Royal Netherlands
Meteorological Institute
Ministry of Infrastructure and the Environment

Large scale spatial statistics
with SPDEs, GMRFs,
and multi-scale component models

Finn Lindgren (`finn.lindgren@ed.ac.uk`)

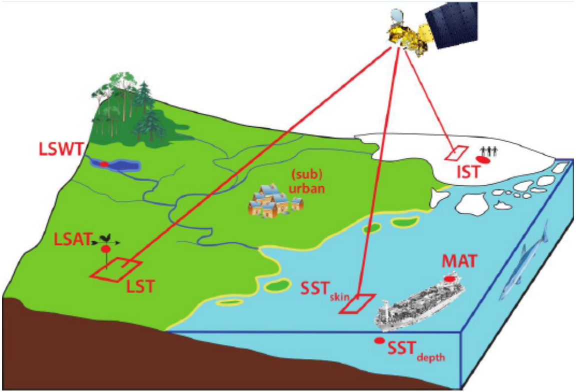
SIAM UQ 2018-04-17

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.



EUSTACE

Quarter degree output grid
365 daily estimates each year
165 years
Two fields: daily mean and range

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

Storing $\sim 10^{11}$ latent variables as `double` takes ~ 1 TB

We want a joint estimate of the entire space-time process
at several time scales (daily, climatological, seasonal)
Methods based on direct covariance calculations are infeasible.

An additive hierarchical stochastic PDE model
and matrix-free iterative solvers
will (hopefully) save us!

Gaussian random field

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

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

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

Generalised Gaussian random field

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

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

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

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

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 $K(\mathrm{s}, \mathrm{s}') = \delta(\mathrm{s} - \mathrm{s}')$. Do not confuse with independent noise, $K(\mathrm{s}, \mathrm{s}') = \mathbb{I}(\mathrm{s} = \mathrm{s}')$, which has non-integrable realisations.

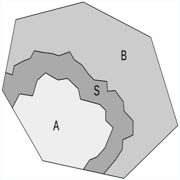
GMRFs: Gaussian Markov random fields

Continuous domain GMRFs

If $x(\boldsymbol{s})$ is a (stationary) Gaussian random field on Ω with covariance kernel $K(\boldsymbol{s}, \boldsymbol{s}')$, it fulfills the *global Markov property*

$$\{x(\mathcal{A}) \perp x(\mathcal{B}) | x(\mathcal{S}),\ \text{for all } \mathcal{AB}\text{-separating sets } \mathcal{S} \subset \Omega\}$$

if the power spectrum can be written as $1/S_x(\boldsymbol{\omega}) =$ polynomial in $\boldsymbol{\omega}$, for some polynomial order p. (Rozanov, 1977)



Generally: Markov iff the precision operator $\mathcal{Q} = \mathcal{R}^{-1}$ is local.

Discrete domain Gaussian Markov random fields (GMRFs)

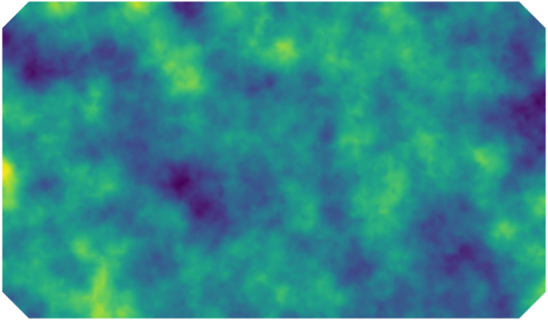
$\boldsymbol{x} = (x_1, \ldots, x_n) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{Q}^{-1})$ is Markov with respect to a neighbourhood structure $\{\mathcal{N}_i, i = 1, \ldots, n\}$ if $Q_{ij} = 0$ whenever $j \neq \mathcal{N}_i \cup i$.

- Continuous domain basis representation with weights: $x(\boldsymbol{s}) = \sum_{k=1}^{n} \psi_k(\boldsymbol{s}) x_k$
- Project the SPDE solution space onto local basis functions: random Markov weights (Lindgren et al, 2011).

