Tellus **000**, 000–000 (0000)   Printed 22 August 2009   (Tellus LaTeX style file v2.2)

On the Equivalence Between Suboptimal 4D-Var and EnKF

Adrian Sandu⋆ , Haiyan Cheng
Computational Science Laboratory
Department of Computer Science
Virginia Polytechnic Institute and State University
2201 Knowledgeworks II, 2202 Kraft Drive, Blacksburg, VA 24060, USA
E-mail: sandu@cs.vt.edu

22 August 2009

ABSTRACT
Two families of methods are widely used in data assimilation: the four dimensional variational (4D-Var) approach, and the ensemble Kalman filter (EnKF) approach. The two families have been developed largely through parallel research efforts, and each method has its advantages and disadvantages. It is of interest to combine the two approaches and develop hybrid data assimilation algorithms. This paper investigates the theoretical equivalence between the suboptimal 4D-Var method (where only a small number of optimization iterations are performed) and the practical EnKF method (where only a small number of ensemble members are used) in a linear Gaussian setting. The analysis motivates a new hybrid algorithm: the optimization directions obtained from a short window 4D-Var run are used to construct the EnKF initial ensemble. Numerical results show that the proposed hybrid ensemble filter method performs better than the regular EnKF method for both linear and nonlinear test problems.
Keywords: Data assimilation, variational methods, ensemble filters, hybrid methods.

1 Introduction

Data assimilation (DA) is a procedure to combine imperfect model predictions with imperfect observations in order to produce coherent estimates of the evolving state of the system, and to improve the ability of models to represent reality. DA is accomplished through inverse analysis by estimating initial, boundary conditions, and model parameters. It has become an essential tool for weather forecasts, climate studies, and environmental analyses.

Two data assimilation methodologies are currently widely used: variational and ensemble filters (Bennett, 2002; Daley, 1991; Evensen, 2007; Kalnay, 2003; Lewis et al., 2005; Rodgers, 2000). While both methodologies are rooted in statistical estimation theory, their theoretical developments and practical implementations have distinct histories. The four dimensional variational (4D-Var) methodology has been used extensively in operational weather prediction centers. In traditional (strong-constrained) 4D-Var a perfect model is assumed; the analysis provides the single trajectory that best fits the background state and all the observations in the assimilation window (Talagrand and Courtier, 1987). The 4D-Var requires the solution of a numerical optimization problem, with gradients provided by an adjoint model; the necessity of maintaining an adjoint model is the main disadvantage of 4D-Var. The ensemble Kalman filter (EnKF) is based on Kalman's seminal work (Kalman, 1960) but uses a Monte Carlo approach to propagate error covariances through the model dynamics. The EnKF corrections are computed in a low dimensional subspace (spanned by the ensemble) and therefore the EnKF analyses are inherently suboptimal. Nevertheless, EnKF performs well in many practical situations Anderson (2003), is easy to implement, and naturally provides estimates of the analysis covariances.

It is known that the fully resolved variational method and the optimal Kalman filter technique compute the same estimate of the posterior mean for linear systems, linear observation operators, and Gaussian uncertainty (Li and Navon, 2001). For very long assimilation windows the 4D-Var analysis at the end of the window is similar to the one produced by running a Kalman filter indefinitely Fisher et al. (2005). In the presence of model errors the weak-constrained 4D-Var and the fixed-interval Kalman smoother are equivalent (Ménard and Daley, 1996). With both methods coming to maturity, new interest in the community has been devoted to assess the relative merits of 4D-Var and EnKF(Kalnay, 2007; Lorenc, 2003). The better understanding of the strengths of each method has opened the possibility to combine them and build hybrid data assimilation methods; relevant work can be found in Bewley et al. (2009); Evensen and van Leeuwen (2000); Hamill and Snyder (2000); Harlim and Hunt (2004); Hunt et al. (2007); Liu

⋆ Corresponding author.
e-mail: sandu@cs.vt.edu

at al. (2008); Tian et al. (2008); Torn and Hakim (2008); Wang et al. (2007); Zhang et al. (2007); Zupanski (2005).

Little attention has been devoted to analyzing the practical situation where only a small number of optimization iterations is performed in 4D-Var, and only a small ensemble is used in EnKF. In this paper we study the relationship between the suboptimal 4D-Var and the practical EnKF methods in a linear Gaussian setting. The close relationship between 4D-Var and EnKF opens the possibility of combining these two approaches, and motivates a new hybrid data assimilation algorithm.

To be specific, consider a forward model that propagates the initial model state $\mathbf{x}(t_0) \in \mathbb{R}^n$ to a future state $\mathbf{x}(t) \in \mathbb{R}^n$.

$$\mathbf{x}(t) = \mathcal{M}_{t_0 \to t}\ (\mathbf{x}(t_0))\ , \quad t_0 \leqslant t \leqslant t_\mathrm{F}. \tag{1}$$

The model solution operator $\mathcal{M}$ represents, for example, a discrete approximation of the partial differential equations that govern the atmospheric or oceanic processes. Realistic atmospheric and ocean models typically have $n \sim 10^7 - 10^9$ variables. Perturbations (small errors $\delta\mathbf{x}$) may be simultaneously evolved according to the tangent linear model:

$$\delta\mathbf{x}(t) = \mathbf{M}_{t_0 \to t}\ (\delta\mathbf{x}(t_0))\ , \quad t_0 \leqslant t \leqslant t_\mathrm{F}. \tag{2}$$

We consider the case where the initial model state is uncertain and a better state estimate is sought for. The model (1) simulation from t_0 to t_n is initialized with a background (prior estimate) $\mathbf{x}_0^B$ of the true atmospheric state $\mathbf{x}_0^t$. The background errors (uncertainties) are assumed to have a normal distribution $(\mathbf{x}_0^B - \mathbf{x}_0^t) \in \mathcal{N}(0, \mathbb{B})$. The background represents the best estimate of the true state *prior* to any measurement being available.

Observations of the state $\mathbf{y}_k = \mathcal{H}_k(\mathbf{x}_k) + \varepsilon_k$ are available at each time instant t_k, $k = 0, \ldots, N_\mathrm{obs} - 1$. These observations are corrupted by measurement and representative errors, which are assumed to have a normal distribution, $\varepsilon_k \in \mathcal{N}(0, \mathbb{R}_k)$. Data assimilation combines the background estimate $\mathbf{x}_0^B$, the measurements $\mathbf{y}_0, \cdots, \mathbf{y}_{N_\mathrm{obs}-1}$, and the model $\mathcal{M}$ to obtain an improved estimate $\mathbf{x}_0^A$ of the true initial state $\mathbf{x}_0^t$. This improved estimate is called the "analysis" (or posterior estimate of the) state.

The four dimensional variational (4D-Var) technique is derived from variational calculus and control theory (**?**). It provides the analysis $\mathbf{x}_0^A$ as the argument which minimizes the cost function:

$$\begin{aligned}
\mathcal{J}(\mathbf{x}_0) &= \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_0^B)^T\, \mathbb{B}^{-1}\, (\mathbf{x}_0 - \mathbf{x}_0^B) \\
&+ \frac{1}{2} \sum_{k=0}^{N_\mathrm{obs}-1} \left(\mathcal{H}_k(\mathbf{x}_k) - \mathbf{y}_k\right)^T\, \mathbb{R}_k^{-1}\, \left(\mathcal{H}_k(\mathbf{x}_k) - \mathbf{y}_k\right) \\
\text{s.t. } \mathbf{x}_k &= \mathcal{M}_{t_0 \to t_k}(\mathbf{x}_0)\,.
\end{aligned} \tag{3}$$

Typically, a gradient-based optimization procedure is used to solve the constrained optimization problem (3) with gradients obtained by adjoint modeling.

