Assimilation Algorithms: EDA and Hybrid Data Assimilation

Massimo Bonavita
ECMWF
Massimo.Bonavita@ecmwf.int

Acknowledgements: Elias Holm, Mats Hamrud

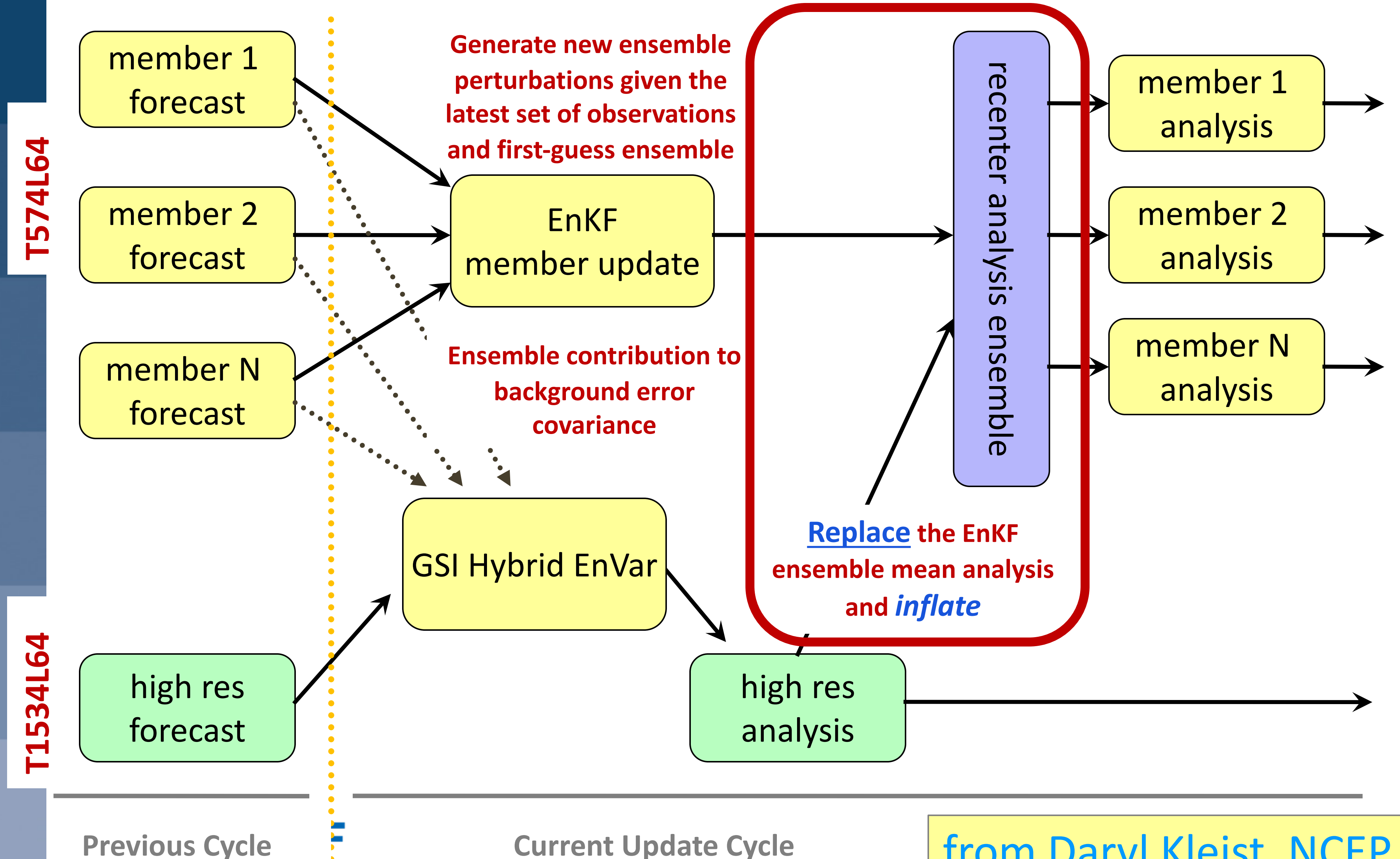
Outline

- A variational implementation of the EnKF: the EDA
- Hybrid Data Assimilation: What it is and why we do it
- Hybrid Data Assimilation systems in global NWP

The Ensemble of Data Assimilations (EDA)

- In the lecture on Ensemble Kalman Filters we have seen that EnKF are commonly used in hybrid DA systems for estimating and cycling error covariance information used by the variational analysis and initialise ensemble prediction systems
- Ensemble sizes of O(100-200) are commonly used. To save on computational cost the EnKF ensemble is run at a reduced spatial resolution (typically double grid point spacing) with respect to the deterministic variational analysis cycle
- The EnKF mean estimate is usually discarded: the EnKF members are re-centred on the high-res variational analysis at each analysis update

Dual-Res Coupled Hybrid Var/EnKF Cycling



from Daryl Kleist, NCEP

Ensemble Data Assimilation

- Can we replicate the error cycling job done by the EnKF using only 4D-Var?
- The answer is yes, by applying the same error simulation concepts used for the stochastic (perturbed observations) EnKF
- Ensemble of Data Assimilations (EDA; Isaksen et al., 2010)

Ensemble Data Assimilation

- For a linear system (linear model **M**, linear observation operator **H**) the data assimilation update can be written as:

$$\mathbf{x}_t^a = \mathbf{x}_t^b + \mathbf{K}\left(\mathbf{y}_t - \mathbf{H}\mathbf{x}_t^b\right)$$
$$\mathbf{x}_{t+1|}^b = \mathbf{M}\mathbf{x}_t^a \qquad (1)$$

- Assuming background (**P**^{b}), observation (**R**) and model errors (**Q**) to be statistically independent, the evolution of the system error covariances is given by:

$$\mathbf{P}_t^a = \left(\mathbf{I} - \mathbf{K}\mathbf{H}\right)\mathbf{P}_t^b\left(\mathbf{I} - \mathbf{K}\mathbf{H}\right)^T + \mathbf{K}\mathbf{R}\mathbf{K}^T$$
$$\mathbf{P}_{t+1}^b = \mathbf{M}\mathbf{P}_t^a\mathbf{M}^T + \mathbf{Q} \qquad (2)$$

- Consider now the evolution of this system if we perturb the observations and the forecast model with random, zero mean noise drawn from the respective error covariances, i.e. $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$, $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$:

$$\tilde{\mathbf{x}}_t^a = \tilde{\mathbf{x}}_t^b + \mathbf{K}\left(\boldsymbol{y}_t + \boldsymbol{\zeta} - \mathbf{H}\tilde{\mathbf{x}}_t^b\right) \qquad (3)$$
$$\tilde{\mathbf{x}}_t^b = \mathbf{M}\tilde{\mathbf{x}}_t^a + \boldsymbol{\eta}$$

where $\tilde{\mathbf{x}}_t^{a/b}$ represent the perturbed system analyses/backgrounds

Ensemble Data Assimilation

- Consider the differences between perturbed states and the unperturbed one:

$$\boldsymbol{\varepsilon}_t^{a/b} = \tilde{\mathbf{x}}_t^{a/b} - \mathbf{x}_t^{a/b}$$

their evolution in a Kalman Filter-type of DA system is obtained by subtracting the unperturbed state evolution equations from the perturbed ones, i.e. (3)-(1):

$$\begin{aligned} \boldsymbol{\varepsilon}_t^{a} &= \boldsymbol{\varepsilon}_t^{b} + \mathbf{K}\left(\boldsymbol{\zeta} - \mathbf{H}\boldsymbol{\varepsilon}_t^{b}\right) \\ \boldsymbol{\varepsilon}_{t+1}^{b} &= \mathbf{M}\boldsymbol{\varepsilon}_t^{a} + \boldsymbol{\eta} \end{aligned} \qquad (4)$$

- We see that the perturbations evolve with same update equations as the system
- What about their covariances?

$$\left\langle \boldsymbol{\varepsilon}_t^{a}\left(\boldsymbol{\varepsilon}_t^{a}\right)^{\mathrm{T}} \right\rangle = \left\langle \boldsymbol{\varepsilon}_t^{b} + \mathbf{K}\left(\boldsymbol{\zeta} - \mathbf{H}\boldsymbol{\varepsilon}_t^{b}\right)\left(\boldsymbol{\varepsilon}_t^{b} + \mathbf{K}\left(\boldsymbol{\eta} - \mathbf{H}\boldsymbol{\varepsilon}_t^{b}\right)\right)^{T} \right\rangle =$$

$$\left\langle \boldsymbol{\varepsilon}_t^{b}\left(\boldsymbol{\varepsilon}_t^{b}\right)^{\mathrm{T}} \right\rangle - \left\langle \boldsymbol{\varepsilon}_t^{b}\left(\boldsymbol{\varepsilon}_t^{b}\right)^{\mathrm{T}} \right\rangle \mathbf{H}^T\mathbf{K}^T - \mathbf{K}\mathbf{H}\left\langle \boldsymbol{\varepsilon}_t^{b}\left(\boldsymbol{\varepsilon}_t^{b}\right)^{\mathrm{T}} \right\rangle + \mathbf{K}\left\langle \boldsymbol{\zeta}(\boldsymbol{\zeta})^{\mathrm{T}} \right\rangle \mathbf{K}^T + \mathbf{K}\mathbf{H}\left\langle \boldsymbol{\varepsilon}_t^{b}\left(\boldsymbol{\varepsilon}_t^{b}\right)^{\mathrm{T}} \right\rangle \mathbf{H}^T\mathbf{K}^T =$$

$$\left(\mathbf{I} - \mathbf{K}\mathbf{H}\right)\left\langle \boldsymbol{\varepsilon}_t^{b}\left(\boldsymbol{\varepsilon}_t^{b}\right)^{\mathrm{T}} \right\rangle \left(\mathbf{I} - \mathbf{K}\mathbf{H}\right)^{\mathrm{T}} + \mathbf{K}\mathbf{R}\mathbf{K}^{\mathrm{T}} \qquad (5)$$

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

Ensemble of Data Assimilation

- We see that the error statistics from a DA system run with perturbed observations and model error perturbations evolve with the same update equations as the error statistics of the unperturbed DA cycle, in the limit of a large ensemble:

$$\left\langle \boldsymbol{\varepsilon}_t^a \left(\boldsymbol{\varepsilon}_t^a\right)^{\mathrm{T}} \right\rangle = (\mathbf{I}-\mathbf{KH})\left\langle \boldsymbol{\varepsilon}_t^b \left(\boldsymbol{\varepsilon}_t^b\right)^{\mathrm{T}} \right\rangle (\mathbf{I}-\mathbf{KH})^{\mathrm{T}} + \mathbf{KRK}^{\mathrm{T}}$$

$$\left\langle \boldsymbol{\varepsilon}_{k+1}^b \left(\boldsymbol{\varepsilon}_{t+1}^b\right)^{\mathrm{T}} \right\rangle = \mathbf{M}\left\langle \boldsymbol{\varepsilon}_t^a \left(\boldsymbol{\varepsilon}_t^a\right)^{\mathrm{T}} \right\rangle \mathbf{M}^{\mathrm{T}} + \mathbf{Q} \qquad (5)$$

$$\mathbf{P}_{\mathbf{t}}^{\mathbf{a}} = (\mathbf{I}-\mathbf{KH})\mathbf{P}_{\mathbf{t}}^{\mathbf{b}}(\mathbf{I}-\mathbf{KH})^{\mathrm{T}} + \mathbf{KRK}^{\mathrm{T}}$$

$$\mathbf{P}_{\mathbf{t+1}}^{\mathbf{b}} = \mathbf{MP}_{\mathbf{t}}^{\mathbf{a}}\mathbf{M}^{\mathrm{T}} + \mathbf{Q} \qquad (2)$$

- For this recipe to work, however, we also require to draw perturbations from the correct error covariance matrices **R** (observation errors) and **Q** (model errors)

- The practical implementation of an Ensemble Data Assimilation system presents us with two challenges:

    a) Computational cost as large ensemble required to control sampling errors;

    b) Scientific challenge of estimation of realistic observation and model errors

Ensemble Data Assimilation

- What does all this mean in practice?

