PREDICTOR-CORRECTOR AND MORPHING ENSEMBLE FILTERS FOR THE ASSIMILATION OF SPARSE DATA INTO HIGH-DIMENSIONAL NONLINEAR SYSTEMS *

Jan Mandel† and Jonathan D. Beezley
University of Colorado at Denver and Health Sciences Center, Denver, CO,
and National Center for Atmospheric Research, Boulder, CO

December 2, 2006

1. INTRODUCTION

A new family of ensemble filters, called predictor-corrector filters, is introduced. The predictor-corrector filters use a proposal ensemble (obtained by some method, called the predictor) with assignment of importance weights to recover the correct statistic of the posterior (the corrector). The proposal ensemble comes from an arbitrary unknown distribution, and it only needs to have a good coverage of the support of the posterior. The ratio of the prior and the proposal densities for calculating the importance weights is obtained by density estimation. Predictor by the ensemble Kalman formulas and corrector by nonparametric density estimation based on the distance in Sobolev spaces are considered. Numerical experiments show that the new predictor-corrector filters combine the advantages of ensemble Kalman filters and particle filters for highly nonlinear systems, and that they are suitable for high dimensional states which are discretizations of smooth functions.

We also propose another class, called morphing ensemble filters, which combine ensemble Kalman filters with the ideas of morphing and registration from image processing. In the morphing ensemble filters, the ensemble members are represented by the composition of one fixed template function with a morphing function, and by adding a residual function. The ensemble Kalman formulas operate on the transformed state consisting of the morphing function and the residual function. This results in filters suitable for problems with sharp thin moving interfaces, such as in wildfire modeling.

The research reported here has been motivated by data assimilation into wildfire models (Mandel et al. 2005, 2006; Douglas et al. 2006). Wildfire modeling presents a challenge to data assimilation because of non-gaussian probability distributions centered around the burning and not burning states, and because of movements of thin reaction fronts with sharp interfaces. Standard EnKF approach (Evensen 2003) fails because it is limited to making linear combinations of states, which results in nonphysical states. This can be ameliorated by regularization (Johns and Mandel 2005), but large ensembles are required and the method does not work reliably when large updates of the state are required due to model errors or infrequent data (Mandel et al. 2006).

The work presented here will be useful in wildfire modeling as well as in data assimilation for other problems with strongly non-gaussian distributions and moving sharp interfaces and features, such as hurricanes.

2. OVERVIEW OF THE STATE SPACE MODEL AND BAYES THEOREM

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,

$$p_a(u) = p(u|d) \propto p(d|u)\, p_f(u). \qquad (1)$$

Here and below, $\propto$ means proportional, $p_f(u)$ 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 model is then advanced until the next analysis time. The data

*11th Symposium on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface (IOAS-AOLS), CD-ROM, Paper 4.12, 87th American Meterological Society Annual Meeting, San Antonio, TX, January 2007. UCDHSC CCM Report 239, November 2006. www.math.cudenver.edu/ccm/reports

†Corresponding author. Department of Mathematical Sciences, University of Colorado at Denver and Health Sciences Center, Denver CO 80217-3364; email: Jan.Mandel@cudenver.edu

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

$$d - h(u) \sim \varepsilon. \tag{2}$$

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.

3. PREDICTOR-CORRECTOR FILTERS

3.1. *Particle Filters*

We first review Particle Filters (PFs) following (Doucet et al. 2001, Ch. 1). PFs approximate the state by a weighted ensemble in the sense of Monte Carlo quadrature as follows. Suppose the probability distribution μ of the state has density $p_\mu(u)$ with respect to some underlying measure ν on the state space U, $(u_k)_{k=1}^N$ is a sample from another distribution π, called the proposal distribution, with density p_π with respect to ν, and f is a real function on U integrable with respect to μ. Then

$$\begin{aligned} \int_U f p d\nu &= \int_U f \frac{p_\mu}{p_\pi} p_\pi d\nu = \int_U f \frac{p_\mu}{p_\pi} d\pi \\ &\approx \sum_{k=1}^N w_k f(u_k), \quad w_k = \frac{1}{N} \frac{p_\mu(u_k)}{p_\pi(u_k)}, \end{aligned} \tag{3}$$

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

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

In this sense, the weighted ensemble $(u_k, w_k)_{k=1}^N$,

$$(u_1, \ldots, u_N) \sim \pi, \quad w_k \propto \frac{p_\mu(u_k)}{p_\pi(u_k)},$$

represents the probability distribution μ, and we will write

$$(u_k, w_k) \sim p_\mu.$$

Now suppose p_f is the forecast density and we are given a sample (u_k^a) from some probability distribution π with density p_π, called *proposal* density. Then, from (3) and (1), the analysis density p_a is represented by the weighted ensemble $(u_k^a, w_k^a) \sim p_a$,

$$w_k^a \propto \frac{p_a(u_k^a)}{p_\pi(u_k^a)} = p(d|u_k^a) \frac{p_f(u_k^a)}{p_\pi(u_k^a)}. \tag{4}$$

While the data likelihood $p(d|u)$ can be readily evaluated from (2), the ratio of the forecast and the proposal densities is in general not available, except in the particular case when the forecast is given as a weighted ensemble $(u_k^f, w_k^f) \sim p_f$ 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.

The problem with SIS 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, most of the analysis weights degenerate to numerical zeros and the effective size of the ensemble decreases. 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 all weights equal. Stochastic advance in time (Markov chain) is then relied upon to spread the ensemble again. In (Kim et al. 2003; Xiong and Navon 2005; Xiong et al. 2006), the resampling is done by first estimating the density p_a, and then generating the new ensemble by random sampling from the estimated density. In (van Leeuwen 2003), the use of Cauchy distribution, which has thicker tails that the normal distribution, is suggested to alleviate the degeneracy problem.

3.2. *Density Ratio Estimation*

When a large change of state is required in the analysis step, SIS with a modestly sized ensemble fails, because it has few or no ensemble members where the posterior is concentrated. But it is possible to get proposal ensembles that can be expected to cover the posterior better by other means, such as the EnKF (Sec. 3.3 below). The EnKF is able to make large changes of the state in the analysis step because it relies on linear algebra to create the analysis ensemble by a least squares fit in the span of the forecast ensemble.

