COMMUN. MATH. SCI.
Vol. 8, No. 1, pp. 27–43

CATASTROPHIC FILTER DIVERGENCE IN FILTERING NONLINEAR DISSIPATIVE SYSTEMS*

JOHN HARLIM† AND ANDREW J. MAJDA‡

Abstract. Two types of filtering failure are the well known filter divergence where errors may exceed the size of the corresponding true chaotic attractor and the much more severe catastrophic filter divergence where solutions diverge to machine infinity in finite time. In this paper, we demonstrate that these failures occur in filtering the L-96 model, a nonlinear chaotic dissipative dynamical system with the absorbing ball property and quasi-Gaussian unimodal statistics. In particular, catastrophic filter divergence occurs in suitable parameter regimes for an ensemble Kalman filter when the noisy turbulent true solution signal is partially observed at sparse regular spatial locations.

With the above documentation, the main theme of this paper is to show that we can suppress the catastrophic filter divergence with a judicious model error strategy, that is, through a suitable linear stochastic model. This result confirms that the Gaussian assumption in the Kalman filter formulation, which is violated by most ensemble Kalman filters through the nonlinearity in the model, is a necessary condition to avoid catastrophic filter divergence. In a suitable range of chaotic regimes, adding model errors is not the best strategy when the true model is known. However, we find that there are several parameter regimes where the filtering performance in the presence of model errors with the stochastic model supersedes the performance in the perfect model simulation of the best ensemble Kalman filter considered here. Secondly, we also show that the advantage of the reduced Fourier domain filtering strategy [A. Majda and M. Grote, Proceedings of the National Academy of Sciences, 104, 1124-1129, 2007], [E. Castronovo, J. Harlim and A. Majda, J. Comput. Phys., 227(7), 3678-3714, 2008], [J. Harlim and A. Majda, J. Comput. Phys., 227(10), 5304-5341, 2008] is not simply through its numerical efficiency, but significant filtering accuracy is also gained through ignoring the correlation between the appropriate Fourier coefficients when the sparse observations are available in regular space locations.

Key words. Kalman filter, Lorenz-96 model, filter divergence.

AMS subject classifications. 93E11, 62M20, 62L12, 65C20.

1. Introduction

The difficulties of many real time prediction problems range from extremely complex physical processes that are hardly understood, to expensive numerical computations when there are interactions between multi scale processes, and to our inability to obtain information or observations of the dynamical variables at every model grid point. Bayesian hierarchical modeling [6] and reduced order filtering strategies [29, 15, 34, 3, 4, 7, 30, 20, 16, 17] have been developed with some success for predicting extremely complex turbulent systems. The basis for such dynamic prediction strategies for complex spatially extended systems is the classical Kalman filtering algorithm [2, 9, 21], which is an optimal strategy that combines the information from the prior forecasts and the observations when the strict assumptions, linearity and Gaussian distributions, are satisfied.

The beauty of the ensemble Kalman filter [14, 7, 3] as one of the computationally cheap strategies for filtering spatially extended partially observed nonlinear dynamical systems lies in its simplicity in implementation, that is, it treats the model as a black box and it does not require any linear tangent or even linear adjoint model as the other alternative strategy, 4D-VAR [13, 11, 32, 31], which is often considered as

*Received: April 4, 2008; accepted (in revised version): July 23, 2008.

†Department of Mathematics and Center for Atmosphere and Ocean Science, Courant Institute of Mathematical Sciences, New York University, NY 10012 (jharlim@cims.nyu.edu).

‡Department of Mathematics and Center for Atmosphere and Ocean Science, Courant Institute of Mathematical Sciences, New York University, NY 10012 (jonjon@cims.nyu.edu).

the state-of-art in the weather prediction community. Instead, it estimates the true solution with an ensemble of solutions that is generated through the Kalman filter formulas whenever the observations are available. However, this supposedly Gaussian distributed ensemble is a sub-optimal sample when the ensemble size is too small or when the Gaussian assumption of the Kalman filter formula is not satisfied due to the nonlinearity of the model (for example, see [5, 29] for the failure of the ensemble Kalman filter in the presence of a bimodal distribution). There are two types of filtering failures: the first type, commonly referred to as filter divergence, is simply poorly filtered solutions with enormous errors that may exceed the size of the chaotic attractor. In many cases, as we will see in some of the numerical experiments below, the filtered solutions simply do not track the true signal. The second type of filtering failure is catastrophic filter divergence. In this case, some of the ensemble members become unbounded (machine infinite in finite time) due to ensemble collapse.

In this paper, we will show that a catastrophic filter divergence may occur even in filtering a chaotic dissipative nonlinear system with the absorbing ball property [10] when the turbulent signal is observed at regularly spaced sparse locations and in a perfect model situation with ensemble size twice the number of model variables. Specifically, we will show the catastrophic filter divergence in filtering the L-96 model [23] exhibited by two ensemble square root filters [33], ETKF (ensemble transform Kalman filter of Bishop [7], reformulated as in [20] for efficient implementation) and EAKF (ensemble adjustment Kalman filter of Anderson [3]), as variants of the ensemble Kalman filter. This dissipative chaotic dynamical system not only has the absorbing ball property but its solutions are unimodally distributed and only weakly skewed from a Gaussian distribution [1], which mimics a typical atmospheric turbulent system [25]. In the perfect model experiment with ETKF, the catastrophic filter divergence is exhibited in almost every chaotic regime ranging from weakly chaotic to fully turbulent when the observations are at most one half of the number of the model grid points. On the other hand, the identical perfect model experiment with EAKF produces very accurate solutions in the weakly chaotic and strongly chaotic regimes. However, EAKF is not immune from catastrophic filter divergence in a suitable sparse observed regime and it exhibits filter divergence with errors larger than the size of the chaotic attractor in stronger chaotic regimes with suitable parameters.

With the above documented filtering failures, the goal of this paper is to show that judicious model error through suitable linear stochastic models [18] is an alternative strategy to avoid a catastrophic filter divergence. We will show that although the filtering skill with these model errors is inferior to those with the perfect model when the nonlinearity is weak, there are parameter sets where this strategy produces more skillful results compared to the solutions with the perfect model simulations. In particular, we shall see that the consistent Gaussian distribution evolution via the linear stochastic model prohibits catastrophic filter divergence. With guaranteed Gaussian prior and posterior distributions, the remaining factors for assessing the filter inaccuracies are the sparse observations and the model errors.

The second point of this paper is to show that the reduced Fourier domain filtering strategies, formulated by one of the authors in [26] and advocated in the recent work of both authors in [8, 19] for filtering both plentifully observed and regularly spaced sparsely observed complex linear stochastic constant coefficient PDEs and for filtering plentifully observed nonlinear dynamical systems [18], is an alternative cheap filtering strategy that produces skillful filtered solutions beyond those produced via more expensive spatially based ensemble filters, ETKF and EAKF, when linear stochastic

models are used. In other words, we advocate that when the observations are sparse and only available at regularly spaced locations, it is not only numerically advantageous to simply ignore the cross correlation between appropriate different Fourier coefficients but it yields better accuracy in the presence of model errors. Furthermore, this Fourier domain imperfect model filter even supersedes the perfect model simulation with ETKF or EAKF in suitable parameter regimes. Finally, we find that this reduced filter offers an alternative efficient filtering strategy, similar to 3D-VAR [22], when the observations are very infrequent in time beyond the model decorrelation time.

