Ensemble Based Error Covariance Matrices for Mesoscale Variational Data Assimilation

M.S.F.V. De Pondeca*, X. Zou
Department of Meteorology, The Florida State University, Tallahassee, Florida

1 Introduction

The evaluation of the background error covariance matrices and their effect on the analysis variables are important aspects of any 3D-Var or 4D-Var data assimilation system, as these matrices are a vehicle for spreading the innovation vectors spatially and to those model variables not explicitly used in the observation operators. It is generally recognized that the shortcomings in the existing operational systems result mainly from deficiencies in specifying the background error covariances, which are modeled through time invariant structure functions or the use of recursive filters. In this work, using the mesoscale MM5 model, we propose a simple method to approximate the background error covariance matrices and compute their effect on the innovation vectors. Using an ensemble of perturbations as a representative sample for the background errors, we approximate the actual (unknown) covariance matrix with the ensemble covariance matrix:

$$\mathbf{B} = \frac{\mathbf{V}\,\mathbf{V}^T}{N} \,, \tag{1.1}$$

where $\mathbf{V}$ is a matrix whose columns are N selected realizations of the background field errors. In our 4D-Var implementation, the background field always coincides with the model first-guess field, and it is chosen either as an analysis field or the output from a short-range forecast ending at the starting point of the assimilation time window. The assumption in eq. (1.1) is that N is large enough so that the background error realizations sampled in $\mathbf{V}$ form a representative, statistically invariant population. Based on the success of ensemble forecasting based on singular vectors (Mureau *et al.*, 1993, Molteni *et al.*, 1996) and on breeding modes (Toth and Kalnay, 1993, 1997), we assume that a handful of perturbations of any of these categories (less than 15) are enough to represent most aspects of the background error statistics relevant to data assimilation. The dimension of $\mathbf{V}$ is M×N, where M is the dimension of the state vector ($\approx 10^5$ - 10^7 for realistic atmospheric prediction models). By construction, $\mathbf{B}$ is a symmetric M×M matrix. We use the Lanczos algorithm to compute the largest eigenvalues and eigenvectors of $\mathbf{B}$. The implementation of that algorithm requires from the user a subroutine that computes the vector $\mathbf{z} = \mathbf{B}\mathbf{y}$, for any given $\mathbf{y}$. Because (1) can be rewritten as

$$\mathbf{B} = \frac{1}{N}\sum_{n}^{N} \mathbf{x}_n\,\mathbf{x}_n^T \,, \tag{1.2}$$

where $\{\mathbf{x}_n\}$ are the vectors of the background error realizations, implementing the subroutine $\mathbf{z} = \mathbf{B}^{-1}\mathbf{y}$ costs very little computer memory and disk space provided that one performs the multiplications within a pre-defined correlation distance.

A reasonable approximation for $\mathbf{B}$ is given by

$$\mathbf{B}_{approx} = \Phi\,\Sigma\,\Phi^T \,, \tag{1.3}$$

where Φ is an $M \times L$ matrix whose columns are the first L recovered (orthonormal) eigenvectors of $\mathbf{B}$ with non-zero eigenvalues. Σ is a diagonal matrix of eigenvalues of $\mathbf{B}$. Clearly, if all the eigenvalues of $\mathbf{B}$ are non-zero, and if one uses all the eigenvectors and eigenvalues, then (1.3) is the exact representation of $\mathbf{B}$ in terms of its orthonormal eigenvectors. In practice, a truncation to L eigenvectors in (1.3) will be good enough if these eigenvectors explain most of the variance of $\mathbf{B}$.

To avoid the use of the inverse of $\mathbf{B}$ in the minimization problem, we adopt the incremental preconditioned approach (see Courtier 1997). The square root of $\mathbf{B}$, needed in that approach, is computed as

$$\mathbf{B}^{1/2}_{approx} = \Phi\,\Sigma^{1/2}\,\Phi^T \,. \tag{1.4}$$

We note that, because $\mathbf{B}$ is a semi-positive definite matrix, its eigenvalues are non-negative. The success of the method proposed here thus rely on a good choice for the ensemble members and the

*Corresponding author address: Dr. M.S.F.V. De Pondeca, Department of Meteorology, The Florida State University, Tallahassee, FL 32306-4520. E-mail: pondeca@met.fsu.edu

number of eigenvectors used to rewrite the ensemble covariance matrix. The advantage of the method is two fold: (1) the covariances are obtained naturally through vector inner products, and (2) the storage of real-size matrices in the computer memory is avoided. Furthermore, our method also allows one to use a combination of perturbations of different categories (singular vectors, breeding modes, multiple analysis perturbations, etc.) to compute a single background error covariance matrix. It seems reasonable to assume that an adequate sampling of the space of possible analysis errors is attainable. If this is the case, then the main shortcoming of our approach is in the approximation expressed through equation (1.3).

2 Results and conclusions

Fig. 1 demonstrates the physical nature of the eigenvectors of **B** defined by eq. (1.1). Shown is the u-component of the wind field for the third eigenvector of **B** at the vertical level σ=0.355 $\approx$ 412 hPa for 0000 UTC 22 February 1998. The structures used in eq. (1.1) are 12 hour ensemble forecasts, whereby the ensemble initial conditions are based on the singular vectors of the MM5 model. The analysis of the sea level pressure for 0000 UTC 22 February 1998 (not shown) is characterized by a strong mid-Pacific cyclone and another developing cyclone adjacent to the US west coast. The structures seen in Fig. 1 seem to adequately capture the auto-correlations in the u-field. Preliminary results for mesoscale 4D-Var data assimilation will be presented at the conference.

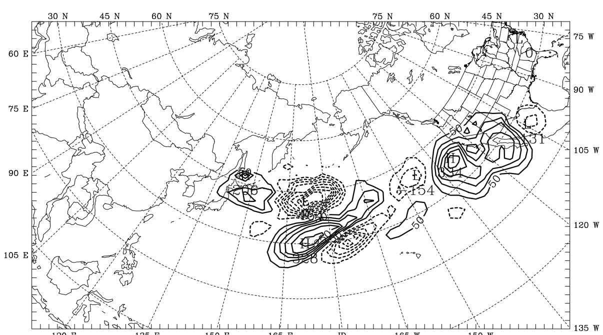

Figure 1: The u-component of the wind field for the third eigenvector of **B** at aproximately 412hPa. The time is 0000UTC 22 February 1998.

References

Courtier, P., 1997: Dual formulation of four dimensional variational assimilation. *Q. J. R. Meteorol. Soc.*, **123**, 2449-2461.

Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73-119.

Mureau, F., F. Molteni, and T. N. Palmer, 1993: Ensemble prediction using dynamically conditioned perturbations. *Quart. J. Roy. Meteor. Soc.*, **119**, 299-323.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317-2330.

Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297-3319.