Met Office

Hybrid Variational/Ensemble Data Assimilation

Dale Barker, Adam Clayton, and Andrew Lorenc

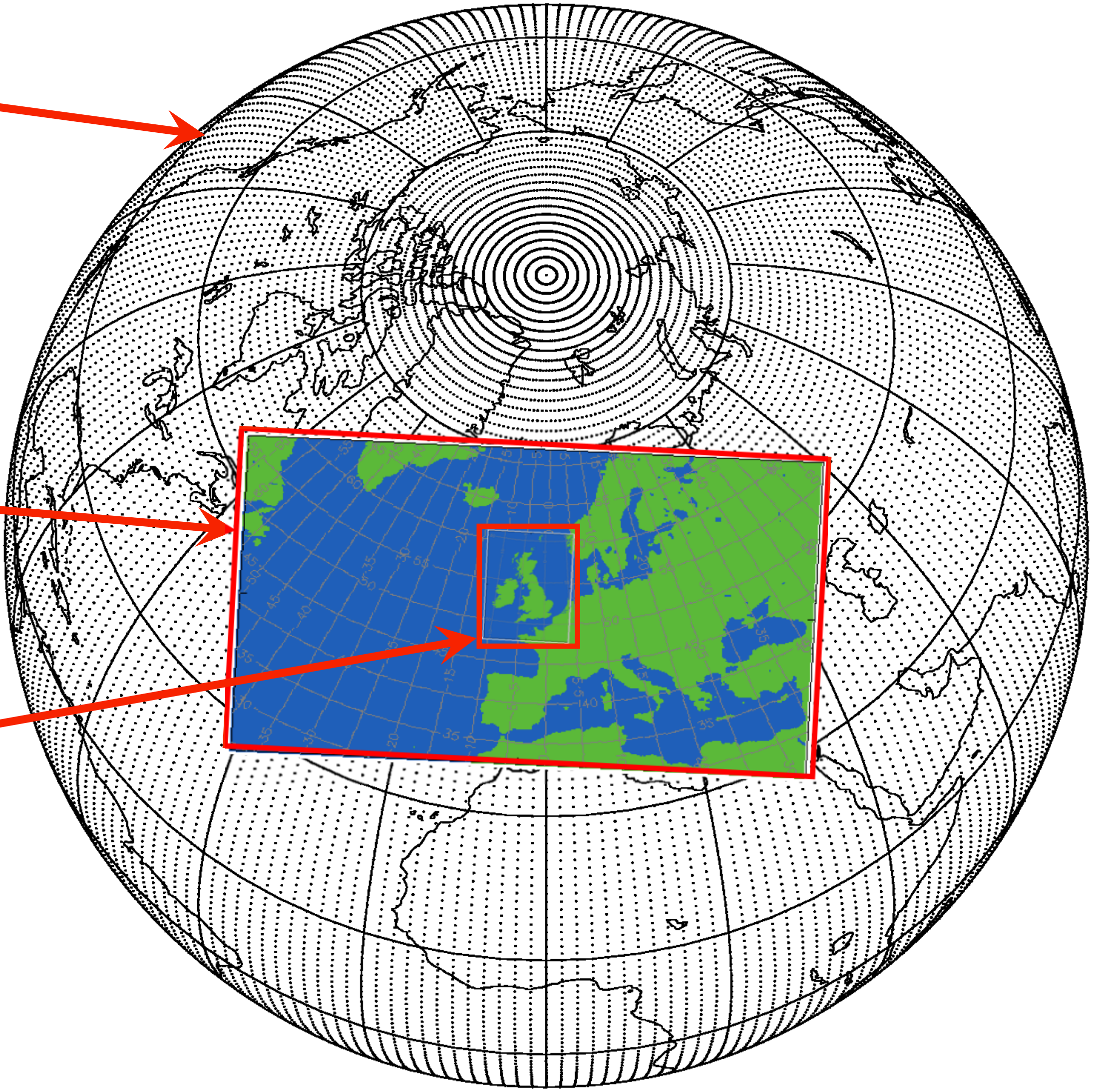
4th EnKF Workshop, 7 April 2010

Operational NWP Models: Jun 2010

Global
- ➢25Km 70 L
- ➢4DVAR (90km inner-loop)
- ➢60 hour forecast twice/day
- ➢144 hour forecast twice/ day
- ➢*+24member EPS at 60km*

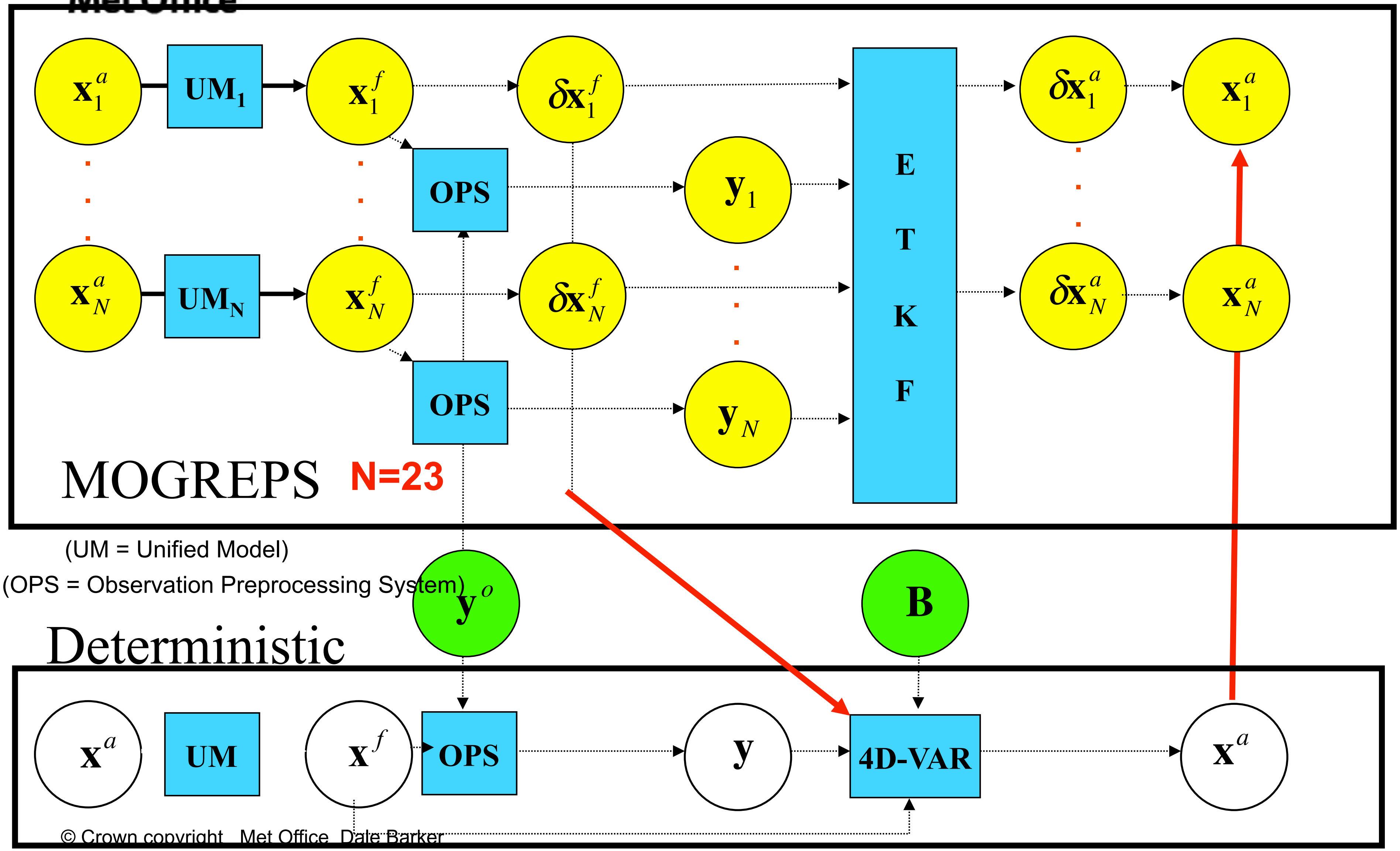
NAE
- ➢12Km 70 L
- ➢4DVAR (36km inner-loop)
- ➢60 hour forecast
- ➢ 4 times per day
- ➢ *+24member EPS at 18km*

UK4
- ➢4km 70 L
- ➢3DVAR
- ➢36 hour forecast
- ➢ 4 times per day



Coupled Deterministic/Ensemble NWP

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



(UM = Unified Model)

(OPS = Observation Preprocessing System)

Met Office

Observation Volumes in 6 hours (20/10/08)

Increased resolution=>Increased Obs usage

Category	Count	% used	Category	Count	% used
TEMPs	637	99%	Satwinds: JMA	26103	4%
PILOTs	307	99%	Satwinds: NESDIS	142478	3%
Wind Profiler	1355	39%	Satwinds: EUMETSAT	220957	1%
Land Synops	16551	99%	Scatwinds: Seawinds	436566	1%
Ships	3034	84%	Scatwinds: ERS	27075	2%
Buoys	8727	63%	Scatwinds: ASCAT	241626	4%
Amdars	64147	23%	SSMI/S	532140	1%
Aireps	7144	12%	SSMI	698048	1%
GPS-RO	776	99%	ATOVS	1127224	3%
			AIRS	75824	6%
			IASI	80280	3%

Computational Cost (4 nodes IBM Power 6)

Component	4D-Var Inner-Loop	MOGREPS (N=24)	Nonlinear model
Resolution (km)	~120km	~90km	~40km
Wall-Clock (W mins)	7*	75	48
Forecast Range (F hrs)	6	72	168
W/F (mins/hr)	1.17	1.04	0.29

- Conclude 4D-Var / 24 member Ensemble Filter costs similar.

