OBSTACLES TO HIGH-DIMENSIONAL PARTICLE FILTERING

12 March 2007

DRAFT. To be submitted to *Mon. Wea. Rev.?*

* The National Center for Atmospheric Research is sponsored by the National Science Foundation.

Corresponding author address: C. Snyder, NCAR, P.O. Box 3000, Boulder, CO 80307-3000;
chriss@ucar.edu

ABSTRACT

1. NOTES

Probably need to introduce the notion of an “effective” dimension from the beginning (Intro). Collapse not determined solely by N_x or N_y but also depends on prior pdf and $\mathbf{H}$.

Obs distributed in time and cycling. Without resampling, collapse just gets worse: more obs and more variability in obs priors.

2. INTRODUCTION

Ensemble methods for data assimilation are presently undergoing rapid development. The ensemble Kalman filter (EnKF), in various forms, has been successfully applied to a wide range of geophysical systems including atmospheric flows from global to convective scales (Whitaker et al. 2004, Snyder and Zhang 2003), oceanography from global to basin scales (refs?), and the land surface (Zhou et al. 2006). Particle filters are another class of ensemble assimilation methods of interest in geophysical applications (refs?). In their purest form, particle filters simply calculate posterior weights for each ensemble member based on their likelihood given present observations. Like the EnKF, particle filters are simple to implement and largely independent of the forecast model, but they have the added attraction that they are, in principle, fully general implementations of Bayes rule and applicable to highly non-Gaussian probability distributions. Unlike the EnKF, particle filters have so far mostly been applied to low-dimensional systems. This paper examines obstacles to applying particle filters in high-dimensional systems.

The goal of data assimilation in its most general form is to estimate the pdf of the state of the system given observations of the system. Both particle filters and the EnKF are Monte-Carlo techniques—they work with samples (i.e., ensembles) rather than directly with pdfs. Naively, one would expect such techniques to require ensemble sizes large compared to the dimension N_x of the state vector **x**. (We defer precise definitions to section 2.) Experience has shown, however, that this requirement does not hold for the EnKF if localization of the sample covariance matrix is employed (Houtekamer and Mitchell 1998, 2001; Hamill et al. 2001). Some theoretical justification for this exists (Furrer and Bengtsson 2007, Bickel and Levina 2007) although many details remain to be understood.

There is much less experience with particle filters in high dimensions. Lit review: Van Leeuwen (2003), Pham (2001), Evensen and van Leeuwen (2000), Xiong et al. (2006), land-surface and hydrology literature (Zhou et al. 2006). Any claims to high-dimensional systems other than van L? Land-surface and hydrology have lots of variables but not coupled horizontally?

We argue here that high-dimensional particle filters face fundamental difficulties. Specifically, we show that, unless the ensemble size N_e increases exponentially with N_x, the particle-filter update suffers from a "collapse" in which with high probability a single member is assigned a posterior weight close to one while all other members have vanishingly small weights. Such collapse of weights has been remarked on previously in the geophyscial literature (Anderson and Anderson 1999, Bengtsson et al. 2003, van Leeuwun 2003) as well as in the particle-filtering literature (refs?), where the phenomenon is also known as "unbalanced weights" (and ...?). Unlike previous studies, however, we emphasize collapse as a fundamental obstacle to particle filtering in high-dimensional systems, in that very large ensembles are required to avoid collapse even for system dimensions of a few tens or hundreds. Demonstrate through simulations, heuristic arguments and reference to results of Li et al. (2007).

Heuristic background on high dimensional pdfs and "curse of dimensionality?"

3. NOTATION AND BACKGROUND ON PARTICLE FILTERS

Our notation will follow that of Ide et al. (1997) except for the dimensions of the state and observation vectors and our use of subscripts to indicate ensemble members.

