COMPARISON OF ENSEMBLE-BASED AND VARIATIONAL-BASED DATA ASSIMILATION SCHEMES IN A QUASI-GEOSTROPHIC MODEL

Shu-Chih Yang¹,²*, Matteo Corazza³, Alberto Carrassi⁴, Eugenia Kalnay¹, and Takemasa Miyoshi⁵

¹ Earth System Science Interdisciplinary Center/Department of Atmospheric and Oceanic Science, University of Maryland, College Park, Maryland, 20742
² Global Modeling Assimilation Office, NASA/ Goddard Space Flight Center, Greenbelt, Maryland, 20770
³ ARPAL – CFMI-PC, V.le Brigate Partigiane 2, 16129, Genova, Italy
⁴ Institut Royal Meteorologique de Belgique, Bruxelles, Belgique
⁵ Numerical Prediction Division, Japan Meteorological Agency, Tokyo, Japan

Corresponding author:
Dr. Shu-Chih Yang
Earth System Science Interdisciplinary Center/ Department of Atmospheric and Oceanic Science. University of Maryland, College Park, MD, 20742.

Email: scyang@atmos.umd.edu
Abstract

Four data assimilation schemes are implemented in a Quasi-geostrophic channel model. Among these, 3D-Var and 4D-Var are variational-based, and the hybrid and the Local Ensemble Transform Kalman Filter (LETKF) are ensemble-based schemes. We compare the quality of their analyses and forecasts using a 12-hour analysis cycle with simulated rawinsonde observations.

The schemes that are able to take into account the flow-dependent errors outperform the 3D-Var under a perfect model scenario. The ensemble-based 3D-Var hybrid scheme has a large advantage over the 3D-Var when ensemble vectors represent the dynamically growing errors and strongly project onto the structures of the background errors, as is the case for bred vectors and final singular vectors. In that case, the hybrid scheme combines the advantages of the 3D-Var and the ensemble Kalman Filter.

We show that given the same observations, the LETKF produces more accurate analyses than 4D-Var with a 12-hour window and that this advantage lasts for 5-day forecast lead-time, but that the performance of 4D-Var benefits substantially from using a longer assimilation window. The characteristics of the corrections made from the LETKF and 4D-Var are compared and discussed in this study.
1. Introduction

The procedure through which all available information on the state of a physical system is combined to obtain the best estimate of its state is known as data assimilation (DA); such a best estimate is usually referred to as an analysis. In geophysics and in particular in the context of numerical weather prediction (NWP) applications, DA algorithms are designed mainly to provide the optimal initial conditions to initialize prediction models. By extracting more information from the observations and physical governing laws, a good DA algorithm should in principle improve the analysis and consequently the forecast skill. Due to the chaotic nature of a nonlinear system such as the atmosphere and the oceans, it is crucial to estimate the initial conditions as accurately as possible in order to obtain reliable forecast states. At the present, most of the operational DA schemes combine in a statistical fashion the available information, i.e., the observations proper and prior information usually in the form of a short model forecast solution.

Three dimensional variational assimilation (3D-Var), which has been used in many operational centers, is considered as an economically and statistically reliable method. Its formulation is derived by assuming that the two sources of information, the forecast and the observations, have errors that are time-independent and Gaussian, described by prescribed static error covariances (Talagrand, 1997). Although these assumptions allow dealing with realistic NWP problems otherwise computationally intractable (Parrish and Derber, 1992), the 3D-Var method does not include the time-dependent, possibly non-normal, error dynamics of non-linear chaotic systems, i.e., the “errors of the day”.

The 3D-Var can be considered as a simplified version of the four-dimensional variational assimilation (4D-Var). The 4D-Var is an advanced technique that seeks the model trajectory that best fits the observations distributed within a given time interval with the dynamical constraint of the model equations (Talagrand and Courtier 1987; Courtier et al., 1994; Rabier et al. 2000). The optimal control theory (LeDimet and Talagrand, 1986) allows the minimization of the 4D-Var cost function (defined over the time interval) to be made with respect to the state at the beginning of the interval by writing its gradient using adjoint model. As a consequence, the computational cost of this method is high and its operational implementation so far has been performed with the simplified incremental form (Courtier et al, 1994). Although it should lead to improvements (Pires et al., 1996), operational centers at the current stage have not yet used an assimilation window beyond one-day for a 6-hour analysis cycle. Also, in 4D-Var, the forecast covariance matrix is implicitly evolved within the assimilation window with a static initial error covariance, and is not available to be used for the next assimilation cycle (Kalnay, 2003).

During the past decade, experience with ensemble forecasting has suggested that ensembles and DA could be combined in a natural way. The first and most celebrated ensemble-based DA scheme is the ensemble Kalman filter (EnKF; Evensen, 1994) that combines estimation theory with a Monte Carlo statistical approach within the conceptual and mathematical framework of Kalman filter equations (KF, Kalman and Bucy, 1961). In essence, most of the current ensemble-based data assimilation schemes have been
developed by aiming at a nonlinear extension of the KF approach while addressing the problem of reducing its computational cost. Several ensemble-based DA algorithms, usually divided into stochastic methods and square root filters, have been developed and the encouraging results obtained so far indicate that they can represent a feasible alternative to 4D-Var (Anderson 2001, Whitaker and Hamill 2002, Evensen, 2003, Ott et al. 2004, Kalnay et al. 2007, Miyoshi and Yamane 2007). In each ensemble-based scheme, a nonlinearly evolved ensemble of trajectories is used to sample the unknown flow-dependent error distribution, and in most practical applications its number is several orders of magnitude smaller than the dimension of the state vector of the system. The use of the full nonlinear model may be beneficial for analyses in situations where nonlinearity is strong and statistics exhibit some non-normality (Hamill 2003; Yang et al. 2005). Recently, Hunt et al. (2004 and 2007) demonstrated that it is feasible and straightforward to use asynoptic observations in the local ensemble Kalman filter scheme (4D-LEKF) without the need of an adjoint model. Kalnay et al. (2007) proposed a no-cost 4D-LEKF smoother that (like 4D-Var) improves the analysis at the beginning of the assimilation window.

Compared with the 4D-Var, an ensemble-based scheme is easy to implement and maintain, since it does not require the development and maintenance of the tangent linear and adjoint models. Moreover, through the analysis ensemble, it naturally provides a set of dynamically consistent states to initialize an ensemble prediction system. On the other hand, the 3D-Var/4D-Var schemes only provide a single deterministic “control” analysis solution and an additional procedure is needed to start a probabilistic forecast. The
information about the flow state, such as the uncertainties associated with flow instabilities, is propagated through the ensemble-based DA cycle, while each DA cycle is an independent cycle in the variational-based methods. Reviews of the ensemble-based data assimilations method are given in Hamill (2003) and Evensen (2003). With a simple perfect Lorenz-96 40-variable model (Lorenz, 1996), Fertig et al. (2007) compared the performance of 4D-Var and 4D-LETKF schemes by assimilating asynchronous observations. In their results, both schemes approach at comparable error amplitude where 4D-LETKF benefits from a short analysis window and 4D-VAR needs a sufficiently long analysis time window. Also, Lorenc (2003) and Kalnay et al. (2007) discuss the pros and cons of ensemble Kalman filter and 4D-Var.

Recently, hybrid schemes that combine the ensemble-based and variational assimilation schemes, suggest that the advantages from both sides can be attained and compensate for the limitation of using a time-independent background error covariance in the 3D/4D-VAR schemes (Hamill and Snyder, 2000, Corazza et al., 2002, Etherton and Bishop 2004, Wang et al. 2006a, b). It is valuable for the community to investigate the potential of the hybrid scheme since most operational centers presently have an ensemble prediction system and the 3D-Var system is also widely used. A hybrid method would thus allow the operational centers upgrade to a better analysis easily.

Through a fair comparison among variational and ensemble-based DA schemes, this work attempts to contribute to the current debate on their merits in terms of the quality of the analyses as well as on the effective computational and implementation cost, and to discuss whether ensemble-based scheme may constitute a competitive alternative to 4D-Var also in an operational context. Two variational and two ensemble-based schemes
have been implemented in a quasi-geostrophic channel model, and we compare their performance when assimilating the same noisy observations. The two variational schemes are 3D-Var (developed by Morss, 1998, Morss et al., 2001, following Parrish and Derber, 1992) and 4D-Var (newly developed for this study). The two ensemble schemes are the Local Ensemble Transform Kalman Filter (LETKF, based on Hunt et al., 2007), and a hybrid system with a set of dynamically evolving ensemble of either bred or singular vectors added to the regular 3DVAR system (Corazza et al, 2002) by augmenting the background error covariance with information on the “errors of the day”. In this study, the perfect model assumption is made for the all the experiments, so the comparison focus on the ability of the DA schemes to control and reduce errors coming from an incorrect estimate of the initial conditions only. The object of this study is to explore the differences between the variational-based and ensemble-based DA methods and discuss considerations that would be applicable to more realistic cases (if implemented operationally). The impact of the model error and its consequence on the performance of the DA schemes would complete the comparison and could constitute the object of future works.

