Stochastic PDEs and Markov random fields with ecological applications

Finn Lindgren

Spatially-varying Stochastic Differential Equations with Applications to the Biological Sciences
OSU, Columbus, Ohio, 2015
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 $D \times T$, $T \subset \mathbb{R}$.
- Random field $u(s)$, $s \in D$, or $u(s, t)$, $(s, t) \in D \times T$.
- Observations $y_i$. In the simplest setting, $y_i = u(s_i) + \epsilon_i$, 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 $y$. 

Finn Lindgren - f.lindgren@bath.ac.uk
Stochastic PDEs and Markov random fields with ecological applications
The Matérn covariance family on \( \mathbb{R}^d \)

\[
\text{Cov}(u(0), u(s)) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} (\kappa \|s\|)^\nu K_\nu(\kappa \|s\|)
\]

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

\[
(\kappa^2 - \nabla \cdot \nabla)^{\alpha/2} u(s) = \mathcal{W}(s), \quad \alpha = \nu + d/2
\]

\( \mathcal{W}(\cdot) \) 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?
White noise on continuous domains, and Brownian motion

Let \( \mathcal{W}^*(f) \) be a Gaussian random measure for functions \( f \in L_2(D) \), such that

\[
E(\mathcal{W}^*(1_A)) = 0, \quad \text{Cov}(\mathcal{W}^*(1_A), \mathcal{W}^*(1_B)) = |A \cap B|_{\text{Leb}(D)}.
\]

for all Lebesgue-measurable \( A, B \subseteq D \).

*Gaussian white noise* \( \mathcal{W}(s) \) is then formally defined by *defining*

\[
\langle f, \mathcal{W} \rangle_D \equiv \mathcal{W}^*(f), \text{ so that }
\]

\[
E(\langle f, \mathcal{W} \rangle_D) = 0,
\]

\[
\text{Cov}(\langle f, \mathcal{W} \rangle_D, \langle g, \mathcal{W} \rangle_D) = \langle f, g \rangle_D \equiv \int_D f(s)g(s) \, ds,
\]

\[
\langle f, \mathcal{W} \rangle_D \sim \mathcal{N}(0, \langle f, f \rangle_D)
\]

The Brownian motion \( \mathcal{W}_t \) on \( \mathbb{R} \) and the associated stochastic differential

\[
d\mathcal{W}_t \text{ from Radu’s lectures are related to } \mathcal{W}^* \text{ and } \mathcal{W} \text{ via}
\]

\[
\mathcal{W}_t = \mathcal{W}^*(1_{[0,t]})
\]

\[
d\mathcal{W}_t = \mathcal{W}(t) \, dt
\]
A **Gaussian random field** $u : D \mapsto \mathbb{R}$ is defined via

$$
E(u(s)) = m(s),
$$
$$
\text{Cov}(u(s), u(s')) = K(s, s'),
$$

$$
[u(s_i), i = 1, \ldots, n] \sim \mathcal{N}(m = [m(s_i), i = 1, \ldots, n],
\Sigma = [K(s_i, s_j), i, j = 1, \ldots, n])
$$

for all finite location sets $\{s_1, \ldots, s_n\}$, and $K(\cdot, \cdot)$ symmetric positive definite.

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

$$
E(\langle f, u \rangle_D) = \langle f, m \rangle_D = \int_D f(s)m(s) \, ds,
$$
$$
\text{Cov}(\langle f, u \rangle_D, \langle g, u \rangle_D) = \langle f, \mathcal{R}g \rangle_D = \iint_{D \times D} f(s)K(s, s')g(s') \, ds \, ds',
$$

$$
\langle f, u \rangle_D \sim \mathcal{N}(\langle f, m \rangle_D, \langle f, \mathcal{R}f \rangle_D)
$$

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

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

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

**Independent Gaussian noise**

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

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

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

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

and the generalised calculus is not applicable.
Spectral properties

Bochner’s theorem on $\mathbb{R}^d$

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

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

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

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

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

White noise on $\mathbb{R}^d$ has spectral density $S_W(\omega) = 1/(2\pi)^d$. 
Let $D = \mathbb{R}^d$, and $\hat{f}(\omega) = (\mathcal{F}f)(\omega) = \int_{\mathbb{R}^d} \exp(i s \cdot \omega) f(s) \, ds$. Very informally, $S_u(\omega) = \mathbb{E}(|\hat{u}(\omega)|^2)$.