So, given an arbitrary proposal ensemble, we replace the ratio of the densities in (4) by the nonparametric estimate

$$\frac{p_f(u_k^a)}{p_\pi(u_k^a)} \approx \frac{\sum_{\ell: \|u_\ell^f - u_k^a\|_H \le h} w_k^f}{\sum_{\ell: \|u_\ell^a - u_k^a\|_H \le h} 1/N_\pi}, \tag{5}$$

where N_π is the number of members of the proposal ensemble, $h > 0$ is a bandwidth parameter, and $\|\cdot\|_H$ is a suitable norm. This estimate uses only the concept of distance and it is inherently dimension independent. We propose to choose the bandwith $h = h(u_k^a)$ so that it is the distance to the $\lfloor N_\pi^{1/2} \rfloor$-th nearest u_ℓ^a in the $\|\cdot\|_H$ norm.

It remains to choose the norm $\|\cdot\|_H$. In the scalar case, the norm $\|\cdot\|_H$ is simply the absolute value. We are particularly interested in the case when the ensemble consists of (discrete representations of) smooth functions. In that case, the norm $\|\cdot\|_H$ is related to the probability distribution of the ensemble as follows. A well-known way (Ruan and McLaughlin 1998; Evensen 1994) to construct smooth random functions for the initial ensemble is from an orthonormal basis $\{\varphi_n\}$ of U and coefficients $\lambda_n > 0$, $\sum_n \lambda_n < \infty$, by

$$u = \sum_n c_n \varphi_n, \quad c_n = \lambda_n d_n, \quad d_n \sim N(0,1). \tag{6}$$

Possible choices of $\{\varphi_n\}$ include a Fourier basis, such as sine or cosine functions, or bred vectors (Kalnay 2003). In any case, with increasing n, the basis functions φ_n are more oscillatory, and they contribute less to the sum since $\lambda_n \to 0$. Because of the fast rate of decay of the coefficients λ_n, the sum (6) defines random smooth functions. The function u defined by (6) is a random variable with values in the state space U distributed according to a gaussian measure on U and λ_n are the eigenvalues of the covariance. The gaussian measure is associated with the so-called Cameron-Martin space (Bogachev 1998; Kuo 1975)

$$V = \left\{ v : \|v\|_V^2 < \infty \right\}$$

where

$$\|u\|_V^2 = \sum_n \frac{|c_n|^2}{\lambda_n}, \quad u = \sum_n c_n \varphi_n. \tag{7}$$

Similarly as above, functions $u \in V$ are smooth because of the fast rate of decay of the coefficients λ_n for large n. It is well known that solutions of partial differential equations are naturally in spaces with the norms of the form (7). For example, the solution of the Laplace equation in 1D, $u'' = f$ in $(0,\pi)$, with $f \in L^2(0,\pi)$ and boundary conditions $u(0) = u(\pi) = 0$ is in the Sobolev space $H^s(0,\pi)$, for $s \leq 2$. The space $H^s(0,\pi)$ is equipped with the norm (7) with $\varphi_n(x) = (1/2)\sin nx$, $x \in (0,\pi)$, and $\lambda_n = 1/\left(1+n^{2s}\right)$, so the condition $\sum_n \lambda_n < \infty$ is satisfied if $s > 1/2$.

We choose the norm $\|\cdot\|_H$ as some norm which is not stronger than the norm $\|\cdot\|_V$, such as

$$\|u\|_H^2 = \sum_n \left| \frac{c_n}{\kappa_n} \right|^2, \quad u = \sum_n c_n \varphi_n, \tag{8}$$

with $\kappa_n \geq \lambda_n$.

We now derive the density ratio estimate (5), which will provide a heuristic motivation for the above choice of the bandwidth and the norm. Mathematical investigation of the convergence of the estimate (5) is an open problem. If $(u_k)_{k=1}^N$ is a sample from a probability distribution μ, which has density f_μ with respect to the measure ν and f_μ is continuous at x, then

$$f_\mu(x) \approx \frac{|\{k : u_k \in B_H(x,h)\}|}{N\nu(B_H(x,h))}, \tag{9}$$

where

$$B_H(x,h) = \{u \in H : \|u - x\|_H \leq h\} \tag{10}$$

is the ball with center x and radius h in the norm $\|\cdot\|_H$. Dividing the estimates (9) for the forecast and the proposal densities gives the density ratio estimate (5). Lebesgue measure does not exist in infinitely dimensional spaces, so one has to choose the underlying measure ν as a gaussian measure. Computing the measure of the ball is in general not possible, but, fortunately, the measure of the ball cancels.

The main assumptions behind deriving the density estimate was that the forecast and the proposal densities with respect to the measure ν exist and are finite, and that the measure of the ball (10) is positive. This is by far not automatic in infinite dimension, and restricts the choice of the norm $\|\cdot\|_H$. By the Hajek-Feldman theorem, every two Gaussian measures on an infinitely dimensional space are either equivalent or mutually singular, and if they are equivalent, they have the same Cameron-Martin space (Bogachev 1998). Therefore, considering the case when p_f and p_a are densities of gaussian measures equivalent to the measure the initial ensemble was sampled from, the measure ν needs to be chosen equivalent to that measure. The requirement that the norm $\|\cdot\|_H$ is not stronger than the norm of the Cameron-Martin space is then given by the requirement that the measure of the ball (10) is positive.

The choice of the bandwidth h is motivated by known convergence results for the density estimate (9). In finite dimension, Loftsgaarden and Quesenberry (1965) have shown that the density estimate (9) converges in probability when h is

chosen as the distance to the $k(N)$-th nearest sample point to y, and

$$\lim_{N\to+\infty} k\,(N) = +\infty, \quad \lim_{N\to+\infty} \frac{k\,(N)}{N} = 0. \qquad (11)$$

Dabo-Niang (2004) proved that for probability distributions on a Banach space, the estimate (9) converges in mean square when

$$\lim_{N\to+\infty} h = 0, \quad \lim_{N\to+\infty} N\nu\,(B_H\,(x,h)) = +\infty, \quad (12)$$

and some other technical conditions are satisfied. It would be interesting to establish which choice of the bandwidth h implies (12) and thus convergence of the density estimate in the infinitely dimensional case.

3.3. *Weighted EnKF*

We first state for reference the Ensemble Kalman Filter (EnKF) following Burgers et al. (1998). 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). However, the Kalman filter involves manipulating and advancing the covariance matrix, which is prohibitively expensive for large problems. In the EnKF, the forecast density p_f is assumed to be gaussian and it is represented by the forecast ensemble with all weights equal. The data likelihood $p\,(d|u)$ is also assumed to be gaussian, $d - Hu \sim N\,(0, R)$. The members of the analysis ensemble are then obtained from

$$u_k^a = u_k^f + K(d_k - Hu_k^f), \qquad (13)$$

where d_k is a randomly perturbed data vector sampled from $N\,(d, R)$, and

$$K = Q_f H^{\mathrm{T}} \left(HQ_f H^{\mathrm{T}} + R\right)^{-1} \qquad (14)$$

is the Kalman gain matrix, with Q_f being the sample covariance of (u_k^f).

