Multiscale spatio-temporal modelling and large scale computation

Finn Lindgren

UNIVERSITY OF BATH

From September: University of Edinburgh

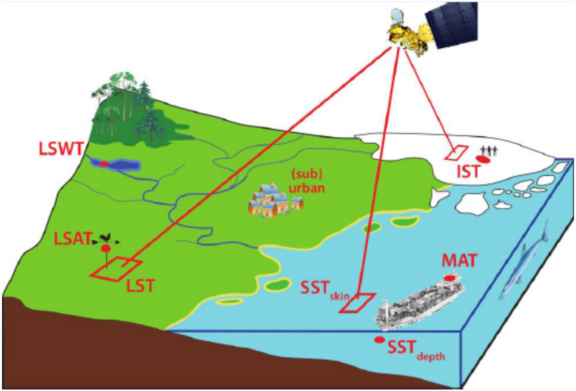
Smögen 2016

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.

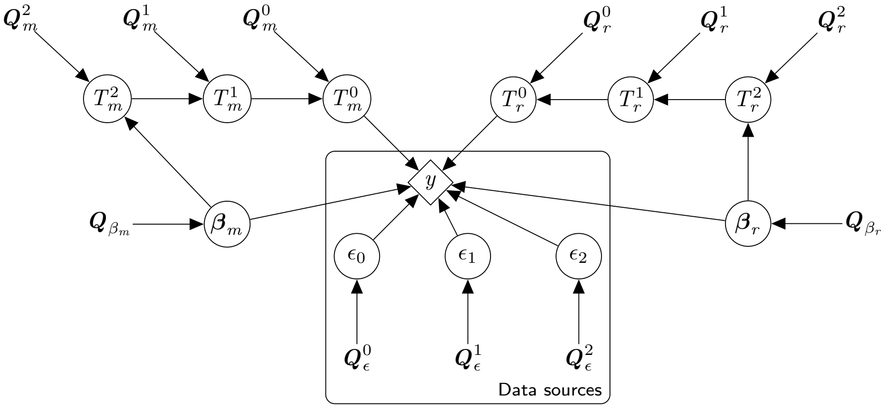


UNIVERSITY OF BATH

EUSTACE

Partial hierarchical representation

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



Conditional specifications, e.g.

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

UNIVERSITY OF BATH

EUSTACE

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 *copula*, or 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).

UNIVERSITY OF BATH

EUSTACE

Data relationships

The grouping of the multiscale components into cumulative sums of increasingly small scale components allows observations to be linked directly only to T_m^0, T_r^0 (or T_r^1 and U_r^0), β_m, and β_r, which helps the computational methods.

The different data sources don't measure the same thing as each other:

Temperature relationship definitions

2 degrees of freedom, but 4 possible types of observations:

$$T_m = \frac{T_n + T_x}{2} \qquad T_n = T_m - T_r/2$$

$$T_r = T_x - T_n \qquad T_x = T_m + T_r/2$$

Note: This *defines* daily mean temperature based on T_n and T_x.
We may also need $T_{\overline{m}} = \frac{1}{|\text{day}|} \int_{\text{day}} T(t)\,\mathrm{d}t = T_m + \epsilon$.

