Kalman Filter Methods

Massimo Bonavita

massimo.bonavita@ecmwf.int

Based on Mike Fisher lecture notes

Outline

- The standard Kalman Filter and its extensions
- Kalman Filters for large dimensional systems
- The Ensemble Kalman Filter
- Hybrid Variational–EnKF algorithms

Standard Kalman Filter

- In a previous lecture it was shown that the linear, unbiased analysis equation had the form:

$$\mathbf{x}^a_k = \mathbf{x}^b_k + \mathbf{K}_k (\mathbf{y}_k - \mathbf{H}_k(\mathbf{x}^b_k))$$

a = analysis; b = background
k = time index (t=0,1,…,k,…)

- It was also shown that the best linear unbiased analysis (a.k.a. Best Linear Unbiased Estimator, BLUE) is achieved when the matrix $\mathbf{K}_k$ (Kalman Gain Matrix) has the form:

$$\mathbf{K}_k = \mathbf{P}^b_k \mathbf{H}^T_k(\mathbf{H}_k \mathbf{P}^b_k \mathbf{H}^T_k + \mathbf{R}_k)^{-1} = ((\mathbf{P}^b_k)^{-1} + \mathbf{H}^T_k \mathbf{R}_k^{-1} \mathbf{H}_k)^{-1} \mathbf{H}^T_k \mathbf{R}_k^{-1}$$

$\mathbf{P}^b$ = covariance matrix of the background error
$\mathbf{R}$ = covariance matrix of the observation error

- Here “best” means the minimum error variance analysis
- An expression for the covariance matrix of the analysis error was also found:

Standard Kalman Filter

- An expression for the covariance matrix of the analysis error was also found:

$$\mathbf{P}^a_k = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}^b_k (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)^T + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T$$

- In most application of data assimilation we want to update our estimate of the state and its uncertainty at later times, as new observations come in: we want to cycle the analysis
- For each analysis in this cycle we require a background $\mathbf{x}^b_k$ (i.e. a prior estimate of the state at time t_k)
- Our best prior estimate of the state at time t_k is given by a forecast from the preceding analysis:

$$\mathbf{x}^b_k = \mathbf{M}_{t_{k-1} \rightarrow t_k}(\mathbf{x}^a_{k-1})$$

- What is the error covariance matrix associated with this background?

Standard Kalman Filter

- What is the error covariance matrix associated with this background?

$$\mathbf{x}^b_k = \mathbf{M}_{t_{k-1}\rightarrow t_k}(\mathbf{x}^a_{k-1})$$

- Subtract the true state $\mathbf{x}^*_k$ from both sides of the equation:

$$\boldsymbol{\varepsilon}^b_k = \mathbf{M}_{t_{k-1}\rightarrow t_k}(\mathbf{x}^a_{k-1}) - \mathbf{x}^*_k$$

- Since $\mathbf{x}^a_{k-1} = \mathbf{x}^*_{k-1} + \boldsymbol{\varepsilon}^a_{k-1}$ we have:

$$\boldsymbol{\varepsilon}^b_k = \mathbf{M}_{t_{k-1}\rightarrow t_k}(\mathbf{x}^*_{k-1} + \boldsymbol{\varepsilon}^a_{k-1}) - \mathbf{x}^*_k =$$

$$\mathbf{M}_{t_{k-1}\rightarrow t_k}(\mathbf{x}^*_{k-1}) + \mathbf{M}_{t_{k-1}\rightarrow t_k}\boldsymbol{\varepsilon}^a_{k-1} - \mathbf{x}^*_k =$$

$$\mathbf{M}_{t_{k-1}\rightarrow t_k}\boldsymbol{\varepsilon}^a_{k-1} + \eta_k$$

- Where we have defined the model error $\eta_k = \mathbf{M}_{t_{k-1}\rightarrow t_k}(\mathbf{x}^*_{k-1}) - \mathbf{x}^*_k$
- We will also assume that $< \boldsymbol{\varepsilon}^a_{k-1} > = < \eta_k > = 0 \Rightarrow < \boldsymbol{\varepsilon}^b_k >$
- The background error covariance matrix will then be given by:

Standard Kalman Filter

$$<\boldsymbol{\varepsilon}^b_k (\boldsymbol{\varepsilon}^b_k)^T> = \mathbf{P}^b_k = <(\mathbf{M}_{t_{k-1}\rightarrow t_k}\boldsymbol{\varepsilon}^a_{k-1} + \eta_k)(\mathbf{M}_{t_{k-1}\rightarrow t_k}\boldsymbol{\varepsilon}^a_{k-1} + \eta_k)^T> =$$

$$\mathbf{M}_{t_{k-1}\rightarrow t_k}<\boldsymbol{\varepsilon}^a_{k-1} (\boldsymbol{\varepsilon}^a_{k-1})^T> (\mathbf{M}_{t_{k-1}\rightarrow t_k})^T + <\eta_k (\eta_k)^T> =$$

$$\mathbf{M}_{t_{k-1}\rightarrow t_k} \mathbf{P}^a_{k-1} (\mathbf{M}_{t_{k-1}\rightarrow t_k})^T + \mathbf{Q}_k$$

- Here we have assumed $< \boldsymbol{\varepsilon}^a_{k-1} (\eta_k)^T> = 0$ and defined the model error covariance matrix $\mathbf{Q}_k = <\eta_k (\eta_k)^T>$
- We now have all the equations necessary to propagate and update the state and its error estimates:

$$\mathbf{x}^b_k = \mathbf{M}_{t_{k-1}\rightarrow t_k}(\mathbf{x}^a_{k-1})$$

$$\mathbf{P}^b_k = \mathbf{M}_{t_{k-1}\rightarrow t_k} \mathbf{P}^a_{k-1} (\mathbf{M}_{t_{k-1}\rightarrow t_k})^T + \mathbf{Q}_k$$

$$\mathbf{x}^a_k = \mathbf{x}^b_k + \mathbf{K}_k (\mathbf{y}_k - \mathbf{H}_k(\mathbf{x}^b_k))$$

$$\mathbf{P}^a_k = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}^b_k (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)^T + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T$$

$$\mathbf{K}_k = \mathbf{P}^b_k \mathbf{H}^T_k(\mathbf{H}_k \mathbf{P}^b_k \mathbf{H}^T_k + \mathbf{R}_k)^{-1}$$

Standard Kalman Filter

$$\mathbf{x}^b_k = \mathbf{M}_{t_{k-1}\rightarrow t_k}(\mathbf{x}^a_{k-1})$$

$$\mathbf{P}^b_k = \mathbf{M}_{t_{k-1}\rightarrow t_k}\,\mathbf{P}^a_{k-1}\,(\mathbf{M}_{t_{k-1}\rightarrow t_k})^T + \mathbf{Q}_k$$

$$\mathbf{x}^a_k = \mathbf{x}^b_k + \mathbf{K}_k\,(\mathbf{y}_k - \mathbf{H}_k(\mathbf{x}^b_k))$$

$$\mathbf{P}^a_k = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}^b_k\,(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)^T + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T$$

$$\mathbf{K}_k = \mathbf{P}^b_k\,\mathbf{H}^T_k(\mathbf{H}_k\,\mathbf{P}^b_k\,\mathbf{H}^T_k + \mathbf{R}_k)^{-1}$$

- Under the assumption that the model $\mathbf{M}_{t_{k-1}\rightarrow t_k}$ and the observation operator $\mathbf{H}_k$ are linear, the Kalman Filter produces an optimal sequence of analysis
- The analysis $\mathbf{x}^a_k$ is the best (minimum variance) estimate of the state at time t_k , given $\mathbf{x}^b_0$ and all observations up to time t_k ($\mathbf{y}_0,\mathbf{y}_1,\ldots,\mathbf{y}_k$).
- Note that Gaussianity of errors is not required. If errors are Gaussian the KF provides the exact conditional probability estimate, i.e. $p(\mathbf{x}^a_k \mid \mathbf{x}^b_0; \mathbf{y}_0,\mathbf{y}_1,\ldots,\mathbf{y}_k)$

Standard Kalman Filter

- If model and/or observation operators are "slightly" nonlinear a modified version of the KF can be used: the Extended Kalman Filter
- The state update and prediction steps use the nonlinear operators:

$$\mathbf{x}^b_k = \mathcal{M}_{t_{k-1}\rightarrow t_k}(\mathbf{x}^a_{k-1})$$

$$\mathbf{x}^a_k = \mathbf{x}^b_k + \mathbf{K}_k (\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}^b_k))$$

- The covariance update and prediction steps use the Jacobians of the model and observation operators, linearized around the analysed/predicted state, i.e.:

$$\mathbf{M}_{t_{k-1}\rightarrow t_k} = \frac{\partial \mathcal{M}}{\partial \boldsymbol{x}}(\mathbf{x}^a_{k-1})$$

$$\mathbf{H}_k = \frac{\partial \mathcal{H}}{\partial \boldsymbol{x}}(\mathbf{x}^b_k)$$

- The EKF is thus a first order linearization of the KF equations around the current state estimates. As such it is as good as the linearization is a good approximation of the full nonlinear system.