Let $\mathbf{x}$ of dimension N_x be the state of the system represented in some discrete basis, such as the values of all prognostic variables on a regular grid. Since it can not be determined exactly given imperfect observations, we consider $\mathbf{x}^t$, the true state of the system, to be a random variable. Our aim is then to estimate $p(\mathbf{x}^t)$, the probability distribution function (pdf) of the true state given all available observations (or at least some of its moments).

The subsequent discussion will focus on the update of $p(\mathbf{x}^t)$ given new observations at some time $t = t_0$. That is, suppose that we have both a vector of observations $\mathbf{y}^o$ of dimension N_y that depends on $\mathbf{x}^t(t = t_0)$ and a prediction $p(\mathbf{x}^t(t = t_0))$. [To be more precise, $p(\mathbf{x}^t(t = t_0))$ is conditioned on all observations prior to $t = t_0$. Since all pdfs here will pertain at $t = t_0$ and will be conditioned on all previous observations, we will suppress explicit reference to t_0 and the previous observations in what follows.] We wish to estimate $p(\mathbf{x}^t|\mathbf{y}^o)$, the pdf of $\mathbf{x}^t$ given the observations $\mathbf{y}^o$, which we will term the posterior pdf.

For simplicity, let the observations have a linear relation to the state and be subject to additive random errors $\boldsymbol{\epsilon}$,

$$\mathbf{y} = \mathbf{H}\mathbf{x}^t + \boldsymbol{\epsilon}. \tag{1}$$

More general observation models are of course possible but (1) suffices for all the points we wish to make in this paper.

The particle filter begins with an ensemble of states $\{\mathbf{x}_i^f,\ i\ =\ 1,\ldots,N_e\}$ that is assumed to be drawn from $p(\mathbf{x}^t)$, where the superscript f (for "forecast") indicates a prior quantity. The ensemble members are also known as particles. The update step approximates $p(\mathbf{x}^t)$ by a sum of delta functions, $N_e^{-1}\sum_{i=1}^{N_e}\delta(\mathbf{x}^t-\mathbf{x}_i^f)$. Applying Bayes rule yields

$$p(\mathbf{x}^t|\mathbf{y}^o)=\frac{p(\mathbf{y}^o|\mathbf{x}^t)p(\mathbf{x}^t)}{\int p(\mathbf{y}^o|\mathbf{x}^t)p(\mathbf{x}^t)\,d\mathbf{x}^t}=\sum_{i=1}^{N_e}w_i\delta(\mathbf{x}^t-\mathbf{x}_i^f), \tag{2}$$

where the posterior weights are given by

$$w_i=\frac{p(\mathbf{y}^o|\mathbf{x}_i^f)}{\sum_{j=1}^{N_e}p(\mathbf{y}^o|\mathbf{x}_j^f)}. \tag{3}$$

In the posterior, each member $\mathbf{x}_i^f$ is weighted according to how likely the observations would be if $\mathbf{x}_i^f$ were the true state.

If the likelihoods $p(\mathbf{y}^o|\mathbf{x}_i^f)$ vary greatly, $\max_i w_i$ will approach one and the particle filter approximates the posterior pdf as single point mass. The particle-filter estimates of posterior expectations, such as the posterior mean

$$E(\mathbf{x}^t|\mathbf{y}^o)=\int\mathbf{x}^t p(\mathbf{x}^t|\mathbf{y}^o)\,d\mathbf{x}^t\approx\sum_{i=1}^{N_e}w_i\mathbf{x}_i^f, \tag{4}$$

are then very poor approximations. We will loosely term this situation, in which one (or a small subset) of the members is given almost all the posterior weight, as "collapse" of the particle filter. The goal of our study is to describe the situations in which collapse occurs, both through rigorous asymptotics as N_y and N_e become large and through simulations and heuristic rules informed by the asymptotics. The large-N_y and N_e results come from Li et al. (2007) and further details and proofs can be found there.

