Computational latent Gaussian process methods for weather and climate reconstruction

Finn Lindgren

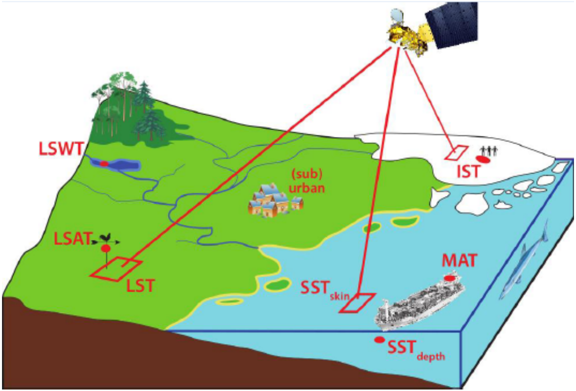
THE UNIVERSITY *of* EDINBURGH

Sheffield 2016-12-08

EUSTACE has received funding from the European Union's Horizon 2020 Programme for Research and Innovation, under Grant Agreement no 640171

EUSTACE
EU Surface Temperatures for All Corners of Earth

EUSTACE will give publicly available daily estimates of surface air temperature since 1850 across the globe for the first time by combining surface and satellite data using novel statistical techniques.



THE UNIVERSITY of EDINBURGH

EUSTACE

Spatial fields, observations, and stochastic models

- Partially observed spatial functions (temperature) or objects related to *latent* spatial functions
- Wanted: estimates of the true values at observed and unobserved locations
- Wanted: quantified uncertainty about those values
- Complex measurement errors can be modeled using hierarchical random effects

Spatial hierarchical model framework

- Observations $\boldsymbol{y} = \{y_i, i = 1, \ldots, n_y\}$
- Latent random field $x(\mathbf{s})$, $\mathbf{s} \in \Omega$
- Model parameters $\boldsymbol{\theta} = \{\theta_j, j = 1, \ldots, n_\theta\}$

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

$$\begin{aligned} \mathsf{E}(x(\mathbf{s})) &= m(\mathbf{s}), \\ \mathsf{Cov}(x(\mathbf{s}), x(\mathbf{s}')) &= K(\mathbf{s}, \mathbf{s}'), \\ \left[x(\mathbf{s}_i), i = 1, \ldots, n\right] &\sim \mathcal{N}\big(\boldsymbol{m} = \left[m(\mathbf{s}_i), i = 1, \ldots, n\right], \\ &\qquad \boldsymbol{\Sigma} = \left[K(\mathbf{s}_i, \mathbf{s}_j), i, j = 1, \ldots, n\right]\big) \end{aligned}$$

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

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

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

for all $f, g \in H_{\mathcal{R}}(D) \equiv \{f : D \mapsto \mathbb{R};\ \langle f, \mathcal{R} f \rangle_D < \infty\}$.

Covariance functions and SPDEs

The Matérn covariance family on

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

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

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

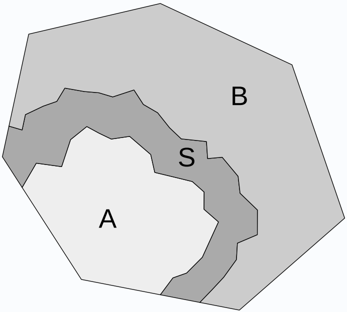
$\mathcal{W}(\cdot)$ white noise, $\nabla \cdot \nabla = \sum_{i=1}^{d} \frac{\partial^2}{\partial s_i^2}$, $\sigma^2 = \frac{\Gamma(\nu)}{\Gamma(\alpha)\kappa^{2\nu}(4\pi)^{d/2}}$

White noise has $K(\mathrm{s}, \mathrm{s}') = \delta(\mathrm{s} - \mathrm{s}')$.

Markov in space

Markov properties

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



Solutions to

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

are Markov when α is an integer.

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

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

$\boldsymbol{Q}_{AB} = \boldsymbol{0}$

$\boldsymbol{Q}_{A|S,B} = \boldsymbol{Q}_{AA}$

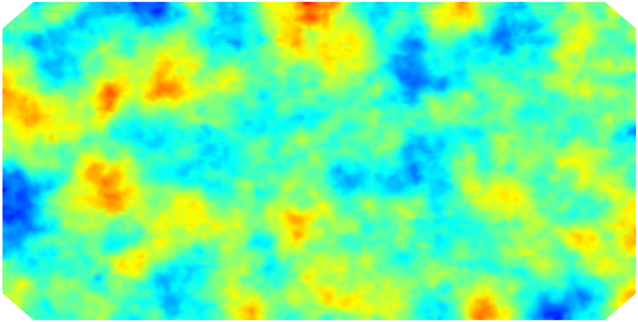
$\boldsymbol{\mu}_{A|S,B} = \boldsymbol{\mu}_A - \boldsymbol{Q}_{AA}^{-1}\boldsymbol{Q}_{AS}(\boldsymbol{u}_S - \boldsymbol{\mu}_S)$

If we use local basis function expansions, we can exploit the continuous Markov property as sparse numerical matrix algebra.

