Hierarchical Modelling for Univariate and Multivariate Spatial Data

Sudipto Banerjee

sudiptob@biostat.umn.edu

University of Minnesota

Introduction

- Researchers in diverse areas such as climatology, ecology, environmental health, and real estate marketing are increasingly faced with the task of analyzing data that are:

Introduction

- Researchers in diverse areas such as climatology, ecology, environmental health, and real estate marketing are increasingly faced with the task of analyzing data that are:
  - highly multivariate, with many important predictors and response variables,

Introduction

- Researchers in diverse areas such as climatology, ecology, environmental health, and real estate marketing 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

Introduction

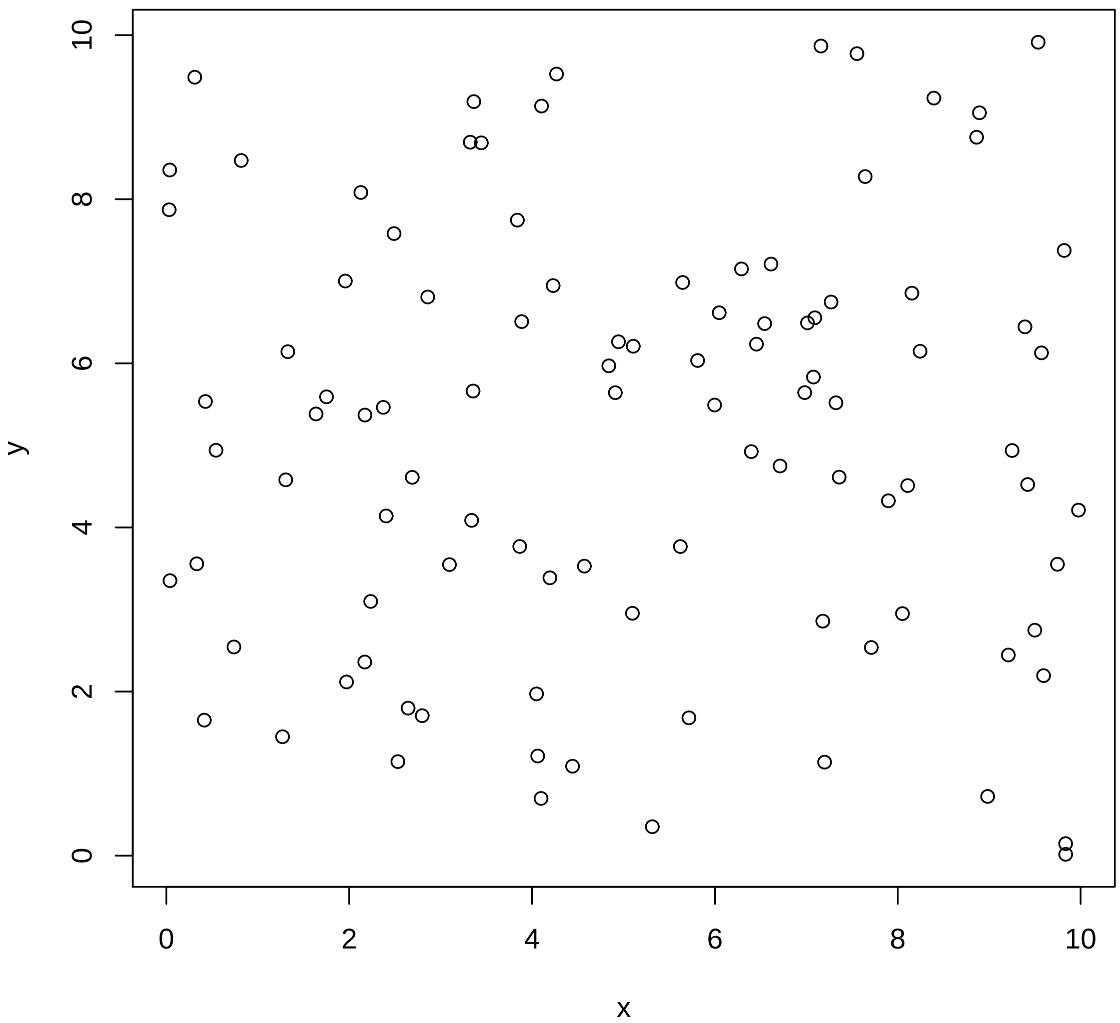
- Researchers in diverse areas such as climatology, ecology, environmental health, and real estate marketing are increasingly faced with the task of analyzing data that are:
  - highly multivariate, with many important predictors and response variables,
  - geographically referenced, and often presented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

Introduction

- Researchers in diverse areas such as climatology, ecology, environmental health, and real estate marketing 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.

Spatial Domain



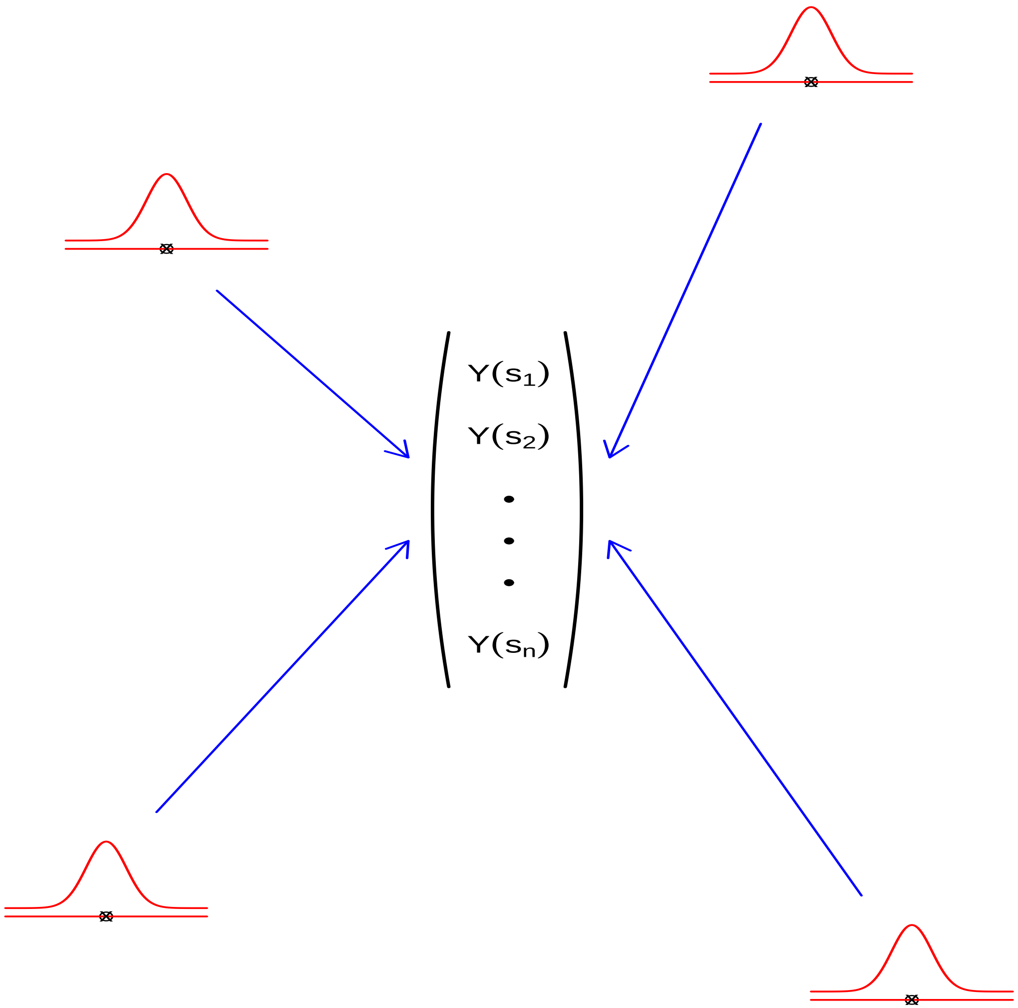
Algorithmic Modelling

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

What is a spatial process?



Mean, Covariance and Error

- Spatial model:

$$Y(\mathrm{s}) = \mu(\mathbf{s}) + w(\mathbf{s}) + \epsilon(\mathbf{s})$$

  - Mean: $\boldsymbol{\mu}(\mathbf{s}) = \mathbf{x}^T(\mathbf{s})\boldsymbol{\beta}$
  - Covariance: $w(\mathbf{s}) \sim GP(\mathbf{0}, \sigma^2\rho(\cdot))$

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

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

- For $\mathcal{S} = \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ let $\mathbf{w} = [w(\mathbf{s}_i)]$. Then

$$\mathbf{w} \sim MVN(\mathbf{0}, \sigma^2 H(\phi));\ H(\phi) = [\rho(\phi; \|\mathbf{s}_i - \mathbf{s}_j\|)]$$

The exponential decay model

- $H(\phi)$ must be p.d.
- $\rho(\cdot)$ must be somewhat *special*
- $\rho(\phi; t) = E_X[e^{i\phi t X}]$, X is symmetric about 0.

The exponential decay model

- $H(\phi)$ must be p.d.
- $\rho(\cdot)$ must be somewhat *special*
- $\rho(\phi; t) = E_X[e^{i\phi t X}]$, X is symmetric about 0.
- Example: exponential decay model

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

The exponential decay model

- $H(\phi)$ must be p.d.
- $\rho(\cdot)$ must be somewhat *special*
- $\rho(\phi; t) = E_X[e^{i\phi t X}]$, X is symmetric about 0.
- Example: exponential decay model

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

- *Effective range:* $t_0 = \arg \rho(\phi; t) = 0.05 \approx 3/\phi$.

The exponential decay model

- $H(\phi)$ must be p.d.
- $\rho(\cdot)$ must be somewhat *special*
- $\rho(\phi; t) = E_X[e^{i\phi t X}]$, X is symmetric about 0.
- Example: exponential decay model

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

- *Effective range:* $t_0 = \arg \rho(\phi; t) = 0.05 \approx 3/\phi$.
- Spatial variance components:
  - The "nugget" τ^2 is a "nonspatial effect variance"
  - The "partial sill" (σ^2) is a "spatial effect variance". The sill is $\sigma^2 + \tau^2$.

