Developments in hybrid ensemble-variational DA for NWP

CliMatNet Conference

Andrew Lorenc, Bath, July 2015.

© Crown copyright Met Office

Content

1. Data Assimilation for Numerical Weather Prediction
2. 4-Dimensional Variational Data Assimilation Methods
3. Research Issues

Data Assimilation for Numerical Weather Prediction

© Crown copyright Met Office

Data Assimilation for (Global) Numerical Weather Prediction

An assimilation cycle is an adaptive filter process:

1. Every ~6 hours we make a forecast from the current best estimate, using best available computer forecast model
2. The DA process also provides an estimate of the error distribution of this forecast
3. We do a Bayesian combination of the prior from 1 & 2 with a batch of observations, to give a new best estimate valid ~6 hours later.
4. We measure the properties of the actual forecast error, and at intervals adapt the algorithm used in 2 (and also in 1 & 3).

Data Assimilation = The Scientific Method

Causes of improvements to NWP
What has been important for getting the best NWP forecast?*(over last 3 decades)*

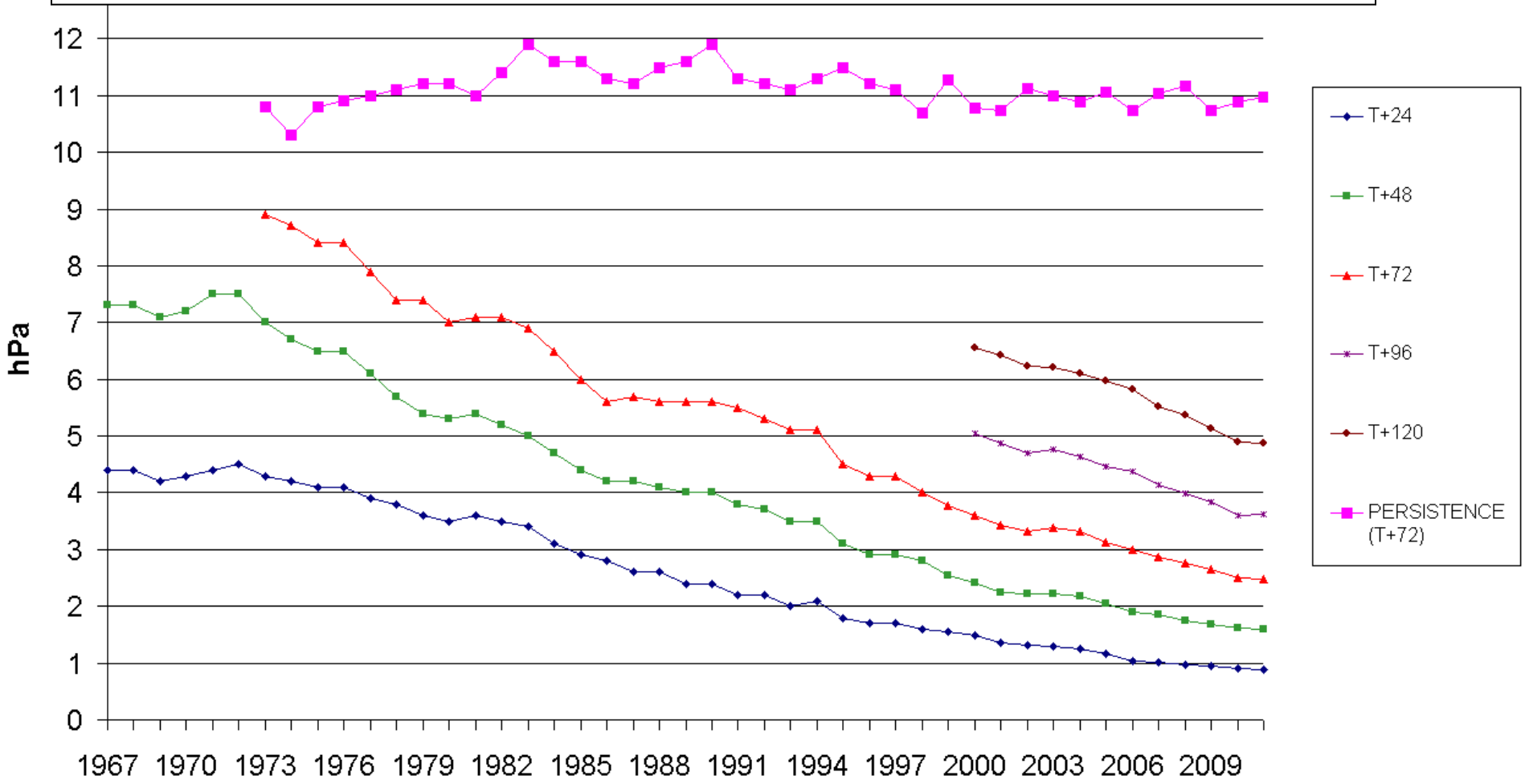
NWP systems are improving by 1 day of predictive skill per decade. This has been due to:

1. *Model improvements, especially resolution.*
2. *Careful use of forecast & observations, allowing for their information content and errors.*
3. *Advanced assimilation using forecast model: 4DVar has been successful.*
4. *Better observations.*

Performance Improvements

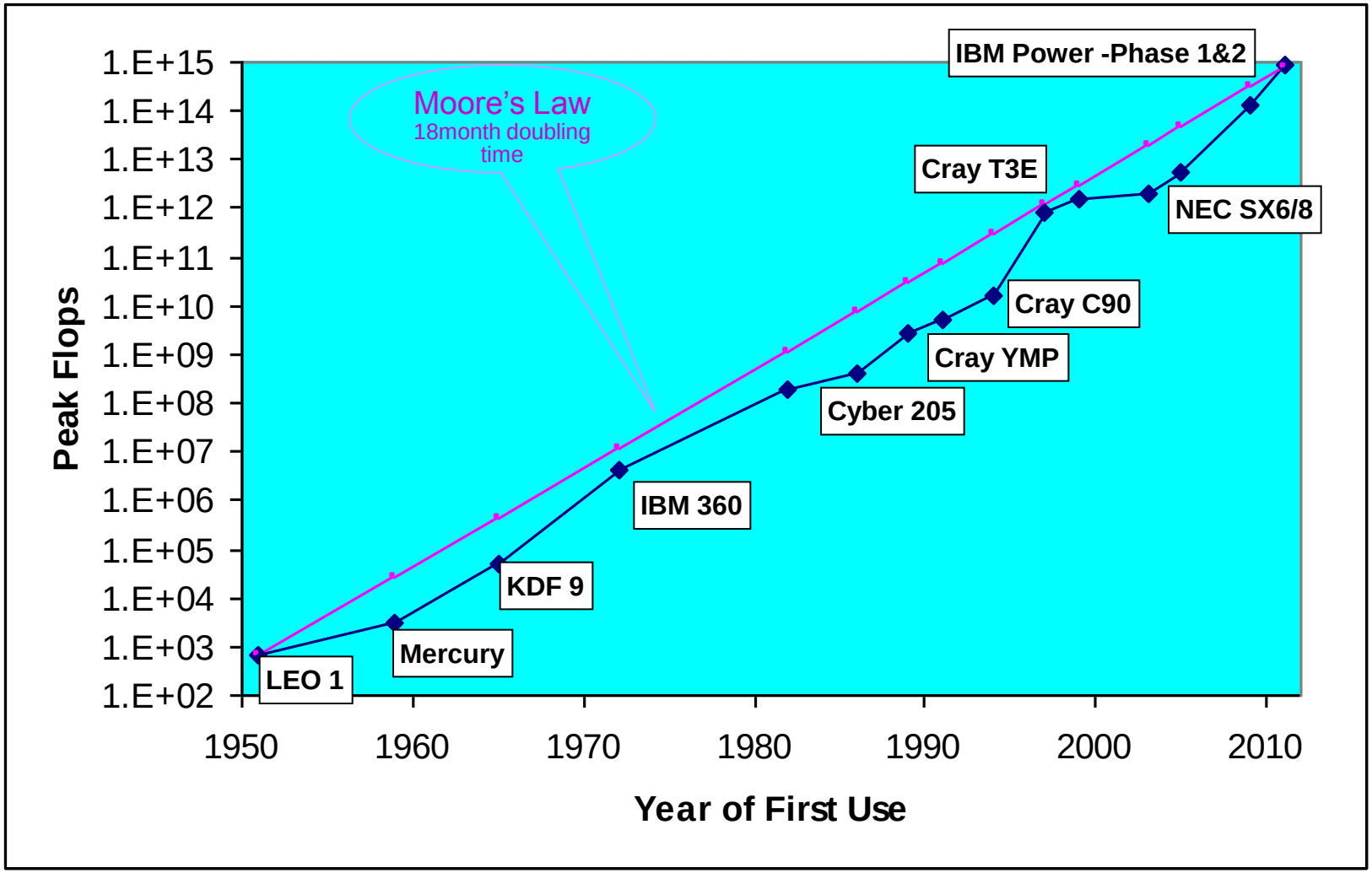
"Improved by about a day per decade"

Met Office RMS surface pressure error over the N. Atlantic & W. Europe



© Crown copyright Met Office Andrew Lorenc 6

60 Years of Met Office Computers



Continuing Improvement of a Complex System

- NWP improvements are due to a synergistic combination of improvements of forecast model, DA and observations.
  - Each helps the other.
- Total NWP system is very large, complex and expensive.
  - We cannot expect to understand it completely as a single entity.
  - We cannot afford thorough tuning of each improvement.
- It is best to base each improvement on scientific insight and analysis of one component.
  - With belief (checked by testing) that theoretically better parts will eventually give a better system.

