

COA 616 Geostatistics in Environmental Sciences

Lecture 11 – Modelling areal data

Wei Wu

November 27, 2018

Modeling areal data (SAR)

Simultaneous autoregressive model (SAR): Uses a regression on the values from the other areas to account for the spatial dependence. The error terms ε are modeled so that they depend on each other in the following way:

$$Y = \mu + e$$

$$e_i = \sum_{j=1}^m b_{ij} e_j + \varepsilon_i$$

ε_i represents residual errors, $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m) \sim N(\vec{0}, \vec{\Lambda})$ with $\vec{\Lambda}$ diagonal

b_{ij} values are used to represent spatial dependence between areas,

b_{ii} must be set to 0 so that each area is not regressed on itself.

If we express the error terms as $e = B(Y - X^T \beta) + \varepsilon$, the model can also be expressed as

$$Y = X^T \beta + B(Y - X^T \beta) + \varepsilon$$

Hence, the model can be formulated in a matrix form as follows :

$$(I - B)(Y - X^T \beta) = \varepsilon$$

where $B = (b_{ij})$, I is the m dimensional identity matrix

$I - B$ must be non - singular

SAR cont.

$$[Z] \sim N(X^T \beta, (I - B)^{-1} \Sigma_\varepsilon (I - B^T)^{-1})$$

Often $\Sigma_\varepsilon = \sigma^2 I$ then

$$[Z] \sim N(X^T \beta, \sigma^2 (I - B)^{-1} (I - B^T)^{-1})$$

Let $B = \lambda W$ where λ is a spatial autocorrelation parameter and W is a matrix that represents spatial dependence (often symmetric), then

$$\text{var}(Z) = \sigma^2 (I - \lambda W)^{-1} (I - \lambda W^T)^{-1}$$

Note Z is Y from the previous slide

Modeling areal data (CAR)

Conditional autoregressive model (CAR): Relies on the conditional distribution of the spatial error term.

$$e_i | e_{j-i} \sim N\left(\frac{\sum_{j-i} c_{ij} e_j}{\sum_{j-i} c_{ij}}, \frac{\sigma_{e_i}^2}{\sum_{j-i} c_{ij}}\right)$$

$$Z(A_i) | Z(A_j) \sim N\left(\mu_i + \sum_{j=1}^n c_{ij} (Z(A_j) - \mu_j), \tau_i^2\right)$$

$$Z \sim N(\vec{\mu}, (\vec{I}_n - \vec{C})^{-1} \vec{T})$$

where c_{ij} are dependence parameters similar to b_{ij}

$$\vec{C} = (c_{ij}), \text{ and } \vec{T} = \text{diag}(\tau_1^2, \tau_2^2, \dots, \tau_n^2)$$

$$\sigma_j^2 c_{ij} = \sigma_i^2 c_{ji}$$

The structure of \mathbf{B} and \mathbf{C} are specified by the shape of lattice. One common way is with a singular parameter that scales neighborhood matrix \mathbf{W} that indicates whether the regions are neighbor or not.

$$w_{ij} = \begin{cases} 1 & \text{if region } A_i \text{ shares the same border or vertex with region } A_j \\ 0 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$