Assimilation Algorithms: Ensemble Kalman Filters

Massimo Bonavita
ECMWF
Massimo.Bonavita@ecmwf.int

Acknowledgements: Mats Hamrud

Outline

- The Standard Kalman Filter
- Kalman Filters for large dimensional systems
- A Monte Carlo implementation of the Kalman Filter: the Ensemble Kalman Filter
- Ensemble Kalman Filters in Hybrid Data Assimilation

Standard Kalman Filter

- In the Overview talk we have seen that, assuming all errors have Gaussian statistics, the posterior (i.e., analysis) distribution $p(\boldsymbol{x}|\boldsymbol{y})$ can also be expressed as a Gaussian distribution:

$$p(\boldsymbol{y}|\boldsymbol{x}) = \frac{1}{(2\pi)^{N/2}|\mathbf{R}|^{1/2}} \boldsymbol{exp}\left(-\frac{1}{2}\left(\boldsymbol{y}-\boldsymbol{H}(\boldsymbol{x})\right)^{\boldsymbol{T}}(\mathbf{R})^{-\mathbf{1}}\left(\boldsymbol{y}-\boldsymbol{H}(\boldsymbol{x})\right)\right)$$

$$p(\boldsymbol{x}_b|\boldsymbol{x}) = \frac{1}{(2\pi)^{N/2}|\mathbf{P}_B|^{1/2}} \boldsymbol{exp}\left(-\frac{1}{2}(\boldsymbol{x_b}-\boldsymbol{x})^{\boldsymbol{T}}(\mathbf{P}_B)^{-\mathbf{1}}(\boldsymbol{x_b}-\boldsymbol{x})\right)$$

$$p(\boldsymbol{x}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x}_b|\boldsymbol{x}) \propto \boldsymbol{exp}\left(-\frac{1}{2}\left(\boldsymbol{y}-\boldsymbol{H}(\boldsymbol{x})\right)^{\boldsymbol{T}}(\mathbf{R})^{-\mathbf{1}}\left(\boldsymbol{y}-\boldsymbol{H}(\boldsymbol{x})\right)-\frac{1}{2}(\boldsymbol{x_b}-\boldsymbol{x})^{\boldsymbol{T}}(\mathbf{P}_B)^{-\mathbf{1}}(\boldsymbol{x_b}-\boldsymbol{x})\right)$$

- Kalman Filter methods are designed to find the mean and covariance of this posterior distribution and to cycle it in time

- Under this Gaussian assumption, knowing the mean and covariance of $p(\boldsymbol{x}|\boldsymbol{y})$ amounts to knowing the full posterior pdf

Standard Kalman Filter

- Let us consider a univariate 1-D example:

Assume we are analysing a single state variable x, whose errors are zero mean and normally distributed around its background forecast x_b:

$$p(x_b|x) = \frac{1}{(2\pi\sigma_b^2)^{1/2}} exp\left(-\frac{1}{2}\frac{(x_b-x)^2}{\sigma_b^2}\right) \sim \mathcal{N}(x_b, \sigma_b^2)$$

We have one observation of the state variable, also with Gaussian errors:

$$p(y|x) = \frac{1}{(2\pi\sigma_o^2)^{1/2}} exp\left(-\frac{1}{2}\frac{(y-x)^2}{\sigma_o^2}\right) \sim \mathcal{N}(y, \sigma_o^2)$$

Applying Bayes theorem we find:

$$p(x|y) \propto p(y|x)p(x_b|x) \propto exp\left(-\frac{1}{2}\frac{(y-x)^2}{\sigma_b^2} - \frac{1}{2}\frac{(x_b-x)^2}{\sigma_b^2}\right) \propto exp\left(-\frac{1}{2}\left(\left(\frac{1}{\sigma_b^2}+\frac{1}{\sigma_o^2}\right)x^2 - 2\left(\frac{x_b}{\sigma_b^2}+\frac{y}{\sigma_o^2}\right)x\right)\right)$$

Comparing to a standard Gaussian distribution:

$$p(x) = \frac{1}{(2\pi\sigma^2)^{1/2}} exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right) = \frac{1}{(2\pi\sigma^2)^{1/2}} exp\left(-\frac{1}{2}\left(\left(\frac{1}{\sigma^2}\right)x^2 - 2\left(\frac{\mu}{\sigma^2}\right)x + \left(\frac{\mu^2}{\sigma^2}\right)\right)\right)$$

we see that the posterior distribution $p(x|y)$ is also Gaussian with mean and variance:

$$Var(x|y) = \sigma^2 = \left(\frac{1}{\sigma_b^2}+\frac{1}{\sigma_o^2}\right)^{-1} = \frac{\sigma_o^2\sigma_b^2}{\sigma_o^2+\sigma_b^2}$$

$$E(x|y) = \mu = \sigma^2\left(\frac{x_b}{\sigma_b^2}+\frac{y}{\sigma_o^2}\right) = \frac{\sigma_o^2\sigma_b^2}{\sigma_o^2+\sigma_b^2}\left(\frac{x_b}{\sigma_b^2}+\frac{y}{\sigma_o^2}\right) = \frac{\sigma_o^2}{\sigma_o^2+\sigma_b^2}x_b + \frac{\sigma_b^2}{\sigma_o^2+\sigma_b^2}\,\mathrm{y}$$

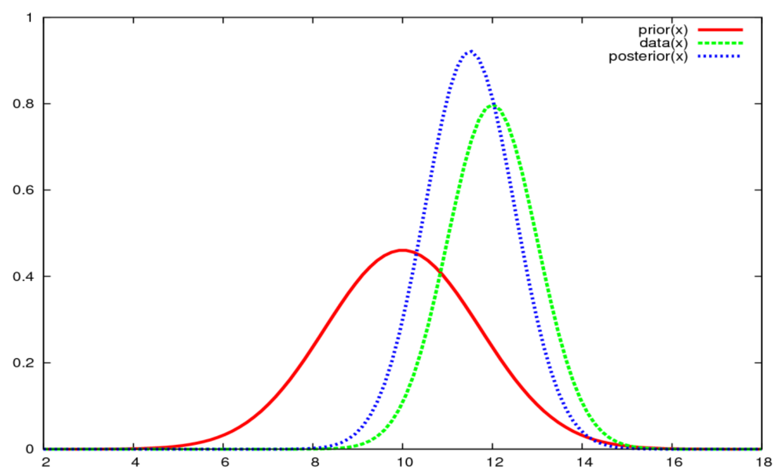
Standard Kalman Filter

Defining the **Kalman gain**: $K \stackrel{\text{def}}{=} \frac{\sigma_b^2}{\sigma_o^2 + \sigma_b^2}$ the equations for the mean and variance can be recast as:

$$Var(x|y) = \sigma_a^2 = \frac{\sigma_o^2 \sigma_b^2}{\sigma_o^2 + \sigma_b^2} = (1 - K)\sigma_b^2$$

$$E(x|y) = x_a = \frac{\sigma_o^2}{\sigma_o^2 + \sigma_b^2} x_b + \frac{\sigma_b^2}{\sigma_o^2 + \sigma_b^2} y = x_b + K(y - x_b)$$

The posterior variance is thus reduced $(1\text{-}K<0)$ with respect to the prior (background) variance, while the posterior mean is a linear, weighted average of the prior (background) and the observation.



Standard Kalman Filter

- These Kalman Filter analysis update equations can be generalised to the multi-dimensional and multivariate case (Wikle and Berliner, 2007; Bromiley, 2014):

$$E(\boldsymbol{x}|\boldsymbol{y}) = \boldsymbol{x_a} = \boldsymbol{x_b} + \mathbf{K}(\boldsymbol{y} - \mathbf{H}(\boldsymbol{x_b}))$$

