AN OVERVIEW OF ENSEMBLE FORECASTING AND DATA ASSIMILATION

Thomas M. Hamill

NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado

1. INTRODUCTION

Many of the talks and posters during this year's conference will discuss how both ensemble forecasting and atmospheric data assimilation can work synergistically together. We detail provide a brief description of the underlying theoretical basis for this research. The unifying idea is that the chaotic nature of the atmosphere can actually be put to use to improve data assimilation. Ensemble forecasts provide flow-dependent estimates of likelihood of a model-forecast state; modern data assimilation theory requires just this sort of estimate in order to determine how to effectively assimilate new observations. Thus, ensemble forecasting and data assimilation can be coupled into a unified theory. It is possible that data assimilation systems around the world 10 years hence will be using ensemble-based methodologies. It is time for this research to emerge from a being fringe discipline to being the central focus for how to improve data assimilation and numerical weather forecasts.

2. UNDERLYING THEORY

The literature of numerical weather prediction over the last 40+ years has generally focused on how to produce a single "best" initial condition, and from that, make the best forecast possible. Data assimilation, in this context, is usually expressed in some equation where a first guess, or background, is adjusted toward new observation to produce an "analysis." However, considering chaotic effects (e.g., Lorenz 1993), a more reasonable goal is to predict the evolution of probability density π, that is, to know the likelihood of any possible state of the weather we are interested in. Notationally, we will use the convention that a capitalized $\mathbf{X}$ and $\mathbf{Y}$ represent random vectors for the model state and observations, respectively. Lowercase $\mathbf{x}$ and $\mathbf{y}$ will represent actual samples of those random vectors.

Say we seek to estimate a discretized representation of the true state of the atmosphere $\mathbf{X}_t^t$ at some time t, perhaps representing the state of the weather (winds, temperature, humidity, etc.) at a finite number of grid points. We have a prior estimate of the probability distribution $\mathbf{X}_t^t$, but now we have newly collected observations $\mathbf{y}^o$, We'd like an updated, sharpened estimate of the probability distribution. In the terminology of Bayesian statistics, the we would

like to find $\pi(\mathbf{X}_t^t = \mathbf{x}_t \mid \mathbf{Y} = \mathbf{y}^o)$, which we shall shorten notationally to $\pi(\mathbf{X}^t|\mathbf{Y} = \mathbf{y}^o)$. That is, we seek the "posterior" distribution of π conditional on (after assimilating) the new data $\mathbf{y}^o$. Bayes' Rule (any statistics text) tells us how to calculate this:

$$\pi(\mathbf{X}_t^t|\mathbf{Y} = \mathbf{y}^o) = \frac{\pi(\mathbf{X}_t^t)\,\pi(\mathbf{Y} = \mathbf{y}^o|\mathbf{X}_t^t)}{\int \pi(\xi)\,\pi(\mathbf{Y} = \mathbf{y}^o|\xi)d\xi} \qquad (1)$$

In Bayesian terminology, $\pi(\mathbf{X}_t^t)$ is called the "prior distribution." An application of Bayes' Rule is illustrated in Fig. 1 for a 2-dimensional system where the observation and the background state are assumed to measure the same quantity. The posterior is basically a normalized multiplication of the prior distribution and the distribution for the observations conditional on the background $\pi(\mathbf{Y} = \mathbf{y}^o|\mathbf{X}_t^t)$.

From this initial, relatively sharp probability distribution, we would like to forecast the evolution of probability, allowing us to estimate the likelihood of future states. In probabilistic terms, our goal is to estimate the probability density $\pi(\mathbf{X}_t^t)$ as t increases. If, perchance, we have a perfect model of the evolution of the atmosphere, i.e., $\frac{d\mathbf{x}_t}{dt} = f(\mathbf{x}_t)$, then the evolution of this probability density π can in principle be modeled with the Liouville equation, a continuity equation for probability density (Ehrendorfer 1994):

