Stochastic multiscale space-time modelling and practical Bayesian inference
Part 1: GRF/SPDE/GMRF in space-time

Finn Lindgren
http://www.maths.ed.ac.uk/~flindgre/siamuq18.html

SIAM UQ 2018

“Big” data

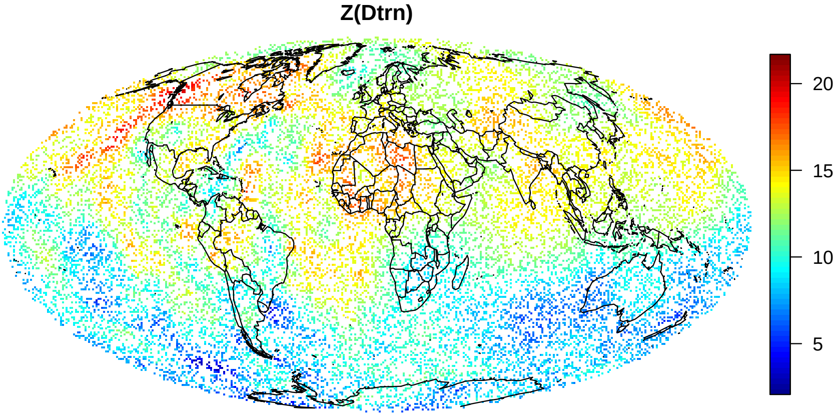
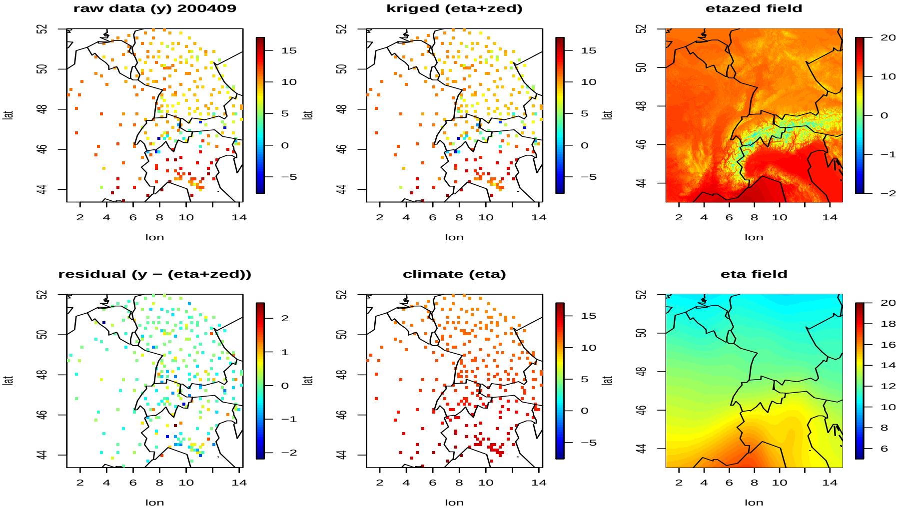


Illustration: Synthetic data mimicking satellite based CO_2 measurements.
Iregular data locations, uneven coverage, features on different scales.

Sparse spatial coverage of temperature measurements



Regional observations: $\approx$ 20,000,000 from daily timeseries over 160 years

Spatio-temporal modelling framework

Spatial statistics framework

- Spatial domain D, or space-time domain $D \times \mathbb{T}$, $\mathbb{T} \subset \mathbb{R}$.
- Random field $u(\boldsymbol{s})$, $\boldsymbol{s} \in D$, or $u(\boldsymbol{s}, t)$, $(\boldsymbol{s}, t) \in D \times \mathbb{T}$.
- Observations y_i. In the simplest setting, $y_i = u(\boldsymbol{s}_i) + \epsilon_i$, but more generally $y_i \sim \text{GLMM}$, with $u(\cdot)$ as a structured random effect.
- Needed: models capturing stochastic dependence on multiple scales
- Partial solution: Basis function expansions, with large scale functions and covariates to capture static and slow structures, and small scale functions for more local variability

Two basic model and method components

- Stochastic models for $u(\cdot)$.
- Computationally efficient (i.e. avoid MCMC whenever possible) inference methods for the posterior distribution of $u(\cdot)$ given data $\boldsymbol{y}$.

Covariance functions and stochastic PDEs

The Matérn covariance family on $\mathbb{R}^d$

$$\mathsf{Cov}(u(\mathbf{0}), u(\boldsymbol{s})) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} (\kappa \|\boldsymbol{s}\|)^\nu K_\nu(\kappa \|\boldsymbol{s}\|)$$

Scale $\kappa > 0$, smoothness $\nu > 0$, variance $\sigma^2 > 0$

Whittle (1954, 1963): Matérn as SPDE solution

Matérn fields are the stationary solutions to the SPDE

$$\left(\kappa^2 - \nabla \cdot \nabla\right)^{\alpha/2} u(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}), \quad \alpha = \nu + d/2$$

$\mathcal{W}(\cdot)$ white noise, $\nabla \cdot \nabla = \sum_{i=1}^d \frac{\partial^2}{\partial s_i^2}$, $\sigma^2 = \frac{\Gamma(\nu)}{\Gamma(\alpha)\kappa^{2\nu}(4\pi)^{d/2}}$

Questions: What is “white noise”? How can we define fractional operators?

How can we deal with SPDEs in hierarchical models? Why do we want to?

Gaussian random field (or Gaussian process)

A *Gaussian random field* $u : D \mapsto \mathbb{R}$ is defined via

$$
\begin{aligned}
\mathsf{E}(u(\mathbf{s})) &= m(\mathbf{s}), \\
\mathsf{Cov}(u(\mathbf{s}), u(\mathbf{s}')) &= K(\mathbf{s}, \mathbf{s}'), \quad \text{(covariance kernel)} \\
\left[u(\mathbf{s}_i), i = 1, \dots, n\right] &\sim \mathcal{N}\Big(\boldsymbol{m} = \left[m(\mathbf{s}_i), i = 1, \dots, n\right], \\
&\qquad \boldsymbol{\Sigma} = \left[K(\mathbf{s}_i, \mathbf{s}_j), i, j = 1, \dots, n\right]\Big)
\end{aligned}
$$

for all finite location sets $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$, and $K(\cdot,\cdot)$ symmetric positive definite.

Generalised random field

A *generalised Gaussian random field* $u : D \mapsto \mathbb{R}$ is defined via a random measure, $\langle f, u\rangle_D = u^*(f) : H_{\mathcal{R}}(D) \mapsto \mathbb{R}$, $\mathcal{R}$ a covariance operator,

$$
\begin{aligned}
\mathsf{E}(\langle f, u\rangle_D) &= \langle f, m\rangle_D = \int_D f(\mathbf{s}) m(\mathbf{s})\,\mathrm{d}\mathbf{s}, \\
\mathsf{Cov}(\langle f, u\rangle_D, \langle g, u\rangle_D) &= \langle f, \mathcal{R} g\rangle_D \equiv \iint_{D\times D} f(\mathbf{s}) K(\mathbf{s}, \mathbf{s}') g(\mathbf{s}')\,\mathrm{d}\mathbf{s}\,\mathrm{d}\mathbf{s}', \\
\langle f, u\rangle_D &\sim \mathcal{N}(\langle f, m\rangle_D, \langle f, \mathcal{R} f\rangle_D)
\end{aligned}
$$

for all $f, g \in H_{\mathcal{R}}(D) \equiv \{f : D \mapsto \mathbb{R};\ \langle f, \mathcal{R} f\rangle_D < \infty\}$.
This allows for singular covariance kernels $K(\cdot,\cdot)$.

White noise vs independent noise

Gaussian white noise on continuous domains

Standard Gaussian white noise $\mathcal{W}(\cdot)$ is a generalised random field, with

$$m(\mathbf{s}) = 0, \qquad K(\mathbf{s}, \mathbf{s}') = \delta_{\mathbf{s}}(\mathbf{s}'), \qquad \langle f, \mathcal{W} \rangle_D \sim \mathcal{N}(0, \langle f, f \rangle_D),$$

for all $f \in L_2(D)$. Since $\langle \delta_{\mathbf{s}}, \delta_{\mathbf{s}} \rangle_D = \infty$ for all $\mathbf{s} \in D$, $\mathcal{W}(\cdot)$ does not have pointwise meaning. We can only do calculus!

Independent Gaussian noise

Spatially independent Gaussian noise $w(\cdot)$ is a random field, with

$$m(\mathbf{s}) = 0, \qquad K(\mathbf{s}, \mathbf{s}') = \mathbf{1}_{\{\mathbf{s}=\mathbf{s}'\}}, \qquad w(\mathbf{s}) \sim \mathcal{N}(0, 1),$$

for all $\mathbf{s}, \mathbf{s}' \in D$. However, for every set $A \subset D$ with $|A|_{\mathsf{Leb(D)}} > 0$,

$$\mathsf{P}(\sup_{\mathbf{s} \in A} w(\mathbf{s}) = \infty) = \mathsf{P}(\inf_{\mathbf{s} \in A} w(\mathbf{s}) = -\infty) = 1,$$

and the generalised calculus is not applicable.

Spectral properties

Bochner's theorem on $\mathbb{R}^d$

A symmetric kernel $K(\mathrm{s}, \mathrm{s}')$, $\mathrm{s}, \mathrm{s}' \in \mathbb{R}^d$, is a positive (semi-)definite stationary covariance kernel if and only if there exists a non-negative spectral measure $S^*(\boldsymbol{\omega})$ such that

$$K(\mathrm{s}, \mathrm{s}') = \int_{\mathbb{R}^d} \exp(i(\mathrm{s}' - \mathrm{s}) \cdot \boldsymbol{\omega}) \,\mathrm{d}S^*(\boldsymbol{\omega})$$

If the measure has a density $S(\boldsymbol{\omega})$,

$$K(\mathrm{s}, \mathrm{s}') = \int_{\mathbb{R}^d} \exp(i(\mathrm{s}' - \mathrm{s}) \cdot \boldsymbol{\omega}) S(\boldsymbol{\omega}) \,\mathrm{d}\boldsymbol{\omega}$$

$$S(\boldsymbol{\omega}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \exp(-i\mathrm{s} \cdot \boldsymbol{\omega}) K(0, \mathrm{s}) \,\mathrm{ds}$$

White noise on $\mathbb{R}^d$ has spectral density $S_{\mathcal{W}}(\boldsymbol{\omega}) = 1/(2\pi)^d$.

