International Journal for Uncertainty Quantification, 5 (6): 491–510 (2015)

AN ERROR SUBSPACE PERSPECTIVE ON DATA ASSIMILATION

***Adrian Sandu*[1,*] *& Haiyan Cheng*[2]**

[1] *Computational Science Laboratory, Department of Computer Science, Virginia Polytechnic Institute and State University, 2201 Knowledgeworks II, 2202 Kraft Drive, Blacksburg, Virginia 24060, USA*

[2] *Department of Computer Science, Willamette University, 900 State Street, Salem, Oregon 97301, USA*

Original Manuscript Submitted: 09/14/2013; Final Draft Received: 08/14/2014

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 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 ensemble filter method performs better than the regular EnKF method for the test problem considered.

KEY WORDS: *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 [1–6]. 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 [7]. 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 [8] 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 [9], is easy to implement, and naturally provides estimates of the analysis covariances.

*Correspond to Adrian Sandu, E-mail: sandu@cs.vt.edu

2152–5080/15/$35.00

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 [10]. In case of Gaussian uncertainty they produce optimal estimates, in a Bayesian sense. 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 [11]. In the presence of model errors the weak-constrained 4D-Var and the fixed-interval Kalman smoother are equivalent [12].

With both methods coming to maturity, new interest in the community has been devoted to assess the relative merits of 4D-Var and EnKF [13, 14]. The better understanding of the strengths of each method has opened the possibility to combine them and build *hybrid data assimilation methods*; In terms of EnKF and 4D-Var hybridization, relevant work can be found in [15–28]. Specifically, in [27], the error subspace statistical estimation (ESSE) and optimal interpolation (OI) schemes are compared to illustrate the advantages and properties of the ESSE. Some hybrid systems aim to improve the background error covariance in the variational system using the information from the ensemble system. In [22], a hybrid background error covariance is used in the operational global data assimilation system at the Met Office, generating an average under 1% reduction of RMS errors. In their hybrid 4D-Var implementation, the background error covariance matrix is a linear combination of the climatological and ensemble covariance. With this update of the background covariance matrix, the authors hope to capture the "error of the day" and improve the forecast system performance. Another attempt to hybridize the ensemble and variational approaches is describes in [23], in which the method is based on the point statistical interpolation (GSI) 3D-Var. The authors experimented various background error covariance updates on the NCEP global forecast system and found that, in general, the hybrid system results in more skillful forecast.

Some hybrid systems investigate the error subspace. In [25], assimilation in the unstable subspace (AUS) algorithm is used in the context of both 4D-Var (4DVar-AUS) and the extended Kalman filter (EKF-AUS). In their work, the state correction is carried out in a unstable and neutral subspace spanned by Lyapunov vectors with positive and zero exponents. Both 4DVar-AUS and EKF-AUS are simpler to implement, less computational demanding, and as efficient as their counterparts. In [26], an improvement is made based on the En4DVar by [29]. The leading singular vectors are extracted from the 4-D ensemble perturbations and are used to construct the analysis increment. The hybrid system results in a significant computational efficiency improvement. An iterative ensemble Kalman smoother (IEnKS) method was implemented in the context of a joint estimation of state variables and parameters on a low-order Lorenz-95 model [28]. The comparisons with EnKF, EnKS and 4D-Var show a systematically outperformance.

A truly hybrid method that couples the 4D-Var and EnKF directly is E4DVAR, proposed in [30]. E4DVAR uses the 4D-Var assimilation procedure to compute analysis means; EnKF is an auxiliary process that provides covariance information. The hybrid scheme proposed in this work is based on EnKF data assimilation; 4D-Var is an auxiliary process that generates the directions of initial ensemble perturbations.

Previous comparative analyses involving ensemble and variational approaches assume large number of ensemble members in EnKF, and many optimization iterations in 4D-Var. 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 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 $n \sim 10^6 - 10^9$ 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 $(\mathbf{x}_0^B - \mathbf{x}_0^t) \in \mathcal{N}(0, \mathbf{B})$. 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 $t_k,\ k = 0, \ldots, N_{\text{obs}} - 1$, 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, $\varepsilon_k \in \mathcal{N}(0, \mathbf{R}_k)$. Data assimilation combines the background estimate $\mathbf{x}_0^B$, the measurements $\mathbf{y}_0, \cdots, \mathbf{y}_{N_{\text{obs}}-1}$, 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 [7]. It provides the analysis $\mathbf{x}_0^A$ as the argument which minimizes the cost function:

