Spatial Models: a brief integration

Douglas Nychka,
National Center for Atmospheric Research

Supported by the National Science Foundation

Outline

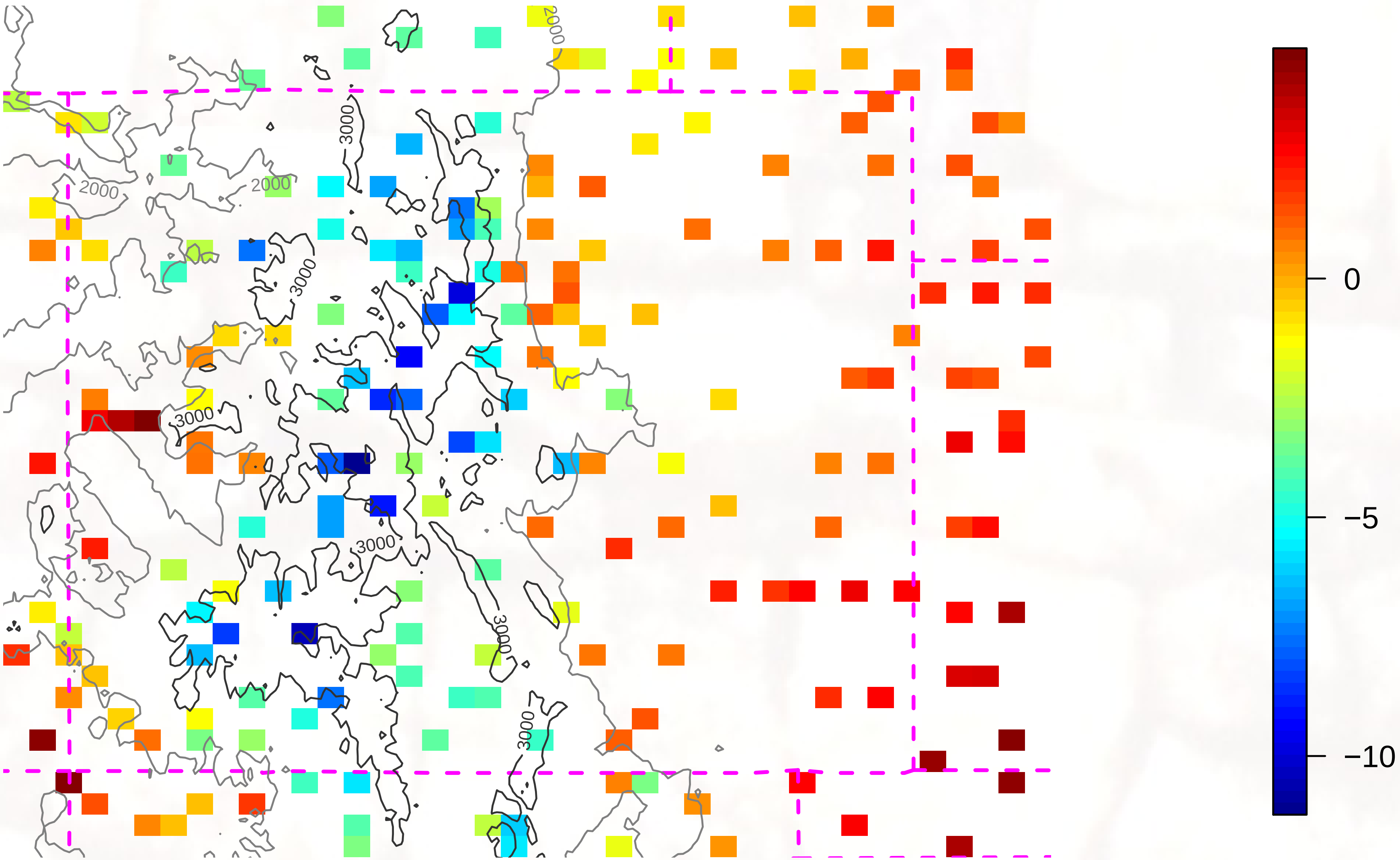
- Colorado springtime temperatures and a spatial analysis
- Easy to use functions from the `fields` R package

Interperting spatial data

- The additive model
- A Bayesian hierarchical model.
- Full conditionals and approximations

Joint distributions, conditional distributions, full conditional distribution.