Implications for DA strategy of *causes of improvements to NWP*

1. Model improvements, especially resolution.

- NWP models are very large (10^9 variables) and expensive to run. Nevertheless there are known shortcomings which could be addressed by making them larger and more expensive.
- DA improvements have to compete with the model for the use of computer upgrades. We cannot expect a large increase in the ***relative*** resources devoted to assimilation.
- Rely on linear DA methods to cope with 10^9 variables and millions of observations. Selected important nonlinearities can be added as corrections.

Implications for DA strategy of *causes of improvements to NWP*

2. Careful use of forecast & observations, allowing for their information content and errors.

- The forecast background summarises information from past observations. It is as accurate as most observations.
- Use **incremental** DA methods which correct the background based on a **Bayesian updating** with observed information.
- Need to understand the real information content of observations, e.g. satellite soundings observe radiances, not temperature profiles.

Implications for DA strategy of *causes of improvements to NWP*

3. Advanced assimilation using forecast model.

- 4D-Var is used by **4** of the top 7 global NWP centres.
  Ensemble Kalman filter methods are popular in other applications.
  Hybrid and Ensemble-Var methods are exciting much interest in operational NWP R&D.
- Multiple forecasts allow the evolution of error covariances and hence the definition of a 4D covariance.
- The NWP model is the best tool for defining the “attractor” of plausible “balanced” states.
  $\Rightarrow$Incremental methods designed such that background is only altered based on observed evidence.

© Crown copyright Met Office Andrew Lorenc 11

4-Dimensional Variational Data Assimilation Methods

© Crown copyright Met Office

4-Dimensional Variational DA Methods

A 4D best-fit to observations & background in a time-window, assuming Gaussian errors. I use underline to denote 4D variables and operators:

$\underline{\mathbf{x}}^b$	background trajectory
$\underline{\mathbf{P}}$	4D error covariance of $\underline{\mathbf{x}}^b$
$\delta\underline{\mathbf{x}}$	4D analysis increment
$\underline{\mathbf{y}} = \underline{H}\left(\underline{\mathbf{x}}^b + \delta\underline{\mathbf{x}}\right)$	model estimate of obs
$J\left(\delta\underline{\mathbf{x}}\right) = \frac{1}{2}\delta\underline{\mathbf{x}}^T\underline{\mathbf{P}}^{-1}\delta\underline{\mathbf{x}} + \frac{1}{2}\left(\underline{\mathbf{y}} - \underline{\mathbf{y}}^o\right)^T\underline{\mathbf{R}}^{-1}\left(\underline{\mathbf{y}} - \underline{\mathbf{y}}^o\right)$	penalty function

$\underline{\mathbf{P}}$ is **big!** We cannot even estimate it fully, let alone compute $\frac{1}{2}\delta\underline{\mathbf{x}}^T\underline{\mathbf{P}}^{-1}\delta\underline{\mathbf{x}}$.

The solution is to model $\underline{\mathbf{P}}$ using a sequence of operations we can compute, then use these to transform $\delta\underline{\mathbf{x}}$ so that $\frac{1}{2}\delta\underline{\mathbf{x}}^T\underline{\mathbf{P}}^{-1}\delta\underline{\mathbf{x}}$ simplifies.

© Crown copyright Met Office

4DVar: using climatological covariance $\mathbf{B}$

Model 3D covariance using transforms	$\mathbf{B} = \mathbf{U}\mathbf{U}^T$
3D analysis increment	$\delta\mathbf{x}_0 = \mathbf{U}\mathbf{v}^c$
made 4D using linear forecast model $\underline{\mathbf{M}}$	$\delta\underline{\mathbf{x}} = \underline{\mathbf{M}}\delta\mathbf{x}_0$
Implicit 4D prior covariance	$\underline{\mathbf{P}} = \underline{\mathbf{M}}\mathbf{B}\underline{\mathbf{M}}^T$
Transformed penalty function	$J(\mathbf{v}^c) = \frac{1}{2}\mathbf{v}^{cT}\mathbf{v}^c + \frac{1}{2}(\underline{\mathbf{y}} - \underline{\mathbf{y}}^o)^T \mathbf{R}^{-1}(\underline{\mathbf{y}} - \underline{\mathbf{y}}^o)$

© Crown copyright Met Office

Weaknesses of 4DVar.

- $\mathbf{B}$ – no flow-dependent **E**rrors **O**f **T**he **D**ay
  1. Use recent ensembles to train a new $\mathbf{B}$
  2. Use current ensemble to supplement $\mathbf{B}$
- Parallelisation – sequential runs of $\underline{\mathbf{M}}$ & $\underline{\mathbf{M}}^T$
  1. Parallelise in time too
  2. Use an ensemble instead (4DEnVar)
- No direct analysis ensemble
  1. Use a perturbed observation ensemble of 4DVars
  2. Use a separate EnKF system

Ensemble covariance filtering

B is big! We need a large ensemble PLUS clever filtering to reduce sampling noise, based on 2 ideas:

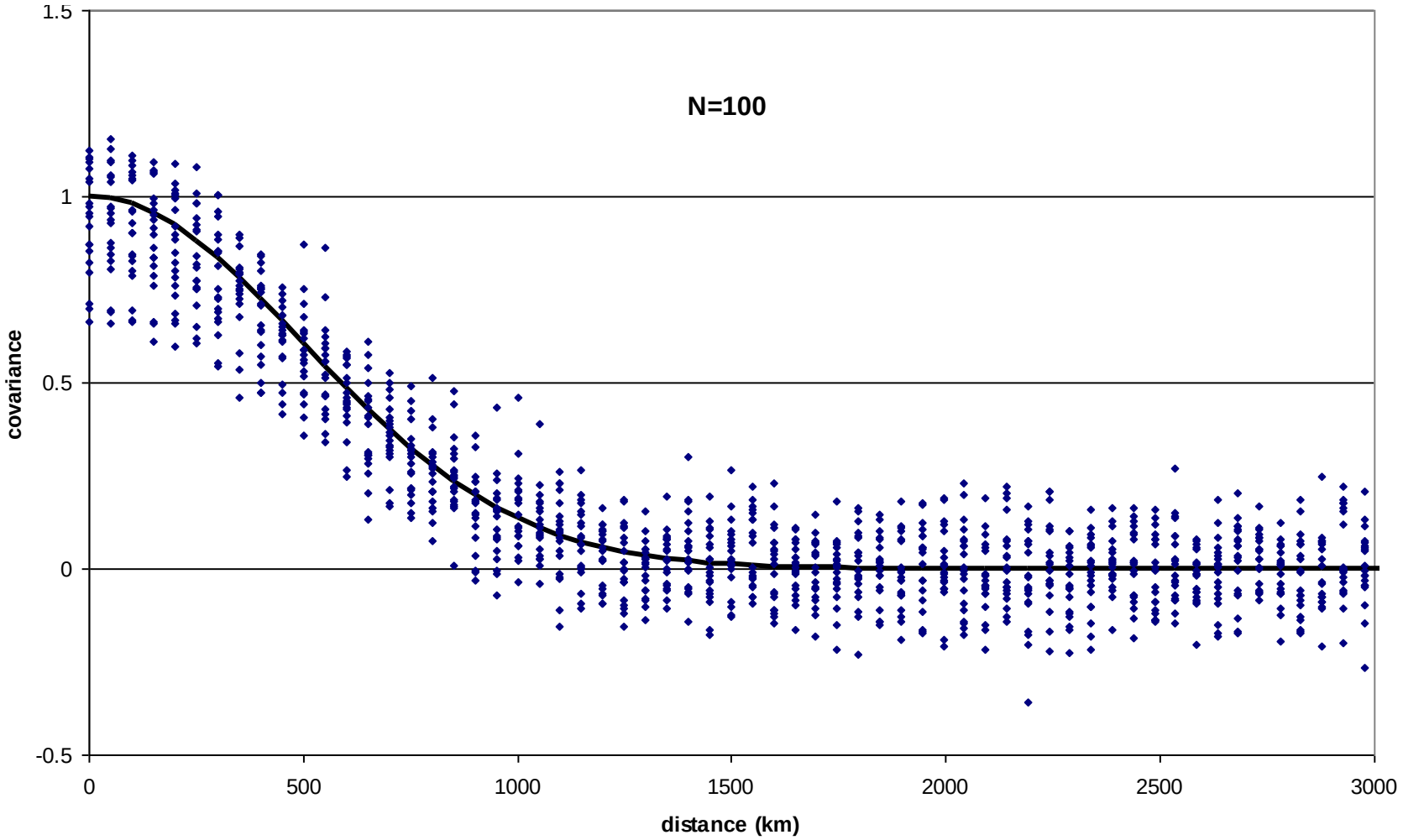
- Assume local homogeneity − apply smoothing:
  horizontal, rotational, and time
- Assume some correlations are near zero, & localise:
  horizontal, vertical, spectral, between transformed variables