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

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

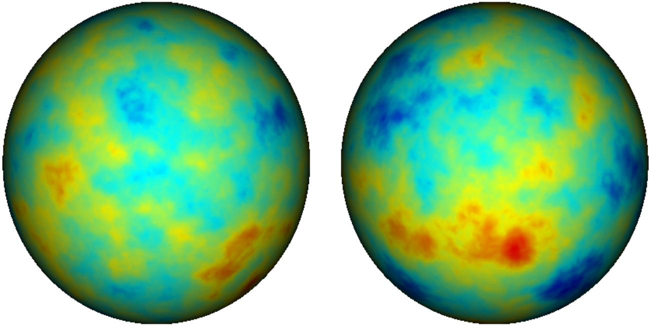
$$(\kappa^2 - \Delta)(\tau\, x(\mathbf{s})) = \mathcal{W}(\mathbf{s}), \quad \mathbf{s} \in \mathbb{R}^d$$



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

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

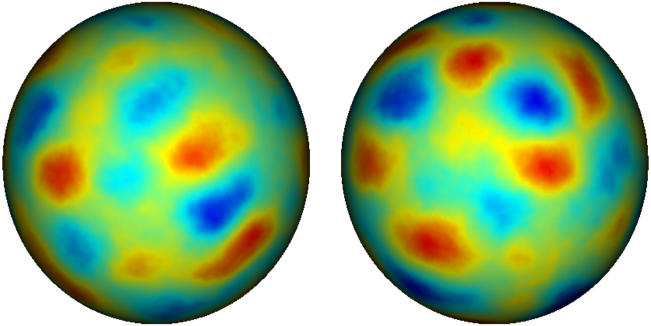
$$(\kappa^2 - \Delta)(\tau\, x(\mathbf{s})) = \mathcal{W}(\mathbf{s}), \quad \mathbf{s} \in \Omega$$



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

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

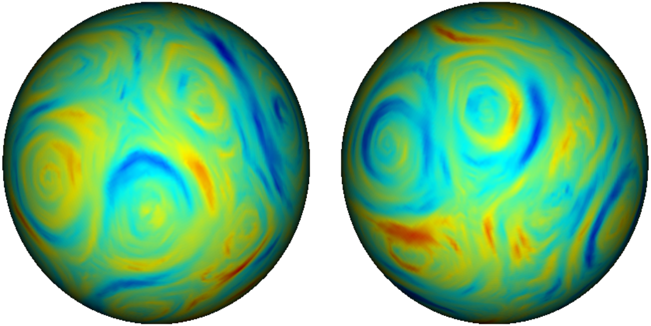
$$(\kappa^2 \, e^{i\pi\theta} - \Delta)(\tau \, x(\mathbf{s})) = \mathcal{W}(\mathbf{s}), \quad \mathbf{s} \in \Omega$$



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

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

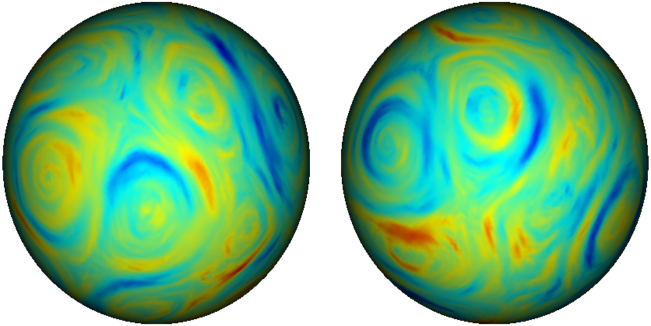
$$(\kappa_{\mathbf{s}}^2 + \nabla \cdot \boldsymbol{m}_{\mathbf{s}} - \nabla \cdot \boldsymbol{M}_{\mathbf{s}} \nabla)(\tau_{\mathbf{s}} x(\mathbf{s})) = \mathcal{W}(\mathbf{s}), \quad \mathbf{s} \in \Omega$$



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

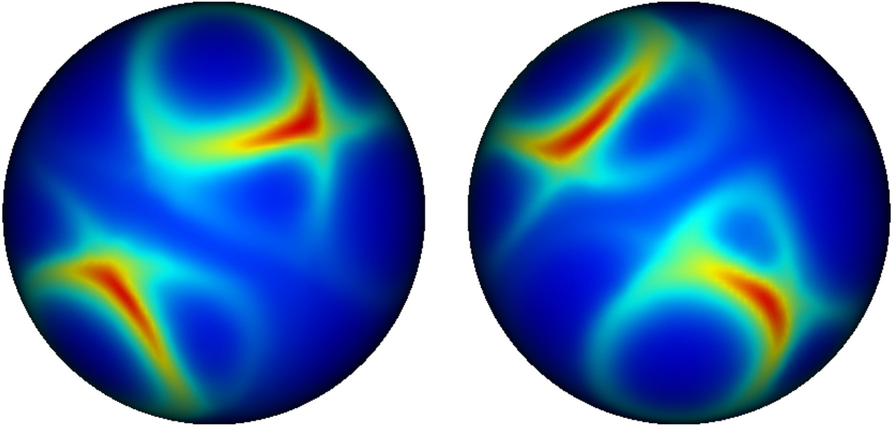
GMRF representations of SPDEs can be constructed for oscillating, anisotropic, non-stationary, non-separable spatio-temporal, and multivariate fields on manifolds.

