Tellus (2008), 60A, 348–360
Printed in Singapore. All rights reserved

TELLUS

A study of ensemble size and shallow water dynamics with the Maximum Likelihood Ensemble Filter

By S. J. FLETCHER* and M. ZUPANSKI, *Cooperative Institute for Research in the Atmosphere, Colorado State University, 1375 Campus Delivery, Fort Collins, CO 80523-1375, USA*

(Manuscript received 26 March 2007; in final form 9 November 2007)

ABSTRACT

In this paper we perform a study using the Maximum Likelihood Ensemble Filter (MLEF), developed at the Cooperative Institute for Research in the Atmosphere (CIRA), Colorado State University (CSU) and Florida State University (FSU), with CSU's two-dimensional shallow water equations model on the sphere. The aim of this study is to find the optimal number of ensemble members, with respect to the root mean square (rms) error of the three state variables in the shallow water equations model, such that the rms error is a minimum. After ascertaining this number, we vary the size of the observations sets for three different Rossby–Haurwitz waves in the shallow water equations model. These waves generate different types of dynamics from fast, shallow motions to fast, tall waves with vortices, to a flow similar to geostrophic balance. We show that, for faster flows, we require less ensemble members than for slow balanced flows. We also present an explanation for this behaviour in the form of hybrid Lyapunov-bred vectors.

1. Introduction

In the field of numerical weather prediction there are two main assimilation methods used in operational or research centres. The first set of methods are the variational (VAR) data assimilation methods (Parish and Derber, 1992; Lorenc et al., 2000; Rabier et al., 2000; Zupanski et al., 2005; Rawlins et al., 2007), where a non-linear cost function is minimized to find the most likely state, mode, of the analysis probability density function (pdf, Lorenc, 1986). The second set of methods are the ensemble and square-root approximations to the Kalman filter (Bishop et al., 2001; Evenson, 2003; Tippett et al., 2003), where the solution is an approximation to the mean of the analysis distribution. Currently these latter methods are more in use in research centres, rather than at the operational centres.

A problem associated with the ensemble based methods is the question of whether or not the optimal number of ensemble members required to obtain a desired order of accuracy in our solutions is dependent on: (1) the dynamic flow of the numerical model or (2) the number of observations taken of the system. Nearly all of the ensemble-based methods are reliant on the sampling structure. Sampling here is meant in the statistical sense where a set of data are taken and used to approximate the moments of the distribution of the random variable. In the ensemble filters an approximation to the true mean is found through an additive average over all the states in the different ensemble members. From this a sample approximation is made to the second order moment, variance/covariance. This implies that the optimal solution could be quite large to fully sample the distribution.

The Maximum Likelihood Ensemble Filter (MLEF, Zupanski, 2005; Zupanski and Zupanski, 2006; Zupanski et al., 2006), does not calculate a sample mean from the ensemble members. The analysis state that the MLEF seeks is the mode, which for linear dynamics and Gaussian statistics is the same as the mean. The mode is found through minimizing a cost function, similar to that in the three-dimensional variational assimilation method (Lorenc, 1986), but projected onto ensemble space, rather than model space. Therefore, this filter is dependent on the underlying dynamics of the flow. More mathematical details about the filter are given in Section 2.

As a result of this difference in the MLEF to other square root and ensemble filters, it is desirable to see how the MLEF performs with more advanced models. The MLEF has been successfully tested with the Korteweg-de Vries-Burgers model, which includes non-linear advection, dispersion and diffusion (Zupanski, 2005), where both linear and non-linear observation operators were used. It has also been tested with Colorado State University's (CSU) shallow water equations model (Heikes and Randall, 1995a,b; Ringler and Randall,

*Corresponding author.
e-mail: fletcher@cira.colostate.edu
DOI: 10.1111/j.1600-0870.2007.00294.x

2002), with test case five from Williamson et al. (1992), which is the geostrophically balanced zonal flow over an isolated conical mountain (Zupanski et al., 2006). In none of these studies has the impact of the number of ensemble members been investigated.

It should be noted that the two test cases above are quite simple models. Therefore, in this paper a study is performed to investigate the impact on the number of ensemble members required for quite accurate analyses in the MLEF, with more advanced initial conditions for the same shallow water equations model. These alternative initial conditions are different Rossby–Haurwitz (RH) waves (Hoskins, 1973; Williamson et al., 1992), which generate different type of flows (Marek, 2002; Fletcher, 2004; Wlasak et al., 2006). These flows differ from a fast, tall wave, which starts to diverge in the model during the time of the assimilation method, to a tall, slow wave, which is similar to geostrophic balance. The reason why this tall, slow wave is similar to geostrophic balance is because the height profile does not change much over the domain and hence it does not have the feature to diverge. The final wave is a short, fast wave which generates divergence in the flow. More mathematical details about the initial conditions are given in Section 4.

In the experiments undertaken, the optimal ensemble size is defined as the number of ensemble members such that the root mean square (rms) error is smallest. A second indicator is the χ^2 statistic, as defined in Zupanski (2005). This statistic should be approximately one if the ensembles are using the observations, given their uncertainties, along with the background state, and its associated uncertainties. These measures are defined in Section 3.

The remainder of the paper is as follows. Section 5 has a description of the two sets of experiments that are performed with the filter. The first experiment investigates the number of ensemble members, given a fixed number of observations, required for the errors to be minimized. The second experiment involves a fixed number for the ensemble size with different sizes of observations sets.

In Section 6, results are presented from the experiments described in Section 5. These results are presented in a series of rms error and χ^2 plots. The results that were obtained were not as expected if the MLEF was a sample based filter. For sample based filters, the true evolution is observed to be different from the members of the ensemble, in which the forecast errors are dominated by the problems in the forecasting system, such as model deficiencies, rather than chaotic growth. Therefore, in Section 7 a brief summary of the definitions of Lyapunov vectors, and bred vectors (Kalnay et al., 2002), is provided. From these definitions, it is shown that these structures are implicitly present in the MLEF. Therefore, the structures that are observed in the MLEF are consistent with these dynamic features.

This paper is finished with some conclusions and plans for further work in Section 8.

2. Maximum Likelihood Ensemble Filter (MLEF)

This section provides a brief summary of the MLEF from Zupanski (2005), as well as the initiation procedure from Zupanski et al. (2006).

2.1. Initiation step

The procedure that is described below is an initiation technique and not an initialization method.

As with all ensemble filters, the method from which the initial analysis error covariance matrix is calculated is a complicated choice. It has been shown in Zupanski et al. (2006), that using random perturbations, which could be correlated or uncorrelated, effects the analysis of the MLEF at latter cycles. In Zupanski et al. (2006) a more structured approach is presented.

In an ensemble filter, it is desirable to have ensemble members that do not excite spurious gravity modes in the model, or to have perturbations that are unphysical. Both approaches could cause the filter to diverge.

A more dynamically based approach to generate the seeds for the ensemble members is described in Zupanski et al. (2006). This technique, which is used for the experiments shown in this paper, is based upon using the Kardar–Parisi–Zhang (KPZ) equation

$$\frac{\partial \phi}{\partial t} = \frac{\partial^2 \phi}{\partial x^2} + \left(\frac{\partial \phi}{\partial x}\right)^2 + \zeta(x, t), \tag{1}$$

where ϕ is the perturbation and ζ is a random forcing. In the experiments presented later, this forcing is a normal distribution with unit variance.

The reason for choosing this equation is that it is related to the dynamic localization of the Lyapunov vectors. Lyapunov vectors can be viewed as the exponentials of the roughened interface (Pikovsky and Politi, 1998). The KPZ equation generates spatially sparse, uncorrelated, random perturbations. These perturbations are then smoothed through a space-limited compactly supported function (Gaspari and Cohn, 1999). The correlation length should be realistic to the spatial scales in the atmosphere, or the associated dynamics, that are being modelled.

An important feature to note here is that this approach does not yield the Lyapunov vectors of the shallow water equations models. This approach generates spatial perturbations which are smoothed to a pre-defined correlation length. A full, more detailed study of this approach with the shallow water equations model initialized with test case 5 from Williamson et al. (1992), can be found in Zupanski et al. (2006).

The perturbations are calculated for each of the prognostic variables of the model at each latitudinal ring, for spherical coordinates, and for each ensemble member.

Once the perturbations are generated and smoothed, as described above, they are added to the ensemble members at a pre determined time before the start of the first assimilation cycle. The ensemble members are then ran to the start of the first assimilation window. At this time the initial analysis square root error covariance matrix is calculated. This is described in the next subsection.

