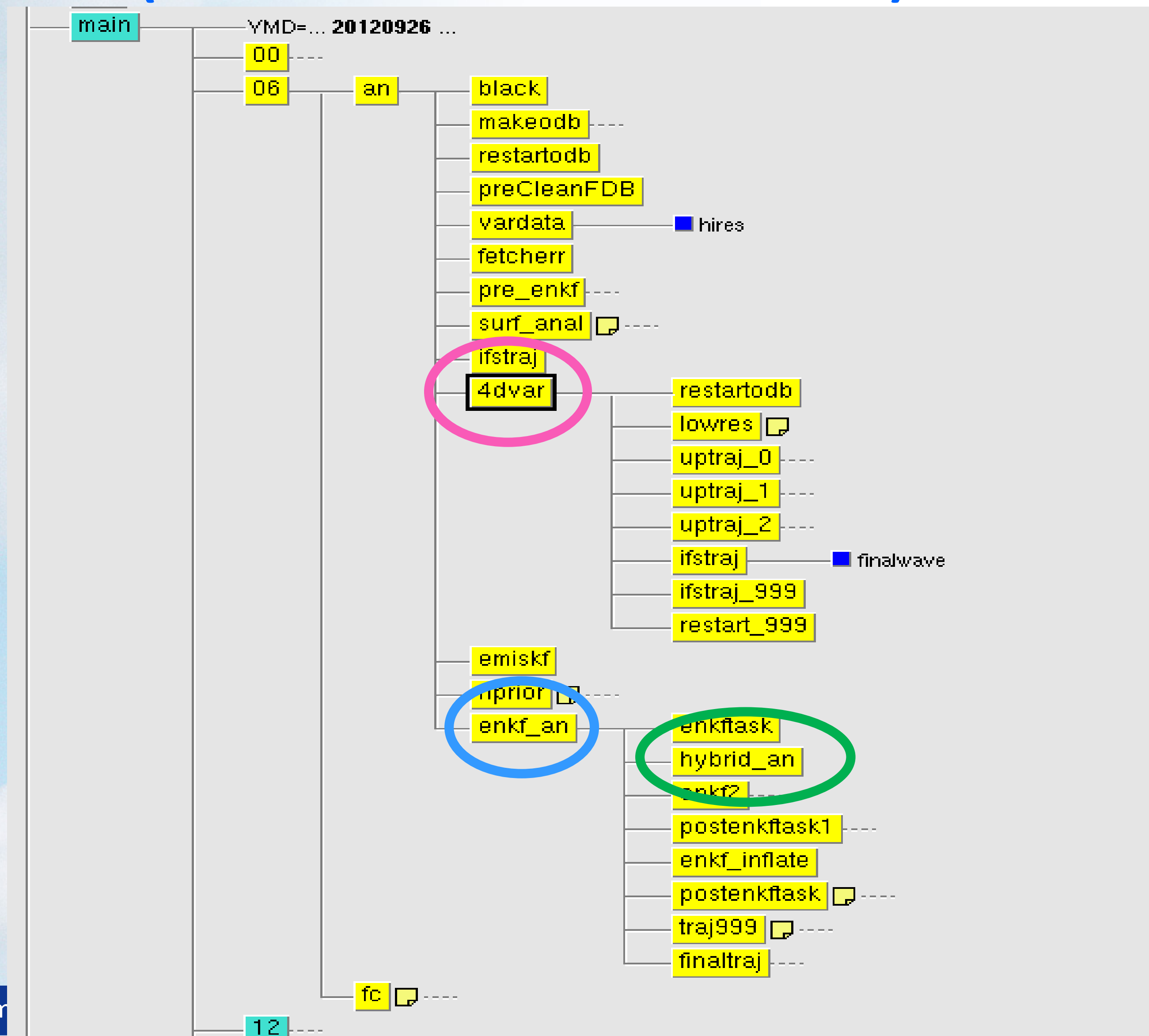
Var and EnKF data assimilation at ECMWF (and combinations thereof!)

Massimo Bonavita

Ack.: Mats Hamrud and Erik Andersson

Variational and EnKF assimilation systems at ECMWF (and combinations thereof!)



Var and EnKF DA systems at ECMWF

What are the main goals of an operational DA system?

1. Coherent analyses of the state of the atmosphere (+ ocean and land surface)
2. Improve the accuracy of weather forecasts
3. Monitoring atmospheric constituents and pollution
4. Producing consistent climate re-analysis: documenting climate change
5. Provide estimates of uncertainty (analysis error) that can be used to initialise EPS

Erik Andersson, 2008

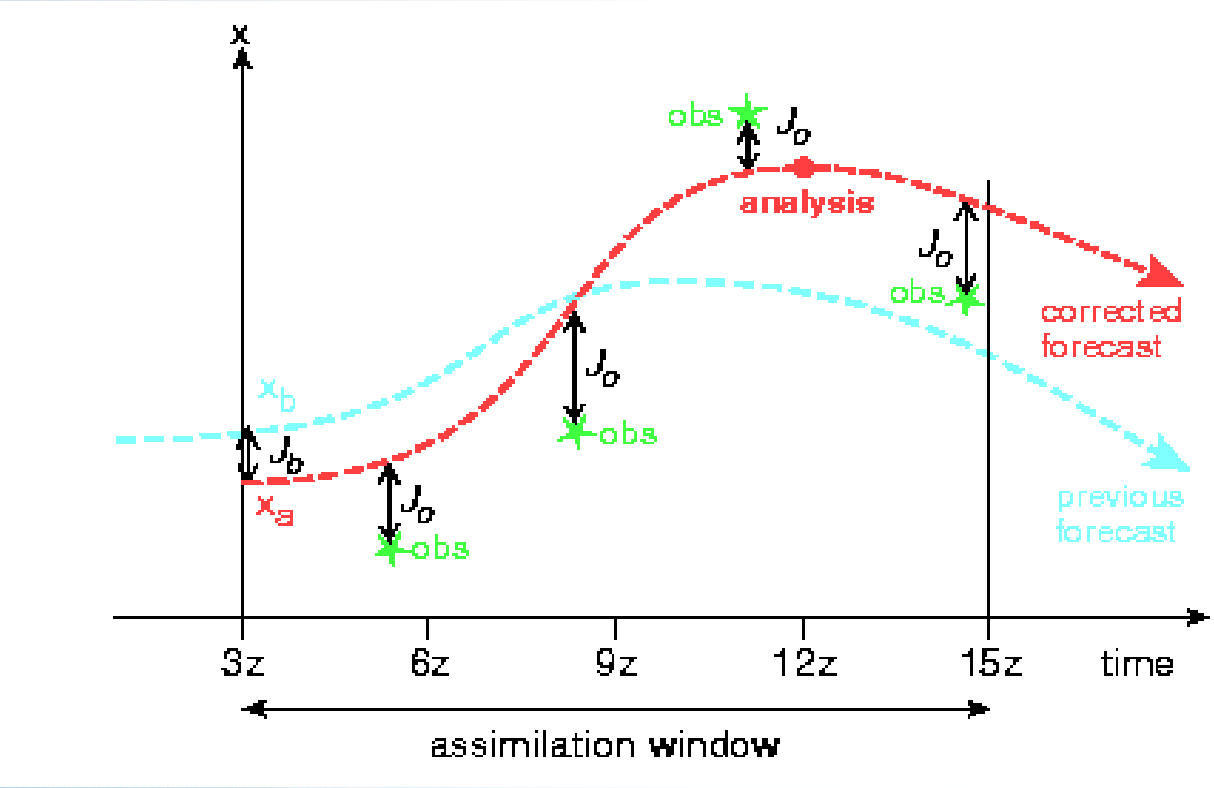
Var and EnKF DA systems at ECMWF

What are the main goals of an operational DA system?

1. Coherent analyses of the state of the atmosphere (+ ocean and land surface)
2. Improve the accuracy of weather forecasts
3. Monitoring atmospheric constituents and pollution
4. Producing consistent climate re-analysis: documenting climate change
5. Provide estimates of uncertainty (analysis error) that can be used to initialise EPS

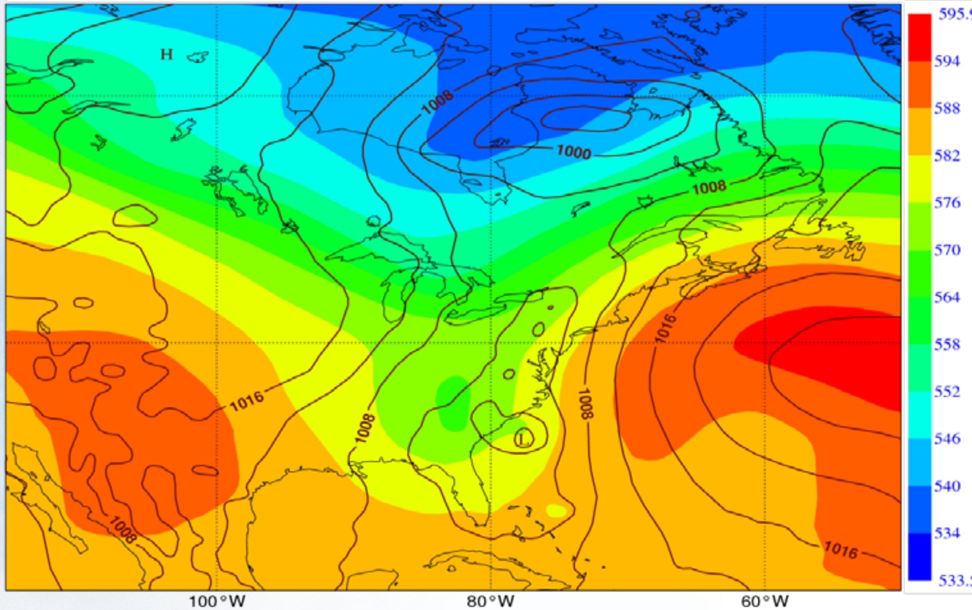
It is not likely that one single methodology will satisfy all the demands of all these application areas.
Variational, ensemble and other approximations of the *Kalman Filter* will all remain in our communal toolbox for many years to come

4DVar

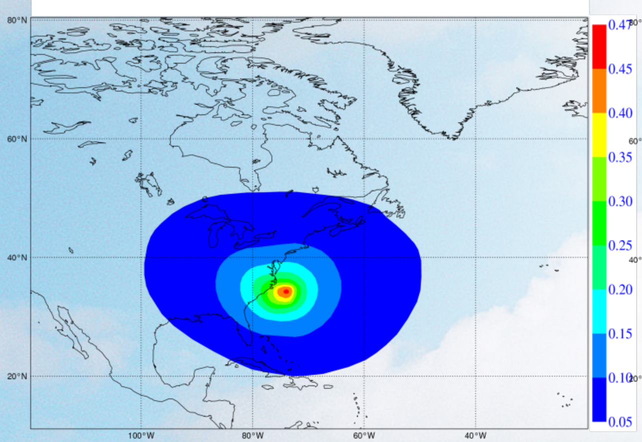


- Incremental, Strong Constraint 4DVar (Courtier et al., 1994; Veersé and Thépaut, 1998) has been the workhorse of ECMWF DA for almost 20 years
- Based on Optimal Control Theory, tries to find the ML trajectory given $\mathbf{x}_b$, $\mathbf{y}$ and their 2nd order error statistics ($\mathbf{B}$, $\mathbf{R}$)
- 4Dvar state estimate is equivalent, for the same $\mathbf{x}_b$, $\mathbf{y}$, $\mathbf{B}$, $\mathbf{R}$, to the Kalman filter solution at the end of the assimilation window (Fisher et al., 2005).

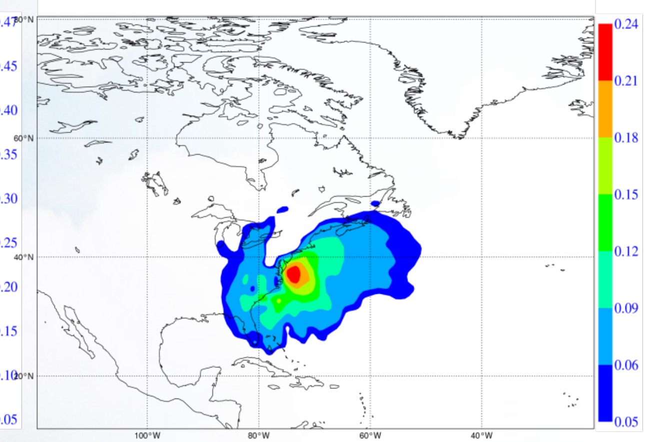
MSLP and 500 hPa Z (shaded) background fcst



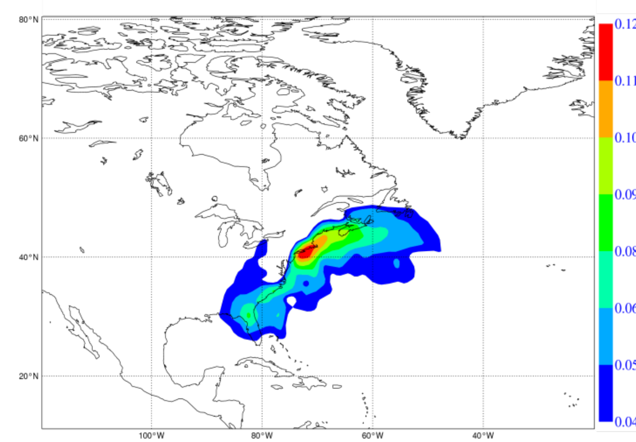
Temperature analysis increments for a single temperature observation at the start of the assimilation window: $x^a(t)-x^b(t) \approx MP^bM^TH^T(y-Hx)/(\sigma_b^2 + \sigma_o^2)$



t=+0h



t=+3h



t=+9h

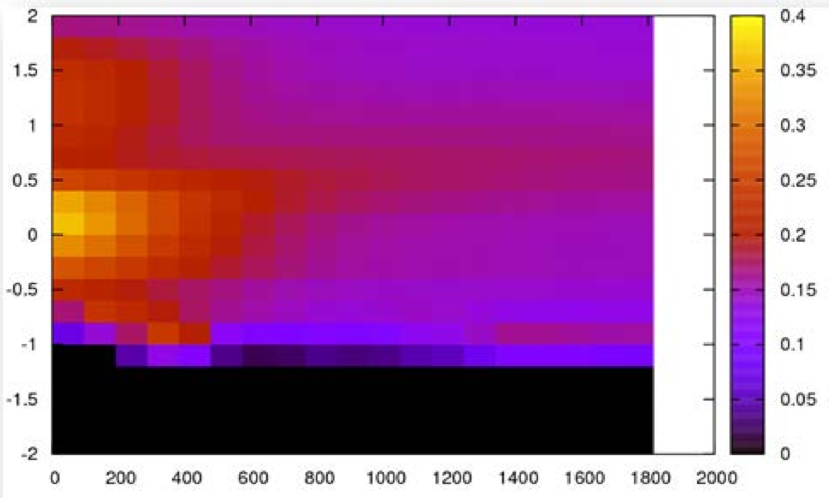
EnKF

- Monte Carlo implementation of the Kalman Filter (Evensen, 1994; Burgers et al., 1998)
- Basic idea: Sample error covariances from ensemble of background fields
- Main selling points: Algorithmic simplicity (No need of complex TL and ADJ model and observation operator codes), non-linear evolution of background errors (but Gaussian update!), ensemble of initial states for EPS, explicit representation of model/system error
- Main drawbacks: Sampling issues due to limited affordable ensemble size, covariance localization, imbalance

