Stochastic PDEs for computationally efficient climate reconstruction

Finn Lindgren

UNIVERSITY OF BATH

Bath 2014-05-20

The many disguises of random fields and the resurrection of useful results

Finn Lindgren

UNIVERSITY OF BATH

Bath 2014-05-20

Spatial statistics on the globe

Hierarchical spatial models (and inverse problems)

Hierarchical model

$\boldsymbol{\theta}$	Model parameters
$\boldsymbol{u}\|\boldsymbol{\theta}$	Random, latent processes; spatial or spatio-temporal fields
$\boldsymbol{y}\|\boldsymbol{\theta}, \boldsymbol{u}$	Measured data

Simple spatial statistics framework

- Spatial domain Ω, or space-time domain $\Omega \times \mathbb{T}$, $\mathbb{T} \subset \mathbb{R}$.
- Random field $u(\boldsymbol{s})$, $\boldsymbol{s} \in \Omega$, or $u(\boldsymbol{s}, t)$, $(\boldsymbol{s}, t) \in \Omega \times \mathbb{T}$.
- Observations $y_i = u(\boldsymbol{s}_i) + \epsilon_i$, with $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon)$.

Two basic model and method components

- We need stochastic models for $u(\cdot)$.
- We need computationally efficient (Bayesian) inference methods for the posterior distributions for $\boldsymbol{\theta}$ and $u(\cdot)$ given data $\boldsymbol{y}$.

Covariance functions and stochastic PDEs

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

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

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

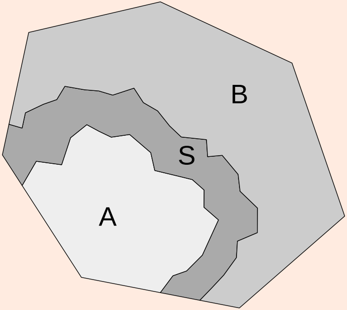
$$\left(\kappa^2 - \nabla \cdot \nabla\right)^{\alpha/2} u(\boldsymbol{s}) = \mathcal{W}(\boldsymbol{s}), \quad \alpha = \nu + d/2$$

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

Spectrum and the continuous global Markov property

Markov condition and spectral densities

Global Markov propery on a manifold:
For any separating set S 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.
(Rozanov, 1977)

Proof of the Matérn/Whittle equivalence and the Markov connection:

$$S(\boldsymbol{\omega}) = \mathcal{F}R(\cdot) = \frac{1}{(2\pi)^d(\kappa^2+\|\boldsymbol{\omega}\|^2)^\alpha}$$

Key fact: For any finite-dimensional Gaussian random field, the non-zero pattern of the precision matrix $\boldsymbol{Q} = \boldsymbol{\Sigma}^{-1}$ defines a graph on which the global Markov property holds. The reverse is also true.

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/GMRF	$\psi_k(\boldsymbol{s})$

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/GMRF	$\propto \sum_k \psi_k(\boldsymbol{s}) u_k$	$\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_\kappa^{-1})$

Continuous domain Markov approximations

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 Gaussian model

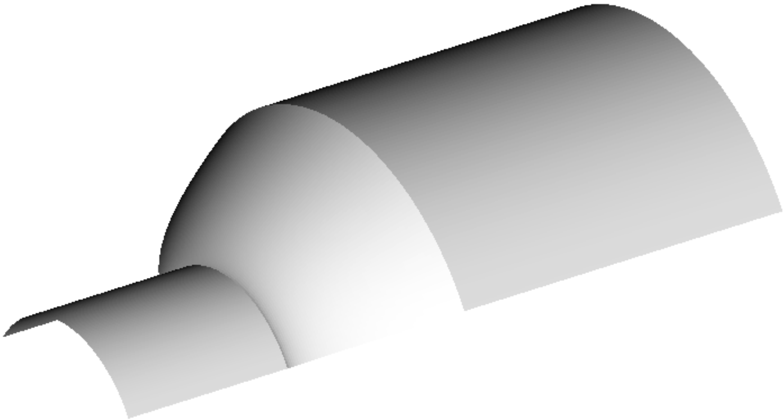
$$\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{\mu}_u, \boldsymbol{Q}_u^{-1}), \quad \boldsymbol{y}|\boldsymbol{u} \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{u}, \boldsymbol{Q}_{y|u}^{-1}) \qquad (A_{ij} = \psi_j(\boldsymbol{s}_i))$$

