Failures of sequential Bayesian filters and the successes of shadowing filters in tracking of nonlinear deterministic and stochastic systems

Kevin Judd and Thomas Stemler
School of Mathematics and Statistics, The University of Western Australia, 35 Stirling Highway, Crawley WA 6009 Perth, Australia
(Received 1 December 2008; revised manuscript received 11 April 2009; published 15 June 2009)

Sequential Bayesian filters, such as particle filters, are often presented as an ideal means of tracking the state of nonlinear systems. Here shadowing filters are demonstrated to perform better than sequential filters at tracking under specific circumstances. The success of shadowing filters is attributed to avoiding both well-known deficiencies of particle filters, and some newly identified problems.

DOI: 10.1103/PhysRevE.79.066206 PACS number(s): 05.45.Tp, 05.10.Gg, 02.50.Tt, 07.05.Kf

State identification is ubiquitous in science and engineering, with a wide range of applications from machine and process control, to navigation and guidance, to tracking moving objects, to weather forecasting. Various procedures called *state estimation*, *filtering*, *tracking*, and *data assimilation* are, from a mathematical point of view, aiming to solve the same problem, perhaps with some differences in emphasis and terminology. In each case there are noisy observations of a nonlinear system, which may involve random processes, and a model of the system, which may be a perfect representation of the system, although in practice never is. If the model were perfect, then the aim is to *estimate* the *true state* of the system. If the model is imperfect, then it is less clear what is to be achieved because there is no longer a true state of the model; the aim is perhaps better described as *tracking* the system state with an appropriate model state. If the goal is forecasting, then the model state ought to be one that obtains useful, or even optimal, forecasts by some criterion; the model need not be perfect to obtain useful forecasts. The term *state estimation* tends to imply there is a true state to estimate, so we prefer the term *state tracking* or *filtering* because they apply equally well in perfect and imperfect model scenarios.

The tracking problem can be formulated for systems evolving in continuous or discrete time [1]. When the observations are made at discrete equally spaced times, then it is sufficient for our purposes to consider only the discrete time case. Let

$$z_{t+1} = f(z_t) + \nu_t, \tag{1}$$

$$s_t = g(z_t) + \epsilon_t, \tag{2}$$

where $z_t \in \mathbb{R}^d$ is the system state at time t, with f defining the change in state between observations, that is, the *dynamics*; ν_t are independent random variables introducing a random element to the change of state, referred to as *dynamical noise*; $s_t \in \mathbb{R}^k$ is an observation of the system state at time t, which is a function g of the state plus noise ϵ_t; ϵ_t are independent random variables, referred to as *observational noise*. If $\nu_t=0$ for all t, then the system is *deterministic*, otherwise it is *nondeterministic* or *stochastic*. Independent in the above means each realization of a random variable at each time is independent of all other variables at any time.

In the perfect model scenario, state tracking requires providing an estimate of the true state z_t given a sequence of past observations s_τ, $\tau \leq t$. For stochastic systems the true state cannot be identified: the best that can be achieved is a probability density of states $p(z_t|s_\tau,\tau\leq t)$, regardless of how many observations are provided [1]. For deterministic systems the situation is more subtle. For a finite sequence of observations there is at best a probability density $p(z_t|s_\tau,t_0 \leq \tau\leq t)$ of states [1], but an infinite sequence of observations can lead to convergence (as $t_0 \to -\infty$) of the probability densities to the true state [2,3]. For deterministic systems that display sensitivity to initial conditions, then the theory of indistinguishable states [2] shows that the best that can be achieved is a probability density of states, even with an infinite sequence of observations. We will refer to procedures that obtain state estimates from observations as *filters*.

In the imperfect model scenario, where the dynamics f and the distribution of the stochastic forcing ν_t may be incorrectly specified in the model, then it is unclear what filters provide. Often it is assumed that model errors can be subsumed into dynamical noise terms, which may be appropriate in some applications. We will return to these difficulties later in Sec. IV, and in the meantime we focus on the perfect model scenario.