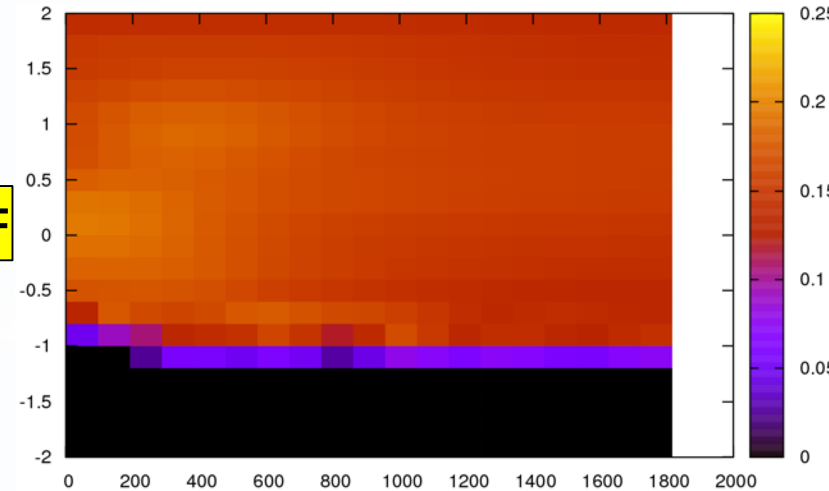
EnKF

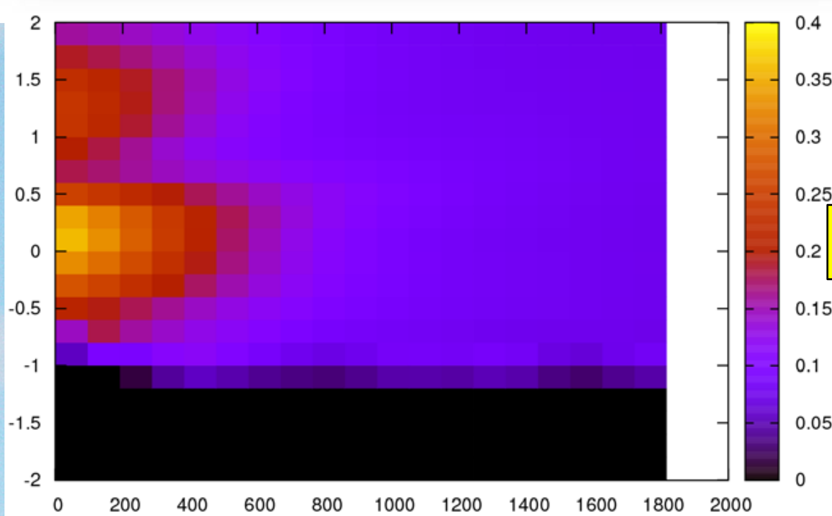
T(500hPa) – BT(AMSUA) abs. correlation

u(500hPa) – BT(AMSUA) abs. correlation

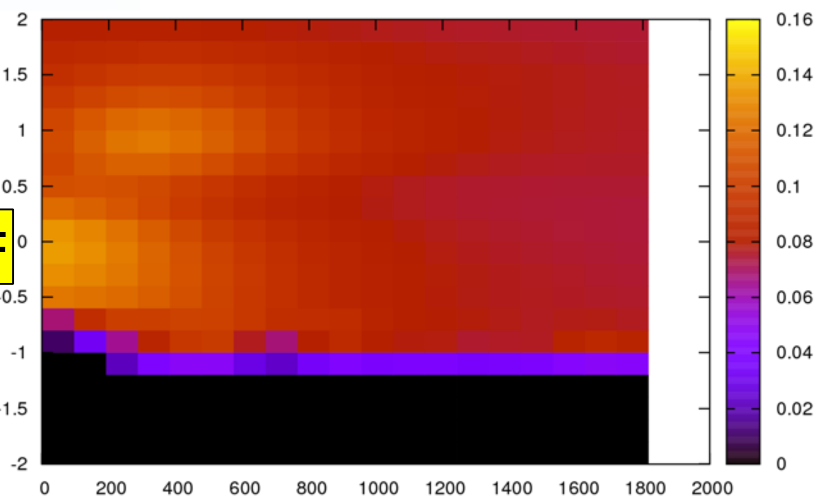


60 EnKF





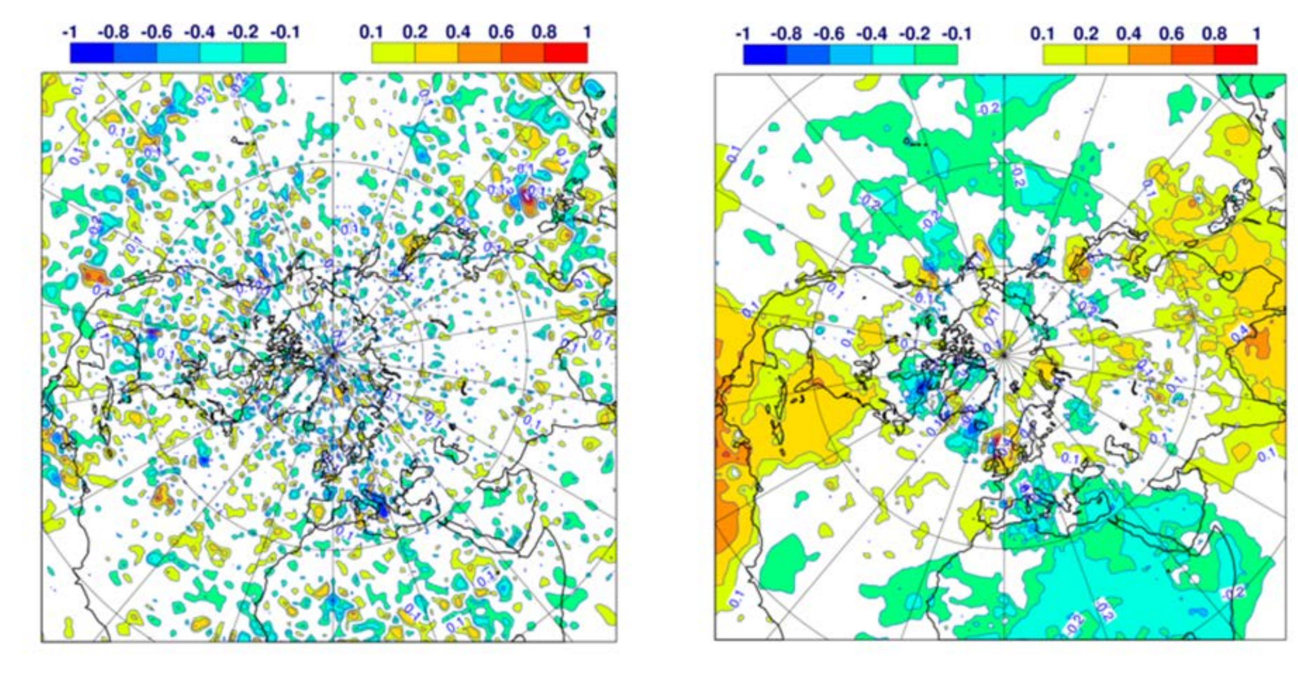
120 EnKF



EnKF

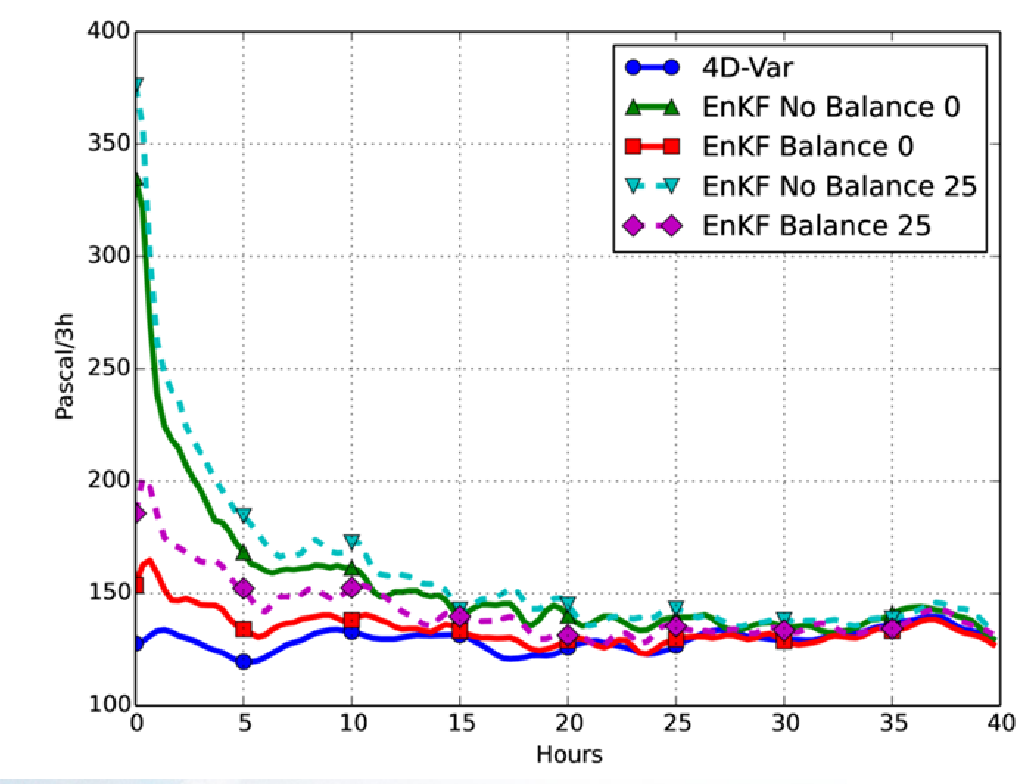
- Balance issues appear to be mainly due to localization in the vertical
- Can be improved by "divergence adjustment" procedure (Hamrud et al, 2015): Analyse surface pressure tendency as additional state vector variable, and adjust column divergence to match

$$\frac{\partial p_s}{\partial t} = -\int_0^1 \nabla \cdot \left(V \frac{\partial p}{\partial \eta} \right) d\eta$$



EnKF

- Balance issues appear to be mainly due to localization in the vertical
- Can be improved by "divergence adjustment" procedure (Hamrud et al, 2015): Analyse surface pressure tendency as additional state vector variable, and adjust column divergence to match



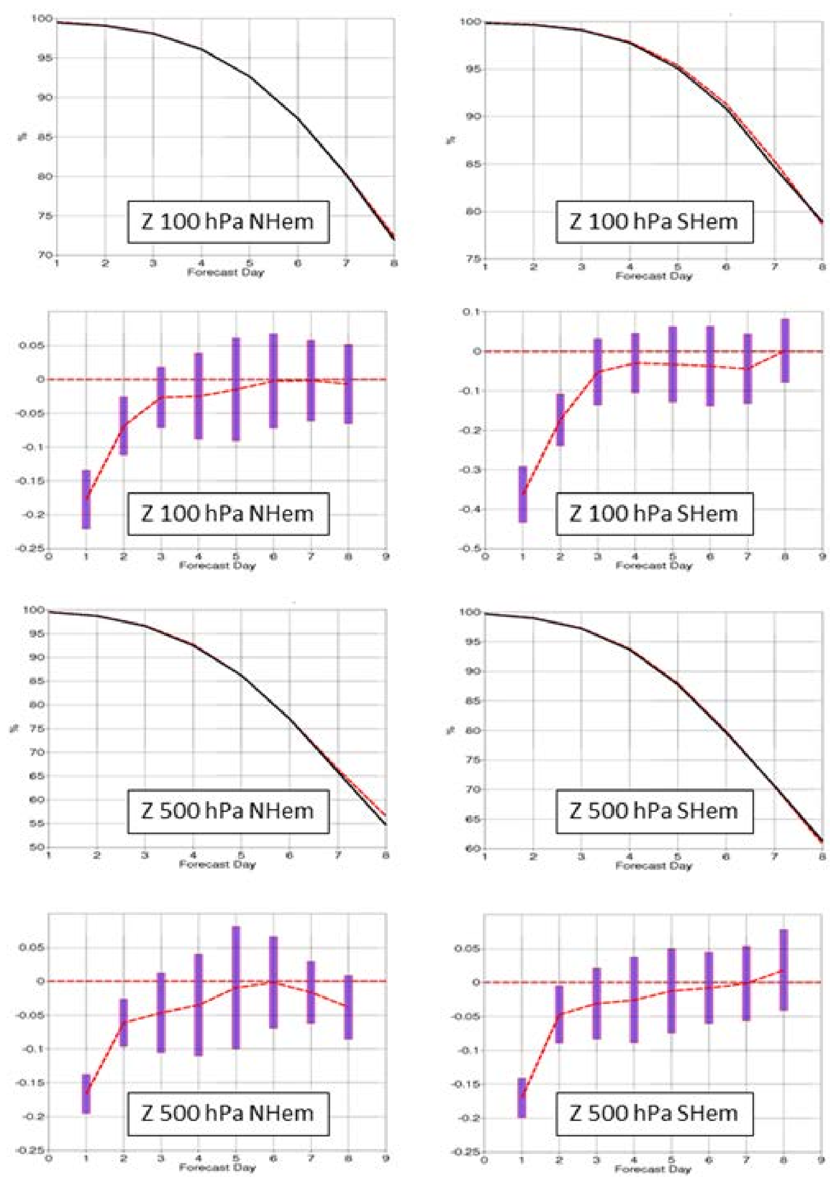
ECMWF

What is the ECMWF EnKF deterministic forecast skill?

TL399 100 member EnKF

TL399 (95/159) 4DVar with static B

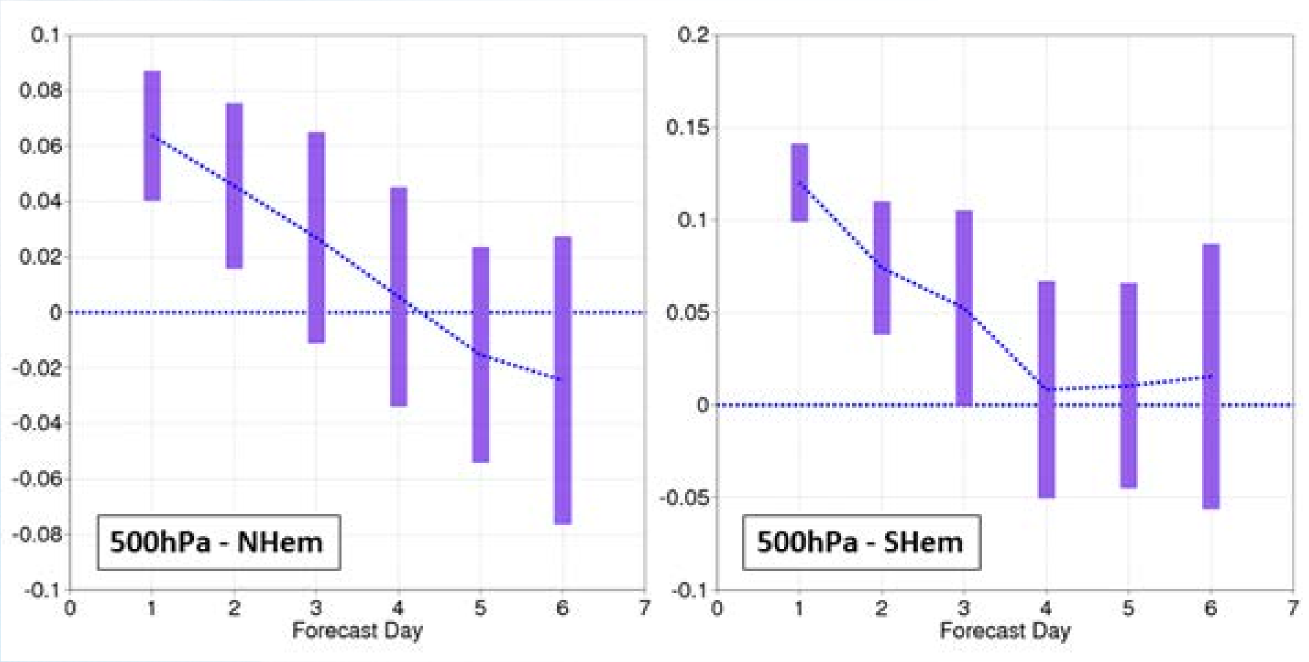
Verification vs ECMWF Operational 4DVar analysis



EnKF

Further developments

- Ensemble size with connected reduction of localization



Relative improvement in Forecast Anomaly Correlation of **240** EnKF with 2*Localization Length/Height scales vs **120** EnKF

EnKF

Further developments

- Ensemble size increase with connected reduction of localization
- Scale-dependent localization
- Cheap (!) algorithmic refinements to improve usage of non-linear observation types (i.e., rainy radiances, scatterometer winds)
- Can we do better than additive inflation or is model error modelling a lost cause?

