Inverse Methods and Data Assimilation in Nonlinear Ocean Models

Geir Evensen
Nansen Environmental and Remote Sensing Center, Bergen, Norway

Physica D, in press

July 6, 1994

Keywords:
Inverse Methods, Data Assimilation, Kalman Filter, Representer Method, Simulated Annealing

Abstract

An overview is given of the current status of inverse methods and data assimilation for nonlinear ocean models. The inverse theory for time dependent dynamical models is formulated and the most promising solution methods like simulated annealing, the representer method, and sequential methods based on Monte Carlo simulations, are discussed with special focus on applications with nonlinear dynamics. A rather general "model independent" presentation has been used to make the methodology more accessible for different scientific areas dealing with dynamical models and data.

phone: +47 55 29 72 88
fax: +47 55 20 00 50
e-mail: geir@fram.nrsc.no

1 Introduction

Over the last decade, an increasing interest in studies of the general ocean circulation and its role in climate processes has motivated the use of inverse methods and data assimilation for improving the knowledge of the ocean. Such methods make it possible to combine the information about the true state which is contained in a set of measurements with an ocean model holding information about the important dynamical processes in the ocean. Inverse methods can be considered as an approach for interpolating or smoothing a data set in space and time where a model acts as a dynamical constraint. This makes it possible to find a more realistic estimate for the true state of the ocean than what can be found from data or model alone. A set of measurements is normally sparse in space and time and does not resolve all the physical time and space scales of interest. A dynamical model only contains information about the interaction between physical processes which are included in the model equations, and the solutions are in general determined from poorly known initial and boundary-conditions and forcing fields. To find a best possible estimate of the true state of the physical system it is necessary to use all available information from both model and measurements in an integrated system.

In meteorology data assimilation has been used for some decades in weather prediction, although with rather simplistic schemes. Both in oceanography and meteorology several contributions have studied applications of varying formulations and methodologies with linear models. Some reviews are given by Lorenc (1986), Ghil (1989), and Ghil and Malanotte-Rizzoli (1991). For nonlinear models the main effort has been focussed on simplistic approaches where the effects of nonlinearities are neglected or approximated. It has become evident that the extension of inverse methods for linear models to fully nonlinear dynamics is nontrivial. Methods developed for linear models have in many cases been applied to nonlinear dynamics by performing certain linearizations, e.g. in the equation for error statistics when using the Kalman filter, or in the adjoint equation when solving the Euler–Lagrange equations.

The significantly smaller spatial scales of the dominant physics in the ocean compared to the atmosphere make the inverse problem for ocean models very challenging. The "weather" in the ocean, i.e., the mesoscale circulation has a spatial scale of typically 10 km, compared to about 1000 km in the atmosphere. Most ocean models are regional because of the higher resolution which is required and one must deal with the ill-posedness of open boundaries in addition to the sparseness of traditional data-sets which poorly resolve the physical scales. The problem of observability is, however, slightly reduced because of the longer time scales in the ocean.

The appearance of new remotely sensed data-sets from several of the environmental monitoring satellites further motivates the use of inverse methods with nonlinear ocean circulation models. Previous inverse calculations for the ocean have mainly been concerned with the smoothing of a specific data set in a regional model and for a limited time period. For operational ocean monitoring and forecasting one will have to rely on remote sensing information.

The most promising remotely sensed information for use in oceanographic data assimilation is the altimeter data collected by satellites such as GEOSAT, TOPEX/POSEIDON and ERS–1, which measure the sea-surface height and thereby provide estimates of the surface geostrophic ocean circulation and its variability. Infrared sensors (IR), e.g. the AVHRR (Advanced Very High Resolution Radiometer) on board the NOAA satellite, or the ATSR (Along Track Scanning Radiometer) carried by ERS–1, measure the sea surface temperature which can be assimilated in ocean models containing thermodynamics, and track frontal positions that can be assimilated in dynamical models. The scatterometers on ERS–1 and Seasat measure the wind-speed and direction at the surface and can be used to improve estimates of the forcing fields which drive the ocean models. Another promising sensor is the SAR (Synthetic Aperture Radar), e.g. from ERS–1 and Seasat, which measures backscatter from the ocean surface. It has proved extremely efficient for ice monitoring (Johannessen *et al.*, 1992), and also provides information about

fronts and structures in the open ocean.

A discussion on the application of remote sensing data for ocean monitoring and modeling is given by Johannessen *et al.* (1993). An important problem with remote sensing sensors is the fact that they only collect information from the sea surface or the upper part of the ocean, while the deep ocean must still be examined by *in situ* observations. A combination of both remote sensing and *in situ* data used together with advanced data assimilation schemes would seem necessary for reconstructing the baroclinic modes and the circulation in the deep ocean in an operational ocean monitoring system.

Even though this work consider the data assimilation problem for nonlinear ocean models, the basic formalism will be identical for many other research areas. Inverse methods and data assimilation have also been used extensively in geophysics, see e.g. Tarantola (1987) and Menke (1984). For an comprehensive introduction to data assimilation and inverse methods in oceanography, see Bennett (1992).

2 Basic Inverse Formulation

The inverse problem can be formulated by allowing the model, the initial and boundary conditions, and the measurements to contain errors. The number of conditions to be fulfilled is increased by adding the data constraints, (the system becomes over-determined), and one then increases the order of the system by including the new unknown error terms. The formulation of the inverse can be illustrated using a nonlinear model on a spatial domain $\mathcal{D}$ with boundary conditions on the boundary $\delta\mathcal{D}$ and initial conditions at $t = 0$, plus a data set,

$$\frac{\partial \psi}{\partial t} = N[\psi] + F + q, \qquad \mathbf{x} \in \mathcal{D},\ 0 \le t \le T, \tag{1}$$

$$\psi = \Psi_0 + a, \qquad \mathbf{x} \in \mathcal{D},\ t = 0, \tag{2}$$

$$\psi = \Psi_B + b, \qquad \mathbf{x} \in \delta\mathcal{D},\ 0 \le t \le T, \tag{3}$$

$$\mathbf{d} = \mathcal{M}[\psi] + \boldsymbol{\epsilon}. \tag{4}$$

Here ψ is the unknown state variable, e.g., a stream function, $N[\,]$ is a nonlinear operator, $F(\mathbf{x}, t)$ is a forcing field, $\mathcal{M}[\,]$ is a vector of linear measurement functionals that relates the measurements $\mathbf{d}$ to the state variable, $\Psi_0(\mathbf{x})$ and $\Psi_B(s(\mathbf{x}), t)$ are the first guess initial and boundary conditions, $s(\mathbf{x})$ denotes the boundary, and $q(\mathbf{x}, t)$, $a(\mathbf{x})$, $b(s(\mathbf{x}), t)$, and $\boldsymbol{\epsilon}$ represent the errors in the model, initial and boundary conditions, and measurements. The extension to the case where ψ is a vector of state variables is straight forward.

In an inverse formulation one attempts to find the state estimate which minimizes the error terms q, a, b, and $\boldsymbol{\epsilon}$, given an appropriate penalty function which in addition includes constraints on realistic amplitudes and smoothness dictated by the true physical scales of the system. An example of such a penalty function is

$$\begin{aligned}
\mathcal{N}[\psi] = &\int_0^T dt_1 \int_0^T dt_2 \int_{\mathcal{D}} d\mathbf{x}_1 \int_{\mathcal{D}} d\mathbf{x}_2\, q(\mathbf{x}_1, t_1) W_{qq}(\mathbf{x}_1, t_1, \mathbf{x}_2, t_2) q(\mathbf{x}_2, t_2) \\
&+ \int_{\mathcal{D}} d\mathbf{x}_1 \int_{\mathcal{D}} d\mathbf{x}_2\, a(\mathbf{x}_1) W_{aa}(\mathbf{x}_1, \mathbf{x}_2) a(\mathbf{x}_2) \\
&+ \int_0^T dt_1 \int_0^T dt_2 \int_{\delta\mathcal{D}} ds_1 \int_{\delta\mathcal{D}} ds_2\ b(s_1, t_1) W_{bb}(s_1, t_1, s_2, t_2) b(s_2, t_2) \\
&+ \boldsymbol{\epsilon}^* \mathbf{w} \boldsymbol{\epsilon},
\end{aligned} \tag{5}$$

where $*$ indicates transpose, and the weights W_{qq}, W_{aa}, W_{bb}, and $\mathbf{w}$ are symmetrical and positive definite. In an obvious simpler notation, (5) can be written as

$$\mathcal{N} = q \circ \bullet W_{qq} \circ \bullet q + a \circ W_{aa} \circ a + b \circ \bullet W_{bb} \circ \bullet b + \boldsymbol{\epsilon}^* \mathbf{w} \boldsymbol{\epsilon}, \tag{6}$$

where $\bullet$ means integration in time, and $\circ$ means integration in space. The weights W_{qq}, W_{aa}, and W_{bb} are functional inverses of the error covariances Q_{qq}, Q_{aa}, and Q_{bb} for the errors in the model, and initial and boundary conditions, e.g.,

$$\int_0^T dt_2 \int_{\mathcal{D}} d\mathbf{x}_2 W_{qq}(\mathbf{x}_1, t_1, \mathbf{x}_2, t_2) Q_{qq}(\mathbf{x}_2, t_2, \mathbf{x}_3, t_3) = \delta(t_1 - t_3)\delta(\mathbf{x}_1 - \mathbf{x}_3). \tag{7}$$

They determine the spatial and temporal scales for the physical problem, and ensure smooth influences from the measurements. The data weight $\mathbf{w}$ is the inverse of the error covariance matrix $\mathbf{w}^{-1}$ for the measurements.