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



Covariances for four reference points

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



Basis function SPDE representations

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 $x(\boldsymbol{s})$	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	$\propto \sum_k \psi_k(\boldsymbol{s}) x_k$	$\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_\kappa^{-1})$

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

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.

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

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

The covariance for the RHS of the SPDE is

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

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

Hierarchical models

Continuous Markovian spatial models (Lindgren et al, 2011)

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

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

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

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

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

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

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

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

The computational GMRF work-horse

Cholesky decomposition (Cholesky, 1924)

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

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

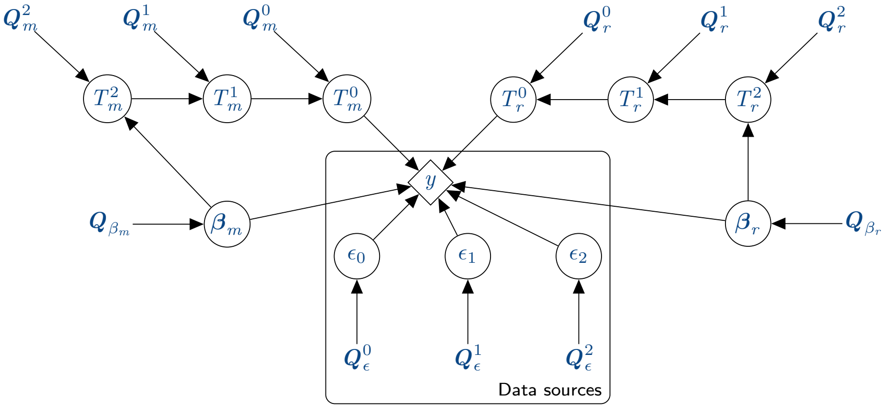
$$\log\det \boldsymbol{Q} = 2\log\det \boldsymbol{L} = 2\sum_i \log L_{ii}$$

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)

Partial hierarchical representation

Observations of *mean*, *max*, *min*. Model *mean* and *range*.



Conditional specifications, e.g.

$$(T_m^0 \mid T_m^1, Q_m^0) \sim \mathcal{N}\left(T_m^1,\ {Q_m^0}^{-1}\right)$$

Basic latent multiscale structure

Let $U_m^k(\mathbf{s},t)$, $U_r^k(\mathbf{s},t)$, $k = 0, 1, 2, S$ be random fields operating on (multi)daily, multimonthly, multidecadal, and cyclic seasonal timescales, respectively, represented by finite element approximations of stochastic heat equations.

Daily mean temperatures

The daily means $T_m(\mathbf{s},t)$ are defined through

$$T_m(\mathbf{s},t) = \underbrace{U_m^0(\mathbf{s},t) + \underbrace{U_m^1(\mathbf{s},t) + \underbrace{U_m^2(\mathbf{s},t) + U_m^S(\mathbf{s},t) + \sum_{i=1}^{N_X} X_i(\mathbf{s},t)\beta_m^{(i)}}_{T_m^2}}_{T_m^1}}_{T_m^0}$$

The β_m coefficients are weights for covariates $X_i(\mathbf{s},t)$ (e.g. elevation, topographical gradients, and land use indicator functions).

Basic latent multiscale structure

Daily temperature range (diurnal range)

The diurnal ranges $T_r(\mathbf{s}, t)$ are defined through

$$g^{-1}[\mu_r(\mathbf{s}, t)] = \underbrace{U_r^1(\mathbf{s}, t) + \underbrace{U_r^2(\mathbf{s}, t) + U_r^S(\mathbf{s}, t) + \sum_{i=1}^{N_X} X_i(\mathbf{s}, t)\beta_r^{(i)}}_{T_r^2}}_{T_r^1},$$

$$T_r(\mathbf{s}, t) = \mu_r(\mathbf{s}, t)\ G\left[U_r^0(\mathbf{s}, t)\right] = \underbrace{g(T_r^1)\ G\left[U_r^0(\mathbf{s}, t)\right]}_{T_r^0},$$

where the slowly varying median process $\mu_r(\mathbf{s}, t)$ is a transformed multiscale model, and G is a non-linear transformation function, controlled by some fixed seasonal fields of distribution scale and shape parameters. The β_m and β_r coefficients are weights for covariates $X_i(\mathbf{s}, t)$ (e.g. elevation, topographical gradients, and land use indicator functions).

