Recent, current, and future issues for large scale space-time and nonlinear INLA

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

THE UNIVERSITY *of* EDINBURGH

Avignon 2018-11-08

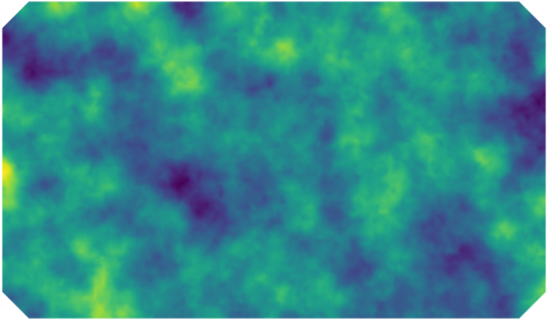
EUSTACE has received funding from the European Union's Horizon 2020 Programme for Research and Innovation, under Grant Agreement no 640171

EUSTACE

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.

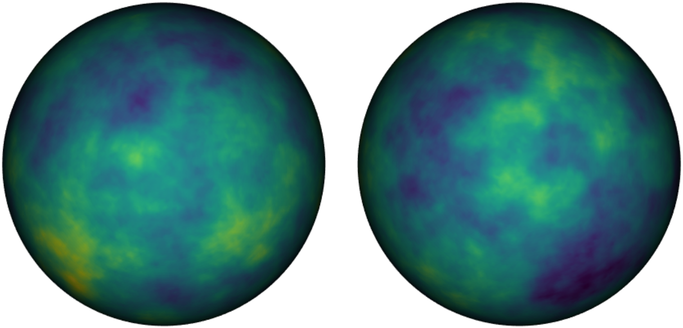
$$(\kappa^2 - \Delta)(\tau\, x(\mathbf{s})) = \mathcal{W}(\mathbf{s}), \quad \mathbf{s} \in \mathbb{R}^d$$



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.

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

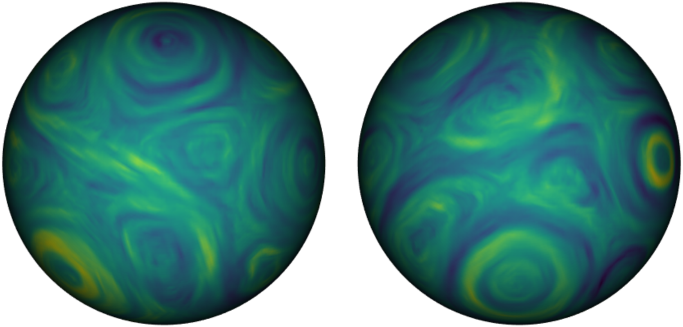


EUSTACE

GMRFs based on SPDEs (Lindgren et al., 2011)

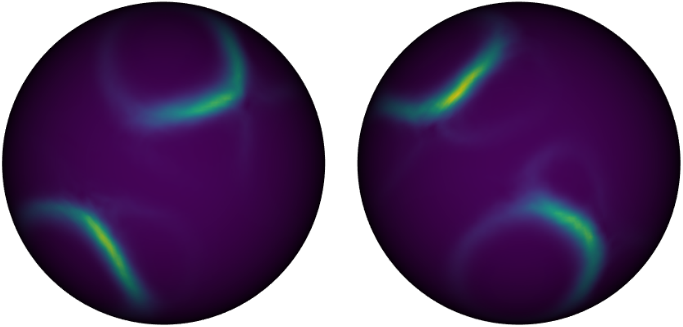
GMRF representations of SPDEs can be constructed for oscillating, anisotropic, non-stationary, non-separable spatio-temporal, and multivariate fields on manifolds.

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



Covariances for four reference points

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



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,

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

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 $\mathcal{W}$.

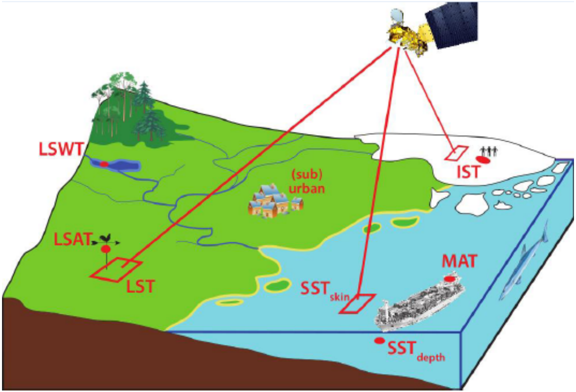
Matching the LHS and RHS distributions leads to the finite element approximation

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

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.



