Practical Estimates of the Ensemble Size Necessary for Particle Filters

Laura Slivinski, *

Woods Hole Oceanographic Institution, Woods Hole, Massachusetts

Chris Snyder

National Center for Atmospheric Research, Boulder, Colorado †

ABSTRACT

Particle filtering methods for data assimilation may suffer from the "curse of dimensionality," where the required ensemble size grows rapidly as the dimension increases. It would therefore be useful to know *a priori* whether a particle filter is feasible to implement in a given system. Previous work provides an asymptotic relation between the necessary ensemble size and an exponential function of τ^2, a statistic that depends on observation-space quantities and that is proportional to the system dimension when the number of observations is large; for linear, Gaussian systems, the statistic τ^2 can be computed from eigenvalues of an appropriately-normalized covariance matrix. Tests with a low-dimensional system show that these asymptotic results remain useful when the system is nonlinear, with either the standard- or optimal-proposal implementation of the particle filter. We also explore approximations to the covariance matrices that facilitate computation in high-dimensional systems, as well as different methods to estimate the accumulated system-noise covariance for the optimal proposal. Since τ^2 may be approximated using an ensemble from a simpler data-assimilation scheme, such as the ensemble Kalman filter, the asymptotic relations thus allow an estimate of the ensemble size required for a particle filter before its implementation. Finally, we demonstrate the improved performance of particle filters with the optimal proposal, relative to those using the standard proposal, in the same low-dimensional system.

1. Introduction

Ensemble methods have been used in a variety of geophysical estimation problems, including atmospheric applications, oceanography, and land surface systems. Recently there has been growing interest in particle filtering methods in particular, as these methods are better able to capture the nonlinearity inherent in many geophysical systems [e.g. the merging particle filter of Nakano et al. (2007), the equivalent-weights filter of Ades and van Leeuwen (2013), and the implicit particle filter (Morzfeld et al. 2012)]. At the same time, particle filters also tend to suffer from the "curse of dimensionality" where the required ensemble size grows very rapidly as the dimension increases. Thus, it would be useful to know *a priori* whether a particle filter is feasible to implement in a given system.

The curse of dimensionality is a well-known problem in density estimation, as Monte-Carlo estimation of high-dimensional probability densities demands notoriously large sample sizes (Silverman 1986). In a series of related papers, Bengtsson et al. (2008), Bickel et al. (2008), and Snyder et al. (2008) show the curse of dimensionality is also manifest in simple particle filters. They demonstrate that the required ensemble size scales exponentially with a statistic related, in part, to the system dimension and which may be considered as an effective dimension. In general, the particle filter allows a choice of proposal distribution from which particles are drawn. Snyder and Bengtsson (2015) (see also Snyder (2012)) showed that the exponential increase of the ensemble size with effective dimension also holds for particle filters using the optimal proposal (Doucet et al. 2001), which we will introduce in more detail in section 5.

We will consider particle filters based on both the proposals above. In the case examined by Bengtsson et al. (2008); Bickel et al. (2008); Snyder et al. (2008), the proposal is the transition distribution for the system dynamics, where new particles are generated by evolving particles from the previous time under the system dynamics. It yields the bootstrap filter of Gordon et al. (1993) and was termed the "standard" proposal by Snyder (2012); Snyder and Bengtsson (2015). The optimal proposal is of interest both because of its relation to the implicit and equivalent-weights particle filters and because it minimizes the degeneracy of weights, as shown in Snyder and Bengtsson (2015), and thereby provides a bound on the performance of other particle filters.

Our ultimate goal is to be able to determine whether a particle filter would be feasible to implement, given that we have statistics from, say, a working ensemble Kalman filter.

For the standard proposal, this is straightforward: the forecast step of ensemble forecasts provides a draw from the proposal and we simply need to compute weights based on the observation likelihood for each member. However, it is harder to use an existing ensemble to assess the feasibility of the particle filter based on the optimal proposal, since it is non-trivial to develop an algorithm to sample from this proposal [c.f. Morzfeld et al. (2012)]. An alternative is to utilize the results of Bengtsson et al. (2008); Bickel et al. (2008); Snyder et al. (2008); Snyder and Bengtsson (2015) that relate the behavior of the weights in the linear, Gaussian case to eigenvalues of certain covariance matrices which may be estimated from an existing ensemble. We evaluate the use of these results in more general nonlinear, non-Gaussian settings. Note that sampling error also presents an issue in applying these results; we investigate these effects and possible methods for overcoming them in this paper as well.

We note here that Chorin and Morzfeld (2013) have investigated a different, but related, effective dimension of a Gaussian data assimilation problem. In particular, they define a "feasibility criterion" to be the Frobenius norm of the steady state posterior covariance matrix (which can be exactly calculated in the linear, Gaussian regime.) While both studies explore potential limitations of particle filtering in high-dimensional systems, their criterion is based on bounding the total posterior error variance as a function of an effective dimension, whereas the studies of Snyder et al. (2008) and Snyder and Bengtsson (2015) quantify the relation between degeneracy of the particle-filter weights and an effective dimension.

The remainder of this paper is organized as follows. In Section 2, we review the ensemble Kalman filter and the particle filter and their respective implementations. Section 3 reviews the previous results of Snyder et al. (2008) regarding the limits of particle filters in high dimensional linear systems. Section 4 verifies the applicability of the results for linear, Gaussian systems and the standard proposal to nonlinear systems; this is specifically useful for understanding the similar extension needed for the optimal proposal. In Section 5, we consider the optimal proposal in a nonlinear system and discuss some of the difficulties that arise, in particular regarding additive model noise in nonlinear systems. Section 6 includes comparisons of performance of the standard and optimal proposals in a nonlinear system. Finally, Section 7 summarizes the results and draws conclusions.

2. Review of Ensemble Methods and Previous Results

Ensemble data assimilation methods approximate probability distributions using an ensemble of states, either weighted or unweighted. Two common ensemble methods are the ensemble Kalman filter (EnKF) and the particle filter. Generally, the traditional EnKF algorithm uses unweighted ensemble members, which are themselves updated when an observation becomes available. On the other hand, the particle filter uses a weighted ensemble. When an observation is available, the particles are drawn from a proposal distribution and then reweighted according to the observation.

In this section, we will first describe the setup and some notation, and then briefly review the standard and optimal proposal implementations of the particle filter as well as the ensemble Kalman filter.

a. Setup and Notation

Assume that our state of interest is given by $\mathbf{x}_k \in \mathbb{R}^{N_x}$, where k indexes time and N_x is the dimension of the state. We will additionally assume that model noise is added at integration time steps (of size dt) indexed by l below, while observations are available every N_t integration steps, indexed by k. Let the time between observations be denoted by $\Delta t = N_t \cdot dt$. The model evolution can be described as:

$$\mathbf{x}_{k,l} = m(\mathbf{x}_{k,l-1}) + \eta_{l-1},\ l = 1...N_t, \tag{1}$$

where $\mathbf{x}_{k,N_t} = \mathbf{x}_{k+1,0} := \mathbf{x}_{k+1}$ and observations are available for $\mathbf{x}_k$, $k = 1, \ldots, N_{obs}$. The η_j, whose dependence on k is suppressed for notational convenience, are i.i.d. random variables that represent the system noise and have a distribution to be specified later. Further define M_{stoch} to be the operator that takes $\mathbf{x}_{k,0}$ to $\mathbf{x}_{k,N_t}$. That is, M_{stoch} includes the nonlinear convolution of the noise between observation times:

$$\mathbf{x}_k = M_{stoch}(\mathbf{x}_{k-1}, \eta_0, \ldots, \eta_{N_t-1}). \tag{2}$$

In particular, note that the model noise is not additive at observation times when m is nonlinear. Next assume that we have linear, noisy observations of the state given by

$$\mathbf{y}_k = \mathbf{H}\mathbf{x}_k + \epsilon_k, \tag{3}$$

where $\mathbf{y}_k \in \mathbb{R}^{N_y}$ is the observation of dimension N_y and $\epsilon_k \sim \mathcal{N}(0, \mathbf{R})$ is the observation error. For the following methods, we denote unweighted ensembles of size N_e as $\{\mathbf{x}_k^i\}_{i=1}^{N_e}$, and weighted ensembles as $\{\mathbf{x}_k^i, w_k^i\}$. Finally, $\mathbf{y}_{k_1:k_2}$ will denote the concatenation of the observations from time t_{k_1} to time t_{k_2}.

