Hierarchical Models for Spatial and Temproal Data

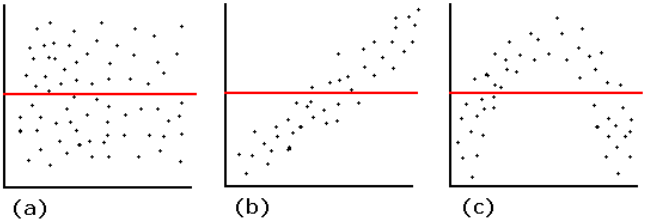
ESS 575 Models for Ecological Data

N. Thompson Hobbs and Mevin Hooten

April 25, 2017

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

$$
\begin{aligned}
\varepsilon_i &= y_i - g(\boldsymbol{\theta}, \mathbf{x_i}) \\
&\varepsilon_i \text{ are iid}
\end{aligned}
$$



What goes wrong if we fail to account for autocorrelation?

The global issue is model checking. More specific issues include:

- Inference is excessively optimistic .
- Model selection favors over-parameterized models.
- Prediction errors increase.
- Your paper will not be published if it goes to a savvy reviewer.

Roadmap: Modeling structure in data

- Temporal processes (briefly)
  - Detecting temporal dependence
  - Modeling temporal dependence
- Continuous spatial processes
  - Detecting spatial dependence
  - Distance matrices
  - Semi-variograms
  - Modeling spatial dependence
- Areal spatial processes (briefly)
  - Detecting spatial dependence
  - Modeling spatial dependence

The problem:

Assume for simplicity that the state is observed perfectly. The simplest model of the change in state with time is

$$y_t = \alpha y_{t-1} + \varepsilon_t \tag{1}$$

where $\mathsf{E}(y_t) = 0$ and $\varepsilon_t \sim \text{normal}(0, \sigma^2)$. We might introduce effects of predictor variables using

$$y_t = g(\boldsymbol{\theta}, \mathbf{x}_t) + \alpha y_{t-1} + \varepsilon_t. \tag{2}$$

What if ε_t depends on previous errors, that is, $e_t = h(e_{t-1})$? In this case, there is structural variation in the data, also called temporal dependence. The assumptions of independent errors does not hold. We have two choices:

1. Improve $g(\boldsymbol{\theta}, \mathbf{x}_t)$ so that the deterministic model accounts for the temporal dependence via the covariates.
2. Model the temporal dependence in the errors directly.

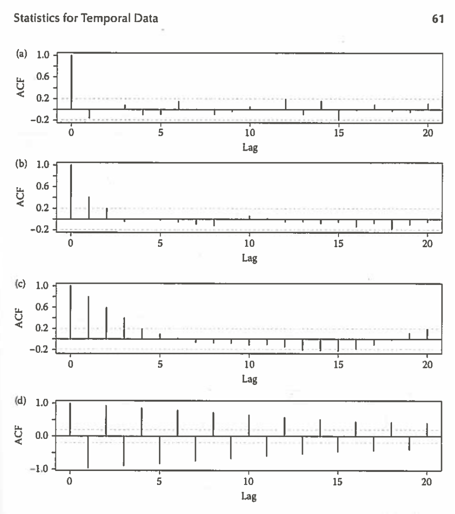
Detecting temporal dependence

The empirical autocorrelation function (ACF):

$$\rho_g = \frac{\sum_{i=1}^{n-g}(\varepsilon_i - \bar{\varepsilon})(\varepsilon_{i+g} - \bar{\varepsilon})}{\sum_{i=1}^{N}(\varepsilon_i - \bar{\varepsilon})^2}$$

where n is the number of steps in the time series and g is the "lag," the number of steps examined for temporal dependence,
$-1 \leq \rho_g \leq 1$
The notation ACF(g) means the correlation between points separated by g time periods.

