OBSTACLES TO HIGH-DIMENSIONAL PARTICLE FILTERING

Chris Snyder

National Center for Atmospheric Research*, Boulder, CO

Thomas Bengtsson

Bell Labs, Murray Hill, NJ

Peter Bickel

Department of Statistics, University of California-Berkeley

Berkeley, CA

Jeff Anderson

National Center for Atmospheric Research, Boulder, CO

Submitted to *Mon. Wea. Rev.*, 2 January 2008

* 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

Particle filters are ensemble-based assimilation schemes that, unlike the ensemble Kalman filter, employ a fully nonlinear and non-Gaussian analysis step to compute the probability distribution function (pdf) of a system's state conditioned on a set of observations. Evidence is provided that the ensemble size required for a successful particle filter scales exponentially with the problem size. For the simple example in which each component of the state vector is independent, Gaussian and of unit variance, and the observations are of each state component separately with independent, Gaussian errors, simulations indicate that the required ensemble size scales exponentially with the state dimension. In this example, the particle filter requires at least 10^{11} members when applied to a 200-dimensional state. Asymptotic results, following the work of Bengtsson, Bickel and collaborators, are provided for two cases: one in which each prior state component is independent and identically distributed, and one in which both the prior pdf and the observation errors are Gaussian. The asymptotic theory reveals that, in both cases, the required ensemble size scales exponentially with the variance of the observation log-likelihood, rather than with the state dimension per se.

1. 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 (Keppenne et al. 2005), and the land surface (Reichle et al. 2002). Particle filters are another class of ensemble-based assimilation methods of interest in geophysical applications. [See Gordon et al. (1993) or Doucet et al. (2001) for an introduction.]

In their simplest form, particle filters calculate posterior weights for each ensemble member based on the likelihood of the observations given that member. 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, however, particle filters have so far mostly been applied to low-dimensional systems. This paper examines obstacles to applying particle filters in high-dimensional systems.

Both particle filters and the EnKF are Monte-Carlo techniques—they work with samples (i.e., ensembles) rather than directly with the underlying probability density function (pdf). Naively, one would expect such techniques to require ensemble sizes large compared to the dimension of the state vector. 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). The feasibility of the EnKF with ensemble sizes much smaller

than the state dimension also has theoretical justification. Furrer and Bengtsson (2007) and Bickel and Levina (2007) examine the sample covariance structure for reasonably natural classes of covariance matrices and demonstrate the effectiveness of localizing the sample covariance matrix.

There is much less experience with particle filters in high dimensions. Several studies have presented results from particle filters and smoothers for very low-dimensional systems, including that of Lorenz (1963) and the double-well potential (Pham 2001, Kim et al. 2003, Moradkhani et al. 2005, Xiong et al. 2006, Chin et al. 2007). Both van Leeuwen (2003) and Zhou et al. (2006), however, apply the particle filter to higher-dimensional systems. Van Leeuwen (2003) considers a model for the Agulhas current with dimension of roughly 2×10^5, while Zhou et al. (2006) use a land-surface model of dimension 684. We will return to the relation of our results to their studies in the concluding section.

We argue here that high-dimensional particle filters face fundamental difficulties. Specifically, we explore the result from Bengtsson et al. (2007) and Bickel et al. (2007) that, unless the ensemble size is exponentially large in a quantity τ^2, 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 . The quantity τ^2 is the variance of the observation log-likelihood, which depends not only on the state dimension but also on the prior distribution and the number and character of observations. As will be discussed later, τ^2 may be considered an effective dimension as it is proportional to the dimension of the state vector in some simple examples.

The tendency for collapse of weights has been remarked on previously in the geophyscial literature (Anderson and Anderson 1999, Bengtsson et al. 2003, van Leeuwen 2003) and is also

well known in the particle-filtering literature, where it is often referred to as "degeneracy," "impoverishment" or "sample attrition." Unlike previous studies, however, we emphasize the collapse of weights 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.[1]

Because of the tendency for collapse, particle filters invariably employ some form of resampling or selection step after the updated weights are calculated (e.g. Liu 2001), in order to remove members with very small weights and replenish the ensemble. We do not analyze resampling algorithms in this paper but rather contend that, whatever their efficacy for systems of small dimension and reasonably large ensemble sizes, they are unlikely to overcome the need for exponentially large ensembles as τ^2 grows. Resampling proceeds from the approximate posterior distribution computed by the particle filter; it does not improve the quality of that approximate posterior.

The outline of the paper is as follows. In section 2, we review the basics of particle filters. Section 3 illustrates the difficulty of particle filtering when τ^2 is not small through simulations for the simplest possible example: a Gaussian prior and observations of each component of the state with Gaussian errors, both of which have identity covariance. In section 4, we derive (following Bengtsson et al. 2007) an asymptotic condition on the ensemble sizes that yield collapse

[1] This obstacle is equally relevant to a related class of "mixture" filters in which the prior ensemble serves as the centers for a kernel density estimate of the prior (Anderson and Anderson 1999, Bengtsson et al. 2003, Smith 2007). These filters also involve the calculation of the weight of each center given observations, and thus are subject to similar difficulties.

when both the prior and observation errors are independent and identically distributed in each component of the state vector. Section 5 extends those results to the more general case of Gaussian priors and Gaussian observation errors. Section 6 briefly discusses the effect of a specific heavy-tailed distribution for the observation error.

2. BACKGROUND ON PARTICLE FILTERS

Our notation will generally 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 $\mathbf{x}$ to be a random variable. Our aim is then to estimate $p(\mathbf{x})$, the pdf of the state given all available observations.