$$\frac{\partial\,\pi(\mathbf{X}_t^t)}{\partial t} + \nabla \cdot \left(\mathbf{X}_t^t\,\pi(\mathbf{X}_t^t)\right) = 0 \qquad (2)$$
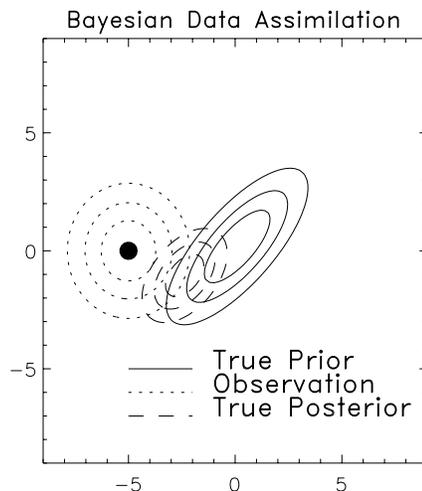

Figure 1. Example of the application of Bayes' Rule in 2 dimensions. Observation probability distribution denoted by dotted line (actual observation is heavy black dot), background (prior) by the solid line, and analysis (posterior) by dashed line. 25, 50, and 75 % of the probability lie outside the respective contours.

Corresponding author address: Dr. Thomas M. Hamill, NOAA-CIRES CDC, R/CDC 1, 325 Broadway, Boulder, CO 80303-3328. e-mail: hamill@cdc.noaa.gov

If our model is imperfect, perhaps our model can expressed stochastically in the form $\frac{d\mathbf{x}_t}{dt} = f(\mathbf{x}_t) + \mathbf{G}(x_t)\mathbf{w}_t$, where $\mathbf{w}_t$ is assumed to be noise from a Weiner process (e.g., Penland 1996). Then, if certain assumptions are made about the stochastic forcing, there is an analogous equation called the Fokker-Planck equation that expresses the evolution of uncertainty (see Gardiner 1990).

In practice, for all but the most trivial systems, neither the Liouville nor the Fokker-Planck equations are tractable, certainly not for million-dimensional state vectors as in current-generation NWP models. However, we can put a handy statistical theorem to use (e.g., Casella and Berger 1990) : Let's shift the time notation so that $t - \delta t$ denotes an initial time, and t the subsequent forecast time. Suppose $\mathbf{x}_{t-\delta t(i)}, i = 1, \ldots, n$ are random samples of the random variable $\mathbf{X}_{t-\delta t}$ and suppose we have some function (e.g., a forecast operator) f such that $\mathbf{x}_{t(i)} = f(\mathbf{x}_{t-\delta t(i)})$. Then $\mathbf{x}_{t(i)}$ is a random sample of $\mathbf{X}_t$. In other words, if we randomly sample the distribution of initial conditions $\pi(\mathbf{X}_{t-\delta t}^t)$ and plug each into the forecast operator, then we get a random sample of forecasts $\pi(\mathbf{X}_t^t)$ (though, buyer beware: we get a random sample reflecting our own NWP forecast operator, not necessarily a random sample of the "real" forecast operator, mother nature).

If we are forced to work with random samples rather than explicitly computing probability distributions, we need to develop a methodology for doing data assimilation that properly utilizes these random sample. It turns out that Bayes' Rule provides the context for how to perform data assimilation, but (1) is not directly useful in highly dimensional systems. The problem in these high-dimensional systems is much simpler if we can make Gaussian assumptions (these assumptions may not be bad if errors are relatively small and still growing linearly). To see how the "Kalman filter" equations (e.g., Lorenc 1986) are derived, let us assume that the prior distribution is Gaussian with mean $\mathbf{x}^b$ and variance-covariance matrix $\mathbf{P}^b$: $\pi(\mathbf{X}_t^t) \sim N(\mathbf{x}^b, \mathbf{P}^b)$. That is,

$$\pi(\mathbf{X}_t^t) \propto exp\left[-\frac{1}{2}\left(\mathbf{X}_t^t - \mathbf{x}^b\right)^T \mathbf{P}^{b-1}\left(\mathbf{X}_t^t - \mathbf{x}^b\right)\right]. \quad (3)$$

