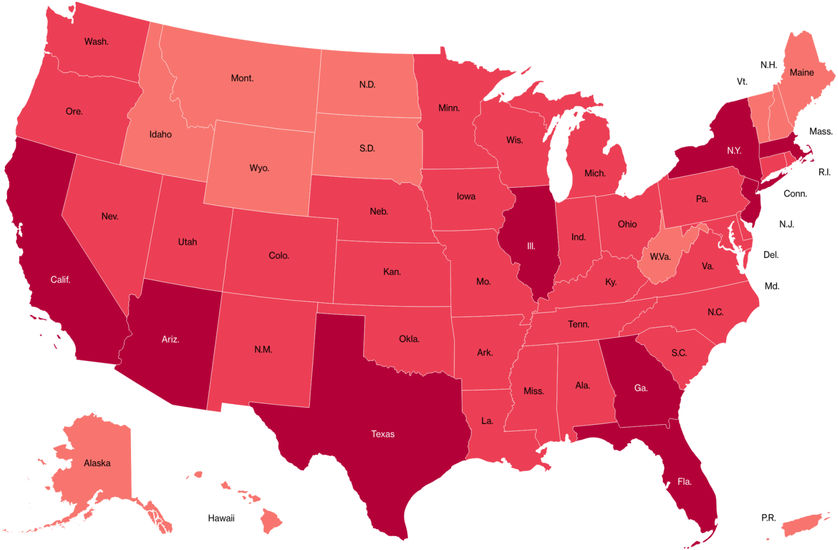
Shiwei Lan[1]

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

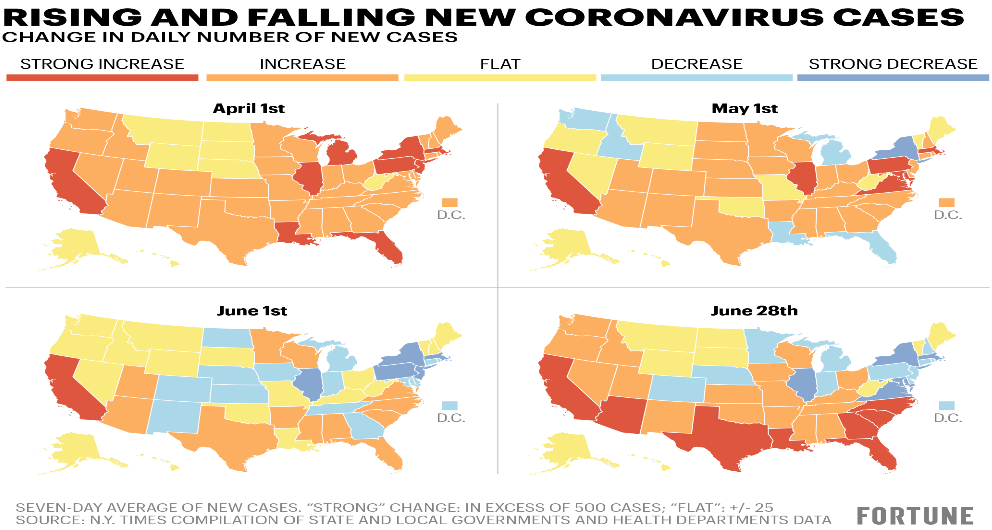
STP598 Spatiotemporal Analysis
Fall 2020

Table of Contents

Spatial Problems



Spatiotemporal Problems



Areal data models

In the context of areal units the general inferential issues are the following:

1. Is there spatial pattern? If so, how strong is it?
2. Do we want to smooth the data? If so, how much?
3. For a new areal unit or set of units, how can we infer about what data values we expect to be associated with these units? This is the so-called *modifiable areal unit problem (MAUP)*.