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.

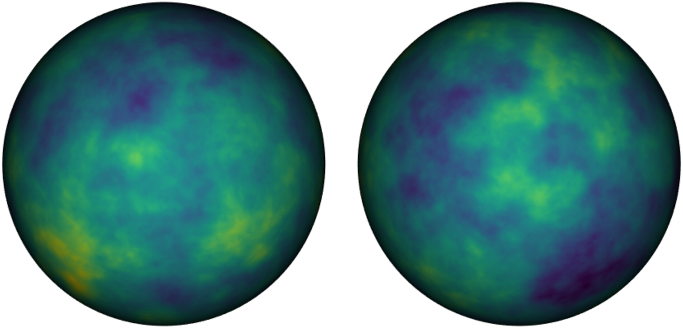
$$(\kappa^2 - \Delta)(\tau\, x(\mathbf{s})) = \mathcal{W}(\mathbf{s}), \quad \mathbf{s} \in \mathbb{R}^d$$



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.

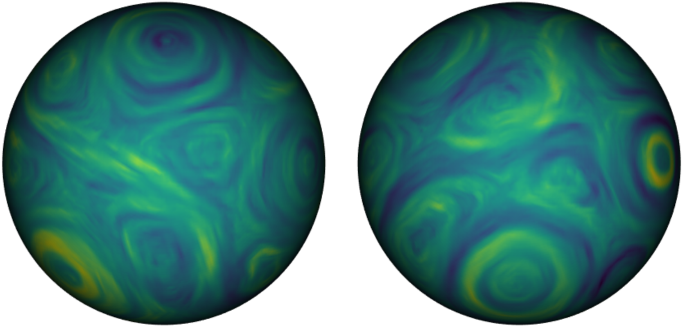
$$(\kappa^2 - \Delta)(\tau\, x(\mathbf{s})) = \mathcal{W}(\mathbf{s}), \quad \mathbf{s} \in \Omega$$



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.

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



Matérn driven heat equation on the sphere

The iterated heat equation is a simple non-separable space-time SPDE family:

$$(\kappa^2 - \Delta)^{\gamma/2} \left[\phi \frac{\partial}{\partial t} + (\kappa^2 - \Delta)^{\alpha/2} \right]^{\beta} x(\mathbf{s}, t) = \mathcal{W}(\mathbf{s}, t)/\tau$$

Fourier spectra are based on eigenfunctions $e_{\boldsymbol{\omega}}(\mathbf{s})$ of $-\Delta$.
On $\mathbb{R}^2$, $-\Delta e_{\boldsymbol{\omega}}(\mathbf{s}) = \|\boldsymbol{\omega}\|^2 e_{\boldsymbol{\omega}}(\mathbf{s})$, and $e_{\boldsymbol{\omega}}$ are harmonic functions.
On $\mathbb{S}^2$, $-\Delta e_k(\mathbf{s}) = \lambda_k e_k(\mathbf{s}) = k(k+1) e_k(\mathbf{s})$, and e_k are spherical harmonics.
The isotropic spectrum on $\mathbb{S}^2 \times \mathbb{R}$ is

$$\widehat{\mathcal{R}}(k, \omega) \propto \frac{2k+1}{\tau^2 (\kappa^2 + \lambda_k)^{\gamma} \left[\phi^2 \omega^2 + (\kappa^2 + \lambda_k)^{\alpha} \right]^{\beta}}$$

