Data assimilation into nonlinear stochastic models

By ROBERT N. MILLER[1]*, EVERETT F. CARTER, Jr.[2] and SALLY T. BLUE[3†], [1]*College of Oceanic and Atmospheric Sciences, Oregon State University, Oceanography Admin. Bldg. 104, Corvallis, OR, USA 97331-5503, [2]Taygeta Scientific Inc., 1340 Munras Ave. Suite 314 Monterey, CA, USA 93940 and [3]College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, OR, USA 97331-5503*

ABSTRACT

With very few exceptions, data assimilation methods which have been used or proposed for use with ocean models have been based on some assumption of linearity or near-linearity. The great majority of these schemes have at their root some least-squares assumption. While one can always perform least-squares analysis on any problem, direct application of least squares may not yield satisfactory results in cases in which the underlying distributions are significantly non-Gaussian. In many cases in which the behavior of the system is governed by intrinsically nonlinear dynamics, distributions of solutions which are initially Gaussian will not remain so as the system evolves. The presence of noise is an additional and inevitable complicating factor. Besides the imperfections in our models which result from physical or computational simplifying assumptions, there is uncertainty in forcing fields such as wind stress and heat flux which will remain with us for the foreseeable future. The real world is a noisy place, and the effects of noise upon highly nonlinear systems can be complex. We therefore consider the problem of data assimilation into systems modeled as nonlinear stochastic differential equations. When the models are described in this way, the general assimilation problem becomes that of estimating the probability density function of the system conditioned on the observations. The quantity we choose as the solution to the problem can be a mean, a median, a mode, or some other statistic. In the fully general formulation, no assumptions about moments or near-linearity are required. We present a series of simulation experiments in which we demonstrate assimilation of data into simple nonlinear models in which least-squares methods such as the (Extended) Kalman filter or the weak-constraint variational methods will not perform well. We illustrate the basic method with three examples: a simple one-dimensional nonlinear stochastic differential equation, the well known three-dimensional Lorenz model and a nonlinear quasigeostrophic channel model. Comparisons to the extended Kalman filter and an extension to the extended Kalman filter are presented.

1. Introduction

The vast majority of the data assimilation methods in use or proposed for application to numerical modeling in oceanography or numerical weather prediction were derived and validated for linear systems with Gaussian noise. In nonlinear systems, even if the errors are initially Gaussian, they do not, in general, remain so. It is therefore important to ask what happens when our linearized methods are applied to stochastically perturbed nonlinear systems.

Least-squares variational methods have been applied with some success in numerical weather forecasting. Solution of the variational problem proceeds by a sequence of adjoint calculations. An extensive bibliography of articles on applications

* Corresponding author.
email: miller@oce.orst.edu
† Present affiliation: Rogue Wave Software, Inc., Corvallis, OR, USA 97333.

of adjoint calculations to meteorological problems was given by Courtier et al. (1993). The incremental form of the variational method which made four dimensional variational methods practical for operational numerical weather prediction was presented by Courtier et al. (1994). The different formulations of the variational inverse problem are well known to be closely related, and all are related to the Kalman filter. For rigorous demonstrations of the relation between the different least-squares methods, the reader may consult Courtier (1997) or Miller (1998), and references therein.

Sophisticated iterative methods have been developed for the solution of the variational inverse problem for nonlinear quasigeostrophic systems (Bennett and Thorburn, 1992; Bennett et al., 1993). The extended Kalman filter (hereafter EKF) has also been applied to nonlinear quasigeostrophic systems (Evensen, 1992; Gauthier et al., 1993). This was a reasonable approach, since the linearizations in time and phase space which are essential to the variational approaches and to the EKF are widely applicable (Lacarra and Talagrand, 1988; Budgell, 1986), but Evensen (1992) found that the estimated error variance grew unreasonably rapidly in a model run without assimilation. Gauthier et al. (1993) found a similar result in data-poor regions, as did Bouttier (1994) in his study of the application of an approximate Kalman filter to an operational numerical weather forecast model.

Bouttier (1994) solved the problem of excessive growth of the error covariance by modifying his approximate Kalman filter to include an explicit saturation mechanism. Evensen (1992) solved the problem at first by modifying the dynamical operator, and later (Evensen, 1994) by implementing a generalization of the extended Kalman filter based on a Monte-Carlo estimate of covariance statistics. This approach was also suggested by Bouttier (1994) who felt at the time that it would not be computationally practical for numerical weather prediction. In most of these systems calculations have been performed over times which are within predictability limits imposed by the dynamics, and regime transitions are not observed. Our purpose here is to approach the problems of intrinsically nonlinear behavior such as bifurcation and chaos directly.

Miller et al. (1994), hereafter MGG, also devised a generalization of the EKF based on Monte-Carlo estimates of covariance statistics and applied it to the unperturbed Lorenz system with good results. We shall see below that Evensen's (1994) method also performs reasonably well on an example with the stochastically perturbed Lorenz system. Evensen and Fario (1997) also demonstrated a smoother for the Lorenz system based on a Monte-Carlo method.

We work directly with stochastic differential equations in order to model systems with random forcing or random parameter values. Such systems arise in practice when forcing functions or parameter values are derived from noisy data. Stochastic terms in the equations are also useful for parameterization of neglected physics, such as sub-gridscale interactions. The probability density function (pdf) associated with a given stochastic differential equation evolves according to the Fokker–Planck equation, a parabolic equation in a number of spatial dimensions equal to the state dimension of the underlying stochastic system. The only practical way to solve equations of this type for models of high state dimension is through the use of Monte-Carlo methods. In cases in which the qualitative behavior of a system is governed by its behavior on some low-dimensional subset of state space, we believe that practical Monte-Carlo methods can be implemented at costs comparable to the EKF.

Within this framework, observations are also considered as pdf's rather than single numerical values. Here we do the simplest thing: we consider observed values as the means of Gaussian distributions, with variances given by estimates of the observation errors. Assimilation of data is then accomplished by using Bayes' formula to combine the prior pdf, calculated from the Fokker–Planck equation with the observation pdf to form a conditional distribution (Lorenc and Hammon, 1988). Statistics including means, medians, modes, standard deviations and confidence intervals may then be calculated directly from the pdf's to produce numerical assimilation products.

Specification of the random components of our stochastic models is one of the most important tasks in the construction of models for the study of the data assimilation process. Little is known of the characteristics of natural noise processes. Simulations with stochastic models require explicit models of the relevant noise processes. This requirement is stronger than is needed for most

data assimilation methods (e.g., the EKF) which require only estimates of the moments. Future data assimilation systems which use techniques based on explicit noise models must be subjected to sensitivity tests and posterior tests of consistency. Similar tests should be applied to any data assimilation system.

Stochastic differential equations do not have solutions in the classical sense. Extensions of classical calculus to stochastic functions were formulated by Itô (1951) and Stratonovich (1966). General background on stochastic calculus may be found in Jazwinski (1970) or Kloeden and Platen (1992). For the purposes of the present discussion we will not need to introduce a full rigorous treatment of the stochastic calculus, beyond stating at this point that we use the Itô stochastic calculus.

We first describe, in the next section, the equations necessary to propagate the pdf of a system in time, given the physics. Section 3 describes data assimilation into nonlinear systems for both time discrete and time continuous observations. In Section 4 we demonstrate the method with a simple one-dimensional system. We then proceed to more complex models. The familiar Lorenz model, a three state variable system which exhibits chaos, is considered in Section 5. Our final example, presented in Section 6, is a nonlinear quasigeostrophic channel model in spectral form. With the resolution chosen, that model has forty-four state variables. Section 7 contains discussion and summary.

2. The Fokker–Planck equation

Consider a system described by the vector Itô stochastic differential equation,

$$\mathrm{d}x_t = f(x_t, t)\,\mathrm{d}t + g(x_t, t)\,\mathrm{d}w_t \tag{1}$$

where x_t is a vector of dimension n, f is a vector function of dimension n, w_t is a k-dimensional Wiener process and g is an $n \times k$ matrix valued function. The notation in (1) is standard in the stochastic differential equations literature. We use it here to remind the reader of the distinction between stochastic differential equations and classical ordinary differential equations. Eq. (1) does not have solutions in the classical sense. Solutions to (1) defined in terms of the stochastic calculus

are everywhere continuous and nowhere differentiable with probability 1.0.

Numerical integrations of individual realizations of stochastic differential equations like (1) are done here with the third-order method of Milshtein (1978). Throughout this paper, we shall view x_t as a vector random process, which has a properly defined pdf. The differential equation that describes the evolution of the pdf, $P(x, t)$, is the Fokker–Planck equation (Gardiner, 1983),

$$\frac{\partial P}{\partial t} = -\nabla \cdot (f(x, t)P) + \sum_{i,j} \frac{\partial^2}{\partial x_i \, \partial x_j} (Q/2)_{ij} P$$

$$= L_{\mathrm{FP}}(P) \tag{2}$$

where $Q = g(x, t)g^{\mathrm{T}}(x, t)$. The Fokker–Planck equation (also known as the forward Kolmogorov equation), is a parabolic partial differential equation with the number of space dimensions equal to the number of state variables of the underlying stochastic differential equation (1). The reader should note that (2) is a deterministic partial differential equation. In its simplest form with $f = 0$ and constant g, (2) becomes the simple diffusion equation which describes the evolution of the concentration of a passive tracer, which in turn consists of particles moving at random according to (1).

The Fokker–Planck equation is an advection-diffusion equation, and thus numerical integration of the Fokker–Planck equation shares many of the difficulties of numerical integration of the Navier–Stokes equations. The Fokker–Planck equation is further complicated by the requirement that for all times the probability density must be nonnegative, and that the total probability density must integrate to one. The obvious approximation of shifting (or truncating) the negative part and renormalizing at each step in the integration introduces appreciable errors into the calculation even after a short integration time. The difficulties in integrating the Fokker–Planck equation can be illustrated with a simple example. Consider the simple scalar Langevin equation:

$$\mathrm{d}x_t = -x_t\,\mathrm{d}t + Q^{1/2}\,\mathrm{d}w_t. \tag{3}$$

The corresponding Fokker–Planck equation is:

$$p_t = (Q/2)p_{xx} + (xp)_x. \tag{4}$$

We can write a simple centered difference

scheme for the steady solution:

$$(Q/2)(p_{j+1} - 2p_j + p_{j-1})/\Delta x^2$$
$$+ ((j+1)p_{j+1} - (j-1)p_{j-1})/2 = 0. \quad (5)$$

The solution to the difference equation can be written:

$$p_j = \prod_{l=1}^{j-1}(1 - l\Delta x^2/Q) \Big/ \prod_{l=1}^{j}(1 + l\Delta x^2/Q). \quad (6)$$

This expression will eventually produce negative values when j becomes large enough. This can be interpreted as the familiar cell Reynolds number criterion from computational fluid dynamics (Fletcher, 1988), with the advection speed identified as $x = l\Delta x$ and Q identified as the viscosity. For smaller Δx, i.e., lower cell Reynolds number, the loss of positivity occurs further from the origin. This is a particularly unhappy circumstance: fine resolution is evidently needed far from the region of interest in order to prevent unrealistic effects from appearing. Different approaches to the solution of the Fokker–Planck equation will be discussed in the section.

3. Nonlinear data assimilation

In this section, the formalism for constructing data assimilation schemes within the framework of pdf's is described. The observations themselves may be viewed as means, medians, modes or some other functionals of a pdf. For the present purposes, we shall view observations as the means of Gaussian distributions with given variances. Data assimilation is then accomplished by constructing a conditional pdf: for a small volume in state space, the conditional pdf is the probability that the system is in that volume, given that we know the observation pdf's and the prior pdf from the Fokker–Planck equation. Formulas can be derived for both discrete and continuous observations.

3.1. Calculation of conditional probabilities

We first consider the case in which the system is described by the Itô stochastic differential equation (1) and the observation process is discrete in time,

$$y_k = h(x_k, t_k) + \sqrt{R}v_k, \quad (7)$$

where y_k is a vector with m components and v_k is a white sequence of Gaussian random variables with covariance given by the $m \times m$ identity. The matrix $\mathbf{R}$ is assumed to be diagonal. Given observations (7) we may examine the evolution of the conditional pdf of the system (1). We already know that in the absence of observations, the pdf will evolve according to the Fokker–Planck equation (2). Given this prior pdf, we can then apply Bayes' theorem (Lorenc and Hammon, 1988; Jazwinski, 1970) to obtain a new one that is conditioned on the observation y_k,

$$P(x, t_k | y_k) = \frac{P(y_k | x)P(x, t_k | y_k^-)}{\int P(y_k | x)P(x, t_k | y_k^-) \, dx} \quad (8)$$

where $P(x, t_k | y_k^-)$ is the pdf given by the Fokker–Planck equation up to but not including the new observation and $P(y_k | x)$ is the probability of seeing the observation y_k given the current state of the system x.

If we assume that the observation noise process is Gaussian white noise with covariance $\mathbf{R}_k$, then we can write (Jazwinski, 1970):

$$P(y_k | x) = \frac{1}{(2\pi)^{m/2}\sqrt{\mathbf{R}_k}}$$
$$\times \exp\left(-\frac{1}{2}[y_k - h(x, t_k)]^T\mathbf{R}_k^{-1}[y_k - h(x, t_k)]\right), \quad (9)$$

where m is the dimension of the observation vector y.

The conditional pdf obtained from (8) contains all of the information available about the stochastic system. Eq. (2) along with (7), (8) and a prior estimate of the observation pdf such as (9) constitute a data assimilation system. Specific estimates of the state of the system can be constructed by computing the mean or the mode of the pdf calculated in (8). This pdf can also be used to calculate confidence intervals for the state estimates. We call this discrete observation assimilation scheme the Bayes' theorem assimilation.