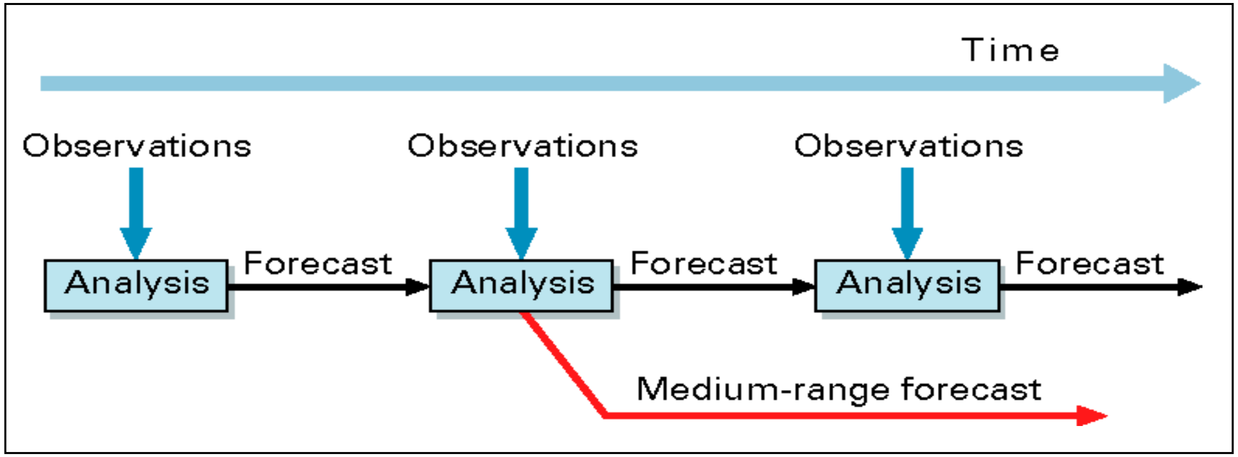
$$\begin{aligned} Var(\boldsymbol{x}|\boldsymbol{y}) = \mathbf{P}^a = \ & (\mathbf{I} - \mathbf{KH})\mathbf{P}^b(\mathbf{I} - \mathbf{KH})^T + \mathbf{KRK}^{\boldsymbol{T}} = \\ & (\mathbf{I} - \mathbf{KH})\mathbf{P}^b - \mathbf{P}^b\mathbf{H}^T\mathbf{K}^T + \mathbf{K}\left(\mathbf{HP}^b\mathbf{H}^T + \mathbf{R}\right)\mathbf{K}^T = \\ & (\mathbf{I} - \mathbf{KH})\mathbf{P}^b \end{aligned}$$

$$\mathbf{K} = \mathbf{P}^b\mathbf{H}^T\left(\mathbf{HP}^b\mathbf{H}^T + \mathbf{R}\right)^{-1} = \left(\left(\mathbf{P}^b\right)^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{R}^{-1}$$

- These are the same update equations for the state $\boldsymbol{x}$ and its uncertainty $\mathbf{P}$ obtained in the lecture on Assimilation Algorithms (1).
- In that context they were derived without making assumptions of Gaussianity, but looking for the analysis estimate which had: 1) the minimum error variance and 2) could be expressed as a linear combination of the background and the observations (we called it the BLUE, Best Linear Unbiased Estimate). Linearity of observation operator and model was invoked.
- If the background and observations are normally distributed, the KF update equations give the mean and the covariance of the posterior distribution. Under these hypotheses the posterior distribution is also Gaussian, so we have completely solved the problem!

Standard Kalman Filter

- In NWP applications 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 assimilation



- For each analysis update in this cycle, we require a background $\mathbf{x}^b_t$ (i.e. a prior estimate of the state valid at time t)

- Usually, our best prior estimate of the state at time t is given by a forecast from the preceding analysis at time t-1 (the "background"):

$$\mathbf{x}^b_t = \mathbf{M}(\mathbf{x}^a_{t-1})$$

- What is the error covariance matrix (=> the uncertainty) associated with this background?

Standard Kalman Filter

$$\mathbf{x}^b{}_t = \mathbf{M}(\mathbf{x}^a{}_{t-1})$$

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

$$\mathbf{x}^b - \mathbf{x}^*{}_t = \boldsymbol{\varepsilon}^b{}_t = \mathbf{M}(\mathbf{x}^a{}_{t-1}) - \mathbf{x}^*{}_t$$

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

$$\boldsymbol{\varepsilon}^b{}_t = \mathbf{M}(\mathbf{x}^*{}_{t-1} + \boldsymbol{\varepsilon}^a{}_{t-1}) - \mathbf{x}^*{}_t \approx$$

$$\mathbf{M}(\mathbf{x}^*{}_{t-1}) + \mathbf{M}\boldsymbol{\varepsilon}^a{}_{t-1} - \mathbf{x}^*{}_t =$$

$$\mathbf{M}\boldsymbol{\varepsilon}^a{}_{t-1} + \boldsymbol{\eta}_t$$

- Here we have defined the model error $\boldsymbol{\eta}_t \stackrel{\text{def}}{=} \mathbf{M}(\mathbf{x}^*{}_{t-1}) - \mathbf{x}^*{}_t$

- We assume small errors (linearisation of model) and no systematic errors are present in our system (or have been separately dealt with!)

$$< \boldsymbol{\varepsilon}^a > = < \boldsymbol{\eta} > = 0 \quad => \quad < \boldsymbol{\varepsilon}^b > = 0$$

Standard Kalman Filter

- The background error covariance matrix will then be given by:

$$\mathbf{P}^b_t \stackrel{\text{def}}{=} <\boldsymbol{\varepsilon}^b_t\,(\boldsymbol{\varepsilon}^b_t)^T> \stackrel{\text{def}}{=} = <(\mathbf{M}\boldsymbol{\varepsilon}^a_{t-1} + \boldsymbol{\eta}_t)\,(\mathbf{M}\boldsymbol{\varepsilon}^a_{t-1} + \boldsymbol{\eta}_t)^T> =$$

$$\mathbf{M}<\boldsymbol{\varepsilon}^a_{t-1}\,(\boldsymbol{\varepsilon}^a_{t-1})^T>\mathbf{M}^T + <\boldsymbol{\eta}_t\,(\boldsymbol{\eta}_t)^T> =$$

$$= \mathbf{M}\,\mathbf{P}^a_{t-1}\,\mathbf{M}^T + \mathbf{Q}_t$$

- Here we have assumed $< \boldsymbol{\varepsilon}^a_{t-1}\,(\boldsymbol{\eta}_t)^T> = 0$ and defined the model error covariance matrix $\mathbf{Q}_t \stackrel{\text{def}}{=} <\boldsymbol{\eta}_t\,(\boldsymbol{\eta}_t)^T>$

- Note how the background error is described as the sum of the errors of the previous analysis propagated by the linear (or linearised) model dynamics to the time of the new update ($\mathbf{M}\,\mathbf{P}^a_{t-1}\,\mathbf{M}^T$, also called "predictability" error covariance) and the new additive errors introduced by the model integration ($\mathbf{Q}_t$)

- We now have all the equations necessary to propagate and update both the state and its error estimates

Standard Kalman Filter

$\mathbf{x}^a_{t-1}$, $\mathbf{P}^a_{t-1}$ — t-1 → $\mathbf{x}^b_t$, $\mathbf{P}^b_t$ — t — $\mathbf{x}^a_t$, $\mathbf{P}^a_t$ → t+1 — $\mathbf{X}^b_{t+1}$, $\mathbf{P}^b_{t+1}$

New Observations

1. Predict the state ahead

   $$\mathbf{x}^b_t = \mathbf{M}(\mathbf{x}^a_{t-1})$$

2. Predict the state error cov.

   $$\mathbf{P}^b_t = \mathbf{M}\,\mathbf{P}^a_{t-1}\,\mathbf{M}^T + \mathbf{Q}_t$$

Propagation

3. Compute the Kalman Gain

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

4. Update state estimate

   $$\mathbf{x}^a_t = \mathbf{x}^b_t + \mathbf{K}\,(\mathbf{y} - \mathbf{H}(\mathbf{x}^b_t))$$

5. Update state error estimate

   $$\mathbf{P}^a_t = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^b_t\,(\mathbf{I} - \mathbf{K}\mathbf{H})^T + \mathbf{K}\mathbf{R}\mathbf{K}^T$$

Update

1. Predict the state ahead

   $$\mathbf{x}^b_{t+1} = \mathbf{M}(\mathbf{x}^a_t)$$

2. Predict the state error cov.

   $$\mathbf{P}^b_{t+1} = \mathbf{M}\,\mathbf{P}^a_t\,\mathbf{M}^T + \mathbf{Q}_{t+1}$$

Propagation

Standard Kalman Filter

- Under the assumption that the model $\mathbf{M}$ and the observation operator $\mathbf{H}$ are linear operators (i.e., they do not depend on the state $\mathbf{x}$), the Kalman Filter produces an optimal sequence of analyses ($\mathbf{x}^a_1$, $\mathbf{x}^a_2$, …, $\mathbf{x}^a_{t-1}$, $\mathbf{x}^a_t$)
- The KF analysis $\mathbf{x}^a_t$ is the best (minimum error variance) linear unbiased estimate of the state at time t, given $\mathbf{x}^b_t$ and all observations up to time t ($\mathbf{y}_0$,$\mathbf{y}_1$,…,$\mathbf{y}_t$).
- Note that Gaussianity of errors has not been invoked. If errors are Gaussian the Kalman Filter provides the correct posterior conditional probability estimate (according to Bayes' Law), i.e. p($\mathbf{x}^a_t$| $\mathbf{x}^b_0$; $\mathbf{y}_0$,$\mathbf{y}_1$,…,$\mathbf{y}_t$). This also implies that if errors are Gaussian then the state estimated with the KF is also the most likely state (the mode of the pdf).