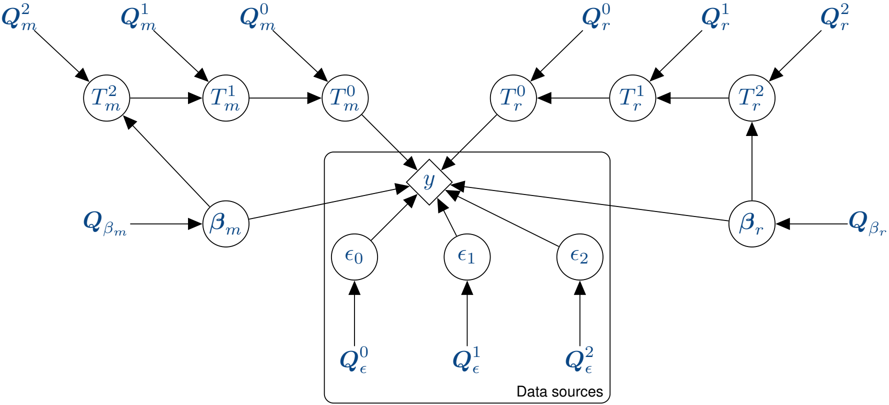
The finite element approximation has precision matrix structure

$$\boldsymbol{Q} = \sum_{i=0}^{\alpha+\beta+\gamma} \boldsymbol{M}_i^{[t]} \otimes \boldsymbol{M}_i^{[\mathbf{s}]}$$

even, e.g., if κ is spatially varying.

Partial hierarchical representation

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



Conditional specifications, e.g.

$$(T_m^0 | T_m^1, \boldsymbol{Q}_m^0) \sim \mathcal{N}\left(T_m^1, {\boldsymbol{Q}_m^0}^{-1}\right)$$

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

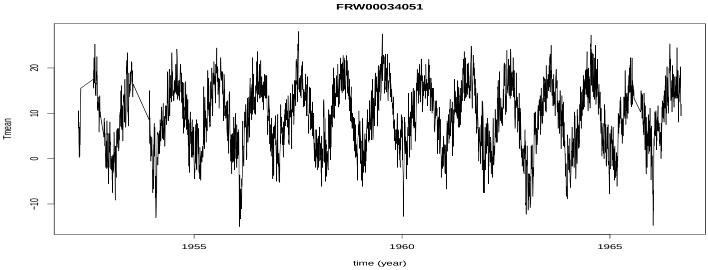
$$g^{-1}[\mu_r(\mathbf{s},t)] = \underbrace{U_r^1(\mathbf{s},t) + \underbrace{U_r^2(\mathbf{s},t) + U_r^S(\mathbf{s},t) + \sum_{i=1}^{N_X} X_i(\mathbf{s},t)\beta_r^{(i)}}_{T_r^2}}_{T_r^1},$$

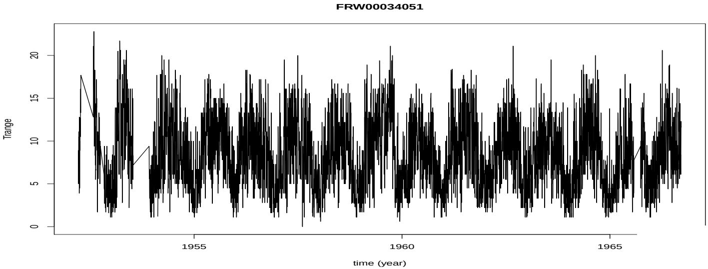
$$T_r(\mathbf{s},t) = \mu_r(\mathbf{s},t)\ G^{-1}\left[U_r^0(\mathbf{s},t)\right] = \underbrace{g(T_r^1)\ G^{-1}\left[U_r^0(\mathbf{s},t)\right]}_{T_r^0},$$

where the slowly varying median process $\mu_r(\mathbf{s},t)$ is a transformed multiscale model, and G^{-1} is a spatially and seasonally varying transformation model. The β_r coefficients are weights for covariates $X_i(\mathbf{s},t)$ (e.g. elevation, topographical gradients, and land use indicator functions).

Observed data

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





Power tail quantile (POQ) model

The quantile function $F_{\boldsymbol{\theta}}^{-1}(p)$, $p \in [0, 1]$, is defined through a quantile blend of left- and right-tailed generalised Pareto distributions.