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

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

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 $p(\mathbf{x})$, where the superscript f (for "forecast") indicates a prior quantity. The ensemble members are also known as particles. The update step makes the approximation of replacing the prior density $p(\mathbf{x})$ by a sum of delta functions, $N_e^{-1} \sum_{i=1}^{N_e} \delta(\mathbf{x} - \mathbf{x}_i^f)$. Applying Bayes rule yields

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

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 one of the likelihoods $p(\mathbf{y}^o|\mathbf{x}_i^f)$ is much larger than the rest, $\max_i w_i$ will be close to 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

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

may then be poor approximations. We will loosely term this situation, in which a single member 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 the rigorous asymptotic results of Bengtsson et al. (2007) for large N_y and N_e and through simulations informed by the asymptotics.

3. FAILURE OF THE PARTICLE FILTER IN A SIMPLE EXAMPLE

We next consider a simple example, in which the prior distribution $p(\mathbf{x})$ is Gaussian with each component of **x** independent and of unit variance, and the observations **y** are of each component of **x** individually with independent, Gaussian errors of unit variance. More concisely, consider $N_y = N_x$, $\mathbf{H} = \mathbf{I}$, $\mathbf{x} \sim N(0, \mathbf{I})$, and $\boldsymbol{\epsilon} \sim N(0, \mathbf{I})$, where the symbol $\sim$ means "is distributed as" and $N(\boldsymbol{\mu}, \mathbf{P})$ is the Gaussian distribution with mean $\boldsymbol{\mu}$ and covariance matrix **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, **x**, $\boldsymbol{\epsilon}$ 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 $\mathbf{x}^a = (\mathbf{x}^f + \mathbf{y}^o)/2$, where the superscript a (for "analysis") indicates a posterior quantity and the prior mean $\mathbf{x}^f = 0$ in this example. The expected squared error of $\mathbf{x}^a$ is $E(|\mathbf{x}^a - \mathbf{x}|^2) = [E(|\mathbf{x}^f - \mathbf{x}|^2 + E(|\mathbf{y}^o - \mathbf{x}|^2)]/4 = N_x/2$, while that of the observations [$E(|\mathbf{y}^o - \mathbf{x}|^2)$] 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 $\mathbf{x}^a$ only for N_x = 10. For N_x = 100, collapse of the weights is pronounced and $\hat{\mathbf{x}}^a$ is 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, which is given by $\sum w_i |\mathbf{x}_i^f - \hat{\mathbf{x}}^a|^2$. The correct posterior variance [tr(cov $\mathbf{x}|\mathbf{y}^o$)] is $N_x/2$, yet the particle-filter estimates (again averaged over 10^3 simulations) 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. We emphasize that this merely requires that the particle-filter estimate of the state is no worse than simply relying on the observations or the prior alone, i.e. that the particle filter "does no harm." 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.

The exponential dependence on N_e is also apparent in other aspects of the problem. Figure 3 shows the minimum N_e such that maximum w_i (averaged over 400 realizations) is less than a specified value. For each of the values 0.6, 0.7 and 0.8, the required N_e increases approximately exponentially with N_x.

4. BEHAVIOR OF WEIGHTS FOR LARGE N_y

The previous example highlights potential difficulties with the particle-filter update but does not permit more general conclusions. Results of Bengtsson 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 Bengtsson et al. for more rigorous and detailed proofs.

a. Approximation of the observation likelihood

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 component of the observation vector given the ith ensemble member. It is convenient to center and scale the argument of the ex-

ponent in (6) by defining

$$S_i = (\sum_{j=1}^{N_y} V_{ij} - \mu)/\tau, \tag{7a}$$

where

$$\mu \equiv \sum_{j=1}^{N_y} E\left(V_{ij}\right), \quad \tau^2 \equiv \text{var}\left(\sum_{j=1}^{N_y} V_{ij}\right). \tag{7b}$$

Then (6) becomes

$$p(\mathbf{y}^o|\mathbf{x}_i^f) = \exp\left(-\mu - \tau S_i\right), \tag{8}$$

where S_i has zero mean and unity variance. The simplest situation (as in the example of section 3) is when the random variables V_{ij}, $j = 1, \ldots, N_y$, are independent given $\mathbf{y}^o$, so that $\tau^2 = \sum_{j=1}^{N_y} \text{var}(V_{ij}|\mathbf{y}^o)$.

Equation (8) together with the approximation $S_i \sim N(0,1)$ are the basis for the asymptotic conditions for collapse derived in the next subsection. They allow statements about the asymptotic behavior of likelihood, and thus of the w_i, for large sample sizes N_e and large numbers of observations N_y, using asymptotic results for large samples from the standard normal distribution.

When V_{ij}, $j = 1, \ldots, N_y$, are independent given $\mathbf{y}^o$, the distribution of S_i approaches the standard Gaussian distribution for large N_y if the Lindeberg condition holds with probability tending to 1 (see Durret 2005, section 2.4a). More generally, the approximate normality of S_i holds for any observation error density $f(\cdot)$ such that $\int f^{1-\epsilon}(t)\, dt$ is finite for some $\epsilon > 0$ and when the V_{ij} are not i.i.d. but have sufficiently similar distributions and are not too dependent. [See Bengtsson et al. (2007).] We note in passing that the requirement that the V_{ij} be not too dependent as N_y increases means that N_x must become large as well and also that the compo-

nents of the state vector are not strongly dependent. We will return to the role of N_x in collapse later.

Showing that the approximation $S_i \sim N(0,1)$ is adequate for our purposes is nontrivial, since the behavior in the tails of the distribution is crucial to the derivations but convergence to a Gaussian is also weakest there. The interested reader will find details and proofs in Bengtsson et al. (2007). In what follows, however, we will assume that $S_i \sim N(0,1)$ holds in a fashion which makes succeeding manipulations valid.

b. Heuristic derivation of conditions for collapse

Using (8), the maximum weight $w_{(N_e)}$ can be expressed as

$$w_{(N_e)} = \left[1 + \sum_{i=2}^{N_e} \exp\left(-\tau(S_{(i)} - S_{(1)})\right)\right]^{-1}, \tag{9}$$

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

$$T = \sum_{i=2}^{N_e} \exp\left(-\tau(S_{(i)} - S_{(1)})\right), \tag{10}$$

we then have

$$w_{(N_e)} = 1/(1+T). \tag{11}$$

Collapse of the particle-filter weights occurs when T approaches zero.

