Available online at www.sciencedirect.com
ScienceDirect
Journal of Computational Physics 227 (2008) 3678–3714
JOURNAL OF COMPUTATIONAL PHYSICS
www.elsevier.com/locate/jcp

Mathematical test criteria for filtering complex systems: Plentiful observations

E. Castronovo, J. Harlim, A.J. Majda *

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

Received 27 February 2007; received in revised form 30 November 2007; accepted 3 December 2007
Available online 1 January 2008

Abstract

An important emerging scientific issue is the real time filtering through observations of noisy turbulent signals for complex systems as well as the statistical accuracy of spatio-temporal discretizations for such systems. These issues are addressed here in detail for the setting with plentiful observations for a scalar field through explicit mathematical test criteria utilizing a recent theory [A.J. Majda, M.J. Grote, Explicit off-line criteria for stable accurate time filtering of strongly unstable spatially extended systems, Proceedings of the National Academy of Sciences 104 (4) (2007) 1124–1129]. For plentiful observations, the number of observations equals the number of mesh points. These test criteria involve much simpler decoupled complex scalar filtering test problems with explicit formulas and elementary numerical experiments which are developed here as guidelines for filter performance. The theory includes information criteria to avoid filter divergence with large model errors, asymptotic Kalman gain, filter stability, and accurate filtering with small ensemble size as well as rigorous results delineating the role of various turbulent spectra for filtering under mesh refinement. These guidelines are also applied to discrete approximations for filtering the stochastically forced dissipative advection equation with very turbulent and noisy signals with either an equipartition of energy or $-5/3$ turbulent spectrum with infrequent observations as severe test problems. The theory and companion simulations demonstrate accurate statistical filtering in this context with implicit schemes with large time step with very small ensemble sizes and even with unstable explicit schemes under appropriate circumstances provided the filtering strategies are guided by the off-line theoretical criteria. The surprising failure of other strongly stable filtering strategies is also explained through these off-line criteria.

Keywords: Finite difference; Kalman filter; Turbulence; Data assimilation

1. Introduction

The need for real time predictions in extended range forecasting of weather and climate drives the development of improved strategies for data assimilation or filtering. Filtering combines the observed features of

* Corresponding author.
E-mail address: jonjon@cims.nyu.edu (A.J. Majda).

0021-9991/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcp.2007.12.016

the chaotic turbulent multi-scale signal together with the evolution of a dynamical model for the coupled atmosphere–ocean system. The dynamical models, general circulation models, often have significant model errors compared with the behavior in the observed signal and both the model and observed signal involve many spatio-temporal scales, rough turbulent energy spectra near the resolved mesh scale, and a large dimensional state space. There is also the inherently difficult practical issue of the "curse of ensemble size" since there is a large computational overhead in propagating the dynamical operator and this restricts the predictions to relatively small ensemble sizes [18]. Another major issue [29] is that one might improve the resolution of the forward model but not improve the filtering skill significantly due to model error or the nature of the observations.

One goal of the present paper is to develop theoretical criteria as guidelines which can address all the above issues for filtering turbulent signals in an idealized context which nevertheless can provide useful insight into these difficult problems. A second goal is to provide alternative computational strategies to filter noisy turbulent signals which have some skill while dealing with the "curse of ensemble size". Thus, a central scientific issue is the following one: How to develop practical mathematical criteria for accurate filtering of such complex systems which can increase computational efficiency while remaining a statistically accurate approximation to the true filtering behavior? This clearly depends on many features of the complex filtering problem such as:

(I.A) The specific underlying dynamics.
(I.B) The energy spectrum at spatial mesh scales of the observed system and the system noise, i.e. decorrelation time, on these scales.
(I.C) The number of observations and the strength of the observational noise.
(I.D) The time scale between observations relative to (I.A) and (I.B).

Central practical computational issues for the filtering in the above context to avoid the "curse of ensemble size" are the following:

(II.A) When is it possible to use for filtering the standard explicit scheme solver of the original dynamic equations by violating the CFL stability condition with a large time step equal (proportional) to the observation time to increase ensemble size yet retain statistical accuracy?
(II.B) When is it possible to use for filtering a standard implicit scheme solver for the original dynamic equations by using a large time step equal to the observation time to increase ensemble size yet retain statistical accuracy?

Clearly resolving the practical issues in II involves the understanding of I in a given context. This paper involves the development of explicit mathematical criteria for filtering complex systems which address all of the subtle issues outlined in I and II for a scalar field with "plentiful" observations, i.e., the number of observation points equals the number of discrete mesh points. The work presented here is based on the mathematical theory developed recently in [22] where the analogue of linearized constant coefficient stability and error analysis for deterministic finite difference schemes [27] has been developed for complex filtering problems incorporating all of the features in I for general $s \times s$ systems of stochastic partial differential equations with either plentiful or sparse observations. In the simplified context of the present paper, there is large model error introduced through standard discretizations which attempt to filter noisy turbulent signals with infrequent observations compared to the local correlation time of the turbulence.

For plentiful observations, this theory rigorously establishes that the understanding of filtering properties for difference approximations to the stochastically forced scalar PDE reduces to understanding the filtering properties of spatio-temporal discrete approximations to a much simpler decoupled complex scalar constant test problem for each spatial wave number. These complex scalar test problems are given by

$$\frac{\mathrm{d}u(t)}{\mathrm{d}t} = \lambda u(t) + \tilde{\sigma}\dot{W}(t), \tag{1}$$

where $u(t) = a(t) + ib(t)$ and $\dot{W}(t) = \dot{W}_1(t) + i\dot{W}_2(t)$ is complex white noise, i.e. $\dot{W}_1(t)$ and $\dot{W}_2(t)$ are independent real white noises in time with variance 1/2. The theory from [22] is summarized in Section 2 for the setting in this paper involving plentiful observations. Thus, to achieve understanding of the issues in I and II and the role of model error for the case of plentiful observations, one needs first to develop a complete understanding of these issues for filtering the simpler test problems in (1); this is developed in Section 3 below through explicit mathematical formulas for the asymptotic Kalman gain and asymptotic filtering stability for (1) as functions of the parameters listed in I as well as an explicit information criteria to reduce model error; the practical computational issues for filtering in II are also addressed in Section 3 for the simpler scalar test problem in (1) which are often necessarily in "stiff regimes" in order to develop meaningful criteria for filtering turbulent signals for the discretized PDE.

In Section 4, we provide rigorous theoretical criteria in this idealized context which provide guidelines into the following important practical questions for operational models: If plentiful observations are available on refined meshes, what is gained by increasing the resolution of the operational model? How does this depend on the nature of the turbulent spectrum? The analysis is elementary but involves an interesting interplay among the conditions in I.

The explicit test criteria developed in Section 3 are applied in Section 5 for filtering turbulent signals with unstable explicit and stable implicit upwind or centered difference approximations to the stochastically forced dissipative advection equation for a variety of turbulent spectra and mesh sizes; also numerical experiments are presented there which demonstrate various practical facets of the theory regarding the issues in I and II. It is worthwhile to note that some operational models utilize implicit schemes for the gravity waves [28] so that our results on implicit schemes have additional interest besides the present context. The computational results in Section 5 establish that the off-line mathematical criteria involving information theory for system noise, the asymptotic Kalman gain, and asymptotic filter stability can be used to both determine and explain the success or failure of various filtering strategies with large model errors for turbulent solutions of the stochastic PDE. These results also establish the utility of the Fourier diagonal filters as guidelines for the behavior of the extended Kalman filter in physical space. They also suggest alternative Fourier diagonal filtering strategies with model error which nevertheless have significant skill in filtering turbulent signals yet avoid the "curse of ensemble size". An application of these ideas to a chaotic turbulent dynamical system with forty degrees of freedom has been developed recently by two of the authors [17].

The present work is motivated by earlier work on Bayesian hierarchical modeling [4] and reduced order filtering strategies [25,13,30,2,3,7,10,11,26,16] that have been developed with some success in these extremely complex systems. The basis for such dynamic prediction strategies for these complex spatially extended systems is the classical Kalman filtering algorithm [8,20,1] which is also utilized here.

Finally, we mention interesting work of Cohn and Dee [9] who developed a theory for checking the observability criteria [8,19,14] for filtering discrete approximations to constant coefficient PDE's. Recently, Grote and Majda [14] developed simple crude mathematical criteria for filtering unstable systems and explicitly demonstrated their importance in stable accurate filtering utilizing an unstable difference scheme for a stochastically forced convection-diffusion equation with identical parameters as in Section 5.2 but with a smoother spectrum. In [14], it was also demonstrated that the simple observability criteria from [9] provide a necessary condition for stable filtering but have limited utility from a practical viewpoint in addressing the central issues in II above; examples are given in [14] where the observability criterion from [9] is satisfied but is practically useless because the asymptotic filter covariance matrix has condition number 10^{13}!

2. Theory for filtering discretizations of stochastic PDE's with plentiful observations

2.1. The perfect model signals

The signals which will be filtered through plentiful observations applied to various spatio-temporal discretization, the perfect truth signals, are determined by solutions of the real valued scalar stochastically forced PDE

$$\frac{\partial u(x,t)}{\partial t} = \mathcal{P}\left(\frac{\partial}{\partial x}\right)u(x,t) - \gamma\left(\frac{\partial}{\partial x}\right)u(x,t) + \sigma(x)\dot{W}(t), \tag{2}$$

$$u(x,0) = u_0(x). \tag{3}$$

Here $\sigma(x)\dot{W}(t)$ is a Gaussian statistically stationary spatially correlated scalar random field and $\dot{W}(t)$ is white noise in time while the initial data u_0 is a Gaussian random field with non-zero mean and covariance. As in usual finite difference linear stability analysis, the problem in (2) is non-dimensionalized to a 2π-periodic domain so that continuous and discrete Fourier series can be utilized in analyzing (2) and the related discrete approximations.

The operators $\mathcal{P}\left(\frac{\partial}{\partial x}\right)$ and $\gamma\left(\frac{\partial}{\partial x}\right)$ are defined through unique symbols at a given wave number k by

$$\begin{aligned}\mathcal{P}\left(\frac{\partial}{\partial x}\right)e^{ik\cdot x} &= \tilde{p}(ik)e^{ikx},\\ \gamma\left(\frac{\partial}{\partial x}\right)e^{ik\cdot x} &= \gamma(ik)e^{ikx}.\end{aligned} \tag{4}$$

We assume that $\tilde{p}(ik)$ is wave-like so that

$$\tilde{p}(ik) = i\omega_k \tag{5}$$

with $-\omega_k$ the real valued dispersion relation while $\gamma(ik)$ represents both explicit and turbulent dissipative processes so that $\gamma(ik)$ is non-negative with

$$\gamma(ik) > 0 \quad \text{for all } k \neq 0. \tag{6}$$

In geophysical applications, it is natural to have a climatological distribution and as discussed below, (5 and 6) are needed in order to guarantee this.

2.2. The stochastically forced dissipative advection equation

The main example of (2) studied in Section 5 as a prototype in this paper is given by the stochastically forced dissipative advection equation

$$\frac{\partial u(x,t)}{\partial t} = -c\frac{\partial u(x,t)}{\partial x} - d\,u(x,t) + \mu\frac{\partial^2 u(x,t)}{\partial x^2} + \sigma(x)\dot{W}(t). \tag{7}$$

In this example, $\tilde{p}(ik) = i\omega_k = -ick$ and the damping symbol $\gamma(ik)$ is given by

$$\gamma(ik) = d + \mu k^2. \tag{8}$$

The slight abuse of notation in (7) and (8) should not confuse the reader. In (8), we require $d \geqslant 0$ and $\mu \geqslant 0$, and at least one of these coefficients to be non-zero in order to satisfy (6). The case with uniform damping, $d > 0$, but without scale dependent damping so that $\mu = 0$ arises often in idealized geophysical problems where d represents radiative damping, Ekman friction or gravity wave absorption [23,24]. In general $\mathcal{P}\left(\frac{\partial}{\partial x}\right)$ can be any differential operator which is a combination of odd-derivatives to satisfy (5) while $\gamma\left(\frac{\partial}{\partial x}\right)$ is a suitable combination of even derivative satisfying (6) (see Chapter 1 of [24] for the precise conditions). The full generality in (4) is important for geophysical equations such as the quasi-geostrophic equations where $\tilde{p}(ik)$ is not a polynomial but is given by $\tilde{p}(ik) = \frac{ik}{k^2+F}$ [24], where F is a non-dimensionalized unit that represents the square of a ratio between the Froude and the Rossby numbers.

