JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 99, NO. C5, PAGES 10,143–10,162, MAY 15, 1994

Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics

Geir Evensen

Nansen Environmental and Remote Sensing Center, Bergen, Norway

Abstract. A new sequential data assimilation method is discussed. It is based on forecasting the error statistics using Monte Carlo methods, a better alternative than solving the traditional and computationally extremely demanding approximate error covariance equation used in the extended Kalman filter. The unbounded error growth found in the extended Kalman filter, which is caused by an overly simplified closure in the error covariance equation, is completely eliminated. Open boundaries can be handled as long as the ocean model is well posed. Well-known numerical instabilities associated with the error covariance equation are avoided because storage and evolution of the error covariance matrix itself are not needed. The results are also better than what is provided by the extended Kalman filter since there is no closure problem and the quality of the forecast error statistics therefore improves. The method should be feasible also for more sophisticated primitive equation models. The computational load for reasonable accuracy is only a fraction of what is required for the extended Kalman filter and is given by the storage of, say, 100 model states for an ensemble size of 100 and thus CPU requirements of the order of the cost of 100 model integrations. The proposed method can therefore be used with realistic nonlinear ocean models on large domains on existing computers, and it is also well suited for parallel computers and clusters of workstations where each processor integrates a few members of the ensemble.

Introduction

The implementation of the extended Kalman filter for data assimilation in a multilayer quasi-geostrophic (QG) model has previously been discussed by *Evensen* [1992, 1993]. The main result from *Evensen's* work [1992] is the finding of an apparent closure problem in the error covariance evolution equation. The extended Kalman filter applies a closure scheme where third- and higher-order moments in the error covariance evolution equation are discarded. This simple closure technique results in a unbounded error variance growth caused by the linearization performed when higher-order moments are neglected. However, a promising result from this work was that the error growth could be avoided by excluding a specific term in the transition matrix that caused the error growth, and this approach led to good results in a simple data assimilation experiment.

Evensen [1993] extended the work to include the possibilities of using open boundaries with in- and outflow with the extended Kalman filter. It was pointed out that even if open boundaries can be used without much difficulty in the pure ocean model, they significantly complicate the treatment of the error evolution equation, where approximate methods must be used. These approximations, although consistent, lead to some serious difficulties in conserving the definiteness of the error covariance matrix during long time integrations. This work showed that even if open boundaries could be handled, they resulted in rather unstable numerical algorithms.

The important conclusions from these two works are that a sequential data assimilation algorithm gives good results in a data assimilation scheme for the nonlinear QG model and that the results improve significantly according to improvements in the error estimate for the model forecast.

The updating scheme in the Kalman filter requires that the error covariance matrices for the model forecast and the measurement vector be known every time measurements are available. Results from *Evensen* [1992, 1993] and the additional fact that the Kalman filter is extremely expensive to compute, even for modest problem sizes, motivate the search for new methods for error covariance evolution or estimation. Such methods should include both the effect of internal error growth caused by the unstable dynamics in nonlinear ocean circulation models and the external error growth associated with the imperfection of the numerical ocean model.

The connection between stochastic dynamic prediction and the error covariance evolution used in the extended Kalman filter is discussed on the basis of the general theory of error evolution and prediction. The approach of stochastic dynamic prediction was first proposed by *Epstein* [1969], and several papers have later extended this theory, mainly in connection with simple spectral models in meteorology; see for example, *Gleeson* [1970], *Fleming* [1971a, b], *Epstein and Pitcher* [1972], *Leith* [1971, 1974], and *Pitcher* [1977].

Paper number 94JC00572.
0148-0227/94/94JC-00572$05.00

These papers discuss both the use of Monte Carlo methods and the approximate stochastic dynamic prediction for finding approximate solutions of the equation for the probability density function. For an application in oceanography, see *Salmon et al.* [1976]. Two more recent applications in meteorology are given by *Seidman* [1981], who examined the predictability of a general circulation model, and *Schubert and Suarez* [1989] who discussed the application of Monte Carlo methods for error prediction in an atmospheric primitive equation model. Different alternative Monte Carlo methods suitable for stochastic dynamic prediction were discussed by *Hoffman and Kalnay* [1983] and *Schubert et al.* [1992].

Here, after an introduction to stochastic dynamic prediction, the approximate equation for the error covariance evolution that is used in the extended Kalman filter and the problem of using a proper closure for this equation are briefly discussed.

It is then shown that the Monte Carlo method can be used as an alternative to the approximate error covariance evolution equation used in the extended Kalman filter to provide the forecast error estimate with a significantly lower computational cost and without any closure problem or even the problem with open boundaries.

Next, a procedure for applying ensemble statistics in the Kalman filter updating scheme for data assimilation is outlined. Finally, the proposed data assimilation method is examined in a twin experiment.

Equations for the Layered Model

The ocean model is the multilayered and nonlinear quasi-geostrophic model on an f plane [see *Pedlosky*, 1987] and has been applied and further discussed by *Haugan et al.* [1991] and *Evensen* [1992, 1993]. It describes conservation of potential vorticity ζ_l in each layer on an f plane. The mean layer thicknesses are D_l, and the density in each layer is ρ_l, where l denotes layer number; $l = 1$ in the upper layer. Ψ_l is the stream function in layer l. The horizontal length scale R_d is the internal Rossby radius of deformation of the upper layer, given by $R_\mathrm{d}^2 = [(\rho_2 - \rho_1)gD_1] \,/\, [\rho_0 f^2]$, where g is gravitational acceleration, ρ_0 is averaged density, and f is the Coriolis parameter. The characteristic horizontal velocity is U, yielding a time scale $T = R_\mathrm{d}/U$. The pressure scale is $\rho_0 f U R_\mathrm{d}$, and the stream function scale is UR_d. The nondimensional quasi-geostrophic equations are

$$\left(\frac{\partial}{\partial t} + u_l\frac{\partial}{\partial x} + v_l\frac{\partial}{\partial y}\right)\zeta_l = 0, \quad l = 1, n_z, \tag{1}$$

where n_z is the number of layers, and the velocities are the geostrophic approximations

$$u_l = -\frac{\partial \Psi_l}{\partial y} \qquad v_l = \frac{\partial \Psi_l}{\partial x}. \tag{2}$$

The vorticity in each layer is given by

$$\zeta_1 = \nabla^2\Psi_1 + fr_{1,2}(\Psi_2 - \Psi_1), \tag{3a}$$

$$\zeta_l = \nabla^2\Psi_l - fr_{l,1}(\Psi_l - \Psi_{l-1}) + fr_{l,2}(\Psi_{l+1} - \Psi_l), \tag{3b}$$

$$\zeta_{n_z} = \nabla^2\Psi_{n_z} - fr_{n_z,1}(\Psi_{n_z} - \Psi_{n_z-1}) + \eta, \tag{3c}$$

where l denote the intermediate layers $l = 2, \ldots, n_z - 1$. When ζ is known, this is a set of coupled Helmholtz equations for the stream function Ψ. The horizontal Laplacian operator is $\nabla^2 = \partial^2/\partial^2 x + \partial^2/\partial^2 y$, and the constants $fr_{l,1}$ and $fr_{l,2}$ are "nondimensional Froude numbers":

$$fr_{l,1} = \frac{D_1}{D_l}\frac{\rho_2 - \rho_1}{\rho_l - \rho_{l-1}}, \qquad fr_{l,2} = \frac{D_1}{D_l}\frac{\rho_2 - \rho_1}{\rho_{l+1} - \rho_l}. \tag{4}$$

The bottom topography term is

$$\eta = \varepsilon^{-1}\frac{h_\mathrm{b}}{D_{n_z}}, \tag{5}$$

with ε as the Rossby number and h as bottom topography. Note that when discretized in space, the QG model can be written as

$$\frac{\partial \boldsymbol{\psi}}{\partial t} = \mathbf{g}(\boldsymbol{\psi}, t), \tag{6}$$

where $\boldsymbol{\psi}$ is the state vector, and $\mathbf{g}$ is a nonlinear vector function.

Theory of Stochastic Dynamic Prediction

Ocean models are integrated forward in time from a specified initial state. The initial condition will normally be a judicious best guess or an estimate interpolated from measurements of the real ocean. It is recognized that the estimated initial state contains errors. Both the sparseness of the measurements, whereby all the scales of the physical system are not resolved, and the interpolation method cause the estimate to depart from the true state. Further, the estimate will be affected by errors in the measurements.

The choice of another interpolation scheme or just different statistical parameters in the interpolation scheme will produce another initial state resulting in a different forecast even if the same deterministic model is applied. It is not possible to say that the forecast based on any of the interpolated initial conditions is right or wrong or better or worse, since each initial state estimate represents an individual member of an infinite ensemble of possible states that are consistent with the data.

In his classic paper, *Epstein* [1969, p. 740], wrote that; "The different analyses will yield different forecasts, even if each were submitted to the same forecast procedure. If there is no way of determining which, if any, analysis is right, and since none is known to be wrong, there is no way of knowing, in any instance, which to believe."

The state of the ocean at a particular time t is represented by a vector of state variables $\boldsymbol{\psi}(t) \in \Re^n$, where the components in the state vector are the values of all dependent variables in the model. Based on the representation used, the components may be values of fields of physical variables such as velocities and density over an array of space mesh points or the amplitudes of an orthogonal function expansion of these physical fields.

The state vector at a specified time $\boldsymbol{\psi}(t)$ can be represented by a single point in an n-dimensional phase space $\mathcal{D}$.

Thus time evolution of the state vector $\psi(t)$ is described by continuous motion of the point along a trajectory in phase space.

The uncertainty in the initial state vector can be represented by a large ensemble of possible initial states, each assigned an individual probability number. Such a cloud of phase points surrounding the analyzed initial estimate $\psi(t = 0)$ (represented by the big dot) is illustrated in Figure 1. Suppose there are N points altogether, where N is a very large number, and dN is their density (points per volume increment) at any location. As the number of such phase points approach infinity, one can define a probability density distribution function

$$\phi(\psi) = \frac{dN}{N}, \tag{7}$$

which can vary throughout the space. The expression

$$\phi(\psi)d\psi \tag{8}$$

is defined as the probability of a phase of the system being located at a certain instant inside the n-dimensional volume element $d\psi$ located around the point ψ. The probability density ϕ must be defined over all of the phase space, and $\phi \geq 0$ for all ψ and t. As the total probability of finding the system in some arbitrary phase is equal to unity, one has

$$\int \phi(\psi)d\psi = 1, \tag{9}$$

where the integration is performed over all of the phase space.

Given a probability density function $\phi(\psi)$, the expected value of a quantity $h(\psi)$ is defined as

$$\overline{h(\psi)} = \int h(\psi)\phi(\psi)d\psi. \tag{10}$$

The following moments can then be defined:

$$\mu_i = \overline{\psi_i} \tag{11}$$

$$P_{ij} = \overline{(\psi_i - \mu_i)(\psi_j - \mu_j)} \tag{12}$$

$$\Theta_{ijk} = \overline{(\psi_i - \mu_i)(\psi_j - \mu_j)(\psi_k - \mu_k)} \tag{13}$$

$$\Gamma_{ijkl} = \overline{(\psi_i - \mu_i)(\psi_j - \mu_j)(\psi_k - \mu_k)(\psi_l - \mu_l)} \tag{14}$$

where μ_i is the mean, P_{ij} is the variance ($i = j$) and covariance ($i \neq j$), and Θ_{ijk} and Γ_{ijkl} are third- and fourth-order moments, respectively. Higher-order moments can also be defined.

It is usually assumed that the probability function for the initial state $\phi(\psi, t = 0)$ is a normal distribution with a specified mean $\mu(t = 0)$ and covariance $P_{ij}(t = 0)$. Note that for the initial state, the mean $\mu(t = 0)$ coincides with the initial estimate $\psi(t = 0)$. The variance of the distribution determines the uncertainty of the initial state, and the probability decreases when one moves away from the mean $\mu(t = 0)$.

In this formalism the time evolution of different ocean states is described by continuous motion of the respective points through the phase space. This is a completely deterministic process, in which a given initial state generates a definite phase path into the future. The time evolution of an ensemble of phase points is illustrated in Figure 2.

During the time integration the central forecast (the single forecast generated from the analyzed initial estimate), may drift away from the forecasted mean state (most probable state). Note also that the ensemble forecast may deform and expand or contract in size. The predicted errors are determined by the variance of the forecasted ensemble.

