Modeling Univariate Spatial Data

September 24, 2019

- Researchers in diverse areas such as ecology, forestry, climatology, and environmental health, are increasingly faced with the task of analyzing data that are:
  - highly multivariate, with many important predictors and response variables,
  - geographically referenced, and often presented as maps, and
  - temporally correlated, as in longitudinal or other time series structures.

$\Rightarrow$ motivates hierarchical modeling and data analysis for complex spatial (and spatiotemporal) data sets.

- point-referenced data, where $y(\mathbf{s})$ is a random vector at a location $\mathbf{s} \in \Re^r$, where $\mathbf{s}$ varies continuously over D, a fixed subset of $\Re^r$ that contains an r-dimensional rectangle of positive volume;

- point-referenced data, where $y(\mathbf{s})$ is a random vector at a location $\mathbf{s} \in \Re^r$, where $\mathbf{s}$ varies continuously over D, a fixed subset of $\Re^r$ that contains an r-dimensional rectangle of positive volume;

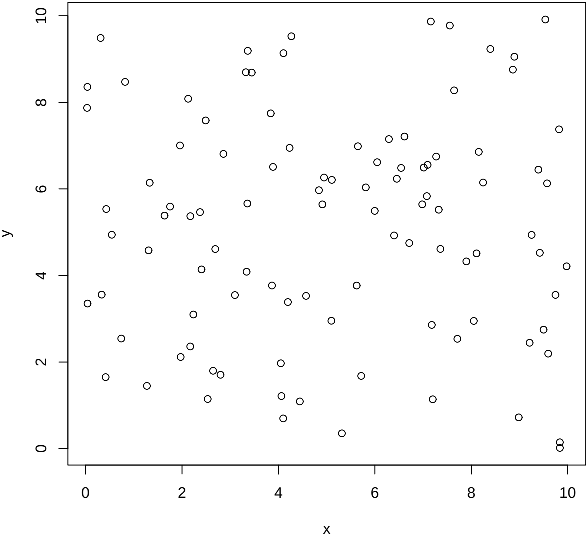
- areal data, where D is again a fixed subset (of regular or irregular shape), but now partitioned into a finite number of areal units with well-defined boundaries;

- point-referenced data, where $y(\mathbf{s})$ is a random vector at a location $\mathbf{s} \in \Re^r$, where $\mathbf{s}$ varies continuously over D, a fixed subset of $\Re^r$ that contains an r-dimensional rectangle of positive volume;

- areal data, where D is again a fixed subset (of regular or irregular shape), but now partitioned into a finite number of areal units with well-defined boundaries;

- point pattern data, where now D is itself random; its index set gives the locations of random events that are the spatial point pattern. $y(\mathbf{s})$ itself can simply equal 1 for all $\mathbf{s} \in D$ (indicating occurrence of the event), or possibly give some additional covariate information (producing a marked point pattern process).

Spatial Domain



Algorithmic Modeling

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

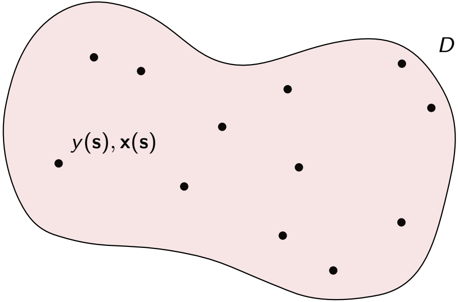
$$f(\mathbf{s}) = \sum_i w_i(\mathcal{S}; \mathbf{s}) f(\mathbf{s}_i)$$

- "Interpolate" by reading off $f(\mathbf{s}_0)$.
- Includes: triangulation, weighted averages, geographically weighted regression (GWR)
- 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: $\mu = \mathbf{x}(\mathbf{s})^\top \beta$
- Error: $\epsilon(\mathbf{s}) \overset{iid}{\sim} N(0, \tau^2)$