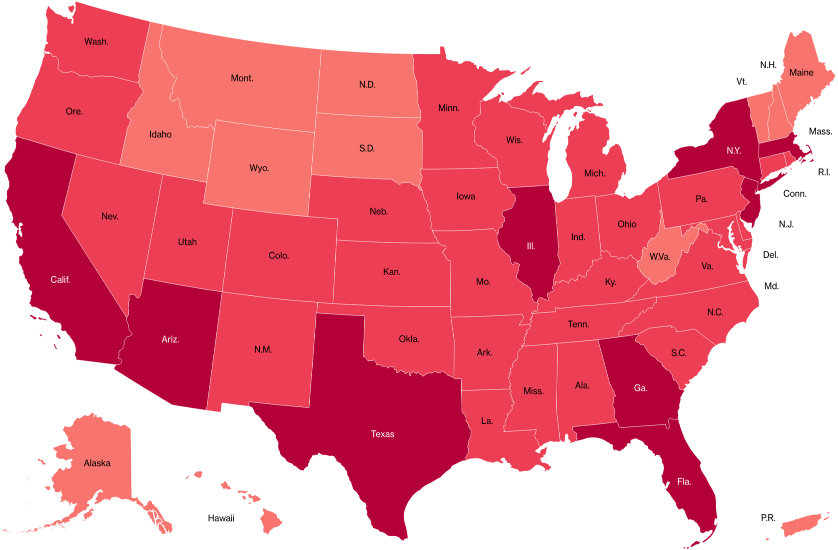
We will explore both descriptive and model-based approaches in this lecture.

Table of Contents

Exploratory data analysis (EDA)

- The primary concept *proximity matrix* W for areal units $1, 2, \cdots, n$ is defined by setting entries w_{ij} spatially connect units i and j ($w_{ii} = 0$).
- Binary choice: $w_{ij} = 1$ if i and j share common boundary; otherwise 0.
- 'Distance': e.g. decreasing function of intercentroidal distance between the units, binary values based on truncated distance or K nearest neighborhood.
- W can be standardized as $\widetilde{W}$ with $\tilde{w}_{ij} = w_{ij}/w_{i+}$ where $w_{i+} = \sum_j w_{ij}$. $\widetilde{W}$ is row stochastic, i.e. $\widetilde{W}1 = 1$.
- Divide distances into bins $(0, d_1], (d_1, d_2], \cdots$ and define k-th order neighbors of unit i as all units with distances in $(d_{k-1}, d_k]$. We can define k-th order proximity matrix $W^{(k)}$ based on k-th order neighbors.

Exploratory data analysis (EDA)



Exploratory data analysis (EDA)

- There are two standard statistics to measure the spatial association (Ripley, 1981).
- Moran's I:

$$I = \frac{n \sum_i \sum_j w_{ij}(Y_i - \bar{Y})(Y_j - \bar{Y})}{\left(\sum_{i \neq j} w_{ij}\right) \sum_i (Y_i - \bar{Y})^2} \tag{1}$$

- Under the null model where Y_i are i.i.d., $I \dot{\sim} N(-1/(n-1), \mathrm{Var}(I))$ with

$$\mathrm{Var}(I) = \frac{n^2(n-1)S_1 - n(n-1)S_2 - 2S_0^2}{(n+1)(n-1)^2 S_0^2}$$

where $S_0 = \sum_{i \neq j} w_{ij}$, $S_1 = \frac{1}{2}\sum_{i \neq j}(w_{ij} + w_{ji})^2$, $S_2 = \sum_k (\sum_j w_{kj} + \sum_i w_{ik})^2$.
- Geary's C:

$$C = \frac{n \sum_i \sum_j w_{ij}(Y_i - Y_j)^2}{\left(\sum_{i \neq j} w_{ij}\right) \sum_i (Y_i - \bar{Y})^2} \tag{2}$$

- $C \dot{\sim} N(1, \mathrm{Var}(C))$ under the null model.