2.2. Forecast step

The MLEF comprises of two different stages. The first, the forecast step, is concerned with the evolution of the forecast error covariances. The starting point for this is from the evolution equation of the discrete Kalman filter (Jazwinski, 1970). This equation is given by

$$\mathbf{P}_f(k) = \boldsymbol{\mathcal{M}}_{k-1,k}\mathbf{P}_a(k-1)\boldsymbol{\mathcal{M}}_{k-1,k}^T + \mathbf{Q}(k-1), \tag{2}$$

where $\mathbf{P}_f$ is the forecast error covariance matrix, k is the time index, $\boldsymbol{\mathcal{M}}$ is the non-linear model evolution operator and $\mathbf{Q}$ is the model error matrix which is assumed to be normally distributed. For the purpose of this work this is assumed to be zero. The discrete index, k, is dropped for the remainder of the paper.

A factorization of $\mathbf{P}_f$ into a square root form can be defined as

$$\mathbf{P}_f = \boldsymbol{\mathcal{M}}\mathbf{P}_a\boldsymbol{\mathcal{M}}^T = \left(\boldsymbol{\mathcal{M}}\mathbf{P}_a^{\frac{1}{2}}\right)\left(\boldsymbol{\mathcal{M}}\mathbf{P}_a^{\frac{1}{2}}\right)^T = \mathbf{P}_f^{\frac{1}{2}}\mathbf{P}_f^{\frac{T}{2}}. \tag{3}$$

The structure of the square-root analysis error covariance matrix, $\mathbf{P}_a^{\frac{1}{2}}$, is

$$\mathbf{P}_a^{\frac{1}{2}} = \begin{pmatrix}\mathbf{p}_1 & \mathbf{p}_2 & \dots & \mathbf{p}_S\end{pmatrix} \quad \text{where} \quad \mathbf{p}_i = \begin{pmatrix} p_{1,i} \\ p_{2,i} \\ \vdots \\ p_{N,i}\end{pmatrix}, \tag{4}$$

where N is the number of state variables, S is the number of ensemble members with the assumption that $S \ll N$.

Upon expanding (4), the square root forecast error covariance matrix, $\mathbf{P}_f^{\frac{1}{2}}$, can be expressed as

$$\begin{aligned}\mathbf{P}_f^{\frac{1}{2}} &= \begin{pmatrix}\mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_S\end{pmatrix} \\ \mathbf{b}_i &= \boldsymbol{\mathcal{M}}\left(\mathbf{x}_{k-1} + \mathbf{p}_i\right) - \boldsymbol{\mathcal{M}}\left(\mathbf{x}_{k-1}\right) \approx \boldsymbol{\mathcal{M}}\mathbf{p}_i,\end{aligned} \tag{5}$$

where $\mathbf{x}_{k-1}$ is the analysis state from the previous assimilation cycle, which is found from the posterior analysis pdf (Lorenc, 1986). Therefore, the MLEF evolves the square root analysis error covariance matrix through the ensembles.

2.3. Analysis step

The analysis step for the MLEF involves solving a non-linear cost function, similar to that of Lorenc (1986), which is based upon a normal assumption for the background variables and the observations.

The associated cost function is defined in terms of $\mathbf{P}_f$ although this matrix is never calculated or stored in the process of the filter. This then results in

$$\begin{aligned}J(\mathbf{x}) = &\frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T\mathbf{P}_f^{-1}(\mathbf{x}-\mathbf{x}_b) \\ &+\frac{1}{2}\left[\mathbf{y}-\mathbf{h}(\mathbf{x})\right]^T\mathbf{R}^{-1}\left[\mathbf{y}-\mathbf{h}(\mathbf{x})\right],\end{aligned} \tag{6}$$

where $\mathbf{y}$ is the vector of observations, $\mathbf{h}$ is the non-linear observation operator, $\mathbf{R}$ is the observational covariance matrix and $\mathbf{x}_b$ is a background state, such that $\mathbf{x}_b = \boldsymbol{\mathcal{M}}(\mathbf{x}_{k-1})$.

To find the minimum of (6), a change of variable is introduced through a Hessian pre-conditioner, defined by

$$\mathbf{x} - \mathbf{x}_b = \mathbf{P}_f^{\frac{1}{2}}(\mathbf{I}+\mathbf{C})^{-\frac{T}{2}}\boldsymbol{\xi}, \tag{7}$$

where $\boldsymbol{\xi}$ is our vector of control variables, defined in ensemble subspace and $\mathbf{C}$ is the Hessian matrix of (6) which is

$$\mathbf{C} = \mathbf{P}_f^{\frac{T}{2}}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{P}_f^{\frac{1}{2}} = \left(\mathbf{R}^{-\frac{1}{2}}\mathbf{H}\mathbf{P}_f^{\frac{1}{2}}\right)^T\left(\mathbf{R}^{-\frac{1}{2}}\mathbf{H}\mathbf{P}_f^{\frac{1}{2}}\right), \tag{8}$$

where $\mathbf{h}$ is the Jacobian matrix of $\mathbf{h}$ evaluated at $\mathbf{x}_b$.

It may be the case that the observation operator is non-linear, difficult to analytically differentiate, or even discontinuous. To overcome this, information from $\mathbf{P}_f$ is used to approximate the square root of $\mathbf{C}$, componentwise, as

$$\begin{aligned}\mathbf{z}_i &= \left(\mathbf{R}^{-\frac{1}{2}}\mathbf{H}\mathbf{P}_f^{\frac{1}{2}}\right)_i = \mathbf{R}^{-\frac{1}{2}}\mathbf{H}\mathbf{b}_i \\ &\approx \mathbf{R}^{-\frac{1}{2}}\left[\mathbf{h}\left(\mathbf{x}+\mathbf{b}_i\right) - \mathbf{h}\left(\mathbf{x}\right)\right].\end{aligned} \tag{9}$$

A new matrix $\mathbf{Z}$ is now defined such that

$$\mathbf{Z} = \begin{pmatrix}\mathbf{z}_1 & \mathbf{z}_2 & \dots & \mathbf{z}_S\end{pmatrix}. \tag{10}$$

The definition above allows $\mathbf{C}$ to be written as $\mathbf{C} = \mathbf{Z}^T\mathbf{Z}$, that is,

$$\mathbf{C} = \begin{pmatrix}\mathbf{z}_1^T\mathbf{z}_1 & \mathbf{z}_1^T\mathbf{z}_2 & \dots & \mathbf{z}_1^T\mathbf{z}_S \\ \mathbf{z}_2^T\mathbf{z}_1 & \mathbf{z}_2^T\mathbf{z}_2 & \dots & \mathbf{z}_2^T\mathbf{z}_S \\ \vdots & \vdots & ots & \vdots \\ \mathbf{z}_S^T\mathbf{z}_1 & \mathbf{z}_S^T\mathbf{z}_2 & \dots & \mathbf{z}_S^T\mathbf{z}_S\end{pmatrix}. \tag{11}$$

To accomplish the inversion of $(\mathbf{I}+\mathbf{C})$, required in (7), the spectral theorem for Hermition matrices (Strang, 1980), is applied. The theorem allows for an orthogonal eigenvalue decomposition of $\mathbf{C}$ in the form

$$\mathbf{C} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^T,$$

where $\mathbf{V}$ is a matrix whose columns are orthogonal eigenvectors, and $\boldsymbol{\Lambda}$ is a diagonal matrix containing the eigenvalues of $\mathbf{C}$.

The final point about the MLEF is the updating of the square root analysis error covariance matrix by

$$\mathbf{P}_a^{\frac{1}{2}} = \mathbf{P}_f^{\frac{1}{2}}\left[\mathbf{I}+\mathbf{C}(\mathbf{x}_{\text{opt}})\right]^{-\frac{T}{2}}, \tag{12}$$

where $\mathbf{x}_{\text{opt}}$ is approximately the minimum of (6). This update to the covariance matrix is similar in appearance to that in Bishop et al. (2001), but the main difference being that we have not restricted the observation operator to be linear. More details about the MLEF and other similar algorithms can be found in Zupanski (2005). Another important difference is that, by using the cost function, we are able to allow for non-Gaussian errors. This is possible through the cost functions defined in Fletcher and Zupanski (2006a,b). One last comment about the difference between the MLEF and some of the other ensemble filters is that, by using (6), then the minimization provides the iterative solution to the non-linear analysis problem, rather than assuming a linear solution, as is the case in the Kalman filter.

