Kalman Filter and Ensemble Kalman Filter

Motivation

Ensemble forecasting : Provides flow-dependent estimate of *uncertainty of the forecast.*

Data assimilation : requires information about *uncertainty in prior forecast* and observations.

More accurate estimate of uncertainty, less error in the analysis. Improved initial conditions, improved ensemble forecasts.

Ensemble forecasting ⬅ ➡ data assimilation.

# Example: where flow-dependent first-guess errors help

The first guess

-5
0
5
10
15

The first guess

14
New Observation

-5
0
5
10
15

4
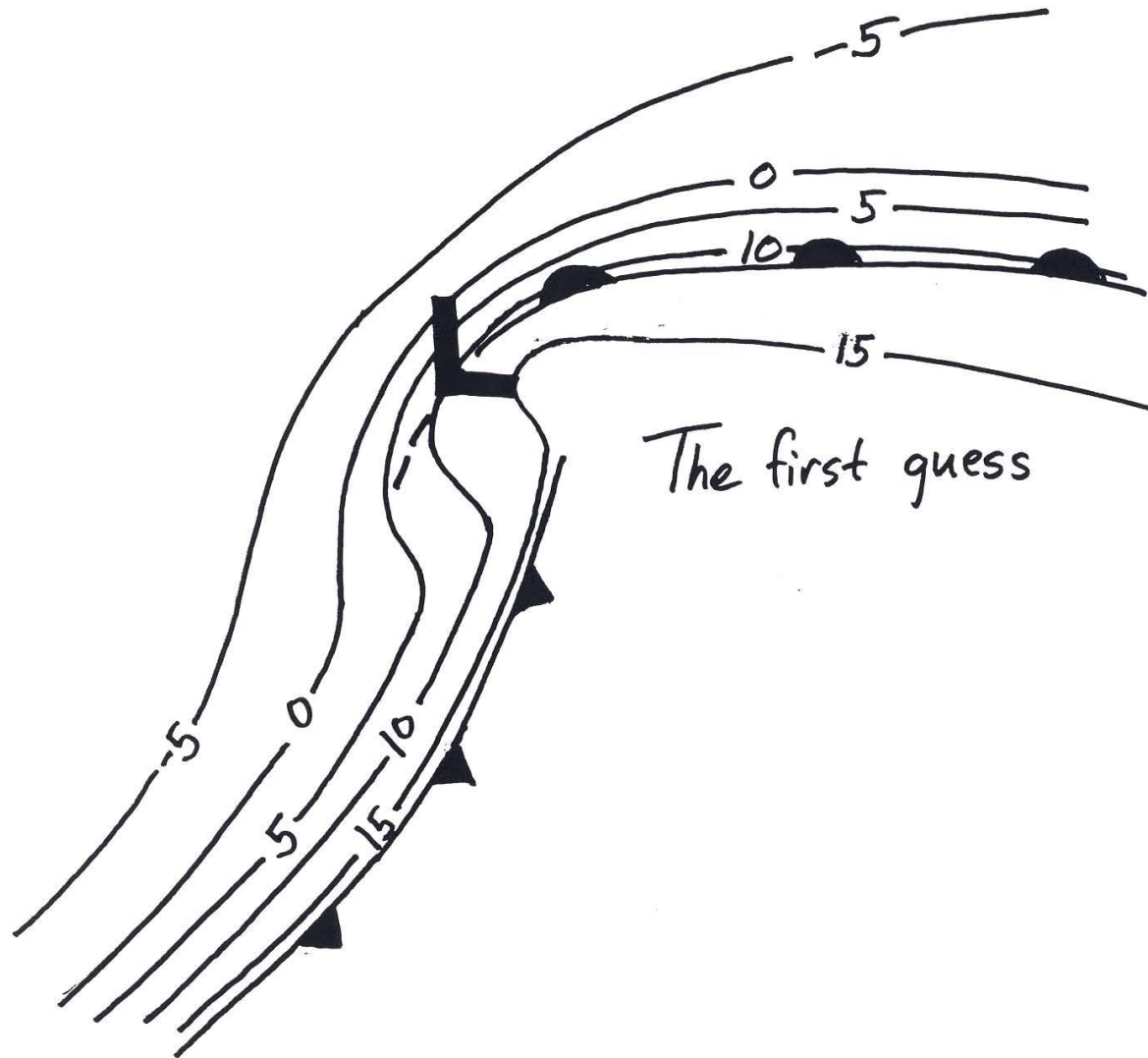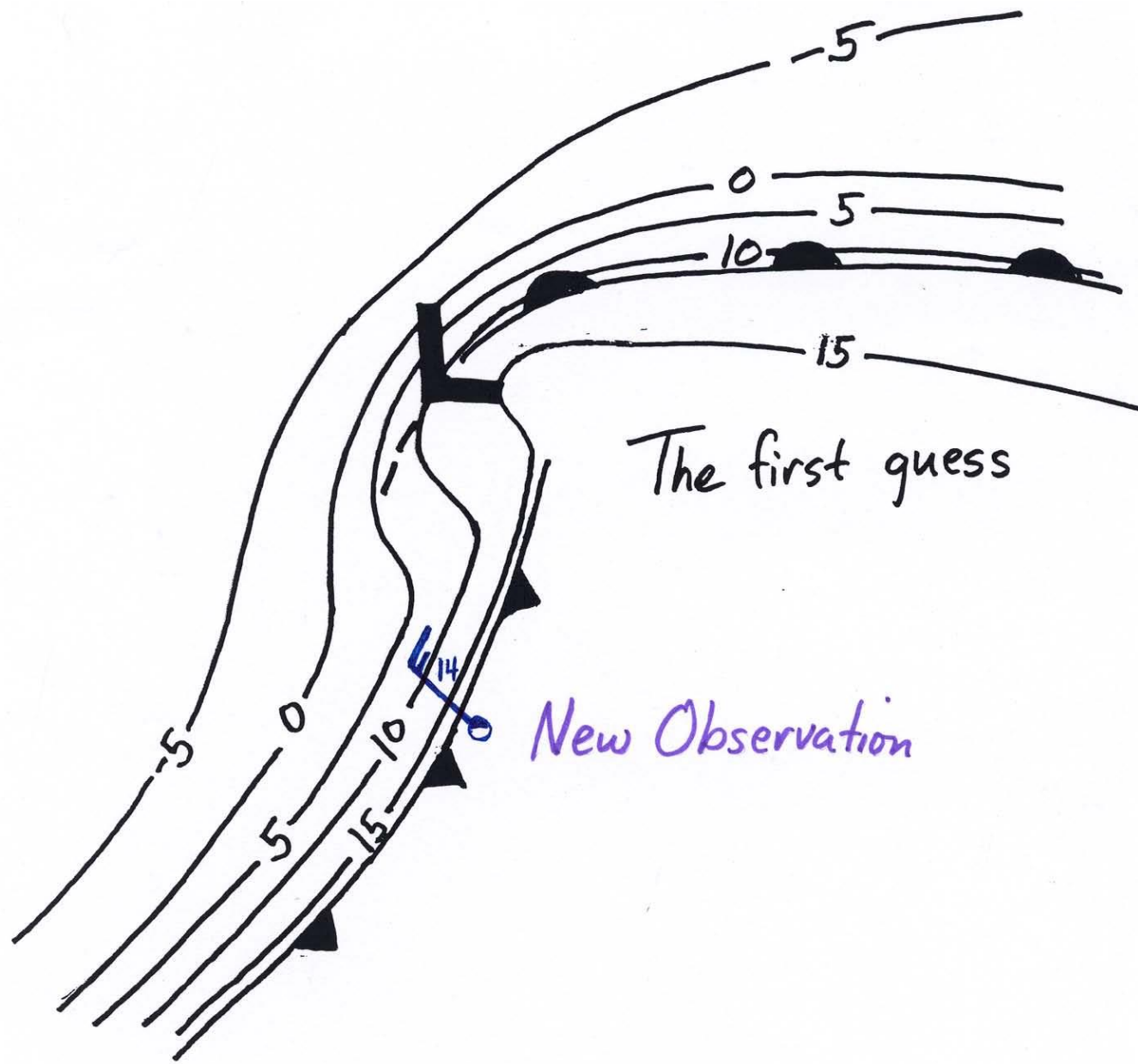
New observation is inconsistent with first guess 3D-Var produces an unappealing analysis, bowing the front out just in the region of the observation.