Exploratory data analysis (EDA)

Figure 3.2 *Rasterized north and south regions (1 km × 1 km) with binary land use classification overlaid.*

Exploratory data analysis (EDA)

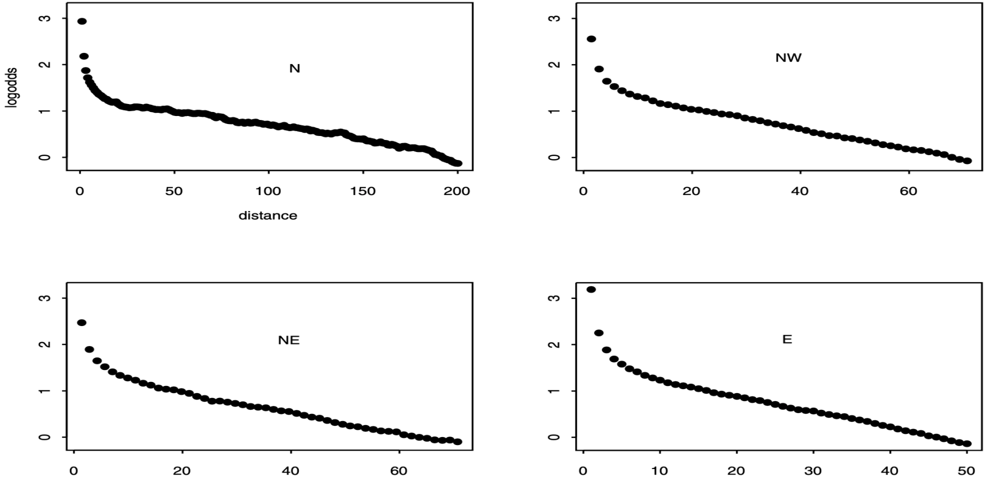


Figure 3.3 *Land use log-odds ratio versus distance in four directions.*

Exploratory data analysis (EDA)

- One could also investigate a choropleth map by smoothing Y_i's.
- The proximity matrix W provides a smoother: $\widehat{Y}_i = \sum_j w_{ij} Y_j / w_{i+}$.
- However, $\widehat{Y}_i$ ignores Y_i. We might revise it to be

$$\widehat{Y}_i^* = (1-\alpha) Y_i + \alpha \widehat{Y}_i \tag{3}$$

  where $\alpha \in (0, 1)$.
- One can refer to general *filters*.

Table of Contents

Full conditional distribution

- Given $p(y_1, \cdots, y_n)$, the so-called *full conditional* distributions, $p(y_i | y_j, j \neq i)$, $i = 1, \cdots, n$, are uniquely determined.
- Brook's lemma (1964) proves the converse and constructively retrieve the unique joint distribution from these full conditionals.
- *Compatible* conditionals, *proper* conditionals (improper joint).
- Brook's lemma:

$$p(y_1, \cdots, y_n) = \frac{p(y_1 | y_2, \cdots, y_n)}{p(y_{10} | y_2, \cdots, y_n)} \cdot \frac{p(y_2 | y_{10}, y_3, \cdots, y_n)}{p(y_{20} | y_{10}, y_3, \cdots, y_n)} \cdots \frac{p(y_n | y_{10}, , \cdots, y_{n-1,0})}{p(y_{n0} | y_{10}, , \cdots, y_{n-1,0})} \cdot p(y_{10}, \cdots, y_{n0}) \tag{4}$$

- Denote ∂_i as the set of neighbors of unit i. An areal process Y_i is referred as a *Markov random field (MRF)* (Besag 1974, Kaiser and Cressie 2000) if

$$p(y_i | y_j, j \neq i) = p(y_i | y_j, j \in \partial_i) \tag{5}$$

Gibbs distribution

- A *clique* is a set of cells (indices) such that each element is a neighbor of every other element.
- A *potential* of order k is a function of k exchangeable arguments.
- $p(y_1, \cdots, y_n)$ is a *Gibbs distribution* if it is a function of the Y_i only through potentials on cliques:

$$p(y_1, \cdots, y_n) \propto \exp\left\{\gamma \sum_k \sum_{\boldsymbol{\alpha}\in\mathcal{M}_k} \phi^{(k)}(y_{\alpha_1}, y_{\alpha_2}, \cdots, y_{\alpha_k})\right\} \tag{6}$$

where $\phi^{(k)}$ is a potential of order k, $\mathcal{M}_k$ is the collection of all subsets of size k from $\{1, 2, \cdots, n\}$, $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_k)$ indexes this set, and $\gamma > 0$ is a scale parameter.