The general solution of (2) is defined through Fourier series. The 2π-periodic solution of (2) is expanded in Fourier series

$$u(x,t) = \sum_{k=-\infty}^{\infty} \hat{u}_k(t)e^{ikx}, \quad \hat{u}_{-k} = \hat{u}_k^*, \tag{9}$$

where $\hat{u}_k(t)$ for $k > 0$ solves the scalar complex coefficient stochastic ODE's [12],

$$d\hat{u}_k = [\tilde{p}(ik) - \gamma(ik)]\hat{u}_k\,dt + \tilde{\sigma}_k\,dW_k, \quad \hat{u}_k(0) = \hat{u}_{k,0}. \tag{10}$$

Here the W_k are independent complex Wiener processes for each k and the independent real and imaginary parts have the same variance 1/2; the coefficients $\hat{u}_{-k}$ for $k > 0$ are defined through the complex conjugate formula $\hat{u}_{-k} = \hat{u}_k^*$ and the constant $k = 0$ Fourier mode is real-valued with a similar single equation with detailed discussion omitted here. Under the natural simplifying assumption that the symbols $\tilde{p}(\mathrm{i}k)$ and $\gamma(\mathrm{i}k)$ satisfy (5) and (6), the statistical equilibrium distribution for (10) exists and is a Gaussian with zero mean and variance, E_k, defining the climatological energy spectrum given by

$$E_k = \frac{\tilde{\sigma}_k^2}{2\gamma(\mathrm{i}k)}, \quad 1 \leqslant k < +\infty. \tag{11}$$

Mathematically, one needs to require $\sum E_k < \infty$ to define the stochastic solution of (2) correctly with a similar requirement on the Gaussian initial data in $u_0(x)$. While distinct Fourier modes with different magnitudes are uncorrelated, the correlation function at a given mode in the statistical steady state is given by

$$\langle u_k(t')u_k^*(t)\rangle = R_k(|t-t'|) \tag{12}$$

$$\langle u_k(t')u_k^*(t)\rangle = E_k \mathrm{e}^{-\gamma_k|t-t'|}\cos(\omega_k(t-t')). \tag{13}$$

In (12) and (13), the damping coefficient, $\gamma_k = \gamma(\mathrm{i}k)$ defines the correlation time, γ_k^{-1}, while $\mathrm{i}\omega_k = \tilde{p}(\mathrm{i}k)$ defines ω_k, the oscillation frequency at wave number k. Clearly, γ_k^{-1} measures the memory in the signal being filtered. As discussed in [22], the noise in (10) and (11) represents the turbulent fluctuations on the mesh scale for both unresolved and resolved features of the non-linear dynamics ([24,23] and references therein) with a given energy spectrum E_k and decorrelation time γ_k at each wave number. In this fashion the features in (I.A) and (I.B) are incorporated in the constant coefficient test problem. In practical problems, quite often the nature of this spectrum is known roughly as well as the decorrelation time, expressed, through the damping coefficient γ_k [24,23,22]. The theory from [14] also applies in the unstable setting with $\gamma_k < 0$ but will not be discussed here. In this paper, one space dimension is not a restriction for any of the results but is utilized to avoid cumbersone notation; the theory below also applies in several variables.

2.3. The discrete approximation

Standard finite difference approximations operate on a family of equispaced $2N+1$ mesh points, $x_j = jh$, $0 \leqslant j \leqslant 2N$, with $(2N+1)h = 2\pi$. If real-valued functions f_j, are defined on mesh points then with the complex inner product,

$$(f, \bar{g})_h = \frac{h}{2\pi}\sum_{j=0}^{2N} f_j g_j, \tag{14}$$

the discrete Fourier coefficients, $\hat{f}_k$ are defined by $\hat{f}_k = (f, \mathrm{e}^{\mathrm{i}kx_j})_h$ for $|k| \leqslant N$, with the well-known properties

$$f_j = \sum_{|k|\leqslant N} \hat{f}_k \mathrm{e}^{\mathrm{i}kx_j}, \quad \hat{f}_{-k} = \hat{f}_k^*, \tag{15}$$

$$(f,f)_h = \sum_{|k|\leqslant N} |\hat{f}_k|^2.$$

For a standard finite difference approximation to (2) without any random noise with a time step Δt, the solution is expressed in standard fashion [27] in terms of the amplification factor, $F_{h,k}$, for integer k with $|k| \leqslant N$. With the finite Fourier expansion

$$u^h = \sum_{|k|\leqslant N} \hat{u}_k^h \mathrm{e}^{\mathrm{i}kx}, \tag{16}$$

the general discrete approximation of (2) is given at the observation times $m\Delta t$ through its Fourier coefficients $\hat{u}_k^h$ by the block diagonal operation,

$$\widehat{u}_{k,m+1|m}^h = F_{h,k}\widehat{u}_{k,m|m}^h + \sigma_{h,k,m+1}. \tag{17}$$

In (17), the zero mean complex Gaussian noises, $\sigma_{h,k,m}$, are uncorrelated in time and their second moment averages satisfy

$$\langle \sigma_{h,k,m}\sigma^*_{h,k',m}\rangle = \delta_{k+k'} r_{h,k}, \quad |k|,\ |k'| \leqslant N \tag{18}$$

with $r_{h,k}$ the variance at wave number k and δ_j the delta function. The consistent notation $F_k = e^{(\tilde{p}(ik)-\gamma(ik))\Delta t}$ is utilized for the amplification factor of the exact solution operator in (10). In order to address the computational efficiency and accuracy issues mentioned in II above, it is not obvious that the best choice of the noise in (17) arises from a straightforward time discretization of the noise; in fact, we will see below in Section 3 that it is often interesting and advantageous to pick the noise in (17) and (18) in a completely different fashion to avoid filter divergence [8,19,1,22] even for a stable filter.

2.4. Plentiful observations

There are $2N+1$ spatial grid points in the finite difference operator; in the case of plentiful observations discussed here, there are $2N+1$ observation points, $\tilde{x}_j$, $1 \leqslant j \leqslant 2N+1$ (not necessarily the grid points) where the signal from (2) is sampled by the solution of the difference equation in (16)–(18) so that

$$v(\tilde{x}_j, m\Delta t) = gu(\tilde{x}_j, m\Delta t) = g\bar{u}^h_{m|m}(\tilde{x}_j) + \sigma^o_{j,m} \tag{19}$$

for $1 \leqslant j \leqslant 2N+1$. In (19) and throughout this paper it is assumed for simplicity in exposition that the observation time Δt, coincides exactly with the finite difference time step (see [14,22] for the more general case). In addressing the issues in II, this is not a major restriction since we are interested in approximate methods with large timesteps. Note that the signal, v, to be observed in (19) is sampled from a (truncated) solution of the stochastic PDE in (2). The observation measurement errors are assumed to be zero mean Gaussian random variables which are uncorrelated from site to site and time to time with variance $r^o = \langle (\sigma^o_j)^2\rangle$. Without loss of generality, we can set $g \equiv 1$ in all the analysis below. However, the value of g is retained below when it is useful to mark the role of the observations in various explicit formulas.

2.5. Reduction of the discrete filter problem to complex scalar test filtering problems

The finite difference approximation defined in (16)–(18) together with the plentiful observations in (19) defines a $2N+1$ dimensional filtering problem. Given a Gaussian distribution as an estimate for $u^h_{m|m}$, the finite difference scheme with noise defined in (16)–(18) is utilized to advance the Gaussian statistical estimate by the dynamics to the state $u^h_{m+1|m}$; this Gaussian state is then utilized as a prior distribution and is constrained by the observations in (19) to produce a new Gaussian distribution $u^h_{m+1|m+1}$ which serves as an estimate for the perfect signal $u(x,(m+1)\Delta t)$ from (2) (see [8,19,20,1,14,22]) by the filtering process. The Kalman filter defines the optimal filter in this setting [8,19,20,1] but there are significant model errors in the choice of discretization and time step in addressing the central issues in I and II.

Theorems 1 and 2 from [22] rigorously guarantee that the above filtering problem, even when the observation points do not coincide with the mesh points, can be analyzed by studying the simpler decoupled complex scalar filtering problems for each different Fourier wave number, k, $0 < k \leqslant N$,

$$\hat{u}^h_{k,m+1|m} = F_{h,k}\hat{u}^h_{k,m|m} + \sigma_{h,k,m+1}, \tag{20}$$

$$\hat{v}_k((m+1)\Delta t) = g\hat{u}_k((m+1)\Delta t) = g\bar{\hat{u}}^h_{k,m+1|m+1} + \hat{\sigma}^o_{k,m}, \tag{21}$$

where the complex observational noise at each time are independent mean zero Gaussian noise with independent real and imaginary parts and

$$\langle \hat{\sigma}^o_{k,m}\hat{\sigma}^o_{k',m}\rangle = \hat{r}^o\delta_{k+k'} = \frac{r^o}{2N+1}\delta_{k+k'}. \tag{22}$$

Note that the variance of physical space observational noise r^o, gets reduced by the factor, $2N+1$, when applied to each of the discrete $2N+1$ Fourier modes. It is established in [22] that the decoupled scalar problems in (20)–(22) are an exact representation of the original filtering problem (Theorem 1) when the observation

points coincide with the discrete mesh points and for general observation points, provide rigorous upper and lower bounds (Theorem 2).

2.5.1. The truth filter

The *truth filter* is the special case of the model in (20) with the choice $F_{h,k} = F_k = \exp((\tilde{p}(\mathrm{i}k) - \gamma(\mathrm{i}k))\Delta t)$ and the complex Gaussian system noise $\sigma_{h,k} = \sigma_k$ with σ_k gotten by exact solution of the stochastic equation in (10) from time step $m\Delta t$ to $(m+1)\Delta t$; thus the real and imaginary parts of σ_k are zero mean independent Gaussian with variance $r_k = E_k(1 - \mathrm{e}^{-2\gamma_k \Delta t})$. The perfect signal $\hat{u}_k(m\Delta t)$, is recovered exactly by the special case of the truth model where the observational noise covariance, r^{o} satisfies $r^{\mathrm{o}} = 0$ so there is no observational noise.

In this section, we have established a simplified theoretical context for addressing all of the central issues for filtering complex systems mentioned in (20)–(22) with rigorous theory guaranteeing a central role for the special complex scalar filtering problems in (2). All of the explicit issues in I and II are studied next in Section 3 for the vastly simpler problems in (2). To simplify notation, the dependence on both spatial wave number, k, and the hat of the discrete Fourier series are omitted in Section 3. This should not confuse the reader.

