IOP PUBLISHING NONLINEARITY

Nonlinearity **21** (2008) 1281–1306 doi:10.1088/0951-7715/21/6/008

# Filtering nonlinear dynamical systems with linear stochastic models

**J Harlim and A J Majda**

Department of Mathematics and Center for Atmosphere and Ocean Science, Courant Institute of Mathematical Sciences, New York University, NY 10012, USA

E-mail: jharlim@cims.nyu.edu

Received 11 July 2007, in final form 4 April 2008
Published 1 May 2008
Online at stacks.iop.org/Non/21/1281

Recommended by C Le Bris

**Abstract**

An important emerging scientific issue is the real time filtering through observations of noisy signals for nonlinear dynamical systems as well as the statistical accuracy of spatio-temporal discretizations for filtering such systems. From the practical standpoint, the demand for operationally practical filtering methods escalates as the model resolution is significantly increased. For example, in numerical weather forecasting the current generation of global circulation models with resolution of 35 km has a total of billions of state variables. Numerous ensemble based Kalman filters (Evensen 2003 *Ocean Dyn.* **53** 343–67; Bishop *et al* 2001 *Mon. Weather Rev.* **129** 420–36; Anderson 2001 *Mon. Weather Rev.* **129** 2884–903; Szunyogh *et al* 2005 *Tellus* A **57** 528–45; Hunt *et al* 2007 *Physica* D **230** 112–26) show promising results in addressing this issue; however, all these methods are very sensitive to model resolution, observation frequency, and the nature of the turbulent signals when a practical limited ensemble size (typically less than 100) is used. In this paper, we implement a radical filtering approach to a relatively low (40) dimensional toy model, the L-96 model (Lorenz 1996 *Proc. on Predictability (ECMWF, 4–8 September 1995)* pp 1–18) in various chaotic regimes in order to address the 'curse of ensemble size' for complex nonlinear systems. Practically, our approach has several desirable features such as extremely high computational efficiency, filter robustness towards variations of ensemble size (we found that the filter is reasonably stable even with a single realization) which makes it feasible for high dimensional problems, and it is independent of any tunable parameters such as the variance inflation coefficient in an ensemble Kalman filter.

This radical filtering strategy decouples the problem of filtering a spatially extended nonlinear deterministic system to filtering a Fourier diagonal system of parametrized linear stochastic differential equations (Majda and Grote 2007 *Proc. Natl Acad. Sci.* **104** 1124–9; Castronovo *et al* 2008 *J. Comput. Phys.* **227** 3678–714); for the linear stochastically forced partial differential equations

  

with constant coefficients such as in (Castronovo *et al* 2008 *J. Comput. Phys.* **227** 3678–714), this Fourier diagonal decoupling is a natural approach provided that the system noise is chosen to be independent in the Fourier space; for a nonlinear problem, however, there is a strong mixing and correlations between different Fourier modes. Our strategy is to radically assume for the purposes of filtering that different Fourier modes are uncorrelated. In particular, we introduce physical model errors by replacing the nonlinearity in the original model with a suitable Ornstein–Uhlenbeck process. We show that even with this 'poor-man's' stochastic model, when the appropriate parametrization strategy is guided by mathematical offline test criteria, it is able to produce reasonably skilful filtered solutions. In the highly turbulent regime with infrequent observation time, this approach is at least as good as trusting the observations while the ensemble Kalman filter implemented in a perfect model scenario diverges. Since these Fourier diagonal linear filters have large model error compared with the nonlinear dynamics, an essential part of the study below is the interplay between this error and the mathematical criteria for a given linear filter in order to produce skilful filtered solutions through the radical strategy.

Mathematics Subject Classification: 93E11, 62M20, 93E11, 62L12, 65C20, 74S25

(Some figures in this article are in colour only in the electronic version)

## 1. Introduction

Many contemporary problems in science ranging from protein folding in molecular dynamics to scale-up of small scale effects in nanotechnology to making accurate real time predictions of the coupled atmosphere–ocean system involve partial observations of extremely complicated systems with many degrees of freedom. Novel mathematical issues arise in the attempt to quantify the behaviour of such complex multi-scale systems [9, 10]. For example, in the coupled atmosphere–ocean system, the current practical models for prediction of both weather and climate involve general circulation models where the physical equations for these extremely complex flows are discretized in space and time and the effects of unresolved processes are parametrized according to various recipes; the result of this process involves a model for the prediction of weather and climate from partial observations of an extremely unstable, chaotic dynamical system with several billion degrees of freedom. These problems typically have many spatio-temporal scales, rough turbulent energy spectra in the solutions near the mesh scale and a very large dimensional state space yet real time predictions are needed.

When the system is low dimensional or when it has a low dimensional attractor, Monte Carlo approaches such as the particle filter [11] with its various up-to-date resampling strategies [12–14] provide better estimates in the presence of strong nonlinearity and highly non-Gaussian distribution. However, with the above practical computational constraint in mind, these accurate nonlinear particle filtering strategies are not feasible since sampling a high dimensional variable is computationally impossible for the foreseeable future. In the second direction, Bayesian hierarchical modelling [15] and reduced order filtering strategies [3, 5, 11, 16–23] based on the Kalman filter [24–26] have been developed with some success in these extremely complex high dimensional nonlinear systems. There is an inherently difficult practical issue of small ensemble size in filtering statistical solutions of these complex problems

due to the large computational overload in generating individual ensemble members through the forward dynamical operator [27].

In our recent work [8], we use the mathematical offline criteria introduced in [7] to understand the effect of the underlying dynamics, energy spectrum, time scale between observations relative to the correlation time and observation noise strength relative to the system noise strength towards the filtering problem where the true signal is a stochastically driven linear constant coefficient partial differential equation with a turbulent solution. In [8], the diagonal filter approach on each Fourier mode is natural since the linear stochastic PDE is completely decoupled into independent Langevin equations provided that spatially dependent white noise is chosen to be completely independent for each Fourier mode. In that paper, we study the effect of model errors due to the numerical temporal and spatial discretization.

In this paper, we implement a similar Fourier diagonal filter on a nonlinear dynamical system with varying degrees of nonlinearity in the true dynamics ranging from weakly chaotic to fully turbulent. This approach introduces physical model error on top of the model errors due to the numerical discretization scheme since the diagonal linear model completely ignores the nonlinear interactions in time that occur in the true dynamics and this is why we call it a radical filtering strategy; we replace the nonlinear terms in the original model by a stochastic noise and a linear damping term (these terms together are an Ornstein–Uhlenbeck process) on each Fourier mode in the hope that these additional terms regenerate the turbulent nature of the true signal. This stochastic modelling is a 'poor-man's' approach for modelling turbulent signals [28]. In other work of the second author and his collaborators, they show that when this 'poor-man's' strategy is adopted in a more elaborate turbulent climate modelling strategy for various deterministic dynamical systems [29–32], they obtained a reasonably good statistical estimate provided that the original signal is highly chaotic or fully turbulent. Our main motivation in this paper is to show that even with this 'poor-man's' model, we can reasonably obtain a skilful filtered solution beyond trusting the observations with appropriate parametrization guided by the mathematical offline test criteria. In the case where the ensemble filter fails, this radical strategy at least trusts the observations. We do not advocate this 'poor-man' strategy as the end product but we hope that the encouraging results shown in this paper serve as a motivation for filtering nonlinear systems in this non-standard fashion; in particular, we hope to convince the readers that ignoring correlations between different Fourier modes of nonlinearly mixed signals is a judicious radical approach and different reduced stochastic models should be tested for more filtering skill beyond the results in this paper.

On the other hand, this radical approach has many attractive features; it is very cheap computationally compared with the current generation of reduced order filtering strategies [1–5] since the filter model is analytically solvable and it is independent of the time step while the classical reduced order filtering such as the extended Kalman filter or the ensemble filter spends most of its computational time in propagating, respectively, the covariance matrix or the ensemble state forward in time. In [8], it was shown that this diagonal filter is very robust towards the variation of ensemble size, which makes this strategy feasible for filtering high dimensional systems such as in the operational numerical weather forecasting which can only afford an ensemble of less than 100 while the current generation of global circulation models has a total of $10^9$ degrees of freedom.

In our numerical experiments, we choose the L-96 model [6] as a test bed since it not only exhibits behaviour ranging from weakly chaotic to strongly turbulent as $F$ is varied [9, 33] and has an inhomogeneous (or anisotropic) spectrum with no power law cascade but it also has an absorbing ball property and its solutions are unimodally distributed and only weakly skewed from a Gaussian distribution [34]. In our numerical experiments, we compare our radical filtering strategy with the ensemble transform Kalman filter (ETKF) [2], an ensemble

Kalman filter which has been shown to be practically feasible in addressing high dimensional problems [4, 5]. The remainder of this paper is organized as follows: in section 2 we state the L-96 model and review a renormalized nondimensional formulation of the L-96 model that has zero mean state and unit energy perturbation [9, 10]. In section 3, we define the stochastic Fourier diagonal filter model, review the basic Kalman filter formulation and discuss several parametrization strategies including the physical based (by fitting the linear model to the climatological statistics or to the observation time correlation) and adaptive approaches (by choosing the filter coefficients from a parameter space). In section 4.1, we state our numerical experimental design. Before showing the filtering results, we quantify the observation time model errors corresponding to different parametrization strategies in section 4.2. In section 4.3, we present the general filter performance and we utilize the mathematical offline criteria and the diagnostic model error to understand the filter performances in various regimes. Finally, we conclude with a short summary in section 5.

## 2. The L-96 model

The L-96 model [6] represents an 'atmospheric variable' $u$ at $J$ equally spaced points around a circle of constant latitude. The $j$th component is propagated in time following differential equation

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

where $j = 0, \ldots, J-1$ 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 $\frac{\mathrm{d}}{\mathrm{d}t}E$) just like many atmospheric models (MW [10]). Following Lorenz [6], MAG [9] and MW [10, 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.

In this paper, we are interested in successfully filtering solutions of the L-96 model in regimes ranging from weakly chaotic $F = 5, 6$ to highly chaotic $F = 8$, to fully turbulent $F = 16$, given only noisy observations. Note that this 40-dimensional model has 9 positive Lyapunov exponents for a weakly chaotic regime with $F = 5$ and as many as 16 positive Lyapunov exponents for a fully turbulent regime with $F = 16$ [33]. For a fair comparison between these regimes, it is then convenient to adopt the rescaling introduced in MAG [9] and MW [10]:

$$u_j = \bar{u} + E_p^{1/2}\tilde{u}_j \qquad \text{and} \qquad t = E_p^{-1/2}\tilde{t}, \tag{2}$$

where $\bar{u}$ represents the (temporal) mean state and $E_p$ is the average variance in the energy fluctuation (see MAG [9] or MW [10] for details on how to compute these quantities). Note that this normalization is chosen such that the rescaled model has zero mean state and a unit energy perturbation.