In many practical applications of filtering, such as process control and target tracking, a full probability density of states is not required or essential; all that is required is a single state, perhaps the maximum-likelihood state or some other best state by a specified criterion. For linear systems the Kalman filter can be optimal at obtaining the maximum-likelihood state [4], and the extended Kalman filter (or one of its many variants) may be adequate when nonlinearity is small relative to observational noise [5].

For general nonlinear systems a theoretically optimal filter for obtaining the state density $p(z_t|s_\tau,\tau\leq t)$ exists [1]: it requires complete knowledge of the previous state density $p(z_{t-1}|s_\tau,\tau< t)$, the change in state density $p(z_{t+1}|z_t)$, and the observation density $p(s_t|z_t)$. This filter employs Bayes rule and the Chapman-Kolmogorov equations to obtain the sequential update equations,

$$p(z_t|s_\tau,\tau \leq t) = \frac{p(s_t|z_t)p(z_t|s_\tau,\tau < t)}{p(s_t|s_\tau,\tau < t)}, \tag{3}$$

where

 ©2009 The American Physical Society

$$p(z_t|s_\tau, \tau < t) = \int p(z_t|z_{t-1})p(z_{t-1}|s_\tau, \tau < t)dz_{t-1}. \quad (4)$$

Despite the optimality of this filter, there are practical difficulties in implementing integral (4), except in special cases, such as linear systems with Gaussian noise, for which the Kalman filter is obtained. If the dimension of the system d is greater than one, then there are significant difficulties in just adequately representing the probability densities $p(z_t|s_\tau, \tau \leq t)$. A common approach is to use *particle filters*, which approximate the densities by an ensemble, or weighted ensemble, of representative states [6–9]. These filters can be very effective, but they have known failings, for example, instabilities that lead to collapse of the filter [10]. Filter collapse through degeneracy of the ensemble is well documented [6,8,10], but later we discuss a more subtle problem that persists even when the usual degeneracy modes of filter collapse are avoided; put simply, sequential filters lack the ability to backtrack and correct past errors.

The first principle point of this paper is that there are other kinds of filters that are not sequential filters, for example, the *shadowing filters* we discuss later. These nonsequential filters can also be optimal, but they optimize different criteria to sequential filters, and these other criteria may be more appropriate in certain applications. A second point of this paper is these nonsequential filters are easily implemented (even in high-dimensional systems) and avoid difficulties inherent in particle-based filters. A third point is that the shadowing filters we discuss generally apply equally well to stochastic and deterministic systems. We present evidence that a significant contribution to a shadowing filter's strength is its focus on dynamical aspects, rather than stochastic aspects, of the tracking problem. Historically, shadowing filters is a descendant of Laplace's, and subsequent other, work, on least-squares methods, which went out of favor in the 1970s with the introduction of probabilistic methods [1].

I. METHODS

To demonstrate our arguments we conduct numerical experiments where the tracking ability of a particular implementation of a sequential Bayesian filter (a particle filter) is compared with the tracking ability of two shadowing filters: one designed for deterministic systems, the other modified for stochastic systems. In order to ensure our results are as generic as possible, we organize tests so that the particle filter is given every reasonable advantage, whereas, the shadowing filters are not optimized and given only the absolute minimum of information required to obtain state estimates.

The numerical experiments were conducted on an Ikeda system [11]. Let $z_t=(u_t,v_t) \in \mathbb{R}^2$ and

$$f(u,v) = \begin{cases} 1 + \mu[u\cos(\theta) - v\sin(\theta)] \\ \mu[u\sin(\theta) + v\cos(\theta)], \end{cases}$$

$$\theta = a - b/(1 + u^2 + v^2), \quad (5)$$

for $a=0.4$, $b=0.6$, $\mu=0.83$, and $g(z)=z$, so $d=k=2$ in the formulation of Eqs. (1) and (2). For noise sources we use isotropic multivariate Gaussians so that if I is the 2×2 identity matrix, then $\nu_t \sim N(0,\eta^2 I)$ and $0.01 \leq \eta \leq 0.1$, and $\epsilon_t \sim N(0,\sigma^2 I)$ and $0.02 \leq \sigma \leq 0.4$.