The rest of this paper is organized as follows: In section 2 we state the L-96 model, review two square root filter strategies (ETKF and EAKF), show examples of catastrophic filter divergence, and review alternative linear stochastic strategies as in [18] as well as the Fourier domain filter reduction strategy as in [26, 19]. We then show numerical results in section 3 and conclude the paper with final discussion in section 4.

2. Filtering regularly spaced sparse observations

In this section, we describe the L-96 model [23] and show the difficulties of filtering this dissipative nonlinear toy model in certain chaotic regimes for regularly spaced sparse observations. In particular, we will show a catastrophic filter divergence with the Ensemble Transform Kalman Filter (ETKF of Bishop et al. [7], reformulated as in [20] for efficient implementation) and a much more stable and often quite accurate scheme, the Ensemble Adjustment Kalman Filter (EAKF of Anderson [3]), that sometimes still suffers from catastrophic filter divergence and sometimes simply diverges with errors larger than the size of the chaotic attractor of the corresponding model. Finally, we briefly describe a stochastic filtering strategy that is immune from catastrophic filter divergence but has model errors.

2.1. The Lorenz-96 model. The L-96 model [23] represents an "atmospheric variable" u at J equally spaced points around a circle of constant latitude. The jth component is propagated in time following differential equation

$$\frac{\mathrm{d}u_j}{\mathrm{d}t} = (u_{j+1} - u_{j-2})u_{j-1} - u_j + F \tag{2.1}$$

where the cyclic indices $j = 0,...,J-1$ represent the spatial coordinates ("longitude"). Note that this model is not a simplification of any atmospheric system, however, it is designed to satisfy three basic properties: it has linear dissipation (the $-u_j$ term) that decreases the total energy defined as $E = \frac{1}{2}\sum_{j=1}^{J} u_j^2$, an external forcing term F that can increase or decrease the total energy, and a quadratic advection term that conserves the total energy (i.e. it does not contribute to $\mathrm{d}E/\mathrm{d}t$) just like many atmospheric models (MW [28]). These properties combine to guarantee the absorbing ball property [10] for this model. Following Lorenz [23], MAG [24], and MW [28, p.239], we set $J = 40$ so that the distance between two adjacent grid points roughly represents the midlatitude Rossby radius ($\approx$ 800 km), assuming the circumference of the midlatitude belt is about 30,000 km.

2.2. Two ensemble square root filters. In this section, we briefly describe two ensemble square root filters which are Kalman filter based filtering methods. The basic idea of the ensemble Kalman filter is to produce a posterior ensemble $\{u^k_{m+1|m+1}, k = 1,\ldots,K\}$ that reflects both the best guess (ensemble mean $\bar{u}_{m+1|m+1}$)

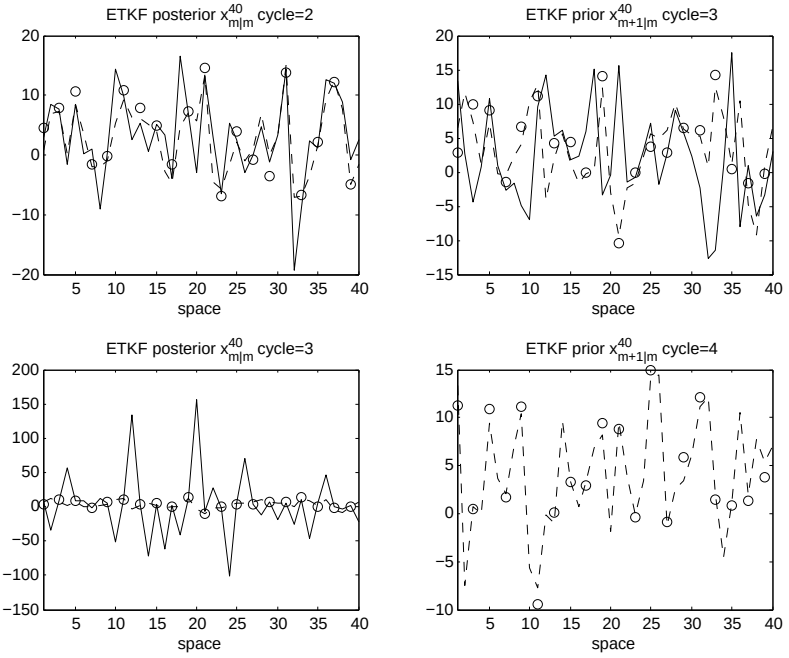

FIG. 2.1. *Snapshots of a single ensemble member exhibiting catastrophic filter divergence from the 80 member ensemble in ETKF at the fourth assimilation cycle. The parameters are* $F=16, p=2, T_{obs}=0.234, r^o=3$. *In each panel, we show the posterior or prior state (solid), the true signal (dashes), and observations (circles), as functions of the model space. In the last panel, the prior state is not plotted since it diverges to* ∞.

and the uncertainty (ensemble error covariance $r_{m+1|m+1}$) in the following fashion

$$u^k_{m+1|m+1} = \bar{u}_{m+1|m+1} + \delta U^k,$$

where δU^k is the k-th perturbation state to be determined. The statistical quantities $\bar{u}_{m+1|m+1}$ and $r_{m+1|m+1}$ are updated through a statistical least squares procedure [2, 9] by accounting for the information from observations or measurements at time step $m+1$ and the prior forecast statistical values $\bar{u}_{m+1|m}$ and $r_{m+1|m}$. This update in the data assimilation community is called the "analysis". The prior forecast statistical values, sometimes also referred as the "background", are obtained through feeding the previous posterior ensemble $\{u^k_{m|m}, k=1,\ldots,K\}$ to the model. This two-step process, forecast and analysis, completes one data assimilation cycle.

The ensemble transform Kalman filter (ETKF of Bishop et al. [7]) chooses

$$\delta U^k = \delta U^k_{m+1|m} T,$$

where perturbation vectors δU^k and $\delta U^k_{m+1|m} = u^k_{m+1|m} - \bar{u}_{m+1|m}$ are the k-th column of matrices δU and $\delta U_{m+1|m}$, respectively, and T is a transformation matrix to be determined such that the ensemble samples the posterior error covariance matrix, i.e.,

$$r_{m+1|m+1} = \frac{1}{K-1}\delta U(\delta U)^T = \frac{1}{K-1}\delta U_{m+1|m} T T^T (\delta U_{m+1|m})^T.$$