Similarly, let us assume that probability distribution for the state of the observations given the background is normally distributed according to

$$\pi(\mathbf{Y} = \mathbf{y}^o | \mathbf{X}_t^t) \propto$$
$$exp\left[-\frac{1}{2}\left(\mathbf{y}^o - \mathbf{HX}_t^t\right)^T \mathbf{R}^{-1}\left(\mathbf{y}^o - \mathbf{HX}_t^t\right)\right]. \quad (4)$$

Here, $\mathbf{H}$ is an operator which converts the model state into the observation type. To find the maximum likelihood estimator for $\pi(\mathbf{X}_t^t | \mathbf{Y} = \mathbf{y}^o)$, we use Bayes' Rule, noting $\pi(\mathbf{X}_t^t | \mathbf{Y} = \mathbf{y}^o) \propto \pi(\mathbf{X}_t^t) \pi(\mathbf{Y} = \mathbf{y}^o | \mathbf{X}_t^t)$. Maximizing this product is equivalent to minimizing the negative of the natural log of the product, i.e., to

minimizing the functional $J(\mathbf{X}_t^t)$

$$J(\mathbf{X}_t^t) = \frac{1}{2}\left[\left(\mathbf{X}_t^t - \mathbf{x}^b\right)^T \mathbf{P}^{b-1}\left(\mathbf{X}_t^t - \mathbf{x}^b\right) + \left(\mathbf{y}^o - \mathbf{HX}_t^t\right)^T \mathbf{R}^{-1}\left(\mathbf{y}^o - \mathbf{HX}_t^t\right)\right] \quad (5)$$

With much algebra, it is possible to show that the posterior distribution has the form $\pi(\mathbf{X}_t^t | \mathbf{Y} = \mathbf{y}^o) \sim N(\mathbf{x}^a, \mathbf{P}^a)$ where $\mathbf{x}^a$ is defined by

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}(\mathbf{y}^o - \mathbf{Hx}^b) \quad (6)$$

and where $\mathbf{K}$ is the Kalman gain matrix,

$$\mathbf{K} = \mathbf{P}^b\mathbf{H}^T(\mathbf{HP}^b\mathbf{H}^T + \mathbf{R})^{-1}. \quad (7)$$

The posterior variance/covariance matrix is predicted by the equation

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{KH})\mathbf{P}^b. \quad (8)$$

It can also be shown that if one has a faulty or simplified estimate of the background error covariance $\hat{\mathbf{P}}^b$ (perhaps an estimate that does not vary with the weather of the day), the expected analysis error covariance is

$$\mathbf{P}^a = \mathbf{P}^b - \hat{\mathbf{K}}\mathbf{HP}^b - \left(\hat{\mathbf{K}}\mathbf{HP}^b\right)^T + \hat{\mathbf{K}}\left(\mathbf{HP}^b\mathbf{H}^T + \mathbf{R}\right)\hat{\mathbf{K}}^T \quad (9)$$

where $\hat{\mathbf{K}} = \hat{\mathbf{P}}^b\mathbf{H}^T(\mathbf{H}\hat{\mathbf{P}}^b\mathbf{H}^T + \mathbf{R})^{-1}$ (Hamill and Snyder 2001). Note that predicting the true $\mathbf{P}^a$ via (9) requires knowledge of the true error statistics $\mathbf{P}^b$, which are likely to be highly flow dependent. Thus, to date, since all operational centers use these simplified estimates $\hat{\mathbf{P}}^b$ and don't know the true $\mathbf{P}^b$, they do not attempt to predict $\mathbf{P}^a$, instead focusing on estimating the most likely state using algorithms akin to (6). Conceptually, one could use an extended Kalman filter (e.g. Cohn 1997, Talagrand 1997), predicting the evolution of $\mathbf{P}^b$ using a linear tangent and adjoint of the forecast model, then using (8) to predict $\mathbf{P}^a$. However, this approach is not computationally feasible for large-dimensional models unless simplifying assumptions are made (e.g., Fisher 1998).

