Revision of TR-09-25: A Hybrid Variational/Ensemble Filter Approach to Data Assimilation

Adrian Sandu[1] and Haiyan Cheng[1]

Computational Science Laboratory
Department of Computer Science
Virginia Polytechnic Institute and State University
2201 Knowledgeworks II, 2202 Kraft Drive, Blacksburg, VA 24060, USA
Phone: 540-231-2193, Fax: 540-231-9218
E-mail: sandu@cs.vt.edu

Department of Computer Science
Willamette University
900 State Street, Salem, OR 97306, USA
Phone: 503-375-5339, Fax: 503-370-6977
E-mail: hcheng@willamette.edu

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. Each method has its advantages and disadvantages. It is of interest to develop hybrid data assimilation algorithms that can combine the relative strengths of the two approaches. This paper proposes a subspace approach to investigate 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. The proposed hybrid method is computationally less expensive than a full 4D-Var, as only short assimilation windows are considered. The hybrid method has the potential to perform better than the regular EnKF due to its look-ahead property. Numerical results show that the proposed hybrid

Preprint submitted to Elsevier *March 16, 2010*

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 [4, 13, 15, 26, 31, 44]. 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 [48]. 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 work [25] 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 [2], 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 [32]. 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 [19]. In the presence of model errors the weak-constrained 4D-Var and the fixed-interval Kalman smoother are equivalent [38].

With both methods coming to maturity, new interest in the community has been devoted to assess the relative merits of 4D-Var and EnKF[27, 35]. 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 [3, 5, 8, 12, 16, 17, 22, 24, 23, 28, 29, 30, 34, 39, 42, 43, 49, 51, 53, 54, 55, 56, 57, 58].

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 $\mathbf{x}(t) \in \mathbb{R}^n$,

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

Here t_0 and t_F are the beginning and the end points of the simulation time interval.

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:

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

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_F 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 true state $\mathbf{y}_k = \mathcal{H}_k(\mathbf{x}_k^t) + \varepsilon_k$ are available at each time instant t_k, $k = 0, \ldots, N_\mathrm{obs} - 1$, where the observation operator $\mathcal{H}_k$ maps the state space to the observation space. 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 $\mathbf{y}_0, \cdots, \mathbf{y}_{N_{\rm obs}-1}$, 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 [48]. It provides the analysis $\mathbf{x}_0^A$ as the argument which minimizes the cost function:

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

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 [4, 31, 41], one linearizes the estimation problem around the background trajectory (the trajectory started from the background initial condition $\mathbf{x}_0^B$ which has a state value $\mathbf{x}_k^B$ at t^k). By expressing the state as the correction over the background state $\mathbf{x}_k = \mathbf{x}_k^B + \Delta\mathbf{x}_k$, $k = 0, \cdots, N_{\rm obs} - 1$, we have

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

where $\Delta\mathbf{x}_k = \mathbf{M}_{t_0 \rightarrow t_k} \, \Delta\mathbf{x}_0$, 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 [25] 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 [18] provides a suboptimal state estimation in the nonlinear case by linearizing the model dynamics and the observation operator.

A typical assumption is that while the state evolves according to nonlinear dynamics 1, small errors evolve according to the linearaized model 2. If the errors in the model state at t_{k-1} have a normal distribution $\mathcal{N}(0, \mathbb{A}_{k-1})$ and propagate according to the linearized model dynamics 2, then the forecast errors at t_k are also normally distributed $\mathcal{N}(0, \mathbb{A}_k)$. The forecast is obtained using

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

where $\mathbf{M}^T$ is the adjoint of the tangent linear model, and $\mathbb{Q}_k$ is the covariance matrix of model errors. In this paper we will consider perfect models, i.e., we will assume $\mathbb{Q}_k = 0$ from now on. The analysis provides the state estimate $\mathbf{x}_k^A$ and the corresponding error covariance matrix $\mathbb{A}_k$

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

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 [18] uses a Monte-Carlo approach to propagate covariances. An ensemble of N_{ens} states (labeled $e = 1, \cdots, N_{\text{ens}}$) 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 $\mathbf{x}_k^f(e)$. If model errors are considered, Gaussian noise 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 ($\{\mathbf{x}_k^f(e)\}_{e=1,\cdots,N_{\text{ens}}}$ and $\{\mathbf{x}_k^A(e)\}_{e=1,\cdots,N_{\text{ens}}}$ respectively). The EnKF approach to data assimilation has attracted considerable attention in meteorology [2, 6] 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 [32], *provided that the 4D-Var solution is computed exactly and an infinitely large number of ensemble members is used in EnKF, and they both use the same covariance matrix*. 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. (In practice, if possible, the number of iterations can be sufficient to ensure that the difference between an exact solution of the minimization problem and the truncated solution is smaller than the statistical uncertainty in the analysis). 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.

The main contribution of this work is conceptual, and proposes a subspace approach to analyze the relationship between the *suboptimal* 4D-Var solution and the *suboptimal* EnKF solution. The analysis motivates a new hybrid filter algorithm for data assimilation which uses intermittent short 4D-Var runs to periodically reinitialize an ensemble filter. Beside the conceptual contribution, the new approach is potentially useful due to the following characteristics:

- The hybrid method is computationally less expensive than the full fledged 4D-Var: instead of solving the 4D-Var problem to convergence over a long assimilation window, one solves it sub-optimally over a short time sub-window; a less expensive hybrid filter then carries out the data assimilation throughout the entire window.

- The hybrid method has the potential to perform better than the regular EnKF. In the first cycle this is due to the special sampling of the initial error space. In subsequent cycles the potential for better performance comes from the look-ahead nature of the hybrid approach: while the regular EnKF continues indefinitely with an error subspace constructed based on *past dynamics and past data*, the hybrid EnKF periodically chooses a new subspace based on *future dynamics and future data*.

Preliminary versions of this work have been reported in Cheng's Ph.D. dissertation [11] and in the technical report [46].

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 model that advances the state $\mathbf{x} \in \mathbb{R}^n$ from t_0 to t_F,

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

We assume that the model $\mathbf{M}$ is perfect (the model error is zero).

We also 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 the 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 measurements is taken at t_F; the measurements are corrupted by unbiased Gaussian errors

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

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. *Consider a linear, perfect model, with a Gaussian distribution of the background initial condition errors. Consider the case of a single assimilation window, with observations taken at only one time (at the end of the window), and with Gaussian observation errors.*

One posterior state estimate is computed by the suboptimal 4D-Var method (truncated after several iterations); another estimate of the posterior mean state is obtained by the suboptimal EnKF method (using only a small number of ensemble members). Both methods use the same background covariance matrix.

There exists a particular initialization of the ensemble for which the suboptimal EnKF mean state estimate is equivalent to the state estimate computed by the suboptimal 4D-Var method.

Comment. The setting of the theorem does not capture the ability of 4D-Var to simultaneously incorporate time distributed observations, the effects of nonlinear dynamics and nonlinear observation operators, and the benefits of EnKF stabilization techniques like covariance inflation and localization.

Nevertheless, the simplified setting allows to draw interesting and useful parallels between 4D-Var and EnKF, and to gain considerable insight.