We now generalize EnKF to weighted ensembles. This makes it possible to use EnKF as the predictor in a predictor-corrector filter, operating on weighted ensembles. Fortunately, all that needs to be done is to use as $Q_f = (q_{ij})$ the weighted sample covariance,

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

with $\bar{u}^f$ being the forecast mean. The weights do not change, $w_k^a = w_k^f$.

For an efficient parallel implementation of EnKF, see Mandel (2006).

3.4. *Numerical Results*

To get an idea of why are predictor-corrector filters 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. We will use predictor by weighted EnKF, and the resulting predictor-corrector filter will be called EnKF-SIS.

3.4.1. Bimodal Prior The first such situation we will consider is that of a bimodal prior. We construct a bimodal prior by first sampling from $N\,(0, 5)$. 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 (a). The data likelihood is 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. The density ratio estimate (5), was then applied to the resulting posterior ensemble to obtain an approximate posterior probability distribution. The results for each method are given in Fig. 1.

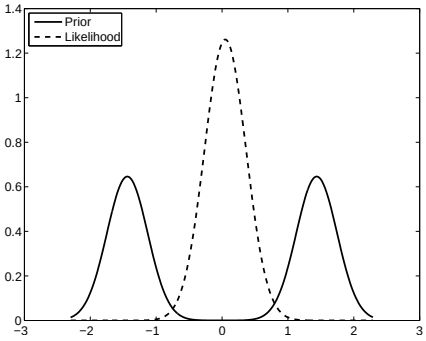
Because the EnKF technique assumes that all distributions are gaussian, it is no surprise that it would fail to capture the non-gaussian features of the posterior. Both SIS and EnKF-SIS were able to represent the nature of the posterior reasonably well.

3.4.2. Filtering in High Dimension Typical results for filtering in the space of functions on $[0, \pi]$ of the form

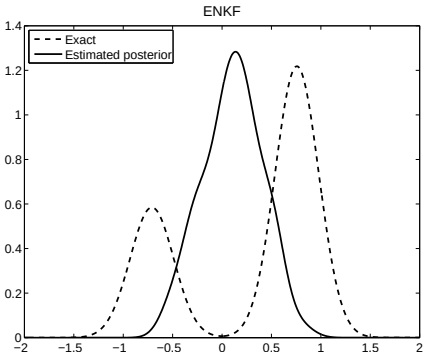
$$u = \sum_{n=1}^{d} c_n \sin\,(nx) \qquad (15)$$

are in Figs. 2 and 3. 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 from (6), and $\kappa_n = n^{-2}$ for density estimation.

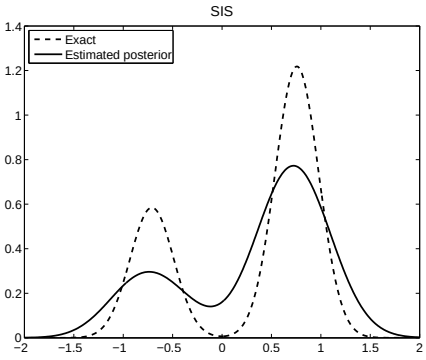
Fig. 2 again shows that EnKF cannot handle bimodal distribution. The prior was constructed by



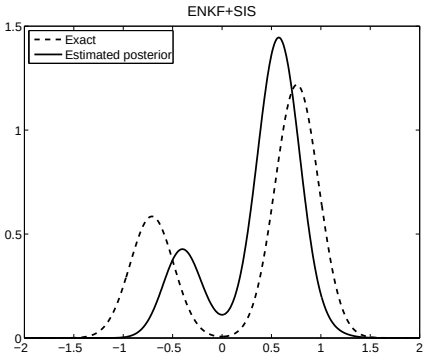
(a) Prior and likelihood distributions



(b) Posterior from EnKF

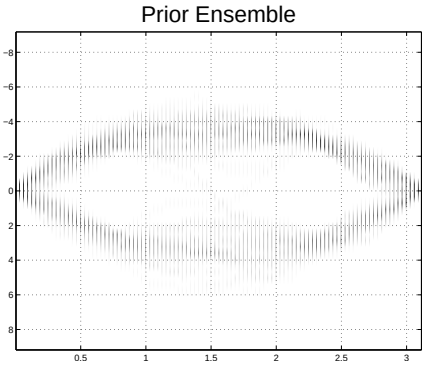


(c) Posterior from SIS

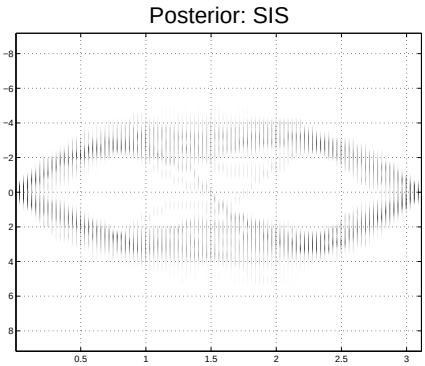


(d) Posterior from EnKF-SIS

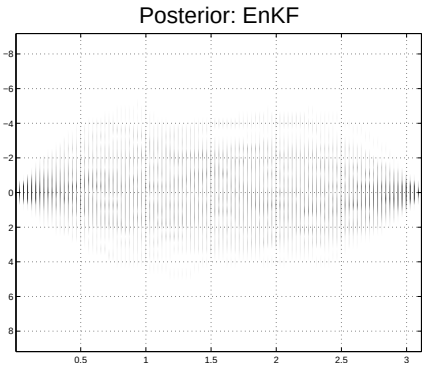
Figure 1: Data assimilation with bimodal prior.



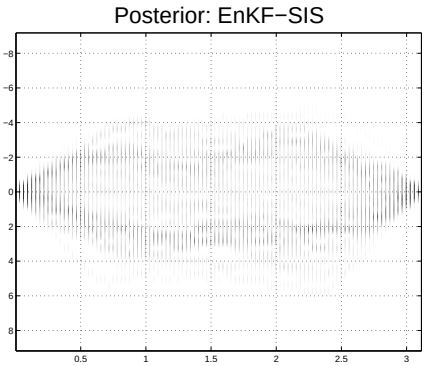
(a)



(b)



(c)



(d)

Figure 2: EnKF does not recognize a non-gaussian prior.