But what might be possible if we have a random sample of $\pi(\mathbf{X}_t^t)$ (an ensemble of forecasts) and new observations? First, it's possible that we can develop a highly accurate estimate of background error covariances $\hat{\mathbf{P}}^b \simeq \mathbf{P}^b$ from the random sample. *The more accurate these background error covariances, the more accurate the analysis, for we will be weighting and spreading the influence of observations as efficiently as possible.* Further, if we have a quality estimate of $\mathbf{P}^b$, then it should be possible to predict the covariance $\mathbf{P}^a$ of the posterior, $\pi(\mathbf{Y} = \mathbf{y}^o | \mathbf{X}_t^t)$. We thus have an estimate of not only the most likely state, but also some estimate of the uncertainty in that state. A random sample

of this distribution is just what we're after in ensemble forecasting.

Let's take this a step further: can we develop a theory so that if we are given a random sample from $\pi(\mathbf{X}_t^t)$, in performing the data assimilation we automatically generate a random sample from the posterior $\pi(\mathbf{Y} = \mathbf{y}^o | \mathbf{X}_t^t)$? Much of the research in ensemble data assimilation is based on this premise.

3. ENSEMBLE KALMAN FILTER DATA ASSIMILATION SYSTEM

In the horse race for the first operational ensemble-based data assimilation method, the ensemble Kalman filter, or "EnKF" has the early lead (Evensen 1994, Houtekamer and Mitchell 1998, 2001, Hamill and Whitaker 2001, and references therein). Many of the talks and posters during the conference will discuss variants on the EnKF. We provide a quick review here.

Start with an ensemble of n analyses at some time t_0 with sufficient spread amongst members. Then repeatedly follow a three-step process for each data assimilation cycle: (1) Make n forecasts to the next analysis time. These forecasts will be used as background fields for n parallel analyses. (2) Given the already imperfect observations at this next analysis time (hereafter called the "control" observations), generate $i = 1, \ldots, n$ independent sets of *perturbed* observations $\mathbf{y}_i^o$ by adding random noise to the control observations $\mathbf{y}^o$. The noise is drawn from the same distribution $\mathbf{R}$ as the observation errors, and the noise is constructed to ensure that the mean of the perturbed observations is equal to the control observation. (3) Perform n objective analyses, updating each of the n background forecasts using the associated set of perturbed observations. The analysis equation for the ith member is

$$\mathbf{x}_i^a = \mathbf{x}_i^b + \hat{\mathbf{P}}^b \mathbf{H}^T \left[\mathbf{H} \hat{\mathbf{P}}^b \mathbf{H}^T + \mathbf{R} \right]^{-1} \left(\mathbf{y}_i^o - \mathbf{H} \mathbf{x}_i^b \right). \quad (10)$$

$\mathbf{x}_i^b$ is the m-dimensional model state vector for the ith member background forecast of an n-member ensemble, and $\mathbf{x}_i^a$ is the subsequently analyzed state for the ith member. $\hat{\mathbf{P}}^b$ is now an approximation of the background error covariances generated from the collection of background forecasts. In its most simple form, $\hat{\mathbf{P}}^b$ is approximated by

$$\hat{\mathbf{P}}^b = \frac{1}{n-1} \sum_{i=1}^{n} \left(\mathbf{x}_i^b - \overline{\mathbf{x}^b} \right) \left(\mathbf{x}_i^b - \overline{\mathbf{x}^b} \right)^T, \quad (11)$$

where $\overline{\mathbf{x}^b} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i^b$ is the ensemble mean. It can be shown (Burgers et al. 1998) that under certain assumptions, the ensemble mean from the EnKF is "optimal" and the posterior covariance calculated from the ensemble of analyses matches the $\mathbf{P}^a$ predicted by (9).

