A convenient representation of a Gaussian process is given by *process convolutions*. Consider a kernel function $k(s)$ and a white noise process $w(s)$, with $E(w(s)) = 0, \mathrm{var}(w(s)) = \sigma^2$ and $\mathrm{cov}(w(s), w(s')) = 0$. Then define

$$x(s) = \int_S k(s-u)w(u)du$$

More formally we define the process as

$$x(s) = \int_S k(s-u)dB(u), \quad \text{where} \quad \int_A dB(u) = B(A) \sim N(0, \sigma^2|A|)$$

and $\mathrm{cov}(B(A), B(C)) = \sigma^2|A \cap C|$, which corresponds to a Brownian motion.

We have that $E(x(s)) = 0$

$$\text{var}(x(s)) = \sigma^2 \int k^2(s-u)du$$

and

$$\text{cov}(x(s), x(s')) = \sigma^2 \int k(s-u)k(s'-u)du = \int k(t)k(t-d)dt$$

where $d = s - s'$. Thus the process $x(s)$ is stationary.

We have that $E(x(s)) = 0$

$$\mathrm{var}(x(s)) = \sigma^2 \int k^2(s-u)du$$

and

$$\mathrm{cov}(x(s), x(s')) = \sigma^2 \int k(s-u)k(s'-u)du = \int k(t)k(t-d)dt$$

where $d = s - s'$. Thus the process $x(s)$ is stationary.

The Fourier transform of the covariance of $x(s)$ is the square of the Fourier transform of the k. Thus, for a given covariance C, the corresponding kernel is the inverse-transform of the root of the spectrum of C:

$$k = IFT(\sqrt{FT(C)})$$

Process Convolutions Examples

- The Gaussian correlation corresponds to a Gaussian kernel.
- The Matèrn correlation corresponds to the kernel

$$k(s) = \left(\frac{s}{\phi}\right)^{\nu-1/2} K_{\nu-1/2}\left(\frac{s}{\phi}\right) \quad \phi > 0, \nu > 1$$

- The exponential correlation corresponds to a spike.

The covariance has twice the number of derivatives that the kernel has.

Process Convolutions Examples

- The Gaussian correlation corresponds to a Gaussian kernel.
- The Matèrn correlation corresponds to the kernel

$$k(s) = \left(\frac{s}{\phi}\right)^{\nu-1/2} K_{\nu-1/2}\left(\frac{s}{\phi}\right) \quad \phi > 0, \nu > 1$$

- The exponential correlation corresponds to a spike.

The covariance has twice the number of derivatives that the kernel has.

A kernel offering varying degrees of smoothing and compact support is the Bezier kernel

$$k(s) = \begin{cases} \left(1 - ||s||^2\right)^{\nu} & \text{if } ||s|| < 1 \\ 0 & \text{otherwise.} \end{cases} \quad \nu > 0$$

In practice we consider a grid of points in S to approximate the kernel convolution. So, for $u_1, \ldots, u_p$ we have that

$$x(s) = \sum_{j=1}^{p} k(s - u_j) w(u_j)$$

We observe hat

$$\text{var}(x(s)) = \sigma^2 \sum_{j=1}^{p} k^2(s - u_j)$$

and

$$\text{cov}(x(s), x(s')) = \sigma^2 \sum_{j=1}^{p} k(s - u_j) k(s' - u_j)$$

which imply that the resulting covariance is NOT stationary.

The continuous version of the kernel convolution can be used to obtain non-stationary Gaussian processes by:

- Kernel dependent on location-varying parameters

$$x(s) = \int_S k(s-u; \theta(u)) w(u) du$$

- Convolving process with covariance function dependent on location-varying parameters.

$$x(s) = \int_S k(s-u) w_{\theta(u)}(s) du$$

Fitting the Model

Given a set of observations $y_1, \ldots, y_m$ at locations $s_1, \ldots, s_m$ we fit the model

$$y_i = \mu(s) + \sum_{j=1}^{p} k(s - u_j; \psi) w_j + \varepsilon_i, \;\; \varepsilon_i \sim N(0, \tau^2)$$

with priors $w_j \sim N(0, \sigma^2), p(\sigma^2), p(\psi)$ and $p(\tau^2)$.

