Comparing ensemble data assimilation methods for the shallow water equations model

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. 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.

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 is novel.

The same model 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 reported in [Hamill et al. 2001], [Constan-

tinescu et al. 07A 07B] 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. We also assess numerically the latest argument of [Snyder et al. 08] that unless the ensemble size 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, a single member is assigned a posterior weight close to one, while all other members have vanishing small weights. In order to render algorithm computationally feasible a parallel MPI implementation is used.

The paper is structured as follows. Section 2 presents the Shallow-Water model 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 the conclusion.

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. Under this condition, conservation of mass implies that the vertical velocity of the fluid is small. It can be shown from the momentum equation that vertical pressure gradients are nearly hydrostatic, and that horizontal pressure gradients are due to the displacement of the pressure surface, implying that the velocity field is nearly constant throughout the depth

of the fluid. Taking the vertical velocity and variations throughout the depth of the fluid to be exactly zero in the Navier-Stokes equations, the shallow water equations are derived. Shallow water equations are especially suitable to model tides which have very large length scales, as well as to model Rossby and Kelvin waves in the atmosphere, rivers, lakes and oceans as well as gravity waves.

The shallow-water equations in spherical geometry are given by

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

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.

The Rossby-Haurwitz waves are steadily propagating solutions of the fully nonlinear nondivergent barotropic vorticity equation on a sphere [Haurwitz 40]. They are described exactly by analytic solutions and so they are useful test cases for numerical models. Although the shallow water equations do not have analogous analytic solution, a Rossby-Haurwitz wave initial condition is expected to evolve in a very similar way to that in the non divergent barotropic vorticity equation case. 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

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

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.

2.1. Solution method

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

$$\phi(\lambda,\mu) = \sum_{m=-M}^{M}\sum_{n=|m|}^{\mathfrak{N}(m)}\phi_{m,n}P_{m,n}(\mu)e^{im\lambda}, \tag{5}$$

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 (5) are determined by

$$\phi_{m,n} = \int_{-1}^{1}\frac{1}{2\pi}\int_{0}^{2\pi}\phi(\lambda,\mu)e^{-im\lambda}P_{m,n}(\mu)d\lambda d\mu \tag{6}$$

The inner integral represents a Fourier transform,

$$\phi_m(\mu) = \frac{1}{2\pi}\int_0^{2\pi} \phi(\lambda,\mu)e^{-im\lambda}d\lambda \tag{7}$$

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

$$\phi_{m,n} = \sum_{j=1}^{J} \phi_m(\mu)P_{m,n}(\mu)\omega_j, \tag{8}$$

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 pseudospectral 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 hyperbolic type of the shallow water equations, the centered leapfrog scheme

$$\frac{\phi_{m,n}^{k+1} - \phi_{m,n}^{k-1}}{2\Delta t} = \mathfrak{F}(\phi_{m,n}^{k}) \tag{9}$$

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 $k\Delta t$.

$$\bar{\phi}^k_{m,n} = \phi^k_{m,n} + \alpha \left[\phi^{k+1}_{m,n} - 2\phi^k_{m,n} + \bar{\phi}^{k-1}_{m,n}\right] \tag{10}$$

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 $1 - 4\alpha \sin^2(\frac{\nu\Delta t}{2})$.

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

$$\dot{\mathbf{x}}_t = f(t, \mathbf{x}_t, u_t, \mathbf{v}_t) \tag{11}$$

$$\mathbf{z}_t = h(t, \mathbf{x}_t, u_t, \mathbf{n}_t) \tag{12}$$

The equation (11) is the state equation or the system model, (12) 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

$$\mathbf{x}_k = f_k(\mathbf{x}_{k-1}, \mathbf{v}_{k-1}), \tag{13}$$

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

$$\mathbf{z}_k = h_k(\mathbf{x}_k, \mathbf{n}_k), \tag{14}$$

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 (13) to obtain the prior PDF of the state at time k via the Chapman-Kolmogorov equation

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1}) d\mathbf{x}_{k-1}. \tag{15}$$

The probabilistic model of the state evolution, $p(\mathbf{x}_k|\mathbf{x}_{k-1})$, is defined by the state equation (13) 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

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) = \frac{p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{1:k-1})}{p(\mathbf{z}_k|\mathbf{z}_{1:k-1})}, \tag{16}$$

where the normalizing constant

$$p(\mathbf{z}_k|\mathbf{z}_{1:k-1}) = \int p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{1:k-1})d\mathbf{x}_k. \tag{17}$$

depends on the likelihood function $p(\mathbf{z}_k|\mathbf{x}_k)$, defined by the measurement equation (14) and the known statistics of $\mathbf{n}_k$.

The relations (15) and (16) 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 and Berliner 07A]) 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. $\exp(\tau^2/2)$,

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 07A 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 $\{\mathbf{x}^i_{k-1}\}_{i=1,\cdots,N}$ 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, $\{q^1_k, \cdots, q^N_k\}$, from the distribution of $\mathbf{v}_k$. Use these to generate a new population of particles, $\{\mathbf{x}^1_{k|k-1}, \mathbf{x}^2_{k|k-1}, \cdots, \mathbf{x}^N_{k|k-1}\}$ via the equation

$$\mathbf{x}^i_{k|k-1} = f_k(\mathbf{x}^i_{k-1}, \mathbf{v}^i_k) \tag{18}$$

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

$$q^i_k = \frac{p(\mathbf{z}_k|\mathbf{x}^i_{k|k-1})}{\sum_{j=1}^N p(\mathbf{z}_k|\mathbf{x}^j_{k|k-1})} \tag{19}$$

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

• **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

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}, \mathbf{z}_k) \propto p(\mathbf{x}_k|\mathbf{z}_{1:k-1})p(\mathbf{z}_k|\mathbf{x}_k), \tag{20}$$

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:

$$p(\mathbf{x}_k|\mathbf{z}_k) = \sum_{i=1}^{N} q_i \delta(\mathbf{x}_k - \mathbf{x}_k{}^i) \tag{21}$$

is replaced by

$$\hat{p}(\mathbf{x}_k|\mathbf{z}_k) = \sum_{i=1}^{N} \frac{1}{N} \delta(\mathbf{x}_k - \mathbf{x}_k{}^\star) = \sum_{i=1}^{N} \frac{n_i}{N} \delta(\mathbf{x}_k - \mathbf{x}_k{}^i) \tag{22}$$

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.

3.2. The Ensemble Kalman Filter

The ensemble Kalman filter (EnKF) was first proposed by Evensen 28 [@] and further developed by [Burgers et al. 98] and [Evensen 03, Evensen07]. 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 (12), 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

$$p(\mathbf{x}|\mathbf{z}) \propto p(\mathbf{z}|\mathbf{x})p(\mathbf{x}). \tag{23}$$

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]. The implementation of the standard EnKF may be divided into three steps, as follows:

- **Setting and matching**

■ Define the ensemble

$$\mathcal{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_N] \tag{24}$$

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

$$\bar{\mathcal{X}} = \mathcal{X} \cdot \mathbf{1}_N, \tag{25}$$

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 $\mathcal{C}$

$$\mathcal{X}' = \mathcal{X} - \frac{1}{N}\bar{\mathcal{X}}, \tag{26}$$

$$\mathcal{C} = \frac{\mathcal{X}'\mathcal{X}'^T}{N-1}, \tag{27}$$

• **Sampling**

■ Generate

$$\mathcal{Z} = [\mathbf{z}_1, \cdots, \mathbf{z}_N] \tag{28}$$

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

$$\mathcal{R} = \frac{\mathcal{Z}\mathcal{Z}^t}{N-1}, \tag{29}$$

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

$$\mathcal{X}^p = \mathcal{X} + \mathcal{C}\mathcal{H}^T(\mathcal{H}\mathcal{C}\mathcal{H}^T + \mathcal{R})^{-1}(\mathcal{Z} - \mathcal{H}\mathcal{X}) \tag{30}$$