Such a "weak constraint" formalism, where the model is allowed to contain errors, was introduced by Sasaki (1958) for an atmospheric model. The minimizing solution will be a smooth field which is close to the measurements without interpolating them and at the same time it will be close to the dynamics imposed by the model.

Most minimization methods are based upon the system of Euler–Lagrange equations which can be found by substituting for $q(\mathbf{x}, t)$, $a(\mathbf{x})$, $b(\mathbf{x}, t)$ and $\boldsymbol{\epsilon}$ from (1–4) in (5) and taking the variation of the penalty function $\mathcal{N}[\psi]$ (Courant and Hilbert, 1953). The system of Euler–Lagrange equations then becomes,

$$\frac{\partial \psi}{\partial t} = N[\psi] + F + Q_{qq} \circ \bullet \lambda, \qquad \mathbf{x} \in \mathcal{D},\ 0 \le t \le T, \tag{8}$$

$$\psi = \Psi_0 + Q_{aa} \circ \lambda, \qquad \mathbf{x} \in \mathcal{D},\ t = 0, \tag{9}$$

$$-\frac{\partial \lambda}{\partial t} = N'^{\dagger}_{\psi} \lambda + \mathcal{M}^*[\delta(\mathbf{x} - X_i)\delta(t - T_i)]\mathbf{w}(\mathbf{d} - \mathcal{M}[\psi]), \qquad \mathbf{x} \in \mathcal{D},\ 0 \le t \le T, \tag{10}$$

$$\lambda = 0, \qquad \mathbf{x} \in \mathcal{D},\ t = T, \tag{11}$$

where $N'^{\dagger}_{\psi}$ is the adjoint of the tangent linear operator of N evaluated in $\psi(\mathbf{x}, t)$, and $\lambda(\mathbf{x}, t)$ is the adjoint variable defined by equation (8). Equation (8) is the dynamical model driven by a term which estimates the model error and contains the adjoint variable, and is integrated forward in time from the initial condition (9). A similar term is also included in the initial condition. The so called adjoint equation (10), which is forced by the data misfits at the space and time locations X_i, T_i where data are available, is integrated backward in time from a final condition (11). The system of Euler–Lagrange equations comprises a two-point boundary value problem in space and time, and the forward and backward equations (8) and (10) are coupled and must be solved simultaneously.

Boundary conditions for the system will be dependent on the model operator $N[\,]$. This can be illustrated using the Burgers' equation on a one dimensional domain $0 \le x \le L$. The model operator, the tangent linear operator, and the adjoint of the tangent linear operator become respectively

$$N[\psi] = -\psi \frac{\partial \psi}{\partial x} + \gamma \frac{\partial^2 \psi}{\partial x^2}, \tag{12}$$

$$N'_{\psi}[\,] = -\left(\frac{\partial \psi}{\partial x}\right) - \psi \frac{\partial}{\partial x} + \gamma \frac{\partial^2}{\partial x^2}, \tag{13}$$

$$N'^{\dagger}_{\psi}[\,] = \psi \frac{\partial}{\partial x} + \gamma \frac{\partial^2}{\partial x^2}, \tag{14}$$

where γ is a diffusion coefficient. Burgers' equation was originally proposed by Burgers to model turbulence, and it was later shown that it can be derived from the Navier–Stokes equations in the limit of a weak shock layer. An analytical solution is given in Kevorkian and Cole (1981). For this problem the boundary conditions (3) for the forward model can be specified as

$$\psi(0,t) = \Psi_B(0,t) + b(0,t), \tag{15}$$

$$\psi(L,t) = \Psi_B(L,t) + b(L,t). \tag{16}$$

The Euler–Lagrange equations are still (8–11), while boundary conditions for the inverse become

$$\psi = \Psi_B + Q_{bb} \bullet \left(\lambda\psi + \gamma\frac{\partial\lambda}{\partial x}\right), \qquad \text{on } x = 0 \text{ for } t \in [0,T], \tag{17}$$

$$\psi = \Psi_B - Q_{bb} \bullet \left(\lambda\psi + \gamma\frac{\partial\lambda}{\partial x}\right), \qquad \text{on } x = L \text{ for } t \in [0,T], \tag{18}$$

for the forward model, and

$$\lambda = 0, \qquad \text{on } x = 0 \text{ for } t \in [0,T], \tag{19}$$

$$\lambda = 0, \qquad \text{on } x = L \text{ for } t \in [0,T], \tag{20}$$

for the adjoint model.

There are two other important points to note here. First, by penalizing the residual in the initial and boundary conditions one ensures uniqueness of the minimizing solution (Bennett and Miller, 1990). If the data weights are zero, i.e. worthless data, the unique solution will be the first guess solution found by solving the forward problem (8) subject to initial and boundary conditions with $\lambda \equiv 0$. On the other hand, if the initial and boundary weights are zero then there are no constraints on the first guess initial and boundary-conditions. The estimated initial and boundary-conditions will then take nonphysical amplitudes and become noisy, and the solution will interpolate the data exactly. Actually, there might be an infinite number of solutions which interpolate the data and give a penalty function equal to zero. Secondly, the use of non-diagonal covariances Q ensures a smooth estimate and has a regularization effect on the solution. The covariance functions should reflect the true physical scales of the system. Diagonal weights would allow noisy initial and boundary conditions and hence noisy estimates.

3 Solution Methods

There are mainly two approaches for solving the inverse problem. One can solve the full generalized inverse and smooth the solution over the data in space and time, i.e. the smoothing problem. Another approach is to use sequential data assimilation algorithms where the data are assimilated during a forward integration of a model and its error statistics, and the solution improves during the integration owing to the increasing amount of accumulated information from measurements. The sequential methods can be derived from the generalized inverse formulation and may be viewed as sub-optimal solution methods for the generalized inverse.

The smoothing problem is convenient when treating a specific data set which is limited in time, say from a field experiment, while the sequential approach is well suited for forecasting purposes. One would assume that sequential methods are less numerically demanding than the smoothing approach, but this is not necessarily true and there are also cases where one would prefer a sequential approach for a time limited data set or a generalized inverse solver for forecasting purposes.

3.1 Generalized Inverse Solvers

Two completely different methods for solving the generalized inverse or the smoothing problem will be illustrated here. The first method is called simulated annealing and is a purely statistical method. It is especially well suited for strongly nonlinear problems where the penalty function often contains multiple local minima and where methods like gradient descent are likely to fail. The second method is the so called representer method, which is the optimal method for solving the generalized inverse for a linear model. It can in some cases also be applied for nonlinear dynamics using an iteration technique. The method is used to decouple the Euler–Lagrange equations and reduces the two point boundary value problem in time to a sequence of initial value problems.

3.1.1 Simulated Annealing

The functional (6) can be minimized directly using simulated annealing (Kirkpatric *et al.*, 1983, Azencott, 1992). The method is extremely simple in its basic formulation and can be illustrated using an example where a penalty function $\mathcal{N}[\psi]$, which may be nonlinear and discontinuous, is to be minimized with respect to the variable ψ:

ψ first guess
for $i = 1 : \ldots$
  $\psi_1 = \psi + \Delta\psi$
  if $(\mathcal{N}[\psi_1] < \mathcal{N}[\psi])$ **then**
    $\psi = \psi_1$
  else
    $\xi \in [0, 1]$ random number
    $p = \exp\{(\mathcal{N}[\psi] - \mathcal{N}[\psi_1])/\theta\} \in [0, 1]$
    if $p > \xi$ **then** $\psi = \psi_1$
  end
  $\theta = f(\theta, i, \mathcal{N}_{min})$
end

Here $\Delta\psi$ might be a normal distributed random number with mean zero and a variance specified to suit the nature of the penalty function. The temperature scheme $\theta = f(\theta, i, \mathcal{N}_{min})$ is used to cool or relax the system and is normally a decreasing function of iteration i. The method has the ability of "climbing" out of local minima since there is a finite probability of accepting a value of ψ that increases the value of the penalty function. It is easily extended to higher dimensional problems, and has been used to invert an inverse for a primitive equation model on a domain with open boundaries by Bennett and Chua (1994). In Barth and Wunsch (1990), the method was used to optimize a data collection scheme.

The conditional uphill climb based on the value of p and ξ is due to Metropolis *et al.* (1953), and is named the Metropolis algorithm. They also gave a proof that the method was ergodic, i.e., that any state can be reached from any other, and that the trials would converge toward a distribution proportional to

$$p_S = \exp(-\mathcal{N}[\psi]/\theta), \tag{21}$$

where again θ is the temperature. By slowly decreasing the value of θ the distribution will approach the delta function at the minimizing value of ψ. The clue then is to choose a temperature scheme where one avoids getting trapped in local minima for too many iterations, or where too many uphill climbs are accepted. In Bohachevsky *et al.* (1986) it was suggested that the temperature should be chosen so

that $p \in [0.5, 0.9]$. Here also a generalized algorithm was proposed where p was calculated according to $p = \exp\{\beta(\mathcal{N}[\psi] - \mathcal{N}[\psi_1])/(\mathcal{N}[\psi] - \mathcal{N}_{min})\}$, where β is approximately 3.5 and $\mathcal{N}_{min}$ is an estimate of the normally unknown minima of the penalty function. Then the probability of accepting a detrimental step tend to zero as the random walk approaches the global minimum. If a value of the cost function is found which is less than $\mathcal{N}_{min}$ this value will replace $\mathcal{N}_{min}$.