Two approaches to hybrid covariances, using these ideas:

1. Train a covariance model using recent ensembles
2. Augment **B** by using localised ensemble perturbations

Horizontal localisation

Errors in sampled ensemble covariances



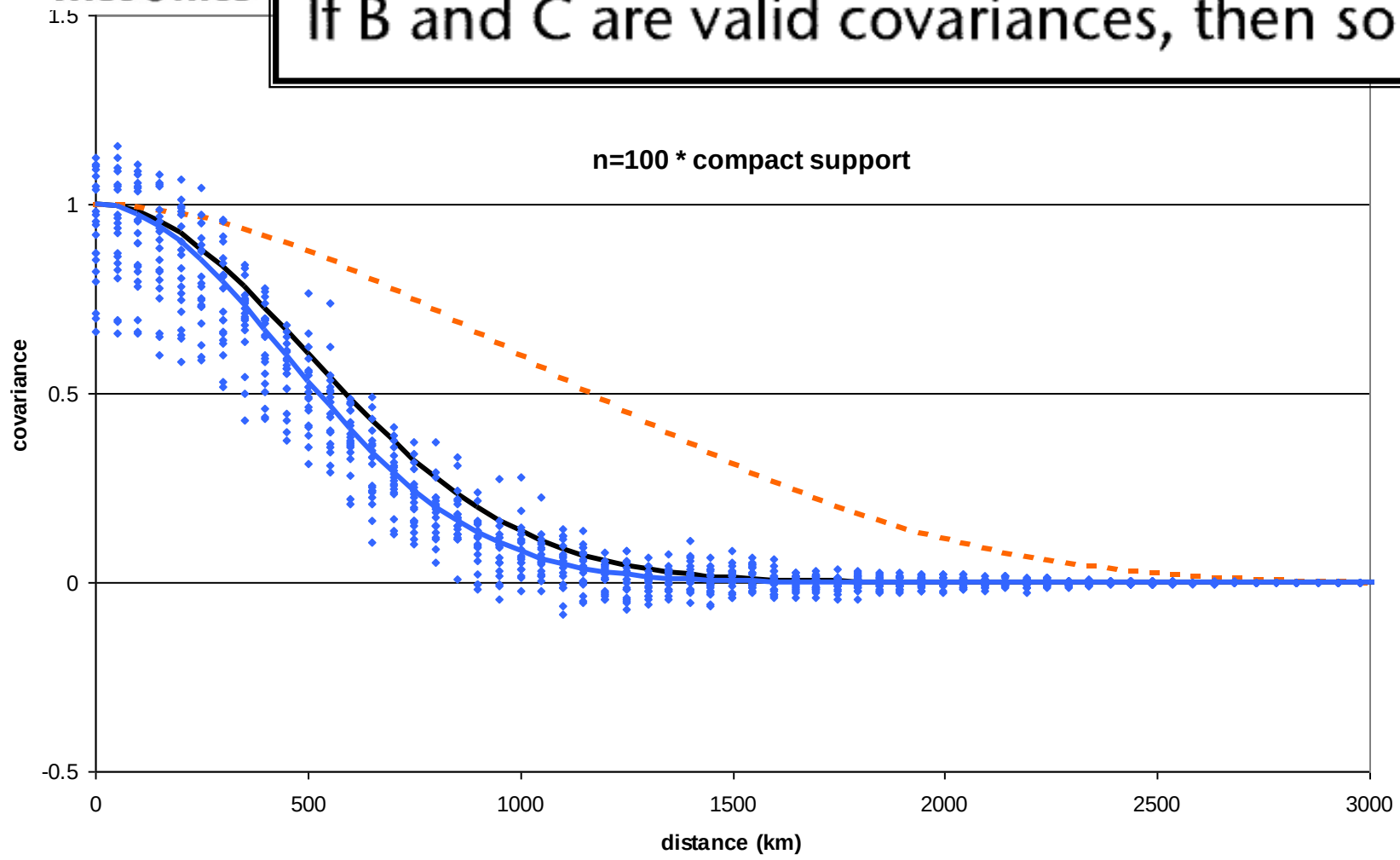
From Lorenc (2003)

The Schur Product

Met Office

$\mathbf{A} = \mathbf{B} \circ \mathbf{C}$ such that $A_{i,j} = B_{i,j} C_{i,j}$.

If B and C are valid covariances, then so is A.



From Lorenc (2003)

En-4DVar: using an ensemble of 3D states which samples background errors

Ensemble perturbation matrix $\mathbf{X} = \left[\mathbf{x}'_1 \ \cdots \ \mathbf{x}'_N \right]$ where $\mathbf{x}'_k = \frac{1}{\sqrt{N-1}} \left(\mathbf{x}_k - \bar{\mathbf{x}} \right)$

Model 3D $\mathbf{P}$ as localised ensemble covariance, $\mathbf{P} = \mathbf{C} \circ \mathbf{X}\mathbf{X}^T$

then model $\mathbf{C}$ using transforms $\mathbf{C} = \mathbf{U}^{\alpha}\mathbf{U}^{\alpha T}$

3D localised linear combination of ensemble perturbations

$$\boldsymbol{\alpha}_k = \mathbf{U}^{\alpha}\mathbf{v}_k^{\alpha}$$
$$\delta\mathbf{x}_0 = \Sigma_{k=1}^{N} \boldsymbol{\alpha}_k \circ \mathbf{x}'_k$$

then linear forecast model $\underline{\mathbf{M}}$ $\delta\underline{\mathbf{x}} = \underline{\mathbf{M}}\delta\mathbf{x}_0$

Localized 4D covariance $\underline{\mathbf{P}} = \underline{\mathbf{M}} \left(\mathbf{C} \circ \mathbf{X}\mathbf{X}^T \right) \underline{\mathbf{M}}^T$

concatenated control vectors $\mathbf{v}^T = \left[\mathbf{v}_1^{\alpha T} \cdots \mathbf{v}_N^{\alpha T} \right]$

Transformed penalty function $J\left(\mathbf{v}\right) = \frac{1}{2}\mathbf{v}^T\mathbf{v} + \frac{1}{2}\left(\underline{\mathbf{y}} - \underline{\mathbf{y}}^o\right)^T \underline{\mathbf{R}}^{-1} \left(\underline{\mathbf{y}} - \underline{\mathbf{y}}^o\right)$

© Crown copyright Met Office Andrew Lorenc

hybrid-4DVar

4D analysis increment $\delta\underline{\mathbf{x}} = \underline{\mathbf{M}}\left(\beta_c\mathbf{U}\mathbf{v}^c + \beta_e \Sigma_{k=1}^{N} \mathbf{U}^\alpha\mathbf{v}_k^\alpha \circ \mathbf{x}_k'\right)$

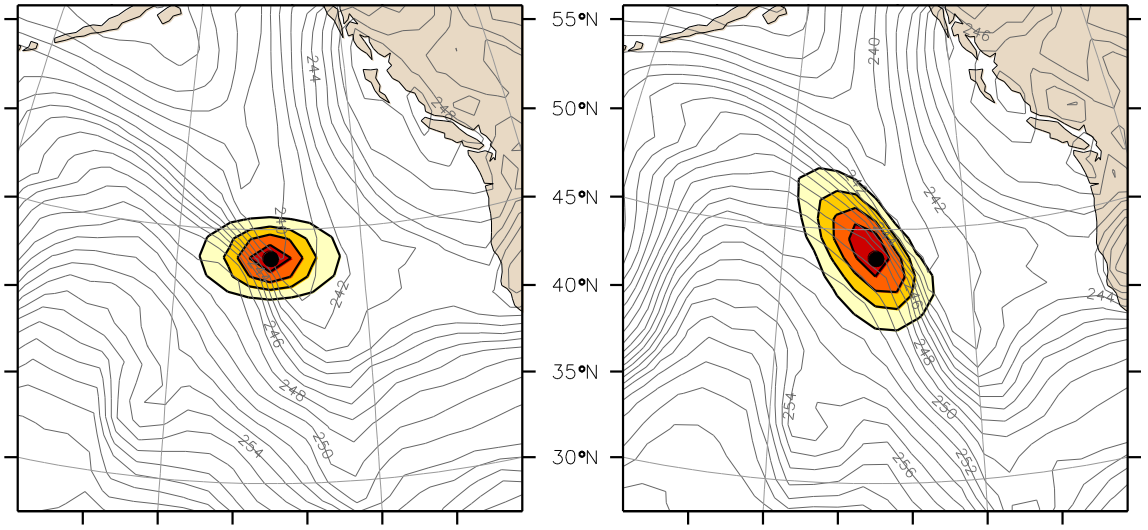
Localized 4D covariance $\underline{\mathbf{P}} = \underline{\mathbf{M}}\left(\beta_c^2\mathbf{B} + \beta_e^2\mathbf{C} \circ \mathbf{X}\mathbf{X}^T\right)\underline{\mathbf{M}}^T$

concatenated control vectors $\mathbf{v}^T = \left[\mathbf{v}^{cT}, \mathbf{v}_1^{\alpha T} \cdots \mathbf{v}_N^{\alpha T}\right]$

1% improvement in rms errors when implemented at Met Office (Clayton *et al*. 2013)

u increments fitting a single u ob at 500hPa, at different times.

