

Adaptive ensemble size reduction

By Uzunoglu B.^{1*}, Navon, I.M^{2†} and Zupanski, M^{3‡},

¹*School of Computational Science and Information Technology, Florida State University, Tallahassee FL, 32306-4120;*

²*Department of Mathematics and School of Computational Science and Information Technology Florida State University Tallahassee, FL 32306-4120;*

³*Cooperative Institute for Research in the Atmosphere, Colorado State University, Foothill Campus, Fort Collins, CO 80523-1375*

25 August 2005

ABSTRACT

The ensemble size in sequential atmospheric ensemble based data assimilation using the Heikes and Randall (1995a,b) global shallow-water model is reduced by projecting the ensemble on a limited number of its leading EOFs. The ensemble size is determined by retaining the modes containing the main directions of variability of the system (most energetic modes of the flow). The efficiency of this approach for adaptively updating the ensemble size in the Maximum Likelihood Ensemble Filter (MLEF) by Zupanski (2005), Zupanski et al. (2005) used for ensemble data assimilation is assessed for different fractions of variability conserved and compared in terms of rms and similarity index error indicators with the full ensemble run. An illustration of the feasibility and effectiveness of the method is presented in framework of twin experiments for the above shallow water model. A reduction of up to a factor of four in the number of members of the ensemble was obtained, yielding comparable ensemble data assimilation (ENSDA) results with the full ensemble run. This novel approach results in sizable computational resource economy for general ensemble data assimilation methods.

1 Introduction

Practical implementation of the Kalman filter scheme is well known to be computationally very expensive. To alleviate this issue two different approaches have been adopted in atmospheric and ocean data assimilation community.

In general, physics of the system can be nonlinear, non-Gaussian, non-stationary therefore in most practical situations, optimal nonlinear filters cannot be applied. Suboptimal approximations can be made to nonlinear filters. Extended Kalman filter, multiple model filters as well as particle

* Uzunoglu, B.

e-mail: uzunoglu@csit.fsu.edu

† Navon, I.M

e-mail: navon@csit.fsu.edu

‡ Zupanski, M.

e-mail: ZupanskiM@cira.colostate.edu

filters are some of the filters in this category. Optimal finite-dimensional Bayesian state estimation in linear-Gaussian case, is the Kalman filter.

A first approach that employs sensible approximations is referred to as suboptimal scheme for instance the one considered by (Todling and Cohn(1994) and Cohn and Todling(1996)). Since error growth in regions of baroclinic or barotropic instability can be explained by leading singular vectors see (Farrell(1989), Moore and Farrell(1993)) a number of suboptimal schemes has emerged based on partial singular value decomposition which utilize a partial singular decomposition (SVD) of the tangent linear dynamics between consecutive observation times and where the tangent linear model is approximated by the leading part of its singular value decomposition. Such is the suboptimal scheme proposed by Cohn and Todling (1996) which generalizes the partial singular value decomposition filter to take more fully into account the dependence of the analysis error covariance upon the observations (PEF). All these schemes can be viewed as low rank approximations of the full Kalman filter (KF). An extensive summary of these suboptimal schemes is provided in Todling et al. (1998) including the partial eigendecomposition filter (PEF) and partial singular-value decomposition retrospective analysis (PSRA/PSRA2) where they employed singular values/vectors of the propagator or eigenvalues/vectors of the error covariances.

A similar approach to the above suboptimal schemes using modified EOF's (MEOF) in a reduced state KF for carrying out data assimilation via a reduced state Kalman filter was illustrated by Cane et al. (1996) which retained only 93 MEOF's while keeping 99 % of the explained variance. They obtained very satisfactory results pointing out that even retaining a much lower number of MEOF's would have achieved practically the same results. Recently Hoteit and Pham.(2004) extended some of these ideas to reduced-state space in adaptively reduced-order extended Kalman filter. See also Fukumori and Malanotte-Rizzoli (1995).

The work of Farrell and Ioannou (2001) also provides for a reduced order Kalman filter based on balanced truncation which was applied to a time dependent Lyapunov unstable quasi-geostrophic model of the forecast tangent linear error system reducing the autonomous forecast error model of order 400 to order 60. Similar ideas were put forward in the work by Heemink et al. (1997) and Verlaan and Heemink (1995,1997).

The second approach aimed at alleviating the computational cost of KF is the ensemble Kalman filter (EnKF) put forward by Evensen (1994). With the advent of ensemble Kalman filter introduced by Evensen (1994), Evensen (2003) the issue of choosing the relevant number of ensemble members for an adequate data assimilation performance is still open. However we know that the ensemble converges like the square root of the number of its members.

Heemink et al. (2001) proposed to combine the EnKF with the reduced-rank approach to reduce the statistical error of the ensemble filter. This is known as variance reduction, referring to the variance of the statistical error of the ensemble approach. Their ensemble filter algorithm consists of two parts: q ensembles in the direction of the q leading eigenvalues of the covariance matrix and N randomly chosen ensembles. In the algorithm, only the projection of the random ensemble members orthogonal to the first ensemble members plus the new random perturbations are used to obtain the filter gain and the state estimate. They call this a partially orthogonal ensemble Kalman filter (POEnKF).

The only errors that can be reduced by the observations are those lying in the manifold spanned by the approximate low rank error covariance matrices. Bishop et al.(2001) introduced a suboptimal

Kalman filter called the ensemble transform Kalman filter by transforming ensemble perturbations into orthonormal vectors and by attaching variances to each of the direction vectors to describe error covariance within the vector subspace of the ensemble perturbations. ETKF became a wide spread tool for using it in adaptive observations and targeting experiments at several operational weather prediction centers (Bishop et al. (2001), Wang and Bishop (2003) and Etherthon and Bishop (2004))

The present paper aims at introducing a method for adaptively reducing the KF/MLEF ensemble size in data assimilation using a Karhounen-Loeve expansion using the full ensemble to generate specific model evaluations (snapshots) from which we extract patterns (EOFs). It is important to stress that each snapshot is an ensemble member. In past approaches, similar concepts were introduced, a projection on leading SVD was used in reduced rank suboptimal Kalman filters by the aforementioned research works in data assimilation.

However to the best knowledge of the authors the present work is the first application for adaptively reducing ensemble size based on Proper Orthogonal Decomposition (POD).For recent reviews of POD method see Rowley et al. (2003) and Rowley (2005).

Implications of the size reduction methodology to the enrichment of ensembles are also addressed since in some instances we may want more ensembles(due to model errors, system complexity , etc.)

Options for ensemble enrichment when the reduction in ensemble size does not go in parallel with the decrease in RMS are also considered. One option is to rerun the procedure at the previous cycle with an increased retained variance percentage criteria which results in a smaller reduction in ensemble size and similarity index with corresponding increase in computational time.

Another option considered is to enrich the number of snapshots (ensemble members) with a number of randomized and initialized ensembles. All these options will be considered in a follow-up work.

Additional possibilities exist to extend the presented ensemble reduction methodology. One way would be to utilize it in initiation of small size regional ensembles from larger number of global ensembles. Similarly, the same ensemble reduction methodology can be used in a multi-model ensemble context to reduce the ensemble size from other sources to a desired small ensemble size. The various possibilities for ensemble enrichment come only as a consequence of the described ensemble reduction method. All these options have important practical implications.

Then our procedure can serve as an adaptive method for both ensemble size inflation and reduction however our main aim in the present work is to reduce rather than inflate the ensemble size without loss of accuracy.These options will be investigated in a follow-up work.(See also work of Ravindran(2002) and Fahl(2000))

POD has been applied to fluid problems by Sirovich (1987) using an SVD basis and other researchers Berkooz *et al* (1993) to understand the important dynamical features or coherent structures seen in fluid flows and has since then been applied in all domains of science.

It has since become the subject of multiple research efforts in fluid dynamics, chemistry , turbulence, hydrological processes and image processing (see Vermeulen et al. (2004)).