A. Particle filter

The particle filter used in the experiments is a general sequential importance sampling (SIS) filter [6–9] using a Monte Carlo-Markov chain (MCMC) approach [12]. A particle filter can represent the density $p(z_t|s_\tau, \tau \leq t)$ either as an ensemble of states that are draws from this density, or as a weighted ensemble to reflect the probability density at each member of an ensemble of states. The latter can provide a significantly more efficient filter than the former, but the filter can be more prone to collapse if not implemented well [6,8,10].

In order to avoid issues of whether or not our particle filter is well implemented, we use an MCMC approach so that our ensemble should be draws of states from $p(z_t|s_\tau, \tau \leq t)$. Our first goal will be to test the tracking ability of the filter by comparing the maximum-likelihood state to the true state. Therefore, we also compute the density $p(z_t|s_\tau, \tau \leq t)$ for each state, to enable easy identification of the maximum-likelihood state; this density weight is not used in updating the ensemble.

Hence, our particle filter computes a weighted ensemble of states $\mathcal{E}_t=\{(z_t^i, w_t^i)\}_{i=1}^N$, where each state z_t^i has its corresponding weight $0 < w_t^i < 1$. Given the ensemble $\mathcal{E}_{t-1}$ and observation s_t, then $\mathcal{E}_t$ is obtained as follows. Define $\rho(z) = e^{-z^2/2\sigma^2}$.

(1) Generate $\mathcal{F}_t=\{(f(z_{t-1}^i), w_{t-1}^i)\}_{i=1}^N$.

(2) Select uniformly $(\zeta,\omega) \in \mathcal{F}_t$.

(3) Generate $z \sim N(\zeta, \eta^2 I)$.

(4) Compute $p=\rho(\|s_t - z\|)$.

(5) Generate uniformly $r \in [0,1]$.

(6) If $p > r$, then include $(z,\hat{w})$ in $\mathcal{E}_t$, calculating a provisional weight $\hat{w}=\omega p$.

(7) Repeat steps 2 to 6 M times or until $|\mathcal{E}_t|=N$.

(8) If $|\mathcal{E}_t| < N$, then repeats steps 2 to 6, skipping step 5 and accepting all $(r=0)$ in step 6.

(9) Normalize the weights: $w_t^i = \hat{w}_t^i / \Sigma_{j=1}^N \hat{w}_t^j$.

The maximum-likelihood state can be approximated by the state with largest weight. An initial ensemble $\mathcal{E}_0$ is constructed with draws from $N(s_0, \sigma^2 I)$. With $M=\infty$ and N sufficiently large, this algorithm should obtain samples from $p(z_t|s_\tau, \tau \leq t)$ after sufficient iteration. SIS filter algorithms often employ an auxiliary distribution, called a *proposal distribution*. The above algorithm does not use a proposal distribution because it directly samples the posterior density, provided the forecast ensemble $\mathcal{F}_t$ is sufficiently dense. Using a proposal distribution can significantly improve the efficiency of a SIS filter, but the performance depends crucially on the choice of proposal distribution. A poor choice of proposal distribution could degrade performance, especially in causing degeneracy problems [6,8,10]. In order to avoid questions of an appropriate proposal density, we use a brute force MCMC evaluation. Our experiments used $N=1000$ and $M=5\times 10^6$. Step 8 is introduced to limit computation time. For each experiment with different σ and η, on a sequence of

2000 observations, step 8 was employed for no more than three states, and in 50% of experiments step 8 was not invoked. We conclude, therefore, that $\mathcal{F}_t$ was nearly always sufficiently dense to achieve adequate sampling of the posterior density to avoid the commonly identified causes of particle filter collapse.