3. Discrete filtering for the complex scalar test problem

The theory developed in Section 2 and in [22] guarantees that successful filtering for the canonical PDE problem with plentiful observations is based on the number of modes used in the numerical approximation and on the uncoupled filtering problem obtained at each wave number. In this section, we focus on the latter, that is on investigating the filtering of the test problem (1). This complex Langevin equation with $\lambda = -\gamma + \mathrm{i}\omega$ and $\mathrm{Re}(\gamma) > 0$, defines a Gaussian stationary process $u(t)$ with zero mean and steady state variance $\tilde{\sigma}^2/2\gamma$. It is useful to remind the reader that the process has a decorrelation timescale given by $T_{\mathrm{corr}} = \gamma^{-1}$ with oscillation timescale given by ω. By letting $\gamma = \gamma(\mathrm{i}k)$ and $\mathrm{i}\omega = p(\mathrm{i}k)$, this test problem in (1) is the evolution equation of the kth mode of the PDE problem in Eq. (2). Below, filtering for suitable discrete approximations of (1) is studied with plentiful observations at time step $m\Delta t$. Denoting by $\bar{u}_{h,m|m}$ the discrete approximation at these observation times, we assume plentiful observations with the form

$$v_m = g\bar{u}^h_{m|m} + \sigma^{\mathrm{o}}_m, \tag{23}$$

where σ^{o}_m are independent Gaussian random variables with zero mean and covariance r^{o}. As in our discussion in Section 2, here we assume that the observations v_m are generated from the truth model in (1), in other words,

$$v_m = gu(m\Delta t), \tag{24}$$

where $u(t)$ is a solution of (1).

The important fact to keep in mind is the following one. While the reduction to (1) is exact, in order to develop meaningful results from the scalar model as guidelines for filtering stochastic PDE in (2), one needs to study the behavior of the discrete filter model in stiff regimes since $r^{\mathrm{o}},\ \omega,\ \gamma$ and the climatological energy E all vary widely with varying wave number and mesh spacing as explained in Section 2. Thus, the emphasis here for the scalar models is to develop comprehensive theory and benchmark simulations to investigate the interplay between the often large model errors created by imperfect discrete filters and their capability in extracting useful information by judicious strategies for these stiff systems.

In this section, a derivation of different filter models based on standard time finite difference schemes is presented. In each finite difference scheme, we discuss two methods of choosing system noise: either directly from finite difference approximations or by the information criteria. Next, we present explicit filtering theory for the test problem in Eq. (1), by obtaining explicit formulas for the asymptotic Kalman gain and the asymptotic variance. We then compare these off-line testing criteria to computational filter performance for each strategy. In our numerical simulations, we also show the performance of the filters for different ensemble sizes.

3.1. Filtering derivation

The general time discretization of Eq. (1) at the observation times $m\Delta t$ takes the form of a first order autoregressive process (20). In Section 2.5.1, we already encountered one discretization of this filter; *the truth model*, obtained from the discretized form of the solution of (1) yields the following amplification and system noise variance

$$A = \mathrm{e}^{\lambda \Delta t}, \tag{25}$$

$$r = \frac{\tilde{\sigma}^2}{2\gamma}(1 - \mathrm{e}^{-2\gamma \Delta t}). \tag{26}$$

In Appendix A, we derive three examples of a discrete evolution operator with either forward Euler, backward Euler or trapezoidal discretization in time. The corresponding amplification factor for each difference method is given by

$$\begin{aligned} A_h &= (1 + \lambda \Delta t) \quad \text{(Forward Euler)}, \\ A_h &= (1 - \lambda \Delta t)^{-1} \quad \text{(Backward Euler)}, \\ A_h &= \left(1 - \lambda \frac{\Delta t}{2}\right)^{-1} \left(1 + \lambda \frac{\Delta t}{2}\right) \quad \text{(Trapezoidal)} \end{aligned} \tag{27}$$

with the evolution operator F_h in (20) given by A_h^p with observation time $T = p\Delta t$. For simplicity in exposition, and since we are interested here primarily in investigating the performance of finite difference filters with long time step, we use $p = 1$ and thus $F_h = A_h$, and $F = A = \mathrm{e}^{\lambda \Delta t}$, in the rest of this paper, with Δt the observation time. Note that forward Euler is strongly unstable for $|1 + \lambda \Delta t| > 1$. The two implicit schemes are stable independent of the time step chosen. One natural way to add noise to approximate the system in (20) is to simply discretize the stochastic component of the evolution equation in (1). The discretization methods in (27) yield noises $\sigma_{h,m+1}$ with covariances r_h given by

$$\begin{aligned} r_h &= \Delta t \tilde{\sigma}^2 \quad \text{(Forward Euler)}, \\ r_h &= |1 - \lambda \Delta t|^{-2} \tilde{\sigma}^2 \Delta t \quad \text{(Backward Euler)}, \\ r_h &= \left|1 - \lambda \frac{\Delta t}{2}\right|^{-2} \tilde{\sigma}^2 \Delta t \quad \text{(Trapezoidal)}, \end{aligned} \tag{28}$$

where $\tilde{\sigma}$ is the original covariance in Eq. (1) (see Appendix A). While the evolution operator is uniquely determined by the choice of time discretization scheme, the model noise variance is a parameter that can be chosen by other systematic criteria. In order to avoid filter divergence, i.e. the situation when the approximate model predicts an unrealistically small covariance compared with the truth model [1], it is reasonable to choose the system noise variance r_h to minimize this difficulty. In [22], an information theory criterion for choosing the model noise covariance is proposed. Associated with the truth model filter we have an asymptotic Kalman gain $K_\infty(\mathrm{e}^{\lambda \Delta t}, r^{\mathrm{o}}, r)$ and a asymptotic variance r_∞. Immediately below in (30), we show how these quantities can be obtained analytically for the test models in (1). The information criterion chooses the appropriate system noise variance to minimize the relative entropy between the probability density obtained from the truth filter and the time approximate filter, i.e. the least biased asymptotic covariance value in the finite difference filter model consistent with the asymptotic truth model [23,24]. In [22] it is shown that r_h can be uniquely determined according to the following criteria:

(A) For the stable case $|F_h| \leqslant 1$, r_h is the unique noise covariance with $K_\infty(\mathrm{e}^{\lambda \Delta t}, r^{\mathrm{o}}, r) = K_\infty(F_h, r^{\mathrm{o}}, r_h)$, which sets the noise to

$$r_h = r^{\mathrm{o}} \frac{K_\infty(\mathrm{e}^{\lambda \Delta t}, r^{\mathrm{o}}, r)(1 - |F_h|^2(1 - K_\infty(\mathrm{e}^{\lambda \Delta t}, r^{\mathrm{o}}, r)g))}{g(1 - K_\infty(\mathrm{e}^{\lambda \Delta t}, r^{\mathrm{o}}, r)g)}. \tag{29}$$

(B) For the unstable case $|F_h| > 1$: if $1 - |F_h|^{-2} \geqslant K_\infty(e^{\lambda\Delta t}, r^o, r)$ use *zero system noise variance*, $r_h \equiv 0$; if $1 - |F_h|^{-2} < K_\infty(e^{\lambda\Delta t}, r^o, r)$, use the equation in (A) to determine r_h uniquely.

For stable F_h, the information theoretic criterion chooses the system noise to avoid filter divergence. In this paper, this criterion will be contrasted and compared with filtering with system noise chosen from the finite difference discretization in (28). It was pointed out in [22] that for unstable modes with $|F_h| > 1$, zeroing the system noise, r_h, is insignificant when the filter weights more toward the observations. Below we will show that when there is no system noise generated by the information criteria, it might be helpful to add some system noise. However, when the filter is weighted more towards the dynamics the violation of controllability due to zero system noise degrades the filter significantly.

3.2. Limiting filter and error statistics

If the truth filter and the approximate filter in (20), (23) and (24) are both observable and controllable, then they are asymptotically stable [8,1,19]. In the test problems discussed here, both criteria are satisfied if there is non-zero system noise. The asymptotic behavior for the variance r_∞ and Kalman gain K_∞ depend on the observation noise variance, r^o, the evolution operator F or F_h, and the system covariance r or r_h. The quantity $0 \leqslant K_\infty \leqslant 1$ expresses the asymptotic weight that the filter assigns to the observations. When $K_\infty = 0$, the filter trusts the dynamics completely while for $K_\infty = 1$ the observations determine the evolution of the state variable. The explicit formula derived in Appendix B for the asymptotic variance and Kalman gain are given by

$$\begin{aligned} &r_\infty = r^o K_\infty, \\ &K_\infty(\tilde{F}, \tilde{r}^o, \tilde{r}) = K_\infty(\tilde{y}, \tilde{z}) = \frac{1 - \tilde{y} - \tilde{z} + \sqrt{(1 - \tilde{y} - \tilde{z})^2 + 4\tilde{y}}}{2g}, \\ &\tilde{y} = \left(\frac{\tilde{r}g}{\tilde{r}^o}\right)|\tilde{F}|^{-2}, \quad \tilde{z} = |\tilde{F}|^{-2}. \end{aligned} \tag{30}$$

The perfect model limiting filter has variance and Kalman gain

$$r_\infty = r^o K_\infty, \quad K_\infty(e^{\lambda\Delta t}, r^o, r) \tag{31}$$

while the approximate model has the limiting filter values

$$r_{h,\infty} = r^o K_{h,\infty}, \quad K_{h,\infty} = K_\infty(F_h, r^o, r_h). \tag{32}$$

The weight the Kalman gain places on the observations depends on the magnitude of the evolution operator and, the ratio of the model noise r_h and the observation noise r^o. As this ratio increases so does the asymptotic Kalman gain.

The limiting Kalman gain also affects the stability of the asymptotic filter [18,19], which is given by the magnitude of

$$S_h = F_h(1 - K_{h,\infty}g) \tag{33}$$

Theoretically, the filter is stable provided $|S_h| < 1$ and clearly this is always satisfied for $|F_h| < 1$ since $0 < |K_\infty g| < 1$. However, marginal stability of the discrete filter i.e., $|F_h| \cong 1$ can lead to practical filter instability when the Kalman gain satisfies $K_\infty \cong 0$ and weights toward the dynamics. In [22], it is showed that the mean model errors for the discrete filter in filtering the truth signal are also controlled by the magnitude of the stability function; in particular, for $|S_h| \cong 1$, the mean model errors decay very slowly. Naively, one might guess that only stability governs practical filter performance; however, many examples below and in Section 5 show that this is not the case for the discrete filters with finite difference noise in (28). The subtle reason for this is practical controllability. While theoretically any non-zero system noise variance $r_h \neq 0$ guarantees controllability, practically, a finite lower threshold for this noise $r_h \geqslant r_* > 0$ is needed for stable difference schemes, the information criteria automatically augment the noise in order to guarantee practical controllability.

Finally, for unstable schemes with zero model noise $r_h = 0$, it is proved in [22] that:

$$1 - K_\infty g = \frac{1}{|F_h|^2}. \tag{34}$$

This implies that the stability of the filter behaves as

$$|S_h| = \frac{1}{|F_h|}, \tag{35}$$

thus as the unstable evolution operator increases in magnitude the asymptotic stability of the filter increases.

3.3. Filter performance on the test problem

In this section, we discuss the performance of various discretization strategies on filtering the scalar stochastic differential equation test problem (1). Our goal is to utilize the off-line testing criteria developed earlier to understand the filtered solutions. In particular, we are interested to see the role of the ensemble size in each filtering strategy and the behavior of the true model and the approximate filtering strategies as we vary the observation time Δt, the frequency ω as well as the observations noise r^o; also the role of the information criteria compared to time discretized system noises.

In each numerical simulation shown below, we generate a true trajectory by evolving a randomly chosen field u_0 with (1) for $L = 200$ steps with time step Δt. With the values of the parameters $\Delta t,\ \omega,\ \gamma,\ L$ utilized below, this is always a time where the mean truth signal relaxes to the climatological state. We simulate each observation by simply adding uncorrelated Gaussian random variables with mean 0 and variance r^o to the true solution at each observation time Δt in accordance with (25) and (26). Without loss of generality we choose $g = 1$. For each assimilation, we simulate with ensembles of size $K = 1$, 10, 50, 100, 250, and 500 where each initial ensemble member is a random state $u^k_{0|0}$, where $k = 1, \ldots, K$.

