Ensemble of Data Assimilations and uncertainty estimation

Massimo Bonavita

ECMWF

Acknowledgments:
Lars Isaksen, Mats Hamrud, Elias Holm, Mike Fisher,
Laure Raynaud, Loik Berre, A. Clayton

Outline

- Why do we need an Ensemble of Data Assimilations?
- Sequential DA methods and Non-Sequential DA methods
- Hybrids methods: the best of both worlds?
- Use of EDA variances in a hybrid DA
- Use of EDA covariances in a hybrid DA
- Conclusions and perspectives

Why do we need an EDA?

A crash course in Data Assimilation!

"*DA is the process through which all the available information is used to estimate as accurately as possible the state of the atmospheric or oceanic flow*" (*Talagrand*, 1997)

A **Bayesian inference problem** (Lorenc, 1986; *Wikle and Berliner*,2007)

If *X* is the state and *Y* our data the full probability model can be factored as:

$$p(x,y) = p(x|y)\,p(y) = p(y|x)\,p(x)$$

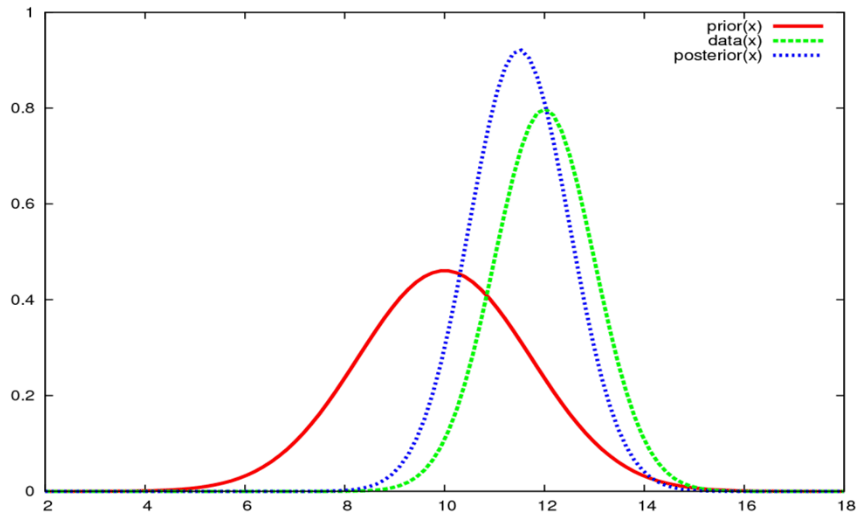
which can be written (*Bayes' Rule*):

$$p(x|y) = \frac{p(y|x)\,p(x)}{p(y)}$$

Why do we need an EDA?

$$p(x\mid y)=\frac{p(y\mid x)\,p(x)}{p(y)}$$

i.e., in order to infer the distribution of the state given the data (*posterior distribution*, *p(x|y)*), we need only form the product of the distributions of measurement errors (*data distribution*, *p(y|x)*) and our prior knowledge about the state (*prior distribution*, *p(x)*). The marginal distribution $p(y)=\int p(y\mid x)\,p(x)dx$ can be thought of as a normalising constant.



ECMWF

Why do we need an EDA?

Let us introduce the time dimension: we want to estimate a set of states $\boldsymbol{X}_{0:t}=[\boldsymbol{X}_0,\boldsymbol{X}_1,..,\boldsymbol{X}_t]$ given all the observations over the same time interval $\boldsymbol{Y}_{1:t}=[\boldsymbol{Y}_1,\boldsymbol{Y}_2,..,\boldsymbol{Y}_t]$, i.e.

$$p(\boldsymbol{x}_{0:t}|\boldsymbol{y}_{1:t}) \propto p(\boldsymbol{y}_{1:t}|\boldsymbol{x}_{0:t})p(\boldsymbol{x}_{0:t})$$

Two hypotheses are commonly introduced:

a) A Markov assumption on the prior distribution, i.e.

$$p(\boldsymbol{x}_{0:t})=p(\boldsymbol{x}_0)\, p(\boldsymbol{x}_1|\boldsymbol{x}_0)\, \ldots p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$$

b) Statistical independence of the observations:

$$p(\boldsymbol{y}_{1:t}|\boldsymbol{x}_{0:t})=p(\boldsymbol{y}_1|\boldsymbol{x}_1)\ldots p(\boldsymbol{y}_t|\boldsymbol{x}_t)$$

This leads to:

$$p(\boldsymbol{x}_{0:t}|\boldsymbol{y}_{1:t}) \propto p(\boldsymbol{x}_0)\, p(\boldsymbol{x}_1|\boldsymbol{x}_0)\, p(\boldsymbol{y}_1|\boldsymbol{x}_1)\ldots\, p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})\, p(\boldsymbol{y}_t|\boldsymbol{x}_{t-1})$$

ECMWF

Why do we need an EDA?

$$p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) \propto p(\mathbf{x}_0)\, p(\mathbf{x}_1|\mathbf{x}_0)\, p(\mathbf{y}_1|\mathbf{x}_1) \ldots p(\mathbf{x}_t|\mathbf{x}_{t-1})\, p(\mathbf{y}_t|\mathbf{x}_{t-1})$$

This form naturally leads to a sequential algorithm, i.e., when new observations are available the state is updated from the previously available estimate.

In real time applications we are mainly concerned with the Filtering problem: Knowing $p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})$ how does a new batch of observations $\mathbf{Y}_t$ change our estimate of the state?

Two step procedure:

1. Compute the forecast distribution at time t (*forecast step*) :

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1})\, p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}) \mathrm{d}\mathbf{x}_{t-1}$$

2. Compute the analysis distribution (*analysis step*) :

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = p(\mathbf{x}_t|\mathbf{y}_t,\mathbf{y}_{1:t-1}) \propto p(\mathbf{y}_t|\mathbf{x}_t,\mathbf{y}_{1:t-1}) p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = p(\mathbf{y}_t|\mathbf{x}_t)\, p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$$

ECMWF

Why do we need an EDA?

For Gaussian error distributions and linear model and observation operators we recover the Kalman Filter (KF) equations:

$$\mathbf{x}_t = \mathbf{M}_{t-1,t}\mathbf{x}_{t-1} + \boldsymbol{\eta}_{t-1,t} \qquad \boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0},\mathbf{Q}_{t-1,t}) \qquad (1)$$

$$\mathbf{y}_t = \mathbf{H}_t\mathbf{x}_t + \boldsymbol{\varepsilon}_t \qquad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0},\mathbf{R}_t) \qquad (2)$$

Models (1) and (2) give the prior p($\boldsymbol{x}_t|\boldsymbol{x}_{t-1}$) and data p($\boldsymbol{y}_t|\boldsymbol{x}_t$) distributions, so that the forecast distribution $\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1} \sim \mathcal{N}(\boldsymbol{x}_{t|t-1},\boldsymbol{P}_{t|t-1})$:

$$\mathbf{x}_{t|t-1} = \mathbf{M}_{t-1,t}\boldsymbol{x}_{t-1} \qquad (3)$$

$$\mathbf{P}_{t|t-1} = \mathbf{M}_{t-1|t}\mathbf{P}_{t|t-1}\mathbf{M}^{\mathrm{T}}_{t-1|t} + \mathbf{Q}_{t-1|t} \qquad (4)$$

The analysis distribution $\boldsymbol{x}_t|\boldsymbol{y}_{1:t} \sim \mathcal{N}(\boldsymbol{x}_{t|t},\boldsymbol{P}_{t|t})$ is given by:

$$\mathbf{x}_{t|t} = \mathbf{x}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t - \mathbf{H}_t\,\mathbf{x}_{t|t-1}) \qquad (5)$$

$$\mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\mathbf{P}_{t|t-1} \qquad (6)$$

$$\mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{H}^{\mathrm{T}}_t(\mathbf{R}_t + \mathbf{H}_t\mathbf{P}_{t|t-1}\mathbf{H}^{\mathrm{T}}_t)^{-1} \qquad (7)$$

ECMWF

Why do we need an EDA?

We may be interested in the distribution $p(\boldsymbol{x}_t|\boldsymbol{y}_{1:T})$ for t=1,…,T, i.e. we want to estimate the state using observations both before and after time t (smoothing distribution). Under the same hypothesis used for the Kalman filter, a Kalman smoother (KS) can be derived (*Cosme et al.*, 2011).

Two aspects need to be emphasised:

a) The Kalman smoother differs from the filter only by using cross-covariances in time to correct the state at time t using observations at future times;

b) At the *end of the assimilation window* (t=T) *the KS and KF estimates are the same*

Why do we need an EDA?

The KF is the optimal solution of the filtering problem for linear, Gaussian systems.

Unfortunately it is impractical for large systems: a current NWP system has $N \sim 10^8$. In the KF we have to compute and evolve in time error covariances of $N x N$ dimension!

Two possible types of solutions:

a) 4D Variational methods

b) Reduced-rank Kalman Filters

Why do we need an EDA?

4D Variational methods

If we neglect model error (perfect model assumption) the smoothing problem of finding the model trajectory that best fits the observations over an assimilation interval (t=0,1,…,T) given a background state $\mathbf{x_b}$ and its error covariance $\mathbf{P^b}$ is the minimum of the cost function:

$$J(\mathbf{x}_0) = (\mathbf{x}_b - \mathbf{x}_o)^T (\mathbf{P}^b)^{-1} (\mathbf{x}_b - \mathbf{x}_o) + \sum_{t=0}^{T} (\mathbf{y}_t - H_t M_{0\to t}(\mathbf{x}_0))^T \mathbf{R}_t^{-1} (\mathbf{y}_t - H_t M_{0\to t}(\mathbf{x}_0))$$

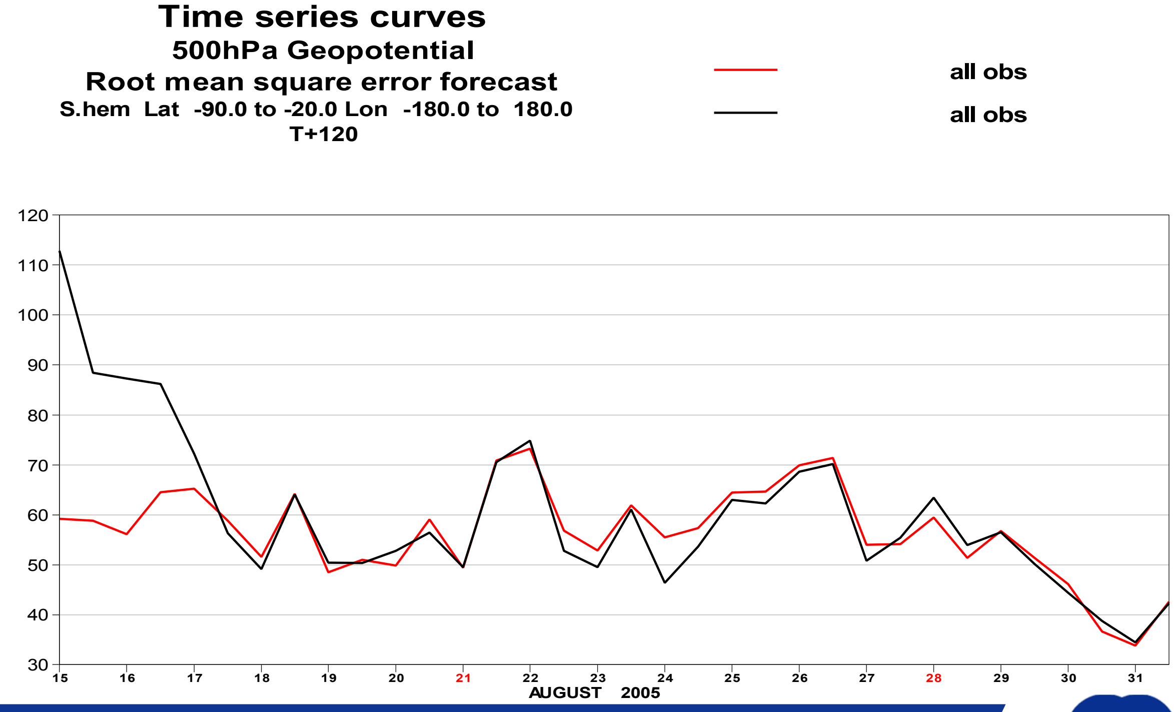
This is equivalent to the Kalman smoother solution over the assimilation interval for the same $\mathbf{x_b}$, $\mathbf{P^b}$ and to the Kalman filter solution at the end of the interval (t=T).
The 4D-Var solution implicitly evolves background error covariances *over the assimilation window* (*Thepaut et al*.,1996), but *does not cycle them*! Information from past observations is only carried forward by $\boldsymbol{x_b}$

Why do we need an EDA?

4D Variational methods

What if we pushed back the start of the assimilation window 'enough' so that the smoothed solution (and the filter solution at the end of the window) would no longer depend on the specified $\boldsymbol{P}^b$?

Enough means 3-5 days for state of art NWP models:



Why do we need an EDA?

4D Variational methods

For assimilation windows > 12h we can not assume the model to be perfect any more. We have to add a model error term to our cost function (Weak-constraint 4D-Var):

$$J(\mathbf{x}_0,\mathbf{x}_1,\ldots,\mathbf{x}_T)=(\mathbf{x}_b-\mathbf{x}_o)^T(\mathbf{P}^b)^{-1}(\mathbf{x}_b-\mathbf{x}_o)+\sum_{t=0}^{T}(\mathbf{y}_t-H_t(\mathbf{x}_t))^T\mathbf{R}_t^{-1}(\mathbf{y}_t-H_t(\mathbf{x}_t))+$$

$$\sum_{t=0}^{T}(\mathbf{x}_t-M_{t-1->t}(\mathbf{x}_{t-1}))^T\mathbf{Q}_t^{-1}(\mathbf{x}_t-M_{t-1->1}(\mathbf{x}_{t-1}))$$

This is an elegant solution, but:

1) Problem is shifted from estimation of $\mathbf{P}^\mathbf{b}$ to estimation of $\mathbf{Q}$.
   $\mathbf{Q}$ can also have a non negligible flow-dependent component
2) It remains difficult in the 4D-Var framework to have realistic estimates of $\mathbf{P}^\mathbf{a}$

ECMWF

Why do we need an EDA?

Reduced-rank Kalman Filters

In order to remain in the sequential paradigm we need to use the Kalman Filter analysis with a low-rank approximation of $\mathbf{P}^{b/a}$

In this framework we look for a low-rank approximation to $\mathbf{P}^b$ of the form

$$\mathbf{P}_t^b = \mathbf{X}^b (\mathbf{X}^b)^T \quad \text{where } \mathbf{X}^b \text{ is } Nxm \text{ and } m << N$$

It then follows that

$$\mathbf{K} = \mathbf{P}_t^b \mathbf{H}^T [\mathbf{H}\mathbf{P}^b\mathbf{H}\mathbf{T} + \mathbf{R}]^{-1} = \mathbf{X}^b (\mathbf{H}\mathbf{X}^b)^T [(\mathbf{H}\mathbf{X}^b)(\mathbf{H}\mathbf{X}^b)^T + \mathbf{R}]^{-1}$$

$$\mathbf{P}_t^a = \mathbf{X}^b [\mathbf{I}_{mxm} + (\mathbf{H}\mathbf{X}^b)^T \mathbf{R} (\mathbf{H}\mathbf{X}^b)] (\mathbf{X}^b)^T$$

$$\mathbf{P}_{t+1}^b = \mathbf{M}_{t->t+1} \mathbf{P}_t^a \mathbf{M}^T_{t->t+1} + \mathbf{Q}_{t->t+1} = \mathbf{M}_{t->t+1} \mathbf{X}^b [\mathbf{A}_{mxm}] (\mathbf{M}_{t->t+1} \mathbf{X}^b)^T + \mathbf{Q}_{t->t+1}$$

i.e., dimension N has been replaced by m in the KF equations!

However…

Why do we need an EDA?

Reduced-rank Kalman Filters

However…

$$\mathbf{x_a}-\mathbf{x_b} = \mathbf{K}\,(\mathbf{y} - \mathrm{H}(\mathbf{x_b})) = \mathbf{X^b}\,(\mathbf{HX^b})^T\,[(\mathbf{HX^b})(\mathbf{HX^b})^T + \mathbf{R}]^{-1}\,(\mathbf{y} - \mathrm{H}(\mathbf{x_b}))$$

It then follows that the analysis increments are confined to the subspace spanned by $\mathbf{X^b}$, which has rank $m<<N$

Reduced-rank KF became popular only with the introduction of the Ensemble Kalman Filter (EnKF, *Evensen*, 1994; *Burgers et al*., 1998)

EnKF is a Monte Carlo approx. of the KF which crucially does not require the Tangent Linear and Adjoints of $\mathbf{M}$ and $\mathbf{H}$.

But the subspace spanned by $\mathbf{P^b}_{\mathbf{ens}}= 1/\sqrt{(N_{ens}-1)}\,\mathbf{X^{b\prime}}(\mathbf{X^{b\prime}})^T$, ($\mathbf{X^{b\prime}}$ are the ensemble perturbations to the ensemble mean) has still dimension $N_{ens}-1 << N$

ECMWF

Why do we need an EDA?

Reduced-rank Kalman Filters

$$\mathbf{P}^{b}_{ens} = 1/\sqrt{(N_{ens}-1)}\ \mathbf{X}^{b\prime}(\mathbf{X}^{b\prime})^{T}\ ,\ N_{ens}-1 << N$$

There are ways to artificially increase the effective ensemble size (Shur product covariance localization, Hamill and Whitaker, 2001; Local analysis, Evensen, 2003; Ott et al., 2004; adaptive localization, Anderson 2007, Bishop and Hodyss, 2007,2009), but they (too!) come at a price:

a) Dynamical balance may be degraded;

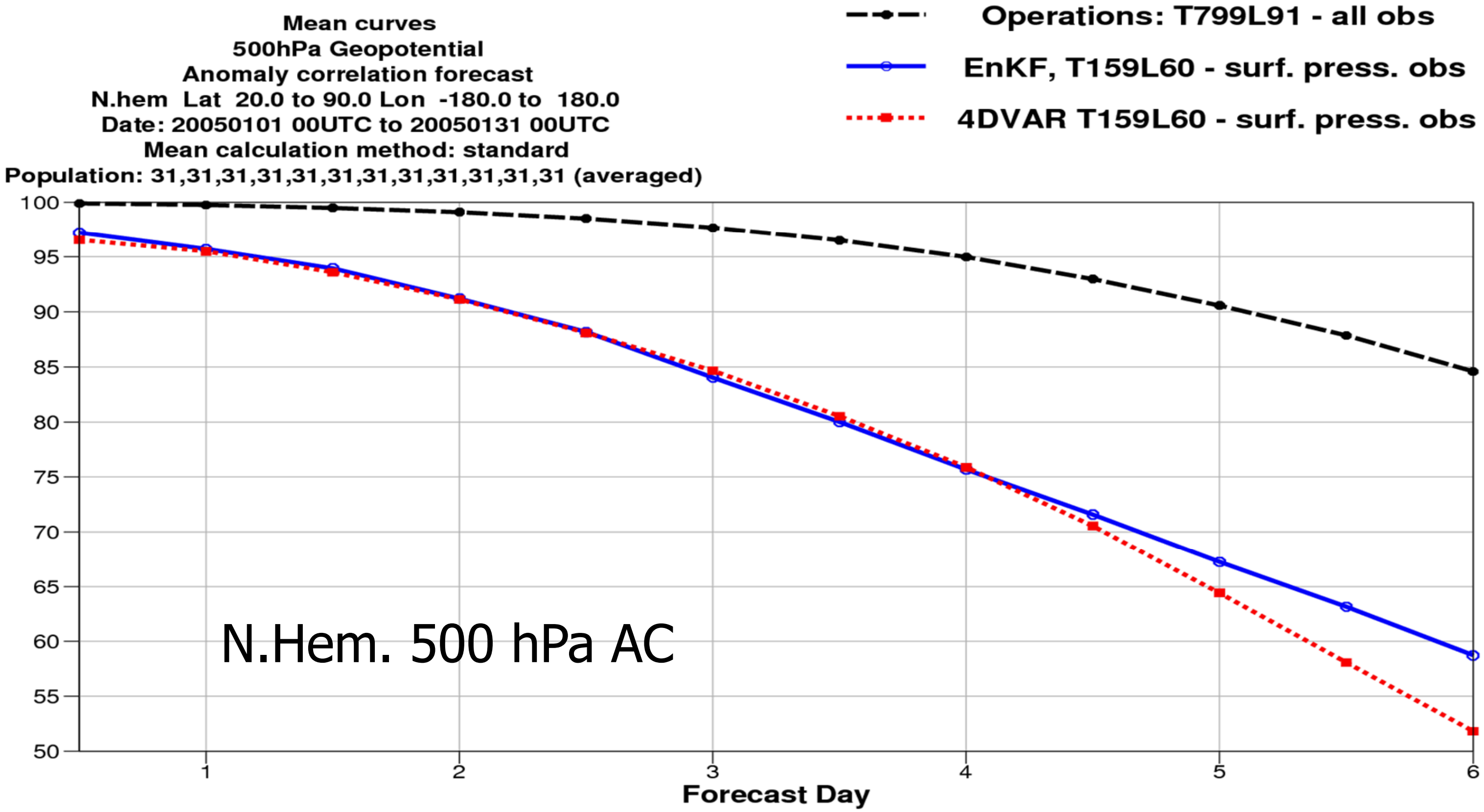
b) Asymptotic optimality of the EnKF lost;

c) More difficult for non-local observations, since usually applied in observation space

Why do we need an EDA?