Extended Standard Kalman Filter

- The extended Kalman Filter (EKF) is a non-linear extension of the Kalman Filter where the model and observation operators are allowed to be nonlinear operators (independent of the state) as in the standard KF:

$$\mathbf{y} = \mathcal{H}(\mathbf{x}_b) + \boldsymbol{\varepsilon}_o \quad \text{(EKF)} \quad \Leftarrow \quad \mathbf{y} = \mathbf{H}\mathbf{x}_b + \boldsymbol{\varepsilon}_o \quad \text{(KF)}$$

$$\mathbf{x}_\mathrm{b} = \mathcal{M}(\mathbf{x}_a) + \boldsymbol{\eta} \quad \text{(EKF)} \quad \Leftarrow \quad \mathbf{x}_b = \mathbf{M}\mathbf{x}_a + \boldsymbol{\eta} \quad \text{(KF)}$$

- The covariance update and prediction steps of the KF equations use the Jacobians of the model and observation operators, linearized around the current state estimate:

$$\mathbf{M} = \frac{\partial \mathcal{M}}{\partial \boldsymbol{x}}(\mathbf{x}_\mathrm{t}), \quad \mathbf{H} = \frac{\partial \mathcal{H}}{\partial \boldsymbol{x}}(\mathbf{x}_\mathrm{t})$$

- The EKF is thus a first order linearization of the KF equations around the current state estimates. Reasonable for systems which are only moderately nonlinear on the timescales of the DA cycle update interval

- A type of EKF has long been used in the land DA community (and ECMWF) in the analysis of soil variables (Simplified Extended Kalman Filter, SEKF). More on this later this week in the Land DA lecture.

Kalman Filters for large dimensional systems

- The Kalman Filter (standard or extended) is unfeasible for large dimensional systems

- The size N of the analysis/background state in the ECMWF 4DVar is $O(10^8)$: the KF requires us to store and evolve in time state covariance matrices ($\mathbf{P}^{a/b}$) of $O(N\mathrm{x}N)$
  - The World's fastest computer can sustain ~ 10^{18} operations per second
  - An efficient implementation of matrix multiplication of two $10^8\mathrm{x}10^8$ matrices requires ~10^{22} ($O(N^{2.8})$)operations: hours on current fast HPCs!
  - Brute force evaluating $\mathbf{P}^b{}_t = \mathbf{M}\,\mathbf{P}^a{}_{t\text{-}1}\,\mathbf{M}^T + \mathbf{Q}_k$ requires $2*N \approx 2*10^8$ model integrations!

- A range of approximate Kalman Filters has been developed for use with large-dimensional systems.

- All of these methods rely on some form of low-rank approximation of the state covariance matrices.

Kalman Filters for large dimensional systems

- Let us assume that $\mathbf{P}^{a/b}$ has rank M<<N (e.g. M≈100). (rank=dim. of vector space spanned by its columns/rows)
- In this case we can write $\mathbf{P}^b = \mathbf{X}^b(\mathbf{X}^b)^T$, where $\mathbf{X}^b{}_k$ is N x M. This decomposition also assures us that the resulting $\mathbf{P}^b$ is positive semi definite. $\mathbf{X}^b$ is a thin matrix!
- The Kalman Gain then becomes:

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

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

$$\mathbf{X}^b\,(\mathbf{H}\mathbf{X}^b)^T(\mathbf{H}\,\mathbf{X}^b(\mathbf{H}\mathbf{X}^b)^T + \mathbf{R})^{-1}$$

- Note that, to evaluate $\mathbf{K}$, we apply $\mathbf{H}$ to the M columns of $\mathbf{X}^b$ rather than to the N columns of $\mathbf{P}^b$!
- The N x N matrices $\mathbf{P}^{a/b}$ have been eliminated from the computation! In their place we have to deal with thin N x M ($\mathbf{X}^b$) matrices in state space and their observation space projections $\mathbf{H}\mathbf{X}^b$ matrices which have dimension L x M (L = number of observations)

Kalman Filters for large dimensional systems

- The approximated KF described above is called Reduced-Rank Kalman Filter (RRKF)
- Unsurprisingly, there is a price to pay for this huge reduction in computational cost
- The analysis increment is a linear combination of the columns of $\mathbf{X}^b$:

$$\mathbf{x}^a - \mathbf{x}^b = \mathbf{K}\,(\mathbf{y} - H(\mathbf{x}^b)) = \mathbf{X}^b\,(\mathbf{H}\mathbf{X}^b)^T\,((\mathbf{H}\mathbf{X}^b)(\mathbf{H}\mathbf{X}^b)^T + \mathbf{R})^{-1}\,(\mathbf{y} - H(\mathbf{x}^b))$$

- The whole blue part of the equation computes to a vector of size M (ie, the number of columns of $\mathbf{X}^b$, which is the rank of $\mathbf{P}^b$)!
- The analysis increments are thus formed as a linear combination of the columns of $\mathbf{X}^b$: they are confined to the column subspace of $\mathbf{X}^b$, which has at most rank $M << N$.
- This severe reduction in rank of $P^{a/b}$ has two main effects:

  1. There are only M (~100) degrees of freedom available to fit the O(10^7) observations available during the analysis window: the analysis will smooth out local detail;
  2. The low-rank approximations of the covariance matrices suffer from spurious long-distance correlations and cross-correlations.

Kalman Filters for large dimensional systems

- Localization of the analysis update is the standard solution to the rank deficiency/sampling noise problem. It comes in two main flavours:

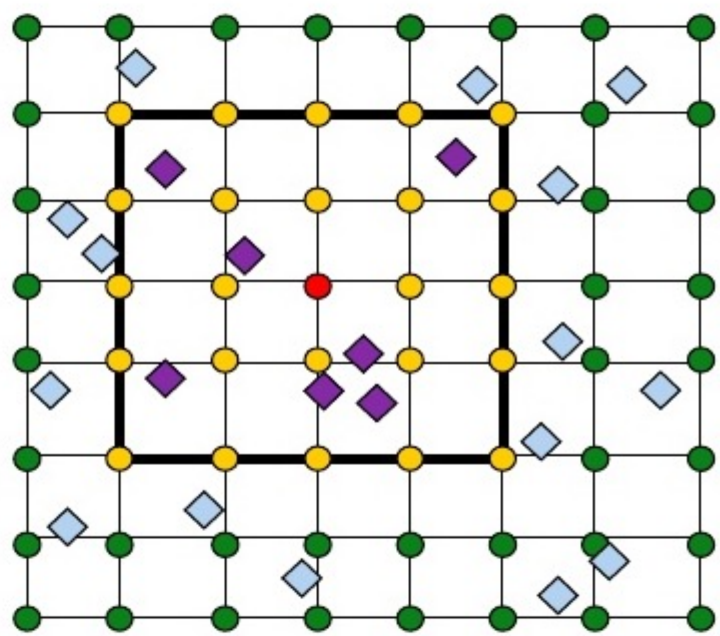
    1. Domain localization (also called "Observation space localization", Houtekamer and Mitchell, 1998; Ott *et al.* 2004);

- Domain localization solves the analysis equations independently for each grid point, or for each of a set of regions covering the domain (very good for parallelisation!)
- Each analysis uses only observations that are local to the grid point (or region) and the observations are usually weighted according to their distance from the analysed grid point (e.g., Hunt *et al.*, 2007)
- This guarantees that the analysis at each grid point (or region/column) is not influenced by distant observations (-> noisy grid point – obs correlations are suppressed)
- The method acts to vastly increase the dimension of the sub-space in which the analysis increment is constructed because each grid point (region/column) is updated by a different linear combination of ensemble perturbations
- However, performing independent analyses for each region can lead to a) reduced skill in the analysis of the large scales and b) in producing balanced analyses.

Kalman Filters for large dimensional systems

Domain localization (e.g. Houtekamer and Mitchell, 1998; Ott *et al.* 2004, Hunt et al., 2007);

- Analysed grid point
- Local observations



Kalman Filters for large dimensional systems

2. Covariance localization (also called "Model space localization", Houtekamer and Mitchell, 2001).

- Covariance localization is performed by element wise (Schur) multiplication of the error covariance matrices with a predefined correlation matrix representing a decaying function of distance (vertical and/or horizontal).

$$\mathbf{P}^b \dashrightarrow \rho_L \circ \mathbf{P}^b$$

- In this way spurious long-range correlations in $\mathbf{P}^b$ are suppressed. ρ_L is often called a "moderation function"
- 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 and largely heuristic. It is easy to modify $\mathbf{P}^b$ in undesirable ways. In particular, physical balance relationships (e.g. geostrophy) may be adversely affected and long-distance correlations will be ignored.
- In order to suppress sampling noise some of the information content of the observations is always lost