  1. We can use an ensemble of perturbed data assimilation cycles to simulate the errors of our reference DA cycle;

  2. The *ensemble of perturbed DAs should be as similar as possible to the reference DA* (i.e., same or similar $\mathbf{K}$ matrix, $\mathbf{M}$, $\mathbf{H}$, and resolution)

  3. The applied *perturbations ζ,η should be drawn from the correct error covariances* (**R, Q**);

  4. There is no need to explicitly perturb the background forecast $\mathbf{x}_b$, if the perturbations are drawn from the correct error covariances (**R, Q**);

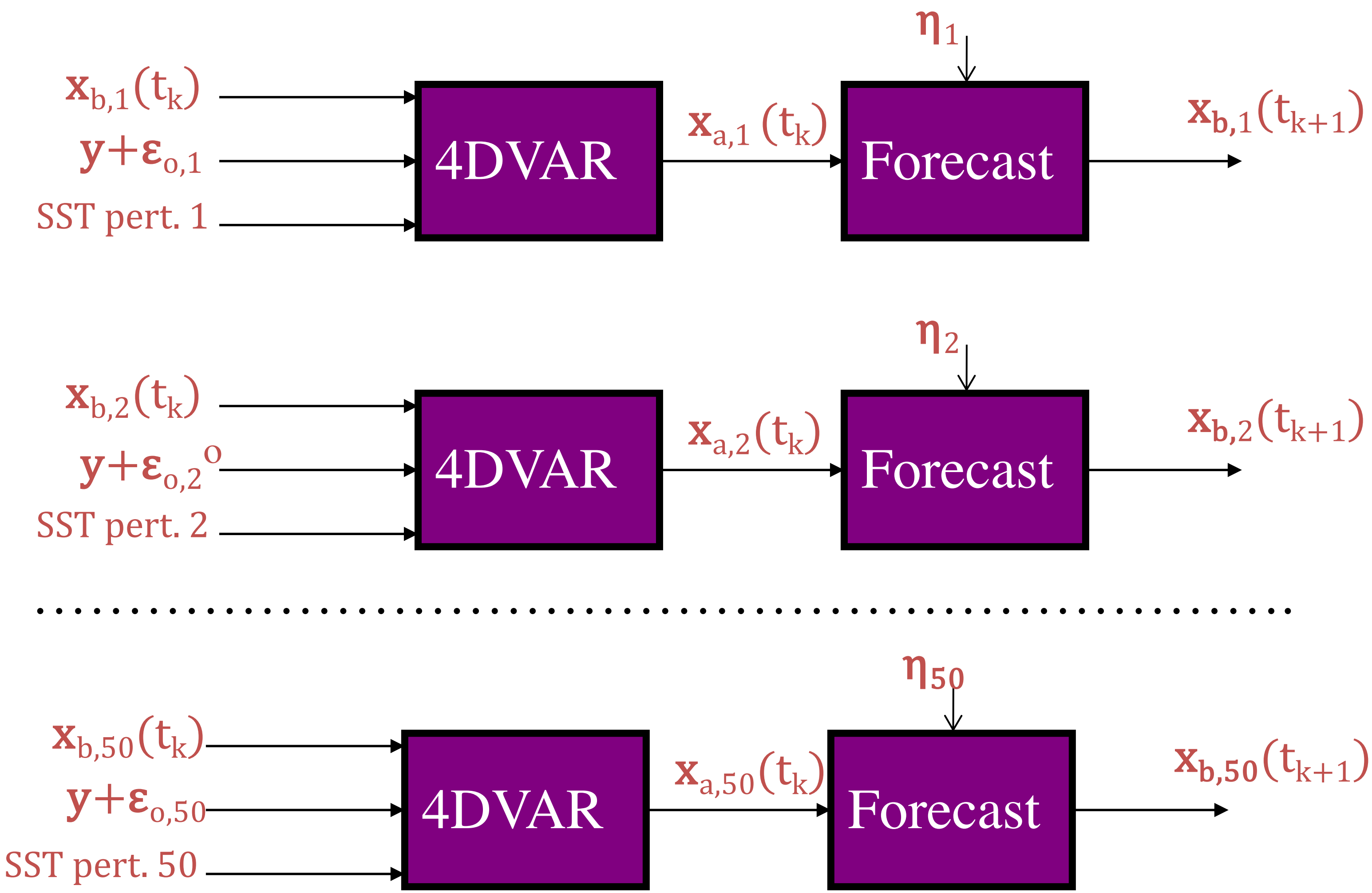
  5. This is a Monte Carlo method: the expectation operators used in (5) imply that results strictly hold for large ensemble sizes. In practice non-negligible sampling errors are to be expected

The Ensemble of Data Assimilations (EDA)

- How do we currently (March 2024) do it?

  1. 50 independent perturbed ensemble members using 4D-Var assimilations at reduced resolution
  2. TCo639 outer loop, TL191/TL191 inner loops. (HRES DA: TCo1279 outer loop, TL255/TL319/TL399/T399 inner loops).
  3. Observations randomly perturbed according to their estimated error covariances $(\mathbf{R})$, currently diagonal is physical space
  4. SST perturbed with climatological error structures (not yet coupled to the Ocean DA)
  5. Model error $(\mathbf{Q})$ represented by stochastic perturbations during the background forecast integration (SPPT, Leutbecher, 2009)

The Ensemble of Data Assimilations (EDA) March 2024



The Ensemble of Data Assimilations (EDA)

- How will we do it from Q3 2024 (IFS Cycle 49r1)?

  1. 50 perturbed ensemble members using 4D-Var assimilations at same outer loop resolution as the high-res 4D-Var control (TCo1279, ~9 km grid spacing)

  2. Perturbed members soft-recentred on high-res control background (plus a few other things, see Holm et al.,

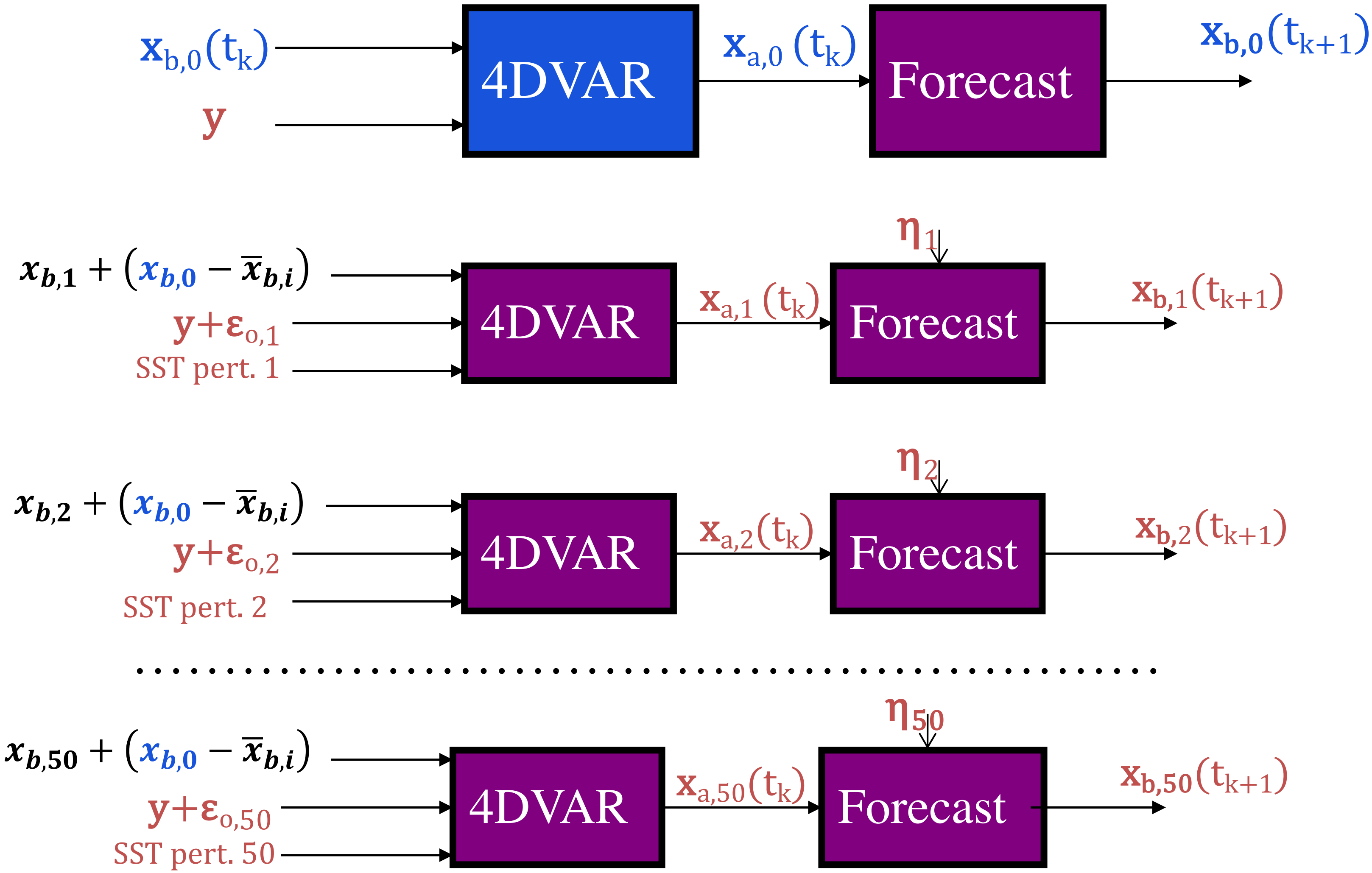
  3. TCo1279 outer loop, TL399 inner loop min.

