A hybrid (variational/Kalman) ensemble smoother for the estimation of nonlinear high-dimensional discretizations of PDE systems

Joseph Cessna, *Student Member, IEEE,* and Thomas Bewley, *Member, IEEE*

***Abstract*—Two classes of state estimation schemes, variational (4DVar) and ensemble Kalman (EnKF), have been developed and used extensively by the weather forecasting community as tractable alternatives to the standard matrix-based Kalman update equations for the estimation of high-dimensional nonlinear systems with possibly nongaussian PDFs. Variational schemes iteratively minimize a finite-horizon cost function with respect to the state estimate, using efficient vector-based gradient descent methods, but fail to capture the moments of the PDF of this estimate. Ensemble Kalman methods represent well the principal moments of the PDF, accounting for the measurements with a sequence of Kalman-like updates with the covariance of the PDF approximated via the ensemble, but fail to provide a mechanism to reinterpret past measurements in light of new data. In this paper, we first introduce a tractable method for updating an ensemble of estimates in a variational fashion, capturing correctly both the estimate (via the ensemble mean) and the leading moments of its PDF (via the ensemble distribution). We then extend this variational ensemble framework to facilitate its consistent hybridization with the ensemble Kalman smoother. Finally, it is shown (on a low-dimensional model problem) that the resulting Hybrid (variational/Kalman) Ensemble Smoother (HEnS), which inherits the tractable extensibility to high-dimensional systems of the component methods upon which it is based, significantly outperforms the existing 4DVar and EnKF approaches used operationally today for high-dimensional state estimation.**

***Index Terms*—state estimation, data assimilation, ensemble Kalman, variational methods, smoothing**

I. Introduction

THE estimation and forecasting of chaotic, multiscale, uncertain fluid systems is one of the most highly visible grand challenge problems of our generation. Specifically, this class of problems includes weather forecasting, climate forecasting, and flow control. The financial impact of a hurricane passing through a major metropolitan center regularly exceeds a billion dollars. Improved atmospheric forecasting techniques provide early and accurate warnings, which are critical to minimize the impact of such events. On longer time scales, the estimation and forecasting of changes in ocean currents and temperatures is essential for an improved understanding of changes to the earth's weather systems. On shorter time scales, feedback control of fluid systems (for reasons such as minimizing drag, maximizing harvested energy, etc.) in mechanical, aerospace, environmental, energy, and chemical engineering settings lead to a variety of similar estimation problems. While this paper makes no claims with regards to addressing the particular details of any of these important applications, it does introduce a new Hybrid (variational/Kalman) Ensemble Smoother (HEnS) for the estimation and forecasting of such multiscale uncertain fluid systems that might well have a transformational effect in all of these areas.

J. Cessna and T. Bewley are with the Department of Mechanical and Aerospace Engineering, University of California San Diego, La Jolla, CA, 92093-0411 USA email: jcessna@ucsd.edu, bewley@ucsd.edu

Manuscript received April ??, 20??; revised January ??, 20??.

A. Historical developments

Much of the research today in state estimation (a.k.a. data assimilation) for multiscale uncertain fluid systems is focused on short- to medium-range weather forecasting. Towards this end, the methods available for this class of problems have matured greatly in the past 25 years. To set the stage, we must first mention a few related developments.

The full, correct answer to the state estimation of nonlinear systems with finite (and, thus, nongaussian) uncertainties dates back to the late 1950s (see [1]). As described clearly on page 164 of [2], it combines two simple steps:
(i) between measurement updates, the full probability density function (PDF) in phase space is propagated via the Kolmogorov forward equation (a.k.a. Fokker-Planck equation);
(ii) at measurement updates, the PDF is updated via application of Bayes' theorem.
During step (i), the PDF stretches and diffuses; during step (ii), the PDF is refocused. An efficient modern implementation of this idea using a grid-based method, leveraging effectively the fact that the PDF is usually nearly zero almost everywhere in phase space, is given in [3]; unfortunately, such methods are numerically intractable for systems with states of order $n \gtrsim 10$, even with modern supercomputing resources.

Particle filters (PF; see [4]) approximate the solution of such Bayesian estimation strategies using a Lagrangian approach. With such methods, a set of candidate state trajectories is followed to track the evolution of the probability distribution in time, and associated with each particle is a weight, which is modified via Bayes' theorem at each measurement update (while normalizing such that the weights always add to one). Unfortunately, application of such updates for successive measurements invariably leads to most weights being driven towards zero as the algorithm proceeds, a phenomenon known as *degeneracy*. To counter this tendency in order to maintain adequate resolution of the significant (nonzero) portion of

the PDF, resampling of the PDF with a new distribution of particles (with equalized weights) is, from time to time, required. A variety of such resampling algorithms have been proposed. When using a large number of particles (necessary when attempting to resolve a nongaussian PDF of the state estimate), the sampling importance resampling algorithm proposed in [5] is commonly used. When using small number of particles N (specifically, for $N = 2n + 1$, used when considering a state estimate of order n with a Gaussian PDF), an unscented transform (see [6], [7]) can be used to resample while preserving exactly the covariance of the original distribution. Unfortunately, PFs are also numerically intractable in large-scale systems.

Kalman filters (see [8], [1], [9], [10], [11]) substantially simplify the full state estimation problem in the common situation in which the random variables are all well approximated by Gaussian PDFs. In this case, the PDF of a given random variable, of order n, can be specified completely by keeping track of its mean (of order n) and its covariance (of order n^2), which enormously simplifies the complexity of the state estimation problem. With modern computational resources, Kalman filters can thus be deployed for systems with states of order up to $n \sim 1000$. Note that Extended Kalman filters, designed for nonlinear systems, are simply Kalman filters, designed based on linearization of the nonlinear system about the expected state trajectory, with the nonlinearity tacked back on in the eleventh hour.

Traditional Kalman and extended Kalman filters were investigated by [12] for atmospheric applications, with nonlinear high-dimensional systems of order $n \gtrsim 10^5$. These applications necessitate the computation of a reduced-rank approximation of the covariance matrix at the heart of the Kalman filter in order to be computationally tractable. Such reduced-rank approximations are known in the controls community as Chandresarkhar's method, and were introduced by [13].

B. Variational methods

Since the mid 1980s, the field of state estimation has seen two revolutionary advancements: variational methods and ensemble Kalman methods. Today, these two classes of methods, in roughly equal proportion worldwide, are used operationally for practical real-time atmospheric forecasting.

The first variational methods introduced were spatial (three-dimensional) variational methods (3DVar; see [14] and [15]), which provide an optimization framework that may be used to fit a large-scale model to a "snapshot" in time of available data. This was soon followed by the development of spatial/temporal (four-dimensional) variational methods (4DVar; see [16] and [17]), in which this optimization framework is extended to account for a time history of observations. This 4DVar framework has the effect of conditioning the resulting estimate on all included data, in a manner consistent with the Kalman Smoother (see [18], [19] and [20]).

Note that 4DVar was developed in parallel, and largely independently, in the controls and weather forecasting communities. In the controls community, the technique is referred to as Moving Horizon Estimation (MHE; see [21]). MHE was developed with low-dimensional ODE systems in mind; implementations of MHE typically search for a small time-varying "state disturbance" or "model error" term in addition to the initial state of the system in order reconcile the measurements with the model over the period of interest. 4DVar, in contrast, was developed with high-fidelity (that is, high-dimensional) discretizations of infinite-dimensional (PDE) systems in mind; in order to maintain numerical tractability, implementations of 4DVar typically do not search for such a time-varying model error term. Both 4DVar and MHE suffer from the fact that they only provide an updated mean trajectory, and not any updated higher-moment statistics. However, during the minimization process, it is possible to build up an approximation to the cost function Hessian. For convex variational problems, this Hessian is directly related to the inverse of the updated covariance matrix. Several of these methods are outlined in [22] and include the randomization method (which uses the statistics of perturbed gradients), the Lanczos method (which exploits the coupling between Lanczos vectors and conjugate gradient directions), and the BFGS method (which explicitly builds up the Hessian during minimization). All three of these methods fail to provide an effective means for *propagating* the updated statistics forward in time, and thus are not typically tractable for variational schemes that cycle over multiple, successive windows.

Another technique that has been introduced to accelerate MHE/4DVar implementations is multiple shooting (see [23]). With this technique, the horizon of interest is split into two or more subintervals. The initial conditions (and, in some cases, the time-varying model error term) for each subinterval are first initialized and optimized independently; these several independent solutions are then adjusted so that the trajectories coincide at the matching points between the subintervals.

C. Ensemble Kalman methods

The more recent development of the Ensemble Kalman Filter (EnKF) (see [24], [25], [26], [27]) has focused much attention on an important refinement of the (sequential) Kalman method in which the estimation statistics are intrinsically represented via the distribution of a cluster or "ensemble" of state estimates in phase space, akin to the particle filters mentioned previously but without separate weights for each ensemble member. As with particle filters, the simultaneous simulation of several perturbed trajectories of the state estimate eliminates the need to propagate the state covariance matrix along with the estimate as required by traditional Kalman and extended Kalman approaches. Instead, this covariance information is approximated based on the spread of the ensemble members (with equal weights) in order to compute Kalman-like updates *to the position of each ensemble member*[1] at the measurement times (for further discussion, see §II-C).

Since its introduction, the EnKF has spawned many variations and modifications that seek to improve both its performance and its numerical tractability. For example, Kalman square-root filters update the analysis only once, in a manner

[1]That is, rather than updating individual weights for each member separately, as done at in particle filters.

different than the traditional perturbed observation method. Some square-root filters introduced include the ensemble adjustment filter of [28], the ensemble transform filter of [29], and the ensemble square-root filter of [30]. Work has also been done (see [31]) to further relax the linear Gaussian assumptions with regards to the interpolation between the observation and the background statistics.

Another essential advancement in the implementation of the EnKF is the idea of covariance localization, as discussed in [32] and [33]. With covariance localization, spurious correlations of the uncertainty covariance over large distances are reduced in an ad hoc fashion in order to improve the overall performance of the estimation algorithm. This adjustment is motivated by the rank-deficiency of the ensemble approximation of the covariance matrix, and facilitates parallel implementation of the resulting algorithm.

The Ensemble Kalman Smoother (EnKS) [34] is the analogous ensemble extension of the standard Kalman Smoother. With the EnKS, updates are performed on past estimates based on future observations in a manner similar to the EnKF. With the EnKS, the smoothed updates are a function of time correlations between two ensemble estimates at the appropriate times. Although each individual update is tractable, it becomes infeasible to update entire trajectories after each new observation; as a result, a fixed-lag or fixed-point EnKS is traditionally used in lieu of a full smoother. Another smoother in this class, the Ensemble Smoother (ES; see [35]), uses ensemble statistics to calculate a variance minimizing estimate, but in practice, for nonlinear systems, performs poorly even when compared to the standard EnKF.

For nonlinear systems, the ensemble Kalman framework is suboptimal due to its reliance on a Kalman-like measurement update formula, which is predicated on a Gaussian distribution of the estimate uncertainty. The more general Particle Filter (PF) method described in §I-A, in contrast, is a full Bayesian approach, with the PDF approximated in a Largrangian fashion akin to the ensemble Kalman framework. The Particle Kalman Filter (PKF) method proposed by [36], which attempts to combine the PF and EnKF approaches in order to inherit the nongaussian uncertainty characterization of the PF method and the numerical tractability of the EnKF method, appears to be promising; this method could potentially benefit from further hybridization with a variational approach, as proposed below.

D. Hybrid methods

The two modern schools of thought in large-scale state estimation for multiscale uncertain systems [namely, space/time variational methods (§I-B) and ensemble Kalman methods (§I-C)] have, for the most part, remained independent, despite their similar theoretical backgrounds. The weather forecasting community has made considerable efforts to compare and contrast both the performance and the theoretical foundation of these two methods (see, e.g., [37], [38], [39], and [40]). While these comparisons are enlightening, it is quite possible that the optimal method for many large-scale state estimation problems cases may well be a hybridization of the two frameworks, as suggested by [40]. We have identified three recent attempts at such hybridization:

1) the 3DVar/EnKF method of [41],
2) the 4DEnKF method of [42], and
3) the E4DVar method of [43].