2.1. Full 4D-Var Solution

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

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

The first order necessary condition $\nabla_{\mathbf{x}_0}\mathcal{J} = 0$ reveals that the optimum increment is obtained by solving the following linear system:

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

where the solution is the deviation of the analysis from the background state, $\mathbf{x}_0^A = \mathbf{x}_0^B + \Delta\mathbf{x}_0$. The system matrix A in (7) is the inverse of the posterior covariance at t_0 [20], $A = \mathbb{A}_0^{-1}$. The right hand side vector b in (7) is the innovation vector 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

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

For nonlinear systems the above procedure (based on linearized dynamics and observation operator) corresponds to the the incremental 4D-Var formulation; this, in general, is only an approximation of full fledged 4D-Var, and it coincides with the Gauss-Newton method for solving the optimality system [21].

2.2. Iterative 4D-Var Solution by the Lanczos Method

In practice (7) is not solved exactly. It is solved within some approximation margin by using an iterative method and performing a number of iterations that is much smaller than the size of the state space. 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 [45] is employed to solve the symmetric linear system (7). The convergence of the Lanczos iterations (and, in general, that of any iterative method) can be improved via preconditioning. The background covariance is known and offers a popular preconditioner. Assume that a Cholesky or a symmetric square root decomposition $\mathbb{B}_0^{1/2}$ of $\mathbb{B}_0$ is available:

$$\mathbb{B}_0 = \mathbb{B}_0^{1/2} \cdot \mathbb{B}_0^{T/2}\,, \quad \mathbb{B}_0^{T/2} = \left(\mathbb{B}_0^{1/2}\right)^T\,.$$

Applying the background covariance square root as a symmetric preconditioner to the original 4D-Var system (7) leads to the following preconditioned 4D-Var system:

$$\begin{aligned} \widetilde{A} \cdot \Delta\mathbf{u}_0 &= \widetilde{b} \\ \widetilde{A} &= \mathbb{B}_0^{T/2}\, A\, \mathbb{B}_0^{1/2} = I_{n\times n} + \mathbb{B}_0^{T/2}\, \mathbf{M}^T H^T \mathbb{R}_F^{-1} H\mathbf{M}\mathbb{B}_0^{1/2} \\ \widetilde{b} &= \mathbb{B}_0^{T/2}\, b = \mathbb{B}_0^{T/2}\, \mathbf{M}^T H^T \mathbb{R}_F^{-1}\, \left(\mathbf{y}_F - H\mathbf{M}\mathbf{x}_0^B\right) \\ \Delta\mathbf{x}_0 &= \mathbb{B}_0^{1/2}\, \Delta\mathbf{u}_0 \end{aligned} \qquad (8)$$

Assume that K Lanczos iterations are performed from the starting point $\Delta\mathbf{u}_0^{[0]} = 0$, i.e., $\Delta\mathbf{x}_0^{[0]} = 0$ ($\mathbf{x}_0^{[0]} = \mathbf{x}_0^B$). Consequently, the first residual is

$r^{[0]} = \widetilde{b}$. The Lanczos method computes a symmetric tridiagonal matrix $\widetilde{T}_K \in \mathbb{R}^{K\times K}$ and a second matrix

$$\widetilde{V}_K = \left[\widetilde{v}_1, \cdots, \widetilde{v}_K\right] \in \mathbb{R}^{n\times K}$$

whose columns form an orthonormal basis of the Krylov space

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

The matrices have the following properties [45]

$$\widetilde{V}_K^T\, \widetilde{V}_K = I_{K\times K}\ , \quad \widetilde{V}_K^T\, \widetilde{A}\, \widetilde{V}_K = \widetilde{T}_K\ .$$

The approximate solution of the preconditioned 4D-Var system (8) obtained after K Lanczos iterations is the exact solution of the system reduced over the Krylov subspace $\mathcal{K}_K$,

$$\begin{aligned}
\widetilde{V}_K^T\, \widetilde{A}\, \widetilde{V}_K \cdot \widetilde{\theta}_K &= \widetilde{V}_K^T\, \widetilde{b}\ ; \quad \Delta\mathbf{u}^{[K]} = \widetilde{V}_K\, \widetilde{\theta}_K \qquad (9)\\
\widetilde{V}_K^T\, \widetilde{A}\, \widetilde{V}_K &= I_{K\times K} + \widetilde{V}_K^T\, \mathbb{B}_0^{T/2}\, \mathbf{M}^T H^T \mathbb{R}_F^{-1} H \mathbf{M} \mathbb{B}_0^{1/2}\, \widetilde{V}_K\\
\widetilde{V}_K^T\, \widetilde{b} &= \widetilde{V}_K^T\, \mathbb{B}_0^{T/2}\, \mathbf{M}^T H^T \mathbb{R}_F^{-1}\, \left(\mathbf{y}_F - H\mathbf{M}\mathbf{x}_0^B\right)\\
\Delta\mathbf{x}_0 &= \mathbb{B}_0^{1/2}\, \Delta\mathbf{u}^{[K]} = \mathbb{B}_0^{1/2}\, \widetilde{V}_K\, \widetilde{\theta}_K\\
\Delta\mathbf{x}_F &= \mathbf{M}\, \Delta\mathbf{x}_0 = \mathbf{M}\, \mathbb{B}_0^{1/2}\, \widetilde{V}_K\, \widetilde{\theta}_K\,.
\end{aligned}$$

An explicit form of the solution (9) can be obtained using the Sherman-Morrison-Woodbury formula [37, 47]

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

with

$$W = I_{K\times K} \quad \text{and} \quad U = V = \widetilde{V}_K^T\, \mathbb{B}_0^{T/2}\, \mathbf{M}^T H^T \mathbb{R}_F^{-1/2}\,.$$

Together with the notation

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

the Sherman-Morrison-Woodbury formula leads to the following solution of (9)

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

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

$$
\mathbf{x}_F^A = \mathbf{x}_F^B + \widetilde{\mathbb{B}}_F\, H^T \left(\mathbb{R}_F + H \widetilde{\mathbb{B}}_F H^T \right)^{-1} \left(\mathbf{y}_F - H \mathbf{x}_F^B \right). \tag{10}
$$

A comparison between (10) and (6) reveals that the 4D-Var update (10) is equivalent to a suboptimal Kalman filter update (10) at time $t_K = t_F$ 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.

Finally, consider the "initial perturbations"

$$\widetilde{\mathbf{X}}_0 = \mathbb{B}_0^{1/2}\, \widetilde{V}_K \ . \tag{11}$$

The system (9) can be rewritten in the following equivalent form

$$\begin{aligned} &\widetilde{\mathbf{X}}_0^T \left(\mathbb{B}_0^{-1} + \mathbf{M}^T H^T \mathbb{R}_F^{-1} H \mathbf{M}\right) \widetilde{\mathbf{X}}_0 \cdot \widetilde{\theta}_K \\ &\qquad = \widetilde{\mathbf{X}}_0^T\, \mathbf{M}^T H^T \mathbb{R}_F^{-1} \left(\mathbf{y}_F - H\mathbf{M}\mathbf{x}_0^B\right) \\ &\Delta\mathbf{x}_0 = \widetilde{\mathbf{X}}_0\, \widetilde{\theta}_K \\ &\Delta\mathbf{x}_F = \mathbf{M}\,\Delta\mathbf{x}_0 = \mathbf{M}\,\widetilde{\mathbf{X}}_0\, \widetilde{\theta}_K \ . \end{aligned} \tag{12}$$

Thus the suboptimal 4D-Var solves the original system (7) by projecting it onto the subspace spanned by $\widetilde{X}_0$.

