Ensemble-Based Data Assimilation

Thomas M. Hamill

NOAA-CIRES Climate Diagnostics Center,
Boulder, Colorado USA
hamill@cdc.noaa.gov

1 Introduction

The purpose of this chapter is to introduce the reader to promising new experimental methods for atmospheric data assimilation involving the use of ensemble forecasts. There is a natural linkage between data assimilation and ensemble forecasting: ensemble forecasts are designed to estimate the flow-dependent uncertainty of the forecast; data assimilation techniques require accurate estimates of forecast uncertainty in order to optimally blend the prior forecast(s) with new observations. Our intent is to demonstrate more concretely how these two endeavours are united and how and why this union may improve the quality of both initial conditions and ensemble forecasts.

Rather than launching into the specifics of recently proposed ensemble-based assimilation techniques, we will take a step back and try to motivate their use by quickly developing them from first principles, noting the approximations that have been made along the way. This will take us from full Bayesian data assimilation, which is conceptually simple but computationally prohibitive, to the Kalman filter, somewhat of a simplification, to the ensemble Kalman filter, which is perhaps more computationally tractable and more accurate. We will then discuss some of the algorithmic techniques and the lessons learned from preliminary experiments with ensemble-based data assimilation systems.

The reader is presumably familiar with the scientific underpinnings of chaos theory (Lorenz 1993). Most specifically, for ensemble-based data assimilation, we will depend on the property that differences between model states in a chaotic system will grow in a greatly reduced dimensionality subspace relative to the number of degrees of freedom in the model. A smaller dimensional subspace in which the variance among model states resides is important for the success of ensemble forecasting and ensemble data assimilation. If errors were likely to grow equally and independently in every possible direction, for a forecast model with 10^7 degrees of freedom, an ensemble of size 10^2 would be nearly useless for modeling the structure of forecast errors, spanning an irrelevant 10^{-5} of the important directions. The more that forecast errors are contained in a small subspace of a few particular directions, the more information in a limited-size ensemble (Tippett et al. 2000). Fortunately, it appears that differences among ensemble member tend to grow in a relatively limited number of directions in weather forecast models.

Why bother with ensemble-based assimilation techniques? Over the past decade, the operational weather prediction facilities have worked very hard to implement four-dimensional variational analysis techniques (4D-Var; Le Dimet and Talagrand 1986, Courtier et al. 1994, Rabier et al. 1998, 2000). These techniques are now considered the state of the art in atmospheric data assimilation. In 4D-Var, we typically seek to find a model trajectory that best fits the observations over some recent period of time, perhaps the last 6-24 h. The model state at the end of this time window is the control initial condition around which an ensemble of forecast initial conditions may be generated. And 4D-Var has shown great promise. In particular, the European Centre for Medium Range Weather Forecasts (ECMWF) has noted dramatic improvements in the accuracy of their analyses and subsequent forecasts, in part due to the use of 4D-Var. Given this, it is worth asking up front: why consider ensemble-based data assimilation techniques? Can they possibly produce better analyses than 4D-Var? Only comparative tests will settle this issue, and so far, such comparisons are lacking. Still, we do know that

4D-Var still has many significant disadvantages that may be easier to address using ensemble techniques. For example, the forecast model used in 4D-Var may not adequately represent the true dynamics of the atmosphere, i.e., model error may be significantly large. Other limitations may include the difficulty of generating an accurate model of background-error covariances used in 4D-Var and the costs of coding and executing accurate tangent-linear and adjoints models. Consequently, 4D-Var may fit a model trajectory that was significantly different than the trajectory of the real atmosphere during that time window. Such problems were suggested in tests with the operational version of 4D-Var at ECMWF (Rabier et al. 2000); when a 6-h assimilation window was used, 4D-Var outperformed 3D-Var, but not when longer assimilation windows were used. In contrast, model imperfections can be dealt with in a straightforward manner in ensemble techniques, and no adjoint nor tangent-linear codes are required. The algorithmic simplicity of these techniques and the fact that they produce an ensemble of analyses make them especially attractive for ensemble forecasting applications.

In subsequent discussion we also will assume that the atmosphere state, which is of course a continuum, can be adequately described in discretized fashion, perhaps by the values of winds, temperature, humidity, and pressure at a set of grid points.

2 Bayesian Data Assimilation

Conceptually, the atmospheric data assimilation problem is a relatively simple one. We would like the best estimate of the probability density function (pdf) for the current atmospheric state given all current and past observations. Much of the material in this section follows the work of Jazwinski (1970). If the reader is interested in further material on the subject, Lorenc (1986) provides a formulation of data assimilation in a Bayesian context, and Talagrand (1997) provides an excellent review of the data assimilation and in particular, the Kalman filter discussed in Section 3. Cohn (1997) provides a more rigorous statistical basis for the problem. We attempt here to use the minimum math possible to guide the reader along the path from Bayesian data assimilation to the ensemble Kalman filter.

Assume the following notational conventions. Random variables for the model state and observations are denoted by capital letters, while the possible values that they may take on are denoted by small letters. Generally, boldface characters will denote vectors or matrices, while use of the normal font denotes a scalar. Thus, if we use the letter "x" to denote the model state, $\mathbf{X}$ indicates the random vector denoting the true model state, $\mathbf{x} = [\mathrm{x}_{(1)}, \ldots, \mathrm{x}_{(n)}]$ denotes a specific n-component vector value, and $P(\mathbf{X} = \mathbf{x})$ indicates the probability density that the true state of the random vector is the value $\mathbf{x}$. Denote the associated random vector for the observations again with capitalized letters, i.e., Ψ_t. This random vector includes observations $\mathbf{Y}_t$ at the most recent time as well as observations at all previous times $\Psi_{t-1} = [\mathbf{Y}_{t-1}, \ldots, \mathbf{Y}_0]$, which we condense to $\Psi_t = [\mathbf{Y}_t, \Psi_{t-1}]$. Denote a *possible* vector value for the observations as $\widetilde{\psi}_t = [\widetilde{\mathbf{y}}, \widetilde{\psi}_{t-1}]$. Also assume we have a collection of *actual* observations $\psi_t = [\mathbf{y}_t, \psi_{t-1}]$, There are M_t observations at time t, i.e., $\mathbf{y}_t = [y_{t(1)}, \ldots, y_{t(M_t)}]$.

Formally, the problem we seek to solve is the following: $P(\mathbf{X}_t = \mathbf{x}_t \mid \Psi_t = \widetilde{\psi}_t)$. That is, we'd like the best probability density estimate of the current atmospheric state provided the current and past observations take on specified values. Bayes' Rule tells us that this quantity can be re-expressed as

$$P(\mathbf{X}_t = \mathbf{x}_t \mid \Psi_t = \widetilde{\psi}_t) \propto P(\Psi_t = \widetilde{\psi}_t \mid \mathbf{X}_t = \mathbf{x}_t)\, P(\mathbf{X}_t = \mathbf{x}_t) \tag{1}$$

Bayes' Rule is usually expressed with a normalization constant in the denominator on the right-hand side of (1); for simplicity, we have dropped that and assume that when coded, the user ensures that probability density integrates to 1.0.

Let us make one hopefully minor assumption: observational errors are independent from one time to the next. Hence, $P(\Psi_t = \widetilde{\psi}_t) = P(\mathbf{Y}_t = \widetilde{\mathbf{y}})\, P(\Psi_{t-1} = \widetilde{\psi}_{t-1})$. This may not be true for observations from satellites, where instrumentation biases may be difficult to remove. Also, errors of observation representativeness (Daley 1991) may be flow-dependent and correlated in time. But under this assumption, (1) is equivalent to

$$P(\mathbf{X}_t = \mathbf{x}_t \mid \Psi_t = \widetilde{\psi}_t) \propto P(\mathbf{Y}_t = \widetilde{\mathbf{y}} \mid \mathbf{X}_t = \mathbf{x}_t)\, P(\Psi_{t-1} = \widetilde{\psi}_{t-1} \mid \mathbf{X}_t = \mathbf{x}_t)\, P(\mathbf{X}_t = \mathbf{x}_t) \tag{2}$$

By Bayes' Rule again, $P(\Psi_{t-1} = \widetilde{\psi}_{t-1} \mid \mathbf{X}_t = \mathbf{x}_t)\ P(\mathbf{X}_t = \mathbf{x}_t) \propto P(\mathbf{X}_t = \mathbf{x}_t \mid \Psi_{t-1} = \widetilde{\psi}_{t-1})$. Hence, (2) simplifies to

$$P(\mathbf{X}_t = \mathbf{x}_t \mid \Psi_t = \widetilde{\psi}_t) \propto P(\mathbf{Y}_t = \widetilde{\mathbf{y}} \mid \mathbf{X}_t = \mathbf{x}_t)\ P(\mathbf{X}_t = \mathbf{x}_t \mid \Psi_{t-1} = \widetilde{\psi}_{t-1}) \tag{3}$$

Equation (3) is elegantly simple. This is a recursive relationship: the "posterior," the pdf for the current model state given all the observations is a product of the the probability distribution for the current observations $P(\mathbf{Y}_t = \widetilde{\mathbf{y}} \mid \mathbf{X}_t = \mathbf{x}_t)$ and the "prior," $P(\mathbf{X}_t = \mathbf{x}_t \mid \Psi_{t-1} = \widetilde{\psi}_{t-1})$ also known as the "background." Typically, it is assumed in (3) that $P(\mathbf{Y}_t = \widetilde{\mathbf{y}} \mid \mathbf{X}_t = \mathbf{x}_t)$ is a probability distribution centered in some manner about the actual observed values; for example, perhaps $P(\mathbf{Y}_t = \widetilde{\mathbf{y}} \mid \mathbf{X}_t = \mathbf{x}_t) \sim N(\mathbf{y}_t, \mathbf{R})$, that is, the pdf is normally distributed about $\mathbf{y}_t$ with some observation-error covariance matrix $\mathbf{R}$. The prior is the pdf of the model state at time t given all the past observations up to time $t-1$. Typically, the prior will have been estimated from a cycle of previous data assimilations and short-term forecasts up to the current time.

Let's now demonstrate Bayesian assimilation with a simple example. Suppose we have an estimate of the prior $P(\mathbf{X}_t = \mathbf{x}_t \mid \Psi_{t-1} = \widetilde{\psi}_{t-1})$ for a two-dimensional model state. This was produced by assimilating all prior observations up to and including time $t-1$ and estimating in some manner how that pdf has evolved in the time interval between $t-1$ and t. We now want to update the pdf given a new scalar observation y, which in this example is observing the same quantity as the first component of the state vector measures. The pdf for the observation $P(Y_t = \widetilde{y}_t \mid \mathbf{X}_t = \mathbf{x}_t)$ is assumed to be distributed normally about the actual observation, $Y_t \sim N(y_t, \sigma^2)$. Here, let $y_t = 58$ and $\sigma^2 = 100$.

Selected contours of the prior are plotted in Fig. 1(a); as shown, the prior is bimodal. The shape of the marginal distributions $P(X_{t_{(1)}} = x_{t_{(1)}} \mid \Psi_{t-1} = \widetilde{\psi}_{t-1})$ and $P(X_{t_{(2)}} = x_{t_{(2)}} \mid \Psi_{t-1} = \widetilde{\psi}_{t-1})$ are plotted on each axis in solid lines. The dashed line denotes the observation probability distribution $P(Y_t = \widetilde{y}_t \mid \mathbf{X}_t = \mathbf{x}_t)$. This probability varies with the value $x_{t_{(1)}}$, but given $x_{t_{(1)}}$ is the same for any value of $x_{t_{(2)}}$. The updated posterior distribution is computed using (3) and is shown in Fig. 1(b). Note that the assimilation of the observation enhanced the probability in the lobe overlapping the observation distribution and decreased it in the other lobe. Overall, the posterior distribution is more sharp (specific) than the prior, as is expected.

Unfortunately, Bayesian data assimilation without some simplification is not practical for real-world numerical weather prediction applications. One problem with modeling a complicated pdf in higher dimensions is the "curse of dimensionality" (e.g., Bellman 1961, Hastie et al. 2001). Were one to try estimate the probability density in a higher-dimensional space using a small ensemble, one would find that the model of probability was very poor unless simplifying assumptions about the form of the distribution were made. Even were this problem surmountable, the computational cost would be extravagant. In the prior example we evaluated the probability density on a 100×100 grid. Suppose a similarly complicated structure for the prior existed in 100 dimensions. Then if we were to keep track of joint probabilities on a similar grid for each dimension, this would involve evaluating and modifying 100^{100} density estimates. Such computations are already prohibitive for a 100-dimensional model state; the problem becomes incomprehensible for model states of $O(10^7)$. Clearly, some simplification is required.

3 The Kalman Filter

3.1 The Discrete Kalman Filter

Non-normality of the prior such as the bimodality in Fig. 1(a) is typically assumed to be uncommon in atmospheric data assimilation. This assumption is probably justified; at least for temperature and winds, the error in the prior is approximately normally distributed, is usually relatively small compared to the error of a random model state, and grows linearly over a short period of time. These assumptions may be bad ones for moisture, cloud cover, and other aspects of the model state that may be very sensitive to motions at small scales, where the time scale of predictability is small and errors grow and saturate rapidly. But let us assume normality and linearity of error growth. Then Bayesian data assimilation computations can be simplified. Accordingly, we

make the following assumptions:

$$P(\mathbf{X}_t = \mathbf{x}_t \mid \Psi_{t-1} = \widetilde{\psi}_{t-1}) \sim N(\mathbf{x}_t^b, \mathbf{P}_t^b) \propto exp\left[-\frac{1}{2}\,(\mathbf{x}_t - \mathbf{x}_t^b)^T \mathbf{P}_t^{b\,-1}(\mathbf{x}_t - \mathbf{x}_t^b)\right] \tag{4}$$

That is, the probability density of the prior is normally distributed with known mean background $\mathbf{x}_t^b$ and background-error covariance matrix $\mathbf{P}_t^b$. Similarly, assume that the observations are distributed normally with mean $\mathbf{y}_t$ and covariance $\mathbf{R}$:

$$P(\mathbf{Y}_t = \widetilde{\mathbf{y}} \mid \mathbf{X}_t = \mathbf{x}_t) \sim N(\mathbf{y}_t, \mathbf{R}) \propto exp\left[-\frac{1}{2}\,(\mathbf{H}\mathbf{x}_t - \mathbf{y}_t)^T\,\mathbf{R}^{-1}(\mathbf{H}\mathbf{x}_t - \mathbf{y}_t)\right] \tag{5}$$

Here, $\mathbf{H}$ is a linear operator that converts the model state to the observation type and location. Applying (3),

$$P(\mathbf{X}_t = \mathbf{x}_t \mid \Psi_t = \widetilde{\psi}_t) \propto exp\left[-\frac{1}{2}\,(\mathbf{x}_t - \mathbf{x}_t^b)^T\,\mathbf{P}_t^{b\,-1}(\mathbf{x}_t - \mathbf{x}_t^b) - \frac{1}{2}\,(\mathbf{H}\mathbf{x}_t - \mathbf{y}_t)^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{x}_t - \mathbf{y}_t)\right] \tag{6}$$

Maximizing (6) is equivalent to minimizing the negative natural log of (6), i.e., to minimizing the functional $J(\mathbf{x}_t)$ according to

$$J(\mathbf{x}_t) = \frac{1}{2}\left[(\mathbf{x}_t - \mathbf{x}_t^b)^T\,\mathbf{P}_t^{b\,-1}(\mathbf{x}_t - \mathbf{x}_t^b) + (\mathbf{H}\mathbf{x}_t - \mathbf{y}_t)^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{x}_t - \mathbf{y}_t)\right] \tag{7}$$

This functional is a common starting point in the derivation of many assimilation schemes, from the Kalman filter to 3-dimensional variational assimilation ("3D-Var;" e.g., Lorenc 1986, Parrish and Derber 1992). We would like to choose the value that minimizes this functional, providing the best fit between the new observations and the prior. Let's call this best fit state the "analysis," or $\mathbf{x}_t^a$. $\mathbf{x}_t^a$ can be found by the differentiating the functional in (7) with respect to $\mathbf{x}_t^a$, setting the result equal to zero, and proceeding with some manipulation. A full derivation is provided in Appendix 1. The resulting "update" equations are

$$\mathbf{x}_t^a = \mathbf{x}_t^b + \mathbf{K}(\mathbf{y}_t - \mathbf{H}\mathbf{x}_t^b) \tag{8}$$

where

$$\mathbf{K} = \mathbf{P}_t^b\mathbf{H}^T(\mathbf{H}\mathbf{P}_t^b\mathbf{H}^T + \mathbf{R})^{-1} \tag{9}$$

The optimal analysis state $\mathbf{x}_t^a$ is estimated by correcting the background $\mathbf{x}_t^b$ toward the "observation increment" $\mathbf{y}_t - \mathbf{H}\mathbf{x}_t^b$, weighted by the Kalman gain matrix $\mathbf{K}$. The effect of the $\mathbf{K}$ is to use the observation increments to correct the background at relevant surrounding grid points. On average, the corrections are typically larger for grid points near to the observation location than for grid points far from the observation location. However, the appeal of the Kalman filter relative to an analysis scheme like 3D-Var is that the structure of the analysis increments $\mathbf{x}_t^b - \mathbf{x}_t^a$ can be quite complicated. This structure depends on the background-error covariances, which are affected by the corrective action of previous observations on previous forecasts as well as the chaotic dynamics since the last update. We will provide some examples in section 8 of complicated background-error structures in the context of an ensemble Kalman filter.

