Inference for non-stationary spatio-temporal models

Finn Lindgren

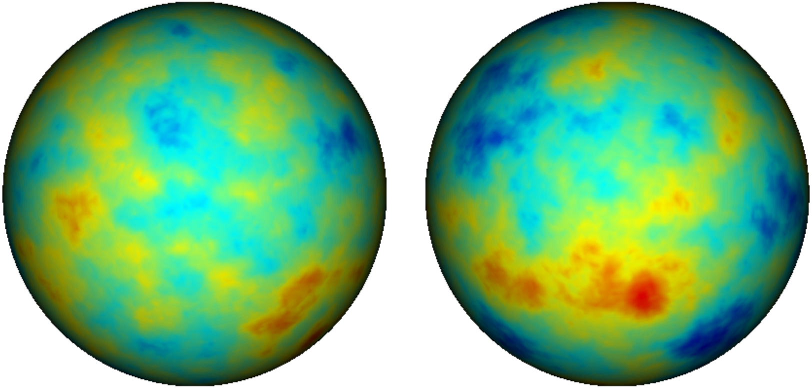
University of Bath

Leeds
1 February 2013

Statistical problem: An irregular, global data set

Temperature measurements from stations on land and ships on oceans
Reconstruct weather and climate, with proper uncertainty estimates

Stochastic non-stationary spatio-temporal models



Here be monsters: We need methods to do calculations!
Even Markov models eventually become too big for direct methods.

Hierarchical spatial models

Hierarchical models

$\boldsymbol{\theta}$	Model parameters
$\mathbf{x}\|\boldsymbol{\theta}$	Latent processes, spatial or spatio-temporal fields
$\mathbf{y}\|\boldsymbol{\theta},\mathbf{x}$	Measured data

Classical spatial models

Spatial field: $x(\mathbf{u}),\ \mathbf{u} \in \mathbb{R}^d, \quad \{x(\mathbf{u}_i)\} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$

Spatial covariance: $\Sigma_{i,j} = \mathrm{Cov}(x(\mathbf{u}_i), x(\mathbf{u}_j))$

Measurements: $y_i = \mathbf{B}_i\boldsymbol{\beta} + x(\mathbf{u}_i) + \epsilon_i, \quad \boldsymbol{\epsilon}|\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon)$

Covariance $\boldsymbol{\Sigma}$: Explicit global dependence

Precision $\mathbf{Q} = \boldsymbol{\Sigma}^{-1}$: Explicit local, implicit global dependence

Describing spatial dependence

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

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

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

Whittle (1954, 1963): Matérn as SPDE solution

Matérn fields are stationary solutions to the SPDE

$$\left(\kappa^2 - \Delta\right)^{\alpha/2} x(\mathbf{u}) = \mathcal{W}(\mathbf{u}), \quad \alpha = \nu + d/2$$

$\sigma^2 = \frac{\Gamma(\nu)}{\Gamma(\alpha)\kappa^{2\nu}(4\pi)^{d/2}}$, Laplacian $\Delta = \sum_{i=1}^d \frac{\partial^2}{\partial u_i^2}$

Piecewise linear Markov models

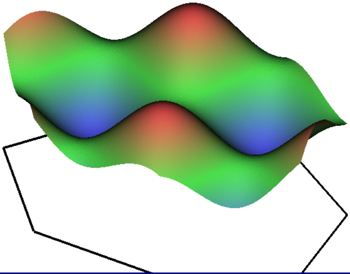
Continuous Markovian spatial models (Lindgren et al, 2011)

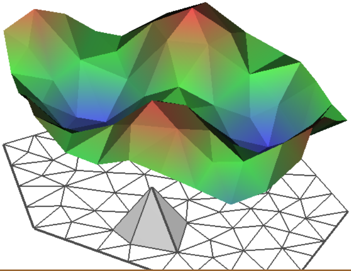
Local basis: $x(\mathbf{u}) = \sum_k \psi_k(\mathbf{u}) x_k$

Basis weights: $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_x^{-1})$, sparse $\mathbf{Q}$

Measurements: $\mathbf{y} = \mathbf{B}\beta + \mathbf{A}\mathbf{x} + \epsilon$, $\epsilon|\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{y|x}^{-1})$

