Royal Netherlands
Meteorological Institute
Ministry of Infrastructure and the Environment

EUSTACE:

A case study in

hierarchical space-time modelling

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

THE UNIVERSITY *of* EDINBURGH

Banff 2017-07-11

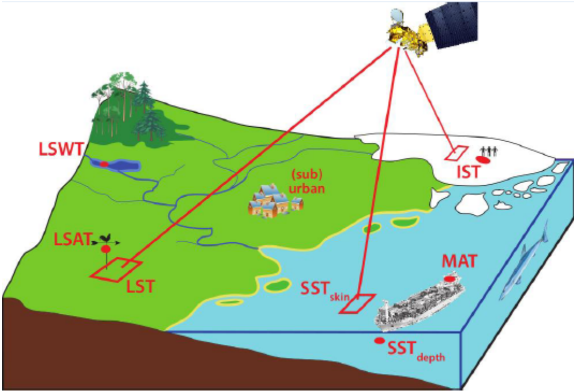
EUSTACE has received funding from the European Union's Horizon 2020 Programme for Research and Innovation, under Grant Agreement no 640171

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.



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

Spatio-temporal hierarchical model framework

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

Gaussian random field

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

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

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, $\langle f, x\rangle_D = x^*(f) : H_{\mathcal{R}}(D) \mapsto \mathbb{R}$,

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

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

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

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.

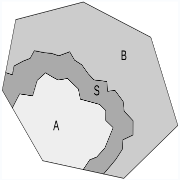
GMRFs: Gaussian Markov random fields

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*

$\{x(\mathcal{A}) \perp x(\mathcal{B}) | x(\mathcal{S})$, for all $\mathcal{AB}$-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 GMRFs

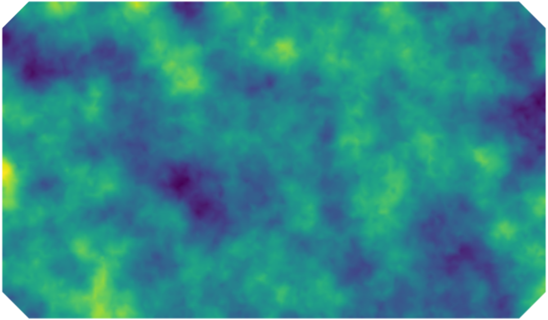
$\boldsymbol{x} = (x_1, \ldots, x_n) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{Q}^{-1})$ is Markov with respect to a neighbourhood structure $\{\mathcal{N}_i, i = 1, \ldots, n\}$ if $Q_{ij} = 0$ whenever $j \neq \mathcal{N}_i \cup i$.

- Continuous domain basis representation with Markov weights: $x(\boldsymbol{s}) = \sum_{k=1}^{n} \psi_k(\boldsymbol{s}) x_k$
- Many stochastic PDE solutions are Markov in continuous space, and can be approximated by Markov weights on local basis functions.

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.

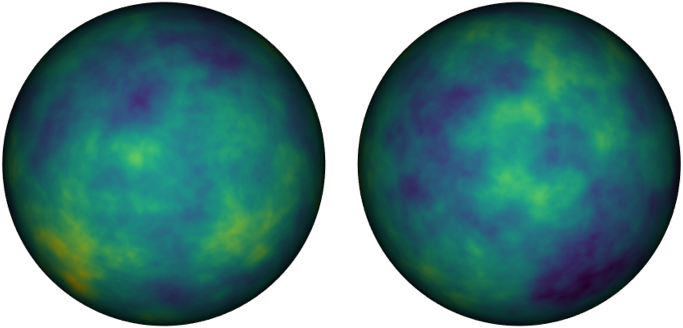
$$(\kappa^2 - \Delta)(\tau\, x(\mathbf{s})) = \mathcal{W}(\mathbf{s}), \quad \mathbf{s} \in \mathbb{R}^d$$



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.

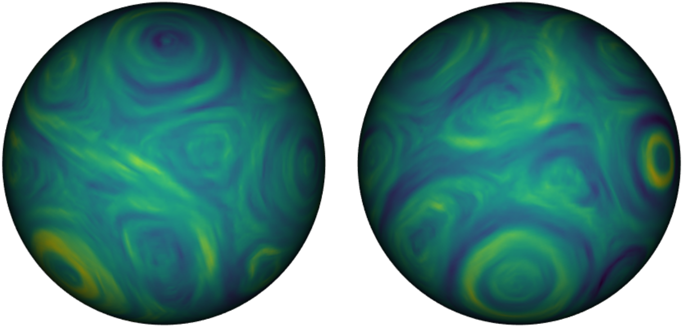
$$(\kappa^2 - \Delta)(\tau\, x(\mathbf{s})) = \mathcal{W}(\mathbf{s}), \quad \mathbf{s} \in \Omega$$



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.