An ensemble based data assimilation Zupanski (2005), named Maximum Likelihood Ensemble Filter(MLEF), is employed in this study. The analysis solution maximizes the likelihood of the posterior probability distribution, obtained by minimization of a cost function that depends on a general nonlinear observation operator.

The MLEF employs the ETKF transformation in Hessian preconditioning. Therefore, one could view the MLEF as a maximum likelihood approach to ETKF. In fact, the size reduction approach is facilitated using that transform.

The paper does not address the issue of adaptive observations. It is organized as follows. In the second section the governing equations of the spherical shallow-water model employed and their discretization are briefly discussed. In the third section, Proper Orthogonal Decomposition (POD) method is presented. The POD method describes the system behavior as an attractor which is a point of evolution for the state space in a subspace of higher dimensions. Numerical results in a twin experiment set-up are discussed in section 4 and the paper concludes with a summary and conclusions section.

2 Governing Equations and Discretization

The shallow water equations have been used customarily to develop new numerical methods for atmospheric models because they exhibit the same wave behavior as more complex baroclinic equations governing the motion of the atmosphere.

The nonlinear shallow-water equations on the sphere assume here the form given in Ringler and Randall (2002a)

$$\frac{\partial}{\partial t} \mathbf{V} = - (f + \zeta) \mathbf{k} \times \mathbf{V} - \nabla[K + g(h + h_s)], \quad (1)$$

$$\frac{\partial}{\partial t} \mathbf{h} = - \nabla \cdot (h \mathbf{V}). \quad (2)$$

Here \mathbf{V} is the horizontal velocity vector, ζ is the relative vorticity, f is the Coriolis parameter, g is the gravitational constant, K is the kinetic energy, h is the fluid depth, and h_s is the surface topography. Alternatively, we can take the curl and divergence of eqn. 1 and 2 to generate equations for vorticity and divergence:

$$\frac{\partial \eta}{\partial t} = - \nabla \cdot (\eta \mathbf{V}), \quad (3)$$

$$\frac{\partial \delta}{\partial t} = - \mathbf{k} \cdot \nabla \times (\eta \mathbf{V}) - \nabla^2 [K + g(h + h_s)]. \quad (4)$$

Here $\eta \equiv f + \zeta$ and $\delta \equiv \nabla \cdot \mathbf{V}$ is the divergence. The vector momentum formulation and the vorticity-divergence formulation are equivalent given the appropriate boundary conditions. The discrete form of the scalar formulation bears a resemblance to the continuous form.

$$\frac{\partial \eta_i}{\partial t} = - [\nabla \cdot \bar{\eta} \mathbf{V}_c]_i, \quad (5)$$

$$\frac{\partial \delta_i}{\partial t} = [\mathbf{k} \nabla \times (\bar{\eta} \mathbf{V}_c)]_i - \{\nabla^2 [K + g(h + h_s)]\}_i. \quad (6)$$

The subscript c denotes quantities defined at the cell corners, while the subscript i denotes quantities defined at grid-cell centers. An overbar denotes quantities that are averaged from cell centers to the cell corners. The prognostic variables vorticity, divergence, and mass are all defined at the grid-cell centers; this is the Z grid introduced by Ringler and Randall (1994).

While the continuous equations allow an infinite number of quantities to be conserved, this is not possible within the discrete system. The form of total energy that is conserved in this discrete system Ringler and Randall (2002b) is

$$\sum_{i=0}^n A_i \{h_i [K_i + g(h_s + \frac{1}{2} h_i)]\} \quad (7)$$

which is consistent with the continuous shallow-water equations, where g is the gravitational

constant, K_i is the kinetic energy, h_i is the fluid depth, h_s is the surface topography and A_i is the area for each discrete cell.

3 Empirical orthogonal functions and proper orthogonal decomposition

The underlying problem we address is that of identifying the ϕ sought structure in a random vector field. It is then natural to look for functions ϕ for which the inner product (z_i, ϕ) exists, i.e. ϕ must be $L^2(\Omega)$. Given an ensemble of random vector fields z_i , a function of ϕ which has a structure typical of the members of the ensemble is sought. In order to resolve this problem, we can try to project the ensemble on ϕ , i.e. (z_i, ϕ) so that ϕ is nearly parallel as possible. The aim is to maximize (z_i, ϕ) while removing the amplitude by normalizing it. In order to include the statistics, we must maximize the expression

$$\frac{(z_i, \phi)}{(\phi, \phi)^{\frac{1}{2}}} \quad (8)$$

in some average sense. We consider the initial ensemble members each of which are viewed as "snapshots".

$$\mathbf{S} = \{z^i : 1 \leq i \leq N\} \quad (9)$$

are solutions of N different ensembles and a function of $\phi \in L^2(\Omega)$ that gives the best representation of S in the sense that it maximizes

$$\frac{1}{N} \sum_{i=1}^N \frac{|(z^i, \phi)|^2}{(\phi, \phi)} \quad (10)$$

A proper orthogonal decomposition method (POD) computes these coherent spatial structures directly and in an optimal sense. The structures computed are optimal for a given data set (see Sirovich (1987) and Berkooz *et al* (1993)). The proper orthogonal decomposition method is a well known analysis technique with the original concept traced to Pearson (1901). Several different names including principal component analysis, Karhounen-Loève decomposition and total-least-squares estimation refer to the same procedure.

We are therefore trying to maximize

$$\frac{1}{N} \sum_{i=1}^N \frac{|(z^i, \phi)|^2}{(\phi, \phi)} = \frac{(\mathbf{K}\phi, \phi)}{(\phi, \phi)} = \lambda \quad (11)$$

where N stands for the number of ensemble solutions. This leads to

$$(\mathbf{K}\phi, \phi) = \lambda(\phi, \phi) \quad (12)$$

The corresponding eigenvalue problem is to find the eigenvalues and eigenfunctions of a symmetric $M \times M$ matrix defined by $\mathbf{K} = \frac{1}{n} \langle z, z^T \rangle$ or $\int_{\Omega} \mathbf{K} d\mathbf{x}$ in $L^2(\Omega)$. In other words one seeks a

function which has the largest mean square projection on the set S . Here we consider the snapshot vectors \mathbf{z}^i and determine the empirical eigenfunctions ϕ as a linear combination of the snapshots given by

$$\phi = \sum_{i=1}^N \alpha \mathbf{z}^i \quad (13)$$

such that equation (12) holds.

In this case, the practical approach for calculating the correlation function is not to determine the $M \times M$ correlation matrix but to use the dual approach on the N snapshots which are individual ensembles in this study (Sirovich (1987)). This method is also known as *sample space setting* Preisendofer (1988). The aim is to reduce the space resolution size to ensemble space size by substituting equation (13) to equation (12). The corresponding eigenvalue problem is then to find the eigenvalues and eigenfunctions of a symmetric $N \times N$ correlation matrix defined by $\mathbf{C} = \frac{1}{n} \langle \mathbf{z}^T, \mathbf{z} \rangle$ or $\int_{\Omega} \mathbf{C} d\mathbf{x}$ in $L^2(\Omega)$

$$(\mathbf{C}\boldsymbol{\alpha}, \boldsymbol{\alpha}) = \lambda(\boldsymbol{\alpha}, \boldsymbol{\alpha}) \quad (14)$$

The POD basis of this eigenvalue problem is orthonormal, i.e., $\phi_i^T \phi_j = 0$ for $i \neq j$ and $\phi_i^T \phi_i = 1$. Since the trace of the matrix \mathbf{C} can represent the total variance (averaged energy retained if \mathbf{C} is redefined for energy) in the snapshots, the variance corresponding to the ensemble data is the sum of the eigenvalues of the correlation matrix, in the sense that

$$E = \sum_{i=1} \lambda_i. \quad (15)$$

A variance percentage can be assigned to each eigenvector based on the eigenvalue associated with the eigenvector, such that

$$E_k = \frac{\lambda_k}{E} \quad (16)$$

Under the assumption that the eigenvalues are arranged in descending order from the largest to the smallest, then we have an ordering of the eigenvectors from largest variance to lowest variance. This will provide us the criteria to define the ensemble size adaptively based on percentage of variance retained.

