Ensemble Data Assimilation for the shallow water equation

M.Jardak,[1] I.M.Navon,[1] and M.Zupanski[2]

————

M.Jardak, Department of Scientific Computing, Florida State University,

Tallahassee, FL 32306-4120, USA

(mjardak@scs.fsu.edu)

I.M. Navon, Department of Scientific Computing, Florida State University,

Tallahassee, FL 32306-4120, USA

(navon@scs.fsu.edu)

M.Zupanski, Cooperative Institute for Research in the Atmosphere,

Colorado State University, 1375 Campus Deliver,

Fort Collins, CO 80523-1375, USA

(ZupanskiM@cira.colostate.edu)

[1]Department of Scientific Computing,

Florida State University, Tallahassee, FL

32306-4120, USA

[2]CIRA,Colorado State University, 1375

Campus Deliver, Fort Collins, CO

80523-1375, USA

Abstract.   Many problems in the geosciences require estimation of the state of a system that changes over time using a sequence of noisy measurements made on the system. Data assimilation is the process of fusing observational data and model predictions to obtain an optimal representation of the state of the atmosphere. Among most challenging problems remaining for EnsDA applications in geosciences is the problem of high nonlinear and nondifferentiable processes and observations. Examples of such processes are cloud, aerosol and precipitation processes, as well as remote sensing (e.g., stellite and radar) observations. Unfortunately, a typical EnsDA analysis equation is linear, being based on the Kalman filter equations. This fundamentally prevents EnsDA from extracting maximum information from such observations, and ultimately limits its applicability to outstanding geosciences problems such as hurricane prediction and high-resolution climate simulation. In order to overcome the limitation of existing EnsDA methodologies in application to highly nonlinear and nondifferentiable geoscience problems,a fundamentally new strategy is required. A new comparison of three frequently used sequential data assimilation methods illuminating their strengths and weaknesses in the presence of linear and nonlinear observation operators is presented. The ensemble Kalman filter (EnKF), the particle filter (PF) and the Maximum Likelihood Ensemble Filter (MLEF) methods were implemented and the spectral shallow water equations model in spherical geometry model was employed using the Rossby-Haurwitz Wave no 4 as initial conditions.

————

Conclusions are drawn as to the performance of the above filters for the above test case with both linear and nonlinear observation operators. Numerical tests conducted reveal that all the three methods perform satisfactory for the linear observation operator for 12 days model integration, whereas the EnKF, even with the Evensen fixture [Evensen 03] for the nonlinear observation operator failed. The particle filter and the hybrid filter (MLEF) both performed satisfactory with slight edge to the PF.

1. Introduction and motivation

Sequential data assimilation fuses observations of the current (and possibly, past) state of a system with results from a mathematical model (the forecast) to produce an analysis, providing "the best" estimate of the current state of the system. Central to the concept of sequential estimation data assimilation is the propagation of flow dependent probability density function (pdf) given an estimate of the initial pdf.

In sequential estimation, the analysis and forecasts can be viewed as probability distributions. The analysis step is an application of the Bayes theorem. Advancing the probability distribution in time, for the general case is done by the Chapman-Kolmogorov equation, but since it is unrealistically expensive, various approximations operating on representations of the probability distributions are used instead. If the probability distributions are normal, they can be represented by their mean and covariance, which gives rise to the Kalman filter (KF). However, due to the high computational and storage overheads required, various approximations based on Monte-Carlo ensemble calculations have been proposed by [Evensen 94]. Research on ensemble Kalman filtering (EnKF) started with the work of [Evensen 94], [Evensen and Van Leeuwen 96], [Burgers et al. 98] and [Houtekamer and Mitchell 98]. The method is essentially a Monte-Carlo approximation of the Kalman filter which avoids evolving the covariance matrix of the pdf of the state vector. A second type of EnKF filter consists of the class of square root filters of [Anderson and Anderson 03] see also [Bishop et al. 01]. The review of [Tippett et al. 03] consists of a single analysis based on the ensemble mean, and where the analysis perturbations are obtained from the square root of the filter analysis error covariance. See also the paper

of [Nerger 05] where the (EnKF) filter, the singular evolutive extended Kalman (SEEK) filter, and the less common singular evolutive interpolated Kalman (SEIK) filter were reviewed and compared.

Particle filters, also known as sequential Monte-Carlo (SCM) methods as well as Bayesian filters, are sophisticated model estimation techniques based on simulation. A precursor to particle filter method tested [Jardak et al. 08], [Xiong et al. 06] is due to the pioneering contribution of [Gordon et al. 93]. Initially the SCM focused on applications to tracking and vision, these techniques are now very widespread and have been used in a variety of applications linked to Bayesian dynamical models see [Doucet et al. 01]. These methods utilize a large number N of random samples named particles to represent the posterior probability distributions. The particles are propagated over time using a combination of sequential importance sampling and resampling steps. Resampling for PF is used to avoid the problem of degeneracy of this algorithm that is, avoiding the situation that all but one of the importance weights are close to zero. The performance of the PF algorithm can be crucially affected by judicious choice of a resampling method. See [Arulampalam et al. 02] for a listing of the most used resampling algorithms. A major drawback of particle filters is that they suffer from sample degeneracy after a few filtering steps. The PF suffers from "the curse of dimensionality" requiring computations that increase exponentially with dimension as pointed out by Silverman [Silverman 86]. This argument was enhanced and amplified by the recent work of [Bengtsson et al. 08] and [Bickel et al. 08] and finally explicitly quantified by [Synder et al. 08]. They indicated that unless the ensemble size is greater than $\exp(\tau^2/2)$, where τ^2 is the variance of the observation log-likelihood, the PF update suffers from a "collapse" in which with high

probability only few members are assigned a posterior weight close to one while all other members have vanishing small weights. This issue become more acute as we move to higher spatial dimensions.

The Maximum Likelihood Ensemble Filter of [Zupanski 05]; [Zupanski and Zupanski 06], does not calculate a sample mean from the ensemble members. The analysis state that the MLEF seeks is the mode, which reduces to the mean in the case of linear dynamics and Gaussian statistics. The mode is found through minimizing a cost function, similar to that in the three-dimensional variational assimilation method, 3D-Var, of [Lorenc 86], but projected into ensemble space, rather than modal space. Hence MLEF can be viewed as a hybrid filter. The method comprises of three steps, a forecast step that is concerned with the evolution of the forecast error covariances, an analysis step based on solving a non-linear cost function and on updating step.

In this paper data assimilation experiments are performed using the ensemble Kalman filter (EnKF), the particle filter (PF) and the Maximum Likelihood Ensemble Filter (MLEF). These methods were tested on the spectral shallow water equations model using the Rossby-Haurwitz test case for both linear and nonlinear observation operators. To the best of our knowledge this contribution in the context presented here is novel.

The spectral shallow water equation model in spherical geometry of Williamson[Williamson 92] and [Jakob et al.95] is used to generate a true solution and to generate forecasts. To improve EnKF analysis errors and avoid ensemble errors that generate spurious corrections, a covariance localization investigated by [Houtekamer and Mitchell 1998,2001]. is incorporated. To improve both the analysis and the forecast results, the forecast ensemble solutions are inflated from the mean as suggested in [Anderson and Anderson 99] and re-

ported in [Hamill et al. 2001], and [Uzunoglu et al. 07]. Finally a parallel implementation similar to the one of [Keppenne 2000] is used to render the algorithm computationally feasible.

Since the resampling is a crucial step for (PF) method, the systematic, multinomial and the merging resampling methods [Arulampalam et al. 02],[Doucet et al. 01] and [Nakano and al. 07] were tested. In order to render algorithm computationally feasible a parallel MPI implementation is used.

The paper is structured as follows. Section 2 presents the spectral shallow-water equations in spherical coordinates and the numerical methods used for its resolution. In section 3 we present each of the data assimilation methods. Section 4 we present the numerical results and discuss them. Finally section 5 is reserved for summary and conclusions.

2. Shallow-Water equations in spherical geometry

The shallow water equations are a set of hyperbolic partial differential equations that describe the flow below a pressure surface in a fluid.

The equations are derived from depth-integrating the Navier-Stokes equations, in the case where the horizontal length scale is much greater than the vertical length scale. The shallow-water equations in spherical geometry are given by

$$
\begin{cases}
\dfrac{\partial u}{\partial t} + \dfrac{u}{a\cos\theta}\dfrac{\partial u}{\partial \lambda} + \dfrac{v}{a}\dfrac{\partial u}{\partial \theta} - \dfrac{\tan\theta}{a}vu - fv = -\dfrac{g}{a\cos\theta}\dfrac{\partial h}{\partial \lambda} \\[2ex]
\dfrac{\partial v}{\partial t} + \dfrac{u}{a\cos\theta}\dfrac{\partial v}{\partial \lambda} + \dfrac{v}{a}\dfrac{\partial v}{\partial \theta} + \dfrac{\tan\theta}{a}u^2 + fu = -\dfrac{g}{a}\dfrac{\partial h}{\partial \theta} \\[2ex]
\dfrac{\partial h}{\partial t} + \dfrac{u}{a\cos\theta}\dfrac{\partial h}{\partial \lambda} + \dfrac{v}{a}\dfrac{\partial h}{\partial \theta} + \dfrac{h}{a\cos\theta}\left[\dfrac{\partial u}{\partial \lambda} + \dfrac{\partial(\cos\theta)}{\partial \theta}\right] = 0
\end{cases}
\tag{1}
$$

where $V = u\vec{i} + v\vec{j}$ is the horizontal velocity vector (with respect to the surface of the sphere), gh is the free surface geopotential, h is the free surface height, g is the gravity acceleration. $f = 2\Omega \sin \theta$ is the Coriolis parameter, Ω is the angular velocity of the earth. θ denotes the angle of latitude, $\mu = \sin \theta$ is the longitude. λ the longitude, and a is the radius of the earth.

One of the major advances in meteorology was the use by Rossby (1939) of the barotropic vorticity equation with the β -plane approximation to the sphericity of the Earth, and the deduction of solutions reminiscent of some large scale waves in the atmosphere. These solutions have become known as Rossby waves. Haurwitz (1940) then produced the equivalent solution for the sphere, now known as Rossby-Haurwitz waves (R-H waves). Rossby-Haurwitz waves are steadily propagating solutions of the fully nondivergent barotropic vorticity equation on a sphere first put forward by Rossby 1939 and Haurwitz (1940).

While the shallow water do not have corresponding analytic solutions they are expected to evolve in a similar way as the aboveR-H equations which explains why they have been widely used to test shallow water numerical models since the seminal paper of Phillips (1959)

Following the research work of Hoskins (1973) Rossby -Haurwitz waves with zonal wave numbers less or equal to 5 are believed to be stable. This makes the R-H zonal wave no 4 a suitable candidate for assessing accuracy of numerical schemes as was evident from its being chosen as a test case by Williamson et al (1992) and by a multitude of other authors.

It has been numerically shown that the R-H wave number 4 breaks down into more turbulent behaviour at late times as recently discovered by Thuburn and Li [20].

The Rossby-Haurwitz waves are analytic solutions of the nonlinear barotropic vorticity equation on the sphere [Haurwitz 40] and R. K. Smith and D. G. Dritschel [Smith 06]. Although they are not analytic solutions of the shallow water equations they have been used so frequently for meteorological tests.

The initial velocity field for the Rossby-Haurwitz wave is defined as

$$\begin{cases} u = a\omega \cos\phi + aK \cos^{r-1}\phi(r\sin^2\phi - \cos^2\phi)\cos(r\lambda) \\ v = -aKr \cos^{r-1}\phi \sin\phi \sin(r\lambda) \end{cases} \tag{2}$$

The initial height field is defined as,

$$h = h_0 + \frac{a^2}{g}[A(\phi) + B(\phi)\cos(r\lambda) + C(\phi)\cos(2r\lambda)] \tag{3}$$

where the variables $A(\phi), B(\phi), C(\phi)$ are given by

$$\begin{cases} A(\phi) = \dfrac{\omega}{2}(2\Omega + \omega)\cos^2\phi + \dfrac{1}{4}k^2\cos^{2r}\phi[(r+1)\cos^2\phi + (2r^2 - 2r - 2) - 2r^2\cos^2\phi] \\[2mm] B(\phi) = \dfrac{2(\Omega + \omega)k}{(r+1)(r+2)}\cos^r\phi[(r^2 + 2r + 2) - (r+1)^2\cos^2\phi] \\[2mm] C(\phi) = \dfrac{1}{4}k^2\cos^{2r}\phi[(r+1)\cos^2\phi - (r+2)] \end{cases} \tag{4}$$

In here, r represents the wave number,h_0 is the height at the poles. The strength of the underlying zonal wind from west to east is given by ω and k controls the amplitude of the wave.

