Tellus (2007), 59A, 210–224
Printed in Singapore. All rights reserved

TELLUS

An adaptive covariance inflation error correction algorithm for ensemble filters

By JEFFREY L. ANDERSON, *NCAR Data Assimilation Research Section, P.O. Box 3000, Boulder, CO 80307-3000, USA*

(Manuscript received 20 March 2006; in final form 14 October 2006)

ABSTRACT

Ensemble filter methods for combining model prior estimates with observations of a system to produce improved posterior estimates of the system state are now being applied to a wide range of problems both in and out of the geophysics community. Basic implementations of ensemble filters are simple to develop even without any data assimilation expertise. However, obtaining good performance using small ensembles and/or models with significant amounts of error can be more challenging. A number of adjunct algorithms have been developed to ameliorate errors in ensemble filters. The most common are covariance inflation and localization which have been used in many applications of ensemble filters. Inflation algorithms modify the prior ensemble estimates of the state variance to reduce filter error and avoid filter divergence. These adjunct algorithms can require considerable tuning for good performance, which can entail significant expense. A hierarchical Bayesian approach is used to develop an adaptive covariance inflation algorithm for use with ensemble filters. This adaptive error correction algorithm uses the same observations that are used to adjust the ensemble filter estimate of the state to estimate appropriate values of covariance inflation. Results are shown for several low-order model examples and the algorithm produces results that are comparable with the best tuned inflation values, even for small ensembles in the presence of very large model error.

1. Introduction

Ensemble filter methods for data assimilation are Monte Carlo algorithms designed to merge prior information from model integrations with observations to produce improved posterior estimates. These methods emerged in ocean (Evensen, 1994) and atmospheric science (Houtekamer and Mitchell, 1998) applications, but have been spreading rapidly to a wide range of fields in geophysics. An application of particular interest is global numerical weather prediction where enormous effort has been expended building high-quality variational assimilation methods (Derber et al., 1991; Courtier et al., 1994; Rabier et al., 2000).

Ensemble methods are appealing because simple implementations require little effort or expert knowledge. Variational methods (Le Dimet and Talagrand, 1986; Talagrand and Courtier, 1987; Derber, 1989), on the other hand, continue to require significant expert knowledge for the development of tangent linear and adjoint versions of models and forward observation operators (Geiring et al., 2005). However, experience has suggested that it is not trivial to match variational assimilation performance with ensemble methods (Houtekamer et al., 2004). Ensemble filters are subject to errors originating from a variety of sources (Lorenc, 2003). Model (Hansen, 2002) and observation representativeness (Hamill and Whitaker, 2004) error are shared with variational techniques (Li and Navon, 2001). Others, like sampling error resulting from small ensembles, are unique to ensemble methods (Anderson, 2007).

Good filter performance, even in low-order perfect model applications, can require tuning several adjunct algorithms designed to ameliorate the impacts of errors. The two most common are covariance inflation (Anderson and Anderson, 1999) and localization (Hamill et al., 2001; Houtekamer and Mitchell, 2001). Most error sources in ensemble filters are expected to lead to underestimates of the ensemble variance (Furrer and Bengtsson, 2006). Insufficient variance can lead to poor performance and, in severe cases, to filter divergence where the filter no longer responds to the observations. Covariance inflation algorithms address this issue by ‘inflating’ the prior ensemble, increasing its variance by pushing ensemble members away from the ensemble mean.

Localization algorithms try to correct for errors in the sample covariance between observations and model state variables (Mitchell and Houtekamer, 2000). These errors arise both from small samples and because most ensemble filters described in the literature assume a least squares fit to represent the relation between an observation and model state variables. These errors are

Correspondence
e-mail: jla@ucar.edu
DOI: 10.1111/j.1600-0870.2006.00216.x

particularly insidious because they lead to systematic underestimates of the posterior ensemble variance. The error is expected to be largest for an observation that is weakly correlated with a given state variable. This can be ameliorated by reducing the impact of an observation on a state variable as a function of the a priori expected correlation between the two. Often, this expected correlation is unknown and ad hoc localization is used. For instance, in numerical weather prediction applications the impact of observations on state variables is normally reduced as a function of the physical distance between the two.

Determining good localization functions in concert with appropriate covariance inflation can make tuning ensemble filters an expensive proposition. Most applications to date have assumed that both localization and inflation are homogeneous in both space and time. In this paper, a hierarchical Bayesian approach leads to an adaptive algorithm that can compute appropriate values for covariance inflation automatically. The same observations that are used in the ensemble filter are used to adjust an estimate of the covariance inflation factor.

The method developed here assumes that inflation values are homogeneous in space but adaptive in time. This serves as a foundation for future work in which inflation can vary adaptively in both space and time. Results demonstrate that the adaptive inflation algorithm can automatically produce results that are nearly as good as the best manually tuned cases. The algorithm can adapt to cases where the ensemble filter errors are very large, for instance when model error is enormous.

Section 2 presents a derivation of the Bayesian filter and Section 3 reviews ensemble filter methods for this problem. Section 4 introduces covariance inflation and Section 5 derives the adaptive inflation algorithm. Section 6 discusses avoiding filter divergence while low-order model results are presented in Section 7.

2. Bayesian filtering

A framework for combining a prior estimate of the state of a system with observations to produce a conditional estimate of the state given the observations can be derived from Bayes theorem (Jazwinksi, 1970). Let the state of the system be a random M-vector, x. Sets of scalar observations, y_k are available at discrete times, t_k, with $t_j > t_i$ if $j > i$. The set of all observations taken up to and including a time t_k is $Y_\tau = \{y_k, k \leq \tau\}$. Assume that at time t_{a-1}, an estimate of the state of the system given Y_{a-1} is $p(x, t_{a-1} \mid Y_{a-1})$, referred to as the posterior, or updated, estimate at t_{a-1}. The next time at which observations are available is t_a, the analysis time in the terminology of atmospheric assimilation. To simplify the problem, it is assumed that the error distributions for observations in different sets are unrelated. Then, the data assimilation problem can be described as the sequential solution to two subproblems. First, given $p(x, t_{a-1} \mid Y_{a-1})$, find the prior estimate of x at t_a given only observations available at times before t_a, $p(x, t_a \mid Y_{a-1})$. Second, given the set of Q observations available at time t_a, $y_a = \{y^1, y^2, \ldots, y^Q\}$, find the posterior estimate at time t_a given Y_a which is $p(x, t_a \mid Y_a)$. The algorithm can be applied iteratively to obtain the prior and posterior estimates at all subsequent observing times.

The first problem, advancing the state in time, requires a model, here a possibly stochastic function, F, that computes a state vector, x_a at time t_a given a state vector x_{a-1} at time t_{a-1},

$$x_a = F(x_{a-1}, t_{a-1}, t_a) = f(x_{a-1}, t_{a-1}, t_a) + g(x_{a-1}, t_{a-1}, t_a), \tag{2.1}$$

where f is a deterministic function and g is a stochastic function. For assimilation, x is a random vector and (2.1) must be applied in an appropriate fashion so that $p(x, t_a \mid Y_{a-1})$ is computed from $p(x, t_{a-1} \mid Y_{a-1})$. This can be done via the Fokker–Planck equation. It can also be done by Monte Carlo methods or in an ad hoc fashion; these two possibilities are addressed below.

Several additional simplifying assumptions are made in solving the second problem. Assume that observations in y_a are related to the state vector by the scalar forward observation operators, h_i, so that

$$y^i = h_i(x) + \nu_i, \quad i = 1, \ldots, Q \tag{2.2}$$

with observational error distribution

$$\nu_i \approx \text{Normal}\left(0, \sigma^2_{o,i}\right) \tag{2.3}$$

and errors for different observations are unrelated (subscript o indicates an 'observational' error variance). Then, h_i produces the expected value of y^i given a state vector. Gaussian kernel methods allow the methods here to be extended to non-Gaussian distributions if desired but with a large computational overhead (Anderson and Anderson, 1999).

The impact of the observations on the prior estimate conditioned on all previous observations is obtained from Bayes rule as

$$\begin{aligned} p(x, t_a \mid Y_a) &= p(x, t_a \mid Y_{a-1}, y_a) \\ &= p(y_a \mid x, Y_{a-1}) p(x, t_a \mid Y_{a-1}) / p(y_a \mid Y_{a-1}). \end{aligned} \tag{2.4}$$

Assuming that observation errors for observations in set y_a are independent of any other set implies that

$$p(y_a \mid x, Y_{a-1}) = p(y_a \mid x). \tag{2.5}$$

This is referred to as the observation likelihood. The denominator of the right-hand side of (2.4) is a normalization and can be written as an integration of the numerator over the range of possible values of the state vector

$$p(y_a \mid Y_{a-1}) = \int p(y_a \mid x) p(x, t_a \mid Y_{a-1}) dx \equiv \text{norm}. \tag{2.6}$$

The result is that (2.4) becomes

$$p(x, t_a \mid Y_a) = p(y_a \mid x) p(x, t_a \mid Y_{a-1}) / \text{norm}. \tag{2.7}$$

The algorithm can be further simplified by assuming that observations within the set y_a also have independent observation errors. The posterior estimate at t_a can be written as

$$p(x, t_a \mid Y_a) = p(x, t_a \mid Y_{a-1}, y_a) \\ = p(x, t_a \mid Y_{a-1}, y^1, y^2, \ldots, y^Q). \quad (2.8)$$

Defining

$$p(x, t_{a,i}) \equiv p(x, t_a \mid Y_{a-1}, \{y^k, k \leq i\}) \quad (2.9)$$

leads to

$$p(x, t_{a,i}) = p(y^i \mid x) p(x, t_{a,i-1})/\text{norm} \quad (2.10)$$

and

$$p(x, t_a \mid Y_a) = p(x, t_{a,Q}). \quad (2.11)$$

Sequential application of 2.10 to each observation in y_a in turn leads to a solution of 2.7 (Anderson, 2003).