Equivalently, this can be thought of as damped singular value decomposition (SVD) solution in which filter factors of zero are used for basis vectors associated with smaller singular values. The most common approach to regularization of numerically rank-deficient problems is to consider the given matrix \mathbf{C} as a noisy representation of numerically rank-deficient matrix, and to replace \mathbf{C} by a matrix that is close to \mathbf{C} and mathematically rank deficient Hansen (1997).

Let $\mathbf{C} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$ denote the SVD of \mathbf{C} and λ_i the i -th largest eigenvalue of \mathbf{C} . then $\sigma_i^2 = n \lambda_i$ for $i = 1, \dots, n$, where σ_i the i -th singular value of \mathbf{C} Volkwein(2004). Using this relation, we can use related Truncated SVD (TSVD) to POD. As a result a \mathbf{C} matrix that does not appeal to energy

criteria can be used to truncate the smaller singular values Hansen(1997). It should be noted that $\phi = U$

The quality of a given cycle ensemble relative to the previous cycle can be judged by the extent to which it preserves the value of the direction of perturbations from previous cycle. This can be used to assess the quality of reduction compared to full-sized ensemble.

To assess the quality of a C matrix in addition to RMS criteria that is produced at the end of each cycle, a similarity matrix (see Buizza (1994), Li *et al* (2005) and Serban *et al* (2005)) can be constructed whose entries

$$m_{ij} = \langle \phi, \tilde{\phi} \rangle^2 \quad (17)$$

are the squared scalar products of the i -th SV of the full model and the j -th SV of the reduced model. These entries represent the amount of variance of the i -th full model SV that can be explained by the j -th reduced model SV. Here the reduced model will be the current cycle while the full model will be the previous cycle. Based on the matrix m_{ij} , a similarity index can be defined as

$$SM = \frac{1}{N} \sum_{i,j=1}^N m_{ij} \quad (18)$$

SM index take values between 0 and 1 and for identical models, it will equal unity.

The RMS may not necessarily decrease with the ensemble size for a given variance percentage criteria defined earlier even though this is not likely given the numerical results. The algorithm must also include ensemble enrichment if the reduction in ensemble size does not go in parallel with the decrease in RMS. One option is to rerun the cycle with an increased variance percentage criteria which might result in a smaller reduction in ensemble size and similarity index with an increase in computational time. Another way will be to increase the correlation length of the C matrix along with a new update of the system by ensemble initialization (see Zupanski et al. 2005). This method can serve as adaptive method for both ensemble size inflation and reduction however our main aim should be to reduce rather than inflate the ensemble size without loss of accuracy. C matrix of the full previous run can be enriched by using correlated ensemble initialization (see Zupanski et al. 2005) if an increase in RMS is observed.

3.1 Adaptive ensemble size reduction and preconditioning algorithm

- **Step1:** Generate the ensemble z^i .
 - **Step2:** Generate the covariance matrix $C = \frac{1}{n} \langle z^{iT}, z^i \rangle$ or the correlation energy matrix and define energy by the new variables $z = \sum_{i=0}^n A_i \{h_i [K_i + g(h_s + \frac{1}{2}h_i)]\}$, and $\int_{\Omega} C dx$.
 - **Step3:** Solve the eigenvalue problem $C \alpha^{(k)} = \lambda_k \alpha^{(k)}$.
 - **Step4:** Calculate variance $E = \sum_{i=1} \lambda_i$.
 - **Step5:** Decide on the amount of variance retained $E_k = \frac{\lambda_k}{E}$ reduce the ensemble size accordingly based on the dominant eigenvalues retained.
 - **Step6:** Check the performance of ensemble reduction $SM = \frac{1}{N} \sum_{i,j=1}^N m_{ij}$ if SM close to unity and if RMS decreases then continue reduction else,
- Step 6a: Rerun the procedure at the previous cycle with an increased retained variance percentage

criteria else,

Step 6b: Reset the new ensemble size from the previous iterations by using extended Proper Orthogonal Decomposition else,

Step 6c: Use \mathbf{z}^i matrix from previous cycle to initialize the next cycle by \mathbf{z}^i by running ensemble data initialization (See Zupanski et al. 2005) for one cycle.

• **Step7**: Update the preconditioner $\mathbf{x} - \mathbf{x}_{k-1} = \mathbf{P}_f(\mathbf{I} + \mathbf{C})^{-T/2} \zeta$ Zupanski (2005), Zupanski et al. 2005).

4 Experiments and Results

The initial conditions for the global shallow-water equation models employ one of the test-cases in Williamson et al. (1992) corresponding to a geostrophically-balanced zonal flow over an isolated conical mountain with strong nonlinearity occurring in the vicinity of the mountain. This set-up is characterized by the excitation of Rossby and gravity waves. The initial zonal flow is 20m/s, and the mountain is centered at 30°N,90°W, with the height of 2000m.

All the experiments use as initial full-ensemble one that consists of 1000 ensemble members. A smaller ensemble size could have been used and this number was chosen for comparison sake.

An unstructured finite-difference mesh was employed whose total number of points-variables (degrees of freedom) is 12802. It includes about 2600 height points, and about 5100 points for each wind component of freedom. The assimilation is performed over an interval of 120 hours which corresponds to 40 cycles of data assimilation, each comprised of a 3-hour interval. The initial conditions for the experimental run are as defined above, but shifted to -6 hours in order to create a set of unbalanced initial conditions.

The observations are created by adding random perturbations from a normal (Gaussian) distribution $N(0,R)$ to a model forecast, which we refer to as the truth. This implies a perfect model assumption, since the same model is used in assimilation. Although the model equations formally predict the velocity potential and the stream function, more conventional wind observations are created, and later assimilated. The observation error covariance R is assumed to be diagonal, i.e. no correlation between observations is assumed. The observation error chosen for the height is 5 m, and for the wind is 0.5ms. There are 1025 observations defined in each analysis cycle, uniformly distributed around the globe. These observations consist of approximately 500 height observations, and 500 wind observations. Since the two wind components (east-west and north-south) are collocated, there is approximately 250 observation points for each wind component. The observations are assimilated every six hours.

Numerical experiments are conducted by comparing RMS height analysis (as well as u and v RMS ,not shown) as well as an error measure (the similarity index SM) for the two set of ensembles tested, namely the adaptively reduced and the full size ensembles, respectively.

Different numerical experiments were conducted differing only in the stage at which the adaptive ensemble reduction is being started, i.e. after either 1, 20, 30 or 40 cycles of the full ensemble run. Based on recently acquired experience in application of reduced order 4-D Var (Robert et al. 2005) we expect a result pointing to the fact that choosing the snapshots from a full assimilation time interval trajectory would yield better results for the adaptively reduced ensemble than if it were to start with a non-assimilated trajectory.

The algorithm used allows the adaptive reduction algorithm to be started from any cycle of the full ensemble run. The different numerical tests of starting the adaptive reduction algorithm at different stages(or number of cycles) of a full ensemble run provide a rigorous test of this procedure and allow a more in-depth understanding and validation of the procedure.

The information displayed in the following figures is comprised of the following information:

a) Cycle number of full ensemble run where the adaptively reduced ensemble procedure is initiated.

b) Percentage of variance explained (two cases of 95 % and 99 % explained variance were chosen for the sake of simplicity).

c) Reduction in the number of ensemble members obtained as a result of applying the adaptive reduced ensemble algorithm.

d) The numerical results then display either the RMS of the full ensemble versus that of the adaptively reduced ensemble as a function of the number of cycles, or the similarity index criterion (explained above in the previous section), for the same two ensemble sets.