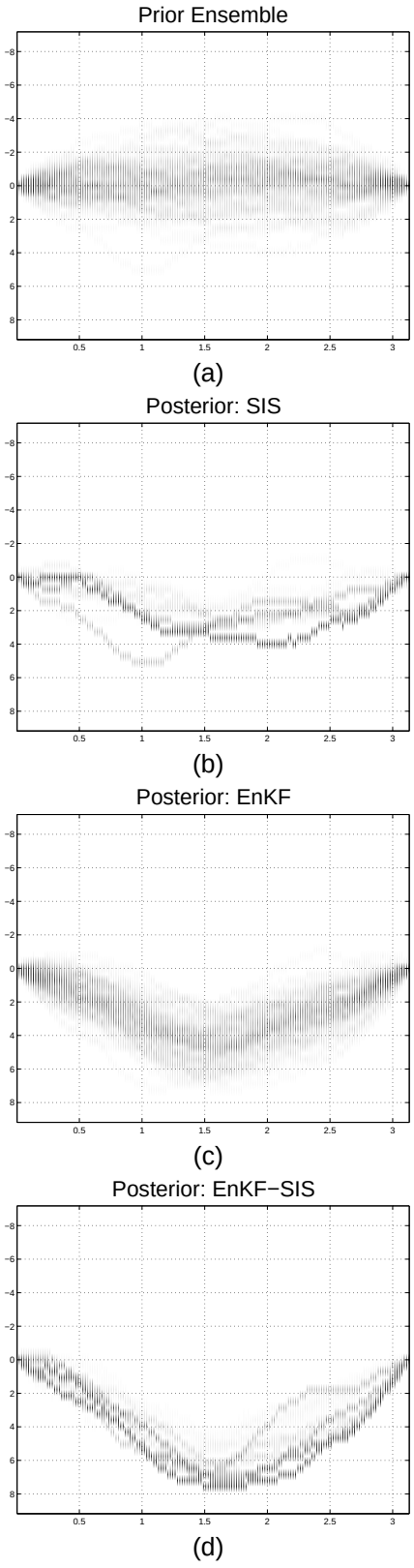


Figure 3: SIS cannot make a large update.

assimilating the data likelihood

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

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

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

was assimilated to obtain the analysis ensemble.

Fig. 3 shows a failure of SIS. The prior ensemble sampled from gaussian distribution with eigenvalues of the covariance $\lambda_n = n^{-3}$ using (6) and (15), and the data likelihood was

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

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.

3.4.3. Filtering for a Stochastic ODE The results given above describe how each filtering technique applies to certain carefully designed synthetic 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 by Kim et al. (2003),

$$\frac{du}{dt} = 4u - 4u^3 + \kappa\eta, \tag{16}$$

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

The deterministic part of this differential equation is of the form $du/dt = -f'\left(u\right)$, where the potential $f(u) = -2u^2 + u^4$, cf., Fig 4. The equilibria are given by $f'\left(u\right) = 0$; there are two stable equilibria at $u = \pm 1$ and an unstable equilibrium at $u = 0$. 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 5) is rare.

The equation (16) is a simple model problem related to fire simulation, cf., (22) below, which adds spatial diffusion; the two equilibria in (16) are related to the burning and not burning states.

A suitable test for an ensemble filter will be determining if it can properly track the model as it

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

Table 1: Data used in assimilation into (16)

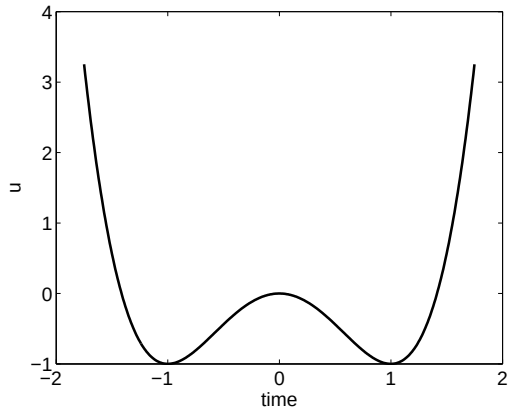


Figure 4: Deterministic potential curve

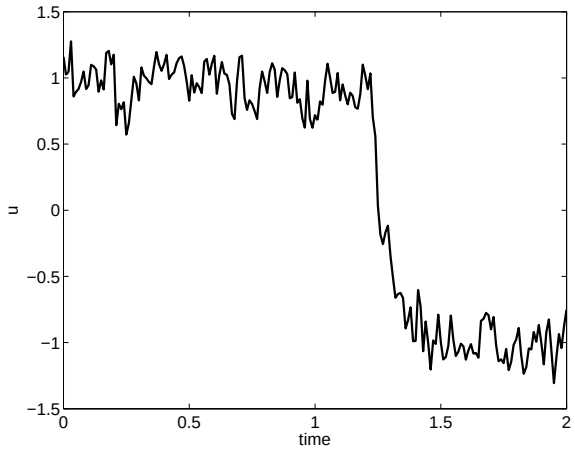


Figure 5: A solution of (16) switching between stable points

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 would be close to the other stable point, so SIS will be even slower in tracking the transition (Kim et al. 2003).

The solution u of (16) is a random variable dependent on time, with density $p(t, u)$. The evolution of the density in time is given by the Fokker-Planck equation (Miller et al. 1999),

$$\frac{\partial p}{\partial t} = \frac{\partial}{\partial u}[4u(u^2 - 1)p] + \frac{\sqrt{\kappa}}{2}\frac{\partial^2 p}{\partial u^2}. \qquad (17)$$

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 u between the bayesian updates by solving the Fokker-Planck equation (17) numerically on a uniform mesh from $u = -3$ to $u = 3$ with the step $\Delta u = 0.01$, using the MATLAB function `pdepe`. At the analysis time, we have multiplied the probability density of u by the data likelihood as in (1) and then scaled so that again $\int p du = 1$, using numerical quadrature by the trapezoidal rule.

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 (16) 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. 6, 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

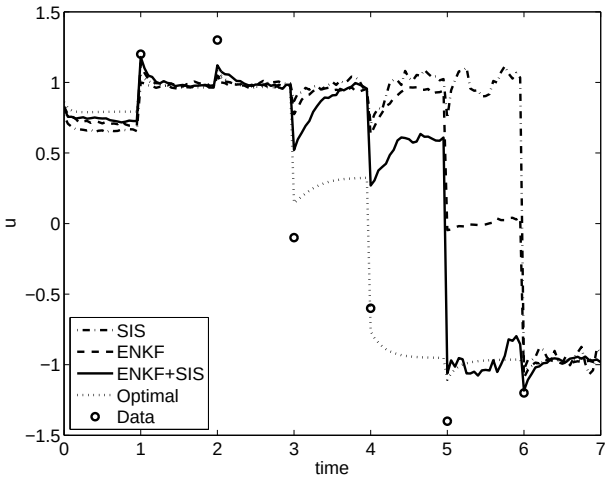


Figure 6: Ensemble filters mean and optimal filter mean for (16)

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.