2.3. EnKF Solution for a Small Ensemble

Consider now a standard formulation of the EnKF with K ensemble members. Let $\langle \mathbf{x} \rangle$ denote the ensemble mean and $\mathbf{x}'(i) = \mathbf{x}(i) - \langle \mathbf{x} \rangle$, $i = 1, \cdots, K$, 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:

$$\mathbf{x}_0'(i) = \mathbb{B}_0^{1/2}\, \xi_i \ , \quad i = 1, \cdots, K \ ; \quad \xi = \left[\xi_1, \cdots, \xi_K\right] \in \left(\mathcal{N}(0,1)\right)^{n \times K} \ . \tag{13}$$

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)

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

Denote the scaled random vectors by

$$\widehat{v}_i = \frac{1}{\sqrt{K-1}}\, \xi_i \ , \quad i = 1, \cdots, K \, ; \quad \widehat{V} = \left[\widehat{v}_1, \cdots, \widehat{v}_K\right] = \frac{1}{\sqrt{K-1}}\, \xi \ ,$$

and the matrix of the scaled initial perturbations by

$$\begin{aligned} \widehat{\mathbf{X}}_0 &= \frac{1}{\sqrt{K-1}} \left[\mathbf{x}_0'(1), \cdots, \mathbf{x}_0'(K)\right] \\ &= \mathbb{B}_0^{1/2} \left[\frac{\xi_1}{\sqrt{K-1}}, \cdots, \frac{\xi_K}{\sqrt{K-1}}\right] \\ &= \mathbb{B}_0^{1/2} \left[\widehat{v}_1, \cdots, \widehat{v}_K\right] . \end{aligned} \tag{14}$$

The ensemble covariance is

$$\begin{aligned} \widehat{\mathbb{B}}_0 &= \frac{1}{K-1}\sum_{i=1}^{K} \mathbf{x}'_0(i)\,\mathbf{x}'_0(i)^T = \widehat{\mathbf{X}}_0 \cdot \widehat{\mathbf{X}}_0^T \\ \widehat{\mathbb{B}}_0^{1/2} &= \widehat{\mathbf{X}}_0 = \mathbb{B}_0^{1/2}\,\widehat{V}\ , \end{aligned} \tag{15}$$

and, for a perfect and linear model,

$$\widehat{\mathbf{X}}_F = \mathbf{M}\cdot\widehat{\mathbf{X}}_0\ ,\quad \widehat{\mathbb{B}}_F = \widehat{\mathbf{X}}_F \cdot \widehat{\mathbf{X}}_F^T\ ,\quad \widehat{\mathbb{B}}_F^{1/2} = \widehat{\mathbf{X}}_F = \mathbf{M}\,\mathbb{B}_0^{1/2}\,\widehat{V}\,. \tag{16}$$

The EnKF analysis updates each member using the formula:

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

Here $\mathbf{y}_F(i)$ is the observation vector $\mathbf{y}_F$ plus a random perturbation vector drawn from the same probability distribution as the observation noise. The ensemble mean values are updated using

$$\langle \mathbf{x}_F^A\rangle = \langle \mathbf{x}_F^B\rangle + \widehat{\mathbb{B}}_F H^T \left(H\widehat{\mathbb{B}}_F H^T + \mathbb{R}_F\right)^{-1}\left(\mathbf{y}_F - H\langle \mathbf{x}_F^B\rangle\right)\ . \tag{17}$$

Comment. Other popular approaches to initializing the ensemble are the breed vectors, the total energy singular vectors, and the Hessian singular vectors [33]. For a linear system the breed vectors are (linear combinations of) the eigenvectors associated with the dominant eigenvalues

$$\mathbf{M}\,v_i = \lambda_i\,v_i\ ,\quad i = 1,\cdots,K\ .$$

Let C_0 and C_F be two positive definite matrices. The total energy singular vectors are defined with respect to the "energy" norms defined by these matrices at t_0 and t_F, respectively:

$$\mathbf{M}^T\,C_F\,\mathbf{M}\,v_i = \lambda'_i\,C_0\,v_i\ ,\quad i = 1,\cdots,K\ .$$

The Hessian singular vectors are the generalized eigenvectors associated with the dominant generalized eigenvalues of the following problem:

$$\mathbf{M}^T\,C_F\,\mathbf{M}\,v_i = \lambda_i\,\nabla^2\mathcal{J}\,v_i = \lambda''_i\,\left(\mathbb{B}_0^{-1} + \mathbf{M}^T H^T \mathbb{R}_F^{-1} H\mathbf{M}\right)\,v_i\,,\quad i = 1,\cdots,K\,.$$

Note that the breed and total energy singular vectors are based only on the model dynamics; the Hessian singular vectors account also for the initial uncertainty (through $\mathbb{B}_0$) and for the observation operator H. None of them accounts for the data $\mathbf{y}_F$. The 4D-Var Lanczos vectors account for the model dynamics, the observation operators, and the data. The cost of computing them is comparable to the cost of computing the Hessian singular vectors over the same time window.

2.4. Comparison of 4D-Var and EnKF Solutions

2.4.1. 4D-Var Solution as a Kalman Update

A comparison of (10) and (17) 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 (10) and the EnKF mean update (17) is in the approximation given to the background covariance matrices. In the EnKF case

$$\widehat{\mathbb{B}}_F^{1/2} = \mathbf{M}\,\mathbb{B}_0^{1/2}\,\widehat{V}\,,$$

while in the 4D-Var case

$$\widetilde{\mathbb{B}}_F^{1/2} = \mathbf{M}\,\mathbb{B}_0^{1/2}\,\widetilde{V}\,,$$

where $\widetilde{V}$ are the orthonormal directions computed by the Lanczos algorithm applied to the preconditioned system (9).