Substituting (2) into (1) and after some algebra, we obtain the rescaled L-96 model

$$\frac{\mathrm{d}\tilde{u}_j}{\mathrm{d}\tilde{t}} = (\tilde{u}_{j+1} - \tilde{u}_{j-2})\tilde{u}_{j-1} + E_p^{-1/2}((\tilde{u}_{j+1} - \tilde{u}_{j-2})\bar{u} - \tilde{u}_j) + E_p^{-1}(F - \bar{u}). \tag{3}$$

The linear term and forcing can be written in Fourier space as

$$\frac{\mathrm{d}\hat{u}_k(\tilde{t})}{\mathrm{d}\tilde{t}} = A_k\hat{u}_k(\tilde{t}) + E_p^{-1}(F - \bar{u})\delta_{k,0}, \tag{4}$$

where $\delta_{k,0} = 0$ unless $k = 0$ where $\delta_{k,0} = 1$. By using the following discrete Fourier expansion:

$$\tilde{u}_j = \sum_{|k| \leqslant J/2} \hat{u}_k \mathrm{e}^{2\pi \mathrm{i} k j/J}$$

in the differential equation in (4), we have

$$A_k = E_p^{-1/2}[(\mathrm{e}^{2\pi \mathrm{i} k/J} - \mathrm{e}^{-4\pi \mathrm{i} k/J})\bar{u} - 1]. \tag{5}$$

## 3. Radical filtering strategies: Fourier diagonal filters

In MAG [9] and MW [10], a detailed linear stability analysis of (4) is developed. Here, we implement the Kalman filter on each separate Fourier mode of the L-96 model. In the introductory discussion, we mentioned that this radical filtering strategy introduces physical model errors whereas in GM [35], MG [7] and CHM [8], the model errors are due to the temporal and spatial numerical discretized approximation alone. To be more specific, we consider filtering noisy observed signal from (1) with the following stochastic differential equation

$$\mathrm{d}\hat{u}_k(\tilde{t}) = E_p^{-1}(F - \bar{u})\delta_{k,0}\,\mathrm{d}\tilde{t} + (A_k - d_k)\,\hat{u}_k(\tilde{t})\,\mathrm{d}\tilde{t} + \sigma_k\,\mathrm{d}\tilde{W}_k(\tilde{t}), \tag{6}$$

where in addition to the linear dynamics (4), we include a linear damping term with damping coefficient $d_k > 0$ and a stochastic white noise term with noise strength $\sigma_k$. In real space, this decoupled stochastic white noise term corresponds to the following correlated stochastic forcing

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

where $\mathrm{d}\tilde{W}_k(\tilde{t}) = 2^{-1/2}(\mathrm{d}\tilde{W}_{1,k}(\tilde{t}) + \mathrm{i}\,\mathrm{d}\tilde{W}_{2,k}(\tilde{t}))$ is a complex Gaussian white noise with each component $\tilde{W}_{j,k}(\tilde{t})$ is an independent Wiener process [36].

These two terms, the linear damping and white noise, are added to mimic the truncated nonlinear term in (3) with the hope that they generate a statistically reasonably approximation to the L-96 model (1). The additional damping term is important because it neutralizes the additional energy from the white noise and hence the two terms together (which is a well-known Ornstein–Uhlenbeck process) statistically conserve the total energy [28], as the replaced discrete advective nonlinear term in the full L-96 model (1). This turbulent approximation is a 'poor-man's' strategy with no mathematical rigorous justification; however, numerical implementations of this approximation in more advanced turbulent climate modelling [29–32] produce reasonably accurate statistical estimates of the original model provided that the original signal from the full nonlinear model is strongly chaotic or fully turbulent.

The solution of equation (6),

$$\hat{u}_k(\tilde{t}) = E_p^{-1}(F - \bar{u})\delta_{k,0} + \hat{u}_k(0)\mathrm{e}^{(A_k - d_k)\tilde{t}} + \sigma_k \int_0^{\tilde{t}} \mathrm{e}^{(A_k - d_k)(\tilde{t} - s)}\,\mathrm{d}\tilde{W}_k(s), \tag{7}$$

is a Gaussian random variable with mean and variance given by

$$\bar{\hat{u}}_k(\tilde{t}) = \langle \hat{u}_k(\tilde{t}) \rangle = E_p^{-1}(F - \bar{u})\delta_{k,0} + \langle \hat{u}_k(0) \rangle \mathrm{e}^{(A_k - d_k)\tilde{t}}, \tag{8}$$

$$\begin{aligned} r_k &= \langle (\hat{u}_k(\tilde{t}) - \bar{\hat{u}}_k(\tilde{t}))(\hat{u}_k(\tilde{t}) - \bar{\hat{u}}_k(\tilde{t}))^* \rangle, \\ &= \frac{\sigma_k^2}{2\,\mathrm{Re}\{d_k - A_k\}}(1 - \mathrm{e}^{-2\,\mathrm{Re}\{d_k - A_k\}\tilde{t}}). \end{aligned} \tag{9}$$

In the discrete formulation, given an observation time $T_{\rm obs} = t_{m+1} - t_m$, we can write our filtering problem for each Fourier mode as

$$\hat{u}_{k,m+1|m} = E_p^{-1}(F - \bar{u})\delta_{k,0} + F_k \hat{u}_{k,m|m} + \eta_{k,m+1}, \tag{10}$$
$$\hat{v}_{k,m} = g\hat{u}_{k,m} + \hat{\sigma}_m^o,$$

where $F_k = \mathrm{e}^{(A_k - d_k)T_{\rm obs}}$ is the deterministic dynamical operator and $\eta_{k,m} \sim \mathcal{N}(0, r_k)$ is a random Gaussian white noise with variance $r_k$ as in equation (9) with $T_{\rm obs}$ replaces $\tilde{t}$. In this paper, we assume that observations $\hat{v}_{k,m}$ are available at every model grid point such that $g = 1$ with observation errors reflected by a Gaussian distribution, i.e. $\hat{\sigma}_m^o \sim \mathcal{N}(0, \hat{r}^o)$. In the real domain, this variance corresponds to $r^o = J\hat{r}^o$.

The classical Kalman filter consists of two steps: it propagates the best estimate of the true state $\hat{u}_{k,m}$ and the error covariance with

$$\bar{\hat{u}}_{k,m+1|m} = E_p^{-1}(F - \bar{u})\delta_{k,0} + F_k \bar{\hat{u}}_{k,m|m}, \tag{11}$$
$$\begin{aligned} r_{k,m+1|m} &= \langle (\hat{u}_{k,m+1} - \bar{\hat{u}}_{k,m+1|m})(\hat{u}_{k,m+1} - \bar{\hat{u}}_{k,m+1|m})^* \rangle \\ &= F_k \langle (\hat{u}_{k,m} - \bar{\hat{u}}_{k,m|m})(\hat{u}_{k,m} - \bar{\hat{u}}_{k,m|m})^* \rangle F_k^* + r_k \\ &= F_k r_{k,m|m} F_k^* + r_k, \end{aligned} \tag{12}$$

where the mean prior state $\bar{\hat{u}}_{k,m+1|m}$ denotes the best estimate before we include observations at time $t_{m+1}$ and $r_{k,m+1|m}$ denotes its corresponding error covariance while the mean posterior state $\bar{\hat{u}}_{k,m|m}$ denotes the state after we include observations at time $t_m$ and $r_{k,m|m}$ denotes the error posterior covariance. Subsequently, it includes the observations through the following update:

$$\bar{\hat{u}}_{k,m+1|m+1} = \bar{\hat{u}}_{k,m+1|m} + K_{k,m+1}(\hat{v}_{k,m+1} - g\bar{\hat{u}}_{k,m+1|m}), \tag{13}$$

where

$$K_{k,m+1} = \frac{g r_{k,m+1|m}}{r^o + g^2 r_{k,m+1|m}}, \tag{14}$$
$$r_{k,m+1,m+1} = (1 - K_{k,m+1}g) r_{k,m+1,m}, \tag{15}$$

are the scalar Kalman gain and the scalar posterior error covariance, consecutively. The analytical asymptotical solution for equations (12), (14), (15) exists since the filtering problem is fully observable $g = 1$ and hence there exists an asymptotic Kalman gain $K_{k,\infty}$ and asymptotic covariance $r_{k,m+1,m+1} \to r_{k,\infty}$ as $m \to \infty$. Interested readers should consult [7, 8] for explicit expressions of these asymptotic quantities. For the best estimate of the true state $\hat{u}_{k,m}$, we simply solve the recursive formula for the mean with (11) and (13). In the numerical experiments below, however, we compare this approach with an ensemble Kalman filter and thus for fair comparison, we purposely generate an ensemble $\{\hat{u}^j_{k,m|m},\ j = 1, \ldots, K\}$ of size $K$ where each ensemble member is generated via (10) and (13) with $\bar{\hat{u}}$ replaced by $\hat{u}^j$ in the latter equation. Since equations (10) and the modified (13) are linear in $\hat{u}^j$, for infinite ensemble size ($K \to \infty$), they will converge to their corresponding theoretical averaged in equations (11) and (13). Thus, the simulations with finite ensemble $K$ introduce sampling errors since the ensemble average $\sum_j \hat{u}^j_k/K$ is not necessarily equal to the theoretical mean $\bar{\hat{u}}_k = \langle \hat{u}_k \rangle$.

With this set-up we commit a considerably 'serious' modelling error in our filtering strategy since the true signal $\hat{u}_{k,m}$ is generated from (1) while the filter model is a linear stochastic model (6). However, we will show that with this approach, we can still obtain skilful results for appropriate parameters $d_k$, $\sigma_k$, $r^o$ and $T_{\rm obs}$.

### 3.1. Climatological stochastic model (CSM)

The simplest strategy to determine the filter coefficients is to fit the linear damping coefficient $d_k$ and the system noise strength $\sigma_k$ to the climatological (or equilibrium) variance

$$E_\infty = \lim_{\tilde{t}\to\infty} r_k = \frac{\sigma_k^2}{2\gamma_k}, \tag{16}$$

where $\gamma_k = d_k - \mathrm{Re}\{A_k\}$ is the effective dampling coefficient and fit the correlation time

$$\begin{aligned} T_{\mathrm{corr}} &= \int_0^\infty \mathrm{Re}\{\mathrm{Corr}(\tau)\}\,\mathrm{d}\tau \\ &= \int_0^\infty \mathrm{Re}\left\{\frac{\langle(\hat{u}_k(t)-\bar{\hat{u}}_k)(\hat{u}_k(\tilde{t}+\tau)-\bar{\hat{u}}_k)^*\rangle}{r_k}\right\}\mathrm{d}\tau \\ &= \int_0^\infty \mathrm{e}^{-\gamma_k\tau}\cos(\omega_k\tau)\,\mathrm{d}\tau \\ &= \frac{\gamma_k}{\omega_k^2+\gamma_k^2}, \end{aligned} \tag{17}$$

where $\omega_k = \mathrm{Im}\{A_k\}$. Here, both the climatological variance ($E_\infty$) and the correlation time ($T_{\mathrm{corr}}$) for each Fourier wave number are estimated from a long trajectory solution of (1) (see MAG [9] and MW [10]). This is the standard strategy of linear stochastic modelling for a turbulence [28], applied to the L-96 model.

### 3.2. General diagonal AR(1) model

Consider an arbitrary time series $\hat{u}_{k,m} \in \mathbb{C}, m = 1,\ldots,T$, where each time step $t_{m+1} - t_m$ corresponds to the observation time $T_{\mathrm{obs}}$. Suppose that we want to model this time series with

$$\hat{u}_{k,m+1} = \mathrm{const} + F_k\hat{u}_{k,m} + \eta_{k,m}, \tag{18}$$

where $F_k = \alpha_k + \beta_k\mathrm{i}$ and white noise $\eta_{k,m} \sim \mathcal{N}(0, r_k)$. Then we have to choose $\alpha_k$, $\beta_k$ and $r_k$ so that the model is statistically relevant to the time series.

In the remainder of this section, we describe a parametrization strategy based on observation time correlation and several adaptive strategies for parametrizing the damping term $\gamma_k$ and system noise strength $\sigma_k$. Our motivation for varying parameters $\gamma_k$ or $\sigma_k$ is to guarantee that the system becomes more controllable [8] as well as simultaneously having small model errors. Recall that controllability for each scalar linear stochastic model (6) implies $r_k \neq 0$ which then requires both $|F_k| \neq 1$ and $\sigma_k \neq 0$.

