Mixing the Elements of 4D-Var and Ensemble Kalman Filters

Building Practical Data Assimilation Algorithms

Istvan Szunyogh

Institute for Physical Science and Technology &
Department of Atmospheric and Oceanic Science
University of Maryland

THORPEX/WWRP Workshop on 4D-Var and Ensemble Kalman Filter Inter-Comparisons

Outline

The Mathematical Model

Behind Both Types of Schemes

The most likely trajectory of the system is the one that minimizes the cost function

$$J^o(\{\mathbf{x}(t)\}) = \left[\mathbf{y}^o - \mathbf{h}\left(\{\mathbf{x}(t)\}\right)\right]^T \mathbf{R}^{-1} \left[\mathbf{y}^o - \mathbf{h}\left(\{\mathbf{x}(t)\}\right)\right]. \qquad (1)$$

Two assumptions are made to obtain the cost function:

- The observation error is Gaussian
- The observed quantities depend on the system trajectory in a known way $\mathbf{h}\left(\{\mathbf{x}(t)\}\right)$

Notation:

- $\mathbf{y}^o$: vector of all observations along the trajectory
- $\mathbf{R}$: observation error covariance matrix
- $\mathbf{x}$: state of the system

The “Canonical” Elements of the Two Schemes

- Common Element: A nonlinear (possibly imperfect) model of the system
- 4D-Var:
  - The cost function is minimized directly using all past and present observations
  - The mapping between tangent spaces along a nonlinear trajectory is represented by the tangent linear map
- EnKF
  - Observations are assimilated sequentially
  - The mapping between the spaces of perturbations along the nonlinear ensemble-mean trajectory is represented by an ensemble of model trajectories (when the perturbations are small, the space of perturbations at a given time is a representation of the tangent space at that time and the mapping is linear)

Synergistic Approaches I

Observation: Whether the observations are assimilated concurrently by a variational approach or sequentially is independent of whether the error propagation is represented by the tangent linear equation or by an ensemble of trajectories.

Corollary: The main components of 4D-Var and EnKF are interchangeable.

Examples for Variational Algorithms that Use an Ensemble to Represent the Error Dynamics:

- Maximum Likelihood Ensemble Filter (MLEF, Zupanski 2005, MWR)
- Data Assimilation using Modulated EnsembleS (DAMES, Bishop, talk on Wednesday)
- Variational LETKF (Harlim and Hunt 2007, Tellus)

General Mathematical Model

Sequential Approach Applied to Short Time Intervals

The total length of time is divided into n time intervals. The cost function is

$$J^o(\{\mathbf{x}(t)\}) = \sum_{j=1}^{n} \left[\mathbf{y}_j^o - \mathbf{h}_j(\{\mathbf{x}(t_j)\})\right]^T \mathbf{R}_j^- 1 \left[\mathbf{y}_j^o - \mathbf{h}_j(\{\mathbf{x}(t_j)\})\right]. \quad (2)$$

The newly added assumption is

- The the errors of the observations assimilated at different times t_j ($j = 1, \ldots, n$) are uncorrelated

The Cost Function

The Effect of All Past Observations is Represented through a Background $\mathbf{x}^b$

The background is obtained by an integration of the model

$$\mathbf{x}_n^b = M_{t_{n-1},t_n}(\mathbf{x}_{n-1}^a), \quad (3)$$

The cost function is

$$\begin{aligned} J_{t_n}^o(\mathbf{x}) &= [\mathbf{x}-\mathbf{x}_n^b]^T(\mathbf{P}_n^b)^{-1}[\mathbf{x}-\mathbf{x}_n^b] + \quad (4) \\ & \left[\mathbf{y}_n^o - \mathbf{h}_n\left(\{\mathbf{x}(t_n)\}\right)\right]^T \mathbf{R}_n^{-1}\left[\mathbf{y}_n^o - \mathbf{h}_n\left(\{\mathbf{x}(t_n)\}\right)\right]. \quad (5) \end{aligned}$$

Notations:

- M_{t_{n-1},t_n}: Model dynamics for the time interval $[t_{n-1}, t_n]$
- $\mathbf{P}_n^b$: Background error covariance matrix

