A TWO-STAGE ENSEMBLE KALMAN FILTER FOR SMOOTH DATA ASSIMILATION

CRAIG J. JOHNS AND JAN MANDEL *

Abstract. The ensemble Kalman Filter (EnKF) applied to a simple fire propagation model by a nonlinear convection-diffusion-reaction partial differential equation breaks down because the EnKF creates nonphysical ensemble member with large gradients. A modification of the EnKF is proposed by adding a regularization term that penalizes large gradients. The method is implemented by applying the EnKF formulas twice, with the regularization term as another observation. The regularization step is also interpreted as a shrinkage of the prior distribution. Numerical results are given to illustrate success of the new method.

KEYWORDS: Data Assimilation, Ensemble Kalman Filter, State-Space Model, Penalty, Tikhonov Regularization, Wildfire, Convection-Reaction-Diffusion, Shrinkage, Bayesian

March 2005, revised September 2005

1. Introduction. The discrete time state-space model in its most general form is an application of the Bayesian update problem: the modeled system is advanced in time until an *analysis* time, when the distribution of the system state before the update, called the *prior* or the *forecast* distribution, and the *data likelihood* are combined to give the new system state distribution, called the *posterior* or the *analysis* distribution. The system is then advanced until the next analysis time. Kalman [20] and Kalman and Bucy [21] provided simple recursive formulas for the system mean and covariance under the assumptions that the probability distributions are normal and the system is linear. The Kalman filter is popular in areas as diverse as medicine [19],

*Department of Mathematics, University of Colorado at Denver and Health Sciences Center, Denver, CO, 80217-3364

economics [29] and geosciences [11]. Variants were devised for, e.g., nonlinear problems [16], missing observations [29], censored observations [18], irregular observation times [19], nonlinear updates [14], and non-Gaussian distributions [7, 23, 27].

Traditional Kalman filters implicitly manipulate the covariance matrix of the state, and they are thus unsuitable for systems with a large number of degrees of freedom, such as in computational models in geophysics. Ensemble Kalman Filters (EnKFs) were developed [10, 15] that represent the distribution of the system state using a random sample, called an ensemble, and do not use the covariance matrix explicitly. The benefit of the EnKF comes in situations where the eigenvalues of the covariance matrix rapidly decay. In that case, even a few ensemble members can reproduce the large-scale behavior of the covariance behavior of the system. This situation is typical of models of governed by partial differential equations, such as in geophysical systems. For related filters relaxing or removing the Gaussian assumption, see [3, 4, 9, 31]. For comprehensive surveys, see [11, 12, 30].

The method proposed in this paper was motivated by the observation that straightforward application of EnKF to a simple wildfire model [24] always fails within a few analysis cycles. Due to statistical variability of the ensemble members, locations with large temperature gradients develop, resulting in bigger fires and some ensemble members move away from the truth. The EnKF update formulas, trying hopelessly to match the observations within the span of the ensemble, result in states that are nonphysical (too large, too small, or too rough) in some places, which causes a complete breakdown of the simulations in subsequent advancements.

We use a Bayesian approach similar to Tikhonov regularization [13] to force the analysis solution to be more spatially smooth. Tikhonov regularization techniques were used in Kalman filters in a different way than described here, to stabilize ill-

conditioned parameter identification [17, 22]. The beauty and utility of the method comes from the implementation; regularization requires simply using the EnKF update formulas twice, thus avoiding the need for a new code.

The paper is organized as follows. In Sec. 2, we set the stage by briefly reviewing the Kalman filter. The Ensemble Kalman filter an its implementation using contemporary numerical software are considered in Sec. 3. In Sec. 4, we add regularization as an independent observation, leading to a two stage EnKF. Finally, results for a simple fire model problem are presented in Sec. 5.

2. The Kalman Filter. We consider the state space model, in which the modeled quantity is the probability distribution of the state vector $\mathbf{x}$. The probability distribution is evolved in time by running the model until the end of an analysis cycle, when it is updated to account for new data. At the end of the cycle, the probability density $p(\mathbf{x})$ of the system state $\mathbf{x}$ before the update (the prior) and the probability density $p(\mathbf{y}|\mathbf{x})$ of the the data $\mathbf{y}$ given an assumed value of the system state $\mathbf{x}$ (the data likelihood) are combined to give the new probability density of the system state $p(\mathbf{x}|\mathbf{y})$ (the posterior) by the Bayes theorem,