In order to obtain asymptotic conditions for collapse, we next derive an expression for $E(T)$ for large N_e and N_y by approximating $E(T|S_{(1)})$ and then taking an expectation over the distribution of $S_{(1)}$. For an expectation conditioned on $S_{(1)}$, the sum in (10) may be replaced by a

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

sum over an unordered ensemble with the condition $S_i > S_{(1)}$. In that case the expectation of each term in the sum will be identical and

$$E(T|S_{(1)}) = (N_e - 1)E\left[\exp(-\tau(\tilde{S} - S_{(1)}))\right], \tag{12}$$

where $\tilde{S}$ is drawn from the same distribution as the S_i but with values restricted to be greater than $S_{(1)}$.

We now proceed with the calculation under the assumption that $S_i \sim N(0,1)$. Then $\tilde{S}$ has the density

$$p(z) = \begin{cases} \varphi(z)/\bar{\Phi}(S_{(1)}), & z > S_{(1)}, \\ 0, & z \leq S_{(1)}, \end{cases}$$

where $\varphi(\cdot)$ is the density for the standard normal distribution and $\bar{\Phi}(x) = \int_x^\infty \varphi(z)\,dz$.

Writing the expectation explicitly with the density of $\tilde{S}$ yields

$$E(T|S_{(1)}) = \frac{N_e - 1}{\Phi(S_{(1)})} \int_{S_{(1)}}^\infty \exp(-\tau(z - S_{(1)}))\varphi(z)\,dz. \tag{13}$$

Next, we replace $\varphi(z)$ by $(2\pi)^{-1/2}\exp(-z^2/2)$ in the integrand in(13), complete the square in the exponent and use the definition of $\bar{\Phi}(x)$ to obtain

$$E(T|S_{(1)}) = \frac{(N_e - 1)\exp\left(\tau S_{(1)} + \tau^2/2\right)\bar{\Phi}(\tau + S_{(1)})}{\Phi(S_{(1)})}. \tag{14}$$

The behavior of Gaussain order statistics, such as the minimum of a sample, are well known (David and Nagaraja 2003). An important result is that[2] , as $N_e \to \infty$,

$$S_{(1)} = -\sqrt{2\log N_e} + o_p(1). \tag{15}$$

Thus, since $S_{(1)}$ is becoming large and negative, $\bar{\Phi}(S_{(1)})$ approaches one and may be ignored in (14) when calculating the asymptotic behavior of $E(T|S_{(1)})$.

[2] A random variable $X = o_p(1)$ as $a \to \infty$ if $\Pr[|X| \geq \delta] \to 0$ for all $\delta \geq 0$

Now suppose that $\tau/\sqrt{\log N_e} \to \infty$ as $N_e \to \infty$. In this limit, $\tau + S_{(1)} \approx \tau(1 - \sqrt{2 \log N_e}/\tau) \to \infty$ and so, by the standard approximation to the behavior of $\bar{\Phi}$ for large positive values of its argument,

$$\bar{\Phi}(\tau + S_{(1)}) = \frac{\varphi(\tau + S_{(1)})}{\tau + S_{(1)}}(1 + o_p(1)), \tag{16}$$

which may be easily derived with integration by parts (e.g. section 6.3 of Bender and Orszag 1978).

Substituting (16) in (14), we conclude after some algebra that

$$E(T|S_{(1)}) \approx (N_e - 1)\frac{\varphi(S_{(1)})}{|S_{(1)}|}\frac{\sqrt{2 \log N_e}}{\tau}. \tag{17}$$

But, reversing the reasoning that led to (16) gives $\varphi(S_{(1)})/|S_{(1)}| \approx \Phi(S_{(1)})$, where $\Phi(x) = 1 - \bar{\Phi}(x)$ is the cumulative distribution function (cdf) for the standard Gaussian. Thus,

$$E(T|S_{(1)}) \approx N_e \Phi(S_{(1)})\frac{\sqrt{2 \log N_e}}{\tau}. \tag{18}$$

as $N_e \to \infty$.

Taking the expectation of (18) over $S_{(1)}$ then gives

$$E(T) = E(1/w_{(N_e)}) - 1 \approx \frac{\sqrt{2 \log N_e}}{\tau}. \tag{19}$$

To see this, recall that evaluating the cdf of a random variable at the value of the random variable, as in $\Phi(S_{(1)})$, yields a random variable with a uniform distribution on [0,1]. This property underlies the use of rank histograms as diagnostics of ensemble forecasts (Hamill 2001 and references therein) and is known in statistics as the "probability integral transform." Thus, $\Phi(S_{(1)})$ is distributed as the minimum of a sample of size N_e from a uniform distribution and $E[\Phi(S_{(1)})] \approx 1/N_e$. In the next section, we will confirm (19) with direct simulations.

Equation (19) implies that the particle filter will suffer collapse asymptotically if $N_e << \exp(\tau^2/2)$. More generally, N_e must increase exponentially with τ^2 in order to keep $w_{(N_e)}$ fixed as τ increases. This exponential dependence of N_e on τ^2 is consistent with the simulation results of section 3, where $\tau^2 \propto N_y$.

In contrast to the most obvious intuition, the asymptotic behavior of $w_{(N_e)}$ given in (19) does not depend directly on the state dimension N_x. Instead, the situation is more subtle: τ^2, a measure of the variability of the observation priors, controls the maximum weight. The dimensionality of the state enters only implicitly, via the approximation that S_i is asymptotically Gaussian, which requires that N_x be asymptotically large. One can then think of τ^2 as an equivalent state dimension, in the sense that τ^2 is the dimension of the identity-prior, identity-observation example (in section 3) that would have the same collapse properties.

