Data Assimilation in Nonlinear Stochastic Models

Robert N. Miller

College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, Oregon

Abstract. 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. We hope to develop data assimilation tools which do not depend on prior assumptions about probability density functions (pdf's) of stochastic systems. Here we regard a pdf, rather than a specific realization of a random process, as "the" answer. This turns out to have applications beyond data assimilation, which may turn out to be more useful in the end.

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. The 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.

The high state dimension of most practical ocean models at first glance makes our task seem all but hopeless, but in many cases of interest, the motions of the model system are confined to some low-dimensional subset of the full state space. We appeal to the modern theory of dynamical systems in order to exploit this useful property of many of our model systems.

Within this framework, observations are also considered as pdf's rather than single numerical values. Here 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.

1. Introduction

The immediate goal of this investigation is to gain insight into data assimilation in highly nonlinear stochastic systems. All data assimilation methods in use or proposed so far have some basis in least squares methodology. You can always do least squares, but our reliance on least squares reflects a Gaussian outlook that may or may not be warranted. In the spirit of this exercise, the initial condition is a random field, and even if the initial probability density function (pdf) is Gaussian, the pdf of the stochastic system will not remain Gaussian as the system evolves.

Least squares methods can be made to track the observations by augmenting the model error covariance. This requires some care, and even if the process is successful, it may not be possible to assign reliable confidence intervals based on covariance calculations alone. In models which result in a pdf with more than one maximum, least squares methods tend to split the difference. This may be useless.

This study began with data assimilation, but this thread will lead to the ensemble methods for forecast validation and predictability that are now common in the numerical weather prediction community. Implications of this work for forecast validation and predictability studies may, in the end, turn out to be more important than implications for data assimilation.

A diverse set of tools is needed for this study. We are picking our way here, and expect to stumble a bit. We will need to do stochastic calculations, and we will need to know about dynamical systems, as well as developing an array of computational tools. Visualization will also be a major issue.

2. Tools

2.1. Nonlinear filtering theory

We consider our models to be nonlinear stochastic differential equations, which we write in the following form:
\[ dx_t = f(x_t, t)dt + gdw_t \]  

(1)

where \( x_t \) is the state vector at time \( t \), \( f \) specifies the deterministic model dynamics and \( dw_t \) represents the noise. \( w \) as usual is a Brownian motion with independent Gaussian increments with unit variance. \( g \) is a predetermined matrix, possibly a matrix-valued function of time, which can be considered as a Choleski factor of the covariance of the additive noise in the model. We only consider noise processes of this form.

Clearly (1) cannot be considered as a differential equation in the usual sense, since the difference quotients of successive values of \( x_t \) will almost surely not converge as the time increments decrease to zero. The tools of ordinary calculus are therefore insufficient to deal rigorously with (1). Stochastic versions of the calculus have been devised by Itô (1951) and successors. There are many texts on this subject; see, e.g., Kloeden and Platen (1992). Jazwinski (1970) provides a treatment of the subject which is less rigorous, but more accessible to the applied scientist.

We consider \( x_t \) as a random variable. Within this framework, given noisy observations of the system, our best estimate of the state of the system is some statistic based on the pdf of the solution to (1). The pdf \( P \) for \( x_t \) evolves according to the *Fokker-Planck or Forward Kolmogorov Equation*:

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

(2)

where \( Q = gg^T \); see, e.g., Jazwinski (1970).

Since (2) is a parabolic equation with a number of space dimension equal to the dimension of the state vector \( x_t \), conventional gridded methods are not practical when (1) corresponds to even the simplest model of the real ocean. We therefore must use Monte Carlo methods for problems of primary interest.

The Fokker-Planck equation (2) is an advection diffusion equation, and therefore one expects complications in its numerical solution similar to those encountered in computational fluid dynamics. Since practical models of the ocean and atmosphere often have state dimensions as high as \( 10^5 \), direct solutions by finite difference or finite element methods are entirely impractical. For more than a few dimensions, Monte Carlo methods must be used.

