Pragmatically ambitious multiscale global temperature reconstruction

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

with Colin Morice, John Kennedy, Christopher Merchant, and the EUSTACE team

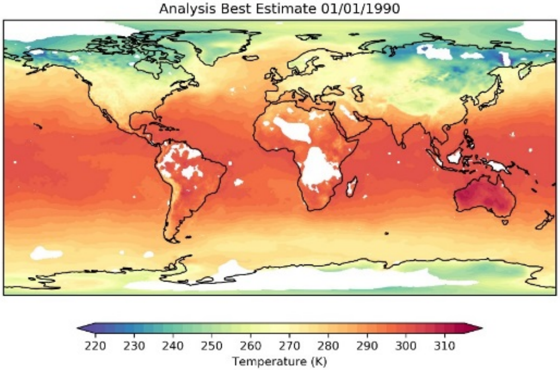
THE UNIVERSITY of EDINBURGH

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

EUSTACE ANALYSIS

Combines in-situ and satellite data sources to derive daily air temperatures across the globe with quantified uncertainties.

- Daily mean air temperature (2 m) estimates from the mid-late 19th century at ¼ degree resolution.
- Observational dataset for use in climate monitoring, services and research.
  - Quantify bias and uncertainty arising from observational sampling (in space and time);
  - Quantify uncertainty from instrumental effects/network changes.
- Higher resolution daily gridded analyses for regional climate
  - Combine in situ and remote sensing data to support high resolution analysis.
  - Absolute temperature rather than anomaly product.



Met Office Hadley Centre

EUSTACE

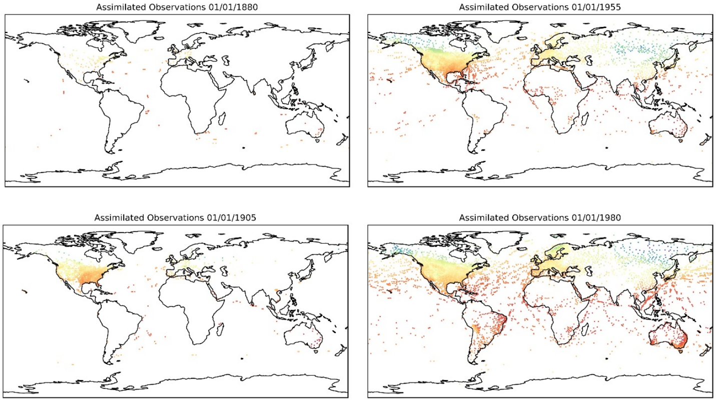
OBSERVATIONS

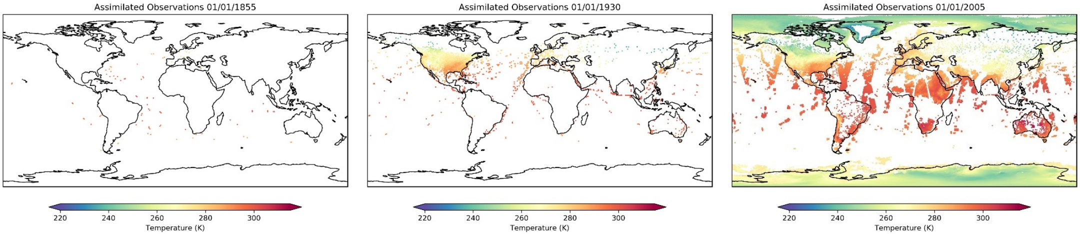
In situ air temperature:

- EUSTACE station dataset (UBERN) (GHCN-D, ECA&D, ISTI, DECADE, ERA-CLIM)
- HadNMAT-2 ship air temperatures (NOCS/Met Office)

Satellite skin temperature derived air temperature:

- Marine: ATSR (ESA CCI SST)
- Land: MODIS (USGS/NASA via ESA GlobTemperature)
- Ice: AVHRR (NOAA/FP7 NACLIM)





Met Office
Hadley Centre

Statistical model and method building blocks

Basic system components

- Temperature *processes on different spatial and temporal scales*
  - Seasonal
  - Slow climate processes
  - Medium-scale variability
  - Daily
- *Vast model size* ($\sim 10^{11}$ unknowns); need computationally efficient tools
- Hierarchical statistical model structure based on Gaussian processes
  - Stochastic PDEs translates to sparse precisions in *Gaussian Markov random fields*
- *Propagated uncertainty* via a Bayesian approach
  - Dependence structure parameters
  - Spatio-temporal process priors
  - Observation models; Multiple *observation sources*, with complex error *uncertainty structure*
- Goals:
  - a *best estimate*,
  - a *collection of samples*, and
  - more precise (and accurate) *uncertainty estimates*.