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

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

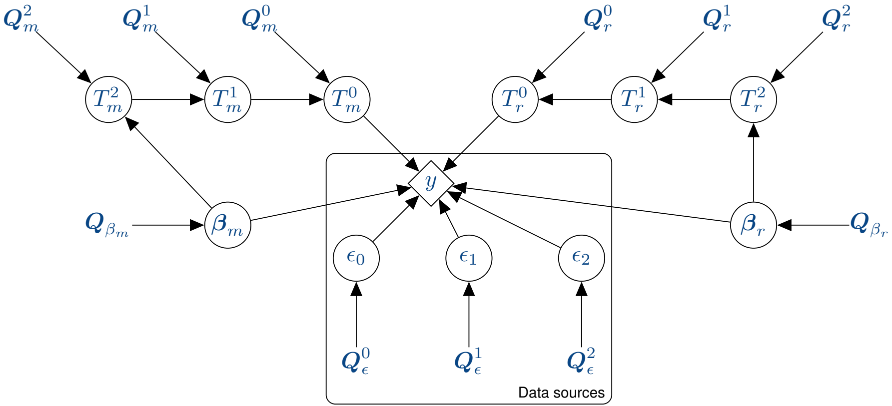
The finite element approximation has precision matrix structure

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

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

Partial hierarchical representation

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



Conditional specifications, e.g.

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

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

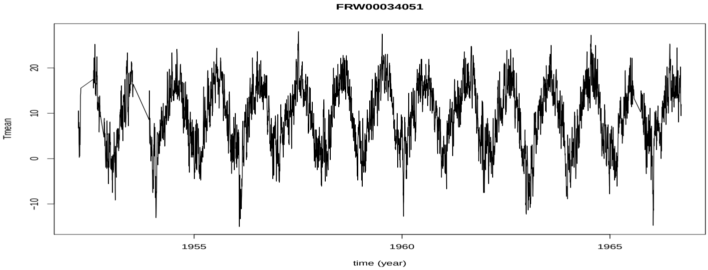
For station k at day t_i

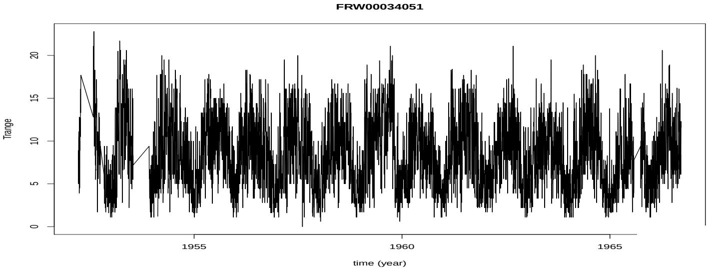
$$y_m^{k,i} = T_m(\mathbf{s}_k, t_i) + \sum_{j=1}^{J_k} H_j^k(t_i) e_m^{k,j} + \epsilon_m^{k,i},$$

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





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 a quantile blend of generalised 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*.

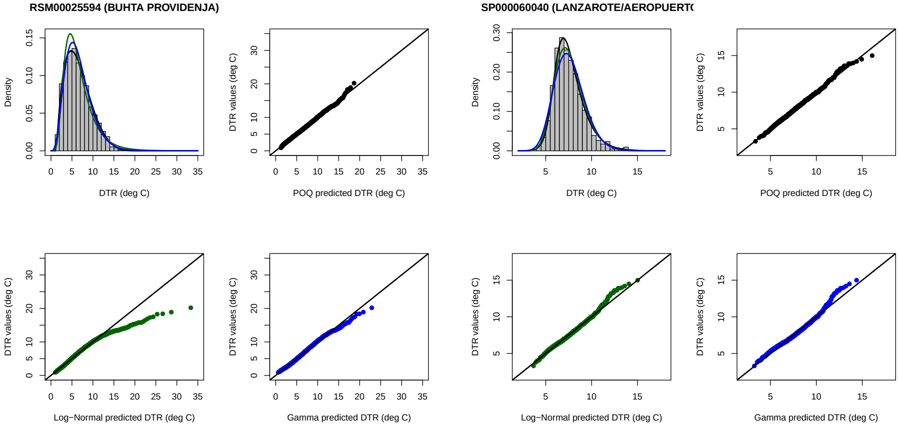
A spatio-temporally dependent Gaussian field $u(\mathbf{s},t)$ with expectation 0 and variance 1 can be transformed into a POQ field by

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

where the parameters can vary with space and time.