Additional complexity is often introduced to the standard EnKF design to deal with the detrimental process known as filter divergence (e.g., Houtekamer and Mitchell 1998, van Leeuwen 1999). In this process, the ensemble progressively ignores observational data more and more in successive cycles, leading to a useless ensemble. One approach is to modifiy backround error covariances by applying a Schur product with a correlation function, as discussed in Houtekamer and Mitchell (2001) and Hamill et al. (2001). Another approach which can ameliorate the tendency toward filter divergence is to use a "double" EnKF (Houtekamer and Mitchell 1998), whereby ensemble members are kept in two separate batches; the covariance model from one batch is used in the assmilation of the other. This can help prevent the feedback cycle toward smaller and smaller background error covariances. Hamill and Snyder (2000) suggested a hybrid EnKF, whereby covariances are modeled as a combination of covariances from the ensemble and from a stationary model like 3D-Var. Anderson and Anderson (1999) suggested increasing background error covariances somewhat by inflating the deviation of background members with respect to their mean by a small amount.

Though covariances are generally modeled in the EnKF as in (11), direct application of (11) is computationally prohibitive. Thus, for computational efficiency, the matrix operations $\hat{\mathbf{P}}^b \mathbf{H}^T$ and $\mathbf{H} \hat{\mathbf{P}}^b \mathbf{H}^T$ in (9) are computed together using data from the ensemble of background states. Define

$$\overline{\mathbf{H} \mathbf{x}}^b = \frac{1}{n} \sum_{i=1}^{n} \mathbf{H} \mathbf{x}_i^b,$$

which represents the mean of the estimate of the observation generated from the background forecasts. Then

$$\hat{\mathbf{P}}^b \mathbf{H}^T = \frac{1}{n-1} \sum_{i=1}^{n} \left(\mathbf{x}_i^b - \overline{\mathbf{x}^b} \right) \left(\mathbf{H} \mathbf{x}_i^b - \overline{\mathbf{H} \mathbf{x}}^b \right)^T, \quad (12)$$

and

$$\mathbf{H} \hat{\mathbf{P}}^b \mathbf{H}^T = \frac{1}{n-1} \sum_{i=1}^{n} \left(\mathbf{H} \mathbf{x}_i^b - \overline{\mathbf{H} \mathbf{x}}^B \right) \left(\mathbf{H} \mathbf{x}_i^b - \overline{\mathbf{H} \mathbf{x}}^B \right)^T. \quad (13)$$

4. ENSEMBLE DATA ASSIMILATION: SIMPLE EXAMPLES

We now discuss some simple example to illustrate how the EnKF works. We take the same data as was used to generate Fig. 1, a 2-D system. However, in actuality we do not know the full prior background-error covariance as assumed to generate Fig. 1. Let us assume that we have an ensemble, a random sample from this distribution (Fig. 2(a), small dots), and a control

observation $\mathbf{y}^o$, denoted by the heavy black dot. We know the observational error covariance $\mathbf{R}$ and generate perturbed observations (diamonds) consistent with these error statistics. A background error covariance, or "sample prior" is estimated from our ensemble; note that since this is estimated from a random sample, it is not the same distribution as the true prior (one dashed contour illustrated for comparison in Fig. 2(a)). Now, parallel assimilation cycles are conducted using eq. (10) updating one background to one perturbed observation. The resulting analyses (dots in Fig 2(b)) are generally consistent with the true posterior distribution (dashed lines in Fig. 2(b), reproduced from Fig. 1). As illustrated in Fig. 2, there is a sampling error in estimating covariances from a random sample of background states. As long as the true distribution is Gaussian, this estimate will improve as more members are added. We note, also, that there is a sampling error associated with the use of perturbed observations. As discussed in Burgers et al. (1998), the perturbed observations were originally added to ensure that the expected value of $\mathbf{P}^a$ from the EnKF matched that predicted by eq. (8). However, especially for small ensemble sizes, the perturbed observations can be spuriously correlated with the background. Fig. 3 (a)-(b) illustrates how, by reassociating certain perturbed observations with certain background members, very different sets of analyses may result. Other presentations during this conference by Anderson and Whitaker will suggest alternative methods to the EnKF which can overcome this particular sampling problem.