The 3DVar/EnKF algorithm introduced by [41] utilizes the ensemble framework to propagate the estimate statistics in a nonlinear setting, but does not exploit the temporal smoothing characteristics of the 4DVar algorithm. The 4DEnKF (4D Ensemble Kalman Filter) introduced by [42] provides a means for assimilating past (and non-uniform) observations in a sequential framework, but does not intrinsically smooth the resulting estimate or fully implement the 4DVar framework. The E4DVar (Ensemble 4DVar) method discussed by [43], which is the closest existing method to the hybrid smoother proposed here, runs a 4DVar and EnKF in parallel, sequentially shifting the mean of the ensemble based on the 4DVar result and providing the background term of the 4DVar algorithm based on the EnKF result; however, this method does not attempt a tighter coupling of the ensemble and variational approaches by using an Ensemble Kalman Smoother to initialize better (and, thus, accelerate) the variational iteration.

The three attempts at hybridization discussed above struggle with the inability of traditional variational iterations to update correctly the statistics of the PDF (covariance, etc.). This is crucial for a consistent[2] hybrid method, thus motivating the precise formulation of ensemble variation methods in §III below. The VAE (Variational Assimilation Ensemble) method of [44] runs a half-dozen perturbed decoupled 4DVar or 3DFgat[3] assimilations in parallel to estimate error covariances, and is the closest existing method we have found in the literature to a true ensemble-variation method. However, to the best of our knowledge, the current paper lays out the first complete mathematical foundation for a pure variational method that provides consistent, updated ensemble statistics upon algorithm convergence.

We can now classify the full taxonomy of ensemble-based methods (see Figure 1). Until now, these methods have been split into two distinct families: ensemble variation methods (suggested previously, but described formally for perhaps the first time in §III) and the well-known ensemble Kalman methods. Each family consists of filter variants (En3DVar and EnKF) and smoother variants (En4DVar and EnKS).

The proposed new algorithm, the Hybrid Ensemble Smoother (HEnS), is a consistent[4] and tightly-coupled hybrid of these two types of ensemble smoothers. HEnS uses the EnKS to precondition an appropriately defined En4DVar iteration. Essentially, the EnKS solution is used as a good initial condition for the ensemble variation problem, which in turn improves upon this smoothed estimate in a manner that would have been impossible using either method independently. In

[2]The word "consistent" is used in a precise fashion in this paper to mean an estimation method that reduces to exactly the Kalman filter in the case that the system happens to be linear, the disturbances happen to be Gaussian, and (in the case of an ensemble-based method) a sufficient number of ensemble members is used.

[3]That is, 3D First Guess at the Appropriate Time (3DFgat), an intermediate variational method with complexity somewhere between that of 3DVar and 4DVar [see [45]].

[4]Again, meaning that it reduces to exactly the Kalman filter under the appropriate assumptions.

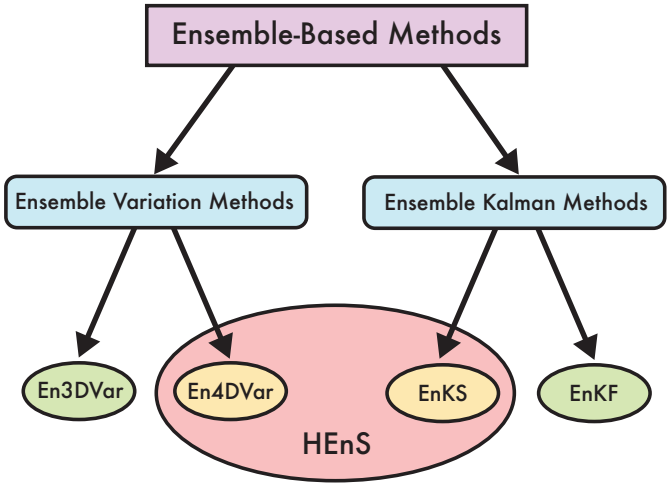


Fig. 1. Ensemble-based methods can be classified into two distinct families. Ensemble Variation Methods (§III) are vector-based methods that iteratively minimize an appropriately-defined cost function to produce either a filtered (En3DVar) or a smoothed (En4DVar) estimate. Ensemble Kalman Methods (§II) use the ensemble statistics to approximate the full (but computationally intractable) matrix-based Kalman updates. The new Hybrid Ensemble Smoother (HEnS) is a consistent hybrid of the smoother variants of these two methods, a described in §IV.

earlier work done by our group (see [46]), the 4DVar/MHE framework was inverted, promoting retrograde time marches (that is, marching the state estimate backward in time and the corresponding adjoint forward in time), which facilitated an adaptive (i.e., multiscale-in-time) receding-horizon optimization framework, dubbed EnVE (Ensemble Variational Estimation). Though the motivation behind this original work was sound, performance suffered, in part as a result of the inability of the variational formulation used to update correctly the higher-moment statistics of the ensemble; the present formulation corrects this significant shortcoming associated with the EnVE formulation.

Section II reviews the general forms of both the ensemble Kalman methods and the traditional variational methods. Section III describes the theoretical foundations for the ensemble variation methods, and identifies their relationship with the well-known KF and KS results. Building upon this, Section IV describes the new hybrid smoother, HEnS, in detail. Finally, Section V contains a comparative example, performed on the low-dimensional chaotic Lorenz system, showing the performance of the various methods in a time-averaged setting. Two follow-up papers are planned which will detail the implementation of the HEnS algorithm on 1D, 2D, and 3D chaotic PDE systems, and introduce a unique adaptive observation algorithm which builds directly upon the hybrid framework discussed here.

II. Theoretical Background

A. Notation

As described above, the Hybrid Ensemble Smoother (HEnS) is a consistent data assimilation method that combines the key ideas of the sequential Ensemble Kalman Smoother (EnKS) and an ensemble variant of the batch (in time) variational method known as 4DVar in the weather forecasting community and as Moving Horizon Estimation (MHE) in the controls community. These methods are thus first reviewed briefly in a fairly standard form. Without loss of generality, the dynamic model used to introduce these methods is a continuous-time nonlinear ODE system with discrete-time measurements:

$$\frac{d\mathbf{x}(t)}{dt} = f(\mathbf{x}(t), \mathbf{w}(t)), \tag{1a}$$

$$\mathbf{y}_k = H\mathbf{x}(t_k) + \mathbf{v}_k, \tag{1b}$$

where the measurement noise $\mathbf{v}_k$ is a zero-mean, white, discrete-time random process with auto-correlation

$$R_{\mathbf{v}}(j;k) = E\{\, \mathbf{v}_{k+j}\, \mathbf{v}_k^T \,\} = R\delta_{j0}, \tag{2}$$

with covariance $R > 0$, and the state disturbance $\mathbf{w}(t)$ is a zero-mean, "nearly"-white[5] continuous-time random process with auto-correlation

$$R_{\mathbf{w}}(\tau;t) = E\{\, \mathbf{w}(t+\tau)\, \mathbf{w}^T(t) \,\} = Q\delta^\sigma(\tau), \tag{3a}$$

$$\text{where} \quad \delta^\sigma(\tau) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\tau^2/(2\sigma^2)}, \tag{3b}$$

with spectral density $Q \geq 0$ and time correlation σ such that $0 < \sigma \ll 1$. Is also assumed that $\mathbf{w}(t)$ and $\mathbf{v}_k$ are uncorrelated.

The noisy measurements $\mathbf{y}_k$ are assumed to be taken at time $t_k = k\Delta t$ for a fixed sample period Δt. For the purpose of analysis, these observations are assumed available for a long history into the past, up to and including the present time of the system being estimated, which is often renormalized to be $t = t_K = T$. It is useful to think of t_K as the time of the most recent available measurement, so, accordingly, this measurement will be denoted $\mathbf{y}_K$ at the beginning of each analysis step. This sets the basis for the indexing notation used in this paper: $k = K$ represents the index of the most recent measurement, $1 \leq k \leq K$ is the set of indices of all available measurements, and $k > K$ indexes observations that are yet to be taken. Continuous-time trajectories, such as $\mathbf{x}(t)$ (the "truth" model), are defined for all time, but are frequently referenced at the observation times only. Hence, the following notation is used:

$$\mathbf{x}(k\Delta t) = \mathbf{x}(t_k) = \mathbf{x}_k. \tag{4}$$

B. Uncertainty Propagation in Chaotic Systems

Estimation, in general, involves the determination of a probability distribution. This probability distribution describes the likelihood that any particular point in phase space matches the truth model. That is, without knowing the actual state of a system, estimation strategies attempt to represent the probability of any given state using only a time history of noisy observations of a subset of the system and an approximate dynamic model of the system of interest. Given this statistical distribution, estimates can be inferred about the "most likely" state of the system, and how much confidence should be placed in that estimate. Unfortunately, in this most general form, the estimation problem is intractable in most systems. However, given certain justifiable assumptions about the nature

[5]The case for infinitesimal σ is sometimes referred to as "continuous-time white noise", but presents certain technical difficulties [47].

of the model and its associated disturbances, simplifications can be applied with regards to how the probability distributions are modeled. Specifically, in linear systems with Gaussian uncertainty of the initial state, Gaussian state disturbances, and Gaussian measurement noise, it can be shown that the probability distribution of the optimal estimate is itself Gaussian [see, e.g., [48]]. Consequently, the entire distribution of the estimate in phase space can be represented exactly by its mean $\bar{\mathbf{x}}$ and its second moment about the mean (that is, its covariance), P, where

$$P = E\big[(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^T\big]. \tag{5}$$

This is the essential piece of theory that leads to the traditional Kalman Filter (KF; see [9] and [10]).

Sequential data assimilation methods provide a method to propagate the mean $\bar{\mathbf{x}}$ and covariance P forward in time, making the appropriate updates to both upon the receipt of each new measurement. Under the assumption of a linear system and white (or, in continuous time, "nearly" white) Gaussian state disturbances and measurement noise, the uncertainty distribution of the optimal estimate is itself Gaussian, and thus is *completely* described by the mean estimate $\bar{\mathbf{x}}$ and the covariance P propagated by the Kalman formulation. It is useful to think of these quantities, at any given time t_k, as being conditioned on a subset of the available measurements. The notation $\bar{\mathbf{x}}_{k|j}$ represents the mean estimate at time t_k given measurements up to and including time t_j. Similarly, $P_{k|j}$ represents the corresponding covariance of this estimate. In particular, $\bar{\mathbf{x}}_{k|k-1}$ and $P_{k|k-1}$ are often called the prediction estimate and prediction covariance, whereas $\bar{\mathbf{x}}_{k|k}$ and $P_{k|k}$ are often called the current estimate and the current covariance. Note that $\bar{\mathbf{x}}_{k|k+K}$, for some $K > 0$, is often called a smoothed estimate, and may be obtained in the sequential setting by a Kalman smoother [see, [19] and [48]].

As mentioned previously, for nonlinear systems with relatively small uncertainties, a common variation on the KF known as the Extended Kalman Filter (EKF) has been developed in which the mean and covariance are propagated, to first-order accuracy, about a linearized trajectory of the full system. Essentially, if a Taylor-series expansion for the nonlinear evolution of the covariance is considered, and all terms higher than quadratic are dropped, what is left is the differential Riccati equation associated with the EKF covariance propagation. Though this approach gives acceptable estimation performance for nonlinear systems when uncertainties are small as compared to the fluctuations of the state itself, EKF estimators often diverge when uncertainties are more substantial, and other techniques are needed.