4. FAILURE OF THE PARTICLE FILTER IN A SIMPLE EXAMPLE

We next consider perhaps the simplest possible example, in which the prior distribution $p(\mathbf{x}^t)$ is Gaussian with each component of $\mathbf{x}^t$ independent and of unit variance, and the observations $\mathbf{y}$ are of each component of $\mathbf{x}^t$ individually with independent, Gaussian errors of unit variance. More concisely, consider $N_y = N_x$, $\mathbf{H} = \mathbf{I}$, $\mathbf{x}^t \sim N(0, \mathbf{I})$, and $\boldsymbol{\epsilon} \sim N(0, \mathbf{I})$, where the symbol $\sim$ means "is distributed as" and $N(\bar{\mathbf{x}}, \mathbf{P})$ is the Gaussian distribution with mean $\bar{\mathbf{x}}$ and covariance matrix $\mathbf{P}$.

Figure 1 shows histograms for $\max_i w_i$ from simulations of the particle-filter update using $N_x = 10$, 30, and 100 and $N_e = 10^3$. In the simulations, $\mathbf{x}^t$, $\boldsymbol{\epsilon}$ and an ensemble $\{\mathbf{x}_i^f,\ i = 1, \ldots, N_e\}$ are drawn from $N(0, \mathbf{I})$. Weights w_i are then computed from (3). The histograms are based on 10^3 realizations for each value of N_x.

The maximum w_i is increasingly likely to be close to one as N_x and N_y increase. Large weights appear occasionally in the case $N_x = 10$, for which $\max_i w_i > 0.5$ in just over 6% of the simulations. Once $N_x = 100$, the average value of $\max_i w_i$ over the 10^3 simulations is greater than 0.8 and $\max_i w_i > 0.5$ with probability 0.9. Collapse of the weights occurs frequently for $N_x = 100$ despite the ensemble size $N_e = 10^3$.

Two comparisons illustrate the detrimental effects of collapse. The correct posterior mean in this Gaussian example is given by $\bar{\mathbf{x}}^a = (\bar{\mathbf{x}}^f + \mathbf{y}^o)/2$, where the overbar denotes expectation and the superscript a (for "analysis") indicates a posterior quantity. The expected squared error of $\bar{\mathbf{x}}^a$ is $E((\bar{\mathbf{x}}^a - \mathbf{x}^t)^2) = [E((\bar{\mathbf{x}}^f - \mathbf{x}^t)^2 + E((\mathbf{y}^o - \mathbf{x}^t)^2)]/4 = N_x/2$, while that of the observations

[$E((\mathbf{y}^o - \mathbf{x}^t)^2)$] is equal to N_x. The posterior mean estimated by the particle filter,

$$\hat{\mathbf{x}}^a = \sum_{i=1}^{N_e} w_i \mathbf{x}_i^f \, ,$$

has squared error of 5.5, 25 and 127 for $N_x = 10, 30, 100$, respectively, when averaged over the simulations. Thus, $\hat{\mathbf{x}}^a$ has error close to that of $\bar{\mathbf{x}}^a$ when $N_x = 10$. The particle filter's performance degrades somewhat for $N_x = 30$, with squared error more than 50% larger than the optimal. For $N_x = 100$, however, collapse of the weights is pronounced. $\hat{\mathbf{x}}^a$ is then a very poor estimator of the posterior mean—it has *larger* errors than either the prior or the observations.

As might be expected, the effects of collapse are also apparent in the particle-filter estimate of posterior variance. The correct posterior variance [tr(cov $\mathbf{x}^a$)] is $N_x/2$, yet the particle-filter estimates are 4.7, 10.5, 19.5 for $N_x = 10, 30, 100$, respectively. Except for $N_x = 10$, the particle-filter update significantly underestimates the posterior variance, especially when compared to the squared error of $\hat{\mathbf{x}}^a$.

The natural question is how large the ensemble must be in order to avoid the complete failure of the update. This example is tractable enough that the answer may be found by direct simulation: for various N_x, we simulate with $N_e = 10 \cdot 2^k$ and increase k until the average squared error of $\hat{\mathbf{x}}^a$ is less than that of the prior or the observations. The N_e required to reach this minimal threshold is shown as a function of N_x (or N_y) in Fig. 2.

