# Multiscale spatio-temporal modelling and large scale computation

**Finn Lindgren**

University of Bath

From September: University of Edinburgh

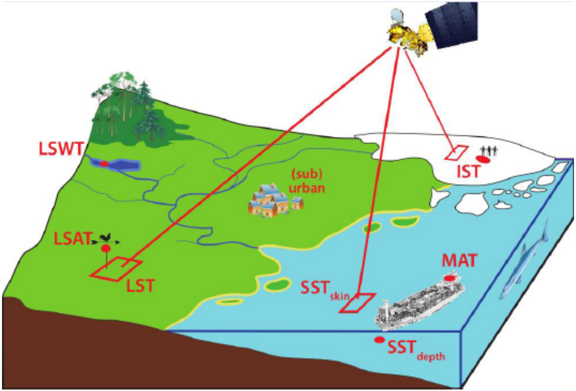
## NORDSTAT 2016

*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.

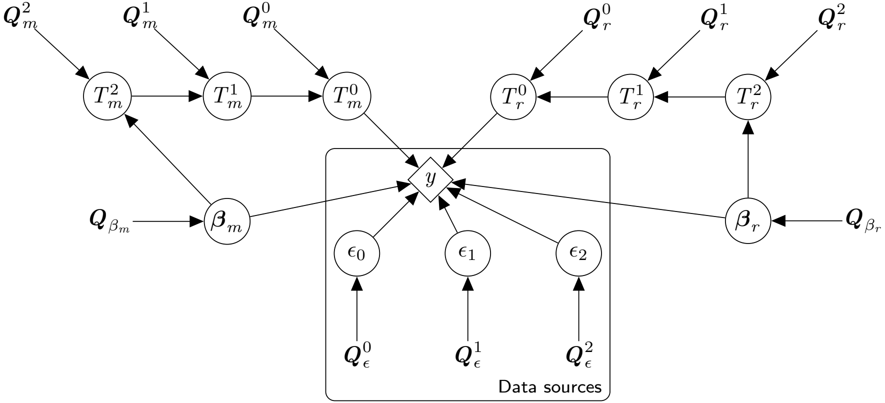



UNIVERSITY OF BATH

EUSTACE

# Partial hierarchical representation

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




Conditional specifications, e.g.

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

# 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. The daily means $T_m(\mathbf{s},t)$ and diurnal ranges $T_r(\mathbf{s},t)$ are defined through

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

$$T_r(\mathbf{s},t) = G\left(U_r^0(\mathbf{s},t);\ \mu_r(\mathbf{s},t)\right),$$

$$\mu_r(\mathbf{s},t) = U_r^1(\mathbf{s},t) + U_r^2(\mathbf{s},t) + U_r^S(\mathbf{s},t) + \sum_{i=1}^{N_X} X_i(\mathbf{s},t)\beta_r^{(i)},$$

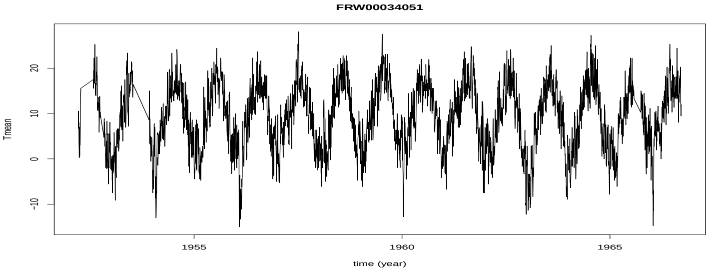
where $G$ is a *copula* or non-linear transformation function, controlled by the slowly varying median process $\mu_r(\mathbf{s},t)$ as well as some fixed seasonal fields of distribution scale and shape parameters. The $\beta_m$ and $\beta_r$ coefficients are weights for covariates $X_i(\mathbf{s},t)$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

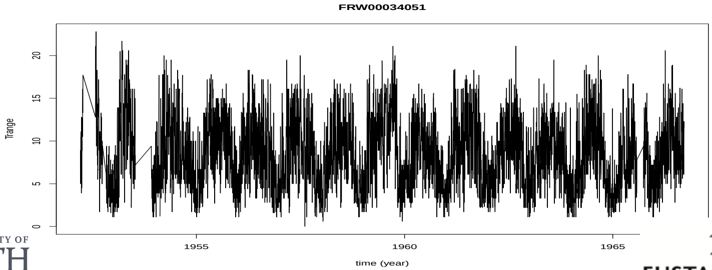
UNIVERSITY OF BATH

EUSTACE

# Observed data

Observed daily $T_{\text{mean}}$ and $T_{\text{range}}$ for station FRW00034051




UNIVERSITY OF BATH

EUSTACE

# 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

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

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

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

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

UNIVERSITY OF BATH

EUSTACE

# 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.

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

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

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

$$F_{\boldsymbol{\theta}}^{-1}(p) = \theta_0 + \tau(1-\gamma) f_{\theta_3}^-(p) + \tau\gamma f_{\theta_4}^+(p),$$

