Hierarchical models and computing with stochastic PDEs

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

The University of Edinburgh, Scotland

with Colin Morice, John Kennedy, and the EUSTACE team,
and also David Bolin, Haavard Rue, Daniel Simpson, Elias Krainski

Expressing and Exploiting Structure in Modeling, Theory, and Computation with Gaussian Processes

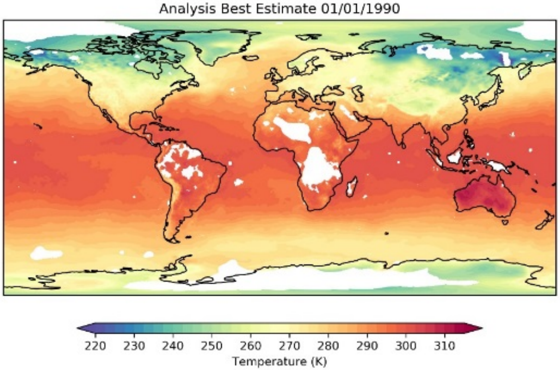
IMSI, Chicago, 29 August – 2 September 2022

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.



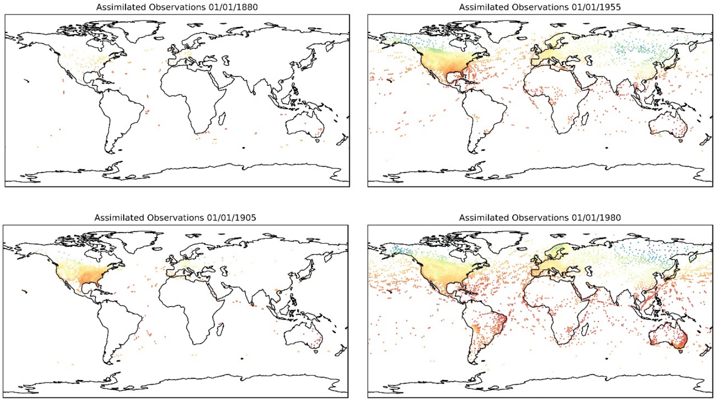
OBSERVATIONS

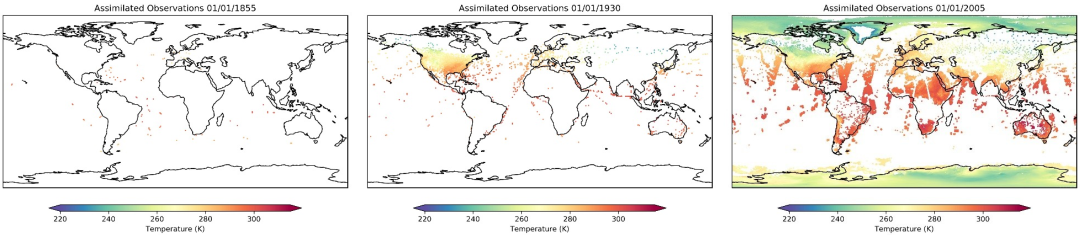
In situ air temperature:

- 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* (GMRFs)
- *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*.
  - Practical, *pragmatic* imlementation, starting with the most essential components.
  - Bayesian spatial analysis with GRMF of size 10^4 takes 90 sec. What scales to 10^{11}?

Example model: Matérn driven heat equation on the sphere

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

$$\left[\phi\frac{\partial}{\partial t} + (\kappa^2 - \Delta)^{\alpha_s/2}\right]^{\alpha_t} x(\mathbf{s},t)\,\mathrm{d}t = \mathrm{d}\mathcal{E}_{(\kappa^2-\Delta)^{\alpha_e}}(\mathbf{s},t)/\tau$$

For constant parameters, $x(\mathbf{s},t)$ has spatial Matérn covariance (for each t) in a Matérn-Whittle sense on $\mathbb{S}^2$.

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).