A Couple of Observations

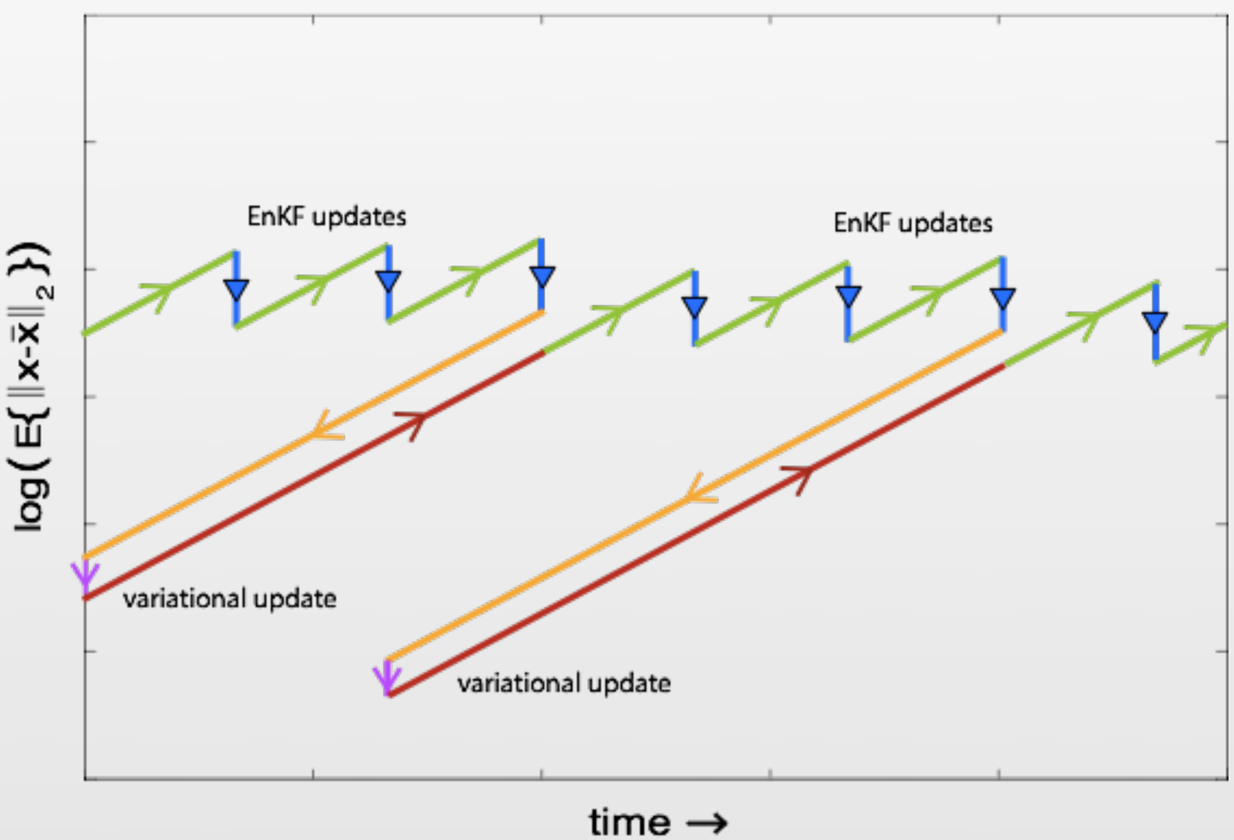
- Correlations between the observation errors within the assimilation window are allowed regardless of which scheme we use to find the minimum of the cost function
- All schemes face the problem of estimating $\mathbf{P}_n^b$ regardless of how we minimize the cost function

Synergistic Approaches II

- Most currently used Ensemble Kalman Filters are 4D (as they interpolate the background ensembles in time to the time of the observations, such evolving $\mathbf{P}^b$ within the assimilation window. (Hunt et al. 2004, Tellus)
- The most popular synergistic approach is to estimate $\mathbf{P}^b$ with an ensemble Kalman filter and to find the minimum of the cost function with a 4D-Var (e.g., E4DVAR, Zhang et al. 2007, UKMO)

EnVE: Ensemble/Variational Estimation

Expected Error (EnVE)



Central Idea

Question: in a hybrid method, can one "use a measurement more than once"?

Answer: YES, but only if done **consistently**.

Scheme must reduce to Kalman filter if system is linear and state disturbances & measurement noise are Gaussian.

Cartoon above shows evolution of the log of the expected estimation error with time.

Sawtooth pattern is EnKF (continuous-time system, discrete-time measurement updates).

Orange/Red is error before/after variational update.

If system is linear, variational update goes to zero, reducing EnVE to EnKF result.

Comprehensive description at http://flowcontrol.ucsd.edu/EnVE.pdf

Concluding Remarks

What will be the optimal mixing of the elements?

My personal prediction of the optimal mix, although I may turn out to be completely wrong:

- Ensemble Kalman Filter based computation of $\mathbf{P}_n^b$ (seems more practical than a weak constrained 4D-Var)
- Variational (global or local) approach to find the minimum of the cost function for a number of reasons
  - Better handling of nonlinearities in the dynamics and the observation operator
  - Easier implementation of balance constraints
  - Possibility of Variational QC
  - Possibility of other considerations that result in a non-quaratic cost function