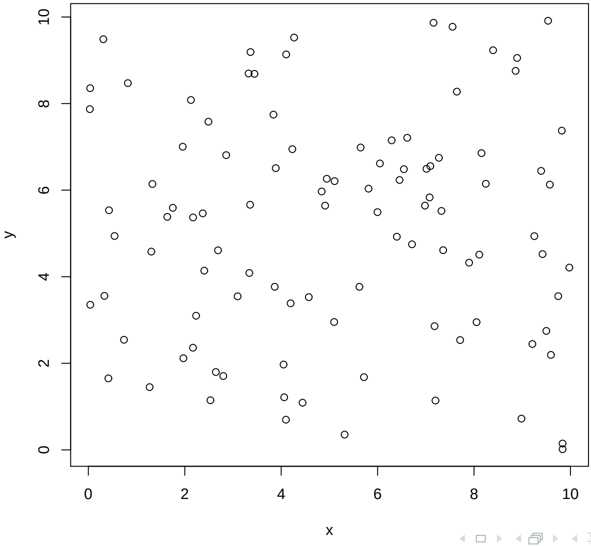
Hierarchical Modeling for Univariate Spatial Data

Andrew O. Finley

Department of Forestry & Department of Geography, Michigan State University, Lansing Michigan, U.S.A.

October 30, 2013

Spatial Domain



Algorithmic Modeling

- Spatial surface observed at finite set of locations $\mathscr{S} = \{\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_n\}$
- Tessellate the spatial domain (usually with data locations as vertices)
- Fit an interpolating polynomial:

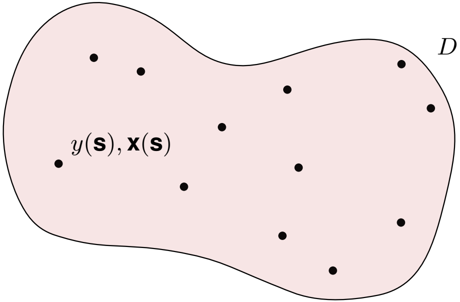
$$f(\mathbf{s}) = \sum_i w_i(\mathscr{S}; \mathbf{s}) f(\mathbf{s}_i)$$

- "Interpolate" by reading off $f(\mathbf{s}_0)$.
- Issues:
  - Sensitivity to tessellations
  - Choices of multivariate interpolators
  - Numerical error analysis

Simple linear model

$$y(\mathbf{s}) = \mu(\mathbf{s}) + \epsilon(\mathbf{s}),$$

- Response: $y(\mathbf{s})$ at location $\mathbf{s}$
- Mean: $\mu = \mathbf{x}(\mathbf{s})'\boldsymbol{\beta}$
- Error: $\epsilon(\mathbf{s}) \stackrel{iid}{\sim} N(0, \tau^2)$

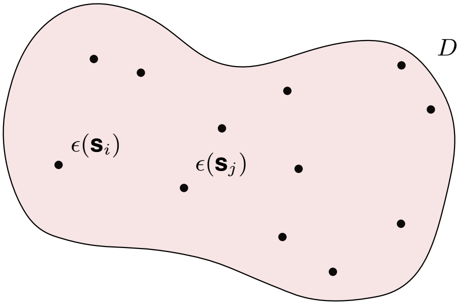


Simple linear model

$$y(\mathbf{s}) = \mu(\mathbf{s}) + \epsilon(\mathbf{s}),$$

Assumptions regarding $\epsilon(\mathbf{s})$:

- $\epsilon(\mathbf{s}) \overset{iid}{\sim} N(0, \tau^2)$

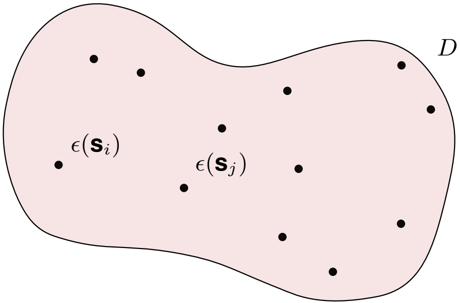


Simple linear model

$$y(\mathbf{s}) = \mu(\mathbf{s}) + \epsilon(\mathbf{s}),$$

Assumptions regarding $\epsilon(\mathbf{s})$:

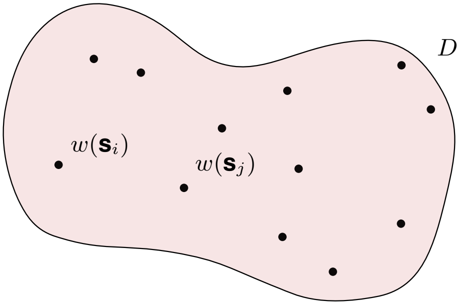
- $\epsilon(\mathbf{s}) \overset{iid}{\sim} N(0, \tau^2)$
- $\epsilon(\mathbf{s}_i)$ and $\epsilon(\mathbf{s}_j)$ are uncorrelated for all $i \neq j$



Spatial Gaussian processes (GP):

- Say $w(\mathbf{s}) \sim GP(0, \sigma^2 \rho(\cdot))$ and

$$Cov(w(\mathbf{s}_1), w(\mathbf{s}_2)) = \sigma^2 \rho\left(\phi; \|\mathbf{s}_1 - \mathbf{s}_2\|\right)$$



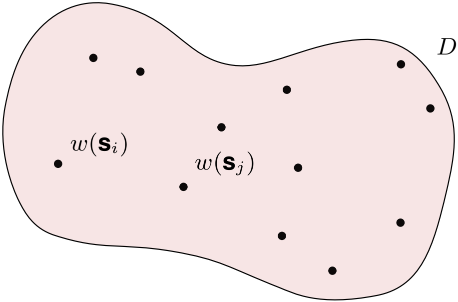
Spatial Gaussian processes (GP):

- Say $w(\mathbf{s}) \sim GP(0, \sigma^2\rho(\cdot))$ and

$$Cov(w(\mathbf{s}_1), w(\mathbf{s}_2)) = \sigma^2\rho\left(\phi; \|\mathbf{s}_1 - \mathbf{s}_2\|\right)$$

- Let $\mathbf{w} = [w(\mathbf{s}_i)]_{i=1}^n$, then

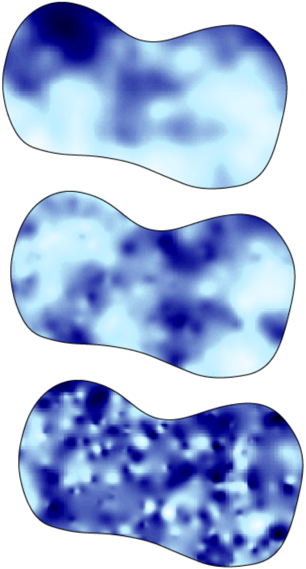
$$\mathbf{w} \sim N(\mathbf{0}, \sigma^2 R(\phi)), \text{ where } R(\phi) = [\rho(\phi; \|\mathbf{s}_i - \mathbf{s}_j\|)]_{i,j=1}^n$$



Realization of a Gaussian process:

- Changing ϕ and holding $\sigma^2 = 1$:

$$\mathbf{w} \sim N(\mathbf{0}, \sigma^2 R(\phi)), \text{ where}$$
$$R(\phi) = [\rho(\phi; \|\mathbf{s}_i - \mathbf{s}_j\|)]_{i,j=1}^{n}$$



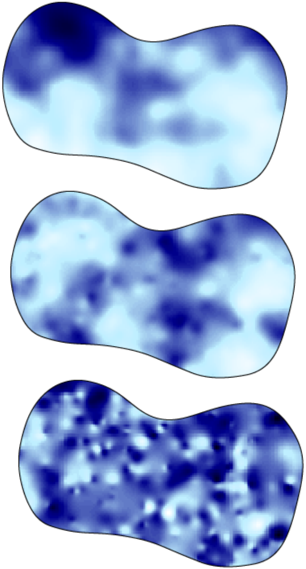
Realization of a Gaussian process:

- Changing ϕ and holding $\sigma^2 = 1$:

$$\mathbf{w} \sim N(\mathbf{0}, \sigma^2 R(\phi)), \text{ where}$$
$$R(\phi) = [\rho(\phi; \|\mathbf{s}_i - \mathbf{s}_j\|)]_{i,j=1}^{n}$$

- Correlation model for $R(\phi)$:
  e.g., exponential decay

$$\rho(\phi; t) = \exp(-\phi t) \text{ if } t > 0.$$



Realization of a Gaussian process:

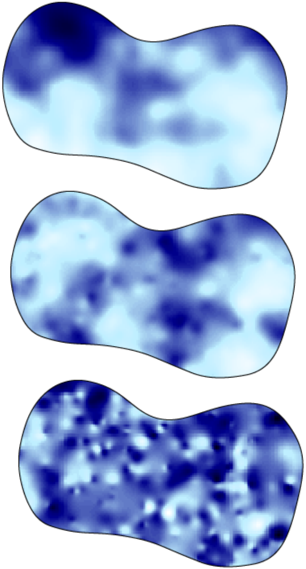
- Changing ϕ and holding $\sigma^2 = 1$:

$$\mathbf{w} \sim N(\mathbf{0}, \sigma^2 R(\phi)), \text{ where}$$
$$R(\phi) = [\rho(\phi; \|\mathbf{s}_i - \mathbf{s}_j\|)]_{i,j=1}^{n}$$

- Correlation model for $R(\phi)$: e.g., exponential decay

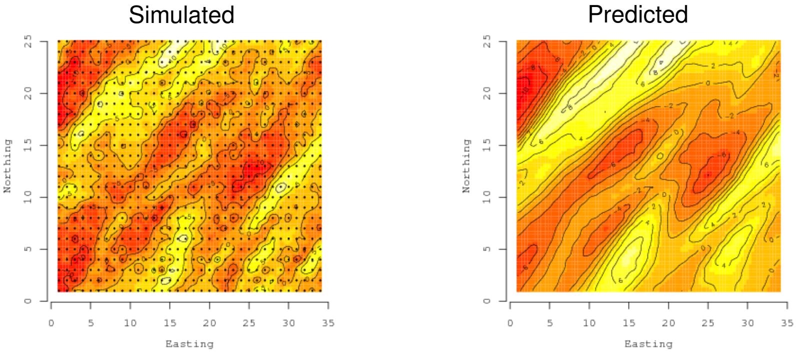
$$\rho(\phi; t) = \exp(-\phi t) \text{ if } t > 0.$$

- Other valid models e.g., Gaussian, Spherical, Matérn.
- Effective range, $t_0 = -ln(0.05)/\phi \approx 3/\phi$



$\mathbf{w} \sim N(\mathbf{0}, \sigma_w^2 R(\phi))$ defines complex spatial dependence structures.

E.g., anisotropic Matérn correlation function:
$\rho(\mathbf{s}_i, \mathbf{s}_j; \phi) = \left(1/\Gamma(\nu)2^{\nu-1}\right)\left(2\sqrt{\nu d_{ij}})^{\nu}\kappa_{\nu}(2\sqrt{\nu d_{ij}}\right)$, where
$d_{ij} = (\mathbf{s}_i - \mathbf{s}_j)'\,\Sigma^{-1}\,(\mathbf{s}_i - \mathbf{s}_j)$, $\Sigma = G(\psi)\Lambda^2 G(\psi)'$. Thus, $\phi = (\nu, \psi, \Lambda)$.