- Operational advantages of ensemble DA in cost-comparison:
  - Operational ensemble partly 'paid for' (but still costs for research/trialling).
  - Ensemble forecasts can be run ahead of obs. cut-off, 4D-Var usually waits.

- Scalability of 4D-Var on next generation machines (10^5 processors) a concern.

* Note ECMWF/NCEP equivalent times are 1hr/30mins respectively!

Hybrid Variational/Ensemble DA Via The 'Alpha Control Variable' Method

- Vector of Ensemble Perturbations $\delta\mathbf{X}_f = \left(\delta\mathbf{x}_{f1}, \delta\mathbf{x}_{f2}, \ldots, \delta\mathbf{x}_{fN}\right)$

- Hybrid analysis increment defined as

$$\delta\mathbf{x}_0 = \delta\mathbf{x}_{clim} + \delta\mathbf{x}_{flow-dep} = \mathbf{B}^{1/2}\mathbf{v} + \delta\mathbf{X}_f \circ \mathbf{a}$$

- Ensemble weights **a** constrained by an additional cost-function,

$$J = \frac{W_b}{2}\delta\mathbf{x}_0^T\mathbf{B}^{-1}\delta\mathbf{x}_0 + \boxed{\frac{W_\alpha}{2}\mathbf{a}^T\mathbf{A}^{-1}\mathbf{a}} + \frac{1}{2}\sum_{i=0}^{n}\left[\mathbf{H}_i\delta\mathbf{x}(t_i) - \mathbf{d}_i\right]^T\mathbf{R}_i^{-1}\left[\mathbf{H}_i\delta\mathbf{x}(t_i) - \mathbf{d}_i\right]$$

- Variance conservation implies $\frac{1}{W_b} + \frac{1}{W_\alpha} = 1$

- W_b=1 is standard 3/4D-Var. W_b=0 fully ensemble covariance (e.g. Liu et al 2008, Buehner et al 2009). **Hybrid is the space in-between!**

Comments on The ACV Method

- ‘Incremental, balance-aware’ covariance localization is trivial, e.g.

$$\delta\psi_0 = \delta\psi_{c\lim} + \delta\psi_f \circ \mathbf{a} \qquad \delta\chi_{u0} = \delta\chi_{uc\lim} + \delta\chi_{uf} \circ \mathbf{a}$$

- Covariance **A** equivalent to model-space covariance localization (Lorenc 2003).

- Localization defined adaptively or empirically, e.g.

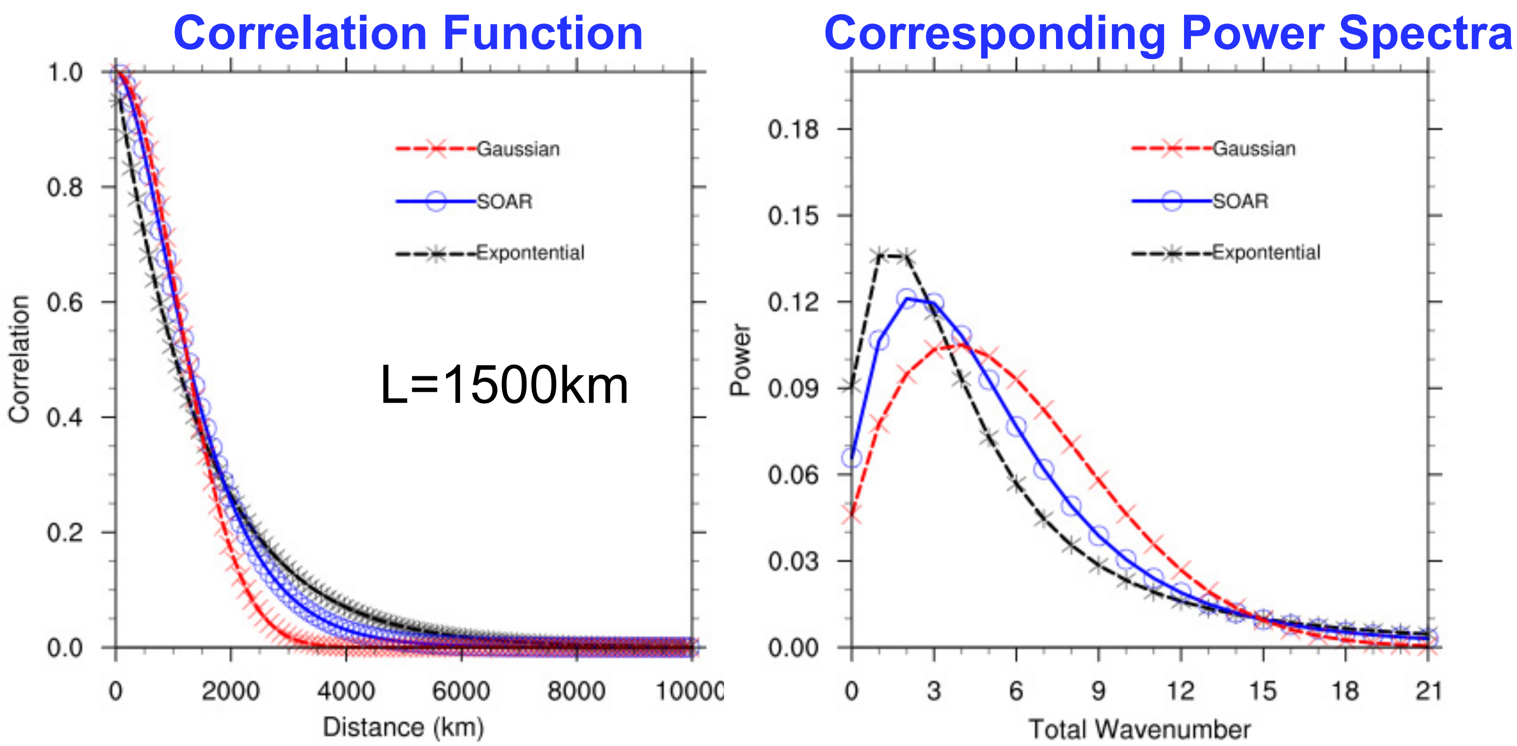
$$\mathbf{a} = \mathbf{A}^{1/2}\alpha \equiv \boldsymbol{\sigma}_\alpha^2 \mathrm{A}_v \mathrm{A}_h \alpha$$

- **A**=**I** implies no localization. Only need one scalar per ensemble member.

- Convenient to use standard control variable transforms for localization operators A_v, A_h, but not essential.