5. 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 $\mathbf{x} \sim N(0, \mathbf{P})$ and the observation error $\boldsymbol{\epsilon} \sim N(0, \mathbf{R})$. We may assume that both $\mathbf{x}$ and $\boldsymbol{\epsilon}$ have mean zero since, if the observations depend linearly on the state, $E(\mathbf{y}) = \mathbf{H}E(\mathbf{x})$ and $p(\mathbf{y}|\mathbf{x})$ is unchanged if $\mathbf{y}$ is replaced by $\mathbf{y} - E(\mathbf{y})$ and $\mathbf{x}$ by $\mathbf{x} - E(\mathbf{x})$.

For Gaussian observation errors $\boldsymbol{\epsilon}$, the transformation $\mathbf{y}' = \mathbf{R}^{-1/2}\mathbf{y}$ also leaves $p(\mathbf{y}|\mathbf{x})$ unchanged but results in $\mathrm{cov}(\boldsymbol{\epsilon}') = \mathrm{cov}(\mathbf{R}^{-1/2}\boldsymbol{\epsilon}) = \mathbf{I}$. 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 and $p(\mathbf{y}|\mathbf{x})$ is again unchanged because $\mathbf{Q}$ is orthogonal. [Anderson (2001) presents a similar transformation that diagonalizes the problem in terms of the state variables, rather than the observation variables.] We therefore assume, without loss of generality, that

$$\mathbf{R} = \mathbf{I}, \quad \mathrm{cov}(\mathbf{H}\mathbf{x}) = \mathbf{H}\mathbf{P}\mathbf{H}^T = \mathrm{diag}(\lambda_1^2, \ldots, \lambda_{N_y}^2), \tag{20}$$

and drop primes in the sequel.

a. Analysis of the observation likelihood

With the assumptions (20), 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 $\mathbf{H}\mathbf{x}_i^f$. Thus,

$$V_{ij} = -\frac{1}{2}\left(y_j^o - (\mathbf{H}\mathbf{x}_i^f)_j\right)^2 + c.$$

The additive constant c results from the normalization of the Gaussian density and may be omitted w.l.o.g. since it cancels in the calculation of the weights w_i.

We wish to approximate the observation likelihood as in (8). This requires $\sum_{j=1}^{N_y} V_{ij}$ to be approximately Gaussian with mean μ and variance τ^2. Leaving aside for the moment the conditions under which the sum is approximately Gaussian, the mean and variance given $\mathbf{y}^o$ of $\sum_{j=1}^{N_y} V_{ij}$ can be calculated directly using (20) together with the properties of the standard normal distribution and the fact that the V_{ij} are independent as j varies [as in (7b)] . This yields

$$\mu = E\left(\sum_{j=1}^{N_y} V_{ij}\right) = -\frac{1}{2}\sum_{j=1}^{N_y}\left(\lambda_j^2 + y_j^{o2}\right), \tag{21a}$$

and

$$\tau^2 = \text{var}\left(\sum_{j=1}^{N_y} V_{ij}\right) = \sum_{j=1}^{N_y}\lambda_j^2\left(\frac{1}{2}\lambda_j^2 + y_j^{o2}\right). \tag{21b}$$

Equations (21) still depend on the specific realization $\mathbf{y}^o$ of the observations. Proceeding rigorously would require taking the expectation of (19) over $\mathbf{y}^o$. Here, we simply assume that expectation may be approximated by replacing τ in (19) by its expectation over $\mathbf{y}^o$. Using the fact that $E(y_j^{o2}) = \lambda_j^2 + 1$, we have

$$E(\mu) = -\frac{1}{2}\sum_{j=1}^{N_y}\left(2\lambda_j^2 + 1\right), \tag{22a}$$

and

$$E(\tau^2) = \sum_{j=1}^{N_y}\lambda_j^2\left(1 + \frac{3}{2}\lambda_j^2\right). \tag{22b}$$

As discussed in section 3 of Bickel et al. (2007), if $\lambda_1 \geq \lambda_2 \geq \ldots$, the distribution of $S_i = (\sum V_{ij} - \mu)/\tau$ converges to a standard Gaussian as $N_y \to \infty$ if and only if

$$\sum_{j=1}^{N_y}\lambda_j^2 \to \infty. \tag{23}$$

That is, S_i converges to a Gaussian when no single eigenvalue or set of eigenvalues dominate the sum of squares: (23) implies that $\max_j(\lambda_j)/\sum\lambda_j \to 0$ as $N_y \to \infty$. The condition (23) also means that $\tau^2 \to \infty$, which in turn leads to collapse if $\log N_e/\tau^2 \to 0$.

On the other hand, in the case that (23) is not satisfied, the unscaled log likelihood converges to a quantity which does *not* have a Gaussian distribution. Collapse does not occur since the updated ensemble empirical distribution converges to the true posterior as $N_e \to \infty$, whatever N_y may be.

b. Simulations

First, we check the asymptotic expression for $E(1/w_{(N_e)}) - 1$ as a function of N_e and N_y, given in (19), for the Gaussian-Gaussian case. For simplicity, let $\lambda_j = 1$, $j = 1, \ldots, N_y$ (as in the example of section 2). Then (22b) implies that $E(\tau^2) = 5N_y/2$ and (19) becomes

$$E(1/w_{(N_e)}) - 1 \approx \sqrt{4/5}\sqrt{\log(N_e)/N_y}. \tag{24}$$

This approximation is valid when N_e is large enough that the sample minimum follows (15) and N_y is large enough that $\log(N_e)/N_y$ is small. To capture the appropriate asymptotic regime, we have performed simulations with $N_e = N_y^\alpha$, $\alpha = 0.75, 0.875, 1.0, 1.25$, N_y varying over a dozen values between 600 and 3000, and $E(1/w_{(N_e)})$ approximated by averaging over 1000 realizations of the experiment. As can be seen from Fig. 4, $1 - E(1/w_{(N_e)})$ has an approximately linear relation to $\sqrt{\log(N_e)/N_y}$, though considerable scatter is present. The best-fit line to the simulation results has a slope of 0.96 with a 95% confidence interval of ± 0.087, which captures the predicted slope of $\sqrt{4/5} \approx 0.89$.