HYBRID Data Assimilation

The case for Hybrid DA

- Strong constraint 4DVar is not able to evolve state errors beyond assimilation window, which cannot be longer than 6-12hours (longer windows lead to nonlinearities, convergence problems, general degradation of scores): No flow-dependent Errors-Of-The-Day
- Weak constraint 4DVar is still not proven with realistic global NWP (It also shifts the problem from $\mathbf{B}$ estimation to $\mathbf{Q}$: is this any simpler?)
- We still need an ensemble of analyses to start an EPS
- We can expect Hybrids to be more robust state estimators than pure EnKF in the presence of sampling errors and significant model error
- We can reuse in Hybrids well tested and efficient components of existing Var systems (Var QC of observations, Var bias correction of satellite radiances)

HYBRID Data Assimilation

- Strong constraint 4DVar: No flow-dependent Errors-Of-The-Day
- Two solutions have been pursued at global NWP Centres:

1. Augment the **B** model of a standard 4DVar with the current, localised ensemble perturbations ("alpha" control variable)

2. Train the **B** model on the current + climatological ensemble perturbations (Hybrid 4DVar-EDA)

HYBRID Data Assimilation

1. Augment the **B** model at the start of the assimilation window of a standard 4DVar with the current, localised ensemble perturbations ("alpha" control variable; Barker, 1999; Lorenc, 2003)

Conceptually adds a flow-dependent term to the model of $\mathbf{P}^b$ ($\mathbf{B}$):

$$\mathbf{B} = \beta_c^2 \mathbf{B}_c + \beta_e^2 \mathbf{P}_e \circ \mathbf{C}_{loc}$$

$\mathbf{B}_c$ is the static, climatological covariance

$\mathbf{P}_e \circ \mathbf{C}_{loc}$ is the localised ensemble sample covariance

In practice this is done through augmentation of the control variable:

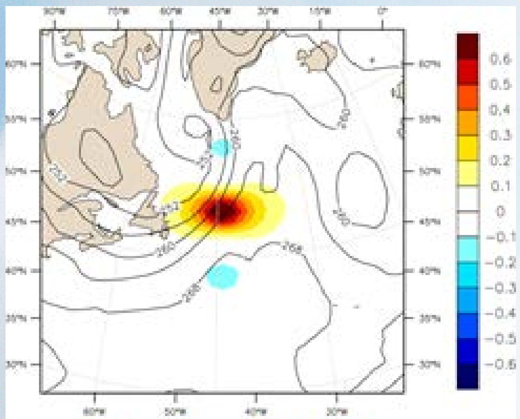
$$\delta\mathbf{x} = \beta_c \mathbf{B}_c^{1/2} \mathbf{v} + \beta_e \mathbf{X}' \circ \boldsymbol{\alpha}$$

and introducing an additional term in the cost function:

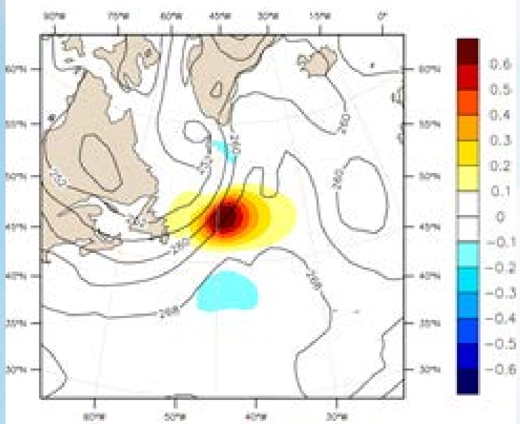
$$J = \frac{1}{2} \boldsymbol{v}^T \boldsymbol{v} + \frac{1}{2} \boldsymbol{\alpha}^T \boldsymbol{C}_{loc}^{-1} \boldsymbol{\alpha} + J_o + J_c$$

Hybrids: α control variable

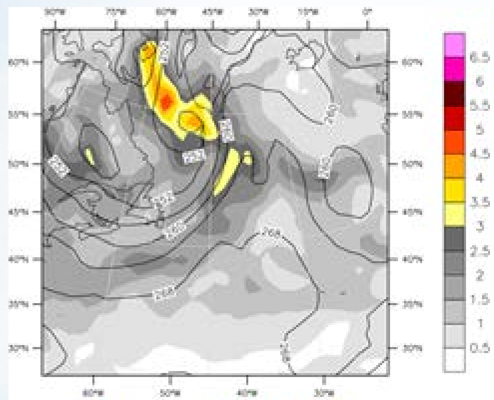
u response to a single u observation at centre of window



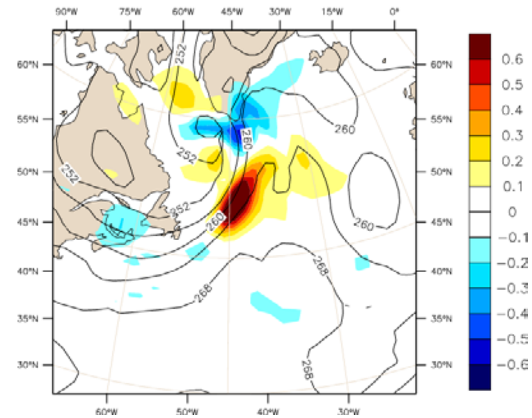
Standard 3D-Var



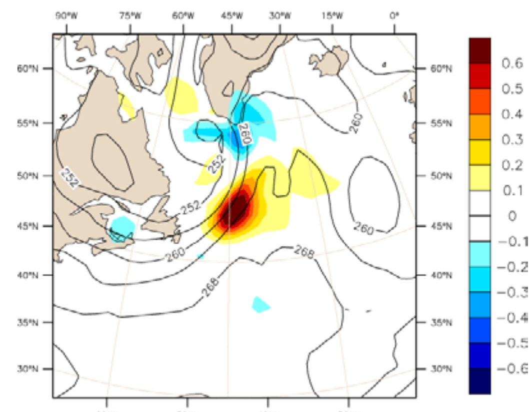
Standard 4D-Var



Ensemble RMS



Pure ensemble 3D-Var



50/50 hybrid 3D-Var

from: A.Clayton

HYBRID Data Assimilation

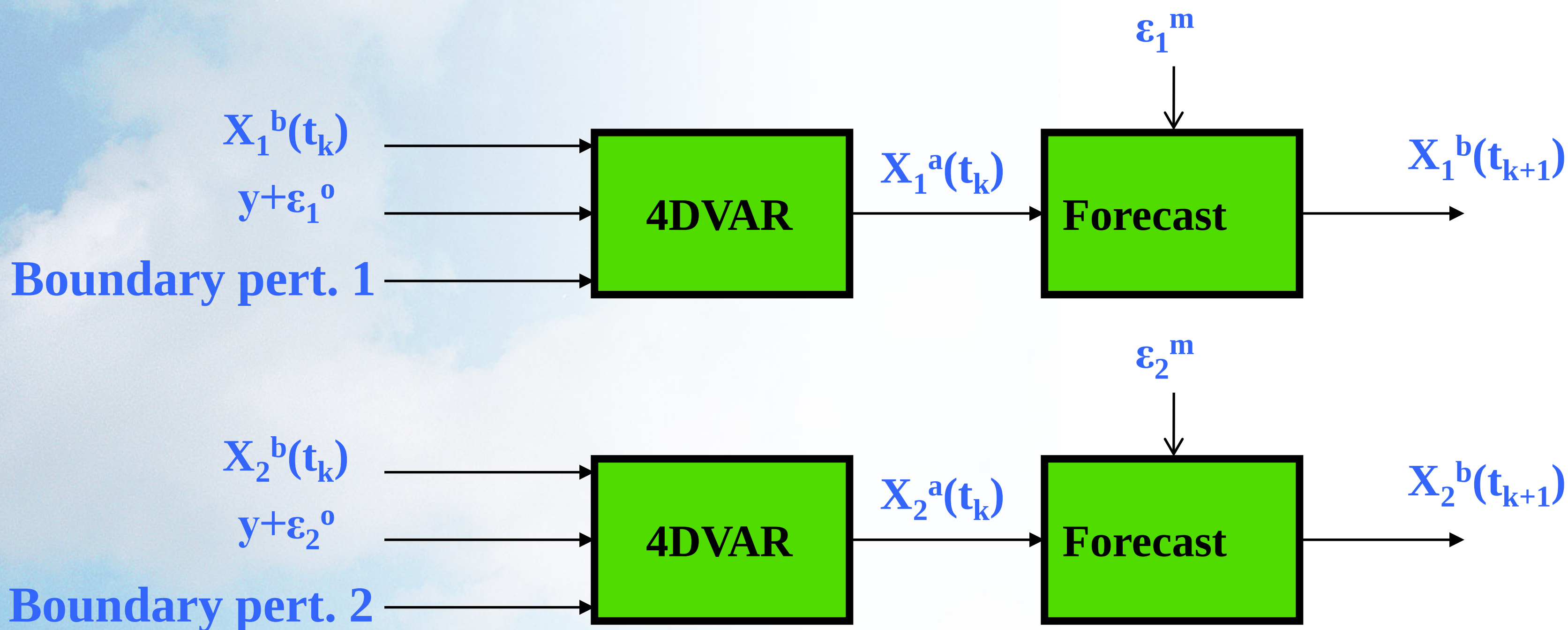
2. Train the **B** model on the current + climatological ensemble perturbations (Hybrid 4DVar-EDA; Isaksen et al, 2010, Bonavita et al., 2010,2012,2015)

There are two specific aspects to the current hybrid DA system at ECMWF:

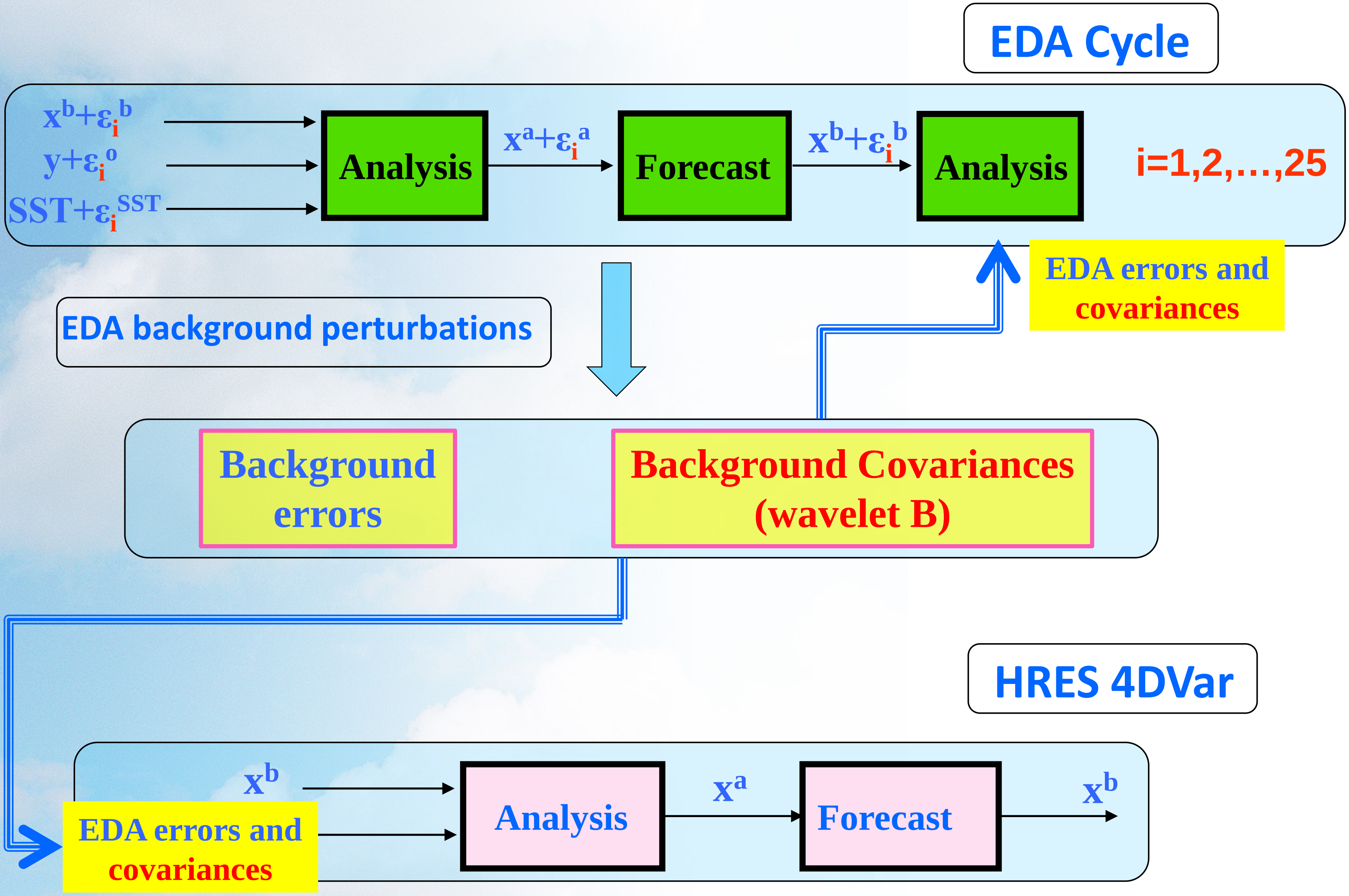
a) The error cycling system (EDA)

b) The **B** model (wavelet **B**)

The Ensemble of Data Assimilations (EDA)



- 25 ensemble members using 4D-Var assimilations at considerably reduced resolution wrt the HRES 4Dvar
- Observations, SST and Model perturbed


EDA Cycle
x^b+ε_i^b
y+ε_i^o
SST+ε_i^SST
Analysis
x^a+ε_i^a
Forecast
x^b+ε_i^b
Analysis
i=1,2,…,25
EDA errors and covariances
EDA background perturbations
Background errors
Background Covariances (wavelet B)
HRES 4DVar
x^b
EDA errors and covariances
Analysis
x^a
Forecast
x^b

Spectral B model

In variational analysis the $\mathbf{B}$ matrix is usually defined implicitly in terms of a transformation from the departure $\delta\mathbf{x}$ in state space to a control variable χ:

$$\delta\mathbf{x} = \mathbf{x}\text{-}\mathbf{x}_b = \mathbf{L}\chi$$

where $\mathbf{L}$ verifies $\mathbf{B}=\mathbf{L}\mathbf{L}^T$

In the spectral formulation (Derber and Bouttier, 1999), the change of variable $\mathbf{L}$ has the form:

$$\mathbf{L} = \mathbf{K}\,\mathbf{B}_u^{1/2}$$

where $\mathbf{K}$ is a balance operator going from the set of "unbalanced " variables $[\zeta, \eta_u, (T,ps)_u, q]$ (the "control vector") to the set of state variables $[\zeta, \eta, (T,ps), q]$

There is a degree of flow-dependence in $\mathbf{K}$ as the balance constraints are linearised about the first-guess trajectory

Spectral B model

$$\delta\mathbf{x} = \mathbf{x}\text{-}\mathbf{x}_\mathrm{b} = \mathbf{L}\chi \qquad \mathbf{L} = \mathbf{K}\,\mathbf{B}_\mathrm{u}^{1/2}$$

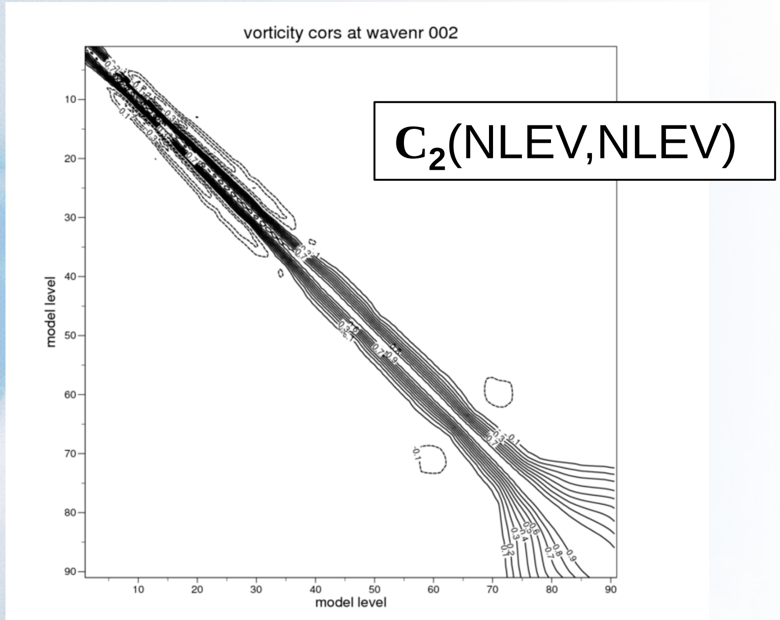
Since we assume that the balance operator accounts for all inter-variable correlations, $\mathbf{B}_\mathbf{u}$ is block diagonal

$$\mathbf{B}_u = \begin{pmatrix} \mathbf{B}_\zeta & 0 & 0 & 0 \\ 0 & \mathbf{B}_{D_u} & 0 & 0 \\ 0 & 0 & \mathbf{B}_{(T,p_s)_u} & 0 \\ 0 & 0 & 0 & \mathbf{B}_q \end{pmatrix}$$

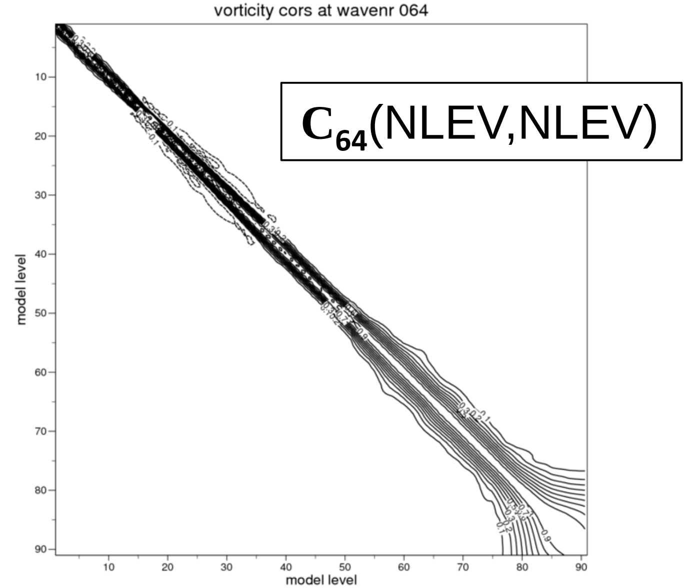
Each block in $\mathbf{B}_\mathbf{u}$ is of the form $\mathbf{\Sigma}^\mathbf{T}\mathbf{C}\mathbf{\Sigma}$

$\mathbf{\Sigma}$ is the gridpoint standard deviation of background errors

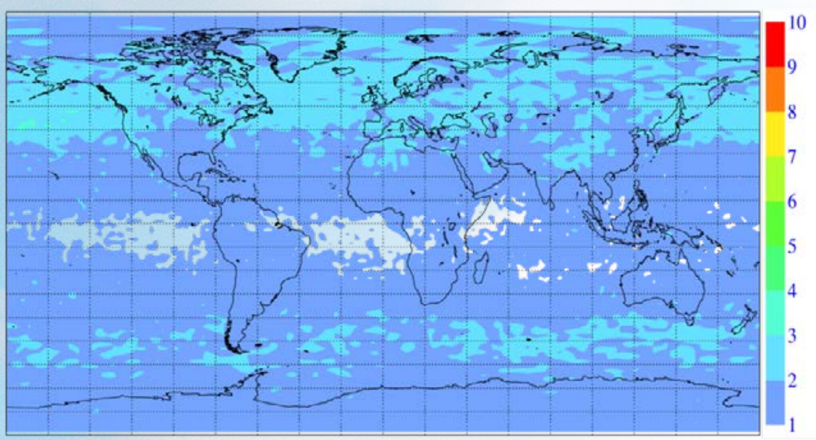
$\mathbf{C}$ models the autocorrelation of the control variables. It is block diagonal with one full vertical correlation matrix for each spectral wavenumber, i.e. C_n(NLEV,NLEV) (non-separable $\mathbf{B}$ model)



