Gaussian Markov Random Fields

Johan Lindström[1]

[1]Centre for Mathematical Sciences
Lund University

Pan-American Advanced Study Institute
Búzios
June 18, 2014

LUND UNIVERSITY

Spatial interpolation — Kriging

Given observations at some locations, $Y(\mathbf{s}_i),\ i = 1 \ldots n$
we want to make statements about the value at unobserved location(s), $X(\mathbf{s})$.

In the simplest case we assume a Gaussian model for the data

$$\begin{bmatrix} \mathbf{Y} \\ \mathbf{X} \end{bmatrix} \sim \mathsf{N}\left(\begin{bmatrix} \mu_y \\ \mu_x \end{bmatrix}, \begin{bmatrix} \Sigma_{yy} & \Sigma_{yx} \\ \Sigma_{yx}^\top & \Sigma_{xx} \end{bmatrix} \right),$$

with some parametric form for the covariance matrix and mean

$$\mathbf{Y} \sim \mathsf{N}\left(\mu(\theta), \Sigma(\theta)\right).$$

The “Big N” problem

The log-likelihood becomes

$$l(\theta|\mathbf{Y}) = -\frac{1}{2}\log|\Sigma(\theta)| - \frac{1}{2}\Big(\mathbf{Y} - \mu(\theta)\Big)^{\top}\Sigma(\theta)^{-1}\Big(\mathbf{Y} - \mu(\theta)\Big).$$

Given (estimated) parameters, predictions at the unobserved locations are given by

$$\mathsf{E}\left(\mathbf{X}\middle|\mathbf{Y},\widehat{\theta}\right) = \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(\mathbf{Y} - \mu_y).$$

The “Big N” problem

Given N observations:

- The covariance matrix has $\mathcal{O}\left(N^2\right)$ unique elements.
- Computations scale as $\mathcal{O}\left(N^3\right)$ (due to $|\Sigma|$ and Σ^{-1}).

Getting around "Big N"

Spectral representation (Whittle, 1954; Fuentes, 2007)
Uses discrete Fourier transforms; limited to regular lattices.

Covariance tapering (Furrer et al., 2006; Kaufman et al., 2008)
Set small values in the covariance matrix to zeros.

Likelihood approximation
Various statistical or numerical approximation methods

- By sequential matrix approx. (Stein et al., 2004)
- Block composite likelihoods (Eidsvik et al., 2014)
- Krylov based numerical approx. (Stein et al., 2013; Aune et al., 2014)

Low rank approximations
Exact computations on a simplified model of reduced rank/size

- Predictive processes (Banerjee et al., 2008; Eidsvik et al., 2012)
- Fixed rank kriging (Cressie and Johannesson, 2008)
- Process convolution or kernel methods (Higdon, 2001)

Gaussian Markov random fields (Rue and Held, 2005)

Let the **neighbours** $\mathcal{N}_i$ to a point $\boldsymbol{s}_i$ be the points $\{\boldsymbol{s}_j,\ j \in \mathcal{N}_i\}$ that are "close" to $\boldsymbol{s}_i$.

Gaussian Markov random field (GMRF)

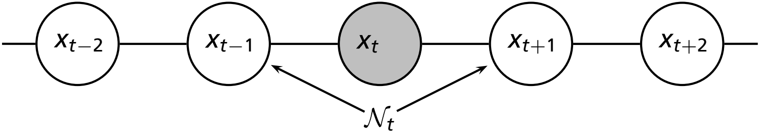
A Gaussian random field $\boldsymbol{x} \sim \mathsf{N}(\mu, \Sigma)$ that satisfies

$$p\left(x_i \middle| \{x_j : j \neq i\}\right) = p\left(x_i \middle| \{x_j : j \in \mathcal{N}_i\}\right)$$

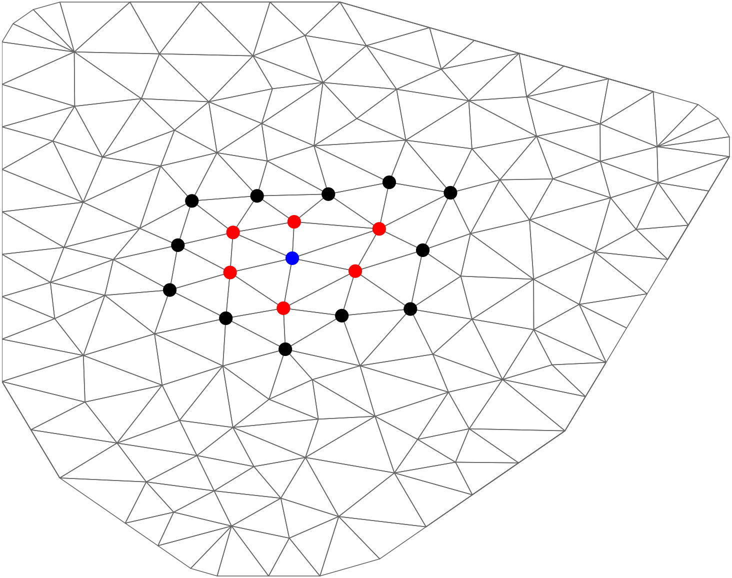
is a Gaussian Markov random field.