The way an observation influences surrounding points is always the same in 3D-Var --- typically concentric rings of decreasing influence the greater the distance from the observation.

-5

0

5

10

15

The first guess

-5

0

5

10

14

0

15

The Front You Want
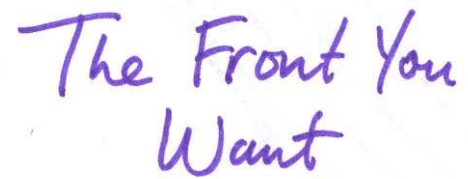
Kalman Filters

We have discussed the Optimal Interpolation, which minimizes the expected analysis error covariance, then the 3D-Var and PSAS methods, which solve essentially the same problem but minimizing a cost function.

In these methods, the forecast (or background) error covariance matrix is estimated once and for all, as if the forecast errors were statistically stationary.

Some research (e.g., Kistler et al, 2001) has shown, however, that there is large day-to-day variability in the model forecast error (with a time scale of a few days) and the variability *is about as large as the average error*. It points to the importance of the "errors of the day", which at the large scales are dominated by baroclinic instabilities of synoptic time scales, and which are ignored when the forecast error covariance is assumed to be constant.

In this section we give a brief introduction to more advanced (and much costlier) schemes that include, at least implicitly, the evolution of the forecast error covariance. A number of papers in Ghil et al (1997) provide more details about the theory and practice of some of these methods. Ide et al (1997) is a brief but extremely clear overview.

Kalman Filter and Extended Kalman Filter

Kalman Filter (KF) is formally very similar to Optimal Interpolation, but with one major difference: the forecast or background error covariance $\mathbf{P}^f(t_i)$ is advanced using the model itself, rather than estimating it as a constant covariance matrix $\mathbf{B}$.

As before, let

$$\mathbf{x}^f(t_i) = M_{i-1}\left[\mathbf{x}^a(t_{i-1})\right]$$

represent the (nonlinear) model forecast that advances from the previous analysis time t_{i-1} to the current t_i.

The model is imperfect; therefore, we assume that for the true atmosphere

$$\mathbf{x}^t(t_i) = M_{i-1}\left[\mathbf{x}^t(t_{i-1})\right] + \mathbf{\eta}(t_{i-1})$$

(6.1)

where $\mathbf{\eta}$ is a noise process with zero mean and covariance matrix $\mathbf{Q}_{i-1} = E\{\mathbf{\eta}_{i-1}\mathbf{\eta}_{i-1}^T\}$ (in other words, when starting from perfect initial conditions,

the forecast error is given by $-\mathbf{\eta}_{i-1}$, where the negative sign is chosen for convenience).