We note here that in the limit as observations become more frequent, the observation process can itself be modeled as a continuous process governed by an Itô stochastic differential equation:

$$dz_t = h(x_t, t) \, dt + \sqrt{R} \, d\eta_t \quad (10)$$

where η_t is a Wiener process, uncorrelated with the system noise process dw. In this case the

corresponding limit of our Bayesian updating process is the Zakai equation: (Zakai, 1969; see also Kushner, 1962, Rozovskii, 1990).

$$dP = L_{FP}(P)\,dt + (\boldsymbol{h}_t - \hat{\boldsymbol{h}}_t)^T \boldsymbol{R}^{-1}(t)(d\boldsymbol{z}_t - \hat{\boldsymbol{h}}_t\,dt)P,$$

$$(11)$$

where $\boldsymbol{h}_t = \boldsymbol{h}(x_t, t)$, $\hat{\boldsymbol{h}}_t = \int \boldsymbol{h}(\boldsymbol{x}, t)P(\boldsymbol{x}, t\,|\,\boldsymbol{Y}_t)\,d\boldsymbol{x}$, $\boldsymbol{Y}_t$ is the time series of observations for all times less than t and $\boldsymbol{R}(t)$ is again a diagonal matrix whose nonzero elements are the error variances of the observation process. Because of the integral in $\hat{\boldsymbol{h}}_t$, eq. (11) is integro-differential. Further, because the innovation $d\boldsymbol{z}_t - \hat{\boldsymbol{h}}_t\,dt$ is random, (11) is a stochastic integro-differential equation. Proof that (11) is the limiting case of the Bayes process was given by Florchinger and LeGland (1990).

3.2. Numerical methods

The Bayes scheme requires a solution to the Fokker–Planck equation (2) as a prior pdf. As shown in Section 2, the simple centered scheme for (2) is unsatisfactory, since it results in negative values of the pdf. Chang and Cooper (1970) proposed an implicit finite difference method for the solution of the Fokker–Planck equation (2) which was specifically designed to conserve probability and preserve positivity. Details of Chang and Cooper's method are given in Subsection 9.1. A finite element method for the Fokker–Planck equation was implemented by Bergman and Spencer (1992); see also Bergman et al. (1996) and Wojtkiewicz et al. (1995). Bergman and colleagues did not encounter the problem of loss of positivity. They attribute this to the fact that they do not retain the Fokker–Planck equation in conservative form; rather they expand the divergence term explicitly (S. Wojtkiewicz, personal communication).

We implemented Chang and Cooper's method only in our one-dimensional example. Gridded methods become increasingly cumbersome as the dimension of the system grows, and direct integration of the Fokker–Planck equation by gridded methods becomes impractical for more than about 5 dimensions. Our primary interest is in models of the atmosphere and ocean, and even the most schematic of these models cannot be represented in terms of a handful of state variables. We therefore used a Monte-Carlo method for all of the results displayed in this paper beyond initial com-

parisons with Chang and Cooper's method for a one-dimensional example.

All Monte-Carlo methods involve the generation of an ensemble of trials. It is therefore necessary to implement some method of calculating values of the pdf at arbitrary points. We use the adaptive kernel method to accomplish this task in our low-dimensional examples, but computational considerations force us to resort to a fixed kernel method for the case of our stochastically perturbed spectral model. A brief description of the adaptive kernel method is given in Subsection 9.2. More detail on kernel methods in general can be found in Silverman (1986).

A generalization of the EKF which is related to the Bayes approach was suggested by Evensen (1994). In Evensen's generalization, an ensemble of random initial conditions is constructed at the initial time, and then allowed to evolve according to a stochastic differential equation with some noise model. When observations become available, the ensemble covariance is calculated and used to calculate the Kalman gain, which is in turn used to update the state estimate according to the standard EKF formula. As Evensen showed, the posterior pdf can be approximated by updating each member of the ensemble by the Kalman filter formula. Once each member of the ensemble has been updated, the cycle continues as before until the next observation becomes available. Evensen found this method to perform better than the EKF in a series of simulated prediction experiments with a quasigeostrophic model at comparable or lower cost.

Another modified EKF based on a Monte-Carlo method was described by MGG. They modeled the local nonlinear effects in the Lorenz system as a stochastic forcing term, and performed a Monte-Carlo calculation to estimate the statistics of that stochastic term. They were then able to calculate an augmented system noise covariance matrix which could be used in the standard EKF.

For a linear model, Evensen's method can be shown to approach the Kalman filter in the limit as the number of members of the ensemble grows. The calculation of the evolution of the ensemble covariance is exactly a Monte-Carlo estimate of the covariance of the solution of the Fokker–Planck equation, from which the standard Kalman–Bucy expression for evolution of the error covariance is easily derived. The Kalman

updating formulas can then be derived from Bayes' theorem (Lorenc and Hammon, 1988). In the Bayes method, no prior assumption is made about the evolution of the higher moments.

4. The double-well

As a simple example of data assimilation in a strongly nonlinear system we will first consider the stochastically forced double-well model (see MGG),

$$dx = -4x(x^2 - 1) \, dt + q^{1/2} \, dw, \quad (12)$$

where w represents a Wiener process. Without stochastic forcing, this equation describes a system with two stable states, separated by an unstable critical point at zero. The stable states can be thought of as potential wells for the potential

$$F(x) = x^2(x^2 - 2). \quad (13)$$

The deterministic form of (12) can then be written:

$$\dot{x} = -\nabla F. \quad (14)$$

If no forcing is present, for any initial condition but the unstable steady state $x(0) = 0$, x will approach one of the stable steady states. In the presence of stochastic forcing, the value of x stays in the vicinity of one or the other of the stable solutions for the noise free system ($x = \pm 1$) for a while, but eventually the random fluctuations drive the system into the potential well for the other stable steady solution. In our experiments, we choose $q = 0.24$. With this value of q one can

calculate the mean residence time within a given well (Gardiner, 1983) to be 4600 time units.

The evolution of the pdf for (12) is given by the Fokker–Planck equation:

$$\frac{\partial P}{\partial t} = \frac{\partial}{\partial x} \left[4x(x^2 - 1)P \right] + \frac{1}{2} q \frac{\partial^2 P}{\partial x^2}. \quad (15)$$

The steady state solution to (15) can be derived by integrating the right-hand-side twice with respect to x, noting that the integration constant must vanish to ensure normalization. The resulting steady-state is given by:

$$P(x) = N \exp\left[-\frac{2x^2(x^2 - 2)}{q} \right], \quad (16)$$

where N is a normalization constant.

Fig. 1 shows the steady solution (16), plotted along with the results of long-time integrations by Chang and Cooper's (1970) difference method and by the Monte-Carlo method. The solid curve plotted in this figure was evaluated from an ensemble of trials by the adaptive kernel method, as noted above. The width of the two peaks is controlled by the noise level q. This solution is clearly unique and can easily shown to be stable (Risken, 1984). A numerical integration of the Fokker–Planck equation (15) will therefore approach this curve for large time, regardless of the initial condition of the system. The finite difference result is obviously smoother than that obtained from the Monte-Carlo calculation. The Monte-Carlo result could be smoothed further by using broader kernels, at a corresponding cost in resolution. Smoother results could also be
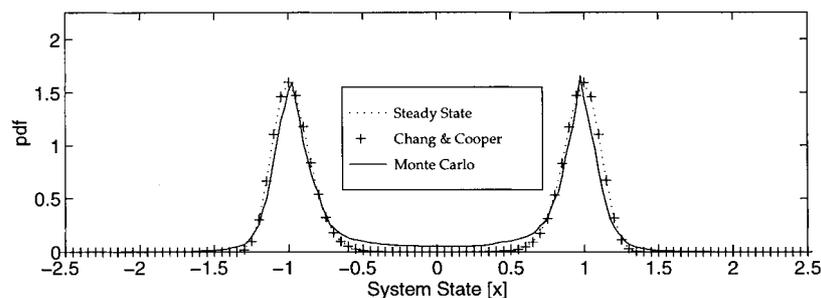

Fig. 1. Comparison of computations of the steady solution of the Fokker–Planck equation for the double-well with stochastic forcing. The forcing has variance $q = 0.24$. Solutions by Chang and Cooper's finite difference method (plus signs) and by a Monte-Carlo simulation with 5000 trials (solid curve) are shown along with the analytical solution (dotted curves). Numerical solutions are for an initial value problem with Gaussian initial conditions with mean zero and variance 0.1. The equation was integrated for a time interval of 1.75.

obtained from the use of a larger ensemble, at greater expense. This smoothness is probably not important for our purposes. The curve shown is the result of an experiment with 5000 trials.

Fig. 2 depicts solutions of (15) for the pdf of the stochastically perturbed double-well, drawn in perspective as a function of x and t, for Gaussian initial condition, with mean 0.0 and variance 0.1.

The evolution of the pdf from a Gaussian pdf to the bimodal equilibrium distribution (16) can be interpreted as reflecting the increasing uncertainty of the randomly perturbed system. Ultimately, the system will approach one equilibrium point or the other, with equal probability. In this view we also see the greater smoothness of the finite difference solution (panel a) compared to the Monte-Carlo
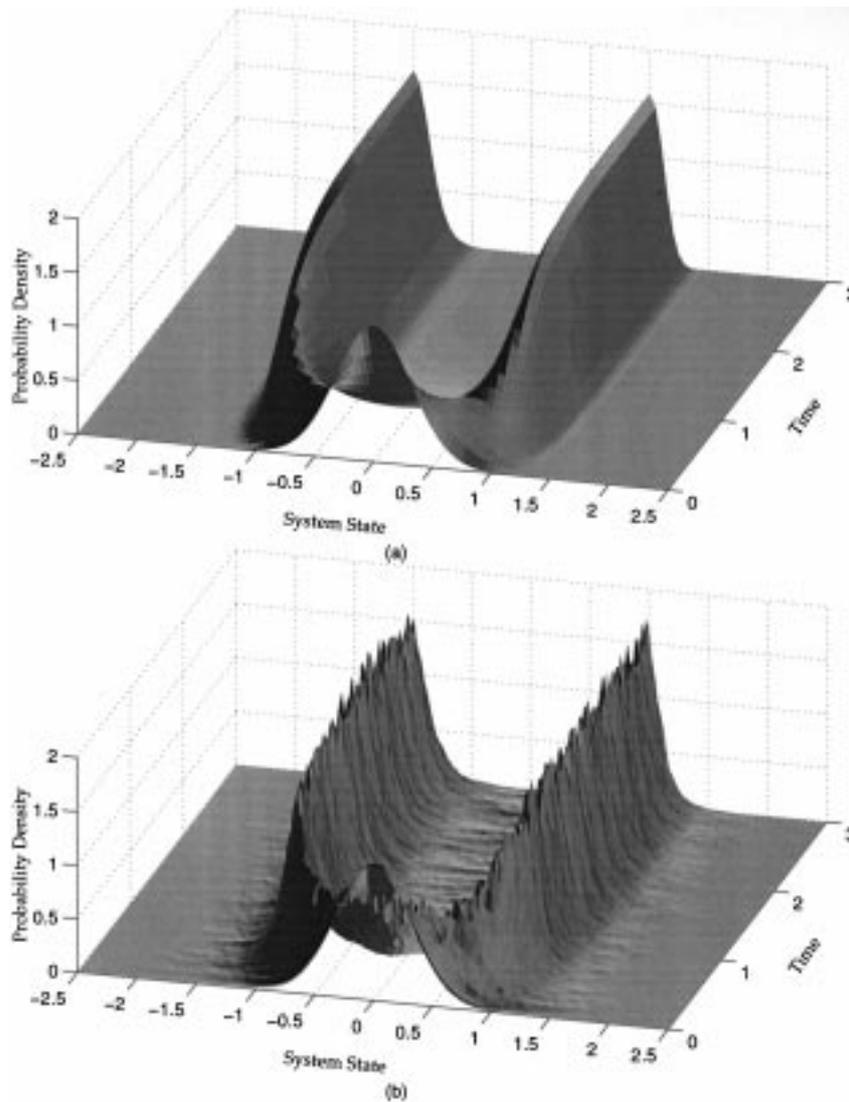
Fig. 2. Solution of the Fokker–Planck equation for the double-well with Gaussian initial condition as in Fig. 1. Note convergence to the bimodal steady state. (a) The Chang and Cooper implicit finite difference method. (b) The Monte-Carlo method.

result (panel b). It is also clear in this view that there is little difference between the modes, medians and confidence intervals in the finite-difference and Monte-Carlo cases.

As Figs. 1 and 2 show, the double-well problem has a pdf that is far from Gaussian. As a consequence the EKF will not perform well, as shown by MGG. A nonlinear assimilation scheme is therefore required. In this section we show comparisons of nonlinear filtering with the EKF and with Evensen's ensemble method.

For the discrete assimilation of data into this system, we will take observations at time intervals of one unit and use an observation noise variance of 0.1. This is a case in which, according to the semi-analytical criterion derived by MGG, the EKF will be expected to fail. The results of this assimilation experiment are shown in Figs. 3, 4.

Fig. 3 shows the conditional pdf from the Bayes' assimilation. The best estimator for the system is the mode of the pdf. Note that the expected value is consistently biased towards zero from the actual value. As long as there is some probability that the system is in the other well, the expected value will be slightly shifted towards the other well. In addition, as Fig. 3 shows, the pdf falls off more slowly on the side towards the other well than it does on the outermost side. Again the result is

that the expected value will be biased towards the opposite well.