Each filtering problem has three parameters: observation time Δt, observation noise r^o, and system noise variance r_h as mentioned in (I.B), (I.C) and (I.D) from Section 1. To fully understand the scalar filter performance, we need to consider the following regimes:

- $\Delta t < T_{\text{corr}};\ \Delta t = T_{\text{corr}};\ \Delta t > T_{\text{corr}}$: Here, the correlation time $T_{\text{corr}} = \gamma^{-1}$ reflects the time when the deterministic part of the system becomes uncorrelated. We are particularly interested to see how the filter performs when the model dynamics is still relevant ($\Delta t < T_{\text{corr}}$), when Δt is equal to the autocorrelation time T_{corr}, and when it is simply a white noise ($\Delta t > T_{\text{corr}}$).
- $r^o < E;\ r^o = E;\ r^o > E$: The steady state energy $E = \sigma^2/2\gamma$ is basically the system noise r (see Eq. (26)) at large observation time $\Delta t \to \infty$. We consider these regimes because we want to understand the filter performance when the observation noise is very accurate while the system noise is huge ($r^o \ll E$), when both are comparable ($r^o = E$), and when the observation is not as accurate as the model $r^o \gg E$.
- $\omega \ll \gamma;\ \omega = \gamma;\ \omega \gg \gamma$: These regimes reflect several type of oscillators: an over-damped oscillator ($\omega \ll \gamma$) to a weakly damped oscillator ($\omega \gg \gamma$). We shall see later that in the latter stiff case, the filter performance is very sensitive to the discretization strategies.

Fortunately, we can fully understand the filter performance in all specified regimes by analyzing only the following three regimes (other regimes gives qualitatively equivalent results):

A. Observation time Δt varies with $T_{\text{corr}} = 10$, $\omega = \gamma$, and $r^o = E$.
B. Observation noise r^o varies with $E = 5$, $\omega = \gamma$, and $\Delta t = T_{\text{corr}}$.
C. The frequency ω varies with $10^{-2} \leqslant \omega \leqslant 10^2$ with damping $\gamma = 1$, $r^o = E$, and $\Delta t = T_{\text{corr}}$.

Regime A. In Fig. 1 (first row), we plot the amplitude of F_h as a function of the observation time Δt: as the observation time is increased, we see that forward Euler (first column) becomes unstable and its amplitude deviates away from the amplitude of the truth operator. The amplitude of the other schemes, backward Euler (second column) and trapezoidal (third column), are relatively similar to the amplitude of the true filter. In the

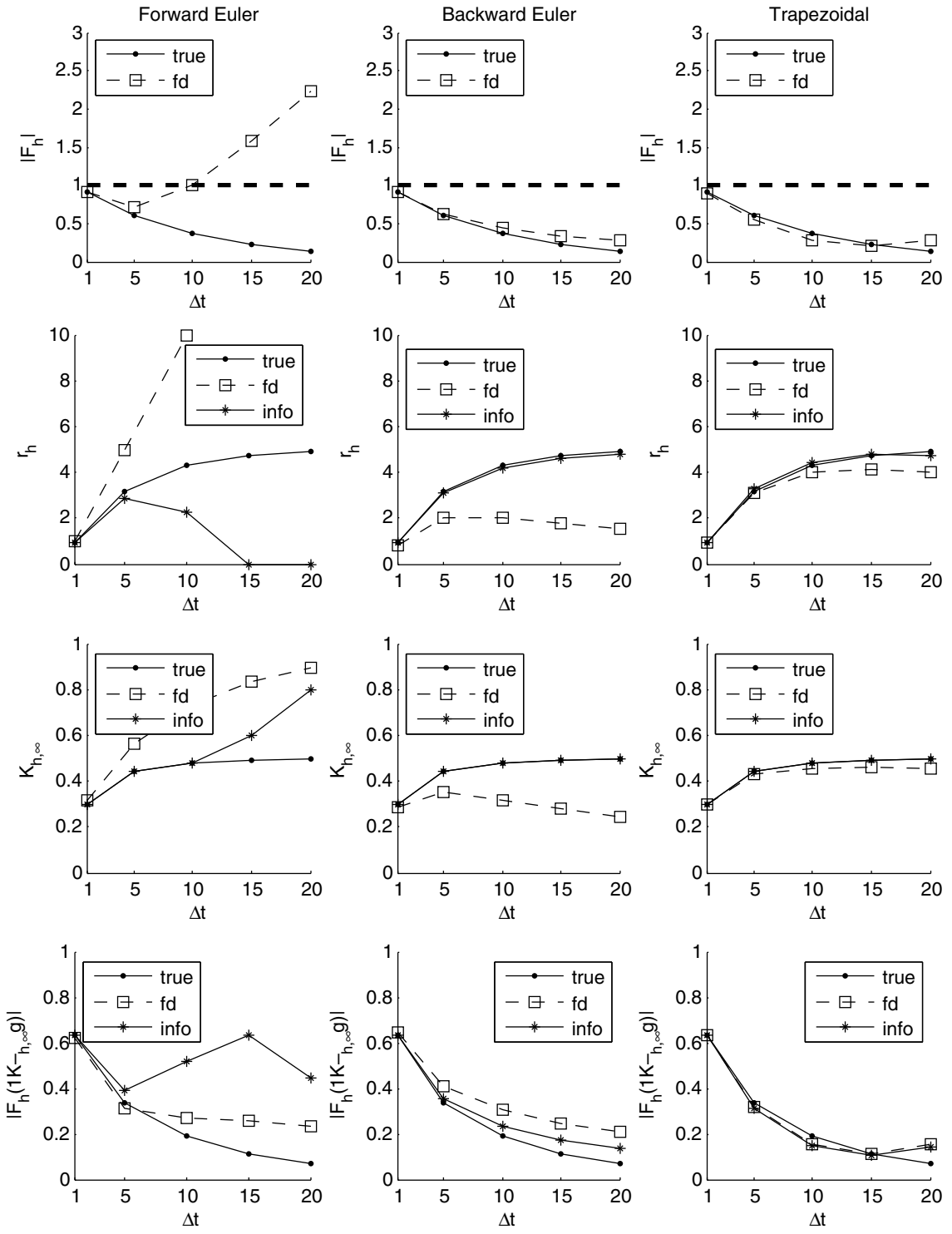

Fig. 1. Off-line testing for Regime A: observations time Δt varies with $T_{\text{corr}} = 10$, $\omega = \gamma$, and $r^o = E$. The panels in the first column depict simulations with forward Euler, second column with backward Euler, and third column with trapezoidal. The first row depicts $|F_h|$, the second row for r_h, the third for $k_{h,\infty}$, and the fourth for stability $|F_h(1 - K_{h,\infty}g)|$. In each panel, 'true' indicates the true filter, 'fd' denotes the finite difference approximate filter, and 'info' denotes the approximate filter with information criterion noise variance.

second row of Fig. 1, we plot the system noise r_h as a function of Δt: in the stable discretized schemes, the information criterion chooses the system noise r_h to be closer to the true system noise r. In the unstable discretized scheme, on the other hand, the system noise r_h is zero when $|F_h| > 1$ (see Section 3.1B). In the third row of Fig. 1, we plot the limiting Kalman gain $K_{h,\infty}$ as function of Δt: when $\Delta t < T_{\text{corr}}$, the Kalman gains of both the true filter and the time discretized filter increase as functions of Δt and they eventually saturate after the system becomes more or less a white noise, i.e. $\Delta t > T_{\text{corr}}$. When the information criteria are used, the Kalman gain of the time discretized filter is similar to that of the true model, except when the system noise is chosen to be zero. In the fourth row of Fig. 1, the stability or $|F_h(1 - K_{h,\infty}g)|$ is plotted as a function of time: as we see, all strategies yield stable filtering since $|F_h(1 - K_{h,\infty}g)| < 1$.

In Fig. 2 (first row), we show the (RMS) average analysis error as a function of Δt for ensemble of size $K = 500$: The average analysis error is defined as follows:

$$\text{error} = \sqrt{\frac{1}{L}\sum_{m=1}^{L}(\bar{u}_{m|m} - u_m)^2}, \tag{36}$$

where $\bar{u}_{m|m}$ denotes the analysis ensemble average. The information criteria improve the time discretized filters when stable schemes are applied since practical controllability is satisfied. For the unstable scheme, the filter degrades as the observation time Δt is increased. In this unstable scenario, the information criteria do not improve the filter because the choice of zero system noise violates the controllability criteria while it also alters the filter to weight more toward the dynamics. In Fig. 2 (second row), the (RMS) average analysis error is plotted as a function of ensemble size K at $\Delta t = 10$: Notice that the larger the ensemble size is, the lower the errors are. However, the filter performs quite well even with a single realization. Fig. 2 (third row) shows the average ensemble error variance as a function of ensemble size: The average ensemble error variance is computed as follows:

$$\text{error variance} = \text{Var}(u^k_{m|m} - u_m), \tag{37}$$

where the variance is taken over each ensemble member and time. This quantity measures the variation of errors of an ensemble member compared to other member. From our results, we see that as we increase the ensemble size, the less varied the error variance is.

Regime B. In Fig. 3 (first row), for reference values we plot the amplitude of F_h: all three discretized schemes are stable although the forward Euler is only marginally stable. Fig. 3 (second row) shows the system noise r_h as a function of r^o. In Fig. 3 (third row), we plot the Kalman gain as a function r^o: our results suggest that the larger the observation noise is, the more the filter trusts the dynamics ($K_{h,\infty}$ decreases). When information criteria are used, the Kalman gain of the true filter is the same as that of any discretized filters since they are all stable schemes. In Fig. 3 (fourth row), the stability is plotted as a function of r^o: we see that quantity $|F_h(1 - K_{h,\infty}g)|$ increases as a function of r^o, which is obvious, since $K_{h,\infty}$ decreases as explained earlier.

In Fig. 4 (first row), we show the (RMS) average analysis error as a function of r^o for ensemble of size $K = 500$: as predicted earlier, the information criterion reduces the error in all time discretized schemes. In Fig. 4 (second row), we plot the (RMS) average analysis error as a function of ensemble size K at $r^o = 5$: as in Fig. 2, we find that the small ensemble size (even with only a single realization) performs quite well. In Fig. 4 (third row), the average ensemble error variance is plotted as a function of ensemble size: from our results, we see that as we increase the ensemble size, the less varied the error variance is.

Regime C. In Fig. 5 (first row), we plot the amplitude of F_h as a function of frequency ω: here, as ω increases, the forward Euler (first column) becomes highly unstable, the backward Euler (second column) is over-damped while the trapezoidal (third column) is right on the instability boundary. Fig. 5 (second row) plots system noise r_h as a function of ω: one very interesting fact is that the time discretized system noises r_h of both backward Euler and trapezoidal schemes tend to zero as ω increases, which imply the practical violation of controllability. When the information criterion is used in the stable schemes, the system noise variance r_h are chosen automatically to be non-zero and the controllability is satisfied. In the unstable schemes, however, the information criteria set $r_h = 0$ and thus the controllability is not satisfied. Fig. 5 (third row) shows the Kalman gain as a function of ω: as ω increases, the unstable scheme tends to trust the observations