The matrix

$$\mathcal{K} = \mathcal{C}\mathcal{H}^T(\mathcal{H}\mathcal{C}\mathcal{H}^T + \mathcal{R})^{-1} \tag{31}$$

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

$$\bar{\mathcal{X}^p} = \mathcal{X}^p + \mathcal{K}\left[\mathbf{z} - (\mathcal{H}\mathcal{X}^p + \mathbf{d})\right], \tag{32}$$

and

$$\mathcal{C}^p = \mathcal{C} - \mathcal{K}\left[\mathcal{H}\mathcal{C}\mathcal{H}^T + \mathcal{R}\right]\mathcal{K}^T, \tag{33}$$

the vector $\mathbf{d}$ which appears in (32) stems from the affine measurement relation

$$h(\mathbf{x}) = \mathcal{H}\mathbf{x} + \mathbf{d}. \tag{34}$$

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).

$$\hat{\mathbf{x}} = \begin{pmatrix} \mathbf{x} \\ \mathcal{H}(\mathbf{x}) \end{pmatrix}. \tag{35}$$

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

$$\hat{\mathcal{H}}\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \mathbf{y} \tag{36}$$

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

$$\hat{\mathcal{K}} = [\rho \circ \mathcal{C}]\,\mathcal{H}^T(\mathcal{H}\,[\rho \circ \mathcal{C}]\,\mathcal{H}^T + \mathcal{R})^{-1} \tag{37}$$

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 horizontal localization [Gaspari and Cohn 99] constructed a Gaussian-shaped function that is actually a fourth-order piece-wise polynomial. [Houtekamer and Mitchell 98] and [Evensen 03] used a cut-off radius so that observations are not assimilated beyond a certain distance from the grid points.

3.3. The Maximum Likelihood Ensemble Filter

The Maximum Likelihood Ensemble Filter (MLEF) proposed by [Zupanski 05], [Zupanski and Zupanski 06] is a hybrid filter combining the 4-D Var 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 70]

$$P_f^k = \mathcal{M}_{k-1,k} P_a^{k-1} \mathcal{M}_{k-1,k}^T + Q_{k-1}, \tag{38}$$

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 (38) could be factorized and written as

$$P_f^k = \overbrace{\left(\mathcal{M}_{k-1,k}(P_a^{k-1})^{1/2}\right)}^{(P_f^{k-1})^{1/2}} \left(\mathcal{M}_{k-1,k}(P_a^{k-1})^{1/2}\right)^T + Q_{k-1} \tag{39}$$

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

$$(P_a^{k-1})^{1/2} = \begin{pmatrix} p_{(1,1)}^{k-1} p_{(2,1)}^{k-1} \cdots p_{(N,1)}^{k-1} \\ p_{(1,2)}^{k-1} p_{(2,2)}^{k-1} \cdots p_{(N,2)}^{k-1} \\ \\ p_{(1,n)}^{k-1} p_{(2,n)}^{k-1} \cdots p_{(N,n)}^{k-1} \end{pmatrix}, \tag{40}$$

as usual N is the number of ensemble members and n the number of state variables. Using equation(40), the square root forecast error covariance matrix $(P_f^{k-1})^{1/2}$ can then

be expressed as

$$(P_f^{k-1})^{1/2} = \begin{pmatrix} b_{(1,1)}^{k-1} b_{(2,1)}^{k-1} \cdots b_{(N,1)}^{k-1} \\ b_{(1,2)}^{k-1} b_{(2,2)}^{k-1} \cdots b_{(N,2)}^{k-1} \\ \\ b_{(1,n)}^{k-1} b_{(2,n)}^{k-1} \cdots b_{(N,n)}^{k-1} \end{pmatrix}, \tag{41}$$

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

$$\mathbf{b}_i^{k-1} = \begin{pmatrix} b_{(i,1)}^{k-1} \\ b_{(i,2)}^{k-1} \\ \vdots \\ b_{(i,n)}^{k-1} \end{pmatrix} = \mathcal{M}_{k-1,k} \begin{pmatrix} x_1^{k-1} + p_{(i,1)}^{k-1} \\ x_2^{k-1} + p_{(i,2)}^{k-1} \\ \vdots \\ x_n^{k-1} + p_{(i,n)}^{k-1} \end{pmatrix} - \mathcal{M}_{k-1,k} \begin{pmatrix} x_1^{k-1} \\ x_2^{k-1} \\ \vdots \\ x_n^{k-1} \end{pmatrix}. \tag{42}$$

The vector $\mathbf{x}^{k-1} = \begin{pmatrix} x_1^{k-1} \\ x_2^{k-1} \\ \vdots \\ x_n^{k-1} \end{pmatrix}$ is the analysis state from the previous assimilation cycle. which is found from the posterior analysis pdf as presented in [Lorenc 86].

• **Analyzing:** The analysis step for the MLEF involves solving a non-linear minimization problem. As in [Lorenc 86], the associated cost function is defined in terms of the forecast error covariance matrix and is given as

$$\mathcal{J}(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \left(P_f^k\right)^{-1} (\mathbf{x} - \mathbf{x}_b) + \frac{1}{2} \left[\mathbf{y} - h(\mathbf{x})\right]^T \mathcal{R}^{-1} \left[\mathbf{y} - h(\mathbf{x})\right] \tag{43}$$

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

$$\mathbf{x}_b = \mathcal{M}_{k-1,k}(\mathbf{x}^{k-1}) + Q_{k-1}. \tag{44}$$

Through a Hessian preconditioner we introduce the change of variable

$$(\mathbf{x} - \mathbf{x}_b) = (P_f^{k-1})^{1/2} (\mathcal{I} + \mathcal{C})^{-T/2} \xi \tag{45}$$

where ξ is vector of control variables, $\mathcal{C}$ is the Hessian matrix of $\mathcal{J}$ and $\mathcal{I}$ is the identity matrix.

The Hessian matrix $\mathcal{C}$ is provided by

$$\mathcal{C} = (P_f^{k-1})^{T/2}\mathcal{H}^T\mathcal{H}\mathcal{R}^{-1}\mathcal{H}(P_f^{k-1})^{T/2} = (\mathcal{R}^{-1/2}\mathcal{H}(P_f^{k-1})^{1/2})^T(\mathcal{R}^{-1/2}\mathcal{H}(P_f^{k-1})^{1/2}) \quad (46)$$

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

To overcome the difficulty of calculating $\mathcal{C}$, which is principally due to the non-linear observation operator, (P_f^{k-1}) is used.