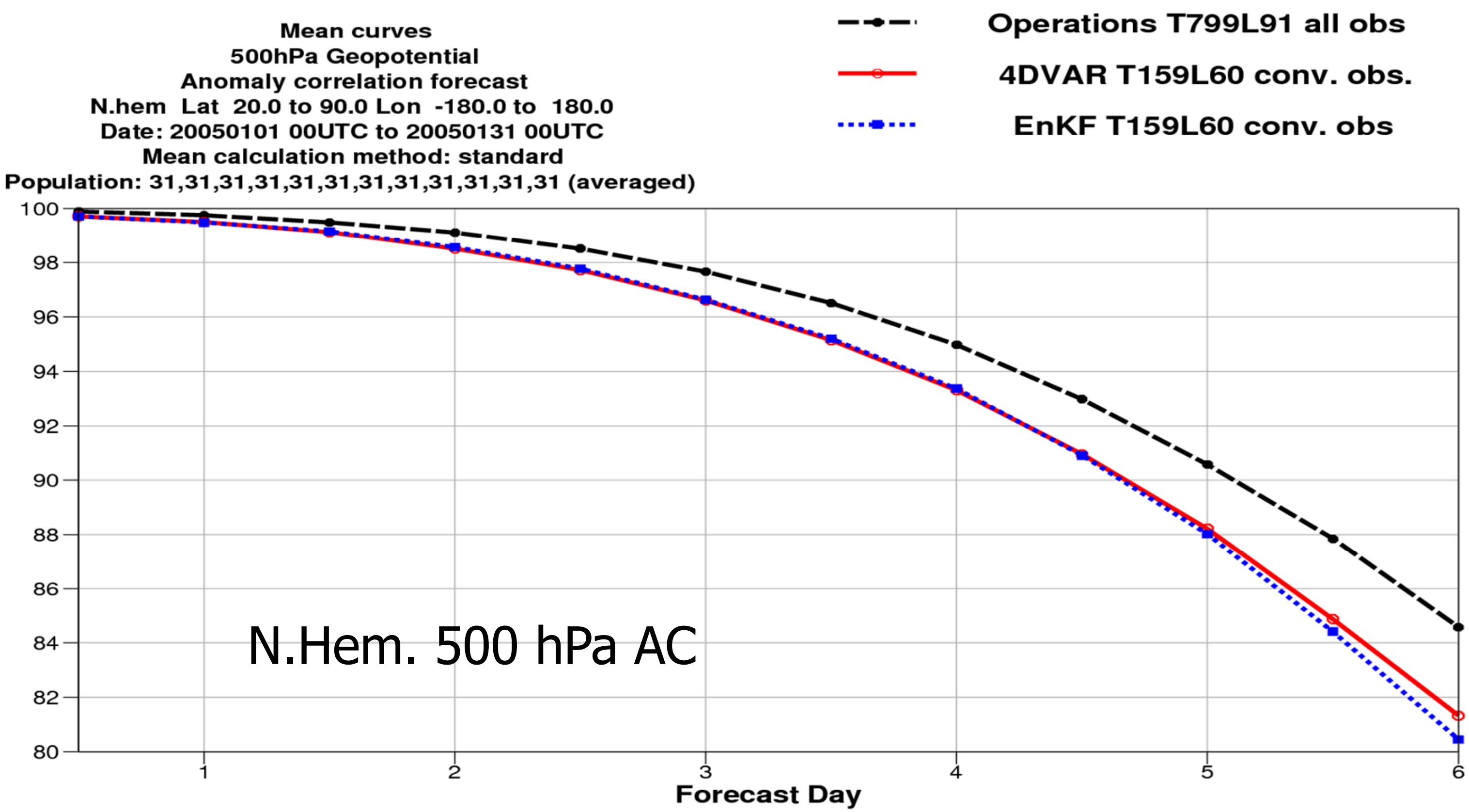
Results with the ECMWF EnKF

Surface Pressure observations only



Why do we need an EDA?

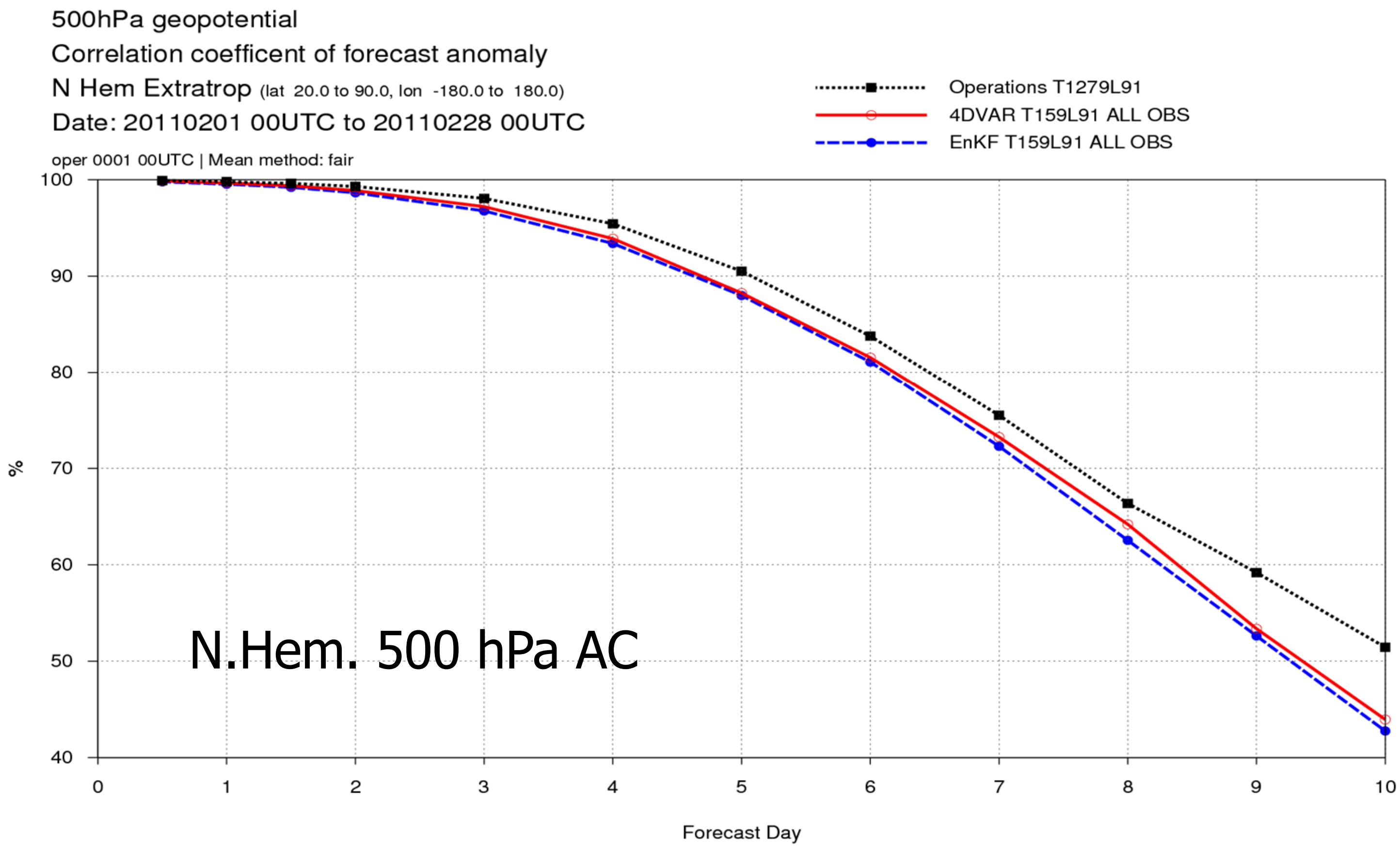
Results with the ECMWF EnKF
Conventional observations only



Why do we need an EDA?

Results with the ECMWF EnKF

All observations



ECMWF

Why do we need an EDA?

Quick recap:

a) Kalman Filter is computationally unfeasible for realistic NWP;

b) Non-sequential approx. (4D-Var) do not cycle state error estimates: work well for short assimilation windows (6-12h), but longer windows have proved more difficult;

c) Sequential approx. (EnKF) cycle low-rank estimates of state error covariances, but analysis increments are confined to perturbations subspace;

….

Hybrid approach: Use flow-dependent state error estimates (from an EnKF/EDA system) in a 3/4D-Var analysis algorithm

Hybrids

Hybrid approx.: Use flow-dependent state error estimates (from an EnKF/EDA system) in a 3/4D-Var analysis algorithm

This solution would:

1) Integrate flow-dependent state error covariance information into the variational analysis
2) Keep the full rank representation of $\mathbf{B}$ and its implicit evolution inside the assimilation window
3) More robust than pure EnKF for limited ensemble sizes and large model errors
4) Allow (eventual) localization of ensemble perturbations to be performed in state space;
5) Allow for flow-dependent QC of observations

Hybrids

In operational use (or in an advanced testing), there are currently two main approaches to doing an hybrid DA in a VAR context:

1. Alpha control variable method (Met Office, NCEP/GMAO)
2. Ensemble of Data Assimilations method (ECMWF, Meteo France)

Hybrids: α control variable

1. Alpha control variable method (Met Office, NCEP/GMAO)

Conceptually add a flow-dependent term to the climatological **B** matrix:

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

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

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

In practice this is done through augmentation of 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:

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

from: A.Clayton

Hybrids: EDA

The Ensemble of Data Assimilations (EDA,Isaksen et al. 2010) can be considered a **flow-dependent extension** of the way the *climatological background error matrix* is estimated (Fisher, 2003).

For a linear system the data assimilation update is:

$$\mathbf{x}_a^k = \mathbf{x}_b^k + \mathbf{K}_k\left(\mathbf{y}^k - \mathbf{H}_k \mathbf{x}_b^k\right)$$

$$\mathbf{x}_b^{k+1} = \mathbf{M}_k \mathbf{x}_a^k$$

$$\mathbf{P}_k^a = \left(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k\right)\mathbf{P}_k^b\left(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k\right)^T + \mathbf{K}_k \mathbf{R}_k \mathbf{K}_k^T$$

$$\mathbf{P}_{k+1}^b = \mathbf{M}_k \mathbf{P}_k^a \mathbf{M}_k^T + \mathbf{Q}_k$$

In our system the *sources of error* are the observations and the forecast model:

$$\mathbf{x}_b^{k+1} = \mathbf{M}_k \mathbf{x}_a^k + \boldsymbol{\eta}_k \qquad \boldsymbol{\eta}_k \sim \mathcal{N}(\mathbf{0},\mathbf{Q}_k)$$

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \zeta_k \qquad \zeta_k \sim \mathcal{N}(\mathbf{0},\mathbf{R}_k)$$

Hybrids: EDA

Consider now the evolution of the same system where we perturb the observations and the model evolution with random noise drawn from the respective error covariances:

$$\tilde{\mathbf{x}}_a^k = \tilde{\mathbf{x}}_b^k + \mathbf{K}_k\left(\mathbf{y}^k + \boldsymbol{\eta}_k - \mathbf{H}_k\tilde{\mathbf{x}}_b^k\right)$$
$$\tilde{\mathbf{x}}_b^{k+1} = \mathbf{M}_k\tilde{\mathbf{x}}_a^k + \zeta_k$$

where $\boldsymbol{\eta}_k \sim \mathrm{N}(\mathbf{0},\mathbf{R})$, $\zeta_k \sim \mathrm{N}(\mathbf{0},\mathbf{Q})$.

If we define the differences between the perturbed and unperturbed state $\boldsymbol{\varepsilon}_a \equiv \tilde{\mathbf{x}}_a - \mathbf{x}_a$ and $\boldsymbol{\varepsilon}_b \equiv \tilde{\mathbf{x}}_b - \mathbf{x}_b$, their evolution is obtained by subtracting the unperturbed state evolution equations from the perturbed ones:

$$\boldsymbol{\varepsilon}_a^k = \boldsymbol{\varepsilon}_b^k + \mathbf{K}_k\left(\boldsymbol{\eta}_k - \mathbf{H}_k\boldsymbol{\varepsilon}_b^k\right)$$
$$\boldsymbol{\varepsilon}_b^{k+1} = \mathbf{M}_k\boldsymbol{\varepsilon}_a^k + \zeta_k$$

Hybrids: EDA

$$\boldsymbol{\varepsilon}_a^k = \boldsymbol{\varepsilon}_b^k + \mathbf{K}_k\left(\boldsymbol{\eta}_k - \mathbf{H}_k\boldsymbol{\varepsilon}_b^k\right)$$
$$\boldsymbol{\varepsilon}_b^{k+1} = \mathbf{M}_k\boldsymbol{\varepsilon}_a^k + \boldsymbol{\zeta}_k$$

i.e., the perturbations evolve with the same update equations of the state

What about the errors?

If we take the statistical expectation of the outer product of the perturbations:

$$\left\langle \boldsymbol{\varepsilon}_k^a\left(\boldsymbol{\varepsilon}_k^a\right)^T \right\rangle = \left(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k\right)\left\langle \boldsymbol{\varepsilon}_k^b\left(\boldsymbol{\varepsilon}_k^b\right)^T \right\rangle\left(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k\right)^T + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T$$
$$\left\langle \boldsymbol{\varepsilon}_{k+1}^b\left(\boldsymbol{\varepsilon}_{k+1}^b\right)^T \right\rangle = \mathbf{M}_k\left\langle \boldsymbol{\varepsilon}_k^a\left(\boldsymbol{\varepsilon}_k^a\right)^T \right\rangle\mathbf{M}_k^T + \mathbf{Q}_k$$

Hybrids: EDA

$$\left\langle \boldsymbol{\varepsilon}_k^a \left(\boldsymbol{\varepsilon}_k^a\right)^T \right\rangle = \left(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k\right) \left\langle \boldsymbol{\varepsilon}_k^b \left(\boldsymbol{\varepsilon}_k^b\right)^T \right\rangle \left(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k\right)^T + \mathbf{K}_k \mathbf{R}_k \mathbf{K}_k^T$$

$$\left\langle \boldsymbol{\varepsilon}_{k+1}^b \left(\boldsymbol{\varepsilon}_{k+1}^b\right)^T \right\rangle = \mathbf{M}_k \left\langle \boldsymbol{\varepsilon}_k^a \left(\boldsymbol{\varepsilon}_k^a\right)^T \right\rangle \mathbf{M}_k^T + \mathbf{Q}_k$$

These are the same equations for the evolution of the system error covariances:

$$\mathbf{P}_k^a = \left(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k\right) \mathbf{P}_k^b \left(\mathbf{I} - \mathbf{K}_k \mathbf{H}_k\right)^T + \mathbf{K}_k \mathbf{R}_k \mathbf{K}_k^T$$

$$\mathbf{P}_{k+1}^b = \mathbf{M}_k \mathbf{P}_k^a \mathbf{M}_k^T + \mathbf{Q}_k$$

provided that:

1. The applied perturbations $\boldsymbol{\eta}_k$, $\boldsymbol{\zeta}_k$ have the required covariances (**R**, **Q**);
2. At some stage in time $\left\langle \boldsymbol{\varepsilon}_k^b \left(\boldsymbol{\varepsilon}_k^b\right)^T \right\rangle = \mathbf{P}_k^b$ (asymptotic convergence, Fisher *et al.*, 2005)

Hybrids: EDA

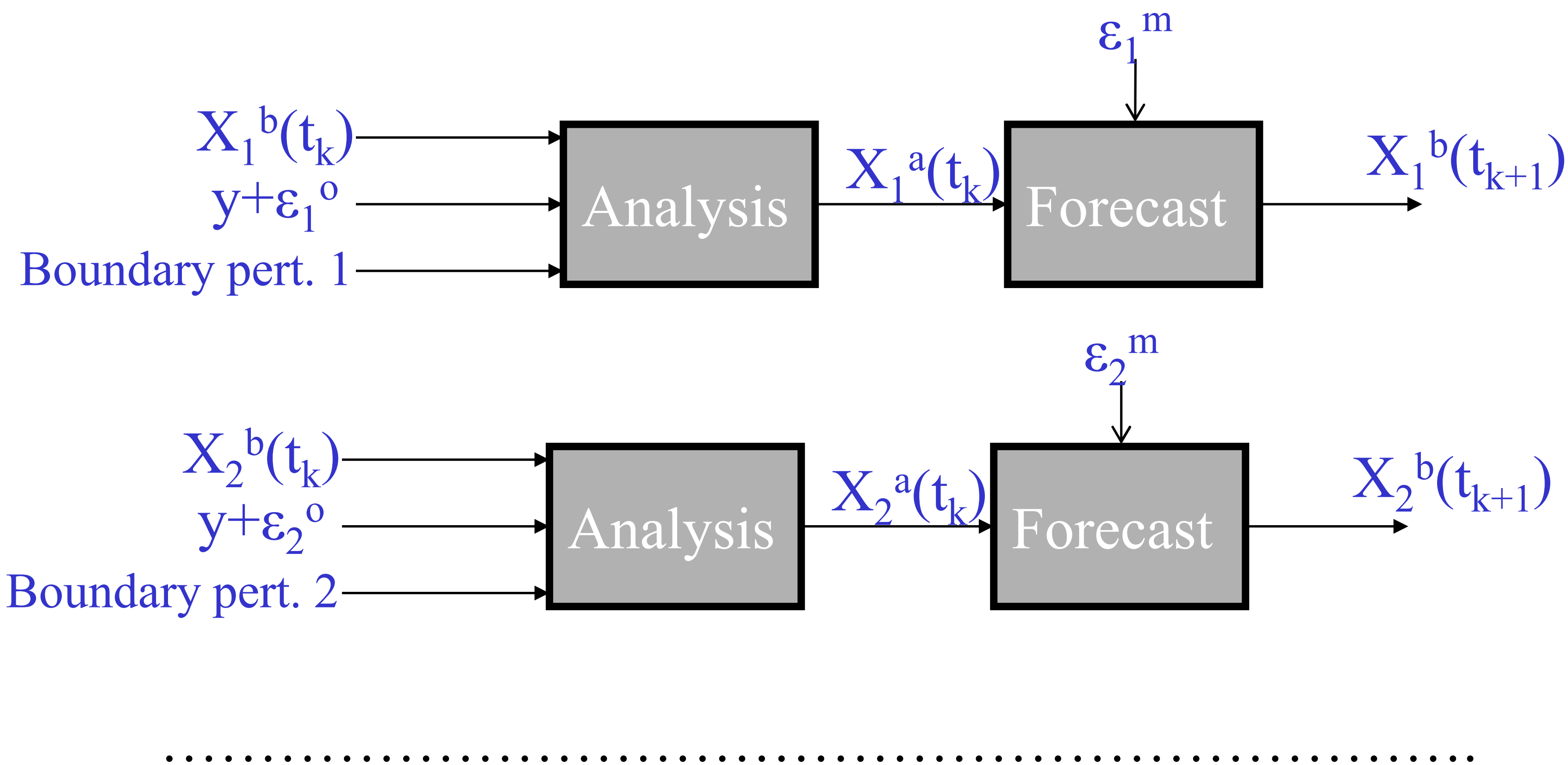
What does all this mean in practice?

1. We can use an ensemble of perturbed 4D-Var to simulate the ***errors*** of our reference high resolution 4D-Var;

2. The ensemble of perturbed DAs should be as similar as possible to the reference DA (i.e., same or similar **K** matrix)

3. The applied perturbations $\boldsymbol{\eta}_k$, ζ_k must have the required error covariances ($\mathbf{R}$, $\mathbf{Q}$); however we do not need an explicit covariance model of $\mathbf{Q}$

Hybrids: EDA



ECMWF

Hybrids: EDA

- **10** ensemble members using 4D-Var assimilations
- **T399** outer loop, T95/T159 inner loops. (Reference DA: T1279 outer loop, T159/T255/T255 inner loops)
- Observations randomly perturbed according to their specified $\mathbf{R}$
- SST perturbed with realistically scaled structures
- Model error represented by stochastic methods (**SPPT,** Leutbecher, 2009)

Hybrids: EDA

The EDA system simulates **the error evolution** of the 4DVar analysis cycle.

As such it has two main applications:

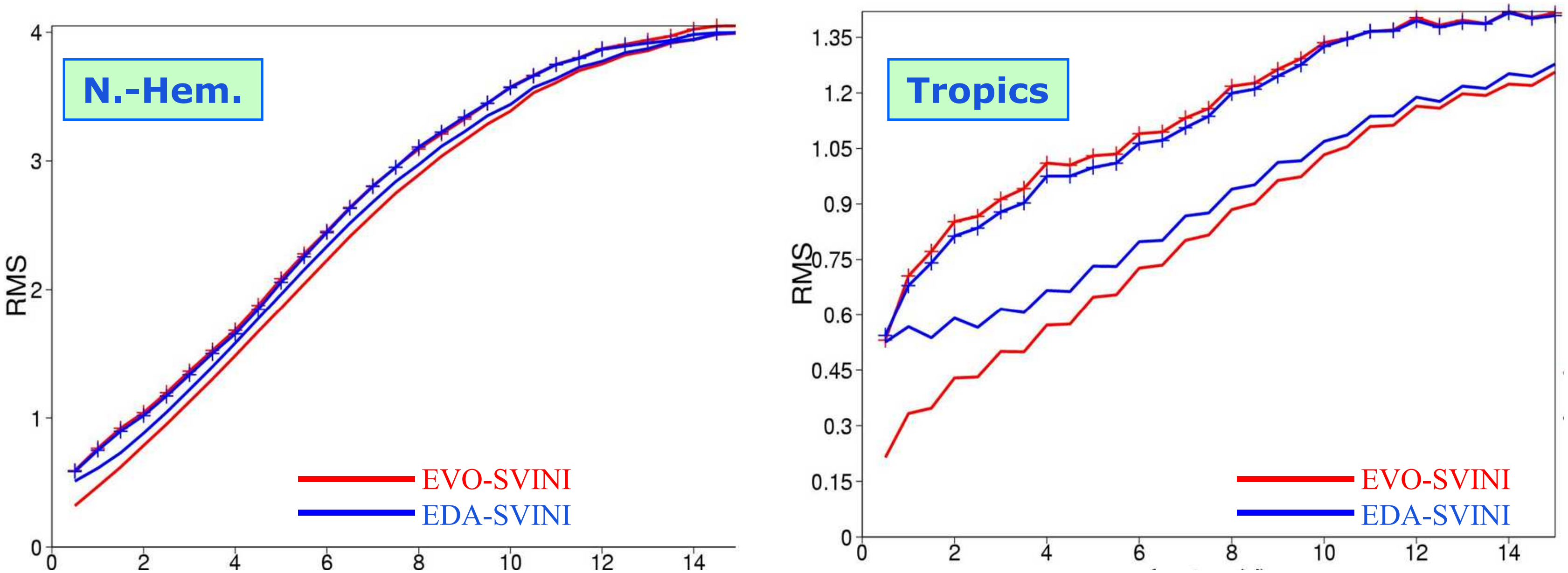
1. Provide a flow-dependent sample of analysis errors to use as initial perturbations for the Ensemble Prediction system (EPS)

2. Provide a flow-dependent sample of background errors at the initial time of the 4D-Var assimilation window

Hybrids: EDA

Improving Ensemble Prediction System by including EDA perturbations for initial uncertainty

The Ensemble Prediction System (EPS) benefits from using EDA based perturbations. Replacing evolved singular vector perturbations by EDA based perturbations improve EPS spread, especially in the tropics.
The Ensemble Mean has slightly lower error when EDA is used.