Observation models

Satellite data error model

The observational&calibration errors are modelled as three error components:
independent (ϵ_0), spatially correlated (ϵ_1), and systematic (ϵ_2), with distributions determined by the uncertainty information from WP1
E.g., $y_i = T_m(\mathbf{s}_i, t_i) + \epsilon_0(\mathbf{s}_i, t_i) + \epsilon_1(\mathbf{s}_i, t_i) + \epsilon_2(\mathbf{s}_i, t_i)$

Station homogenisation

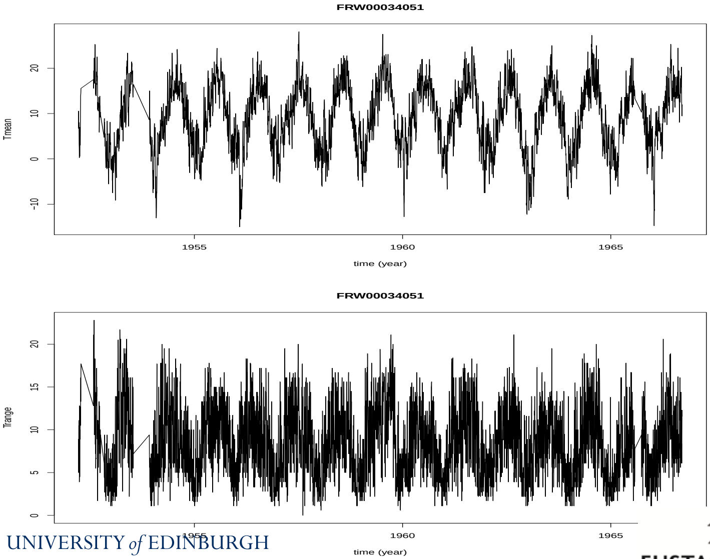
For station k at day t_i

$$y_m^{k,i} = T_m(\mathbf{s}_k, t_i) + \sum_{j=1}^{J_k} H_j^k(t_i) e_m^{k,j} + \epsilon_m^{k,i},$$

where $H_j^k(t)$ are temporal step functions, $e_m^{k,j}$ are latent bias variables, and $\epsilon_m^{k,i}$ are independent measurement and discretisation errors.

Observed data

Observed daily T_{mean} and T_{range} for station FRW00034051



Power tail quantile (POQ) model

The quantile function (inverse cumulative distribution function) $F_{\boldsymbol{\theta}}^{-1}(p)$, $p \in [0,1]$, is defined through

$$f_\theta^-(p) = \begin{cases} \frac{1-(2p)^{-\theta}}{2\theta}, & \theta \neq 0, \\ \frac{1}{2}\log(2p), & \theta = 0, \end{cases}$$

$$f_\theta^+(p) = -f_\theta^-(1-p) = \begin{cases} \frac{(2(1-p))^{-\theta}-1}{2\theta}, & \theta \neq 0, \\ -\frac{1}{2}\log(2(1-p)), & \theta = 0. \end{cases}$$

$$F_{\boldsymbol{\theta}}^{-1}(p) = \theta_0 + \frac{\tau}{2}\left[(1-\gamma)f_{\theta_3}^-(p) + (1+\gamma)f_{\theta_4}^+(p)\right],$$

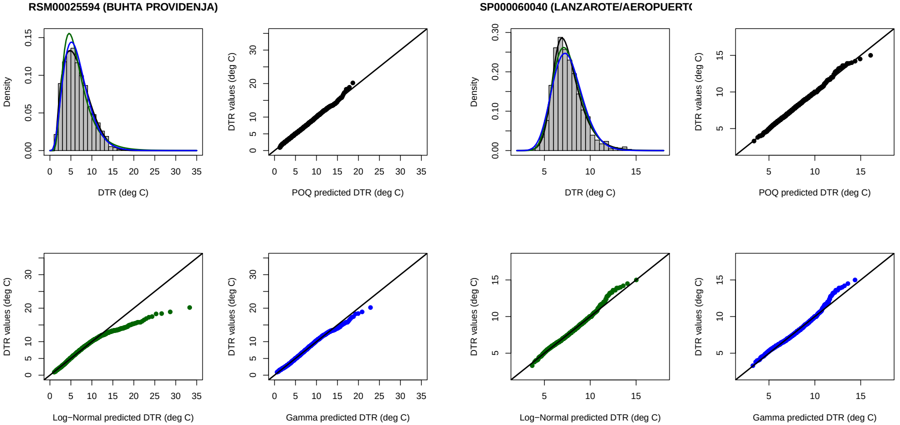
The parameters $\boldsymbol{\theta} = (\theta_0, \theta_1 = \log\tau, \theta_2 = \operatorname{logit}[(\gamma+1)/2], \theta_3, \theta_4)$ control the median, spread/scale, skewness, and the left and right tail shape. This model is also known as the *five parameter lambda model*.

A spatio-temporally dependent Gaussian field $u(\mathbf{s}, t)$ with expectation 0 and variance 1 can be transformed into a POQ field by

$$\widetilde{u}(\mathbf{s}, t) = F_{\boldsymbol{\theta}(\mathbf{s},t)}^{-1}(\Phi(u(\mathbf{s}, t))),$$