Equation (19) also implies that asymptotic collapse of the particle filter depends only on τ rather than the specific sequence $\{\lambda_j, j = 1, \ldots, N_y\}$. To illustrate that τ does control collapse, we consider various λ sequences by setting $\lambda_j^2 = cj^{-\theta}$. In this case, the simulations fix $N_y = 4 \times 10^3$ and $N_e = 10^5$ while θ takes the values 0.3, 0.5 and 0.7 and c is varied such that

substituting (22b) in (19) gives $0.01 < E(1/w_{(N_e)}) - 1 < 0.075$. These values are again chosen to capture the appropriate asymptotic regime where the normalized log-likelihood S_i is approximately Gaussian. The expectation $E(1/w_{(N_e)})$ is approximated by averaging over 400 realizations of the experiment.

Figure 5 shows results as a function of $\sqrt{2 \log N_e}/\tau$. As predicted by (19), $E(1/w_{(N_e)})$ depends mainly on τ rather than on the specific λ sequence . The simulations thus confirm the validity of (19) and, in particular, the control of the maximum weight by τ. Nevertheless, some limited scatter around the theoretical prediction remains, which arises from weak dependence of $E(1/w_{(N_e)})$ on the λ sequence for finite τ. We defer to a subsequent study a more detailed examination of the behavior of the maximum weight for finite τ and N_e and the limits of validity of (19).

6. MULTIVARIATE CAUCHY OBSERVATION-ERROR DISTRIBUTION

Van Leeuwen (2003) proposes the use of a multivariate Cauchy distribution for the observation error to avoid collapse and gives some numerical results supporting his claim. It is argued in Bengtsson et al. (2007) that collapse still occurs but more slowly with such an observation-error distribution. Specifically, they show collapse is formally governed by

$$\frac{\log N_e}{N_y} \log \left| \frac{\log N_e}{N_y} \right| \to \infty. \tag{25}$$

This condition is of course equivalent to $\log N_e/N_y \to 0$, since $te^{-t} \to 0$ if $t \to \infty$, but it suggests slow collapse.

Intiutively, what happens is that if $\boldsymbol{\epsilon}$ has a multivariate Cauchy distribution, then $\boldsymbol{\epsilon}$ can be written as

$$\boldsymbol{\epsilon} = |z_{N_y+1}|^{-1}(z_1, \ldots, z_{Ny}),$$

where $z_1, \ldots, z_{Ny}$ are i.i.d. $N(0, 1)$. For given z_{N_y+1} close to 0, the errors have very large scale Gaussian tails. This makes collapse harder because the true posterior resembles the prior, implying that the observations have relatively little information.

7. CONCLUSIONS

Particle filters have a well-known tendency for the particle weights to collapse, with one member receiving a posterior weight close to unity. We have illustrated this tendency through simulations of the particle-filter update for the simplest example, in which the priors for each of N_x state variables are i.i.d. and Gaussian, and the observations are of each state variable with independent, Gaussian errors. In this case, avoiding collapse and its detrimental effects can require very large ensemble sizes even for moderate N_x. The simulations indicate that the ensemble size N_e must increase exponentially with N_x in order for the posterior mean from the particle filter to have expected error smaller than either the prior or the observations. For N_x = 100, the posterior mean will typically be worse than either the prior or the observations unless $N_e > 10^6$.

Asymptotic analysis, following Bengtsson et al. (2007) and Bickel et al. (2007), provides precise conditions for collapse either in the case of i.i.d. observation likelihoods or when both the prior and the observation errors are Gaussian and the observation operator is linear. The asymptotic result holds when N_e is large and τ^2, the variance of the observation log-likelihood

defined in (7b), becomes large and has an approximately Gaussian distribution. Then, in the limit that $\tau^{-1}\sqrt{N_e} \to 0$, the maximum weight $w_{(N_e)}$ satisfies $E(1/w_{(N_e)}) \approx 1 + \tau^{-1}\sqrt{2 \log N_e}$. The maximum weight therefore approaches one (and collapse occurs) as τ increases unless the ensemble size N_e grows exponentially with τ.

It is thus not the state dimension per se that matters for collapse, but rather τ, which depends on both the variability of the prior and the characteristics of the observations. Still, one may think of τ^2 as an effective dimension, as it gives the dimension of the identity-prior, identity-observation Gaussian system (as in section 3) that would have the same collapse properties. This analogy is only useful, however, when the normalized observation log-likelihood S_i defined in (7a) has an approximately Gaussian distribution, which requires that N_x be large.

Our results thus point to a fundamental obstacle to the application of particle filters in high-dimensional systems. Nevertheless, some limitations of this study will need to be addressed before the potential of particle filtering in high dimensions is completely clear.

First, the simulations and asymptotic theory presented here have not dealt with the most general situation, namely, when the prior and observations are non-Gaussian and have non-trivial dependencies among their components. There is no obvious reason to expect that the general case should have less stringent requirements on N_e and we speculate that the Gaussian-Gaussian results of section 5 will still be informative even for non-Gaussian systems.

Second, the asymptotic theory pertains to the behavior of the maximum weight but says nothing about how the tendency for collapse might degrade the quality of the particle-filter update. Indeed, the update may be poor long before the maximum weight approaches unity, as illustrated by Figs. 2 and 3. What is needed is practical guidance on ensemble size for a given

problem with finite N_x, N_y and τ. Though rigorous asymptotic analysis will be difficult, we anticipate that simulations may provide useful empirical rules to guide the choice of ensemble size.

