Accounting for Skewness in Ensemble Data Assimilation

Daniel Hodyss

Naval Research Laboratory, Marine Meteorology Division, Monterey, California

Abstract

A new framework is presented for understanding how a non-normal probability density function (pdf) may affect a state estimate and how one might usefully exploit the non-normal properties of the pdf when constructing a state estimate. A Bayesian framework is constructed that leads naturally to an expansion of the expected forecast error in a polynomial series consisting of powers of the innovation vector. This polynomial expansion in the innovation reveals a new view of the geometric nature of the state estimation problem. It is shown that this expansion in powers of the innovation provides a direct relationship between a non-normal pdf describing the likely distribution of states and a normal pdf determined by powers of the forecast error. A practical data assimilation algorithm is presented that explicitly accounts for skewness in the prior distribution. The algorithm operates as a global-solve (all observations are considered at once) using a minimization-based approach and Schur/Hadamard (element-wise) localization. The central feature of this technique is the squaring of the innovation and the ensemble perturbations so as to create an extended state-space that accounts for the second, third and fourth moments of the prior distribution.

1. Introduction

Ensemble-based data assimilation (DA) is rapidly becoming the technique of choice for the estimation of the state of a geophysical system. This popularity is largely due to the significant ease of implementation afforded by the use of an Ensemble-based Kalman Filter (EnKF) DA system. The EnKF is a Monte-Carlo state-estimation technique for estimating the posterior mean and for generating random draws from the posterior distribution. The application of this technique in the meteorological community has been met with considerable success in a wide-range of applications (e.g., Houtekamer et al. 2005, Whitaker et al. 2008, Anderson et al. 2009). There are, however, unresolved issues with the application of the EnKF to the highly nonlinear dynamics inherent to meteorological flows at high resolution. Situations in which the EnKF is known to have some difficulty, and where nonlinearity may be the culprit, include: the assimilation of vortex position (Lawson and Hansen 2005, Chen and Snyder 2007), radar observations (Dowell et al. 2011), parameter estimation (Hacker et al. 2011), and observations over a long assimilation window (Khare et al. 2008). We speculate that one reason state and parameter estimation in these situations is sometimes difficult is the fact that the relationship between the prior estimates of the observed variables and the state-vector may be nonlinear. This nonlinearity may come about from the nonlinearity in the model operator (i.e. model dynamics) or from the nonlinearity in the observation operator used to observe the system. In either case, this nonlinear relationship will generally lead to skewed (non-zero third moment) posterior distributions that result in suboptimal behavior from the EnKF.

2. Issues with Skewness

Hodyss (2011) showed that whenever the posterior is skewed the posterior mean is a nonlinear (curved) function of the innovation. Hodyss (2011) showed that the posterior mean, $\bar{x}$, is related to posterior third moment through:

$$\frac{d^2\bar{x}}{dv^2} = \frac{T}{R^2}$$

where T is the posterior third moment, R is the observation error variance, and v is the innovation. Because the EnKF estimate of the posterior mean is a linear function of the innovation this leads to significant errors in its estimate of the posterior mean.

Hodyss (2011) also showed that ensemble generation is also difficult when there exists a posterior third moment because

$$\frac{dP_a}{dv} = \frac{T}{R}$$

which implies that the analysis error variance is a function of the most recent innovation. However, the EnKF algorithm assumes the analysis error variance is independent of innovation.

3. Data Assimilation through Bayes' Rule

We imagine the true state, x, to be drawn from a distribution $\rho(x)$ that we will refer to as the prior. At the present time we have available an observation y drawn from a distribution we refer to as the observation likelihood such that we may use Bayes' rule to obtain a density that describes the combined knowledge of the distribution of possible true states:

$$\rho(x|y) \sim \rho(y|x)\rho(x)$$

A standard estimation technique for the true state given the posterior density is to find its mean, i.e.