3. Ensemble filters

Ensemble filters use Monte Carlo methods to approximate the solution to (2.7, 2.10) and to advance estimates of the state probability distribution function in time using (2.1). Assume that an N-member sample, referred to as an ensemble, of the posterior distribution, $p(x, t_{a-1} \mid Y_{a-1})$ is available. Prior estimates of $p(x, t_a \mid Y_{a-1})$ are computed by applying (2.1) independently to each posterior ensemble state. If F has a stochastic component, g, an independent sample of the stochastic component should be used when advancing each ensemble member.

Ensemble methods for computing (2.7, 2.10) can be derived starting with the Kalman filter (Kalman, 1960) and applying a Monte Carlo approximation (Burgers et al., 1998). Methods can also be derived directly from (2.7, 2.10) as in Anderson (2003) resulting in identical algorithms in many cases. Here, the sequential Ensemble Adjustment Filter derived in (Anderson, 2001) is described briefly to highlight error sources that require correction. This filter is a member of the class of ensemble square root filters (Whitaker and Hamill, 2002; Tippet et al., 2003). Other ensemble filter algorithms in the literature (Pham, 2001; Keppenne and Rienecker, 2002; Ott et al., 2004) may have some different properties but are still subject to these errors.

To further simplify notation, subscripts indexing the time and the observation number (in the set) are dropped from (2.10). Assimilation of a single scalar observation, y, suffices to describe the whole algorithm.

$$p(x \mid Y, y) = p(y \mid x) p(x, \mid Y)/\text{norm}. \quad (3.1)$$

A prior ensemble estimate of y is created by applying the forward operator h (subscript dropped since a single observation is being discussed) to each sample of the prior state. An updated ensemble estimate of y conditioned on the observation can be computed from the prior ensemble estimate of y, the observed value y^o and the observation's error variance, σ_0^2 using (3.1). The result can be used to compute an increment for each of the N prior ensemble estimates of y. Details of the computation of the increments distinguish most ensemble filter variants in the atmospheric and oceanic literature (Anderson, 2003).

In the Ensemble Adjustment Filter, the prior ensemble estimate of y is approximated as Normal($\bar{y}_p, \sigma_p^2$) where $\bar{y}_p$ and σ_p^2 are the sample mean and variance. The product of Normal($\bar{y}_p, \sigma_p^2$) and Normal (y^o, σ_o^2) in (3.1) is computed resulting in a Gaussian updated distribution for y, Normal($\bar{y}_u, \sigma_u^2$) with

$$\sigma_u^2 = \left[\left(\sigma_p^2\right)^{-1} + \left(\sigma_o^2\right)^{-1} \right]^{-1} \quad (3.2)$$

and

$$\bar{y}_u = \sigma_u^2 \left[\left(\sigma_p^2\right)^{-1} \bar{y}_p + \left(\sigma_o^2\right)^{-1} y^o \right]. \quad (3.3)$$

The prior ensemble distribution of y is then shifted and linearly compacted to create an updated ensemble with sample statistics $\bar{y}_u$ and σ_u^2. Increments are

$$\Delta y_i = \sqrt{\sigma_u^2 / \sigma_p^2} (y_{p,i} - \bar{y}_p) + \bar{y}_u - y_{p,i}, i = 1, \ldots, N, \quad (3.4)$$

where i subscripts the ensemble member. Other ensemble filters calculate these increments differently, for instance via the perturbed observation approach of the ensemble Kalman filter (Evensen, 1994).

Finally, increments for each component of the prior state vector are computed independently using linear regression with the prior joint sample statistics. Increments for the jth component are computed as

$$\Delta x_{j,i} = \left(\sigma_{p,j} / \sigma_p^2\right) \Delta y_i, \quad j = 1, \ldots, M, \quad i = 1, \ldots, N, \quad (3.5)$$

where $\sigma_{p,j}$ is the prior sample covariance of the observed variable, y, and the jth element of the state vector, x, and M is the size of the model state vector.

Error is associated with each step in ensemble filter algorithms. Error in the model can be a large term in many applications (Buizza et al., 1999; Dee and Todling, 2000). When computing the forward operators, errors of representativeness (Daley, 1993; Dee et al., 1999) can be introduced if the model state does not adequately represent all the spatial/temporal scales and physical processes that take place in the system being observed. For instance, global atmospheric models do not resolve small scale convection, but radiosonde observations can be significantly impacted by the presence of a thunderstorm. Additional errors may be involved with more complicated forward observation operators like those for satellite radiances (Eyre et al., 1993).

The observations themselves are often contaminated by errors that are not adequately represented by a specified Gaussian distribution. For instance, 'gross' errors made through human error plague operational numerical weather prediction and atmospheric reanalyses (Kistler et al., 2001; Uppsala et al., 2005).

When the prior observation ensemble and the observation likelihood are combined using (3.1), sampling error from finite ensembles is a source of error. Characterizing this error can be extremely complex, even when using very simple models in perfect model settings. For large models, the nature of these sampling errors remains largely unknown. Assuming that the prior distribution and the observation likelihood are Gaussian in the observation space update step may also introduce error.

Errors can also occur during the regression that is used to update state variables given increments for the observation prior eq. (3.5). Sampling error is again an issue since ensemble statistics are used to compute the regression coefficients. Assuming a linear relation between the observation and the state variables may be inappropriate leading to additional regression errors. If the model is not perfect, the prior joint statistics may be unable to represent the appropriate relation between an observation and state variables even if linearity is valid and sampling error is negligible. Again, it is very difficult to characterize these errors in large model applications.

4. State space covariance inflation

All error sources are normally expected to produce ensemble estimates with insufficient variance which can in turn lead to increased error in the mean. Unrealistically confident priors cause observations to receive insufficient weight in the update eq. (3.1) leading to further errors in the mean and increasingly underestimated error variance. Prior estimates may become so confident that further observations are essentially ignored and the filter solution depends only on the model. This is referred to as filter divergence.

To avoid filter divergence and improve assimilation quality, ensemble filters can employ heuristic methods to increase variance estimates. One common method is 'covariance inflation' (Anderson and Anderson, 1999) in which the prior ensemble state covariance is increased by linearly inflating each scalar component of the state vector before assimilating observations:

$$x_{j,i}^{\inf} = \sqrt{\lambda}(x_{j,i} - \bar{x}_j) + \bar{x}_j,\ j = 1, ..., M; i = 1, \dots N, \qquad (4.1)$$

where j indexes the state vector component and i the ensemble member, the overbar is an ensemble mean and N is the ensemble size, M is the model size, and λ is referred to as a covariance inflation factor. The sample variance of each component x_j is increased by λ while sample correlations between any pair of components remain unchanged. Covariance inflation has been applied in many studies with both low-order models and large atmospheric prediction models. Frequently, ensemble filters diverge without covariance inflation. Good assimilations can require careful tuning of λ, an expensive proposition in large systems. Other methods of increasing variance have been explored more recently including using additive instead of multiplicative inflation (Tom Hamill 2005, personal communication).

5. Adaptive inflation

Covariance inflation can be viewed as a simple model to correct the unknown deficiencies in an ensemble filter that lead to underestimated prior variance. The inflation factor, λ, can be viewed as a one-dimensional state vector for this model of variance error, and observations in conjunction with Bayes theorem can be used to improve the estimate of λ. In an ensemble filter, the prior ensemble estimate of an observation, y, (the result of applying the forward operator to each ensemble member), the observation, y^o, and the observational error variance, σ_o^2, can be used to estimate whether λ is too big or too small. The Bayesian filtering update eq. (2.7) can be used to compute probability distributions for λ given a sequence of observations. The approach here is closely related to adaptive error variance algorithms developed for non-ensemble assimilation systems (Dee, 1995; Dee and Da Silva, 1999; Dee et al., 1999).

Using the notation of Section 2, a prior distribution for λ

$$p(\lambda, t_a \mid Y_{a-1}) \qquad (5.1)$$

is required along with a method for computing a prior distribution from a posterior,

$$p(\lambda, t_{a-1} \mid Y_{a-1}), \qquad (5.2)$$

at the previous observation time

$$\lambda_a = F_\lambda(\lambda_{a-1}, t_{a-1}, t_a) = f_\lambda(\lambda_{a-1}, t_{a-1}, t_a) + g_\lambda(\lambda_{a-1}, t_{a-1}, t_a)\,. \qquad (5.3)$$

Here, the prior distribution of λ is assumed to be normal

$$p(\lambda, t_a \mid Y_{a-1}) = \mathrm{Normal}\left(\bar{\lambda}_p, \sigma_{\lambda,p}^2\right) \qquad (5.4)$$

with mean $\bar{\lambda}_p$and variance $\sigma_{\lambda,p}^2$. Initially, it is assumed that the 'model' time tendency for λ is 0. In other words, the prior distribution at t_a is identical to the posterior distribution at t_{a-1}; $f_\lambda(\lambda, t_1, t_2) = \lambda$ and $g_\lambda(\lambda, t_1, t_2) = 0, \forall \{\lambda, t_1, t_2\}$. The only changes to the distribution for λ are caused by the impact of observations.

