Spatial Statistics: Beyond a Textbook

Doug Nychka, Soutir Bandyopadhyay, William Kleiber
Colorado School of Mines

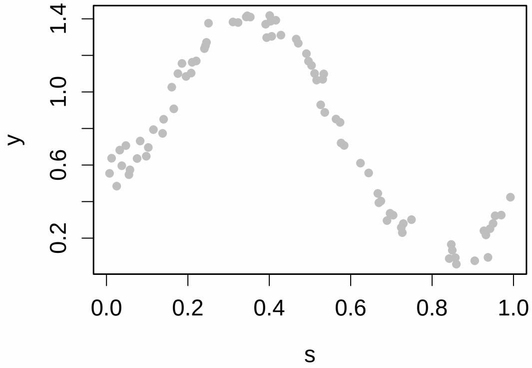
Outline

- Gaussian process models
- A 1D example
- The precision matrix
- Approximation for large data - basis functions
- Change of support
- Non-Gaussian observations

WaldoTobler:
"everything is related to everything else, but near things are more related than distant things."

. . . but the distant things become less important given the near things . . .

Some data



Goal: Inference for the maximum – where is it? How high? What is the uncertainty?

Gaussian Process model

$g(s)$ Gaussian process

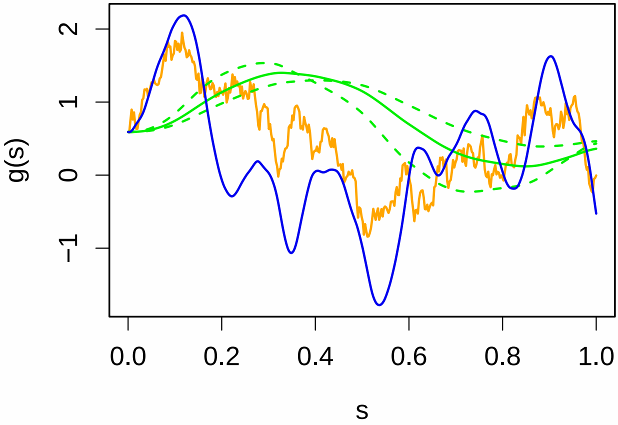
- $E(g(s)) = 0$
- . $Cov(g(s), g(s') = k(s, s')$

Some examples of stationary $k(.,.)$

$d = \|s - s'\|/\alpha$ Distance scaled by a parameter.

- Exponential: $k(s, s') = e^{-d}$
- Gaussian: $k(s, s') = e^{-d^2}$
- Matern ($\nu = 2.5$): $k(s, s') = (1 + d + (1/3)d^2)e^{-d}$
- Wendland: $k(s, s') = \begin{cases} (1-d)^6(35d^2 + 18d + 3)/3 & d \leq 0 \\ 0 & d > 0 \end{cases}$

Gaussian Process realizations



Exponential, Matern, Wendland

The data

$$Y_i = g(s_i) + e_i$$

$\{e_i\}$ Independent $N(0, \tau^2)$
Assume that $VAR(g(s_i)) = \sigma^2$

The Big Three:

- Correlation range, α
- Variance of process, σ^2
- Variance of measurement error, τ^2

What we do

Goal: estimate/infer g from $\mathbf{z}$ (but we will need the Big Three to do it).

The program:

- Estimate the big three by maximum likelihood (or Bayes).
- Sample from the conditional distribution of g given the observations.

Kriging – finding g

Recall

$$\mathbf{z} = \mathbf{g} + \mathbf{e}$$

$$\mathbf{g} = \begin{pmatrix} g(s_1) \\ g(s_2) \\ \vdots \\ g(s_n) \end{pmatrix}$$

Assumptions

- $[\mathbf{z}|\mathbf{g}]$ is $MN(\mathbf{g}, \tau^2 I)$
- $[\mathbf{g}]$ is $MN(0, \sigma^2 K)$
- $[\mathbf{z}|\mathbf{g}][\mathbf{g}] = [\mathbf{z}, \mathbf{g}]$ is jointly multivariate normal.

Note: we can fill in K for any set of locations because we have the covariance function.