b. Particle filter

The particle filter estimates the true Bayesian probability distributions using a weighted ensemble of states. When an observation is available at time t_k, we are interested in $p(\mathbf{x}_k|\mathbf{y}_{0:k}) \approx \sum_{i=1}^{N_e} w_k^i \delta(\mathbf{x}_k - \mathbf{x}_k^i)$. We will briefly review the derivation of the weight update based on importance sampling (Doucet et al. 2000; Snyder 2012; Snyder

and Bengtsson 2015), which more generally also includes dependence on the state at the previous time: $\mathbf{x}_{k-1}$. To this end, suppose we want to sample from the distribution $p(\mathbf{x}_k, \mathbf{x}_{k-1}|\mathbf{y}_{0:k})$, which is unknown. Instead we sample from a proposal distribution denoted $\pi(\mathbf{x}_k, \mathbf{x}_{k-1}|\mathbf{y}_{0:k})$, which we can choose, and then assign weights to each member of the sample. We will choose our proposal distribution to be of the form of the product $\pi(\mathbf{x}_{k-1})\pi(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{y}_k)$, which will allow for a recursive definition of the weights. This weighted ensemble is then a sample from the correct distribution, if the weights are given by

$$w_k^i \propto \frac{p(\mathbf{x}_k^i, \mathbf{x}_{k-1}^i|\mathbf{y}_{0:k})}{\pi(\mathbf{x}_k, \mathbf{x}_{k-1}|\mathbf{y}_{0:k})} \tag{4}$$

and normalized to sum to 1. Since we will always be conditioning on $\mathbf{y}_{0:k-1}$, we omit this term in what follows, and consider only $\mathbf{y}_k$.

1) Standard proposal

The simplest choice for a proposal density is the standard proposal, in which $\pi(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{y}_k)$ is chosen to be $p(\mathbf{x}_k|\mathbf{x}_{k-1})$. Then, noting that
$p(\mathbf{x}_k, \mathbf{x}_{k-1}|\mathbf{y}_k) \propto p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1})$,

$$w_k^i \propto \frac{p(\mathbf{y}_k|\mathbf{x}_k^i)p(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i)}{\pi(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i, \mathbf{y}_k)} w_{k-1}^i \tag{5}$$

$$\propto p(\mathbf{y}_k|\mathbf{x}_k^i) w_{k-1}^i \tag{6}$$

where the constant of proportionality is determined so that all weights sum to 1.

2) Optimal proposal

Doucet et al. (2000) discuss the so-called optimal proposal, which includes information about the previous state as well as the current observation: $\pi(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{y}_k) = p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{y}_k)$. In this case, the weights are updated according to:

$$w_k^i = p(\mathbf{y}_k|\mathbf{x}_{k-1}^i) w_{k-1}^i. \tag{7}$$

Sampling from this proposal is discussed in more detail in the appendix, but note that drawing from the optimal proposal and updating the weights are both more complicated in the case of the optimal proposal than the standard proposal. Despite this added computational effort, there are cases in which the optimal proposal has significant performance gain over the standard proposal, with the same number of particles (see Section 6 below.) Thus the optimal proposal may be more computationally tractable than the standard proposal in terms of the number of particles needed for an acceptable error level.

c. Ensemble Kalman filter

Evensen (1994) introduced the ensemble Kalman filter as an approximation of the Kalman filter which, like the particle filter, uses an ensemble of realizations of the system state to represent probability distributions. Unlike the particle filter, the ensemble Kalman filter uses an unweighted (or equally weighted) ensemble of states. Suppose the ensemble at time t_k is given by $\{\mathbf{x}_k^{f,i}\}_{i=1}^{N_e}$, where f stands for "forecast," and a will represent "analysis." If an observation is also available at time t_k, each ensemble member is updated according to

$$\mathbf{x}_k^{a,i} = \mathbf{x}_k^{f,i} - \mathbf{K}(\mathbf{y}_k - \mathbf{x}_k^{f,i} + \epsilon_k^i), \tag{8}$$

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

where $\epsilon_k^i \sim \mathcal{N}(0, \mathbf{R})$ and $\mathbf{P}^f$ is the ensemble covariance of the forecast:

$$\mathbf{P}^f = \frac{1}{N_e - 1}\sum_{i=1}^{N_e}\left(\mathbf{x}_k^{f,i} - \bar{\mathbf{x}}_k^f\right), \tag{10}$$

$$\bar{\mathbf{x}}^f = \frac{1}{N_e}\sum_{i=1}^{N_e}\mathbf{x}_k^{f,i}. \tag{11}$$

This is the so-called perturbed-observation formulation of the EnKF (Evensen 2003), in which each observation is viewed as a random variable. In the update step, we replace $\mathbf{y}_k$ with $\mathbf{y}_k + \epsilon_k$ where ϵ_k has the same statistics as the observation error noise. This formulation is shown to give the correct posterior covariance; otherwise, the covariance is overly tightened [see Burgers et al. (1998); Houtekamer and Mitchell (1998)].

The EnKF is a linear method, and thus will be suboptimal for problems that are significantly non-Gaussian even if N_e is large. However, for distributions which are close to Gaussian, the EnKF works well with relatively few ensemble members, though often it requires localization and inflation [see Houtekamer and Mitchell (1998, 2001); Anderson and Anderson (1999); Hamill et al. (2001)]. In the experiments in this paper, we use the perturbed observation formulation of the EnKF with covariance localization using the compactly supported correlation function of Gaspari and Cohn (1999) and a small but fixed inflation.

d. Review of Previous Asymptotic Results

Snyder et al. (2008) prove, in certain regimes, an exponential relationship between the variance of the observation log-likelihood and the inverse of the maximum weight. In the linear Gaussian case, this variance can be calculated as a sum of eigenvalues of an explicit function of covariances. First, we give some definitions.

Define the weight update factor in Equation (5) to be

$\tilde{w}_k^i$; that is,

$$\tilde{w}_k^i = \frac{p(\mathbf{y}_k|\mathbf{x}_k^i)p(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i)}{\pi(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i,\mathbf{y}_k)}. \tag{12}$$

Next define τ^2 to be the variance of the log of these factors conditioned on the observations:

$$\tau^2 = \mathrm{var}(\log(\tilde{w}_k^i)|\mathbf{y}). \tag{13}$$

Let w_{max} denote the maximum weight over the ensemble. Snyder et al. (2008) show that

$$E[1/w_{max}] - 1 \approx \frac{\sqrt{2\log N_e}}{\tau}, \tag{14}$$

under the following assumptions: first, that the observation errors are spatially and temporally independent; second, that N_y and N_x are large; third, that N_e and $\tau/\sqrt{\log N_e}$ are large, in the sense that $\tau/\sqrt{\log N_e} \to \infty$ as $N_e \to \infty$; and finally, that the distribution of $\log(\tilde{w}_k^i)$ over draws of $\mathbf{x}^i$ from the proposal is sufficiently close to Gaussian. The first three assumptions are easily verified and generally hold in the systems of interest to this work. The final assumption is less obvious, and below we investigate situations in which this assumption may not hold.

Snyder et al. (2008) apply these asymptotic results to the particle filter with the standard proposal, where $\tilde{w}_k^i = p(\mathbf{y}_k|\mathbf{x}_k^i)$. Snyder (2012); Snyder and Bengtsson (2015) note that similar arguments apply to the optimal proposal, where $\tilde{w}_k^i = p(\mathbf{y}_k|\mathbf{x}_{k-1}^i)$. Snyder and Bengtsson (2015) also note that the asymptotic theory developed for the optimal proposal provides bounds for the implicit particle filter, as the implicit particle filter reduces to the optimal proposal particle filter in the case where observations are linear and taken every step (Morzfeld et al. 2012). Although we will consider nonlinear, non-Gaussian systems, the asymptotics developed in the linear, Gaussian case are of interest here because they provide an explicit expression for τ^2. Additionally, in the linear, Gaussian case, the conditions for $\log \tilde{w}_k^i$ to be Gaussian (and thus for validity of the asymptotic theory) are straightforward.