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

$$\mathcal{Z} = \begin{pmatrix} z_{(1,1)}z_{(2,1)}\cdots z_{(N,1)} \\ z_{(1,2)}z_{(2,2)}\cdots z_{(N,2)} \\ \\ z_{(1,n)}z_{(2,n)}\cdots z_{(N,n)} \end{pmatrix}, \mathbf{z}_i = \begin{pmatrix} z_{(i,1)} \\ z_{(i,2)} \\ \vdots \\ z_{(i,n)} \end{pmatrix} = \mathcal{R}^{-1/2}\mathcal{H}\mathbf{b}_i^{k-1}, \quad (47)$$

using the approximation

$$\mathbf{z}_i \approx \mathcal{R}^{-1/2}\left[h(\mathbf{x}+\mathbf{b}_i^{k-1}) - h(\mathbf{x})\right], \quad (48)$$

the matrix $\mathcal{C}$ can then be approximated by

$$\mathcal{C} =< \mathcal{Z}\mathcal{Z}^T >= \begin{pmatrix} {z_1}^T z_1 {z_1}^T z_2 \cdots {z_1}^T z_N \\ {z_2}^T z_1 {z_2}^T z_2 \cdots {z_2}^T z_N \\ \\ {z_N}^T z_1 {z_N}^T z_2 \cdots {z_N}^T z_N \end{pmatrix} \quad (49)$$

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

$$(P_a^k)^{T/2} = (P_f^{k-1})^{T/2}(\mathcal{I} + \mathcal{C}(\mathbf{x}_{opt}))^{-T/2}, \quad (50)$$

where $\mathbf{x}_{opt}$ is approximately the minimum of the cost function $\mathcal{J}$ given by (43).

$$R^2(\tilde{u}, u^t) = \frac{\{N_u \sum_{i=1}^{N_u}(\tilde{u}_i u_i^t) - \sum_{i=1}^{N_u}\tilde{u}_i \sum_{i=1}^{N_u} u_i^t\}^2}{\{N_u \sum_{i=1}^{N_u}\tilde{u}_i^2 - (\sum_{i=1}^{N_u}\tilde{u}_i)^2\}\{N_u \sum_{i=1}^{N_u} u_i^{t2} - (\sum_{i=1}^{N_u} u_i^t)^2\}} \quad (51)$$

Acknowledgments. The research of Prof. I.M. Navon and Dr. M. Jardak is supported by the National Science Foundation (NSF), grant ATM-03727818. The authors also acknowledge the support by NASA Modeling, Analysis, and Prediction Program under Award NNG06GC67G. The first author is thankful to W.H.Hanya for the continuous help.

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" *manuscript,* (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**, 456–460 (2006) .

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 lter 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.,* accepted (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. Jardak, I. M. Navon and M. Zupanski, "Comparison of Sequential data assimilation methods for the Kuramoto-Sivashinsky equation," (Submitted to IJNMF)

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).

E. M. Constantinescu, A. Sandu, T. Chai and G. R. Carmichael, "Ensemble-based chemical data assimilation.I: General approach," *Q.J.R.Meteorol.Soc.,* **133** 1229-1243 (2007).

E. M. Constantinescu, A. Sandu, T. Chai and G. R. Carmichael, "Ensemble-based chemical data assimilation.II: Covariance localization," *Q.J.R.Meteorol.Soc.,* **133** 1245-1256 (2007).

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. Apte1 , C. K. R. T. Jones and A. M. Stuart, "A Bayesian approach to Lagrangian data assimilation," *Tellus A,***60**, 336–347 (2008)