3. Sequential Bayesian Filter- theoretical setting

This applies to all 3 sequential data assimilation methods discussed herein. The sequential Bayesian filter employing a large number N of random samples advanced in time by a stochastic evolution equation, to approximate the probability densities. In order to analyze and make inference about the dynamic system at least a model equation along with an observation operator are required. First, a model describing the evolution of the state

with time, and an observation operator for noisy observations of the state. Generically, stochastic filtering problem is a dynamic system that assumes the form

$$\dot{\mathbf{x}}_t = f(t, \mathbf{x}_t, u_t, \mathbf{v}_t) \tag{5}$$

$$\mathbf{z}_t = h(t, \mathbf{x}_t, u_t, \mathbf{n}_t) \tag{6}$$

The equation (5) is the state equation or the system model, (6) is the observation operator equation, $\mathbf{x}_t$ is the state vector, $\mathbf{z}_t$ the observation vector and $\mathbf{u}_t$ is the system input vector serving as the driving force. $\mathbf{v}_t$ and $\mathbf{n}_t$ are the state and observation noises, respectively. In practical application, however, we are more concerned about the discrete-time filtering, and we consider the evolution of the state sequence $\{\mathbf{x}_k, k \in \mathbb{N}\}$, given by

$$\mathbf{x}_k = f_k(\mathbf{x}_{k-1}, \mathbf{v}_{k-1}), \tag{7}$$

where the deterministic mapping $f_k : \mathbb{R}^{n_x} \times \mathbb{R}^{n_d} \longrightarrow \mathbb{R}^{n_x}$ is a possibly non-linear function of the state $\mathbf{x}_{k-1}$, $\{\mathbf{v}_{k-1}, k \in \mathbb{N}\}$ is an independent identically distributed (i.i.d) process noise sequence, n_x, n_d are dimensions of the state and process noise vectors, respectively, and $\mathbb{N}$ is the set of the natural numbers. The objective is to recursively estimate $\mathbf{x}_k$ from observations

$$\mathbf{z}_k = h_k(\mathbf{x}_k, \mathbf{n}_k), \tag{8}$$

where $h_k : \mathbb{R}^{n_{\mathbf{x}}} \times \mathbb{R}^{n_n} \longrightarrow \mathbb{R}^{n_z}$ is a possibly non-linear function, $\{\mathbf{n}_k, k \in \mathbb{N}\}$ is an i.i.d. observation noise sequence, and n_x, n_n are dimensions of the state and observation noise vectors, respectively.

We denote by $\mathbf{z}_{1:k}$ the set of all available observations $\mathbf{z}_i$ up to time $t = k$, $\mathbf{z}_{1:k} = \{\mathbf{z}_i | i = 1, \cdots, k\}$. From a Bayesian point of view, the problem is to recursively calculate some degree of belief in the state $\mathbf{x}_k$ at time $t = k$, taking different values, given the

data $\mathbf{z}_{1:k}$ up to the time $t = k$. Then the Bayesian solution would be to calculate the pdf $p(\mathbf{x}_k|\mathbf{z}_{1:k})$. This density will encapsulate all the information about the state vector $\mathbf{x}_k$ that is contained in the observations $\mathbf{z}_{1:k}$ and the prior distribution for $\mathbf{x}_k$.

Suppose that the required PDF $p(\mathbf{x}|\mathbf{z}_{1:k-1})$ at time $k - 1$ is available. The prediction stage uses the state equation (7) to obtain the prior PDF of the state at time k via the Chapman-Kolmogorov equation

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})d\mathbf{x}_{k-1}. \tag{9}$$

The probabilistic model of the state evolution, $p(\mathbf{x}_k|\mathbf{x}_{k-1})$, is defined by the state equation (7) and the known statistics of $\mathbf{v}_{k-1}$.

At time $t = k$, a measurement $\mathbf{z}_k$ becomes available, and it may be used to update the prior via the Bayes rule

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) = \frac{p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{1:k-1})}{p(\mathbf{z}_k|\mathbf{z}_{1:k-1})}, \tag{10}$$

where the normalizing constant

$$p(\mathbf{z}_k|\mathbf{z}_{1:k-1}) = \int p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{1:k-1})d\mathbf{x}_k. \tag{11}$$

depends on the likelihood function $p(\mathbf{z}_k|\mathbf{x}_k)$, defined by the measurement equation (8) and the known statistics of $\mathbf{n}_k$.

The relations (9) and (10) form the basis for the optimal Bayesian solution. This recursive propagation of the posterior density is only a conceptual solution. One cannot generally obtain an analytical solution. Solutions exist only in a very restrictive set of cases like that of the Kalman filters for instance. (namely, if f_k and h_k are linear and both v_k and n_k are Gaussian). Particle filters provide a direct approximation to the Bayes rule outlined above.

3.1. Particle Filters

Particle filters see [Doucet 2000, Doucet 2001, Arulampalam 02 ,Berliner 07 A] and recently the review paper of [Van Leeuwen 09] approximate the posterior densities by population of states. These states are called "particles". Each of the particles has an assigned weight, and the posterior distribution can then be approximated by a discrete distribution which has support on each of the particles. The probability assigned to each particle is proportional to its weight. The different (PF) algorithms differ in the way that the population of particles evolves and assimilates the incoming observations. A major drawback of particle filters is that they suffer from sample degeneracy after a few filtering steps.

In this 2-D plus time problem the common remedy is to resample the prior PDF whenever the weights focus on few members of the ensemble assuming that our experiments do not encounter the exponential collapse. Here we use several strategies such as Systematic Resampling (SR), Residual Resampling (RR) and or the Bayesian bootstrap filter of Gordon et al. [Gordon 93] see also [Berliner and Wikel 07 A and 07B]. Multinomial Resampling (MR) The SIR algorithm generates a population of equally weighted particles to approximate the posterior at some time k. This population of particles is assumed to be an approximate sample from the true posterior at that time instant.

The PF algorithm proceeds as follows:

• **Initialization:** The filter is initialized by drawing a sample of size N from the prior at the initial time. The algorithm is then started with the filtering step.

• **Preliminaries:** Assume that $\{\mathbf{x}_{k-1}^i\}_{i=1,\cdots,N}$ is a population of N particles, approximately distributed as in an independent sample from $p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})$

• **Prediction** Sample N values, $\{q_k^1, \cdots, q_k^N\}$, from the distribution of $\mathbf{v}_k$. Use these to generate a new population of particles, $\{\mathbf{x}_{k|k-1}^1, \mathbf{x}_{k|k-1}^2, \cdots, \mathbf{x}_{k|k-1}^N\}$ via the equation

$$\mathbf{x}_{k|k-1}^i = f_k(\mathbf{x}_{k-1}^i, \mathbf{v}_k^i) \tag{12}$$

• **Filtering:** Assign each $\mathbf{x}_{k|k-1}^i$, a weight q_k^i. This weight is calculated by

$$q_k^i = \frac{p(\mathbf{z}_k|\mathbf{x}_{k|k-1}^i)}{\sum_{j=1}^{N} p(\mathbf{z}_k|\mathbf{x}_{k|k-1}^j)} \tag{13}$$

This defines a discrete distribution which, for $i \in \{1, 2, \cdots, N\}$, assigns probability mass q_k^i to element $\mathbf{x}_{k|k-1}^i$

• **Resampling:** Resample independently N times, with replacement, from the distribution obtained in the filtering stage. The resulting particles, $\{\mathbf{x}_k^i\}_{i=1,\cdots,N}$, form an approximate sample from $p(\mathbf{x}_k|\mathbf{z}_{1:k})$.

The method outlined above can be justified as follows. If the particles at time $t = k-1$ were an i.i.d sample from the posterior at time $t = k-1$, then the predictive stage just produces an i.i.d. sample from the prior at time $t = k$. The filtering stage can be viewed as an importance sampling approach to generate an empirical distribution which approximates the posterior.

The proposal density is just the prior $p(\mathbf{x}_k|\mathbf{z}_{1:k-1})$, and as a result of Bayes formula, we obtain

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}, \mathbf{z}_k) \propto p(\mathbf{x}_k|\mathbf{z}_{1:k-1})p(\mathbf{z}_k|\mathbf{x}_k), \tag{14}$$

the weights are proportional to the likelihood $p(\mathbf{z}_k|\mathbf{x}_k)$. As N tends to infinity, the discrete distribution which has probability mass q_i at point $\mathbf{x}_{k|k-1}^i$, converges weakly to the true posterior. The resampling step is a crucial and computationally expensive part in a particle filter. It is used to generate equally weighted particles aimed at avoiding the problem

of degeneracy of the algorithm, that is, avoiding the situation that all but one of the weights are close to zero. The resampling step modifies the weighted approximate density $p(\mathbf{x}_k|\mathbf{z}_k)$ to an unweighted density $\hat{p}(\mathbf{x}_k|\mathbf{z}_k)$ by eliminating particles having low importance weights and by multiplying particles having highly importance weights. Formally:

$$p(\mathbf{x}_k|\mathbf{z}_k) = \sum_{i=1}^{N} q_i \delta(\mathbf{x}_k - \mathbf{x}_k{}^i) \qquad (15)$$

is replaced by

$$\hat{p}(\mathbf{x}_k|\mathbf{z}_k) = \sum_{i=1}^{N} \frac{1}{N} \delta(\mathbf{x}_k - \mathbf{x}_k{}^{\star}) = \sum_{i=1}^{N} \frac{n_i}{N} \delta(\mathbf{x}_k - \mathbf{x}_k{}^i) \qquad (16)$$

where n_i is the number of copies of particle $\mathbf{x}_k{}^i$ in the new set of particles $\{\mathbf{x}_k{}^{\star}\}$. Generically, it is implemented as follows:

- Draw N particles $\{\tilde{\mathbf{x}}_k^i\}_{i=1,\cdots,N}$ from the uniform distribution.

- Assign the resampled particles $\{\tilde{\mathbf{x}}_k^i\}_{i=1,\cdots,N}$ to $\{\mathbf{x}_k^i\}_{i=1,\cdots,N}$ and assign equal weights $\frac{1}{N}$ to each particle.

For extensive review of particle filtering see [Van Leeuwen 09].

3.2. The Ensemble Kalman Filter

The ensemble Kalman filter (EnKF) was first proposed by Evensen [Evensen 94] and further developed by [Burgers et al. 98] and [Evensen 03, Evensen07]. For the current status and the potential of the EnKF we refer to [Kalnay 09]. It is related to particle filters in the context that a particle is identical to an ensemble member. EnKF is a sequential filter method, which means that the model is integrated forward in time and, whenever observations are available, these are used to reinitialize the model before the integration continues. The EnKF originated as a version of the Extended Kalman Filter (EKF) [Jazwinski 70],[Bucy 65] for large problems. The classical KF [Kalman 60] method

is optimal in the sense of minimizing the variance only for linear systems and Gaussian statistics.

Similar to the particle filter method, the EnKF stems from a Monte Carlo integration of the Fokker-Planck equation governing the evolution of the PDF that describes the prior, forecast, and error statistics. In the analysis step, each ensemble member is updated according to the KF scheme and replaces the covariance matrix by the sample covariance computed from the ensemble. However, the EnKF presents two potential problems namely:

1) Even though the EnKF uses full non-linear dynamics to propagate the forecast error statistics, the EnKF assumes that all probability distributions involved are Gaussian.

2) The updated ensemble preserves only the first two moments of the posterior.

Let $p(\mathbf{x})$ denote the Gaussian prior probability density distribution of the state vector $\mathbf{x}$ with mean μ and covariance $\mathcal{Q}$

$$p(\mathbf{x}) \propto exp\left(\frac{-1}{2}(\mathbf{x} - \mu)^T \mathcal{Q}^{-1}(\mathbf{x} - \mu)\right)$$

We assume the data $\mathbf{z}$ to have a Gaussian PDF with covariance $\mathcal{R}$ and mean $\mathcal{H}\mathbf{x}$, where $\mathcal{H}$ is the so-called the observation matrix, is related to h of equation (6), and where the value $\mathcal{H}\mathbf{x}$ assumes the value of the data $\mathbf{z}$ would be for the state $\mathbf{x}$ in absence of observation errors. Then the conditional probability or likelihood $p(\mathbf{z}|\mathbf{x})$ assumes the form

$$p(\mathbf{z}|\mathbf{x}) \propto exp\left(\frac{-1}{2}(\mathbf{z} - \mathcal{H}\mathbf{x})^T \mathcal{R}^{-1}(\mathbf{z} - \mathcal{H}\mathbf{x})\right).$$

According to the Bayes theorem the posterior probability density follows from the relation

$$p(\mathbf{x}|\mathbf{z}) \propto p(\mathbf{z}|\mathbf{x})p(\mathbf{x}). \tag{17}$$

There are many variants of implementing the EnKF of various computational efficiency and in what follow we employ standard formulation of the EnKF for linear and nonlinear observation operators with covariance localization. See [Evensen 94, Burgers et al. 98, Mandel 06, Mandel 07 and Lewis et al.06], also see [Nerger et al. 05] and Sakov and Oke [Sakov 08] The implementation of the standard EnKF may be divided into three steps, as follows:

- **Setting and matching**

■ Define the ensemble

$$\mathcal{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_N] \tag{18}$$

be an $n_x \times N$ matrix whose columns are a sample from the prior distribution. N being the number of the ensemble members.

■ Form the ensemble mean

$$\bar{\mathcal{X}} = \mathcal{X} \cdot \mathbf{1}_N, \tag{19}$$

where $\mathbf{1}_N \in \mathbb{R}^{N \times N}$ is the matrix where each element is equal to 1.

■ Define the ensemble perturbation matrix $\mathcal{X}'$ and set the $\mathbb{R}^{n_x \times n_x}$ ensemble covariance matrix $\mathcal{C}$

$$\mathcal{X}' = \mathcal{X} - \frac{1}{N}\bar{\mathcal{X}}, \tag{20}$$

$$\mathcal{C} = \frac{\mathcal{X}'\mathcal{X}'^T}{N - 1}, \tag{21}$$

- **Sampling**

■ Generate

$$\mathcal{Z} = [\mathbf{z}_1, \cdots, \mathbf{z}_N] \tag{22}$$

be an $n_z \times N$ matrix whose columns are a replicate of the measurement vector $\mathbf{z}$ plus a random vector from the normal distribution $\mathcal{N}(0, \mathcal{R})$.