In the incremental formulation of 4D-Var (Bennett, 2002; Lewis et al., 2005), one linearizes the estimation problem around the background trajectory. By expressing the state as $\mathbf{x}_k = \mathbf{x}_k^B + \delta\mathbf{x}_k$, $k = 0, \cdots, N_\mathrm{obs} - 1$, we have

$$\begin{aligned}
\mathcal{J}'(\delta\mathbf{x}_0) &= \frac{1}{2}\, \delta\mathbf{x}_0{}^T\, \mathbb{B}^{-1}\, \delta\mathbf{x}_0 \\
&+ \frac{1}{2} \sum_{k=0}^{N_\mathrm{obs}} \left(H_k \delta\mathbf{x}_k - d_k^B\right)^T \mathbb{R}_k^{-1} \left(H_k \delta\mathbf{x}_k - d_k^B\right)\,, \\
d_k^B &= \mathcal{H}_k\left(\mathbf{x}_k^B\right) - \mathbf{y}_k\,,
\end{aligned} \tag{4}$$

where $\delta\mathbf{x}_k = \mathbf{M}_{t_0 \to t_k}(\mathbf{x}_0)$, and H_k is the linearized observational operator around $\mathbf{x}_k^B$ at time t_k. The incremental 4D-Var problem (4) uses linearized operators and leads to a quadratic cost function $\mathcal{J}'$. The incremental 4D-Var estimate is $\mathbf{x}_0^A = \mathbf{x}_0^B + \delta\mathbf{x}_0^A$. A new linearization can be performed about this estimate and the incremental problem (4) can be solved again to improve the resulting analysis.

Ensemble filters are based on the Kalman Filter (Kalman, 1960) theory, which gives an optimal estimate of the true state under the assumption that probability densities of all errors are Gaussian, and the model dynamics and observation operators are all linear. The extended Kalman filter (Fisher, 2002) provides a suboptimal state estimation in the nonlinear case by linearizing the model dynamics and the observation operator.

If the errors in the model state at t_{k-1} have a normal distribution $\mathcal{N}(0, P_{k-1}^A)$ and propagate according to the linearized model dynamics, then the forecast errors at t_k are also normally distributed $\mathcal{N}(0, P_k^f)$. The forecast is obtained using

$$\begin{aligned}
\mathbf{x}_k^f &= \mathcal{M}_{t_{k-1} \to t_k}\left(\mathbf{x}_{k-1}^A\right)\,, \\
P_k^f &= \mathbf{M}_{t_{k-1} \to t_k}\, P_{k-1}^A\, \mathbf{M}_{t_{k-1} \to t_k}^T + Q_k\,,
\end{aligned} \tag{5}$$

where $\mathbf{M}^T$ is the adjoint of the tangent linear model, and Q_k is the covariance matrix of model errors. The analysis provides the state estimate $\mathbf{x}_k^A$ and the corresponding error covariance matrix P_k^A

$$\begin{aligned}
\mathbf{x}_k^A &= \mathbf{x}_k^f + \mathbf{K}_k \left(y_k - \mathcal{H}_k(\mathbf{x}_k^f)\right)\,, \\
P_k^A &= P_k^f - \mathbf{K}_k\, H_k\, P_k^f\,, \\
\mathbf{K}_k &= P_k^f\, H_i^T \left(H_k\, P_k^f\, H_k^T + \mathbb{R}_k\right)^{-1}\,,
\end{aligned} \tag{6}$$

where $\mathbf{K}_k$ is the Kalman gain matrix.

The extended Kalman filter is not practical for large systems because of the prohibitive computational cost needed to invert large matrices and to propagate the covariance matrix in time. Approximations are needed to make the EKF computationally feasible. The ("perturbed observations" version of the) ensemble Kalman filter (Fisher, 2002) uses a Monte-Carlo approach to propagate covariances. An ensemble of N_ens states (labeled $e = 1, \cdots, N_\mathrm{ens}$) is used to sample the probability distribution of the background error. Each member of the ensemble (with state $\mathbf{x}_{k-1}^A(e)$ at t_{k-1}) is propagated to t_k using the nonlinear model (1) to obtain the "forecast" ensemble $\mathbf{x}_k^f(e)$. Gaussian noise $\eta_i \in \mathcal{N}(0, Q_k)$ is added to the forecast to account for the effect of model errors. Each member of the forecast is analyzed separately using the state equation in (6). The forecast and the analysis error covariances (P_k^f and P_k^A) are estimated from the statistical samples ($\{\mathbf{x}_k^f(e)\}_{e=1,\cdots,N_\mathrm{ens}}$ and $\{\mathbf{x}_k^A(e)\}_{e=1,\cdots,N_\mathrm{ens}}$ respectively). The EnKF approach to data assimilation has attracted considerable attention in meteorology (Anderson, 2003; Burgers et al., 1998) due to its many attractive features.

It has been established that the 4D-Var and the EnKF techniques are equivalent for linear systems with Gaussian uncertainty (Li and Navon, 2001), *provided that the 4D-Var solution is computed exactly and an infinitely large number of ensemble members is used in EnKF*. By *equivalent* we mean that the two approaches provide the same estimates of the posterior mean. In practice, the dynamical systems of interest for data assimilation are very large – for example, typical models of the atmosphere have $n \sim 10^7 - 10^9$ variables. As a consequence, the numerical optimization problem in 4D-Var (3) can only be solved approximately, by an iterative procedure stopped after a relatively small number of iterations. Similarly, in an ensemble based approach, the number of ensemble members is typically much smaller than the state space dimension and the sampling is inherently suboptimal. In this work we seek to better understand the relationship between the *suboptimal* 4D-Var solution and the *suboptimal* EnKF solution. This analysis motivates a new hybrid filter algorithm for data assimilation.

The paper is organized as follows. Section 2 performs a theoretical analysis that reveals subtle similarities between the suboptimal 4D-Var and EnKF solutions in the linear Gaussian case, and for one observation time. This analysis motivates a new hybrid filter algorithm for data assimilation, which is discussed in Section 3. Numerical experiments presented in Section 4 reveal that the proposed algorithm performs better than the traditional EnKF for both linear and nonlinear problems.

2 Comparison of Suboptimal 4D-Var and EnKF Solutions in the Linear, Gaussian Case with a Single Observation Time

Consider a linear system that advances the state from t_0 to t_F,

$$\mathbf{x}_F = \mathbf{M} \cdot \mathbf{x}_0 \,.$$

We assume $\mathbf{M}$ to be an invertible matrix (i.e., the system is time-reversible) so that we can uniquely map any solution at t_F to the corresponding solution at t_0, $\mathbf{x}_F = \mathbf{M}^{-1} \cdot \mathbf{x}_0$.

We assume the initial state uncertain, and the prior distribution of uncertainty is Gaussian, $\mathbf{x}_0^t \in \mathcal{N}\left(\mathbf{x}_0^B, \mathbb{B}_0\right)$. Consequently, the uncertainty in the background state at the final time t_F is also Gaussian. The mean background state and the background covariance at final time are

$$\mathbf{x}_F^B = \mathbf{M} \cdot \mathbf{x}_0^B \,, \quad \mathbb{B}_F = \mathbf{M} \cdot \mathbb{B}_0 \cdot \mathbf{M}^T \,.$$

A single set of noisy measurements are taken at t_F

$$\mathbf{y}_F = H \cdot \mathbf{x}_F + \varepsilon_F \,, \quad \varepsilon_F \in \mathcal{N}\left(0, \mathbb{R}_F\right) \,.$$

We consider the assimilation window $[t_0, t_F]$. Under the above assumptions, the posterior distribution of the true state is Gaussian, with mean $\mathbf{x}^A$ and posterior covariance matrix $\mathbb{A}$

$$\mathbf{x}_0^t \in \mathcal{N}\left(\mathbf{x}_0^A, \mathbb{A}_0\right) \,, \quad \mathbf{x}_F^t \in \mathcal{N}\left(\mathbf{x}_F^A, \mathbb{A}_F\right) \,.$$

We use both 4D-Var and EnKF methods to estimate the posterior initial condition $\mathbf{x}_0^A$. Each method is applied in a suboptimal formulation: only a small number of iterations is used to obtain the 4D-Var solution, and only a small number of ensemble members is used in EnKF.

