# Models for Spatially Dependent Areal Data

ESS 575 Models for Ecological Data

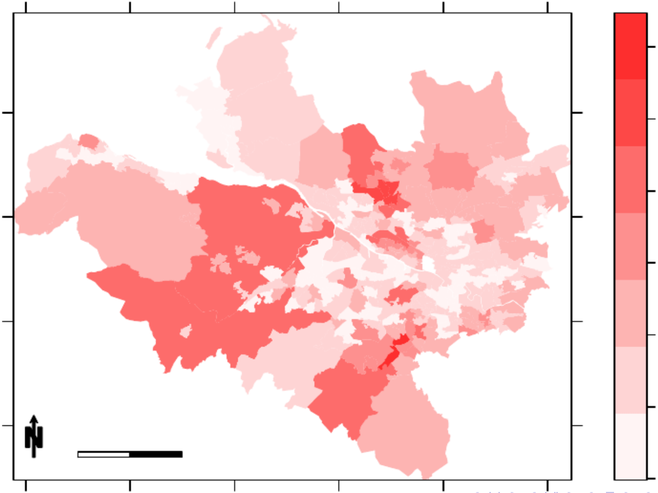
N. Thompson Hobbs

May 7, 2019

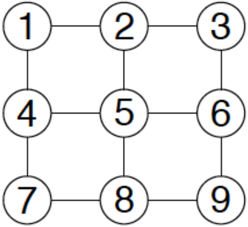
Graduate Degree Program in Ecology
Department of Ecosystem Science and Sustainability
Colorado State University

# Most ecological data are spatial

## Areal spatial processes



# Areal data and proximity




$$\mathbf{W} = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$

Possibilities include, but are not limited to:

- $w_{ij} = 1$ if $i, j$ share a common boundary (possibly a common vertex)
- $w_{ij} = 1$ for $m$ nearest neighbors.

# Measures of regularity, clustering

Moran's I: similar to covariogram ( see `spdep` package ).

$$\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta} \tag{1}$$

$$\mathbf{u} = \rho \mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon},\ \boldsymbol{\varepsilon} \text{ multivariate normal}(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}) \tag{2}$$

Moran's I: (3)

$$I = \frac{\hat{u}'}{||\hat{\mathbf{u}}||^2} \tag{4}$$

- $E(I) = -\frac{1}{n-1}$
- $I > E(I)$ implies clustering.
- $I < E(I)$ implies regularity.

# Modeling areal data

Two general types of *spatial autoregressive models:*

- Simultaneous autoregressive models (SAR): less common in Bayesian analysis. Some references at end of lecture.
- Conditional autoregressive models (CAR): The probability of values estimated at any given location are conditional on neighboring values.

# Conditional autoregressive model

$$\begin{aligned}
\mathbf{y} &= g(\boldsymbol{\theta}, \mathbf{X}) + \boldsymbol{\eta} + \boldsymbol{\varepsilon} \\
\boldsymbol{\eta} &\sim \text{multivariate normal}(\mathbf{0}, \boldsymbol{\Sigma}) \\
\boldsymbol{\Sigma} &= \overbrace{\underbrace{\sigma^2 (\mathbf{I} - \rho \mathbf{W})}_{\text{precision matrix}}{}^{-1} \mathbf{I}}^{\text{covariance matrix}} \\
\boldsymbol{\varepsilon} &\sim \text{multivariate normal}(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I})
\end{aligned}$$

1. $\rho$ is an autocorrelation parameter.
2. Proximity matrix $\mathbf{W}$ must be symmetric.

# Alternative notation

$$
\begin{aligned}
\mathbf{y} &\sim \text{multivariate normal}(g(\boldsymbol{\theta}, \mathbf{X}) + \boldsymbol{\eta}, \sigma_\varepsilon^2 \mathbf{I}) \\
\boldsymbol{\eta} &\sim \text{multivariate normal}(0, \boldsymbol{\Sigma})
\end{aligned}
$$

# Conditional autoregressive model with row standardization

$$
\begin{aligned}
\mathbf{y} &\sim \text{multivariate normal}(g(\boldsymbol{\theta},\mathbf{X})+\boldsymbol{\eta},\sigma_\varepsilon^2\mathbf{I}) \\
\boldsymbol{\eta} &\sim \text{multivariate normal}(0,\boldsymbol{\Sigma}) \\
\boldsymbol{\Sigma} &= (\text{diag}(\mathbf{W1})-\rho\mathbf{W})^{-1}
\end{aligned}
$$