At its core, the linear thinking associated with the uncertainty propagation in the KF and EKF breaks down in chaotic systems. Chaotic systems are characterized by stable manifolds or "attractors" in n-dimensional phase space. Such attractors are fractional-dimensional subsets (a.k.a. "fractal" subsets) of the entire phase space. Trajectories of chaotic systems are stable with respect to the attractor in the sense that initial conditions off the attractor converge exponentially to the attractor, and trajectories on the attractor remain on the attractor. On the attractor, however, trajectories of chaotic

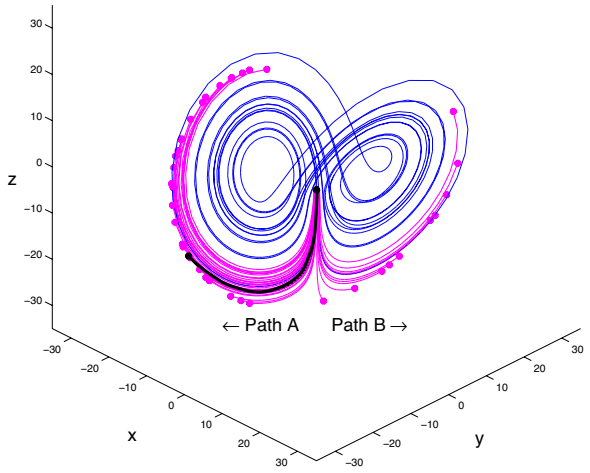


Fig. 2. Non-Gaussian uncertainty propagation in the Lorenz system. The black point in the center shows a typical point located in a sensitive area of this chaotic system's attractor in phase space, representing a current estimate of the state. The thick black line represents the evolution in time of the trajectory from this estimate. If the uncertainty of the estimate is modeled as a very small cloud of points, centered at the original estimate with an initially Gaussian distribution, then the additional magenta lines show the evolution of each of these perturbed points in time. A Gaussian model of the resulting distribution of points is, clearly, completely invalid.

systems are characterized by an *exponential divergence*–along the attractor–of slightly perturbed trajectories. That is, two points infinitesimally close on the attractor at one time will diverge exponentially from one another as the system evolves until they are effectively uncorrelated.

Just as an individual trajectories diverge along the attractor, so does the uncertainty associated with them. This uncertainty diverges in a highly non-Gaussian fashion when such uncertainties are not infinitesimal (see Figure 2). Estimation techniques that attempt to propagate probability distributions under linear, Gaussian assumptions fail to capture the true uncertainty of the estimate in such settings, and thus improved estimation techniques are required. The Ensemble Kalman Filter, in contrast, accounts properly for the nonlinearities of the chaotic system when propagating estimator uncertainty. This idea is a central component of the hybrid ensemble/variational method proposed in the present work, and is thus reviewed next.

C. Ensemble Kalman Filtering

The Ensemble Kalman Filter (EnKF) is a sequential data assimilation method useful for nonlinear multiscale systems with substantial uncertainties. In practice, it has been shown repeatedly to provide significantly improved state estimates in systems for which the traditional EKF breaks down. Unlike in the KF and EKF, the statistics of the estimation error in the EnKF are not propagated via a covariance matrix, but rather are approximated implicitly via the appropriate nonlinear propagation of several perturbed trajectories ("ensemble members") centered about the ensemble mean, as illustrated in Figure 2. The collection of these ensemble members (itself called the "ensemble"), propagates the statistics of the estimation error exactly in the limit of an infinite number of

ensemble members. Realistic approximations arise when the number of ensemble members, N, is (necessarily) finite. Even with a finite ensemble, the propagation of the statistics is still consistent with the nonlinear nature of the model. Conversely, the EKF propagates only the lowest-order components of the second-moment statistics about some assumed trajectory of the system. This difference is a primary strength of the EnKF.

In practice, the ensemble members $\hat{\mathbf{x}}^j$ in the EnKF are initialized with some known statistics about an initial mean estimate $\bar{\mathbf{x}}$. The ensemble members are propagated forward in time using the fully nonlinear model equation (1a), incorporating random forcing $\mathbf{w}^j(t)$ with statistics consistent with those of the actual state disturbances $\mathbf{w}(t)$ [see (3)]:

$$\frac{d\hat{\mathbf{x}}^j(t)}{dt} = f(\hat{\mathbf{x}}^j(t), \mathbf{w}^j(t)). \tag{6}$$

At the time t_k (for integer k), an observation $\mathbf{y}_k$ is taken [see (1b)]. Each ensemble member is updated using this observation, incorporating random forcing $\mathbf{v}_k^j$ with statistics consistent with those of the actual measurement noise, $\mathbf{v}_k$ [see (2)]:

$$\mathbf{d}_k^j = \mathbf{y}_k + \mathbf{v}_k^j. \tag{7}$$

Given this perturbed observation $\mathbf{d}_k^j$, each ensemble member is updated in a manner consistent[6] with the KF and EKF:

$$\hat{\mathbf{x}}^j_{k|k} = \hat{\mathbf{x}}^j_{k|k-1} + P^e_{k|k-1} H^T (H P^e_{k|k-1} H^T + R)^{-1} (\mathbf{d}_k^j - H\hat{\mathbf{x}}^j_{k|k-1}), \tag{8}$$

Unlike the EKF, in which the entire covariance matrix P is propagated using the appropriate Riccati equation, the EnKF estimate covariance P^e is computed "on the fly" using the second moment of the ensembles from the ensemble mean:

$$P^e = \frac{(\delta\widehat{X})\ (\delta\widehat{X})^T}{N-1}, \text{ where } \delta\widehat{X} = \begin{bmatrix} \delta\hat{\mathbf{x}}^1 & \delta\hat{\mathbf{x}}^2 & \cdots & \delta\hat{\mathbf{x}}^N \end{bmatrix},$$
$$\delta\hat{\mathbf{x}}^j = \hat{\mathbf{x}}^j - \bar{\mathbf{x}}, \quad \text{and} \quad \bar{\mathbf{x}} = \frac{1}{N}\sum_j \hat{\mathbf{x}}^j, \tag{9}$$

where N is the number of ensemble members, and the time subscripts have been dropped for notational clarity[7].

Thus, like the KF and EKF, the EnKF is propagated with a forecast step (6) and an update step (8). The ensemble members $\hat{\mathbf{x}}^j(t)$ are propagated forward in time using the system equations [with state disturbances $\mathbf{w}^j(t)$] until a new measurement $\mathbf{y}_k$ is obtained, then each ensemble member $\hat{\mathbf{x}}^j(t_k) = \hat{\mathbf{x}}_k^j$ is updated to include this new information [with measurement noise $\mathbf{v}_k^j$]. The covariance matrix is not propagated explicitly, as its evolution is implicitly represented by the evolution of the ensemble itself.

It is convenient to think of the various estimates during such a data assimilation procedure in terms of the set of measurements that have been included to obtain that estimate. Just as it is possible to propagate the ensemble members forward in time accounting for new measurements, ensemble members can also be propagated *backward* in time, either retaining the effect of each measurement or subtracting this information back off. In the case of a linear system, the former approach is equivalent to the Kalman smoother, while the later approach simply retraces the forward march of the Kalman filter backward in time. In order to make this distinction clear, the notation $\widehat{X}_{j|k}$ will represent the estimate ensemble at time t_j given measurements up to and including time t_k. Similarly, $\bar{\mathbf{x}}_{j|k}$ will represent the corresponding ensemble mean; that is, the average of the ensemble and the "highest-likelihood" estimate of the system.

While the EnKF significantly outperforms the more traditional EKF for chaotic systems, further approximations need to be made for multiscale systems such as atmospheric models. When assimilating data for 3D PDEs, the discretized state dimension n is many orders of magnitude larger than the number of ensemble members N that is computationally feasible (i.e., $N \ll n$). The consequences of this are twofold. First, the ensemble covariance matrix P^e is guaranteed to be singular, which can lead to difficulty when trying to solve linear systems constructed with this matrix. Second, this singularity combined with an insufficient statistical sample size produces directions in phase space in which no information is gained through the assimilation. This leads to spurious correlations in the covariance that would cause improper updates across the domain of the system. This problem can be significantly diminished via the ad hoc method of "covariance localization" mentioned previously, which artificially suppresses these spurious correlations using a distance-dependent damping function.

D. Ensemble Kalman Smoother (EnKS)

The Ensemble Kalman Smoother (EnKS) is built upon the theoretical foundations of the EnKF. The key difference lies in its ability to update past estimates based on future observations. Thus, we end up with smoothed estimates $\hat{\mathbf{x}}^j_{p|k}$, where p is not necessarily less than k. Given a new observation $\mathbf{y}_k$ at time t_k and forecasted ensemble $\hat{\mathbf{x}}^j_{k|k-1}$ at that time, the smoothed estimate $\hat{\mathbf{x}}^j_{p|k}$ is given by the following update equation:

$$\hat{\mathbf{x}}^j_{p|k} = \hat{\mathbf{x}}^j_{p|k-1} + S^e_{k-1} H^T (H P^e_{k|k-1} H^T + R)^{-1} (\mathbf{d}_k^j - H\hat{\mathbf{x}}^j_{k|k-1}), \tag{10a}$$

where S^e_{k-1} is the time covariance matrix between the estimate at the observation time t_k and the estimate at the smoothing time t_p, which is given by

$$S^e_{k-1} = \frac{(\delta\widehat{X}_{p|k-1})\ (\delta\widehat{X}_{k|k-1})^T}{N-1}, \tag{10b}$$

with the definitions for $\delta\widehat{X}$ given in (9). Note that, when $t_p = t_k$, the time covariance matrix S^e_{k-1} reduces to the standard covariance matrix $P^e_{k|k-1}$, and thus the EnKS update (10a) reduces appropriately to the standard EnKF update (8). This highlights an important property of the EnKS: even for highly chaotic, nonlinear systems, the EnKS provides the same estimate at the most recent measurement as the EnKF (in the

[6]Note that some authors (see, e.g., [27]) prefer to replace R in (8) with R^e, where

$$R^e = \frac{(V_k)\,(V_k)^T}{N-1} \quad \text{and} \quad V_k = \begin{bmatrix} \mathbf{v}_k^1 & \mathbf{v}_k^2 & \cdots & \mathbf{v}_k^N \end{bmatrix}.$$

Our current research has not revealed any clear advantage for using this more computationally expensive form.

[7]Note also that the factor $N-1$ (instead of N) is used in (9) to obtain an unbiased estimate of the covariance matrix [see [47]].

limit of an infinite number of ensemble members). This result is expected in the linear setting, but is a major shortcoming of the EnKS when applied to the typical nonlinear systems. This shortcoming is rectified by the hybrid method presented in Section IV.

E. Variational Methods

For high-dimensional systems in which matrix-based methods are computationally infeasible, vector-based variational methods are preferred for data assimilation. 3DVar is a vector-based equivalent to the KF. In both 3DVar and KF, the cost function being minimized is a (quadratic) weighted combination of the uncertainty in the background term and the uncertainty in the new measurement. If the system is linear, the optimal update to the state estimate can be found analytically, though this solution requires matrix-based arithmetic (specifically, the propagation of a Riccati equation), and is the origin of the optimal update gain matrix for the KF. When this matrix is too large for direct computation, a local gradient can instead be found using vector-based arithmetic only; 3DVar uses this local gradient information to determine the optimal update iteratively.

Similarly, 4DVar is the vector-based equivalent to the Kalman Smoother. In 4DVar, a finite time window (or "batch process") of a history of measurements is analyzed together to improve the estimate of the system at one edge of this window (and, thus, the corresponding trajectory of the estimate over the entire window). Unlike sequential methods, a smoother uses all available data over this finite time window to optimize the estimates of the system. This has the consequence of refining past estimates of the system based on future measurements, whereas with sequential methods any given estimate is only conditioned on previous observations.