Posterior: Local observations $\implies$ Markovian posterior for $\mathbf{x}$





The best piecewise linear approximation $\sum_k \psi_k(\mathbf{u})\, x_k$

Projection of the SPDE: Linear systems of equations ($\alpha = 2$)

$$\sum_j (\kappa^2 \underbrace{\langle \psi_i,\, \psi_j \rangle}_{\mathbf{C}_{ij}} + \underbrace{\langle \psi_i,\, -\Delta\psi_j \rangle}_{\mathbf{G}_{ij}}) x_j \overset{D}{=} \langle \psi_i,\, \mathcal{W} \rangle \quad \text{jointly for all } i.$$

$\mathbf{C}$ and $\mathbf{G}$ are as sparse as the triangulation neighbourhood

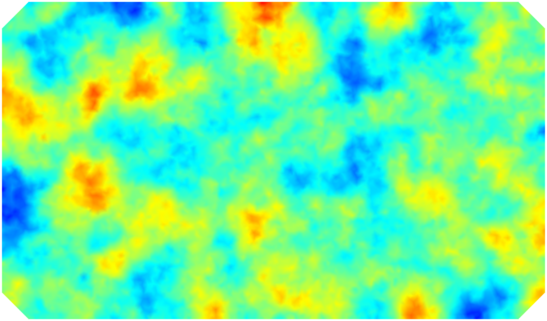
Constructing the precision matrices

$\mathbf{K} = \kappa^2\mathbf{C} + \mathbf{G}$	$\alpha = 1$	$\alpha = 2$	$\alpha = 3, 4, \ldots$
$\mathbf{Kx}$	$\mathcal{N}(\mathbf{0}, \mathbf{K})$	$\mathcal{N}(\mathbf{0}, \mathbf{C})$	$\mathcal{N}\left(\mathbf{0}, \mathbf{C}\mathbf{Q}_{\mathbf{x},\alpha-2}^{-1}\mathbf{C}\right)$
$\mathbf{Q}_{\mathbf{x},\alpha}$	$\mathbf{K}$	$\mathbf{K}^\mathsf{T}\mathbf{C}^{-1}\mathbf{K}$	$\mathbf{K}^\mathsf{T}\mathbf{C}^{-1}\mathbf{Q}_{\mathbf{x},\alpha-2}\mathbf{C}^{-1}\mathbf{K}$

Simulations with precisions via finite element calculations

The approach can in a straightforward way be extended to oscillating, anisotropic, non-stationary, non-separable spatio-temporal, and multivariate fields on manifolds.

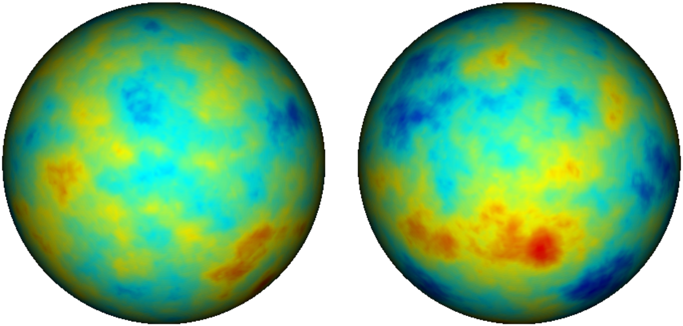
$$(\kappa^2 - \Delta)(\tau\, x(\mathbf{u})) = \mathcal{W}(\mathbf{u}), \quad \mathbf{u} \in \mathbb{R}^d$$



Simulations with precisions via finite element calculations

The approach can in a straightforward way be extended to oscillating, anisotropic, non-stationary, non-separable spatio-temporal, and multivariate fields on manifolds.

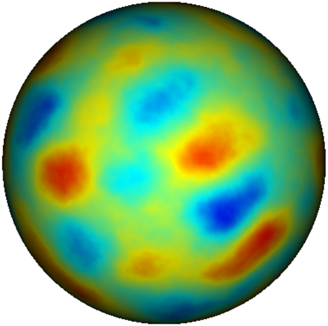
$$(\kappa^2 - \Delta)(\tau\, x(\mathbf{u})) = \mathcal{W}(\mathbf{u}), \quad \mathbf{u} \in \Omega$$

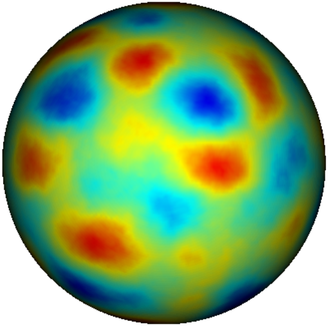


Simulations with precisions via finite element calculations

The approach can in a straightforward way be extended to oscillating, anisotropic, non-stationary, non-separable spatio-temporal, and multivariate fields on manifolds.