  4. Observations randomly perturbed according to their estimated error covariances $(\mathbf{R})$, currently diagonal is physical space

  5. SST perturbed with climatological error structures (not yet coupled to the Ocean DA)

  6. Model error $(\mathbf{Q})$ represented by stochastic perturbations during the background forecast integration (SPP, Stochastically Perturbed Parametrisations. Lang et al., 2021)

The Ensemble of Data Assimilations (EDA) Cy49r1

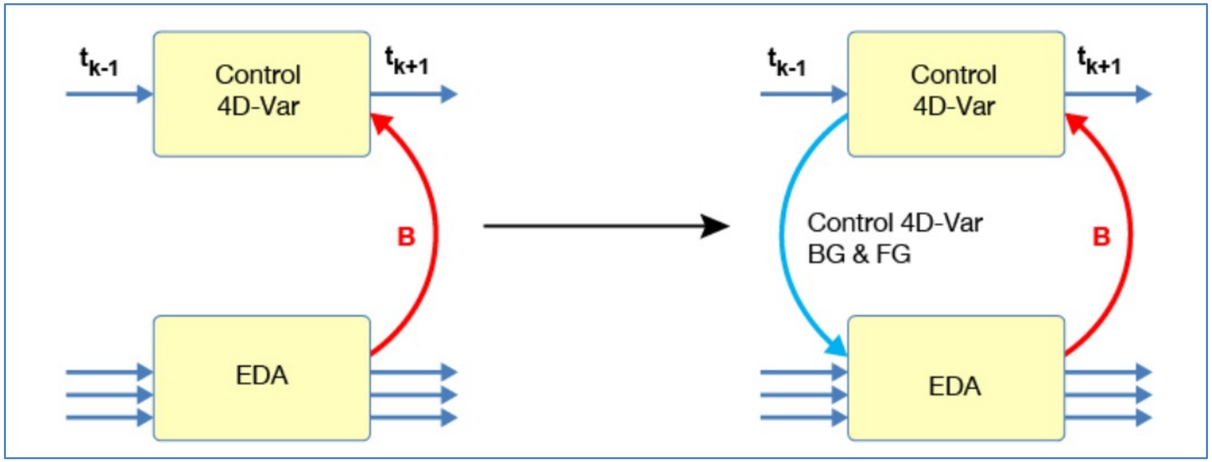


Soft-centred EDA

- A new, computationally cheaper and more effective version of the ECMWF EDA is scheduled to go into operations in Q3 2024 (IFS Cycle 49r1): Soft-centred EDA (Hólm, Bonavita and Lang, 2022)

- Current EDA: All members are independent including the control (unperturbed member). The control member is only different from the perturbed members because no observation and model stochastic perts are applied

- Soft-centred EDA: 1) Control member run with more and higher resol. inner loops than perturbed members 2) Control member provides bg and fg for perturbed members



Soft-centred EDA

- In standard "hard-centred" hybrid DA the ensemble mean analysis is discarded and the ensemble analysis is centred on the control (unperturbed, higher res) analysis
- In Soft-centred EDA the ensemble mean background is discarded in favour of the control background -> each member perform an independent analysis -> vastly reduces imbalances in IC, helps maintain spread
- Computational costs reduced ~30%; impact on ensemble prediction performance largely positive

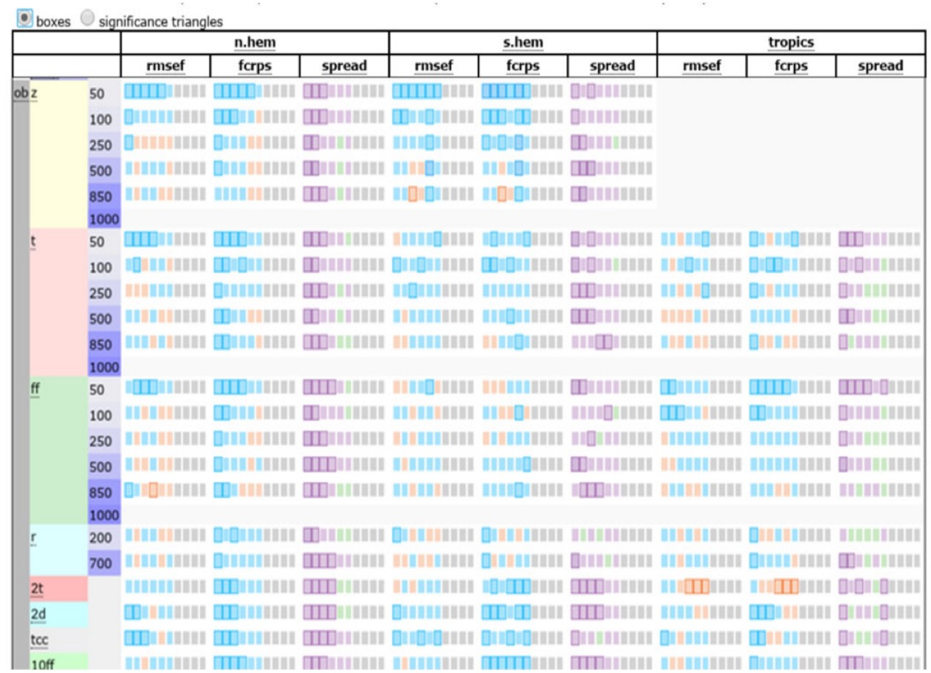


Figure 46: Scorecard measuring the forecast performance difference between two ENS prediction experiments, one using initial conditions from the standard EDA and the other from soft-centred EDA. Blue/orange squares indicate better/worse performance of the ENS started from the soft-centred EDA, squares indicate statistical significance at the 95% confidence level. Purple/green colours indicate larger/smaller spread for the soft-centred EDA ENS experiment. Verification against observations.

ECMWF Scientific Advisory Committee 2021, Weather and Earth System science report

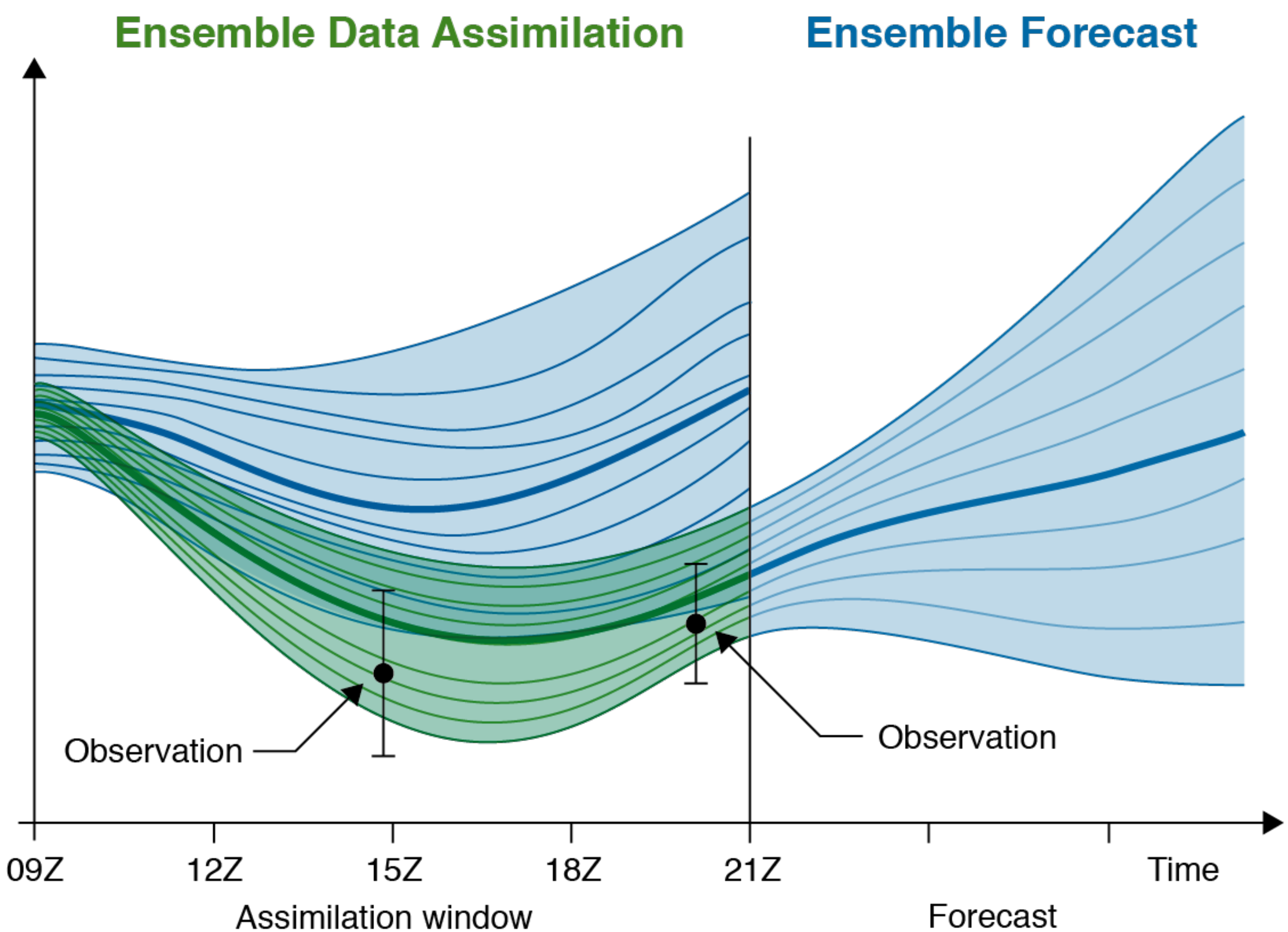
The Ensemble of Data Assimilations (EDA)

- The EDA simulates the error evolution of the 4DVar analysis cycle. As such it has two main applications:

    1. Provide an ensemble of initial conditions to initialize the ensemble prediction system (EPS)

    2. Provide a flow-dependent estimate of background error covariances for use in the 4D-Var assimilation (both for the HRES DA system and the EDA members themselves)

The Ensemble of Data Assimilations (EDA)



The Ensemble of Data Assimilations (EDA)

- The EDA is the system used at ECMWF to simulate the error evolution of the 4DVar analysis cycle.
- It is conceptually similar and it is based on the same assumptions of the Perturbed Observations (Stochastic) EnKF.
- There are advantages for ECMWF to using an ensemble of 4DVars to simulate the error of a reference high resolution 4DVar:

    1. The two systems are more similar to one another in terms of Kalman Gain than an EnKF and a 4DVar; error estimates should thus be more accurate
    2. There are obvious technical and maintenance synergies

- There are also disadvantages. In particular running an ensemble of 4DVar is computationally more expensive than running an EnKF. Current efforts are aimed at reducing the computational costs of the EDA

Hybrid Data Assimilation

•

Hybrid Data Assimilation: Motivation

If we neglect model error (perfect model assumption) the 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 $\mathbf{P}^b$ can be solved by finding the minimum of the 4D-Var cost function:

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

The 4D-Var solution is equivalent, for the same $\mathbf{x}_b$, $\mathbf{P}^b$, and linear $\mathbf{H}$, $\mathbf{M}$, to the Kalman Filter solution at the end of the assimilation window (t=T) (Fisher *et al*, 2005).

Hybrid Data Assimilation: Motivation

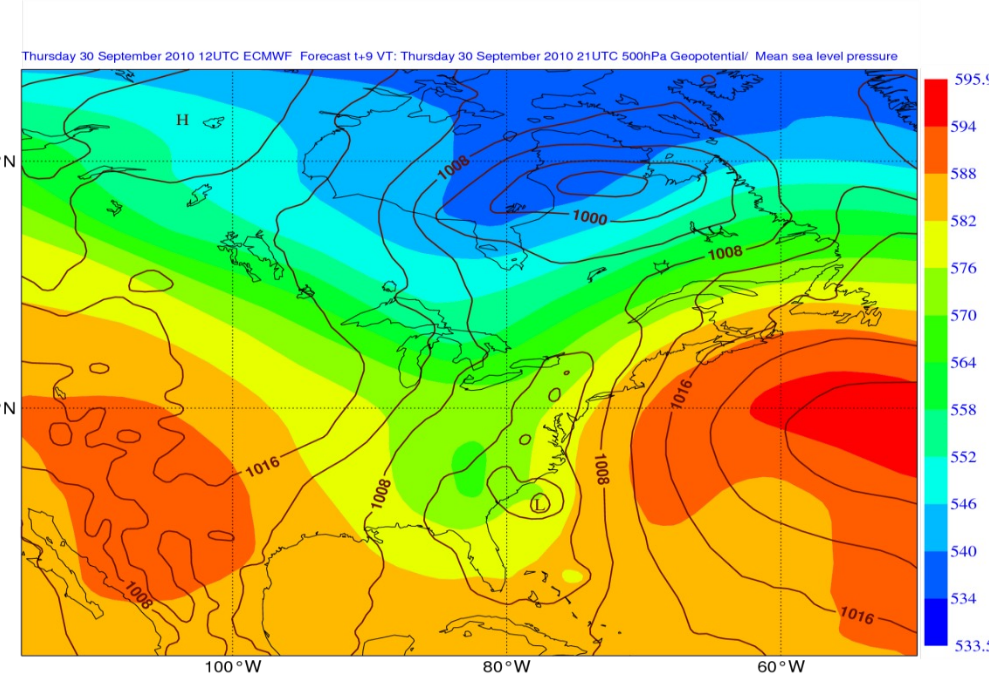
The 4D-Var solution implicitly evolves the initial background error covariances *over the length of the assimilation window* (Thepaut *et al*.,1996) with the tangent linear dynamics:

$$\mathbf{P}^{\mathrm{b}}(\mathrm{t}) \approx \mathbf{M}\mathbf{P}^{\mathrm{b}}(\mathrm{t}{=}0)\mathbf{M}^{\mathrm{T}}$$

