

Ensemble Kalman Filter: Current Status and Potential

Eugenia Kalnay

University of Maryland, College Park, MD, 20742-2425, USA,
ekalnay@atmos.umd.edu

1 Introduction

In this chapter we give an introduction to different types of Ensemble Kalman filter, describe the Local Ensemble Transform Kalman Filter (LETKF) as a representative prototype of these methods, and several examples of how advanced properties and applications that have been developed and explored for 4D-Var (four-dimensional variational assimilation) can be adapted to the LETKF without requiring an adjoint model. Although the Ensemble Kalman filter is less mature than 4D-Var, its simplicity and its competitive performance with respect to 4D-Var suggest that it may become the method of choice.

The mathematical foundation of data assimilation is reviewed by Nichols (chapter *Mathematical Concepts of Data Assimilation*). Ide *et al.* (1997) concisely summarized the *sequential* and *variational* approaches in a paper introducing a widely used notation that we follow here, with bold low-case letters and bold capitals representing vectors and matrices, respectively. Non-linear operators are, however, represented in bold Künster script (as in other chapters in this book). Since variational methods (chapter *Variational Assimilation*, Talagrand) and sequential methods basically solve the same problem (Lorenç 1986; Fisher *et al.* 2005) but make different approximations in order to become computationally feasible for large atmospheric and oceanic problems, it is particularly interesting to compare them whenever possible.

In this chapter we briefly review the most developed advanced sequential method, the Ensemble Kalman filter (EnKF) and several widely used formulations (Section 2). In Section 3 we compare the EnKF with the corresponding most advanced variational approach, 4D-Var (see chapter *Variational Assimilation*, Talagrand). Because 4D-Var has a longer history (*e.g.* Talagrand and Courtier 1987; Courtier and Talagrand 1990; Thépaut and Courtier 1991), and has been implemented in many operational centers (*e.g.* Rabier *et al.* 2000), there are many innovative ideas that have been developed and explored in the context of 4D-Var, whereas the EnKF is a newer and less mature approach. We therefore present in Section 3 examples of how specific approaches explored in the context of 4D-Var can be simply adapted to the EnKF. These include the 4D-Var smoothing property that leads to a faster spin-up, the outer loop that increases the analysis accuracy in the presence of non-linear observation operators, the adjoint sensitivity of the forecasts to the observations, the use of lower resolution analysis grids, and the treatment of model errors. Section 4 is a summary and discussion.

2 Brief review of ensemble Kalman filtering

The Kalman filter equations are discussed by Nichols (chapter *Mathematical Concepts of Data Assimilation*, Section 1.3.1). Here we summarize key points of an alternative derivation of the Kalman filter equations for a linear perfect model due to Hunt *et al.* (2007) based on a maximum likelihood approach which provides additional insight about the role that the background term plays in the variational cost function (see Nichols, Chapter *Mathematical Concepts of Data Assimilation*, Section 1.2; Talagrand, Chapter *Variational Assimilation*, Section 2).

We start by assuming that the analysis $\bar{\mathbf{x}}_{n-1}^a$ valid at time t_{n-1} has Gaussian errors with covariance \mathbf{P}_{n-1}^a so that the likelihood of the true state \mathbf{x}^t is

$$\rho(\mathbf{x}^t - \bar{\mathbf{x}}_{n-1}^a) \propto \exp\left\{-\frac{1}{2}(\mathbf{x}^t - \bar{\mathbf{x}}_{n-1}^a)^T [\mathbf{P}_{n-1}^a]^{-1} (\mathbf{x}^t - \bar{\mathbf{x}}_{n-1}^a)\right\},$$

where the overbar represents the expected value (*cf.* Chapter *Mathematical Concepts of Data Assimilation*, Nichols, Section 1.2.4). The past observations \mathbf{y}_j from time t_1 to t_{n-1} (i.e. $j=1, \dots, n-1$) are also assumed to have a Gaussian distribution with error covariances \mathbf{R}_j , so that the likelihood of a trajectory of states $\{\mathbf{x}(t_j) \mid j=1, \dots, n-1\}$ given the past observations is proportional to

$$\prod_{j=1}^{n-1} \exp\left[-\frac{1}{2}(\mathbf{y}_j - \mathbf{H}_j \mathbf{x}(t_j))^T \mathbf{R}_j^{-1} (\mathbf{y}_j - \mathbf{H}_j \mathbf{x}(t_j))\right],$$

where \mathbf{H}_j is the linear observation operator that transforms the model into the corresponding observation. To maximize the likelihood function, it is more convenient, however, to write the likelihood function as a function of the state at a single time rather than for the whole trajectory. Let $\mathbf{M}_{i,j}$ be the linear forecast model that advances a state from $\mathbf{x}(t_i)$ to $\mathbf{x}(t_j)$, we can then express the likelihood function as a function of the state \mathbf{x} at a single time say t_{n-1} , as follows

$$\prod_{j=1}^{n-1} \exp\left[-\frac{1}{2}(\mathbf{y}_j - \mathbf{H}_j \mathbf{M}_{n-1,j} \mathbf{x}_{n-1})^T \mathbf{R}_j^{-1} (\mathbf{y}_j - \mathbf{H}_j \mathbf{M}_{n-1,j} \mathbf{x}_{n-1})\right].$$

Note that in this derivation we allow t_j to be less than t_{n-1} , although integrating the model backward in time is problematic, it is used here only to derive the algorithm – in the end the algorithm will not require the backward integration of the model. Such an issue about time integration is found in the derivation of most Kalman smoother algorithms (see for instance Jazwinski 1970).

The analysis $\bar{\mathbf{x}}_{n-1}^a$ and its covariance \mathbf{P}_{n-1}^a are the mean and covariance of a Gaussian probability distribution representing the relative likelihood of a state \mathbf{x}_{n-1} given all previous observations, so that taking logarithms of the likelihoods, for some constant c ,

$$\sum_{j=1}^{n-1} [\mathbf{y}_j^o - \mathbf{H}_j \mathbf{M}_{n-1,j} \mathbf{x}_{n-1}]^T \mathbf{R}_j^{-1} [\mathbf{y}_j^o - \mathbf{H}_j \mathbf{M}_{n-1,j} \mathbf{x}_{n-1}] = [\mathbf{x}_{n-1} - \bar{\mathbf{x}}_{n-1}^a]^T (\mathbf{P}_{n-1}^a)^{-1} [\mathbf{x}_{n-1} - \bar{\mathbf{x}}_{n-1}^a] + c \quad (1)$$

The Kalman filter determines $\bar{\mathbf{x}}_n^a$ and \mathbf{P}_n^a such that an equation analogous to Eq. 1 holds at time t_n . In the **forecast step** of the Kalman filter the analysis $\bar{\mathbf{x}}_n^a$ and its covariance are propagated to time t_n with the linear forecast model $\mathbf{M}_{n-1,n}$ and its adjoint $\mathbf{M}_{n-1,n}^T$ creating the background state and its covariance:

$$\begin{aligned}\bar{\mathbf{x}}_n^b &= \mathbf{M}_{n-1,n} \bar{\mathbf{x}}_{n-1}^a \\ \mathbf{P}_n^b &= \mathbf{M}_{n-1,n} \mathbf{P}_n^a \mathbf{M}_{n-1,n}^T\end{aligned}\quad (2)$$

Propagating Eq. 1, using Eq. 2, we get a relationship valid for states at time t_n (see Hunt *et al.* 2007 for further details), showing that the background term represents the Gaussian probability distribution of a state, given the past observations up to t_{n-1} :

$$\sum_{j=1}^{n-1} [\mathbf{y}_j^o - \mathbf{H}_j \mathbf{M}_{n,j} \mathbf{x}_n]^T \mathbf{R}_j^{-1} [\mathbf{y}_j^o - \mathbf{H}_j \mathbf{M}_{n,j} \mathbf{x}_n] = [\mathbf{x}_n - \bar{\mathbf{x}}_n^b]^T (\mathbf{P}_n^b)^{-1} [\mathbf{x}_n - \bar{\mathbf{x}}_n^b] + c \quad (3)$$

When the new observations at time t_n are obtained, we use Eq. 3 to obtain an expression equivalent to Eq. 1 valid at time t_n , for another constant c' :

$$[\mathbf{x}_n - \bar{\mathbf{x}}_n^b]^T (\mathbf{P}_n^b)^{-1} [\mathbf{x}_n - \bar{\mathbf{x}}_n^b] + [\mathbf{y}_n^o - \mathbf{H}_n \mathbf{x}_n]^T \mathbf{R}_n^{-1} [\mathbf{y}_n^o - \mathbf{H}_n \mathbf{x}_n] = [\mathbf{x}_n - \bar{\mathbf{x}}_n^a]^T (\mathbf{P}_n^a)^{-1} [\mathbf{x}_n - \bar{\mathbf{x}}_n^a] + c' \quad (4)$$

The analysis state that minimizes the variational cost function

$$J(\mathbf{x}_n) = [\mathbf{x}_n - \bar{\mathbf{x}}_n^b]^T (\mathbf{P}_n^b)^{-1} [\mathbf{x}_n - \bar{\mathbf{x}}_n^b] + [\mathbf{y}_n^o - \mathbf{H}_n \mathbf{x}_n]^T \mathbf{R}_n^{-1} [\mathbf{y}_n^o - \mathbf{H}_n \mathbf{x}_n]$$

is the state with maximum likelihood given all the observations (*cf.*, chapter *Mathematical Concepts of Data Assimilation*, Section 1.2.4). Equation 3 shows that in this cost function the background term represents the Gaussian distribution of a state with the maximum likelihood trajectory (history), *i.e.*, $\bar{\mathbf{x}}_n^b$ is the analysis/forecast trajectory that best fits the past data available until t_{n-1} .