Differential operators can also be interpreted spectrally:

$$
\begin{array}{c|cccc}
\mathcal{L}f & f & \nabla f & -\nabla \cdot \nabla f & \mathcal{L}^{\alpha/2} f \\
\mathcal{L}\hat{f} \equiv \mathcal{F}(\mathcal{L}f) & \hat{f} & i\omega\hat{f} & ||\omega||^2\hat{f} & |\mathcal{L}|^{\alpha/2}\hat{f} \\
\end{array}
$$

The rightmost column is a definition of a fractional operator!

For the Whittle-Matérn SPDE, informally,

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

$$(\kappa^2 + ||\omega||^2)^{\alpha/2} \hat{u}(\omega) = \hat{\mathcal{W}}(\omega)$$

$$\mathbb{E}((\kappa^2 + ||\omega||^2)^{\alpha/2} \hat{u}(\omega))^2) = \mathbb{E}(|\hat{\mathcal{W}}(\omega)|^2)$$

$$(\kappa^2 + ||\omega||^2)^\alpha S_u(\omega) = S_{\mathcal{W}}(\omega)$$

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

Whittle (1954, 1963) showed that $(\mathcal{F}S_u(\cdot))(s' - s)$ 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 $u(s, t), (s, t) \in \mathbb{R}^d \times \mathbb{R}$, and $S_u(k, \omega), (k, \omega) \in \mathbb{R}^d \times \mathbb{R}$.

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

$$
\left\{ \gamma \frac{\partial}{\partial t} + \kappa^2 - \nabla_s \cdot \nabla_s \right\} u(s) = E(s, t),
$$

$$(\kappa^2 - \nabla_s \cdot \nabla_s)^{\alpha/2} E(s, t) = \mathcal{W}(s, t).$$

The Fourier domain version is

$$
\left\{ i \gamma \omega + \kappa^2 + \|k\|^2 \right\} \hat{u}(k, \omega) = \hat{E}(k, \omega),
$$

$$(\kappa^2 + \|k\|^2)^{\alpha/2} \hat{E}(k, \omega) = \hat{\mathcal{W}}(k, \omega),$$

and

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

How differentiable are the realisations?
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

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

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

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 $$K(s, s') = \frac{1}{4\pi \gamma} \frac{1}{2\pi} K_0(\kappa \|s' - s\|).$$
To help understand the temporal properties, take the Fourier transform in only the spatial directions:

\[
\left\{ \gamma \frac{\partial}{\partial t} + \kappa^2 + \|k\|^2 \right\} \tilde{u}(k, t) = \frac{\tilde{W}(k, t)}{(\kappa^2 + \|k\|^2)^{\alpha/2}},
\]

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

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

There is one more property we need to understand: Markov in space.
Markov properties on discrete and continuous domains

Directed Markov on discrete time
\[ p(x_t | x_{<t}) = p(x_t | x_{t-1}) \]

2nd order Markov

Undirected Markov
\[ p(x_t | x_{\neq t}) = p(x_t | x_{t-1}, x_{t+1}) \]

For all sep. sets \( S \):
\[ p(x_A, x_B | x_S) = p(x_A | x_S) p(x_B | x_S) \]
Markov in time, filtration $\sigma$-algebras:

\[
\begin{align*}
    a & \in \mathcal{F}^{\sigma}_{(\infty,t]} \equiv \sigma(u(s), s \leq t) \\
    b & \in \mathcal{F}^{\sigma}_{[t,\infty)} \equiv \sigma(u(s), s \geq t)
\end{align*}
\]

\[
P(a \cap b \mid u(t)) = P(a \mid u(t))P(b \mid u(t))
\]

Markov for regular random fields on spatial and spatio-temporal domains: $A, B, S \subset D$, such that $S$ separates $A$ and $B$.

\[
\mathcal{F}^{\sigma}_S \equiv \sigma(u(s), s \in S),
\]

\[
a \in \mathcal{F}^{\sigma}_A, \quad b \in \mathcal{F}^{\sigma}_B,
\]

\[
P(a \cap b \mid \mathcal{F}^{\sigma}_S) = P(a \mid \mathcal{F}^{\sigma}_S)P(b \mid \mathcal{F}^{\sigma}_S)
\]