ECMWF

Kalman Filters for Large Dimensional Systems

- The Kalman Filter is impractical for large dimensional systems
- Assuming our state is $O(10^8)$ (which is the order of magnitude of the analysis state in ECMWF 4DVar) the KF requires us to store and evolve in time state covariance matrices ($\mathbf{P}^{a/b}$) of $O(N x N)$
  - The World's fastest computers can sustain ~ 10^{15} operations per second
  - An efficient implementation of matrix multiplication of two $10^8 X 10^8$ matrices requires ~10^{22} operations: about 4 months on the fastest computer!
  - Evaluating $\mathbf{P}^b_k = \mathbf{M}_{t_{k-1}\rightarrow t_k} \mathbf{P}^a_{k-1} (\mathbf{M}_{t_{k-1}\rightarrow t_k})^T + \mathbf{Q}_k$ requires N~10^8 model integrations.
- A range of approximate Kalman Filters has been developed for use with large systems.
- All of these methods rely on a low-rank approximation of the covariance matrices of background and analysis error.

Kalman Filters for Large Dimensional Systems

- Assume (big assumption!!) that $\mathbf{P}^b_k$ has rank M<<N (e.g. M 100).
- Then we can write $\mathbf{P}^b_k = \mathbf{X}^b_k(\mathbf{X}^b_k)^T$, where $\mathbf{X}^b_k$ is N x M.
- The Kalman Gain then becomes:

$$\mathbf{K}_k = \mathbf{P}^b_k \mathbf{H}^T_k(\mathbf{H}_k \mathbf{P}^b_k \mathbf{H}^T_k + \mathbf{R}_k)^{-1} =$$

$$\mathbf{X}^b_k(\mathbf{X}^b_k)^T\mathbf{H}^T_k(\mathbf{H}_k \mathbf{X}^b_k(\mathbf{X}^b_k)^T \mathbf{H}^T_k + \mathbf{R}_k)^{-1} =$$

$$\mathbf{X}^b_k (\mathbf{H}_k\mathbf{X}^b_k)^T(\mathbf{H}_k \mathbf{X}^b_k(\mathbf{H}_k\mathbf{X}^b_k)^T + \mathbf{R}_k)^{-1}$$

- Note that, to evaluate $\mathbf{K}$, we apply $\mathbf{H}_k$ to the M columns of $\mathbf{X}^b_k$ rather than to the N columns of $\mathbf{P}^b_k$.
- The N x N matrices $\mathbf{P}^{a/b}_k$ have been eliminated from the computation! In their place we have N x M ($\mathbf{X}^b_k$) and L x M ($\mathbf{H}_k\mathbf{X}^b_k$) matrices ($L$ = number of observations)

Kalman Filters for Large Dimensional Systems

- The analysis error covariance matrix becomes:

$$\mathbf{P}^a_k = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}^b_k (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)^T + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T =$$

$$= (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}^b_k = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\, \mathbf{X}^b_k(\mathbf{X}^b_k)^T =$$

$$\mathbf{X}^b_k(\mathbf{X}^b_k)^T - \mathbf{K}_k\mathbf{H}_k\, \mathbf{X}^b_k(\mathbf{X}^b_k)^T$$

- Both terms in this expression for $\mathbf{P}^a_k$ contain an initial $\mathbf{X}^b_k$ and a final $(\mathbf{X}^b_k)^T$ so that $\mathbf{P}^a_k = \mathbf{X}^b_k\mathbf{W}_k(\mathbf{X}^b_k)^T$ for some M x M matrix $\mathbf{W}_k$
- Finally the covariance matrix is propagated by:

$$\mathbf{P}^b_k = \mathbf{M}_{t_{k-1}\rightarrow t_k}\, \mathbf{P}^a_{k-1}\, (\mathbf{M}_{t_{k-1}\rightarrow t_k})^T + \mathbf{Q}_k =$$

$$\mathbf{M}_{t_{k-1}\rightarrow t_k}\mathbf{X}^b_k\mathbf{W}_k(\mathbf{X}^b_k)^T(\mathbf{M}_{t_{k-1}\rightarrow t_k})^T + \mathbf{Q}_k =$$

$$\mathbf{M}_{t_{k-1}\rightarrow t_k}\mathbf{X}^b_k\mathbf{W}_k\, (\mathbf{M}_{t_{k-1}\rightarrow t_k}\mathbf{X}^b_k)^T + \mathbf{Q}_k$$

- This requires only M integrations of the linearized model $\mathbf{M}_{t_{k-1}\rightarrow t_k}$
- $\mathbf{Q}_k$ can be approximated by a suitable projection on M-dim subspace

ECMWF

Kalman Filters for Large Dimensional Systems

- The algorithm described above is called Reduced-rank Kalman Filter
- All these gains in computational efficiency have a price, however ☹
- The analysis increment is a linear combination of the columns of $\mathbf{X}^b_k$:

$$\mathbf{x}^a_k - \mathbf{x}^b_k = \mathbf{K}_k (\mathbf{y}_k - \mathcal{H}_k(\mathbf{x}^b_k)) = \mathbf{X}^b_k (\mathbf{H}_k\mathbf{X}^b_k)^T ((\mathbf{H}_k\mathbf{X}^b_k)(\mathbf{H}_k\mathbf{X}^b_k)^T + \mathbf{R})^{-1} (\mathbf{y}_k - \mathcal{H}_k(x_b))$$

- Thus the increments are confined to the subspace spanned by $\mathbf{X}^b_k$, which has at most rank $M << N$.
- The severe reduction in rank manifests itself in two forms:
    1. There are too few degrees of freedom available to fit the ~10^7 observations: the analysis is too "smooth";
    2. The low-rank approximations of the covariance matrices suffer from spurious long-distance correlations. These cause spurious increments in regions where there are no observations.

Kalman Filters for Large Dimensional Systems

- There are two ways around the rank deficiency problem:
    1. Domain localization (e.g. Evensen 2003; Ott *et al.* 2004);
- Domain localization solves the analysis equations independently for each gridpoint, or for each of a set of regions covering the domain.
- Each analysis uses only observations that are local to the gridpoint (or region) and the observations are usually weighted according to their distance from the analysed gridpoint (e.g., Hunt *et al.*, 2007)
- This guarantees that the analysis at each gridpoint (or region) is not influenced by distant observations.
- In effect, the method acts to vastly increase the dimension of the sub-space in which the analysis increment is constructed.
- However, performing independent analyses for each region is not optimal, e.g. poor analysis of the large scales, and difficulties in producing balanced analyses.

Kalman Filters for Large Dimensional Systems

- There are two ways around the rank deficiency problem:

  2. Covariance localization (e.g. Houtekamer and Mitchell 2001).

- Covariance localization is performed by element wise (Schur) multiplication of the error covariance matrices with a predefined covariance matrix representing a decaying function of distance.
- In this way spurious long range correlations in $\mathbf{P}^{a/b}_k$ are suppressed.
- As for domain localization, the method acts to vastly increase the dimension of the sub-space in which the analysis increment is constructed.
- Choosing the product function is non-trivial. It is easy to modify $\mathbf{P}^{a/b}_k$ in undesirable ways. In particular, balance relationships may be adversely affected.