The basic statistics of the Kalman filter update can be understood if one considers a one-dimensional state vector updated to a single observation. Assume H=1. Under these assumptions, (8) simplifies to

$$\begin{aligned} x_t^a &= x_t^b - \frac{P_t^b}{P_t^b + R}(y_t - x_t^b) \\ &= \frac{R}{P^b + R}x_b^t + \frac{P^b}{P^b + R}y_t \end{aligned} \tag{10}$$

The analysis state thus is a weighted linear combination of the background and the observation (Daley 1991). The smaller the observation error is relative to the background, the more the analysis is drawn toward the observation. Analogies for multiple dimensions will be discussed in later sections.

Equations (8) and (9) indicate how to predict the most likely state, but the accuracy of these calculations depends on the accuracy of $\mathbf{P}_t^b$ and $\mathbf{R}$. For atmospheric data assimilation, the latter, $\mathbf{R}$, is usually derived from extensive calibration and validation field experiments. Error statistics for $\mathbf{R}$ are usually assumed to be independent of the flow; the same error statistics are used for different observations of the same type, regardless of location, time, and synoptic conditions. Background-error statistics are explicitly updated in the Kalman filter. Given $\mathbf{P}_t^b$, $\mathbf{R}$, and the observation locations (implied in $\mathbf{H}$), the analysis-error covariance $\mathbf{P}_t^a$ is predicted. A forecast is then made of how analysis-error covariances will evolve in time until the next assimilation cycle. The equation for the expected $\mathbf{P}_t^a$ is

$$\begin{aligned}\mathbf{P}_t^a &= (\mathbf{I}-\mathbf{KH})\mathbf{P}_t^b \\ &= \mathbf{P}_t^b - \mathbf{KHP}_t^b \\ &= \mathbf{P}_t^b - \mathbf{P}_t^b\mathbf{H}^T(\mathbf{HP}_t^b\mathbf{H}^T+\mathbf{R})^{-1}\mathbf{HP}_t^b\end{aligned} \tag{11}$$

The derivation of this is also provided in Appendix 1.

The form of (11) indicates that the analysis-error covariance is a correction to the background error covariance. The magnitude of covariances are reduced, with the amount that they are reduced reflected in $\mathbf{KHP}_t^b$. If we consider the one-dimensional system again as in (10), we find that

$$\begin{aligned}P_t^a &= P_t^b - P_t^b\frac{P_t^b}{P_t^b+R} \\ &= P_t^b\left(\frac{R}{P_t^b+R}\right)\end{aligned} \tag{12}$$

Hence, the smaller the magnitude of R (the more accurate the observation) relative to P_t^b, the greater the fractional reduction of error covariance.

Given $\mathbf{x}_t^a$ and $\mathbf{P}_t^a$, we will need to evolve the expected analysis and the covariances forward in some fashion to produce an estimate of background-error covariances at time $t+1$. In the discrete Kalman filter, we suppose that the true model state evolves according to the equation

$$\mathbf{x}_{t+1} = \mathbf{M}\mathbf{x}_t + \eta \tag{13}$$

That is, the forecast evolution can be expressed as the sum of a linear operation on the current model state $\mathbf{M}\mathbf{x}_t$ plus an unknown error η, also known as the "system noise." $\mathbf{M}$ is an $m \times m$ matrix, often called the *transition matrix* between times t and $t+1$. Since we describe the discrete Kalman filter only to motivate use of the ensemble Kalman filter, algorithmic details like the method of calculation of $\mathbf{M}$ will be skipped; these details are not crucial to the understanding of ensemble filters. η is assumed to have an expected value of zero $\left\langle \eta \right\rangle = 0$ and to be uncorrelated in time, with expected covariance $\mathbf{Q}$: $\left\langle \eta_p \eta_q^T \right\rangle = \mathbf{Q}\delta_{pq}$ where δ is the Kronecker delta, and $_p$ and $_q$ denote two assimilation times. In practice, accurately determining the statistics of $\mathbf{Q}$ may be somewhat involved (Dee 1995).

Given our model in (13), if we are looking for the best estimate of the evolution of the mean state (e.g., Talagrand 1997), this is simply (13) without the noise term:

$$\mathbf{x}_{t+1}^b = \mathbf{M}\mathbf{x}_t^a \tag{14}$$

We also require an estimate of the background-error covariances at the next assimilation time. Assuming that the system noise is uncorrelated with the tangent-linear dynamics, $\left\langle \mathbf{M}(\mathbf{x}_t^a - \mathbf{x}_t)\ \eta^T \right\rangle = 0$, we get

$$\begin{aligned}
\mathbf{P}_{t+1}^b &= \left\langle (\mathbf{x}_{t+1}^b - \mathbf{x}_{t+1})\ (\mathbf{x}_{t+1}^b - \mathbf{x}_{t+1})^T \right\rangle \\
&= \left\langle (\mathbf{M}\mathbf{x}_t^a - \mathbf{M}\mathbf{x}_t - \eta)\ (\mathbf{M}\mathbf{x}_t^a - \mathbf{M}\mathbf{x}_t - \eta)^T \right\rangle \\
&= \left\langle (\mathbf{M}(\mathbf{x}_t^a - \mathbf{x}_t) - \eta)\ (\mathbf{M}(\mathbf{x}_t^a - \mathbf{x}_t) - \eta)^T \right\rangle \\
&= \mathbf{M}\mathbf{P}_t^a\mathbf{M}^T + \mathbf{Q} \\
&= \mathbf{M}\ (\mathbf{M}\mathbf{P}_t^a)^T + \mathbf{Q}
\end{aligned} \tag{15}$$

Given an operator $\mathbf{M}$, this is how analysis-error covariances are evolved in the discrete Kalman filter.

3.2 The extended Kalman filter

Before considering ensemble data assimilation methods, we touch briefly on an extension to the discrete Kalman filter called the *extended Kalman filter*, where some of the assumptions of linearity are relaxed. First, suppose the assumption of linearity in (13) is a poor one; perhaps were one to use a fully nonlinear model operator $\mathscr{M}$, i.e.,

$$\mathbf{x}_{t+1} = \mathscr{M}\mathbf{x}_t + \eta \tag{16}$$

then η would be much smaller. Accordingly, in the extended Kalman filter, one assumes that the mean state will be evolved according to

$$\mathbf{x}_{t+1}^b = \mathscr{M}\mathbf{x}_t^a \tag{17}$$

intead of using (14). If differences $\mathbf{x}_t^a - \mathbf{x}_t$ are small, then the evolution of these difference should be approximately linear:

$$\mathscr{M}\mathbf{x}_t^a - \mathscr{M}\mathbf{x}_t \simeq \mathbf{M}(\mathbf{x}_t^a - \mathbf{x}_t) \tag{18}$$

where here $\mathbf{M}$ is the Jacobian matrix of $\mathscr{M}$: $\mathbf{M} = \frac{\partial \mathscr{M}}{\partial \mathbf{x}}$.

Let's also relax the restriction that the forward operator $\mathbf{H}$ in (5) be linear. We replace it with a (perhaps nonlinear) $\mathscr{H}$. Again, we presume that differences like $\mathbf{x}_t^b - \mathbf{x}_t$ are small enough so that the innovation vector $\mathbf{y}_t - \mathscr{H}\mathbf{x}_t^b = \mathscr{H}\mathbf{x}_t - \mathscr{H}\mathbf{x}_t^b + \varepsilon$ can be approximated with $\mathbf{H}(\mathbf{x}_t - \mathbf{x}_t^b) + \varepsilon$, where now $\mathbf{H} = \frac{\partial \mathscr{H}}{\partial \mathbf{x}}$.

Given these relaxed assumptions, one can proceed to derive an alternate form of the Kalman filter update equations. The derivation of these is beyond the scope of this note but is discussed in Jazwinski (1970). We will be content here to simply note the changes. In addition to assuming that the mean state is evolved nonlinearly in (17), the update equations is changed. (8) and (9) are replaced by

$$\mathbf{x}_t^a = \mathbf{x}_t^b + \mathbf{K}(\mathbf{y}_t - \mathscr{H}\mathbf{x}_t^b) \tag{19}$$

where

$$\mathbf{K} = \mathbf{P}_t^b\mathbf{H}^T(\mathbf{H}\mathbf{P}_t^b\mathbf{H}^T + \mathbf{R})^{-1} \tag{20}$$

Covariance propagation is done as in (15).

3.3 Considerations in use of Kalman filters

Though Kalman filters provide a dramatic reduction in the computational cost relative to full Bayesian data assimilation, for a highly dimensional state vector, the computational costs in weather prediction models may still be extravagant. Consider the last line in (14). For an n-dimensional model state vector, it will require $2n$ applications of $\mathbf{M}$ to forecast the error covariances. Some reductions of computational expense may be

possible. For example, there have been suggestions that this computation may be more practical if the tangent-linear calculations are performed in a subspace of the leading singular vectors (Fisher 1998, Farrell and Ioannou 2001).

There is also the potential disadvantage of the assumption of linearity of error growth. While errors for large-scale variables may reasonably be assumed to grow linearly over a typical period between assimilation cycles of 3-6 h, some aspects, especially moist thermodynamic variables, may have errors which saturate on this time scale.

Much more can be said about the Kalman filter, such as its equivalence to 4D-Var under certain assumptions (perfect model, linearity, Gaussianity), the manner of computing $\mathbf{M}$, iterated extensions of the basic extended Kalman filter (Jazwinski 1970, Cohn 1997), and the properties of its estimators (which, in the case of the dicrete filter, if assumptions hold, provide the Best Linear Unbiased Estimate, or BLUE; see Talagrand 1997). These Kalman filters, however, are here but stepping stones toward the assimilation method we want to focus on, the ensemble Kalman filter.

4 The Ensemble Kalman Filter

Many research groups have sought practical remedies for the computational expense of the Kalman filter. One promising approach is a technique that has been dubbed the *ensemble Kalman filter*, or "EnKF." The EnKF's modern roots go back to Evensen (1994), though similar ensemble filters have been used for engineering and aerospace applications as far back as the 1960's (Potter 1964, Maybeck 1979). Here, we will provide a derivation of the EnKF. The subsequent sections will describe more fully some practical algorithmic details concerning how to implement ensemble filters as well as some demonstrations of their usefulness.

The main difference between the EnKF and the Kalman filter is that background-error covariances for the data assimilation are estimated from a finite sample of fully nonlinear ensemble forecasts. These ensemble forecasts start from a set of initial conditions that were created by the previous iteration of the EnKF. The EnKF is both an approximation to and an extention of the Kalman filter. Since covariances are estimated from a finite sample, they may be worse than those predicted by the Kalman filter. However, the Kalman filter itself makes the approximation that error covariances evolve according to linear dynamics, whereas the EnKF makes no such assumption. Hence, if error dynamics are strongly nonlinear, the EnKF may produce a more accurate estimate relative to the Kalman filter. However, to the extent that error dynamics are linear, we would like the behavior of the EnKF to converge to that obtained from the extended Kalman filter as the ensemble size increases.

The body of literature on ensemble-based data assimilation for atmospheric and oceanographic applications has blossomed in recent years (e.g., Evensen 1994, Evensen and van Leeuwen 1996, Burgers et al. 1998, Houtekamer and Mitchell 1998, 2001, van Leeuwen 1999, Lermusiaux and Robinson 1999, Anderson and Anderson 1999, Hamill and Snyder 2000, Keppenne 2000, Mitchell and Houtekamer 2000, Heemink et al. 2001, Hamill et al. 2001, Anderson 2001, Pham 2001, Whitaker and Hamill 2002, Reichle et al. 2002, Tippett et al. 2002, Mitchell et al. 2002). These articles describe the EnKF as well as various proposed extensions or variants on its basic design. Hopefully, this literature will be more readily accessible with an understanding of the basics provided here.

4.1 Update equations

We first demonstrate how an ensemble of background forecasts are updated in the EnKF and how they will have the correct asymptotic properties as ensemble size increases.

For notational simplicity, in this section we will drop the t time subscript used in previous sections; it is assumed unless noted otherwise that we are interested in the analysis at time t. We start off by assuming that we have an ensemble of forecasts that randomly sample the model background errors. Let's denote this ensemble as $\mathbf{X}^b$,

defined by

$$\mathbf{X}^{\mathrm{b}} = (\mathbf{x}_1^{\mathrm{b}}, \ldots, \mathbf{x}_m^{\mathrm{b}}) \tag{21}$$

where the subscript now denotes the ensemble member. $\mathbf{X}^{\mathrm{b}}$ is thus a matrix whose columns are comprised of ensemble member's state vectors. We define the ensemble mean $\overline{\mathbf{x}}^{\mathrm{b}}$ as

$$\overline{\mathbf{x}}^{\mathrm{b}} = \frac{1}{m}\sum_{i=1}^{m}\mathbf{x}_i^{\mathrm{b}} \tag{22}$$

The perturbation from the mean for the ith member is $\mathbf{x}_i'^{\mathrm{b}} = \mathbf{x}_i^{\mathrm{b}} - \overline{\mathbf{x}}^{\mathrm{b}}$. Let's define an ensemble of perturbations $\mathbf{X}'^{\mathrm{b}}$ as

$$\mathbf{X}'^{\mathrm{b}} = (\mathbf{x}_1'^{\mathrm{b}}, \ldots, \mathbf{x}_m'^{\mathrm{b}}) \tag{23}$$

Let $\hat{\mathbf{P}}^{\mathrm{b}}$ represent an estimate of $\mathbf{P}^{\mathrm{b}}$ from a finite ensemble:

$$\hat{\mathbf{P}}^{\mathrm{b}} = \frac{1}{m-1}\mathbf{X}'^{\mathrm{b}}\mathbf{X}'^{\mathrm{b}\mathrm{T}} \tag{24}$$

Also, let $\mathscr{M}$ denote the full, nonlinear forecast model operator that integrates the model state from time t to $t+1$.

Our goal is to construct a data assimilation system such that if the ensemble size is infinite and if the dynamics are linear, i.e., if $\mathscr{M}(\mathbf{x}_1^{\mathrm{a}}) - \mathscr{M}(\overline{\mathbf{x}}^{\mathrm{a}}) = \mathbf{M}(\mathbf{x}_1^{\mathrm{a}} - \overline{\mathbf{x}}^{\mathrm{a}})$, where $\mathbf{M} = \frac{\partial \mathscr{M}}{\partial \mathbf{x}}$ is the Jacobian of $\mathscr{M}$, we will converge to the extended Kalman filter solution. To wit, the core idea of the the EnKF is to run an *ensemble* of parallel data assimilation cycles, $i = 1, \ldots, m$, with each member updated to a somewhat different realization of the observations, as illustrated in Fig. 2:

$$\mathbf{x}_i^{\mathrm{a}} = \mathbf{x}_i^{\mathrm{b}} + \hat{\mathbf{K}}(\mathbf{y}_i - \mathscr{H}\mathbf{x}_i^{\mathrm{b}}) \tag{25}$$

Here, $\hat{\mathbf{K}} = \hat{\mathbf{P}}^{\mathrm{b}}\mathscr{H}^{\mathrm{T}}(\mathscr{H}\hat{\mathbf{P}}^{\mathrm{b}}\mathscr{H}^{\mathrm{T}} + \mathbf{R})^{-1}$. $\mathscr{H}$ is the observation operator, which is permitted here to be a nonlinear operator. For future reference, as with the extended Kalman filter, let's define $\mathbf{H}$ as the Jacobian of $\mathscr{H}$: $\mathbf{H} = \frac{\partial \mathscr{H}}{\partial \mathbf{x}}$. In (25), the $\mathbf{y}_i = \mathbf{y} + \mathbf{y}_i'$ are "perturbed observations," defined such that $\mathbf{y}_i' \sim N(0, \mathbf{R})$ and we ensure that

$$\frac{1}{m}\sum_{i=1}^{m}\mathbf{y}_i' = 0.$$

The m sets of perturbed observations are thus created to update the m different background fields.