We take the linear, Gaussian system to be

$$\mathbf{x}_k = \mathbf{M}\mathbf{x}_{k-1} + \gamma_k, \tag{15}$$

where $\gamma_k \sim \mathcal{N}(0, \mathbf{Q})$ and the observations are as defined in (3). Note that, for the linear Gaussian case, the dynamics are only written for observation times t_k, in contrast to (1).

With the standard proposal, we first need to calculate the eigenvalues λ_j^2 of the matrix

$$\mathbf{C}_s = \mathbf{R}^{-1/2}\mathbf{H}\mathrm{cov}(\mathbf{x}_\mathrm{k})\mathbf{H}^\mathrm{T}\mathbf{R}^{-1/2}. \tag{16}$$

Snyder et al. (2008) and references therein derive the relation

$$E(\tau^2) = \sum_{j=1}^{N_y} \lambda_j^2 \left(1 + \frac{3}{2}\lambda_j^2\right), \tag{17}$$

where the expectation is taken over $\mathbf{y}_k$. Moreover, Bickel et al. (2008) show that $\log \tilde{w}_k^i$ is asymptotically Gaussian (over draws from the proposal), and the relation (14) is valid, as long as no eigenvalue(s) dominate the sum of squares above. In the case of the optimal proposal, the same expression (17) and the same conditions for validity hold, but using the eigenvalues of

$$\mathbf{C}_o = (\mathbf{R}+\mathbf{H}\mathbf{Q}\mathbf{H}^T)^{-1/2}\mathbf{H}\mathbf{M}\mathrm{cov}(\mathbf{x}_{\mathrm{k}-1})\mathbf{M}^\mathrm{T}\mathbf{H}^\mathrm{T}(\mathbf{R}+\mathbf{H}\mathbf{Q}\mathbf{H}^\mathrm{T})^{-1/2}. \tag{18}$$

In a system where each degree of freedom is independent and independently observed, these expressions simplify and show that τ^2 will be proportional to the number of observations. A similar but more informal derivation of this result also appears in Ades and van Leeuwen (2013).

3. Model and Experimental Setup

In all experiments in this paper, we will restrict our attention to the nonlinear dynamical system of Lorenz (1996). The deterministic form of these equations is given by:

$$\frac{dx_{(j)}}{dt} = (x_{(j+1)} - x_{(j-2)})x_{(j-1)} - x_{(j)} + F, \tag{19}$$

for $j = 1, \ldots, N_x$ and $F = 8$ here. The subscripts (j) indicate the spatial location in a one-dimensional, periodic domain and should be understood $\mod N_x$.

We solve a discrete-time, stochastic version of this equation, cast in the form (1). Fixing an integration time step dt and an observation time step $\Delta t = N_t dt$, we compute $m(\mathbf{x}_{k,l-1})$ by integrating (19) over a single time step dt using a fourth-order Runge-Kutta scheme and draw η_{l-1} from $\mathcal{N}(0, dt\sigma_{sys}^2\mathbf{I})$. Except where noted below, all results employ $dt = 0.01$. Alternatively, we could have started from a continuous time stochastic differential equation by including noise directly in Eqn. (19); this distinction is not crucial to any of the results we present. The observing network in the experiments will consist of full observations, so that $\mathbf{H} = \mathbf{I}$, and the observation error covariance is $\mathbf{R} = \sigma_{obs}^2\mathbf{I}$.

We will need an example ensemble DA scheme to calculate the statistics necessary to test the asymptotic theory. Since our goal is to demonstrate that these statistics may be used in practice to determine the applicability of the particle filter, we will use the EnKF, a common method for high-dimensional problems, to calculate the statistics. While the EnKF is suboptimal in the nonlinear case, we wish to show that the method is reasonably effective across a wide range of parameters for this system. The spread-skill relation (Table 1) indicates that this is true. For this experiment, we fix $N_x = 100$, let the system noise be fixed with $\sigma_{sys} = 0.01$, and vary the observation error noise σ_{obs}^2. For each value of observation error, we run the EnKF for 200 sequential observations. Table 1 shows the forecast mean squared error and forecast variance over the last 190 observations, for each value of observation error. For these

results, $\Delta t = 0.1$. As the results show, the EnKF is working well with the chosen values of inflation and localization, since the forecast mean squared error (MSE) and variance are comparable and neither blows up.

TABLE 1. Forecast error and variance of the working EnKF, with varying values of observation noise.

σ^2_{obs}	forecast MSE	avg forecast var
0.0001	0.0021	0.0011
0.0005	0.0024	0.0014
0.001	0.0027	0.0018
0.003	0.0037	0.0027
0.005	0.0044	0.0035
0.007	0.0052	0.0041
0.009	0.0057	0.0048
0.02	0.0082	0.0075
0.05	0.0145	0.0142
0.1	0.0236	0.0243
0.5	0.0736	0.0830
1	0.1448	0.1695

In Section 6, we will run a sequential particle filter for many observations to compare the overall performance of different proposal algorithms. In this case, we will need to resample in order to prevent weight collapse: here, we test two different resampling thresholds. The first is the resampling threshold defined in Kong et al. (1994), in which the filter is set to resample when the effective sample size $N_{\text{eff}} = 1/\sum_{i=1}^{N_e}(w^i)^2$ falls below a fixed ensemble size N_t. This is the threshold suggested by Arulampalam et al. (2002). The second threshold is based on the maximum weight, in which the filter resamples when $w_{\max}$ exceeds a certain value (here we use 0.5.) We then use a Monte Carlo Metropolis-Hastings resampling technique [see (Hastings 1970; Robert and Casella 2004) for an introduction and (van Leeuwen 2009) for a description applied to particle filters], followed by resetting the weights to be equal.

4. Extension to Nonlinear Case: Standard Proposal

Our goal is to show how to use an existing DA ensemble to determine whether it would be feasible to use a particle filter for a given nonlinear system, and if so, how many particles would be necessary to avoid filter collapse. In the case of the standard proposal, it is straightforward to directly calculate the weights without implementing the particle filter and quantify the statistics of the maximum weight directly. Alternatively, if we know $\mathbf{R}$ and $\text{cov}(\mathbf{x}_k)$, we could use Eqn. (17) to estimate τ^2 and then predict $E(1/w_{max})$ from Eqn. (14). This alternative approach to predict the behavior of w_{max} is especially useful in the case of the optimal proposal, where computing the weights directly requires sampling from the optimal proposal, which can be difficult. Thus we first numerically demonstrate the theory in the simpler case with the standard proposal, but with a nonlinear model, before moving on to the optimal proposal. Note that the asymptotics have been verified numerically for linear, Gaussian systems in Snyder et al. (2008).

We consider the Lorenz (1996) equations with $N_x = 100$, fix the system noise as $\sigma_{sys} = 0.01$, and vary the observation error variance σ^2_{obs}. The existing DA scheme we use is the EnKF as described in Section 3.

First, to demonstrate the degree of nonlinearity in this system of equations, we study the difference in perturbations after evolving two initial points forward under the fully nonlinear Lorenz equations as well as a linearized system. Specifically, we choose a random observation time in the EnKF experiment, choose two random ensemble members as our initial perturbation, and linearize the system about one of them. We evolve each ensemble member under the linearized dynamics to get $\{\mathbf{x}^i_{lin}\}_{i=1,2}$ and under the full dynamics to get $\{\mathbf{x}^i_{full}\}_{i=1,2}$; we then measure the linearity of the system with

$$err = \frac{\left| |\mathbf{x}^1_{full} - \mathbf{x}^2_{full}| - |\mathbf{x}^1_{lin} - \mathbf{x}^2_{lin}| \right|}{|\mathbf{x}^1_{full} - \mathbf{x}^2_{full}|}, \qquad (20)$$

where $|\mathbf{x}^1 - \mathbf{x}^2| = \left[\sum_{j=1}^{N_x} \left(x^1_{(j)} - x^2_{(j)}\right)^2\right]^{1/2}$. This will be close to 0 if the full system is close to linear. Additionally, note that this traditional version of the Lorenz (1996) equations with $F = 8$ has a doubling time of 2.1 days, where one model time unit corresponds to 5 days; thus, the doubling time is 0.42 model time units. Table 2 shows results with a fixed $dt = 0.005$, and variable integration time Δt, averaged over 100 randomly chosen observation times. Note that in the experiments in this paper, we vary the time between observations as $\Delta t = 0.1$ (standard proposal experiment) or $\Delta t = 0.4$ (optimal proposal experiment, following section.)