Markov for generalised random fields on spatial and spatio-temporal domains:

\[
\mathcal{F}^{\sigma}_S \equiv \sigma(\langle f, u \rangle_S, f \in H_R(S)),
\]

\[
a \in \mathcal{F}^{\sigma}_A, \quad b \in \mathcal{F}^{\sigma}_B,
\]

\[
P(a \cap b \mid \mathcal{F}^{\sigma}_S) = P(a \mid \mathcal{F}^{\sigma}_S)P(b \mid \mathcal{F}^{\sigma}_S)
\]
Markov in space

Markov properties

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

Solutions to

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

are Markov when $\alpha$ is an integer.

(Generally, when the reciprocal of the spectral density is a polynomial, Rozanov, 1977)

Discrete representations ($Q = \Sigma^{-1}$):

$Q_{AB} = 0$

$Q_{A|S,B} = Q_{AA}$

$\mu_{A|S,B} = \mu_A - Q_{AA}^{-1} Q_{AS}(u_S - \mu_S)$

Weak Markov: $\varepsilon > 0$

Strong/strict Markov: $\varepsilon = 0$
# Markov random fields

**Rozanov (1977)**

\[ u(s) \] is stationary Markov \( \iff S_u(k) \propto P(k)^{-1} \)

where \( P(k) \geq 0 \) is a symmetric polynomial

Matérn/Whittle:

\[ S_u(k) \propto (\kappa^2 + \|k\|^2)^{-\alpha} \]

<table>
<thead>
<tr>
<th>GMRF</th>
<th>Covariance (( \mathbb{R}^2 ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAR(1)</td>
<td>( \propto \kappa |u| K_1(\kappa |u|) )  Whittle (1954)</td>
</tr>
<tr>
<td>CAR(2)</td>
<td>( \propto \kappa |u| K_1(\kappa |u|) )  Whittle (1954)</td>
</tr>
<tr>
<td>CAR(1)</td>
<td>( \frac{1}{2\pi} K_0(\kappa |u|) )  Besag (1981)</td>
</tr>
<tr>
<td>ICAR(1)</td>
<td>( -\frac{1}{2\pi} \log(|u|) )  Besag &amp; Mondal (2005)</td>
</tr>
</tbody>
</table>

On lattices, classical CAR \( \rightarrow \) Matérn models (limits of).

Can extend to non-stationary SPDE models, on irregular triangulations.
Hilbert space approximation

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

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

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

A finite basis expansion

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

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 $\nabla f$ or $-\nabla \cdot \nabla g$ are as generalised as white noise.

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

$$\left\langle \psi_i, (\kappa^2 - \nabla \cdot \nabla) \sum_j \psi_j u_j \right\rangle_D = \sum_j \left\{ \kappa^2 \langle \psi_i, \psi_j \rangle_D + \langle \nabla \psi_i, \nabla \psi_j \rangle_D \right\} u_j$$

$$= (\kappa^2 C + G)u$$

The covariance for the RHS of the SPDE is

$$\text{Cov}(\langle \psi_i, \mathcal{W} \rangle_D, \langle \psi_j, \mathcal{W} \rangle_D) = [\langle \psi_i, \psi_j \rangle_D] = C$$

by the definition of $\mathcal{W}$. 
We seek $u \sim \mathcal{N}(0, \Sigma)$ such that $\text{Var}\{(\kappa^2 \mathbf{C} + \mathbf{G})u\} = \mathbf{C}$:

$$(\kappa^2 \mathbf{C} + \mathbf{G}) \Sigma (\kappa^2 \mathbf{C} + \mathbf{G}) = \mathbf{C}$$