Ensemble spread and Ensemble mean RMSE for 850hPa T

Hybrids: EDA

The EDA system simulates **the error evolution** of the 4DVar analysis cycle.

As such it has two main applications:

1. Provide a flow-dependent sample of analysis errors to use as initial perturbations for the Ensemble Prediction system (EPS)

2. Provide a flow-dependent sample of background errors at the initial time of the 4D-Var assimilation window

Hybrids: EDA

In the ECMWF 4D-Var, the $\mathbf{B}$ matrix is defined implicitly in terms of a transformation from the background departure $(\mathbf{x}\text{-}\mathbf{x_b})$ to a control variable χ:

$$(\mathbf{x}\text{-}\mathbf{x_b}) = \mathbf{L}\chi$$

So that the implied $\mathbf{B}=\mathbf{L}\mathbf{L}^{\mathbf{T}}$.

In the current wavelet formulation (Fisher, 2003), the variable transform can be written as:

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

$\mathbf{T}$ is the balance operator

$\mathbf{\Sigma}_\mathrm{b}$ is the gridpoint variance of background errors

$\mathbf{C_j}(\lambda,\varphi)$ is the vertical covariance matrix for wavelet index j

ψ_j are the set of radial basis function that define the wavelet transform.

Hybrids: EDA

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

$\mathbf{C}_j(\lambda,\varphi)$ are full vertical covariance matrices, function of (λ,φ). They determine both the horizontal and vertical background error *correlation structures*;

In standard 4D-Var $\mathbf{T}$ and $\mathbf{C}_j$ are computed off-line using a climatology of EDA perturbations.

$\mathbf{\Sigma}_b$ is computed by random sampling of the static $\mathbf{B}$ matrix (randomization procedure, Fisher and Courtier, 1995)

How do we make this error covariance model flow-dependent?

We look for flow-dependent EDA estimates of $\mathbf{\Sigma}_b$ and $\mathbf{C}_j(\lambda,\varphi)$

EDA variances

Σ_b is defined in grid space; it can be directly sampled from the EDA background forecasts:

$$\boldsymbol{\Sigma}_b(i,j,k)=\frac{1}{N_{EDA}-1}\sum_{l=1}^{N_{EDA}}\left(\mathbf{x}_b^l(i,j,k)-\overline{\mathbf{x}}_b(i,j,k)\right)^2$$

However the sampled variance estimates are affected by two errors:

a) Sampling Noise due to the small EDA dimensionality (N_{eda}=10):

$$\hat{\sigma}_{\Sigma_b}=\sqrt{\frac{2}{N_{EDA}-1}}\boldsymbol{\Sigma}_b$$

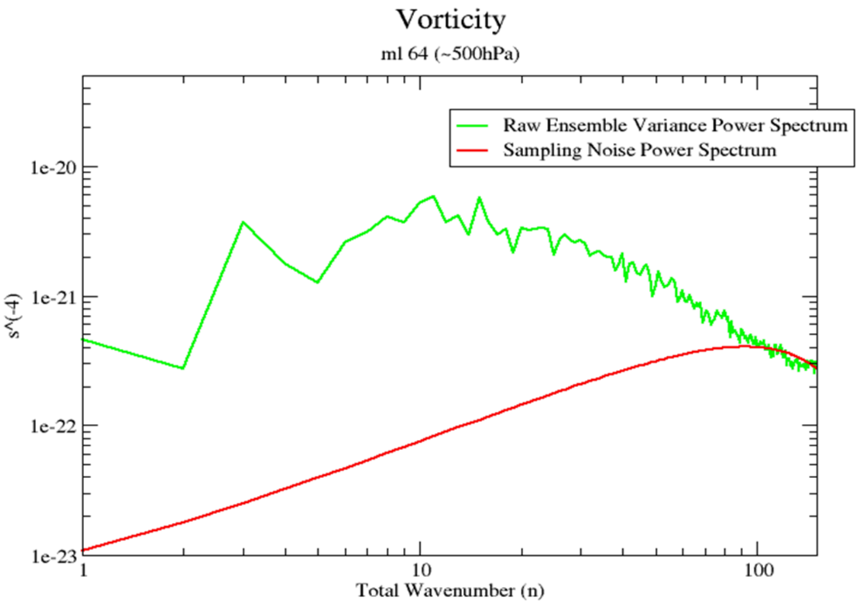
b) Systematic errors due to incorrect specification of error sources in the EDA (i.e., mis-specification of $\mathbf{R}$, $\mathbf{Q}$, uncertainties in the boundary conditions)

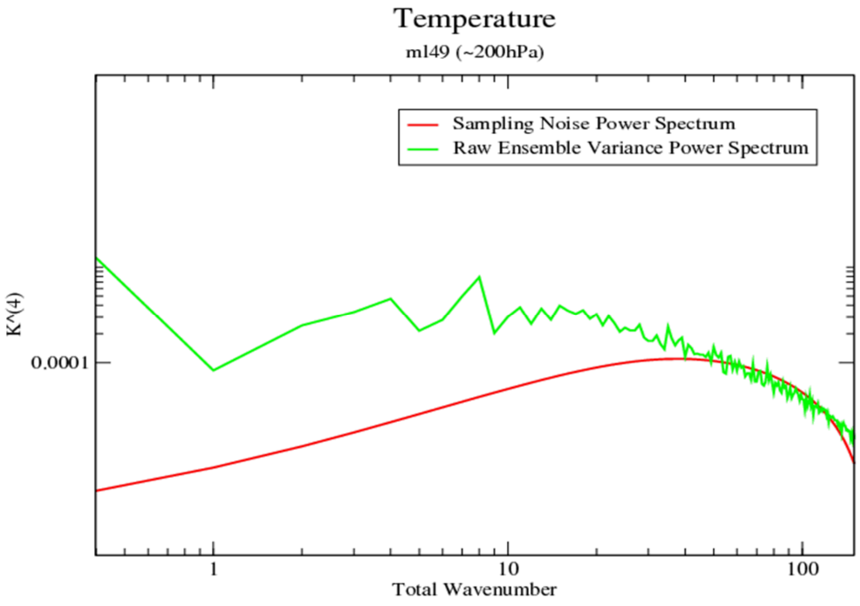
EDA variances

a) Sampling Noise due to the small EDA dimensionality (N_{eda}=10)

The key insight is to recognise that *sampling noise is small scale with respect to the error variance field* (Raynaud *et al.*, 2008)

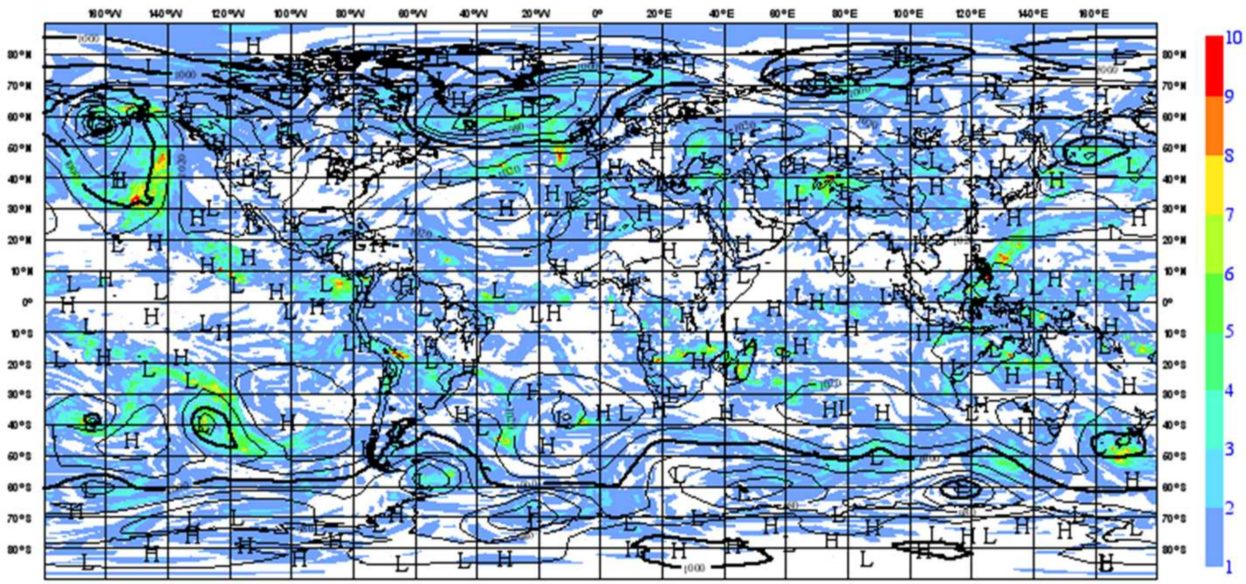
We may use a **spectral filter** to disentangle noise error from the signal



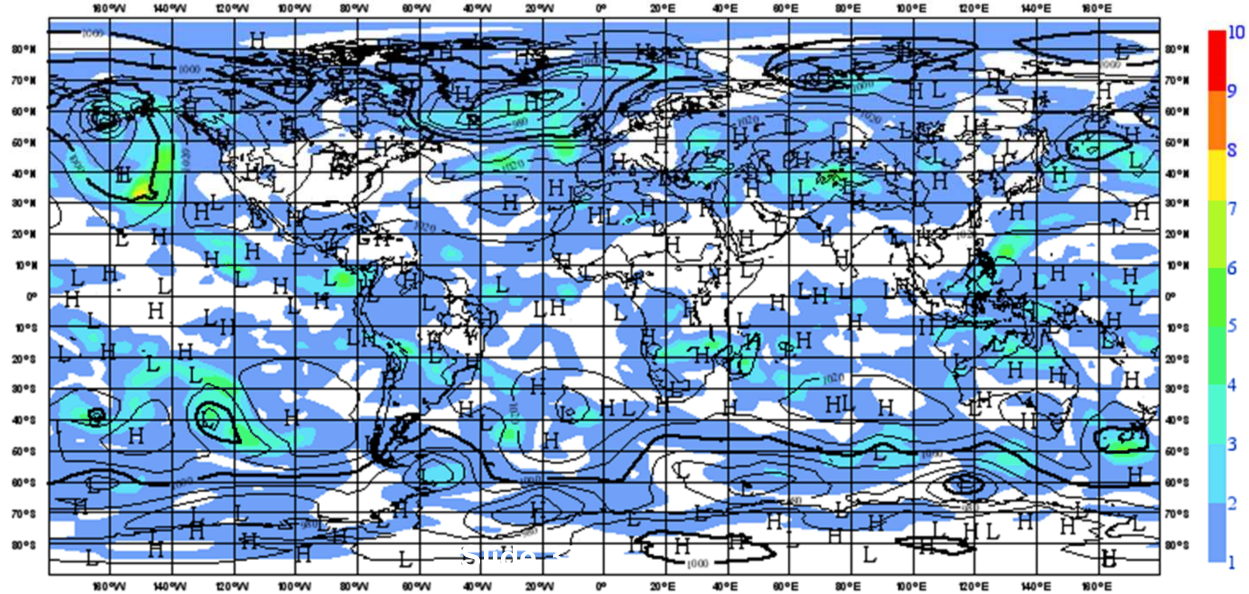


ECMWF

EDA variances

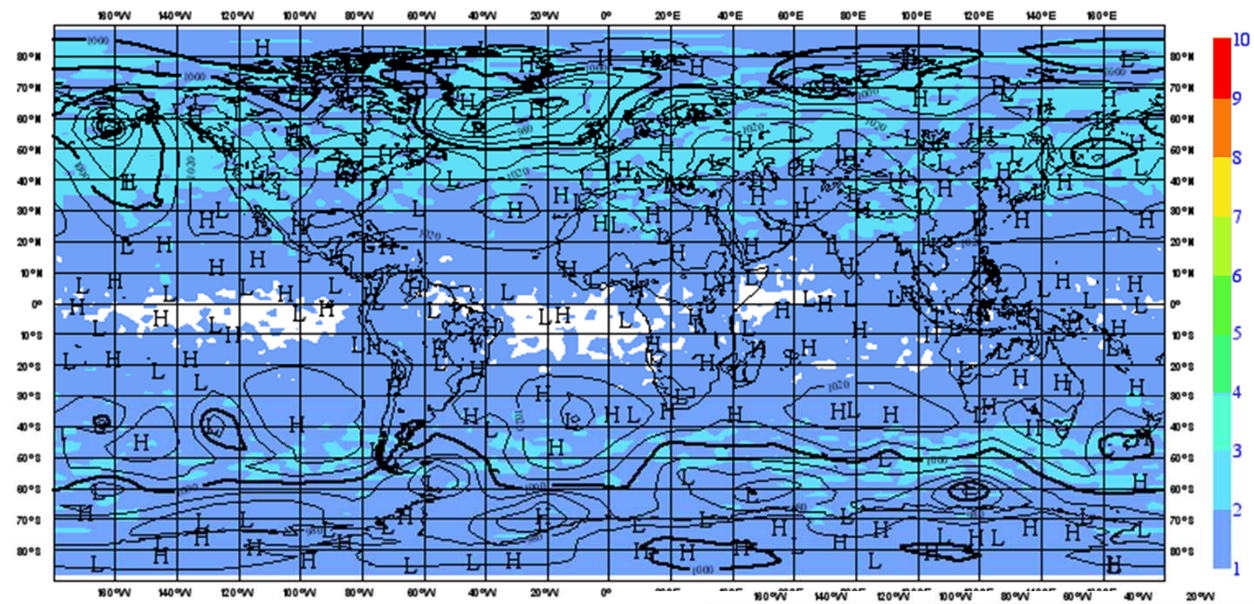


Raw Ensemble StDev
VO ml64



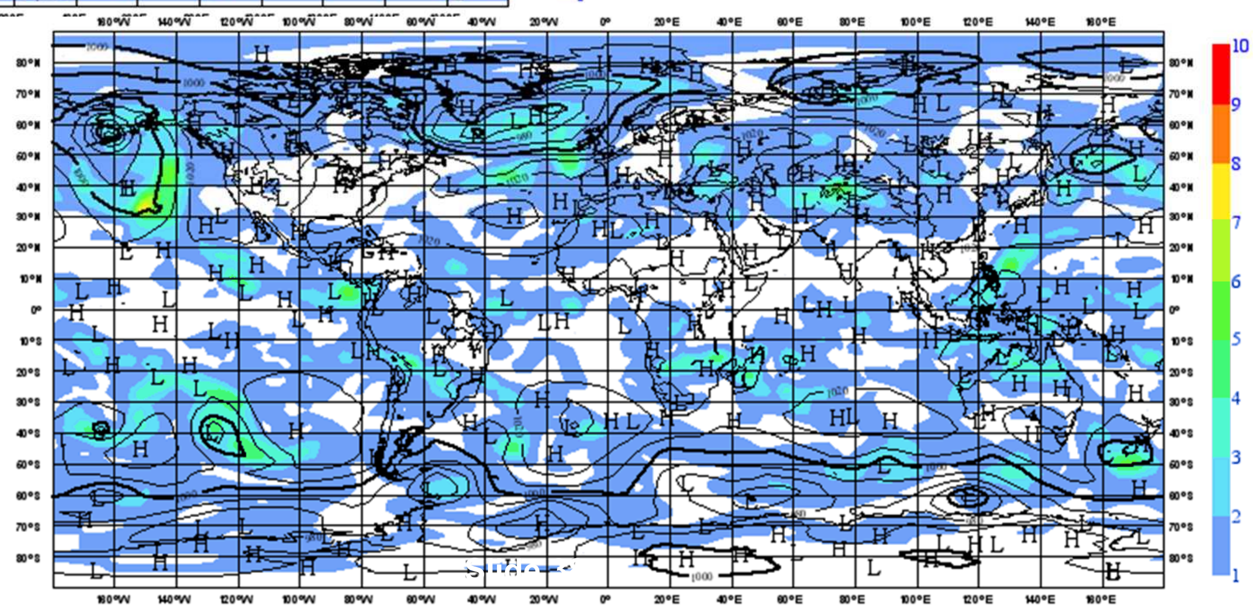
Filtered Ensemble StDev
VO ml64

EDA variances



Operational StDev
Random. Method
(Fisher & Courtier, 1995)

Filtered Ensemble StDev
VO ml64

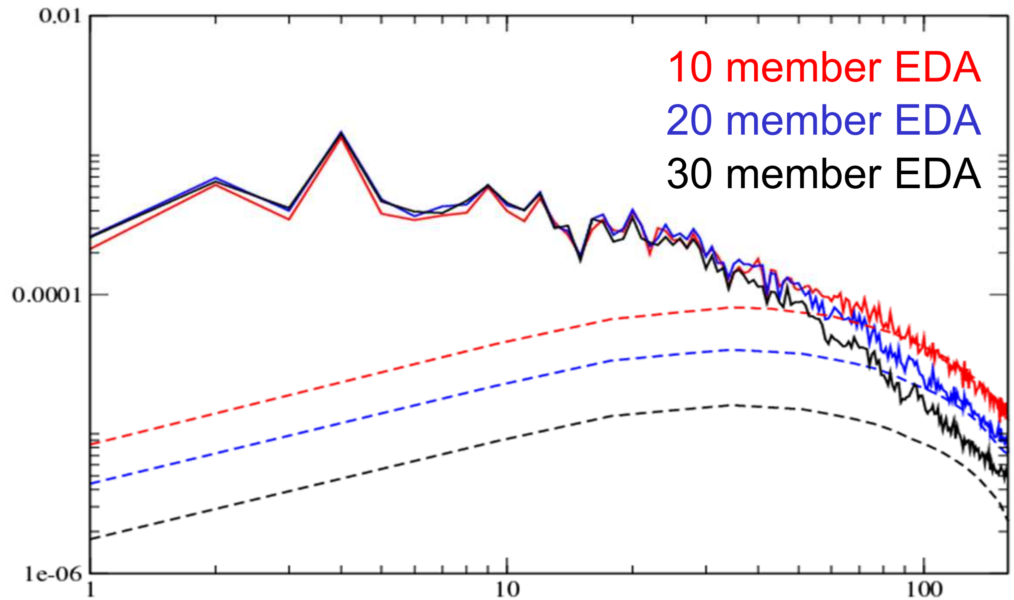


EDA variances: Ensemble Size

The sampling noise effectively limits the scales that we can robustly estimate from the EDA.

The effective spatial resolution of the diagnosed errors is much coarser than the nominal EDA resolution (T399) and is primarily determined by the ensemble size (Bonavita et al., 2010)

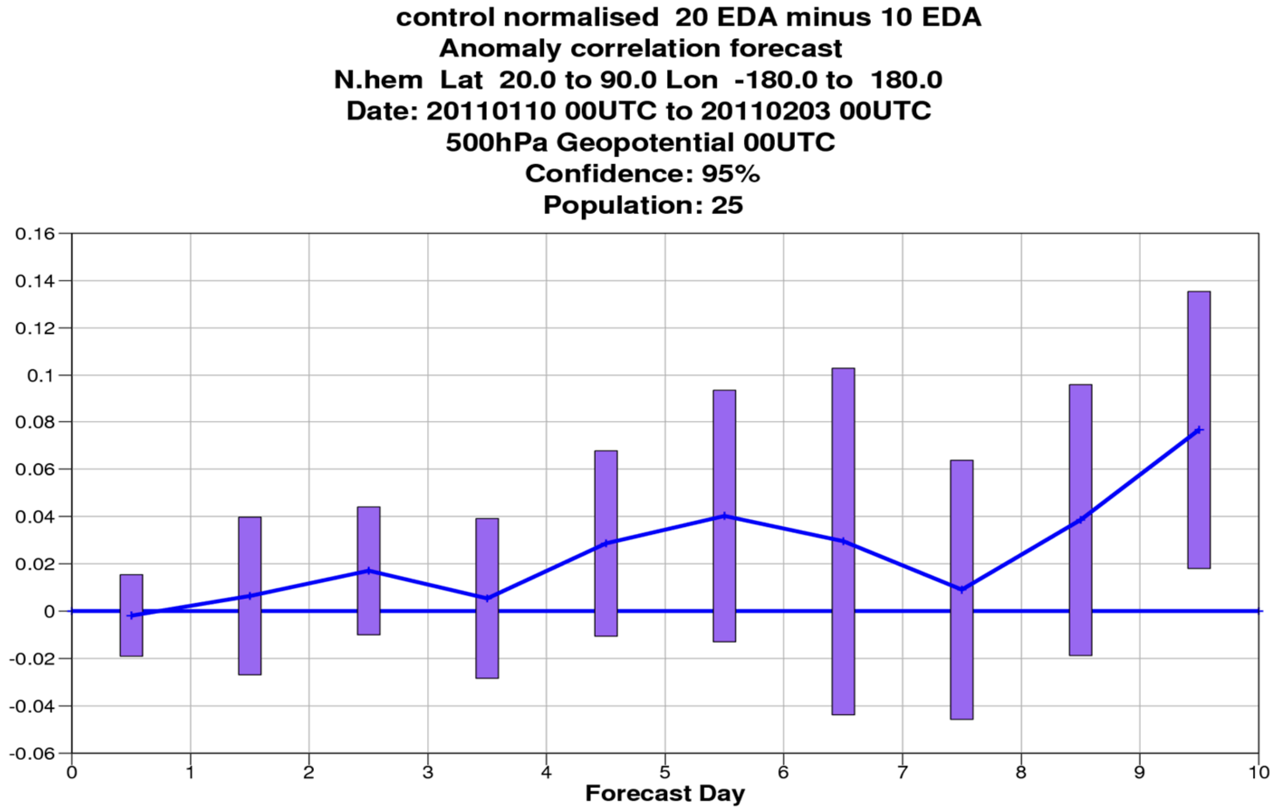
Temperature ml 49 (~200 hPa)



EDA variances: Ensemble Size

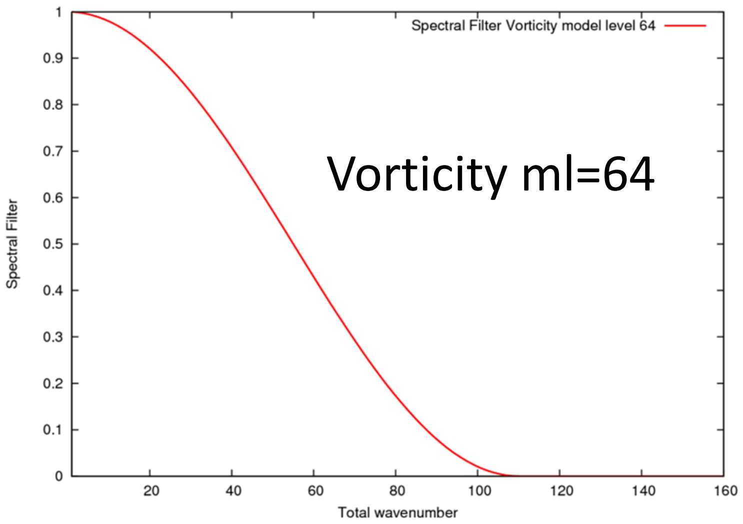
A larger EDA effectively allows the sampling of errors at finer resolutions.