As in Section 2, observations can be assimilated sequentially given independence of the observational errors. In this case, an algorithm for applying (2.10) for a single scalar observation is sufficient to describe the entire Bayesian update for λ.

The state vector x in (2.10) is replaced by λ, giving

$$p(\lambda, t_{a,i}) = p(y^i \mid \lambda) p(\lambda, t_{a,i-1})/\mathrm{norm} \qquad (5.5)$$

and a similar transformation can be made to (2.11). The ensemble samples of the x state vector are parameters of the assimilation problem for λ, just as λ is a parameter of the ensemble assimilation problem for x when covariance inflation is used.

An expression for the observation likelihood, the first term in the numerator of (5.5) is required. To simplify notation, use of only a single observation is described and subscripts referencing the observation number/time are dropped. Suppose the prior observation ensemble is given by

$$y_{p,k} = h(x_k),\ k = 1, \dots, N \qquad (5.6)$$

with sample mean and variance $\bar{y}_p$and σ_p^2. The observation is y^o with error variance σ_o^2. The observation likelihood expresses the probability that y^o would be observed given a value of the inflation, λ.

Let θ be the expected value of the distance between $\bar{y}_p$ and y^o, given a value of λ. If the prior observation distribution is unbiased as assumed in (2.3),

$$\theta = \sqrt{\lambda\sigma_p^2 + \sigma_o^2}. \quad (5.7)$$

This implicitly assumes that inflating all the state variable components in x using λ leads to an inflation of the observation prior variance, σ_p^2, by the same factor. This is equivalent to the assumption of gaussianity that is made when the linear regression is used to update state variables given observation variable increments (3.5) in the ensemble filter. This assumption is violated if, for instance, the relation between the observed variable prior and the state variables is not linear over the range of the prior ensemble samples.

The actual distance between the prior ensemble mean and the observed value is $D = |\bar{y}_p - y^o|$. The assumption that the prior sample and observation are unbiased implies that D is drawn from a distribution that is Normal (0, θ^2). By the definition of the normal

$$p(y^o \mid \lambda) = (\sqrt{2\pi}\theta)^{-1} \exp(-D^2/2\theta^2) \quad (5.8)$$

is the observation likelihood term. In this case, (5.5) becomes

$$p(\lambda, t_{a,i}) = (\sqrt{2\pi}\theta)^{-1} \exp(-D^2/2\theta^2)\, N\big(\bar{\lambda}_p, \sigma_{\lambda,p}^2\big)/\text{norm}. \quad (5.9)$$

Although both the observation likelihood (5.8) and the prior (5.4) are Gaussian, the former has a variance that is a function of λ while the latter has a mean that is a function of λ so the product is not Gaussian (unlike 3.1 for the ensemble filter). Nevertheless, the algorithm here *assumes* a Gaussian for the posterior since it is equivalent to the next prior which is assumed Gaussian.

Figure 1a shows an example of an observation likelihood (5.8) and a prior probability distribution for λ (prior divided by 300 and shifted upwards by 0.195 so that it can be viewed on the same axes). This case is from an experiment described in more detail in Section 7. The likelihood is skewed toward larger values of λ and is extremely flat compared to the prior. If not divided by 300 the prior would be a high, narrow spike compared to the likelihood, so variations in the posterior are dominated by the prior. Figure 1b shows the posterior minus the prior. The difference is very small, consistent with Fig. 1a, with the mean of the posterior shifted towards the mode of the observation likelihood as expected. In this case, the distance D between the prior ensemble mean and the observed value was larger than expected so the distribution of λ is shifted slightly towards larger values. The variance of the posterior is also slightly reduced although this is not clearly visible in the figure. The fact that the observation likelihood is so flat compared to the prior indicates that there is very little information about λ available from a single observation.

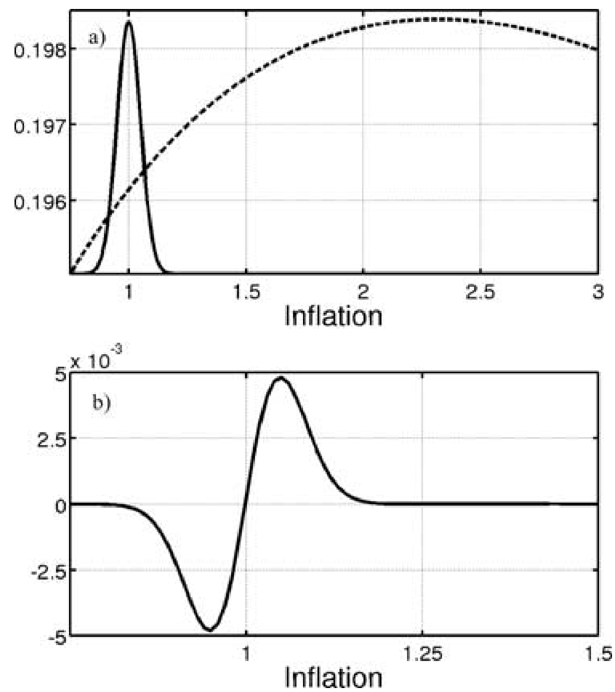

Fig. 1. The prior distribution function divided by 300 and shifted upward by adding 0.195 (solid) and the observation likelihood (dashed) for the adaptive inflation for the first observation assimilated at the assimilation time 2001 of a 10-member ensemble assimilation of 40 randomly located observations in the Lorenz-96 model is in panel (a). Panel (b) shows the posterior minus the prior probability distribution functions for this observation.

Although the posterior distribution computed via (5.9) is not Gaussian, it is generally very nearly Gaussian as shown by the small differences in Fig. 1b. A posterior mean and variance must be found so that Normal($\bar{\lambda}_u$, $\sigma_{\lambda,u}^2$) closely approximates the result of (5.9). One could try to find the mean of the product through analytic integration, but this is difficult because λ is associated with the mean in the prior and the variance in the likelihood terms. Instead, the algorithm used here sets $\bar{\lambda}_u$ equal to the mode of the exact posterior from (5.9). The mode can be found analytically by differentiating the numerator of (5.9) and setting the result equal to 0 (Appendix A). If the resulting cubic equation has a single real root (by far the most common case in the results here), this root is the mode. If there are multiple real roots, the root closest to $\bar{\lambda}_p$ is chosen. Computing $\bar{\lambda}_u$ requires a number of multiplications, several square roots, and two cube roots.

The variance of the posterior is found by a naïve numerical technique. The ratio, *R,* of the value of the numerator of (5.9) at $\bar{\lambda}_u + \sigma_{\lambda,p}$ to the value at $\bar{\lambda}_u$ is computed by evaluating the numerator of (5.9) at both points. It is assumed that the posterior is Normal($\bar{\lambda}_u$, $\sigma_{\lambda,u}^2$) and expressions for the values of the normal evaluated at $\bar{\lambda}_u + \sigma_{\lambda,p}$ and $\bar{\lambda}_u$ can be computed using the definition of the normal density. Taking the ratio of these two expressions, one can solve for the updated variance as

$$\sigma^2_{\lambda,u} = -\sigma^2_{\lambda,p}/2\ln(R). \tag{5.10}$$

Other methods could be used to approximate the updated variance. One could compute ratios as just outlined for a large number of points and compute the mean value of the approximate variances to account for the impact of higher order moments in the posterior. Another method uses quadrature and an optimization to find the value of the variance so that the integrated area between the exact posterior and the Gaussian approximation is minimized. The approximate method used here is biased to produce slightly larger values of $\sigma^2_{\lambda,u}$ in the mean over many observations. This is a convenient property as noted in the discussion of filter divergence in the next section.

The most direct implementation of the ensemble filter algorithm computes the adaptive inflation filter independently before the ensemble filter as follows:

1. The ensemble estimate of the model state, *x,* is advanced to the time of the next observation using the model.
2. The posterior estimate of the inflation from the time of the previous observations is advanced to the time of the next observation using a model. As noted earlier, one may often not know how to advance the inflation in time and this step may simply entail using the previous posterior inflation estimate as the new prior estimate.
3. Each observation is processed sequentially to update the prior inflation estimate.

(a) The forward observation operator is computed for each ensemble member.
(b) The expected distance between the prior ensemble mean of the observation and the observed value is computed with (5.7).
(c) The cubic formula is used to solve (A.8) for x and (A.7) is used to compute the approximate posterior mean for the inflation.
(d) The estimated variance of the updated inflation distribution is computed using (5.10).

Once the inflation distribution has been updated with all observations at this time, the prior state estimate is inflated using $\bar{\lambda}_u$ in 4.1. The ensemble filter is then used to sequentially assimilate the observations available at this time.

A variety of modifications can be made to the algorithm to enhance computational efficiency. For certain parallel implementation of filters and for some types of observations, it can be very expensive to compute forward observation operators at different points in the filter algorithm. Instead, the algorithm can be modified to compute the ensembles of forward observation operators for all observations available at a given time at once before beginning the sequential assimilation algorithms. An implementation of this variant of the hierarchical filter proceeds as follows:

1. The state vector ensemble is advanced to the time of the next observation and the prior distribution for the inflation is computed.
2. The ensemble is inflated using $\bar{\lambda}_p$.
3. All forward observation operators are computed for each ensemble member.
4. The prior variance of each observation is computed and 'uninflated' to account for the fact that inflation has already been done in state space.
5. Each observation is then processed sequentially.