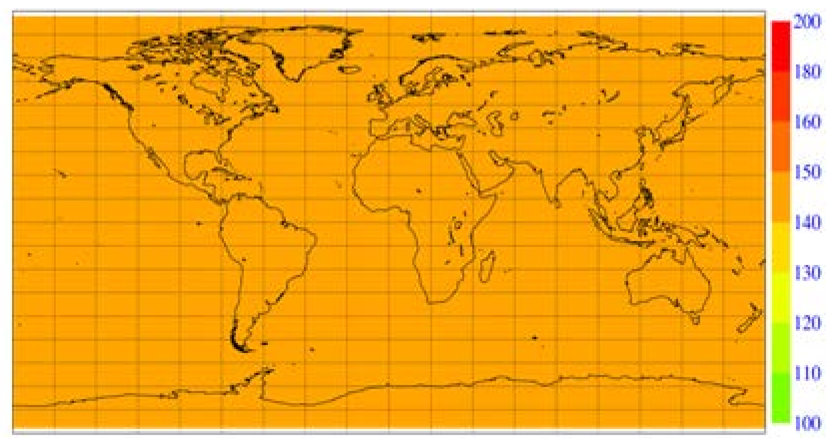
Vorticity correl. wavenum=2



Vorticity correl. wavenum=64



Vorticity bg error stdev, 500hPa



Vorticity bg error corr. Lscale, 500hPa

From Spectral to Wavelet B model

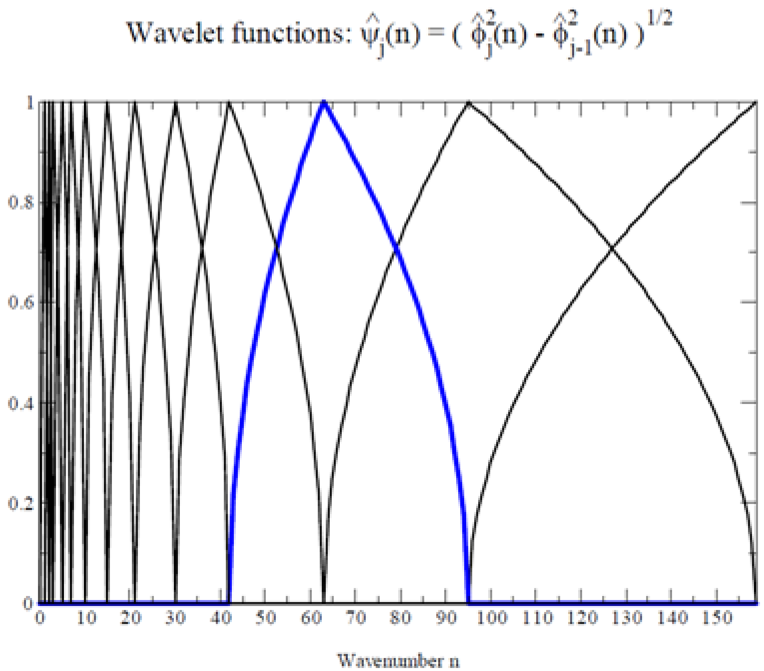
- The spectral **B** model is one end of the spectrum: full resolution of the variation of vertical correlation with horizontal scale, but it allows no horizontal variability of the vertical/horizontal correlations

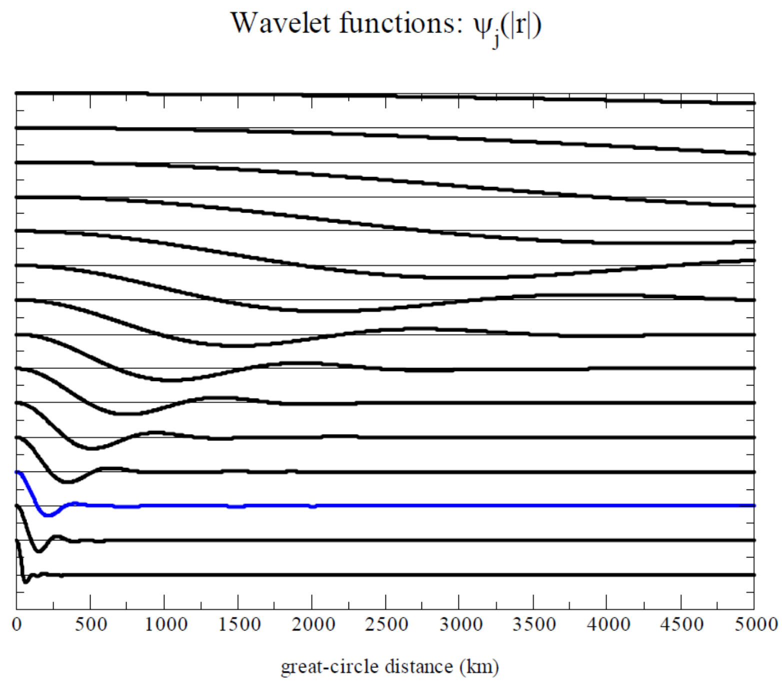
- The other end of the spectrum is represented by the separable formulation which allows full horizontal variation of the correlations (we may specify a different vertical covariance matrix for each horizontal grid point), but has no variation of vertical correlation with horizontal scale

- The wavelet **B** (Fisher, 2003) is a compromise between these two extremes and allows a degree of variation of correlation with both wavenumber and horizontal location

Wavelet B model

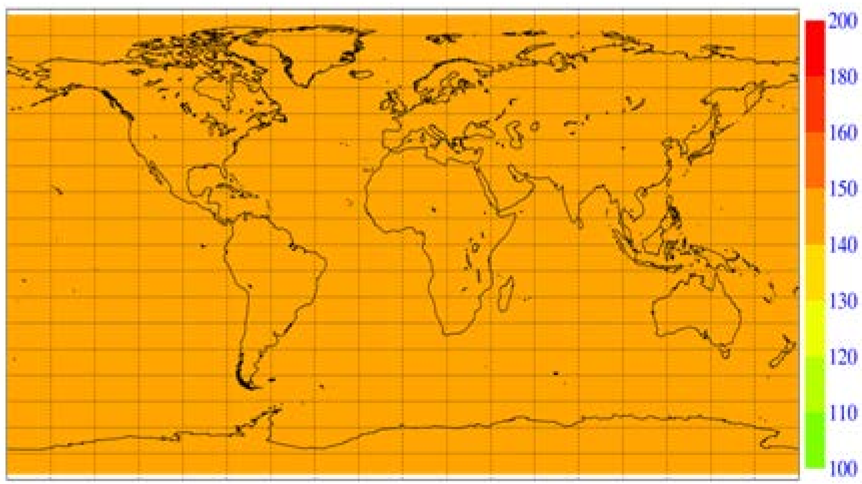
- The wavelet $\mathbf{B}$ is based on a wavelet expansion on the sphere.
- The basis functions (wavelets) are chosen to be band-limited and, to a good approximation, spatially localized



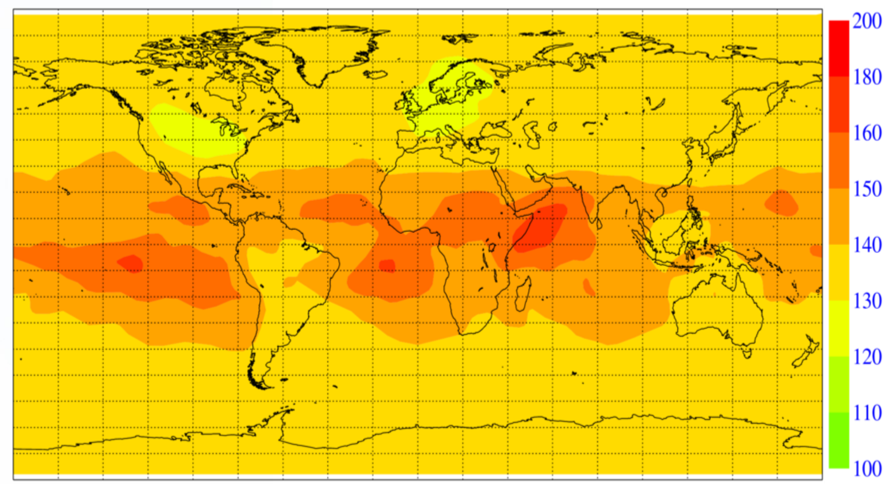


Wavelet B model

- The correlation matrices $\mathbf{C_n}[N_{lev}\mathrm{x}N_{lev}]$ are now of the form $\mathbf{C_j}[N_{lev}\mathrm{x}N_{lev}](\lambda,\varphi)$, where $\mathbf{j}$ is now the index of the wavelet component
- The choice of the wavelet bandwidths $[N_j, N_{j+1}]$ determines the trade-off between spectral and spatial resolution. If the bands are narrow, the corresponding wavelet functions are not spatially localized, and vice versa



Climat. Spectral **B**
Vorticity bg error corr. Lscale, 500hPa



Climat. Wavelet **B**
Vorticity bg error corr. Lscale, 500hPa

Flow-dependent wavelet B model

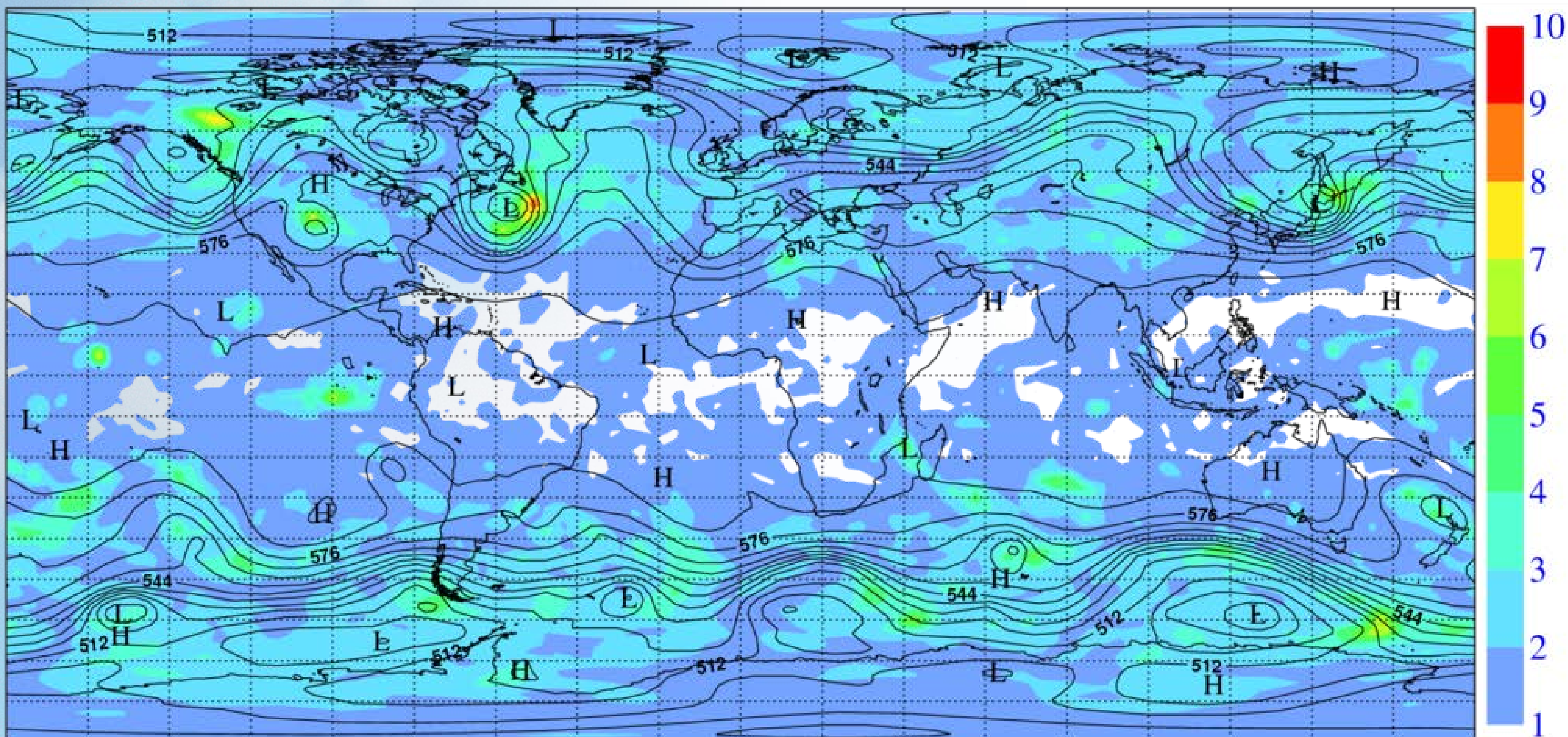
The wavelet $\mathbf{B}$ formulation:

$$\left(\mathbf{x}-\mathbf{x}_b\right)=\mathbf{L}\boldsymbol{\chi}=\mathbf{K}\boldsymbol{\Sigma}_b^{1/2}\sum_j \psi_j \otimes\left[\mathbf{C}_j^{1/2}(\lambda,\phi)\chi_j\right]$$

can be made flow-dependent by obtaining flow-dependent estimates of the background error variances ($\boldsymbol{\Sigma}_b$) and correlations ($\mathbf{C}_j(\lambda,\varphi)$) from the EDA background perturbations

Flow-dependent background errors

Tuesday 3 April 2012 12UTC ECMWF Forecast t+9 VT: Tuesday 3 April 2012 21UTC 500hPa Geopotential



- EDA estimate of background error standard deviation for Vorticity at 500 hPa after spatial filtering and calibration (Units=10^{-5} s^{-1})

Flow-dependent wavelet B model

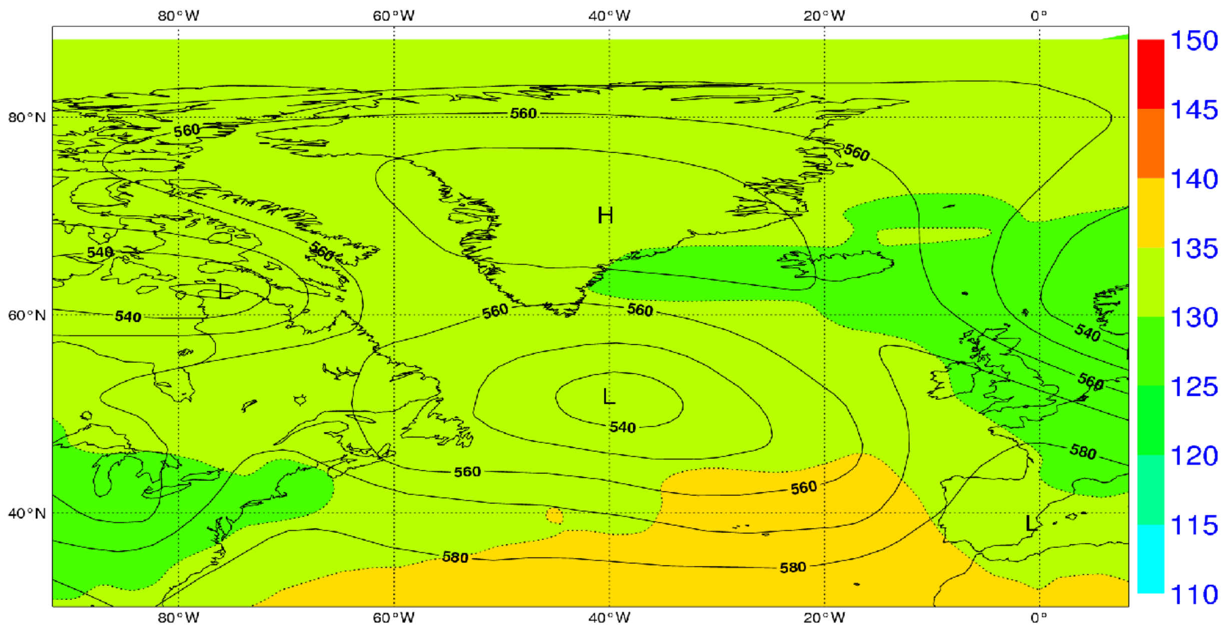
$$\left(\mathbf{x}-\mathbf{x}_b\right)=\mathbf{L}\boldsymbol{\chi}=\mathbf{K}\boldsymbol{\Sigma}_b^{1/2}\sum_j \psi_j \otimes\left[\mathbf{C}_j^{1/2}(\lambda,\phi)\chi_j\right]$$

The computation of the wavelet B (i.e., the correlations ($\mathbf{C}_j(\lambda,\varphi)$) requires considerably more EDA perturbations than those available from the latest EDA. For this reason they are estimated through a linear combination of a climatological wavelet B and perturbations from the latest EDA:

$$\mathbf{C}_{hybrid}=(1-\alpha)\mathbf{C}_{static}+\alpha\mathbf{C}_{online}$$

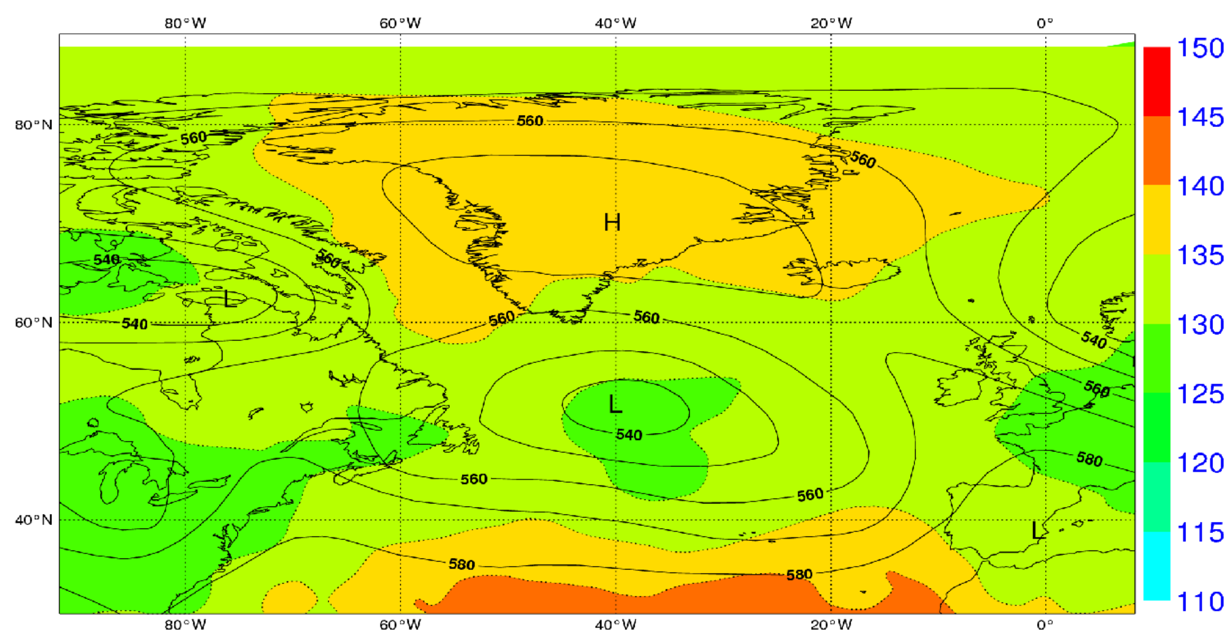
alpha is currently set at 0.3.