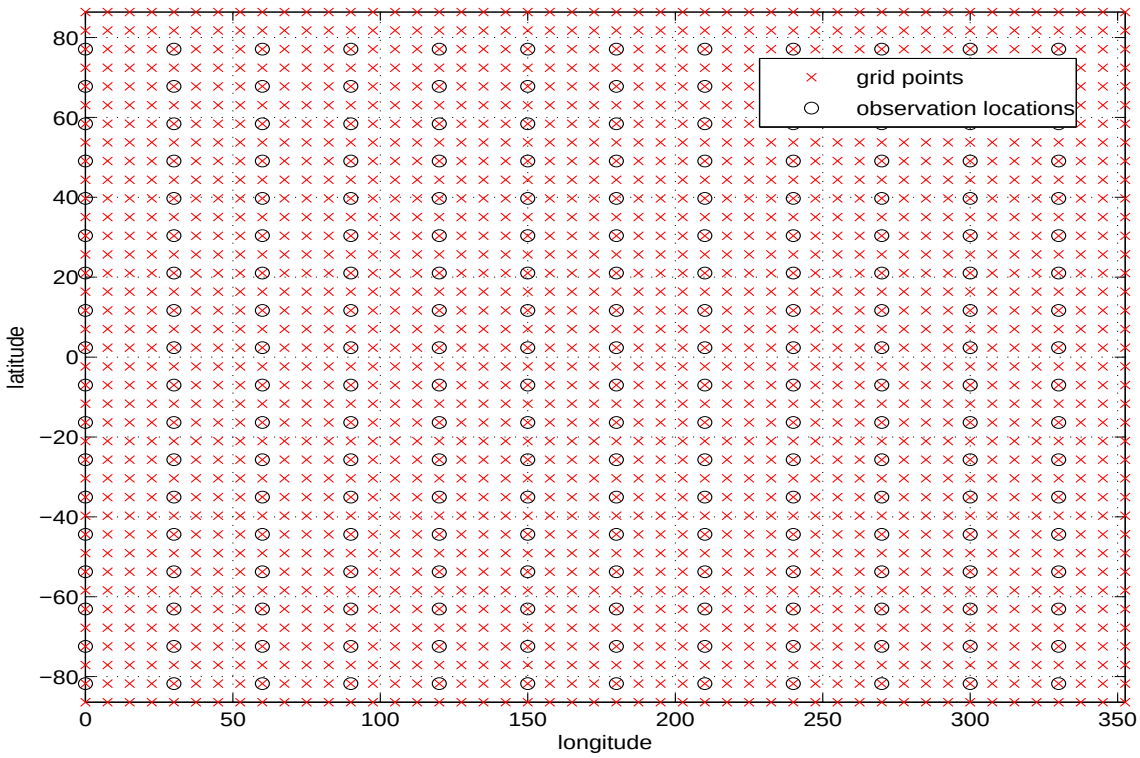

Figure 1. The locations of the grid and observation points

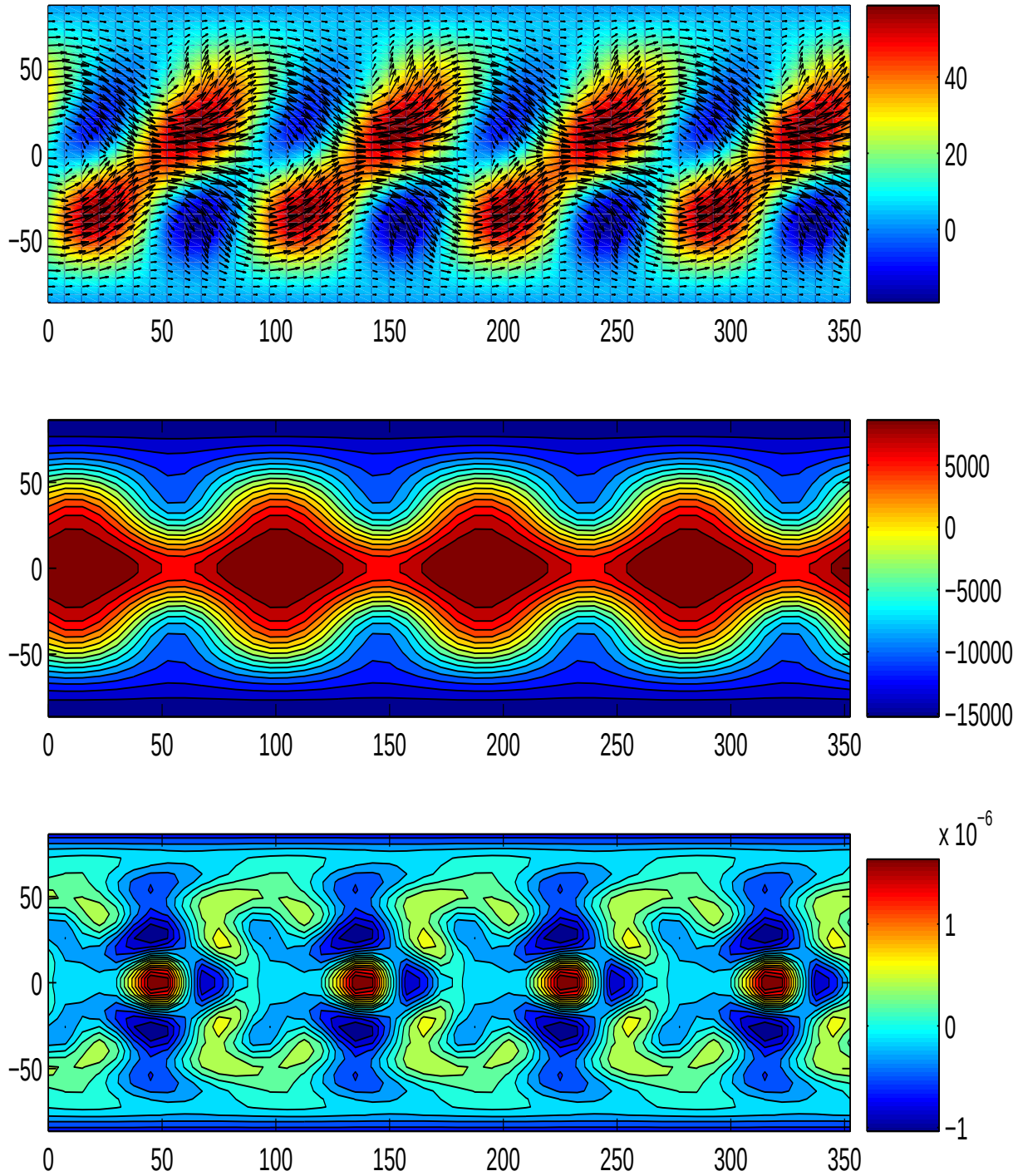

Figure 2. EnKF filter: velocity field, geopotential, and divergence after 1 day. 200 ensemble members and 1% perturbation

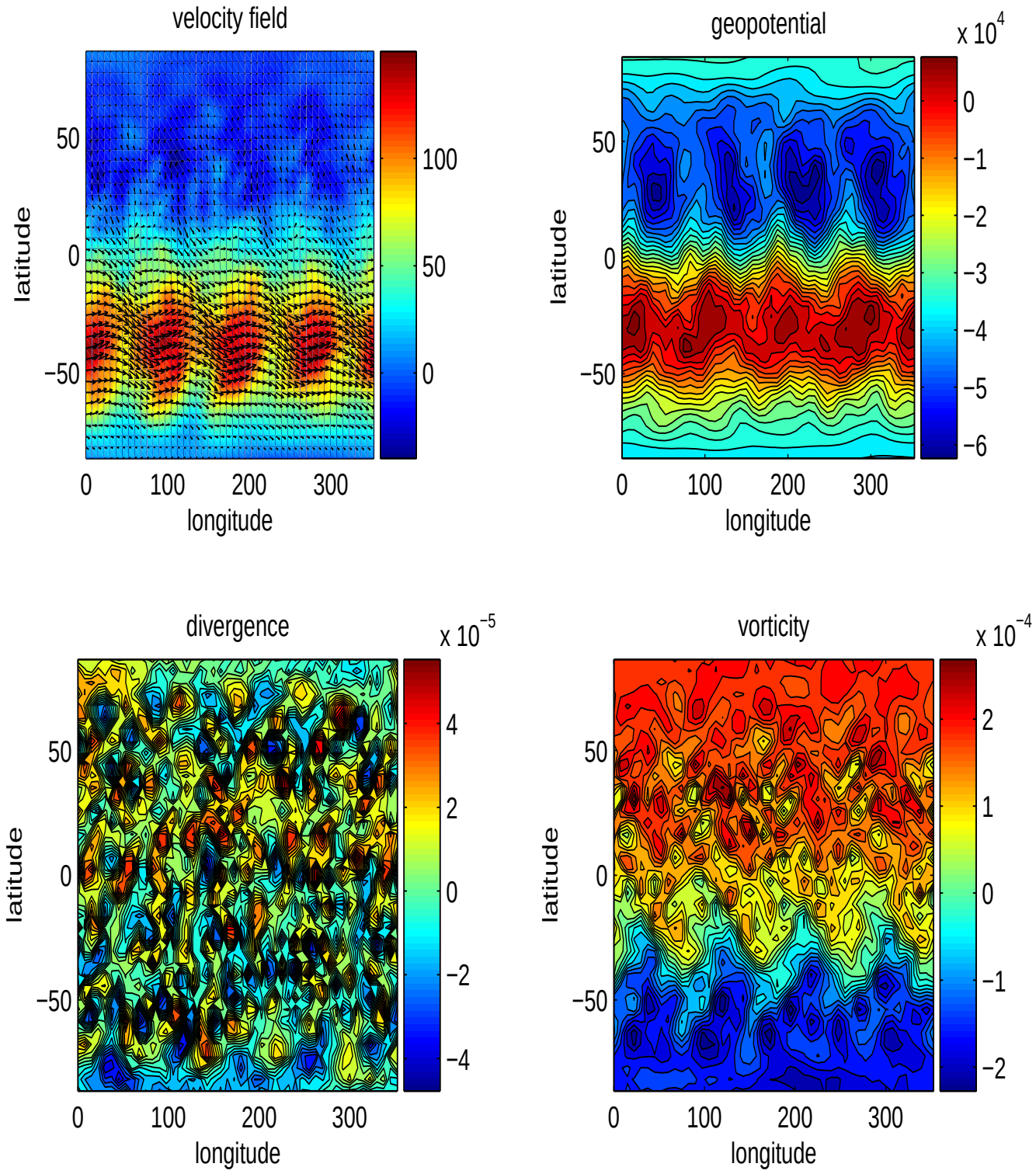

Figure 3. EnKF filter: velocity field, geopotential, divergence and vorticity after 1 day.