(a) The inflation distribution is updated using the 'uninflated' prior variance computed in 4).
(b) The ensemble filter is used to compute observation increments for this observation.
(c) All state variables are updated by regression with these increments.
(d) The ensemble distributions of all remaining observation priors are also updated by regression with these increments.

6. Avoiding inflation filter divergence

The adaptive inflation algorithm consists of two inter-related filters: an ensemble filter for the state vector *x*; a continuous filter for λ. Adaptive inflation is used to avoid filter divergence and improve performance for the state vector ensemble filter. Although the λ filter uses a continuous probability function representation, it is still subject to the same concerns about filter divergence as the ensemble filter. Every time an observation is assimilated in the λ filter, the variance estimate σ^2_λ is expected to decrease (Chui and Chen, 1987). In the discussion so far, the prior value of λ at the next time is equal to the posterior at the previous time, so there will be a systematic loss of variance that can lead to filter divergence.

Figure 2 shows $\sigma^2_{\lambda,p}$ and $\bar{\lambda}_p$ as a function of assimilation time from an experiment described in detail in Section 7. The initial $\sigma^2_{\lambda,p}$ is 0.25 and after assimilating observations at 2000 different times, this is reduced to about 0.06 (Fig. 2a). If the assimilation continued, $\sigma^2_{\lambda,p}$ would continue to decrease, and the adaptive inflation filter would give progressively less weight to observations. This behavior is also apparent in the time evolution of $\bar{\lambda}_p$ (Fig. 2b), which shows that the timescale of variability in the estimate of $\bar{\lambda}_p$is increasing as σ^2_λ decreases. Eventually, observations would have negligible impact on the estimate, and $\bar{\lambda}_p$ would become nearly constant in time. This would not be a problem if the underlying errors in the ensemble filter were constant in time. The filter would presumably converge to the correct value and stay there. In general, however, the ensemble filter assimilation errors may not be constant in time and one might prefer that the estimate of $\bar{\lambda}_p$ continue to evolve in response to additional observations. In other words, the systematic loss of variance for λ could eventually lead to filter divergence.

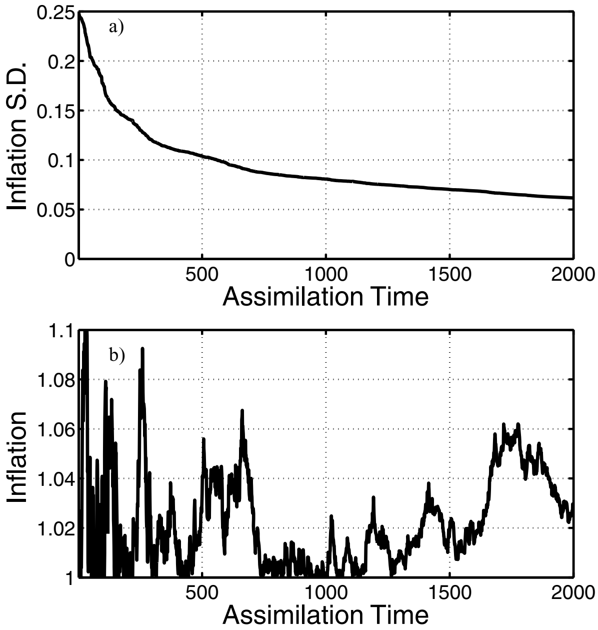

Fig. 2. Value of standard deviation (a) and mean (b) of adaptive inflation distribution during first 2000 steps of a 10-member ensemble assimilation of 40 randomly located observations in the Lorenz-96 model.

If desired, this loss of variance could be addressed in several ways. One could apply covariance inflation to the λ filter and construct yet another filter, which would itself be subject to filter divergence, to determine the values for this second inflation factor. This hierarchical Bayesian approach is somewhat reminiscent of hierarchical turbulence closures in atmospheric modeling (Mellor and Yamada, 1982); it has to be closed somewhere. One could also use some sort of variance growing model for f and g in (5.3) to compute the prior from the previous posterior for λ. For most results here, $\sigma^2_{\lambda,p}$ is fixed in time,

$$\sigma^2_{\lambda,p} = \sigma^2_c, \tag{6.1}$$

where the subscript c indicates a value that is constant in time. σ^2_c must be selected empirically and this only makes sense if it is less difficult and costly to find appropriate values than it is to find good values of λ itself. Results below indicate that good performance can be obtained for a wide range of ensemble filter errors using a single value of σ^2_c. Roughly equivalent performance is found for a wide range of choices for σ^2_c meaning that tuning is not generally required. This is not always the case for empirically tuning λitself as seen in Section 7.5.

7. Low-order model results

The adaptive inflation algorithm is applied to synthetic observation experiments using the 40-variable model of Lorenz (Lorenz, 1996; Lorenz and Emanuel, 1998; Appendix B) to see how it responds to various sources and magnitudes of ensemble filter errors. Small ensemble assimilations in this model are subject to large sampling error from the regression step. Localization is used to reduce this sampling error. The 40 state variables are defined to be equally spaced on a periodic (0, 1) domain. Observing stations are placed at locations on (0, 1) and forward operators are linear interpolation between the two nearest state variable locations. When the regression of increments from an observation located at z_o onto a state variable at z_s is performed with (3.5), the regression coefficient is multiplied byζ (d, c) where

$$d = \min\left(|z_o - z_s|\,,\ 1 - |z_o - z_s|\right),$$

is the distance between the observation and state variable and ζ is the fifth order polynomial function of Gaspari and Cohn (1999) with half-width c. For $d \geq 2c$, the observation has no impact on the state variable. For $d < 2c$, ζ approximates a Gaussian.

A set of observing stations with spatial locations randomly drawn from U(0, 1) and fixed in time is used in all experiments. Synthetic observations are taken every time step by linearly interpolating to each observation location and adding a draw from Normal (0, σ^2_o)to simulate observational error. The specified observational error variance σ^2_o is set to 1.0 in all experiments.

Initial conditions for the ensemble assimilation are 'climatological', randomly chosen states from long free integrations of the model used to generate the observations. The assimilating model (not always the same as the model used to generate observations) is used to assimilate for 4000 observation times but results are only reported for the last 2000 times. Transient behavior appears to have disappeared after the 2000 initial steps in all experiments. When the adaptive inflation algorithm is used, the initial value for $\bar{\lambda}$ is 1.0 and $\bar{\lambda}$ is constrained to be no less than 1. Allowing $\bar{\lambda}$ to become less than 1.0 has negligible impact on the results shown here.

With one exception, the deterministic square root filter (Tippett et al., 2003) update method called the ensemble adjustment filter described in Section 2 is used for the observation space scalar update. The primary metric of filter performance is the time mean (prior) ensemble mean rms innovation in observation space

$$I = \sqrt{\sum_{j=1}^{M}\left[\sum_{i=1}^{N} h_j\left(x_i\right)/N - y_{o,j}\right]^2 \Big/ M} \tag{7.1}$$

and the associated expected value of the rms innovation given the prior ensemble

$$S = \sqrt{\sum_{j=1}^{M}\sum_{k=1}^{N}\left\{\left[\sum_{i=1}^{N} h_j(x_i)/N - h_j(x_k)\right]^2 + \sigma^2_{o,j}\right\} \Big/ MN} \tag{7.2}$$

where N is the ensemble size and M is the total number of observations in the 2000 assimilation step verification period. The quantity in (7.2) is referred to as the 'innovation spread'. For an optimally performing filter, the innovation values should be consistent with the spread. These statistics are computed by applying the forward observation operator for all observations available at a given time before any of the observations from that time are assimilated. The prior rms innovation compares only to independent observations and can be evaluated in real assimilation experiments where the 'truth' is not known. Values of the ensemble mean time mean rms error from the truth for the state vector (7.3) (not available in real assimilations) are given in some cases for comparison to previous studies on the Lorenz-96 model.

7.1. *Varying ensemble size*

The first results examine the response of the adaptive inflation algorithm to varying the ensemble size while holding other parameters constant. Forty randomly located observing stations are used and the localization half-width c is 0.15. This c is too small (large) for optimal filter performance for large (small) ensembles. Ensemble sizes of $\{5,6,7,8,9,10,12,14,16,18,20,25,30,35,40,50,60,70,80,90,100\}$ were tested; Fig. 3 plots the time mean value of $\bar{\lambda}$ over 2000 assimilation steps, the maximum and minimum values of λ over the 2000 steps, and the rms innovation and innovation spread as a function of ensemble size. As the ensemble gets smaller, sampling error in the ensemble filter increases and λ increases (Fig. 3a). The range of the values of $\bar{\lambda}$ also increases as the ensemble size decreases. The rms innovation and innovation spread generally have roughly consistent values and are approximately constant for ensemble sizes between 8 and 100. Ensembles smaller than 4 were unstable while no significant change occurred for ensembles larger than 100.