The simplest example of a GMRF is the AR(1)-process

$$x_t = a x_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{N}(0, \sigma^2) \text{ and independent.}$$



Good neighbours



Let me introduce: The precision matrix $\mathbf{Q} = \Sigma^{-1}$

Using the **precision matrix** the model becomes

$$\mathbf{X} \sim \mathsf{N}\left(\mu, \mathbf{Q}^{-1}\right) \qquad \text{cf.} \qquad \mathbf{X} \sim \mathsf{N}\left(\mu, \Sigma\right)$$

With conditional expectation

$$\mathsf{E}\left(\mathbf{X}|\mathbf{Y}\right) = \mu_x - \mathbf{Q}_{xx}^{-1}\mathbf{Q}_{xy}(\mathbf{Y} - \mu_y).$$

instead of

$$\mathsf{E}\left(\mathbf{X}|\mathbf{Y}\right) = \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(\mathbf{Y} - \mu_y).$$

The conditional expectation for a single location is

$$\begin{aligned}\mathsf{E}\left(x_i \middle| x_j, j \neq i\right) &= \mu_i - \frac{\sum_{j\neq i} Q_{ij}(x_j - \mu_j)}{Q_{ii}} \\ &= \mu_i - \frac{1}{Q_{ii}}\left(\sum_{j\in\mathcal{N}_i} Q_{ij}(x_j - \mu_j) + \sum_{j\notin\{\mathcal{N}_i,i\}} Q_{ij}(x_j - \mu_j)\right)\end{aligned}$$

If $Q_{ij} = 0$ for all $j \notin \mathcal{N}_i$ then

$$\mathsf{E}\left(x_i \middle| x_j, j \neq i\right) = \mathsf{E}\left(x_i \middle| x_j, j \in \mathcal{N}_i\right).$$

The precision matrix is sparse

Elements in the precision matrix of a Gaussin Markov random field are **non-zero only for neighbours** and diagonal elements.

$$j \notin \{i, \mathcal{N}_i\} \iff Q_{ij} = 0.$$

Computational details

If $\boldsymbol{Q}$ is a sparse matrix then (under mild conditions) the Cholesky factorisation $\boldsymbol{Q} = \boldsymbol{R}^\top \boldsymbol{R}$ will also be sparse.

- Simulation of $\boldsymbol{X} \in \mathsf{N}\left(\mu, \boldsymbol{Q}^{-1}\right)$.

$$\boldsymbol{X} = \mu + \boldsymbol{R}^{-1}\boldsymbol{E}, \quad \boldsymbol{E} \in \mathsf{N}(\boldsymbol{0}, \boldsymbol{I})$$

- Conditional expectations

$$\mathsf{E}(\boldsymbol{X}|\boldsymbol{Y}) = \mu_x - \boldsymbol{R}_{xx}^{-1}\left(\boldsymbol{R}_{xx}^{-\top}\left(\boldsymbol{Q}_{xy}(\boldsymbol{Y} - \mu_y)\right)\right)$$

- Computing the determinant

$$\frac{1}{2}\log|\boldsymbol{Q}| = \frac{1}{2}\log\left|\boldsymbol{R}^\top\boldsymbol{R}\right| = \log|\boldsymbol{R}| = \sum_i \log \boldsymbol{R}_{ii}$$

- Never compute $\boldsymbol{R}^{-1}$, use back substitution for triangular systems instead.

How to create Q?

The Matérn covariance family (Matérn, 1960)

The covariance between two points at distance $\|\boldsymbol{h}\|$ is

$$r(\|\boldsymbol{h}\|) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(\varkappa \|\boldsymbol{h}\|)^{\nu} K_{\nu}(\varkappa \|\boldsymbol{h}\|), \quad \boldsymbol{h} \in \mathbb{R}^d$$

Fields with Matérn covariances are solutions to a **Stochastic Partial Differential Equation (SPDE)** (Whittle, 1954, 1963),

$$\left(\varkappa^2 - \Delta\right)^{\alpha/2} x(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}).$$

Here $\mathcal{W}(\boldsymbol{s})$ is white noise, $\Delta = \sum_i \frac{\partial^2}{\partial s_i^2}$, and $\alpha = \nu + d/2$.

Does the Matérn covariance produce Markov fields?

The Matérn covariance has wave number spectrum

$$R(\boldsymbol{k}) \propto \frac{1}{(\varkappa^2 + \|\boldsymbol{k}\|^2)^\alpha} \qquad \text{cf.} \qquad \left(\varkappa^2 - \Delta\right)^{\alpha/2} x(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}).$$

Spectral density for Markov fields

According to Rozanov (1977) a stationary field is **Markov** if and only if the **spectral density** is a **reciprocal of a polynomial**.