Hierarchical modelling

- First stage:

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

Hierarchical modelling

- First stage:

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

- Second stage:

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

Hierarchical modelling

- First stage:

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

- Second stage:

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

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

Hierarchical modelling

- First stage:

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

- Second stage:

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

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

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

Hierarchical modelling

- First stage:

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

- Second stage:

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

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

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

- Note: Spatial process parametrizes Σ:

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon},\ \boldsymbol{\epsilon} \sim N\left(\mathbf{0}, \Sigma\right),\ \Sigma = \sigma^2 H(\phi) + \tau^2 I$$

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 $\mathbf{w}$. 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
  - $\sigma^2 H(\phi) + \tau^2 I$ is more stable than $\sigma^2 H(\phi)$.

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
  - $\sigma^2 H(\phi) + \tau^2 I$ is more stable than $\sigma^2 H(\phi)$.
- But what about $H^{-1}(\phi)$?? EXPENSIVE!.

Where are the 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 $\Omega^{(1)}, \ldots, \Omega^{(G)} \sim [\Omega|\mathbf{y}, X]$

Where are the w's?

- Interest often lies in the spatial surface **w**$|$**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 $\mathbf{w}^{(g)} \sim [\mathbf{w}|\Omega^{(g)}, \mathbf{y}, X]$

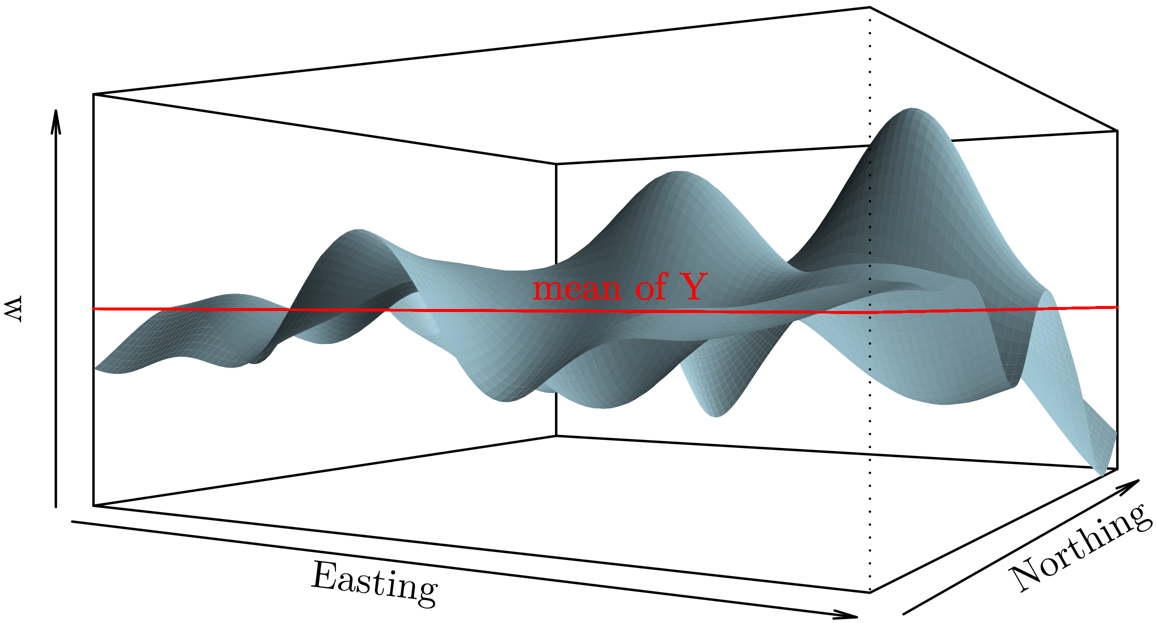
Where are the w's?

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

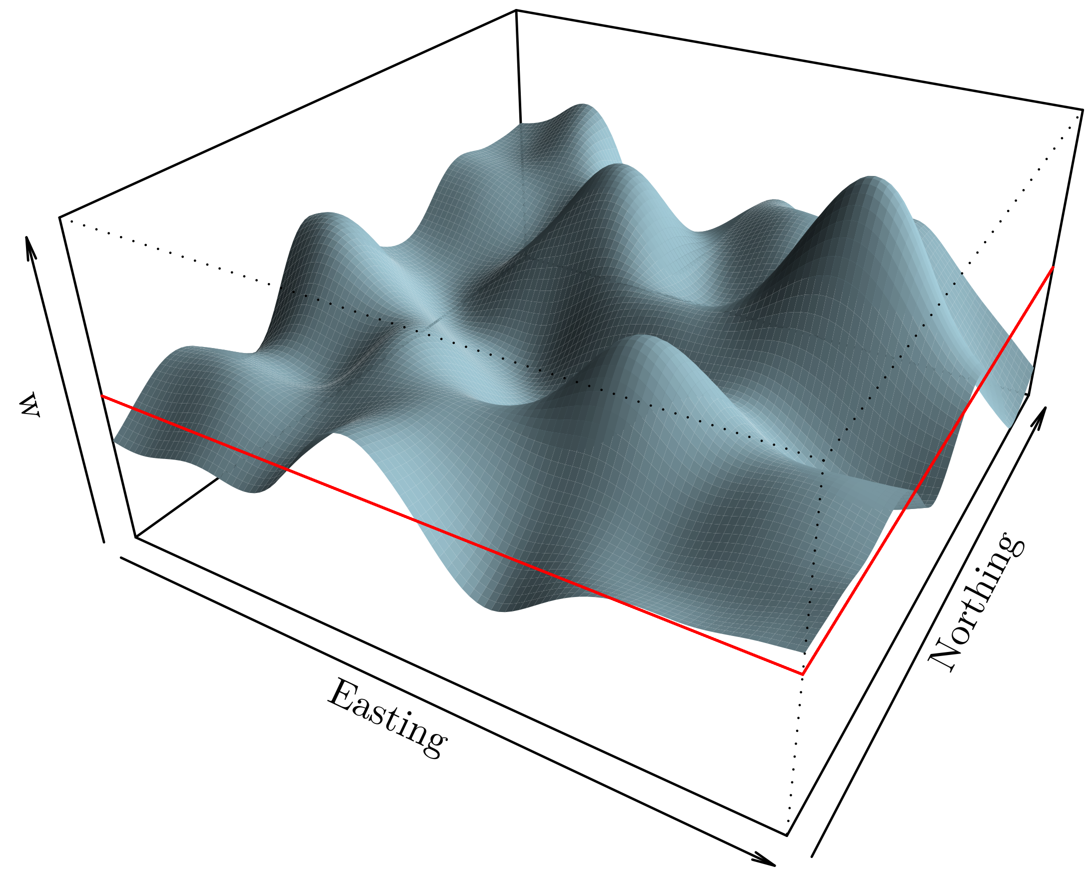
$$[\mathbf{w}|\mathbf{y}, X] = \int [\mathbf{w}|\Omega, \mathbf{y}, X] \times [\Omega|\mathbf{y}, X] d\Omega$$

  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.

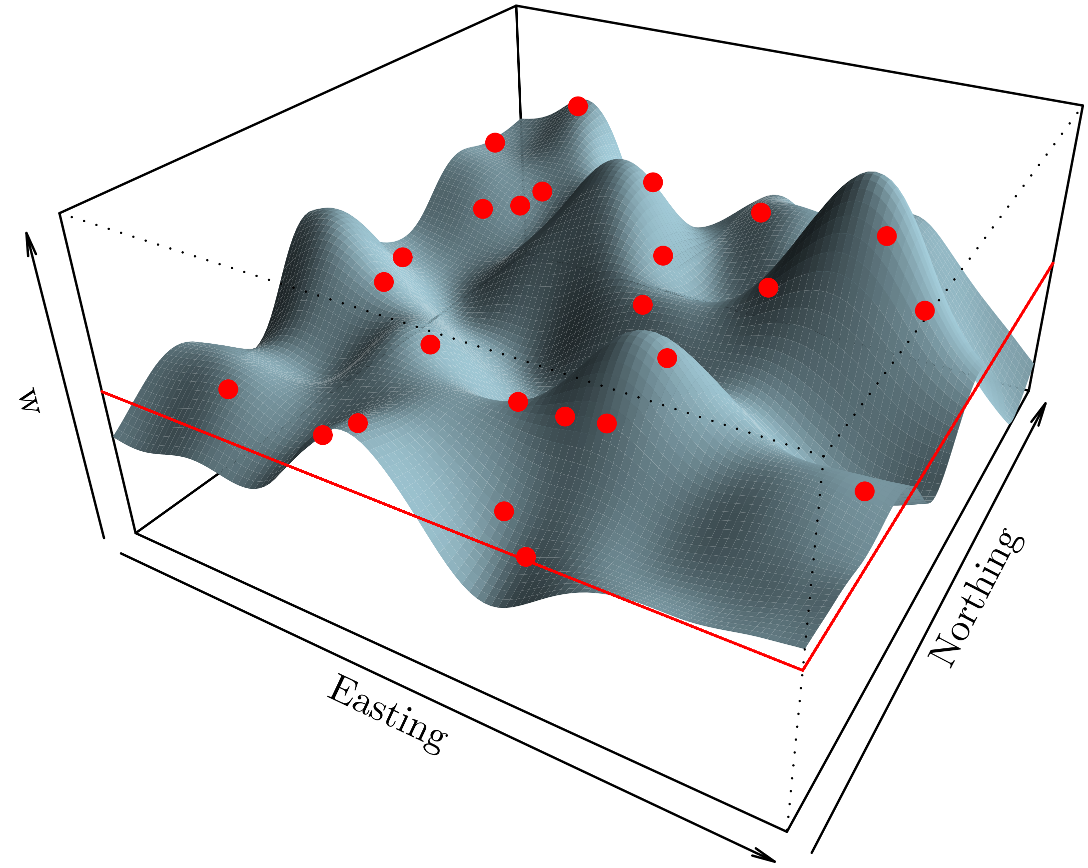
Residual plot: $[w(\mathbf{s})|\mathbf{y}]$