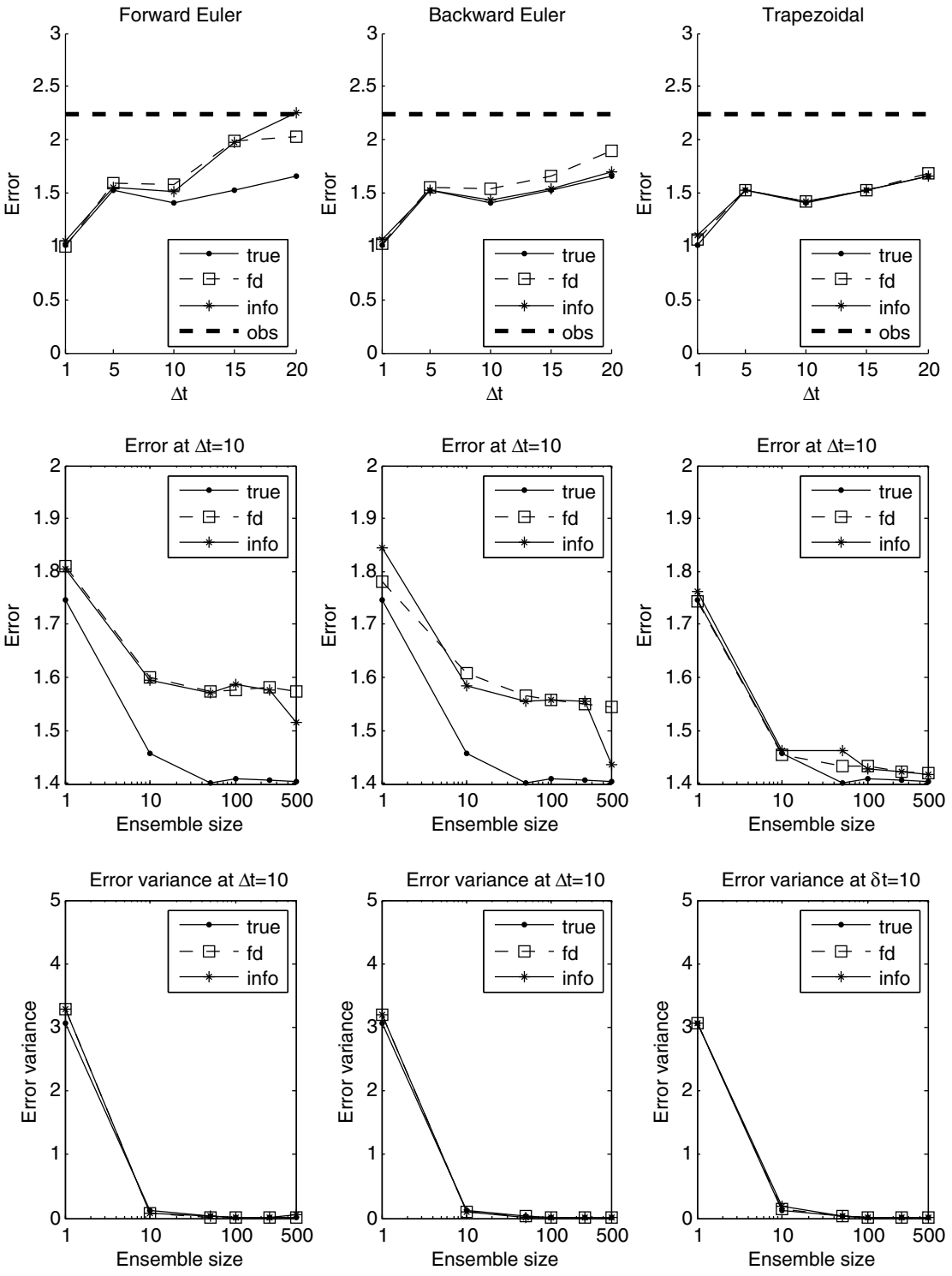

Fig. 2. Filtering solution for Regime A: observations time Δt varies with $T_{\text{corr}} = 10$, $\omega = \gamma$, and $r^o = E$. The panels in the first column depict simulations with forward Euler, second column with backward Euler, and third column with trapezoidal. The first row depicts error as function of observation time, the second and third rows plot the error and the ensemble error variance (both for $\Delta t = 10$), consecutively, as functions of ensemble size. In each panel, 'true' indicates the true filter, 'fd' denotes the finite difference approximate filter, 'info' denotes the approximate filter with information criterion noise variance, and 'obs' denotes the observation error.

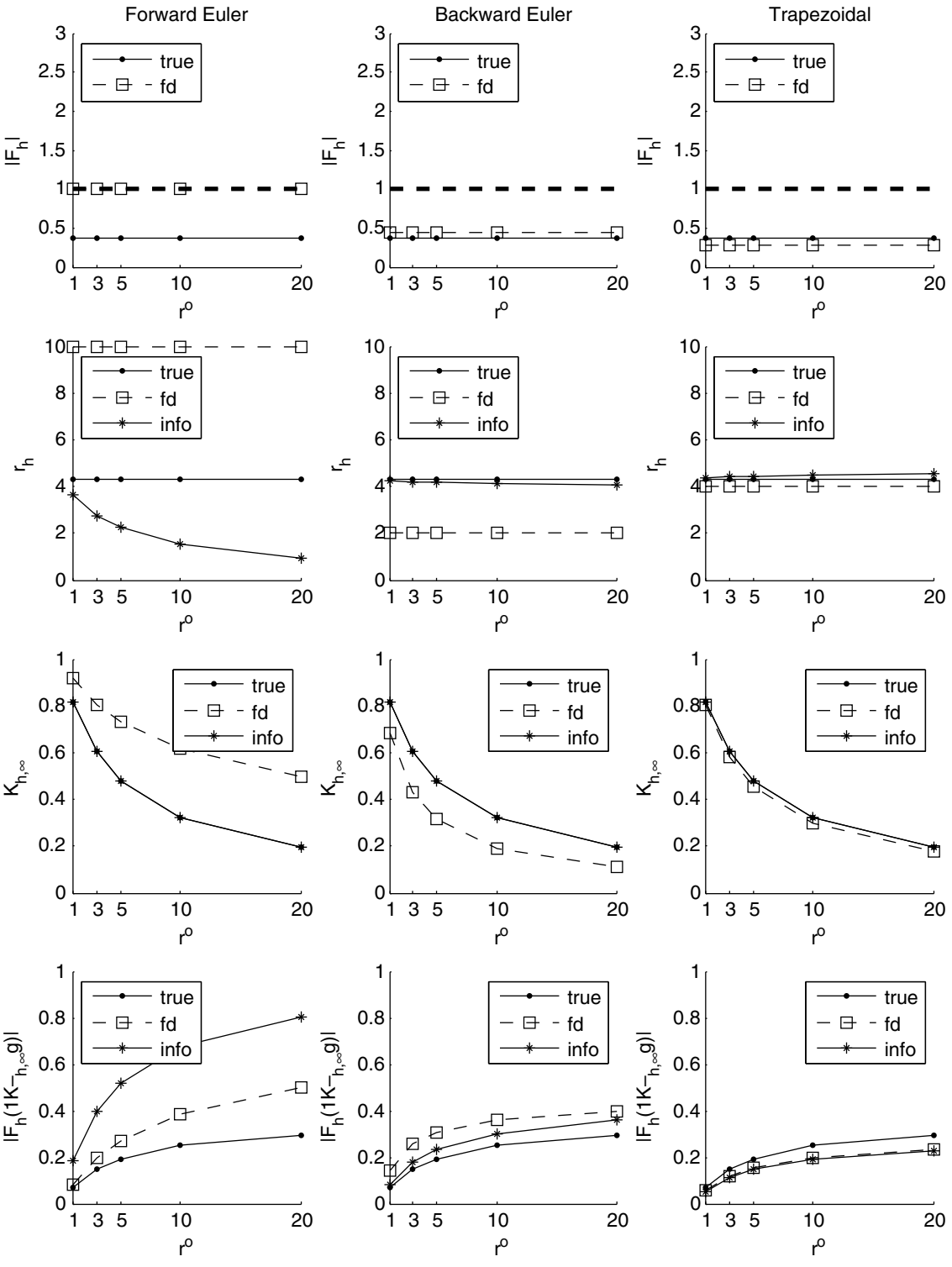

Fig. 3. Off-line testing for Regime B: observations noise r^o varies with $E = 5$, $\omega = \gamma$, and $\Delta t = T_{\text{corr}}$. The panels in the first column depict simulations with forward Euler, second column with backward Euler, and third column with trapezoidal. The first row depicts $|F_h|$, the second row for r_h, the third for $k_{h,\infty}$, and the fourth for stability $|F_h(1 - K_{h,\infty}g)|$. In each panel, 'true' indicates the true filter, 'fd' denotes the finite difference approximate filter, and 'info' denotes the approximate filter with information criterion noise variance.

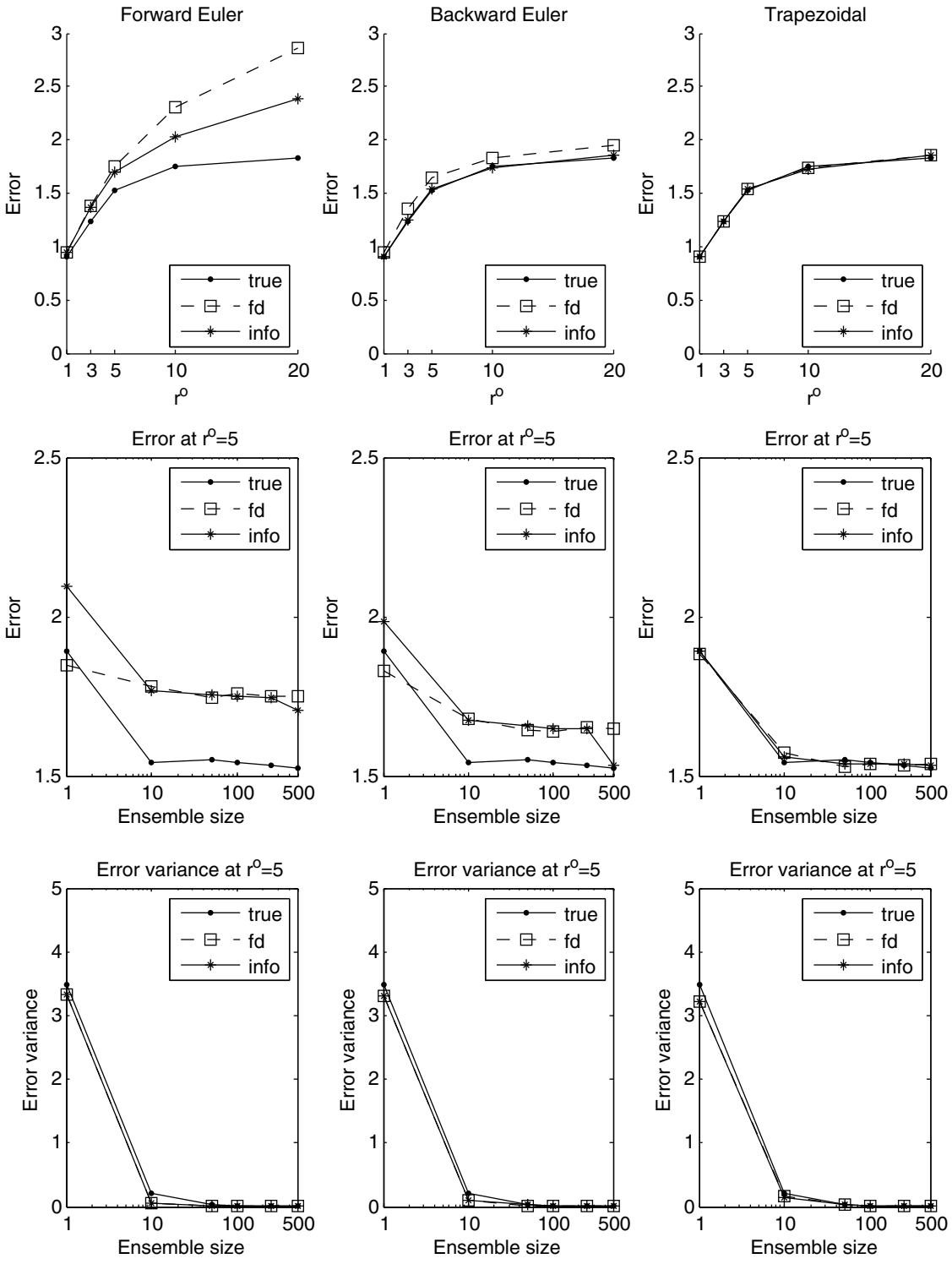

Fig. 4. Filtering solution for Regime B: observations noise r^{o} varies with $E = 5$, $\omega = \gamma$, and $\Delta t = T_{\mathrm{corr}}$. The panels in the first column depict simulations with forward Euler, second column with backward Euler, and third column with trapezoidal. The first row depicts error as function of observation noise variance, the second and third rows plot the error and the ensemble error variance (both for $r^{o} = 5$), consecutively, as functions of ensemble size. In each panel, 'true' indicates the true filter, 'fd' denotes the finite difference approximate filter, 'info' denotes the approximate filter with information criterion noise variance, and 'obs' denotes the observation error.