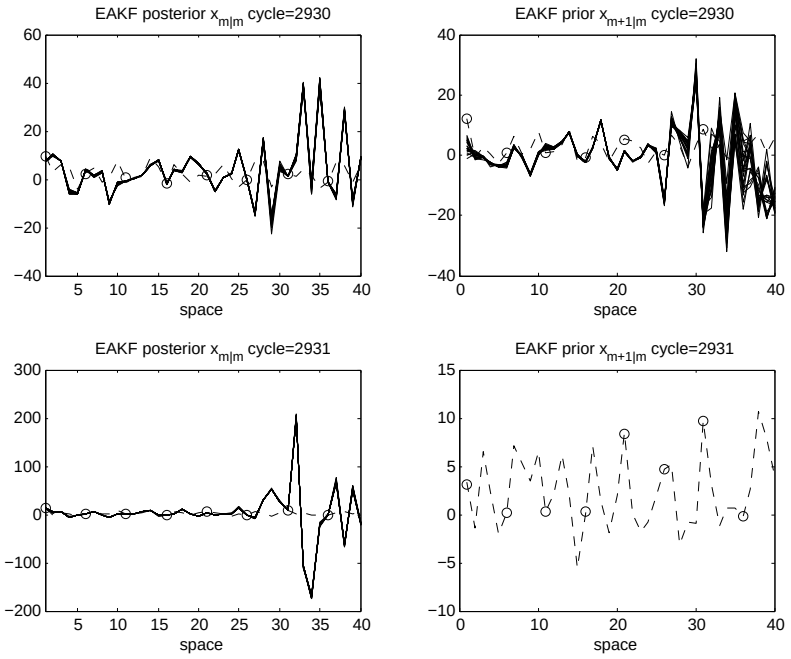

FIG. 2.2. *Snapshots of 27 members exhibiting catastrophic filter divergence from the 80 member ensemble in EAKF at assimilation cycle 2931. The parameters are* $F=8, p=5, \Delta t=0.234, r^o=0.01$. *All the posterior ensemble members on the left panels are collapsing on the same exploding large amplitude state. In each panel, we show the posterior or prior states (in solid black), the true signal (dashes), and observations (circles). In the last panel, the prior state is not plotted since it diverges to* ∞.

Practically, in the algorithm $r_{m+1|m+1}$ is never computed. Instead this form is inserted into the Kalman gain matrix and covariance update matrix as constraints to specify T through a singular value decomposition (SVD).

Anderson [3], on the other hand, introduced an alternative square root factor A and chose

$$\delta U^k = A\delta U^k_{m+1|m}$$

such that

$$r_{m+1|m+1} = \frac{1}{K-1}\delta U(\delta U)^T = \frac{1}{K-1}A\delta U_{m+1|m}(\delta U_{m+1|m})^T A^T.$$

This choice of ensemble update is called the ensemble adjustment Kalman filter (EAKF). Tippett et al. [33] showed the similarity between the two schemes by particular choices of A and T and the two schemes have identical compuational costs. In our implementation of EAKF, however, we follow [20] to avoid calculating a singular value decomposition on the prior covariance matrix $r_{m+1|m}$. Thus, EAKF is numerically more expensive than the reformulated ETKF. In real applications, however, EAKF is implemented sequentially [4] to avoid explicit computation of the SVD by processing the observations in smaller collected subset, or even individually.

In our simulations with ensemble size $K=80$ (double the model state space $J=40$), fixed variance inflation 5% to avoid ensemble collapse [5, 35], and 5000 assim-

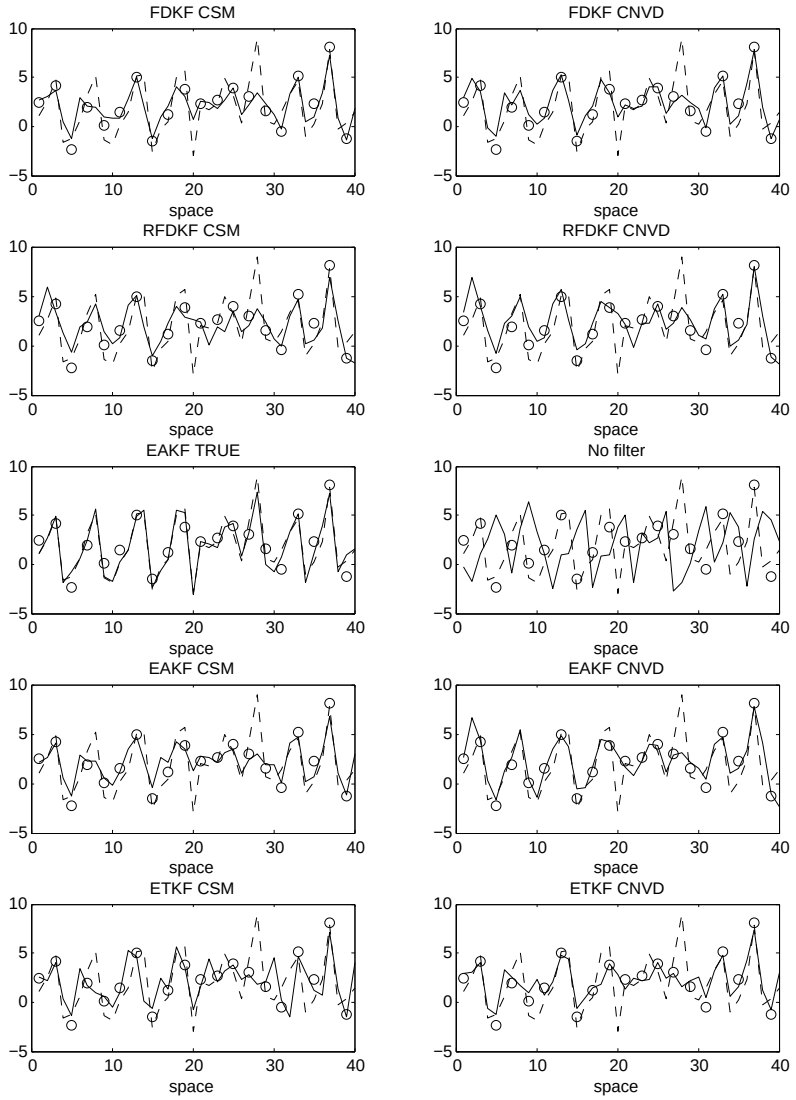

FIG. 2.3. *Snapshots of the filtered solutions at time 1172.1 (or after 5000 assimilation cycles) for* $F=6$, $p=2$, $r^o=1.96$, *and* $n=15$. *In each panel, the true signal is denoted in dashes, the observations in circles, and the filtered solution in the solid line.*

ilation cycles, we find that EAKF is a better scheme compared to ETKF when dealing with sparse observation. In particular, we find that ETKF suffers from catastrophic filter divergence when the observation error variance is at least $1/4$ of the chaotic attractor size for any sparse observed network ($p \geq 2$, i.e., the number of observations are at most $1/p$, that is, $1/2$ of the model grid points), regardless of whether the dynamics is weakly chaotic $F=6$, strongly chaotic $F=8$, or fully turbulent $F=16$. In figure 2.1, we show an example of catastrophic filter divergence exhibited by one of its ensemble members among a total of 80 ensemble members after only 3 steps of a data assimilation cycle when filtering the L-96 model with $F=16$, $p=2$, observation