Another look: $[w(\mathbf{s})|\mathbf{y}]$



Another look: $[w(\mathbf{s})|\mathbf{y}]$



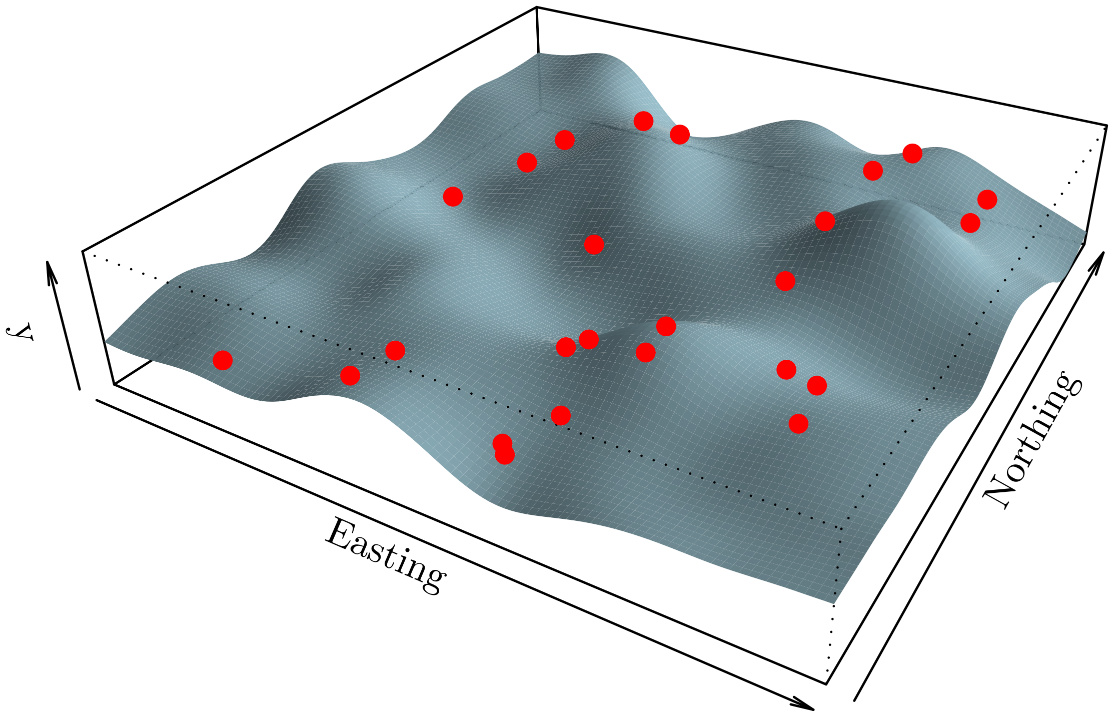
Spatial Prediction

- 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 $\tilde{\mathbf{y}}^{(g)} \sim [\tilde{\mathbf{y}}|\mathbf{y}, \Omega^{(g)}, X, \tilde{X}]$.

Prediction: Summary of $[Y(\mathbf{s})|\mathbf{y}]$



Colorado data illustration

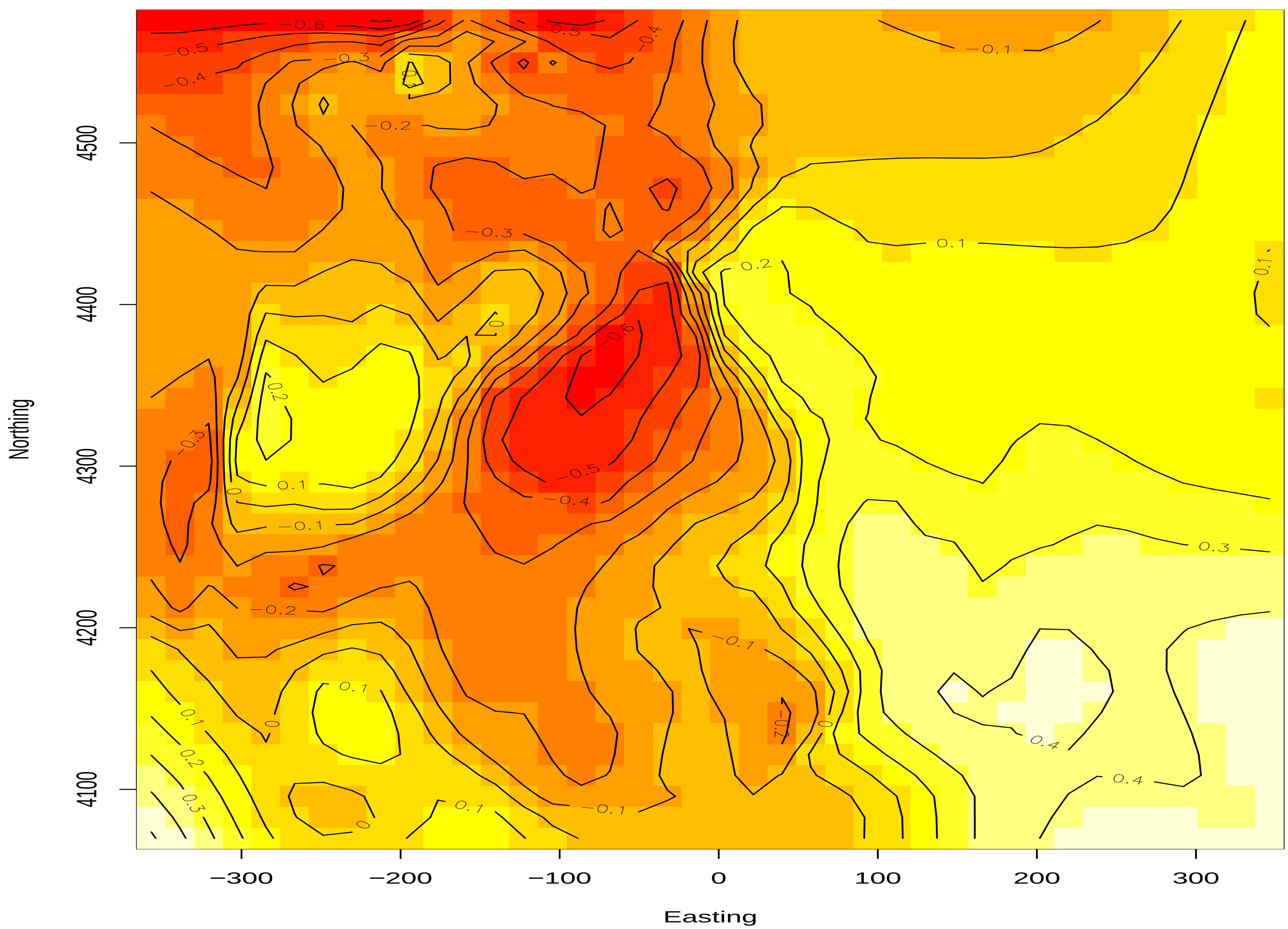
- Modelling temperature: 507 locations in Colorado.
- Simple spatial regression model:

$$Y(\mathbf{s}) = \mathbf{x}^T(\mathbf{s})\boldsymbol{\beta} + w(\mathbf{s}) + \epsilon(\mathbf{s})$$

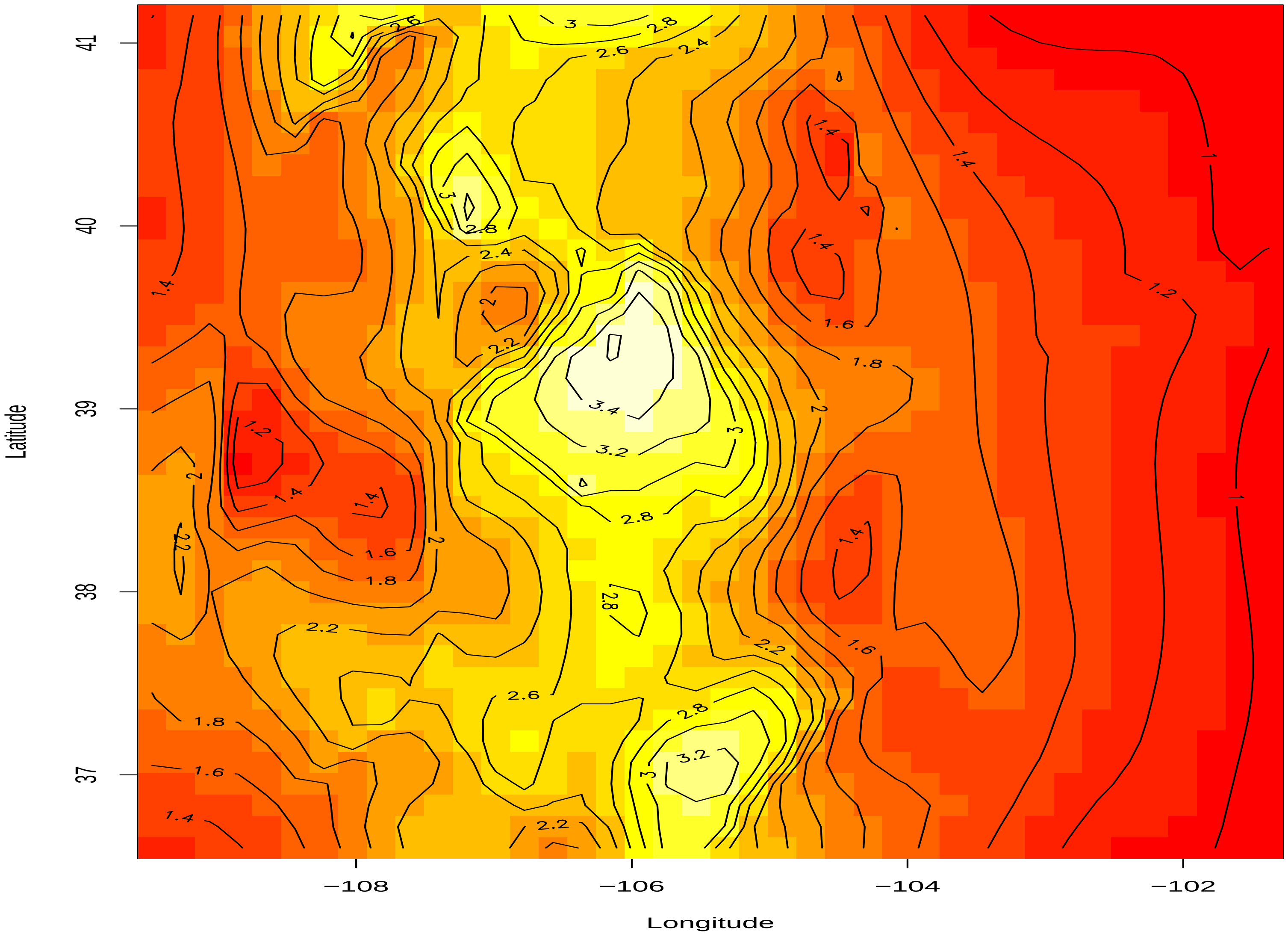
- $w(\mathbf{s}) \sim GP(0, \sigma^2\rho(\cdot;\phi,\nu))$; $\epsilon(\mathbf{s}) \overset{iid}{\sim} N(0,\tau^2)$