UNIVERSITY OF BATH

EUSTACE

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

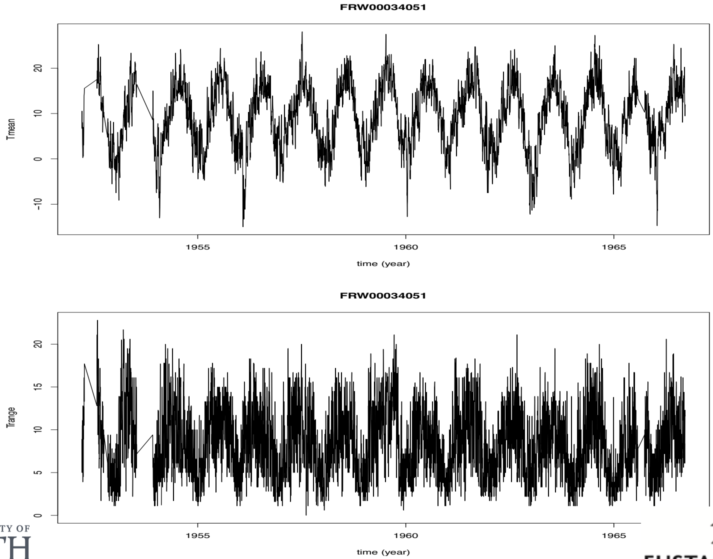
For station k at day t_i

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

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



UNIVERSITY OF BATH

EUSTACE

Power tail quantile (POQ) model

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

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

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

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

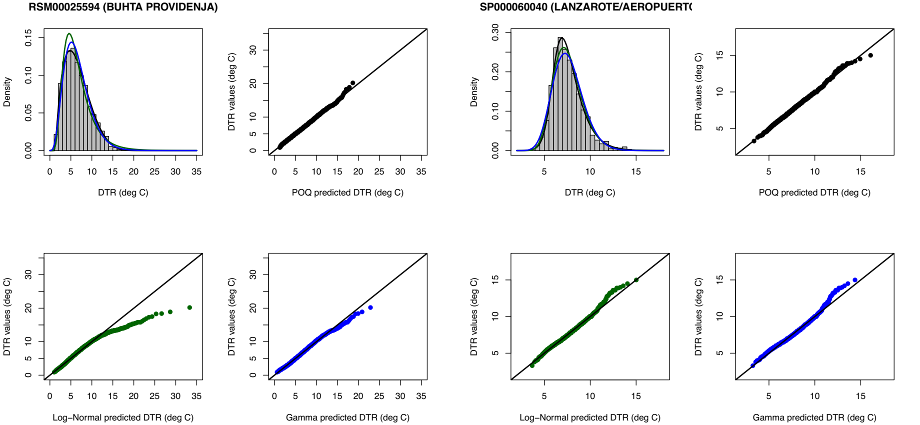
The parameters $\boldsymbol{\theta} = (\theta_0, \theta_1 = \log\tau, \theta_2 = \operatorname{logit}[(\gamma+1)/2], \theta_3, \theta_4)$ control the median, spread/scale, skewness, and the left and right tail shape. This model is also known as the *five parameter lambda model*.

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

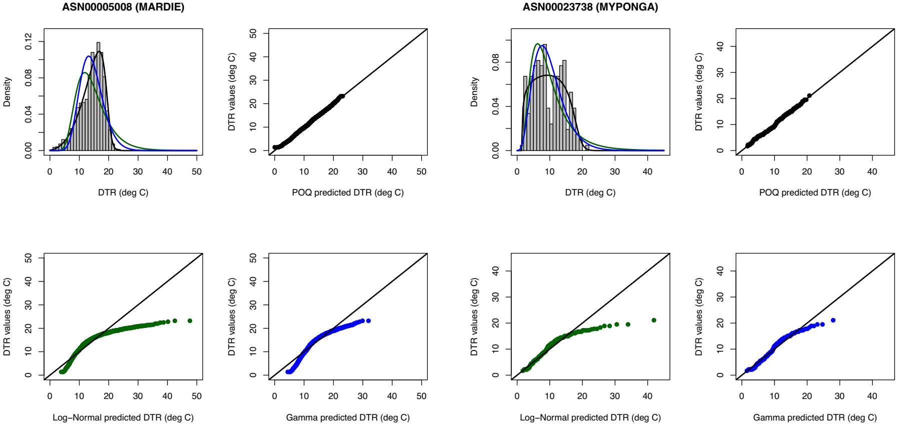


For these stations, POQ does a slightly better job than a Gamma distribution.