Even when solutions are obtained, it is not clear how to interpret a function of \( 10^5 \) variables once we get it. Our real hope is that the essential behavior of the system is governed by a subsystem with a manageable number of dimensions.

2.2. How to Assistalte Data

We now wish to devise data assimilation schemes based on our assumption that the system is described by a nonlinear stochastic differential equation. Beginning with an initial pdf we integrate the Fokker-Planck equation somehow. When data become available, we can use Bayes’ theorem to construct the posterior pdf:

Assume that at some time, data become available. Observations are assumed to be related to the state \( x_t \) of the system by

\[ y_k = h_k(x_{t_k}, t_k) + \nu_k^{1/2}v_k \]

Observations are themselves random variables; assimilation proceeds by combining pdf’s. Bayes’ Theorem gives us a recipe for doing this. Let \( P(x, t_k | y_k^-) \) be the pdf for the system, conditioned on all observations at times up to but not including \( t_k \). Bayes’ theorem may be written

\[ P(x, t_k | y_k) = \frac{P(y_k | x)P(x, t_k | y_k^-)}{\int P(y_k | x)P(x, t_k | y_k^-) dx} \]

(3)

\( P(x, t_k | y_k^-) \) is known as the prior pdf. The result \( P(x, t_k | y_k) \) is known as the posterior pdf. This is how we actually assimilate data. Bayes’ theorem forms the basis of the theory of nonlinear filtering; see, e.g., Zakai (1969).

If the observation noise \( v \) is Gaussian

\[ P(y_k | x) = \frac{1}{(2\pi)^{m/2} R_k^{1/2}} \exp\left(-\frac{1}{2}(y_k - h(x, t_k))^T R_k^{-1}(y_k - h(x, t_k))\right) \]

Since we expect to calculate pdf’s from ensembles of trials in Monte Carlo experiments, the explicit prior pdf’s needed are calculated by a kernel method. This is essentially a generalization of a simple histogram; see, e.g., Silverman (1986).

**Dynamical Systems**

Our best hope of learning something from nonlinear filtering is the case in which the essential behavior of the system is captured by a low-dimensional subsystem. This is the case in which we expect the Monte Carlo method to be most successful. We shall appeal to the modern theory of dynamical systems, especially the tools of bifurcation theory to help investigate the structure of these low-dimensional subsystems.

As an example of a case in which bifurcation theory can be expected to contribute to our understanding of models of the ocean and atmosphere, we present a simple model of the Kuroshio off the coast of Japan. The Kuroshio is an intense current which occurs in the North Pacific, perhaps best described to the nonspecialist in physical oceanography as the dynamical analogue of the Gulf Stream. Unlike the Gulf Stream, the Kuroshio exhibits two distinct regimes near the coast of Japan. Once either regime is established, fluctuations
about the mean state are fairly small. Both regimes are apparently stable and persist for years; see, e.g., Taft (1972, 1978).

Our model is a very simple one, with quasi-geostrophic barotropic physics similar to that of Chao (1984). Computational details are given by Kumaran and Miller (1995) and Kumaran et al. (1996). Steady solutions of this model show the observed bimodal behavior of the current.

Figure 1 shows contour maps of the large and small meander states, as well as the unstable intermediate state, which is not observed. The maps shown are shaded contour maps of the transport streamfunction. If you consider, say, the 30 Sv contour, then there are 30 Sv total (net backflows) that flow between that contour and the shoreline. We note that 1 Sv = 10⁶ m³/sec; to give this some scale, the total flow of all the rivers in the world combined is about 1 Sv.