$$\bar{x}(y) = \int_{-\infty}^{\infty} x\rho(x|y)\,dx$$

Because the observation and the innovation are linearly related we may equally well condition on the innovation

$$\bar{x}(v) = \bar{x}_f + \int_{-\infty}^{\infty}(x-\bar{x}_f)\rho(x|v)\,dx$$

This equation shows that the correction to the prior mean should be the expected forecast error given todays innovation. Because the posterior may be a nonlinear function of the innovation a DA system that aims to deliver the posterior mean must also be a nonlinear function of the innovation. One way to accomplish this task is through nonlinear regression. We may expand the integral above into a Taylor-series to perform nonlinear polynomial regression:

$$\bar{x}(v) - \bar{x}_f = \int_{-\infty}^{\infty}(x-\bar{x}_f)\rho(x|v)\,dx = M_1 v + M_2\left(v^2 - \langle v^2\rangle\right) + \ldots$$

$$M_1 = K - (I-K)T_f\Pi^{-1}T_f\left(P_f+R\right)^{-1}$$

$$M_2 = (I-K)T_f\Pi^{-1} \qquad K = P_f\left(P_f+R\right)^{-1}$$

$$\Pi = F_f + 6RP_f - T_f^2\left(P_f+R\right)^{-1} - \langle v^2\rangle^2$$

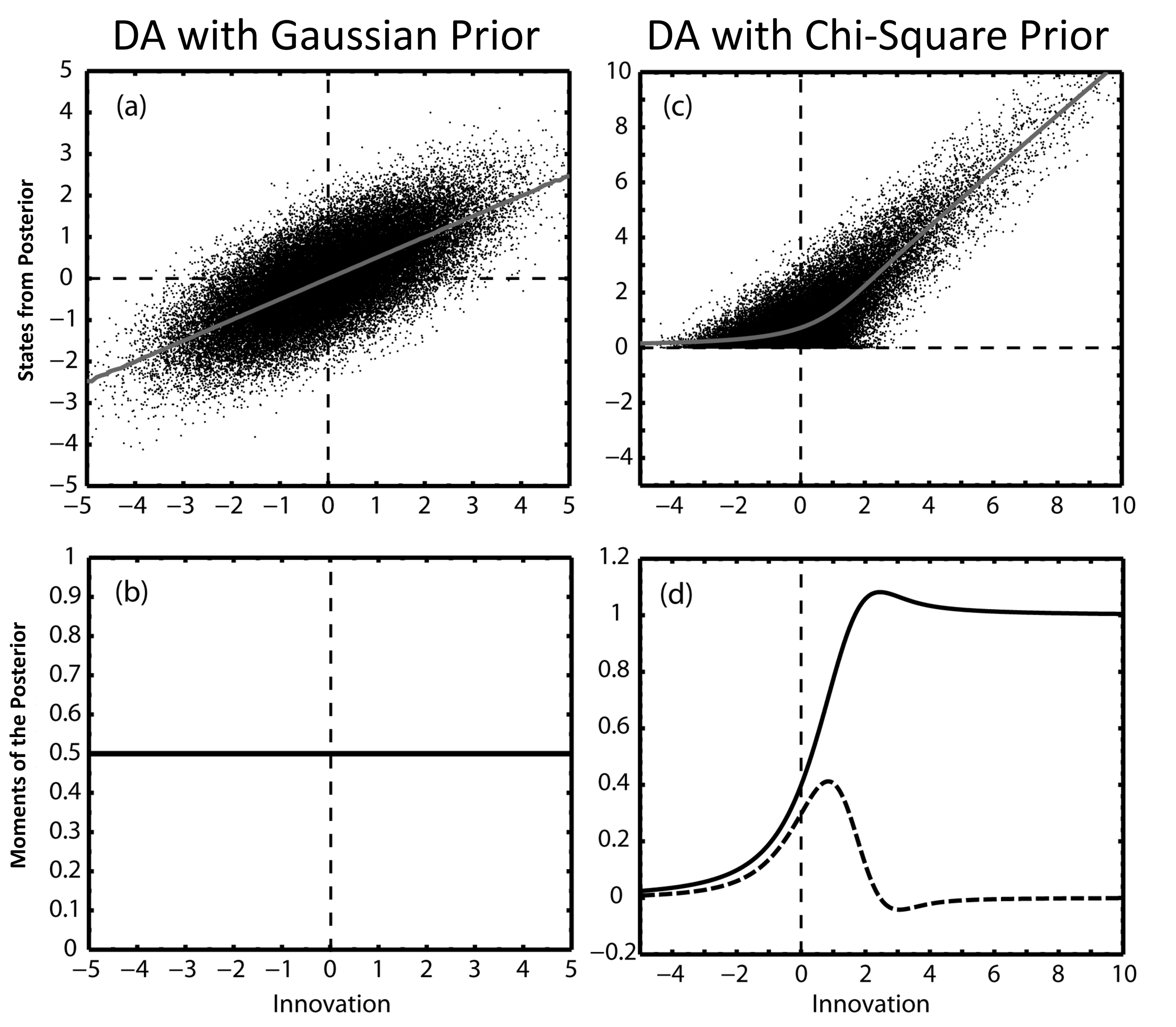

Figure 1. Two posterior distributions as represented by a scatter diagram of realizations of the truth as a function of the innovation. In (a) and (c) are scatter diagrams of a posterior derived from a normal and chi-square prior, respectively. The thick gray line is the mean of the distribution. In (b) and (d) are listed the second (solid) and third (thick dashed) moments of the two distributions in (a) and (c), respectively.

4. A Global-Solve Algorithm

Define an extended innovation vector as

$$\hat{\mathbf{v}}' = \left[\mathbf{v}^T \quad (\mathbf{v}\odot\mathbf{v})^T\right]^T - \left[\langle\mathbf{v}\rangle^T \quad \langle\mathbf{v}\odot\mathbf{v}\rangle^T\right]^T$$