*3.2.1. Observation time correlation and climatology (OTCC).* With the form in (18), the variance of $\hat{u}_{k,m}$ evolves as follows:

$$\langle|\hat{u}_{k,m+1} - \bar{\hat{u}}_{k,m+1}|^2\rangle = F_k\langle|\hat{u}_{k,m} - \bar{\hat{u}}_{k,m}|^2\rangle F_k^* + r_k, \tag{19}$$

and it converges to a stationary variance $E_\infty = \lim_{m\to\infty}\langle|\hat{u}_{k,m+1} - \bar{\hat{u}}_{k,m+1}|^2\rangle$. Taking $m \to \infty$, we can rewrite (19) as

$$E_\infty = \frac{r_k}{1-|F_k|^2}. \tag{20}$$

A standard first-order autoregressive model simply matches the time correlation at lag $T_{\mathrm{obs}}$. From the difference between deviations (from the mean, i.e. $x' = x - \bar{x}$) at two different times, we have

$$\hat{u}'_{k,m+1} - \hat{u}'_{k,m} = \hat{u}_{k,m+1} - \bar{\hat{u}}_{k,m+1} - (\hat{u}_{k,m} - \bar{\hat{u}}_{k,m}) = (F_k - 1)(\hat{u}_{k,m} - \bar{\hat{u}}_{k,m}) + \eta_{k,m}. \tag{21}$$

Taking the expectation of the square of (21) yields

$$\begin{aligned}
&\langle|\hat{u}'_{k,m+1} - \hat{u}'_{k,m}|^2\rangle = |F_k - 1|^2\langle|\hat{u}_{k,m} - \bar{\hat{u}}_{k,m}|^2\rangle + r_k \\
&2E_\infty - 2\,\mathrm{Re}\{\langle(\hat{u}_{k,m} - \bar{\hat{u}}_{k,m})(\hat{u}_{k,m+1} - \bar{\hat{u}}_{k,m+1})^*\rangle\} = |F_k - 1|^2 E_\infty + r_k \qquad (22)\\
&2E_\infty - 2\,\mathrm{Re}\{(\hat{u}_{k,m} - \bar{\hat{u}}_{k,m})(\hat{u}_{k,m+1} - \bar{\hat{u}}_{k,m+1})^*\rangle\} = (2 - 2\alpha_k)E_\infty.
\end{aligned}$$

The second line in (22) sets the covariance to be fixed at the climatology value and the third line is obtained by substituting (20) and the following identity:

$$|1 - F_k|^2 = 1 + |F_k|^2 - 2\alpha_k.$$

We can rewrite (22) as

$$\alpha_k = \frac{\mathrm{Re}\{(\hat{u}_{k,m} - \bar{\hat{u}}_{k,m})(\hat{u}_{k,m+1} - \bar{\hat{u}}_{k,m+1})^*\rangle\}}{E_\infty} \qquad (23)$$

where the numerator is basically the cross correlation at lag $T_{\mathrm{obs}}$.

In our stochastic model, the discrete operator $F_k$ is the deterministic exponential solution of the linear differential equation (6), i.e. $F_k = \mathrm{e}^{-(\gamma_k + \mathrm{i}\omega_k)T_{\mathrm{obs}}}$, where $\gamma_k = d_k - \mathrm{Re}\{A_k\}$ and $\omega_k = \mathrm{Im}\{A_k\}$. Hence, given $\alpha_k$ (from (23)) and $\omega_k$ (from (5)), we can find the corresponding total damping term $\gamma_k$ by solving

$$\alpha_k = \mathrm{e}^{-\gamma_k T_{\mathrm{obs}}}\cos(\omega_k T_{\mathrm{obs}}), \qquad (24)$$

and thus $\beta_k$ can be easily obtained. The discrete time system noise variance $r_k$ is then obtained by solving (20) which is effectively equivalent to solving (9).

*3.2.2. Climate noise varying damping (CNVD) model.* Above, we describe two strategies for stochastic modelling with physical reasoning, i.e. based either on the climatology or on the observation time correlation. The climate noise varying damping (CNVD) model is implemented by fixing the phase $\omega_k$ to be exactly the linear phase or the imaginary part of (5) and fixing the system noise variance $r_k$ as solution of (20) with climatological variance $E_\infty$. The damping term $\gamma_k$ is chosen so that $|F_k| < 1$ and so that the RMS difference between the filtered solution and the true signal is minimum for each wave number. For the rest of this paper, we will refer to this RMS difference as error and we will describe in section 4 on how to obtain this RMS difference or error. In the optimization strategy, we fix $|F_k| = 0.01, 0.1, 0.2, \ldots, 0.9, 0.95$ and pick the one that produces the lowest error.

*3.2.3. Climate damping varying noise (CDVN) model.* The CDVN model uses the CSM damping term, that is, $\gamma_k$ is chosen by solving (17) with the same linear phase $\omega_k$ as used in our previous strategies. The only difference from the CSM is that this approach uses the following system noise variance

$$r_k = c_k E_\infty(1 - |F_k|^2)$$

with $c_k > 1$ chosen such that the filtered solution yields the lowest errors.

*3.2.4. Observation time correlation damping varying noise model (OTCDVN).* The OTCDVN model is similar to CDVN except that it chooses the damping term $\gamma_k$ by solving (23) and (24) with linear phase $\omega_k$.

## 4. Performance on L-96 test bed

In this section, we begin by describing the experimental design set-up. We then discuss and quantify the 'observation time model errors' of each method following Tribbia and Baumhefner [37]. Subsequently, we show results from numerical experiments of implementing the filtering strategies discussed earlier and we use the offline test criteria [8] to understand the performance of various methods. As in [8], we check the robustness of our results by adding a second source of model error that involves implicit numerical scheme. Finally, we show the consistency of the Fourier diagonal filter with different parameters and its computational savings.

### *4.1. Experimental design*

In each numerical simulation, a true trajectory $u^t$ is generated by integrating the L-96 model (1) with the Runge–Kutta method with time step $\Delta t = 1/64$. The observations are simulated by adding uncorrelated Gaussian noise at every model grid point with variance $r^o$ to the true trajectories at observation time $T_{\text{obs}} = n\Delta t$. In our experiments, we check three observation times with $n = 2, 5, 15$. When $F = 8$, Lorenz suggests that 0.05 non-dimensionalized units are equivalent to 6 h based on doubling time in a global weather model [6]. Thus, our choices of $n$'s correspond to roughly 4 h, 9 h and 28 h, respectively.

For our numerical experiments, we vary the observation noise variances $r^o$ and compare the filter performance when the variance is smaller or larger relative to the square of the size of the chaotic attractor that can be roughly estimated by taking a root-mean-square (RMS) average difference between two long trajectories initiated from two almost identical model states. In our numerical results, we call this average difference the errors due to 'no filter' and we use it for a benchmark of the filter performance in addition to the observation error $\sqrt{r^o}$.

We define an RMS average error or difference between $u(x_j, t_i)$ and $v(x_j, t_i)$ as follows:

$$\text{rms}(u, v) = \sqrt{\frac{1}{J(T - T_o)} \sum_{i=T_o+1,\, j=1}^{T,J} (u(x_j, t_i) - v(x_j, t_i))^2}, \tag{25}$$

where $T_o = 100$ and $T = 500$ and we measure the filter performance with the RMS difference between the ensemble average posterior state and the true signal, that is $\text{rms}(\bar{u}_{m|m}, u^t)$, where $\bar{u} = \sum_{j=1}^{K} u^j_{m|m}/K$ is an ensemble average for the remainder of this paper. Thus, the no filter RMS error is nothing else but $\text{rms}(u, u + \epsilon)$, where $\epsilon$ is an arbitrary random perturbation. When details are necessary, we also compute the RMS average only spatially and present this quantity as a function of time.

We also compare these radical filtering strategies with the ensemble transform Kalman filter (ETKF) of Bishop *et al* [2], implemented with symmetric square root ensemble perturbations [23, 38, 39], with ensemble size $K = J = 40$, and variance inflation coefficient $r = 0.05$. The main reason why we choose ETKF is that it is easily expandable for high dimensional systems [4, 5] and it produces a comparable accuracy to the extended Kalman filter [22] for L-96 model when $K \geqslant J$.

### *4.2. Observation time model error*

Let $u_i$, $i = 1, \ldots, T = 500$ be a trajectory solution of the L-96 model (1), sampled discretely at every observation time $t_{i+1} - t_i = T_{\text{obs}}$. At each observation time interval, we integrate differential equation (6) with initial condition $u_i$ and time step $T_{\text{obs}}$ to obtain $\tilde{u}_{i+T_{\text{obs}}}$. We define the observation time model error $\text{rms}(u, \tilde{u})$ as the RMS average difference between solutions

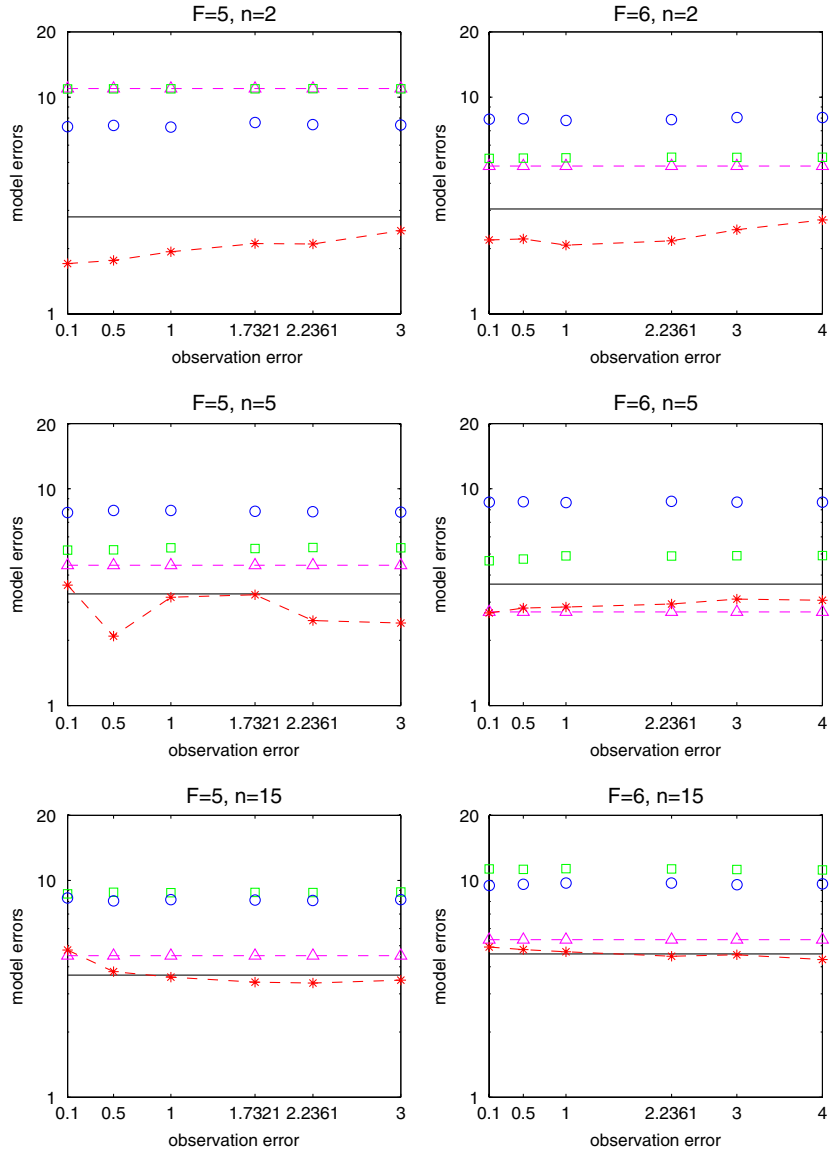


**Figure 1.** Observation time model errors as functions of observation errors ($\sqrt{r^o}$): the first column is for $F = 5$, the second column is for $F = 6$. In each column, results for $n = 2, 5$ and 15 are shown. In each plot, CSM by black solid, OTCC by magenta with triangle, CNVD by red dashes with '$*$', OTCDVN with green box and CDVN by blue circle.

of L-96 model (1) and solutions of the stochastic differential equation (6) after $T_{\text{obs}}$ time. In earlier discussions (in sections 1 and 2) we referred to this quantity as (physical) model error so for the rest of the paper, we shall be consistent and understand it as the model error after $T_{\text{obs}}$ time.