Equating the terms in Eq. 4 that are quadratic and linear in \mathbf{x} , the Kalman filter equations for the **analysis step** are obtained:

$$\mathbf{P}_n^a = \left[(\mathbf{P}_n^b)^{-1} + \mathbf{H}_n^T \mathbf{R}_n^{-1} \mathbf{H}_n \right]^{-1} = \left[\mathbf{I} + \mathbf{P}_n^b \mathbf{H}_n^T \mathbf{R}_n^{-1} \mathbf{H}_n \right]^{-1} \mathbf{P}_n^b \quad (5)$$

$$\bar{\mathbf{x}}_n^a = \mathbf{P}_n^a \left[(\mathbf{P}_n^b)^{-1} \bar{\mathbf{x}}_n^b + \mathbf{H}_n^T \mathbf{R}_n^{-1} \mathbf{y}_n^o \right] = \bar{\mathbf{x}}_n^b + \mathbf{P}_n^a \mathbf{H}_n^T \mathbf{R}_n^{-1} (\mathbf{y}_n^o - \mathbf{H}_n \bar{\mathbf{x}}_n^b) \quad (6)$$

The Remark 1 of Ide *et al.* (1997) “[In sequential methods] observations are processed whenever available and then discarded” follows from the fact that the background term is the most likely solution given all the past data, *i.e.*, *if the Kalman filter has already spun-up from the initial conditions*, the observations are to be used only once (but see the discussion on spin-up in Section 3).

The Kalman gain matrix that multiplies the observational increment $\mathbf{y}_n^o - \mathbf{H}_n \bar{\mathbf{x}}_n^b$ in Eq. 6 can be written as

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

For non-linear models $\mathbf{M}_{n-1,n}$, the *Extended Kalman filter* (EKF) approximation uses the non-linear model in the forecast step to advance the background state, but the covariance is advanced using the model linearized around the trajectory $\bar{\mathbf{x}}_n^b$, and its adjoint (e.g. Ghil and Rizzoli 1991; Nichols, chapter *Mathematical Concepts of Data Assimilation*, Section 1.3):

$$\begin{aligned} \bar{\mathbf{x}}_n^b &= \mathbf{M}_{n-1,n}(\bar{\mathbf{x}}_{n-1}^a), \\ \mathbf{P}_n^b &= \mathbf{M}_{n-1,n} \mathbf{P}_{n-1}^a \mathbf{M}_{n-1,n}^T \end{aligned} \quad (7)$$

The cost of advancing the background error covariance with the linear tangent and adjoint models in Eq. 7 makes the EKF computationally unfeasible for any atmospheric model of realistic size without major simplifications.

Evensen (1994) suggested that Eq. 7 could be computed more efficiently with an *Ensemble Kalman filter* (EnKF) for non-linear models. The ensemble is created running K forecasts, where the size of the forecast ensemble is much smaller than n , the dimension of the model, $K \ll n$. Then Eq. 7 can be replaced by

$$\begin{aligned} \bar{\mathbf{x}}_n^b &= \mathbf{M}_{n-1,n}(\bar{\mathbf{x}}_{n-1}^a), \bar{\mathbf{x}}_n^b = \frac{1}{K} \sum_{k=1}^K \mathbf{x}_{n,k}^b \\ \mathbf{P}_n^b &\approx \frac{1}{K-1} \sum_{k=1}^{K-1} (\mathbf{x}_{n,k}^b - \bar{\mathbf{x}}_n^b)(\mathbf{x}_{n,k}^b - \bar{\mathbf{x}}_n^b)^T \end{aligned} \quad (8)$$

where the overbar now represents the *ensemble average*.

Because the background error covariance is estimated from a relatively small ensemble, there are sampling errors at long distances, so that Houtekamer and Mitchell (2001) and Hamill *et al.* (2001) introduced the idea of *localizing* \mathbf{P}_n^b , i.e., multiplying each term of the covariance by an approximation of the Gaussian function $\exp(-r_{ij}^2 / 2L^2)$ (Gaspari and Cohn 1999). Here, r_{ij} is the distance between two grid points i,j , and L is the localization scale, so that the effect of localization is that long distance correlations are damped to zero. Mitchell *et al.* (2002) pointed out that this localization introduces imbalances in the analysis. Hunt (2005) and Miyoshi (2005) used an alternative localization multiplying the inverse of the observation error covariance \mathbf{R}^{-1} by the Gaussian function, thus assuming that long distance observations have larger errors and reducing their impact on the grid point analyses. Because, unlike \mathbf{P}_n^b , \mathbf{R} is generally either diagonal or block diagonal, this ‘‘observation localization’’ may be less prone to generate imbalances (Greybush 2009).

There are two basic approaches to the EnKF, perturbed observations and square-root filters. In the *perturbed observations* EnKF, Burgers *et al.* (1998), Houtekamer

and Mitchell (1998), Keppenne (2000) and others used ensembles of data assimilation systems with randomly perturbed observations. Perturbing the observations assimilated in different ensembles is required in this approach in order to avoid an underestimation of the size of the analysis error covariance, but it may introduce an additional source of sampling errors (Whitaker and Hamill 2002).

An alternative to the perturbed observations (or stochastic) approach are the *ensemble square-root filters* that generate an analysis ensemble mean and covariance satisfying the Kalman filter equations for linear models (Tippett *et al.* 2003; Bishop *et al.* 2001; Anderson 2001; Whitaker and Hamill 2002; Ott *et al.* 2004; Hunt *et al.* 2007). We will focus in the rest of the chapter on square-root (or deterministic) filters. Houtekamer and Mitchell (2001) pointed out that observations with uncorrelated errors can be assimilated serially (one at a time), with the background for a new observation being the analysis obtained when assimilating the previous observation. Tippett *et al.* (2003) discuss the differences between several square-root filters that derive computational efficiency by assimilating observations serially. Another Monte Carlo method that avoids using perturbed observations is described in Pham (2001).

Different square-root filters are possible because different analysis ensemble perturbations can have the same analysis error covariance. Of the three schemes discussed in Tippett *et al.* (2003), the Ensemble Adjustment Kalman Filter (EAKF) of Anderson (2001) has been implemented into the flexible Data Assimilation Research Testbed (DART) infrastructure and has been applied to many geophysical problems (<http://www.image.ucar.edu/DARes/Publications/>). The square-root filter of Whitaker and Hamill (2002) results in simple scalar assimilation equations when observations are assimilated serially, and has also been adopted for a number of problems, such as the assimilation of surface observations (Whitaker *et al.* 2004), and for the regional EnKF of Torn and Hakim (2007) where only non-satellite data are assimilated. We note that the application of EnKF to regional models requires including appropriate perturbations in the boundary conditions to avoid a reduction in variance in the interior (Nutter *et al.* 2004). Torn *et al.* (2006) showed that in the absence of a global EnKF system to provide consistent perturbed boundary conditions, several perturbation methods could give results comparable to those obtained with a global ensemble boundary conditions, thus making regional EnKF practically feasible for many groups without access to global EnKF. The third square-root filter discussed in Tippett *et al.* (2003) is the Ensemble Transform Kalman Filter, ETKF (Bishop *et al.* 2001), which introduced the computation of the analysis covariance by a transform method also adopted by Hunt *et al.* (2007). Zupanski (2005) proposed the Maximum Likelihood Ensemble Filter (MLEF) where a 4D-Var cost function with possibly non-linear observation operators is minimized within the subspace of the ensemble forecasts. In this system, the control forecast is allowed to have higher resolution than the rest of the ensemble. A review of EnKF methods is presented in Evensen (2003), and a comparison of EnKF with 4D-Var results for several models in Kalnay *et al.* (2007a).

Ott *et al.* (2002, 2004), and Hunt *et al.* (2007) developed an alternative type of square-root EnKF without perturbed observations by performing the analyses *locally in space*, as did Keppenne (2001). This is computationally efficient because the analyses at different grid points are independent and thus can be done in parallel. Since observations are assimilated simultaneously, not serially, it is simple to account for observation error correlations.

In this chapter we present results mostly based on the Local Ensemble Transform Kalman Filter (LETKF) as a representative prototype of EnKF. The LETKF algorithm is summarized below (see Hunt *et al.* 2007, for full details).

LETKF algorithm:

This summary description is written as if all the observations are at the analysis time (*i.e.*, for the 3D-LETKF), but the algorithm is essentially the same for the 4D-LETKF (Hunt *et al.* 2007). In 4D-LETKF (discussed below) the observations are in a time window that includes the analysis time and the non-linear observation operator \mathbf{H} is evaluated at the observation time. \mathbf{M} is the non-linear model forecast.

- a) LETKF **forecast step** (done globally) for each ensemble member k :

$$\mathbf{x}_{n,k}^b = \mathbf{M}_{n-1,n}(\mathbf{x}_{n-1,k}^a), \quad k = 1, \dots, K$$

- b) LETKF **analysis step** (at time t_n , so the subscript n is dropped):

$$\begin{aligned} \mathbf{X}^b &= [\mathbf{x}_1^b - \bar{\mathbf{x}}^b, \dots, \mathbf{x}_K^b - \bar{\mathbf{x}}^b]; \quad \mathbf{P}^b = \mathbf{X}^b (\mathbf{X}^b)^T \\ \mathbf{y}_k^b &= \mathbf{H}(\mathbf{x}_k^b); \quad \mathbf{Y}^b = [\mathbf{y}_1^b - \bar{\mathbf{y}}^b, \dots, \mathbf{y}_K^b - \bar{\mathbf{y}}^b] \end{aligned}$$

These computations can be done locally or globally, whichever is more efficient. Here the overbar represents the ensemble average.