Diurnal range distributions

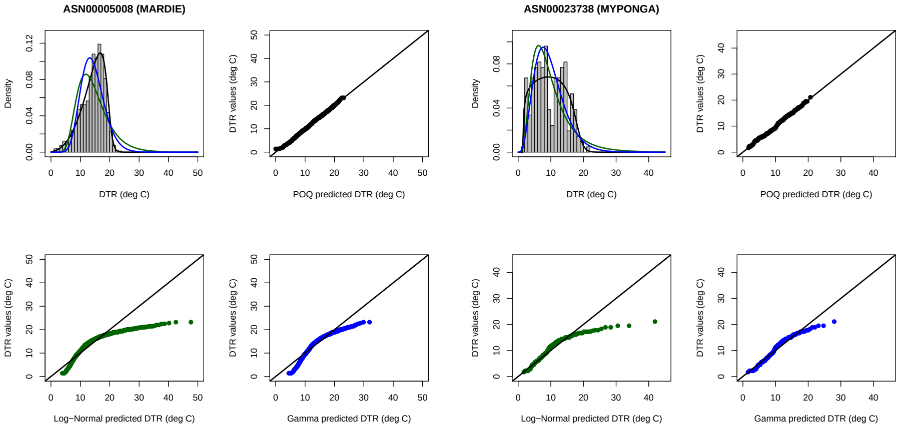
After seasonal compensation:



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

Diurnal range distributions; quantile model

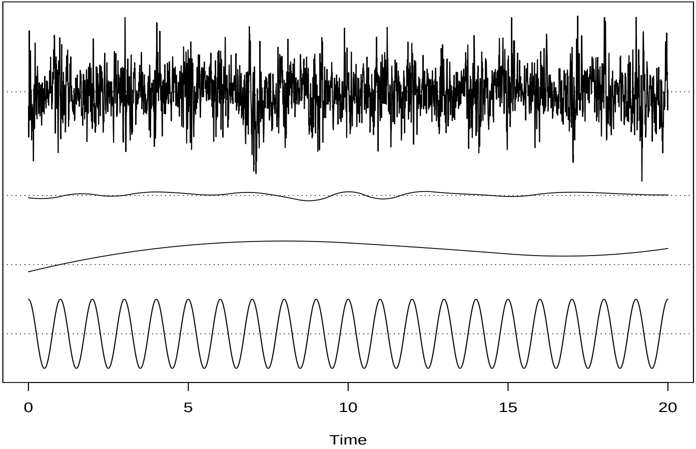
After seasonal compensation:



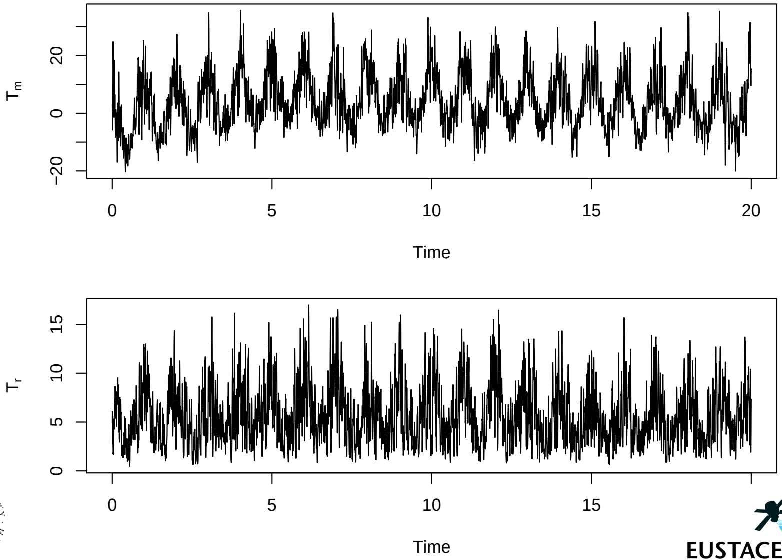
For these stations only POQ comes close to representing the distributions.