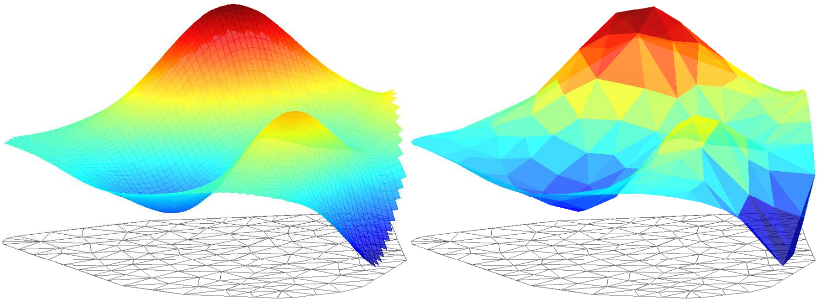
For the SPDE this implies $\alpha \in \mathbb{Z}$ (or $\nu \in \mathbb{Z}$ for $\mathbb{R}^2$).

Basic idea

Construct a discrete approximation of the continious field using basis functions, $\{\psi_k\}$, and weights, $\{w_k\}$,

$$x(\mathbf{s}) = \sum_k \psi_k(\mathbf{s}) w_k$$

Find the distribution of w_k by solving $\left(\kappa^2 - \Delta\right)^{\alpha/2} x(\mathbf{s}) = \mathcal{W}(\mathbf{s})$



Solving the SPDE

A **stochastic weak solution** to the SPDE is given by

$$\left[\left\langle \varphi_k,\, (\varkappa^2 - \Delta)^{\alpha/2} \boldsymbol{x} \right\rangle\right]_{k=1,\ldots,n} \overset{D}{=} \left[\langle \varphi_k,\, \mathcal{W}\rangle\right]_{k=1,\ldots,n}$$

for each set of **test functions** $\{\varphi_k\}$

Replacing $\boldsymbol{x}$ with $\sum_k \psi_k w_k$ gives

$$\left[\left\langle \varphi_i,\, (\varkappa^2 - \Delta)^{\alpha/2} \psi_j \right\rangle\right]_{i,j} \boldsymbol{w} \overset{D}{=} \left[\langle \varphi_k,\, \mathcal{W}\rangle\right]_k$$

Study the case $\alpha = 2$ and $\varphi_i = \psi_i$ (Galerkin)

$$\left(\varkappa^2 \underbrace{\left[\langle \psi_i,\, \psi_j\rangle\right]}_{\boldsymbol{C}} + \underbrace{\left[\langle \psi_i,\, -\Delta\psi_j\rangle\right]}_{\boldsymbol{G}}\right) \boldsymbol{w} \overset{D}{=} \underbrace{\left[\langle \psi_k,\, \mathcal{W}\rangle\right]}_{\mathsf{N}(0,\boldsymbol{C})}$$

Solution to the SPDE

A weak solution to the SPDE

$$\left(\kappa^2 - \Delta\right) x(\mathbf{s}) = \mathcal{W}(\mathbf{s}).$$

is given by

$$x(\mathbf{s}) = \sum_k \psi_k(\mathbf{s}) w_k \qquad \text{where} \qquad \left(\kappa^2 \mathbf{C} + \mathbf{G}\right) \mathbf{w} \sim \mathsf{N}\left(0, \mathbf{C}\right)$$

The precision of the weights, $\mathbf{w}$, is

$$\mathsf{V}\left(\mathbf{w}\right)^{-1} = \mathbf{Q}_2 = \left(\kappa^2 \mathbf{C} + \mathbf{G}\right)^\top \mathbf{C}^{-1} \left(\kappa^2 \mathbf{C} + \mathbf{G}\right)$$

$$\begin{aligned}
\mathbf{Q}_1 &= \kappa^2 \mathbf{C} + \mathbf{G} \\
\mathbf{Q}_2 &= \left(\kappa^2 \mathbf{C} + \mathbf{G}\right)^\top \mathbf{C}^{-1} \left(\kappa^2 \mathbf{C} + \mathbf{G}\right) \\
\mathbf{Q}_\alpha &= \left(\kappa^2 \mathbf{C} + \mathbf{G}\right)^\top \mathbf{C}^{-1} \mathbf{Q}_{\alpha-2} \mathbf{C}^{-1} \left(\kappa^2 \mathbf{C} + \mathbf{G}\right), \quad \alpha = 3, 4, 5, \ldots
\end{aligned}$$

What's a good basis?