This definition plays an important part in the filter and more explanation is given in Section 7. The final part of the filter is to use the columns of the $\mathbf{P}_a^{\frac{1}{2}}$ matrix to initiate the ensemble members for the next forecast cycle.

An important feature to note here is that the MLEF does not calculate an ensemble mean at any point.

3. Performance statistics

As mentioned in the introduction, there are two statistics that are used to identify the optimal ensemble size. The definitions for these two statistics are presented in this section.

3.1. Root mean square error

The first statistic is the root mean square (rms) error. This statistic is calculated globally for the three components of the shallow water equations model, geopotential height, h, and the two horizontal wind components, u and v, as

$$e_h = \sqrt{\frac{1}{N_h}\sum_{i=1}^{N_h}\left(h_i^n - h_i^t\right)^2}, \tag{13}$$

$$e_u = \sqrt{\frac{1}{N_u}\sum_{i=1}^{N_u}\left(u_i^n - u_i^t\right)^2}, \tag{14}$$

$$e_v = \sqrt{\frac{1}{N_v}\sum_{i=1}^{N_v}\left(v_i^n - v_i^t\right)^2}, \tag{15}$$

where the superscript n represents the discrete model output from the analysis step of the MLEF, and t represents the 'true' solution, which is a run of the numerical model with no errors added.

If the filter is performing well then it is expected that these statistics should be tending to zero.

3.2. χ^2 statistic

The version that is used with the MLEF is defined in observational space (Ménard et al., 2000; Zupanski, 2005), as

$$\chi^2 = \frac{1}{N_o}\left[\mathbf{y}_k - \mathbf{h}\left(\mathbf{x}_k^n\right)\right]^T\left(\mathbf{HP}_f\mathbf{H}^T + \mathbf{R}\right)^{-1}\left[\mathbf{y}_k - \mathbf{h}\left(\mathbf{x}_k^n\right)\right], \tag{16}$$

which can be rewritten as

$$\chi^2 = \frac{1}{N_o}\left\{\mathbf{R}^{-\frac{1}{2}}\left[\mathbf{y}_k - \mathbf{h}\left(\mathbf{x}_k^n\right)\right]\right\}^T\mathbf{G}^{-1}\left\{\mathbf{R}^{-\frac{1}{2}}\left[\mathbf{y}_k - \mathbf{h}\left(\mathbf{x}_k^n\right)\right]\right\}, \tag{17}$$

where

$$\mathbf{G} = \left(\mathbf{I} + \mathbf{R}^{-\frac{1}{2}}\mathbf{HP}_f\mathbf{HR}^{-\frac{T}{2}}\right). \tag{18}$$

The matrix $\mathbf{G}$ in (18) is manipulated in Zupanski (2005) to be in the form of the $\mathbf{C}$ matrix's components defined in the Hessian pre-conditioner. Therefore, the inversion of $\mathbf{G}$ can be obtained through the eigenvalue decomposition, which is already calculated in the process of the filter.

The reason to use the χ^2 statistic is to test how well the observed quantities are matching the expected values from the model Ménard et al. (2000). This is defined in terms of the innovation statistic, $\mathbf{y} - \mathbf{h}(\mathbf{x})$, and the associated covariance matrix, $\mathbf{HP}_f\mathbf{H}^T + \mathbf{R}$. If the innovation statistics are Gaussian distributed, as assumed in the experiments here, then the statistic, χ^2, is chi-squared distributed with degrees of freedom equal to the number of observations (Ménard et al., 2000). Therefore, by normalizing with respect to the number of observations, the statistic should be equal to one, if the covariance matrices are properly defined, the analysis is quite accurate and is matching the observations.

However, it could be the case that the Gaussian assumption may not be true, or that the covariance matrices have not been correctly defined. Therefore, it could appear that this statistic is diverging. However, it is the rms error statistic that is the primary metric that is used to asses the optimal ensemble size.

4. Shallow water equations model and initial conditions

In the experiments that are presented in this paper, the spherical version of the shallow water equations are used. These comprise of the horizontal momentum equation, and the continuity equations which are given by

$$\frac{\partial u}{\partial t} + \frac{u}{a\cos\theta}\frac{\partial u}{\partial\lambda} + \frac{v}{a}\frac{\partial u}{\partial\theta} - \frac{\tan\theta}{a}vu - fv = -\frac{g}{a\cos\theta}\frac{\partial h}{\partial\lambda}, \tag{19}$$

$$\frac{\partial v}{\partial t} + \frac{u}{a\cos\theta}\frac{\partial v}{\partial\lambda} + \frac{v}{a}\frac{\partial v}{\partial\theta} + \frac{\tan\theta}{a}u^2 + fu = -\frac{g}{a}\frac{\partial h}{\partial\theta}, \tag{20}$$

$$\frac{\partial h}{\partial t} + \frac{u}{a\cos\theta}\frac{\partial h}{\partial\lambda} + \frac{v}{a}\frac{\partial h}{\partial\theta} + \frac{h}{a\cos\theta}\left[\frac{\partial u}{\partial\lambda} + \frac{\partial(\cos\theta)}{\partial\theta}\right] = 0, \tag{21}$$

where a is the radius of the Earth, h is the geopotential height, u and v are the horizontal wind components, g is acceleration due to gravity, θ is the angle of latitude, where $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, λ is the angle of longitude, where $\lambda \in [0, 2\pi)$, and f is the Coriolis parameter, given by $f = 2\Omega \sin\theta$, where Ω is the Earth's rotation rate.

It is the discrete versions of (19)–(21) that is used in the experiments. The numerical model that is used is CSU's model. This model uses a geodesic grid, as described in Heikes and Randall (1995a,b), with the improved physics (Ringler and Randall, 2002). The numerical grid comprises of tessellating hexagons and pentagons to represent a sphere. The three components, geopotential height and the horizontal wind components, are stored at different parts of the grid. The heights are stored in the centres, while the winds are stored at the corners of the polygons.

The initial condition that is used in the experiments is test case 6 from Williamson et al. (1992). This test case is the Rossby–Haurwitz wave. The initial profile of the wave is given by

$$h = \frac{1}{g}\left[gh_0 + a^2 A(\theta) + a^2 B(\theta)\cos\lambda + a^2 C(\theta)\cos 2R\lambda\right], \tag{22}$$

$$u = a\omega\cos\theta + aK\cos^{R-1}\theta\left(R\sin^2\theta - \cos^2\theta\right)\cos R\lambda, \tag{23}$$

$$v = -aKR\cos^{R-1}\theta\sin\theta\sin R\lambda, \tag{24}$$

where

$$\begin{aligned} A(\theta) &= \frac{\omega}{2}(2\Omega+\omega)\cos^2\theta + \frac{1}{4}K^2\cos^{2R}\theta\left[(R+1)\cos^2\theta\right. \\ &\quad \left. + \left(2R^2 - R - 2\right) - 2R^2\cos^{-2}\theta\right], \\ B(\theta) &= \frac{2(\Omega+\omega)K}{(R+1)(R+2)}\cos^R\theta[(R^2+2R+2) \\ &\quad -(R+1)^2\cos^2\theta], \\ C(\theta) &= \frac{K^2}{4}\cos^{2R}\theta[(R+1)\cos^2\theta - (R+2)]. \end{aligned}$$

The four parameters in the definitions above, h_0, K, R and ω determine different flows of the wave. The first parameters is the initial height field at the poles, h_0. This value is the minimum of the height field across the whole domain. The strength of the underlying zonal winds, from West to East, is given by ω, K controls the amplitude of the wave, while R is the wavenumber.

The wave is not a solution of the shallow water equations, but is the analytical solution of the barotropic vorticity equations. In the barotropic vorticity model the wave is non-divergent. However, this is not the case when the wave is used as an initial condition for the shallow water equations model.

5. Experiments

In this section, the experiments that are performed to find the optimal number of ensemble members required for the MLEF, given the three different RH waves, is outlined. For all three waves we use $R = 4$, due to this being the highest, stable wavenumber for the RH wave (Hoskins, 1973). The three waves that are used, and their associated parameter values, are presented in Table 1.