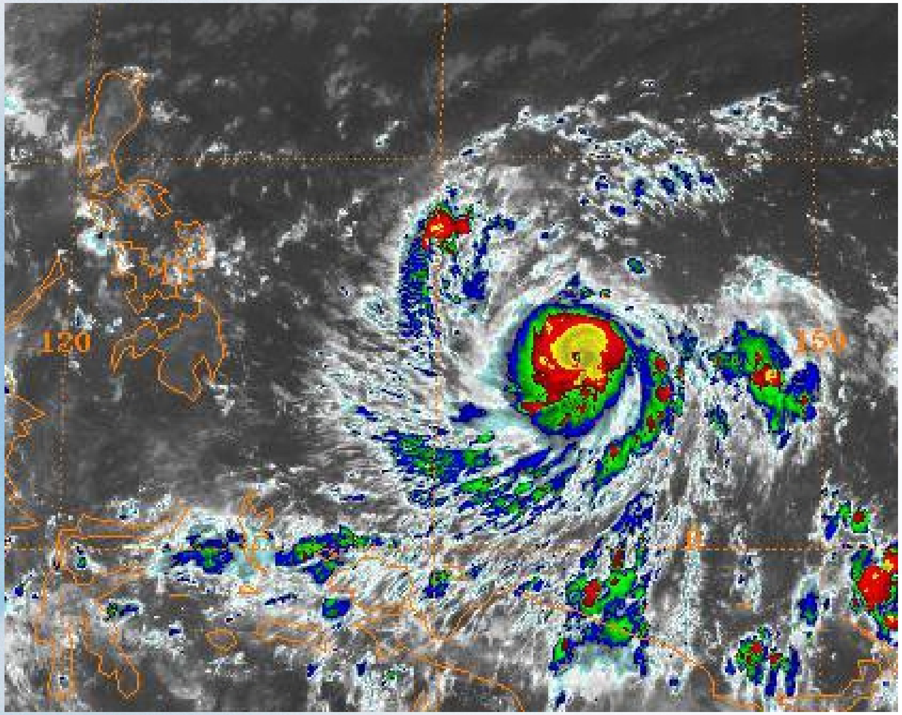
Error Correlation L-scales for Vorticity, 500 hPa



Static wavelet B

Hybrid wavelet B

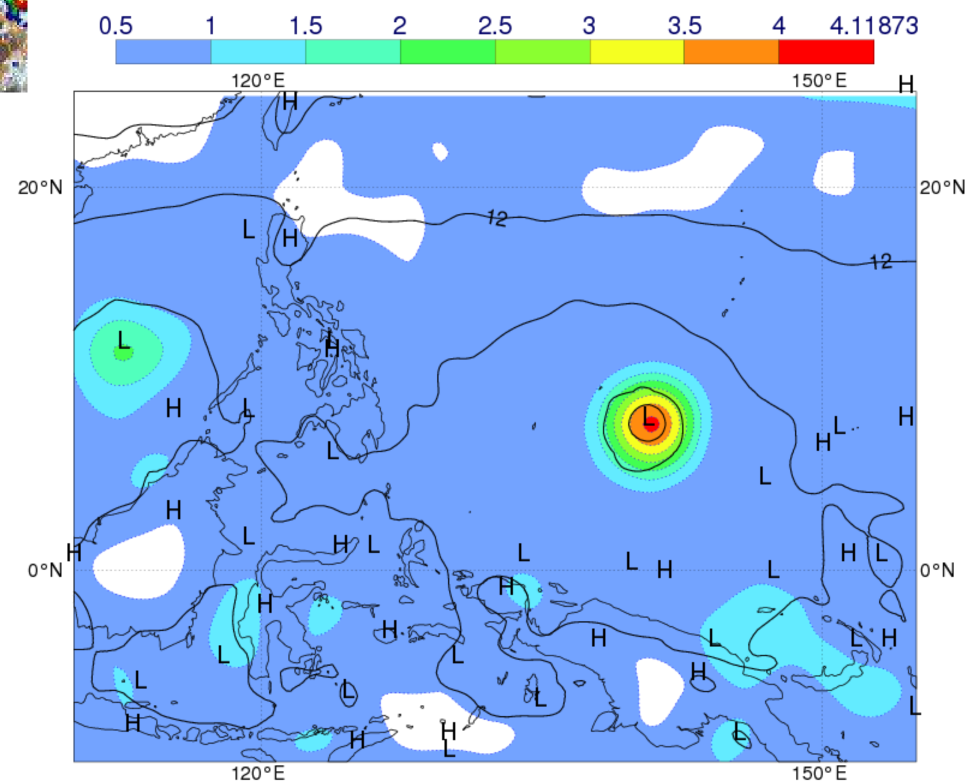




TYPHOON HAIYAN

MTSAT IR

2013-11-05 21UTC

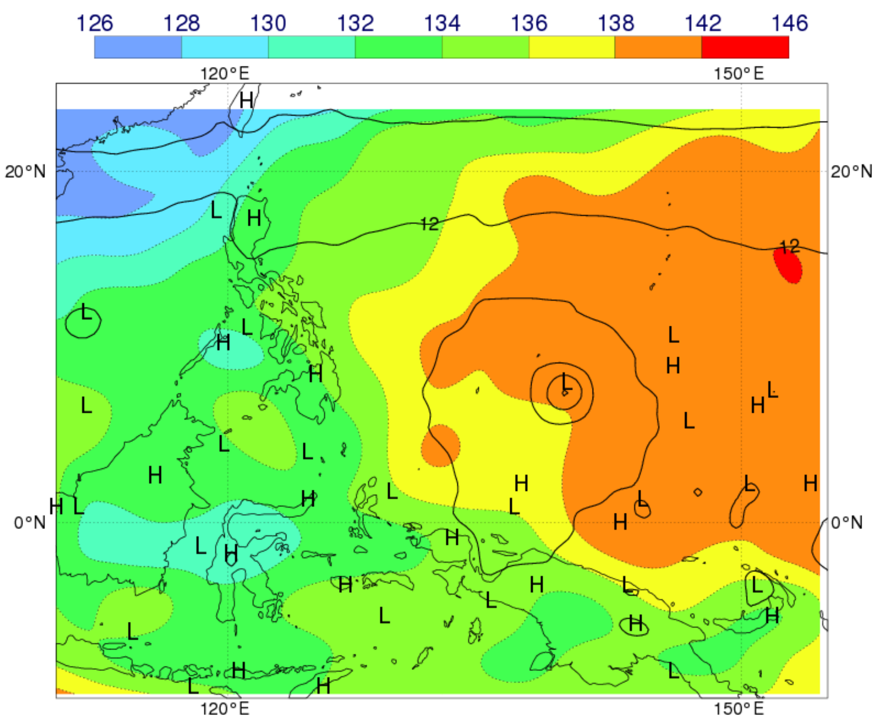


Z1000 BG (isolines)

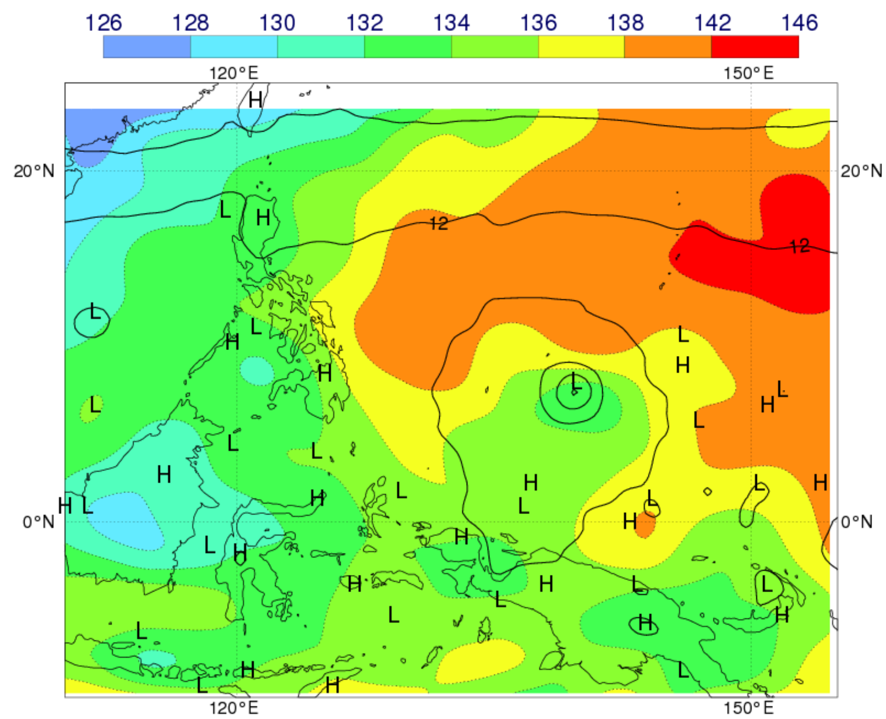
EDA Vorticity Spread (shaded) $10^{-5}s^{-1}$

valid at 2013-11-05 21UTC

Climat. Wavelet B



Hybrid Wav. B (α=0.3)

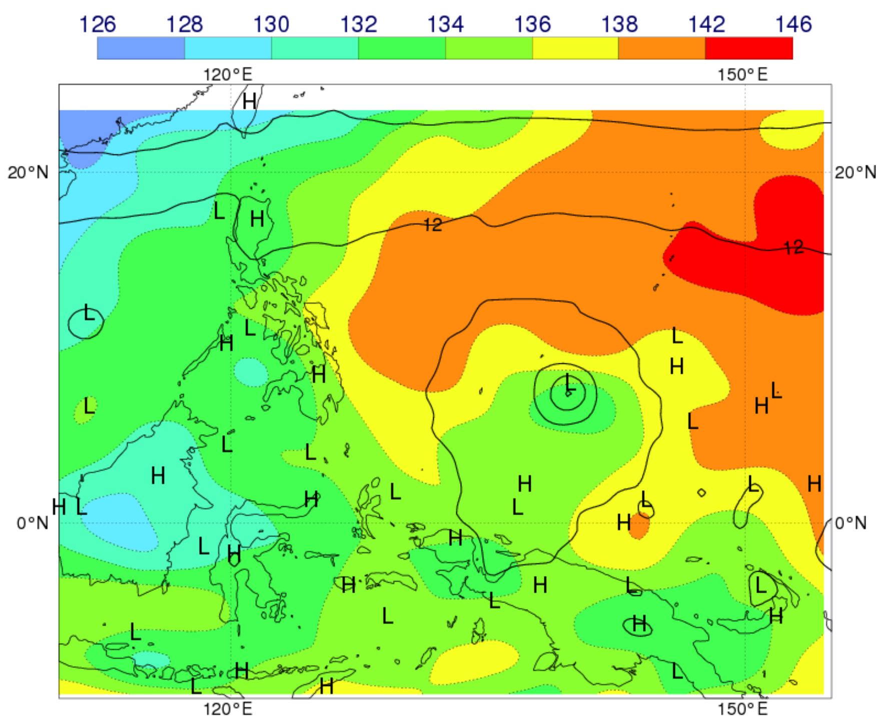


Vorticity errors length scale at the surface (shaded)

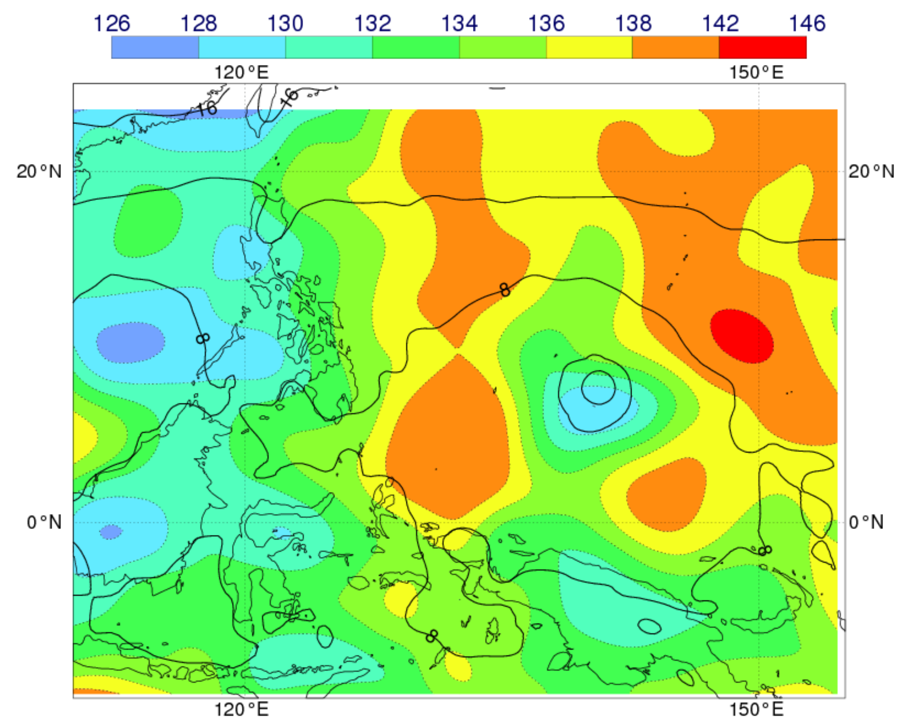
Geopotential height at 1000hPA (isolines)

“What if we had a 100 member ensemble DA?”