The results in Table 2 show that the measure of nonlinearity is very close to 0 for a single integration step, but quickly increases for longer time windows. This implies that the system is well-approximated by a linear model after just one integration step, but the nonlinearity increases as the length of integration increases. We have therefore chosen observation frequencies for the following experiments which guarantee nonlinear behavior of the model between observations in order to test the theory. Additionally, note that although we are operating in a regime where the system is fully observed, based on theory, we expect the same results to hold in the more realistic situation of inhomogeneous spatial observation coverage. In particular, fewer observations will lead to more strongly non-Gaussian

TABLE 2. Measure of nonlinearity of the Lorenz 96 system, for varying lengths of time.

Δt	0.005	0.05	0.1	0.2	0.4	0.8	1.0
σ^2_{obs}							
5e-5	0.011	0.009	0.018	0.002	0.028	0.298	0.755
1e-4	0.006	0.012	0.011	0.003	0.009	0.346	0.336
5e-4	0.006	0.019	0.004	0.019	0.007	0.362	1.419
1e-3	0.001	0.001	0.006	0.025	0.012	0.305	0.526
3e-3	0.002	0.007	0.006	0.002	0.011	0.381	1.279
5e-3	0.003	0.001	0.011	0.013	0.067	0.433	0.815
7e-3	0.0002	0.005	0.009	0.0002	0.035	0.460	0.950
9e-3	0.003	0.011	0.007	0.003	0.044	0.469	0.812
0.02	0.001	0.005	0.001	0.012	0.070	0.442	1.305
0.05	0.001	0.004	0.001	0.023	0.103	0.656	1.244

probability distributions; however, we are testing the effects of non-Gaussian distributions by ensuring the time between observations is long enough to display nonlinear behavior.

Next, to test the asymptotic theory on the calculation of τ^2 and its relationship to w_{max}, we run the EnKF with a localization radius of 5 and a covariance inflation of 1.05 on the Lorenz equations with 100 variables for 3000 sequential observations; at each observation time, before the EnKF analysis, we calculate what the true maximum weight would be if we were implementing the particle filter. We also calculate τ^2 using the approximation defined in the linear case. In order to have an accurate estimation of the covariance matrices, we run the EnKF with a large number of ensemble members ($N_{e,cov} = 1000$) to estimate the covariances, then draw a smaller ensemble ($N_e = 100$) to calculate the weights directly. The ensemble size N_e is then used in the term $(2\log(N_e))^{1/2}/\tau$ in the numerics. In this experiment, we fix $\Delta t = 0.1$, $N_x = 100$, and system noise $\sigma_{sys} = 0.01$, and vary the observation error σ^2_{obs} from 5×10^{-5} to 0.05. Note that varying the observation error leads to different values of τ, and thus different data points, since the estimate of τ^2 involves the eigenvalues of a matrix proportional to $\mathbf{R}^{-1}$. Thus, small values of σ^2_{obs} lead to larger values of τ and will result in ensembles that are close to collapse. Intuitively, this can be understood by thinking about a one-dimensional case: if the variance of the observation likelihood is very small, then the support of the probability distribution is very narrow, and all particles except the one closest to the observation will have very small weight.

Figure 1 shows the results of the asymptotic theory of collapse, where each data point is averaged over the last 2990 steps of the filter and the error bars represent 95% confidence intervals. Note that the observation error is increasing as we move in the positive x-axis direction. Results using the full covariance to calculate the eigenvalues are given in blue. They agree well with the theory in the regime near the origin, where the theory is formally valid, but deviations from the theory increase as $(\log N_e)^{1/2}/\tau$ increases.

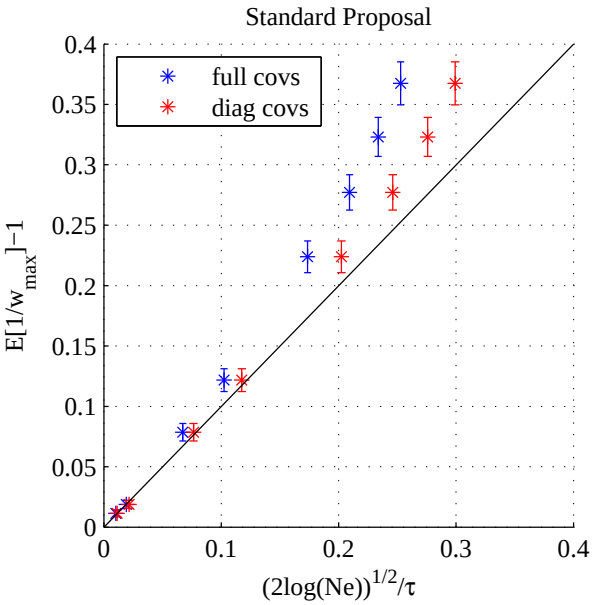

FIG. 1. Numerical estimation of $(2\log(N_e))^{1/2}/\tau$ versus the time average and 95% confidence interval of $E[1/w_{max}]-1$ calculated using the standard proposal. Blue represents calculating τ from the true eigenvalues, and red represents calculations based on the diagonal entries of the matrix. The black line represents the theoretical relationship between $(2\log(N_e))^{1/2}/\tau$ and $E[1/w_{max}] - 1$.

There are several additional reasons for deviation from the theory within the asymptotic regime as well. In particular, the theory relies on the assumption that $\{\log \tilde{w}^i_k\}$ is an approximate sample from a Gaussian distribution. This assumption is satisfied provided $\log \tilde{w}^i_k$ is a sum of a large number of sufficiently independent random variables. However, evolving the ensemble under the state dynamics generally concentrates the state variance into a few growing structures, which increases spatial correlations and makes observation-space quantities more dependent. This leads to $\log \tilde{w}^i_k$ effectively being a sum over fewer independent random variables, which (all else being equal) in turn leads to $\log \tilde{w}^i_k$ being less Gaussian.

To test whether the non-Gaussian nature of $\log \tilde{w}^i_k$ may be a factor in the deviation of the numerics from the theory, we also investigate the degree to which $\log \tilde{w}^i_k$ may be skewed in this system. In particular, we look at the skewness of the ensembles of $\log \tilde{w}^i_k$ for the standard proposal experiment in this section; results are given in Table 3. If

the skewness is far from 0, then the sample distribution is far from Gaussian. The skewnesses are averaged over the last 2990 observations. As the results in Table 3 show, larger observation error generally leads to higher skewness values; since large observation error corresponds to larger $2\log N_e/\tau^2$, this may explain why the data points do not follow the theory as well further from the asymptotic regime. On the other hand, the observed increase in skewness is not very strong, and thus may not be the only cause of the deviation between numerics and theory. However, the frequent changes of variables needed to derive τ^2 prevent a more detailed analysis of this deviation.

TABLE 3. Skewness of the ensembles of $\log \tilde{w}_k^i$ after one evolution under the Lorenz 96 model for varying magnitude of observation error; standard proposal experiment. Mean and 95% confidence intervals over final 2990 time steps.

σ^2_{obs}	mean skewness
5e-5	0.251 ± 0.009
1e-4	0.252 ± 0.009
5e-4	0.271 ± 0.009
1e-3	0.270 ± 0.009
3e-3	0.281 ± 0.009
5e-3	0.290 ± 0.009
7e-3	0.302 ± 0.009
9e-3	0.300 ± 0.009
0.02	0.320 ± 0.009

In practice, there are difficulties using Eqn. (17) to estimate τ^2. First, computing a covariance matrix from a small sample typically yields an eigenvalue spectrum that is artificially steep, with too much variance in leading directions. The corresponding calculation of τ^2 will then be too large, since it is a sum of higher powers of the eigenvalues. We have therefore chosen a large ensemble ($N_e \geq N_x$) in this experiment in order to estimate the covariances accurately and avoid this effect. Second, for large numbers of observations and large ensembles, calculating eigenvalues of these matrices may be computationally prohibitive.