Table 1. Description of the three parameters in the Rossby–Haurwitz wave

Test case	Initial height, h_0	Wave strength, ω	Wave amplitude, K
TC1	75 m	$7.848 \times 10^{-7}\,\mathrm{s}^{-1}$	$7.848 \times 10^{-7}\,\mathrm{s}^{-1}$
TC2	8000 m	$7.848 \times 10^{-6}\,\mathrm{s}^{-1}$	$7.848 \times 10^{-6}\,\mathrm{s}^{-1}$
TC3	8000 m	$7.848 \times 10^{-7}\,\mathrm{s}^{-1}$	$7.848 \times 10^{-7}\,\mathrm{s}^{-1}$

The flows that the three test cases generate have different Burger numbers associated with them. This number, given by

$$B_u = \frac{gH}{f^2L^2},$$

where g is acceleration due to gravity, H is some vertical length scale, f is the Coriolis parameter and L is a horizontal length scale, represents the relative importance of the effects of stratification, and rotation. The larger the number, the more stratified the fluid. For flows with Burger number less than one, then the flow is dominated by rotation.

The other measure that can be considered for the three flows presented is the Rossby number, given by

$$R_o = \frac{U}{fL},$$

where U is a velocity scale. The significance of this number is to measure the effects the rotation of the Earth has on the flow. If $R_o \ll 1$ then the flow is dominated by geostrophic balance (Holton, 1992). Geostrophic balance is interpreted as rotational, non-divergent flow.

If we consider the three test cases that are used in the experiments, then for TC1, B_u is less than 1, but R_o is less than one. Therefore we have a stratisfied flow, near to geostrophic balance. For TC2, $B_u \gg 1$ and $R_o > 1$, therefore the flow is stratified and divergent. This is due to the faster velocities for this test case. The flow that is generated by TC3 has $B_u \gg 1$ and $R_o \ll 1$. Therefore this flow is a balanced, geostrophic flow.

The initial conditions for TC1 generates a fast, shallow wave. The shalllowness of the wave is referring to the base height of the wave at the pole being 75 m. Therefore, the maximum height value is 325 m. The slowness referred to here arises for the range of the values the winds fields have, which for the u component this is from 0 to 10 ms^{-1}. For the v component the range is from -5 to 5 ms^{-1}.

The initial conditions for TC2 generates a tall, fast wave. The tallness referred to here is due to the base height at the pole here being 8000 m. The maximum value for the height field is 10 500 m. The fastness referred to here is again due to the range

Table 2. Description of the uncertainties in the observations

Test case	$\sigma_{o,h}$	$\sigma_{o,u}$	$\sigma_{o,v}$
TC1	15 m	0.5 ms^{-1}	0.5 ms^{-1}
TC2	500 m	5 ms^{-1}	5 ms^{-1}
TC3	15 m	0.5 ms^{-1}	0.5 ms^{-1}

of the speed of the horizontal winds, which for the u component this varies from 0 to 100 ms^{-1}, while the v component varies from -50 to 50 ms^{-1}.

The initial conditions for TC3 generates a tall, slow wave. As with TC2, this wave's base height is 8000 m but only has a maximum height of 8250 m. The range of the wind speeds for this wave are the same as for TC1. This generates a flow which is almost non-divergent and hence is similar to geostrophic flow.

The first experiment is to identify the optimal number of ensemble members, given a set of observation of size $N_0 = 1025$. This set of observations comprises of 513 height observations and 512 wind observations. This is seen as the control run for all three test cases. The uncertainties, $\sigma_{o,h}$, for the height observations and $\sigma_{o,u}$ and $\sigma_{o,v}$ for the two horizontal wind component observations, for all three test cases, are summarized in Table 2.

The experiment consists of running the MLEF using the initiation procedure, as described in Zupanski et al. (2006). The following choices for the number of ensemble members S: 40, 80, 120, 160, 200, 400 and 800 are used. The reason for the multiples of 40 is so that the nodes on the BLUESKY supercomputer at the National Center for Atmospheric Research are efficiently used. These choices also give a wide range of approximations, but will also help to illustrate that the MLEF is not a sample based filter, that is, we should not obtain better results with more ensemble members.

In the second set of experiments the observational errors, as defined in Table 2, are still used for the three test cases, but now the number of observations is changed. The breakdown of the observation sets can be found in Table 3, where $N_{o,h}$ is the number of height observations, and $N_{o,u}$ and $N_{o,v}$ are the number of u and v wind components observations respectively. In the experiment

Table 3. Description of the sets of observations for Experiment 2

N_o	$N_{o,h}$	$N_{o,u}$	$N_{o,v}$
769	513	128	128
1025	513	256	256
1366	342	512	512
1537	513	512	512

the observation operator is linear, that is, we are observing the state variables directly, but sparsely.

The number of observations may again appear random, however, this is due to the icosahedral grid that the discrete approximation to the shallow water equations model uses. As a consequence of this, it is not possible to exactly halve the number of height observations. The control run set of observations is of size $N_o = 1025$, as used in Experiment 1.

The observations for both sets of experiments are generated from a run of the model without any error added to the initial conditions. Moreover, the error is added to the output at the time of the analysis. The errors are sampled from a Gaussian distribution. It is assumed that the observational errors are uncorrelated, both in time and space. Therefore **R** is a diagonal matrix.

6. Results

In this section we present plots of the height rms errors, along with the u and v wind field errors, first, followed by the χ^2 plots, for all three test cases.

6.1. Experiment 1

We start with TC1 where, from Fig. 1, it is quite clear that the errors associated with all of the prognostic variables for the smaller ensemble sizes, have an oscillatory structure to them, $S = 40$, 80 and 120. This structure is not present when there are 200, or more, ensemble members.

The first criterion that is used to asses the optimal ensemble size is that the rms error should be less than the observational error, Table 2. We see that for all of the ensemble sizes, the rms error is less than the observational error within a few cycles for the height field, Fig. 1a. For the two wind components, Fig. 1b and c, we see that it is only for the larger ensemble sizes that rms error is less than or equal to the observational error. The second criterion is the obvious one, which is the ensemble size with the smallest rms error. For TC1 it appears that the optimal ensemble size is $S = 400$ for all three prognostic fields.

The χ^2 plot for this test case is presented in Fig. 4a. From this figure it is clear that only when $S = 400$ does the statistic stay approximately equal to 1. This is suggesting that the ensemble members are using the information from the observations.

As has been mentioned before, the wave generated in TC2 has quite different dynamics to that of TC1. This Rossby–Haurwitz wave is a fast, tall wave. The surprising result for this test case is that the optimal ensemble size is significantly smaller than for TC1, Fig. 2. For all three of the prognostic variables the smallest rms errors appears to be occurring when $S = 80$. The other surprising feature is the increase in the rms error as the ensemble size is increased. A possible explanation for this feature is given in the next section.

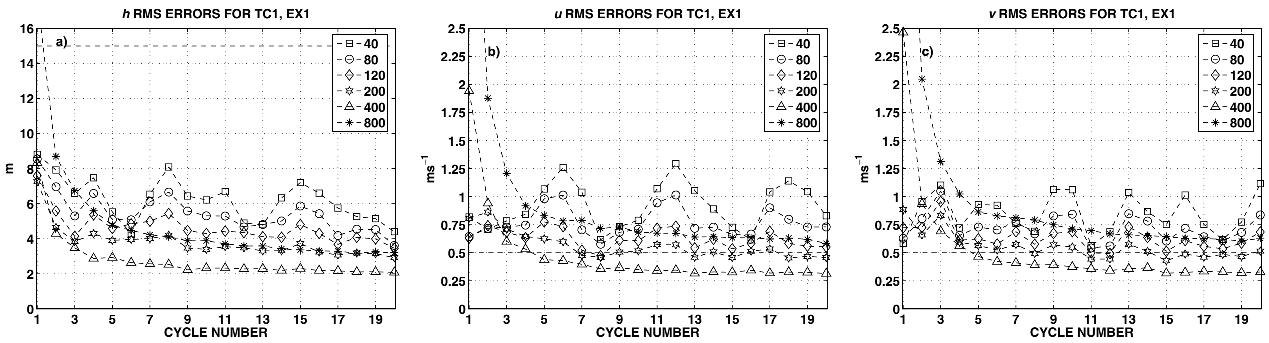

Fig. 1. Experiment 1: Plots of rms errors for (a) h field, (b) u field and (c) v field for TC1.