Let $D = \mathbb{R}^d$, and $\widehat{f}(\boldsymbol{\omega}) = (\mathcal{F}f)(\boldsymbol{\omega}) = \int_{\mathbb{R}^d} \exp(i\mathbf{s}\cdot\boldsymbol{\omega}) f(\mathbf{s})\,\mathrm{d}\mathbf{s}$.
Very informally, $S_u(\boldsymbol{\omega}) = \mathsf{E}(|\widehat{u}(\boldsymbol{\omega})|^2)$.
Differential operators can also be interpreted spectrally:

$\mathcal{L}f$	f	∇f	$-\nabla\cdot\nabla f$	$\mathcal{L}^{\alpha/2} f$
$\widehat{\mathcal{L}}\widehat{f} \equiv \mathcal{F}(\mathcal{L}f)$	$\widehat{f}$	$i\boldsymbol{\omega}\widehat{f}$	$\|\boldsymbol{\omega}\|^2\widehat{f}$	$\lvert\widehat{\mathcal{L}}\rvert^{\alpha/2}\widehat{f}$

The rightmost column is a *definition* of a fractional operator!

For the Whittle-Matérn SPDE, informally,

$$
\begin{aligned}
(\kappa^2 - \nabla\cdot\nabla)^{\alpha/2} u(\mathbf{s}) &= \mathcal{W}(\mathbf{s}) \\
(\kappa^2 + \|\boldsymbol{\omega}\|^2)^{\alpha/2} \widehat{u}(\boldsymbol{\omega}) &= \widehat{\mathcal{W}}(\boldsymbol{\omega}) \\
\mathsf{E}(|(\kappa^2 + \|\boldsymbol{\omega}\|^2)^{\alpha/2} \widehat{u}(\boldsymbol{\omega})|^2) &= \mathsf{E}(|\widehat{\mathcal{W}}(\boldsymbol{\omega})|^2) \\
(\kappa^2 + \|\boldsymbol{\omega}\|^2)^{\alpha} S_u(\boldsymbol{\omega}) &= S_{\mathcal{W}}(\boldsymbol{\omega}) \\
S_u(\boldsymbol{\omega}) &= \frac{1}{(2\pi)^d(\kappa^2 + \|\boldsymbol{\omega}\|^2)^{\alpha}}
\end{aligned}
$$

Whittle (1954, 1963) showed that $(\mathcal{F}S_u(\cdot))(\mathbf{s}'-\mathbf{s})$ is equal to the Matérn covariance (up to a known scaling constant), with smoothness $\nu = \alpha - d/2$.

Simple heat equation

For space-time fields, we write $u(\mathbf{s},t)$, $(\mathbf{s},t) \in \mathbb{R}^d \times \mathbb{R}$, and $S_u(\boldsymbol{k},\omega)$, $(\boldsymbol{k},\omega) \in \mathbb{R}^d \times \mathbb{R}$.

We drive a heat equation with a noise process $\mathcal{E}$ that is white noise in time and Matérn noise in space, with parameters matching the heat operator:

$$
\begin{aligned}
\left\{\gamma \frac{\partial}{\partial t} + \kappa^2 - \nabla_s \cdot \nabla_s\right\} u(\mathbf{s}) &= \mathcal{E}(\mathbf{s},t), \\
(\kappa^2 - \nabla_s \cdot \nabla_s)^{\alpha/2} \mathcal{E}(\mathbf{s},t) &= \mathcal{W}(\mathbf{s},t).
\end{aligned}
$$

The Fourier domain version is

$$
\begin{aligned}
\left\{i\gamma\omega + \kappa^2 + \|\boldsymbol{k}\|^2\right\} \widehat{u}(\boldsymbol{k},\omega) &= \widehat{\mathcal{E}}(\boldsymbol{k},\omega), \\
(\kappa^2 + \|\boldsymbol{k}\|^2)^{\alpha/2} \widehat{\mathcal{E}}(\boldsymbol{k},\omega) &= \widehat{\mathcal{W}}(\boldsymbol{k},\omega),
\end{aligned}
$$

and

$$
S_u(\boldsymbol{k},\omega) = \frac{1}{(2\pi)^{d+1}(\gamma^2\omega^2 + (\kappa^2 + \|\boldsymbol{k}\|^2)^2)(\kappa^2 + \|\boldsymbol{k}\|^2)^{\alpha}}
$$

How differentiable are the realisations?

Simple heat equation (cont)

Using that, in the standardised Whittle-Matérn SPDE, the variance is

$$\sigma^2 = \frac{\Gamma(\nu)}{\Gamma(\alpha)\kappa^{2\nu}(4\pi)^{d/2}}, \quad \nu = \alpha - d/2,$$

the marginal spatial spectrum for the heat model is

$$S_u(\boldsymbol{k}) = \int_{\mathbb{R}} S_u(\boldsymbol{k},\omega)\,\mathrm{d}\omega = \frac{1}{4\pi\gamma}\frac{1}{(2\pi)^d(\kappa^2+\|\boldsymbol{k}\|^2)^{\alpha+1}},$$

which is a scaled Whittle spectrum for a Matérn covariance with smoothness $\nu = \alpha + 1 - d/2$.

A generalised generalised case

If $\alpha = 0$, $d = 2$, then $\nu = 0$, which is just outside of the allowed range of the Matérn family. However, for every t, $u(\cdot,t)$ is a generalised random field with singular kernel $K(\mathbf{s},\mathbf{s}') = \frac{1}{4\pi\gamma}\frac{1}{2\pi}K_0(\kappa\|\mathbf{s}'-\mathbf{s}\|)$.

Simple heat equation (cont)

To help understand the temporal properties, take the Fourier transform in only the spatial directions:

$$\left\{\gamma\frac{\partial}{\partial t}+\kappa^2+\|\boldsymbol{k}\|^2\right\}\widetilde{u}(\boldsymbol{k},t)=\frac{\widetilde{\mathcal{W}}(\boldsymbol{k},t)}{(\kappa^2+\|\boldsymbol{k}\|^2)^{\alpha/2}},$$

so for each spatial frequency $\boldsymbol{k}$, the temporal evolution of $\widetilde{u}(\boldsymbol{k},t)$ is an Ornstein-Uhlenbeck process with covariance

$$\frac{1}{4\pi\gamma(\kappa^2+\|\boldsymbol{k}\|^2)^{\alpha+1}}\exp\left(-|t|\frac{\kappa^2+\|\boldsymbol{k}\|^2}{\gamma}\right).$$

There is one more property we need to understand: Markov in space

First order Markov in time

Filtration σ-algebras:

$$a \in \mathcal{F}^\sigma_{(-\infty,t]} \equiv \sigma(u(s), s \leq t), \quad b \in \mathcal{F}^\sigma_{[t,\infty)} \equiv \sigma(u(s), s \geq t)$$
$$\mathsf{P}(a \cap b \mid u(t)) = \mathsf{P}(a \mid u(t))\mathsf{P}(b \mid u(t))$$

Higher order Markov on spatial and spatio-temporal domains

Let $A, B, S \subset D$, such that S separates A and B.

$$\mathcal{F}^\sigma_S \equiv \sigma(u(\mathbf{s}), \mathbf{s} \in S), \quad a \in \mathcal{F}^\sigma_A, \quad b \in \mathcal{F}^\sigma_B,$$
$$\mathsf{P}(a \cap b \mid \mathcal{F}^\sigma_S) = \mathsf{P}(a \mid \mathcal{F}^\sigma_S)\mathsf{P}(b \mid \mathcal{F}^\sigma_S)$$

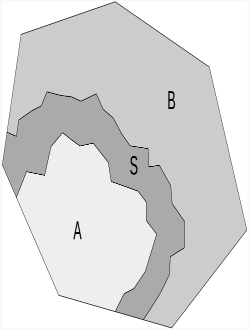
Markov for generalised random fields

$$\mathcal{F}^\sigma_S \equiv \sigma(\langle f, u\rangle_S, f \in H_{\mathcal{R}}(S)), \quad a \in \mathcal{F}^\sigma_A, \quad b \in \mathcal{F}^\sigma_B,$$
$$\mathsf{P}(a \cap b \mid \mathcal{F}^\sigma_S) = \mathsf{P}(a \mid \mathcal{F}^\sigma_S)\mathsf{P}(b \mid \mathcal{F}^\sigma_S)$$

Markov in space

Markov properties

S is a separating set for A and B: $u(A) \perp u(B) \mid u(S)$



Solutions to

$$\left(\kappa^2 - \nabla \cdot \nabla\right)^{\alpha/2} u(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s})$$

are Markov when α is an integer.

More generally, when the reciprocal of the spectral density is a polynomial, Rozanov, 1977

In graphs with no edges between A and B ($\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}$):

$$\boldsymbol{Q}_{AB} = \boldsymbol{0}$$
$$\boldsymbol{Q}_{A|S,B} = \boldsymbol{Q}_{AA}$$
$$\boldsymbol{\mu}_{A|S,B} = \boldsymbol{\mu}_A - \boldsymbol{Q}_{AA}^{-1}\boldsymbol{Q}_{AS}(\boldsymbol{u}_S - \boldsymbol{\mu}_S)$$

Generally: Markov iff the precision *operator* $\mathcal{Q} = \mathcal{R}^{-1}$ is local.

Precision matrix block structure: $\begin{bmatrix} \boldsymbol{Q}_{AA} & \boldsymbol{Q}_{AS} & \boldsymbol{0} \\ \boldsymbol{Q}_{SA} & \boldsymbol{Q}_{SS} & \boldsymbol{Q}_{SB} \\ \boldsymbol{0} & \boldsymbol{Q}_{BS} & \boldsymbol{Q}_{BB} \end{bmatrix}$

Hilbert space approximation

We want to construct finite dimensional approximations to the distribution of $u(\cdot)$, where

$$\left[\left\langle f_i, (\kappa^2 - \nabla\cdot\nabla)^{\alpha/2} u(\cdot)\right\rangle_D, \, i = 1, \ldots, m\right] \overset{d}{=} \left[\left\langle f_i, \mathcal{W}(\cdot)\right\rangle_D, \, i = 1, \ldots, m\right]$$

for all finite collections of test functions $f_i \in H_{\mathcal{R}}(D)$.

A finite basis expansion

$$u(\mathbf{s}) = \sum_{j=1}^{n} \psi_j(\mathbf{s}) u_j$$

can only hope to achieve this for a subspace of size n.

Two main approaches:

- Galerkin: $\{f_i = \psi_i, \, i = 1, \ldots, n\}$
- Least squares: $\{f_i = (\kappa^2 - \nabla\cdot\nabla)^{\alpha/2}\psi_i, \, i = 1, \ldots, n\}$

We use least squares for $\alpha = 1$, Galerkin for $\alpha = 2$, and a recursion for $\alpha \geq 3$.

Stochastic Green's first identity

On any sufficiently smooth manifold domain D,

$$\langle f, -\nabla\cdot\nabla g\rangle_D = \langle \nabla f, \nabla g\rangle_D - \langle f, \partial_n g\rangle_{\partial D}$$

holds, even if either ∇f or $-\nabla\cdot\nabla g$ are as generalised as white noise.