Hybrid Wavel. B (α=0.3)

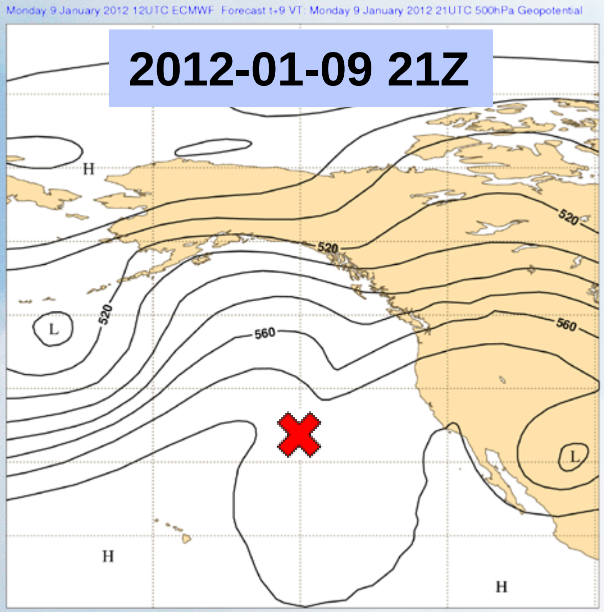


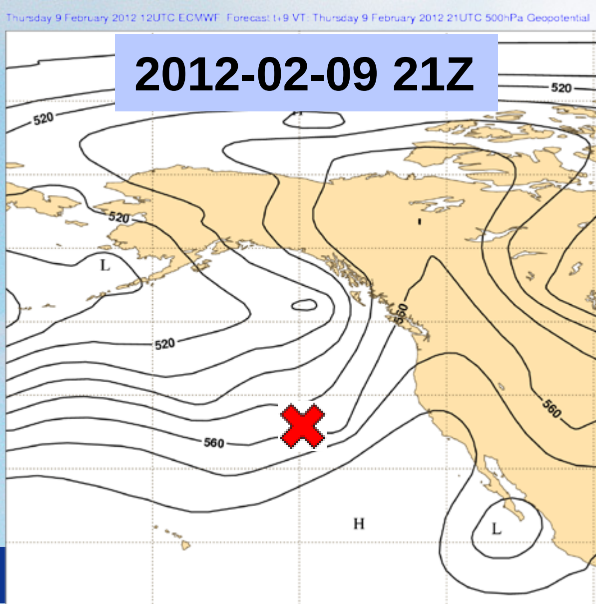
Hybrid Wavel. B (α=0.7)

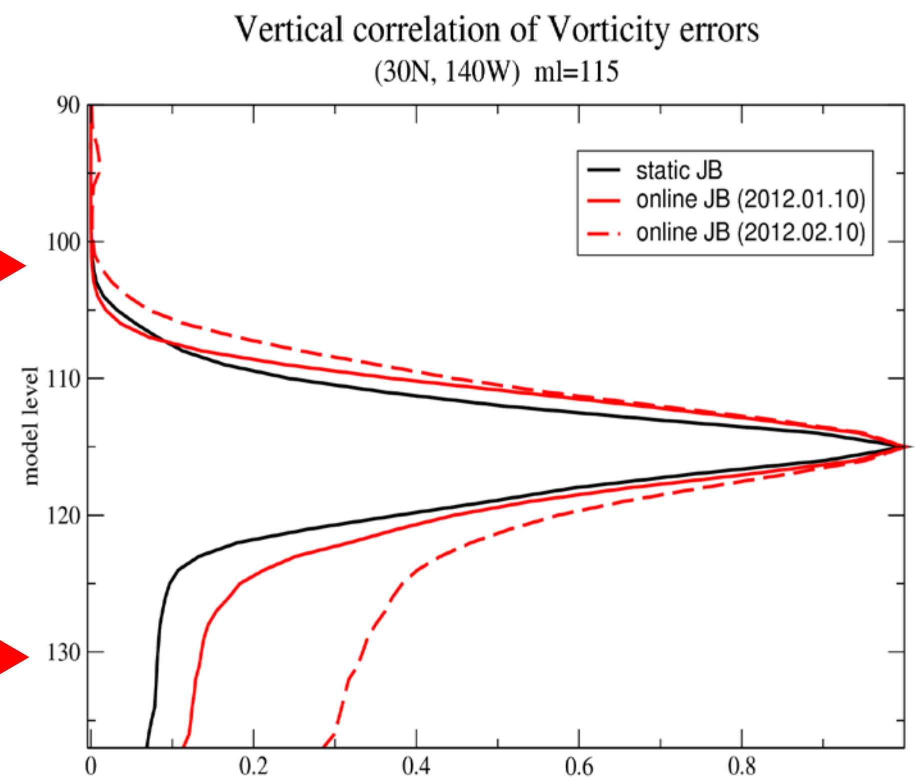


Vorticity errors length scale at the surface (shaded)
Geopotential height at 1000hPA (isolines)

Vertical Error Correlation - Vorticity, 850 hPa

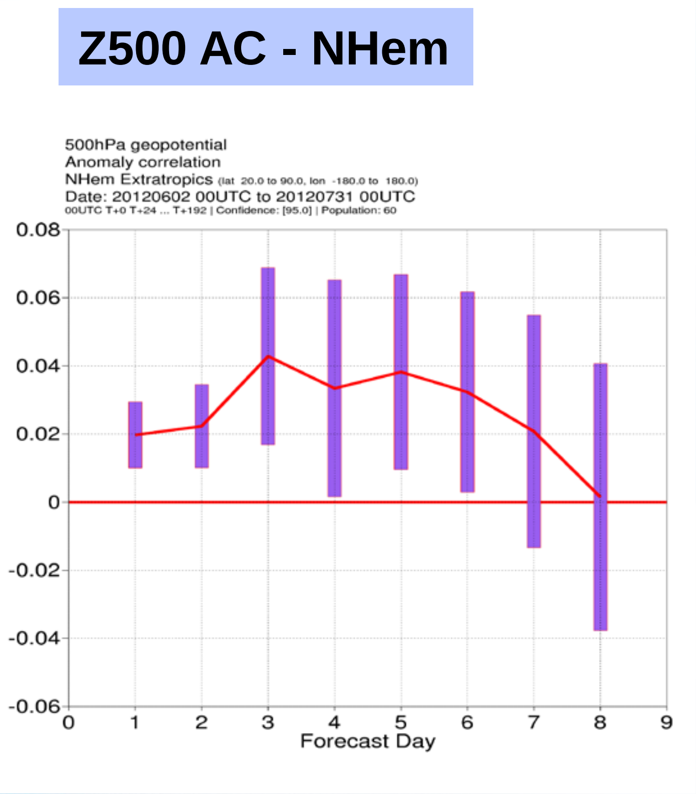


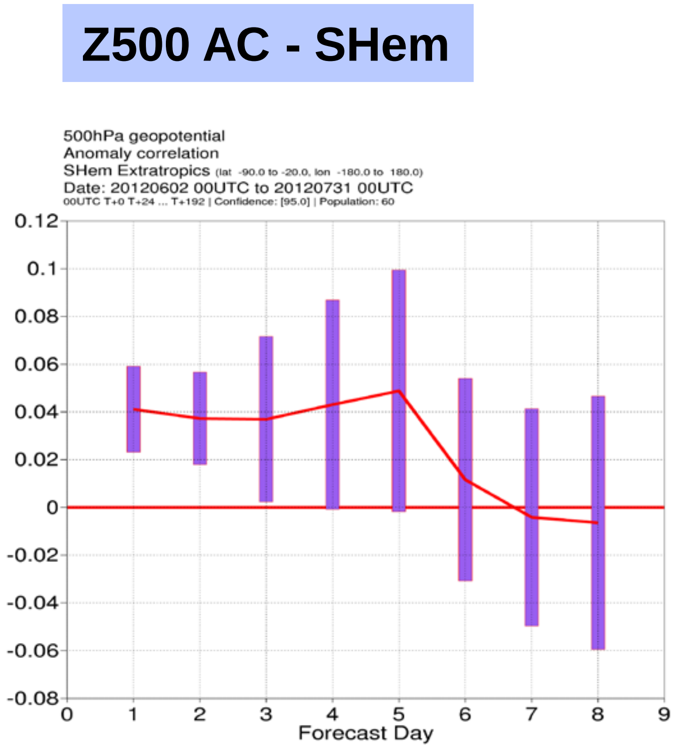




ECMWF

Skill of Hybrid 4DVar-EDA over 4DVar with climatological errors and correlations





Current Developments

1. 4D-Ensemble-Var (Liu et al., 2008)
2. Hybrid Gain EnDA (Hamrud et al., 2015; Bonavita et al., 2015)

4D-Ensemble-Var

In the standard hybrid 4DVar based on the alpha control variable method ensemble perturbations are used to augment the $\mathbf{B}$ model at the start of the assimilation window. This $\mathbf{B}$ model is still implicitly propagated in time by the TL and ADJ of the forecast model:

$$\mathbf{B}(t) = \mathbf{M}\left(\beta_c^2 \mathbf{B}_c + \beta_e^2 \mathbf{P}_e \circ \mathbf{C}_{loc}\right)\mathbf{M}^{\mathrm{T}}$$

$$\mathbf{P_e} = \frac{1}{N-1}\mathbf{X}\mathbf{X}^T$$

In 4D-En-Var the 4D $\mathbf{B}$ is directly modelled by the localised ensemble covariances, i.e.:

$$\mathbf{B}(t) = \mathbf{P}_e \circ \mathbf{C}_{loc}$$

4D-Ensemble-Var

In 4D-En-Var the analysis increments are thus 4D localised linear combinations of ensemble perturbations:

$$\delta\mathbf{x} = \sum_{k=1,N} \boldsymbol{\alpha}_k \circ \mathbf{x}'_k(t)$$

This is fundamentally the same mean state update procedure of the LETKF version of EnKF (Hunt et al., 2007).

Is there any fundamental reason why 4D-En-Var should perform better than LETKF?

Current Developments

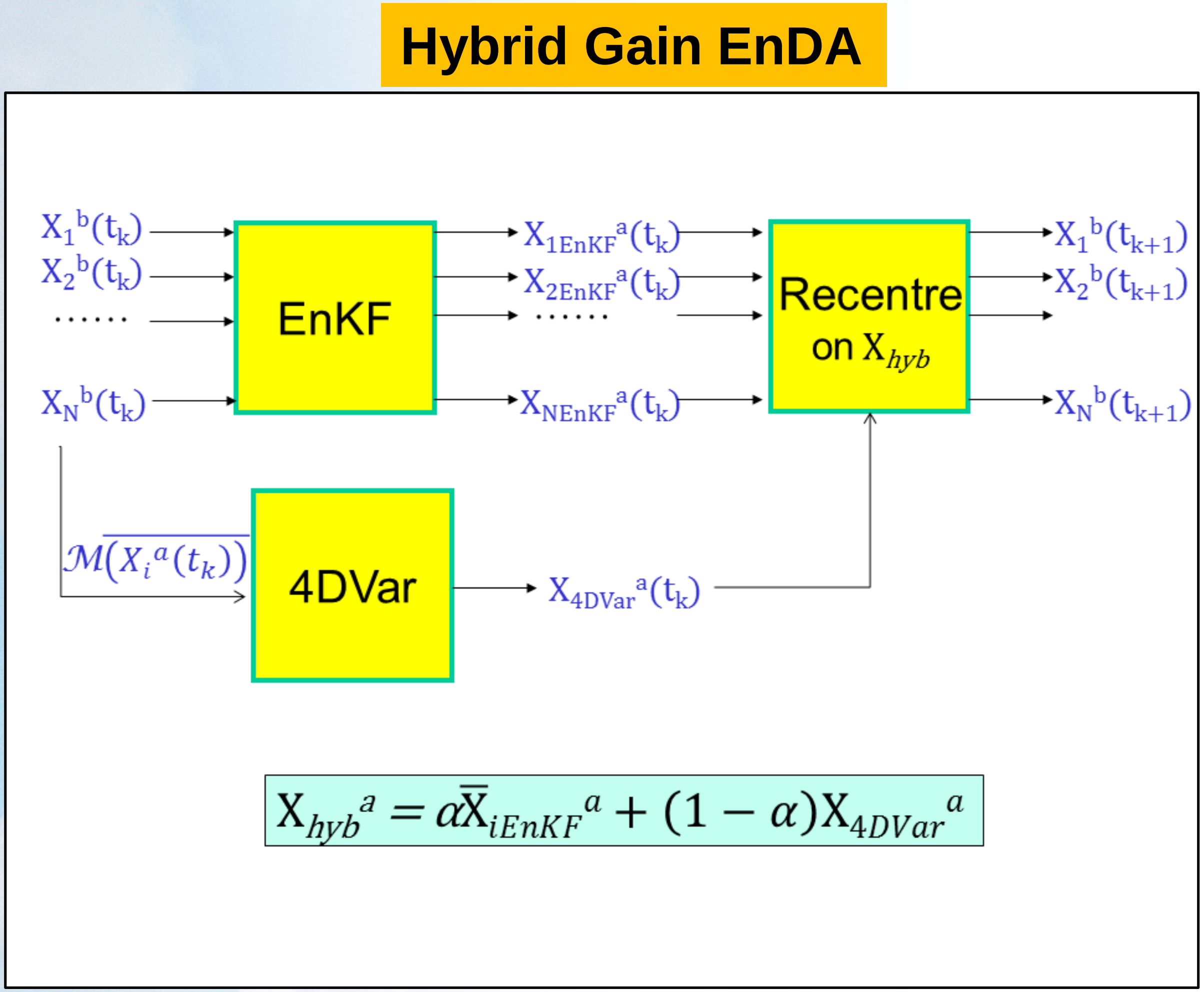
1. 4D-Ensemble-Var (Liu et al., 2008)

2. Hybrid Gain EnDA (Hamrud et al., 2015; Bonavita et al., 2015)

- Based on ideas from Penny (2013)
- Majority of proposed Hybrid DA systems use ensemble to construct/augment/blend the $\mathbf{B}$ model used in a variational analysis update with current ensemble perturbations
- We have seen that EnKF and 4DVar (with static B) have comparable accuracy (at least at ECMWF!)
- We could as well try blending the complete Kalman Gain matrices of the two systems (EnKF and 4DVar) in an EnKF framework

Hybrid Gain EnDA

- Can we improve by blending two analysis system of similar quality inside the EnKF framework?



$$\mathrm{X}_{hyb}{}^{a} = \alpha\overline{\mathrm{X}}_{iEnKF}{}^{a} + (1-\alpha)\mathrm{X}_{4DVar}{}^{a}$$