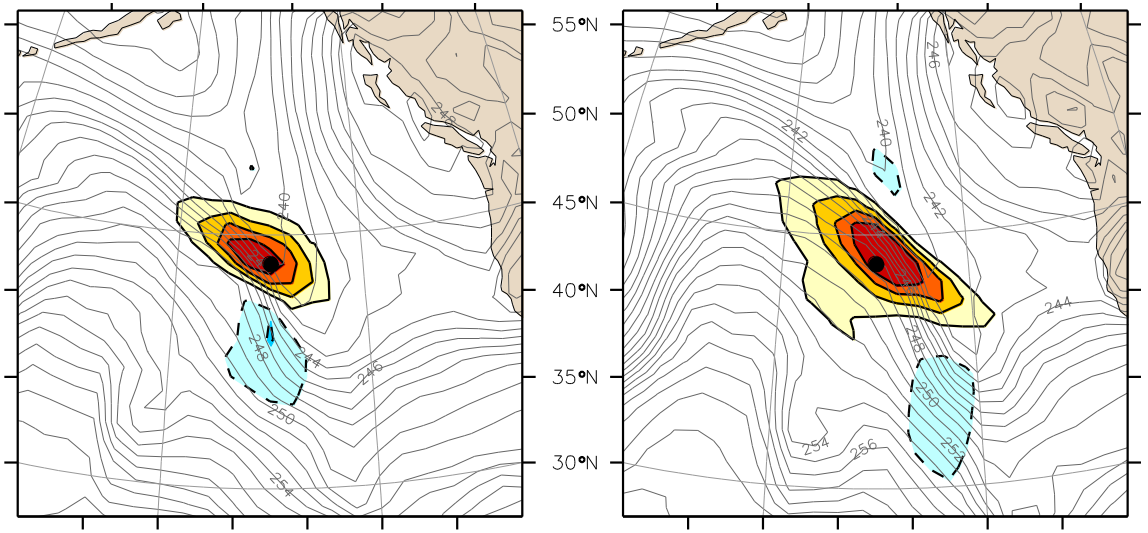
4D-Var



at start of window

at end of 6-hour window

Hybrid 4D-Var



Unfilled contours show T field.

Clayton *et al.* 2013

4DEnVar: using an ensemble of 4D trajectories which samples background errors

Ensemble trajectory matrix $\underline{\mathbf{X}} = \left[\underline{\mathbf{x}}'_1 \ \cdots \ \underline{\mathbf{x}}'_N \right]$ where $\underline{\mathbf{x}}'_k = \frac{1}{\sqrt{N-1}} \left(\underline{\mathbf{x}}_k - \bar{\underline{\mathbf{x}}} \right)$

Model 4D $\underline{\mathbf{P}}$ directly, as localised ensemble covariance,

$$\underline{\mathbf{P}} = \underline{\mathbf{C}} \circ \underline{\mathbf{X}}\underline{\mathbf{X}}^T$$

then model $\underline{\mathbf{C}}$ using transforms

$$\underline{\mathbf{C}} = \underline{\mathbf{U}}^\alpha \underline{\mathbf{U}}^{\alpha T}$$

4D localised linear combination of ensemble trajectories

$$\underline{\boldsymbol{\alpha}}_k = \underline{\mathbf{U}}^\alpha \mathbf{v}_k^\alpha$$

$$\delta\underline{\mathbf{x}} = \Sigma_{k=1}^{N} \underline{\boldsymbol{\alpha}}_k \circ \underline{\mathbf{x}}'_k$$

concatenated control vectors

$$\mathbf{v}^T = \left[\mathbf{v}_1^{\alpha T} \cdots \mathbf{v}_N^{\alpha T} \right]$$

Transformed penalty function

$$J(\mathbf{v}) = \tfrac{1}{2}\mathbf{v}^T\mathbf{v} + \tfrac{1}{2}\left(\underline{\mathbf{y}} - \underline{\mathbf{y}}^o\right)^T \underline{\mathbf{R}}^{-1} \left(\underline{\mathbf{y}} - \underline{\mathbf{y}}^o\right)$$

© Crown copyright Met Office

4DEnVar: using an ensemble of 4D trajectories which samples background errors

Extra Details

model $\underline{\mathbf{C}}$ using transforms

$$\underline{\mathbf{C}} = \underline{\mathbf{U}}^{\alpha}\underline{\mathbf{U}}^{\alpha T}$$

It is common to use a 3D $\mathbf{C}$

$$\mathbf{C} = \mathbf{U}^{\alpha}\mathbf{U}^{\alpha T}$$

and persistence in time: $\underline{\mathbf{I}}$

$$\underline{\mathbf{C}} = \underline{\mathbf{I}}\mathbf{C}\underline{\mathbf{I}}^{T}$$

4D localised linear combination of ensemble trajectories

$$\delta\underline{\mathbf{x}} = \Sigma_{k=1}^{N}\,\underline{\boldsymbol{\alpha}}_k \circ \underline{\mathbf{x}}'_k$$

can be built from 3D localised perturbations and constant $\boldsymbol{\alpha}_k$.

$$\boldsymbol{\alpha}_k = \mathbf{U}^{\alpha}\mathbf{v}_k^{\alpha}$$

$$\delta\mathbf{x}\left(t\right) = \Sigma_{k=1}^{N}\boldsymbol{\alpha}_k \circ \mathbf{x}'_k\left(t\right)$$

Matrix notation:

$$\mathbf{A} = \left[\,\boldsymbol{\alpha}'_1\ \cdots\ \boldsymbol{\alpha}'_N\,\right]$$

$$\delta\mathbf{x}\left(t\right) = \left(\mathbf{A} \circ \mathbf{X}\left(t\right)\right)\mathbf{1}_N$$

$\mathbf{1}_N$ is a column vector of N 1s

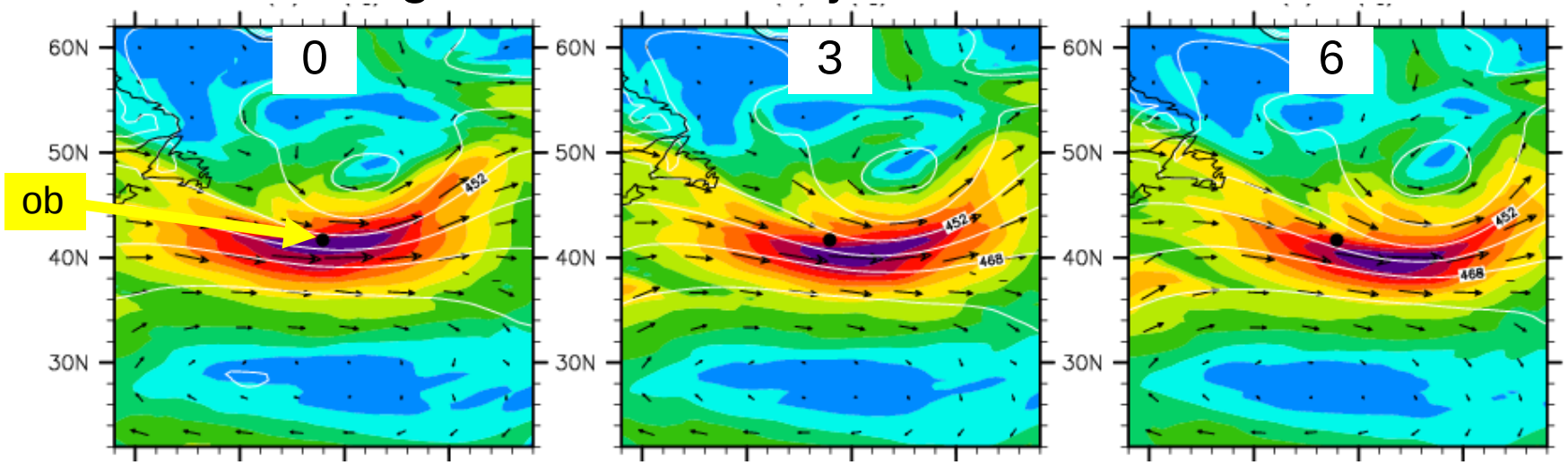
© Crown copyright Met Office

An new form of linear model

En-4DVar analysis increment $\delta\underline{\mathbf{x}} = \underline{\mathbf{M}}\,\Sigma_{k=1}^{N}\,\boldsymbol{\alpha}_k \circ \mathbf{x}'_k$