Third, we have not addressed the possible effects of sequentially cycling the particle filter given observations at multiple instants in time. Overall, cycling must increase the tendency for collapse of the particle filter. The quantitative effect, however, will depend on the resampling strategy, which again makes analytic progress unlikely.

Finally, we have not considered resampling algorithms, which are frequently employed to counteract the particle filter's tendency for collapse of the ensemble. We emphasize that resampling strategies that do not alter the update step are unlikely to overcome the need for very large N_e, since they do not improve the estimate of the posterior distribution but merely avoid carrying members with very small weights further in the algorithm. It is conceivable that the required N_e might be reduced by splitting a large set of observations valid at a single time into several batches, and then assimilating the batches serially with resampling after each update step. Alternatively, one might identify states in the past that will evolve under the system dynamics to become consistent with present observations, thereby reducing the need for large ensembles of present states when updating given present observations. Gordon et al. (1993) term this process "editing," and a similar idea is employed by Pitt and Shephard (1999). Such a scheme, however, would likely demand very large ensembles of past states.

As noted in the introduction, both van Leeuwen (2003) and Zhou et al. (2006) have applied particle filters to systems of dimension significantly larger than 100. In Zhou et al., however, each update is based on only a single observation (and only 28 observations total are assimilated); assuming that the prior uncertainty is comparable to the observation variance, $\tau^2 < 28$

in their case and their ensemble sizes of $O(1000)$ would be adequate based on Fig. 3. Based on the characteristics of the sea-surface height observations assimilated by van Leeuwen, we estimate that the particle-filter update uses $O(100)$ observations at each (daily) analysis. Allowing for the possibility that nearby observations are significantly correlated owing to the relatively large scales emphasized by sea-surface height, then van Leeuwen's use of 500–1000 ensemble members would seem to be at the edge of where our results would indicate collapse to occur. Consistent with this, van Leeuwen notes a strong tendency for collapse.

Fundamentally, the particle filter suffers collapse in high-dimensional problems because the prior and posterior distributions are nearly mutually singular, so that any sample from the prior distribution has exceptionally small probability under the posterior distribution. For example, in the Gaussian, i.i.d. case, the prior and posterior distributions have almost all their mass confined to the neighborhood of hyper-spheres with different radii and different centers.

Another way of looking at the cause of collapse is that the weights of different members for any chosen state variable are influenced by *all* observations, even if those observations are nearly independent of the particular state variable. The particle filter thus inherently overestimates the information available in the observations and underestimates the uncertainty of the posterior distribution. Similar problems occur for the EnKF and, for spatially distributed systems with finite correlation lengths (such as most geophysical systems), can be reduced by explicitly restricting any observation's influence to some spatially local neighborhod. This motivates the development of nonlinear, non-Gaussian ensemble assimilation schemes that perform spatially local updates, as in Bengtsson et al. (2003) or Harlim and Hunt (2007).

Acknowledgments. T. Hamill first introduced the lead author to the potential problems with the particle-filter update in high dimensions. This work was supported in part by NSF grant 0205655.

REFERENCES

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

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

Bender, M., and S. Orszag, 1978: *Advanced Mathematical Methods for Scientists and Engineers*. McGraw-Hill, pp. 593.

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

Bengtsson, T., P. Bickel and B. Li, 2007: Curse of dimensionality revisited: Collapse of the particle filter in very large scale systems. *IMS Lecture Notes—Monograph Series in Probability and Statistics: Essays in Honor of David A. Freedman*, D. Nolan and T. Speed, Eds., Institute of Mathematical Statistics, in press, 316–334.

Bickel, P., and E. Levina, 2007: Regularized estimation of large covariance matrices. *Ann. Statist.*, to appear.

Bickel, P., B. Li and T. Bengtsson, 2007: Sharp failure rates for the bootstrap filter in high dimensions. To appear in *IMS Lecture Notes—Monograph Series: Essays in Honor of J. K. Ghosh*, Institute of Mathematical Statistics, to appear.

David, H. A., and H. N. Nagaraja, 2003: *Order Statistics, Third Edition*, Wiley, 458 pp.

Doucet A., N. de Freitas, N. Gordon, 2001: An introduction to sequential Monte Carlo methods. *Sequential Monte Carlo Methods in Practice.* A. Doucet, N. de Freitas, and N. Gordon, Eds., Springer-Verlag, 2–14.

Durret, R., 2005: *Probability: Theory and Examples, Third Edition*, Duxbury Press, 512 pp

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.

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

Hamill, T. M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.,* **129**, 550-560.

Harlim, J., and B. R. Hunt, 2007: A non-Gaussian ensemble filter for assimilating infrequent noisy observations *Tellus*, **59A**, 225–237.

Keppenne, C. L., M. M. Rienecker, N. P. Kurkowski and D. A. Adamec, 2005: Ensemble Kalman filter assimilation of temperature and altimeterdata with bias correction and application to seasonal prediction. *Nonlinear Processes in Geophysics*, **12**, 491–503.

Liu, J. S., 2001: *Monte Carlo strategies in scientific computing.* Springer-Verlag, New York, pp. 364.

Moradkhani, H., K.-L. Hsu, H. Gupta and S. Sorooshian, 2005: Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. *Water Resources Res.,* **41**, W05012, doi:10.1029/2004WR003604.

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

Pitt, M. K., and N. Shephard, 1999: Filtering via simulation: Auxilliary particle filters. *J. Amer. Statistical Assoc.*, **94**, 590–599.

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

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, Y., 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.