The standard EnKF initialization (13) is based on the random vectors ξ sampled from a normal distribution. If the vectors ξ are chosen such that

$$\frac{1}{K-1}\xi\xi^T = \widehat{V}\,\widehat{V}^T = \widetilde{V}\,\widetilde{V}^T \tag{18}$$

then the covariances are the same

$$\widehat{\mathbb{B}}_F = \mathbf{M}\,\mathbb{B}_0^{1/2}\,\widehat{V}\,\widehat{V}^T\,\mathbb{B}_0^{1/2}\,\mathbf{M}^T = \mathbf{M}\,\mathbb{B}_0^{1/2}\,\widetilde{V}\,\widetilde{V}^T\,\mathbb{B}_0^{1/2}\,\mathbf{M}^T = \widetilde{\mathbb{B}}_F$$

and the EnKF analysis mean (17) coincides with the 4D-Var analysis (10). An ensemble satisfying (18) will be called an *equivalent initial ensemble.*

2.4.2. EnKF as an Optimization Algorithm

EnKF looks for an increment in the subspace of ensemble deviations from mean

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

where the vector of coefficients $\widehat{\theta}$ is obtained as the minimizer of the function [40]:

$$\mathcal{J}^{\rm ens}(\widehat{\theta}) = \frac{1}{2}\widehat{\theta}^T\widehat{\theta} + \frac{1}{2}\left(d_F^B - H\widehat{\mathbf{X}}_F\widehat{\theta}\right)^T \, \mathbb{R}_F^{-1} \, \left(d_F^B - H\widehat{\mathbf{X}}_F\widehat{\theta}\right) \qquad (19)$$

with

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

The optimality condition $\nabla_{\widehat{\theta}}\mathcal{J}^{\rm ens} = 0$ is equivalent to the linear system

$$\left(I_{K\times K} + \widehat{\mathbf{X}}_F^T H^T \mathbb{R}_F^{-1} H \widehat{\mathbf{X}}_F\right) \cdot \widehat{\theta} = \widehat{\mathbf{X}}_F^T H^T \mathbb{R}_F^{-1} \, d_F^B. \qquad (20)$$

Using the Serman-Woodbury-Morrison formula to "invert" the system matrix in (20) leads to the following closed form solution:

$$\begin{aligned} \widehat{\theta} &= \widehat{\mathbf{X}}_F^T H^T \left(\mathbb{R}_F + H\widehat{\mathbf{X}}_F\widehat{\mathbf{X}}_F^T H^T\right)^{-1} \cdot d_F^B \qquad (21) \\ \langle \mathbf{x}_F^A \rangle &= \langle \mathbf{x}_F^B \rangle + \widehat{\mathbf{X}}_F \cdot \widehat{\theta} \\ &= \langle \mathbf{x}_F^B \rangle + \widehat{\mathbb{B}}_F H^T \left(\mathbb{R}_F + H\widehat{\mathbb{B}}_F H^T\right)^{-1} \\ &\qquad \cdot \left(\mathbf{y}_F - H\langle \mathbf{x}_F^B \rangle\right) \, . \end{aligned}$$

This confirms that the EnKF analysis formula provides the minimizer for (19).

Using (15)–(16) the system (20) becomes:

$$\begin{aligned} &\left(I_{K\times K} + \widehat{\mathbf{X}}_0^T \mathbf{M}^T H^T \mathbb{R}_F^{-1} H \mathbf{M} \widehat{\mathbf{X}}_0\right) \cdot \widehat{\theta} \qquad (22) \\ &\qquad = \widehat{\mathbf{X}}_0^T \mathbf{M}^T H^T \mathbb{R}_F^{-1} d_F^B \\ &\Delta\langle \mathbf{x}_F \rangle = \langle \mathbf{x}_F^A \rangle - \langle \mathbf{x}_F^B \rangle = \widehat{\mathbf{X}}_F \cdot \widehat{\theta} = \mathbf{M}\mathbb{B}_0^{1/2}\widehat{V} \cdot \widehat{\theta} \\ &\Delta\langle \mathbf{x}_0 \rangle = \langle \mathbf{x}_0^A \rangle - \langle \mathbf{x}_0^B \rangle = \widehat{\mathbf{X}}_0 \cdot \widehat{\theta} = \mathbb{B}_0^{1/2}\widehat{V} \cdot \widehat{\theta} \end{aligned}$$

We can rewrite (22) as

$$\begin{aligned}
&\widehat{\mathbf{X}}_0^T \left(\widehat{\mathbb{B}}_0^{\#} + \mathbf{M}^T H^T \mathbb{R}_F^{-1} H \mathbf{M}\right) \widehat{\mathbf{X}}_0 \cdot \widehat{\theta} \\
&\qquad = \widehat{\mathbf{X}}_0^T \mathbf{M}^T H^T \mathbb{R}_F^{-1} d_F^B \\
&\Delta\langle \mathbf{x}_F \rangle = \mathbf{M}\, \widehat{\mathbf{X}}_0 \cdot \widehat{\theta} \\
&\Delta\langle \mathbf{x}_0 \rangle = \widehat{\mathbf{X}}_0 \cdot \widehat{\theta}
\end{aligned} \tag{23}$$

where $\widehat{\mathbb{B}}_0^{\#}$ is the pseudo-inverse of the initial ensemble background covariance

$$\widehat{\mathbb{B}}_0 = \widehat{\mathbf{X}}_0 \widehat{\mathbf{X}}_0^T \ , \quad \widehat{\mathbf{X}}_0 = U \Sigma V^T \ , \quad \widehat{\mathbb{B}}_0^{\#} = U \Sigma^{-2} U^T \ .$$

A comparison of the EnKF system (22) with the 4D-Var system solved by K Lanczos iterations (12) reveals that the two formulas are nearly identical. EnKF solves a modified 4D-Var problem, with the inverse of background covariance replaced by the pseudo-inverse of the ensemble background covariance; the system is solved via a reduction over the ensemble subspace.

Note that a reduction of the original 4D-Var system (7) onto the subspace of randomly sampled ensemble deviations does not give correct results since

$$\widehat{\mathbf{X}}_0^T \left(\mathbb{B}_0^{-1} + \mathbf{M}^T H^T \mathbb{R}_F^{-1} H \mathbf{M}\right) \widehat{\mathbf{X}}_0 \ \approx \ \frac{n-1}{K-1}\, I_{K\times K} + \widehat{\mathbf{X}}_0^T \mathbf{M}^T H^T \mathbb{R}_F^{-1} H \mathbf{M} \widehat{\mathbf{X}}_0$$

which is (considerably) different than the system matrix in (22).

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 in the first step, and is given by the assimilation history in subsequent steps.

2.5. Statistical Properties of the Equivalent Initial Ensemble

We now consider the construction of an equivalent initial ensemble, i.e., the choice of vectors ξ such that (18) holds. A first idea is to use the initial perturbations (11), i.e., to replace the random vectors by the scaled Lanczos directions:

$$\begin{aligned}
\xi &= \sqrt{K-1} \cdot \widetilde{V} \ , \\
\mathbf{x}'_i &= \mathbb{B}_0^{1/2}\, \xi_i = \sqrt{K-1} \cdot \mathbb{B}_0^{1/2}\, \widetilde{v}_i \ , \quad i = 1, \cdots, K \ .
\end{aligned} \tag{24}$$

Note that the initial perturbations in the regular EnKF have zero mean. On the other hand the Lanczos orthonormal directions $\widetilde{v}_i$ are independent, and therefore their ensemble mean is nonzero,

$$\langle \widetilde{v} \rangle = \frac{1}{K} \sum_{i=1}^{K} \widetilde{v}_i = \frac{1}{K} \widetilde{V} \cdot \mathbb{1}_K \neq 0$$

Consequently the ensemble (24) is biased and performs an adjustment of the initial mean state. The bias can be removed by constructing a double-sized ensemble of symmetric perturbations using $\xi \in \mathbb{R}^{n \times 2K}$ as follows

$$\xi_i = \sqrt{\frac{K-1}{2}}\, \widetilde{v}_i \ , \quad \xi_{K+i} = -\sqrt{\frac{K-1}{2}}\, \widetilde{v}_i \ , \quad i = 1, \cdots, K \ . \tag{25}$$

The mean is zero and the equivalence property (18) holds exactly

$$\frac{1}{K-1}\, \xi\, \xi^T \ = \ \left[\frac{1}{\sqrt{2}}\widetilde{V}, -\frac{1}{\sqrt{2}}\widetilde{V}\right] \cdot \left[\frac{1}{\sqrt{2}}\widetilde{V}, -\frac{1}{\sqrt{2}}\widetilde{V}\right]^T = \widetilde{V}\, \widetilde{V}^T \, .$$

The existence of an equivalent initial ensemble completes the proof of Theorem 1.