Given a set of observations $y_1, \ldots, y_m$ at locations $s_1, \ldots, s_m$ we fit the model

$$y_i = \mu(s) + \sum_{j=1}^{p} k(s - u_j; \psi) w_j + \varepsilon_i, \quad \varepsilon_i \sim N(0, \tau^2)$$

with priors $w_j \sim N(0, \sigma^2), p(\sigma^2), p(\psi)$ and $p(\tau^2)$.

This is a hierarchical linear model where the design matrix is defined by the kernel.

$$Y = \boldsymbol{\mu} + K(\psi)\boldsymbol{w} + \boldsymbol{\varepsilon}$$

The dimension of $\boldsymbol{w}$ is p, which corresponds to the size of the grid that is used for $w(s)$. As p is much smaller than m, this results is an important reduction in the dimension of the problem for computational purposes. Usually $\mu(s) = z'(s)\beta$.

Non-Gaussian Processes

In many situations we need to consider spatial processes that are not Gaussian. Here are several options:

- Generalized linear models. Use a Gaussian process to model the mean function of the observations transformed using the link function.
- Non-linear transformations and "clipping". This is useful, for example, for binary data.
- Process convolutions for non-Gaussian latent processes. $w(u_i)$ can be taken as realizations of distributions other than Gaussian. For example, a positive distribution will provide a positive-valued random field.

Non-Gaussian Process Convolutions

Elaborating on the previous slide

$$x(s) = \sum_{j=1}^{p} k(s - u_j; \psi) w_j, \quad w_j \sim F$$

where F is a distribution with support in $\mathbb{R}^+$, then $x(s) \geq 0$, $\forall s$.

Non-Gaussian Process Convolutions

Elaborating on the previous slide

$$x(s) = \sum_{j=1}^{p} k(s - u_j; \psi) w_j, \quad w_j \sim F$$

where F is a distribution with support in $\mathbb{R}^+$, then $x(s) \geq 0$, $\forall s$.

Consider the normalized kernel

$$k^*(s - u_j; \psi) = \frac{k(s - u_j; \psi)}{\sum_l k(s - u_l; \psi)}$$

then

$$x(s) = \sum_{j=1}^{p} k^*(s - u_j; \psi) w_j, \quad w_j \sim F$$

where F has $(0, 1)$ support, like a beta distribution, then $x(s) \in (0, 1)$, $\forall s$, as $x(s)$ is a convex combination of w_j.

A process with spatially varying kernel can be written as

$$x(\boldsymbol{s}) = \sum_{i=1}^{p} b(\boldsymbol{s} - \boldsymbol{u}_i; \boldsymbol{\psi}(s)) w_i$$

where

$$b(\boldsymbol{s} - \boldsymbol{j}; \boldsymbol{\psi}) \equiv \begin{cases} \left(1 - ||\boldsymbol{s} - \boldsymbol{j}||_{\boldsymbol{\Sigma}}^2\right)^{\psi_1} & \text{if } ||\boldsymbol{s} - \boldsymbol{j}||_{\boldsymbol{\Sigma}} < 1 \\ 0 & \text{otherwise.} \end{cases}$$

and $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_4)$.