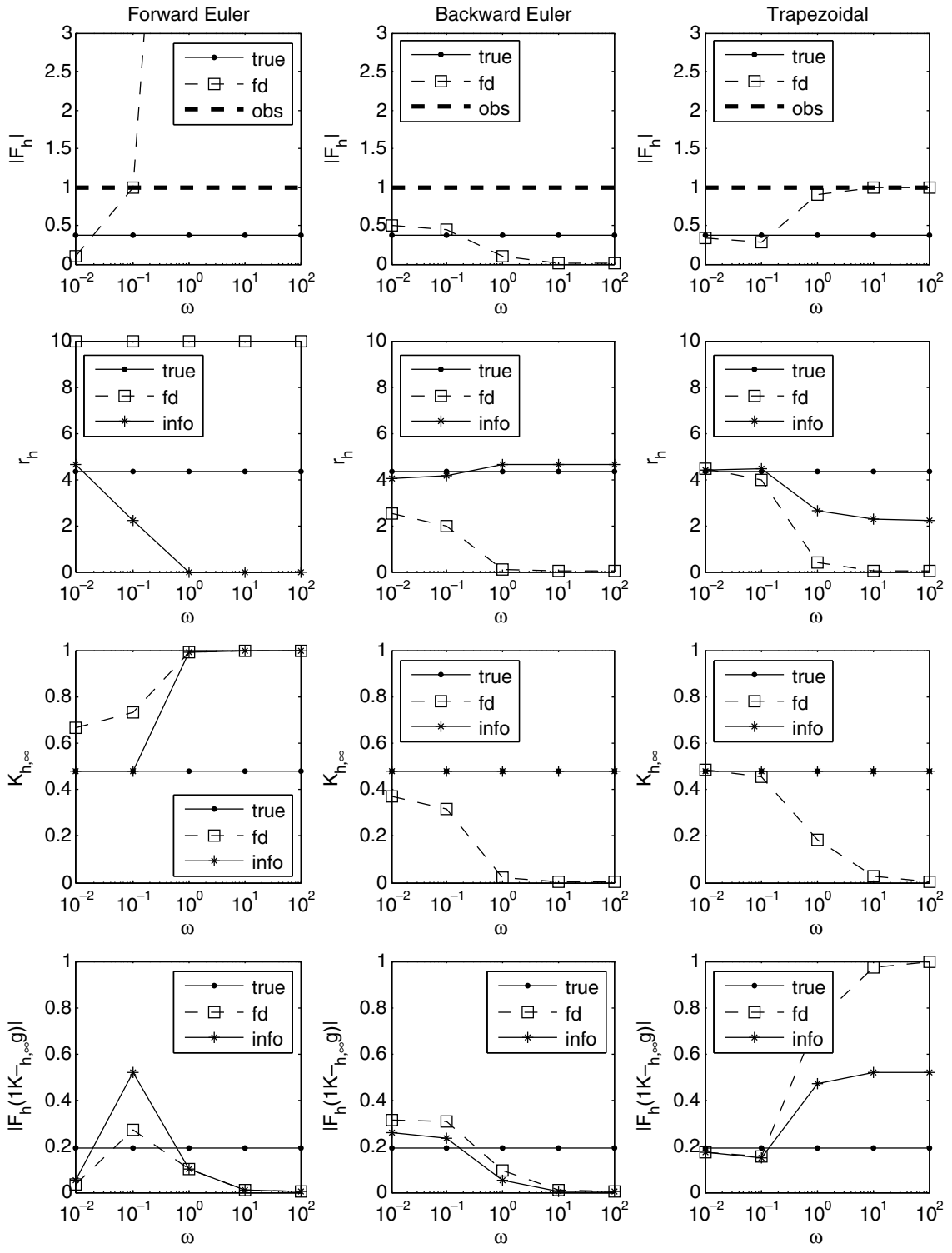

Fig. 5. Off-line testing for Regime C: the frequency ω varies with $10^{-2} \leqslant \omega \leqslant 10^{2}$ with damping $\gamma = 1$, $r^{o} = E$, and $\Delta t = T_{\text{corr}}$. The panels in the first column depict simulations with forward Euler, second column with backward Euler, and third column with trapezoidal. The first row depicts $|F_h|$, the second row for r_h, the third for $k_{h,\infty}$, and the fourth for stability $|F_h(1 - K_{h,\infty}g)|$. In each panel, 'true' indicates the true filter, 'fd' denotes the finite difference approximate filter, and 'info' denotes the approximate filter with information criterion noise variance.

more while the stable schemes with standard discretized noise trust the highly inaccurate dynamics more. With the information criteria and the stable schemes, of course, the Kalman gain is automatically the same as the truth model. In Fig. 5 (fourth row), we plot the stability as a function of ω: here, we see an indication that the trapezoidal scheme may not perform as well as all the other strategies when the time discretized system noise is used since $|F_h(1 - K_{h,\infty}g)| \approx 1$.

In Fig. 6 (first row), we plot the (RMS) average analysis error as a function of ω for ensemble of size $K = 500$: we see that the unstable scheme fully trusts the observations as suggested by the off-line test criteria. The backward Euler scheme with augmented noise from the information criterion performs almost as well as the true filter robustly in ω despite the fact that it strongly over damps the dynamics; on the other hand, backward Euler with the standard discretization noise performs poorly for large ω.

When the trapezoidal scheme is used, the most interesting regime occurs for large ω when $|F_h| \approx 1$ and the system noise $r_h \approx 0$ (see the second row in Fig. 5), where the information criteria makes a significant difference. Here, we see that for time discretized noise (28), the limiting Kalman gain is close to zero and the violation of the controllability condition together with a large weight in the dynamics, again, yields unstable filtering for large ω. In this case, the information criteria improve the filter by choosing larger system noise. In Figs. 6 (second and fourth rows), we show the (RMS) average analysis errors as functions of ensemble size K for $\omega = 1$ and 100, respectively. When $\omega = 1$, forward Euler simply trusts the observations regardless of how the system noises are chosen. Both the backward Euler and the trapezoidal scheme produce comparable errors as the true filter when the information criteria are used. In an oscillatory case (e.g., $\omega = 100$), the information criteria reduce the errors significantly for the trapezoidal scheme even with only one realization. In both cases ($\omega = 1$ and 100), the performances for small ensemble members (or even with one realization) are comparable to those with $K = 500$. Figs. 6 (third and fifth rows) show the average ensemble error variances as functions of ensemble size for $\omega = 1$ and 100: in both cases, we find that the ensemble error variances of all three discretized schemes are close to that of the truth when the information criteria are used. As the ensemble size is increased, the backward Euler scheme with time discretized system noise has the lowest ensemble error variance, even lower than the truth model (this is also true for forward Euler scheme with $\omega = 100$). In general, the ensemble error variances seems to be significantly reduced when large ensemble size is used, except for the trapezoidal scheme with time discretized noise when $\omega = 100$. In this case, the information criteria reduces the ensemble error variance significantly.

Summary. From our results, we conclude that when we use big time steps (large Δt), both the backward Euler and trapezoidal schemes are the better approximate schemes, especially when the information criteria are used. The trapezoidal scheme with discretized system noise does not work well when $|F_h| \approx 1$ with nearly zero system noise variance r_h. This case occurs when the system is highly oscillating with weak damping ($\omega \gg \gamma$), which happens in many applications (Section 5). The main difficulties are practical violation of controllability and filter stability $|F_h(1 - K_{h,\infty}g)| \approx 1$ which reflects nearly unstable filtering. In this situation, we show that the information criterion is a possible remedy. When the information criterion is used so that practical controllability is guaranteed, the best scheme for practical filter performance among all the stable approximate schemes can be predicted by theoretical guidelines by looking at the asymptotic stability factor $|F_h(1 - K_{h,\infty}g)|$; a smaller magnitude for this quantity implies a more stable scheme.

When using an explicit scheme (such as forward Euler), it is an unstable scheme when the damping is weak compared to the oscillation frequency or when the observation time is larger than the correlation time. In this case, the filter will be weighted toward the observations and hence the filtered solutions are more or less equal to simply trusting the observations. The information criterion does not improve the skill here and even has the potential to put more weight on the (unphysical) dynamics although the practical significance is small (see Section 5 below). However, if the filter is fully weighted toward the observations, the violation of the controllability is irrelevant in our numerical experiments.

We see that, in general, the filter is better when the ensemble size is larger. However, the filter improvement is not significant in the true filter and in all time discretized filters with increased ensemble size. We found that even with a single realization, we can get a reasonably good filter solution. Furthermore, the ensemble error variances decrease as functions of ensemble size, and the errors produced by one ensemble member are not much different than those produced by other members.

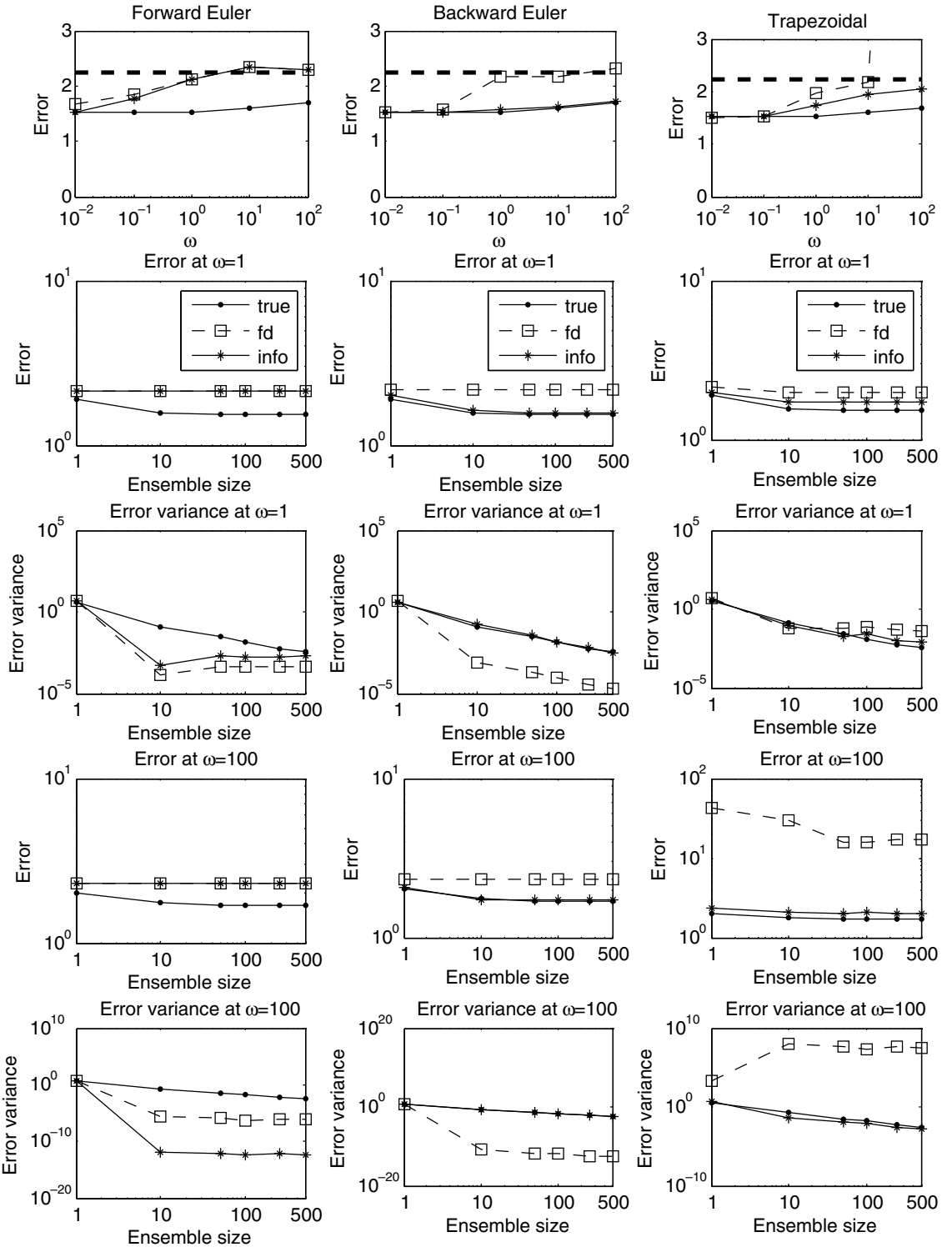

Fig. 6. Filtering solution for Regime C: the frequency ω varies with $10^{-2} \leqslant \omega \leqslant 10^{2}$ with damping $\gamma = 1$, $r^{o} = E$, and $\Delta t = T_{\text{corr}}$. The panels in the first column depict simulations with forward Euler, second column with backward Euler, and third column with trapezoidal. The first row depicts error as function of frequency, the second and third rows plot the error and the ensemble error variance (both for $\omega = 1$), consecutively, as functions of ensemble size. The fourth and fifth rows plot the error and the ensemble error variance (both for $\omega = 100$), consecutively, as functions of ensemble size. In each panel, 'true' indicates the true filter, 'fd' denotes the finite difference approximate filter, 'info' denotes the approximate filter with information criterion noise variance, and 'obs' denotes the observation error.