In practice one seeks to avoid the construction of large ensembles, which are computationally expensive. We now discuss other approaches to generate unbiased initial ensembles with a smaller number of members, and for which the equivalence property (18) holds within some approximation margin.

1. Remove the bias by subtracting the mean from each Lanczos direction.

$$\xi_i = \sqrt{K-1} \cdot \left(\widetilde{v}_i - \frac{1}{K} \widetilde{V}\, \mathbb{1}_K \right) \ , \quad i = 1, \cdots, K. \tag{26}$$

   In this case the resulting initial ensemble covariance is

$$\widehat{\mathbb{B}}_0 \ = \ \mathbb{B}_0^{1/2}\, \widetilde{V} \left(I_{K \times K} - \frac{1}{K}\, \mathbb{1}_K\, \mathbb{1}_K^T \right) \widetilde{V}^T\, B_0^{1/2}$$

   Alternatively, the bias can be removed by adding one additional ensemble member initialized using

$$\xi_i = \sqrt{K-1} \cdot \widetilde{v}_i \ , \quad i = 1, \cdots, K \ ; \quad \xi_{K+1} = -\sqrt{K-1} \cdot \widetilde{V} \cdot \mathbb{1}_K \ . \tag{27}$$

   In this case the initial ensemble covariance reads

$$\widehat{\mathbb{B}}_0 = \mathbb{B}_0^{1/2}\, \widetilde{V} \, \left(I_{K \times K} + \mathbb{1}_K\, \mathbb{1}_K^T \right) \widetilde{V}^T\, \mathbb{B}_0^{T/2}$$

2. The orthonormal Lanczos directions do not provide a random sample. In order to preserve the statistical interpretation of the EnKF the initialization can be performed by a random sampling *of the Lanczos subspace*:

$$\zeta \;\in\; \Big(\mathcal{N}(0,1)\Big)^{K\times K} \;, \quad \xi = \widetilde{V}\cdot\zeta\;. \tag{28}$$

With this choice the initial ensemble is unbiased, and the equivalence property (18) holds in a statistical sense:

$$\begin{aligned}\mathbb{E}\left[\xi\right] &= \widetilde{V}\cdot\mathbb{E}\left[\zeta\right] = 0_{n\times K}\;,\\ \mathbb{E}\left[\frac{1}{K-1}\,\xi\,\xi^T\right] &= \widetilde{V}\cdot\mathbb{E}\left[\frac{1}{K-1}\,\zeta\,\zeta^T\right]\cdot\widetilde{V}^T = \widetilde{V}\cdot\widetilde{V}^T\;.\end{aligned}$$

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.

3.1. The Hybrid Algorithm

We now describe in detail the hybrid data assimilation algorithm; even if the motivation comes from a linear analysis, the algorithm below can be applied to nonlinear systems as well.

1. 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 $j = 1,\ldots,\ell$.
2. Construct $\mathcal{S}_{t_0}$, a matrix whose columns are the normalized 4D-Var differences between adjacent iterations:

$$\mathcal{S}_{t_0} = \left[\frac{\mathbf{x}_0^{(j)}-\mathbf{x}_0^{(j-1)}}{\left\|\mathbf{x}_0^{(j)}-\mathbf{x}_0^{(j-1)}\right\|}\right]_{j=1,\ldots,\ell} \in \mathbb{R}^{n\times\ell}\;. \tag{29}$$

In the linear symmetric case the solution increment belongs to the subspace spanned by the Lanczos vectors. In the nonlinear case the solution increment belongs to the subspace spanned by successive search directions. Therefore, the normalized differences between adjacent iterations play the role of the Lanczos vectors in the general case. Note, however, that they are not orthogonal.

3. Perform a singular value decomposition of $\mathcal{S}_{t_0}$:

$$\mathcal{S}_{t_0} = U\Sigma V^T \, , \tag{30}$$

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 (24) to generate the initial EnKF ensemble.

Alternatively, a less expensive Gram-Schmidt procedure can be used to orthogonalize the columns of $\mathcal{S}_{t_0}$; in this case one chooses (the first) K directions out of the set of ℓ orthogonal vectors.

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

We next discuss qualitatively several aspects of the proposed hybrid approach.

3.2. The 4D-Var Perspective on the Hybrid Approach

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. Instead of solving the 4D-Var problem to convergence over the entire assimilation window, one solves it sub-optimally over a short time sub-window; the less expensive hybrid EnKF then carries out the data assimilation for the entire length of the assimilation window. From a computational standpoint the hybrid algorithm is an ensemble filter, with intermittent short 4D-Var runs used to re-generate the ensemble subspace.

A practical question is how to choose the length of the short 4D-Var windows in relation to the total length of the assimilation window. The answer is likely to depend on the particular dynamics of the underlying model. Therefore, a practical implementation would require an algorithm to monitor the performance of the ensemble filter, and to decide on-line when to regenerate the ensemble subspace by running a new 4D-Var. A theoretical basis for such an adaptive approach is not available, and needs to be the focus of future research.

3.3. The EnKF perspective on the Hybrid Approach

The hybrid method is expected to perform better than the randomly initialized EnKF due to the special sampling of the initial error space. 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.

Another popular approach to initializing the EnKF is to place the initial perturbations along the "bred vectors" (BVs) [52]. The bred vectors share similar properties with the Lyapunov vectors [1, 7]; they have finite amplitude, finite time, and have local properties in space. The BVs capture the maximum error growth directions in the model. For example, for linear systems, the bred vectors are (well approximated by) linear combinations of the dominant eigenvectors (see the numerical experiments described in section 4.1). While the bred error subspace depends only on the model dynamics, the hybrid subspace takes into account both the model dynamics and the data over the short 4D-Var window. In his regard the hybrid initialization has the potential to provide better results than the bred vector initialization.

Other methods of formulating/initializing the EnKF using special basis vectors have been proposed in the literature, including the use of an internal coordinate system [40], and the use of orthogonal bases [34, 50]. A comprehensive comparison of the hybrid approach with other initialization methods is outside the scope of this paper; future research will elucidate the similarities and differences.

In a longer run the error subspace sampled by (any flavor of) EnKF is given by the previous analyses. Thus, over a long assimilation time window, the differences between the analyses given by different versions of EnKF will likely fade away. The hybrid method periodically resamples the error space. (Note that this is also a common practice with particle filters). The resampling involves a short 4D-Var run over the *next* small time interval. At the resampling time the two filters take very different approaches. While the regular EnKF continues with an error subspace constructed based on *past dynamics and past data*, the hybrid EnKF chooses a subspace based on *future dynamics and future data*. Past information is used in the form of the background covariance matrix. Due to this look-ahead property the hybrid EnKF has the potential to perform better than the regular EnKF.