Fig 1. displays the isoline global plot of the height analysis RMS error for both the full ensemble (1000 members) as well as for the adaptively reduced ensemble between cycles 12 and 13 of the MLEF filter whose size decreases from 380 to 280 members for explained variance of 95 %. The global shallow water equations model was run for initial conditions consisting of test case 5 of Williamson et al. (1992) describing zonal flow over an isolated conical mountain.

Fig 2 displays the height analysis for same model and initial conditions running MLEF full ensemble for cycle 13 (1000 members) along with the adaptively reduced ensemble (280 members) for same MLEF filter cycle and for explained variance of 95 %.

Fig. 3 displays the isoline height analysis RMS error difference fields of the adaptively reduced ensemble minus full ensemble for the same abovementioned initial conditions at 3 consecutive MLEF filter cycles (Cycle 12-cycle 14). The results were scaled to appear 10 times larger for illustration purposes. The results are displayed on the global earth domain and the difference height analysis RMS errors are of the order of a maximum of 2 meters. This for explained variance of 95 %.

Fig 4 shows in three vertical columns, the RMS height analysis, the similarity index criterion and the evolution of ensemble size for an explained variance of 99% , respectively.

The first column of the figure shows that the height RMS for case (1-40) i.e where adaptively reduced ensemble starts at cycle 1, yields the worst results, whereas as we increase the cycle number where adaptively reduced ensemble is initiated, the RMS error of the adaptively reduced ensemble dramatically decreases and is very similar to the one obtained with a full ensemble run.

Remark: We could improve the height RMS analysis results for the reduced ensemble by increasing the percentage of variance explained as a criterion for the adaptive ensemble reduction. (for instance going to an explained variance of 99.5% versus one of 99%).

The results for 10-40, 20-40 and 30-40 i.e. for the adaptively reduced ensemble starting at cycle 10, 20 or 30 of the full ensemble run respectively show dramatic improvement in the RMS error compared with that of the full ensemble run as the adaptively reduced ensemble procedure is initiated at later stages of the MLEF full ensemble data assimilation.

The second column displays the similarity index measure between 2 subsequent cycles of both the full ensemble run and the adaptively reduced runs, respectively. When the adaptive reduced ensemble is activated at a given cycle of the full ensemble run we observe a dip in the similarity

index plot that recovers after 3-4 cycles. This dip occurs at exactly the cycle number where the adaptive reduced ensemble is first introduced.

The similarity index provides a criterion related to the performance of the full ensemble run versus the adaptive ensemble run.

Finally the third column in the composite figure displays the corresponding reduction in the ensemble size versus the cycles (time) when the adaptive reduced ensemble procedure is initiated at different cycle numbers of the full MLEF ensemble run.

The reduced ensemble size is changing in the experiments 1-40, 10-40, etc. namely the ensemble size tends to some constant value which assumes a lower value as the stage at which adaptive reduction is initiated is occurring later in the data assimilation cycle. The limiting ensemble size in each of the above four experiments is 700, 480, 424 and 416 respectively.

For a given percentage of variance explained the adaptively reduced ensemble contains pertinent information on the variability of the full ensemble. As shown also in Fig 5 some ensemble members do not contribute information during later cycles of full ensemble MFLEF filter data assimilation or in other words that the variability of the system can be described by a lower dimensional space.

Fig 5 displays the magnitude of the eigenvalues of the matrix \mathbf{C} as a function of the number of spectrum modes for the different cycles in the full MLEF ensemble run and the adaptively reduced runs scaled by the largest eigenvalue of \mathbf{C} at each consecutive cycle. This for explained variance of 99 %.

As the MLEF assimilation filter cycle number increases, the scaled magnitude of the eigenvalues of \mathbf{C} decreases respectively - exhibiting a rank reduction of matrix \mathbf{C} as the assimilation process proceeds. This justifies the use of the adaptively reduced ensemble procedure since as number of cycles in full ensemble run increases- an increasing number of modes in spectrum of \mathbf{C} tend to zero, i.e. there are more ensemble members that do not contribute information to the MLEF filter assimilation.

The computational economy realized using adaptively reduced ensemble for cases where the RMS height analysis difference error is small compared to that of the full ensemble run exhibits a factor of 2-3 economy in CPU time(corresponding roughly with a factor of 2-4 reduction of the number of ensemble members for the adaptively reduced ensemble).

In all the experiments with both full ensemble and reduced ensemble runs the number of observations is identical.

5 Summary and Conclusions

A novel adaptive methodology for reducing ensemble size by projecting the ensemble on a limited number of its leading EOFs was successfully applied in the framework of MLEF for data assimilation.

The methodology is general enough to be applied and extended to general ensemble filter data assimilation techniques such as EnKF and its various flavors.

A test case was run in the framework of twin experiments for a global shallow water equations model (Randall et al. 1995) using a typical relevant test case taken from the suite of Williamson (1992) (the mountain test case) test cases.

The idea used here originates in reduced order modeling theory as well as in regularization theory

(Tikhonov regularization in particular) and allows a sizable economy in application of ensemble methods for data assimilation, resulting typically in a reduction of up to a factor of four in the number of members of ensemble required for successful implementation.

Numerical experiments conducted in above MLEF framework of twin experiments validate that the adaptive reduction in ensemble size, assessed for different fractions of variability conserved (in the range of 95 % to 99.9 %) and compared in terms of RMS and similarity index error indicators as well as in terms of impact of resulting assimilated geopotential and velocity fields only marginally affects results obtained when compared to the results of the full ensemble run. This was illustrated by comparing differences in geopotential and velocity fields between the full and adaptively reduced ensembles. This is further explained by considering the spectrum of eigenvalues of the matrix \mathbf{C} in the MLEF ensemble filter.

A discussion was presented related to the impact of using the adaptively reduced ensemble and generalization of our results to sequential ensemble data assimilation in general.

The goal is not necessarily to reduce, or enlarge the ensemble size, rather to find an optimal utilization of ensembles, given the computing limitations and the complexity of the problem. In that sense, the reduction is an important part of the whole picture. Once we know why, and how to do that, we can come up with enrichment strategy (e.g., resampling), if deemed relevant for the modeling-observation system we use.

Coremmelin and Majda(2004) pointed out to some limitations of EOFs in particular for systems that exhibit sudden transitions between different states (i.e., bursting behavior). While our test case and in general short period atmospheric circulation do not have such behavior, this limitation should be borne in mind and other optimal bases such as principal interaction patterns (PIPs) should then be considered.

Future research will be aimed at applying the adaptive ensemble size reduction to non-Gaussian ensembles as well as extending the present methodology to typical EnKF data assimilation with realistic models.

Aspects of increasing the adaptively reduced ensemble size when dictated by criteria of increase in RMS error will be investigated in a follow-up work by returning to a previous cycle and adding freshly randomized ensemble members.

For ensemble size inflation, the computed eigensets derived from snapshots (members) of the ensemble depend on the initial and final portions of the ensemble set under consideration. For this purpose the POD method can be extended as described by Glezer *et al* (1989), Uzunoglu *et al* (2002) and Tan *et al* (2001). For statistically stationary data, the extended POD(EPOD) is equivalent to the classical POD. For ensemble enrichment by ensemble size inflation, the reduced ensemble model can be inflated using the ensemble generated by the filter at previous cycles. The possibility to carry out hybrid runs of adaptively reduced ensemble followed by performing a few cycles with a full ensemble size at the end of the reduced ensemble assimilation will be considered in future research work.

A comparison of our results with results obtained by reduced-order 4-D Var data assimilation Hoteit *et al.*, (2003) and Robert *et al.* (2005) is also considered as a follow-up effort.

We emphasize that the adaptive reduced ensemble methodology (and MLEF, as well as ETKF) are well suited to handle efficiently large number of observations.

Sizable reduction in the computational load of ensemble simulations is possible by extending

this approach of adaptive ensemble size reduction to realistic context operational EnKF runs in 3-D .(See Houtekamer et al.2005).

