EnVE: A new estimation algorithm for weather forecasting and flow control

Joseph Cessna,* Chris Colburn, and Thomas Bewley

Flow Control Laboratory, Dept. of MAE, UC San Diego, La Jolla, CA 92093, USA

Chaotic systems are characterized by long-term unpredictability. Existing methods designed to estimate and forecast such systems, such as Extended Kalman filtering (a "sequential" or "incremental" matrix-based approach) and 4Dvar (a "variational" or "batch" vector-based approach), are essentially based on the assumption that Gaussian uncertainty in the initial state, state disturbances, and measurement noise leads to uncertainty of the state estimate at later times that is well described by a Gaussian model. This assumption is not valid in chaotic systems with appreciable uncertainties. A new method is thus proposed that combines the speed and LQG optimality of a sequential-based method, the non-Gaussian uncertainty propagation of an ensemble-based method, and the favorable smoothing properties of a variational-based method. This new approach, referred to as Ensemble Variational Estimation (EnVE), is an extension of algorithms currently being used by the weather forecasting community. EnVE is a hybrid method leveraging sequential preconditioning of the batch optimization steps, simultaneous backwards-in-time marches of the system and its adjoint (eliminating the checkpointing normally required by 4Dvar), a receding-horizon optimization framework, and adaptation of the optimization horizon based on the estimate uncertainty at each iteration. If the system is linear, EnVE is consistent with the well-known Kalman filter, with all of its well-established optimality properties. The strength of EnVE is its remarkable effectiveness in highly uncertain nonlinear systems, in which EnVE consistently uses and revisits the information contained in recent observations with batch (that is, variational) optimization steps, while consistently propagating the uncertainty of the resulting estimate forward in time.

I. Introduction

The estimation and forecasting of chaotic, multiscale, uncertain fluid systems is one of the most highly visible computational 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 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, and chemical engineering settings lead to a variety of similar estimation problems[1] . While this paper makes no claims with regards to solving such important problems, it does address a new hybrid technique for the estimation and forecasting of such multiscale uncertain fluid systems that might one day have a significant impact in all of these areas.

The two most prevalent data assimilation strategies for the multiscale uncertain systems of interest are the Ensemble Kalman Filter[2] (EnKF) and the 4DVar[3] method. The Ensemble Kalman Filter 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 more traditional Extended Kalman Filter breaks down. The statistics of the estimation error in the EnKF are not propagated via a covariance matrix, but rather are implicitly approximated via the appropriate nonlinear propagation of several perturbed trajectories ("ensemble members") centered about the ensemble mean. The collection of these ensemble members (itself called the "ensemble"), propagates the statistics of the estimation error accurately in many problems even when a relatively small number of ensemble members is used. The 4Dvar method is a batch or "variational" method which propagates state and sensitivity or "adjoint" simulations back and forth across a time horizon of interest. An optimization is performed based on these marches order to minimize a cost function balancing the misfit of the estimate with the measurements, together with a "background" term which accounts for the measurements and corresponding estimate obtained before

*Corresponding author. e-mail: jcessna@ucsd.edu

this time window. For more information see our associated full paper on this topic[4] and the extensive review contained therein.

II. The EnVE Algorithm

The new Ensemble Variational Estimation (EnVE) algorithm is now presented as a consistent hybrid of the two aforementioned assimilation schemes, EnKF and 4DVar. Assume, without loss of generality, that an EnKF estimate $\widehat{\mathbf{X}}_{-j|-j}$ exists at some past time t_{-j}. This ensemble represents the best estimate at time t_{-j} given measurements up to and including $\mathbf{y}_{-j}$. At this point, available measurements up to t_0 are considered. The EnVE algorithm is initialized via a traditional sequential march of the EnKF up to the time of the most current measurement, t_0. This provides the current, best-estimate ensemble, $\widehat{\mathbf{X}}_{0|0}$, and all of its corresponding implicit statistics. The mean of the estimate is denoted $\bar{\mathbf{x}}_{0|0}$, and is found by taking the average of all the ensemble members. This is the best estimate at time t_0 given measurements up to and including this time. Doing a traditional KF march over this interval for a linear system would produce the optimal estimate at t_0. However, errors due to the nonlinearity of the chaotic system and approximations due to the finite size of the ensemble ultimately lead to a suboptimal estimate via the EnKF approach.