UNIVERSITY OF BATH

EUSTACE

Diurnal range distributions

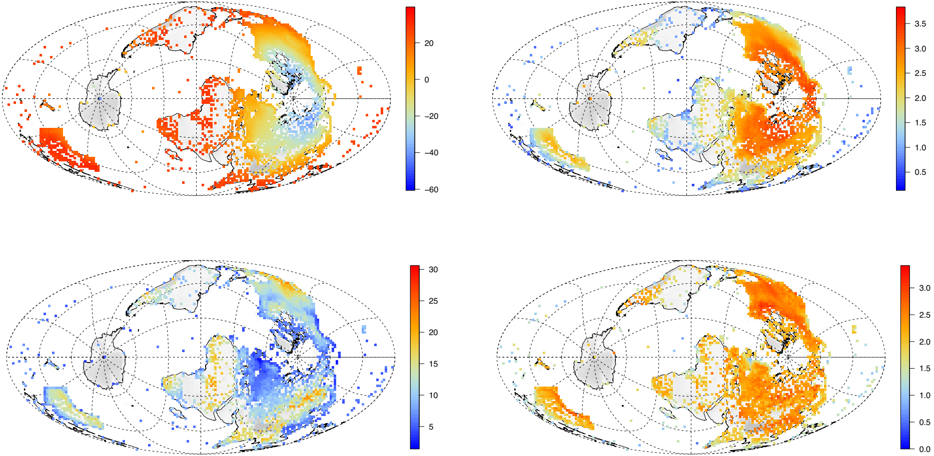


For these stations only POQ comes close to representing the distributions.
Note: Some of the mixture-like distribution shapes may be an effect of unmodeled station inhomogeneities.

UNIVERSITY OF BATH

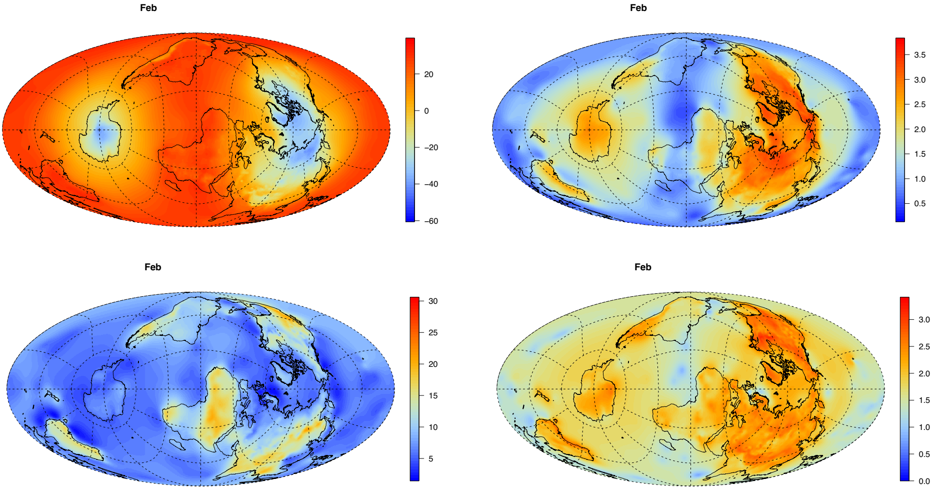
EUSTACE

Median & scale for daily means and ranges



February climatology

Estimates of median & scale for T_m and T_r

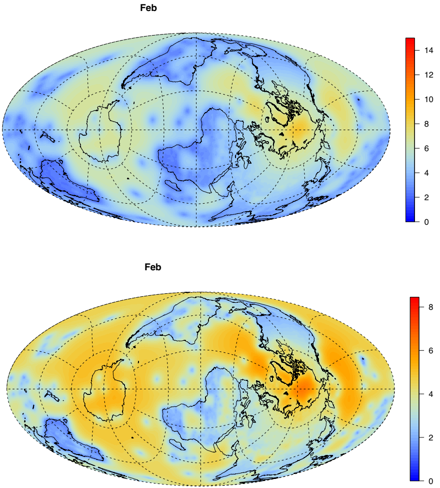


February climatology

UNIVERSITY OF BATH

EUSTACE

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



February climatology

Linearised inference

Spatio-temporal latent random processes (u), geographical effects ($\boldsymbol{\beta}$), station and other persistent effects ($\boldsymbol{b}$).