The original Kalman filter was developed for linear prediction model. When a nonlinear model M is used for the state variable prediction step, the filter is called Extended Kalman Filter (EKF).

Although we are assuming that the mean error is zero, in reality model errors have significant biases that should be taken into account. Dee and DaSilva (1997) show how to estimate and remove these model biases.

In the Extended Kalman Filter (EKF), the forecast error covariance is obtained linearizing the model about the nonlinear trajectory of the model between t_{i-1} and t_i, so that if we introduce a perturbation in the initial conditions, the final perturbation is given by

$$\mathbf{x}(t_i) + \delta\mathbf{x}(t_i) = M_{i-1}\left[\mathbf{x}(t_{i-1}) + \delta\mathbf{x}(t_{i-1})\right] = M_{i-1}\left[\mathbf{x}(t_{i-1})\right] + \mathbf{L}_{i-1}\delta\mathbf{x}(t_{i-1}) + O(|\delta\mathbf{x}|^2)$$

$$(6.2)$$

where $\mathbf{L}_{i-1}$ is the linear tangent model defined earlier.

As we did with OI and 3D-Var, observations are assumed to have random errors with zero mean and an observational error covariance matrix $\mathbf{R}_i = E\{\boldsymbol{\varepsilon}_i^o \boldsymbol{\varepsilon}_i^{oT}\}$, where

$$\mathbf{y}_i^o = H(\mathbf{x}^t(t_i)) + \boldsymbol{\varepsilon}_i^o \tag{6.3}$$

and H is the forward or observation operator.

Note that the forecast error over a forecast depends on the initial (analysis) error and on the errors introduced by the forecast model during that period:

$$\begin{aligned}
\boldsymbol{\varepsilon}_i^f &= M_{i-1}(\mathbf{x}_{i-1}^t) + \boldsymbol{\eta}_{i-1} - M_{i-1}(\mathbf{x}_{i-1}^a) \\
&= M_{i-1}(\mathbf{x}_{i-1}^a + \mathbf{x}_{i-1}^t - \mathbf{x}_{i-1}^a) + \boldsymbol{\eta}_{i-1} - M_{i-1}(\mathbf{x}_{i-1}^a) \qquad (6.4) \\
&\approx \mathbf{L}_{i-1}\boldsymbol{\varepsilon}_{i-1}^a + \boldsymbol{\eta}_{i-1}
\end{aligned}$$

where we neglected higher order terms.

The analysis and forecast error covariances are defined, as usual, from their corresponding errors at the appropriate time:

$$\mathbf{P}_i = E\{\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T\}. \tag{6.5}$$

Plugging $\boldsymbol{\varepsilon}_i^f$ defined in (6.4) into the following, we obtain the prediction equation for $\mathbf{P}_i^f$:

$$
\begin{aligned}
\mathbf{P}_i^f &= E\{\boldsymbol{\varepsilon}_i^f \boldsymbol{\varepsilon}_i^{fT}\} = E\{(\mathbf{L}_{i-1}\boldsymbol{\varepsilon}_{i-1}^a + \boldsymbol{\eta}_i)(\mathbf{L}_{i-1}\boldsymbol{\varepsilon}_{i-1}^a + \boldsymbol{\eta}_i)^T\} \\
&= E\{\mathbf{L}_{i-1}\boldsymbol{\varepsilon}_{i-1}^a \boldsymbol{\varepsilon}_{i-1}^{aT}\mathbf{L}_{i-1}^T\} + E\{\mathbf{L}_{i-1}\boldsymbol{\varepsilon}_{i-1}^a \boldsymbol{\eta}_i^T\} + E\{\boldsymbol{\eta}_i\boldsymbol{\varepsilon}_{i-1}^{aT}\mathbf{L}_{i-1}^T\} + E\{\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T\} \\
&= \mathbf{L}_{i-1}E\{\boldsymbol{\varepsilon}_{i-1}^a \boldsymbol{\varepsilon}_{i-1}^{aT}\}\mathbf{L}_{i-1}^T + E\{\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T\} \\
&= \mathbf{L}_{i-1}\mathbf{P}_{i-1}^a\mathbf{L}_{i-1}^T + \mathbf{Q}_i
\end{aligned}
$$

From these equations we can define EKF which consists of a "forecast step" that advances the forecast and the forecast error covariance, followed by an "analysis" or update step, a sequence analogous to OI. After the forecast step, an optimal weight matrix or Kalman gain matrix is calculated as in OI, and this matrix is used in the analysis step.

The forecast step is

$$\mathbf{x}^f(t_i) = M_{i-1}\left[\mathbf{x}^a(t_{i-1})\right],$$

$$\mathbf{P}^f(t_i) = \mathbf{L}_{i-1}\mathbf{P}^a(t_{i-1})\mathbf{L}^T_{i-1} + \mathbf{Q}(t_{i-1}).$$

(6.6)

The analysis step is written as in OI, with

$$\mathbf{x}^a(t_i) = \mathbf{x}^f(t_i) + \mathbf{K}_i\mathbf{d}_i,$$

$$\mathbf{P}^a(t_i) = (\mathbf{I} - \mathbf{K}_i\mathbf{H}_i)\mathbf{P}^f(t_i),$$

(6.7)

where

$$\mathbf{d}_i = \mathbf{y}_i^o - H(\mathbf{x}^f(t_i))$$

(6.8)

is the observational increment or innovation.