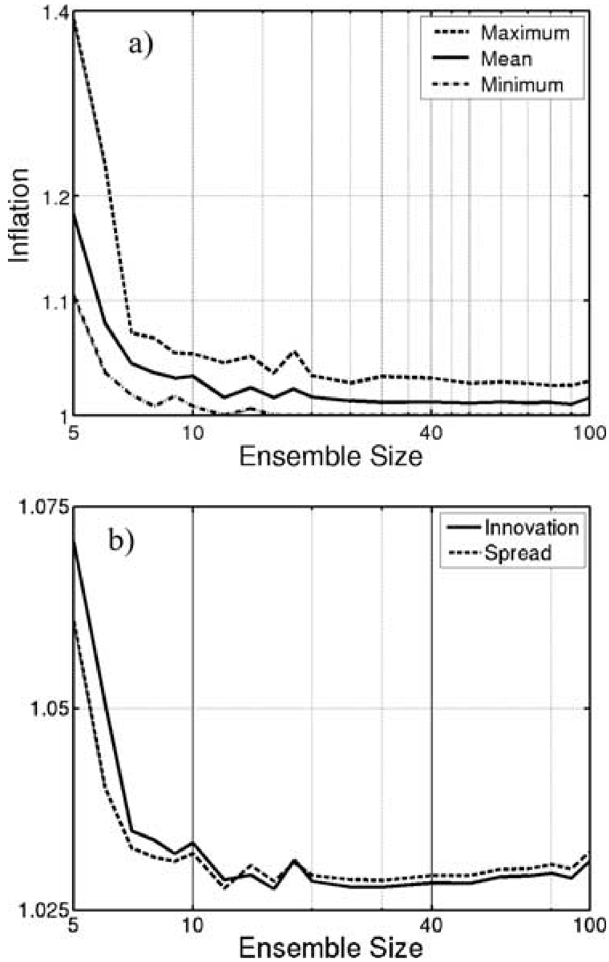

Fig. 3. Panel (a) shows the minimum, mean and maximum values of the adaptive inflation, λ, as a function of ensemble size when assimilating 40 randomly located observations in the Lorenz-96 model. Results are averaged over the last 2000 steps of a 4000 step assimilation. Panel (b) shows the time-mean rms innovation and ensemble innovation spread for the prior ensemble estimates of the observations.

7.2. *Varying localization for fixed ensemble size*

Next, the ensemble size is fixed at 10 and the localization half-width, c, is varied among $\{0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 \text{ and } 1.0\}$. The optimal c for this ensemble size and observation set is about 0.15. Larger values lead to increased sampling error in the regression (3.5) while smaller values lead to increased error in the mean because valid observational information is ignored and state space increments are not appropriately correlated for adjacent state variables. Figure 4 shows the 2000 step statistics of λ and the time mean observation space error and spread as c is varied. λ is smallest when the half-width is 0.15 and increases for larger and smaller values. The values of $\bar{\lambda}$ vary from approximately 1.05 to 1.10 over the range in which rms innovation is roughly constant.

7.3. *Model error*

The Lorenz-96 model is a forced dissipative model with a parameter F that controls the strength of the forcing (B.1). Integer values of F greater than 3 produce chaotic time-series, values from 1 to 3 produce periodic time-series, and $F = 0$ produces a damped time-series. The time evolution of model state variable x_1 for $F = 0$, 3, 5, 8 and 11 is shown in Fig. 5, all cases starting from identical initial conditions at $t = 0$. The model behavior is quite different for these different values of F.

To simulate model error, observations at the 40 randomly located stations are produced by a model with $F = 8$. These observations are then assimilated by 20 member ensemble filters that use models with $F = 4, 5, \ldots, 12$. Cases with $F < 3$ and > 13 led to models that became numerically unstable. Figure 6 plots the minimum, maximum and mean of λ over 2000 assimilation steps and the rms innovation and innovation spread for the 9 cases. As F becomes increasingly different from 8, the values of λ and the innovation and spread grow. The adaptive algorithm is able to

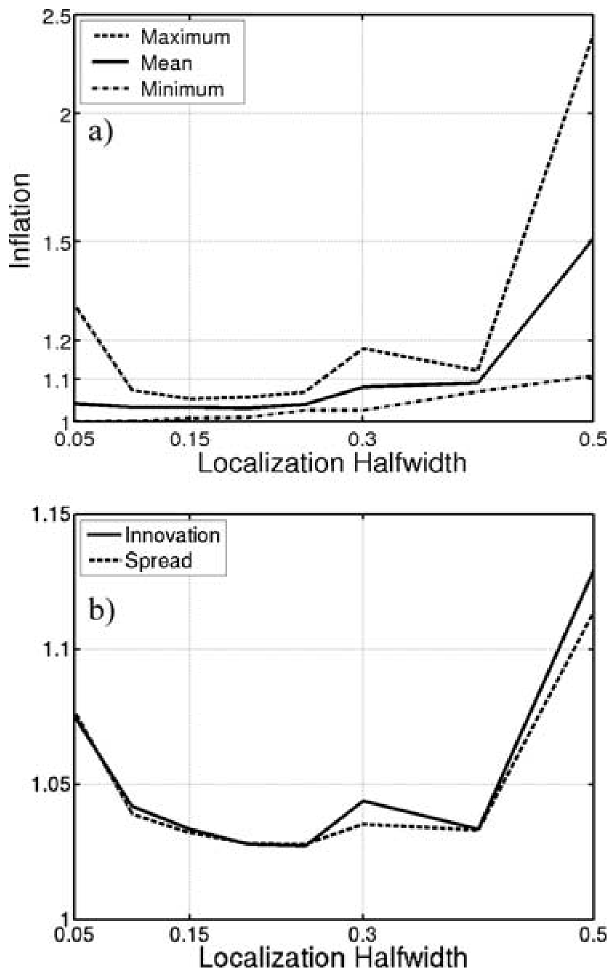

Fig. 4. As in Fig. 3, but as a function of the Gaspari-Cohn localization half-width in a 10-member ensemble assimilation.

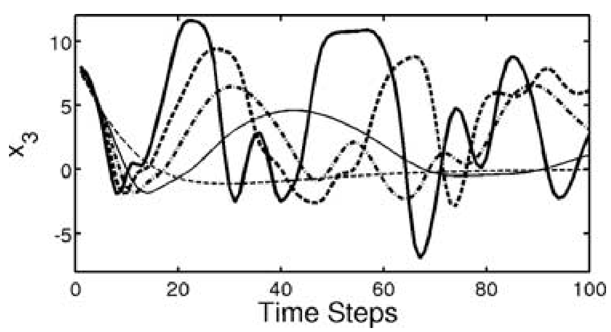

Fig. 5. Time evolution of state variable x_3 for different values of forcing F: $F = 11$ (thick solid), 8 (thick dashed), 5 (dot-dashed), 3 (thin solid) and 0 (thin dashed).

produce successful assimilations over a wide range of model errors. Searching for time constant inflation factors for these cases (see Section 7.5) required a large number of iterations for each value of F.

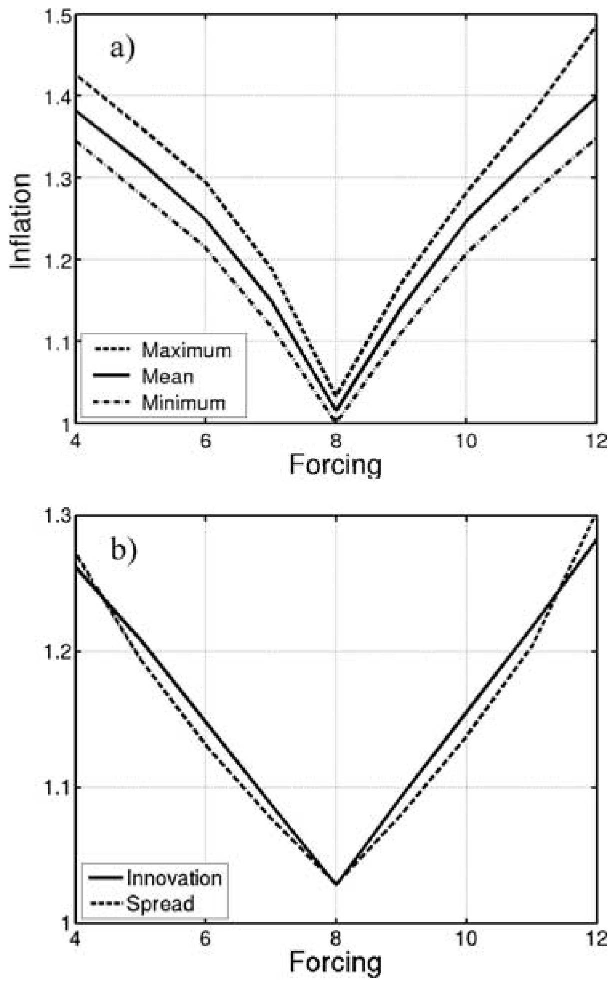

Fig. 6. As in Fig. 3 but as a function of forcing, F, in assimilating model for 20-member ensemble with 40 randomly located observations.

Since the uncertainty, and hence the range of the prior ensemble, gets larger as the error in F is increased, the error in the regression step is also expected to increase. The linear approximation used for the regression becomes increasingly suspect as the range of the prior ensemble sample becomes larger. Also, the relation between state variables in the imperfect assimilating models becomes increasingly different from the relation in the model that produced the observations. The result is that the assimilating model is unable to produce the correct relation between the state and observed variables (Mitchell et al., 2002).