For now, we'll impose deterministic Neumann boundary conditions, informally $\partial_n u(\mathbf{s}) = 0$ for all $\mathbf{s} \in \partial D$. For $\alpha = 2$ and Galerkin,

$$\begin{aligned}\left\langle \psi_i, (\kappa^2 - \nabla\cdot\nabla) \sum_j \psi_j u_j \right\rangle_D &= \sum_j \left\{ \kappa^2 \langle \psi_i, \psi_j\rangle_D + \langle \nabla\psi_i, \nabla\psi_j\rangle_D \right\} u_j \\ &= (\kappa^2 \boldsymbol{C} + \boldsymbol{G})\boldsymbol{u}\end{aligned}$$

The covariance for the RHS of the SPDE is

$$\left[\mathsf{Cov}(\langle \psi_i, \mathcal{W}\rangle_D, \langle \psi_j, \mathcal{W}\rangle_D)\right] = \left[\langle \psi_i, \psi_j\rangle_D\right] = \boldsymbol{C}$$

by the definition of $\mathcal{W}$.

We seek $\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ such that $\mathsf{Var}\{(\kappa^2\boldsymbol{C} + \boldsymbol{G})\boldsymbol{u}\} = \boldsymbol{C}$:

$$(\kappa^2\boldsymbol{C} + \boldsymbol{G})\boldsymbol{\Sigma}(\kappa^2\boldsymbol{C} + \boldsymbol{G}) = \boldsymbol{C}$$
$$\boldsymbol{\Sigma} = (\kappa^2\boldsymbol{C} + \boldsymbol{G})^{-1}\boldsymbol{C}(\kappa^2\boldsymbol{C} + \boldsymbol{G})^{-1}$$

If ψ_i are piecewise linear on a triangulation of D, then $\boldsymbol{C}$ and $\boldsymbol{G}$ are both very sparse, and in addition, $\boldsymbol{C} = \mathrm{diag}(\langle\psi_i, 1\rangle_D)$ is a valid approximation. Then, the *precision* matrix is also sparse,

$$\boldsymbol{Q} = (\kappa^2\boldsymbol{C} + \boldsymbol{G})\boldsymbol{C}^{-1}(\kappa^2\boldsymbol{C} + \boldsymbol{G})$$

and $\boldsymbol{u}$ is Markov on the adjacency graph given by the non-zero structure of $\boldsymbol{Q}$.

Least squares and Galerkin recursion gives precisions for all $\alpha = 1, 2, \ldots$:

- $\boldsymbol{Q}_1 = (\kappa^2\boldsymbol{C} + \boldsymbol{G})$
- $\boldsymbol{Q}_2 = (\kappa^2\boldsymbol{C} + \boldsymbol{G})\boldsymbol{C}^{-1}(\kappa^2\boldsymbol{C} + \boldsymbol{G}) = \kappa^4\boldsymbol{C} + 2\kappa^2\boldsymbol{G} + \boldsymbol{G}\boldsymbol{C}^{-1}\boldsymbol{G}$
- $\boldsymbol{Q}_\alpha = (\kappa^2\boldsymbol{C} + \boldsymbol{G})\boldsymbol{C}^{-1}\boldsymbol{Q}_{\alpha-2}\boldsymbol{C}^{-1}(\kappa^2\boldsymbol{C} + \boldsymbol{G})$
- Any $\alpha \geq 0$: $\boldsymbol{Q}_\alpha = \boldsymbol{C}^{1/2}\left\{\boldsymbol{C}^{-1/2}(\kappa^2\boldsymbol{C} + \boldsymbol{G})\boldsymbol{C}^{-1/2}\right\}^\alpha \boldsymbol{C}^{1/2}$ (non-sparse for non-integer α)

Basis function representations for Gaussian Matérn fields

Basis definitions

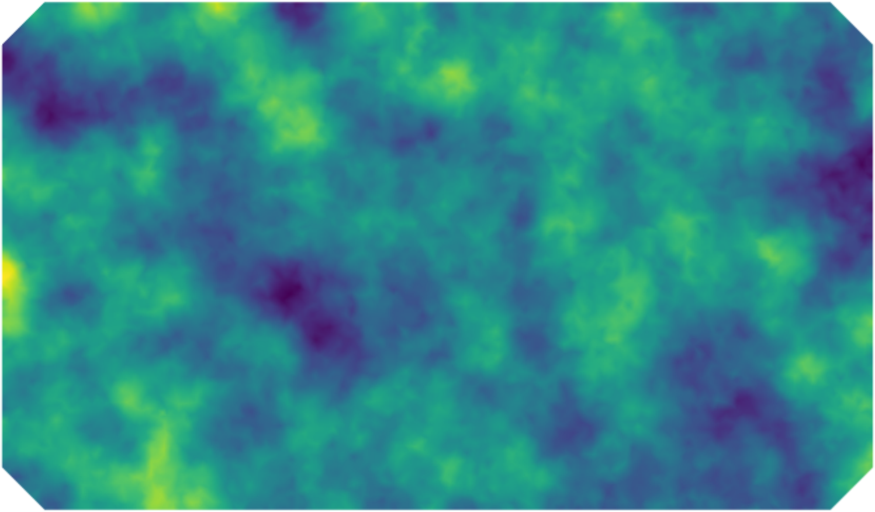
	Finite basis set ($k = 1, \ldots, n$)
Karhunen-Loève	$(\kappa^2 - \nabla\cdot\nabla)^{-\alpha} e_{\kappa,k}(\boldsymbol{s}) = \lambda_{\kappa,k} e_{\kappa,k}(\boldsymbol{s})$
Fourier	$-\nabla\cdot\nabla e_k(\boldsymbol{s}) = \lambda_k e_k(\boldsymbol{s})$
Convolution	$(\kappa^2 - \nabla\cdot\nabla)^{\alpha/2} g_\kappa(\boldsymbol{s}) = \delta(\boldsymbol{s})$
General	$\psi_k(\boldsymbol{s})$

Field representations

	Field $u(\boldsymbol{s})$	Weights
Karhunen-Loève	$\propto \sum_k e_{\kappa,k}(\boldsymbol{s}) z_k$	$z_k \sim \mathcal{N}(0, \lambda_{\kappa,k})$
Fourier	$\propto \sum_k e_k(\boldsymbol{s}) z_k$	$z_k \sim \mathcal{N}(0, (\kappa^2 + \lambda_k)^{-\alpha})$
Convolution	$\propto \sum_k g_\kappa(\boldsymbol{s} - \boldsymbol{s}_k) z_k$	$z_k \sim \mathcal{N}(0, \lvert\text{cell}_k\rvert)$
General	$\propto \sum_k \psi_k(\boldsymbol{s}) u_k$	$\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_\kappa^{-1})$

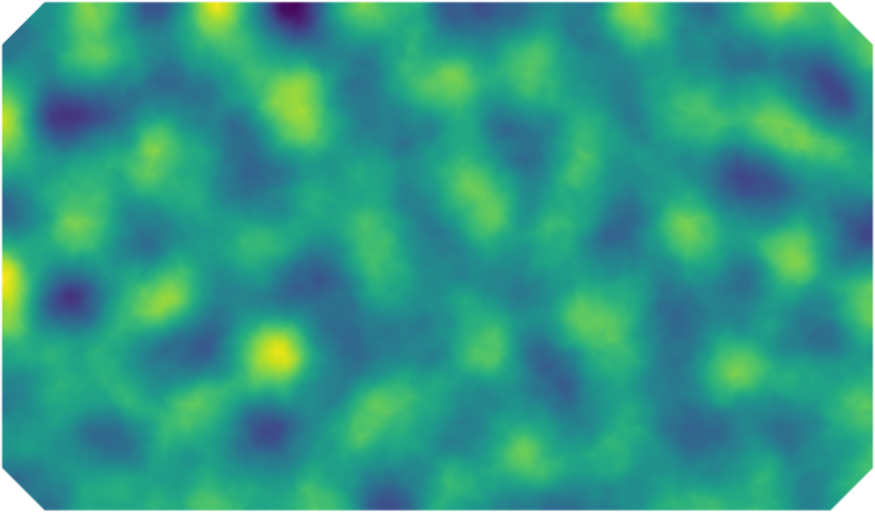
Note: Harmonic basis functions (as in the Fourier approach) give a diagonal $\boldsymbol{Q}_\kappa$, but lead to dense posterior precision matrices.

SPDE/GMRF realisations and non-stationary models



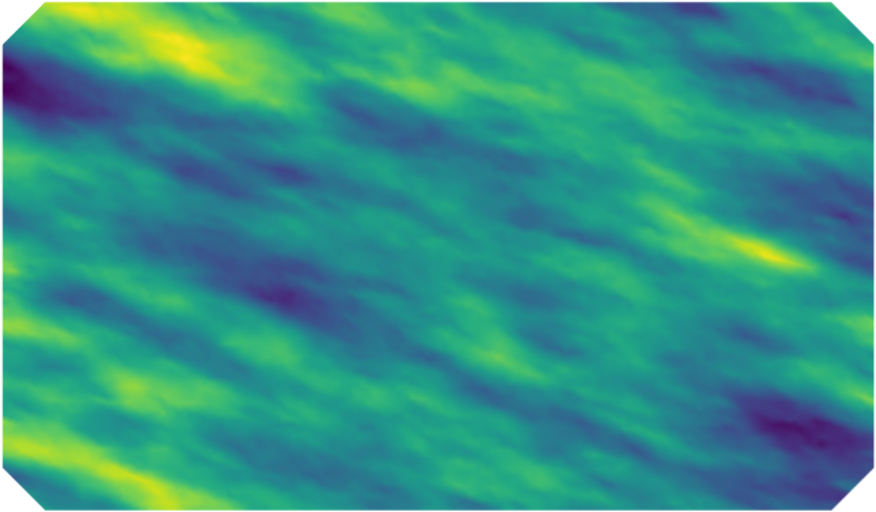
$$(\kappa^2 - \nabla \cdot \nabla) u(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}), \quad \boldsymbol{s} \in D$$

SPDE/GMRF realisations and non-stationary models



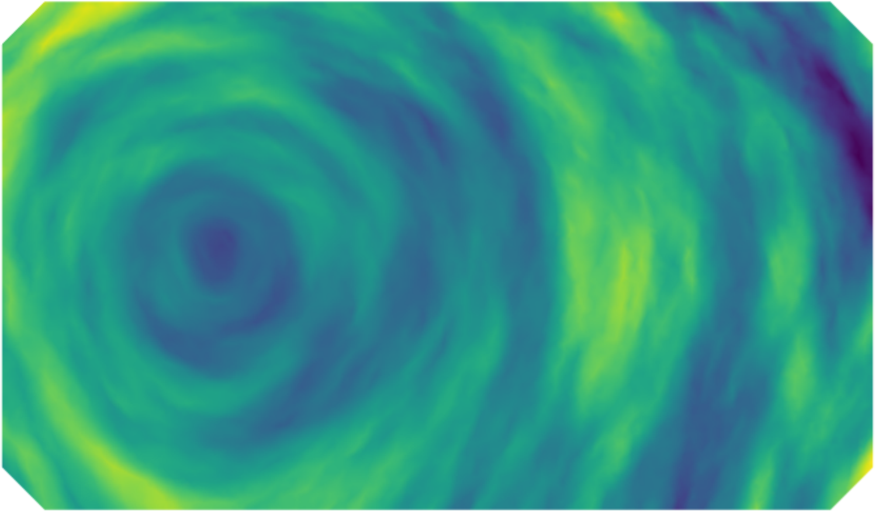
$(\kappa^2 \exp(i\theta) - \nabla \cdot \nabla)u(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s})$, $\boldsymbol{s} \in D$, Re(u) independent of Im(u)