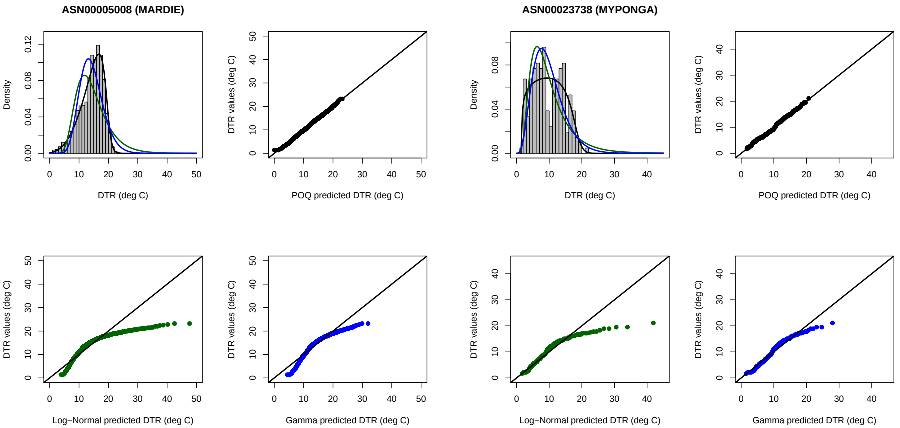
where the parameters can vary with space and time.

Diurnal range distributions



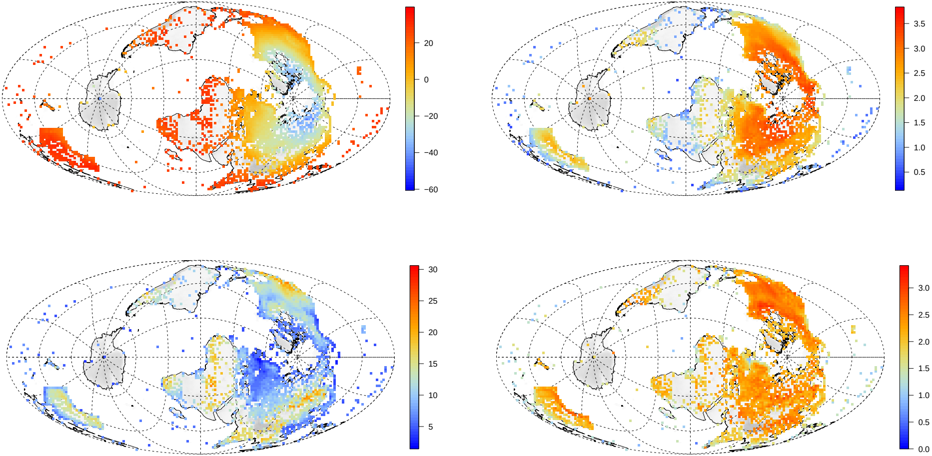
For these stations, POQ does a slightly better job than a Gamma distribution.

Diurnal range distributions



For these stations only POQ comes close to representing the distributions. Note: Some of the mixture-like distribution shapes may be an effect of unmodeled station inhomogeneities.