Localization: choose for each grid point the observations to be used. Compute for each grid point the local analysis error covariance $\hat{\mathbf{P}}^a$ and analysis perturbations \mathbf{W}^a in ensemble space:

$$\begin{aligned} \hat{\mathbf{P}}^a &= [(K-1)\mathbf{I} + (\mathbf{Y}^b)^T \mathbf{R}^{-1} \mathbf{Y}^b] \\ \mathbf{W}^a &= [(K-1)\hat{\mathbf{P}}^a]^{1/2} \end{aligned}$$

The square-root required for the matrix of analysis perturbations in ensemble space is computed using the symmetric square root (Wang *et al.* 2004). This square-root has the advantage of having a zero mean and being closer to the identity than the square-root matrix obtained by Cholesky decomposition. As a result the analysis perturbations (chosen in different ways in different EnKF schemes) are also close to the background perturbations (Ott *et al.* 2002). Note that \mathbf{W}^a can also be considered a *matrix of weights* since multiplying the forecast ensemble perturbations at each grid point by \mathbf{W}^a gives the grid point analysis ensemble perturbations.

Local analysis mean increment in ensemble space:

$$\bar{\mathbf{w}}^a = \hat{\mathbf{P}}^a (\mathbf{Y}^b)^T \mathbf{R}^{-1} (\mathbf{y}^o - \bar{\mathbf{y}}^b)$$

Note that the forecast ensemble at each grid point multiplied by the *vector of weights* $\bar{\mathbf{w}}^a$ gives the grid point analysis $\bar{\mathbf{x}}^a$. The ensemble space analysis $\bar{\mathbf{w}}^a$ is added to each column of \mathbf{W}^a to get the analysis ensemble in ensemble space: $\mathbf{W}^a \leftarrow \mathbf{W}^a \oplus \bar{\mathbf{w}}^a$

The new ensemble analyses are the K columns of

$$\mathbf{X}^a = \mathbf{X}^b \mathbf{W}^a + \bar{\mathbf{x}}^b$$

Global analysis ensemble:

The analysis ensemble columns for each grid point are gathered together to form the new global analysis ensemble $\mathbf{x}_{n,k}^a$, and the analysis cycle can proceed.

3 Adaptation of 4D-Var techniques into EnKF

4D-Var and EnKF are essentially solving the same problem since they minimize the same cost function in Eq. 2 using different computational methods. These differences lead to several advantages and disadvantages for each of the two methods (see, for example, Lorenc 2003; Table 7 of Kalnay *et al.* 2007a; discussion of Gustafsson 2007; response of Kalnay *et al.* 2007b).

A major difference between 4D-Var and the EnKF is the dimension of the subspace of the analysis increments (analysis minus background). 4D-Var corrects the background forecast in a subspace that has the dimension of the linear tangent and the adjoint models used in the minimization algorithm, and this subspace is generally much larger than the local subspace of corrections in the EnKF of dimension $K - 1$ determined by the ensemble size. It would be impractical to try to overcome this apparent EnKF disadvantage by using a very large ensemble size. Fortunately, the localization of the error covariances carried out in the EnKF in order to reduce long distance covariance sampling errors, substantially addresses this problem by greatly increasing the number of degrees of freedom available to fit the data. As a result, experience has been that the quality of the EnKF analyses with localization increases with the number of ensemble members, but that there is little further improvement when the size of the ensemble is increased beyond about 100. The observation that 50-100 ensemble members are sufficient for the EnKF seems to hold for atmospheric problems ranging from the storm-scales and mesoscales to the global-scales (*Fuqing Zhang, pers. comm.*).

There are a number of attractive properties of 4D-Var developed over the years. They include the ability to assimilate observations at their right time (Talagrand and Courtier 1987); the fact that within the data assimilation window 4D-Var acts as a smoother (Thépaut and Courtier 1991); the availability of an adjoint model allowing the estimation of the impact of observations on the analysis (Cardinali *et al.* 2004) and on the forecasts (Langland and Baker 2004); the ability to use long assimilation windows (Pires *et al.* 1996); the computation of outer loops correcting the background state when computing non-linear observation operators and the ability to use a lower resolution simplified model in the inner loop (see Fig. 3 discussed later); and the possibility of accounting for model errors by using the model as a weak constraint (Trémolet 2007). In the rest of this section we discuss how these advantageous methods that have been developed and implemented for 4D-Var systems can also be adapted and used in the LETKF, a prototype of EnKF.

3.1 4D-LETKF and no-cost smoother

Hunt *et al.* (2004) developed an extension of the Local Ensemble Kalman Filter (LEKF; Ott *et al.* 2004) to four dimensions (4-D), taking advantage of the fact that the observational increments are expressed as linear combinations (weights) of the

forecast ensemble perturbations at the time of the observation¹. This allows using the same coefficients to “transport” the observational increments either forward or backward in time to the time of the analysis. We note that within this 4-D formulation it is possible to account for observation errors correlated in time, as Järvinen *et al.* (1999) have done within 4D-Var. Hunt *et al.* (2007) showed that the 4-D extension is particularly simple within the LETKF framework, requiring the concatenation of observations performed at different times within the assimilation window into the vectors $\mathbf{y}^o, \bar{\mathbf{y}}^b$ and the vertical columns of \mathbf{Y}^b and of a block error covariance \mathbf{R} with blocks corresponding to the same observations. Note that 4D-LETKF determines the linear combination of ensemble forecasts valid at the end of the assimilation window that best fits the data throughout the assimilation window.

This property allows creating a “cost-free” smoother for the LETKF with analogous smoothing properties as 4D-Var (Fig. 1): the same weighted combination of the forecasts with weights given by the vector $\bar{\mathbf{w}}^a$ is valid at any time of the assimilation interval. It provides a smoothed analysis mean that (as in 4D-Var) is more accurate than the original analysis because it uses the future data available within the assimilation window (Kalnay *et al.* 2007b; Yang *et al.* 2009a). As in 4D-Var, the smoothed analysis at the beginning of the assimilation window is an improvement over the filtered analysis computed using only past data. At the end of the assimilation interval only past data is used so that (as in 4D-Var) the smoother coincides with the analysis obtained with the filter.

It should be noted that in the same way we can use the weights $\bar{\mathbf{w}}^a$ to provide a mean smoother solution as a function of time, we can use the matrix \mathbf{W}^a and apply it to the forecast perturbations $\mathbf{X}^b \mathbf{W}^a$ to provide an associated uncertainty evolving with time (Ross Hoffman, *pers. comm.*). The updating of the uncertainty is critical for the “Running in Place” method described next, but the uncertainty is not updated in the “outer loop” approach.

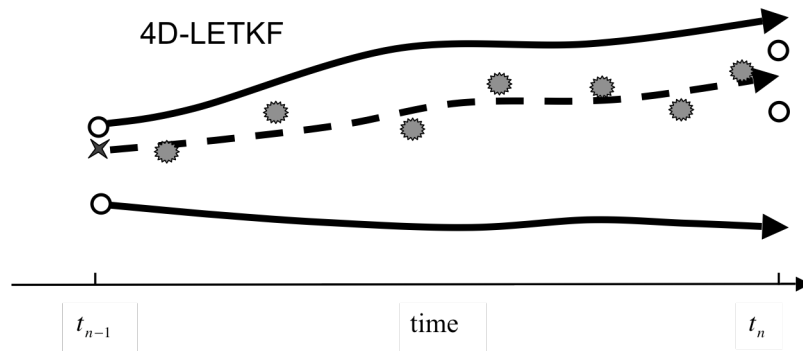


Fig. 1. Schematic showing that the 4D-LETKF finds the linear combination of the ensemble forecasts at t_n that best fits the observations *throughout* the assimilation

¹ Strictly speaking the combinations are not linear since the weights depend on the forecasts (Nerger *et al.* 2005).

window $t_{n-1} - t_n$. The white circles represent the ensemble of analyses (whose mean is the analysis $\bar{\mathbf{x}}^a$), the full lines represent the ensemble forecasts, the dashed line represents the linear combination of the forecasts whose final state is the analysis, and the grey stars represent the asynchronous observations. The cross at the initial time of the assimilation window t_{n-1} is a *no-cost Kalman smoother*, i.e., an analysis at t_{n-1} improved using the information of “future” observations within the assimilation window by weighting the ensembles at t_{n-1} with the weights obtained at t_n . The smoothed analysis ensemble at t_{n-1} (not shown in the schematic) can also be obtained at no cost using the same linear combination of the ensemble forecasts valid at t_n given by \mathbf{W}^a (Adapted from Kalnay *et al.* 2007b.).

3.2 Application of the no-cost smoother to the acceleration of the spin-up

4D-Var has been observed to spin up faster than EnKF (*e.g.* Caya *et al.* 2005), presumably because of its smoothing properties that allow finding the initial conditions at the beginning of the assimilation window that will best fit all the observations. The fact that we can compute a no-cost smoother allows the development of a new algorithm, called *Running in Place* by Kalnay and Yang (2009) that should be useful in rapidly evolving situations. For example, at the time radar measurements first detect the development of a severe storm, the available EnKF estimate of the atmospheric state and its uncertainty are no longer very useful. In other words, *while formally the EnKF members and their average are still the most likely state and best estimate of the uncertainty given all the past data, these EnKF estimates are no longer likely at all.* At the start of severe storm convection, the dynamics of the system changes substantially, and the statistics of the processes become non-stationary. In this case, as in the spin-up case in which there are no previous observations available, the running in place algorithm ignores the rule “use the data and then discard it” and repeatedly recycles the new observations.

Running in place algorithm:

This algorithm is applied to each assimilation window during the spin-up phase. The LETKF is “cold-started” with any initial ensemble mean and perturbations at t_0 . The “running in place” loop at time t_n (initially t_0) is as follows:

1. Integrate the ensemble from t_n to t_{n+1} , perform a standard LETKF analysis and obtain the analysis weights for the interval $[t_n, t_{n+1}]$, saving the mean square observations minus forecast (OMF) computed by the LETKF;
2. Apply the no-cost smoother to obtain the smoothed analysis ensemble at t_n by using these weights;
3. Perturb the smoothed analysis ensemble with small zero-mean random Gaussian perturbations, a method similar to additive inflation. Typically, the perturbations have an amplitude equal to a small percentage of the climate variance;
4. Integrate the perturbed smoothed ensemble to t_{n+1} . While the forecast fit to the observations continues to improve according to a criterion such as

$$\frac{OMF^2(ite\text{r}) - OMF^2(ite\text{r} + 1)}{OMF^2(ite\text{r})} > \varepsilon,$$

go to step 2 and perform another iteration. If not, replace t_n with t_{n+1} and go to step 1;

5. If no additional iteration beyond the first one is needed, the **running in place**

analysis is the same as the standard EnKF. When the system converges, no additional iterations are needed, so that if several assimilation cycles take place without invoking a second iteration, the **running in place** algorithm can be switched off and the system returns to a normal EnKF.