Acknowledgments

The research of Prof. Michael Navon and Dr. Bahri Uzunoglu was sponsored by the National Science Foundation Collaboration in Mathematical Geosciences , Grant ATM-0327818. Dr. Milija Zupanski's work was supported by the National Science Foundation Collaboration in Mathematical Geosciences , Grant ATM-0327651. Our gratitude is also extended to National Center for Atmospheric Research, which is sponsored by the National Science Foundation, for the computing time used in this research. The authors would also like to thank Prof. Gordon Erlebacher for his valuable discussions.

REFERENCES

- Berkooz, G., Holmes, P. and Lumley, J.L., 1993. The generalized inverse of a nonlinear quasi-geostrophic ocean circulation model. *Ann. Rev. Fluid Mech.* **25**, 539-575.
- Bishop, C.H., Etherton, B.J. and Majumdar, S.J., 2001. Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects. *Mon. Wea. Rev.* **129**, 420-436.
- Buizza, R. 1994. Sensitivity of optimal unstable structures. . *Q. J. R. Meteorol. Soc.* **120**, 429-451.
- Cane, M.A., Kaplan, A., Miller, R.N., Tang, B.Y., Hackert, E.C. and Busalacchi, A.J., 1996. Mapping tropical Pacific sea level: Data assimilation via a reduced state space Kalman filter. *J. Geophys. Res-Oceans* **101**, 22599-22617.
- Cohn, S.E. and Todling, R., 1996. Approximate data assimilation schemes for stable and unstable dynamics. *J. Meteorol. Soc Jpn* **74**, 63-75.
- Cohn, S.E. and Todling, R., 1995. Approximate Kalman filters for unstable dynamics. *Second Int. Symp. on Assimilation of Observations in Meteorology and Oceanography, Tokyo, Japan, WMO* 241-246.
- Crommelin, D.T. and Majda, A. J., 2004. Strategies for model reduction: Comparing different optimal bases. *J. Atmos. Sci.* **61**, 2206-2217.
- Evensen, G., 1994. Sequential data assimilation with a non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**(C5), 10143-10162.
- Evensen, G., 2003. The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics* **53**, 343-367.
- Etherton, B.J. and Bishop, C.H., 2004. Resilience of Hybrid Ensemble/3DVAR Analysis Schemes to Model Error and Ensemble Covariance Error. *Mon. Wea. Rev.* **132**, 1065-1080.
- Fahl, M. 2000. Trust-region methods for flow control based on reduced order modelling. *Doctoral Dissertation Universitat Trier*, 147pp.
- Farrell, B.F. 1989. Optimal Excitation of Baroclinic Waves. *J. Atmos. Sci* **46**, 1193-1206.
- Farrell, B.F. and Ioannou, P.J. 2001. Accurate Low-Dimensional Approximation of the Linear Dynamics of Fluid Flow. *J. Atmos. Sci* **58**, 2771-2789.
- Fukumori, I. and Malanotte-Rizzoli, P. 1995. An approximate Kalman Filter for ocean data assimilation- An example with an idealized Gulf-Stream Model. *J. Geophys. Res-Oceans* **100**, 6777-6793.

- Glezer, A. , Kadioglu, Z. and Pearlstein, A. 1989. Development of an extended proper orthogonal decomposition and its application to a time periodically forced plane mixing layer. *Phys. Fluids A* **1**, 1363-1373.
- Hansen, O., 1997. In: *Rank-Deficient and Discrete Ill-Posed Problems*, SIAM, Philadelphia.
- Heemink, A.W., Bolding, K. and Verlaan, M. 1997. Storm surge forecasting using Kalman filtering. *J. Meteorol. Soc. Jpn.* **75**, 305-318.
- Heemink, A.W., Verlaan, M. and Segers, J.A. 2001. Variance Reduced Ensemble Kalman Filtering. *Mon. Wea. Rev.* **129**, 1718-1728.
- Heikes, R. and Randall, D.A., 1994. Numerical Integration of the Shallow-Water Equations on a Twisted Icosahedral Grid. Part I: Basic Design and Results of Tests. *Mon. Wea. Rev.* **123**, 1862-1187
- Hoteit, I., Khol, A., Stammer, D. and Heimbach, P., 2003. A reduced order optimization strategy for four dimensional variational data assimilation. . *Proceedings of EGS-AGU-EUG Joint Assembly, 611 April 2003, Nice, France.*
- Hoteit, I. and Pham, D.T., 2004. An adaptively reduced-order extended Kalman filter for data assimilation in the tropical Pacific . *J. Marine Syst.* **45(3-4)**, 173-188
- Houtekamer, P.L., Mitchell, H.L., Pellerin, G., Buehner, M., Charron, M., Spacek, L. and Hansen ,M. 2005. Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations . *Mon. Wea. Rev.* **133**, 604-620
- Jazwinski, A.H., 1970. In: *Stochastic Processes and Filtering Theory*, Academic Press, New York.
- Li, Zhijin, Navon, I.M. and Hussaini, M.Y., 2005. Analysis of the singular vectors of the full-physics Florida State University Global Spectral Model. *Tellus* **57A**, 560-574.
- Moore, A.M. and Farrell, B.F., 1993. Rapid perturbation growth in spatially and temporally varying oceanic flows as determined by an adjoint method: application to the Gulf Stream. *J. Phys. Ocean.* **23**, 1682-1702.
- Pearson, K., 1901. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine* **2**, 559-572.
- Pham, D. T., 2001. Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Mon. Wea. Rev.* **129**, 1194-1207.
- Preisendorfer, R., 1988. In: *Principal component analysis in meteorology and oceanography*, Elsevier, Amsterdam.
- Ringler, D.A. and Randall, D.A., 1994. Geostrophic adjustment and the finite-difference shallow-water equations. *Mon. Wea. Rev.* **122**, 1371-1377.
- Ravindran, S.S. 2002. Adaptive reduced-order controllers for a thermal flow system using proper orthogonal decomposition . *SIAM J. Sci Comput.* **28**, 1924-1942.
- Robert, C., 2004. Developpement et comparaison de methodes d'assimilation de donnees de rang reduit dans un modele de circulation oceanique : application a l'ocean Pacifique Tropical. These de doctorat (21 Decembre 2004). *Universite Joseph-Fourier - Grenoble I.*
- Robert, C., Durbiano, S., Blayo, E., Verron, J., Blum, J. and Le Dimet, F.-X., 2005. A reduced -order strategy for 4-D Var Data assimilation. *J. Marine Syst.* , **57**, 70-82.
- Ringler, T.D. and Randall, D.A., 2002a. The ZM Grid: An alternative to the Z Grid. *Mon. Wea. Rev.* **129**, 1411-1422.
- Ringler, T.D and Randall, D.A., 2002b. A potential Enstrophy and Energy Conserving Numerical