- Row standardization assures that $|\rho|<1$
- $\mathbf{1}$ is a column vector of $1's$.
- $\mathbf{W1}$ is the sums of the rows
- $\text{diag}(\mathbf{W1})$ is a matrix with the sums of the rows on the diagonal and zeros elsewhere.
- Equivalent to dividing each element of $\mathbf{W}$ by the sum of the rows to obtain $\mathbf{W}_+$ and using $\boldsymbol{\Sigma}=\boldsymbol{\sigma}^{\mathbf{2}}(\mathbf{I}-\rho\mathbf{W}_+)^{-\mathbf{1}}$
- $\rho\sim\text{Beta}(18,2)$ to favor values close to 1
- $\sigma^2\sim\text{IG}(r,q)$.

# CAR for non-negative observations

Let $\boldsymbol{\Sigma} = \sigma^2 \text{diag}(\mathbf{W}\mathbf{1} - \rho \mathbf{W})^{-1}$

$$\begin{aligned} g(\boldsymbol{\beta}, \mathbf{X}) &= \exp(\mathbf{X}\boldsymbol{\beta}) && (5) \\ \log(y) &\sim \text{multivariate normal}(\log(g(\boldsymbol{\beta}, \mathbf{X})) + \boldsymbol{\eta}, \boldsymbol{\sigma}_{\varepsilon}^2 \mathbf{I}) && (6) \end{aligned}$$

# CAR for counts

Let $\boldsymbol{\Sigma} = \sigma^2 \text{diag}(\mathbf{W}\mathbf{1} - \rho\mathbf{W})^{-1}$

$$
\begin{aligned}
g(\boldsymbol{\beta}, \mathbf{X}) &= \exp(\mathbf{X}\boldsymbol{\beta}) && (7)\\
y_i &\sim \text{Poisson}(\lambda_i) && (8)\\
\log(\boldsymbol{\lambda}) &\sim \text{multivariate normal}(\log(g(\boldsymbol{\beta}, \mathbf{X})) + \boldsymbol{\eta}, \boldsymbol{\sigma}_\varepsilon^2 \mathbf{I}) && (9)
\end{aligned}
$$

## CAR for binary observations

Let $\Sigma = \sigma^2 \text{diag}(\mathbf{W1} - \rho \mathbf{W})^{-1}$

$$
\begin{aligned}
g(\boldsymbol{\beta}, \mathbf{X}) &= \text{logit}^{-1}(\mathbf{X}\boldsymbol{\beta}) && (10) \\
y_i &\sim \text{Bernoulli}(p_i) && (11) \\
\text{logit}(\mathbf{p}) &\sim \text{multivariate normal}(\text{logit}(g(\boldsymbol{\beta}, \mathbf{X})) + \boldsymbol{\eta}, \sigma_\varepsilon^2 \mathbf{I}) && (12)
\end{aligned}
$$

# Take home

- Data taken over time or space are likely to be structured by physical and biological processes.
- Our deterministic model may account for this structure. However, if the *residuals* show correlation over time and/ or space, then we are obliged to model their covariance to assure that iid assumptions are met.
- Doing so requires estimating only a few more parameters, in most cases one or two, relative to the aspatial model.
- Fitting spatial models is computationally challenging.
- Deciding whether to use a spatial or aspatial model should probably be treated as a problem in model selection.

# Further study


- Ver Hoef, J. M., E. E. Peterson, M. B. Hooten, E. M. Hanks, and M. J. Fortin. 2018. Spatial autoregressive models for statistical inference from ecological data. Ecological Monographs 88:36-59.
- Peterson, E.E., E.M. Hanks, M.B. Hooten, J.M. Ver Hoef, and M.-J. Fortin. (2019). Spatially structured statistical network models for landscape genetics. Ecological Monographs, 89: e01355.
- Finley, A. O., S. Banerjee, and B. P. Carlin. 2007. spBayes: An R package for univariate and multivariate hierarchical point-referenced spatial models. Journal of Statistical Software 19.
- Finley, A. O., S. Bsnerjee, and A. E. Gelfand. 2015. spBayes for Large Univariate and Multivariate Point-Referenced Spatio-Temporal Data Models. Journal of Statistical Software 63:1-28.
- Lee, D. 2013. CARBayes: An R Package for Bayesian Spatial Modeling with Conditional Autoregressive Priors. Journal of Statistical Software 55:1-24.
- Cressie, N., and C. K. Wikle. 2011. Statistics for spatio-temporal data. Wiley.