Thus, we also tested this theory using a computationally feasible approximation for the eigenvalues of $\mathbf{R}^{-1/2}\mathbf{H}\mathrm{cov}(\mathbf{x}_k)\mathbf{H}^T\mathbf{R}^{-1/2}$: we assume $\mathbf{R}$ and $\mathbf{H}\mathrm{cov}(\mathbf{x}_k)\mathbf{H}^T$ are diagonal, so that the eigenvalues are simply the product of the corresponding diagonal elements of $\mathbf{R}^{-1}$ and $\mathbf{H}\mathrm{cov}(\mathbf{x}_k)\mathbf{H}^T$. Results with τ^2 approximated in this way are also shown in Fig. 1. The approximation systematically underestimates τ^2 data points with the approximation always lie to the right of those using eigenvalues of the full matrix (16). [1] Nevertheless, using the approximation of τ^2 in the asymptotic relation gives reasonable predictions of $E(1/w_{max})$, often better than with the unapproximated τ^2, because the underestimation by the diagonal approximation compensates for the overestimation of $E(1/w_{max})$ that is, empirically, a property of the asymptotic relation when $(2\log N_e)^{1/2}/\tau$ is not small. It is not clear whether this compensation will be equally effective in other problems.

[1] Since τ^2 is a sum of squares of the eigenvalues of (16) [see (17)] and because the sum of the eigenvalues equals the sum of the diagonal elements, τ^2 will be underestimated by the diagonal approximation whenever the eigenvalue spectrum is steeper than the sorted list of diagonal elements. We expect this to be true in many problems involving spatial correlations, with spatially correlated but nearly spatially homogeneous processes being a prime example.

5. Optimal Proposal

Next, we follow the approach of the previous section, but apply the asymptotic theory to the optimal proposal. Specifically, we wish to use an existing ensemble to evaluate the feasibility of a particle filter using the optimal proposal. As in the case of the standard proposal above, the evaluation will be limited by the fact that it applies results from linear, Gaussian systems in a nonlinear, non-Gaussian setting, and by sampling errors in estimating the necessary covariance matrices from a finite ensemble. We will check these limitations with numerical simulations using the Lorenz (1996) system. For the optimal proposal, there is also an additional issue, in that some of the covariance matrices involved in the definition (18) of $\mathbf{C}_o$ do not appear explicitly in the nonlinear problem. We turn to this issue first.

a. Model noise in nonlinear systems

The matrix $\mathbf{C}_o$, whose eigenvalues determine τ^2 for the optimal proposal via (17) in the linear, Gaussian case, involves the covariance matrices $\mathbf{HM}\mathrm{cov}(\mathbf{x}_{k-1})\mathbf{M}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}$ and $\mathbf{HQH}^T$. Since the equations (1)-(3) for the nonlinear system do not specify these quantities, we take the approach of defining them through more general expressions that reduce to the correct result for the linear, Gaussian case.

To compute the covariance involving the linear dynamics $\mathbf{M}$, we first define $M_{det}(\mathbf{x}) = M_{stoch}(\mathbf{x}, 0, \ldots, 0)$ [recalling that M_{stoch} in (2) is a function of the state $\mathbf{x}$ as well as the realizations of the noise at each integration step $\eta_0, \ldots, \eta_{N_t-1}$]. In the linear case, $M_{det}(\mathbf{x}) = \mathbf{M}\mathbf{x}$ and a more general definition for the desired covariance is

$$\mathbf{HM}\,\mathrm{cov}(\mathbf{x}_{k-1})\mathbf{M}^T\mathbf{H}^T = \mathrm{cov}(\mathbf{H}M_{det}(\mathbf{x}_{k-1})). \qquad (21)$$

We estimate the right hand side for the nonlinear system by evolving an ensemble of initial conditions from t_{k-1} using M_{det}, applying $\mathbf{H}$ and computing the sample covariance.

For the covariance $\mathbf{HQH}^T$, there are at least two possible definitions that generalize to the nonlinear system. The first uses

$$\mathbf{HQH}^T = \mathrm{cov}(\mathbf{H}\mathbf{x}_k|\mathbf{x}_{k-1}), \qquad (22)$$

which is an identity in the linear case and gives a quantity that, in the nonlinear case, will depend on $\mathbf{x}_{k-1}$. We can estimate the covariance on the right hand side by starting from a given $\mathbf{x}^i_{k-1}$ and computing an ensemble $M_{stoch}(\mathbf{x}^i_{k-1}, \eta_0, \ldots, \eta_{N_t-1})$ over realizations $\eta_0, \ldots, \eta_{N_t-1}$ of the system noise. Let $\mathbf{Q}^i$ be the state-space covariance estimated in this way. (Recall from Section 3 that $\mathbf{H} = \mathbf{I}$ in our experiments.) A further step would be to compute $\bar{\mathbf{Q}}$ by averaging the $\mathbf{Q}^i$ over an ensemble of $\mathbf{x}^i_{k-1}$.

The second possible definition relies on

$$\mathbf{HQH}^T = \mathrm{cov}(\mathbf{H}M_{stoch}(\mathbf{x}_{k-1}, \eta_0, \ldots, \eta_{N_t-1}) - \mathbf{H}M_{det}(\mathbf{x}_{k-1})). \tag{23}$$

This expression is again an identity in the linear case – $\mathbf{Q}$ can be written as the sum over contributions from the noise in (1) at each of the N_t model time steps between t_{k-1} and t_k. Beginning from an ensemble of realizations of $\mathbf{x}_{k-1}$, we estimate the covariance on the right hand side above by evolving each member from t_{k-1} to t_k with both M_{det} and M_{stoch}, with independent realizations of the system noise in M_{stoch}, and then taking the sample covariance of the differences in $\mathbf{x}_k$. We denote this estimate $\tilde{\mathbf{Q}}$.

It is not immediately obvious whether one of these definitions is to be preferred. They will agree in the limit of linear dynamics and may differ as nonlinearity increases. We have therefore explored the behavior of both approaches in the case with $N_x = 100$, $\Delta t = 0.1$, and with varying model noise σ_{sys} and initial ensemble size σ_{ens}. The test consists of evolving the particles forward from time 0 to time Δt and estimating $\mathbf{Q}$ in the three ways described above. First, we calculate $\mathbf{Q}^i$ for each particle; second, we take the average $\bar{\mathbf{Q}}$ of these $\mathbf{Q}^i$s; finally, we estimate $\tilde{\mathbf{Q}}$ as above. We found that the variations of $\mathbf{Q}^i$ about $\bar{\mathbf{Q}}$ were negligible relative to the magnitude of elements of $\bar{\mathbf{Q}}$. Similarly we found good agreement between $\bar{\mathbf{Q}}$ and $\tilde{\mathbf{Q}}$ in these cases. Thus, the effects of nonlinearity in estimating the effective model noise covariance are small in these experiments; in particular, they are much smaller than sampling error in estimates of $\mathbf{Q}$ with ensembles of size 100, which we use in the following experiments.