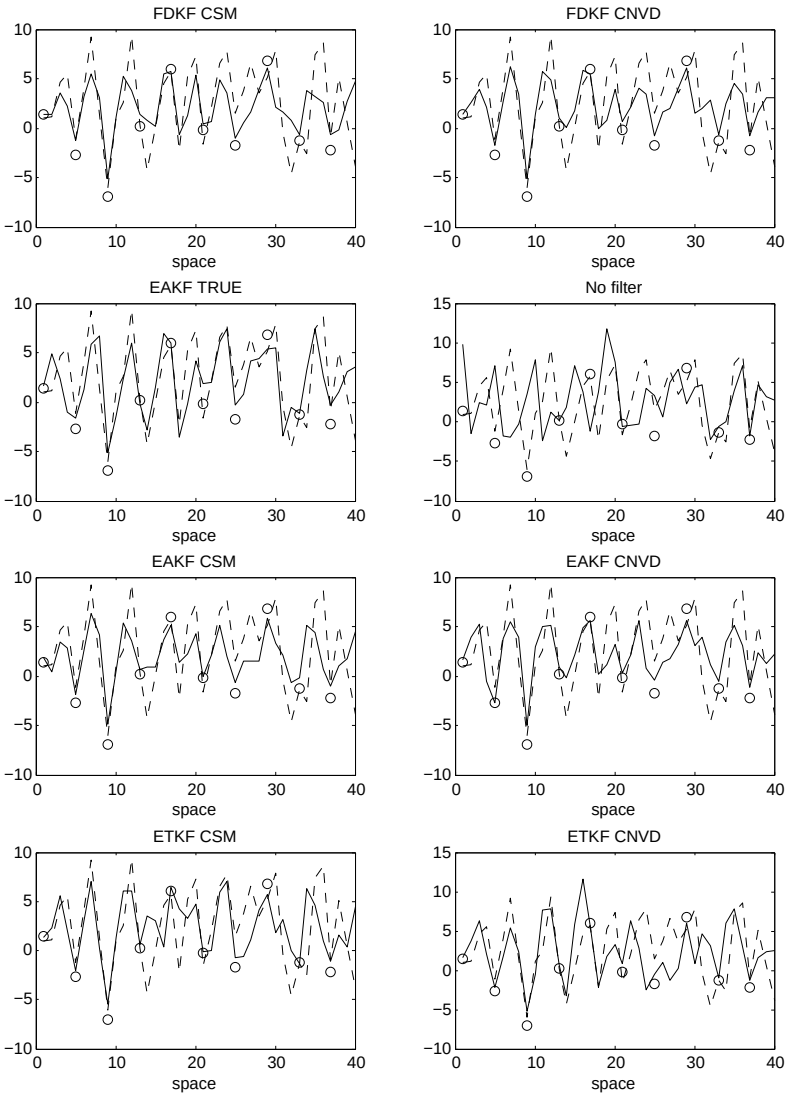

FIG. 2.4. *Snapshots of the filtered solutions at time 1172.1 (or after 5000 assimilation cycles) for* $F=8$, $p=4$, $r^o=3.24$, *and* $n=15$. *In each panel, the true signal is denoted in dashes, the observations in circles, and the filtered solution in the solid line.*

time $T_{obs}=0.234$ (corresponds to 28 hours according to the doubling time [23]), and observation noise size $r^o=3$. In this simulation, the initial ensemble is sampled from a Gaussian distribution with the true state as the mean and the equilibrium energy E as the variance.

In parallel simulations, EAKF is very skillful in almost every regime except

I) When observations are very sparse ($p=5$ or $1/5$ of the model grid points), the observation time $T_{obs}=0.234$ is infrequent, with very small observation noise variance $r^o=0.01$, EAKF suffers from a catastrophic filter divergence as well. In figure 2.2, we show snapshots of 27 ensemble members exhibiting catas-

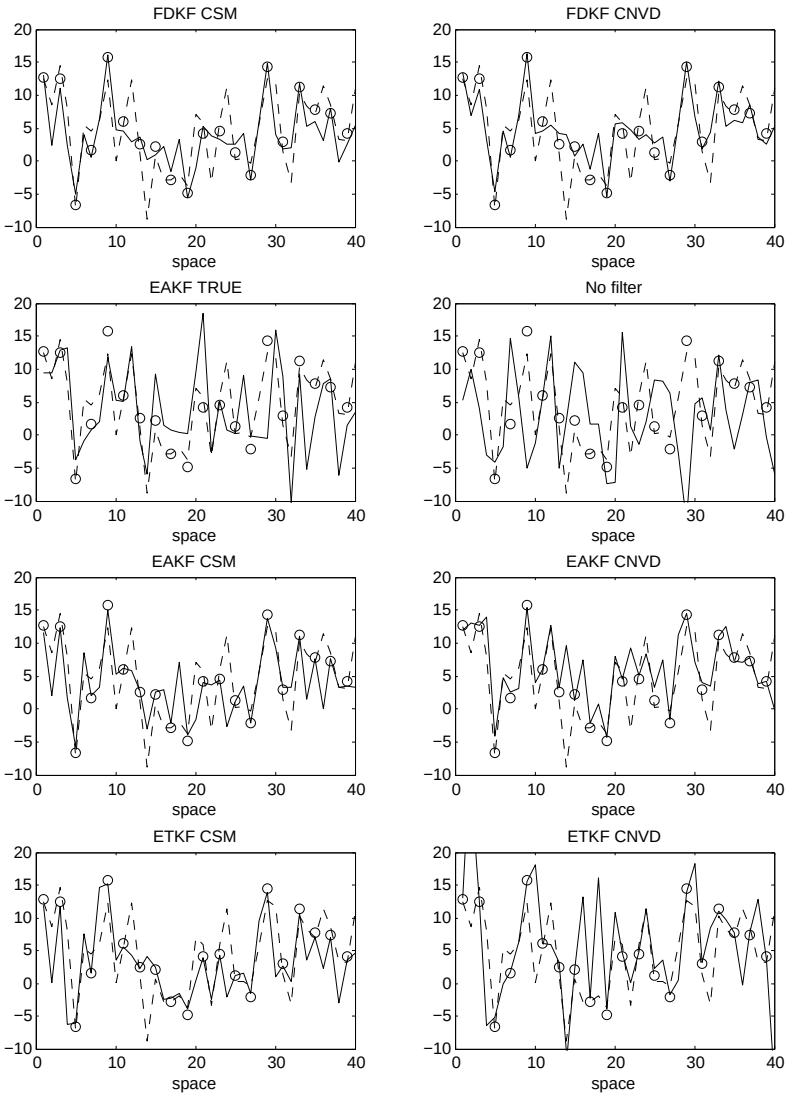

FIG. 2.5. *Snapshots of the filtered solutions at time 390.7 (or after 5000 assimilation cycles) for* $F=16$, $p=2$, $r^o=0.81$, *and* $n=5$. *In each panel, the true signal is denoted in dashes, the observations in circles, and the filtered solution in the solid line.*

trophic filter divergence from the 80 member ensemble in EAKF at assimilation cycle 2931. All the posterior ensemble members on the left panels are collapsing onto the same exploding large amplitude state.