A practical question is whether it is possible to optimally combine the regular subspace, which contains past information, with the hybrid subspace, which contains future information. For example the random sampling (28) could involve $2K$ basis vectors from both subspaces. Future research is needed to fully answer this question.

4. Numerical Experiments

We now illustrate the proposed hybrid data assimilation algorithm using two test problems, one linear and one nonlinear. The performance of the hybrid approach is compared with the regular EnKF as well as with the EnKF with the breeding initialization. While it is difficult to extrapolate the results from simple test problems to complex systems, the numerical results are encouraging and point to the potential usefulness of the hybrid approach.

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

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

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. For our test case, since the total state dimension is seven, with only two major directions where the error grows, three 4D-Var iterations (and three ensemble members, respectively) are sufficient to capture these directions and hence obtain a sub-optimal solution.

The background covariance is constructed with a correlation distance $L = 1$, where each element is computed as follows: follows:

$$\mathbb{B}_0(i,j) = \sigma_i \cdot \sigma_j \cdot \exp\left(-\frac{|i-j|^2}{L^2}\right)\,, \quad i,\, j = 1,\, \ldots,\, n\,, \tag{31}$$

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 $\mathbb{R}_i = \mathrm{diag}\{0.01\}$.

Since the system is linear, the cost function is quadratic, and the 4D-Var solution is obtained by solving a linear system for the six time units. 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. The PCG and the Lanczos approaches are equivalent [45], however, as discussed in [10], practical applications favor PCG due to its low overhead and disk storage requirements.

We use three ensemble members for the EnKF. Covariance inflation could be used to correct for under-sampling errors. Since the test system has only seven states, covariance inflation is not used here. Several versions of the EnKF are implemented as follows:

1. EnKF-Regular. The ensemble is initialized using normal random samples and the perturbed observations version of the algorithm implemented in [14].
2. 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 [52].
3. EnKF-Hybrid. A "short window" 4D-Var solution is obtained by using only the observations t the end of the first 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. The initial bias is removed by subtracting the mean (26).

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:

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

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.

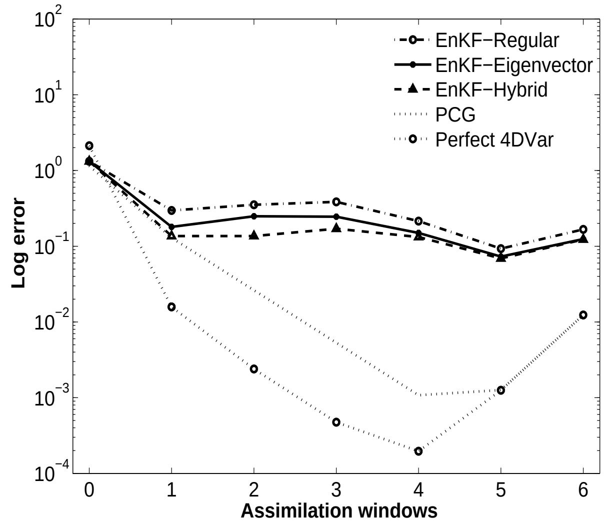

Figure 1: Comparison of analysis errors for several data assimilation methods applied to the linear test problem. Among the EnKF methods the hybrid version is the most accurate.

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 [36]. This chaotic model has $n = 40$ states and is described by the following equations:

$$\begin{aligned} \frac{dx_j}{dt} &= -x_{j-1}\ (x_{j-2} - x_{j+1}) - x_j + F\ ,\quad j = 1, \ldots, n\ , \\ x_{-1} &= x_{n-1}\ ,\quad x_0 = x_n\ ,\quad x_{n+1} = x_1\ . \end{aligned} \tag{32}$$

The forcing term is $F = 8.0$. The Lorenz-96 model has been used to compare 4D-Var and 4D EnKF in [17].

The conventional EnKF method implementation follows the algorithm described in [14]. We compare the following methods:

1. EnKF-Regular: sample normal random numbers to form the perturbation ensemble, then add the perturbations ensemble to the initial best guess (the background initial condition in the first window).

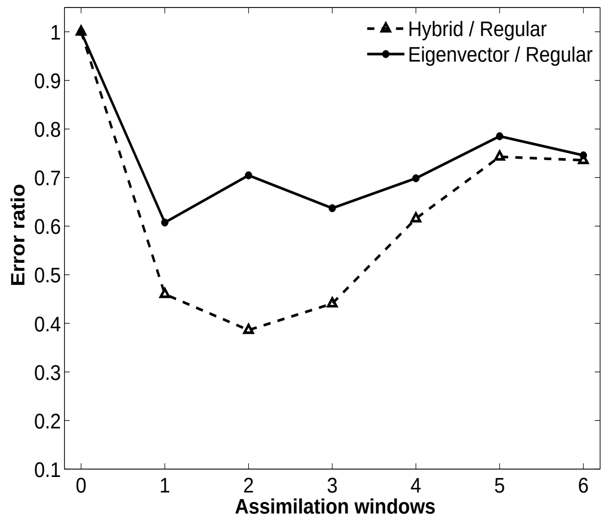

Figure 2: Ratios of analysis errors obtained with different assimilation methods for the linear test. EnKF-eigenvector over regular EnKF (solid), and unbiased EnKF-Hybrid over regular EnKF (dashed).

2. EnKF-Breeding. The "breeding" technique described in [52] is used to capture the maximum error growth directions of the system. The initial ensemble perturbations are set along the bred vectors.
3. 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. The initial bias is removed by subtracting the mean (26).

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). This is very short compared to the total assimilation window of 5 time units.

The background covariance is generated using (31) with $L = 1.0$ and standard deviations equal to 1% of the initial reference values. The breeding EnKF implementation follows the description in [52], 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.

The ensemble filters use neither covariance inflation nor covariance localization. While these techniques are useful in large scale systems, they are not needed here as the number of samples is relatively large compared to the size of the state space. Covariance inflation and localization could be used with any of the ensemble filters if under-sampling becomes an issue.

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 during most of the intervals [0,1.5] and [2.5,3]. The breeding analysis error is smaller on most of the interval [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.

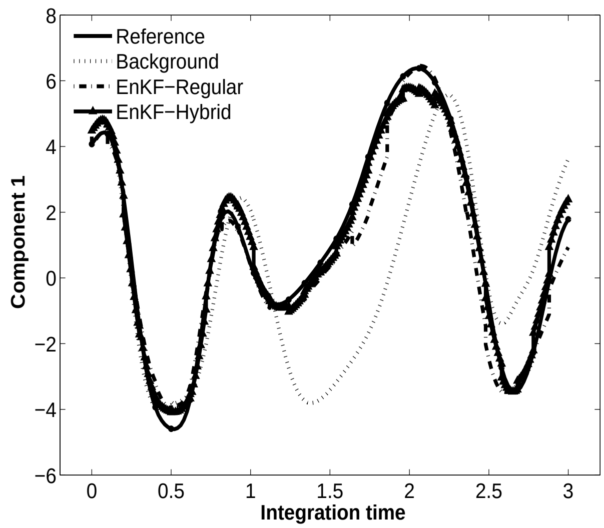

Figure 3: Time evolution of the first Lorenz-96 component for different solutions. Reference (solid line), background (dashed line), analysis with regular EnKF, 10 members (dash dotted line), and analysis with hybrid EnKF, 10 members (solid line with triangles).