200 ensemble members and 10% perturbation

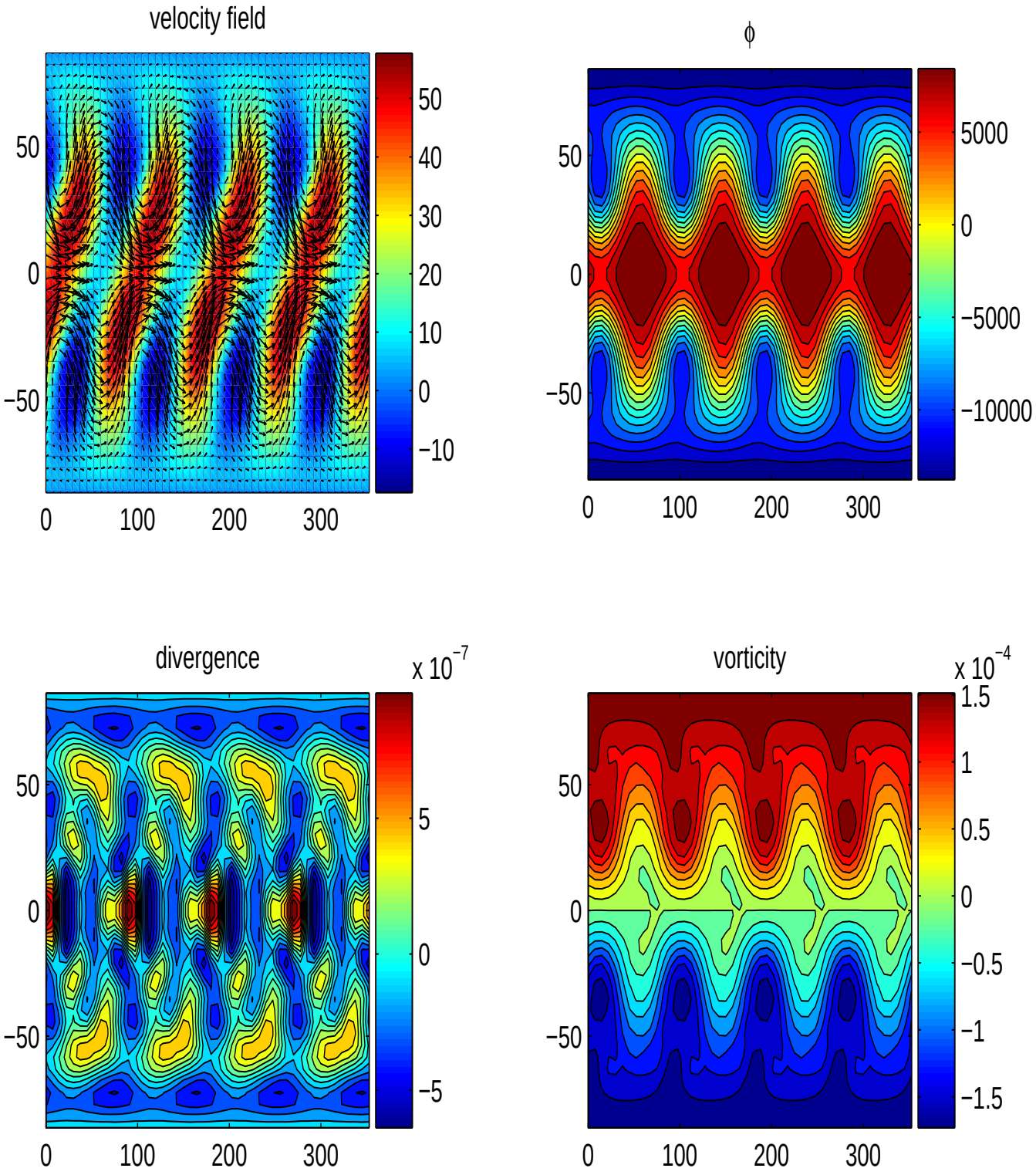

Figure 4. EnKF filter: velocity field, geopotential,divergence and vorticity after 5 days. 200 ensemble members and 1% perturbation

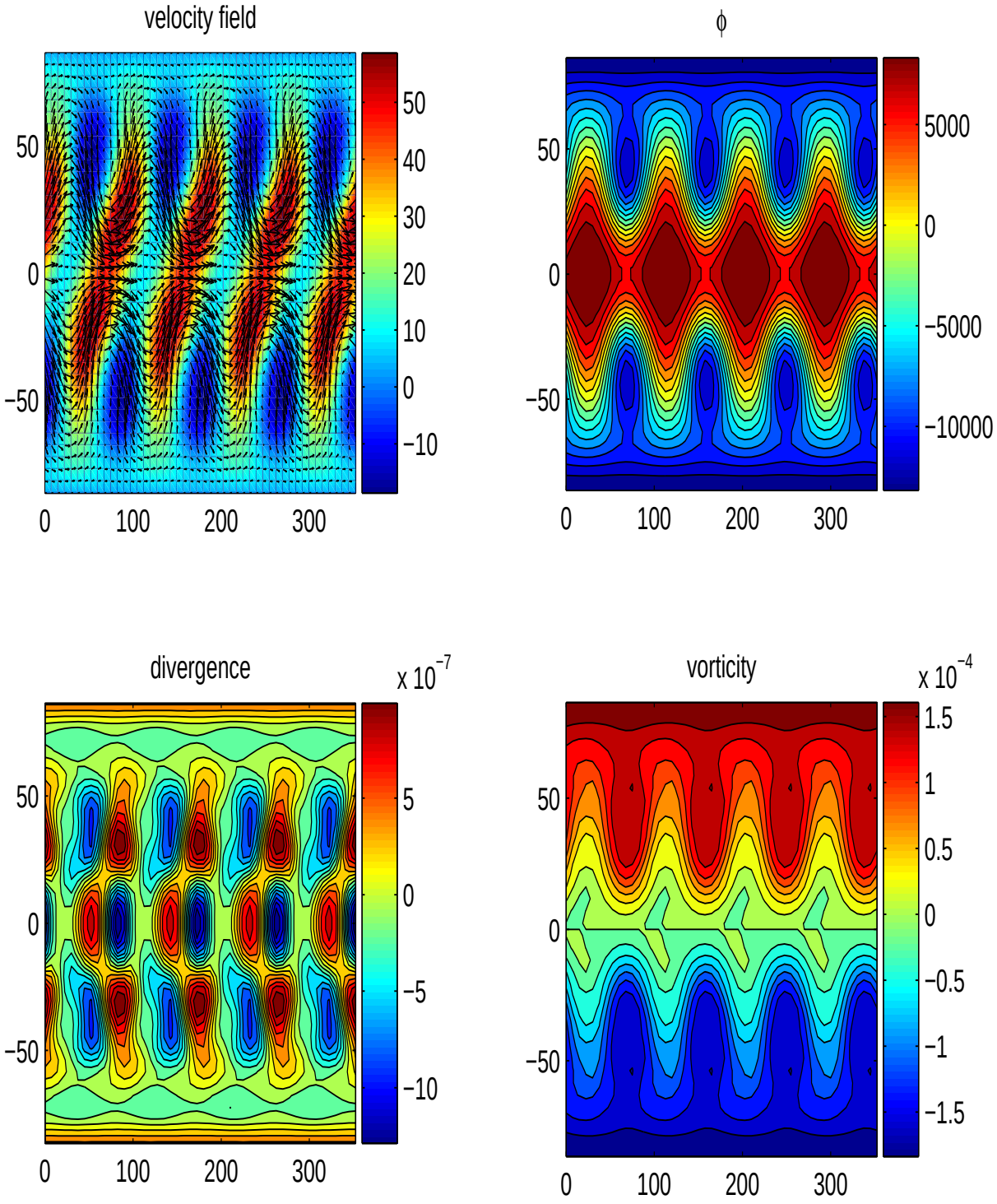

Figure 5. EnKF filter: velocity field, geopotential,divergence and vorticity after 10 days. 200 ensemble members and 1% perturbation