The parameters $\boldsymbol{\theta} = (\theta_0, \theta_1 = \log\tau, \theta_2 = \text{logit}[(\gamma+1)/2], \theta_3, \theta_4)$ 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 POQ copula 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

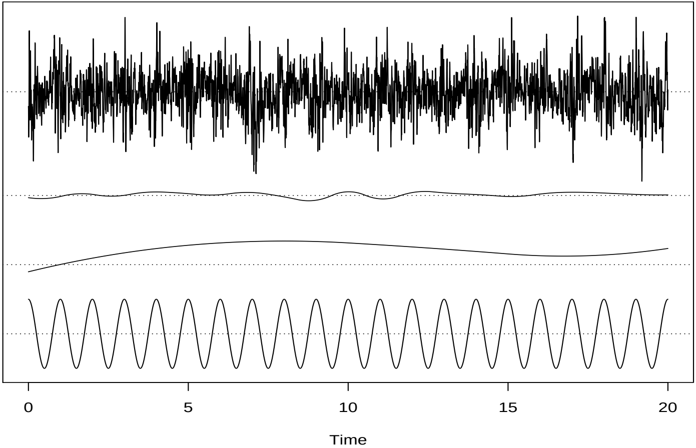
$$\widetilde{u}(\mathbf{s}, t) = G^{-1}[u(\mathbf{s}, t)] = F_{\boldsymbol{\theta}(\mathbf{s},t)}^{-1}(\Phi(u(\mathbf{s}, t)),$$

where the parameters can vary with space and time.

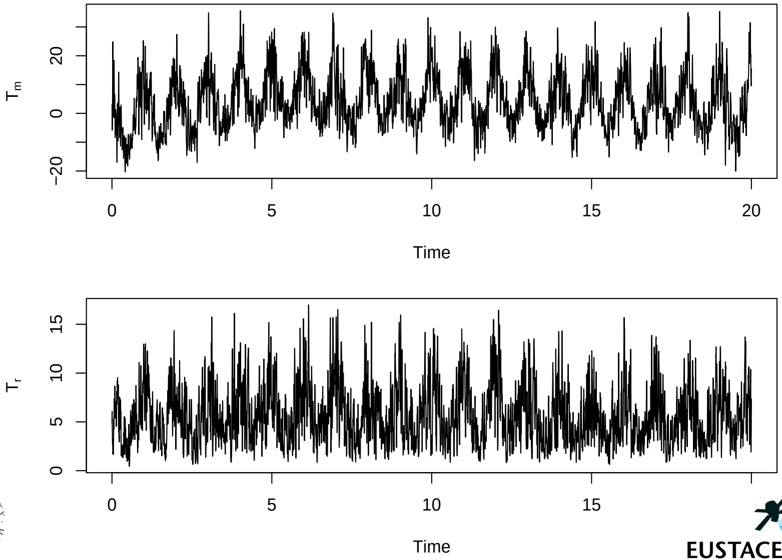
Due to the large size of the problem, we estimate parameters in a two-step procedure:

1. Estimate seasonal POQ and temporal covariance parameters for separate time series
2. With a basic spatial-seasonal random field prior, find the posterior mean parameter field

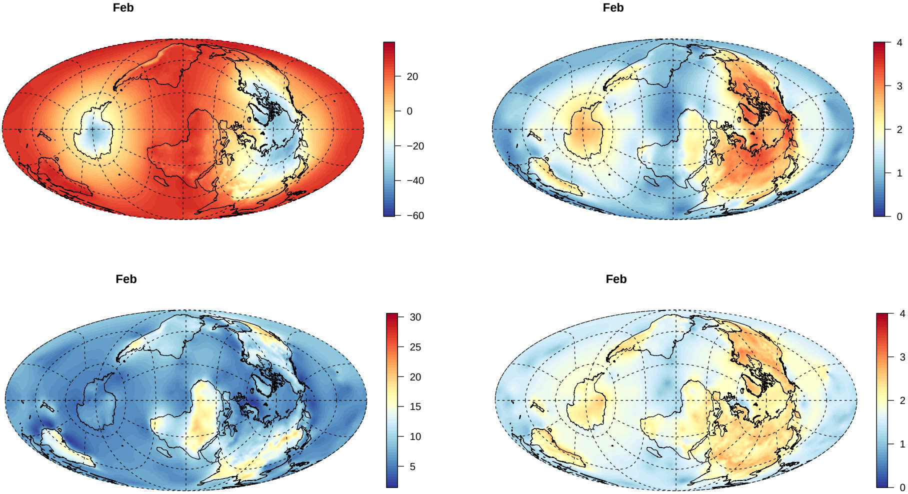
Multiscale model component samples



Combined model samples for T_m and T_r



Estimates of median & scale for T_m and T_r



February climatology

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 $\boldsymbol{Q}_x$:

$$(\boldsymbol{x} \mid \boldsymbol{\theta}) \sim \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{Q}_x^{-1}) \qquad \text{(Prior; Only used in pre-estimation in EUSTACE)}$$

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

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

Linear Gaussian observations

The conditional posterior distribution is

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

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

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

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 $\boldsymbol{Q}_x$:

$$(\boldsymbol{x} \mid \boldsymbol{\theta}) \sim \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{Q}_x^{-1}) \qquad \text{(Prior; Only used in pre-estimation in EUSTACE)}$$

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

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

Non-linear and/or non-Gaussian observations

For a non-linear $h(\boldsymbol{A}\boldsymbol{x})$ with Jacobian $\boldsymbol{J}$ at $\boldsymbol{x} = \widetilde{\boldsymbol{\mu}}$, iterate:

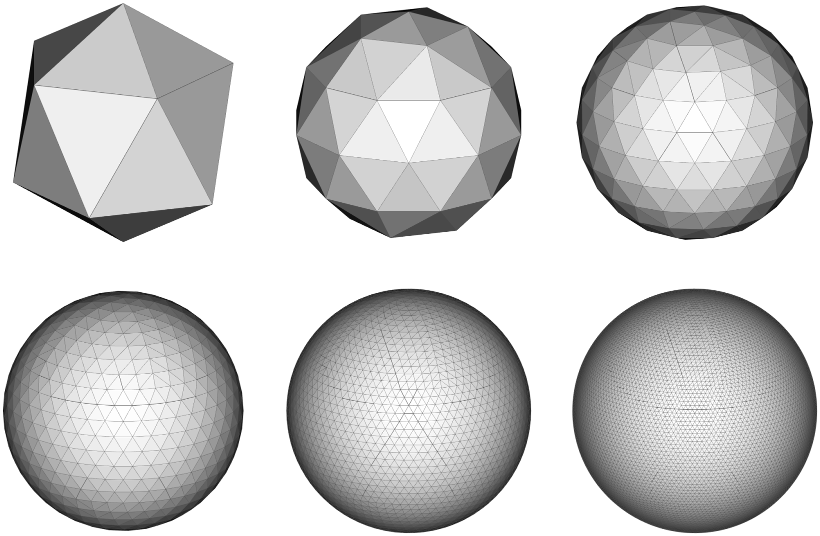
$$(\boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{\theta}) \overset{\text{approx}}{\sim} \mathcal{N}(\widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{Q}}^{-1}) \qquad \text{(Approximate conditional posterior)}$$

$$\widetilde{\boldsymbol{Q}} = \boldsymbol{Q}_x + \boldsymbol{J}^\top \boldsymbol{Q}_{y|x} \boldsymbol{J}$$

$$\widetilde{\boldsymbol{\mu}}' = \widetilde{\boldsymbol{\mu}} + a \widetilde{\boldsymbol{Q}}^{-1} \left\{ \boldsymbol{J}^\top \boldsymbol{Q}_{y|x} [\boldsymbol{y} - h(\boldsymbol{A}\widetilde{\boldsymbol{\mu}})] - \boldsymbol{Q}_x (\widetilde{\boldsymbol{\mu}} - \boldsymbol{\mu}_x) \right\}$$

for some $a > 0$ chosen by line-search.