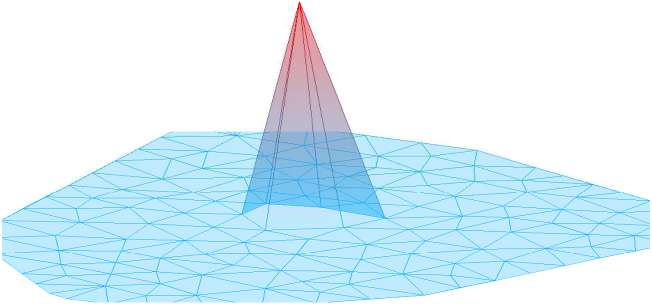
Several possible basis functions exist:

Harmonic functions gives a spectral representation (Thon et al., 2012; Sigrist et al., 2012).

Eigenfunctions of the covariance matrix give Karhunen-Loéve.

Piecewise linear basis gives (almost) a GMRF (Lindgren et al., 2011).

Waveletts also gives a GMRF (Bolin and Lindgren, 2013).



But it's not a Markov field! — Yet

Using a piecewise linear basis only **neighbouring basis functions overlap**, so both

$$\mathbf{G}_{ij} = \langle \psi_i, -\Delta\psi_j \rangle \qquad \text{and} \qquad \mathbf{C}_{ij} = \langle \psi_i, \psi_j \rangle$$

are **sparse**. However, $\mathbf{C}^{-1}$ is **not sparse**.

GMRF approximation

To obtain sparse precision matrices we replace the $\mathbf{C}$-matrix with a diagonal matrix $\widetilde{\mathbf{C}}$ with elements

$$\widetilde{\mathbf{C}}_{i,i} = \int \psi_i(\mathbf{s})\, \mathrm{d}\mathbf{s}$$

The resulting approximation error is small (Bolin and Lindgren, 2013; Simpson et al., 2010)

Regular lattice in $\mathbb{R}^2$

Solving the SPDE gives a GMRF with precision elements

Order $\alpha = 1$ ($\nu = 0$):

$$\kappa^2 \begin{bmatrix} & & \\ & 1 & \\ & & \end{bmatrix} + \underbrace{\begin{bmatrix} & -1 & \\ -1 & 4 & -1 \\ & -1 & \end{bmatrix}}_{\approx -\Delta}$$

Order $\alpha = 2$ ($\nu = 1$):

$$\kappa^4 \begin{bmatrix} & & \\ & 1 & \\ & & \end{bmatrix} + 2\kappa^2 \underbrace{\begin{bmatrix} & -1 & \\ -1 & 4 & -1 \\ & -1 & \end{bmatrix}}_{\approx -\Delta} + \underbrace{\begin{bmatrix} & & 1 & & \\ & 2 & -8 & 2 & \\ 1 & -8 & 20 & -8 & 1 \\ & 2 & -8 & 2 & \\ & & 1 & & \end{bmatrix}}_{\approx \Delta^2}$$

Observations— The $\boldsymbol{A}$-matrix

The field is created as a **weighted sum of basis functions**.

$$x(\boldsymbol{s}) = \sum_{k=1}^{N} \psi_k(\boldsymbol{s})\, w_k,$$

The locations of the basis functions do **not** need to match observation locations.

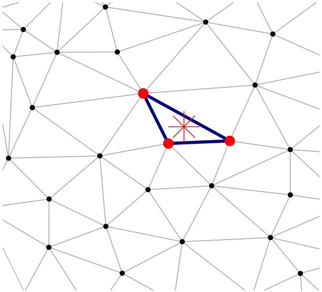
Observations

$$\begin{aligned} y(\boldsymbol{s}) &= x(\boldsymbol{s}) + \varepsilon \\ &= \sum_k \psi_k(\boldsymbol{s}) w_k + \varepsilon \end{aligned}$$

We introduce a sparse matrix

$$\boldsymbol{A}_{i\cdot} = \begin{bmatrix} \psi_1(\boldsymbol{s}_i) & \cdots & \psi_N(\boldsymbol{s}_i) \end{bmatrix}$$

linking the field to the observation.



Conditional expectation

Given a matrix A with rows

$$A_{i\cdot} = \begin{bmatrix} \psi_1(\mathbf{s}_i) & \cdots & \psi_N(\mathbf{s}_i) \end{bmatrix}$$

we can write the observation equation on matrix form as

$$\mathbf{Y}|\mathbf{w} \in \mathsf{N}\left(\mathbf{A}\mathbf{w}, \mathbf{Q}_{\varepsilon}^{-1}\right) \qquad \mathbf{w} \in \mathsf{N}\left(\mu, \mathbf{Q}^{-1}\right)$$

Kriging with GMRF

$$\mathsf{E}\left(\mathbf{w}|\mathbf{y}\right) = \mu + \mathbf{Q}_{w|y}^{-1}\mathbf{A}^{\top}\mathbf{Q}_{\varepsilon}\left(\mathbf{y} - \mathbf{A}\mu\right)$$

$$\mathsf{V}\left(\mathbf{w}|\mathbf{y}\right) = \mathbf{Q}_{w|y}^{-1} = \left(\mathbf{Q} + \mathbf{A}^{\top}\mathbf{Q}_{\varepsilon}\mathbf{A}\right)^{-1}$$

Bayesian hierarchical modelling using GMRF

Data model, $p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta})$: Describing how **observations** arise assuming a **known latent field $\boldsymbol{x}$**.