For forecasting applications, the most important estimate is the one at the most recent measurement time t_0, because it is this which is used as an initial condition for any forecasting calculation. With a linear system, any type of smoothing at this stage in the EnKF algorithm would have no effect on the estimate at t_0. The smoother would simply reduce the error in the past estimates, for some time $t < t_0$, using the information in the observations between t and t_0. However, for a nonlinear system, smoothing affects the entire estimate trajectory, even the most recent estimate at t_0. This is due to the dependence of the evolution of the estimate uncertainty on the trajectory of the estimate itself. For a linear system, the covariance propagation is independent of trajectory, but for a nonlinear system, changes in a past estimate (via smoothing) will impact the future trajectory of the estimate and its associated covariance. This motivates the consistent revisiting of past measurements to help improve the resulting forecast.

To this end, the ensemble $\widehat{\mathbf{X}}_{0|0}$ is marched backwards, using only the model equations. In so doing, the estimate retains the information captured by the measurements during the forward EnKF march. Thus, any point on this resulting trajectory is conditioned on all available measurements. At the conclusion of this backwards march, the ensemble mean and implicit statistics are known at some past time, say t_{-K}. This retrograde march is monitored in such a way as to define the width of the observation window for the variational step of the EnVE algorithm. If the initial estimate at t_0 is poor, then a lot of useful information may be deduced from a small time window containing only a few observations. Including more observations in this case is superfluous, and in fact unnecessarily increases the complexity of the optimization surface. Conversely, if the initial estimate at t_0 is very accurate, then a significantly longer variational window can, and should, be included in the analysis.

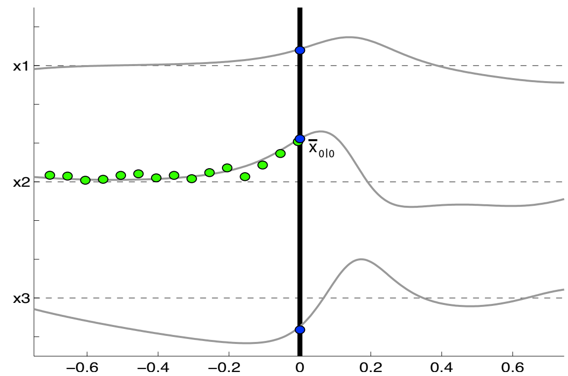

Figure 1. EnVE is initialized by marching a traditional EnKF forward through the available observations, making the appropriate updates. This provides the current, best-estimate of the state of the system $\bar{\mathbf{x}}_{0|0}$. At this point, it may be beneficial to revisit past measurements to update the trajectory of the estimate in light of the more recent measurements.

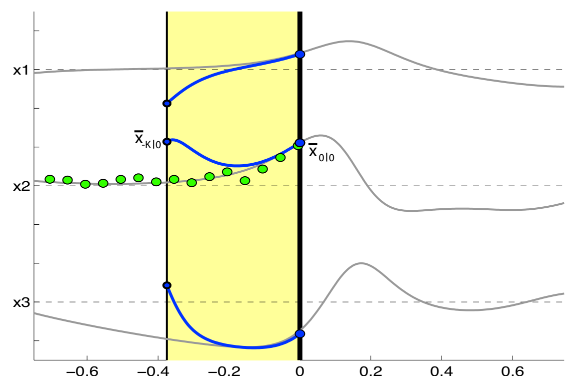

Figure 2. To determine the accuracy of the current estimate (that is, its correlation with the recent measurements), the ensemble at the current time is marched backwards using the system equations until the trajectory of the ensemble mean is significantly divergent from the observations. This gives the current best estimate at the past time, $\bar{\mathbf{x}}_{-K|0}$.