The method is also extremely sensitive to the choice of the perturbation $\Delta\psi$. In a higher dimensional problem with a vector state, $\boldsymbol{\psi}$, all components could be perturbed simultaneously by constructing a vector $\mathbf{r}$ of length one where the components are given by $r_i = s_i/||\mathbf{s}||_2$. Here the components s_i are taken from a normal distribution and $||\mathbf{s}||_2$ is the length of $\mathbf{s}$ in a 2-norm. The vector of perturbations then becomes $\Delta\boldsymbol{\psi} = \alpha\mathbf{r}$ where α is a scale factor.

In Bennett and Chua (1994) an alternative to a random walk, which provided a significantly faster convergence, was used for calculating the inverse of a nonlinear open ocean shallow water model. The algorithm is due to Duane *et al.* (1987), and it is based on constructing the Hamiltonian

$$\mathcal{H}[\boldsymbol{\psi}, \boldsymbol{\pi}] = \mathcal{N}[\boldsymbol{\psi}] + \frac{1}{2}\boldsymbol{\pi}^*\boldsymbol{\pi}, \tag{22}$$

and then deriving the canonical equations of motion in $(\boldsymbol{\psi}, \boldsymbol{\pi})$ phase space, with respect to a pseudo time variable τ,

$$\frac{\partial\psi_i}{\partial\tau} = \frac{\partial\mathcal{H}}{\partial\pi_i} = \pi_i, \tag{23}$$

$$\frac{\partial\pi_i}{\partial\tau} = -\frac{\partial\mathcal{H}}{\partial\psi_i} = -\frac{\partial\mathcal{N}}{\partial\psi_i}. \tag{24}$$

This system is integrated for a pseudo time interval, $\tau \in [0, \tau_1]$, using the previously accepted value of $\boldsymbol{\psi}$ and a random guess for $\boldsymbol{\pi}(0)$ as initial conditions. The Metropolis algorithm can then be used for the new guess $\boldsymbol{\psi}(\tau_1)$. In Duane *et al.* (1987), it was proved that this algorithm also preserved detailed balance, i.e.,

$$p_S(\boldsymbol{\psi}_1)p_M(\boldsymbol{\psi}_1 \to \boldsymbol{\psi}_2) = p_S(\boldsymbol{\psi}_2)p_M(\boldsymbol{\psi}_2 \to \boldsymbol{\psi}_1), \tag{25}$$

where p_M is the transition probability function for the Markov process where the canonical equations are integrated from a random initial momentum. Detailed balance is essential for showing that a long sequence of random trials will converge toward the distribution (21).

After a minimum of the variational problem has been found, the posterior error statistics can be estimated by raising the annealing temperature and collecting samples of nearby states (Bennett and Chua, 1994).

Such a substitution algorithm is extremely computationally expensive. However, it can handle strongly nonlinear problems, and it also suits the SIMD (Single Instruction Multiple Data) massively parallel computer architectures, e.g. the Thinking Machines CM–200 or the MasPar MP–2. When using massively parallel computers, each processor will integrate the canonical equations for the state variable in its respective grid point, and the estimate will be perturbed simultaneously all over the grid.

Using simulated annealing a solution is found without integrating any of the original dynamical equations, and the method is therefore convenient for problems where the inverse is well-posed while the actual integration of the model equations may be ill-posed, (Bennett and Chua, 1994). The method is promising, but a lot of work still has to be done in tuning the parameters in the iterations.

3.1.2 Representer method

For linear models it can be shown that a unique solution of the inverse problem is given by a first guess field plus a linear combination of influence functions, one for each of the measurements. For illustrational purposes the following linear system of Euler–Lagrange equations is considered:

$$\frac{\partial \psi}{\partial t} - L\psi = F(\mathbf{x},t) + Q_{qq} \circ \bullet\lambda, \tag{26}$$

$$\psi = \Psi_0 + Q_{aa} \circ \lambda \qquad t = 0, \tag{27}$$

$$\frac{\partial \lambda}{\partial t} + L^\dagger \lambda = -\mathcal{M}[\delta(\mathbf{x} - X_i)\delta(t - T_i)]\mathbf{w}(\mathbf{d} - \mathcal{M}[\psi]) \tag{28}$$

$$\lambda = 0, \qquad t = T, \tag{29}$$

where periodic boundary conditions have been assumed and L is now a linear operator. If the first guess solution, $\psi_F(\mathbf{x},t)$, is defined by integrating equation (26) subject to (27) and with $\lambda = 0$, a solution of the coupled Euler–Lagrange equations can be found using the expansion

$$\psi(\mathbf{x},t) = \psi_F(\mathbf{x},t) + \mathbf{p}^*\mathbf{r}(\mathbf{x},t), \tag{30}$$

i.e., the solution is expressed as the first guess field plus a linear combination of representer fields $\mathbf{r}(\mathbf{x},t)$, one for each measurement. This reduces the problem to a sequence of initial value problems for each of the representer fields and their adjoints $\boldsymbol{\alpha}(\mathbf{x},t)$,

$$\frac{\partial \mathbf{r}}{\partial t} - L\mathbf{r} = Q_{qq} \circ \bullet\boldsymbol{\alpha}, \tag{31}$$

$$\mathbf{r} = Q_{aa} \circ \boldsymbol{\alpha}, \qquad t = 0, \tag{32}$$

$$\frac{\partial \boldsymbol{\alpha}}{\partial t} + L^\dagger \boldsymbol{\alpha} = -\mathcal{M}^*[\delta(\mathbf{x} - X_i)\delta(t - T_i)], \tag{33}$$

$$\boldsymbol{\alpha} = 0, \qquad t = T, \tag{34}$$

where the coupling to ψ in (28) has been removed. The coefficients $\mathbf{p}$ can be found by solving the system

$$\left(\mathcal{M}[\mathbf{r}^*] + \mathbf{w}^{-1}\right)\mathbf{p} = \mathbf{d} - \mathcal{M}[\psi_F]. \tag{35}$$

By solving the $2n_o$ initial value problems (31–34), where n_o is the number of measurements one can calculate $\mathbf{r}(\mathbf{x},t)$ and also construct the matrix $\mathcal{M}[\mathbf{r}^*] + \mathbf{w}^{-1}$ by operating on each of the representer fields with the measurement functional and then adding the measurement error covariance matrix.

Note that each of the influence functions or representer fields contained in $\mathbf{r}(\mathbf{x},t)$ can also be expressed by a measurement functional operating on the representer or space-time covariance function $\Gamma(\mathbf{x}_1,t_1,\mathbf{x}_2,t_2)$, i.e.,

$$\mathbf{r}(\mathbf{x},t) = \mathcal{M}_{(\mathbf{x}_2,t_2)}[\Gamma(\mathbf{x},t,\mathbf{x}_2,t_2)], \tag{36}$$

and note also the similarity to the objective analysis technique (Bretherton *et al.*, 1976, and McIntosh, 1990), which in this case has been extended to include the time dimension. By integrating the system (31–34) one avoids the actual calculation of the representer $\Gamma(\mathbf{x}_1,t_1,\mathbf{x}_2,t_2)$.

The posterior error covariance for the minimizing solution, (Bennett, 1992), is

$$C(\mathbf{x}_1,t_1,\mathbf{x}_2,t_2) = \Gamma(\mathbf{x}_1,t_1,\mathbf{x}_2,t_2) - \mathbf{r}(\mathbf{x}_1,t_1)^*\left(\mathcal{M}[\mathbf{r}^*] + \mathbf{w}^{-1}\right)^{-1}\mathbf{r}(\mathbf{x}_2,t_2). \tag{37}$$

However, the calculation of $\Gamma(\mathbf{x}_1,t_1,\mathbf{x}_2,t_2)$ is an extremely expensive task. Actually, for calculating

the error variance estimate $C(\mathbf{x}, t, \mathbf{x}, t)$, only the variance $\Gamma(\mathbf{x}, t, \mathbf{x}, t)$ is required and this is normally calculated by statistical simulation, a method which is described in Section 3.2.4 below.

For nonlinear dynamics, an expansion like (30) does not exist in general. However, it has been illustrated by Bennett and Thorburn (1992), and Bennett *et al.* (1993) that for a nonlinear quasi-geostrophic (QG) model, an upper bounded iteration can be constructed for the Euler–Lagrange equations if all the advecting velocities are evaluated at the previous iterate. It is then possible to decouple each of the linear two-point boundary value problems in the sequence of linear iterates using a representer expansion, and a unique solution can be found for each of the linear iterates.