Median & scale for daily means and ranges

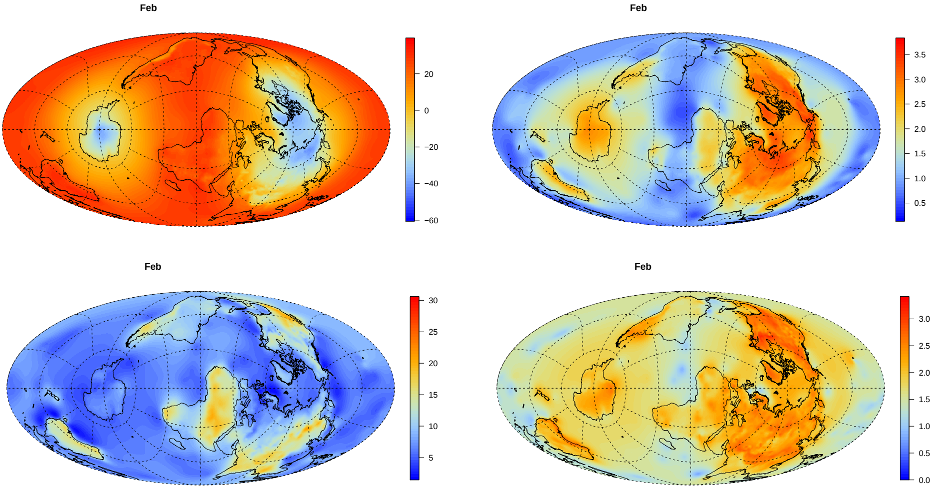


February climatology

THE UNIVERSITY of EDINBURGH

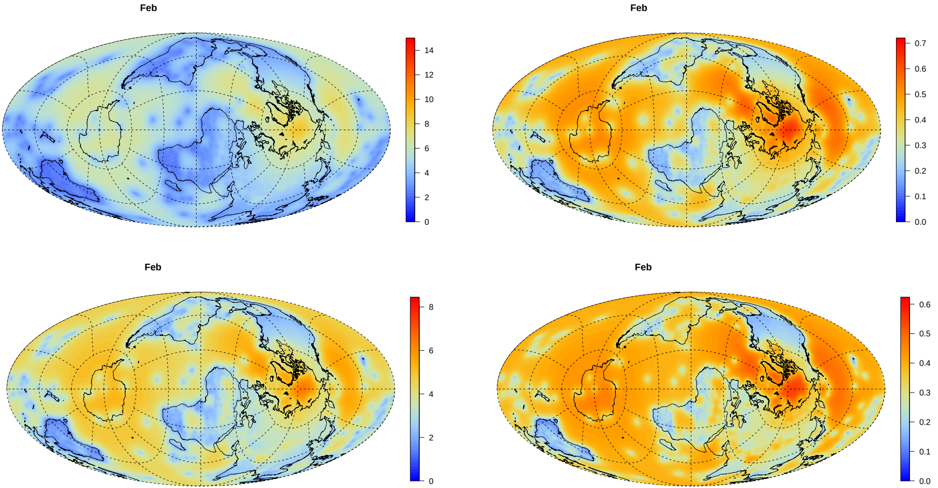
EUSTACE

Estimates of median & scale for T_m and T_r



February climatology

Std.dev. of median & scale for T_m and T_r



February climatology

THE UNIVERSITY of EDINBURGH

EUSTACE

Linearised inference

All Spatio-temporal latent random processes combined into $\boldsymbol{x} = (\boldsymbol{u}, \boldsymbol{\beta}, \boldsymbol{b})$, with joint expectation $\boldsymbol{\mu}_x$ and precision $\boldsymbol{Q}_x$:

$$(\boldsymbol{x} \mid \boldsymbol{\theta}) \sim \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{Q}_x^{-1}) \qquad \text{(Prior)}$$

$$(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}) \sim \mathcal{N}(h(\boldsymbol{x}), \boldsymbol{Q}_{y|x}^{-1}) \qquad \text{(Observations)}$$

$$p(\boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{\theta}) \propto p(\boldsymbol{x} \mid \boldsymbol{\theta})\, p(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}) \qquad \text{(Posterior)}$$

Linear Gaussian observations

In a linear model with $h(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$,

$$(\boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{\theta}) \sim \mathcal{N}(\widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{Q}}^{-1}) \qquad \text{(Posterior)}$$

$$\widetilde{\boldsymbol{Q}} = \boldsymbol{Q}_{x\beta b} + \boldsymbol{A}^\top \boldsymbol{Q}_{y|x} \boldsymbol{A}$$

$$\widetilde{\boldsymbol{\mu}} = \boldsymbol{\mu}_x + \widetilde{\boldsymbol{Q}}^{-1} \boldsymbol{A}^\top \boldsymbol{Q}_y (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\mu}_x)$$

Linearised inference

All Spatio-temporal latent random processes combined into $\boldsymbol{x} = (\boldsymbol{u}, \boldsymbol{\beta}, \boldsymbol{b})$, with joint expectation $\boldsymbol{\mu}_x$ and precision $\boldsymbol{Q}_x$:

$$(\boldsymbol{x} \mid \boldsymbol{\theta}) \sim \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{Q}_x^{-1}) \qquad \text{(Prior)}$$

$$(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}) \sim \mathcal{N}(h(\boldsymbol{x}), \boldsymbol{Q}_{y|x}^{-1}) \qquad \text{(Observations)}$$

$$p(\boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{\theta}) \propto p(\boldsymbol{x} \mid \boldsymbol{\theta})\, p(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}) \qquad \text{(Posterior)}$$

Non-linear and/or non-Gaussian observations

Linearise at $\widetilde{\boldsymbol{\mu}}$ and iterate:

$$(\boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{\theta}) \overset{\text{approx}}{\sim} \mathcal{N}(\widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{Q}}^{-1}) \qquad \text{(Approximate posterior)}$$

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

$$\widetilde{\boldsymbol{Q}} = \boldsymbol{Q}_x - \nabla_{\boldsymbol{x}}^2 \ln p(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}) \big|_{\boldsymbol{x}=\widetilde{\boldsymbol{\mu}}}$$

Products of transformed processes

Assume that $\boldsymbol{u}$ is a large scale process and $\boldsymbol{v}$ is a small scale process, so that they are statistically identifiable from observations of the form

$$y_i = h_u(u_i) \cdot h_v(v_i) + \epsilon_i, \quad h_u \text{ and } h_v \text{ non-linear transformations.}$$

Write $\boldsymbol{h}_u$, $\boldsymbol{h}'_u$, $\boldsymbol{h}''_u$ for the vectors of transformed values and derivatives of h_u at the u_i values, and similarly for $\boldsymbol{v}$. Then

$$\begin{aligned}
C - \log p(\boldsymbol{y} \mid \boldsymbol{u}, \boldsymbol{v}) &= \frac{1}{2}(\boldsymbol{y} - \boldsymbol{h}_u \odot \boldsymbol{h}_v)^\top \boldsymbol{Q}_\epsilon (\boldsymbol{y} - \boldsymbol{h}_u \odot \boldsymbol{h}_v) \\
-\frac{\partial}{\partial \boldsymbol{v}} \log p(\boldsymbol{y} \mid \boldsymbol{u}, \boldsymbol{v}) &= -\operatorname{diag}(\boldsymbol{h}_u \odot \boldsymbol{h}'_v) \boldsymbol{Q}_\epsilon (\boldsymbol{y} - \boldsymbol{h}_u \odot \boldsymbol{h}_v) \\
-\frac{\partial^2}{\partial \boldsymbol{v}^2} \log p(\boldsymbol{y} \mid \boldsymbol{u}, \boldsymbol{v}) &= \operatorname{diag}(\boldsymbol{h}_u \odot \boldsymbol{h}'_v) \boldsymbol{Q}_\epsilon \operatorname{diag}(\boldsymbol{h}_u \odot \boldsymbol{h}'_v) \\
&\quad - \operatorname{diag}(\operatorname{diag}(\boldsymbol{h}_u \odot \boldsymbol{h}''_v) \boldsymbol{Q}_\epsilon (\boldsymbol{y} - \boldsymbol{h}_u \odot \boldsymbol{h}_v))
\end{aligned}$$

