Royal Netherlands
Meteorological Institute
Ministry of Infrastructure and the Environment

Large Spatio-temporal Modelling and Computing for Past Weather and Climate

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

THE UNIVERSITY *of* EDINBURGH

ISBA 2018, 2018-06-27

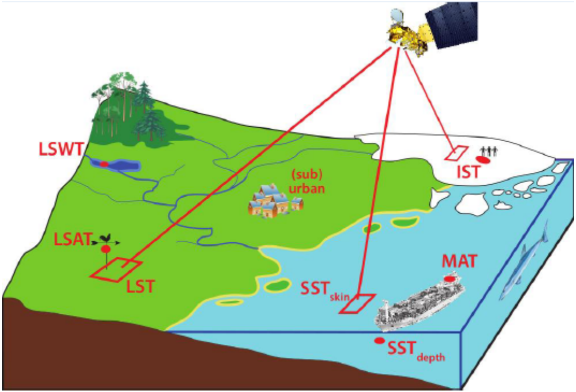
EUSTACE has received funding from the European Union's Horizon 2020 Programme for Research and Innovation, under Grant Agreement no 640171

EUSTACE

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.



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

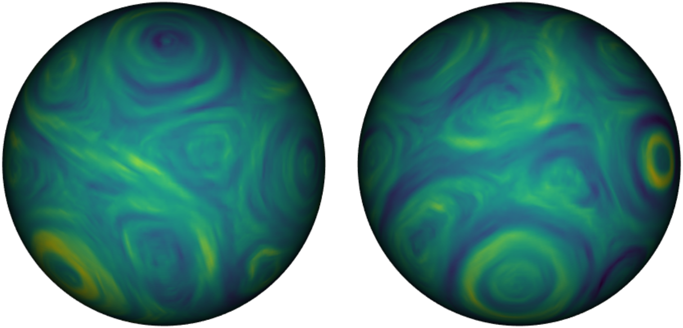
We want a joint estimate of the entire space-time process
at several time scales (daily, climatological, seasonal)
Methods based on direct covariance calculations are infeasible.

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

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.

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



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

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

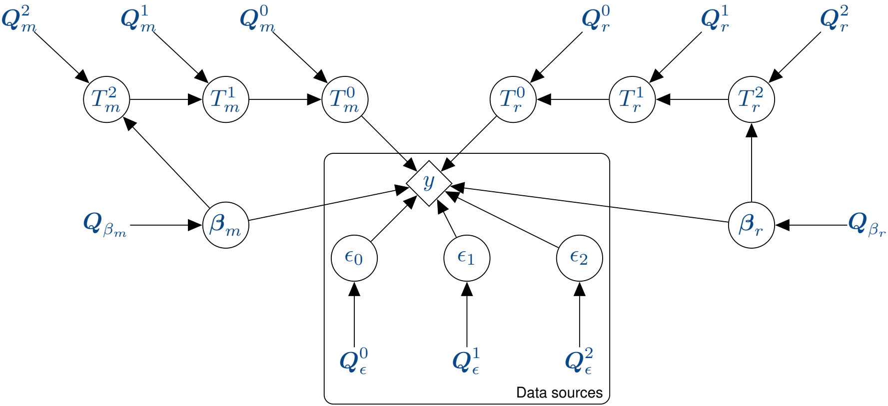
A finite element approximation has structure

$$x(\boldsymbol{s}) = \sum_{k=1}^{n} \psi_k(\boldsymbol{s}) x_k, \quad \boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}^{-1}), \quad \boldsymbol{Q} = \sum_{i=0}^{\alpha+\beta+\gamma} \boldsymbol{M}_i^{[t]} \otimes \boldsymbol{M}_i^{[\mathbf{s}]}$$

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

Partial hierarchical representation

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



Conditional specifications, e.g.

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

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

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

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

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).