$$(\boldsymbol{u}, \boldsymbol{\beta}, \boldsymbol{b} \mid \boldsymbol{\theta}) \sim \mathcal{N}(\boldsymbol{\mu}_{u\beta b}, \boldsymbol{Q}_{u\beta b}^{-1}) \qquad \text{(Prior)}$$

$$(\boldsymbol{y} \mid \boldsymbol{u}, \boldsymbol{\beta}, \boldsymbol{b}) \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{u} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{b}, \boldsymbol{Q}_y^{-1}) \qquad \text{(Observations)}$$

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

$$\widetilde{\boldsymbol{Q}} = \boldsymbol{Q}_{u\beta b} + \begin{bmatrix}\boldsymbol{A} & \boldsymbol{X} & \boldsymbol{Z}\end{bmatrix}^\top \boldsymbol{Q}_y \begin{bmatrix}\boldsymbol{A} & \boldsymbol{X} & \boldsymbol{Z}\end{bmatrix}$$

$$\widetilde{\boldsymbol{\mu}} = \boldsymbol{\mu}_{u\beta b} + \widetilde{\boldsymbol{Q}}^{-1} \begin{bmatrix}\boldsymbol{A} & \boldsymbol{X} & \boldsymbol{Z}\end{bmatrix}^\top \boldsymbol{Q}_y \left(\boldsymbol{y} - \begin{bmatrix}\boldsymbol{A} & \boldsymbol{X} & \boldsymbol{Z}\end{bmatrix} \boldsymbol{\mu}_{u\beta b}\right)$$

Gaussian posterior approximation for non-linear observations

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

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

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

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

Posterior calculations

Simpified 2-step multiscale precision matrix block structure:

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

can be pseudo-Cholesky-factorised:

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

Posterior expectation, samples, and marginal variances:

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

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

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

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

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

Posterior calculations

Heat equation precision:

$Q = \sum_{k=0}^{p} \theta_k Q_t^{(k)} \otimes Q_s^{(k)}$

Simpified 2-step multiscale precision matrix block structure:

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

can be pseudo-Cholesky-factorised:

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

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!

Preconditioning for iterative solvers

Solving $Qx = b$ is equivalent to solving $M^{-1}Qx = M^{-1}b$. Choosing M^{-1} as an approximate inverse to Q gives a less ill-conditioned system. Only the *action* of M^{-1} is needed, e.g. one or more fixed point iterations:

Block Jacobi and Gauss-Seidel preconditioning

Matrix split: $\boldsymbol{Q}_{x|y} = \boldsymbol{L} + \boldsymbol{D} + \boldsymbol{L}^\top$

Jacobi: $\boldsymbol{x}^{(k+1)} = \boldsymbol{D}^{-1}\left(-(\boldsymbol{L} + \boldsymbol{L}^\top)\boldsymbol{x}^{(k)} + \boldsymbol{b}\right)$

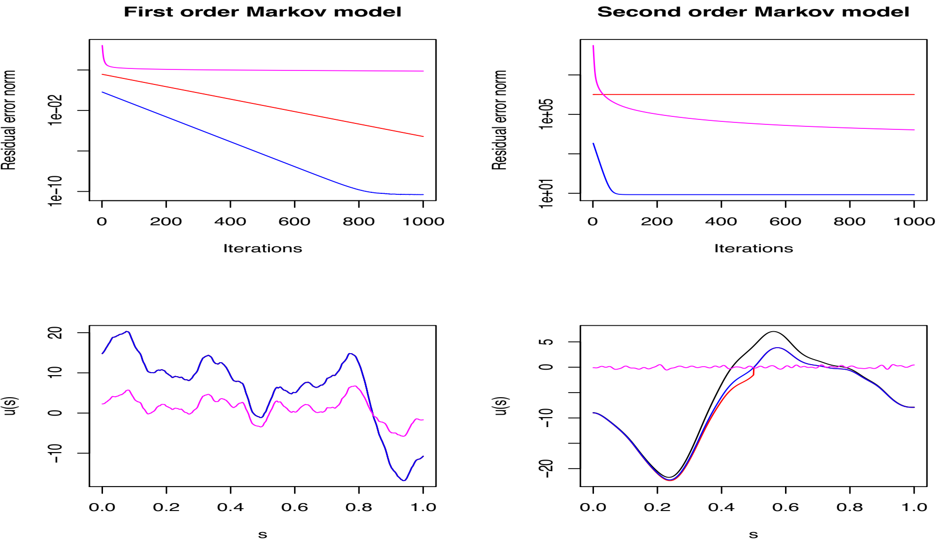
Gauss-Seidel: $\boldsymbol{x}^{(k+1)} = (\boldsymbol{L} + \boldsymbol{D})^{-1}\left(-\boldsymbol{L}^\top\boldsymbol{x}^{(k)} + \boldsymbol{b}\right)$