Ensemble Kalman Filters

- Ensemble Kalman Filters (EnKF, Evensen, 1994; Burgers et al., 1998) are Monte Carlo implementations of the Reduced-rank KF
- In EnKF error covariances are constructed as sample covariances from an ensemble of background/analysis fields, i.e.:

$$\mathbf{P}^{a/b}_{k} = \frac{1}{M-1}\sum_{m=1,M\text{-}1}(\mathbf{x}^{b}_{k,m} - \langle\mathbf{x}^{b}_{k,m}\rangle)(\mathbf{x}^{b}_{k,m} - \langle\mathbf{x}^{b}_{k,m}\rangle)^{T} =$$

$$= \mathbf{X}^{b}_{k}(\mathbf{X}^{b}_{k})^{T}$$

- $\mathbf{X}^{b}_{k}$ is the *N* x *M* matrix of background perturbations, i.e.:

$$\mathbf{X}^{b}_{k} = \frac{1}{\sqrt{M-1}}((\mathbf{x}^{b}_{k,1} - \langle\mathbf{x}^{b}_{k,m}\rangle), (\mathbf{x}^{b}_{k,2} - \langle\mathbf{x}^{b}_{k,m}\rangle), .., (\mathbf{x}^{b}_{k,M} - \langle\mathbf{x}^{b}_{k,m}\rangle))$$

- Note that the full covariance matrix is never formed explicitly: The error covariances are usually computed locally for each gridpoint in the *M* x *M* ensemble space

Ensemble Kalman Filters

- In the (extended) KF the error covariances are explicitly propagated using the tangent linear and adjoint of the model and observation operators, i.e.:

$$\mathbf{K}_k = \mathbf{P}^b{}_k\, \mathbf{H}^T{}_k(\mathbf{H}_k\, \mathbf{P}^b{}_k\, \mathbf{H}^T{}_k + \mathbf{R}_k)^{-1}$$

$$\mathbf{P}^b{}_k = \mathbf{M}_{t_{k-1}\rightarrow t_k}\, \mathbf{P}^a{}_{k-1}\, (\mathbf{M}_{t_{k-1}\rightarrow t_k})^T + \mathbf{Q}_k$$

- In the EnKF the error covariances are implicitly propagated in time through the ensemble forecasts and the observation operators linearizations are computed as:

$$\mathbf{P}^b{}_k\mathbf{H}_k{}^T = \mathbf{X}^b{}_k(\mathbf{X}^b{}_k)^T\, \mathbf{H}_k{}^T = \mathbf{X}^b{}_k(\mathbf{H}_k\, \mathbf{X}^b{}_k)^T =$$

$$\frac{1}{M-1}\Sigma_{m=1,M-1}(\mathbf{x}^b{}_{k,m} - <\mathbf{x}^b{}_{k,m}>)\ (\mathbf{x}^b{}_{k,m} - <\mathcal{H}(\mathbf{x}^b{}_{k,m})>)^T$$

$$\mathbf{H}_k\mathbf{P}^b{}_k\mathbf{H}_k{}^T = \mathbf{H}_k\, \mathbf{X}^b{}_k(\mathbf{H}_k\, \mathbf{X}^b{}_k)^T =$$

$$\frac{1}{M-1}\Sigma_{m=1,M-1}(\mathbf{x}^b{}_{k,m} - <\mathcal{H}(\mathbf{x}^b{}_{k,m})>)\ (\mathbf{x}^b{}_{k,m} - <\mathcal{H}(\mathbf{x}^b{}_{k,m})>)^T$$

- Not having to code TL and ADJ operators is a major advantage!

Ensemble Kalman Filters

- The Ensemble Kalman Filter requires us to generate a sample $\{\mathbf{x}^b_{k,m}; m=1,..,M\}$ drawn from the p.d.f. of background error: how to do this?
- We can generate this from a sample $\{\mathbf{x}^a_{k-1,m}; m=1,..,M\}$ from the p.d.f. of analysis error for the previous cycle:

$$\mathbf{x}^b_{k,m} = \mathcal{M}_{t_{k-1}\rightarrow t_k}(\mathbf{x}^a_{k-1,m}) + \boldsymbol{\eta}_{k,m}$$

where $\boldsymbol{\eta}_{k,m}$ is a sample drawn from the p.d.f. of model error.

- The question is then: How do we generate a sample from the analysis p.d.f.? Let us look at the analysis update again:

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}\,(\mathbf{y} - \mathbf{H}(\mathbf{x}^b)) = (\mathbf{I}\text{-}\mathbf{KH})\,\mathbf{x}^b + \mathbf{Ky}$$

- If we subtract the true state $\mathbf{x}^*$ from both sides (and assume $\mathbf{y}^*=\mathbf{Hx}^*$)

$$\mathbf{e}^a = (\mathbf{I}\text{-}\mathbf{KH})\,\mathbf{e}^b + \mathbf{Ke}^o$$

- i.e., the errors have the same update as the state; note that this holds also for suboptimal **K**

Ensemble Kalman Filters

- Consider now an ensemble of analysis where all the inputs to the analysis have been perturbed according to their error p.d.f.:

$$\mathbf{x}^{a\prime} = (\mathbf{I}-\mathbf{KH})\,\mathbf{x}^{b\prime} + \mathbf{Ky}'$$

- If we subtract the unperturbed analysis $\mathbf{x}^a = (\mathbf{I}-\mathbf{KH})\,\mathbf{x}^b + \mathbf{Ky}$

$$\boldsymbol{\varepsilon}^a = (\mathbf{I}-\mathbf{KH})\,\boldsymbol{\varepsilon}^b + \mathbf{K}\boldsymbol{\varepsilon}^o$$

- Note that the observations (during the update step) and the model (during the forecast step) are perturbed explicitly.
- The background is implicitly perturbed , i.e.:

$$\mathbf{x}^b{}_{k,m} = \mathcal{M}_{t_{k-1}\rightarrow t_k}(\mathbf{x}^a{}_{k-1,m}) + \boldsymbol{\eta}_{k,m}$$

- Hence, one way to generate a sample drawn from the p.d.f. of analysis error is to perturb the observations with perturbations characteristic of observation error.
- The EnKF based on this idea is called Perturbed Observations EnKF (Houtekamer and Mitchell, 1998). It is also the basis of ECMWF EDA

Ensemble Kalman Filters

- Another way to construct the analysis sample without perturbing the observations is to make a linear combination of the background sample:

$$\mathbf{X}^a_k=\mathbf{X}^b_k\mathbf{T}$$

where T is a *M* x *M* matrix chosen such that:

$$\mathbf{X}^a_k(\mathbf{X}^a_k)^T = (\mathbf{X}^b_k\mathbf{T})\,(\mathbf{X}^b_k\mathbf{T})^T = \mathbf{P}^a_k = (\mathbf{I}-\mathbf{K}_k\mathbf{H}_k)\mathbf{P}^b_k$$