We first state the main result of this section; the detailed analysis and the proof follow.

Theorem 1. Under the assumption that the model is linear and the errors are Gaussian, with observations at only one time, the state estimate computed by the suboptimal 4D-Var method is equivalent to that obtained by the EnKF method with a small number of ensemble members.

2.1 Full 4D-Var Solution

The 4D-Var analysis is obtained as the minimizer of the function:

$$\begin{aligned} J(\mathbf{x}_0) \;&=\; \frac{1}{2}\left(\mathbf{x}_0 - \mathbf{x}_0^B\right)^T \mathbb{B}_0^{-1} \left(\mathbf{x}_0 - \mathbf{x}_0^B\right) \\ &\quad + \frac{1}{2}\left(H\mathbf{M}\mathbf{x}_0 - \mathbf{y}_F\right)^T \mathbb{R}_F^{-1} \left(H\mathbf{M}\mathbf{x}_0 - \mathbf{y}_F\right) \,. \end{aligned}$$

The first order necessary condition $\nabla J = 0$ reveals that the optimum increment is obtained by solving the following linear system:

$$\begin{aligned} A \cdot \Delta\mathbf{x}_0 &= b \\ A &= \left(\mathbb{B}_0^{-1} + \mathbf{M}^T H^T \mathbb{R}_F^{-1} H\mathbf{M}\right) \\ b &= \mathbf{M}^T H^T \mathbb{R}_F^{-1} \left(\mathbf{y}_F - H\mathbf{M}\mathbf{x}_0^B\right) \end{aligned} \tag{7}$$

where the solution is the deviation of the analysis from the background state

$$\Delta\mathbf{x}_0 = \mathbf{x}_0 - \mathbf{x}_0^B \,.$$

The system matrix A in (7) is the inverse of the posterior covariance at t_0 (Gejadze et al., 2008), $A = \mathbb{A}_0^{-1}$. The right hand side vector b in (7) is the increment corresponding to the background state $d_F = \mathbf{y} - H\mathbf{M}\mathbf{x}_0^B$ scaled by the inverse covariance and "pulled back" to t_0 via the adjoint model

$$b = \mathbf{M}^T H^T \mathbb{R}_F^{-1} \, d_F \,.$$

Since $\mathbf{M}$ is invertible, the 4D-Var system (7) can be formulated in terms of the solution at t_F using the relation $\mathbb{B}_0^{-1} = \mathbf{M}^T \mathbb{B}_F^{-1} \mathbf{M}$

$$\begin{aligned} \left(\mathbb{B}_F^{-1} + H^T \mathbb{R}_F^{-1} H\right) \cdot \Delta\mathbf{x}_F &= H^T \mathbb{R}_F^{-1} \left(\mathbf{y}_F - H\mathbf{x}_F^B\right) \\ \Delta\mathbf{x}_F &= \mathbf{x}_F - \mathbf{x}_F^B \,. \end{aligned} \tag{8}$$

2.2 Iterative 4D-Var Solution by the Lanczos Method

In practice (7) is not solved exactly, but is solved in an approximate sense by using an iterative method and performing a number of iterations that is much smaller than the size of the state space n. We are interested in the properties of this suboptimal algorithm. In the nonlinear case a relatively small number of iterations are performed with a numerical optimization algorithm.

Assume that the Lanczos algorithm (Saad, 2003) is employed to solve the symmetric linear system (7). Specifically, K Lanczos iterations are performed from the starting point $\mathbf{x}_0^{[0]} = \mathbf{x}_0^B$, i.e., $\Delta\mathbf{x}_0^{[0]} = 0$. Consequently, the first residual is $r^{[0]} = b$.

The Lanczos method computes a symmetric tridiagonal matrix $T_K \in \mathbb{R}^{K\times K}$ and a second matrix

$$V_K = \left[v_1, \cdots, v_K\right] \in \mathbb{R}^{n\times K}$$

whose columns form an orthonormal basis of the Krylov space

$$\mathcal{K}_K(A, r^{[0]}) \quad = \quad \left\{r^{[0]}, Ar^{[0]}, A^2 r^{[0]}, \cdots, A^{K-1} r^{[0]}\right\} .$$

The matrices have the following properties (Saad, 2003)

$$V_K^T\, V_K = I \;, \quad V_K^T\, A\, V_K = T_K \;.$$

The approximate solution of the system (7) after K iterations is the exact solution of the system reduced over the Krylov subspace $\mathcal{K}_K$,

$$V_K^T\, A\, V_K \cdot \theta = T_K \cdot \theta = V_K^T\, b \;, \quad \Delta\mathbf{x}^{[K]} = V_K\, \theta \,. \qquad (9)$$

The convergence of the Lanczos iterations can be improved via preconditioning. The background covariance is available and offers a popular preconditioner. Assume that a Cholesky or matrix square root decomposition of $\mathbb{B}_0$ is available:

$$\mathbb{B}_0 = \mathbb{B}_0^{1/2} \cdot \mathbb{B}_0^{T/2}.$$

Applying the background covariance square root as a symmetric preconditioner to the system (7) leads to:

$$\mathbb{B}_0^{T/2}\, A\, \mathbb{B}_0^{1/2} \cdot \Delta\mathbf{u} = \mathbb{B}_0^{T/2}\, b \;, \quad \Delta\mathbf{x} = \mathbb{B}_0^{1/2}\, \Delta\mathbf{u}$$

The preconditioned 4D-Var system reads:

$$\left(I + \mathbb{B}_0^{T/2}\, \mathbf{M}^T H^T \mathbb{R}_F^{-1} H \mathbf{M} \mathbb{B}_0^{1/2}\right) \cdot \Delta\mathbf{u} = \mathbb{B}_0^{T/2}\, \mathbf{M}^T H^T \mathbb{R}_F^{-1} \left(\mathbf{y}_F - H\mathbf{M}\mathbf{x}_0^B\right) \qquad (10)$$

The application of Lanczos iterations to the preconditioned system (10) leads to a new Krylov space (for the new matrix and right hand side vector) and a new orthonormal basis $\widetilde{V}_K$. The approximate solution (9) obtained after K iterations reads:

$$\left(I + \widetilde{V}_K^T\, \mathbb{B}_0^{T/2}\, \mathbf{M}^T H^T \mathbb{R}_F^{-1} H \mathbf{M} \mathbb{B}_0^{1/2} \widetilde{V}_K\right) \cdot \widetilde{\theta}_K \qquad (11)$$
$$= \widetilde{V}_K^T\, \mathbb{B}_0^{T/2}\, \mathbf{M}^T H^T \mathbb{R}_F^{-1} \left(\mathbf{y}_F - H\mathbf{M}\mathbf{x}_0^B\right)$$
$$\Delta\mathbf{u} = \widetilde{V}_K\, \widetilde{\theta}_K \;, \qquad \Delta\mathbf{x}_0 = \mathbb{B}_0^{1/2}\, \widetilde{V}_K\, \widetilde{\theta}_K \,.$$

An explicit form of the solution (11) can be obtained using the Sherman-Morrison-Woodbury formula (Sherman and Morrison, 1950; Mandel, 2007)

$$\left(W + UV^T\right)^{-1} = W^{-1} - W^{-1} U \left(I + V^T W^{-1} U\right)^{-1} V^T W^{-1}$$

with $W = I$ and

$$U = V = \widetilde{V}_K^T\, \mathbb{B}_0^{T/2}\, \mathbf{M}^T H^T \mathbb{R}_F^{-1/2} \mathbb{R}_F^{-1/2} H \mathbf{M} \mathbb{B}_0^{1/2} \widetilde{V}_K \,.$$