In figures 1 and 2, we show the model errors for various approaches as functions of observation errors $\sqrt{r^o}$. The CSM (section 3.1) and OTCC (section 3.2.1) are independent of observation noise; however, the other three strategies, CNVD (section 3.2.2), CDVN

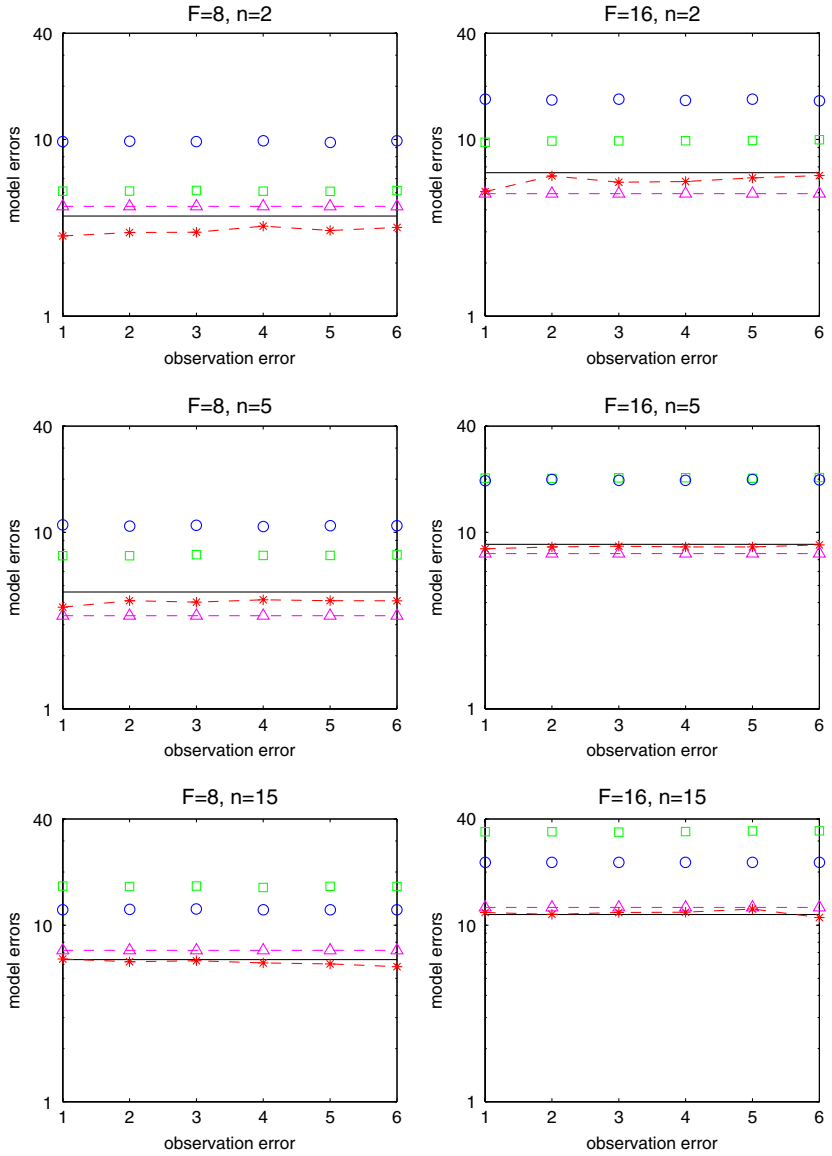


**Figure 2.** Observation time model errors as functions of observation errors ($\sqrt{r^o}$): the first column is for $F = 8$, the second column is for $F = 16$. In each column, results for $n = 2, 5$ and 15 are shown. In each plot, CSM by black solid, OTCC by magenta with triangle, CNVD by red dashes with '$*$', OTCDVN with green box and CDVN by blue circle.

(section 3.2.3), and OTCDVN (section 3.2.4), depend on the observation noise since each choice of noise or damping coefficients in these filters is optimized differently for different observation noises and observation times.

From figures 1 and 2, we see that the model errors of CSM increase as functions of both time and external forcing $F$. The model error increment as a function of time is well understood since a linear approximation of a nonlinear function typically holds only for a sufficiently short time. The model error increment as a function of $F$ is also obvious since the dynamical system

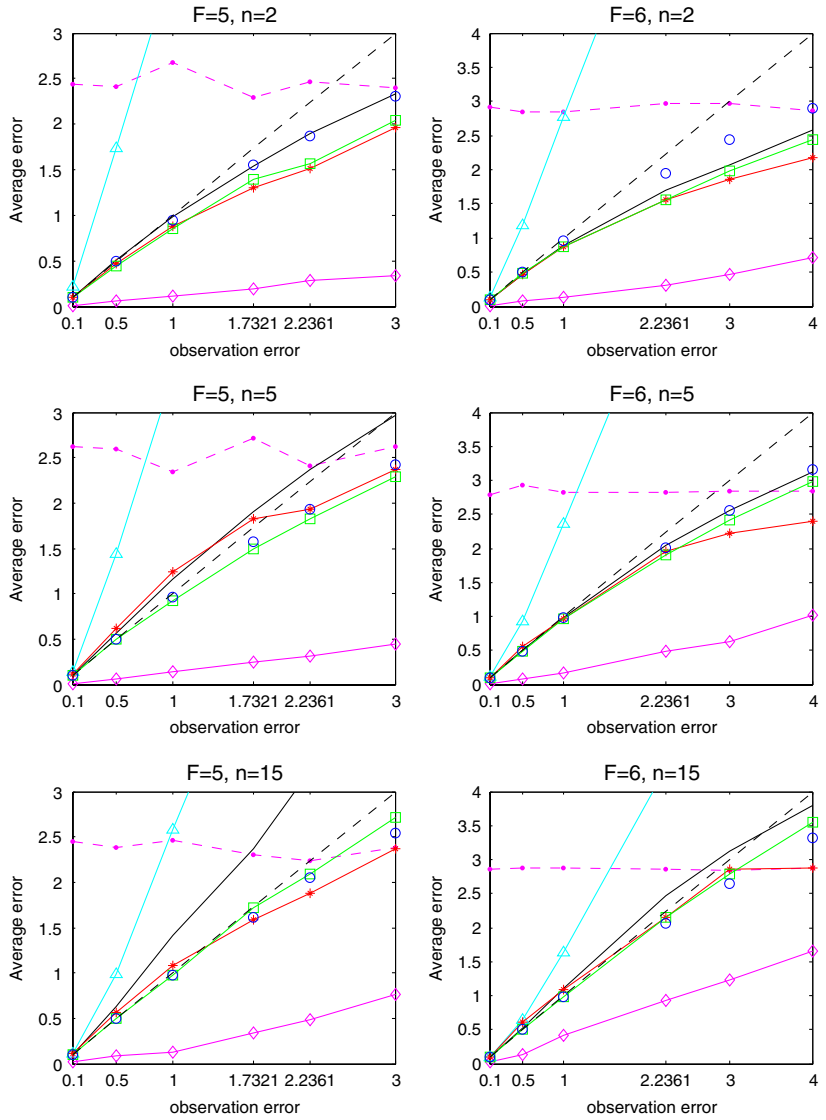


**Figure 3.** RMS errors as functions of observation errors ($\sqrt{r^o}$): the first column is for $F = 5$, the second column is for $F = 6$. In each column, results for $n = 2, 5$ and 15 are shown. In each plot, ETKF is denoted by magenta with diamond, no filter by dashed–dotted magenta, observation errors by black dashes, CSM by black solid, OTCC by light blue with triangle, CNVD by red with '$*$', OTCDVN with green box and CDVN by blue circle.

becomes more chaotic and the number of positive Lyapunov exponents increases as a function of $F$ [33]. Model errors of OTCC (magenta dashes with triangle) for short time vary; in the weakly chaotic system $F = 5, 6$ they are large while in the more turbulent systems, the model error is less than that of CSM. As observation time increases, the model error of OTCC becomes comparable and sometimes larger than that of CSM. We also find that in almost every regime CNVD (red dashes with '$*$') has the least model error. The two other adaptive strategies CDVN (blue circle) and OTCDVN (green box) have the largest model errors which

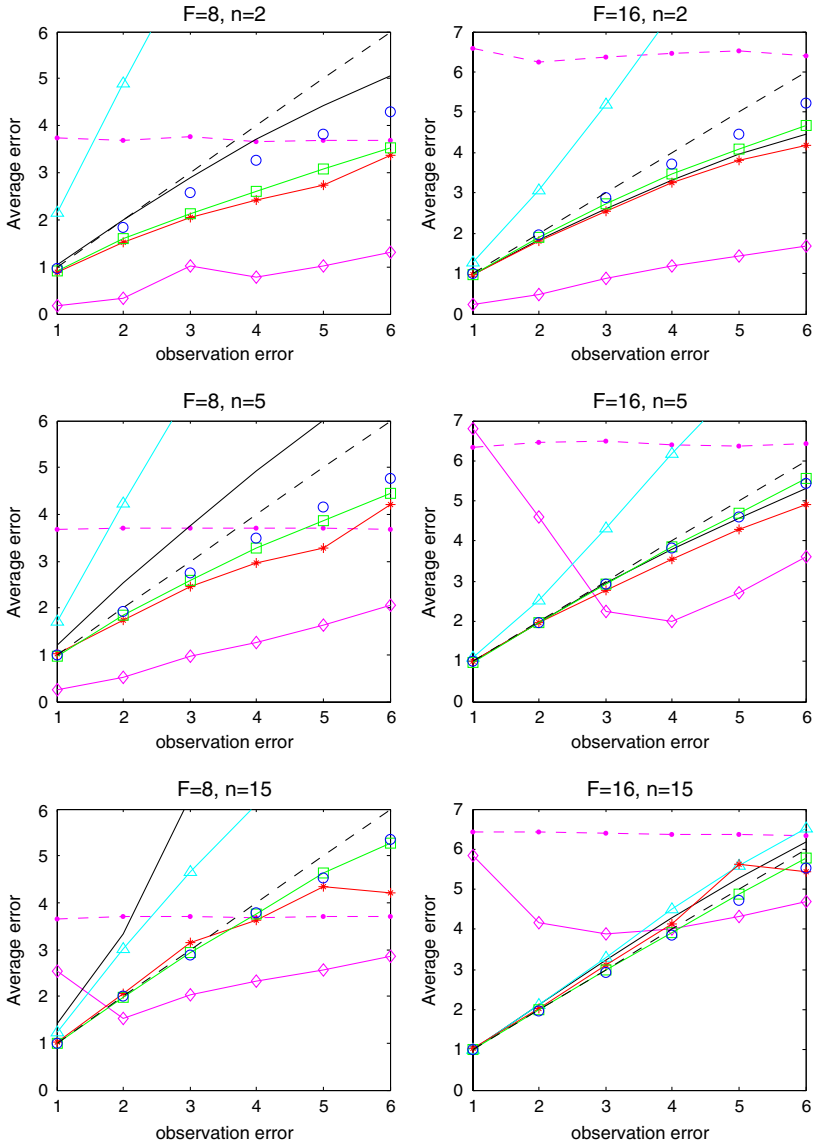


**Figure 4.** RMS errors as functions of observation errors ($\sqrt{r^o}$): the first column is for $F = 8$, the second column is for $F = 16$. In each column, results for $n = 2$, 5 and 15 are shown. In each plot, ETKF is denoted by solid magenta with diamond, no filter by dashed–dotted magenta, observation errors by black dashes, CSM by black solid, OTCC by light blue with triangle, CNVD by red with '$*$', OTCDVN with green box and CDVN by blue circle.

is mostly contributed by the stochastic term of the model when the system noise variance is large.