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})\, p(\mathbf{x}), \tag{2.1}$$

where $\propto$ means proportionality. Eq. (2.1) determines the posterior density $p(\mathbf{x}|\mathbf{y})$ completely because $\int p(\mathbf{x}|\mathbf{y})\, d\omega(\mathbf{x}) = 1$. Consider the case of linear observation operator H: given system state $\mathbf{x}$, the data value, $\mathbf{y}$, would be $H\mathbf{x}$ if the model and the data were perfect with no errors. Of course, in general, the given data $\mathbf{y} \neq H\mathbf{x}$, so discrepancies are modelled with the likelihood $p(\mathbf{y}|\mathbf{x})$. Assume that the prior has normal distribution with mean μ and covariance Q, and the data likelihood is normal

with mean $H\mathbf{x}$ and covariance R,

$$p(\mathbf{x}) \propto \exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^{\mathrm{T}}Q^{-1}(\mathbf{x}-\mu)\right),$$
$$p\left(\mathbf{y}|\mathbf{x}\right) \propto \exp\left(-\frac{1}{2}(\mathbf{y}-H\mathbf{x})^{\mathrm{T}}R^{-1}(\mathbf{y}-H\mathbf{x})\right).$$

Denote the posterior system state by $\hat{\mathbf{x}}$ instead of $\mathbf{x}|\mathbf{y}$. It can be shown by algebraic manipulations [1] that the posterior is also normal,

$$p\left(\hat{\mathbf{x}}\right) \propto \exp\left(-\frac{1}{2}(\hat{\mathbf{x}}-\hat{\mu})^{\mathrm{T}}P^{-1}(\hat{\mathbf{x}}-\hat{\mu})\right),$$

where the posterior mean $\hat{\mu}$ and covariance P are given by the update formulas

$$\hat{\mu} = \mu + K\left(\mathbf{y} - H\mu\right), \quad P = \left(I - KH\right)Q, \tag{2.2}$$

$$K = QH^{\mathrm{T}}\left(HQH^{\mathrm{T}} + R\right)^{-1}. \tag{2.3}$$

The matrix K is called the Kalman gain matrix. The observation [26, 28] in the following Lemma interprets the Kalman filter as least squares: the posterior mean $\hat{\mu}$ is obtained by trying to match the observation, $H\hat{\mu} \approx \mathbf{y}$, as well as to preserve the mean, $\hat{\mu} \approx \mu$.

LEMMA 2.1. *If $\hat{\mu}$ is defined by (2.2) and (2.3), then $\hat{\mu}$ is the solution $\mathbf{x}$ of the least-squares problem*

$$S\left(\mathbf{x}\right) = (\mathbf{x}-\mu)^{\mathrm{T}}Q^{-1}(\mathbf{x}-\mu) + (\mathbf{y}-H\mathbf{x})^{\mathrm{T}}R^{-1}(\mathbf{y}-H\mathbf{x}) \rightarrow \min_{\mathbf{x}}. \tag{2.4}$$

Proof. At the minimum,

$$\nabla S(\mathbf{x}) = 2Q^{-1}(\mathbf{x}-\mu) - 2H^{\mathrm{T}}R^{-1}(\mathbf{y}-H\mathbf{x}) = 0,$$

which gives $\mathbf{x} = \tilde{P}(Q^{-1}\mu + H^{\mathrm{T}}R^{-1}\mathbf{y})$, where

$$\tilde{P} = (Q^{-1} + H^{\mathrm{T}}R^{-1}H)^{-1} = \left[Q - QH^{\mathrm{T}}(HQH^{\mathrm{T}} + R)^{-1}HQ\right] = (I-KH)Q = P.$$

Consequently, $\mathbf{x} = P(Q^{-1}\mu + H^{\mathrm{T}}R^{-1}\mathbf{y}) = \mu + K(\mathbf{y} - H\mathbf{x})$. ∎

The next Lemma is an elementary consequence of the Bayes theorem (2.1).