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



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.

For $\alpha = 2$ in the Matérn SPDE,

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

The covariance for the RHS of the SPDE is

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

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

Matching the LHS and RHS distributions leads to the finite element approximation

$$\boldsymbol{x}\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{Q}=\kappa^4\boldsymbol{C}+2\kappa^2\boldsymbol{G}+\boldsymbol{G}\boldsymbol{C}^{-1}\boldsymbol{G})$$

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

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

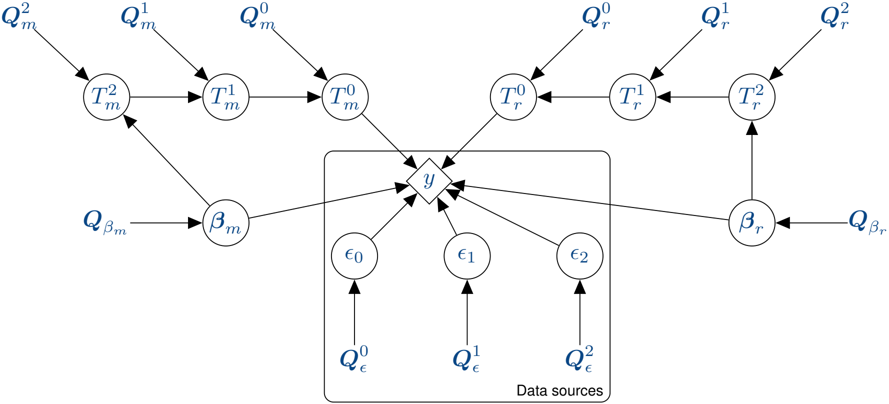
The finite element approximation has precision matrix structure

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

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

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

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

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

EUSTACE

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

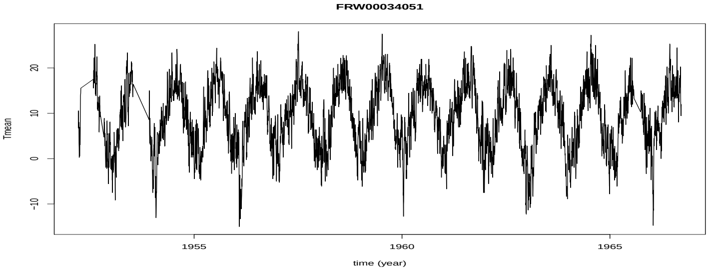
For station k at day t_i

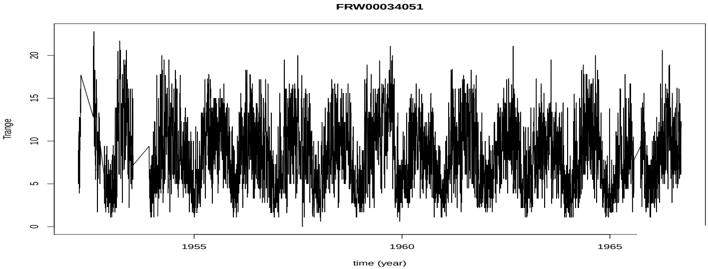
$$y_m^{k,i} = T_m(\mathbf{s}_k, t_i) + \sum_{j=1}^{J_k} H_j^k(t_i) e_m^{k,j} + \epsilon_m^{k,i},$$

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

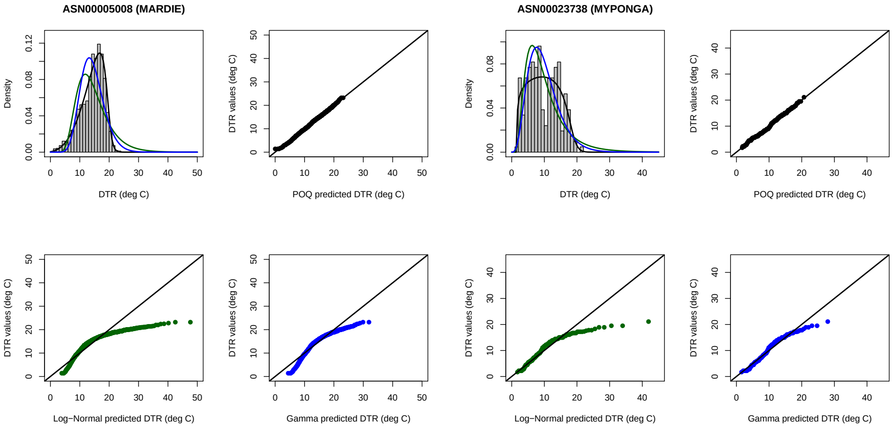
Observed daily T_{mean} and T_{range} for station FRW00034051