Fig. 4 shows a comparison of the general non-linear filter by Bayes' theorem, the EKF and Evensen's ensemble method. For Bayes' assimilation, we choose the mode of the pdf as our estimate. As expected, the EKF failed to follow the transition from the well at $x = -1$ to that at $x = 1$. Both the ensemble method and the Bayes' assimilation track the transition successfully, but the ensemble method is one observation behind the Bayes method.

5. A stochastic Lorenz model

A somewhat more relevant but still highly simplified model is a stochastic version of Lorenz' (1963) now classic spectral truncation of the Boussinesq equations,

$$dX = \sigma(Y - X)\,dt + g_1\,dw_1, \tag{17}$$

$$dY = (\rho X - Y - XZ)\,dt + g_2\,dw_2, \tag{18}$$

$$dZ = (XY - \beta Z)\,dt + g_3\,dw_3, \tag{19}$$

where σ is the Prandtl number, ρ is a normalized Rayleigh number and β is a non-dimensional wavenumber. We choose the parameters first used
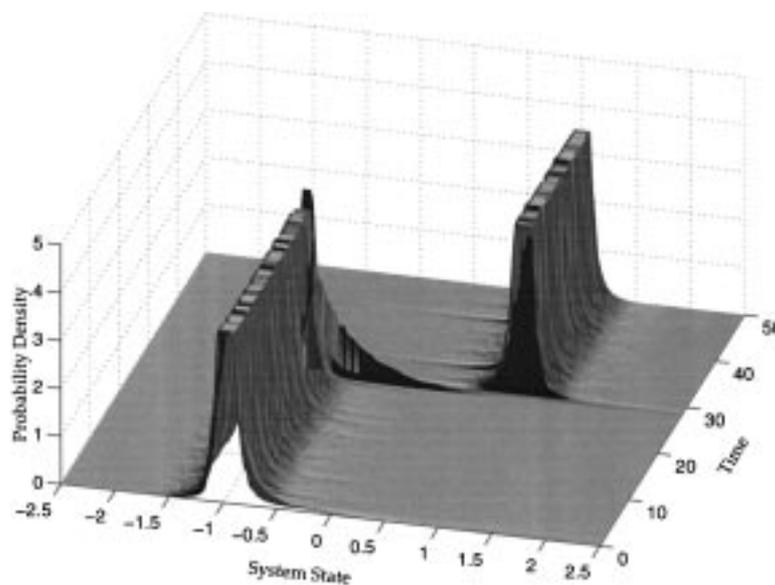

Fig. 3. Evolution of the conditional pdf for the double-well with observations, computed by Bayes' assimilation.
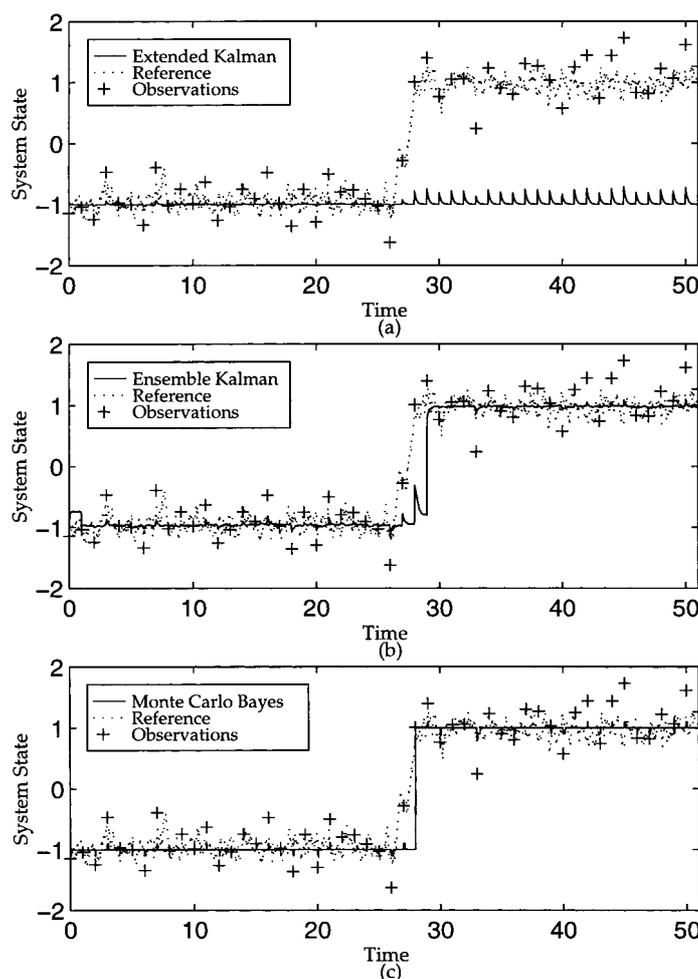
Fig. 4. Comparison of the EKF, the ensemble method and nonlinear filtering by Bayes' theorem for the double-well problem.

by Lorenz to obtain chaotic solutions: $\sigma = 10.0$, $\rho = 28.0$ and $\beta = 8/3$. The stochastic forcing given by: $d\boldsymbol{w} = (dw_1, dw_2, dw_3)^\mathrm{T}$ is a stationary white noise process with covariance given by the identity. We choose $g_1 = g_2 = g_3 = (0.5)^{1/2}$. The Fokker–Planck equation for this system can be obtained by substituting the vector valued function $\boldsymbol{f}(X, Y, Z) = (\sigma(Y - X), \rho X - Y - XZ, XY - \beta Z)^\mathrm{T}$ into (2), along with $\boldsymbol{Q} = \mathrm{diag}(0.5, 0.5, 0.5)$. We constructed a reference solution by solving (17)–(19) by Milshtein's (1978) method (see also Kloeden and Platen, 1992) with initial conditions given by $(X, Y, Z) = (-5.91652, -5.52332, 24.5723)$, and noise gener-

ated by applying the Box–Muller transformation (Box and Muller, 1958) to uniformly distributed pseudo-random numbers obtained from the random number generator R250 (Kirkpatrick and Stoll, 1981). The duration of that integration was 45.0 time units. The reference solution was sampled at intervals of 0.48, and Gaussian noise with variance 2.0 was added to the samples to form a set of simulated noisy observations. With the parameters chosen, the solution is dominated by oscillations of roughly unit period. These simulated observations were then used to perform data assimilation experiments.

A Monte-Carlo simulation with 1750 trials was

used to calculate a solution to the Fokker–Planck equation for the stochastically perturbed Lorenz system (17)–(19). The pdf was calculated explicitly by the adaptive kernel method; cf. Subsection 9.2 or Silverman (1986). Fig. 5 shows a contour map of the pdf on the $z = 33$ plane. In this representation, the bimodality of the attractor, sometimes referred to in terms of the two wings of a butterfly, is clear. A more familiar view of the butterfly wing shape can be found in Ghil and Childress (1987). A similar figure of the unperturbed system, along with an analysis of the Lorenz system subject to systematic perturbation was given by Palmer (1993). As we shall see later on, $Z = 33$ is near the maximum attained by the system in any given oscillation The maxima in this picture correspond to those values of X and Y where the corresponding $\dot{Z}$ is small.

A different view of the pdf of the Lorenz system is shown in Fig. 6. In this figure, the integral of the pdf over all z is shown as a function of x and y. Here the bimodal structure is not apparent. Examination of (17)–(19) shows that the Z-axis

is invariant in the unperturbed system; near the origin of the X–Y plane, $\dot{X}$ and $\dot{Y}$ are small. This accounts for the maximum at the origin in this plot. Careful examination of a well resolved plot of this quantity shows that the slopes of the slice curves along the major axes of the elliptical contours depicted in this plot flatten out significantly. Rather than resembling a bell curve, as one might guess from this figure, the slice curves have distinct shoulders. This is what one would expect from a system of two harmonic oscillators with overlapping ranges, in which the decision of which harmonic oscillator is active is determined by the toss of a fair coin. A figure corresponding to a slice curve of Fig. 6 for the unperturbed system appears in Fig. B2 of MGG.

A slice along $Y = 0$ is shown in the central panel of Fig. 7. This figure exhibits the expected symmetry. As noted above, $\dot{X}$ is small near the Z axis and the maximum at $X = 0$ is expected. It is also expected that the maximum density be reasonably far from the highly unstable saddle at the origin.

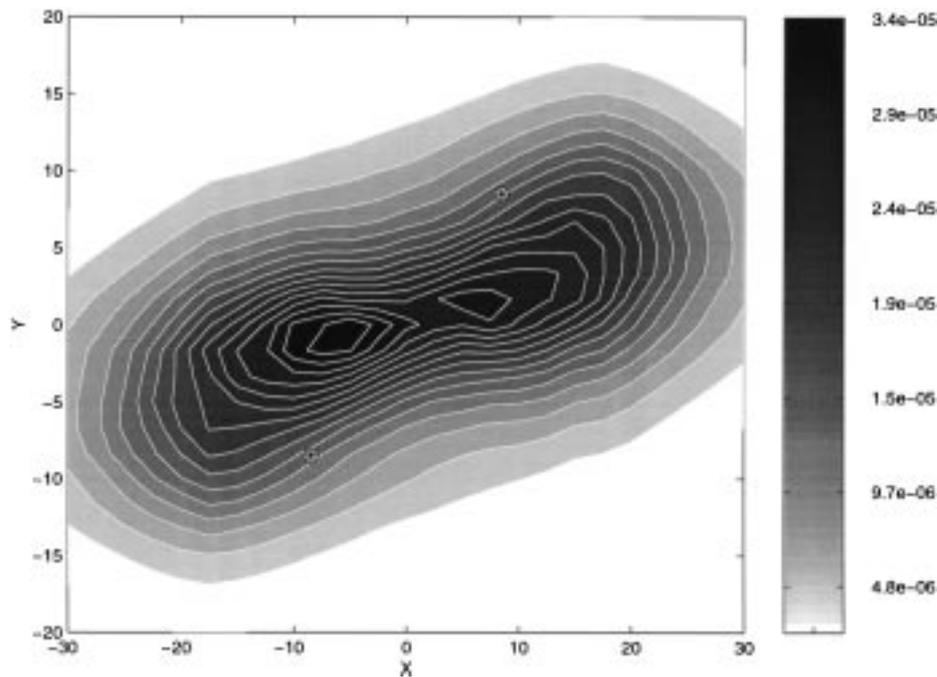Three methods were used to assimilate data in

Fig. 5. Contour map of a slice of the probability density function for the stochastically perturbed Lorenz equations. A map in X–Y coordinates on the plane $z = 33$ is shown following an integration of 46 non-dimensional time units. At this time this picture is relatively steady. Asterisks "*" denote the locations of the critical points in the X–Y plane.
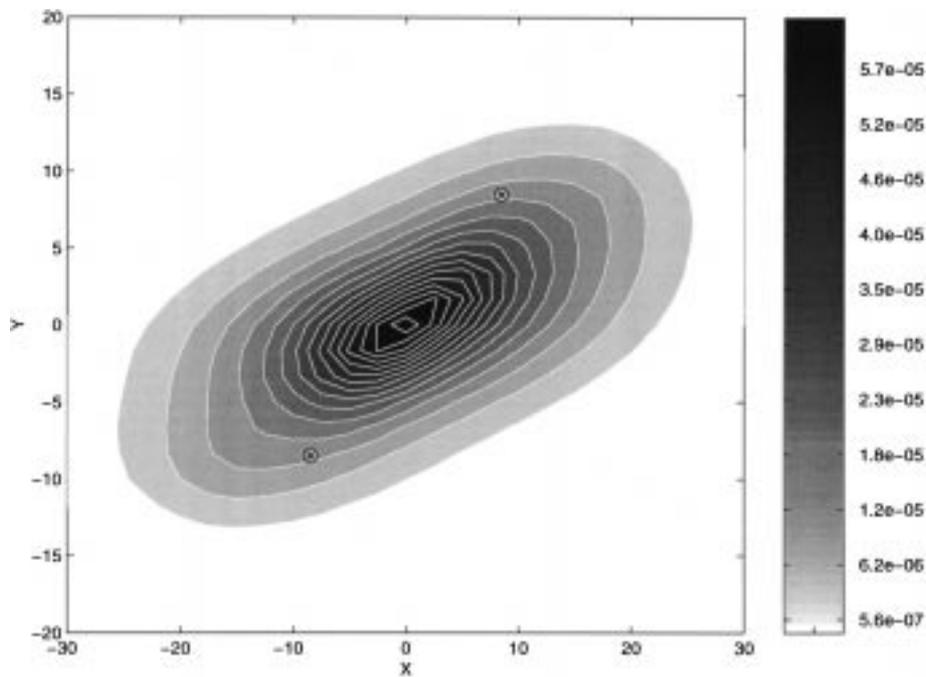
Fig. 6. Contour map of the probability density function for the stochastically perturbed Lorenz equations, integrated in the Z direction. The quantity depicted on this map is actually $\int P(X, Y, Z)\,dZ$ following an integration of 46 non-dimensional time units.

the stochastically perturbed Lorenz system: the EKF, Evensen's ensemble method and the Monte-Carlo–Bayes method, as before. In all three cases, the noise characteristics were assumed to be known exactly, and data were assimilated for all three state variables.

The top and bottom panels of Fig. 7 show the explicit results of the Bayes' assimilation scheme. These panels reflect another view of the butterfly wing structure of the attractor. Panel (a) shows a slice along $Y = 0$ of the conditional pdf in a case in which the system is observed on the left butterfly wing. Panel (c) shows a similar picture, but for a different time in the simulation. At that time, the system was observed in the right butterfly wing. When there is no specific information about the state of the system at a given time, there is no way of knowing whether the system state point is in the left or right wing. This is the meaning of the symmetry in the central panel.