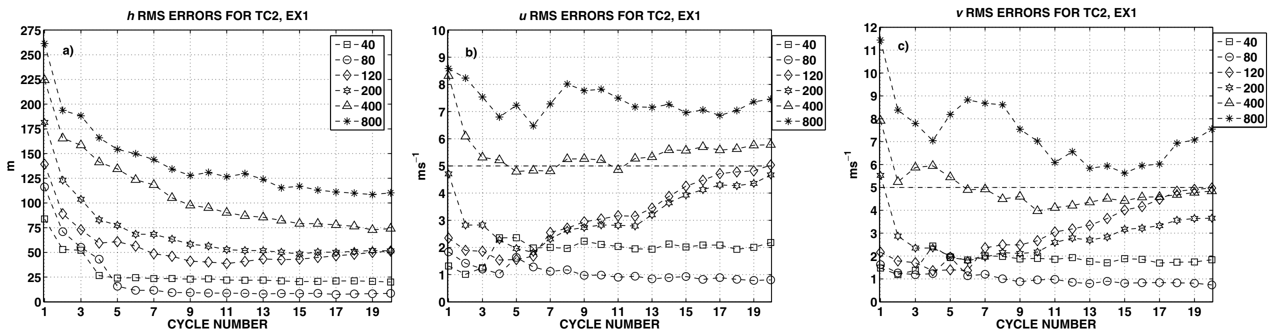

Fig. 2. Experiment 1: Plots of rms errors for (a) h field, (b) u field and (c) v field for TC2.

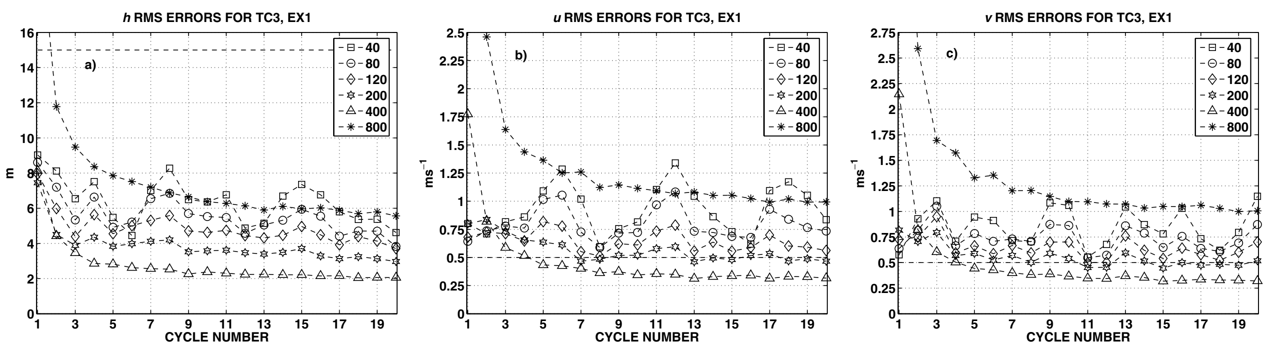

Fig. 3. Experiment 1: Plots of rms errors for (a) h field, (b) u field and (c) v field for TC3.

The χ^2 plot for this test case is presented in Fig. 4b, where, unlike for TC1, Fig. 4a, as the ensemble size increases there is no dramatic decrease in the statistic. Another important point to note here is that the optimal ensemble size, where χ^2 stays approximately 1, is $S = 80$ which is consistent with the results from the rms error analysis.

In the final test case, TC3, the associated wave is more of a non-divergent wave, similar to geostrophic balance. The surprising feature here is that a fairly large size ensemble is required to capture the dynamics associated with this wave. From the rms error plots, Fig. 3, the optimal ensemble size appears to be $S = 400$, the same as TC1. The other interesting feature is the drastic increase in the rms error when the ensemble size is increased to $S = 800$. This effect is present in all three error plots for TC3, Fig. 3. The same effect is also observed in the χ^2 plot for this test case, Fig. 4. The statistic is only approximately 1 for $S = 400$.

The results that have been presented here, for this first experiment, raises many questions which shall be addressed in the next section. However, we now present results from the second experiment, where the number of ensemble members is kept constant at the optimal number identified in the first experiment, but now the total number of observations is varied.

6.2. Experiment 2

The second set of experiments are concerned with the same model and initial conditions as for the first experiment, but now

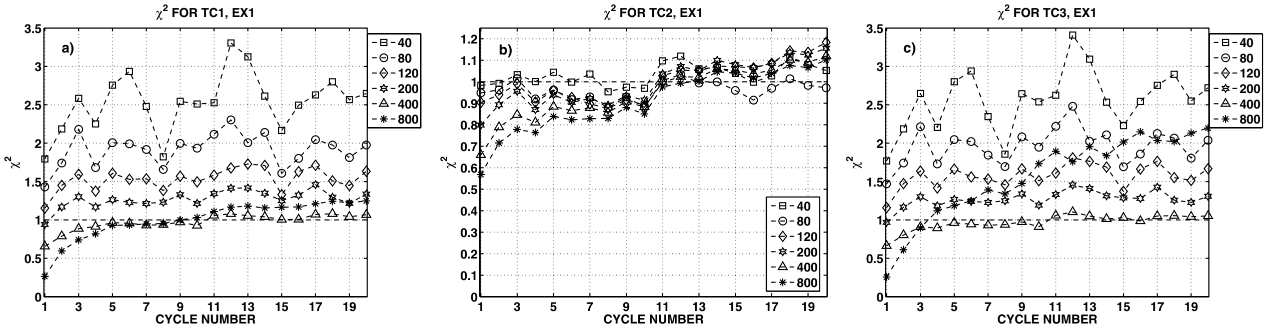

Fig. 4. Experiment 1: Plot of the χ^2 statistic for (a) TC1, (b) TC2 and (c) TC3.

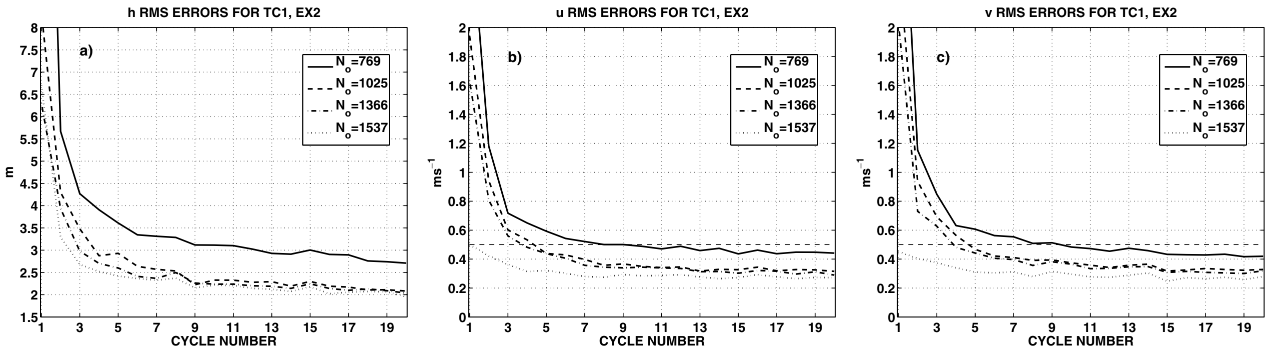

Fig. 5. Experiment 2: Plots of rms errors for (a) h field, (b) u field and (c) v field for TC1.

the impact on the filter when the number of observations of both the height and wind fields is allowed to vary, is investigated.

Starting with TC1, Fig. 5, it is clear that the rms error for the height field, given the different sets of observations, is substantially less than that of the observational error. When $N_o = 769$, it appears that there may not be quite enough information for 400 ensemble members, but the error is still reduced more than by increasing the ensemble size.

However, by increasing the number of observations to $N_o = 1366$, there appears to be a slight reduction, relative to the results from the first experiment. An interesting result comes from the wind fields when $N_o = 1537$. The rms error for the wind fields, Fig. 5a and b, starts below the observational error from the first cycle, and continues to out perform the other sets of observations, given the optimal choice for the ensemble size from the first experiment.

The χ^2 results for this test case, Fig. 8a, show that this statistic is approximately 1 for $N_o = 1025$ and 1366, but is not so for the other two sets of observations. This suggests that the observations are not being used efficiently, given the background uncertainties.

For TC2, Fig. 6, it is noticeable that the smallest rms error for all three prognostic variables is when $N_0 = 1025$. This suggests that the height observations are the more important observations for this test case. The rms error results from all four sets of observations are less than the observational errors.