where $(\mathbf{v}\odot\mathbf{v})^T = \left[v_1^2 \quad v_2^2 \quad \cdots \quad v_p^2\right]$

Similarly, the observation operator is extended

$$\hat{\mathbf{H}} = \begin{bmatrix} \mathbf{H} & \mathbf{0}_{pxN} \\ \mathbf{0}_{pxN} & \mathbf{H} \end{bmatrix}$$

We define an extended covariance matrix as

$$\hat{\mathbf{P}}_f = \begin{bmatrix} \mathbf{P}_f & \mathbf{T}_f \\ \mathbf{T}_f^T & \mathbf{F}_f - \mathbf{p}_d\mathbf{p}_d^T \end{bmatrix} \approx \hat{\mathbf{Z}}_r\hat{\mathbf{Z}}_r^T$$

where

$$\hat{\mathbf{Z}}_r = \frac{1}{\sqrt{K-1}}\begin{bmatrix} \boldsymbol{\varepsilon}_1 & \boldsymbol{\varepsilon}_2 & \cdots & \boldsymbol{\varepsilon}_K \\ \boldsymbol{\varepsilon}_1\odot\boldsymbol{\varepsilon}_1 - \mathbf{p}_d & \boldsymbol{\varepsilon}_2\odot\boldsymbol{\varepsilon}_2 - \mathbf{p}_d & \cdots & \boldsymbol{\varepsilon}_K\odot\boldsymbol{\varepsilon}_K - \mathbf{p}_d \end{bmatrix}$$

The extended observation error covariance matrix is

$$\hat{\mathbf{R}} = \begin{bmatrix} \mathbf{R} & \mathbf{0}_{pxp} \\ \mathbf{0}_{pxp} & \mathbf{R}_4 - \mathbf{r}_d\mathbf{r}_d^T + 4\mathbf{R}\odot\left(\mathbf{HP}_f\mathbf{H}^T\right) \end{bmatrix}$$

With these definitions we may perform nonlinear quadratic regression using the following formula

$$\bar{\mathbf{x}}_a = \bar{\mathbf{x}}_f + \mathbf{Zw} \qquad \mathbf{w} = \hat{\mathbf{Z}}^T\hat{\mathbf{H}}^T\left[\hat{\mathbf{H}}\hat{\mathbf{P}}_f\hat{\mathbf{H}}^T + \hat{\mathbf{R}}\right]^{-1}\hat{\mathbf{v}}'$$

The most difficult part of the calculation is the matrix inversion in the calculation of the weights. The most efficient way to do this in a global-solve (all observations assimilated at once) is through a minimization approach.