For a complex numerical weather prediction model with a high-dimensional state vector, explicitly forming $\hat{\mathbf{P}}^{\mathrm{b}}$ as in (24) would be computationally prohibitive; for example, in a model with 10^6 degrees of freedom, storing and readily accessing the 10^{12} elements of $\hat{\mathbf{P}}^{\mathrm{b}}$ is not possible. However, in the EnKF, $\hat{\mathbf{K}}$ can be formed without ever explicitly computing the full $\hat{\mathbf{P}}^{\mathrm{b}}$. Instead, the components of $\hat{\mathbf{P}}^{\mathrm{b}}\mathscr{H}^{\mathrm{T}}$ and $\mathscr{H}\hat{\mathbf{P}}^{\mathrm{b}}\mathscr{H}^{\mathrm{T}}$ of $\hat{\mathbf{K}}$ are computed separately. Define

$$\overline{\mathscr{H}\mathbf{x}^{\mathrm{b}}} = \frac{1}{m}\sum_{i=1}^{n}\mathscr{H}\mathbf{x}_{\mathrm{i}}^{\mathrm{b}},$$

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

$$\hat{\mathbf{P}}^{\mathrm{b}}\mathscr{H}^{\mathrm{T}} = \frac{1}{m-1}\sum_{i=1}^{m}\left(\mathbf{x}_{\mathrm{i}}^{\mathrm{b}} - \overline{\mathbf{x}^{\mathrm{b}}}\right)\left(\mathscr{H}\mathbf{x}_{\mathrm{i}}^{\mathrm{b}} - \overline{\mathscr{H}\mathbf{x}^{\mathrm{b}}}\right)^{\mathrm{T}}, \tag{26}$$

and

$$\mathscr{H}\hat{\mathbf{P}}^{\mathrm{b}}\mathscr{H}^{\mathrm{T}} = \frac{1}{m-1}\sum_{i=1}^{m}\left(\mathscr{H}\mathbf{x}_{\mathrm{i}}^{\mathrm{b}} - \overline{\mathscr{H}\mathbf{x}^{\mathrm{b}}}\right)\left(\mathscr{H}\mathbf{x}_{\mathrm{i}}^{\mathrm{b}} - \overline{\mathscr{H}\mathbf{x}^{\mathrm{b}}}\right)^{\mathrm{T}}. \tag{27}$$

If departures from nonlinearity are again small, then $\mathscr{H}\mathbf{x}_{\mathrm{i}}^{\mathrm{b}} - \overline{\mathscr{H}\mathbf{x}^{\mathrm{b}}} \simeq \mathbf{H}(\mathbf{x}_{\mathrm{i}}^{\mathrm{b}} - \overline{\mathbf{x}}^{\mathrm{b}})$. In this case,

$$\hat{\mathbf{P}}^{\mathrm{b}}\mathscr{H}^{\mathrm{T}} \simeq \frac{1}{m-1}\sum_{i=1}^{m}\left(\mathbf{x}_{\mathrm{i}}^{\mathrm{b}} - \overline{\mathbf{x}}^{\mathrm{b}}\right)\left(\mathbf{x}_{\mathrm{i}}^{\mathrm{b}} - \overline{\mathbf{x}}^{\mathrm{b}}\right)^{\mathrm{T}}\mathbf{H}^{\mathrm{T}} = \hat{\mathbf{P}}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}}, \tag{28}$$

and

$$\mathscr{H}\hat{\mathbf{P}}^{\mathrm{b}}\mathscr{H}^{\mathrm{T}} \simeq \frac{1}{m-1}\sum_{i=1}^{m}\mathbf{H}\left(\mathbf{x}_{\mathrm{i}}^{\mathrm{b}}-\overline{\mathbf{x}}^{\mathrm{b}}\right)\left(\mathbf{x}_{\mathrm{i}}^{\mathrm{b}}-\overline{\mathbf{x}}^{\mathrm{b}}\right)^{\mathrm{T}}\mathbf{H}^{\mathrm{T}} \;=\; \mathbf{H}\hat{\mathbf{P}}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}}. \tag{29}$$

Hence, $\hat{\mathbf{K}} \rightarrow \mathbf{K}$ of the extended Kalman filter when $\hat{\mathbf{P}}^{\mathrm{b}} \rightarrow \mathbf{P}^{\mathrm{b}}$, which occurs as $m \rightarrow \infty$. Note that essentially what we are doing in (26) and (27) is using the ensemble to make a square-root representation of the background-error covariance. This trait will be shared among ensemble filters; see Tippett et al. (2002) for more discussion of this.

Why run a set of parallel data assimilation cycles, assimilating perturbed observations? We now would like to show that in the limit of infinite ensemble size, we end up with the same update equations in the EnKF that would be produced by the extended Kalman filter. Again, we assume that departures from nonlinearity in the observation operator are relatively small so that $\overline{\mathscr{H}\mathbf{x}^{\mathrm{b}}} \simeq \mathscr{H}\overline{\mathbf{x}}^{\mathrm{b}}$. The EnKF's update for the ensemble mean state is thus

$$\begin{aligned}
\overline{\mathbf{x}}^{\mathrm{a}} &= \frac{1}{m}\sum_{i=1}^{m}\mathbf{x}_i^{\mathrm{a}} \\
&= \frac{1}{m}\left(\sum_{i=1}^{m}\mathbf{x}_i^{\mathrm{b}} + \hat{\mathbf{K}}(\sum_{i=1}^{m}\mathbf{y}_i - \sum_{i=1}^{m}\mathscr{H}\mathbf{x}_i^{\mathrm{b}}\right) \\
&= \overline{\mathbf{x}}^{\mathrm{b}} + \hat{\mathbf{K}}(\mathbf{y} - \overline{\mathscr{H}\mathbf{x}^{\mathrm{b}}}) \\
&\simeq \overline{\mathbf{x}}^{\mathrm{b}} + \hat{\mathbf{K}}(\mathbf{y} - \mathscr{H}\overline{\mathbf{x}}^{\mathrm{b}})
\end{aligned} \tag{30}$$

Thus, (30) is equivalent to the extended Kalman filter update equation (19) as $m \rightarrow \infty$, assuming $\mathscr{H}$ is approximately linear.

The equation to show the asymptotic equivalence of the analysis-error covariances is much more tedious to derive. Interested readers can see the full derivation in Appendix 2. For here, we are content to note that again, as $m \rightarrow \infty$, $\hat{\mathbf{P}}^{\mathrm{a}} = \frac{1}{m-1}\left\langle \mathbf{X}'^{\mathrm{a}}\mathbf{X}'^{\mathrm{a}\mathrm{T}}\right\rangle \rightarrow \mathbf{P}^{\mathrm{a}} = (\mathbf{I}-\mathbf{K}\mathbf{H})\mathbf{P}^{\mathrm{b}}$ as was defined for the extended Kalman filter. As shown in Appendix 2, the covariance estimates from the extended and ensemble Kalman filters are equivalent only if the observations are perturbed.

Ideally, we would hope to produce a quality ensemble of analyses even when m is small; the more members, the larger the computational cost of the EnKF. To the extent that the error covariances can be approximated reasonably by a small sample $m << n$, the EnKF will have a large computational advantage over the extended Kalman filter.

4.2 Error covariance propagation and model-error parameterization

We have not discussed yet how to evolve the analysis-error covariances forward to produce an estimate of background-error covariances for the next assimilation time. In the discrete and extended Kalman filters, analysis-error covariances were propagated using the linear tangent and adjoint of the forecast model (eq 15), and a covariance $\mathbf{Q}$ was added to account for the uncertainty due to model error. In the ensemble Kalman filter, we'd like to take advantage of the potential increase in accuracy that may result from estimating covariances from an ensemble propagated with the fully nonlinear forecast model. If forecast-error dynamics are in fact quite nonlinear and saturate quickly, then the assumption of linearity in the discrete and extended Kalman filters was inappropriate, and some accuracy may be gained relative to the Kalman filters by estimating covariances from a sample of fully nonlinear model forecasts (Miller et al. 1999).

As we will see, we cannot just estimate background-error covariances at the next assimilation cycle by conducting an ensemble of forecasts forward from the current cycle's analyses. To understand why, let's return to (16). Because of model deficiencies, even if the true state of the atmosphere is perfectly known, the resulting

forecast will be imperfect: $\mathbf{x}_{(t+1)} = \mathscr{M}\mathbf{x}_{(t)} + \eta$, where here we denote the time index in parentheses. Let's first assume that our forecast model is unbiased $\left\langle \eta \right\rangle = 0$, again with model-error covariance $\left\langle \eta\eta^{\mathrm{T}} \right\rangle = \mathbf{Q}$. In practice, the assumption of no bias is probably not justified, and if the bias can be determined, the forecasts ought to be corrected (Dee and Todling 2000). In any case, we propagate an analysis state estimate forward and examine the error covariance at the next assimilation time. Assume again that forecast error and model error are uncorrelated $\left\langle \left(\mathscr{M}(\mathbf{x}^{\mathrm{a}}_{\mathrm{i}(t)}) - \mathscr{M}(\mathbf{x}_{(t)}) \right) \eta^{\mathrm{T}} \right\rangle = 0$, and assume linearity of the error growth $\mathscr{M}(\mathbf{x}^{\mathrm{a}}_{\mathrm{i}(t)}) - \mathscr{M}(\mathbf{x}_{(t)}) \simeq \mathbf{M}\left(\mathbf{x}^{\mathrm{a}}_{\mathrm{i}(t)} - \mathbf{x}_{(t)} \right)$. Then the true background-error covariance at the next assimilation time is

$$\begin{aligned}
\left\langle (\mathbf{x}^{\mathrm{b}}_{\mathrm{i}(t+1)} - \mathbf{x}_{(t+1)}) \, (\mathbf{x}^{\mathrm{b}}_{\mathrm{i}(t+1)} - \mathbf{x}_{(t+1)})^{\mathrm{T}} \right\rangle &= \left\langle (\mathscr{M}\mathbf{x}^{\mathrm{a}}_{\mathrm{i}(t)} - \mathscr{M}\mathbf{x}_{(t)} - \eta) \, (\mathscr{M}\mathbf{x}^{\mathrm{a}}_{\mathrm{i}(t)} - \mathscr{M}\mathbf{x}_{(t)} - \eta)^{\mathrm{T}} \right\rangle \\
&\simeq \left\langle \mathbf{M}\left(\mathbf{x}^{\mathrm{a}}_{\mathrm{i}(t)} - \mathbf{x}_{(t)} \right) \left(\mathbf{x}^{\mathrm{a}}_{\mathrm{i}(t)} - \mathbf{x}_{(t)} \right)^{\mathrm{T}} \mathbf{M}^{\mathrm{T}} \right\rangle + \left\langle \eta\eta^{\mathrm{T}} \right\rangle \\
&= \mathbf{M}\mathbf{P}^{\mathrm{a}}_{\mathrm{t}}\mathbf{M}^{\mathrm{T}} + \mathbf{Q} \qquad (31)
\end{aligned}$$

Consider what happens when we estimate covariances directly from an ensemble of forecasts. Say we propagate our ensemble of $i = 1, \ldots, m$ analyses forward with the fully nonlinear forecast model

$$\mathbf{x}^{\mathrm{b}}_{\mathrm{i}(t+1)} = \mathscr{M}(\mathbf{x}^{\mathrm{a}}_{\mathrm{i}(t)}), \qquad (32)$$

Calculating the covariance, we get

$$\begin{aligned}
\left\langle (\mathbf{x}^{\mathrm{b}}_{\mathrm{i}(t+1)} - \overline{\mathbf{x}}^{\mathrm{b}}_{(t+1)})(\mathbf{x}^{\mathrm{b}}_{\mathrm{i}(t+1)} - \overline{\mathbf{x}}^{\mathrm{b}}_{(t+1)})^{\mathrm{T}} \right\rangle &= \left\langle \left(\mathscr{M}(\mathbf{x}^{\mathrm{a}}_{\mathrm{i}(t)}) - \mathscr{M}(\overline{\mathbf{x}}^{\mathrm{a}}_{(t)}) \right) \left(\mathscr{M}(\mathbf{x}^{\mathrm{a}}_{\mathrm{i}(t)}) - \mathscr{M}(\overline{\mathbf{x}}^{\mathrm{a}}_{(t)}) \right)^{\mathrm{T}} \right\rangle \\
&\simeq \left\langle \mathbf{M}\left(\mathbf{x}^{\mathrm{a}}_{\mathrm{i}(t)} - \overline{\mathbf{x}}^{\mathrm{a}}_{(t)} \right) \left(\mathbf{x}^{\mathrm{a}}_{\mathrm{i}(t)} - \overline{\mathbf{x}}^{\mathrm{a}}_{(t)} \right)^{\mathrm{T}} \mathbf{M}^{\mathrm{T}} \right\rangle \\
&\simeq \mathbf{M}\hat{\mathbf{P}}^{\mathrm{a}}_{(t)}\mathbf{M}^{\mathrm{T}} \qquad (33)
\end{aligned}$$

Comparing (31) and (33), we can see that an ensemble of analyses that are simply propagated forward with the nonlinear forecast model will have too small an expected amount of spread, missing the extra covariance $\mathbf{Q}$. Let us define some hypothetical set of background forecasts at time $t+1$ that *do* have the correct covariance, i.e., define $x^{\mathrm{b}}_{\mathrm{i}(t+1)}$ such that $\left\langle (x^{\mathrm{b}}_{\mathrm{i}(t+1)} - \overline{x}^{\mathrm{b}}_{(t+1)})(x^{\mathrm{b}}_{\mathrm{i}(t+1)} - \overline{x}^{\mathrm{b}}_{(t+1)})^{\mathrm{T}} \right\rangle = \mathbf{M}\hat{\mathbf{P}}^{\mathrm{a}}_{(t)}\mathbf{M}^{\mathrm{T}} + \mathbf{Q}$. Such an ensemble is possible if we add noise to our existing ensemble:

$$x^{\mathrm{b}}_{\mathrm{i}(t+1)} = \mathbf{x}^{\mathrm{b}}_{\mathrm{i}(t+1)} + \xi_{\mathrm{i}} \qquad (34)$$

where $\left\langle \xi_{\mathrm{i}}\xi_{\mathrm{i}}^{\mathrm{T}} \right\rangle = \mathbf{Q}$ and $\left\langle \xi_{\mathrm{i}} \right\rangle = 0$.

Several methods have been considered for incorporating noise into the ensemble of forecasts so that they account for model error. First, one could actually make the forecast model stochastic, adding terms to the prognostic equations to represent interactions with unresolved scales and/or parameterized effects; in essence, $\mathscr{M}$ is changed so that the ensemble of forecasts integrates random noise in addition to the determininstic forecast dynamics. Over an assimilation cycle, this additional variance added to the ensemble as a result of integrating noise should increase the covariance by the missing $\mathbf{Q}$. Another possibility is that one may choose to run a forecast model without integrating noise but to add noise to each member at the data assimilation time so as to increase the ensemble variance appropriate to the missing $\mathbf{Q}$. Third, it may be possible to use a multi-model ensemble to estimate covariances.

Little work has yet been done on the first of these three approaches. Buizza et al. (1999) demonstrated a simple technique for integrating noise to account for deterministic sub-gridscale parameterizations. Under their methodology, the parameterized terms in the prognostic equations were multiplied by a random number. The method was quite heuristic but appears to have increased the spread in the ensemble forecasts somewhat

and increased their skill. Penland (2002) outlines a more general approach for integrating system noise in numerical models. To date, however, a comprehensive noise integration scheme has not yet been demonstrated in an operational weather prediction model. Still, this is a promising avenue for more research: the problem of deficient spread in ensemble forecasts will of course affect longer-range forecasts as well as shorter ones between data assimilation cycles. If the methodology can be designed for data assimilation purposes, it could be used to ameliorate spread deficiencies in subsequent ensemble forecasts as well.

The second general approach is to change the model of covariances at the data assimilation time to reflect model errors. Mitchell and Houtekamer (2000) detail such an approach. Following a methodology outlined in Dee (1995), innovation statistics $\mathbf{y} - \mathscr{H}\overline{\mathbf{x}}^{\mathrm{b}}$ from the data assimilation are used to estimate parameters of a simple model of the system noise covariances $\mathbf{Q}$. Random vectors that are consistent with the system noise covariance are added to each ensemble member, producing an ensemble with a larger spread. A disadvantage of this approach is that it is tailored specifically to making ensemble forecasts for data assimilation purposes, not for treating model error in ensemble forecasts in general.

The third approach, use of multiple forecast models for generating the ensemble of background forecasts (e.g., Houtekamer et al. 1996, Harrison et al. 1999, Evans et al. 2000, Ziehmann 2000, Richardson 2000, Hou et al. 2001), is appealing for its simplicity. A wider range of forecasts is typically generated when different weather forecast models are used to forecast the evolution of different ensemble members. Unfortunately, initial experimentation has shown that the multi-model ensembles tend to produce unrealistic estimates of error covariances. Forecast errors ought to be mostly balanced, but when estimated from multi-model ensembles, preliminary results suggest that the errors are excessively out of balance, with detrimental effects on the subsequent assimilation (personal communication, M. Buehner).