LEMMA 2.2. *Let* $\mathbf{y}$ *and* $\mathbf{z}$ *be observations such that, conditional on* $\mathbf{x}$*, the error distributions are independent. Then assimilating the observations* $\mathbf{y}, \mathbf{z}$ *jointly gives the same result as first assimilating the observation* $\mathbf{y}$ *to obtain the posterior* $\propto p(\mathbf{y}|\mathbf{x})\, p(\mathbf{x})$*, then taking this posterior to be the new prior and assimilating* $\mathbf{z}$*.*

Proof. From the Bayes theorem,

$$p(\mathbf{x}|\mathbf{y},\mathbf{z}) \propto p(\mathbf{y},\mathbf{z}|\mathbf{x})\, p(\mathbf{x}) = p(\mathbf{z}|\mathbf{x})\, p(\mathbf{y}|\mathbf{x})\, p(\mathbf{x}) = p(\mathbf{z}|\mathbf{x})\, [p(\mathbf{y}|\mathbf{x})\, p(\mathbf{x})], \qquad (2.5)$$

since $\mathbf{y}$ and $\mathbf{z}$ are conditionally independent random variables. □

3. Ensemble Kalman Filter. The EnKF is a Monte Carlo implementation of the Kalman filter, which avoids evolving the covariance matrix of the distribution of the state vector $\mathbf{x}$. Instead, the distribution is represented by a sample, called an ensemble. Ensemble members are evolved in time until the end of the analysis cycle, when the ensemble is updated from the Bayes theorem.

Let $\mathbf{X}$ be a p by n matrix whose columns are ensemble members before the Bayesian update, that is, a random sample from the prior distribution for the state vector $\mathbf{x}$. The prior distribution is assumed to be normal with covariance Q. Further, replicate the observations $\mathbf{y}$ into matrix $\mathbf{Y}$ with d rows and n columns so that each column $\mathbf{y}_k$ consists of the observed vector $\mathbf{y}$ plus a random vector from $N(0, R)$. Then it follows from (2.2) and (2.3) that the columns of

$$\hat{\mathbf{X}} = \mathbf{X} + K(\mathbf{Y} - H\mathbf{X})$$

form a random sample from the posterior distribution.

The ensemble filter involves two approximations. First, in the Kalman gain matrix $K = QH^{\mathrm{T}} \left(HQH^{\mathrm{T}} + R\right)^{-1}$, the state covariance Q is unknown, so it is replaced by

the sample covariance computed from the ensemble members,

$$\hat{Q} = \frac{\mathbf{E}\mathbf{E}^{\mathrm{T}}}{n-1}, \quad \mathbf{E} = \mathbf{X} - \mathbf{X}\frac{\mathbf{e}_n\mathbf{e}_d^{\mathrm{T}}}{n},$$

where $\mathbf{e}_k$ is column vector of all ones of size k. This gives the matrix form of the analysis ensemble

$$\hat{\mathbf{X}} \approx \hat{\mathbf{X}}_{ens} = \mathbf{X} + \hat{Q}H^{\mathrm{T}}\left(H\hat{Q}H^{\mathrm{T}} + R\right)^{-1}(\mathbf{Y} - H\mathbf{X}). \tag{3.1}$$

See [6] for more details.

The second approximation results from the fact that if the state evolution is nonlinear, then the prior distribution is not necessarily normal. Nevertheless, in practice, it is usually hoped that the distribution from which analysis ensemble members in $\hat{\mathbf{X}}_{ens}$ are drawn from is a good approximation of the posterior and the ensemble filter formulas are used anyway. In general, the approximations are practical in the setting where a model is approximately linear in each model advancement step.

REMARK 3.1. *For EnKF, observations with independent error distributions can be again assimilated sequentially according to Lemma 2.2, because the analysis ensemble is a sample from a distribution that approximates the posterior in the limit for a large ensemble. However, the approximation is different when the observations are assimilated sequentially and when they are assimilated jointly because of the sample covariance approximation. In particular, the resulting analysis ensembles are in general different. For the example presented, we found that differences were quantitative rather than qualitative.*

We now consider an efficient implementation of the EnKF. We write (3.1) as

$$\hat{\mathbf{X}}_{ens} = \mathbf{X} + \frac{\mathbf{E}(H\mathbf{E})^{\mathrm{T}}}{n-1}\underbrace{\left(\frac{(H\mathbf{E})(H\mathbf{E})^{\mathrm{T}}}{n-1} + R\right)}_{B}^{-1}(\mathbf{Y} - H\mathbf{X}),$$

where the matrix

$$H\mathbf{E} = \mathbf{HX} - \mathbf{HX}\frac{\mathbf{e}_n\mathbf{e}_d^{\mathrm{T}}}{n}$$

can be computed using the already known product **HX**, and note that efficient Choleski decomposition of the symmetric matrix B is possible because R is positive definite. The dominant operations are full matrix-matrix operations efficiently implemented in the Level 3 BLAS [8] and LAPACK [2] routines, and also readily parallelizable by SCALAPACK [5]. In many cases, the computational cost can be further reduced as H is usually sparse and/or highly structured, since a component of $\mathbf{y}$ represents a characteristic of the state vector at a single point or the integrated value over some small region. In addition, if the observation errors are independent, the computation may be split into assimilating a part of the observations at a time as noted above.