II) When the model is fully turbulent ($F=16$) and the observation time is shorter than its decorrelation time, the filtered solution diverges (but not catastrophically) and becomes unskillful. See the panel denoted by "EAKF true" in figure 2.5 where we show the snapshot after 5000 assimilation cycles of simulation with $F=16$, $p=2$, $T_{obs}=0.078$, $r^o=0.81$. Also notice the poor average RMS error and low average spatial correlation in Table 3.3 for this corresponding

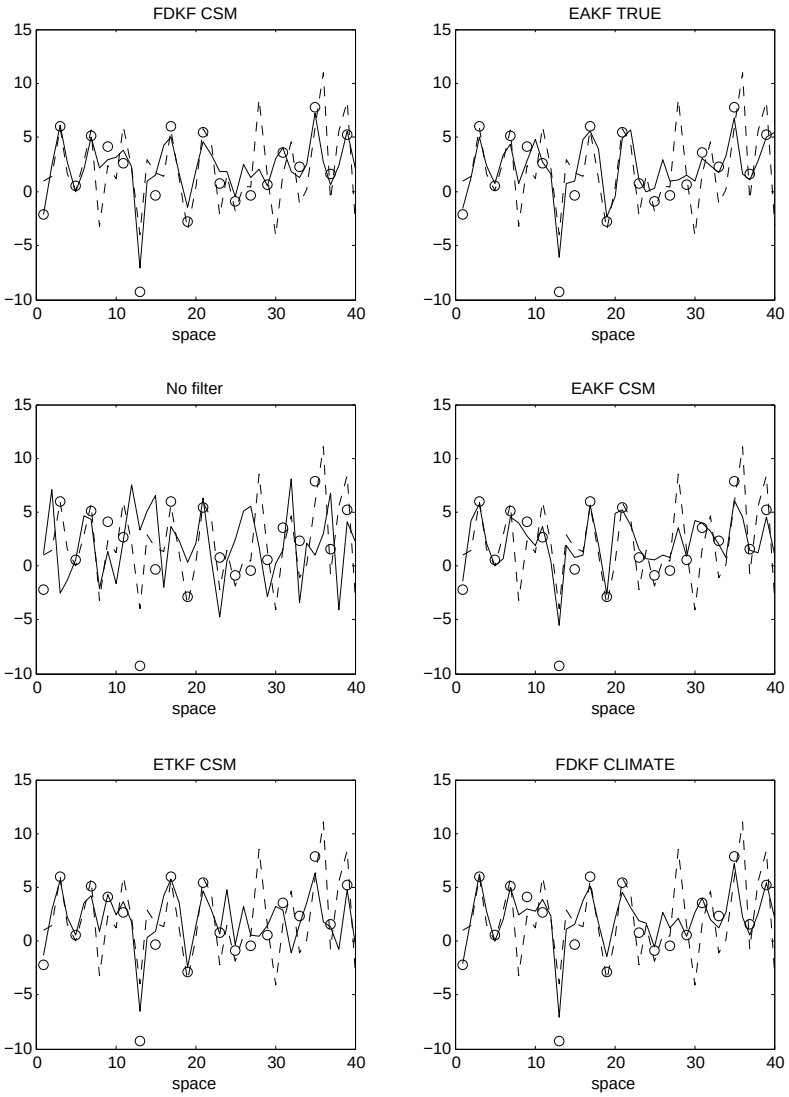

FIG. 2.6. *Snapshots of the filtered solutions at time 11,721.10 (or after 5000 assimilation cycles) for* $F=8$, $p=2$, $r^o=3.24$, *and* $n=150$. *In each panel, the true signal is denoted in dashes, the observations in circles, and the filtered solution in the solid line.*

perfect model simulation.

2.3. Linear Stochastic Model. An alternative approach is to introduce radical model errors through stochastic linear models [18]. Instead of using the full L-96 model (2.1), for each Fourier mode, we propose a filter model that solves a temporal discretization of the following stochastic differential equation

$$\frac{\mathrm{d}\hat{u}_k(\tilde{t})}{\mathrm{d}\tilde{t}} = E_p^{-1}(F-\bar{u}) + (A_k - d_k)\,\hat{u}_k(\tilde{t}) + \sigma_k \dot{\tilde{W}}_k(\tilde{t}), \tag{2.2}$$

where $\hat{u}_k(\tilde{t})$ is the Fourier coefficient of the renormalized L-96 model [18] through

$$\tilde{u} = \frac{u-\bar{u}}{\sqrt{E_p}}, \quad \tilde{t} = \sqrt{E_p}t \tag{2.3}$$

such that $\tilde{u}$ has zero mean and unit energy variance. In (2.2)–(2.3), E_p is the average variance in the energy fluctuation (see MAG [24] or MW [28] for details on how to compute these quantities), $\bar{u}$ represents the (temporal) mean state, and

$$A_k = E_p^{-1/2}\left[(\mathrm{e}^{2\pi \mathrm{i}k/J} - \mathrm{e}^{-4\pi \mathrm{i}k/J})\bar{u} - 1\right] \tag{2.4}$$

is the linear dynamical operator. This linear model (2.2) also includes a linear damping term $d_k > 0$ and a stochastic white noise term with noise strength σ_k; these two terms are added to reproduce the truncated nonlinearity with hope that they generate a statistically reasonably approximate solution for filtering the L-96 model (2.1) following the simplest stochastic modelling strategies for chaotic signals [18, 12]. In real space, the decoupled stochastic white noise term corresponds to the following correlated stochastic forcing

$$\sigma \circ \dot{W} = \sum_{|k| \le J/2} \sigma_k \mathrm{e}^{2\pi \mathrm{i}kj/J} \dot{\tilde{W}}_k, \tag{2.5}$$

where $\tilde{W}_k = (\tilde{W}_{k,1} + \mathrm{i}\tilde{W}_{k,2})/\sqrt{2}$ where $\tilde{W}_{k,j}$ is a Wiener process and notation $\sigma \circ \dot{W}$ describes the inverse Fourier transform of the white noise.

In discrete form, given an observation time $T_{obs} = t_{m+1} - t_m$ we can write our filter model for each Fourier mode as

$$\hat{u}_{k,m+1|m} = const. + F_k \hat{u}_{k,m|m} + \eta_{k,m}, \tag{2.6}$$

where the prior state $\hat{u}_{k,m+1|m}$ denotes the kth Fourier coefficient before we include observations at time t_{m+1}, while the posterior state $\hat{u}_{k,m|m}$ denotes the the kth Fourier coefficient after we include observations at time t_m (or after analysis). These two states are defined as an estimate of the true signal $\hat{u}_{k,m}$ generated by the nonlinear L-96 model that is hidden from the filter.

The filter model (2.6) has three terms: a constant term $E_p^{-1}(F-\bar{u})(A_k - d_k)^{-1}(1-\mathrm{e}^{(A_k-d_k)T_{obs}})$ only for the zeroth mode, a dynamical operator F_k, and a noise term $\eta_{k,m} \sim \mathcal{N}(0, r_k)$ where the variance r_k is given as follows

$$r_k = \frac{\sigma_k^2}{2Re\{d_k - A_k\}}\left(1 - \mathrm{e}^{-2Re\{d_k - A_k\}T_{obs}}\right). \tag{2.7}$$

This leaves us to wonder which values of d_k and σ_k yield better filtered solutions. In this paper, we show results with fitting the model to the climatological statistical values (as in [18], we referred to this parametrization strategy as the climatological stochastic model or CSM). In this strategy, the variances σ_k and the damping d_k are chosen uniquely to match the climatological energy spectrum and correlation time and are independent of the observation time, Δt. This is the simplest linear stochastic modelling strategy [12]. A second parametrization strategy is to fit only the noise term σ_k to the climatological statistics but the damping term is chosen to minimize the RMS errors at a given observation time, Δt, when plentiful observations are available

(again, see [18] for details). We referred to this method as the climate noise varying damping model or simply CNVD.