5 Extensions to deal with filter divergence

In the EnKF, algorithmic modifications are necessary to ensure that background-error covariances are not systematically underestimated. When they are, the filter will assume that the background is relatively accurate compared to the observations, and consequently the analysis is insufficiently corrected toward the observations. This underweighting of new observations can feed back on itself, the ensemble progressively ignoring observational data more and more in successive cycles, leading eventually to a useless ensemble. This detrimental process is known as "filter divergence" (e.g., Houtekamer and Mitchell 1998, van Leeuwen 1999, Hamill et al. 2001). For the EnKF, filter divergence can be caused by inappropriately treating model error, as discussed in the previous section. It can also be caused by using a finite ensemble to represent background error statistics. Hence, filter divergence can occur even if the forecast model is perfect.

Two potential sources of filter divergence are illustrated in Fig. 3. The first is an underestimate of background-error covariances. Fig. 3a illustrates a hypothetical posterior probability distribution when background-error covariances are estimated correctly (in this example, the covariance between the two state components is zero). If background errors are underestimated, the observation is comparatively underweighted (Fig. 3b) and the posterior distribution unduly resembles the prior. Similarly, if there are directions in phase space where the ensemble underestimates the true background covariances because of sampling errors, or at its worst assumes no variance at all because of the limited span of a finite number of ensemble members, then the background is not sufficiently corrected back toward the observation in these directions.

Another problem may be that the magnitude of background-error covariances between an observation location and a secondary, far-removed grid point are over-estimated due to sampling errors. If so, the posterior probability distribution at this secondary grid point will be adjusted too much (Fig. 3c). This can generate a posterior probability distribution that is biased and/or has too little variance. In probabilistic terms, the posterior distribution has insufficient probability in the region in phase space near to the true state.

The problems illustrated in Figs. 3b-c can be expected to show up when applied to more complex models, too.

Following Houtekamer and Mitchell (1998), Figures 4 a-b illustrate maps of covariances of 300 hPa temperature estimated from background ensembles of various sizes. Here, a dry, low-resolution general circulation model and an ensemble assimilation scheme (Whitaker and Hamill 2002) were used; these will be described in more depth in Section 8. Each map represents the spatial pattern of covariances of the ensemble background forecasts at each location with a location over Europe, marked by a dot. Note two general characteristics: first, for the 25-member ensemble in Fig. 4a, there are many locations worldwide that appear to co-vary with the ensemble at the observation location. Note especially the large covariances in the Southern Hemisphere oceans, where there are few observations and background-error variance is large (see Section 8). Second, note that the extent of the undesirably noisy covariances is diminished somewhat when the ensemble size is increased (Fig. 4b). If the Kalman gain calculation uses background-error covariances that are directly estimated from the ensemble, at grid points where there are larger covariances, the EnKF will modify the analysis proportionately. These unwanted corrections can bias the resulting mean analysis and result in too little variance in the analysis, contributing to filter divergence.

There are several possible measures that can be taken to prevent filter divergence. First, some adequate treatment of model error is necessary, perhaps along the lines of algorithms suggested in section 4b. By treating model error, the spread in the ensemble will be larger, resulting in a Kalman gain that draws the analysis away from the background forecasts and closer to the observations. By (11) and (12), increasing the background-error covariances also preserves more variance in the subsequent analysis. Another possible set of remedies are made possible by modifying the model of background-error covariances estimated from the ensemble. For example, perhaps one believes that the covariances are more trustworth near the observation but less so far from them (Houtekamer and Mitchell 2001, Hamill et al. 2001). The covariance estimate from the ensemble could be used more the near the observation location and damped to zero or a small value away from it. This is the essence of one technique to ameliorate filter divergence, a technique called "covariance localization." Under this, the covariance estimate from the ensemble is multiplied point by point with a correlation function that is 1.0 at the observation location and decreases monotonically with increasing distance. Mathematically, the Kalman gain equation $\hat{\mathbf{K}} = \hat{\mathbf{P}}^{\mathrm{b}}\mathscr{H}^{\mathrm{T}}(\mathscr{H}\hat{\mathbf{P}}^{\mathrm{b}}\mathscr{H}^{\mathrm{T}} + \mathbf{R})^{-1}$ is replaced by a modified gain

$$\hat{\mathbf{K}} = \left(\rho_S \circ \hat{\mathbf{P}}^{\mathrm{b}}\mathscr{H}^{\mathrm{T}}\right)\left(\mathscr{H}(\rho_S \circ \hat{\mathbf{P}}^{\mathrm{b}})\mathscr{H}^{\mathrm{T}} + \mathbf{R}\right)^{-1} \tag{35}$$

where the operation $\rho_S \circ$ in (35) denotes a Schur product (an element-by-element multiplication) of a correlation matrix **S** with the covariance model generated by the ensemble. The Schur product of matrices **A** and **B** is a matrix **C** of the same dimension, where $C_{ij} = A_{ij}B_{ij}$. The application of the Schur product to the covariance estimate from the 100-member ensemble is shown in Fig. 4c. Note the damping of covariances everywhere but in the region around the observation.

See Hamill et al. (2001) for more details on implementation of covariance localization, and see Mitchell and Houtekamer (2002) for a discussion of balance issues related to the length scale of the covariance localization function.

Another problem, illustrated in Fig. 3b, was the detrimental effect of underestimating the variances. As a remedy, 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. Before the first observation is assimilated in a new cycle, background forecasts deviation from the mean are inflated by an amount r, slightly greater than 1.0:

$$\mathbf{x}_{\mathrm{i}}^{\mathrm{b}} \leftarrow r\left(\mathbf{x}_{\mathrm{i}}^{\mathrm{b}} - \overline{\mathbf{x}}^{\mathrm{b}}\right) + \overline{\mathbf{x}}^{\mathrm{b}} \tag{36}$$

Here, the operation $\leftarrow$ denotes a replacement of the previous value of $\mathbf{x}_{\mathrm{i}}^{\mathrm{b}}$.

Houtekamer and Mitchell (1998) have proposed the use of a "double" ensemble Kalman filter that eliminates the necessity of using an inflation factor. The ensemble is split into two separate sub-ensembles, where one ensemble is used to estimate error covariances for the other. Systematic underestimation of covariances are much less likely to happen in the double EnKF, though for the same accuracy of covariance estimates, twice as many ensemble members are needed.

6 A Simple Demonstration of the Update Step in the Ensemble Kalman Filter

Let us return to the data assimilation problem illustrated back in Fig. 1, when we were discussing Bayesian data assimilation. There, we were seeking to update a 2-D probability distribution given an observation in one dimension. Let's explore the characteristics of the EnKF applied to this problem.

We sidestep the issue of system noise and model error discussed in the previous section. We simply assume that we can generate a realistic random sample from the prior distribution for purposes of demonstrating the data assimilation methodology. Accordingly, we start by generating an $m = 50$-member random sample from the distribution in Fig. 1(a). These are denoted by the black dots in Fig. 3(a). We'll keep special track of the assimilation for one particular member, denoted by the larger black dot.

The EnKF assumes that the background-error probability distribution is normally distributed. Estimated from this random sample,

$$\hat{\mathbf{P}}^{\mathrm{b}} = \begin{pmatrix} \sigma^2(x^{\mathrm{b}}_{(1)}) & Cov(x^{\mathrm{b}}_{(1)}, x^{\mathrm{b}}_{(2)}) \\ Cov(x^{\mathrm{b}}_{(1)}, x^{\mathrm{b}}_{(2)}) & \sigma^2(x^{\mathrm{b}}_{(2)}) \end{pmatrix} = \begin{pmatrix} 121.03 & 115.47 \\ 115.47 & 232.72 \end{pmatrix}$$

Here, since our observation measures the same aspect as the first component of our state variable, $\mathscr{H} = [1, 0]$. We assume $\mathbf{R} = 100$, so $\mathscr{H}\hat{\mathbf{P}}^{\mathrm{b}}\mathscr{H}^{\mathrm{T}} + \mathbf{R} = 121.03 + 100 = 221.03$. $\hat{\mathbf{P}}^{\mathrm{b}}\mathscr{H}^{\mathrm{T}} = [121.03, 115.47]^{\mathrm{T}}$, and hence $\hat{\mathbf{K}} = \mathbf{P}^{\mathrm{b}}\mathscr{H}^{\mathrm{T}}(\mathscr{H}\mathbf{P}^{\mathrm{b}}\mathscr{H}^{\mathrm{T}} + \mathbf{R})^{-1} \simeq [.547, .512]^{\mathrm{T}}$. We then apply (25), updating background samples to their associated perturbed observations, generating analysis samples. For example, the heavy black dot in Fig. 3(a) denotes one such background sample. This particular background is updated to the perturbed observation marked with the "*". The resulting analysis sample is the large black dot in Fig. 3(b). The analysis has been adjusted from the background in both directions, consistent with the error statistics built into $\hat{\mathbf{K}}$. Here, the first component of the background state was much less than the mean, and the perturbed observation was greater than the mean background state. The resulting analysis nudged the posterior state toward the mean in both components.

Compare the EnKF random samples of the posterior from Fig. 3b with the Bayesian posterior in Fig. 1b. The EnKF assumed from the start that the prior distribution was normally distributed, when in fact it was bimodal. Consequently, the samples from Fig. 3b do not appear to randomly sample the posterior from Fig. 1b, which can be seen by comparing the fitted normal distribution for the EnKF posterior to the Bayesian posterior. The EnKF posterior is shifted slightly toward lower values in both components. The underlying problem is both in the assumption of normality and in the limited ensemble size. Hopefully, scenarios with non-normal distributions such as in Fig. 1(a) are relatively rare, as more involved techniques would then be required (e.g., Gordon et al. 1993).

7 Making the Ensemble Kalman Filter Computationally Tractable

Though the EnKF is computationally less expensive than the Kalman filter, it is still a much more expensive calculation than 3D-Var and it is probably roughly the same order of magnitude as 4D-Var. The computational expense of the EnKF will scale with the number of observations times the number of ensemble members times the dimension of the state vector. Practically, though, the expense may be more determined by factors such as the extent of parallelization and the complexity of the operator $\mathscr{H}$. For example, with observations such as radiances, $\mathscr{H}$ may be computationally very expensive, while with observations like rawinsonde temperature data, simple interpolations to the observation location may be all that is needed. Since $\mathscr{H}$ is calculated for each ensemble member, if $\mathscr{H}$ is particularly expensive, this may dramatically increase the overall cost of the ensemble filter.

How can we minimize the computational time of these ensemble filters? We have already discussed one short-cut, avoiding the formulation of $\hat{\mathbf{P}}^{\mathrm{b}}$ directly and calculating the gain components directly from the ensemble.

(eqs. 26 and 27). We now consider two other methods for simplifying the calculations, serial (sequential) processing of observations and parallel-processing methods.

7.1 Serial Processing of Observations

Gelb (1974) demonstrated in the Kalman filter that a batch of independent observations can be assimilated either simultaneously or sequentially (serially). Starting with a background forecast and background error-covariances, a single observation may be assimilated. The analysis state and analysis-error covariance resulting from the assimilation of the first observation may be used as the background and background-error covariance in the assimilation of the second observation. Regardless of whether the observation are assimilated simultaneously or serially, the same expected analysis and analysis-error covariance will result. Consequentially, the order in which observations are serially assimilated is unimportant.

Serial processing of observations may be desirable, as both the coding and the computations are somewhat less demanding. Most of the ensemble-based algorithms described in the literature serially assimilate observations. When observations are assimilated serially, for each observation that is assimilated, $\mathscr{H}\hat{\mathbf{P}}^{\mathrm{b}}\mathscr{H}^{\mathrm{T}}$ and $\mathbf{R}$ become scalars. Thus, their inverse $(\mathscr{H}\hat{\mathbf{P}}^{\mathrm{b}}\mathscr{H}^{\mathrm{T}}+\mathbf{R})^{-1}$ in the gain matrix is trivial to compute. Also, the application of the covariance localization in (35) is much more straightforward.

The equivalence of serial and simultaneous assimilation is often assumed in the ensemble Kalman filter. Unfortunately, unlike the Kalman filter, somewhat different analyses may result if observation 1 is processed before observation 2 instead of 2 before 1, so the equivalence is not strictly correct. The reasons for this and an alternative algorithm to the EnKF where observation order is irrelevant are discussed in Whitaker and Hamill (2002). Regardless, for a reasonably large ensemble (50-100 members), order is nearly irrelevant in the EnKF, introducing errors that are much smaller than sampling errors in the covariance estimate.

The equivalence of serial and simultaneous processing is only true if observations have independent errors. Practically, however, many observations may have vertically or horizontally correlated errors. Consider two alternatives to deal with this. First, if the size of a batch of observations with correlated errors is relatively small, these batches can be processed simultanteously without much more computational expense; the matrix inverse of $(\mathscr{H}\hat{\mathbf{P}}^{\mathrm{b}}\mathscr{H}^{\mathrm{T}}+\mathbf{R})^{-1}$ should not be prohibitively expensive. Another option is to transform the observations and the update equation so that the observations are effectively independent. The method for doing this is derived in Appendix 3.

7.2 Parallel Processing

Many modern computers today are "massively parallel" computers consisting of individual central processing units (CPUs) with their own memory. Such computers often have relatively fast computational speed on each CPU, but the overall speed of a calculation can be slowed significantly when new data must be continually exchanged between CPUs. Hence, to speed up the processing speed of the EnKF on such computers, the parallelization ought to be designed so a minimum of data need be exchanged among the CPUs.

There are two general components to the EnKF; first, for an *m*-member EnKF, *m* parallel data assimilation cycles will need to be computed, and second, from the *m* resulting analyses, *m* forecasts forward to the next assimilation time. The forecast component is easily parallelizable. Each member of the ensemble can be forecast in parallel on a separate CPU(s), since no information need be swapped between CPUs during the forecast step. However, parallelizing the data assimilation component is considerably more difficult. The update of each member background forecast to the new observations requires information from all the other ensemble members in formulating the gain matrix $\hat{\mathbf{K}}$. It would be highly inefficient to simply parallelize so that different members were updated on different processors, since each processor would be duplicating the same computationally expensive gain calculation. How, then, might we speed up the computation of parallel analyses?

One reasonably simple thing to do is to parallelize over widely separated observations or batches of observations. Envision two observations on opposite sides of the world. With covariance localization, each observation corrects the background at a mutually exclusive set of grid points. In this case, regardless of whether the two observations are processed serially or simultaneously in parallel, the same analyses will result. This is one simple example of how the computations in the EnKF may be parallelized. Houtekamer and Mitchell (2001) discuss the design and testing of a parallel ensemble Kalman filter exploiting this algorithm.

How else might the work of the ensemble filter be split up among several CPUs? Typically, one of the most computationally expensive steps during the EnKF is the calculation of the $\hat{\mathbf{K}}$, and in particular, the computation of the term $\hat{\mathbf{P}}^{\mathrm{b}}\mathscr{H}^{\mathrm{T}}$. If this could be parallelized, it could result in dramatic cost savings. Following (26), when observations are processed serially, this term is computed from a product of the $n \times m$ matrix of ensemble perturbations and the m-dimensional vector of observation operators applied to the ensemble perturbations. Consider, now, that the m-dimensional state vector is split up over a number of processors. Perhaps the ensemble of northern hemisphere grid points are processed on CPU 1, the southern hemisphere on CPU 2. Say we are to process two observations serially, one just north of the equator and one south of it. For the first observation, the ensemble of $\left(\mathscr{H}\mathbf{x}_{\mathrm{i}}^{\mathrm{b}} - \overline{\mathscr{H}\mathbf{x}^{\mathrm{b}}}\right)^{\mathrm{T}}$ from (26) could be computed on CPU 1 and then transferred to processor 2. CPU 1 would then calculate $\hat{\mathbf{P}}^{\mathrm{b}}\mathscr{H}^{\mathrm{T}}$ for the northern hemisphere, CPU 2 for the southern hemisphere. Following that, the update (25) would similarly be split amongst the two CPUs. When the second observation in the southern hemisphere is ready to be processed, the ensemble of $\left(\mathscr{H}\mathbf{x}_{\mathrm{i}}^{\mathrm{b}} - \overline{\mathscr{H}\mathbf{x}^{\mathrm{b}}}\right)^{\mathrm{T}}$ can be computed on CPU 2 and that data shipped back to CPU 1. The computationally expensive part of the gain calculation (26) and the update (25) are then again computed in parallel. This general method of parallelization could be split up over any arbitrarily large number of CPUs. As long as the operator $\mathscr{H}$ is relatively simple, then the step of computing the $\mathscr{H}\mathbf{x}_{\mathrm{i}}^{\mathrm{b}}$ should proceed relatively quickly on one processor (not leaving the other processors idle for long), and the amount of data shipped between processors should be minimal.

8 Demonstration of ensemble-based data assimilation

We now demonstrate an ensemble-based data assimilation methodology in a simplified general circulation model. Specifically, we attempt to show how the use of flow-dependent background-error covariances from the ensemble filter can lead to dramatic reductions in the analysis errors. We will not focus on comparisons against other assimilation schemes; for such results, see, for example, Hamill and Snyder (2000), Anderson (2001), and Hamill and Snyder (2002).