$$(\kappa^2 e^{i\pi\theta} - \Delta)(\tau x(\mathbf{u})) = \mathcal{W}(\mathbf{u}), \quad \mathbf{u} \in \Omega$$

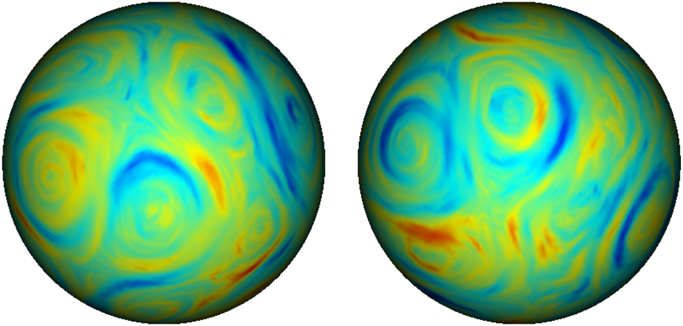




Simulations with precisions via finite element calculations

The approach can in a straightforward way be extended to oscillating, anisotropic, non-stationary, non-separable spatio-temporal, and multivariate fields on manifolds.

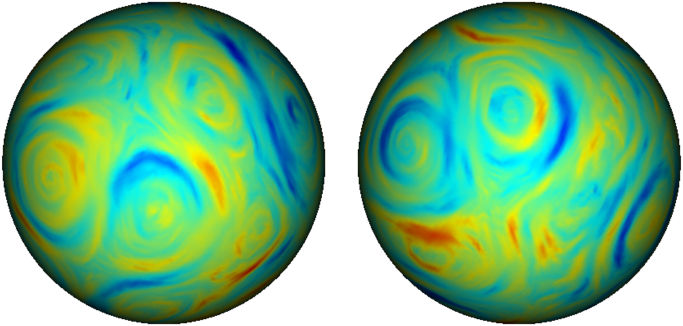
$$(\kappa_{\mathbf{u}}^2 + \nabla \cdot \mathbf{m}_{\mathbf{u}} - \nabla \cdot \mathbf{M}_{\mathbf{u}} \nabla)(\tau_{\mathbf{u}} x(\mathbf{u})) = \mathcal{W}(\mathbf{u}), \quad \mathbf{u} \in \Omega$$



Simulations with precisions via finite element calculations

The approach can in a straightforward way be extended to oscillating, anisotropic, non-stationary, non-separable spatio-temporal, and multivariate fields on manifolds.

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



Direct Bayesian inference (r-inla.org)

Conditional distribution in a Gaussian model

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}_x, \mathbf{Q}_x^{-1}), \quad \mathbf{y}|\mathbf{x} \sim \mathcal{N}(\mathbf{A}\mathbf{x}, \mathbf{Q}_{y|x}^{-1})$$

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

$$\mathbf{Q}_{x|y} = \mathbf{Q}_x + \mathbf{A}^T \mathbf{Q}_{y|x} \mathbf{A}$$

$$\boldsymbol{\mu}_{x|y} = \boldsymbol{\mu}_x + \mathbf{Q}_{x|y}^{-1} \mathbf{A}^T \mathbf{Q}_{y|x} (\mathbf{y} - \mathbf{A}\boldsymbol{\mu}_x)$$

Direct Bayesian inference with INLA

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto \left. \frac{p(\boldsymbol{\theta}) p(\mathbf{x}|\boldsymbol{\theta}) p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})}{p_G(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta})} \right|_{\mathbf{x}=\mathbf{x}^*}$$

$$p(\mathbf{x}_i|\mathbf{y}) \propto \int p_G(\mathbf{x}_i|\mathbf{y}, \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{y}) \mathrm{d}\boldsymbol{\theta}$$

Point pattern data

Log-Gaussian Cox processes

Point intensity:

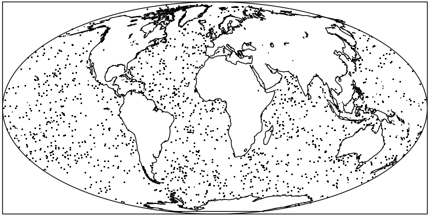
$$\lambda(\mathbf{u}) = \exp\left(\sum_i b_i(\mathbf{u})\beta_i + x(\mathbf{u})\right)$$

