University of Colorado at Denver and Health Sciences Center

Predictor-Corrector Ensemble Filters for the Assimilation of Sparse Data into High Dimensional Nonlinear Systems

Jan Mandel and Jonathan D. Beezley

June 2006 UCDHSC/CCM Report No. 232

CENTER FOR COMPUTATIONAL MATHEMATICS REPORTS

Predictor-Corrector Ensemble Filters for the Assimilation of Sparse Data into High-Dimensional Nonlinear Systems*

Jan Mandel† Jonathan D. Beezley‡

June 7, 2006

Abstract

An ensemble particle filter is proposed which is suitable for very large systems with smooth state, such as arizing from discretization of partial differential equations. The proposal ensemble comes from an arbitrary unknown distribution, and it is selected to have a good coverage of the support of the posterior. Proposal ensembles from the ensemble Kalman filter and from deterministic nudging to randomly perturbed observation data are considered. The ratio of the the prior and the proposal densities for calculating the importance weights is obtained by density estimation in Sobolev spaces, which are infinitely dimensional, and so the density estimate does not deteriorate in high dimension. Numerical experiments show that the new filter combines the advantages of ensemble Kalman filters and particle filters.

Keywords: Ensemble Kalman filters, data assimilation, density estimation, particle filters, sequential Monte-Carlo, Sobolev spaces, probability on functional spaces.

*This research was supported by NSF grant CNS-0325314 and by the NCAR Faculty Fellowship Program.

†Center for Computational Mathematics, Department of Mathematical Sciences, and Department of Computer Science, University of Colorado at Denver and Health Sciences Center, Denver CO 80217-3364. Visiting faculty, NCAR, Boulder, CO.

‡Department of Mathematical Sciences, University of Colorado at Denver and Health Sciences Center, Denver CO 80217-3364. Visiting graduate student, NCAR, Boulder, CO.

1. Introduction

The members of ensemble or particle filters play two roles: *(i)* the members carry information about the state distribution as a sample from the distribution (in ensemble filters) or weighted sample (in particle filters), and *(ii)* the members serve to create a representation of the posterior distribution by forming linear combinations in Ensemble Kalman filters (EnKFs) or by assignment of weights and optionally resampling in particle filters (PFs).

The role of the ensemble to be a sample from the state distribution is preserved by advancing the ensemble in time. However, the posterior distribution may not be well represented in the above way if the prior ensemble is not rich enough. In particular, if the posterior is much different from the prior, particle filters result in very small weights for many members (thus severely decreasing the effective size of the ensemble), and ensemble filters may produce far off analysis ensemble in a hopeless attempt to find a linear combination of ensemble members to match the observation. Ensemble filters have the remarkable ability to make big changes from the prior to the posterior, as long as a good posterior ensemble can be found in the span of the prior ensemble. In addition, if the system is nonlinear, the fact that ensemble filter algorithms rely on assumption about a specific form of state distribution (such as gaussian or gaussian mixture) results in an analysis ensemble that is not a sample from the posterior.

We propose to decouple the two roles of the ensemble. At analysis time, we use two ensembles. The *forecast ensemble* is obtained by advancing in time, and it represent the prior. The *proposal ensemble* are states that are a good supply of material to build a representation of the posterior; they are obtained by another method, called the *predictor*. The predictor can be a version of EnKF, or deterministic techniques such as nudging, which tend to produce system states that are a better match for the data. The *corrector* then uses density estimation to assign weights to the members of the proposal ensemble. The weighted and optionally resampled proposal ensemble will become the *analysis ensemble*, which represents the posterior.

The mathematical approach in this paper is heuristic. A rigorous analysis will be studied elsewhere.

This paper is organized as follows. In Sec. 2, we introduce the concept of weighted ensemble and briefly review the EnKF and PF. Description of the new class of methods is in Sec. 3, and Sec. 4 contains the results of numerical experiments. Sec. 5 is the conclusion. The necessary concepts of probability on infinitely dimensional spaces and nonparametric density estimation, which motivate the new methods, are briefly reviewed in Appendix A.

2. Preliminaries

We first need to collect some notation and concepts.

2a. *Importance Monte Carlo Quadrature and Weighted Ensemble*

Given a probability measure μ on $\mathbb{R}^n$, consider the integral $\int f d\mu$ with the function f scalar valued. Suppose that π is another probability measure, called the importance measure, and p_μ, p_π are the densities of μ and π relative to some underlying measure ν. Suppose that $(u_1, \ldots, u_N)$ is a sample from π, that is, a realization of N independent random variables with the same distribution π. A sample will be also called an *ensemble*. The Monte Carlo quadrature method, e.g., (Doucet et al. 2001, Ch. 1), gives

$$\begin{aligned} \int_\Omega f d\mu &= \int_\Omega f p_\mu d\nu = \int_\Omega f \frac{p_\mu}{p_\pi} p_\pi d\nu = \int_\Omega f \frac{p_\mu}{p_\pi} d\pi \\ &\approx \sum_{k=1}^N w_k f(u_k), \quad w_k = \frac{1}{N} \frac{p_\mu(u_k)}{p_\pi(u_k)}, \end{aligned} \tag{1}$$

with error $O(N^{-1/2})$ in mean square, if

$$\int_\Omega \left| f \frac{p_\mu}{p_\pi} \right|^2 d\pi < +\infty.$$

In this sense, the weighted ensemble (u_k, w_k)

$$(u_1, \ldots, u_N) \sim \pi, \quad w_k \propto \frac{p_\mu(u_k)}{p_\pi(u_k)}, \quad \sum_{k=1}^N w_k = 1 \tag{2}$$

represents the probability distribution μ, and we will write

$$(u_k, w_k) \sim \mu \text{ or } (u_k, w_k) \sim p_\mu \tag{3}$$

to indicate (2). In (2) and in the rest of this paper, $\propto$ means proportional.

If $(u_k, w_k) \sim \mu$, an unbiased estimate of the mean $\bar{u}$ and covariance Q of μ is given by the weighed sample mean and covariance,

$$\bar{u} = (\bar{u}_i) = \sum_{k=1}^N w_k u_k \tag{4}$$

$$Q = [q_{ij}], \tag{5}$$

$$q_{ij} = \frac{\sum_{k=1}^N w_k (u_{ik} - \bar{u}_i)(u_{jk} - \bar{u}_j)}{1 - \sum_{k=1}^N w_k^2}, \tag{6}$$

where u_{ik} is the i-th entry of ensemble member u_k. Note that if all $w_k = 1/N$, then (4–6) reduce to the usual formulas for sample mean and covariance. Weighted ensembles are related to biased samples in statistics (Efromovich 2004).

To advance a weighted ensemble in time, all ensemble members are advanced and the weights remain the same (Doucet et al. 2001, Ch. 1).