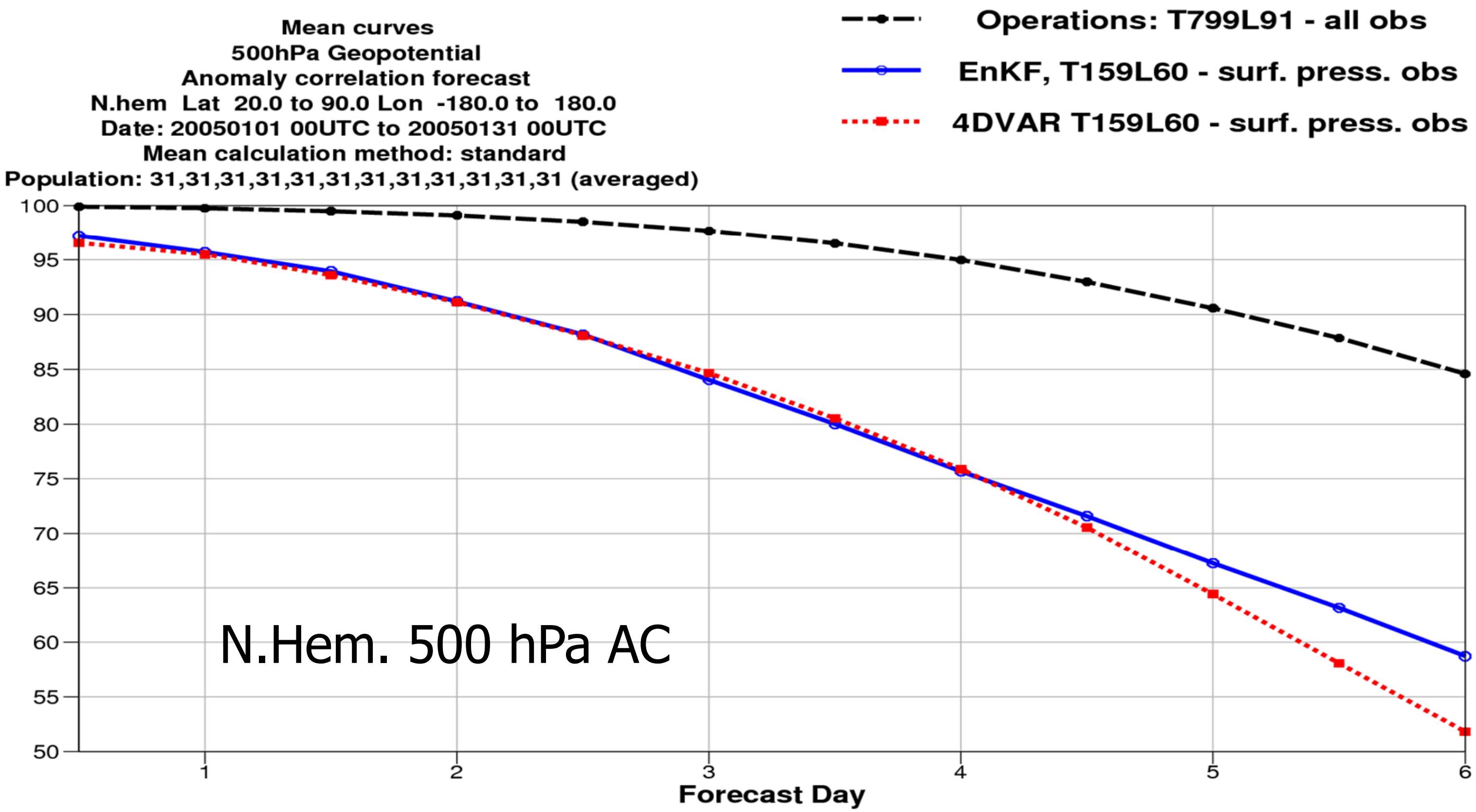
- Note that the choice of **T** is not unique: Any orthonormal transformation **Q** ($\mathbf{QQ}^T=\mathbf{Q}^T\mathbf{Q}=\mathbf{I}$) can be applied to **T** and give a valid analysis sample
- Implementations also differ on the treatment of observations (i.e., local patches, one at a time)
- Consequently there are a number of different, functionally equivalent, implementations of the Deterministic EnKF (ETKF, Bishop *et al.*, 2001; LETKF, Ott *et al.*, 2004, Hunt *et al.*, 2007; EnSRF, Whitaker and Hamill, 2002; EnAF, Anderson, 2001;…)

Ensemble Kalman Filters

- How does the EnKF compare with standard 4DVar?
- The short answer: It depends!