A_h: Horizontal Covariance Localization

- Example: Truncated power spectrum from empirical correlation with scale *L:*

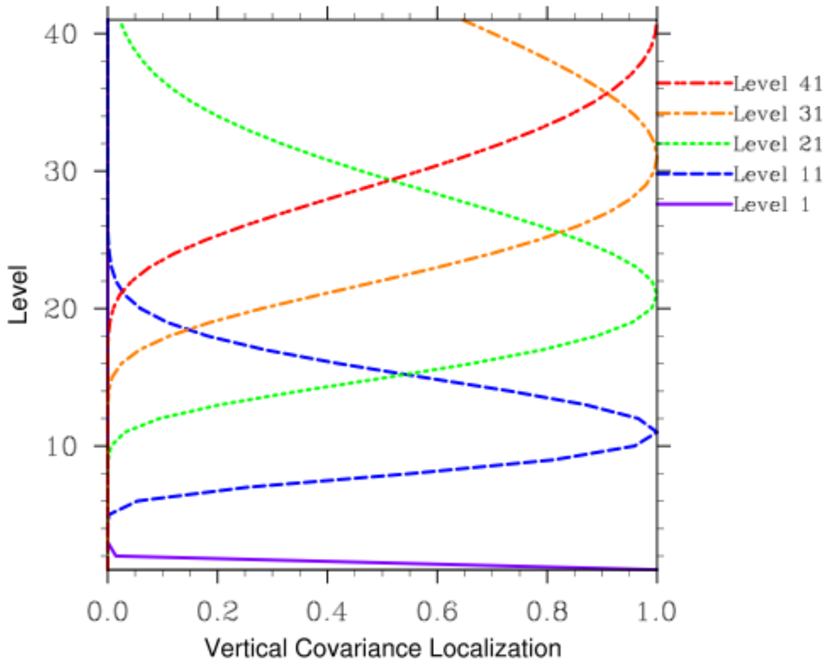


A_v: Vertical Covariance Localization

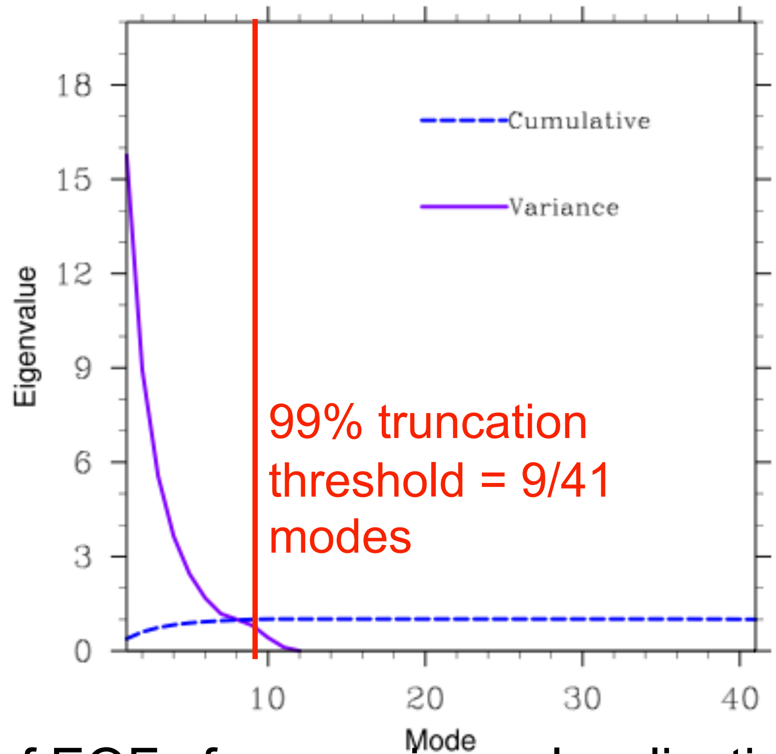
Example: Gaussian with level-dependent localization scale:

$$\rho(k-k_c)=\exp\left[-\left(k-k_c\right)^2/L_c^2\right]$$
$$L_c=20k_c/41$$

Correlation Function

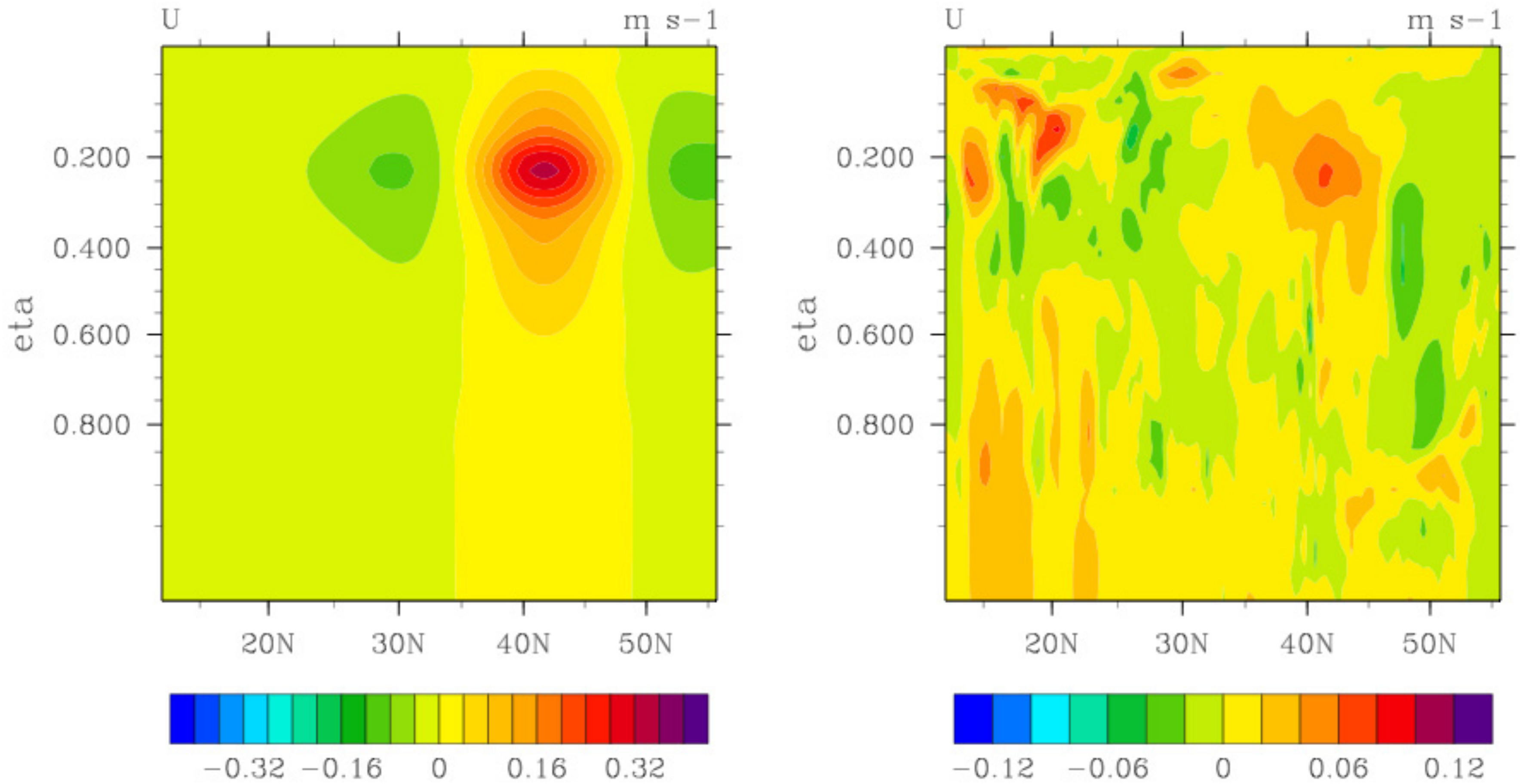


Eigenvalues



Conclusion: 75% data compression via use of EOFs for covariance localization

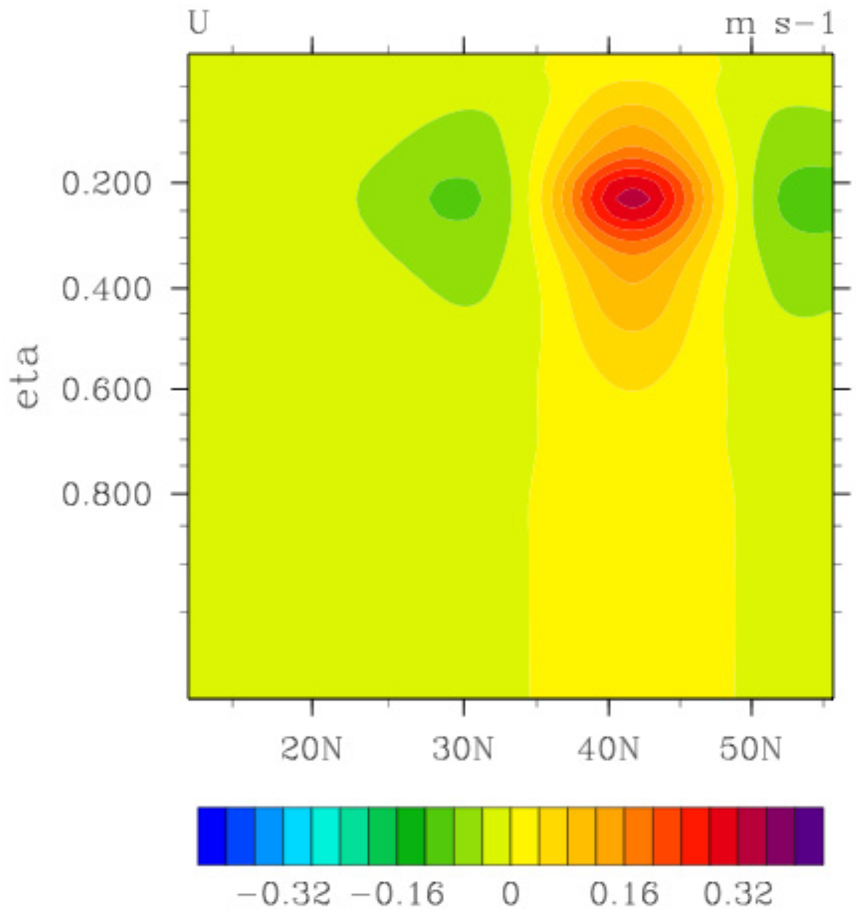
ACV Single ob test
(250hPa u, O-B=1m/s, sigma_o=3.3m/s)