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

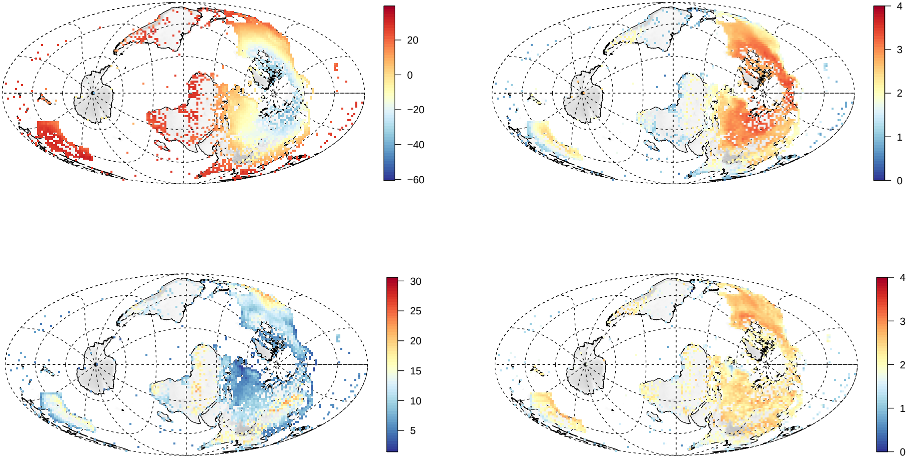
Multiscale model component samples



Combined model samples for T_m and T_r

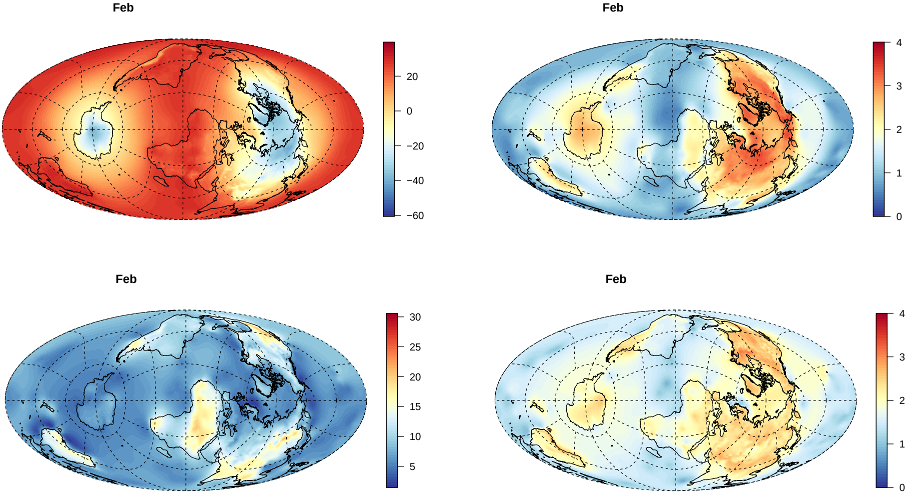


Median & scale for daily means and ranges



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 $\boldsymbol{Q}_x$:

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

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

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

Linear Gaussian observations

The conditional posterior distribution is

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

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

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

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 $\boldsymbol{Q}_x$:

$$(\boldsymbol{x} \mid \boldsymbol{\theta}) \sim \mathcal{N}(\boldsymbol{\mu}_x, \boldsymbol{Q}_x^{-1}) \qquad \text{(Prior)}$$
$$(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}) \sim \mathcal{N}(h(\boldsymbol{A}\boldsymbol{x}), \boldsymbol{Q}_{y|x}^{-1}) \qquad \text{(Observations)}$$
$$p(\boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{\theta}) \propto p(\boldsymbol{x} \mid \boldsymbol{\theta})\, p(\boldsymbol{y} \mid \boldsymbol{x}, \boldsymbol{\theta}) \qquad \text{(Conditional posterior)}$$

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:

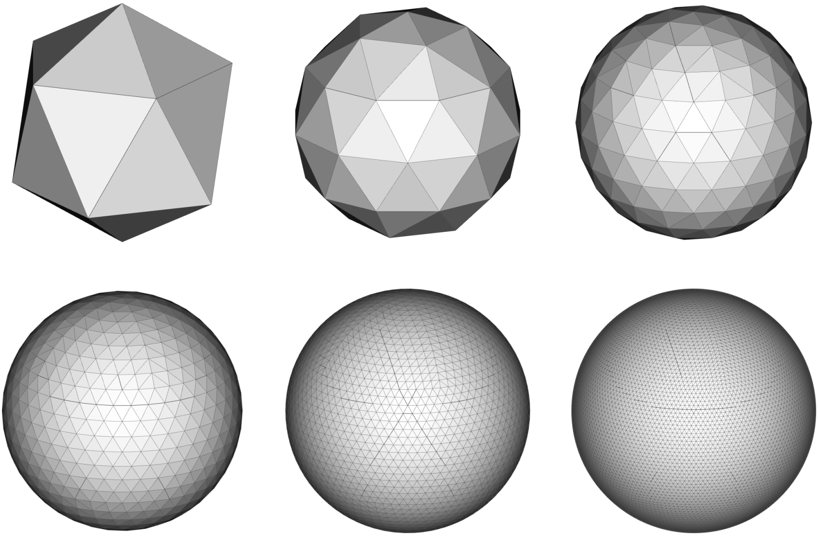
$$(\boldsymbol{x} \mid \boldsymbol{y}, \boldsymbol{\theta}) \overset{\text{approx}}{\sim} \mathcal{N}(\widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{Q}}^{-1}) \qquad \text{(Approximate posterior)}$$
$$\widetilde{\boldsymbol{Q}} = \boldsymbol{Q}_x + \boldsymbol{J}^\top \boldsymbol{Q}_{y|x} \boldsymbol{J}$$
$$\widetilde{\boldsymbol{\mu}}' = \widetilde{\boldsymbol{\mu}} + a\widetilde{\boldsymbol{Q}}^{-1} \left\{ \boldsymbol{J}^\top \boldsymbol{Q}_y \left[\boldsymbol{y} - h(\boldsymbol{A}\widetilde{\boldsymbol{\mu}})\right] - \boldsymbol{Q}_x(\widetilde{\boldsymbol{\mu}} - \boldsymbol{\mu}_x) \right\}$$