The required N_e appears to increase exponentially in N_x. The limitations this increase places on implementations of the particle filter are profound. For $N_x = N_y = 90$, somewhat more than 3×10^5 ensemble members are needed. Ensemble sizes for larger systems can be estimated from the best-fit line shown in Fig. 2. Increasing N_x and N_y to 100 increases the necessary ensemble size to just under 10^6, while $N_x = N_y = 200$ would require 10^{11} members. We emphasize that ensembles of these sizes merely avoid the update being worse than the prior or observations.

5. BEHAVIOR OF WEIGHTS FOR LARGE N_y AND I.I.D. OBSERVATION ERRORS

It is not clear how general the foregoing example is. Results of Li et al. (2007), outlined in this section and the next, provide further guidance on the behavior of the particle-filter weights. Our discussion will be largely heuristic; we refer the reader to Li et al. for more rigorous and detailed proofs.

Suppose that each component ϵ_j of $\boldsymbol{\epsilon}$ is i.i.d. (independent and identically distributed) with density $f(\cdot)$. Then for each member $\mathbf{x}_i^f$, the observation likelihood can be written as

$$p(\mathbf{y}^o|\mathbf{x}_i^f) = \prod_{j=1}^{N_y} f\left(y_j^o - (\mathbf{H}\mathbf{x}_i^f)_j\right), \tag{5}$$

where y_j^o and $(\mathbf{H}\mathbf{x}_i^f)_j$ are the jth components of $\mathbf{y}^o$ and $\mathbf{H}\mathbf{x}_i^f$, respectively. An elementary consequence of (5) is that, given $\mathbf{y}^o$, the likelihood depends only on N_y, $f(\cdot)$ and the prior as reflected in the observed variables $\mathbf{H}\mathbf{x}$. There is no direct dependence on the state dimension N_x.

Defining $\psi(\cdot) = \log f(\cdot)$,

$$p(\mathbf{y}^o|\mathbf{x}_i^f) = \exp\left(\sum_{j=1}^{N_y} \psi\left(y_j^o - (\mathbf{H}\mathbf{x}_i^f)_j\right)\right) = \exp\left(-\sum_{j=1}^{N_y} V_{ij}\right), \tag{6}$$

where $V_{ij} = -\psi\left(y_j^o - (\mathbf{H}\mathbf{x}_i^f)_j\right)$, the log-likelihood of the jth observation component given the ith ensemble member. The simplest situation (as in the example of section 3) is when the random variables V_{ij}, $j = 1, \ldots, N_y$, are i.i.d. given $\mathbf{y}^o$. By the central limit theorem (? ref?), the sum in (6) then has an approximately Gaussian distribution for large N_y with mean $N_y\mu$ and variance $N_y\sigma^2$, where $\mu = E(V_{ij}|\mathbf{y}^o)$ and $\sigma^2 = \mathrm{var}(V_{ij}|\mathbf{y}^o)$. Thus, the likelihood can be approximated as

$$p(\mathbf{y}^o|\mathbf{x}_i^f) \approx \exp\left(-\mu N_y - \sigma N_y^{1/2} Z_i\right), \tag{7}$$

where $Z_i \sim N(0,1)$. [As discussed in Li et al. (2007) ?, (7) also holds when the V_{ij} are not i.i.d. but have sufficiently similar distributions (given $\mathbf{y}^o$) and are not too dependent, although general expressions for μ and σ^2 are not possible.]

The maximum weight $w_{(N_x)}$ can then be expressed as

$$w_{(N_x)} \approx \left[1 + \sum_{i=1}^{N_e} \exp(-\sigma N_y^{1/2}(Z_{(i)} - Z_{(1)}))\right]^{-1},$$

where $Z_{(i)}$) is the ith order statistic of the sample $\{Z_i,\ i = 1, \ldots, N_e\}$[1].