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)

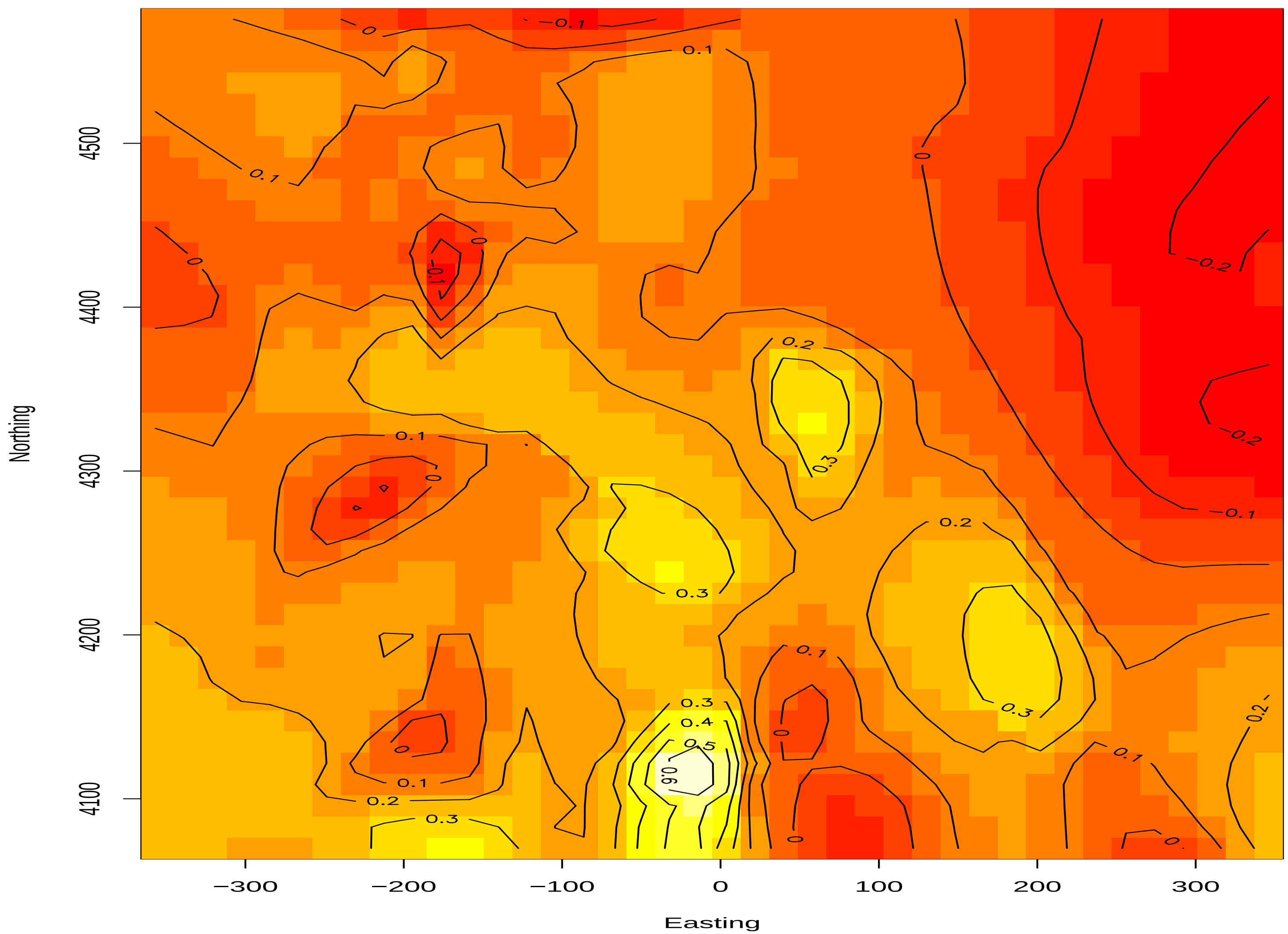
Temperature residual map



Elevation map



Residual map with elev. as covariate



Spatial Generalized Linear Models

- Often data sets preclude Gaussian modelling: $Y(\mathbf{s})$ may not even be continuous

Spatial Generalized Linear Models

- Often data sets preclude Gaussian modelling: $Y(\mathbf{s})$ may not even be continuous
- Example: $Y(\mathbf{s})$ is a binary or count variable

Spatial Generalized Linear Models

- Often data sets preclude Gaussian modelling: $Y(\mathbf{s})$ may not even be continuous
- Example: $Y(\mathbf{s})$ is a binary or count variable
  - rainfall was measurable or not

Spatial Generalized Linear Models

- Often data sets preclude Gaussian modelling: $Y(\mathbf{s})$ may not even be continuous
- Example: $Y(\mathbf{s})$ is a binary or count variable
  - rainfall was measurable or not
  - number of insurance claims by residents of a single family home at $\mathbf{s}$

Spatial Generalized Linear Models

- Often data sets preclude Gaussian modelling: $Y(\mathbf{s})$ may not even be continuous
- Example: $Y(\mathbf{s})$ is a binary or count variable
  - rainfall was measurable or not
  - number of insurance claims by residents of a single family home at **s**
  - price is high or low for home at location **s**

Spatial Generalized Linear Models

- Often data sets preclude Gaussian modelling: $Y(\mathbf{s})$ may not even be continuous
- Example: $Y(\mathbf{s})$ is a binary or count variable
  - rainfall was measurable or not
  - number of insurance claims by residents of a single family home at $\mathbf{s}$
  - price is high or low for home at location $\mathbf{s}$
- Replace Gaussian likelihood by exponential family member

Spatial Generalized Linear Models

- Often data sets preclude Gaussian modelling: $Y(\mathbf{s})$ may not even be continuous
- Example: $Y(\mathbf{s})$ is a binary or count variable
  - rainfall was measurable or not
  - number of insurance claims by residents of a single family home at $\mathbf{s}$
  - price is high or low for home at location $\mathbf{s}$
- Replace Gaussian likelihood by exponential family member
- Diggle Tawn and Moyeed (1998)

Spatial GLM (contd.)

- First stage: $Y(\mathbf{s}_i)$ are conditionally independent given $\boldsymbol{\beta}$ and $w(\mathbf{s}_i)$, so $f(y(\mathbf{s}_i)|\boldsymbol{\beta}, w(\mathbf{s}_i), \gamma)$ equals

  $$h(y(\mathbf{s}_i), \gamma) \exp\left(\gamma[y(\mathbf{s}_i)\eta(\mathbf{s}_i) - \psi(\eta(\mathbf{s}_i))]\right)$$

  where $g(E(Y(\mathbf{s}_i))) = \eta(\mathbf{s}_i) = \mathbf{x}^T(\mathbf{s}_i)\boldsymbol{\beta} + w(\mathbf{s}_i)$ (canonical link function) and γ is a dispersion parameter.

Spatial GLM (contd.)

- First stage: $Y(\mathbf{s}_i)$ are conditionally independent given $\boldsymbol{\beta}$ and $w(\mathbf{s}_i)$, so $f(y(\mathbf{s}_i)|\boldsymbol{\beta}, w(\mathbf{s}_i), \gamma)$ equals

$$h(y(\mathbf{s}_i), \gamma) \exp\left(\gamma[y(\mathbf{s}_i)\eta(\mathbf{s}_i) - \psi(\eta(\mathbf{s}_i))]\right)$$

  where $g(E(Y(\mathbf{s}_i))) = \eta(\mathbf{s}_i) = \mathbf{x}^T(\mathbf{s}_i)\boldsymbol{\beta} + w(\mathbf{s}_i)$ (canonical link function) and γ is a dispersion parameter.

- Second stage: Model $w(\mathbf{s})$ as a Gaussian process:

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

Spatial GLM (contd.)

- First stage: $Y(\mathbf{s}_i)$ are conditionally independent given $\boldsymbol{\beta}$ and $w(\mathbf{s}_i)$, so $f(y(\mathbf{s}_i)|\boldsymbol{\beta}, w(\mathbf{s}_i), \gamma)$ equals

$$h(y(\mathbf{s}_i), \gamma) \exp\left(\gamma[y(\mathbf{s}_i)\eta(\mathbf{s}_i) - \psi(\eta(\mathbf{s}_i))]\right)$$

  where $g(E(Y(\mathbf{s}_i))) = \eta(\mathbf{s}_i) = \mathbf{x}^T(\mathbf{s}_i)\boldsymbol{\beta} + w(\mathbf{s}_i)$ (canonical link function) and γ is a dispersion parameter.

- Second stage: Model $w(\mathbf{s})$ as a Gaussian process:

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

- Third stage: Priors and hyperpriors.

Spatial GLM (contd.)

- First stage: $Y(\mathbf{s}_i)$ are conditionally independent given $\boldsymbol{\beta}$ and $w(\mathbf{s}_i)$, so $f(y(\mathbf{s}_i)|\boldsymbol{\beta}, w(\mathbf{s}_i), \gamma)$ equals

$$h(y(\mathbf{s}_i), \gamma) \exp\left(\gamma[y(\mathbf{s}_i)\eta(\mathbf{s}_i) - \psi(\eta(\mathbf{s}_i))]\right)$$

  where $g(E(Y(\mathbf{s}_i))) = \eta(\mathbf{s}_i) = \mathbf{x}^T(\mathbf{s}_i)\boldsymbol{\beta} + w(\mathbf{s}_i)$ (canonical link function) and γ is a dispersion parameter.

- Second stage: Model $w(\mathbf{s})$ as a Gaussian process:

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

- Third stage: Priors and hyperpriors.
- No process for $Y(\mathbf{s})$, only a valid joint distribution

Spatial GLM (contd.)