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

Problem: $\sim 10^{11}$ latent variables

Solution: Iterative solvers

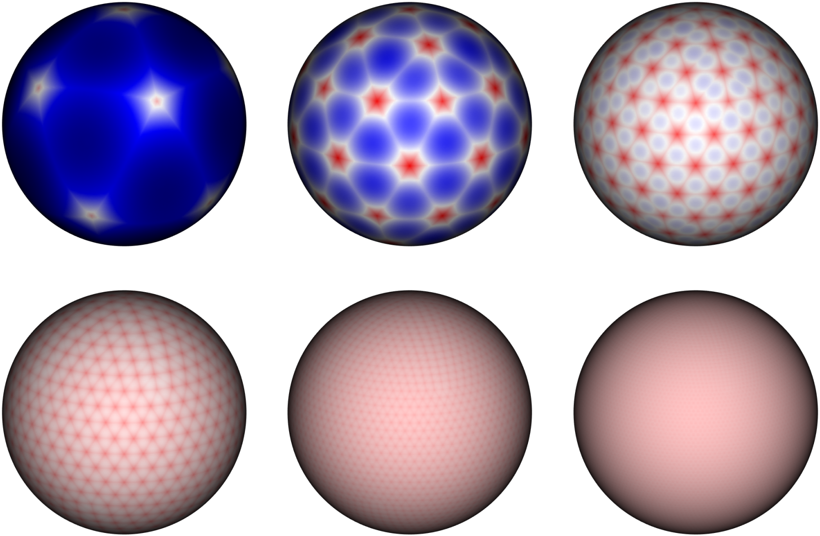
Triangulations for all corners of Earth



THE UNIVERSITY OF EDINBURGH

EUSTACE

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

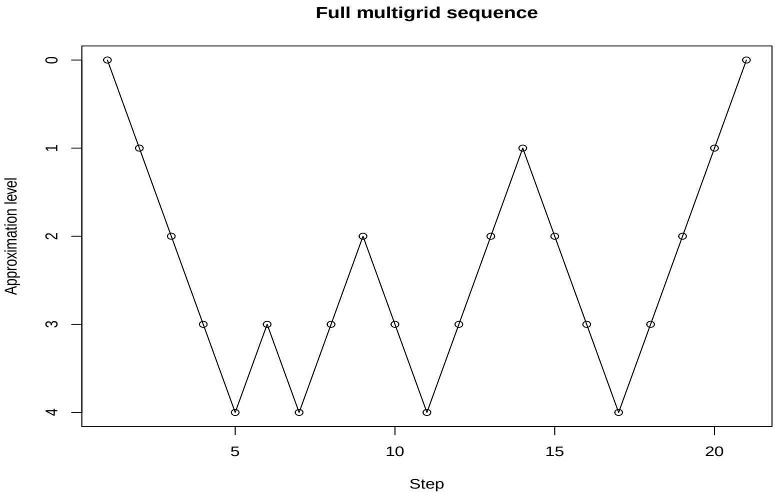
$$\boldsymbol{M}^{-1}\boldsymbol{x} = \sum_{k=1}^{K} \boldsymbol{W}_k \boldsymbol{B}_k (\boldsymbol{B}_k^\top \boldsymbol{Q} \boldsymbol{B}_k)^{-1} \boldsymbol{B}_k^\top \boldsymbol{W}_k \boldsymbol{x}$$

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 hierarchy of scales and preconditioning ($\boldsymbol{x}_0 = \boldsymbol{B}\boldsymbol{x}_1 +$ fine scale variability):