More on Kriging

Estimate is the conditional expectation of $g(s)$ given $\mathbf{z}$.

$$\hat{g}(s) = E[g(s)|\mathbf{z}] = \sum_{i=1}^{n} k(s, s_i)\mathbf{c}_i$$

with $\mathbf{c}$ is the solution to the linear system

$$COV(\mathbf{z}, \mathbf{z})\mathbf{c} = \mathbf{z}$$

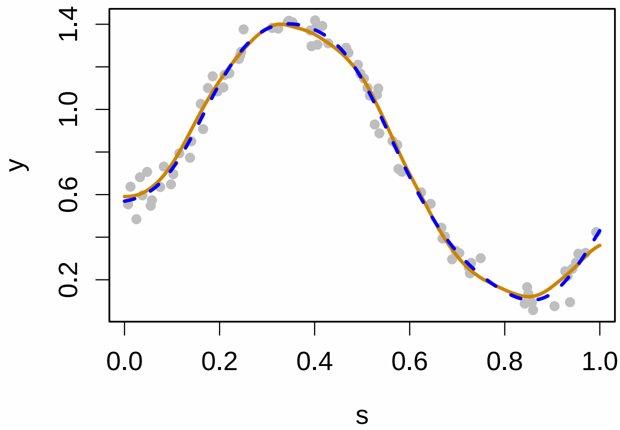
$$(\sigma^2 K(\alpha) + \tau^2 I)\mathbf{c} = \mathbf{z}$$

Prediction at the data is a smoother:

$$\hat{h} = K(\alpha)(K(\alpha) + \lambda I)^{-1}\mathbf{z}$$

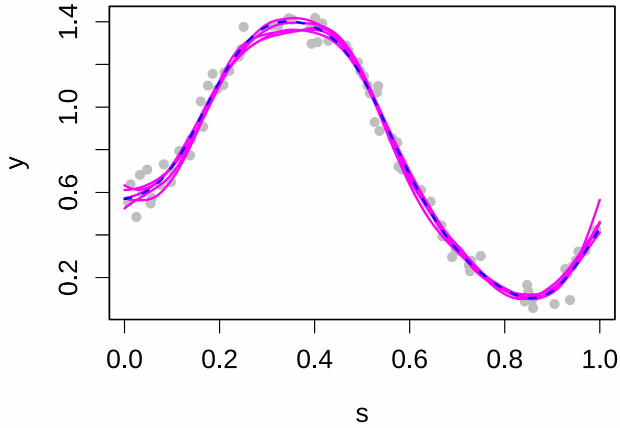
$\lambda = \tau^2/\sigma^2$

Estimate



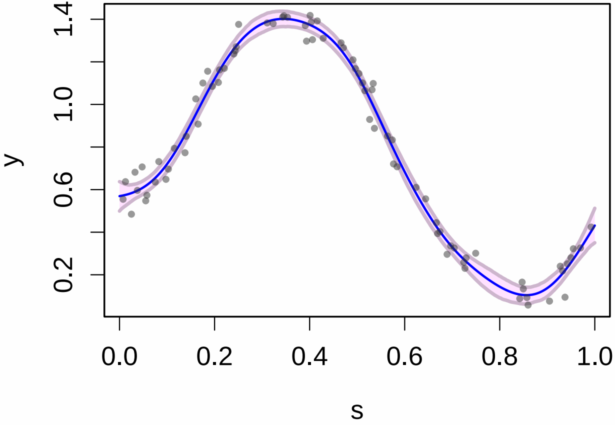
true curve

5 Draws from $[\mathbf{g}|\mathbf{z}]$



$\hat{g}$

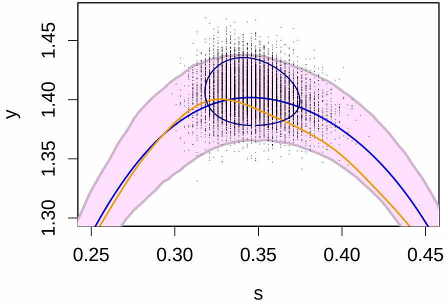
95% envelope



Inference for the maximum

≈ 95% confidence region for maximum X and Y

Points are 1000 realizations of the maxima from conditional samples



True curve, Estimate

We are 95% confident the maximum of g and its location are in this region.