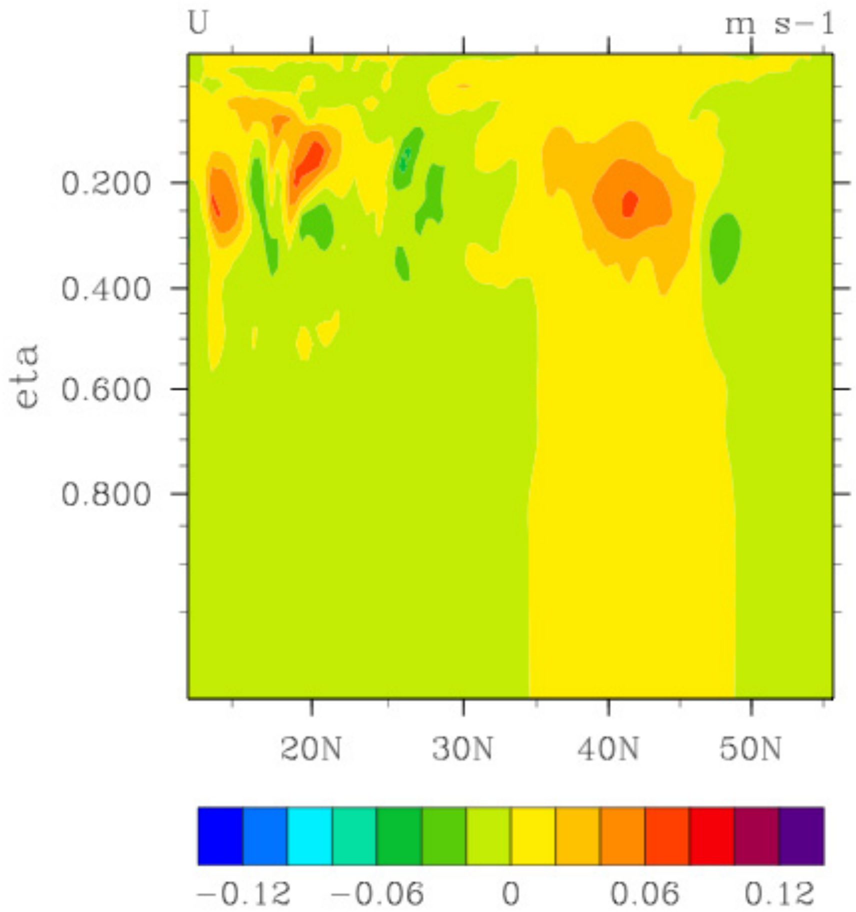
$W_b=1$ $W_b=0$, $\mathbf{A}=\mathbf{I}$

ACV Single ob test (250hPa u, O-B=1m/s, sigma_o=3.3m/s)

Impact of Vertical Localization



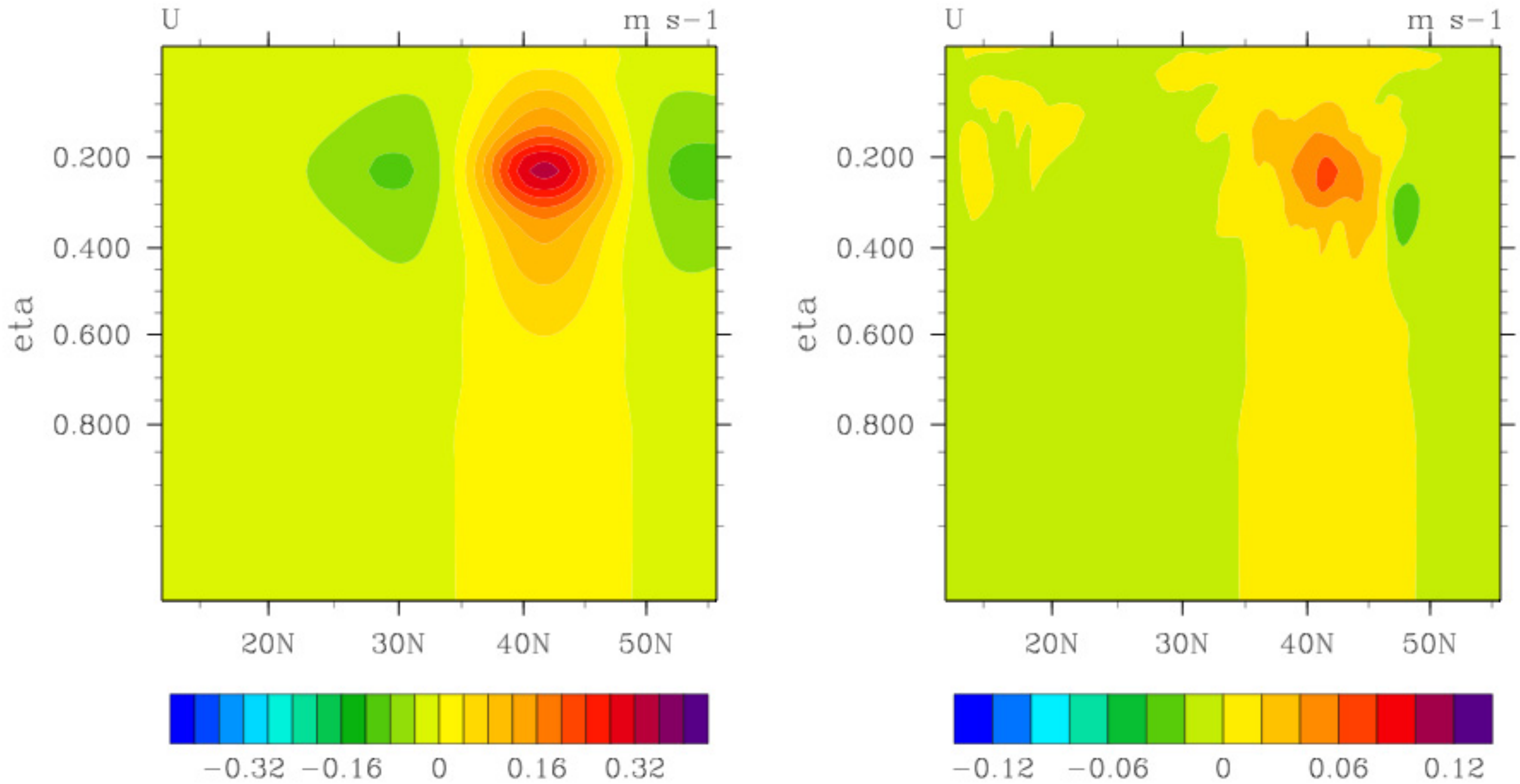
$W_b=1$



$W_b=0$, $\mathbf{A}=A_v$

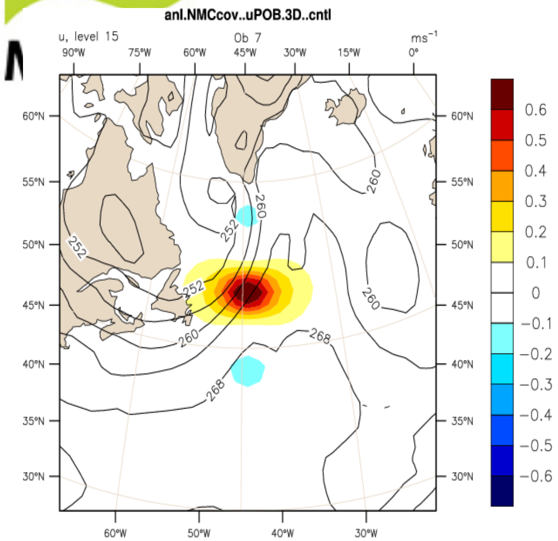
ACV Single ob test
(250hPa u, O-B=1m/s, sigma_o=3.3m/s)

Impact of Vertical + Horizontal Localization

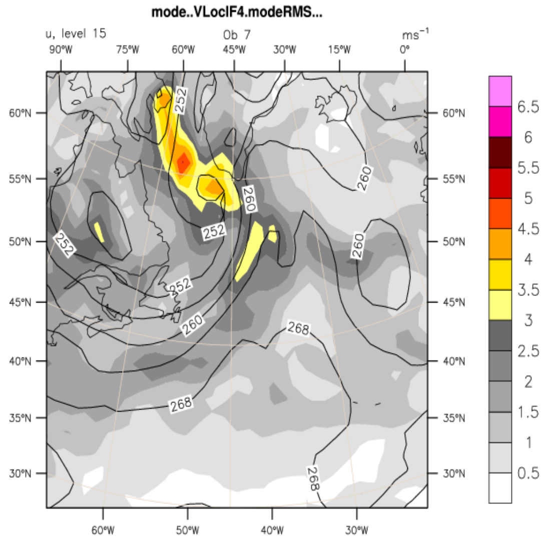


$W_b=1$ $\qquad$ $W_b=0$, $\mathbf{A}=A_vA_h$

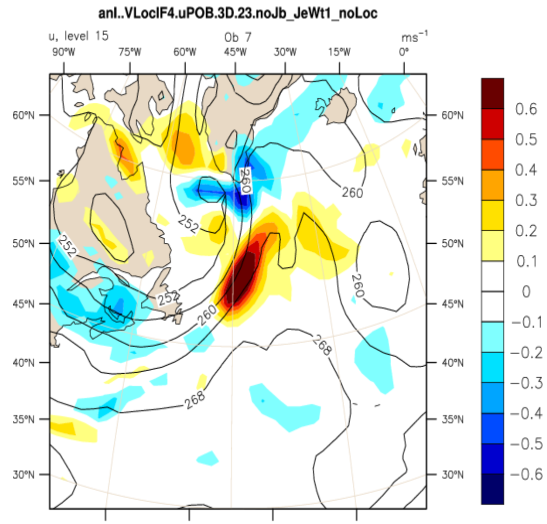
Single Observation Tests: $\langle u\ u \rangle_h$