Observation models

Common satellite derived data error model framework

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

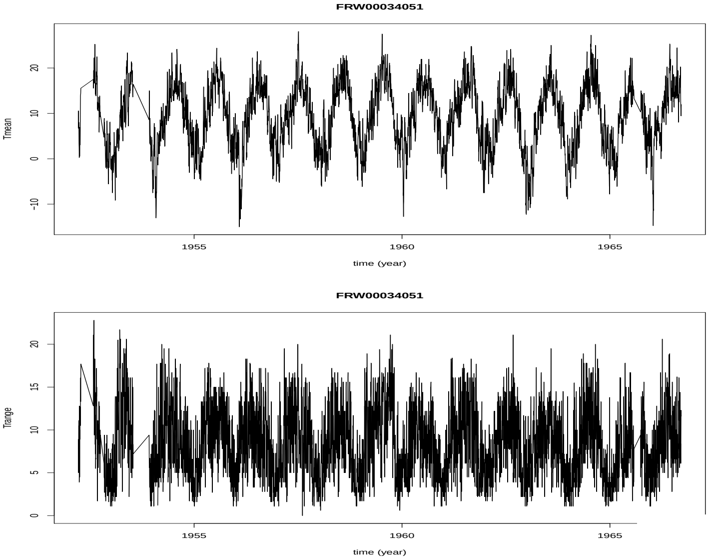
For station k at day t_i

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

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 $F_{\boldsymbol{\theta}}^{-1}(p)$, $p \in [0, 1]$, 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* (Gilchrist, 2000).

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

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

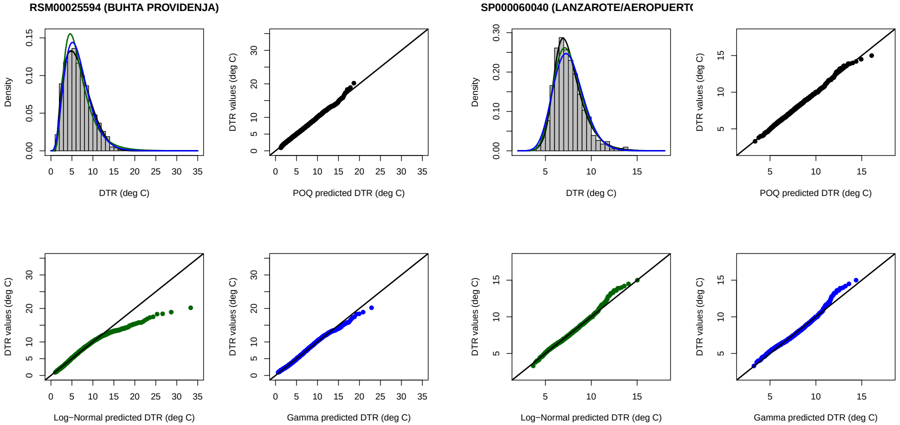
where the parameters can vary with space and time.

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

Diurnal range distributions

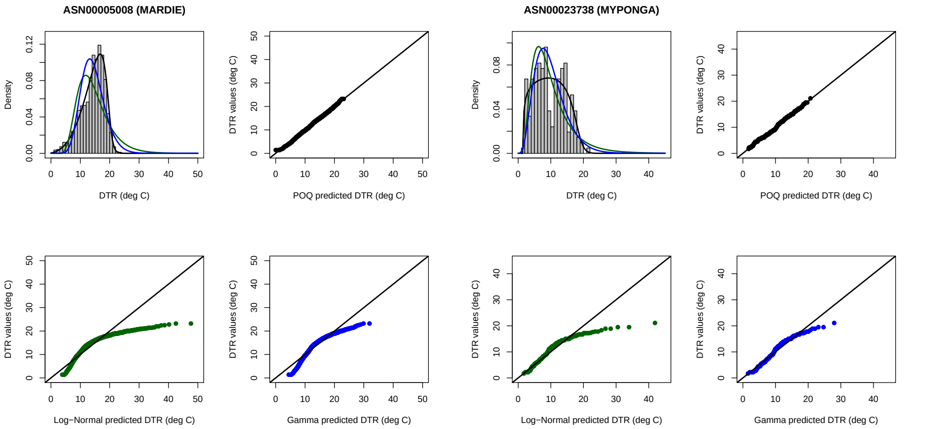
After seasonal compensation:



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

Diurnal range distributions; quantile model

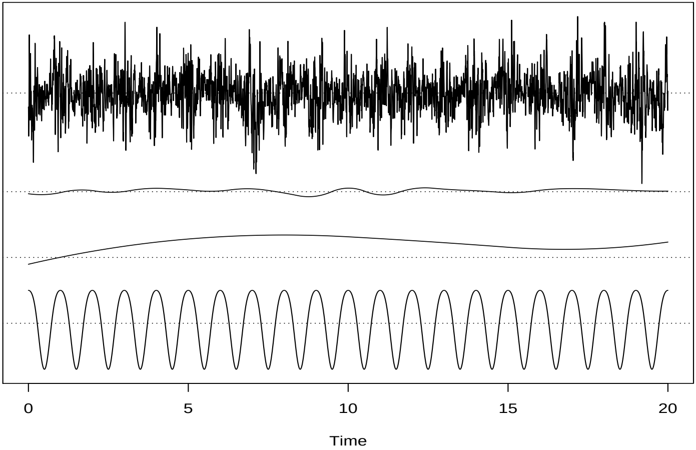
After seasonal compensation:



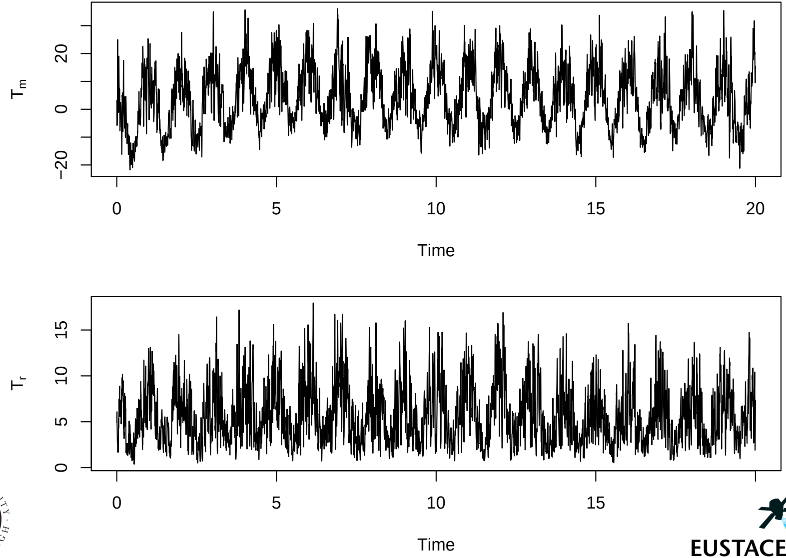
For these stations only POQ comes close to representing the distributions.

Note: Some of the mixture-like distribution shapes may be an effect of unmodeled station inhomogeneities as well as temporal shift effects.

Multiscale model component samples

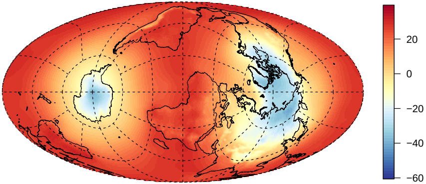


Combined model samples for T_m and T_r

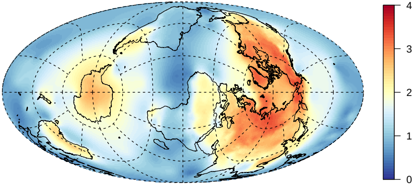


Estimates of median & scale for T_m and T_r

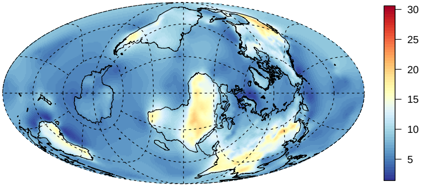
Feb



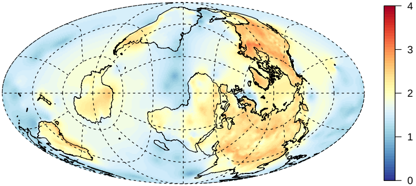
Feb



Feb



Feb



February climatology

(Preliminary estimates, using only in-situ land station data)

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 $\boldsymbol{Q}_x$:

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

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

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

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 $\boldsymbol{Q}_x$:

$$(\boldsymbol{x} \mid \boldsymbol{\theta}) \sim \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{Q}_x^{-1})$$ (Prior; Parameters pre-estimated in EUSTACE)

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

$$p(\boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{\theta}) \propto p(\boldsymbol{x} \mid \boldsymbol{\theta})\, p(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta})$$ (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:

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

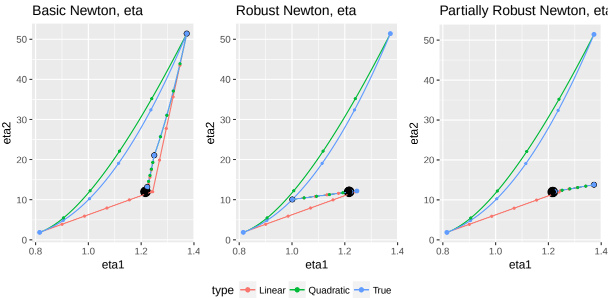
$$\widetilde{\boldsymbol{Q}} = \boldsymbol{Q}_x + \boldsymbol{J}^\top \boldsymbol{Q}_{y|x} \boldsymbol{J}$$