The experiment is conducted under perfect-model assumptions; that is, the same forecast model is used to generate both a synthetic true state and in the conduct of the ensemble forecasts. To generate a time series of the true state, we started with a random perturbation superimposed upon a resting state. The model is then integrated for 280 days. The first 100 days were discarded, and the remaining 180 days comprise the time series of the true state used in this experiment. Hereafter, day 100 is considered the starting point, the day 0 for all further experiments.

Errors will be measured in a total energy norm. We follow the definition of total energy from Ehrendorfer and Errico (1995):

$$\| \cdot \| = \sqrt{\frac{1}{2}\int_D \int_0^1 \left[u^2 + v^2 + \frac{c_p}{T_r}T^2 + R_d T_r \left(\frac{p_s}{p_r}\right)^2\right] d\sigma \, dD} \tag{37}$$

where D indicates the horizontal domain, σ is the vertical coordinate, T_r is a reference temperature (here, 300 K), R_d is the gas constant for dry air (287 $J\,K^{-1}\,kg^{-1}$), and c_p is the specific heat of dry air at constant pressure (1004 $J\,K^{-1}\,kg^{-1}$), p_s is the surface pressure, and p_r is a reference pressure (1000 hPa).

8.1 Forecast model

A T51 L15 dry, primitive equation spectral model will be used in the following experiments. There is no terrain nor surface variations. The model has 60,996 degrees of freedom. The prognostic variables are vorticity, divergence, temperature, and surface pressure. Except for a minor modification to the forcing, described in Hamill et al. (2002), the model is essentially equivalent to the model of Held and Suarez (1994). The model has an error-doubling time of $\sim$ 3.1 days.

8.2 Observations

Synthetic rawinsondes (raobs) were assimilated every 12 h. The observations consisted of a surface pressure measurement and winds and temperatures at 7 of the sigma levels, located approximately at 900, 766, 633, 500, 366, 233, and 100 hPa. Observations have error characteristics derived from Parrish and Derber (1992), and observation errors were assumed uncorrelated in the vertical. Observation locations are shown in Fig. 6; they were chosen to provide a crude analog to the operational raob network, with more observations over the land than the ocean.

8.3 Assimilation methodology

The ensemble-based data assimilation methodology used here is a variant of the ensemble Kalman filter, a technique we have called the *ensemble square-root filter*, or "EnSRF." It is very similar to the EnKF, but the algorithm is designed in a manner where each of the parallel data assimilation cycles assimilates the same unperturbed observations. We describe the scheme briefly here; for more information, see Whitaker and Hamill (2002).

Like the EnKF, the EnSRF conducts a set of parallel data assimilation cycles. It is convenient in the EnSRF to update the equations for the ensemble mean (denoted by an overbar) and the deviation of the *i*th member from the mean separately:

$$\overline{\mathbf{x}}^{\mathrm{a}} = \overline{\mathbf{x}}^{\mathrm{b}} + \hat{\mathbf{K}}(\mathbf{y} - \mathscr{H}\overline{\mathbf{x}}^{\mathrm{b}}), \tag{38}$$

$$\mathbf{x}_i'^{\mathrm{a}} = (\mathbf{I} - \widetilde{\mathbf{K}}\mathscr{H})\mathbf{x}_{\mathbf{i}}'^{\mathrm{b}}. \tag{39}$$

Here, $\hat{\mathbf{K}}$ is the traditional Kalman gain as in Eq. (25), and $\widetilde{\mathbf{K}}$ is the "reduced" gain used to update deviations from the ensemble mean.

When sequentially processing independent observations, $\hat{\mathbf{K}}$, $\widetilde{\mathbf{K}}$, $\mathscr{H}\hat{\mathbf{P}}^{\mathrm{b}}$ and $\hat{\mathbf{P}}^{\mathrm{b}}\mathscr{H}^{\mathrm{T}}$ are all vectors with the same length as the model state vector, and $\mathscr{H}\hat{\mathbf{P}}^{\mathrm{b}}\mathscr{H}^{\mathrm{T}}$ and $\mathbf{R}$ are scalars. Thus, as first noted by Potter (1964), when observations are processed one at a time,

$$\widetilde{\mathbf{K}} = \left(1 + \sqrt{\frac{\mathbf{R}}{\mathscr{H}\hat{\mathbf{P}}^{\mathrm{b}}\mathscr{H}^{\mathrm{T}} + \mathrm{R}}}\right)^{-1} \hat{\mathbf{K}}. \tag{40}$$

The quantity multiplying $\hat{\mathbf{K}}$ in Eq. (40) is thus a scalar between 0 and 1. This means that, in order to obtain the correct analysis-error covariance with unperturbed observations, one uses a modified Kalman gain to update deviations from the ensemble mean that is reduced in magnitude relative to the traditional Kalman gain. Consequently, deviations from the mean are reduced less in the analysis using $\widetilde{\mathbf{K}}$ than they would be using $\hat{\mathbf{K}}$. In the EnKF, the excess variance reduction caused by using $\hat{\mathbf{K}}$ to update deviations from the mean is compensated for by the introduction of noise to the observations. In the EnSRF, the mean and departures from the mean are updated independently according to Eqs. (38) and (39). If observations are processed one at a time, the EnSRF requires about the same computation as the traditional EnKF with perturbed observations.

The general analysis methodology is thus as follows: generate a set of perturbed initial conditions. Make n forecasts forward to the next data assimilation time. Perform $n+1$ parallel data assimilation cycles, updating

the mean state using (38) and the n perturbations using (39) and (40). Repeat the process. In each data assimilation cycle, observations are assimilated serially.

Our experiment was conducted over an 180-day period and used a 100-member EnSRF data assimilation system. In this implementation of the EnSRF, covariances were localized using a Schur product of ensemble covariances with an $\sim$ Gaussian-shaped function with local support (Gaspari and Cohn 1999) reaching a zero value at 4000 km distance from the observation. Before each data assimilation cycle, ensemble deviation from the mean were inflated by 1.2 % following (33). The ensemble is initialized with an ensemble consisting 80 percent of the true state and 20 percent of a random model state.

8.4 Results

First, consider a time series of analysis errors in an energy norm. This initial ensemble is not very accurate, but the data assimilation rather quickly reduces the error in the analysis (Fig. 7). On average, the errors remain largest in the southern hemisphere, where there are few observations (Fig. 8).

As the ensemble cycles forward, the covariances begin to reflect more and more the flow-dependent error structures. Figure 9 provides an illustration of the background-error covariances used in the assimilation around four selected observation locations. Notice that the covariances have very complicated structures. Consider the covariances around the grid point in the eastern Pacific. Here, the observation is just north of the front, and the model of covariances indicates that covariances are not largest at the observation location, but rather slightly southeast of the observation, in the region of the front. Further, the covariances are elongated along the isotherms rather than across it. When considering the assimilation of a single observation, the Kalman gain $\hat{\mathbf{K}}$ will be directly proportional to $\hat{\mathbf{P}}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}}$, that is, proportional to the covariance with the observation location. Hence, this indicates that a small observation increment at this location in the Pacific will make larger corrections to the background state in the region of the nearby front than at the observation location itself. This is very different than the corrections that 3D-Var would make, which would be largest at the observation and decrease with increasing distance from the observation. The background-error covariance model from the ensemble filter may make more synoptic sense; a small change in the position of the front may result in a large change of temperature there, but that change is lessened outside the frontal zone (Fig. 10).

Notice also in Fig. 9 that the background-error covariances have much different structures around each of the observation locations. For example, that the covariances around the observation in eastern Europe are much smaller in magnitude. Part of the reason covariances are smaller is because this observation is in a more data-rich region. As a result of the smaller covariances, the analysis in this region will be drawn less to new observation than it will for the observation in the eastern Pacific.

9 Where Next?

The field of ensemble-based atmospheric data assimilation is a very new one. To this date, because of the computational expense and the need for a basic understanding of these approaches, most of the experimentation with ensemble filters have been done in simple models, often using assumptions that are unrealistic for practical numerical weather prediction (for example, the assumption of a perfect model). The results with these simple experiments have been uniformly impressive, indicating that testing in more realistic scenarios is warranted. That is the state of the field in 2002. Several groups, most notably researchers at the Canadian Meteorological Centre, are testing the methodology with real observational data.

We still have much to learn about ensemble-based data assimilation methodologies. We are still learning how to parameterize model error. Head-to-head comparisons against 4D-Var have not yet been performed in operational weather prediction models, so the appeal is more theoretical than evidence-based. The extent to which the underlying assumptions of these filters are met (such as Gaussianity) are not well known. Many

practical problems such as ensuring balanced initial conditions may need to be addressed. Despite the problems, the potential upside of ensemble assimilation methodologies is huge. Since no tangent-linear or adjoint models are needed, the code is relatively simple to develop. Unlike 4D-Var, linearity of error growth is not assumed, and model error should be able to be treated in an intelligent fashion.

Because of the upside, expect a burgeoning literature on ensemble-based data assimilation in the coming years.

10 Appendix 1: Derivation of the Discrete Kalman Filter Update equations

10.1 *a. Discrete Kalman filter state update equation*

Given the functional

$$J(\mathbf{x}_t) = \frac{1}{2}\left[(\mathbf{x}_t - \mathbf{x}_t^b)^T\, \mathbf{P}_t^{b\,-1}(\mathbf{x}_t - \mathbf{x}_t^b) + (\mathbf{H}\mathbf{x}_t - \mathbf{y}_t)^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{x}_t - \mathbf{y}_t)\right] \tag{A1}$$

we seek to find the model state that minimizes this functional, providing the best fit between the observations and background. Let's call this model state "the analysis," or $\mathbf{x}_t^a$. Differentiating the functional with respect to $\mathbf{x}_t^a$ and setting the result equal to zero to find the expected minimum, we get

$$\mathbf{P}_t^{b\,-1}(\mathbf{x}_t^a - \mathbf{x}_t^b) + \mathbf{H}^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{x}_t^a - \mathbf{y}_t) = 0 \tag{A2}$$

Rearranging terms, we get

$$\mathbf{x}_t^a = \left(\mathbf{P}_t^{b\,-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\left[\mathbf{P}_t^{b\,-1}\mathbf{x}_t^b + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}_t\right] \tag{A3}$$

We now put to use the Sherman-Morrison-Woodbury formula (Golub and van Loan 1989). For matrices $\mathbf{A}$, $\mathbf{U}$, and $\mathbf{V}$,

$$(\mathbf{A} + \mathbf{U}\mathbf{V}^T)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{I} + \mathbf{V}^T\mathbf{A}^{-1}\mathbf{U})^{-1}\mathbf{V}^T\mathbf{A}^{-1}. \tag{A4}$$

Using this, where $\mathbf{A} = \mathbf{P}_t^{b\,-1}$, $\mathbf{U} = \mathbf{H}^T\mathbf{R}^{-1}$, and $\mathbf{V}^T = \mathbf{H}$, we get

$$\mathbf{x}_t^a = \left(\mathbf{P}_t^b - \mathbf{P}_t^b\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{I} + \mathbf{H}\mathbf{P}_t^b\mathbf{H}^T\mathbf{R}^{-1})^{-1}\mathbf{H}\mathbf{P}_t^b\right)\left[\mathbf{P}_t^{b\,-1}\mathbf{x}_t^b + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}_t\right] \tag{A5}$$

Next, using the identity $\mathbf{A}^{-1}\mathbf{B}^{-1} = (\mathbf{B}\mathbf{A})^{-1}$ for two invertible matrices $\mathbf{A}$ and $\mathbf{B}$, where $\mathbf{A} = \mathbf{R}$ and $\mathbf{B} = \mathbf{I} + \mathbf{H}\mathbf{P}_t^b\mathbf{H}^T\mathbf{R}^{-1}$ (A5) becomes

$$\mathbf{x}_t^a = \left(\mathbf{P}_t^b - \mathbf{P}_t^b\mathbf{H}^T(\mathbf{H}\mathbf{P}_t^b\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{P}_t^b\right)\left[\mathbf{P}_t^{b\,-1}\mathbf{x}_t^b + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}_t\right] \tag{A6}$$

Expanding (A6), we get

$$\begin{aligned}
\mathbf{x}_t^a = {} & \mathbf{x}_t^b - \mathbf{P}_t^b\mathbf{H}^T(\mathbf{H}\mathbf{P}_t^b\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{x}_t^b + \mathbf{P}_t^b\mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}_t - \mathbf{P}_t^b\mathbf{H}^T(\mathbf{H}\mathbf{P}_t^b\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{P}_t^b\mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}_t \\
= {} & \mathbf{x}_t^b - \mathbf{P}_t^b\mathbf{H}^T(\mathbf{H}\mathbf{P}_t^b\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{x}_t^b + \mathbf{P}_t^b\mathbf{H}^T(\mathbf{R}^{-1} - (\mathbf{H}\mathbf{P}_t^b\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{P}_t^b\mathbf{H}^T\mathbf{R}^{-1})\mathbf{y}_t \\
= {} & \mathbf{x}_t^b - \mathbf{P}_t^b\mathbf{H}^T(\mathbf{H}\mathbf{P}_t^b\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}\mathbf{x}_t^b + \mathbf{P}_t^b\mathbf{H}^T(\mathbf{R}^{-1} - (\mathbf{H}\mathbf{P}_t^b\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{H}\mathbf{P}_t^b\mathbf{H}^T + \mathbf{R} - \mathbf{R})\mathbf{R}^{-1})\mathbf{y}_t
\end{aligned} \tag{A7}$$

Using the vector identity $\mathbf{B}^{-1} = \mathbf{A}^{-1} - \mathbf{B}^{-1}(\mathbf{B} - \mathbf{A})\mathbf{A}^{-1}$, (Golub and Van Loan 1989) to $\mathbf{R}^{-1} - (\mathbf{H}\mathbf{P}_t^b\mathbf{H}^T + \mathbf{R})^{-1}(\mathbf{H}\mathbf{P}_t^b\mathbf{H}^T + \mathbf{R} - \mathbf{R})\mathbf{R}^{-1}$, where $\mathbf{A} = \mathbf{R}^{-1}$ and $\mathbf{B} = \mathbf{H}\mathbf{P}_t^b\mathbf{H}^T + \mathbf{R}$, (A7) simplifies to

$$\mathbf{x}_t^a = \mathbf{x}_t^b + \left[\mathbf{P}_t^b\mathbf{H}^T(\mathbf{H}\mathbf{P}_t^b\mathbf{H}^T + \mathbf{R})^{-1}\right](\mathbf{y}_t - \mathbf{H}\mathbf{x}_t^b) \tag{A8}$$

This is update equation for the model state in the Kalman filter. Commonly we define a matrix $\mathbf{K}$, the Kalman gain matrix, as

$$\mathbf{K} = \mathbf{P}_t^b \mathbf{H}^T (\mathbf{H}\mathbf{P}_t^b \mathbf{H}^T + \mathbf{R})^{-1} \quad (A9)$$

so (A8) is expressed more compactly as

$$\mathbf{x}_t^a = \mathbf{x}_t^b + \mathbf{K}(\mathbf{y}_t - \mathbf{H}\mathbf{x}_t^b) \quad (A10)$$

Interpretation of this equation is described in the main text.