4. Two-Stage Kalman Filter. In order to stabilize the EnKF, we now combine EnKF with a technique related to Tikhonov regularization. Tikhonov regularization for the algebraic least squares problem $\|Ax - b\|^2 \to \min$ consists of solving instead

$$\|Ax - b\| + \lambda^2 \|Lx\|^2 \to \min_x . \tag{4.1}$$

The added term $\lambda^2 \|Lx\|^2$ incorporates a priori assumptions about the size and smoothness of the desired solution x, in the form of the quadratic $\|Lx\|^2$ [13, p. 100]. The parameter λ should be chosen so that both $\|Ax - b\|^2$ and $\|Lx\|^2$ are close to their minimal values as functions of λ; see [13, p. 84] for details. In practice, λ is often determined by trial and error and can be considered a smoothing parameter.

From (3.1) and Lemma 2.1, it follows that each column $\hat{\mathbf{x}}_k$ of the posterior ensemble $\hat{\mathbf{X}}_{ens}$ is the solution of the least squares problem

$$(\hat{\mathbf{x}}_k - \mathbf{x}_k)^{\mathrm{T}}\hat{Q}^{-1}(\hat{\mathbf{x}}_k - \mathbf{x}_k) + (\mathbf{y}_k - H\hat{\mathbf{x}}_k)^{\mathrm{T}}R^{-1}(\mathbf{y}_k - H\hat{\mathbf{x}}_k) \to \min_{\hat{\mathbf{x}}_k} . \tag{4.2}$$

Note that in deriving (4.2), Lemma 2.1 is applied purely algebraically, with $\mathbf{x}_k$, $\hat{\mathbf{x}}_k$, and $\mathbf{y}_k$ playing the roles of μ, $\hat{\mu}$, and $\mathbf{y}$, respectively.

We want to add to EnKF the assumption that $L\mathbf{x}$ does not vary much from $L\mu$, which, analogously to (4.1), leads to a modification of (4.2), where the columns of the posterior ensemble are found as the solutions $\hat{\mathbf{x}}_k$ of

$$\begin{aligned}(\hat{\mathbf{x}}_k - \mathbf{x}_k)^{\mathrm{T}} Q^{-1} (\hat{\mathbf{x}}_k - \mathbf{x}_k) + (\mathbf{y}_k - H\hat{\mathbf{x}}_k)^{\mathrm{T}} R^{-1} (\mathbf{y}_k - H\hat{\mathbf{x}}_k) \\ + (\mathbf{r}_k - L\hat{\mathbf{x}}_k)^{\mathrm{T}} D^{-1} (\mathbf{r}_k - L\hat{\mathbf{x}}_k) \to \min_{\hat{\mathbf{x}}_k}.\end{aligned} \tag{4.3}$$

Here, D is a given symmetric positive definite covariance matrix, D^{-1} plays the generalized role of the parameter λ, and $\mathbf{r}_k$ is sampled from $N(L\mu, D)$. According to Lemma 2.1, (4.3) is equivalent to assimilating two independent observations, $\mathbf{y} = H\mathbf{x}$ with the error covariance R, and $L\mathbf{x} = L\mu$ with error covariance D. From Lemma 2.2 and Remark 3.1, the Bayesian update corresponding to (4.3) can be implemented as a two stage EnKF simply by applying the EnKF formulas (3.1) twice for the two observations $H\mathbf{x} = \mathbf{y}$ and $L\mathbf{x} = L\mu$.