For analysis, let the variational window be defined on $t \in (\,0, T\,]$. Additionally, let there be K measurements in this interval, with measurement indices given by the set

$$M = \{\, k \mid t_k \in (\,0, T\,]\,\} \quad \Rightarrow \quad M = \{\,1\,, 2\,,\ \cdots\ ,\, K\,\}. \tag{11}$$

Without loss of generality, it will be assumed that there are measurements at the right edge of the window (at $t_{_K} = T$), but not at the left (at $t_0 = 0$). Then, the cost function $J(\mathbf{u})$ that 4DVar minimizes (with respect to $\mathbf{u}$) is defined as follows:

$$\begin{aligned} J(\mathbf{u}) =& \frac{1}{2}\,(\,\mathbf{u} - \bar{\mathbf{x}}_{_{0|0}})^T\, P_{_{0|0}}^{-1}\,(\,\mathbf{u} - \bar{\mathbf{x}}_{_{0|0}}) + \\ & \frac{1}{2}\sum_{k=1}^{K} \big(\,\mathbf{y}_k - H\,\tilde{\mathbf{x}}_k\,\big)^T\, R^{-1}\,\big(\,\mathbf{y}_k - H\,\tilde{\mathbf{x}}_k\,\big), \end{aligned} \tag{12}$$

where the optimization variable $\mathbf{u}$ is the initial condition on the refined state estimate $\tilde{\mathbf{x}}$ on the interval $t \in (\,0, T\,]$; that is,

$$\frac{d\tilde{\mathbf{x}}(t)}{dt} = f(\tilde{\mathbf{x}}(t), 0), \tag{13a}$$

$$\tilde{\mathbf{x}}_{_0} = \mathbf{u}. \tag{13b}$$

The first term in the cost function (12), known as the "background" term, summarizes the fit of $\mathbf{u}$ with the current probability distribution before the optimization (i.e., the effect of all past measurement updates). Like with the KF, $\bar{\mathbf{x}}_{_{0|0}}$ is the estimate at time $t_{_0}$ not including any of the new measurements in the window, and the covariance $P_{_{0|0}}$ quantifies the second moment of the uncertainty in that estimate. Assuming an apriori Gaussian probability distribution of this uncertainty, the background mean and covariance exactly describe this distribution. The second term in the cost function (12) summarizes the misfit between the estimated system trajectory and the observations within the variational window. Thus, the solution $\mathbf{u}$ to this optimization problem is the estimate that best "fits" the observations over the variational window while also accounting for the existing information from observations prior to the variational window.

In practice, a 4DVar iteration is usually initialized with the background mean, $\mathbf{u} = \bar{\mathbf{x}}_{_{0|0}}$. Given this initial guess for $\mathbf{u}$, the trajectory $\tilde{\mathbf{x}}(t)$ may be found using the full nonlinear equations for the system (13). To find the gradient of the cost function (12), consider a small perturbation of the optimization variable, $\mathbf{u} \leftarrow \mathbf{u} + \mathbf{u}'$, and the resulting perturbed trajectory, $\tilde{\mathbf{x}}(t) \leftarrow \tilde{\mathbf{x}}(t) + \tilde{\mathbf{x}}'(t)$, and perturbed cost function, $J(\mathbf{u}) \leftarrow J(\mathbf{u}) + J'(\mathbf{u}')$. The local gradient of (12), $\nabla J(\mathbf{u})$, is defined here as the sensitivity of the perturbed cost function $J'(\mathbf{u}')$ to the perturbed optimization variable $\mathbf{u}'$:

$$J'(\mathbf{u}') = \big[\nabla J(\mathbf{u})\,\big]^T\,\mathbf{u}'. \tag{14}$$

The derivation included in the Appendix illustrates how to write $J'(\mathbf{u}')$ in this simple form, leveraging the definition of an appropriate adjoint field $\mathbf{r}(t)$ on $t \in (\,0, T\,]$, providing the following gradient:

$$\nabla J(\mathbf{u}) = P_{_{0|0}}^{-1}\,(\,\mathbf{u} - \bar{\mathbf{x}}_{_{0|0}}) - \mathbf{r}_{_0}. \tag{15}$$

The resulting gradient can then be used iteratively to update the current estimate via a suitable minimization algorithm (steepest descent, conjugate gradient, limited-memory BFGS, etc.).

Being vector based makes 4DVar well suited for multiscale problems, and as a result is currently used extensively by the weather forecasting community. However, it has several key disadvantages. Most significantly, upon convergence, the algorithm provides an updated mean estimate $\bar{\mathbf{x}}_{_{0|K}}$, but provides no clear formula for computing the updated estimate uncertainty covariance or its inverse, $P_{_{0|K}}^{-1}$. That is, the statistical distribution of the estimate probability is not contained in the output of a traditional 4DVar algorithm. It can be shown that, upon full convergence for a linear system, the resulting analysis covariance $P_{_{0|K}}$ is simply the Hessian of the original cost function (12) [see, e.g., [49]]. However, this is merely an analytical curiosity; computing the analysis covariance in this fashion requires as much matrix algebra as would be required to propagate a sequential filter through the entire variational window, defeating the purpose of the vector-based method.

Additionally, as posed above, the width of the variational window is fixed in the traditional 4DVar formulation. Thus, the cost function and associated n-dimensional minimization surface are also constant throughout the iterations. For nonlinear systems, especially chaotic systems, this makes traditional 4DVar extremely sensitive to initial conditions. Because of

the chaotic nature of these systems, the optimization surface, especially if T is large, is highly irregular and nonconvex (that is, fraught with local minima). The gradient-based algorithms associated with 4DVar are only guaranteed to converge to local minima. Thus, if the initial background estimate is located in the region of attraction of one of these local minima, the solution of the 4DVar algorithm will tend to converge to a suboptimal estimate.

III. Ensemble Variation Methods

As pointed out in Section II-E, one of the major weaknesses of the standard variational assimilation schemes is the inability of these methods to update the higher moment estimate statistics. Given both a background mean and covariance, 3DVar and 4DVar simply return an updated mean; the corresponding updated covariance has thus far only been found via computationally involved Hessian analysis [49] or schemes coupled with a Kalman-like covariance propagation [43]. Here, we lay out the mathematical foundations for a consistent class of variational methods that, much like the Ensemble Kalman methods, use a finite cloud of points to represent implicitly both the background and analysis estimate statistics. Unlike the Ensemble Kalman methods, this new class of Ensemble Variation methods uses an *ensemble of variational problems* to solve iteratively the complete data assimilation via the minimization of an appropriately defined cost function. It is shown that, under the standard assumptions of linear dynamics and Gaussian noise and disturbances, these Ensemble Variation methods reduce to the well-known optimal results of the standard Kalman Filter and Kalman Smoother.

A. Ensemble 3D Variational Assimilation (En3DVar)

Given a measurement $\mathbf{y}_k$ at time t_k, we will represent our estimate statistics with a finite ensemble of N members such that the sample mean and sample covariance are consistent (in the limit as $N \to \infty$) with the (assumed) known background mean and covariance. Thus, we have a collection of ensemble members $\hat{\mathbf{x}}^j_{k|k-1}$ conditioned on all prior observations $\{\, \mathbf{y}_p \mid p < k \,\}$ that build a sample covariance given by $P^e_{k|k-1}$. With this, we can define an En3DVar component cost function for each ensemble member as:

$$J_j(\mathbf{u}^j) = \frac{1}{2}\,(\mathbf{u}^j - \hat{\mathbf{x}}^j_{k|k-1})^T\,(P^e_{k|k-1})^{-1}\,(\mathbf{u}^j - \hat{\mathbf{x}}^j_{k|k-1}) + \frac{1}{2}\,(\mathbf{d}^j_k - H\mathbf{u}^j)^T\,R^{-1}\,(\mathbf{d}^j_k - H\mathbf{u}^j), \tag{16}$$

where the control variable $\mathbf{u}^j$ for each ensemble member is (at the minimum) the updated estimate $\hat{\mathbf{x}}^j_{k|k}$, now conditioned on the new measurement $\mathbf{y}_k$. As with the EnKF and EnKS, each ensemble member is assimilated with additional noise added onto the (already noisy) measurement, i.e.,

$$\mathbf{d}^j_k = \mathbf{y}_k + \mathbf{v}^j_k. \tag{17}$$

The total cost function J is given as the sum of the component cost functions J_j for each ensemble member j. Because the component cost functions are only coupled through the specified (and fixed) background ensemble members $\hat{\mathbf{x}}^j_{k|k}$ and the covariance matrix which they approximate, $P^e_{k|k-1}$, each J_j can be minimized *independently*, creating an optimization problem that is trivial to parallelize on modern high performance computing hardware. Similar to traditional 3DVar, each component cost function is minimized by finding the local gradient and then using a suitable descent algorithm; again, these component-wise minimizations are completely decoupled from one ensemble member to the next.

In summary, En3DVar is performed at a given time t_k by assimilating an ensemble of 3DVar problems, one for each ensemble member. Each individual component 3DVar problem is uniquely characterized by its own perturbed background state $\hat{\mathbf{x}}^j_{k|k-1}$ and its own perturbed measurement $\mathbf{d}^j_k$. The component cost functions are coupled through the background ensemble covariance matrix $P^e_{k|k-1}$, the measurement noise covariance matrix R, and the original, unperturbed (but still noisy) measurement $\mathbf{y}_k$. It is shown in the following section that the unique solution to this problem (in the limit as $N \to \infty$) is a new ensemble with corresponding sample statistics (mean and covariance) that are consistent with the well-known optimal Kalman results.

Theorem 1 (Equivalence of En3DVar to the Kalman Filter): In the limit of an infinite number of ensemble members (i.e., $N \to \infty$), the En3DVar problem defined above converges to the equivalent Kalman filter solution.

Proof: Because each component cost function is minimized independently, we will examine the unique solution of just one for the purpose of this proof. Note that (16) is convex in $\mathbf{u}^j$. The gradient of the j^{th} component cost function with respect to the initial state $\mathbf{u}^j$ is given by

$$\nabla J_j = (P^e_{k|k-1})^{-1}\,(\mathbf{u}^j - \hat{\mathbf{x}}^j_{k|k-1}) - H^T\,R^{-1}\,(\mathbf{d}^j_k - H\mathbf{u}^j). \tag{18}$$

Typically, the cost function is minimized iteratively via a gradient descent method, but for the purpose of analysis here, we can find the minimum directly by setting the $\nabla J_j = 0$ and solving for the updated estimate $\mathbf{u}^j = \hat{\mathbf{x}}^j_{k|k}$ at the minimum:

$$0 = (P^e_{k|k-1})^{-1}\,(\hat{\mathbf{x}}^j_{k|k} - \hat{\mathbf{x}}^j_{k|k-1}) - H^T\,R^{-1}\,(\mathbf{d}^j_k - H\hat{\mathbf{x}}^j_{k|k}) \tag{19a}$$

$$0 = (P^e_{k|k-1})^{-1}\,(\hat{\mathbf{x}}^j_{k|k} - \hat{\mathbf{x}}^j_{k|k-1}) + H^T\,R^{-1}\,H\,(\hat{\mathbf{x}}^j_{k|k} - \hat{\mathbf{x}}^j_{k|k-1}) - H^T\,R^{-1}\,(\mathbf{d}^j_k - H\hat{\mathbf{x}}^j_{k|k-1}) \tag{19b}$$

$$(\hat{\mathbf{x}}^j_{k|k} - \hat{\mathbf{x}}^j_{k|k-1}) = \big[\,(P^e_{k|k-1})^{-1} + H^T\,R^{-1}\,H\,\big]^{-1} H^T\,R^{-1}\,(\mathbf{d}^j_k - H\hat{\mathbf{x}}^j_{k|k-1}) \tag{19c}$$