Kalman Filters for large dimensional systems

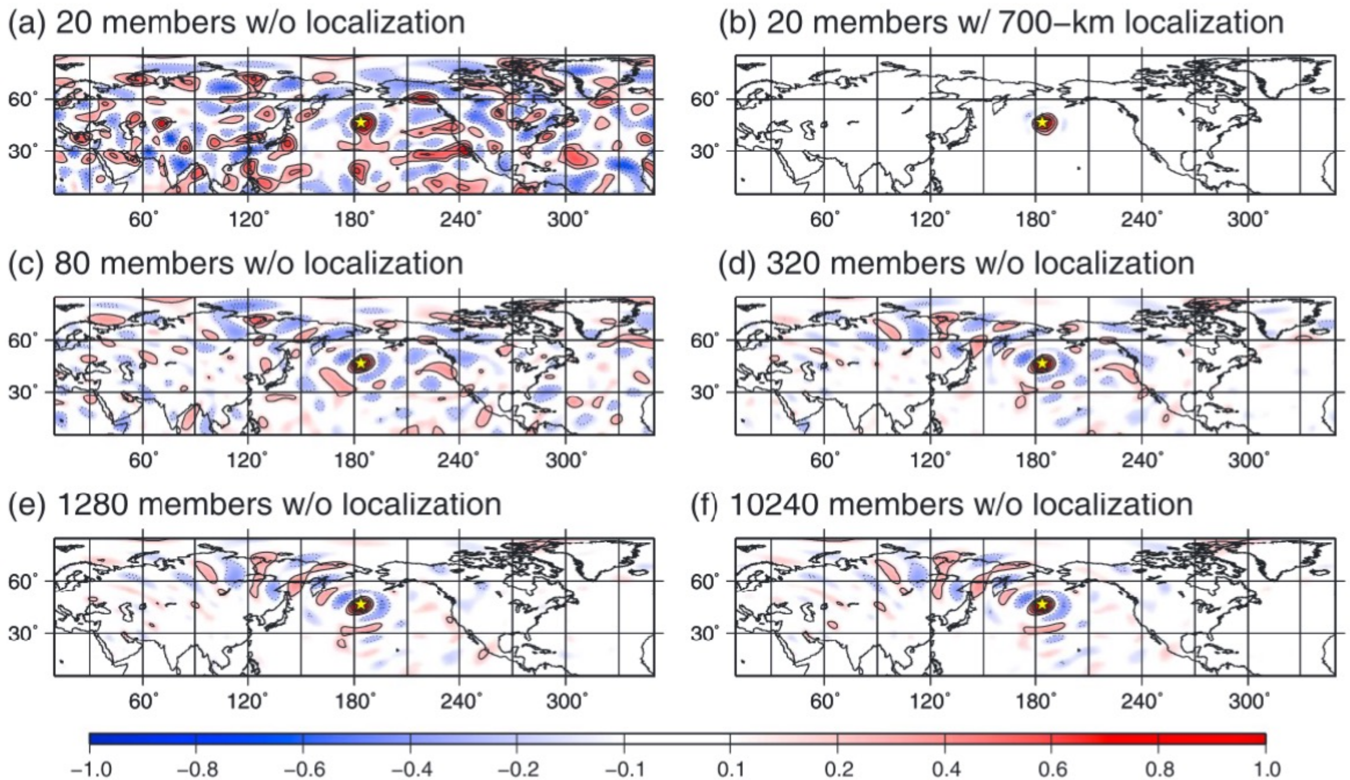
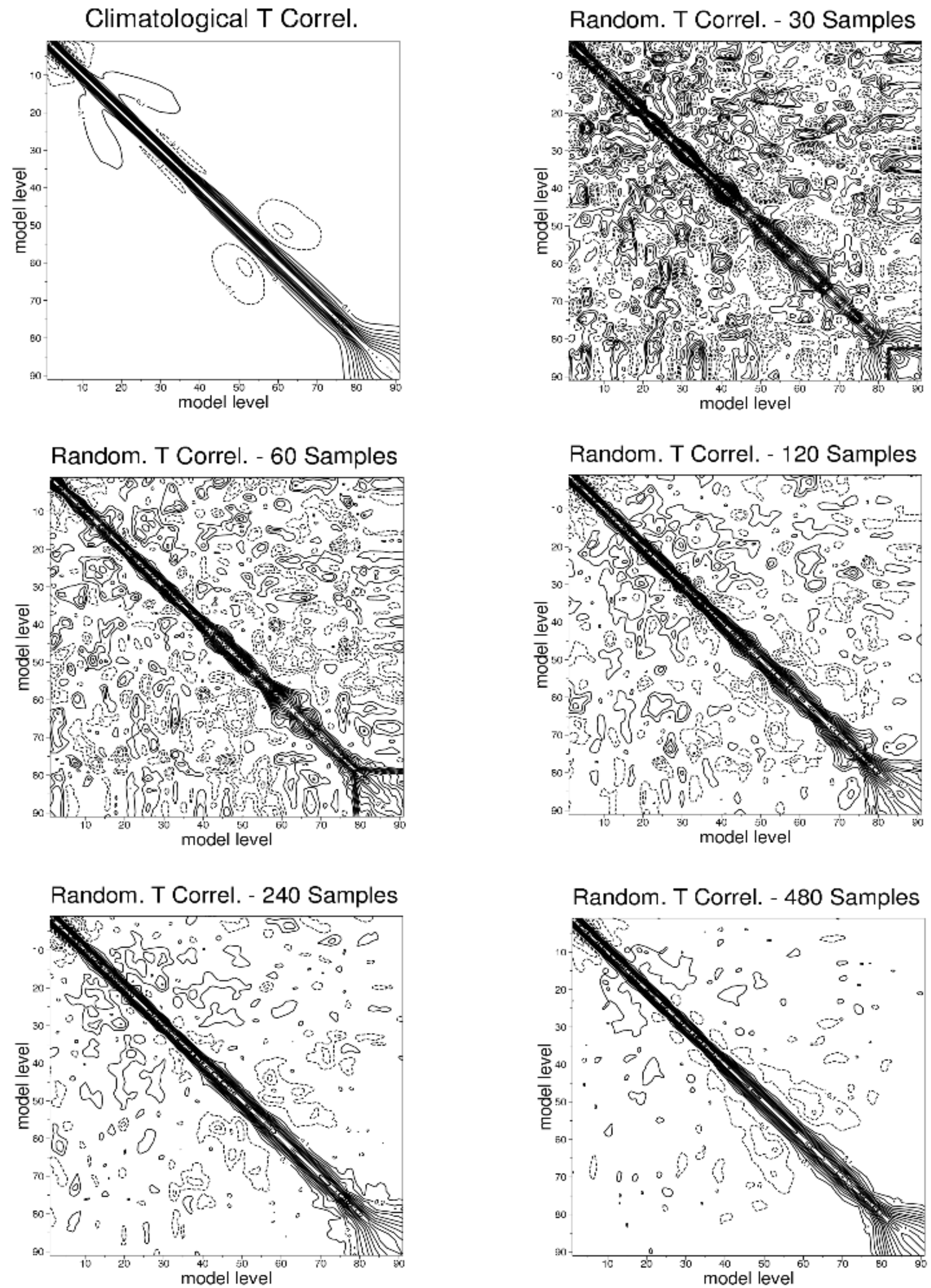


Figure 4. Similar to Figure 1 but at 00:00 UTC 18 January with the yellow star point at 46.389°N, 176.25°W and for different ensemble sizes ((a) 20, (c) 80, (d) 320, (e) 1280, and (f) 10,240 members) and (b) with localization for 20 members.

Miyoshi et al., 2014

Random sampling of vertical background error correlation matrix for different ensemble sizes.

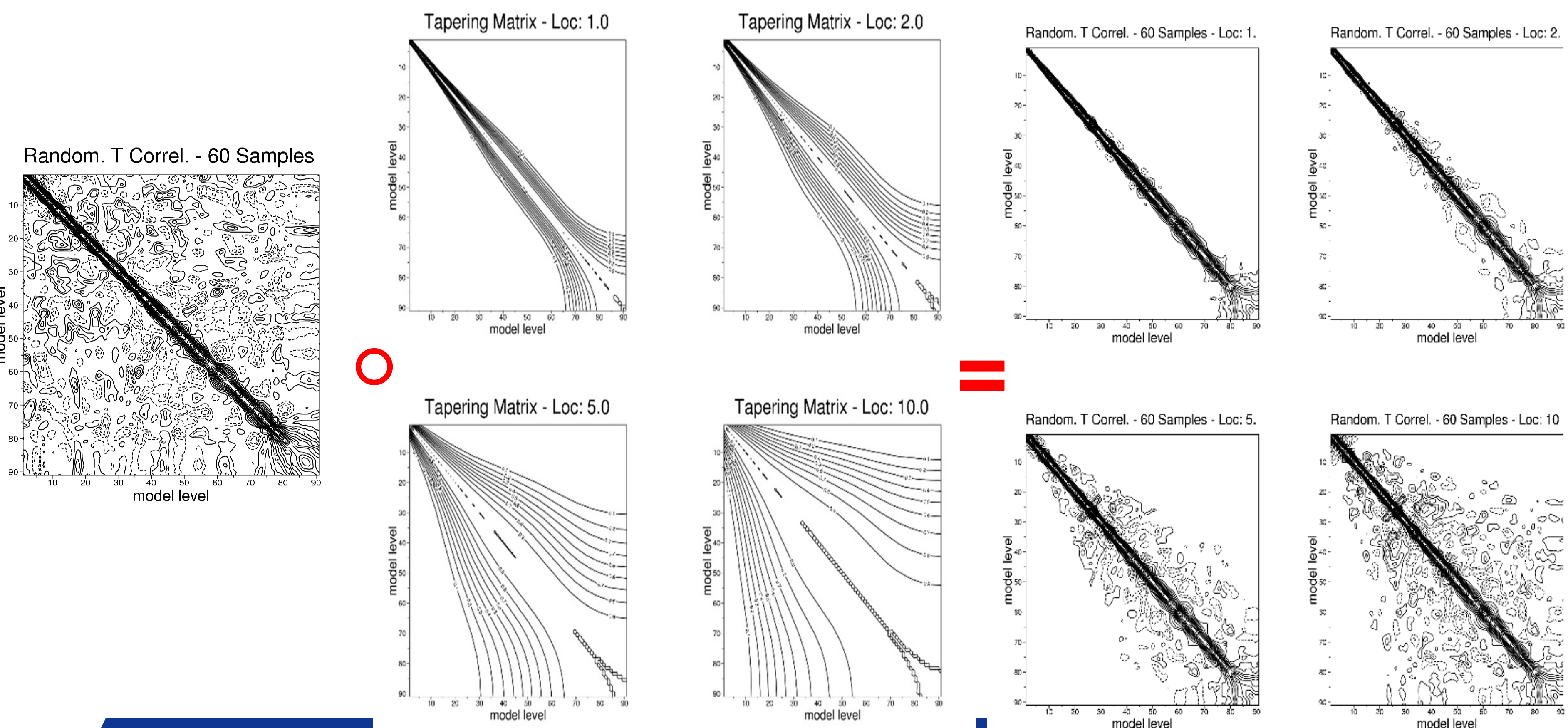
Note how sampling noise decreases slowly with ensemble size $O(M^{1/2})$