REMARK 4.1. *Assimilating the observation $L\mathbf{x} = L\mu$ first, it is easy to see that the proposed two-stage Kalman filter is equivalent to a hierarchical Bayesian update using the observation $L\mathbf{x} = L\mu$ to modify the prior. This first update step has the effect of "shrinkage," reducing the spread of $L\mathbf{x}$ around $L\mu$. The observation $\mathbf{y} = H\mathbf{x}$ is then applied to the modified prior in the usual manner.*

5. The Fire Model. A simple example useful for demonstrating the utility of the regularized algorithm proposed above is the simplified model of wildfire by a reaction-convection-diffusion equation [25],

$$\begin{aligned}\frac{\partial T}{\partial t} &= -\nabla \cdot (k \nabla T) - c_1 \cdot \nabla T - c_2 (T - T_a) + c_3 \frac{\partial S}{\partial t} \\ \frac{\partial S}{\partial t} &= -c_4 \max\{0, T - T_i\}^{\alpha} S\end{aligned}$$

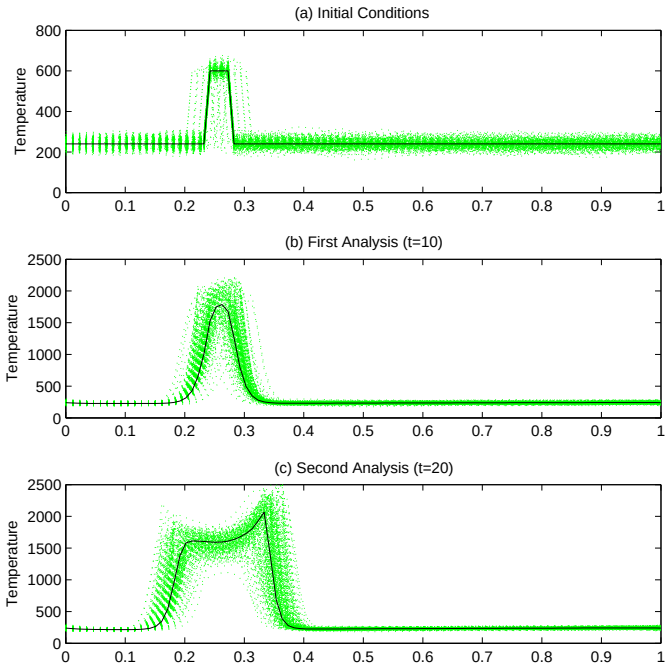

FIG. 5.1. *Panel (a) shows the temperature profiles of the reference solution (solid, dark line) as well as initial distribution of ensemble members (dotted, light lines) for the fire model. Panels (b) and (c) are similar and show the effect of forward propagation on the reference solution (solid dark line) as well as the ensemble members (dotted, light lines) to the first (t=10) and second (t=20) analysis times, respectively.*

on the spatial domain $[0, 1]$, with Dirichlet boundary conditions $T(0) = T(1) = T_a$. The first equation is the heat balance, where T is the temperature, $-\nabla \cdot (k\nabla T)$ is the diffusion of heat, $-c_1 \cdot \nabla T$ is the heat transport by wind, $-c_2(T - T_a)$ is the heat escaping to the environment with the ambient temperature T_a, and $c_3 \frac{\partial S}{\partial t}$ is the heat generated by burning. The second equation models the fuel supply; its right-hand side is the intensity of burning. This is a very simplified model and we do not use any physical data, yet it appears to capture some essential qualitative fire behavior. All coefficients $c_1, c_2, c_2, c_4, \alpha$ are positive. The

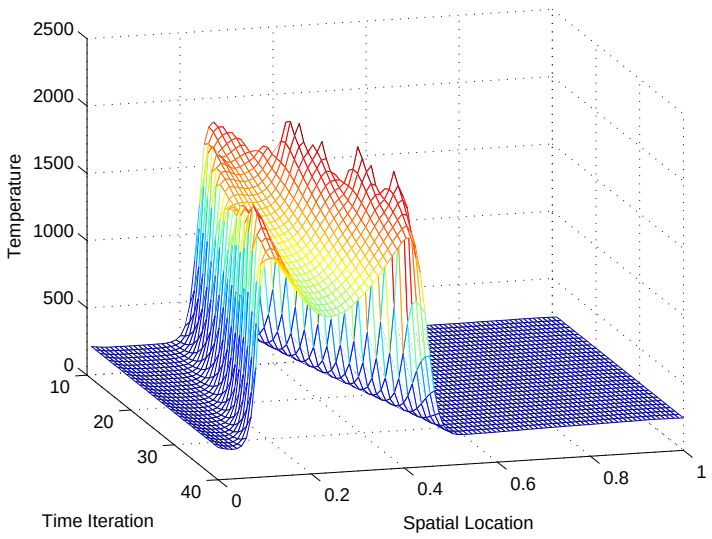

FIG. 5.2. *The reference solution of temperature for the fire model over 30 time iterations starting from the first analysis time period (t=10).*

variables and constants are dimensionless. The model is discretized by standard finite differences on the mesh $[x_0 = 0, x_1, \ldots, x_N = 1]$ with uniform spacing $h = x_{k+1} - x_k = 1/N$ and mesh size $N = 100$. The MATLAB code is available from `http://www-math.cudenver.edu/~cjohns/fire1d`.