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

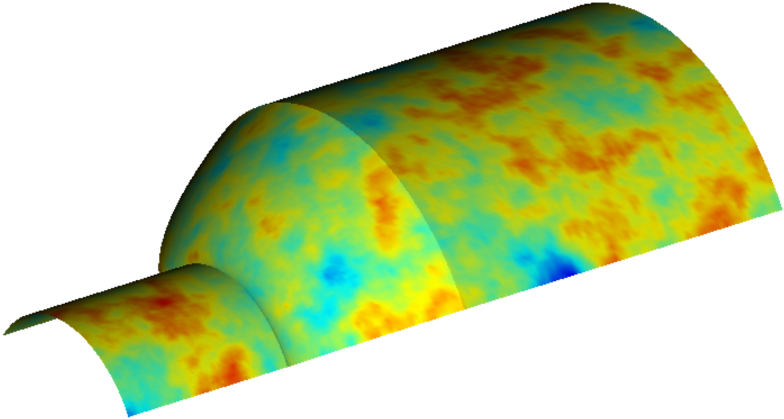
$$\boldsymbol{Q}_{u|y} = \boldsymbol{Q}_u + \boldsymbol{A}^T \boldsymbol{Q}_{y|u} \boldsymbol{A}$$ (∼"Sparse iff ψ_k have compact support")

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

Connection with the deformation method for non-stationarity

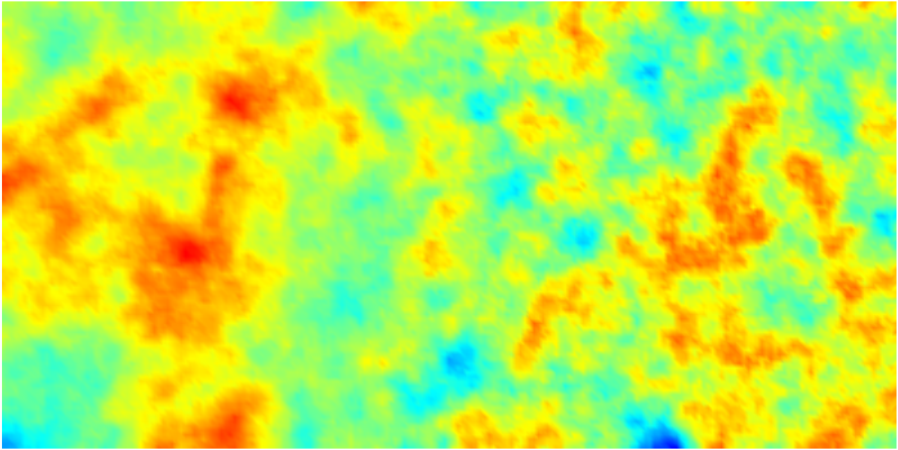


"Stationary" field on a deformed manifold $\widetilde{\Omega}$



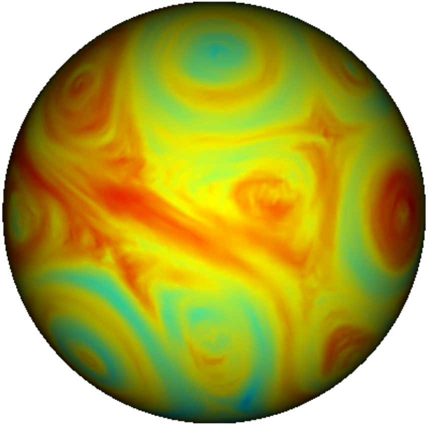
$$(1 - \widetilde{\nabla} \cdot \widetilde{\nabla})\widetilde{u}(\widetilde{\boldsymbol{s}}) = \widetilde{\mathcal{W}}(\widetilde{\boldsymbol{s}}), \quad \widetilde{\boldsymbol{s}} \in \widetilde{\Omega}$$