The two definitions do, however, differ substantially in their computational demands, as the computation of the $\mathbf{Q}^i$ and $\bar{\mathbf{Q}}$ requires an ensemble of integrations for each $\mathbf{x}^i_{k-1}$, while a single integration for each $\mathbf{x}^i_{k-1}$ suffices for $\tilde{\mathbf{Q}}$. In all following experiments, we therefore use $\tilde{\mathbf{Q}}$ to estimate the model noise covariance, as it is the most computationally efficient.

b. Numerical Results

Snyder and Bengtsson (2015) have rigorously shown that the asymptotics developed in Bengtsson et al. (2008); Snyder et al. (2008) also hold for the optimal proposal. Here, we numerically show how these results extend to the nonlinear system of Lorenz (1996). As in the experiment with the standard proposal, we run the EnKF with a localization radius of 5 and a covariance inflation of 1.05 on the Lorenz equations with 100 variables for 3000 sequential observations. We fix $\Delta t = 0.4$ and the system noise $\sigma_{sys} = 0.01$, and vary the observation error σ^2_{obs} from 5×10^{-3} to 1. In this experiment, we use the approximation $\tilde{\mathbf{Q}}$ described above when calculating both τ and the exact weights. The size of the ensemble used to calculate $\tilde{\mathbf{Q}}$ is $N_{e,cov} = 1000$, but we take a subsample of size $N_e = 100$ when calculating the weights themselves (and, as above, use $N_e = 100$ in the theoretical value $(2\log(N_e))^{1/2}/\tau$.) We approximate sampling from the optimal proposal by sampling from the distribution derived for the linear, Gaussian case given in Equation (A3), replacing all $\mathbf{Q}$'s with $\tilde{\mathbf{Q}}$.

The results are given in Figure 2. Clearly, the data

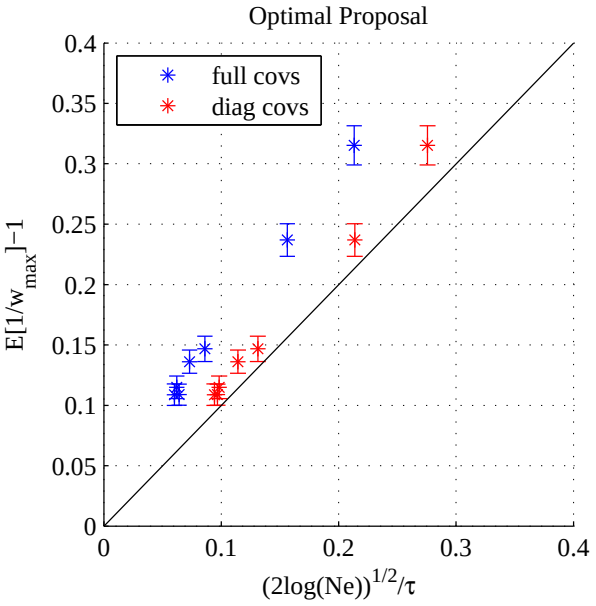

FIG. 2. Numerical estimation of $(2\log(N_e))^{1/2}/\tau$ versus the time average and 95% confidence interval of $E[1/w_{max}] - 1$ calculated using the optimal proposal, with approximations as described in the text. Blue represents calculating τ from the true eigenvalues, and red represents calculations based on the diagonal entries of the matrix. The black line represents the theoretical relationship between $(2\log(N_e))^{1/2}/\tau$ and $E[1/w_{max}] - 1$.

points do not agree with the theory as well as in the experiment with the standard proposal. This is likely due to the parameter choices in this experiment. When using the optimal proposal, we empirically found that we needed to increase the time between observations in order to satisfy the assumption that the filter is close to collapse (that is, that $2\log(N_e)/\tau^2$ is close to 0.) However, as mentioned previously, this also leads to a steep spectrum of

the covariance matrices, which in turn leads to violation of the assumption that $\log \tilde{w}_k^i$ is approximately Gaussian. We again investigate the values of skewness for this experiment; these results are given in Table 4. Note that the longer observation time window in the optimal proposal experiment here leads to higher values of skewness than for the shorter time window standard proposal experiment in the previous section.

TABLE 4. Skewness of the ensembles of $\log \tilde{w}_k^i$ after one evolution under the Lorenz 96 model for varying magnitude of observation error; optimal proposal experiment. Mean and 95% confidence intervals over final 2990 time steps.

σ^2_{obs}	mean skewness
5e-3	0.423 ± 0.010
0.01	0.485 ± 0.011
0.05	0.706 ± 0.014
0.1	0.781 ± 0.016
0.3	0.791 ± 0.016
0.5	0.718 ± 0.014
1	0.577 ± 0.013

Thus, we would expect worse agreement with the asymptotics in the optimal proposal experiments, because $\log \tilde{w}_k^i$ is less Gaussian than in the standard proposal experiments. The data points for which the full covariances were used (in blue) fall almost entirely above the theoretical line in solid black. On the other hand, since approximating the eigenvalues by the diagonal elements leads to underestimating τ, the data points for which this approximation was used (in red) are much closer to the theoretical line. That is, the underestimation of τ by the diagonal approximation compensates for the overestimation of τ by the theory due to the steep spectrum. But, as in the case of the standard proposal, these approximations are more accurate in the asymptotic regime (close to the origin) while the data deviates from the theory away from this regime.

6. Performance of Standard and Optimal Proposals

Recently, there has been a focus in the particle filtering community on the optimal proposal as an improvement over the standard proposal (Doucet et al. 2000; Arulampalam et al. 2002; Bocquet et al. 2010; Snyder 2012; Snyder and Bengtsson 2015). Intuitively, sampling from a distribution conditioned on the new observations should perform better than a distribution conditioned on the previous observations. The form of the weight update should also provide intuition behind the performance gain: the standard proposal weight update involves the distribution of the observations conditioned on the state at the current time $p(\mathbf{y}_k|\mathbf{x}_k)$, whereas the optimal proposal weight update is conditioned on the state at the previous time: $p(\mathbf{y}_k|\mathbf{x}_{k-1})$. Since uncertainty generally increases with a longer prediction window, the likelihood $p(\mathbf{y}_k|\mathbf{x}_{k-1})$ will tend to be broader than $p(\mathbf{y}_k|\mathbf{x}_k)$, and thus there will be less variance across the weights for the optimal proposal update.

In a review of non-Gaussian data assimilation methods, Bocquet et al. (2010) performed a simple comparison between the standard and optimal proposal implementation of the particle filter and found that the optimal proposal results in lower mean squared errors for smaller ensemble sizes, and has comparable performance to the standard proposal for large ensemble sizes. Here, we perform experiments which not only compare the mean squared errors of these methods, but we also consider the frequency at which resampling occurs as well as the maximum weight of each method after a single step.

To test the usefulness of the optimal proposal, experiments were run with the Lorenz (1996) system with 5, 10, and 20 variables, with full observations once per time step for 300 time steps, using both the standard and optimal proposal distributions. The observation error variance, system noise variance, and initial ensemble variance are fixed at $\sigma^2_{obs} = 0.5$, $\sigma^2_{sys} = 0.01$, $\sigma^2_{ens} = 1.0$, respectively. We test two resampling thresholds: first, when the effective sample size falls below $0.1N_e$; and second, when the maximum weight exceeds 0.5. After resampling, the weights are reset to $1/N_e$ and a small amount of jitter (with variance 0.01) is added to each particle. The errors are averaged over the last 200 time steps, but the resample counts are over the entire 300-step window.

TABLE 5. Number of times each method resampled in a window of 300 assimilation steps, for varying ensemble sizes and state dimensions. Top: resample when effective sample size (N_{eff}) falls below $0.1N_e$. Bottom: resample when maximum weight ($w_{\max}$) exceeds 0.5.