Diurnal range distributions; quantile model

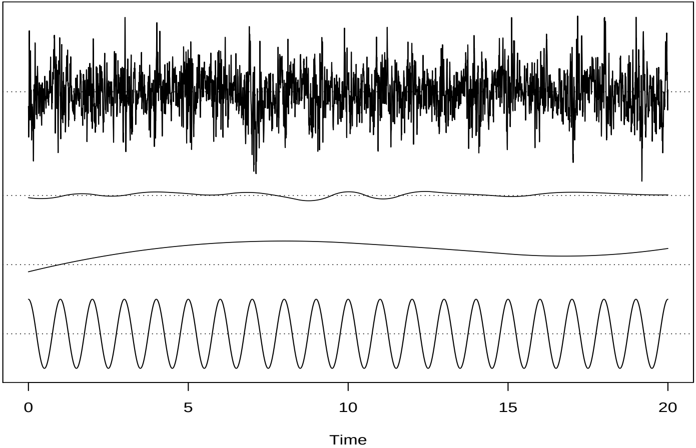
After seasonal compensation:



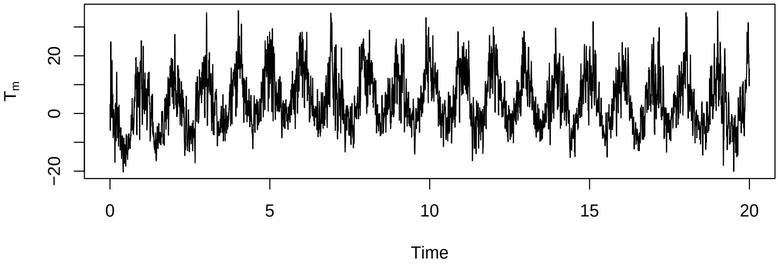
For these stations only POQ comes close to representing the distributions.

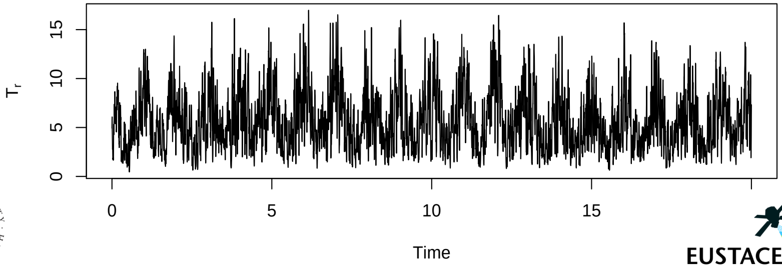
Note: Some of the mixture-like distribution shapes may be an effect of unmodeled station inhomogeneities as well as temporal shift effects.