The purpose of adding perturbations in step 3 is twofold: it avoids reaching the same analysis as in the previous iteration, and it increases the chances that the ensemble will explore unstable directions of error growth missed by the unperturbed ensemble and not be “trapped” in the “unlikely” subspace of the initial perturbations.

Running in place was tested with the LETKF in a quasi-geostrophic, QG, model (Fig. 2, adapted from Kalnay and Yang 2009). When starting from a 3D-Var (three dimensional variational) analysis mean, the LETKF converges quickly (not shown), but from random initial states it takes 120 cycles (60 days) to reach a point in which the ensemble perturbations represent the “errors of the day” (black line in Fig. 2). From then on the ensemble converges quickly, in about 60 more cycles (180 cycles total).

By contrast, the 4D-Var started from the same initial mean state, but using as background error covariance the 3D-Var \mathbf{B} scaled down with an optimal factor, converges twice as fast, in about 90 cycles (blue line in Fig. 2). The running in place algorithm with $\varepsilon = 5\%$ (red line) converges about as fast as 4D-Var, and it only takes about 2 iterations per cycle (*i.e.*, one additional assimilation for each window). The green line is also for $\varepsilon = 5\%$, but with $K=20$ ensemble members, not $K=40$ as used in the other experiments and also gives good results, but experiments with $K=10$ failed to spin-up faster with this technique. With $\varepsilon = 1\%$ (not shown) the initial convergence (in real time) is faster, but it requires about 5 times more iterations. It is interesting that when the number of iterations is *fixed* to 10 (not shown), the data are over-fitted so that the system quickly converges to a final level of error about twice as large than when the iterations are chosen adaptively.

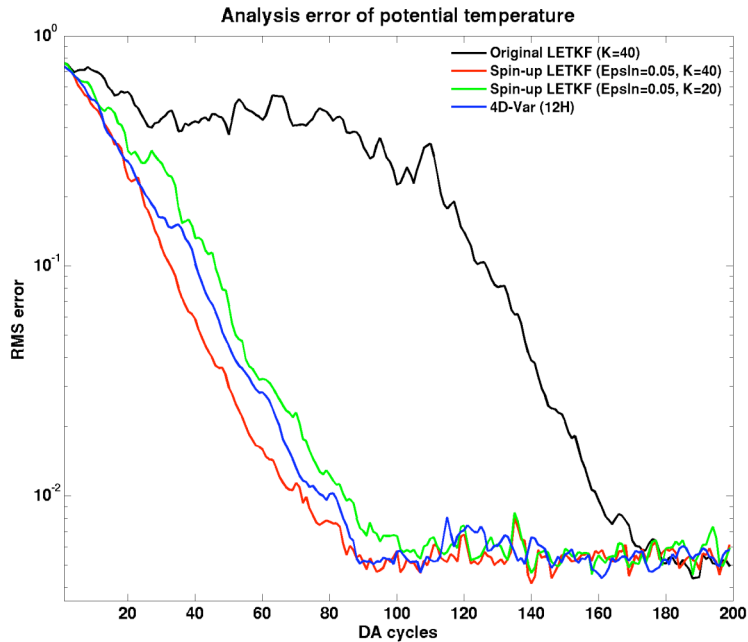


Fig. 2. Comparison of the spin-up of a quasi-geostrophic model simulated data assimilation when starting from random initial conditions. Observations (simulated radiosondes) are available every 12 hours, and the analysis root-mean-square (RMS) errors are computed by comparing with a nature run (see the chapter *Observing System Simulation Experiments*, Masutani *et al.*). Black line: original LETKF with 40 ensemble members, and no prior statistical information, blue line: optimized 4D-Var, red line: LETKF “running in place” with $\varepsilon = 5\%$ and 40 ensemble members, green line: as the red line but with 20 ensemble members.

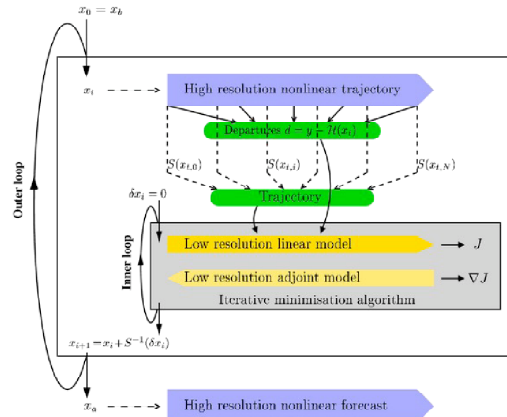
3.3 “Outer loop” and dealing with non-linear ensemble perturbations

A disadvantage of the EnKF is that the Kalman filter equations used in the analysis assume that the ensemble perturbations are Gaussian, so that when windows are relatively long and perturbations become non-linear, this assumption breaks down and the EnKF is not optimal (Harlim and Hunt 2007a, b). By contrast, 4D-Var is recomputed within an assimilation window until the initial conditions that minimize the cost function for the non-linear model integration in that window are found. In many operational centres (*e.g.* the National Centers for Environmental Prediction, NCEP, and the European Centre for Medium-Range Weather Forecasts, ECMWF) the minimization of the 3D-Var or 4D-Var cost function is done with a linear “inner loop” that improves the initial conditions minimizing a cost function that is quadratic in the perturbations. In the 4D-Var “outer loop” the non-linear model is integrated from the initial state improved by the inner loop and the linearized observational increments are recomputed for the next inner loop (Fig. 3).

The ability of including an outer loop increases significantly the accuracy of both 3D-Var and 4D-Var analyses (Arlindo da Silva, *pers. comm.*), so that it would be important to develop the ability to carry out an equivalent “outer loop” in the

LETKF. This can be done by considering the LETKF analysis for a window as an “inner loop” and, using the no-cost smoother, adapting the 4D-Var outer loop algorithm to the EnKF. As in 4D-Var, we introduce into the LETKF the freedom of the inner loop to improve the initial analysis (*i.e.*, the mean of the ensemble) but keep constant the background error covariance, given by the ensemble initial perturbations. This re-centres the initial ensemble forecasts about the value improved by the inner loop, and another “outer loop” with full non-linear integrations can be carried out². Note that this algorithm is identical to “running in place”, except that only the mean is updated, not the perturbations about the mean at t_n .

Incremental 4D-Var



Incremental 4D-Var - 7



Fig. 3. Schematic of how the 4D-Var cost function is minimized in the ECMWF system. (From Yannick Trémolet, August 2007 class on Incremental 4D-Var at University of Maryland summer Workshop on Applications of Remotely sensed data to Data Assimilation.)

Table 1: Comparison of the RMSE (RMS error, non-dimensional units) for 4D-Var and LETKF for the Lorenz (1963) 3-variable model. 4D-Var has been simultaneously optimized for the window length (Kalnay *et al.* 2007a; Pires *et al.* 1996) and the background error covariance, and the full non-linear model is used in the minimization. LETKF is performed with 3 ensemble members (no localization is needed for this problem), and

² Takemasa Miyoshi (*pers. comm.*) has pointed out that Jazwinski (1970) proposed the same “outer loop” algorithm for Extended Kalman filter (see footnote on page 276).

inflation is optimized. For the 25 steps case, “running in place” further reduces the error to about 0.39.

| Experiment details | 4D-Var | LETKF (inflation factor) | LETKF with less than 3 “outer loop” iterations (inflation factor) |
|---|--------|-----------------------------|---|
| Window=8 steps (perturbations are approximately linear) | 0.31 | 0.30 (1.05) | 0.27 (1.04) |
| Window=25 steps (perturbations are non-linear) | 0.53 | 0.66 (1.28) | 0.48 (1.08) |

This algorithm for an outer loop within the EnKF was tested with the Lorenz (1963) model for which comparisons between LETKF and 4D-Var were made, optimizing simultaneously the background error covariance and the length of the window for 4D-Var (Kalnay *et al.* 2007a). For short assimilation windows, the 3-member LETKF gives analysis errors similar or smaller than 4D-Var, but with long assimilation windows of 25 steps, when the perturbations grow non-linearly, Kalnay *et al.* (2007a) were not able to find an LETKF configuration competitive with 4D-Var. However, as shown in Table 1 above, the LETKF with an outer loop is able to beat 4D-Var. We note that “running in place” (with up to one additional analysis) can further improve the results for the case of 25 steps, reducing the RMS (root-mean-square) analysis error of 0.48 obtained using the outer loop to about 0.39, with inflation of 1.05. As in the case of the spin-up, this re-use of observations is justified by the fact that for long windows and non-linear perturbations, the background ensemble ceases to be Gaussian, and the assumption of statistical stationarity is no longer viable.

These experiments suggest that it should be possible to deal with non-linearities and obtain results comparable or better than 4D-Var by methods such as an outer loop and running in place.

3.4 Adjoint forecast sensitivity to observations without adjoint model

Langland and Baker (2004) proposed an adjoint-based procedure to assess the observation impact on short-range forecasts without carrying out data-denial experiments. This adjoint-based procedure can evaluate the impact of any or all observations assimilated in the data assimilation and forecast system on a selected measure of short-range forecast error. In addition, it can be used as a diagnostic tool to monitor the quality of observations, showing which observations make the forecast worse, and can also give an estimate of the relative importance of observations from different sources. Following a similar procedure, Zhu and Gelaro (2008) showed that this adjoint-based method provides accurate assessments of the forecast sensitivity with respect to most of the observations assimilated. Unfortunately, this powerful and efficient method to estimate observation impact requires the adjoint of the forecast model which is complicated to develop and not always available, as well as the adjoint of the data assimilation algorithm.