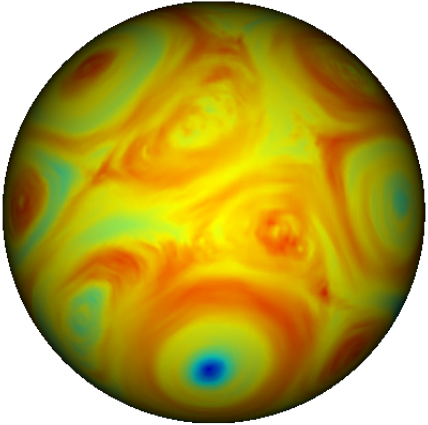
Non-stationary field on original manifold Ω



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

Anisotropic field on a globe via change of manifold metric

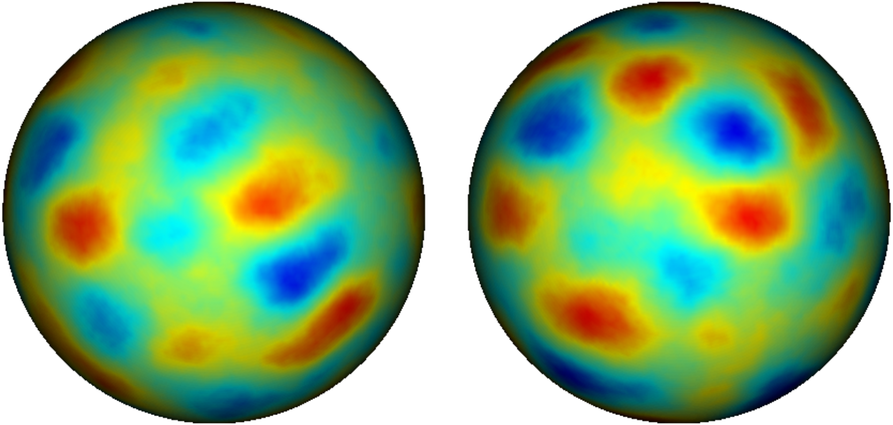




Corresponds to a non-stationary SPDE operator:

$$(\kappa_{\boldsymbol{s}}^2 + \nabla \cdot \boldsymbol{m}_{\boldsymbol{s}} - \nabla \cdot \boldsymbol{M}_{\boldsymbol{s}} \nabla)(\tau_{\boldsymbol{s}} u(\boldsymbol{s})) = \gamma_{\boldsymbol{s}} \mathcal{W}(\boldsymbol{s})$$

Oscillating fields



$$(\kappa^2 \, e^{i\pi\theta} - \nabla \cdot \nabla)(u_R(\boldsymbol{s}) + i u_I(\boldsymbol{s})) = \mathcal{W}_R(\boldsymbol{s}) + i\mathcal{W}_I(\boldsymbol{s})$$

Simulation free inference with Laplace approximations

Quadratic posterior log-likelihood approximation

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

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

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

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

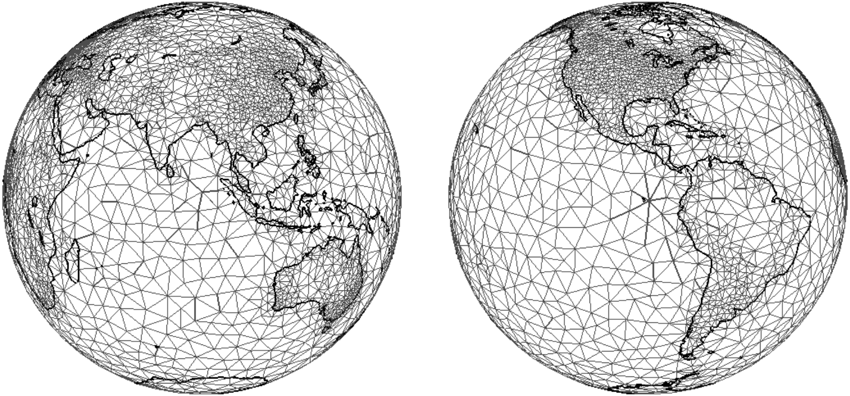
Direct Bayesian inference with INLA (`r-inla.org`)

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

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

Key observation: No sampling is required, in principle.