THE UNIVERSITY OF EDINBURGH

EUSTACE

Matérn driven heat equation on the sphere

The iterated heat equation is a simple non-separable space-time SPDE family:

$$(\kappa^2 - \Delta)^{\gamma/2} \left[\phi \frac{\partial}{\partial t} + (\kappa^2 - \Delta)^{\alpha/2} \right]^{\beta} x(\mathbf{s}, t) = \mathcal{W}(\mathbf{s}, t)/\tau$$

For constant parameters, $x(\mathbf{s}, t)$ has spatial Matérn covariance (for each t).

Discrete domain Gaussian Markov random fields (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 $j \neq \mathcal{N}_i \cup i$.

- Project the SPDE solution space onto local basis functions: random Markov dependent basis weights (Lindgren et al, 2011).

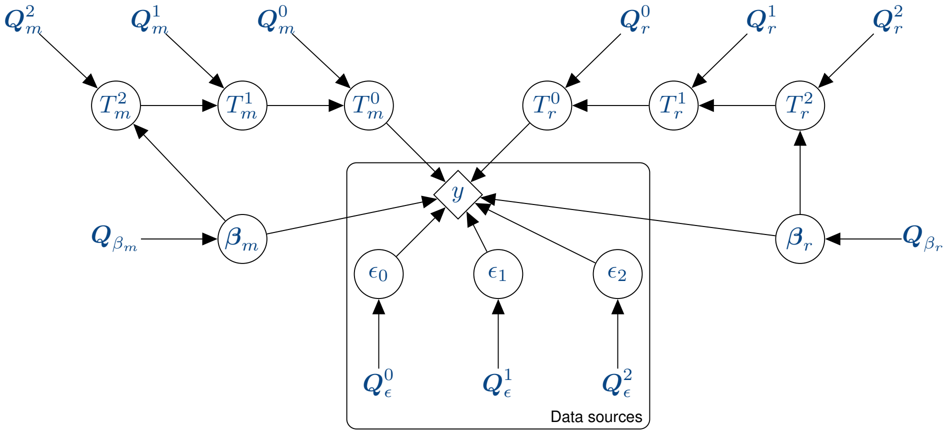
A finite element approximation has structure

$$x(\boldsymbol{s}, t) = \sum_{i,j} \psi_i^{[s]}(\boldsymbol{s}) \psi_j^{[t]}(t) x_{ij}, \quad \boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}^{-1}), \quad \boldsymbol{Q} = \sum_{k=0}^{\alpha+\beta+\gamma} \boldsymbol{M}_k^{[t]} \otimes \boldsymbol{M}_k^{[\mathbf{s}]}$$

even, e.g., if the spatial scale parameter κ is spatially varying.

Partial hierarchical representation

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



Conditional specifications, e.g.

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

$$T_r^0 = \exp(T_r^1)\; G^{-1}\left[U_r^0(\mathbf{s}, t)\right], \quad U_r^0 \sim \mathcal{N}\left(\mathbf{0}, {\boldsymbol{Q}_r^0}^{-1}\right)$$

THE UNIVERSITY OF EDINBURGH

EUSTACE

Standardised observation uncertainty models

- Each data source may have complicated dependence structure
- To facilitate information blending, use a common error term structure

Common satellite derived data error model framework

The observational&calibration errors are modelled as three error components:

- independent (ϵ_0),
- spatially and/or temporally correlated (ϵ_1), and
- systematic (ϵ_2),

with distributions determined by the uncertainty information from satellite calibration models.
E.g., $y_i = T_m(\mathbf{s}_i, t_i) + \epsilon_0(\mathbf{s}_i, t_i) + \epsilon_1(\mathbf{s}_i, t_i) + \epsilon_2(\mathbf{s}_i, t_i)$

In practice, each data source might have several different components of each type; independent components can be merged, but not necessarily correlated or systematic components.

THE UNIVERSITY OF EDINBURGH

EUSTACE

Station observation&homogenisation model

Daily means

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.