The method is very well suited to parallel computers with the MIMD (Multiple Instruction Multiple Data) architecture, e.g. the Thinking Machines CM–5 or the Intel Paragon, where each of the representer fields can be calculated simultaneously in parallel on different processors (Bennett and Baugh, 1992).

The method has been applied to a meteorological problem where a barotropic nonlinear (QG) model was used to forecast several tropical cyclones (Bennett *et al.*, 1993), and it was found that the inverse with such a model generated forecasts with higher accuracy than the official meteorological forecasts.

There are also other alternative solution methods for the generalized inverse, e.g., the sweep algorithm or the Kalman smoother (Gelfand and Fomin, 1963, Bennett and Budgell, 1989, Moiseenko and Saenko, 1992), which again reduces the Euler–Lagrange equations to a number of initial-value problems, although the method is rather expensive because an equation for the spatial error covariances must be integrated. Further, for nonlinear dynamics the solutions will no longer be optimal because linearized equations are used to propagate the error statistics (Evensen, 1992).

For more advanced nonlinear primitive equation models, additional problems occur. First of all, it is not clear that the adjoint operator exists for all models. If it really exists, and can be found, it is still necessary to have a well-posed formulation for both the forward and backward model. Oceanographic models are often regional with open boundaries, and are often ill-posed in the sense that the boundary conditions are unknown and may be over or under-specified (Bennett and Chua, 1994). Note that the adjoint of the tangent linear operator is approximate, and in Evensen (1992) it was found that because of the linearization dynamical instabilities failed to saturate and it was not clear how this could be handled.

The problems arising from strongly nonlinear models suggest the use of direct substitution algorithms where one avoids the actual integration of the forward and backward model equations, e.g. algorithms based on simulated annealing or gradient descent.

3.2 Sequential Data Assimilation

A completely different approach is to apply sequential data assimilation algorithms. In that case the model is integrated forward in time and the model solution is updated whenever measurements are available. The updating scheme normally replaces the forecast by a linear combination of the predicted model state and a vector of measurements. The linear combination is chosen to minimize the error variance of the analyzed estimate in a least squares sense, and is based on estimates of the error statistics for the model forecast and the measurements. For linear models the optimal sequential algorithm is the Kalman filter (Kalman, 1960). It can be derived both from a statistical formulation (Gelb, 1974), and from the Euler–Lagrange equations (Bennett and Budgell, 1987).

3.2.1 Kalman Filter

The Kalman filter for a linear discrete model with periodic boundary conditions can be summerized as follows; the true state is assumed to evolve according to

$$\boldsymbol{\psi}_k^{\mathrm{t}} = L_k \boldsymbol{\psi}_{k-1}^{\mathrm{t}} + F_k + \mathbf{q}_k, \qquad \mathbf{x} \in \mathcal{D},\ 0 \le t \le T, \tag{38}$$

$$\boldsymbol{\psi}_0^{\mathrm{t}} = \boldsymbol{\Psi}_0 + \mathbf{a}, \qquad \mathbf{x} \in \mathcal{D},\ t = 0, \tag{39}$$

where $\mathbf{q}_k$ and $\mathbf{a}$ are vectors of unknown random white noise which represent the errors in the model and initial conditions, L_k is the discrete linear model operator, and F_k is the forcing at time t_k. A vector of measurements $\mathbf{d}_k$ at time t_k is related to the true state vector by the measurement equation

$$\mathbf{d}_k = M_k \boldsymbol{\psi}_k^{\mathrm{t}} + \boldsymbol{\epsilon}_k, \tag{40}$$

where M_k is the discrete form of $\mathcal{M}[\,]$ for the measurements at time t_k, i.e. a matrix operator. The superscripts a, f, and t denote analysis, forecast, and true respectively. The model forecast at time t_k is generated from the analyzed estimate at time t_{k-1} from

$$\boldsymbol{\psi}_k^{\mathrm{f}} = L_k \boldsymbol{\psi}_{k-1}^{\mathrm{a}} + F_k, \qquad \mathbf{x} \in \mathcal{D},\ 0 \le t \le T, \tag{41}$$

starting with the initial condition

$$\boldsymbol{\psi}_0^{\mathrm{f}} = \boldsymbol{\Psi}_0, \qquad \mathbf{x} \in \mathcal{D},\ t = 0. \tag{42}$$

The analyzed estimate is given by

$$\boldsymbol{\psi}_k^{\mathrm{a}} = \begin{cases} \boldsymbol{\psi}_k^{\mathrm{f}} + K_k(\mathbf{d}_k - M_k \boldsymbol{\psi}_k^{\mathrm{f}}), & \text{if } t_k = T_i, \\ \boldsymbol{\psi}_k^{\mathrm{f}}, & \text{if } t_k \neq T_i. \end{cases} \tag{43}$$

Further, the error statistics of the model forecast is described by the error covariance matrix $P_k^{\mathrm{f}} = \overline{(\boldsymbol{\psi}_k^{\mathrm{f}} - \boldsymbol{\psi}_k^{\mathrm{t}})(\boldsymbol{\psi}_k^{\mathrm{f}} - \boldsymbol{\psi}_k^{\mathrm{t}})^*}$, which can be calculated from

$$P_k^{\mathrm{f}} = L_{k-1} P_{k-1}^{\mathrm{a}} L_{k-1}^* + Q_{k-1}, \qquad \mathbf{x} \in \mathcal{D},\ 0 \le t \le T, \tag{44}$$

starting from the initial condition

$$P_0^{\mathrm{f}} = A, \qquad \mathbf{x} \in \mathcal{D},\ t = 0, \tag{45}$$

where $Q_k = \overline{\mathbf{q}_k \mathbf{q}_k^*}$ is the model error covariance matrix and $A = \overline{\mathbf{a}_k \mathbf{a}_k^*}$ is the error covariance matrix for the initial conditions. The analyzed error statistics is given by

$$P_k^{\mathrm{a}} = \begin{cases} (I - K_k M_k) P_k^{\mathrm{f}}, & \text{if } t_k = T_i, \\ P_k^{\mathrm{f}}, & \text{if } t_k \neq T_i. \end{cases} \tag{46}$$

The measurements at time t_k have error statistics described by the measurement error covariance matrix $\mathbf{w}_k^{-1} = \overline{\boldsymbol{\epsilon}_k \boldsymbol{\epsilon}_k^*}$. Given the error statistics for both the model forecast and the measurements, the optimal linear analysis is given by the Kalman gain

$$K_k = P_k^{\mathrm{f}} M_k^* \left(M_k P_k^{\mathrm{f}} M_k^* + \mathbf{w}_k^{-1} \right)^{-1}. \tag{47}$$

In summary the Kalman filter integrates forward in time both the ocean model and an equation for the error covariance matrix, and at every time step when measurements are available the model state and the error statistics are updated to reflect the information contained in the measurements. For a linear model it can be shown, (Bennett, 1992), that the Kalman filter provides an optimal solution at the end of the integration time where all the measurements have been assimilated, i.e. the solution equals the result found by solving the generalized inverse (5).

The Kalman filter was originally designed for engineering purposes, and it becomes extremely expensive to compute for multi-dimensional ocean models. If n is the number of state variables on the numerical grid, typically a large number of order 10 000 – 100 000, then the number of elements in the error covariance matrix P is n^2. Further, the computational cost is determined by $2n$ model integrations in (44). Despite the enormous numerical load, the optimality of the filter for linear dynamics has motivated an extensive study of the method for lower dimensional and simpler models, e.g, Miller (1986), Miller (1989), Miller and Cane (1989), Miller (1990), Bennett and Budgell (1987), Budgell (1987) for some oceanographic applications and Ghil *et al.* (1981), Cohn and Parrish (1991), Cohn (1993), Daley (1992d, 1992a, 1992c, 1992b, 1993), Todling and Ghil (1994), and Dee (1991) for some meteorological applications.

3.2.2 Comparison of methods

It is instructive to compare the Kalman filter with the representer method, and it is then convenient to rewrite the equation for the analyzed state (43) as

$$\boldsymbol{\psi}_k^{\mathrm{a}} = \boldsymbol{\psi}_k^{\mathrm{f}} + \mathbf{p}^* M_k P_k^{\mathrm{f}}, \tag{48}$$

where the optimal choice of the vector $\mathbf{p}_k$ is found by solving the system

$$(M_k P_k^{\mathrm{f}} M_k^* + \mathbf{w}_k^{-1})\mathbf{p}_k = (\mathbf{d}_k - M_k \boldsymbol{\psi}_k^{\mathrm{f}}). \tag{49}$$

Note the similarity with the representer expansion (30), (36) and (35). Here the analysis at time t_k is given by adding a linear combination of "representers" $M_k P_k$, one for each measurement to the forecast or "first guess" $\boldsymbol{\psi}_k^{\mathrm{f}}$.

The expression for the analyzed error covariance matrix (46) can be rewritten as

$$P_k^{\mathrm{a}} = P_k^{\mathrm{f}} - (M_k P_k^{\mathrm{f}})^* \left(M_k P_k^{\mathrm{f}} M_k^* + \mathbf{w}_k^{-1} \right)^{-1} M_k P_k^{\mathrm{f}}. \tag{50}$$

If (36) is used in (37) the similarity between the analyzed error covariance matrix (46) and the the posterior error estimate provided by the representer method is striking.