Liu and Kalnay (2008) proposed an ensemble-based sensitivity method able to assess the same forecast sensitivity to observations as in Langland and Baker (2004),

but without adjoint. Following Langland and Baker (2004), they define a cost function $\Delta \mathbf{e}_t^2 = (\mathbf{e}_{t|0}^T \mathbf{e}_{t|0} - \mathbf{e}_{t|-6}^T \mathbf{e}_{t|-6})$ that measures the forecast sensitivity at time t of the observations assimilated at time 0. Here $\mathbf{e}_{t|0} = \bar{\mathbf{x}}_{t|0}^f - \bar{\mathbf{x}}_t^a$ is the perceived error of the forecast started from the analysis at $t = 0$, verified against the analysis valid at time t , and $\mathbf{e}_{t|-6} = \bar{\mathbf{x}}_{t|-6}^f - \bar{\mathbf{x}}_t^a$ is the corresponding error of the forecast starting from the previous analysis at $t = -6$ hours. The difference between the two forecasts is only due to the observations \mathbf{y}_0^o assimilated at $t = 0$: $\bar{\mathbf{x}}_0^a - \bar{\mathbf{x}}_{0|-6}^b = \mathbf{K}(\mathbf{y}_0^o - \mathbf{H}(\bar{\mathbf{x}}_{0|-6}^b))$, where \mathbf{K} is the gain matrix of the data assimilation system. There is a slight error in Eq. 10 in Liu and Kalnay (2008) so that we re-derive here the forecast sensitivity equation (*Hong Li, pers. comm.*):

$$\begin{aligned} \Delta \mathbf{e}_t^2 &= (\mathbf{e}_{t|0}^T \mathbf{e}_{t|0} - \mathbf{e}_{t|-6}^T \mathbf{e}_{t|-6}) = (\mathbf{e}_{t|0}^T - \mathbf{e}_{t|-6}^T)(\mathbf{e}_{t|0} + \mathbf{e}_{t|-6}) = (\bar{\mathbf{x}}_{t|0}^f - \bar{\mathbf{x}}_{t|-6}^f)^T (\mathbf{e}_{t|0} + \mathbf{e}_{t|-6}) \\ \Delta \mathbf{e}_t^2 &\approx [\mathbf{M}(\bar{\mathbf{x}}_0^a - \bar{\mathbf{x}}_{0|-6}^b)] (\mathbf{e}_{t|0} + \mathbf{e}_{t|-6}) = [\mathbf{M}\mathbf{K}(\mathbf{y} - \mathbf{H}(\bar{\mathbf{x}}_{0|-6}^b))] (\mathbf{e}_{t|0} + \mathbf{e}_{t|-6}) \end{aligned}$$

where \mathbf{M} is the linear tangent forecast model that advances a perturbation from 0-hours to time t .

Langland and Baker (2004) compute this error sensitivity by using the adjoint of the model and of the data assimilation scheme:

$$\Delta \mathbf{e}_{t, LB}^2 = [\mathbf{y} - \mathbf{H}(\bar{\mathbf{x}}_{0|-6}^b)] \mathbf{K}^T \mathbf{M}^T (\mathbf{e}_{t|0} + \mathbf{e}_{t|-6})$$

In the EnKF we can take advantage of the fact that the Kalman gain is computed as $\mathbf{K} = \mathbf{P}^a \mathbf{H}^T \mathbf{R}^{-1} = (K - 1)^{-1} \mathbf{X}^a (\mathbf{X}^a)^T \mathbf{H}^T \mathbf{R}^{-1}$, so that $\mathbf{M}\mathbf{K} = \mathbf{M}\mathbf{X}^a (\mathbf{X}^a)^T \mathbf{H}^T \mathbf{R}^{-1} / (K - 1) \approx \mathbf{X}_{t|0}^f (\mathbf{Y}^a)^T \mathbf{R}^{-1} / (K - 1)$, with $\mathbf{X}_{t|0}^f$, the forecast differences at time t computed non-linearly rather than with the linear tangent model. As a result, for EnKF the forecast sensitivity is computed as

$$\Delta \mathbf{e}_{t, EnKF}^2 = [\mathbf{y} - \mathbf{H}(\bar{\mathbf{x}}_{0|-6}^b)]^T \mathbf{R}^{-1} \mathbf{Y}^a (\mathbf{X}_{t|0}^f)^T (\mathbf{e}_{t|0} + \mathbf{e}_{t|-6}) / (K - 1)$$

Because the forecast perturbation matrix $\mathbf{X}_{t|0}^f$ is computed non-linearly, the forecast sensitivity and the ability to detect bad observations remains valid even for forecasts longer than 24 hours, for which the adjoint sensitivity based on the adjoint model \mathbf{M}^T ceases to be accurate. As in Langland and Baker (2004) and Zhu and Gelaro (2008), it is possible to split the vector of observational increments $\mathbf{y} - \mathbf{H}(\bar{\mathbf{x}}_{0|-6}^b)$ into any subset of observations and obtain the corresponding forecast sensitivity.

Figure 4 shows the result of applying this method to the Lorenz (1996) 40-variables model. In this case observations were made at every point every 6 hours, created from a ‘‘nature’’ run by adding Gaussian observational errors of mean zero and standard deviation 0.2. The left panel shows that both the adjoint and the EnKF sensitivity methods are able to estimate quite accurately the day-to-day variability in the 24-hour forecast sensitivity to the observations when all the observations have similar Gaussian errors. A ‘‘bad station’’ was then simulated at grid point 11 by

increasing the standard deviation of the errors to 0.8 without “telling” the data assimilation system about the observation problems in this location. The right panel of Fig. 4 shows the time average of the forecast sensitivity for this case, indicating that both the adjoint and the ensemble-based sensitivity are able to identify that the observations at grid point 11 have a deleterious impact on the forecast. They both show that the neighbouring points improved the forecasts more than average by partially correcting the effects of the 11th point observations.

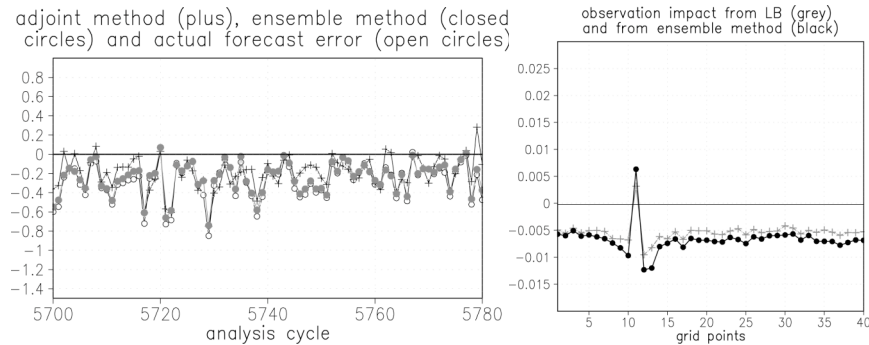


Fig. 4. Left: Domain average variability in the forecast impact estimated by the adjoint method (plus symbols), the EnKF sensitivity (closed circles) and the actual forecast sensitivity. Right: Time average (over the last 7000 analysis cycles) of the contribution to the reduction of the 1-day forecast errors from each observation location. The observation at the 11th grid point has $\sigma^o = 8$ random errors rather than the specified value of 0.2. Adjoint sensitivity (grey plus), EnKF sensitivity (black). Adapted from Liu and Kalnay (2008).

The cost function in this example was based on the Eulerian norm, appropriate for a univariate problem, but the method can be easily extended to an energy norm, allowing the comparison of the impact of winds and temperature observations on the forecasts. Although for short (one-day) forecasts the adjoint and ensemble sensitivities have similar performances (Fig. 4), the (linear) adjoint sensitivity ceases to identify the wrong observations if the forecasts are 2-days or longer. The ensemble sensitivity, which is based on non-linear integrations, continues to identify the observations having a negative impact even on long forecasts (not shown).

We note that Liu *et al.* (2009) also formulated the sensitivity of the analysis to the observations as in Cardinali *et al.* (2004) and showed that it provides a good qualitative estimate of the impact of adding or denying observations on the analysis error, without the need to run these costly experiments. Since the Kalman gain matrix is available in ensemble space, complete cross-validations of each observation can be computed exactly within the LETKF without repeating the analysis.

3.5 Use of a lower resolution analysis

The inner/outer loop used in 4D-Var was introduced in subsection 3.3 above, where we showed that a similar outer loop can be carried out in EnKF. We now point out that it is common practice to compute the inner loop minimization, shown

schematically in Fig. 3, using a simplified model (Lorenz 2003), which usually has lower resolution and simpler physics than the full resolution model used for the non-linear outer loop integration. The low-resolution analysis correction computed in the inner loop is interpolated back to the full resolution model (Fig. 3). The use of lower resolution in the minimization algorithm of the inner loop results in substantial savings in computational cost compared with a full resolution minimization, but it also degrades the analysis.

Yang *et al.* (2009b) took advantage that in the LETKF the analysis ensemble members are a weighted combination of the forecasts, and that the analysis weights \mathbf{W}^a are much smoother (they vary on a much larger scale) than the analysis increments or the analysis fields themselves. They tested the idea of interpolating the weights but using the full resolution forecast model on the same quasi-geostrophic model discussed before. They performed full resolution analyses and compared the results with a computation of the LETKF analysis (*i.e.*, the weight matrix \mathbf{W}^a) on coarser grids, every 3×3 , 5×5 and 7×7 grid points, corresponding to an analysis grid coverage of 11%, 4% and 2%, respectively, as well as interpolating the analysis increments. They found that interpolating the weights did not degrade the analysis compared with the full resolution, whereas interpolating the analysis increments resulted in a serious degradation (Fig. 5).