### *4.3. Filter performance*

In figures 3 and 4, we show the RMS average errors for all filtering strategies discussed in section 3 and that of the ensemble transform Kalman filter (ETKF). From this error

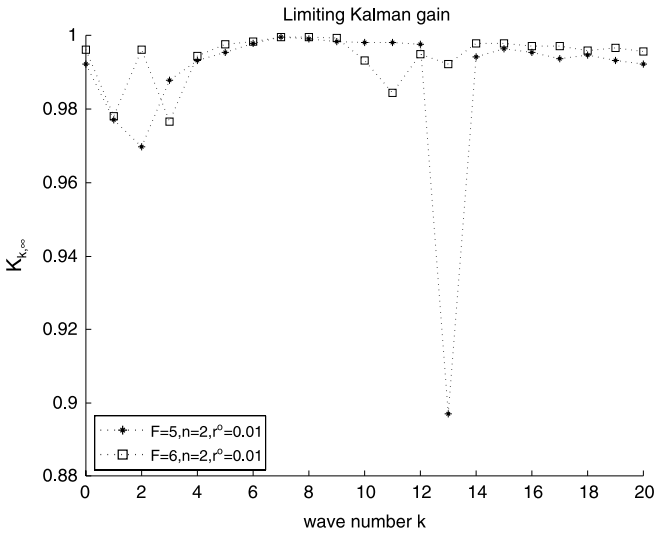


**Figure 5.** Limiting Kalman gain of CSM as a function of wave numbers for short time $n = 2$, small observation noise variance $r^o = 0.01$ and weakly chaotic system $F = 5$ and 6.

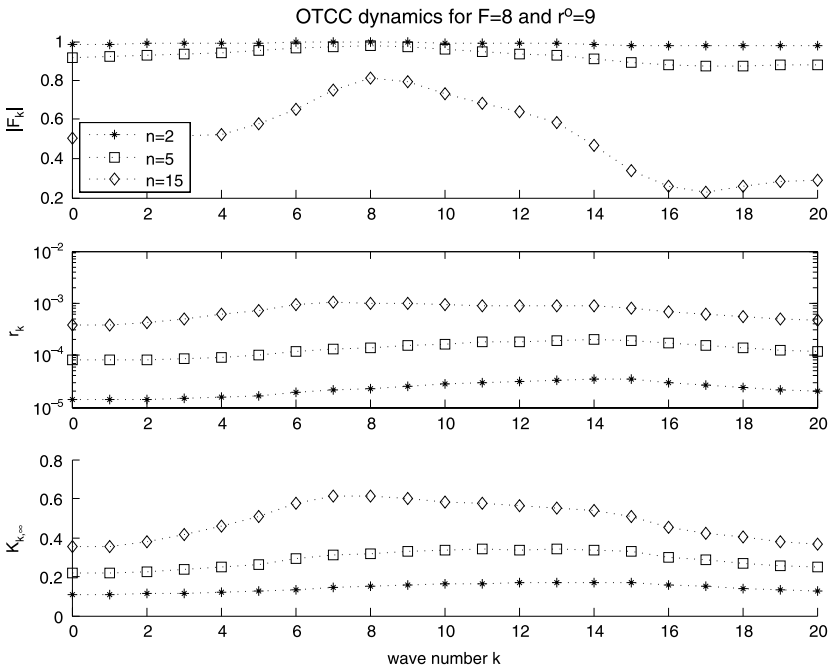


**Figure 6.** Amplitude of $F_k$, system noise variance $r_k$ and limiting Kalman gain $K_{k,\infty}$ of OTCC as functions of wave numbers for regime $F = 8$, $r^o = 9$ and three different observation times corresponding to $n = 2$, 5 and 15.

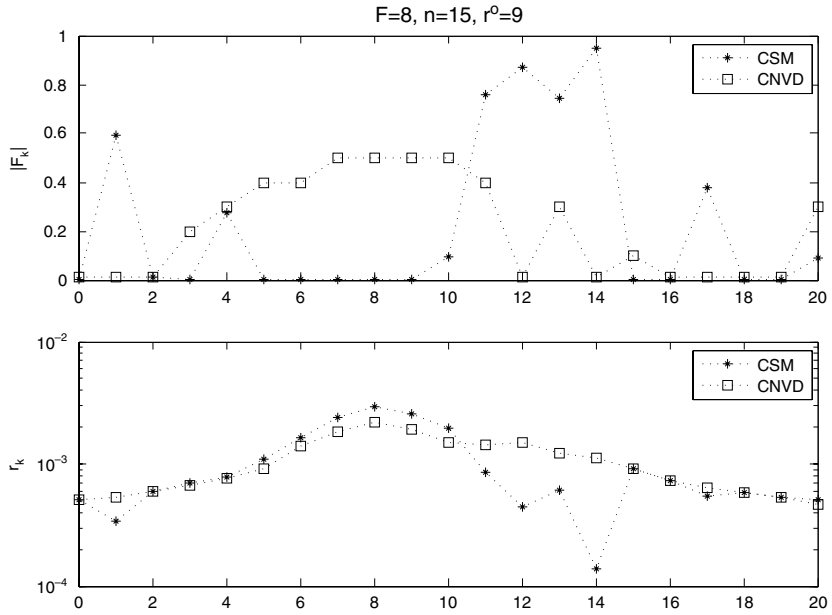


**Figure 7.** Amplitude of $F_k$ and system noise variance $r_k$ of regime $F = 8, n = 5, r^o = 9$ for CSM ('$*$') and CNVD (in square).

measure (25), the overall performance for various regimes in the Fourier diagonal filtering strategies from the best to the worst rank as follows: CNVD, OTCDVN, CDVN, CSM and OTCC. In almost every regime, the first three strategies always produce better filtered solutions compared with trusting observations especially when the observation error is large. In the remainder of this section, we will use the offline test criteria [8] and model error estimates (of section 4.2) to understand the performance of our proposed strategies in various regimes.

The CSM filter (see section 3.1) always produces filtered solutions that are better than or at least comparable to simply trusting the observations for short observation time with $n = 2$ (see the first row of figures 3 and 4). When the observation error is small, the filter weights more towards the observations. To demonstrate this fact, in figure 5 we present the limiting Kalman gain $K_{k,\infty}$, which is close to 1 for almost every wave number for regimes $F = 5$ and 6, $n = 2$, $r^o = 0.01$. Note that the limiting Kalman gain is a scalar quantity bounded by $0 \leqslant K_{k,\infty} \leqslant 1$ with properties that $K_{k,\infty} = 0$ when the filter fully trusts the dynamics and $K_{k,\infty} = 1$ when the filter fully trusts the observations [8]. As we increase the observation error variance $r^o$, the filter works remarkably well for short observation times since the filter weights partially to the dynamics and the model errors are not yet significant. As we increase the observation time, the model errors become significant (see figures 1 and 2) and hence the filter becomes increasingly inaccurate. We also notice that the filter tolerance towards the size of model error increases as a function of external forcing $F$ since the size of the chaotic attractor increases as a function of $F$ as well [33]. For example when $F = 5$, an RMS model error $\gtrapprox 3$ yields an inaccurate filtered solution when $n = 5$ whereas the filter with RMS model error $\approx 4$ for $F = 6$ is still accurate (see figure 1).

The OTCC filter (see section 3.2.1) simply does not work in every regime we test. For short time, the inaccurate filtering is due to the violation of controllability. As mentioned at

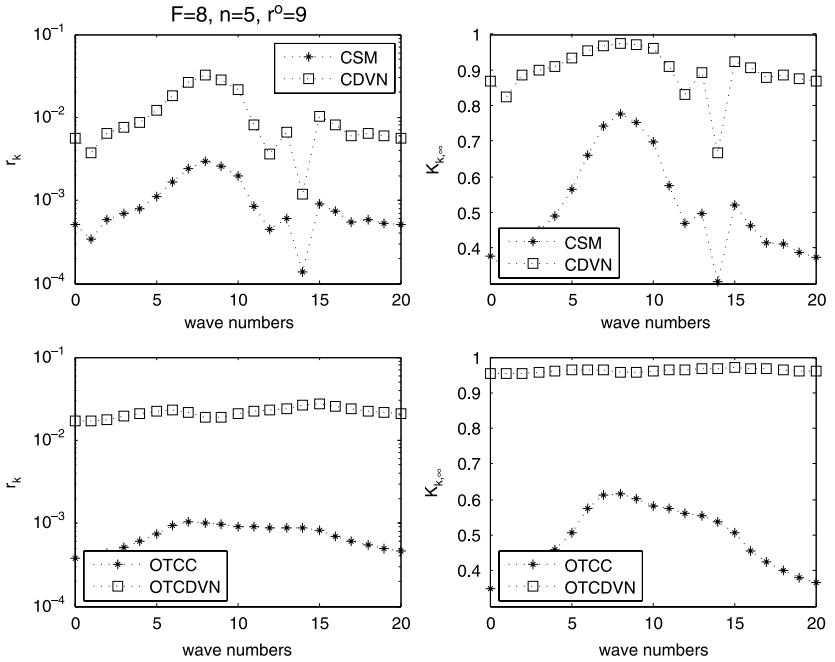


**Figure 8.** System noise variance $r_k$ (first column) and limiting Kalman gain (second column) as functions of wave numbers for regime $F = 8, n = 5, r^o = 9$. The top figures show the noise and Kalman gain inflations for CSM while the bottom figures for OTCC.

the beginning of section 3.2, for a scalar filter, the violation of controllability is characterized by a nearly unit amplitude of the filter dynamical operator $|F_k| \approx 1$ and small stochastic noise variance $r_k$. In figure 6, we present an example of this phenomenon for regime $F = 8, n = 2, 5, 15$, and $r^o = 9$ where $r_k \approx 10^{-5}$ when $n = 2$. For longer time steps, the filtered solution RMS error is smaller since the filter is more controllable ($|F_k| < 0.8$ and $r_k \approx 10^{-3}$ when $n = 15$, again see figure 6), however the presence of model errors is too significant (see figure 2 for this regime) so that the dynamics are not trustable while the limiting Kalman gain for almost every wave number is $K_{k,\infty} < 0.5$, which suggests that the filter weights more toward the linear dynamics with large model error.

The CNVD filter (see section 3.2.2) performs the best overall since its model error is the smallest among all methods we present (see figures 1 and 2). Secondly by allowing the deterministic part of the model (that is, the damping term) to be chosen among the available constraints $|F_k| < 1$, this scheme avoids the violation of controllability. For example, in regime $F = 8, n = 15, r^o = 9$, where the CSM filter does not work, the CNVD readjusts the damping term such that they are not overdamped nor too weak while retaining comparable noise (see figure 7).