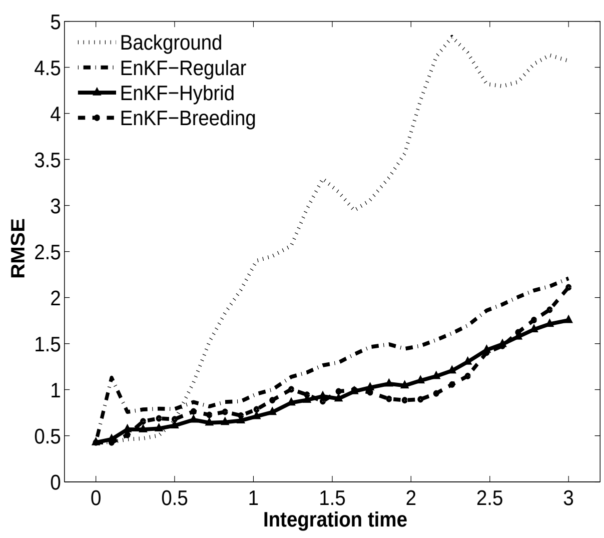

Figure 4: Root mean square error evolution for background, regular EnKF, hybrid EnKF, and breeding EnKF solutions.

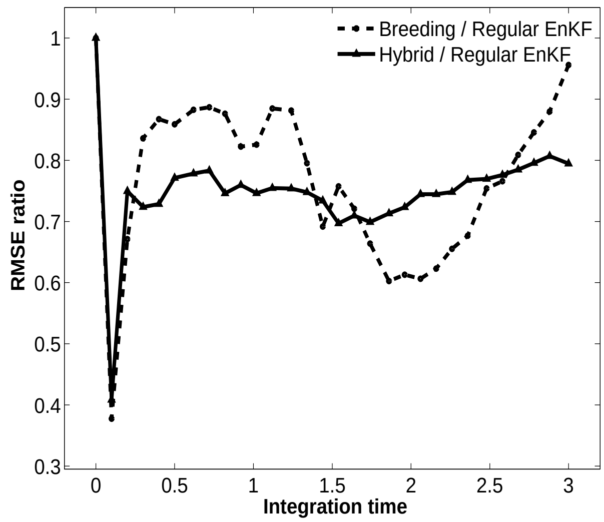

Figure 5: Ratios of analysis errors obtained with different assimilation methods for the nonlinear test. Breeding over regular EnKF (dashed) and hybrid over regular EnKF (solid).

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.

5. Summary

This work takes a subspace perspective on different data assimilation methods. Based on this it establishes 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 subtle relationship 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. The proposed hybrid method is computationally less expensive than a full 4D-Var, as only short assimilation windows are considered, and only a relatively small number of iterations is performed. The hybrid method has the potential to perform better than the regular EnKF due to its look-ahead property. While the regular EnKF uses an error subspace that is either randomly chosen, or constructed based on past dynamics and past data, the hybrid EnKF uses a subspace based on future dynamics and future data. The cost for the hybrid method is the more complex infrastructure required including an adjoint model.

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 will be performed to further understand the properties of hybrid data assimilation approaches.

Several extensions of the present work are possible and needed in order to make the hybrid approach useful in real calculations. A theoretical basis for choosing the length of the short 4D-Var windows in relation to the total length of the assimilation window, and for deciding on-line when to regenerate the ensemble subspace by running a new 4D-Var, is needed. Moreover, one should investigate the possibility to optimally combine the regular subspace, which contains past information, with the hybrid subspace, which contains future information, and to assess the implications of this approach.

Acknowledgements

This work has been supported in part by NSF through the awards NSF CCF 0635194 and DMS 0915047, and by NASA through the AIST 2005 SOX project led by Dr. Meemong Lee. The authors would like to thank Dr. Mohamed Jardak for many fruitful discussions on the topic, for pointing us to very relevant literature, and for his feedback on the manuscript.

References

[1] K. T. Alligood, T. D. Sauer, and J. A. Yorke. *Chaos: an introduction to dynamical systems*. Springer-Verlag, New York, 1996.

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

[3] D. Barker. Hybrid variational/ensemble data assimilation. Buenos Aires, Argentina, November 2008.

[4] A. F. Bennett. *Inverse modeling of the ocean and atmosphere*, volume 352. Cambridge University Press, 2002.

[5] T. Bewley, J. Cessna, and C. Colburn. EnVE:A consistent hybrid ensemble/variational estimation strategy for multiscale uncertain systems. *13th Conference on Integrated Observing and Assimilation Systems for Atmosphere, Oceans, and Land Surface (IOAS-AOLS)*, http://ams.confex.com/ams/89annual/techprogram/program_531.htm, 2009.

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

[7] M. Cai, E. Kalnay, and Z. Toth. Bred vectors of the Zebiak–Cane model and their potential application to ENSO predictions. *Journal of Climate*, 16 (1):40–56, January 2003.

[8] A. Caya, J. Sun, and C. Snyder. A Comparison between the 4D-Var and the EnKF techniques for radar data assimilation. *Monthly Weather Review*, 133:3081–3094, 2005.

[9] A. Clayton, D. Barker, and A. Lorenc. Development of a hybrid variational/ensemble data assimilation system at U.K. Met Office. *Centre for Australian Weather and Climate Research Workshop on Ensemble Prediction and Data Assimilation*, February 2009.

[10] A. Continho, J.L.D. Alves, L. Landau, N.F.F. Ebecken, and L.M. Troina. Comparison of Lanczos and conjugate gradients for the element-by-element solution of finite element equations on the IBM 3090 vector computer. *Computers & Structures*, 39 (1/2):47–55, 1991.

[11] H. Cheng. Uncertainty quantification and uncertainty reduction methods in large scale simulations. *Ph.D. dissertation*, Computer Science Department, Virginia Polytechnic Institute and State University, August 2009.

[12] H. Cheng, M. Jardak, M. Alexe, and A. Sandu. A hybrid approach to estimating error covariances in variational data assimilation. *Tellus A*, in print, 2010.

[13] R. Daley. *Atmospheric Data Analysis*. Cambridge University Press, 1991.

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

[15] G. Evensen. *Data Assimilation:The ensemble Kalman filter*. Springer, Berlin, 2007.

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

[17] E.J. Fertig, J. Harlim, and Brian R. Hunt. A Comparative study of 4D-Var and a 4D-EnKF: perfect model simulations with Lorenz-96. *Tellus*, 59A:96–100, 2007.

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

[19] M. Fisher, M. Leutbecher, and G. A. Kelly. On the equivalence between Kalman smoothing and weak-constraint four-dimensional variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 131: 3235–3246, 2005.