The formula for the Kalman gain or weight matrix is obtained by minimizing the analysis error covariance $\mathbf{P}_i^a$ and is given by the same formula derived for OI,

but with the constant background error covariance $\mathbf{B}$ replaced by the evolved forecast error covariance $\mathbf{P}^f(t_i)$:

$$\mathbf{K}_i = \mathbf{P}^f(t_i)\mathbf{H}_i^T(\mathbf{H}_i\mathbf{P}^f(t_i)\mathbf{H}^T + \mathbf{R}_i)^{-1} \tag{6.9}$$

Entended Kalman Filter – how good or useful is it?

- The EKF was the "gold standard" of data assimilation.

- Even if a system starts with a poor initial guess of the state of the atmosphere, the EKF may go through an initial transient period of a week or two, after which it should provide the best linear unbiased estimate (BLUE) of the state of the atmosphere *and* its error covariance.

- However, if the system is very unstable, and the observations are not frequent enough, it is possible for the linearization to become inaccurate, and the EKF may drift away from the true solution.
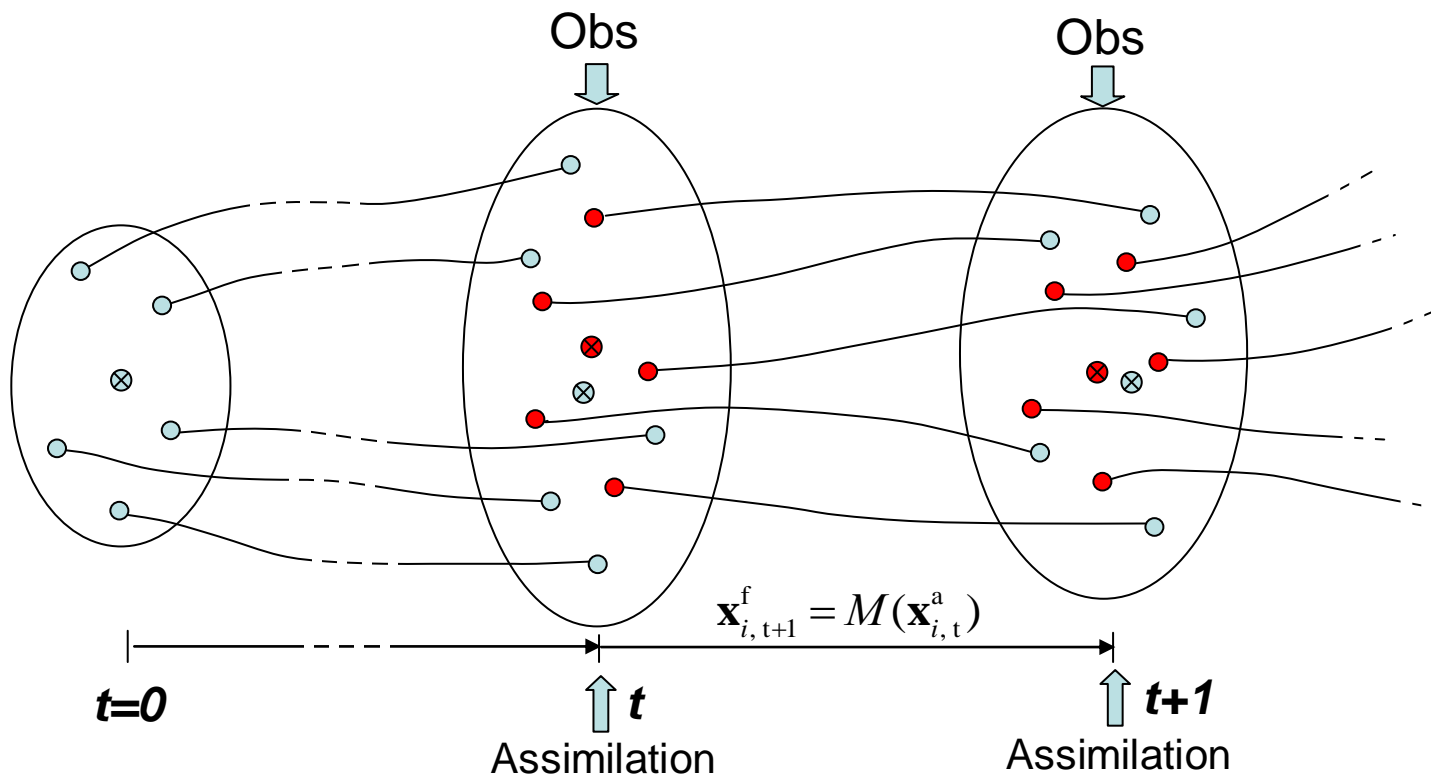
- The updating of the forecast error covariance matrix ensures that the analysis takes into account flow-dependent errors

- Unfortunately the EKF is exceedingly expensive, since the linear model matrix $\mathbf{L}_{i-1}$ has size n, the number of d.o.f. of a modern model (more than 10^6) and updating the error covariance is equivalent to performing $O(n)$ model integrations. For this reason, this step has been replaced by the use of simplifying assumptions (e.g., a lower order model and/or infrequent updating).

Reference:

Evensen, G., 1992: Using the extended Kalman filter with a multi-layer quasi-geostrophic ocean model. *J. Geophys. Res.*, **97(C11),** 17 905-17 924.