The next issue concerns the time evolution of the probability density which contains all the statistical information about the ensemble. The ensemble of possible states moves through the phase space governed by dynamic laws. The density of the ensemble has an associated probability, so that the dynamic equations can be used to generate probability predictions as well as predictions of the physical state itself. Each of the initial states in a volume element $d\psi$ will evolve

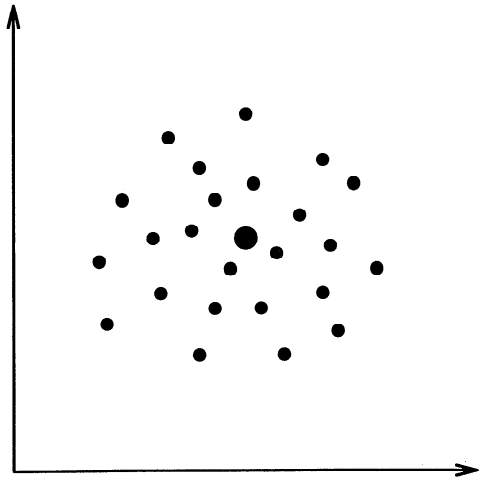

Figure 1. An ensemble of possible initial states of the ocean model can be represented in an n-dimensional phase space as a cloud of points, where each point represents an individual state.

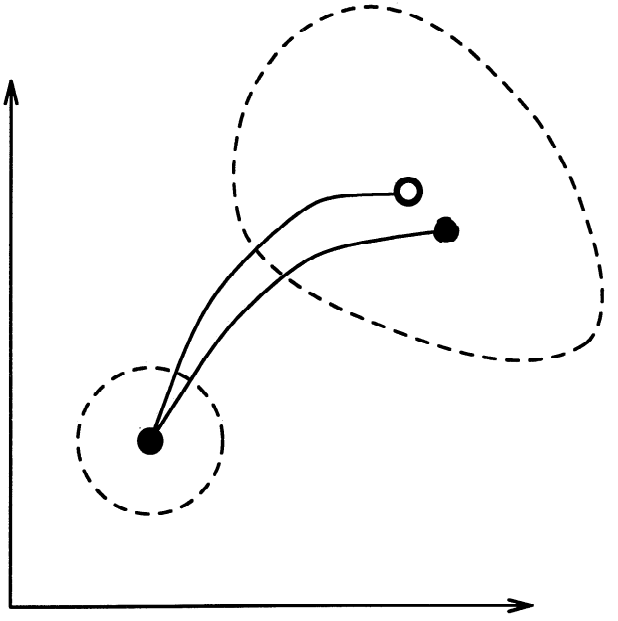

Figure 2. The time evolution of an ensemble of initial states in phase space is illustrated.

according to the same deterministic model, and no members of the phase points in $d\boldsymbol{\psi}$ may be created or lost. This is simply the conservation law of probability and may be expressed as

$$\frac{\partial \phi}{\partial t} + \sum_{i=1}^{n} \frac{\partial \dot{\psi}_i \phi}{\partial \psi_i} = 0, \tag{15}$$

where $\dot{\psi}_i = \partial \psi_i / \partial t$ is the "velocity" component in the ψ_i direction as determined by model equation (6). This equation states that the local change of probability density with time in phase space $\partial \phi / \partial t$ balances the "probability flux" represented by the remaining terms in the summation. Note the similarity to the continuity equation for mass density in hydrodynamics. Given appropriate boundary conditions for ϕ (e.g., $\phi \rightarrow 0$ as $\psi_i \rightarrow \pm\infty$) and an initial probability density $\phi(\boldsymbol{\psi}, t = 0)$, this equation can be integrated to obtain $\phi(\boldsymbol{\psi}, t)$ for $t \geq 0$. However, the integration by direct numerical methods is impractical because of the size of the problem. The continuity equation for the probability density ϕ provides a natural criterion for error growth/decrease, and it can be written as

$$\frac{\partial \phi}{\partial t} + \sum_{i=1}^{n} \dot{\psi}_i \frac{\partial \phi}{\partial \psi_i} = -\phi \sum_{i=1}^{n} \frac{\partial \dot{\psi}_i}{\partial \psi_i}. \tag{16}$$

The right-hand side of this equation is a divergence term which describes either increase or decrease in the probability density or, alternatively, contraction or expansion of the cloud of phase points. With the right-hand side equal to zero, (16) reduces to the Liouville's theorem for conservative systems in statistical mechanics (e.g., *Delcroix* [1968]).

If $d\phi/dt$, the change of ϕ following a point in the ensemble, is positive (negative), the ensemble is converging (diverging) and the errors are decreasing (increasing). This process can be illustrated using a simple example with a one-dimensional advection and diffusion equation. Consider the linear advection-diffusion equation discretized in space:

$$\frac{\partial \psi_i}{\partial t} = -\frac{\psi_{i+1} - \psi_{i-1}}{2\Delta x} + \frac{\psi_{i+1} - 2\psi_i + \psi_{i-1}}{\Delta x^2}. \tag{17}$$

Equation (16) then becomes

$$\frac{\partial \phi}{\partial t} + \sum_{i=1}^{n} \dot{\psi}_i \frac{\partial \phi}{\partial \psi_i} = \frac{2n\phi}{\Delta x^2} \geq 0 \tag{18}$$

The advection term in (17) has no contribution on the right-hand side in (16). This represents a property of conservative systems where the probability density is conserved with the motion in phase space [see *Gleeson*, 1970]. The diffusion term causes the system to lose energy; that is, the variance is decreasing. This leads to increasing probability in (18), which is equivalent to increasing predictability.

The divergence term on the right-hand side is zero for a conservative system like the QG model [see *Salmon et al.*, 1976]. Equation (16) then becomes Liouville's theorem, which states that the density of the ensemble is conserved with the motion along a path in phase space. Consider the volume element $d\mathcal{D}$ in phase space at time t, containing a cloud of points representing states with specific probabilities. At time $t + \delta t$, the points fill another volume element $d\mathcal{D}'$ of the phase space. Liouville's theorem then states that $d\mathcal{D} = d\mathcal{D}'$, and noting that no points are lost or gained from the volume during the integration, the probability of finding the system in $d\mathcal{D}'$ is equal to the probability of finding it in $d\mathcal{D}$. The total predictability of the state of the ocean then neither improves nor deteriorates in time. Note that the volume is still freely deformable, diverging in one dimension and converging in another, and the variance of the ensemble may change at different positions even though the energy for each of the ensemble members is conserved.

The formulation and methodology in the field of inverse methods and data assimilation traditionally assume that the dynamical model contains errors. These errors are assumed to be random white processes with a specified variance. When including the model noise in the ocean model (6), it becomes

$$d\boldsymbol{\psi} = \mathbf{g}(\boldsymbol{\psi}, t)dt + d\mathbf{b}, \tag{19}$$

where $d\mathbf{b} \in \Re^n$ is a vector of random white noise with mean zero. This equation is an Itô stochastic differential equation describing a Markov process. The evolution of the probability density for this equation is described by the forward Kolmogorov's equation (also called the Fokker-Planck equation):

$$\frac{\partial \phi}{\partial t} + \sum_{i=1}^{n} \frac{\partial g_i \phi}{\partial \psi_i} = \sum_{i,j=1}^{n} \frac{Q_{ij}}{2} \frac{\partial^2 \phi}{\partial \psi_i \partial \psi_j}, \tag{20}$$

where $Q = \overline{\mathbf{b}\mathbf{b}^{\mathrm{T}}}$ is the covariance matrix for the model errors. A derivation of this equation is given by *Jazwinski* [1970, p. 129]. For the QG model it can be simplified to

$$\frac{\partial \phi}{\partial t} + \sum_{i=1}^{n} g_i \frac{\partial \phi}{\partial \psi_i} = \sum_{i,j=1}^{n} \frac{Q_{ij}}{2} \frac{\partial^2 \phi}{\partial \psi_i \partial \psi_j}. \tag{21}$$

The stochastic forcing introduces a diffusion term that tends to flatten the probability density function (spreading the ensemble) during the integration; that is, the probability decreases and the errors increase.

If this equation could be solved for the probability density function, it would be possible to calculate statistical moments like the mean state and the error covariances at different time levels. Actually, an analytic steady state solution of Kolmogorov's equation for the double-well problem, containing only one state variable, was found by *Miller* [1994], but for multidimensional ocean models it is not a realistic task to solve it.

A linear ocean model for a Gauss-Markov process in which the initial condition is assumed to be taken from a normal distribution $N(\mu(0), P(0))$ will have a probability density which is completely characterized by its mean $\mu(t)$ and covariance matrix $P(t)$ for all times. One can then derive exact equations for the evolution of the mean and the covariance matrix as a simpler alternative than solving the full Kolmogorov's equation. Such moments of Kolmogorov's equation are easy to derive, and several methods are illustrated by *Jazwinski* [1970, examples 4.19–4.21].

For a nonlinear model, the mean and covariance matrix will not in general characterize $\phi(\boldsymbol{\psi}, t)$. They do, however, determine the mean path and the dispersion about that path,

and it is possible to solve approximate equations for the moments, which is the procedure characterizing the extended Kalman filter. In the next section such equations will be derived for the nonlinear QG model discretized both in space and time.

Approximate Stochastic Dynamic Prediction

The multilayer QG model given by *Evensen* [1992, 1993] was written in vector notation as

$$L\boldsymbol{\psi}_{k+1} = \mathbf{f}(\boldsymbol{\psi}_k) - L\mathbf{v}_{k+1}, \tag{22}$$

where L is the discretized Helmholtz operator, and $\mathbf{f}(\boldsymbol{\psi}_k)$ is a nonlinear vector function containing the advected vorticity and the boundary stream function values. The last term contains random white noise due to errors in the imperfect model. An equation for the propagation of the mean can be derived by taking the expected value of the ocean model (22):

$$L\boldsymbol{\mu}_{k+1} = \overline{\mathbf{f}(\boldsymbol{\psi}_k)}. \tag{23}$$

The nonlinearity of $\mathbf{f}(\boldsymbol{\psi}_k)$ makes it impossible to evaluate the exact value of the right-hand side of (23). Thus for nonlinear models, it is impossible to derive a closed finite set of equations for the moments, and approximative methods must be applied. It is appropriate to express $\mathbf{f}(\boldsymbol{\psi}_k)$ as a Taylor expansion about the mean $\boldsymbol{\mu}_k$. With the definitions of the tangent linear operator

$$\mathcal{F} = \mathcal{F}_{ij} = \left.\frac{\partial f_i(\boldsymbol{\psi})}{\partial \psi_j}\right|_{\boldsymbol{\psi}=\boldsymbol{\mu}} = \nabla \mathbf{f}(\boldsymbol{\psi})|_{\boldsymbol{\psi}=\boldsymbol{\mu}}, \tag{24}$$

the Hessian

$$\mathcal{H} = \mathcal{H}_{ijk} = \left.\frac{\partial^2 f_i(\boldsymbol{\psi})}{\partial \psi_k \psi_j}\right|_{\boldsymbol{\psi}=\boldsymbol{\mu}} = \nabla\nabla \mathbf{f}(\boldsymbol{\psi})|_{\boldsymbol{\psi}=\boldsymbol{\mu}}, \tag{25}$$

and

$$\mathcal{T} = \mathcal{T}_{ijkl} = \left.\frac{\partial^3 f_i(\boldsymbol{\psi})}{\partial \psi_j \psi_k \psi_l}\right|_{\boldsymbol{\psi}=\boldsymbol{\mu}} = \nabla\nabla\nabla \mathbf{f}(\boldsymbol{\psi})|_{\boldsymbol{\psi}=\boldsymbol{\mu}}, \tag{26}$$

where i, j, k, and l are dummy indices, the third-order Taylor expansion for the vector function $\mathbf{f}(\boldsymbol{\psi})$ about the mean $\boldsymbol{\mu}$ can then be written

$$\mathbf{f}(\boldsymbol{\mu} + \Delta\boldsymbol{\psi}) = \mathbf{f}(\boldsymbol{\mu}) + \mathcal{F}\Delta\boldsymbol{\psi} + \frac{1}{2}\mathcal{H}(\Delta\boldsymbol{\psi}\Delta\boldsymbol{\psi}^{\mathrm{T}}) + \frac{1}{6}\mathcal{T}(\Delta\boldsymbol{\psi}\Delta\boldsymbol{\psi}\Delta\boldsymbol{\psi}) + \cdots, \tag{27}$$

where $\Delta\boldsymbol{\psi} = \boldsymbol{\psi} - \boldsymbol{\mu}$. Using this expansion in the right-hand sides of (22) and (23) yields, respectively,

$$L\boldsymbol{\psi}_{k+1} = \mathbf{f}(\boldsymbol{\mu}_k) + \mathcal{F}_k\Delta\boldsymbol{\psi}_k - L\mathbf{v}_{k+1} + \frac{1}{2}\mathcal{H}_k(\Delta\boldsymbol{\psi}_k\Delta\boldsymbol{\psi}_k^{\mathrm{T}}) + \frac{1}{6}\mathcal{T}_k(\Delta\boldsymbol{\psi}_k\Delta\boldsymbol{\psi}_k\Delta\boldsymbol{\psi}_k) + \cdots, \tag{28}$$

$$L\boldsymbol{\mu}_{k+1} = \mathbf{f}(\boldsymbol{\mu}_k) + \frac{1}{2}\mathcal{H}_k P_k + \frac{1}{6}\mathcal{T}_k\Theta_k + \cdots, \tag{29}$$

where k now is the time index. By dropping the last term containing the third-order moment in (29), one has the standard equation for the evolution of the mean. Thus an approximate equation for evolution of the first moment (mean of the ensemble) is dependent on the second-order moment (covariance matrix), and also higher-order moments if more terms are included from the Taylor expansion (27).