and similarly for $\frac{\partial}{\partial \boldsymbol{u}}$, $\frac{\partial^2}{\partial \boldsymbol{u} \partial \boldsymbol{v}}$, and $\frac{\partial^2}{\partial \boldsymbol{u}^2}$. The problematic term in the Hessian involving $\boldsymbol{y}$ disappears in Fisher scoring:
$\mathsf{E}_{\boldsymbol{y}|\boldsymbol{u},\boldsymbol{v}} \left(-\nabla^2_{(\boldsymbol{u},\boldsymbol{v})} \ln p(\boldsymbol{y} \mid \boldsymbol{u}, \boldsymbol{v}) \right)$ is positive definite.

Posterior calculations

Simpified 2-step multiscale precision matrix block structure:

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

can be pseudo-Cholesky-factorised:

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

Posterior expectation, samples, and marginal variances:

$$\widetilde{\boldsymbol{A}} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{0} \end{bmatrix},$$

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

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

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

$$\mathsf{Var}(x_i|\boldsymbol{y}) = \texttt{diag(inla.qinv(}\boldsymbol{Q}_{x|y}\texttt{))} \quad \text{(requires Cholesky)}$$

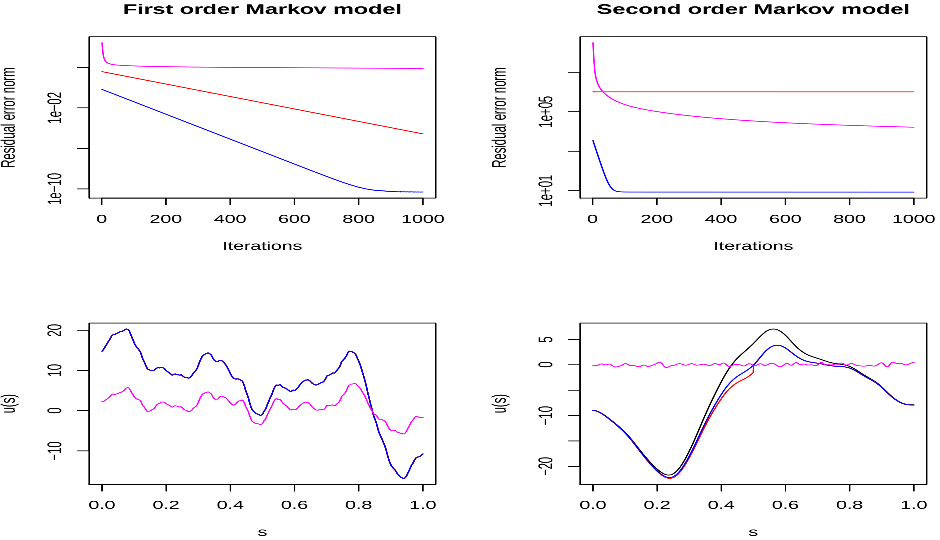
Quarter degree output grid
365 daily estimates each year
165 years
Two fields

$$360 \cdot 180 \cdot 4^2 \cdot 365 \cdot 165 \cdot 2 = 124,882,560,000$$

Storing $\sim 10^{11}$ latent variables as floats takes ~ 500 GB
(And that just covers the finest scale)

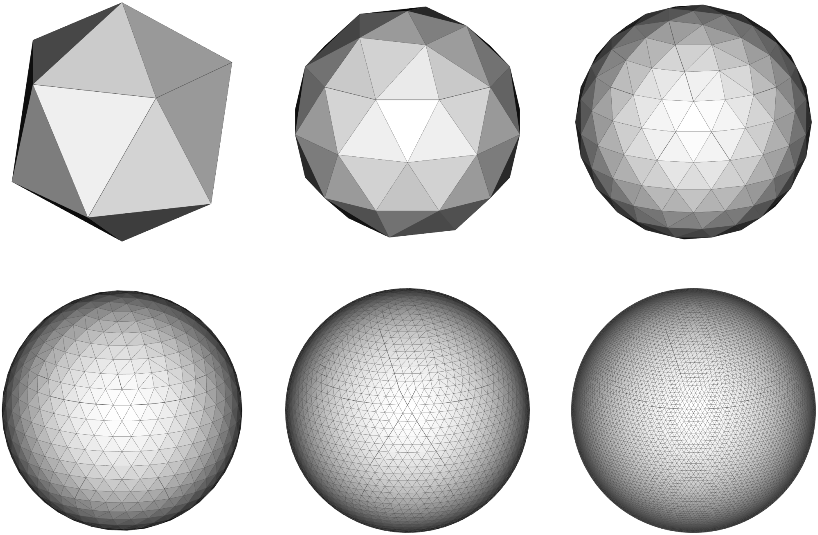
To store the data (> 10 TB), model information, and estimated uncertainties we need a computing cluster with lots of RAM and fast temporary parallell disk access.