SPDE/GMRF realisations and non-stationary models



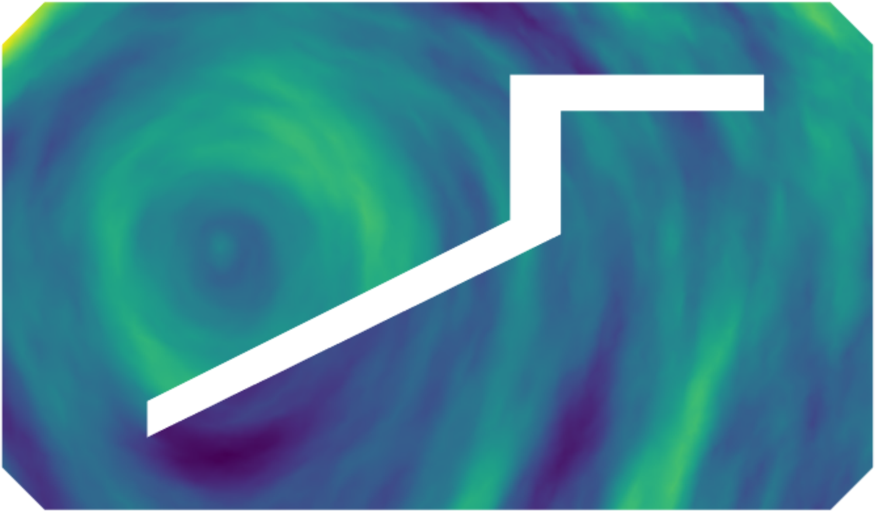
$$(\kappa^2 - \nabla \cdot \boldsymbol{H} \nabla) u(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}), \quad \boldsymbol{s} \in D$$

SPDE/GMRF realisations and non-stationary models



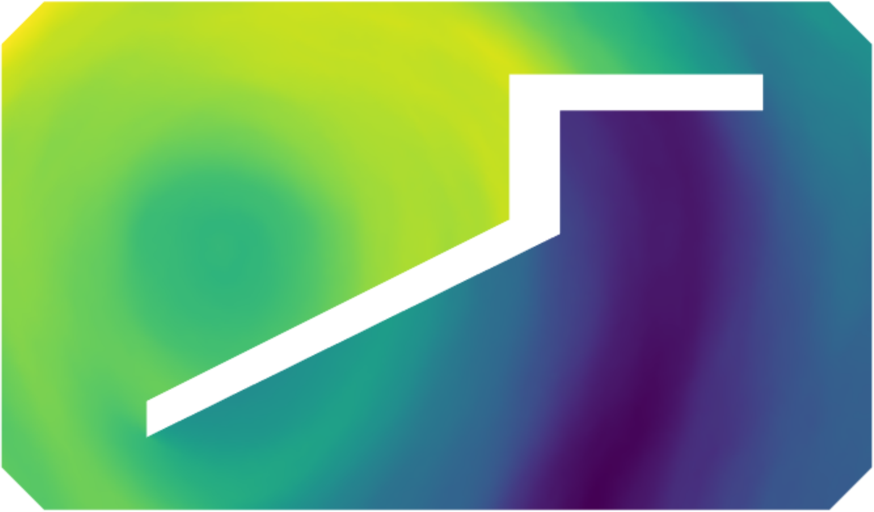
$$(\kappa^2 - \nabla \cdot \boldsymbol{H}(\boldsymbol{s})\nabla)u(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}), \quad \boldsymbol{s} \in D$$

SPDE/GMRF realisations and non-stationary models



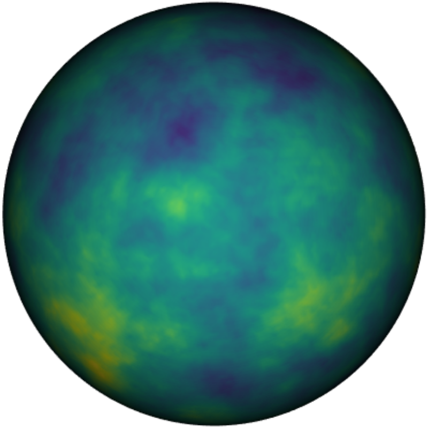
$$(\kappa^2 - \nabla \cdot \nabla)u(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}), \quad \boldsymbol{s} \in D$$

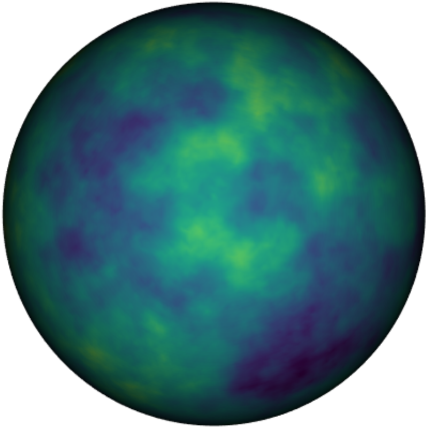
SPDE/GMRF realisations and non-stationary models



$$(\kappa^2 - \nabla \cdot \nabla)u(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}), \quad \boldsymbol{s} \in D$$

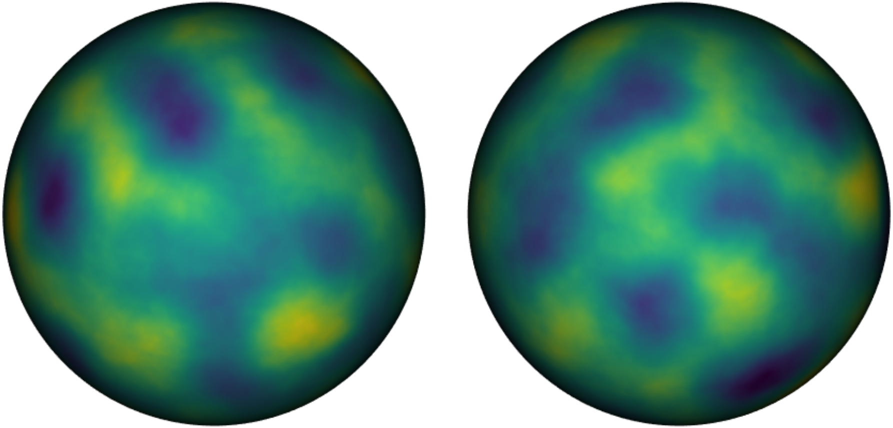
SPDE/GMRF realisations and non-stationary models





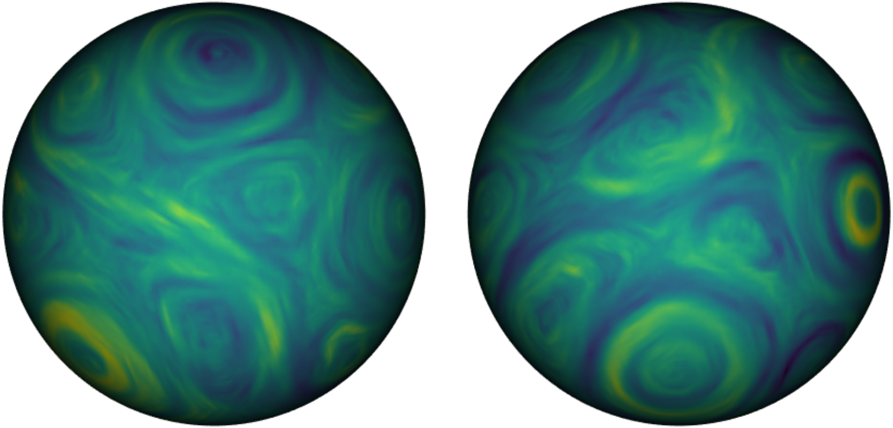
$$(\kappa^2 - \nabla \cdot \nabla)u(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}), \quad \boldsymbol{s} \in D = \mathbb{S}^2$$

SPDE/GMRF realisations and non-stationary models



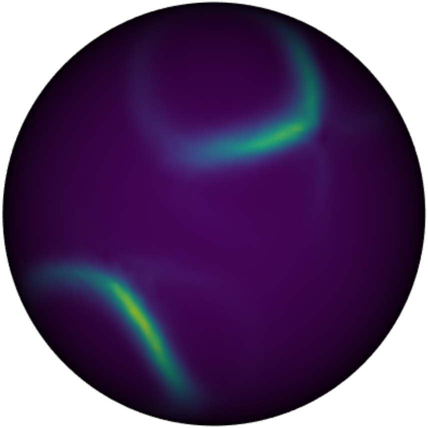
$$(\kappa^2 \exp(i\theta) - \nabla \cdot \nabla) u(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}), \quad \boldsymbol{s} \in D = \mathbb{S}^2$$

Markov does *not* mean local dependence



$$(\kappa(\boldsymbol{s})^2 - \nabla \cdot \boldsymbol{H}(\boldsymbol{s})\nabla)u(\boldsymbol{s}) = \kappa(\boldsymbol{s})\mathcal{W}(\boldsymbol{s}), \quad \boldsymbol{s} \in \Omega$$

Covariances for four reference points



Hierarchical models

Continuous Markovian spatial models (Lindgren et al, 2011)

Local basis: $u(\boldsymbol{s}) = \sum_k \psi_k(\boldsymbol{s}) u_k$, (compact, piecewise linear)

Basis weights: $\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}^{-1})$, sparse $\boldsymbol{Q}$ based on an SPDE

Special case: $(\kappa^2 - \nabla \cdot \nabla) u(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s})$, $\boldsymbol{s} \in \Omega$

Precision: $\boldsymbol{Q} = \kappa^4 \boldsymbol{C} + 2\kappa^2 \boldsymbol{G} + \boldsymbol{G}_2$ $(\kappa^4 + 2\kappa^2 |\boldsymbol{\omega}|^2 + |\boldsymbol{\omega}|^4)$

Conditional distribution in a jointly Gaussian model

$$\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{\mu}_u, \boldsymbol{Q}_u^{-1}), \quad \boldsymbol{y}|\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{u}, \boldsymbol{Q}_{y|u}^{-1}) \qquad (A_{ij} = \psi_j(\boldsymbol{s}_i))$$

$$\boldsymbol{u}|\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}_{u|y}, \boldsymbol{Q}_{u|y}^{-1})$$

$$\boldsymbol{Q}_{u|y} = \boldsymbol{Q}_u + \boldsymbol{A}^T \boldsymbol{Q}_{y|u} \boldsymbol{A}$$ (∼"Sparse iff ψ_k have compact support")

$$\boldsymbol{\mu}_{u|y} = \boldsymbol{\mu}_u + \boldsymbol{Q}_{u|y}^{-1} \boldsymbol{A}^T \boldsymbol{Q}_{y|u} (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\mu}_u)$$