Simple linear model + random spatial effects

$$y(\mathbf{s}) = \mu(\mathbf{s}) + w(\mathbf{s}) + \epsilon(\mathbf{s}),$$

- Response: $y(\mathbf{s})$ at some site
- Mean: $\mu = \mathbf{x}(\mathbf{s})'\beta$
- Spatial random effects: $w(\mathbf{s}) \sim GP(0, \sigma^2 \rho(\phi; \|\mathbf{s}_1 - \mathbf{s}_2\|))$
- Non-spatial variance: $\epsilon(\mathbf{s}) \overset{iid}{\sim} N(0, \tau^2)$

Hierarchical modeling

- First stage:

$$\mathbf{y}|\boldsymbol{\beta}, \mathbf{w}, \tau^2 \sim \prod_{i=1}^{n} N(y(\mathbf{s}_i)\,|\,\mathbf{x}(\mathbf{s}_i)'\boldsymbol{\beta} + w(\mathbf{s}_i), \tau^2)$$

Hierarchical modeling

- First stage:

$$\mathbf{y}|\boldsymbol{\beta}, \mathbf{w}, \tau^2 \sim \prod_{i=1}^{n} N(y(\mathbf{s}_i) \,|\, \mathbf{x}(\mathbf{s}_i)'\boldsymbol{\beta} + w(\mathbf{s}_i), \tau^2)$$

- Second stage:

$$\mathbf{w}|\sigma^2, \phi \sim N(\mathbf{0}, \sigma^2 R(\phi))$$

Hierarchical modeling

- First stage:

$$\mathbf{y}|\boldsymbol{\beta},\mathbf{w},\tau^2 \sim \prod_{i=1}^{n} N(y(\mathbf{s}_i)\,|\,\mathbf{x}(\mathbf{s}_i)'\boldsymbol{\beta}+w(\mathbf{s}_i),\tau^2)$$

- Second stage:

$$\mathbf{w}|\sigma^2,\phi \sim N(\mathbf{0},\sigma^2 R(\phi))$$

- Third stage: Priors on $\Omega = (\boldsymbol{\beta},\tau^2,\sigma^2,\phi)$

Hierarchical modeling

- First stage:

$$\mathbf{y}|\boldsymbol{\beta}, \mathbf{w}, \tau^2 \sim \prod_{i=1}^{n} N(y(\mathbf{s}_i) \,|\, \mathbf{x}(\mathbf{s}_i)'\boldsymbol{\beta} + w(\mathbf{s}_i), \tau^2)$$

- Second stage:

$$\mathbf{w}|\sigma^2, \phi \sim N(\mathbf{0}, \sigma^2 R(\phi))$$

- Third stage: Priors on $\Omega = (\boldsymbol{\beta}, \tau^2, \sigma^2, \phi)$
- Marginalized likelihood:

$$\mathbf{y}|\Omega \sim N(X\boldsymbol{\beta}, \sigma^2 R(\phi) + \tau^2 I)$$

Hierarchical modeling

- First stage:

$$\mathbf{y}|\boldsymbol{\beta}, \mathbf{w}, \tau^2 \sim \prod_{i=1}^{n} N(y(\mathbf{s}_i) \,|\, \mathbf{x}(\mathbf{s}_i)'\boldsymbol{\beta} + w(\mathbf{s}_i), \tau^2)$$

- Second stage:

$$\mathbf{w}|\sigma^2, \phi \sim N(\mathbf{0}, \sigma^2 R(\phi))$$

- Third stage: Priors on $\Omega = (\boldsymbol{\beta}, \tau^2, \sigma^2, \phi)$
- Marginalized likelihood:

$$\mathbf{y}|\Omega \sim N(X\boldsymbol{\beta}, \sigma^2 R(\phi) + \tau^2 I)$$

- Note: Spatial process parametrizes Σ:
  $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon},\ \boldsymbol{\epsilon} \sim N\left(\mathbf{0}, \Sigma\right),\ \Sigma = \sigma^2 R(\phi) + \tau^2 I$

Bayesian Computations

- Choice: Fit $[\mathbf{y}|\Omega] \times [\Omega]$ or $[\mathbf{y}|\boldsymbol{\beta}, \mathbf{w}, \tau^2] \times [\mathbf{w}|\sigma^2, \phi] \times [\Omega]$.