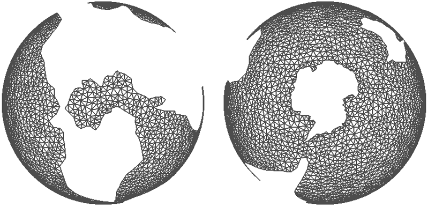
Inhomogeneous Poisson process likelihood:

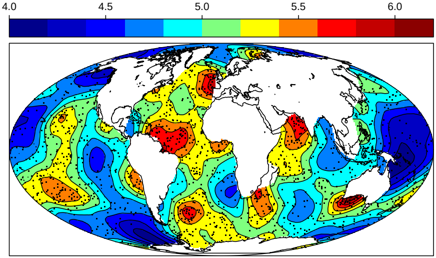
$$p(\mathbf{y}_1, \ldots, \mathbf{y}_n \mid \boldsymbol{\lambda}) = \exp\left(|\Omega| - \int_\Omega \lambda(\mathbf{u})\, d\mathbf{u}\right) \prod_{k=1}^n \lambda(\mathbf{y}_k)$$

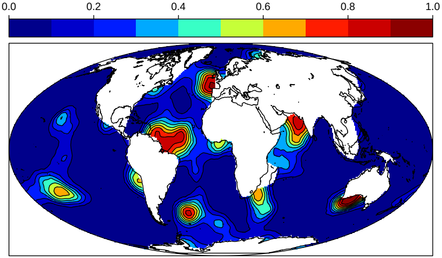
The likelihood can be approximated efficiently using the Markov property.

log-Gaussian Cox point process on a manifold









Linear model for weather observations

Weather = Climate + Anomaly

$$\mathbf{z} \sim \mathrm{N}(0, \mathbf{Q}_z^{-1}) \quad \text{(climate: space-time model)}$$

$$z(t, \mathbf{s}) = \sum_k B_k(t) \mathbf{z}_k(\mathbf{s}) \quad \text{(basis function representation)}$$

$$\mathbf{a} \sim \mathrm{N}(0, \mathbf{I} \otimes \mathbf{Q}_a^{-1}) \quad \text{(anomaly: spatial model, indep. in time)}$$

$$w(t, \mathbf{s}) = a(t, \mathbf{s}) + z(t, \mathbf{s}) \quad \text{(weather)}$$

$$y_i = \text{altitude effect } + w(t_i, \mathbf{s}_i) + \epsilon_i \quad \text{(observations)}$$

$$\epsilon \sim \mathrm{N}(0, \mathbf{Q}_\epsilon^{-1})$$

$$\mathbf{y} = \mathbf{A}(\mathbf{a} + (\mathbf{B} \otimes \mathbf{I})\mathbf{z}) + \epsilon$$

Stochastic weather anomaly model

Non-stationary spatial SPDE

$$
\begin{aligned}
(\kappa(\mathbf{s})^2 - \Delta)(\tau(\mathbf{s})a(\mathbf{s})) &= \mathcal{W}(\mathbf{s}) \\
\log \kappa(\mathbf{s}) &= \sum B_k^\kappa(\mathbf{s})\theta_k \\
\log \tau(\mathbf{s}) &= \sum B_k^\tau(\mathbf{s})\theta_k
\end{aligned}
$$

Precision

$$
\begin{aligned}
&\mathbf{K}_{ii} = \kappa(\mathbf{s}_i) \quad \mathbf{T}_{ii} = \tau(\mathbf{s}_i) \\
&\mathbf{Q}_a = \mathbf{T}\left(\mathbf{K}^2\mathbf{C}\mathbf{K}^2 + \mathbf{K}^2\mathbf{G} + \mathbf{G}\mathbf{K}^2 + \mathbf{G}\mathbf{C}^{-1}\mathbf{G}\right)\mathbf{T}
\end{aligned}
$$

Stochastic climate model

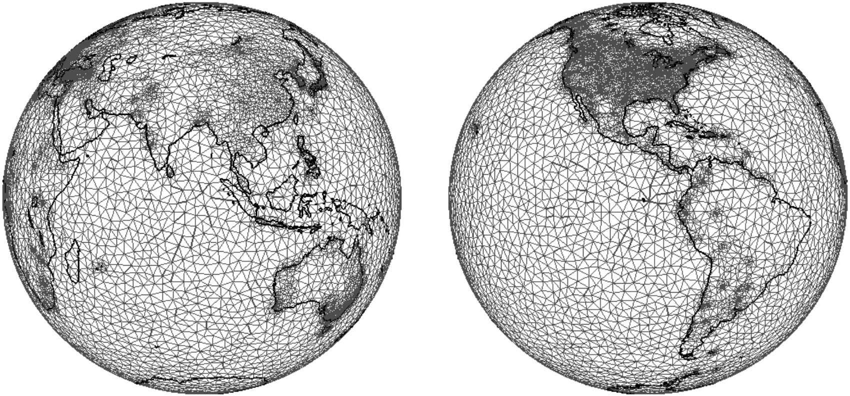
Simplified heat equation with spatially correlated noise