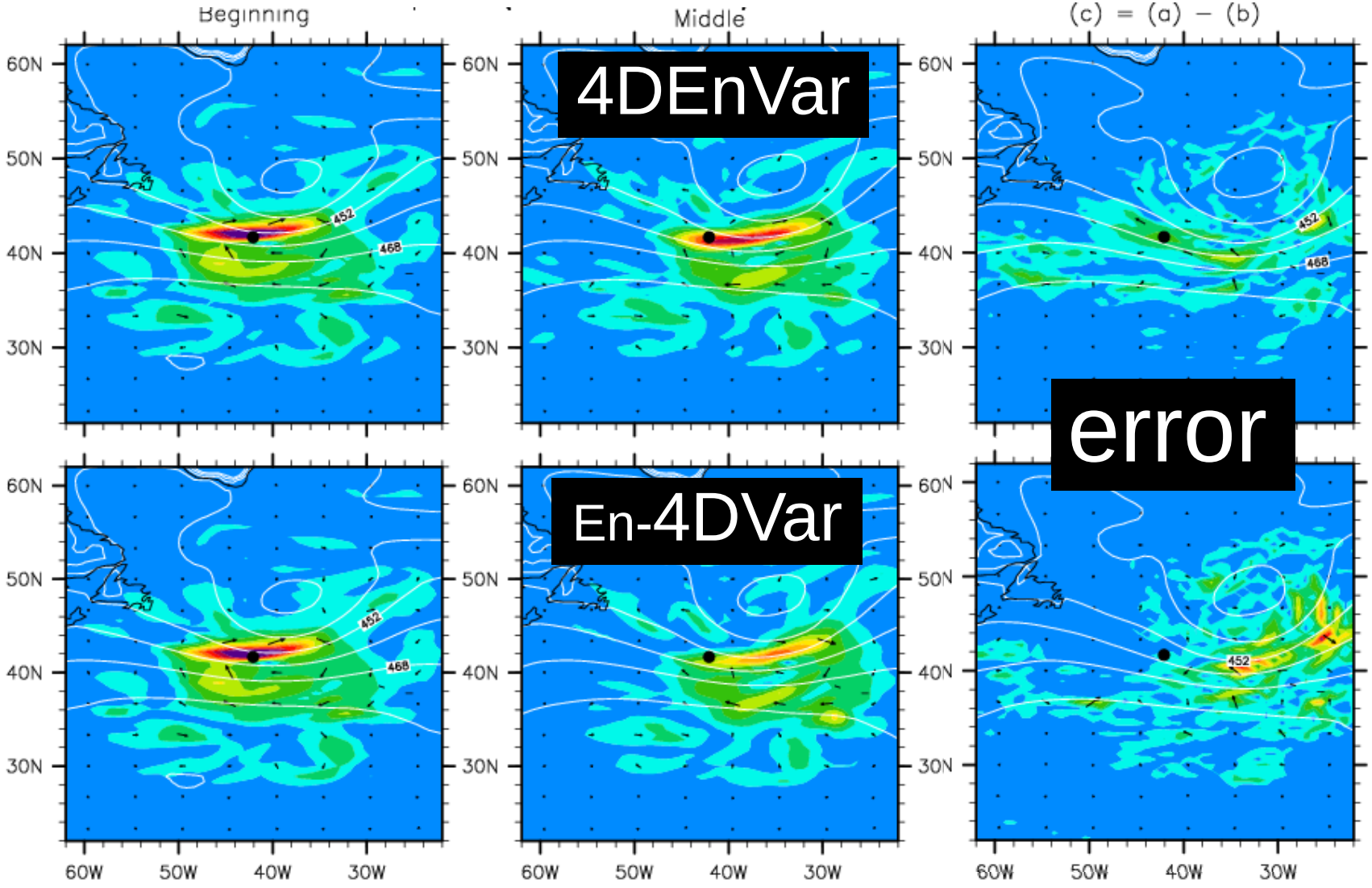
4DEnVar analysis increment $\delta\underline{\mathbf{x}} = \Sigma_{k=1}^{N}\,\underline{\boldsymbol{\alpha}}_k \circ \underline{\mathbf{x}}'_k$

Test with a single wind ob, in a jet, at the start of the window



© Crown copyright Met Office

100% ensemble
1200km localization scale



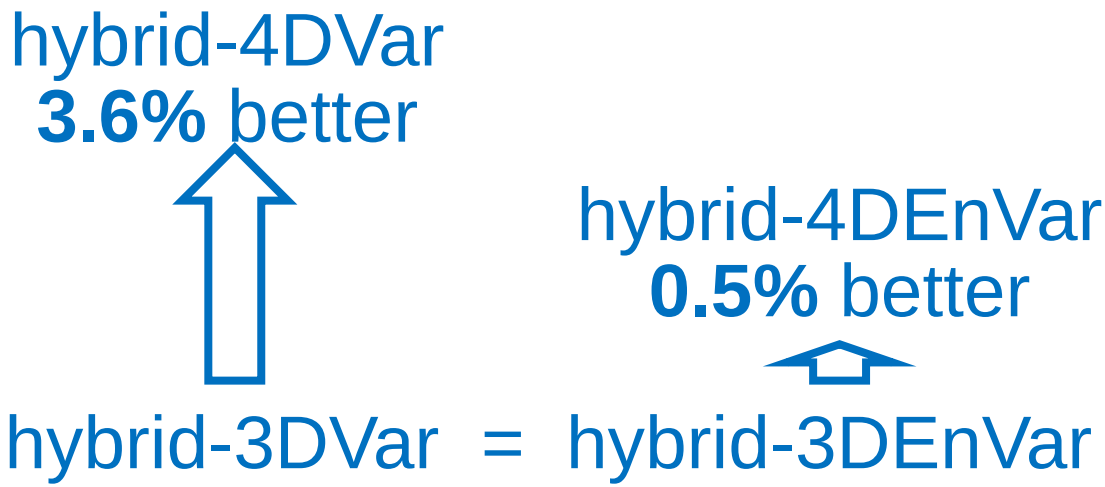
Met Office trial of 4DEnVar

Lorenc et al. (2014)

Our first trial copied settings from the hybrid-4DVar:

- C with localisation scale 1200km,
- hybrid weights $\beta_c^2=0.8$, $\beta_e^2=0.5$

Results were disappointing:



The reason was the large weight given to the climatological covariance, which is treated like 3DVar in 4DEnVar

© Crown copyright Met Office Andrew Lorenc

Why?

1. $\mathbf{B}^c$ was given large weight (80%) and treated differently: propagated by model in 4DVar, persisted in hybrid-4DEnVar.

2. Non-propagation of localisation.

3. Initialisation: 4DVar's J_c ensured that its increment was a balanced trajectory of the PF model.

4. Insufficient degrees of freedom: more are needed before increasing weight on $\mathbf{B}^e$, *even more* for $\underline{\mathbf{B}}^e$.

5. Something else?
   Why did hybrid-4DVar beat hybrid-3DVar by so much?

hybrid-4DEnVar

4D analysis increment $\delta\underline{\mathbf{x}} = \beta_c\underline{\mathbf{I}}\delta\mathbf{x}_0 + \beta_e \Sigma_{k=1}^{N} \underline{\boldsymbol{\alpha}}_k \circ \underline{\mathbf{x}}'_k$

Localized 4D covariance $\underline{\mathbf{P}} = \beta_c^2\underline{\mathbf{I}}\mathbf{B}\underline{\mathbf{I}}^T + \beta_e^2\underline{\mathbf{C}} \circ \underline{\mathbf{X}}\underline{\mathbf{X}}^T$