The paper is organized as follow: in section 2 the setup of the model and of the observation network are described, in section 3 brief descriptions about the DA schemes used in this study are given and the results are presented in section 4. Finally, a summary and discussions are presented in section 5.

2. Observing Systems Simulation Experiments setup with a Quasi-geostrophic model

2.1. The Quasi-geostrophic model, tangent-linear and adjoint models
All the data assimilation schemes are implemented in the quasi-geostrophic (QG) model developed by Rotunno and Bao (1996). It is a periodic channel model on a beta-plane. At the resolution used in this study, it has 64 grid points in the zonal direction, 33 grid points in the meridional direction, and 7 vertical levels; the total number of degrees of freedom is 14784. Physical processes include advection, diffusion, relaxation and Ekman pumping at the bottom level. The model variables are non-dimensional potential temperature at the bottom and top levels, and non-dimensional potential vorticity at the 5 inner levels. The integration time step is 30 minutes. Details about the governing equations can be found in Rotunno and Bao (1996) and Morss (1998). This model has been widely used for testing and comparing problems related to data assimilation and adaptive observations (e.g., Morss 1998, Hamill and Snyder 2000, Corazza et al., 2002, 2003, Kim et al. 2003, Carrassi et al. 2007, and Corazza et al., 2007). In this study, the model is assumed to be perfect and the true state, from which observations are extracted, is represented by a reference trajectory integrated from this QG model. Hereafter, the Corazza et al. (2007) will be referred to as Corazza07 because the experimental configuration of the LETKF follows tightly and is related to their work.

In order to implement the 4D-Var scheme, the tangent linear and adjoint models of this nonlinear QG model have been developed. The evolution of the linear perturbation inherits the boundary conditions used in the nonlinear model. In this study, the Tangent Linear and Adjoint Model Compiler (TAMC, Giering, 1998 and 1999) was used to generate the tangent linear and adjoint codes. Although it was straightforward to build up the codes with TAMC, they did not fulfill the boundary conditions automatically and
several very subtle corrections (i.e., a long debugging effort) were required for the TAMC generated linear and adjoint models in order to avoid the accumulation of extreme values at the meridional and vertical walls and zonal periodic boundaries. Verification checks for tangent linear and adjoint codes, following Navon et al. (1992), indicate the linear regime is valid for 5 days.

2.2. The observing system configuration

The simulated “rawinsonde observations” consist of the velocity components and temperature at all levels. They are generated from the true state through a linear observation operator, $H$, mapping from model variables into observation variables (Morss, 1999). 64 rawinsonde observations are used and their locations are randomly chosen at the beginning and remain fixed afterward. The observation locations are on the model grid points (about 3% coverage of the domain), thus no interpolation is involved in the observation operator. Observations are available every 12 hours and the analysis cycle is also performed every 12 hours.

Observation errors are generated by adding white random noise sampled by a Gaussian-typed distribution consistent with the given observational error covariance matrix (Morss, 1999, Morss et al., 2001). The observation error covariance matrix is constructed following Dey and Morone (1985): the observation error is assumed to be uncorrelated between observations and between different variables; only vertical correlation for the same variable is considered. The wind and temperature observation error variances are adapted from Parrish and Derber (1992); the corresponding values are provided in Morss (1999). The default zonal wind error has 80% of the total wind error
and the meridional wind error has 50% of the total wind error in order to reflect the model geometry.

3 Data assimilation schemes

3.1 3D-Var

The 3D-Var system implemented for the QG model was developed by Morss (1999), based on the same framework as in Parrish and Derber (1992). The time-independent background error covariance matrix \( B \) (in grid-points coordinates) has separable horizontal and vertical structures and it is assumed to be diagonal in horizontal spectral coordinates. Thus, the background error covariance in the spectral space \( C \), can be written as

\[
C = \hat{C} \hat{C}^{1/2} V \hat{C}^{1/2},
\]

where \( \hat{C} \) is the matrix of horizontal background error covariance at each level and \( V \) is the matrix of background error correlations between different levels for each horizontal spectral components. The diagonal property in \( \hat{C} \) reflects the assumption that structures of different wavelengths are uncorrelated. The grid-points space matrix \( B \) is then recovered through a grid-point – spectral space transformation, according to \( B = S C S^{T} \), where \( S \) is the transformation matrix. Further details on the 3D-Var configuration setup are found in Morss (1999).

The 3D-Var is implemented and solved in the incremental cost-function form as shown in (1). The control variable for the minimization of the associated cost-function is the analysis increment \( \delta x_{a} = x_{a} - x_{b} \) where \( x_{a} \) and \( x_{b} \) denote the analysis and the background state vectors respectively. The 3D-Var incremental cost-function is given by:
\[
J(\delta x_a) = \frac{1}{2} [\delta x_a^T B^{-1} \delta x_a + (y^o - H(x_b))^T R^{-1} (y^o - H(x_b))] 
\]

(1)

where \( y^o, R \) and \( H \) indicates the observational vector, the observation error covariance matrix and \( H \) the observation operator respectively.

The minimum of (1) is obtained by computing the solution of (2) with the zero gradient of the cost function.

\[
[I + BH^T R^{-1} H] (x_a - x_b) = BH^T R^{-1} (y^o - H x_b) 
\]

(2)

In (2), \( H \) is the linearized observation operator; we point out that in all the applications described in this work, the observation operator is inherently linear. The analysis increment is solved with the iterative conjugate gradient method.

In this study, the 3D-Var provides the benchmark for comparison with other assimilation schemes.

3.2 4D-Var

The 4D-Var cost-function generalizes that of the 3D-Var to consider observations distributed within a time interval. Its minimization provides the initial condition (at the beginning of the time interval) leading to the forecast trajectory that best fits all the observations within the assimilation window (Courtier et al. 1994). By using the forward model integration the differences between observations and their background estimates (innovations) are computed at the appropriate times.

As the generalized form of the 3D-Var incremental cost function, we make use of the 4D-Var in incremental form (Courtier et al, 1994). Equations (3) and (4) show the 4D-
Var cost-function and its gradient respectively as a function of the analysis increment at the initial time $\delta x(t_0)$ in the grid-points coordinates:

$$J(\delta x(t_0)) = \frac{1}{2} (\delta x(t_0))^T B^{-1} (\delta x(t_0)) + \frac{1}{2} \sum_{i=0}^{n} [H(t_0, t_i) \delta x(t_0) - d(t_i)]^T R^{-1} [H(t_0, t_i) \delta x(t_0) - d(t_i)] \quad (3)$$

$$\nabla J(\delta x(t_0)) = B^{-1} (\delta x(t_0)) + \sum_{i=0}^{K} L(t_i, t_0) H(t_0, t_i) U^{-1} [H(t_0, t_i) \delta x(t_0) - d(t_i)] \quad (4)$$

In these equations, $d(t_i)$ is the innovation (the difference between the background state and observations) at observing time $t_i$. $L(t_0, t_i)$ is the tangent linear (forward) model advancing a perturbation from $t_0$ to $t_i$ and $L^T(t_i, t_0)$ is the adjoint (backward) operator.

The cost-function (3) has been preconditioned through a variable transformation that allows avoiding the inversion of $B$, to compute the cost-function directly and make the minimization more efficient. The transformation operator, $U$, is chosen to be the square root of the inverse of $B$, $B^{-1} = U^T U$. With the preconditioned variable, $\delta \hat{v}$, the analysis increment is expressed as $\delta x = U^{-1} \delta \hat{v}$ and the cost function and its gradient are reformulated, as

$$J(\delta \hat{v}(t_0)) = \frac{1}{2} (\delta \hat{v}(t_0))^T (U^{-1} \delta \hat{v}(t_0)) + \frac{1}{2} \sum_{i=0}^{n} [H(t_0, t_i) U^{-1} \delta \hat{v}(t_0) - d(t_i)]^T R^{-1} [H(t_0, t_i) U^{-1} \delta \hat{v}(t_0) - d(t_i)] \quad (5)$$

$$\nabla J(\delta \hat{v}(t_0)) = \delta \hat{v}(t_0) + \sum_{i=0}^{K} (U^{-1})^T L(t_i, t_0) H(t_0, t_i) U^{-1} [H(t_0, t_i) U^{-1} \delta \hat{v}(t_0) - d(t_i)] \quad (6)$$

The minimization is now made with respect to the variable $\delta \hat{v}$, initially set to zero. Following the assumptions in Morss (1999) and Parrish and Derber (1992), $B$ is built by constructing the horizontal error covariances at each level and linking them through a
vertical correlation matrix, assuming no correlation between different wave numbers. Therefore, we can define $U^{-1}$ as:

$$U^{-1} = S \hat{C}^{\frac{1}{2}} \sqrt{V},$$

(7)

where $S$ is the operator for the transformation from spectral to model coordinates as defined in Sec. 3.1, $\hat{C}^{\frac{1}{2}}$ is the square root of the horizontal background error variance of each wave-number, and $\sqrt{V}$ is the square root of the vertical correlation matrix. In this application, $\hat{C}^{\frac{1}{2}}$ is diagonal and $\sqrt{V}$ is a 7x7 matrix.