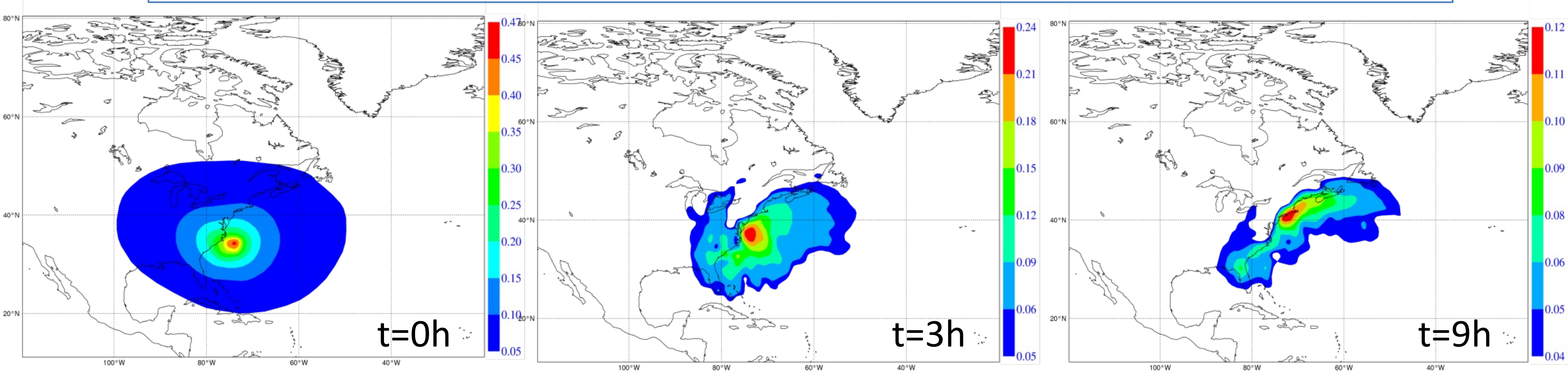
This effect can be seen most easily looking at the evolution of the analysis increment for a single observation during an assimilation window:

$$\mathbf{x}_a(t) - \mathbf{x}_b(t) \cong \mathbf{M}\mathbf{P}^b(t=0)\mathbf{M}^T\mathbf{H}^T\big(y - H(\mathbf{x}_b)\big)/\big(\sigma_b^2 + \sigma_o^2\big)$$

MSLP (solid lines)
500 hPa Z (shaded)
background forecast



Temperature analysis increments for a single temperature observation at the start of the assimilation window



Hybrid Data Assimilation: Motivation

The 4D-Var solution implicitly evolves the initial background error covariances *over the length of the assimilation window* (Thepaut *et al*.,1996) with the tangent linear dynamics:

$$\mathbf{P}^{b}(t) \approx \mathbf{M}\mathbf{P}^{b}\mathbf{(t=0)}\mathbf{M}^{T}$$

but it does not propagate error information from one assimilation cycle to the next. $\mathbf{P}^{b}$ is not evolved according to Kalman Filter equations (i.e., $\mathbf{P}^{b} = \mathbf{M}\mathbf{P}^{a}\mathbf{M}^{T} + \mathbf{Q}$) but is reset to a climatological, stationary estimate at the beginning of each assimilation window.

In standard 4D-Var only information about the state $(\mathbf{x}_{b})$ is propagated from one cycle to the next.

Hybrid Data Assimilation: Motivation

Hybrid Data Assimilation: Use an EnKF/EDA system to produce flow-dependent error covariance information to be used in the high resolution Variational analysis

The hybrid approach would have the benefit of:

1) Integrate flow-dependent state error covariance information into the variational analysis

2) Keep the full rank representation of $\mathbf{P}^b$ and its implicit evolution inside the assimilation window

3) More robust than pure EnKF/EnVar for limited ensemble sizes and large model errors

4) Allow for flow-dependent quality control of observations

Hybrid Data Assimilation: Applications

The next question to address is: how do we integrate the flow-dependent error covariance information from the EnKF/EDA systems into the variational analysis?

1. Augmented control variable method (Met Office)
2. 4D-Ensemble-Var (NCEP, CMC, DWD)
3. Hybrid EDA 4D-Var (ECMWF, Météo France)
4. Other forms of hybrid systems: Hybrid-Gain

Hybrid Data Assimilation: Applications

Augmented (alpha) control variable (Lorenc, 2003)

Conceptually it adds a flow-dependent term to the background error model:

$$\mathbf{P}^b = \beta_c^2 \mathbf{P}^b{}_{c\,\mathrm{lim}} + \beta_e^2 \mathbf{P}_{ens} \circ \mathbf{C}_{loc}$$

$\mathbf{P}^b{}_{\mathrm{clim}}$ is the static, climatological background error covariance

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

In practice this hybrid covariance model is done through augmentation of the control variable (more on this in the **B** modelling lecture):

$$\delta \boldsymbol{x} = \boldsymbol{\beta}_c \boldsymbol{P}^{b\;1/2}_{clim} \boldsymbol{\chi} + \boldsymbol{\beta}_e \mathbf{X}' \circ \boldsymbol{\alpha} \begin{pmatrix} 1 \\ .. \\ 1 \end{pmatrix}$$

and introducing an additional term in the 4D-Var cost function:

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

Hybrid Data Assimilation: Applications

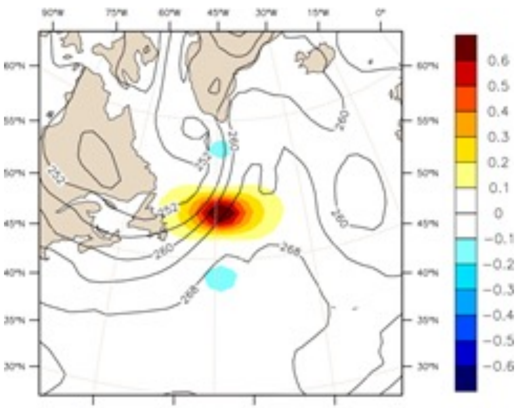
Augmented (alpha) control variable

$$\delta \boldsymbol{x} = \boldsymbol{\beta}_c \boldsymbol{P}^{\boldsymbol{b}\,1/2}_{clim} \boldsymbol{\chi} + \boldsymbol{\beta}_e \mathbf{X}' \circ \boldsymbol{\alpha} \begin{pmatrix} 1 \\ .. \\ 1 \end{pmatrix} = \boldsymbol{\delta x}_{clim} + \boldsymbol{\delta x}_{ens}$$

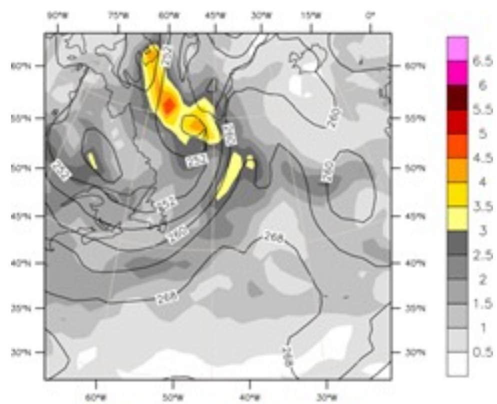
- The analysis increment is now a weighted sum of a component from the static, climatological $\mathbf{P}^{b}_{clim}$ and a component from the flow-dependent, ensemble based $\mathbf{P}^{b}_{ens}$
- *The flow-dependent increment is a linear combination of the ensemble background perturbations X', spatially modulated by the α fields of coefficients*
- If the $\boldsymbol{\alpha}$ fields were homogeneous $\delta\mathbf{x}_{ens}$ could only span N_{ens}-1 degrees of freedom; instead $\boldsymbol{\alpha}$ are spatially varying fields, which effectively increases the available degrees of freedom since at different grid points the increment will be a different linear combination of ensemble perturbations
- $\mathbf{C}_{loc}$ is a covariance (localization) model for the flow-dependent increments: it controls the spatial variation of $\boldsymbol{\alpha}$

Hybrid Data Assimilation: Applications

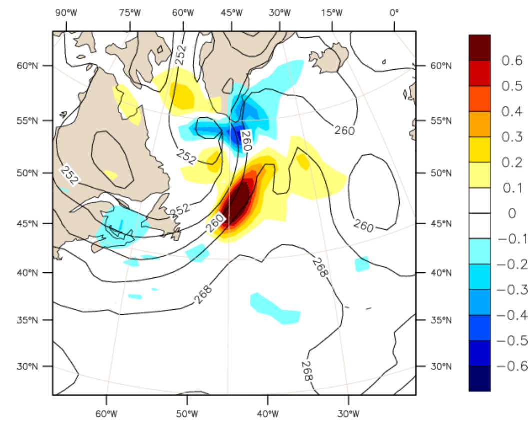
Extended (alpha) control variable



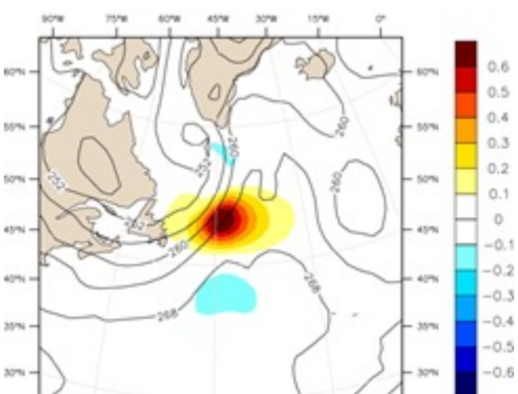
Standard 3D-Var



Ensemble RMS

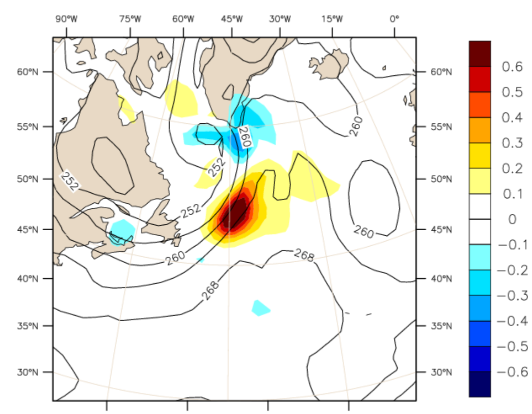


Pure ensemble 3D-Var



Standard 4D-Var

from A.Clayton (MO)