A process with spatially varying kernel can be written as

$$x(\boldsymbol{s}) = \sum_{i=1}^{p} b(\boldsymbol{s} - \boldsymbol{u}_i; \boldsymbol{\psi}(s)) w_i$$

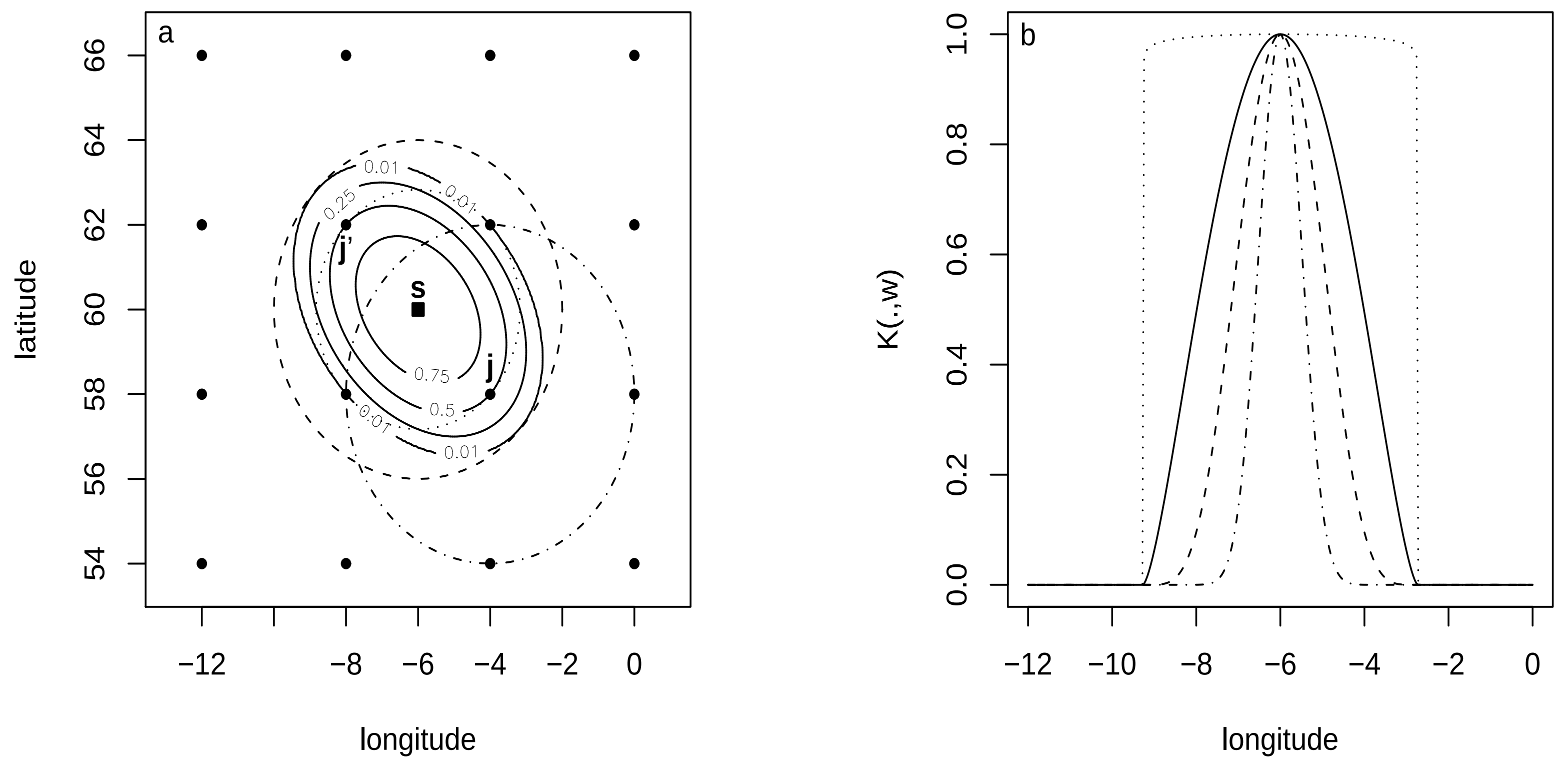
where

$$b(\boldsymbol{s} - \boldsymbol{j}; \boldsymbol{\psi}) \equiv \begin{cases} \left(1 - ||\boldsymbol{s} - \boldsymbol{j}||_{\boldsymbol{\Sigma}}^2\right)^{\psi_1} & \text{if } ||\boldsymbol{s} - \boldsymbol{j}||_{\boldsymbol{\Sigma}} < 1 \\ 0 & \text{otherwise.} \end{cases}$$

and $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_4)$.