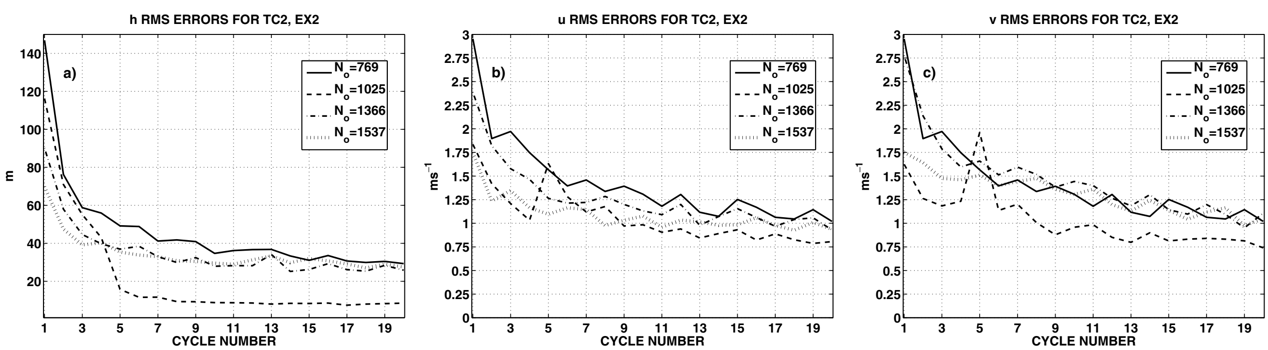

Fig. 6. Experiment 2: Plots of rms errors for (a) h field, (b) u field and (c) v field for TC2.

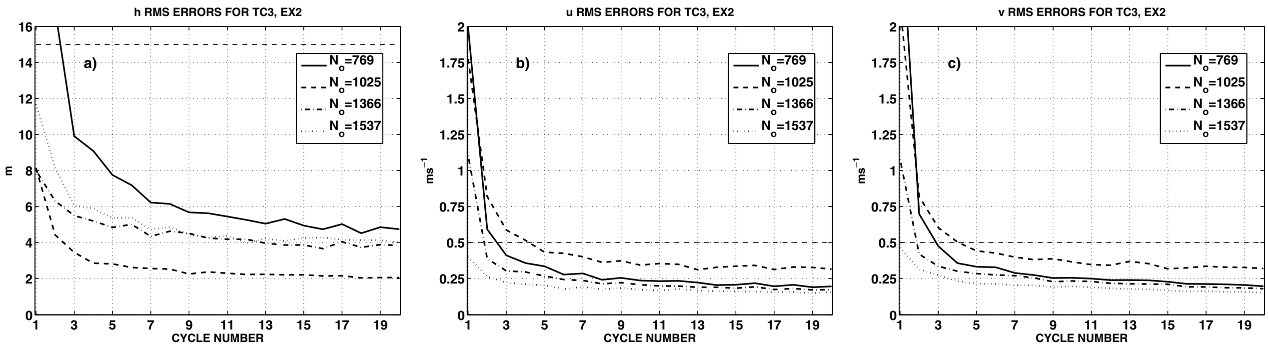

Fig. 7. Experiment 2: Plots of rms errors for (a) h field, (b) u field and (c) v field for TC3.

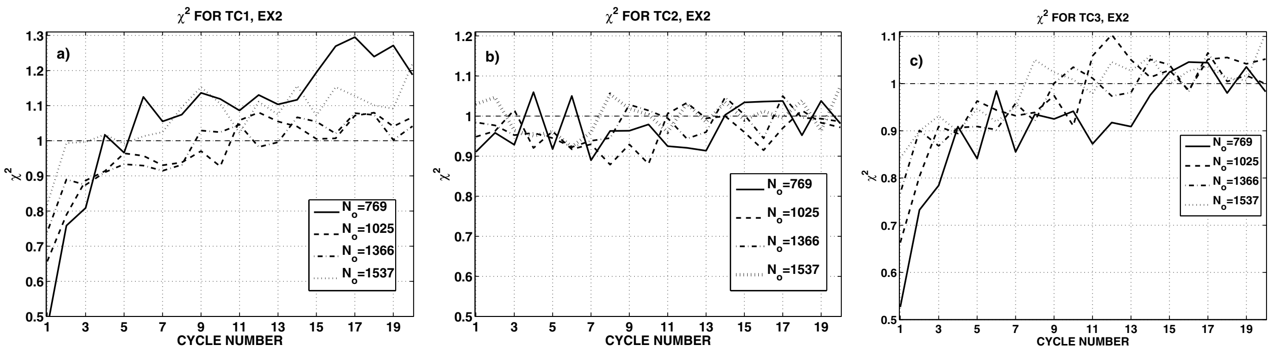

Fig. 8. Experiment 2: Plots of the χ^2 statistic for (a) TC1, (b) TC2 and (c) TC3.

The associated χ^2 results for this test case are presented in Fig. 8b). The value of the statistic fluctuates around 1 for all four sets of observations, but does not appear to be diverging. This then implies that the observations are being used efficiently, and are consistent with the background state.

In the final sets of results, Fig. 7, there appears to be a conflict as to what is the best set of observations to minimize the rms errors. From the height rms error plot, Fig. 7a, the best set of observations appears to be when there are approximately equal number of height and wind observations. However, for the wind field rms error plots, Fig. 7b and c, it appears that having twice the number of wind observations, $N_o = 1366$ and $N_o = 1537$, reduces the u and v rms errors noticeably more than when $N_o = 1025$.

Finally, considering the χ^2 results, Fig. 8c, we see that the statistic is suggesting that the filter is able to use most of the observations effectively. However, at cycle 20 we see that the χ^2 value is increasing for $N_o = 1537$, but is still approximately 1 as the flows is approximately linear.

6.3. *Summary*

In this section we have presented results from using the MLEF with CSU's shallow water equations model with three different Rossby–Haurwitz waves as initial conditions. The three waves each generate different types of flows, from geostrophic, TC3, to fast, shallow wave, TC1, to tall, fast dynamic wave, TC2. In Experiment 1 the optimal number of ensemble members, S, has been investigated with respect to the root mean square error for the three prognostic variables, geopotential height, h, and horizontal wind components, u and v. For TC1 the optimal value appears to be $S = 400$. This value also appears to be the true for TC3. However, the interesting feature is for TC2, where it appears the optimal value is $S \approx 80$.

Recalling the definition, and interpretation, of the Burger and Rossby numbers, along with their associations with the three test cases, we see that for flows that have a strong geostrophic component, then the MLEF requires more members in the ensemble to capture the flow correctly. However, for flows that are not dominated by a form of goestrophic balance, then the required number of ensemble members is much smaller. A possible explanation for this behaviour, given the underlying flows, is presented in the next section.

The second set of experiments that we have presented results for in this section are associated with taking the optimal ensemble size, as identified in Experiment 1, and changing the number of observations that are used. We saw for TC1 that doubling the number of wind observations reduced the rms error for all three fields. For TC2 we saw that, for all three fields, the set of observations we had used for Experiment 1, gave the best results. This suggests that this test case is more dependent on the height observations, rather than the wind observations.

For the final test case there appeared to be a slight conflict. When there was more height observations, then h rms error was smaller, however, there was a significant drop in the rms errors for the u and v components when there was more wind observations. This suggests that best set of observation is when there are equal number of wind and height observations, for this test case.

7. MLEF and hybrid Lyapunov-bred vectors

In this brief section some mathematical insights are presented into the structure of the MLEF, which was presented in Section 2, but here we highlight the implicit workings of the filter. We start with a brief summary of Lyapunov and bred vectors. The summary of the two types of vectors comes from Kalnay et al. (2002).

7.1. Lyapunov and bred vectors

Bred vectors were first presented as the finite amplitude, finite time extension of Lyapunov vectors in Toth and Kalnay (1993, 1997). A clearer mathematical description can be found in Kalnay et al. (2002), where a comparisons is made between bred vectors and Lyapunov vectors. The two different vectors are briefly summarized below.

7.1.1. Lyapunov vector generation. For Lyapunov vectors it is assumed that there is an evolving basic solution, $\mathbf{f}(\mathbf{x}, t)$, that satisfies a set of non-linear model equations that have been discretized in time and space. The associated numerical intergration scheme is represented by

$$\mathbf{f}(\mathbf{x}, t + \Delta t) = \mathcal{M}[\mathbf{f}(\mathbf{x}, t)].$$

If the initial conditions are perturbed, then the linear evolution of that perturbation is defined by

$$\delta\mathbf{f}(\mathbf{x}, t + \Delta t) = \mathbf{L}(\mathbf{f}, t, \Delta t)\,\delta\mathbf{f}(\mathbf{x}, t), \tag{25}$$

where

$$\mathbf{L} = \frac{\partial \mathcal{M}}{\partial \mathbf{x}}, \tag{26}$$

is the tangent linear model, TLM, or propagator.