ACF plots


Statistics for Temporal Data
61
(a)
(b)
(c)
(d)
ACF
Lag
1.0
0.6
0.2
−0.2
0.0
−1.0
0
5
10
15
20

ACF in MCMC

$\boldsymbol{\mu}_t = g(\boldsymbol{\theta}, z_{t-1}, \mathbf{x}_{t-1})$

1. Compute residuals at each MCMC iteration, $e_t^{(k)} = y_t - \boldsymbol{\mu}_t^{(k)}$
2. Compute $\boldsymbol{\rho}_g^{(k)}$ at each MCMC iteration and plot posterior means of $\boldsymbol{\rho}_g^{(k)}$ as a function of g.
3. Or, better and easier, sample from MCMC output for $e_t^{(k)}$, use `acf()` function in R to find posterior distributions of $\boldsymbol{\rho}_g$. Make statements like "Mean autocorrelation was .21 (BCI = .23,.18) at lag 3, revealing minimal temporal dependence in the residuals."

Modeling temporal dependence

Let $\eta_t \sim \text{normal}(\alpha\eta_{t-1}, \sigma^2)$. The quantity η_t represents time dependent, structured variation such that

$$y_t = g(\boldsymbol{\theta}, \mathbf{x}_t) + \eta_t. \tag{3}$$

We would also like to include variation that does not depend on time, the unstructured variation $\varepsilon_t \sim \text{normal}(0, \sigma^2)$. Substituting $\alpha\eta_{t-1} + \varepsilon_t$ for η_t in 3:

$$y_t = g(\boldsymbol{\theta}, \mathbf{x}_t) + \alpha\eta_{t-1} + \varepsilon_t. \tag{4}$$

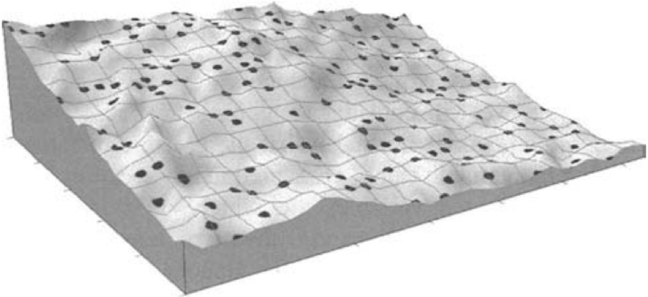
Setting time to $t-1$, solving 3 for η_{t-1} and substituting for η_{t-1} in 4:

$$\begin{aligned} y_t &= g(\boldsymbol{\theta}, \mathbf{x}_t) + \alpha\big(y_{t-1} - g(\boldsymbol{\theta}, \mathbf{x}_{t-1})\big) + \varepsilon_t && (5) \\ &= g(\boldsymbol{\theta}, \mathbf{x}_t) - \alpha g(\boldsymbol{\theta}, \mathbf{x}_{t-1}) + \alpha y_{t-1} + \varepsilon_t && (6) \end{aligned}$$

Equation 6 demonstrates the role of temporal dependence. When autocorrelation is strong $|\alpha| > 0$, inference shifts away from the direct effect of $\mathbf{x}_t$ on the response and shifts toward the effect of a *change* in covariates over time.

Most ecological data are spatial

Continuous spatial processes



Data for continuous spatial processes

All data points include a spatial reference.

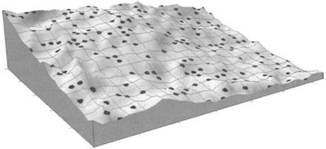
$$\text{aspatial data point} \quad : \quad y_i \tag{7}$$

$$\text{spatially referenced data point} \quad : \quad y(\mathbf{s}_i) \tag{8}$$

Where $\mathbf{s}_i$ is a vector of spatial coordinates of length 1, 2, or 3. The data are said to be continuous because they can occur at any point $(\mathbf{s}_i)$ in one, two, or three dimensional space. This does not mean that the value at that point $(y(s_i))$ can not be discrete.