To separate the impact of increased regression error from model error, the model error experiments are repeated with 100 randomly located observing stations. The increased observation density reduces the assimilation error in all cases and therefore the impact of the regression errors. Fig. 7 plots the time mean values of λ and the rms innovation and innovation spread for the denser observation case. Again, all increase as the error in F increases, but the rms innovation increases more slowly than in

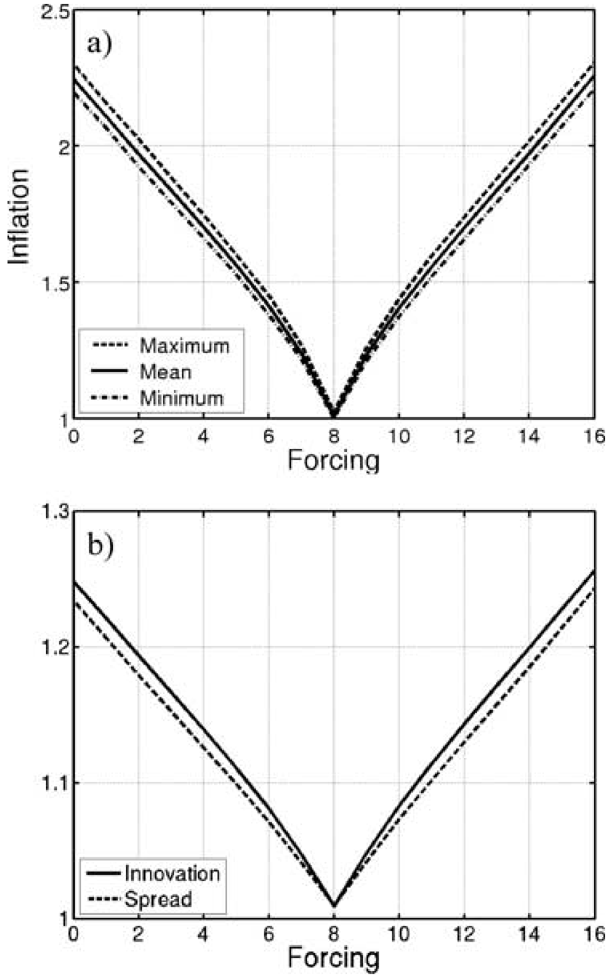

Fig. 7. As in Fig. 6 but for assimilation of 100 randomly located observations. Note the extended horizontal axis when comparing to Fig. 6.

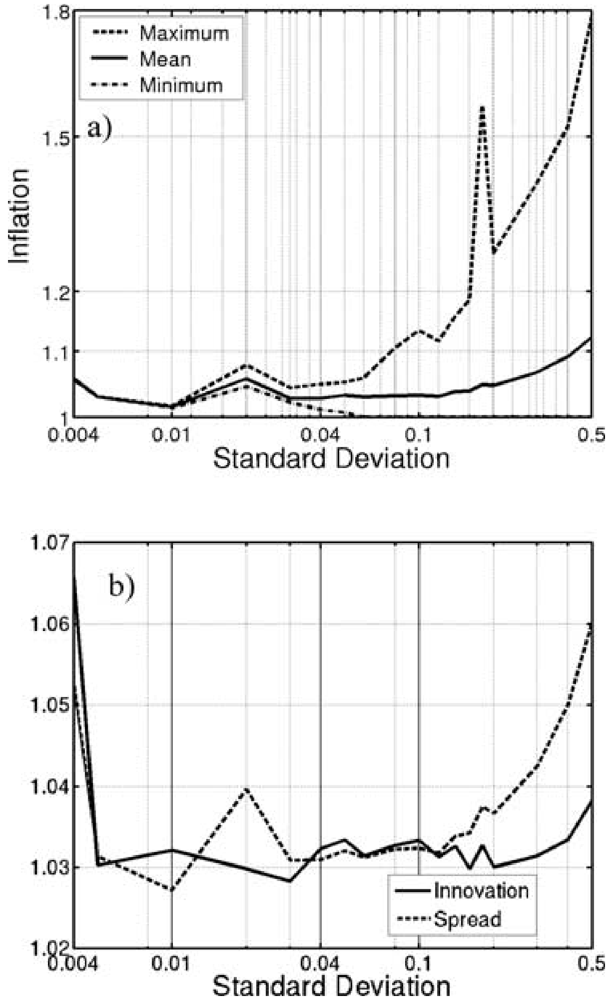

Fig. 8. As in Fig. 3 but as a function of the standard deviation of λ for assimilations with 10 members with 40 randomly located observations.

the 40 observation case. On the other hand, the inflation is larger due to the increased number of observations which lead to more spurious reduction of the variance estimates. Results are shown for $F = 1, 2, \ldots, 16$ in this case.

7.4. Sensitivity to fixed inflation variance

As noted in Section 6, the inflation variance, σ_c^2 (6.1), is held fixed in each experiment. In all cases discussed in Section 7 so far, it has been set to 0.05. Assimilations of 40 randomly located observing stations with a 10 member ensemble were performed for values of $\sigma_c = \{0.004, 0.006, 0.008, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.12, 0.15, 0.2, 0.25, 0.3, 0.4, 0.5\}$. The time mean values of λ and the rms innovation and innovation spread are plotted in Fig. 8. While there is some fluctuation in the values, the filter performance is roughly the same through nearly two orders of magnitude change in σ_c, or four orders of magnitude for σ_c^2. σ_c^2 controls how sensitively

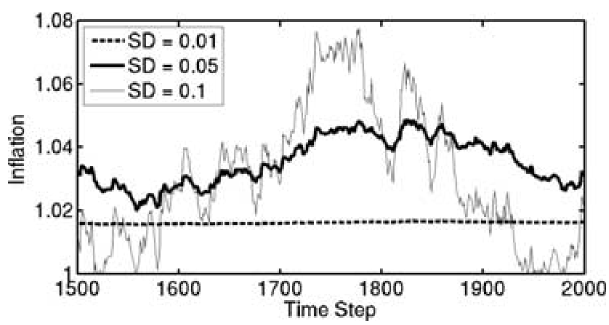

Fig. 9. Time sequence of values of adaptive inflation λ over last 500 steps of assimilation for different values of standard deviation of λ.

$\bar{\lambda}$ responds to observations. The smaller σ_c^2, the less rapid the response as more weight is given to the prior distribution. The larger range of minimum to maximum values of λ in Fig. 8a as σ_c^2 is increased is indicative of this sensitivity. Fig. 9 plots the time-series of $\bar{\lambda}$ values from the runs with $\sigma_c^2 = \{0.01, 0.05,$

0.1}. As σ_c^2 increases, the time-series become smoother while smaller σ_c^2 leads to relatively noisy time-series.

As noted in Section 6, there would be little purpose in using adaptive observation space inflation if it required as much work to tune σ_c^2 as to tune λ for a fixed inflation filter. In cases like these model error experiments, it requires a large number of trials to find values of λ that produce assimilations that are of high quality while the adaptive inflation filter is able to produce relatively high-quality assimilations without tuning.

7.5. Comparison to fixed tuned physical space inflation

Trial and error was used to find good time constant values of covariance inflation for the 100-observing-station model error experiments described in Section 7.3. The optimal time-constant λ and the time mean adaptive λ are plotted in Fig. 10a; the time-constant values are always larger. The rms innovation from the adaptive inflation case along with the rms innovation and spread from the time-constant cases are shown in Fig. 10b. The innovations for the optimal time-constant cases are smaller and become progressively more so as the model error is increased. The innovation spread from the time-constant cases is significantly larger than the rms innovation and is very close to the rms innovation values for the adaptive cases.

Figure 11 displays the time-mean rms errors and spread in state space for the 100-observation model error experiments (compare to Fig. 10b for innovations in observation space). The state space rms error is defined as the 2000-step time average of E, the ensemble mean rms difference from the truth for the state vector

$$E = \sqrt{\sum_{j=1}^{M} \left[\sum_{i=1}^{N} x_{j,i} \Big/ N - \hat{x}_j \right]^2 / M} \tag{7.3}$$

and the associated prior spread

$$S = \sqrt{\sum_{j=1}^{M} \sum_{k=1}^{N} \left[\sum_{i=1}^{N} x_{j,i}/N - x_{j,k} \right]^2 \Big/ M(N-1)} \tag{7.4}$$

where the hat represents the true value from the model integration that generated the observations and M is now the model size. The best tuned time-constant value still produces smaller errors, but the differences are smaller than for innovations in observation space. This reflects the fact that the fixed inflation values were tuned to minimize innovations in observation space. The tuned inflation values also result in state space spread that is much too large. The adaptive spread is also larger than the adaptive rms error, but it is not nearly as inconsistent as in the fixed inflation case.

The time-constant results are expected to be better because the adaptive cases know nothing about errors that occur when computing how much a change in λ should inflate the observed variable prior variance. The observation likelihood term (5.8) implicitly assumes that the prior relation between the observation and the state variables is Gaussian and that the prior ensemble sample statistics accurately summarize the linear relation. Errors here naturally reduce the variance of the state variables compared to what would be expected in the limit of large ensembles. If the assumption of a linear relation between the observation priors and the state variable priors is incorrect, these errors lead to errors in the state variable estimates which are also consistent with too little spread in state space. The trial and error cases overcome this error source by overinflating to correct these errors. The observations are then weighted more heavily than in the adaptive filters which acts to reduce the impact of the model and assimilation system error.