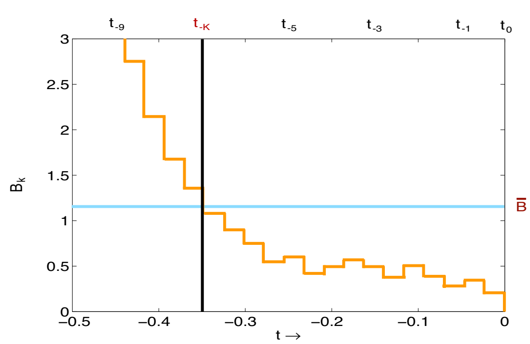

Figure 3. The accumulation of "bias" between the estimate trajectory and the observations is shown as the original estimate is marched backwards. Upon reaching a critical bias $\bar{B}$, the retrograde march is stopped. This time t_{-K} defines the width of the subsequent variational window.

The backwards march defines the window width by looking at the correlation between the initial estimate's trajectory and the past measurement history. Poor estimates diverge quickly from the measurements and should be analyzed with short optimization windows; conversely, accurate estimates march much further back in time before they begin to diverge from the measurements, and should be analyzed with longer optimization windows. To quantify this divergence, a "bias" measure, $B_k = \left\| \sum_{j=0}^{-k} (\mathbf{y}_j - H\,\overline{\mathbf{x}}_{j|0}) \right\|_1$, is calculated during the backward march. Through experimentation, a critical bias $\overline{B}$ is defined such that the ensemble mean is deemed significantly divergent from the observations past this period. With the variational window $[t_{-K}, t_0]$ so defined, the initial best smoothed estimate of the state $\overline{\mathbf{x}}_{-K|0}$ is given as the mean of the ensemble $\widehat{\mathbf{X}}_{-K|0}$. At this point, variational methods are used to improve this estimate in a consistent manner. To this end, the traditional 4DVar cost function is defined with a background estimate and covariance at t_{-K}. The background term of the cost function must now be defined carefully, as the correct background term is essential for EnVE to be consistent. In other words, properly defining the background terms in the variational cost function guarantees that erroneous updates are not made by using an observation more than once, and ensures that the result obtained reduces to that obtained by the Kalman Filter in the special case that the system considered happens to be linear.

Therefore, the background terms must be determined by returning to the original ensemble, $\widehat{\mathbf{X}}_{0|0}$, and marching it backwards again to the left edge of the window t_{-K}, this time *removing* the effect of the measurements along the way.

A suitable formula for removing the effect of a measurement can be found by rearranging the standard forward KF update equations for the mean and covariance.[4] This removal formula (combined with a suitable backwards march of the system) may be used to produce the background ensemble $\widehat{\mathbf{X}}_{-K|-K}$ at the left edge of the variational window. From this background ensemble, the background mean $\overline{\mathbf{x}}_{-K|-K}$ and background covariance $\mathcal{P}^e_{-K|-K}$ can be extracted and used to define the variational cost function. Because the background terms of the cost function are consistently defined (in that, in the linear setting, they incorporate no information from the observations in the variational window), the corresponding n-dimensional optimization surface is, in the linear case, identical to what would have been used had no sequential march through those observations been completed. The global minimum of this surface is independent of any previous updates to the estimate within the variational window that have been computed.

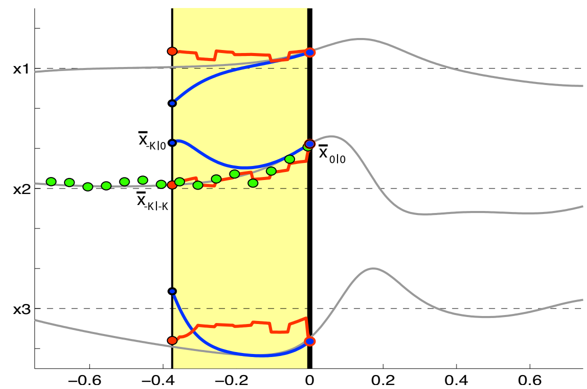

Figure 4. In order to fully define the variational cost function, the background terms at t_{-K} must be recalled. This is done by marching the original ensemble $\widehat{\mathbf{X}}_{0|0}$ backwards through the window, sequentially removing the effect of each measurement. This march results in a background ensemble $\widehat{\mathbf{X}}_{-K|-K}$ at the left edge of the window of interest. From this ensemble, the background mean and covariance can be inferred.