Distance matrices

$n \times n$ matrix, i indexes rows, j indexes columns



$$\begin{pmatrix} 0 & d_{1,2} & d_{1,3} & . & . & d_{1,n} \\ d_{2,1} & 0 & d_{2,3} & . & . & d_{2,n} \\ d_{3,1} & d_{3,2} & 0 & . & . & d_{3,n} \\ . & . & . & 0 & . & . \\ . & . & . & . & 0 & . \\ d_{n,1} & d_{n,2} & . & . & d_{n,n-1} & 0 \end{pmatrix}$$

Assessing spatial correlation

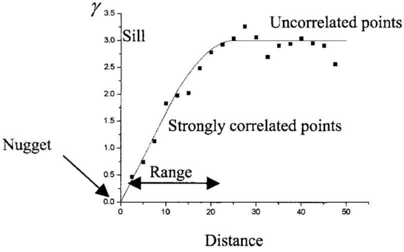
Let $\boldsymbol{\mu}_i = g(\boldsymbol{\theta}, \mathbf{x}_i)$

1. Assume $\mathbf{y}$ is measured at n spatial locations.
2. Compute the residuals: $\mathbf{e} = \mathbf{y} - \boldsymbol{\mu}$.
3. Examine the residuals $\mathbf{e}$ for spatial correlation (i.e., autocorrelation).

Assessing spatial correlation

Empirical semi-variogram

$$\hat{\gamma}(d) = \frac{\sum_{i,j \in N(d)} (e_i - e_j)^2}{2N(d)}$$



The y axis is the average squared difference between pairs of residuals at a given distance d divided by two. The x axis is the distance between pairs. Distances can be binned into categories.

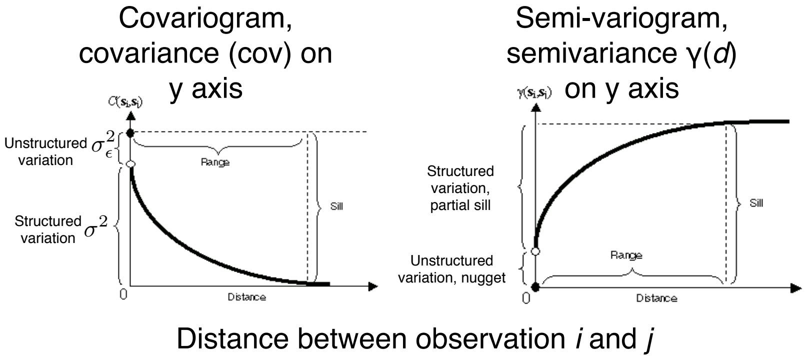
In MCMC, compute residuals at each iteration, compute $\gamma(d)^{(k)}$ and plot variogram using posterior mean of $\gamma(d)$.

Or, better, sample MCMC output for residuals in R, use R functions (gstat?) to find variogram with credible intervals.

Modeling spatial structure with two sources of variation

1. **Correlated error:** The structured, process component. Varies with distance between points. Process variance here.
2. **Uncorrelated error:** The unstructured, site specific component. It includes effects of fine scale heterogeneity and measurement error.

Modeling spatial structure with two sources of variation



$$\text{cov}(d) = \text{cov}(0) - \gamma(d)$$

Figures modified from ESRI ArcGIS Desktop online manual

Remember the covariance matrix $\boldsymbol{\Sigma}$

Imagine a vector of 3 random variables, $(z_i, z_2, z_3)'$ The covariance between any two of these random variables is simply an unstandardized version of the correlation between them– it is correlation measured in the units of the random variables. The covariance matrix (aka variance covariance matrix) of the random variable is:

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \mathrm{Cov}_{1,2} & \mathrm{Cov}_{1,3} \\ \mathrm{Cov}_{2,1} & \sigma_2^2 & \mathrm{Cov}_{2,3} \\ \mathrm{Cov}_{3,1} & \mathrm{Cov}_{3,2} & \sigma_3^2 \end{pmatrix} \tag{9}$$

Generalizing, a $m \times m$ covariance matrix has the variances of the random variable on the diagonal and the covariance on the off diagonal. The covariance between random variable i and j is $\mathrm{Cov}_{ij} = \rho \sigma_i \sigma_j$ where ρ is the correlation coefficient, which takes on values between -1 and 1. Covariance can take on values between $-\infty$ and $+\infty$.