The use of a symmetric square-root in the LETKF ensures that the interpolated analysis has the same linear conservation properties as the full resolution analysis. The results suggest that interpolating the analysis weights computed on a coarse grid without degrading the analysis can substantially reduce the computational cost of the LETKF. Although the full resolution ensemble forecasts are still required, they are also needed for ensemble forecasting in operational centres.

We note that the fact that the weights vary on large scales, and that the use of a coarser analyses with weight interpolation actually improves slightly the analysis in data sparse regions, suggest that smoothing the weights is a good approach to filling data gaps such as those that appear in between satellite orbits. (Yang *et al.* 2009b; Lars Nerger, *pers. comm.*). Smoothing the weights, both in the horizontal and in the vertical may also reduce sampling errors and increase the accuracy of the EnKF analyses.

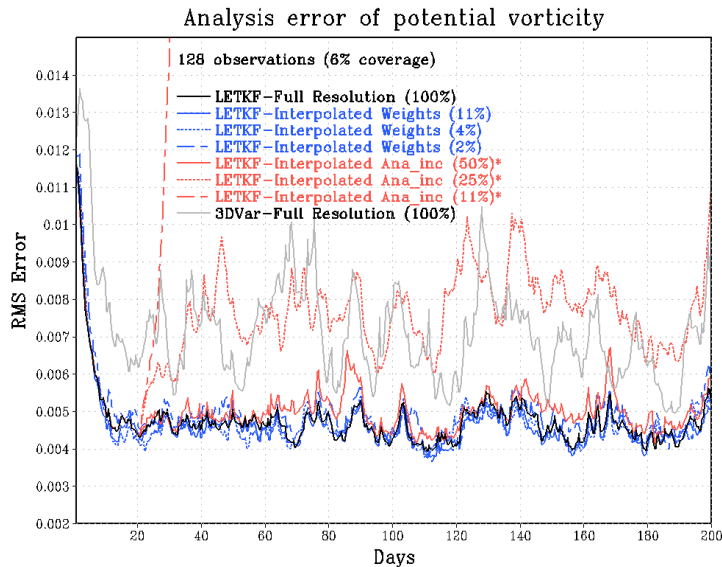


Fig. 5. The time series of the RMS analysis error in terms of the potential vorticity from different data assimilation experiments. The LETKF analysis from the full-resolution is denoted as the black line and the 3D-Var derived at the same resolution is denoted as the grey line. The LETKF analyses derived from weight-interpolation with different analysis coverage are indicated with blue lines. The LETKF analyses derived after the first 20 days from increment-interpolation with different analysis coverage are indicated with the red lines. Adapted from Yang *et al.* (2009b).

3.6 Model and observational error

Model error can seriously affect the EnKF because, among other reasons, the presence of model biases cannot be detected by the EnKF original formulation, and the ensemble spread is the same with or without model bias (Li 2007). For this reason, the most widely used method for imperfect models is to increase the multiplicative or additive inflation (*e.g.* Whitaker *et al.* 2007). Model biases can also be taken into account by estimating the bias as in Dee and da Silva (1998) or its simplified approximation (Radakovich *et al.* 2001) — see also chapter *Bias Estimation* (Ménard). More recently, Baek *et al.* (2007) pointed out that model bias could be estimated accurately augmenting the model state with the bias, and allowing the error covariance to eventually correct the bias. Because in this study the bias was assumed to be a full resolution field, this required doubling the number of ensemble members in order to reach convergence.

In the standard 4D-Var, the impact of model bias cannot be neglected within longer windows because the model (assumed to be perfect) is used as a strong constraint in the minimization (*e.g.* Andersson *et al.* 2005). Trémolet (2007) has developed several techniques allowing for the model to be a weak constraint in order to estimate and correct model errors. Although the results are promising, the methodology for the weak constraint is complex, and still under development.

Li (2007, 2009a) compared several methods to deal with model bias (Fig. 6), including a “Low-dimensional” method based on an independent estimation of the

bias from averages of 6-hour estimated forecast errors started from a reanalysis (or any other available good quality analysis). This method was applied to the SPEEDY (Simplified Parameterizations primitive-Equation Dynamics) model assimilating simulated observations from the NCEP-NCAR (National Centers for Environmental Prediction-National Center for Atmospheric Research) reanalysis, and it was found to be able not only to estimate the bias, but also the errors in the diurnal cycle and the model forecast errors linearly dependent on the state of the model (Danforth *et al.* 2007; Danforth and Kalnay 2008).

The results obtained by Li (2009a) accounting for model errors within the LETKF, presented in Fig. 6, indicate that: a) additive inflation is slightly better than multiplicative inflation; and b) methods to estimate and correct model bias (*e.g.* Dee and da Silva 1998; Danforth *et al.* 2007) should be combined with inflation, which is more appropriate in correcting random model errors. The combination of the low-dimensional method with additive inflation gave the best results, and was substantially better than the results obtained assuming a perfect model (Fig. 6).

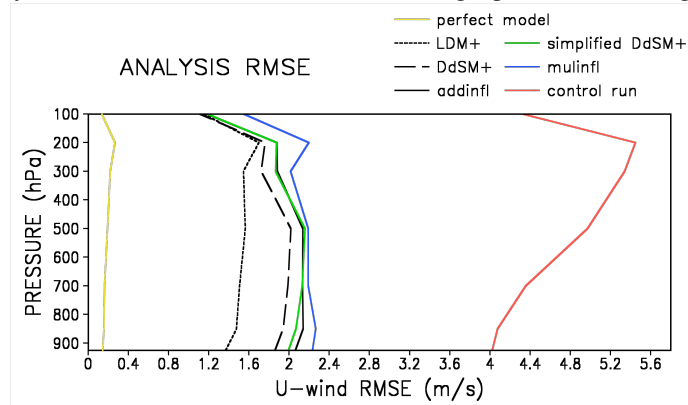


Fig. 6. Comparison of the analysis error averaged over two months for the zonal velocity in the SPEEDY model for several simulations with the radiosonde observations available every other point. The yellow line corresponds to a perfect model simulation with the observations extracted from a SPEEDY model “nature run” (see chapter *Observing System Simulation Experiments*, Masutani *et al.*). The red is the control run, in which the observations were extracted from the NCEP-NCAR reanalysis, but the same multiplicative inflation was used as in the perfect model case. The blue line and the black solid lines correspond to the application of optimized multiplicative and additive inflation, respectively. The long-dashed line was obtained correcting the bias with the Dee and da Silva (1998) method, and combining it with additive inflation. The short-dashed is as the long-dashed but using the Danforth *et al.* (2007) low-dimensional method to correct the bias, and the green line is as the long-dashed line but using the simplified Dee and da Silva method. Adapted from Li (2007).

We note that the approach of Baek *et al.* (2007) of correcting model bias by augmenting the state vector with the bias can be used to determine other parameters, such as surface fluxes, observational bias, nudging coefficients, *etc.* It is similar to increasing the control vector in the variational approach, and is only limited by the number of degrees of freedom that are added to the control vector and sampling errors in the augmented background error covariance.

With respect to observation error estimations, Desroziers *et al.* (2005) derived statistical relationships between products of observations minus background, observations minus analysis, and analysis minus forecasts and the background and observational error covariances. Li (2007) took advantage of these relationships to develop a method to adaptively estimate both the observation errors variance and the optimal inflation of the background error covariance. This method has been successfully tested in several models (Li *et al.* 2009b; Reichle *et al.* 2008).

4 Summary and discussion

4D-Var and the EnKF are the most advanced methods for data assimilation. 4D-Var has been widely adopted in operational centres, with excellent results and much accumulated experience. EnKF is less mature, and has the disadvantage that the corrections introduced by observations are done in spaces of lower dimension that depend on the ensemble size, although this problem is ameliorated by the use of localization. The main advantages of the EnKF are that it provides an estimate of the forecast and analysis error covariances, and that it is much simpler to implement than 4D-Var. A recent WWRP/THORPEX Workshop in Buenos Aires, 10-13 November 2008, was dedicated to 4D-Var and Ensemble Kalman Filter Inter-comparisons with many papers and discussions (<http://4dvarenkf.cima.fcen.uba.ar/>). Buehner *et al.* (2008) presented “clean” comparisons between the operational 4D-Var and EnKF systems in Environment Canada, using the same model resolution and observations, showing that their forecasts had essentially identical scores in the Northern Hemisphere, whereas a hybrid system based on 4D-Var but using a background error covariance based on the EnKF gave a 10-hour improvement in the 5-day forecasts in the Southern Hemisphere. This supports the statement that the best approach should be a hybrid that combines “the best characteristics” of both EnKF and 4D-Var (*e.g.* Lorenc 2003; Barker 2008). This would also bring the main disadvantage of 4D-Var to the hybrid system, *i.e.*, the need to develop and maintain an adjoint model. This makes the hybrid approach attractive to operational centres that already have appropriate linear tangent and adjoint models, but less so to other centres.

In this review we have focused on the idea that the advantages and new techniques developed over the years for 4D-Var, can be adapted and implemented within the EnKF framework, without requiring an adjoint model. The LETKF (Hunt *et al.* 2007) was used as a prototype of the EnKF. It belongs to the square-root or deterministic class of the EnKF (*e.g.* Whitaker and Hamill 2002) but simultaneously assimilates observations locally in space, and uses the ensemble transform approach of Bishop *et al.* (2001) to obtain the analysis ensemble as a linear combination of the background forecasts.