Gibbs distribution

- The *Hammersley-Clifford Theorem* (Clifford 1990) demonstrates that if we have an MRF, then its joint distribution is a Gibbs distribution.
- Geman and Geman (1984) provides essentially the converse of the Hammersley-Clifford theorem: A Gibbs distribution determines an MRF.
- Sampling a MRF is reduced to sampling its associated Gibbs distribution, hence coining the term 'Gibbs sampler'.
- With cliques of order 1, we consider for continuous data on $\mathbb{R}^1$

$$p(y_1, \cdots, y_n) \propto \exp\left\{ -\frac{1}{2\tau^2} \sum_{i,j} (y_i - y_j)^2 I(i \sim j) \right\} \tag{7}$$

- It is a Gibbs distribution on potentials of order 1 and 2 and that

$$p(y_i | y_j, j \neq i) = N\left(\sum_{j \in \partial_i} y_j / m_i, \tau^2 / m_i \right) \tag{8}$$

Table of Contents

Conditionally autoregressive (CAR) models

- We begin with the Gaussian (*autonormal*) case. Suppose

$$Y_i | y_j, j \neq i \sim N\left(\sum_j b_{ij} y_j, \tau_i^2\right), \quad i = 1, \cdots, n \tag{9}$$

- By Brook's Lemma, we have

$$p(y_1, \cdots, y_n) \propto \exp\left\{-\frac{1}{2}\mathrm{y}' D^{-1}(I - B)\mathrm{y}\right\} \tag{10}$$

where $B = (b_{ij})$ and $D = \mathrm{diag}\{\tau_i^2\}$.

- $\mathrm{Y} \sim N(0, \Sigma_{\mathrm{y}} = (I - B)^{-1} D)$?

$$\frac{b_{ij}}{\tau_i^2} = \frac{b_{ji}}{\tau_j^2} \quad \textit{for all } i, j \tag{11}$$

Conditionally autoregressive (CAR) models

- Setting $b_{ij} = w_{ij}/w_{i+}$ and $\tau_i^2 = \tau^2/w_{i+}$, we have

$$p(y_i|y_j, j \neq i) = N\left(\sum_j w_{ij}y_j/w_{i+}, \tau^2/w_{i+}\right) \tag{12}$$

- Therefore we have the joint distribution (intrinsically autoregressive, IAR)

$$p(y_1, \cdots, y_n) \propto \exp\left\{-\frac{1}{2\tau^2}\mathrm{y}'(D_w - W)\mathrm{y}\right\} = \exp\left\{-\frac{1}{2\tau^2}\sum_{i\neq j} w_{ij}(y_i - y_j)^2\right\} \tag{13}$$

where $D_w = \mathrm{diag}\{w_{i+}\}$.

- $(D_w - W)1 = 0$. $\Sigma_{\mathrm{y}} =$?
- Redefine $\Sigma_{\mathrm{y}}^{-1} = D_w - \rho W > 0$ for chosen $\rho \in (1/\lambda_{(1)}, 1/\lambda_{(n)})$.

Conditionally autoregressive (CAR) models

- Rewriting the autonormal model

$$\mathrm{Y} = B\mathrm{Y} + \epsilon \tag{14}$$

- If $p(\mathrm{y})$ is proper, then:
  - $\mathrm{Y} \sim N(0, (I-B)^{-1}D)$, $\epsilon \sim N(0, D(I-B)^T)$, and $\mathrm{Cov}(\epsilon, \mathrm{Y}) = D$.
  - $1/(\Sigma_{\mathrm{y}}^{-1})_{ii} = \mathrm{Var}(Y_i | Y_j, j \neq i) = \tau_i^2$.
  - $(\Sigma_{\mathrm{y}}^{-1})_{ij} = b_{ij} = 0$ implies $Y_i \perp Y_j | Y_k, k \neq i, j$. We have control on conditional independence (by setting $w_{ij} = 0$)!
- One can introduce regression component to CAR.
- Considering a vector of dependent areal units leads to MCAR model.
- CAR model can be applied to point-level data.