B. Two shadowing filters

Shadowing filters for deterministic systems attempt to find a trajectory of the system that *shadows*, that is, remains in close proximity to, a sequence of observations of the system state [5,13–15]. For stochastic systems one looks for a shadowing pseudo-orbit, rather than a trajectory [15,16]. A straightforward way to achieve shadowing is gradient descent of the *indeterminism* [2,14,17]. Let $S=(s_1,\ldots,s_n)$ be a sequence of observations of the system state, $X=(x_1,\ldots,x_n)\in\mathbb{R}^{nd}$ be a sequence of states, and $Y=(y_1,\ldots,y_{n-1})\in\mathbb{R}^{(n-1)d}$ be a sequence of quantities to be called *mismatches*. Given model dynamics f, the *generalized indeterminism* of (X,Y) is defined by

$$I(X,Y)=\frac{1}{2}\sum_{i=1}^{n-1}\|x_{i+1}-f(x_i)-y_i\|^2. \qquad (6)$$

If $I(X,0)=0$, then X is a trajectory of f, and for deterministic systems, $I(X,0)$ is a measure of how far the sequence of states X is from being a trajectory. Consequently, starting from $X=S$, and moving down the gradient of $I(X,0)$ until I is zero, is a means of finding a trajectory of f in close proximity to S; this is the essence of the shadowing filter for deterministic systems [2,14]. This idea can be generalized to a shadowing filter for nondeterministic systems, by starting from $X=S$ and $Y=0$, then moving down the gradient of $I(X,Y)$ until I is zero [15–17]. Gradient descent of I can be approximated by an iterative procedure as follows.

Define sequences $\delta X=(\delta x_1,\ldots,\delta x_n)$ and $\delta Y=(\delta y_1,\ldots,\delta y_{n-1})$ by

$$\delta y_i = x_{i+1}-f(x_i)-y_i,$$

$$\delta x_i=\begin{cases}-A(x_i)\delta y_i, & i=1\\ \delta y_{i-1}-A(x_i)\delta y_i, & 1<i<n\\ \delta y_{i-1}, & i=n,\end{cases} \qquad (7)$$

where $A(x)$ is the adjoint (transpose of Jacobian) of f at x. For a nondeterministic system, one step of size Δ down the gradient of I is $(X,Y)\mapsto(X-\Delta\delta X, Y-\Delta\delta Y)$. (For a deterministic system, $X\mapsto X-\Delta\delta X$ with $Y=0$ for all steps.) The iteration can be continued until convergence is achieved, and the final X provides an estimate of a shadowing trajectory, or shadowing pseudo-orbit. In our experiments the shadowing filters we applied used a fixed 300 iterations, with $\Delta=0.1$ and $n=16$.

II. TRACKING EXPERIMENTS

One aim of the numerical experiments is to compare the performance of the stated particle filter and shadowing filters in the task of tracking the state of a stochastic nonlinear system. Our first aim is to estimate the true state. The particle filter provides an estimate of a probability density for the system state, and hence for the purposes of tracking we can use the maximum-likelihood state as the state estimate, which can be approximated by the ensemble member with largest weight. For the shadowing filters we can use the last state x_n of the shadowing trajectory, or shadowing pseudo-orbit, as the state estimate. Although the test system is stochastic, we applied both the nondeterministic and deterministic forms of the shadowing filter.