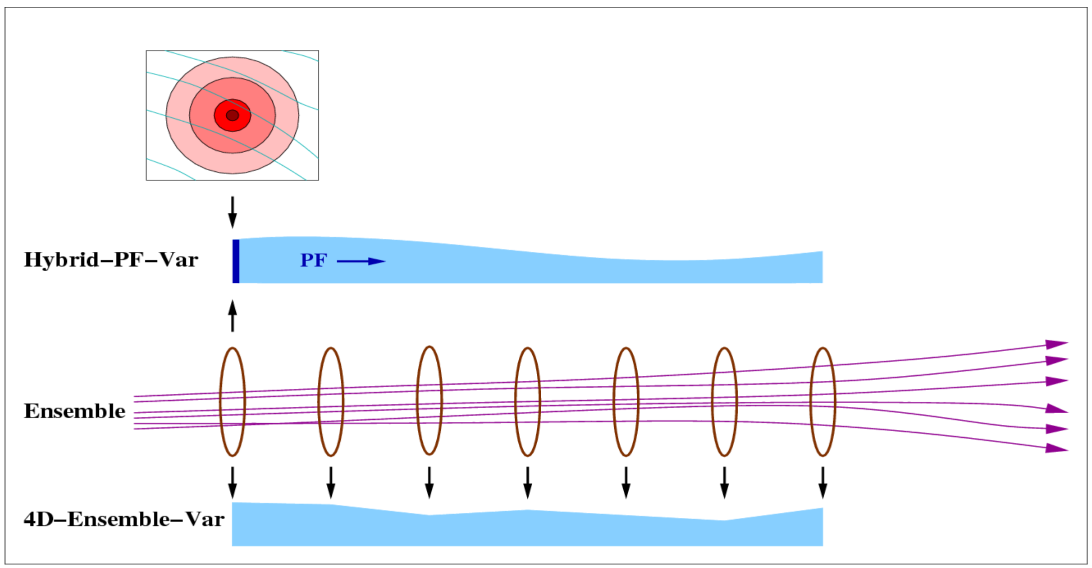


50/50 hybrid 3D-Var

Hybrid Data Assimilation: Applications

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

- In the alpha control variable method one uses the ensemble perturbations to estimate $\mathbf{P}^{b}$ only at the start of the 4D-Var assimilation window: the evolution of $\mathbf{P}^{b}$ inside the window is done by the tangent linear dynamics ($\mathbf{P}^{b}(t) \approx \mathbf{M}\mathbf{P}^{b}\mathbf{M}^{T}$)
- In 4D-En-Var $\mathbf{P}^{b}$ is sampled from ensemble trajectories throughout the assimilation window:



from D. Barker

Hybrid Data Assimilation: Applications

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

- The 4D-Ens-Var analysis increment is a localised linear combination of ensemble trajectories' perturbations:

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

$$\mathbf{x}_k^{'}(t) = \mathbf{x}_\mathbf{k}(t) - \overline{\mathbf{x}_\mathbf{k}(t)}$$

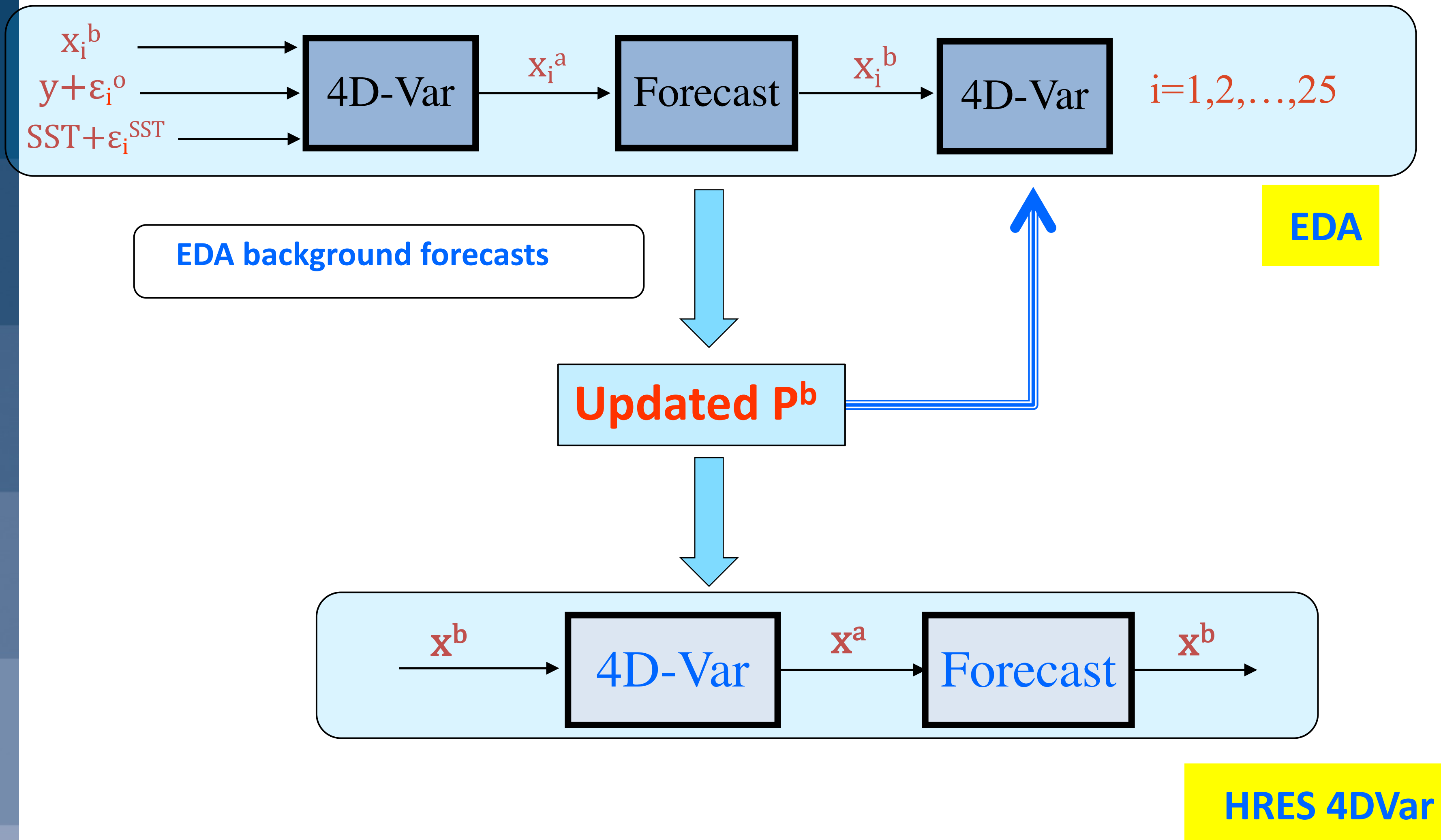
- This is fundamentally the same state update procedure of the LETKF version of EnKF (Hunt *et al.*, 2007)
- While traditional 4D-Var requires repeated, sequential runs of **M**, $\mathbf{M}^\mathbf{T}$, ensemble trajectories from the previous assimilation time can be pre-computed in parallel
- However these ensemble trajectories need to be stored and read-in: we are trading computational cost for I/O and memory cost
- As in the EnKF, 4D-Ens-Var does not require developing and maintaining the TL and Adjoint models, which makes it popular!

Hybrid Data Assimilation: Applications

Hybrid EDA 4D-Var (Bonavita *et al.*, 2012, 2015)

- In Hybrid 4D-Var we use the perturbations from the EDA background forecasts to update the background error covariance model used in 4D-Var
- The ensemble perturbations are not used directly to construct the analysis increments, but to update the modelled $\mathbf{P}^b(t=0)$ used in 4D-Var

Hybrid Data Assimilation: Applications



Hybrid Data Assimilation: Applications

How does the background error model update works?

- In variational DA, the background error covariance matrix **B** is usually defined implicitly in terms of a transformation from an increment defined in terms of model variables ($\mathbf{x}\text{-}\mathbf{x}_b$) to one defined in terms of a control variable χ:

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

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

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

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

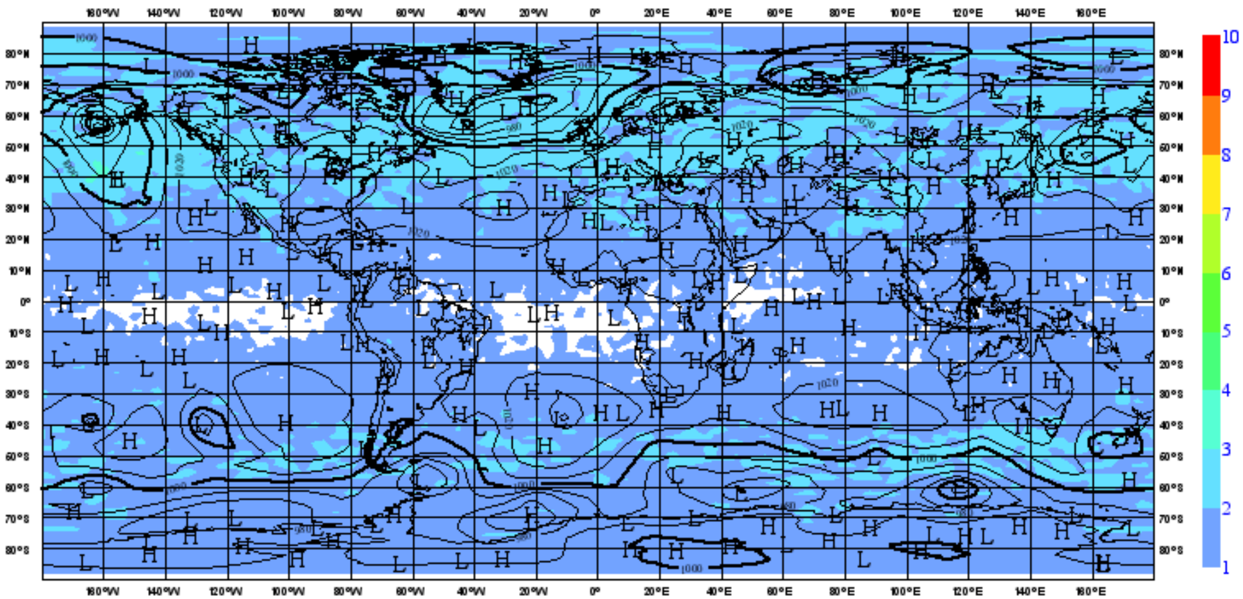
1. $\mathbf{K}$ is the balance operator, i.e. the operator that links the control variables to the model variables
2. $\boldsymbol{\Sigma}_b$ is the grid point variance of background errors
3. $\mathbf{C}_j(\lambda,\phi)$ is the vertical correlation matrix for wavelet index j
4. The wavelet transform is defined by the set of basis functions ψ_j

Hybrid Data Assimilation: Applications

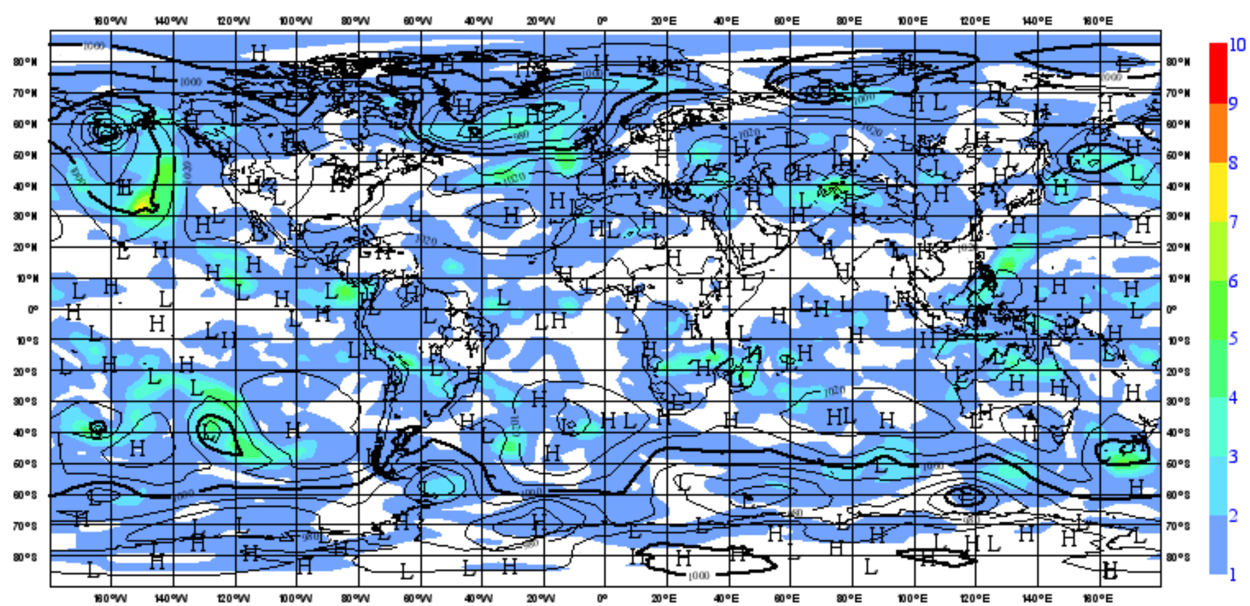
How does the background error model update works?