The distance is given as

$$||\boldsymbol{s} - \boldsymbol{j}||_{\boldsymbol{\Sigma}} \equiv \sqrt{((x_s - x_j), (y_s - y_j))^T \boldsymbol{\Sigma}^{-1} ((x_s - x_j), (y_s - y_j))}.$$



a
latitude
longitude
s
j'
j
0.01
0.25
0.5
0.75
b
K(.,w)
longitude

The ellipsoidal shape is controlled by the parameters in

$$\boldsymbol{\Sigma}^{-1} \equiv \begin{pmatrix} \Psi_1 + \Psi_2 \cos 2\pi\psi_4 & \Psi_2 \sin 2\pi\psi_4 \\ \Psi_2 \sin 2\pi\psi_4 & \Psi_1 - \Psi_2 \cos 2\pi\psi_4 \end{pmatrix}$$

$$\boldsymbol{\Psi} = \frac{1}{2}\left(\frac{1}{a^2} + \frac{1}{A^2}, \frac{1}{a^2} - \frac{1}{A^2}\right)$$

$$a = L + \psi_2(U - L), \quad A = a + \psi_3(U - a), \quad \psi_2, \psi_3 \in (0, 1)$$

So the semi-minor and semi-major axes a and A belong to (L, U).

The ellipsoidal shape is controlled by the parameters in

$$\boldsymbol{\Sigma}^{-1} \equiv \begin{pmatrix} \Psi_1 + \Psi_2 \cos 2\pi\psi_4 & \Psi_2 \sin 2\pi\psi_4 \\ \Psi_2 \sin 2\pi\psi_4 & \Psi_1 - \Psi_2 \cos 2\pi\psi_4 \end{pmatrix}$$

$$\boldsymbol{\Psi} = \frac{1}{2}\left(\frac{1}{a^2} + \frac{1}{A^2}, \frac{1}{a^2} - \frac{1}{A^2}\right)$$