Assuming that all inverses indicated exist, the identity

$$\big[\,(P^e_{k|k-1})^{-1} + H^T\,R^{-1}\,H\,\big]^{-1} H^T\,R^{-1} = P^e_{k|k-1} H^T\,\big[\,H\,P^e_{k|k-1}\,H^T + R\,\big]^{-1} \tag{19d}$$

can be substituted into (19c) to get the form

$$\hat{\mathbf{x}}^j_{k|k} = \hat{\mathbf{x}}^j_{k|k-1} + P^e_{k|k-1} H^T \left[H\, P^e_{k|k-1}\, H^T + R \right]^{-1} (\mathbf{d}^j_k - H\hat{\mathbf{x}}^j_{k|k-1}), \quad (20a)$$

$$K \equiv P^e_{k|k-1} H^T \left[H\, P^e_{k|k-1}\, H^T + R \right]^{-1}, \quad (20b)$$

$$\hat{\mathbf{x}}^j_{k|k} = \hat{\mathbf{x}}^j_{k|k-1} + K\,(\mathbf{d}^j_k - H\hat{\mathbf{x}}^j_{k|k-1}). \quad (20c)$$

Recall that (20c) is the unique solution for the j^{th} ensemble member. Thus, we can think of the ensemble of solutions $\hat{\mathbf{x}}^j_{k|k}$ as a random variable that is itself conditioned on two other random variables, $\hat{\mathbf{x}}^j_{k|k-1}$ and $\mathbf{d}^j_k$. Note that the gain matrix K is identical to that of the Kalman Filter. To see the rest of the equivalence with the Kalman Filter, we take the sample mean of the result.

$$\begin{aligned}\bar{\mathbf{x}}_{k|k} &= \frac{1}{N}\sum_{j=1}^{N} \hat{\mathbf{x}}^j_{k|k} \\ &= \frac{1}{N}\sum_{j=1}^{N} \hat{\mathbf{x}}^j_{k|k-1} + K\left(\frac{1}{N}\sum_{j=1}^{N}\mathbf{d}^j_k - H\,\frac{1}{N}\sum_{j=1}^{N}\hat{\mathbf{x}}^j_{k|k-1}\right) \\ &= \bar{\mathbf{x}}_{k|k-1} + K\,(\,\mathbf{y}_k - H\,\bar{\mathbf{x}}_{k|k-1}) \end{aligned} \quad (21)$$

Although we obtain the Kalman update (21) for the estimate mean by using En3DVar, it is important to note that the traditional 3DVar algorithm (involving only an iterative update of the mean) would also have provided us with this result. The real strength of En3Dvar lies in its ability to also implicitly update the estimate covariance, something that was not possible with traditional 3DVar. To see this equivalence, we take the sample covariance of the updated ensemble $\hat{\mathbf{x}}^j_{k|k}$,

$$P^e_{k|k} = \frac{1}{N-1}\sum_{j=1}^{N} (\,\hat{\mathbf{x}}^j_{k|k} - \bar{\mathbf{x}}_{k|k})\,(\,\hat{\mathbf{x}}^j_{k|k} - \bar{\mathbf{x}}_{k|k})^T. \quad (22a)$$

Substituting in the definitions of $\hat{\mathbf{x}}^j_{k|k}$ from (20c) and $\bar{\mathbf{x}}_{k|k}$ from (21) and simplifying, we get

$$\begin{aligned} P^e_{k|k} &= (I - KH)\,P^e_{k|k-1}(I - KH)^T + K\,R^e\,K^T + \Phi + \Phi^T, \\ \Phi &= \frac{1}{N-1}\sum_{j=1}^{N}\Big\{(\,I - K\,H\,)\,(\,\hat{\mathbf{x}}^j_{k|k-1} - \bar{\mathbf{x}}_{k|k-1}) \\ &\qquad \cdot (\,\mathbf{d}^j_k - \mathbf{y}_k)^T\,K^T\Big\}. \end{aligned} \quad (22b)$$

The final terms $\Phi + \Phi^T$ in (22b) arise from spurious correlations between the background error and the measurement noise. In a similar manner to the EnKF, as the number of ensembles increase, these terms disappear, leaving the expected Kalman Filter covariance update equation, i.e.,

$$\lim_{N\to\infty} \Phi = 0, \quad (23a)$$

$$P^e_{k|k} = (\,I - K\,H\,)\,P^e_{k|k-1}(\,I - K\,H\,)^T + K\,R^e\,K^T. \quad (23b)$$

Thus, we have shown that, by iteratively assimilating an ensemble of 3DVar problems with both perturbed background states and perturbed measurements, we are able to compute both the analysis mean *and* covariance. This algorithm is both tractable for high dimensional systems (in the sense that it is vector-based) and very easily parallelized (in the sense that each individual problem can be solved independently). ■

B. Ensemble 4D Variational Assimilation (En4DVar)

Given a time history of measurements $\{\,\mathbf{y}_k \mid t_k \in (\,0,\,T\,]\,\}$, as with the En3DVar case, we will represent our estimate statistics with a finite ensemble of N members such that the sample mean and sample covariance are consistent (in the limit as $N \to \infty$) with the (assumed) known background mean and covariance at the left edge of the time window t_0. We can then define an analogous En4DVar cost function over the window, for the j^{th} ensemble member, that balances the misfit between a set of perturbed observations and the deviation from a perturbed background initial condition as follows:

$$\begin{aligned} J_j(\mathbf{u}^j) &= \frac{1}{2}\,(\mathbf{u}^j - \hat{\mathbf{x}}^j_{0|0})^T\,(P^e_{0|0})^{-1}\,(\mathbf{u}^j - \hat{\mathbf{x}}^j_{0|0}) \\ &\quad + \frac{1}{2}\sum_{k=1}^{K}(\mathbf{d}^j_k - H\,\tilde{\mathbf{x}}^j_k)^T\,R^{-1}\,(\mathbf{d}^j_k - H\,\tilde{\mathbf{x}}^j_k). \end{aligned} \quad (24)$$

Muck like traditional 4DVar, each ensemble member is constrained over the window by the model, and the control variable $\mathbf{u}^j$ serves as the initial condition for its trajectory.

$$\frac{d\tilde{\mathbf{x}}^j(t)}{dt} = f(\tilde{\mathbf{x}}^j(t), 0), \quad (25a)$$

$$\tilde{\mathbf{x}}^j_0 = \mathbf{u}^j. \quad (25b)$$

Each initial ensemble member $\hat{\mathbf{x}}^j_{0|0}$ acts as its own perturbed background, and each ensemble member is assimilated with its own set of perturbed measurements $\{\,\mathbf{d}^j_k = \mathbf{y}_k + \mathbf{v}^j_k \mid t_k \in (\,0,\,T\,]\,\}$.

The total cost function J is given as the sum of the component cost functions for each ensemble member. Because the component cost functions are only coupled through the specified (and fixed) background ensemble members $\hat{\mathbf{x}}^j_{0|0}$ and the covariance matrix which they approximate, $P^e_{0|0}$, each J_j can be minimized *independently*, creating an embarrassingly parallel optimization problem. Similar to traditional 4DVar, each cost function is minimized by finding the local gradient and using a suitable descent algorithm. Finding the gradient of (24) requires the use of an appropriately-defined adjoint field. The derivation parallels that of standard 4DVar (as illustrated in the Appendix), and gives the j^{th} gradient as:

$$\nabla J_j(\mathbf{u}^j) = (P^e_{0|0})^{-1}\,(\,\mathbf{u}^j - \hat{\mathbf{x}}^j_{0|0}) - \mathbf{r}^j_0, \quad (26)$$

where $\mathbf{r}^j_0$ is the initial condition of the j^{th} adjoint field found via a background march from t_K to t_0 of the adjoint equations. Thus each iteration of En4DVar requires a forward march of the ensemble through the optimization window followed by a backward march of an *ensemble of adjoints* to find each component gradient. The decoupled nature of these marches is what facilitates the efficient parallel global solution.

Due to the fact that En4DVar accounts for all observations within the assimilation window, it is, by nature, a smoother. Upon completion of the minimization, we are provided with a new ensemble of points $\hat{\mathbf{x}}^j_{0|K}$, conditioned on these measurements. From this ensemble, statistical measures such as the sample mean and covariance can be extracted.

Theorem 2 (Equivalence of En4DVar to the Kalman Smoother): In the limit of an infinite number of ensemble members (i.e. $N \to \infty$) and under the assumptions of linear dynamics

and Gaussian noise and disturbances, the En4DVar problem defined above converges to the equivalent Kalman smoother solution.

Proof: The proof is straightforward and follows directly from that of Theorem 1, however, due to the addition of the time dynamics, it tends to become notationally cumbersome. In the interest of brevity, we have elected to omit it here. ■

IV. Hybrid Ensemble Smoother (HEnS)

As was initially illustrated in Figure 1, we have identified two families of ensemble-based assimilation methods: the ensemble variational methods (consisting of En3DVar and En4DVar) and the ensemble Kalman methods (consisting of the EnKF and the EnKS). In theory, both families address the same problem. Further, as we have shown in the simplified case with linear dynamics and Gaussian uncertainties, they converge to the same solution. However, when the system is highly nonlinear, all bets on optimality of the solutions are off, and we do not necessarily expect each method to provide identical solutions.

In the case of nonlinear systems, one might thus wonder which method typically provides the best answer. The best answer, however, may in fact not come from one individual method, but rather from a consistent hybrid of both methods. This is the idea behind the development of the Hybrid Ensemble Smoother (HEnS), which is a consistent hybrid between the two smoothers, En4DVar and EnKS.

A key motivation for HEnS is the iterative nature of the En4DVar method. Once the component cost functions (24) are defined appropriately using the known background ensemble, any initial condition $\mathbf{u}^j$ can be used to begin the gradient-based minimization. Typically, the best guess we have at the start of an iteration is the background itself (i.e., $\mathbf{u}^j = \hat{\mathbf{x}}^j_{0|0}$), because no other information is known. However, if we were to first run the EnKS through the entire window $(0, T]$, we would develop an intermediate smoothed estimate $\hat{\mathbf{x}}^j_{0|K}$. That is, the output of the EnKS at the left edge of the window is the best estimate at that time, given all measurements in the optimization window *as determined by the EnKS framework*. Again, under the appropriate assumptions, this smoothed estimate would be optimal, but, due to the nonlinear nature of the system and the necessarily finite ensemble size, the EnKS typically finds a suboptimal solution to the smoothing problem. Consequently, this intermediate smoothed estimate can then be used as the initial condition for the specified En4DVar minimization problem in lieu of the background state. If there is any more information to be extracted from the observations, this iterative minimization will attempt to do just that.

Theorem 3 (Consistency of HEnS): In the limit of an infinite number of ensemble members (i.e. $N \to \infty$), and under the assumptions of linear dynamics and Gaussian noise and disturbances, the HEnS formulation described above converges to both the Kalman smoother solution and the equivalent En4DVar solution.

Proof: Under the above assumptions, the En4DVar cost function is convex, and thus contains only one minimum. Provided the cost function is defined appropriately, the En4DVar iteration will converge to this global minimum, regardless of initial condition–even if the output from the EnKS is used to initialize the minimization, as done with HEnS. Therefore, HEnS will converge to the same solution as En4DVar under the assumptions stated. The proof of the equivalence to the Kalman smoother then follows immediately from Theorem 2. ■

An important consequence of Theorem 3 is that the HEnS framework will do no worse than the EnKS solution alone.

In summary, HEnS is a consistent hybrid of both EnKS and En4DVar. Essentially, HEnS uses a (typically) suboptimal smoothed estimate from the EnKS to initialize an En4DVar minimization. Because the output from the EnKS is much closer to the expected minimum than the original background estimate, the En4DVar iteration is less likely to converge to spurious local minima, far from the optimal estimate, and thus produces more a more accurate estimate than either smoother would by itself. HEnS can be implemented in three straightforward steps:

1) Given a set of measurements $\{\mathbf{y}_1, \cdots, \mathbf{y}_K\}$ on the window $(0, T]$ and a background ensemble $\hat{\mathbf{x}}^j_{0|0}$ at the left edge of the window, define the appropriate En4DVar component cost functions.
2) March a fixed-point EnKS through the window, smoothing the estimate at the left edge of the window to produce an intermediate smoothed estimate $\hat{\mathbf{x}}^j_{0|K}$ conditioned on all observations in the window.
3) Use the intermediate smoothed estimate output from the EnKS as the initial condition $\mathbf{u}^j$ for an En4DVar minimization over the same window, using the previously-defined component cost functions. This will potentially provide a better smoothed estimate over the entire window for the full nonlinear system.