Multiscale model component samples

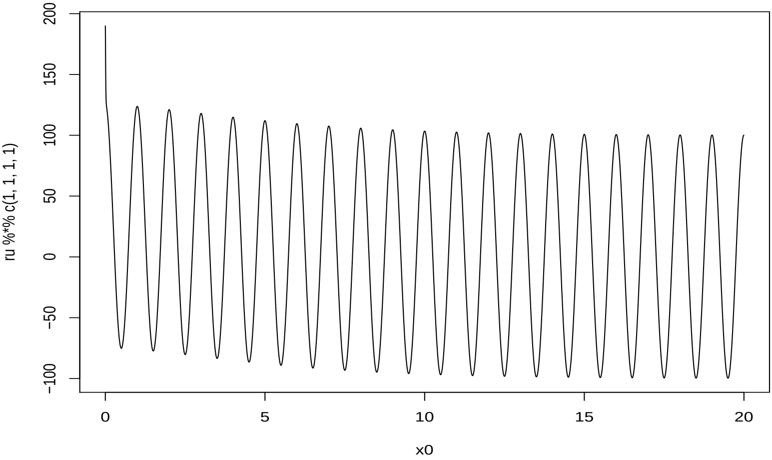


Combined model samples for T_m and T_r

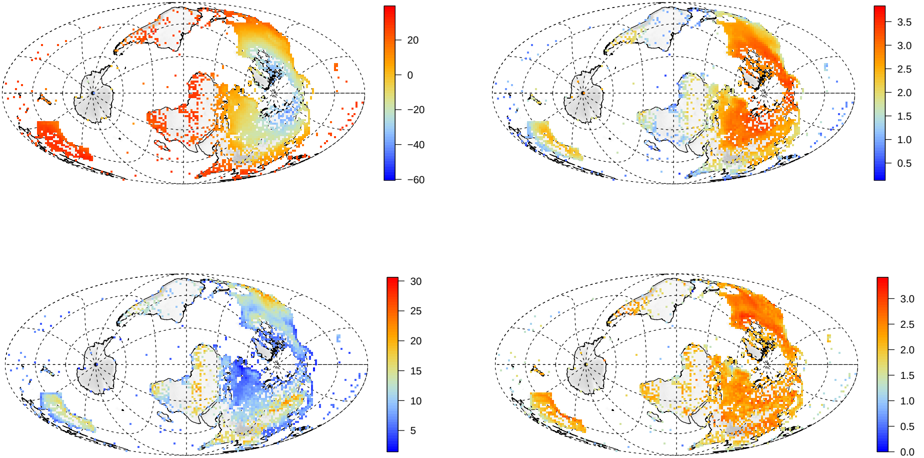




Combined covariance function

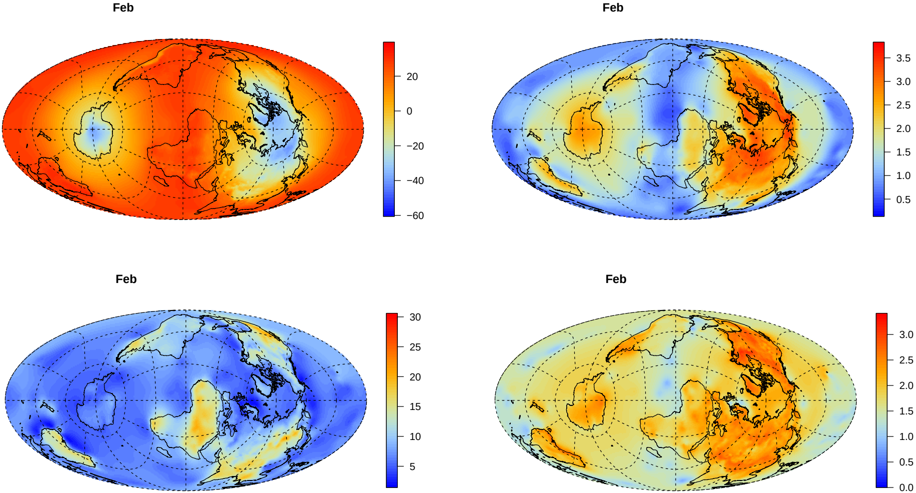


Median & scale for daily means and ranges



February climatology

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 $\boldsymbol{Q}_x$:

$$
\begin{aligned}
(\boldsymbol{x} \mid \boldsymbol{\theta}) &\sim \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{Q}_x^{-1}) && \text{(Prior)} \\
(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}) &\sim \mathcal{N}(h(\boldsymbol{x}), \boldsymbol{Q}_{y|x}^{-1}) && \text{(Observations)} \\
p(\boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{\theta}) &\propto p(\boldsymbol{x} \mid \boldsymbol{\theta})\, p(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}) && \text{(Posterior)}
\end{aligned}
$$

Linear Gaussian observations

For a linear $h(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$,

$$
\begin{aligned}
(\boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{\theta}) &\sim \mathcal{N}(\widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{Q}}^{-1}) && \text{(Posterior)} \\
\widetilde{\boldsymbol{Q}} &= \boldsymbol{Q}_x + \boldsymbol{A}^\top \boldsymbol{Q}_{y|x} \boldsymbol{A} \\
\widetilde{\boldsymbol{\mu}} &= \boldsymbol{\mu}_x + \widetilde{\boldsymbol{Q}}^{-1} \boldsymbol{A}^\top \boldsymbol{Q}_y (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\mu}_x)
\end{aligned}
$$