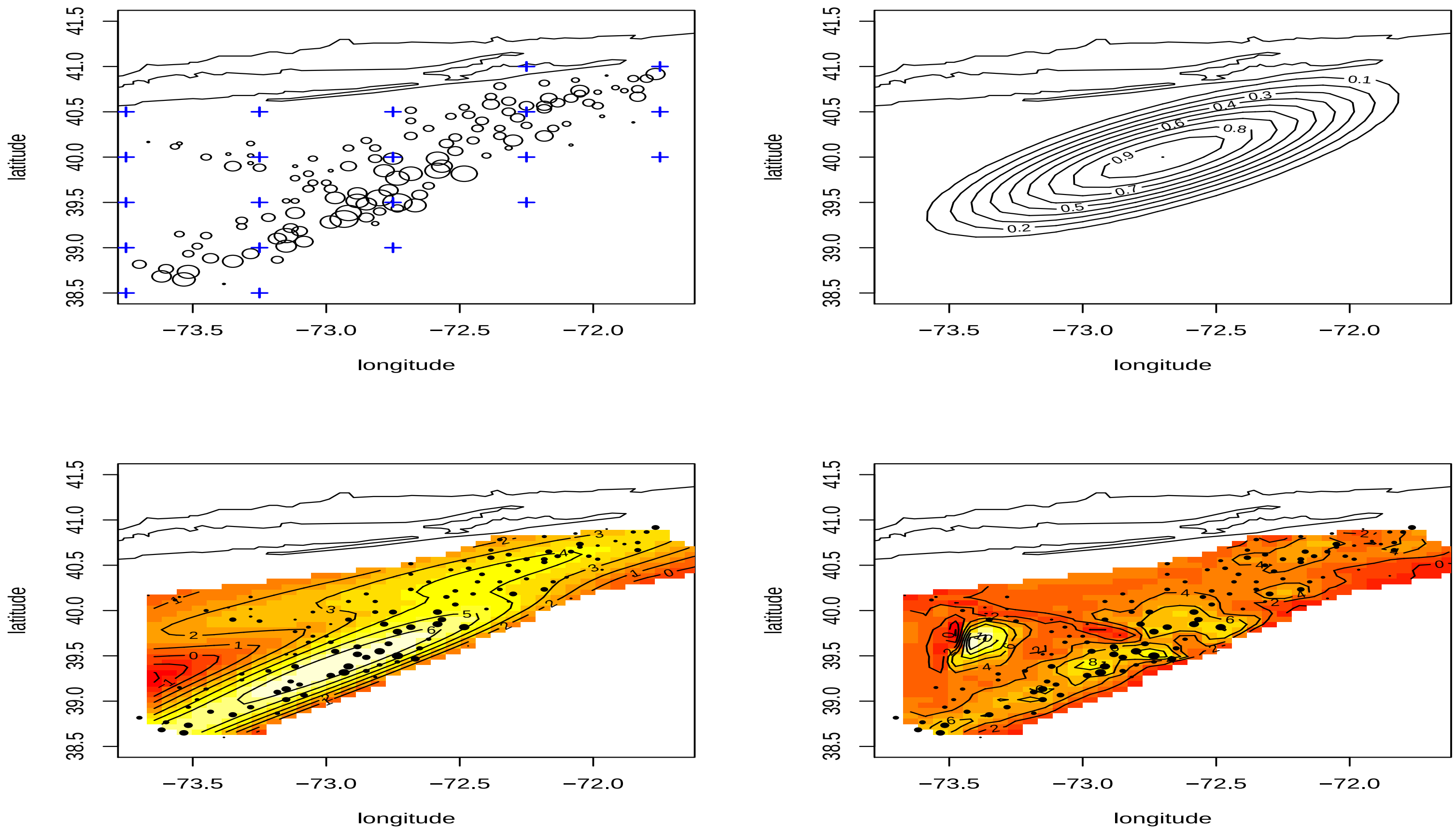
$$a = L + \psi_2(U - L), \quad A = a + \psi_3(U - a), \quad \psi_2, \psi_3 \in (0, 1)$$

So the semi-minor and semi-major axes a and A belong to (L, U).

The spatial variation of $\boldsymbol{\psi}$ is obtained, with a normalized b, as