- Scheme for Solution of the Shallow-Water Equations on a Geodesic Grid. *Mon. Wea. Rev.* **130**, 1397–1410.
- Rowley, C.W. Colonius, T., and Murray R.M. 2004. Model reduction for compressible flows using POD and Galerkin projection . *Physica D* **189**, 115–129.
- Rowley, C.W. 2005. Model reduction for fluids, using balanced proper orthogonal decomposition . *Int. J. of Bifurcat. Chaos* **15**, 997–1103.
- Serban, R., Homescu, C. and Petzold, L.R., 2005. The Effect of problem perturbations on nonlinear dynamical systems and their reduced order models. *Submitted to SIAM Jour. Sci. Stat. Comput.*
- Sirovich, L., 1987 Turbulence and the dynamics of coherent structures. Parts I-III. *Q. Appl. Math.*, **XLV**, No. 3, 561-590.
- Stratonovich, R.L., 1966. A new representation for stochastic integrals and equations. *SIAM J. Control* **4**, 362–371.
- Tan, M., Uzunoglu, B., Price, W.G. and Rogers, E. 2002. Reduced Models for Statistically Stationary and Non-stationary flows with control applications. *Proceedings of the Institute of Mechanical Engineers Part M* **216**, 95-102.
- Todling, R., Cohn, S.E. and Sivakumaran, N.S.1998. Suboptimal schemes for retrospective data assimilation based on the fixed-lag Kalman smoother. *Mon. Wea. Rev.* **126**, 2274-2286.
- Todling, R., and Cohn, S.E., 1994. Suboptimal schemes for atmospheric data assimilation based on the Kalman filter . *Mon. Wea. Rev.* **122**, 2530-2557.
- Uzunoglu, B., and Nair, P., 2001. Optimal Flow Control Framework For Adaptive Variable Fidelity-Reduced Modelling . *Global Flow Instability and Control Symposium, Crete, Greece, June 11-13 2001*
- Verlaan, M. and Heemink, A.W., 1995. Data assimilation schemes for non-linear shallow water flow models. *Proc. Second Int. Symp. on Assimilation of Observations, Tokyo, Japan, WMO, 247252.*
- Vermeulen, P.T.M. and Heemink, A.W. and te Stroet, C.B.M. 2004. Low Dimensional Modelling of Numerical Groundwater Flow. *Hydrol. Process.* **18**, 1487–1504.
- Verlaan, M. and Heemink, A.W., 1997. Tidal flow forecasting using reduced-rank square root filters. *Stochastic Hydro. Hydraul.* **11**, 349–368.
- Volkwein, S., 2004. Interpretation of proper orthogonal decomposition as singular value decomposition and HJB-based feedback design. *Proceedings of the Sixteenth International Symposium on Mathematical Theory of Networks and Systems (MTNS), Leuven, Belgium July 5-9,2004*
- Wang, C.H. and Bishop, C.H., 2003. A Comparison of Breeding and Ensemble Transform Kalman Filter Ensemble Forecast Schemes. *J. Atmos. Sci* **60**, 1140-1158.
- Williamson, D.L., Drake, J.B., Hack, J.J., Jakob, R. and Swarztrauber, P.N. 1992. A standard test set for numerical approximations to the shallow water equations in spherical geometry. *J. Comput. Phys.* **102**, 211–224.
- Zupanski, M., 2005. Maximum Likelihood Ensemble Filter: Theoretical Aspects. *Mon. Wea. Rev.* **133**, 1710-1726.
- Zupanski, M., Fletcher, S.J., Navon, I.M.,Uzunoglu, B., Heikes, R.P., Randall, D.A., Ringler, T.D.and Daescu, D.N. 2005. Initiation of ensemble data assimilation . *Submitted to Tellus.*

- **Figure 1** RMS error isolines of height field analysis scaled 10 times larger for plotting. Initial conditions using Williamson *et al.* (1992) test case 5 ,i.e. zonal flow over an isolated conical mountain. Full ensemble (left column) and reduced ensemble with 95 % variance retained(right column) for (a) Cycle 24; Full ensemble size, 1000 members (b) Cycle 24; Reduced ensemble size, 380 members and (c) Cycle 26; Full ensemble size, 1000 members and (d) Cycle 26; Reduced ensemble size 280 members, respectively.

- **Figure 2** Isolines of height field analysis of Williamson *et al.* (1992) test case 5 ,i.e. zonal flow over an isolated conical mountain for (a) full ensemble,comprised of 1000 members and (b) reduced ensemble, comprised of 280 members for cycle 26 of the MLEF data assimilation.

- **Figure 3** RMS error isolines of height analysis difference fields between reduced ensemble and full ensemble runs for cycle (a) 24 (b) 26 (c) 28 with retained variance of 95 %. The results are scaled 10 times larger for plotting. Initial conditions using Williamson *et al.* (1992) test case 5 ,i.e. zonal flow over an isolated conical mountain.

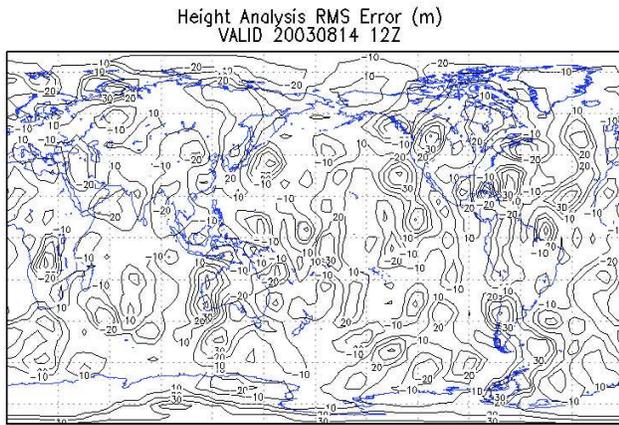
- **Figure 4** Column A illustrates the evolution of height field RMS error for full and reduced ensembles as a function of the MLEF cycle where adaptive ensemble reduction is initiated from 1-st , 10-th, 20-th and 30-th cycle of the full MLEF ensemble run consisting of 40 cycles respectively, for an explained variance of 99 %.

Column B illustrates the evolution of the similarity index for full and reduced ensembles where adaptive ensemble reduction is initiated from 1-st , 10-th, 20-th and 30-th cycle of the full ensemble MLEF cycle, respectively, for an explained variance of 99 %.

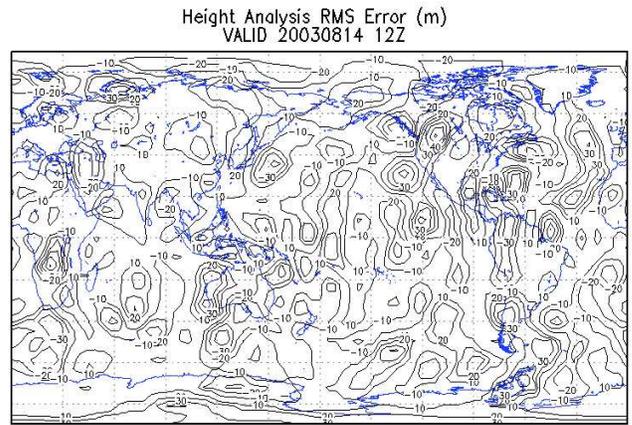
Column C illustrates the evolution of the ensemble size of the reduced ensembles where adaptive ensemble reduction is initiated from 1-st , 10-th, 20-th and 30-th of the full ensemble MLEF cycle, respectively, for an explained variance of 99 %.

- **Figure 5** Magnitude of the eigenvalue spectrum of the \mathbf{C} matrix versus the total number of modes for each analysis cycle from cycle 2-28, for a retained variance of 99 %, where adaptive ensemble reduction is initiated from 1-st , 10-th, 20-th and 30-th cycle of the full MLEF ensemble run consisting of 40 cycles, respectively. (a) Full Ensemble (b) Reduced Ensemble

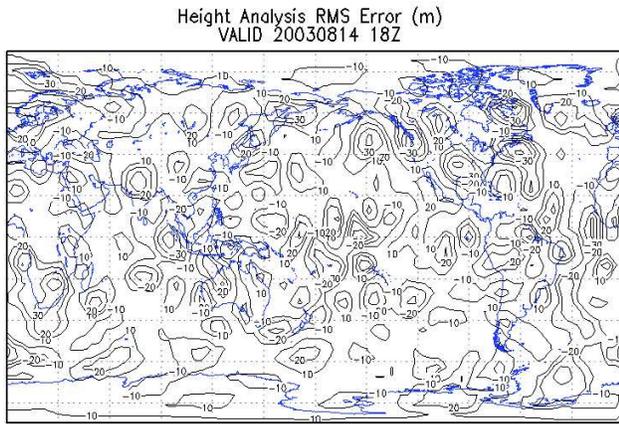
Full Ensemble (Left Column) and Reduced Ensemble (Right Column)