N_e	$N_x = 5$	$N_x = 5$	$N_x = 10$	$N_x = 10$	$N_x = 20$	$N_x = 20$
(N_{eff})	std	opt	std	opt	std	opt
20	254	81	297	150	298	237
50	264	103	299	190	299	272
100	269	113	299	195	299	277
500	266	115	299	208	299	289
1000	276	127	299	209	299	287
($w_{\max}$)						
20	272	112	297	197	297	197
50	239	85	297	154	297	154
100	202	76	291	129	291	129
500	127	49	284	98	284	98
1000	125	41	265	89	265	89

Figure 3 shows the root mean squared error of the posterior mean as a function of ensemble size. For the system

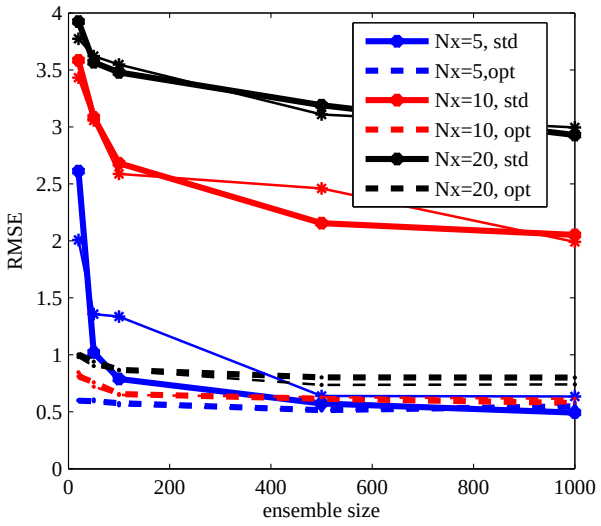

FIG. 3. Average errors for standard proposal (solid line) and optimal proposal (dashed line), as a function of ensemble size and for varying state dimensions: $N_x = 5$ (blue), $N_x = 10$ (red) and $N_x = 20$ (black). Thick lines represent the resampling threshold determined by effective ensemble size ($N_{\text{eff}} < 0.1N_e$) and thin lines represent the resampling threshold determined by maximum weight $w_{\max} > 0.5$).

with the smallest state dimension ($N_x = 5$) and increasing ensemble size, results for the standard proposal converge quickly toward those from the optimal proposal. When N_x is larger, however, the results for the standard proposal do not appear to converge over the range of ensemble sizes considered, and the root mean squared errors remain much larger than from the optimal proposal even at the largest ensemble size. Note also that the standard proposal improves slightly over the optimal proposal for $N_x = 5$ and large ensembles. We believe this reflects the approximations in our implementation of the optimal proposal. Similar errors result from both resampling thresholds, with the exception of small ensemble sizes for small state dimension ($N_x = 5$), in which case the threshold determined by effective sample size results in smaller errors.

A further difference is that the filter using the standard proposal resamples much more often than that of the optimal proposal, with both resampling thresholds. (See Table 5.) This may help explain why the optimal proposal has better error values: since resampling loses information in some sense, resampling less frequently should be preferable to resampling often.

Note that under the effective sampling size threshold, the number of times the filter resampled increases with ensemble size for a fixed state dimension. This may be due to the dependence of the threshold on the ensemble size, leading to increased resampling frequency with N_e. Alternatively, the optimal proposal density may be too narrow in relation to the posterior density, resulting in particles in the tail of the proposal with high posterior probability, and thus a low effective sample size. To test this, we also tried inflating the proposal variance as in Del Moral and Murray (2015). For these results, the resampling frequency still increased with increasing N_e, though not as drastically. Additionally, the errors were not affected by inflation, and so we have not included the results here. On the other hand, the threshold determined by the maximum weight results in decreasing resampling frequency as N_e grows for a fixed N_x, without inflating the proposal variance. This would suggest that even if the effective sample size is small, the weights are well-distributed across these particles. Then, even though the effective sample size may be increasing at a slower rate than N_e, resulting in higher resampling frequency with larger N_e, the weights are still distributed across more particles. While resampling methods comprise a rich area of research, they are not the focus of this work, and will not be investigated further here.

In addition to having smaller errors over time, the optimal proposal is less likely to experience collapse than the standard proposal. A hint to this behavior is given by the lower number of necessary resampling steps for the optimal proposal than the standard proposal; however, this effect can be studied directly by comparing the maximum weight after one step for each proposal. Results are shown in Figure 4. All parameters are fixed at the same value for each proposal, except the state dimension which varies as shown. The ensemble size is fixed at $N_e = 1000$, the data are averaged over 100 trials, and the error bars show 95% confidence intervals based on this sample. These results demonstrate that, for fixed ensemble size and state dimension, the optimal proposal consistently provides a lower maximum weight, and thus less variance across the weights. In this experiment, the improvement is especially clear in the regime where the state dimension is between 10 and 50.

The proposal density for the optimal proposal converges to that of the standard proposal when system noise becomes negligible. Without system noise, $\mathbf{x}_k$ becomes a deterministic function of $\mathbf{x}_{k-1}$ and both $p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{y}_k)$ and $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ are delta functions at $M_{det}(\mathbf{x}_{k-1})$. This can also be seen in the system (A1)-(A2) discussed in the appendix: when $\mathbf{Q}$ is very small, so is $\mathbf{K}$ in (A6) and the mean and covariance of $p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{y}_k)$ approach $M_{det}(\mathbf{x}_{k-1})$ and $\mathbf{Q}$, respectively, which are the same as the standard proposal mean and covariance. Thus, the gain that the optimal proposal affords over the standard proposal will be dependent on the size of the system noise. Table 6 includes results for the Lorenz system with 5 variables and 500 particles, with varying system noise. The resampling threshold is determined by the effective sample size (specifically, when N_{eff} falls below $0.1N_e$.) The particle filter with each

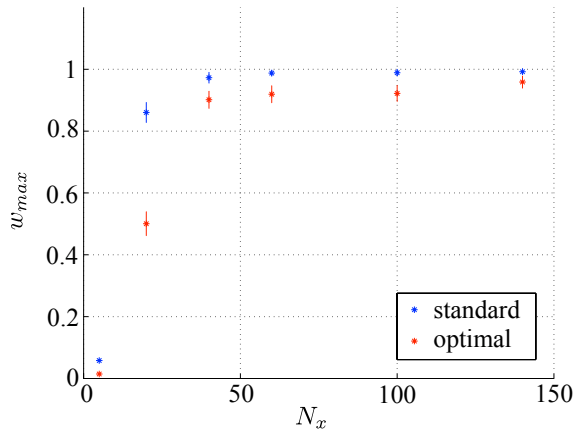

FIG. 4. Comparison of maximum weight after one assimilation step as a function of state dimension, using the standard proposal (blue) and the optimal proposal (red).

proposal distribution was run over 300 time steps, with observations at each time step; the table includes the ratio of the means of the errors over the last 200 time steps.

As Table 6 shows, the difference in errors between the standard and optimal proposal increases as the system noise increases. For the smallest system noise, the ratio of the optimal error to the standard error is very close to 1, but for larger noise, the optimal proposal yields a significant decrease in error over the standard proposal. Since the optimal proposal requires more computational effort than the standard proposal, if the problem of interest has very small system noise, then the standard proposal should be used. Lin et al. (2013) present the optimal proposal particle filter as one method in a class of "lookahead" algorithms, and investigate other such algorithms in the context of computational expense for various types of problems.

TABLE 6. Comparison of performance of standard and optimal proposal for varying size of model system noise.

σ_{sys}	(opt error)/ (std error)	PF, std resamp.	PF, opt resamp.
1e-3	0.916	59	62
5e-3	0.6877	105	94
1e-2	0.630	93	69
5e-2	0.473	127	48
0.1	0.473	160	39
0.5	0.374	258	11

7. Discussion & Conclusions

In this work, we attempted to answer the question of whether one could predict collapse of the optimal particle filter without building a scheme to sample from the optimal proposal. We have shown that this is possible in the Lorenz (1996) system, using results from Snyder et al. (2008) and their extension to the optimal proposal in Snyder and Bengtsson (2015). The results of the former demonstrate how to use eigenvalues of matrices from a linear, Gaussian system to calculate the effective dimension τ^2, which can then be used to assess the feasibility of the particle filter in that system. One key issue is determining the extent to which these results are valid in more general settings (e.g., nonlinear systems.) To this end, we have shown that the asymptotic approximations and results found in Snyder et al. (2008); Snyder and Bengtsson (2015) are useful also in the nonlinear regime with both the standard proposal and the optimal proposal.

Another key issue regarding the extension to nonlinear systems with the optimal proposal involves estimating an "effective" system-noise covariance corresponding to the additive Gaussian noise at observation times assumed in Snyder et al. (2008); Snyder and Bengtsson (2015). We have discussed several different approximations of this covariance, and shown that the asymptotic results are also valid with the optimal proposal when these approximate system-noise covariances are used.

Additionally, the eigenvalue decompositions necessary to estimate the effective dimension will be costly for large systems (and large ensembles.) Thus, in practice, we will need to find computationally feasible approximations. In this work, we have chosen to approximate the matrices as diagonal to simplify these eigenvalue calculations. This approximation appears to be effective in the idealized system considered here, though it also tends to overestimate the degree of collapse. The margin of this overestimation decreases as the system gets closer to collapse.