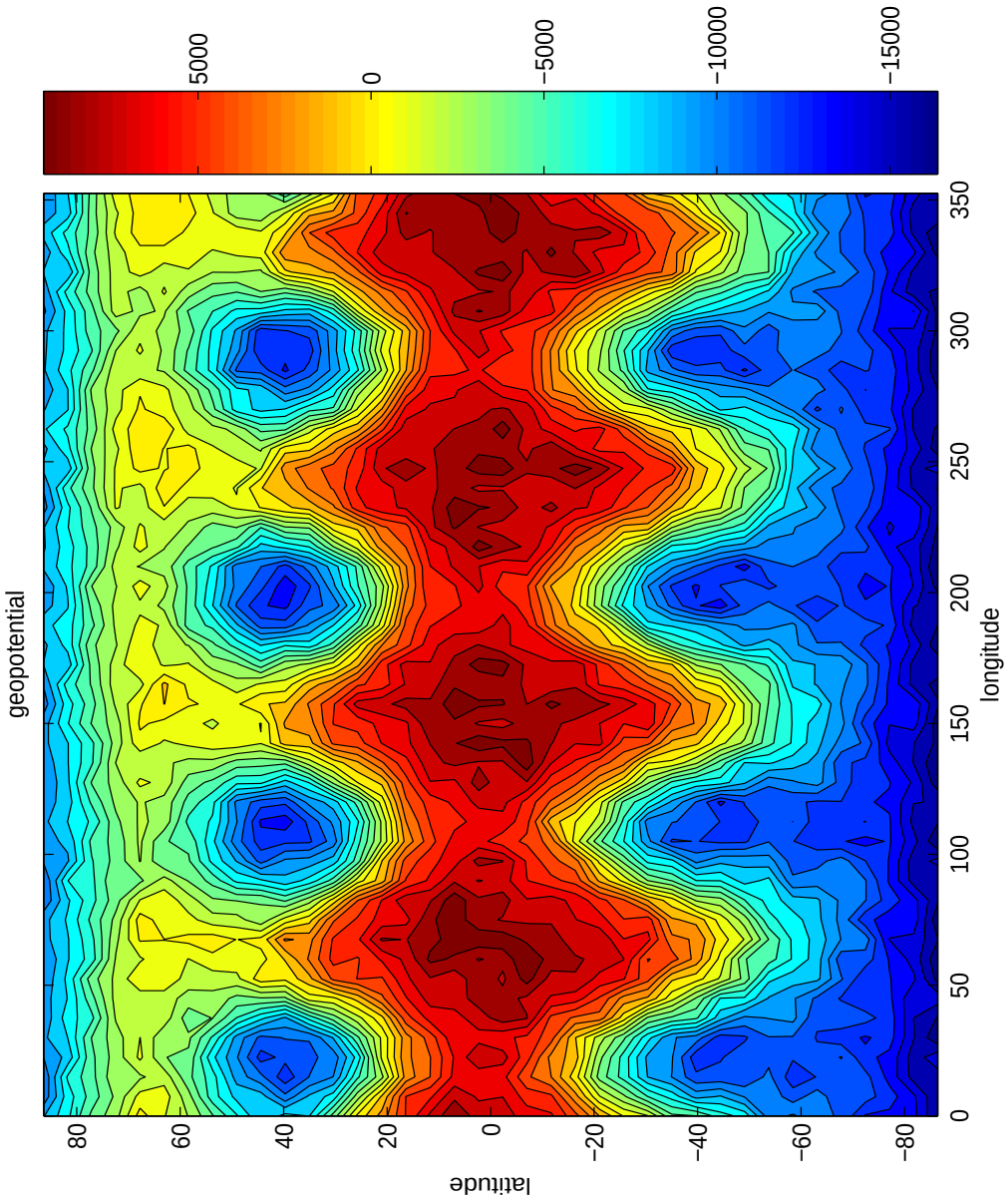

Figure 6. EnKF geopotential, linear observation operator case after 12 days

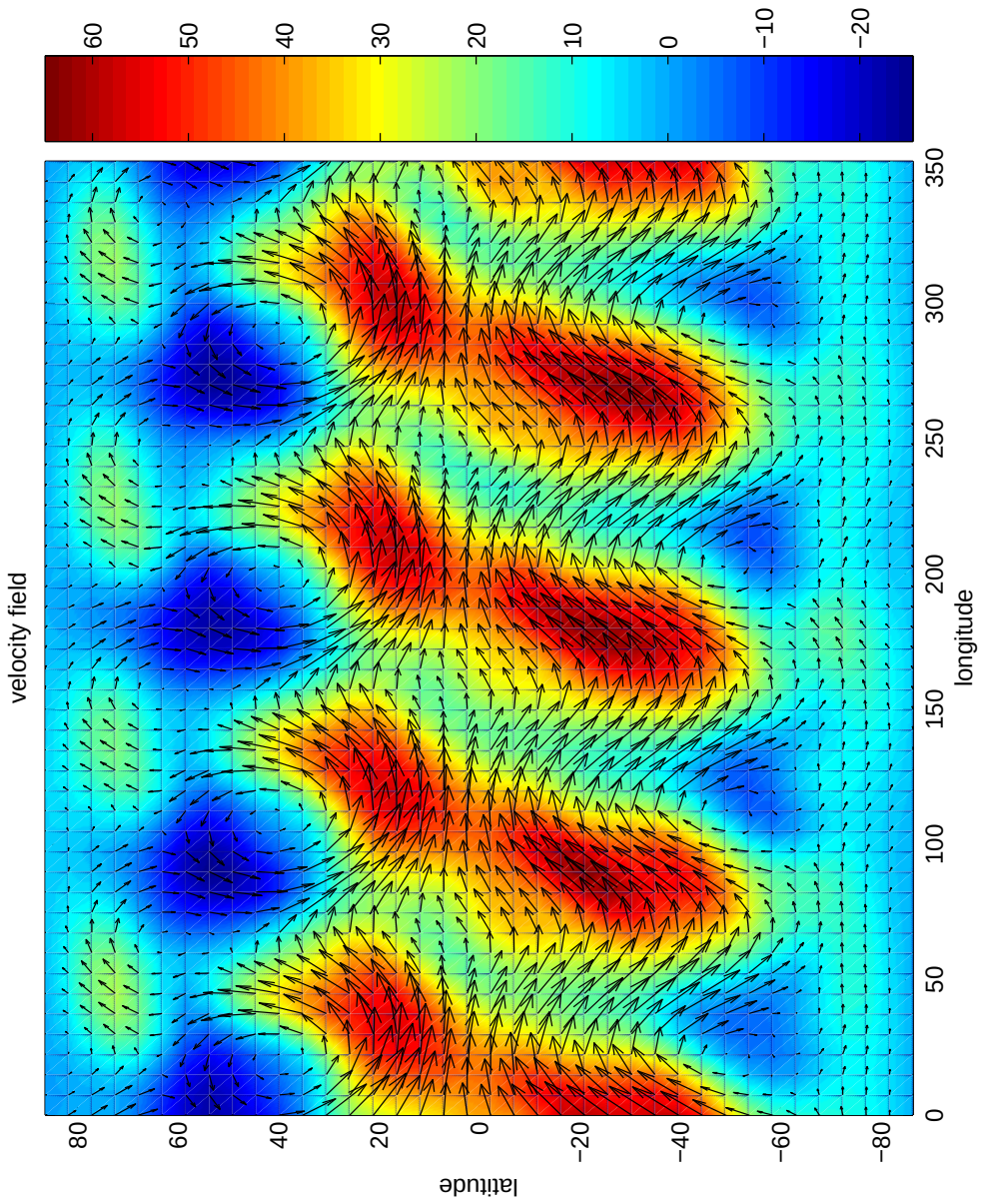

Figure 7. EnKF analysis velocity field, linear observation operator case after 12 days