- First stage: $Y(\mathbf{s}_i)$ are conditionally independent given $\boldsymbol{\beta}$ and $w(\mathbf{s}_i)$, so $f(y(\mathbf{s}_i)|\boldsymbol{\beta}, w(\mathbf{s}_i), \gamma)$ equals

$$h(y(\mathbf{s}_i), \gamma) \exp\left(\gamma[y(\mathbf{s}_i)\eta(\mathbf{s}_i) - \psi(\eta(\mathbf{s}_i))]\right)$$

  where $g(E(Y(\mathbf{s}_i))) = \eta(\mathbf{s}_i) = \mathbf{x}^T(\mathbf{s}_i)\boldsymbol{\beta} + w(\mathbf{s}_i)$ (canonical link function) and γ is a dispersion parameter.

- Second stage: Model $w(\mathbf{s})$ as a Gaussian process:

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

- Third stage: Priors and hyperpriors.
- No process for $Y(\mathbf{s})$, only a valid joint distribution
- Not sensible to add a pure error term $\epsilon(\mathbf{s})$

Spatial GLM: comments

- We are modelling with spatial random effects

Spatial GLM: comments

- We are modelling with spatial random effects
- Introducing these in the transformed mean encourages means of spatial variables at proximate locations to be close to each other

Spatial GLM: comments

- We are modelling with spatial random effects
- Introducing these in the transformed mean encourages means of spatial variables at proximate locations to be close to each other
- Marginal spatial dependence is induced between, say, $Y(\mathbf{s})$ and $Y(\mathbf{s}')$, but observed $Y(\mathbf{s})$ and $Y(\mathbf{s}')$ need not be close to each other

Spatial GLM: comments

- We are modelling with spatial random effects
- Introducing these in the transformed mean encourages means of spatial variables at proximate locations to be close to each other
- Marginal spatial dependence is induced between, say, $Y(\mathbf{s})$ and $Y(\mathbf{s}')$, but observed $Y(\mathbf{s})$ and $Y(\mathbf{s}')$ need not be close to each other
- Second stage spatial modelling is attractive for spatial explanation in the *mean*

Spatial GLM: comments

- We are modelling with spatial random effects
- Introducing these in the transformed mean encourages means of spatial variables at proximate locations to be close to each other
- Marginal spatial dependence is induced between, say, $Y(\mathbf{s})$ and $Y(\mathbf{s}')$, but observed $Y(\mathbf{s})$ and $Y(\mathbf{s}')$ need not be close to each other
- Second stage spatial modelling is attractive for spatial explanation in the *mean*
- First stage spatial modelling more appropriate to encourage proximate observations to be close.

Binary spatial regression: Home prices

Here we illustrate a non-Gaussian model for point-referenced spatial data:

- Data: Observations are home values (based on recent real estate sales) at 50 locations in Baton Rouge, Louisiana, USA.

Binary spatial regression: Home prices

Here we illustrate a non-Gaussian model for point-referenced spatial data:

- Data: Observations are home values (based on recent real estate sales) at 50 locations in Baton Rouge, Louisiana, USA.
- The response $Y(\mathbf{s})$ is a binary variable, with

$$Y(\mathbf{s}) = \begin{cases} 1 & \text{if price is “high” (above the median)} \\ 0 & \text{if price is “low” (below the median)} \end{cases}$$

Binary spatial regression: Home prices

Here we illustrate a non-Gaussian model for point-referenced spatial data:

- Data: Observations are home values (based on recent real estate sales) at 50 locations in Baton Rouge, Louisiana, USA.
- The response $Y(\mathbf{s})$ is a binary variable, with

$$Y(\mathbf{s}) = \begin{cases} 1 & \text{if price is “high” (above the median)} \\ 0 & \text{if price is “low” (below the median)} \end{cases}$$

- Observed covariates include the house’s age and total living area

Binary spatial regression: Home prices

- We fit a generalized linear model where
  $Y(\mathbf{s}_i) \sim Bernoulli(p(\mathbf{s}_i)), \;\; logit(p(\mathbf{s}_i)) = \mathbf{x}^T(\mathbf{s}_i)\boldsymbol{\beta} + w(\mathbf{s}_i)$

Binary spatial regression: Home prices

- We fit a generalized linear model where $Y(\mathbf{s}_i) \sim Bernoulli(p(\mathbf{s}_i)), \;\; logit(p(\mathbf{s}_i)) = \mathbf{x}^T(\mathbf{s}_i)\boldsymbol{\beta} + w(\mathbf{s}_i)$
- Assume vague priors for $\boldsymbol{\beta}$, a Uniform$(0, 10)$ prior for ϕ, and an Inverse Gamma$(0.1, 0.1)$ prior for σ^2.

Binary spatial regression: Home prices

- We fit a generalized linear model where $Y(\mathbf{s}_i) \sim Bernoulli(p(\mathbf{s}_i)), \;\; logit(p(\mathbf{s}_i)) = \mathbf{x}^T(\mathbf{s}_i)\boldsymbol{\beta} + w(\mathbf{s}_i)$
- Assume vague priors for $\boldsymbol{\beta}$, a Uniform$(0, 10)$ prior for ϕ, and an Inverse Gamma$(0.1, 0.1)$ prior for σ^2.
- The WinBUGS code:

```
for (i in 1:N) {
    Y[i] ~ dbern(p[i])
    logit(p[i]) <- w[i]
    mu[i] <- beta[1]+beta[2]*LivingArea[i]/1000+beta[3]*Age[i] }
for (i in 1:3) beta[i] ~ dnorm(0.0,0.001)
w[1:N] ~ spatial.exp(mu[], x[], y[], spat.prec, phi, 1)
phi ~ dunif(0.1,10)
spat.prec ~ dgamma(0.1, 0.1)
sigmasq <- 1/spat.prec
```