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

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 [1, 5, 31], 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 $\mathbf{x}_k = \mathbf{x}_k^B + \Delta\mathbf{x}_k$, $k = 0, \cdots, N_{\text{obs}} - 1$, we have

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

where $\Delta\mathbf{x}_k = \mathbf{M}_{t_0 \to t_k}\ \Delta\mathbf{x}_0$, 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 [8] theory, which gives a minimum variance estimate of the true state under the assumption that the model dynamics and observation operators are all linear. If all errors are additive and Gaussian, then the Kalman filter estimate is optimal in a Bayesian sense. The extended Kalman filter [32] provides a suboptimal state estimation in the nonlinear case by linearizing the model dynamics and the observation operator. We note that in typical large-scale applications (e.g., from meteorology and oceanography) errors are distinctly non-Gaussian [33]. Nevertheless, Kalman filter methods—based on linear theory (and on Gaussian analysis, from a Bayesian point of view)—produce useful results in non-Gaussian and mildly non-linear meteorological and oceanographical applications.

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

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

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

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

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 [32] 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 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 [9, 34] 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 [10], *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 $n \sim 10^6 - 10^9$ 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 suboptimally 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 PhD dissertation [35] and in the technical report [36].

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 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 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 is drawn from a distribution with mean $\mathbb{E}[\mathbf{x}_0^t] = \mathbf{x}_0^B$ and covariance $\mathbb{C}\text{ov}[\mathbf{x}_0^t] = \mathbf{B}_0$. The mean background state and the background covariance at the final time are

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

A single set of measurements is taken at t_F; the measurements are corrupted by unbiased random errors

$$\mathbf{y}_F = H \cdot \mathbf{x}_F^t + \varepsilon_F, \quad \mathbb{E}[\varepsilon_F] = 0, \quad \mathbb{C}\text{ov}[\varepsilon_F] = \mathbf{R}_F.$$

We consider the assimilation window $[t_0, t_F]$. Data assimilation produces a posterior distribution for the initial values, with mean $\mathbb{E}[\mathbf{x}_0^t] = \mathbf{x}_0^A$ and posterior covariance matrix $\mathbb{C}\text{ov}[\mathbf{x}_0^t] = \mathbf{A}_0$. The best posterior estimate $\mathbf{x}_0^A$ is called the *analysis* state. The mean analysis state and the analysis covariance matrix at the final time are

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

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 the case of a linear, perfect model, with observations taken at only one time—at the end of the assimilation window. 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 error covariance matrices for the background and observations.*

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 1. 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 us to draw interesting and useful parallels between 4D-Var and EnKF, and to gain considerable insight.

Comment 2. In a Bayesian framework it is typically assumed that the prior distribution of uncertainty is Gaussian, $\mathbf{x}_0^t \in \mathcal{N}\left(\mathbf{x}_0^B, \mathbf{B}_0\right)$, and the observation errors are normally distributed, $\varepsilon_F \in \mathcal{N}\left(\mathbf{0}, \mathbf{R}_F\right)$. Consequently, the uncertainty in the background state at the final time $\mathbf{x}_F^B$ is also Gaussian. Moreover, the Bayesian posterior probability density is Gaussian, $\mathbf{x}_0^t \in \mathcal{N}\left(\mathbf{x}_0^A, \mathbf{A}_0\right)$. Both the full 4D-Var and the full Kalman filter yield the *optimal* analysis $\mathbf{x}_0^A$. The error normality assumption is not necessary for establishing the equivalence between 4D-Var and EnKF. Without the normality assumption, the analyses produced by the two methods are equivalent, but not optimal in Bayesian sense.

Comment 3. The proof of Theorem 1 is constructive: we show how a particular initialization of the filter can be obtained, such that the filter solution coincides with the suboptimal 4D-Var solution. This construction will then be used in a hybrid version of the filter.

2.1 Full 4D-Var Solution

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

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

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:

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

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 [37], $A=\mathbf{A}_0^{-1}$. 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

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

For nonlinear systems the above procedure (based on linearized dynamics and observation operator) corresponds to 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 [38].