10.2 *b. Discrete Kalman filter covariance update equation*

The crucial difference between the Kalman filter compared with 3D-Var or other analysis schemes is that the error covariances are explicitly evolved in time through the data assimilation and through the subsequent forecast. Assume that we know the background-error covariance $\mathbf{P}_t^b$ and seek to know the resulting analysis-error covariance $\mathbf{P}_t^a$ that results from the assimilation of observations. We start with the definition of $\mathbf{P}_t^a$:'

$$\mathbf{P}_t^a = \left\langle (\mathbf{x}_t - \mathbf{x}_t^a)(\mathbf{x}_t - \mathbf{x}_t^a)^T \right\rangle. \quad (A11)$$

Subtracting both sides of (A10) from $\mathbf{x}_t$ gives

$$\mathbf{x}_t - \mathbf{x}_t^a = \Big(\mathbf{I} - \mathbf{K}\mathbf{H}\Big)(\mathbf{x}_t - \mathbf{x}_t^b) - \mathbf{K}(\mathbf{y}_t - \mathbf{H}\mathbf{x}_t). \quad (A12)$$

As previously, we assume $\mathbf{y}_t = \mathbf{H}\mathbf{x}_t + \varepsilon$, where $\varepsilon \sim N(0, \mathbf{R})$. In this case, we can write (A12) as

$$\begin{aligned} \mathbf{x}_t - \mathbf{x}_t^a &= \mathbf{x}_t - \mathbf{x}_t^b - \mathbf{K}(\varepsilon + \mathbf{H}\mathbf{x}_t - \mathbf{H}\mathbf{x}_t^b) \\ &= (\mathbf{I} - \mathbf{K}\mathbf{H})(\mathbf{x}_t - \mathbf{x}_t^b) - \mathbf{K}\varepsilon \end{aligned} \quad (A13)$$

Now, let us form the covariance matrix (A11). Assuming that observation and background errors are uncorrelated, i.e., $\left\langle (\mathbf{x}_t - \mathbf{x}_t^b)\ \varepsilon^T \right\rangle = 0$, and noting that $\mathbf{P}_t^b = \mathbf{P}_t^{bT}$ and using the matrix identity $\mathbf{A}^T\mathbf{B}^T = (\mathbf{A}\mathbf{B})^T$, we get

$$\begin{aligned} \mathbf{P}_t^a &= \left\langle (\mathbf{x}_t - \mathbf{x}_t^a)(\mathbf{x}_t - \mathbf{x}_t^a)^T \right\rangle \\ &= \left\langle (\mathbf{I} - \mathbf{K}\mathbf{H})(\mathbf{x}_t - \mathbf{x}_t^b)(\mathbf{x}_t - \mathbf{x}_t^b)^T (\mathbf{I} - \mathbf{K}\mathbf{H})^T + \mathbf{K}\varepsilon\varepsilon^T\mathbf{K}^T \right\rangle \\ &= \Big(\mathbf{I} - \mathbf{K}\mathbf{H}\Big)\mathbf{P}_t^b\Big(\mathbf{I} - \mathbf{K}\mathbf{H}\Big)^T + \mathbf{K}\mathbf{R}\mathbf{K}^T \\ &= \mathbf{P}_t^b - \mathbf{K}\mathbf{H}\mathbf{P}_t^b - \Big(\mathbf{K}\mathbf{H}\mathbf{P}_t^b\Big)^T + \mathbf{K}\Big(\mathbf{H}\mathbf{P}_t^b\mathbf{H}^T + \mathbf{R}\Big)\mathbf{K}^T. \end{aligned} \quad (A14)$$

Inserting the definition of $\mathbf{K}$ from (A9) and expanding, we simplify to

$$\begin{aligned} \mathbf{P}_t^a &= \mathbf{P}_t^b - \mathbf{P}_t^b\mathbf{H}^T\Big(\mathbf{H}\mathbf{P}_t^b\mathbf{H}^T + \mathbf{R}\Big)^{-1}\mathbf{H}\mathbf{P}_t^b \\ &= \Big(\mathbf{I} - \mathbf{K}\mathbf{H}\Big)\mathbf{P}_t^b. \end{aligned} \quad (A15)$$

This is the update equation for the covariances. We note in passing that the assumption that observation and background errors are uncorrelated is not always realistic. The observation-error covariance $\mathbf{R}$ is commonly thought of as representing both random errors from the instrument as well as "representativeness" errors, denoting the fact that an observation typically measures a point value not a grid volume; hence the point value may not be representative of the average value for that grid volume. One may imagine that representativeness errors may be larger in situations where there is a strong gradient in the background, such as near a front. In this manner, representativeness errors may in fact have some correlation with the background errors, one being larger when the other is. For simplicity, we will neglect this source of error, but it may be helpful to keep in mind where such simplifying assumptions have been made.

11 Appendix 2: Derivation of the Ensemble Kalman Filter covariance update

Here we show that the analysis error covariance estimated from the ensemble converges to that which would be obtained from the Kalman filter as ensemble size increases. That is, for an m-member ensemble, if $\mathbf{X}'^{\mathrm{a}}$ is defined as $\mathbf{X}'^{\mathrm{a}} = [\mathbf{x}_1^{\mathrm{a}} - \overline{\mathbf{x}}^{\mathrm{a}}, \ldots, \mathbf{x}_m^{\mathrm{a}} - \overline{\mathbf{x}}^{\mathrm{a}}]$, then $\hat{\mathbf{P}}^{\mathrm{a}} = \frac{1}{m-1}\mathbf{X}'^{\mathrm{a}}\mathbf{X}'^{\mathrm{a}\mathrm{T}}$, and $\hat{\mathbf{P}}^{\mathrm{a}} \rightarrow \mathbf{P}^{\mathrm{a}}$ as $m \rightarrow \infty$, where $\mathbf{P}^{\mathrm{a}} = \mathbf{P}_{\mathrm{t}}^{\mathrm{a}}$ in (11).

¿From (25) it is apparent that the EnKF requires the up-front computation of a set of perturbed observations, one set associated with each member forecast. The perturbed observations are generated by adding random noise $\sim N(0, \mathbf{R})$ to the control observations. Let's make a matrix of column vectors from the perturbed observations:

$$\mathbf{Y}' = (\mathbf{y}'_1 - \mathbf{y}, \ldots, \mathbf{y}'_m - \mathbf{y}). \tag{A16}$$

With some simple algebra, one can show that

$$\hat{\mathbf{P}}^{\mathrm{a}} = \frac{1}{m-1}\left[\mathbf{X}'^{\mathrm{b}} + \hat{\mathbf{K}}\mathbf{Y}' - \hat{\mathbf{K}}\mathscr{H}\mathbf{X}'^{\mathrm{b}}\right]\left[\mathbf{X}'^{\mathrm{b}} + \hat{\mathbf{K}}\mathbf{Y}' - \hat{\mathbf{K}}\mathscr{H}\mathbf{X}'^{\mathrm{b}}\right]^{\mathrm{T}}. \tag{A17}$$

$\mathbf{X}'^{\mathrm{b}}$ was defined in (23) of section 4, and $\hat{\mathbf{K}} = \hat{\mathbf{P}}^{\mathrm{b}}\mathscr{H}^{\mathrm{T}}(\mathscr{H}\hat{\mathbf{P}}^{\mathrm{b}}\mathscr{H}^{\mathrm{T}} + \mathbf{R})^{-1}$. Let's assume that the observations and the background are uncorrelated, i.e., $\left\langle \mathbf{X}'^{\mathrm{b}}\mathbf{Y}'^{\mathrm{T}} \right\rangle = 0$ and $\left\langle \mathscr{H}\mathbf{X}'^{\mathrm{b}}\mathbf{Y}'^{\mathrm{T}} \right\rangle = 0$. Using the definition of $\hat{\mathbf{P}}^{\mathrm{b}}$ from (24), one can then show that (A17) can be expressed as

$$\begin{aligned}
\hat{\mathbf{P}}^{\mathrm{a}} &= \frac{1}{m-1}\left(\mathbf{X}'^{\mathrm{b}}\mathbf{X}'^{\mathrm{b}\mathrm{T}} - \mathbf{X}'^{\mathrm{b}}(\hat{\mathbf{K}}\mathscr{H}\mathbf{X}'^{\mathrm{b}})^{\mathrm{T}} + \hat{\mathbf{K}}\mathbf{Y}'\mathbf{Y}'^{\mathrm{T}}\hat{\mathbf{K}}^{\mathrm{T}} - \hat{\mathbf{K}}\mathscr{H}\mathbf{X}'^{\mathrm{b}}\mathbf{X}'^{\mathrm{b}\mathrm{T}} + (\hat{\mathbf{K}}\mathscr{H}\mathbf{X}'^{\mathrm{b}})(\hat{\mathbf{K}}\mathscr{H}\mathbf{X}'^{\mathrm{b}})^{\mathrm{T}}\right) \\
&= \hat{\mathbf{P}}^{\mathrm{b}} - \hat{\mathbf{P}}^{\mathrm{b}}\mathscr{H}^{\mathrm{T}}\hat{\mathbf{K}}^{\mathrm{T}} + \hat{\mathbf{K}}\mathbf{R}\hat{\mathbf{K}}^{\mathrm{T}} - \hat{\mathbf{K}}\mathscr{H}\hat{\mathbf{P}}^{\mathrm{b}} + \hat{\mathbf{K}}\mathscr{H}\hat{\mathbf{P}}^{\mathrm{b}}\mathscr{H}^{\mathrm{T}}\hat{\mathbf{K}}^{\mathrm{T}} \\
&= \hat{\mathbf{P}}^{\mathrm{b}} - \hat{\mathbf{P}}^{\mathrm{b}}\mathscr{H}^{\mathrm{T}}\hat{\mathbf{K}}^{\mathrm{T}} - \hat{\mathbf{K}}\mathscr{H}\hat{\mathbf{P}}^{\mathrm{b}} + \hat{\mathbf{K}}(\mathscr{H}\hat{\mathbf{P}}^{\mathrm{b}}\mathscr{H}^{\mathrm{T}} + \mathbf{R})\hat{\mathbf{K}}^{\mathrm{T}}.
\end{aligned} \tag{A18}$$

By inserting the definition of $\hat{\mathbf{K}}$, then $\hat{\mathbf{K}}(\mathscr{H}\hat{\mathbf{P}}^{\mathrm{b}}\mathscr{H}^{\mathrm{T}} + \mathbf{R})\hat{\mathbf{K}}^{\mathrm{T}} = \hat{\mathbf{P}}^{\mathrm{b}}\mathscr{H}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}$, so

$$\hat{\mathbf{P}}^{\mathrm{a}} = \hat{\mathbf{P}}^{\mathrm{b}} - \hat{\mathbf{K}}\mathscr{H}\hat{\mathbf{P}}^{\mathrm{b}}. \tag{A19}$$

If $\hat{\mathbf{P}}^{\mathrm{b}} = \mathbf{P}^{\mathrm{b}}$ as it should be as $m \rightarrow \infty$, and if $\hat{\mathbf{P}}^{\mathrm{b}}\mathscr{H}^{\mathrm{T}} = \hat{\mathbf{P}}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}}$ and $\mathscr{H}\hat{\mathbf{P}}^{\mathrm{b}}\mathscr{H}^{\mathrm{T}} = \mathbf{H}\hat{\mathbf{P}}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}}$ as in (28) and (29), then $\hat{\mathbf{K}} = \mathbf{K}$, where $\mathbf{K}$ is the Kalman gain of the extended Kalman filter, and (A19) is equivalent to (11).

12 Appendix 3: Serial processing of non-independent observations in the ensemble Kalman filter

As previously outlined, if observations have independent errors, then they can be processed serially, which may be of computational advantage. If the observations are not independent, the EnKF update equation (19) can be transformed so that serial processing of observations can occur.

Recall that we assume $\mathbf{y}_{\mathrm{t}} = \mathscr{H}\mathbf{x}_{\mathrm{t}} + \varepsilon$, where $\varepsilon \sim N(0, \mathbf{R})$. Presumably $\mathbf{R}$ is not diagonal here, so observations are not independent. However, $\mathbf{R}$ is symmetric and positive definite, so it has a decomposition of the form $\mathbf{R} = \mathbf{Q}_{\mathrm{R}}\Lambda_{\mathrm{R}}\mathbf{Q}_{\mathrm{R}}^{\mathrm{T}}$, where $\mathbf{Q}_{\mathrm{R}}$ is a unitary matrix with properties that $\mathbf{Q}_{\mathrm{R}}\mathbf{Q}_{\mathrm{R}}^{\mathrm{T}} = \mathbf{I}$ and $\mathbf{Q}_{\mathrm{R}}^{\mathrm{T}} = \mathbf{Q}_{\mathrm{R}}^{-1}$. Λ_{R} is a diagonal matrix of associated eigenvalues.

Let's denote a pseudo-observation $\widetilde{\mathbf{y}} = \mathbf{Q}_{\mathrm{R}}^{\mathrm{T}}\mathbf{y}$, or alternately, $\mathbf{y} = \mathbf{Q}_{\mathrm{R}}\widetilde{\mathbf{y}}$. Then $\widetilde{\mathbf{y}} = \mathbf{Q}_{\mathrm{R}}^{\mathrm{T}}\mathscr{H}\mathbf{x}_{\mathrm{t}} + \mathbf{Q}_{\mathrm{R}}^{\mathrm{T}}\varepsilon$. Hence

$$\left\langle \mathbf{Q}_{\mathrm{R}}^{\mathrm{T}}\varepsilon\,(\mathbf{Q}_{\mathrm{R}}^{\mathrm{T}}\varepsilon)^{\mathrm{T}} \right\rangle = \mathbf{Q}_{\mathrm{R}}^{\mathrm{T}}\left\langle \varepsilon\varepsilon^{\mathrm{T}} \right\rangle\mathbf{Q}_{\mathrm{R}} = \Lambda_{\mathrm{R}}. \tag{A20}$$

Define $\widetilde{\mathscr{H}} = \mathbf{Q}_{\mathrm{R}}^{\mathrm{T}}\mathscr{H}$, or equivalently $\mathscr{H} = \mathbf{Q}_{\mathrm{R}}\widetilde{\mathscr{H}}$. Substituting this definition of $\mathscr{H}$ and $\mathbf{y}$ into the EnKF update equation (25), we get

$$
\begin{aligned}
\mathbf{x}_i^a &= \mathbf{x}_i^b + \hat{\mathbf{P}}^b \mathscr{H}^T (\mathscr{H} \hat{\mathbf{P}}^b \mathscr{H}^T + \mathbf{R})^{-1} (\mathbf{y}_i - \mathscr{H} \mathbf{x}_i^b) \\
&= \mathbf{x}_i^b + \hat{\mathbf{P}}^b (\mathbf{Q}_R \widetilde{\mathscr{H}})^T (\mathbf{Q}_R \widetilde{\mathscr{H}} \mathbf{P}^b \widetilde{\mathscr{H}}^T \mathbf{Q}_R^T + \mathbf{Q}_R \Lambda_R \mathbf{Q}_R^T)^{-1} (\mathbf{Q}_R \tilde{\mathbf{y}}_i - \mathbf{Q}_R \widetilde{\mathscr{H}} \mathbf{x}_i^b) \\
&= \mathbf{x}_i^b + \hat{\mathbf{P}}^b \widetilde{\mathscr{H}}^T \mathbf{Q}_R^T \mathbf{Q}_R (\widetilde{\mathscr{H}} \hat{\mathbf{P}}^b \widetilde{\mathscr{H}}^T + \Lambda_R)^{-1} \mathbf{Q}_R^T \mathbf{Q}_R (\tilde{\mathbf{y}}_i - \widetilde{\mathscr{H}} \mathbf{x}_i^b) \\
&= \mathbf{x}_i^b + \hat{\mathbf{P}}^b \widetilde{\mathscr{H}}^T (\widetilde{\mathscr{H}} \hat{\mathbf{P}}^b \widetilde{\mathscr{H}}^T + \Lambda_R)^{-1} (\tilde{\mathbf{y}} - \widetilde{\mathscr{H}} \mathbf{x}_i^b). \qquad \text{(A21)}
\end{aligned}
$$

Thus, given a batch of observations with correlated errors and known observation-error covariance matrix $\mathbf{R}$ for these observations, one determines the eigenvectors $\mathbf{Q}_R$ and eigenvalues Λ_R of $\mathbf{R}$, forms the transformed perturbed observations $\tilde{\mathbf{y}}$ and operator $\widetilde{\mathscr{H}}$ and then solves the last line of (A21) can be used to serially process observations.

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.

Anderson, J. L., 2001: An ensemble adjustment filter for data assimilation. *Mon. Wea. Rev.,* **129**, 2884-2903.

Bellman, R. E., 1961: *Adaptive Control Processes*. Princeton University Press.

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

Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.,* **126**, 1719-1724.

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

Courtier, P., J.-N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart. J. Roy. Meteor. Soc.,* **120**, 1367-1387.

Daley, R., 1991: *Atmospheric Data Analysis*. Cambridge University Press. 457 pp.

Dee, D. P., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.,* **123**, 1128-1145.

Dee, D. P., and R. Todling, 2000: Data assimilation in the presence of forecast bias: the GEOS moisture analysis. *Mon. Wea. Rev.*, **128**, 3268-3282.

Ehrendorfer, M., and R. M. Errico, 1995: Mesoscale predictability and the spectrum of optimal perturbations. *J. Atmos. Sci.,* **52**, 3475-3500.

Evans, R. E., M. S. J. Harrison, and R. J. Graham, 2000: Joint medium-range ensembles from the Met. Office and ECMWF systems. *Mon. Wea. Rev.,* **128**, 3104-3127.

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.

Farrell, B. F., and P. J. Ioannou, 2001: State estimation using a reduced- order Kalman filter. *J. Atmos. Sci.*, **58**, 3666-3680.

Fisher, M., 1998: Development of a simplified Kalman filter *ECMWF Research Department Techical Memo-*

randum 260. European Centre for Medium-Range Weather Forecasts. 16 pp. Available from Library, ECMWF, Shinfield Park, Reading, Berkshire, RG2 9AX, England.

Gaspari, G. and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723-757.

Gelb, A. (ed.), 1974: *Applied optimal estimation.* MIT Press, 374 pp.

Golub, G. H., and C. F. Van Loan, 1989: *Matrix Computations (second edition).* Johns Hopkins University Press, 642 pp.

Gordon, N. J., D. J. Salmond, and A. F. M. Smith, 1993: Novel approach to nonlinear/non-Gaussian Bayesian state estimation. it IEEE Proceedings - F, **140**, 107-113.

Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background-error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, **129**, 2776-2790.

———— , and C. Snyder, 2000. A hybrid ensemble Kalman filter / 3d-variational analysis scheme. *Mon. Wea. Rev.*, **128**, 2905-1919.

———— , and ———— , 2002: Using improved background-error covariances from an ensemble Kalman filter for adaptive observations. *Mon. Wea. Rev.*, **130**, 1552-1572.

Harrison, M. S. J., T. N. Palmer, D. S. Richardson, and R. Buizza, 1999: Analysis and model dependencies in medium-range ensembles: two transplant case studies. *Quart. J. Roy. Meteor. Soc.*, **126**, 2487-2515.

Hastie, T., R. Tibshirani, and J. Friedman, 2001: *The Elements of Statistical Learning*. Springer, 533 pp.

Held, I. M., and M. J. Suarez, 1994: A proposal for the intercomparison of the dynamical cores of atmospheric general circulation models. *Bull. Amer. Meteor. Soc.*, **75**, 1825-1830.

Heemink, A. W., M. Verlaan, and A. J. Segers, 2001: Variance-reduced ensemble Kalman filtering. *Mon. Wea. Rev.*, **129**, 1718-1728.

Hou, D., E. Kalnay, and K. K. Droegemeier, 2001: Objective verification of the SAMEX-98 ensemble forecasts. *Mon. Wea. Rev.*, **129**, 73-91.

Houtekamer, P. L., J. Derome, H. Ritchie, and H. L. Mitchell, 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, **126**, 1225-1242.

———— , 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.

Jazwinski, A. H., 1970: Stochastic processes and filtering theory. Academic Press, 376 pp.

Keppenne, C. L., 2000: Data assimilation into a primitive equation model with a parallel ensemble Kalman filter. *Mon. Wea. Rev.*, **128**, 1971-1981.

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

Lermusiaux, P. F. J., and A. R. Robinson, 1999: Data assimilation via error subspace statistical estimation. Part 1: theory and schemes. *Mon. Wea. Rev.*, **127**, 1385-1407.

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.

Maybeck, P. S., 1979: *Stochastic models, estimation, and control*. Academic Press, volume 1, chapter 7, 368-409.

Miller, R. N., E. F. Carter, and S. T. Blue, 1999: Data assimilation into nonlinear stochastic models. *Tellus*, , **51A**, 167-194.

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

——— , and ——— , and G. Pellerin, 2002: Ensemble size, balance, and model-error representation in an ensemble Kalman filter. *Mon. Wea. Rev.*, in press.

Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's Spectral Statistical Interpolation Analysis System. *Mon. Wea. Rev.*, **120**, 1747-1763.

Penland, C., 2002: A stochastic approach to nonlinear dynamics: a review. *Bull. Amer. Meteor. Soc.*, , accepted.

Pham, D. T., 2001 : Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Mon. Wea. Rev.*, **129**, 1194-1207.

Potter, J., 1964: W matrix augmentation. *M. I. T. Instrumentation Laboratory Memo SGA 5-64*, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Rabier, F., J.-N. Thepaut, and P. Courtier, 1998: Extended assimilation and forecast experiments with a four-dimensional variational assimilation system. *Quart. J. Roy. Meteor. Soc.*, **124**, 1-39.

——— , H. Järvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. I: experimental results with simplified physics. *Quart. J. Roy. Meteor. Soc.*, **126**, 1143-1170.

Reichle, R. H., D. B. McLaughlin, and D. Entekhabi, 2002: Hydrologic data assimilation with the ensemble Kalman filter. *Mon. Wea. Rev.*, **130**, 103-114.

Richardson, D. S., 2000: Ensembles using multiple models and analyses. *Quart. J. Roy. Meteor. Soc.*, **127**, 1847-1864.

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

Tippett, M. K., S. E. Cohn, R. Todling, and D. Marchesin, 2000. Low-dimensional representation of error covariance. *Tellus*, **52 (A)**, 533-553.

——— , J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2002: Ensemble square root filters. *Mon. Wea. Rev.*, accepted.

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

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed obsevations. *Mon. Wea. Rev.*, **130**, 1913-1924.

Ziehmann, C, 2000: Comparison of a single-model EPS with a multi-model ensemble consisting of a few operational models. *Tellus*, **52a**, 280-299.

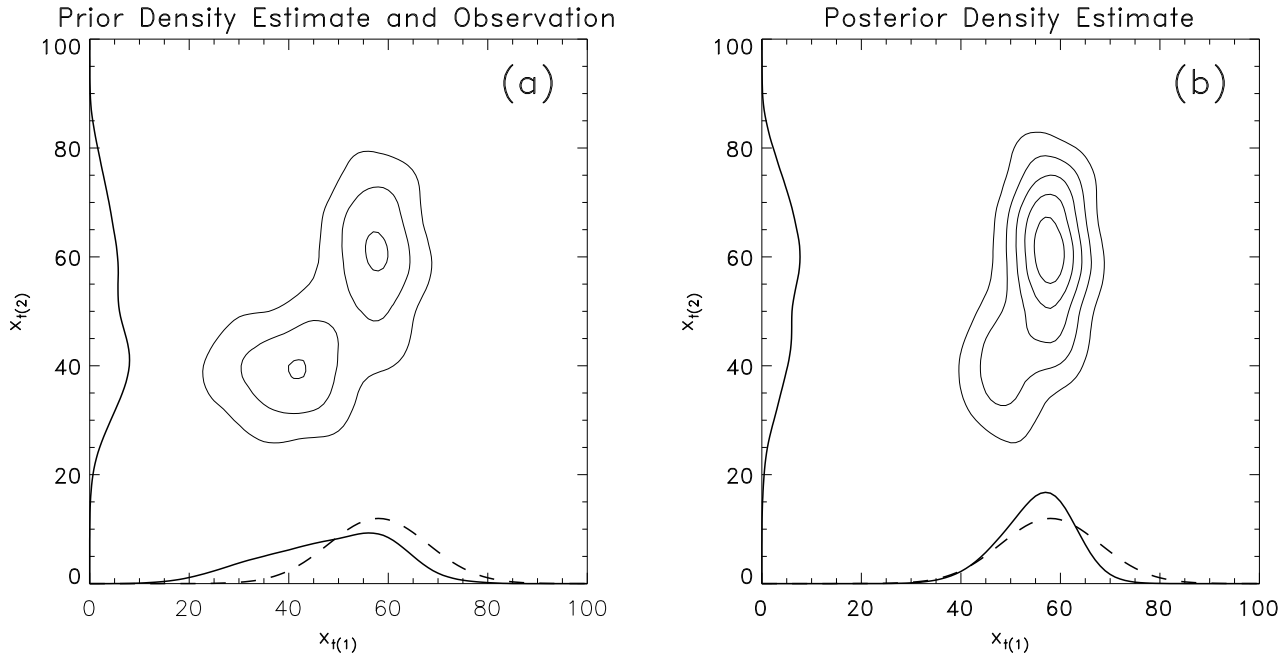

Figure 1. Example of Bayesian data assimilation. Here the model state is two dimensional and a single observation is assimilated. This observation measures the same variable as the first component of the model state. (a) Probability density for prior distributions (solid) and sample observation distribution (dashed). (b) Probability density for posterior distributions.

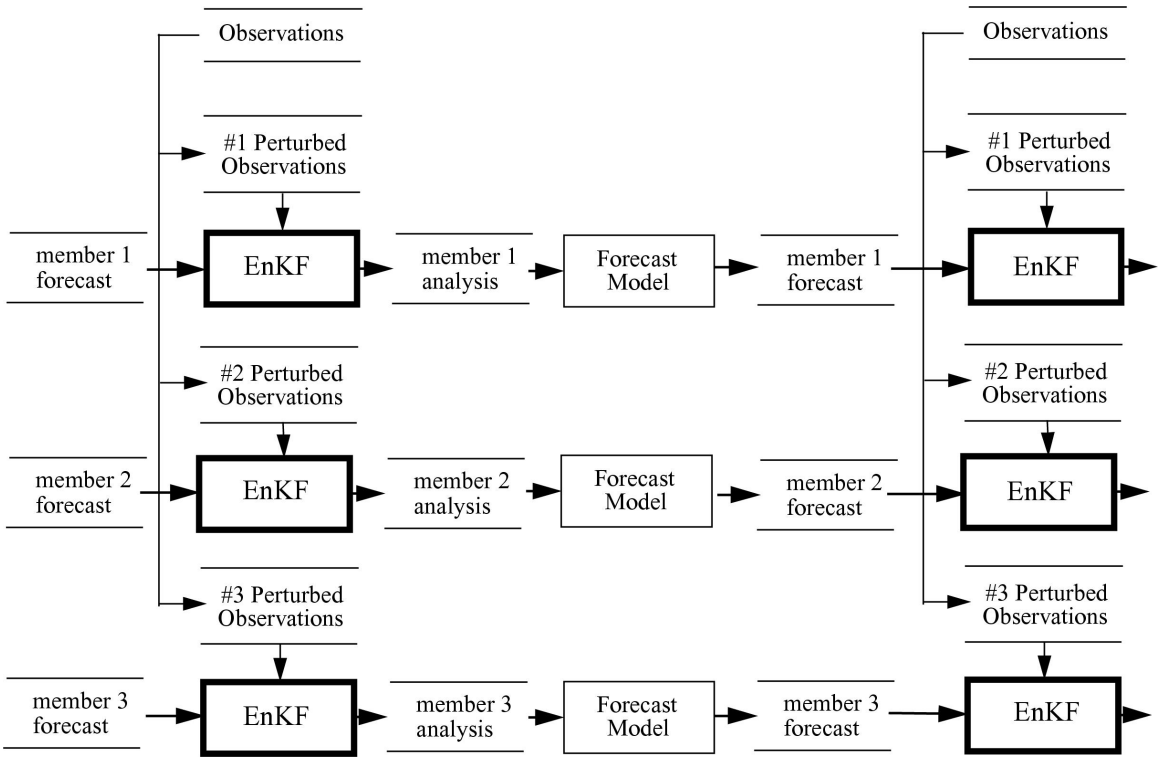

Figure 2. Schematic of the parallel data assimilation cycles in the ensemble Kalman filter, with each parallel cycle assimilating distinct perturbed observations. Note that the figure is somewhat misleading, in that the information from all ensemble members is input into the update step for each member, since the ensemble members are used to model background-error covariances.

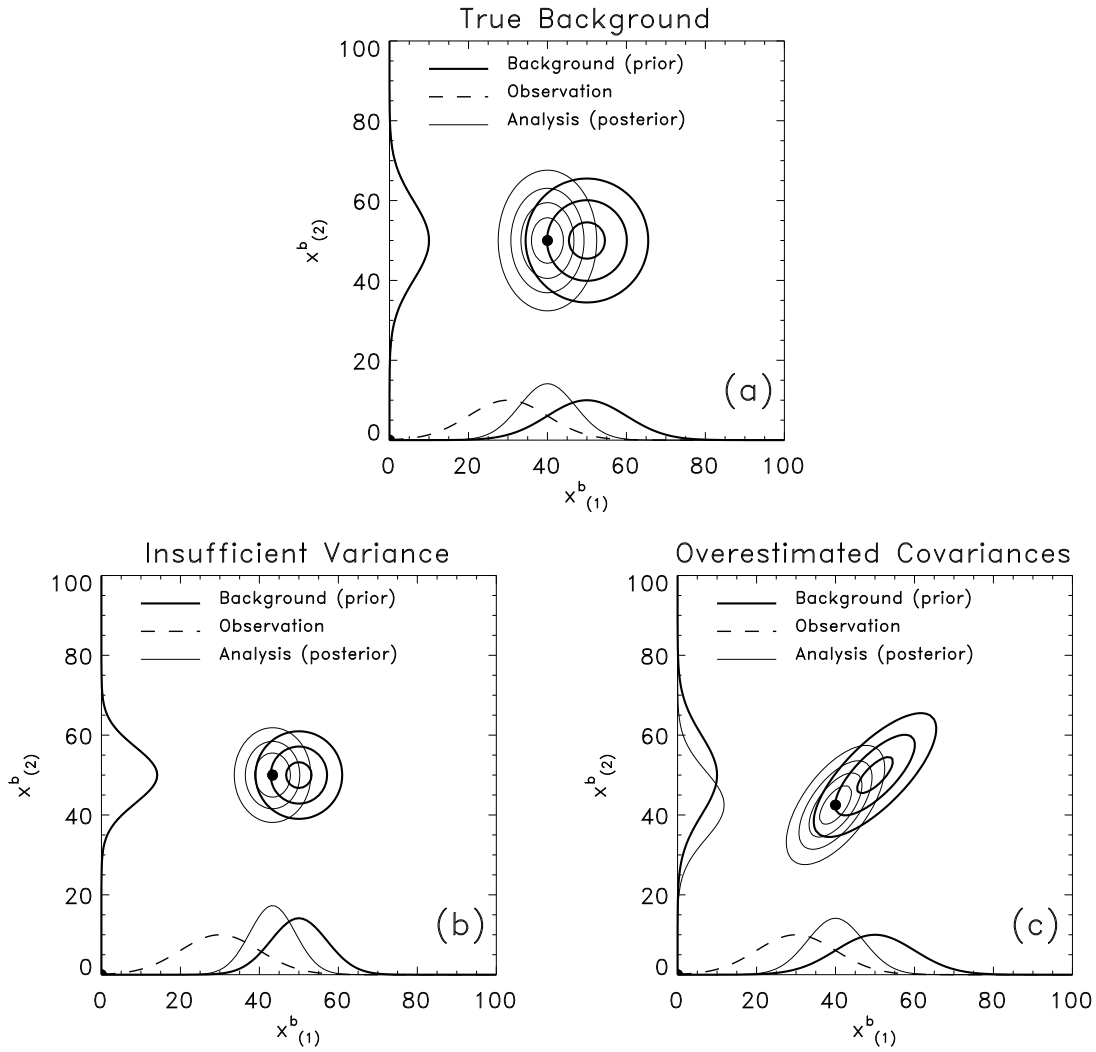

Figure 3. (a) Hypothetical data assimilation for two-dimensional state vector with an observation in only the $\mathbf{x}^b_{(1)}$ component. Heavy lines denote the true background error distribution, or prior (marginal distributions plotted along each axis). Dashed line denotes marginal distribution for observation. (b) As in (a), but assuming the the background error distribution is underestimated in magnitude. Note the posterior is shifted very little from the prior. (c) As in (a), but where correlations between the two components are overestimated, so the posterior of $\mathbf{x}^b_{(2)}$ is inappropriately shifted.

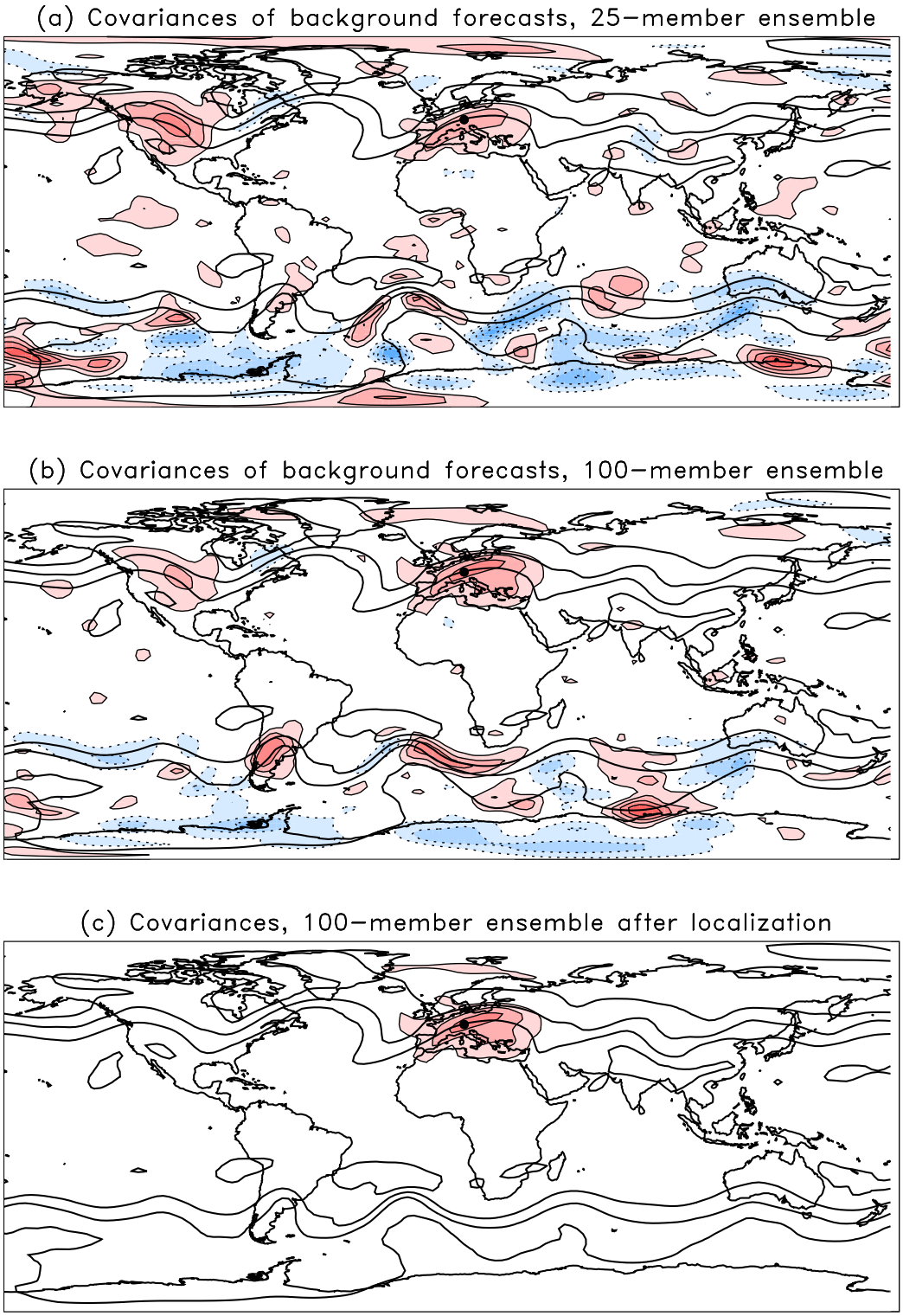

Figure 4. Covariances of ensemble of 300 hPa background temperature forecasts with background temperature forecasts at grid point marked with dot. Background forecasts are taken from an ensemble data assimilation system in a low-resolution, dry general circulation model. Solid lines denote ensemble-mean 300 hPa temperature (contours every 5 K). Colors denote covariances, with deep red colors indicating large positive covariance and blue negative. (a) Covariances directly from 25-member ensemble; (b) from 100-member ensemble; (c) from 100-member ensemble with Schur product of correlation function applied.