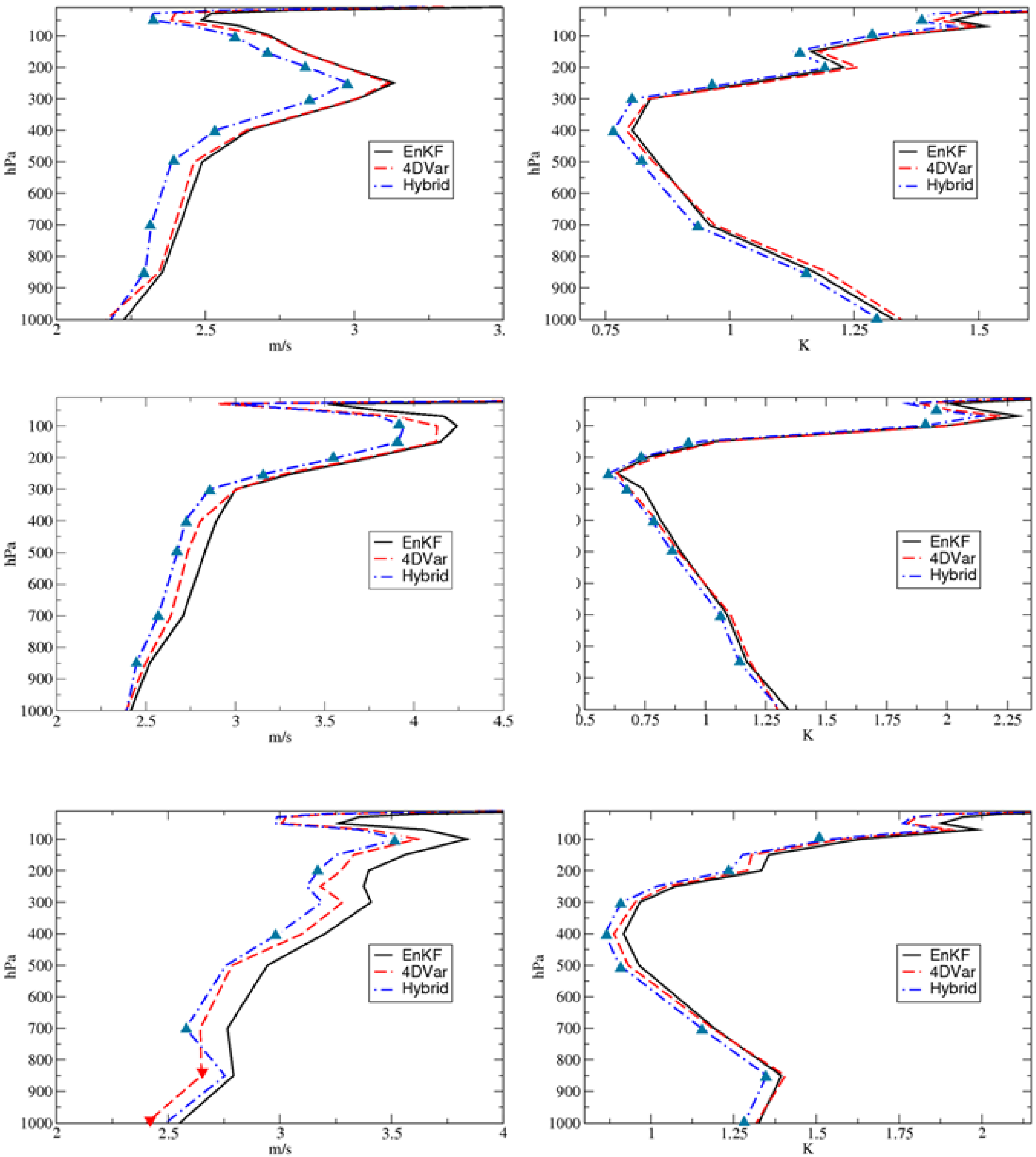
RAOB u wind **RAOB Temp.**

EnKF
4DVar
Hybrid Gain

NHem

Trop

SHem



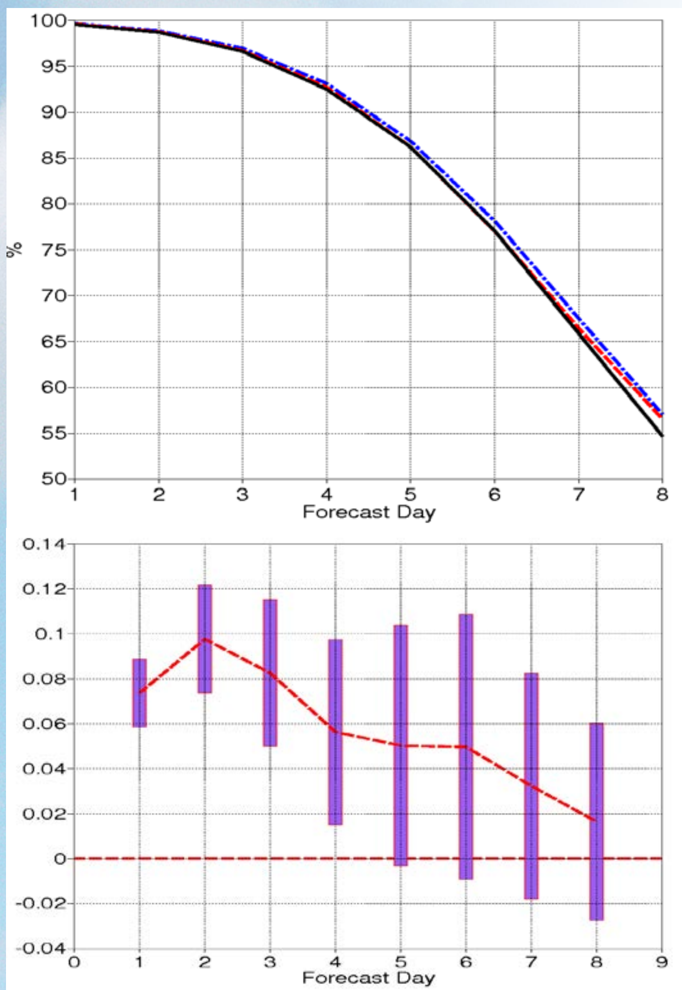
mo Bonavita ECMWF

Hybrid Gain EnDA

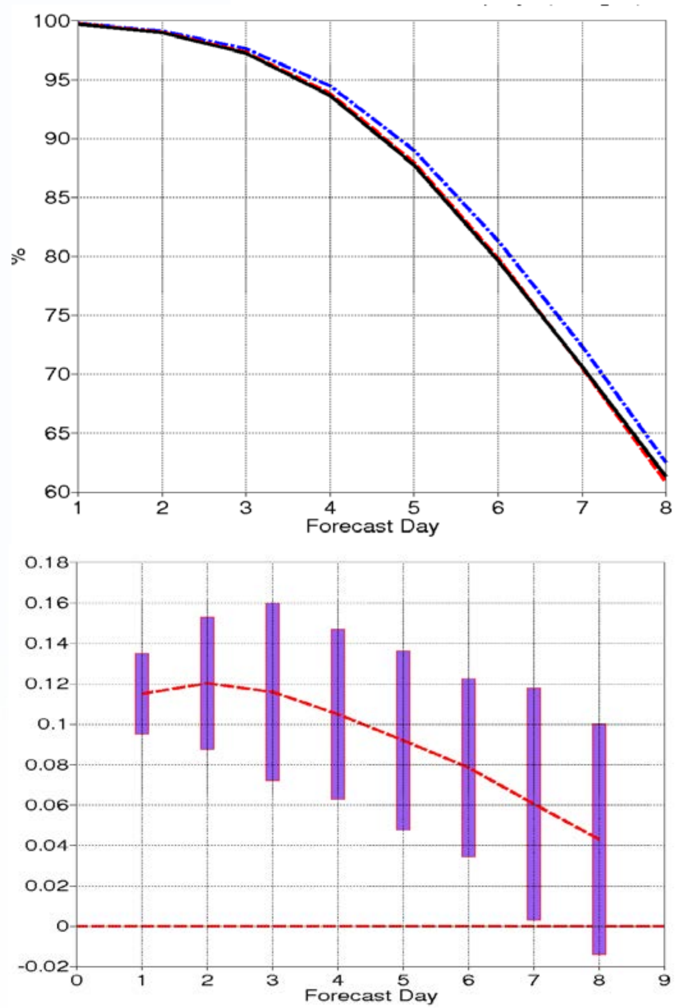
TL399 100 member EnKF

TL399 100 member Hyb. Gain EnDA

TL399 4DVar – static B



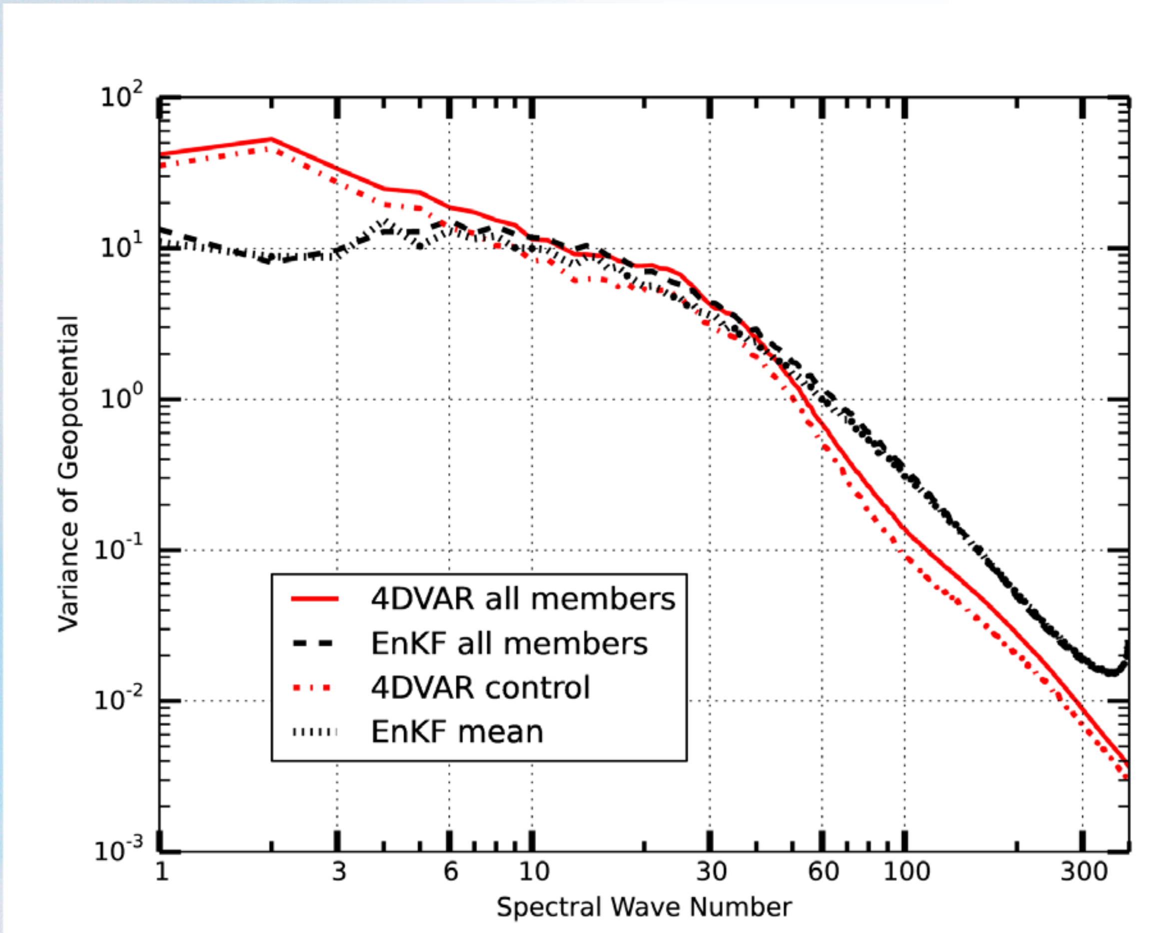
Z500 hPa AC - NHem



Z500 hPa AC - SHem

Hybrid Gain EnDA

- Hybrid Gain EnDA works surprisingly well. But why?

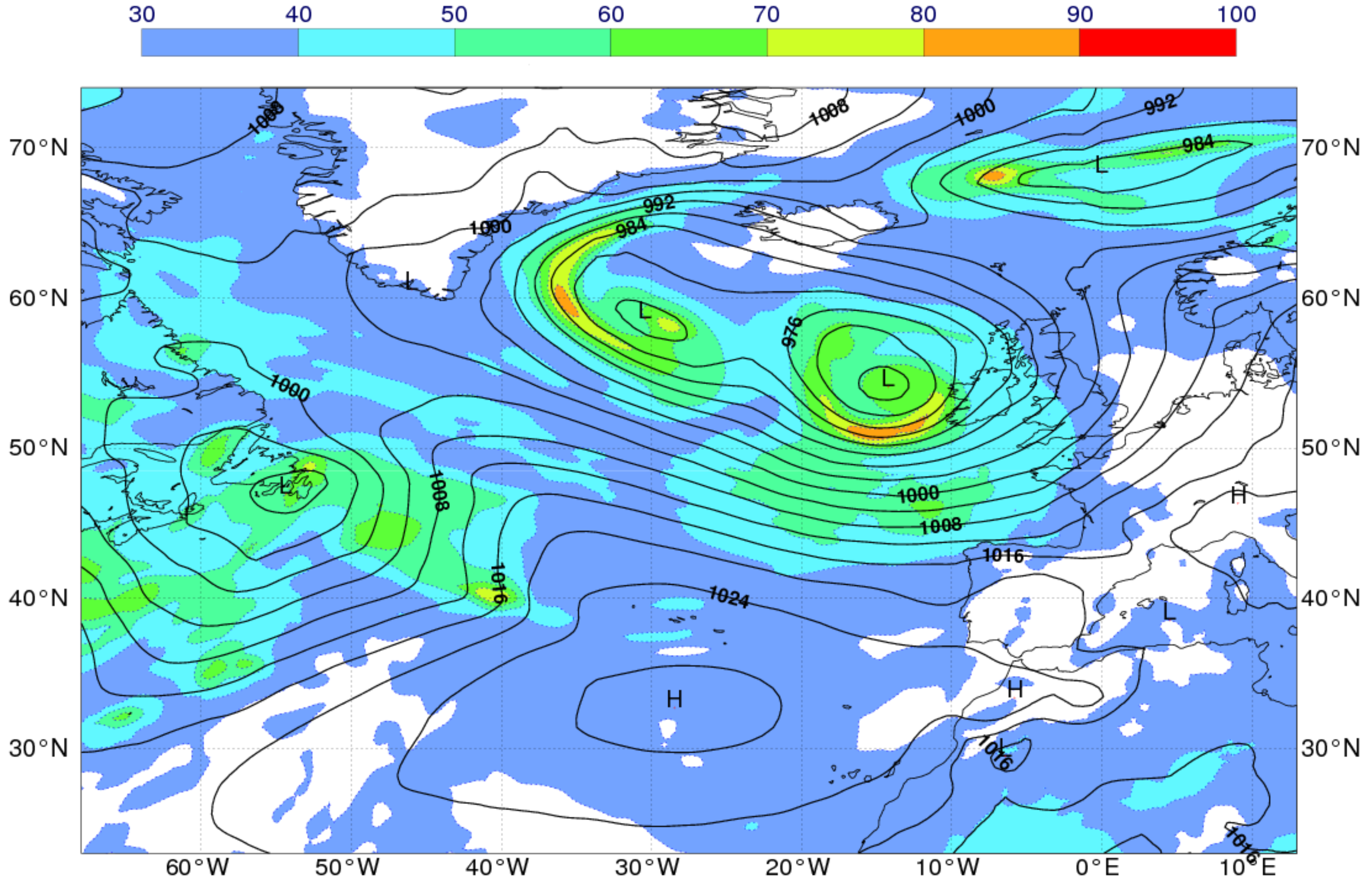


Power spectra of Z500 hPa analysis increments

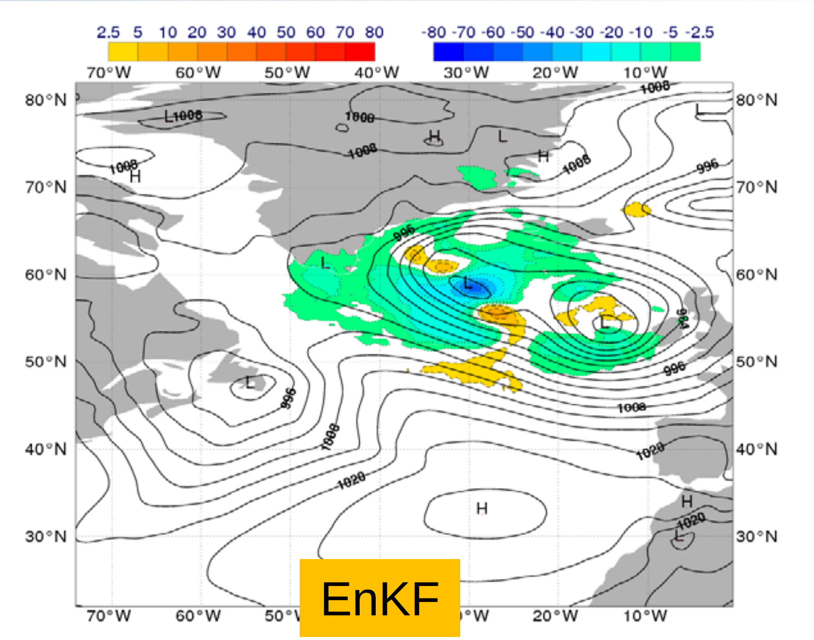
Hybrid Gain EnDA

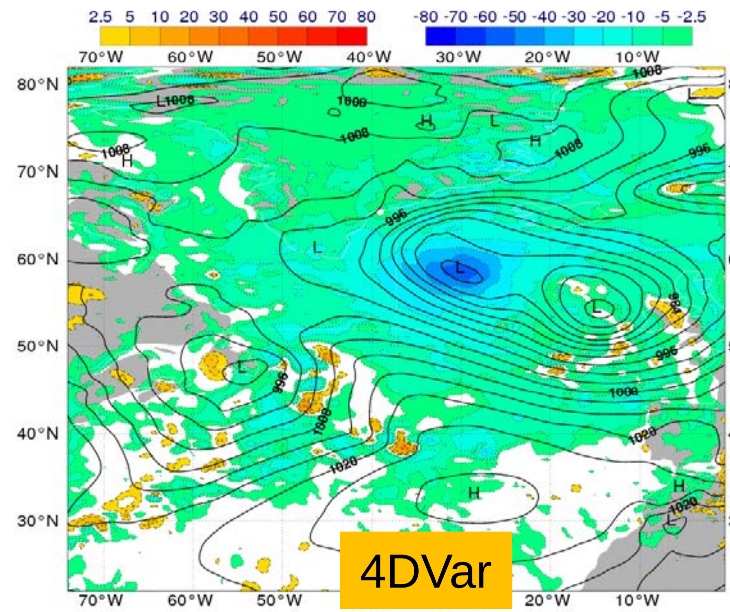
- Hybrid Gain EnDA works surprisingly well. But why?

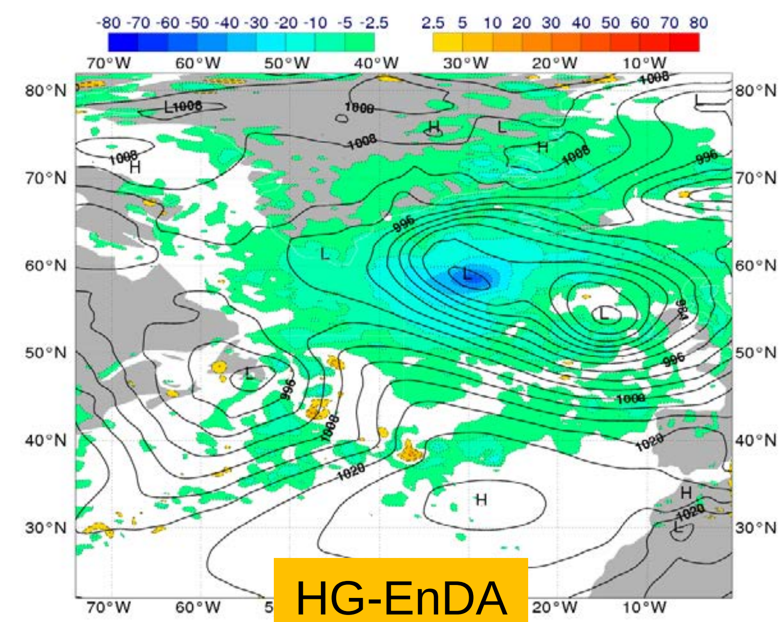
MSLP t+6h fcst and MSLP Ensemble stdev (shaded)
SP obs at (58.5N, 30.3W), middle of window, y-H(x)=-1hPa



SP obs at (58.5N, 30.3W), middle of window, y-H(x)=-1hPa







Hybrid Gain EnDA

- The positive effects of the HG-EnDA seem to originate from:
  1) Mitigating the effects of localization in the EnKF increments
  2) Introducing climatological information in the EnKF covariance estimates

- Ideally we would like to keep more of the EnKF flow-dependent structures near the observation location and gradually revert to the climatological covariances of 4DVar farther away: we are now looking at a scale-dependent blending of the two analyses ($\alpha = \alpha(n)$)

Final thoughts on Hybrid DA

- Hybrid DA has been one of the main drivers of progress in operational DA over the past 10 years
- As (hopefully!) shown in this presentation, there are many possible ways of setting up a Hybrid DA system
- So far, I do not see fundamental reasons to favour one over the others
- Choice will be most likely dictated by practical considerations:
  a) Computational efficiency and scalability on emerging computing architectures
  b) Size of affordable ensemble
  c) Complexity of development and maintenance of codebase
  d) Availability of well tested, reusable components

Thanks for your attention!
Questions?