Figure 1. Steady states of the barotropic quasi-geostrophic Kuroshio model. Inlet transport is 45 Sverdrups. All parameters are identical in all three panels. Top: stable small meander state; Center: unstable intermediate state; Bottom: stable large meander state.
Figure 2 shows the bifurcation diagram for this model, with the total transport as the bifurcation parameter; the y-axis is the maximum excursion from the shoreline of the half-transport streamline. The curves shown in this figure were calculated by a pseudo-arclength continuation method; see, e.g., Legras and Ghil (1985). You can see that the large meander state is born in a saddle-node bifurcation below 30 Sv, and the small meander state becomes unstable in a Hopf bifurcation above 45 Sv or so.

Given a bifurcation diagram of this form one is tempted to draw the picture picture shown here as Figure 3 to characterize the behavior of the system. In fact, there are theorems which state that this picture is reasonably accurate, but we don’t know what else is lurking out there in this high dimensional space, in this case about 2300. The stable solutions are stable spirals; the oscillations probably correspond to Rossby waves.

Examples

In this section we present examples of application of the tools of nonlinear filtering and bifurcation theory to increasingly complex systems.

The Double Well

Our Kuroshio model has > 2000 state variables; we’re not ready for that yet. Let’s begin with some simple examples. A simple nontrivial example is the double well.

This is a scalar system with three equilibrium points, one at zero and the others at ±1. The two at ±1 are stable and the one at zero is not.

\[ dx = -4x(x^2 - 1)dt + q^{1/2}dw \]  \hspace{1cm} (4)

This system was investigated by Miller et al. (1994) from the point of view of evaluating the performance of methods such as the extended Kalman filter, which work well in linear systems, to strongly nonlinear systems. Examination of (4) shows that the system has three equilibrium points zero and ±1. The ±1 states are stable, and the zero state is not. The stochastic system tends to remain in the neighborhood of one or the other of the two stable equilibrium points, with transitions occurring when the random process \( q^{1/2}dw \) produces a long series of values which tend to drive the system from one equilibrium to the other. In the absence of specific information about which well the system is in, our ability to guess which well system is in diminishes with time. Eventually, in the absence of further information, the system is equally likely to be in either well, and unlikely to be near the origin. We therefore expect the pdf to be symmetric and bimodal.

The corresponding Fokker-Planck equation is a simple scalar advection diffusion equation:

\[ \frac{\partial P}{\partial t} = \frac{\partial}{\partial x} \left[ 4x(x^2 - 1)P \right] + \frac{1}{2} q \frac{\partial^2 P}{\partial x^2} \]  \hspace{1cm} (5)

The steady solution of (5) is shown in Figure 4. Figure 5 shows the solution of the Fokker-Planck equation for the double-well with Gaussian initial condition. Note convergence to the bimodal steady state.
The evolution of the pdf from a Gaussian pdf to the bimodal equilibrium distribution 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 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.

3.2. A Stochastic Lorenz Model

We next examine a slightly more complex example, Lorenz’ (1963) truncation of the Boussinesq equations:
\[
\begin{align*}
    dX &= \sigma(Y - X)dt + g_1 dw_1 \\
    dY &= (\rho X - Y - XZ)dt + g_2 dw_2 \\
    dZ &= (XY - \beta Z)dt + g_3 dw_3
\end{align*}
\]  

The pdf is now a scalar-valued function of 3 variables. We use the original parameters from Lorenz’ original paper: \( \sigma \) is the Prandtl number, \( \rho \) is a normalized Rayleigh number and \( \beta \) is a nondimensional wavenumber. We choose the values first used by Lorenz to obtain chaotic solutions: \( \sigma = 10.0, \rho = 28.0 \) and \( \beta = 8/3 \). The stochastic forcing is given as in (1):
\[
dw = (dw_1, dw_2, dw_3)^T
\]
which 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 \( f(X,Y,Z) = (\sigma(Y - X),\rho X - Y - XZ,XY - \beta Z)^T \) into (2), along with \( Q = \text{diag}(0.5,0.5,0.5) \). We constructed a reference solution by solving (6) - (8) by Milstein’s (1978) method (see also Kloeden and Platen, 1992) with initial conditions given by
\[
(X,Y,Z) = (-5.91652,-5.52332,24.5723),
\]
(9)