2b. *Data Assimilation*

Sequential statistical estimation techniques have become known in meteorology and oceanography as data assimilation (Bennett 1992, p. 67). The discrete time state-space model is an application of the bayesian update problem. The state of the model is an approximation of the probability distribution of the system state u, usually written in terms of its density $p(u)$. The probability distribution is advanced in time until an *analysis* time, when new data is incorporated into the probability distribution by an application of Bayes theorem,

$$p_a(u) = p(u|d) \propto p(d|u)\, p_f(u). \tag{7}$$

Here $p_f(u)$ is the probability density before the update, called the *prior* or the *forecast* density, the conditional probability density $p(d|u)$ is the *data likelihood,* and $p_a(u)$ is the *posterior* or the *analysis* density. The system is then advanced until the next analysis time. The data likelihood is the probability that the measurement value is d if the true state of the system is u. The data likelihood is found from the data error distribution ε, which is assumed to be known (every measurement must be accompanied by an error estimate), and from an observation function h, by

$$d - h(u) \sim \varepsilon. \tag{8}$$

The value $h(u)$ of the observation function would be the value of the measurements if the system state u and the measurements d were exact. In the rest of this paper, we consider only a linear observation function h, given by $h(u) = Hu$.

2c. *Ensemble Kalman Filter*

If the forecast distribution and the data error distributions are gaussian, and the observation function is linear, then the analysis distribution can be computed by linear algebra. This is the famous *Kalman filter* (Kalman 1960; Kalman and Bucy 1961): If the forecast density p_f is gaussian with mean $\bar{u}_f$ and covariance Q_f and, for any fixed u, the data likelihood $p(d|u)$ is gaussian with mean Hu and covariance R, then the analysis density p_a is also gaussian and the analysis mean $\bar{u}_a$ can be computed by

$$\bar{u}_a = \bar{u}_f + K(d - H\bar{u}_f), \tag{9}$$

where

$$K = Q_f H^{\mathrm{T}} \left(H Q_f H^{\mathrm{T}} + R\right)^{-1} \tag{10}$$

is the Kalman gain matrix. The analysis covariance is $Q_a = (I - KH)\, Q_f$. See (Anderson and Moore 1979) for details.

Because Kalman filter and its variations need to advance in time the covariance matrix of the state, they are unsuitable for systems with a large number of degrees of freedom, such as arising from numerical solution of PDEs. EnKFs (Evensen 1994; Houtekamer and Mitchell 1998) represent the distribution of the system state using an ensemble, and replace the covariance matrix Q_f in (10) by the sample covariance computed from the forecast ensemble members. In the EnKFs in the literature, the ensemble is not weighted. To obtain correct statistics of the analysis ensemble, the update needs to use the data d with a perturbation sampled from the data error distribution for each ensemble member (Burgers et al. 1998); cf., Algorithm 1 below. There are also versions of EnKF that do not involve randomization of data, and form the analysis ensemble deterministically (Anderson 1999; Tippett et al. 2003).

The principal advantage of EnKF over PF is that now the analysis distribution and the forecast distribution can be very different; the update formula gives the correct ensemble (in the limit for $N \to \infty$, of course), as long as the forecast ensemble is a sample from a gaussian distribution. Therefore, the analysis cycles can be much longer and the ensembles much smaller than for PFs. However, the assumption that the distributions are gaussian is a serious limitation. There have been attempts to relax the gaussian assumption by gaussian mixtures (Anderson and Anderson 1999; Bengtsson et al. 2003; Chen and Liu 2000), but filters relying on the gaussian assumption are still used in practice.

In general, *EnKF works by forming the analysis ensemble as linear combinations of the forecast ensemble* (Evensen 2004; Tippett et al. 2003). This raises two concerns, especially in highly nonlinear models: 1. if the change of state in the bayesian update is large, there may not be suitable forecast members to take linear combinations of in order to match the data; 2. a linear combination of realizable states may not be a realizable state. The former concern again results in the need for large ensembles and frequent small updates. The latter concern can be addressed, to some extent, by penalizing non-realizable states (Johns and Mandel in print). Similar concerns arise in any method that resamples the ensemble from estimated density, such as (Kim et al. 2003; Pham 2001; Xiong and Navon Tellus, submitted), because resampling may not yield a realizable state (Pham 2001).

For surveys of EnKF techniques, see Evensen (2003,

2004) and Tippett et al. (2003).

2d. *Particle Filters*

PFs approximate the probability of the state $p(u)$ by a weighted ensemble $(u_k, w_k) \sim p$. To incorporate new data following the Bayes theorem (7), suppose that we are given the forecast density p_f and a sample (u_k^a) from a *proposal* density π, which is hopefully close to the analysis density. Then, from (2) and (7), the analysis density p_a is represented by the weighted ensemble

$$(u_k^a, w_k^a) \sim p_a,\, w_k^a \propto \frac{p_a(u_k^a)}{p_\pi(u_k^a)} = p\left(d|u_k^a\right) \frac{p_f\left(u_k^a\right)}{p_\pi\left(u_k^a\right)}. \quad (11)$$

While the data likelihood $p\left(d|u\right)$ can be readily evaluated from (8), the ratio of the densities is in general not available, except in the particular case when the forecast is given as a weighted ensemble and the proposal analysis ensemble is taken the same as forecast ensemble, $(u_k^a) = (u_k^f)$; then the analysis weights are simply

$$w_k^a \propto p(d|u_k^f) w_k^f.$$

This is the Sequential Importance Sampling (SIS) method; see Doucet et al. (2001, Ch. 1) for an overview. In this paper, we will propose an *efficient way to estimate the ratio of the densities in (11), using nonparametric density estimation*.

The problem with PFs is that only members where the weights w_k^a are large contribute to the accuracy of the representation. Because the regions where analysis density p_a is large and where the forecast density p_f is large in general differ, the effective size of the ensemble decreases, and, in few analysis cycles, the filter fails. Therefore, the SIS update is followed by resampling to construct an analysis ensemble with all weights equal. In the SIR method (the bootstrap filter) (Gordon and Smith 1993), a new ensemble member u_k^a is obtained by selecting u_ℓ^f with probability w_ℓ^a. This results in an ensemble with repeated members, and stochastic advance in time (Markov chain) is relied upon to spread the ensemble again. In (Kim et al. 2003; Xiong and Navon Tellus, submitted), the resampling is done by first estimating the density p_a, and then generating the new ensemble by random sampling from the estimated density. Setting likelihood weights by kernel estimation of the analysis density was also used in a somewhat different context (de Valpine 2004).

3. The New Predictor-Corrector Filters

The new method proposed in this article consists of

1. generating the initial ensemble and advancing it in time,
2. selecting a proposal ensemble that has concentration of members close to the analysis distribution (the *predictor*),
3. computing the importance weights from the forecast density p_f and proposal density p_π by density estimation (the *corrector*), and
4. optionally resampling the proposal ensemble

3a. *Initial Ensemble*

The initial ensemble is created by the discrete Fourier transform in a standard manner. Given initial state u_0 and a Fourier basis $\{e_n\}_{n=1}^d$ of the state space, such as sine, cosine, or complex exponential function, and coefficients $\lambda_n > 0$ such that $\sum_{n=1}^\infty 1/\lambda_n < +\infty$, generate the initial ensemble members by

$$u_k = u_0 + \sum_{n=1}^d \frac{v_n}{\lambda_n} e_n, \quad \nu_n \sim N\;(0,1)\,. \quad (12)$$

The smoothness of the initial ensemble is controlled by the growth rate of the coefficients λ_n. This process can be understood as a finite dimensional version of random sampling from a gaussian measure on an infinitely dimensional space, cf., Appendix Ab.