ECMWF EUROPEAN CEI

- Standard Error of sample correlation ≈ $(1-\rho^2)/\sqrt{(N_{ens}-1)}$
- For small ρ, N_{ens} SE becomes ~ ρ (e.g. $\rho=0.1$, $N_{ens}=40 \Rightarrow \text{stderr}(\rho)\approx 0.16$)
- Since ρ becomes small for large horiz./vert. distances, we apply distance based covariance localization on the sample $\mathbf{P}^f$

$$P^f_{sampled} \circ \rho_L = P^f_{local}$$



Kalman Filters for large dimensional systems

- Domain/Covariance localization is a practical requirement for using the KF in large dimensional applications
- Finding the right amount of localization is an (expensive) tuning exercise: a good trade-off needs to be found between computational effort, controlling sampling error and not losing observational information
- Finding the "optimal" localization scales as functions of the system characteristics is an area of active research (e.g., Flowerdew, 2015; Periáñez et al., 2014; Menetrier et al., 2014; Bishop, 2017)
- Recent ideas on how to combine domain and covariance localisation will also play a role (eg, domain loc. in the horizontal, covariance loc. in the vertical, Farchi and Bocquet, 2019)

Ensemble Kalman Filters

- Ensemble Kalman Filters (EnKF, Evensen, 1994; Houtekamer and Mitchell, 1998; Burgers et al., 1998, **Houtekamer and Zhang, 2016**) 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, of size M<<N:

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

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

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

$$\mathbf{X}^{b} = \frac{1}{\sqrt{M-1}}((\mathbf{x}^{b}{}_{1} - <\mathbf{x}^{b}>), (\mathbf{x}^{b}{}_{2} - <\mathbf{x}^{b}>), .., (\mathbf{x}^{b}{}_{M} - <\mathbf{x}^{b}>))$$

- Note that the full covariance matrix is never formed explicitly: The error covariances are usually computed locally for each grid point (or column) in the *M* x *M* ensemble space

Ensemble Kalman Filters

- In the standard KF the error covariances are explicitly computed and propagated in time using the tangent linear and adjoint of the model and observation operators, i.e.:

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

$$\mathbf{P}^b = \mathbf{M}\mathbf{P}^a\mathbf{M}^T + \mathbf{Q}$$

- In the EnKF the error covariances are sampled from the ensemble forecasts ($M(x_m)$) and their observation equivalents ($H\big(M(x_m)\big)$) and the huge matrix $\mathbf{P}^b$ is never explicitly formed:

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

$$\frac{1}{M-1}\Sigma_m(\mathbf{x}^b{}_m - <\mathbf{x}^b{}_m>)\ (\mathbf{H}\mathbf{x}^b{}_m - <\mathbf{H}(\mathbf{x}^b{}_m)>)^T$$

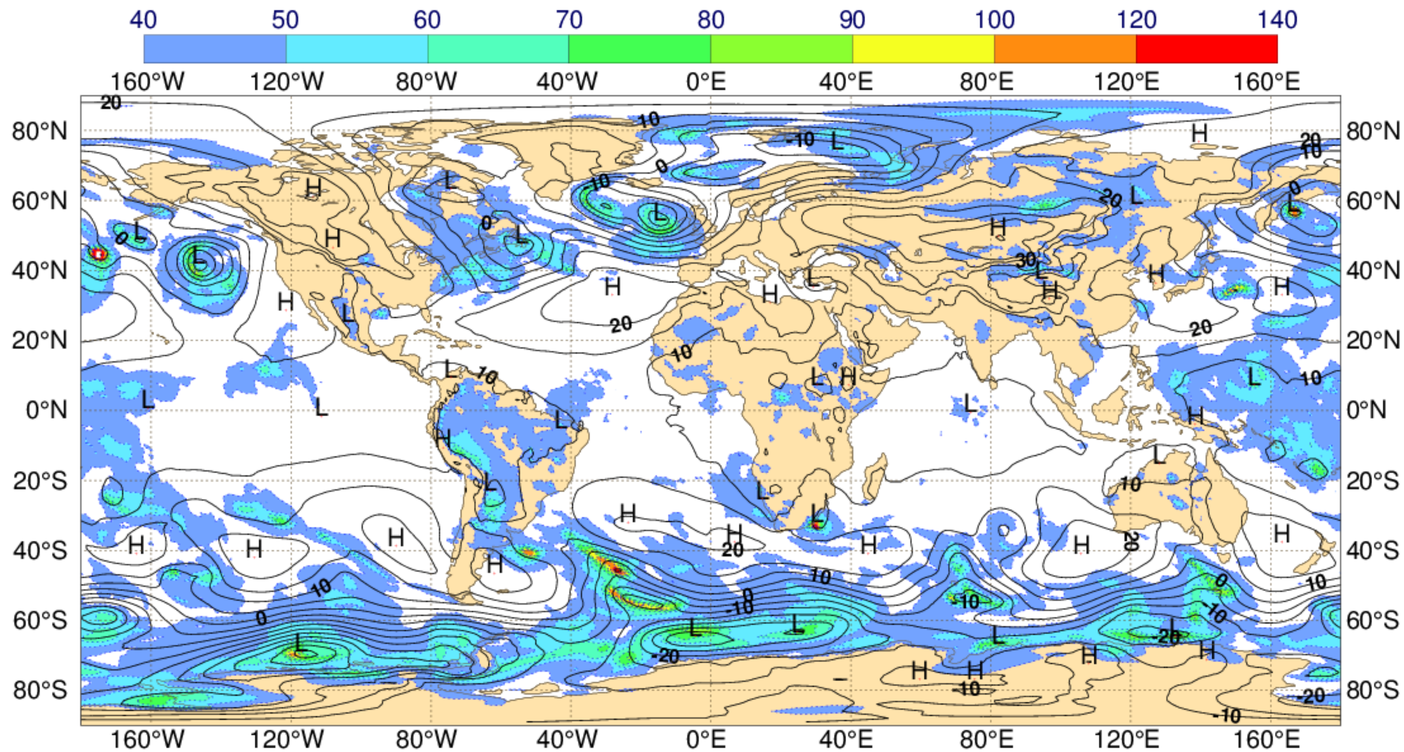
$$\mathbf{H}\mathbf{P}^b\mathbf{H}^T = \mathbf{H}\mathbf{X}^b(\mathbf{H}\mathbf{X}^b)^T \approx$$

$$\frac{1}{M-1}\Sigma_m(\mathbf{H}\mathbf{x}^b{}_m - <\mathbf{H}(\mathbf{x}^b{}_m)>)\ (\mathbf{H}\mathbf{x}^b{}_m - <\mathbf{H}(\mathbf{x}^b{}_m)>)^T$$

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

Ensemble Kalman Filters

- In the EnKF the error covariances are sampled from the ensemble forecasts. They reflect uncertainties in the state of the atmospheric flow



Standard deviation of surface pressure background t+6h fcst (shaded, Pa)
Z1000 background t+6h fcst (black isolines)

Ensemble Kalman Filters

- The Ensemble Kalman Filter is a Monte Carlo technique: it requires us to generate a sample $\{\mathbf{x}^b_m;\ m=1,..,M\}$ drawn from the pdf of background error: how to do this?