Ensemble Kalman Filtering

One promising simplification of Kalman Filtering is Ensemble Kalman Filtering (EnKF). In this approach, an ensemble of N data assimilation cycles is carried out simultaneously.

$$\mathbf{x}^{f}_{i, t+1} = M(\mathbf{x}^{a}_{i, t})$$

Obs   **Obs**

t=0   **t**   **t+1**

Assimilation   Assimilation

$\circ$ $\mathbf{x}^{f}_{i}$ $\mathbf{x}^{a}_{k} = \mathbf{x}^{f}_{k} + \mathbf{K}[\mathbf{y}^{o}_{k} - H(\mathbf{x}^{f}_{k})]$

$\bullet$ $\mathbf{x}^{a}_{i}$ $\mathbf{K} = \mathbf{P}^{f}\mathbf{H}^{T}[\mathbf{H}\mathbf{P}^{f}\mathbf{H}^{T} + \mathbf{R}]^{-1}$

Schematics of EnKF Cycles

With ensemble Kalman filter, the forecast error covariance $\mathbf{P}^f(t_i)$ is not predicted by the covariance equation (6.6), but estimated using a forecast ensemble.

After completing the ensemble of analyses at time t_{i-1}, and an ensemble of N forecasts $\mathbf{x}_k^f(t_i) = M_{i-1}^k(\mathbf{x}_k^a(t_{i-1}))$, one can obtain an estimate of the forecast error covariance from the K forecasts $\mathbf{x}_k^f(t_i)$, according to

$$\mathbf{P}^f \approx \frac{1}{K-1}\sum_{k=1}^{K}\left(\mathbf{x}_k^f - \overline{\mathbf{x}}^f\right)\left(\mathbf{x}_k^f - \overline{\mathbf{x}}^f\right)^T,$$

(6.10)

where the overbar represents the ensemble average.

Remark 1: Note that according to the definition of error covariance, $\overline{\mathbf{x}}^f$ in the above equation really should have been $\mathbf{x}^t$, i.e., the true state vector. Since true state is never known, we are using $\overline{\mathbf{x}}^f$ as it's best estimate – this assumption is a source of error.

Remark 2: Equation (6.10) tends to underestimate the variance of the forecast errors because every forecast is used to compute the estimate of its own error

covariance. Houtekamer and Mitchell (1998) and Hamill and Snyder (2000) suggest instead to compute the forecast error covariance for ensemble member l from an ensemble that excludes the forecast l:

$$\mathbf{P}_l^f \approx \frac{1}{N-2} \sum_{k \neq l} \left(\mathbf{x}_k^f - \bar{\mathbf{x}}_l^f \right) \left(\mathbf{x}_k^f - \bar{\mathbf{x}}_l^f \right)^T \qquad (6.11)$$

Hamill and Snyder (2000) also suggest a hybrid between 3D-Var and EnKF, where the forecast error covariance is obtained from a linear combination of the (constant) 3D-Var covariance $\mathbf{B}_{3DVAR}$:

$$\mathbf{P}_l^{f\,(hybrid)} = (1-\alpha)\mathbf{P}_l^f + \alpha\mathbf{B}_{3DVAR} \qquad (6.12)$$

where α is a tunable parameter that varies from 0, pure EnKF from (6.11) to 1, pure 3D-Var.

In practice, we don't directly evaluate $\mathbf{P}^f$, but rather evaluate the following two terms that are needed by the Kalman gain $\mathbf{K}$:

$$\mathbf{P}^f \mathbf{H}^\mathrm{T} = \frac{1}{K-1} \sum_{k=1}^{K} [\mathbf{x}_k^f - \overline{\mathbf{x}}^f][H(\mathbf{x}_k^f) - \overline{H(\mathbf{x}^f)}]^\mathrm{T}, \tag{6.13}$$

$$\mathbf{H}\mathbf{P}^f \mathbf{H}^\mathrm{T} = \frac{1}{K-1} \sum_{k=1}^{K} [H(\mathbf{x}_k^f) - \overline{H(\mathbf{x}^f)}][H(\mathbf{x}_k^f) - \overline{H(\mathbf{x}^f)}]^\mathrm{T}, \tag{6.14}$$

After the above two covariances are obtained, they are plugged into the optimal solution, as given in (6.7) to obtain an analysis for **x**. In fact, this analysis is performed for every member of the ensemble. They are several variations to the actually implementation, for the analysis of individual members. Here we will discuss two commonly used algorithms, the classical EnKF algorithm and the ensemble square-root filter (EnSRF) algorithm.

Classical or stochastic EnKF algorithm with perturbed observations
(Evensen 1994, 2003)

As shown by Burgers et al. (1998) and Whitaker and Hamill (2002), in order for the analysis error covariance $\mathbf{P}^a$ to have correct values, special measures have to be taken.

With the traditional Kalman filter update equation,

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}[\mathbf{y}^o - H(\mathbf{x}^f)] \qquad (6.14a)$$

where

$$\mathbf{K} = \mathbf{P}^f \mathbf{H}^T [\mathbf{H}\mathbf{P}^f \mathbf{H}^T + \mathbf{R}]^{-1}, \qquad (6.14b)$$

As we saw in the OI section, based on definition $\mathbf{P}^a = E\{(\mathbf{x}^a - \mathbf{x}^t)(\mathbf{x}^a - \mathbf{x}^t)^T\}$, the analysis error covariance is (replacing $\mathbf{W}$ there with $\mathbf{K}$ here)

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^f(\mathbf{I} - \mathbf{K}\mathbf{H})^T + \mathbf{K}\mathbf{R}\mathbf{K}^T = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^f \qquad (6.14c)$$

where $\mathbf{K} = \mathbf{P}^f \mathbf{H}^T [\mathbf{H}\mathbf{P}^f \mathbf{H}^T + \mathbf{R}]^{-1}$ was used.