Triangulation partly adapted to the data density



Linear model for weather observations

Weather = Climate + Anomaly

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

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

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

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

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

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

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

Stochastic weather anomaly model

Non-stationary spatial SPDE

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

Precision

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

Stochastic climate model

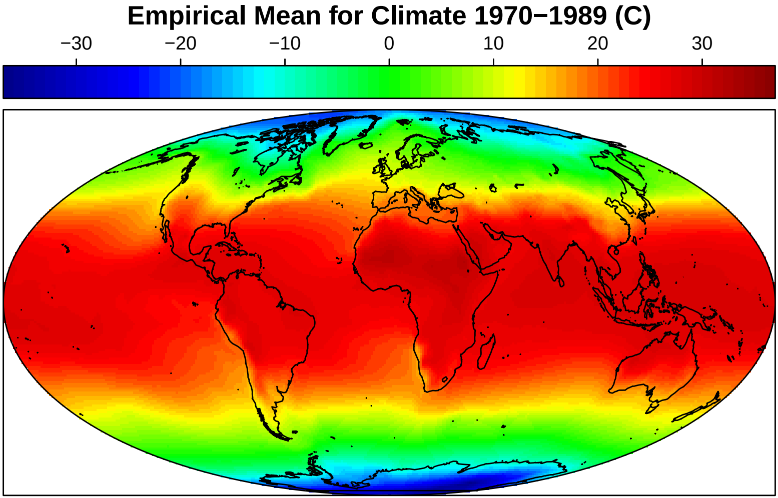
Simplified heat equation

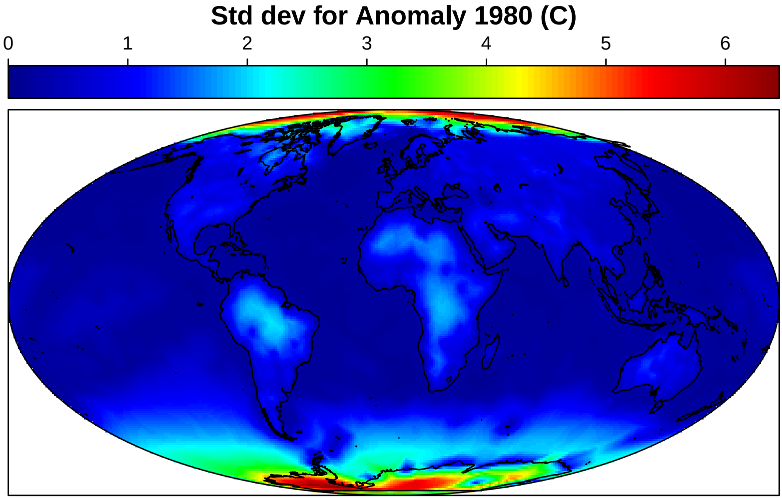
$$\gamma_t \dot{z}(\mathbf{s}, t) - \Delta z(\mathbf{s}, t) = \gamma_s^{-1/2} \mathcal{E}(\mathbf{s}, t)$$
$$\mathcal{E}(\mathbf{s}, \delta t) - \gamma_{\mathcal{E}} \Delta \mathcal{E}(\mathbf{s}, \delta t) = \mathcal{W}_{\mathcal{E}}(\mathbf{s}, \delta t)$$

Note: An *iterated* heat equation fits the same framework.

Precision

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


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

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

Approximate Schur-complement is an alternative.

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 $\widetilde{\boldsymbol{\Sigma}}$.