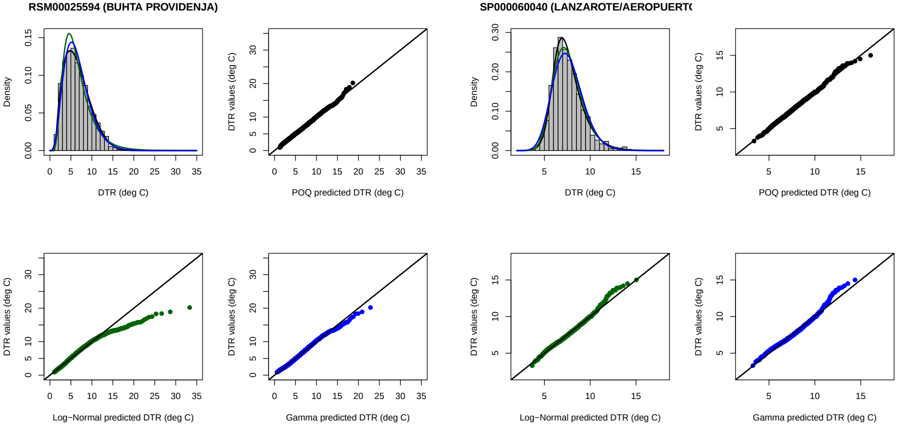
The parameters $\boldsymbol{\theta} = (\theta_0, \theta_1 = \log\tau, \theta_2 = \operatorname{logit}(\gamma), \theta_3, \theta_4)$ control the median, spread/scale, skewness, and the left and right tail shape. This model is also known as the *five parameter lambda model*.

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

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

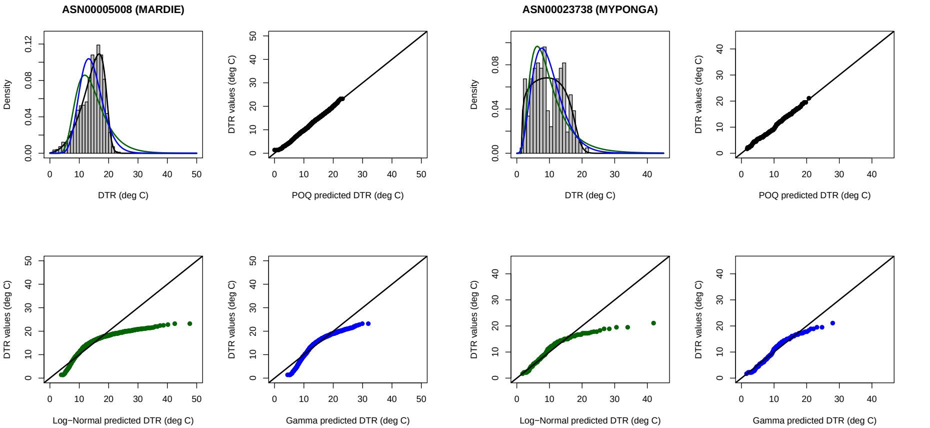
where the parameters can vary with space and time.