It is easy to show that the update (43) can be determined by minimizing a quadratic functional similar to (5), i.e.,

$$\mathcal{J}[\psi_k^{\mathrm{a}}] = (\boldsymbol{\psi}_k^{\mathrm{a}} - \boldsymbol{\psi}_k^{\mathrm{f}})^* \left(P_k^{\mathrm{f}} \right)^{-1} (\boldsymbol{\psi}_k^{\mathrm{a}} - \boldsymbol{\psi}_k^{\mathrm{f}}) + (\mathbf{d}_k - M_k \psi_k^{\mathrm{a}})^* \mathbf{w}_k (\mathbf{d}_k - M_k \psi_k^{\mathrm{a}}). \tag{51}$$

3.2.3 Extended Kalman Filter

For nonlinear dynamics only a few applications of the so-called extended Kalman filter exist, i.e. Budgell (1986), Evensen (1992, 1993), Miller *et al.* (1994), and Gauthier *et al.* (1993). The problems with a nonlinear model

$$\boldsymbol{\psi}_k^{\mathrm{f}} = \mathbf{f}(\boldsymbol{\psi}_{k-1}^{\mathrm{a}}), \tag{52}$$

is that the error covariance equation becomes approximate and unclosed of the form

$$\begin{aligned} P_{k+1}^{\mathrm{f}} &= \mathcal{F}_k P_k \mathcal{F}_k^{\mathrm{T}} + Q_k \\ &+ \mathcal{F}_k \Theta_k \mathcal{H}_k^{\mathrm{T}} + \frac{1}{4}\mathcal{H}_k \Gamma_k \mathcal{H}_k^{\mathrm{T}} + \frac{1}{3}\mathcal{F}_k \Gamma_k \mathcal{T}_k^{\mathrm{T}} \\ &- \frac{1}{4}\mathcal{H}_k P_k P_k^{\mathrm{T}} \mathcal{H}_k^{\mathrm{T}} - \frac{1}{6}\mathcal{H}_k P_k \Theta_k^{\mathrm{T}} \mathcal{T}_k^{\mathrm{T}} - \frac{1}{36}\mathcal{T}_k \Theta_k \Theta_k^{\mathrm{T}} \mathcal{T}_k^{\mathrm{T}} + \cdots \end{aligned} \tag{53}$$

where $\mathcal{F}_k$ is the tangent linear operator of $\mathbf{f}(\boldsymbol{\psi})$ calculated in $\boldsymbol{\psi}_k^{\mathrm{a}}$. In addition to the second order statistical moment, i.e. the error covariance matrix P_k, the equation also refers to higher order moments of which the third and fourth order moments Θ and Γ are included here. The operator $\mathcal{H}_k$ is now the Hessian of $\mathbf{f}(\boldsymbol{\psi})$, and $\mathcal{T}_k$ contains third order derivatives of $\mathbf{f}(\boldsymbol{\psi})$, both calculated in $\boldsymbol{\psi}_k^{\mathrm{a}}$. It is also possible to derive equations for Θ and Γ, which refer to even higher order moments.

In the so called extended Kalman filter the hierarchy of equations for the different statistical moments are closed by neglecting all moments with higher order than the error covariance matrix, i.e., a statistical linearization is performed. One is then left with the first line in equation (53) and the equation is similar to equation (44) except that an approximate tangent linear operator is used.

In Evensen (1992) it was shown that the overly simplified closure used in the extended Kalman filter resulted in unbounded error variance growth in the error covariance equation, and certain approximations had to be used to ensure stability of the system. A more consistent fix for the problem would be to apply a higher order closure for the error covariance equation. Such higher order turbulence closure schemes have been discussed by, e.g., Leith (1971), Leith and Kraichnan (1972), Fleming (1971a, and 1971b); however, these will be too computationally expensive for ocean models. The unboundedness of such an internal or dynamical instability was not a promising result with respect to the possibilities of using the Kalman filter in operational applications with realistic ocean models. In Evensen (1993) it was also pointed out that the use of open boundary conditions leads to "numerical ill-posedness" for the integration of the error covariance matrix. The optimality of the Kalman filter with linear models is also lost when the dynamics becomes nonlinear because of the approximate error evolution.

Two important conclusions from Evensen (1992, 1993) are that a sequential data assimilation algorithm gives good results in a data assimilation scheme for a multilayered nonlinear QG model, (see Appendix A), and that the results improve significantly according to improvements in the predicted error statistics. Results from Evensen (1992, 1993) and the additional fact that the Kalman filter is extremely expensive to compute, even for modest problem sizes, have motivated 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. Fortunately, there are other more consistent methods for calculating the error covariances than integrating the approximate error covariance equation (53).

3.2.4 Forecasting Error Statistics by Monte Carlo Methods

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 e.g., Gleeson (1970), Fleming (1971a, 1971b), Epstein and Pitcher (1972), Leith (1971, 1974), and Pitcher (1977). These papers discuss both the use of Monte Carlo methods and approximate stochastic dynamic prediction for finding approximate solutions of the equation for the probability density function. An application in oceanography has been discussed by 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).

The state vector at a specified time, $\boldsymbol{\psi}$, can be represented by a single point in an n-dimensional phase space $\mathcal{P}$. Thus, time evolution of the state vector $\boldsymbol{\psi}$ is described by continuous motion of the point along a trajectory in phase space. The uncertainty in the state vector can be represented by a large ensemble of possible states, each assigned an individual probability number. 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 approaches infinity, one can define a probability density distribution function

$$\phi(\boldsymbol{\psi}) = \frac{dN}{N}, \tag{54}$$

which can vary throughout the space. In Evensen (1994) it was shown that the QG model could be written in discrete form as an Itô stochastic differential equation describing a Markov process

$$d\boldsymbol{\psi} = \mathbf{g}(\boldsymbol{\psi}, t)dt + d\mathbf{q}, \tag{55}$$

where $\boldsymbol{\psi}$ is the state vector and $\mathbf{g}$ is a nonlinear vector function, and $d\mathbf{q} \in \Re^n$ is a vector of random white noise with zero mean. The evolution of the probability density for this equation is described by the Kolmogorov's equation, (also called the Fokker-Plank equation) which for the QG model simplifies to

$$\frac{\partial \phi}{\partial t} + \sum_{i=1}^{n} g_i \frac{\partial \phi}{\partial \psi_i} = \sum_{i,j}^{n} \frac{Q_{ij}}{2} \frac{\partial^2 \phi}{\partial \psi_i \partial \psi_j}, \tag{56}$$

where $Q = \overline{\mathbf{qq}^*}$ is the covariance matrix for the model errors. A derivation of this equation can be found in Jazwinski (1970). The probability density function represents the density of an infinite ensemble of possible ocean states, each having an associated probability number. The width of the probability density function corresponds to the variance of the ensemble and represents the errors in the predicted solution.

The stochastic forcing introduces a diffusion term that tends to flatten the probability density function (spreading the ensemble) during the integration, i.e. 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. An analytic steady state solution of the Kolmogorov's equation was found by Miller (1994) for the double well problem containing only one state variable and this simple example illustrates some of the difficulties of dealing with nonlinear dynamics.

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. An alternative to integration of the approximate error covariance equation is to use Monte Carlo methods. A large cloud of ocean states, i.e. 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 as a particle method in phase space. When the size N of the ensemble 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, not by the unboundedness of dynamical instabilities like in the extended Kalman filter.

When the Monte Carlo method is applied one first calculates a best guess initial condition based on available information from data and statistics. The model solution based on this initial state is denoted the central forecast. The uncertainty in the best guess initial condition is represented by the initial variance. An ensemble of initial states is then generated where the mean equals the best guess initial condition, and the variance is specified based on knowledge of the uncertainty in the first guess initial state. The covariance or smoothness of the ensemble members should reflect the true scales of the system, e.g., the internal Rossby radius is the physical scale for a QG model. A procedure for generating such pseudo random fields with a specified variance and covariance was outlined in Evensen (1994).

The effect of external error growth must be included to give reliable estimates for the evolution of errors. One way to do this is to force the model with smooth pseudo random fields with a specified variance. This will provide a realistic increase in the ensemble variance, provided that the estimates of the model error variance is reasonably good.

In Evensen (1994) it was illustrated how one could calculate the covariance functions required to construct the Kalman gain (47) from the ensemble. Further it was shown that by operating directly on each individual member of the ensemble with the equation for the analyzed state estimate (43) one recovered an ensemble with the correct statistics as determined by (46).

4 Example

To illustrate a data assimilation process an example of sequential data assimilation is presented where the error statistics of the model forecast is generated using ensemble statistics. The method was briefly outlined in the previous section and has been extensively discussed in Evensen (1994). The model is a two layer QG model set up for a region including the Halten bank on the Norwegian continental shelf. This area has been subject to an extensive study during the Norwegian Continental Shelf Experiment (NORCSEX), (Haugan *et al.*, 1991). A 81 times 121 grid has been used and the size of the domain is about 324 km along shore and 216 km cross-shore. The internal Rossby radius which is the characteristic length scale for mesoscale circulation on the shelf is 5.4 km and is resolved by the grid. The boundary along the coast, i.e., the eastern boundary at $y = 0.0$, is closed and there is a specified inflow of the Norwegian coastal current and the Atlantic current at the southern boundary, i.e. $x = 0.0$. Realistic smoothed bottom topography has been used and the domain is shown in Figure 1.