In this paper, we assume that observations $\hat{v}_{k,m}$ (the Fourier coefficients of the rescaled observations, $\tilde{v}=(v-\bar{u})/\sqrt{E_p}$) are sparsely available at every p model grid point with observation errors reflected by a Gaussian distribution, i.e., $\hat{\sigma}^o_m \sim \mathcal{N}(0,\hat{r}^o)$. In the real domain, this variance corresponds to $r^o = JE_p\hat{r}^o/p$. By Theorem 3 in [26] or as in [19], the Fourier space representation of the observation is given as follows

$$\hat{v}_{\ell,m} = \sum_{k\in\mathcal{A}(\ell)} \hat{u}_{k,m} + \hat{\sigma}^o_m, \tag{2.8}$$

where

$$\mathcal{A}(\ell) = \{k | k = \ell + q(J/p), |k| \le \frac{J}{2}, q\in\mathbb{Z}\} \tag{2.9}$$

is the aliasing set for mode ℓ. For simplicity, we assume that J/p, which represents the total number of observations, is an integer value. As an example, consider $J=40$ as in the L-96 model and $p=2$; then the model has 21 Fourier modes whereas the observations have 11 Fourier modes. According to formula (2.9), there are a total of 11 aliasing sets: $\mathcal{A}(0)=\{0,-20\}$, $\mathcal{A}(1)=\{1,-19\},\ldots,\mathcal{A}(9)=\{9,-11\}$, and $\mathcal{A}(10)=\{10\}$. Thus, the filtering problem is decoupled into 10 two-dimensional filters with scalar observations for Fourier modes in the aliasing sets 0 to 9 and a scalar filter on the 10th mode. This filtering reduction simply ignores the correlation between the Fourier coefficients in different aliasing sets. As in [18, 19], we referred to this filtering reduction as the Fourier domain Kalman filter or simply FDKF.

We will also show a variant of FDKF which simply filters only the most energetic mode in each aliasing set. Therefore, this approximate filter reduces to uncoupled scalar filters in which the unfiltered Fourier coefficients trust the dynamics fully. We call this approach the reduced Fourier domain Kalman filter or RFDKF following [19].

3. Numerical Results

In each numerical simulation, a true trajectory u^t isgenerated by integrating the L-96 model (2.1) with the Runge-Kutta method with time step $\Delta t = 1/64$. The sparse observations are simulated by adding uncorrelated Gaussian noises at every p model grid point with variance r^o to the true trajectories at every observation time $T_{obs} = n\Delta t$. When $F=8$, Lorenz suggests that 0.05 non-dimensionalized units are equivalent to 6 hours based on doubling time in a global weather model [23]. Thus, our choices of n's correspond to roughly 9 hours for $n=5$ and 28 hours for $n=15$; both observation times are still within the 3 days decorrelation time. The observation noise variance r^o is chosen relative to the fraction of the size of the chaotic attractor that can be roughly estimated by taking the temporal average of the RMS (root-mean-square) difference between two long trajectories initiated from two almost identical model states. In our numerical results, we call this average difference as the errors due to "no filter". The ETKF and EAKF are initiated by an exactly similar ensemble of initial conditions that is randomly chosen from a Gaussian distribution with the true state u^t_o as the mean and average variance of the energy spectrum E_p as the covariance. To be consistent with the simulation with ETKF and EAKF, we initiate the Fourier domain filters exactly from the true state.

In each experiment, we run the data assimilation cycle for 5000 cycles and to measure the performance we compute the temporal average of the RMS error and the average spatial correlation, both, between the true state u^t_m and the mean posterior

state $\bar{u}_{m|m}$. As a benchmark, we show results from a perfect model simulation with EAKF (which we denoted as "EAKF true" in the remainder of this paper) and the unfiltered solution. We do not show the 'ETKF true', that is ETKF with perfect model, since its solution blows up, i.e., it always suffers from catastrophic filter divergence, in the parameter regimes considered below. In this paper we will not compare the filtered solutions to the observation time model error as in [18], instead we provide simulations of EAKF and ETKF for both linear stochastic models CSM and CNVD. In this sense, we can compare how well FDKF performs compared to both ensemble methods in the presence of model errors.

In the first numerical simulation, we consider a weakly chaotic regime $F=6$ with a long observation time $n=15$ (i.e., $T_{obs}=0.234$). We consider sparse observations with $p=2$ and error variance $r^o=1.96$ to be the square of one-half of our rough estimate of the size of the chaotic attractor, which is 2.8 for this weakly chaotic regime [24, 28]. Figure 2.3 shows snapshots of the filtered solutions after 5000 assimilation cycles from various approaches including the perfect model simulation EAKF true, the four filtering strategies EAKF, ETKF, FDKF, and RFDKF in the presence of model errors through CSM and CNVD, and the unfiltered solutions. Notice that the filtered solutions of the Fourier domain filters with model errors are less skillful than those with the EAKF true; in the region with rather large amplitude of oscillations, the presence of model errors hurts the filter when observations are not available. However, the Fourier domain FDKF shows substantial skill (much better than the unskillful solutions) and notice that the model errors through CSM and CNVD also hurt both ensemble methods, EAKF and ETKF, at similar unobserved locations (compare snapshots in figure 2.3). In fact, the two ensemble filters with model errors produce larger average RMS errors and lower average spatial correlations compared to the exactly identical assimilation with FDKF (see Table 3.1). It is interesting that the catastrophic filter divergence in ETKF vanishes when both linear stochastic models (CSM and CNVD), which essentially include model errors, are implemented. We also notice that there is filtering skill even with the cheapest scheme RFDKF with solutions comparable to those produced by ETKF. The surprising skill of RFDKF can be justified as follows: the strongly energetic modes in the weakly chaotic L-96 model are concentrated within modes 6 to 10 [24, 28] and the structure of the aliasing set defined in (2.9) allows these energetic modes to be distributed in different aliasing sets so that the unfiltered modes (modes where RFDKF fully trusts the dynamics) are the weakly energetic modes. On the other hand, when the model is strongly chaotic $F=8$ or fully turbulent $F=16$, the energy spectrum becomes more homogeneous so that ignoring some of the energetic modes may hurt the filter substantially. In other unreported experiments, we confirm the declining skill with RFDKF for stronger chaotic regimes. In the second experiment, we consider sparser observations with $p=4$. The results in figure 2.4 are from a simulation in the strongly chaotic regime $F=8$ with observation noise size $r^o=3.24$ (i.e., square of a one-half of the chaotic attractor size 3.6) and observation time $n=15$ ($T_{obs}=0.234$). Qualitatively, we find that the filtered solutions underestimate the unobserved peaks in the true signal even with perfect model simulation. The presence of model errors through linear stochastic models do not magnify the RMS errors as significantly as we saw in the earlier experiments; CNVD increases errors by about 1.17 times in this experiment, compared to 3 times in the earlier experiment. In this experiment, the filtered solution with the perfect model simulation, EAKF true, produces filtered solutions with the lowest average RMS error of 2.69 and the highest average spatial correlation of 0.68. In the presence of model

TABLE 3.1. *Average RMS errors and spatial correlations for simulations with $F=6$, $p=2$ $r^o=1.96$, and $n=15$ (corresponding to $T_{obs}=0.234$). This is a regime where EAKF true is superior.*