Step 1. Solve $\left(\hat{\mathbf{H}}\hat{\mathbf{P}}_f\hat{\mathbf{H}}^T + \hat{\mathbf{R}}\right)\mathbf{u} = \hat{\mathbf{v}}'$ for **u**. We may write a better conditioned version as

$$\left(\mathbf{Z}_N\mathbf{Z}_N^T + \mathbf{I}\right)\mathbf{a} = \hat{\mathbf{R}}^{-1/2}\hat{\mathbf{v}}' \qquad \mathbf{u} = \hat{\mathbf{R}}^{-1/2}\mathbf{a} \qquad \mathbf{Z}_N = \hat{\mathbf{R}}^{-1/2}\hat{\mathbf{H}}\hat{\mathbf{Z}}$$

Typically, one would solve this equation using a technique like conjugate gradient.

Step 2. Solve for the weights of the prior through $\mathbf{w} = \mathbf{Z}_N^T\mathbf{a}$

Step 3. Solve for the mean update through $\bar{\mathbf{x}}_a = \bar{\mathbf{x}}_f + \mathbf{Zw}$

Note: The difference between this algorithm and a traditional global-solve method, such as Buehner (2005), is simply the extension of the length of the state-vectors to include the quadratic perturbations. Therefore, this method does not require a new system to be developed from scratch. Rather, this algorithm can be seen as a modification to the system the user presently has constructed.

5. Application

Here we apply the algorithm to the left to a 2-d shear layer simulation using the nonlinear Boussinesq equations. Localization is applied consistent with Bishop et al. (2011). The state vector is of length 8448 elements. This system is of high enough dimension that both localization and prior inflation were required to prevent filter divergence at the ensemble sizes considered here. Both localization and prior inflation are tuned separately for both the EnKF as well as the quadratic ensemble filter. Ensemble generation was performed with the method referred to as perturbed observations. Three different cycling intervals of 200, 300, and 400 model time steps are tested. Observations of zonal wind and temperature at 10 equally spaced vertical soundings with 32 observations in each vertical sounding are taken. We cycle for 320 cycles, throw away the first 20, and calculate statistics on the remaining 300 cycles.

In figure 2 we see the RMS analysis error for the different experiments as function of ensemble size. In all experiments the quadratic ensemble filter outperforms the traditional EnKF for all cycling intervals and ensemble sizes shown. Tests with a cycling interval of 100 model time steps showed no improvement over the EnKF and tests with an ensemble size of 32 showed degradation. Hence, we conclude that given a sufficient ensemble size and a sufficiently long cycling interval (longer cycling intervals have stronger non-Gaussian distributions) then the quadratic ensemble filter is superior to a state-of-the-art EnKF.