3b. *The Predictor*

Given a forecast ensemble (u_k^f, w_k^f), the predictor generates a proposal ensemble (u_k^π, w_k^π). It is enough to know that the proposal ensemble is obtained from some probability distribution that has continuous density;

it is not necessary to know what that distribution exactly is. The proposal ensemble should be concentrated about where the posterior density is expected to be concentrated, and the data likelihood on the proposal ensemble members should not be too small. Here are some possible choices.

i. EnKF as the Predictor The proposal ensemble can be obtained from adding weights to the version of the EnKF algorithm by Burgers et al. (1998).

Algorithm 1 (Weighted EnKF) *Given weighted forecast ensemble* (u_k^f, w_k^f)*, data* d*, and observation matrix* H*,*

1. *compute the weighted forecast mean* $\bar{u}_f$ *and covariance* Q_f *from (4) and (6),*
2. *generate* d_k *by random perturbation of the data,* $d_k = d + e_k$*,* $e_k \sim N(0, R)$*,*
3. *compute the proposal ensemble* (u_k^π, w_k^π) *from*

$$u_k^\pi = u_k^f + K\left(d_k - Hu_k^f\right), \quad w_k^\pi = w_k^f.$$

EnKF as the predictor has the advantage that it is asymptotically exact in the gaussian case, and it often works quite well otherwise. The resulting predictor-corrector filter retains a great deal of the efficiency of the EnKF, but without the gaussian assumption.

ii. Predictor by Nudging Another possibility is an ensemble obtained by nudging of the forecast ensemble members to randomly perturbed data. First generate perturbed data (d_k) by sampling from the error distribution ε. Then evolve the system to achieve states u_k such that $h(u_k) \approx d_k$ by nudging: instead of evolving u_k from the differential equation $u' = F(u)$, use

$$u' = F(u) - cG(h(u) - d_k) \tag{13}$$

where G is such that the solution is attracted to a states u where $h(u) \approx d_k$. If the observation function is linear, $h(u) = Hu$, then G should be a right approximate inverse of H: $HG \approx I$. The constant $c > 0$ is small enough not to affect the dynamics significantly. Typically, H implements interpolation from mesh nodes to the measurement location, while G spreads the residual to mesh nodes within a small "circle of influence" (Kalnay 2003). The view of nudging as an application of a one-sided inverse, presented here, seems to be new.

3c. *The Corrector*

The role of the corrector is to assign importance weights from (11) by density estimation.

Consider the kernel $\varphi(\|u\|_U)$, where $\|\cdot\|_U$ is a norm on the state space, for example

$$\|v\|_U^2 = \sum_{n=1}^{d} \frac{|v_n|^2}{\lambda_n}, \quad v = \sum_{n=1}^{d} v_n e_n.$$

Then, using (1) and (31), we have for the forecast density

$$\sum_{k=1}^{M} w_k^f \varphi\left(\frac{\|u_k^f - y\|_U}{h}\right) \tag{14}$$

$$\approx \int p_f(u)\, \varphi\left(\frac{\|u-y\|_U}{h}\right) d\nu(u) \tag{15}$$

$$\approx p_f(y) \int_U \varphi\left(\frac{\|u\|_U}{h}\right) d\nu(u), \tag{16}$$

and similarly for the proposal density p_π, which yields an estimate for the ratio of the densities in (11),

$$\frac{p_f(y)}{p_\pi(y)} \approx \frac{\sum_{k=1}^{M} w_k^f \varphi(\|u_k^f - y\|_U/h)}{\sum_{\ell=1}^{N} w_\ell^\pi \varphi\left(\|u_\ell^\pi - y\|_U/h\right)}. \tag{17}$$

Note that the unknown integral from (16) cancels.

Algorithm 2 *Given a kernel function* φ*, bandwidth* h*, and norm* $\|\cdot\|_U$ *on the state space, for a forecast ensemble* $(u_k^f, w_k^f)_{k=1}^M$*proposal ensemble* $(u_\ell^\pi, w_\ell^\pi)_{k=1}^N$*, compute the analysis ensemble* (u_k^a, w_k^a) *from*

$$\begin{aligned} u_k^a &= u_k^\pi, \\ w_k^a &\propto p(d|u_k^\pi) \frac{\sum_{k=1}^{M} w_k^f \varphi(\|u_k^f - u_k^\pi\|_U/h)}{\sum_{\ell=1}^{N} w_\ell^\pi \varphi\left(\|u_\ell^\pi - u_k^\pi\|_U/h\right)}, \\ \sum_{k=1}^{M} w_k^a &= 1. \end{aligned}$$

This is the SIS method with the ratio of the densities replaced by an estimate. So, when EnKF is used as the predictor, we will call the resulting filter EnKF-SIS.

4. Numerical Results

To get an idea of why this new method of filtering is beneficial, we will consider some situations where standard filtering techniques are unsuitable. Such conditions are frequently encountered when considering nonlinear problems, or when it is technically unfeasible to use a sufficiently large ensemble to approximate the distributions.

In all tests, density estimation with the bandwidth determined by the k_N-th nearest was used with $k_N = [\sqrt(N)]$, cf., Appendix Ac.

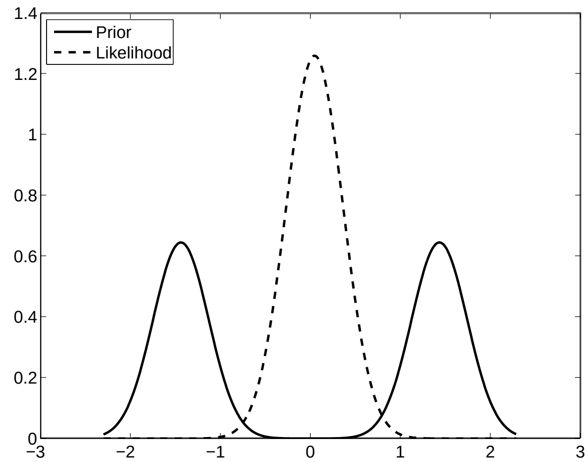

Figure 1: Prior and likelihood distributions

4a. *Bimodal Prior*

The first such situation we will consider is that of a bimodal prior. With a suitable likelihood distribution, it is possible to obtain a highly non-gaussian posterior. We construct a bimodal prior by first sampling from a gaussian distribution with variance 5. These samples are then weighted by the function

$$w_f(x_i) = e^{-5(1.5-x_i)^2} + e^{-5(-1.5-x_i)^2}$$

representing the sum of two gaussian distributions with mean ± 1.5 and variance 0.1. The resulting weighted ensemble can then be considered a weighted sample from the prior probability distribution function shown in Fig 1. The likelihood is also gaussian with mean shifted slightly to the right.