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 [39] 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. For this particular approximation we chose, the background covariance is known, and this covariance offers a popular preconditioner. Assume that a Cholesky or a symmetric square root decomposition $\mathbf{B}_0^{1/2}$ of $\mathbf{B}_0$ is available:

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

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:

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

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 [39]

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

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 $\mathcal{K}_K$,

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

An explicit form of the solution (9) can be obtained using the Sherman-Morrison-Woodbury formula [40, 41]

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

with

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

Together with the notation

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

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

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

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

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

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

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

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

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

where the last relation follows from another application of the Sherman-Morrison-Woodbury formula when $\widetilde{\mathbf{B}}_F$ is invertible.

Finally, consider the "initial perturbations"

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

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

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

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 $\mathcal{N}(0, \mathbf{B}_0)$. Equivalently, they are obtained via a variable transformation from the standard normal vectors ξ_i as follows:

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

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

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

and the matrix of the scaled initial perturbations by

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

The ensemble covariance is

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

and, for a perfect and linear model,

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

The EnKF analysis updates each member using the formula

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

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

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

Comment 4. Other popular approaches to initialize the ensemble are the breed vectors, the total energy singular vectors, and the Hessian singular vectors [42]. 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:

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

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 $\mathbf{B}_0$) 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

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

while in the 4D-Var case

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

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

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

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 [43]:

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

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

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

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

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

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

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

$$\left(\mathbf{I}_{K\times K} + \widehat{\mathbf{X}}_0^T\mathbf{M}^T H^T\mathbf{R}_F^{-1}H\mathbf{M}\widehat{\mathbf{X}}_0\right)\cdot\widehat{\theta} = \widehat{\mathbf{X}}_0^T\mathbf{M}^T H^T\mathbf{R}_F^{-1}d_F^B \tag{22}$$
$$\Delta\langle\mathbf{x}_F\rangle = \langle\mathbf{x}_F^A\rangle - \langle\mathbf{x}_F^B\rangle = \widehat{\mathbf{X}}_F\cdot\widehat{\theta} = \mathbf{M}\mathbf{B}_0^{1/2}\widehat{V}\cdot\widehat{\theta}$$
$$\Delta\langle\mathbf{x}_0\rangle = \langle\mathbf{x}_0^A\rangle - \langle\mathbf{x}_0^B\rangle = \widehat{\mathbf{X}}_0\cdot\widehat{\theta} = \mathbf{B}_0^{1/2}\widehat{V}\cdot\widehat{\theta}$$

We can rewrite (22) as

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

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

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

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

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

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:

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

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

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

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.

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

In this case the resulting initial ensemble covariance is

$$\widehat{\mathbf{B}}_0 = \mathbf{B}_0^{1/2}\,\widetilde{V}\,\left(\mathbf{I}_{K\times K} - \frac{1}{K}\,\mathbf{1}_K\,\mathbf{1}_K^T\right)\,\widetilde{V}^T\,B_0^{1/2}. \tag{27}$$

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

$$\xi_i = \sqrt{K-1}\cdot\widetilde{v}_i,\;\; i=1,\cdots,K;\quad \xi_{K+1} = -\sqrt{K-1}\cdot\widetilde{V}\cdot\mathbf{1}_K. \tag{28}$$

In this case the initial ensemble covariance reads

$$\widehat{\mathbf{B}}_0 = \mathbf{B}_0^{1/2}\,\widetilde{V}\,\left(\mathbf{I}_{K\times K} + \mathbf{1}_K\,\mathbf{1}_K^T\right)\,\widetilde{V}^T\,\mathbf{B}_0^{T/2}. \tag{29}$$

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*:

$$\zeta \in \Big(\mathcal{N}(0,1)\Big)^{K\times K},\quad \xi = \widetilde{V}\cdot\zeta. \tag{30}$$

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

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

3. A HYBRID APPROACH TO DATA ASSIMILATION

The above analysis reveals a subtle similarity between the 4D-Var and EnKF analyses for the linear 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 directions. A version of the proposed hybrid algorithm could be based on using distinct observation sets in the 4D-Var and in the EnKF calculations.

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:

$$\mathcal{S}_{t_0} = \left[\frac{\mathbf{x}_0^{(j)} - \mathbf{x}_0^{(j-1)}}{\left\| \mathbf{x}_0^{(j)} - \mathbf{x}_0^{(j-1)} \right\|} \right]_{j=1,\ldots,\ell} \in \mathbb{R}^{n \times \ell}. \tag{31}$$

   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 $\mathcal{S}_{t_0}$:

$$\mathcal{S}_{t_0} = U \Sigma V^T, \tag{32}$$

   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 suboptimally over a short time subwindow; 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 regenerate 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 initialize the EnKF is to place the initial perturbations along the "bred vectors" (BVs) [44]. The bred vectors share similar properties with the Lyapunov vectors [45, 46]; 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. 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 this 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 [43], and the use of orthogonal bases [19, 47]. In [48], the author has generated EnKF ensembles initialized with rescaled breeding random perturbations (BGV), balanced random perturbations (BAL), and "gridpoint" random perturbations (RDM). Numerical experiments have revealed that the corresponding error covariance matrices have similar structure.