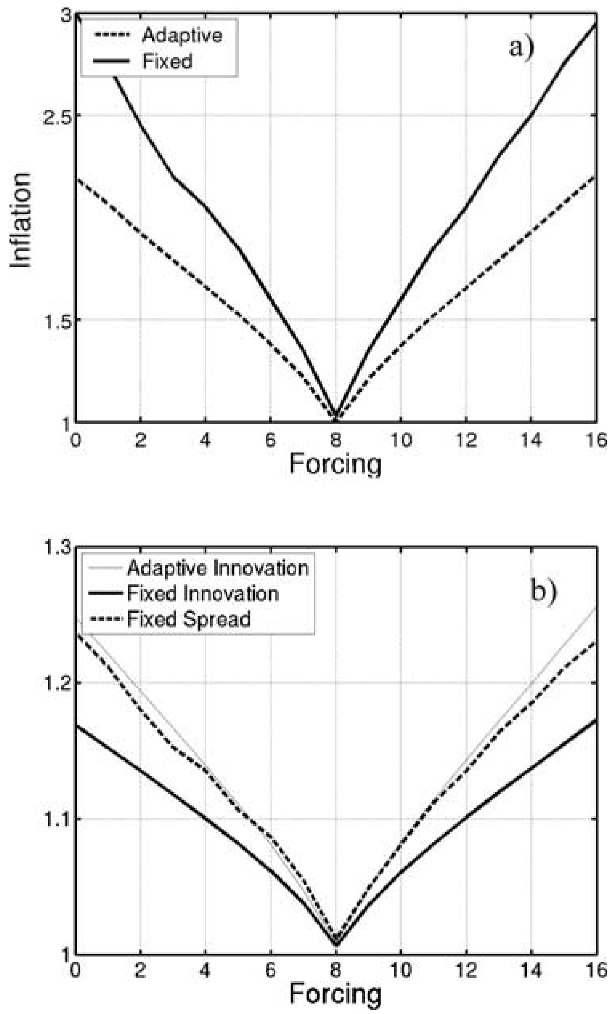

Fig. 10. Panel (a) shows the time mean value of adaptive inflation and the value of optimal fixed state space inflation as a function of forcing. Panel (b) shows the time mean rms innovation for the adaptive filter and the rms innovation and innovation spread for the optimal fixed inflation as a function of forcing. Both are for the case with 100 randomly located observations and 20 member ensembles.

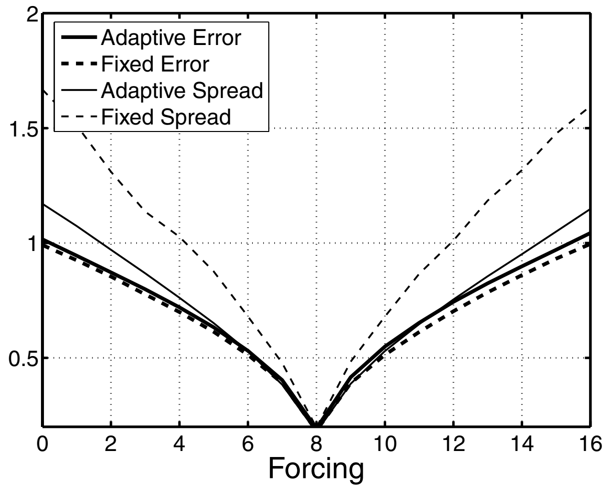

Fig. 11. Time mean prior rms error and spread in state space for same case shown in Figure 10 for an adaptive inflation filter and the best fixed inflation filter.

7.6. *Ensemble Kalman filter*

The adaptive inflation algorithm can be applied in concert with other observation space update methods like ensemble kernel filters or the classical stochastic update algorithm, the perturbed observations ensemble Kalman filter (EnKF) (Evensen, 1994). All of the cases performed here were repeated with the EnKF instead of the ensemble adjustment filter and the results were compared. For almost all cases, the time mean values of λ were slightly less, the rms state space error and observation space innovations were slightly larger, and the spreads in state space and observation space were slightly lower for the EnKF. It is unclear why the EnKF performed in this fashion. Previous research has shown that the relative performance of the EnKF and the ensemble adjustment filter can vary between different models and observation sets. The results here should not be interpreted as indicating that one method is superior, but do indicate that the adaptive algorithm can be applied with a variety of ensemble filter algorithms.

8. Discussion

The adaptive inflation algorithm developed here is able to produce improved assimilations and variance estimates for problems with a wide range of error magnitudes and sources. Assimilating in the presence of significant model error is especially significant. This is a challenging test for assimilation methodologies in general and filters in particular. Many filters as developed in the literature have assumed that models are perfect, clearly far from the case for many applications like numerical weather prediction. More traditional algorithms like 3D-var and earlier optimal interpolation algorithms implicitly introduced a capability to deal with model error through specified background error statistics. These statistics were tuned for particular applications and ended up correcting for a variety of error sources including background error. For ensemble filters to be applied without a need for extensive application-specific tuning, adaptive methods that detect and compensate for model error are essential. The fact that the filter can adapt to such a wide range of errors using a single value of the parameter σ_c^2 is important. Especially for the model error cases, it is non-trivial to find values of time-constant inflation that are as good as the adaptive filter. In large model applications, the cost of repeated assimilations to determine good values of fixed inflation may be prohibitive.

The cost of the adaptive inflation algorithm is small compared to the cost of the rest of an ensemble filter for many applications. The cost of the standard filter for one observation is O(NC) where N is the ensemble size and C is the number of state variables that are impacted by the observation. The computation tends to be dominated by the cost of computing N forward observation operators, but even more by computing C regression coefficients from N member samples. The additional cost of the inflation algorithm for one observation is constant and is dominated by the cost of solving (A.8) with the cubic formula and (5.10). The cost of solving (A.8) is dominated by evaluating a cube root and two trigonometric functions while the cost of solving (5.10) is dominated by a natural logarithm. As ensemble size N and/or the number of state variables impacted by each observation C increases, the relative cost of the inflation will decrease. In the experiments described here with the Lorenz-96 model, the inflation algorithm takes less than 5% of the total computation time. In experiments assimilating 200,000 real observations a day with a 5 million variable atmospheric GCM, the inflation cost is less than 1% of the total computation cost.

Ensemble filters should be expected to return sample information not only about the mean value of the state variables given the observations, but also about higher order moments, in particular the variance and covariance of the state variables. Because of a variety of errors, the variance in state estimates tends to be underestimated and covariance inflation is an attempt to rectify this. In the inflation algorithms discussed here, the prior distribution is inflated just before the application of forward operators. However, it is obvious that the posterior distribution and any forecasts generated from analyses will still have too little variance due to errors in the assimilation. In order to generate analyses (posteriors) with appropriate variance, inflation would have to be performed *after* the assimilation. It is not clear that this can be done without with-holding some portion of the observations and using these independent observations to adjust the variance.

Doing inflation at more than one point in the ensemble algorithm is a possible extension. Errors can be partitioned into those coming from the assimilation and those from model error. Inflating after the assimilation step to generate sufficient posterior variance and after the model advance to generate sufficient prior variance should be possible. This would also help to isolate the magnitude of model error which could then be used to produce ensemble forecasts with more realistic variance. Obviously, model error is a function of the time between observations. If this time is not constant, a more sophisticated model for the time tendency of inflation would be required.

A major challenge in adaptive inflation algorithms is selecting an appropriate prior distribution for the inflation. As discussed above, for a particular application it is not clear if the prior inflation variance should be allowed to asymptote to zero or not. In some applications, it may be appropriate to try to find a time-constant inflation. In other applications, it may be essential for inflation to adapt with time, for instance in experiments where the observation set changes with time or in experiments where an assimilation must be started from some sort of 'climatological' or non-equilibrium initial ensemble members. Fixing the inflation variance has worked for a variety of problems here, but other solutions might be appropriate. Some model of the way in which the inflation distribution changes with time (5.3) would be useful in this context but it might be difficult to develop one.

Selecting the appropriate prior becomes even harder if the assimilation problem is not homogeneous in time. For instance, if a large number of observations is available infrequently while smaller sets are available at intermediate times. After assimilating the large set, the spurious variance reduction in the ensemble is expected to be larger. A prior conditioned on a sequence of smaller observation sets might have inappropriately small inflation. This is the same issue that requires the algorithmic approach discussed when many observations are available at the same time. One could attempt to do the inflation in a purely sequential algorithm. Each scalar observation would update the inflation. Then the state would be inflated before computing the next forward operator. The first prior would have to include inflation to account for model error. The second would not.

In the sample problems above, it has been assumed that there exists a (possibly) temporally varying inflation that is appropriate for all state variables. This is clearly not the case in more realistic applications. In fact, one of the problems with traditional fixed covariance inflation is that it can lead to excessive variance for poorly observed model state variables. Assimilations of radiosonde observations with a tropospheric general circulation model provide an example. Areas over the south Pacific may be far removed from any observations while areas over North America may have state variables that are close to a large number of observations. Inflation is likely to be required for state variables over North America to avoid filter divergence caused by errors in regression coefficients. Over the south Pacific, repeated application of inflation is never countered by the impact of observations. Eventually, ensemble members can be inflated to unrealistic or numerically unacceptable values.