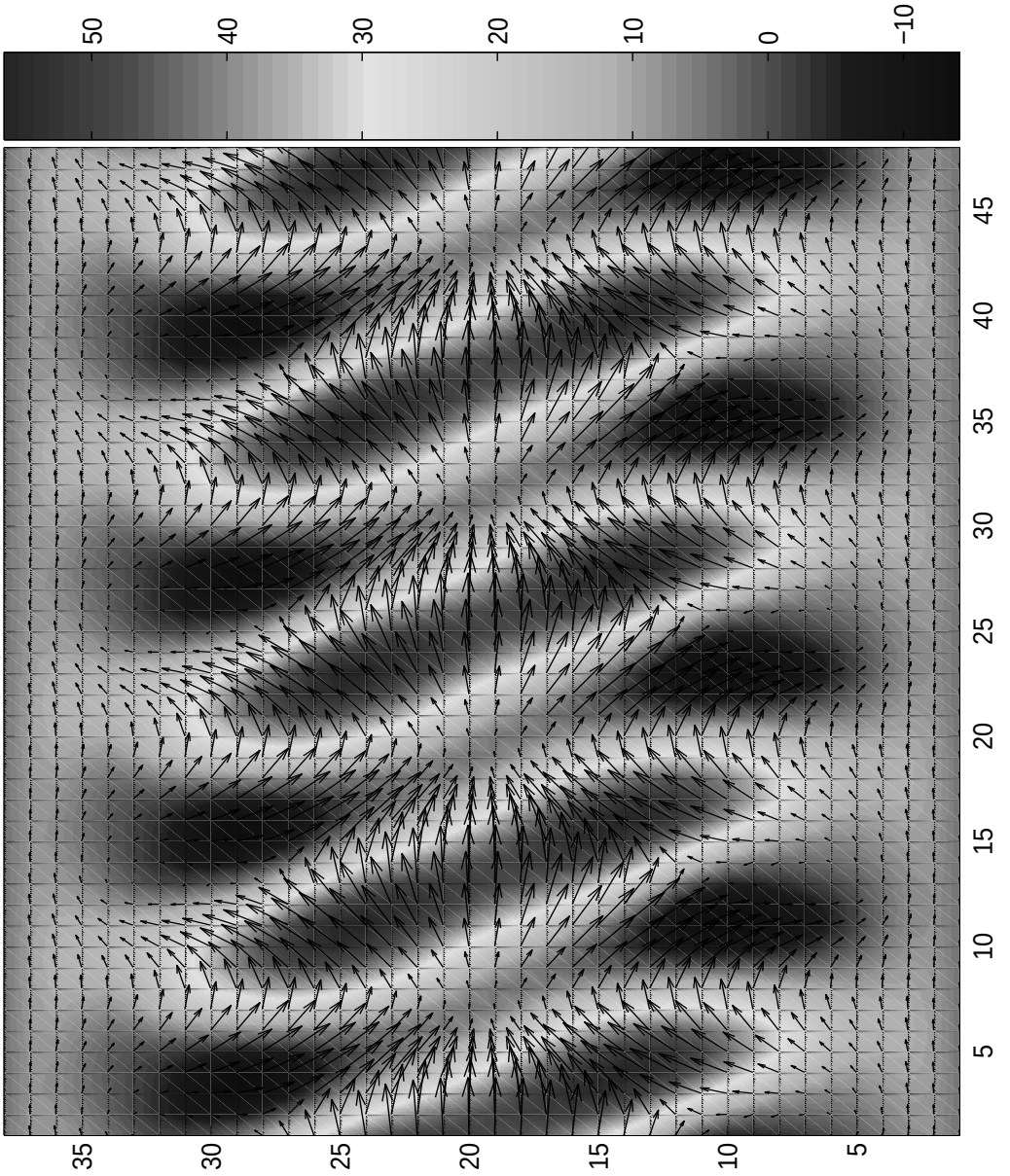

Figure 8. MLEF background velocity field, linear observation operator case after 5 days

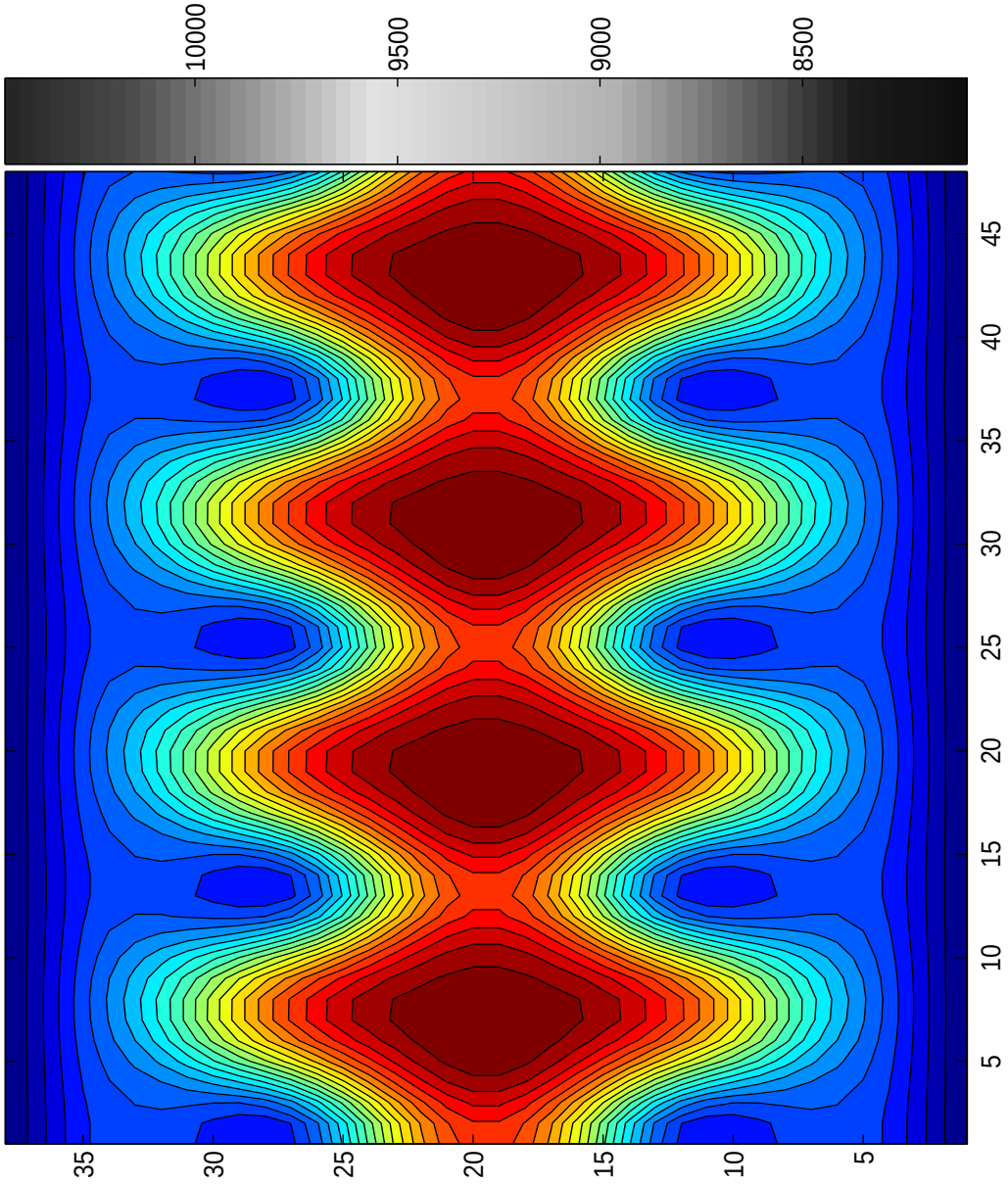

Figure 9. MLEF background geopotential, linear observation operator case after 5 days