Remember the identity matrix $\mathbf{I}$

Using a 3 x 3 matrix to illustrate:

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{10}$$

$$\sigma_\varepsilon^2 \mathbf{I} = \begin{pmatrix} \sigma_\varepsilon^2 & 0 & 0 \\ 0 & \sigma_\varepsilon^2 & 0 \\ 0 & 0 & \sigma_\varepsilon^2 \end{pmatrix} \tag{11}$$

Modeling spatial structure with two sources of error

$\mu_i = g(\boldsymbol{\theta}, x_i)$, a model of an ecological process that can take on real values (for now).
$\boldsymbol{\mu} = g(\boldsymbol{\theta}, \mathbf{X})$, note that $\boldsymbol{\mu}$ is a vector with length $=$ number of observations (n) and $\mathbf{X}$ is a data matrix with number of rows $= n$ and number of columns $=$ number of predictor variables.

$$\mathbf{y} \sim \text{multivariate normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma} + \boldsymbol{\sigma}_{\varepsilon}^2 \mathbf{I}) \tag{12}$$

$\boldsymbol{\Sigma}$ is an $n \times n$ matrix with structured variance $(\boldsymbol{\sigma}^2)$ at distance 0 on the diagonal and the covariance between observation i and observation j on the off diagonals $(i \neq j)$. $\mathbf{I}$ an $n \times n$ matrix with ones on the diagonal and zeros elsewhere. $\boldsymbol{\sigma}_{\varepsilon}^2$ is unstructured (uncorrelated) variance.

Alternative notation: random effects approach

$$\mathbf{y} = g(\boldsymbol{\theta}, X) + \boldsymbol{\eta} + \boldsymbol{\varepsilon}$$

1. Correlated Error: $\boldsymbol{\eta} \sim$ multivariate normal$(\mathbf{0}, \boldsymbol{\Sigma})$
2. Uncorrelated Error: $\boldsymbol{\varepsilon} \sim$ multivariate normal$(\mathbf{0}, \sigma_{\varepsilon}^2 \mathbf{I})$

Alternative notation: hierarchical approach

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

1. Correlated Error: $\boldsymbol{\eta}$
2. Uncorrelated Error: σ_e^2

Modeling spatial structure with two sources of error

These both imply:

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

Modeling spatial structure with two sources of error

Do we really need to predict $\frac{1}{2}(n^2 - n)$ covariances? No. Instead, we model[1] them as a function of distance using parametric covariance functions. [2]:

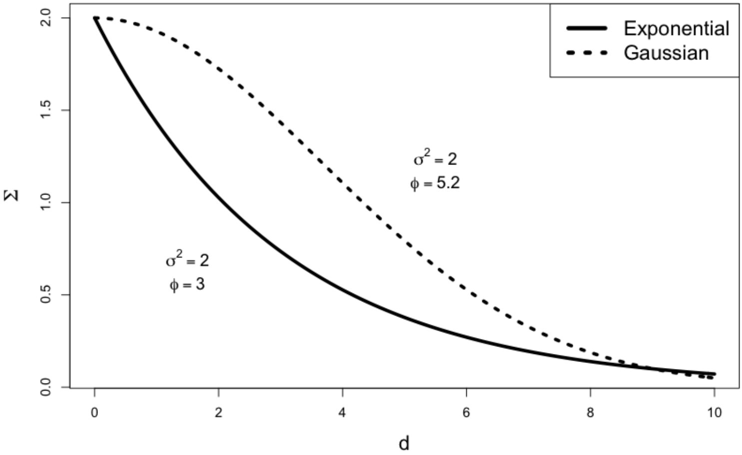
- Exponential: $\Sigma_{i,j} = \boldsymbol{\sigma}^2 \exp\left(-\frac{d_{i,j}}{\phi}\right)$
- Gaussian: $\Sigma_{i,j} = \boldsymbol{\sigma}^2 \exp\left(-\frac{d_{i,j}^2}{\phi^2}\right)$