An approximate equation for the second-order moment can be derived by inserting (28) and (29) in the definition of the covariance matrix at time t_k and taking the expectation of the resulting equation, which gives

$$LP_{k+1}L^{\mathrm{T}} = \mathcal{F}_k P_k \mathcal{F}_k^{\mathrm{T}} + LQ_{k+1}L^{\mathrm{T}} + \mathcal{F}_k\Theta_k\mathcal{H}_k^{\mathrm{T}} + \frac{1}{4}\mathcal{H}_k\Gamma_k\mathcal{H}_k^{\mathrm{T}} + \frac{1}{3}\mathcal{F}_k\Gamma_k\mathcal{T}_k^{\mathrm{T}} - \frac{1}{4}\mathcal{H}_k P_k P_k^{\mathrm{T}}\mathcal{H}_k^{\mathrm{T}} - \frac{1}{6}\mathcal{H}_k P_k\Theta_k^{\mathrm{T}}\mathcal{T}_k^{\mathrm{T}} - \frac{1}{36}\mathcal{T}_k\Theta_k\Theta_k^{\mathrm{T}}\mathcal{T}_k^{\mathrm{T}} + \cdots, \tag{30}$$

where $Q_{k+1} = \overline{\mathbf{v}\mathbf{v}^{\mathrm{T}}}$ is the model error covariance matrix. Of course, for solving (29) and (30) it is also necessary to include equations for Θ and Γ, which again requires references to higher-order moments. This has been illustrated for the case with a zero dimensional model by *Miller et al.* [1994].

It is quite apparent that the equation for the evolution of the mean (29) and the equation for the covariance matrix (30) are an unclosed system of equations. The expansion of the nonlinear term contain references to infinitely many higher-order statistical moments. Thus to integrate these equations, a closure scheme must be applied. In the standard form of the extended Kalman filter, one will normally close the system by neglecting all moments with higher orders than the covariances. However, higher-order turbulence closure schemes have been discussed by, for example, *Leith* [1971, 1974], *Leith and Kraichnan* [1972], and *Fleming* [1971a, b]. An example is the eddy-damped quasi-normal closure where a "damped" equation for the third moment Θ is included and the fourth-order moment is expressed in terms of products of the second-order moment, which is a valid approximation for a multivariate normal distribution. Variants of this scheme were tested and compared with results from Monte Carlo simulations by *Fleming* [1971a]. He found that the higher-order closure significantly improved the results, and it should also improve the results for the evolution of the error statistics in the extended Kalman filter for the QG model. Note that these studies were all performed using spectral models including only a few wave modes. For today's standard ocean models it is not practical to store higher order moments on existing computers.

Another issue for discussion is whether one should integrate the model equation for the central forecast or the equation for the mean. In previous applications of the extended Kalman filter, the pure model equation has been used in favor of the equation for the mean. However, in a recent paper, *Cohn* [1993] suggested that for weather forecasting the equation for the ensemble mean should be used. This argument was based on analytical calculations with the Burgers equation, where it was found that biasing the model also saturated the error growth in the error covariance equation.

To saturate the error growth in the QG model, the vorticity is required to approach zero, since nonzero vorticity

leads to a nonzero error growth term (see the discussion by *Evensen* [1992]). For a long time integration of the conservative QG model, one would expect the mean to approach zero, at least for a case with flat bottom topography and no external forcing. This would then result in a vanishing error growth term, and this effect might correspond to the result found by *Cohn* [1993] when using the equation for the mean saturated the error growth. It is believed that a physically more acceptable error growth saturation must be the result of a higher-order closure in the actual error covariance equation. However, it would be interesting to see an experiment in which the equation for the mean is used.

If the mean is used as the estimate for the forecast, this might in some cases lead to nonphysical results. As an example, one could consider the double-well problem discussed by *Miller et al.* [1994] and *Miller* [1994]. In that case there are two stable solutions (one in each of the wells) and one unstable solution between them. The mean solution would here be given by the unstable equilibrium in between the physically realistic solutions. A similar problem arises when one studies the two stable equilibria of the Kuroshio, where the mean again would correspond to a nonphysical solution that never exists somewhere in between the two stable meanders [see *Miller*, 1994]. This problem might occur in all cases where the ensemble has a nonnormal probability distribution. See, however, the discussion below in the section containing the data assimilation experiment for the case where the ensemble is converging owing to assimilation of data.

Note that in this discussion it is the covariance matrix of the ensemble that has been referred to and not the error covariance matrix, which is used in the extended Kalman filter. However, by using the equation for the evolution of the true state ψ^{t},

$$L\psi^{\mathrm{t}}_{k+1} = \mathbf{f}(\psi^{\mathrm{t}}_{k}) - L\mathbf{v}_{k+1}, \tag{31}$$

and the equation for a model forecast ψ^{f}_{k+1} calculated from the analyzed estimate ψ^{a}_{k},

$$L\psi^{\mathrm{f}}_{k+1} = \mathbf{f}(\psi^{\mathrm{a}}_{k}), \tag{32}$$

and then performing a Taylor expansion similar to (27) but this time about the analyzed state estimate, one can derive an equation for the analyzed/forecast error covariance matrix $P^{\mathrm{a,f}} = \overline{(\psi^{\mathrm{a,f}} - \psi^{\mathrm{t}})(\psi^{\mathrm{a,f}} - \psi^{\mathrm{t}})}$. This equation will be identical to the covariance evolution equation above except that the operators (24) – (26) are calculated at the analyzed model state instead of the mean.

Monte Carlo Methods

An alternative to the approximate stochastic dynamic approach for solving Kolmogorov's equation and predicting the error statistics is to use Monte Carlo methods. A large cloud of states, that is, points in phase space, can be used to represent a specific probability density function. By integrating such an ensemble of states forward in time, it is easy to calculate approximate estimates for moments of the probability density function at different time levels. In this context the Monte Carlo method might be considered a particle method in phase space. When the ensemble size N increases, the errors in the solution for the probability density will approach zero at a rate proportional to $1/\sqrt{N}$. For practical ensemble sizes, say, $O(100)$, the errors will be dominated by statistical noise, rather than by dynamical errors as in approximate stochastic dynamic prediction.

In the Monte Carlo method, one first calculates a best guess initial condition based on information available from data and statistics. The model solution based on this initial state is denoted the central forecast. The uncertainty in this best guess initial condition is represented by the initial variance. An ensemble of initial states is then generated in which the mean equals the best guess initial condition and the variance is specified on the basis of knowledge of the uncertainty in the first-guess initial state. The covariance or smoothness of the ensemble should reflect the true scales of the system, that is, the internal Rossby radius for the QG model. A procedure for generating such pseudo random fields with a specified variance and covariance is outlined in the appendix.

The effect of external error growth must be included to give reliable estimates for the evolution of errors. In the approximate stochastic dynamic method, this can be done rather simply by adding the system error covariance matrix every time step. However, in the Monte Carlo method, another approach must be used. Little is actually known about model errors, but a crude way to model them is to force the QG model with smooth pseudo random fields with a specified variance and covariance. This will provide a realistic increase in the ensemble variance, provided that the estimate of the model error variance is reasonably good.

Error Prediction Experiment

Cohn [1993] suggested that the approximate stochastic dynamic equations should be used to predict the evolution of the error covariances. The main argument was that these equations are the governing equations for the error covariance evolution and give a good approximation for the error covariances for short time integrations, an argument which was also supported by *Lacarra and Talagrand* [1988] and *Budgell* [1986]. In the next examples it will be illustrated how the approximate error covariance solution departs from an "exact" solution based on Monte Carlo methods.

The case used for illustrations here is essentially the same as that used by *Evensen* [1992]. A square 17×17 grid and two layers with closed boundaries and a linearly sloping bottom with increasing depth in the positive y direction are used. The initial stream function field consists of an anticyclonic eddy in the center of the domain with the same amplitudes in both the upper and the lower layers (see upper left plot in Figure 4).

The initial error variance field is assumed to have zero variance at the closed boundaries where the solution is assumed to be exactly known and has a smooth increase in error variance to an amplitude of 0.25 at a couple of Rossby radii from the boundaries (see upper left plot in Figure 5).

The structures of the error covariance functions are determined by the same exponential relation as that used by *Evensen* [1992], with a horizontal decorrelation scale of 2.0 and a vertical decorrelation factor between the two layers

equal to $e^{-1.0}$ (see also the appendix). The initial covariance function for the upper layer is shown in Figure 7.

Evensen [1992] showed that the transition matrix could be written as a sum of two terms:

$$\mathcal{F}_k = \Gamma(\psi_k^a)L + \left.\frac{\partial\Gamma(\psi)}{\partial\psi}\right|_{\psi=\psi_k^a}(L\psi^a + \eta). \qquad (33)$$

Here L is the Helmholtz operator, $\Gamma(\psi)$ is the nonlinear advection operator, and η is the bottom topography term. The first term in the transition matrix then describes pure advection of the covariance vorticity by the model estimated velocity field.

The second term is a result of the nonlinearity of the advection operator and it contains the derivatives of the advection operator times the stream function vorticity. It is this term that causes the internal error growth in the extended Kalman filter. See also the extensive discussion by *Evensen* [1992].

Description of Cases

The following cases are run to illustrate the differences and similarities between the approximate stochastic error covariance evolution (ASDP) and the method based on Monte Carlo simulations (MCS). Except for the last case MCS100n no model error is included since these experiments are designed for studying the dynamical evolution of the error covariances.

Case ASDP. The full transition matrix has been used to integrate the standard error covariance equation, that is, all moments with an order higher than 2 in (30) have been set to zero. The internal error growth is included, but it becomes unbounded because of the use of a too simplified closure.

Case ASDPs. Case ASDPs is a simplified case in which only the first part of the transition matrix has been used to integrate the error covariance equation. The term resulting in the error growth is neglected, and only the effect of advection of the error covariance vorticity in the model estimated stream function field is included. This simplified approach was shown by *Evensen* [1992] to work well with the extended Kalman filter.

Case MCS100. The error evolution is computed using the Monte Carlo method with an ensemble size of 100 members. The initial ensemble is generated using the procedure outlined in the appendix.

Case MCS500. Case MCS500 uses the same parameters as case MCS100, but the ensemble size has been increased to 500 members for verification of the accuracy of the Monte Carlo method.

Case MCS100n. Case MCS100n is the same as case MCS100, but the ensemble is forced by smooth pseudo random fields with variance equal to 0.0001 to simulate system noise.

Ensemble Solutions

In Figure 3 the first six stream function fields from the ensemble in case MCS100 are shown at the initial time $t = 0$ and the final integration time $t = 200$ in both the upper and lower layers. Several interesting conclusions can be drawn from these results.

Note first the specified correlation between the stream function in the upper and lower layers for the initial time and the significantly stronger correlation at the final time. An explanation for the increase in correlation is that the model is forced only through interaction with the bottom topography and that the upper layer stream function adjusts to the lower layer stream function.

When the ensemble solutions at $t = 200$ are examined, there seem to exist two "quasi-equilibrium" states, one in which most of the energy is contained in an anticyclonic eddy close to the boundary $y = 1$ in the shallow part of the domain and another in which the energy is contained in a cyclonic eddy close to the boundary $y = 17$ in the deep part of the domain. By examining a larger part of the ensemble solutions, it was found that about one third of the ensemble members approached the cyclonic state, while about two thirds ended in the anticyclonic state, except for a few cases in which other states were approached, for example, the sixth ensemble member in Figure 3. Thus even this simple example seems to evolve toward a non-Gaussian distribution function.