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.

Note that 4D-Var plays the role of an auxiliary process to initialize EnKF. It identifies a subspace of large errors; concentrating the computational effort to this subspace decreases the sampling error. Note that identifying a particular subspace does not change the prior probability density. The EnKF data assimilation procedure (by itself) continues to follow the fundamental rules of Bayes statistics. The choice of the EnKF initialization comes down to selecting a subspace where to represent the errors, and building a reduced rank filter in that subspace. All choices are consistent with the Bayesian view in the sense that they lead to the full Bayesian approach in the limit of the subspace dimension approaching the state space dimension. Different choices, however, affect the accuracy of the resulting (approximate) analysis when the subspace dimension is small.

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 (30) 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 a nonlinear test problem. 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 (where variational and ensemble algorithms show similar performance [49, 50]), the numerical results are encouraging and point to the potential usefulness of the hybrid approach.

The nonlinear test is carried out with the Lorenz-96 model [51]. This chaotic model has $n = 40$ states and is described by the following equations:

$$\begin{aligned} &\frac{dx_j}{dt} = -x_{j-1}\ (x_{j-2} - x_{j+1}) - x_j + F\ , \quad j = 1, \ldots, n, \\ &x_{-1} = x_{n-1}, \quad x_0 = x_n, \quad x_{n+1} = x_1. \end{aligned} \tag{33}$$

The forcing term is $F = 8.0$. The Lorenz-96 model has been used to compare 4D-Var and 4D EnKF in [52].

The conventional EnKF method implementation follows the algorithm described in [53]. 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).

2. EnKF-Breeding. The "breeding" technique described in [44] 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 0.1 time units with five observations. The directions generated by the L-BFGS numerical optimization routine are used to initialize the hybrid ensemble as explained in Section 3. This ensemble regeneration procedure is repeated periodically for every 50 observations. The initial bias is removed by subtracting the mean (26).

The 40 components of the state are first drawn from a uniform distribution in $[-2, 2]$, and a spin-up run of 10 time units is performed. The state of the model at the end of the spin-up phase is the reference initial condition; the reference solution is obtained by integrating (33) through the 10 time units simulation window using Matlab's ode45 with tight relative and absolute tolerances ($RelTol = AbsTol = 10^{-9}$). Synthetic observations are obtained every 0.02 time units from the reference solution, with added normal random noise. The standard deviation of the noise is $\sigma_i^{\text{obs}} = 0.005\,\mathbf{x}_i^{\text{reference}}$. We consider two test cases, as follows:

1. the "dense data" case includes observations of all 40 states, while

2. the "sparse data" case includes observations of every third state only, i.e., of states $1, 4, 7, \ldots, 37, 40$.

The discrete model advances the state in time increments of 0.02 time units, which corresponds to four time steps of the underlying fourth-order Runge-Kutta numerical integration scheme. The difference between the fixed step Runge-Kutta solution and the variable step, tight tolerance reference solution represents the model error.

The background correlation is a weighted average between the identity matrix and an exponential decay correlation matrix.

$$\mathbf{B}_0(i,j) = \sigma_i^{\text{b}} \cdot \sigma_j^{\text{b}} \cdot \left(0.1\,\delta_{i,j} + 0.9\,\exp\left(-\frac{|i-j|^2}{L^2} \right) \right), \quad i,\, j = 1,\, \ldots,\, n, \tag{34}$$

where $\delta_{i,j}$ is the Kronecker delta symbol, $L = 4$, and $\sigma_i^{\text{b}} = 0.03\,(\mathbf{x}_0)_i$. This is very similar with the one chosen in [54].

The background state is the reference initial condition, plus a perturbation drawn from the corresponding unbiased normal distribution. A more informative background covariance matrix for 4D-Var can be constructed using the ensemble information from the previous assimilation window.