[20] I. Yu Gejadze, F. X. Le Dimet, and V. Shutyaev. On analysis error covariances in variational data assimilation. *SIAM Journal on Scientific Computing*, 30 (4):1847–1874, 2008. ISSN 1064-8275. doi: http://dx.doi.org/10.1137/07068744X.

[21] S. Gratton, A. S. Lawless, and N. K. Nicholas. Approximate gauss-newton methods for nonlinear least squares problems. *SIAM Journal on Optimization*, 18 (1):106–132, February 2007. ISSN 1052-6234.

[22] T. M. Hamill and C. Snyder. A hybrid ensemble Kalman filter-3D variational analysis scheme. *Monthly Weather Review*, 128:2905–2919, 2000.

[23] J. Harlim and B. R. Hunt. Four-dimentional local ensemble transform Kalman filter:numerical experiments with a global circulation model. *Tellus A*, (59A):731–748, 2004.

[24] B. R. Hunt, E. Kalnay, E. J. Kostelich, E. Ott, D. J. Patil, T. D. Sauer, I. Szunyogh, J. A. Yorke, and A. V. Zimin. Four-dimensional ensemble Kalman filtering. *Tellus A*, 56A:273–277, 2004.

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

[26] E. Kalnay. *Atmospheric modelling, data assimilation and predictability*. Cambridge University Press, 2003.

[27] E. Kalnay, H. Li, T. Miyoshi, S.-C. Yang, and J. Ballabrera-Poy. 4D-Var or ensemble Kalman filter? *Tellus A*, 59 (5):758–773, 2007.

[28] M. Krysta, E. Blayo, E. Cosme, C. Robert, J. Verron, and A. Vidard. Hybridisation of data assimilation methods for applications in oceanography. Ocean Sciences Meeting, Orlando, American Geophysical Union , 2008.

[29] M. Krysta, E. Blayo, E. Cosme, C. Robert, J. Verron, and A. Vidard. A 4D-Var SEEK smoother hybrid. Towards applications in oceanography. WWRP/THORPEX Workshop on 4D-Var and Ensemble Kalman filter inter-comparisons, Buenos Aires, Argentina, 2008.

[30] M. Krysta, E. Blayo, E. Cosme, C. Robert, J. Verron, and A. Vidard. Adapting reduced-size control subspace in hybrid data assimilation. GODAE Final Symposium, Nice, France, 2008.

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

[32] Z. Li and I. M. Navon. 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, January 2001.

[33] W. Liao, A. Sandu, G.R. Carmichael, and T. Chai. Total energy singular vector analysis for atmospheric chemical transport models. *Monthly Weather Review*, 134 (9):2443–2465, 2006.

[34] C. Liu, Q. Xiao, and B. Wang. An ensemble-based four-dimensional variational data assimilation scheme. part I:Technical formulation and preliminary test. *American Weather Review*, 136:3363–3373, 2008.

[35] A. C. Lorenc. The potential of the Ensemble Kalman Filter for NWP-A comparison with the 4D-Var. *Quarterly Journal of Royal Meteorology Society*, 129:3183–3203, 2003.

[36] E. N. Lorenz. Predictabilty:A problem partly solved. In *Seminar on Predictability, ECMWF*, Shinfield Park, Reading UK., 1996. European Centre for Medium-Range Weather Forecasting.

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

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

[39] T. Miyoshi, Y. Sato, and T. Kadowaki. Ensemble Kalman filter and 4D-Var inter-comparison with the Japanese operational global analysis and prediction system. *Monthly Weather Review*, in press, 2010.

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

[41] F. Rabier and P. Courtier. Four-dimensional assimilation in the presence of baroclinic instability. *Quarterly Journal of the Royal Meteorological Society*, 118:649–672, 1992.

[42] C. Robert, E. Blayo, and J. Verron. Comparison of reduced-order sequential, variational and hybrid data assimilation methods in the context of a Tropical Pacific ocean model. *Ocean Dynamics* 56:624–633, 2006.

[43] C. Robert, E. Blayo, and J. Verron. Reduced-order 4D-Var: a preconditioner for the full 4D-Var data assimilation method. *Geophysical Research Letters* 33(L18609), `doi:10.1029/2006GL026555`, 2006.

[44] C. D. Rodgers. *Inverse methods for atmospheric sounding :Theory and Practice*, volume Series on Atmospheric Oceanic and Planetary Physics. World Scientific Publishing Company, 2000.

[45] Y. Saad. *Iterative Methods for Sparse Linear Systems*. SIAM, Philadelphia, 2nd edition, 2003.

[46] A. Sandu and H. Cheng. A hybrid variational/ensemble filter approach to data assimilation. Technical Report TR-09-25, Computer Science Department, Virginia Polytechnic Institute and State University, 2009. `http://eprints.cs.vt.edu/archive/00001090`.

[47] J. Sherman and W. J. Morrison. Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *The Annals of Mathematical Statistics*, 21:124–127, 1950.

[48] O. Talagrand and P. Courtier. Variational assimilation of meteorological observations with the adjoint equations Part I:Theory. *Quarterly Journal of Royal Meteorology Society*, 113:1311–1328, 1987.

[49] X. Tian, Z. Xie, and A. Dai. An ensemble-based explicit four-dimensional variational assimilation method. *Journal of Geophysical Research*, 113 (D21124), Nov 2008.

[50] M. K. Tippett, J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker Ensemble square-root filters. *Monthly Weather Review* 131, 1485–1490, 2003.

[51] R. Torn and G. Hakim. Ensemble-based sensitivity analysis. *Monthly Weather Review*, 136 (2):663–677, 2008.

[52] Z. Toth and E. Kalnay. Ensemble forecasting at NCEP and the breeding method. *Monthly Weather Review*, 127:3297–3318, 1997.

[53] X. Wang, C. Snyder, and T. M. Hamill. On the theoretical equivalence of differently proposed ensemble–3D-Var hybrid analysis schemes. *Monthly Weather Review*, 135 (1):222–227, 2007.

[54] X. Wang, D. Barker, C. Snyder, and T. M. Hamill. A hybrid ETKF–3D-Var data assimilation scheme for the WRF model. Part I: Observing system simulation experiment. *Monthly Weather Review* 136:5116–5131. `DOI:10.1175/2008MWR2444.1`, 2008.

[55] X. Wang, D. Barker, C. Snyder, and T. M. Hamill. A hybrid ETKF–3D-Var data assimilation scheme for the WRF model. Part II: Real observation experiments. *Monthly Weather Review* 136:5132–5147. DOI:10.1175/2008MWR2445.1, 2008.

[56] X. Wang, T.M. Hamill, J.S. Whitaker, and C.H. Bishop. A comparison of the hybrid and EnSRF analysis schemes in the presence of model errors due to unresolved scales. *Monthly Weather Review*137(10):3219–3232, 2009.

[57] F. Zhang, M. Zhang, and J. A. Hansen. Coupling ensemble Kalman filter with four-dimensional variational data assimilation. *Advances in Atmospheric Sciences*, 26 (1):1–8, 2009.

[58] M. Zupanski. Maximum likelihood ensemble filter:Theoretical aspects. *American Weather Review*, 133:1710–1726, 2005.