Fig. 8 shows the reference solution along with the results of the three data assimilation experiments for the X state variable. As in MGG, the

EKF fails to follow the reference solution. The results of the EKF experiment appear in panel (a). This panel shows several missed transitions between $t = 20$ and $t = 30$. The reader should note that this experiment has less frequent observations than those shown in MGG by nearly a factor of two. In a preliminary experiment similar to the one shown in MGG, i.e., with observations taken at intervals of 0.24 but with the stochastic terms in (17)–(19) present, the EKF performed better than it did in the noise-free case described in MGG.

The results of the experiment with the ensemble Kalman filter are shown in panel (b). The ensemble method performed better than the EKF. The only obvious qualitative error in panel (b) is the spike in the reference solution near $t = 22$ which the ensemble method does not follow. It does not, however, appear to ignore the data, as the EKF does near $t = 26$. It is curious that the EKF does reproduce the spike near $t = 22.0$, despite the fact that its performance does not match that of the ensemble method for the experiment as a whole.
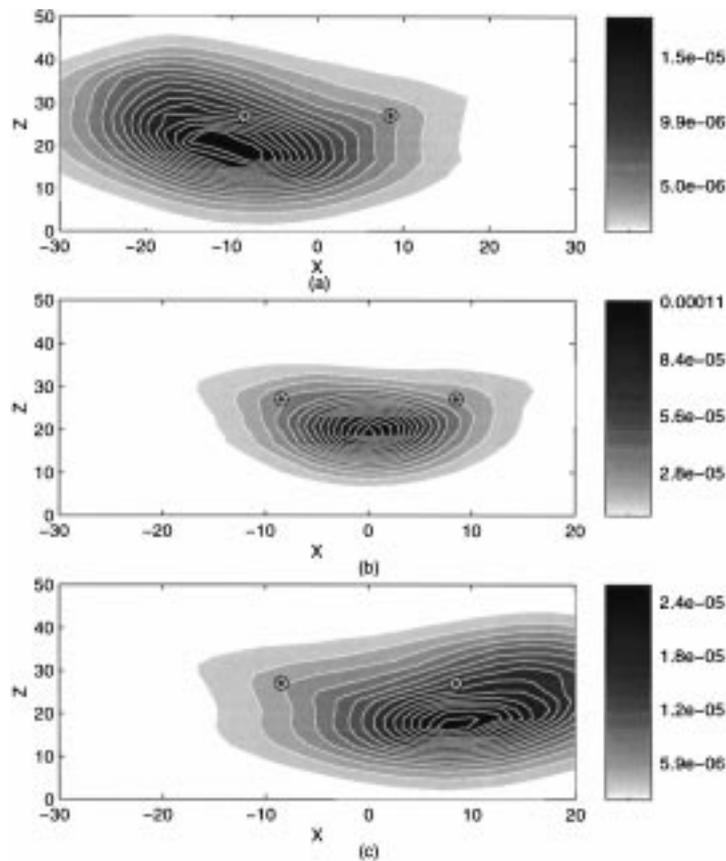
Fig. 7. This figure shows slice curves of contour surfaces of the pdf's for the Lorenz equation in three different situations. Slice curves through $Y = 0$ are shown. The top panel is the conditional pdf for the system observed near the equilibrium point shown on the left in this view. The middle panel is the equilibrium pdf for the system with no observations, and the bottom panel is like the top, but for the opposite butterfly wing.

The results of the Bayes' assimilation technique are shown in panel (c). Here the mode of the pdf determined by the Bayes' assimilation is taken as the estimate of the system state. This scheme captures the spike which is not present in the ensemble case. The lack of smoothness in panel (c) is due to the relatively coarse temporal interval on which the mode was calculated. A kernel calculation is necessary to evaluate the mode, and these are relatively expensive. When the mean instead of the mode of the pdf derived from the Bayes calculation was used as the estimate of the system state, the spike in the reference solution near $t = 22.0$ was also missed.

Fig. 9 is similar to Fig. 8, but for the Z state variable. Regime transitions are usually indicated

in the Z variable, and here also the difference between the EKF and the ensemble method is visible. The EKF captures, but the ensemble method misses the peak in the reference Z near $t = 22$ which signals the downward spike in the X variable. This is the occasion of the worst performance of the ensemble method. While the analysis goes nearly through the data point, the ensemble method underestimates the maximum of Z in that vicinity. On the other hand, it follows nicely for most of the interval, while the EKF is distinctly wrong in both X and Z variables from about $t = 23$ to $t = 28$. No such errors are present in the Bayes' assimilation. All three methods follow closely from about $t = 28$ to $t = 35$.

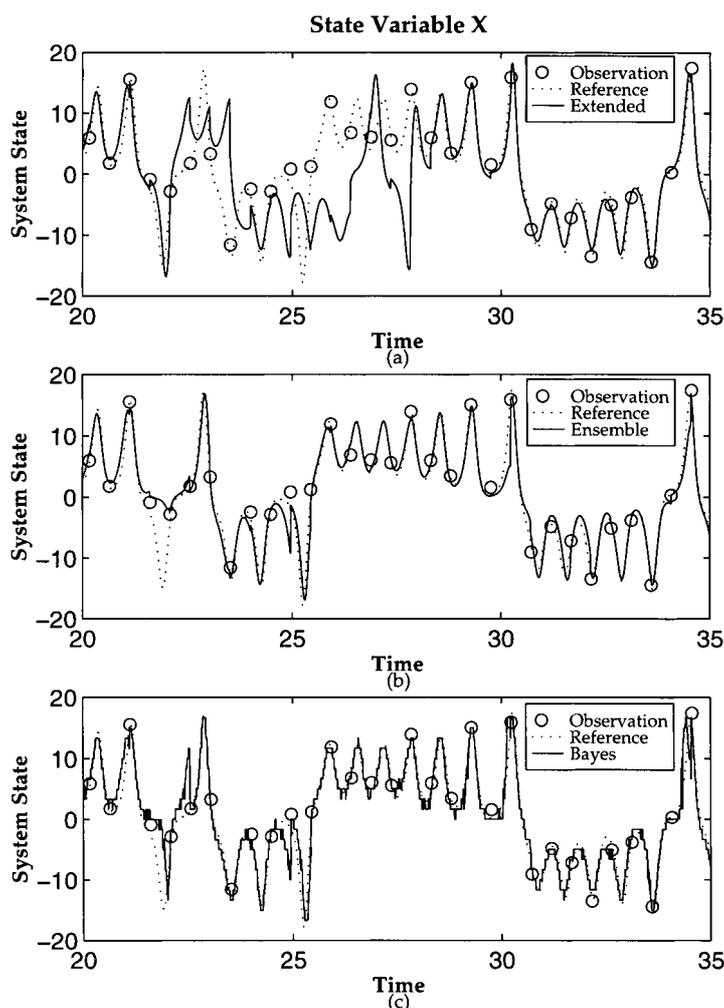Fig. 10 shows a view in state space of the

Fig. 8. Results of data assimilation experiments. State variable *X*, reference solution, observations, and output of data assimilation scheme. (a) The extended Kalman filter (EKF). (b) The ensemble EKF. (c) The mode of the conditional distribution, calculated by Monte-Carlo and Bayes' theorem.

operation of the Bayes' assimilation method. Contour surfaces are shown in shaded perspective of the solutions to the Fokker–Planck equation (i.e., no updating) and the Bayes' assimilation run. The larger shaded surfaces in the center of Fig. 10 show the solution to the Fokker–Planck equation, near equilibrium. Note that these surfaces contain very little of the structure of the attractor. The reader should note further that a plane perpendicular to the *z* axis at $z = 33$ would intersect the largest of the shaded surfaces in a curve similar to one of the contour lines in Fig. 5.

One contour surface of the Bayes pdf is shown as the small dark shape at the upper left of Fig. 10. This surface contains nearly the entire support of the conditional pdf, i.e., the integral of the Bayes pdf over that volume is very close to 1.0. The observation is shown as the black polygon inside the dark shape. A movie of these contour surfaces would show the dark surfaces following the observations as they trace out the shape of the attractor. The trajectory would appear similar to that shown in Figs. 5.10 to 5.12 of Ghil and Childress (1987).
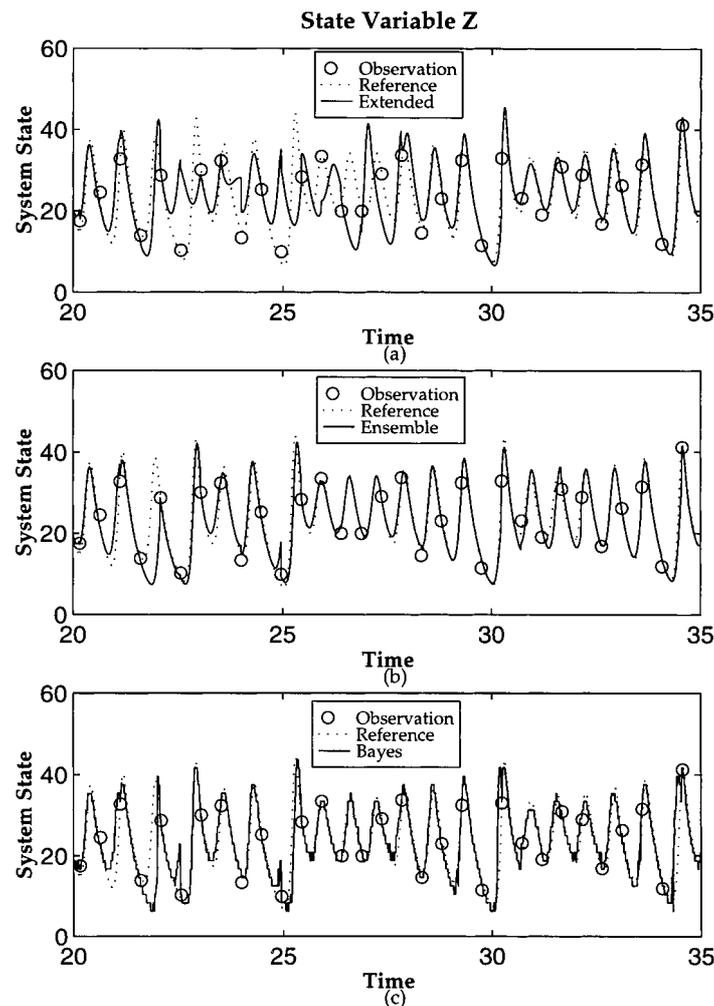
State Variable Z

Fig. 9. Similar to Fig. 6, but for state variable *Z*.

6. A truncated spectral barotropic model

The examples presented so far were chosen to suggest specific aspects of the expected behavior of models of the ocean and atmosphere. In this section we examine a truncated spectral model of a barotropic system in a periodic channel with sinusoidal topography on a β-plane. While still a simplified system, this example is drawn directly from a model of the mid-latitude atmosphere. We therefore expect it to exhibit the features of practical oceanic and atmospheric models more faithfully than the previous examples.

Similar systems have been investigated since the work of Charney and DeVore (1979) and Pedlosky (1981) for the purpose of elucidating the occurrence of multiple equilibria in oceanic and atmospheric models. For more recent treatments, including baroclinic extensions, see Jin and Ghil (1990) and references therein. Here we work with a system described by Gravel and Derome (1993).

The model describes the evolution of the deviation from uniform flow with speed u^*. The total streamfunction is given by $\Psi = -u^*y + \phi(x, y, t)$. Dissipation is by Rayleigh friction with time constant τ. H is the mean depth of the fluid, and the
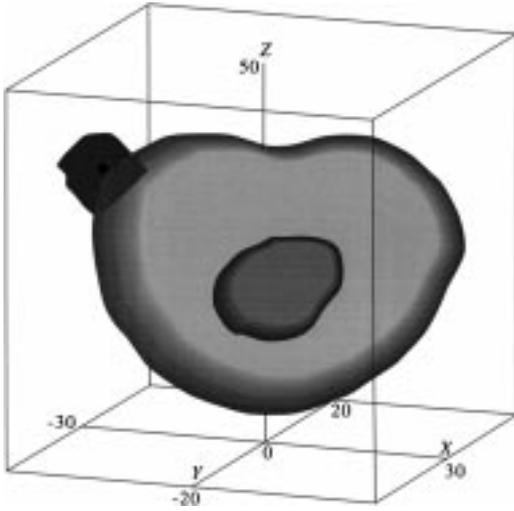
$$u^* \sim U, \tag{24}$$

$$h \sim H. \tag{25}$$

In this scaling (20) becomes:

$$\frac{\partial}{\partial t} \nabla^2 \phi + J(\phi, \nabla^2 \phi + (\beta^* \alpha/R_0)y + (\alpha/R_0)h)$$

$$+ u^* \frac{\partial}{\partial x} (\nabla^2 \phi + (\alpha/R_0)h) = -\frac{1}{\tau^*} \nabla^2 \phi, \tag{26}$$

where:

$$\alpha = D/L \qquad \text{the aspect ratio,} \tag{27}$$

$$\beta^* = \beta D/f_0 \qquad \text{dimensionless } \beta, \tag{28}$$

$$R_0 = U/f_0 L = \text{the Rossby number,} \tag{29}$$

$$\tau^* = U\tau/L \qquad \text{the dimensionless decay time.} \tag{30}$$

The Laplacian operator becomes $\nabla^2 = \alpha^2 \, \partial^2/\partial x^2 + \partial^2/\partial y^2$.