4. MORPHING ENSEMBLE FILTERS

4.1. *Position Errors, Registration, and Morphing*

Adjusting the simulation state is often more naturally done by moving the position of a distinct feature, such as a fireline or a vortex, rather than employing an additive correction to the state. Also, while the position of the feature may have error distribution that is approximately gaussian, this is not necessarily the case for the value of the state at a given point. For this reason, alternative error models including the position of features were considered in the literature. Hoffman et al. (1995) proposed to decompose the difference between two states into a transformation of the space, a multiplication of the amplitude, and an additive residual error. Such decomposition is not unique, and an optimization approach is needed to make it at least locally unique, e.g., as in (21) below. Davis et al. (2006) evaluated the difference by the location and characteristics of features treated as objects. While an additive approach to data assimilation can be successful if the movement is not too drastic (Chen and Snyder 2006; Mandel et al. 2006), a number of works emerged that achieve more efficient movement of the features by using a spatial transformation as the field to which additive corrections are made. Alexander et al. (1998) proposed a transformation of the space by a global low order polynomial mapping to achieve alignment. Lawson and Hansen (2005) proposed a two-step model to treat alignment and additive errors separately. Ravela et al. (2006) used an alignment preprocessing step to 3DVAR.

The concept of position error is closely related to registration and morphing in image processing (Brown 1992). Suppose two images are represented as functions u and v, respectively, over a spatial domain D. The registration problem is to find a registration mapping T such that v becomes u after a transformation of D by $I+T$, that is,

$$\widetilde{r}(x) = u(x) - v(x + T(x) \approx 0, \quad x \in D,$$

or, with $\circ$ denoting composition of mappings,

$$\widetilde{r} = u - v \circ (I + T) \approx 0.$$

The function u can be represented in terms of v as

$$u = (v + r) \circ (I + T) \tag{18}$$

where

$$r = \widetilde{r} \circ (I + T)^{-1} = u \circ (I + T)^{-1} - v,$$

assuming that $(I+T)^{-1}$ exists. We can avoid the issue whether $I+T$ maps D into D by extending u and v beyond D. The representation (18) can be then used to create images that are intermediate between u and v,

$$u_\lambda = (v + \lambda r) \circ (I + \lambda T), \quad 0 \le \lambda \le 1, \tag{19}$$

cf., Fig. 7. Clearly, $u_0 = u$ and $u_1 = v$.

4.2. *Morphing EnKF Algorithm*

The key observation is that (18) with fixed v allows replacing linear combinations of states by linear combinations of the morphing representations $[T, r]$. The overall scheme can be stated as follows.

1. Choose one of the forecast ensemble members and copy it to v, which will be called a template.
2. Represent all forecast ensemble members by registration with the template as

   $$u_k^f = (v + r_k^f) \circ (I + T_k^f),$$

   using T_k^a from the previous analysis cycle as the initial values.
3. Apply the EnKF formulas (Sec. 3.3) to the transformed forecast ensemble $([T_k^f; r_k^f])_{k=1}^N$, writing the observation function also in terms

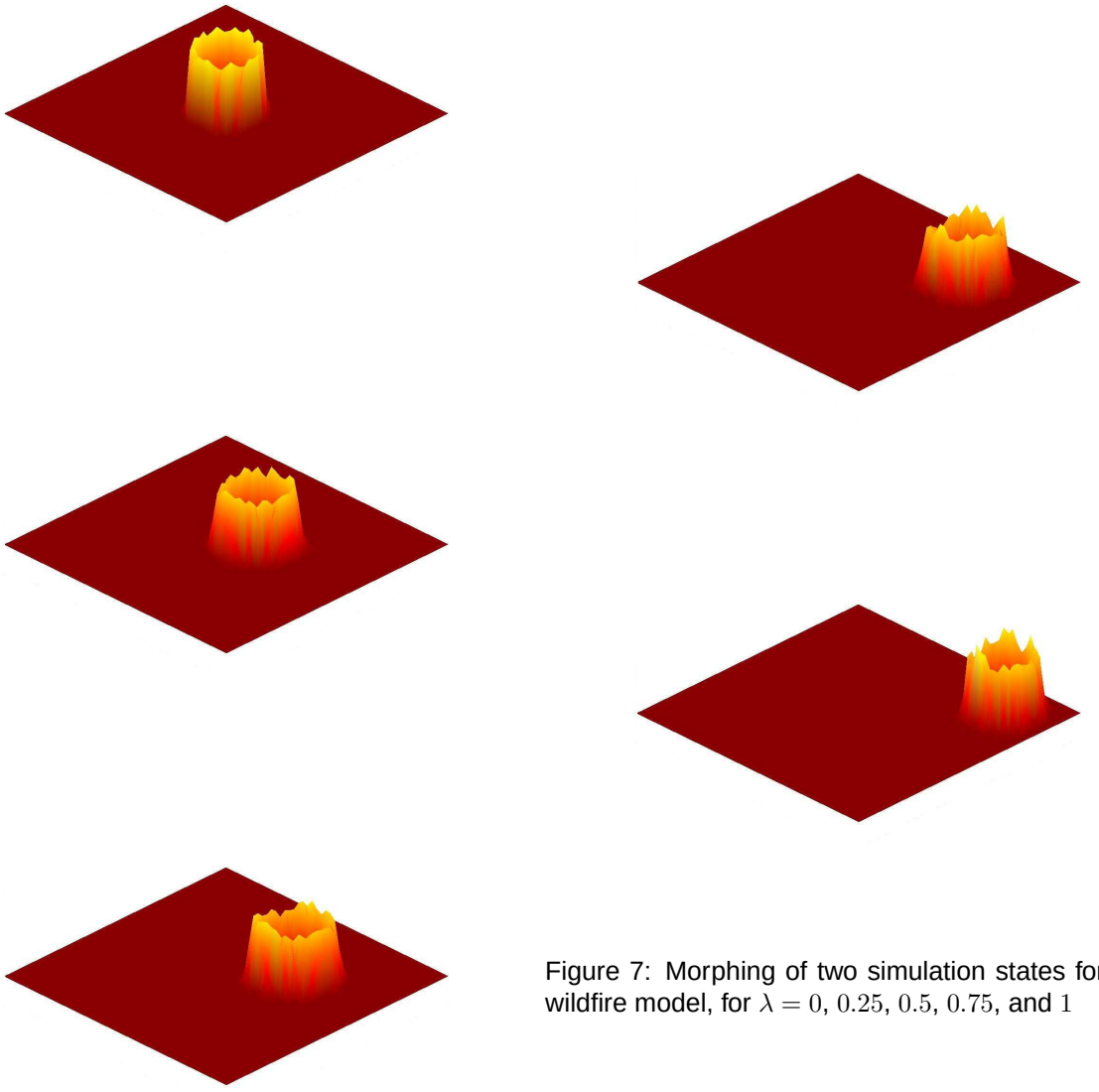


Figure 7: Morphing of two simulation states for the wildfire model, for $\lambda = 0$, 0.25, 0.5, 0.75, and 1