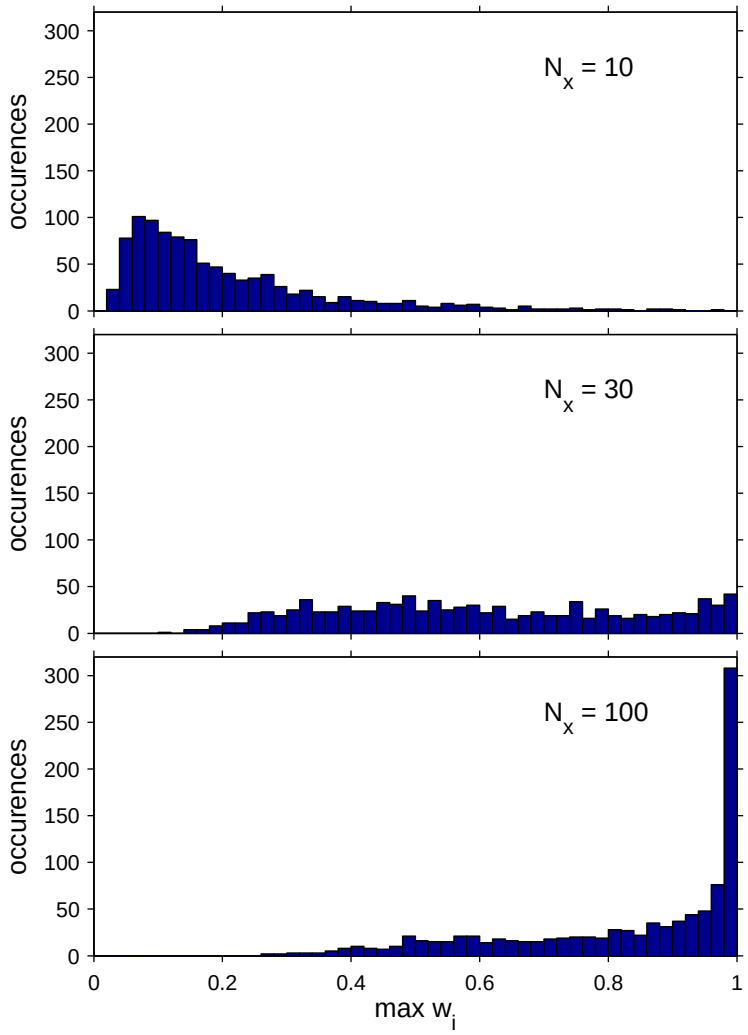

Figure 1. Histograms of max w_i for N_x = 10, 30, 100 and N_e = 10^3 from the particle-filter simulations described in text [N_e = 10^3, $\mathbf{x} \sim N(0, \mathbf{I})$, $N_y = N_x$, $\mathbf{H} = \mathbf{I}$ and $\boldsymbol{\epsilon} \sim N(0, \mathbf{I})$].

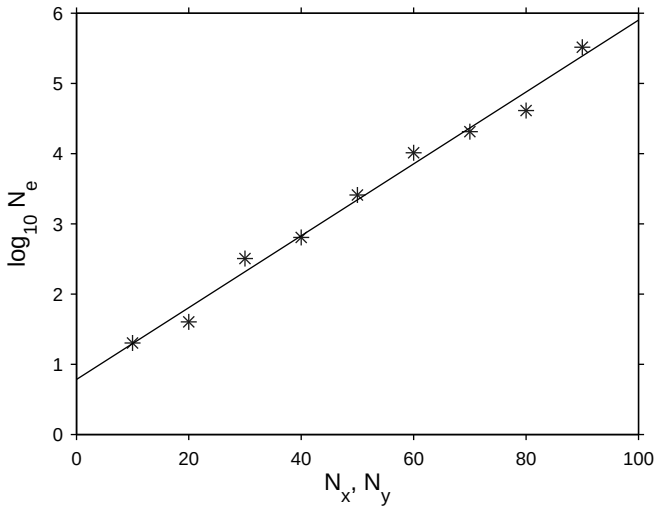

Figure 2. The ensemble size N_e as a function of N_x (or N_y) required if the posterior mean estimated by the particle filter is to have average squared error less than the prior or observations, in the simple example considered in the text. Asterisks show the simulation results, averaged over 400 realizations.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

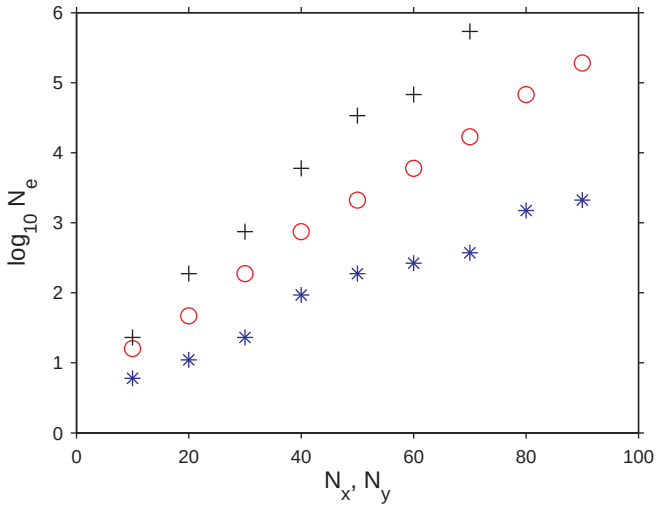

Figure 3. The ensemble size N_e as a function of N_x (or N_y) such that max w_i averaged over 400 realizations is less than 0.6 (plus signs), 0.7 (circles) and 0.8 (asterisks) in the simple example considered in the text.