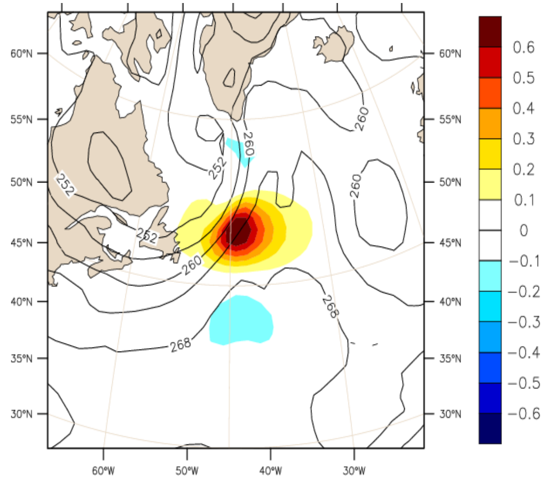
3D-Var Increment



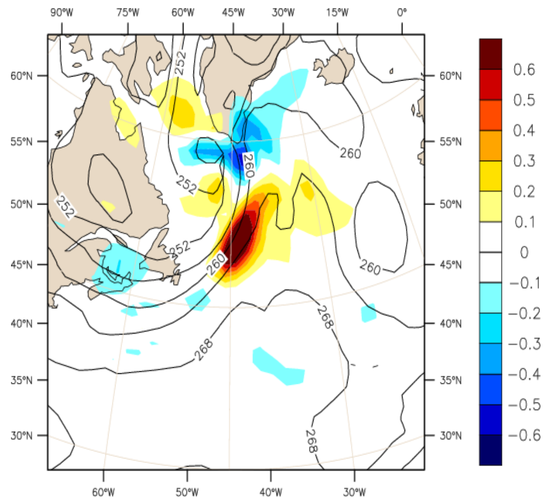
Ensemble Spread



Ensemble Increment, **A**=**I**

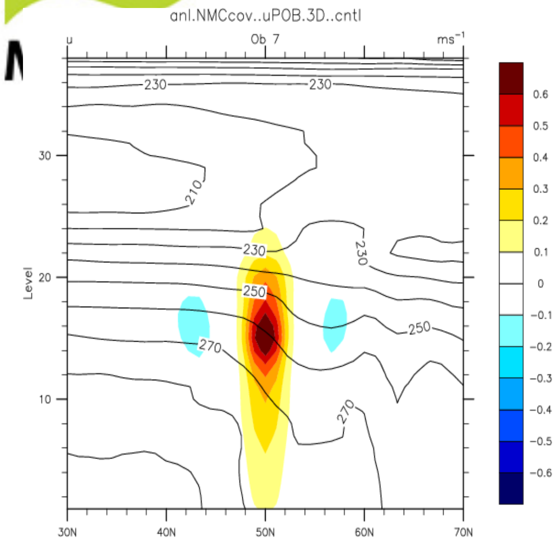


4D-Var Increment (middle of 6hr time window)

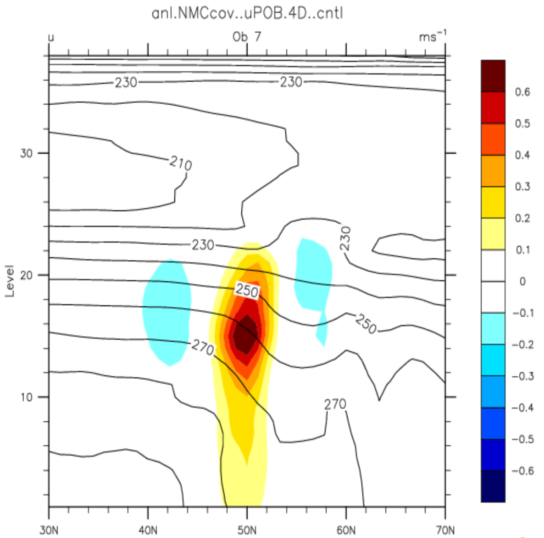


Ensemble Increment, **A**=A_h

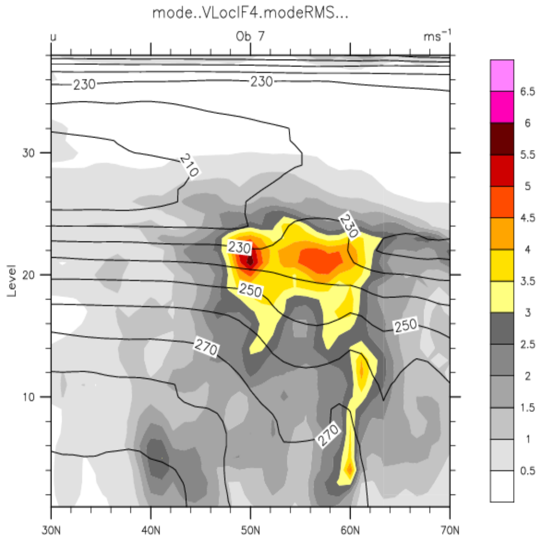
Single Observation Tests: $<u\ u>_v$



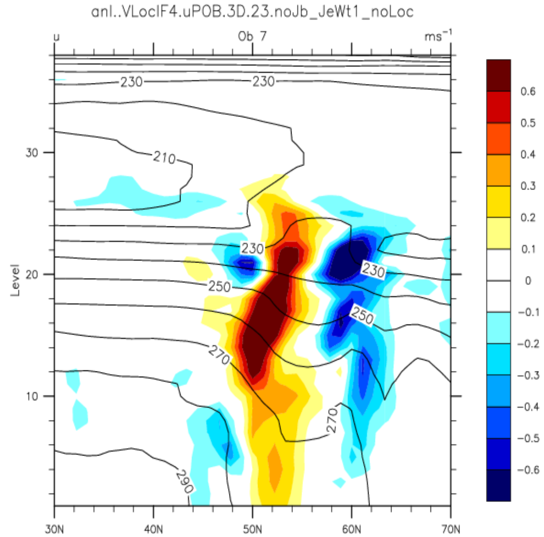
3D-Var Increment



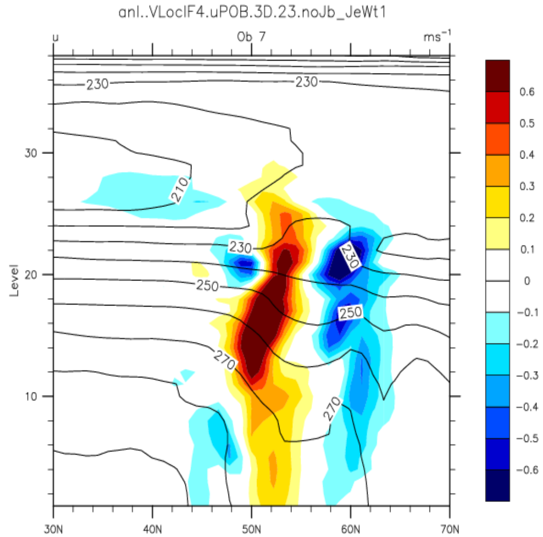
4D-Var Increment (middle of 6hr time window)



Ensemble Spread



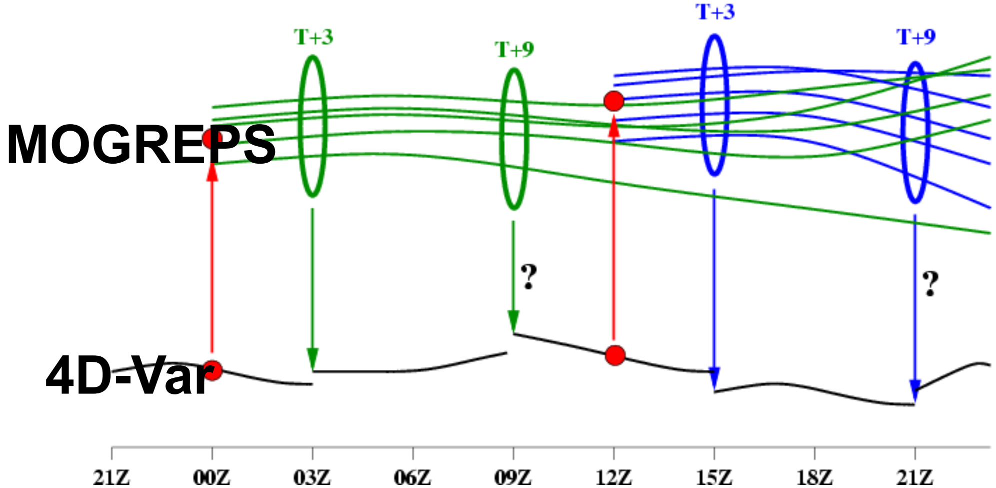
Ensemble Increment, **A**=I



Ensemble Increment, **A**=A_h

3/4D-Var Hybrid Trial Configurations (Adam Clayton)

