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 of 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 the powers of the forecast error. A practical data assimilation algorithm is presented that explicitly accounts for skewness in the prior distribution.

4. A Global-Solve Algorithm

Define an extended innovation vector as
\[ \mathbf{x}^e = \mathbf{x} + \mathbf{v} \]

where \( \mathbf{x}^e \) is the extended state vector containing the innovation vector \( \mathbf{v} \).

Similarly, the observation operator is extended
\[ \mathbf{H}^e = \mathbf{H} + \mathbf{R} \]

We define an extended covariance matrix as
\[ \mathbf{P}^e = \mathbf{P} + \mathbf{R} \]

where \( \mathbf{P}^e \) is the extended state covariance matrix.

The extended observation error covariance matrix is
\[ \mathbf{R}^e = \mathbf{R} + \mathbf{R}_{\text{prior}} + \mathbf{R}_{\text{post}} \]

4.1. Data Assimilation through Bayes’ Rule

The state estimate is the posterior mean, which is a linear function of the innovation. This leads to a minimization-based approach. The problem of finding the posterior mean is a nonlinear (curved) function of the innovation.

4.2. Issue 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 whenever the posterior is skewed, \( \mathbf{x} \), is related to posterior third moment through:
\[ \mathbf{x} = \mathbf{x}^e - \mathbf{E}(\mathbf{x}^e) \]

5. Application

Here we apply the algorithm to a 2-d shear layer simulation using the nonlinear Boussinesq equations. Localization was applied consistent with Hodyss et al. (2011). The state vector is of length 8448 elements. The system is of high enough dimension that both localization and prior inflation were required to prevent filter divergence at the ensemble size 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 were 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 120 cycles, throw away the first 20, and calculate statistics on the remaining 100 cycles.

6. Acknowledgments

...