There is also an increase in the horizontal correlation scale; that is, the final states are smoother than the initial states. This is to be expected for QG models on the basis of two-dimensional geostrophic turbulence which has been discussed by *Pedlosky*, [1987, pp. 169–177]. For motion subject to quasi-geostrophic dynamics there will be a transformation of energy to larger scales (smaller wave numbers), accompanied by a transfer of enstrophy to smaller scales (larger wave numbers).

The large difference in the stream function solutions for the members of the ensemble also provides a warning against having too much faith in a single forecast of an ocean model. When a simple model like the QG model has such a complicated behavior, one should certainly not expect more sophisticated primitive equation models to provide results with any greater predictability.

Evolution of the Central Forecast and Mean

In Figure 4 the upper layer stream function is shown at different time levels for the central forecast in the first column, and the ensemble means from the three Monte Carlo cases MCS100, MCS500, and MCS100n are shown in the next three columns. Note first the similarity between the solutions in cases MCS100 and MCS500. This actually suggests that an ensemble size of $O(100)$ is sufficient to calculate the mean with reasonable accuracy. In these two cases the mean solutions have some structures similar to those of the central forecast. However, the amplitudes are lower, and the mean solution contains less energy. Note again the non-Gaussian distribution of the final ensemble, which explains the lower amplitudes in the forecast mean. Actually, if the ensemble members were distributed equally between the two states, the mean solution would be close to zero.

The case which is forced by external errors produces a significantly different mean solution which is more energetic than the unforced cases. The external forcing therefore has a significant impact on the ensemble, and it also seems to

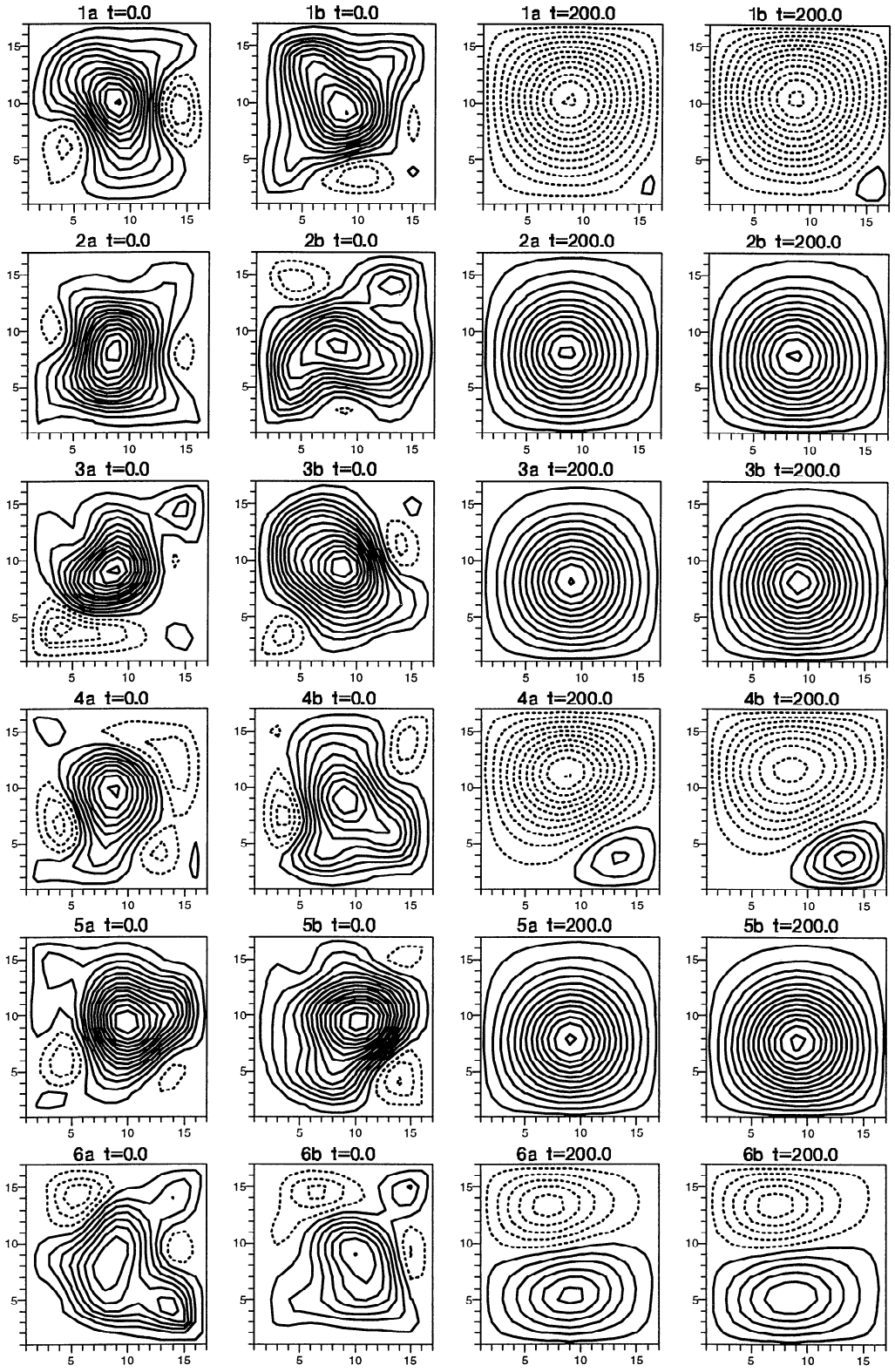

Figure 3. Here the first six members of the ensemble are shown for $t = 0$ in the upper and lower layers (first and second column) and for $t = 200$ in the upper and lower layers (third and fourth column). The contour interval is 0.2.

provide a different final distribution function, with a larger fraction of the members in the anticyclonic state.

Evolution of Error Variance

Evensen [1992] discussed the instability in the error covariance equation and proposed a few approaches for handling it. One could use an extensive data set in the Kalman filter to control the error growth. This was shown to work if one had "complete" data coverage, but in both oceanography and meteorology, there will be data void regions where the errors grow. Another approach was to neglect the term in the transition matrix that generated the error growth, that is,

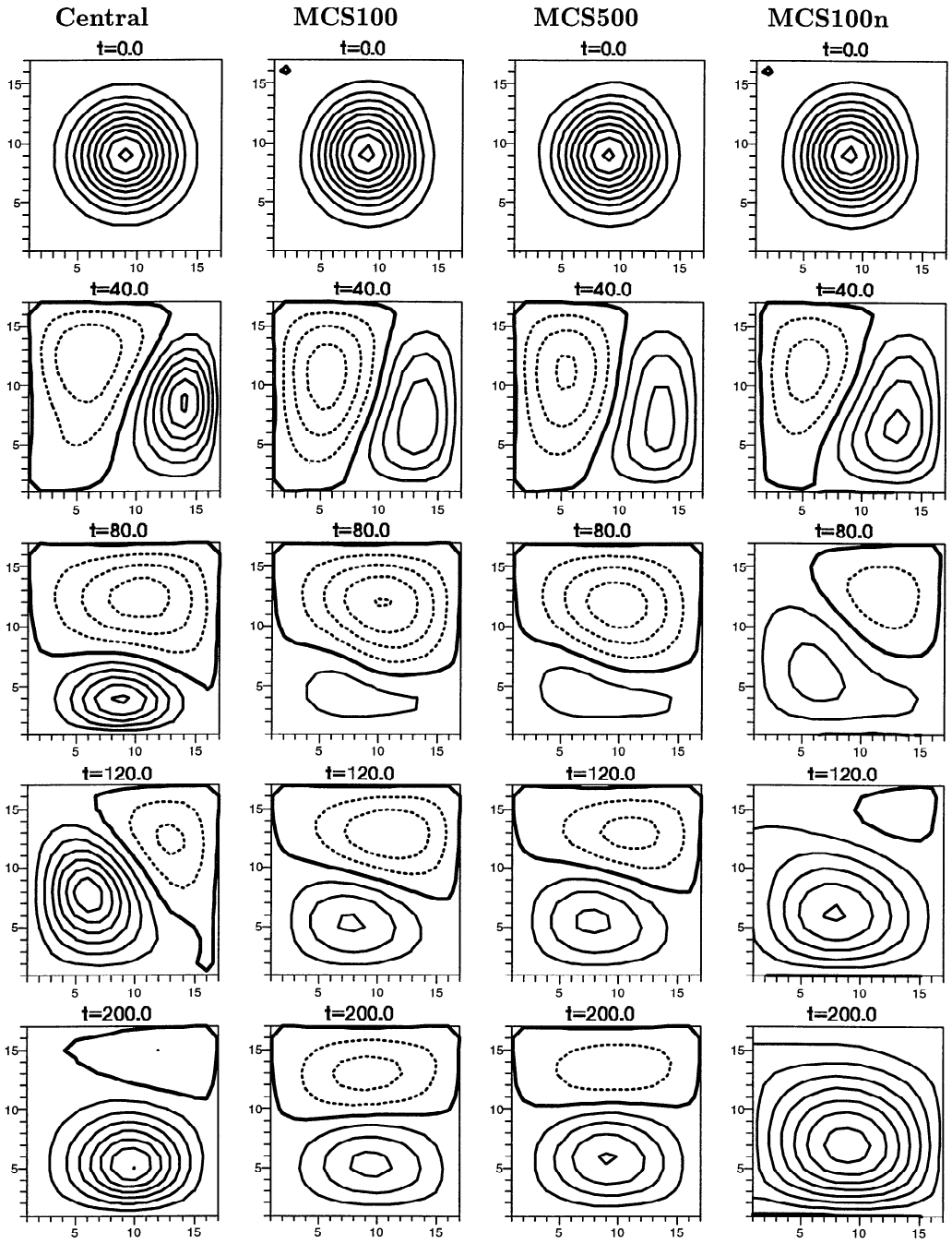

Figure 4. Evolution of the central forecast and mean generated by Monte Carlo simulations. The contour interval is 0.2.

case ASDPs. This eliminated the unboundedness of the error growth, but the true effect of internal error growth caused by the inherent dynamical instability in the model was lost.

Figure 5 shows the evolution of error variance for the cases described above. Note that the contour interval in case ASDPs is 0.025, while the rest of the plots in this figure use an interval of 0.1. In case ASDP the errors start growing exponentially at an early stage, and after time $t = 20$ the errors are unrealistically high. For even longer time integrations the errors grow exponentially until the gradients in the covariance functions become too steep and the numerical schemes are unable to solve the error covariance equation with any accuracy. This causes the error growth to be limited by the diffusion and dispersion in the numerical schemes. In case ASDPs it is clearly seen how the errors decrease because of diffusion in the numerical schemes and there is no error growth.

The error variance evolution in the Monte Carlo method seems to be more realistic. The errors increase rather quickly in the first stage of the integration, but after a while the error growth saturates and a stable error level is approached. Note the similar structures in the errors for cases ASDP and MCS500 at time $t = 10$ and $t = 20$. The locations of the error growth are essentially the same but the error growth is