For EnKF, we use both covariance inflation and covariance localization mechanism with the following setups. The EnKF background covariance matrix at the initial time is (34), and is used to generate the background ensemble. A covariance inflation with a factor of $\alpha = 1.05$ is included at each cycle to alleviate the filter divergence problem. The tests use ensembles with different numbers of members. The same inflation and localization strategy is applied for the breeding and hybrid filters.

The hybrid filter runs EnKF for a long assimilation window of one time unit. After one time unit, a new short window 4D-Var run is performed, the ensemble is reinitialized, and EnKF is used to perform data assimilation for the subsequent long window. The EnKF ensemble reinitialization process uses the mean state computed from the end of the previous long window. The 4D-Var runs for reinitializing the ensemble directions use short windows of length 0.1 time units. Each short window contains five observations. The optimization stops after a number of iterations equal the number of ensemble members plus one, and L-BFGS stores 10 vector pairs. The intermediate solutions generated by the optimization are saved as $x_0^{(j)}$, and relation (31) is used to construct the subspace spanned by successive search directions $x_0^{(j)} - x_0^{(j-1)}$. We then apply (32) to identify the error reduction directions and use these directions to initialize the EnKF ensemble as shown in (24).

The breeding EnKF implementation follows the description in [44], where the perturbations are propagated with the system for one time unit and rescaled. The propagation and rescaling are carried out 10 times in succesion.

Both the breeding and the hybrid filters regenerate the EnKF ensemble directions every two time units, i.e., at simulation times 2, 4, 6, and 8 units.

The full-fledged 4D-Var run uses assimilation windows that are one time unit long, and processes all 50 observations within this window. The L-BFGS optimization stores the last 10 vector pairs, and stops when both the projected gradient norm and the cost function values have decreased by a factor of 1000. The number of iterations is larger in the first window, but decreases after the spin up period.

To assess the accuracy of each assimilation method we use the following relative error metric:

$$\text{Err} = \sqrt{\frac{\sum_{i=1}^{n} \left(x_i^{\text{analysis}} - x_i^{\text{reference}} \right)^2}{\sum_{i=1}^{n} \left(x_i^{\text{reference}} \right)^2}}. \tag{35}$$

The analyses relative errors for the dense and the sparse data test cases are shown in Figs. 1 and 2, respectively. Table 1 shows the computational times of various methods for the two test cases. The Matlab code is not optimized, and therefore the results can only be interpreted in a qualitative sense. In particular the cost of 4D-Var, based on direct nonlinear minimization, is very large. The 4D-Var CPU time can be considerably improved by tuning the code, and by using advanced algorithms such as sequential quadratic programming (solving a sequence of incremental 4D-Var problems) and preconditioning.

In the "dense data" scenario the numbers of ensemble members are chosen such that the total CPU times are similar for all methods. The relative errors of the corresponding analyses are shown in Fig. 1. We notice that the performance of all ensemble members is roughly similar in terms of accuracy. However, the hybrid filter achieves this accuracy using only nine ensemble members. This is a promising result, especially for practical large models where only a small ensembles are feasible.

In the "sparse data" scenario all methods use 15 ensemble members. EnKF is the fastest, while the breeding and the hybrid methods have about the same CPU time in this setting. The relative errors of the corresponding analyses are shown in Fig. 2; as expected, the errors are larger than in the "dense data" scenario. The performance of the methods is similar except for the last two time units, where EnKF is slightly less accurate. We notice an actual increase in error after resampling at time 8 units for both the hybrid and the breeding approaches. This indicates that the new directions computed at this particular point are not of sufficient quality. In this case keeping the old ensemble (like

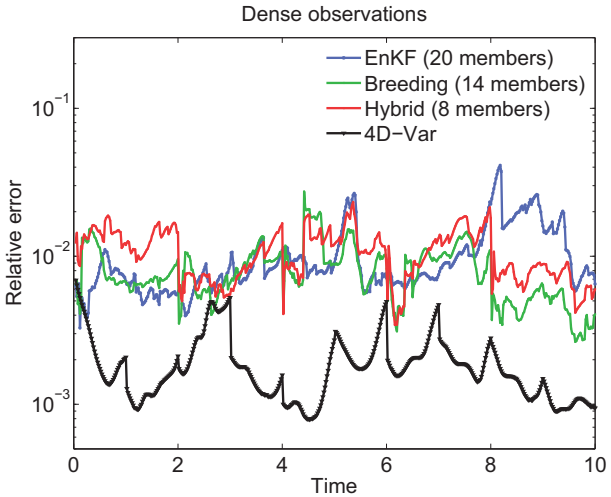

FIG. 1: Relative errors for EnKF, Breeding, Hybrid, and 4D-Var analyses in the "dense data" test case (all state components are observed; observation error std is 0.005 times the reference solution).