"Expert: An ordinary man away from home giving advice"

Oscar Wilde

More details in:

Hamrud, Bonavita and Isaksen, 2015: "EnKF and Hybrid Gain Data Assimilation Part I: EnKF implementation", Mon. Wea. Rev., under revision

Bonavita, Hamrud and Isaksen, 2015: "EnKF and Hybrid Gain Data Assimilation Part II: EnKF and Hybrid Gain Results", Mon. Wea. Rev., submitted

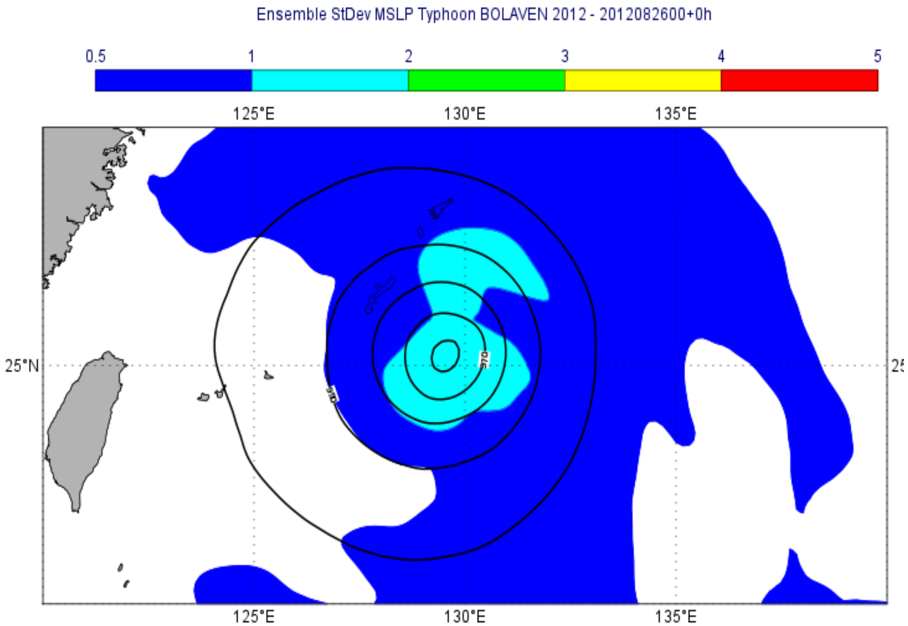
Bonavita, Holm, Isaksen and Fisher, 2014: "The evolution of the ECMWF hybrid data assimilation system", Q.J.R. Mets, submitted

Additional Slides

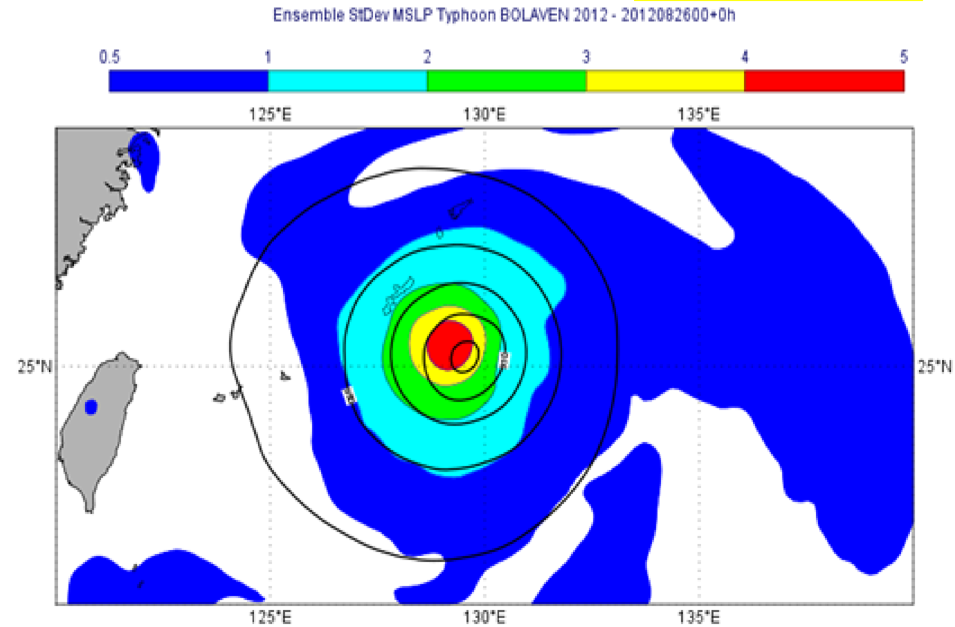
1. The EDA is an effective and theoretically well grounded system for the error cycling of the High Resolution 4DVar, but is very expensive for current (and future!) resources
2. There are obvious advantages from running a larger ensemble at higher resolution

Typhoon BOLAVEN 2012 – MSLP ENS StDev

T399 EDA

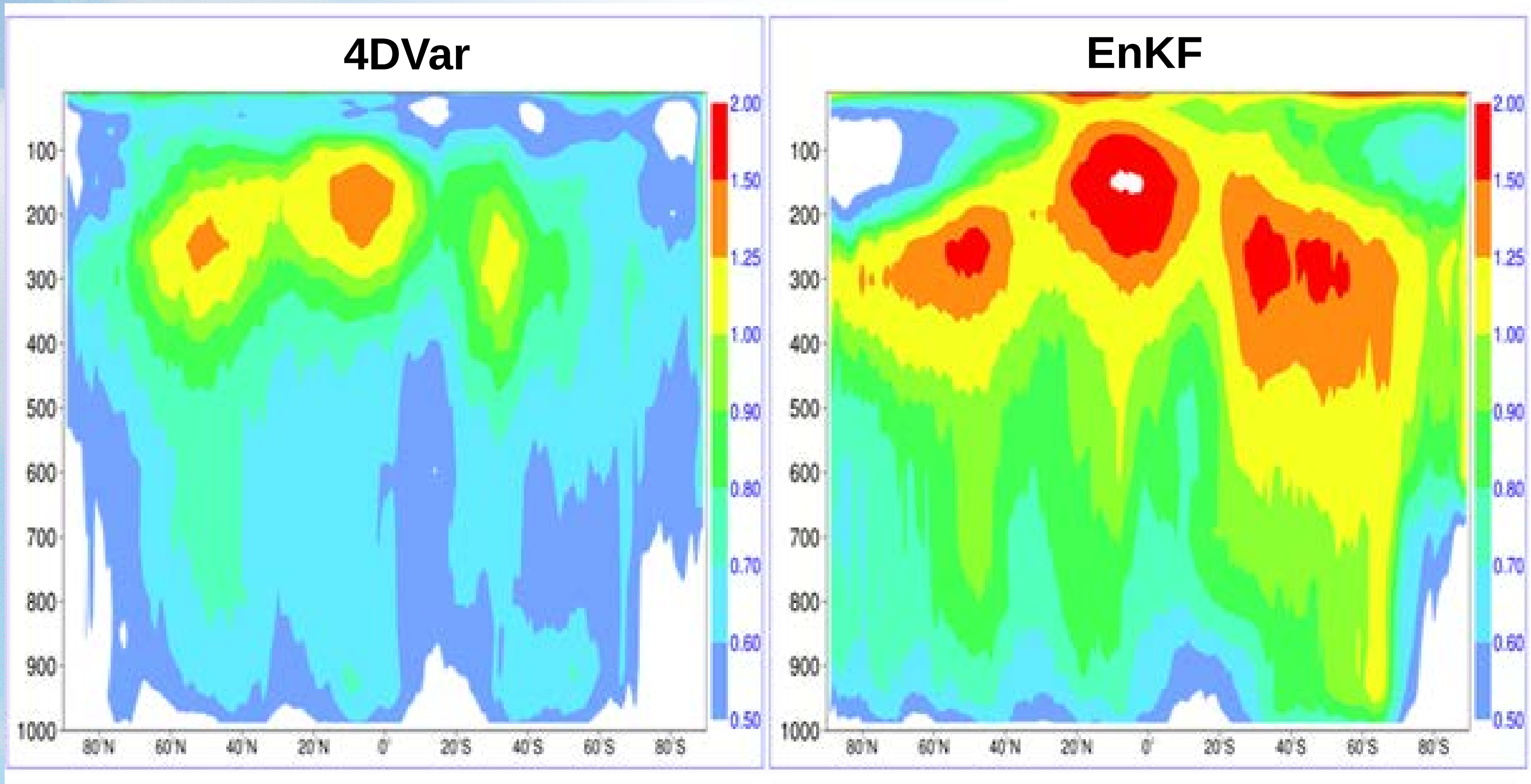


T639 EDA



Hybrid Gain EnDA

- What do 4DVar and EnKF do differently?

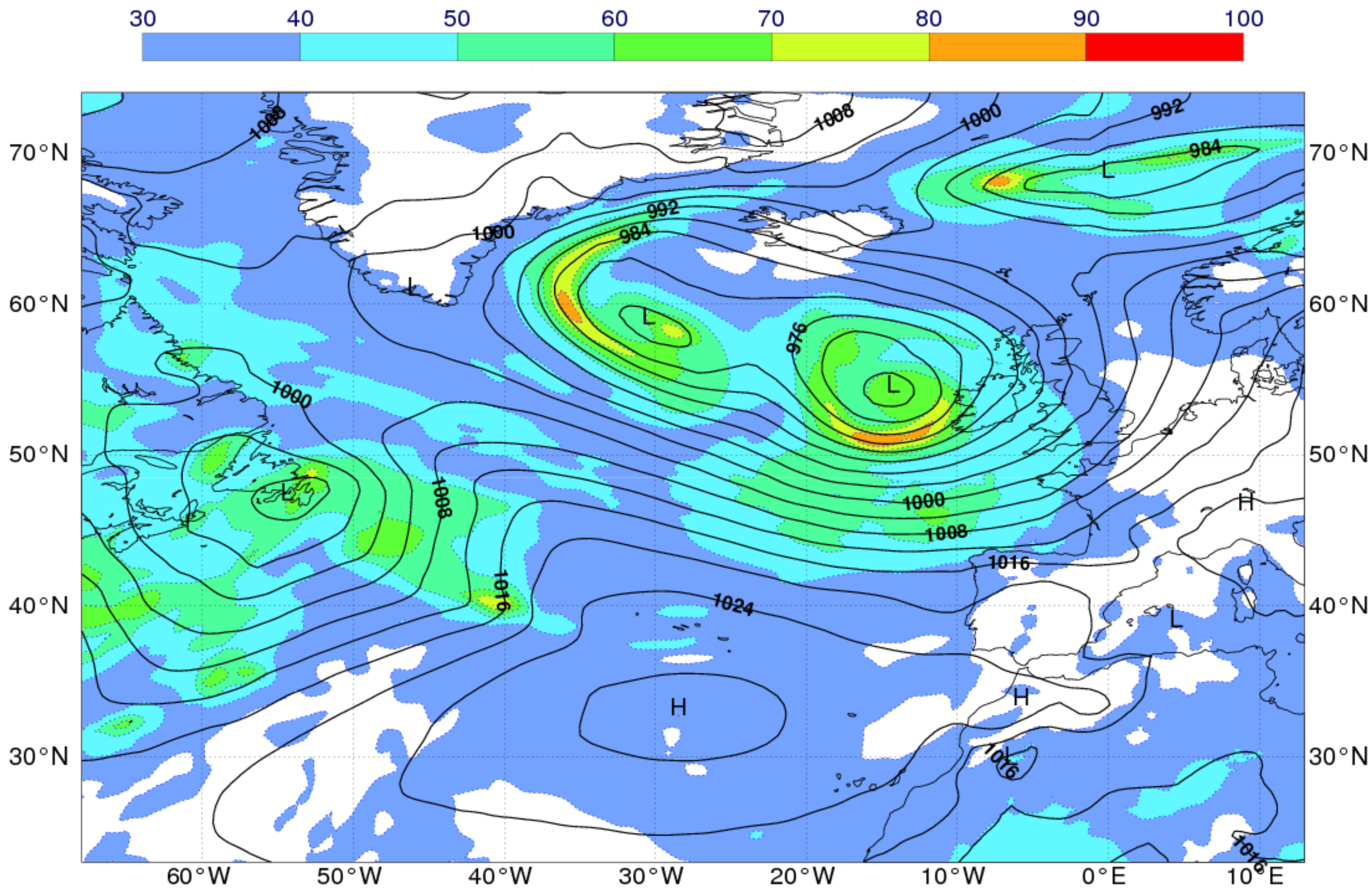


Zonal averages of the standard deviation of the zonal wind analysis increm.

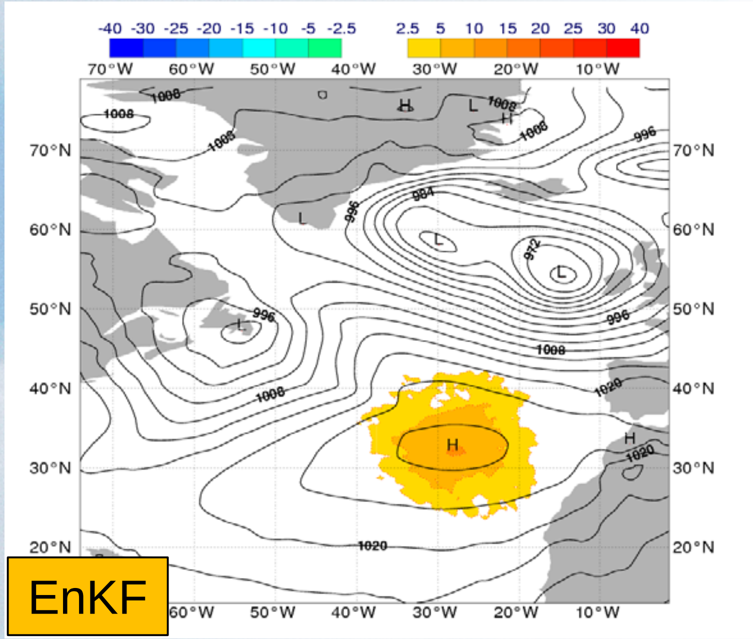
Hybrid Gain EnDA

- Hybrid Gain EnDA works surprisingly well. But why?

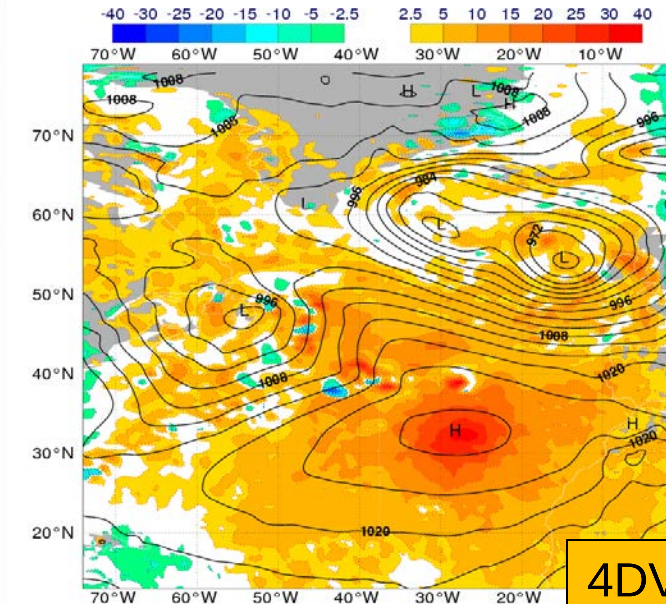
MSLP t+6h fcst and MSLP Ensemble stdev (shaded)
SP obs at (32.0N, 28.0W), middle of window, y-H(x)=+1hPa



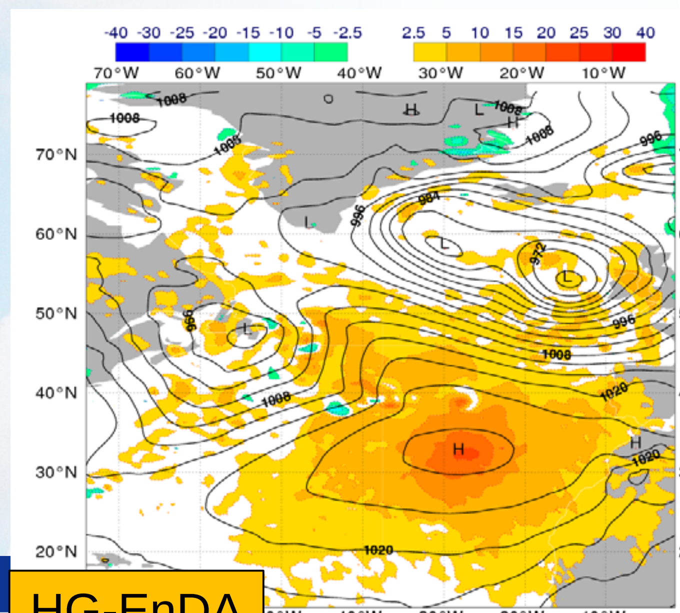
SP obs at (32.0N, 28.0W), middle of window, y-H(x)=+1hPa



EnKF



4DVar



HG-EnDA