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

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

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

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

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

- $Q_1 = (\kappa^2 \mathbf{C} + \mathbf{G})$
- $Q_2 = (\kappa^2 \mathbf{C} + \mathbf{G}) \mathbf{C}^{-1} (\kappa^2 \mathbf{C} + \mathbf{G}) = \kappa^4 \mathbf{C} + 2\kappa^2 \mathbf{G} + \mathbf{G} \mathbf{C}^{-1} \mathbf{G}$
- $Q_\alpha = (\kappa^2 \mathbf{C} + \mathbf{G}) \mathbf{C}^{-1} Q_{\alpha-2} \mathbf{C}^{-1} (\kappa^2 \mathbf{C} + \mathbf{G})$
- Any $\alpha \geq 0$: $Q_\alpha = C_{1/2}^{1/2} \left\{ C_{-1/2}^{-1/2} (\kappa^2 \mathbf{C} + \mathbf{G}) C_{-1/2}^{-1/2} \right\}^\alpha C_{1/2}^{1/2}$

(non-sparse for non-integer $\alpha$)

Finn Lindgren - f.lindgren@bath.ac.uk
Stochastic PDEs and Markov random fields with ecological applications
### Basis definitions

<table>
<thead>
<tr>
<th>Basis definitions</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finite basis set (k = 1, \ldots, n)</td>
<td>((\kappa^2 - \nabla \cdot \nabla)^{-\alpha} e_{\kappa, k}(s) = \lambda_{\kappa, k} e_{\kappa, k}(s))</td>
</tr>
<tr>
<td>Karhunen-Loève</td>
<td>(-\nabla \cdot \nabla e_k(s) = \lambda_k e_k(s))</td>
</tr>
<tr>
<td>Fourier</td>
<td>((\kappa^2 - \nabla \cdot \nabla)^{\alpha/2} g_{\kappa}(s) = \delta(s))</td>
</tr>
<tr>
<td>Convolution</td>
<td>(\psi_k(s))</td>
</tr>
<tr>
<td>General</td>
<td></td>
</tr>
</tbody>
</table>

### Field representations

<table>
<thead>
<tr>
<th>Field (u(s))</th>
<th>Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Karhunen-Loève</td>
<td>(z_k \sim \mathcal{N}(0, \lambda_{\kappa, k}))</td>
</tr>
<tr>
<td>Fourier</td>
<td>(z_k \sim \mathcal{N}(0, (\kappa^2 + \lambda_k)^{-\alpha}))</td>
</tr>
<tr>
<td>Convolution</td>
<td>(z_k \sim \mathcal{N}(0,</td>
</tr>
<tr>
<td>General</td>
<td>(u \sim \mathcal{N}(0, Q_{\kappa}^{-1}))</td>
</tr>
</tbody>
</table>

Note: Harmonic basis functions (as in the Fourier approach) give a diagonal \(Q_{\kappa}\), but lead to dense posterior precision matrices.
SPDE/GMRF realisations and non-stationary models

\[(\kappa^2 - \nabla \cdot \nabla) u(s) = \mathcal{W}(s), \quad s \in D\]
SPDE/GMRF realisations and non-stationary models

\[(\kappa^2 \exp(i\theta) - \nabla \cdot \nabla) u(s) = \mathcal{W}(s), \ s \in D, \ \text{Re}(u) \text{ independent of } \text{Im}(u)\]
Beyond Matérn: Non-stationary field

\[(\kappa(s)^2 - \nabla \cdot \nabla) u(s) = \kappa(s) \mathcal{W}(s), \quad s \in \Omega\]
SPDE/GMRF realisations and non-stationary models

\[(\kappa^2 - \nabla \cdot H \nabla) u(s) = \mathcal{W}(s), \quad s \in D\]
SPDE/GMRF realisations and non-stationary models

\[(\kappa^2 - \nabla \cdot H(s)\nabla)u(s) = \mathcal{W}(s), \quad s \in D\]
SPDE/GMRF realisations and non-stationary models

\[(\kappa^2 - \nabla \cdot \nabla) u(s) = \mathcal{W}(s), \quad s \in D\]
SPDE/GMRF realisations and non-stationary models

\[( \kappa^2 - \nabla \cdot \nabla ) u(s) = \mathcal{W}(s), \quad s \in D \]
SPDE/GMRF realisations and non-stationary models

\[(\kappa^2 - \nabla \cdot \nabla) u(s) = \mathcal{W}(s), \quad s \in D = \mathbb{S}^2\]
SPDE/GMRF realisations and non-stationary models

\[(\kappa^2 \exp(i\theta) - \nabla \cdot \nabla) u(s) = \mathcal{W}(s), \quad s \in D = S^2\]
Markov does *not* mean local dependence

$$\left(\kappa(s)^2 - \nabla \cdot H(s) \nabla \right) u(s) = \kappa(s) \mathcal{W}(s), \quad s \in \Omega$$
Covariances for four reference points
Hierarchical models