This helps improve analysis and forecast skill!



EDA variances

The current noise filter is spectral:



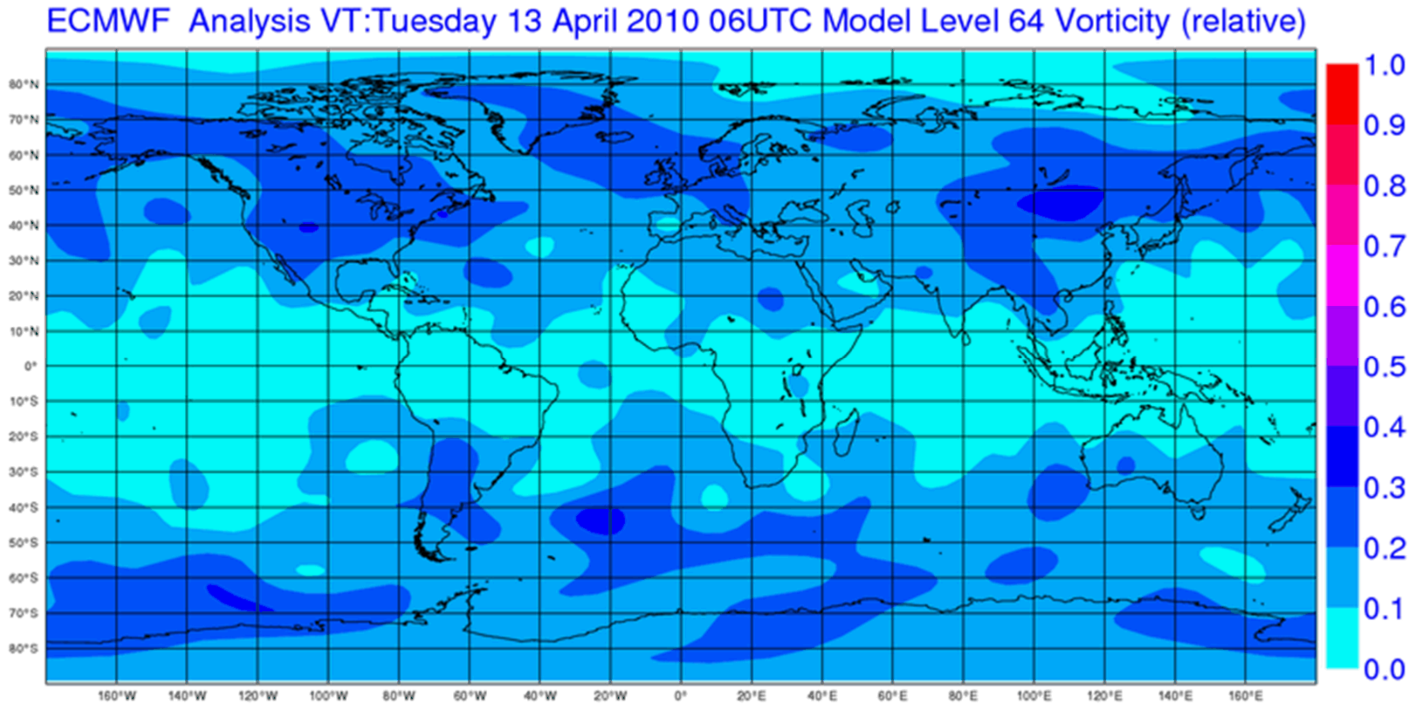
This means that there is full resolution in terms of scale but none in physical space (i.e., the same filtering function is applied everywhere on the globe).

A wavelet filter would trade in some spectral resolution in exchange for spatial resolution:

EDA variances

A wavelet filter would trade in some spectral resolution in exchange for spatial resolution.

Filter for Vorticity (ml=64), wavelet 14 (wavenumbers 95-127)



EDA variances

b) Systematic errors due to incorrect specification of error sources in the EDA (i.e., mis-specification of $\mathbf{R}$, $\mathbf{Q}$, uncertainties in the boundary conditions)

A statistically consistent ensemble should satisfy:

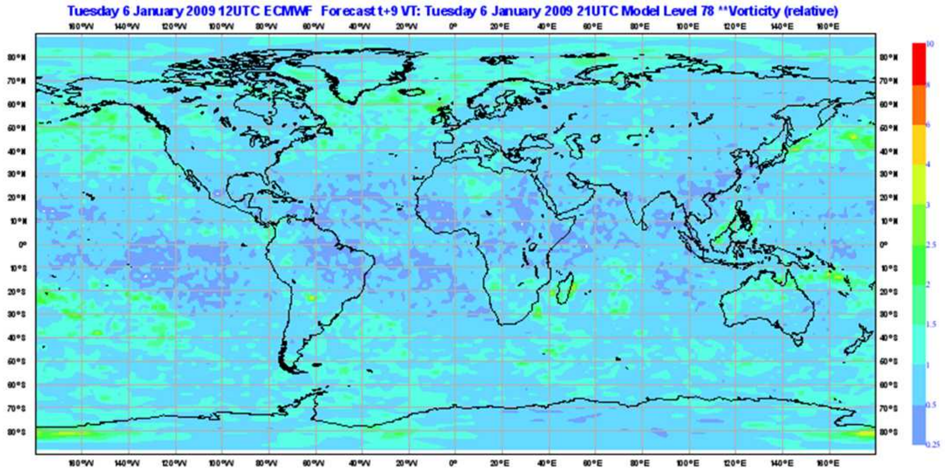
$$\left(1-\frac{1}{N_{ens}}\right)^{-1}\left\langle \frac{1}{N_{ens}}\sum_{i=1}^{N_{ens}}\left(x_i-\overline{x}\right)^2\right\rangle = \left(1+\frac{1}{N_{ens}}\right)^{-1}\left\langle\left(\overline{x}-x^*\right)^2\right\rangle$$

<ensemble variance> ≈ <squared ensemble mean error>

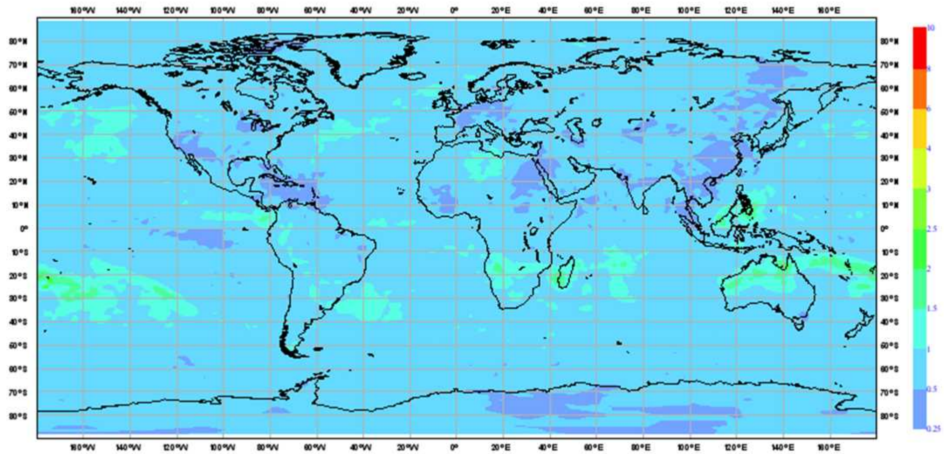
EDA variances

Vorticity ml 78 (~850hPa)

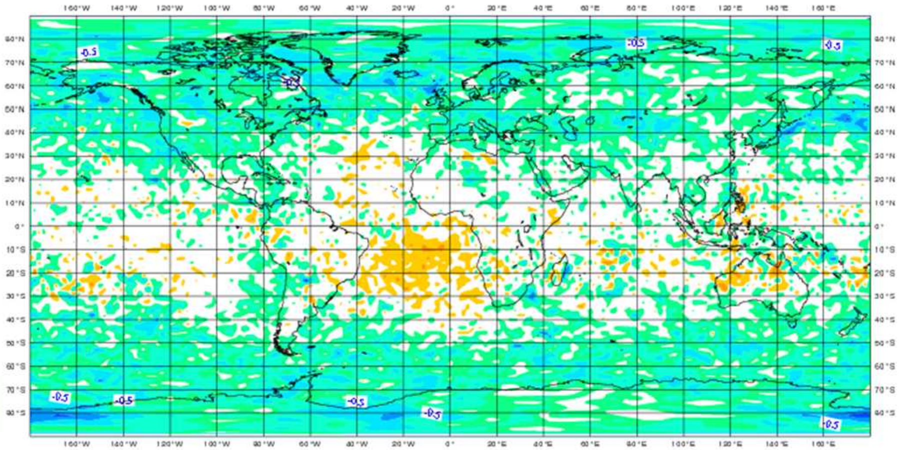
Ensemble Error



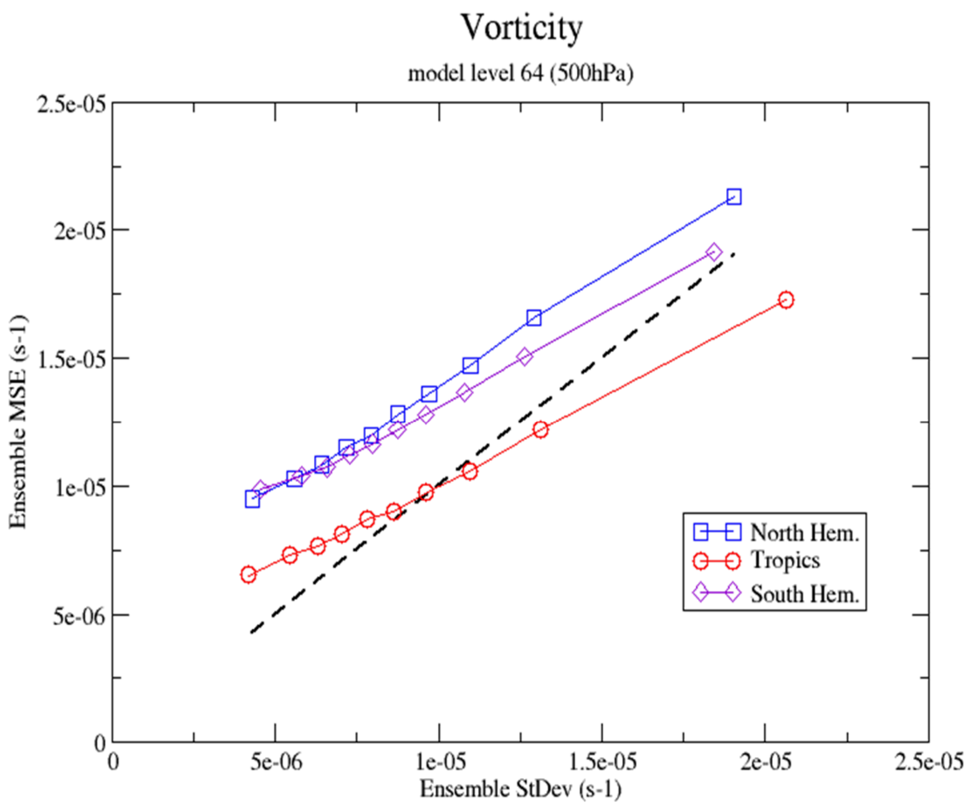
Ensemble Spread



Spread - Error



EDA variances



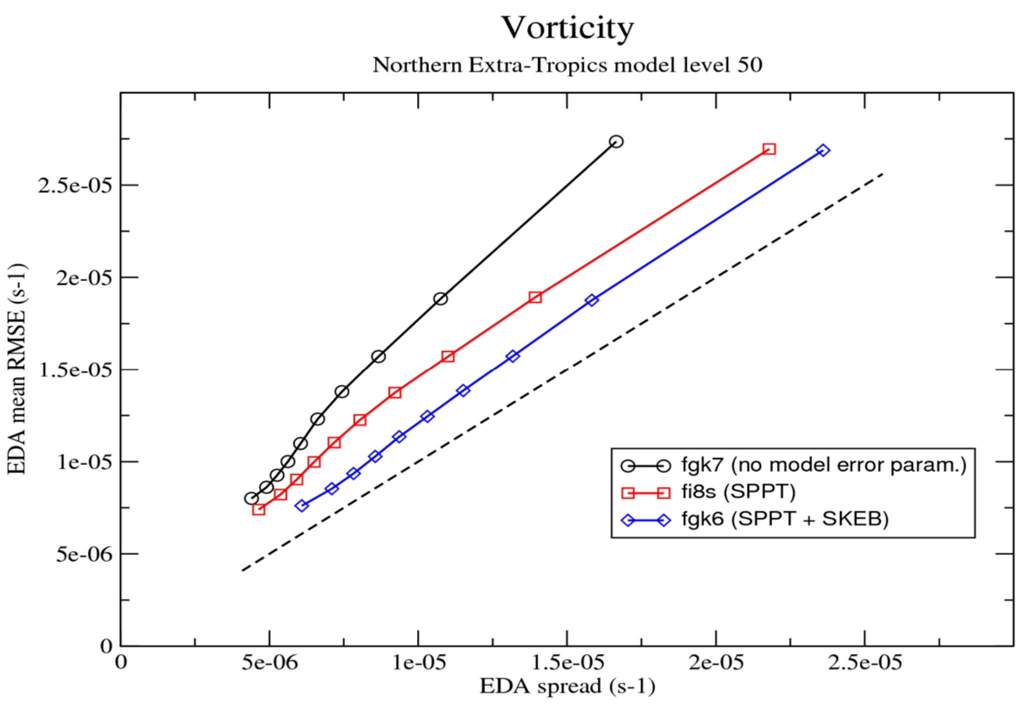
Conditional distribution of the EDA mean background RMS error for given EDA background standard deviation

ECMWF

EDA variances

"Spread-Skill" regressions of the type shown serve two purposes:

1. Diagnose the progress (or lack thereof!) in the modelling of system uncertainties in the EDA



ECMWF

EDA variances

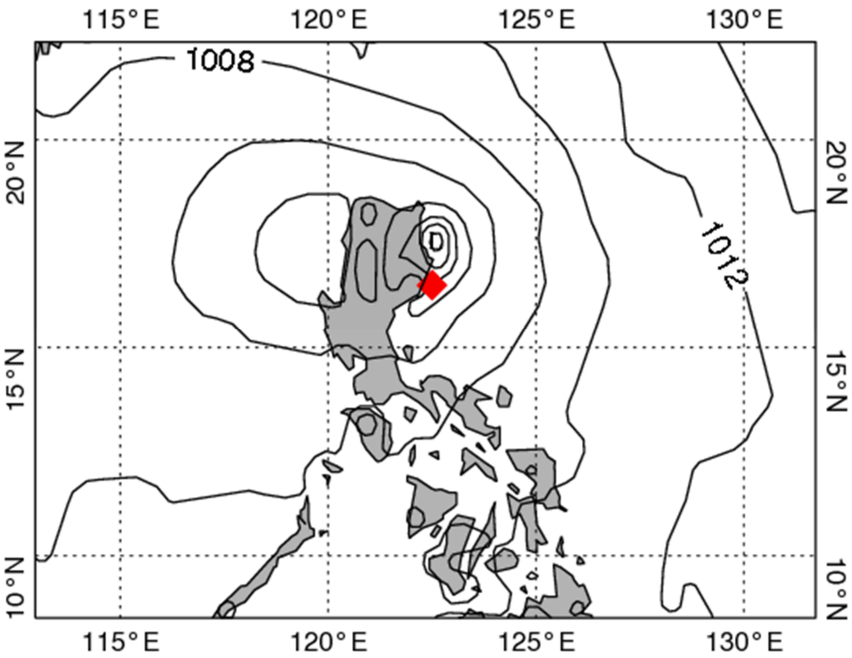
"Spread-Skill" regressions of the type shown serve two purposes:

1. Diagnose the progress (or lack thereof!) in the modelling of system uncertainties in the EDA
2. Calibrate on-line the EDA sample variances to obtain realistic estimates of background errors (Ensemble Variance Calibration, *Kolczynsky et al.,* 2009, 2011; *Bonavita et al*., 2011)

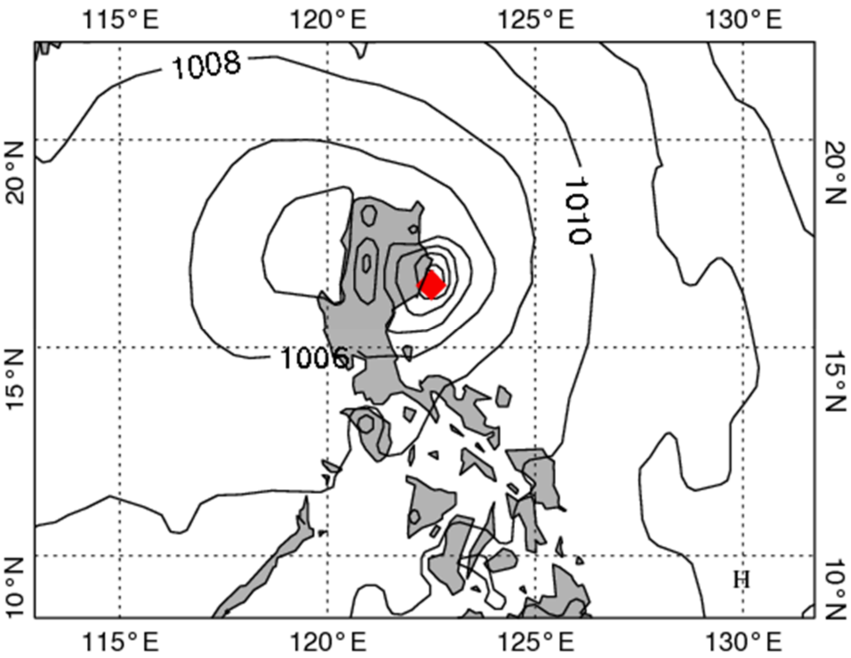
EDA variances

Tropical Storm Aere, 9 May 2011 00UTC:

Operational Analysis

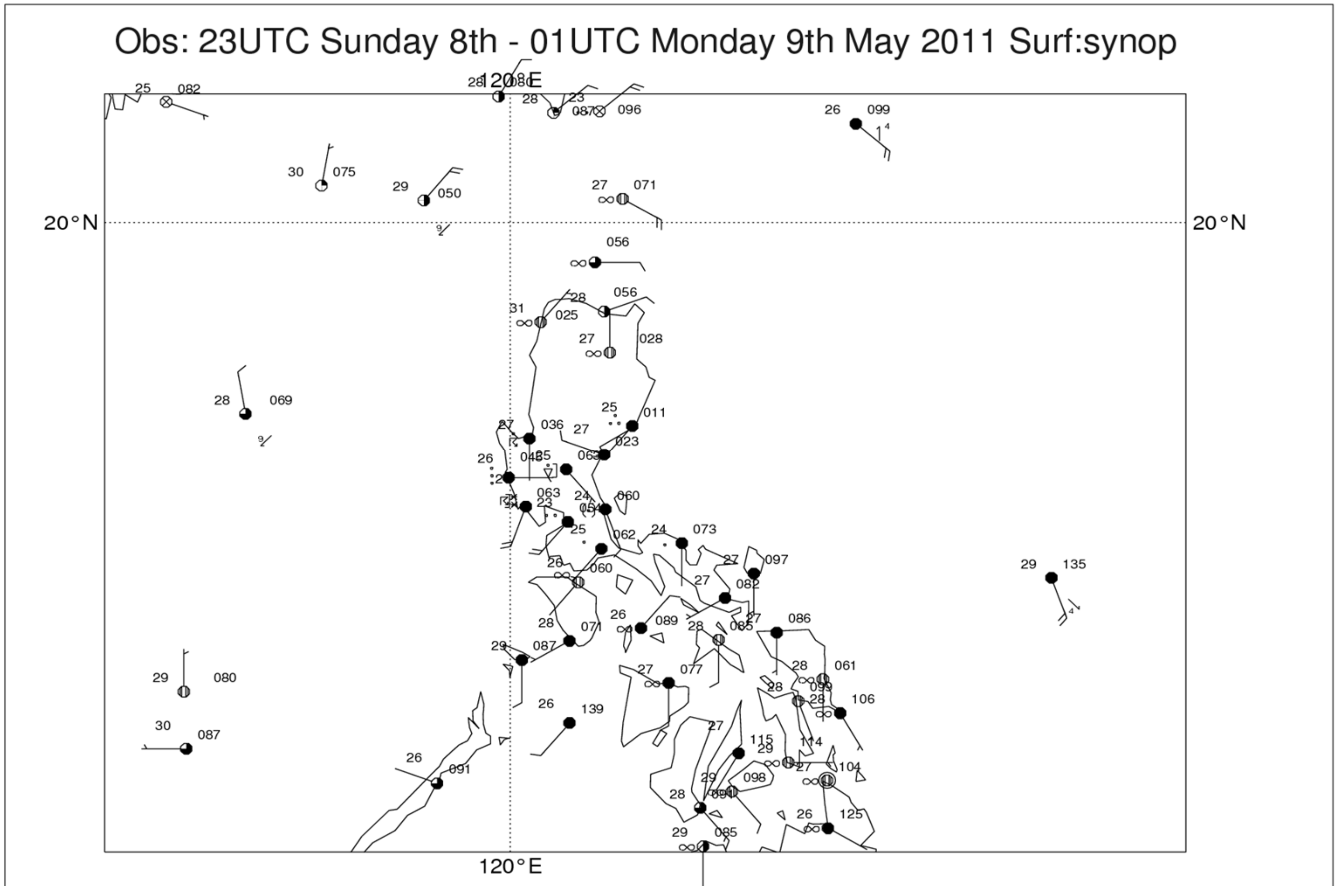


E-suite Analysis



EDA variances

Tropical Storm Aere, 9 May 2011 00UTC:



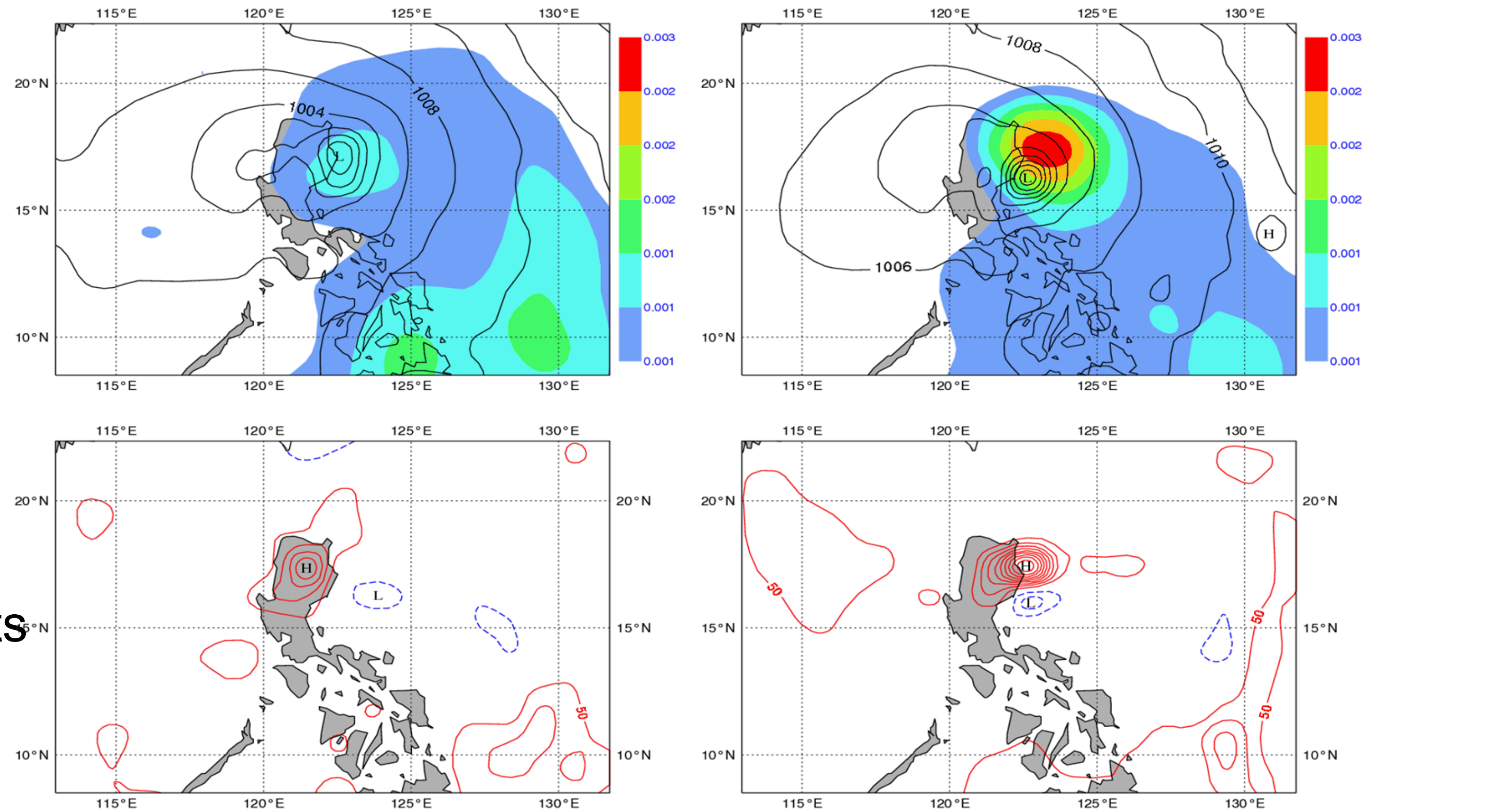
EDA variances

Tropical Storm Aere, 9 May 2011 00UTC:

Operational Analysis | E-suite Analysis

BG Error StDev

Analysis Increments



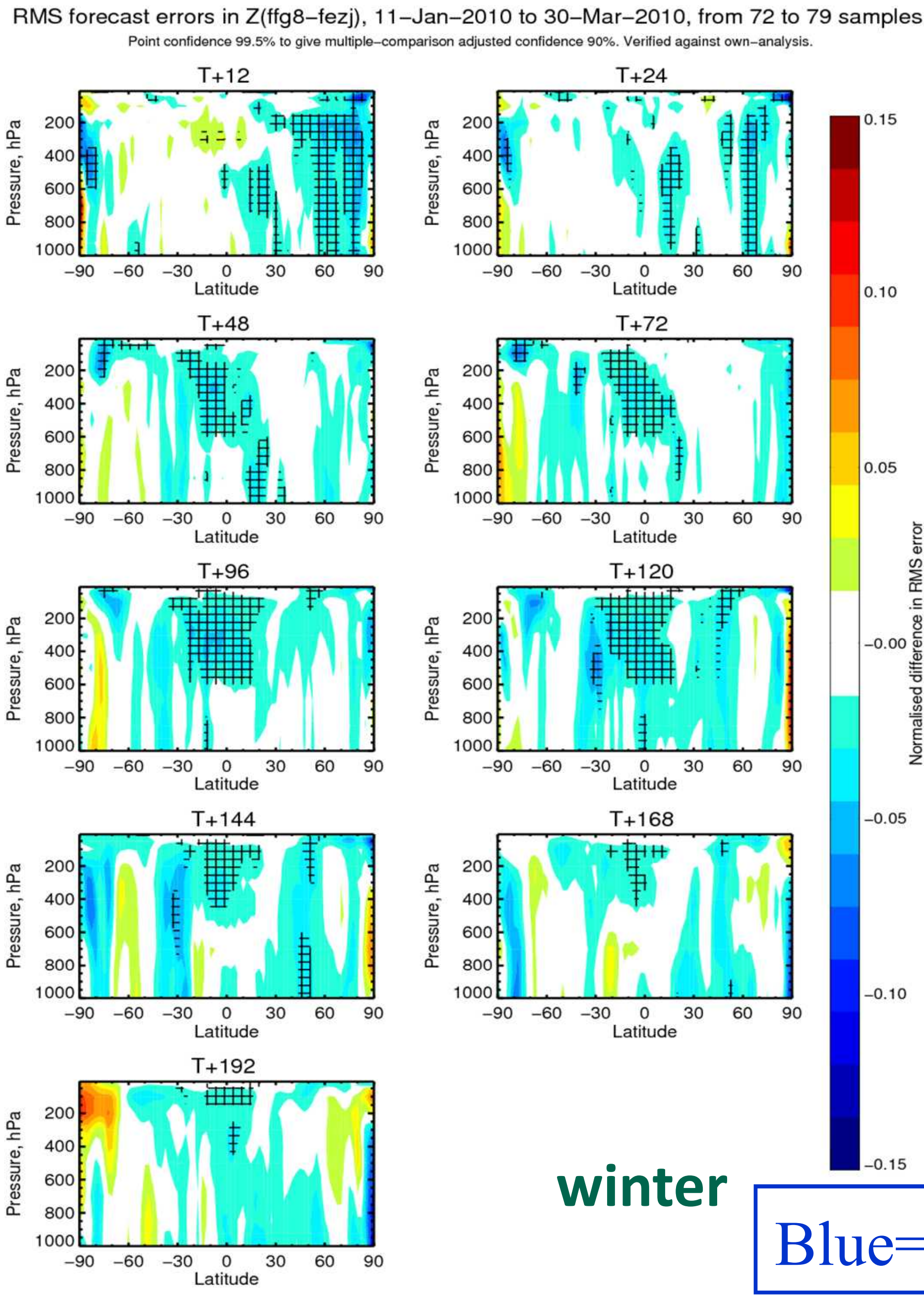
EDA variances

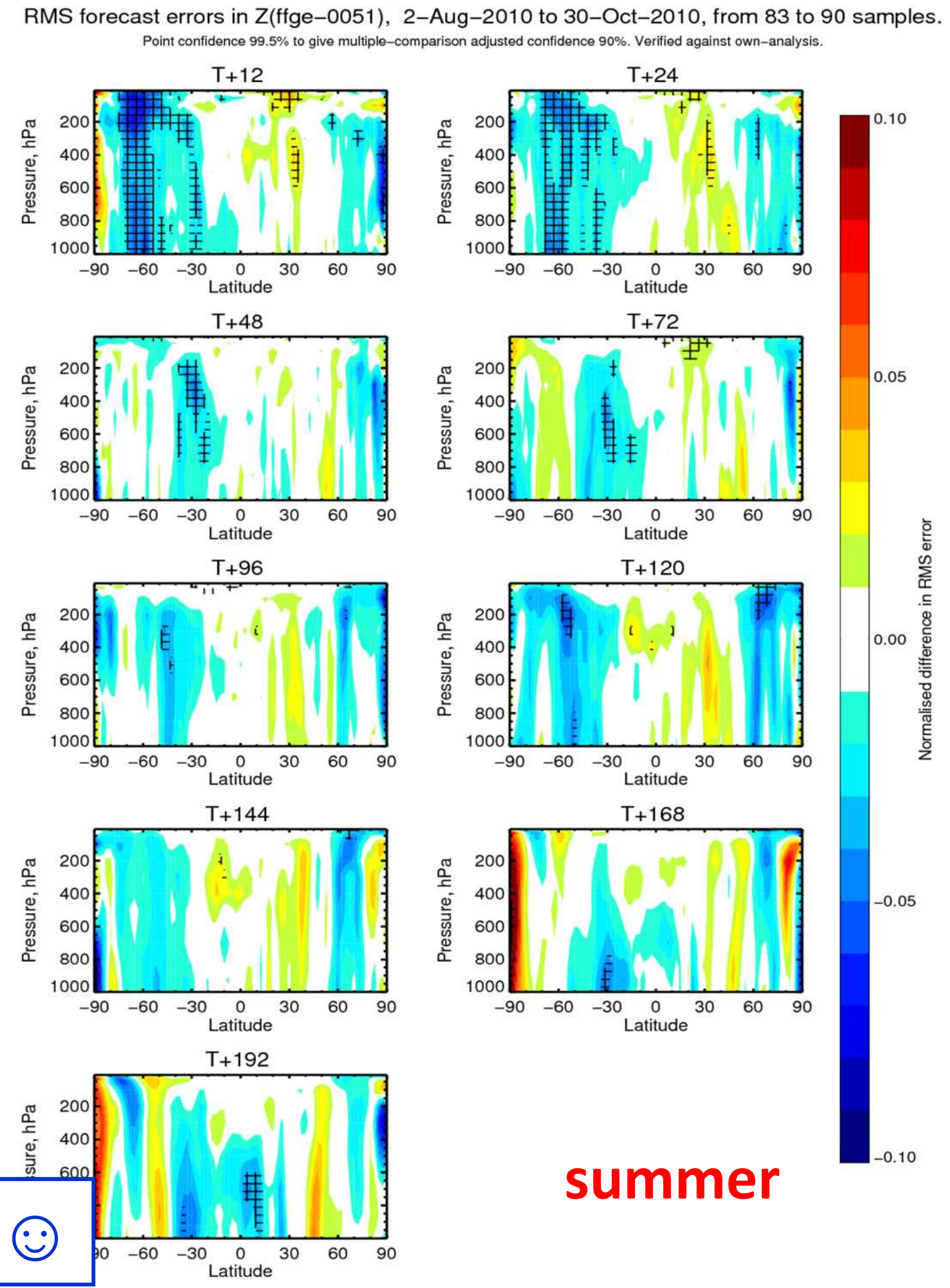
What is the impact of flow-dependent EDA variances on the IFS scores?

CY36R4, T1279L91

- ffg8 20100111 - 20100331 (control: fezj): WINTER
- ffge 20100802 – 20101030 (control: 0051): SUMMER

Geopotential RMSE reduction





winter

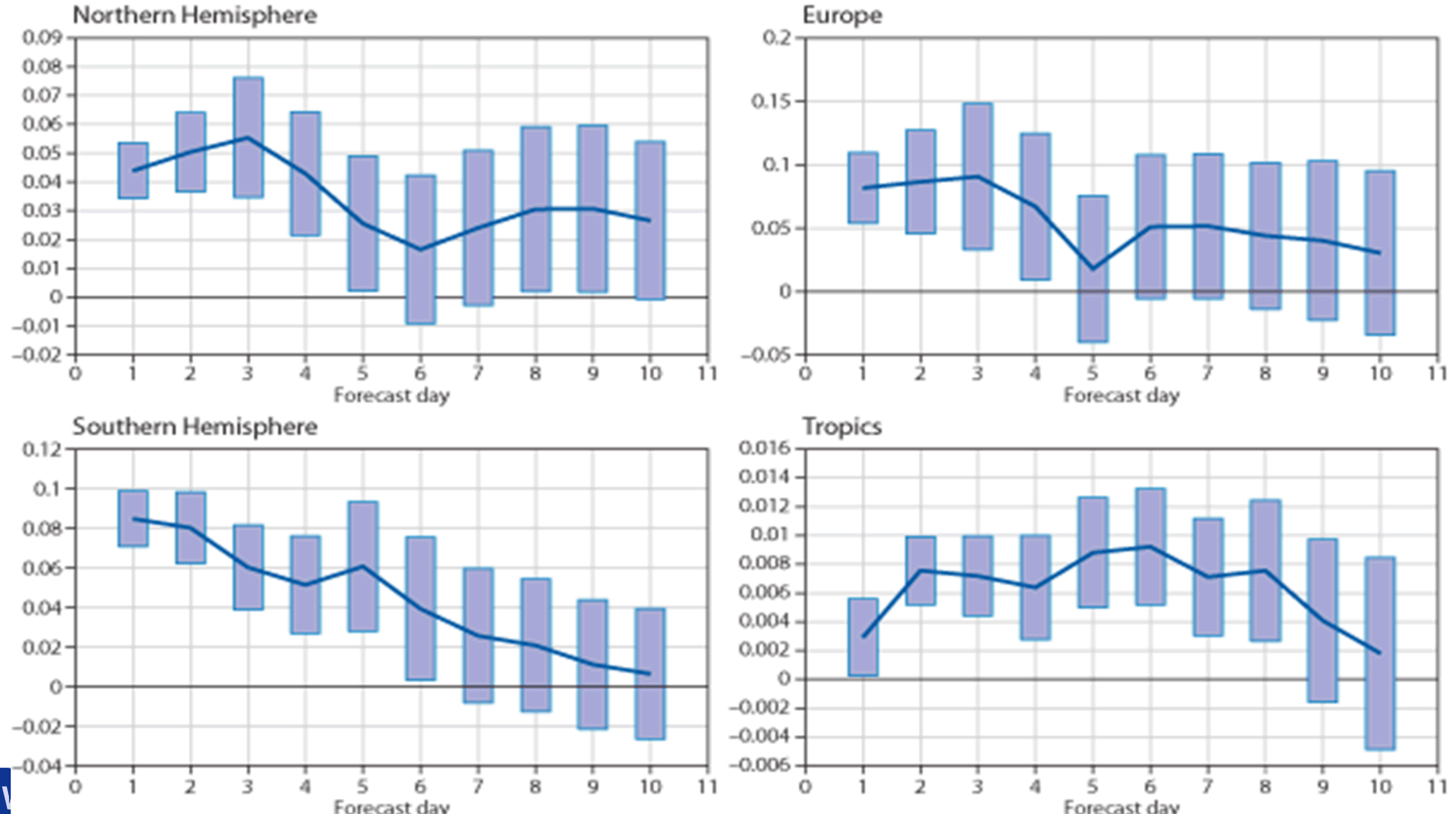
summer

Blue= ☺

EDA variances

CY37R2 (18 May 2011):

- ❑ Use of EDA Variances in 4D-Var
- ❑ Reduction of AMSU-A observation errors



Hybrids: EDA Covariances

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

$\mathbf{C}_j(\lambda,\varphi)$ are full vertical covariance matrices, function of (λ,φ). They determine both the horizontal and vertical background error *correlation structures*;

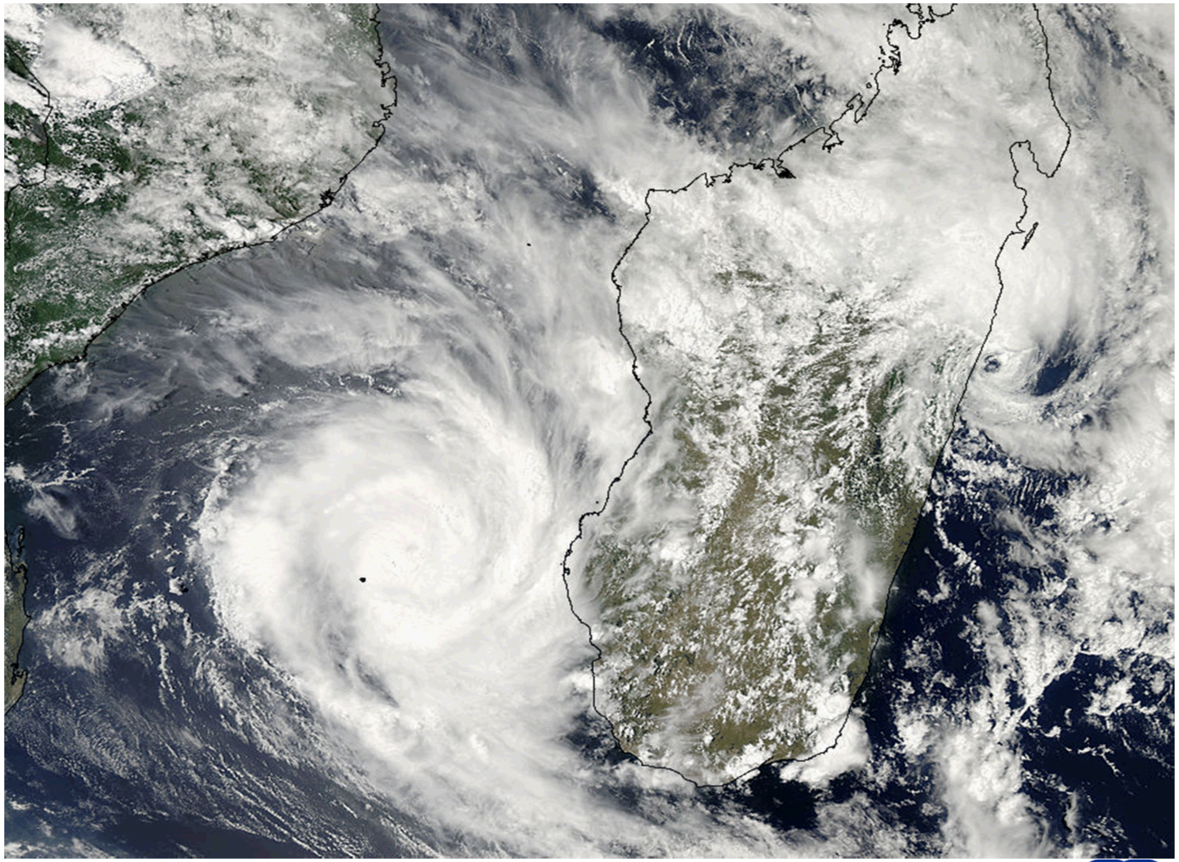
How do we make this error covariance model flow-dependent?