Colorado MAM average tmin



Goals:

1. Prediction: Determine the climate at locations where there are not stations.

2. Uncertainty: Quantify the error in the predictions.

3. Summarize the spatial/temporal structure: One tool for comparing observations to models and models to models for physical insight

The additive statistical model:

Observation = climate covariates

+ Smooth function (location) + error

Given n pairs of observations (x_i, y_i), $i = 1, \ldots, n$

$$y_i = \boldsymbol{Z}_i \boldsymbol{\beta} + g(x_i) + \epsilon_i$$

- ϵ_i's are random errors
- z_i climate covariates (e.g. elevation, latitude)
- parameters ($\boldsymbol{\beta}$)
- g is an unknown smooth function.

Need more assumptions to draw inferences about g and β

A hierarchical model

A sequence of at least three conditional models: (DATA, PROCESS, PRIOR)

DATA LEVEL: observations given the process

$$\boldsymbol{y}_i = \boldsymbol{Z}_i\boldsymbol{\beta} + g(x_i) + e_i$$

$$\boldsymbol{y} = \boldsymbol{Z}\boldsymbol{\beta} + \boldsymbol{g} + e_i$$

Random: $\boldsymbol{e}$ "measurement" error $N(0, \sigma^2)$
Held fixed: $\boldsymbol{\beta}$ and $\boldsymbol{g} = (g(x_1), ..., g(x_n))$

Conditional distribution of $\boldsymbol{y}$ given other components is $N(\boldsymbol{Z}\boldsymbol{\beta} + \boldsymbol{g}, \sigma^2 I)$
In Gelfand notation: $[\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{g}, \sigma^2]$

A hierarchical model continued

PROCESS LEVEL: process given priors

g(x) is a mean Gaussian Process with a specified covariance function.

$$E(g(x)) = 0$$

$$COV(g(\boldsymbol{x}_1), g(\boldsymbol{x}_2)) = \rho k_\theta(\boldsymbol{x}_1, \boldsymbol{x}_2)$$

e.g $COV(g(x_1), g(x_2) = \rho e^{\text{distance}(x_1,x_2)/\theta}$

NOTE: It is natural to separate g into the points where there is direct information ($\boldsymbol{g} = (g(x_1), ..., g(x_n))$ and the other points.

In the final analysis we will first find the posterior distribution of the $\boldsymbol{g}$ and then find the conditional distribution of $g(x)$ given $\boldsymbol{g}$

- In Gelfand notation: $[\boldsymbol{g}|\rho, \theta]$

A hierarchical model continued

PRIOR LEVEL: physical and statistical parameters Joint distribution for the parameters some standard choices • $\boldsymbol{\beta} \sim$ uniform $[0, \infty]$

- $\sigma^2 \sim$ Inverse Gamma (a_1, b_1)
- $\rho \sim$ Inverse Gamma (a_2, b_2)
- $\theta \sim$ uniform

In Gelfand notation: $[\boldsymbol{\beta}, \rho, \sigma^2, \theta]$ usually $[\boldsymbol{\beta}][\rho][\sigma^2][\theta]$

The joint density and full conditionals:

Schematically we have $[y|\beta, \boldsymbol{g}, \sigma^2], [\boldsymbol{g}|\rho, \theta], [\boldsymbol{\beta}, \sigma^2, \rho, \theta]$ Each is a probability density function.

The joint density is the product:

$$[y|\beta, \boldsymbol{g}, \sigma^2][\boldsymbol{g}|\rho, \theta][\boldsymbol{\beta}, \sigma^2, \rho, \theta]$$

- Although this has many parts one can usually write down this expression in closed form

- The posterior is proportional to this expression

- The full conditionals can be identified from this product. For a given component all collect terms that have that component. What ever you get is proportional to the full conditional density.