The computational GMRF work-horse

Cholesky decomposition (Cholesky, 1924)

$$\boldsymbol{Q} = \boldsymbol{L}\boldsymbol{L}^\top, \quad \boldsymbol{L} \text{ lower triangular } (\sim \mathcal{O}(n^{(d+1)/2}) \text{ for } d = 1, 2, 3)$$

$$\boldsymbol{Q}^{-1}\boldsymbol{x} = \boldsymbol{L}^{-\top}\boldsymbol{L}^{-1}\boldsymbol{x}, \quad \text{via forward/backward substitution}$$

$$\log\det \boldsymbol{Q} = 2\log\det \boldsymbol{L} = 2\sum_i \log L_{ii}$$

André-Louis Cholesky (1875–1918)

"He invented, for the solution of the condition equations in the method of least squares, a very ingenious computational procedure which immediately proved extremely useful, and which most assuredly would have great benefits for all geodesists, if it were published some day." (Euology by Commandant Benoit, 1922)

Laplace approximations for non-Gaussian observations

Quadratic posterior log-likelihood approximation

$$p(\boldsymbol{u} \mid \boldsymbol{\theta}) \sim \mathcal{N}(\boldsymbol{\mu}_u, \boldsymbol{Q}_u^{-1}), \quad \boldsymbol{y} \mid \boldsymbol{u}, \boldsymbol{\theta} \sim p(\boldsymbol{y} \mid \boldsymbol{u})$$

$$p_G(\boldsymbol{u} \mid \boldsymbol{y}, \boldsymbol{\theta}) \sim \mathcal{N}(\widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{Q}}^{-1})$$

$$\boldsymbol{0} = \nabla_{\boldsymbol{u}} \left\{\ln p(\boldsymbol{u} \mid \boldsymbol{\theta}) + \ln p(\boldsymbol{y} \mid \boldsymbol{u})\right\}\Big|_{\boldsymbol{u}=\widetilde{\boldsymbol{\mu}}}$$

$$\widetilde{\boldsymbol{Q}} = \boldsymbol{Q}_u - \nabla_{\boldsymbol{u}}^2 \ln p(\boldsymbol{y} \mid \boldsymbol{u})\Big|_{\boldsymbol{u}=\widetilde{\boldsymbol{\mu}}}$$

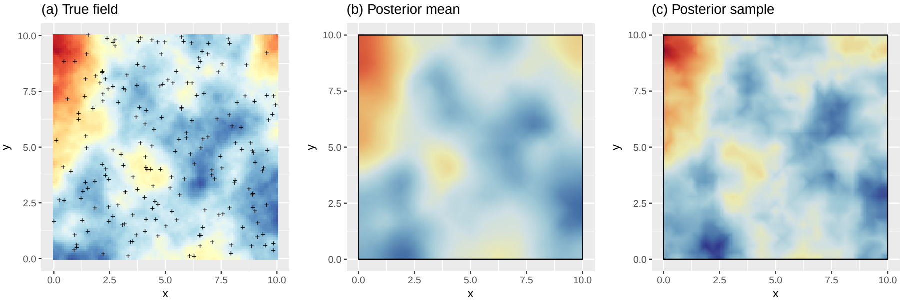
Direct Bayesian inference with INLA (`r-inla.org` & `inlabru.org`)

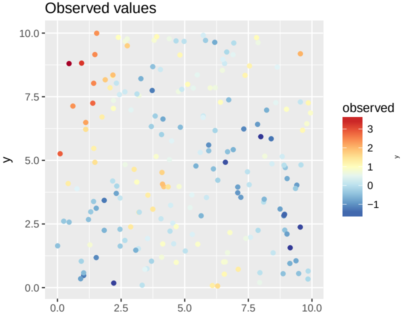
$$\widetilde{p}(\boldsymbol{\theta} \mid \boldsymbol{y}) \propto \left. \frac{p(\boldsymbol{\theta}) p(\boldsymbol{u} \mid \boldsymbol{\theta}) p(\boldsymbol{y} \mid \boldsymbol{u}, \boldsymbol{\theta})}{p_G(\boldsymbol{u} \mid \boldsymbol{y}, \boldsymbol{\theta})} \right|_{\boldsymbol{u}=\widetilde{\boldsymbol{\mu}}}$$

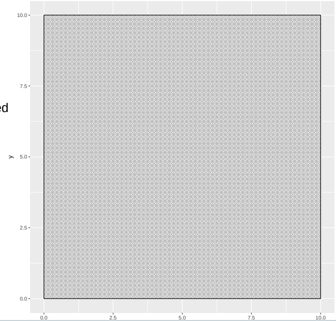
$$\widetilde{p}(\boldsymbol{u}_i \mid \boldsymbol{y}) \propto \int p_{GG}(\boldsymbol{u}_i \mid \boldsymbol{y}, \boldsymbol{\theta}) \widetilde{p}(\boldsymbol{\theta} \mid \boldsymbol{y}) \, d\boldsymbol{\theta}$$

The main practical limiting factors for the INLA method are the number of latent variables and the number model parameters.

Example: 2D georeferenced data







A multiscale model example

- A temporally slow, stochastic heat equation (non-separable)

$$\begin{aligned}\frac{\partial}{\partial t} z(\boldsymbol{s},t) + \gamma_z(1-\gamma_{\mathcal{E}}\nabla\cdot\nabla)z(\boldsymbol{s},t) &= \mathcal{E}(\boldsymbol{s},t)\\ (1-\gamma_{\mathcal{E}}\nabla\cdot\nabla)^{1/2}\mathcal{E}(\boldsymbol{s},t) &= \mathcal{W}_{\mathcal{E}}(\boldsymbol{s},t)\end{aligned}$$

- A temporally quick, spatially non-stationary SPDE/GMRF (separable) $\left(\frac{\partial}{\partial t} + \gamma_t\right)\left(\kappa(\boldsymbol{s})^2 - \nabla\cdot\nabla\right)\left(\tau(\boldsymbol{s})a(\boldsymbol{s},t)\right) = \mathcal{W}_a(\boldsymbol{s},t)$
- Measurements
  $y_i = a(\boldsymbol{s}_i,t_i) + z(\boldsymbol{s}_i,t_i) + \epsilon_i$, discretised into
  $\boldsymbol{y} = \boldsymbol{A}(\boldsymbol{a} + (\boldsymbol{B}\otimes\boldsymbol{I})\boldsymbol{z}) + \boldsymbol{\epsilon}$, $\boldsymbol{\epsilon}\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{Q}_\epsilon^{-1})$
  where $\boldsymbol{B}$ maps from long-term basis functions to short-term, and $\boldsymbol{A}$ maps from short-term basis functions to the observations.

The posterior precision can be formulated for $(a+z,z)|y$:

$$\boldsymbol{Q}_{(a+z,z)|y} = \begin{bmatrix} \boldsymbol{Q}_t\otimes\boldsymbol{Q}_a + \boldsymbol{A}^\top\boldsymbol{Q}_\epsilon\boldsymbol{A} & -\boldsymbol{Q}_t\boldsymbol{B}\otimes\boldsymbol{Q}_a \\ -\boldsymbol{B}^\top\boldsymbol{Q}_t\otimes\boldsymbol{Q}_a & \boldsymbol{Q}_z + \boldsymbol{B}^\top\boldsymbol{Q}_t\boldsymbol{B}\otimes\boldsymbol{Q}_a \end{bmatrix}$$

Locally isotropic non-stationary precision construction

Finite element construction of basis weight precision

Non-stationary SPDE:

$$\left(\kappa(\boldsymbol{s})^2 - \nabla \cdot \nabla\right)\left(\tau(\boldsymbol{s})u(\boldsymbol{s})\right) = \mathcal{W}(\boldsymbol{s})$$

The SPDE parameters are constructed via spatial covariates:

$$\log \tau(\boldsymbol{s}) = b_0^\tau(\boldsymbol{s}) + \sum_{j=1}^p b_j^\tau(\boldsymbol{s})\theta_j, \quad \log \kappa(\boldsymbol{s}) = b_0^\kappa(\boldsymbol{s}) + \sum_{j=1}^p b_j^\kappa(\boldsymbol{s})\theta_j$$

Finite element calculations give

$$\boldsymbol{T} = \mathrm{diag}(\tau(\boldsymbol{s}_i)), \quad \boldsymbol{K} = \mathrm{diag}(\kappa(\boldsymbol{s}_i))$$

$$C_{ii} = \int \psi_i(\boldsymbol{s})\,d\boldsymbol{s}, \quad G_{ij} = \int \nabla\psi_i(\boldsymbol{s}) \cdot \nabla\psi_j(\boldsymbol{s})\,d\boldsymbol{s}$$

$$\boldsymbol{Q} = \boldsymbol{T}\left(\boldsymbol{K}^2\boldsymbol{C}\boldsymbol{K}^2 + \boldsymbol{K}^2\boldsymbol{G} + \boldsymbol{G}\boldsymbol{K}^2 + \boldsymbol{G}\boldsymbol{C}^{-1}\boldsymbol{G}\right)\boldsymbol{T}$$

Combining this with an AR(1) discretisation of the temporal operator, we get $\boldsymbol{Q}_t \otimes \boldsymbol{Q}_a$.

GMRF precision for stochastic heat equation

$$\begin{aligned}
\boldsymbol{Q}_z &= \boldsymbol{M}_2^{(t)} \otimes \boldsymbol{M}_0^{(s)} + \boldsymbol{M}_1^{(t)} \otimes \boldsymbol{M}_1^{(s)} + \boldsymbol{M}_0^{(t)} \otimes \boldsymbol{M}_2^{(s)} \\
\boldsymbol{M}_0^{(s)} &= \boldsymbol{C} + \gamma_{\mathcal{E}} \boldsymbol{G} \\
\boldsymbol{M}_1^{(s)} &= \gamma_z (\boldsymbol{C} + \gamma_{\mathcal{E}} \boldsymbol{G}) \boldsymbol{C}^{-1} (\boldsymbol{C} + \gamma_{\mathcal{E}} \boldsymbol{G}) \\
\boldsymbol{M}_2^{(s)} &= \gamma_z^2 (\boldsymbol{C} + \gamma_{\mathcal{E}} \boldsymbol{G}) \boldsymbol{C}^{-1} (\boldsymbol{C} + \gamma_{\mathcal{E}} \boldsymbol{G}) \boldsymbol{C}^{-1} (\boldsymbol{C} + \gamma_{\mathcal{E}} \boldsymbol{G})
\end{aligned}$$