An inflation solution to this problem requires values that vary spatially, perhaps going so far as to allow each state variable to have its own inflation value. Adaptive algorithms can be constructed for this case. Another possibility is to associate inflation with the observations themselves and perform the inflation in observation space when the prior estimate of an observation variable is computed. In this case, only state variables being impacted by observations would be subject to inflation. Research on these two approaches will be an extension to the results discussed here.

9. Acknowledgments

The author is indebted to Nancy Collins, Tim Hoar, Hui Liu and Kevin Raeder for their work on the Data Assimilation Research Testbed. Comments from Doug Nychka, Chris Snyder, Shree Khare and two anonymous reviewers significantly improved this manuscript.

10. Appendix A: Finding the mode of the posterior distribution

The numerator of eq. (5.7) is

$$(\sqrt{2\pi}\theta)^{-1}\exp(-D^2/2\theta^2)\,\mathrm{Normal}\left(\bar{\lambda},\sigma_\lambda^2\right) \tag{A.1}$$

The mode of eq. (5.9) is at the maximum value of (A.1) which can be differentiated with respect to λ. This is

$$\frac{\partial}{\partial\lambda}(uvw) = uv\frac{\partial w}{\partial\lambda} + uw\frac{\partial v}{\partial\lambda} + vw\frac{\partial u}{\partial\lambda} \tag{A.2}$$

where

$$\begin{aligned} u &= (\sqrt{2\pi}\theta)^{-1},\ v = \exp(-D^2/2\theta^2),\\ w &= (\sqrt{2\pi}\sigma_\lambda)^{-1}\exp\left[-(\lambda-\bar{\lambda})^2/2\sigma_\lambda^2\right]. \end{aligned} \tag{A.3}$$

The individual partial derivatives are:

$$\frac{\partial u}{\partial\lambda} = -\frac{1}{2}\sigma_p^2\theta^{-2}u \tag{A.4}$$

$$\frac{\partial v}{\partial\lambda} = \frac{1}{2}\sigma_p^2 D^2\theta^{-4}v \tag{A.5}$$

$$\frac{dw}{d\lambda} = -\sigma_\lambda^{-2}\left(\lambda-\bar{\lambda}\right)\,w. \tag{A.6}$$

Substituting for λ with

$$\lambda = \left(\theta^2-\sigma_o^2\right)/\sigma_p^2 \tag{A.7}$$

gives a sixth order polynomial in θ which can be rewritten as a cubic polynomial in $x = \theta^2$. Setting this to 0 and removing a common monomial factor leaves

$$x^3 - \left(\sigma_o^2+\bar{\lambda}\sigma_p^2\right)x^2 + \frac{1}{2}\sigma_\lambda^2\sigma_p^4 x - \frac{1}{2}\sigma_\lambda^2\sigma_p^4 D^2 = 0 \tag{A.8}$$

which can be solved for x by the cubic formula. Substituting the value of x into eq. (A.7) gives a maximum value of λ;

11. Appendix B: The Lorenz-96 model

The L96 (Lorenz, 1996) model has N state variables, $X_1, X_2, \ldots, X_N$, and is governed by the equation

$$dX_i\big/dt = (X_{i+1} - X_{i-2})\,X_{i-1} - X_i + F, \tag{B.1}$$

where $i = 1, \ldots, N$ with cyclic indices. Here, N is 40, $F = 8.0$ unless otherwise noted, and a fourth-order Runge–Kutta time step with d$t = 0.05$ is applied as in Lorenz and Emanuel (1998).

References

Anderson, J. L. 2001. An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.* **129**, 2894–2903.

Anderson, J. L. 2003. A local least squares framework for ensemble filtering. *Mon. Wea. Rev.* **131**, 634–642.

Anderson, J. L. 2007. Exploring the need for localization in ensemble data assimilation using an hierarchical ensemble filter. Physica D, in press.

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

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

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

Chui, C. K. and Chen G. 1987. *Kalman Filtering*, Springer, Berlin.

Courtier, P., Thepaut J.-N. and Hollingsworth A. 1994. A strategy for operational implementation of 4D-var using an incremental approach. *Quart. J. Roy. Meteor. Soc.* **120**, 136701387.

Daley, R. 1993. Estimating the observation error statistics for atmospheric data assimilation. *Ann. Geophys.* **11**, 634–647.

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 da Silva, A. M. 1999. Maximum-likelihood estimation of forecast and observation error covariance parameters. Part I: methodology. *Mon. Wea. Rev.* **127**, 1822–1834.

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

Dee, D. P., Gaspari, G., Redder, C., Rukhovets, L. and da Silva, A. M. 1999. Maximum-likelihood estimation of forecast and observation error covariance parameters. Part II: applications. *Mon. Wea. Rev.* **127**, 1835–1849.

Derber, J. C. 1989. A variational continuous assimilation technique. *Mon. Wea. Rev.* **117**, 2437–2446.

Derber, J. C., Parrish, D. F. and Lord, S. J. 1991. The new global operational analysis system at the National Meteorological Center. *Wea. Forecasting* **6**, 538–547.

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

Eyre, J. R., Kelly, G. A., McNally, A. P., Andersson, E. and Persson A. 1993. Assimilation of TOVS radiance information through one-dimensional variational analysis. *Quart. J. Roy. Meteor. Soc.* **119**, 1427–1463.

Furrer, R. and Bengtsson T. 2006. Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. *J. Multivariate Anal* **98**, 227–255.

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

Geiring, R., Kaminski, T. and Slawig T. 2005. Generating efficient derivative code with TAF: adjoint and tangent linear Euler flow around an airfoil. *Future Generation Comp. Syst.* **21**, 1345–1355.

Hamill, T. M. and Whitaker, J. S. 2004. Accounting for error due to unresolved scales in ensemble data assimilation: a comparison of different approaches. *Mon. Wea. Rev.* **133**, 3132–3147.

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

Hansen, J. A. 2002. Accounting for model error in ensemble-based state estimation and forecasting. *Mon. Wea. Rev.* **130**, 2373–2391.

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

Houtekamer, P. L., and Mitchell, H. L. 2001. A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.* **129**, 123–137.

Houtekamer, P. L., Mitchell, H. L. Pellerin G. Buehner M. Charron M., and co-authors. 2004. Atmospheric data assimilation with the ensemble Kalman filter: results with real observations. *Mon. Wea. Rev.* **133**, 604–620.

Jazwinski, A. H. 1970. *Stochastic Processes And Filtering Theory*. Academic Press, New York, 376 pp.

Kalman, R. E., 1960. A new approach to linear filtering and prediction problems. *Trans. AMSE J. Basic Eng.* **82D**, 35–45.

Keppenne, C. L. and Rienecker, M. M. 2002. Initial testing of a massively parallel ensemble Kalman filter with the Poseidon isopycnal ocean general circulation model. *Mon. Wea. Rev.* **130**, 2951–2965.

Kistler, R., Collins W. Saha S. White, G. and Woolen J. 2001. The NCEP-NCAR 50-year reanalysis: monthly means CD-ROM and documentation. *Bull. Amer. Met. Soc.* **82**, 247–267.

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

Li, Z. and Navon, I. M. 2001. Optimality of variational data assimilation and its relationship with the Kalman filter and smoother. *Quart. J. Roy. Meteor. Soc.* **127**, 661–683.

Lorenc, A. C., 2003. The potential of the ensemble Kalman filter for NWP. *Quart. J. Roy. Meteor. Soc.* **129**, 3183–3204.

Lorenz, E. N., 1996. Predictability: a problem partly solved. Proc. ECMWF Seminar on Predictability, Vol. 1, Reading, United Kingdom, ECMWF, 1–18.

Lorenz, E. N. and Emanuel, K. A. 1998. Optimal sites for supplementary weather observations: simulation with a small model. *J. Atmos. Sci.* **55**, 399–414.

Mellor, G. L. and Yamada T. 1982. Development of a turbulent closure model for geophysical fluid problems. *Rev. Geophys. Space Phys.* **20**, 851–875.

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

Mitchell, H. L., Houtekamer, P. L. and Pellerin G. 2002. Ensemble size, balance, and model-error representation in an ensemble Kalman filter. *Mon. Wea. Rev.* **130**, 2791–2808.

Ott, E., Hunt B., Szunyogh, I., Zimin, A., Kostelich, and co-authors. 2004. A local ensemble Kalman filter for atmospheric data assimilation. *Tellus* **56A**, 415–428.

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

Rabier, F., Jarvinen, H., Klinker, E., Mahfouf, J. F. and Simmons A. 2000. The ECMWF operational implementation of four-dimensional variational assimilation. I: experimental results with simplified physics. *Quart. J. Roy. Meteor. Soc.* **126**, 1148–1170.

Talagrand, O. and Courtier P. 1987. Variational assimilation of meteorological observations with the adjoint vorticity equation. Part I: theory. *Quart. J. Roy. Meteor. Soc.* **113**, 1311–1328.

Tippett, M. K., Anderson, J. L. Bishop, C. H. Hamill, T. M. and Whitaker, J. S. 2003. Ensemble square root filters. *Mon. Wea. Rev.* **131**, 1485–1490.

Uppsala, S. M. and 45 contributing authors 2005. The ERA-40 re-analysis. *Quart. J. Roy. Meteor. Soc.* **131,** 2961–3012.

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