In the EnKF, $\mathbf{P}^f$ is approximated using the sample covariance from an ensemble of model forecasts. Hereafter, the symbol $\mathbf{P}$ is used to denote the sample covariance calculated from the ensemble according to Eq.(6.10), and $\mathbf{K}$ is understood to be computed using sample covariances. Expressing the variables as an ensemble mean (denoted by an overbar) and a deviation from the mean (denoted by a prime), the update equations for the EnKF may be written as

$$\overline{\mathbf{x}}^a = \overline{\mathbf{x}}^f + \mathbf{K}[\overline{\mathbf{y}}^o - H(\overline{\mathbf{x}}^f)], \qquad\qquad (6.14d)$$

$$\mathbf{x}'^a_k = \mathbf{x}'^f_k + \tilde{\mathbf{K}}[\mathbf{y}'^o_k - H(\mathbf{x}'^f_k)]. \qquad\qquad (6.14e)$$

The above is the result of applying the standard Kalman filter equation to each of the ensemble member, and assuming that the observations used by each ensemble member has been perturbed by adding $\mathbf{y}'^o_k$. Also, for the traditional EnKF algorithm $\tilde{\mathbf{K}} = \mathbf{K}$.

In the above, if all members are updated using the same observations ($\mathbf{y}''^o_k$), then the covariance of the analyzed ensemble can be shown to be

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{KH})\mathbf{P}^f(\mathbf{I} - \mathbf{KH})^T. \qquad\qquad (6.14f)$$

Compared to the correct $\mathbf{P}^a$ derived earlier, $\mathbf{KRK}^T$ term is missing, causing $\mathbf{P}^a$ to be systematically under-estimated.

If random noise is added to the observations so that $\mathbf{y}''^o_k \neq 0$ then the analyzed ensemble covariance is

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{KH})\mathbf{P}^f(\mathbf{I} - \mathbf{KH})^T + \mathbf{K}\left[E\{\mathbf{y}''^o \mathbf{y}''^{oT}\} - E\{\mathbf{Hx}'^f \mathbf{y}''^{oT}\} - E\{\mathbf{y}''^o \mathbf{x}'^{fT}\mathbf{H}^T\} \right]\mathbf{K}^T$$

$$+ E\{\mathbf{x}'^f \mathbf{y}''^{oT}\}\mathbf{K}^T + \mathbf{K}E\{\mathbf{y}''^o \mathbf{x}'^{fT}\}$$

If the added observation noise is defined such that

$$E\{\mathbf{y}''^o \mathbf{y}''^{oT}\} = \mathbf{R}$$

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{KH})\mathbf{P}^f (\mathbf{I} - \mathbf{KH})^T + \mathbf{KRK}^T \qquad\qquad (6.14\text{g})$$

the same as the correct $\mathbf{P}^a$ given in (6.14c) without missing the $\mathbf{KRK}^T$ term. The above assumes that the background perturbations are uncorrelated with the added observation noise.

The above is exactly the same reason for adding perturbations to the observations, as originally pointed out by Burgers et al. (1998).

In summary, to avoid under-estimating the analysis covariance $\mathbf{P}^a$, we add different sets of *random perturbations to the observations* that are assimilated in each member of the ensemble, and this algorithm is often call **the perturbed observation or the stochastic EnKF method**. The equations can also be written as follows. To keep things more general, we use superscript b instead of f in the equations.

For each ensemble, k, the analysis equation is

$$\mathbf{x}_k^a = \mathbf{x}_k^b + \mathbf{P}^b\mathbf{H}^T[\mathbf{HP}^b\mathbf{H}^T + \mathbf{R}]^{-1}[\mathbf{y}_k^o - H(\mathbf{x}_k^b)], \qquad (6.15)$$

where

$$\mathbf{y}_k^o = \mathbf{y}^o + \varepsilon_k, \tag{6.16}$$

are the perturbed sets of observations, with the perturbations ε_k drawn from a normal distribution of covariance $\mathbf{R}$, i.e.,

$$\left\langle \varepsilon \varepsilon^T \right\rangle = \mathbf{R}.$$

In practice, when the observations are uncorrelated, the observations can be assimilated serially, or one at a time. In such as case, the covariances in (6.13) and (6.14) are recalculated after the assimilation of each observation, and the new analysis becomes the background of the next analysis. In this sense, superscript b is more general than f in the above equations. When we are dealing with a single observation, $\mathbf{R}$ becomes a scalar, so does $\mathbf{HP}^b\mathbf{H}^T$, therefore this is no matrix inversion involved when evaluating the gain matrix.

Tong and Xue (2005) is an example of using this perturbed observation method for assimilating radar data.

Ensemble Square-root Kalman Filter Algorithm

Whitaker and Hamill (2002) pointed out that the stochastic algorithm is subject to sampling error. A 'deterministic algorithm' not involving perturbing the observations is preferred.

There is actually a class of algorithms that do not involve perturbation the observations, the include the ensemble square root filter (EnSRF, Whitaker and Hamill 2002), ensemble adjustment filter (EAKF, Anderson 2001) and ensemble transform Kalman filter (ETKF, Bishop et al. 2001). These algorithms are discussed together in Tippett et al. (2003). Not perturbing observations avoid sampling errors introduced by the addition of observation perturbations.

Whitaker and Hamill (2002) proposed the ensemble square-root Kalman filter (EnSRF) algorithm that does not require perturbing observations. It is also a serial algorithm that analyzes uncorrelated observations, one after another.