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

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

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

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

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

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

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

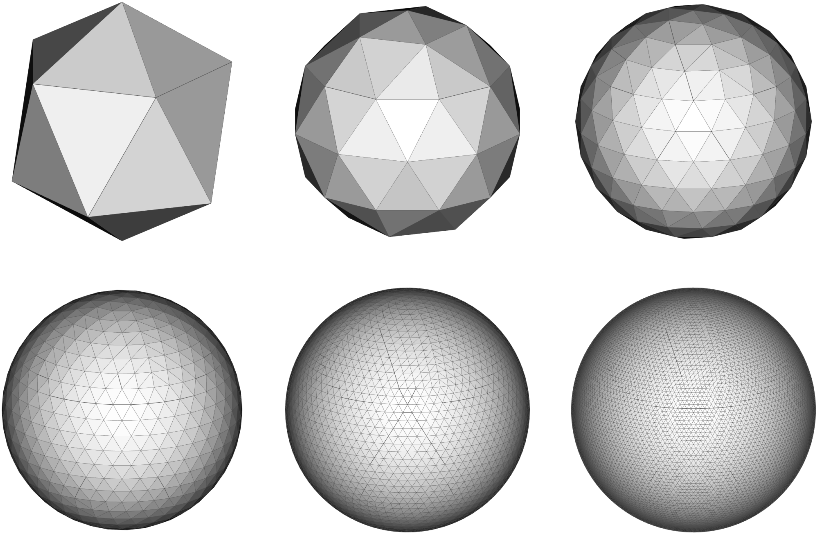
$$360 \cdot 180 \cdot 4^2 \cdot 365 \cdot 165 \cdot 2 = 124,882,560,000$$

Storing $\sim 10^{11}$ latent variables as `double` takes ~ 1 TB
(And that just covers the finest scale)

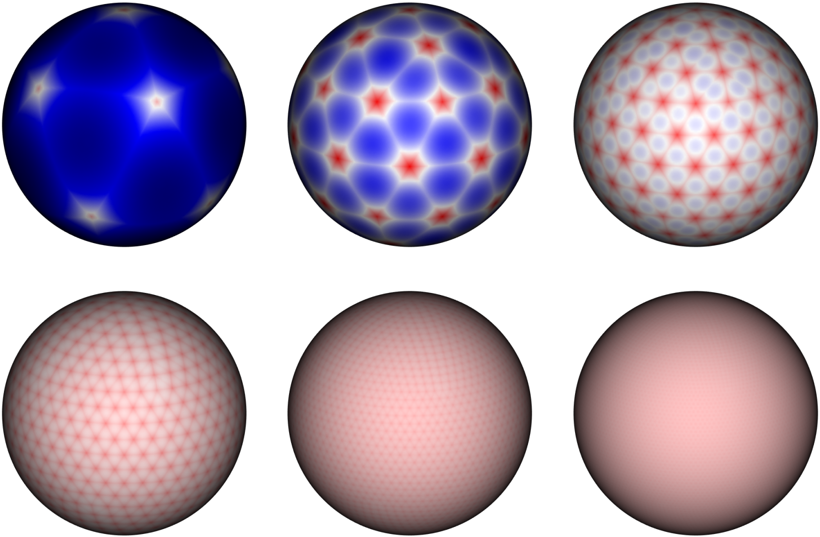
To store the data (> 10 TB), model information, and estimated uncertainties we need a computing cluster with lots of RAM and fast temporary parallell disk access.

Matrix-free iterative solvers will be our saviours!