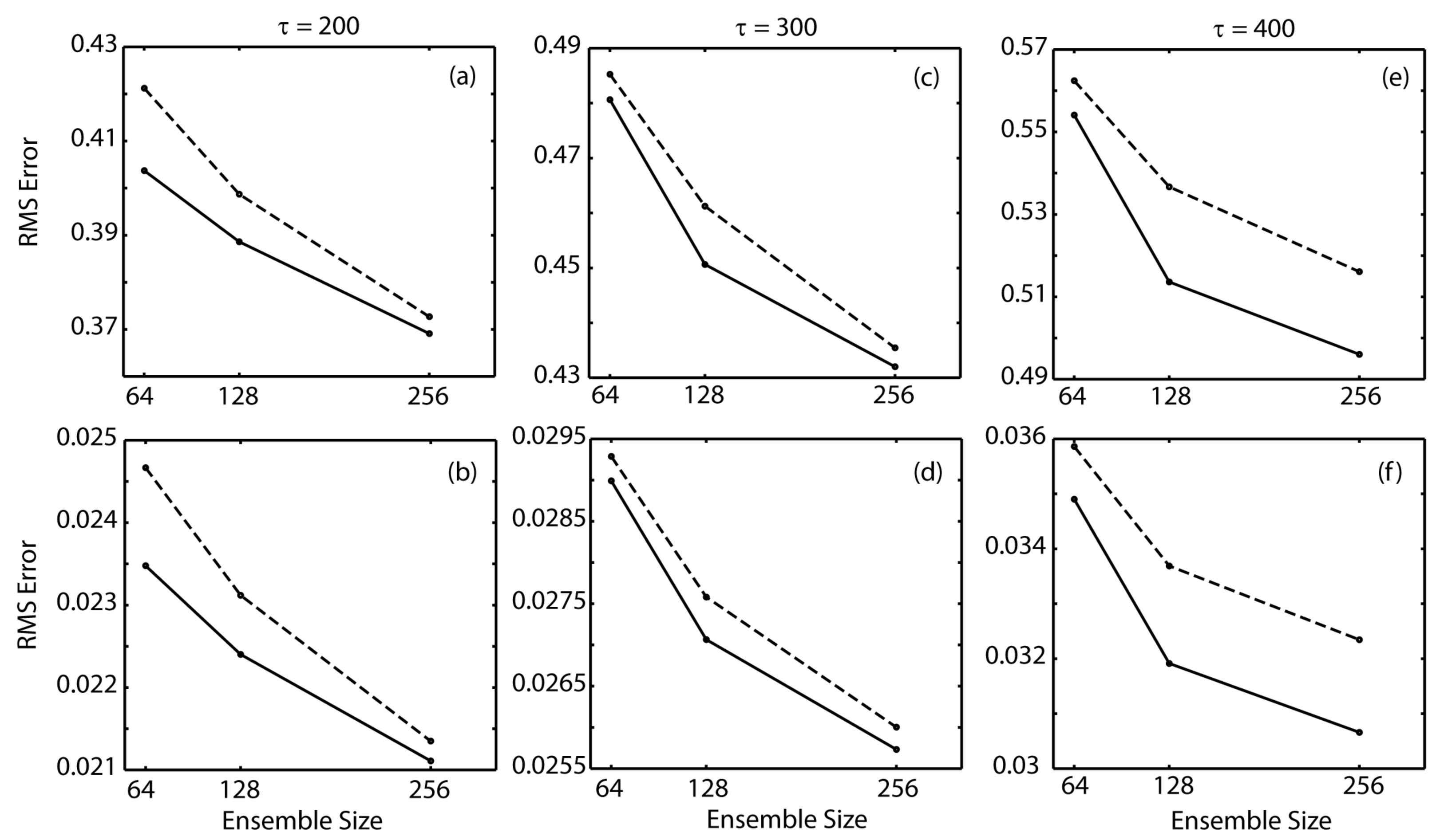

Figure 2 RMS analysis error. In (a), (c), (e) [(b), (d), (f)] are the RMS analysis errors for the vorticity [temperature] for a cycling interval of τ = 200, 300, and 400, respectively. Solid lines are for the Quadratic Ensemble Filter and dashed lines are for the EnKF.

Acknowledgement: *DH gratefully acknowledges support from the NRL Base Program through BE-4287-0-0-5.*

Please see the following articles for further information:

Hodyss, D., 2011: Ensemble state estimation for nonlinear systems using polynomial expansions in the innovation. *Mon. Wea. Rev.*, **139**, 3571-3588.
Hodyss, D., 2012: Accounting for skewness in ensemble data assimilation. *Mon. Wea. Rev.*, **140**, 2346-2358.

Contact email: daniel.hodyss@nrlmry.navy.mil