The perturbation streamfunction ϕ is expanded spectrally:

$$\phi(x, y, t) = \sum_i x_i(t) G_i(x, y), \tag{31}$$

where the basis functions G are from the set,

$$G_i \in \{\sqrt{2} \cos \pi ny, \ 2 \sin 2\pi mx \sin \pi ny,$$
$$2 \cos 2\pi mx \sin \pi ny\},$$
$$m = 1, 2, ..., M, \qquad n = 1, 2, ..., N. \tag{32}$$

We refer to the three forms of the G_i, $\sqrt{2} \cos \pi ny$, $2 \sin 2\pi mx \sin \pi ny$, $2 \cos 2\pi mx \sin \pi ny$ as the G_α, G_β and G_γ respectively.

The resulting equations for the spectral components are:

$$\frac{dx_i}{dt} = \frac{1}{a_i^2 R_0} \left\{ u^* \sum_j b_{ji}(\alpha h_j - R_0 a_j^2 x_j) \right.$$

$$+ \sum_{j,k} c_{ijk} x_j (\alpha h_k - R_0 a_k^2 x_k)$$

$$\left. + \beta^* \alpha \sum_j b_{ji} x_j - (R_0/\tau^*) a_i^2 x_i \right\}. \tag{33}$$

The coefficients a_i, b_{ji} and c_{ijk} are given by:

$$a_i^2 = 4\alpha^2 m^2 \pi^2 + n^2 \pi^2, \tag{34}$$

$$b_{ji} = \int_x \int_y G_j \frac{\partial G_i}{\partial x} \, dy \, dx, \tag{35}$$

$$c_{kij} = \int_x \int_y G_k J(G_i, G_j) \, dy \, dx, \tag{36}$$

Fig. 10. Contour surfaces of pdf's of Lorenz equations from Monte-Carlo and Monte-Carlo/Bayes' experiments. Outer surface: 1.28×10^{-5} contour. Inner surface: 1.09×10^{-4} contour. The small dark surface at the upper left is the $p = 0$ contour of the conditional pdf, determined by Bayes' theorem. This surface contains the entire support of the conditional pdf, i.e., all Monte-Carlo–Bayes' trials fall within it. The small black surface contained in the dark contour is centered at the observation point.

variable topography is given by h. The perturbation streamfunction ϕ then evolves according to the equation of conservation of potential vorticity, which we write in the form:

$$\frac{\partial}{\partial t} \nabla^2 \phi + J\left(\phi, \nabla^2 \phi + \beta y + \frac{f_0}{H} h\right)$$

$$+ u^* \frac{\partial}{\partial x}\left(\nabla^2 \phi + \frac{f_0}{H} h\right) = -\frac{1}{\tau} \nabla^2 \phi$$

Following Gravel and Derome, we chose the parameters of our channel to mimic the mid-latitude atmosphere at 45°N. Our channel has length $L = 28\,000$ km and width $D = 4000$ km. At this latitude, we have the Coriolis parameter $f_0 = 1.03 \times 10^{-4}$ s^{-1} and β, the meridional rate of change of the Coriolis parameter given by: $\beta = 1.6 \times 10^{-8}$ s^{-1} km^{-1}.

We adopt the following scaling:

$$U = \text{scale speed,} \tag{21}$$

$$x \sim L, \qquad y \sim D, \qquad \phi \sim UD, \tag{22}$$

$$t \sim L/U, \qquad \text{the advective time scale} \tag{23}$$

where m and n in (34) are the zonal and meridional wavenumbers corresponding to the ith spectral component.

Following Gravel and Derome, we use a spectral expansion with 5 zonal wavenumbers and 4 meridional wavenumbers ($M = 5$ and $N = 4$), and we choose parameter values that mimic the mid-latitude atmosphere. In this model, $H = 10$ km and h is chosen to have peak amplitude of 500 m with functional form given by $2 \sin(2\pi x/L) \sin(\pi y/D)$. With $\tau = 22$ days, the qualitative structure of the system is highly complex.

The stability of the equilibrium solutions is governed by Hopf bifurcations in the parameter u^*. We follow Gravel and Derome, who used the arclength continuation method described by Legras and Ghil (1985). The bifurcation diagram for the equilibria is shown in Fig. 11. In this figure, for steady solutions of (33), the RMS kinetic energy of the perturbation streamfunction ϕ is plotted against u^*. As expected, the perturbation streamfunction ϕ draws energy from the mean flow most efficiently at speeds near the linear Rossby wave resonance at 24 m s^{-1}.

The equilibria of the system lose stability through a Hopf bifurcation near 18 m s^{-1} and regain stability through another Hopf bifurcation near 28 m s^{-1}. For this choice of parameters, we found the same result as Gravel and Derome, i.e.,
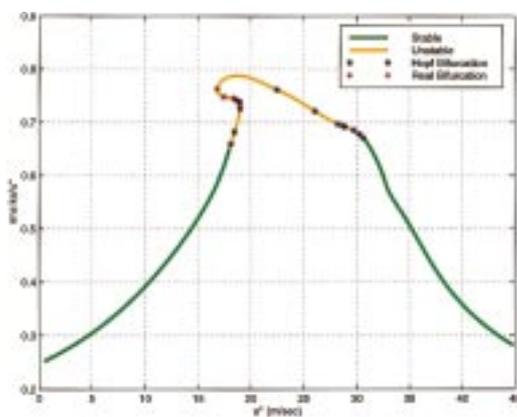
there is no choice of u^* for which multiple stable equilibria were found.

There are, however, stable limit cycles in much of the range in which the equilibria are unstable, and at least one case in which multiple stable limit cycles coexist at the same parameter values. We calculated these limit cycles and determined their stability by a method similar to that described by Strong et al. (1995). The bifurcation diagram for the limit cycles is shown in Fig. 12. This figure shows that there are at least two branches of stable limit cycles. The interval in the parameter u^* in which the branches overlap is small, but not empty. Near $u^* = 20$ m s^{-1} the bifurcation diagram for limit cycles becomes very complex.

Examples of limit cycles projected into three dimensions are shown in Fig. 13. Two of the limit cycles shown here are stable, and the other two are not. The bifurcation diagram shown in Fig. 12 is by no means complete. We believe that there are still other limit cycles at this value of u^*, but did not pursue the calculation of other branches.

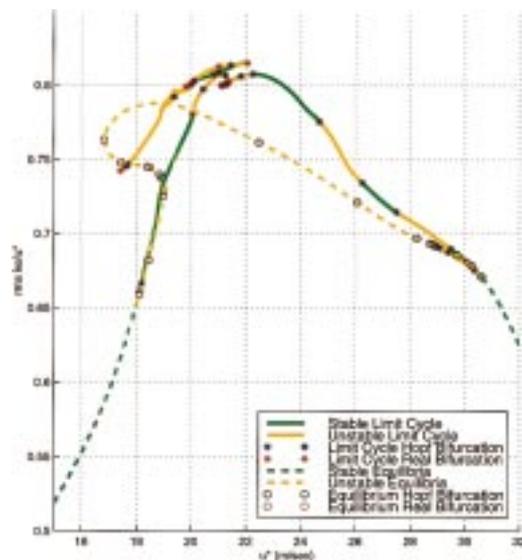Fig. 12 shows that Hopf bifurcations of branches of limit cycles occur at a number of

Fig. 11. Bifurcation of equilibria for truncated spectral approximation to the β-plane channel with sinusoidal topography. Abscissa is the speed of the mean flow; ordinate is the ratio of the RMS kinetic energy of the perturbation to the speed of the mean flow.

Fig. 12. Bifurcation diagram of limit cycles for truncated spectral approximation to the β-plane channel with sinusoidal topography. Axes are similar to those in Fig. 11 but perturbation kinetic energy for a periodic orbit is averaged over a period. The bifurcation of equilibria is shown as a dashed line.
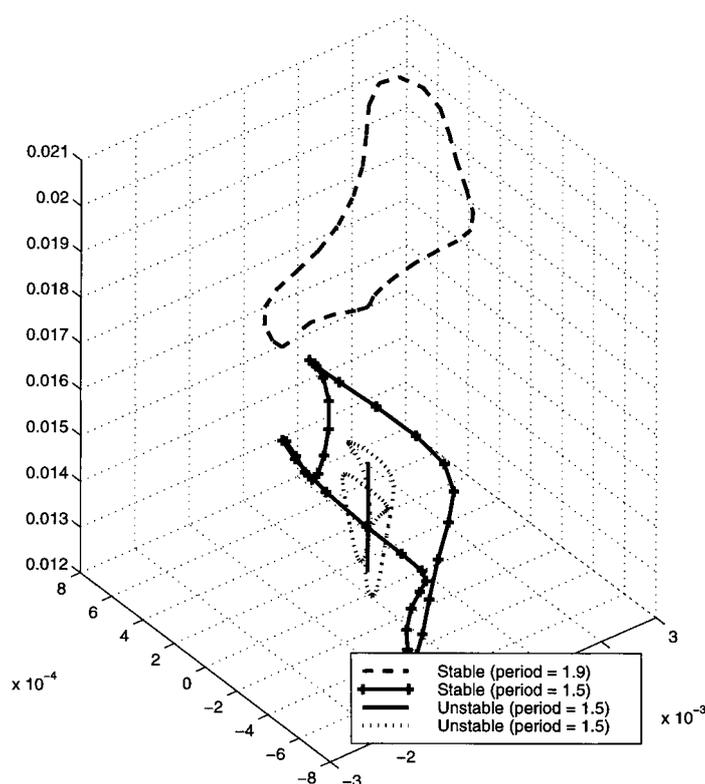
Fig. 13. Four limit cycles. Two are stable and the other two are not. Two of the coordinates are the real and imaginary parts of the eigenvector corresponding to the most unstable mode of the equilibrium point at this parameter value $u^* = 19.98$. That equilibrium point is the origin of the present plot. The third dimension was constructed by calculating the vector average of the differences between points on the most stable limit cycle, in the case the one shown on top, and the equilibrium point, and then applying the Gram–Schmidt process to determine a direction orthogonal to the other two.

points on the diagram. These bifurcation points in some portions of this diagram mark regions in which we have not been able to find stable equilibria or limit cycles. Long integrations of the model in those parameter ranges result in complicated patterns which we have been unable to resolve into limit cycles or equilibria. A Hopf bifurcation from a limit cycle signals either the existence of a new limit cycle whose period is an integer multiple of the original one, or an invariant torus, depending on whether the argument of the eigenvalue of the return map which crosses the unit circle is a rational or irrational multiple of π. As one would expect, it is therefore all but impossible to distinguish an invariant torus from a tortuous periodic orbit which arose by a period multiplying bifurcation from a simple limit cycle. Fig. 14 shows

an explicit example of a trajectory we believe to lie on an invariant torus. We have not investigated the stability of the invariant torus itself. We cannot rule out the possibility that this trajectory arose from a period multiplying bifurcation of a limit cycle at a nearby parameter value, and is therefore a long limit cycle.

We now consider the system (33) with stochastic perturbations added. We expect this system to behave qualitatively like the double-well when u^* is chosen in a regime with multiple stable limit cycles. Before we look explicitly for transitions from one limit cycle to another, let us examine the local behavior of the system, without assimilation, in the neighborhood of one of the limit cycles.

For a stochastic system whose deterministic part is characterized by an attractor consisting of
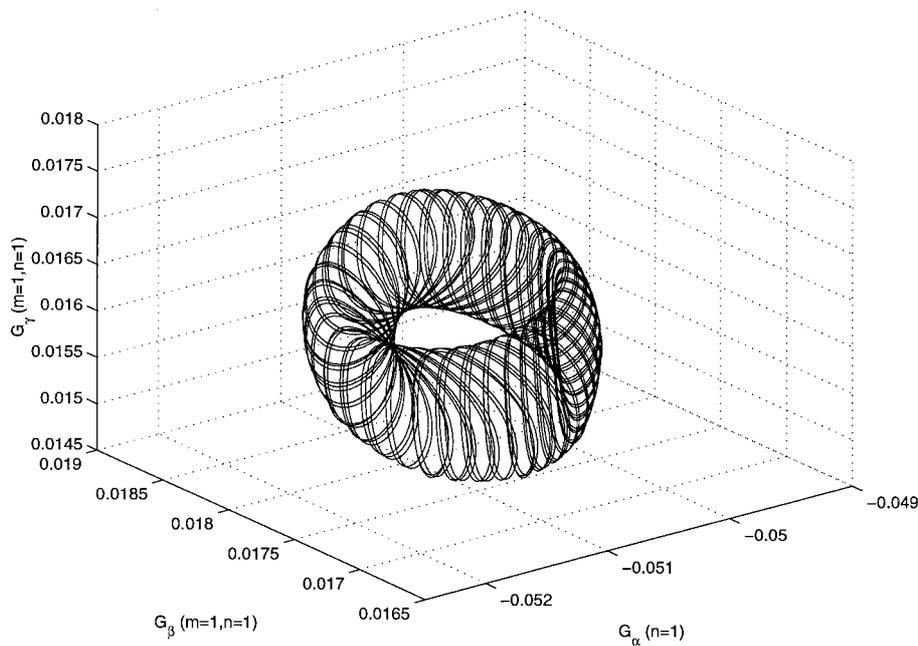
Fig. 14. Projection into three dimensions of a trajectory believed to lie on the surface of an invariant torus. The three coordinates are amplitudes of the spectral components shown on the axis labels. Here $u^* = 24 \text{ m s}^{-1}$. We cannot entirely rule out the possibility that this is a long period limit cycle.

a single stable limit cycle, we expect the equilibrium pdf to be supported in a neighborhood of the attractor itself, but distributed in some fashion around the limit cycle. Perturbations in directions transverse to the limit cycle will tend to decay, while the system will be neutrally stable to perturbations tangential to the limit cycle. Therefore, we expect that any initial phase information will eventually be lost. Eventually, we expect the highest probability regions of state space to be near the limit cycle, but we will have trouble determining exactly where on the limit cycle the system is.