where $d_{i,j} =$ distance between locations i and j. Note that an aspatial model would require approximating the posterior distribution of a single variance parameter $\boldsymbol{\sigma}^2$. The spatial equivalent requires three: $\boldsymbol{\sigma}^2, \phi$ and $\boldsymbol{\sigma}_\varepsilon^2$. Also note that when $i = j$ such that we are "at" a location, $d_{i,j} = 0$and $\Sigma_{i,j} = \boldsymbol{\sigma}^2$.

[1] This is a great illustration of the main purpose of science: dimension reduction.

[2] There are many others, but these are used most frequently.

Modeling spatial structure with two sources of error



Important assumptions

- **Stationarity**: spatial structure does not vary with location, which means that the spatial correlation does not change within the area being analyzed.
- **Isotropy**: spatial structure does not vary with direction, which means the spatial correlation does not change with direction.

Toy illustration for 3 data points and simple linear regression

$$
\begin{aligned}
\mathbf{y} &= (y(\mathbf{s}_1), y(\mathbf{s}_2), y(\mathbf{s}_3)) \\
\boldsymbol{\beta} &= (\beta_0, \beta_1) \\
\mathbf{X} &= \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \end{pmatrix} \\
\mathbf{D} &= \begin{pmatrix} d_{1,1} & d_{1,2} & d_{1,3} \\ d_{2,1} & d_{2,2} & d_{2,3} \\ d_{3,1} & d_{3,2} & d_{3,3} \end{pmatrix}
\end{aligned}
$$

$$
\begin{aligned}
\boldsymbol{\Sigma} &= \begin{pmatrix} \sigma^2 & \sigma^2 e^{-\frac{d_{1,2}}{\phi}} & \sigma^2 e^{-\frac{d_{1,3}}{\phi}} \\ \sigma^2 e^{-\frac{d_{2,1}}{\phi}} & \sigma^2 & \sigma^2 e^{-\frac{d_{2,3}}{\phi}} \\ \sigma^2 e^{-\frac{d_{3,1}}{\phi}} & \sigma^2 e^{-\frac{d_{3,,2}}{\phi}} & \sigma^2 \end{pmatrix} \\
\sigma_\varepsilon^2 \mathbf{I} &= \begin{pmatrix} \sigma_\varepsilon^2 & 0 & 0 \\ 0 & \sigma_\varepsilon^2 & 0 \\ 0 & 0 & \sigma_\varepsilon^2 \end{pmatrix} \\
\boldsymbol{\Sigma} + \sigma_\varepsilon^2 \mathbf{I} &= \begin{pmatrix} \sigma^2 + \sigma_\varepsilon^2 & \sigma^2 e^{-\frac{d_{1,2}}{\phi}} & \sigma^2 e^{-\frac{d_{1,3}}{\phi}} \\ \sigma^2 e^{-\frac{d_{2,1}}{\phi}} & \sigma^2 + \sigma_\varepsilon^2 & \sigma^2 e^{-\frac{d_{2,3}}{\phi}} \\ \sigma^2 e^{-\frac{d_{3,1}}{\phi}} & \sigma^2 e^{-\frac{d_{3,,2}}{\phi}} & \sigma^2 + \sigma_\varepsilon^2 \end{pmatrix}
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{y} &\sim \text{multivariate normal}(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma} + \sigma_\varepsilon^2 \mathbf{I}) \\
[\boldsymbol{\beta}, \sigma^2, \sigma_\varepsilon^2, \phi \mid \mathbf{y}] &\propto \propto [\mathbf{y} \mid \mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma} + \sigma_\varepsilon^2 \mathbf{I}] \\
&\times [\boldsymbol{\beta}][\sigma^2][\sigma_\varepsilon^2][\phi]
\end{aligned}
$$