The precision structure can be used to formulate sampling as

$$\boldsymbol{Q}_z \boldsymbol{z} = \widetilde{\boldsymbol{L}}_z \boldsymbol{w}, \quad \boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$$

where $\widetilde{\boldsymbol{L}}_z$ is a pseudo Cholesky factor,

$$\begin{aligned}
\widetilde{\boldsymbol{L}}_z = \Big[&\left[\boldsymbol{L}_2^{(t)} \otimes \boldsymbol{L}_{\boldsymbol{C}}, \quad \boldsymbol{L}_1^{(t)} \otimes \boldsymbol{L}_{\boldsymbol{G}}, \quad \boldsymbol{L}_0^{(t)} \otimes \boldsymbol{G} \boldsymbol{L}_{\boldsymbol{C}}^{-\top} \right], \\
&\gamma_{\mathcal{E}}^{1/2} \left[\boldsymbol{L}_2^{(t)} \otimes \boldsymbol{L}_{\boldsymbol{G}}, \quad \boldsymbol{L}_1^{(t)} \otimes \boldsymbol{G} \boldsymbol{L}_{\boldsymbol{C}}^{-\top}, \quad \boldsymbol{L}_0^{(t)} \otimes \boldsymbol{G} \boldsymbol{C}^{-1} \boldsymbol{L}_{\boldsymbol{G}} \right] \Big]
\end{aligned}$$

Posterior calculations

Write $x = (a + z, z)$ for the full latent field.

$$\boldsymbol{Q}_{x|y} = \begin{bmatrix} \boldsymbol{Q}_t \otimes \boldsymbol{Q}_a + \boldsymbol{A}^\top \boldsymbol{Q}_\epsilon \boldsymbol{A} & -\boldsymbol{Q}_t \boldsymbol{B} \otimes \boldsymbol{Q}_a \\ -\boldsymbol{B}^\top \boldsymbol{Q}_t \otimes \boldsymbol{Q}_a & \boldsymbol{Q}_z + \boldsymbol{B}^\top \boldsymbol{Q}_t \boldsymbol{B} \otimes \boldsymbol{Q}_a \end{bmatrix}$$

can be pseudo-Cholesky-factorised:

$$\boldsymbol{Q}_{x|y} = \widetilde{\boldsymbol{L}}_{x|y} \widetilde{\boldsymbol{L}}_{x|y}^\top, \qquad \widetilde{\boldsymbol{L}}_{x|y} = \begin{bmatrix} \boldsymbol{L}_t \otimes \boldsymbol{L}_a & \boldsymbol{0} & \boldsymbol{A}^\top \boldsymbol{L}_\epsilon \\ -\boldsymbol{B}^\top \boldsymbol{L}_t \otimes \boldsymbol{L}_a & \widetilde{\boldsymbol{L}}_z & \boldsymbol{0} \end{bmatrix}$$

Posterior expectation, samples, and marginal variances (with $\widetilde{\boldsymbol{A}} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{0} \end{bmatrix}$):

$$\boldsymbol{Q}_{x|y}(\boldsymbol{\mu}_{x|y} - \boldsymbol{\mu}_x) = \widetilde{\boldsymbol{A}}^\top \boldsymbol{Q}_\epsilon (\boldsymbol{y} - \widetilde{\boldsymbol{A}} \boldsymbol{\mu}_x),$$

$$\boldsymbol{Q}_{x|y}(\boldsymbol{x} - \boldsymbol{\mu}_{x|y}) = \widetilde{\boldsymbol{L}}_{x|y} \boldsymbol{w}, \quad \boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}), \quad \text{or}$$

$$\boldsymbol{Q}_{x|y}(\boldsymbol{x} - \boldsymbol{\mu}_x) = \widetilde{\boldsymbol{A}}^\top \boldsymbol{Q}_\epsilon (\boldsymbol{y} - \widetilde{\boldsymbol{A}} \boldsymbol{\mu}_x) + \widetilde{\boldsymbol{L}}_{x|y} \boldsymbol{w}, \quad \boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}),$$

$$\mathsf{Var}(x_i | \boldsymbol{y}) = \left(\boldsymbol{Q}_{x|y}^{-1} \right)_{ii} \qquad \text{(Ouch! Don't do this!! Use Takahashi!!!)}$$

Overlapping block preconditioning

For ease of notation, write the two-level model posterior precision as

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{Q}_0 + \boldsymbol{A}^\top \boldsymbol{Q}_\epsilon \boldsymbol{A} & -\boldsymbol{Q}_0 \boldsymbol{B} \\ -\boldsymbol{B}^\top \boldsymbol{Q}_0 & \boldsymbol{Q}_1 + \boldsymbol{B}^\top \boldsymbol{Q}_0 \boldsymbol{B} \end{bmatrix}$$

Overlapping block preconditioning

Let $\boldsymbol{D}_k^\top$ be a restriction matrix to subdomain Ω_k, and let $\boldsymbol{W}_k$ be a diagonal weight matrix. Then a useful additive Schwartz preconditioner is

$$\boldsymbol{M}^{-1}\boldsymbol{x} = \sum_{k=1}^{K} \boldsymbol{W}_k \boldsymbol{D}_k (\boldsymbol{D}_k^\top \boldsymbol{Q} \boldsymbol{D}_k)^{-1} \boldsymbol{D}_k^\top \boldsymbol{W}_k \boldsymbol{x}$$

The domain overlap may need to be substantial:

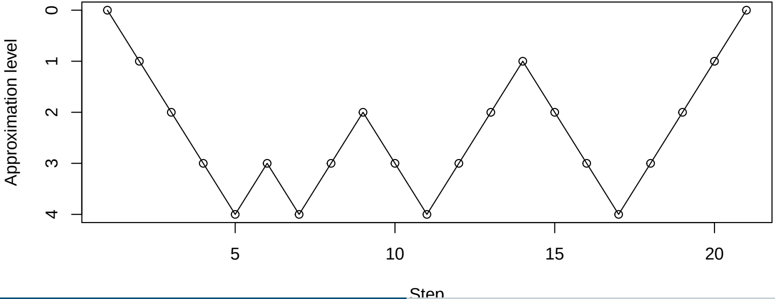
- Typical off-the-shelf preconditioning is aimed at at most 2nd order operators (Laplacian)
- In the example model, the spatial precision operator order is 6
- In a hierarchical triangle subdivision mesh, neighbouring hexagonal macro-domains overlap by 2 macro-triangles

Multigrid

Multigrid

Let $\boldsymbol{B}_c^\top$ be a projection matrix to a coarse approximative model. Then a basic multigrid step for $\boldsymbol{Q}\boldsymbol{x} = \boldsymbol{b}$ is

1. Apply high frequency preconditioner to get $\widehat{\boldsymbol{x}}_0$, let $\boldsymbol{r}_0 = \boldsymbol{b} - \boldsymbol{Q}\widehat{\boldsymbol{x}}_0$
2. Project the problem to the coarser model: $\boldsymbol{Q}_c = \boldsymbol{B}_c^\top \boldsymbol{Q}\boldsymbol{B}_c$, $\boldsymbol{r}_c = \boldsymbol{B}_c^\top \boldsymbol{r}_0$
3. Apply multigrid to $\boldsymbol{Q}_c\boldsymbol{x}_c = \boldsymbol{r}_c$
4. Update the solution: $\widehat{\boldsymbol{x}}_1 = \widehat{\boldsymbol{x}}_0 + \boldsymbol{B}_c\widehat{\boldsymbol{x}}_c$
5. Apply high frequency preconditioner to get $\widehat{\boldsymbol{x}}_2$



The hierarchy of scales and preconditioning ($\boldsymbol{x}_0 = \boldsymbol{B}\boldsymbol{x}_1 +$ fine scale variability):

Multiscale Schur complement approximation

Solving $\boldsymbol{Q}_{\boldsymbol{x}|\boldsymbol{y}}\boldsymbol{x} = \boldsymbol{b}$ can be formulated using two solves with the upper (fine) block $\boldsymbol{Q}_0 + \boldsymbol{A}^\top\boldsymbol{Q}_\epsilon\boldsymbol{A}$, and one solve with the *Schur complement*

$$\boldsymbol{Q}_1 + \boldsymbol{B}^\top\boldsymbol{Q}_0\boldsymbol{B} - \boldsymbol{B}^\top\boldsymbol{Q}_0\left(\boldsymbol{Q}_0 + \boldsymbol{A}^\top\boldsymbol{Q}_\epsilon\boldsymbol{A}\right)^{-1}\boldsymbol{Q}_0$$

By mapping the fine scale model onto the coarse basis used for the coarse model, we get an *approximate* (and sparse) Schur solve via

$$\begin{bmatrix} \widetilde{\boldsymbol{Q}}_B + \boldsymbol{B}^\top\boldsymbol{A}^\top\boldsymbol{Q}_\epsilon\boldsymbol{A}\boldsymbol{B} & -\widetilde{\boldsymbol{Q}}_B \\ -\widetilde{\boldsymbol{Q}}_B & \boldsymbol{Q}_1 + \widetilde{\boldsymbol{Q}}_B \end{bmatrix} \begin{bmatrix} \text{ignored} \\ \boldsymbol{x}_1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} \\ \widetilde{\boldsymbol{b}} \end{bmatrix}$$

where $\widetilde{\boldsymbol{Q}}_B = \boldsymbol{B}^\top\boldsymbol{Q}_0\boldsymbol{B}$.
The block matrix can be interpreted as the precision of a bivariate field on a common, coarse spatio-temporal scale, and the same technique applied to this system, with $\boldsymbol{x}_{1,1} = \boldsymbol{B}_{1|2}\boldsymbol{x}_{1,2} +$ finer scale variability.

For realistic problems we need to combine all three techniques.

Variance calculations

Sparse partial inverse: Takahashi recursions postprocesses Cholesky

Takahashi recursions compute $\boldsymbol{S}$ such that $\boldsymbol{S}_{ij} = (\boldsymbol{Q}^{-1})_{ij}$ for all $Q_{ij} \neq 0$.
Postprocessing of the (sparse) Cholesky factor.