Multiscale Schur complement approximation

Solving $\boldsymbol{Q}_{\boldsymbol{x}|\boldsymbol{y}}\boldsymbol{x} = \boldsymbol{b}$ can be formulated using two solves with the upper (fine) block $\boldsymbol{Q}_0 + \boldsymbol{A}^\top\boldsymbol{Q}_\epsilon\boldsymbol{A}$, and one solve with the *Schur complement*

$$\boldsymbol{Q}_1 + \boldsymbol{B}^\top\boldsymbol{Q}_0\boldsymbol{B} - \boldsymbol{B}^\top\boldsymbol{Q}_0\left(\boldsymbol{Q}_0 + \boldsymbol{A}^\top\boldsymbol{Q}_\epsilon\boldsymbol{A}\right)^{-1}\boldsymbol{Q}_0$$

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

$$\begin{bmatrix}\widetilde{\boldsymbol{Q}}_B + \boldsymbol{B}^\top\boldsymbol{A}^\top\boldsymbol{Q}_\epsilon\boldsymbol{A}\boldsymbol{B} & -\widetilde{\boldsymbol{Q}}_B \\ -\widetilde{\boldsymbol{Q}}_B & \boldsymbol{Q}_1 + \widetilde{\boldsymbol{Q}}_B\end{bmatrix}\begin{bmatrix}\text{ignored} \\ \boldsymbol{x}_1\end{bmatrix} = \begin{bmatrix}\boldsymbol{0} \\ \widetilde{\boldsymbol{b}}\end{bmatrix}$$

where $\widetilde{\boldsymbol{Q}}_B = \boldsymbol{B}^\top\boldsymbol{Q}_0\boldsymbol{B}$.
The block matrix can be interpreted as the precision of a bivariate field on a common, coarse spatio-temporal scale, and the same technique applied to this system, with $\boldsymbol{x}_{1,1} = \boldsymbol{B}_{1|2}\boldsymbol{x}_{1,2} +$ finer scale variability.

Also applies to the station data bias homogenisation coefficients.

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 $\Omega_k \subset \Omega$,

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

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

Efficient if $\boldsymbol{a}\boldsymbol{a}^\top$ sparsity matches $\boldsymbol{S}$ for each subdomain.
Sidén et al (2018, JCGS): Iterated blockwise Takahashi

EUSTACE

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
- Within each scale, approximate multigrid solves (not implemented)
- Overlapping space-time domain decomposition within each multigrid level
- Direct Monte Carlo sampling: add suitable randomness to the RHS of the $Q_{x|y}$ solves for $\widetilde{\mu}$.
- Rao-Blackwellised variance estimation

Parameter estimation:

In the project, several ad hoc methods are used;

Timeseries subsets used for diurnal range distributions and temporal correlation parameters.

Local estimation of spatial dependence paramters blended into a full spacetime SPDE model

EUSTACE

inlabru, a 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)
```

inlabru features (with paraphrased code)

- Automated model structure mapping, sp spatial objects

```
components <-
  myeffect(map = covariate(x, y), ...) +
  field(map = coordinates, model = spde)
```

No user-side `inla.stack` or `inla.spde.make.A` calls needed!

- Nonlinear predictors, advanced likelihood wrappers, iterated INLA calls

```
formula <- y ~ myeffect + exp(field)
formula <- coordinates ~ myeffect + exp(field)
fit <- bru(components, like(family = "cp", formula = formula, ...), ...)
```

- Posterior sampling and prediction of (nearly) arbitrary expressions