T and r. Replace the the multiplication of observation matrix and vector of deviations by

$$H[T_k^f - \overline{T}_k^f; r_k^f - \overline{r}_k^f] \approx h([T_k^f; r_k^f]) - \frac{1}{N}\sum_{j=1}^{N} h([T_j^f; r_j^f]). \quad (20)$$

Obtain the transformed analysis ensemble $[T_k^a; r_k^a]$.

4. Construct the analysis ensemble by transforming back,

$$u_k^a = (v + r_k^a) \circ (I + T_k^a)$$

Since registration methods are iterative, starting from a good approximation in step 2 above results in a faster method. Therefore, the values of T_k and r_k from the previous forecast are kept and used as initial values.

It should be noted that the ensemble mean is not suitable as a template: instead of average position of the features of interest, the mean will have the superposition of the features at different locations.

The treatment of a nonlinear observation function as in (20) is known in ensemble filtering practice (Chen and Snyder 2006). This approximation is exact when the transformed observation function is linear. However, when the observation function is given in terms of values of u at a set of points (station observations), the transformed observation function will be far from being linear, and the approximation may miss the significant features completely. Therefore, it is preferable to transform the data by registration first if possible, cf., (24) below.

4.3. *The Registration Algorithm*

We have used the method by Gao and Sederberg (1998) with some modifications. We determine T by minimizing the objective function

$$C_r||u - v \circ (I + T)\,||_1 + C_T||T||_1 + C_g||\nabla T||_1. \quad (21)$$

In our application, the domain D is a rectangle. For simplicity of exposition, suppose that $D = [0,1] \times [0,1]$. The functions u and v are piecewise bilinear functions, identified with their values on a rectangular mesh, and T is given by its values on the same mesh. The composition $\circ$ is implemented by bilinear interpolation. The $\|\cdot\|_1$ norm of a vector is the sum of absolute values of its entries divided by the number of the entries. The gradient ∇T is computed by central differencing. The function v is extended outside of D by a constant, equal to the boundary condition.

The method proceeds by from coarser to finer grid. Starting with $i = 1$, the domain D is discretized by a $(2^i + 1) \times (2^i + 1)$ uniform grid. The mapping T is stored as two $(2^i + 1) \times (2^i + 1)$ matrices T_x and T_y, so that for (x, y) on the grid, $T : (x, y) \mapsto (T_x(x, y), T_y(x, y))$. To guarantee that $I + T$ is one-to-one, we will seek mappings such that $x+T_x(x, y_0)$ and $y + T_y(x_0, y)$ are strictly increasing for all x_0, y_0.

The values $(x_j + X, y_k + Y)$ will be called vertex position. The basic optimization approach to sweep through all $x_j = j/2^i$, $y_k = k/2^i$, $j, k = 0, ..., 2^i$ and, for each vertex j, k in turn, update the vertex position $X = T_x(x_j, y_k)$, $Y = T_y(x_j, y_k)$ to the value that minimizes a version of the objective function (21), defined by the current mesh quantities, with the vertex constrained to the local rectangle

$$\begin{aligned} x_{j-1} + T_x(x_{j-1}, y_k) &< x_j + X \\ &< x_{j+1} + T_x(x_{j+1}, y_k), \\ y_{k-1} + T_y(x_j, y_{k-1}) &< y_k + Y \\ &< y_{k+1} + T_y(x_j, y_{k+1}), \end{aligned}$$

using the current values for all other vertices.

Because there may be many local minima within this region, the objective function is first evaluated on a uniform grid within the local rectangle. The size of this grid depends on the mesh level, with more positions tested at the coarsest grid because the objective function will be more nonlinear over larger distances. Once the objective function is minimized on the local grid, the vertex position is further refined using a standard library nonlinear optimization method (`fminbnd` in MATLAB). We make up to 3 sweeps, depending on stopping conditions such as if the relative improvement in the objective function over the last iteration is small.

The values of T_x are T_y are then interpolated to a mesh with half the stepsize and the method repeats with i incremented by 1, until the desired mesh size is reached.

Because we are only interested in changes in the objective function when finding the values (X, Y) above, only the terms in the objective function that change are evaluated for efficiency. Also, because we really interested minimizing the residual, if the maximum value of the residual is below a certain threshold within this region we can skip the optimization on the current local rectangle.

In order to improve registration in the case when the features of interest are smaller than the current grid, we apply gaussian smoothing before the

optimization sweeps. This amounts to applying a stencil to every element of the matrices u and v. This stencil is a discretized two dimensional gaussian distribution:

$$\exp(-x^2/\alpha - y^2/\alpha).$$

The parameters α determines the amount of smoothing along each coordinate axis and it is tuned so that there is more smoothing in the coarser grids. This allows for the mapping to track large perturbations in the domain on coarser grids even for a thin feature such as a fireline, while maintaining small scale accuracy on finer grids.

The differences between the method described here and the method by Gao and Sederberg (1998) are: the refinement of each vertex position by nonlinear optimization; the image smoothing on coarse levels by convolution; the gradient term in the objective function; and allowing the vertices to leave the domain.

4.4. *Numerical Results*

A good application for the morphing ensemble filter is when there is significant uncertainty in the physical location in the data or in the system state, such as caused by a firefront propagating with a different speed than anticipated. Standard ensemble filters do no represent this uncertainty accurately, particularly when dealing with models that exhibit large gradients, such as in fires or hurricanes. We present here the results of applying the morphing ensemble filter to a simple wildland fire model in the form of the reaction-convection-diffusion equation,

$$\frac{du}{dt} = \nabla \cdot (k\nabla u) + \overrightarrow{v} \cdot \nabla u - C\,(u - u_a) + Af(u, S), \tag{22}$$

$$\frac{dS}{dt} = -f(u, S), \quad u > u_a, \tag{23}$$

where u and S are the state variables representing temperature and fuel supply respectively. The function, f, represents the rate of the reaction. The ambient wind and temperature are given by $\overrightarrow{v}$ and u_a respectively. The constants are determined by matching the solution to real fire data. See (Mandel et al. 2006) for details.