Argument that leads to $w_{(N_x)} \to 1$ if $(\log N_e)/N_y \to 0$:

NEED to discuss/emphasize role of σ in determining rate of collapse.

Collapse occurs if

$$\sum_{i=1}^{N_e} \exp(-\sigma N_y^{1/2}(Z_{(i)} - Z_{(1)})) \to 0$$

as N_e and N_y increase.

Largest term in this sum is the first, $\exp(-\sigma N_y^{1/2}(Z_{(2)} - Z_{(1)}))$. Extreme values of samples from Gaussian distribution have been thoroughly studied. As the sample size N_e increases, $Z_{(2)} - Z_{(1)} \sim 1/\log^{1/2} N_e$ (REF?). The largest term in the sum approaches zero for large N_e and N_y if $\log N_e/N_y \to 0$ and thus $\log N_e \geq N_y$ is sufficient to avoid collapse.

[Simulations suggest this is also a necessary condition (i.e., collapse occurs if $\log N_e/N_y \to 0$). Figure 3 shows the ensemble size necessary to have $E(\max w_i) < 0.6$, 0.7 or 0.8 as a func-

---

[1] In other words, $Z_{(1)}$ is the minimum, $Z_{(2)}$ is the next smallest element of the sample, and so on until the maximum, $Z_{(N_e)}$.

tion of N_y and N_x. Or, sufficient just to refer to Fig. 2? Emphasize that even larger ensemble sizes will be needed for PF to provide accurate estimates of posterior expectations—as shown by simulations also.]

Rigorous results proven in Li et al. (2007) are weaker, showing only that collapse occurs if $(\log N_e)/N_y^{1/3} \rightarrow 0$.

6. THE GAUSSIAN-GAUSSIAN CASE

The analysis in the previous section focused on situations in which the log likelihoods for the observations (considered as random functions of the prior) were mutually independent and identically distributed. In general, however, the observation likelihoods need not be i.i.d., since the state variables are correlated in the prior distribution and observations may depend on multiple state variables. In this section, we consider the case of a Gaussian prior, Gaussian observation errors and linear **H**, where analytic progress is possible even for general prior covariances and general **H**.

Let the prior $\mathbf{x}^t \sim N(\bar{\mathbf{x}}, \mathbf{P})$ and the observation error $\boldsymbol{\epsilon} \sim N(\bar{\boldsymbol{\epsilon}}, \mathbf{R})$. Since the observations depend linearly on the state, $E(\mathbf{y}) = \mathbf{H}E(\mathbf{x}^t)$ and $p(\mathbf{y}|\mathbf{x}^t)$ is unchanged if $\mathbf{y}$ is replaced by $E(\mathbf{y})$ and $\mathbf{x}^t$ by $\mathbf{x}^t - E(\mathbf{x}^t)$. We may therefore assume that both $\mathbf{x}^t$ and $\boldsymbol{\epsilon}$ have mean zero.

We can also assume that $\mathbf{R} = \mathbf{I}$ since, for general $\mathbf{R}$, the observations may always be transformed as $\mathbf{y}' = \mathbf{R}^{-1/2}\mathbf{y}$ so that $\mathrm{cov}(\boldsymbol{\epsilon}') = \mathrm{cov}(\mathbf{R}^{-1/2}\boldsymbol{\epsilon}) = \mathbf{I}$. Further simplification comes from diagonalizing $\mathrm{cov}(\mathbf{R}^{-1/2}\mathbf{H}\mathbf{x})$ via an additional orthogonal transformation in the observation space. Let $\mathbf{y}'' = \mathbf{Q}^T\mathbf{y}'$, where $\mathbf{Q}$ is the matrix of eigenvectors of $\mathrm{cov}(\mathbf{R}^{-1/2}\mathbf{H}\mathbf{x})$ with corresponding eigenvalues λ_j^2, $j = 1, \ldots, N_y$; then $\mathrm{cov}(\mathbf{Q}^T\mathbf{R}^{-1/2}\mathbf{H}\mathbf{x}) = \mathrm{diag}(\lambda_1^2, \ldots, \lambda_{N_y}^2)$ while $\boldsymbol{\epsilon}'' = \mathbf{Q}^T\boldsymbol{\epsilon}'$ still has identity covariance. We therefore assume, without loss of generality, that