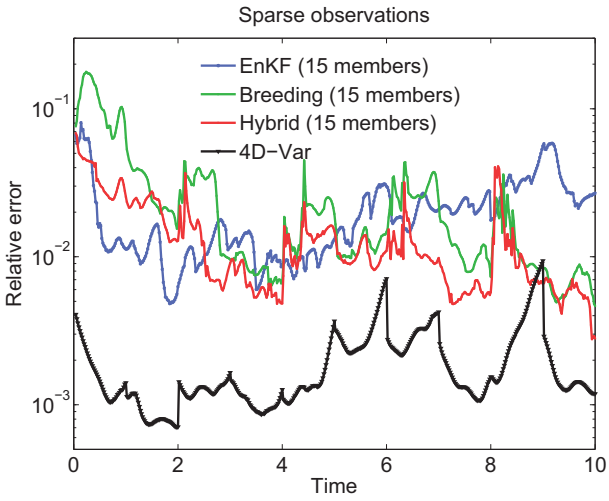

FIG. 2: Relative errors for EnKF, Breeding, Hybrid, and 4D-Var analyses in the "sparse data" test case (every third state components are observed; observation error std is 0.005 times the reference solution).

TABLE 1: Computational times for different data assimilation methods on the two scenarios. One CPU time unit is the wall-clock time needed to perform one Lorenz-96 model integration for the entire simulation interval

Method	Dense data		Sparse data	
	Nens	CPU time	Nens	CPU time
EnKF	20	22.2	15	16.7
Breeding	14	19.6	15	21.1
Hybrid	8	20.4	15	21.8
4D-Var	—	260	—	280

EnKF does) would be a better strategy. More research is needed to find algorithms that integrate the look-ahead directions computed by breeding or 4D-Var with the old (EnKF) directions for improved analysis.

The numerical tests in the Lorenz '96 chaotic nonlinear case show the hybrid method provides similar analyses as the regular EnKF but requires fewer ensemble members. The hybrid solution takes significantly less computational time than the full 4D-Var solution. The relative costs of running additional ensemble members, versus running short 4D-Var sequences, determines the relative performance of the two approaches.

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 show that the proposed hybrid approach provides an analysis accuracy similar to the regular EnKF, but uses a smaller number of ensemble members. For the same number of members, 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. Future work is also needed to improve the computational efficiency of the hybrid method.

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.

ACKNOWLEDGMENTS

This work has been supported in part by NSF through the awards NSF CCF-0635194, NSF CCF-0916493, NSF OCI-0904397, NSF CMMI-1130667, and DMS-0915047, by AFOSR through the award FA9550–12–1–0293–DEF (DDDAS program), and by the Computational Science Laboratory at Virginia Tech. The authors would like to thank Dr. Mohamed Jardak for discussions on the topic. The authors would also like to thank the anonymous reviewers whose feedback helped improve this work.

REFERENCES

1. Bennett, A. F., *Inverse Modeling of the Ocean and Atmosphere*, Vol. 352, Cambridge University Press, Cambridge, 2002.
2. Daley, R., *Atmospheric Data Analysis*, Cambridge University Press, Cambridge, 1991.
3. Evensen, G., *Data Assimilation: The Ensemble Kalman Filter*, Springer, Berlin, 2007.
4. Kalnay, E., *Atmospheric Modelling, Data Assimilation and Predictability*, Cambridge University Press, Cambridge, 2003.
5. Lewis, J., Lakshmivarahan, S., and Dhall, S., *Dynamic Data Assimilation: A Least Squares Problem*, Cambridge University Press, Cambridge, 2005.
6. Rodgers, C. D., *Inverse Methods for Atmospheric Sounding: Theory and Practice*, Series on Atmospheric Oceanic and Planetary Physics, World Scientific Publishing Company, 2000.
7. Talagrand, O. and Courtier, P., Variational assimilation of meteorological observations with the adjoint equations Part I: Theory, *Q. J. R. Meteorol. Soc.*, 113:1311–1328, 1987.
8. Kalman, R. E., A new approach to linear filtering and prediction problems, *Trans. ASME: J. Basic Eng.*, 82:35–45, 1960.
9. Anderson, J. L., A local least squares framework for ensemble filtering, *Mont. Weather Rev.*, 131(4):634–642, 2003.
10. Li, Z., and Navon, I. M., Optimality of 4D-Var and its relationship with the Kalman filter and Kalman smoother, *Q. J. R. Meteorol. Soc.*, 127(572):661–684, 2001.
11. Fisher, M., Leutbecher, M., and Kelly, G. A., On the equivalence between Kalman smoothing and weak-constraint four-dimensional variational data assimilation, *Q. J. R. Meteorol. Soc.*, 131:3235–3246, 2005.
12. Menard, R., and Daley, R., The application of Kalman smoother theory to the estimation of 4D-Var error statistics, *Tellus A*, 48:221–237, 1996.
13. Kalnay, E., Li, H., Miyoshi, T., Yang, S.-C., and Ballabrera-Poy, J., 4D-Var or ensemble Kalman filter, *Tellus A*, 59(5):758–773, 2007.
14. Lorenc, A. C., The potential of the ensemble Kalman filter for NWP-A comparison with the 4D-Var, *Q. J. R. Meteorol. Soc.*, 129:3183–3203, 2003.