Each filter (EnKF, SIS, and EnKF-SIS) was applied to this problem with an ensemble size of 100. Density estimation, given by (15), was then applied to the resulting posterior ensemble to obtain an approximate posterior probability distribution. The results for each method are given in Figs 2, 3, and 4.

Because the EnKF technique assumes that all distributions are gaussian, it is no surprise that it would fail to capture the non-gaussian features of the posterior. Both SIS and EnKF-SIS were able to represent the nature of the posterior reasonably well.

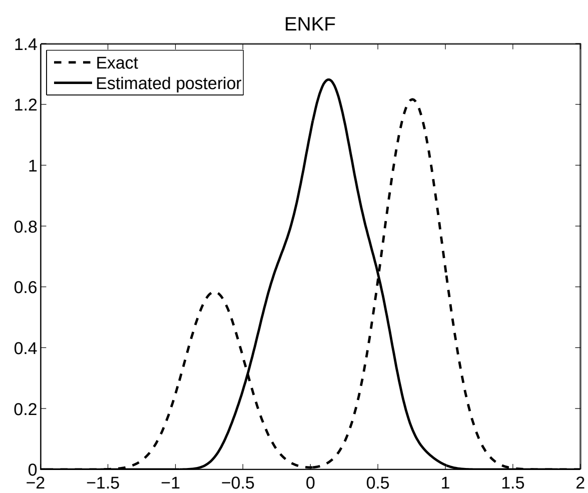

Figure 2: Posterior from EnKF

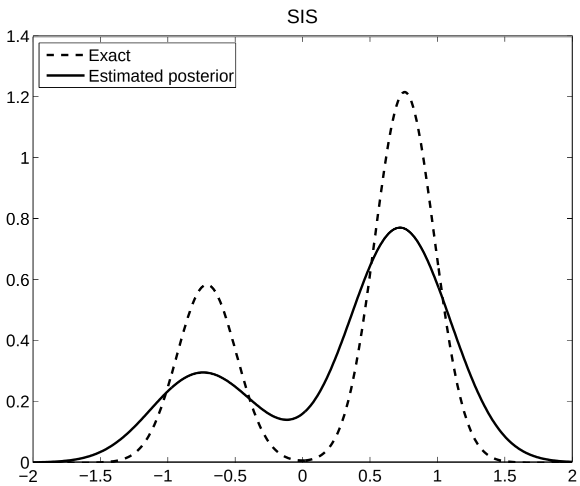

Figure 3: Posterior from SIS

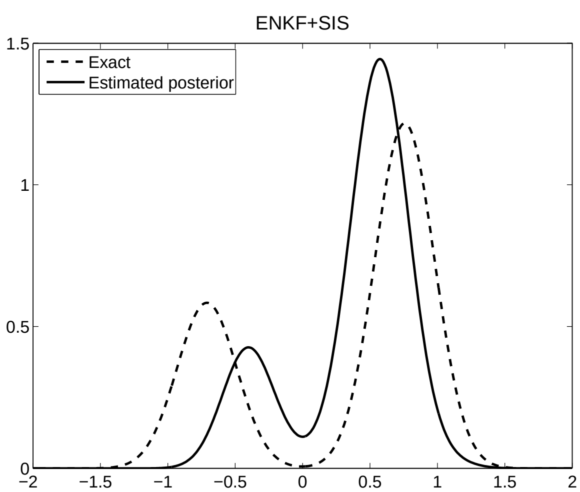

Figure 4: Posterior from EnKF-SIS

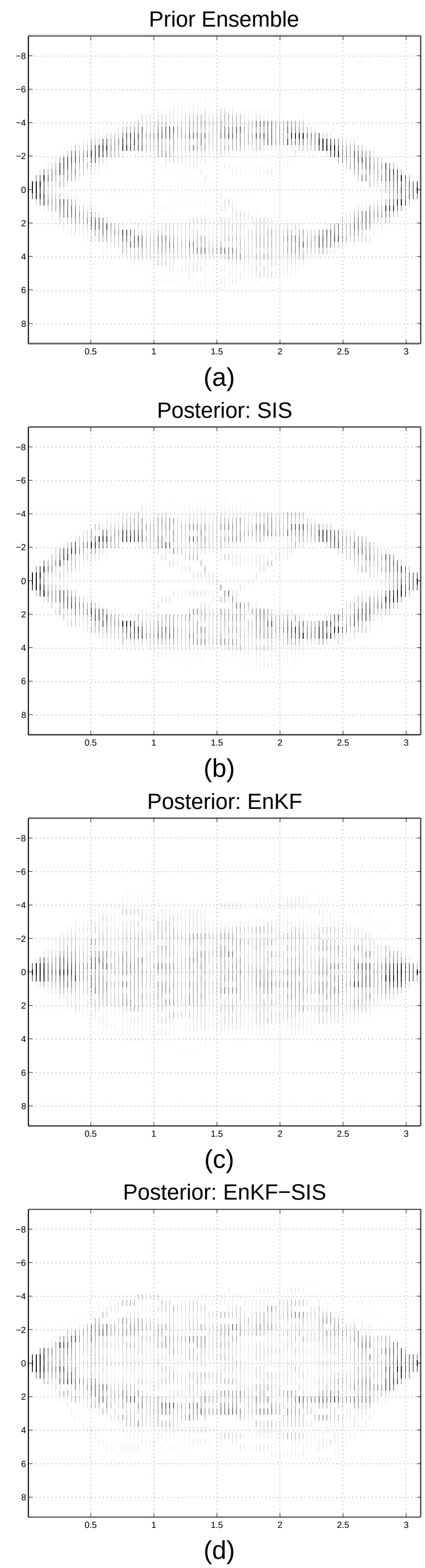

Figure 5: EnKF does not recognize a non-gaussian prior.