■ Form the $\mathbb{R}^{n_z \times n_z}$ measurement error covariance

$$\mathcal{R} = \frac{\mathcal{Z}\mathcal{Z}^t}{N-1}, \tag{23}$$

- **Updating** Obtain the posterior $\mathcal{X}^p$ by the linear combinations of members of the prior ensemble

$$\mathcal{X}^p = \mathcal{X} + \mathcal{C}\mathcal{H}^T(\mathcal{H}\mathcal{C}\mathcal{H}^T + \mathcal{R})^{-1}(\mathcal{Z} - \mathcal{H}\mathcal{X}) \tag{24}$$

The matrix

$$\mathcal{K} = \mathcal{C}\mathcal{H}^T(\mathcal{H}\mathcal{C}\mathcal{H}^T + \mathcal{R})^{-1} \tag{25}$$

is the Kalman gain matrix. Since $\mathcal{R}$ is always positive definite(i.e. covariance matrix), the inverse $(\mathcal{H}\mathcal{C}\mathcal{H}^T + \mathcal{R})^{-1}$ exists. An easy computation shows that the mean and covariance of the posterior or updated ensemble are given by

$$\bar{\mathcal{X}}^p = \mathcal{X}^p + \mathcal{K}\left[\mathbf{z} - (\mathcal{H}\mathcal{X}^p + \mathbf{d})\right], \tag{26}$$

and

$$\mathcal{C}^p = \mathcal{C} - \mathcal{K}\left[\mathcal{H}\mathcal{C}\mathcal{H}^T + \mathcal{R}\right]\mathcal{K}^T, \tag{27}$$

the vector $\mathbf{d}$ which appears in (26) stems from the affine measurement relation

$$h(\mathbf{x}) = \mathcal{H}\mathbf{x} + \mathbf{d}. \tag{28}$$

In the case of nonlinear observation operators, a modification to the above algorithm is advised. As presented in Evensen [Evensen 03], let $\hat{\mathbf{x}}$ the augmented state vector made of the state vector and the predicted observation vector (nonlinear in this case).

$$\hat{\mathbf{x}} = \begin{pmatrix} \mathbf{x} \\ \mathcal{H}(\mathbf{x}) \end{pmatrix}. \tag{29}$$

Define the linear observation operator $\hat{\mathcal{H}}$ by

$$\hat{\mathcal{H}} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \mathbf{y} \tag{30}$$

and carry out the steps of the EnKF formulation in augmented state space $\hat{\mathbf{x}}$ and $\hat{\mathcal{H}}$ instead of $\mathbf{x}$ and $\mathcal{H}$. Superficially, this technique appears to reduce the nonlinear problem to the previous linear observation operator case. However, whilst the augmented problem, involving linear observation problem, is a reasonable way of formulating the EnKF, it is not as well-founded as the linear case, which can be justified as an approximation to the exact and optimal KF.

To prevent the occurrence of filter divergence usually due to the background-error covariance estimates from small number of ensemble members as pointed out in [Houtekamer and Mitchell 98], the use of covariance localization was suggested. Mathematically, the covariance localization increases the effective rank of the background error covariances. See the work of [Gaspari and Cohn 99] also [Hamill and Snyder 2000, 2006] and [Ehrendorfer 07]. The covariance localization consists of multiplying point by point the covariance estimate from the ensemble with a correlation function that is 1.0 at the observation location and zero beyond some prescribed distance. Mathematically, to apply covariance

localization, the Kalman gain

$$\mathcal{K} = \mathcal{C}\mathcal{H}^T(\mathcal{H}\mathcal{C}\mathcal{H}^T + \mathcal{R})^{-1}$$

is replaced by a modified gain

$$\hat{\mathcal{K}} = [\rho \circ \mathcal{C}]\,\mathcal{H}^T(\mathcal{H}\,[\rho \circ \mathcal{C}]\,\mathcal{H}^T + \mathcal{R})^{-1} \tag{31}$$

where $\rho\circ$ denotes the Schur product (The Schur product of matrices $\mathcal{A}$ and $\mathcal{B}$ is a matrix $\mathcal{D}$ of the same dimension, where $d_{ij} = a_{ij}b_{ij}$) of a matrix $\mathcal{S}$ with local support with the covariance model generated by the ensemble. Various correlation matrices have been used. For the present model, we used the usual Gaussian correlation function

$$\rho(D) = \exp[-[\frac{D}{l}]^2, l = 200\text{length units}] \tag{32}$$

In addition, the additive covariance inflation of Anderson and Anderson [Anderson 99] with $r = 1.001$ has been employed.

3.3. The Maximum Likelihood Ensemble Filter

The Maximum Likelihood Ensemble Filter (MLEF) proposed by Zupanski [Zupanski 05], and Zupanski and Zupanski [ZupanskiD 06] is a hybrid filter combining the 4-D variational method with the EnKF. It maximizes the likelihood of posterior probability distribution which justifies its name. The method comprises three steps, a forecast step that is concerned with the evolution of the forecast error covariances, an analysis step based on solving a non-linear cost function and updating step.

- **Forecasting:** It consists of evolving the square root analysis error covariance matrix through the ensembles. The starting point is from the evolution equation of the discrete

Kalman filter described in Jazwinski [Jazwinski 70]

$$P_f^k = \mathcal{M}_{k-1,k} P_a^k \mathcal{M}_{k-1,k}^T + Q_{k-1}, \tag{33}$$

where $P_f^(k)$ is the forecast error covariance matrix at time k, $\mathcal{M}_{k-1,k}$ the non-linear model

evolution operator from time $k-1$ to time k, and Q_{k-1} is the model error matrix which

is assumed to be normally distributed. Since P_a^{k-1} is positive matrix for any k, equation

(33) could be factorized and written as

$$P_f^k = \overbrace{\left(\mathcal{M}_{k-1,k} (P_a^k)^{1/2} \right)}^{(P_f^k)^{1/2}} \left(\mathcal{M}_{k-1,k} (P_a^k)^{1/2} \right)^T + Q_{k-1} \tag{34}$$

where $(P_a^k)^{1/2}$ is of the form

$$(P_a^k)^{1/2} = \begin{pmatrix} p_{(1,1)}^k p_{(2,1)}^k \cdots p_{(N,1)}^k \\ p_{(1,2)}^k p_{(2,2)}^k \cdots p_{(N,2)}^k \\ \\ p_{(1,n)}^k p_{(2,n)}^k \cdots p_{(N,n)}^k \end{pmatrix}, \tag{35}$$

as usual N is the number of ensemble members and n the number of state variables. The

lower case $p_{i,j}^k$ are obtained by calculating the square root of (P_a^k). Using equation(35),

the square root forecast error covariance matrix $(P_f^k)^{1/2}$ can then be expressed as

$$(P_f^k)^{1/2} = \begin{pmatrix} b_{(1,1)}^k b_{(2,1)}^k \cdots b_{(N,1)}^k \\ b_{(1,2)}^k b_{(2,2)}^k \cdots b_{(N,2)}^k \\ \\ b_{(1,n)}^k b_{(2,n)}^k \cdots b_{(N,n)}^k \end{pmatrix}, \tag{36}$$

where for each $1 \leq i \leq N$

$$\mathbf{b}_i^k = \begin{pmatrix} b_{(i,1)}^k \\ b_{(i,2)}^k \\ \vdots \\ b_{(i,n)}^k \end{pmatrix} = \mathcal{M}_{k-1,k} \begin{pmatrix} x_1^k + p_{(i,1)}^k \\ x_2^k + p_{(i,2)}^k \\ \vdots \\ x_n^k + p_{(i,n)}^k \end{pmatrix} - \mathcal{M}_{k-1,k} \begin{pmatrix} x_1^k \\ x_2^k \\ \vdots \\ x_n^k \end{pmatrix}. \tag{37}$$

The vector $\mathbf{x}^k = \left(x_1^k x_2^k \cdots x_n^k \right)^T$ is the analysis state from the previous assimilation cycle.

which is found from the posterior analysis pdf as presented in Lorenc 92 [@].

  • **Analyzing:** The analysis step for the MLEF involves solving a non-linear minimiza-

tion problem. As in Lorenc [Lorenc 86], the associated cost function is defined in terms

of the forecast error covariance matrix and is given as

$$\mathcal{J}(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \left(P_f^k\right)^{-1}(\mathbf{x} - \mathbf{x}_b) + \frac{1}{2}\left[\mathbf{y} - h(\mathbf{x})\right]^T \mathcal{R}^{-1}\left[\mathbf{y} - h(\mathbf{x})\right] \tag{38}$$

where $\mathbf{y}$ is the vector of observations, h is the non-linear observation operator, $\mathcal{R}$ is the observational covariance matrix and $\mathbf{x}_b$ is a background state given by

$$\mathbf{x}_b = \mathcal{M}_{k-1,k}(\mathbf{x}^k) + Q_{k-1}. \tag{39}$$

Through a Hessian preconditioner we introduce the change of variable

$$(\mathbf{x} - \mathbf{x}_b) = (P_f^k)^{1/2}(\mathcal{I} + \mathcal{O})^{-T/2}\xi \tag{40}$$

where ξ is vector of control variables, $\mathcal{O}$ is referred to as the observation information matrix and $\mathcal{I}$ is the identity matrix. The matrix $\mathcal{O}$ is provided by

$$\mathcal{O} = (P_f^k)^{T/2}\tilde{\mathcal{H}}^T\tilde{\mathcal{H}}\mathcal{R}^{-1}\tilde{\mathcal{H}}(P_f^k)^{T/2} = (\mathcal{R}^{-1/2}\tilde{\mathcal{H}}(P_f^k)^{1/2})^T(\mathcal{R}^{-1/2}\tilde{\mathcal{H}}(P_f^k)^{1/2}) \tag{41}$$

here $\tilde{\mathcal{H}}$ is the Jacobian matrix of the non-linear observation operator h evaluated at the background state $\mathbf{x}_b$.

Let $\mathcal{Z}$ the matrix defined by

$$\mathcal{Z} = \begin{pmatrix} z_{(1,1)} z_{(2,1)} \cdots z_{(N,1)} \\ z_{(1,2)} z_{(2,2)} \cdots z_{(N,2)} \\ z_{(1,n)} z_{(2,n)} \cdots z_{(N,n)} \end{pmatrix}, \mathbf{z}_i = \begin{pmatrix} z_{(i,1)} \\ z_{(i,2)} \\ \vdots \\ z_{(i,n)} \end{pmatrix} = \mathcal{R}^{-1/2}\tilde{\mathcal{H}}\mathbf{b}_i^k, \tag{42}$$

using the following approximations

$$\mathbf{z}_i \approx \mathcal{R}^{-1/2}\left[h(\mathbf{x} + \mathbf{b}_i^k) - h(\mathbf{x})\right], \tag{43}$$

and

$$\mathcal{O} \approx \mathcal{Z}\mathcal{Z}^T. \tag{44}$$

one can use an eigenvalue decomposition of the of symmetric positive definite matrix

$\mathcal{I} + \mathcal{O}$ to calculate the inverse square root matrix necessary to the updating step. Its

worth mentioning that the approximation (43) is not necessary and a derivation of the

MLEF not involving (43) has been recently developed in Zupanski et al. [Zupanski 08]

• **Updating:** The final point about MLEF is to update the square root analysis error

covariance matrix. In order to estimate the analysis error covariance at the optimal point,

the optimal state $\mathbf{x}_{opt}$ minimizing the cost function $\mathcal{J}$ given by (38)is substituted

$$(P_a^k)^{T/2} = (P_f^k)^{T/2}(\mathcal{I} + \mathcal{O}(\mathbf{x}_{opt}))^{-T/2}, \tag{45}$$

4. Numerical results

4.1. Set-up of numerical experiments

The following table summarizes the logic of the numerical experiments conducted in

this paper

Filter	Features	Observation operator type
PF	resampling	linear and nonlinear
EnKF	standard and with localization and inflation	linear and nonlinear
MLEF	standard	linear and nonlinear

4.2. Model set-up

As in [Williamson 92,97,07] and [Jakob et al.95], the grid representation for any arbitrary

variable ϕ is related to the following spectral decomposition

$$\phi(\lambda, \mu) = \sum_{m=-M}^{M} \sum_{n=|m|}^{\mathfrak{N}(m)} \phi_{m,n} P_{m,n}(\mu) e^{im\lambda}, \tag{46}$$

where $P_{m,n}(\mu)e^{im\lambda}$ are the spherical harmonic functions [Boyd 01]. $P_{m,n}(\mu)$ stands for

the Legendre polynomial. M is the highest Fourier wavenumber included in the east-west

representation, $\mathfrak{N}(m)$ is the highest degree of the associated Legendre polynomials for

longitudinal wavenumber m.

The coefficients of the spectral representation (46) are determined by

$$\phi_{m,n} = \int_{-1}^{1} \frac{1}{2\pi} \int_{0}^{2\pi} \phi(\lambda, \mu) e^{-im\lambda} P_{m,n}(\mu) d\lambda d\mu \qquad (47)$$

The inner integral represents a Fourier transform,

$$\phi_m(\mu) = \frac{1}{2\pi} \int_{0}^{2\pi} \phi(\lambda, \mu) e^{-im\lambda} d\lambda \qquad (48)$$

which is evaluated using a fast Fourier transform routine. The outer integer is evaluated using Gaussian quadrature on the transform grid.

$$\phi_{m,n} = \sum_{j=1}^{J} \phi_m(\mu) P_{m,n}(\mu) \omega_j, \qquad (49)$$

where μ_j denotes the Gaussian grid points in the meridional direction and ω_j is the Gaussian weight at point μ_j. The meridional grid points are located at the Gaussian latitudes θ_j, which are the J roots of the Legendre polynomial $P_j(\sin \theta_j) = 0$. The number of grid points in the longitudinal and meridional directions are determined so as to allow the unaliased representation of quadratic terms,

$$\begin{cases} I \geq 3M + 1 \\ J \geq \dfrac{3N + 1}{2} \end{cases}$$

where N is the highest wavenumber retained in the latitudinal Legendre representation $N = \max \mathfrak{N}(m) = M$. The pseudo spectral method, also known as the spectral transform method in the geophysical community, of [Orszag 1969, 1970] and [Eliassen et al. 1970] has been used to tackle the nonlinearity.