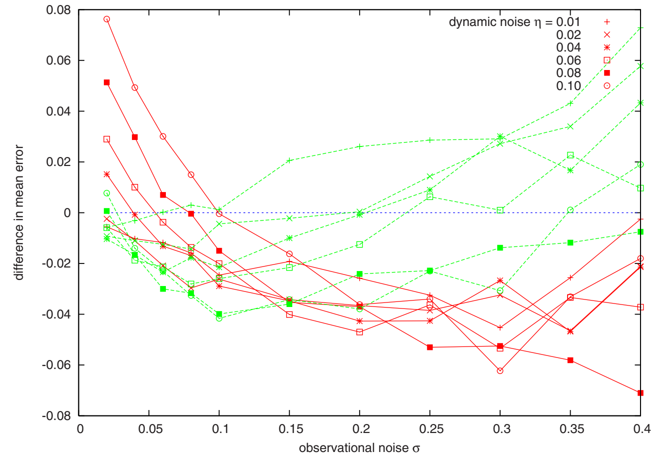

FIG. 1. (Color online) Mean, over 2000 states, of the difference in the distance between state estimate and true state for each shadowing filter and particle filter, plotted against σ and one line for each η value. Solid (red) line deterministic shadowing filter, dashed (green) line for nondeterministic shadowing filter. The standard deviation of the means is on the order of $\sqrt{(\sigma^2+\nu^2)/2000}$, which is about the vertical tick mark spacing for $\sigma=0.4$, and hence, the differences are significant.

Tests were conducted for all combinations of noise parameters $\sigma\in\{0.02,0.04,0.06,0.08,0.1,0.2,0.3,0.4\}$ and $\eta\in\{0.01,0.02,0.04,0.06,0.08,0.1\}$, where $\sigma=0.4$ is a signal-to-noise ratio of ≈ 1 (0 dB). In each case 16 observations were used to initialize the shadowing filters and particle filter, with $\mathcal{E}_0$ for the particle filter draws from $N(s_0,\sigma I)$. Then for the next 2000 observations the Euclidean distance between the state estimates and the true state were computed. Figure 1 plots the mean difference between shadowing filter errors and the particle filter errors so that negative values mean the shadowing filter performed better than the particle filter on average. The results of these experiments can be summarized as follows:

(i) If $\sigma>\eta$, then the deterministic shadowing filter performed better than the particle filter.

(ii) If $\sigma>2\eta$, then the deterministic shadowing filter outperformed the nondeterministic shadowing filter.

(iii) If σ is sufficiently small relative to η, certainly for $\sigma<2\eta$, then the nondeterministic shadowing filter generally performed better than particle filter.

III. COMMENTS

In these experiments the particle filter is given knowledge of the σ and η. In practical tracking problems σ may be known because it is a property of the measurement instruments, but η is usually unknown. Although, in principle, σ and η can be estimated from data, incorrect values of these will degrade the performance of the filter. On the other hand, the shadowing filters we implemented do not require this information; these shadowing filters exploit only knowledge of the dynamics f of the system, not knowledge of the randomness ϵ_t and ν_t.

For all test cases at least one of the two shadowing filters performed better than the particle filter, except for two cases where $\sigma=0.02$ and $\eta>0.06$. In these exceptional cases the dynamics of f are essentially irrelevant because the raw observation s_t gives an almost perfect state estimate. Furthermore, the dynamical noise ν_t effectively dominates, so that the dynamics f provides no useful additional information beyond that obtained from s_t.

Possibly the most surprising result of this experiment is that the deterministic shadowing filter performs best of all the filters, except for when $\eta>\sigma$, that is, treating the system as though it were a deterministic system gives the best tracking performance, except for when the dynamic noise exceeds the observational noise. The conclusion we draw from this is that if the observational noise exceeds the dynamical noise, then the nondeterministic effects are irrelevant and exploiting the dynamics f returns the greatest reward in terms of tracking performance. Knowledge of the dynamics f is therefore making its greatest contribution in its removal of the effects of observational noise ϵ_t.