Priors on ϕ

Choices for range parameter ϕ:

- $\phi \sim \text{gamma}(\gamma_1, \gamma_2)$
- $\log(\phi) \sim \text{normal}(\mu_\phi, \sigma_\phi^2)$
- $\phi \sim \text{Half-Cauchy}(\gamma)$

Bayesian kriging: predictions at unobserved locations

- We seek to predict $y(\mathbf{s}_u)$ at unobserved location $\mathbf{s}_u$, given the model and the data $y(\mathbf{s}_i)$ for $i = 1, \ldots, n$.

$$y(\mathbf{s}_i) = g(\boldsymbol{\theta}, \mathbf{x}(\mathbf{s}_i)) + \eta(\mathbf{s}_i) + \varepsilon_i$$

where $\eta(\mathbf{s}_i)$ is a random variable representing structured variation ε_i represents unstructured variation.
- We need the posterior predictive distribution:

$$[y_u|\mathbf{y}] = \int\int\int\int [y_u|\mathbf{y}, \boldsymbol{\theta}, \sigma^2, \sigma_\varepsilon^2, \phi][\boldsymbol{\theta}, \sigma^2, \sigma_\varepsilon^2, \phi|\mathbf{y}]\, d\boldsymbol{\beta}\, d\sigma^2\, d\sigma_\varepsilon^2\, d\phi$$

- Approximation:
  - Compose $\mathbf{D_u}$ for distances between observed $(\mathbf{y}_o)$ and unobserved $(\mathbf{y}_u)$.
  - At each MCMC iteration k,
    - Use $\mathbf{D_u}$ with values $\sigma^{(k)}, \sigma_\varepsilon^{(k)}$ and $\phi^{(k)}$ to compute covariance matrix and $\Sigma_u^{(k)} + \sigma_\varepsilon^{2(k)}\mathbf{I}$
    - Compute value at new location using

$$\mathbf{y}_u^k = \mathbf{X}_u \boldsymbol{\beta}^{(k)} + \frac{(\Sigma_u^{(k)} + \sigma_\varepsilon^{2(k)}\mathbf{I})}{(\Sigma^{(k)} + \sigma_\varepsilon^{2(k)}\mathbf{I})}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^{(k)})$$

- This allows you to put credible intervals on values at locations.

General spatial models

- Real valued, non-negative

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

- Counts

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

- Binary

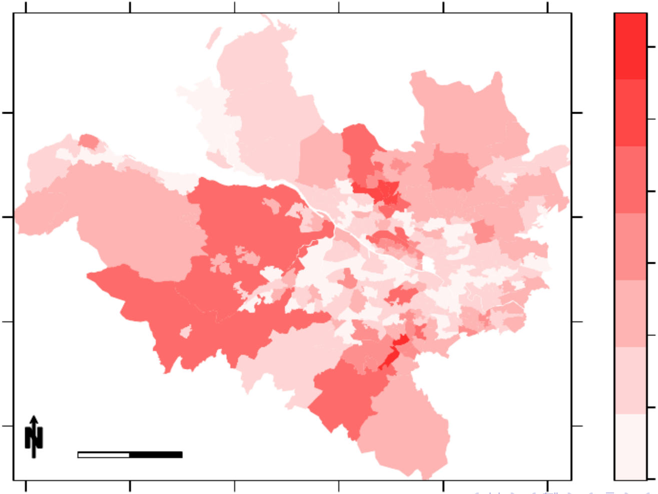
$$\begin{aligned} g(\boldsymbol{\beta},\mathbf{X}) &= \text{logit}^{-1}(\mathbf{X}\boldsymbol{\beta}) \\ \text{logit}(\mathbf{p}) &\sim \text{multivariate normal}(\text{logit}(g(\boldsymbol{\beta},\mathbf{X}),\boldsymbol{\Sigma}+\sigma_\varepsilon^2\mathbf{I})) \\ y_i &\sim \text{Bernoulli}(p_i) \end{aligned}$$