- Maximum likelihood is maximizing just part

$$[y|\beta, \boldsymbol{g}, \sigma^2][\boldsymbol{g}|\rho, \theta]$$

or assuming all the priors are uniform

Recalling the Gibbs sampler

If we had a density $f(u, v, w)$ then to generate a sample from (U, V, W)

- $U \sim f(u|V, W) = f(u, V, W)/f(V, W)$
- $V \sim f(v|U, W) = f(v, U, W)/f(U, W)$
- $W \sim f(w|U, V) = f(w, U, V)/f(U, V)$

Keep iterating among these. As this converges will be a multivariate sample.

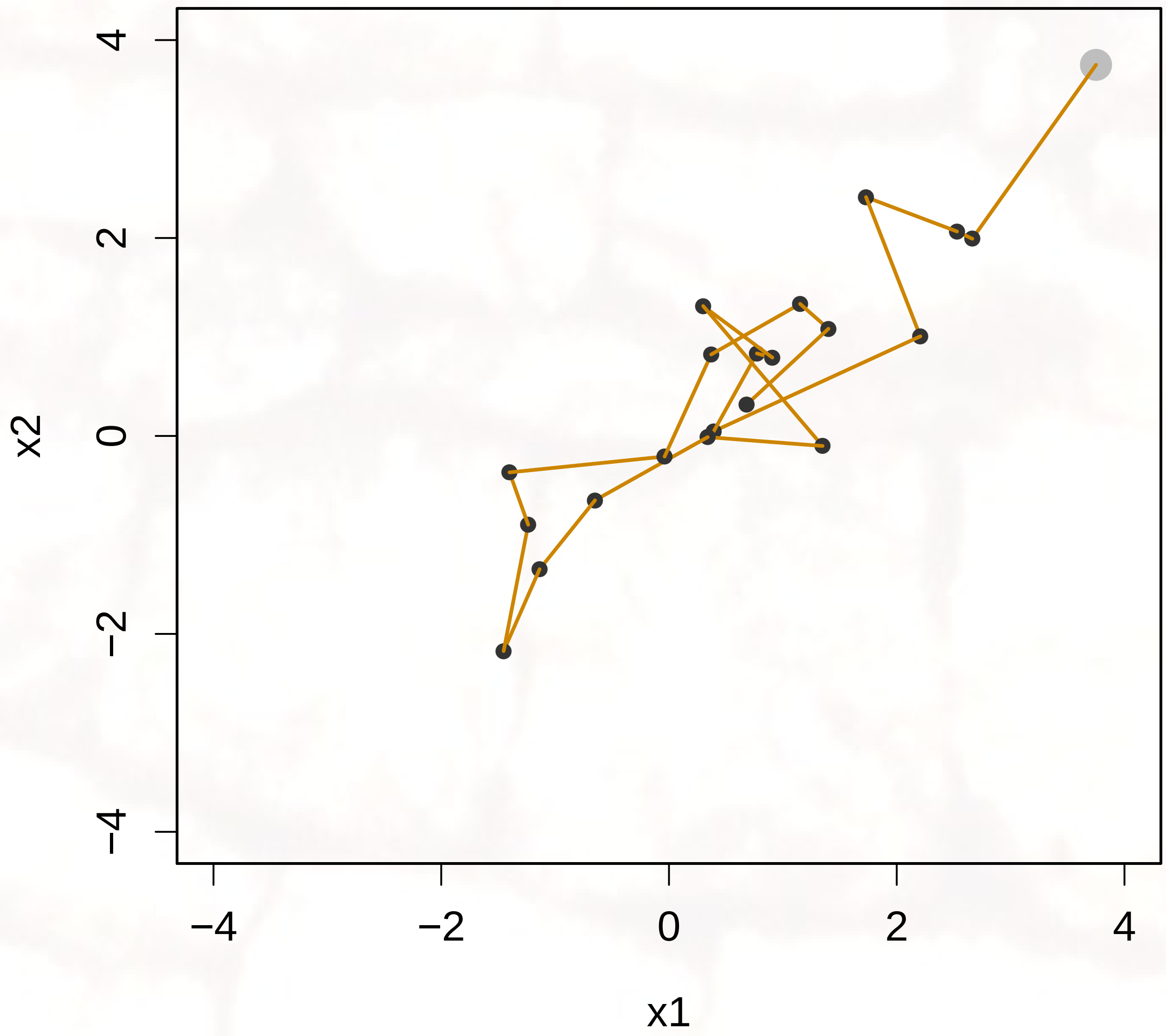
- the U, V, W are parameters or model components

- can shortcut computation by substituting some fixed estimate in place of sampling and skip a step.

e.g. just use $\widehat{U}$ at every iteration.