Whenever CSM (or OTCC) produces inaccurate filtered solutions, the noise inflation in CDVN (see section 3.2.3) or OTCDVN (section 3.2.4), respectively, improves the filtering controllability significantly. In figure 8, we show the system noise variance $r_k$ and the limiting Kalman gain $K_{k,\infty}$ for the case $F = 8, n = 15, r^o = 9$ for these four methods and we see that the noise inflation, which also increases the model errors significantly (see figure 2), forces the filter to weight more towards the observations. In other words, the deterministic part of the model is so different from the true dynamics that a higher uncertainty (reflected by a higher

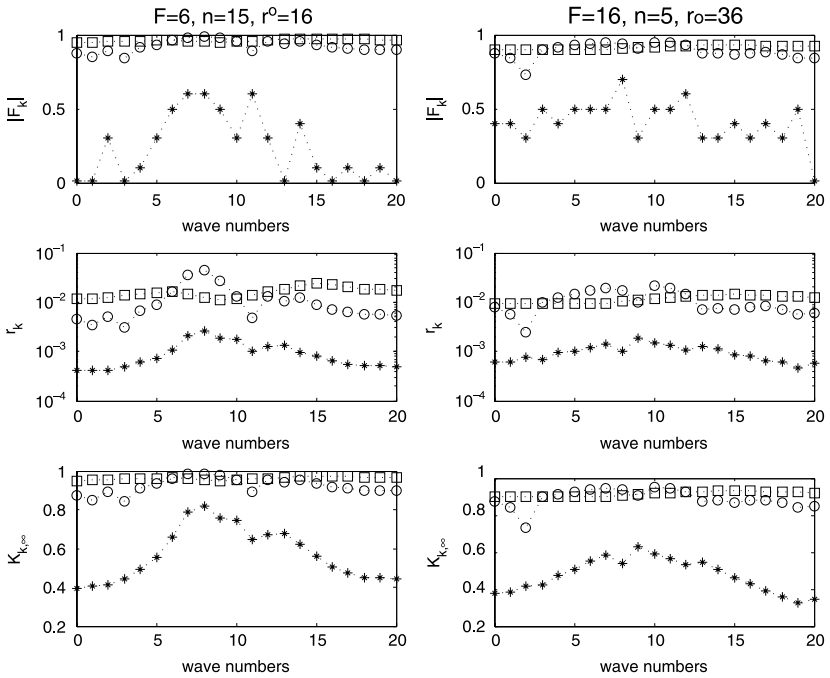


**Figure 9.** In each column, we plot the amplitude of $F_k$, the system noise variance $r_k$ (first column) and the limiting Kalman gain $K_{k,\infty}$ as functions of wave numbers of CNVD ('$*$'), OTCDVN (square), and CDVN (circle). The first column for regime $F = 6, n = 15, r^o = 16$ and the second column for $F = 16, n = 5, r^o = 36$.

system noise variance $r^o$) is required in order to push the prior state to be closer to the unknown true state.

In certain regimes such as $F = 6, n = 15, r^o = 16$ or $F = 16, n = 5, r^o = 36$, CNVD is slightly better than CDVN and OTCDVN since in the last two cases, the filters weight almost fully towards the observations (see figure 9) but in these regimes the observation errors are relatively large compared with the errors of no filter. On the other hand, the CNVD filter weights more towards the dynamics that has the smallest model error (see figures 1 and 2).

For shorter time and the weakly chaotic system, the best Fourier diagonal filter performance has skill, but the filter error is significantly larger than those for ETKF (see RMS errors as functions of time and snapshots of filtered solutions at the end of the assimilation time for regime $F = 6, n = 2, r^o = 5$ in figures 10 and 11). Here, the model errors are not too large whereas the linear analysis suggests that the filter satisfies controllability, hence the filter skill is not so surprising. For large observation time, the Fourier diagonal filtering strategies supersede ETKF when filtering strongly chaotic and turbulent signals (see figures 12 and 13). Here, we start to see the flaws of the Gaussian assumption forced by the ensemble Kalman filter. From the dynamical system point of view, with such a long observation time the prior state $\hat{u}_{k,m|m}$ is close to (or even in) the chaotic attractor set of (1). In this case, the nonlinearity of the chaotic evolution operator and the underestimation of the uncertainties in the background forecasts due to small ensemble size (for all experiments, we used 40 ensemble member) deteriorate the filter performance. The sensitivity of ETKF is even more revealing for $F = 16$ and $n = 5, 15$ (see second column of figure 4). There, all of the Fourier diagonal filtering strategies steadily trust the observations while the ETKF fails when $r^o$ is small.

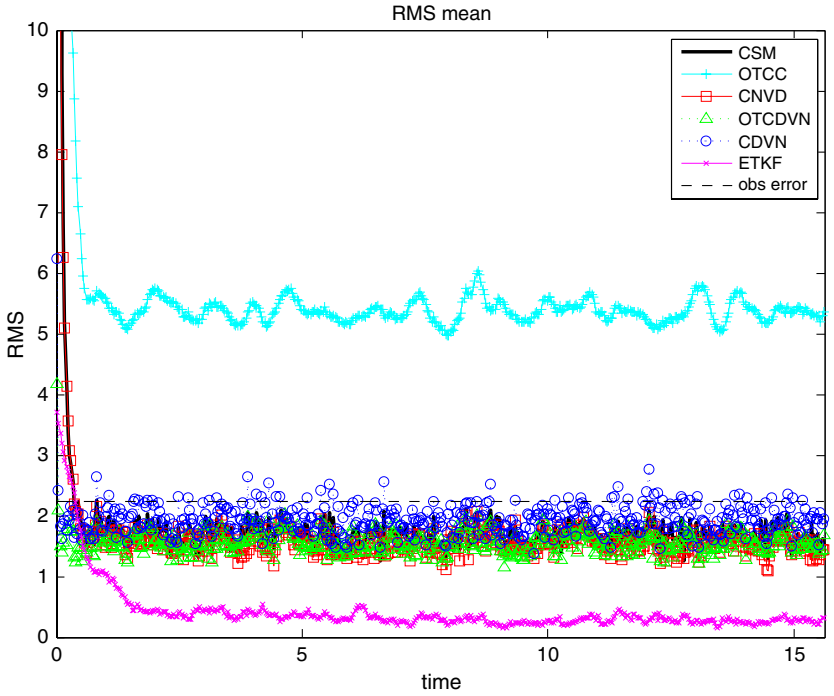


**Figure 10.** Regime $F = 6$, $n = 2$, $r^o = 5$, RMS temporal average errors as functions of time.

### *4.4. Implicit time-discretized scheme*

In a real time prediction problem, another source of model error is from the numerical integrator. Typically, the numerical solution of the governing equation is integrated with a temporal discretized scheme. In [8], the authors found surprising filter skill for turbulent signals in various regimes by utilizing implicit schemes with variable noise chosen by an information criterion. In the following, as in [8] we implement the Fourier diagonal filtering strategy with temporal implicit schemes such as the backward Euler and the trapezoidal methods. Mathematically, instead of using the exact solution of (6) and noise variance (9), we have the following pairs:

$$F_k = (1 - (A_k - d_k)T_{\text{obs}})^{-1},$$
$$r_k = \sigma_k^2 T_{\text{obs}} |1 - (A_k - d_k)T_{\text{obs}}|^{-2}$$

for backward Euler and

$$F_k = \frac{1 + (A_k - d_k)T_{\text{obs}}/2}{1 - (A_k - d_k)T_{\text{obs}}/2},$$
$$r_k = \sigma_k^2 T_{\text{obs}} |1 - (A_k - d_k)T_{\text{obs}}/2|^{-2}$$

for the trapezoidal method.

In this paper, we show results implementing these time-discretized schemes to the CSM model and we compare it with the adaptive approach as discussed in section 3.2.3. That is, we retain $F_k$ from these implicit schemes and we choose the best filtered solution among the rescaled time-discretized variance $c_k r_k$ with $c_k > 1$. Thus, we are mimicking the strategies from [8] where for a given implicit scheme we optimized the noise by an information criterion.

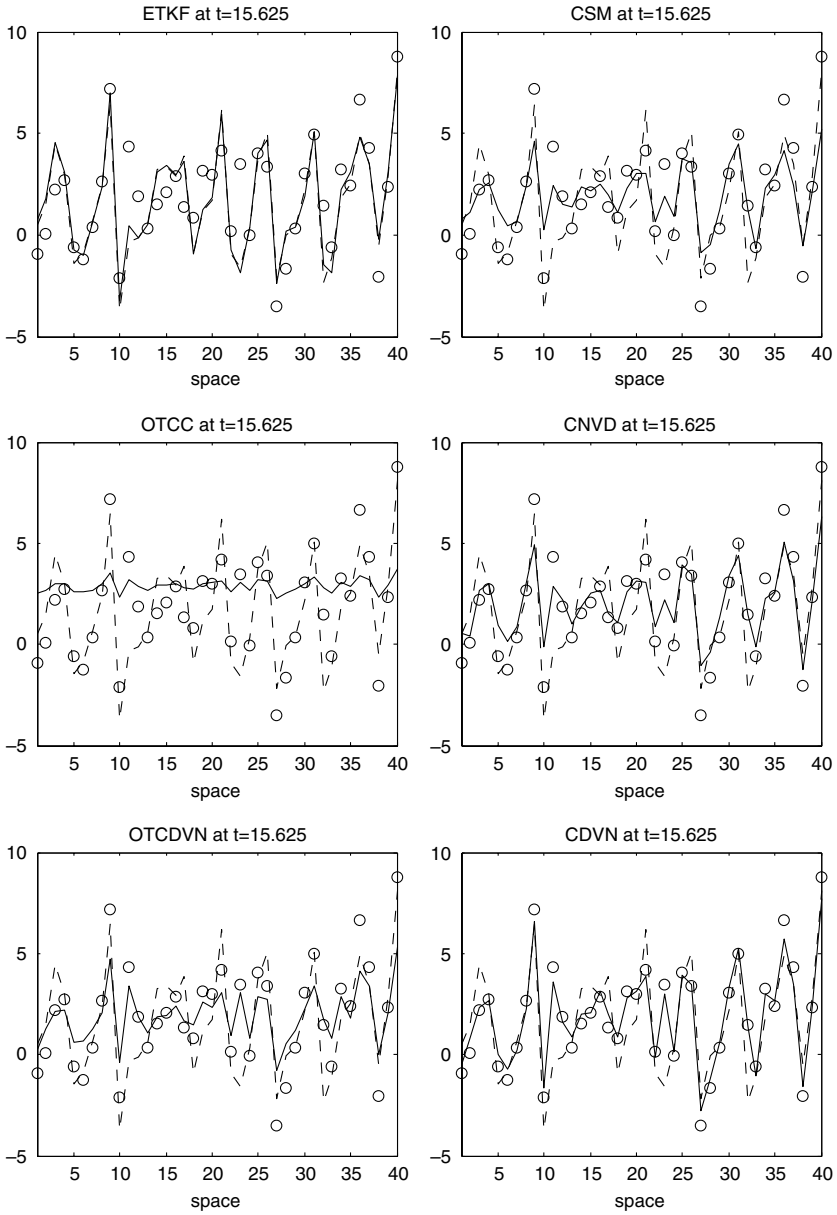


**Figure 11.** Regime $F = 6$, $n = 2$, $r^o = 5$, filtered solution (solid), true signal (dashes) and observations (circle) as functions of model space after 500 assimilation cycles.

In figure 14, the simulation for the case $F = 6, n = 2, r^o = 5$ suggests that filtering with the trapezoidal scheme is the best among these two implicit schemes. In this regime, the model error of the trapezoidal scheme is 3.4 in RMS average, which is comparable to the model error of the analytical solution of the CSM model with RMS error 3.2 (see figure 1), whereas the model error of the backward Euler scheme is even smaller with RMS average 2.3. From our offline criteria, we found that the system noise variance of the backward Euler is $r_k \approx 10^{-4}$ while the trapezoidal is $r_k \approx 10^{-3}$. When the adaptive approach is implemented

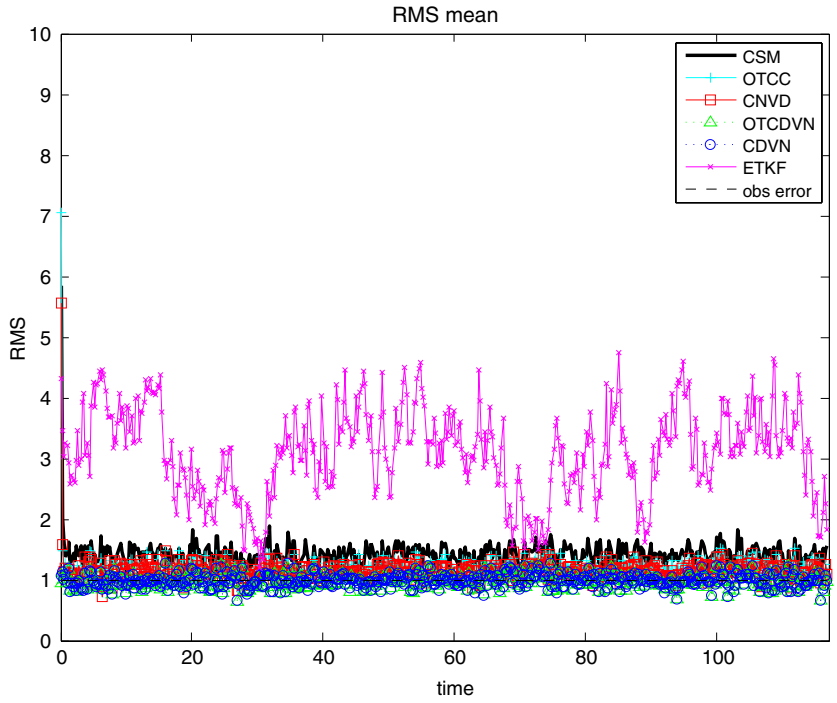


**Figure 12.** Regime $F = 8$, $n = 15$, $r^o = 1$, RMS temporal average errors as functions of time.

to backward Euler, the system noise variance is raised to $r_k \approx 10^{-3}$. Therefore, the failure of the straightforward backward Euler scheme is due to the violation of controllability since the model errors in all these approaches are relatively small when the time step is small and suitable noise inflation is needed [8].