Without perturbing the observations, the earlier perturbation updating equation becomes

$$\mathbf{x}_k^{'a} = \mathbf{x}_k^{'b} - \tilde{\mathbf{K}}\mathbf{H}\mathbf{x}_k^{'b} = (\mathbf{I} - \tilde{\mathbf{K}}\mathbf{H})\mathbf{x}_k^{'b} \qquad (6.16a)$$

We seek a definition for that will result in an ensemble whose analysis error covariance satisfies the correct equation for $\mathbf{P}^a$ as given in (6.14c). The equation used to solve for $\tilde{\mathbf{K}}$ is therefore

$$(\mathbf{I} - \tilde{\mathbf{K}}\mathbf{H})\mathbf{P}^b(\mathbf{I} - \tilde{\mathbf{K}}\mathbf{H})^T = (\mathbf{I} - \tilde{\mathbf{K}}\mathbf{H})\mathbf{P}^b \qquad (6.16b)$$

which has a solution

$$\tilde{\mathbf{K}} = \mathbf{P}^b\mathbf{H}^T\left[(\sqrt{\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R}})^{-1}\right]^T\left[\sqrt{\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R}} + \sqrt{\mathbf{R}}\right]^{-1}. \qquad (6.16c)$$

The above algorithm is actually a Monte Carlo implementation of a square root filter (Maybeck 1979); the algorithm is therefore called the square root algorithm or EnSRF. The above equation involves finding the square root and

inversion of matrixes show dimension is defined by the number of observations so can be expensive for a large number of observations. We again can employ the serial processing strategy where observations processed one by one, assuming **R** is diagonal.

For a single observation $\mathbf{HP}^b\mathbf{H}^T$ and **R** reduce to scalars, and Eq.(6.16b) can be written as

$$\frac{\mathbf{HP}^b\mathbf{H}^T}{\mathbf{HP}^b\mathbf{H}^T + \mathbf{R}}\tilde{\mathbf{K}}\tilde{\mathbf{K}}^T - \mathbf{K}\tilde{\mathbf{K}}^T - \tilde{\mathbf{K}}\mathbf{K}^T + \mathbf{K}\mathbf{K}^T = 0. \qquad (6.16d)$$

If $\tilde{\mathbf{K}} = \alpha\mathbf{K}$, where α is a constant, then $\mathbf{K}\mathbf{K}^T$ can be factored out of the above equation, resulting in a scalar quadratic equation for α,

$$\frac{\mathbf{HP}^b\mathbf{H}^T}{\mathbf{HP}^b\mathbf{H}^T + \mathbf{R}}\alpha^2 - 2\alpha + 1 = 0. \qquad (6.16e)$$

Since we want the deviations from the ensemble mean to be reduced in magnitude while maintaining the same sign, we choose the solution to Eq. (12) that is between 0 and 1. This solution is

$$\alpha = \left[1 + \sqrt{\mathbf{R}(\mathbf{HP}^b\mathbf{H}^T + \mathbf{R})^{-1}} \right]^{-1}.$$

Here, $\mathbf{HP}^b\mathbf{H}^T$ and $\mathbf{R}$ are scalars representing the background and observational-error variance at the observation location. Using $\tilde{\mathbf{K}}$ to update deviations from the mean, the analysis-error covariance is guaranteed to be exactly equal to Eq. (6.14c).

In summary, with the EnSRF algorithm, the analysis equations for ensemble mean state, $\bar{\mathbf{x}}$, and the ensemble deviation from the mean, $\mathbf{x}'_k$, are, respectively:

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^b + \mathbf{K}[\mathbf{y}^o - H(\bar{\mathbf{x}}^b)], \tag{6.17}$$

$$\mathbf{x}'^a_k = \beta(\mathbf{I} - \alpha\mathbf{KH})\mathbf{x}'^b_k, \tag{6.18}$$

where

$$\mathbf{K} = \rho \circ \mathbf{P}^b \mathbf{H}^T (\rho \circ \mathbf{H}\mathbf{P}^b \mathbf{H}^T + \mathbf{R})^{-1} \tag{6.19}$$

is the regular Kalman gain matrix, $\mathbf{P}^b$ is the background or prior error covariance matrix.

In Eq.(6.18), β is a **covariance inflation factor** that will be discussed later.

In Eq.(6.19), ρ is the **localization coefficient factor** which is often defined as function of distance between grid points (for $\mathbf{P}^b$) or between the observation and the grid point being updated (for $\mathbf{P}^b\mathbf{H}^T$). Symbol $\circ$ represents the Schur product or element-wise product of matrices. The localization effectively modified the original $\mathbf{P}$ calculated from the ensemble; usually more distance covariance is reduced to a smaller value or zeroed out completely (see Houtekamer and Mitchell 2001).

Eq. (4.10) of Gaspari and Cohn (1999) is most often used to define ρ.

Gaspari and Cohn (1999) give a number of examples of correlation functions with local support. Many studies define ρ as a compactly supported fifth order piecewise rational function, as given by their Eq. (4.10).

Gaspari, G. and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125,** 723-757.

The function is isotropic and decreases monotonically with distance at a rate that depends on a single length-scale parameter (denoted c by Gaspari and Cohn). As shown in Fig. 6 of the Gaspari and Cohn paper, the form of ρ is very similar to that of a Gaussian (i.e., negative-squared exponential) function.

Multiplication of the covariances calculated from the ensemble by ρ has several effects. Since ρ has compact support, it filters out the small (and noisy) correlations associated with remote observations. This localization strategy greatly improves the conditioning of the matrices $P^b H^T$ and $H P^b H^T$.