Daily mean/max/min

For station k at day t_i,

$$y_m^{k,i} = T_m(\mathbf{s}_k, t_i) + \widetilde{H}_m^k(t_i) + \epsilon_m^{k,i},$$

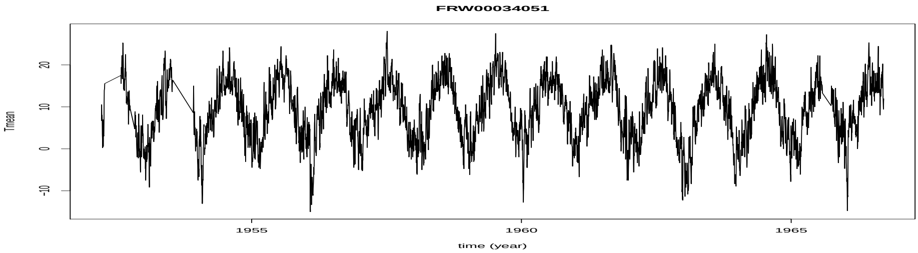
$$y_x^{k,i} = T_m(\mathbf{s}_k, t_i) + \frac{\exp[\widetilde{H}_r^k(t_i)]}{2} T_r(\mathbf{s}_k, t_i) + \epsilon_x^{k,i},$$

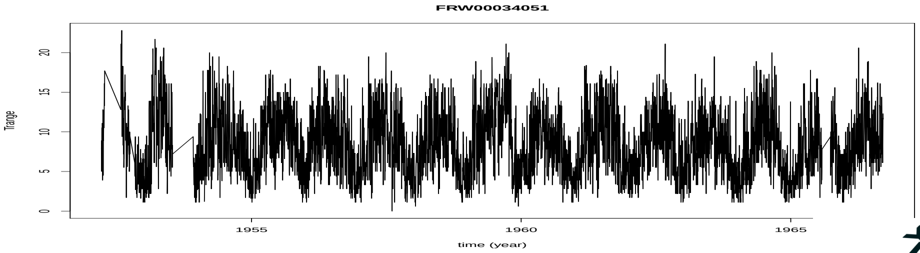
$$y_n^{k,i} = T_m(\mathbf{s}_k, t_i) - \frac{\exp[\widetilde{H}_r^k(t_i)]}{2} T_r(\mathbf{s}_k, t_i) + \epsilon_n^{k,i},$$

where $\widetilde{H}_\cdot^\cdot$ are the total bias correction variables for each observation.