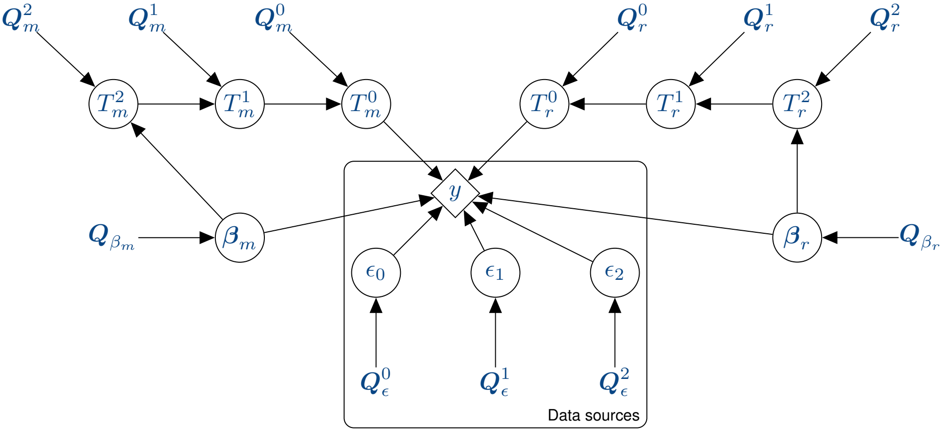
A finite element approximation has structure

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

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.

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

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

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.

Before satellites you had to go measure in person

"The Discovery", Dundee, Scotland (Photos: Finn Lindgren, August 2022)

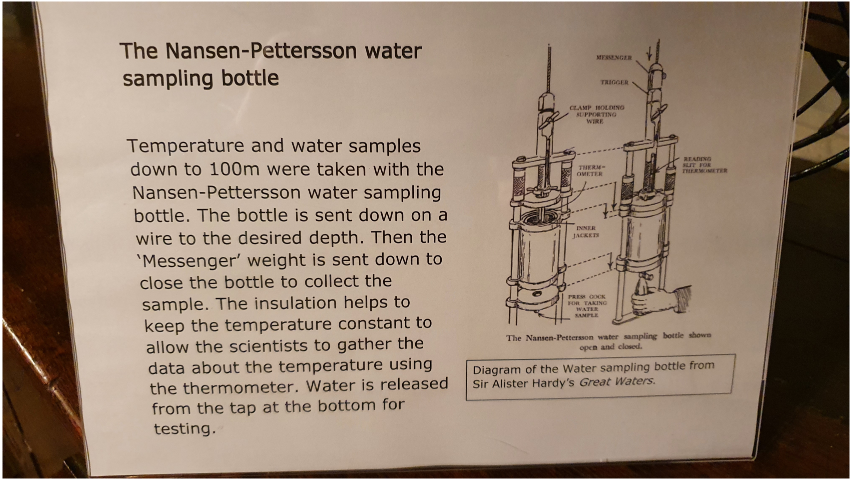
Hydrology lab from the 1925-27 Antarctic ocean expedition

"The Discovery", Dundee, Scotland (Photos: Finn Lindgren, August 2022)

What’s that in the corner?

"The Discovery", Dundee, Scotland (Photos: Finn Lindgren, August 2022)

It’s a Nansen-Pettersson water sampling bottle!



"The Discovery", Dundee, Scotland (Photos: Finn Lindgren, August 2022)

EUSTACE

Station observation & homogenisation model

Daily mean air temperature measurements

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

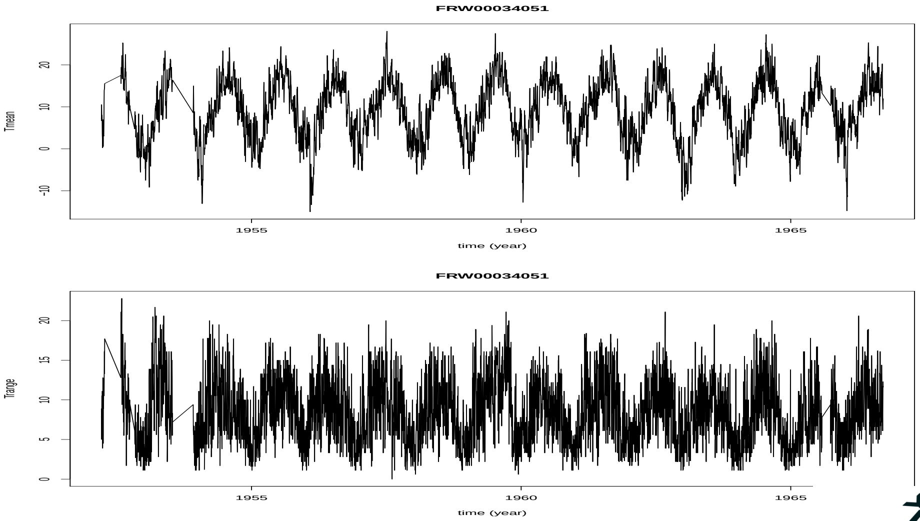
$$y_x^{k,i} = T_m(\mathbf{s}_k, t_i) + \widetilde{H}_{r,m}^k(t_i) + \frac{\widetilde{H}_{r,r}^k(t_i)}{2} T_r(\mathbf{s}_k, t_i) + \epsilon_x^{k,i},$$

$$y_n^{k,i} = T_m(\mathbf{s}_k, t_i) + \widetilde{H}_{r,m}^k(t_i) - \frac{\widetilde{H}_{r,r}^k(t_i)}{2} T_r(\mathbf{s}_k, t_i) + \epsilon_n^{k,i},$$