```
# Sample and compute nonlinear functionals:
generate(fit, formula = ~ exp(myeffect + exp(field)), ...)
# Sample and compute posterior summaries,
# here a data level posterior probability function:
predict(
  fit,
  formula = ~ dpois(0:200, sum(weight * exp(myeffect + exp(field)))),
  ...)
```

Optimisation algorithms

Best estimate, with Newton iteration

1. Pick a linearisation point $\boldsymbol{x}_0$
2. Setup RHS from the gradient of the minimisation problem
3. Find search direction $\boldsymbol{d}_k$ with linear solve
4. Find a new linearisation point $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k$
5. Repeat from 2

Linear solve, with PCG iteration

The linear system for the best estimate and for samples have structure $\boldsymbol{Q}\boldsymbol{x} = \boldsymbol{b}$.

1. Start at some $\boldsymbol{x}_0$
2. Compute residual $\boldsymbol{r}_k = \boldsymbol{b} - \boldsymbol{Q}\boldsymbol{x}_k$
3. Apply preconditioner: $\boldsymbol{d}_k = \boldsymbol{M}^{-1}\boldsymbol{r}_k$
4. Scale and rotate $\boldsymbol{d}_k$ to conjugate search direction $\widetilde{\boldsymbol{d}}_k$
5. Find new point $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \widetilde{\boldsymbol{d}}_k$
6. Repeat from 2

THE UNIVERSITY OF EDINBURGH

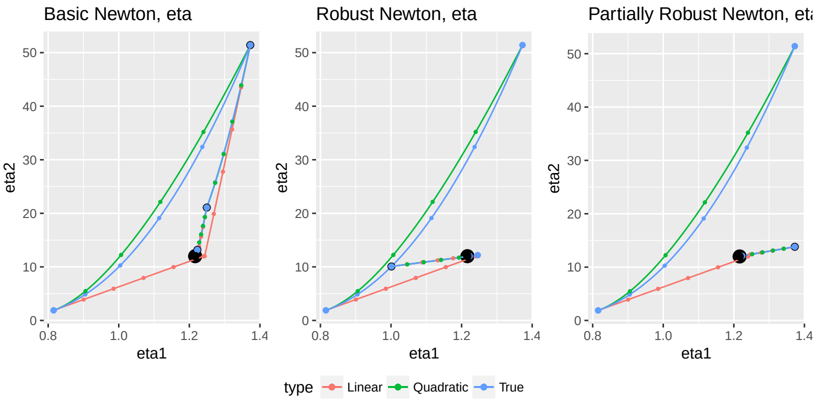
EUSTACE

A toy problem for the diurnal range model

The choice of linearisation point is important.

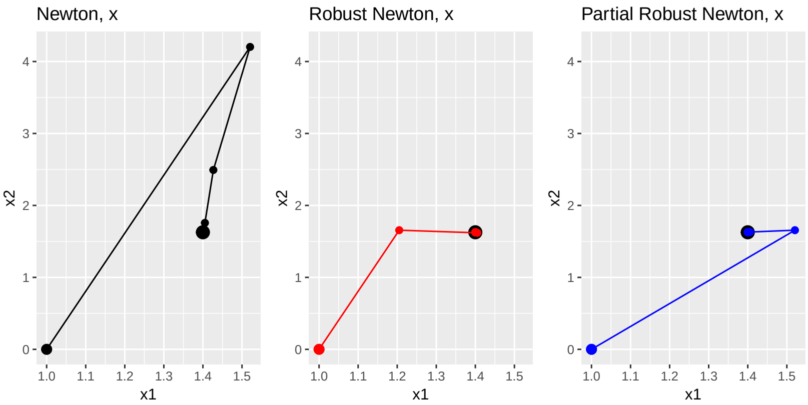
- The linearised model is only close to the full model for short distances
- Large scale component x_1, daily component x_2
- $\eta_1 = \exp(x_1) \cdot 0.3$, $\eta_2 = \exp(x_1) G^{-1}(x_2)$
- Linearise at some $\boldsymbol{x}_0$
- Plain Newton may overshoot the target
- With a simple line search minimising the distance between the current η-estimate and the next linearisation point the method is robustified
- Only the daily component is local, and is the least defined; partially robustify blockwise
- Robustify for each overlapping spatiotemporal block separately, and do blended weighting

Tracing the model predictor



- The plain Newton method overshoots the target, ending up in an extreme place.
- Shorter steps in all or some components accelerates convergence

Tracing the model components



- The plain Newton method overshoots the target, ending up in an extreme place.
- Shorter steps in all or some components accelerates convergence

Partly solved and unsolved problems

- High order operator preconditioners for iterative "matrix free" solvers; overlapping block calculations.
- Multivariate likelihoods, e.g. Dirichlet:
  Current work with Joaquín Martínez Minaya to convert Dirichlet likelihoods to independent Gaussians to "fool" INLA. Postprocess for full Laplace approximation.

Quadratisation of Dirichlet likelihoods

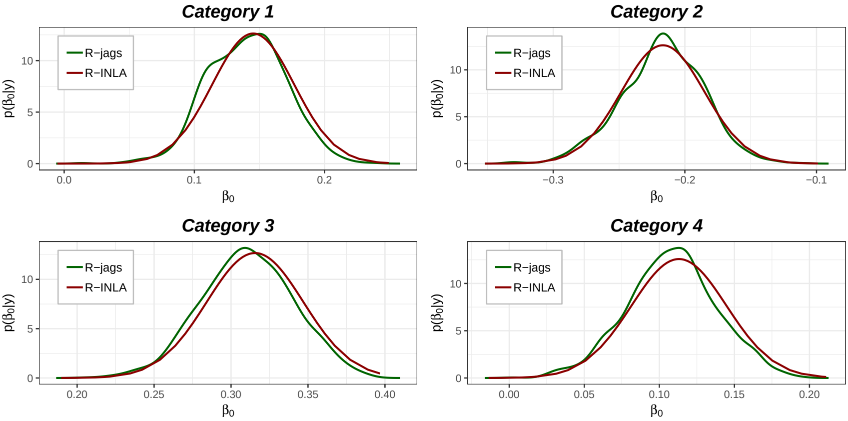
- Multivariate likelihood for $y_c \in (0, 1)$, $\sum_{c=1}^{C} y_c = 1$:

$$(y_1, \ldots, y_C | \boldsymbol{\alpha}) \sim \text{Dirichlet}(\alpha_1, \ldots, \alpha_C), \quad 0 < \alpha_c,$$

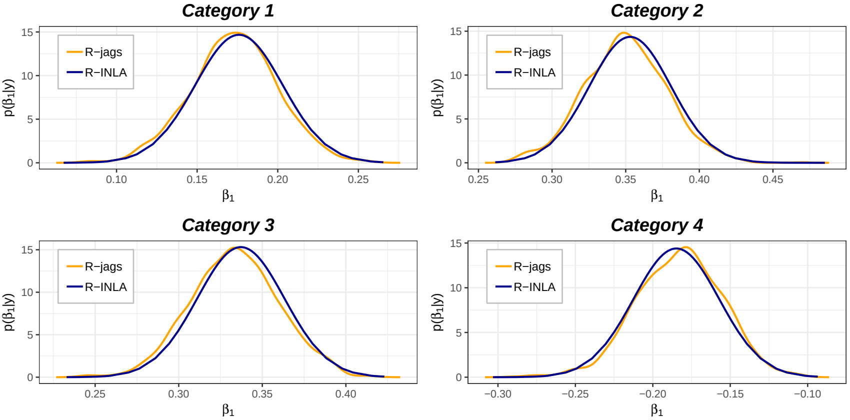
$$p(y_1, \ldots, y_C | \boldsymbol{\alpha}) = \frac{1}{B(\boldsymbol{\alpha})} \prod_{c=1}^{C} y_c^{\alpha_c - 1}$$

- Example model: $\alpha_{ic} = \beta_{c0} + X_{ic}\beta_{c1}$
- R-INLA restriction: $y_i | \boldsymbol{\eta}$ conditionally independent and only dependent on η_i.
- Iterative approximation solution:
  1. Construct a multivariate Gaussian approximation to the likelihood
  2. Rotate & scale the observations $\longrightarrow$ Conditionally independent $\mathsf{N}(0, 1)$ pseudo-observations.
  3. Run INLA, and repeat the approximation at the new estimate

Iterated-INLA and JAGS posterior densities: intercepts



Iterated-INLA and JAGS posterior densities: slopes



Partly solved and unsolved problems

- High order operator preconditioners for iterative "matrix free" solvers; overlapping block calculations.
- Multivariate likelihoods, e.g. Dirichlet:
  Current work with Joaquín Martínez Minaya to convert Dirichlet likelihoods to independent Gaussians to "fool" INLA. Postprocess for full Laplace approximation.
- Explicit stochastic boundary precision construction (ongoing work with Daniel Simpson and David Bolin)
- Blending local stationary SPDE parameter estimates to globally non-stationary models
- Making manifold SPDEs more accessible; brains and other internal organs!
- Software testing; INLA has no automated test suite!
- Generalising `inlabru`: New backend code to allow easier extensions, and complete INLA support;
  improve programmability of INLA features without breaking existing code
- Documentation! Progressing, but slowly. Can now use `roxygen2` also in INLA.
- "Large INLA"; a hypothetical new INLA implementation aimed at laaaaaarge models! Missing piece: fast and accurate log-determinants

EUSTACE

Practical preconditioner (extra slide)

Minimise data reading:

- Sweep through time for each spatial subregion
- For each temporal scale at the same time; accumulate information until a macro space-time block is constructed
- Approximate solve for a macro space-time block
- Homogenisation biases form their own blocks

Notes:

- Nested Schur complement alternative:
  Experiments show that approximate nested Schur complements (Fine-Coarse-Fine) leads to a robust and efficient preconditioner, but requires re-reading all the data multiple times (2 fine scale steps)
- Separate preconditioning for each scale is faster per iteration (1 fine scale step):
  Not having to re-read data multiple times is faster; offsets the need for more iterations
- Tradeoff unknown

Spatial fields, observations, and stochastic models

- Partially observed spatial functions 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 $\boldsymbol{\theta} = \{\theta_j, j = 1, \ldots, n_\theta\}$