Together with the notation

$$\begin{aligned}
\widetilde{\mathbb{B}}_0^{1/2} &= \mathbb{B}_0^{1/2} \widetilde{V}_K \;, \\
\widetilde{\mathbb{B}}_0 &= \mathbb{B}_0^{1/2} \widetilde{V}_K \widetilde{V}_K^T\, \mathbb{B}_0^{T/2} \;, \\
\widetilde{\mathbb{B}}_F^{1/2} &= \mathbf{M} \widetilde{\mathbb{B}}_0^{1/2} = \mathbf{M} \mathbb{B}_0^{1/2} \widetilde{V}_K \;, \\
\widetilde{\mathbb{B}}_F &= \mathbf{M} \mathbb{B}_0^{1/2} \widetilde{V}_K \widetilde{V}_K^T\, \mathbb{B}_0^{T/2}\, \mathbf{M}^T \;,
\end{aligned}$$

the Sherman-Morrison-Woodbury formula leads to the following solution of (11)

$$\begin{aligned}
\widetilde{\theta}_K &= \left(I - \widetilde{\mathbb{B}}_F^{T/2}\, H^T \left(\mathbb{R}_F + H\widetilde{\mathbb{B}}_F H^T\right)^{-1} H \widetilde{\mathbb{B}}_F^{1/2}\right) \\
&\quad \cdot \widetilde{\mathbb{B}}_F^{T/2} H^T \mathbb{R}_F^{-1} \left(\mathbf{y} - H\mathbf{x}_F^B\right) \\
\Delta\mathbf{x}_F &= \mathbf{M} \mathbb{B}_0^{1/2}\, \widetilde{V}_K\, \widetilde{\theta}_K = \widetilde{\mathbb{B}}_F^{1/2}\, \widetilde{\theta}_K \\
&= \widetilde{\mathbb{B}}_F^{1/2} \Big(I - \widetilde{\mathbb{B}}_F^{T/2}\, H^T \left(\mathbb{R}_F + H\widetilde{\mathbb{B}}_F H^T\right)^{-1} H \widetilde{\mathbb{B}}_F^{1/2}\Big) \\
&\quad \cdot \widetilde{\mathbb{B}}_F^{T/2} H^T \mathbb{R}_F^{-1} \left(\mathbf{y}_F - H\mathbf{x}_F^B\right) \\
&= \left(\widetilde{\mathbb{B}}_F - \widetilde{\mathbb{B}}_F\, H^T \left(\mathbb{R}_F + H\widetilde{\mathbb{B}}_F H^T\right)^{-1} H \widetilde{\mathbb{B}}_F\right) \\
&\quad \cdot H^T \mathbb{R}_F^{-1} \left(\mathbf{y}_F - H\mathbf{x}_F^B\right) \\
&= \widetilde{\mathbb{B}}_F\, H^T \left(\mathbb{R}_F + H\widetilde{\mathbb{B}}_F H^T\right)^{-1} \left(\mathbf{y}_F - H\mathbf{x}_F^B\right).
\end{aligned}$$

The above relation gives the 4D-Var update formula at t_F:

$$\mathbf{x}_F^A = \mathbf{x}_F^B + \widetilde{\mathbb{B}}_F\, H^T \left(\mathbb{R}_F + H\widetilde{\mathbb{B}}_F H^T\right)^{-1} \left(\mathbf{y}_F - H\mathbf{x}_F^B\right). \qquad (12)$$

A comparison between (12) and (6) reveals that the 4D-Var update (12) is equivalent to a suboptimal Kalman filter update with

$$\mathbf{K}_F = \widetilde{\mathbb{B}}_F\, H^T \left(\mathbb{R}_F + H\widetilde{\mathbb{B}}_F H^T\right)^{-1} .$$

Consequently the analysis covariance associated with the 4D-Var estimate is:

$$\begin{aligned}
\widetilde{\mathbb{A}}_F &= \widetilde{\mathbb{B}}_F - \widetilde{\mathbb{B}}_F\, H^T \left(\mathbb{R}_F + H\widetilde{\mathbb{B}}_F H^T\right)^{-1} H\, \widetilde{\mathbb{B}}_F \\
&= \widetilde{\mathbb{B}}_F - \widetilde{\mathbb{B}}_F\, H^T\, \mathbb{R}_F^{-1/2} \left(I + \mathbb{R}_F^{-1/2}\, H\widetilde{\mathbb{B}}_F H^T\, \mathbb{R}_F^{-1/2}\right)^{-1} \\
&\quad \mathbb{R}_F^{-1/2}\, H\, \widetilde{\mathbb{B}}_F \\
&= \left(\widetilde{\mathbb{B}}_F^{-1} + H^T\, \mathbb{R}_F^{-1}\, H\right)^{-1} ,
\end{aligned}$$

where the last relation follows from another application of the Sherman-Morrison-Woodbury formula.

2.3 *EnKF Solution with Small Ensemble*

Consider now a standard formulation of the EnKF with K ensemble members. Let $\langle\mathbf{x}\rangle$ denote the ensemble mean and $\delta\mathbf{x}(i) = \mathbf{x}(i) - \langle\mathbf{x}\rangle$, $i = 1, \cdots, K$, denote the deviations from the mean. The initial set of K ensemble perturbations are drawn from the normal distribution $\mathcal{N}\left(0, \mathbb{B}_0\right)$. Equivalently, they are obtained via a variable transformation from the standard normal vectors ξ_i as follows:

$$\delta\mathbf{x}_0(i) = \mathbb{B}_0^{1/2}\, \xi_i \;, \quad \xi_i \in \left(\mathcal{N}(0,1)\right)^n \;, \quad i = 1, \cdots, K \,. \qquad (13)$$

The ensemble covariance is

$$\widehat{\mathbb{B}}_0 = \frac{1}{K-1} \sum_{i=1}^{K} \delta\mathbf{x}_0(i)\, \delta\mathbf{x}_0(i)^T \approx \mathbb{B}_0 \,.$$

The perturbations are propagated to the final time via the tangent linear model (this holds true for perturbations of any magnitude for the linear model dynamics assumed here)

$$\delta\mathbf{x}_F(i) = \mathbf{M} \cdot \delta\mathbf{x}_0(i) \;, \quad i = 1, \cdots, K \,.$$

Denote the scaled initial perturbations by

$$\begin{aligned} \mathbf{X}_0 &= \frac{1}{\sqrt{K-1}} \left[\delta \mathbf{x}_0(1), \cdots, \delta \mathbf{x}_0(K)\right] \qquad (14) \\ &= \mathbb{B}_0^{1/2} \left[\frac{\xi_1}{\sqrt{K-1}}, \cdots, \frac{\xi_K}{\sqrt{K-1}}\right], \end{aligned}$$

we have that

$$\widehat{\mathbb{B}}_0 = \mathbf{X}_0 \cdot \mathbf{X}_0^T ,$$

and

$$\mathbf{X}_F = \mathbf{M} \cdot \mathbf{X}_0 , \quad \widehat{\mathbb{B}}_F = \mathbf{X}_F \cdot \mathbf{X}_F^T .$$

The EnKF analysis updates each member using the formula:

$$\begin{aligned} \mathbf{x}_F^A(i) &= \mathbf{x}_F^B(i) + \widehat{\mathbb{B}}_F H^T \left(H \widehat{\mathbb{B}}_F H^T + \mathbb{R}\right) \\ &\quad \cdot \left(\mathbf{y}_F(i) - H \mathbf{x}_F^B(i)\right) , \quad i = 1, \cdots, K . \end{aligned}$$

The ensemble mean values are updated using

$$\langle \mathbf{x}_F^A \rangle = \langle \mathbf{x}_F^B \rangle + \widehat{\mathbb{B}}_F H^T \left(H \widehat{\mathbb{B}}_F H^T + \mathbb{R}\right)^{-1} \left(\mathbf{y}_F - H \langle \mathbf{x}_F^B \rangle\right) . \qquad (15)$$

2.4 Comparison of 4D-Var and EnKF Solutions

2.4.1 4D-Var Solution as a Kalman Update A comparison of (12) and (15) reveals an interesting conclusion. The suboptimal 4D-Var (in the linear case, with one observation time) leads to a Kalman-like update of the state at the final time. The difference between the 4D-Var update (12) and the EnKF mean update (15) is in the approximation given to the background covariance matrices.