In conjunction with the spatial discretization described before, the time discretization, two semi-implicit time steps have been used for the initialization. Because of the hyper-

bolic type of the shallow water equations, the centered leapfrog scheme

$$\frac{\phi_{m,n}^{k+1} - \phi_{m,n}^{k-1}}{2\Delta t} = \mathfrak{F}(\phi_{m,n}^{k}) \tag{50}$$

has been invoked for the subsequent time steps. After the leapfrog time-differencing scheme is used to obtain the solution at $t = (k+1)\Delta t$, a slight time smoothing is applied to the solution at time $k\Delta t$ (Asselin filter) .

$$\bar{\phi}_{m,n}^{k} = \phi_{m,n}^{k} + \alpha \left[\phi_{m,n}^{k+1} - 2\phi_{m,n}^{k} + \bar{\phi}_{m,n}^{k-1} \right] \tag{51}$$

replacing the solution at time k. The added term acts as a smoother in time. It reduces the amplitude of different frequencies ν by a factor $1 - 4\alpha \sin^2(\frac{\nu\Delta t}{2})$.

4.3. EnKF set-up and results

In all our numerical experiments a time step of $\Delta t = 600 sec$ has been employed. The observations were provided at a frequency consisting of one observation every $36\Delta t$.

4.3.1. Impact of the number of observations and the number of ensemble members

The impact of the number of observations follows the findings of [Fletcher and Zupanski 2008]. In fact after attaining a number of observations threshold, all the root mean square errors RMSE for all the fields (namely the components of the velocity and geopotential) coincide. In our test the number of the observations threshold was attained at around 912 observations. The observational grid is a subset of the rhomboidal spectral truncation (48x38) and consists of observations being distributed at every grid points in the longitude and every 2 points in the latitude.

The impact of the number of ensemble members on the results has been also examined. We have conducted several experiments with different number of ensemble mem-

bers, namely, 50, 100, 200 and 300 ensemble members respectively. Our results reveal that employing only 100 ensemble members was sufficient enough to successfully perform the data assimilation and that using higher number of ensemble members had no impact on the the ensuing results.

4.3.2. Linear case - description of the EnKF results

Figure 1 gives an overview of the unperturbed shallow water solution using the Rossby-Haurwitz Wave no 4 as initial conditions after 10 days of time integration.

In Figures 2 and 3 we present a 1% random perturbation around the mean of the geopotential and the divergence field respectively. These perturbed fields along with the perturbed velocity and vorticity will serve as initialization for the observations.

In figure 4 we present an overview of the results obtained using the 10 days EnKF time integration. The linear observation operator $\mathcal{H}(u) = u$ has been employed, 100 ensemble members were used. and 1% random perturbation around the mean accounting for the standard deviation for the model error covariance matrix has been employed Figures 5 and 6 present the resulting geopotential and divergence fields of the EnKF R-H wave 4 in the presence of a linear observation operator. We observe that the R-H wave 4 preserves its global characteristic shapes with changes in the center of the recirculation zones. The divergence field, which is the most sensitive field for this test case, displays a distortion as compared with the non-perturbed R-H wave 4.

To gain a deeper understanding of the impact of data assimilation upon the evolution of conservation of the quadratic integral invariants of the shallow water equations model, we considered a time evolution of the normalized mass, total energy and potential enstrophy respectively. The test was conducted for 10 days of EnKF in the presence of a linear

observation operator, 100 ensemble members and 1% random perturbation around the

mean. The result presented in Figure 7 reveals that while the evolution in time of the

normalized potential enstropy is oscillatory but bounded, the normalized mass integral

invariant displays a slight increase for the window of data assimilation under discussion.

The normalized total energy exhibits a consistent increase amounting to a factor of 2.5 at

the end of the 10 days window of data assimilation.

4.3.3. Nonlinear case - description of the EnKF results

We are employing $\mathcal{H}(u) = u^2$ as a representative of the nonlinear observation operator (

we have also tested $\mathcal{H}(u) = u^4$) The number of observation points is 912 (their distribution

in space and time being similar to the linear observation operator EnKF case).

In figure 8 we present an overview of the results obtained using the 15 days EnKF time

integration. The nonlinear observation operator $\mathcal{H}(u) = u^2$ has been employed, Figures

9 and 10 are the geopotential and the divergence fields respectively with nonlinear obser-

vation operator after 15 days of data assimilation and where 100 ensemble members are

used. The filter appears to distort the fields. In fact the divergence field, the most sensi-

tive field, lost its typical pattern formation and became grossly distorted accompanied by

an increase in its values. The geopotential field still has the 4 cells pattern but its shape

is distorted - a sign of filter divergence but very out of shape after the 15 days of data

assimilation. Similar remarks apply when the degree on nonlinearity was increased from

$\mathcal{H}(u) = u^2$ to $\mathcal{H}(u) = u^4$. More distortion and typical cell pattern breakdown can been

seen in Figures 11 and 12.

The result presented in Figure 13 correspond to the time evolution of the normalized mass, total energy and potential enstrophy respectively. 15 days of EnKF in the presence of the nonlinear observation operator $\mathcal{H}(u) = u^4$ was used. Again 100 ensemble members and 1% random perturbation around the mean was employed. In contrast with the case of linear observation operator, the time evolution of each of the invariants is now oscillatory with consistent increase in their amplitude.

References

D. L. Williamson, J. B. Drake, J. J. Hack, R. Jacob and P. N. Swarztrauber, "A standard test set for numerical approximations to shallow-water equations in spherical geometry," *J. Comput. Phys.* **102** 211–224 (1992).

D. L. Williamson "Climate Simulations with a spectral semi-Lagrangian model with linear grid," Numerical Methods in Atmospheric and Ocean Modelling. *Canadian Meteorological and Oceanographic Society* 279–292 (1997).

D. L. Williamson "The Evolution of Dynamical Cores for Global Atmospheric Models," *Journal of the Meteorological Society of Japan* **85B** 241–269 (2007)

R. Jakob-Chien, J. J. Hack and D. L. Williamson, "Spectral transform solutions to the shallow water test set," *J. Comput. Phys.* **119** 164–187 (1995).

J. P. Boyd , "Chebyshev and Fourier Spectral Methods, Second Edition (Revised)," *Dover Publishing.* (2001).

M. L. Berliner and C. K. Wikle, "A Bayesian tutorial for data assimilation," *Physica D,* **30** 1–16 (2007).

M. L. Berliner and C. K. Wikle, "Approximate importance sampling Monte Carlo for data assimilation,' *Physica D,* **30** 37-49 (2007)

A. Doucet, A. , S. Godsill, S. and C. Andrieu, "On Sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing,* **10 (3)** 197–208 (2000).

A. Doucet, J. F. G. de Freitas and N. J. Gordon, "An introduction to sequential Monte Carlo methods," *Sequential Monte Carlo Methods in Practice, eds Doucet, A. , de Freitas, J.F.G,and Gordon, N.J , Springer-Verlag, New York,* (2001).

M. S. Arulampalam, S. Maskell, N. Gordon and T. Clapp, "A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking," *IEEE transactions on signal processing,* **Vol(150) no.2,** 174–188 (2002).

N. J. Gordon, D. J. Salmond and A. F. M. Smith, "Novel approach to nonlinear non-Gaussian Bayesian state estimate," *IEEE Proc. F,* **140** 107–113 (1993).

J. S. Liu and R. Chen, "Sequential Monte Carlo Methods for Dynamic Systems," *Journal of the American Statistical Association,* **Vol.93, No.442** 1032-1044 (1998).

R.E. Kalman, "A new approach to linear filtering and prediction problems," *Transaction of the ASME - Journal of Basic Engineering,* **Series D, 82** 35–45 (1960).

G. Evensen, "Sequential data assimilation in non-linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics," *J. Geophys Res,* **99** (C5):10 108– (1994).

G. Evensen and P. J. Van Leeuwen, "Assimilation of Geosat altimeter data for Agulhas Current using the Ensemble Kalman filter with quasi-geostrophic model," *Monthly Weather Review,* **124**(1) 85-96 (1996).

G. Burgers, P. J. Van Leeuwen and G. Evensen, "Analysis scheme in the ensemble Kalman Filter," *Monthly Weather Review,* **126** 1719-1724 (1998).

G. Evensen, "The Ensemble Kalman Filter: theoretical formulation and practical implementation," *Ocean Dynamics,* **53** 343-367 (2003).

G. Evensen, "Data Assimilation: The ensemble Kalman filter," *Springer, Berlin,* (2007).

J. Mandel, "A brief Tutorial on the ensemble Kalman filter" *UCDHSC/CCM Report No.242* (2007).

J. Mandel, "Efficient Implementation of the Ensemble Kalman Filter," *CCM Report 231,* (2006).

A. H.Jazwinski, "Stochastic processes and filtering theory," *Academic Press, New York,* (1970).

R. S. Bucy, "Recurrent sets," *Ann. Math. Statist.,* **36** 535–545 (1965).

J. L. Anderson and S. L. Anderson, "A Monte Carlo Implementation of the Nonlinear Filtering Problem to Produce Ensemble Assimilations and Forecasts," *Mon. Wea. Rev.,* **127** 2741–2785 (1999).

J. L. Anderson, "A local least squares framework for ensemble filtering," *Monthly Weather Review,* **131** 634–642 (2003).

C. H. Bishop, B. J. Etherton and S. J. Majumdar, "Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects.," *Monthly Weather Review,* **129**(3) 420–436 (2001).

M. K.,Tippett, J. L. Anderson, C. H. Bishop and T.M. Hamill, " Ensemble square root filters," *Monthly Weather Review,* **131** 1485–1490 (2003).

J. L. Lewis, S. Lakshivarahan and S. K. Dhall, "Dynamic Data Assimilation: A least squares approach," *Cambridge University Press, in the series on Encyclopedia of Mathematics and its Applications,* Volume **104** (2006)

E. N. Lorenz, "Deterministic nonperiodic flow," *J. Atmos. Sci.,* **20** 130–141 (1963).

E. N. Lorenz, "The predictability of a flow which posses many scales of motion," *Tellus* **21** 289–307 (1969).

S. E. Cohn, "An introduction to estimation theory," *Journal of the Meteorological Society of Japan,* Vol.**75**, No.1B 257–288 (1997)

X. Xiong, I. M. Navon and B. Uzunoglu, *A Note on the Particle Filter with Posterior Gaussian Resampling* Tellus 58A, (2006) 456–460.

P. Houtekamer and H. Mitchell, "Data assimilation using an ensemble Kalman filter technique," *Mon. Wea. Rev.,* **126**, 796–811(1998).

P. L. Houtekamer and H. L. Mitchell, "A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation," *Mon.Wea.Rev.,***129,** 123–137 (2000).

S. L. Dance, "Issues in high resolution limited area data assimilation for quantitative precipitation forecasting," *Physica D* **196** 1–27 (2004).

G. Kotecha and P. M. Djurić, "Gaussian particle filtering," *IEEE Trans. Signal Processing,* **51**, 2592–2601 (2003).

D. E. Goldberg, "Genetic algorithms in search, optimization and machine learning," *Addison-Wesley, Reading,* (1989).

B. W. Silverman, "Density Estimation for Statistics and Data Analysis," *Chapman and Hall,* (1986).

T. Bengtsson, P. Bickel, and B. Li, "Curse-of-dimensionality revisited:Collapse of the particle filter in very large scale systems," *Probability and Statistics: Essays in Honor of David A. Freedman'* **Vol. 2,** 316–334 (2008).

P. Bickel, B. Li and T. Bengtsson, "Sharp Failure Rates for the Bootstrap Particle Filter in High Dimensions," *IMS Collections: Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh,* **Vol.3**, 318-329 (2008).

C. Snyder, T. Bengtsson, P. Bickel, and J. L. Anderson, "Obstacles to high-dimensional particle filtering," *Mon. Wea. Rev.,* **Vol 136**, No12 ,4629–4640 (2008).

D. Zupanski, "A general weak constraint applicable to operational 4DVAR data assimilation systems," *Mon.Wea.Rev.,* **125,** 2274-2292 (1997)

M. Zupanski, "Maximum Likelihood Ensemble Filter. Part1: the theoretical aspect," *Mon.Wea.Rev.,* **133,** 1710–1726 (2005).

D. Zupanski, M. Zupanski, "Model error estimation employing ensemble data assimilation approach," *Mon.Wea.Rev.,* **134,** 137–1354 (2006).

M. Zupanski, I. M. Navon and D. Zupanski, "The Maximum Likelihood Ensemble Filter as a non-differentiable minumization algorithm" *Q.J.R. Meteorol.* **134**: 1039–1050 (2008)

M. Jardak, I. M. Navon and M. Zupanski, "Comparison of Sequential data assimilation methods for the Kuramoto-Sivashinsky equation," In early view *International Journal for Numerical Methods in Fluids* DOI:10.1002/fld.2020

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

T. M. Hamill, J. S. Whitaker and C. Snyder, " Distance-Dependent Filtering of Background Error Covariance Estimates in an Ensemble Kalman Filter," *Mon.Wea.Rev.,***129,** 2776–2790 (2001).