Continuous Markovian spatial models (Lindgren et al, 2011)

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

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

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

Precision: \( Q = \kappa^4 C + 2\kappa^2 G + G_2 \) \( (\kappa^4 + 2\kappa^2 |\omega|^2 + |\omega|^4) \)

Conditional distribution in a jointly Gaussian model

\( u \sim \mathcal{N}(\mu_u, Q_{u|u}^{-1}) \), \( y|u \sim \mathcal{N}(Au, Q_{y|u}^{-1}) \) \( (A_{ij} = \psi_j(s_i)) \)

\( u|y \sim \mathcal{N}(\mu_{u|y}, Q_{u|y}^{-1}) \)

\( Q_{u|y} = Q_u + A^T Q_{y|u} A \) \( (\sim”Sparse \text{ iff } \psi_k \text{ have compact support”}) \)

\( \mu_{u|y} = \mu_u + Q_{u|y}^{-1} A^T Q_{y|u}(y - A\mu_u) \)
The computational GMRF work-horse

Cholesky decomposition (Cholesky, 1924)

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

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

\[ \log \det Q = 2 \log \det L = 2 \sum_i \log L_{ii} \]

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**

\[
p(u \mid \theta) \sim \mathcal{N}(\mu_u, Q_u^{-1}), \quad y \mid u, \theta \sim p(y \mid u)
\]

\[
p_G(u \mid y, \theta) \sim \mathcal{N}(\tilde{\mu}, \tilde{Q}^{-1})
\]

\[
0 = \nabla_u \{\ln p(u \mid \theta) + \ln p(y \mid u)\}|_{u=\tilde{\mu}}
\]

\[
\tilde{Q} = Q_u - \nabla_u^2 \ln p(y \mid u)|_{u=\tilde{\mu}}
\]

**Direct Bayesian inference with INLA (r-inla.org)**

\[
\tilde{p}(\theta \mid y) \propto \frac{p(\theta)p(u \mid \theta)p(y \mid u, \theta)}{p_G(u \mid y, \theta)}|_{u=\tilde{\mu}}
\]

\[
\tilde{p}(u_i \mid y) \propto \int p_{GG}(u_i \mid y, \theta)\tilde{p}(\theta \mid y) \, d\theta
\]

The main practical limiting factors for the INLA method are the number of latent variables and the number model parameters.
**Example: Point process data**

**Log-Gaussian Cox processes**

Point intensity:

\[
\lambda(s) = \exp \left( \sum_i b_i(s) \beta_i + u(s) \right)
\]

Inhomogeneous Poisson process log-likelihood:

\[
\ln p(\{y_k\} | \lambda) = |D| - \int_D \lambda(s) \, ds + \sum_{k=1}^{N} \ln \lambda(y_k)
\]

The likelihood can be approximated numerically, e.g.

\[
\int_D \lambda(s) \, ds \approx \sum_{j=1}^{n} \lambda(s_j) w_j,
\]

where \(s_j\) are mesh nodes, and \(w_j = \langle \psi_j, 1 \rangle_D\)

Finn Lindgren - f.lindgren@bath.ac.uk
Stochastic PDEs and Markov random fields with ecological applications
Example: Point process data (cont)

Discretised field and likelihood:

\[
\lambda(s) = \exp \left( \sum_i b_i(s) \beta_i + \sum_j \psi_j(s) u_j \right)
\]

\[
\ln p(\{y_k\} \mid \lambda) \approx |D| - \sum_{j=1}^n \lambda(s_j) w_j + \sum_{k=1}^n \ln \lambda(y_k)
\]