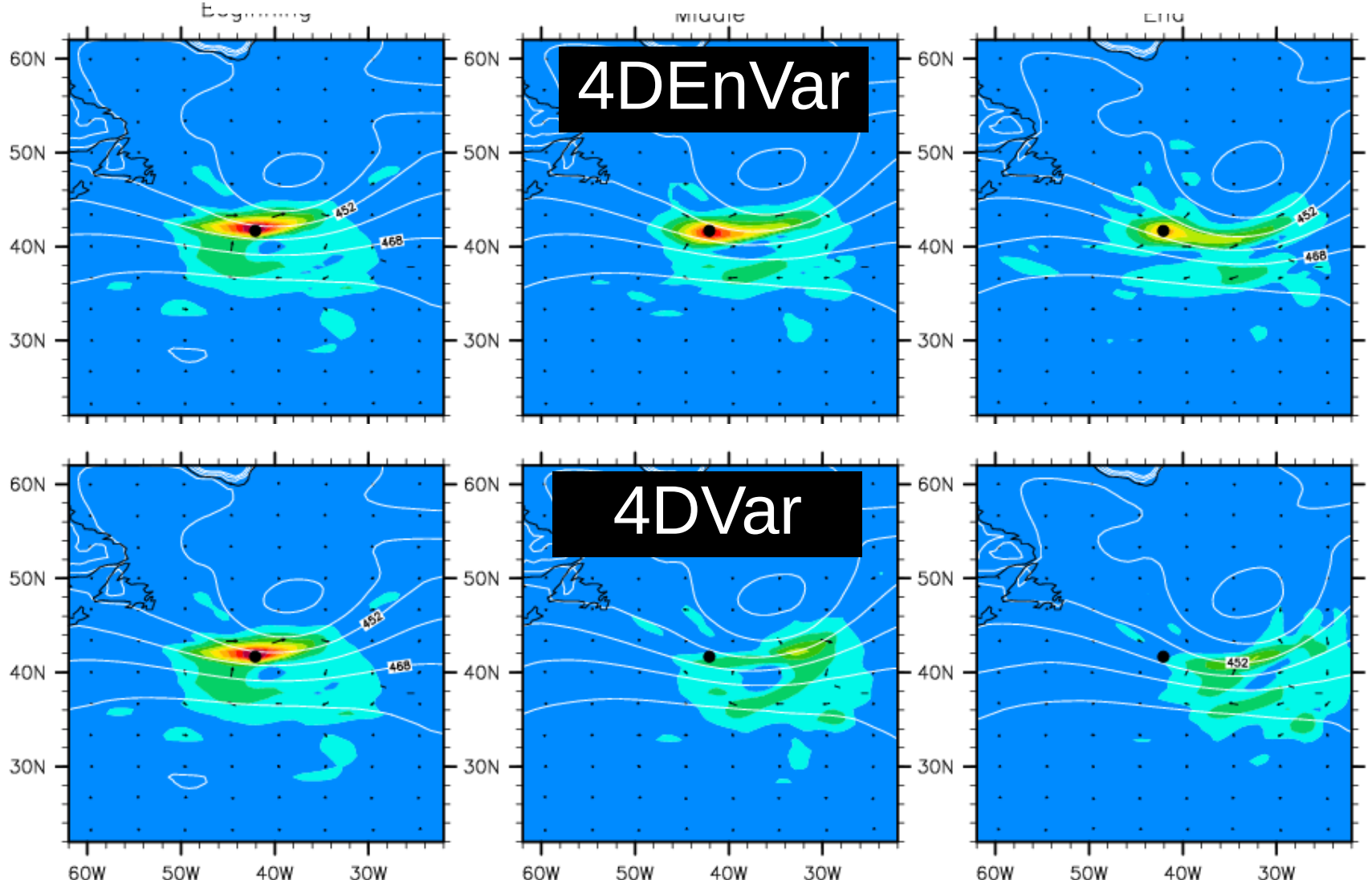
hybrid-4DVar

4D analysis increment $\delta\underline{\mathbf{x}} = \underline{\mathbf{M}}\left(\beta_c\delta\mathbf{x}_0 + \beta_e \Sigma_{k=1}^{N} \boldsymbol{\alpha}_k \circ \mathbf{x}'_k\right)$

Localized 4D covariance $\underline{\mathbf{P}} = \underline{\mathbf{M}}\left(\beta_c^2\mathbf{B} + \beta_e^2\mathbf{C} \circ \mathbf{X}\mathbf{X}^T\right)\underline{\mathbf{M}}^T$

© Crown copyright

50-50% hybrid
1200km localization scale



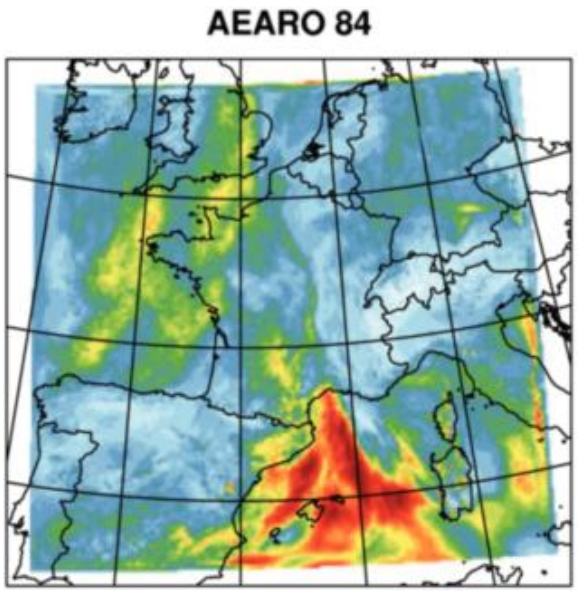
Research Issues

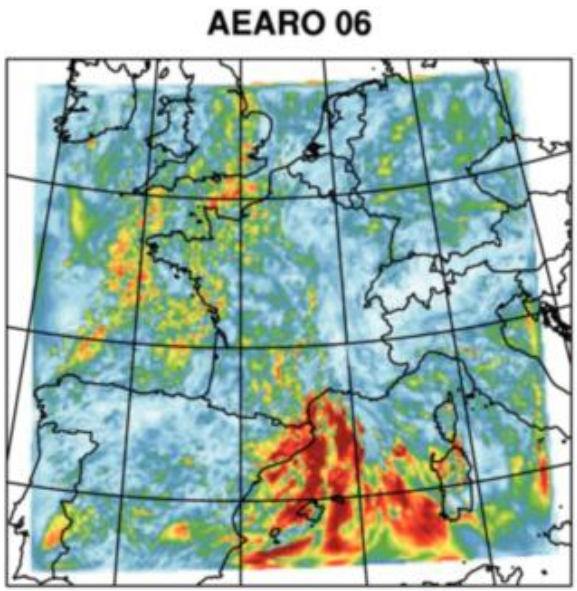
© Crown copyright Met Office

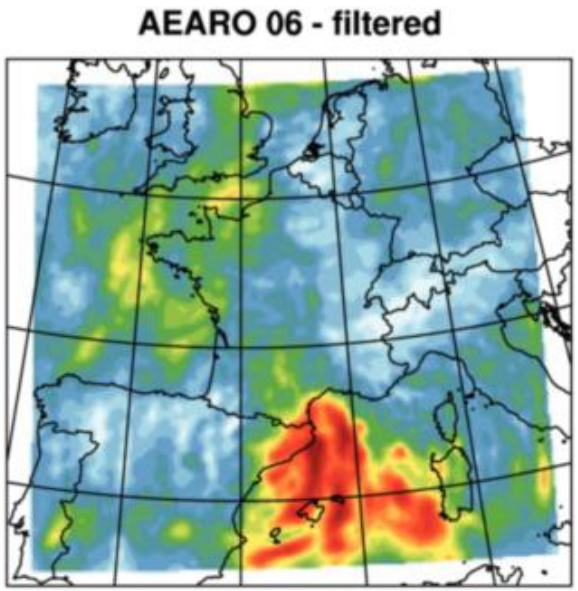
Research Issues

- 4DEnVar & ensemble covariance filtering
- Ensemble generation
- Model constraint & representation of model errors
- Nonlinearities & non-Gaussianity
- Massively parallel computers, advanced software systems, modularisation & separation of concerns.

Convective-scale ensemble s.d. of humidity at 945hPa







Large ensemble (84 members)

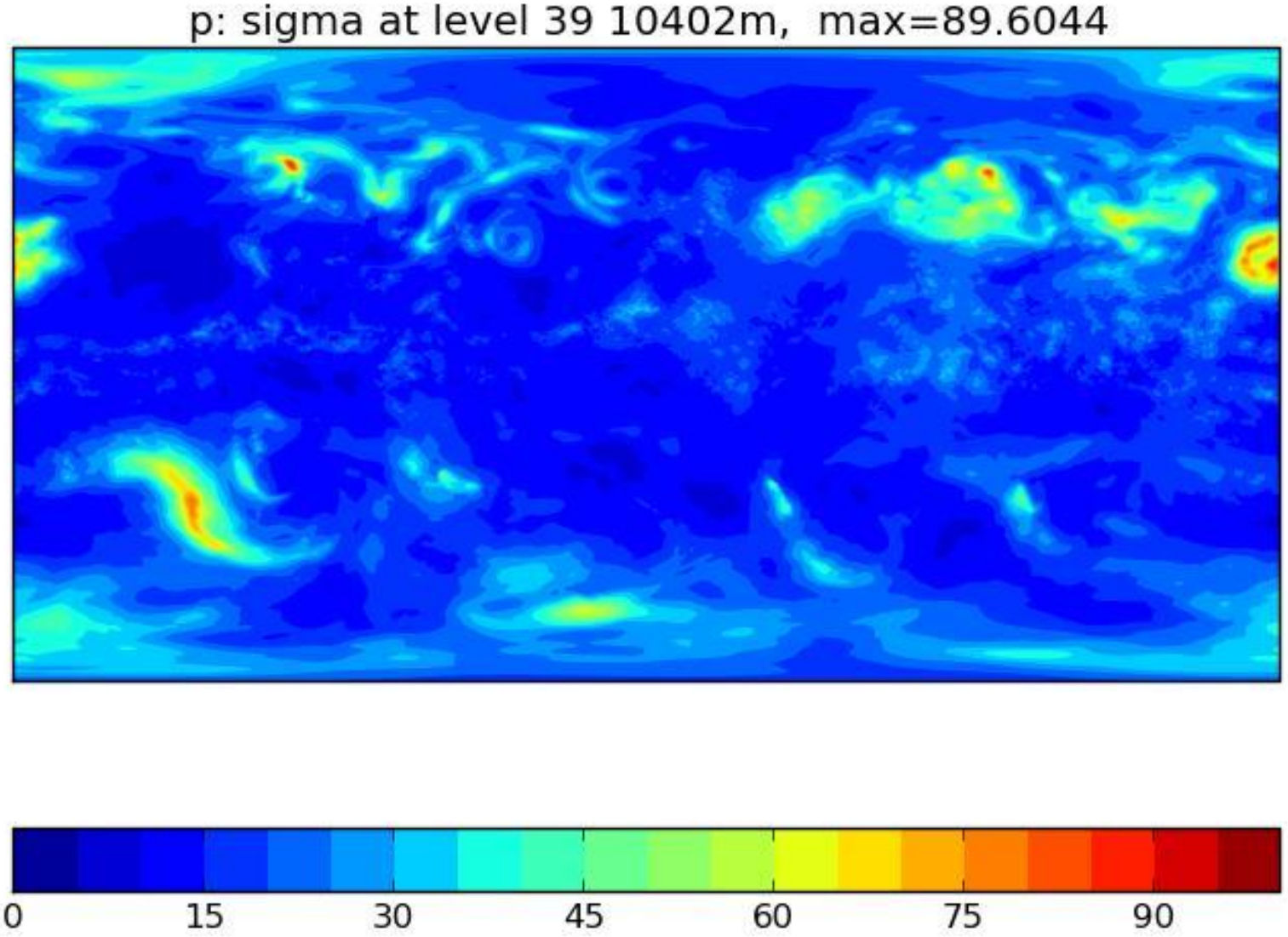
Small ensemble (6 members)