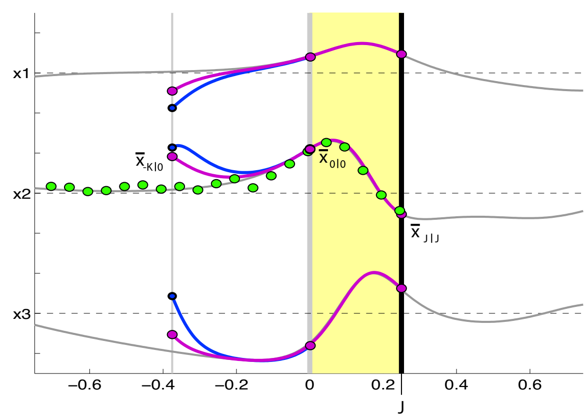

Figure 5. Upon completion of a variational iteration, the improved ensemble estimate $\widehat{\mathbf{X}}_{-K|0}$ at the back edge of the window is propagated forward to the old current time t_0. No measurement updates are done during this march, as all observations have been accounted for. Upon reaching the front edge of the window, the new ensemble estimate $\widehat{\mathbf{X}}_{0|0}$ is sequentially marched forward using the EnKF to account for any additional measurements received during the computation time required for the previous iteration.

With the cost function appropriately defined in this manner, a variational iteration can now be performed, similar to 4DVar. With traditional 4DVar, the first iteration is typically initialized using the background term, $\mathbf{u} = \overline{\mathbf{x}}_{-K|-K}$. However, with EnVE, a better estimate than this is already known, namely $\overline{\mathbf{x}}_{-K|0}$. This is one of the strengths of EnVE; it initializes the variational iteration with an estimate that is known to be significantly better than the background. In either case, the optimization surface is identical, but with EnVE, the initial estimate for $\mathbf{u}$ is much closer to the global minimum than the original background term. Consequently, if any significant improvement can be made upon the original best estimate, it will be discovered in the first variational iteration. Further, the original estimate is more likely to be in the region of attraction of this global minimum, so the probability of erroneous convergence to spurious local minima can be substantially reduced.

In practice, the estimate only exists as the average of the ensemble $\widehat{\mathbf{X}}_{-K|0}$. Hence, once the gradient and step size is determined from the variational step, each member of the ensemble is "shifted" in phase space the same amount, effectively shifting the ensemble mean without affecting the higher-order ensemble statistics. In fact, with an adjusted estimate of this sort, a modified, if not improved covariance $\mathcal{P}^e_{-K|0}$ would be expected as well. However, as variational methods do not provide a means for tracking these changes, EnVE must simply use this shifted ensemble representation, which is a bit conservative. Note, though, that this is a significant improvement over 4Dvar in which rigorous methods to march $\mathcal{P}$ are essentially unavailable. In contrast, with EnVE, the covariance associated with the original smoothed estimate is available, so it can be utilized. Though this is a conservative estimate of the covariance that does not account for the correction to the estimate due to the variational step, it correctly captures the main features of the covariance matrix, including the principle directions of estimate uncertainty.

To cycle the algorithm, the updated ensemble is marched forward to the front of the window. Note that the ensemble already accounts for the measurement in the window, so each ensemble member is propagated forward using the system equations only, with no additional measurement updates. This gives an improved best estimate at t_0, $\widehat{\mathbf{X}}_{0|0}$. During the period of this iteration, some new measurements $\{\mathbf{y}_1 \cdots \mathbf{y}_J\}$ will usually become available due to the computational time required to complete the variational step. The ensemble $\widehat{\mathbf{X}}_{0|0}$ can thus be marched forward again, using the EnKF to account for these new measurements, until the new current time t_J is reached. At this point, time is reset $t_0 \leftarrow t_J$, and the algorithm is repeated. Note that a significant computational burden can be avoided by storing the updated estimate at the previous current time, $\widehat{\mathbf{X}}_{0|0}$. This point can serve as the initial condition for finding the background terms of the variational cost function, as opposed to using $\widehat{\mathbf{X}}_{J|J}$. Depending on the relative widths of the next variational window and the time elapsed during the current variational step, using this saved background initial condition will result in either a shorter backwards EnKF march (very beneficial due to the ill-posed nature of such a march) or possibly even a forward EnKF march (a well-posed march) to the left edge of the new variational window. This simple storage trick reduces the computational cost of the algorithm significantly and shortens (or removes altogether) one of the ill-posed backwards marches.