Basic Rao-Blackwellisation of sample estimators

Let $\boldsymbol{x}^{(j)}$ be samples from a Gaussian posterior and let $\boldsymbol{a}^\top \boldsymbol{x}$ be a linear combination of interest. Then, for any subdomain $\Omega_k \subset \Omega$,

$$\mathsf{E}(\boldsymbol{a}^\top \boldsymbol{x}) = \mathsf{E}\left[\mathsf{E}(\boldsymbol{a}^\top \boldsymbol{x} \mid \boldsymbol{x}_{\Omega_k^*})\right] \approx \frac{1}{J}\sum_{j=1}^{J} \mathsf{E}(\boldsymbol{a}^\top \boldsymbol{x} \mid \boldsymbol{x}_{\Omega_k^*}^{(j)})$$

$$\begin{aligned}\mathsf{Var}(\boldsymbol{a}^\top \boldsymbol{x}) &= \mathsf{E}\left[\mathsf{Var}(\boldsymbol{a}^\top \boldsymbol{x} \mid \boldsymbol{x}_{\Omega_k^*})\right] + \mathsf{Var}\left[\mathsf{E}(\boldsymbol{a}^\top \boldsymbol{x} \mid \boldsymbol{x}_{\Omega_k^*})\right] \\ &\approx \mathsf{Var}(\boldsymbol{a}^\top \boldsymbol{x} \mid \boldsymbol{x}_{\Omega_k^*}^{(j)}) + \frac{1}{J}\sum_{j=1}^{J}\left[\mathsf{E}(\boldsymbol{a}^\top \boldsymbol{x} \mid \boldsymbol{x}_{\Omega_k^*}^{(j)}) - \mathsf{E}(\boldsymbol{a}^\top \boldsymbol{x})\right]^2\end{aligned}$$

Efficient if $\boldsymbol{a}\boldsymbol{a}^\top$ sparsity matches $\boldsymbol{S}_k$ on each subdomain:

$$\mathsf{Var}(\boldsymbol{a}^\top \boldsymbol{x}) = \boldsymbol{a}^\top \boldsymbol{Q}^{-1} \boldsymbol{a} = \operatorname{tr}(\boldsymbol{Q}^{-1}\boldsymbol{a}\boldsymbol{a}^\top) = \sum_{ij}\left[\boldsymbol{Q}^{-1} \odot \boldsymbol{a}\boldsymbol{a}^\top\right]_{ij} = \sum_{ij}\left[\boldsymbol{S} \odot \boldsymbol{a}\boldsymbol{a}^\top\right]_{ij}$$

Summary so far

- Translation between GRF/SPDE/GMRF; they are all the same Gaussian process
- Precision matrices are sparse for Markov models
- Local basis functions $+$ projection $\longrightarrow$ Success
- Small problems solved using Cholesky factorisation
- Large problems need iterative methods with preconditioners:
  - Overlapping blocks (high order operators)
  - Multigrid (mostly within-field multiscale)
  - Approximate Schur complements (between-fields multiscale)
- Sparse partial inversion for (conditional) (co)variances
- The INLA method for Bayesian parameter estimation works for small enough problems
- Next: What about priors, and likelihoods for large problems?

Not covered:

- Domain boundary effects
- Excursion sets

References (1/3)

- F. Lindgren, H. Rue and J. Lindström (2011), *An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach (with discussion)*, JRSSB, 73(4):423–498. Code available in `R-INLA`, see http://r-inla.org/
- Preliminary draft book chapter on GMRF computation basics: http://www.maths.ed.ac.uk/~flindgre/tmp/gmrf.pdf
- R. Ingebrigtsen, F. Lindgren, I. Steinsland (2014), *Spatial models with explanatory variables in the dependence structure*, Spatial Statistics, 8:20–38. http://dx.doi.org/10.1016/j.spasta.2013.06.002
- G-A. Fuglstad, F. Lindgren, D. Simpson, H. Rue (2015), *Exploring a new class of non-stationary spatial Gaussian random fields with varying local anisotropy*, Statistica Sinica, 25:115–133. http://arxiv.org/abs/1304.6949
- D. Bolin, F. Lindgren (2015), *Excursion and contour uncertainty regions for latent Gaussian models*, JRSSB, 77(1):85–106.http://arxiv.org/abs/1211.3946 CRAN package: `excursions` http://arxiv.org/abs/1211.3946

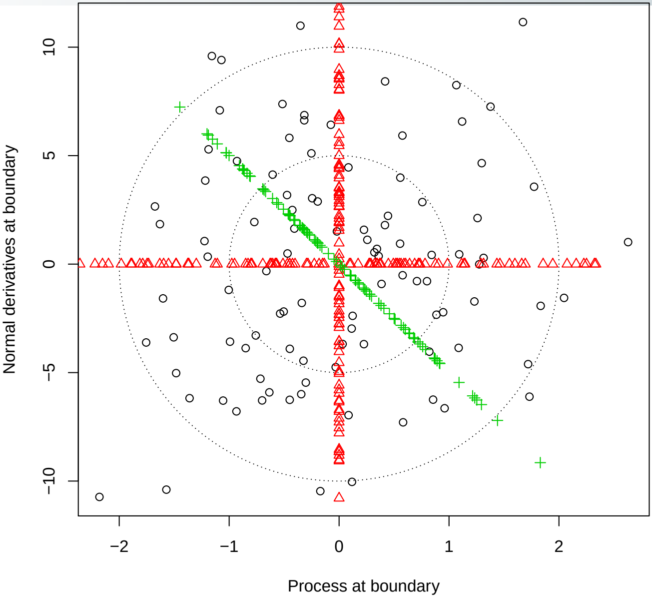
References (2/3)

- J. Besag (1974),
  Spatial interaction and the statistical analysis of latice systems (with discussion), JRSSB, 36(2):192–225.
- J. Besag and C. Kooperberg (1995),
  On conditional and intrinsic autoregressions, Biometrika, 82(4):733–746.
- J. Besag (1981),
  On a system of two-dimensional recurrence equations, JRSSB, 43(3):302–309.
- J. Besag and D. Mondal (2005),
  First-order intrinsic autoregressions and the de Wij process, Biometrika, 92(4):909–920.
- Y. A. Rozanov (1977),
  Markov random fields and stochastic partial differential equations, Mathematics of the USSR – Sbornik, 32(4):515–534
- Y. A. Rozanov (1977),
  On the paper "Markov random fields and stochastic partial differential equations", Mathematics of the USSR – Sbornik, 35(1):157–164
- Y. A. Rozanov (1982, Russian original 1980),
  Markov Random Fields, Springer Verlag, New York.

References (3/3)

- Y. A. Rozanov (2010, reprint of 1st ed. 1998), *Random fields and stochastic partial differential equations*, Springer Verlag, Mathematics and its Applications.
- R. J. Adler and J. Taylor (2007), *Random Fields and Geometry*, Springer Monographs in Mathematics
- G. Lindgren (2012), *Stationary Stochastic Processes: Theory and applications*, Chapman & Hall/CRC, Texts in Statistical Science.
- P. Whittle (1954), *On stationary processes in the plane*, Biometrika, 41(3/4):434–449.
- P. Whittle (1963), *Stochastic processes in several dimensions*, Bull. Inst. Internat. Statist., 40:974–994.
- B. Matérn (1960), *Spatial variation. Stochastic models and their application to some problems in forest surveys and other sampling investigations.* (Swedish: Stokastiska modeller och deras tillämpning på några problem i skogstaxering och andra samplingundersökningar.), Stockholm University, PhD thesis in Mathematical Statistics. You may be lucky and find someone who has a copy!

All deterministic boundary conditions are ‘inappropriate’



Stationary stochastic boundary adjustment (current work)

Recall the Matérn generating SPDE

$$(\kappa^2 - \nabla \cdot \nabla)^{\alpha/2} u(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s})$$

RKHS inner product/precision operator for Matérn fields on $\mathbb{R}^d$:

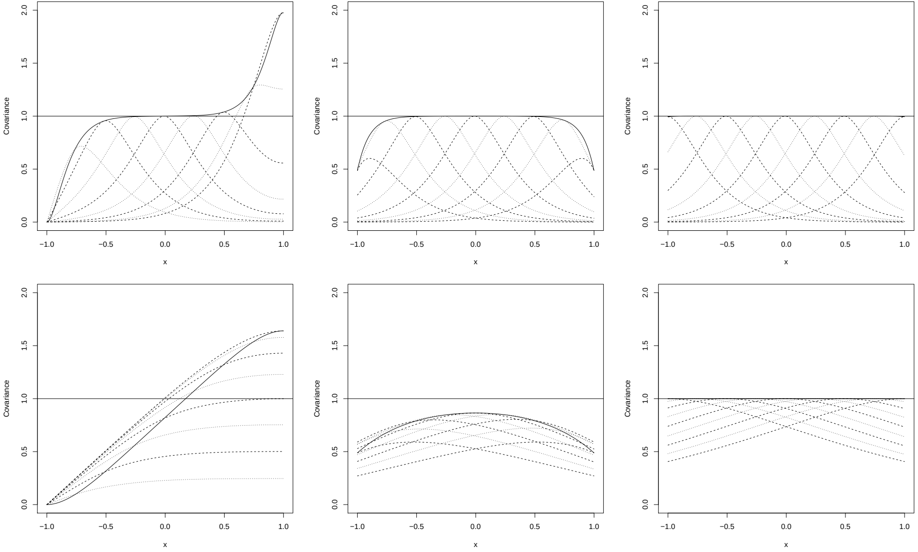
$$\langle f, \mathcal{Q}_{\mathbb{R}^d} g \rangle_D = \sum_{k=0}^{\alpha} \binom{\alpha}{k} \kappa^{2\alpha-2k} \left\langle \nabla^k f, \nabla^k g \right\rangle_D$$

Boundary adjusted precision operator on a compact subdomain, where $\mathcal{P}$ is a conditional expectation operator:

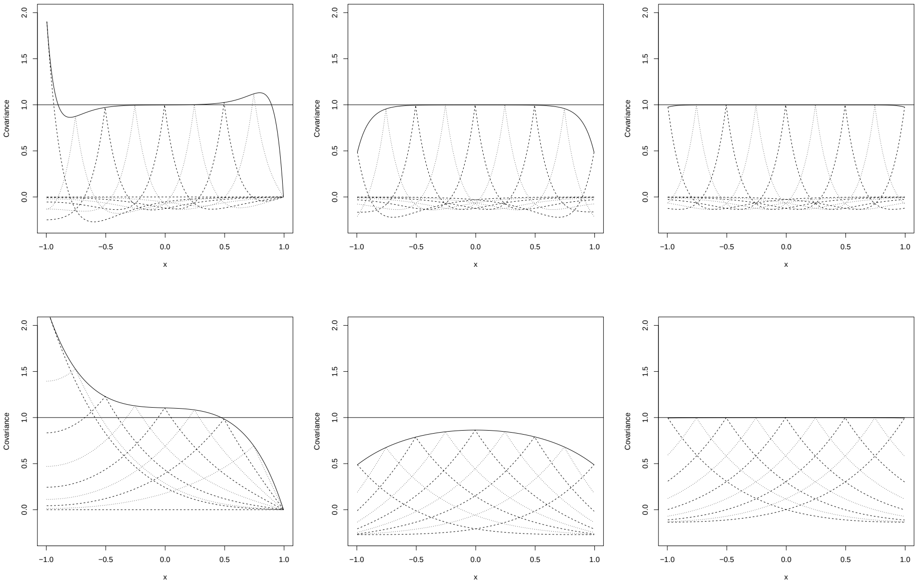
$$\begin{aligned} \langle f, \mathcal{Q}_D g \rangle_D &= \langle f, \mathcal{Q}_{\mathbb{R}^2} g \rangle_D - \langle \mathcal{P} f, \mathcal{Q}_{\mathbb{R}^2} \mathcal{P} g \rangle_D + \langle f, \mathcal{Q}_{\partial D} g \rangle_{\partial D} \\ &= \langle f - \mathcal{P} f, \mathcal{Q}_{\mathbb{R}^2} (g - \mathcal{P} g) \rangle_D + \langle f, \mathcal{Q}_{\partial D} g \rangle_{\partial D}, \end{aligned}$$

where the boundary precision operator may involve normal derivatives of f and g.