We look for flow-dependent EDA estimates of $\mathbf{\Sigma}_\mathrm{b}$ and $\mathbf{C}_j(\lambda,\varphi)$

Hybrids: EDA Covariances

Diagnosing the Background Error Correlation Length-Scales

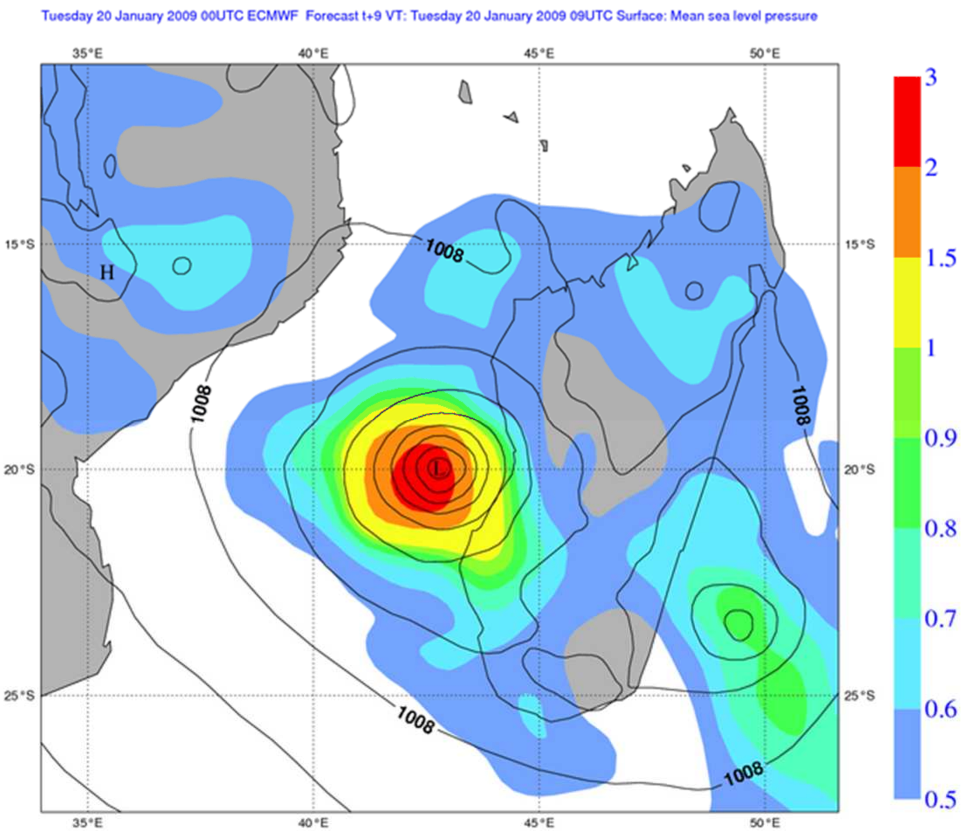
Hurricane Fanele, 20 January 2009

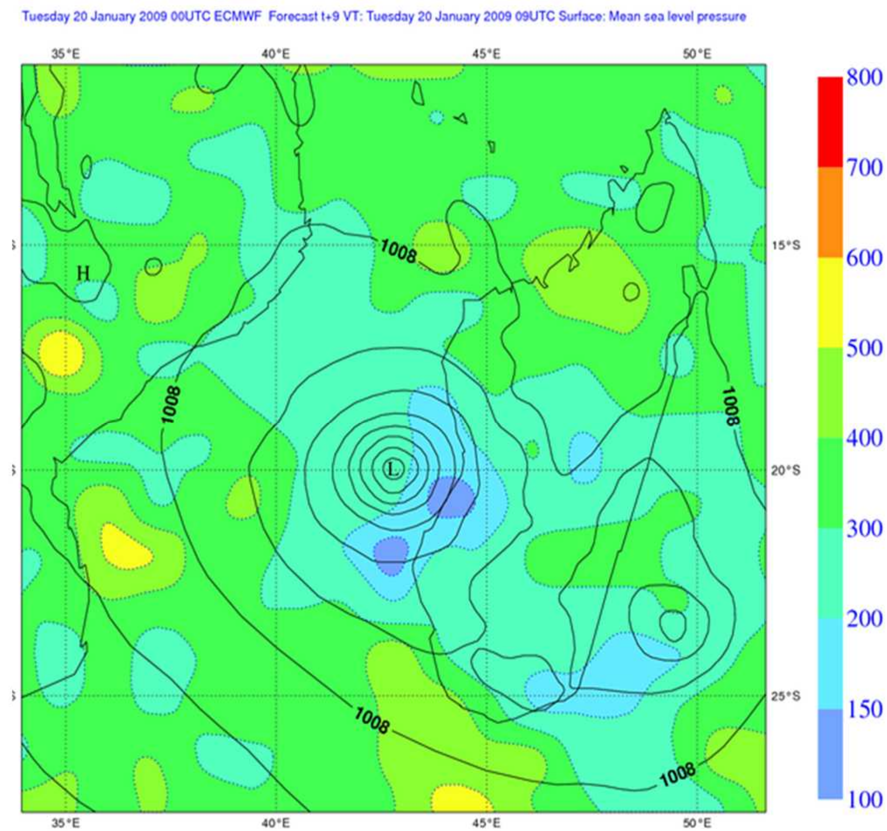


Hybrids: EDA Covariances

20 member EDA

Surf. Press. Background Err. St.Dev.   Surf. Press. BG Err. Correlation L. Scale





Hybrids: EDA Covariances

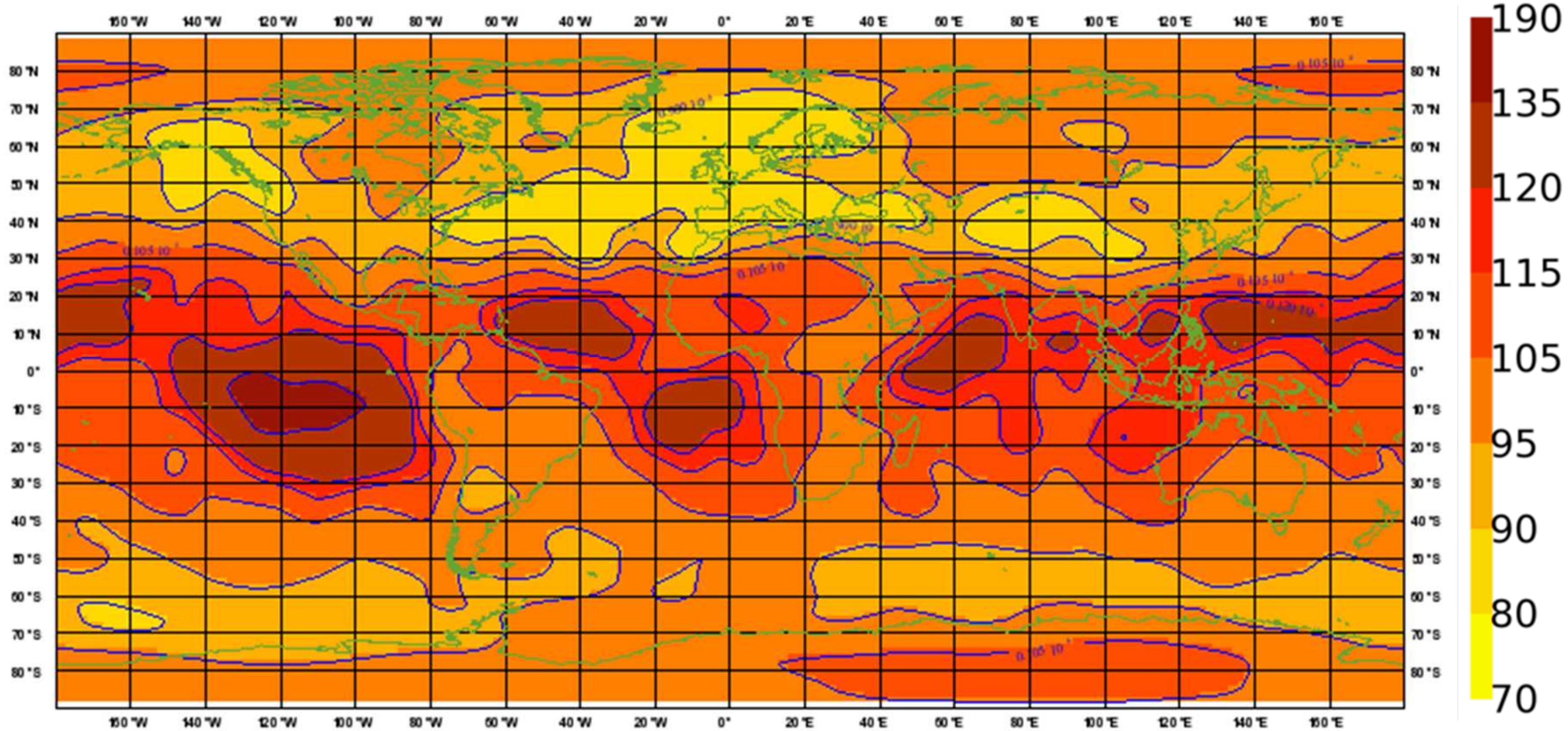
BG Error Length Scale field has more high frequency spatial structure than BG error StDev -> need for larger ensemble

Off-line estimates of $\mathbf{C}_j(\lambda,\varphi)$ are computed over a period of 2 months.

Simplest approach to introduce flow-dependency in the correlation structures **is through use of an evolving, on-line estimation of** $\mathbf{C}_j(\lambda,\varphi)$ **over a short calibration period** (*Varella et al*., 2011)

Hybrids: EDA Covariances

wavelet-implied length-scales of wind near 500 hPa
3-week average, from 15/2 to 7/3 2010

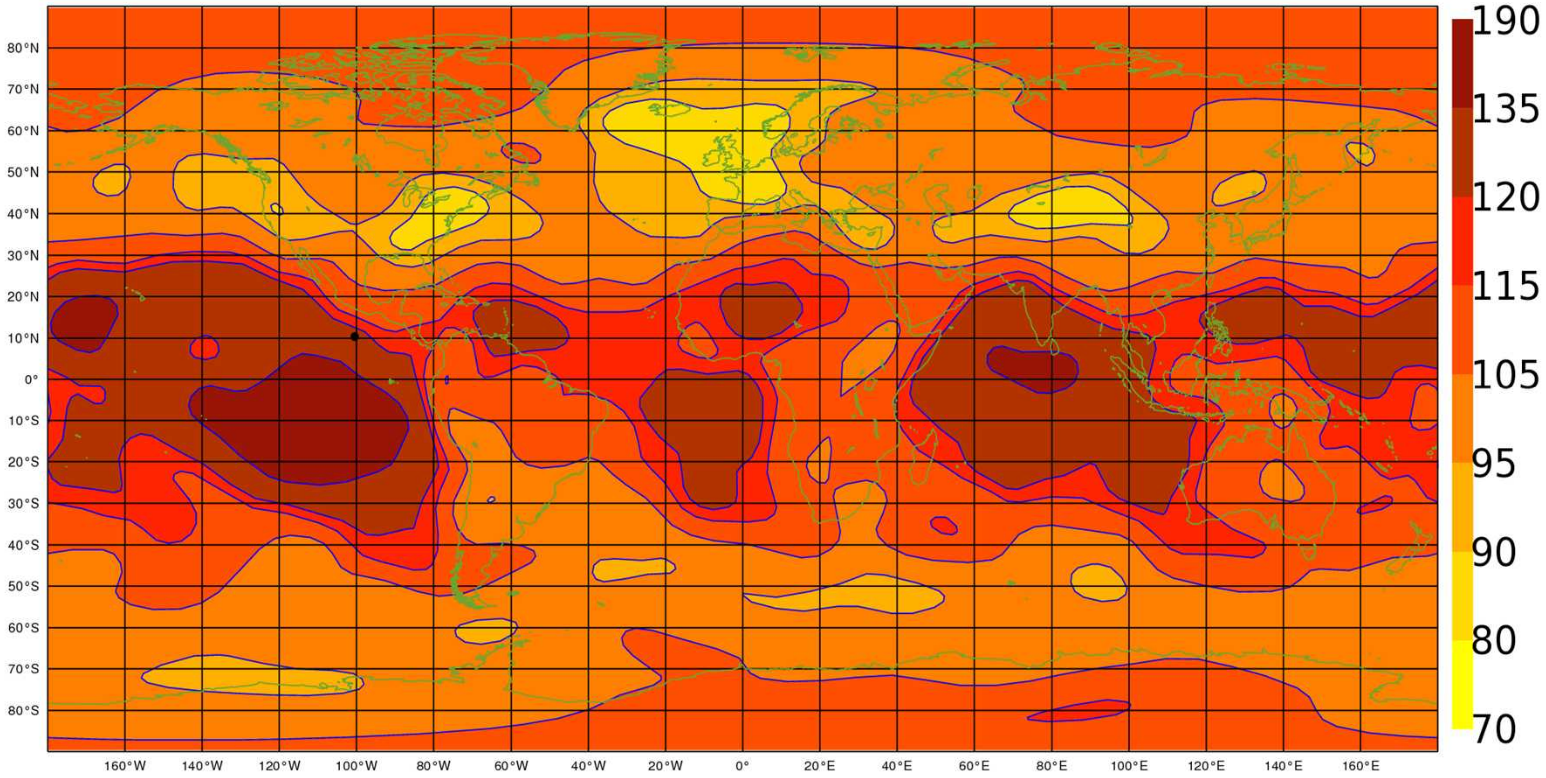


from: L.Berre

ECMWF

Hybrids: EDA Covariances

wavelet-implied length-scales of wind near 500 hPa
4-day average, from 24/2 to 27/2 2010

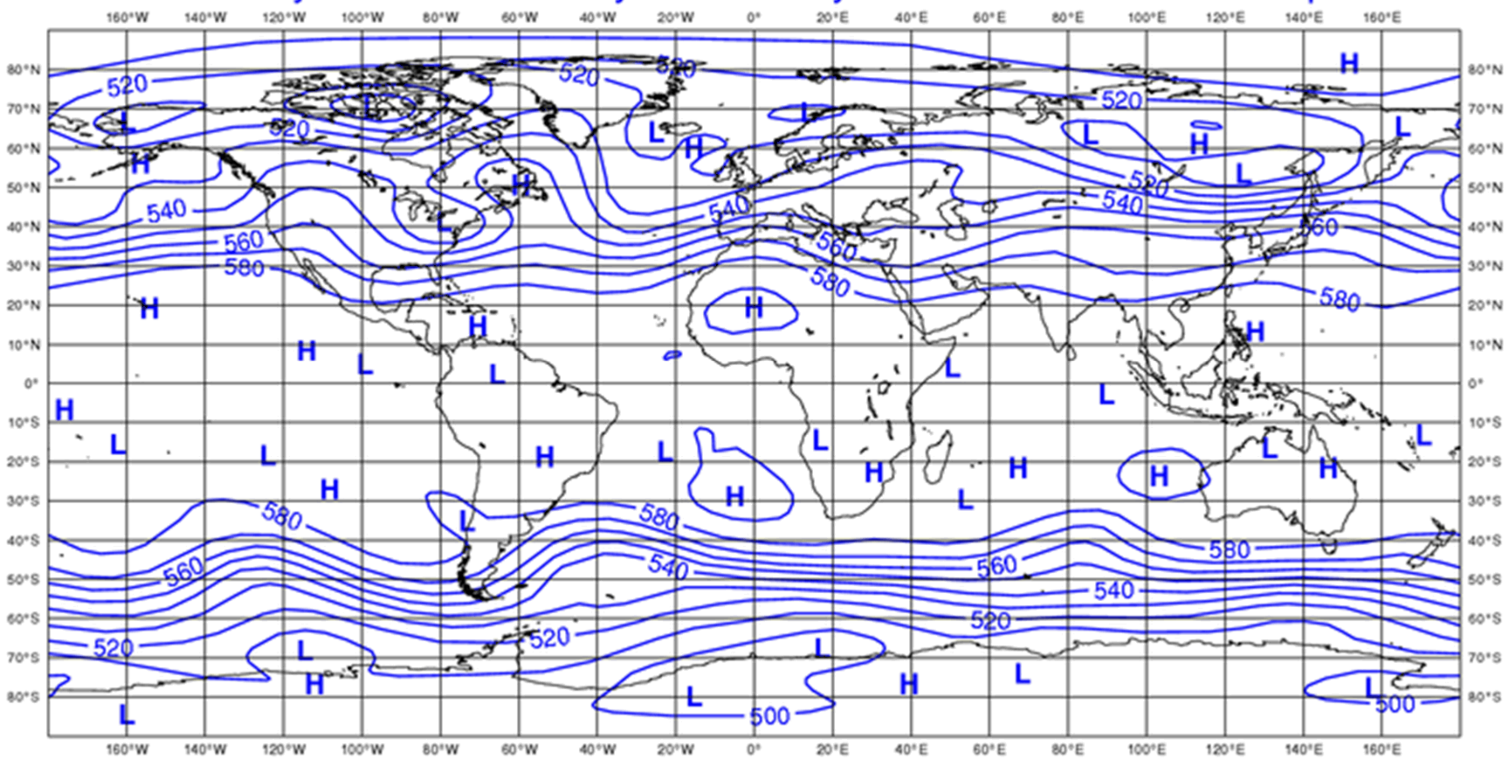


from: L.Berre

Hybrids: EDA Covariances

Mean Geopotential Height at 500 hPa
4-day average, from 24/2 to 27/2 2010

ECMWF Analysis VT:Wednesday 24 February 2010 00UTC 500hPa Geopotential



ECMWF

Hybrids: EDA Covariances

Regularization of the on-line correlation estimates through temporal averaging and the implicit spatial averaging of the wavelet representation

Larger ensemble would allow for a larger flow-dependent component to be retained

Other forms of regularization of the on-line correlation estimates can be envisaged (i.e., convex combinations of on-line and off-line $\mathbf{C}_j(\lambda,\varphi)$ estimates)

Conclusions and Perspectives

- The Kalman Filter/Smoother is still the gold standard of atmospheric global NWP data assimilation, but practically unfeasible

- Non-sequential approx. to KF (4D-Var):

  1. Keeps full-rank representation of **B** matrix and its implicit evolution during the assimilation window;
  2. Unable to cycle **B** estimates;
  3. Difficult to access realistic estimates of $\mathbf{P}^a$;
  4. Long-window weak-constraint 4D-Var would potentially solve issue 2. but still to be demonstrated in realistic NWP settings

Conclusions and Perspectives

- Low-rank, Monte Carlo, Sequential approx. to KF (EnKF):
  1. Explicit evolution and cycling of low-rank approximation of $\mathbf{B}$ matrix (and $\mathbf{P}^a$);
  2. Computationally scalable and efficient;
  3. The EnKF analysis is restricted to the error subspace spanned by the ensemble perturbations. This is unrealistically small and requires covariance localization/inflation to keep the EnKF from diverging;
  4. EnKF performance degrades with respect to 4D-Var when N_{obs} in the local analysis patch is >> N_{ens} and observations are non-local (satellite radiances). Can this problem be cured with larger but affordable ensemble size and more careful observation selection?

Conclusions and Perspectives

- Hybrid approx. to KF: try to combine the strengths of the sequential and non-sequential approaches

  a) Low-rank, Monte Carlo error representation through cycling EnKF/EDA system;

  b) State estimate from full-rank 4D-Var analysis where static $\mathbf{B}$ at the start of the window is (partially) replaced by EnKF/EDA flow-dependent $\mathbf{B}$

- Hybrid can be done by adding an ensemble, flow-dependent component to the static $\mathbf{B}$ used in 4D-Var (alpha control var.)

- Hybrid can also be done by using an EDA/EnKF to get an on-line, flow-dependent estimate of parameterised $\mathbf{B}$ (EDA approach)

Conclusions and Perspectives

- Use of hybrids consistently improves deterministic analysis and forecast skill w.r.to pure sequential (EnKF) and non-sequential (4D-Var) solutions;
- EDA/EnKF, possibly re-centred around deterministic analysis, provide improved sampling of initial errors for Ensemble Prediction
- We can expect growing ensemble use in 4D-Var:
  1. A larger ensemble (both in the EDA and EnKF) improves error characterization and ultimately skill scores;
  2. 4D background error covariances sampled from an EDA/EnKF could be used over the all 4D-Var assimilation window (not only at the start!): En-4D-Var (Liu et al., 2008; Buehner et al., 2010). This would remove the need of developing and maintaining a TL and Adjoint version of the forecast model

Conclusions and Perspectives

- We can expect growing ensemble use in 4D-Var:

3. Weak-constraint Long-window 4D-Var revolves around the estimation of $\mathbf{Q}$: It is conceivable that an EDA will provide a way of effectively sampling $\mathbf{Q}$

4. The EnKF is more computationally efficient than an ensemble of 4D-Var analysis (EDA): **if** it can be shown to be as accurate as standard 4D-Var with the full observing system, then it will provide a relatively cheap and efficient way of cycling error estimates in a hybrid system

Questions and Answers!

Additional Slides

Randomization

Randomization procedure, Fisher and Courtier, 1995

Define N random vector in control-variable space, with independent elements drawn from a Gaussian distribution with zero mean and unit variance $\boldsymbol{\xi}_i \sim \mathcal{N}(\mathbf{0},\mathbf{I})$. Then $\boldsymbol{L\xi}_i$ will be drawn from the distribution $\mathcal{N}(\mathbf{0},\mathbf{B})$

A grid point estimate of background error variances can then be computed from:

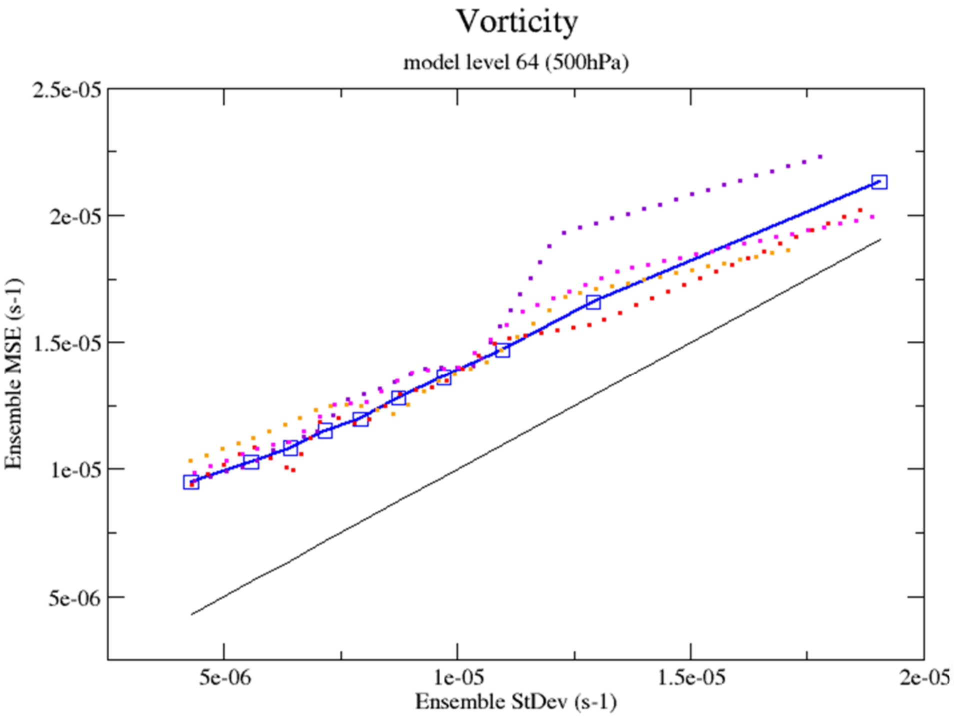
$$\hat{B}_g = \frac{1}{N}\sum_{i=1}^{N}\left(S^{-1}\mathbf{L}\xi_i\right)\left(S^{-1}\mathbf{L}\xi_i\right)^T$$

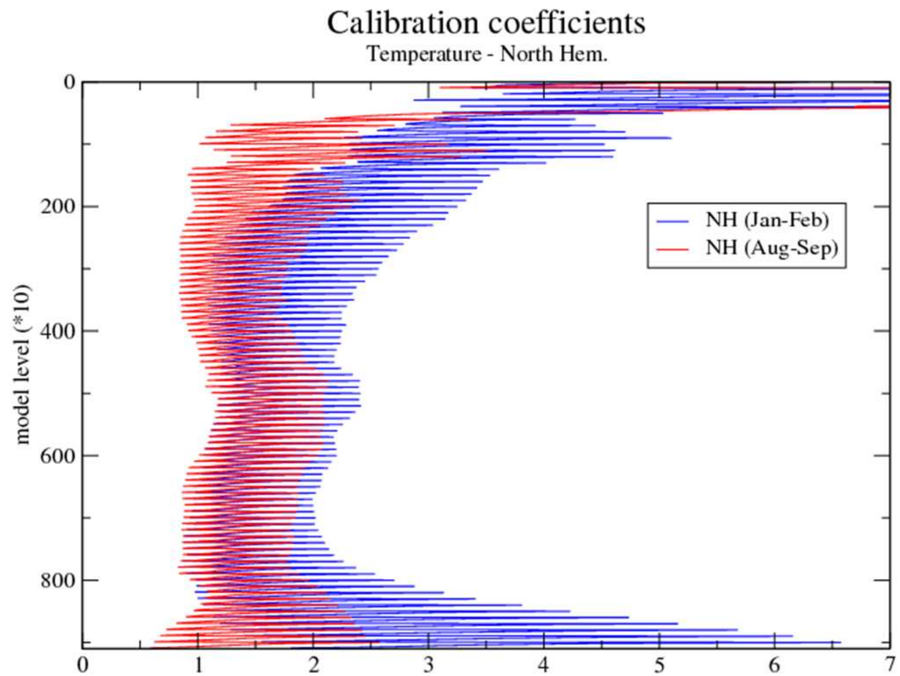
Where S^{-1} denotes the inverse transform from Spectral space.