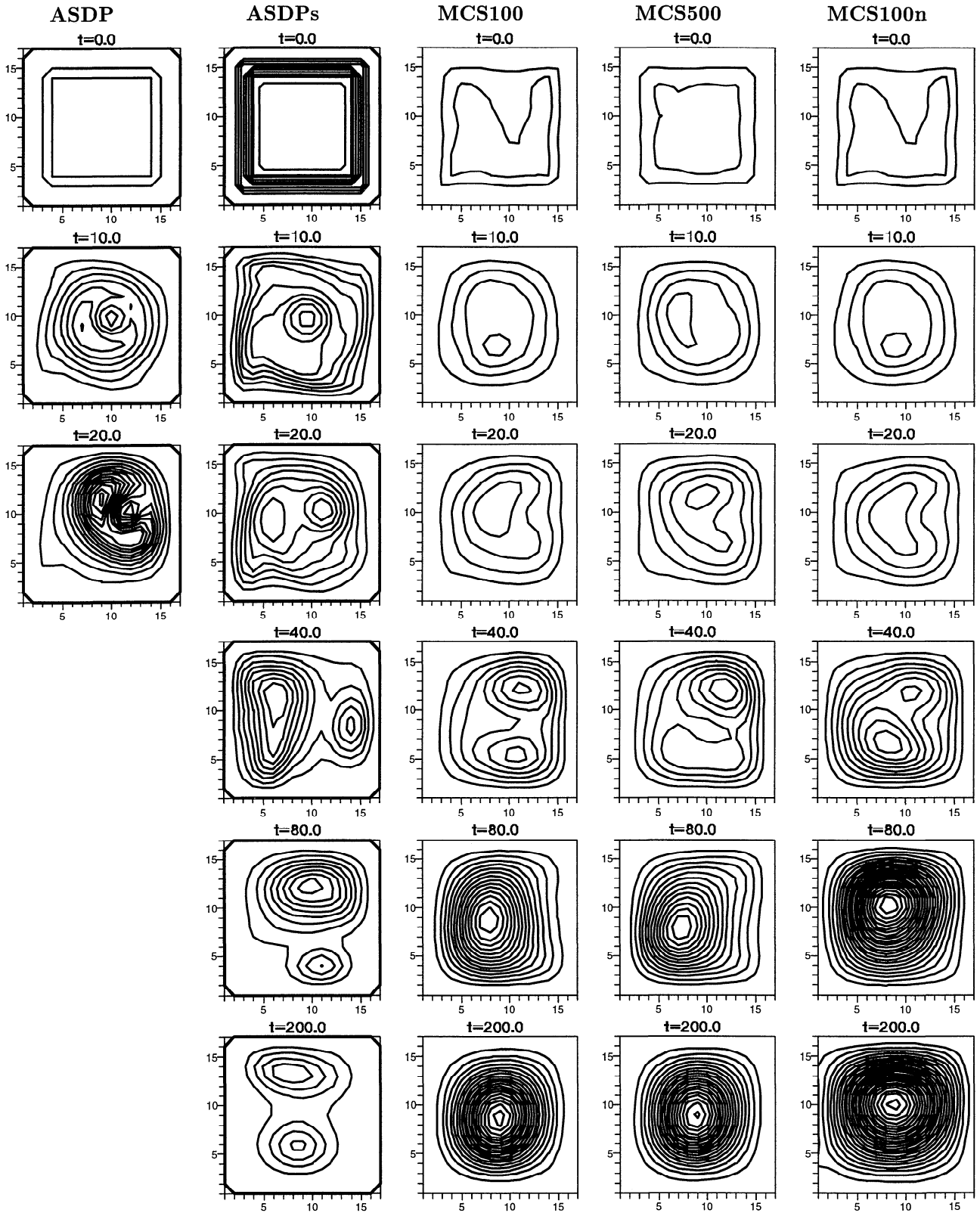

Figure 5. Error variance fields at different time levels from the stochastic dynamic approach including/excluding the second part of the transition matrix and from the Monte Carlo simulations. the contour interval is 0.025 in the second column and 0.1 in the rest of the columns.

stronger for case ASDP. The error variance maps for cases MCS100 and MCS500 are also rather similar, and an ensemble of $O(100)$ members should be large enough to estimate the error variance evolution with reasonable accuracy.

From Kolmogorov's equation (21) it can be seen that the effect of external error growth is unbounded, that is, the diffusion term will tend to smooth or widen the probability density function for infinitely long time intervals. However,

the effect of the diffusion term becomes very small after the variance has increased a certain amount, and the contribution of the external error growth will then be less important. Comparing case MCS100n to case MCS100 illustrates how system noise leads to a stronger error growth.

The evolution of the mean square errors are shown in Figure 6. One first notices the strong error growth for $t \in [0, 40]$ for case ASDP and then the failure of the numerical schemes which produces results with no physical reality. The two unforced Monte Carlo runs, MCS100 and MCS500, have similar behaviors and stay close all through the integration. Note also the saturation of the error growth. The forced Monte Carlo run, MCS100n, has a stronger error growth in the beginning and seems to saturate but with larger errors. The simplified case ASDPs predicts an unrealistic decrease in errors which is caused by diffusive effects.

Evolution of Error Covariance Functions

To illustrate how the error covariances evolve in time for the different cases discussed above, the covariance function corresponding to grid-point (6, 6) in the upper layer is shown in Figure 7 for the same time levels and cases as in Figure 5. Actually, it is hard to compare covariances between the different cases because of the dependency on the respective variance fields, that is, the amplitudes should not be compared. The covariance functions at time $t = 10.0$ have a rather similar structure for all the cases. At $t = 20.0$ it is hard to compare cases ASDP and ASDPs with the Monte Carlo cases because the error variance fields are rather different. However, for the three Monte Carlo cases, one notices the similarities in structure and horizontal scale. Note also the increasing horizontal scale of the covariance functions. This is also reflected by the smoother solutions in the final ensemble solutions in Figure 3.

An Alternative to the Extended Kalman Filter

The results from the previous sections on error prediction using Monte Carlo methods are encouraging with respect to applying them for error evolution in the Kalman filter. It is actually possible to calculate reasonable statistics using an ensemble size of $O(100)$ members. This corresponds to a numerical load where the required CPU equals $O(100)$ model integrations and the storage is $O(100)$ model state vectors of dimension n. This should be compared to the extended Kalman filter, which requires $2n$ model integrations and storage of the error covariance matrix of dimension n^2. Note also that the required storage and CPU increase linearly with the size of the state vector for the Monte Carlo approach.

There are, then, two main issues to consider. A method is required for assimilating the measurements and calculating an updated state estimate. One also has to modify the error statistics by operating on the ensemble in some way.

Updating the State Estimate

When the error statistics of the measurements $\mathbf{z}_k$ and the forecast state estimate $\boldsymbol{\psi}_k^{\mathrm{f}}$ are described by their respective error covariance matrices R_k and P_k^{f}, the optimal linear combination of measurements and state estimate becomes

$$\boldsymbol{\psi}_k^{\mathrm{a}} = \boldsymbol{\psi}_k^{\mathrm{f}} + K_k(\mathbf{z}_k - H_k\boldsymbol{\psi}_k^{\mathrm{f}}), \tag{34}$$

where the Kalman gain K_k is given by

$$K_k = P_k^{\mathrm{f}} H_k^{\mathrm{T}}[H_k P_k^{\mathrm{f}} H_k^{\mathrm{T}} + R_k]^{-1}. \tag{35}$$

Covariance functions can be calculated directly from the ensemble, and it is therefore a simple process to generate a gain matrix from formula (35) using ensemble statistics and without storing the full error covariance matrix P. The measurement matrix will normally contain only a few nonzero elements in each row; that is, each row relates the stream function field to a measurement.

The product PH^{T} can be generated by calculating only the covariance functions (columns in P), which correspond to nonzero elements in the measurement matrix H. Normally, only a small number of covariance functions are needed, and this number is always less than or equal to the number of nonzero elements in the measurement matrix. When the

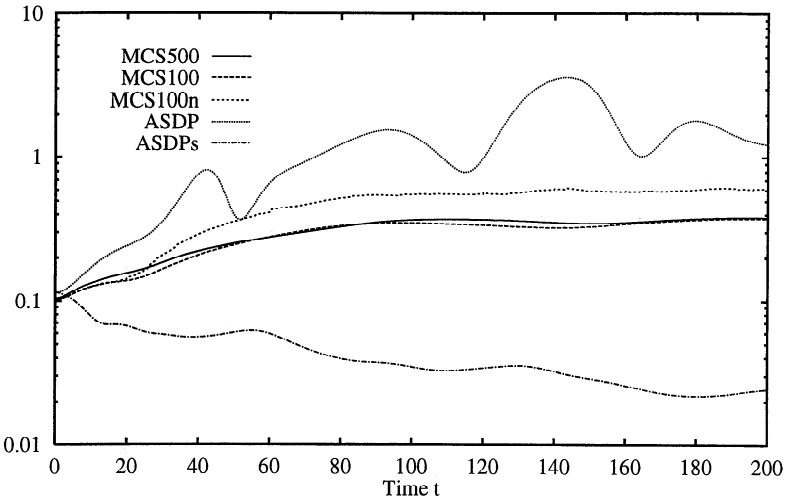

Figure 6. Time evolution of the mean square errors. Note the stronger growth of error amplitudes in the approximative stochastic approach ASDP and also the similarity between the Monte Carlo forecasts with different sizes on the ensembles.

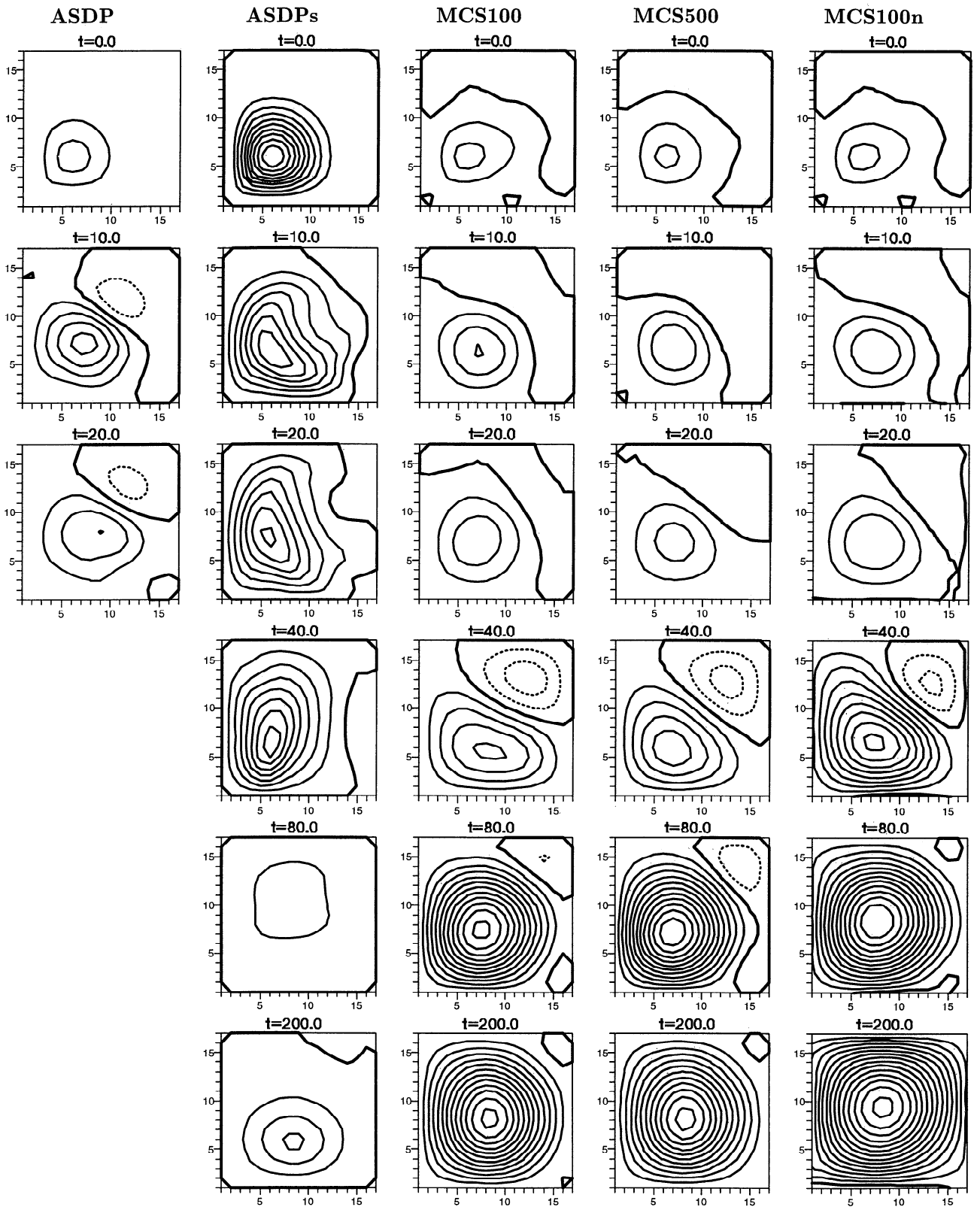

Figure 7. Error covariance fields from the same cases at the same times as in Figure 5. Again the contour interval is 0.025 in the second column and 0.1 for the rest of the plots.

product PH^{T} is calculated, the gain K can be found by solving a sequence of linear systems.

Updating the error statistics

The next issue concerns the update of the statistical ensemble to correct the error statistics after an analysis with measurements. Thus, after an analysis the ensemble variance should be correctly reduced at the measurement locations and also have the correct covariances.

Consider a scalar state described by a model forecast w_k^{f} and a measurement z_k at a time t_k. The model forecast and the measurements contain errors $\tilde{w}_k^{\mathrm{f}}$ and $\tilde{z}_k$ and therefore give only an approximate estimate of the state. The estimates w_k^{f}, w_k^{a}, and z_k are related to the true state w_k^{t} through the expressions

$$w_k^{\mathrm{f}} = w_k^{\mathrm{t}} + \tilde{w}_k^{\mathrm{f}}, \tag{36}$$

$$w_k^{\mathrm{a}} = w_k^{\mathrm{t}} + \tilde{w}_k^{\mathrm{a}}, \tag{37}$$

$$z_k = w_k^{\mathrm{t}} + \tilde{z}_k, \tag{38}$$

where the expectations $\overline{\tilde{w}_k^{\mathrm{f}}} = \overline{\tilde{w}_k^{\mathrm{a}}} = \overline{\tilde{z}_k} = 0$.