This point was illustrated by Spencer and Bergman (1993) for the van der Pol oscillator, perhaps the most extensively studied limit cycle oscillator. Their solution of the Fokker–Planck equation for the stochastically perturbed van der Pol oscillator shows the expected behavior. Their initial pdf is chosen to be Gaussian, centered outside the limit cycle. The initial pdf is rapidly attracted to the limit cycle, and is, in general, elongated in the direction tangential to the limit cycle and compressed in the transverse direction. The details of its behavior as it proceeds around

the limit cycle are complex, but there are two features to notice. First, after 1.25 periods, the longest time shown in Spencer and Bergman's figure, the pdf has not spread out very far along the limit cycle; in other words, not much phase information has been lost due to the stochastic perturbation. Second, Spencer and Bergman's calculation of the stationary solution to the Fokker–Planck equation, illustrated in their Fig. 9, shows the expected roughly cylindrical distribution. The highest points on the perspective graph of the pdf lie on the limit cycle, but some are higher than others, due to the fact that the angular speed of the phase point around the limit cycle is not uniform, and the system spends more time in some quadrants than in others.

Spencer and Bergman solved the Fokker–Planck equation in two space dimensions. For calculations in systems with very high dimension, the representation of the Fokker–Planck equation is critical in determining whether the calculation can be implemented on any modern computer. Clearly, a regular grid of points can only be stored if the dimension of the lattice is small. We can,

however, calculate ensembles of trajectories of the stochastic system, and evaluate the functionals of the pdf directly on regular grids imposed on low-dimensional subspaces of state space. Fig. 15 shows the result of such a calculation. We performed that calculation with the initial pdf chosen to be multivariate Gaussian, centered at a point on the more stable of the limit cycles shown in Fig. 13. We generated an ensemble of 10 000 points from that pdf, and integrated forward for ten periods. We then chose the two-dimensional subspace of the full 44 dimensional state space determined by the amplitudes of two spectral components. These particular components were chosen by examining the projections of the limit cycle into two-dimensional subspaces of this form, and choosing a pair of coordinates which gave a good graphic representation of the limit cycle. It is then a simple matter to impose a regular grid on this two-dimensional subspace, and perform a

kernel calculation of the pdf resulting from the distribution of these two coordinates of the ensemble, ignoring the other 42. This amounts to an approximation of the integral of the pdf over the other 42 dimensions. Fig. 15 is shown normalized so that the largest value is unity. It is likely that some information is lost in this representation. Comparison of Figs. 5 and 6 shows that this was the case for the pdfs of the stochastic Lorenz model.

At that point in the integration, ten periods from the initial time, the pdf was still localized at one end of the trajectory, i.e., phase information from the initial ensemble remains. Further insight into this figure can be gained by comparison to Fig. 8 of Spencer and Bergman (1993). We clearly cannot say from this figure whether there is significant probability of transition from the basin of attraction of this limit cycle to the basin of attraction of another stable limit cycle.
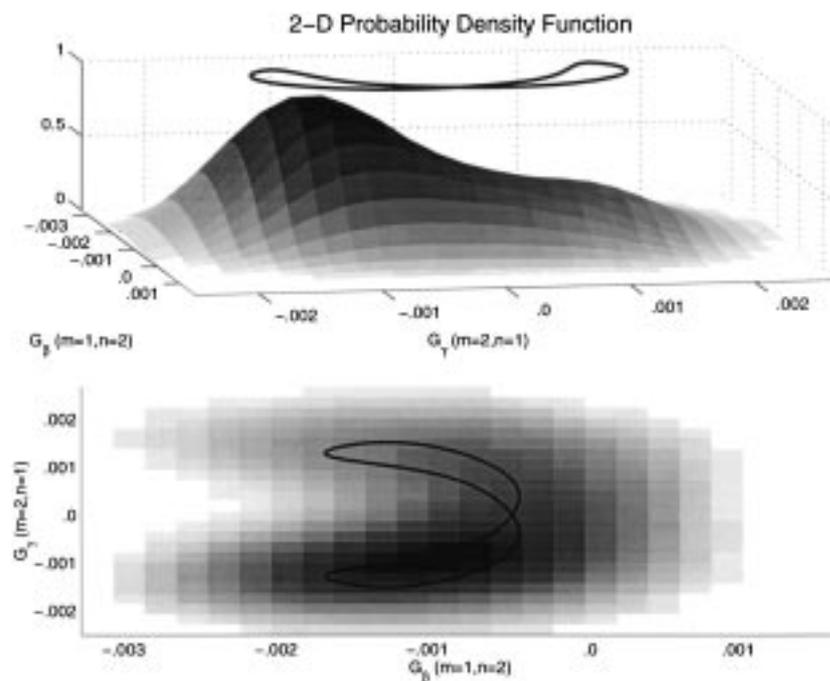

Fig. 15. Two views of the marginal pdf of the truncated spectral model in the neighborhood of the more stable limit cycle, following integration for 10 periods. The horizontal axes are the amplitudes of two spectral components which exhibit the limit cycle clearly. The values plotted at each gridpoint represent the integral at that gridpoint over the other 42 dimensions. Top: three-dimensional plot, shown in perspective of the marginal pdf. The limit cycle projected into this coordinate plane is shown as a solid line above. Bottom: two-dimensional view of marginal pdf. The limit cycle is shown superimposed as a solid line.

While projections of the pdf into selected subspaces can give us insight into the nature of the evolution of the pdf, they are not sufficient for study of the assimilation process. If we look at eqs. (8) or (11), we see that a high dimensional integration will be necessary in order to proceed with the assimilation. Such integrals will be done using Monte-Carlo sampling on a quasi-random set of points (quasi-random points are used instead of pseudo-random points, because this gives considerably improved integral estimates for a given number of points) (Press and Teukolsky, 1989). This means that we will only use the values of the pdf $P(x, t_k | y_k)$ on the quasi-random lattice. We therefore calculate the pdf only on these points. As a result, the number of points at which we calculate the pdf is limited by the available memory of the computers being used and is not directly related to the dimensionality of the problem. We further adapt the size of the domain (keeping the number of points fixed) as the calculation proceeds. The result of this is that the density of sampled points goes up as the probability distribution sharpens, so the resolution improves where it is most needed.

While the actual implementation of these calculations on a quasi-random lattice requires careful software design, we have encountered no fundamental limitation to this type of implementation with the models that we have attempted to date. We clearly have a long way to go in the area of visualization. A major obstacle to understanding of high dimensional stochastic systems is the difficulty of graphical representation of the results.

In order to make our data assimilation experiments as realistic as possible, we have performed our simulated observations in physical space. Our simulated observations were constructed by sampling the full perturbation streamfunction ϕ at nine points in physical space. The array is shown in the top panel of Fig. 16.

The center and bottom panels of Fig. 16 show the results of observing the two stable limit cycles shown in Fig. 13. From these figures we see that a clear distinction between the two stable limit cycles appears in observations of the system in physical space. There is therefore some hope that we will be able to use data assimilation to track the system through a transition from one regime to the other.

The stochastic differential equation does indeed exhibit transitions from one stable limit cycle to another. Two views of such a transition are shown in Fig. 17. The top panel of that figure shows the transition in phase space, projected into two dimensions. The bottom panel shows the result of observing the transition at point number 2 in the physical space array shown in Fig. 16. The example illustrates the fact that the system behaves qualitatively like the double-well. We therefore expect to encounter the same type of problems in assimilating data into this model as we did in the double-well.

From our experience with the double-well, we expect the extended Kalman filter to be most likely to fail in the case of small system noise and large observation noise (see MGG and section of the present manuscript). For our simulations, we chose the system noise to be the smallest which would produce a transition in a reasonable amount of time, and we chose the observation noise to be as large as we could make it and still be small enough that the observations would distinguish the basins of attraction of the stable limit cycles from one another. For the experiments presented here, we chose the system noise to have RMS amplitude equal to 20% of the RMS amplitude of the more stable limit cycle, and the same spectral characteristics in wavenumber space as that limit cycle.

Observations were taken at the nine points shown in Fig. 16 at a time interval of one half the period of the less stable of the two limit cycles. The periods of the two limit cycles are similar, and the less stable one was chosen because the system started out near it. Noise was added to the synthetic observations. The observation noise was assumed to be white in time and uncorrelated from one observation to another. The observation noise variance was chosen by a rough subjective criterion. We examined graphs of noisy observations, i.e., the points marked by stars in the lower panels of Figs. 18 and 19, and chose the observation noise variance to be the largest value consistent with our ability to distinguish between the two limit cycles by inspection.

Fig. 18 shows the result of the application of the extended Kalman filter to the truncated spectral system. The extended Kalman filter was implemented in the square root form (see, e.g., Bierman, 1977). The square root filter is equivalent to the extended Kalman filter, but is more stable numer-
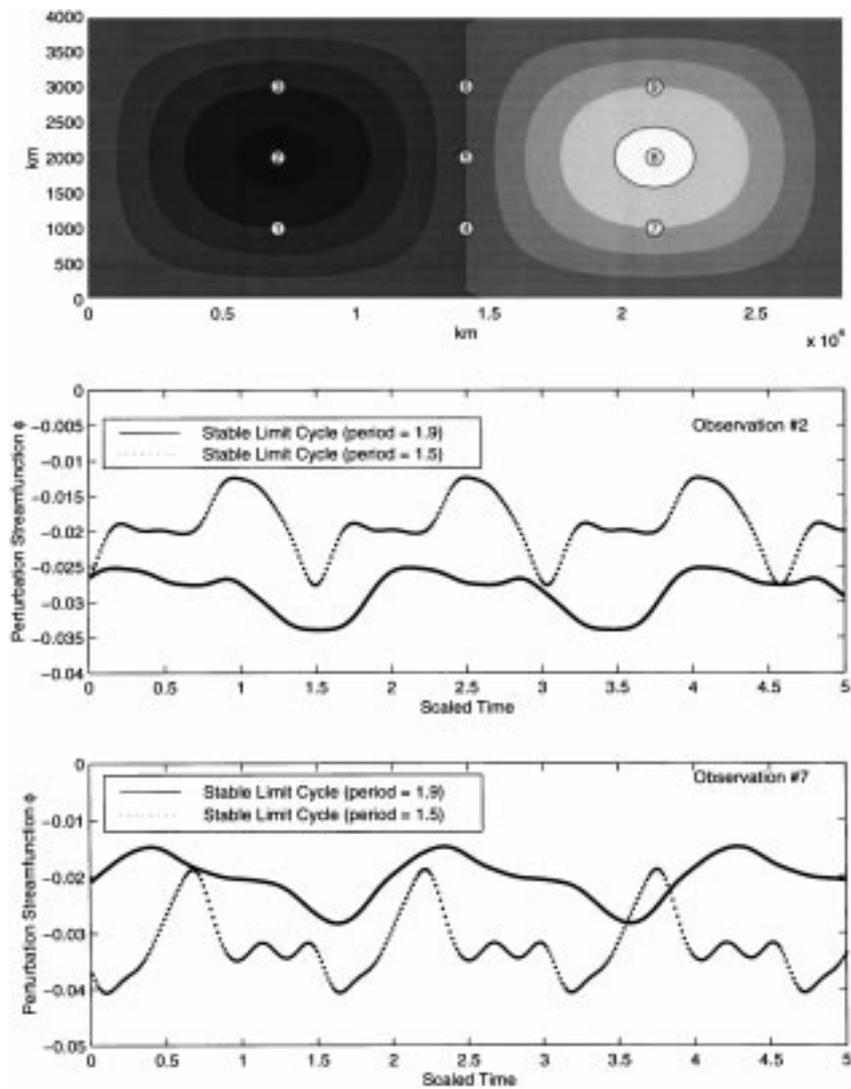
Fig. 16. Simulated observations for the periodic channel: Top: the nine observation points, superimposed on a shaded map of the bathymetry. Center: result of observation of the stable limit cycles at point number 2. The blue curve shows the result of observing the two stable limit cycles shown in Fig. 13. No observation noise is imposed here. Bottom: similar to center panel, but for observation at the point number 7.

ically. We did not observe the instabilities reported by Evensen (1992), Gauthier et al. (1993) or Bouttier (1994), probably because of our fairly dense uniformly spaced observing array.

The top and bottom panels of Fig. 18 show clearly that the filter estimate does not follow the state transition, even though this transition is clearly indicated in the observations. The reader should compare these panels to Figs. 4 and 8, which show similar failures of the extended Kalman filter.