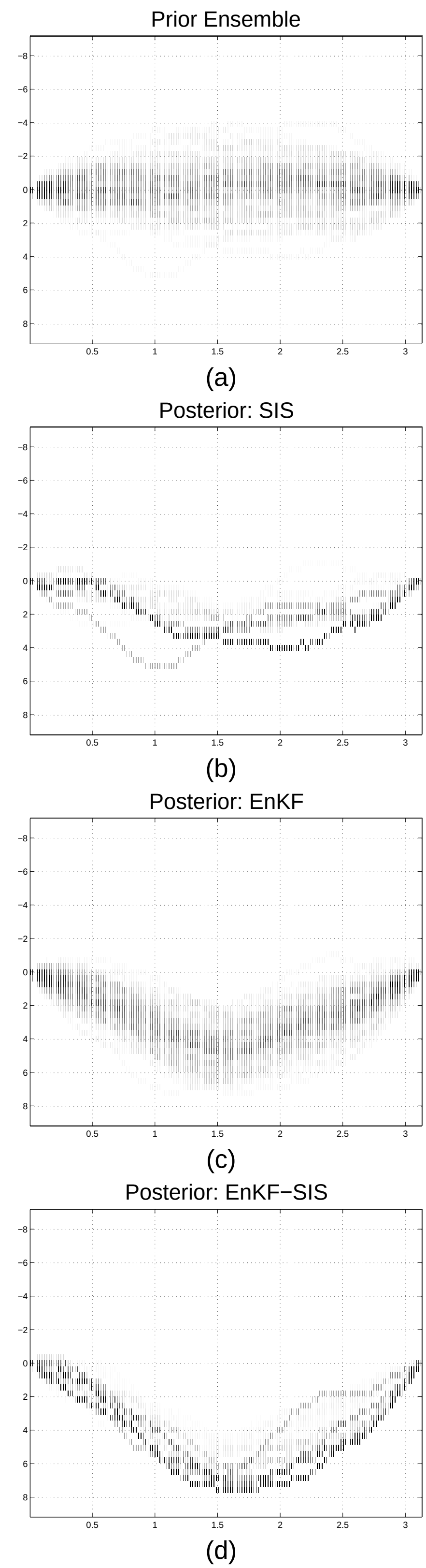

Figure 6: SIS cannot make a large update.

4b. *Filtering in High Dimension*

Typical results for filtering in the space of functions on $[0, \pi]$ of the form

$$u = \sum_{n=1}^{d} c_n \sin(nx) \tag{18}$$

are in Figs. 5 and 6. Note that (18) is a discrete Fourier transform, which maps the d-dimensional space of function on a uniform 1D mesh and zero boundary conditions on the space of the coefficients c_n. In all panels, the horizontal axis is the spatial coordinate $x \in [0, \pi]$. The vertical axis is the value of u. The level of shading on each vertical line is the marginal density of u at a fixed x, computed from a histogram with 50 bins. The ensemble size was $N = 100$ and the dimension of the state space was $d = 500$. The eigenvalues of the covariance were chosen $\lambda_n = n^{-3}$ to generate the initial ensemble and $\lambda_n = n^{-2}$ for density estimation. The initial ensemble was obtained from (12) with $u_0 = 0$.

Fig. 5 again shows that EnKF cannot handle bimodal distribution. The prior was constructed by assimilating the data likelihood

$$p(d|u) = \begin{cases} 1/2 \text{ if } u(\pi/4) \text{ and } u(3\pi/4) \in (-2,-1) \cup (1,2) \\ 0 \text{ otherwise} \end{cases}$$

into a large initial ensemble (size 50000) and resampling to the obtain the forecast ensemble size 100 with a non-gaussian density. Then the data likelihood

$$u(\pi/2) - 0.1 \sim N(0,1)$$

was assimilated to obtain the analysis ensemble.

Fig. 6 shows a failure of SIS. The prior ensemble was same as the initial ensemble, that is gaussian, and the data likelihood was

$$u(\pi/2) - 7 \sim N(0,1).$$

We observe that while EnKF and EnKF-SIS create ensembles that are attracted to the point $(\pi/2, 7)$, SIS cannot reach so far because there are no such members in this relatively small ensemble of size 100.

4c. *Filtering for a Stochastic ODE*

The results given above describe how each filtering technique applies to certain carefully designed synthetic

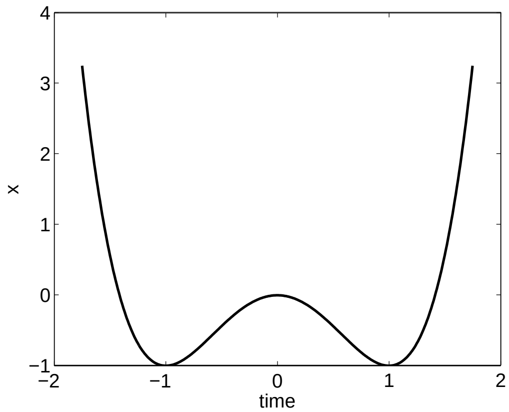

Figure 7: Deterministic potential curve

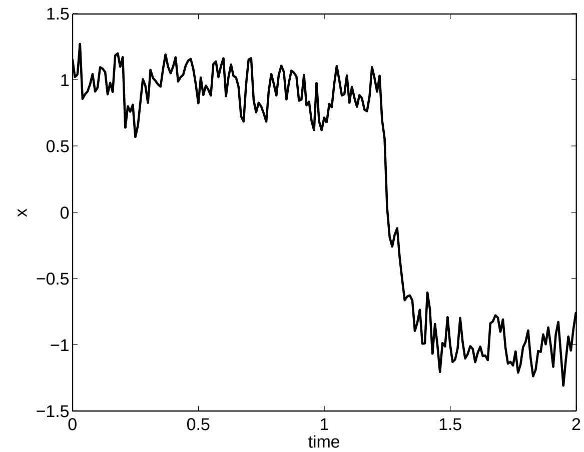

Figure 8: A solution of (19) switching between stable points

distributions. It remains to be seen how these filters work when applied to an actual model. We have implemented a stochastic differential equation as used in (Kim et al. 2003),

$$\dot{x} = 4x - 4x^3 + \kappa\eta, \tag{19}$$

where $\eta(t)$ is a stochastic variable representing white noise, independent samples from a gaussian distribution with zero mean and covariance $E[\eta(t)\eta(t')] = \delta(t - t')$, which represents independent observations in time. The parameter κ controls the magnitude of the stochastic term.

The deterministic part of this differential equation is of the form $\dot{x} = -f'(x)$, where the potential $f(x) = -2x^2 + x^4$, cf., Fig 7. The equilibria are given by $f'(x) = 0$; there are two stable equilibria at $x = \pm 1$ and an unstable equilibrium at $x = 0$. The stochastic term of the differential equation makes it possible for the state to flip from one stable equilibrium point to another; however, a sufficiently small κ insures that such an event (Fig 8) is rare.

A suitable test for an ensemble filter will be determining if it can properly track the model as it transitions from one stable fixed point to the other. From the previous results, it is clear that EnKF will not be capable of describing the bimodal nature of the distributions involved so it will not perform well when tracking the transition. When the initial ensemble is centered around one stable point, it is unlikely that some ensemble members be close to the other stable point, so SIS will be even slower in tracking the transition (Kim et al. 2003).