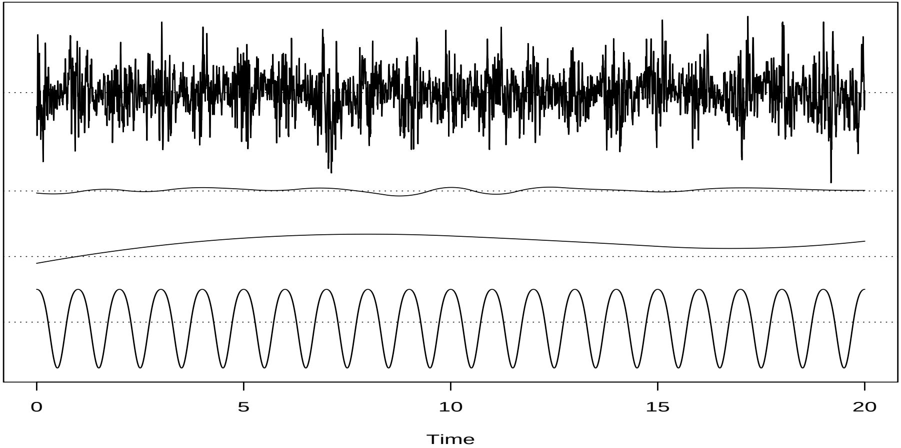
Observed data

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



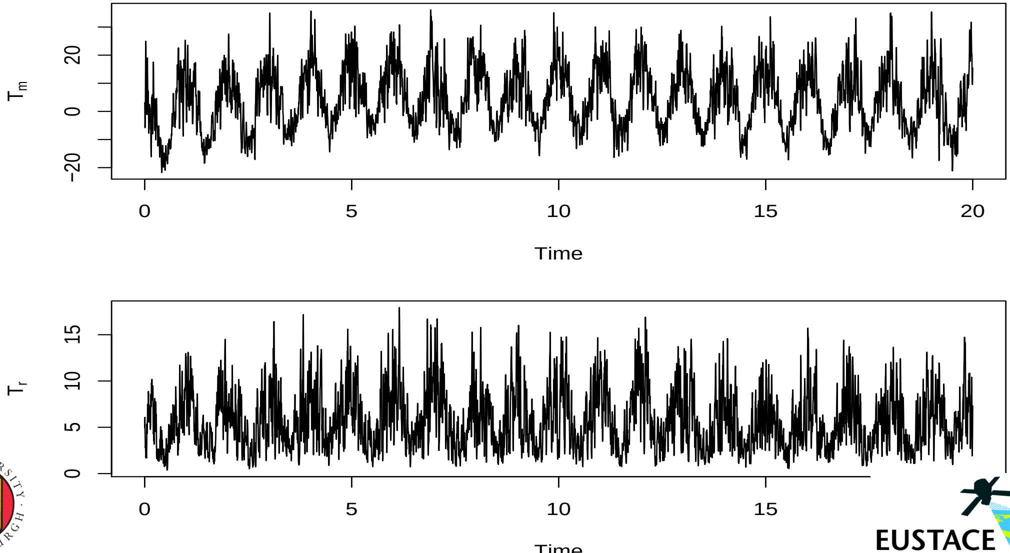


Multiscale model component samples



Combined model samples for T_m and T_r

(Proof of concept; no actual data was involved in this figure)



Modelling non-Gaussian quantities

Power tail quantile (POQ) model

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

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

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

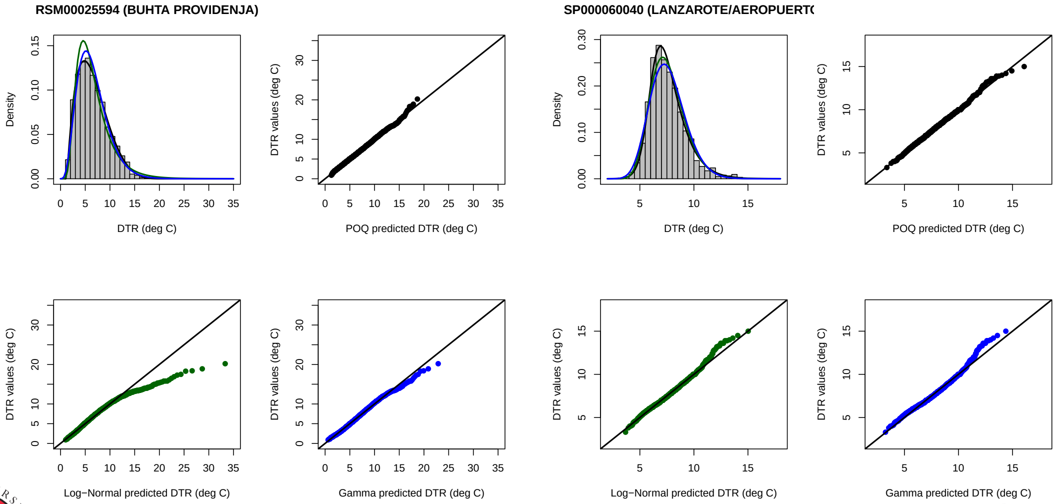
$$F_{\boldsymbol{\theta}}^{-1}(p) = \theta_0 + \frac{\tau}{2}\left[(1-\gamma)f_{\theta_3}^-(p) + (1+\gamma)f_{\theta_4}^+(p)\right].$$

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* (Gilchrist, 2000).

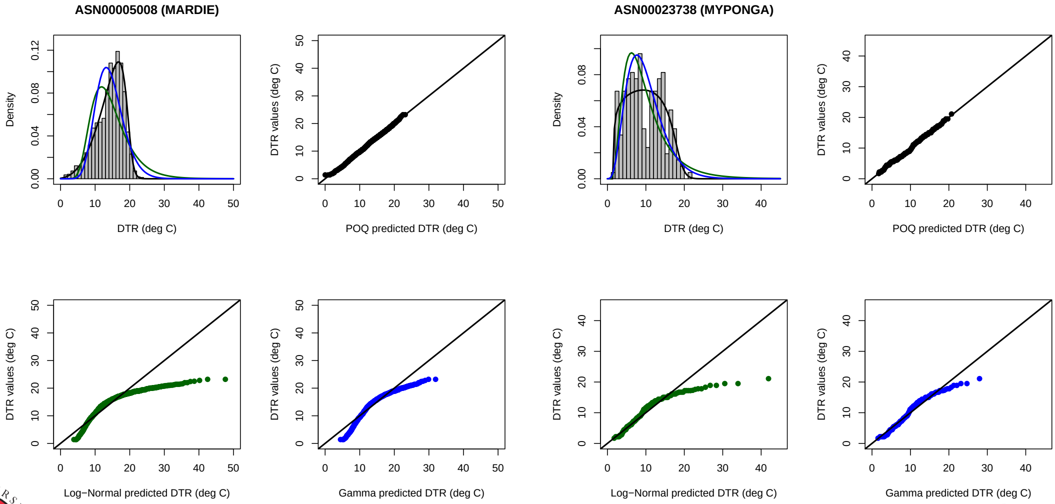
Copula transformation: $G^{-1}[u(\mathbf{s},t)] = F_{\boldsymbol{\theta}(\mathbf{s},t)}^{-1}\{\Phi[u(\mathbf{s},t)]\}$