II.A. EnVE Consistency

Ultimately, sequential methods (EnKF) and variational methods (4DVar) are used to solve the same problem. Both methods work to minimize a cost function to optimize the estimate at t_0 conditioned on all available measurements. Thus, when these cost functions are defined appropriately, it is possible to switch back and forth between sequential and variational methods consistently, as EnVE does. For a linear system with a set of measurements defined on $[t_{-K}, t_0]$, the smoothed KF estimate at the back edge of the window, $\bar{\mathbf{x}}_{-K|0}$ (found by marching a KF forward through the observations and marching the resulting estimate backwards to t_{-K}) is identical to the solution of a converged 4DVar algorithm with appropriately defined background terms. In other words, the optimal smoothed KF estimate $\bar{\mathbf{x}}_{-K|0}$ is the global minimum of the 4DVar cost function in the case of a linear system. For nonlinear systems, this relationship is still true, but the optimal smoothed KF estimate can not necessarily be found via a sequential estimator.

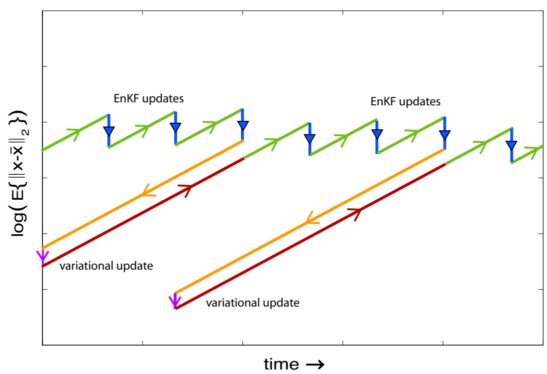

Figure 6. A cartoon illustrating the expected error for EnVE performed on a chaotic system. Exponential growth (linear growth in semi-log coordinates) in the expected error occurs during forward marches. Discrete reduction in the expected error occur at both the sequential updates and the variational update. Note that with a linear system, the variational update is necessarily zero, returning the estimate to its original value upon completion of the variational step.

This relationship is what EnVE attempts to exploit to improve the estimate. Marching an Ensemble Kalman Smoother (EnKS) will not produce the optimal smoothed estimate $\bar{\mathbf{x}}_{-K|0}$ because of the nonlinearities in the system and the approximations required for the ensemble framework. However, by removing the effect of the measurements and appropriately defining the 4DVar cost function background terms, this sub-optimal smoothed estimate can be used as an initial condition for the variational step. If the smoothed estimate $\bar{\mathbf{x}}_{-K|0}$ happens to be optimal, then the variational iteration is already converged and will produce a zero update to the estimate. Thus, EnVE uses the EnKS to initialize the 4DVar optimization, but does not reuse the information in the observations inconsistently. Thus, EnVE reduces to the expected optimal results of the Kalman Smoother (KS) for a linear system with both Gaussian measurement noise and disturbances.

A cartoon of the expected estimation error as EnVE progresses for a typical chaotic system is shown in Figure 6. Due to the chaotic nature of the system, any forward march of an estimate will lead to expected exponential growth of the estimation error (shown linearly in semi-log coordinates). Each EnKF measurement update creates a discrete drop in the expected estimation error. When a variational iteration is performed, the estimate is marched backwards. This causes an exponential decrease in the expected error as trajectories of the chaotic system will converge (along the attractor) during the backwards march. Then, a variational update is made, further reducing the expected error, and the resulting estimate is propagated forward again to the next available measurement. Recall that with a linear system, the update due to the variational step will have zero length, thus returning the estimate back to its original state to continue the sequential march. This helps illustrate the consistent nature of EnVE.

III. Advantages

By combining the statistical capabilities of the EnKF along with the batch processing/smoothing capabilities of a variational method, EnVE builds a better estimate of the system, possibly in real-time, at a justifiable computational cost. Using the EnKF to initialize a 4DVar-like iteration allows for fewer iterations because full convergence is not required and the initial estimate is more accurate than the background estimate alone. The intrinsic ability of the EnKF to represent the statistical properties of the estimate allows EnVE to repeatedly and consistently revisit past measurements and update the central trajectory of the ensemble about which the system can be linearized when considering its covariance evolution, based on new measurements.