In this paper, we consider the differential equation in one spatial dimension. The model has ignition point $T_i = 300$. The initial values of S is 1 on the spatial domain between 0 and 1, except that $S = 0$ at a fuel break between 0.45 and 0.50. The reference solution profile of the initial conditions for temperature is shown in the top panel of Figure 5.1. Panels (b) and (c) show how the temperature profile of the reference solution propagates to time periods 0.05 and further to 0.10.

An initial ensemble of size 250 was generated by perturbing the reference solution temperature and fuel supply values with spatially correlated normal deviates. Six of the ensemble members were generated by shifting the reference solution right and

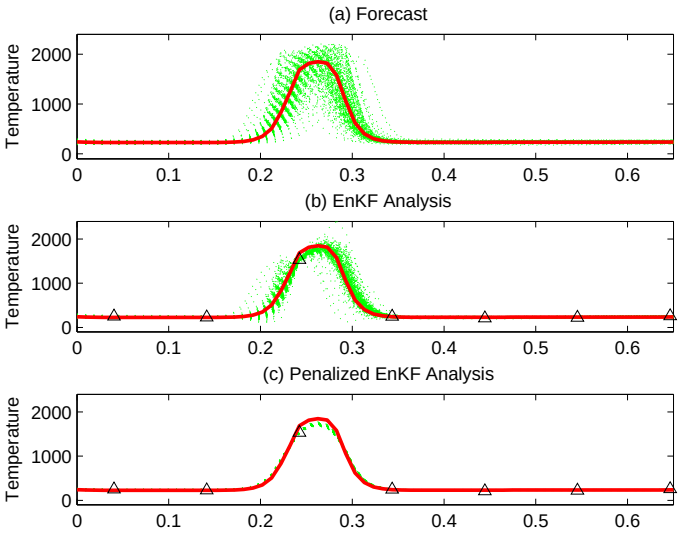

FIG. 5.3. *Panel (a) shows the temperature profiles of 250 ensemble members (dotted lines) forecast to the first analysis cycle (10 time iterations from initial conditions) and the corresponding reference solution (solid line). Panel (b) shows the temperature profiles (dotted lines) in the analysis ensemble corresponding to the EnKF update and includes the information in the data (△). Panel (c) compares the regularized analysis ensemble with the reference solution where the second analysis smooths the temperature via an approximate spatial derivative constraint.*

left spatially, and then adding perturbations. Each ensemble member is propagated forward ten time steps via the fire model to reach the time period shown in Figure 5.2 to generate the forecast or prior ensemble. Panel (a) of Figure 5.3 shows the 250 members of the forecast ensemble (dotted lines) and the associated true temperature profile (solid). Synthetic data (△) collected every 10 spatial units are shown in Panel (b) along with the ensemble (dotted lines) updated to the EnKF analysis stage. Note that the analysis step does a good job of cinching the temperature profiles to the few observations; however, the data information does not carry over to nearby locations (e.g. the analysis ensemble near 0.20 or 0.30 on the x-axis in panel (b).) This cinching phenomenon is a byproduct of the least-squares approach

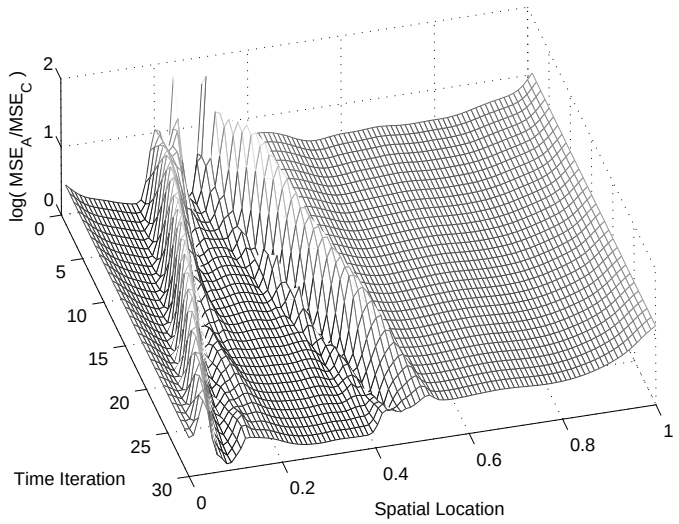

FIG. 5.4. *Log of the ratio of Mean-squared error (with respect to reference solution) of the analysis ensemble (MSE_A) and regularized analysis ensemble (MSE_R). Values greater than zero support the use of the two-stage shrinkage method.*