Part of the reason for the different performance of the particle filter and shadowing filters is that they approximate different maximum-likelihood states. It can be shown that the shadowing filter is an approximation of the maximum-likelihood trajectory given the observations [2,15,17], that is, finds $(x_0,\ldots,x_n)$ that maximizes

$$p(x_0,\ldots,x_n|s_\tau,\tau\le n)=\left(\prod_{i=1}^{n}p(s_i|x_i)p(x_i|x_{i-1})\right)p(x_0|s_0). \tag{8}$$

On the other hand, from Eqs. (3) and (4),

$$p(x_t|s_\tau,\tau\le t)=\int\frac{\prod_{i=1}^{t}p(s_i|x_i)p(x_i|x_{i-1})}{\prod_{j=1}^{t}p(s_j|s_\tau,\tau<j)}p(x_0|s_0)dx_0\ldots dx_{t-1}. \tag{9}$$

Hence, the shadowing filter estimates x_t by maximizing the joint density of $(x_{t-n},\ldots,x_t)$, whereas a sequential Bayesian filter maximizes the marginal density of x_t. For linear systems with Gaussian noise sources, the joint and marginal densities are Gaussian with identical means, and hence both methods return the same maximum-likelihood state, which is a classical result on the equivalence of Kalman filters and variational methods [18]. For non-Gaussian densities the joint and marginal densities need not have the same maximum-likelihood state, and generally maximizing the joint density is better [19]. The maximization of joint density (8), rather than marginal density (9), may provide an explanation for the better performance of the shadowing filter observed in the experiments and displayed in Fig. 1.

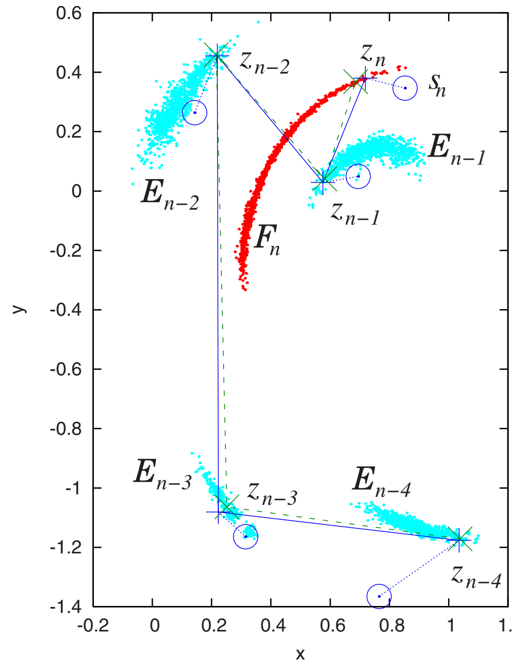

FIG. 2. (Color online) State estimates for test system with $\sigma=0.1$ and $\eta=0.01$. Five consecutive true states z_t (+) and observations s_t (○), particle filter ensembles $\mathcal{E}_t$ (·), and forward iterate $\mathcal{F}_n$ of $\mathcal{E}_{n-1}$. Also shown are the last five states x_t (×) of the deterministic shadowing filter trajectory valid at $t=n$.

It is well known that in general the performance of a particle filter is strongly dependent on the choice of proposal densities, and that any particle filter algorithm will fail if the ensemble size is not sufficiently large [6,8]. A subtle aspect to this failure is illustrated in Fig. 2 using data from our experiments. Here a sequence of observations is such that the forecast ensemble $\mathcal{F}_n$ assigns very low probability to the true state z_n. Consequently, very few ensemble members are close to the observation s_n, in this case, so few that the algorithm realizes less than 150 candidate states in $M=5\times10^6$ trials. This kind of problem occurred for around 1 in 30 states. If the ensemble size is smaller, the failure is more marked and frequent. The essential mechanism of failure is endemic to ensemble-based sequential Bayesian filters, and can be summarized as follows. Any filter may obtain a poor estimate of a current state because the arrangement of errors in recent observations gives a misleading picture of where the true state lies. Sequential filters can then find themselves in an untenable situation, in that they are unable to propagate any ensemble members forward to high probability posterior states.