Latent model, $p(\boldsymbol{x}|\boldsymbol{\theta})$: Describing how the **latent field** behaves.

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{w} + \boldsymbol{B}\beta \qquad \boldsymbol{w} \sim \mathsf{N}\left(\boldsymbol{0}, \boldsymbol{Q}^{-1}(\theta)\right)$$

Parameters, $p(\boldsymbol{\theta})$: Describing our, sometimes vauge, **prior knowledge** of the parameters.

For INLA we require that

$$p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\theta}) = \prod_i p(y_i|x_i,\boldsymbol{\theta})$$

Inference

Given a Bayesian hierarchical model we are interested in

Posteriors for the parameters

$$\mathrm{p}\left(\boldsymbol{\theta}|\boldsymbol{y}\right) \propto \mathrm{p}\left(\boldsymbol{y}|\boldsymbol{\theta}\right)\mathrm{p}\left(\boldsymbol{\theta}\right)$$

Posteriors for the latent field

$$\mathrm{p}\left(\boldsymbol{x}|\boldsymbol{y}\right) \propto \int \mathrm{p}\left(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}\right)\mathrm{p}\left(\boldsymbol{\theta}|\boldsymbol{y}\right)\,\mathrm{d}\boldsymbol{\theta}$$

Computing the posterior — A "trick"

Conditional distributions provide the equality

$$p(\boldsymbol{y}|\boldsymbol{x},\theta)\,p(\boldsymbol{x}|\theta) = p(\boldsymbol{y},\boldsymbol{x}|\theta) = p(\boldsymbol{x}|\boldsymbol{y},\theta)\,p(\boldsymbol{y}|\theta)$$

This gives

$$p(\theta|\boldsymbol{y}) \propto \underbrace{\frac{p(\boldsymbol{y}|\boldsymbol{x},\theta)\,p(\boldsymbol{x}|\theta)}{p(\boldsymbol{x}|\boldsymbol{y},\theta)}}_{p(\boldsymbol{y}|\theta)} \cdot p(\theta) \quad \text{for any } \boldsymbol{x}.$$

For non-Gaussian observations we need a good approximation of

$$p(\boldsymbol{x}|\boldsymbol{y},\theta)$$

Approximating the posterior

For non-Gaussian observations we have that

$$\log \mathrm{p}(\boldsymbol{x}|\boldsymbol{y},\theta) = \log \mathrm{p}(\boldsymbol{y}|\boldsymbol{x},\theta) + \log \mathrm{p}(\boldsymbol{x}|\theta) + \text{const.}$$

Using a second order **Taylor approximation** of

$$f(\boldsymbol{x}) = \log \mathrm{p}(\boldsymbol{y}|\boldsymbol{x},\theta) \qquad \text{around } \boldsymbol{x}_0$$

we can obtain a Gaussian approximation $\mathrm{p}_G(\boldsymbol{x}|\boldsymbol{y},\theta)$, with

$$\mathsf{E}_{x_0}(\boldsymbol{x}|\boldsymbol{y},\theta) \approx \left(\boldsymbol{Q} - \operatorname{diag}(f''(\boldsymbol{x}_0))\right)^{-1}\left(\boldsymbol{Q}\mu + f'(\boldsymbol{x}_0) - f''(\boldsymbol{x}_0)\boldsymbol{x}_0\right)$$

$$\mathsf{V}_{x_0}(\boldsymbol{x}|\boldsymbol{y},\theta) \approx \left(\boldsymbol{Q} - \operatorname{diag}(f''(\boldsymbol{x}_0))\right)^{-1}$$

Integrated Nested Laplace Approximation – INLA

To evaluate $\mathrm{p}(\boldsymbol{\theta}|\boldsymbol{y})$ using the Taylor/Laplace approximation do:

1. For a given $\boldsymbol{\theta}$ find the mode
$$\boldsymbol{x}_0 = \operatorname*{argmax}_{\boldsymbol{x}} \mathrm{p}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta})$$
2. Compute the Taylor expansion of $f(\boldsymbol{x})$ around $\boldsymbol{x}_0$
3. The approximation of $\mathrm{p}(\boldsymbol{\theta}|\boldsymbol{y})$ is
$$\widetilde{\mathrm{p}}(\boldsymbol{\theta}|\boldsymbol{y}) \propto \frac{\mathrm{p}(\boldsymbol{y}|\boldsymbol{x}_0,\boldsymbol{\theta})\,\mathrm{p}(\boldsymbol{x}_0|\boldsymbol{\theta})\,\mathrm{p}(\boldsymbol{\theta})}{\mathrm{p}_G(\boldsymbol{x}_0|\boldsymbol{y},\boldsymbol{\theta})}$$