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

Use conditional distributions

Block-Rao-Blackwellised Monte Carlo integration

$$\begin{aligned}\mathsf{Var}(\mathbf{x}_1) &= \mathsf{E}(\mathsf{Var}(\mathbf{x}_1 \mid \mathbf{x}_2)) + \mathsf{Var}(\mathsf{E}(\mathbf{x}_1 \mid \mathbf{x}_2)) \\ &\approx \mathsf{Var}(\mathbf{x}_1 \mid \mathbf{x}_2) + \frac{1}{N}\sum_{k=1}^{N}\Big(\mathsf{E}(\mathbf{x}_1 \mid \mathbf{x}_2^{(k)}) - \mathsf{E}(\mathbf{x}_1)\Big)^2\end{aligned}$$

for samples $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(N)}$.

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 obtained from a pre-computed "$\widetilde{\Sigma}$-method" for each sub-block, or pre-computed sub-block solves.

Rao-Blackwellisation of linear combinations

Which cross-products give the smallest MC error?

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

$$
\mathsf{Var}\left(\begin{bmatrix} z_1 \\ z_2 \\ e_1 \\ e_2 \end{bmatrix} \right) = \begin{bmatrix} s_{11} & s_{12} & e_{11} & s_{12} \\ s_{12} & s_{22} & s_{12} & e_{22} \\ e_{11} & s_{12} & e_{11} & e_{12} \\ s_{12} & e_{22} & e_{12} & e_{22} \end{bmatrix}
$$

There's an adorable *partially observed Wishart* inference problem hiding here!

Example: Linear regression

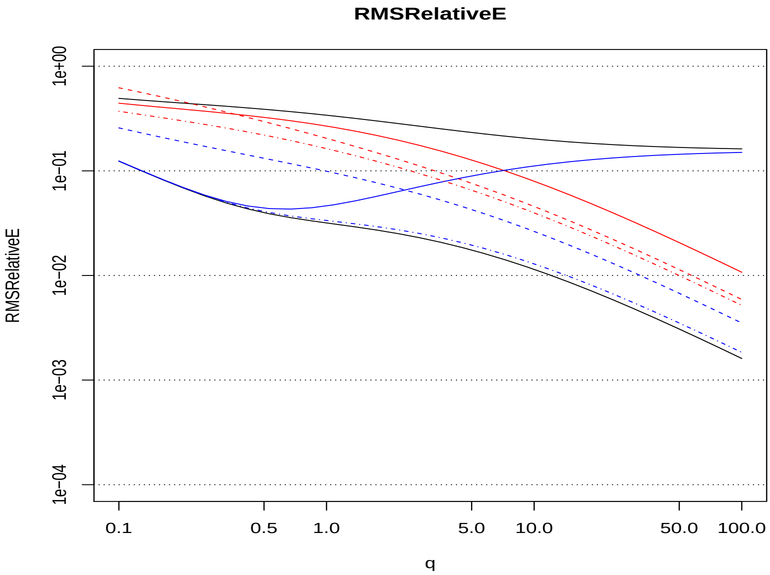
A toy example with structure similar to the climate model

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

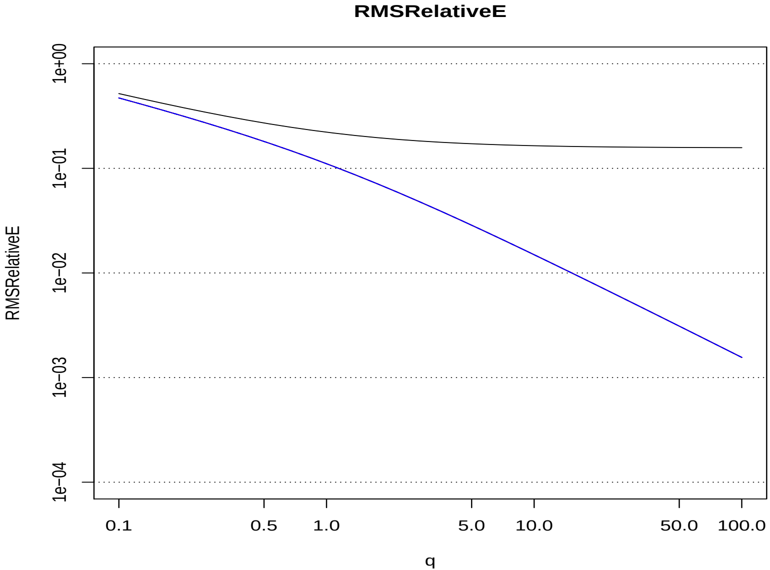
$$\mathbf{Q}_{\mathbf{x}|\mathbf{y}} = \begin{bmatrix} (\tau_1 + q)\mathbf{I}_n & -\tau_1 \boldsymbol{B} \\ -\tau_1 \boldsymbol{B}^\top & \tau_2\mathbf{I}_3 + \tau_1 \boldsymbol{B}^\top \boldsymbol{B} \end{bmatrix}$$