Matrix-free iterative solvers will be our saviours!

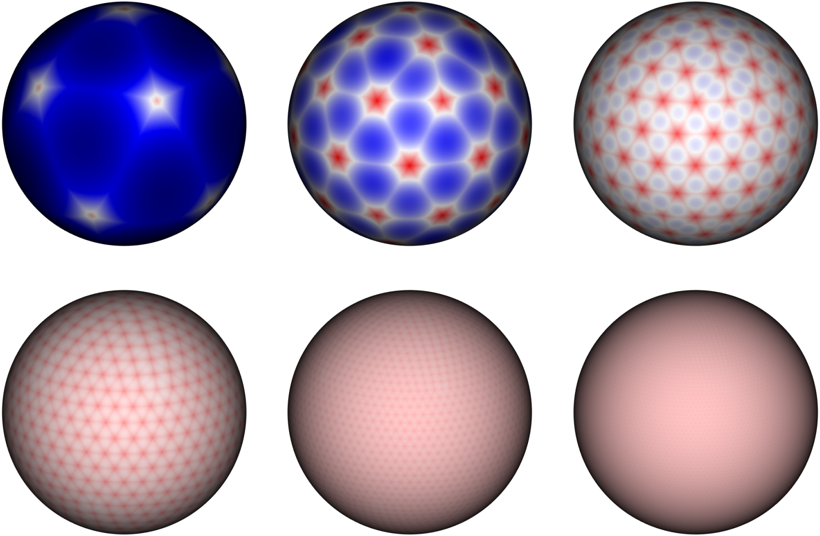


Residual norms and results after 1000 iterations for Block Jacobi (red), block Gauss-Seidel (blue), and single site Gauss-Seidel (magenta). Convergence is spectacularly slow for higher order operators!

Triangulations for all corners of Earth



Triangulations for all corners of Earth



Domain decomposition

Use *overlapping blocks* distributed over many computing nodes, and add an approximate global step:

Overlapping subdomains

Let $\boldsymbol{B}_k^\top$ be the restriction matrix to subdomain Ω_k, and let $\boldsymbol{B}_c^\top$ be a projection onto a coarse basis. Then an additive Schwartz preconditioner with coarse correction is given by

$$\boldsymbol{M}^{-1}\boldsymbol{x} = \boldsymbol{B}_c(\boldsymbol{B}_c^\top \boldsymbol{Q}\boldsymbol{B}_c)^{-1}\boldsymbol{B}_c^\top \boldsymbol{x} + \sum_{k=1}^{K} \boldsymbol{B}_k(\boldsymbol{B}_k^\top \boldsymbol{Q}\boldsymbol{B}_k)^{-1}\boldsymbol{B}_k^\top \boldsymbol{x}$$

The different timescales can be handled with repeated multiscale preconditioning:

Multiscale Schur complement approximation

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

$$\boldsymbol{Q}_z + \boldsymbol{B}^\top \boldsymbol{Q}_t \boldsymbol{B} \otimes \boldsymbol{Q}_s - \boldsymbol{B}^\top \boldsymbol{Q}_t \otimes \boldsymbol{Q}_s \left(\boldsymbol{Q}_t \otimes \boldsymbol{Q}_s + \boldsymbol{A}^\top \boldsymbol{Q}_\epsilon \boldsymbol{A} \right)^{-1} \boldsymbol{Q}_t \boldsymbol{B} \otimes \boldsymbol{Q}_s$$

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

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

where $\widetilde{\boldsymbol{B}} = \boldsymbol{B} \otimes \boldsymbol{I}$, $\widetilde{\boldsymbol{Q}}_B = \boldsymbol{B}^\top \boldsymbol{Q}_t \boldsymbol{B} \otimes \boldsymbol{Q}_s$, and the block matrix can be interpreted as the precision of a bivariate field on a common, coarse spatio-temporal scale.

Variance calculations

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 $\Omega_k \subset \Omega$,

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

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

Method overview

- Hierarchical timescale combination of space-time random fields
- Preprocessing to estimate model parameters and non-Gaussianity
- Iterated linearisation
- Distributed Preconditioned Conjugate Gradient solves
- Information is passed between the scales via approximate Schur complements
- Overlapping space-time domain decomposition within each scale
- Rao-Blackwellised variance estimation

References

- Rue, H. and Held, L.: Gaussian Markov Random Fields; Theory and Applications; *Chapman & Hall/CRC*, 2005
- Lindgren, F.: Computation fundamentals of discrete GMRF representations of continuous domain spatial models; preliminary book chapter manuscript, 2015, http://www.maths.ed.ac.uk/~flindgre/tmp/gmrf.pdf
- Lindgren, F., Rue, H., and Lindström, J.: An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach (with discussion); *JRSS Series B*, 2011
  Non-CRAN package: `R-INLA` at http://r-inla.org/