Note that the control variable $\delta v(t_0)$ is expressed in spectral coordinates. Given an initial guess for $\delta v(t_0)$, the cost function and its gradient are computed using (5) and (6) solved with a L-BFGS quasi-Newton minimizer. The process is repeated till the minimization criterion is satisfied with a threshold that the gradient of the cost function is smaller than a chosen tolerance value of $10^{-3}$. In addition, we approximate the innovation vector with the full nonlinear model as in (8) based on the assumption that the analysis state is close to the background state.

$$H_L(t_0, t_i) U^{-1} \delta v - d(t_i) = H_M(x(t_0)) - y(t_i)$$

(8)

It is common to assume that the background error covariance $B$ in (3) has a correlation structure similar to the background error covariance optimized for the 3D-Var. In a perfect model, this is a reasonable assumption, but the performance of 4D-Var is also sensitive to the amplitude of $B$. Since 4D-Var is expected to provide a better first guess than 3D-Var, using $B=B_{3D-Var}$ overestimates the actual background error (Kalnay et al, 2007). Therefore, we retained the structure of $B_{3D-Var}$ but tuned its amplitude: we found that for a 12 hr assimilation window, $B=0.02B_{3D-Var}$ was optimal. In the following
results, this tuned $B$ is used as the initial background error covariance for all the 4D-Var experiments.

Table 1 indicates the time and domain root-mean-square (RMS) analysis errors, expressed by the potential enstrophy norm, for different assimilation window length ranging from 12 hours to 5 days, averaged over 30 days (60 analysis cycles). The results indicate that longer windows, although computationally costlier, are very beneficial for 4D-Var, confirming the results of Pires et al. (1996) and Kalnay et al. (2007). We also notice that for longer windows the sensitivity to the amplitude of $B$ decreases (not shown). Most of the improvement is gained when the assimilation window increases from 12 hours to 1 day, while beyond 1 day the improvement is small. By comparing the experiment on a Linux PC machine with a Pentium 2.66 GHz processor and a 1GB of memory, the computational time needed for a 5-day window is 9 times larger than for a 12-hour window. As a consequence, it may become impractical to use a very long assimilation window with the 4D-Var scheme in the framework of an operational context.

3.3 Hybrid scheme

In the 3D-Var scheme, the background error covariance is time-independent and therefore does not include “errors of the day”. Hamill and Snyder (2000) proposed a hybrid DA scheme by combining a 3D-Var with an Ensemble Square Root Kalman Filter scheme (Whitaker and Hamill, 2002), showing sensible improvements in the analysis due to the flow-dependent information provided by the ensemble-based estimate of the background error covariance matrix. They showed that the ensemble-based background
error covariance can completely replace the background error covariance of 3D-Var if the ensemble size is large enough. In their hybrid scheme, the assimilation process is necessary for each ensemble member with perturbed observation at every analysis step, and therefore the computational cost may be unaffordable if a large ensemble size is required. On the other hand, Patil et al. (2000) showed that the flow dependent atmospheric instabilities (the errors of the day) tend to be locally confined to a low dimensional subspace with respect to the global system phase space. Such evidence supports the construction of an ensemble to span the unstable subspace of the system by representing such space with an ensemble size much smaller than the model dimension, thus contributing to reduce the computational cost dramatically. Here, we used a rather efficient hybrid scheme proposed by Corazza et al. (2002) where a set of bred (or singular) vectors is used to augment the background error covariance of the 3D-Var scheme. As in Hamill and Snyder (2000), the new background error covariance (9) is a linear combination of the standard 3D-Var background error covariance \( B_{3DVAR} \) and an additional error covariance \( B' \) constructed from an ensemble of perturbations\(^1\). As shown in (10), the additional covariance used to augment the 3D-Var covariance is given by the outer product of ensemble vectors \( \mathbf{v} \) that represent flow dependent growing perturbations and taken as the square root of the ensemble-based background error covariance, \( c \) denotes a normalization factor that ensures the error covariance has an optimal magnitude with a variance comparable to the analysis error variance from the 3D-Var.

\[
B = (1 - \alpha)B_{3DVAR} + \alpha B'
\]

\(^{1}\) Wang et al. (2006) indicated that solving the minimum of the cost function with this hybrid background error covariance is equivalent to augmenting the state vector with another set of control variables associated with the square root of the ensemble covariance.
\[ B' = cvv^T \]  

Corazza et al. (2002) proposed an efficient approach to solve the augmented part of the background error covariance within the 3D-Var framework (1). The analysis increment is computed as the correction from the regular 3D-Var (B) and from the ensemble-based error covariance (B') system, as shown in (11).

\[
[I + ((1 - \alpha)B_{3DVAR} + \alpha cvv^T)H^T R^{-1}H](x_a - x_b) = \\
[(1 - \alpha)B_{3DVAR} + \alpha cvv^T]H^T R^{-1}(y - Hx_b)
\]

During the minimization procedure, the differences between LHS and RHS of (11) are iteratively evaluated. Since the LHS and RHS of (11) both start from a vector, there is no need to calculate B' explicitly. Also, the hybrid scheme is performed in the 3D-Var framework with the same conjugate residual solver. This approach also allows easily exploring the impact of using different types of ensemble vectors. In the following experiments, we will demonstrate, as could be expected, that the hybrid scheme can improve the 3D-Var scheme when the ensemble vectors provide a good representation of the background errors (the 12-hour forecast error).

With a 20-member ensemble of bred vectors, Corazza et al. (2002) found that the optimal value for the weighting coefficient, \(\alpha\), was 0.4, but \(\alpha\) needs to be carefully optimized if the observation error or the ensemble size vary. They also found that adding small amplitude random noise to the bred vectors during the breeding cycle significantly reduces the RMS analysis error and accelerates its convergence to a low and stable value. The size of the random perturbations is chosen to have the same variance as the observation errors. This method was confirmed by Corazza07 for the LEKF (section 3.4).
In this study, we localize the $\mathbf{v} \mathbf{v}^T$ in (11) in order to reduce the spurious error rising from an incorrect estimate of the long-distance correlations sampled with a finite size ensemble (Houtekamer and Mitchell, 2000). More specifically, the background error covariance constructed from the ensemble vectors is localized by means of a Gaussian masking function. As shown in (12), a compactly supported function $\rho$ is applied to limit the extent of the covariance, where $[\bullet]$ is the Schur product. $\rho$ is given in (13), where $r_d$ is the de-correlation length and $r_{i,j}$ is the distance between grid points $i$ and $j$.

$$ \mathbf{B}' = \rho \bullet \mathbf{v} \mathbf{v}^T $$

(12)

$$ \rho_{i,j} = e^{-\frac{r_{i,j}^2}{2r_d^2}} $$

(13)

After optimization, the de-correlation length has been set equal to a distance of 10 grid points and the correlation is assumed to be zero if $\rho_{i,j}$ is smaller than 0.1. This distance is inherently related to the spatial scale of the dominant instabilities of the system. The localization improves the performance of the hybrid scheme so that the dynamic errors from ensemble vectors can be more effectively used. Unfortunately, the localization while significantly improving the performance of the hybrid scheme, also makes it more computationally expensive (Table 4).

Our experiments confirm that the positive impact of the localization process on using the ensemble vectors information more efficiently: with the same hybrid coefficient $\alpha = 0.4$ , a ensemble size of 100 members without localization is required to have results comparable to those obtained with localization using an ensemble of size 20.
3.4 Local Ensemble Transform Kalman Filter (LETKF)

The Local Ensemble Kalman Filter (LEKF) was first proposed by Ott et al. (2004), who solved the Ensemble Kalman Filter equations in local patches exploiting the low-dimension properties of the atmospheric instabilities (Patil et al. 2000). Szunyogh et al. (2005) successfully tested this scheme in a large realistic atmospheric primitive equation model with complicated physics (NCEP GFS). Corazza07 studied its performance and made a comparison with a 3D-Var with the same QG model used here. The LEKF scheme was modified by Hunt et al. (2007) into the LETKF, an equivalent but more efficient approach. This scheme has been proven to provides essentially identical results as the ensemble square root filter of Whitaker and Hamill (2002) but it is computationally much more efficient when the number of observations is large (Whitaker et al. 2007). In the following, we briefly discuss the main characteristics of the LETKF concerning the implementation in the present QG model framework.

The LETKF/LEKF scheme is based on the use of an ensemble of model solutions and on the representation of the state in local domains, allowing both the data assimilation step and the construction of the new vectors of the ensemble to be performed locally and in parallel, in contrast to the variational methods (3D/4D-Var).