Remark: Block Gibbs sampling for a GMRF posterior

With $\boldsymbol{Q} = \boldsymbol{Q}_{x|y}$, $\boldsymbol{b} = \boldsymbol{A}^\top\boldsymbol{Q}_\epsilon(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\mu}_x)$ and $\widetilde{\boldsymbol{x}} = \boldsymbol{x} - \boldsymbol{\mu}_x$,

$$\widetilde{\boldsymbol{x}}^{(k+1)} = (\boldsymbol{L} + \boldsymbol{D})^{-1}\left(-\boldsymbol{L}^\top\widetilde{\boldsymbol{x}}^{(k)} + \boldsymbol{b} + \widetilde{\boldsymbol{L}}_D\boldsymbol{w}\right), \quad \boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$$

Gauss-Seidel and Gibbs are both inefficient on their own, but G-S leads to useful preconditioners. Convergence testing is much easier for linear solvers than for MCMC.



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!

UNIVERSITY OF BATH

EUSTACE

Use *overlapping blocks* distributed over many computing nodes, and apply approximate 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 $\boldsymbol{Q}_t \otimes \boldsymbol{Q}_a + \boldsymbol{A}^\top \boldsymbol{Q}_\epsilon \boldsymbol{A}$, and one solve with the *Schur complement*

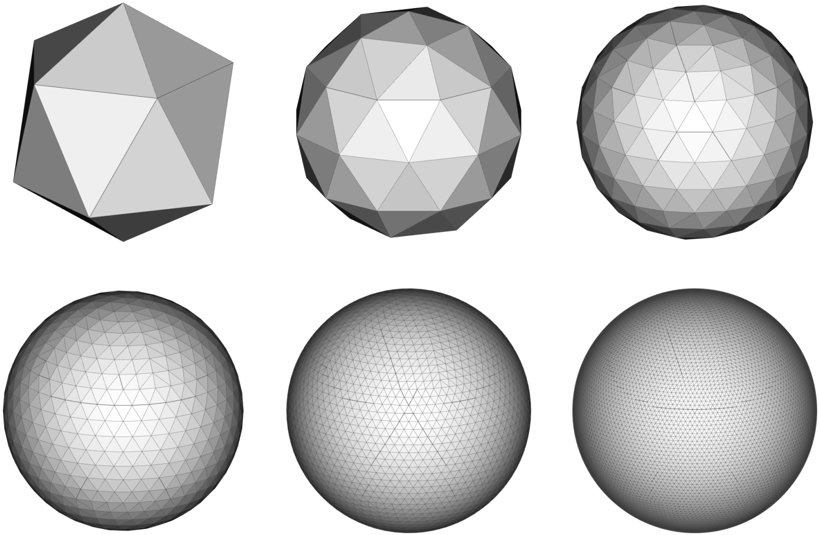
$$\boldsymbol{Q}_z + \boldsymbol{B}^\top \boldsymbol{Q}_t \boldsymbol{B} \otimes \boldsymbol{Q}_a - \boldsymbol{B}^\top \boldsymbol{Q}_t \otimes \boldsymbol{Q}_a \left(\boldsymbol{Q}_t \otimes \boldsymbol{Q}_a + \boldsymbol{A}^\top \boldsymbol{Q}_\epsilon \boldsymbol{A} \right)^{-1} \boldsymbol{Q}_t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