T. M. Hamill and C. Snyder, "A Hybrid Ensemble Kalman Filter-3D Variational Analysis Scheme," *Mon.Wea.Rev.,***128,** 2905–2919 (2000).

T. M. Hamill, "Ensemble-based atmospheric data assimilation," *Predictability of Weather and Climate, Cambridge Press,* 124–156 (2006).

L. Nerger, ,W. Hiller, J. Schrater, "A comparison of error subspace Kalman filters".*Tellus A,***57 (5)***, 715–735 (2005)

G. Gaspari, S. E. Cohn, "Construction of correlation functions in two and three dimensions " *The Quarterly Journal of the Royal Meteorological Society*, **Vol 125 Issue 554**,723-757 (1999).

B. Uzunoglu, S.J. Fletcher, M. Zupanski and I. M Navon, "Adaptive ensemble reduction and inflation," *Quarterly Journal of the Royal Meteorological Society,* **133-626** 1281-1294 (2007)

A. Apte1 , C. K. R. T. Jones, A. M. Stuart and J. Voss, "Data assimilation: Mathematical and statistical perspectives," *Int. J. Numer. Meth. Fluids,* **56** 1033–1046 (2008)

A. Apte , C. K. R. T. Jones and A. M. Stuart, "A Bayesian approach to Lagrangian data assimilation," *Tellus A,***60**, 336–347 (2008)

R. K. Smith and D. G. Dritschel, "Revisiting the Rossby–Haurwitz wave test case with contour advection, " *Journal of Computational Physics,* Volume 217, Issue 2, Pages 473–484 (2006)

P. J. van Leeuwen, "Particle filtering in geophysical systems" *Mon.Wea.Rev.,*-Early online Releases.

E. Kalnay , "Ensemble Kalman Filter: Current Status and Potential" submitted as Chapter 4 of "Data Assimilation: Making sense of observations" Richard Menard, Editor, Springer, (2008)

P. Sakov and P. Oke, "Implications of the Form of the Ensemble Transformation in the Ensemble Square Root Filters" *Mon.Wea.Rev.,***136,** 1042–1052 (2008).

B. Haurwitz, "The motion of atmospheric disturbances on the spherical earth" *J. Mar. Res.* **3** 254–267 (1940).

J. Thuburn and Y. Li, "Numerical simulations of RossbyHaurwitz waves", *Tellus A* **52** (2) 181-189 (2000).

B. J. Hoskins, "Stability of the RossbyHaurwitz wave" *Quart. J. Roy. Meteorol. Soc.***99**, 723-745 (1973).

N.A. Phillips, "Numerical integration of the primitive equations on the hemisphere" *Mon. Wea. Rev.***87**, 333-345 (1959).

C. G. Rossby and collaborators. 1939 "Relation between variations in the intensity of the zonal circulation of the atmosphere and the displacements of the semi-permanent centres of action" *J. Marine Res.*, **2**, 38–55 (1939).
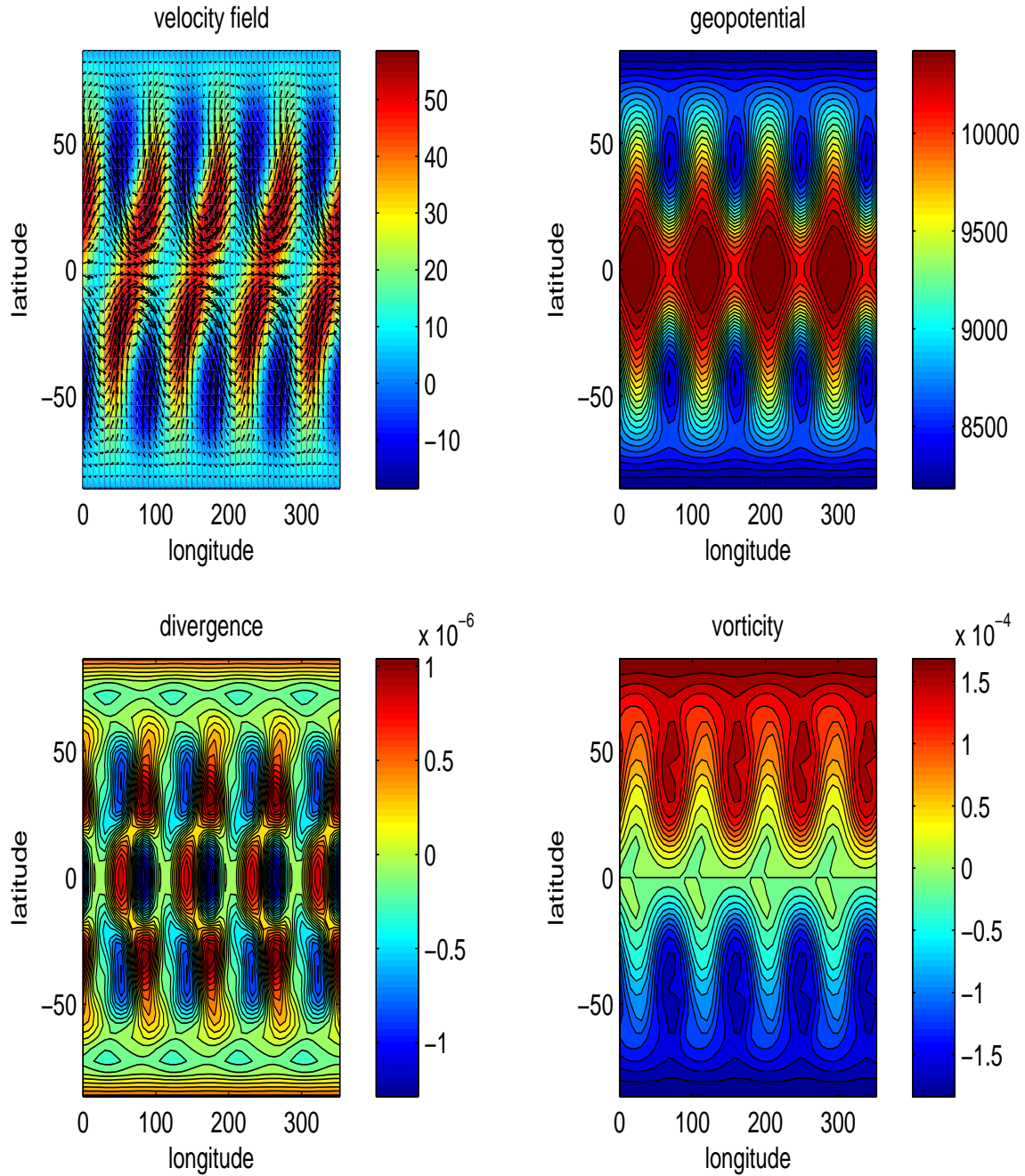
Figure 1. Non-perturbed shallow water solution R-H wave 4 case: velocity, geopotential, and divergence fields after 10 days of time integration
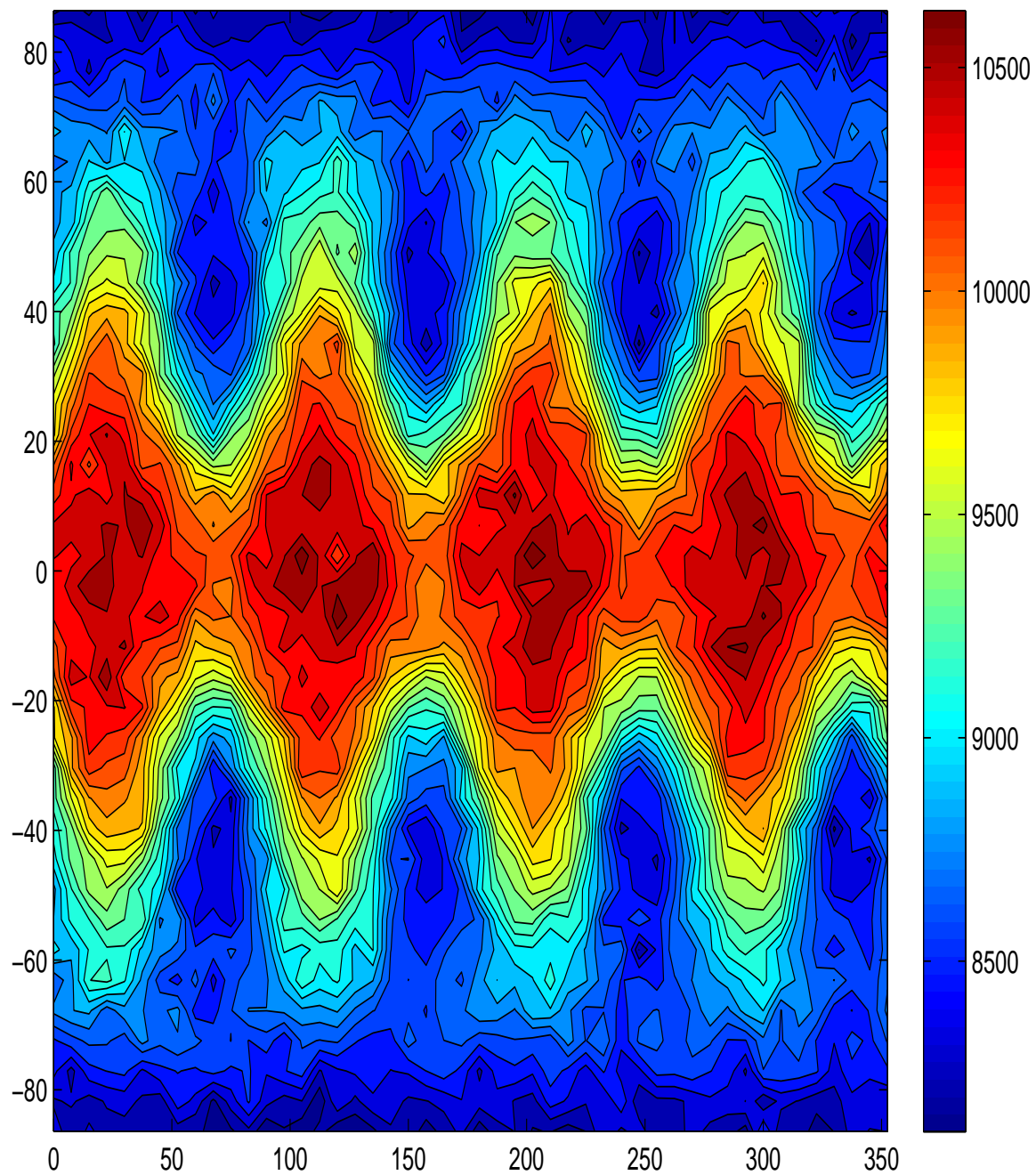
Figure 2. EnKF filter linear observation operator case : initial perturbation geopotential field
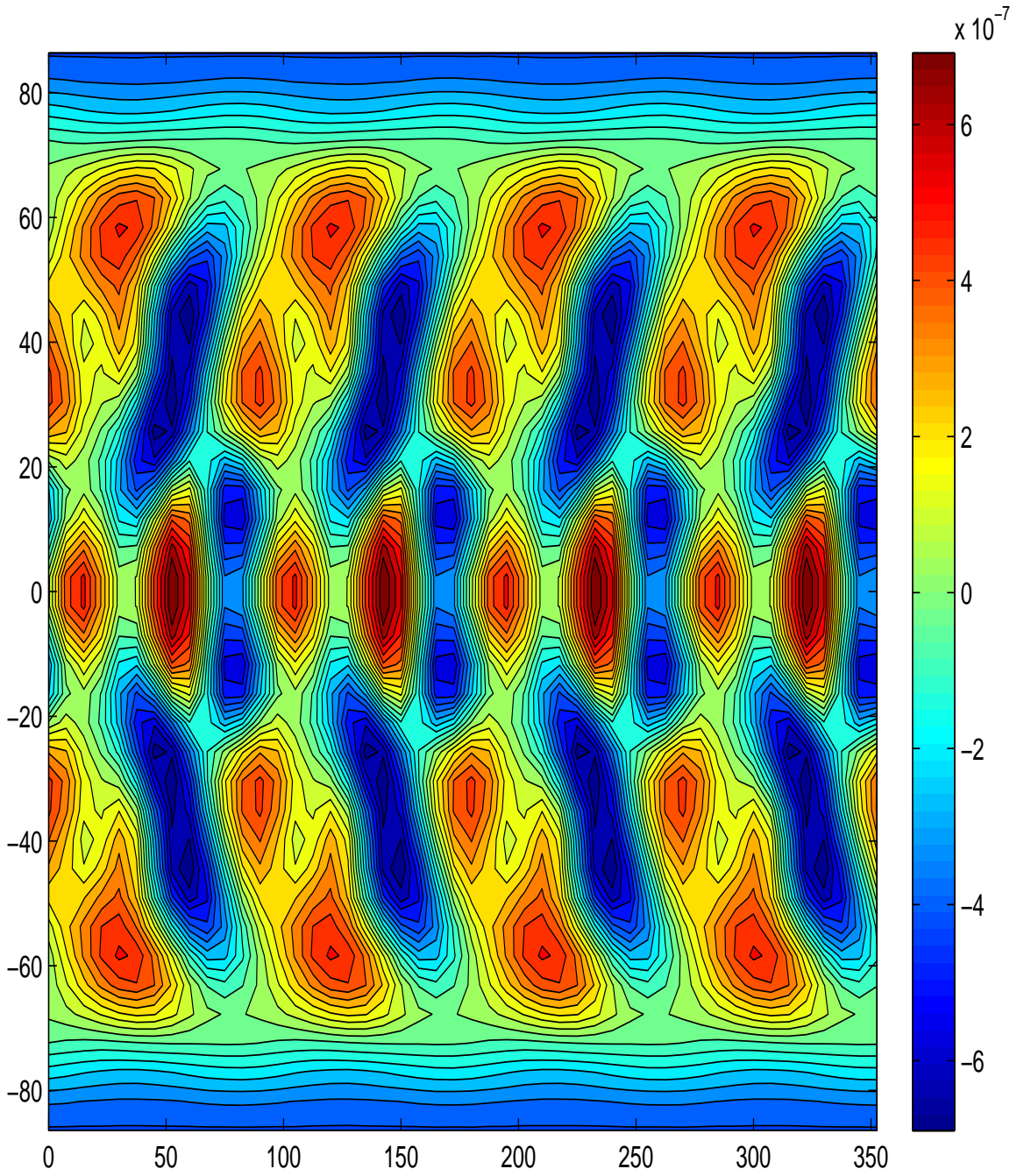
Figure 3. EnKF filter linear observation operator case : initial perturbation divergence