5. CONCLUSIONS

Ensemble-based data assimilation schemes are a very promising new approach which may dramatically improve data assimilation. By providing flow-dependent information on the errors in the background state, the influence of new observations can be used much more effectively. Data assimilation and ensemble forecasting, now considered distinctly different parts of the forecast process, may be unified, improving the quality of each.

Research into ensembles and data assimilation is growing, but there are still a host of issues that need to be explored and resolved before the ensemble-based data assimilation techniques will be ready for operational use. Among the expected problems that will need to be addressed are:

• How do we minimize the computational expense of ensemble-based data assimilation schemes? Generally, the ensemble schemes are more accurate with more members, but the schemes scale in cost with the size of the ensemble. It may be possible to reduce costs using dynamically constrained perturbations (e.g., Heemink et al. 2001), and clever methods of exploiting parallelism may reduce the computational demands (e.g., Houtekamer and Mitchell 2001).
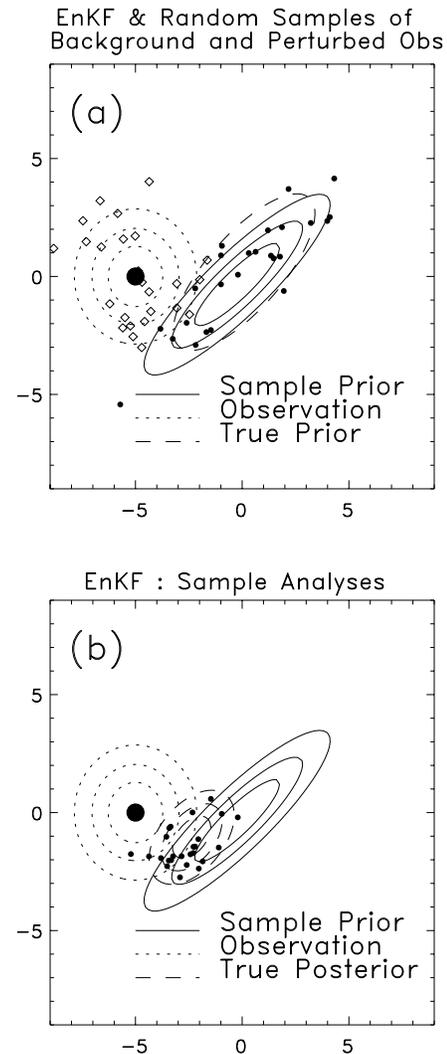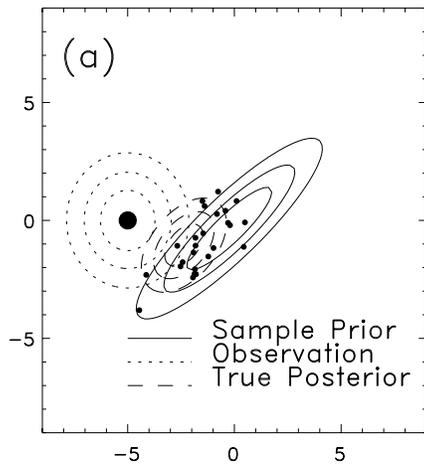

Figure 2. (a) Example of EnKF algorithm. Perturbed observations (diamonds) created by adding noise to control observation (heavy dot). Background states (dots) here assumed to be randomly sampled from true prior distribution in Fig. 1. Sample prior distribution from prior overplotted in solid, and observational error distribution with dots. (b) Posterior (analyzed) states from EnKF (dots), with true posterior distribution from Fig. 1 plotted over top (dashed).

• How do we deal with non-Gaussian statistics? To what extent are the distributions really non-Gaussian? This is an active but relatively new area of research.

• Will there be complications (such as balance issues) when the technique is applied to full primitive-equation models with complex physical parameterizations?