We showed how the LETKF could be modified to include some of the most important 4D-Var advantages. In particular, the 3D-LETKF or 4D-LETKF can be used as a smoother that is cost-free beyond the computation of the filter and storing the weights. This allows a faster spin-up in the “running in place” method, so that the LETKF spins up as fast as 4D-Var. This is important in situations such as the forecast of severe storms, which cannot wait for a slow ensemble spin-up. Long assimilation windows and the consequent non-linearity of the perturbations typically result in non-Gaussianity of the ensemble perturbations and, as a result, a poorer performance of LETKF compared to 4D-Var. The no-cost smoothing method can be used to perform the equivalent of the 4D-Var “outer loop” and help deal with the problem of non-linearity. One of the most powerful applications of the adjoint model is the ability to estimate the impact of a class of observations on the short range forecast (Langland and Baker 2004). Liu and Kalnay (2008) have shown how to

perform the same “adjoint sensitivity” within the LETKF without an adjoint model. Yang *et al.* (2009b) showed that the analysis weights created by the LETKF vary smoothly on horizontal scales much larger than the analyses or the analysis increments, so that the analyses can be performed on a very coarse grid and the weights interpolated to the full resolution grid. Because these weights are applied to the full resolution model, Yang *et al.* (2009b) found that the weight interpolation from a coarse resolution grid did not degrade the analysis, suggesting that the weights vary on large scales, and that smoothing the weights can increase the accuracy of the analysis. Li *et al.* (2009b) compared several methods used to correct model errors and showed that it is advantageous to combine methods that correct the bias, such as that of Dee and da Silva (1998) and the low-dimensional method of Danforth *et al.* (2007), with methods like inflation that are more appropriate to account for random model errors. This is an alternative to the weak constraint method (Trémolet 2007) to deal with model errors in 4D-Var, and involves the addition of a relatively small number of degrees of freedom to the control vector. Li *et al.* (2009a) also showed how observation errors and background error inflation can be estimated adaptively within EnKF.

In summary, we have emphasized that the EnKF can profit from the methods and improvements that have been developed in the wide research and operational experience acquired with 4D-Var. Given that operational tests comparing 4D-Var and the LETKF indicate that the performance of these two methods is already very close (*e.g.* Miyoshi and Yamane 2007; Buehner *et al.* 2008), and that the LETKF and other EnKF methods are simpler to implement, their future looks bright. For centres that have access to the model adjoint, hybrid 4D-Var-EnKF may be optimal.

Acknowledgments

I want to thank the members of the Chaos-Weather group at the University of Maryland, and in particular to Profs. Shu-Chih Yang, Brian Hunt, Kayo Ide, Eric Kostelich, Ed Ott, Istvan Szunyogh, and Jim Yorke. My deepest gratitude is to my former students at the University of Maryland, Drs. Matteo Corazza, Chris Danforth, Hong Li, Junjie Liu, Takemasa Miyoshi, Malaquías Peña, Shu-Chih Yang, Ji-Sun Kang, Matt Hoffman, and present students Steve Penny, Steve Greybush, Tamara Singleton, Javier Amezcua and others, whose creative research allowed us to learn together. Interactions with the thriving Ensemble Kalman Filter community members, especially Ross Hoffman, Jeff Whitaker, Craig Bishop, Kayo Ide, Joaquim Ballabrera, Jidong Gao, Zoltan Toth, Milija Zupanski, Tom Hamill, Herschel Mitchell, Peter Houtekamer, Chris Snyder, Fuqing Zhang and others, as well as with Michael Ghil, Arlindo da Silva, Jim Carton, Dick Dee, and Wayman Baker, have been a source of inspiration. Richard Ménard, Ross Hoffman, Kayo Ide, Lars Nerger and William Lahoz made important suggestions that improved the review and my own understanding of the subject.

References

- Anderson, J.L., 2001. An Ensemble Adjustment Kalman Filter for Data Assimilation. *Mon. Weather Rev.*, **129**, 2884-2903.
- Andersson, E., M. Fisher, E. Hólm, L. Isaksen, G. Radnoti and Y. Trémolet, 2005. Will the 4D-Var approach be defeated by nonlinearity? *ECMWF Tech Memo 479*.
- Baek, S.-J. B.R. Hunt, E. Kalnay, E. Ott and I. Szunyogh, 2006. Local ensemble Kalman filtering in the presence of model bias. *Tellus A*, **58**, 293–306.

- Barker, D.M., 2008. How 4DVar can benefit from or contribute to EnKF (a 4DVar perspective). Available from http://4dvarenkf.cima.fcen.uba.ar/Download/Session_8/4DVar_EnKF_Barker.pdf.
- Bishop, C.H., B.J. Etherton and S.J. Majumdar, 2001. Adaptive Sampling with Ensemble Transform Kalman Filter. Part I: Theoretical Aspects. *Mon. Weather Rev.*, **129**, 420-436.
- Buehner, M., C. Charente, B. He, *et al.*, 2008. Intercomparison of 4-D Var and EnKF systems for operational deterministic NWP. Available from http://4dvarenkf.cima.fcen.uba.ar/Download/Session_7/Intercomparison_4D-Var_EnKF_Buehner.pdf.
- Burgers, G., P.J. van Leeuwen and G. Evensen, 1998. On the analysis scheme in the Ensemble Kalman Filter. *Mon. Weather Rev.*, **126**, 1719-1724.
- Cardinali, C., S. Pezzulli and E. Andersson, 2004. Influence-matrix diagnostic of a data assimilation system. *Q. J. R. Meteorol. Soc.*, **130**, 2767-2786.
- Caya, A., J. Sun and C. Snyder, 2005. A comparison between the 4D-VAR and the ensemble Kalman filter techniques for radar data assimilation. *Mon. Weather Rev.*, **133**, 3081-3094.
- Corazza, M., E. Kalnay, D. J. Patil, S.-C. Yang, R. Morss, M. Cai, I. Szunyogh, B.R. Hunt and J.A. Yorke, 2003. Use of the breeding technique to estimate the structure of the analysis "error of the day". *Nonlinear Processes in Geophysics*, **10**, 233-243.
- Corazza, M., E. Kalnay and S.-C. Yang, 2007. An implementation of the Local Ensemble Kalman Filter in a quasigeostrophic model and comparison with 3D-Var. *Nonlinear Processes in Geophysics*, **14**, 89-101.
- Courtier, P. and O. Talagrand, 1990. Variational assimilation of meteorological observations with the direct and adjoint shallow water equations. *Tellus*, **42A**, 531-549.
- Danforth, C.M. and E. Kalnay, 2008. Using Singular Value Decomposition to parameterize state dependent model errors. *J. Atmos. Sci.*, **65**, 1467-1478.
- Danforth, C.M., E. Kalnay and T. Miyoshi, 2006. Estimating and Correcting Global Weather Model Error. *Mon. Weather Rev.*, **134**, 281-299.
- Dee, D.P. and A.M. da Silva, 1998. Data assimilation in the presence of forecast bias. *Q. J. R. Meteorol. Soc.*, **124**, 269-295.
- Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10,143-10,162.
- Evensen, G., 1997. Advanced data assimilation for strongly nonlinear dynamics. *Mon. Weather Rev.*, **125**, 1342-1354.
- Evensen, G., 2003. The ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343-367.
- Evensen, G. and P.J. van Leeuwen, 1996. Assimilation of Geosat altimeter data for the Agulhas current using the ensemble Kalman Filter with a quasi-geostrophic model. *Mon. Weather Rev.*, **124**, 85-96.
- Fertig, E., J. Harlim and B. Hunt, 2007a. A comparative study of 4D-Var and 4D Ensemble Kalman Filter: Perfect Model Simulations with Lorenz-96. *Tellus*, **59**, 96-101.
- Fisher, M., M. Leutbecher and G. Kelly, 2005. On the equivalence between Kalman smoothing and weak-constraint four-dimensional variational data assimilation. *Q. J. R. Meteorol. Soc.*, **131**, 3235-3246.
- Gaspari, G. and S.E. Cohn, 1999. Construction of correlation functions in two and three dimensions. *Q. J. R. Meteorol. Soc.*, **125**, 723-757.
- Ghil, M. and P. Malanotte-Rizzoli, 1991. Data assimilation in meteorology and oceanography. *Adv. Geophys.*, **33**, 141-266.
- Greybush, S., E. Kalnay, T. Miyoshi, K. Ide and B. Hunt, 2009: EnKF Localization Techniques and Balance. Presented at the WMO 5th International Symposium on Data Assimilation. Melbourne, Australia, 6-9 October 2009, submitted to *Mon. Weather Rev.*, available at http://www.weatherchaos.umd.edu/papers/Greybush_Melbourne2009.ppt.