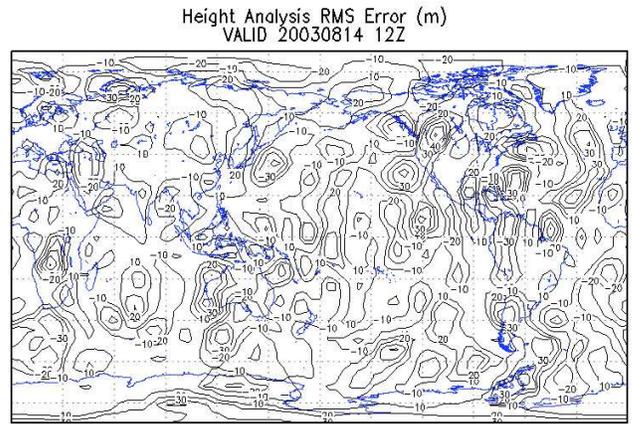
(a) Cycle24; Ensemble Size 1000



(b) Cycle24; Ensemble Size 380

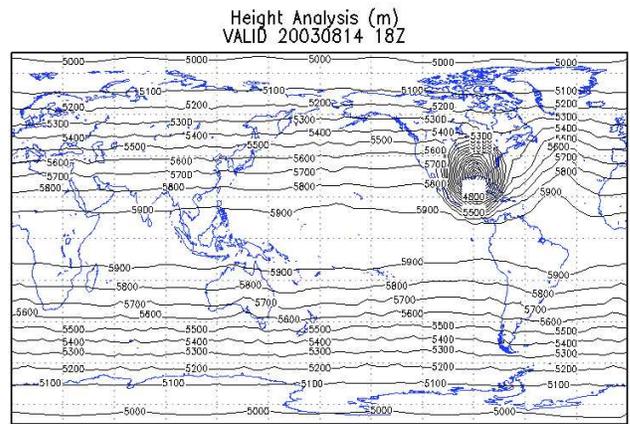
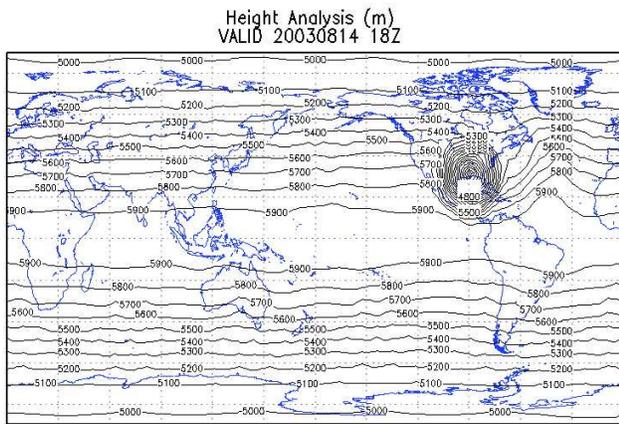


(c) Cycle26; Ensemble Size 1000



(d) Cycle26; Ensemble Size 280

Figure 1. RMS error isolines of height field analysis scaled 10 times larger for plotting. Initial conditions using Williamson *et al.* (1992) test case 5 ,i.e. zonal flow over an isolated conical mountain. Full ensemble (left column) and reduced ensemble with 95 % variance retained(right column) for (a) Cycle 24; Full ensemble size, 1000 members (b) Cycle 24; Reduced ensemble size, 380 members and (c) Cycle 26; Full ensemble size, 1000 members and (d) Cycle 26; Reduced ensemble size 280 members, respectively.

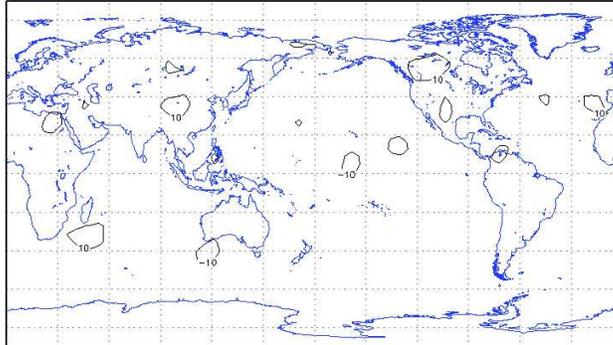


(a) Full ensemble for cycle26; Ensemble Size 1000

(b) Reduced ensemble for cycle26; Ensemble Size 280

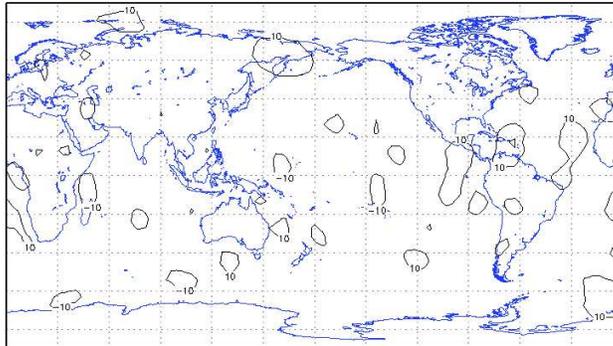
Figure 2. Isolines of height field analysis of Williamson *et al.* (1992) test case 5 ,i.e. zonal flow over an isolated conical mountain for (a) full ensemble, comprised of 1000 members and (b) reduced ensemble, comprised of 280 members for cycle 26 of the MLEF data assimilation with 95 % variance retained.

Height Analysis RMS Error (m)
VALID 20030814 12Z



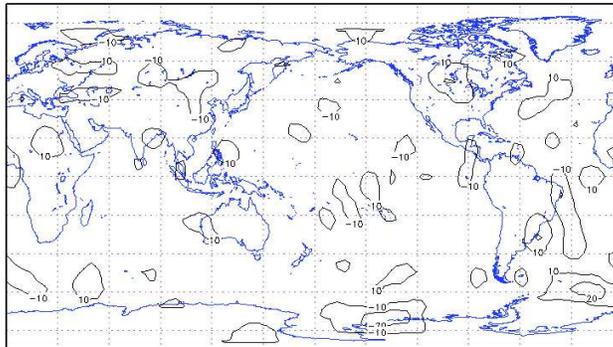
(a) Cycle 24

Height Analysis RMS Error (m)
VALID 20030814 18Z



(b) Cycle 26

Height Analysis RMS Error (m)
VALID 20030815 00Z



(c) Cycle 28

Figure 3. RMS error isolines of height analysis difference fields between reduced ensemble and full ensemble runs for cycle (a) 24 (b) 26 (c) 28 with retained variance of 95 %. The results are scaled 10 times larger for plotting. Initial conditions using Williamson *et al.* (1992) test case 5 ,i.e. zonal flow over an isolated conical mountain.

