Royal Netherlands Meteorological Institute
Ministry of Infrastructure and the Environment

Simultaneous modelling and estimation of climate and weather

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

THE UNIVERSITY *of* EDINBURGH

TIES, Bergamo 2017-07-24

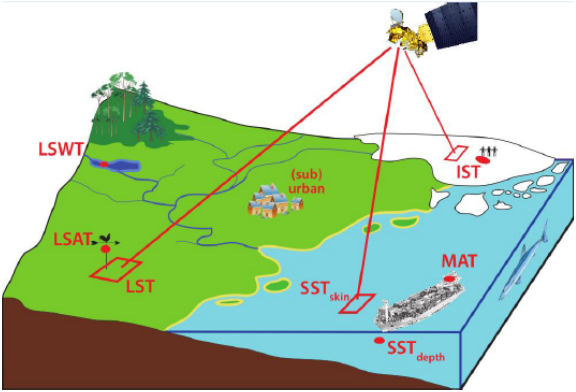
EUSTACE has received funding from the European Union's Horizon 2020 Programme for Research and Innovation, under Grant Agreement no 640171

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.



Covariance functions and SPDEs

The Matérn covariance family on $\mathbb{R}^d$

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

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.

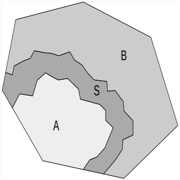
GMRFs: Gaussian Markov random fields

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*

$\{x(\mathcal{A}) \perp x(\mathcal{B}) | x(\mathcal{S})$, 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

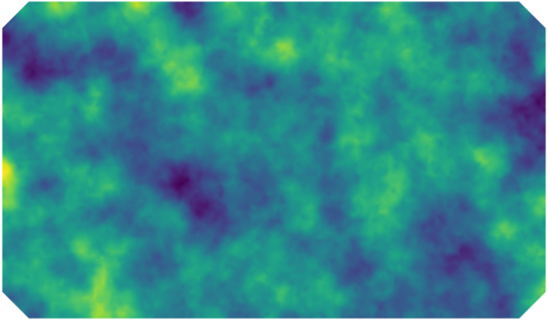
$\boldsymbol{x} = (x_1, \ldots, x_n) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{Q}^{-1})$ is Markov with respect to a neighbourhood structure $\{\mathcal{N}_i, i = 1, \ldots, n\}$ if $Q_{ij} = 0$ whenever $j \neq \mathcal{N}_i \cup i$.

- 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.

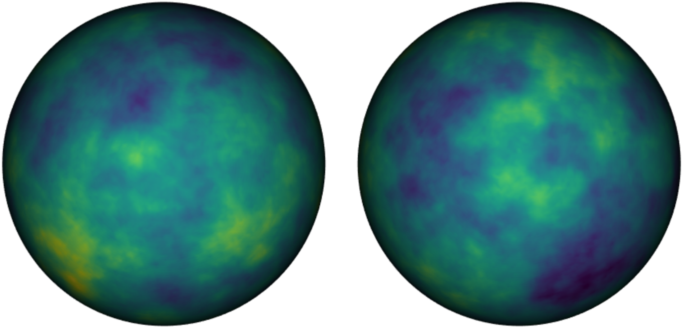
$$(\kappa^2 - \Delta)(\tau\, x(\mathbf{s})) = \mathcal{W}(\mathbf{s}), \quad \mathbf{s} \in \mathbb{R}^d$$



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.

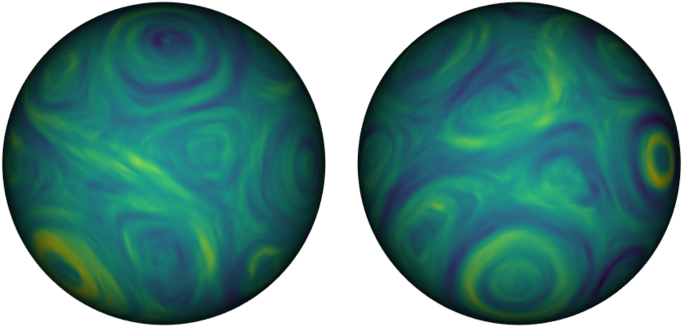
$$(\kappa^2 - \Delta)(\tau\, x(\mathbf{s})) = \mathcal{W}(\mathbf{s}), \quad \mathbf{s} \in \Omega$$



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.