In the ensemble Kalman filter case:

$$\begin{aligned} \widehat{\mathbb{B}}_F^{1/2} &= \mathbf{M} \mathbb{B}_0^{1/2} \Big[\frac{\xi_1}{\sqrt{K-1}}, \cdots, \frac{\xi_K}{\sqrt{K-1}}\Big] \\ &= \mathbf{M} \mathbb{B}_0^{1/2} \Big[\widehat{v}_1, \cdots, \widehat{v}_K\Big] , \end{aligned}$$

while in the 4D-Var case

$$\widetilde{\mathbb{B}}_F^{1/2} = \mathbf{M} \mathbb{B}_0^{1/2} \Big[\widetilde{v}_1, \cdots, \widetilde{v}_K\Big] .$$

The standard EnKF initialization (13) is based on the random vectors ξ_i sampled from a normal distribution. Instead of the random vectors one can use the following initial perturbations:

$$\begin{aligned} \xi_i &\in \left(\mathcal{N}(0,1)\right)^n \underset{\text{replaced by}}{\longrightarrow} \xi_i = \sqrt{K-1} \cdot \widetilde{v}_i , \\ i &= 1, \cdots, K , \qquad (16) \end{aligned}$$

where $\widetilde{v}_i$ are the orthonormal directions computed by the Lanczos algorithm applied to the preconditioned system (11). With this initialization the EnKF analysis for the mean (15) is precisely the 4D-Var solution (12).

Note that while the initial perturbations in the regular EnKF have zero mean, the perturbations (16) along the Lanczos directions have nonzero mean. Thus formula (16) constructs a biased initial ensemble, which performs an adjustment of the initial state before the filtering occurs.

2.4.2 EnKF as an Optimization Algorithm EnKF looks for an increment in the space of ensemble deviations

$$\langle \mathbf{x}_F^A \rangle = \langle \mathbf{x}_F^B \rangle + \mathbf{X}_F \cdot \mu,$$

where the vector of coefficients μ is obtained as the minimizer of the function (Ott et al., 2004):

$$J^{\rm ens}(\mu) = \frac{1}{2} \mu^T \mu + \frac{1}{2} \left(d_F^B - H \mathbf{X}_F \mu\right)^T \mathbb{R}_F^{-1} \left(d_F^B - H \mathbf{X}_F \mu\right) \qquad (17)$$

with

$$d_F^B = \mathbf{y}_F - H \langle \mathbf{x}_F^B \rangle .$$

The solution is obtained by solving the linear system

$$\nabla_\mu J^{\rm ens}(\mu) = 0$$

which is equivalent to

$$\left(I + \mathbf{X}_F^T H^T \mathbb{R}_F^{-1} H \mathbf{X}_F\right) \cdot \mu = \mathbf{X}_F^T H^T \mathbb{R}_F^{-1} d_F^B . \qquad (18)$$

Using the Serman-Woodbury-Morrison formula to "invert" the system matrix in (18) leads to the following closed form solution:

$$\begin{aligned} \mu &= \mathbf{X}_F^T H^T \left(\mathbb{R}_F + H \mathbf{X}_F \mathbf{X}_F^T H^T\right)^{-1} \cdot d_F^B \qquad (19) \\ \langle \mathbf{x}_F^A \rangle &= \langle \mathbf{x}_F^B \rangle + \mathbf{X}_F \cdot \mu \\ &= \langle \mathbf{x}_F^B \rangle + \mathbf{X}_F \mathbf{X}_F^T H^T \left(\mathbb{R}_F + H \mathbf{X}_F \mathbf{X}_F^T H^T\right)^{-1} \cdot d_F^B \\ &= \langle \mathbf{x}_F^B \rangle + \widehat{\mathbb{B}}_F H^T \left(\mathbb{R}_F + H \widehat{\mathbb{B}}_F H^T\right)^{-1} \\ &\quad \cdot \left(\mathbf{y}_F - H \langle \mathbf{x}_F^B \rangle\right) . \end{aligned}$$

This confirms that the EnKF analysis formula provides the minimizer for (17).

Let $\mathbf{X}_F = \mathbf{M} \mathbb{B}_0^{1/2} \widehat{V}$, where $\widehat{V} = (K-1)^{-1/2}\, \xi$ and $\xi \in \mathbb{R}^{n \times K}$ is a matrix of $\mathcal{N}(0,1)$ independent random numbers. Then the system (18) becomes:

$$\begin{aligned} &\left(I + \widehat{V}^T \mathbb{B}_0^{T/2} \mathbf{M}^T H^T \mathbb{R}_F^{-1} H \mathbf{M} \mathbb{B}_0^{1/2} \widehat{V}\right) \cdot \mu \\ &\qquad = \widehat{V}^T \mathbb{B}_0^{T/2} \mathbf{M}^T H^T \mathbb{R}_F^{-1} d_F^B \qquad (20) \\ &\Delta \langle \mathbf{x}_F \rangle = \langle \mathbf{x}_F^A \rangle - \langle \mathbf{x}_F^B \rangle = \mathbf{X}_F \cdot \mu = \mathbf{M} \mathbb{B}_0^{1/2} \widehat{V} \cdot \mu \\ &\Delta \langle \mathbf{x}_0 \rangle = \langle \mathbf{x}_0^A \rangle - \langle \mathbf{x}_0^B \rangle = \mathbf{X}_0 \cdot \mu = \mathbb{B}_0^{1/2} \widehat{V} \cdot \mu \end{aligned}$$

A comparison of the EnKF system (20) with the 4D-Var system solved by K Lanczos iterations (11) reveals that the two formulas are nearly identical. In the suboptimal 4D-Var approach the full preconditioned 4D-Var system (10) is reduced by the orthonormal matrix $\widetilde{V}_K \in \mathbb{R}^{n \times K}$ (whose columns are the Lanczos vectors). When the preconditioned 4D-Var system (10) is *partially reduced* by the matrix of independent normal random numbers $\xi \in \mathbb{R}^{n \times K}$, the result is the system (20) that provides the EnKF analysis. By *partially reduced* we mean that the reduction is applied only to the second matrix in the parenthesis, and that the identity is taken with the appropriate dimensions. Note that the full reduction leads to the first term equal to $\widehat{V}^T \widehat{V} \approx (n-1)/(K-1)\, I$; this term approximately equals a scaled identity matrix, with a scale factor much larger than one in practice.

Loosely speaking, an important difference between 4D-Var and EnKF is the choice of subspace where the full system is reduced. In 4D-Var the subspace is carefully chosen by the iterative procedure, while in EnKF this subspace is chosen randomly.

3 A Hybrid Approach to Data Assimilation

The above analysis reveals a subtle similarity between the 4D-Var and EnKF analyses for the linear, Gaussian case with one observation window. If the initial ensemble is constructed using perturbations along the directions chosen by the 4D-Var solver, the EnKF yields the same mean analysis as the 4D-Var yield. This result motivates a hybrid assimilation algorithm, where 4D-Var is run for a short window; the 4D-Var search directions are used to construct an initial ensemble, and then EnKF is run for a longer time window. The procedure can be repeated periodically, i.e., additional short window 4D-Var runs can be used from time to time to regenerate the ensemble. We now describe in detail the hybrid data assimilation algorithm; even if the motivation comes from a linear analysis, the algorithm below can also be applied to nonlinear systems.

(i) Starting from $\mathbf{x}_0^{(0)} = \mathbf{x}_0^B$, run 4D-Var for a short time window. The iterative numerical optimization algorithm generates a sequence of intermediate solutions $\mathbf{x}_0^{(j)}$ for each iteration $j = 1, \ldots, \ell$.

(ii) Construct $\mathcal{S}_{t_0}$, a matrix whose columns are the normalized 4D-Var increments:

$$\mathcal{S}_{t_0} = \left[\frac{\mathbf{x}_0^{(j)} - \mathbf{x}_0^{(j-1)}}{\left\| \mathbf{x}_0^{(j)} - \mathbf{x}_0^{(j-1)} \right\|} \right]_{j=1,\ldots,\ell} \in \mathbb{R}^{n\times\ell} \,. \qquad (21)$$

The normalized increments play the role of the Lanczos vectors in the general case. Note, however, that they are not orthogonal.