15. Barker, D., *Hybrid Variational/Ensemble Data Assimilation*, Buenos Aires, Argentina, 2008.
16. Cheng, H., Jardak, M., Alexe, M., and Sandu, A., Dynamic meteorology and oceanography, *Tellus A*, 62(3):288–297, 2010.
17. Hunt, B. R., Kalnay, E., Kostelich, E. J., Ott, E., Patil, D. J., Sauer, T. D., Szunyogh, I., Yorke, J. A., and Zimin, A. V., Four-dimensional ensemble Kalman filtering, *Tellus A*, 56A:273–277, 2004.
18. Harlim, J., and Hunt, B. R., Four-dimentional local ensemble transform Kalman filter: Numerical experiments with a global circulation model, *Tellus A*, 59A:731–748, 2004.
19. Liu, C., Xiao, Q., and Wang, B., An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formulation and preliminary test, *Am. Weather Rev.*, 136:3363–3373, 2008.
20. Robert, C., Blayo, E., and Verron, J., Reduced-order 4D-Var: A preconditioner for the full 4D-Var data assimilation method, *Geophys. Res. Lett.*, 33:L18609, 2006.
21. Tian, X., Xie, Z., and Dai, A., An ensemble-based explicit four-dimensional variational assimilation method, *J. Geophys. Res.*, 113:D21124, 2008.
22. Clayton, A. M., Operational implementation of a hybrid ensemble/4D-Var global data assimilation system at the met office, *Q. J. R. Meteorol. Soc.*, 139:1445–1461, 2013.
23. Wang, X., Parrish, D., Kleist, D., and Whitaker, J., GSI 3D-Var-based ensemble-variational hybrid data assimilation for the NCEP global forecast system: Single resolution experiments, *Mon. Weather Rev.*, 141:4098–4117, 2013.
24. Jardak, M. and Talagrand, O., Ensemble variational assimilation and Bayesian estimation, *DARC Meetings*, Reading, UK, June, 2014.
25. Palatella, L., Carrassi, A., and Trevisan, A., Lyapunov vectors and assimilation in the unstable subspace: Theory and applications, *J. Phys. Math. Theor.*, 46:254020, 2013.
26. Shao, A., Xi, S., Qiu, C., and Xu, Q., A hybrid space approach for ensemble-based 4-D variational data assimilation, *J. Geoph. Res.*, 114:D17114, 2009.
27. Lermusiaux, P. and Robinson, A., Data assimilation via error subspace statistical estimation. Part I: Theory and schemes, *Mon. Weather Rev.*, 127:1385–1407, 1999.
28. Bocquet, M. and Sakov, P., Joint state and parameter estimation with an iterative ensemble Kalman smoother, *Nonlin. Proc. Geophys.*, 20(5):803–818, 2013.
29. Qiu, C., Shao, A., Xu, Q., and Li, W., Fitting model fields to observations by using singular value decomposition: An ensemble-based 4DVar approach, *J. Geophys. Res.*, 112:D11, 2007.
30. Zhang, F., Zhang, M., and Hansen, J. A., Coupling ensemble Kalman filter with four-dimensional variational data assimilation, *Adv. Atmos. Sci.*, 26(1):1–8, 2009.
31. Rabier, F. and Courtier, P., Four-dimensional assimilation in the presence of baroclinic instability, *Q. J. R. Meteorol. Soc.*, 118:649–672, 1992.
32. Fisher, M., Assimilation techniques (5): Approximate Kalman filters and singular vectors, *European Center Medium-Range Weather Forecast*, 2002.
33. Pires, C. A., Talagrand, O., and Bocquet, M., Diagnosis and impacts of non-Gaussianity of innovations in data assimilation, *Phys. D: Nonlin. Phenomena*, 239(17):1701–1717, 2003.
34. Burgers, G., van der Leeuwen, P. J., and Evensen, G., Analysis scheme in the ensemble Kalman filter, *Mon. Weather Rev.*, 126:1719–1724, 1998.
35. Cheng, H., *Uncertainty Quantification and Uncertainty Reduction Methods in Large Scale Simulations*, PhD Dissertation, Computer Science Department, Virginia Polytechnic Institute and State University, 2009.
36. Sandu, A. and Cheng, H., A hybrid variational/ensemble filter approach to data assimilation, Technical Report TR-09-25, Computer Science Department, Virginia Polytechnic Institute and State University, 2009.
37. Gejadze, I. Yu., Le Dimet, F. X., and Shutyaev, V., On analysis error covariances in variational data assimilation, *SIAM J. Sci. Comput.*, 30(4):1847–1874, 2008.
38. Gratton, S., Lawless, A. S., and Nicholas, N. K., Approximate Gauss-Newton methods for nonlinear least squares problems, *SIAM J. Optimiz.*, 18(1):106–132, 2007.
39. Saad, Y., *Iterative Methods for Sparse Linear Systems*, 2nd ed., SIAM, Philadelphia, 2003.