The local domain referred to a local volume surrounds the analysed grid point, with a total number of grid points in this volume equal to \((2l + 1)^2 \times (2l_z + 1)\), where \((2l+1)\) is the chosen horizontal local length and \((2l_z+1)\) is the chosen vertical local length. The corresponding background state in this volume is then arranged into a local ensemble vector with their ensemble mean denoted \(\vec{x}_f\) and the ensemble perturbation denoted \(\delta x_f\).
A similar notation is adopted for the ensemble of analysis states, with the mean denoted \( \bar{x}_a \) and the deviations denoted \( \delta x_a \). The observations available in this local volume are denoted as \( y_o \). The standard Kalman Filter formula (14) is used to update the analysis mean state and solved locally.

\[
\bar{x}_a = \bar{x}_f + P_f H^T (H P_f H^T + R)^{-1} (y_o - H \bar{x}_f)
\]  

(14)

In (14), \( H \) is the same linear observation operator discussed for other schemes, and \( P_f \) is the local forecast error covariance in the model grid coordinate (physical space). In LETKF, the error statistics are sampled from the K ensemble perturbations and represented in the K-dimensional space. A transform matrix is used to map from the K-dimensional space back to the physical space (in this case, the dimension of the local vector, \( N = (2l + 1)^2 \times (2l_z + 1) \)). The local forecast error covariance (a \( N \times N \) matrix) is estimated using the local ensemble forecast perturbations as

\[
P_f^\prime = \frac{1}{K-1} [\delta x_1, \ldots, \delta x_K] [\delta x_1, \ldots, \delta x_K]^T.
\]

(15)

With an appropriate transform matrix, the LETKF transforms the ensemble background perturbation into the analysis ensemble perturbations in the K-dimension ensemble space (Hunt et al. 2007): provided the formula for the as

\[
\bar{P}^\prime = [(H E_f^\prime)^T R^{-1} (H E_f^\prime) + (K - 1) I / \Delta]^{-1},
\]

(16)

where \( \bar{P}^\prime \) is the transform matrix, \( E_f^\prime \) equals to \( \frac{1}{\sqrt{K - 1}} [\delta x_{i,i=1,..,K}] \), a \( N \times K \) matrix whose columns are the background ensemble perturbations, and \( \Delta \) is a parameter for multiplicative variation inflation. The analysis error covariance in model space is then given by
\[ \mathbf{P}_a = \mathbf{E}_f \hat{\mathbf{P}}_a \mathbf{E}_f^T. \]  

(17)

and the analysis ensemble perturbations are

\[ \delta \mathbf{x}_a = \mathbf{E}_f (K - 1) \hat{\mathbf{P}}_a  \]

(18)

Equation (14) is then written as

\[ \bar{\mathbf{x}}_a = \bar{\mathbf{x}}_f + \mathbf{E}_f \hat{\mathbf{P}}_a (\mathbf{H} \mathbf{E}_f)^T \mathbf{R}^{-1} (\mathbf{y}^o - \mathbf{H} \bar{\mathbf{x}}_f). \]

(19)

Equations (18) and (19) provide the full analysis ensemble. See Hunt et al. (2007) for more details and discussion.

After the local analysis procedure is completed, the center point of the local volume is used to update the analysis state in the global domain. We adopted the additive variance inflation from Corazza07 who suggested that this method is able to efficiently improve the performance of the system by “refreshing” the ensemble vectors, thus preventing its collapse into a too small space, which causes a rank problem in the ensemble-based DA schemes (Wang and Bishop, 2003, Etherton and Bishop, 2004). A small amount of random perturbations (of the order of 2% of the field variability) is generated in a similar way as discussed in the hybrid scheme in Section 3.3 and added to the part of ensemble perturbations (deviations from the ensemble mean). Both the initial and the additive ensemble perturbations are required to fulfil the boundary conditions of the model.

In addition, an observation localization method, suggested by Hunt (2005, personal communication) and used by Miyoshi (2005), is applied by multiplying the observation error covariance by the inverse of a Gaussian localization operator. As a result, the weighting of the observations located further away from the center of the
volume is reduced. This localization scheme has a significant impact when the observation density is sparse and large local patches are required.

For our application, we modified the multiplicative variance inflation in (16) to have vertical dependence (Table 2) due the error characteristic (see section 4.3). This reduces the RMS analysis error by about 4% in the potential vorticity at the lower levels and 6.5% in the potential temperature at the bottom level compared to the result with a vertically-constant inflation variance.

Our results confirm that the performance of LETKF is comparable to that of the LEKF tested in Corazza07 but more computationally efficient. The size of the local patch needs to be optimized depending on the ensemble size and the observing density. For this model, a local patch with a horizontal domain of 7x7 \((l=3)\) produces a reliable analysis with 64 observations (Table 4) reducing by 53% the RMS analysis errors obtained from the 3D-Var system. Improvements are obtained when enlarging the size of the local patch; however, this impact is saturated beyond a certain size (see Table 4 and the discussion in Corrazza07). The optimal local (horizontal) patch using 64 observations and 40 ensemble members is 19×19 \((l=9)\) which allows the inclusion of 3~4 observations per local patch. Also, in our experiments, all the vertical levels are included \((2l_z+1=7)\), without vertical localization. Table 2 lists the experiment configuration of the LETKF used in this study.

4. Results
4.1 3D-Var vs. ensemble-based hybrid schemes

In this subsection, we examine the performance of hybrid schemes with different types of ensemble vectors. The hybrid scheme is implemented by using either bred or singular vectors, both related to the dynamically evolving errors, as well as flow-independent random vectors for comparison. Bred vectors and singular vectors are operationally used for initializing the ensemble prediction system at NCEP and ECMWF respectively (Toth and Kalnay, 1997; Molteni et al, 1996).

Bred vectors (BV), defined as the differences between perturbed and non-perturbed nonlinear runs, represent the fast growing dynamical instabilities of the evolving flow and naturally carry information on “errors of the day” (Toth and Kalnay, 1993, 1997). As mentioned in 3.3 and 3.4, a small amount of random perturbations are added to BVs at every breeding cycle in order to “refresh” BVs and avoid the tendency of BVs to converge to a too small dimensional subspace (Wang and Bishop, 2003). We notice that the “refreshing” plays an important role in accelerating the BV convergence to the subspace of growing instability during the transition period and also increases the BVs’ ability to capture the subspace of the evolving background errors (Appendix B, Yang, 2005).

Singular vectors (SVs, Palmer et al. 1998, Errico and Ehrendorfer, 1995 and Kim and Morgan 2002) are the directions that maximize the growth of perturbations at the end of a chosen optimization time with a chosen norm. The definition of growth is given in (20), where $\|C\|^2$ is the square of the L2 norm defined with the chosen norm C. In our
case, both the initial and final norms are the potential enstrophy norm\(^2\) and \(C\) is simply
the identity matrix, since the model is appropriately scaled. The initial SV are then
derived by solving the eigensystem(21). The SVs are obtained by using the linear
operator, \(L\), to forward integrate the initial singular vectors for the chosen optimization
time (i.e. the time interval \(t_0\) to \(t\)).

\[
\lambda^2 = \frac{\|x(t)\|_C^2}{\|x(t_0)\|_C^2} = \frac{\langle x(t_0), L^T C L x(t_0) \rangle}{\langle x(t_0), C x(t_0) \rangle} \tag{20}
\]

\[
L^T C L x(t_0) = \lambda^2 x(t_0) \tag{21}
\]

Figure 1 shows the leading initial and final SVs derived from different chosen
optimization times as well as the BV with the corresponding 12-hour forecast errors
overimposed in contours. From Fig. 1, we see that the leading initial SVs are dominated
by large-scale features because of the choice of the norm (the potential enstrophy norm in
this study), while the final SVs show rather local structures, which resemble the shapes of
the BVs. Both the final SVs and BVs project strongly on the forecast errors. As the
optimization time is lengthened, features relative to higher wavenumbers appear in the
initial SV as indicated in Fig 1(g). In all these experiments the SVs and BVs are derived
based on the corresponding analysis states obtained from the hybrid system described in
Section 3.3.

4.1.1 Flow-dependent vectors: bred vectors vs. singular vectors

\(\text{Potential enstrophy norm} = \sum_{i=1}^{64} \sum_{j=1}^{i-1} \left( \sum_{k=1}^{5} (\delta \eta_{i,j,k})^2 + \delta \Phi_{i,j}^2 + \delta \theta_{i,j}^2 \right)\), where \(\delta \eta\) is the perturbation of potential
vorticity, \(\delta \Phi\) and \(\delta \theta\) are the perturbation of potential temperatures at bottom and top levels.

\cite{23}
In this section, we compare the performance of the BVs-hybrid with a SVs-hybrid scheme. We first examine the relationship between BVs and initial/final SVs with the actual background error.