$$\mathbf{R} = \mathbf{I}, \quad \mathrm{cov}(\mathbf{Hx}) = \mathbf{HPH}^T = \mathrm{diag}(\lambda_1^2, \ldots, \lambda_{N_y}^2), \tag{XX}$$

and drop primes in the sequel.

a. Analysis of the observation likelihood

With the assumptions (XX), the observation errors are independent, so $p(\mathbf{y}^o|\mathbf{x}_i^f)$ can be written in terms of a sum over the log-likelihoods V_{ij} as in (6). In addition, the pdf for each component of the observations is Gaussian with unit variance and, given $\mathbf{x}_i^f$, mean $\mathbf{Hx}_i^f$. Thus,

$$V_{ij} = -\left(y_j^o - (\mathbf{Hx}_i^f)_j\right)^2/2, \tag{Y}$$

where we have omitted an additive constant that results from the normalization of the Gaussian density and which cancels in the calculation of the weights w_i.

We wish to approximate the observation likelihood as in (7). This requires $\sum_{j=1}^{N_y} V_{ij}$ to be approximately Gaussian with mean μN_y and variance $\sigma^2 N_y$ for appropriate μ and σ. Leaving aside for the moment whether the sum is approximately Gaussian, its mean and variance given $\mathbf{y}^o$ can be calculated directly using (XX), the fact that the V_{ij} are independent as j varies, and the properties of the standard normal distribution. This yields

$$\mu N_y = E\left(\sum_{j=1}^{N_y} V_{ij}\right) = \sum_{j=1}^{N_y} \lambda_j^2 + y_j^{o2}, \tag{Za}$$

and

$$\sigma^2 N_y = \mathrm{var}\left(\sum_{j=1}^{N_y} V_{ij}\right) = \sum_{j=1}^{N_y} \lambda_j^2\left(\lambda_j^2 + 2y_j^{o2}\right). \tag{Zb}$$

Is this where to make point that λ_j^2 determine variability of likelihoods and thus control collapse? : Thus, obs likelihoods and weights in general Gaussian-Gaussian case are controlled by sum in (Zb). Key is not state dimension N_x or obs dimension N_y, but value of σ^2.

Conditions under which $\sum_{j=1}^{N_y} V_{ij}$ is asymptotically (as $N_y \rightarrow \infty$): λ_j bounded above and away from zero (A1 of Li et al.) and σ^2 in (Zb) approaches limiting value (A2 of Li et al.). Explain these in more detail?

Case in which these conditions are not met and $\sum_{j=1}^{N_y} V_{ij}$ is not close to Gaussian will be considered separately (subsection c)? Or discuss here?

b. Simulations

Simulations: Set $\lambda_j = j^\theta$. Vary N_y, N_e and θ.

Collapse occurs. Show that collapse is determined by σ^2 defined in (Zb)—even for different θ, collapse is the same if σ^2 is unchanged.

What diagnostics to show? If Peter has specific asymptotic predictions, those would be most natural. Or, compare rms error of posterior mean against optimal posterior and against obs error, as in first example? PF posterior can be poor even before collapse is complete?

Discuss subtleties that arise when λ_i decreases rapidly with i. [Then log likelihood is not close to Gaussian and (7) does not hold.] Extreme example is scalar state with many obs. Simulations to illustrate.

7. NON-GAUSSIAN PRIOR AND/OR OBSERVATION DISTRIBUTIONS

Summarize Li et al. results for this case.