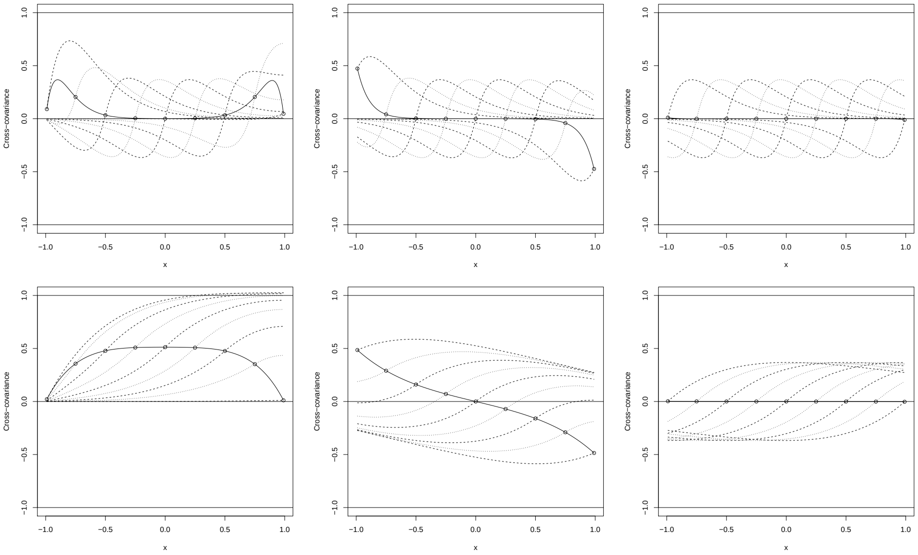
Covariances (D&N, Robin, Stoch) for $\kappa = 5$ and 1



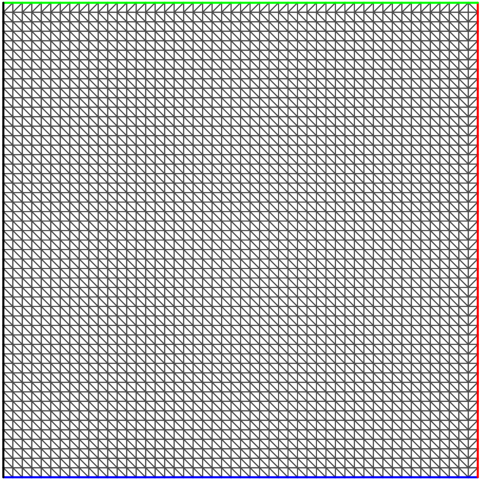
Derivative covariances (D&N, Robin, Stoch) for $\kappa = 5$ and 1



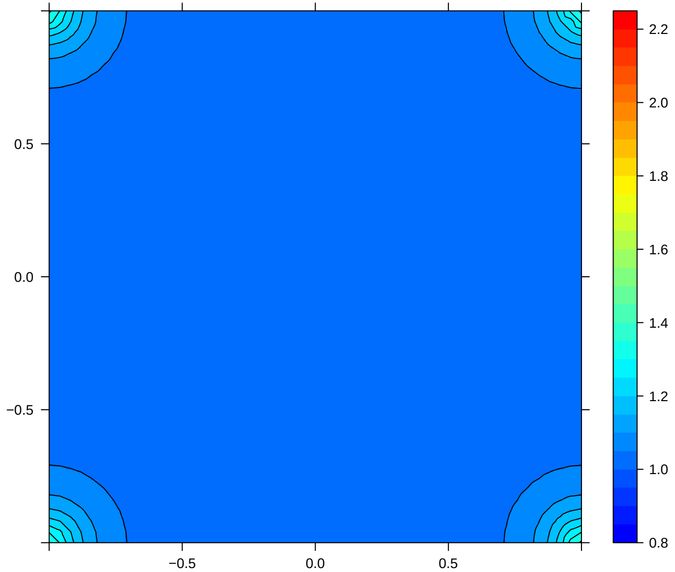
Process-derivative cross-covariances (D&N, Robin, Stoch)



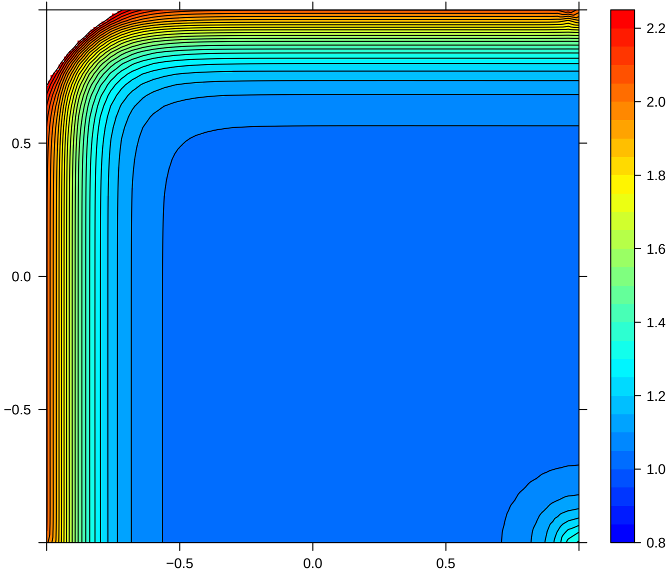
Square domain, basis triangulation



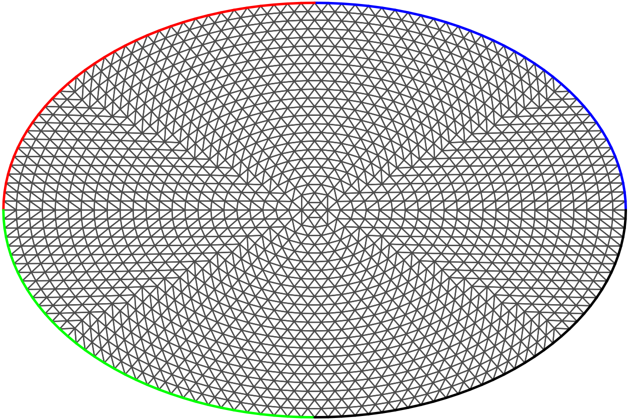
Square domain, stochastic boundary (variances)



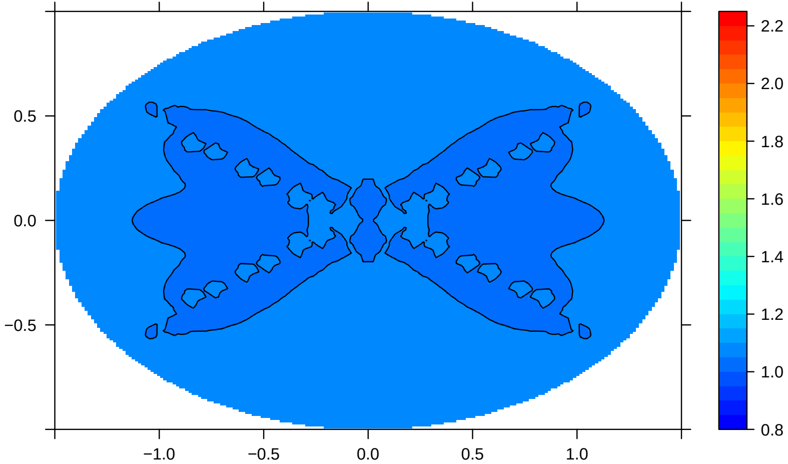
Square domain, mixed boundary (variances)



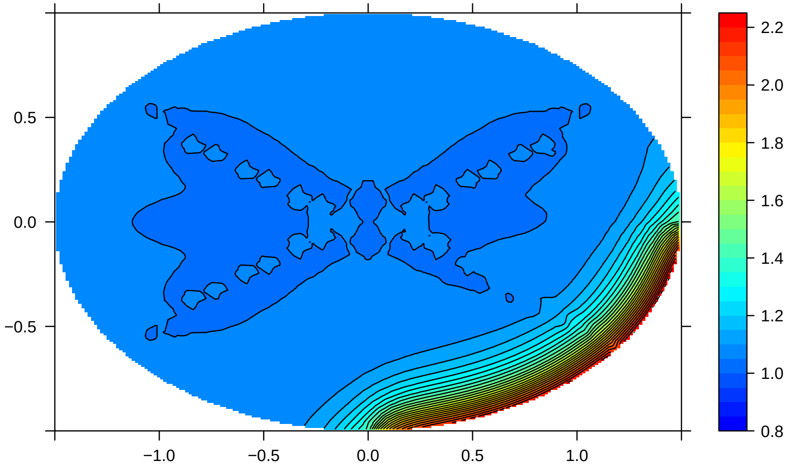
Elliptical domain, basis triangulation



Elliptical domain, stochastic boundary (variances)



Elliptical domain, mixed boundary (variances)



Excursion sets for random fields

Excursion sets

Let $u(\mathbf{s})$, $\mathbf{s} \in D$ be a random process. The positive and negative level u excursion sets with probability $1-\alpha$ are

$$E^+_{u,\alpha}(x) = \operatorname*{argmax}_D\{|D| : \mathsf{P}(D \subseteq A^+_u(x)) \geq 1-\alpha\}.$$
$$E^-_{u,\alpha}(x) = \operatorname*{argmax}_D\{|D| : \mathsf{P}(D \subseteq A^-_u(x)) \geq 1-\alpha\}.$$

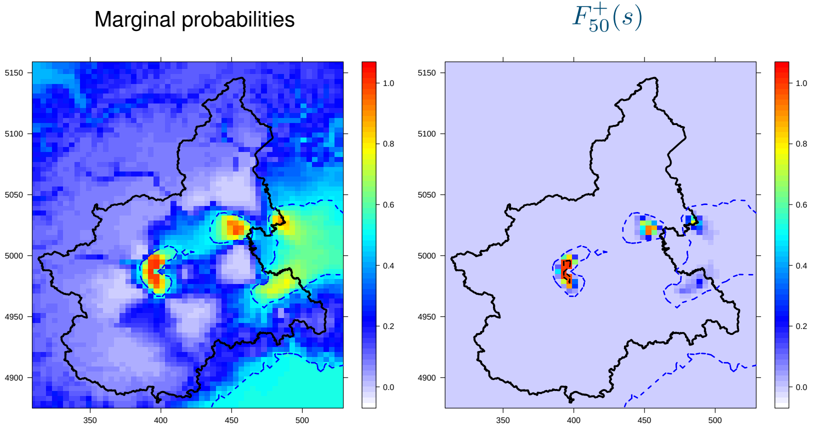
These are sets with high probability for excursions *in the entire set*.

Excursion functions

The positive and negative u excursion functions are given by

$$F^+_u(s) = \sup\{1-\alpha; s \in E^+_{u,\alpha}\},$$
$$F^-_u(s) = \sup\{1-\alpha; s \in E^-_{u,\alpha}\}.$$

PM_{10} exceedances in Piemonte, January 30, 2006



Model estimated with INLA, result passed onward to `excursions()`, evaluating high dimensional GMRF probabilities and finding credible regions. Latest version has user friendly options for continuous domain interpretations.