For this example the data have been generated by running a reference case from which the solution has been stored at specific data locations in space and time. A regular mesh of 9×6 data points in the upper layer only are generated every $\Delta t = 1.0$ which corresponds to about 316 minutes. The distance between the data points is about 35 km which means the Rossby radius is not resolved by the data, although eddies and meanders normally extends over a few Rossby radii.

The initial condition for the reference case was found by adding a pseudo random field to the initial condition used for the assimilation run, and then integrating the model for about three weeks, where the final stream function was used as initial conditions for the reference case.

An ensemble of 400 statistically independent ocean states was generated by adding pseudo random fields, with variance equal to 0.1 and covariance corresponding to the physical scales of the problem, to the initial condition for the assimilation run. The ensemble members were integrated forward in time and updated every time measurements were available according to equation (43). The covariance functions required for calculating the Kalman gain were generated as statistical moments of the ensemble. This required a few hours of CPU time on a CRAY Y-MP.

The initial error variance is 0.1 in the interior domain and on the inflow boundary, and it is 0.0 at the closed and outflow boundaries. A finite error variance on the inflow boundary generates perturbations in

the ensemble members that propagate downstream, i.e. the inflow boundary is an error source. At the closed boundary along the coast the solution is known and the errors should be zero. The error variance at the eastern boundary is zero and kept constant to avoid the inflow of energy in an area known to be rather quiet. At the northern boundary the initial variance is zero, but a radiation conditions has been applied so that eddies and meanders in the coastal current are able to escape.

The results of the assimilation are shown in Figures 2 and 3. When comparing the time series of the stream function fields from the reference case (left) and the central forecast for the assimilation experiment in Figure 2, it is seen how the central forecast approaches the reference stream function during the integration. All the main structures like eddies and meanders are reproduced although with some differences in amplitude and phase. In Figure 3, time series of the mean (left) and the error variance (right) of the ensemble is shown. At the final time the mean is slightly closer to reference case than the central forecast. In the time series of error variance plots it is seen how the assimilation of measurements reduces the errors at the measurement locations, i.e. the ensemble converges. Further, the errors are decreasing faster in the Atlantic current where the information from the measurements are propagating downstream.

5 Discussion

A brief discussion has been given of the properties of the most promising methods for inverse calculations and data assimilation in nonlinear models. Traditional modeling of time dependent systems assumes a dynamical model with appropriate initial and boundary conditions stated in a way which in mathematical terms ensures well-posedness, thus a unique solution can be found. However, for many physical systems, additional information from measurements may also be available and this information should be used to improve the solutions and the knowledge of the true state of the dynamical system. The information contained in the measurements can not be directly imposed on the well-posed mathematical problem since this results in an over-specified problem where in general no solution exists. However, by allowing the model equations to contain errors, the over-specified system can be reformulated as a least-squares problem which can be solved for a unique minimizing solution which is close to both the model dynamics and the measurements.

Inverse methods are also applied for optimal design of observation schemes, including the design of repeat cycles of monitoring satellites. The criterion to be used is to maximize the condition of the linear system (35). The condition is mainly determined by the interdependence between the measurements, i.e., if the measurements are clustered or the errors in the collected information are correlated in space or time, that will reduce the condition of the problem. Examples are given by Bennett (1985), McIntosh (1987), Barth and Wunsch (1990), and Ghil and Ide (1994). Inverse methods have also been used for parameter estimation in ocean models by Smedstad and O'Brien (1991) and Yu and O'Brien (1991).

There is also an extensive literature on so called suboptimal methods. Such methods can be derived by applying certain assumptions on the generalized inverse formulation or on the error statistics used in the Kalman filter.

A much used variational method is based on the assumption that the dynamics imposed by the model is perfect, i.e., the model errors are zero. This assumption is not valid in general but it decouples the system of Euler–Lagrange equations (8–11). Exploiting the fact that the adjoint variable at $t = 0$ is the gradient of the penalty function with respect to the initial conditions, a gradient descent algorithm can be used to iterate the initial conditions so that the forward model integration results in a best possible fit to the data. In each iteration the perfect model is integrated forward in time followed by a backward integration of the adjoint equation (10), to obtain the gradient. This strong constraint approach which is often called "the adjoint method" was first discussed by Lewis and Derber (1985), Le Dimet and

Talagrand (1986), and Talagrand and Courtier (1987), and it has been reviewed by e.g. Schröter (1989). See also the special issue on adjoint applications in dynamic meteorology in Tellus, **45A**:5, (1993). Posterior error estimates can be generated by inverting the Hessian of the penalty function, but this is an enormous numerical task. If the assumption of a perfect nonlinear model is invalid it is very difficult to get a good fit to the data if the assimilation time interval is several times the prediction limit of the dynamics. The penalty function will in fact contain multiple local minima and the gradient descent method is likely to converge to one of the local minima. By solving the weak constraint problem, e.g., using simulated annealing or the representer method, the model errors allows the system to diminish the weight on information from data outside a certain time interval determined by the prediction limit of the dynamics, i.e. the measurements will have a bell shaped influence in time, determined by the time covariance functions for the dynamics.

Another simplifying approach which has been widely used, is sequential methods where certain assumptions are made for the error covariances, e.g., steady error statistics. These methods are normally called optimal interpolation (OI) and they have been used extensively in operational weather forecasting for the last decades (Daley, 1991). A review on OI methods is given by Todling and Cohn (1993). It should be noted that beside the loss of optimality this method does not directly provide posterior error estimates.

Within sequential methods, Monte Carlo methods can be used to forecast the error covariances, and this resolves many of the crucial problems with the extended Kalman filter. Ensemble forecasting can also be used to calculate the representer fields (36), and in that case the backward integration of the linearized equations in (33) is avoided. Note that the ensemble forecast must be performed for each of the linear iterates when the model is nonlinear.

All the methods discussed in Section 3 can be considered as statistical or stochastic, and they are well suited for exploiting the new parallel computer architectures. They also include model errors to reflect the approximations used when the model equations are derived, and they provide estimates of the posterior error statistics.

A Multilayered Quasi-geostrophic Model

The ocean model is the multilayered and nonlinear quasi-geostrophic (QG) model on an f plane, (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 $R_{\rm d}$ is the internal Rossby radius of deformation of the upper layer, given by $R_{\rm d}^2 = [(\rho_2 - \rho_1)gD_1] / [\rho_0 f^2]$, where g is the gravitational acceleration, ρ_0 is averaged density and f is the Coriolis parameter. The characteristic horizontal velocity is denoted U, yielding a time scale $T = R_{\rm d}/U$. The pressure scale is $\rho_0 f U R_{\rm d}$, and the stream function scale is $UR_{\rm d}$. The nondimensional quasi-geostrophic equations are

$$(\frac{\partial}{\partial t} + u_l\frac{\partial}{\partial x} + v_l\frac{\partial}{\partial y})\zeta_l = 0, \quad l = 1, n_z, \tag{57}$$

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

$$u_l = -\frac{\partial \Psi_l}{\partial y} \qquad v_l = \frac{\partial \Psi_l}{\partial x}. \tag{58}$$

The vorticity in each layer is given by

$$\zeta_1 = \nabla^2\Psi_1 + fr_{1,2}(\Psi_2 - \Psi_1), \tag{59}$$

$$\zeta_l = \nabla^2\Psi_l - fr_{l,1}(\Psi_l - \Psi_{l-1}) + fr_{l,2}(\Psi_{l+1} - \Psi_l), \tag{60}$$

$$\zeta_{n_z} = \nabla^2\Psi_{n_z} - fr_{n_z,1}(\Psi_{n_z} - \Psi_{n_z-1}) + \eta, \tag{61}$$

where $l = 2, n_z-1$. When ζ is known, this is a set of coupled Helmholtz equations for the stream function Ψ. The Laplacian 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"

$$fr_{l,1} = \frac{D_1}{D_l}\frac{\rho_2 - \rho_1}{\rho_l - \rho_{l-1}}, \qquad fr_{l,2} = \frac{D_1}{D_l}\frac{\rho_2 - \rho_1}{\rho_{l+1} - \rho_l}. \tag{62}$$

The bottom topography term is

$$\eta = \varepsilon^{-1}\frac{h_\mathrm{b}}{D_{n_z}}, \tag{63}$$

with ε as the Rossby number and h as bottom topography.

References

Azencott, R. (1992), *Simulated Annealing*, John Wiley, New York.

Barth, N. and Wunsch, C. (1990), Oceanographic experiment design by simulated annealing, *J. Phys. Oceanogr.*, **20**:1249–1263.

Bennett, A. F. (1985), Array design by inverse methods, *Prog. Oceanog.*, **15**:129–156.

Bennett, A. F. (1992), *Inverse methods in Oceanography*, Cambridge University Press.

Bennett, A. F. and Baugh, J. R. (1992), A parallel algorithm for variational assimilation in oceanography and meteorology, *J. Atmos. Ocean. Tech.*, **9**:426–433.

Bennett, A. F. and Budgell, W. P. (1987), Ocean data assimilation and the Kalman filter: Spatial regularity, *J. Phys. Oceanogr.*, **17**(10):1583–1601.