- We can generate this from a sample $\{\mathbf{x}^a_{t-1,m};\ m=1,..,M\}$ of the pdf of analysis error for the previous cycle:

$$\mathbf{x}^b_{t,m} = \mathcal{M}(\mathbf{x}^a_{t-1,m}) + \boldsymbol{\eta}_m$$

where $\boldsymbol{\eta}_m$ is a sample drawn from the pdf of model error.

- This shifts the problem to: How do we generate a sample from the analysis pdf? 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 equation as the state

Ensemble Kalman Filters

- Consider now an ensemble of analysis where all the inputs to the analysis (i.e., the background forecast and the observations) have been perturbed according to their errors:

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

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

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

- Note that the **observations** (during the update step) and the **model** (during the forecast step) are **perturbed explicitly** (i.e., we add random numbers with prescribed statistics).

- The **background** is **implicitly perturbed** , i.e.:

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

- Hence, one way to generate a sample drawn from the pdf of analysis error is to perturb the observations and the model with perturbations drawn from their error covariances.

- The EnKF based on this idea is called Perturbed Observations (Stochastic) EnKF (Houtekamer and Mitchell, 1998). It is also the basis of ECMWF EDA (more on this later)

Ensemble Kalman Filters

- Another way to construct the analysis sample without perturbing the observations (but still perturb the model!) is to make a linear combination of the background sample:

$$\mathbf{X}^a=\mathbf{X}^b\mathbf{T}$$

where $\mathbf{T}$ is a linear transformation $(M \times M)$ chosen so that it produces the correct analysis covariance when applied to $\mathbf{X}^b$:

$$\mathbf{X}^a(\mathbf{X}^a)^T = (\mathbf{X}^b\mathbf{T})\,(\mathbf{X}^b\mathbf{T})^T = \mathbf{P}^a = (\mathbf{I}\text{-}\mathbf{KH})\mathbf{P}^b$$

- Note that the choice of $\mathbf{T}$ is not unique: Any orthonormal transformation $\mathbf{Q}$ $(\mathbf{Q}\mathbf{Q}^T=\mathbf{Q}^T\mathbf{Q}=\mathbf{I})$ can be applied to $\mathbf{T}$ and give another 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; EAF, Anderson, 2001;…)

Ensemble Kalman Filters

- We might want to ask the questions:
    1. How good is the EnKF for state estimation?
    2. How does it compare with 4D-Var?

Ensemble Kalman Filters

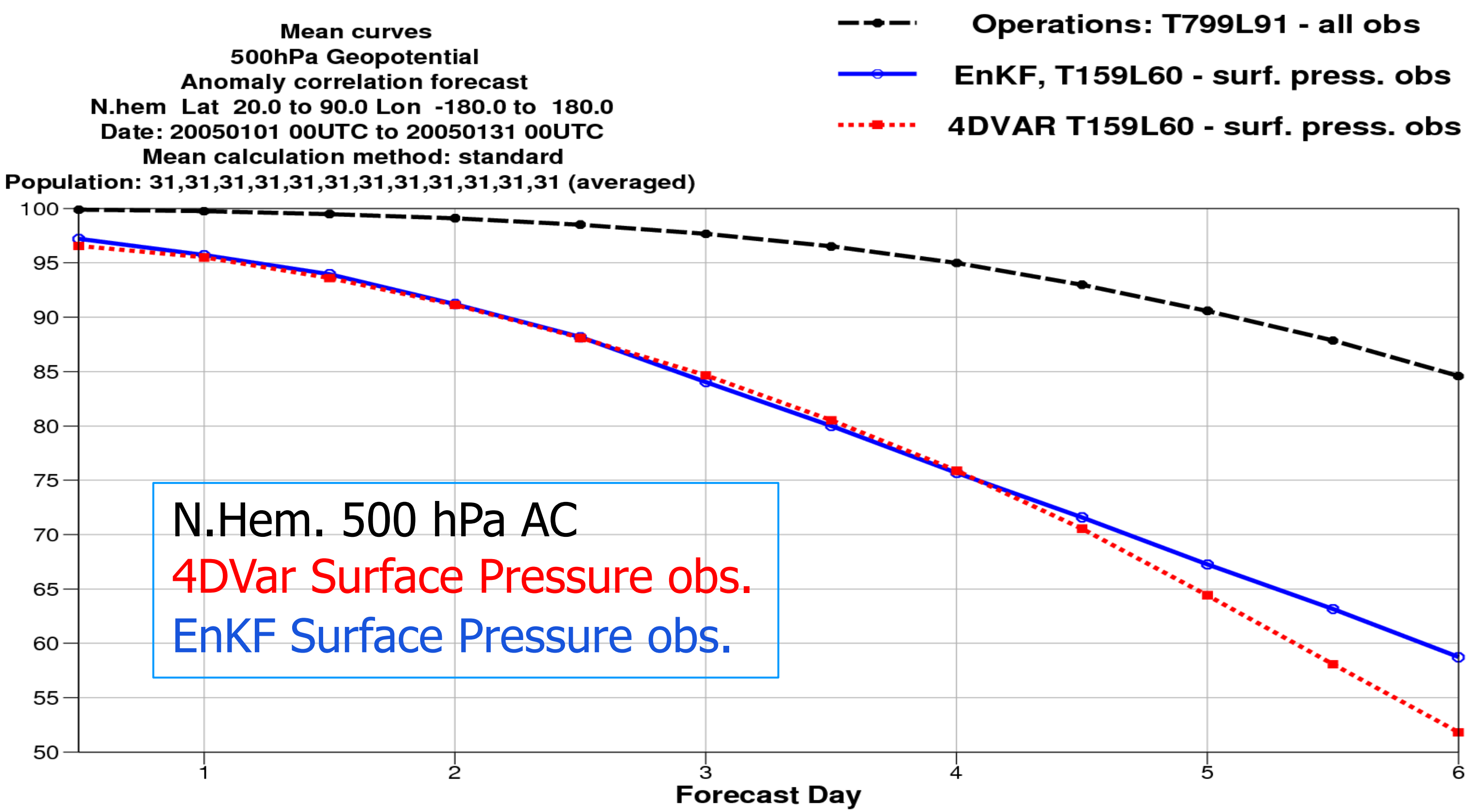
- We might want to ask the questions:
  1. How good is the EnKF for state estimation?
  2. How does it compare with 4D-Var?

- The short answer: it depends…

(more detailed answers in Hamrud et al, 2015; Bonavita et al, 2015; Bonavita et al, 2020)

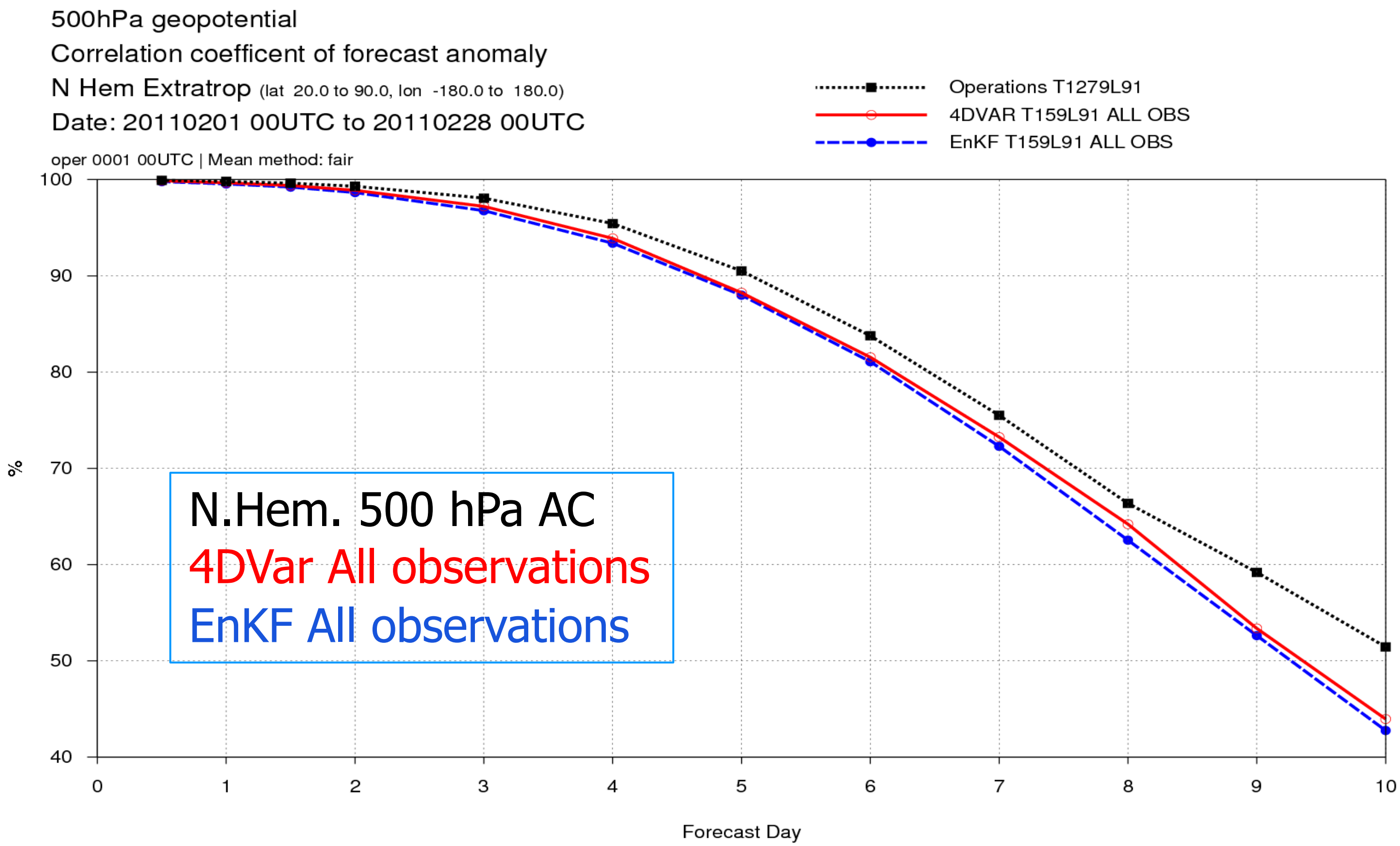
Ensemble Kalman Filters

- For sparsely observed systems the EnKF works quite well:



Ensemble Kalman Filters

- For densely observed systems the EnKF works not quite as well as 4D-Var:



Ensemble Kalman Filters

- Advantages:
  1. Background/Analysis error estimates reflect state of the flow
  2. Provides an ensemble of analyses: can use to directly initialise ensemble prediction
  3. Competitive with 4D-Var for sparsely observed systems (eg early period reanalysis, Ocean/Land DA, etc.)
  4. Excellent scalability properties
  5. Relative ease of coding and maintenance (No TL and ADJ models!)

Ensemble Kalman Filters

- Disadvantages:

  1. The affordable ensembles are relatively small (O(100)), thus sampling noise and rank deficiency of the sampled error covariances become a performance limiting factor for the EnKF (Stop the press! ML models are very cheap to run, could make order of magnitudes larger ensemble possible…)

  2. Careful localization of sampled covariances becomes necessary: This is an on-going research topic for both EnKF and Ensemble Variational systems. Note that localisation reduces amount of information that can be extracted from observations

  3. Vertical covariance localization becomes conceptually and practically more difficult for observations (e.g., satellite radiances) which are non-local, i.e. they sample a layer of the atmosphere (Campbell *et al.*, 2010).

Ensemble Kalman Filters

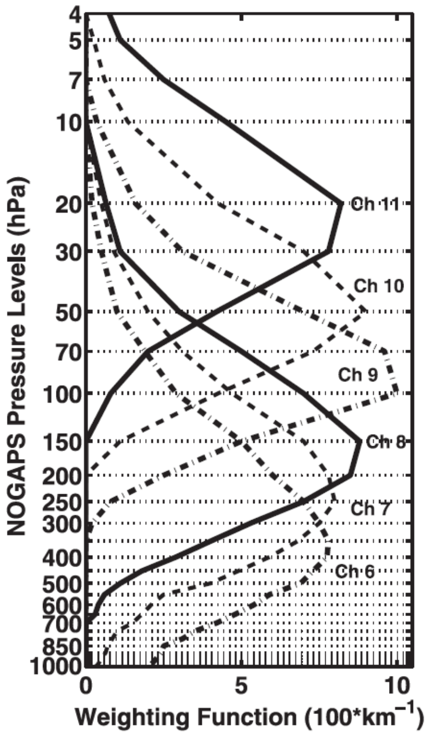


FIG. 1. AMSU-A weighting functions for channels 6–11 projected onto the 30 levels of NOGAPS.

from Campbell et al., 2010

- It is more effective (and mathematically consistent) to do model space localisation in the vertical for nonlocal observations like those from satellite sounders (Campbell *et al.*, 2010).
- Model space localisation can be implemented also in LETKF type of EnKF, using an expanded ensemble in the analysis step ("modulated ensemble". Whitaker, 2016, Bishop et al., 2017). This is more computationally expensive (larger ensemble) but the additional cost can be absorbed by updating the whole vertical column at once
- Further details in Lei et al, 2018 and references therein; Farchi and Bocquet, 2019

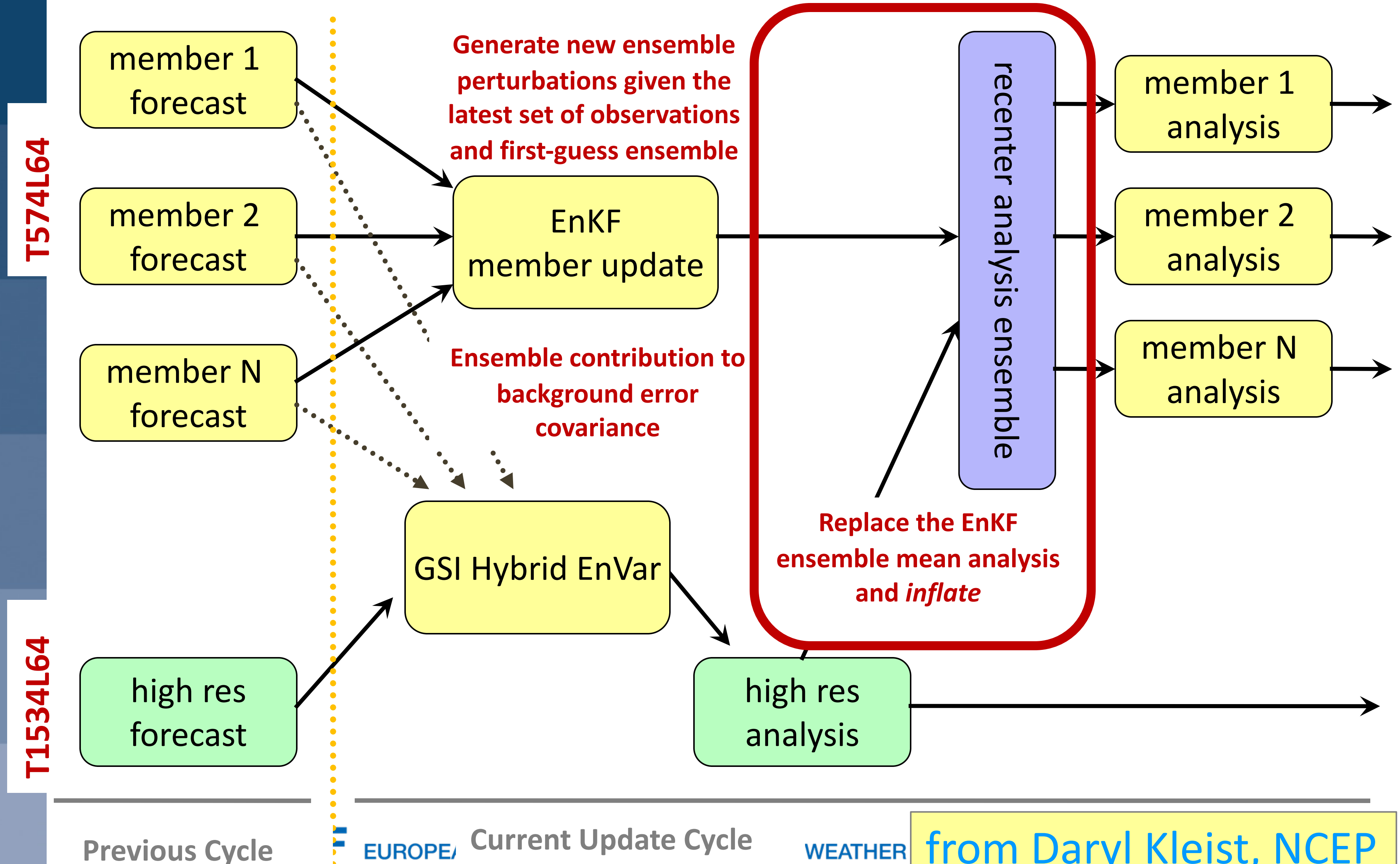
Ensemble Kalman Filters

- Minuses:
  4. EnKF produces linear analysis updates. However high-resolution models and new cloud/precip. sensitive observations (e.g., rain radar, cloud/rain affected radiances, etc) are increasingly nonlinear. What to do?
     1. Iterate the analysis, similar to what incremental 4D-Var does (e.g., Iterative EnKF, Sakov, 2012; Sakov et al., 2018). Conceptually easy but computationally expensive;
     2. Extend the Gaussian framework of the EnKF to classes of non-Gaussian pdfs. This can be done in a variety of ways, e.g. Gaussian Mixture models (Andersson and Andersson, 1999; Bengtsson et al. 2003; Hoteit et al. 2008, 2012; Stordal et al. 2011; Frei and Künsch, 2013), GIGGS Filter (Bishop, 2016); expansion to higher orders in the innovations (Hodyss, 2011);
     3. A combination of 1 and 2 (e.g., Posselt and Bishop, 2018);
     4. Rank Histogram Filters (Andersson, 2010, 2020; Metref et al, 2014)
     5. Employ some combination/hybrid of EnKF and Particle Filter (see for example: Van Leeuwen, Y. Cheng and S. Reich, 2015; Carrassi et al, 2017; Google this for even more recent results!)
     6. Let the EnKF give up gracefully in presence of increasing nonlinearity (Bonavita, Geer and Hamrud, 2020)
     7. Increase frequency of analysis updates (6h->3h->1h->…)