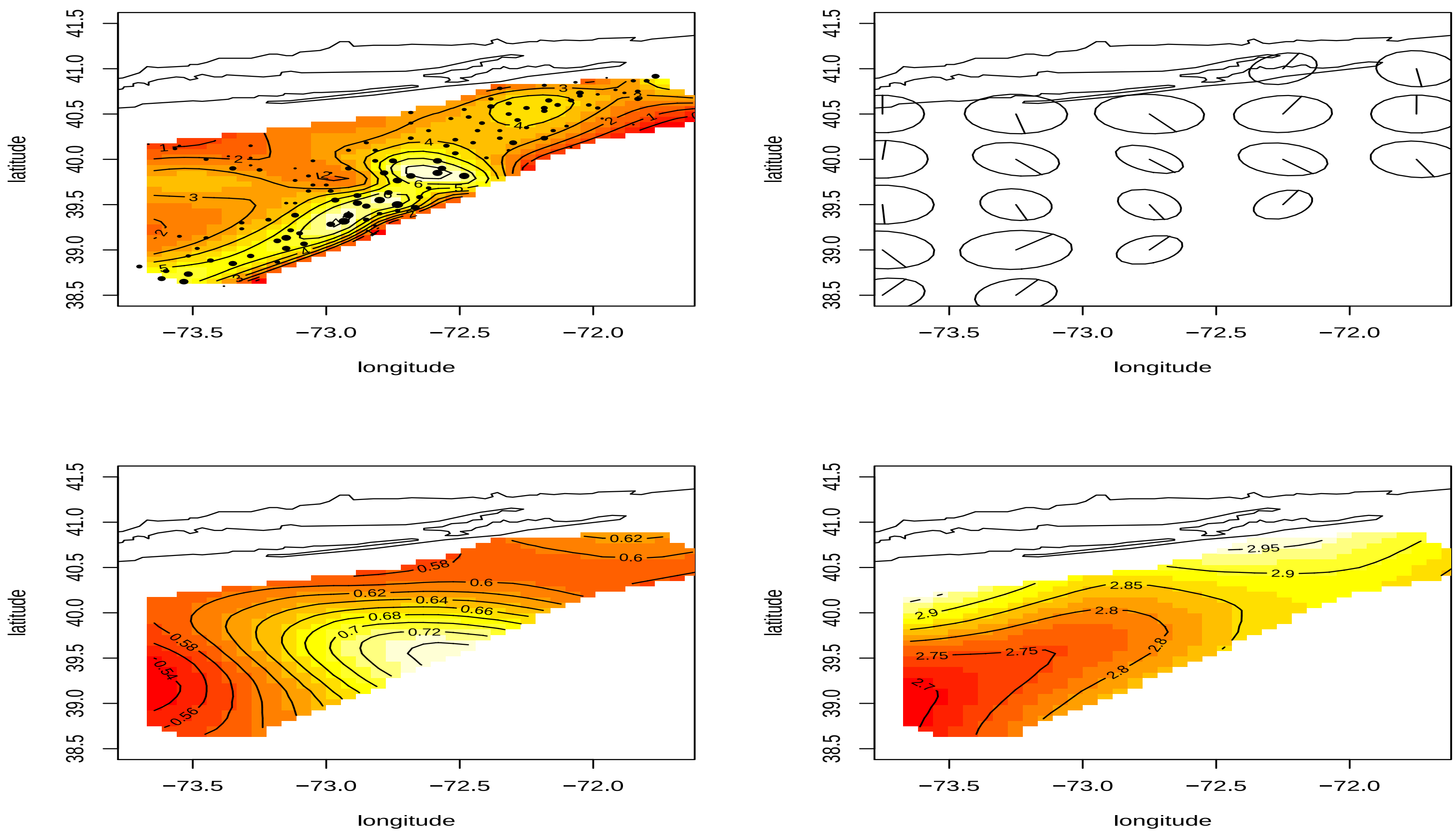
$$\boldsymbol{\psi}(\boldsymbol{s}) = \sum_{i=1}^{p} b(\boldsymbol{s} - \boldsymbol{u}_i; \boldsymbol{\phi})\boldsymbol{\rho}(\boldsymbol{u}_i) \quad \boldsymbol{\phi} = (2, 1, 0, 0)$$

with appropriate uniform priors on each $\rho_k(\boldsymbol{u}_i), k = 1, \ldots, 4$.



Using the DPC with fixed parameter Bézier kernels on the scallops data.

We use a fixed ellipsoidal kernel that follows the coastline.



Letting the four parameters be DPCs with spherical kernels we have space-varying ellipsoidal shapes and smoothness. Lower panels: eccentricity (left) and smoothness (right).

Modeling issues

- In the previous model we add measurement error and usually a linear drift.
- We need to specify a prior for $\boldsymbol{w}$. Usually $p(\boldsymbol{w}) \propto 1$ works fine.
- The compact support of $b(\cdot)$ produces sparse matrices in the resulting design matrix of the linear model. This can be used to speed up calculations.
- How many knots do we use? Where do we put them? There are no good answers to these questions. We can use model comparison methods to choose between different configurations. Nevertheless, intuitively, the space-varying nature of the support should compensate for small or sparse grids.