The solution x of (19) is a random variable dependent on time, with density $p(t, x)$,

$$\Pr(x(t) < x_0) = \int_{-\infty}^{x_0} p(t, x) dx.$$

The evolution of the density in time is given by the Fokker-Planck equation (Miller et al. 1999),

$$\frac{\partial p}{\partial t} = \frac{\partial}{\partial x}[4x(x^2 - 1)p] + \frac{\sqrt{\kappa}}{2}\frac{\partial^2 p}{\partial x^2}. \tag{20}$$

To obtain the exact (also called optimal) solution to the bayesian filtering problem, up to a numerical truncation error, we have advanced the probability density of x between the bayesian updates by solving the Fokker-Planck equation (20) numerically on a uniform mesh from $x = -3$ to $x = 3$ with the step $\Delta x = 0.01$, using the MATLAB function pdepe. At the analysis time, we have multiplied the probability density of x by the data likelihood as in (7) and then scaled so that again $\int p dx = 1$, using numerical quadrature by the trapezoidal rule.

t	1	2	3	4	5	6
x	1.2	1.3	-0.1	-0.6	-1.4	-1.2

Table 1: Data used in assimilation into (19)

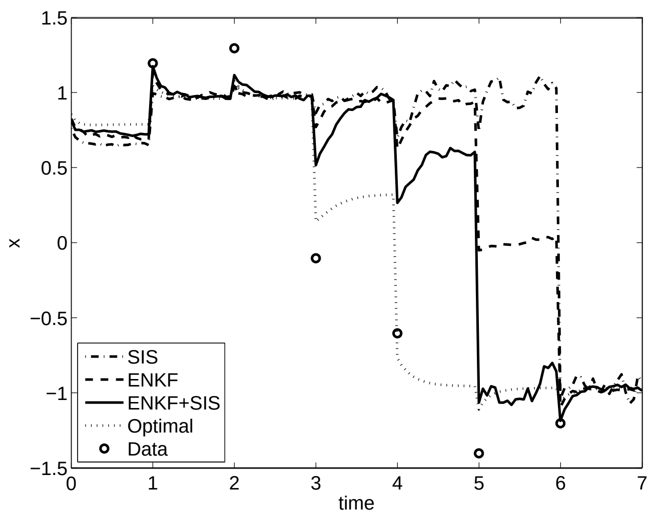

Figure 9: Ensemble filters mean and optimal filter mean for (19)

The data points (Table 1) were taken from one solution of this model, called a reference solution, which exhibits a switch at time $t \approx 1.3$. The data error distribution was normal with the variance taken to be 0.1 at each point. To advance the ensemble members and the reference solution, we have solved (19) by the explicit Euler method with a random perturbation from $N(0, (\Delta t)^{1/2})$ added to the right hand side in every step (Higham 2001). The simulation was run for each method with ensemble size 100, and assimilations performed for each data point.

Examining the results in Fig. 9, it is clear that EnKF-SIS was able to approximate the optimal solution better than either SIS or EnKF alone. EnKF provided the smoothest approximation; however, it is unable to track the data quickly as it switches. SIS suffers from a lot of noise as only a small number of ensemble members contribute to the posterior. In addition, SIS provides even worse tracking of the data than EnKF. The hybrid EnKF-SIS method provides the best tracking in this case, but it exhibits some noise in the solution, because the proposal and forecast distributions are far apart. This noise is similar to that seen with SIS, but less severe.

5. Conclusion

We have introduced a new class of ensemble filters, combining the EnKF and the SIS Filter. We have motivated the new filters by probability on infinitely dimensional spaces of smooth functions, which occur naturally in the solution of partial differential equations. We have demonstrated the potential of the new filters to perform better than existing filters for non-gaussian distributions, large updates, and in tracking of a solution switching between two stable states with sparse data.

We have also observed that on spaces of smooth functions, the SIS method itself often works quite well already, as long as the updates are not too large, even in high dimension. The analysis of this fact will be pursued elsewhere.

6. Acknowledgement

The authors are grateful to Doug Nychka, Craig Johns and Tolya Puhalskii for useful discussions.

A. Probability on Infinitely Dimensional Spaces

A random function (a random field, or a stochastic process) $X(x, \omega)$ is a function of a deterministic variable $x \in D$ and a stochastic event $\omega \in \Omega$, where D is a domain in space, time, or both, or a finite set, and Ω is a probability event space. We assume that for a fixed ω, the function $x \mapsto X(x, \omega)$ is an element $X(\omega)$ of a normed linear space V of functions on $\mathcal{D}$, and that $\omega \mapsto X(\omega)$ is a measurable function; that is, X is a random variable with values in V. The random variable X induces a probability measure μ on V by $\mu(S) = \Pr(\{\omega : X(\omega) \in S\})$. The mean $E(x)$ and the covariance $C(x,y)$ of the random field X and of the measure μ are same, and they are given by

$$\begin{aligned} E(x) &= \int_\Omega X(x,\omega)\,d\omega, \\ C(x,y) &= \int_\Omega (X(x,\omega)-E(x))\,(X(y,\omega)-E(y))\,d\omega, \end{aligned}$$

for $x,y\in\mathcal{D}$, cf., e.g., Loève (1978).

Aa. *Gaussian Measures on Infinitely Dimensional Spaces*

A gaussian probability measure μ on a Hilbert space V can be defined from the eigenvalues and eigenvectors of its covariance C in terms of a norm on a compactly embedded subspace, called the Cameron-Martin space (Bogachev 1998), as follows. Suppose that λ_n, e_n,

$$\int_{\mathcal{D}} C(x,y)\,e_n(x)\,dx = \lambda_n e_n(y)\,, \quad y\in\mathcal{D}, \tag{21}$$

are all eigenvalues and eigenvectors of C, and

$$\lambda_n \geq 0, \quad \sum_{n=1}^{\infty}\lambda_n < +\infty.$$

The eigenvectors e_n can be always chosen orthonormal in V. The Cameron-Martin space $U\subset V$ of the measure μ is the Hilbert space

$$\begin{aligned} U &= \left\{v\in V : \|v\|_U^2 < +\infty\right\} \qquad (22) \\ \|v\|_U^2 &= \sum_{n=1}^{\infty}\frac{|v_n|^2}{\lambda_n}, \quad v=\sum_{n=1}^{\infty}v_n e_n. \end{aligned}$$

For $d\in\mathbb{N}$, define finite dimensional subspaces of U,

$$U_d = \left\{v\in V : v=\sum_{n=1}^{d}v_n e_n,\ v_n\in\mathbb{R}\right\}\subset U\subset V. \tag{23}$$

and, for a Borel set $S_d\subset U_d$, let $\widehat{S}_d$ be the cylindrical set

$$\widehat{S}_d = \left\{\sum_{n=1}^{\infty}v_n e_n\in V : \sum_{n=1}^{d}v_n e_n\in S_d\right\}.$$

A gaussian measure γ on V centered at zero is then the unique extension of the measure on cylindrical sets, which is defined by the corresponding finite dimensional gaussian measures of the cylinder bases S_d,

$$\begin{aligned} \gamma\left(\widehat{S}_d\right) &= \mu(S_d) = \int_{S_d}\prod_{n=1}^{d}\frac{e^{-\frac{|v_n|^2}{2\lambda_n}}}{\sqrt{2\pi\lambda_n}}dv_1\ldots dv_d \qquad (24) \\ &= c_d\int_{S_d}\varphi\left(\|w\|_U\right)dw, \qquad (25) \end{aligned}$$

where

$$\begin{aligned} c_d &= \prod_{m=1}^{d}\frac{1}{\sqrt{2\pi\lambda_m}}, \\ \varphi\left(\|w\|_U\right) &= e^{-\frac{\|w\|_U^2}{2}}, \quad w\in U. \qquad (26) \end{aligned}$$

cf., (Bogachev 1998; Kuo 1975).