field divergence
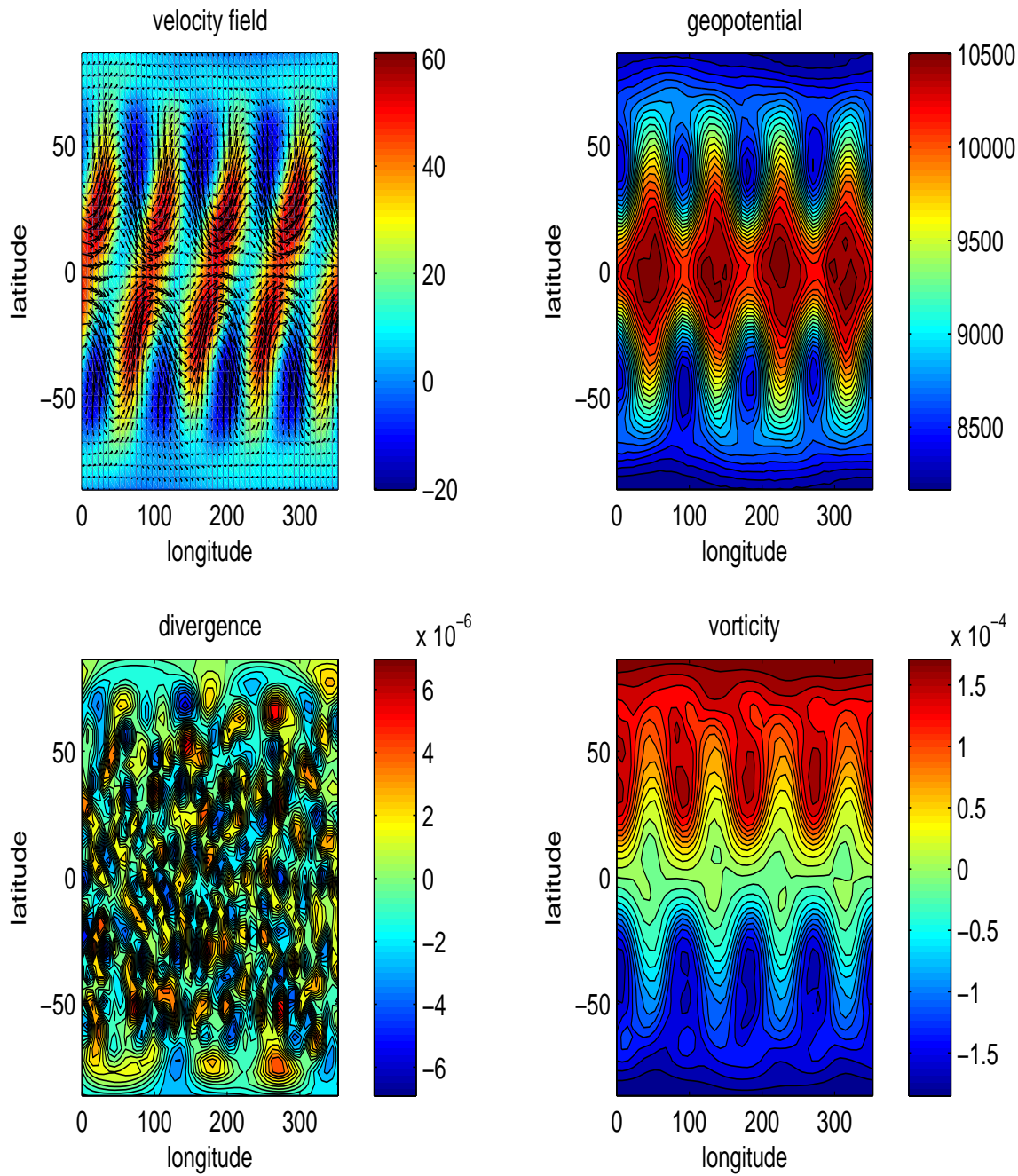
Figure 4. EnKF filter linear observation operator case : velocity, geopotential, and divergence fields after 10 days of time integration. 100 ensemble members and 1% random perturbation around the mean
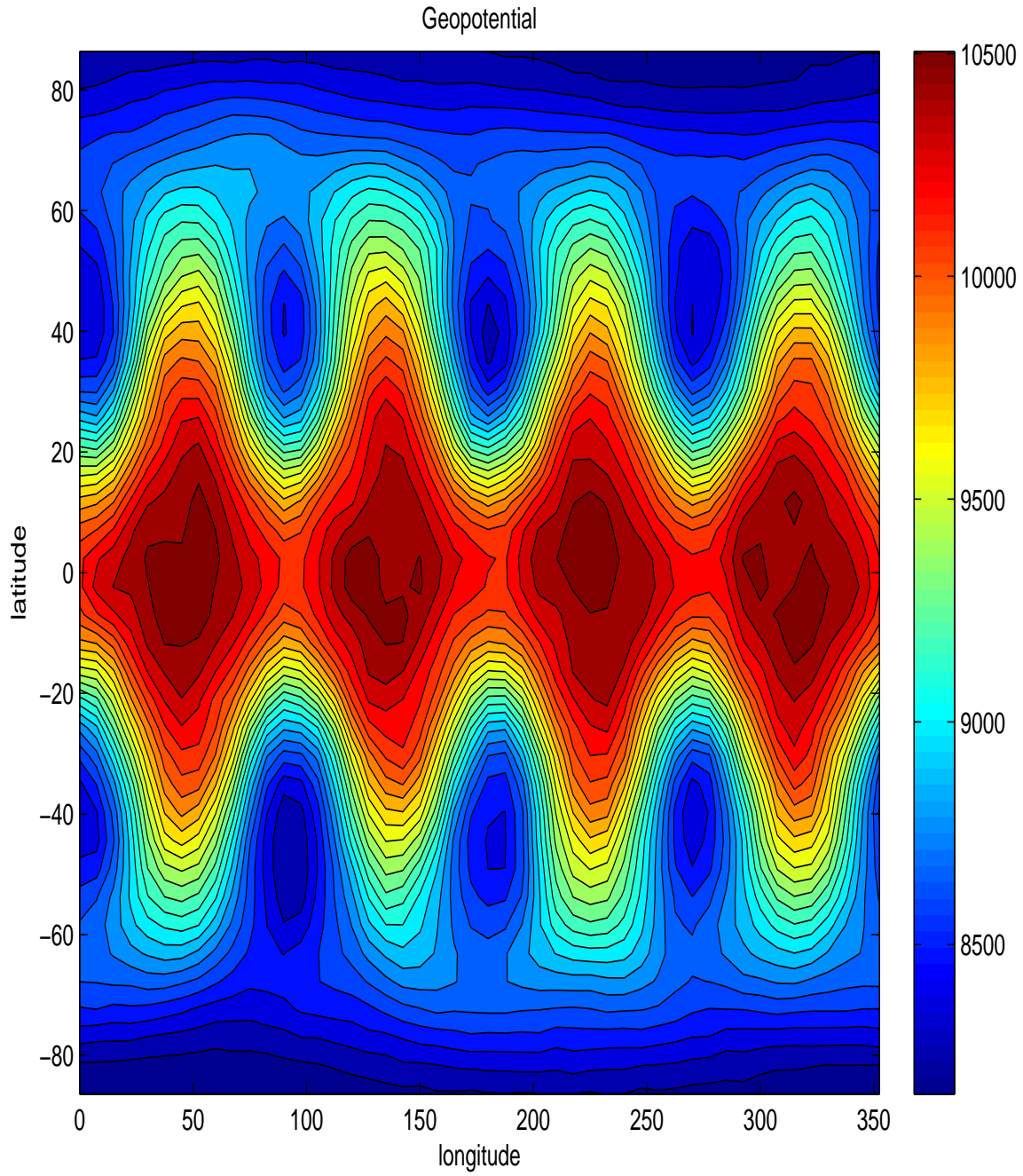
Figure 5. EnKF filter linear observation operator case: geopotential field, 100 ensemble members and 1% random perturbation around the mean after 10 days of time integration
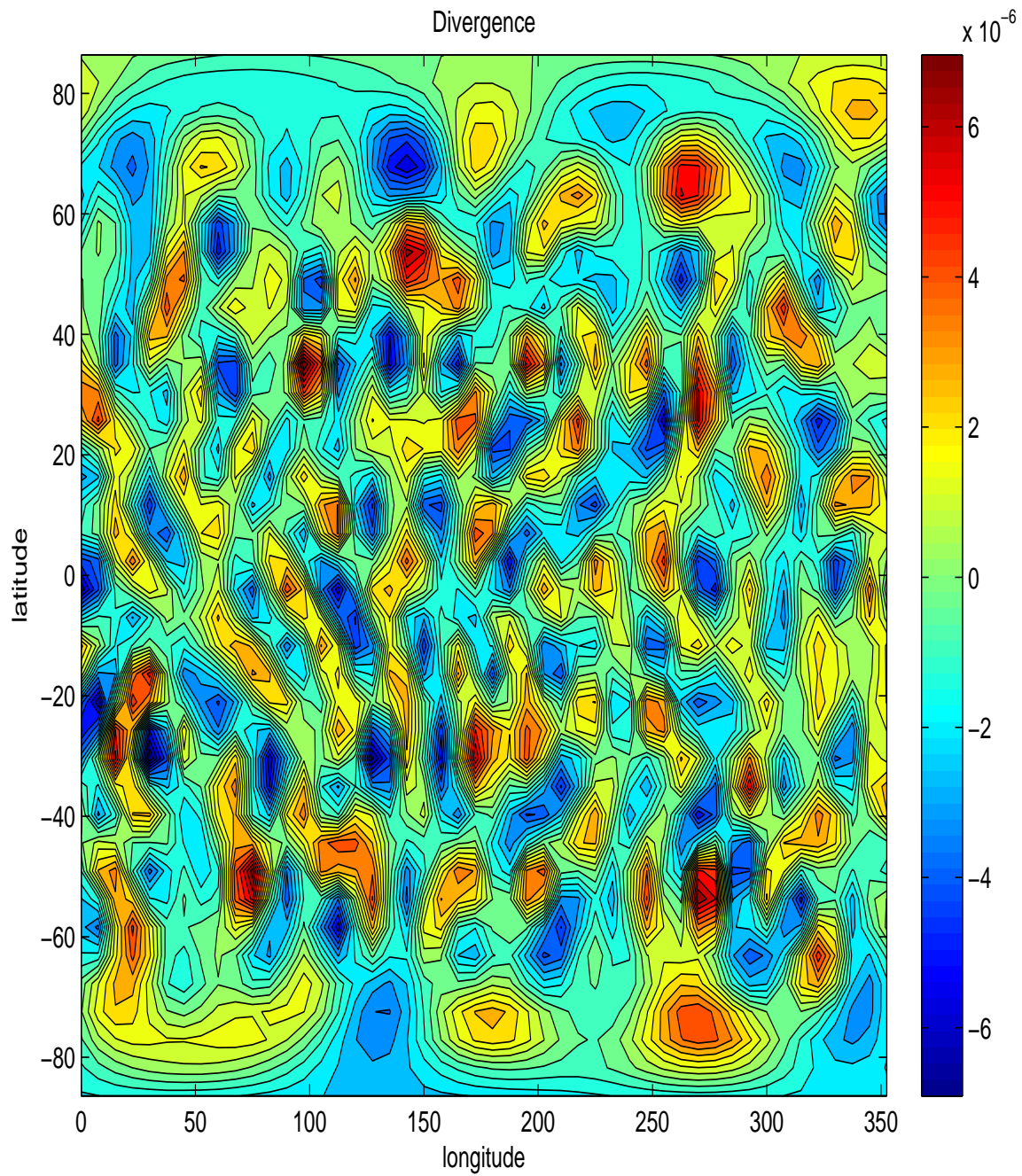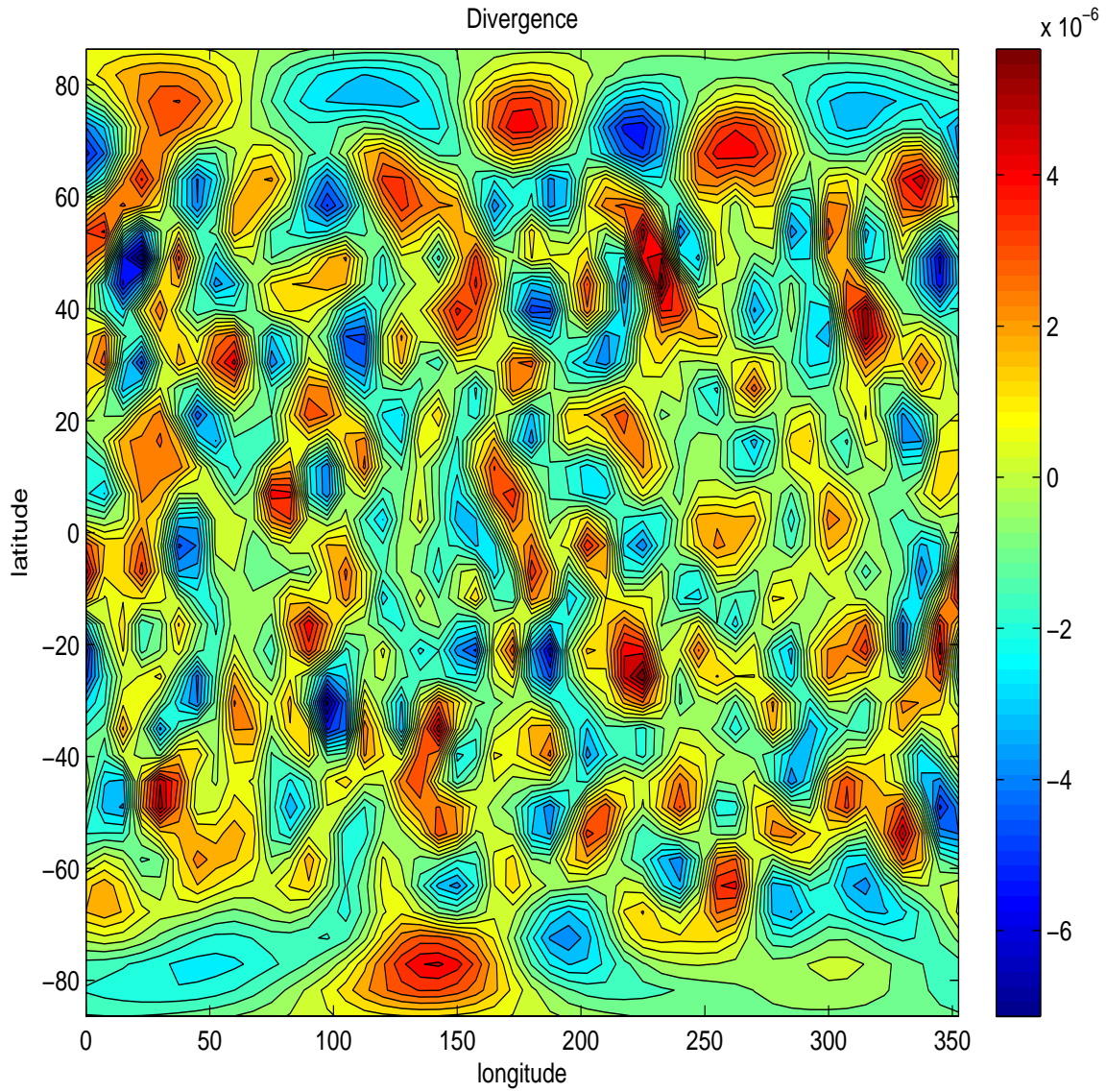
Figure 6. EnKF filter linear observation operator case: divergence field, 100 ensemble members and 1% random perturbation around the mean after 10 days of time integration