- **Global** **~90km L38** model
  - Incremental 3/4D-Var (~120km)
  - 24m ETKF ensemble (~90km)

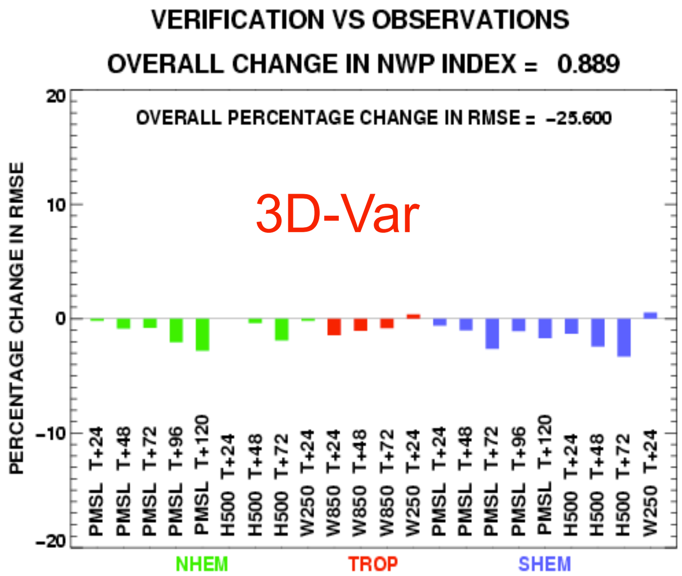


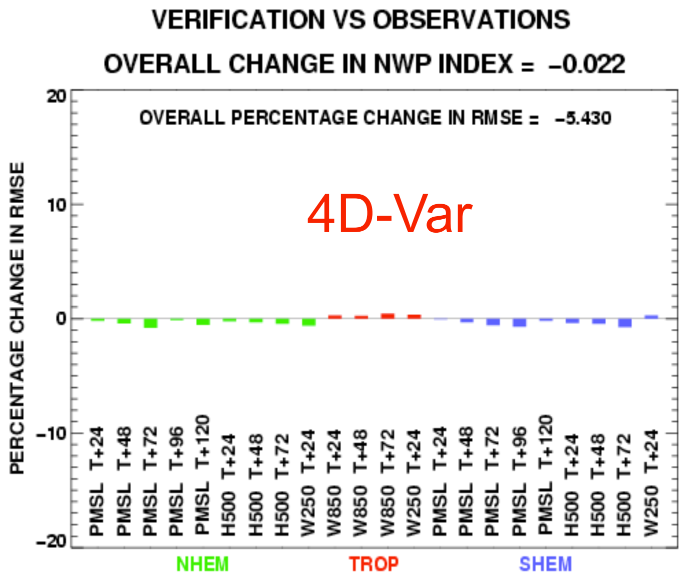
- **Trial period**: 5 – 31st May 2008

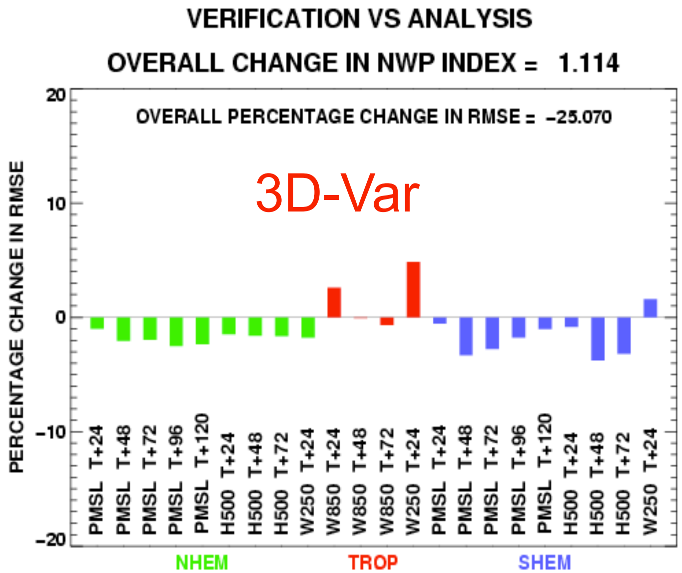
- **Observations:** Surface, Scatwind, Satwind, Aircraft, Sonde, ATOVS, SSMI, AIRS, GPSRO, SSMIS, IASI

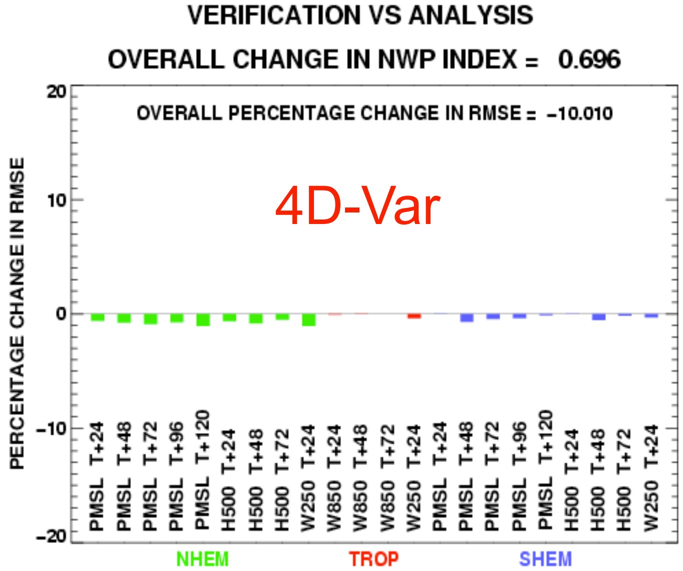
- **Hybrid configuration:**
  - Localization: 1500km (T10) Gaussian. Horizontal only. 'Balance-aware'.
  - Climatological/Ensemble Covariances Given Equal Weight (W_b=2).
  - No tuning as yet!

May 2008 3/4D-Var Trial Results









Hybrid Results Summary

Met Office

- Hybrid demonstrated with model-space, 3D, ‘balance-aware’ localization.

- Hybrid shown to compensate for rank-deficiency in ensemble covariance.

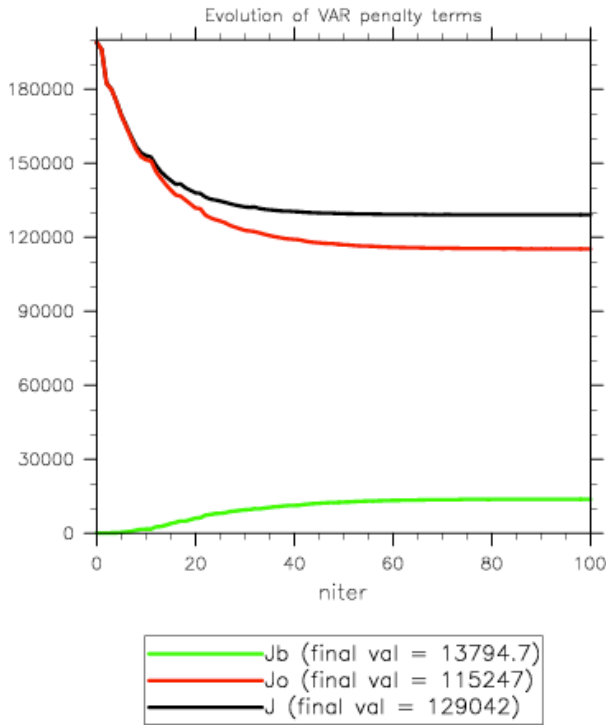
- First tests of impact of hybrid in a full observation 3/4D-Var: Positive benefit vs. 3D-Var mode, neutral in 4D-Var. Pleasing result given no tuning yet.