Image plot of $w(\mathbf{s})$ surface

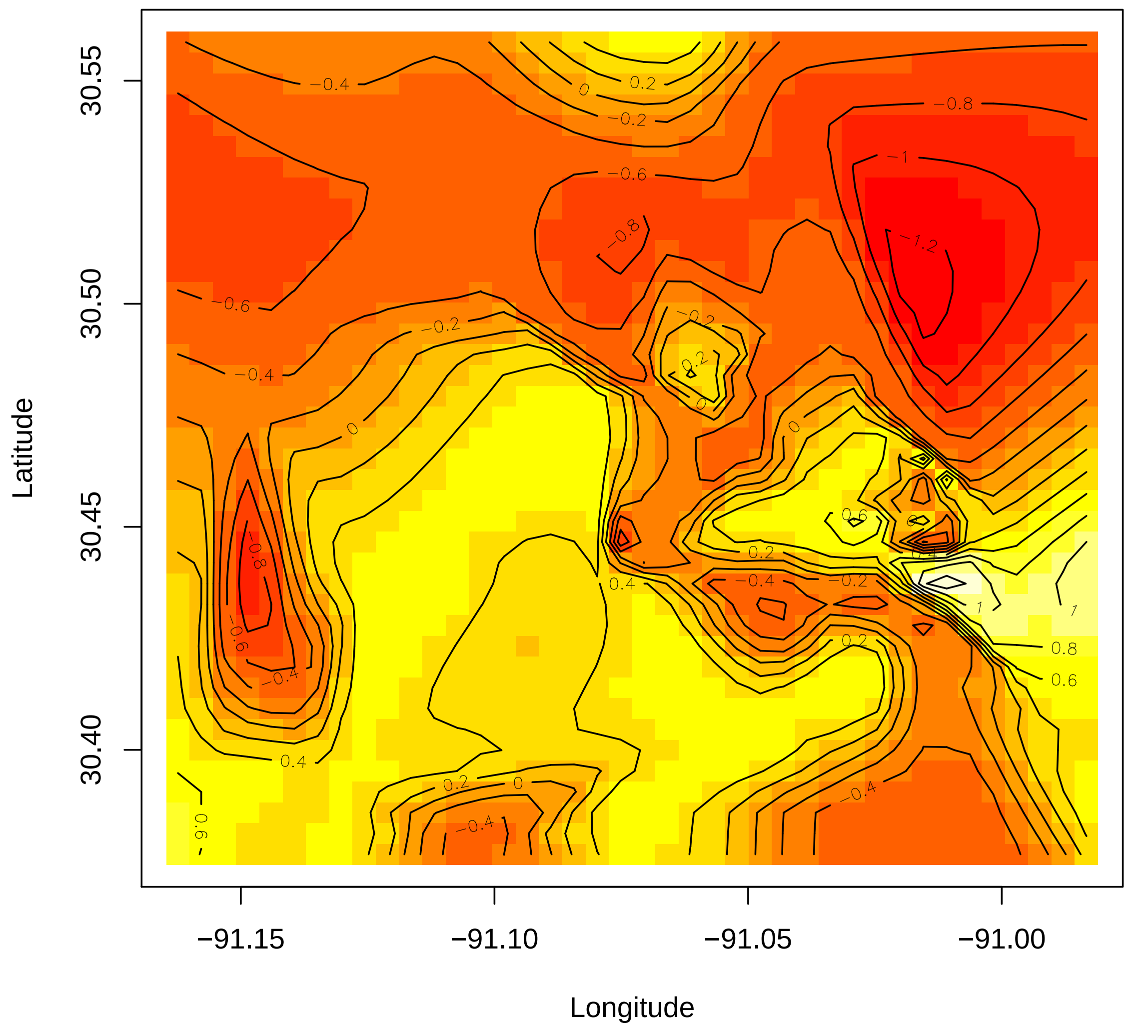


Image plot of $w(\mathbf{s})$ surface

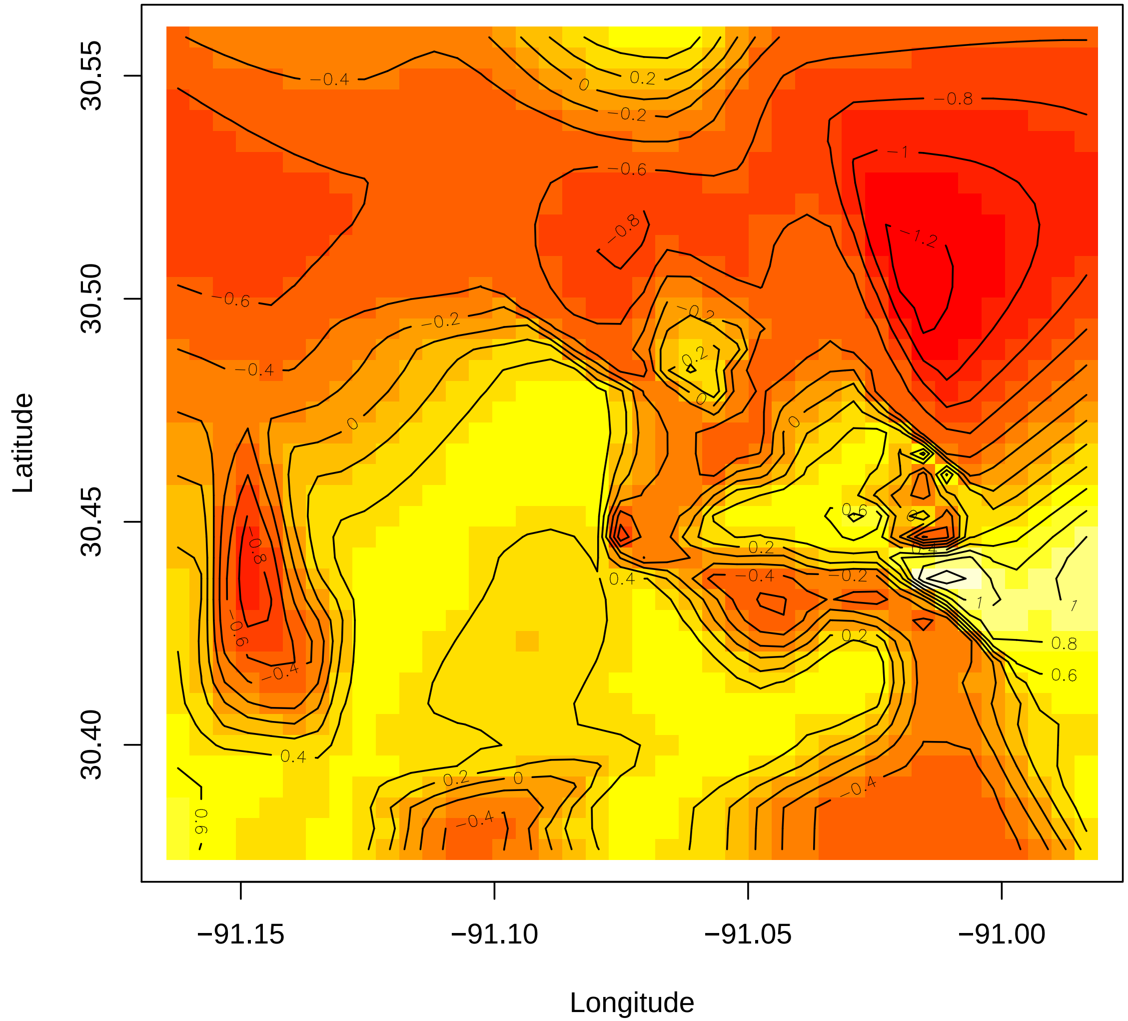
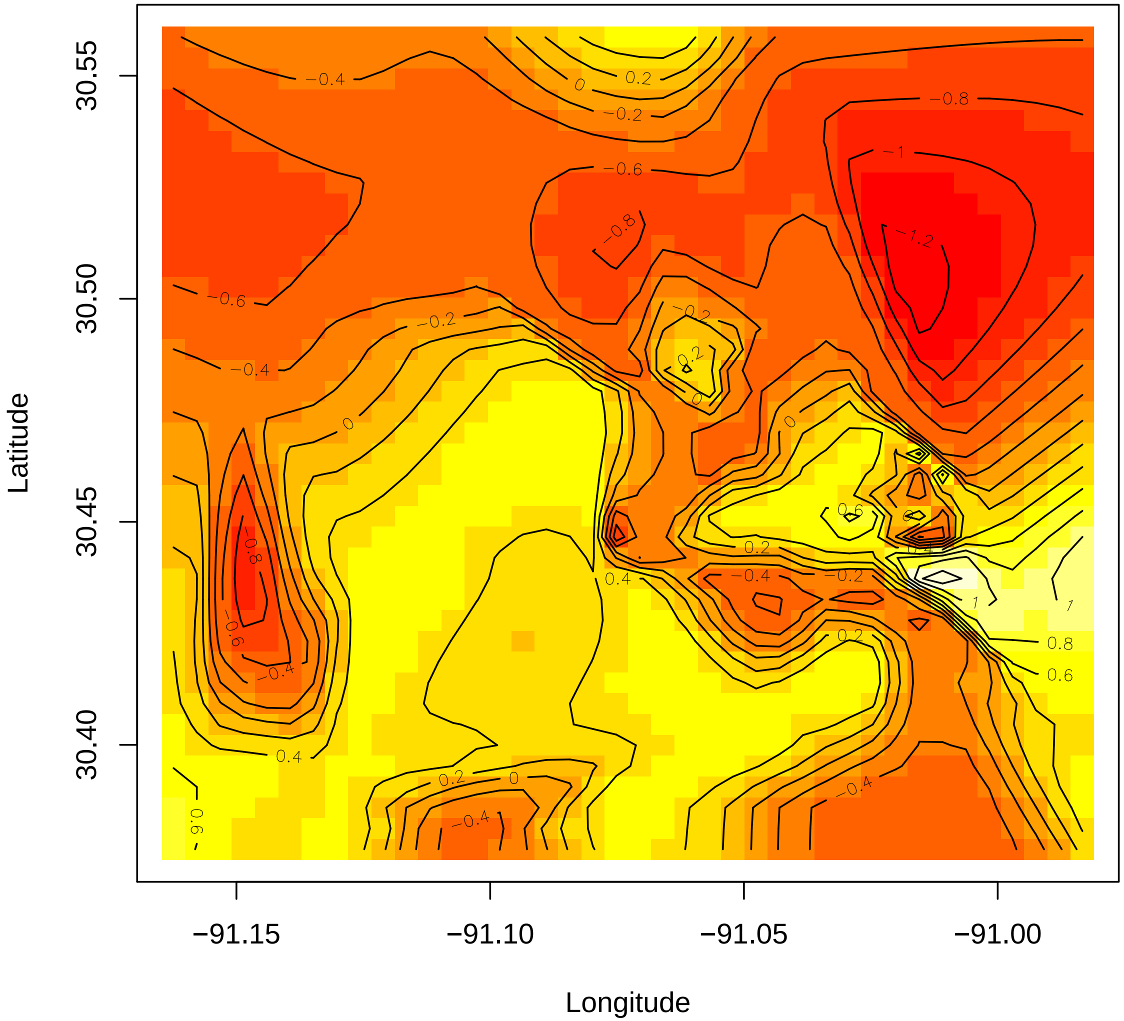


Image plot of $w(\mathbf{s})$ surface



Posterior parameter estimates

Parameter estimates (posterior medians and upper and lower .025 percentiles):

Parameter	50%	(2.5%, 97.5%)
β_1 (intercept)	–1.096	(–4.198, 0.4305)
β_2 (living area)	0.659	(–0.091, 2.254)
β_3 (age)	0.009615	(–0.8653, 0.7235)
ϕ	5.79	(1.236, 9.765)
σ^2	1.38	(0.1821, 6.889)

The covariate effects are generally uninteresting, though living area seems to have a marginally significant effect on price class.

Multivariate spatial modeling

- Point-referenced spatial data often come as multivariate measurements at each location

Multivariate spatial modeling

- Point-referenced spatial data often come as multivariate measurements at each location
- Examples:

Multivariate spatial modeling

- Point-referenced spatial data often come as multivariate measurements at each location
- Examples:
  - Environmental monitoring stations yield measurements on ozone, NO, CO, $PM_{2.5}$, etc.

Multivariate spatial modeling

- Point-referenced spatial data often come as multivariate measurements at each location
- Examples:
  - Environmental monitoring stations yield measurements on ozone, NO, CO, $PM_{2.5}$, etc.
  - In atmospheric modeling at a given site we observe surface temperature, precipitation and wind speed

Multivariate spatial modeling

- Point-referenced spatial data often come as multivariate measurements at each location
- Examples:
  - Environmental monitoring stations yield measurements on ozone, NO, CO, $PM_{2.5}$, etc.
  - In atmospheric modeling at a given site we observe surface temperature, precipitation and wind speed
  - In real estate modeling for an individual property we observe selling price and total rental income

Multivariate spatial modeling

- Point-referenced spatial data often come as multivariate measurements at each location
- Examples:
  - Environmental monitoring stations yield measurements on ozone, NO, CO, $PM_{2.5}$, etc.
  - In atmospheric modeling at a given site we observe surface temperature, precipitation and wind speed
  - In real estate modeling for an individual property we observe selling price and total rental income
- We anticipate dependence between measurements

Multivariate spatial modeling

- Point-referenced spatial data often come as multivariate measurements at each location
- Examples:
  - Environmental monitoring stations yield measurements on ozone, NO, CO, $PM_{2.5}$, etc.
  - In atmospheric modeling at a given site we observe surface temperature, precipitation and wind speed
  - In real estate modeling for an individual property we observe selling price and total rental income
- We anticipate dependence between measurements
  - at a particular location

Multivariate spatial modeling

- Point-referenced spatial data often come as multivariate measurements at each location
- Examples:
  - Environmental monitoring stations yield measurements on ozone, NO, CO, $PM_{2.5}$, etc.
  - In atmospheric modeling at a given site we observe surface temperature, precipitation and wind speed
  - In real estate modeling for an individual property we observe selling price and total rental income
- We anticipate dependence between measurements
  - at a particular location
  - across locations

Multivariate spatial regression

- Each location contains m spatial regressions

$$Y_k(\mathbf{s}) = \mu_k(\mathbf{s}) + w_k(\mathbf{s}) + \epsilon_k(\mathbf{s}),\; k = 1, \ldots, m.$$

  - Mean: $\boldsymbol{\mu}(\mathbf{s}) = [\mu_k(\mathbf{s})]_{k=1}^m = [\mathbf{x}_k^T(\mathbf{s})\boldsymbol{\beta}_k]_{k=1}^m$
  - Cov: $\mathbf{w}(\mathbf{s}) = [w_k(\mathbf{s})]_{k=1}^m \sim MVGP(\mathbf{0}, \Gamma_{\mathbf{w}}(\cdot, \cdot))$

$$\Gamma_{\mathbf{w}}(\mathbf{s}, \mathbf{s}') = [Cov(w_k(\mathbf{s}), w_{k'}(\mathbf{s}'))]_{k,k'=1}^m$$

  - Error: $\boldsymbol{\epsilon}(\mathbf{s}) = [\epsilon_k(\mathbf{s})]_{k=1}^m \sim MVN(\mathbf{0}, \Psi)$

- Ψ is an $m \times m$ p.d. matrix, e.g. usually $Diag(\tau_k^2)_{k=1}^m$.