- In standard 4D-Var the background error variances ($\mathbf{\Sigma}_b$) and the background error correlations ($\mathbf{C}_j(\lambda,\varphi)$) are computed offline from a climatology of EDA background perturbations.
- In Hybrid EDA 4D-Var these quantities (variances and correlations) are updated online using the latest set of EDA background perturbations
- In this way the **B** model is continuously updated and is able to represent the "errors of the day"

Hybrid Data Assimilation: Applications



500 hPa Vorticity errors estimated from **climat. B**

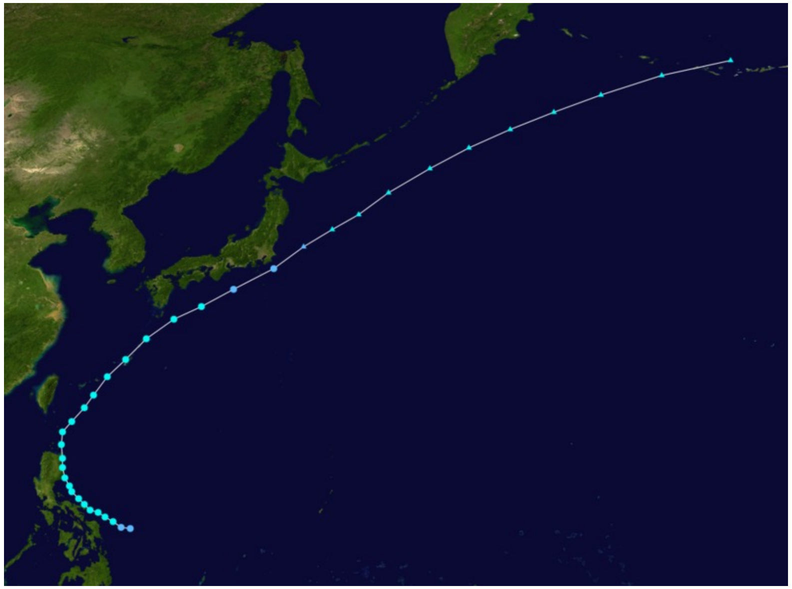


500 hPa Vorticity errors estimated from online **B**

Hybrid Data Assimilation: Applications

- Inside 4D-Var EDA derived background error estimates change the shape and size of analysis increments

Tropical Cyclone Aere, Philippines 8-9 May 2011.

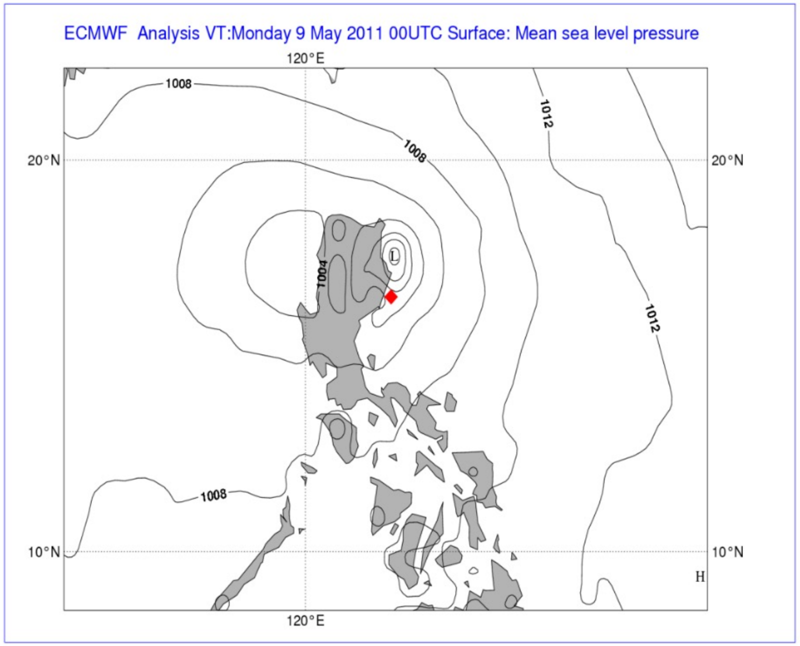


Hybrid Data Assimilation: Applications

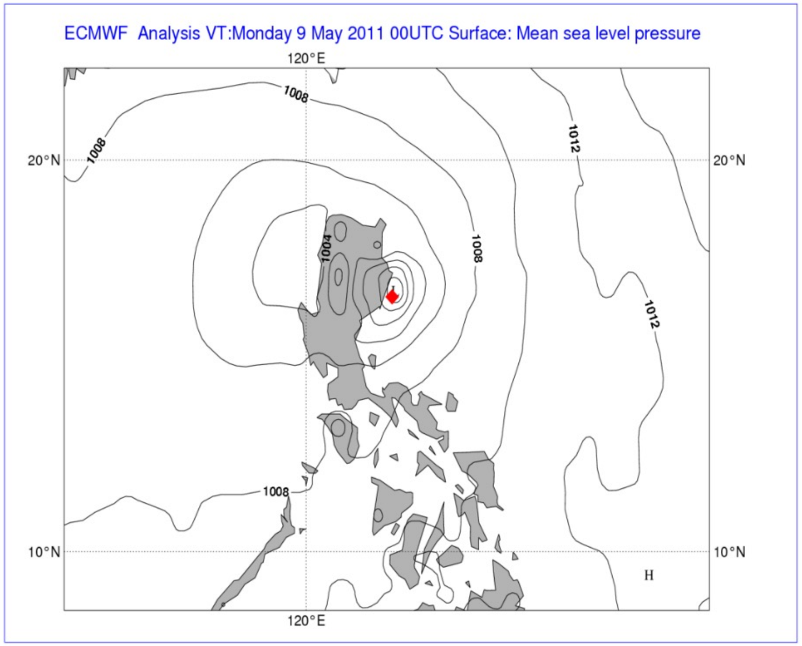
- Inside 4D-Var EDA derived background error estimates change the shape and size of analysis increments

Significant operational analysis error, corrected by 4DVar with EDA variances

4DVar with Static errors



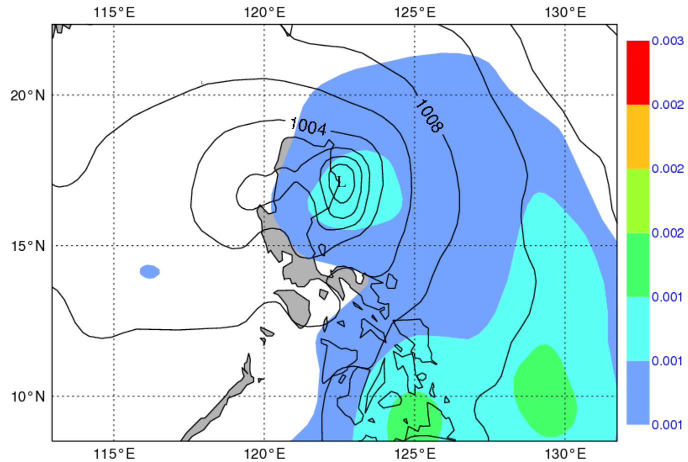
4DVar with EDA errors



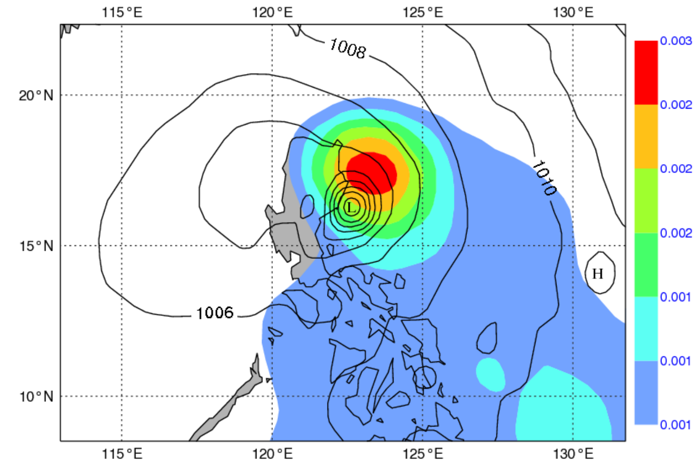
ECMWF EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS

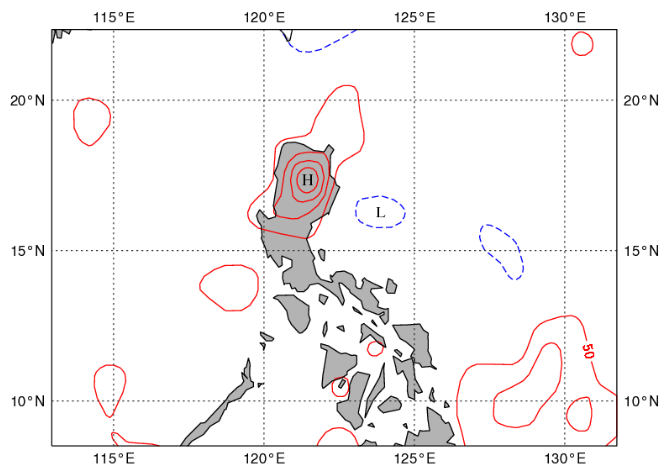
Hybrid Data Assimilation: Applications

Static errors

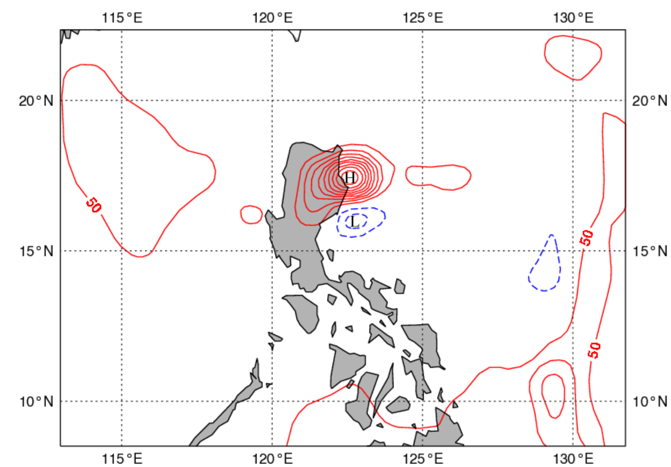


EDA errors





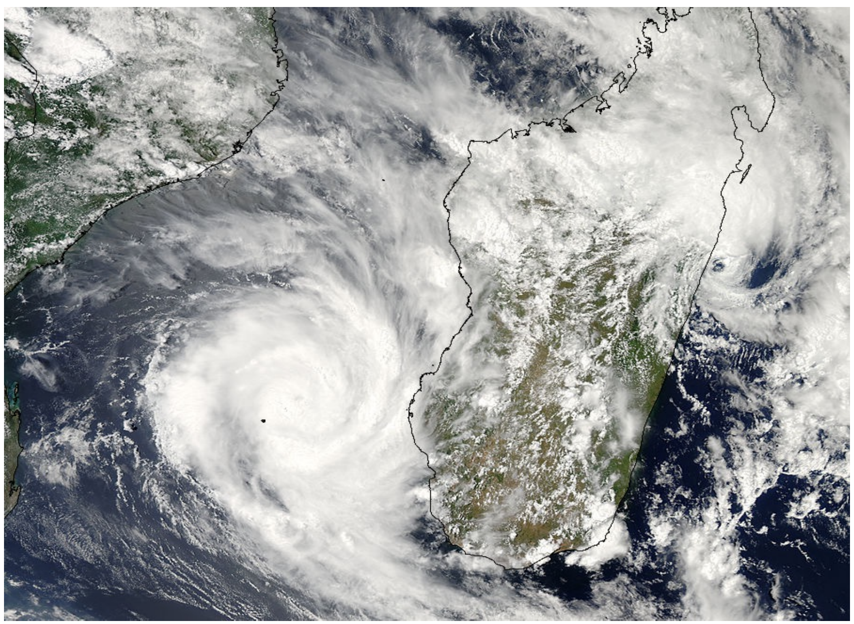
Static SP ana incr.