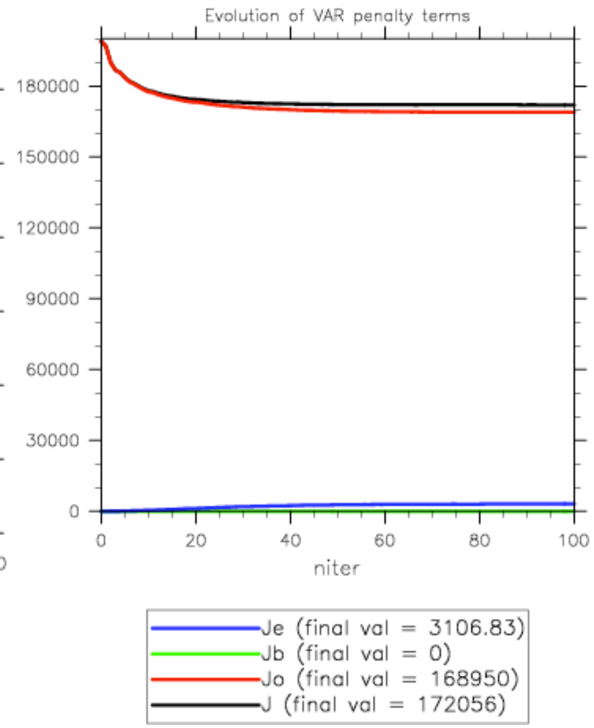
- Putting things in perspective:

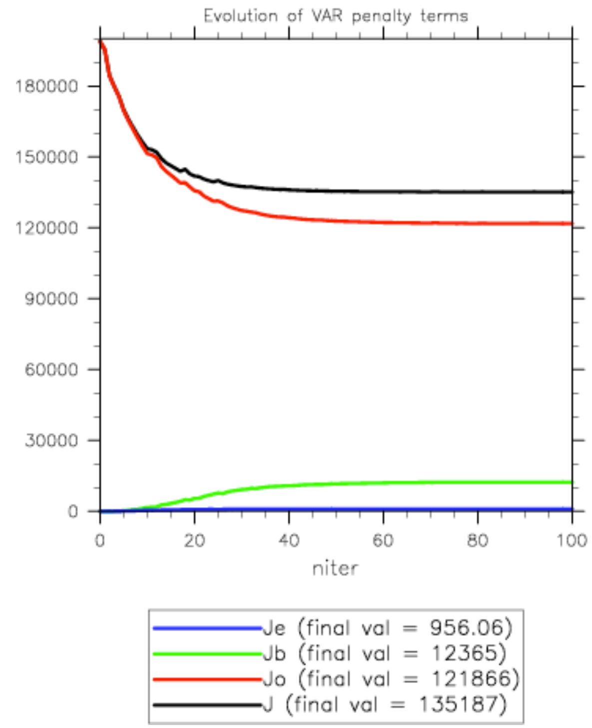
Impact on NWP Index	3D-Var Hybrid vs. 3D-Var	4D-Var Hybrid vs. 4D-Var	4D-Var vs. 3D-Var
Verification vs. Obs	+0.78%	-0.39%	+2.7%
Verification vs. Analysis	+0.94%	+1.33%	+1.3%

Convergence Of 3D-Var Cost Function

Met Office







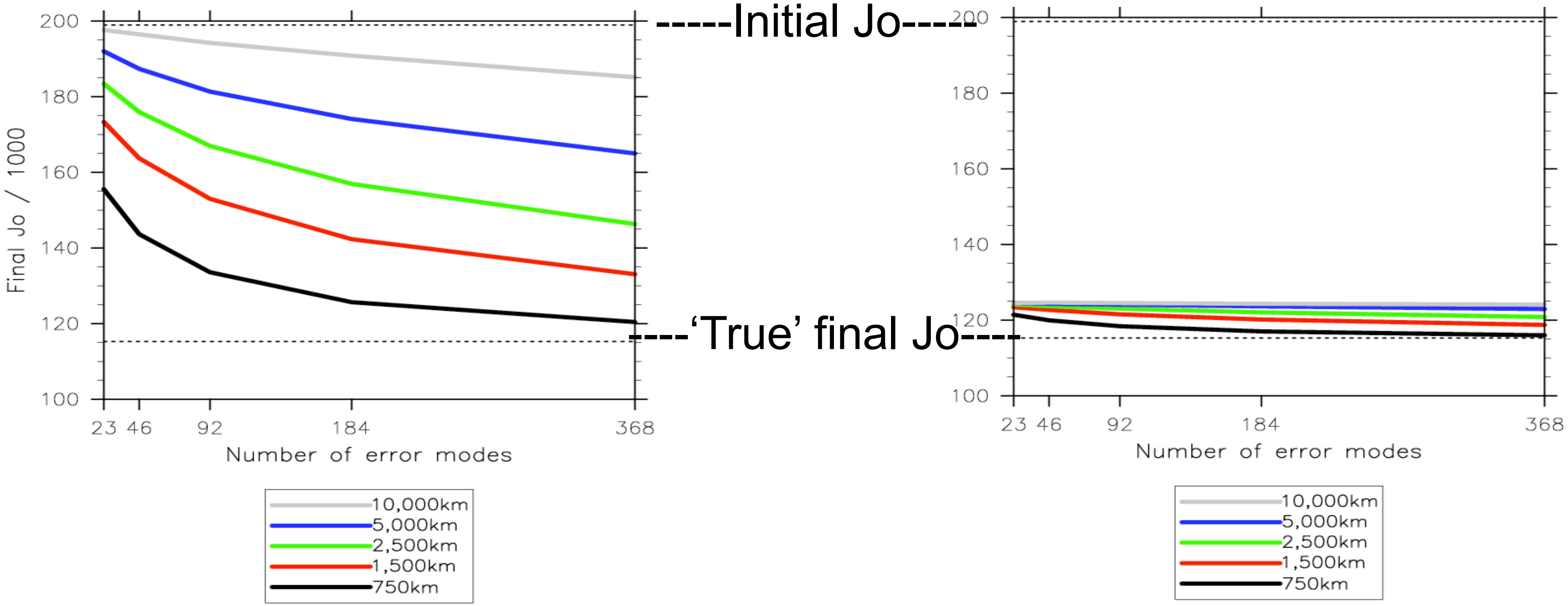
Conclusion: Ensemble covariances from 24 member ETKF are significantly rank-deficient, and underfitting observations. Hybrid helps.

Estimation of Ensemble Sampling Error

Method: Simulate ensemble by sampling climatological **B**. Study effect of ensemble size, localization, hybrid on minimization.

Pure Ensemble Covariance

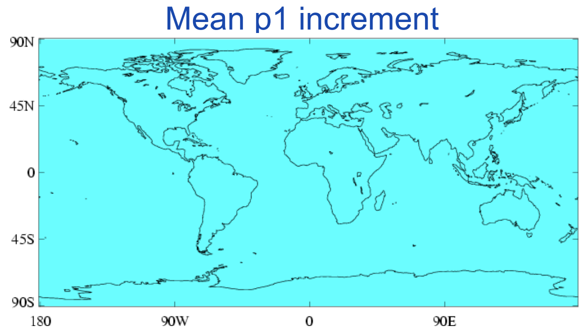
Hybrid Covariance

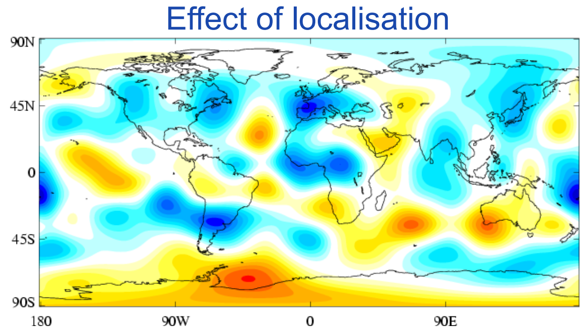


Conclusion: Pure ensemble covariance significantly rank-deficient, even with larger ensembles, and shorter localization scales. Hybrid vital!

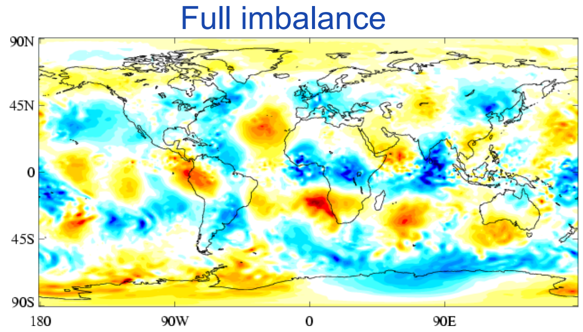
Balance and localisation

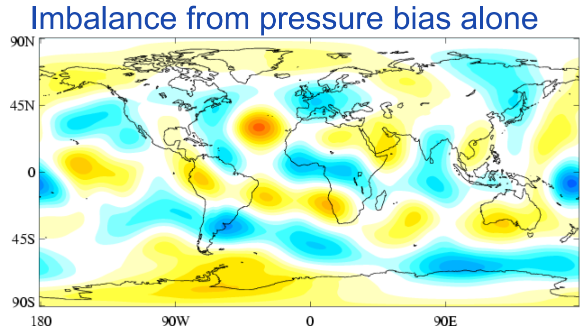
- Cause is large-scale increments within the MOGREPS modes, which get projected onto the localisation scale:





- Mean p1 increment alone explains significant proportion of imbalance:

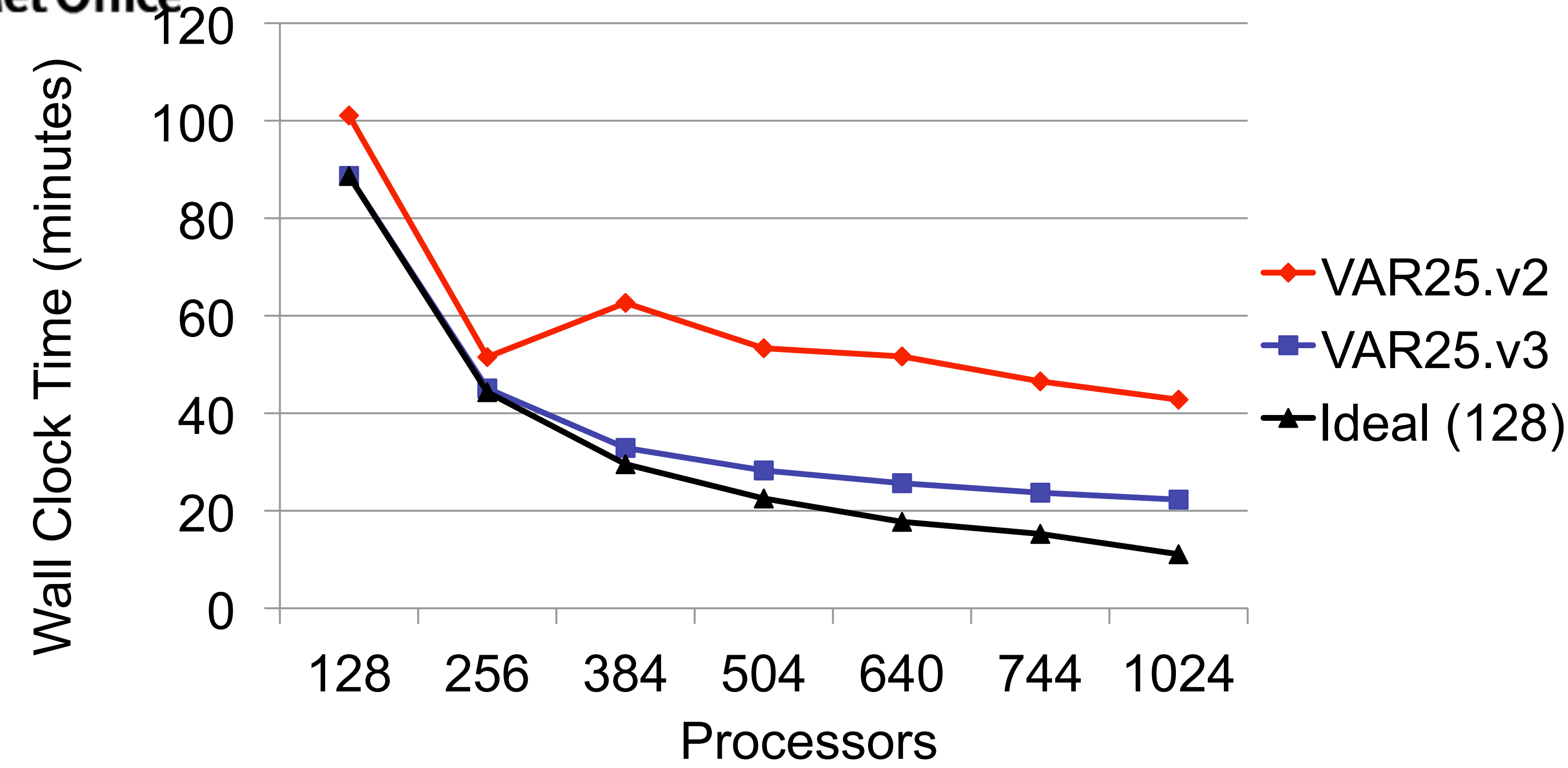




- Fix: Remove scales larger than the localisation scale from the error modes.

4D-Var Scalability (N216/~75km Inner Loop)

Met Office



- Results for single outer loop.
- Operational wall-clock time is ~10 minutes.
- Algorithmic changes required to maintain affordability after 2012 upgrade.

Data: Rick Rawlins

4D-Var Algorithmic Changes

- Algorithmic changes 1: Hessian preconditioning, GCR mods, etc.
  - Default N216 4D-Var = ~21mins
  - Optimized N216 4D-Var = ~13mins.
- Algorithmic changes 2: 'Multigrid' 4D-Var:
  - T-60mins: 4D-Var (120km), T-30mins: 4D-Var (75km)
  - Finally at T+0: 4D-Var (50km – <10 iterations)
  - Permits affordable N320 4D-Var for 2011 computer upgrade.
  - After that, all bets are off!

Met Office

Operational NWP Models: Sep 2010

Global

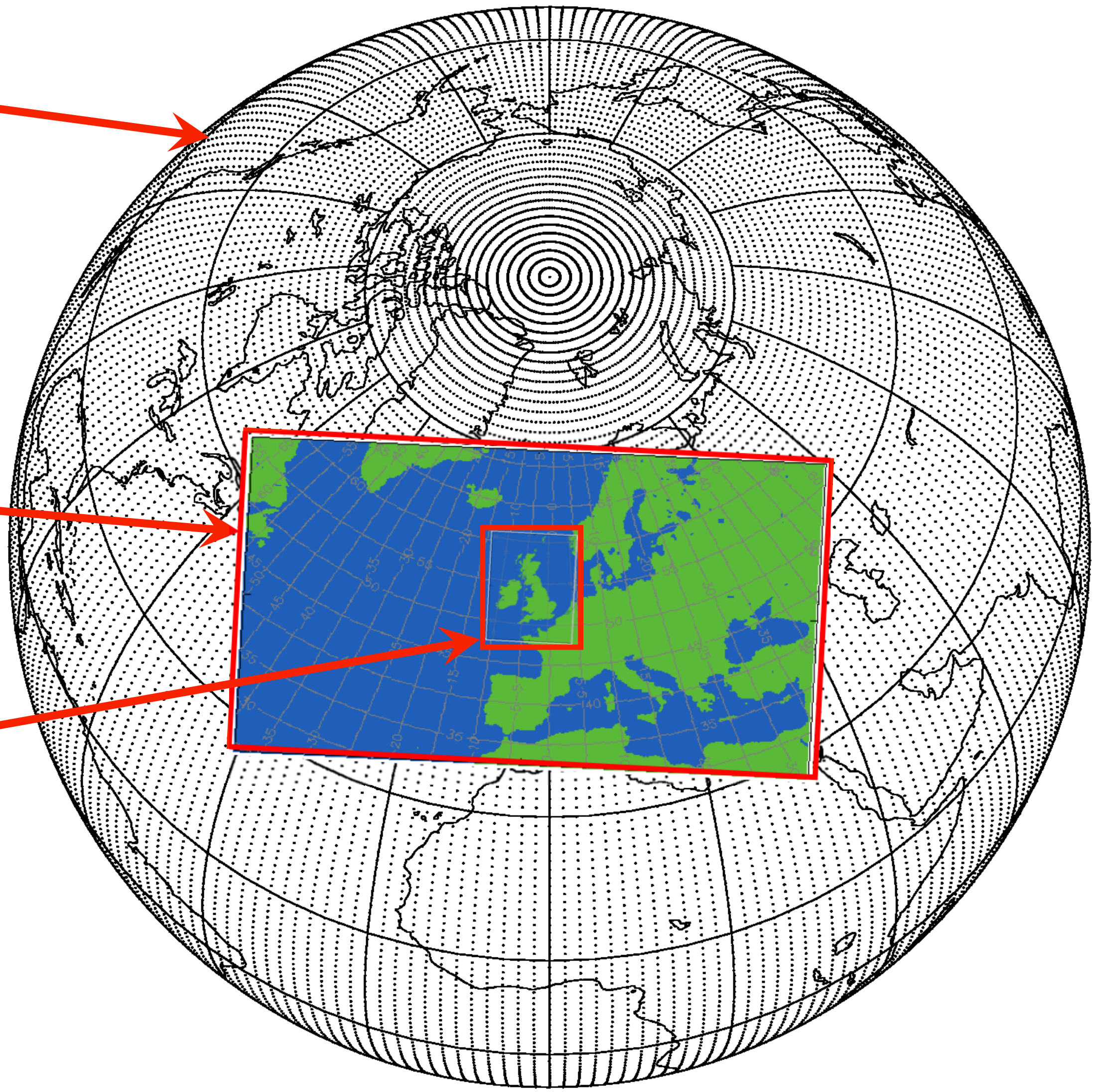
- 25Km 70 L
- Hybrid 4DVAR (75km IL)
- 60 hour forecast twice/day
- 144 hour forecast twice/ day
- *+24member EPS at 60km*

NAE

- 12Km 70 L
- 4DVAR (36km inner-loop)
- 60 hour forecast
- 4 times per day
- *+24member EPS at 18km*

UKV

- 1.5km 70 L (Variable resolution)
- 3DVAR (hourly)
- 36 hour forecast
- 4 times per day



Hybrid Summary

- MetO hybrid implementation similar to WRF (ACV method) but first application in 4D-Var and full observation system.
- Results indicate hybrid improves 3D-Var. Neutral in 4d-Var mode. Note: no tuning so far.
- Studies indicated localization imbalance remains a problem – hybrid needs to be scale-selective.
- MetO short/medium term strategy for coupled variational/ensemble DA but may need to look at more significant changes long-term (>5 yrs).

Questions?