# Diurnal range distributions




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

# Diurnal range distributions

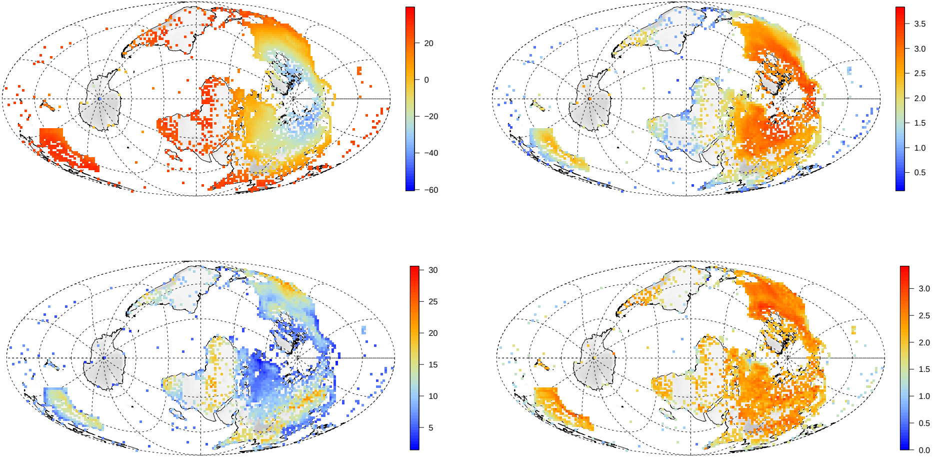



For these stations only POQ comes close to representing the distributions.
Note: Some of the mixture-like distribution shapes may be an effect of unmodeled station inhomogeneities.

UNIVERSITY OF BATH

EUSTACE

# Median & scale for daily means and ranges




June climatology

UNIVERSITY OF BATH

EUSTACE

# Posterior calculations

Example 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 (with $\widetilde{\boldsymbol{A}} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{0} \end{bmatrix}$):

$$\begin{aligned} \boldsymbol{Q}_{x|y}(\boldsymbol{\mu}_{x|y} - \boldsymbol{\mu}_x) &= \widetilde{\boldsymbol{A}}^\top \boldsymbol{Q}_\epsilon (\boldsymbol{y} - \widetilde{\boldsymbol{A}} \boldsymbol{\mu}_x), \\ \boldsymbol{Q}_{x|y}(\boldsymbol{x} - \boldsymbol{\mu}_{x|y}) &= \widetilde{\boldsymbol{L}}_{x|y} \boldsymbol{w}, \quad \boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}), \quad \text{or} \\ \boldsymbol{Q}_{x|y}(\boldsymbol{x} - \boldsymbol{\mu}_x) &= \widetilde{\boldsymbol{A}}^\top \boldsymbol{Q}_\epsilon (\boldsymbol{y} - \widetilde{\boldsymbol{A}} \boldsymbol{\mu}_x) + \widetilde{\boldsymbol{L}}_{x|y} \boldsymbol{w}, \quad \boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}), \\ \mathsf{Var}(x_i | \boldsymbol{y}) &= \texttt{diag(inla.qinv(}\boldsymbol{Q}_{x|y}\texttt{))} \quad \text{(requires Cholesky)} \end{aligned}$$

UNIVERSITY OF BATH

EUSTACE

# Preconditioning for iterative solvers

Solving $\boldsymbol{Q}\boldsymbol{x} = \boldsymbol{b}$ is equivalent to solving $\boldsymbol{M}^{-1}\boldsymbol{Q}\boldsymbol{x} = \boldsymbol{M}^{-1}\boldsymbol{b}$. Choosing $\boldsymbol{M}^{-1}$ as an approximate inverse to $\boldsymbol{Q}$ gives a less ill-conditioned system. Only the *action* of $\boldsymbol{M}^{-1}$ 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: $\boldsymbol{x}^{(k+1)} = \boldsymbol{D}^{-1}\left(-(\boldsymbol{L} + \boldsymbol{L}^\top)\boldsymbol{x}^{(k)} + \boldsymbol{b}\right)$

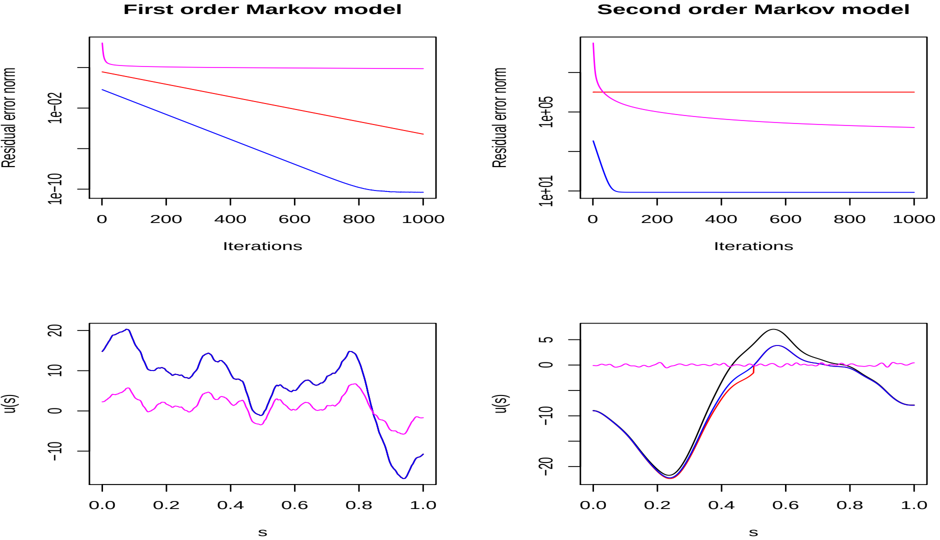
Gauss-Seidel: $\boldsymbol{x}^{(k+1)} = (\boldsymbol{L} + \boldsymbol{D})^{-1}\left(-\boldsymbol{L}^\top\boldsymbol{x}^{(k)} + \boldsymbol{b}\right)$