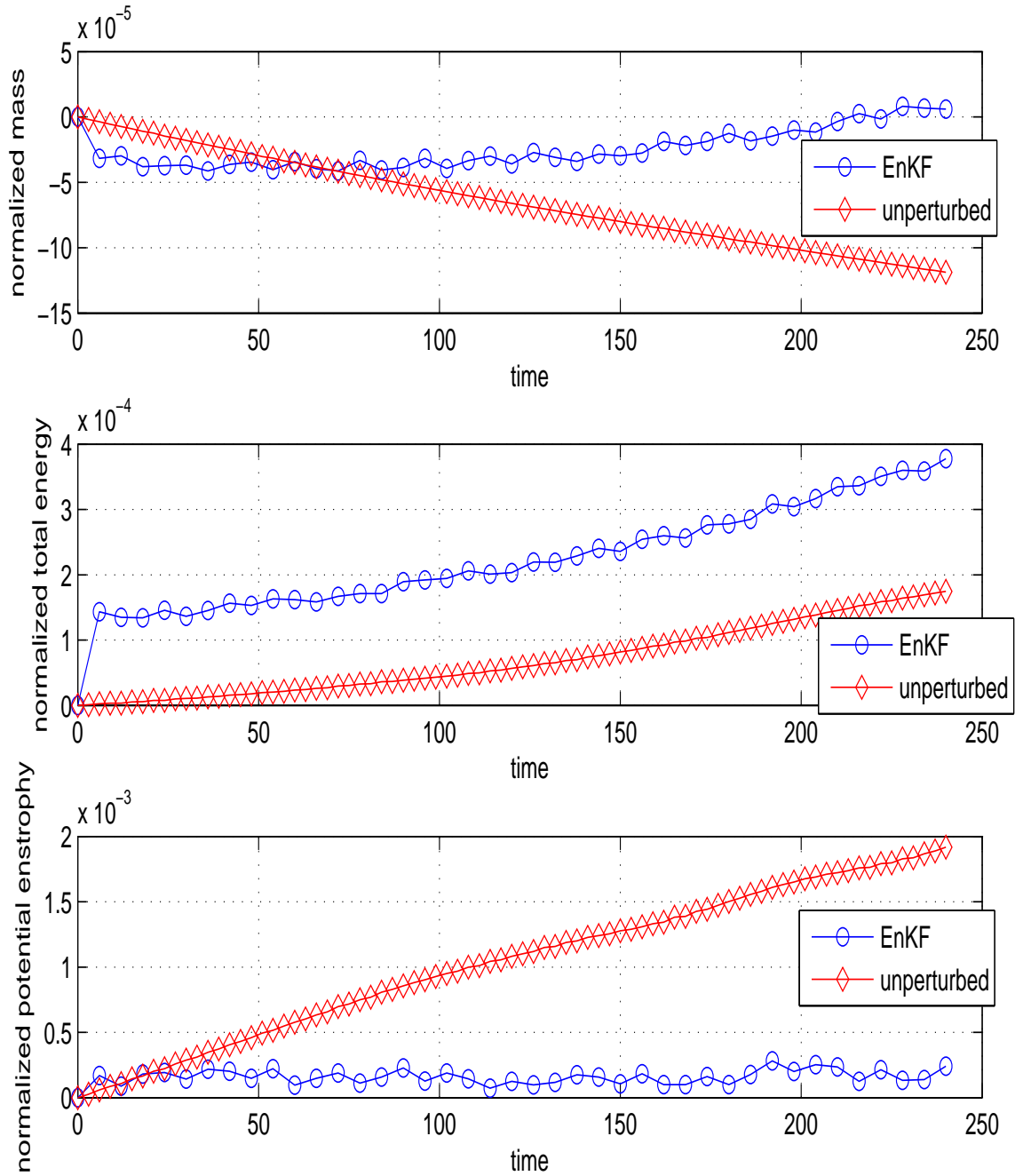
Figure 7.   Non-perturbed and EnKF filter with linear observation operator integral invariants: normalized mass, total energy and potential enstrophy, after 10 days of time integration, 100 ensemble members.
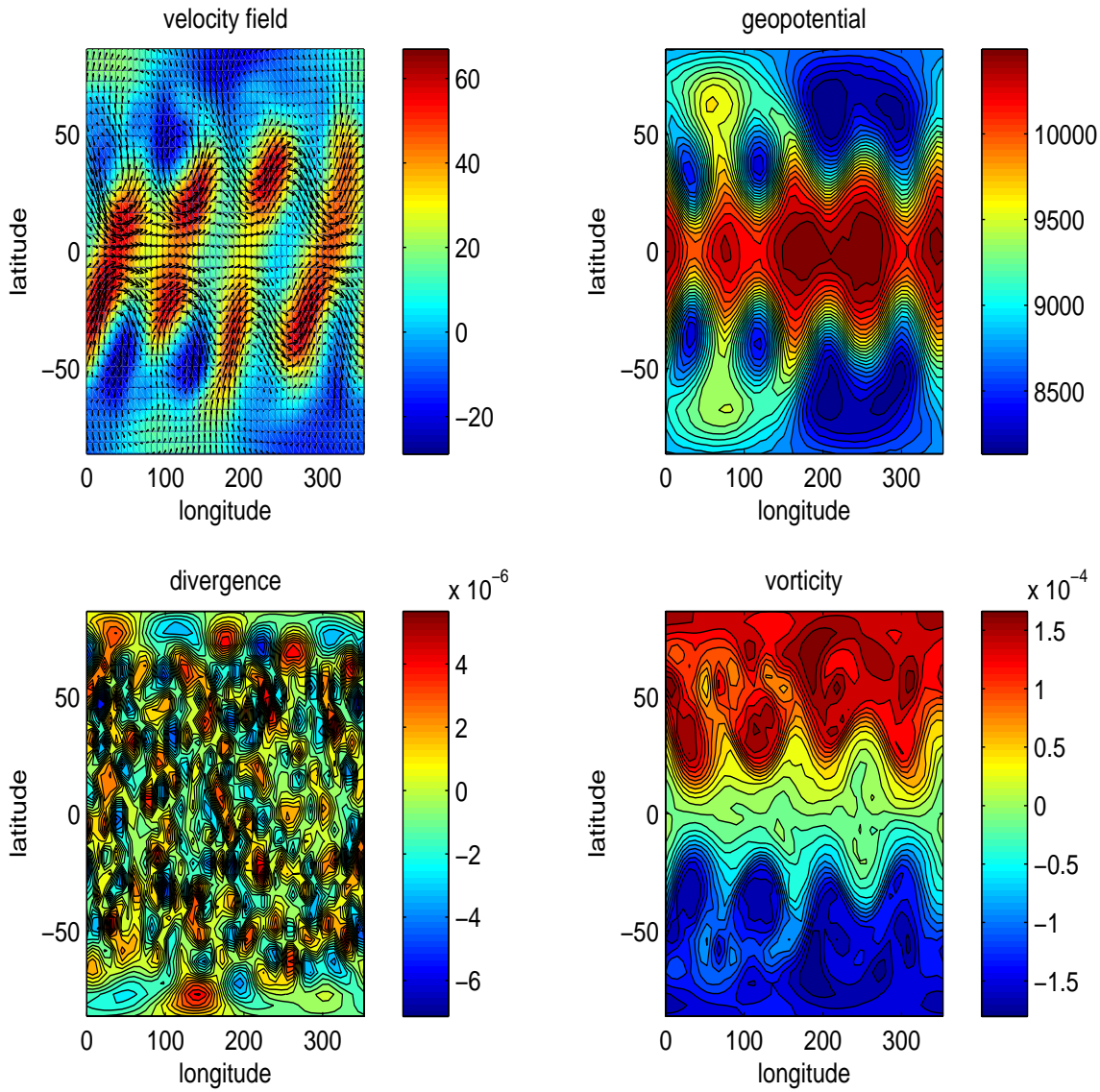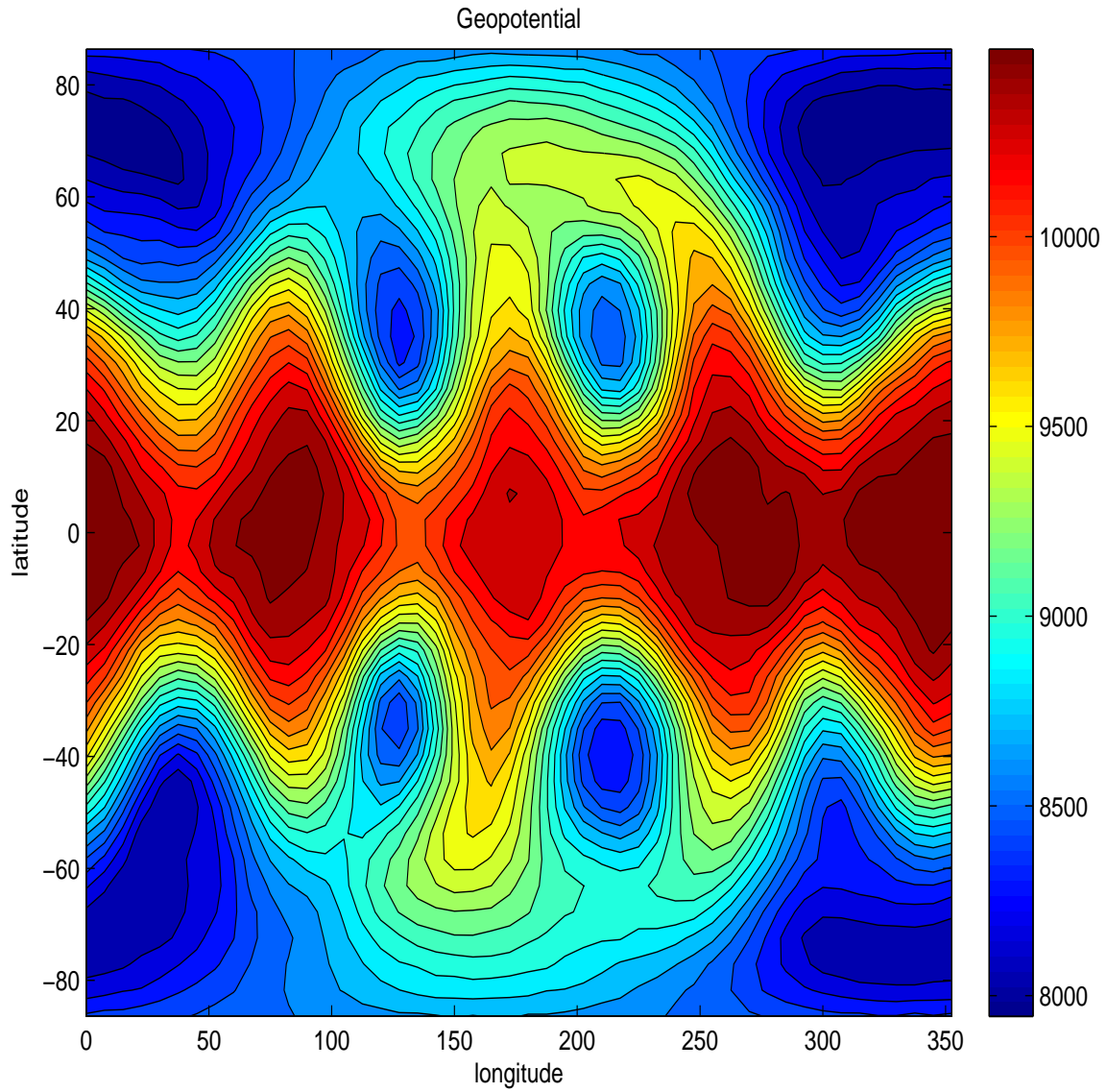
Figure 8. EnKF filter non-linear observation operator case $\mathcal{H}(u) = u^2$: velocity, geopotential, and divergence fields after 15 days of time integration. 100 ensemble members and 1% random perturbation around the mean
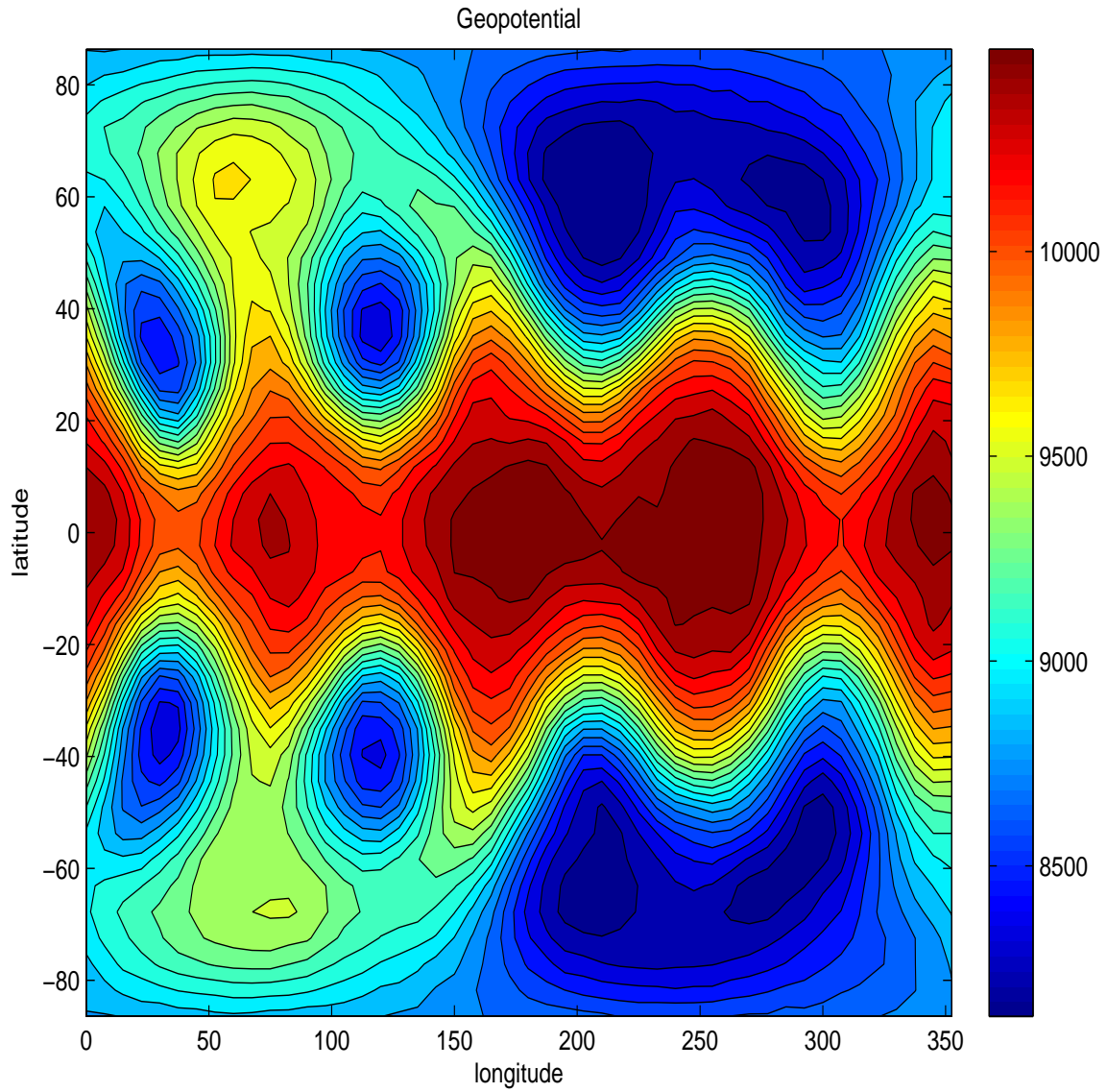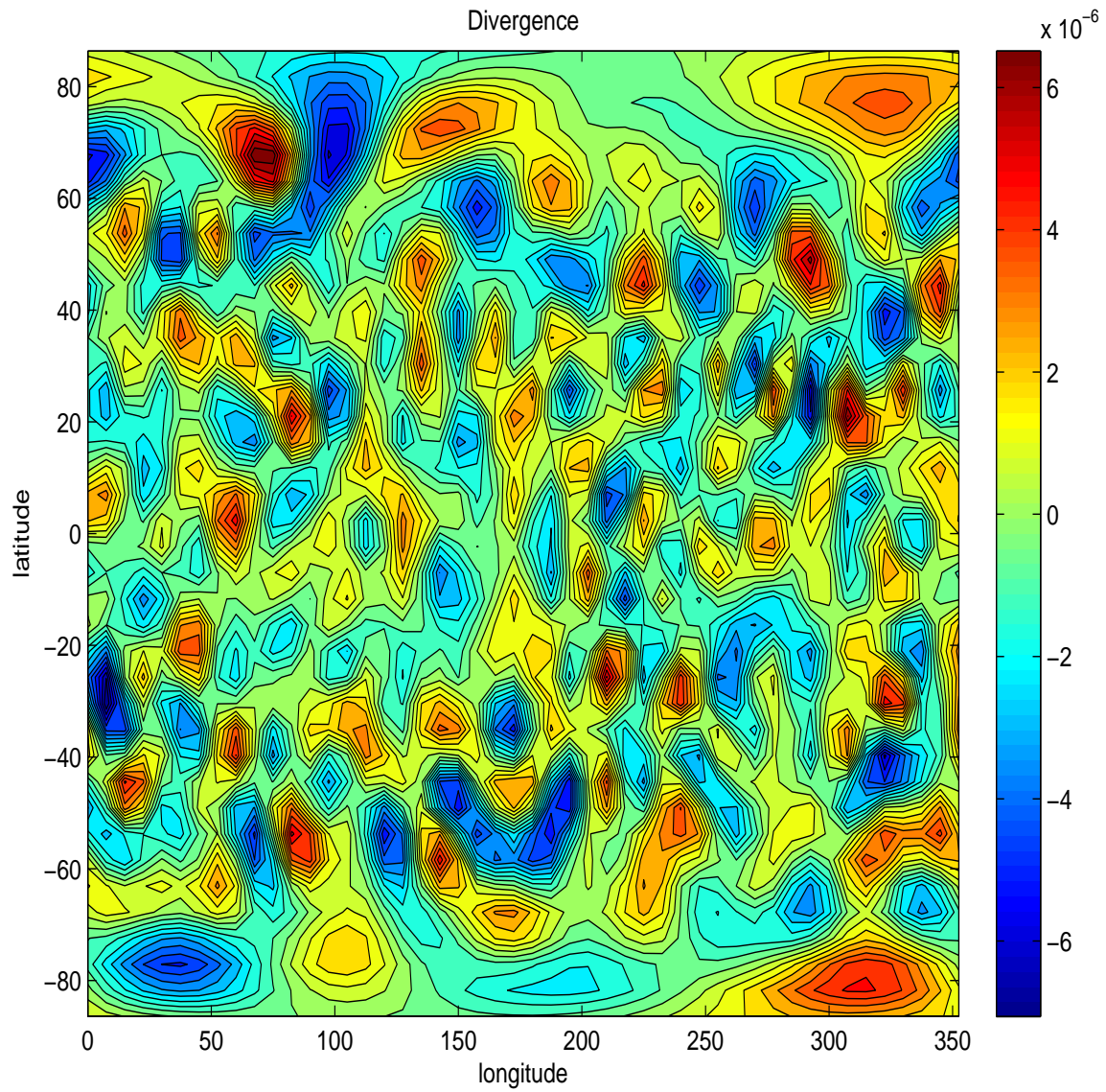