Diurnal range distributions



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

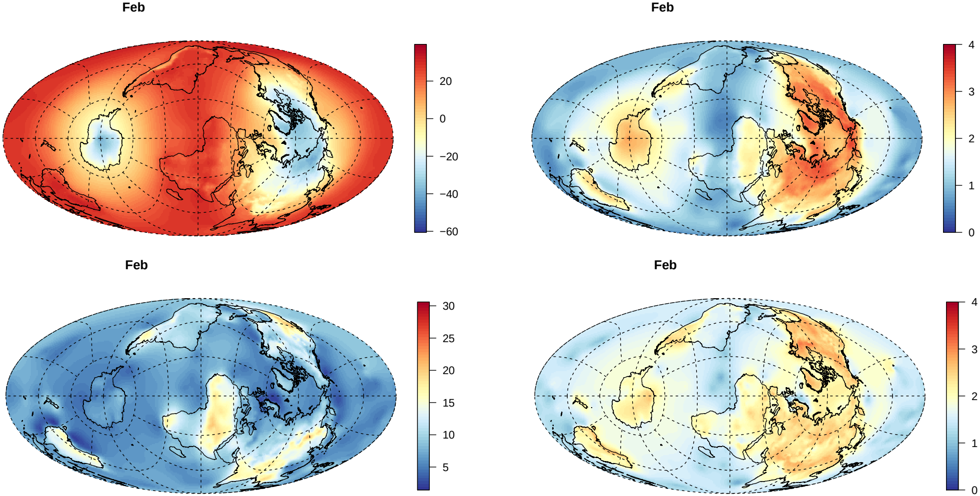
Diurnal range distributions



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

Note: Some shapes may be due to unmodeled station inhomogeneities.

Estimates of median & scale for T_m and T_r



February climatology

(Preliminary estimates, using only in-situ land station data)

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:

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

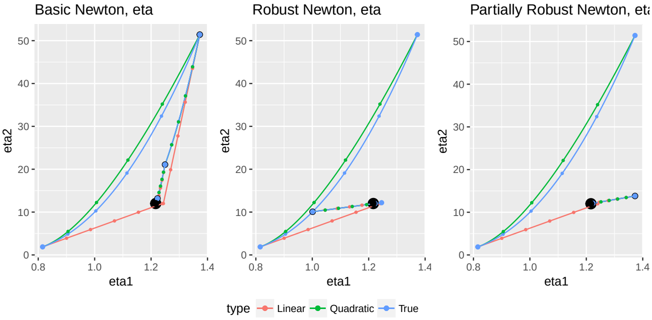
$$\widetilde{\boldsymbol{Q}} = \boldsymbol{Q}_x + \boldsymbol{J}^\top \boldsymbol{Q}_{y|x} \boldsymbol{J}$$