Then, with \( \lambda_D = [\lambda(s_i)] \), \( A_D = [\psi_j(s_i)] \), and \( A_y = [\psi_j(y_i)] \),

\[
\nabla_u \ln p(\{y_k\} \mid \lambda) \approx -A_D^\top \text{diag}(w) \lambda_D + A_y^\top 1
\]

\[
\nabla^2_u \ln p(\{y_k\} \mid \lambda) \approx -A_D^\top \text{diag}(w) \text{diag}(\lambda_D) A_D
\]

and similarly for \( \nabla_\beta \), \( \nabla^2_\beta \), and \( \nabla_u \nabla_\beta \).
SPDE based inference for point process data
A multiscale model example

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

\[
\frac{\partial}{\partial t} z(s, t) - \gamma_z \nabla \cdot \nabla z(s, t) = \mathcal{E}(s, t)
\]

\[
(1 - \gamma_E \nabla \cdot \nabla) \mathcal{E}(s, t) = \mathcal{W}_E(s, t)
\]

- A temporally quick, spatially non-stationary SPDE/GMRF (separable)

\[
\left( \frac{\partial}{\partial t} + \gamma_t \right) \left( \kappa(s)^2 - \nabla \cdot \nabla \right) (\tau(s) a(s, t)) = \mathcal{W}_a(s, t)
\]

- Measurements

\[
y_i = a(s_i, t_i) + z(s_i, t_i) + \epsilon_i, \text{ discretised into}
\]

\[
y = A(a + (B \otimes I)z) + \epsilon, \quad \epsilon \sim \mathcal{N}(0, Q_{\epsilon}^{-1})
\]

where \(B\) maps from long-term basis functions to short-term, and \(A\) maps from short-term basis functions to the observations.

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

\[
Q_{(a+z, z)|y} = \begin{bmatrix}
Q_t \otimes Q_a + A^\top Q_\epsilon A & -Q_t B \otimes Q_a \\
-B^\top Q_t \otimes Q_a & Q_z + B^\top Q_t B \otimes Q_a
\end{bmatrix}
\]
Locally isotropic non-stationary precision construction

Finite element construction of basis weight precision

Non-stationary SPDE:
\[
(\kappa(s)^2 - \nabla \cdot \nabla) (\tau(s) u(s)) = \mathcal{W}(s)
\]

The SPDE parameters are constructed via spatial covariates:
\[
\log \tau(s) = b_0^\tau(s) + \sum_{j=1}^{p} b_j^\tau(s) \theta_j, \quad \log \kappa(s) = b_0^\kappa(s) + \sum_{j=1}^{p} b_j^\kappa(s) \theta_j
\]

Finite element calculations give
\[
T = \text{diag}(\tau(s_i)), \quad K = \text{diag}(\kappa(s_i))
\]
\[
C_{ii} = \int \psi_i(s) \, ds, \quad G_{ij} = \int \nabla \psi_i(s) \cdot \nabla \psi_j(s) \, ds
\]
\[
Q = T \left( K^2 C K^2 + K^2 G + GK^2 + GC^{-1}G \right) T
\]

Combining this with an AR(1) discretisation of the temporal operator, we get
\[
Q_t \otimes Q_a.
\]
GMRF precision for simplified stochastic heat equation

\[ Q_z = M_2^{(t)} \otimes M_0^{(s)} + M_1^{(t)} \otimes M_1^{(s)} + M_0^{(t)} \otimes M_2^{(s)} \]

\[ M_0^{(s)} = C + \gamma \varepsilon G \]
\[ M_1^{(s)} = G + \gamma \varepsilon GC^{-1} G \]
\[ M_2^{(s)} = GC^{-1} G + \gamma \varepsilon GC^{-1} GC^{-1} G \]

Ignoring the degenerate aspect of the model, the precision structure can be used to formulate sampling as

\[ Q_z z = \tilde{L}_z w, \quad w \sim \mathcal{N}(0, I) \]

where \( \tilde{L}_z \) is a pseudo Cholesky factor,

\[ \tilde{L}_z = \left[ \begin{bmatrix} L_2^{(t)} \otimes L_C, & L_1^{(t)} \otimes L_G, & L_0^{(t)} \otimes GL_C^{-\top} \end{bmatrix}, \right. \\
\left. \gamma^{1/2}_\varepsilon \begin{bmatrix} L_2^{(t)} \otimes L_G, & L_1^{(t)} \otimes GL_C^{-\top}, & L_0^{(t)} \otimes GC^{-1} L_G \end{bmatrix} \right] \]
Posterior calculations