The variances are then rescaled based on an estimate of analysis errors from the leading eigenvectors of the Hessian matrix.

Finally an error growth model (Savijärvi, 1995) is applied to account for error growth over the short range forecast

Use of EDA variances in 4DVar

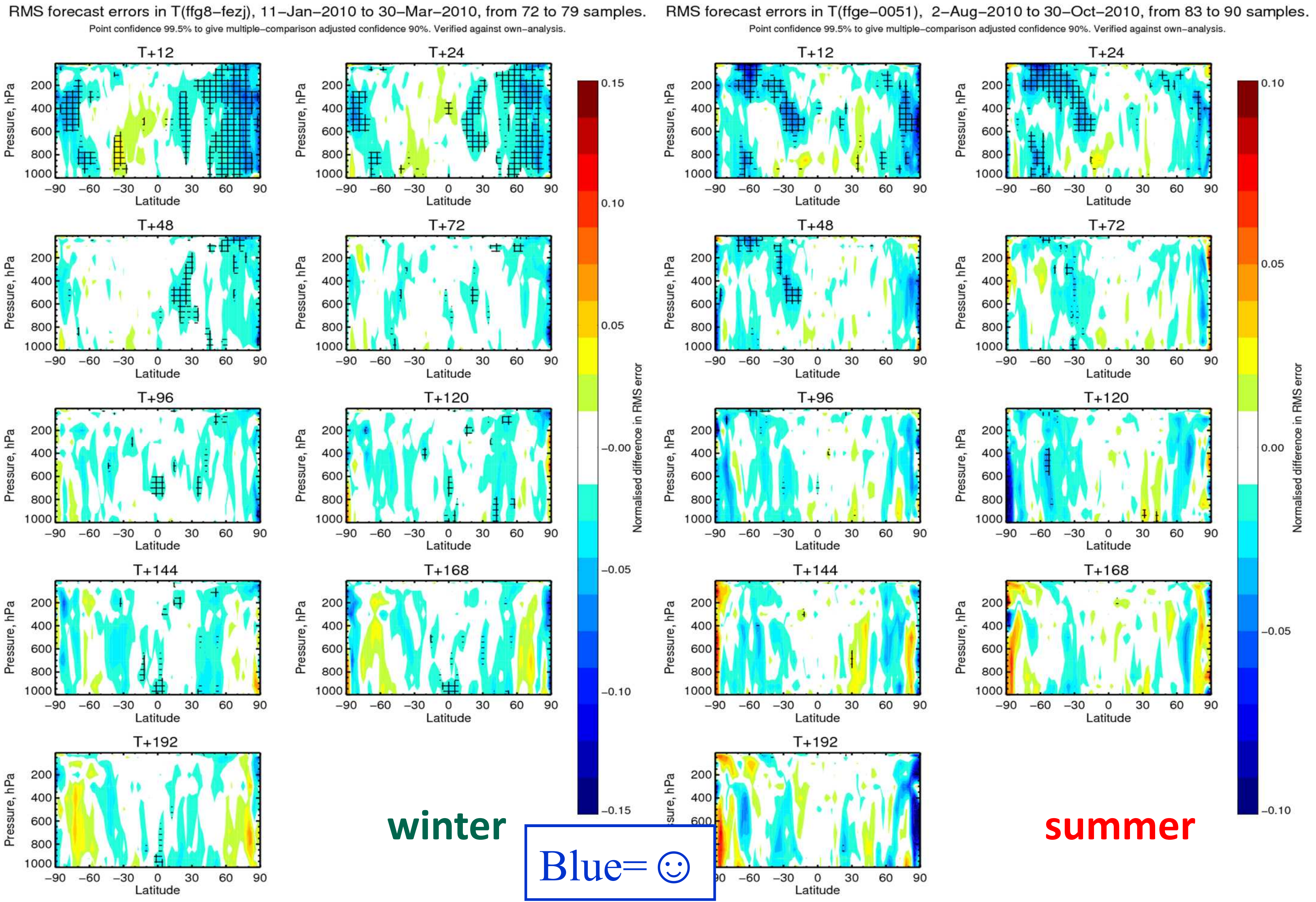




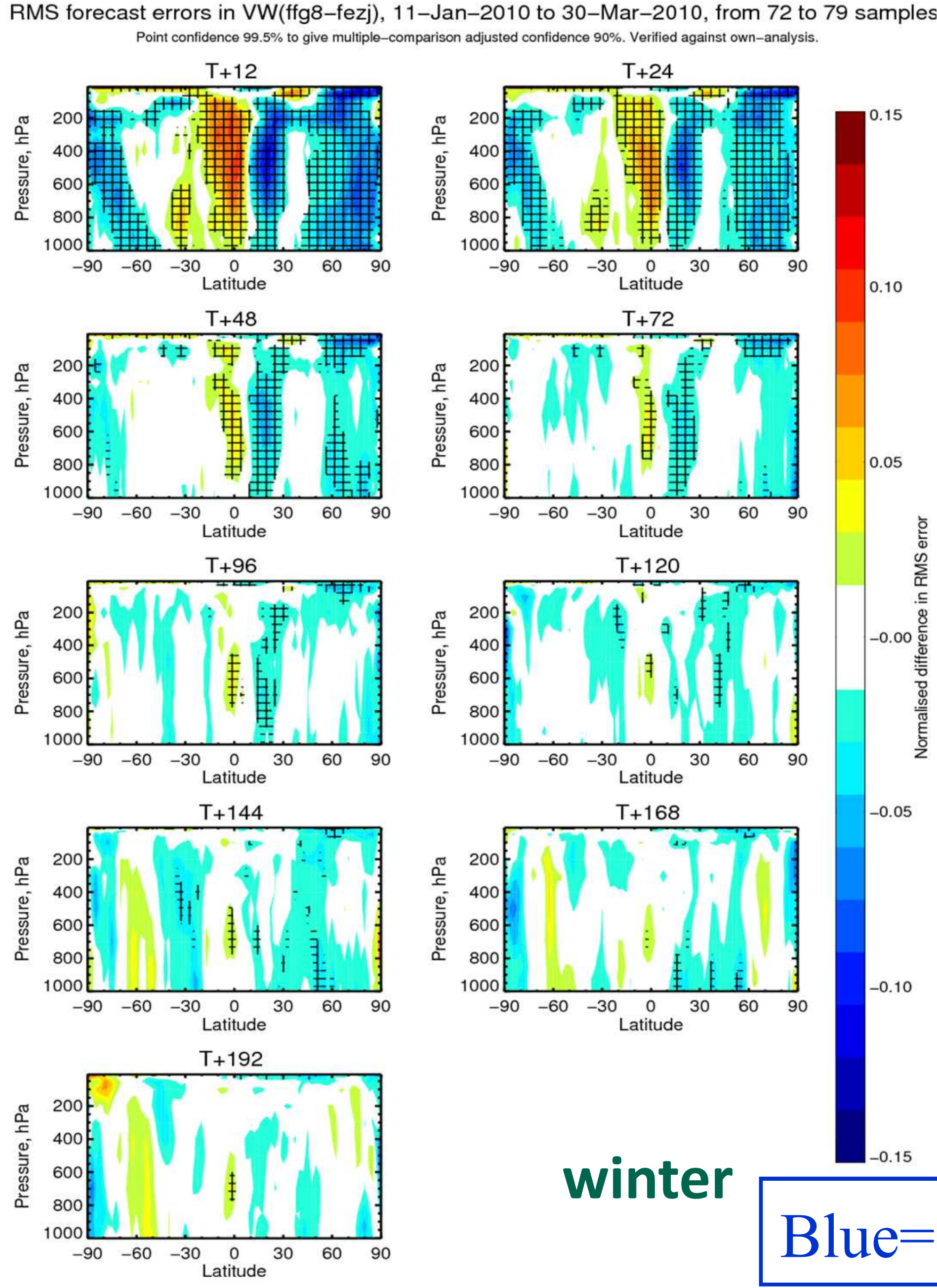
- There is not much variability on daily-weekly scales but **seasonal variability** is important
- General solution: **slowly varying adaptive** calibration coefficients

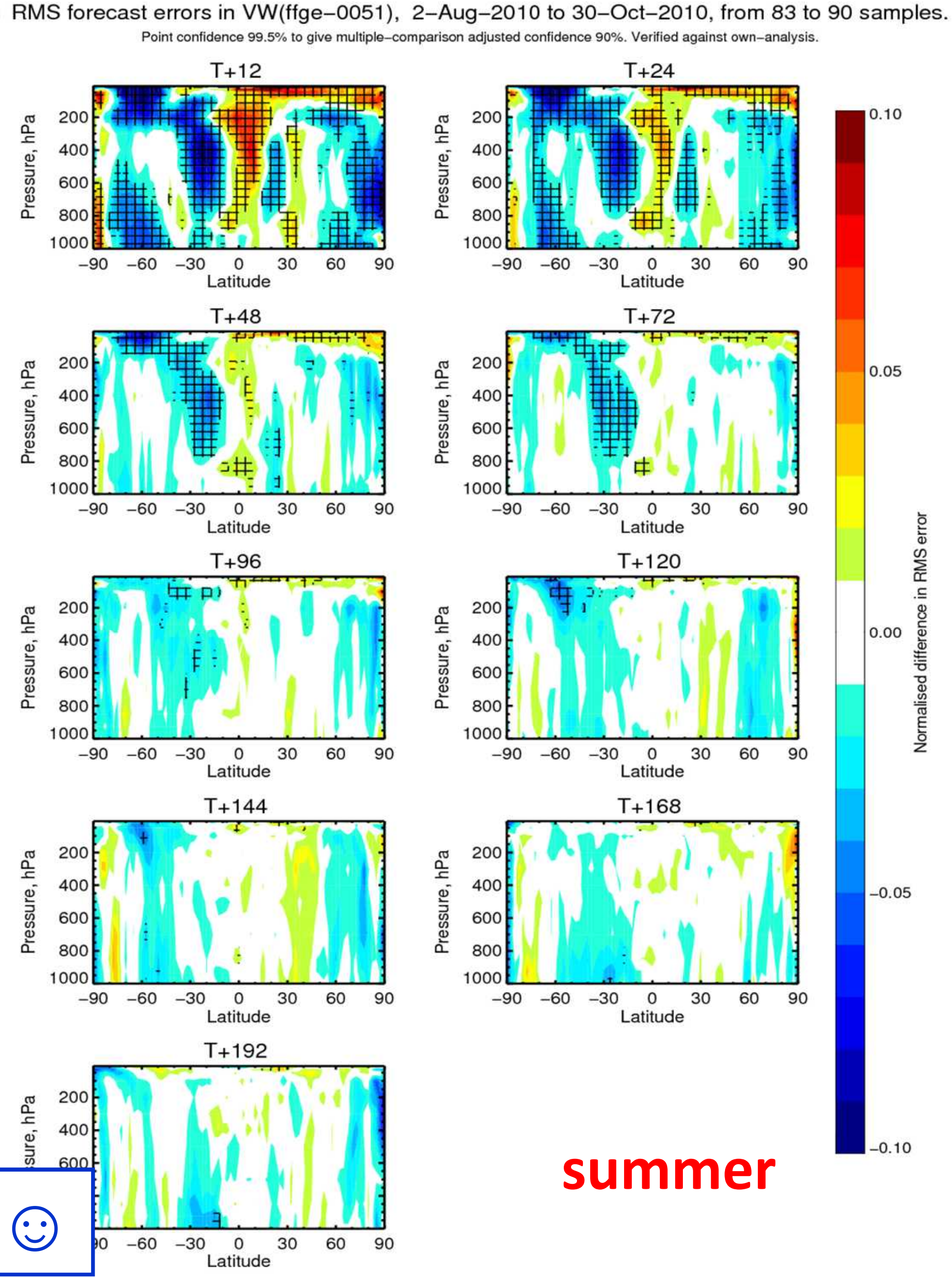
ECMWF

Temperature RMSE reduction

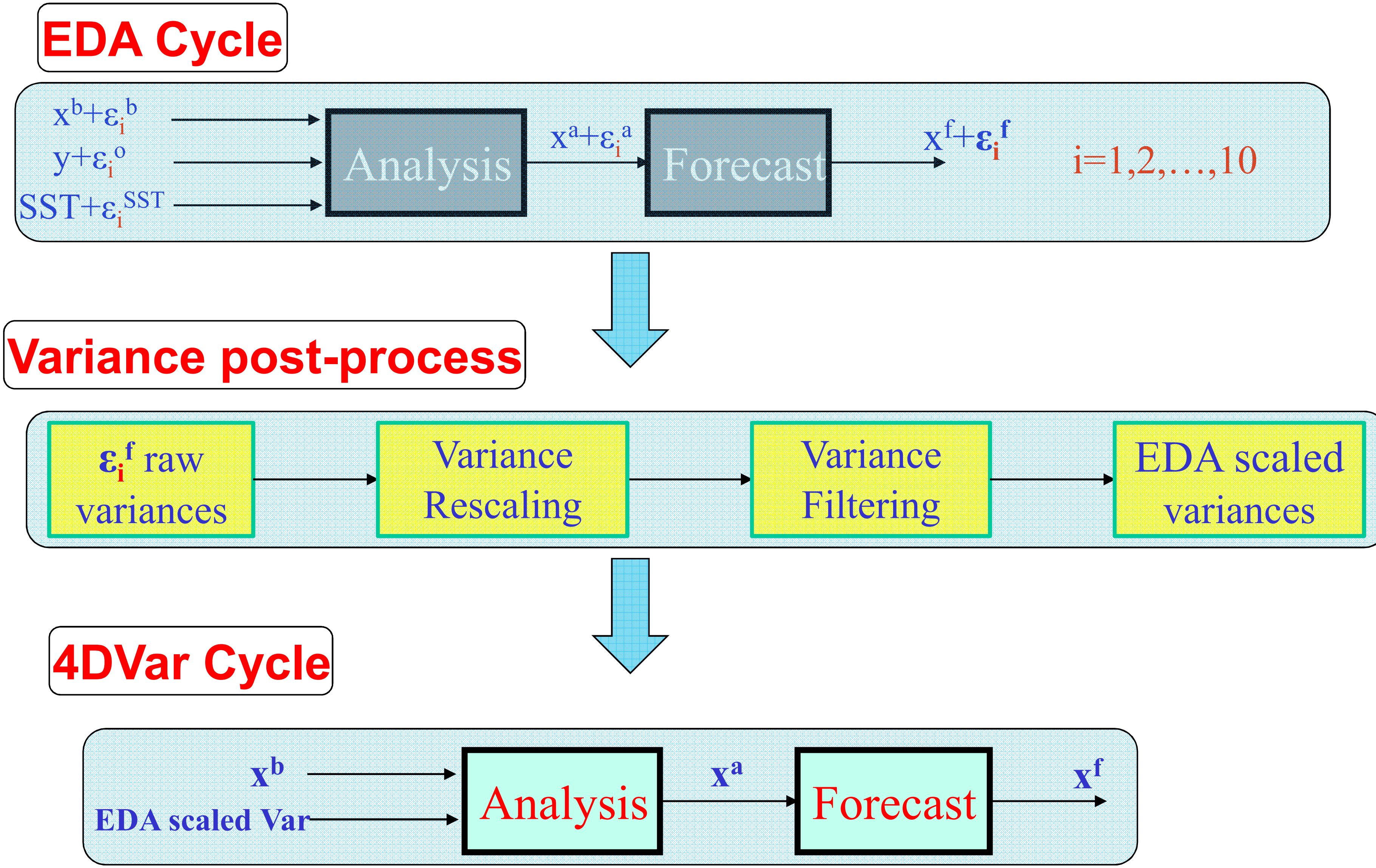


Wind Vector RMSE reduction





Blue= ☺


EDA Cycle
$x^b+\varepsilon_i^b$
$y+\varepsilon_i^o$
$SST+\varepsilon_i^{SST}$
Analysis
$x^a+\varepsilon_i^a$
Forecast
$x^f+\varepsilon_i^f$
i=1,2,…,10
Variance post-process
ε_i^f raw variances
Variance Rescaling
Variance Filtering
EDA scaled variances
4DVar Cycle
x^b
EDA scaled Var
Analysis
x^a
Forecast
x^f

EDA variances

a) Sampling Noise due to the small EDA dimensionality (N_{eda}=10)

The key insight is to recognise that *sampling noise is small scale with respect to the error variance field* (Raynaud *et al.*, 2008)

Define $G^e(i)$ as the sampling error in the estimated ensemble variance at gridpoint *i*: $G^e(i) \equiv \widetilde{B}_{ii} - E\left[\widetilde{B}_{ii}\right]$

Then the covariance of the sampling noise can be shown to be a simple function of the expectation of the ensemble-based covariance matrix:

$$E\left[G^e(i)G^e(j)\right] = \frac{2}{N-1}\left(E\left[\widetilde{B}_{ij}\right]\right)^2 \qquad (1)$$

A consequence of (1) is that $L_{G^e}(i)=L_{\varepsilon^b}(i)/\sqrt{2}$, i.e., sampling noise is smaller scale than background error.
The variance field varies on larger scales then the background error, so we may use a **spectral filter** to disentangle noise error from the variance field

ECMWF

EDA variances

There is indeed scale separation between signal and sampling noise!

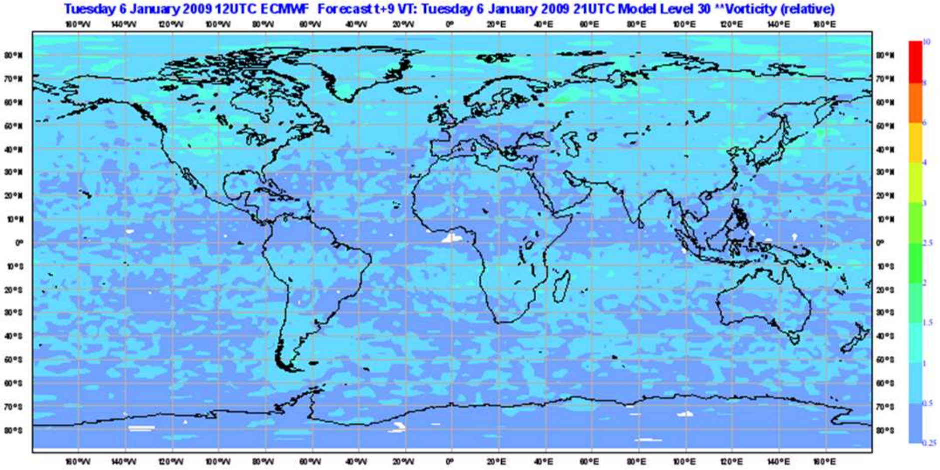
Truncation wavenumber is determined by **maximizing signal-to-noise** ratio of filtered variances (details in Raynaud *et al*., 2009; Bonavita *et al*., 2011)

Optimal truncation wavenumber depends on parameter and model level

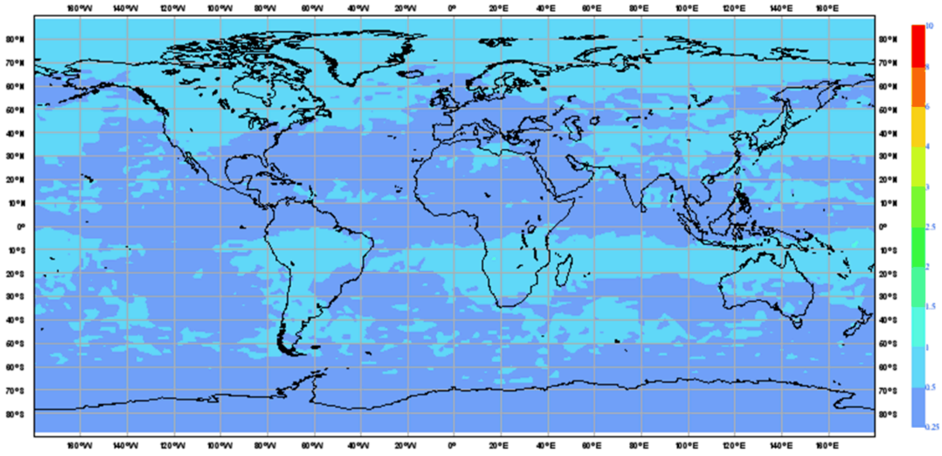
EDA variances

Vorticity ml 30 (~50hPa)

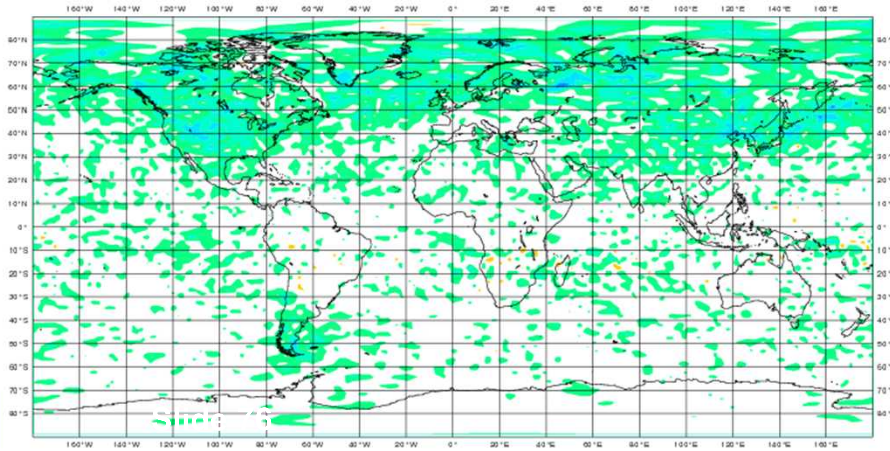
Ensemble Error



Ensemble Spread

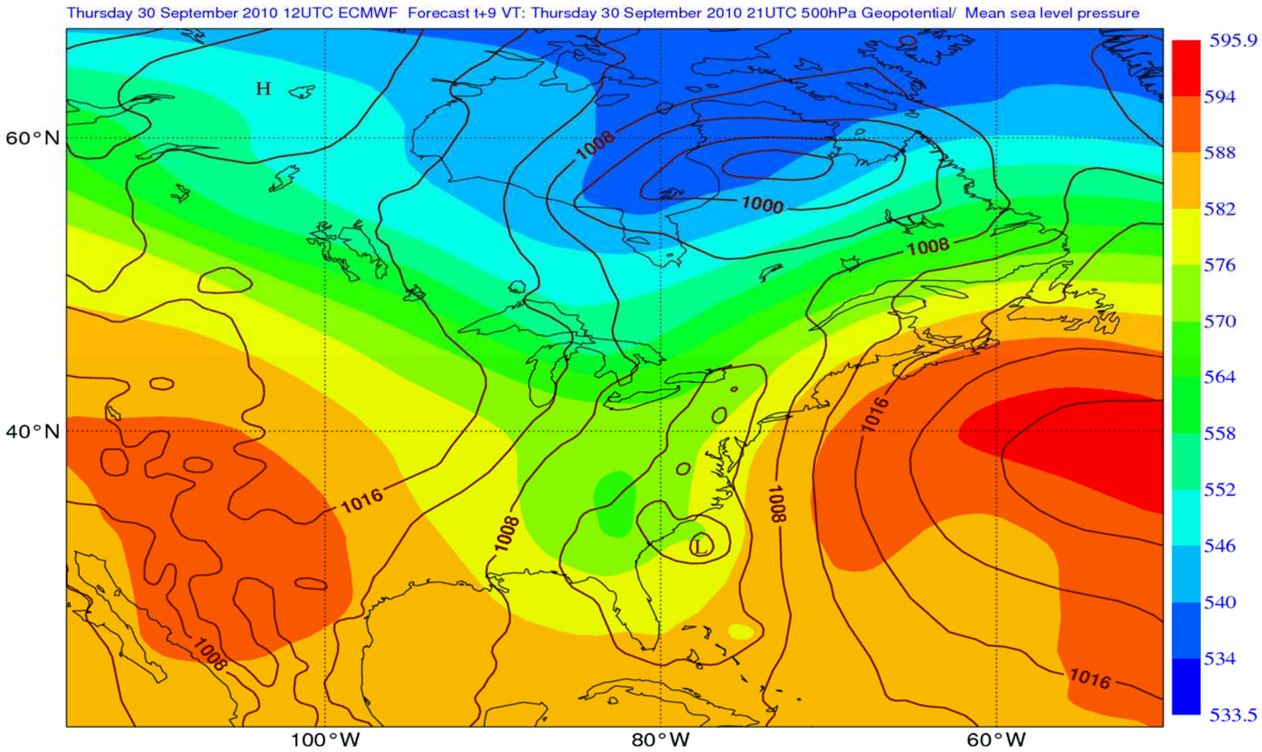


Spread - Error



EDA variances

How does a flow-dependent error variance estimate change the 4D-Var analysis?