Figure 4. Comparison of computations of the steady solution of the Fokker-Planck equation for the double-well with stochastic forcing. The forcing has variance \( q = .24 \). Solutions by Chang and Cooper’s (1970) finite difference method (dashed curves) and by two Monte Carlo simulations (solid curves) are shown along with the analytical solution (dotted curves). All numerical solutions are for an initial value problem with Gaussian initial conditions with mean zero and variance 0.1 after a time interval of 1.75. Numbers in parentheses denote the number of trials in the Monte Carlo computations.
and noise from the random number generator R250 (Kirkpatrick and Stoll, 1981) for an interval of 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.

The result of our Monte Carlo experiment can be examined in a few ways. For an ensemble calculation with no data assimilation, we may estimate the pdf by the kernel method, and examine the result. Figure 6 shows the values of the equilibrium pdf on the plane $z = 33$. For this choice of parameters, there are three critical points, one at the origin and the other two on the plane $z = 27$. Note that the butterfly structure is clearly visible in this picture.

Results of data assimilation by application of (3) to the pdf calculated from the ensemble are shown in Figure 7. This figure shows slices through $Y = 0$. The top panel is the conditional pdf for the system observed near the equilibrium point shown. The middle panel is the equilibrium pdf with no observations, and the bottom panel is like the top, but for the opposite butterfly wing. A contour-surface representation of the pdf of the stochastic Lorenz system is shown in Figure 8. Here, the butterfly wings are not so obvious. If you squint, you can sort of see the tips of the wings at the top of the figure. We can also show an example of a conditional pdf given an observation. This is clearly more localized, and it shows the state point at a fairly improbable location in state space.
Figure 6. A slice through the equilibrium pdf at $Z = 33$. Filled circles denote the position of the critical points in the $X - Y$ plane.

Figure 7. Result of data assimilation. X-Z plots with $Y = 0$. a: posterior pdf for system observed near critical point with X and Y negative b: equilibrium pdf, no obs. c: similar to a;, but for other critical point.
3.3. A Truncated Spectral Barotropic model

Our last example is a quasigeostrophic $\beta$-plane channel model. The model describes the evolution of the deviation from uniform flow with speed $u^*$. This model was presented by Gravel and Derome (1993) as a schematic model of mid-latitude atmospheric circulation. It is adapted from a very similar model proposed by Charney and DeVore (1979) for the study of multiple stable regimes in the atmosphere. Since this is a quasigeostrophic model, the flow is given in terms of a streamfunction, and the evolution of the streamfunction is determined by the evolution of the potential vorticity field.

The total streamfunction is given by

$$\Psi = -u^* y + \phi(x, y, t)$$

Dissipation is by Rayleigh friction with time constant $\tau$. $H$ and $h$ are the mean depth of the fluid and the variable topography respectively. $\phi$ evolves according to the equation

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

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

We perform our experiments on a spectral truncation of this model, with 5 along-channel and 4 cross-channel wavenumbers. This results in a state space with 44 dimensions.

The bifurcation diagram for equilibria of the 44 dimensional system follow appears as Figure 9. From this picture, we can see that, while there are multiple equilibria, there is only one stable equilibrium point for any choice of the undisturbed velocity $u^*$ in this parameter range.

There are, however, stable limit cycles in 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 Figure 10.

When limit cycles go through a Hopf bifurcation, the result is either period multiplication or bifurcation to an invariant torus, depending on whether the argument of the Floquet multiplier is rational; see, e.g., Ruelle (1989). Analysis of a system whose qualitative behavior depends on whether some computed number is rational or not is clearly problematical. An example of an invariant torus, projected into the space spanned by the three lowest wavenumber components is shown explicitly in Figure 11. Then again, the trajectory illustrated may be a result of a period multiplication by some fairly large factor.