Patil et al. (2000) pointed out the local structure and low dimensionality of forecast errors. Here we measure the forecast error projection on the BVs or SVs by means of the local explained variance as in Corazza et al., 2002. The local explained variance is defined as the square of the cosine of the local angle between the vector representing the background error on a local patch with a size of 5x5 grid at the mid (3rd) level, and the subspace of BVs/SVs within the same patch. Figure 2 shows this quantity as a function of time for 20 BVs and both final and initial SVs. It is evident that the background error locally projects very well on 20 ensemble vectors (BV or final SV). In particular, from Figure 2 and Table 3, we see that the BVs and the final SVs with the longest optimization period (48 hours) have the highest local explained variance while all the initial SVs have lower values, especially those with the longest optimization period. Clearly, as the optimization period increases, the final singular vectors become more similar to bred vectors. Even though the initial SV have the least correlation with the background errors, the final SV evolve into local structures that strongly project on the background error.

Table 3 summarizes the explained variance results as well as the local E-dimension** of the space spanned by both 10 and 20 ensembles (Patil et al., 2000). The

** The E-dimension (Patil et al. 2000) is computed as \(\frac{\sum_{i=1}^{k} \sigma_i^2}{\sum_{i=1}^{k} \sigma_i^2}\), where \(\sigma_i, i=1,\ldots, k\), are the singular values from applying the singular value decomposition on a matrix whose column is the ensemble vectors from a local patch with a size 5x5 grid points and with k ensemble.
subspaces spanned by the BVs and the SVs are both characterized by locally low dimensions much smaller than the ensemble size, suggesting that growing directions of the forecast errors also locally project on the ensembles. The BVs, refreshed with a small amount of random perturbations, effectively capture more growing directions as indicated by an increased E-dimension compared to the unperturbed BVs (not shown). The 12-hour and 24-hour initial SVs have slightly smaller E-dimensions than the BVs, indicating that while the SVs are globally orthogonal, locally, they have similar shapes and form a locally low-dimensional space. The results presented in Table 3 suggest that the BVs, dominated by the local flow instabilities, have an advantage in capturing locally growing errors. In summary, the results indicate that 20 BVs lead to a good estimation of forecast errors (highest explained variance), slightly overperforming the 48-hours final SVs, but at a lower computational cost.

It is reasonable to expect that when we use these ensemble vectors to augment the 3D-Var background error covariance, the hybrid system will show the most benefits from the ensembles that best represent the structures of background errors.

In the following, we augment the 3D-Var background error covariance with the ensemble-based background error covariance built by 20 BVs and 20 SVs (both the initial and final ones at the corresponding analysis time) following the method described in Section 3.3. Figure 3 is the time series of the RMS analysis errors of the hybrid systems with the result of the regular 3D-Var shown as the benchmark. Due to computational limitations, the ensemble size is limited to 20. All the hybrid experiments start from the same 3D-Var analysis. We illustrate here the influence of flow-dependent
ensemble vectors on the analysis without the localization procedure; the hybrid coefficient is set to 0.4.

The results shown in Fig.3 indicate that the hybrid schemes with BVs or with 24-, 48-hours final SVs lead to large improvements with respect to the regular 3D-Var system while the use of 12-hours final SVs show least positive influences; the use of initial SVs has large negative impacts on the DA performance in this context indicating they are not sufficiently representative of the background error structures. Best results are obtained by using the set of vectors that better describe the forecast error space and time features. As expected from the explained local variance (Figure 2), the BVs hybrid scheme performs similarly to the 48-hours final SVs. However, there is a large computational time difference between 20 BVs and 20 48-hour SVs: the former takes 1.1 seconds to complete one analysis cycle, while 7.2 seconds are necessary for the hybrid SV system. Figure 3 also indicates that large error spikes occasionally occur when the hybridized ensemble vectors are the final SVs with a short (12-hour) optimization period and a longer optimization period is required to remove such spurious spikes. By contrast, the error spikes are much less evident in the hybrid-BV systems, where the rapid error growth due to uncorrected fast growing errors is inhibited. We emphasize that these results were obtained “refreshing” the bred perturbations to keep the ensemble from collapsing into a small subspace.

The hybrid assimilation experiments (in the presence of model errors) conducted by Etherton et al (2004) showed that their hybrid system provides the best results by using perturbed observations (with 16 ensembles members) or with singular vectors while bred vectors are shown to provide the least improvement compared to the 3D-Var
system. This is because in their experiment, the BVs are not refreshed, and evolve towards similar directions, which leads to a BV-based covariance with a limited capability to correct the fast growing modes. Such limitation is not a concern here; the hybrid system with a regularly refreshed BVs has an improved performance that reduces 20% of the 3D-Var analysis errors compared to only 9% improvement obtained when the BVs are used without small random perturbations.

4.1.2 Flow-dependent vectors vs. flow-independent vectors

Figure 4 shows the results from the hybrid scheme in terms of the time averaged RMS analysis error (potential enstrophy norm) using 20 bred (or random) vectors as a function of the weighting coefficient $\alpha$ ranging from 0.1 to 1.0. The hybrid system with bred vectors without localization diverges if the weighting coefficients are larger than 0.9. The localization improves the performance of the scheme for any value of $\alpha$ and prevents the system from divergence even when the background error covariance comes only from the BV-based ensemble ($\alpha=1.0$), although in this case the system performs slightly worse than 3D-Var. With $\alpha = 0.9$, the localized-hybrid schemes provide an analysis comparable to the 4D-Var with a 12-hour assimilation window and also the LETKF scheme with a small local patch of $l=3$ (table 4). Because the BVs are dominated by locally growing errors, such a large improvement takes place due to efficiently taking the advantages of the local properties of BVs and to the reduction of sampling errors related to spurious long-distance correlations. This gives the hybrid scheme high confidence to allow the background error covariance to rely on the BVs’ structures.
These results indicate the potential advantage of this hybrid scheme, an easy-to-implement combination of the regular 3D-Var and ensemble-based DA scheme. In practice, the fact that this hybrid scheme can also include the 3D-Var background error covariance is an appealing advantage to implicitly take into account the model error covariance since the 3D-Var background error covariance is estimated to describe the statistics of the errors of a real, imperfect model.

The overall improvement gained by the hybrid schemes is due to its reliable flow-dependent estimate of the forecast errors or at least of their most unstable and dynamically consistent features. In fact, as demonstrated by Corazza et al. (2003), the forecast error strongly projects on the subspace spanned by the BVs. This allows to correct the background state toward the observations along the estimated unstable subspace spanned by the BVs. On the other hand, the hybrid system based on random vectors, which do not provide any dynamical information about the flow instability so that its performance is always worse than the regular 3D-Var system.

We also notice that the hybrid system with random vectors tends to worsen the large RMS analysis error spikes of the 3D-Var analysis (not shown). Large error spikes occur when the flow undergoes a rapid growth of instabilities, which are not taken into account by the constant background error covariance matrix $B$ of the 3D-Var and by the random vectors hybrid scheme. In contrast, such large analysis error growths are efficiently reduced by the hybrid scheme with BVs or final SVs, indicating that the error growths due to the time-dependent flow instabilities are properly accounted.
4.2 4D-Var and LETKF

In this subsection, we focus on the comparison between the two most advanced data assimilation schemes included in this study considered to be the new-generation in operational systems: 4D-Var and LETKF.

In Table 4, we compare the performance of LETKF with the different sizes of local patches with 4D-Var 12 and 24 hour assimilation window, in terms of the temporally and spatially averaged RMS analysis errors. Table 4 also contains the results of the regular 3D-Var and the localized hybrid system with 20 BVs. We first confirm the results of Corazza07, showing that the performance of the LETKF improves with the size of the local patch but the improvement saturates beyond a certain size of the local patch. Briefly, LETKF with the smaller local patch of 7x7 ($l=3$) outperforms the 3D-Var and a local patch of 11x11 ($l=5$) outperforms the 12-hour 4D-Var. Results obtained from local patches larger then 11x11 are on the average slightly worse than those of the 24-hour 4D-Var.

As pointed out by Miyoshi (2005), the localization of the observations error covariance reduces the impact of observations far from the center of the local domain and this becomes important when the local patch gets larger. In the present case, the observation coverage is 3% of the model domain, which may be too sparse compared to be a realistic configuration. Therefore, a larger local patch is expected to show better results when compared with 4D-Var scheme with a long assimilation window.

Table 4 also lists the computational time required for one analysis cycle for 3D-Var, hybrid, 4D-Var and LETKF with the same (serial) computational environment used in
section 3.2. The computing time does not include the time for ensemble forecasts since in practice the ensemble forecasts are included in either type of the assimilation scheme. It is important to note that the LETKF \((l=5)\) produce more accurate analyses than the 12-hour 4D-Var. On a serial computer its computational cost increases with the increase of the size of local patch but a significant advantage of the LETKF is that it is intrinsically parallel (Hunt et al. 2007) while the parallelization of variational schemes is more difficult since the minimization process in 4D-Var is solved globally.