Two of the main objectives for the development of EnVE was the desire for a multiscale-in-time algorithm combined with a receding horizon optimization framework. The advantages of these properties are highlighted in the following section. Combined, these two exclusive properties of EnVE create a dynamic optimization surface that tends to have desirable convergence properties for highly nonlinear systems.

III.A. Multiscale in Time

Because the variational window in EnVE is defined from the right (current time) by marching the current estimate backwards until divergence, the width of this window can be selected during the iteration. In contrast, with traditional 4DVar, this window width must be specified in advance. The variable variational window widths of EnVE can be used as a tool to precondition the optimization problem appropriately by coordinating this width with the accuracy of the initial estimate, as discussed previously.

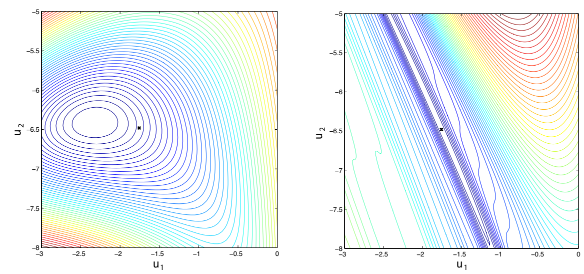

Figure 7. A cartoon illustrating the change in complexity of the optimization surfaces for a short variational window (left) and a long variational window (right). Also shown is the known truth model global minimum, which is more closely related to the global minimum of the highly irregular optimization surface of the longer window.

Due to the noise in the measurements, a short window containing only a few observations is prone to inaccuracy. That is, the global minimum of the cost function defined over only a few observations is likely to deviate significantly from the 'truth'. However, because only a few measurements are included in this short window (with corresponding short marches of the chaotic system) this optimization surface tends to be more regular with a larger region of attraction for the global minimum. The size of the region of attraction is important with gradient-based algorithms, as they are prone to converge to local minima.

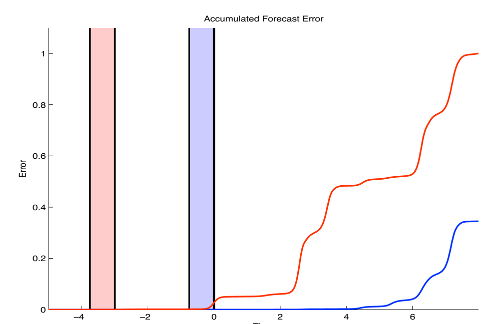

Figure 8. The accumulated forecast error from two forecasts is shown. The left-most variational window is due to a typical 4DVar without a receding horizon framework. The right-most window is due to EnVE with a receding horizon framework. Note the difference in the accumulated errors of each of forecast is due in large part to the time the forecast is ahead of the latest optimization window used. As this time is significantly reduced in the receding-horizon framework, forecasts made a certain amount of time into the future are greatly improved.

As the estimate improves, longer windows with more included observations can be utilized. This will tend to make the optimization surface more irregular and shrink the region of attraction for the global minimum, and thus this extension of the variational window needs to be done gradually enough that the improved estimate remains in this reduced region of attraction. Because more measurements are included in this window, the effect of sensor noise is diminished from the shorter window, making this global minimum more accurate with respect to the 'truth'.

III.B. Receding Horizon

A receding-horizon approach is defined by nudging the variational window forward in time to incorporate the most recent measurements obtained during each iteration of the variational optimization. Simplistic approaches to variational data assimilation leave the optimization window fixed until convergence. In contrast, EnVE redefines the optimization problem slightly at each iteration, updating it to include the newly-obtained measurements. As this modification causes the optimization surface to constantly shift, the algorithm never completely converges. However, the receding-horizon optimization framework updates the current estimate at each iteration with maximal efficiency, as it is constantly using the most up-to-date information available. Further, the resulting dynamic evolution of the optimization surface in fact helps to nudge the estimate out of the local minima into which it might otherwise settle.