The (approximate) maximum likelihood estimate is

$$\boldsymbol{\theta}_{\mathrm{ML}} \approx \operatorname*{argmax}_{\boldsymbol{\theta}} \widetilde{\mathrm{p}}(\boldsymbol{\theta}|\boldsymbol{y})$$

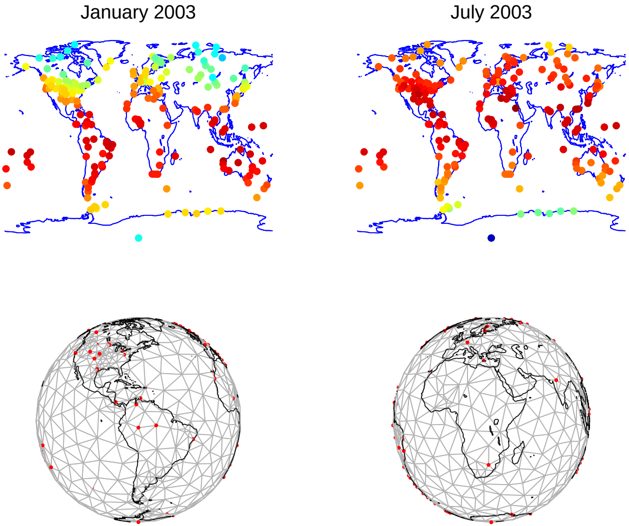
Use numerical integration over $\boldsymbol{\theta}$ to obtain posteriors for $[x|y]$

$$\mathrm{p}(x_i|\boldsymbol{y}) = \int \mathrm{p}(x_i|\boldsymbol{\theta},\boldsymbol{y})\,\mathrm{p}(\boldsymbol{\theta}|\boldsymbol{y})\,\mathrm{d}\boldsymbol{\theta} \approx \sum_k \mathrm{p}_{\mathrm{G}}(x_i|\boldsymbol{\theta}_k,\boldsymbol{y})\,\widetilde{\mathrm{p}}(\boldsymbol{\theta}_k|\boldsymbol{y})$$

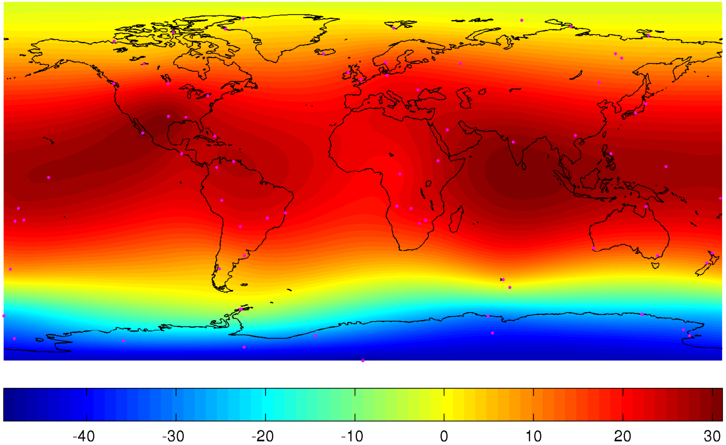
INLA (Rue et al., 2009)

- Posteriors computed include
  - $\mathrm{p}(\theta_i|\boldsymbol{y})$
  - $\mathrm{p}(\boldsymbol{\theta}|\boldsymbol{y})$
  - $\mathrm{p}(x_i|\boldsymbol{y})$
  - Some linear combinations of the field.
- The full joint posterior $\mathrm{p}(\boldsymbol{x}|\boldsymbol{y})$ is **not** computed.
- Errors due to the Taylor expansion and numerical integration are usually smaller than the MCMC errors.
- R-package: `http://www.r-inla.org`
- The package is fast, but **memory intensive** for large problems.
- See also Lindgren and Rue (2013) for applications to the SPDE-model.

Global Temperature Data



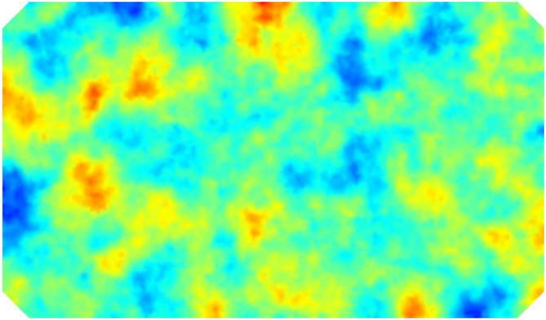
Global Temperature Data



GMRFs based on SPDEs (Lindgren et al., 2011)

GMRF representations of SPDEs can be constructed for to oscillating, anisotropic, non-stationary, non-separable spatio-temporal, and multivariate fields on manifolds.