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

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

Iterative solutions for $\sim 10^{11}$ latent variables

- Nonlinear Newton iteration with robust line-search

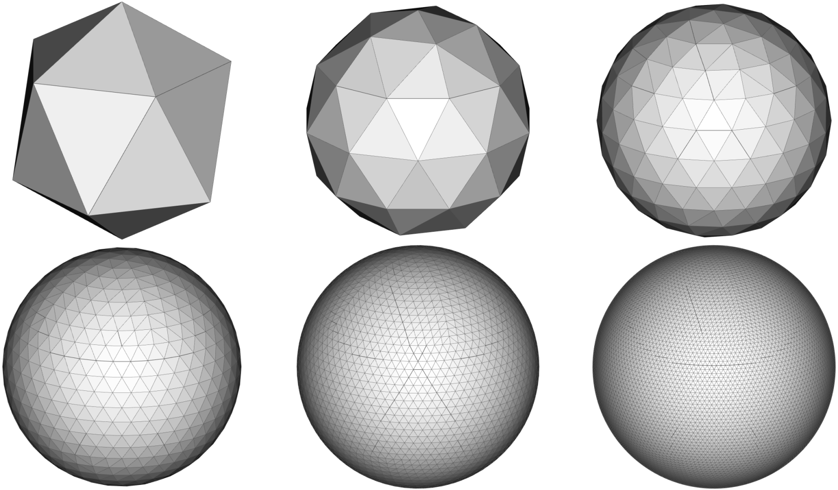


- Preconditioned conjugate gradient (PCG) iteration for
  $\boldsymbol{Q}(\boldsymbol{\mu} - \widehat{\boldsymbol{\mu}}) = \boldsymbol{r} = \boldsymbol{b} - \boldsymbol{Q}\widehat{\boldsymbol{\mu}}$
- Local and multiscale approximations for preconditioning: $\boldsymbol{M}^{-1}\boldsymbol{Q} \approx \boldsymbol{I}$
- Sampling with PCG: $\boldsymbol{Q}(\boldsymbol{x} - \widehat{\boldsymbol{\mu}}) = \boldsymbol{L}\boldsymbol{w}$
  Requires only a rectangular pseudo-Cholesky factorisation $\boldsymbol{L}\boldsymbol{L}^\top = \boldsymbol{Q}$.
  Possible due to the kronecker product sum precision structure.

THE UNIVERSITY OF EDINBURGH

EUSTACE

Triangulations for all corners of Earth



Variance calculations

Sparse partial inverse: Takahashi recursions postprocesses Cholesky

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.

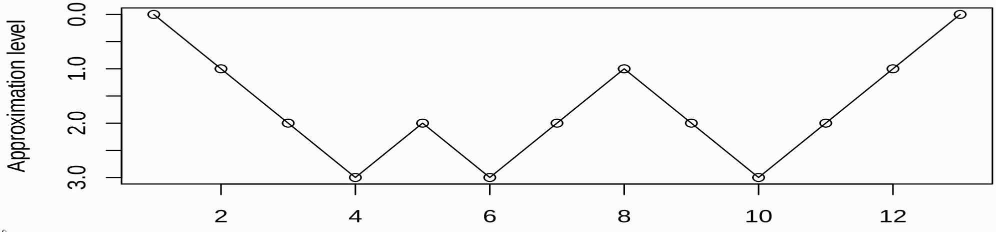
Overlapping blocks and multigrid

Overlapping block preconditioning

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

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

Multigrid and/or approximate multiscale Schur complements



Data reading issue: Schur complements vs conditional block updating

MULTI-SCALE ANALYSIS MODEL

Statistical model for temperature variations and different scales (space and time):

- **Climatological variation**: local seasonal cycle with effects of latitude, altitude and coastal influence.
- **Large-scale variation**: Slowly varying climatological mean temperature field. Station homogenisation.
- **Daily Local**: daily variability associated with weather. Satellite retrieval biases.

Simultaneously estimates observational biases of known bias structures:

- e.g. satellite biases, station homogenisation.

Processed on STFC's LOTUS cluster www.jasmin.ac.uk:

- Largest solves processed on 20 core/256GB RAM node.
- Highly parallel observation pre-processing.

Element	Resolution	N Variables
Seasonal	Bimonthly x 1° SPDE	245,772
Slow-scale*	5 year x 5° SPDE	107,604
Latitude	0.5° latitude SPDE	721
Altitude	(0.25° grid)	1
Coastal	(0.25° grid)	1
Grand mean	Analysis mean	1

Element	Resolution	N Variables
Large-scale	3 monthly x 5° SPDE	1,752,408
Station bias	NA	82,072

Element	Resolution	N Variables per day
Daily local	~0.5 degree SPDE	162,842
Satellite bias (marine)	Global	1
Satellite bias (land)	Global + 2.5 degree SPDE	1 + 40,962
Satellite bias (ice)	Hemispheric + 2.5 degree SPDE*	2 + 40,962

Met Office
Hadley Centre

EUSTACE

ITERATIVE SOLUTION

Condition on expected value of other components

(1) Climatology:
Global mean T
Seasonal cycle
Latitude
Altitude
Water fraction
$\boldsymbol{T}^{clim}(\boldsymbol{s},t)$

(2) Large-scale:
Multi-annual SPDE
Multi-month SPDE
Station biases
$\boldsymbol{T}^{large}(\boldsymbol{s},t)$

(3) Daily local:
Daily fields
Satellite land bias
Satellite marine bias
Satellite ice bias
$\boldsymbol{T}^{local}(\boldsymbol{s},t)$