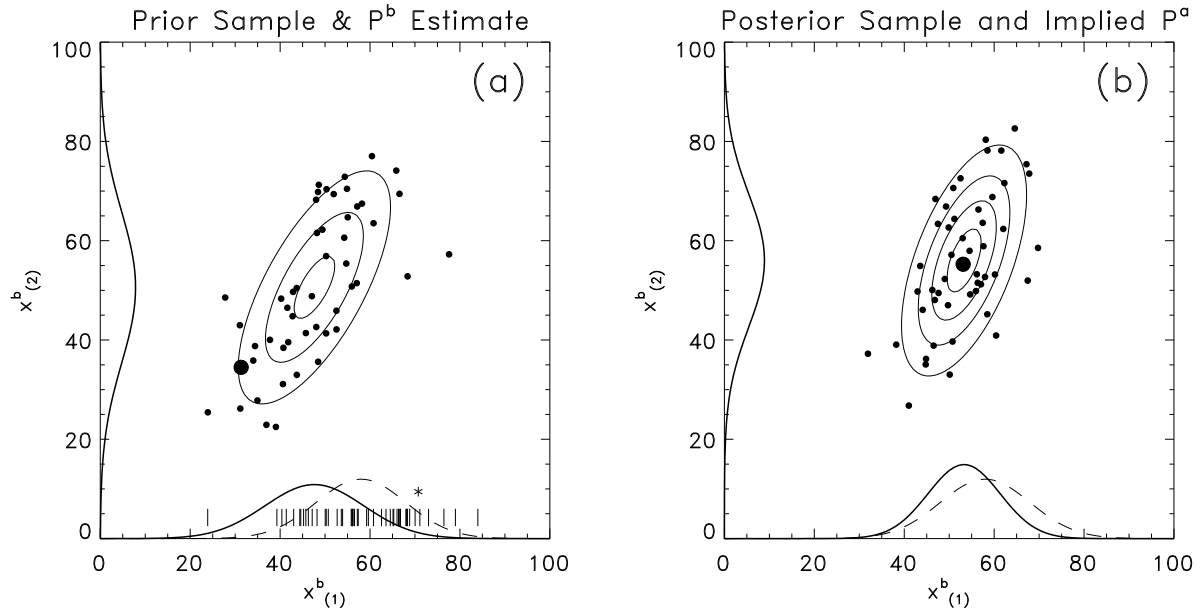

Figure 5. Illustration of the EnKF with a two-dimensional state variable and observations observing the same as $\mathbf{x}^{b}_{(1)}$. (a) Random samples (black dots) from the probability distribution in (1). Implied bivariate normal probability background-error covariance distribution contoured in black, and the observation sampling distribution (dashed). Solid vertical lines denote individual perturbed observations sampled from this distribution. The one black dot and the perturbed observation marked with a star denote the sample discussed in the text. (b) Random samples from the EnKF assimilation scheme (dots) and the implied analysis-error covariance from this sample (solid lines).

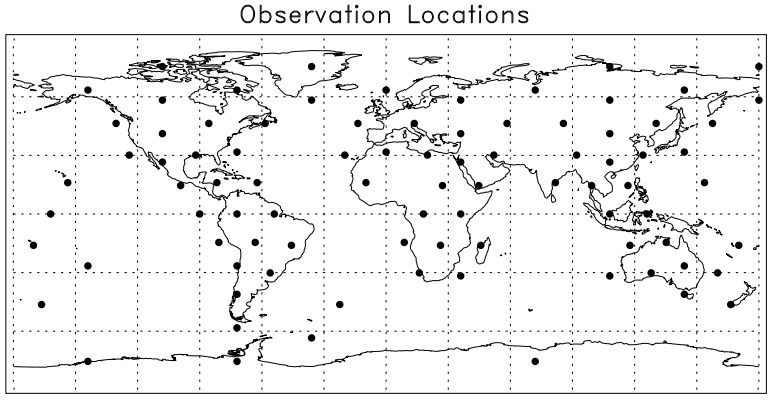

Figure 6. Observations used in the data assimilation experiment.

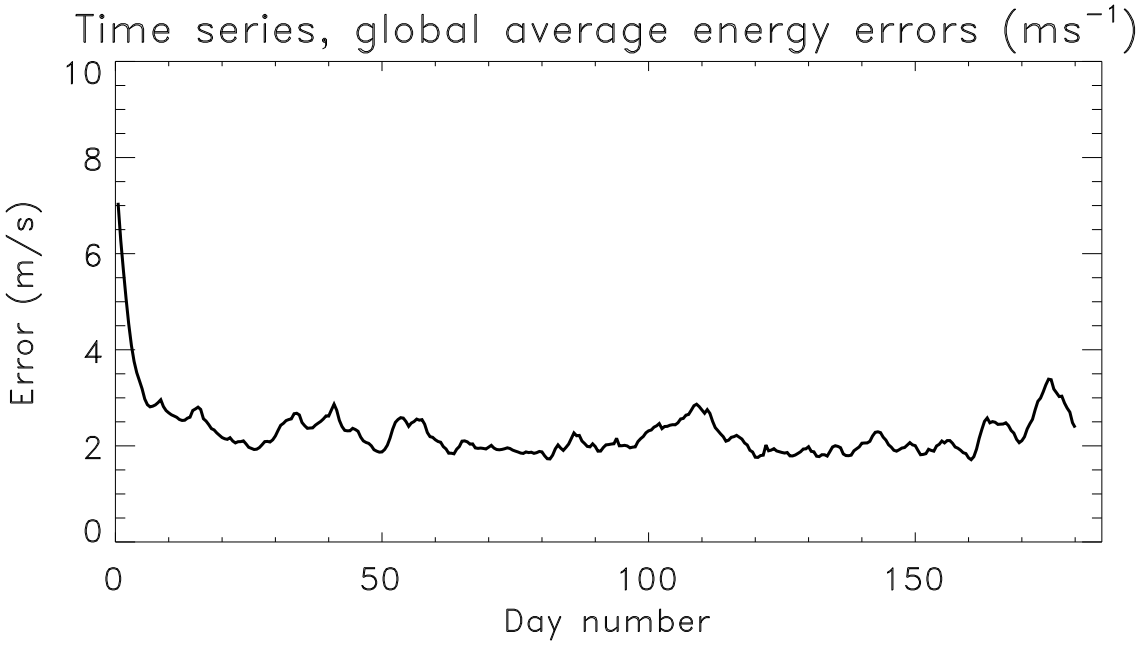

Figure 7. Time series of domain-averaged analysis errors.

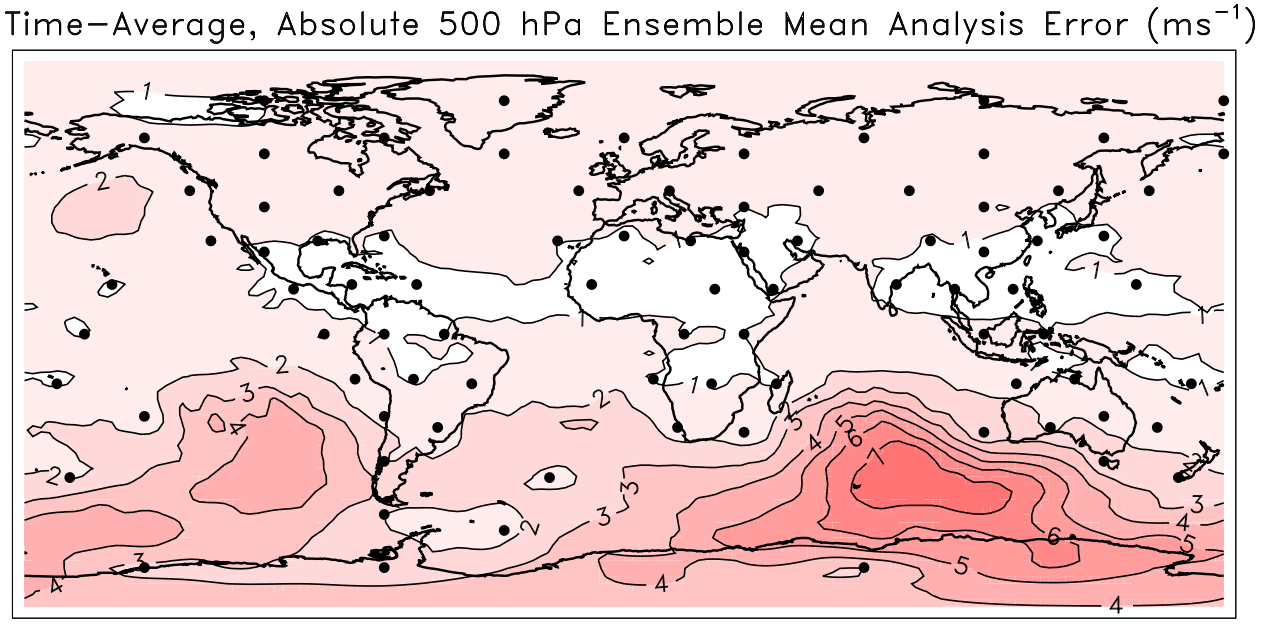

Figure 8. Horizonal map of time-averaged analysis errors in a total-energy norm at 500 hPa. Time average was computed from days 17.5 to 180 of the time series. Dots again indicate observation locations.

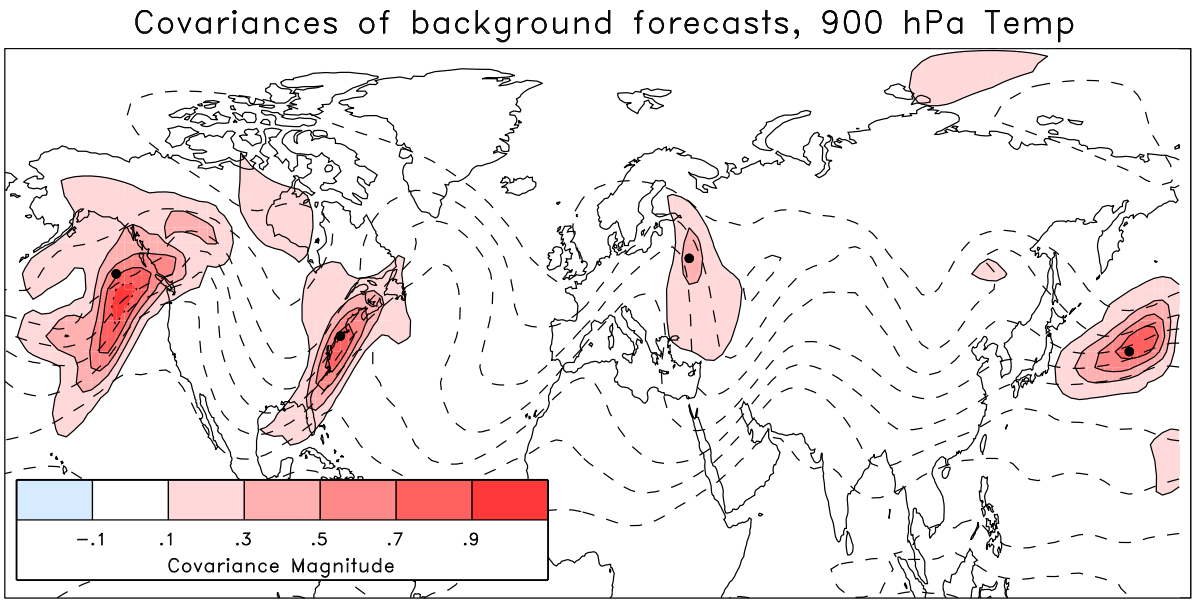

Figure 9. Examples of ensemble based covariance estimates. Covariances of 900 hPa temperatures in the vicinity of four observation locations, denoted by dots, are shown. Covariances were estimated from 100-member ensemble, with covariance localization applied. Dashed lines are contours of 900 hPa ensemble mean background temperatures; contours are plotted every 5K. Covariances are normalized so that the largest covariance in the figure is assigned a non-dimensional value of 1.0.

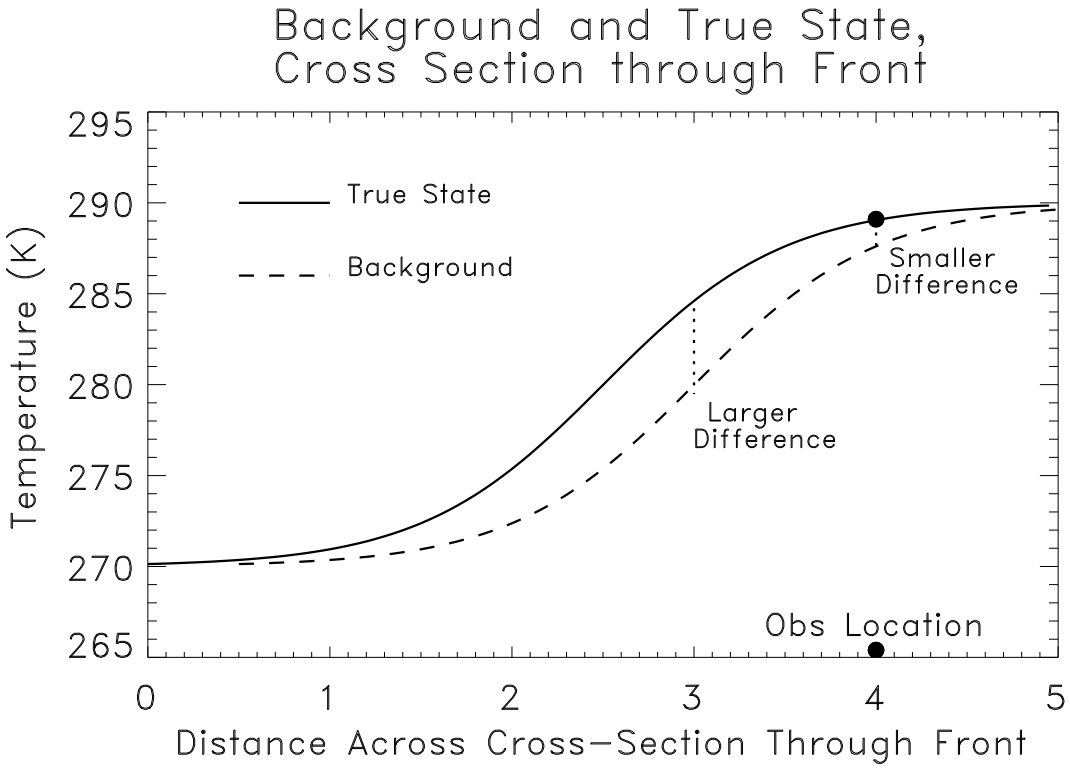

Figure 10. Example of why corrections in ensemble filter can be largest at locations somewhat distant from the observation. Consider the a mistakenly analyzed cold front (dashed line) and the true state of the front. If the only observation is available in a region outside the frontal zone, a small change in temperature at this location may reasonably imply a much greater correction is needed in the nearby frontal zone.