The ensemble Kalman filter devised by Evensen and colleagues (Evensen, 1994; Evensen and Fario, 1997; and references therein) also fails in this case. We used 100 members in our ensemble, as Evensen (1994) did, and found that
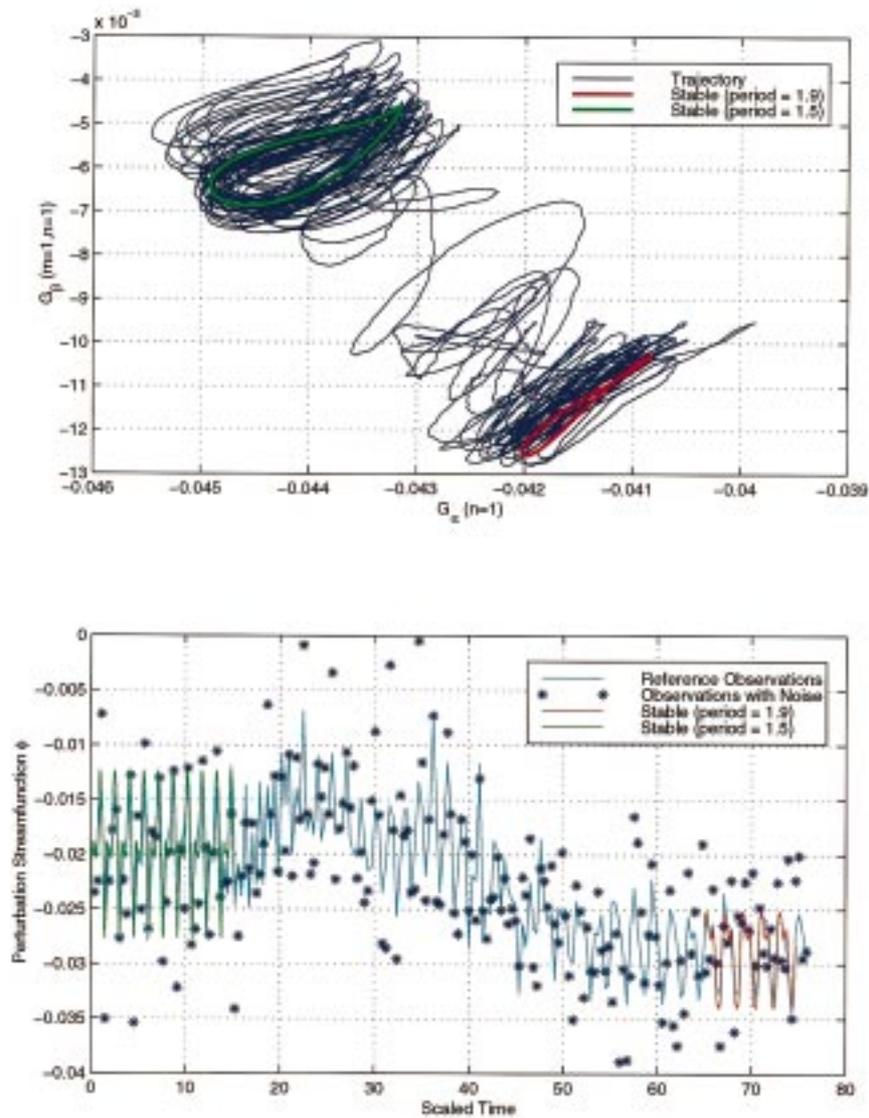
Fig. 17. Top: trajectory in phase space of a realization of the stochastic system undergoing a transition from the basin of attraction of one stable limit cycle to that of another, shown here projected into two dimensions. The trajectory is shown in blue and the limit cycles are shown in green and red. The system begins in the basin of attraction of the limit cycle shown in green. Bottom: observations taken continuously at point number 2 of the top panel of Fig. 16 of the same transition shown in the upper panel. No noise is added to the reference observations. Observations of the exact unperturbed limit cycles are shown in green and red to emphasize the transition.

the method failed in exactly the same fashion as the EKF. This may be due to the relatively small ensemble, but as noted here, this ensemble size is typical of applications of this method reported in the literature. Graphs corresponding to Fig. 18 for the ensemble Kalman filter are nearly identical to those resulting from the EKF and are therefore not shown here.

Results of the Bayes' assimilation are shown in Fig. 19. The estimate of the state given by the
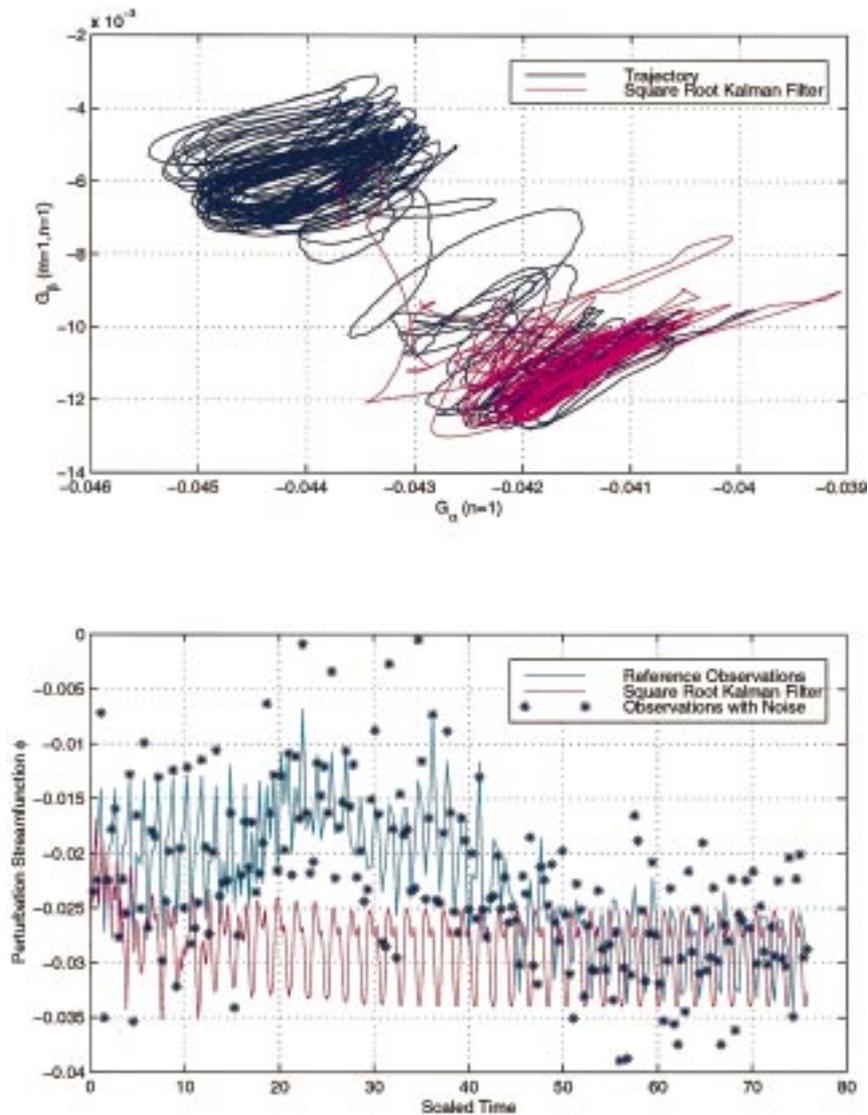
Fig. 18. Results of extended Kalman filter experiment for the periodic channel. Top: projection of result of the square root implementation of the EKF into state space, shown in magenta, along with the reference trajectory, shown in blue. Bottom: output of the EKF, along with the reference solution and the observations for observation at point number 2 in Fig. 16. Observations are taken at time intervals of about 0.375, and contaminated with random noise.

mode of the conditional pdf stays roughly in the general vicinity of the correct limit cycle, as opposed to losing track entirely, but fails to track the observations faithfully. There are two possible causes for this: either the errors in our approximation of the conditional pdf's in the Bayes run were insufficiently accurate, or the noisy observa-

tions did not contain sufficient information to track the phase space trajectory faithfully. We used 10 000 trials in the estimate of the pdf, and this may be too few. Results may also be improved by fine tuning of the kernel method, but previous experience does not lead us to be optimistic about this approach.
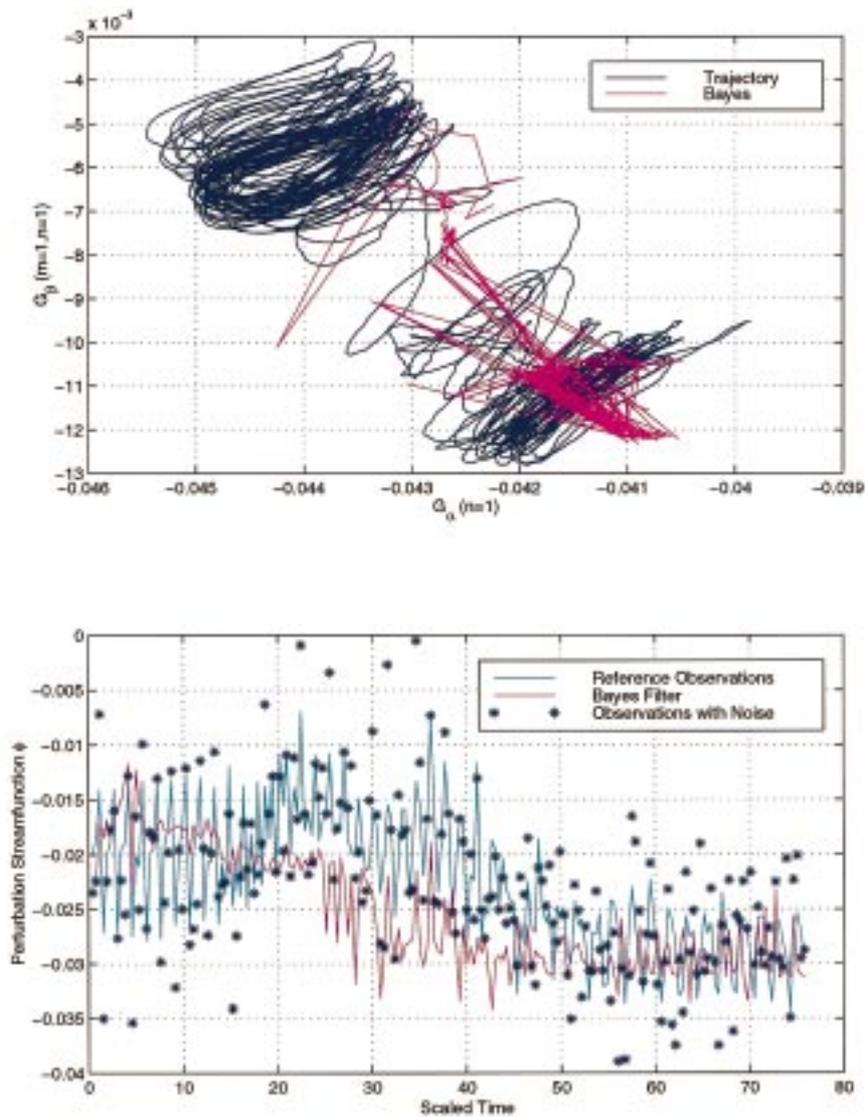
Fig. 19. Result of Bayes' assimilation experiment. Legend similar to Fig. 18. The results of the assimilation shown here are the modes of the conditional pdf's computed at each observation.

We believe that the apparently poor performance of the Bayes system in this case is due to the information content of the noisy observations. Fig. 20 shows the limit cycles and the observations in a plane defined by two of the measurements. In the case of vanishing or very small system noise, one would expect the noise-free observations to cluster about the limit cycles. Examination of the upper panel of Fig. 20 shows that while the observations are concentrated in the general area of the limit cycles, they do not appear to follow the limit cycles closely. Evidently the system noise is too large for that. Given this picture, it should not be surprising that the best we can do with the Bayes' assimilation is to stay in the right general part of phase space. The addition of observation
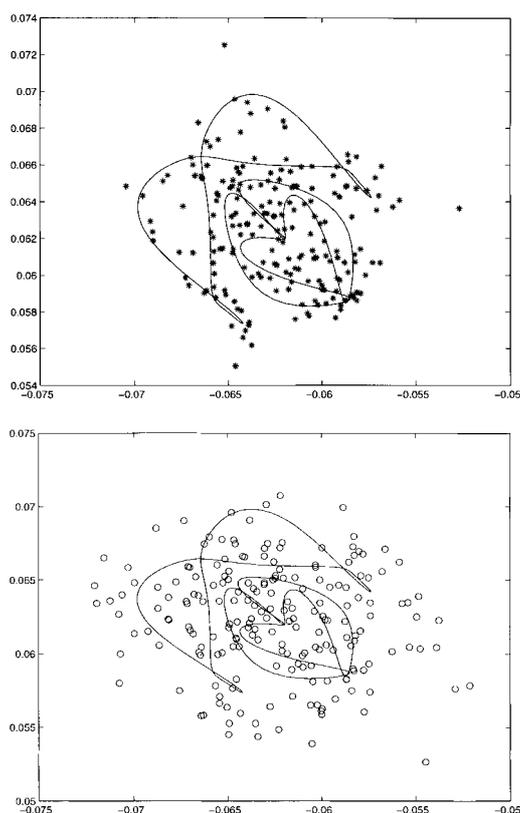
Fig. 20. Phase plot of Bayes' experiment. The horizontal axis is observation no. 1 and the vertical axis is observation no. 9; see Fig. 16. The solid curves depict the limit cycles. The curve to the upper left is the more stable limit cycle, and the one to the lower right is the less stable. Top panel: limit cycles shown with noise-free observations of the stochastic system, shown here as "*". Bottom panel: limit cycles shown with noisy observations, shown here as open circles.

noise only makes the situation worse, as shown in the lower panel of Fig. 20.

7. Discussion

Nonlinear filtering, viz. calculation of the pdf of the stochastic system, conditioned upon observations, was implemented by explicit calculation of an approximate solution to the Fokker–Planck equation between updates, and updating by Bayes' theorem when observations became available. This technique was applied to a series of increasingly complex examples, beginning with a simple scalar system and ending with a coarsely resolved spectral truncation of a barotropic channel model with topography on the mid-latitude β plane.

This data assimilation method is general enough to handle nonlinear systems and observation processes. It can also work with complicated noise processes such as correlated or multiplicative noise.

For high dimension systems the Fokker–Planck part of the calculation must be implemented using Monte-Carlo methods. Because of this, it is most practical to implement the method with either massively parallel computers or with distributed processor systems such as PVM (Geist et al., 1991). In high dimension the interpretation/visualization problem is not trivial.