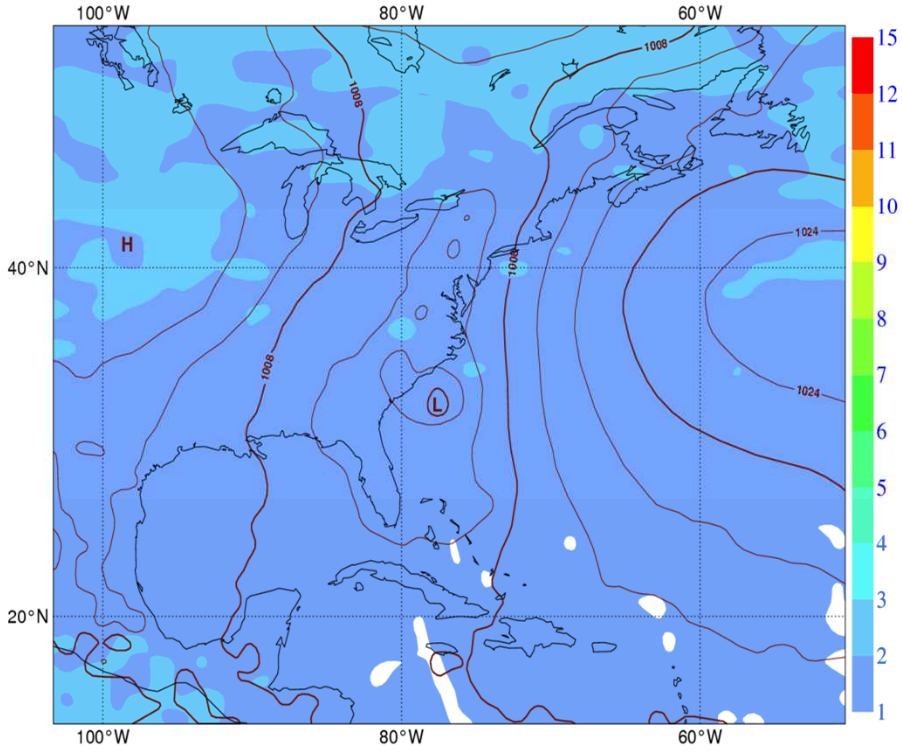


Z500 Geopotential (shaded) and MSLP 30-09-2010 21Z

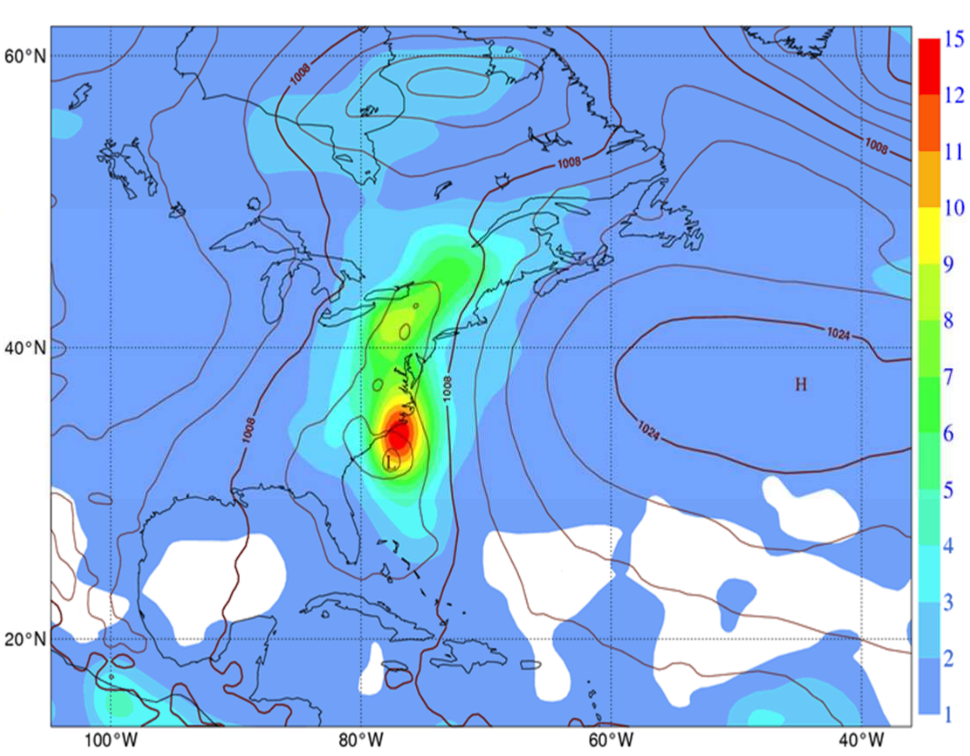
EDA variances

Vorticity Background Error ml=78 (850hPa)

“Randomization method”



EDA

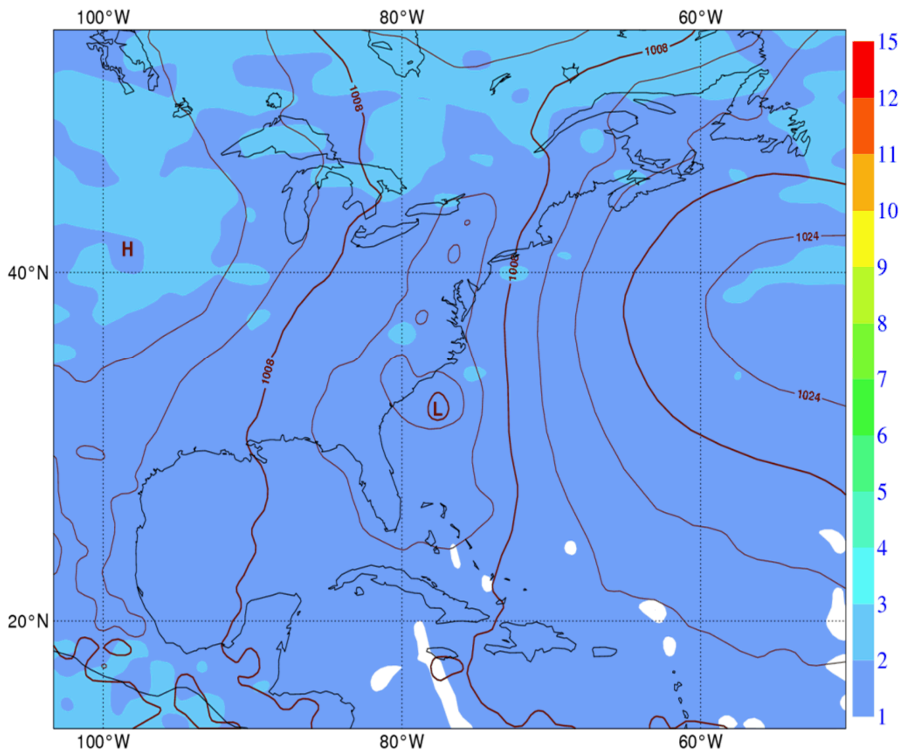


ECMWF

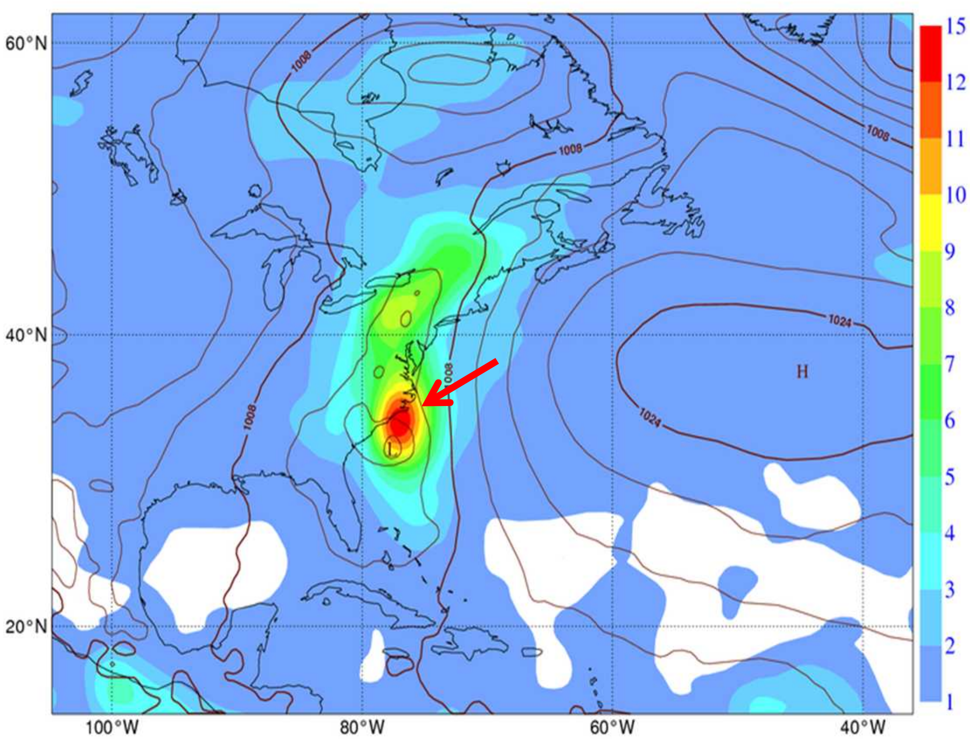
EDA variances

Single obs. experiment: T_{obs}-T_{fg}=+1K, (34N,74W), 900 hPa

“Randomization method”



EDA



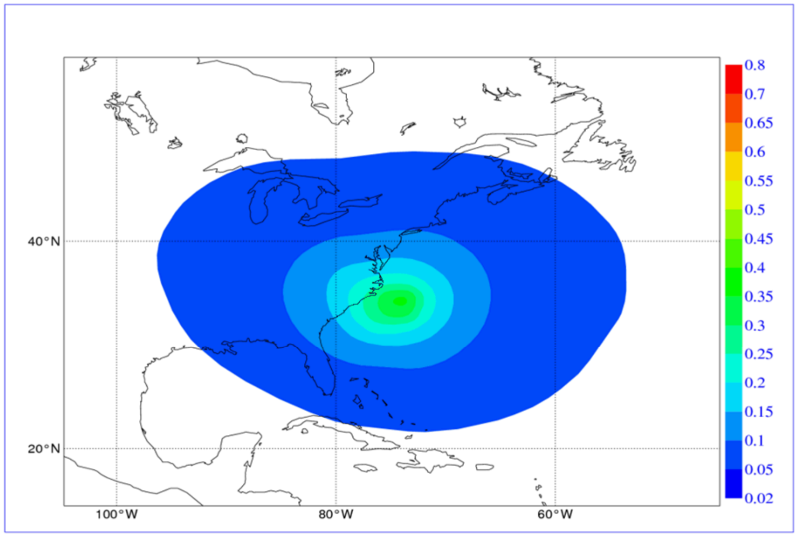
ECMWF

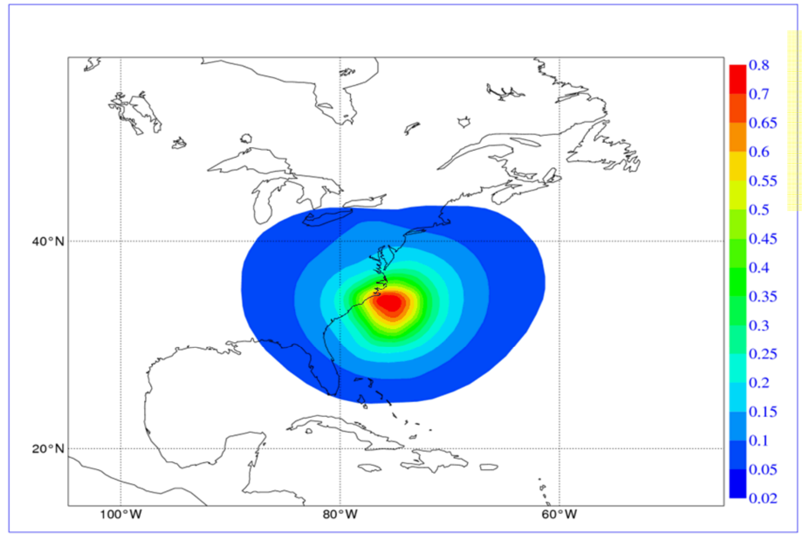
EDA variances

Single obs. experiment: $T_{obs}-T_{fg}$=+1K, (34N,74W), 900 hPa

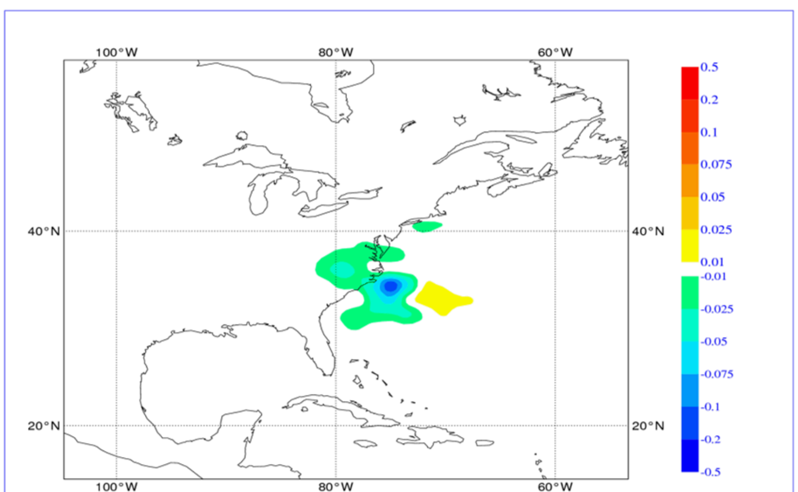
Randomization

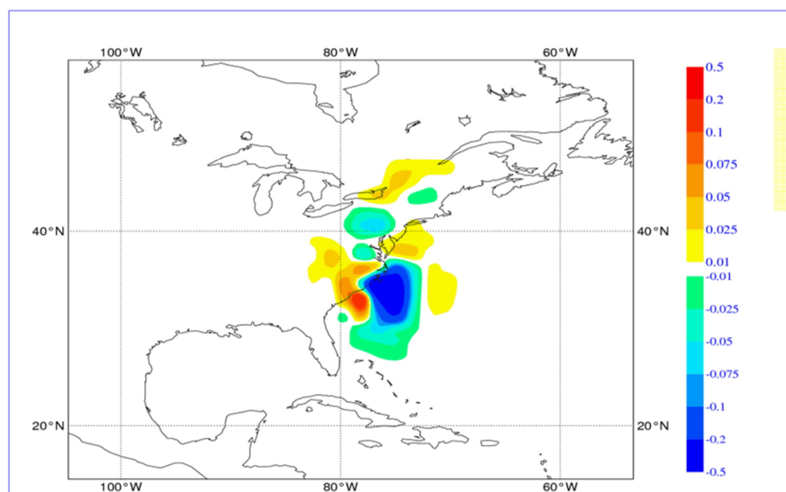
EDA





$T_{ana}-T_{fg}$ 900 hPa





$VO_{ana}-VO_{fg}$ 850 hPa

EDA variances

1. The observation weight in the analysis is increased in the area of large background uncertainty:

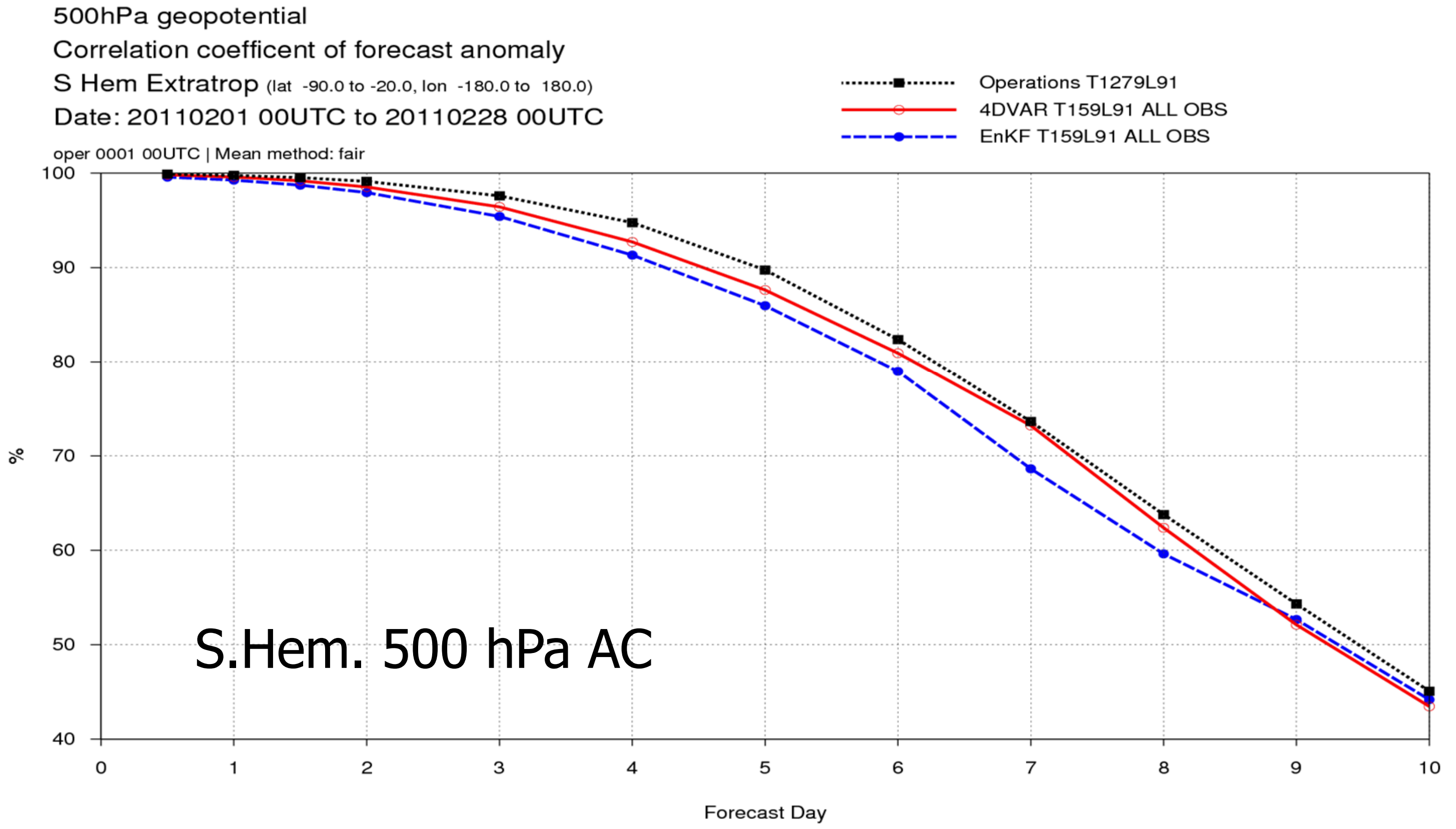
	EDA	Randomiz.
ΔT	0.75 K	0.37 K
ΔVO	5.E-5 s^{-1}	1.3E-5 s^{-1}

2. The EDA analysis increments show a degree of flow-dependency

Why do we need an EDA?

Results with the ECMWF EnKF

All observations



ECMWF