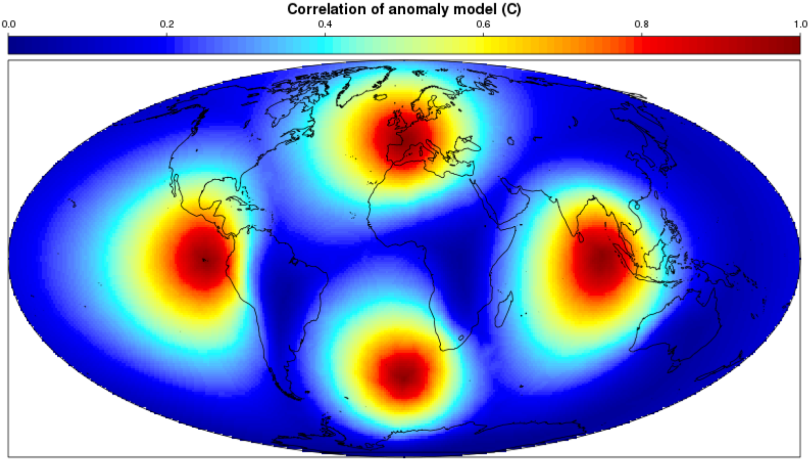
$$\gamma_t \dot{z}(\mathbf{s}, t) - \Delta z(\mathbf{s}, t) = \gamma_s^{-1/2} \mathcal{E}(\mathbf{s}, t)$$
$$\mathcal{E}(\mathbf{s}, \delta t) - \gamma_{\mathcal{E}} \Delta \mathcal{E}(\mathbf{s}, \delta t) = \mathcal{W}_{\mathcal{E}}(\mathbf{s}, \delta t)$$