What about the big three?

Integrate out the latent Gaussian process.

$$\int [\mathbf{z}, \mathbf{g} | \sigma^2, \tau^2, \alpha] d\mathbf{g} = [\mathbf{z} | \sigma^2, \tau^2, \alpha] \sim MN(0, \sigma^2 K(\alpha) + \tau^2 I)$$

log likelihood is

$$-\frac{1}{2}\mathbf{z}^T(\sigma^2 K(\alpha) + \tau^2 I)^{-1}\mathbf{z} - \frac{1}{2} log(|\sigma^2 K(\alpha) + \tau^2 I|) + constant$$

Can concentrate over $\lambda = \tau^2/\sigma^2$ and α for a 2D optimization.

For a few thousand locations (in `fields` package)

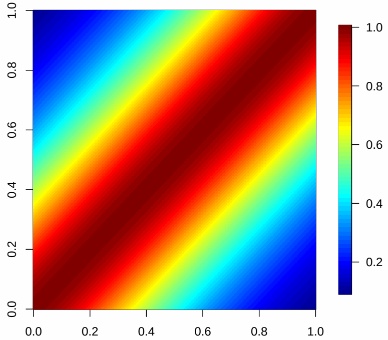
```
obj <- spatialProcess(s,z, smoothness=2.5)
```

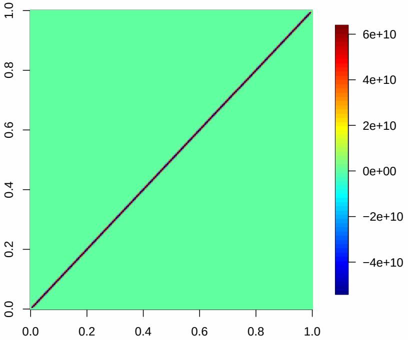
Exact linear algebra increases as $O(n^3)$ with n the number of locations.
$n = 3500$, takes 30 seconds $\rightarrow$ $n = 8 * 3500$, takes $\approx$ 4 hours

More about covariance matrices

Covariance used in the example

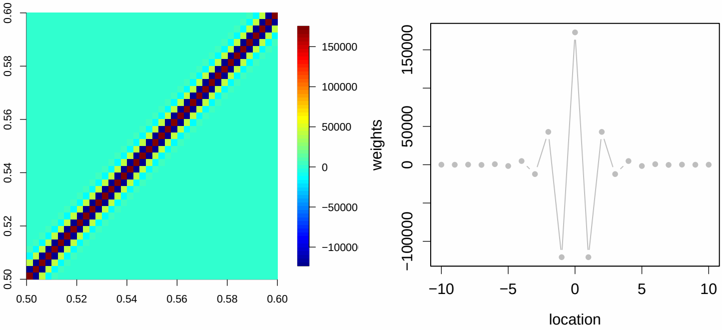
Covariance matrix (Matern $\nu = 2.5$) and its inverse





Spatial AR

Symmetric Inverse square root of covariance matrix



If $cov(\mathbf{g}) = \Sigma$ then $\Sigma^{-1/2}\mathbf{g}$ has an identity covariance matrix.
I.e.

$$cov(\Sigma^{-1/2}\mathbf{g}) = \Sigma^{-1/2}(\Sigma)\Sigma^{-1/2} = I$$

- $\Sigma^{-1/2}$ a recipe for decorrelation.

Fixed Rank Kriging (FRK): Basic Idea

- Model:

$$\mathbf{z}_i = \mathbf{X}_i \beta + \mathbf{g}(s_i) + \epsilon_\mathbf{i}$$

- The spatial process $g(\cdot)$ is approximated by a linear combination of L fixed basis functions and random coefficients.