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

\[
Q_{x|y} = \begin{bmatrix}
Q_t \otimes Q_a + A^T Q_\epsilon A & -Q_t B \otimes Q_a \\
-B^T Q_t \otimes Q_a & Q_z + B^T Q_t B \otimes Q_a
\end{bmatrix}
\]

can be pseudo-Cholesky-factorised:

\[
Q_{x|y} = \tilde{L}_{x|y} \tilde{L}_{x|y}^T, \quad \tilde{L}_{x|y} = \begin{bmatrix}
L_t \otimes L_a & 0 & A^T L_\epsilon \\
-B^T L_t \otimes L_a & \tilde{L}_z & 0
\end{bmatrix}
\]

Posterior expectation, samples, and marginal variances (with \( \tilde{A} = [A \ 0] \)):

\[
Q_{x|y}(\mu_{x|y} - \mu_x) = \tilde{A}^T Q_\epsilon (y - \tilde{A} \mu_x),
\]

\[
Q_{x|y}(x - \mu_{x|y}) = \tilde{L}_{x|y} w, \quad w \sim \mathcal{N}(0, I), \quad \text{or}
\]

\[
Q_{x|y}(x - \mu_x) = \tilde{A}^T Q_\epsilon (y - \tilde{A} \mu_x) + \tilde{L}_{x|y} w, \quad w \sim \mathcal{N}(0, I),
\]

\[
\text{Var}(x_i|y) = \text{diag}(\text{inla.qinv}(Q_{x|y})) \quad \text{(requires Cholesky)}
\]
Preconditioning for e.g. conjugate gradient solutions

Solving \( Qx = b \) is equivalent to solving \( M^{-1}Qx = M^{-1}b \). Choosing \( M^{-1} \) as an approximate inverse to \( Q \) gives a less ill-conditioned system. Only the action of \( M^{-1} \) is needed, e.g. one or more fixed point iterations:

**Block Jacobi and Gauss-Seidel preconditioning**

Matrix split: \( Qx|y = L + D + L^\top \)

Jacobi: \( x^{(k+1)} = D^{-1} \left( -(L + L^\top)x^{(k)} + b \right) \)

Gauss-Seidel: \( x^{(k+1)} = (L + D)^{-1} \left( -L^\top x^{(k)} + b \right) \)

**Remark: Block Gibbs sampling for a GMRF posterior**

With \( Q = Qx|y, b = A^\top Q_\epsilon (y - A\mu_x) \) and \( \tilde{x} = x - \mu_x \),

\[
\tilde{x}^{(k+1)} = (L + D)^{-1} \left( -L^\top \tilde{x}^{(k)} + b + \tilde{L}_D w \right), \quad w \sim \mathcal{N}(0, I)
\]

Gauss-Seidel and Gibbs are both very inefficient on their own.
Finite element mesh

Triangulation mesh
Finite element mesh (spot the R map plotting surprise!)
All deterministic boundary conditions are ‘inappropriate’
Stationary stochastic boundary adjustment (current work)

Recall the Matérn generating SPDE

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

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

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

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

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

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

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

\[
E_{u,\alpha}^+(x) = \arg\max_{D} \{|D| : P(D \subseteq A_{u}^+(x)) \geq 1 - \alpha\}.
\]

\[
E_{u,\alpha}^-(x) = \arg\max_{D} \{|D| : P(D \subseteq A_{u}^-(x)) \geq 1 - \alpha\}.
\]

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 \texttt{excursions()}, evaluating high dimensional GMRF probabilities and finding credible regions. Latest version has user friendly options for continuous domain interpretations.
References (2/3)


- Preliminary draft book chapter on GMRF computation basics: http://people.bath.ac.uk/fl353/tmp/gmrf.pdf


CRAN package: excursions
References (2/3)

- J. Besag (1974),
  *Spatial interaction and the statistical analysis of lattice systems (with discussion)*, JRSSB, 36(2):192–225.

- J. Besag and C. Kooperberg (1995),

- J. Besag (1981),

- J. Besag and D. Mondal (2005),

- Y. A. Rozanov (1977),

- Y. A. Rozanov (1977),

- Y. A. Rozanov (1982, Russian original 1980),
References (3/3)


- 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!