Model Factorisation

The model is factorised into $m = 1, ..., 3$ components that are estimated interatively, substituting $\tilde{\boldsymbol{y}}_m$ for $\boldsymbol{y}$:

$$\tilde{\boldsymbol{y}}_m = \boldsymbol{y} - \sum_{n \neq m} \boldsymbol{J}_n \boldsymbol{\mu}_{\boldsymbol{x}_n|\tilde{\boldsymbol{y}}_n}$$

Met Office
Hadley Centre

EUSTACE

MULTI-SCALE ANALYSIS MODEL

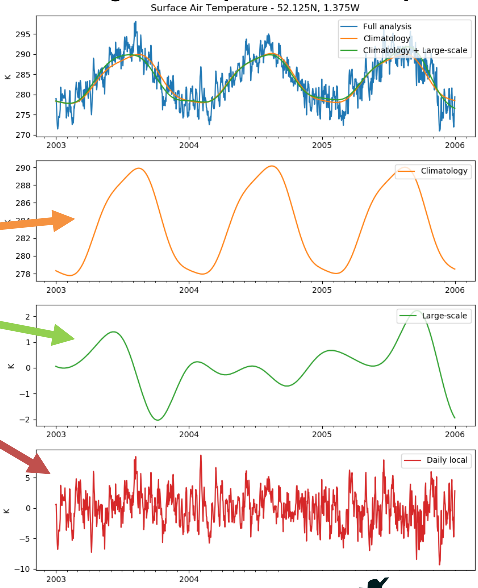
Statistical model for temperature variations and different scales (space and time):

- **Climatological variation**: local seasonal cycle with effects of latitude, altitude and coastal influence.
- **Large-scale variation**: Slowly varying climatological mean temperature field.
- **Daily Local**: daily variability associated with weather.

Simultaneously estimates observational biases of known bias structures:

- e.g. satellite biases, station homogenisation.

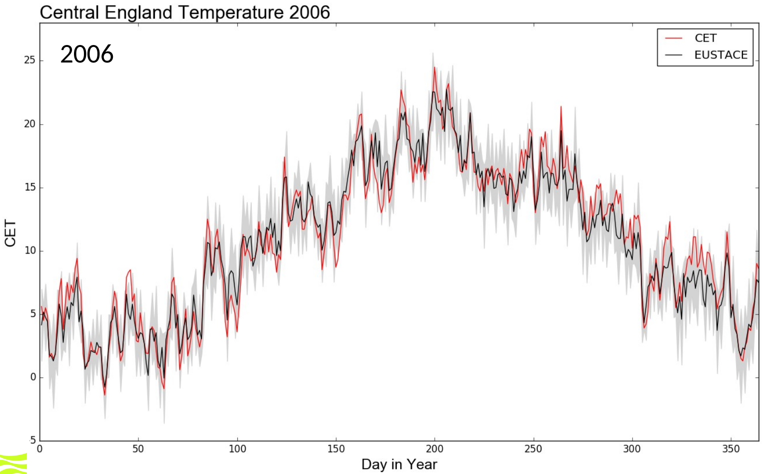
Central England Temperature Decomposition