As with any assimilation strategy, the output of HEnS from a previous window can be used as the background estimate for a subsequent window, cycling the algorithm.

A. Advantages in forecasting

In data assimilation applications that involve forecasting, the most important estimate is always the most recent one. This is, of course, the estimate that is used as an initial condition for any open-loop forecast (into the future).

In the linear setting, the most recent filtered estimate is identical to the most recent smoothed estimate because they have both been conditioned on the same set of measurements (which are all necessarily in the past). It is for this reason alone that smoothers have been largely ignored by the operational weather forecasting community. Up until now, their computational expense has not been justified by providing an improved estimate upon which a forecast could be made.

Although this conclusion is certainly true in the linear world, much information is to be gained, even at the most recent time, by revisiting old measurements in light of new data. Even in the EnKF setting, the suboptimal updates made at any given time are a function of the ensemble member trajectories used. If those trajectories were improved (say, through the use of a smoother), then it might be possible to perform more accurate

updates, which in turn would increase the accuracy of the estimate at a future time.

Unfortunately, the formulation of the EnKS does not leverage this idea, and (in the limit of an infinite ensemble size) returns exactly the same estimate at the right edge of the window. HEnS, however, is not so constrained. By improving the smoothed estimate initial condition at the left edge of the window, HEnS also improves the estimate throughout the window, and specifically reduces the error in the most recent estimate, providing a more accurate long-term forecast.

B. Hybrid Ensemble Filter (HEnF)

It is worth noting that, in a manner analogous to the hybridization of the two smoothers (En4DVar and the EnKS), a hybrid filter can be defined by combining En3DVar and the EnKF. The resulting algorithm, appropriately dubbed HEnF, would use the EnKF to precondition an En3DVar iteration. That is, at a given measurement time, the EnKF update would be used as an initial condition for the subsequent En3DVar minimization. However, because neither filter update necessarily takes into account the dynamics of the system, it is not apparent to the authors that such additional computation would provide a better solution. As a result, we have neglected to highlight the HEnF in this discussion.

V. Representative Example

The two primary new ideas described in this paper, En4DVar and HEnS, are now compared, via computational experiments, to both EnKF and EnKS. We have already shown analytically that, in the linear setting, all four methods provide consistent solutions; however, in a nonlinear setting with significant uncertainties, these different approaches provide substantially different results.

The Lorenz equation (see [50]) is used here as a simple model of a nonlinear system with self-sustained chaotic unsteadiness[8] in order to perform this comparison. The Lorenz equation is a system of three coupled, nonlinear ordinary differential equations given by:

$$\frac{d\mathbf{x}(t)}{dt} = \begin{pmatrix} \sigma\left(x_2 - x_1\right) \\ -x_2 - x_1 x_3 \\ -\beta x_3 + x_1 x_2 - \beta\phi \end{pmatrix}, \tag{27}$$

where σ, β, and ϕ are tunable parameters. Solutions of these equations approach a well-defined manifold or *attractor* of dimension slightly higher than two. Perturbed trajectories converge exponentially back onto the attractor, while adjacent trajectories diverge exponentially within the attractor, creating the familiar chaotic motion in the Lorenz system. Note that, for convenience, the system equations of (27) are transformed slightly from the traditional form such that the attractor is approximately centered at the origin.

In this comparison, we quantify the time-averaged statistics of the four assimilation methods under consideration. That is, we run a very large number of trials and calculate both the average error of the estimate as well as the average energy of the estimation error (a.k.a. the variance) over all of these trials. A typical run simulates a truth trajectory over a fixed window, taking noisy measurements at set intervals; then, each method (EnKF, EnKS, En4DVar, and HEnS) is used on the resulting data set, initialized with its own independent background. Appealing to the ergodicic nature of the Lorenz system, the output at the right edge of the window for each method (and the truth model) is then used as the input for the next run on the subsequent time window. Consequently, a series of assimilation windows are evaluated on various intervals all over the attractor. The statistics of this process are then used to calculate the expected performance characteristics on this nonlinear chaotic system.

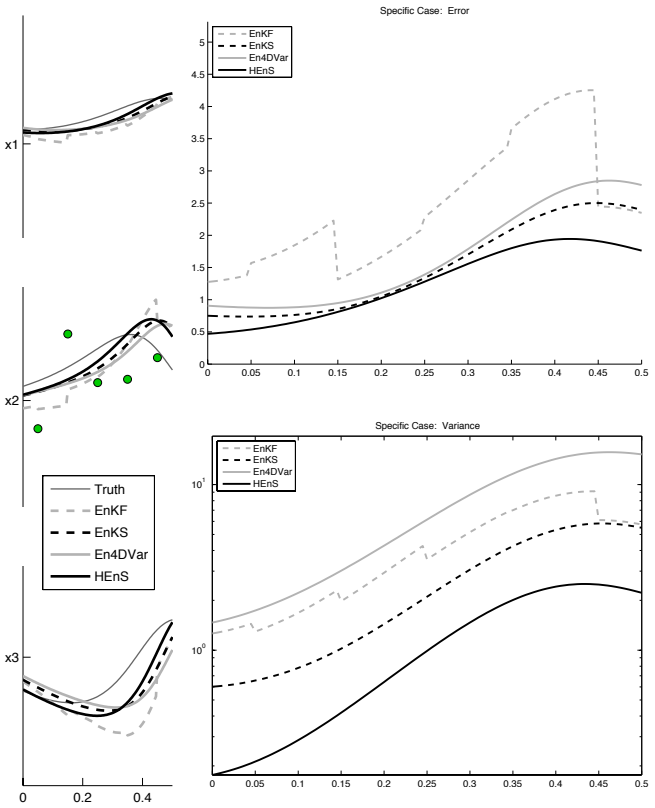


Fig. 3. A typical run for the Lorenz test problem. On the left, each assimilation method is compared to the truth trajectory and the very noisy measurements taken of x_2. From this, the estimate error (top-right) and error energy (bottom-right) are calculated. Note that the EnKF and EnKS are not necessarily equivalent at the right edge of the window due to the finite ensemble size.

For the results shown, the model parameters used are $\sigma = 4$, $\beta = 1$, and $\phi = 48$. Only the second state x_2 is measured, and the measurement noise variance used $R = 5$ (which, as seen in Figure 3, is quite substantial). The assimilation window has width $T = 0.5$, and five observations are taken (at intervals of $\Delta t = 0.1$) centered in this window. The starting conditions for the truth and estimate ensemble backgrounds are not significant, as the converged statistics are not a function of the starting point used in the simulation. The nonlinear model is assumed to be perfect (that is, $Q = 0$), and thus any ensemble forecasting would be done with a simple evolution of the unforced equations from the most recent state estimate. All

[8]That is, the system considered maintains its nonperiodic, finite, bounded unsteady motion with no externally-applied stochastic forcing.

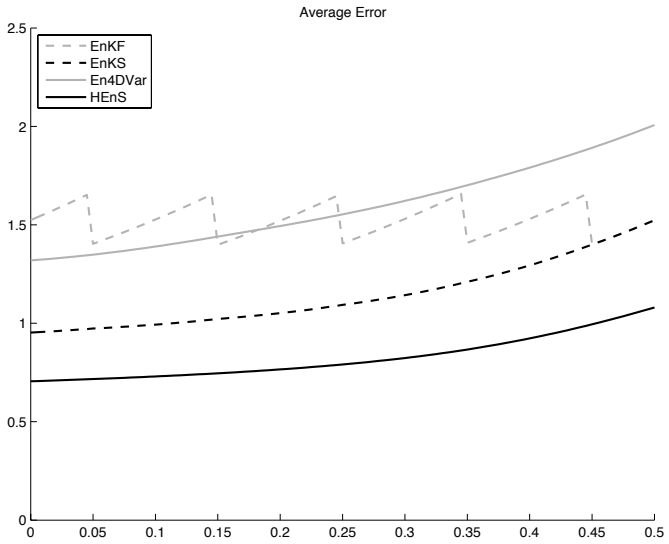


Fig. 4. This is the converged error plot for the Lorenz test case on EnKF, EnKS, En4DVar, and HEnS. At the left edge of the window, we see that En4DVar decreases the error from the background, but convergence to local minima prevents it from competing with EnKS and HEnS. By initializing an En4DVar minimization with the output from the EnKS, we can see that, statistically, HEnS reduces the error by an additional 50%.

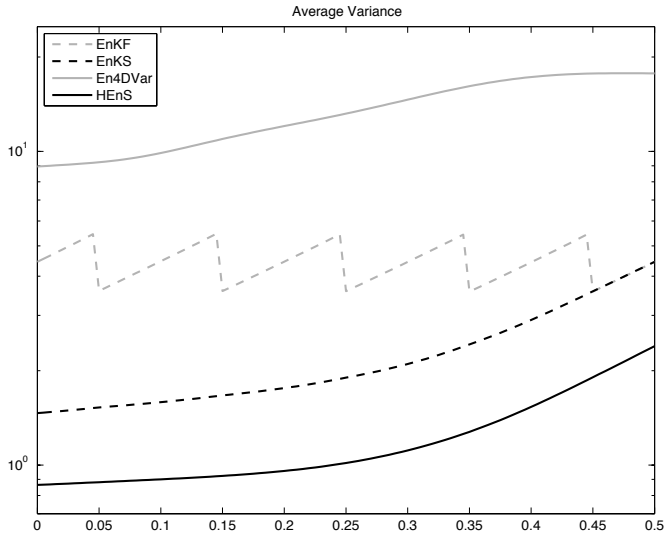


Fig. 5. This is the converged variance plot for the experimental Lorenz test case on EnKF, EnKS, En4DVar, and HEnS. The trace of the covariance matrix for each method is plotted as a function of time (averaged over several runs). En4DVar on its own does not perform as well as the other methods. However, when En4DVar is combined with EnKS to make HEnS, a substantially improved time-averaged performance is realized.

of the cases were run with $N = 300$ ensemble members, but similar results were found with significantly fewer ensemble members. A typical run is shown in Figure 3.

Statistical steady state was achieved via several weeks of statistical averaging with an efficient yet single-threaded implementation of the HEnS algorithm on a modern desktop computer (3 GHz Core Duo). The converged statistics are illustrated in Figures 4 and 5. One can see immediately that, on average, HEnS *significantly* outperforms the other three methods considered in terms of both accuracy (lower error) and precision (lower variance).

VI. Conclusion

A new, hybrid method (dubbed the Hybrid Ensemble Smoother, or HEnS) for state estimation in nonlinear chaotic systems has been proposed. This new method is based on a new variational formulation of the ensemble smoother, dubbed En4DVar, initialized using the (traditional, non-variational) formulation of the ensemble Kalman smoother (EnKS). The methods introduced in this work are proven to be *consistent*, meaning that they all reduce to the Kalman filter under the appropriate simplifying assumptions (that is, linear system, Gaussian state disturbances and measurement noise, and a sufficiently large number of ensemble members). Finally, the new HEnS algorithm has been shown, on average, to *significantly* outperform the leading scalable existing methods in a representative estimation problem related to a model nonlinear chaotic system. [Note that application of the HEnS method to more complex models is currently ongoing, and will be reported in future work.]