- Gustafsson, N., 2007. Response to the discussion on “4-D-Var or EnKF?”. *Tellus A*, **59**, 778-780.
- Hamill, T.M., J.S. Whitaker and C. Snyder, 2001. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Weather Rev.*, **129**, 2776–2790.
- Harlim, J. and B.R. Hunt, 2007a. A non-Gaussian ensemble filter for assimilating infrequent noisy observations. *Tellus A*, **59**, 225-237.
- Harlim, J. and B.R. Hunt, 2007b. Four-dimensional local ensemble transform Kalman filter: variational formulation and numerical experiments with a global circulation model. *Tellus A*, **59**, 731-748.
- Houtekamer, P.L. and H.L. Mitchell, 1998. Data Assimilation Using an Ensemble Kalman Filter Technique. *Mon. Weather Rev.*, **126**, 796-811.
- Houtekamer, P.L. and H.L. Mitchell, 2001. A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation. *Mon. Weather Rev.*, **129**, 123–137.
- Houtekamer, P.L., H.L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek and B. Hansen, 2005. Atmospheric Data Assimilation with an Ensemble Kalman Filter: Results with Real Observations. *Mon. Weather Rev.*, **133**, 604-620.
- Houtekamer, P.L. and H.L. Mitchell, 2005. Ensemble Kalman filtering. *Q. J. R. Meteorol. Soc.*, **131**, 3269-3290.
- Hunt, B.R., 2005. An efficient implementation of the local ensemble Kalman filter. Available at <http://arxiv.org/abs/physics/0511236>.
- Hunt, B.R., E. Kalnay, E.J. Kostelich, *et al.*, 2004. Four-dimensional ensemble Kalman filtering. *Tellus*, **56A**, 273-277.
- Hunt, B.R., E.J. Kostelich and I. Szunyogh, 2007. Efficient data assimilation for spatiotemporal chaos: a Local Ensemble Transform Kalman Filter. *Physica D*, **230**, 112-126.
- Ide, K., P. Courtier, M. Ghil and A. Lorenc, 1997. Unified notation for data assimilation: operational, sequential and variational. *J. Meteor. Soc. Jpn.*, **75**, 181-189.
- Järvinen, H., E. Andersson and F. Bouttier, 1999. Variational assimilation of time sequences of surface observations with serially correlated errors. *Tellus*, **51A**, 469-488.
- Jazwinski, A.H., 1970. *Stochastic processes and filtering theory*. Academic Press, 376 pp.
- Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. *J. Basic Engineering*, **82**, 35-45.
- Kalnay, E., 2003. *Atmospheric modeling, data assimilation and predictability*. Cambridge University Press, Cambridge, 341 pp.
- Kalnay, E., H. Li, T. Miyoshi, S.-C. Yang and J. Ballabrera-Poy, 2007a. 4D-Var or Ensemble Kalman Filter? *Tellus A*, **59**, 758-773.
- Kalnay, E., H. Li, T. Miyoshi, S.-C. Yang and J. Ballabrera-Poy, 2007b. Response to the Discussion on “4D-Var or EnKF?” by Nils Gustafsson. *Tellus A*, **59**, 778-780.
- Kalnay, E. and S-C. Yang, 2008. Accelerating the spin-up in EnKF. *Arxiv: physics:Nonlinear/0.806.0180v1*.
- Keppenne, C.L., 2000. Data assimilation into a primitive-equation model with a parallel ensemble Kalman filter. *Mon. Weather Rev.*, **128**, 1971-1981.
- Keppenne, C. and Rienecker, H. 2002. Initial testing of a massively parallel ensemble Kalman filter with the Poseidon isopycnal ocean general circulation model. *Mon. Weather Rev.*, **130**, 2951-2965.
- Langland, R.H. and N.L. Baker, 2004. Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. *Tellus*, **56A**, 189-201.
- Li, H., 2007. Local ensemble transform Kalman filter with realistic observations. Ph. D. thesis. Available at <http://hdl.handle.net/1903/7317>.
- Li H., E. Kalnay, T. Miyoshi and C.M. Danforth 2009a. Accounting for model errors in ensemble data assimilation. *Mon. Weather Rev.*, **137**, 3407-3419.

- Li, H., E. Kalnay and T. Miyoshi, 2009b. Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Q. J. R. Meteorol. Soc.*, **135**, 523-533.
- Liu, J. and E. Kalnay, 2008. Estimating observation impact without adjoint model in an ensemble Kalman Filter. *Q. J. R. Meteorol. Soc.*, **134**, 1327-1335.
- Liu, J., E. Kalnay, T. Miyoshi and C. Cardinali, 2009: Analysis sensitivity calculation in an ensemble Kalman Filter. *Q. J. R. Meteorol. Soc.*, **135**, in press.
- Lorenc, A.C., 1986. Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.*, **112**, 1177-1194.
- 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.
- Lorenz, E., 1963. Deterministic Non-periodic Flow. *J. Atmos. Sci.*, **20**, 130-141.
- Mitchell, H.L., P.L. Houtekamer and G. Pellerin, 2002. Ensemble Size, Balance, and Model-Error Representation in an Ensemble Kalman Filter. *Mon. Weather Rev.*, **130**, 2791-2808.
- Miyoshi, T., 2005. Ensemble Kalman filter experiments with a Primitive-Equation global model. Doctoral dissertation, University of Maryland, College Park, 197 pp. Available at <https://drum.umd.edu/dspace/handle/1903/3046>.
- Miyoshi, T. and S. Yamane, 2007. Local ensemble transform Kalman filtering with an AGCM at a T159/L48 resolution. *Mon. Weather Rev.*, **135**, 3841-3861.
- Molteni, F., 2003. Atmospheric simulations using a GCM with simplified physical parameterizations. I: model climatology and variability in multi-decadal experiments. *Clim. Dyn.*, **20**, 175-191.
- Nerger, L., W. Hiller and J. Scroeter, 2005. A comparison of error subspace Kalman filters. *Tellus*, **57A**, 715-735.
- Nutter, P., M. Xue and D. Stensrud, 2004. Application of lateral boundary condition perturbations to help restore dispersion in limited-area ensemble forecasts. *Mon. Weather Rev.*, **132**, 2378-2390.
- Ott, E., B.R. Hunt, I. Szunyogh, A.V. Zimin, E.J. Kostelich, M. Corazza, E. Kalnay, D.J. Patil and J.A. Yorke, 2004. A local ensemble Kalman filter for atmospheric data assimilation. *Tellus*, **56A**, 415-428.
- Patil, D.J., B.R. Hunt, E. Kalnay, J.A. Yorke and E. Ott, 2001. Local low dimensionality of atmospheric dynamics. *Phys. Rev. Lett.*, **86**, 5878-5881.
- Pham, D.T., 2001. Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Mon. Weather Rev.*, **129**, 1194-1207.
- Pires, C., R. Vautard and O. Talagrand, 1996. On extending the limits of variational assimilation in chaotic systems. *Tellus*, **48A**, 96-121.
- Rabier, F., H. Järvinen, E. Klinker, J.-F. Mahfouf and A. Simmons, 2000. The ECMWF operational implementation of four-dimensional variational physics. *Q. J. R. Meteorol. Soc.*, **126**, 1143-1170.
- Radakovich, J.D., P.R. Houser, A.M. da Silva and M.G. Bosilovich, 2001. Results from global land-surface data assimilation methods. *Proceeding of the fifth symposium on integrated observing systems*, 14-19 January 2001, Albuquerque, NM, pp 132-134.
- Reichle, R.H., W.T. Crow and C.L. Keppenne, 2008. An adaptive ensemble Kalman filter for soil moisture data assimilation. *Water Resour. Res.*, **44**, W03423, doi:10.1029/2007WR006357.
- Rotunno, R. and J.W. Bao, 1996. A case study of cyclogenesis using a model hierarchy. *Mon. Weather Rev.*, **124**, 1051-1066.
- Szunyogh, I., E. Kostelich, G. Gyarmati, E. Kalnay, B.R. Hunt, E. Ott, E. Satterfield and J.A. Yorke, 2008. Assessing a local Ensemble Kalman Filter: Assimilating Real Observations with the NCEP Global Model. *Tellus*, in press.
- Szunyogh, I., E.J. Kostelich, G. Gyarmati, D.J. Patil, B.R. Hunt, E. Kalnay, E. Ott, and J.A. Yorke, 2005. Assessing a local ensemble Kalman filter: Perfect model experiments with

- the NCEP global model. *Tellus*, **57A**, 528-545.
- Talagrand O. and P. Courtier, 1987. Variational assimilation of meteorological observations with the adjoint vorticity equation I: theory. *Q. J. R. Meteorol. Soc.*, **113**, 1311–1328.
- Thépaut, J.-N. and P. Courtier, 1991. Four-dimensional data assimilation using the adjoint of a multilevel primitive equation model. *Q. J. R. Meteorol. Soc.*, **117**, 1225-1254.
- Tippett, M.K., J.L. Anderson, C.H. Bishop, T.M. Hamill and J.S. Whitaker, 2003. Ensemble Square Root Filters. *Mon. Weather Rev.*, **131**, 1485-1490.
- Torn, R.D., G.J. Hakim and C. Snyder, 2006. Boundary Conditions for Limited-Area Ensemble Kalman Filters. *Mon. Weather Rev.*, **134**, 2490-2502.
- Trémolet, Y., 2007. Model-error estimation in 4D-Var. *Q. J. R. Meteorol. Soc.*, **133**, 1267–1280.
- Wang, X., C.H. Bishop and S.J. Julier, 2004. Which is better, an ensemble of positive-negative pairs or a centered spherical simplex ensemble? *Mon. Weather Rev.*, **132**, 1590–1605.
- Whitaker, J.S., G.P. Compo, X. Wei and T.M. Hamill, 2004. Reanalysis without radiosondes using ensemble data assimilation. *Mon. Weather Rev.*, **132**, 1190–1200.
- Whitaker, J.S. and T.M. Hamill, 2002. Ensemble data assimilation without perturbed observations. *Mon. Weather Rev.*, **130**, 1913-1924.
- Whitaker, J.S., T.M. Hamill, X. Wei, Y. Song and Z. Toth, 2008. Ensemble data assimilation with the NCEP global forecast system. *Mon. Weather Rev.*, **136**, 463-482.
- Yang, S-C, M. Corazza, A. Carrassi, E. Kalnay and T. Miyoshi, 2009a. Comparison of ensemble-based and variational-based data assimilation schemes in a quasi-geostrophic model. *Mon. Weather Rev.*, **137**, 693-709.
- Yang, S-C, E. Kalnay, B. Hunt and N. Bowler, 2009b. Weight interpolation for efficient data assimilation with the Local Ensemble Transform Kalman Filter. *Q. J. R. Meteorol. Soc.*, **135**, 251-262.
- Zhang, F., C. Snyder and J. Sun, 2004. Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Weather Rev.*, **132**, 1238–1253.
- Zhu, Y. and R. Gelaro, 2008. Observation Sensitivity Calculations Using the Adjoint of the Gridpoint Statistical Interpolation (GSI) Analysis System. *Mon. Weather Rev.*, **136**, 335–351.
- Zupanski, M., 2005. Maximum likelihood ensemble filter: theoretical aspects. *Mon. Weather Rev.*, **133**, 1710-1726.
- Zupanski M., S.J. Fletcher, I.M. Navon, *et al.*, 2006. Initiation of ensemble data assimilation. *Tellus*, **58A**, 159-170.