For the case $F = 8$, $n = 15$, $r^o = 9$ (see figure 15), the model errors from both discretized schemes are larger than 5 in RMS average. The additional noise variance from the adaptive strategies increases the model error tremendously; however these model errors are mostly contributed by the stochastic term of the model. Consequently, it pushes the filter to weight more towards the observations and from our offline calculation, the limiting Kalman gain is $K_{k,\infty} \approx 1$ for trapezoidal and $K_{k,\infty} \approx 0.8$ for the backward Euler. From this finding, we can also improve the backward Euler by inflating the system noise variance more. These behaviours are analogues to the noise inflations in CDVN and OTCDVN over the CSM and OTCC, respectively.

### *4.5. Sensitivity to ensemble size and computational costs*

In all previous experiments, we fixed the ensemble size to be $K = 40$ and for the ETKF the filter is implemented with variance inflation coefficient $r = 0.05$ (see, e.g. [39, 23] for details on implementing ETKF with variance inflation). Now, we consider the sensitivity of the Fourier diagonal filtering strategy, that is the Fourier mode-by-mode decoupled scalar filter with linear stochastic dynamics, towards the variation of ensemble size. In this paper, as an example, we show the sensitivity for two regimes with the CSM model: $F = 6$, $n = 2$, $r^o = 5$ and case $F = 16$, $n = 15$, $r^o = 9$. The first case is when the ETKF supersedes the CSM filter while the second case is when the CSM filter is better than the ETKF (see figures 3 and 4).

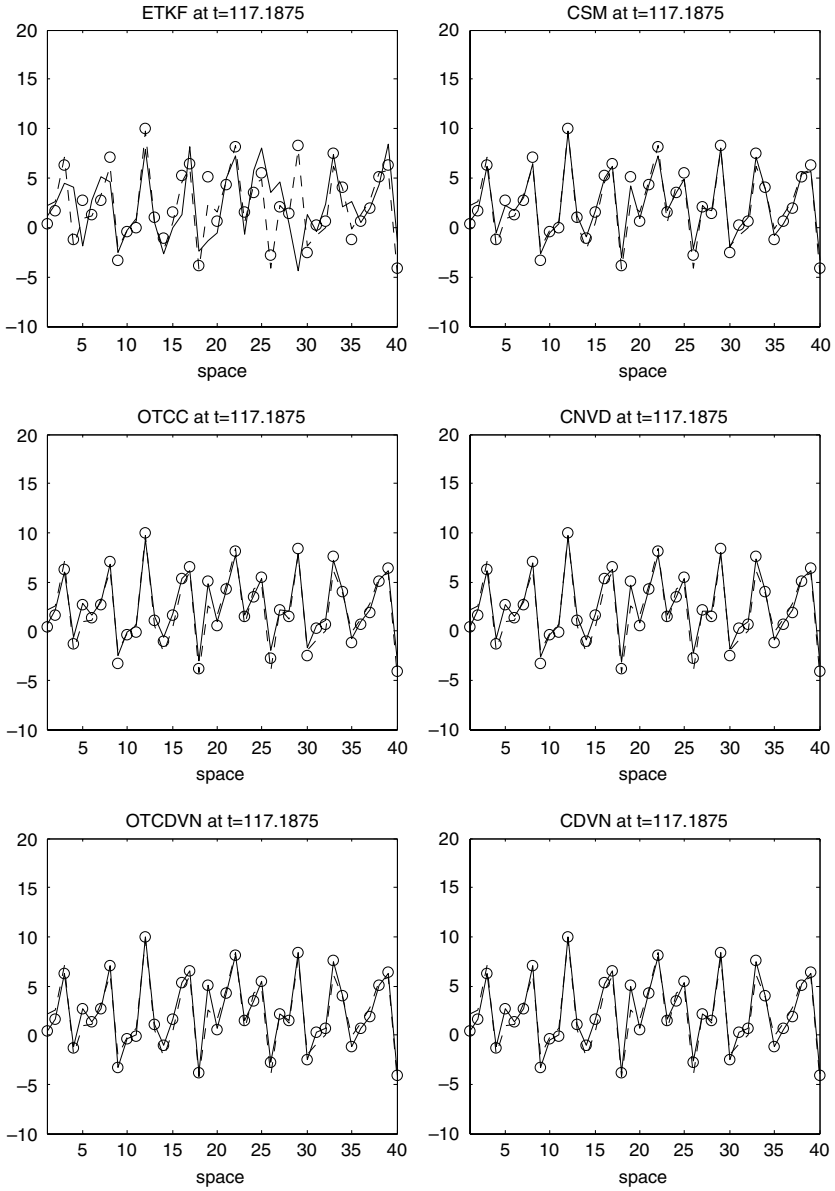


**Figure 13.** Regime $F = 8$, $n = 15$, $r^o = 1$, filtered solution (solid), true signal (dashes) and observations (circle) as functions of model space after 500 assimilation cycles.

In figure 16, we plot the RMS errors as functions of ensemble size (left column). This result suggests that ETKF is very sensitive to ensemble size. In particular, the performance is significantly better for large ensemble size and extremely poor for smaller ensemble size. On the other hand, the CSM filter performance for a single realization does not differ very much compared with that of using ensemble size $K = 100$. Secondly, the CSM filter does not depend on any tunable parameter like variance inflation coefficient $r$. In these experiments, the ETKF uses $r = 0.4, 0.2, 0.1, 0.05$ and $0.01$ for ensemble size $K = 5, 10, 20, 40$ and $100$, respectively.

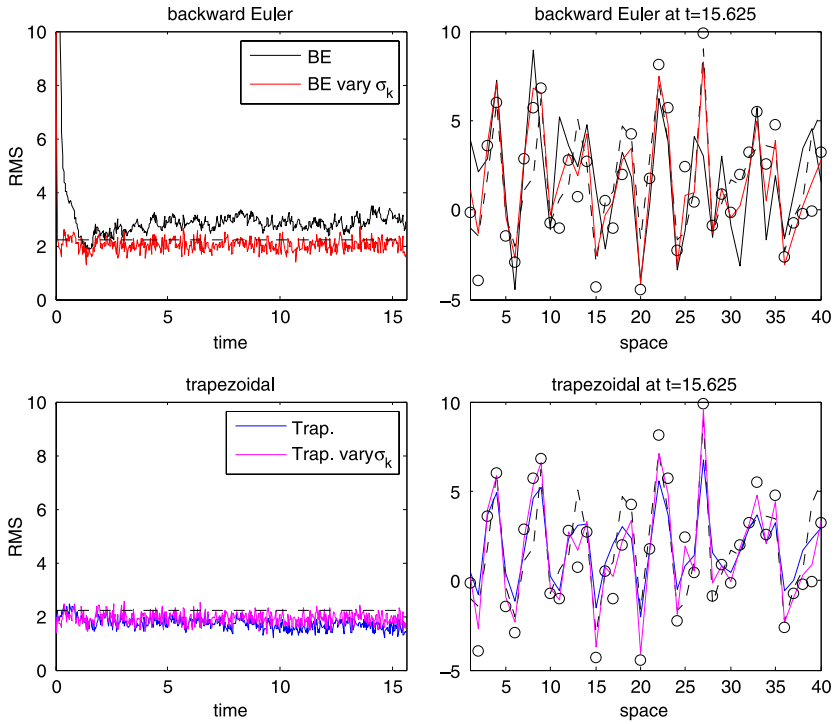


**Figure 14.** Filtering with implicit schemes for regime $F = 6$, $n = 2$, $r^o = 5$, we show the RMS temporal average errors as functions of time (first column) and the filtered solutions as functions of model space (second column), compared with the true signal (dashes) and observations (circle).

The highest time consumption in data assimilation is typically in the evolution steps, as one can see in our example, the ETKF with $n = 15$ is much slower than that with $n = 2$ (see right column of figure 16). In general this varies for different models. With the Fourier diagonal filtering strategy, however, the filter spends the same amount of time on propagating each state independently of $n$ since the linear stochastic model is solved analytically. Secondly, in each correction step (or the step where the filter includes the observations to produce the posterior state) ETKF requires an inverse of a $K \times K$ matrix to obtain the mean posterior state and a matrix square root of a $K \times K$ matrix to generate the posterior ensemble state [23, 39], while the Fourier diagonal filtering strategy reduces to filtering $J/2$ independent wave numbers with each correction involving only scalar operations. In figure 16 (right column), we show an example of computational cost comparison by just looking at the CPU wall clock time for each simulation (beware that these quantities can be very different for different CPU and coding style, but we show a reasonably optimized code for both schemes run with an Intel Pentium 3Ghz processor and 2Gb RAM), where each simulation includes 500 assimilation cycles.

## 5. Summary

In this paper, we introduce a radical strategy for filtering signals of nonlinear dynamical systems that exhibit a wide range of behaviour from weakly chaotic to highly turbulent. Computationally, we reduce the problem of filtering a nonlinear spatially extended system to

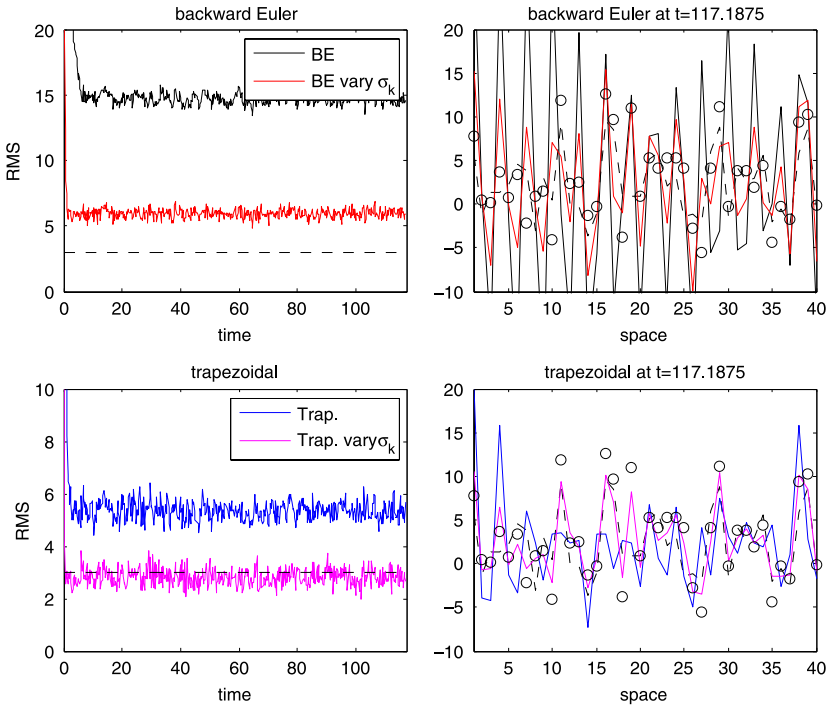


**Figure 15.** Filtering with implicit schemes for the case $F = 16$, $n = 15$, $r^o = 9$, RMS temporal average errors as functions of time (first column) and the filtered solutions as functions of model space (second column), compared with the true signal (dashes) and observations (circle).

filtering decoupled one-dimensional parametric linear stochastic models. We then parametrize the linear stochastic model such that it statistically resembles the physical characteristic of the nonlinear dynamical system. Our approaches include fitting the model to the climatological variance, the correlation time and the observation time correlation, which yield CSM (see section 3.1) and OTCC (section 3.2.1) filters.