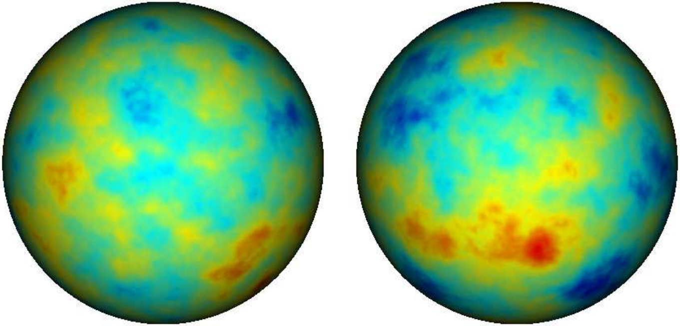
$$(\kappa^2 - \Delta)(\tau x(\boldsymbol{s})) = \mathcal{W}(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathbb{R}^d$$



GMRFs based on SPDEs (Lindgren et al., 2011)

GMRF representations of SPDEs can be constructed for to oscillating, anisotropic, non-stationary, non-separable spatio-temporal, and multivariate fields on manifolds.

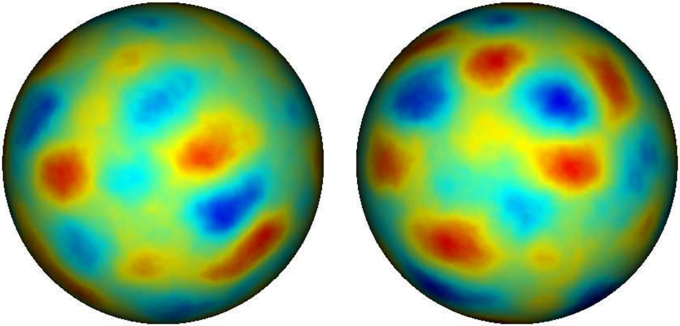
$$(\varkappa^2 - \Delta)(\tau x(\boldsymbol{s})) = \mathcal{W}(\boldsymbol{s}), \quad \boldsymbol{s} \in \Omega$$



GMRFs based on SPDEs (Lindgren et al., 2011)

GMRF representations of SPDEs can be constructed for to oscillating, anisotropic, non-stationary, non-separable spatio-temporal, and multivariate fields on manifolds.

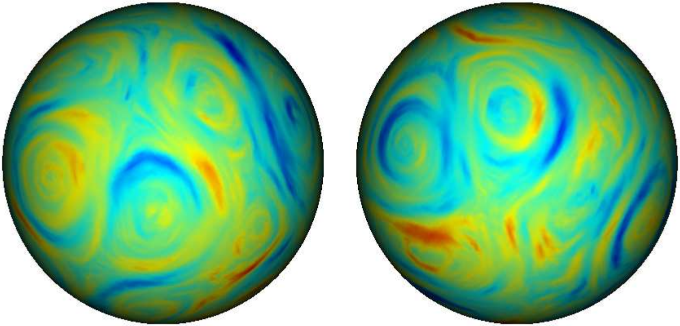
$$(\varkappa^2 e^{i\pi\theta} - \Delta)(\tau x(\boldsymbol{s})) = \mathcal{W}(\boldsymbol{s}), \quad \boldsymbol{s} \in \Omega$$



GMRFs based on SPDEs (Lindgren et al., 2011)

GMRF representations of SPDEs can be constructed for to oscillating, anisotropic, non-stationary, non-separable spatio-temporal, and multivariate fields on manifolds.

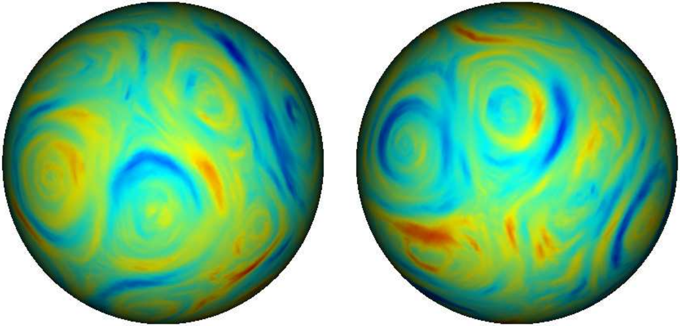
$$(\kappa_{\mathbf{s}}^2 + \nabla \cdot \boldsymbol{m}_{\mathbf{s}} - \nabla \cdot \boldsymbol{M}_{\mathbf{s}} \nabla)(\tau_{\mathbf{s}} x(\mathbf{s})) = \mathcal{W}(\mathbf{s}), \quad \mathbf{s} \in \Omega$$



GMRFs based on SPDEs (Lindgren et al., 2011)

GMRF representations of SPDEs can be constructed for to oscillating, anisotropic, non-stationary, non-separable spatio-temporal, and multivariate fields on manifolds.

$$\left(\frac{\partial}{\partial t} + \kappa^2_{\mathbf{s},t} + \nabla \cdot \mathbf{m}_{\mathbf{s},t} - \nabla \cdot \mathbf{M}_{\mathbf{s},t} \nabla\right) \left(\tau_{\mathbf{s},t} x(\mathbf{s}, t)\right) = \mathcal{E}(\mathbf{s}, t), \quad (\mathbf{s}, t) \in \Omega \times \mathbb{R}$$



Questions?