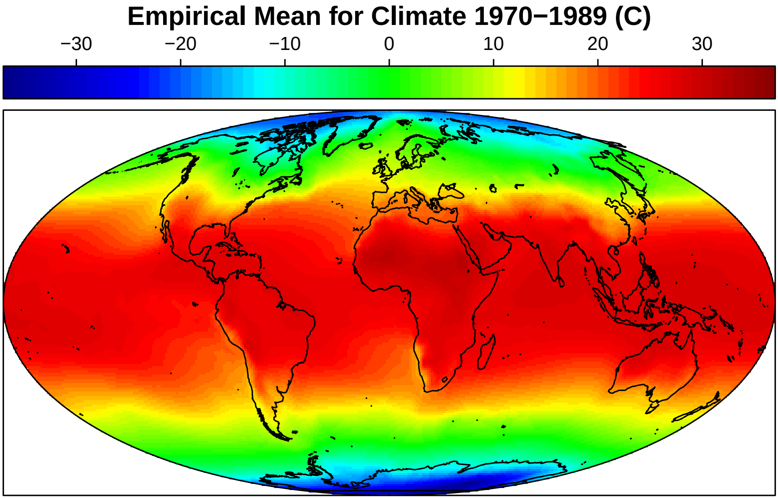
Precision

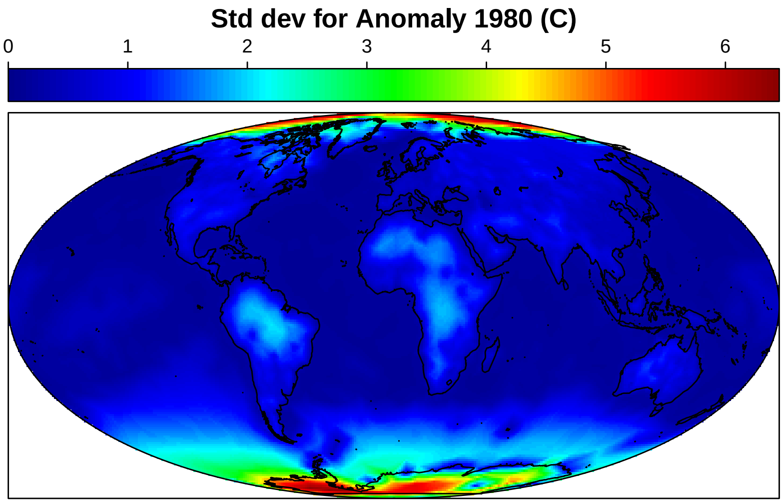
$$\mathbf{Q}_z = \gamma_s \left(\gamma_t^2 \mathbf{M}_0 + 2\gamma_t \mathbf{M}_1 + \mathbf{M}_2\right)$$
$$\mathbf{M}_0 = \mathbf{M}_2^{(t)} \otimes \mathbf{C}(\mathbf{I} + \gamma_{\mathcal{E}} \mathbf{C}^{-1}\mathbf{G})$$
$$\mathbf{M}_1 = \mathbf{M}_1^{(t)} \otimes \mathbf{G}(\mathbf{I} + \gamma_{\mathcal{E}} \mathbf{C}^{-1}\mathbf{G})$$
$$\mathbf{M}_2 = \mathbf{M}_0^{(t)} \otimes \mathbf{G}\mathbf{C}^{-1}\mathbf{G}(\mathbf{I} + \gamma_{\mathcal{E}} \mathbf{C}^{-1}\mathbf{G})$$
$$\mathbf{Q}_x = \phi^2 \mathbf{M}_0^{(t)} + 2\phi \mathbf{M}_1^{(t)} + \mathbf{M}_2^{(t)}, \quad \dot{x}(t) + \phi x(t) = \mathcal{W}(t)$$

Spherical triangulation GMRF/SPDE




Correlation of anomaly model (C)
0.0
0.2
0.4
0.6
0.8
1.0


Empirical Mean for Climate 1970−1989 (C)
−30
−20
−10
0
10
20
30


Std dev for Anomaly 1980 (C)
0
1
2
3
4
5
6

Practical computations: Precision structure

Problem: Large, ill-conditioned precision with interlocking blocks

Reparameterisation gives a more well behaved matrix

$$\mathbf{Q}_{(\mathbf{a},\mathbf{z})|\mathbf{y}} = \begin{bmatrix} \mathbf{I} \otimes \mathbf{Q}_a & 0 \\ 0 & \mathbf{Q}_z \end{bmatrix} + \begin{bmatrix} \mathbf{A}^T \\ (\mathbf{B}^T \otimes \mathbf{I})\mathbf{A}^T \end{bmatrix} \mathbf{Q}_\varepsilon \begin{bmatrix} \mathbf{A} & \mathbf{A}(\mathbf{B} \otimes \mathbf{I}) \end{bmatrix}$$

$$\mathbf{Q}_{(\mathbf{z}+\mathbf{a},\mathbf{z})|\mathbf{y}} = \begin{bmatrix} \mathbf{I} \otimes \mathbf{Q}_a + \mathbf{A}^T \mathbf{Q}_\varepsilon \mathbf{A} & -\mathbf{B} \otimes \mathbf{Q}_a \\ -\mathbf{B}^T \otimes \mathbf{Q}_a & \mathbf{Q}_z + (\mathbf{B}^T\mathbf{B}) \otimes \mathbf{Q}_a \end{bmatrix}$$

Block-diagonal preconditioner for iterative methods

$$\mathbf{M} = \begin{bmatrix} \mathbf{I} \otimes \mathbf{Q}_a + \mathbf{A}^T \mathbf{Q}_\varepsilon \mathbf{A} & 0 \\ 0 & \mathbf{Q}_z + (\mathbf{B}^T\mathbf{B}) \otimes \mathbf{Q}_a \end{bmatrix}$$

Variances of linear combinations

Using whatever can be computed

For precisions with sparse Cholesky factors, there is an algorithm to compute all covariances between neighbouring nodes $\widetilde{\boldsymbol{\Sigma}}$.

$$\mathrm{Var}(\mathbf{w}^\mathsf{T}\mathbf{x}) = \mathbf{w}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{w} = \mathbf{w}^\mathsf{T}\widetilde{\boldsymbol{\Sigma}}\mathbf{w}, \quad \text{if } w_i w_j = 0 \text{ for all } i \not\sim j$$