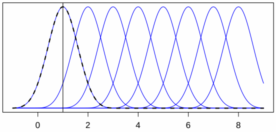
$$g(s) = \psi_1(s)c_1 + \psi_2(s)c_2 + \ldots + \psi_L(s)c_L$$

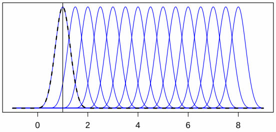
- Assume that **c** is multivariate normal, mean zero and covariance matrix, Q^{-1}.
- Sherman-Morrison-Woodbury formula reduces the computations to linear algebra in Q ($L \times L$).
  If Q is dense then L small is helpful.
  If Q is sparse then L can be larger than the number of observations.
- $Q = B^T B$ then the rows of B are a recipe for transforming the coefficients to independent $N(0, 1)$

More on basis functions and coefficients

$$g(s) = \psi_1(s)c_1 + \psi_2(s)c_2 + \ldots + \psi_L(s)c_L$$

In practice, $\{\psi_j\}$ have different resolutions and are zero beyond a fixed range.





Model for the coefficients, j^{th} row of B:

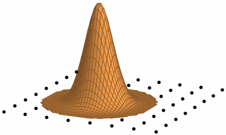
$$\ldots \quad 0 \quad 0 \quad \text{-1} \quad a_j \quad \text{-1} \quad 0 \quad 0 \quad \ldots$$

The recipe for decorrelation:

$$a_j \mathbf{c}_j - \mathbf{c}_{j-1} - \mathbf{c}_{j+1} \sim N(0,1)$$

Note $a_j > 2$ for stationarity

LatticeKrig 2 D Implementation

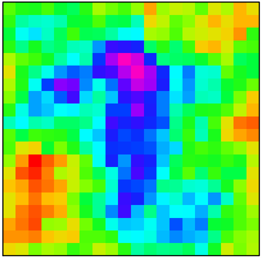


A 2-d basis function

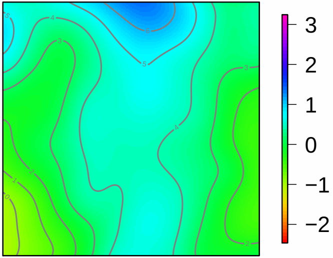
.	.	.	.	.
.	.	-1	.	.
.	-1	a_j	-1	.
.	.	-1	.	.
.	.	.	.	.

Need $a_j > 4$

16^2 Coefficients



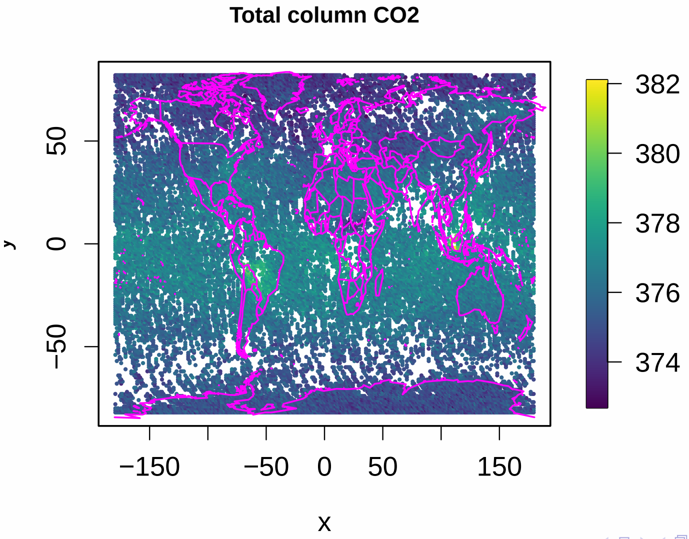
Expanding with basis functions



$$\mathbf{c}_j \quad \rightarrow \quad \sum \psi_j(x)\mathbf{c}_j$$

LatticeKrig Demo in R

CO2 data set ≈ 27K observations in PPM.



Change of support

Observations of area averages

$$g(s) = \psi_1(s)c_1 + \ldots + \psi_L(s)c_L$$
$$\int_A g(s)ds = \left(\int_A \psi_1(s)ds\right) c_1 + \ldots + \left(\int_A \psi_L(s)ds\right) c_L$$

- Observations:

z_i average over region A_i

$$\mathbf{z} = W\mathbf{c} + \mathbf{e}$$

$W_{i,j}$ integral of region i over basis function, j.