Bennett, A. F. and Budgell, W. P. (1989), The Kalman smoother for a linear quasigeostrophic model of ocean circulation, *Dyn. Atmos. Oceans*, **13**(3-4):219–267.

Bennett, A. F. and Chua, B. S. (1994), Open-ocean modeling as an inverse problem: The primitive equations, *Mon. Weather Rev.*, In press.

Bennett, A. F., Leslie, L. M., Hagelberg, C. R., and Powers, P. E. (1993), Tropical cyclone prediction using a barotropic model initialized by a generalized inverse method, *Mon. Weather Rev.*, **121**:1714–1729.

Bennett, A. F. and Miller, R. N. (1990), Weighting initial conditions in variational assimilation schemes, *Mon. Weather Rev.*, **119**:1098–1102.

Bennett, A. F. and Thorburn, M. (1992), The generalized inverse of a nonlinear quasigeostrophic ocean circulation model, *J. Phys. Oceanogr.*, **22**:213–230.

Bohachevsky, I. O., Johnson, M. E., and Stein, M. L. (1986), Generalized simulated annealing for function optimization, *Technometrics*, **28**(3):209–217.

Bretherton, F. P., Davis, R. E., and Fandry, C. B. (1976), A technique for objective analyses and design of oceanographic experiments applied to MODE-73, *Deep-Sea Res.*, **23**:559–582.

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

Budgell, W. P. (1987), Stochastic filtering of linear shallow water wave processes, *SIAM J. Sci. Stat. Comput.*, **8**(2):152–170.

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

Cohn, S. E. and Parrish, D. F. (1991), The behavior of forecast error covariances for a Kalman filter in two dimensions, *Mon. Weather Rev.*, **119**:1757–1785.

Courant, R. and Hilbert, D. (1953), *Methods of mathematical physics*, volume 1, Wiley–Interscience, New York, London, Sydney.

Daley, R. (1991), *Atmospheric Data Analysis*, Cambridge University Press.

Daley, R. (1992a), The effect of serially correlated observation and model error on atmospheric data assimilation, *Mon. Weather Rev.*, **120**:164–177.

Daley, R. (1992b), Estimating model-error covariances for application to atmospheric data assimilation, *Mon. Weather Rev.*, **120**:1736–1746.

Daley, R. (1992c), Forecast-error statistics for homogeneous and inhomogeneous observation networks, *Mon. Weather Rev.*, **120**:627–643.

Daley, R. (1992d), The lagged innovation covariance: A performance diagnostic for atmospheric data assimilation, *Mon. Weather Rev.*, **120**:178–196.

Daley, R. and Ménard, R. (1993), Spectral characteristics of Kalman filter systems for atmospheric data assimilation, *Mon. Weather Rev.*, **121**:1554–1565.

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

Duane, S., Kennedy, A. D., Pendleton, B. J., and Roweth, D. (1987), Hybrid Monte Carlo, *Phys. Lett. B.*, **195**:216–222.

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

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

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

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

Evensen, G. (1994), Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, **99**(C5):10,143–10,162.

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

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

Gauthier, P., Courtier, P., and Moll, P. (1993), Assimilation of simulated wind lidar data with a Kalman filter, *Mon. Weather Rev.*, **121**:1803–1820.

Gelb, A. (1974), *Applied Optimal Estimation*, MIT Press Cambridge.

Gelfand, I. M. and Fomin, S. V. (1963), *Calculus of Variations*, Prentice-Hall, Englewood Cliffs, N. J.

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

Ghil, M., Cohn, S., Tavantzis, J., Bube, K., and Isaacson, E. (1981), Applications of estimation theory to numerical weather prediction, in *Dynamic Meteorology: Data Assimilation Methods*, edited by Bengtsson, L., Ghil, M., and Kallén, E., pp. 139–224, Springer Verlag, New York.

Ghil, M. and Ide, K. (1994), Extended Kalman filtering for vortex systems: An example of observing-system design., in *Data Assimilation: A New Tool for Modelling the Ocean in a Global Change Perspective*, edited by Nihoul, J. C. H., Springer Verlag.

Ghil, M. and Malanotte-Rizzoli, P. (1991), Data assimilation in meteorology and oceanography, *Adv. Geophys.*, **33**:141–266.

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

Haugan, P. M., Evensen, G., Johannessen, J. A., Johannessen, O. M., and Pettersson, L. (1991), 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.

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

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

Johannessen, J. A., Røed, L. P., Johannessen, O. M., Evensen, G., Hackett, B., Petterson, L. H., Haugan, P. M., Sandven, S., and Shuchman, R. (1993), Monitoring and modeling of the marine coastal environment, *Photogrametric Eng. and Remote Sensing*, **59**(3):351–361.

Johannessen, O. M., Campbell, W. J., Shuchman, R., Sandven, S., Gloersen, P., Johannessen, J. A., Josberger, E. G., and Haugan, P. M. (1992), *Microwave study programs of air–ice–ocean interactive processes in the seasonal ice zone of the Greenland and Barents seas*, volume 68 of *Geophysical Monograph*, pp. 261–289, American Geophysical Union, Washington, DC.

Kalman, R. E. (1960), A new approach to linear filter and prediction problems, *J. Basic. Eng.*, **82**:35–45.

Kevorkian, J. and Cole, J. D. (1981), *Perturbation Methods in Applied Mathematics*, Applied Math. Sci., Springer-Verlag, New York.

Kirkpatric, S., Gelatt, C. D., and Vecchi, M. P. (1983), Optimation by simulated annealing, *Science*, **220**(4598):671–680.

Le Dimet, F.-X. and Talagrand, O. (1986), Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects, *Tellus, Ser. A*, **38**:97–110.

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

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

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

Lewis, J. M. and Derber, J. C. (1985), The use of adjoint equations to solve a variational adjustment problem with advective constraints, *Tellus, Ser. A*, **37**:309–322.

Lorenc, A. C. (1986), Analysis methods for numerical weather prediction, *Q. J. R. Meteorol. Soc.*, **112**:1177–1194.

McIntosh, P. C. (1987), Systematic design of observational arrays, *J. Phys. Oceanogr.*, **17**:885–902.

McIntosh, P. C. (1990), Oceanic data interpolation: Objective analysis and splines, *J. Geophys. Res.*, **95**(C8):13,529–13,541.

Menke, W. (1984), *Geophysical Data Analysis: Discrete Inverse Theory*, Academic Press, Orlando.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953), Equation of state calculations by fast computing machines, *J. Chem. Phys.*, **21**(6):1087–1092.

Miller, R. N. (1986), Toward the application of the Kalman filter to regional open ocean modeling, *J. Phys. Oceanogr.*, **16**:72–86.

Miller, R. N. (1989), Direct assimilation of altimetric differences using the Kalman filter, *Dyn. Atmos. Oceans*, **13**(3-4):317–333.

Miller, R. N. (1990), Tropical data assimilation experiments with simulated data: The impact of the tropical ocean and global atmosphere thermal array for the ocean, *J. Geophys. Res.*, **95**(C7):11,461–11,482.

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

Miller, R. N. and Cane, M. A. (1989), A Kalman filter analyses of sea level height in the tropical Pacific, *J. Phys. Oceanogr.*, **19**:773–790.

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

Moiseenko, V. A. and Saenko, O. A. (1992), Using the Kalman smoother to analyse and process oceanographic data, *Russ. J. Numer. Anal. Math. Modelling*, **7**(4):291–294.

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

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

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

Sasaki, Y. (1958), An objective analysis based on the variational method, *J. Meteor. Soc. Japan*, **33**(6):77–88.

Schröter, J. (1989), Driving of non-linear time-dependent ocean models by observation of transient tracers—a problem of constrained optimisation, in *Oceanic Circulation Models: Combining Data and Dynamics*, edited by Anderson, D. L. T. and Willebrand, J., pp. 257–285, Kluwer Academic Publishers.

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

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

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

Smedstad, O. M. and O'Brien, J. J. (1991), Variational data assimilation and parameter estimation in an equatorial Pacific Ocean model, *Prog. Oceanog.*, **26**:179–241.

Talagrand, O. and Courtier, P. (1987), Variational assimilation of meteorological observations with the adjoint vorticity equation I: Theory, *Q. J. R. Meteorol. Soc.*, **113**:1311–1328.

Tarantola, A. (1987), *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*, Elsevier: Amsterdam.

Todling, R. and Cohn, S. E. (1993), Suboptimal schemes for atmospheric data assimilation based on the Kalman filter, *Mon. Weather Rev.*, **122**, Submitted.

Todling, R. and Ghil, M. (1994), Tracking atmospheric instabilities with the Kalman filter. Part I: Methodology and one-layer results, *Mon. Weather Rev.*, pp. 183–204.

Yu, L. and O'Brien, J. J. (1991), Variational estimation of the wind stress drag coefficient and the oceanic eddy viscosity profile, *J. Phys. Oceanogr.*, **21**:709–719.

Acknowledgments

The author would like to thank P.J. van Leeuwen for valuable comments on the manuscript, and A.F. Bennett and R.N. Miller for interesting discussions on inverse methods and data assimilation. The work and computer time 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.