The reason for this remarkable success is that *HEnS provides an effective mechanism for revisiting past measurements in light of new data*, leveraging a smoother effectively to reinterpret past measurements based upon a more refined past state estimate, and thereby improving significantly the present state estimate. In essence, HEnS combines the powerful retrospective analysis of a variational method with the effective synthesis of the principal directions of uncertainty, as summarized by an ensemble-based method. An important ingredient to the method's operational effectiveness is the initialization of the variational analysis with the solution from a EnKS computation, which is far better than initializing this variational analysis simply with an EnKF computation based on older measurements.

Note finally that the HEnS method is based solely on variational and ensemble Kalman components that are already used heavily for operational weather forecasting, and are applied routinely, in real time, to systems with state dimension larger than 10^7. Thus, HEnS naturally inherits this effective scalability to very large scale chaotic systems, and holds significant promise to improve the accuracy of such practical estimation and forecasting efforts.

Appendix
Mixed Discrete/Continuous Adjoint Derivation

The full derivation of the gradient $\nabla J(\mathbf{u})$ is included here due to the unusual setting considered (that is, of a continuous-time system with discrete-time measurements). Perturbing the nonlinear model equations (1a) and linearizing about $\tilde{\mathbf{x}}(t)$ gives:

$$\frac{d\tilde{\mathbf{x}}'(t)}{dt} = A\big(\tilde{\mathbf{x}}(t)\big)\,\tilde{\mathbf{x}}'(t) \quad \text{with} \quad \tilde{\mathbf{x}}'_{-K} = \mathbf{u}' \tag{28}$$

$$\Rightarrow \quad \mathcal{L}\,\tilde{\mathbf{x}}' = 0 \quad \text{where} \quad \mathcal{L} = \frac{d}{dt} - A\big(\tilde{\mathbf{x}}(t)\big). \tag{29}$$

Similarly, the perturbed cost function is:

$$J'(\mathbf{u}') = (\mathbf{u} - \bar{\mathbf{x}}_{0|0})^T\, P_{0|0}^{-1}\, \mathbf{u}' - \sum_{k=1}^{K} (\mathbf{y}_k - H\tilde{\mathbf{x}}_k)^T\, R^{-1}\, H\tilde{\mathbf{x}}'_k. \tag{30}$$

The perturbed cost function (30) is not quite in the form necessary to extract the gradient, as illustrated in (14). However, there is an implicitly-defined linear relationship between $\mathbf{u}'$ and $\tilde{\mathbf{x}}'(t)$ on $t \in (0, T]$ given by (28). To re-express this relationship, a set of K adjoint functions $\mathbf{r}^{(k)}(t)$ are defined over the measurement intervals such that, for all $k \in [1, K]$, the adjoint function $\mathbf{r}^{(k)}(t)$ is defined on the closed interval $t \in \left[t_{k-1}, t_k\right]$. These adjoint functions will be used to identify the gradient. Towards this end, a suitable duality pairing is defined here as:

$$\langle\, \mathbf{r}^{(k)}\,,\, \tilde{\mathbf{x}}'\,\rangle = \int_{t_{k-1}}^{t_k} (\mathbf{r}^{(k)})^T\, \tilde{\mathbf{x}}'\, dt. \tag{31}$$

Then, the necessary adjoint identity is given by

$$\langle\, \mathbf{r}^{(k)}\,,\, \mathcal{L}\,\tilde{\mathbf{x}}'\,\rangle = \langle\, \mathcal{L}^*\,\mathbf{r}^{(k)}\,,\, \tilde{\mathbf{x}}'\,\rangle + b^{(k)}. \tag{32a}$$

Using the definition of the operator $\mathcal{L}$ given by (29) and the appropriate integration by parts, it is easily shown that

$$\mathcal{L}^*\,\mathbf{r}^{(k)} = -\frac{d\mathbf{r}^{(k)}(t)}{dt} - A\big(\tilde{\mathbf{x}}(t)\big)^T\,\mathbf{r}^{(k)}(t), \tag{32b}$$

$$b^{(k)} = (\mathbf{r}_k^{(k)})^T\,\tilde{\mathbf{x}}'_k - (\mathbf{r}_{k-1}^{(k)})^T\,\tilde{\mathbf{x}}'_{k-1}. \tag{32c}$$

Returning to the perturbed cost function, (30) can be rewritten as:

$$\begin{aligned} J'(\mathbf{u}') =& (\,\mathbf{u} - \bar{\mathbf{x}}_{0|0}\,)^T\,P_{0|0}^{-1}\,\mathbf{u}' - J'_K \\ &- \sum_{k=1}^{K-1} (\,\mathbf{y}_k - H\,\tilde{\mathbf{x}}_k\,)^T\,R^{-1}\,H\,\tilde{\mathbf{x}}'_k, \end{aligned} \tag{33a}$$

$$J'_K = \big[H^T\,R^{-1}\,(\,\mathbf{y}_K - H\,\tilde{\mathbf{x}}_K\,)\big]^T\,\tilde{\mathbf{x}}'_K. \tag{33b}$$

Looking at the adjoint defined over the last interval, $\mathbf{r}^{(K)}(t)$, the following criteria is enforced:

$$\mathcal{L}^*\,\mathbf{r}^{(K)} = 0 \quad\Rightarrow\quad \langle\, \mathcal{L}^*\,\mathbf{r}^{(K)}\,,\, \tilde{\mathbf{x}}'\,\rangle = 0, \tag{34a}$$

$$\mathbf{r}_K^{(K)} = H^T R^{-1}\,(\,\mathbf{y}_K - H\,\tilde{\mathbf{x}}_K\,). \tag{34b}$$

Substituting (29) and (34a) into (32a) for $k = K$ gives:

$$\begin{aligned} &b^{(K)} = 0 \\ &\Rightarrow (\mathbf{r}_K^{(K)})^T\,\tilde{\mathbf{x}}'_K - (\mathbf{r}_{K-1}^{(K)})^T\,\tilde{\mathbf{x}}'_{K-1} = 0, \\ &\Rightarrow \big[H^T\,R^{-1}\,(\mathbf{y}_K - H\,\tilde{\mathbf{x}}_K)\big]^T\,\tilde{\mathbf{x}}'_K = (\mathbf{r}_{K-1}^{(K)})^T\,\tilde{\mathbf{x}}'_{K-1}, \end{aligned} \tag{35}$$

which allows us to re-express J'_K in (33b) as

$$J'_K = (\mathbf{r}_{K-1}^{(K)})^T\,\tilde{\mathbf{x}}'_{K-1}. \tag{36}$$

Note that (34a) and (34b) give the full evolution equation and terminal condition for the adjoint $\mathbf{r}^{(K)}$ defined on the interval $t \in [\,t_{K-1}, t_K\,]$. Hence, a backward march over this interval will lead to the term $\mathbf{r}_{K-1}^{(K)}$ contained in (36).

The perturbed cost function (33a) can now be rewritten such that

$$\begin{aligned} J'(\mathbf{u}') =& (\,\mathbf{u} - \bar{\mathbf{x}}_{0|0}\,)^T\,P_{0|0}^{-1}\,\mathbf{u}' - J'_{K-1} \\ &- \sum_{k=1}^{K-2} (\,\mathbf{y}_k - H\,\tilde{\mathbf{x}}_k\,)^T\,R^{-1}\,H\,\tilde{\mathbf{x}}'_k, \end{aligned} \tag{37a}$$

$$\Rightarrow\quad J'_{K-1} = \big[\,H^T\,R^{-1}\,(\,\mathbf{y}_{K-1} - H\,\tilde{\mathbf{x}}_{K-1}\,) + \mathbf{r}_{K-1}^{(K)}\,\big]^T\,\tilde{\mathbf{x}}'_{K-1}. \tag{37b}$$

Enforcing the following conditions [cf. (34)] for the adjoint on this interval, $\mathbf{r}^{(K-1)}(t)$,

$$\mathcal{L}^*\,\mathbf{r}^{(K-1)} = 0, \tag{38a}$$

$$\mathbf{r}_{K-1}^{(K-1)} = H^T\,R^{-1}\,(\mathbf{y}_{K-1} - H\,\tilde{\mathbf{x}}_{K-1}) + \mathbf{r}_{K-1}^{(K)}, \tag{38b}$$

it can be shown via a derivation similar to (35) that

$$J'_{K-1} = (\mathbf{r}_{K-2}^{(K-1)})^T\,\tilde{\mathbf{x}}'_{K-2}, \tag{39}$$

which is of identical form as (36). Thus, it follows that each of the adjoints can be defined in such a way as to collapse the sum in the perturbed cost function (30) as above, until the last adjoint equation $\mathbf{r}^{(1)}$ reduces the perturbed cost function to the following:

$$J'(\mathbf{u}') = (\,\mathbf{u} - \bar{\mathbf{x}}_{0|0}\,)^T\,P_{0|0}^{-1}\,\mathbf{u}' - (\mathbf{r}_0^{(1)})^T\,\tilde{\mathbf{x}}'_0 \tag{40}$$

with the adjoints over the K intervals being defined as:

$$\begin{aligned} \frac{d\mathbf{r}^{(K)}(t)}{dt} &= -A\big(\tilde{\mathbf{x}}(t)\big)^T\,\mathbf{r}^{(K)}(t), \quad \text{where} \\ \mathbf{r}_K^{(K)} &= \;0\; + H^T\,R^{-1}\,(\,\mathbf{y}_K - H\,\tilde{\mathbf{x}}_K\,), \end{aligned}$$

$$\begin{aligned} \frac{d\mathbf{r}^{(K-1)}(t)}{dt} &= -A\big(\tilde{\mathbf{x}}(t)\big)^T\,\mathbf{r}^{(K-1)}(t), \quad \text{where} \\ \mathbf{r}_{K-1}^{(K-1)} &= \mathbf{r}_{K-1}^{(K)} + H^T\,R^{-1}\,(\,\mathbf{y}_{K-1} - H\,\tilde{\mathbf{x}}_{K-1}\,), \\ &\vdots \\ \frac{d\mathbf{r}^{(1)}(t)}{dt} &= -A\big(\tilde{\mathbf{x}}(t)\big)^T\,\mathbf{r}^{(1)}(t), \quad \text{where} \\ \mathbf{r}_1^{(1)} &= \mathbf{r}_1^{(2)} + H^T\,R^{-1}\,(\,\mathbf{y}_1 - H\,\tilde{\mathbf{x}}_1\,). \end{aligned} \tag{41}$$

Upon further examination, the system of adjoints (41) all have the same form. Each backward-marching adjoint variable $\mathbf{r}^{(k)}$ is endowed with a terminal condition that is the initial condition of the previous adjoint march $\mathbf{r}^{(k+1)}$ plus a correction due to the discrete measurement $\mathbf{y}_k$ at the measurement time t_k. Thus, the total adjoint march can be thought of as one continuous-time march of a single adjoint variable $\mathbf{r}(t)$ backward over the window $[\,t_0, t_K\,]$, with discrete "jumps" in $\mathbf{r}$ at each measurement time t_k. That is, (41) can be rewritten as:

$$\frac{d\mathbf{r}(t)}{dt} = -A\big(\tilde{\mathbf{x}}(t)\big)^T\,\mathbf{r}(t), \tag{42a}$$

which is marched backward over the entire interval $t \in [\,t_0, t_K\,]$ with $\mathbf{r}_K = 0$. At the measurement times (t_k for $k \in M$) the adjoint is updated such that

$$\mathbf{r}_k \leftarrow \mathbf{r}_k + H^T\,R^{-1}\,(\,\mathbf{y}_k - H\,\tilde{\mathbf{x}}_k\,). \tag{42b}$$

Then, this definition of the adjoint can be substituted into (40) to give:

$$J'(\mathbf{u}') = (\,\mathbf{u} - \bar{\mathbf{x}}_{0|0}\,)^T\,P_{0|0}^{-1}\,\mathbf{u}' - \mathbf{r}_0^T\,\tilde{\mathbf{x}}'_0, \tag{43}$$

$$\Rightarrow\quad J'(\mathbf{u}') = \left[\,P_{0|0}^{-1}\,(\,\mathbf{u} - \bar{\mathbf{x}}_{0|0}\,) - \mathbf{r}_0\,\right]^T\mathbf{u}', \tag{44}$$

where (44) is found by noting that $\tilde{\mathbf{x}}'_{-K} = \mathbf{u}'$. Then finally, from (14) and (44), the gradient sought may be written as:

$$\nabla J(\mathbf{u}) = P_{0|0}^{-1}\,(\,\mathbf{u} - \bar{\mathbf{x}}_{0|0}\,) - \mathbf{r}_0. \tag{45}$$

The resulting gradient[9] can then be used iteratively to update the current estimate via a suitable minimization algorithm (steepest descent, conjugate gradient, limited-memory BFGS, etc.).