EnKF vs 4DVar

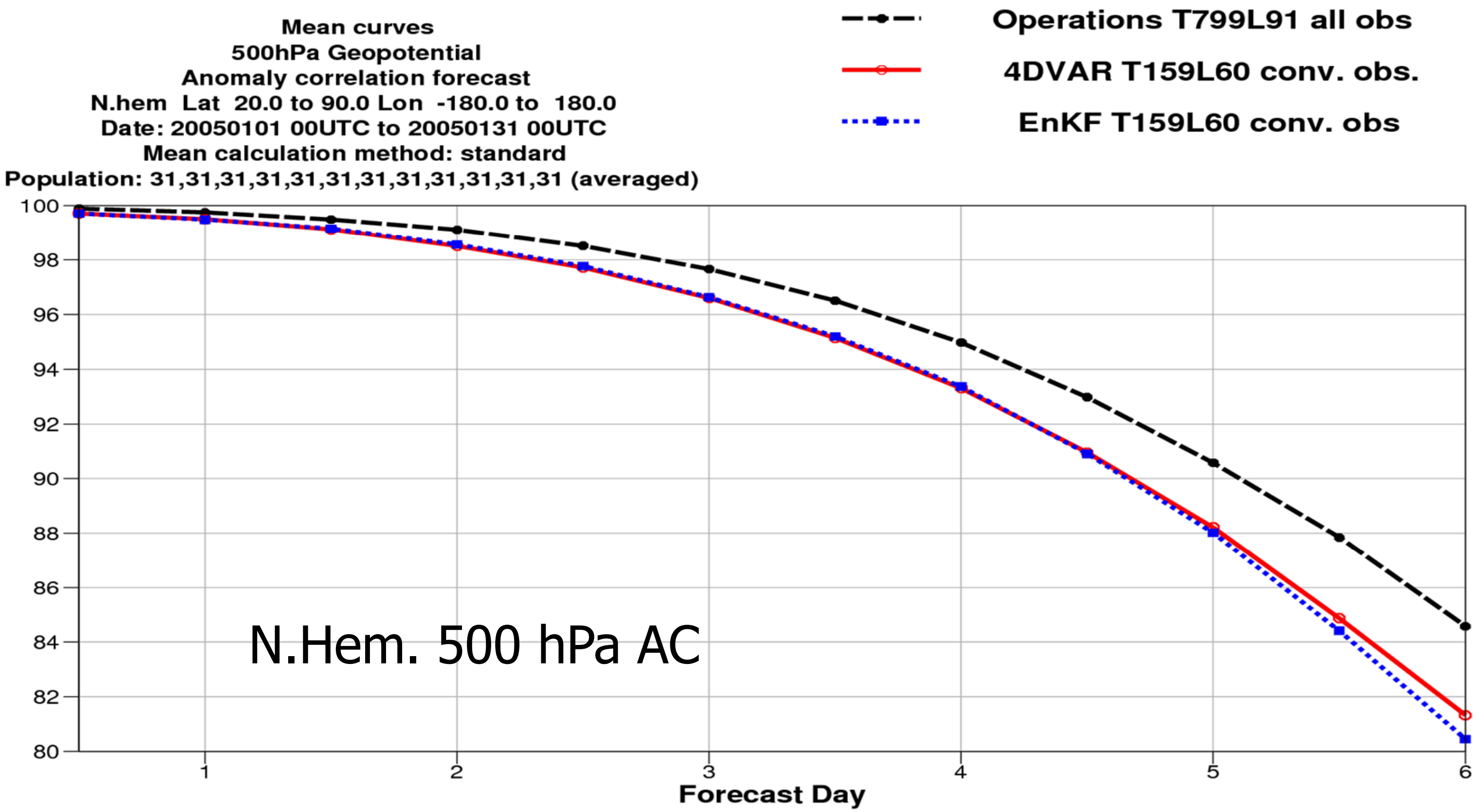
Surface Pressure observations only



ECMWF

EnKF vs 4DVar

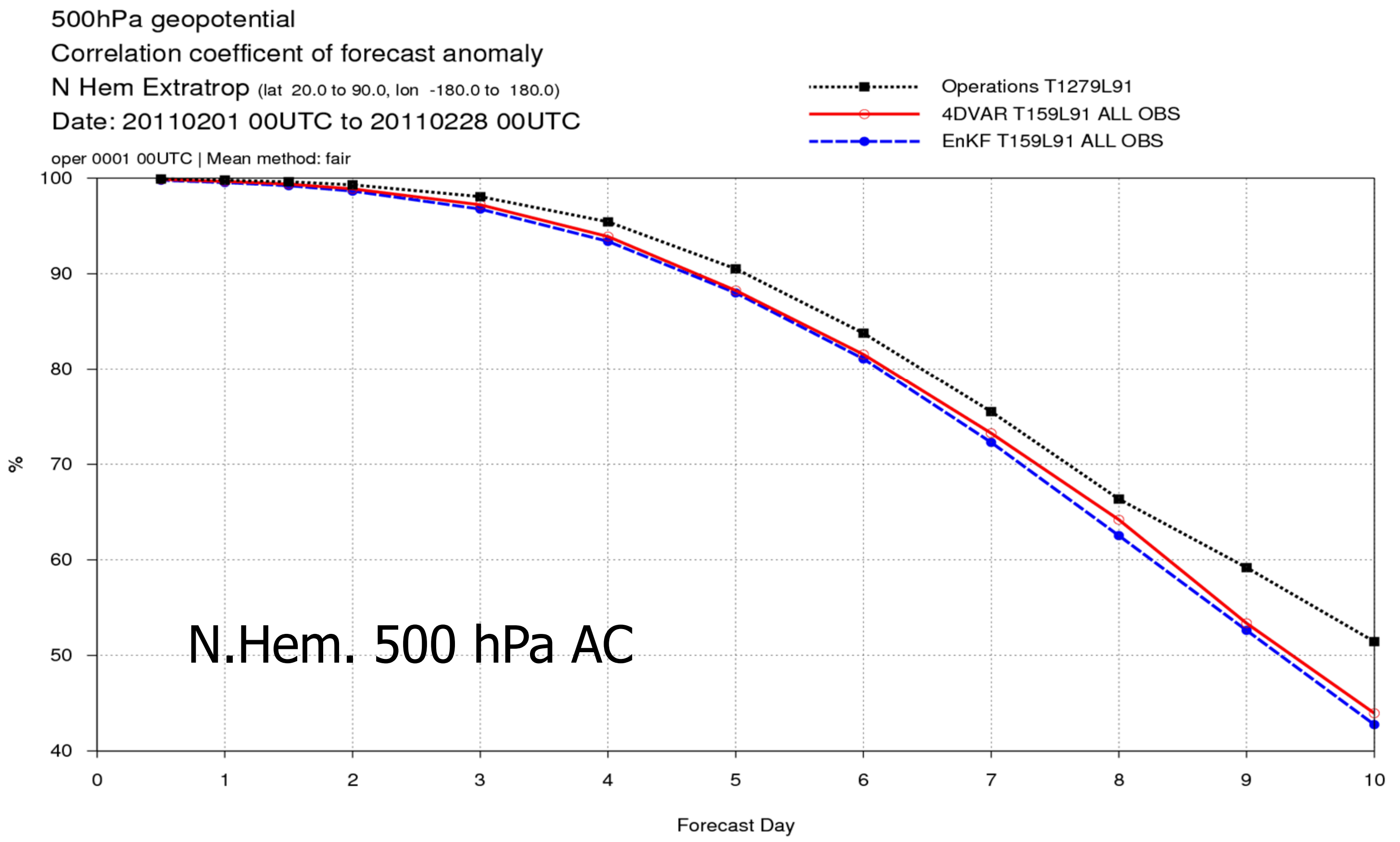
Conventional observations only



ECMWF

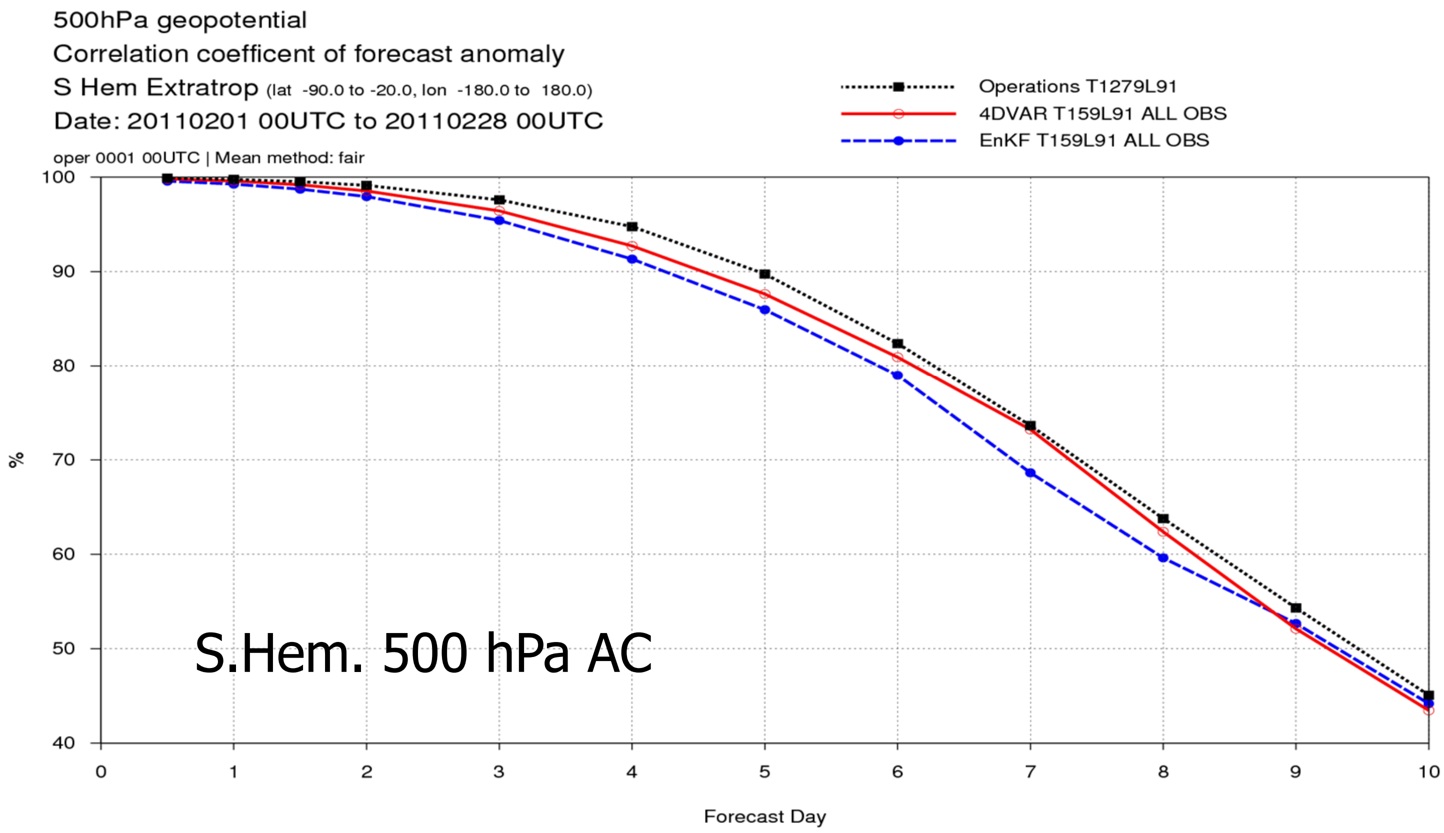
EnKF vs 4DVar

All observations



EnKF vs 4DVar

All observations



Ensemble Kalman Filters

- The rank deficiency of the sampled error covariances is not an issue when the observations are few, i.e. of the order of ensemble size
- The rank deficiency of the sampled error covariances becomes problematic when the observations are orders of magnitude more than the ensemble size
- In this latter case, careful localization of sampled covariances becomes crucial: This is an on-going research topic for EnKF
- Note how covariance localization becomes conceptually and practically more difficult for observations (satellite radiances) which are non-local by nature (Campbell *et al.*, 2010)

Hybrid Variational–EnKF algorithms

4D Variational methods

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 cost function:

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

This is equivalent, for the same $\mathbf{x}_b$, $\mathbf{P}^b$, to the Kalman filter solution at the end of the assimilation window (t=T) (*Fisher et al.*, 2005).