Spatial domain scheme	RMS	corr.	Fourier domain scheme	RMS	corr.
EAKF true	0.82	0.95			
EAKF CSM	2.20	0.64	FDKF CSM	2.07	0.69
EAKF CNVD	2.09	0.70	FDKF CNVD	1.99	0.73
ETKF true	∞	-			
ETKF CSM	2.50	0.55	RFDKF CSM	2.39	0.60
ETKF CNVD	2.39	0.62	RFDKF CNVD	2.38	0.62

errors, the RMS errors with CNVD is lower than those with CSM. Comparing both the average RMS error and the average spatial correlation for linear stochastic model CNVD, the highest skill is obtain through FDKF, follows by EAKF, and ETKF (see Table 3.2). In the third numerical experiment (see figure 2.5 for the snapshots of

TABLE 3.2. *Average RMS errors and spatial correlations for simulations with $F=8$, $p=4$, $r^o=3.24$, and $n=15$ (corresponds to $T_{obs}=0.234$). This is a regime where EAKF true is mildly superior.*

Spatial domain scheme	RMS	corr.	Fourier domain scheme	RMS	corr.
EAKF true	2.69	0.68			
EAKF CSM	3.28	0.46	FDKF CSM	3.16	0.51
EAKF CNVD	3.15	0.52	FDKF CNVD	3.06	0.56
ETKF true	∞	-			
ETKF CSM	4.18	0.32			
ETKF CNVD	3.66	0.43			

the filtered solution after 5000 assimilation cycles and Table 3.3 for its corresponding average RMS errors and spatial correlations), we consider a fully turbulent regime $F=16$ with a shorter observation time $n=5$ ($T_{obs}=0.078$), observation density $p=2$, and observation error $r^o=0.81$ (which is the square of a quarter of it's chaotic attractor size 6.4). In this regime, the perfect model experiment, EAKF true, is completely unskillful with average RMS error 7.55, larger than the attractor size 6.4. On the other hand, the model errors either through CSM or CNVD improve the filtering skill (again, see Table 3.3). The model errors also improve the EAKF substantially, but the numerically cheaper FDKF still has the lowest average errors and highest average correlation. Finally, ETKF, which suffers from a catastrophic filter divergence with perfect model, has some skill when radical model errors through CSM are introduced but still has conventional filter divergence with CNVD (although not catastrophically). The better filtering skill through the linear stochastic model compared to the EAKF true, as discussed in this numerical experiment is also observed when the L-96 is strongly chaotic with $F=8$ and intermediate observation times with $n=5$ to $n=7$. Finally, we consider a super-long observation time with $n=150$ ($T_{obs}=2.34$, this corresponds to 12 days which is far beyond the 3 days decorrelation time of the L-96 model with $F=8$ [23]). In this numerical experiment, we set all parameters as in the second experiment: $F=8, r^o=3.24$, but with $p=2$ instead of $p=4$. We find that (see also figure 2.6 for snapshots) the average RMS error and the average spatial

TABLE 3.3. *Average RMS errors and spatial correlations for simulations with $F=16$, $p=2$, $r^o=0.81$, and $n=5$ (corresponding to $T_{obs}=0.078$). This is a regime where FDKF is superior.*

Spatial domain scheme	RMS	corr.	Fourier domain scheme	RMS	corr.
EAKF true	7.55	0.48			
EAKF CSM	5.15	0.61	FDKF CSM	4.80	0.66
EAKF CNVD	5.87	0.57	FDKF CNVD	4.78	0.68
ETKF true	∞	-			
ETKF CSM	5.80	0.54			
ETKF CNVD	7.45	0.45			

correlation of the perfect model simulation EAKF true are, respectively, 2.96 and 0.59, which are slightly worse than those of the FDKF CSM (RMS=2.82, corr=0.64), as expected given the earlier trends, EAKF CSM (RMS=3.03, corr=0.57) and ETKF CSM (RMS=3.21, corr=0.53), do not produce better filtered solutions compared to FDKF or EAKF true. In each filtering strategy, we ignore simulations with CNVD since its filtering skills are not much different from CSM. In fact, we find similar filtering skill if we simply use the climatological mean and covariance as the prior statistics (FDKF CLIMATE in figure 2.6 produces RMS=2.81,corr=0.64). Thus, for a very long observation time beyond the decorrelation time, a dynamic-less filter akin to the 3D-VAR [22] scheme is an alternative strategy.

4. Summary

In this paper we show that in filtering the dissipative nonlinear L-96 model [23] with an absorbing ball property and a unimodal quasi-Gaussian distribution, when the number of observations are at most one half of the model spatial dimension and when these observations are located at regularly spaced grid points the perfect model simulation with ETKF (reformulated as in [20]) exhibits catastrophic filter divergence in various turbulent regimes. Moreover, the perfect model simulation with EAKF, which is a more stable and very skillful filtering scheme compared to ETKF, is not immune from catastrophic filter divergence. Our numerical simulations show that in several chaotic regimes where the perfect model simulation with EAKF does not suffer from a catastrophic filter divergence, this method produces unskillful results with errors larger than their corresponding attractor sizes which is a more conventional filter divergence.

As an alternative, we implement the Fourier domain Kalman filter (FDKF), an innovative filter reduction strategy for filtering turbulence as shown in [19] and formulated in [26], to linear stochastic models [18] of the L-96 model. We compare this strategy with the perfect model simulations and simulations with model errors (through the same linear stochastic models used in the Fourier domain filters) with EAKF and ETKF. From these experiments, we find that in the weakly chaotic regime, the perfect model simulation with EAKF is superb and unbeatable compared to any simulations with model errors. Here, ETKF suffers from a catastrophic filter divergence. Secondly, among the simulations with model errors we find that FDKF is a better alternative strategy and this suggests that ignoring the correlations between Fourier coefficients in different aliasing sets (which is not done in both ensemble methods EAKF and ETKF) improves the filtered solutions. Most importantly, our numerical experiments also show that the improvement through this approach is robust throughout all regimes we tested, which confirms the superiority of this reduced

filtering strategy. Thirdly, we find that trusting fully the dynamics in the weakly energetic modes within each aliasing set, which is prescribed in RFDKF, produces comparable results as those with the computationally more expensive ETKF in the weakly chaotic regime, $F=6$. However, when the system is strongly chaotic or fully turbulent, since the energy spectrum is more uniformly distributed, then trusting fully the dynamics (as in RFDKF) in the potentially strongly energetic truncated modes can lead to very poor filtering skill.

We find that in the fully turbulent regime, $F=16$, and with short observation time (much shorter than the decorrelation time), the model errors through linear stochastic models improve the filtering skill; FDKF is the most skillful filter for this situation; in EAKF, the unskillful filter divergence with the perfect model is improved to a reasonably more skillful result; for ETKF with the stochastic models, the catastrophic filter divergence with the perfect model is avoided, however, the filtered solution still diverges when CNVD is used. The fact that FDKF supersedes the perfect model simulation with EAKF is rather interesting. The nonlinearity in the fully turbulent L-96 model, in fact, only skews the prior ensemble $\{x^k_{m+1|m}\}$ sample slightly from a Gaussian distribution [1] and what is surprising is that this quasi-Gaussian prior density is sufficient to cause the EAKF to diverge and the ETKF to exhibit catastrophic filter divergence. On the other hand, the model errors through the linear stochastic model guarantee a Gaussian prior ensemble $\{x^k_{m+1|m}\}$ distribution. We also find that it is beneficial to consider a cheap dynamics-less filter (which is a cheaper version of the 3D-VAR [22] with climatological prior distribution) whenever the observation time is longer than the model decorrelation time. Finally, it is important to note here that filtering sparse regular observation in Fourier space with the climate stochastic model, CSM, always gives essentially comparable and sometimes superior results as the CNVD model yet requires only minimal information of the climatological variances and correlation times in fitting this model. Thus, CSM would be the practical algorithm of choice for the context here.