Simulating data for a continuous spatial process

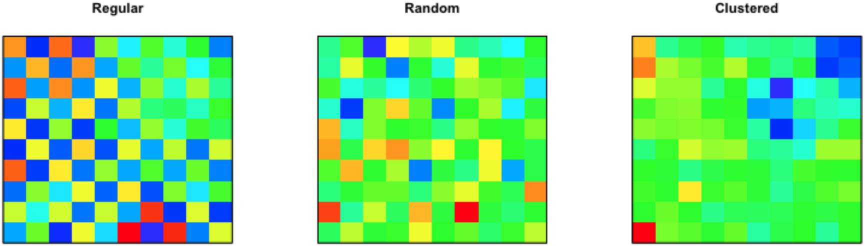
1. Choose locations $\mathbf{s}_i$ for $i = 1, \ldots, n$.
2. Choose the mean $\boldsymbol{\mu}$. This could be a scalar or it could vary spatially. It could be the output of a model with parameter values that you choose and $\mathbf{x}$ data.
3. Choose the unstructured variance σ_ε^2.
4. Choose range parameter ϕ and variance component σ^2.
5. Compute distance matrix $\mathbf{D}$ between all n locations of interest.
6. Calculate covariance matrix $\boldsymbol{\Sigma} = \sigma^2 \exp\left(-\frac{\mathbf{D}}{\phi}\right)$.
7. Sample the n-dimensional vector $\mathbf{y} \sim$ multivariate normal$(\boldsymbol{\mu}, \boldsymbol{\Sigma} + \sigma_\varepsilon^2 \mathbf{I})$.

Most ecological data are spatial

Areal spatial processes



Regular, random, or clustered?



Measures of regularity, clustering

- Moran's I: similar to covariogram.

$$I = \frac{n}{(n-1)s^2 w_{\cdot\cdot}} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j}(y(a_i) - \bar{y})(y(a_j) - \bar{y})$$

- Geary's C: similar to variogram (or Durbin-Watson statistic in time series).

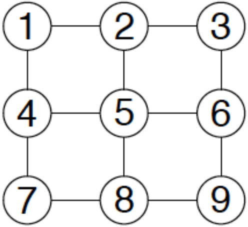
$$C = \frac{1}{2s^2 w_{\cdot\cdot}} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j}(y(a_i) - y(a_j))^2$$

s^2 =sample variance

Testing for regularity, clustering

- Moran's I:
  - $E(I) = -\frac{1}{n-1}$
  - $I > E(I)$ implies clustering.
  - $I < E(I)$ implies regularity.
- Geary's C:
  - $E(C) = 1$
  - $C > 1$ implies negative autocorrelation (regularity).
  - $C < 1$ implies positive autocorrelation (clustering).
  - $0 < C < 2$.

Areal data and proximity



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

Choices for elements of W, the "weights"

Possibilities include, but are not limited to:

- $w_{ij} = 1$ if *i*,j share a common boundary (possibly a common vertex)
- w_{ij} is an inverse distance between units
- $w_{ij} = 1$ if distance between units is $\leq K$
- $w_{ij} = 1$ for m nearest neighbors.

$$y(a_i) = \mathbf{x}(a_i)'\boldsymbol{\beta} + \boldsymbol{\varepsilon}(a_i)$$

- Similar to that for continuous spatial modeling, except:
  - Covariance is not parameterized in terms of distance.
  - Usually not stationary (variance changes spatially).
- We need a modeling framework that accounts for these issues.

Modeling areal data

Two general types of *spatial autoregressive models:*

- Simultaneous autoregressive models (SAR): not commonly used in Bayesian analysis. Some notes at end.
- Conditional autoregressive models (CAR): The probability of values estimated at any given location are conditional on neighboring values.