ACKNOWLEDGMENT

The authors gratefully acknowledge the generous financial support of the National Security Education Center (NSEC) at Los Alamos National Laboratory (LANL) and numerous helpful discussions with Chris Colburn (UCSD), David Zhang (UCSD), Dr. Frank Alexander (LANL), Prof. Daniel Tartakovsky (UCSD), and Prof. Bruce Cornuelle (SIO).

REFERENCES

[1] R. Stratonovich, "Data assimilation and inverse methods in terms of a probabilistic formulation," *Radiofizika*, vol. 2, pp. 892—901, 1959.

[2] A. Jazwinski, *Stochastic Processes and Filtering Theory*. Academic Press, 1970.

[3] T. Bewley and A. Sharma, "A tractable framework for grid-based bayesian estimation of nonlinear low-dimensional systems with sparse nongaussian pdfs," *Automatica*, p. (submitted).

[4] N. G. S. Arulampalam, S. Maskell and T. Clapp, "A tutorial on particle filters for on-line non-linear/non-gaussian bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, pp. 174–188, 2002.

[5] N. G. A. Doucet and V. Krishnamurthy, "Particle filters for state estimation of jump markov linear systems," *IEEE Trans. Signal Processing*, vol. 49, pp. 613–624, 1993.

[6] S. Julier and J. Uhlmann, "A new extension of the kalman filter to nonlinear systems," *Proc. AeroSense: 11th Int. Symp. Aerospace/Defence Sensing, Simulation and Controls*, pp. 182–193, 1997.

[7] ——, "Unscented filtering and nonlinear estimation," *Proc. of the IEEE*, vol. 92, pp. 401–422, 2004.

[8] P. Swerling, "A proposed stagewise differential correction procedure for satellite tracking and prediction," *RAND Technical Report, Santa Monica, California*, 1958.

[9] R. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[10] R. Kalman and R. Bucy, "New results in linear filtering and prediction theory," *Journal of Basic Engineering*, vol. 83, no. 3, pp. 95–108, 1961.

[11] B. Anderson and J. Moore, *Optimal Filtering*. Dover, 1979.

[12] M. Ghil, S. Cohn, J. Tavantzis, K. Bube, and E. Isaacson, "Applications of estimation theory to numerical weather prediction," *Dynamic meteorology, data assimilation methods*, pp. 139–224, 1981.

[13] T. Kailath, "Some new algorithms for recursive estimation in constant linear systems," *IEEE Transactions on Information Theory*, vol. 19, pp. 750–760, 1973.

[14] A. Lorenc, "Analysis methods for numerical weather prediction," *Quarterly Journal of the Royal Meteorological Society*, vol. 112, no. 474, pp. 1177–1194, 1986.

[15] D. Parrish and J. Derber, "The National Meteorological Center's spectral statistical-interpolation analysis system," *Monthly Weather Review*, vol. 120, pp. 1747–1763, 1992.

[16] F.-X. L. Dimet and O. Talagrand, "Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects," *Tellus*, vol. 38A, p. 97, 1986.

[17] F. Rabier, J.-N. Thepaut, and P. Courtier, "Extended assimilation and forecast experiments with a four-dimensional variational assimilation system," *Quarterly Journal of the Royal Meteorological Society*, vol. 124, no. 550, pp. 1861–1887, July 1998.

[18] Z. Li and I. Navon, "Optimality of 4D-Var and its relationship with the Kalman filter and Kalman smoother," *Quarterly Journal of the Royal Meteorological Society*, vol. 127, no. 572, pp. 661–684, January 2001.

[9]Omitted in this gradient derivation is the substantial flexibility in the choice of the gradient definition (14) and the duality pairing (31). There is freedom in the choice of these inner products (e.g. by incorporating derivative and/or integral operators as well as weighting factors) that can serve to better precondition the optimization problem at hand without affecting its minimum points. This ability to precondition the adjoint problem is discussed at length in [51].

[19] H. Rauch, F. Tung, and C. Striebel, "Maximum likelihood estimates of linear dynamic systems," *AIAA Journal*, vol. 3, pp. 1445–1450, 1965.

[20] S. Cohn, N. Sivakumaran, and R. Todling, "A fixed-lag Kalman smoother for retrospective data assimilation," *Monthly Weather Review*, vol. 122, pp. 2838–2867, April 1994.

[21] H. Michalska and D. Mayne, "Moving horizon observers and observer-based control," *IEEE Transactions on Automatic Control*, vol. 40, pp. 995–1006, 1995.

[22] M. Fisher and P. Courtier, "Estimating the covariance matrices of analysis and forecast error in variational data assimilation," European Centre for Medium-Range Weather Forecasts, Tech Memo 220, 1995.

[23] T. Kraus, P. Kühl, L. Wirsching, H. Bock, and M. Diehl, "A moving horizon state estimation algorithm applied to the Tennessee Eastman Benchmark Process," in *Proceedings of the IEEE Conference on Multisensor Fusion and Integration for Intelligent Systems*, 2006, pp. 377–382.

[24] G. Evensen, "Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics," *Journal of Geophysical Research*, vol. 99, pp. 10 143–10 162, May 1994.

[25] P. Houtekamer and H. Mitchell, "Data assimilation using an ensemble Kalman filter technique," *Monthly Weather Review*, vol. 126, pp. 796–811, March 1998.

[26] ——, "A sequential ensemble Kalman filter for atmospheric data assimilation," *Monthly Weather Review*, vol. 129, pp. 123–137, November 2001.

[27] G. Evensen, "The ensemble Kalman filter: theoretical formulation and practical implementation," *Ocean Dynamics*, vol. 53, pp. 343–367, 2003.

[28] J. Anderson, "An ensemble adjustment Kalman filter for data assimilation," *Monthly Weather Review*, vol. 129, pp. 2884–2903, 2001.

[29] C. Bishop, B. Etherton, and S. Majumdar, "Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects," *Monthly Weather Review*, vol. 129, pp. 420–436, 2001.

[30] J. Whitaker and T. Hamill, "Ensemble data assimilation without perturbed observations," *Monthly Weather Review*, vol. 130, no. 7, pp. 1913–1924, 2002.

[31] S. Kim, G. Eyink, J. Restrepo, F. Alexander, and G. Johnson, "Ensemble filtering for nonlinear dynamics," *Monthly Weather Review*, vol. 131, pp. 2586–2594, November 2003.

[32] T. Hamill, J. Whitaker, and C. Snyder, "Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter," *Monthly Weather Review*, vol. 129, pp. 2776–2790, 2001.

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

[34] G. Evensen and P. van Leeuwen, "An ensemble Kalman smoother for nonlinear dynamics," *Monthly Weather Review*, vol. 128, pp. 1852–1867, June 2000.

[35] P. van Leeuwen and G. Evensen, "Data assimilation and inverse methods in terms of a probabilistic formulation," *Monthly Weather Review*, vol. 124, pp. 2898–2913, 1996.

[36] I. Hoteit, D. Pham, G. Triantafyllou, and G. Korres, "Particle Kalman filtering for data assimilation in meteorology and oceanography," *Under Preparation*, 2008.

[37] A. Lorenc, "The potential of the ensemble Kalman filter for NWP-a comparison with 4D-Var," *Quarterly Journal of the Royal Meteorological Society*, vol. 129, no. 595, pp. 3183–3203, 2003.

[38] A. Caya, J. Sun, and C. Snyder, "A comparison between the 4DVar and the ensemble Kalman filter techniques for radar data assimilation," *Monthly Weather Review*, vol. 133, no. 11, pp. 3081–3094, 2005.

[39] E. Kalnay, H. Li, T. Miyoshi, S. Yang, and J. Ballabrera-Poy, "4d-var or ensemble kalman filter?" *Tellus*, vol. 59A, pp. 758–773, 2007.

[40] N. Gustaffson, "Discussions on '4DVar of EnKF?'," *Tellus*, vol. 59A, pp. 774–777, 2007.

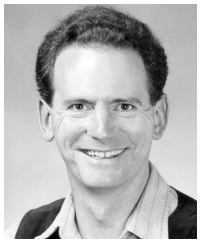
[41] T. Hamill and C. Snyder, "A hybrid ensemble Kalman filter–3D variational analysis scheme," *Monthly Weather Review*, vol. 128, no. 8, pp. 2905–2919, 2000.

[42] B. Hunt, E. Kalnay, E. Kostelich, E. Ott, D. Patil, T. Sauer, I. Szunyogh, J. Yorke, and A. Zimin, "Four-dimensional ensemble Kalman filtering," *Tellus*, vol. 56A, pp. 273–277, April 2004.

[43] M. Zhang, F. Zhang, and J. Hansen, "Coupling ensemble Kalman filter with four dimensional variational data assimilation," in *22nd Conference on Weather Analysis and Forecasting/ 18th Conference on Numerical Weather Prediction*, 2007.

[44] L. Berre, O. Pannekoucke, G. Desroziers, and G. Ştefănescu, "A variational assimilation ensemble and the spatial filtering of its error covariances," in *Proceedings of the ECMWF workshop on Flow-dependent aspects of Data Assimilation*. Reading, 2007, pp. 151–168.
[45] M. Fisher, "Assimilation techniques (4): 4dvar, April 2001," in *Meteorological training course lecture series*. ECMWF, 2002.
[46] J. Cessna, C. Colburn, and T. Bewley, "Multiscale retrograde estimation and forecasting of chaotic nonlinear systems," in *Proceedings of the 46th IEEE Conference on Decision and Control*, New Orleans, LA, USA, 2007.
[47] T. Bewley, *Numerical Renaissance: Simulation, Optimization, and Control*. under preparation, 2009.
[48] B. Anderson and J. Moore, *Optimal Filtering*. Dover Publications, 1979.
[49] F. Bouttier and P. Courtier, "Data assimilation concepts and methods march 1999," *Meteorological training course lecture series. ECMWF*, 2002.
[50] E. Lorenz, "Deterministic nonperiodic flow," *Journal of the Atmospheric Sciences*, vol. 20, pp. 130–141, January 1963.
[51] B. Protas and T. Bewley, "A computational framework for the regularization of adjoint analysis in multiscale pde systems," *Journal of Computational Physics*, vol. 195, pp. 49–89, 2004.

Joseph Cessna is a PhD candidate in the Mechanical and Aerospace Engineering department at UC San Diego. He is part of the interdisciplinary program in Computational Science, Mathematics, and Engineering (CSME). Mr. Cessna got his BS from the University of Wisconsin in Engineering Mechanics & Astronautics and Mathematics. He received his MS in Engineering Sciences from UC San Diego. His current research is centered around the development of novel data assimilation schemes for the estimation of chaotic, multiscale systems.



Thomas Bewley directs the Flow Control and Coordinated Robotics Labs at UC San Diego, and also co-directs UCSD's interdisciplinary program in Computational Science, Mathematics, and Engineering. Prof. Bewley got his BS and MS at Caltech and his PhD at Stanford. His recent work focuses on the the design optimization, feedback control, and real-time forecasting of fluid-mechanical systems in both the aeronautical and environmental settings, as well as the coordinated observation of such systems with autonomous unmanned vehicles.