ECMWF

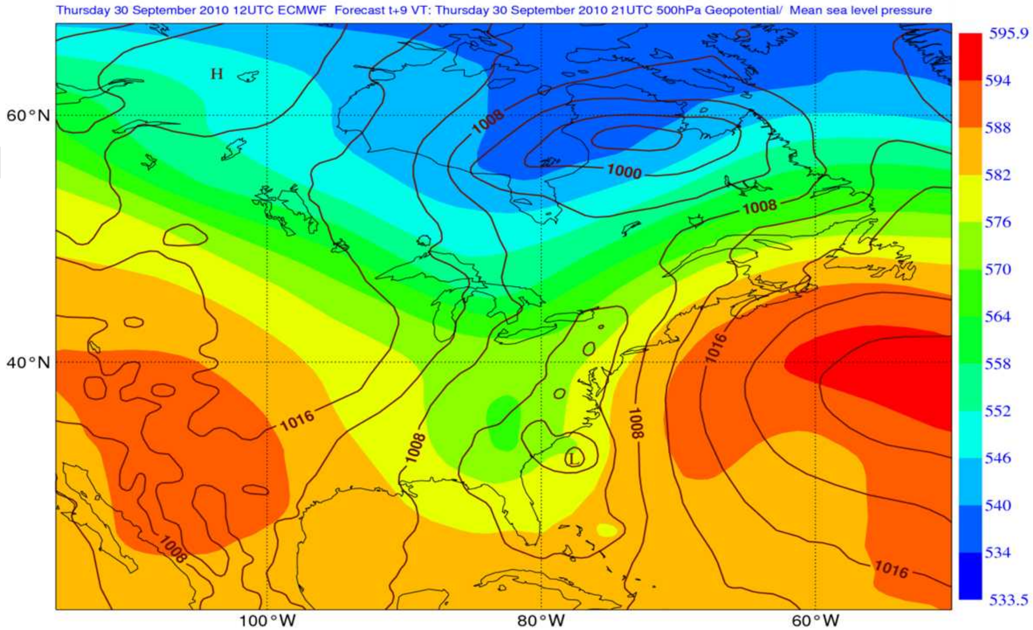
Hybrid Variational–EnKF algorithms

4D Variational methods

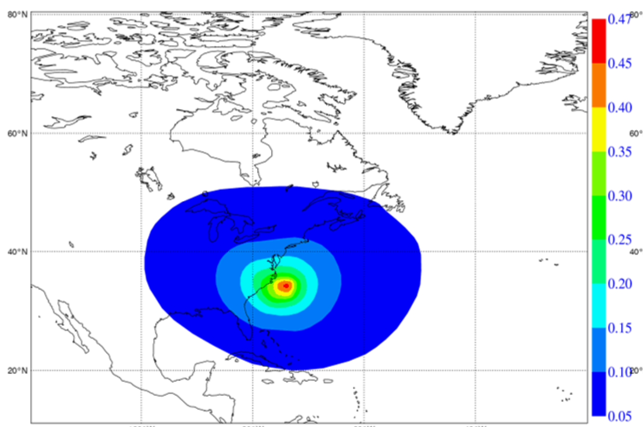
The 4D-Var solution implicitly evolves background error covariances *over the assimilation window* (Thepaut *et al*.,1996) with the tangent linear dynamics:

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

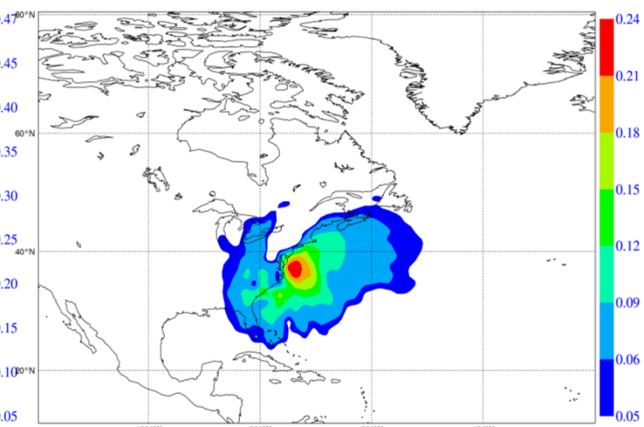
MSLP and 500 hPa Z (shaded) background fcst



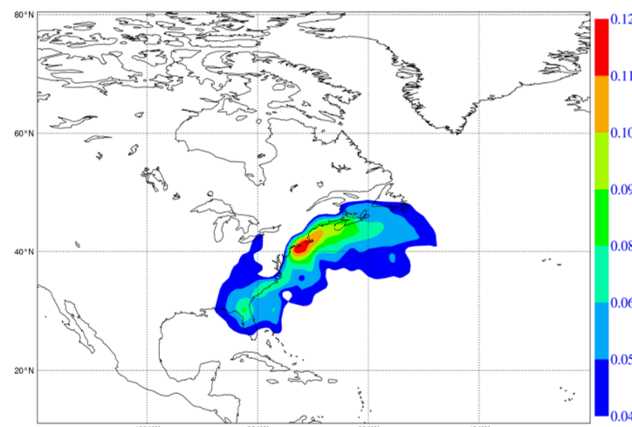
Temperature analysis increments for a single temperature observation at the start of the assimilation window: $x^a(t)\text{-}x^b(t) \approx MP^bM^TH^T(y\text{-}Hx)/(\sigma_b^2 + \sigma_o^2)$



t=+0h



t=+3h



t=+9h

ECMWF

Hybrid Variational–EnKF algorithms

4D Variational methods

- The 4D-Var solution implicitly evolves background error covariances *over the assimilation window* with the tangent linear dynamics:

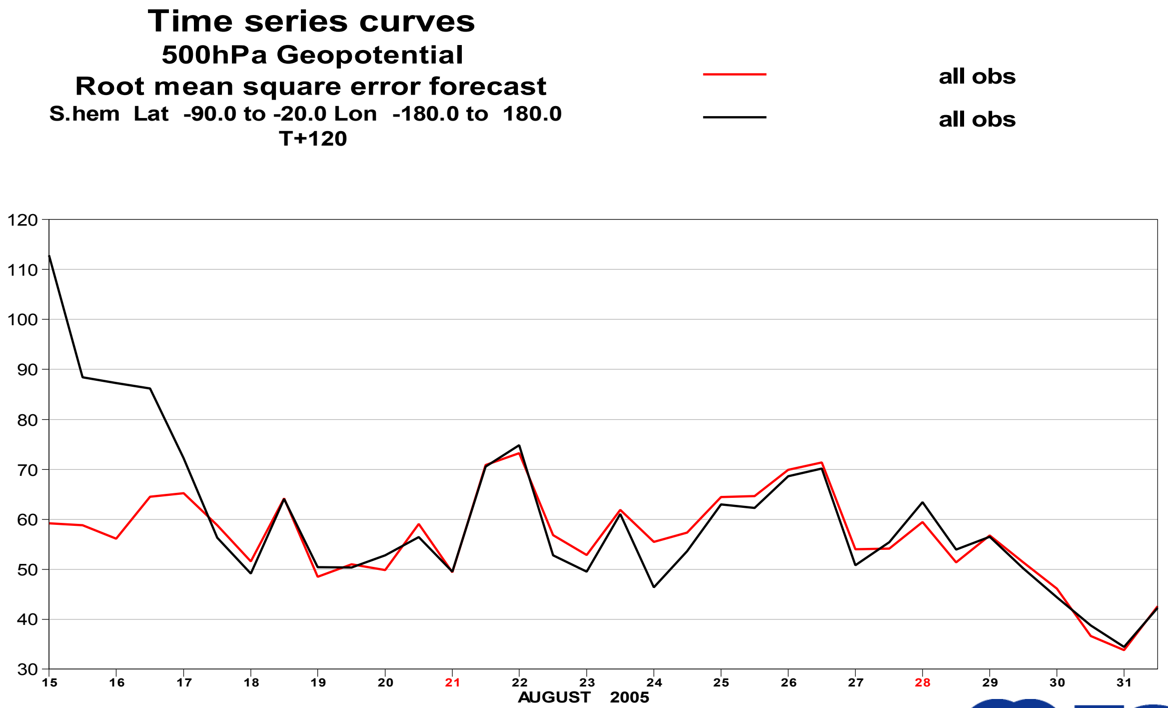
$$\mathbf{P}^b(t) \approx \mathbf{M}\mathbf{P}^b\mathbf{M}^T$$

- But it does not propagate error information from one assimilation cycle to the next: $\mathbf{P}^b$ is not evolved according to KF 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.
- Only information about the state ($\mathbf{x}_b$) is propagated from one cycle to the next.