(iii) Perform a singular value decomposition of $\mathcal{S}_{t_0}$:

$$\mathcal{S}_{t_0} = U \Sigma V^T \,, \qquad (22)$$

and retain only the first K right singular vectors $u_1, \ldots u_K$ that correspond to the largest K singular values $\sigma_1, \ldots, \sigma_K$.

The directions $\widetilde{v}_i = u_i$, $i = 1, \cdots, K$, are used in (16) to generate the initial EnKF ensemble.

(iv) EnKF initialized as above is run for a longer time period, after which the ensemble is reinitialized using another short window 4D-Var run.

Note that the initial perturbations in the regular EnKF have zero mean. On the other hand the orthonormal directions $\widetilde{v}_i$ obtained in step (iii) above are independent, and therefore their mean is nonzero. We can remove the bias by subtracting the mean from each direction, or by adding one additional ensemble member initialized using $\widetilde{v}_{K+1} = -\sum_{i=1}^{K} \widetilde{v}_i$.

The proposed hybrid method is computationally less expensive than the full fledged 4D-Var, as only short assimilation windows are considered, and only a relatively small number of iterations is performed. The method is expected to perform better than the regular EnKF due to the special selection of how the state space is initially sampled. Note that the application of the hybrid method requires the 4D-Var machinery to be in place (and in particular, requires an adjoint model). The infrastructure is thus more complex than that required by regular EnKF; the complexity is similar to the case where the total energy singular vectors (or the Hessian singular vectors) are computed and used to initialize the ensemble.

Note that a popular approach to initializing the EnKF is to place the initial perturbations along the "bred vectors" (BVs) (Toth and Kalnay, 1997). The bred vectors share similar properties with the Lyapunov vectors (LVs); they have finite amplitude, finite time, and have local properties in space. The BVs are used to capture the maximum error growth directions in the model. We next implement the proposed hybrid approach and compare with the regular EnKF as well as the EnKF with the breeding technique, and show the numerical test results in section 4.

4 Numerical Experiments

4.1 Linear Test Case

To test the proposed hybrid approach, we first use a simple linear model with $n = 7$ states. Define the diagonal eigenvalue matrix

$$D = \mathrm{diag}\{10,\ 9.9,\ 0.2,\ 0.1,\ 0.01,\ 0.001,\ 0.0001\}\,,$$

and the tridiagonal eigenvector matrix V:

$$V = \begin{bmatrix} 2 & 1 & 0 & \cdots \\ 1 & 2 & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ \cdots & 0 & 1 & 2 \end{bmatrix}.$$

The linear model is defined by the matrix

$$\mathbf{M} = V \cdot D \cdot V^{-1}$$

such that a multiplication by $\mathbf{M}$ advances the state in time by one time unit. The linear model has two directions along which the error is amplified (corresponding to the eigenvalues greater than one). The two dimensional subspace of error growth can be spanned by only three ensemble members in EnKF, and by the first three directions generated by the 4D-Var iterative optimization routine.

The background covariance is constructed with a correlation distance $L = 1$ as follows:

$$\mathbb{B}_0(i,j) = \sigma_i \cdot \sigma_j \cdot \exp\left(-\frac{|i-j|^2}{L^2} \right), \quad i,\, j = 1,\, \ldots,\, n\,, \quad (23)$$

with the standard deviations $\sigma_i = 0.1$.

The linear model is run for six time units. The "true" solution $\mathbf{x}_i^t = 0$ is zero at all times. Synthetic observations at the end of each time unit are obtained by adding random noise with mean zero and covariance $\mathbb{R}_i = 0.01$.

Since the system is linear, the cost function is quadratic, and the 4D-Var solution is obtained by solving a linear system. We compute the perfect 4D-Var solution by solving this linear system exactly. We also compute a suboptimal 4D-Var solution by applying a preconditioned conjugate gradients (PCG) method with three iterations.

We use three ensemble members for the EnKF. Several versions of the EnKF are implemented as follows:

(i) EnKF-Regular. The ensemble is initialized using normal random samples and the perturbed observations version of the algorithm implemented in (Evensen, 2003).

(ii) EnKF-Eigenvector. The initial ensemble perturbations are placed along the three dominant eigenvectors of the linear system, i.e., the initial ensemble spans the directions of maximal error growth. This approach represents the initialization along the bred vectors (Toth and Kalnay, 1997).

(iii) EnKF-Hybrid. A "short window" 4D-Var solution is obtained by using only the observations after one time unit, and by applying three PCG iterations. The directions generated by the short window 4D-Var are used to initialize the hybrid EnKF method.

To assess the effectiveness of each assimilation method we compute the 2-norm of the analysis error $\|\mathbf{x}_i^a - \mathbf{x}_i^t\|$ (analysis minus truth) at the end of each time unit $i = 1, \cdots, 6$. The results of the EnKF-Regular method depend on the particular draw of normal random numbers used to initialize the ensemble. To remove the random effects from the comparison, we perform multiple EnKF-Regular experiments (each initialized with a different random draw) and report the average errors from 1,000 converging runs.

In the regular EnKF ensemble generation, the ensemble of initial perturbations has zero mean. In the EnKF-Hybrid approach we take an extra step to eliminate the bias by subtracting the mean from each perturbation direction before constructing the initial ensemble. The same procedure is applied in the EnKF-Eigenvector case. The evolution of the analysis errors for different assimilation methods is shown in Figure 1.

The smallest errors are associated with the perfect 4D-Var solution, followed by the suboptimal 4D-Var solution (three PCG iterations). The errors keep decreasing until the end of window 4. The suboptimal 4D-Var solution also shows small errors, approaching the perfect 4D-Var solution.

Among the three EnKF methods the largest errors are associated with the regular version, which uses a random initial ensemble. The medium errors are associated with the case where the initial perturbations are along the dominant eigenvectors. Finally, the EnKF-Hybrid solution shows the smallest errors. This indicates that the initial ensemble generated with 4D-Var directions is more effective than initial ensembles obtained through either random sampling or breeding.

To quantify the improvement provided by the hybrid approach we compute the ratio between the analysis errors with hybrid EnKF and the regular EnKF as follows:

$$\text{error ratio} = \frac{\left\|\mathbf{x}^{\text{EnKF}-\text{Hybrid}} - \mathbf{x}^t\right\|}{\left\|\mathbf{x}^{\text{EnKF}-\text{Regular}} - \mathbf{x}^t\right\|}.$$

A similar metric is used for the analysis errors of EnKF-Eigenvector. The error ratios are presented in Figure 2. The results indicate that both the eigenvector and the hybrid versions of EnKF provide smaller errors than the regular (randomly initialized) EnKF. The hybrid error is consistently smaller than the eigenvector error, showing the power of the proposed hybrid approach.

We have performed additional experiments where the "short window 4D-Var" used to initialize the hybrid ensemble spans two time units. The results are similar to those obtained from only one window, and are not reported here.

4.2 Nonlinear Test Case

The nonlinear test is carried out with the Lorenz-96 model. This chaotic model has $n = 40$ states and is described by the following equations:

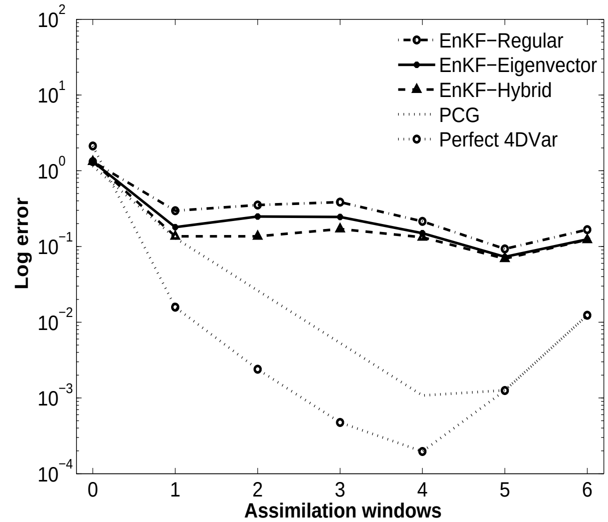

Figure 1. Comparison of analysis errors for several data assimilation methods applied to the linear test problem. Among the EnKF methods the hybrid version is the most accurate.