Now what? Clearly a regular grid on 44 dimensional space is not practical.
Figure 9. Bifurcation diagram of equilibria for the truncated spectral model.

Figure 10. Bifurcation diagram of limit cycles for the truncated spectral model.
3.4. Considerations for very high dimensional problems

We can clearly generate ensembles in high dimensions, but making sense of the pdf is very difficult. We expect to do best when the ensemble remains near a low-dimensional subset of the state space, as we expect it to in most problems of interest, but many other problems arise as simple consequences of the high dimension. Problems that are not of serious concern in two or three dimensions become extremely serious when the dimension of the state space increases by an order of magnitude. We expect the kernel method to give reliable results, but the kernels themselves must be evaluated on a random grid, which, in turn, must be chosen with some attention to the dynamical system, or the representation will be extremely inefficient. In addition, kernel widths must be nearly equal; otherwise normalization will be a serious problem. Note that $0.8^{44} \approx 10^{-4}$ and $1.3^{44} \approx 10^5$. How can we interpret a function of 44 variables? It’s difficult to make sense of our 44 dimensional results, so we shall back up: What does a Gaussian pdf look like in 44 dimensions? We expect the level surfaces of a Gaussian pdf to be ellipsoids. Figure 12 shows the projection of a simple 44 dimensional Gaussian pdf into 3 dimensions. For comparison, Figure 13 shows an approximation of that same Gaussian pdf, this time generated by a kernel method on a random grid of the size we plan to use for our solutions to the Fokker-Planck equation for (10).

Each coordinate of the random grid was chosen from a uniform distribution set according to the maxima and minima of the ensemble in each dimension. In Bayes runs or runs in calculations of solutions to the Fokker-Planck equation in which the evolution is to be displayed, the random grid is re-generated in order to adapt to the ensemble. In this case we have 7500 lattice points and 10,000 ensemble points. Comparison of Figures 12 and 13 shows that approximation of a Gaussian pdf by the kernel method on a random grid gives rise to a reasonable approximation, at least for visual inspection purposes.

4. Summary

We have attempted to approach the problem of data assimilation in highly nonlinear stochastic systems by applying the tools of nonlinear filtering theory. Within this framework, we regard the pdf as the solution to the problem. Direct estimates of the state of the system are derived as maximum likelihood. Viewed in this way, data assimilation becomes an exercise in Bayesian statistics: we calculate a prior pdf, and then data assimilation proceeds by explicit calculation of a conditional pdf, conditioned upon observations.

Practical models of the ocean and atmosphere have state spaces reckoned in the tens or hundreds of thousands, which makes explicit calculation of pdf’s on a grid in state space impossible. However, most models of the ocean or atmosphere have only a small number
Figure 12. Projection into 3 dimensions of a 44 dimensional Gaussian pdf on a random grid

Figure 13. Projection into 3 dimensions of a pdf generated from a 44 dimensional Gaussian ensemble
of independent degrees of freedom, and the qualitative behavior of these models may be determined by a low-dimensional subsystem. We apply the tools of the modern theory of dynamical systems in order to examine these low-dimensional systems.

Actually finding these low-dimensional subspaces explicitly is a problem of considerable difficulty, one on which we have made little progress, even in our relatively simple truncated spectral model. Still, our bifurcation analyses allow us to characterize these low-dimensional systems with some precision, and this gives us hope.

Finally, just working at all with systems of this size generates problems of its own. In these cases, we find that once we have the solution, we don’t know how to interpret it or what to do with it. The world has very little experience with systems with tens of dimensions, and in this as in the rest of this project, we are picking our way. Visualization is certainly one of the most important and difficult problems.

Major weather centers now generate ensembles operationally to validate forecasts; see, e.g., Molteni et al. (1996). We expect our work to lead us to the best way to generate these ensembles, and the ways to make best use of them.

Acknowledgments. This work was supported by Office of Naval Research contract N00014-90-J-1125. Sally Blue, Everett Carter and Santhosh Kumaran all contributed to this work.

References