4. Theoretical guidelines for filter performance under mesh refinement for turbulent signals

In many situations such as local regional numerical weather prediction over populated areas in developed countries, there are plentiful observations available for many successively refined discrete meshes. Also the nature of the turbulent spectra in the true dynamics might be white noise in space, reflecting physical processes such as moist convection over a few kilometers, or another steeper power law in the upper troposphere, reflecting gravity wave activity. Thus, a very interesting and relevant question for operational models is the following one: If plentiful observations are available, what is gained in filter performance by increasing the resolution of the operational model? How does this depend on the nature of the turbulent spectrum? Here, we provide theoretical guidelines for these important practical issues by answering the above question for filtering the general turbulent signals from the model in (2) with the truth model itself.

We begin with the case of selective damping, Thus, a basic issue of practical interest in filtering a system like that in (2) is the following one: given the system in (2) with the scale selective damping $\gamma(ik)$, i.e. damping with increasing wave number, and with the energy spectrum E_k, what is the effect of increasing the number of grid points with plentiful observations on the Kalman gain? Under what circumstances should most of the weight be given to the observations alone for filtering at high spatial wave numbers and when should most of the weight be given to the dynamics alone at high spatial wave numbers? As mentioned above, these are important issues to consider as guidelines for mesh refinement of the filtering problem in turbulent systems. Here we answer the above question completely for the asymptotic Kalman gain in the truth model by processing the explicit formula in (30) in elementary analytic fashion. The answer involves an interesting quantitative interplay among the factors listed in (I.B), (I.C) and (I.D). Assume that the selective scale damping $\gamma(ik)$ and the energy spectrum $E(k) = E_k$ are both given by power laws,

$$\gamma(ik) = \gamma_0|k|^{\alpha}, \quad 0 < \alpha < +\infty, \tag{38}$$

$$E(k) = E_0|k|^{-\beta}, \quad 0 \leqslant \beta < +\infty. \tag{39}$$

Note that for the moment we require that $\alpha > 0$ in (38) so that there is actually increased selective scale damping as $|k|$ increases. Recall from (12) that in the statistical steady state, the decorrelation time of the kth wave number is given by

$$T_{\mathrm{corr}}(k) = (\gamma(ik))^{-1} = \gamma_0^{-1}|k|^{-\alpha}. \tag{40}$$

The filtering properties are determined by the observation time, Δt, and the observational noise variance, r^{o}. Consider the explicit formula for the asymptotic Kalman gain in (30) for the truth model; for $2N + 1$ discrete modes, we have

$$\tilde{z} = |F_k|^{-2} = \mathrm{e}^{2\gamma_0|k|^{\alpha}\Delta t}, \tag{41}$$

$$\tilde{z} = |F_k|^{-2} = \mathrm{e}^{2(T_{\mathrm{corr}}(k))^{-1}\Delta t} \tag{42}$$

while

$$\tilde{y} = A(k)\tilde{z} \tag{43}$$

with $A(k)$ the ratio of system noise to observational noise at wave number k; by using (11) so that for the truth filter, $r_k = E(k)[1 - \mathrm{e}^{-2T_{\mathrm{corr}}^{-1}(k)\Delta t}]$, we have

$$A(k) = \frac{r_k(2N+1)}{r^{o}} = (2N+1)\frac{E_0}{r^{o}}|k|^{-\beta}[1 - \mathrm{e}^{-2T_{\mathrm{corr}}^{-1}(k)\Delta t}] \tag{44}$$

for $0 \leqslant |k| \leqslant N$. Note that in (44) we utilized the fact that the observational noise per mode decreases the observation noise by $2N + 1$. Our intuition for filtering the truth model suggests that if $A(k) \to 0$ as $|k|$ increases, there is more observational noise compared with decreasing system noise so that the filter should trust the dynamics alone; on the other hand, for $A(k) \to \infty$ the observation noise is relatively small and we should trust the additional observations alone in the filtering problem. Next we evaluate (44) asymptotically for

$N \to \infty$ for the wave numbers $N/2 \leqslant |k| \leqslant N$ (in the argument below, the lower bound with 1/2 can be any factor smaller than one). From (44), we have the dichotomy

A. For $1 - \beta < 0$

$$\max_{\frac{N}{2} \leqslant |k| \leqslant N} A(k) \leqslant \frac{E_0}{r^o} 2^{\beta}(2N+1)N^{-\beta} \sim N^{1-\beta} \to 0, \tag{45}$$

as $N \to \infty$.

B. For $1 - \beta > 0$

$$\min_{\frac{N}{2} \leqslant |k| \leqslant N} A(k) \geqslant \frac{E_0}{r^o} 2^{-\beta}(2N+1)N^{-\beta} \sim N^{1-\beta} \to \infty, \tag{46}$$

as $N \to \infty$.

This is a quantitative estimate for the intuition mentioned earlier. This intuition is confirmed by the following.

Theorem 1. *For the selective damping with $\alpha > 0$ and the energy spectrum (39), there are two different universal regimes of behavior for the filtering problem with plentiful observations for high wave numbers $\frac{N}{2} \leqslant |k| \leqslant N$ as $N \to \infty$.*

(A) For $\beta > 1$ the asymptotic Kalman gain matrix tends to zero uniformly for $\frac{N}{2} \leqslant |k| \leqslant N$ as $N \to \infty$. Thus, in this regime even with plentiful observations, one can trust the dynamics alone on the large wave numbers, with a refined mesh.

(B) For $\beta < 1$ the asymptotic Kalman gain matrix tends to one uniformly for $\frac{N}{2} \leqslant |k| \leqslant N$ as $N \to \infty$. Thus, in this regime with plentiful observations, one should primarily trust the observations on the large wave numbers in the filtering problem with a refined mesh.

Thus, for turbulent signals with $\beta < 1$, increasing mesh resolution only weights toward the additional observations while for $\beta > 1$, increased resolution improves the dynamics but the additional observations are not significant.

As with all asymptotic statements, the validity with a given discrete mesh depends crucially on whether the quantitative numbers in (45) and (46) are, respectively, large or small and also the growth factor

$$\min_{\frac{N}{2} \leqslant |k| \leqslant N} e^{2\frac{\Delta t}{T_{\mathrm{corr}}(k)}} = e^{2^{1-\alpha}\gamma_0 |N|^{\alpha}\Delta t}. \tag{47}$$

This is the only result in this paper which depends on the spatial dimension; for d-spatial dimensions, the proposition remains valid but the dichotomy $1 - \beta < 0$ or $1 - \beta > 0$ is replaced by $d - \beta < 0$ or $d - \beta > 0$. This arises simply because for plentiful observations in d-dimension we have the factor $(2N+1)^d$ in (22), (45) and (46) replacing $2N + 1$. We state this in the following corollary.

Corollary 1. *In d-space dimensions, Theorem 1 remains valid with $d - \beta$ replacing $1 - \beta$.*

With the above background from (42)–(47), the proof of the proposition is elementary through evaluating the explicit Kalman gain formula in (30). In the situation from (A) with $\beta > 1$, we have from (45) and (47) the uniform asymptotic behavior for $\frac{N}{2} \leqslant |k| \leqslant N$

$$\tilde{z} \to \infty, \quad \tilde{y} = \epsilon\tilde{z}, \quad \epsilon(N) \to 0 \tag{48}$$

with $\epsilon(N)$ determined from (45) so that the explicit Kalman gain satisfies

$$K(F_k, \tilde{y}, \tilde{z}) \equiv \mathcal{O}(\epsilon(N)) \to 0 \tag{49}$$

as $N \to \infty$ uniformly for $\frac{N}{2} \leqslant |k| \leqslant N$. In the situation from (B) with $\beta < 1$, we have from (46) and (47), the uniform asymptotic behavior for $\frac{N}{2} \leqslant |k| \leqslant N$

$$\tilde{y} \to \infty, \quad \tilde{z} = \epsilon\tilde{y}, \quad \epsilon(N) \to 0 \tag{50}$$

with $\epsilon(N)$ determined from (46) (B) so that the explicit Kalman gain matrix satisfies

$$K(F_k, \tilde{y}, \tilde{z}) \equiv 1 - \mathcal{O}(\epsilon(N)) \to 1, \tag{51}$$

uniformly for $\frac{N}{2} \leqslant |k| \leqslant N$.

Is Theorem 1 still valid with $\alpha = 0$ in (38) so that there is uniform damping at all wave numbers? In this case, $\tilde{z} = e^{2\gamma_0 \Delta t} > 1$ remains a bounded fixed constant but the dichotomy in (A) and (B) remains valid. For $\beta > 1$, we have $\tilde{y} = \epsilon\tilde{z}$; since $\tilde{z}$ is bounded, the asymptotic Kalman gain formula from (30) is given by

$$1 - \tilde{z} + |1 - \tilde{z}| = 0, \tag{52}$$

so part (A) of the Proposition is valid. For $\beta < 1$, we have $\tilde{z} = \epsilon\tilde{y}$ so the identical argument for part (B) is valid. Thus, we have the following corollary.