Also displayed on Fig. 2 is the final part of the shadowing trajectory estimate obtained at time $t=n$. Observe how this shadowing trajectory accurately estimates each true state at this time, including the final state. Figure 3, on the other hand, shows with large crosshairs the state estimates valid at each previous time, that is, the final states of the shadowing pseudo-orbits using data only up to that time. It can be seen that at times $t=n-2$ and $t=n-1$ that the shadowing filter was as equally mislead by the observations as the particle filter.

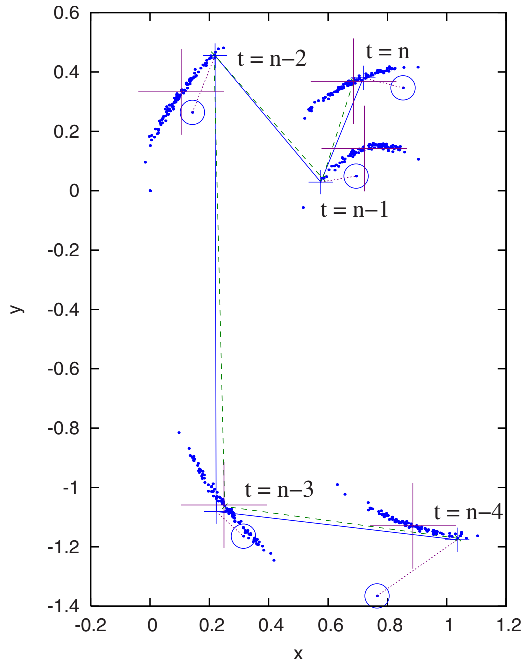

FIG. 3. (Color online) Same situation as Fig. 2 showing 100 member ensembles of indistinguishable states $(\cdot)$ valid at each t. Final shadowing filter state valid at each t indicated by crosshairs.

However, on obtaining the observation s_n the shadowing filter is able to recognize that the recent sequence of states is no longer tenable, and then recovers accurate estimates of each state, as seen in Fig. 2.

Moving beyond the tracking problem, there are situations where a probability density for the true states is useful, for example, in forecasting. The theory of indistinguishable states provides means to equip shadowing trajectories with probability densities of states [2,17]. Comparing the performance of probability density estimates is difficult. Figure 3 also shows the corresponding indistinguishable state ensembles valid at each time, that is, this is a random selection of the indistinguishable states based upon their indistinguishability density Q [2,17]. We observe that these ensembles are essentially equivalent to the particle filter ensembles in this case, although more compact on the attractor. It is important to note that these ensembles are generated from the shadowing trajectory, and hence can be of arbitrary size without affecting the performance of the shadowing filter. In contrast, particle filters require a sufficiently large ensemble to achieve reliable results.

IV. CONCLUSIONS

We have compared the tracking ability of a sequential Bayesian filter, implemented as a particle filter, with that of deterministic and nondeterministic shadowing filters. We found that the shadowing filters perform better than the particle filter, except when the observational noise σ is very much smaller than the dynamical noise η, which is a situation where the dynamical noise is the dominant source of uncertainty. We conclude that the principle value the dynamics f is limiting the uncertainty due to observational noise; hence, if the observational noise ϵ_t is larger than the dynamical noise ν_t, then the nonlinear dynamics f of a model is more important to the tracking of the system than the stochastic element ν_t of the model.

We have also discussed how particle filters have an endemic problem in ensemble collapse, not only in the traditionally recognized problem of degeneracy, but also in a subtle newly recognized form resulting from the inability to backtrack and correct past mistakes. These kind of problems do not affect shadowing filters. Furthermore, indistinguishable states of shadowing trajectories appear to provide ensemble estimates of state probabilities at least as good as those obtained from particle filters.