LE40 simulations? Use these to generate non-Gaussian priors. Calculate λ_j's. Estimate effective dimension as for Gaussian-Gaussian. Evaluate usefulness of Gaussian-Gaussian results for predicting collapse in this case.

8. SIMULATIONS FOR “CYCLED” PARTICLE FILTERS

Heuristic summary of issues introduced by dynamics and cycling. (CS: not quite sure what these are or how to explain yet. DO WE NEED THIS SECTION?)

Use low-order model (Lorenz-Emanuel), simple atmospheric model or just specify simple (diagonal) linear dynamics?

Basic message: Collapse at least as bad with cycling. Can we say this more rigorously/precisely? (Are there pathological exceptions?)

9. FURTHER ISSUES

a. Van Leeuwen (2003) results

b. Non-Gaussian prior or observation errors

Now covered in section 5?

c. Nonlinear

Also section 5?

10. CONCLUSIONS

11. QUESTIONS ON PRESENTATION AND CONTENT

Discuss PF embellishments?: sampling strategies, etc. While such techniques can improve the performance of PF, we can find no evidence that they remove the exponential increase in the required N_e as the system dimension increases.

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.

Bengtsson T., C. Snyder, and D. Nychka, 2003: Toward a nonlinear ensemble filter for high-dimensional systems. *J. Geophys. Res.*, **108(D24)**, 8775–8785.

Bickel, P., and E. Levina, 2007: Regularized estimation of large covariance matrices, in preparation.

Evensen G., and P. J. van Leeuwen, 2000: An ensemble Kalman smoother for nonlinear dynamics. *Mon. Wea. Rev.*, **128**, 1852–1867.

Furrer, R., and T. Bengtsson, 2007: Estimation of high-dimensional prior and posteriori covariance matrices in Kalman filter variants. *Journal of Multivariate Analysis*, **98 (2)**, 227–255.

Li, B., T. Bengtsson and P. Bickel, 2007: Curse of dimensionality revisited: Collapse of the particle filter in very large scale systems. To appear in ??

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

Snyder, C. and F. Zhang, 2003: Assimilation of simulated Doppler radar observations with an ensemble Kalman filter. *Mon. Wea. Rev.*, **131**, 1663–1677.

Whitaker, J. S., G. P. Compo, X. Wei, T. M. Hamill, 2004: Reanalysis without radiosondes using ensemble data assimilation. *Mon. Wea. Rev.*, **132**, 1190–1200.

van Leeuwen, P. J.. 2003: A variance-minimizing filter for large-scale applications. *Mon. Wea. Rev.*, **131**, 2071–2084.

Xiong, X., I. M. Navon and B. Uzunoglu, 2006: A note on the particle filter with posterior Gaussian resampling. *Tellus*, **58A**, 456–460.

Zhou, Yuhua, D. McLaughlin and D. Entekhabi, 2006: Assessing the performance of the ensemble Kalman filter for land surface data assimilation. *Mon. Wea. Rev.*, **134**, 2128–2142.

Figure 1. Histograms of $\max w_i$ for $N_x = 10, 30, 100$ and $N_e = 10^3$ from the particle-filter simulations described in text [$N_e = 10^3$, $\mathbf{x}^t \sim N(0, \mathbf{I})$, $N_y = N_x$, $\mathbf{H} = \mathbf{I}$ and $\boldsymbol{\epsilon} \sim N(0, \mathbf{I})$].

Figure 2. The ensemble size N_e as a function of N_x (or N_y) required if the posterior mean estimated by the particle filter is to have average squared error less than the prior or observations, in the simple example considered in the text. Asterisks show the simulation results, averaged over 400 realizations. The best fit line is given by $\log_{10} N_e = 0.05N_x + 0.78$.

Figure 3. The ensemble size N_e as a function of N_x (or N_y) such that the expected max w_i, computed over 400 realizations, is less than 0.6 (plus signs), 0.7 (circles) and 0.8 (asterisks) in the simple example considered in the text.