To prepare the forecast ensemble, we have generated temperature profiles by raising a $25\ m \times 25\ m$ square above the ignition temperature and advancing the solution by $500\ s$. Two such profiles were generated, one with the ignition in the center of the domain, as shown in Fig. 8, called the comparison solution, and one with the ignition in the lower right corner, called the reference solution (Fig. 9).

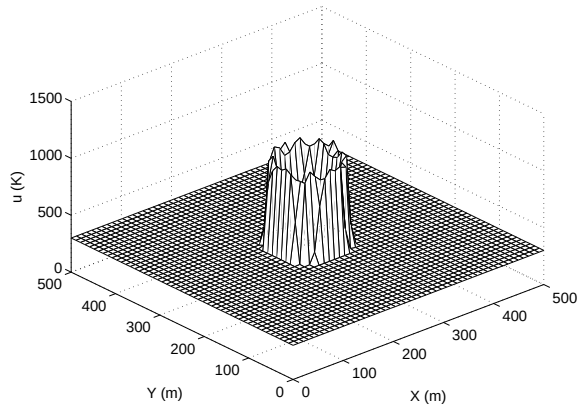


Figure 8: The temperature profile used for the comparison solution (to generate the ensemble).

We have used smooth random fields

$$\widetilde{u}(x, y) = \sum_{n=1}^{d} \sum_{m=1}^{d} \lambda_{m,n} d_{m,n} \sin m\pi x \sin n\pi y$$

with $d_{m,n} \sim N(0, 1)$ and

$$\lambda_{m,n} = \left(1 + \frac{\sqrt{m^2 + n^2}}{d - 1}\right)^{-2}.$$

A forecast ensemble with 100 members was generated first by adding a smooth random field to each state variable of the comparison solution; the temperature of k^{th} ensemble member is given by $\widehat{u}_k = u_0 + c_T \tilde{u}_k$, where the scalar c_T controls the magnitude of this perturbation. Then the ensemble was moved spatially in both x and y directions by,

$$T_k(x, y) = \widehat{u}_k(x + c_x \tilde{u}_{i_k}(x, y), y + c_y \tilde{u}_{j_k}(x, y)).$$

Here, c_x and c_y control the magnitude of the shift in each coordinate, bilinear interpolation is used to determine $\hat{T}$ on off-grid points, and temperature outside of the computational domain is assumed to be at the ambient temperature. The given simulation was run with the initialization parameters $c_T = 5$ and $c_x = c_y = 150$.

The location of the perturbed firelines of all 100 ensemble members are shown in Fig. 10. The ensemble mean location of the fireline is located approximately $300\ m$ away from that of the data and are spread out over most of the domain. This represents a high degree of prior

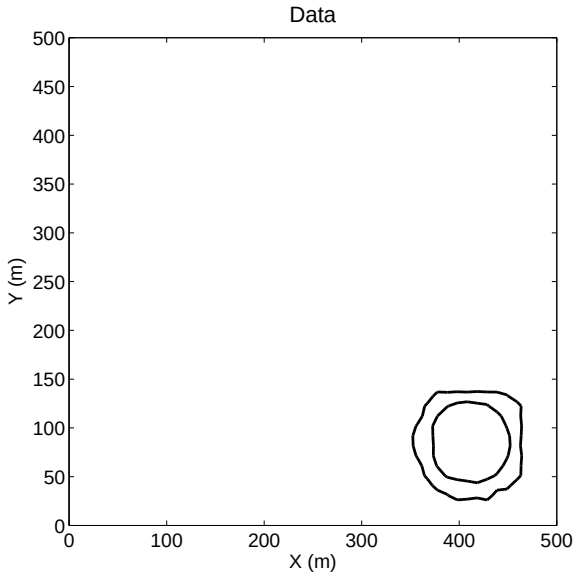


Figure 9: Contours at $800K$, indicating the location of the fireline in the reference solution (the simulated data). The reaction zone is between the two curves.

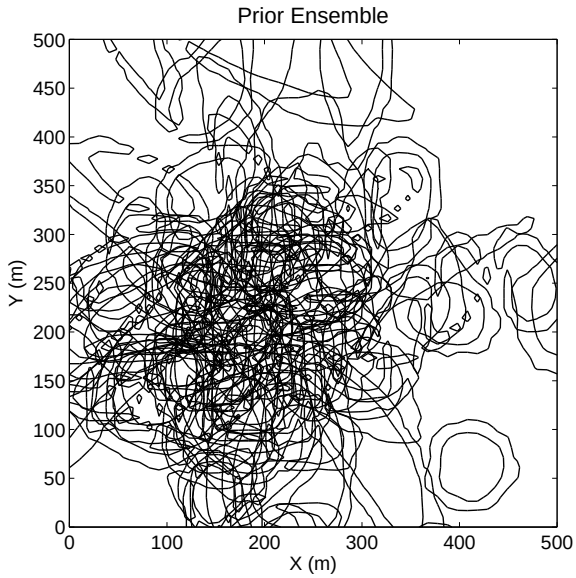


Figure 10: Contours at $800K$, indicating the location of the firelines of the ensemble before applying the data assimilation.

uncertainty in the location of the fire. The data assimilation was performed with residual, mapping, and gradient variances of $1\ K$, $.01\ m$, and $10\ K/m$ respectively. The data was taken to be the reference solution itself, and it was transformed to the registration representation $[T_d; r_d]$. Regularization by penalization of the gradient of T was added to the observation function as well; thus, the observation equation was

$$h\left(\begin{bmatrix} T \\ r \end{bmatrix}\right) = \begin{bmatrix} T \\ r \\ \nabla T \end{bmatrix} = \begin{bmatrix} T_d \\ r_d \\ 0 \end{bmatrix} \qquad (24)$$

with data variances of $.05$ for T, 1 for r, and 10 for ∇T. The constants in the objective function of the registration algorithm were $C_T = .75$, $C_r = .07$, and $C_g = .01$. The posterior ensemble, as shown in Fig. 11, has far less variance in the mappings and has shifted closer to the data.