While this method failed to track the phase of the limit cycles in the most complicated of our examples, it did not lose track entirely, as the extended Kalman filter did. The extended Kalman filter is based on the assumption that the dynamics of paths near the reference trajectory are well described by linearizing the full evolution operator about the current estimate. It is possible to discard this assumption without abandoning the least squares formulations entirely, and thus gain some of the benefits of the simplicity of least squares systems. The simplest of these systems are the ensemble methods devised by Evensen and colleagues (Evensen, 1994; Evensen and Fario, 1997). These schemes can work well when the extended Kalman filter fails. Failures of ensemble schemes of this type, when they occur, are as likely to be the result of the ensemble mean being a less satisfactory estimator than the maximum likelihood as they are to result from the underlying least squares assumptions in the Kalman gain formula. One must also consider the fact that most of the examples in the literature of ensemble calculations were performed with small ensembles, i.e., hundreds of trials at most, while our Bayes calculations were performed with 10 000 trials. This still may be too small for detailed calculation of pdf's in cases of practical interest.

Once we accept the stochastic nature of the systems we deal with, we must conclude that the pdf's conditioned upon observations contain all of the information available to us about the system. We are free to choose the statistic which best suits our purpose as our estimate of "the"

state of the system — we favor the mode — but the pdf, as opposed to any time series of estimated states, is really the solution to the problem.

The theory of stochastic nonlinear filtering is well developed but little studied in the oceanic and atmospheric science community. We speculate that this is due to the high dimensionality of even the simplest relevant problems, which renders these calculations cumbersome in the extreme. To our knowledge, our spectral model of the barotropic β plane channel is the most relevant example to the geophysical fluid dynamics community to be studied by these techniques.

In many cases the choice of a specific method for use in a practical data assimilation system is a tactical decision, based on data interfaces and on the details of model implementations. Most of the methods under widespread consideration are closely related if not precisely equivalent (Courtier, 1997). Larger questions of error statistics and representation of the measurement process, i.e., the explicit form of h in (7), are the defining characteristics of practical systems. In this work we have not systematically explored the consequences of uncertainties in these quantities; rather we have focused on the implications for data assimilation of the interaction between nonlinearity and noise. The description of the various noise sources must be considered carefully in evaluating the practical applicability of results of simulation experiments.

It is not our purpose to advocate nonlinear filtering as a competitor to other data assimilation methods for operational use. At this time, it is far too resource intensive for that, but we propose it as a conceptual tool, to be used to gain insight into the performance of approximate schemes in highly nonlinear settings. The reader should be reminded that this was the status of the Kalman filter a decade ago. While this remains so to a large extent today, the most cursory glance at the literature shows application of the Kalman filter and other weak constraint schemes to problems of complexity far beyond anything envisioned by the early investigators of those methods.

9. Appendix: details of numerical methods

9.1. Chang and Cooper's numerical method for the Fokker–Planck equation

Given the one-dimensional Fokker–Planck equation:

$$\frac{\partial p}{\partial t} = \frac{\partial}{\partial x}\left[f(x,t)p + (Q/2)\frac{\partial p}{\partial x}\right], \tag{37}$$

we can find a solution by Chang and Cooper's finite difference scheme in three steps.

First, solve the ordinary differential equation

$$\frac{\partial}{\partial x}\left[f(x,t)p + (Q/2)\frac{\partial p}{\partial x}\right] = 0. \tag{38}$$

Second, calculate the δ_j^{n+1}s where $\delta_j^n = \delta(j\Delta x, n\Delta t)$:

$$\left[(1 - \delta_j^{n+1})f_{j+1/2}^n + \frac{1}{\Delta x}(Q/2)\right] p_{j+1}^{n+1}$$

$$- \left(\frac{1}{\Delta x}(Q/2) - \delta_j^{n+1}f_{j+1/2}^n \right) p_j^{n+1} = 0. \tag{39}$$

Last, solve the following implicit difference equation by back substitution.

$$\frac{1}{\Delta t}(p_j^{n+1} - p_j^n)$$

$$= \frac{1}{\Delta x}\left\{ \left[(1 - \delta_j^{n+1})f_{j+1/2}^n + \frac{1}{\Delta x}(Q/2)\right] p_{j+1}^{n+1} \right.$$

$$- \left[\frac{1}{\Delta x}Q + (1 - \delta_{j-1}^{n+1})f_{j-1/2}^n - \delta_j^{n+1}f_{j+1/2}^n \right] p_j^{n+1}$$

$$\left. + \left[\frac{1}{\Delta x}(Q/2) - \delta_{j-1}^{n+1}f_{j-1/2}^n \right] p_{j-1}^{n+1} \right\}. \tag{40}$$

9.2. Adaptive kernel method

The adaptive kernel approach first requires an initial estimate of the probability density function $f(t)$. This inital pdf can be obtained by the kernel or nearest neighbor method, however, since the

adaptive kernel method is insensitive to the inital guess, any estimate can be used. This estimate is used to calculate the local bandwidths. The bandwidths determine the adaptive kernel estimate. The following steps describe the process in detail.

Step 1: Find an initial estimate to the probability density function $f(t)$ such that $f(x_i) > 0$ for all i.

Step 2: Define the local bandwidth factors λ_i by

$$\lambda_i = \{f(x_i)/g\}^{-\alpha}, \qquad (41)$$

where g is the geometric mean of $f(x_i)$,

$$\log g = n^{-1} \sum f(x_i). \qquad (42)$$

The sensitivity parameter α must satisfy the condition $0 \leqslant \alpha \leqslant 1$.

Step 3: Define the adaptive kernel estimate $\hat{f}(t)$ by

$$\hat{f}(x) = n^{-1} \sum_{i=1}^{n} h^{-d} \lambda_k^{-d} K\{h^{-1}\lambda_i^{-1}(t - x_i)\}, \qquad (43)$$

where K is the kernel function and h is the bandwidth.

In this paper, we used the Epanechnikov kernel as the kernel function. The Epanechnikov kernel is defined as

$$\begin{array}{ll} \frac{3}{4}(1 - \frac{1}{5}t^2)/\sqrt{5}, & \text{for } |t| < \sqrt{5}; \\ 0, & \text{otherwise} \end{array} \qquad (44)$$

The sensitivity parameter α was set to $\frac{1}{2}$. As α increases, the differences in bandwidths also increases. The resulting adaptive kernel method is more sensitive to the initial estimate. As α decreases, the bandwidths become uniform, until, for $\alpha = 0$, the adaptive kernel method is reduced to the fixed width kernel method.

REFERENCES

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

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. Wea. Rev.* **121**, 1714–1729.

Bergman, L. A., Wojtkiewicz, S. F., Johnson, E. A. and Spencer, B. F. 1996. A robust numerical solution of the Fokker–Planck equation for second order dynamical systems under parametric and external white noise excitation. *Fields Inst. Comm.* **9**, 23–37.

Bergman, L. A. and Spencer, B. F. Jr. 1992. *Robust numerical solution of the transient Fokker–Planck equation for nonlinear dynamical systems.* IUTAM Symposium, Turin. N. Bellomo and F. Casciati, eds., Springer-Verlag, Berlin, Heidelberg, 49–59.

Bierman, G. J. 1977. *Factorization methods for discrete sequential estimation.* Academic Press, New York. 241 pp.

Bouttier, F. 1994. A dynamical estimation of forecast error covariances in an assimilation system. *Mon. Wea. Rev.* **122**, 2376–2390.

Box, G. E. P and Muller, M. E. 1958. A note on the generation of random normal deviates. *Annals Math. Stat.* **29**, 610–611.

Budgell, W. P. 1986. Nonlinear data assimilation for shallow water equations in branched channels. *J. Geophys. Res.* **91**, 10633–10644.

Chang, J. S. and Cooper, G. 1970. A practical difference scheme for Fokker–Planck equations. *J. Comput. Phys.* **6**, 1–16.

Charney, J. G. and Devore, J. G. 1979. Multiple flow equilibria in the atmosphere and blocking. *J. Atmos. Sci.* **36**, 1205–1216.

Courtier, P. 1997. Dual formulation of four-dimensional variational assimilation. *Q. J. R. Meteorol. Soc.* **123**, 2449–2461.

Courtier, P., Derber, J., Errico, R., Louis, J.-F. and Vukicevic, T. 1993. Important literature on the use of adjoint, variational methods and the Kalman filter in meteorology. *Tellus* **45A**, 342–357.

Courtier, P., Thépaut, J.-N. and Hollingsworth, A. 1994. A strategy for operational implementation of 4D-Var, using an incremental approach. *Q. J. R. Meteorol. Soc.* **120**, 1367–1387.

Evensen, G. 1992. Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model. *J. Geophys. Res.* **97**, 17905–17924.

Evensen, G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**, 10143–10162.

Evensen, G. and Fario, N. 1997. Solving for the generalized inverse of the Lorenz model. *J. Meteor. Soc. Japan* **75**, 229–243.

Fletcher, C. A. J. 1988. *Computational techniques for fluid dynamics*, vol I. Springer Series in Computational Physics. Springer-Verlag, Berlin, Heidelberg. 409 pp.

Florchinger, P. and LeGland, F. 1990. Time-discretization of the Zakai equation for diffusion processes observed in correlated noise. *9th Conference on Analysis and optimization*, A. Bensoussan and J. L. Lions, eds. Lecture Notes in Control and Information Sciences 144, Springer-Verlag.

Gardiner, C. W. 1983. *Handbook of stochastic methods,*

for physics, chemistry, and the natural sciences. Springer-Verlag, Berlin, 442 pp.

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

Geist, G. A, and Sunderam, V. S. 1991. *The PVM System: supercomputing level concurrent computations on a heterogeneous network of workstations.* 6th Distributed Memory Computing Conference Proceedings, IEEE, Portland, OR, April/May 1991, 258–261.

Ghil, M. and Childress, S. 1987. *Topics in geophysical fluid dynamics: atmospheric dynamics, dynamo theory and climate dynamics.* Springer-Verlag, 485 pp.

Gravel, S. and Derome, J. 1993. A study of multiple equilibria in a β-plane and a hemispheric model of a barotropic atmosphere. *Tellus* **45A**, 81–98.

Itô, K. 1951. *On stochastic differential equations.* American Mathematical Society Memoirs #4, 51 pp.

Jazwinski, A. H. 1970. *Stochastic processes and filtering theory.* Academic Press, New York, 376 pp.

Jin, F.-F. and Ghil, M. 1990. Intraseasonal oscillations in the extratropics: Hopf bifurcations and topographic instabilities. *J. Atmos. Sci.* **47**, 3007–3022.

Kirkpatrick, S. and Stoll, E. 1981. A very fast Shift-Register Sequence Random Number generator. *J. Comput. Phys.* **40**, 517–526.

Kloeden, P. and Platen, E. 1992. *Numerical solution of stochastic differential equations.* Springer-Verlag, Berlin. 519 pp.

Kushner, H. J. 1962. On the differential equations satisfied by conditional probability densities of Markov processes, with applications. *SIAM J. Control, Series A* **2**, 106–119.

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

Legras, B. and Ghil, M. 1985. Persistent anomalies, blocking and variations in atmospheric predictability. *J. Atmos. Sci.* **42**, 433–471.

Lorenc, A. C. and Hammon, O. 1988. Objective quality control of observations using Bayesian methods — theory and a practical implementation. *Q. J. R. Meteorol. Soc.* **114**, 515–543.

Lorenz, E. N. 1963. Deterministic nonperiodic flow. *J. Atmos. Sci.* **20**, 448–464.

Miller, R. N. 1998. Introduction to the Kalman filter. In: *Proceedings of the ECMWF Seminar on Data assimilation.* European Centre for Medium Range Weather Forecasting, Shinfield Park, Reading, RG2 9AX UK, 9–11 September, 1996, pp. 47–59.

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

Milshtein, G. N. 1978. A method of second-order accuracy integration of stochastic differential equations. *Theory Prob. Appl.* **23**, 396–401.

Palmer, T. N. 1993. Extended range atmospheric prediction and the Lorenz model. *Bull. Amer. Met. Soc.* **74**, 49–65.

Pedlosky, J. 1981. Resonant topographic waves in barotropic and baroclinic flows. *J. Atmos. Sci.* **38**, 2626–2641.

Press, W. H. and Teukolsky, S. A. 1989. Quasi- (that is, sub-) random numbers. *Computers in Physics* **3**, 76–79.

Risken, H. 1984. *The Fokker–Planck equation.* Springer-Verlag, New York, 454 pp.

Rozovskii, B. L. 1990. *Stochastic evolution systems: linear theory and applications to nonlinear filtering.* Kluwer Academic, 315 pp.

Silverman, B. W. 1986. *Density estimation for statistics and data analysis.* Chapman and Hall, London. 175 pp.

Spencer, B. F. Jr. and Bergman, L. A. 1993. On the numerical solution of the Fokker–Planck equation for nonlinear stochastic systems. *Nonlinear Dynamics* **4**, 357–372.

Stratonovich, R. L. 1966. A new form of representing stochastic integrals and equations. *SIAM J. Control* **4**, 362–371.

Strong, C., Jin, F.-F. and Ghil, M. 1995. Intraseasonal oscillations in a barotropic model with annual cycle, and their predictability. *J. Atmos. Sci.* **52**, 2627–2642.

Wojtkiewicz, S. F., Bergman, L. A. and Spencer, B. F. 1995. Numerical solution of some three-state random vibration problems. *DE-Vol 84-1, 1995 Design Engineering Technical Conferences*, Volume 3, Part A, ASME 1995, 939–947.

Zakai, M. 1969. On the optimal filtering of diffusion processes. *Z. Wahrsch. verw. Gebiete* **11**, 230–243.