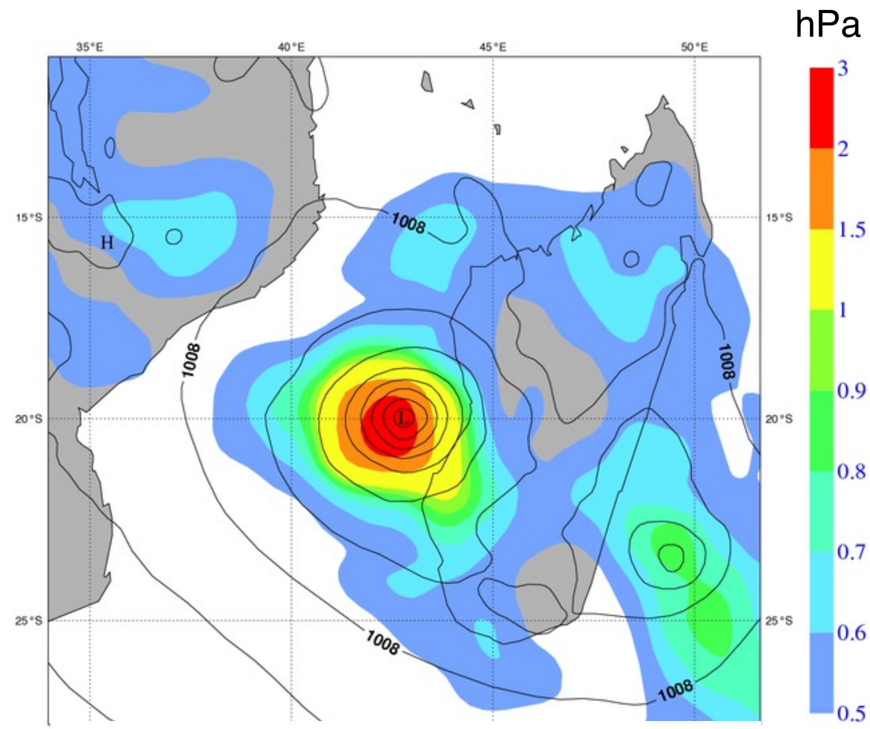
EDA SP ana incr.

Hybrid Data Assimilation: Applications

- The online update of B involves not only the background error variances ($\mathbf{\Sigma}_b$) but also the background error correlations ($\mathbf{C}_j(\lambda,\varphi)$)



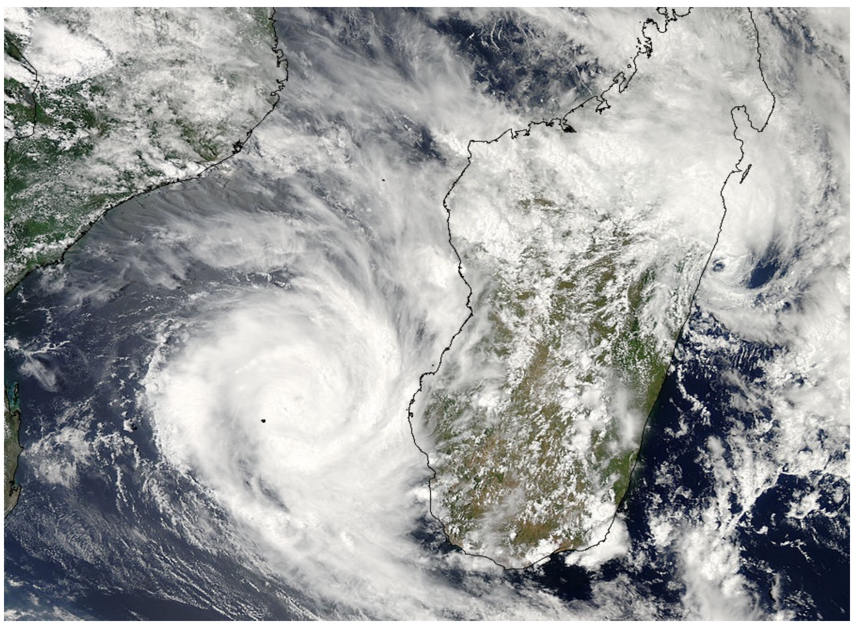
Hurricane Fanele, 20 January 2009



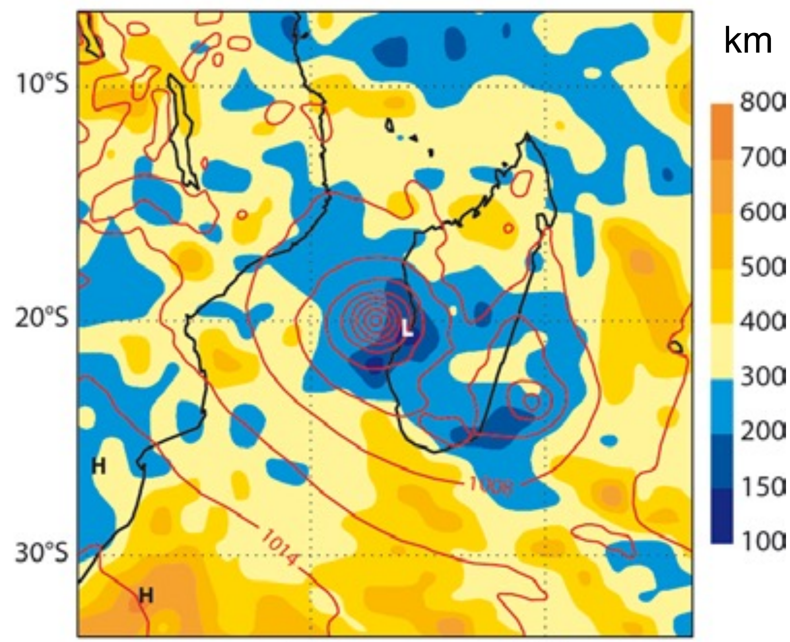
EDA derived background error variance for Surface pressure

Hybrid Data Assimilation: Applications

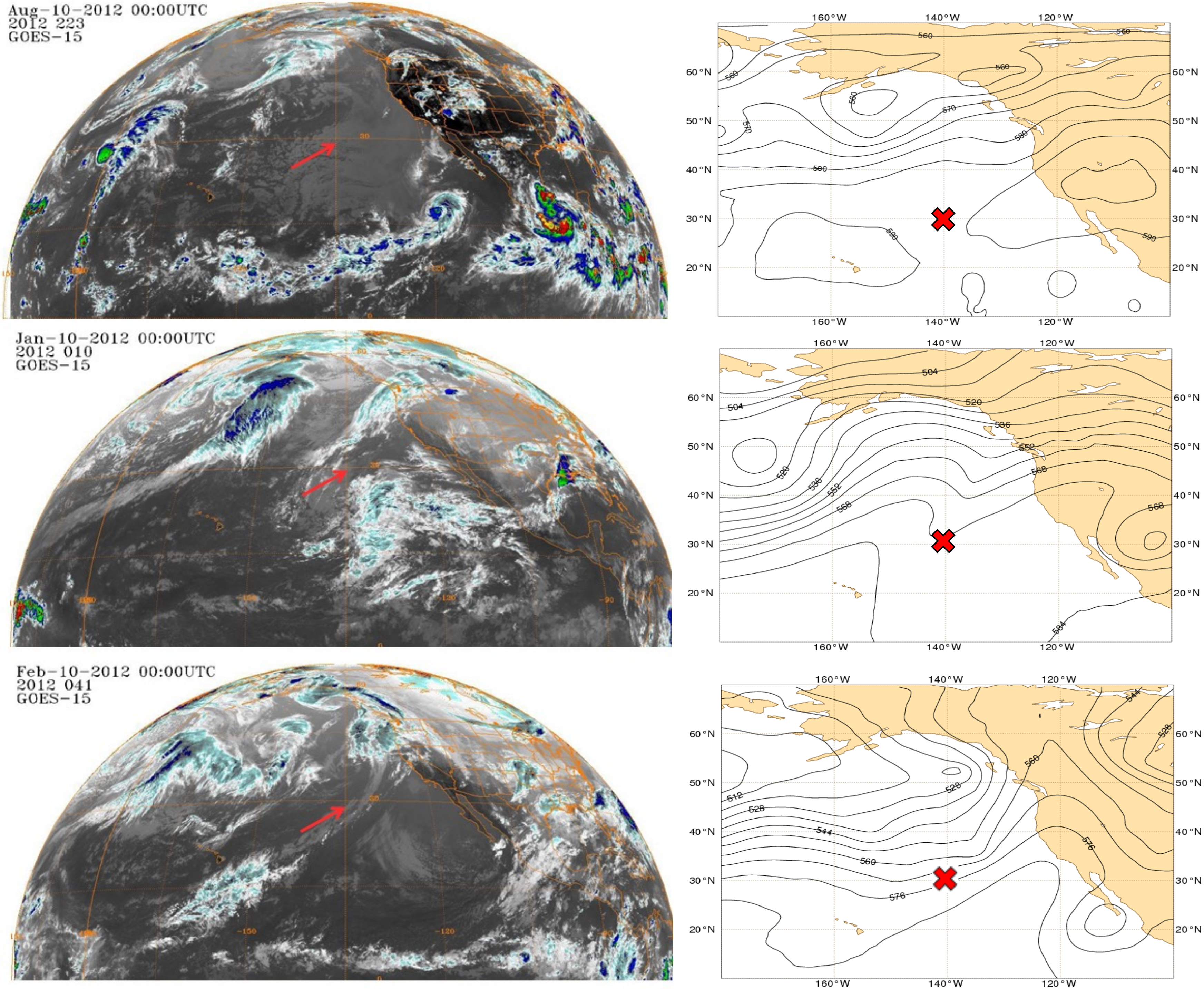
- The online update of B involves not only the background error variances ($\boldsymbol{\Sigma}_b$) but also the background error correlations ($\mathbf{C}_j(\lambda,\varphi)$)