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



EUSTACE

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

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

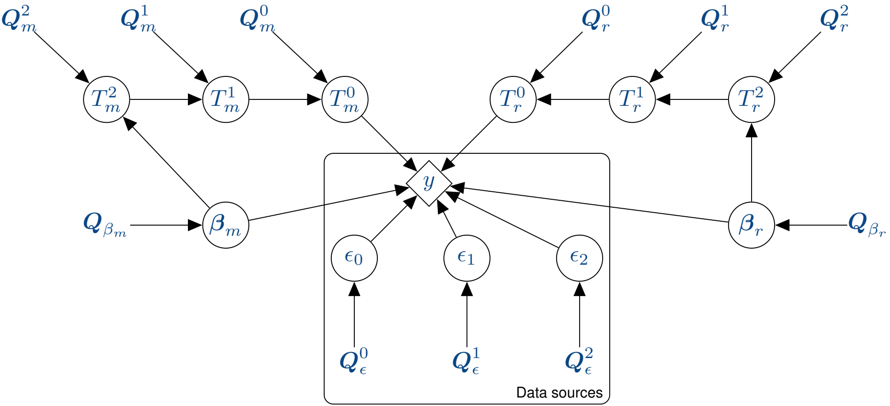
  which leads to Matérn covariances marginally in space, and in time for each spatial frequency.
- The finite element approximation has precision matrix structure

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

  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

Daily mean temperatures

The daily means $T_m(\mathbf{s},t)$ are accumulation of independent fields and covariate effects,

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

Daily temperature range (diurnal range)

The diurnal ranges $T_r(\mathbf{s},t)$ are defined through

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

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

where G^{-1} is a spatially and seasonally varying quantile model.

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, 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

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

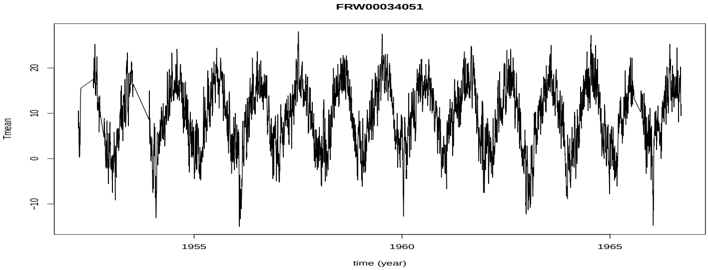
where $H_j^k(t)$ are temporal step functions, $e_m^{k,j}$ are latent bias variables, and $\epsilon_m^{k,i}$ are independent measurement and discretisation errors.

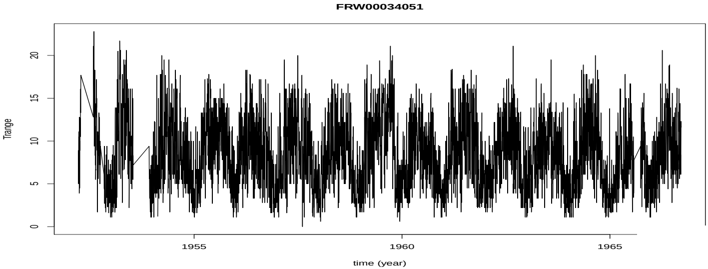
THE UNIVERSITY OF EDINBURGH

EUSTACE

Observed data

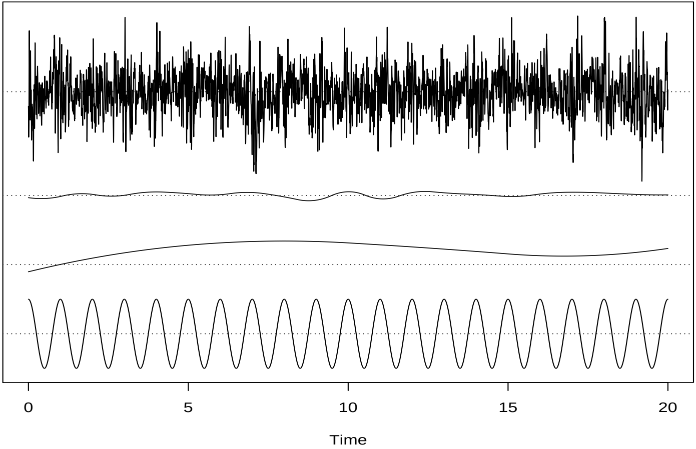
Observed daily T_{mean} and T_{range} for station FRW00034051