• How do we deal with model errors? Do we have models cast in a stochastic framework (e.g., Buizza et al. 1999, Penland 1996) or add system noise during the data assimilation (e.g., Mitchell and Houtekamer 2000)?

EnKF : Non-Random Perturbed Obs

(a)

—— Sample Prior
.... Observation
— — True Posterior

EnKF : Non-Random Perturbed Obs

(b)

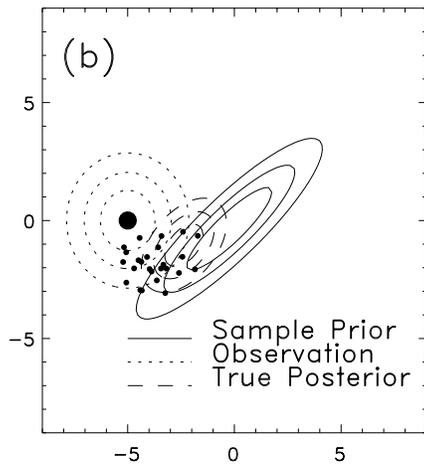—— Sample Prior
.... Observation
— — True Posterior

Figure 3. Two more examples of samples of the EnKF analyses (dots) for different associations of the perturbed observations with the background states.

REFERENCES

Anderson, J. L., and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.*, **127**, 2741-2758.

Buizza, R., M. Miller, and T. N. Palmer, 1999: Stochastic simulation of model uncertainty in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887-2908.

Casella, G., and R. L. Berger, 1990: *Statistical Inference.* Duxbury Press, 650 pp.

Cohn, S. E., 1997: An introduction to estimation theory. *J. Meteor. Soc. Jap.*, **75(1B)**, 257-288.

Ehrendorfer, M., 1994: The Liouville equation and its potential usefulness for the prediction of forecast skill. part I: theory. *Mon. Wea. Rev.*, **122**, 703-713.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99 (C5)**, 10143-10162.

——, and P. J. van Leeuwen, 1996: Assimilation of Geosat altimeter data for the Agulhas current using the ensemble Kalman filter with a quasigeostrophic model. *Mon. Wea. Rev.*, **124**, 85-96.

Fisher, M., 1998: Development of a simplified Kalman filter *ECMWF Research Department Techical Memorandum 260.* European Centre for Medium-Range Weather Forecasts. 16 pp. Available from Library, ECMWF, Shinfield Park, Reading, Berkshire, RG2 9AX, England.

Gardiner, C. W., 1990: *Handbook of stochastic methods.* Springer-Verlag, 442 pp.

Hamill, T. M., and C. Snyder, 2000. A hybrid ensemble Kalman filter - 3D variational analysis scheme. *Mon. Wea. Rev.*, **128**, 2905-2919.

——, and ——, 2001: Using improved background error covariance estimates from an ensemble Kalman filter for adaptive observations. *Mon. Wea. Rev.*, submitted. Available from hamill@cdc.noaa.gov.

——, J. S. Whitaker, and ——, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, accepted.

Heemink, A. W., M. Verlaan, and A. J. Segers, 2000: Variance-reduced ensemble Kalman filtering. *Mon. Wea. Rev.*, in press.

Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796-811.

——, and ——, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123-137.

Lorenc, A. C., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **112**, 1177-1194.

Lorenz, E. N., 1993: *The essence of chaos.* University of Washington Press.

Mitchell, H. L., and P. L. Houtekamer, 2000: An adaptive ensemble Kalman filter. *Mon. Wea. Rev.*, **128**, 416-433.

Penland, C., 1996: A stochastic model of IndoPacific sea surface temperature anomalies. *Physica D* **98**, 534-558.

Talagrand, O., 1997: Assimilation of observations: an introduction. *J. Meteor. Soc. Jap.*, **75 (1B)**, 191-209.

van Leeuwen, P. J., 1999: Comment on "Data assimilation using an ensemble Kalman filter technique." *Mon. Wea. Rev.*, **127**, 1374-1377.