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

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

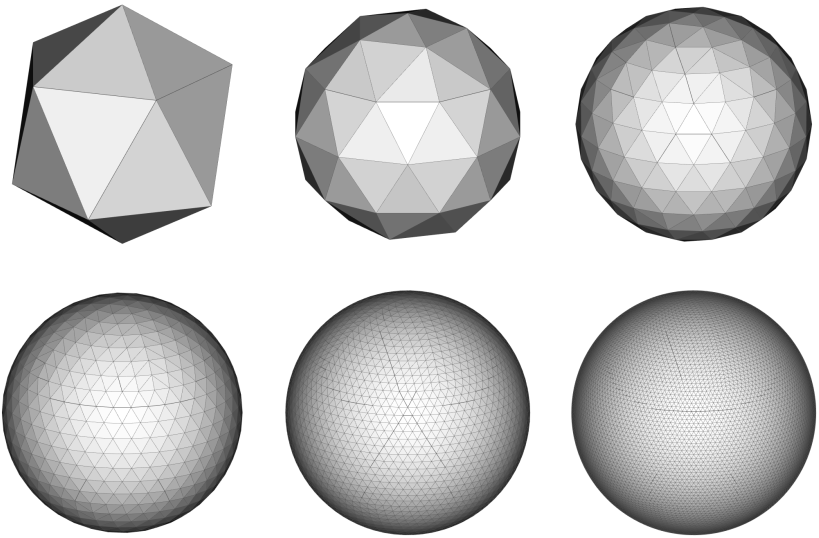
Iterative solutions

- Nonlinear Newton iteration with robust line-search



- Preconditioned conjugate gradient (PCG) iteration for $\boldsymbol{Q}(\boldsymbol{\mu}-\widehat{\boldsymbol{\mu}}) = \boldsymbol{r} = \boldsymbol{b} - \boldsymbol{Q}\widehat{\boldsymbol{\mu}}$
- Local and multiscale approximations for preconditioning: $\boldsymbol{M}^{-1}\boldsymbol{Q} \approx \boldsymbol{I}$
- Sampling with PCG: $\boldsymbol{Q}(\boldsymbol{x}-\widehat{\boldsymbol{\mu}}) = \boldsymbol{L}\boldsymbol{w}$
  Requires only a rectangular pseudo-Cholesky factorisation $\boldsymbol{L}\boldsymbol{L}^\top = \boldsymbol{Q}$.

Triangulations for all corners of Earth



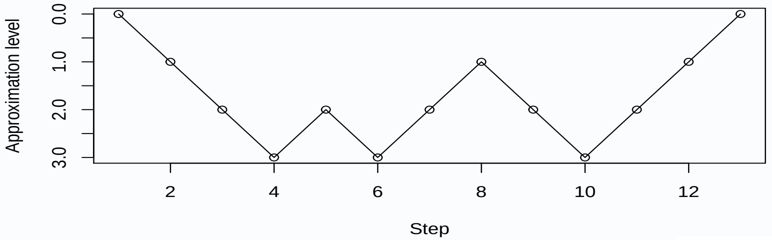
Overlapping blocks and multigrid

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 and approximate multiscale Schur complements



Variance calculations

Sparse partial inverse: Takahashi recursions postprocesses Cholesky

Takahashi recursions compute $\boldsymbol{S}$ such that $\boldsymbol{S}_{ij} = (\boldsymbol{Q}^{-1})_{ij}$ for all $Q_{ij} \neq 0$.
Postprocessing of the (sparse) Cholesky factor.

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 $\Omega_k \subset \Omega$,

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

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

Efficient if $\boldsymbol{a}\boldsymbol{a}^\top$ sparsity matches $\boldsymbol{S}$ for each subdomain.

Summary and further developments

- "Big" data $\not\Rightarrow$ abundant information; imbalanced sparse data collection
  Real temperatures are the primary interest, not the model parameters
- Hierarchical timescale combination of transformed space-time random fields
- Translation between GRF/SPDE/GMRF; they are all Gaussian processes
- Know how to solve smaller problems; overlapping domains for preconditioning
- Small problems solvable; Use multiscale structure for global solution
- 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 would likely require multigrid
- Spatial covariance parameter estimation should take advantage of the non-stationarity; don't need a global, joint Bayesian parameter estimate; estimate locally, and blend to a coherent global model.
- Iterative global-Cholesky-free Rao-Blackwellisation:
  Efficient Covariance Approximations for Large Sparse Precision Matrices
  Sidén et al, arXiv:1705.08656, to appear in JCGS