Note that the Lebesgue measure does not exist on an infinitely dimensional space (Kuo 1975), and so one cannot have a density with respect to the Lebesgue measure as in finite dimensional spaces. However, *the kernel* (26) *still plays the role of a density kernel even in infinite dimension*. The Sobolev norm $\|v\|_{H^s(D)}$ can be used as the distance which is naturally related to a gaussian measure through the density kernel (26). Other measures can be then studied by using their densities with respect to a gaussian measure, which is a common approach in probability on abstract spaces (Araujo and Giné 1980; Ledoux and Talagrand 1991).

Ab. *Random Functions and Sobolev Spaces*

Generation of smooth random functions by Fast Fourier Transform (FFT), well known in geosciences (Evensen 1994; Ruan and McLaughlin 1998), is in fact sampling from a gaussian distribution on a space of smooth functions, with the degree of smoothness determined by the Fourier basis e_n and the decay of the eigenvalues λ_n of the covariance from (21). A random variable u_d distributed according to the gaussian measure (24) on the finite dimensional space U_d is

$$u_d = \sum_{n=1}^{d}\frac{v_n e_n}{\lambda_n}, \quad \nu_n\sim N(0,1)\,, \tag{27}$$

where ν_n are i.i.d. In the limit for the dimension $d\to\infty$, we have the random variable u on the infintely dimensional space V,

$$u = \sum_{n=1}^{\infty} \frac{v_n e_n}{\lambda_n}, \quad \nu_n \sim N(0,1).$$

Thus, $u \in U$ almost surely.

Sobolev spaces and their norms (Adams 1975) can be written in the form (22). For example, consider the space of functions on $(0,\pi)$ with square integrable generalized derivatives of order up to $k > 0$ and zero boundary values, equipped with the Sobolev norm of order k, defined by

$$\|u\|_{H^k}^2 = \int_0^\pi |u|^2 + |u^{(k)}|^2 dx.$$

Expanding u in the Fourier series

$$u(x) = \sum_{n=1}^{\infty} u_n \sin(nx),$$

it is easy to see that

$$\|u\|_{H^k}^2 = \frac{2}{\pi} \sum_{n=1}^{\infty} \left(1+n^{2k}\right) |u_n|^2, \quad u \in U, \tag{28}$$

which is a norm of the form (22), with $\lambda_n = 1/\left(1+n^{2k}\right)$.

Ac. *Nonparametric Density Estimation*

Density estimation in high dimension is notoriously difficult and considered intractable in dimensions higher than about 10 (Terrell and Scott 1992). However, there are nonparametric approaches that depend only on the concept of distance and are applicable in the infinite dimensional case as well, though they involve deep issues in probability theory. Consider a normed linear space V equipped with a measure μ and Cameron-Martin space $U \subset V$. Given a suitable kernel function $\varphi : \mathbb{R} \to \mathbb{R}$ consider the kernel

$$\varphi(\|w\|_U).$$

We will assume that

$$\varphi \geq 0, \qquad \int_U \varphi(\|w\|_U)\, d\mu < +\infty. \tag{29}$$

Standard examples include $\varphi(x) = e^{-x^2/2}$, which gives the gaussian kernel (26), and the "naive" kernel function,

$$\varphi(x) = \begin{cases} 1 \text{ if } |x| \leq 1, \\ 0 \text{ if } |x| > 1. \end{cases} \tag{30}$$

Given a probability measure ν on V with density p_ν with respect to ν and a sample $(u_1, \ldots, u_N) \sim \nu$, consider the density estimate

$$p_{N.h}(y) = \frac{\sum_{k=1}^N \frac{1}{N} \varphi\left(\frac{\|y-u_k\|_U}{h}\right)}{\int_U \varphi\left(\frac{\|w\|_U}{h}\right) d\mu(w)} \approx p_\nu(y). \tag{31}$$

The parameter $h = h(N) > 0$ is called bandwidth.

In the finite dimensional case, with μ the Lebesgue measure on $\mathbb{R}^d$, it holds that

$$\int_U \varphi\left(\frac{\|w\|_U}{h}\right) d\mu(w) = h^d \int_U \varphi(\|w\|_U)\, d\mu,$$

and so it is customary to assume in the literature that φ is scaled so that $\int_U \varphi(\|w\|_U)\, d\mu = 1$. Convergence of the kernel estimate and the choice of the bandwidth in this case have been extensively studied, see Scott (1992); Scott and Sain (2005); Terrell and Scott (1992), and references therein.

For the naive kernel given by (30),

$$\begin{aligned} \int_U \varphi\left(\frac{\|w\|_U}{h}\right) d\mu(w) &= \mu(B_U(h)), \\ B_U(h) &= \{w \in U : \|w\|_U \leq h\}. \end{aligned}$$

In the infinite dimensional case, Lebesgue measure does not exist, so another measure μ must be chosen, such as a gaussian measure. Then (29) holds for any bounded φ because $\mu(U) = 1$. However, for other measures than the Lebesgue measure, $\mu(B_U(h))$ cannot be easily evaluated. The study of the behavior of $\mu(B_U(h))$ as $h \to 0$ is known as the small deviations problem or the small ball probability problem in probability theory (Christensen 1978; Lifshits and Linde 2005; Li and Linde 1999), and it is an active area of research.

Dabo-Niang (2004) has proved that for the naive kernel, the estimate (31) converges in mean square when

$$\lim_{N\to+\infty} h = 0, \quad \lim_{N\to+\infty} N\mu(B_U(h)) = +\infty.$$

In finite dimension, Loftsgaarden and Quesenberry (1965) have shown that the estimate (15) converges in probability when h is chosen as the distance to the $k(N)$-th nearest sample point to y, and

$$\lim_{N\to+\infty} k\,(N) = +\infty, \quad \lim_{N\to+\infty} \frac{k\,(N)}{N} = 0.$$

A common choice is $k(N) \approx N^{\alpha}$, $0 < \alpha < 1$. For other convergence results for density estimation methods, see Bosq and Lecoutre (1987).

References

Adams, R. A., 1975: *Sobolev spaces*. Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London, xviii+268 pp., pure and Applied Mathematics, Vol. 65.

Anderson, B. D. O. and J. B. Moore, 1979: *Optimal filtering*. Prentice-Hall, Englewood Cliffs, N.J., x, 357 pp.

Anderson, J. L., 1999: An ensemble adjustment Kalman filter for data assimilation. *Monthly Weather Review*, **129**, 2884–2903.

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

Araujo, A. and E. Giné, 1980: *The central limit theorem for real and Banach valued random variables*. Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, New York-Chichester-Brisbane, xiv+233 pp.

Bengtsson, T., C. Snyder, and D. Nychka, 2003: Toward a nonlinear ensemble filter for high dimensional systems. *Journal of Geophysical Research - Atmospheres*, **108(D24)**, STS 2–1–10.

Bennett, A. F., 1992: *Inverse Methods in Physical Oceanography*. Cambridge University Press.