Given the descriptions in (25) and (26), then the procedure to calculate the leading Lyapunov vector is to start with the arbitrary perturbation, $\delta\mathbf{f}(\mathbf{x}, t)$, and evolve this perturbation through the TLM, (25). After a sufficiently long time the perturbation converges to the leading Lyapunov vector. To find the other Lyapunov vectors other perturbations can be used, but must be orthogonalized at each time step to prevent convergence to the leading Lyapunov vector.

7.1.2. Bred vector generation. The procedure to generate bred vectors starts with an arbitrary perturbation, $\delta\mathbf{f}(\mathbf{x}, t)$, of some size, defined by an arbitrary norm. The perturbation is then added to the control solution. This perturbed initial condition is then integrated with the non-linear model to the next time step, and then subtracted from the trajectory of the control run. The evolved perturbation, $\widehat{\delta\mathbf{f}}$, is given by

$$\widehat{\delta\mathbf{f}}(\mathbf{x}, t + \Delta t) = \mathcal{M}[\mathbf{f}(\mathbf{x}, t) + \delta\mathbf{f}(\mathbf{x}, t + \Delta t)] - \mathcal{M}[\mathbf{f}(\mathbf{x}, t)]. \tag{27}$$

The next step is to measure the increase in the size of the evolved perturbation, and then re-scale to create the perturbation for the next non-linear run. In the process of calculating the bred vectors we do not orthogonalize the vectors if we have multiple perturbations. Therefore all of the bred vectors are related to the leading Lyapunov vector (Kalnay et al., 2002).

We now consider the structure of the MLEF to compare to the techniques described above.

7.2. Hybrid Lyapunov-bred vectors

The two main matrices for for the MLEF were given in Section 2. These were the square root analysis error covariance matrix, $\mathbf{P}_a^{\frac{1}{2}}$, (12) with (8), and the square root forecast error covariance matrix, $\mathbf{P}_f^{\frac{1}{2}}$, (5). Considering the square root forecast error covariance matrix first, we have

$$\begin{aligned}\mathbf{P}_f^{\frac{1}{2}} &= (\mathbf{b}_1 \quad \mathbf{b}_2 \quad \dots \quad \mathbf{b}_S),\\ \mathbf{b}_j &= \mathcal{M}_{i-1,i}\left(\mathbf{x}_i + \mathbf{p}_j\right) - \mathcal{M}_{i-1,i}(\mathbf{x}_i) \quad j = 1, 2, \dots, S\end{aligned}$$

where $\mathbf{p}_j$s are the columns of $\mathbf{P}_a^{\frac{1}{2}}$ from the previous cycle. Given the definition for bred vectors in (27), then it is clear that the columns of $\mathbf{P}_f^{\frac{1}{2}}$ are a series of non-linearly evolved perturbations. We must note here that the perturbations that are evolved are those from the square-root analysis error covariance matrix. We now consider how these perturbations are calculated to see how they fit into either the rescaling criterion of the bred vectors generation, or the orthogonalization criterion of separating Lyapunov vectors, or both.

We start be recalling the definition of the Hessian update, (12),

$$\mathbf{P}_a^{\frac{1}{2}} \equiv \mathbf{P}_f^{\frac{1}{2}}[\mathbf{I} + \mathbf{C}(\mathbf{x}_{opt})]^{-\frac{1}{2}}, \tag{28}$$

where

$$\mathbf{C} = \mathbf{P}_f^{\frac{T}{2}}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\mathbf{P}_f^{\frac{1}{2}}. \tag{29}$$

As we can see in (28), the inverse of the Hessian matrix of the cost function is required. As mentioned in Section 2, this inversion is accomplished through an orthogonal eigenvalue decomposition. Therefore, (28) can be rewritten as

$$\mathbf{P}_a^{\frac{1}{2}} = \mathbf{P}_f^{\frac{1}{2}}\mathbf{V}(\mathbf{I} + \Lambda)^{-\frac{1}{2}}\mathbf{V}^T, \tag{30}$$

where $\mathbf{V}$ is the matrix containing the orthogonal basis of the Hessian of the cost function, projected onto the ensemble subspace.

If we now consider the sequence of matrix multiplications in (30), we see that first we apply an orthogonal transformation to

$\mathbf{P}_f^{\frac{1}{2}}$. This transformation could be in the form of any orthogonal transformation, that is, rotation, reflection, etc., see Strang (1980) or Golub and Van Loan (1996) for more information about the different types of orthogonal transforms. To these new transformed states we rescale with respect to the inverse square root of the Hessian matrix in ensemble subspace. This is then telling us the directions in which the maximum energy/variability occurs and scale the increments accordingly. We finally transform this new state again by the inverse transform matrix, that is, $\mathbf{V}^T$.

Therefore, the reason why the filter is referred to as using a form of hybrid Lyapunov-bred vectors is because, the evolution of the perturbations are through the non-linear model, similar to bred vectors, but when we consider the Hessian update we both rescale the perturbations with respect to the directions with the most variance, that is, the inverse eigenvalues, and apply orthogonal transforms to the perturbations, which is similar to Lyapunov vector generation.

As it appears that the filter has a Lyapunov vector generation component implicitly built in, we start to notice that the performance of the filter is consistent with previous studies with Lyapunov and bred vectors. As mentioned in Kalnay et al. (2002), the more stable the dynamics, that is, high Burger number, low Rossby number, TC1, or low Burger number, low Rossby number (geostrophic balance), TC3, then there are more equally dominant Lyapunov vectors. The other case study, fast, tall wave with more dynamics involved, showed us that we require fewer vectors/ensemble members, which is again consistent with the behaviour of Lyapunov vectors, as seen in previous studies (Lorenz, 1996).

Recently a technique has been developed for the MLEF, and effectively the Ensemble Transform Kalman Filter (ETKF), that allows for the inflation of the size of the ensemble through using the perturbations at the previous analysis cycle, augmented to the current ensemble. This then introduces temporal information into the ensemble, as well as information about the dominant directions at the previous analysis time (Uzunoglu et al., 2007). Also in Uzunoglu et al. (2007), a technique is developed to reduce the size of the ensemble when there are ensemble members who's contribution to the total variance is nearly zero. This is consistent with ensemble members pointing in the same Lyapunov direction, therefore, not adding much to the analysis.

As we see with the larger ensemble sizes, we have the possibility that some of the members may be tending to the same Lyapunov vectors, and hence causing a form of degeneracy in the $\mathbf{C}$ matrix. This would imply that we could have an ill-conditioned, but still be symmetric positive definite, $\mathbf{C}$ matrix. If an eigenvalue was to be zero, then there would be two ensemble members pointing in the same direction. This could lead to problems with the inversion of the $\mathbf{C}$ matrix. The new technique in Uzunoglu et al. (2007) enables us to remove these members, thus resizing and reconditioning the $\mathbf{C}$ matrix to remove degeneracy.

8. Conclusions and further work

In this paper we have performed experiments to investigate, and explain more about, the characteristics of the MLEF (Zupanski, 2005). This filter is different from other ensemble filter in that it never calculates the ensemble, sample, mean.

As we never average over the ensemble members, we allow for a more dynamically based statistic, in the form of the mode, to be used to perform the analysis. We have seen that the three different types of flows presented here, TC1; shallow, fast wave, TC2; tall, fast wave and TC3, tall, slow wave (geostrophic balance), require different optimal number of ensemble members to reduce the root mean square (rms) error to a reasonable level. We have seen that for TC1 and TC3 the optimal number appears to be $S \approx 400$, while for the fast wave of TC2, we see that optimal number is $S \approx 80$.

In the last section we provided an explanation for this impact, as it is assumed, in sample theory, that the more ensemble members that you have, the better the approximations to the true moments of the distributions are. We have shown that this is not the case for this filter. The reason for this is the structure of the filter not being based upon sample theory, but rather on a form of hybrid Lyapunov-bred vectors. Therefore, we have a structure to the filter that is dependent on Lyapunov exponent directions.

We also investigated the impact of changing the number of observations with the identified optimal number of ensemble members from the first experiment. We saw that impact was minimal for some of the test cases, but having more height observations did have an impact when the height field drove the flow of the wave, see Marek (2002) for detailed analysis of the dominant features of the three waves used.

Although it may appear that we are using a sample approximation to the covariance matrices, we are actually using an singular value decomposition like approximation to the error covariance matrices. This is more consistent with a principal component analysis expansion of a covariance matrix (Sharma, 1996). We never average over the ensemble members to calculate the error covariance matrices, which is done in the ETKF (Bishop et al., 2001),

As a consequence of the hybrid Lyupanov/bred vector structure, the optimal size of the ensemble appears to be reliant on the Lyapunov vector structure in the dynamics that the filter is trying to predict. This an important result as this would not be the case if we were taking a statistical sample.