in the filter and can produces temperature profiles not compatible with common sense, theory, and numerical approximation schemes. Assuming a spatial correlation function amongst the observations can eliminate some of the cinching effect in the updates and produce smoother update ensemble members. However, we found that the strength of correlation required to make significant improvements were impractical and counterintuitive. Furthermore, we are interested in the particular situation where observations are accurate but spatially sparse.

To force a measure of smoothness upon the analysis, we impose a penalty on the spread of the first spatial derivative on the temperature analysis field using a second application of the EnKF code as described in Sec. 4. We choose the observation

function L by

$$L\mathbf{x} = L\begin{bmatrix} T \\ \\ S \end{bmatrix} = \begin{bmatrix} (T_1 - T_0)/h \\ \vdots \\ (T_N - T_{N-1})/h \end{bmatrix}$$

where $T_i \approx T(x_i)$ is the discrete temperature variable at node i. So, $L\mathbf{x}$ is the numerical derivative of the temperature field in the system state. In this example, D is diagonal with non-zero elements $d_{ii} = |z_i|/(2h^2)$, where $\mathbf{z} = L\mu$. That is, at each node, the variance expected in the temperature gradient is proportional to the average temperature gradient at that node in the prior. The values of d_{ii} are large enough not to influence reasonable smooth simulations perceptibly but they will suppress simulations that start going unstable and exhibit large swings on the scale of the mesh step h.

The regularized analysis field is shown in panel (c) of Figure 5.3. We now compare the forecasts from the EnKF analysis (A) and the regularized analysis (R). The ensembles shown in panels (b) and (c) of Figure 5.3 were propagated forward 30 time steps and the pointwise squared errors (MSE) from the reference solution were averaged over each spatial location to simulate a prediction. The log of the ratio of MSE's for the two sets of predictions, MSE_A/MSE_R, is shown in Figure 5.4. In every case, the regularized predictions had a smaller MSE value than those based on the original EnKF analysis ensemble. Because the reference solution is the basis for this MSE analysis, the effects of bias are included in the calculations via the usual $bias^2 + variance$ formula. A graphical comparison of the bias showed that the bias for both methods were of roughly the same magnitudes and generally smaller for the regularized method.

6. Conclusion. Including a regularization step with the EnKF can be considered changing the prior distribution and it was implemented by running EnKF update twice, once on the actual observation and once on the regularization term as another artificial observation. This additional update adds information just as the information added by the data likelihood, and it is justified by a belief about the properties of states that should result from the numerical simulations. In nonlinear problems, the region of valid simulation states (valid region) may be quite small and a linear ensemble Kalman filter update will take simulation states out of the valid region, particularly if the variance of the prior or the data likelihood are large, and a numerical breakdown results. The proposed regularization helps to keep the states within the valid region by statistically imposing restraints on the gradient. This two-step update is similar to a Bayesian hierarchical update and can be viewed as a shrinkage step.

We have applied the proposed improvement to EnKF for a simple fire model, which breaks down numerically because updated states under the usual Kalman filter results in large, nonphysical values at a later time. Since a sign of the problem is the emergence of spikes in the fire temperature, and since detecting sharp gradients is more sensitive than detecting large values, we have chosen as the regularization term the numerical derivative of the temperature field. The covariance of the regularization term was chosen to penalize large swings on the spatial mesh scale.

In small systems with fast updates, the effect of choosing the covariance of the regularization term can be considered similar to choosing a smoothing parameter in a nonparametric regression setting.

Acknowledgements. This research was supported by the National Science Foundation under grants 0325314 and 0417909.

REFERENCES

[1] B. Anderson and J. Moore, *Optimal Filtering*, Prentice Hall, Englewood Cliffs, NJ, 1979.

[2] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen, *LAPACK Users' Guide*, Society for Industrial and Applied Mathematics, Philadelphia, PA, third ed., 1999.

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

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

[5] L. S. Blackford, J. Choi, A. Cleary, E. D'Azevedo, J. Demmel, I. Dhillon, J. Dongarra, S. Hammarling, G. Henry, A. Petitet, K. Stanley, D. Walker, and R. C. Whaley, *ScaLAPACK Users' Guide*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1997.