Multivariate Gaussian Process

- $\mathbf{w}(\mathbf{s}) \sim MVGP(\mathbf{0}, \Gamma_{\mathbf{w}}(\cdot))$ with

$$\Gamma_{\mathbf{w}}(\mathbf{s}, \mathbf{s}') = [Cov(w_k(\mathbf{s}), w_{k'}(\mathbf{s}'))]_{k,k'=1}^{m}$$

- Example: with $m = 2$

$$\Gamma_{\mathbf{w}}(\mathbf{s}, \mathbf{s}') = \begin{pmatrix} Cov(w_1(\mathbf{s}), w_1(\mathbf{s}')) & Cov(w_1(\mathbf{s}), w_2(\mathbf{s}')) \\ Cov(w_2(\mathbf{s}), w_1(\mathbf{s}')) & Cov(w_2(\mathbf{s}), w_2(\mathbf{s}')) \end{pmatrix}$$

- For finite set of locations $\mathcal{S} = \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$:

$$Var\left([\mathbf{w}(\mathbf{s}_i)]_{i=1}^{n}\right) = \Sigma_{\mathbf{w}} = [\Gamma_{\mathbf{w}}(\mathbf{s}_i, \mathbf{s}_j)]_{i,j=1}^{n}$$

contd.

- Properties:
  - $\Gamma_{\mathbf{w}}(\mathbf{s}', \mathbf{s}) = \Gamma_{\mathbf{w}}^T(\mathbf{s}, \mathbf{s}')$
  - $\lim_{\mathbf{s} \to \mathbf{s}'} \Gamma_{\mathbf{w}}(\mathbf{s}, \mathbf{s}')$ is p.d. and $\Gamma_{\mathbf{w}}(\mathbf{s}, \mathbf{s}) = Var(\mathbf{w}(\mathbf{s}))$.
  - For sites in any finite collection $\mathcal{S} = \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{u}_i^T \Gamma_{\mathbf{w}}(\mathbf{s}_i, \mathbf{s}_j) \mathbf{u}_j \geq 0 \text{ for all } \mathbf{u}_i, \mathbf{u}_j \in \Re^m.$$

- Any *valid* $\Gamma_{\mathbf{w}}$ must satisfy the above conditions.
- The last property implies that $\Sigma_{\mathbf{w}}$ is p.d.
- In complete generality:
  - $\Gamma_{\mathbf{w}}(\mathbf{s}, \mathbf{s}')$ need *not* be symmetric.
  - $\Gamma_{\mathbf{w}}(\mathbf{s}, \mathbf{s}')$ need *not* be p.d. for $\mathbf{s} \neq \mathbf{s}'$.

Modelling cross-covariances

- Moving average or kernel convolution of a process:
  - Let $Z(\mathbf{s}) \sim GP(0, \rho(\cdot))$. Use kernels to form:

$$w_j(\mathbf{s}) = \int \kappa_j(\mathbf{u}) Z(\mathbf{s}+\mathbf{u}) d\mathbf{u} = \int \kappa_j(\mathbf{s}-\mathbf{s}') Z(\mathbf{s}') d\mathbf{s}'$$

  - $\Gamma_{\mathbf{w}}(\mathbf{s}-\mathbf{s}')$ has (i,j)-th element:

$$[\Gamma_{\mathbf{w}}(\mathbf{s}-\mathbf{s}')]_{i,j} = \int\int \kappa_i(\mathbf{s}-\mathbf{s}'+\mathbf{u}) \kappa_j(\mathbf{u}') \rho(\mathbf{u}-\mathbf{u}') d\mathbf{u} d\mathbf{u}'$$

- Convolution of Covariance Functions:
  - $\rho_1, \rho_2, ...\rho_m$ are valid covariance functions. Form:

$$[\Gamma_{\mathbf{w}}(\mathbf{s}-\mathbf{s}')]_{i,j} = \int \rho_i(\mathbf{s}-\mathbf{s}'-\mathbf{t}) \rho_j(\mathbf{t}) d\mathbf{t}.$$

Constructive approach

- Let $v_k(\mathbf{s}) \sim GP(0, \rho_k(\mathbf{s}, \mathbf{s}'))$, for $k = 1, \ldots, m$ be m independent GP's with unit variance.
- Form the simple multivariate process $\mathbf{v}(\mathbf{s}) = [v_k(\mathbf{s})]_{k=1}^m$:

$$\mathbf{v}(\mathbf{s}) \sim MVGP(\mathbf{0}, \Gamma_{\mathbf{v}}(\cdot, \cdot))$$

with $\Gamma_{\mathbf{v}}(\mathbf{s}, \mathbf{s}') = Diag(\rho_k(\mathbf{s}, \mathbf{s}'))_{k=1}^m$.
- Assume $\mathbf{w}(\mathbf{s}) = A(\mathbf{s})\mathbf{v}(\mathbf{s})$ arises as a *space-varying* linear transformation of $\mathbf{v}(\mathbf{s})$. Then:

$$\Gamma_{\mathbf{w}}(\mathbf{s}, \mathbf{s}') = A(\mathbf{s})\Gamma_{\mathbf{v}}(\mathbf{s}, \mathbf{s}')A^T(\mathbf{s}')$$

is a valid cross-covariance function.

contd.

- When $\mathbf{s} = \mathbf{s}'$, $\Gamma_{\mathbf{v}}(\mathbf{s}, \mathbf{s}) = I_m$, so:

$$\Gamma_{\mathbf{w}}(\mathbf{s}, \mathbf{s}) = A(\mathbf{s})A^T(\mathbf{s})$$

- $A(\mathbf{s})$ identifies with any square-root of $\Gamma_{\mathbf{w}}(\mathbf{s}, \mathbf{s})$. Can be taken as lower-triangular (Cholesky).
- $A(\mathbf{s})$ is unknown!
  - Should we first model $A(\mathbf{s})$ to obtain $\Gamma_{\mathbf{w}}(\mathbf{s}, \mathbf{s})$?
  - Or should we model $\Gamma_{\mathbf{w}}(\mathbf{s}, \mathbf{s}')$ first and derive $A(\mathbf{s})$?
  - $A(\mathbf{s})$ is completely determined from within-site associations.

contd.

- If $A(\mathbf{s}) = A$:
  - $\mathbf{w}(\mathbf{s})$ is stationary when $\mathbf{v}(\mathbf{s})$ is.
  - $\Gamma_{\mathbf{w}}(\mathbf{s}, \mathbf{s}')$ is symmetric.
  - $\Gamma_{\mathbf{v}}(\mathbf{s}, \mathbf{s}') = \rho(\mathbf{s}, \mathbf{s}') I_m \Rightarrow \Gamma_{\mathbf{w}} = \rho(\mathbf{s}, \mathbf{s}') A A^T$
- Last specification is called intrinsic and leads to separable models:

$$\Sigma_{\mathbf{w}} = H(\phi) \otimes \Lambda;\ \Lambda = A A^T$$

Hierarchical modelling

- Let $\mathbf{y} = [\mathbf{Y}(\mathbf{s}_i)]_{i=1}^{n}$ and $\mathbf{w} = [\mathbf{W}(\mathbf{s}_i)]_{i=1}^{n}$.

Hierarchical modelling

- Let $\mathbf{y} = [\mathbf{Y}(\mathbf{s}_i)]_{i=1}^n$ and $\mathbf{w} = [\mathbf{W}(\mathbf{s}_i)]_{i=1}^n$.
- First stage:

$$\mathbf{y}|\boldsymbol{\beta}, \mathbf{w}, \Psi \sim MVN(X\boldsymbol{\beta} + \mathbf{w}, I \otimes \Psi)$$

Hierarchical modelling

- Let $\mathbf{y} = [\mathbf{Y}(\mathbf{s}_i)]_{i=1}^n$ and $\mathbf{w} = [\mathbf{W}(\mathbf{s}_i)]_{i=1}^n$.
- First stage:

$$\mathbf{y}|\boldsymbol{\beta}, \mathbf{w}, \Psi \sim MVN(X\boldsymbol{\beta} + \mathbf{w}, I \otimes \Psi)$$

- Second stage:

$$\mathbf{w}|\boldsymbol{\theta} \sim MVN(\mathbf{0}, \Sigma_{\mathbf{w}}(\Phi))$$

where $\Sigma_{\mathbf{w}}(\Phi) = [\Gamma_{\mathbf{w}}(\mathbf{s}_i, \mathbf{s}_j; \Phi)]_{i,j=1}^n$.

Hierarchical modelling

- Let $\mathbf{y} = [\mathbf{Y}(\mathbf{s}_i)]_{i=1}^n$ and $\mathbf{w} = [\mathbf{W}(\mathbf{s}_i)]_{i=1}^n$.
- First stage:

$$\mathbf{y}|\boldsymbol{\beta}, \mathbf{w}, \Psi \sim MVN(X\boldsymbol{\beta} + \mathbf{w}, I \otimes \Psi)$$

- Second stage:

$$\mathbf{w}|\boldsymbol{\theta} \sim MVN(\mathbf{0}, \Sigma_{\mathbf{w}}(\Phi))$$

where $\Sigma_{\mathbf{w}}(\Phi) = [\Gamma_{\mathbf{w}}(\mathbf{s}_i, \mathbf{s}_j; \Phi)]_{i,j=1}^n$.

- Third stage: Priors on $\Omega = (\boldsymbol{\beta}, \Psi, \Phi)$.