Triangulations for all corners of Earth



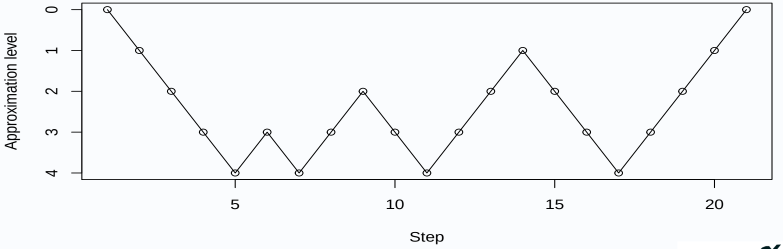
Overlapping blocks and multigrid

Overlapping block preconditioning

Let $\boldsymbol{D}_k^\top$ be a restriction matrix to subdomain Ω_k, and let $\boldsymbol{W}_k$ be a diagonal weight matrix. Then an additive Schwartz preconditioner is

$$\boldsymbol{M}^{-1}\boldsymbol{x} = \sum_{k=1}^{K} \boldsymbol{W}_k \boldsymbol{D}_k (\boldsymbol{D}_k^\top \boldsymbol{Q} \boldsymbol{D}_k)^{-1} \boldsymbol{D}_k^\top \boldsymbol{W}_k \boldsymbol{x}$$

Multigrid



The hierarchy of scales and preconditioning ($\boldsymbol{x}_0 = \boldsymbol{B}\boldsymbol{x}_1 +$ fine scale variability):

Multiscale Schur complement approximation

Solving $\boldsymbol{Q}_{\boldsymbol{x}|\boldsymbol{y}}\boldsymbol{x} = \boldsymbol{b}$ can be formulated using two solves with the upper (fine) block $\boldsymbol{Q}_0 + \boldsymbol{A}^\top\boldsymbol{Q}_\epsilon\boldsymbol{A}$, and one solve with the *Schur complement*

$$\boldsymbol{Q}_1 + \boldsymbol{B}^\top\boldsymbol{Q}_0\boldsymbol{B} - \boldsymbol{B}^\top\boldsymbol{Q}_0\left(\boldsymbol{Q}_0 + \boldsymbol{A}^\top\boldsymbol{Q}_\epsilon\boldsymbol{A}\right)^{-1}\boldsymbol{Q}_0$$

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

$$\begin{bmatrix} \widetilde{\boldsymbol{Q}}_B + \boldsymbol{B}^\top\boldsymbol{A}^\top\boldsymbol{Q}_\epsilon\boldsymbol{A}\boldsymbol{B} & -\widetilde{\boldsymbol{Q}}_B \\ -\widetilde{\boldsymbol{Q}}_B & \boldsymbol{Q}_1 + \widetilde{\boldsymbol{Q}}_B \end{bmatrix}\begin{bmatrix} \text{ignored} \\ \boldsymbol{x}_1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} \\ \widetilde{\boldsymbol{b}} \end{bmatrix}$$

where $\widetilde{\boldsymbol{Q}}_B = \boldsymbol{B}^\top\boldsymbol{Q}_0\boldsymbol{B}$.
The block matrix can be interpreted as the precision of a bivariate field on a common, coarse spatio-temporal scale, and the same technique applied to this system, with $\boldsymbol{x}_{1,1} = \boldsymbol{B}_{1|2}\boldsymbol{x}_{1,2} +$ finer scale variability.

Also applies to the station data bias homogenisation coefficients.

Summary and further developments

- Hierarchical timescale combination of space-time random fields
- Translation between GRF/SPDE/GMRF; they are all the same Gaussian process
- Know how to solve smaller problems; overlapping domains for preconditioning
- Multiscale model structure used for effective preconditioning
- Direct Monte Carlo sampling: add suitable randomness to the RHS of the system
- Improve posterior variance estimates with Rao-Blackwellisation

Current status and future developments:

- Implementation for smaller region than global is in progress
- Full global solve will likely require multigrid
- The full approximate Schur complement method would require multiple data read for the preconditioner; Is there a better alternative than separate block-preconditioning?
- Spatial covariance parameter estimation should take advantage of the non-stationarity; a global, joint Bayesian parameter estimate would be overkill; estimate locally, and blend to a coherent global model.