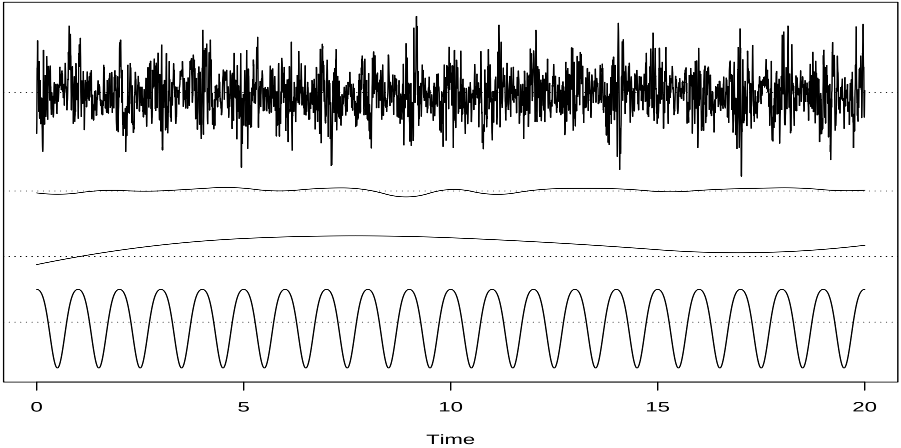
where $\widetilde{H}_\cdot$ are the total bias correction variables for each observation.

Observed data

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

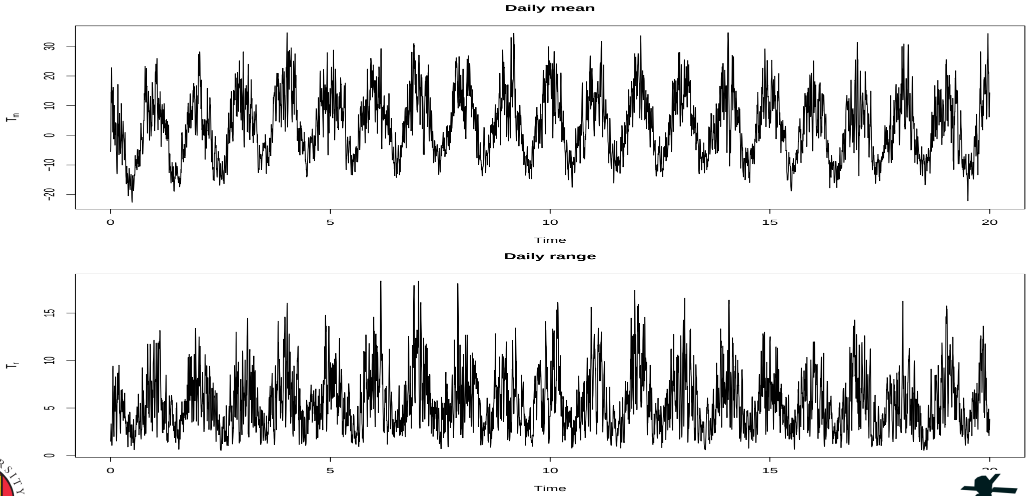


Multiscale model component samples

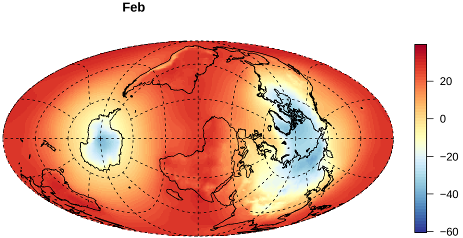


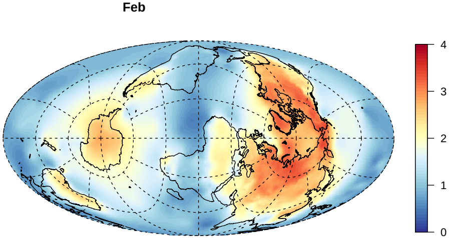
Combined model samples for T_m and T_r

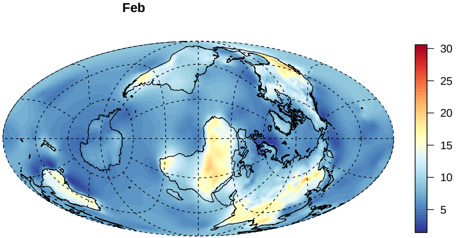
(Proof of concept; no actual data was involved in this figure)