## 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*

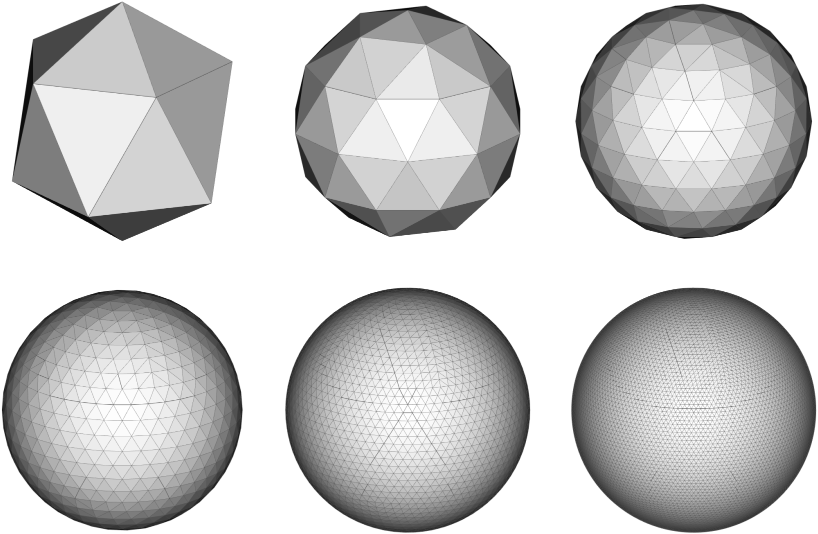
$$\boldsymbol{Q}_z + \boldsymbol{B}^\top \boldsymbol{Q}_t \boldsymbol{B} \otimes \boldsymbol{Q}_a - \boldsymbol{B}^\top \boldsymbol{Q}_t \otimes \boldsymbol{Q}_a \left( \boldsymbol{Q}_t \otimes \boldsymbol{Q}_a + \boldsymbol{A}^\top \boldsymbol{Q}_\epsilon \boldsymbol{A} \right)^{-1} \boldsymbol{Q}_t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

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.

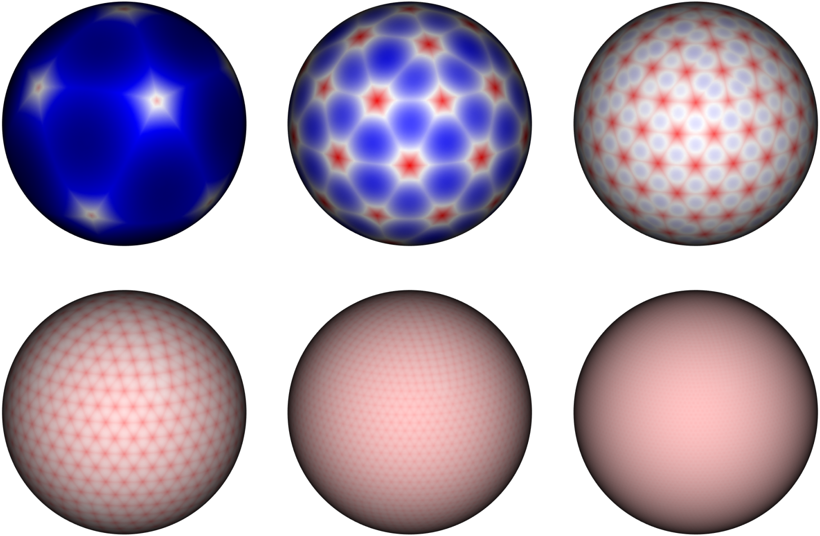
# Triangulations for all corners of Earth



UNIVERSITY OF BATH

EUSTACE

# Triangulations for all corners of Earth



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/`

# Observation models

## Satellite data error model

The observational&calibration errors are modelled as three error components:
independent ($\epsilon_0$), spatially correlated ($\epsilon_1$), and systematic ($\epsilon_2$), 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.

UNIVERSITY OF BATH

EUSTACE