Finally, motivated by the results of Snyder (2012) which demonstrate the benefits of the optimal proposal implementation over the standard proposal in a simple example, we investigated the performance gain of the optimal proposal over the standard proposal in a nonlinear system. We have shown that the optimal proposal not only collapses less frequently than the standard proposal in the same regimes, but also results in quicker error convergence as a function of increasing particles. Thus, for systems in which the particle filter may work, utilizing the optimal proposal can provide increased performance with fewer particles than the standard proposal. The optimal proposal, however, is not trivial to implement and its benefits disappear in the limit of small system noise.

There are several remaining challenges regarding particle filters in nonlinear systems. Experiments still need

to be done to determine how the filters behave when applied sequentially; all the experiments in this paper study the degree of collapse after one assimilation step. However, this does not preclude the possibility of the particle filter collapsing after two or more steps. In addition, it could be useful to know whether the numerical results in this paper have an analytical analogue, as in the linear Gaussian case. Finally, further work should be done to investigate the optimal proposal, particularly in regards to approximations of the model noise covariance.

Acknowledgments.

Slivinski was supported by the NSF through grants DMS-0907904 and DMS-1148284, and by NCAR's Advanced Study Program during a collaborative visit to NCAR.

APPENDIX

Sampling from Optimal Proposal

Recall that the optimal proposal requires conditioning on the current observation: $p(\mathbf{x}_k|\mathbf{x}_{k-1},\mathbf{y}_k)$ (Doucet et al. 2000; Snyder 2012). Consider the case of additive Gaussian noise and a linear observation operator, where the system is given by

$$\mathbf{x}_k = M(\mathbf{x}_{k-1}) + \eta_k, \tag{A1}$$

$$\mathbf{y}_k = \mathbf{H}\mathbf{x}_k + \epsilon_k \tag{A2}$$

with $\eta_k \sim \mathcal{N}(0,\mathbf{Q})$ and $\epsilon_k \sim \mathcal{N}(0,\mathbf{R})$. Then

$$\mathbf{x}_k|\mathbf{x}_{k-1},\mathbf{y}_k \sim \mathcal{N}(\bar{\mathbf{x}}_k,\mathbf{P}), \tag{A3}$$

$$\bar{\mathbf{x}}_k = (\mathbf{I}-\mathbf{K}\mathbf{H})M(\mathbf{x}_{k-1}) + \mathbf{K}\mathbf{y}_k, \tag{A4}$$

$$\mathbf{P} = (\mathbf{I}-\mathbf{K}\mathbf{H})\mathbf{Q}, \tag{A5}$$

$$\mathbf{K} = \mathbf{Q}\mathbf{H}^T(\mathbf{H}\mathbf{Q}\mathbf{H}^T+\mathbf{R})^{-1}. \tag{A6}$$

In this case, the weights have an analytic update expression, since

$$\mathbf{y}_k|\mathbf{x}_{k-1} \sim \mathcal{N}(\mathbf{H}M(\mathbf{x}_{k-1}), \mathbf{H}\mathbf{Q}\mathbf{H}^T+\mathbf{R}). \tag{A7}$$

Thus, the particles at time t_k are first sampled from (A3), and then their weights are updated according to

$$w_k^i \propto \exp\left\{-\frac{1}{2}J(\mathbf{x}_{k-1}^i,\mathbf{y}_k)\right\} w_{k-1}^i, \tag{A8}$$

$$J(\mathbf{x}_{k-1}^i,\mathbf{y}_k) = \left(\mathbf{y}_k - \mathbf{H}M(\mathbf{x}_{k-1}^i)\right)^T \cdot \left(\mathbf{H}\mathbf{Q}\mathbf{H}^T+\mathbf{R}\right)^{-1}\left(\mathbf{y}_k - \mathbf{H}M(\mathbf{x}_{k-1}^i)\right). \tag{A9}$$

REFERENCES

Ades, M. and P. van Leeuwen, 2013: An exploration of the equivalent weights particle filter. *Quarterly Journal of the Royal Meteorological Society*, **139 (672)**, 820–840.

Anderson, J. and S. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.*, **127 (12)**, 2741–2758.

Arulampalam, M., S. Maskell, N. Gordon, and T. Clapp, 2002: A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, **50 (2)**, 174–188.

Bengtsson, T., P. Bickel, and B. Li, 2008: Curse of dimensionality revisted: Collapse of the particle filter in very large scale systems. *Probability and Statistics: Essays in Honor of David A. Freedman*, **2**, 316–334, D. Nolan and T. Speed, Eds., Institute of Mathematical Statistics.

Bickel, P., B. Li, and T. Bengtsson, 2008: Sharp failure rates for the bootstrap particle filter in high dimensions. *Pushing the Limits of Contemporary Statistics: Contributions in Honor of Jayanta K. Ghosh*, **3**, 318–329, b. Clarke and S. Ghosal Eds., Institute of Mathematical Statistics.

Bocquet, M., C. A. Pires, and L. Wu, 2010: Beyond Gaussian statistical modeling in geophysical data assimilation. *Mon. Wea. Rev.*, **138 (8)**, 29973023.

Burgers, G., P. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Monthly Weather Review*, **126**, 1719–1724.

Chorin, A. J. and M. Morzfeld, 2013: Conditions for successful data assimilation. *J. Geophys. Res. Atmos.*, **118 (20)**, 11,522–11,533, doi:10.1002/2013JD019838.

Del Moral, P. and L. Murray, 2015: Sequential Monte Carlo with highly informative observations. (in prep.).

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

Doucet, A., S. Godsill, and C. Andrieu, 2000: On sequential Monte Carlo sampling methods for Bayesian filtering. *Statist. Comput.*, **10**, 197–208.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.: Oceans*, **99 (C5)**, 10 143–10 162.

Evensen, G., 2003: The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics*, **53**, 343–367.

Gaspari, G. and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, **125 (554)**, 723–757.

Gordon, N., D. Salmond, and A. Smith, 1993: Novel approach to nonlinear-non-Gaussian Bayesian state estimation. *IEE Proceedings F (Radar and Signal Processing)*, **140 (2)**, 107–113.

Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Monthly Weather Review*, **129 (11)**, 2776–2790.

Hastings, W. K., 1970: Monte Carlo sampling methods using Markov Chains and their applications. *Biometrika*, **57**, 97–109.

Houtekamer, P. L. and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, **126**, 796–811.

Houtekamer, P. L. and H. L. Mitchell, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, **129**, 123–137.

Kong, A., J. Liu, and W. Wong, 1994: Sequential imputations and Bayesian missing data problems. *Journal of the American Statistical Association*, **89**, 278–288.

Lin, M., R. Chen, and J. Liu, 2013: Lookahead strategies or sequential Monte Carlo. *Statistical Science*, **28 (1)**, 69–94.

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

Morzfeld, M., X. Tu, E. Atkins, and A. J. Chorin, 2012: A random map implementation of implicit filters. *Journal of Computational Physics*, **231 (4)**, 2049–2066, doi:10.1016/j.jcp.2011.11.022.

Nakano, S., G. Ueno, and T. Higuchi, 2007: Merging particle filter for sequential data assimilation. *Nonlin. Processes Geophys.*, **14 (4)**, 395–408.

Robert, C. P. and G. Casella, 2004: *Monte-Carlo Statistical Methods*. Springer-Verlag, New York.

Silverman, B., 1986: *Density estimation for statistics and data analysis*, Vol. 26. CRC press.

Snyder, C., 2012: Particle filters, the "optimal" proposal and high-dimensional systems. *ECMWF Seminar on Data Assimilation for Atmosphere and Ocean*, Shinfield, UK, 161–170.

Snyder, C. and T. Bengtsson, 2015: Performance bounds for particle filters using the optimal proposal. (in prep.).

Snyder, C., T. Bengtsson, P. Bickel, and J. Anderson, 2008: Obstacles to high-dimensional particle filtering. *Monthly Weather Review*, **136**, 4629–4640.

van Leeuwen, P., 2009: Particle filtering in geophysical systems. *Monthly Weather Review*, **137**, 4089–4114.