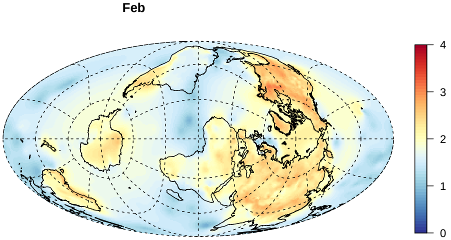


Estimates of median & scale for T_m and T_r









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 $\boldsymbol{Q}_x$:

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

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

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

Non-linear and/or non-Gaussian observations

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

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

$$\widetilde{\boldsymbol{Q}} = \boldsymbol{Q}_x + \boldsymbol{J}^\top \boldsymbol{Q}_{y|x} \boldsymbol{J} \qquad \text{(Generally: } \boldsymbol{Q}_x - \nabla_x \nabla_x^\top \log p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})\text{)}$$

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

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

Full non-linear solution 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/grid 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 $\boldsymbol{L}\boldsymbol{L}^\top = \boldsymbol{Q} = \boldsymbol{Q}_x + \boldsymbol{J}^\top \boldsymbol{Q}_{y|x}\boldsymbol{J}$.
  Possible due to the kronecker product sum precision structure: $\boldsymbol{L} = \left[\ldots, \boldsymbol{L}_k^{[s]} \otimes \boldsymbol{L}_k^{[t]}, \ldots, \boldsymbol{J}^\top \boldsymbol{L}_\epsilon\right]$

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

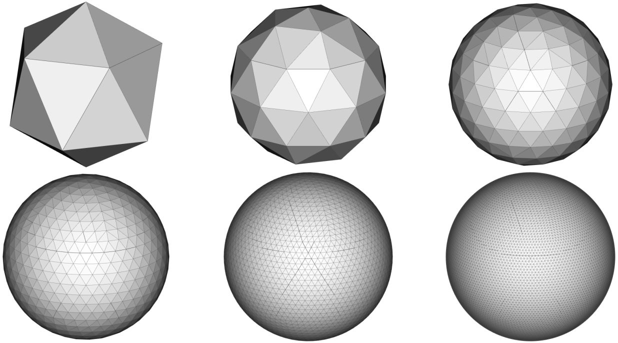
$$\boldsymbol{M}^{-1}\boldsymbol{x} = \sum_{k=1}^{K} \boldsymbol{W}_k \boldsymbol{D}_k (\boldsymbol{D}_k^\top \boldsymbol{Q} \boldsymbol{D}_k)^{-1} \boldsymbol{D}_k^\top \boldsymbol{W}_k \boldsymbol{x}$$

THE UNIVERSITY OF EDINBURGH

EUSTACE

EUSTACE pragmatic implementation

- Daily mean temperature only
- $\sim 60,000$ conditionally independent days (on the fine temporal scale): embarrassingly parallel daily direct solves
- Multiscale component grouped into three superblocks
- Reduced spatial resolution



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

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 $m = 1, \ldots, 3$ 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

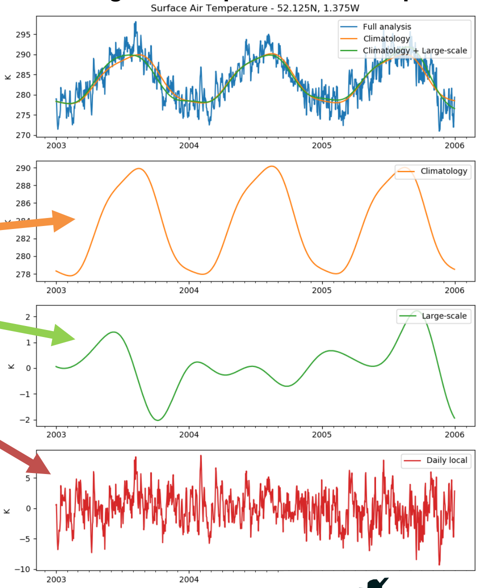
Statistical model for temperature variations and different scales (space and time):