From the viewpoint of nonlinear analysis, in filtering sparse observations the catastrophic filter divergence of ETKF and EAKF in a chaotic dynamical system with the absorbing ball property clearly needs further mathematical theory. The catastrophic divergence in figures 2.1 and 2.2 resembles classical nonlinear instability for finite difference schemes (see p.66 of [27]) but new mechanisms occur here.

Acknowledgment. The research of A.J. Majda is partially supported by the National Science Foundation grant DMS-0456713, the Office of Naval Research grant N00014-05-1-0164, and the Defense Advanced Research Projects Agency. John Harlim is supported as a postdoctoral fellow through the last two agencies. The authors thank Jeff Anderson for his help and insights with EAKF in the early stages of this work.

REFERENCES

[1] R. Abramov and A. Majda, *Quantifying uncertainty for non-Gaussian ensembles in complex systems*, SIAM J. Sci. Comput., 25(2), 411–447, 2004.
[2] B. Anderson and J. Moore, *Optimal Filtering*, Prentice-Hall, Englewood Cliffs, NJ, 1979.
[3] J. Anderson, *An ensemble adjustment Kalman filter for data assimilation*, Monthly Weather Review, 129, 2884–2903, 2001.
[4] J. Anderson, *A local least squares framework for ensemble filtering*, Monthly Weather Review, 131(4), 634–642, 2003.
[5] J. Anderson and S. Anderson, *A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts*, Monthly Weather Review, 127, 2741–2758, 1999.

[6] L. Berliner, R. Milliff and C. Wikle, *Bayesian hierarchical modeling of air-sea interaction*, J. Geophysical Research, 108, 3104–3120, 2003.
[7] C. Bishop, B. Etherton and S. Majumdar, *Adaptive sampling with the ensemble transform Kalman filter part I: the theoretical aspects*, Monthly Weather Review, 129, 420–436, 2001.
[8] E. Castronovo, J. Harlim and A. Majda, *Mathematical criteria for filtering complex systems: plentiful observations*, J. Comput. Phys., 227(7), 3678–3714, 2008.
[9] C. Chui and G. Chen, *Kalman Filtering*, Springer New York, 1999.
[10] P. Constantin, C. Foias, B. Nicolaenko and R. Temam, *Integral Manifolds and Inertial Manifolds for Dissipative Partial Differential Equations*, Appl. Math. Sci., Springer, 70, 1988.
[11] P. Courtier, J.N. Thépaut and A. Hollingsworth, *A strategy for operational implementation of 4D-VAR, using an incremental approach*, Quarterly J. Royal Meteorological Society, 120, 1367–1387, 1994.
[12] T. Delsole, *Stochastic model of quasigeostrophic turbulence*, Surveys in Geophysics, 25(2), 107–149, 2004.
[13] F.X.L. Dimet and O. Talagrand, *Variational algorithm for analysis and assimilation of meteorological observations: theoretical aspects*, Tellus A, 38, 97–110, 1986.
[14] G. Evensen, *Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics*, J. Geophysical Research, 99, 10143–10162, 1994.
[15] M. Ghil and P. Malanotte-Rizolli, *Data assimilation in meteorology and oceanography*, Advances in Geophysics, 33, 141–266, 1991.
[16] J. Harlim and B. Hunt, *A non-Gaussian ensemble filter for assimilating infrequent noisy observations*, Tellus A, 59(2), 225–237, 2007.
[17] J. Harlim and B. Hunt, *Four-dimensional local ensemble transform Kalman filter: numerical experiments with a global circulation model*, Tellus A, 59(5), 731–748, 2007.
[18] J. Harlim and A. Majda, *Filtering nonlinear dynamical systems with linear stochastic models*, Nonlinearity, 21(6), 1281–1306, 2008.
[19] J. Harlim and A. Majda, *Mathematical strategies for filtering complex systems: regularly spaced sparse observations*, J. Comput. Phys., 227(10), 5304–5341, 2008.
[20] B. Hunt, E. Kostelich and I. Szunyogh, *Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter*, Physica D, 230, 112–126, 2007.
[21] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, Springer New York, 2005.
[22] A. Lorenc, *Analysis methods for numerical weather prediction*, Quarterly J. Royal Meteorological Society, 112, 1177–1194, 1986.
[23] E. Lorenz, *Predictability - a problem partly solved*, Proceedings on predictability, held at ECMWF on 4-8 September 1995, 1–18, 1996.
[24] A. Majda, R. Abramov and M. Grote, *Information theory and stochastics for multiscale nonlinear systems*, CRM Monograph Series v.25, American Mathematical Society, Providence, Rhode Island, USA, 2005.
[25] A. Majda, C. Franzke and B. Khouider, *An applied mathematics perspective on stochastic modelling for climate*, Philos Transact A Math. Phys. Eng. Sci., 366(1875), 2429–2455, 2008.
[26] A. Majda and M. Grote, *Explicit off-line criteria for stable accurate time filtering of strongly unstable spatially extended systems*, Proceedings of the National Academy of Sciences, 104, 1124–1129, 2007.
[27] A. Majda and I. Timofeyev, *Statistical mechanics for truncations of the Burgers-Hopf equation: a model for intrinsic behavior with scaling*, Milan J. of Mathematics, 70, 39–96, 2002.
[28] A. Majda and X. Wang, *Nonlinear Dynamics and Statistical Theories for Basic Geophysical Flows*, Cambridge University Press, UK, 2006.
[29] R. Miller, E. Carter and S. Blue, *Data assimilation into nonlinear stochastic models*, Tellus A, 51(2), 167–194, 1999.
[30] E. Ott, B. Hunt, I. Szunyogh, A. Zimin, E. Kostelich, M. Corrazza, E. Kalnay and J. Yorke, *A local ensemble Kalman filter for atmospheric data assimilation*, Tellus A, 56, 415–428, 2004.
[31] F. Rabier, H. Jarvinen, J.F. Mahfouf and A. Simmons, *The ECMWF operational implementation of four-dimensional variational assimilation: experimental results with simplified physics*, Quarterly J. Royal Meteorological Society, 126, 1143–1170, 2000.
[32] F. Rabier, J.N. Thépaut and P. Courtier, *Extended assimilation and forecast experiments with a four-dimensional variational assimilation system*, Quarterly J. Royal Meteorological Society, 124, 1–39, 1998.
[33] M. Tippett, J. Anderson, C. Bishop, T. Hamill and J. Whitaker, *Ensemble square-root filters*, Monthly Weather Review, 131, 1485–1490, 2003.

[34] R. Todling and M. Ghil, *Tracking atmospheric instabilities with the Kalman filter. Part I: methodology and one-layer results*, Monthly Weather Review, 122(1), 183–204, 1994.
[35] J. Whitaker and T. Hamill, *Ensemble data assimilation without perturbed observations*, Monthly Weather Review, 130, 1913–1924, 2002.