ECMWF

Hybrid Variational–EnKF algorithms

- What if we pushed back the start of the assimilation window ‘enough’ so that the filter solution at the end of the window would no longer depend on the specified initial $\mathbf{P}^b$?
- How long is enough? 3-5 days in the troposphere for current NWP models, longer in the stratosphere



Hybrid Variational–EnKF algorithms

4D Variational methods

For assimilation windows > 12h it is not accurate to assume the model to be perfect over the assimilation window. For long windows we have to add a model error term to our cost function (Weak-constraint 4D-Var):

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

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

Two caveats:

1. Problem is shifted from estimation of $\mathbf{P^b}$ to estimation of **Q**: this is not any easier!
2. It is difficult in the variational framework to produce good estimates of $\mathbf{P^a}$: this is important for ensemble prediction!

Hybrid Variational–EnKF algorithms

Quick recap:

a) Kalman Filter is computationally unfeasible for large dimensional systems (e.g., operational NWP);

b) Variational (4D-Var) do not cycle state error estimates: work well for short assimilation windows (6-12h). Longer windows, where **Q** is required, have proved more difficult;

c) Reduced rank KF (EnKF) cycle reduced-rank estimates of state error covariances: need for spatial localization to combat rank deficiency, degrades dynamical balance, problematic for non-local observations (radiances);

….

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

Hybrid Variational–EnKF algorithms

Hybrid approach: Use cycled, 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 a 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 for limited ensemble sizes and large model errors
4) Allow consistent localization of ensemble perturbations to be performed in state space (advantageous for radiances);
5) Allow for flow-dependent QC of observations

Hybrid Variational–EnKF algorithms

In operational use (or under test), there are currently three main approaches to doing hybrid DA in a VAR context:

1. Alpha control variable method (Met Office, NCEP/GMAO, CMC)
2. 4D-Ens-Var
3. Ensemble of Data Assimilations method (ECMWF, Meteo France)

ECMWF

Hybrids: α control variable

1. Alpha control variable method (Barker, 1999; Lorenc, 2003)

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

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

from: A.Clayton

ECMWF

Hybrids: α control variable

1. Alpha control variable method

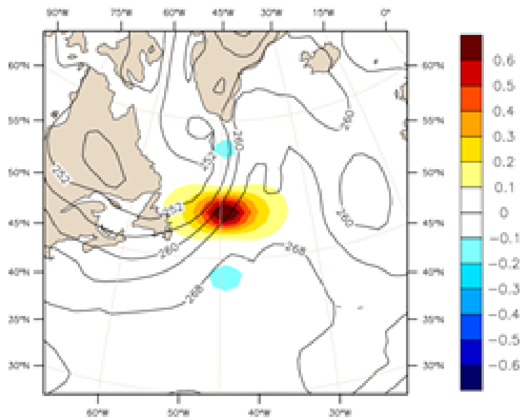
$$\delta \mathbf{x} = \beta_c \mathbf{B}_c^{1/2} \mathbf{v} + \beta_e \mathbf{X}' \circ \boldsymbol{\alpha} = \delta \mathbf{x}_{c\lim} + \delta \mathbf{x}_{ens}$$

- The increment is now a weighted sum of the static $\mathbf{B}$ component and the flow-dependent, ensemble based $\mathbf{B}$
- The flow-dependent increment is a linear combination of ensemble perturbations $\mathbf{X}'$, modulated by the $\boldsymbol{\alpha}$ fields
- If the $\boldsymbol{\alpha}$ fields were homogeneous $\delta \mathbf{x}_{ens}$ could only span N_{ens}-1 degrees of freedom; $\boldsymbol{\alpha}$ fields are then smoothly varying fields, which effectively increases the degrees of freedom
- $\mathbf{C}_{loc}$ is a covariance (localization) model for the flow-dependent increments: it controls the spatial variation of $\boldsymbol{\alpha}$

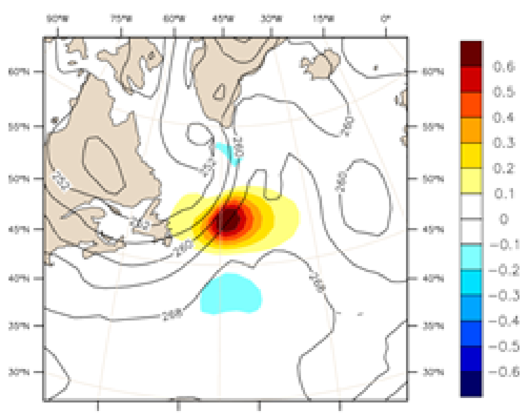
ECMWF

Hybrids: α control variable

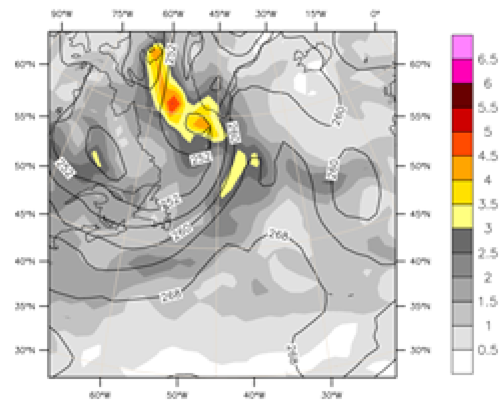
u response to a single u observation at centre of window