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

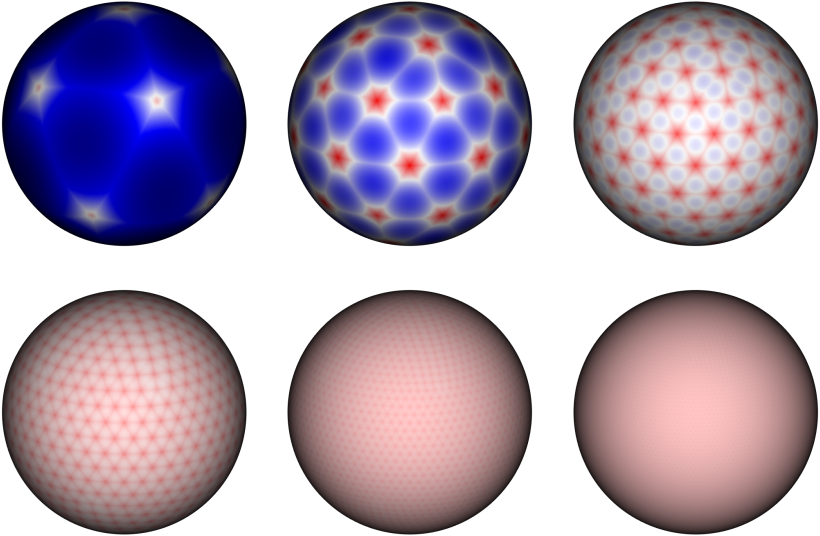
$$\begin{bmatrix} \widetilde{\boldsymbol{Q}}_B + \widetilde{\boldsymbol{B}}^\top \boldsymbol{A}^\top \boldsymbol{Q}_\epsilon \boldsymbol{A} \widetilde{\boldsymbol{B}} & -\widetilde{\boldsymbol{Q}}_B \\ -\widetilde{\boldsymbol{Q}}_B & \boldsymbol{Q}_z + \widetilde{\boldsymbol{Q}}_B \end{bmatrix} \begin{bmatrix}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

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

Triangulations for all corners of Earth



Triangulations for all corners of Earth

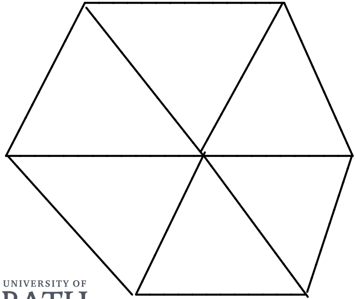


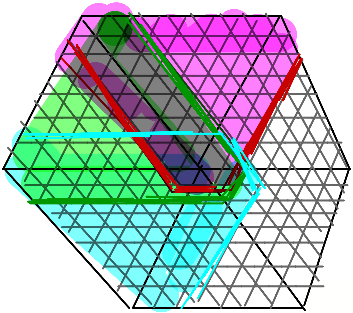
Domain decomposition

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 the additive Schwartz preconditioner with coarse correction is given by

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





UNIVERSITY OF BATH

EUSTACE

Variance calculations

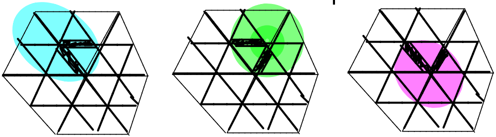
Basic Rao-Blackwellisation of sample estimators

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 $\Omega_k \subset \Omega$,

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

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

New idea from `S=inla.qinv(Q)` ($S_{ij} = \left(\boldsymbol{Q}^{-1}\right)_{ij}$ for all $\{i, j; Q_{ij} \neq 0\}$):
Iterative solver for the covariances on subdomain interfaces, with boundary conditioned $\boldsymbol{S}$-evaluations as preconditioner.



UNIVERSITY OF BATH

EUSTACE

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://people.bath.ac.uk/fl353/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/`

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

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

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.

UNIVERSITY OF BATH

EUSTACE