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

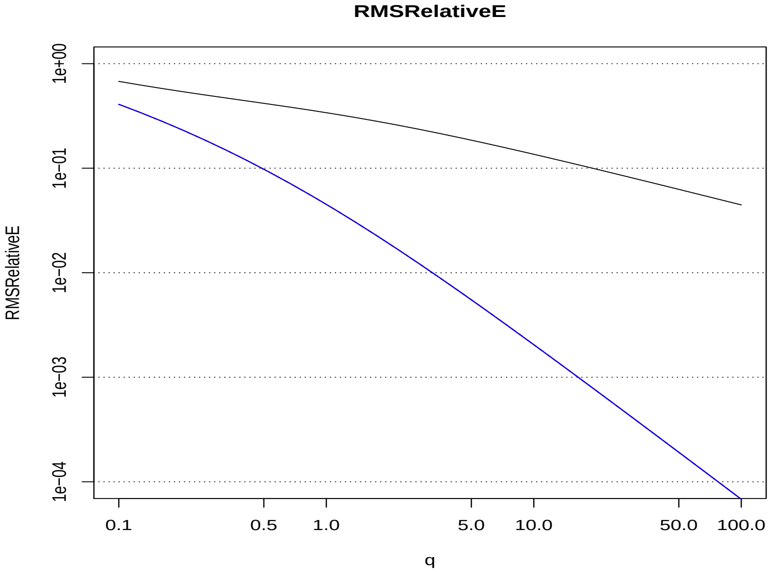
Root mean square of relative MC errors



MC-RMSE for “Anomaly uncertainty”, $\mathbf{w}_2 = \mathbf{0}$



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



Current and future challenges

- Stochastic boundary conditions
- Higher order basis functions (improved approximation accuracy; state-space formulation for smaller Markov structure)
- Sums of Markov models; Multiresolution methods
  - LatticeKrig (CRAN): Local, smooth basis functions, in a multiresolution hierarchy; uses direct solvers
  - Climate and weather modelling; multiple temporal and spatial scales;
- Multigrid methods for very large space-time problems; *In theory*, $\mathcal{O}(n^{3/2})$ (space) and $\mathcal{O}(n^2)$ (space-time) becomes $\mathcal{O}(n)$
- Issue: Marginal variances are available from Cholesky factors, but not directly from iterative solvers; pure Monte Carlo estimators too expensive and/or imprecise.
- Practical non-isotropic non-stationarity

References

References (see also `r-inla.org`)

- 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.
- D. Bolin, F. Lindgren (2013), *Excursion and contour uncertainty regions for latent Gaussian models*, JRSSB, in press. Accepted version online and at arXiv:1211.3946. CRAN package: `excursions`
- R. Ingebrigtsen, F. Lindgren, I. Steinsland (2013), *Spatial models with explanatory variables in the dependence structure*, Spatial Statistics, In Press (available online).
- G-A. Fuglstad, F. Lindgren, D. Simpson, H. Rue (2013), *Exploring a new class of non-stationary spatial Gaussian random fields with varying local anisotropy*, accepted for Statistica Sinica, arXiv:1304.6949