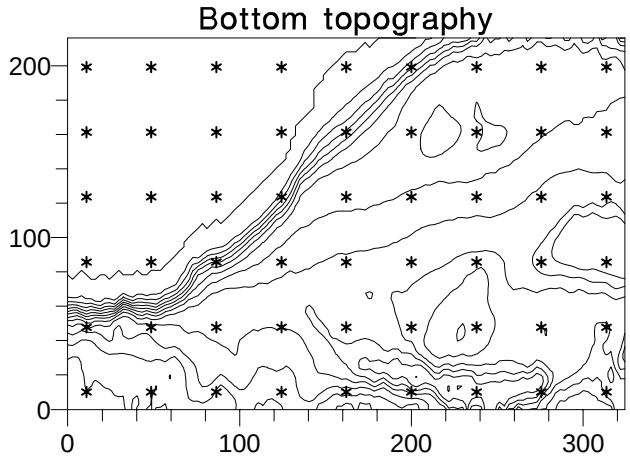

Figure 1: The smoothed bottom topography. Note the steep shelf break where contours down to 700 m are included, and the varying depths on the shelf especially the Halten bank with the deep trench inside it. The locations of the data points are shown as black stars. The contour interval is 50 m and the upper contour is 100 m. Axis numbers denote km.

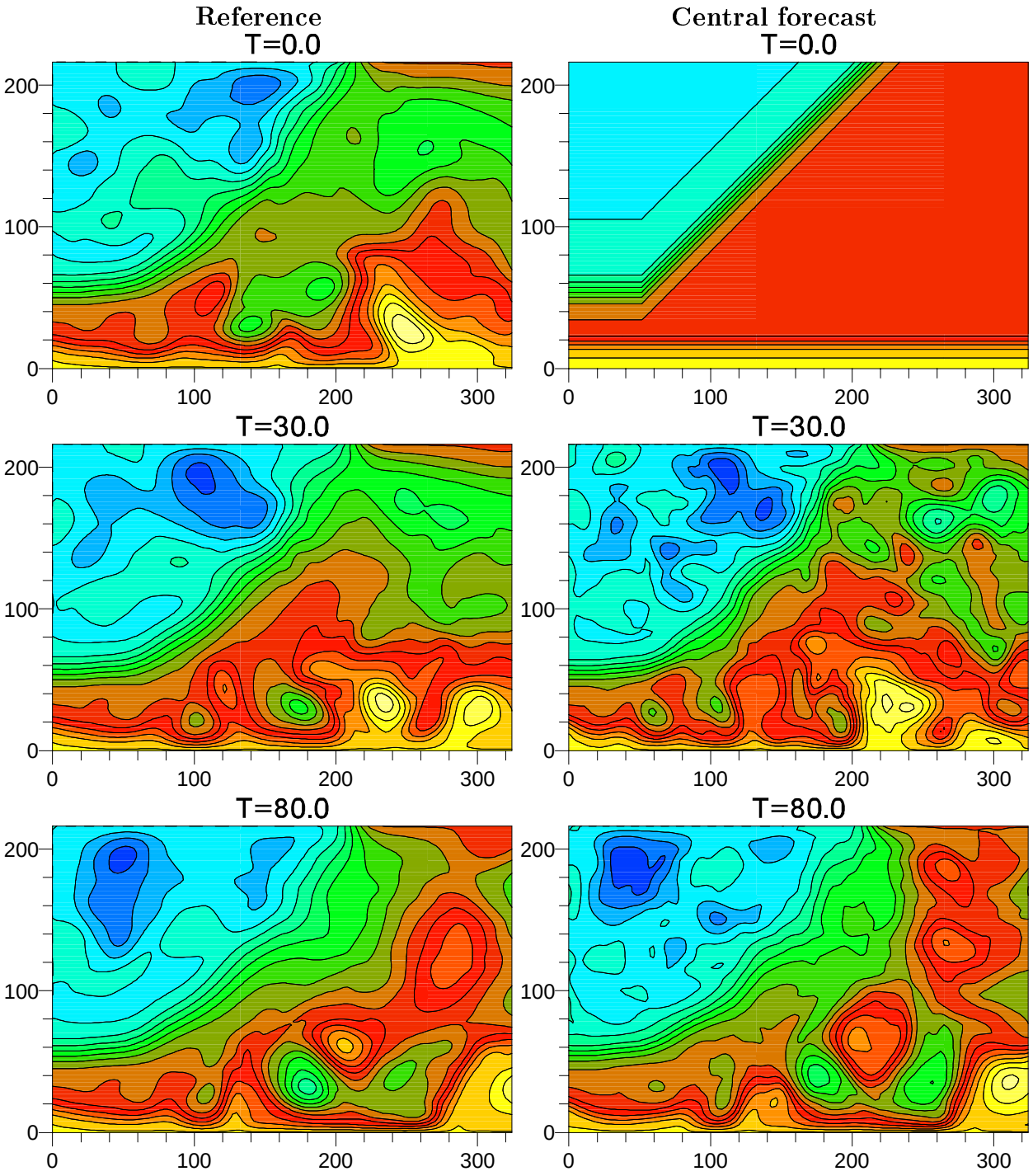

Figure 2: The reference stream function in the left column and the central forecast generated by the assimilation in the right column. Contour intervals are 0.5.

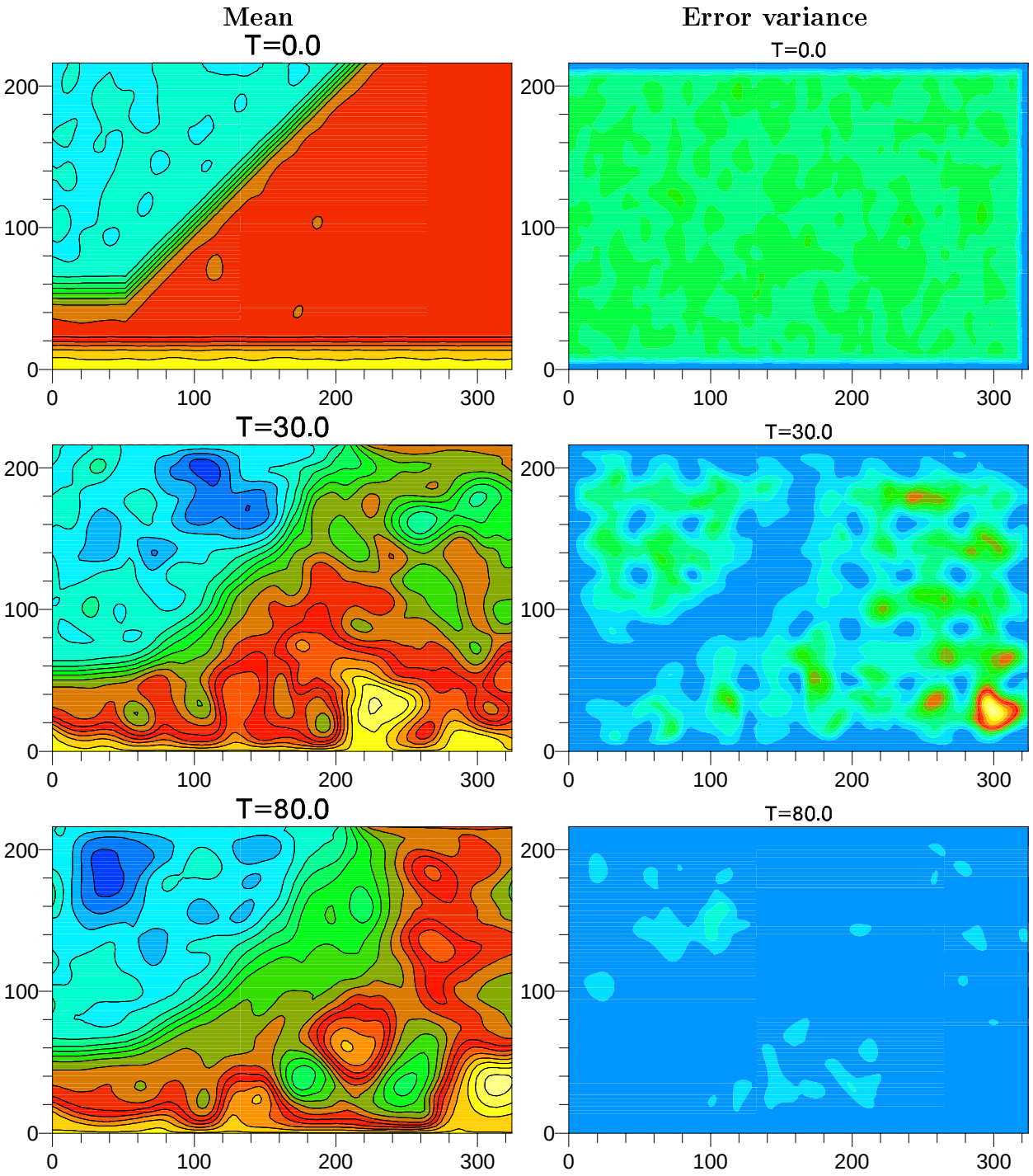

Figure 3: The statistical moments for the assimilation run; mean stream function in the left column, and error variance in the right column. Contour intervals are 0.5 for the mean and 0.025 for the error variance plots.