Bibliography I

Aune, E., Simpson, D. P., and Eidsvik, J. (2014), "Parameter estimation in high dimensional Gaussian distributions," *Stat. Comput.*, 24, 247–263.

Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008), "Gaussian predictive process models for large spatial data sets," *J. R. Stat. Soc. B*, 70, 825–848.

Bolin, D. and Lindgren, F. (2013), "A comparison between Markov approximations and other methods for large spatial data sets," *Comput. Stat. Data Anal.*, 61, 7–21.

Cressie, N. and Johannesson, G. (2008), "Fixed rank kriging for very large spatial data sets," *J. R. Stat. Soc. B*, 70, 209–226.

Eidsvik, J., Finley, A. O., Banerjee, S., and Rue, H. (2012), "Approximate Bayesian inference for large spatial datasets using predictive process models," *Comput. Stat. Data Anal.*, 56, 1362–138.

Eidsvik, J., Shaby, B. A., Reich, B. J., Wheeler, M., and Niemi, J. (2014), "Estimation and Prediction in Spatial Models With Block Composite Likelihoods," *J. Comput. Graphical Stat.*, 23, 295–315.

Bibliography II

Fuentes, M. (2007), "Approximate Likelihood for Large Irregularly Spaced Spatial Data," *J. Am. Stat. Assoc.*, 102, 321–331.

Furrer, R., Genton, M. G., and Nychka, D. (2006), "Covariance Tapering for Interpolation of Large Spatial Datasets," *J. Comput. Graphical Stat.*, 15, 502–523.

Higdon, D. (2001), "Space and space time modeling using process convolutions," Tech. rep., Institute of Statistics and Decision Sciences, Duke University, Durham, NC, USA.

Kaufman, C. G., Schervish, M. J., and Nychka, D. W. (2008), "Covariance Tapering for Likelihood-Based Estimation in Large Spatial Data Sets," *J. Am. Stat. Assoc.*, 103, 1545–1555.

Lindgren, F. and Rue, H. (2013), "Bayesian Spatial and Spatio-temporal Modelling with R-INLA," Submitted to J. Stat. Softw.

Lindgren, F., Rue, H., and Lindström, J. (2011), "An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach," *J. R. Stat. Soc. B*, 73, 423–498.

Bibliography III

Matérn, B. (1960), "Spatial variation. Stochastic models and their application to some problems in forest surveys and other sampling investigations." Ph.D. thesis, Stockholm University, Stockholm, Sweden.

Rozanov, Y. A. (1977), "Markov random fields and stochastic partial differential equations," *Math. USSR Sb.*, 32, 515–534.

Rue, H. and Held, L. (2005), *Gaussian Markov Random Fields; Theory and Applications*, vol. 104 of *Monographs on Statistics and Applied Probability*, Chapman & Hall/CRC.

Rue, H., Martino, S., and Chopin, N. (2009), "Approximate Bayesian inference for hierarchical Gaussian Markov random field models," *J. R. Stat. Soc. B*, 71, 1–35.

Sigrist, F., Künsch, H. R., and Stahel, W. A. (2012), "Stochastic partial differential equation based modeling of large space-time data sets," Tech. Rep. arXiv:1204.6118v5, arXiv preprint.

Simpson, D., Lindgren, F., and Rue, H. (2010), "In order to make spatial statistics computationally feasible, we need to forget about the covariance function," *Environmetrics*, 23, 65–74.

Bibliography IV

Stein, M. L., Chen, J., and Anitescu, M. (2013), "Stochastic approximation of score functions for Gaussian processes," *Ann. Stat.*, 7, 1162–1191.

Stein, M. L., Chi, Z., and Welty, L. J. (2004), "Approximating likelihoods for large spatial data sets," *J. R. Stat. Soc. B*, 66, 275–296.

Thon, K., Rue, H., Skrøvseth, S. O., and Godtliebsen, F. (2012), "Bayesian multiscale analysis of images modeled as Gaussian Markov random fields," *Comput. Stat. Data Anal.*, 56, 49–61.

Whittle, P. (1954), "On Stationary Processes in the Plane," *Biometrika*, 41, 434–449.

— (1963), "Stochastic processes in several dimensions," *Bull. Int. Stat. Inst.*, 40, 974–994.