[6] G. Burgers, P. J. van Leeuwen, and G. Evensen, *Analysis scheme in the ensemble Kalman filter*, Monthly Weather Review, 126 (1998), pp. 1719–1724.

[7] J. B. Carlin, N. Polson, and D. S. Stoffer, *A Monte Carlo approach to nonnormal and nonlinear state-space modeling*, Journal of the American Statistical Association, (1992), pp. 494–500.

[8] J. J. Dongarra, J. D. Croz, S. Hammarling, and I. Duff, *A set of level 3 Basic Linear Algebra Subprograms*, ACM Transactions on Mathematical Software, 16 (1990), pp. 1–17.

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

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

[11] G. Evensen, *The ensemble Kalman filter: Theoretical formulation and practical implementation*, Ocean Dynamics, 53 (2003), pp. 343–367.

[12] ———, *Sampling strategies and square root analysis schemes for the EnKF*, Ocean Dynamics, (2004), pp. 539–560.

[13] P. C. Hansen, *Rank-deficient and discrete ill-posed problems*, SIAM Monographs on

Mathematical Modeling and Computation, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1998.

[14] A. C. Harvey, *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge, 1989.

[15] P. Houtekamer and H. L. Mitchell, *Data assimilation using an ensemble Kalman filter technique*, Monthly Weather Review, 126 (1998), pp. 796–811.

[16] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic Press, 1970.

[17] T. A. Johansen, *On Tikhonov regularization, bias and variance in nonlinear system identification*, Automatica, 33 (1997), pp. 441–446.

[18] C. J. Johns and R. H. Shumway, *A state-space model for censored data*, Environmetrics, (2005), pp. 167–180.

[19] R. H. Jones, *Fitting multivariate models to unequally spaced data*, in Time Series Analysis of Irregularly Spaced Data, E. Parzen, ed., no. 25 in Lecture Notes in Statistics, Springer-Verlag, 1984, pp. 158–188.

[20] R. E. Kalman, *A new approach to linear filtering and prediction problems*, Trans. ASME J. Basic Eng., 82 (1960), pp. 35–45.

[21] R. E. Kalman and R. S. Bucy, *New results in filtering and prediction theory*, Trans. ASME J. Basic Eng., 83 (1961), pp. 95–108.

[22] K. Y. Kim, S. I. Kang, M. C. Kim, S. Kim, Y. J. Lee, and M. Vauhkonen, *Dynamic image reconstruction in electrical impedance tomography with known internal structures*, IEEE Transactions on Magnetics, 38 (2002), pp. 1301–1304.

[23] G. Kitagawa, *Non-Gaussian state-space modeling of nonstationary time series, (with discussion)*, Journal of the American Statistical Association, 400 (1987), pp. 1032–1044.

[24] J. Mandel, M. Chen, L. P. Franca, C. Johns, A. Puhalskii, J. L. Coen, C. C. Douglas, R. Kremens, A. Vodacek, and W. Zhao, *A note on dynamic data driven wildfire modeling*, in Computational Science - ICCS 2004, M. Bubak, G. D. van Albada, P. M. A. Sloot, and J. J. Dongarra, eds., vol. 3038 of Lecture Notes in Computer Science, Springer, 2004, pp. 725–731.

[25] ———, *Dynamic data driven wildfire modeling*, in Dynamic Data Driver Application System, F. Darema, ed., Kluwer, 2005, p. in print. UCD CCM Report 208, March 2004, http://www-math.cudenver.edu/ccm/reports/rep208.pdf.

[26] R. Meinhold and N. Singpurwalla, *Understanding the Kalman filter*, American Statistician, 37 (1983), pp. 123–127.

[27] ———, *Robustification of Kalman filter models*, Journal of the American Statistical Society, 84 (1989), pp. 479–486.

[28] C. C. Paige and M. A. Saunders, *Least squares estimation of discrete linear dynamic systems using orthogonal transformations*, SIAM J. Numer. Anal., 14 (1977), pp. 180–193.

[29] R. H. Shumway and D. S. Stoffer, *An approach to time series smoothing and forecasting using the EM algorithm*, Journal of Time Series Analsysis, 3 (1982), pp. 253–264.

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

[31] P. van Leeuwen, *A variance-minimizing filter for large-scale applications*, Monthly Weather Review, 131 (2003), pp. 2071–2084.