Bayesian Computations

- Choice: Fit $[\mathbf{y}|\Omega] \times [\Omega]$ or $[\mathbf{y}|\boldsymbol{\beta}, \mathbf{w}, \tau^2] \times [\mathbf{w}|\sigma^2, \phi] \times [\Omega]$.
- Conditional model:
  - conjugate full conditionals for σ^2, τ^2 and **w** – easier to program.

Bayesian Computations

- Choice: Fit $[\mathbf{y}|\Omega] \times [\Omega]$ or $[\mathbf{y}|\boldsymbol{\beta}, \mathbf{w}, \tau^2] \times [\mathbf{w}|\sigma^2, \phi] \times [\Omega]$.
- Conditional model:
  - conjugate full conditionals for σ^2, τ^2 and **w** – easier to program.
- Marginalized model:
  - need Metropolis or Slice sampling for σ^2, τ^2 and ϕ. Harder to program.
  - But, reduced parameter space $\Rightarrow$ faster convergence
  - $\sigma^2 R(\phi) + \tau^2 I$ is more stable than $\sigma^2 R(\phi)$.

Bayesian Computations

- Choice: Fit $[\mathbf{y}|\Omega] \times [\Omega]$ or $[\mathbf{y}|\boldsymbol{\beta}, \mathbf{w}, \tau^2] \times [\mathbf{w}|\sigma^2, \phi] \times [\Omega]$.
- Conditional model:
  - conjugate full conditionals for σ^2, τ^2 and **w** – easier to program.
- Marginalized model:
  - need Metropolis or Slice sampling for σ^2, τ^2 and ϕ. Harder to program.
  - But, reduced parameter space $\Rightarrow$ faster convergence
  - $\sigma^2 R(\phi) + \tau^2 I$ is more stable than $\sigma^2 R(\phi)$.
- But what about $R^{-1}(\phi)$?? EXPENSIVE!

Where are the $\mathbf{w}$'s?

- Interest often lies in the spatial surface $\mathbf{w}|\mathbf{y}$.

Where are the **w**'s?

- Interest often lies in the spatial surface $\mathbf{w}|\mathbf{y}$.
- They are recovered from

$$[\mathbf{w}|\mathbf{y}, X] = \int [\mathbf{w}|\Omega, \mathbf{y}, X] \times [\Omega|\mathbf{y}, X] d\Omega$$

using posterior samples:

Where are the **w**'s?

- Interest often lies in the spatial surface $\mathbf{w}|\mathbf{y}$.
- They are recovered from

$$[\mathbf{w}|\mathbf{y}, X] = \int [\mathbf{w}|\Omega, \mathbf{y}, X] \times [\Omega|\mathbf{y}, X] d\Omega$$

  using posterior samples:
  - Obtain $\Omega^{(1)}, \ldots, \Omega^{(G)} \sim [\Omega|\mathbf{y}, X]$
  - For each $\Omega^{(g)}$, draw $\mathbf{w}^{(g)} \sim [\mathbf{w}|\Omega^{(g)}, \mathbf{y}, X]$

Where are the **w**'s?

- Interest often lies in the spatial surface $\mathbf{w}|\mathbf{y}$.
- They are recovered from

$$[\mathbf{w}|\mathbf{y}, X] = \int [\mathbf{w}|\Omega, \mathbf{y}, X] \times [\Omega|\mathbf{y}, X] d\Omega$$

  using posterior samples:
  - Obtain $\Omega^{(1)}, \ldots, \Omega^{(G)} \sim [\Omega|\mathbf{y}, X]$
  - For each $\Omega^{(g)}$, draw $\mathbf{w}^{(g)} \sim [\mathbf{w}|\Omega^{(g)}, \mathbf{y}, X]$
- NOTE: With Gaussian likelihoods $[\mathbf{w}|\Omega, \mathbf{y}, X]$ is also Gaussian. With other likelihoods this may not be easy and often the conditional updating scheme is preferred.