40. Mandel, J., 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)*, 2007.
41. Sherman, J. and Morrison, W. J., Adjustment of an inverse matrix corresponding to a change in one element of a given matrix, *Annals Math. Stat.*, 21:124–127, 1950.
42. Liao, W., Sandu, A., Carmichael, G. R., and Chai, T., Total energy singular vector analysis for atmospheric chemical transport models, *Mon. Weather Rev.*, 134(9):2443–2465, 2006.
43. Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., Corazza, M., Kalnay, E., Patil, D. J., and Yorke, J. A., A local ensemble Kalman filter for atmospheric data assimilation, *Tellus*, 56A:415–428, 2004.
44. Toth, Z. and Kalnay, E., Ensemble forecasting at NCEP and the breeding method, *Mon. Weather Rev.*, 127:3297–3318, 1997.
45. Alligood, K. T., Sauer, T. D., and Yorke, J. A., *Chaos: An Introduction to Dynamical Systems*, Springer-Verlag, New York, 1996.
46. Cai, M., Kalnay, E., and Toth, Z., Bred vectors of the Zebiak–Cane model and their potential application to ENSO predictions, *J. Climate*, 16(1):40–56, 2003.
47. Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M., and Whitaker, J. S., Ensemble square-root filters, *Mon. Weather Rev.*, 131:1485–1490, 2003.
48. Zhang, F., Dynamics and structure of mesoscale error covariance of a winter cyclone estimated through short-range ensemble forecasts, *Mon. Weather Rev.*, 133:2876–2893, 2005.
49. Buehner, M., Houtekamer, P. L., Charette, C., Mitchell, H. L., and He, B., Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: Description and single-observation experiments, *Mon. Weather Rev.*, 138:1550–1566, 2010.
50. Buehner, M., Houtekamer, P. L., Charette, C., Mitchell, H. L., and He, B., Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part II: One-month experiments with real observations, *Mon. Weather Rev.*, 138:1567–1586, 2010.
51. Lorenz, E. N., Predictabilty: A problem partly solved, In *Seminar on Predictability, ECMWF*, Shinfield Park, European Centre for Medium-Range, 1996.
52. Fertig, E. J., Harlim, J., and Hunt, B. R., A Comparative study of 4D-Var and a 4D-EnKF: Perfect model simulations with Lorenz-96, *Tellus*, 59A:96–100, 2007.
53. Evensen, G., The ensemble Kalman filter: Theoretical formulation and practical implementation, *Ocean Dyn.*, 53:343–367, 2003.
54. Navon, I. M., Daescu, D. N., and Liu, Z., The impact of background error on incomplete observations for 4D-Var data assimilation with the FSU GSM, *Comput. Sci. ICCS*, 3515:837–844, 2005.