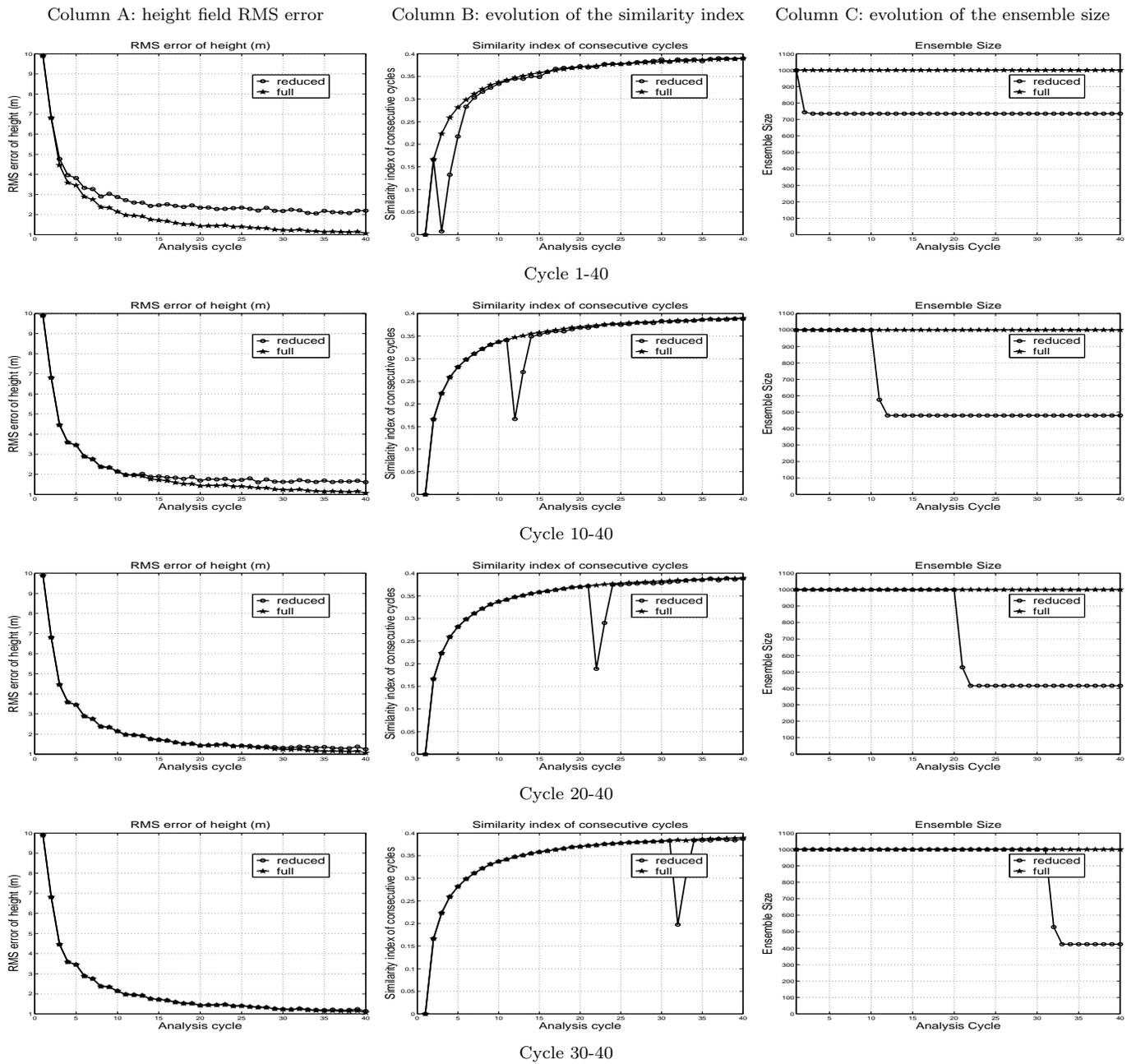
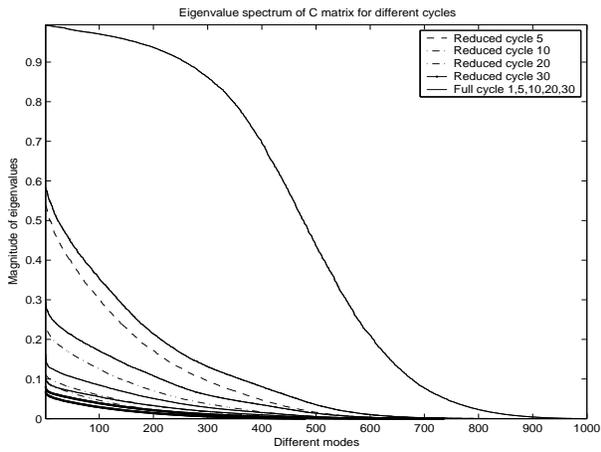
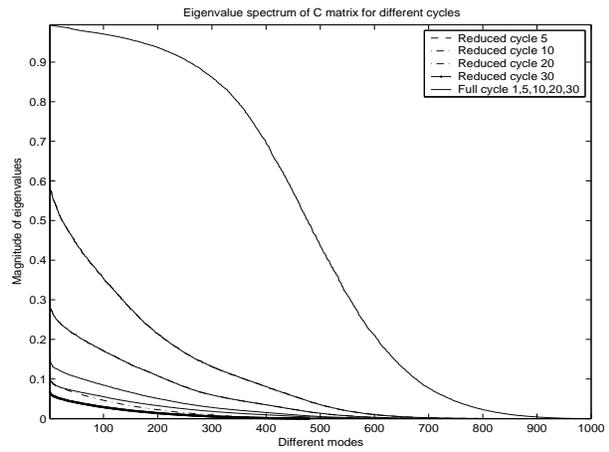


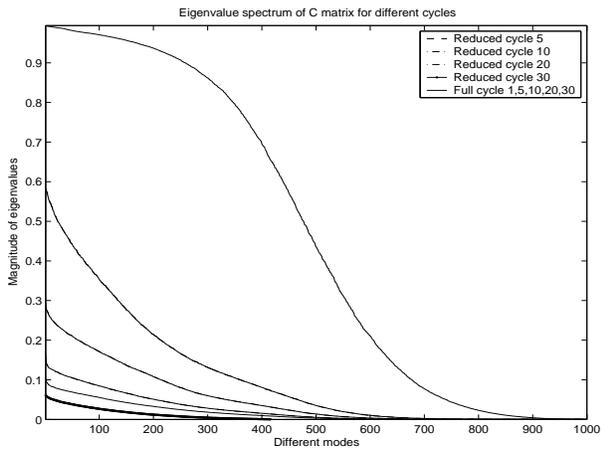
Figure 4. Column A illustrates the evolution of height field RMS error for full and reduced ensembles as a function of the MLEF cycle where adaptive ensemble reduction is initiated from 1-st , 10-th, 20-th and 30-th cycle of the full MLEF ensemble run consisting of 40 cycles respectively, for an explained variance of 99 %. Column B illustrates the evolution of the similarity index for full and reduced ensembles where adaptive ensemble reduction is initiated from 1-st , 10-th, 20-th and 30-th cycle of the full ensemble MLEF cycle, respectively, for an explained variance of 99 %. Column C illustrates the evolution of the ensemble size of the reduced ensembles where adaptive ensemble reduction is initiated from 1-st , 10-th, 20-th and 30-th of the full ensemble MLEF cycle, respectively, for an explained variance of 99 %.



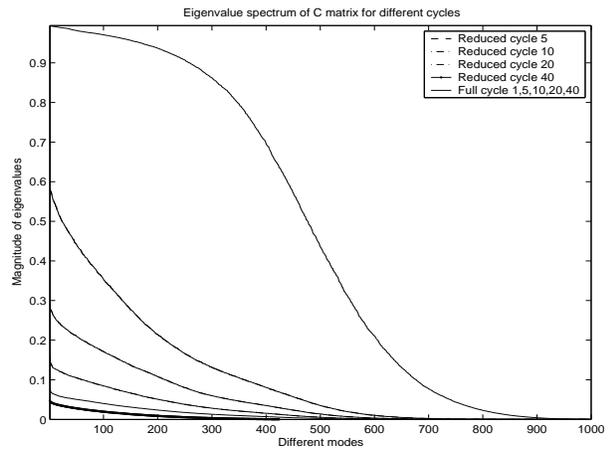
(a) Full ensemble versus reduced ensemble Cycles 1-40.



(b) Full ensemble versus reduced ensemble Cycles 10-40:
Note except cycles 20-30, all cycles coincide.



(c) Full ensemble versus reduced ensemble Cycles 20-40:
Note except cycle 30, all cycles coincide.



(d) Full ensemble versus reduced ensemble Cycles 30-40:
Note except cycle 30, all cycles coincide.

Figure 5. Magnitude of the eigenvalue spectrum of the C matrix versus the total number of modes for each analysis cycle from cycle 1-40, for a retained variance of 99 %, where adaptive ensemble reduction is initiated from (a) 1-st , (b) 10-th, (c) 20-th and (d) 30-th cycle of the full MLEF ensemble run consisting of 40 cycles