Other parametrization strategies such as the CNVD, CDVN and OTCDVN (sections 3.2.2–3.2.4) are motivated from the offline test criteria, introduced in [7] and developed in [8]. That is, we parametrize the model such that the filter becomes more controllable as well as to have the least RMS model error. Thus, we are following the general strategy, first put forward in [37], to assess both model errors and filter performance with the given model simultaneously.

This linear approximation obviously introduces physical modelling errors since we replace the nonlinear discrete advective term in the L-96 model with an Ornstein–Uhlenbeck process (linear damping + white noise) in each Fourier mode. In our numerical experiments, we see that even with this rough approximation, the radical filtering strategy produces a reasonably skilful filtered solution in various regimes when the model errors are relatively small. These encouraging results open new research directions. In future work, we will introduce a more systematic parametrization strategy that minimizes the model error in a probabilistic sense utilizing only the statistics of the dynamics. Another direction is to keep our radical strategy (of ignoring the correlation between different dynamics) but use other stochastic processes to replace the nonlinear mixing terms. In the above results, similar performance is also seen when the filter includes model errors from the numerical solver where in our experiments we show results using two implicit schemes: backward Euler and trapezoidal. From our numerical

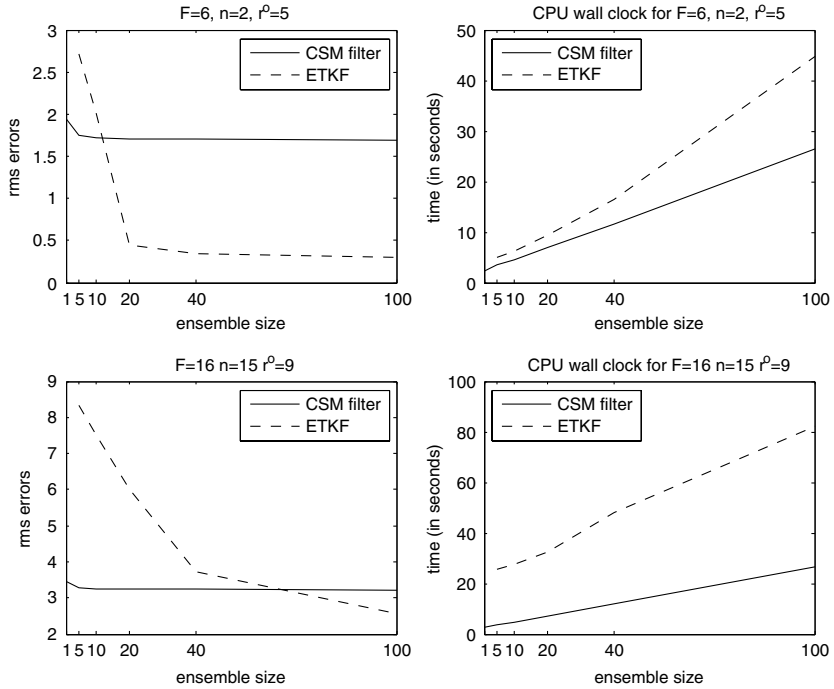


**Figure 16.** The first column shows the RMS errors as functions of ensemble size and the second column shows their corresponding CPU wall clock time for each simulation. The top figures are comparison in regime case I and the bottom figures are comparison for the case $F = 16, n = 15, r^o = 9$.

simulations, we conclude that when the model errors are small, the linear analysis using the offline test criteria suggests that satisfying controllability condition is the central role for accurate filtering. When the model error is large and mostly contributed by the deterministic part of the filtering model, the linear analysis suggests that filtering failure is because the filter over weights towards an extremely inaccurate model. However, when the model error is enormous and mostly contributed by the stochastic part of the filter model, then this volatility forces the filter to weight more towards the observations. In future work, we plan to test the radical diagonal filtering strategies compared with ETKF in the severe environment of sparse regular observations where it is known [22] that ETKF often suffers filter divergence.

Practically, the Fourier diagonal filtering strategy is very attractive for high dimensional problems since it is independent of tunable parameters, and importantly, it is very robust towards the variation of ensemble size. Our crude strategy of modelling the nonlinear dynamical system allows one to generate the prior forecast state instantly while ensemble filters spend most of their computational time generating an ensemble of prior forecast states for estimating the forecast uncertainty. In contrast, we show that ETKF is very sensitive towards the ensemble size and also to the variance inflation coefficient. From our numerical simulations, we found that in a highly chaotic regime with turbulent dynamics, the Fourier diagonal filtering strategy supersedes the ETKF when the observation time interval is large. There, the Fourier diagonal filtering strategies force the filter to trust the observations, whereas in ETKF, the Gaussian assumption is no longer valid since the filter model is highly nonlinear (even when the solutions are still unimodally distributed and weakly skewed from a Gaussian distribution [34]) for such a long observation time.

## Acknowledgments

The research of Andrew 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 grant N00014-07-1-0750. John Harlim is supported as a postdoctoral fellow through the last two agencies.

## References

[1] Evensen G 2003 The ensemble Kalman filter: theoretical formulation and practical implementation *Ocean Dyn.* **53** 343–67
[2] Bishop C H, Etherton B and Majumdar S J 2001 Adaptive sampling with the ensemble transform Kalman filter: I. The theoretical aspects *Mon. Weather Rev.* **129** 420–36
[3] Anderson J L 2001 An ensemble adjustment Kalman filter for data assimilation *Mon. Weather Rev.* **129** 2884–903
[4] Szunyogh I, Kostelich E J, Gyarmati G, Patil D J, Hunt B R, Kalnay E, Ott E and Yorke J A 2005 Assessing a local ensemble Kalman filter: perfect model experiments with the NCEP global model *Tellus* A **57** 528–45
[5] Hunt B R, Kostelich E J and Szunyogh I 2007 Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter *Physica* D **230** 112–26
[6] Lorenz E N 1996 Predictability—a problem partly solved *Proc. on Predictability (ECMWF, Reading, UK, 4–8 September 1995)* pp 1–18
[7] Majda A J and Grote M J 2007 Explicit off-line criteria for stable accurate time filtering of strongly unstable spatially extended systems *Proc. Natl Acad. Sci.* **104** 1124–9
[8] Castronovo E, Harlim J and Majda A J 2008 Mathematical criteria for filtering complex systems: plentiful observations *J. Comput. Phys.* **227** 3678–714
[9] Majda A J, Abramov R V and Grote M J 2005 *Information Theory and Stochastics for Multiscale Nonlinear Systems* (*CRM Monograph Series* vol 25) (Providence, RI: American Mathematical Society)
[10] Majda A J and Wang X 2006 *Nonlinear Dynamics and Statistical Theories for Basic Geophysical Flows* (Cambridge: Cambridge University Press)
[11] Chorin A J and Krause P 2004 Dimensional reduction for a bayesian filter *Proc. Natl Acad. Sci.* **101** 15013–7
[12] Del Moral P 1996 Nonlinear filtering: interacting particle solutions *Markov Process. Relat. Fields* **2** 555–80
[13] Del Moral P and Jacod J 2001 Interacting particle filtering with discrete observation *Sequential Monte-Carlo Methods in Practice* (*Statistics for Engineering and Information Science*) ed A Doucet *et al* (Berlin: Springer) pp 43–75
[14] Rossi V and Vila J-P 2006 Nonlinear filtering in discrete time: a particle convolution approach *Ann. Inst. Stat. Univ. Paris* **3** 71–102
[15] Berliner L M, Milliff R F and Wikle C K 2003 Bayesian hierarchical modeling of air-sea interaction *J. Geophys. Res.* **108** 3104–20
[16] Miller R N, Carter E F and Blue S T 1999 Data assimilation into nonlinear stochastic models *Tellus* A **51** 167–94
[17] Ghil M and Malanotte-Rizolli P 1991 Data assimilation in meteorology and oceanography *Adv. Geophys.* **33** 141–266
[18] Todling R and Ghil M 1994 Tracking atmospheric instabilities with the kalman filter: I. Methodology and one-layer results *Mon. Weather Rev.* **122** 183–204
[19] Anderson J L 2003 A local least squares framework for ensemble filtering *Mon. Weather Rev.* **131** 634–42
[20] Farrell B F and Ioannou P J 2001 State estimation using a reduced-order kalman filter *J. Atmos. Sci.* **58** 3666–80
[21] Farrell B F and Ioannou P J 2005 Distributed forcing of forecast and assimilation error systems *J. Atmos. Sci.* **62** 460–75
[22] Ott E, Hunt B R, Szunyogh I, Zimin A V, Kostelich E J, Corrazza M, Kalnay E and Yorke J A 2004 A local ensemble Kalman filter for atmospheric data assimilation *Tellus* A **56** 415–28
[23] Harlim J and Hunt B R 2007 Four-dimensional local ensemble transform Kalman filter: numerical experiments with a global circulation model *Tellus* A **59** 731–48
[24] Anderson B D and Moore J B 1979 *Optimal Filtering* (Englewood Cliffs, NJ: Prentice-Hall)
[25] Chui C K and Chen G 1999 *Kalman Filtering* (New York: Springer)
[26] Kaipio J P and Somersalo E 2005 *Statistical and Computational Inverse Problems* (New York: Springer)
[27] Haven K, Majda A J and Abramov R 2005 Quantifying predictability through information theory: small sample estimation in a non-gaussian framework *J. Comput. Phys.* **206** 334–62
[28] Delsole T 2004 Stochastic model of quasigeostrophic turbulence *Surv. Geophys.* **25** 107–49

[29] Majda A J, Timofeyev I and Vanden-Eijnden E 2001 A mathematical framework for stochastic climate models *Commun. Pure Appl. Math.* **54** 891–974
[30] Majda A J and Timofeyev I 2004 Low dimensional chaotic dynamics versus intrinsic stochastic chaos: a paradigm model *Physica* D **199** 339–68
[31] Majda A J, Timofeyev I and Vanden-Eijnden E 2006 Stochastic models for selected slow variables in large deterministic systems *Nonlinearity* **19** 769–94
[32] Franzke C and Majda A J 2006 Low-order stochastic mode reduction for a prototype atmospheric GCM *J. Atmos. Sci.* **63** 457–79
[33] Abramov R V and Majda A J 2007 Blended response algorithm for linear fluctuation-dissipation for complex nonlinear dynamical systems *Nonlinearity* **20** 2793–821
[34] Abramov R and Majda A J 2004 Quantifying uncertainty for non-gaussian ensembles in complex systems *SIAM J. Sci. Comput.* **25** 411–47
[35] Grote M J and Majda A J 2006 Stable time filtering of strongly unstable spatially extended systems *Proc. Natl Acad. Sci.* **103** 7548–53
[36] Gardiner C W 1997 *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences* (New York: Springer)
[37] Tribbia J J and Baumhefner D P 1988 The reliability of improvements in deterministic short-range forecasts in the presence of initial state and modeling deficiencies *Mon. Weather Rev.* **116** 2276–88
[38] Wang X, Bishop C H and Julier S J 2004 Which is better, an ensemble of positive-negative pairs or a centered spherical simplex ensemble? *Mon. Weather Rev.* **132** 1590–605
[39] Harlim J and Hunt B R 2007 A non-Gaussian ensemble filter for assimilating infrequent noisy observations *Tellus* A **59** 225–37