Conditional autoregressive model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

- where, $\boldsymbol{\varepsilon} \sim \text{normal}(\mathbf{0}, \sigma^2(\mathbf{I} - \rho\mathbf{W})^{-1})$.
- ρ is an autocorrelation parameter.
- Note: Proximity matrix $\mathbf{W}$ must be symmetric.
- Derivation requires Hamersley-Clifford theorem, Brooks lemma

Conditional autoregressive model with row standardization

$$\mathbf{y} \sim \text{normal}(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{R})$$

- $\boldsymbol{\beta} \sim \text{normal}(\boldsymbol{\mu}_\beta, \Sigma_\beta)$.
- $\boldsymbol{\Sigma} = \sigma^2\mathbf{R}$
- $\mathbf{R} = (\text{diag}(\mathbf{W1}) - \rho\mathbf{W})^{-1}$
  - $\mathbf{1}$ 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.
  - Row standardization assures that $|\rho| < 1$
  - Equivalent to dividing each element of $\mathbf{W}$ by the sum of the rows to obtain $\mathbf{W}_+$ and using $\boldsymbol{\Sigma} = \boldsymbol{\sigma^2}(\mathbf{I} - \rho\mathbf{W}_+)^{-\mathbf{1}}$
- $\rho \sim \text{Beta}(18, 2)$ to favor values close to 1
- $\sigma^2 \sim \text{IG}(r, q)$.

Hierarchical CAR models

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

We presume we have a data model with parameters subscripted by d

Real numbers

$$
\begin{aligned}
y_i &\sim \text{normal}(\boldsymbol{\mu}, \boldsymbol{\sigma}_d^2) && (13)\\
\boldsymbol{\mu} &\sim \text{multivariate normal}(\mathbf{X}\boldsymbol{\beta}, \mathbf{\Sigma}) && (14)
\end{aligned}
$$

Non-negative, continuous

$$
\begin{aligned}
g(\boldsymbol{\beta}, \mathbf{X}) &= \exp(\mathbf{X}\boldsymbol{\beta}) && (15)\\
y_i &\sim \text{gamma}\left(\frac{\boldsymbol{\mu}^2}{\boldsymbol{\sigma}_d^2}, \frac{\boldsymbol{\mu}}{\boldsymbol{\sigma}_d^2}\right) && (16)\\
\log(\boldsymbol{\mu}) &\sim \text{multivariate normal}(\log(g(\boldsymbol{\beta}, \mathbf{X})), \mathbf{\Sigma}) && (17)
\end{aligned}
$$

Counts

Simultaneous autoregressive model (SAR)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

- $\boldsymbol{\varepsilon} = \rho\mathbf{W}\boldsymbol{\varepsilon} + \boldsymbol{\nu}$
  - ρ is an autocorrelation parameter.
  - $E(\boldsymbol{\nu}) = \mathbf{0}$.
  - $E(\boldsymbol{\nu}\boldsymbol{\nu}') = \sigma^2\mathbf{I}$.

The $\boldsymbol{\varepsilon}$ are autoregressive because they appear on both sides of the $=$. They are spatially structured by $\rho\mathbf{W}$.

Some algebra needed for next slide

$$\begin{aligned} \boldsymbol{\varepsilon} &= \rho \mathbf{W} \varepsilon + \boldsymbol{v} && (24) \\ \boldsymbol{v} &= \boldsymbol{\varepsilon} - \rho \mathbf{W} \boldsymbol{\varepsilon} && (25) \\ \boldsymbol{v} &= \varepsilon(\mathsf{I} - \rho \mathbf{W}) && (26) \\ \varepsilon &= v(\mathsf{I} - \rho \mathbf{W})^{-1} && (27) \end{aligned}$$

Simultaneous autoregressive model with covariates

$$
\begin{aligned}
\mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\
&= \mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \rho\mathbf{W})^{-1}\boldsymbol{\nu}
\end{aligned}
$$

- Typically $\boldsymbol{\nu} \sim \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I})$.
- Note: $\mathrm{cov}(\boldsymbol{\varepsilon}) = \sigma^2((\mathbf{I} - \rho\mathbf{W})'(\mathbf{I} - \rho\mathbf{W}))^{-1}$.