Figure 5 shows the time series of the RMS analysis error, in the potential enstrophy norm, for the 12-hour 4D-Var, the 24-hour 4D-Var, the LETKF \((l=9)\), the 3D-Var and the hybrid system with 20 BVs (with localization). Hereafter, results of the LETKF will be shown with configuration using the local patch of \(l=9\).

In Fig. 5, the 4D-Var, LETKF and the hybrid schemes all outperform the 3D-Var and successfully avoid the occurrence of large error spikes (e.g. day 60 in Fig. 5) but large error variations occur occasionally in the hybrid scheme. We observed that the spin-up time of the ensemble-based schemes depends on the accuracy of the initial background state at the first analysis cycle while it has little influence for 4D-Var with a 24-hour assimilation window. The spin-up is much longer if LETKF analysis cycle is initialized with the climatology mean state (not shown). This is because the corrections from the ensemble-based schemes require that the ensemble perturbations be representative of the relevant background growing errors, and this can be fulfilled only when they are capable to depict the structures of the flow-dependent background dynamical instabilities. This is not a problem when we choose as starting initial condition the 3D-Var solution which is
sufficiently close to the true state, and LETKF spin-up time (~10 days) is even shorter than for the 12-hour 4D-Var (~15 days), as indicated in Fig. 5.

Figure 6 shows the analysis and forecast error as a function of forecast lead-time up to 5 day and from all the schemes. Combining Figs. 5 and 6, the results confirm that LETKF and 4D-Var have provided better corrections to the background field. The performance of LETKF is in between the results of 4D-Var with 12-hour and 24-hour assimilation window. The slope of the 3D-Var curve is the highest (errors grow faster), indicating that the corrections made to the background state in the other data assimilation scheme affect mostly the growing unstable components of the forecast errors. It is relevant to point out that the advantage of the LETKF with respect to the 12-hour 4D-Var comes through its “flow-dependent” background error covariance (both schemes assimilate exactly the same observations). By providing the 3D-Var system with the ensemble mean background state from the LETKF system, we found that the 3D-Var analysis from the modified background states are much accurate than the original 3D-Var analysis but slightly worse than the LETKF analysis (the red line in Figure 7). This result might be interpreted as implying that the effect of ensemble average in LETKF is more important than the information on "the errors of the day". However, this would not be a correct interpretation since it is this information that allows the LETKF to minimize the errors of the day and thus provide an excellent first guess to the 3D-Var analysis. Once 3D-Var is allowed to continue without further information from the LETKF background, the errors of the day start growing and the 3D-Var analysis error increases over a few days to its normal value (the green line in Figure 7). Also, the ensemble-based background error covariance that contains the structures of the growing errors indeed has
better ability to correct the background state (ensemble mean) with the available observations. Such accumulated process makes the LETKF perform better than 3D-Var.

This aspect can be further examined by comparing the spatial structure of ensemble spread obtained from LETKF with the forecast errors from both assimilation schemes. In Figure 8 (a) and (b), an arbitrarily selected time of the LETKF and the 12-hour 4D-Var background error at the 3rd model level are shown with color shading respectively; and the corresponding ensemble spread from LETKF is overlapped with contour lines on Fig. 8(a). The distributions of the background errors from LETKF and 4D-Var are very similar because the background errors are dominated by the dynamically growing errors associated with the true state. First, it is clear that LETKF has an error variance smaller than the one from the 4D-Var with 12-hour window, especially in the regions with large amplitudes, i.e. rapidly growing errors. Second, the ensemble spread has shapes corresponding to the structures of the background error, which supports that the ensemble properly represents the dynamically growing errors well. Therefore, we are justified in making corrections derived from the ensemble space even though its dimension is much smaller than the model dimension.

The success of the 24-hour assimilation window 4D-Var can be attributed to its ability to improve the background trajectory due to the increased amount of observation information within the assimilation window. Also, among the results in Figure 5, 4D-Var with 24-hour window has the least fluctuations, followed by the LETKF. The relatively large fluctuations of 4D-Var with 12-hour window reflects the fact that the correction from 4D-Var with a short window like 12 hours is largely influenced by the constant
structure of the time-independent background error covariance at initial time. As the window lengthens, the influence of the background error covariance is reduced.

To gain further insight into the differences in the analysis corrections produced by these two schemes, we compare at an arbitrary time the analysis increments from LETKF and 4D-Var (at the beginning and end of the assimilation window) to the corresponding BVs and SVs. In Figure 9, the analysis increments (color shades) from LETKF and 4D-Var are superimposed to the BVs and initial and final corresponding leading SVs for both 12 and 24 hr assimilation windows (contours) respectively. We also include a modified initial increment for the previous analysis from the LETKF (Figure 9(a)). The modified initial analysis (equivalent to a no-cost LETKF smoother) is defined using the weighted average of the ensemble as the analysis at the end of the assimilation window. This process is similar to the 4D-Var analysis increment at the beginning of the window, so that evolving this modified analysis forward will end up with the analysis directly computed from the LETKF (Kalnay et al., 2007, also see Appendix A).

From Figure 9(a) and (b), both the smoother initial increment and the analysis increment from LETKF have local structures related to the BVs structures at the corresponding time. This indicates that the shapes of the corrections and their evolutions are strongly influenced by the dynamical instabilities, which are characterized by the local properties.

The 4D-Var analysis increment at the end of the 24-hour assimilation window strongly projects onto the dominant final SV (Figure 9(f)); the 12-hour window 4D-Var analysis increment shows a slightly looser relationship with the corresponding 12-hour final SV (Figure 9(d)). The analysis increments at the initial time from 4D-Var 12 and
24-hour windows both have large-scale corrections and are more isotropic for the 12-hour window due to its stronger dependence on the initial 3DVar-like isotropic background error covariance. In other words, the 4D-Var analysis increment with long assimilation window appears to more tightly correct the growing errors within the subspace of the leading SVs.

4.3 Vertical distribution of the analysis errors from different assimilation schemes

We observe that the vertical profile of the analysis error from the regular 3D-Var scheme exhibits a strong vertical dependence with largest amplitudes at the 3rd, bottom and top levels. In the 3D-Var scheme, the background error covariance is stored in horizontal spectral coordinates and then linked by the vertical correlation among different levels, so that the performance of the 3D-Var is directly influenced by the vertical correlation matrix. In the ensemble-based schemes (like the LETKF and the hybrid schemes), on the other hand, the vertical correlation of the forecast errors is naturally considered in the background error covariance. In this section, we examine the vertical profiles of the analysis errors from all the assimilation schemes considered here.

First, we observe that the vertical profiles of the analysis errors from different schemes are all similar to the one from the 3D-Var scheme. The vertical correlation matrix in 3D-Var has small values near the innermost levels, suggesting a situation-dependent (more uncertainties) behavior (table 3.3 in Morss, 1999), and largest variance at the mid-level. Figure 10 shows the relative improvement in vertical of other schemes.
compared to the 3D-Var in terms of the incorporated potential vorticity*. Our results suggest that the hybrid, LETKF and 4D-Var all improve near the mid-level implying that all these schemes have a better ability in capturing the changing background states. Also, the vertical profiles from the 4D-Var scheme have more uniform improvements for all the levels. The relative improvement from the LETKF shows slightly less improvement at the lower levels than at the upper levels, related to the fact that we did not apply a local decorrelation in vertical as we did for the horizontal. The use of a vertically dependent inflation (Table 2) improves this result.

5. SUMMARY AND DISCUSSION

In this study, data assimilation schemes related to variational and ensemble methods are implemented in a quasi-geostrophic model. Four different schemes are discussed: 3D-Var, 4D-Var, 3D-Var hybridized with dynamically evolving vectors, and the LETKF. The goal of this study is not only to compare individual performances but also to try to understand advantages and disadvantages for practical implementation in operational systems, as we know that 4D-Var has already been implemented in several operational centers but with a rather short assimilation window, and Ensemble Kalman Filters are regarded as a possible candidate for the next phase data assimilation system. Given the same rawinsonde observations, our results are discussed in terms of the error structure in

\* The potential temperatures at the bottom and top levels are included in the PV.

\[ q_{1} = q_{1} + \frac{\partial \theta}{\partial z}, \quad q_{5} = q_{5} - \frac{\partial \theta}{\partial z} \]
different quantities for all the data assimilation experiments and results from 3D-Var are regarded as the benchmark.