In summary: The MLEF has shown to have a quite different behaviour to that of a sample based ensemble filter. We have shown that using a very large size ensemble does not always improve the filter's performance, given the finite set of observations that we have used. We have identified a Lyapunov/bred vector structure to the filter that shows why the number of ensemble members needed to be optimal, with respect to the rms error, does not necessarily have to be too big.

The plan for the future are to test this hypothesis about the Lyapunov-bred vector basis of the filter with more advanced models, like the Weather, Research and Forecasting model, WRF, with the situation described in Kalnay et al. (2002) to see if we have the same structures as mention in that paper.

9. Acknowledgments

We are grateful for many useful conversations about the structure of the MLEF, and how to interpret the results that we have shown, with Drs. Andrew Jones and Dusanka Zupanski. We are also grateful to the Super Computing Department at the National Center for Atmospheric Research, which is supported by the National Science Foundation, NSF, for the use of BLUESKY with the numerical experiments presented here. We are grateful for very useful comments from the two anonymous reviewers of this manuscript. This work was part supported by the NSF grant for the collaboration in the mathematical geoscience grant # 0327651 and by the DoD Center for Geoscience/Atmospheric Research at Colorado State University under Cooperative Agreement (W911NF-060200015) with the Army Research Laboratory.

References

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.

Daley, R. 1996. Atmospheric Data Analysis. In: *Cambridge Atmospheric and Space Science Series*, (eds J. Houghton, M. J. Rycroft and A. J. Dessler), Cambridge University Press, Cambridge, UK., 457 pp.

Evensen, G. 2003. The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dyn.* **53**, 343–367.

Fletcher, S. J. 2004. *Higher Order Balance Conditions using Hamiltonian Dynamics for Numerical Weather Prediction*. PhD Thesis, University of Reading, U.K.

Fletcher, S. J. and Zupanski, M. 2006a. A data assimilation method for log-normally distributed observational errors. *Quart. J. Roy. Meteor. Soc.* **132**, 2505–2519.

Fletcher, S. J. and Zupanski, M. 2006b. A hybrid normal and lognormal distribution for data assimilation. *Atmos. Sci. Lett.* **7**, 43–46.

Gaspari, G. and Cohn, S. E. 1999. Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.* **125**, 723–757.

Golub, G. H. and Van Loan, C. F. 1996. *Matrix Computations* 3rd Edition. The Johns Hopkins University Press, London, UK., 694 pp.

Hamill, T. M. and Synder, C. 2000. A hybrid ensemble Kalman filter-3D variational analysis scheme. *Mon. Wea. Rev.* **128**, 2905–2919.

Heikes, R. and Randall, D. A. 1995a. Numerical Integration Of The Shallow Water Equations On A Twisted Icosahedral Grid. Part I: basic design and results of tests. *Mon. Wea. Rev.* **123**, 1862–1880.

Heikes, R. and Randall, D. A. 1995b. Numerical integration of the shallow water equations on a twisted icosahedral grid. Part II: a detailed description of the grid and an analysis of numerical accuracy. *Mon. Wea. Rev.* **123**, 1881–1887.

Holton, J. R. 1992. *An Introduction to Dynamic Meteorology*, Academic Press, San Diego, USA., 511 pp.

Hoskins, B. J. 1973. Stability of the Rossby–Haurwitz wave. *Quart. J. Roy. Meteor. Soc.* **99**, 723–745.

Houtekamer, P. L. and Mitchell, H. L. 2001. A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.* **129**, 123–137.

Jazwinski, A. H. 1970. *Stochastic Processes and Filtering Theory*, Academic Press, New York, USA., 376 pp.

Kalnay, E. 2003. *Atmospheric Modeling, Data Assimilation and Predictability*, Cambridge University Press, Cambridge, UK., 341 pp.

Kalnay, E., Corazza, M. and Cai, M. 2002. Are bred vectors the same as Lyapunov vectors?. In: *Proceedings of the Atmospheric Modeling, Data Assimilation and Predictability*, American Meteorological Society, Boston, MA, USA, 173–177

Lorenc, A. C. 1986. Analysis methods for numerical weather prediction. *Q. J. Roy. Meteor. Soc* **112**, 1177–1194.

Lorenc, A. C., Ballard, S. P., Bell, R. S., Ingleby, N. B., Andrews, P. L. F. and co-authors. 2000. The Met. Office global three-dimensional variational data assimilation scheme. *Quart. J. Roy. Meteor. Soc.* **126**, 2991–3012.

Lorenz, E. N. 1996. Predictability–a problem partly solved. *ECMWF Proc. Seminar on Predictability, Reading, UK.* 1–18.

Ménard, R., Cohn, S. E., Chang, L.-P., and Lyster, P. M. 2000. Assimilation of stratospheric chemical tracer observations using a Kalman filter. Part I: formulation. *Mon. Wea. Rev.* **128**, 2654–2671.

Parrish, D. F. and Derber, J. C. 1992. The National Meteorological Center's spectral statistical-interpolation analysis system. *Mon. Wea. Rev.* **120**, 1747–1763.

Pikovsky, A. and Politi, A. 1998. Dynamic localization of Lyapunov vectors in spacetime chaos. *Nonlinearity* **11**, 1049–1062.

Rabier, F., Jarvinen, H., Klinker, E., Mahfouf, J.-F. and Simmons, A. 2000. The ECMWF implementation of four dimensional variational assimilation. Part I: experimental results with simplified physics. *Quart. J. Roy. Meteor. Soc.* **126A**, 1143–1170.

Rawlins, F., Ballard, S. P., Bovis, K. J., Clayton, A. M., Li, D. and co-authors. 2007. The Met Office global four-dimensional variational data assimilation scheme. *Quart. J. Roy. Meteor. Soc* **133**, 347–362.

Reichle, R. H., McLaughlin, D. B. and Entekhabi, D. 2002. Hydrologic data assimilation with Kalman filter. *Mon. Wea. Rev.* **130**, 103–114.

Ringler, T. D. and Randall, D. A. 2002. A potential enstrophy and energy conserving numerical scheme for solutions of the shallow water equations on a geodesic grid. *Mon. Wea. Rev.* **130**, 1397–1410.

Sharma, S. 1996. *Applied Multivariate Techniques* John Wileys and Sons, Inc., New York, USA, 493 pp.

Strang, G. 1980. *Linear Algebra and its Applications* 2nd Edition. Academic Press, New York, USA, 414 pp.

Tippett, M., Anderson, J. L., Bishop, C. H., Hamill, T. M. and Whitaker, J. S. 2003. Ensemble square root filters. *Mon. Wea. Rev.* **131**, 1485–1490.

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

Toth, Z. and Kalnay, E. 1997. Ensemble forecasting at NCEP: the breeding method. *Mon. Wea. Rev.* **125**, 3297–3318.

Uzunoglu, B., Fletcher, S. J., Zupanski, M. and Navon, I. M. 2007. Adaptive ensemble reduction and inflation. *Quart. J. Roy. Meteor. Soc.* **133**, 1281–1294.

Williamson, D. L., Drake, J. B., Hack, J. J., Jakob, R. and Swartrauber, P. N. 1992. A standard test set for numerical approximations to the shallow water equations in spherical geometry. *J. Comp. Phys.* **102**, 211–224.

Wlasak, M. A. 2002. *The Examination of Balanced and Unbalanced Flow using Potential Vorticity in Atmospheric Modelling*. PhD Thesis, University of Reading, U.K.

Wlasak, M. A., Nichols, N. K. and Roulstone, I. 2006. Use of potential vorticity for incremental data assimilation. *Quart. J. Roy. Meteor. Soc.* **132**, 2867–2886.

Zupanski, D. and Zupanski, M. 2006. Model error estimation employing ensemble data assimilation approach. *Mon. Wea. Rev.* **134**, 1337–1354.

Zupanski, M., 2005. Maximum Likelihood Ensemble Filter. Part I: theoretical Aspects. *Mon. Wea. Rev.* **133**, 1710–1726.

Zupanski, M., Fletcher, S. J., Navon, I. M., Uzunoglu, B., Heikes, R. P. and co-authors. 2006. A method for initiation of ensemble data assimilation. *Tellus* **58A**, 159–170.

Zupanski, M., Zupanski, D., Vukicevic, T., Eis, K. and Vonder Haar, T. 2005. CIRA/CSU Four-dimensional variational data assimilation scheme. *Mon. Wea. Rev.* **133**, 829–843.