The analyzed estimate is given by

$$w_k^{\mathrm{a}} = w_k^{\mathrm{f}} + \frac{P_k^{\mathrm{f}}}{P_k^{\mathrm{f}} + R_k}(z_k - w_k^{\mathrm{f}}), \tag{39}$$

where $P_k^{\mathrm{f}} = \overline{(\tilde{w}^{\mathrm{f}})^2}$ and $R_k = \overline{(\tilde{z})^2}$ are the error variance of the forecast and the measurement, respectively. See, for example, *Ghil* [1989] for an introduction to linear estimation theory.

The error covariance for the analyzed estimate becomes

$$\begin{aligned} P^{\mathrm{a}} &= \overline{(\tilde{w}^{\mathrm{a}})^2} = \overline{(w^{\mathrm{a}} - w^{\mathrm{t}})^2} \\ &= \overline{\left(w^{\mathrm{f}} + \frac{P^{\mathrm{f}}}{P^{\mathrm{f}} + R}(z - w^{\mathrm{f}}) - w^{\mathrm{t}}\right)^2} \\ &= \overline{\left(\tilde{w}^{\mathrm{f}} + \frac{P^{\mathrm{f}}}{P^{\mathrm{f}} + R}(\tilde{z} - \tilde{w}^{\mathrm{f}})\right)^2} \\ &= \left(I - \frac{P^{\mathrm{f}}}{P^{\mathrm{f}} + R}\right) P^{\mathrm{f}} \\ &\quad + 2\frac{P^{\mathrm{f}}}{P^{\mathrm{f}} + R}\left(1 - \frac{P^{\mathrm{f}}}{P^{\mathrm{f}} + R}\right)\overline{\tilde{z}\tilde{w}^{\mathrm{f}}}, \end{aligned} \tag{40}$$

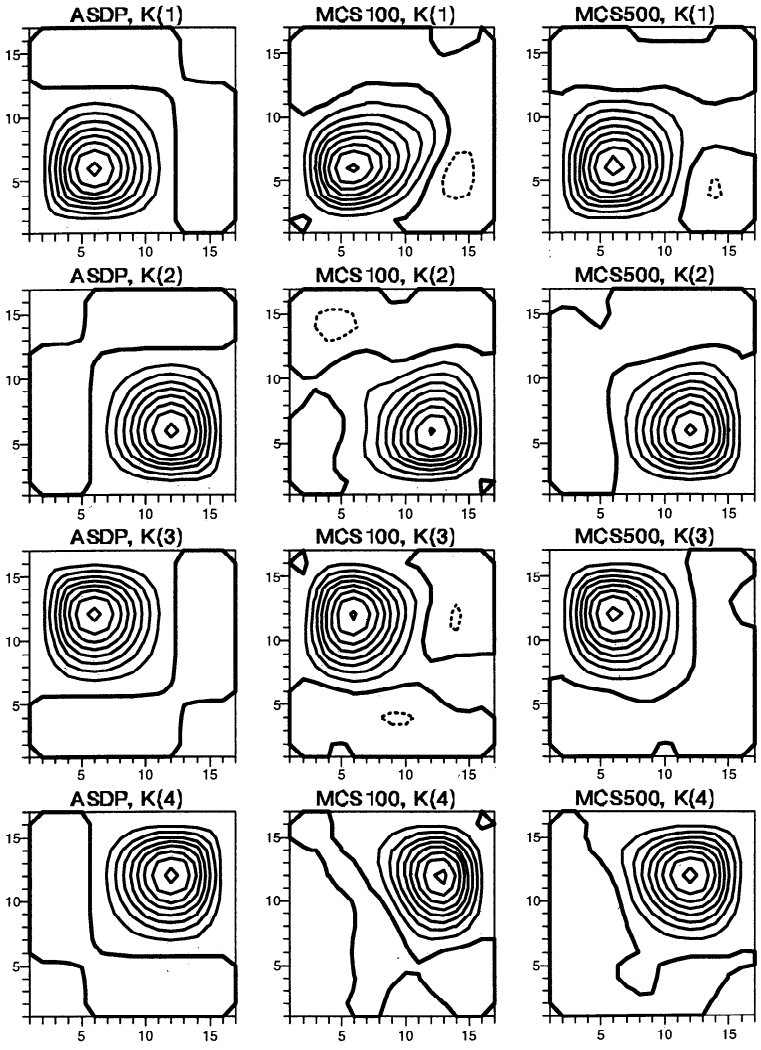

Figure 8. Kalman gains for case ASDP and gains from cases MCS100 and MCS500 for four measurement points.

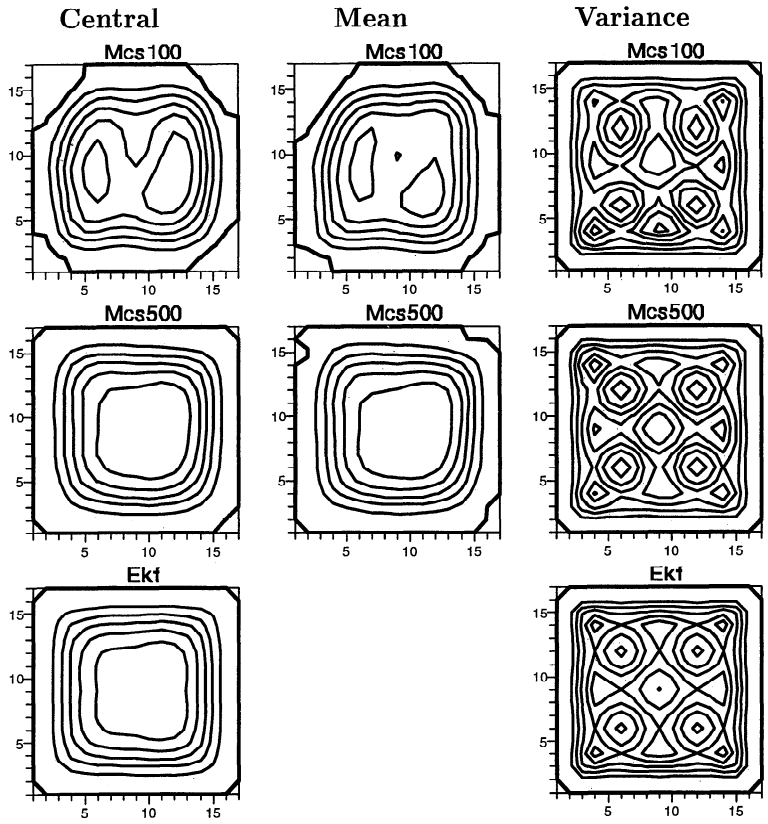

Figure 9. Analyzed central forecasts and means in the two left columns and analyzed error variance in the right column for the cases MCS100, MCS500, and ASDP after one single update with measurements.

where the time indices have been dropped. In the Kalman filter one assumes the measurement errors to be uncorrelated with the forecast errors $\overline{\tilde{z}\tilde{w}^{\mathrm{f}}} = 0$. The analyzed error covariance matrix can then be calculated from the remaining terms, which for the vector case becomes

$$P^{\mathrm{a}} = (I - KH)\,P^{\mathrm{f}}. \tag{41}$$

The derivation (40) also shows through the second line that the analyzed error covariance matrix can be calculated by operating directly on the ensemble itself. Thus, by calculating the Kalman gain and then updating all the members in the ensemble using (34), the resulting ensemble will have the same error covariances as one would get from (41), which is used in the standard Kalman filter.

The error covariance matrix is defined as the variance and covariance about the true state, which unfortunately is unknown. One therefore has to decide whether one should calculate the covariances about the ensemble mean or the central forecast. The two choices will give slightly different estimates for the error statistics, and the quality will depend on which is the best estimate of the true state. At this stage the ensemble mean has been used.

Example

Now a simple example will be used to compare the Kalman gains, analyzed means and central forecast, and error variance fields as calculated from the Kalman filter and the method based on ensemble statistics. Four measurements taken from the initial stream function in case ASDP at four different grid points are used to update a zero stream function once. There is no dynamical evolution in this case. The Kalman gains for the four measurements are shown in Figure 8, calculated both by using the ensemble statistics with ensemble sizes of 100 and 500 and by using the analytically prescribed covariance functions that are initially used in the extended Kalman filter. As expected, these Kalman gains are rather similar for the three cases. The analyzed central forecast, mean, and variance fields are shown in Figure 9. The conclusions thus far are that the Monte Carlo method provides realistic error forecasts and that the outlined procedure for generating analyzed fields gives promising results.

Data Assimilation Experiment

The previous sections have discussed the fundamental theory required for applying the Monte Carlo method and ensemble statistics as an alternative to solving the error covariance equation in the extended Kalman filter. The central forecast stream function from the evolution experiments shown in Figure 4 in the time interval $t \in [0, 60]$ is now used to generate measurements in four grid points in the upper layer every $\Delta t = 2.0$. These measurements are then used in an assimilation experiment in which the initial stream function has been set to zero all over the domain. The initial variance field is shown in, for example, Figure 5 and is set to 0.25 in the interior of the domain and to zero at the closed boundaries where the stream function is assumed to be known. For

the error covariance functions, the horizontal decorrelation length is $r_h = 2.0$, and the factor for the vertical correlation is $e^{-0.25}$ (see the appendix). The simulated measurements are assumed to have a variance of 0.02 to reduce their influence, although they are generated without noise. The ensemble is forced by a system noise with variance 0.0001, which is the same as was used in case MCS100n. For a case with real dynamics with the QG model set up in a realistic domain, the system noise should certainly be higher, but in this twin experiment, where the QG model has been used to generate the data, such a low value might account for numerical truncation errors. Here an ensemble size of 500 has been used, although results from the previous examples suggest that an ensemble size of 100 would be sufficient. For this case on such a small domain, only a few minutes are required on a CRAY Y-MP.

The results from the single data assimilation experiment are shown in Figure 10 for the upper layer solution and in Figure 11 for the lower layer solution. The time evolution for the reference stream function is shown in the leftmost column, and the results from the assimilation procedure for the central forecast and the mean are shown in the second and third columns. Both the central forecast and the mean are converging toward the reference solution, and it is not easy to visually determine which gives the estimate closest to the reference case. At $t = 30$, both the main structures and their amplitudes are quite well regenerated, both for the central forecast and the mean. Examining the numbers in Table 1, shows that for the upper layer, the solution for the central forecast is slightly better than the mean, while for the lower layer, the mean gives a better fit than the central forecast to the reference case.

The assimilation procedure performs very well, and the results from this example are significantly better than those from the experiments of *Evensen* [1992] in which the full extended Kalman filter was used. However, the results are rather similar to the simplified case 2A from *Evensen* [1992]. See Table 1 also for a comparison of the errors in the solution from the extended Kalman filter and the mean and central forecast solutions resulting from the Monte Carlo method.

One could probably say that the mean and the central forecast are about equally good in this experiment. In practical data assimilation applications, though, it is not obvious whether one should use the central forecast or the mean of the ensemble as the "best" estimate. Intuitively, the mean should be used if the ensemble has converged (reflected by a very low variance in the ensemble). A problem might be that the ensemble is converging in data-dense regions while the variance remains quite large in data-sparse regions. In that case the mean could be used in the regions with low variance, while the central forecast and each of the ensemble members will provide equally bad but at least physically acceptable solutions in the regions with high variance. However, the mean might be physically unacceptable if the ensemble members have a nonnormal distribution in the regions with high variance.

This experiment, in which data are assimilated only in the upper layer in a two-layer model, also raises the question of vertical projection of surface information. The experience gained so far, based on applications with the Kalman filter by *Evensen* [1992, 1993] and the Monte Carlo approach in this paper, suggests that the dynamical evolution of the error statistics also provides the correct vertical influence functions, and that data in the upper layer will be sufficient to reconstruct the barotropic and the first baroclinic modes. This in an important issue for assimilation of altimeter data in models for large-scale ocean circulation, and the results support the use of advanced data assimilation methods for handling the vertical projection problem.

In the error variance fields (fourth column in Figures 10 and 11), it is seen how the error variance estimates are reduced faster in the upper layer because of the assimilation of data than in the lower layer, where the data have less influence. By studying the variance plots in detail, one can also observe the advection of error variance in the background velocity field, for example, the high amplitude near the right boundary in the lower layer at $t = 10$–30, which is moved toward the lower boundary. The evolution of the mean square errors for this experiment is shown in Figure 12. It is again seen how the errors in layer one are lower than those in layer two. The time evolution is similar to what one would expect from previous examples with the Kalman filter for linear models. Of course, here the slight increase in errors between assimilation times is a result of both the system noise and the dynamical error growth. Note also that the too strong error growth from the data assimilation experiments of *Evensen* [1992] is absent. From the curves in Figure 12 it is expected that the errors would continue to decrease slightly if the assimilation process was continued for longer time.