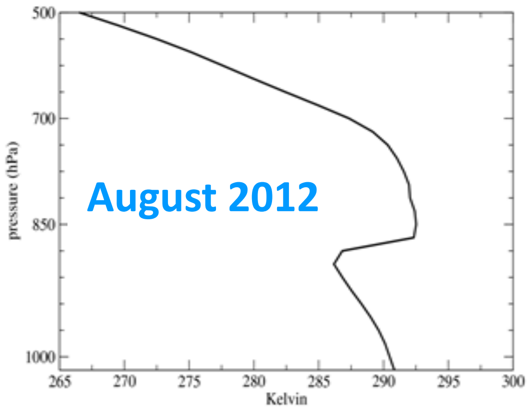
Hurricane Fanele, 20 January 2009

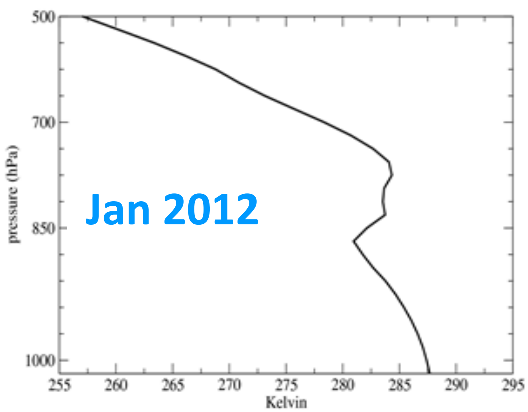


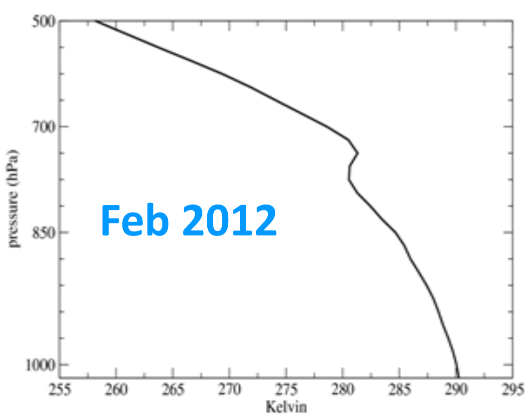
EDA derived background error correlation length scale for surface pressure



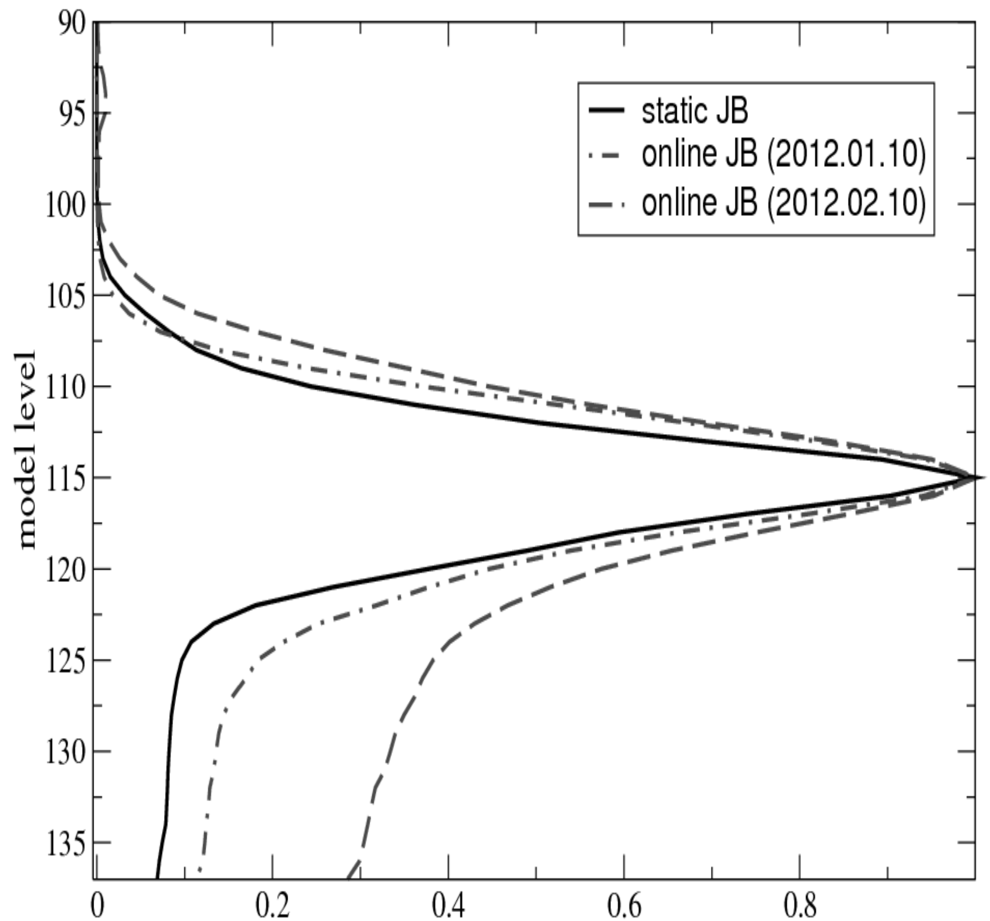
Aug-10-2012 00:00UTC
2012 223
GOES-15
Jan-10-2012 00:00UTC
2012 010
GOES-15
Feb-10-2012 00:00UTC
2012 041
GOES-15
160°W
140°W
120°W
60°N
50°N
40°N
30°N
20°N
560
570
580
590
504
520
536
552
568
584
512
528
544
576







Vertical bg error correlation for Vorticity, ~850hPa



Hybrid Data Assimilation: Hybrid Gain

- Another possible way of hybridizing an EnKF with a variational DA system is to simply perform a linear combination of their respective analyses (Penny, 2014):

$$\mathbf{x}^{\boldsymbol{a}}_{Hyb} = \alpha\overline{\mathbf{x}^{\boldsymbol{a}}_{iEnKF}} + (1-\alpha)\mathbf{x}^{\boldsymbol{a}}_{4DVar}$$

- In a linear framework this is equivalent to doing a linear combination of the Kalman gains of the two analyses (hence the name):

$$\mathbf{x}^{\boldsymbol{a}}_{Hyb} = \overline{\mathbf{x}^{\boldsymbol{b}}_{iHyb}} + (\boldsymbol{\alpha}\mathbf{K}_{EnKF} + (1-\alpha)\mathbf{K}_{4DVar})\left(\boldsymbol{y} - \overline{\mathbf{x}^{\boldsymbol{b}}_{iHyb}}\right)$$

- How does it work? Pretty well, actually, both in a deterministic sense (Bonavita et al., 2015) and in terms of ensemble performance (Houtekamer et al., 2019).
- Open question: How to optimally combine the two systems? See Barbieri De Azevedo et al., 2020 (EnKF posterior error cov.); Chang et al., 2020, (subspace orthogonal to ensemble) for provisional answers

Summary

- The EDA is a variational implementation of the Perturbed Observations (Stochastic) EnKF.
- It is used at ECMWF to estimate the state error covariances in order to a) initialise the ensemble prediction system and b) to provide estimates of the background error covariances for 4D-Var analysis
- Advantages: closer to reference 4D-Var, simpler to maintain and update. Disadvantages: computational cost
- Hybrid DA: 3/4D-Var in combination with EnKF/EDA for error estimation and cycling
- Better results than stand-alone 4D-Var or EnKF
- Various flavours of Hybrid DA possible: a) with direct use of ensemble perturbations (extended control variable, 4D-Ens-Var); b) updating a **B** model (hybrid EDA 4D-Var)
- Common issue: limited affordable ensemble size introduces sampling problems. Different techniques to tackle them (localisations, spatial averaging, time averaging, etc.).
- Estimates of $\mathbf{P}^{a/b}$ only as good as our knowledge of **R**, **Q** => improvements in error modelling improve forecasts!

Bibliography

1. Barbieri De Azevedo, H. et al., (2020) Dynamically weighted hybrid gain data assimilation: perfect model testing, Tellus A: Dynamic Meteorology and Oceanography, 72:1, 1-11, DOI: 10.1080/16000870.2020.1835310
2. Bonavita M., L. Raynaud and L. Isaksen, 2010: Estimating background-error variances with the ECMWF Ensemble of Data Assimilations system: the effect of ensemble size and day-to-day variability. Q. J. R. Meteorol. Soc., 137: 423–434.
3. Bonavita M., L. Isaksen and E. Holm, 2012: On the use of EDA background error variances in the ECMWF 4D-Var. Q. J. R. Meteorol. Soc.. Early Online Release, doi: 10.1002/qj.1899
4. Bonavita M., E. Holm, L. Isaksen and M. Fisher, 2015: The evolution of the ECMWF hybrid data assimilation system, Q.J.R. Meteorol. Soc., 142: 287–303. DOI: 10.1002/qj.2652.
5. Bonavita, M., Hamrud, M., & Isaksen, L. (2015). EnKF and Hybrid Gain Ensemble Data Assimilation. Part II: EnKF and Hybrid Gain Results, Monthly Weather Review, 143(12), 4865-4882
6. Chih-Chien Chang, Stephen G. Penny, Shu-Chih Yang, Hybrid Gain Data Assimilation Using Variational Corrections in the Subspace Orthogonal to the Ensemble, Monthly Weather Review, 10.1175/MWR-D-19-0128.1, 148, 6, (2331-2350), (2020).
7. Fisher, M. Leutbecher M, Kelly GA. 2005 On the equivalence between Kalman smoothing and weak-constraint four-dimensional variational data assimilation. Q. J. R. Meteorol. Soc. 131 3235–3246.
8. Fisher, M., 2003: Background error covariance modelling. Proceedings of the ECMWF Seminar on recent developments in data assimilation for atmosphere and ocean, ECMWF, pages 45–63. (Available from: http://www.ecmwf.int/publications/)

Bibliography

9. Hólm, E, Bonavita, M. and S. Lang: Soft re-centring Ensemble of Data Assimilations, ECMWF Newsletter, Spring 2022, available at https://www.ecmwf.int/en/newsletter/171/news/soft-re-centring-ensemble-data-assimilations
10. Houtekamer, PL, Buehner, M, De La Chevrotière, M. Using the hybrid gain algorithm to sample data assimilation uncertainty. Q J R Meteorol Soc. 2019; 145 (Suppl. 1): 35– 56. https://doi.org/10.1002/qj.3426
11. Isaksen, L., M. Bonavita, R. Buizza, M. Fisher, J. Haseler, M. Leutbecher and Laure Raynaud, 2010: Ensemble of data assimilations at ECMWF. ECMWF Technical Memorandum No. 636. (Available from: http://www.ecmwf.int/publications/)
12. Hunt, B. R., Kostelich, E. J. and Szunyogh, I., 2007. Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter. Physica D, 230, 112–126.
13. Lang STK, Lock S-J, Leutbecher M, Bechtold P, Forbes RM. Revision of the Stochastically Perturbed Parametrisations model uncertainty scheme in the Integrated Forecasting System. *Q J R Meteorol Soc*. 2021; 147: 1364–1381. https://doi.org/10.1002/qj.3978
14. Liu C, Xiao Q, Wang B. 2008. An ensemble-based four-dimensional variational data assimilation scheme. part i: Technical formulation and preliminary test. Mon. Weather Rev. 136: 3363–3373.
15. Lorenc, A.C.,2003: The potential of the ensemble Kalman filter for NWP—A comparison with 4D-VAR. Q. J. R. Meteorol. Soc., 129: 3183–3203.
16. Ott E, Hunt BR, Szunyogh I, Zimin AV, Kostelich EJ, Corazza M, Kalnay E, Patil DJ. 2004. A local ensemble Kalman filter for atmospheric data assimilation. Tellus 56A: 415–428.
17. Penny, G. S., 2014: The hybrid local ensemble transform Kalman filter. Mon. Wea. Rev., 142, 2139–2149, doi:10.1175/MWR-D-13-00131.1

Bibliography

17. Raynaud, L., L. Berre, and G. Desroziers, 2008: Spatial averaging of ensemble-based background-error variances. Q. J. R. Meteorol. Soc., 134, 1003–1014.
18. Thepaut, J.-N., Courtier, P., Belaud, G. and Lemaıtre, G., 1996: Dynamical structure functions in a four-dimensional variational assimilation: A case study. Q. J. R. Meteorol. Soc., 122, 535–561