- Spatial process:

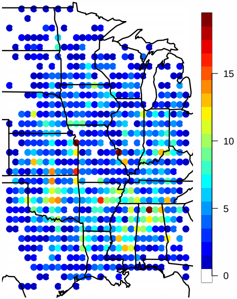
g(s) follows *same* fixed rank Kriging model, $\mathbf{c} \sim MN(0, Q^{-1})$

Non Gaussian observations

Poission counts with a spatial intensity

- $\mathbf{z}_i \sim Poisson(\gamma_i)$
- $log(\gamma(\mathbf{s})) \sim$ Gaussian process with σ^2 and α and mean $X\boldsymbol{\beta}$.
- Prior on σ^2, α and $\boldsymbol{\beta}$

US Tornado counts 1950-2020 (F level 3 or higher)



X = lon, lat, and elevation

As a Hierarchical Model

Observational model: $[\mathbf{z}_i|\mathbf{g}, \beta, \alpha, \sigma^2] \sim Poisson(e^{X_i\beta+g(s_i)})$

Process: $[\mathbf{g}|\beta, \alpha, \sigma^2] \sim N(X\beta, \sigma^2 K(\alpha))$

Priors: $[\beta, \alpha, \sigma^2]$

Key technical step is to "integrate out" $\mathbf{g}$

$$[\mathbf{z}_i|\beta, \alpha, \sigma^2] = \int [\mathbf{z}_i|\mathbf{g}, \beta, \alpha, \sigma^2] d\mathbf{g}$$

- Needed to find MLEs or Bayesian inference.

- Use a Laplace approximation for this step – this is the basis of the INLA framework.

Some details for $g(s)$

Poisson case:

$$[\mathbf{g}, \alpha, \sigma^2, \boldsymbol{\beta} | \mathbf{z}] \propto [\mathbf{z} | \boldsymbol{\beta}, \mathbf{g}][\mathbf{g} | \alpha, \sigma^2]$$

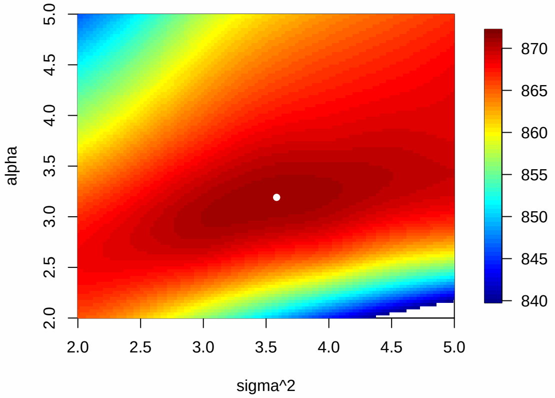
$$\propto \left[\prod_{i=1}^{n} (\gamma_i^{z_i}) e^{-\gamma_i} \right] \left[\frac{e^{-(\mathbf{g}-X\boldsymbol{\beta})^T (\sigma^2 K(\alpha))^{-1} (\mathbf{g}-X\boldsymbol{\beta})/2}}{|\sigma^2 K(\alpha)|^{1/2}} \right]$$

and $\gamma_i = e^{x_i^T \beta + g_i}$ or $log(\gamma) = X\beta + \mathbf{g}$

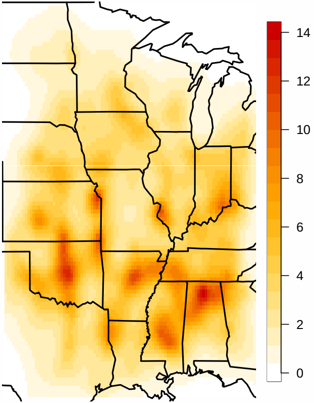
- For fixed parameters can maximize over **g** – this is a crucial computation giving – $\hat{\mathbf{g}}$
- Note that $\hat{\mathbf{g}}$ depends on the data and other parameters.
- Expand the log of the likelihood in a Taylors series about $\hat{\mathbf{g}}$ giving a multivariate normal approximation.

Intensity of tornados

Laplace approximation to log likelihood surface for σ^2 and α



Intensity of tornados



Predicted intensity surface – is this biased towards population centers?

Conclusions

- Spatial analysis tied to multivariate normal that describes dependence
- Conditional sampling from the spatial model is a useful for inference.
- Dependence can be introduced into a basis function/ coefficient way to handle large data and complicated observation types.

Methods in the `fields` and `LatticeKrig` R packages.

Thank you!