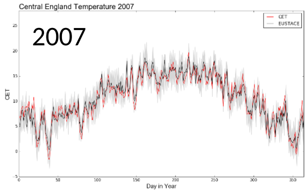


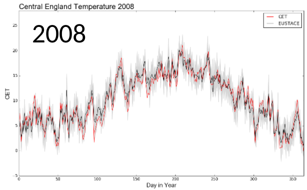
Met Office Hadley Centre

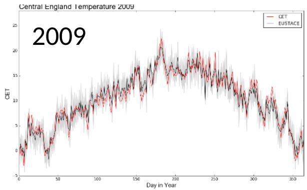
EUSTACE

COMPARING EUSTACE WITH CENTRAL ENGLAND TEMPERATURE









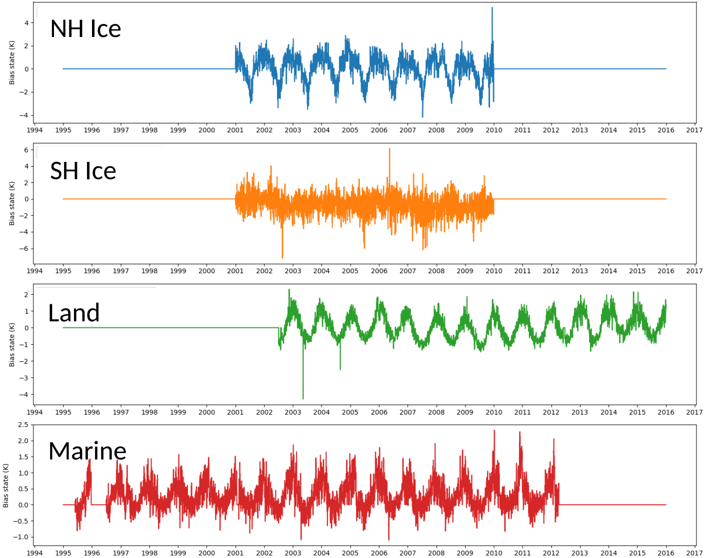
Met Office
Hadley Centre

EUSTACE

SATELLITE BIAS MODELS

- Simplified model of known error structures in satellite air temperature retrievals:
  - Global/hemispheric systematic bias covariates.
  - Daily estimates of spatially varying bias as a spatial random field.

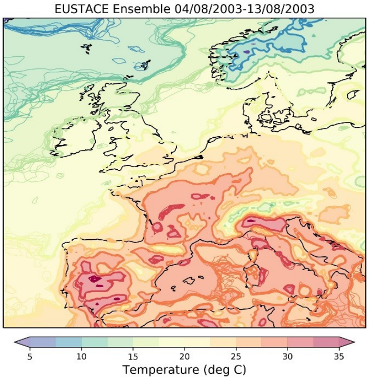
- Estimated jointly with daily temperature variability.

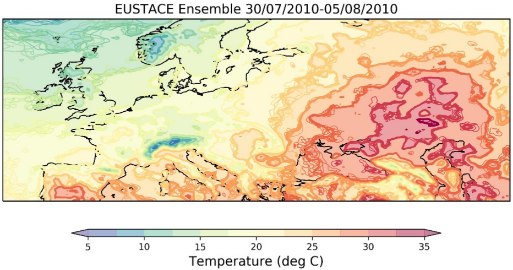


Met Office
Hadley Centre

ENSEMBLE ANALYSIS

- Samples drawn from joint posterior distribution of temperature and bias variables.
- Temperature model samples projected onto analysis grid.
- Spatial/temporal correlation in analysis errors is encoded into the ensemble.
- Summary statistics can be derived from the ensemble. Expected value, total uncertainty and observation constraint information also available.

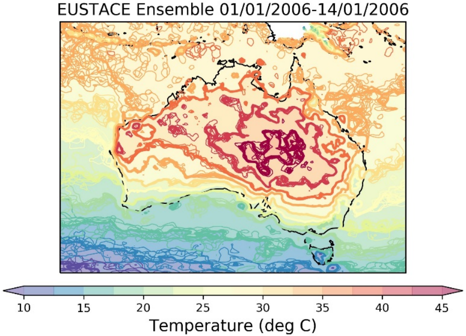


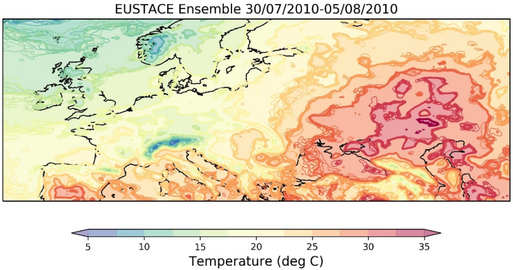


Met Office
Hadley Centre

ENSEMBLE ANALYSIS

- Samples drawn from joint posterior distribution of temperature and bias variables.
- Temperature model samples projected onto analysis grid.
- Spatial/temporal correlation in analysis errors is encoded into the ensemble.
- Summary statistics can be derived from the ensemble. Expected value, total uncertainty and observation constraint information also available.





Met Office
Hadley Centre

Summary

Only partially covered in this talk:

- Pure conditional block updates risk getting stuck;
  need for convergence acceleration
- Overlapping space-time blocks for preconditioning
- Non-stationary random field parameter estimation
- Direct&iterative variance calculations to eliminate or reduce
  Monte Carlo error in the reconstruction uncertainties
- Fast approximate handling of correlated error components

Summary:

- Challenging statistical problem, in both size and complexity
- Approximate calculation techniques allows some of the complexity
  to be handled with reasonable computational resources
  - SPDEs and Gaussian Markov random fields
  - Fast local sparse solves
  - Global multiscale block iteration
- Close collaboration between climate scientistis, statisticians,
  and software engineers is essential

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.