Bogachev, V. I., 1998: *Gaussian measures*, volume 62 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, xii+433 pp.

Bosq, D. and J.-P. Lecoutre, 1987: *Theorie de l'estimation fonctionelle*. Economica, Paris.

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

Chen, R. and J. S. Liu, 2000: Mixture Kalman filters. *J. of the Royal Statistical Society: Series B*, **62**, 493–508.

Christensen, J. P. R., 1978: The small ball theorem for Hilbert spaces. *Math. Ann.*, **237**, 273–276.

Dabo-Niang, S., 2004: Density estimation by orthogonal series in an infinite dimensional space: application to processes of diffusion type I. *J. Nonparametr. Stat.*, **16**, 171–186, the International Conference on Recent Trends and Directions in Nonparametric Statistics.

de Valpine, P., 2004: Monte Carlo state-space likelihoods by weighted posterior kernel density estimation. *J. Amer. Statist. Assoc.*, **99**, 523–536.

Doucet, A., N. de Freitas, and N. Gordon, eds., 2001: *Sequential Monte Carlo in Practice*. Springer.

Efromovich, S., 2004: Distribution estimation for biased data. *J. Statist. Plann. Inference*, **124**, 1–43.

Evensen, G., 1994: Sequential data assimilation with nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, **99 (C5)**, 143–162.

— 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, **53**, 343–367.

— 2004: Sampling strategies and square root analysis schemes for the EnKF. *Ocean Dynamics*, 539–560.

Gordon, N. J. and A. F. M. Smith, 1993: Approximate non-Gaussian Bayesian estimation and modal consistency. *J. Roy. Statist. Soc. Ser. B*, **55**, 913–918.

Higham, D. J., 2001: An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Rev.*, **43**, 525–546.

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

Johns, C. J. and J. Mandel, in print: A two-stage ensemble Kalman filter for smooth data assimilation, environmental and Ecological Statistics. Conference on New Developments of Statistical Analysis in Wildlife, Fisheries, and Ecological Research, Oct 13-16, 2004, Columbia, MI.

Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *Transactions of the ASME – Journal of Basic Engineering, Series D*, **82**, 35–45.

Kalman, R. E. and R. S. Bucy, 1961: New results in filtering and prediction theory. *Transactions of the ASME – Journal of Basic Engineering*, **83**, 95–108.

Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press.

Kim, S., G. L. Eyink, J. M. Restrepo, F. J. Alexander, and G. Johnson, 2003: Ensemble filtering for nonlinear dynamics. *Monthly Weather Review*, **131**, 2586–2594.

Kuo, H. H., 1975: *Gaussian measures in Banach spaces*. Springer-Verlag, Berlin, vi+224 pp., lecture Notes in Mathematics, Vol. 463.

Ledoux, M. and M. Talagrand, 1991: *Probability in Banach spaces*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, xii+480 pp.

Li, W. V. and W. Linde, 1999: Approximation, metric entropy and small ball estimates for Gaussian measures. *Ann. Probab.*, **27**, 1556–1578.

Lifshits, M. A. and W. Linde, 2005: Small deviations of weighted fractional processes and average non-linear approximation. *Trans. Amer. Math. Soc.*, **357**, 2059–2079.

Loève, M., 1978: *Probability theory. II*, volume 46 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, fourth edition, xvi+413 pp.

Loftsgaarden, D. O. and C. P. Quesenberry, 1965: A nonparametric estimate of a multivariate density function. *Ann. Math. Statist.*, **36**, 1049–1051.

Miller, R. N., E. F. Carter, and S. T. Blue, 1999: Data assimilation into nonlinear stochastic models. *Tellus*, **51A**, 167–194.

Pham, D. T., 2001: Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Monthly Weather Review*, **129**, 1194–1207.

Ruan, F. and D. McLaughlin, 1998: An efficient multivariate random field generator using the fast Fourier transform. *Advances in Water Resources*, **21**, 385–399.

Scott, D. W., 1992: *Multivariate density estimation*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons Inc., New York, xiv+317 pp., theory, practice, and visualization, A Wiley-Interscience Publication.

Scott, D. W. and S. R. Sain, 2005: Multidimensional density estimation. *Handbook of Statististics 24: Data Mining and Visualization*, C. R. Rao, E. J. Wegman, and J. L. Solka, eds., Elsevier, Amsterdam, 229–263.

Terrell, G. R. and D. W. Scott, 1992: Variable kernel density estimation. *Ann. Statist.*, **20**, 1236–1265.

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Monthly Weather Review*, **131**, 1485–1490.

Xiong, X. and I. Navon, Tellus, submitted: Ensemble particle filter with posterior Gaussian resampling.

CENTER FOR COMPUTATIONAL MATHEMATICS REPORTS

University of Colorado at Denver and Health Sciences Center
P.O. Box 173364, Campus Box 170
Denver, CO 80217-3364
Fax: (303) 556-8550
Phone: (303) 556-8442
http://www-math.cudenver.edu/ccm

221 Craig Johns and Jan Mandel, "A Two-Stage Ensemble Kalman Filter for Smooth Data Assimilation." April 2005

222 I. Lashuk, M. E. Argentati, E. Ovchinnikov and A. V. Knyazev, "Preconditioned Eigensolver LOBPCG in hypre and PETSc." September 2005.

223 A. V. Knyazev and M. E. Argentati, "Majorization for Changes in Angles Between Subspaces, Ritz values, and graph Laplacian spectra." September 2005.

224 D. W. Dean, "A Semimartingale Approach For Modeling Multiphase Flow In Heterogeneous Porous Media." November 2005.

225 D. W. Dean, T.H. Illangasekare and T.F. Russell, "A Stochastic Differential Equation Approach For Modeling Of NAPL Flow In Heterogeneous Porous Media." November 2005.

226 J. Mandel and B. Sousedík, "Adaptive Coarse Space Selection in the BDDC and the FETI-DP Iterative Substructuring Methods: Optimal Face Degrees of Freedom." November 2005.

227 J. Mandel and B. Sousedík, "Adaptive Selection of Face Coarse Degrees of Freedom in the BDDC and the FETI-DP Iterative Substructuring Methods." November 2005.

228 Craig C. Douglas, Jonathan D. Beezley, Janice Coen, Deng Li, Wei Li, Alan K. Mandel, Jan Mandel, Guan Qin, and Anthony Vodacek, "Demonstrating the Validity of a Wildfire DDDAS." February 2006.

229 Lynn Schreyer Bennethum, "Flow and Deformation: Understanding the Assumptions and Thermodynamics, Updated." February 2006.

230 Lynn Schreyer Bennethum, "Compressibility Moduli for Porous Materials Incorporating Volume Fraction with Detailed Computations." February 2006.

231 Jan Mandel, "Efficient Implementation of the Ensemble Kalman Filter." May 2006.

232 Jan Mandel and Jonathan D. Beezley, "Predictor-Corrector Ensemble Filters for the Assimilation of Sparse Data into High Dimensional Nonlinear Systems." June 2006.