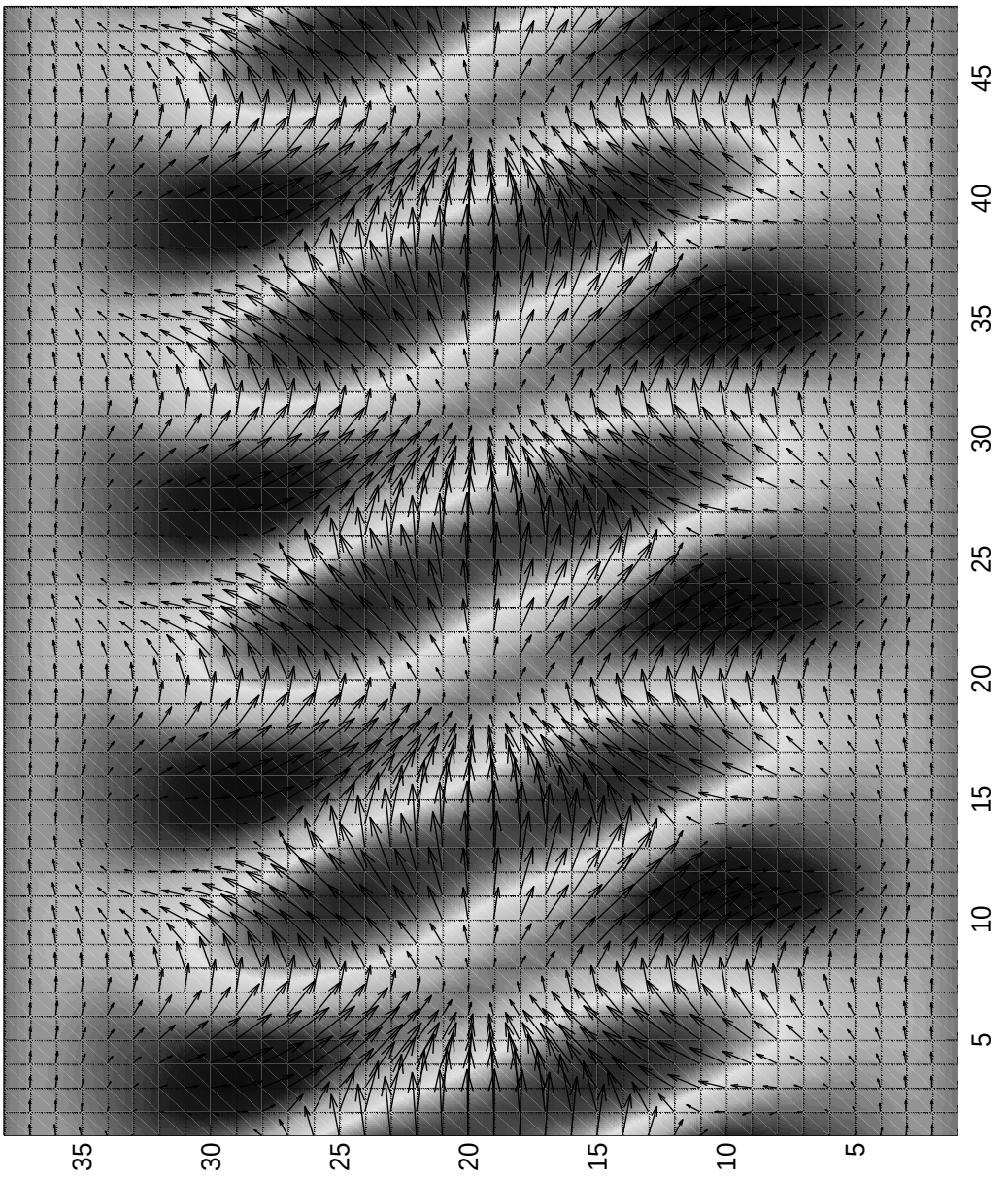

Figure 10. MLEF analysis velocity field, linear observation operator case after 5 days

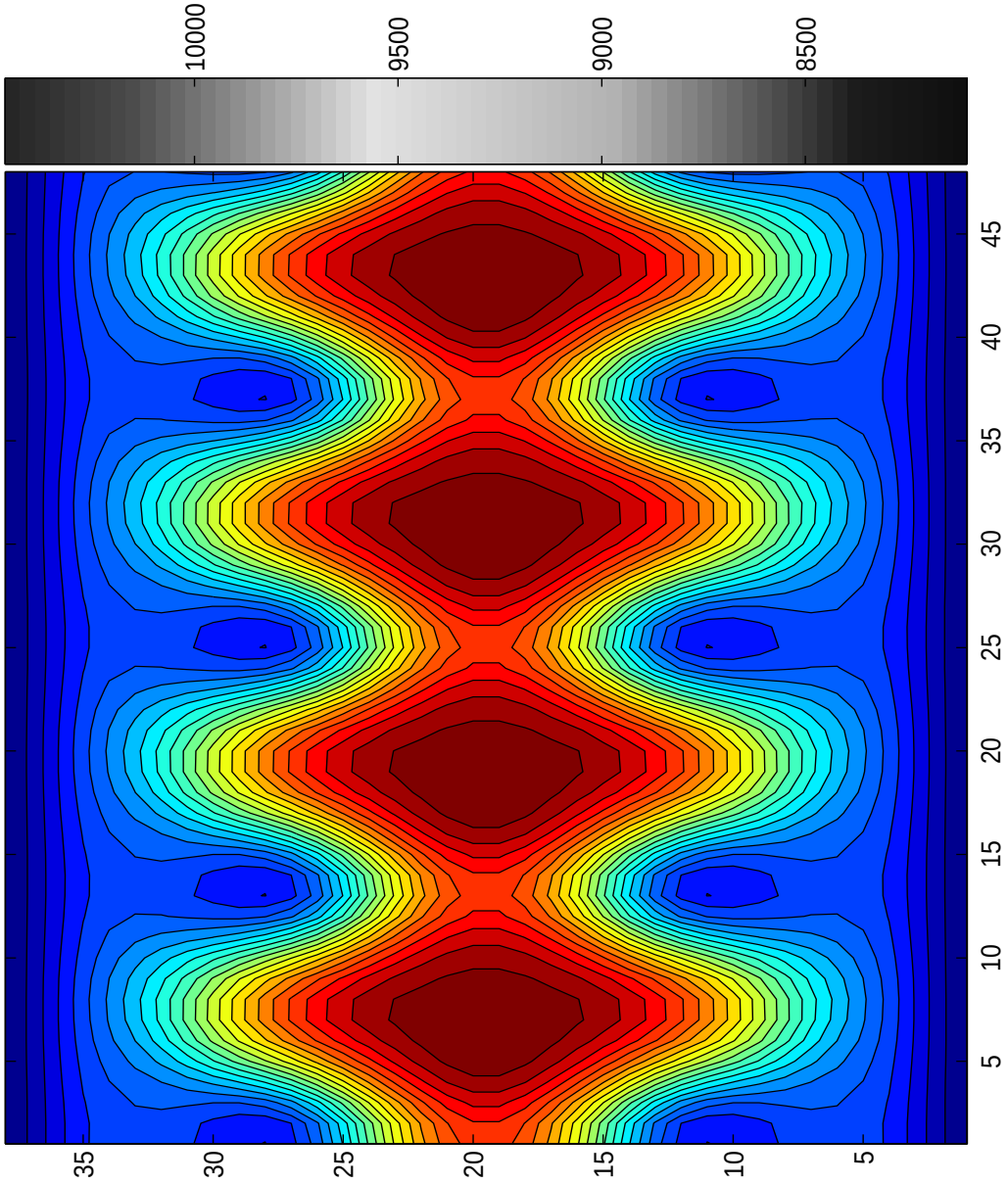

Figure 11. MLEF analysis geopotential, linear observation operator case after 5 days