Conditionally autoregressive (CAR) models

- We could also consider non-Gaussian case.

$$p(y_i|y_j, j \neq i) = \exp\left(\{\psi(\theta_i y_i - \chi(\theta_i))\}\right) \tag{15}$$

where $\theta_i = \sum_{j\neq i} b_{ij} y_j$.

- Autologistic model:

$$\log \frac{P(Y_i = 1)}{P(Y_i = 0)} = \mathrm{x}_i^T \gamma + \psi \sum w_{ij} y_j \tag{16}$$

which implies

$$p(y_1, \cdots, y_n) \propto \exp\left(\gamma^T (\sum_i y_i \mathrm{x}_i) + \psi \sum_{i,j} w_{ij} y_i y_j\right) \tag{17}$$

- Potts model

$$P(Y_i = l|Y_j, j \neq i) \propto \exp\left(\psi \sum_{i,j} w_{ij} I(Y_j = l)\right) \tag{18}$$

Table of Contents

Simultaneous autoregressive (SAR) models

- Now we start from $\boldsymbol{\epsilon} \sim N(0, \tilde{D})$ with $\tilde{D} = \mathrm{diag}\{\sigma_i^2\}$. Then

$$\mathrm{Y} = B\mathrm{Y} + \boldsymbol{\epsilon} \sim N(0, (I-B)^{-1}\tilde{D}(I-B)^{-T}) \tag{19}$$

  where $\mathrm{Cov}(\boldsymbol{\epsilon}, \mathrm{Y}) = \tilde{D}(I-B)^{-1}$.
- $(I - B)$ must be full rank:
  1. $B = \rho W$, W the contiguity matrix $w_{ij} = I(i \sim j)$. $Y_i = \rho \sum_j Y_j I(j \in \partial_i) + \epsilon_i$ with *spatial autoregressive parameter* $\rho \in (1/\lambda_{(1)}, 1/\lambda_{(n)})$.
  2. $B = \alpha \widetilde{W}$, with *spatial autocorrelation parameter* $\alpha \in (-1, 1)$.
- SAR model is introduced in a regression context and is applied to the *residuals* $\mathrm{U} = \mathrm{Y} - X\boldsymbol{\beta}$:

$$\mathrm{U} = B\mathrm{U} + \boldsymbol{\epsilon} \tag{20}$$

Simultaneous autoregressive (SAR) models

- The overall model is then written as follows with B interpolating between an OLS ($B = 0$) regression and a purely spatial model:

$$\mathrm{Y} = B\mathrm{Y} + (I - B)X\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{21}$$

- Assuming $\tilde{D} = \sigma^2 I$, the log-likelihood can be efficiently calculated thus amenable to MLE

$$\frac{1}{2}\log|\sigma^{-1}(I - B)| - \frac{1}{2\sigma^2}(\mathrm{Y} - X\boldsymbol{\beta})^T(I - B)(I - B)^T(\mathrm{Y} - X\boldsymbol{\beta}) \tag{22}$$

- Extendable to Bayesian setting. No convenient form for full conditional distributions as in CAR.

CAR vs SAR

- Both are spatial models for areal data.
- They are equivalent iff

$$(I - B)^{-1} D = (I - B)^{-1} \tilde{D} (I - B)^{-T} \tag{23}$$

- Cressie (1993) shows that any SAR model can represented as a CAR model; but not vice versa.
- The first-order neighbor correlations increase at a slower rate as a function of ρ in the CAR model than in SAR model.
- Gibbs sampler is usually used for CAR but likelihood based inference is used for SAR.