Horizontally filtered small ensemble

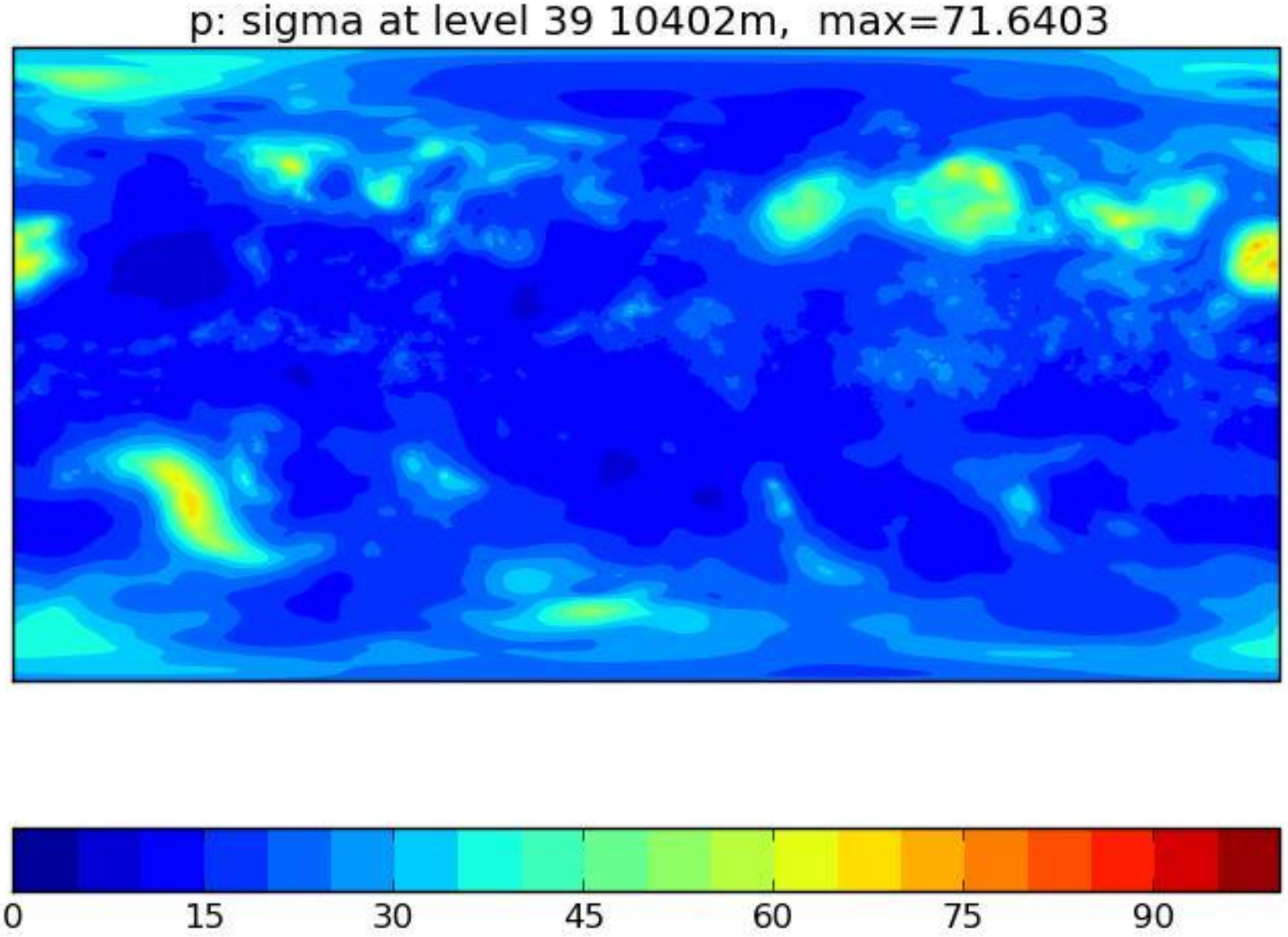
Using the AROME ensemble (Ménétrier et al. 2014).

© Crown copyright Met Office Andrew Lorenc 32

Sampled raw ensemble s.d.

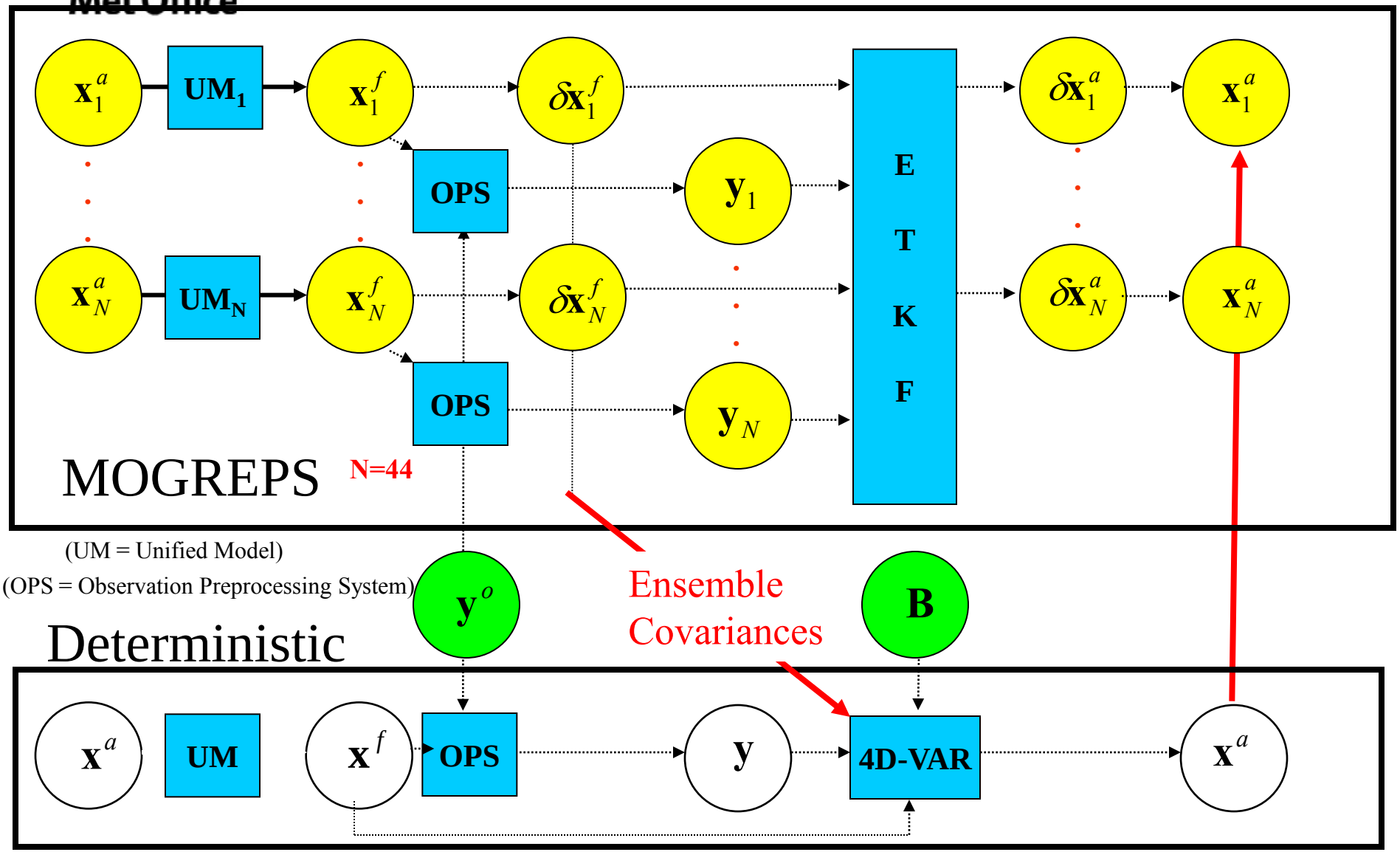


s.d. after spectral localization



2-Way Coupled DA/E

$$\mathbf{y}_n = H(\mathbf{x}_n^f), \sigma_o, \ldots \qquad \mathbf{x}_n^a = \mathbf{x}^a + \delta\mathbf{x}_n^a$$