Triangulations for all corners of Earth



Triangulations for all corners of Earth



Domain decomposition and multigrid

Overlapping domain decomposition

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

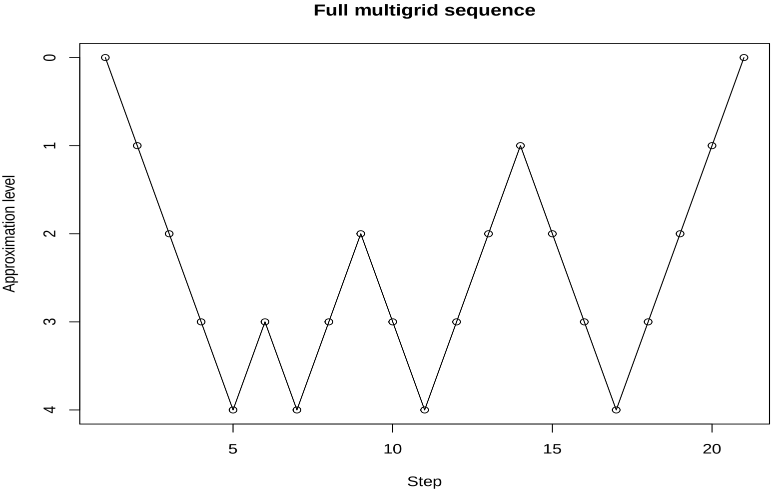
$$\boldsymbol{M}^{-1}\boldsymbol{x} = \sum_{k=1}^{K} \boldsymbol{W}_k \boldsymbol{B}_k (\boldsymbol{B}_k^\top \boldsymbol{Q} \boldsymbol{B}_k)^{-1} \boldsymbol{B}_k^\top \boldsymbol{W}_k \boldsymbol{x}$$

Multigrid

Let $\boldsymbol{B}_c^\top$ be a projection matrix to a coarse approximative model. Then a basic multigrid step for $\boldsymbol{Q}\boldsymbol{x} = \boldsymbol{b}$ is

1. Apply high frequency preconditioner to get $\widehat{\boldsymbol{x}}_0$, let $\boldsymbol{r}_0 = \boldsymbol{b} - \boldsymbol{Q}\widehat{\boldsymbol{x}}_0$
2. Project the problem to the coarser model: $\boldsymbol{Q}_c = \boldsymbol{B}_c^\top \boldsymbol{Q} \boldsymbol{B}_c$, $\boldsymbol{r}_c = \boldsymbol{B}_c^\top \boldsymbol{r}_0$
3. Apply multigrid to $\boldsymbol{Q}_c \boldsymbol{x}_c = \boldsymbol{r}_c$
4. Update the solution: $\widehat{\boldsymbol{x}}_1 = \widehat{\boldsymbol{x}}_0 + \boldsymbol{B}_c \widehat{\boldsymbol{x}}_c$
5. Apply high frequency preconditioner to get $\widehat{\boldsymbol{x}}_2$

Full multigrid



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

Sparse partial inverse

Takahashi recursions compute $\boldsymbol{S}$ such that $\boldsymbol{S}_{ij} = (\boldsymbol{Q}^{-1})_{ij}$ for all $\boldsymbol{Q}_{ij} \neq 0$. 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.

Method overview

- Hierarchical timescale combination of space-time random fields
- Preprocessing to estimate model parameters and non-Gaussianity
- Iterated linearisation in approximate Newton optimisation
- Distributed Preconditioned Conjugate Gradient solves
- Information is passed between the scales with the aid of approximate Schur complements
- Within each scale, approximate multigrid solves
- Overlapping space-time domain decomposition within each multigrid level
- Direct Monte Carlo sampling: add suitable randomness to the RHS of the $\boldsymbol{Q}_{x|y}$ solves for $\widetilde{\boldsymbol{\mu}}$.
- Rao-Blackwellised variance estimation

Parameter estimation:
In the project, several ad hoc methods are used.
In general, several approaches to get $\log\det\boldsymbol{Q}$ and/or $\frac{\partial}{\partial\theta}\log\det\boldsymbol{Q} = \operatorname{tr}\left(\boldsymbol{S}\frac{\partial\boldsymbol{Q}}{\partial\theta}\right)$, but much more work is needed to handle complex models.

Gratuitous commercial: `inlabru`, the friendlier INLA interface

R-INLA

```
A.data <- inla.spde.make.A(...)
A.pred <- inla.spde.make.A(...)
stack.data <- inla.stack(data=..., A=list(A.data, ...), effects=...)
stack.pred <- inla.stack(data=..., A=list(A.pred, ...), effects=...)
stack <- inla.stack(stack.data, stack.pred)
formula <- y ~ ... + f(field, model=spde)
result <- inla(...)
## Linear prediction:
prediction <- result$summary.fitted.values[some.indices, "mean"]
```

`http://inlabru.org`

```
components <- ~ ... + field(map=coordinates, model=spde)
formula <- y ~ ... + field
result <- bru(...)
## Non-linear prediction (via direct posterior sampling)
prediction <- predict(..., cos(field))
## Extra: non-linear formulas and marked LGCP capabilities
formula <- y ~ field1 * exp(field2)
formula <- coordinates + size ~ field1 + dnorm(size, field2, sd=exp(theta),
                                               log=TRUE)
```