- Often we need to predict $y(\mathbf{s})$ at a *new* set of locations $\{\tilde{\mathbf{s}}_0, \ldots, \tilde{\mathbf{s}}_m\}$ with associated predictor matrix $\tilde{X}$.

- Sample from predictive distribution:

$$\begin{aligned}[\tilde{\mathbf{y}}|\mathbf{y}, X, \tilde{X}] &= \int [\tilde{\mathbf{y}}, \Omega|\mathbf{y}, X, \tilde{X}]d\Omega \\ &= \int [\tilde{\mathbf{y}}|\mathbf{y}, \Omega, X, \tilde{X}] \times [\Omega|\mathbf{y}, X]d\Omega,\end{aligned}$$

$[\tilde{\mathbf{y}}|\mathbf{y}, \Omega, X, \tilde{X}]$ is multivariate normal. Sampling scheme:

  - Obtain $\Omega^{(1)}, \ldots, \Omega^{(G)} \sim [\Omega|\mathbf{y}, X]$
  - For each $\Omega^{(g)}$, draw $\tilde{\mathbf{y}}^{(g)} \sim [\tilde{\mathbf{y}}|\mathbf{y}, \Omega^{(g)}, X, \tilde{X}]$.

Colorado data illustration

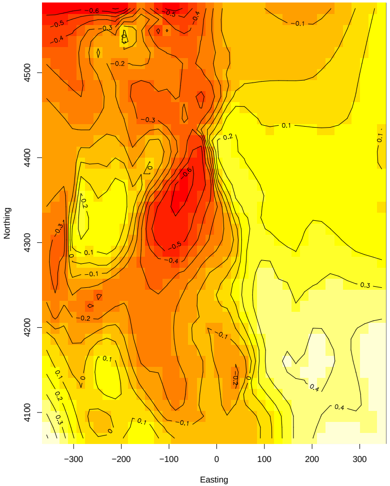
- Modeling temperature: 507 locations in Colorado.
- Simple spatial regression model:

$$y(\mathbf{s}) = \mathbf{x}(\mathbf{s})'\boldsymbol{\beta} + w(\mathbf{s}) + \epsilon(\mathbf{s})$$

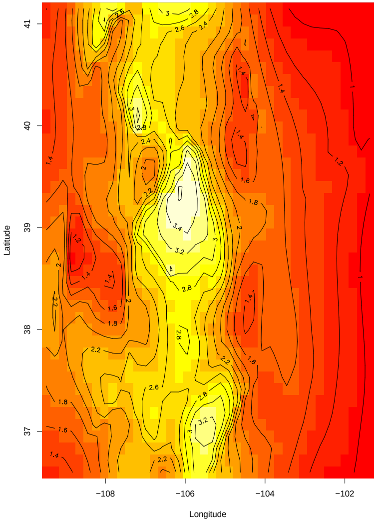
- $w(\mathbf{s}) \sim GP(0, \sigma^2\rho(\cdot;\phi,\nu))$; $\epsilon(\mathbf{s}) \stackrel{iid}{\sim} N(0, \tau^2)$

Parameters	50% (2.5%,97.5%)
Intercept	2.827 (2.131,3.866)
[Elevation]	-0.426 (-0.527,-0.333)
Precipitation	0.037 (0.002,0.072)
σ^2	0.134 (0.051, 1.245)
ϕ	7.39E-3 (4.71E-3, 51.21E-3)
Range	278.2 (38.8, 476.3)
τ^2	0.051 (0.022, 0.092)

Temperature residual map



Elevation map



Residual map with elev. as covariate