Ensemble Kalman Filters in hybrid DA

- While the pure EnKF is not currently competitive with variational methods for state estimation in global NWP, its good scalability properties and ease of maintenance make it a popular choice as a Monte Carlo system to estimate and cycle the error covariances ($P^{a/b}$) needed in a variational analysis system and to initialise an ensemble prediction system: hybrid Variational-EnKF analysis systems (NCEP, CMC, UKMO, JMA)

Dual-Res Coupled Hybrid Var/EnKF Cycling



EUROPE WEATHER

from Daryl Kleist, NCEP

Summary

- For linear model $\mathbf{M}$ and the observation operators $\mathbf{H}$ the Kalman Filter produces an optimal (minimum error variance) sequence of analyses ($\mathbf{x}^a{}_1$, $\mathbf{x}^a{}_2$, …, $\mathbf{x}^a{}_{t-1}$, $\mathbf{x}^a{}_t$)

- Under the additional assumption of Gaussian errors the Kalman Filter provides the exact posterior probability estimate, p($\mathbf{x}^a{}_t$| $\mathbf{x}^b{}_0$; $\mathbf{y}_0$,$\mathbf{y}_1$,…,$\mathbf{y}_t$).

- Kalman Filters are impractical for large-dimensional systems like in NWP, due to the impossibility of storing and evolving the state error covariance matrices ($\mathbf{P}^{a/b}$)

- We need to use reduced-rank representations of the state error covariance matrices: this can be done, but has other drawbacks (need for localisation, physical imbalances, etc.)

- The Ensemble Kalman Filter is a Monte Carlo implementation of the reduced-rank Kalman Filter. It works well for sparsely observed systems, but for well observed systems the severe rank reduction can be a performance limiting factor

- The EnKF (and its variants) are currently used in many global NWP Centres as the error cycling component of a hybrid Variational-EnKF system

Bibliography

1. Anderson, J. L., and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. Mon. Wea. Rev., 127, 2741–2758, doi:https://doi.org/10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2
2. Anderson, J. L., 2001. An ensemble adjustment Kalman filter for data assimilation. Mon. Wea. Rev. 129, 2884–2903.
3. Anderson, J. L.: A non-Gaussian ensemble filter update for data assimilation, Mon. Weather Rev., 138, 4186–4198, 2010.
4. Anderson, J. L., 2020: A Marginal Adjustment Rank Histogram Filter for Non-Gaussian Ensemble Data Assimilation. *Mon. Wea. Rev.*, **148**, 3361–3378, https://doi.org/10.1175/MWR-D-19-0307.1.
5. Bengtsson, T., C. Snyder, and D. Nychka, 2003: Toward a nonlinear ensemble filter for high-dimensional systems. J. Geophys. Res., 108, 8775, doi:https://doi.org/10.1029/2002JD002900.
6. 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.
7. Bishop, C.H., 2016: The GIGG-EnKF: ensemble Kalman filtering for highly skewed non-negative uncertainty distributions. Q. J. R. Meteorol. Soc., 142, 1395–1412.
8. Bishop, C. H., J. Whiatker and L. Lei, 2017: Gain Form of the Ensemble Transform Kalman Filter and Its Relevance to Satellite Data Assimilation with Model Space Ensemble Covariance Localization. Mon, Wea. Rev., https://doi.org/10.1175/MWR-D-17-0102.1
9. Bromiley, P.A., 2014: Products and Convolutions of Gaussian Probability Density Functions, TINA Memo N. 2003-003. Available at http://www.tina-vision.net/docs/memos/2003-003.pdf
10. 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.
11. 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.
12. Carrassi, A., Bocquet, M., Bertino, L., et al., 2017. Data assimilation in the geosciences—an overview on methods, issues and perspectives. arXiv:1709.02798.

Bibliography

13. 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.

14. Evensen, G . 2004 . Sampling strategies and square root analysis schemes for the EnKF . No.: 54 . p.: 539-560 . Ocean Dynamics

14. Farchi, A. and M. Bocquet, 2019: On the Efficiency of Covariance Localisation of the Ensemble Kalman Filter Using Augmented Ensembles. Front. Appl. Math. Stat. https://doi.org/10.3389/fams.2019.00003

15. 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.

16. Flowerdew, J. 2015. Towards a theory of optimal localisation. Tellus A 2015, 67, 25257, http://dx.doi.org/10.3402/tellusa.v67.25257

17. Frei, M., and H. R. Künsch, 2013: Mixture ensemble Kalman filters. Comput. Stat. Data Anal., 58, 127–138, doi:https://doi.org/10.1016/j.csda.2011.04.013.

18. Hodyss, D., 2011: Ensemble State Estimation for Nonlinear Systems Using Polynomial Expansions in the Innovation. Mon. Wea. Rev., 139, 3571–3588, https://doi.org/10.1175/2011MWR3558.1

19. Hoteit, I., X. Luo, and D.-T. Pham, 2012: Particle Kalman filtering: A nonlinear Bayesian framework for ensemble Kalman filters. Mon. Wea. Rev., 140, 528–542, doi:https://doi.org/10.1175/2011MWR3640.1.

20. Houtekamer, P. L. and Mitchell, H. L., 1998. Data assimilation using an ensemble Kalman filter technique. Mon. Wea. Rev. 126, 796–811.

21. Houtekamer, P. L. and Mitchell, H. L., 2001. A sequential ensemble Kalman filter for atmospheric data assimilation. Mon. Wea. Rev. 129, 123–137.

22. Houtekamer, P. L., & Zhang, F. (2016). Review of the Ensemble Kalman Filter for Atmospheric Data Assimilation, Monthly Weather Review, 144(12), 4489-4532.

23. 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.

Bibliography

24. Lei, L., Whitaker, J. S., & Bishop, C. (2018). Improving assimilation of radiance observations by implementing model space localization in an ensemble Kalman filter. Journal of Advances in Modeling Earth Systems, 10, 3221–3232. https://doi.org/10.1029/2018MS001468
25. 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.
26. 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.
27. Ménétrier, B., T. Montmerle, Y. Michel and L. Berre, 2015: Linear Filtering of Sample Covariances for Ensemble-Based Data Assimilation. Part I: Optimality Criteria and Application to Variance Filtering and Covariance Localization. Mon. Wea. Rev, 143, 1622–1643
28. Metref, S., E. Cosme, C. Snyder, and P. Brasseur, 2014: A non-Gaussian analysis scheme using rank histograms for ensemble data assimilation. Nonlinear Processes Geophys., 21, 869–885, doi:https://doi.org/10.5194/npg-21-869-2014.
29. Miyoshi, T., K. Kondo, and T. Imamura, 2014: The 10,240-member ensemble Kalman filtering with an intermediate AGCM, Geophys. Res. Lett., 41, 5264–5271, doi:10.1002/2014GL060863.
30. 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.
31. Periáñez, Á., Reich, H. and Potthast, R. 2014. Optimal localization for ensemble Kalman filter systems. J. Met. Soc. Japan. 62, 585 597.
32. Posselt DJ, Bishop CH. Nonlinear data assimilation for clouds and precipitation using a gamma inverse-gamma ensemble filter. Q J R Meteorol Soc. 2018;144:2331–2349. https://doi.org/10.1002/qj.3374
33. Sakov, P., D.S. Oliver, and L. Bertino, 2012: An Iterative EnKF for Strongly Nonlinear Systems. Mon. Wea. Rev., 140, 1988–2004, https://doi.org/10.1175/MWR-D-11-00176.1
34. Sakov, P., Haussaire, J.-M. and Bocquet, M., 2018: An iterative ensemble Kalman filter in the presence of additive model error. Q.J.R. Meteorol. Soc.. doi:10.1002/qj.3213

Bibliography

35. Stordal, A. S., H. A. Karlsen, G. Nævdal, H. J. Skaug, and B. Vallès, 2011: Bridging the ensemble Kalman filter and particle filters: The adaptive Gaussian mixture filter. Comput. Geosci., 15, 293–305, doi:https://doi.org/10.1007/s10596-010-9207-1.

36. 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

37. van Leeuwen, P. J., Y. Cheng, and S. Reich,Nonlinear data assimilation, vol. 2, Springer, 2015

38. Whitaker, J. S. and Hamill, T. M., 2002. Ensemble data assimilation without perturbed observations. Mon. Wea. Rev. 130, 1913–1924.

39. Wikle, C.K. and Berliner, L.M., 2007: A Bayesian tutorial for data assimilation. Physica D: Nonlinear Phenomena, vol. 230, no. 1-2, pp. 1–16