It is interesting to see the similarities between the estimated error variance fields and the calculated residuals (rightmost columns in Figures 10 and 11), especially in the lower layer for $t = 20$, 30, and 40, that is, highest variance in regions with largest residuals. The residuals are defined as reference case minus central forecast assimilation case. From the low variance and residuals at the final time $t = 60$, it is clear that the assimilation process is converging. This was expected with such good data coverage.

Discussion

A new sequential data assimilation algorithm has been outlined and discussed. It has been illustrated how instead of solving the error covariance equation which is used in the extended Kalman filter, one can solve the "full" Kolmogorov's equation, which is the exact equation describing the evolution of the error statistics using Monte Carlo methods. An ensemble of probable initial states is generated with a specified mean, variance, and covariance. By integrating the ensemble forward in time and by properly updating the ensemble when data are available, all the error statistics required in the extended Kalman filter can be calculated directly from the ensemble.

Evensen [1992] reported a serious closure problem for the approximate stochastic evolution equation, and *Evensen* [1993] also discussed the problem of posing consistent open boundary conditions for the approximate stochastic evolution equation. The method outlined here resolves both of these problems. By solving the full Kolmogorov's equation, no approximate closure schemes need be used, and as long as

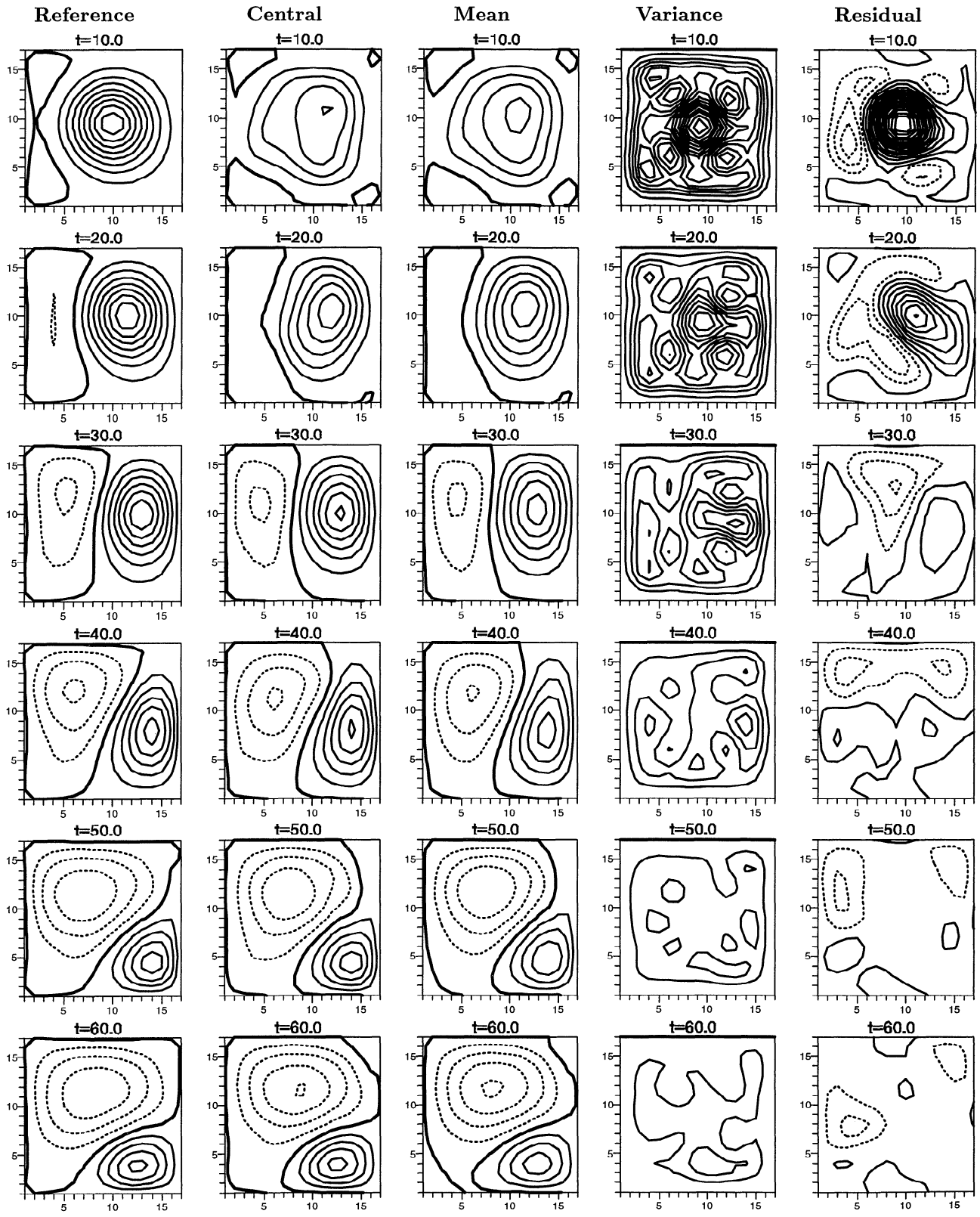

Figure 10. The reference stream function is shown in the left column. The next two columns contain the central and mean forecasts for the stream functions that result from the assimilation process. The contour interval is 0.2 for all the stream function plots. The two rightmost columns contain the variance plots and the residual defined by the reference stream function minus the central forecast generated by the assimilation process.

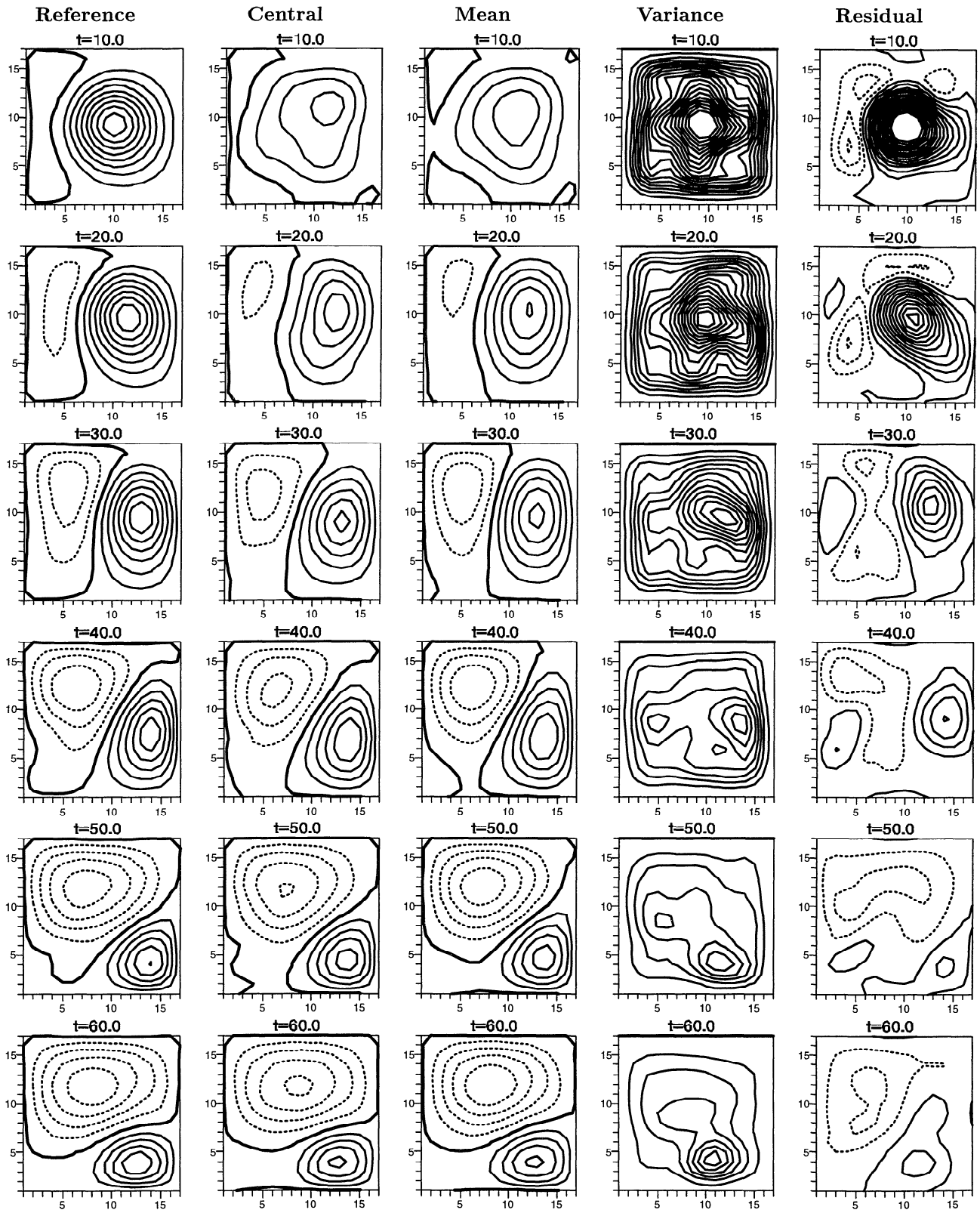

Figure 11. Same as Figure 10 but for the lower layer.

the boundary conditions are well posed for the ocean model, this will normally also be true for each of the ensemble members.

The standard Kalman filter has several major drawbacks. It carries a huge amount of information forward in time; that is, it stores the covariance functions for all the state variables in every grid point, even though only the covariance functions referenced by the measurement matrix are required. One would also expect the covariances to approach zero at a certain distance. This expectation has motivated attempts

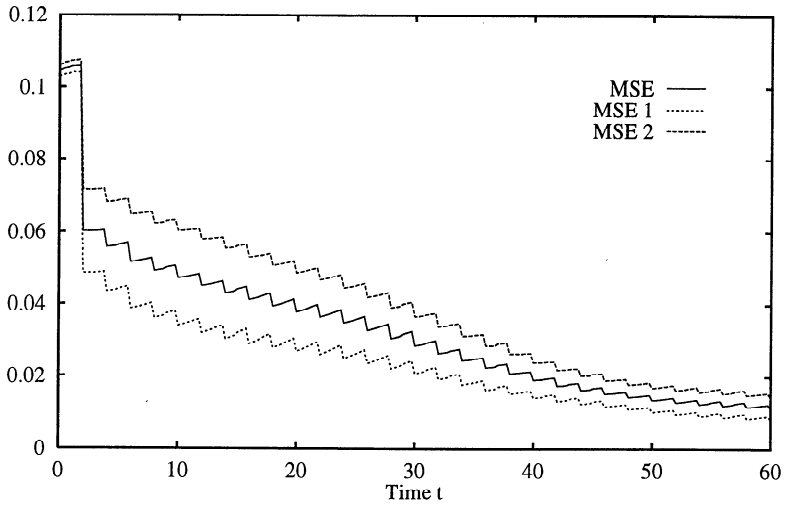

Figure 12. Time evolution of the mean square errors for the data assimilation experiment using Monte Carlo methods to forecast error statistics. Both the evolution of the total mean square errors for the upper and lower layers (middle curve) and the mean square errors for each of the individual layers are shown.

by several scientists (including the author of this paper) to store only a banded covariance matrix for reducing the computational load. The results so far seem not to be especially useful, mainly because any numerical approximation used for the error covariance evolution equation results in unstable numerical algorithms owing to loss of positive definiteness.

Some authors have tried to simplify the dynamical equations for the error statistics: for example, *Dee* [1991] projected the error statistics on the slow manifold, and *Cohn* [1993] derived an approximate system of partial differential equations which described the qualitative behavior of the error covariance evolution.

The approach presented in this paper resolves several of the main issues concerning applications of the extended Kalman filter with a nonlinear quasi-geostrophic (QG) model. The next stage would be to test the method with realistic primitive equation models.

Besides the points mentioned above, the method also has a significantly lower computational load. Only the ensemble, which has a size of, say, $100n$ where n is the number of state variables in the ocean model, needs to be stored. Note that storage increases linearly with n in contrast to the standard Kalman filter, for which storage is proportional to n^2. The required CPU corresponds to $O(100)$ model forecasts if $O(100)$ members are used in the ensemble, in contrast to $2n$ model forecasts in the extended Kalman filter. The method is very simple to implement, it should be feasible for full primitive equation models on existing computers, and it also provides better results than the traditional extended Kalman filter. The proposed method is also well suited for parallel computers on which each member of the ensemble could be integrated independently by a single processor.