Figure 9. EnKF filter non-linear observation operator case $\mathcal{H}(u) = u^2$: geopotential field after 15 days of time integration
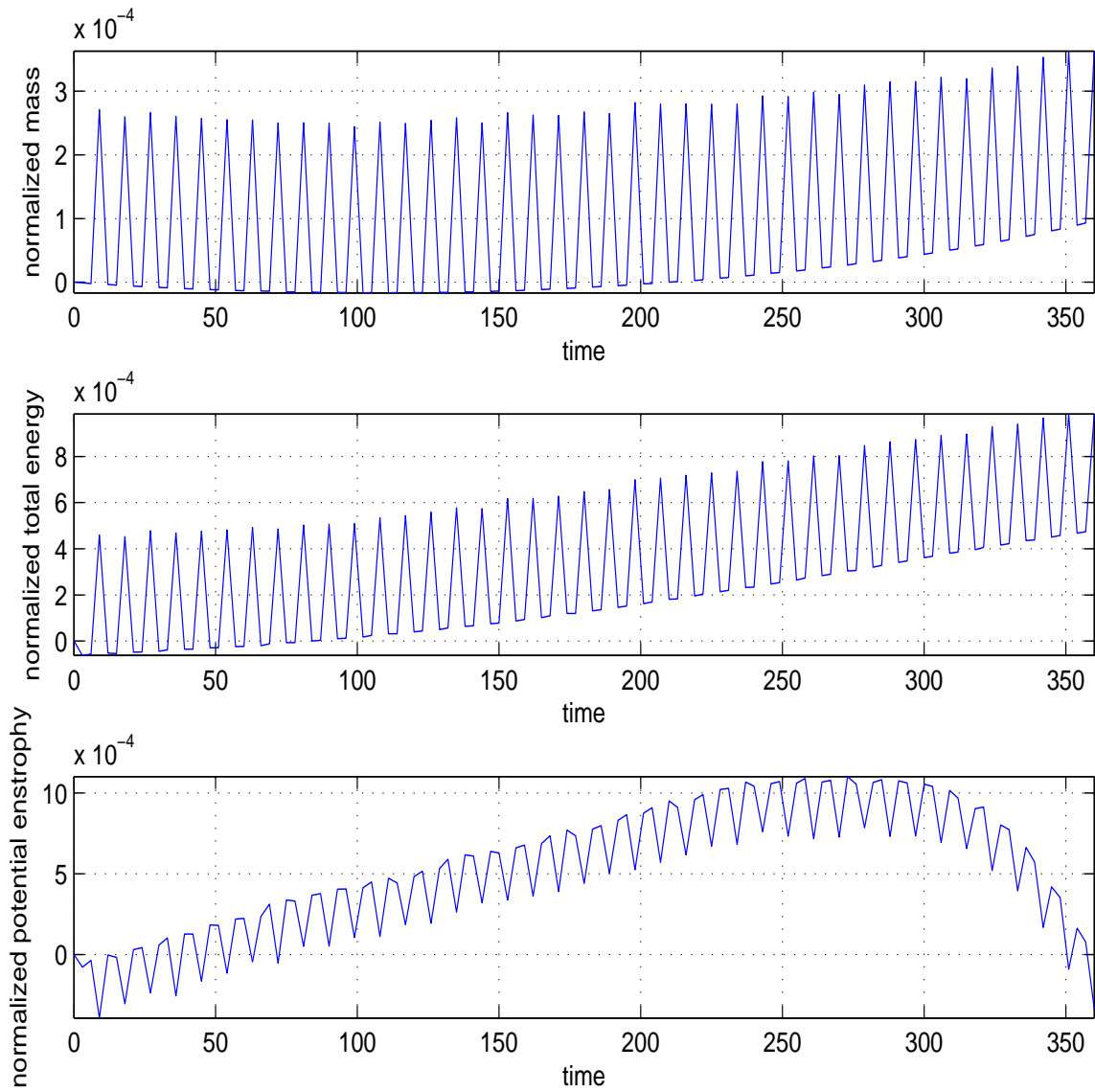
Figure 10. EnKF filter non-linear observation operator case $\mathcal{H}(u) = u^2$: divergence field after 15 days of time integration

Figure 11. EnKF filter non-linear observation operator case $\mathcal{H}(u) = u^4$: geopotential field after 15 days of time integration

Figure 12.   EnKF filter non-linear observation operator case $\mathcal{H}(u) = u^4$: divergence field after 15 days of time integration

Figure 13. EnKF filter non-linear observation operator case $\mathcal{H}(u) = u^4$: normalized mass, total energy and potential enstrophy after 15 days of time integration, 100 ensemble members.