Hierarchical modelling

- Let $\mathbf{y} = [\mathbf{Y}(\mathbf{s}_i)]_{i=1}^{n}$ and $\mathbf{w} = [\mathbf{W}(\mathbf{s}_i)]_{i=1}^{n}$.
- First stage:

$$\mathbf{y}|\boldsymbol{\beta}, \mathbf{w}, \Psi \sim MVN(X\boldsymbol{\beta} + \mathbf{w}, I \otimes \Psi)$$

- Second stage:

$$\mathbf{w}|\boldsymbol{\theta} \sim MVN(\mathbf{0}, \Sigma_{\mathbf{w}}(\Phi))$$

where $\Sigma_{\mathbf{w}}(\Phi) = [\Gamma_{\mathbf{w}}(\mathbf{s}_i, \mathbf{s}_j; \Phi)]_{i,j=1}^{n}$.

- Third stage: Priors on $\Omega = (\boldsymbol{\beta}, \Psi, \Phi)$.
- Marginalized likelihood:

$$\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\theta}, \Psi \sim MVN(X\boldsymbol{\beta}, \Sigma_{\mathbf{w}}(\Phi) + I \otimes \Psi)$$

Bayesian Computations

- Choice: Fit as $[\mathbf{y}|\Omega] \times [\Omega]$ or as $[\mathbf{y}|\boldsymbol{\beta}, \mathbf{w}, \Psi] \times [\mathbf{w}|\Phi] \times [\Omega]$.

Bayesian Computations

- Choice: Fit as $[\mathbf{y}|\Omega] \times [\Omega]$ or as $[\mathbf{y}|\boldsymbol{\beta}, \mathbf{w}, \Psi] \times [\mathbf{w}|\Phi] \times [\Omega]$.
- Conditional model:
  - Conjugate distributions are available for Ψ and other variance parameters. Easy to program.

Bayesian Computations

- Choice: Fit as $[\mathbf{y}|\Omega] \times [\Omega]$ or as $[\mathbf{y}|\boldsymbol{\beta}, \mathbf{w}, \Psi] \times [\mathbf{w}|\Phi] \times [\Omega]$.
- Conditional model:
  - Conjugate distributions are available for Ψ and other variance parameters. Easy to program.
- Marginalized model:
  - need Metropolis or Slice sampling for most variance-covariance parameters. Harder to program.
  - But reduced parameter space (no **w**'s) results in faster convergence
  - $\Sigma_{\mathbf{w}}(\Phi) + I \otimes \Psi$ is more stable than $\Sigma_{\mathbf{w}}(\Phi)$.

Bayesian Computations

- Choice: Fit as $[\mathbf{y}|\Omega] \times [\Omega]$ or as $[\mathbf{y}|\boldsymbol{\beta}, \mathbf{w}, \Psi] \times [\mathbf{w}|\Phi] \times [\Omega]$.
- Conditional model:
  - Conjugate distributions are available for Ψ and other variance parameters. Easy to program.
- Marginalized model:
  - need Metropolis or Slice sampling for most variance-covariance parameters. Harder to program.
  - But reduced parameter space (no $\mathbf{w}$'s) results in faster convergence
  - $\Sigma_{\mathbf{w}}(\Phi) + I \otimes \Psi$ is more stable than $\Sigma_{\mathbf{w}}(\Phi)$.
- But what about $\Sigma_{\mathbf{w}}^{-1}(\Phi)$?? Matrix inversion is EXPENSIVE $O(n^3)$.

Recovering the w's?

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

Recovering 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:

Recovering 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 $\Omega^{(1)}, \ldots, \Omega^{(G)} \sim [\Omega|\mathbf{y}, X]$

Recovering 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 $\mathbf{w}^{(g)} \sim [\mathbf{w}|\Omega^{(g)}, \mathbf{y}, X]$

Recovering 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.

Spatial Prediction

- 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:

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

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

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

The "Big N" problem

- The multivariate spatial regression model:

$$\mathbf{Y}(\mathbf{s}) = \mathbf{X}^T(\mathbf{s})\boldsymbol{\beta} + \mathbf{w}(\mathbf{s}) + \boldsymbol{\epsilon}(\mathbf{s})$$

  with m spatial regressions at each $\mathbf{s}$.

- Fitting a fully Bayes model over sites $\mathcal{S} = \{\mathbf{s}_1, \ldots, \mathbf{s}_n\}$ involves matrix decompositions and determinants for the $mn \times mn$ $\Sigma_{\mathbf{w}}(\Phi)$. When Φ is estimated this must be done in *each* MCMC iteration.
- This can be prohibitive when n is large and is known as the Big-N problem in geostatistics.
- Approach: Replace $\mathbf{w}(\mathbf{s})$ by a process $\tilde{\mathbf{w}}(\mathbf{s})$ that will somehow project the model into a smaller subspace.

The “Kriging” Equation

- Consider “knots” $\mathcal{S}^* = \{\mathbf{s}_1^*, \ldots, \mathbf{s}_p^*\}$ with $p << n$.
- Predict $\mathbf{w}(\mathbf{s})$ based upon $\mathcal{S}^*$ at $\mathbf{s}_0$:

$$\begin{aligned}\tilde{\mathbf{w}}(\mathbf{s}_0) &= E[\mathbf{w}(\mathbf{s}_0)|\mathbf{w}^*] \\ &= Cov(\mathbf{w}(\mathbf{s}_0), \mathbf{w}^*)Var(\mathbf{w}^*)^{-1}\mathbf{w}^*;\end{aligned}$$

$$\begin{aligned}\mathbf{w}^* &= [\mathbf{w}(\mathbf{s}_i^*)]_{i=1}^p, \\ Cov(\mathbf{w}(\mathbf{s}_0), \mathbf{w}^*) &= [\Gamma_{\mathbf{w}}(\mathbf{s}_0, \mathbf{s}_1^*; \Phi), \ldots, \Gamma_{\mathbf{w}}(\mathbf{s}_0, \mathbf{s}_p^*; \Phi)], \\ Var(\mathbf{w}^*) &= [\Gamma_{\mathbf{w}}(\mathbf{s}_i^*, \mathbf{s}_j^*; \Phi)]_{i,j=1}^p.\end{aligned}$$

The Predictive Process

- For generic **s**, $\tilde{\mathbf{w}}(\mathbf{s}) \sim MVGP(\mathbf{0}, \Gamma_{\tilde{\mathbf{w}}}(\cdot,\cdot))$

$$\Gamma_{\tilde{\mathbf{w}}}(\mathbf{s}, \mathbf{s}') = Cov(\mathbf{w}(\mathbf{s}), \mathbf{w}^*) Var(\mathbf{w}^*)^{-1} Cov^T(\mathbf{w}(\mathbf{s}), \mathbf{w}^*).$$

- We can write $\tilde{\mathbf{w}}(\mathbf{s}) = Z(\mathbf{s})\mathbf{w}^*$
  - $Z(\mathbf{s}) = Cov(\mathbf{w}(\mathbf{s}), \mathbf{w}^*) Var(\mathbf{w}^*)^{-1}$ is an $m \times mp$ matrix.
- Process realizations over $\mathcal{S} = \mathbf{s}_1, \ldots, \mathbf{s}_n$:

$$\begin{aligned} \tilde{\mathbf{w}} &= [\tilde{\mathbf{w}}(\mathbf{s}_i)]_{i=1}^n = \mathbf{Z}\mathbf{w}^*; \\ &\sim MVN(\mathbf{0}, \mathbf{Z}\Sigma_{\mathbf{w}^*}^{-1}\mathbf{Z}^T) \end{aligned}$$

  with $\mathbf{Z} = [Z(\mathbf{s}_i)]_{i=1}^n$ being $mn \times mp$.
- Realization is *singular* as soon as n exceeds m.

Reducing the dimension

- We fit the reduced model:

$$\begin{aligned} \mathbf{Y}(\mathbf{s}) &= \mathbf{X}^T(\mathbf{s})\boldsymbol{\beta} + \tilde{\mathbf{w}}(\mathbf{s}) + \boldsymbol{\epsilon}(\mathbf{s}) \\ &= \mathbf{X}^T(\mathbf{s})\boldsymbol{\beta} + Z(\mathbf{s})\mathbf{w}^* + \boldsymbol{\epsilon}(\mathbf{s}). \end{aligned}$$

- Computations involve $\Sigma_{\mathbf{w}^*}$ ($mp \times mp$) instead of $\Sigma_{\mathbf{w}}$ ($np \times np$). Can make a huge difference with $p << n$.
- Concerns:
  - How many knots will suffice?
  - How do we choose the knots?
  - Impact of knot selection in practical data analysis?

Attractions

- Greater flexibility with knots compared to other methods. Does not have to be on a regular grid as for spectral methods and other likelihood approximation methods.

Attractions

- Greater flexibility with knots compared to other methods. Does not have to be on a regular grid as for spectral methods and other likelihood approximation methods.
- Derives from the original process. No need to specify kernels or other covariance functions.

Attractions

- Greater flexibility with knots compared to other methods. Does not have to be on a regular grid as for spectral methods and other likelihood approximation methods.
- Derives from the original process. No need to specify kernels or other covariance functions.
- Adapts seamlessly to spatiotemporal and other settings – helps being a “model” rather than an approximation!

Attractions

- Greater flexibility with knots compared to other methods. Does not have to be on a regular grid as for spectral methods and other likelihood approximation methods.
- Derives from the original process. No need to specify kernels or other covariance functions.
- Adapts seamlessly to spatiotemporal and other settings – helps being a "model" rather than an approximation!
- Given $\mathcal{S}^*$ you *cannot* do better in a (reverse entropy) Kullback-Leibler paradigm with any process of the form $H(\mathbf{s})\mathbf{w}^*$.

Attractions

- Greater flexibility with knots compared to other methods. Does not have to be on a regular grid as for spectral methods and other likelihood approximation methods.
- Derives from the original process. No need to specify kernels or other covariance functions.
- Adapts seamlessly to spatiotemporal and other settings – helps being a “model” rather than an approximation!
- Given $\mathcal{S}^*$ you *cannot* do better in a (reverse entropy) Kullback-Leibler paradigm with any process of the form $H(\mathbf{s})\mathbf{w}^*$.
- When $\mathcal{S}^* = \mathcal{S}$ we recover the original model.