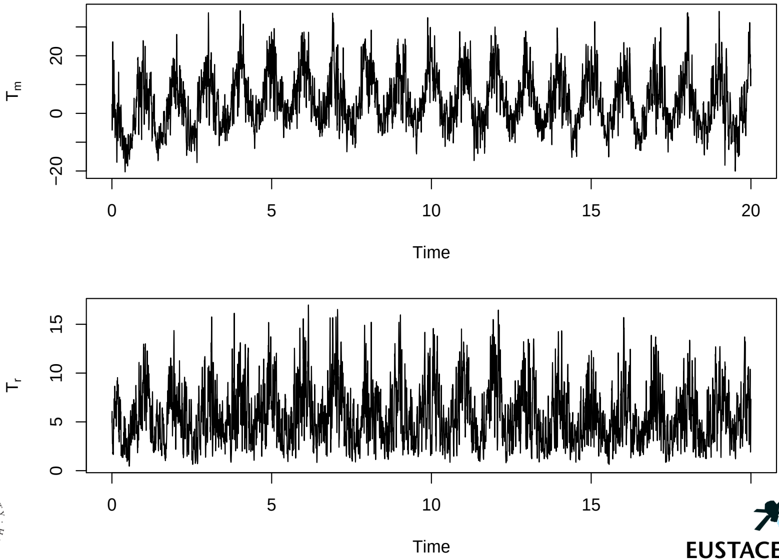


Multiscale model component samples

The result of sampling based on hand-picked temporal process parameters

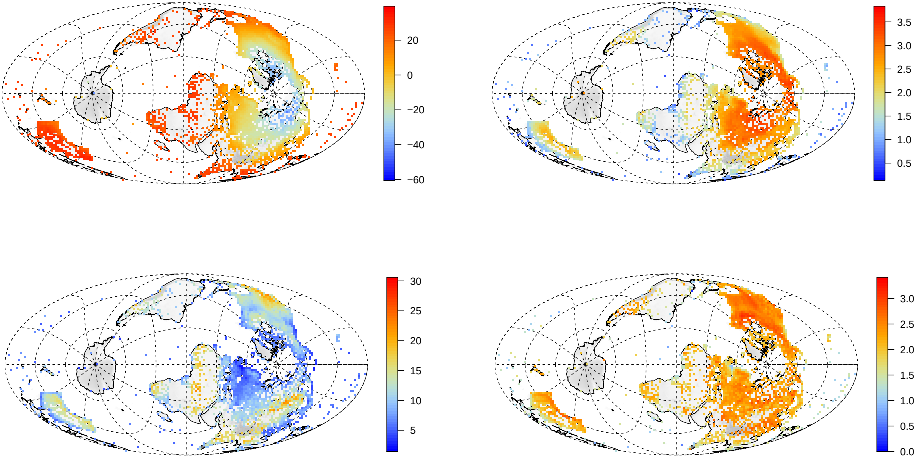


Combined model samples for T_m and T_r



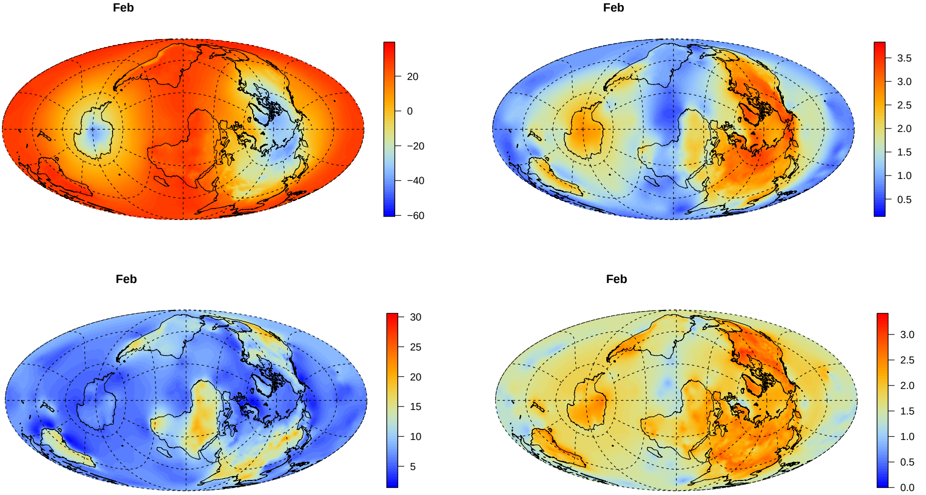
EUSTACE

Median & scale for daily means and ranges



February climatology

Estimates of median & scale for T_m and T_r



February climatology

Linear 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 $\boldsymbol{Q}_x$:

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

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

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

Linear Gaussian observations

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

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

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

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

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 $\boldsymbol{Q}_x$:

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

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

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

Non-linear and/or non-Gaussian observations

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

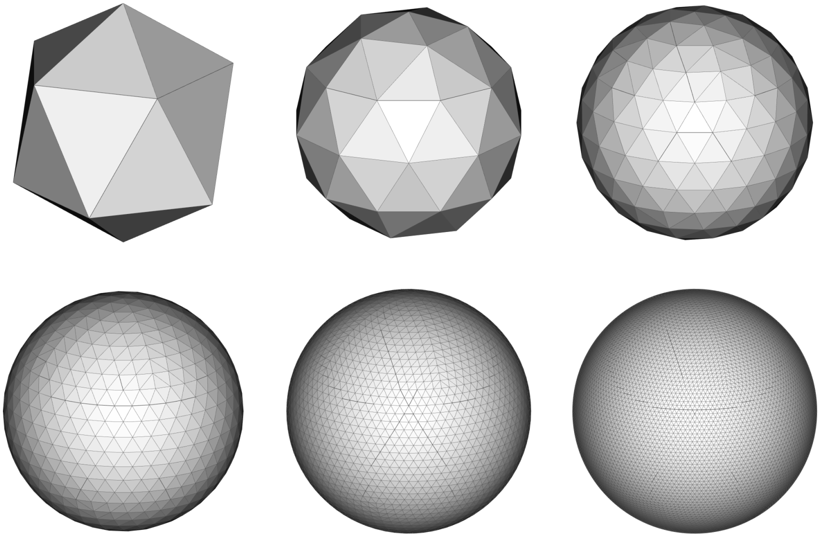
$$(\boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{\theta}) \overset{\text{approx}}{\sim} \mathcal{N}(\widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{Q}}^{-1}) \qquad \text{(Approximate posterior)}$$

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

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

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

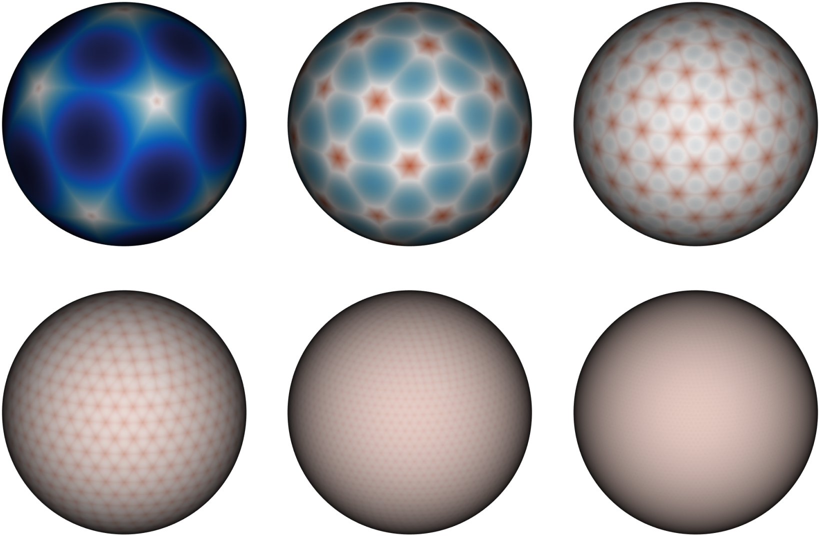
Triangulations for all corners of Earth



THE UNIVERSITY OF EDINBURGH

EUSTACE

Triangulations for all corners of Earth

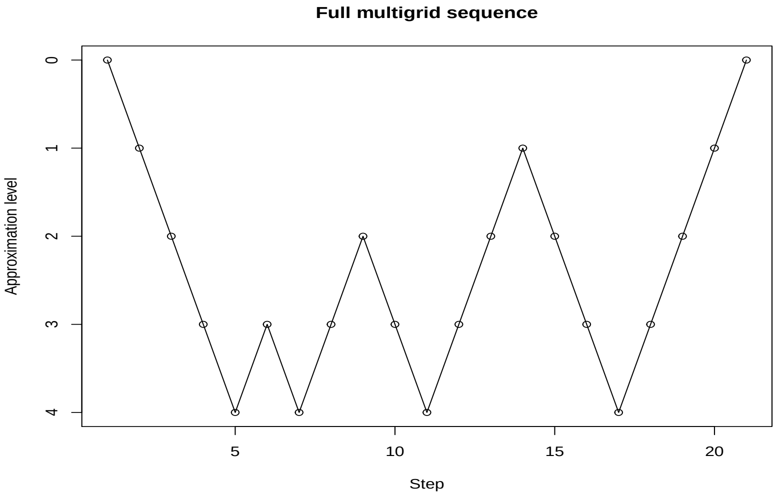


Iterative solver components

Space and time nodes: $360 \cdot 180 \cdot 4^2 \cdot 365 \cdot 165 \cdot 2 = 124,882,560,000$, or ~ 1TB! Full precision matrix storage and direct factorisation not realistic.