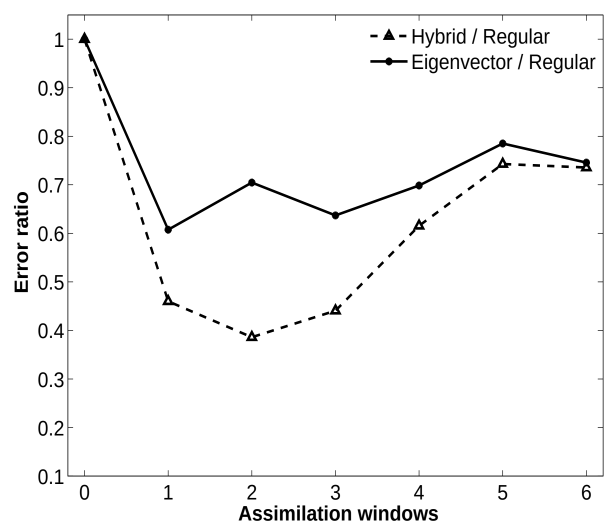

Figure 2. Ratios of analysis errors obtained with different assimilation methods for the linear test. EnKF-eigenvector over regular EnKF (solid), and unbiased EnKF-Hybrid over regular EnKF (dashed).

$$\frac{dx_j}{dt} = -x_{j-1}(x_{j-2} - x_{j+1} - x_j) + F\,, \quad j = 1, \ldots, n \quad (24)$$

with periodic boundary conditions. The forcing term is $F = 8.0$. The conventional EnKF method implementation follows the algorithm described in (Evensen, 2003). We compare the following methods:

(i) EnKF-Regular: sample normal random numbers to form the perturbation ensemble, then add the perturbations ensemble to the initial best guess.

(ii) EnKF-Breeding. The "breeding" technique described in (Toth and Kalnay, 1997) is used to capture the maximum error growth directions of the system. The initial ensemble perturbations are set along the bred vectors.

(iii) EnKF-Hybrid. A 4D-Var assimilation is run in a short window of 0.2 time units. The directions generated by the L-BFGS numerical optimization routine are used to initialize the hybrid ensemble as explained in Section 3.

Each method uses an ensemble of 10 members. The total simulation time is three time units of the Lorenz model. There are 15 equidistant observation times; synthetic observations for all states are obtained from the reference solution. The 4D-Var short window run used to initialize the hybrid ensemble spans 0.2 time units (one observation time). The background covariance is generated using (23) with $L = 1.0$ and standard deviations equal to 1% of the initial reference values. The breeding EnKF implementation follows the description in (Toth and Kalnay, 1997), where the perturbations are propagated with the system for one time unit and rescaled. The propagation and rescaling are carried out ten times. We use three resulting bred vectors as maximum error reduction directions to construct the perturbation ensembles.

In order to alleviate the impact of different random choices of the initial conditions in EnKF-Regular, and of different random noise levels used to perturb the observations, we run 100 independent tests with each method to obtain the average solutions. Without loss of generality, we plot the first component of the Lorenz chaotic state. Figure 3 shows the first component of the solutions obtained with different methods. The reference solution is represented with a solid line, with circles on it indicating the observations. The background solution is represented with a dashed line. The EnKF-regular solution is represented with dash dotted line, and the EnKF-Hybrid solution is represented by a solid line with triangles. Both EnKF analyses are in good agreement with the reference solution.

To better assess the accuracy of each method, we compute the root mean square error (RMSE) of the average solution obtained from 100 runs for each method, and plot the error in Figure 4. The dotted line shows the background RMSE error. The EnKF-Regular RMSE is shown with a dash dotted line. The EnKF-Hybrid RMSE is the solid line with triangles. The dashed line with circle on it represents the EnKF-Breeding RMSE. We observe that both the EnKF-Hybrid RMSE and the EnKF-Breeding RMSE are smaller than the EnKF-Regular RMSE, showing improvements of both methods over the regular sampling method for EnKF ensemble generation.

Figure 5 reports the ratio of the EnKF-Hybrid RMSE over the EnKF-Regular RMSE, and the ratio of the EnKF-Breeding RMSE over the EnKF-Regular RMSE. Both ratios are well below one throughout the simulation interval, indicating that both methods perform better than EnKF-Regular. The hybrid analysis error is smaller for the first 1.5 time units, and for the time interval [2.5,3] units. The breeding analysis error is smaller between [1.5,2.5] time units. The hybrid RMSE is about 70% of that of the regular EnKF. We conclude that, for some time interval after initialization (here, 1.5 units) the hybrid ensemble method works better than the breeding method. After this interval a new short window 4D-Var may be necessary to reinitialize the ensemble. More work is needed to formulate and test this resampling strategy.

The numerical tests in both linear and nonlinear cases show the hybrid method improves the analysis solution when compared to the regular EnKF solution. The implementation requires running a 4D-Var for a short time window in order to collect the directions used to initialize the ensemble. Tests also show that the proposed hybrid approach performs better than the breeding method for some time interval after the initialization.

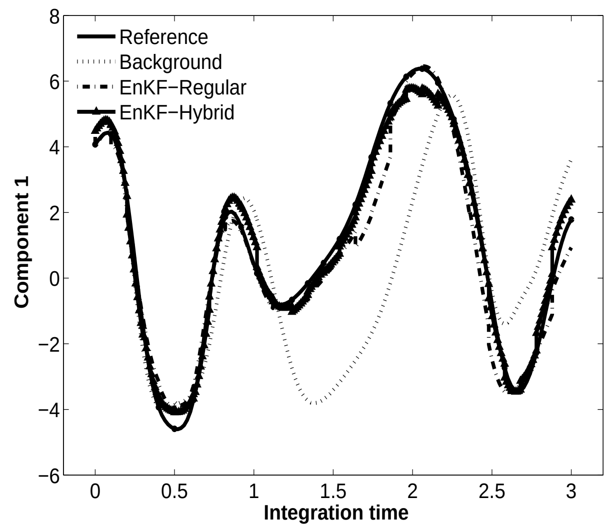

Figure 3. Time evolution of the first Lorenz-96 component for different solutions. Reference (solid line), background (dashed line), analysis with regular EnKF, 10 members (dash dotted line), and analysis with hybrid EnKF, 10 members (solid line with triangles).

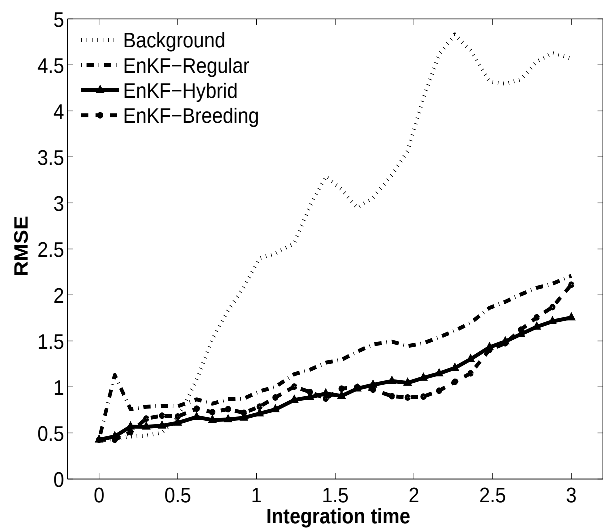

Figure 4. Root mean square error evolution for background, regular EnKF, hybrid EnKF, and breeding EnKF solutions.