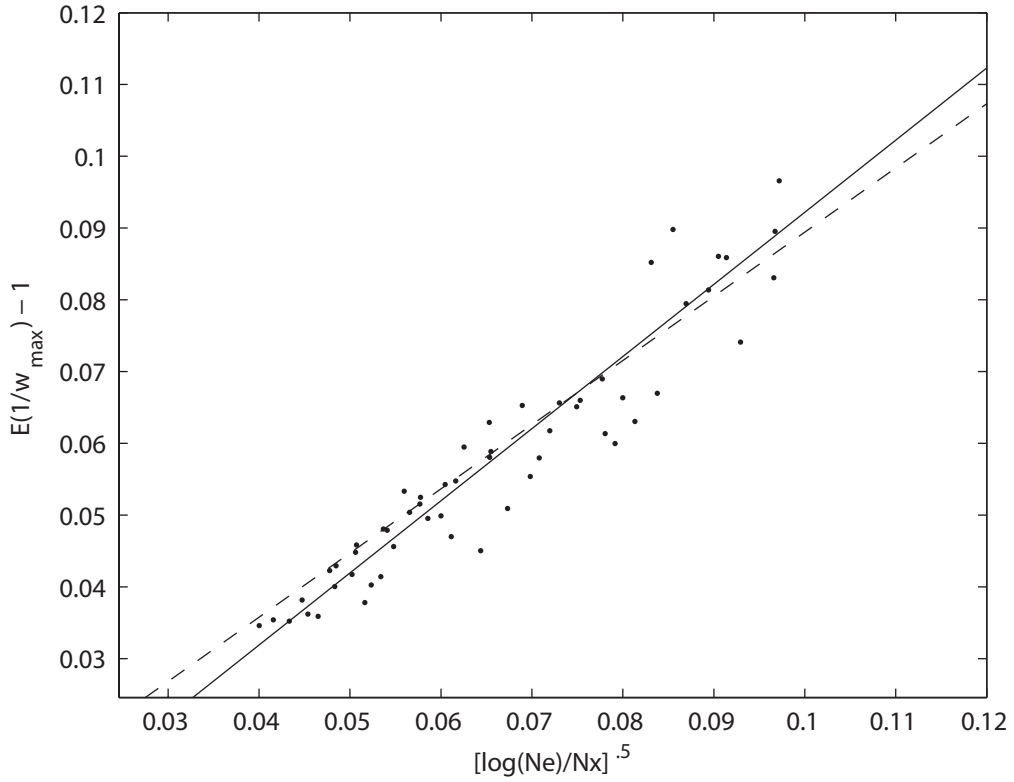

Figure 4. Evaluation of (19) against simulations in the case $\lambda_j^2 = 1$, $j = 1, \ldots, N_y$. For each of 60 (N_y, N_e) pairs as detailed in the text, $E(1/w_{(N_e)})$ was estimated from an average of 1000 realizations of the particle-filter update. The best-fit line to the data, given by $E(1/w_{(N_e)}) - 1 = -0.006 + 0.964\sqrt{\log(N_e)/N_y}$, is indicated by the solid line, while the prediction (24) is shown by a dashed line.

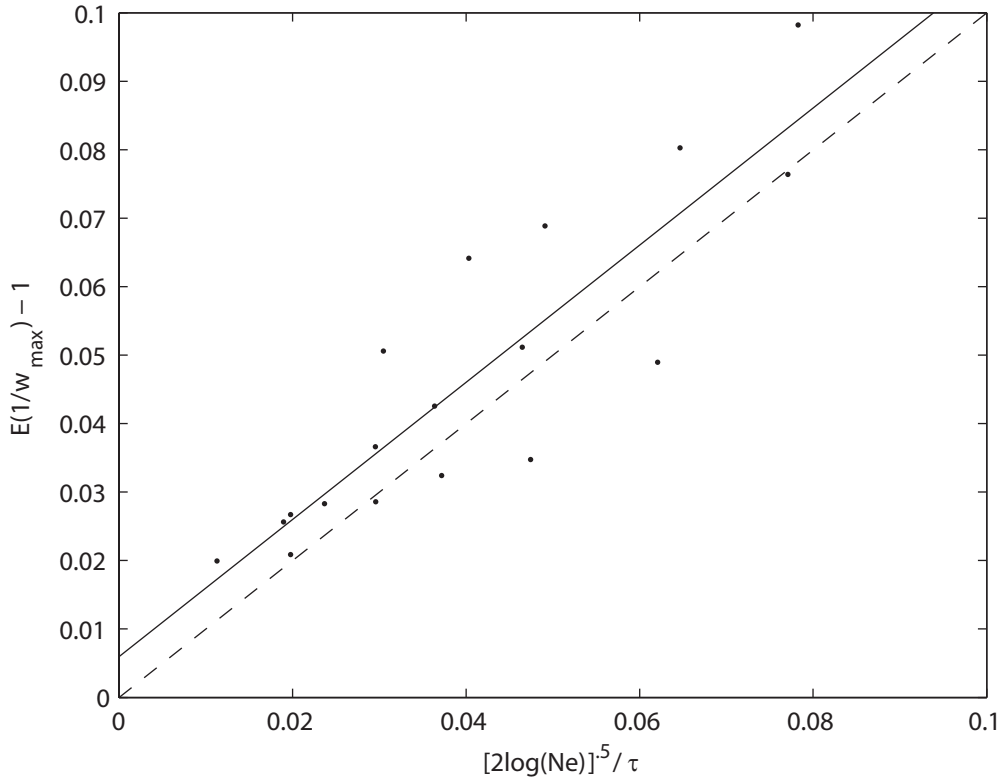

Figure 5. Evaluation of (19) against simulations in the case $\lambda_j^2 = cj^\theta$, $j = 1, \ldots, N_y$. The parameters θ and c are varied as described in the text, while $N_y = 4 \times 10^3$ and $N_e = 10^5$ are fixed. The expectation $E(1/w_{(N_e)})$ was estimated by averaging over 400 realizations of the particle-filter update. The best-fit line to the data, given by $E(1/w_{(N_e)}) - 1 = 0.006 + 1.008\sqrt{2 \log N_e}/\tau$ is indicated by the solid line, while the prediction (19) is shown by a dashed line.