Use conditional distributions

Block-Rao-Blackwellised Monte Carlo integration

$$\mathrm{Var}(\mathbf{x}_1) \approx \mathrm{Var}(\mathbf{x}_1 \mid \mathbf{x}_2) + \frac{1}{N}\sum_{k=1}^{N}\left(\mathrm{E}(\mathbf{x}_1 \mid \mathbf{x}_2^{(k)}) - \mathrm{E}(\mathbf{x}_1)\right)^2$$

Rao-Blackwellisation of linear combinations

For ease of notation, let $\mathsf{E}(\mathbf{x}) = \mathbf{0}$

Use the model block structure

$$
\begin{aligned}
z = \mathbf{w}^\top \mathbf{x} &= \mathbf{w}_1^\top \mathbf{x}_1 + \mathbf{w}_2^\top \mathbf{x}_2 = z_1 + z_2 \\
\mathsf{Var}(z) &= \mathsf{E}(z_1^2 + z_2^2 + 2 z_1 z_2) \\
&= \mathsf{E}(v_1 + e_1^2 + z_2^2 + 2 e_1 z_2) \\
&= \mathsf{E}(v_1 + e_1^2 + v_2 + e_2^2 + 2 e_1 z_2) \\
v_1 &= \mathsf{Var}(z_1 | \mathbf{x}_2), \quad v_2 = \mathsf{Var}(z_2 | \mathbf{x}_1) \\
e_1 &= \mathsf{E}(z_1 | \mathbf{x}_2), \quad e_2 = \mathsf{E}(z_2 | \mathbf{x}_1)
\end{aligned}
$$

The conditional variances can be computed by applying the "$\widetilde{\mathbf{\Sigma}}$-method" to the precision sub-blocks.

Rao-Blackwellisation of linear combinations

Which cross-products give the smallest MC error?

$$
\begin{aligned}
e_{11} &= \mathsf{E}(e_1 e_1), \quad s_{11} = \mathsf{E}(z_1 z_1) = v_1 + e_{11} \\
e_{12} &= \mathsf{E}(e_1 e_2), \quad s_{12} = \mathsf{E}(z_1 z_2) \\
e_{22} &= \mathsf{E}(e_2 e_2), \quad s_{22} = \mathsf{E}(z_2 z_2) = v_2 + e_{22} \\
\mathsf{Var}(z) &= s_{11} + s_{22} + 2 s_{12} \\
\mathsf{Var}\left(\begin{bmatrix} z_1 \\ e_1 \\ z_2 \\ e_2 \end{bmatrix} \right) &= \begin{bmatrix} s_{11} & e_{11} & s_{12} & s_{12} \\ e_{11} & e_{11} & s_{12} & e_{12} \\ s_{12} & s_{12} & s_{22} & e_{22} \\ s_{12} & e_{12} & e_{22} & e_{22} \end{bmatrix}
\end{aligned}
$$

Example: Linear regression

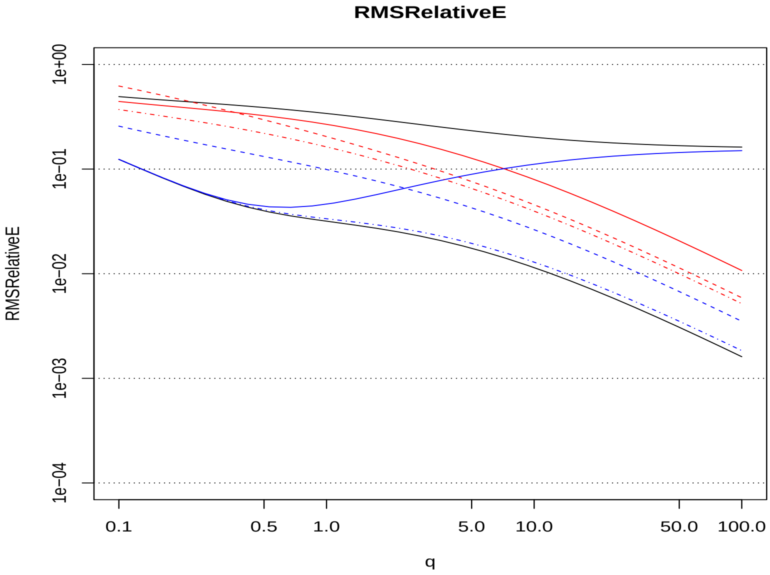
A toy example with structure similar to the climate model