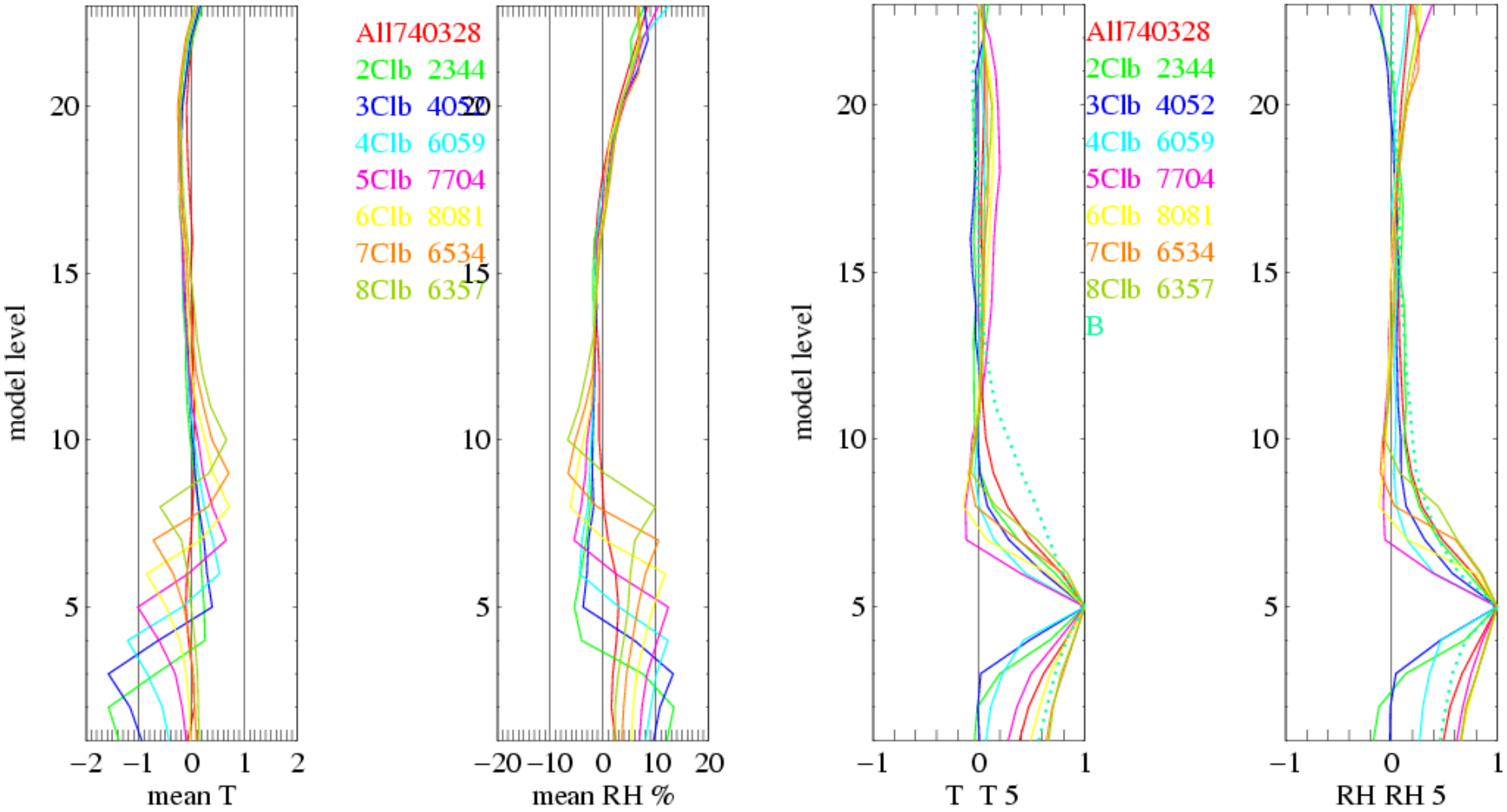


Met Office

Error PDFs composited by cloudy inversion level:

Left: mean of T & RH, ⇨ large bias

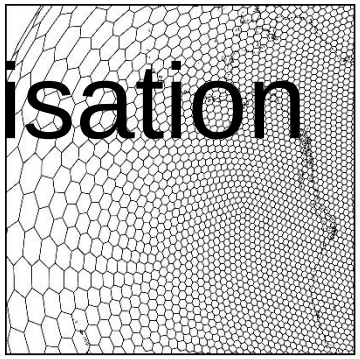
Right: correlations with level 5, ⇨ ~0 across inversion

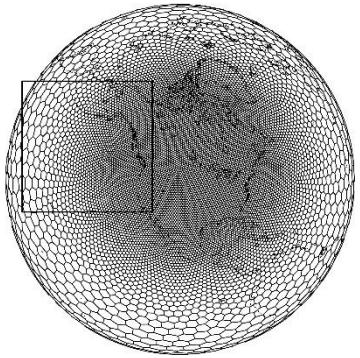


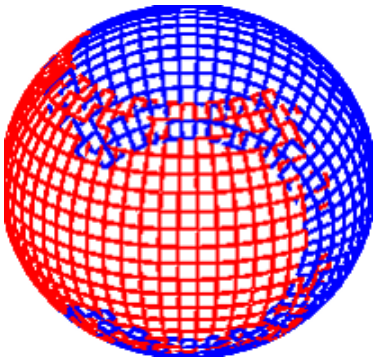
Lorenc (2007)

© Crown copyright Met Office Andrew Lorenc 36

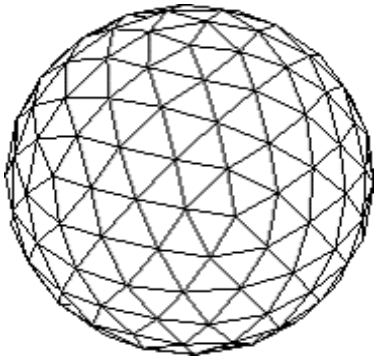
Parallelisation

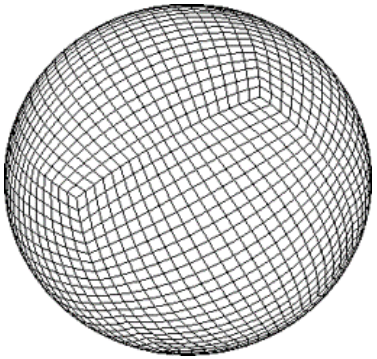






4DVar has several potential problems looming in the next decade - their timing for each centres will depend on their computers and models:





1. Need new design to use millions of parallel threads, especially in sequential runs of linear (PF) and Adjoint models.

2. Forecast models are being redesigned to address this − a maintenance issue for the PF and Adjoint models.

A simple solution is to use the ensemble trajectories, pre-calculated in parallel, instead of the models inside 4DVar.

If Fourier filters and Poisson solvers are not available then the LETKF is an easier approach.

© Crown copyright Met Office Andrew Lorenc 37

Questions?

References

Clayton AM, Lorenc AC, Barker DM. 2013. Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the Met Office. *Q. J. R. Meteorol. Soc.* **139**: 1445–1461, doi:http://dx.doi.org/10.1002/qj.2054.

Lorenc AC. 2003. The potential of the ensemble Kalman filter for NWP - a comparison with 4D-Var. *Q. J. R. Meteorol. Soc.* **129**: 3183–3203, doi:http://dx.doi.org/10.1256/qj.02.132.

Lorenc AC. 2007. A study of o-b monitoring statistics from radiosondes, composited for low-level cloud layers. Met Office Forecasting Research Tech. Rept. 504, Met Office, URL `http://www.metoffice.gov.uk/archive/forecasting-research-technical-report-504`.

Lorenc AC, Bowler NE, Clayton AM, Fairbairn D, Pring SR. 2015. Comparison of hybrid-4DEnVar and hybrid-4DVar data assimilation methods for global NWP. *Mon. Weather Rev.* **143**: 212–229, doi:10.1175/MWR-D-14-00195.1.

Ménétrier B, Montmerle T, Berre L, Michel Y. 2014. Estimation and diagnosis of heterogeneous flow-dependent background-error covariances at the convective scale using either large or small ensembles. *Q. J. R. Meteorol. Soc.* doi:10.1002/qj.2267.