Sampling bivariate normal

Correlation of .8 and 20 iterations



Finding the full conditionals

$\boldsymbol{\beta}$

Only depends on the first term and prior (which we assume is uniform

$$[\boldsymbol{\beta}|\text{everything else}] \sim e^{-\frac{(\boldsymbol{y}-\boldsymbol{Z}\boldsymbol{\beta}-\boldsymbol{g})^T(\boldsymbol{y}-\boldsymbol{Z}\boldsymbol{\beta})}{2\sigma^2}}$$

(Remember only $\boldsymbol{\beta}$ varies in this expression)

After some algebra ...

$$[\boldsymbol{\beta}|\text{everything else}] = MN((Z^TZ)^{-1}(\boldsymbol{y}-\boldsymbol{g}),(\sigma^2Z^TZ)^{-1})$$

- Notice this only depends on $\boldsymbol{g}$ and σ
- Standard to simulate from this distribution.

Finding the full conditionals (cont.)

$\boldsymbol{g}$

Only depends on the first term and second terms

$$[\boldsymbol{g}|\text{everything else}] \sim e^{-\frac{(\boldsymbol{y}-\boldsymbol{Z\beta}-\boldsymbol{g})^T(\boldsymbol{y}-\boldsymbol{Z\beta})}{2\sigma^2}} e^{-\frac{-(\boldsymbol{g})^T\Omega_\theta^{-1}(\boldsymbol{g})}{2\rho}}$$

(Remember only $\boldsymbol{g}$ varies in this expression)

After some algebra ...

$$[\boldsymbol{g}|\text{everything else}] = MN(\boldsymbol{A}(\boldsymbol{y} - X\boldsymbol{\beta}), \rho\Omega(I - \boldsymbol{A}))$$

$\boldsymbol{A} = \rho\Omega(\rho\Omega + \sigma^2 I)^{-1}$

- This only depends on all the other components because $\boldsymbol{g}$ is also in the process level.
- "Standard to simulate from this distribution."

A more exotic full conditional

ρ

$$[\rho|\text{everything else}] \sim e^{-\frac{-(\boldsymbol{g})^T \Omega_\theta^{-1}(\boldsymbol{g})}{2\rho}} \rho^{-n/2}[\rho]$$

- If the prior is inverse gamma that this density is again proportional to an inverse gamma.

- "Standard to simulate from this distribution."

Where there are problems

- Typically θ does not have a simple distribution and one has to use Metropolis-Hasting to sample.

- Ω_θ can be a large matrix
evaluations of $(\rho\Omega + \sigma^2 I)^{-1}$ and $|\rho\Omega + \sigma^2 I|$ can be prohibitive for large data sets.

What is Kriging?

On the good side

It is simple to work the full conditional for $\boldsymbol{\beta}$ and $\boldsymbol{g}$ together.

- For fixed values of the covariance and variance parameters Kriging = sampling the full conditional

- More efficient that sampling individually

Where one cheats

The spatial model can be collapsed to a linear model:

$$\boldsymbol{y} \sim MN(X\beta, \rho\Omega_\theta + \sigma^2 I)$$

or

$$\boldsymbol{y} \sim MN(X\beta, \rho(\Omega_\theta + \lambda I))$$

$\lambda = \sigma^2/\rho$

- Find all the parameters by maximum likelihood.
- Need to find some of these numerically. (e.g. θ and λ)

Summary

- Bayesian spatial models naturally separate into three levels.
- The basic parameters can all be sampled exactly as full conditionals except for the range.
- Kriging is the sampling of the full conditionals of β and g with the covariance parameters fixed.
- Can estimate covariance parameters using maximum likelihood.

Thank you