- Coefficients for trend and a nuisance covariate: $\mathbf{x}_2 \sim N(0, \tau_2^{-1}\mathbf{I}_3)$
- True values: $(\mathbf{x}_1|\mathbf{x}_2) \sim N(\mathbf{B}\mathbf{x}_2, \tau_1^{-1}\mathbf{I}_n)$
- Measurements: $(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) \sim N(\mathbf{x}_1, q^{-1}\mathbf{I}_n)$
- Posterior precision ($\tau_1 = 1$, $\tau_2 = 0.01$)

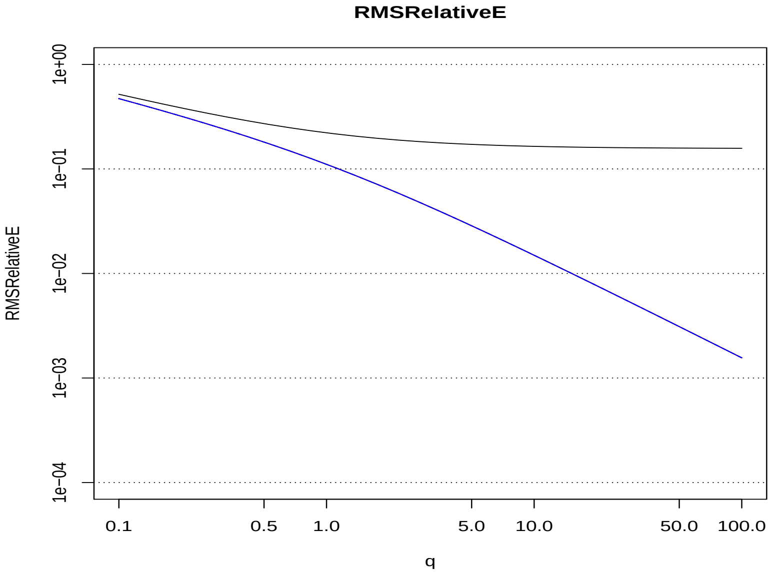
$$\mathbf{Q}_{\mathbf{x}|\mathbf{y}} = \begin{bmatrix} (\tau_1 + q)\mathbf{I}_n & -\tau_1\mathbf{B} \\ -\tau_1\mathbf{B}^\mathsf{T} & \tau_2\mathbf{I}_3 + \tau_1\mathbf{B}^\mathsf{T}\mathbf{B} \end{bmatrix}$$

- Linear combination weights $\mathbf{w}_1 = (1, 0, 0, \ldots, 0)$, $\mathbf{w}_2 = (B_{11}, B_{12}, 0)$

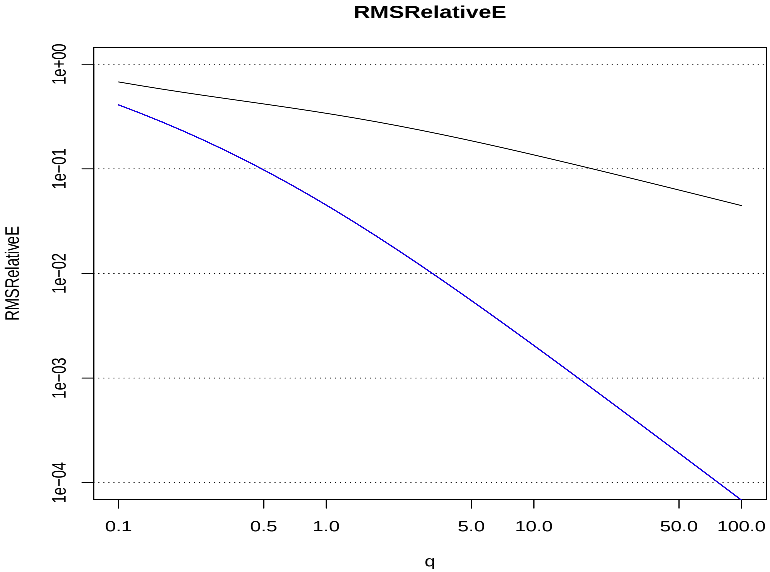
Root mean square of relative MC errors



MC-RMSE for "Anomaly uncertainty", $\mathbf{w}_1 = \mathbf{0}$



MC-RMSE for "Climate uncertainty", $\mathbf{w}_2 = \mathbf{0}$



References

- 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)*, Journal of the Royal Statistical Society, Series B, 73(4), 423–498.
- D. Simpson, F. Lindgren and H. Rue (2012), *In order to make spatial statistics computationally feasible, we need to forget about the covariance function*, Environmetrics, 23: 65–74.
- `http://www.r-inla.org/`