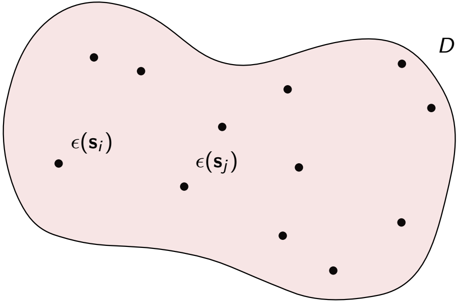


Simple linear model

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

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

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

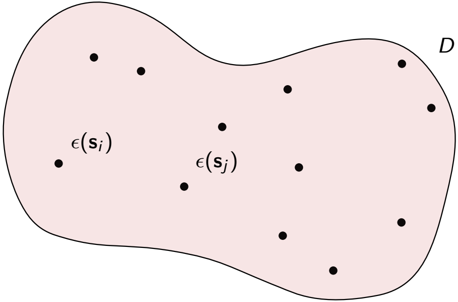


Simple linear model

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

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

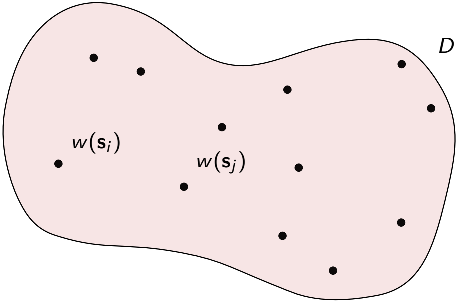
- $\boldsymbol{\epsilon}(\mathbf{s}) \overset{iid}{\sim} N(0, \tau^2)$
- $\epsilon(\mathbf{s}_i)$ and $\epsilon(\mathbf{s}_j)$ are uncorrelated for all $i \neq j$



Spatial Gaussian processes (GP):

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

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



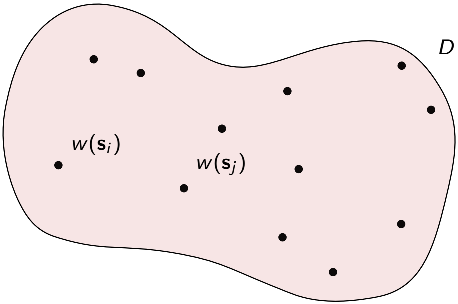
Spatial Gaussian processes (GP):

- 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

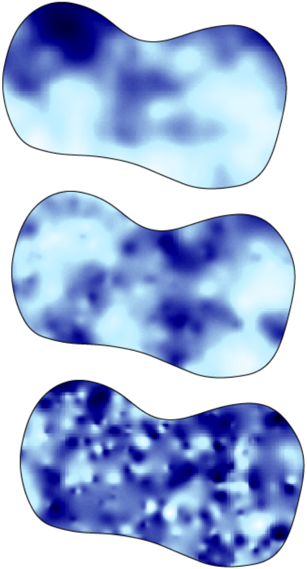
$$\mathbf{w} \sim N(\mathbf{0}, \sigma^2\mathbf{R}(\phi)), \text{ where } \mathbf{R}(\phi) = [\rho(\phi; \|\mathbf{s}_i - \mathbf{s}_j\|)]_{i,j=1}^n$$



Realization of a Gaussian process:

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

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



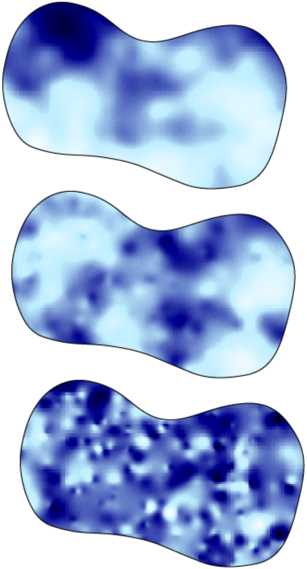
Realization of a Gaussian process:

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

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

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

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



Realization of a Gaussian process:

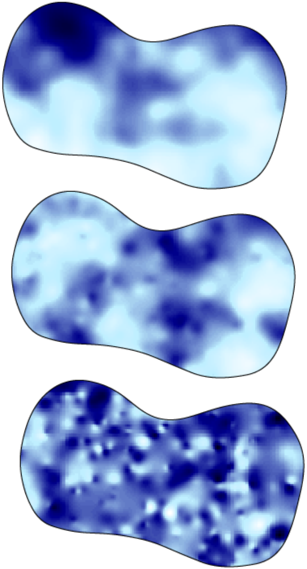
- Changing ϕ and holding $\sigma^2 = 1$:

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

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

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

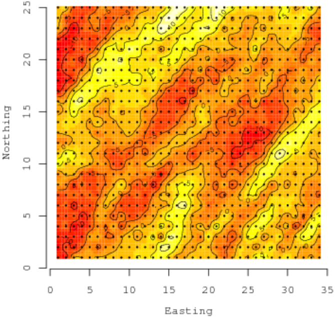
- Other valid models e.g., Gaussian, Spherical, Matérn.
- Effective range, $t_0 = -ln(0.05)/\phi \approx 3/\phi$



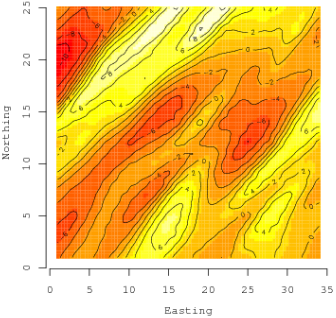
$\mathbf{w} \sim N(\mathbf{0}, \sigma^2 \mathbf{R}(\phi))$ provides complex spatial dependence through simple structured dependence.

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

Simulated



Predicted



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: $\mu = \mathbf{x}(\mathbf{s})^\top \boldsymbol{\beta}$
- Spatial random effects: $w(\mathbf{s}) \sim GP(0, \sigma^2 \rho(\phi; \|\mathbf{s}_1 - \mathbf{s}_2\|))$
- Non-spatial variance: $\epsilon(\mathbf{s}) \overset{iid}{\sim} N(0, \tau^2)$. Interpretation as pure error, measurement error, replication error, microscale error.

Hierarchical modeling

- First stage:

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

Hierarchical modeling

- First stage:

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

- Second stage:

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

Hierarchical modeling

- First stage:

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

- Second stage:

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

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

Hierarchical modeling

- First stage:

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

- Second stage:

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

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

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

Hierarchical modeling

- First stage:

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

- Second stage:

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

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

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

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

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 $\boldsymbol{\beta}$, σ^2, τ^2 and $\mathbf{w}$ – easier to program.

Bayesian Computations

- Choice: Fit $[\mathbf{y}|\Omega] \times [\Omega]$ or $[\mathbf{y}|\beta, \mathbf{w}, \tau^2] \times [\mathbf{w}|\sigma^2, \phi] \times [\Omega]$.
- Conditional model:
  - conjugate full conditionals for β, σ^2, τ^2 and $\mathbf{w}$ – 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
  - $\sigma^2\mathbf{R}(\phi) + \tau^2\mathbf{I}$ is more stable than $\sigma^2\mathbf{R}(\phi)$.

Bayesian Computations

- Choice: Fit $[\mathbf{y}|\Omega] \times [\Omega]$ or $[\mathbf{y}|\beta, \mathbf{w}, \tau^2] \times [\mathbf{w}|\sigma^2, \phi] \times [\Omega]$.
- Conditional model:
  - conjugate full conditionals for β, σ^2, τ^2 and $\mathbf{w}$ – 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
  - $\sigma^2\mathbf{R}(\phi) + \tau^2\mathbf{I}$ is more stable than $\sigma^2\mathbf{R}(\phi)$.
- But what about $\mathbf{R}^{-1}(\phi)$?? EXPENSIVE!

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

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

Where are the $\mathbf{w}$'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 $\Omega^{(1)}, \ldots, \Omega^{(G)} \sim [\Omega|\mathbf{y}, \mathbf{X}]$
  - For each $\Omega^{(g)}$, draw $\mathbf{w}^{(g)} \sim [\mathbf{w}|\Omega^{(g)}, \mathbf{y}, \mathbf{X}]$
- NOTE: With Gaussian likelihoods $[\mathbf{w}|\Omega, \mathbf{y}, \mathbf{X}]$ is also Gaussian. With other likelihoods this may not be a standard distribution; conditional updating scheme is preferred.

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

- Sample from predictive distribution:

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

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