Table 1. Error Estimates

	Maximum Error	MSE
Upper layer case 2A	0.15	0.0019
Lower layer case 2A	0.15	0.0033
Upper layer central	0.12	0.0010
Lower layer central	0.13	0.0028
Upper layer mean	0.17	0.0022
Lower layer mean	0.11	0.0013

The true errors in the estimated stream functions at $t = 60$. Case 2A results are from a similar experiment by *Evensen* [1992], in which the extended Kalman filter was used with simplified error covariance evolution. Central and mean refer to the central forecast and mean forecast in the data assimilation experiment. MSE is mean square error. Note that there are some small differences in parameters between case 2A and the current experiment so case 2A is not exactly comparable to the current experiment.

Appendix

Generating Pseudo Random Fields

A method for generating two-dimensional pseudo random fields $q = q(x, y)$ with a specific mean, variance, and covariance is illustrated. A continuous field $q = q(x, y)$ can be described by its Fourier transform

$$q(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} d\mathbf{k}\, \widehat{q}(\mathbf{k})\, e^{i\mathbf{k}\cdot\mathbf{x}}. \qquad \text{(A1)}$$

By using a grid with dimensions N and M and defining $\mathbf{k} = (\kappa_l, \lambda_p)$, the expression above can be written as

$$q(x_n, y_m) = \sum_{l,p} \widehat{q}(\kappa_l, \lambda_p)\, \Delta\mathbf{k}\, e^{i(\kappa_l x_n + \lambda_p y_m)} \qquad \text{(A2)}$$

where $x_n = n\Delta x$, $y_m = m\Delta y$, $\kappa_l = \frac{2\pi l}{N\Delta x}$, $\lambda_p = \frac{2\pi p}{M\Delta y}$, and $\Delta\mathbf{k} = \Delta\kappa\Delta\lambda = \frac{(2\pi)^2}{NM\Delta x\Delta y}$.

Wave Amplitudes

Define the function

$$\widehat{q}(\kappa_l, \lambda_p) = \frac{c}{\sqrt{\Delta\mathbf{k}}} e^{-(\kappa_l^2+\lambda_p^2)/\sigma^2} e^{2\pi i\phi_{l,p}} \qquad \text{(A3)}$$

where $\phi_{l,p} \in [0,1]$ is a random number that introduces a random phase shift. The exponential function means that the frequency spectrum is decreasing exponentially with increasing wave number. The formula for the pseudo random fields is then

$$q(x_n, y_m) = \Delta \mathbf{k} \sum_{l,p} \frac{c}{\sqrt{\Delta \mathbf{k}}} e^{-(\kappa_l^2+\lambda_p^2)/\sigma^2} e^{2\pi i \phi_{l,p}} e^{i(\kappa_l x_n + \lambda_p y_m)}. \quad \text{(A4)}$$

To ensure that the pseudo random fields are real, the condition

$$\widehat{q}(\kappa_l, \lambda_p) = \widehat{q}(\kappa_{-l}, \lambda_{-p})^* \quad \text{(A5)}$$

or, equivalently, $\phi_{-l,-p} = -\phi_{l,p}$ must be satisfied. In addition we must have $\operatorname{Im} \widehat{q}(\kappa_0, \lambda_0) = 0$.

The formula (A4) can be used to generate an ensemble of pseudo random fields with a specific covariance determined by the choice of parameters c and σ. The covariance of the ensemble is

$$\overline{q(x_1,y_1)\, q(x_2,y_2)} = \sum_{l,p,r,s} (\Delta \mathbf{k})^2 \widehat{q}(\kappa_l, \lambda_p)\, \widehat{q}(\kappa_r, \lambda_s)^* e^{i(\kappa_l x_1 - \kappa_r x_2 + \lambda_p y_1 - \lambda_s y_2)} \quad \text{(A6)}$$

or

$$\overline{q(x_1,y_1)\, q(x_2,y_2)} = \sum_{l,p,r,s} \Delta \mathbf{k} c^2 e^{-(\kappa_l^2+\lambda_p^2+\kappa_r^2+\lambda_s^2)/\sigma^2} e^{2\pi i(\phi_{l,p}-\phi_{r,s})} e^{i(\kappa_l x_1 - \kappa_r x_2 + \lambda_p y_1 - \lambda_s y_2)}. \quad \text{(A7)}$$

The fields should be δ-correlated in wave space, so that

$$\overline{q(x_1,y_1)\, q(x_2,y_2)} = \sum_{l,p} \Delta \mathbf{k} c^2 e^{-2(\kappa_l^2+\lambda_p^2)/\sigma^2} e^0 e^{i(\kappa_l(x_1-x_2)+\lambda_p(y_1-y_2))}. \quad \text{(A8)}$$

The variance in a location (x, y) is

$$\overline{q(x,y)\, q(x,y)} = \Delta \mathbf{k}\, c^2 \sum_{l,p} e^{-2(\kappa_l^2+\lambda_p^2)/\sigma^2}. \quad \text{(A9)}$$

By requiring that the variance be equal to 1 in (A9) and using the condition $\overline{q(0)\, q(r_h)} = e^{-1}$ in (A8), one can solve for the two unknowns c and σ. If one applies the isotropy imposed when assuming δ-correlated fields, this reduces to solving, for example,

$$1 = \Delta \mathbf{k}\, c^2 \sum_{l,p} e^{-2(\kappa_l^2+\lambda_p^2)/\sigma^2}, \quad \text{(A10)}$$

$$e^{-1} = \Delta \mathbf{k} c^2 \sum_{l,p} e^{-2(\kappa_l^2+\lambda_p^2)/\sigma^2} \cos(\kappa_l r_h). \quad \text{(A11)}$$

for c and σ. Note that c can be eliminated, and the problem reduces to a nonlinear scalar equation for σ, which can be solved using a numerical algorithm. Here a routine, "fmm/zeroin.f" from Netlib, which contains a modified implementation of the algorithm "zero" given by *Brent* [1973] is used.

This generates an ensemble of pseudo random fields with a variance of 1 and covariance determined by the decorrelation length r_h used in the isotropic covariance function

$$\overline{q(0)\, q(r)} = \exp\left(-\frac{r^2}{r_h^2}\right). \quad \text{(A12)}$$

An extremely efficient method for generating pseudo random fields is provided by using two-dimensional fast Fourier transform routines. It would not be practical to use formula (A4) directly for larger computational grids and realistic applications. To avoid using doubly periodic fields, the inverse FFT is calculated on a grid which is a few characteristic lengths larger than the computational domain. The subdomain corresponding to the computational domain will then have nonperiodic fields.

Imposing Vertical Correlation

In the QG model there will always be a specific vertical correlation between the stream functions in the different layers. Such a correlation should also be contained in the ensemble of initial states and system noise fields. Using the procedure outlined above, one generates horizontal pseudo random fields with variance equal to 1 and with zero correlation between the layers. Given a set of pseudo random fields $q_i(x, y, z_1)$ and $q_i(x, y, z_2)$ for the upper and lower layers in the two-layer model that has been used here, a new set $q_i'(x, y, z_1)$ and $q_i'(x, y, z_2)$ with a specific covariance α between the layers can be generated by the linear combination

$$q'(x,y,z_1) = \sqrt{1-\alpha} q(x,y,z_1) + \alpha q(x,y,z_2) \quad \text{(A13)}$$

$$q'(x,y,z_2) = q(x,y,z_2). \quad \text{(A14)}$$

Such equations are also easy to derive for a higher number of layers by assuming linear combinations between the new and old fields in the different layers and then using the conditions of specified variances in and covariances between the different layers.

Acknowledgments. The author would like to thank R. N. Miller and A. F. Bennett for interesting discussions during part of this work. Thanks are also given to P. J. van Leeuwen for comments on the manuscript. The work was supported by the Research Council of Norway and has received support from the Norwegian Super Computing Committee (TRU) through a grant of computing time.

References

Brent, R. P., *Algorithms for Minimization Without Derivatives*, Prentice-Hall, Englewood Cliffs, N. J., 1973.

Budgell, W. P., Nonlinear data assimilation for shallow water equations in branched channels, *J. Geophys. Res.*, *91*(C9), 10,633–10,644, 1986.

Cohn, S. E., Dynamics of short term univariate forecast error covariances, *Mon. Weather Rev.*, *121*, 3123–3149, 1993.

Dee, D. P., Simplification of the Kalman filter for meteorological data assimilation, *Q. J. R. Meteorol. Soc.*, *117*, 365–384, 1991.

Delcroix, J. L., *Plasma Physics*, vol. 1, John Wiley, New York, 1968.

Epstein, E. S., Stochastic dynamic prediction, *Tellus, Ser. A*, *21*, 739–759, 1969.

Epstein, E. S., and E. J. Pitcher, Stochastic analysis of meteorological fields, *J. Atmos. Sci.*, *29*, 244–257, 1972.

Evensen, G., Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model, *J. Geophys. Res.*, *97*(C11), 17,905–17,924, 1992.

Evensen, G., Open boundary conditions for the extended Kalman filter with a quasi-geostrophic model, *J. Geophys. Res.*, *98*(C9), 16,529–16,546, 1993.

Fleming, R. J., On stochastic dynamic prediction, I, The energetics of uncertainty and the question of closure, *Mon. Weather Rev.*, *99*, 851–872, 1971a.

Fleming, R. J., On stochastic dynamic prediction, II, Predictability and utility, *Mon. Weather Rev.*, *99*, 927–938, 1971b.

Ghil, M., Meteorological data assimilation for oceanographers, I, Description and theoretical framework, *Dyn. Atmos. Oceans*, *13*(3-4), 171–218, 1989.

Gleeson, T. A., Statistical-dynamical predictions, *J. Appl. Met.*, *9*, 333–344, 1970.

Haugan, P. M., G. Evensen, J. A. Johannessen, O. M. Johannessen, and L. Pettersson, Modeled and observed mesoscale circulation and wave-current refraction during the 1988 Norwegian continental shelf experiment, *J. Geophys. Res.*, *96*(C6), 10,487–10,506, 1991.

Hoffman, R. N., and E. Kalnay, Lagged average forecasting, an alternative to Monte Carlo forecasting, *Tellus, Ser. A*, *35*, 100–118, 1983.

Jazwinski, A. H., *Stochastic Processes and Filtering Theory*, Academic, San Diego, Calif., 1970.

Lacarra, J.-F., and O. Talagrand, Short-range evolution of small perturbations in a barotropic model, *Tellus, Ser. A*, *40*, 81–95, 1988.

Leith, C. E., Atmospheric predictability and two-dimensional turbulence, *J. Atmos. Sci.*, *28*, 145–161, 1971.

Leith, C. E., Theoretical skill of Monte Carlo forecasts, *Mon. Weather Rev.*, *102*, 409–418, 1974.

Leith, C. E., and R. H. Kraichnan, Predictability of turbulent flows, *J. Atmos. Sci.*, *29*, 1041–1058, 1972.

Miller, R. N., Perspectives on advanced data assimilation in strongly nonlinear systems, *Proc. NATO ARW, Liegé, Belgium May 1993*, in press 1994.

Miller, R. N., M. Ghil, and F. Gauthiez, Advanced data assimilation in strongly nonlinear dynamical systems, *J. Atmos. Sci.*, *51*, 1037–1056, 1994.

Pedlosky, J., *Geophysical Fluid Dynamics*, 2nd ed., Springer–Verlag, New York, 1987.

Pitcher, E. J., Application of stochastic dynamic prediction to real data, *J. Atmos. Sci.*, *24*, 3–21, 1977.

Salmon, R., G. Holloway, and M. C. Hendershott, The equilibrium statistical mechanics of simple quasi-geostrophic models, *J. Fluid. Mech.*, *75*(Part 4), 691–703, 1976.

Schubert, S., M. Suarez, J.-K. Schemm, and E. Epstein, Dynamically stratified Monte Carlo forecasting, *Mon. Weather Rev.*, *120*, 1077–1088, 1992.

Schubert, S. D., and M. Suarez, Dynamical predictability in a simple general circulation model: Average error growth, *J. Atmos. Sci.*, *46*, 353–370, 1989.

Seidman, A. N., Averaging techniques in long-range weather forecasting, *Mon. Weather Rev.*, *109*, 1367–1379, 1981.

Geir Evensen, Nansen Environmental and Remote Sensing Center, Edvard Griegsvei 3a, N–5037 Solheimsviken/Bergen, Norway. (e-mail: geir@fram.nrsc.no)

(Received August 12, 1993; revised December 30, 1993; accepted January 21, 1994.)