Experiments with different lengths of assimilation window time confirm the results of Pires et al. (1996) that the 4D-Var performance benefits from a long assimilation window by improving the background trajectories with the given time-dependent observations. The largest improvement is obtained by increasing the assimilation window from 12 to 24-hour, beyond which the performance remains similar. With a short window time like 12-hour, 4D-Var is very sensitive to the amplitude of the initial background error. The initial background error covariance plays an important role in the corrections for the initial state and is critical for the 4D-Var performances when using a relatively short window. Concerning the computational cost in operational applications, 4D-Var with a long assimilation window time is more expensive and thus, operational centers have been using a window time short than 12 hours and reduced the model resolution or with a simplified governing physics during the minimization process. As a result, these factors limit the performance of 4D-Var. Recent studies showed that allowing the initial background error covariance to be flow-dependent has positive impact for the 4D-Var system. Beck and Ehrendorfer (2005) showed that the 4D-Var analysis can be more accurate by providing the dynamical error-covariance evolution with the singular vectors derived from a reduced-rank approach. However, the disadvantage is that such approach still requires a very large size of available singular vectors in order to represent well the dynamical evolutions in global model space.
In our study, the ensemble-based hybrid system is solved within the 3D-Var framework and the ensemble vectors are used to augment the time-independent background error. We confirm that the hybrid-3D-Var scheme can work as a shortcut to combine the advantages of 3D-Var and Ensemble Kalman Filter. If the hybridized ensemble vectors are chosen to represent well the structure of dynamically evolving background errors, the scheme can show positive impact over the 3D-Var with a small ensemble size of 20 ensemble vectors. Our results show that the hybrid system gains most improvements from using either BVs or the final SVs derived from an optimization time longer than 24 hours, both project strongly on the background errors. Augmenting the background error covariance with the initial singular vectors does not provide a positive impact to the hybrid system, even though the initial SVs are also flow-dependent and indicate the potentially growing directions after the optimization time. Therefore, we can claim that for this hybrid scheme it is the structures related to the background errors that matter in this hybrid scheme. Although BVs and final SVs provide similar benefits in the hybrid system, SVs are computationally much more expensive than BVs. We note that “refreshing” the BVs with small random perturbations (akin to additive inflation in ensemble Kalman Filter) avoids the BV collapse into a subspace that is too small and that led to the poorer results presented in Etherton et al (2004).

By allowing the system to use the ensemble information locally, the hybrid scheme performs very well with 90% of the background error covariance provided from the ensemble dynamically-evolving structures. In particular, the hybrid system can improve the regular 3D-Var system by suppressing the spurious error spikes. By contrast, random
perturbations without any information of the evolving flow make those errors even larger since rapidly growing errors associated with the initial condition are not removed.

Our results also suggest that the hybrid scheme with localization is able to achieve satisfactory results comparable to those of the 12-hour 4D-Var, at a low computational cost and with easy implementation within the 3D-Var framework. We also show that it is comparable to LETKF with a small local patch of $l=3$ with an ensemble size of 40. The success of the hybrid scheme using a rather small ensemble suggests that including the background error covariance from the 3D-Var provides an important basis for the representation of the average behavior of the background error. This may be also helpful for methods using ensemble Kalman filters that suffer from rank-deficiency or sampling problems. In addition the hybrid scheme with full model resolution can be expected to work well even in the presence of model error since the statistics of the model errors can be naturally included in the 3D-Var related background error covariance. Therefore, our results suggest a hybrid scheme could be considered as a fast upgrade in an operational center with little added cost if an ensemble forecast system is already established.

Concerning the advanced assimilation schemes, we compared the performance between the LETKF and the 4D-Var schemes. Our results indicate that given the same observations, the LETKF with a local patch of 11x11 performs better than 12-hour 4D-Var but with a lower computational cost, and its advantage remains at long forecast lead times, indicating its ability to remove growing errors from the initial states. With a local patch of 19x19, the LETKF provides a satisfactory result when compared with the 24 hour 4D-Var, but the computational cost of the LETKF is doubled when the chosen
length the local patch is doubled. This is not serious disadvantage when implementing the scheme in a parallel framework.

The structures of the analysis errors and analysis increments from these schemes were examined in this study. The analysis increment and the smoother initial increment from the LETKF both have local structures characterized with the local instabilities. Those structures also strongly related to the BVs at the corresponding times. The initial increment from 4D-Var shows projections on the dominant initial SVs but characterized with large-scale features even though the analysis increment (at the end of the window) exhibits strong similarities on the corresponding final SVs. This suggests that what needs to be corrected in the background state (at the analysis time) is strongly related to the structures of BVs and final SVs, i.e. the fast growing errors corresponding to the analysis time.

In summary, it would be natural to combine the data assimilation and ensemble forecasting process for the numerical weather prediction in order to obtain the best description of flow-dependent error covariance prediction and extract the maximum information from the observations. Ensemble forecast provides valuable accessibility to probabilistic prediction. The 4D-Var scheme, which provides a deterministic analysis state and derive the analysis error covariance implicitly, shows a significant advantage for a long assimilation window. In this study, we provide indication that it is possible for an operational center to derive a reliable analysis from their ensemble forecasting system with an affordable computational cost compared to the 4D-Var schemes. In addition, the ensemble information can be useful to transform the static error covariance of the variational schemes into flow-dependent error covariance without propagating the full-
rank error covariance. For future studies related to the comparisons between the ensemble-based and variational-based schemes, issues like the inclusion and estimation of the model errors should be carefully addressed for operational implementation.

Acknowledgements
We are deeply grateful to Rebecca Morss for providing the QG-model and the 3D-Var system and to Joaquim Ballabrera for helping to correct the adjoint model. We are very grateful to Profs. Istvan Szunyogh, Ed Ott, Brian Hunt and the other members of the Chaos and Weather Group at the University of Maryland, as well as Jeff Whitaker from NOAA/OAR/PSD for many helpful interactions. S.-C. Yang was supported by NASA grants NNG004GK78A and NNG06GB77G. A. Carrassi was supported by the Belgian Federal Science Policy Program under contract MO/34/017.

6. REFERENCES


Carrassi A., A. Trevisan and F. Uboldi, 2007: Adaptive observations and assimilation in


Szunoygh, I., Kostelich, E. J., Gyarmati, G., Patil, D. J., Kalnay, E., Ott, E. and Yorke, J.


----------------, T. M. Hamill, X. Wei, Y. Song and Z. Toth, 2007: Ensemble Data


Appendix: No-cost LETKF smoother

From (19), the mean analysis computed from the LETKF at time $i$ is:

$$\bar{x}_a^i = \bar{x}_f^i + E_f^i W^i$$  \hspace{1cm} (A.1)

where $W^i$ is a matrix of weights (A.2) and $E_f^i$ is the matrix whose columns are the ensemble background perturbations at time $i$.

$$W^i = \hat{P}_a^i (H E_f^i)^T R^{-1} (y_r^i - H(\bar{x}_f^i))$$  \hspace{1cm} (A.2)

Since the weight matrix provides the combination of background perturbations that best fit the observations within the assimilation window, we can improve the analysis at the beginning of the time window (at time $i-1$), by applying a smoother with the same weights, $W^i$, to the ensemble perturbations at time $i-1$:

$$\bar{x}_a^{i-1} = \bar{x}_a^{i-1} + E_a^{i-1} W^i$$  \hspace{1cm} (A.3)

where $\bar{x}_a^{i-1}$ and $E_a^{i-1}$ are mean analysis the ensemble analysis perturbations at time $i-1$ and $\bar{x}_a^{i-1}$ is the modified analysis at the same time. Although this smoothed analysis $\bar{x}_a^{i-1}$ at the beginning of the window does slightly improve the original analysis $\bar{x}_a^{i-1}$ because it uses the observations at time $i$, it does not change the analysis at time $i$, as indicated in (A.3). Such no-cost smoothing would be useful in the context of reanalysis (Kalnay et al. 2007).

Note that in (A.3), $M$ is the nonlinear model advancing the state from time $i-1$ to $i$ and $L$ is the tangent linear model defined in section 3.2.

$$M(\bar{x}_a^{i-1}) = M(\bar{x}_a^{i-1} + E_a^{i-1} W^i)$$

$$\approx M(\bar{x}_a^{i-1}) + L E_a^{i-1} W^i = \bar{x}_f^i + E_f^i W^i = \bar{x}_a^i$$
Table Captions

**Table 1** Mean RMS analysis error (potential vorticity) of the 4D-Var with different assimilation window lengths averaged over 30 days (60 analysis cycles).

**Table 2** Settings adopted for the LETKF system for the simulations described in the text.

**Table 3** Time and domain mean explained variance and E-dimension (see the definition in the text) of different types of ensemble vectors of potential vorticity at the 3rd level. The values are computed on a local patch with a size of 5x5 grid points.

**Table 4** Mean RMS analysis error (potential vorticity at the 3rd-level) for the 3D-Var, the Hybrid, the 4D-Var and the LETKF averaged over 30 days (60 analysis times) and the corresponding computational times required to finish one analysis cycle.
Figure Captions

Figure 1 12-hour forecast error (contour) vs. leading initial SV, final SV and BV (shaded). (a) the 1st initial SV with the optimization period of 12 hours, (b) the 1st final SV corresponding to (a), (c) the BV at the same time relevant to (b). (d)-(i) are arranged the same as (a)-(c), except the SVs are with the optimization period of 24 hours in (d) and (e), and with the optimization period of 48 hours in (g) and (h). Note BVs are chosen with no preference, and their differences are simply using different random perturbations to start the first breeding cycle (the contour lines of forecast errors in all the plots are using the same isolines).