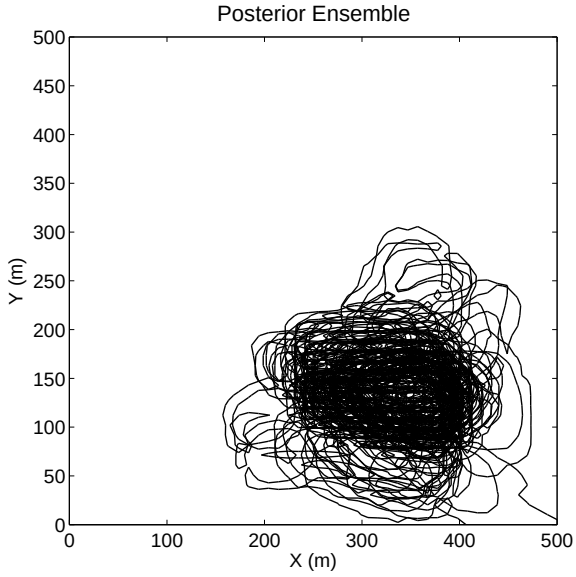


Figure 11: Contours at $800K$, indicating the location of the firelines of the ensemble after applying the data assimilation.

5. CONCLUSION AND FUTURE

We have demonstrated the potential of predictor-corrector and morphing ensemble filters to perform a successful bayesian update in the presence of non-gaussian distributions, large number of degrees of freedom, large change of the state distribution, and position errors.

For the predictor-corrector filters, open questions include convergence of the filters when applied to

multiple updates over time, mathematical convergence proofs for the density estimation and for the bayesian update, and performance of the filters when applied to systems with a large number of different physical variables, as is common in atmospheric models. 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. This appears due to the fact that the "effective dimension" of a space of smooth functions is quite small, regardless of the number of grid points.

The open questions for morphing filters include handling of sparse station observations, resulting in strongly nonlinear observation function. One possible path is to go back to the formulation of the EnKF as least squares and using nonlinear minimization. The registration algorithm can be also formulated as a version of multilevel optimization (Mandel 1984; Gelman and Mandel 1990).

Combination of predictor-corrector and morphing filters would be also of interest.

These issues will be pursued elsewhere.

Acknowledgement The authors are grateful to Doug Nychka, Craig Johns, Tolya Puhalskii, and Yongsheng Chen for useful discussions. This research was supported by NSF grant CNS-0325314 and by the NCAR Faculty Fellowship Program.

References

Alexander, G. D., J. A. Weinman, and J. L. Schols, 1998: The use of digital warping of microwave integrated water vapor imagery to improve forecasts of marine extratropical cyclones. *Mon. Wea. Rev.*, **126**, 1469–1496.

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

Bogachev, V. I., 1998: *Gaussian measures*. Mathematical Surveys and Monographs, Vol. 62, American Mathematical Society, Providence, RI.

Brown, L. G., 1992: A survey of image registration techniques. *ACM Comput. Surv.*, **24**, 325–376.

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

Chen, Y. and C. Snyder, 2006: Assimilating vortex position with an ensemble Kalman filter. Monthly Weather Review, to appear.

Dabo-Niang, S., 2004: Kernel density estimator in an infinite-dimensional space with a rate of convergence in the case of diffusion process. *Appl. Math. Lett.*, **17**, 381–386.

Davis, C., B. Brown, and R. Bullock, 2006: Object-based verification of precipitation forecasts. part i: Methodology and application to mesoscale rain areas. *Mon. Weather Rev.*, **134**, 1772–1784.

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

Douglas, C. C., J. D. Beezley, J. Coen, D. Li, W. Li, A. K. Mandel, J. Mandel, G. Qin, and A. Vodacek, 2006: Demonstrating the validity of a wildfire DDDAS. *Computational Science ICCS 2006: 6th International Conference, Reading, UK, May 28-31, 2006, Proceedings, Part III*, V. N. Alexandrov, D. G. van Albada, P. M. A. Sloot, and J. Dongarra, eds., Springer, Berlin/Heidelberg, volume 3993 of *Lecture Notes in Computer Science*, 522–529.

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.

Gao, P. and T. W. Sederberg, 1998: A work minimization approach to image morphing. *The Visual Computer*, **14**, 390–400.

Gelman, E. and J. Mandel, 1990: On multilevel iterative methods for optimization problems. *Math. Programming*, **48**, 1–17.

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.

Hoffman, R. N., Z. Liu, J.-F. Louis, and C. Grassoti, 1995: Distortion representation of forecast errors. *Monthly Weather Review*, **123**, 2758–2770.

Johns, C. J. and J. Mandel, 2005: A two-stage ensemble Kalman filter for smooth data assimilation. CCM Report 221,

www.math.cudenver.edu/ccm/reports, Environmental and Ecological Statistics, in print. 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.

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*. Lecture Notes in Mathematics, Vol. 463, Springer-Verlag, Berlin.

Lawson, W. G. and J. A. Hansen, 2005: Alignment error models and ensemble-based data assimilation. *Monthly Weather Review*, **133**, 1687–1709.

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

Mandel, J., 1984: A multilevel iterative method for symmetric, positive definite linear complementarity problems. *Appl. Math. Optim.*, **11**, 77–95.

— 2006: Efficient implementation of the ensemble Kalman filter. CCM Report 231, University of Colorado at Denver and Health Sciences Center, 2006; http://www.math.cudenver.edu/ccm/reports.

Mandel, J., L. S. Bennethum, J. D. Beezley, J. L. Coen, C. C. Douglas, L. P. Franca, M. Kim, and A. Vodacek, 2006: A wildfire model with data assimilation. CCM Report 233, http://www.math.cudenver.edu/ccm/reports, submitted to Mathematics and Computers in Simulation.

Mandel, J., L. S. Bennethum, M. Chen, J. L. Coen, C. C. Douglas, L. P. Franca, C. J. Johns, M. Kim, A. V. Knyazev, R. Kremens, V. Kulkarni, G. Qin, A. Vodacek, J. Wu, W. Zhao, and A. Zornes, 2005: Towards a dynamic data driven application system for wildfire simulation. *Computational Science - ICCS 2005*, V. S. Sunderam, G. D. van Albada, P. M. A. Sloot, and J. J. Dongarra, eds., Springer, volume 3515 of *Lecture Notes in Computer Science*, 632–639.

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

Ravela, S., K. A. Emanuel, and D. McLaughlin, 2006: Data assimilation by field alignment. Physica D, to appear.

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

van Leeuwen, P., 2003: A variance-minimizing filter for large-scale applications. *Monthly Weather Review*, **131**, 2071–2084.

Xiong, X. and I. Navon, 2005: Ensemble particle filter with posterior Gaussian resampling. Tellus, submitted.

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