In addition, since ρ is smooth and monotonically decreasing, the Schur product tends to reduce and smooth the effect of those observations at intermediate

distances. The result is smooth analysis increments, in contrast to those produced by algorithms which use a cutoff radius.

The ensemble mean analysis, $\overline{\mathbf{x}}^a$, is obtained first from Eq. (6.17), the deviation from the mean by the kth ensemble member is then given be Eq. (6.18), in which β is a covariance inflation factor that is usually slightly larger than 1, and,

$$\alpha = \left[1 + \sqrt{\mathbf{R}(\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R})^{-1}} \right]^{-1}. \tag{6.20}$$

Equation (6.20) is only valid for single observation analysis and therefore both the numerator and denominator inside the square root are scalars and the evaluation of α is easy.

As before, the background error covariances $\mathbf{P}^b\mathbf{H}^T$ and $\mathbf{H}\mathbf{P}^b\mathbf{H}^T$ are estimated from the background ensemble, according to

$$\mathbf{P}^b\mathbf{H}^T = \frac{1}{N-1}\sum_{k}^{N}(\mathbf{x}_k^b - \overline{\mathbf{x}}^b)(H(\mathbf{x}_k^b) - \overline{H(\mathbf{x}^b)})^T, \tag{6.21}$$

$$\mathbf{HP}^{b}\mathbf{H}^{\mathrm{T}} = \frac{1}{N-1}\sum_{k}^{N}(H(\mathbf{x}_{k}^{b}) - \overline{H(\mathbf{x}^{b})})(H(\mathbf{x}_{k}^{b}) - \overline{H(\mathbf{x}^{b})})^{\mathrm{T}}, \qquad (6.22)$$

where N is the ensemble size, H is the observation operator which can be nonlinear (and is in this study).

For a single observation, $\mathbf{P}^{b}\mathbf{H}^{\mathrm{T}}$ is a vector having the length of vector x and $\mathbf{HP}^{b}\mathbf{H}^{\mathrm{T}}$ is a scalar. In practice, because of covariance localization, all elements in $\mathbf{P}^{b}\mathbf{H}^{\mathrm{T}}$ are not calculated; those outside the influence range of a given observation are assumed to be zero. After the analysis of one observation is completed, the analysis becomes the new background ($\mathbf{x}^{a}$ becomes $\mathbf{x}^{b}$) for the next observation and the analysis is repeated. After all observations at a given time are analyzed, an ensemble of forecasts proceeds from the analysis ensemble until the time of new observation(s), at that time the analysis cycle is repeated.

Rank Deficiency Problem

In (6.11) the EnKF covariance is estimated from only a limited sample of ensemble members $K-1$ compared to a much larger number of d.o.f of the

model, it is therefore rank deficient (which basically means that the matrix does not contain smaller pieces of independent information than the order of the matrix).

The combination with the 3D-Var, computed from many estimated forecast errors (using for example the method of Parrish and Derber, 1992) may ameliorate this sampling problem and "fill out" the error covariance.

In the experiments of Hamill and Snyder (2000) the best results were obtained for low values of α, between 0.1 and 0.4, indicating good impact of the use of the ensemble-evolved forecast error covariance. They found that 25-50 ensemble members were enough to provide the benefit of EnKF (but this may be different when using a more complex model than the quasi-geostrophic model used here).

Another way of ameliorating the problem associated with rank deficiency is to modify the estimated covariance, by reducing remote covariances to zero. This is usually achieved by multiplying the estimated covariance by a factor that decreases with distance from the observation. This is usually called 'covariance localization', which is found to increase performance of EnKF with the typically ensemble sizes.

Covariance Inflation

Because of the typically relatively smaller ensemble size, the covariance estimated from the ensemble is usually smaller that the true covariance (this shows up as the under-dispersion of the ensemble or the spread being too small). In practice, the estimated covariance usually needs to be artificially increased, by a few to tens of percent. Such procedures are usually called covariance inflation.

Difference covariance inflation methods.

The advantages of EnKF approach

a) K is of the order of 10-100, so that the computational cost (compared to OI or 3D-Var) is increased by a 10-100. Although this increased cost may seem large, it is small compared to EKF, which requires a cost increase of the order of the number of d.o.f. of the model.

b) EnKF does not require the development of a tangent linear and adjoint model.

c)   It does not require linearizing the evolution of the forecast error covariance.

d)   It may provide excellent initial perturbations for ensemble forecasting. Despite these advantages, no operational center has yet implemented this system at this time.

Lorenc (2003) is an excellent review on the key characteristics of 4DVAR and EnKF.

References:

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129,** 2884-2903.
Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129,** 420.
Burgers, G., P. J. v. Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126,** 1719-1724.
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.
Evensen, G., 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, **53,** 343-367.
Hamill, T. M. and C. Snyder, 2000: A hybrid ensemble Kalman filter - 3D variational analysis scheme. *Mon. Wea. Rev.*, **128,** 2905-2919.
Houtekamer, P. L. and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126,** 796-811.

Ide, K., P. Courtier, M. Ghil, and A. Lorenc, 1997: Unified notation for data assimilation: Operational, sequential and variational. *J. Meteor. Soc. Japan*, **75,** 181-189.

Lorenc, A., 2003: The potential of the ensemble Kalman filter for NWP - a comparison with 4D-Var. *Quart. J. Roy. Meteor. Soc.*, **129,** 3183-3204.

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.*, **131,** 1485-1490.

Tong, M. and M. Xue, 2005: Ensemble Kalman filter assimilation of Doppler radar data with a compressible nonhydrostatic model: OSS Experiments. *Mon. Wea. Rev.*, **133,** 1789-1807.

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