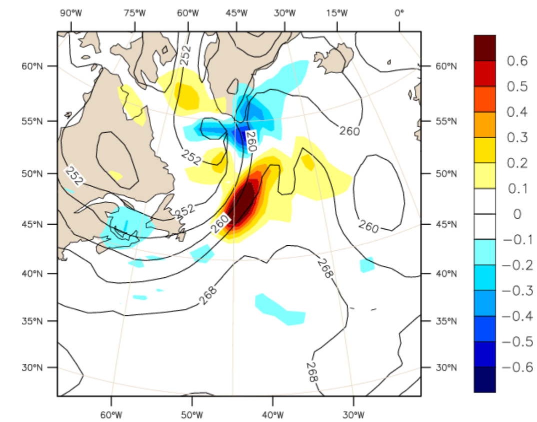
Standard 3D-Var



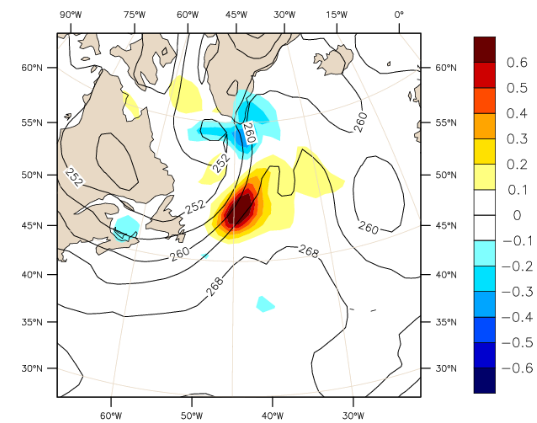
Standard 4D-Var



Ensemble RMS



Pure ensemble 3D-Var



50/50 hybrid 3D-Var

from: A.Clayton

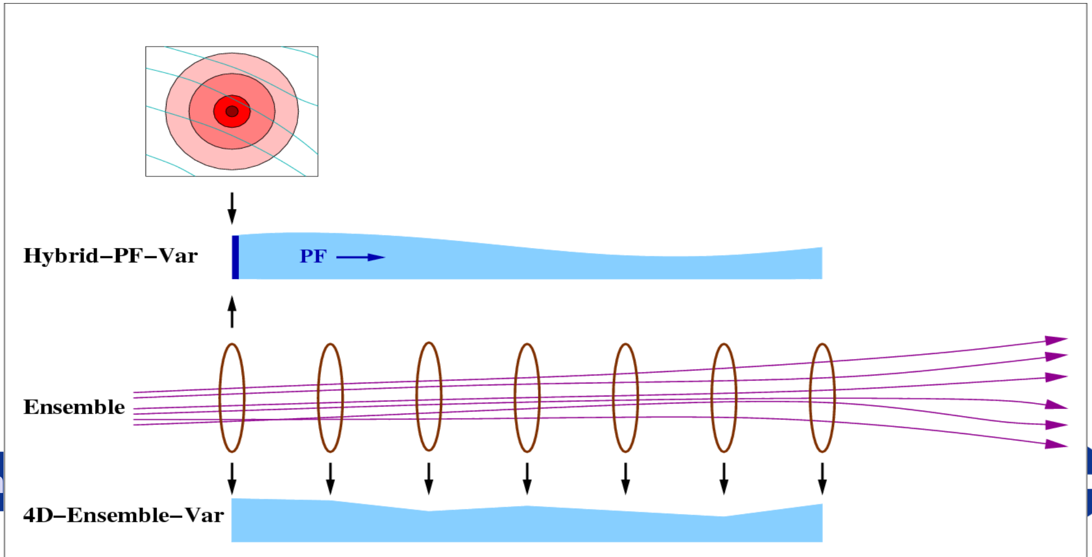
ECMWF

Hybrids: 4D-Ens-Var

2. 4D-Ensemble-Var method (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 4DVar assimilation window: the evolution of $\mathbf{P}^b$ inside the window is due to the tangent linear dynamics ($\mathbf{P}^b(t) \approx \mathbf{M}\mathbf{P}^b\mathbf{M}^T$)

- In 4D-Ens-Var $\mathbf{P}^b$ is sampled from ensemble trajectories throughout the assimilation window:



from: D. Barker

Hybrids: 4D-Ens-Var

2. 4D-Ensemble-Var method (Liu et al., 2008)

- The 4D-Ens-Var analysis is thus a localised linear combination of ensemble trajectories perturbations: conceptually very close to a pure EnKF
- While traditional 4DVar requires repeated, **sequential** runs of $\mathbf{M}$, $\mathbf{M}^T$, ensemble trajectories from the previous assimilation time can be pre-computed in parallel
- Developing and maintaining the TL and Adjoint models requires substantial resources and it is technically demanding: 4D-Ens-Var does not need them

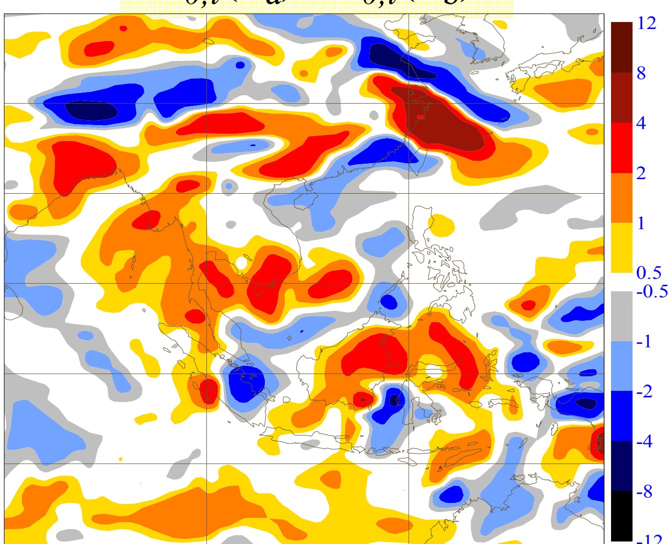
ECMWF

Hybrids: 4D-Ens-Var

- However 4D-Ens-Var requires all ensemble trajectories to be stored in memory: increasingly difficult for larger ensemble sizes/resolutions
- It is typically more accurate to evolve an initial estimate of $\mathbf{P}^b$ by the model TL dynamics than sampling it from an ensemble of trajectories

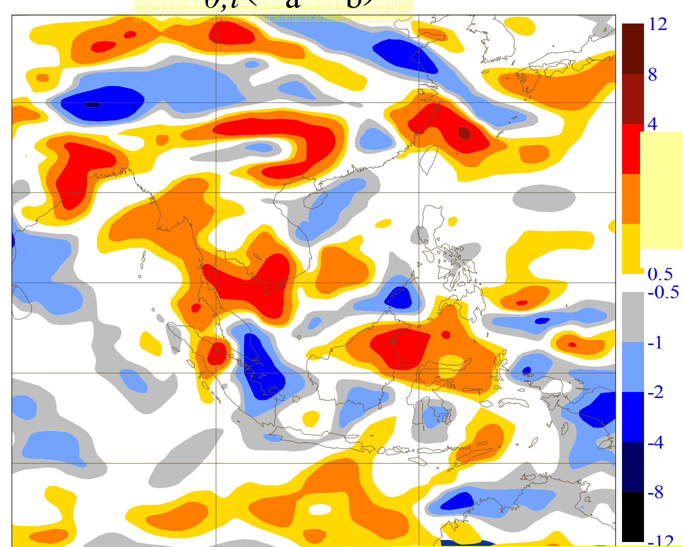
$\mathcal{M}_{0,t}(\mathbf{x}_a)-\mathcal{M}_{0,t}(\mathbf{x}_b)$

Non-linear finite difference



$\mathbf{M}_{0,t}(\mathbf{x}_a-\mathbf{x}_b)$

TL integration



u-wind increments
fc t+12, ~700 hPa

from: M. Janiskova

Hybrids: EDA method

3. Ensemble of Data Assimilations method

- To be continued…

References

1. Anderson, J. L., 2001. An ensemble adjustment Kalman filter for data assimilation. Mon. Wea. Rev. 129, 2884–2903.
2. Bishop, C. H., Etherton, B. J. and Majumdar, S. J., 2001. Adaptive sampling with ensemble transform Kalman filter. Part I: theoretical aspects. Mon. Wea. Rev. 129, 420–436.
3. Burgers, G., Van Leeuwen, P. J. and Evensen, G., 1998. On the analysis scheme in the ensemble Kalman filter. Mon. Wea. Rev. 126, 1719–1724.
4. Campbell, W. F., C. H. Bishop, and D. Hodyss, 2010: Vertical covariance localization for satellite radiances in ensemble Kalman Filters. Mon. Wea. Rev.,138,282–290.
5. Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. J. Geophys. Res. 99(C5), 10 143–10 162.
6. Evensen, G . 2004 . Sampling strategies and square root analysis schemes for the EnKF . No.: 54 . p.: 539-560 . Ocean Dynamics
7. Fisher, M., Leutbecher, M. and Kelly, G. A. 2005. On the equivalence between Kalman smoothing and weak-constraint four-dimensional variational data assimilation. Q.J.R. Meteorol. Soc., 131: 3235–3246.

References

8. Houtekamer, P. L. and Mitchell, H. L., 1998. Data assimilation using an ensemble Kalman filter technique. Mon. Wea. Rev. 126, 796–811.
9. Houtekamer, P. L. and Mitchell, H. L., 2001. A sequential ensemble Kalman filter for atmospheric data assimilation. Mon. Wea. Rev. 129, 123–137.
10. 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.
11. 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.
12. 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.
13. Ott, E., Hunt, B. H., Szunyogh, I., Zimin, A. V., Kostelich, E. J. and co-authors. 2004. A local ensemble Kalman filter for atmospheric data assimilation. Tellus 56A, 415–428.
14. Thépaut, J.-N., Courtier, P., Belaud, G. and Lemaître, G. 1996. Dynamical structure functions in a four-dimensional variational assimilation: A case-study. Q. J. R. Meteorol. Soc., 122, 535–561
15. Whitaker, J. S. and Hamill, T. M., 2002. Ensemble data assimilation without perturbed observations. Mon. Wea. Rev. 130, 1913–1924.