A typical contrast between the error of two forecasts (one generated with a 4DVar algorithm and the other with EnVE) is shown in Figure 8. Unlike EnVE, due to the computation required for convergence of the 4DVar algorithm, the corresponding variational window has slipped into the past. Because of the chaotic nature of the systems of interest, any forecast will begin to exponential diverge. Consequently, much of the relevant range of the 4DVar forecast is wasted predicting events that have already taken place.

III.C. Parallel State/Adjoint Marches

Another advantageous side effect of posing the variational optimization problem in a retrograde setting deals with the numerical implementation of EnVE. The adjoint equation is marched backwards in time (from t_0 to t_{-K}) forced using the trajectory $\tilde{\mathbf{x}}(t)$. Typically, this trajectory is found by marching the initial condition $\tilde{\mathbf{x}}_{-K} = \mathbf{u}$ forward through the window (from t_{-K} to t_0). Especially for the multiscale systems of interest, this poses a large storage constraint on the problem because the adjoint is forced by the whole trajectory, but in reverse order. In other words, the trajectory of $\tilde{\mathbf{x}}(t)$ needs to be computed and saved over the entire interval before the adjoint march can begin. Attempts to circumvent this problem for large atmospheric scale systems include the checkpointing algorithm, in which the trajectory is stored only on coarse time grid points, and is either recomputed or linearly interpolated onto the fine time grid points when necessary. However, checkpointing still requires a substantial amount of storage and also significantly increases the computation required to compute the adjoint. Note that with EnVE, though, this required trajectory has already been computed in the second phase during the retrograde march of the original estimate. In fact, because the estimate trajectory is determined backwards in time, coupled with the fact that neither the background terms nor the width of the variational window need to be known a priori, a parallel march of all three systems (the estimate with the measurements, the adjoint, and the estimate without the measurements) is facilitated. Computationally, this is extremely efficient as there are no additional storage requirements for the adjoint march. Because they are marched in parallel, the estimate trajectory is immediately available to appropriately force the adjoint 'on the fly'. Then, when the mean of the ensemble diverges significantly from the observations, the parallel march can be halted, immediately providing the necessary gradient information from the adjoint, which is calculated at the same time.

IV. Summary and Conclusions

In this paper, a new hybrid data assimilation method is summarized: Ensemble Variational Estimation (EnVE). For a more detailed discussion, please see the associated journal article.[4] The new method leverages the nonlinear statistical propagation properties of the sequential EnKF/EnKS to initialize and properly define an appropriate variational iteration, similar to 4DVar. This variational iteration is posed in such a way as to allow for a multiscale-in-time, receding-horizon optimization framework. The smoothed estimate from the EnKF is used as an accurate initial condition for the variational iteration, thus improving its overall performance. The multiscale-in-time framework is achieved via a retrograde march of the current estimate over the available observations, and appropriately preconditions the variational step. This also allows for a concurrent, parallel march of the appropriate adjoint equation, which is forced by the backwards march of the estimate. Thus, no additional storage is required for the gradient computation, as is typical with a 4DVar implementation. Because the variational window width is a function of the accuracy of the estimate, EnVE tends to update poor estimates with short windows and more accurate estimates with longer windows. Finally, EnVE is a consistent and convenient hybrid of the basic EnKF and 4DVar algorithms already in wide use that reduces to the KF in the linear setting. Thus, much of the current work in the EnKF and 4DVar may be applied to the EnVE algorithm while maintaining its desirable properties and consistency. It is only with such combined efforts that it may be possible to develop significantly improved large-scale data assimilation algorithms in the years to come.

References

[1] Kim, J. and Bewley, T. (2007) A linear systems approach to flow control. *Annual Review of Fluid Mechanics* **39**, 383–417.

[2] Evensen, G. (1994) Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Review* **99**, 10143–10162.

[3] Le Dimet, F.-X. and Talagrand, O. (1986) Variational algorithms for analysis and assimilation of meteorological observations; theoretical aspects. *Tellus* **38A**, 97.

[4] Bewley, T., Cessna, J. and Colburn, C. (2008) EnVE: A consistent hybrid ensemble/variational estimation strategy for multiscale uncertain systems. *Under preparation*.