- **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



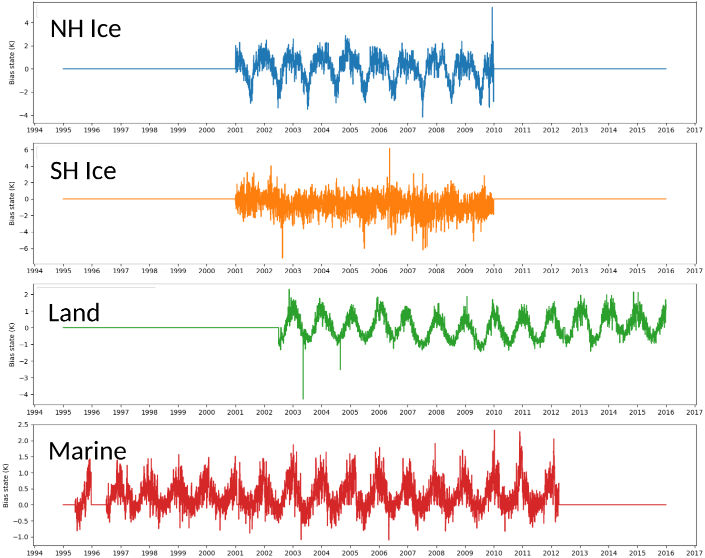
Met Office
Hadley Centre

EUSTACE

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.

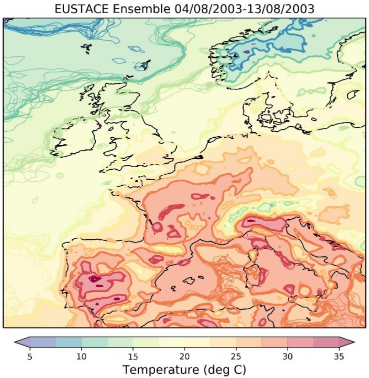
- Estimated jointly with daily temperature variability.

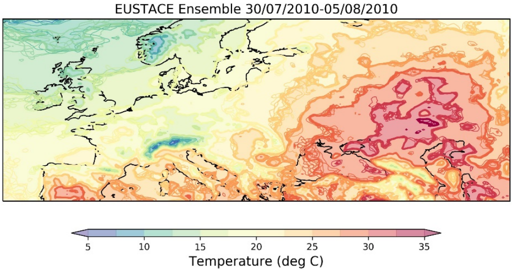


Met Office
Hadley Centre

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.

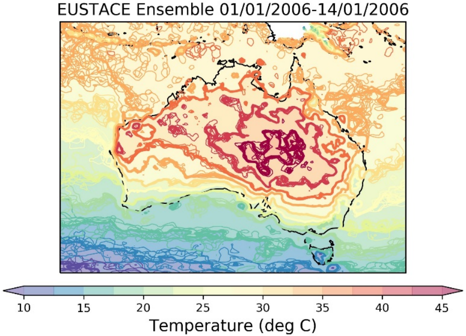


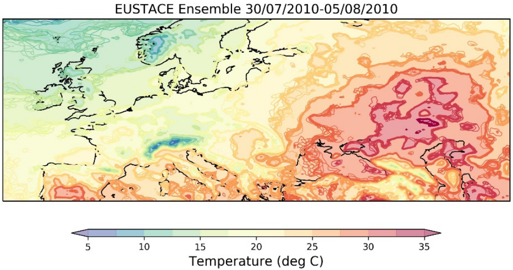


Met Office
Hadley Centre

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

Hierarchichal model challenges: Ideas to take home

- Real-life data behaviour introduces complex long-term dependence
- Methods for individual Gaussian fields are insufficient
- Efficient representations (Markov/SPDE/NNGP/Vecchia/Low rank/Blockwise/Incomplete Cholesky/etc) need to be coupled with proper iterative solvers
- Preconditioning needs to handle highly heterogeneous data
- We can handle up to $\lesssim 10^6$ latent variables exactly; use as preconditioner building blocks
- Multigrid/level methods appear highly promising for hierarchical space-time structures
- Need for flexible geography-induced non-stationarity modelling (not just estimation)
- Does increasing non-stationary reduce the need for global log-determinants?
  Should exploit local and multi-level hierarchy structure.

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://www.maths.ed.ac.uk/~flindgre/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/
- Lindgren, F., Bolin, D., and Rue, H.: The SPDE Approach for Gaussian and Non-Gaussian Fields: 10 Years and Still Running; *Spatial Statistics, Special Issue: The Impact of Spatial Statistics*, 50:100599. https://arxiv.org/abs/2111.01084
- Links to EUSTACE project reports and data: https://www.eustaceproject.org/
- Video illustrating the results, produced by Philip Brohan:
  https://twitter.com/philipbrohan/status/1253411283598073867
  https://player.vimeo.com/video/403663259