Figure 2 Local explained variance for the background (12 hour forecast) error by 20 bred vectors (BV, denoted as the red line) and 20 singular vectors (SV, initial ones are denoted by the black lines and final ones are denoted by the blue lines) with different optimization periods (12, 24 and 48 hours), represented by different line types.

Figure 3 Time series of RMS analysis error in potential vorticity from 3D-Var (the black line) and the hybrid scheme with 20 BVs (the red line), 20 initial SVs (the dashed blue line) and 20 final SVs (the solid blue line), with the optimization time of 12 and 24 hours.

Figure 4 Mean RMS analysis error (enstrophy norm) for global hybrid 3DVAR system with 20 bred vectors (blue solid line: no localization, blue dash line: with localization) and 20 random noise vectors (red line) as a function of the coefficient \( \alpha \). The black solid
line denotes the RMS analysis error of the 3D-Var and the solid line denotes the 4D-Var RMS error with a 12-hour assimilation window. Results are averaged for horizontal and vertical domain and from day 45 to day 100. The hybrid scheme returns to regular 3D-Var as $\alpha$ equals to 0.

**Figure 5** Time series of RMS analysis errors in potential vorticity from the 3D-Var, the Hybrid, the 4D-Var (12 hour and 24 hour window time) and LETKF schemes.

**Figure 6** Analysis and forecast errors in terms of the potential enstrophy as a function of leading time (days) from different assimilation schemes.

**Figure 7** Time series of RMS analysis errors in potential vorticity from the 3D-Var, the 3D-Var but replaced with the LETKF background state, the 3D-Var but replaced with the LETKF background state during the first 50days and the LETKF schemes.

**Figure 8** (a) Background error (color shades) and the ensemble spread (the contour lines) from the LETKF scheme and (b) background error from the 4D-Var with 12-hour assimilation window at Day 41 00Z.

**Figure 9 (a)** The modified (smoothed, see text) initial increments of LETKF (color shades) and BVs (contours) at day24 12Z, (b) the same as (a) but for the LETKF analysis increments and BVs at day 25 00Z, (c) 12-hour 4D-Var, initial increment and initial SV at day 24 12Z, (d) 12-hour 4D-Var analysis increment and final SV at day25 00Z and (e) and (f) are the same as (c) and (d) but for the 24-hour 4D-Var at the corresponding time.
Figure 10 RMS analysis error improvement relative to 3D-Var in terms of the incorporated potential vorticity for all levels.
Table 1 Mean RMS of analysis errors (potential vorticity) of the 4D-Var with different assimilation window lengths averaged over 30 days (60 analysis cycles).

<table>
<thead>
<tr>
<th></th>
<th>3D-Var</th>
<th>4D-Var</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>12 hr</td>
</tr>
<tr>
<td>Ana. Error</td>
<td>0.011</td>
<td>0.0046</td>
</tr>
</tbody>
</table>
Table 2 Settings adopted for the LETKF system for the simulations described in the text

<table>
<thead>
<tr>
<th>Horizontal dimension of the local domain</th>
<th>19x19 ($l=9$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of ensemble members</td>
<td>$K=40$</td>
</tr>
<tr>
<td>Method to update the global field</td>
<td>Center point</td>
</tr>
<tr>
<td>Amplitude of the random perturbations</td>
<td>2% vectors amplitude</td>
</tr>
<tr>
<td>Multiplicative variance inflation</td>
<td>1.08 1.06 1.07 1.08 1.075 1.07 1.14</td>
</tr>
<tr>
<td>Decorrelation length</td>
<td>$r_d=7$</td>
</tr>
</tbody>
</table>
Table 3 Time and domain mean explained variance and E-dimension (see the definition in the text) of different types of ensemble vectors of potential vorticity at the 3rd level. The values are computed on a local patch with a size of 5x5 grid points.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>10 ensemble</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explained variance</td>
<td>0.85</td>
<td>0.78</td>
<td>0.73</td>
<td>0.69</td>
<td>0.80</td>
<td>0.81</td>
<td>0.83</td>
</tr>
<tr>
<td>E-dim</td>
<td>3.13</td>
<td>2.56</td>
<td>2.53</td>
<td>2.88</td>
<td>2.62</td>
<td>2.65</td>
<td>2.90</td>
</tr>
<tr>
<td>20 ensemble</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explained variance</td>
<td>0.99</td>
<td>0.96</td>
<td>0.95</td>
<td>0.94</td>
<td>0.97</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>E-dim</td>
<td>3.54</td>
<td>3.15</td>
<td>3.14</td>
<td>3.54</td>
<td>3.02</td>
<td>3.08</td>
<td>3.41</td>
</tr>
</tbody>
</table>
Table 4 Mean RMS analysis errors of (potential vorticity at the 3rd-level) for 3D-Var, Hybrid, 4D-Var and LETKF schemes averaged over 30 days (60 analysis cycles) and the corresponding computational times required to finish one analysis cycle.

<table>
<thead>
<tr>
<th></th>
<th>3D-Var</th>
<th>HYBD</th>
<th>4D-Var</th>
<th>LETKF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>l=3</td>
<td>l=5</td>
</tr>
<tr>
<td>RMS error ((\times 10^{-2}))</td>
<td>1.44</td>
<td>0.70</td>
<td>0.56</td>
<td>0.35</td>
</tr>
<tr>
<td>Time ((\text{minutes}))</td>
<td>0.15</td>
<td>1.5</td>
<td>1.8</td>
<td>2.5</td>
</tr>
</tbody>
</table>
Figure 112-hour forecast error (contour) vs. leading initial SV, final SV and BV (shaded). (a) the 1st initial SV with the optimization period of 12 hours, (b) the 1st final SV corresponding to (a), (c) the BV at the same time relevant to (b). (d)-(i) are arranged the same as (a)-(c), except the SVs are with the optimization period of 24 hours in (d) and (e), and with the optimization period of 48 hours in (g) and (h). Note BVs are chosen with no preference, and their differences are simply using different random perturbations to start the first breeding cycle (the contour lines of forecast errors in all the plots are using the same isolines).
Figure 2 Local explained variance for the background (12 hour forecast) error by 20 bred vectors (BV, denoted as the red line) and 20 singular vectors (SV, initial ones are denoted by the black lines and final ones are denoted by the blue lines) with different optimization periods (12, 24 and 48 hours), represented by different line types. Local explained variance is defined as the square of the cosine of the local angle between the background errors and the subspace spanned by BVs/SVs.
Figure 3 Time series of RMS analysis error in potential vorticity from 3D-Var (the black line) and the hybrid scheme with 20 BVs (the red line), 20 initial SVs (the dashed blue line) and 20 final SVs (the solid blue line), with the optimization time of 12 and 24 hours.
Figure 4 Mean RMS analysis error (enstrophy norm) for global hybrid 3DVAR system with 20 bred vectors (blue solid line: no localization, blue dash line: with localization) and 20 random noise vectors (red line) as a function of the coefficient $\alpha$. The black solid line denotes the RMS analysis error of the 3D-Var and the solid line denotes the 4D-Var RMS error with a 12-hour assimilation window. Results are averaged for horizontal and vertical domain and from day 45 to day 100. The hybrid scheme returns to regular 3D-Var as $\alpha$ equals to 0.
Figure 5 Time series of RMS analysis errors in potential vorticity from the 3D-Var, the Hybrid, the 4D-Var systems (12 hour and 24 hour window time) and the LETKF schemes.
Figure 6 Analysis and forecast errors in terms of the potential enstrophy and as a function of leading time (days) from different assimilation schemes.
Figure 7 Time series of RMS analysis errors in potential vorticity from the 3D-Var, the 3D-Var but replaced with the LETKF background state, the 3D-Var but replaced with the LETKF background state during the first 50 days and the LETKF schemes.
Figure 8 (a) Background error (color shades) and the ensemble spread (the contour lines) from the LETKF scheme and (b) background error from the 4D-Var with 12-hour assimilation window at Day 41 00Z.
Figure 9 (a) The modified (smoothed, see text) initial increments of LETKF (color shades) and BVs (contours) at day 24 12Z, (b) the same as (a) but for the LETKF analysis increments and BVs at day 25 00Z, (c) 12-hour 4D-Var, initial increment and initial SV at day 24 12Z, (d) 12-hour 4D-Var analysis increment and final SV at day 25 00Z and (e) and (f) are the same as (c) and (d) but for the 24-hour 4D-Var at the corresponding time.
Figure 10 RMS analysis error improvement relative to 3D-Var in terms of the incorporated potential vorticity for all levels.