5 Summary

This work is the first to establish the equivalence between the EnKF with a small ensemble and the suboptimal 4D-Var method in the linear Gaussian case, and for a single observation time within one assimilation window. The rela-

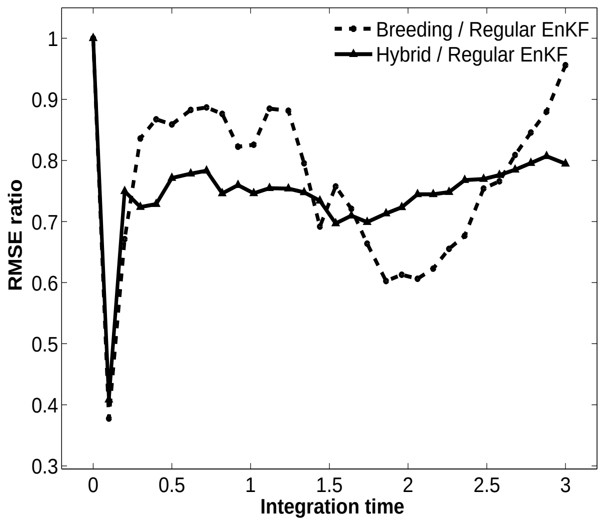

Figure 5. Ratios of analysis errors obtained with different assimilation methods for the nonlinear test. Breeding over regular EnKF (dashed) and hybrid over regular EnKF (solid).

tionship between these two methods motivates a new hybrid data assimilation approach: the directions identified by an iterative solver for a short window 4D-Var problem are used to construct the EnKF initial ensemble. Numerical tests on both linear and nonlinear cases show that the proposed hybrid approach improves the analysis accuracy of the regular EnKF. The overall increase in computational cost over regular EnKF is moderate, as short window 4D-Var problems are solved infrequently, and only a small number of iterations is performed each time. The hybrid method requires that a model adjoint is available. The proposed approach brings together two different families of methods, variational and ensemble filtering. More detailed tests on complex systems are planned as future work.

Acknowledgements

This work has been supported in part by NSF through the awards NSF CCF 0635194 and DMS 0915047. The authors would like to thank Mohamed Jardak for many fruitful discussions on the topic, for pointing us to very relevant literature, and for his feedback on the manuscript.

REFERENCES

Anderson, J. L. 2003. A local least squares framework for ensemble filtering. *Monthly Weather Review* **131**(4), 634–642.

Bewley, T., Cessna, J., and Colburn, C. 2009. EnVE: A new estimation algorithm for weather forecasting and flow control. 4th Flow Control Conference, 23 - 26 June 2008, Seattle, Washington.

Bennett, A. F. 2002. Inverse modeling of the ocean and atmosphere Bennett, A. F. Vol. 352 Cambridge University Press.

Burgers, G., van der Leeuwen, P. J. and Evensen, G. 1998. Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review* **126**, 1719–1724.

Daley, R. 1991. Atmospheric data analysis Daley, R. Cambridge University Press.

Evensen, G., and van Leeuwen, P.J. 2000. An ensemble Kalman smoother for nonlinear dynamics. *Monthly Weather Review*, **128**, 1852–1867.

Evensen, G. 2003. The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics* **53**, 343–367.

Evensen, G. 2007. Data assimilation: The ensemble Kalman filter. Evensen, G. Springer, Berlin.

Fisher, M. 2002. Assimilation techniques (5): approximate Kalman filters and singular vectors. *European Center for Medium-Range Weather Forecast* .

Fisher, M. Leutbecher, M. and Kelly, G. A. 2005. On the equivalence between Kalman smoothing and weak-constrained four-dimensional variational data assimilation *Quarterly Journal of the Royal Meteorological Society* **131**(572), 3235–3246.

Gejadze, I. Y., Le Dimet, F. X. and Shutyaev, V. 2008. On analysis error covariances in variational data assimilation. *SIAM Journal on Scientific Computing* **30**(4), 1847–1874. `http://dx.doi.org/10.1137/07068744X`.

Hamill. T. M. and Snyder, C. 2000. A hybrid EnKF-3D-Var analysis scheme. *Monthly Weather Review*, **128**, 2905–2919.

Harlim, J., and Hunt, B.R. 2004. Four-dimensional local ensemble transform Kalman filter: numerical experiments with a global circulation mode. *Tellus*, **59A**, 731–748.

Hunt, B.R. et al. 2007. Four dimensional ensemble Kalman filtering. *Tellus*, **56A**, 273–277.

Kalman, R. E. 1960. A new approach to linear filtering and prediction problems. *Transaction of the ASME -Journal of Basic Engineering* **82**(Series D), 35–45.

Kalnay, E. 2003. Atmospheric modelling, data assimilation and predictability Kalnay, E. Cambridge University Press.

Kalnay, E., Li, H., Miyoshi, T., Yang, S.-C. and Ballabrera-Poy, J. 2007. 4D-Var or Ensemble Kalman Filter? *Tellus A* **59**, 758–773

Lewis, J., Lakshmivarahan, S. and Dhall, S. 2005. Dynamic data assimilation: A least squares problem. *Cambridge University Press* .

Li, Z. and Navon, I. M. 2001. Optimality of 4D-Var and its relationship with the Kalman filter and Kalman smoother. *Quarterly Journal of the Royal Meteorological Society* **127**(572), 661–684.

Liu, C., Xiao, Q., and Wang, B. 2008. An Ensemble-Based 4D-Var Data Assimilation scheme. Part I: Technical formulation and preliminary test. *American Weather Review*, **136**, 3363–3373, 2008.

Lorenc, A. C. 2003. The potential of the ensemble Kalman filter for NWP – a comparison with 4D-Var. Quarterly Journal of the Royal Meteorological Society, **129**(595), 3183–3203.

Mandel, J. 2007. Predictor-corrector and morphing ensemble filters for the assimilation of sparse data into high-dimensional nonlinear systems. *Proceeding of the 11th Symposium on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface (IOAS-AOLS)* .

Ménard, R. and Daley, R. 1996. The application of Kalman smoother theory to the estimation of 4D-Var error statistics. *Tellus A* **48**, 221–237.

Ott, E., Hunt, B.R., Szunyogh, I., Zimin, A.V., Kostelich, E.J., Corazza, M., Kalnay, E., Patil, D.J., and Yorke, J.A. 2004. A local ensemble Kalman filter for atmospheric data assimilation. *Tellus* **56A**, 415–428.

Rodgers, C. D. 2000. Inverse methods for atmospheric sounding: Theory and practice. Series on Atmospheric Oceanic and Planetary Physics, World Scientific Publishing Company.

Saad, Y. 2003. Iterative methods for sparse linear systems. Second edition, SIAM, Philadelphia.

Sherman, J. and Morrison, W. J. 1950. Adjustment of an inverse matrix corresponding to a change in one element of a given

matrix. *The Annals of Mathematical Statistics* **21**, 124–127.
Talagrand, O. and Courtier, P. 1987. Variational assimilation of meteorological observations with the adjoint equations Part I: Theory. *Quarterly Journal of Royal Meteorology Society* **113**, 1311–1328.
Tian, X., Xie, J., and Dai, A. 2008. An ensemble-based explicit 4D-Var assimilation method. *Journal of Geophysical Research*, **113**(D21124).
Torn, R.D., and Hakim, G.J. 2008. Ensemble-based sensitivity analysis. *Monthly Weather Review*, **136**(2), 663–677.
Toth, Z. and Kalnay, E. 1997. Ensemble forecasting at NCEP and the breeding method. *Monthly Weather Review* **127**, 3297–3318.
Wang, X., Snyder, C., and Hamill. T. M. 2007. On the Theoretical Equivalence of Differently Proposed Ensemble–3D-Var Hybrid Analysis Schemes. *Monthly Weather Review*, **135**(1), 222–227.
Zhang, M., Zhang, F., and Hansen, J. 2007. Coupling ensemble Kalman filter with 4D-Var data assimilation. *22nd Conference on Weather Analysis and Forecasting/ 18th Conference on Numerical Weather Prediction, Park City*.
Zupanski, M. 2005. Maximum Likelihood Ensemble Filter: Theoretical aspects. *American Weather Review*, **133**, 1710–1726.