This paper has only identified specific problems with particle filters; it has not attempted to investigate the precise circumstances when one filter performs better the others. Future research ought to address performance for both the maximum-likelihood state estimation and probabilistic forecasting based on ensembles. There are many features of a system that can affect performance of the filters, for example, the magnitude of noise sources relative to each other and relative to the diameter of the attractor, the signal-to-noise ratio, the strength of the nonlinearity of the system, and the largest Lyapunov exponent. Each of these factors plays a role in limiting performance, and these factors are not all independent. There are additional issues of computation resources required to achieve a prescribed level of reliability of filters. At this stage no clear indication can be given of what features affect performance most in any circumstance. It is certainly known that Bayesian methods fail in the limit of deterministic systems [20]. The authors are preparing guides to the successful implementation of shadowing filters [21] and ensembles of indistinguishable states [22].

The results of this paper should also be viewed within the context of appropriate application of Bayesian methods to deterministic nonlinear systems [20], and fundamental limitations to likelihood methods in trajectory estimation [23]. A more important issue, touched on earlier, is the fact that in practice models are imperfect, whereas Bayesian methods are usually applied within an implicitly perfect model scenario. Although shadowing techniques have been applied in imperfect model scenarios [15,17], there is still only a basic level of theoretical understanding of the effects of model error on filters and the interpretation of their forecasts [15,16,24].

ACKNOWLEDGMENTS

The authors acknowledge the useful comments of the reviewers. This research was supported by ARC through Grants No. DP0662841 and No. DP0984659.

[1] A. H. Jazwinski, *Stochastic Processes and Filtering Theory* (Academic, New York, 1970), Vol. 64.
[2] K. Judd and L. A. Smith, Physica D **151**, 125 (2001).
[3] D. Ridout and K. Judd, Physica D **165**, 26 (2002).
[4] R. E. Kalman, J. Basic Eng. **82**, 35 (1960).
[5] K. Judd, Physica D **183**, 273 (2003).
[6] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, IEEE Trans. Signal Process. **50**, 174 (2002).
[7] P. Del Moral, Theor. Probab. Appl. **40**, 690 (1995).
[8] A. Doucet, S. Godsill, and C. Andrieu, Stat. Comput. **10**, 197 (2000).
[9] D. B. Rubin, in *Bayesian Statistics 3*, edited by J. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith (Oxford University Press, New York, 1988).
[10] F. Le Grand and N. Oudjane, Ann. Appl. Probab. **14**, 144 (2004).
[11] K. Ikeda, Opt. Commun. **30**, 257 (1979).
[12] W. R. Gilks and C. Berzuini, J. R. Stat. Soc. Ser. B (Stat. Methodol.) **63**, 127 (2001).
[13] J. Bröcker, U. Parlitz, and M. Ogorzałek, Proc. IEEE **90**, 898 (2002).
[14] M. Davies, Int. J. Bifurcation Chaos Appl. Sci. Eng. **3**, 113 (1993).
[15] K. Judd, Physica D **237**, 216 (2008).
[16] K. Judd, J. Nonlinear Sci. **18**, 57 (2008).
[17] K. Judd and L. A. Smith, Physica D **196**, 224 (2004).
[18] H. E. Rauch, F. Tung, and C. T. Striebel, AIAA J. **3**, 1445 (1965).
[19] Consider a density $p(x_1,x_2)$ zero everywhere, except on the set $\{(x_1,x_2)\in\mathbb{R}^2 : x_2=x_1^2\}$, and is unimodal on this line with a maximum at (1, 1). The marginal density on x_1 is unimodal with a maximum at $x_1=1$, but the marginal density on x_2 can have two peaks at $x_2=1$ and $x_2=0$. Depending on how the density lies on the line, the peak at $x_2=0$ can be larger than the peak at $x_2=1$; hence, a marginal density can be misleading. Consequently, it is generally better to maximize a joint density to avoid this kind of situation.
[20] K. Judd, Phys. Rev. E **67**, 026212 (2003).
[21] T. Stemler and K. Judd, Physica D **238**, 1260 (2009).
[22] T. Stemler and K. Judd (unpublished)..
[23] K. Judd, Phys. Rev. E **75**, 036210 (2007).
[24] K. Judd and T. Stemler, Philos. Trans. R. Soc. London (to be published).