- Overlapping domain decomposition (DD)
  Macro-triangles linked to coarse nodes, small enough for (nearly) exact computations
- Multigrid (MG)
  A sequence of fine to coarse models; DD applied at each level
- Approximate Schur complents (Schur)
  Solve the fast timescale block with MG, then project the model to the next timescale
  Recurse through the timescales, with increasingly multivariate blocks
- Preconditioned conjugate gradients (PCG)
  Use the above methods as preconditioner, find approximate solution
- Non-linear least squares Newton optimisation
  Linearise the model, find search direction with PCG, perform simple line search, and iterate

Full multigrid



Variance calculations

Sparse partial inverse

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.

EUSTACE

- 3 Met-offices, 5 universities, 1 data storage facility, 2 spatio-temporal infilling methods
- 165 years of daily temperature observations from stations, ships, and satellites
- Multiscale stochastic weather and climate model based on SPDEs and finite element GMRFs
- Multiple iterative matrix solver techniques, exploiting the model structure
- Output:
  - Point estimates of daily mean, minimum, and maximum temperatures on a high resolution grid
  - Associated uncertainty estimates
  - Sample from the posterior distributions of the temperature fields
- Project stage: method software implementation in progress, results to be validated and released in 2018.

Gratuitous commercial

inlabru, the friendlier INLA interface. More from Fabian Bachl, COMP2, Wednesday

R-INLA, http://r-inla.org/

```
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 and non-linear prediction:
prediction <- result$summary.fitted.values[some.indices, "mean"]
prediction <- lapply(inla.posterior.sample(n=..., result),
                     function(x) cos(x$latent$field))
```

inlabru, 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)
```

Power tail quantile (POQ) model

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

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

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

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

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

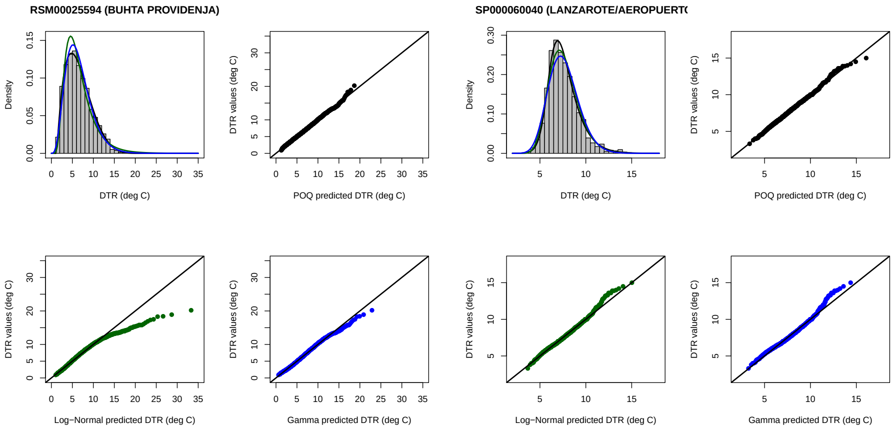
$$\widetilde{u}(\mathbf{s}, t) = F_{\boldsymbol{\theta}(\mathbf{s},t)}^{-1}(\Phi(u(\mathbf{s}, t))),$$

where the parameters can vary with space and time.

EUSTACE

Diurnal range distributions

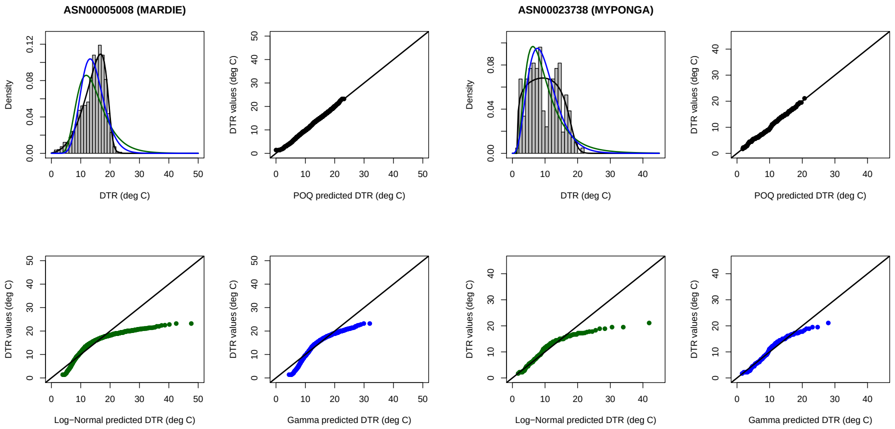
After seasonal compensation:



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.