Corollary 2. *Proposition* 1 *remains valid with uniform damping at all wave numbers.*

In Fig. 7, we plot the asymptotic Kalman gain as a function of wave number k for different resolutions N for two spectra with the two sets of parameter values in the dissipative advection equation utilized in Sections 5.1 and 5.2 below. Theorem 1 is confirmed as one sees (in the first row of Fig. 7) that the asymptotic Kalman

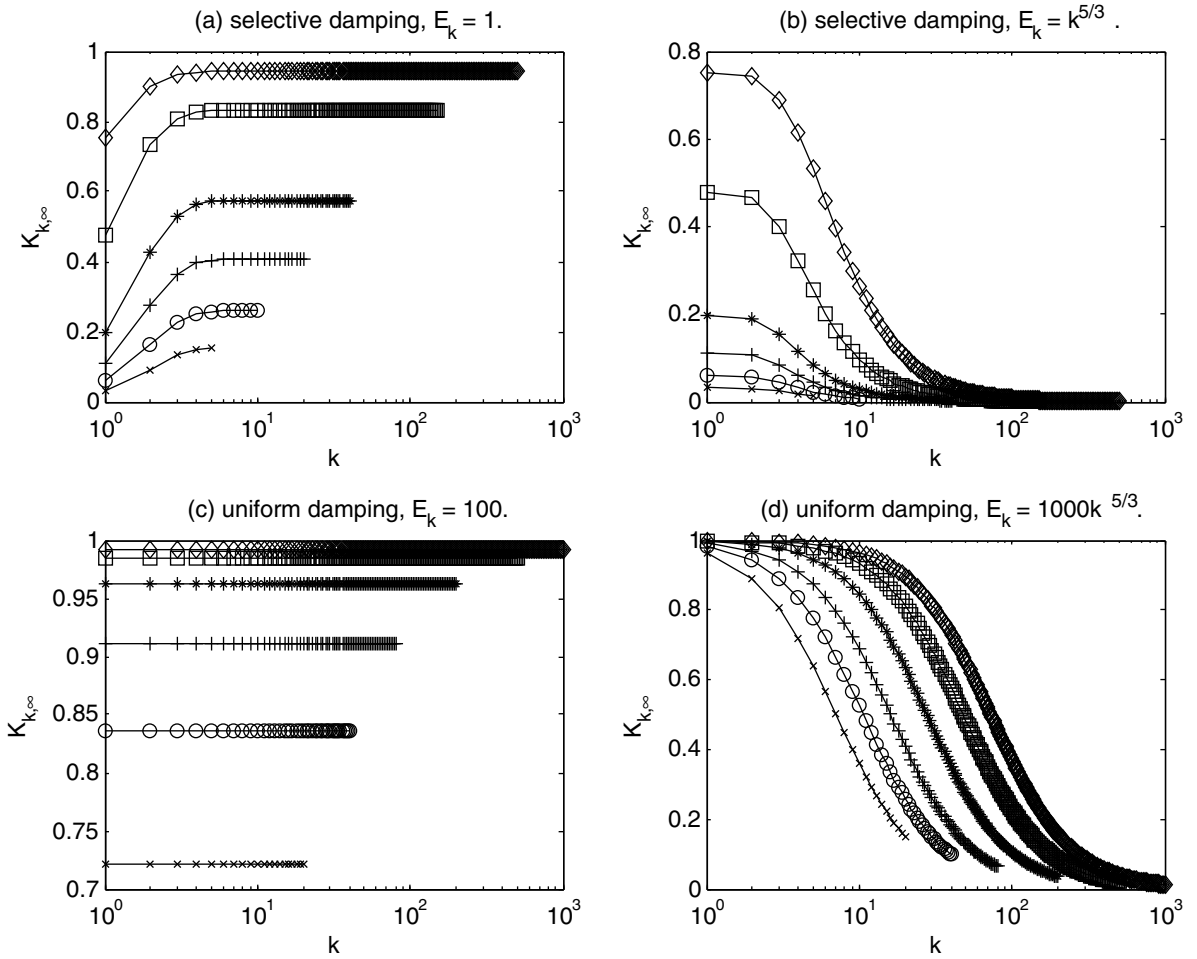

Fig. 7. Kalman gains as function of wave numbers. The first row is for selective damped signal with $r^o = 60$ and $\Delta t = 1$; in this regime, the model resolution is varied with wave numbers $N = 5$, 10, 20, 40, 150 and 500. The second row is for uniform damped signal with $r^o = 1000$ and $\Delta t = 50$; in this regime, the model resolution is varied with wave numbers $N = 20$, 40, 80, 200, 500 and 1000. The parameter values for selective decay and uniform damping are the same as in Sections 5.1 and 5.2.

gain converges to one for large enough wave numbers when $\beta = 0$ (see the first column of Fig. 7) and converges to zero for $\beta = 5/3$ (see the second column of Fig. 7). Similar trends are found for uniform damping case (see second row) which confirm Corollary 2. In this last case, there are large prefactors in front of the power law so much larger value of N are needed to realize the theory.

5. Discrete filtering for the stochastically forced dissipative advection equation

Here we study filter performance for discretizations of the stochastically forced dissipative advection equation in (7). One goal is to illustrate how rough turbulent signals generated by (7) can be filtered successfully by suitable discretization strategies with significant model error which respect the theoretical and computational guidelines established in Section 3 for the scalar test problem; these guidelines also explain inaccurate filtering with strongly stable filters. A second goal is to illustrate and compare these Fourier filters with the extended Kalman filter in physical space with the same approximation which generates errors which do not respect the Fourier diagonal covariance structure. Finally, the third goal is to point toward the potential practical use of Fourier diagonal filters with significant model errors (here these model errors are generated through spatio-temporal discretization) which are guided by the mathematical criteria in Sections 2 and 3 as alternatives to ETKF to overcome the "curse of ensemble size"; tests on a family of forty dimensional non-linear systems with chaotic instability are developed elsewhere by two of the authors [17].

5.1. Filtering strongly turbulent signals with uniform damping and infrequent observation

Here we consider discrete filter performance for the equation in (5) with uniform damping and without selective decay as a stringent test case. Thus, we generate truth signals for filtering (7) with $d > 0$ but $\mu = 0$. We utilize parameter values $c = 1,\ d = 0.01$, and $\Delta t = 50 = T_{\text{corr}}/2$ where $T_{\text{corr}} = 1/d = 100$ is the decorrelation time at each wave number. For this uniformly damped setting, the amplification factors at each wave numbers, F_k, satisfy $|F_k| = \mathrm{e}^{-d\Delta t} = \mathrm{e}^{-1/2} < 1$ so there is strong asymptotic stability in the perfect model filter in this regime. We consider truth signals generated from two extremely turbulent spectra, an equipartition spectrum with $E_k = 100$ and a $-5/3$ spectrum with $E_k = 1000k^{-5/3}$.

Next we consider a family of upwind discretizations as discrete filters. We consider a simple upwinding scheme

$$\frac{\partial}{\partial x}u(x_j,\cdot) \sim \frac{u_j - u_{j-1}}{h}. \tag{53}$$

Utilizing the discrete Fourier transform defined in (14), we can rewrite the discrete spatial derivative as

$$\frac{\partial}{\partial x}u(x_j,\cdot) \to \frac{1-\mathrm{e}^{-\mathrm{i}kh}}{h}\hat{u}_k. \tag{54}$$

As in Section 3 for the scalar test problem, we approximate the time derivative with the same three different time discretized schemes (27) with

$$\lambda_k = -c\frac{1-\mathrm{e}^{-\mathrm{i}kh}}{h} - d. \tag{55}$$

For implicit Euler and trapezoidal schemes, the condition $|F_{h,k}| < 1$ is always satisfied for any resolution. In the trapezoidal schemes, however, almost every mode is marginally stable so that $|F_{h,k}| \cong 1$ for most modes. For the unstable forward Euler scheme, the amplitude satisfies $|F_{h,k}| > 1$ for the coarse mesh $N = 20$ and as we increase N, the magnitude of $F_{h,k}$ increases sharply.

5.1.1. Off-line test criteria

Here we utilize the theoretical off-line test criteria developed in Section 3 as a guideline for filter performance with these rough spectra. Fig. 8 shows these off-line criteria for the backward Euler (panels a–d) and trapezoidal methods (panels e and f).

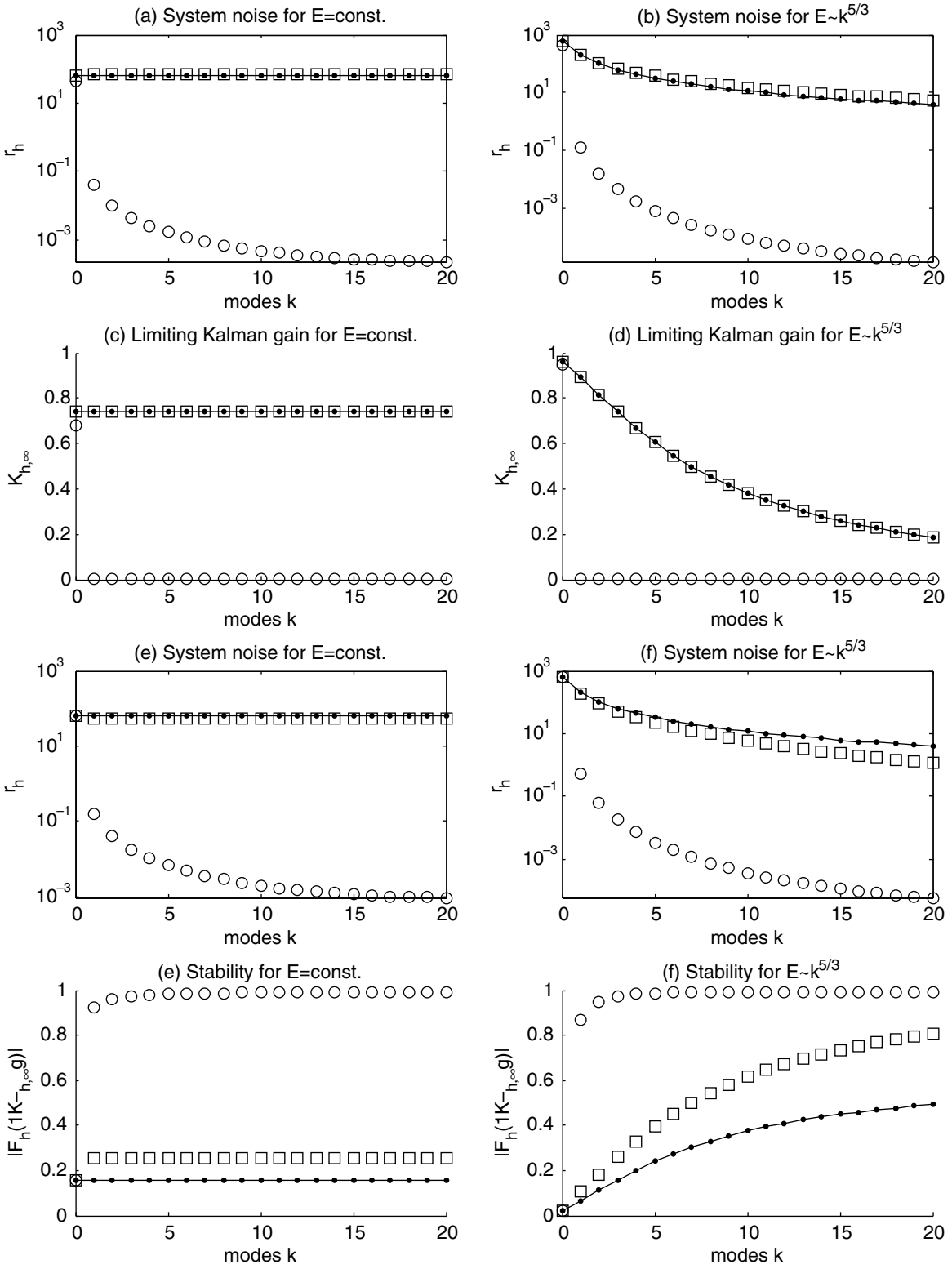

Fig. 8. Off-line testing of the backward Euler (a–d) and trapezoidal method (e–h) with $\Delta t = T_{\text{corr}}/2$. The first column shows the off-line testing with white noise spectrum $E_k = 100$, the second column shows testing with smooth spectrum $E_k = 1000k^{-5/3}$. In each subfigure, we show the truth model (solid line with dots), approximate scheme with time discretized noise (circle) and with information criteria (square).

In our off-line testing, we predict that the information criteria will improve filtering for both backward Euler and trapezoidal schemes by giving practical controllability at all wave numbers through augmented system noise. In this regime, the Kalman gain $K_{k,\infty}$ is close to 1 for the constant energy spectrum but it is much less than 1 for smooth spectrum $E_k \sim k^{-\frac{5}{3}}$ (see panels c and d in Fig. 8). Thus, our theory predicts that the true filter is not much better than simply trusting the observations for the constant spectrum while it is better for the $-5/3$ spectrum. When we employ the information criteria, the better filter among the two schemes is predicted by the asymptotic stability amplitude $|F_k(1 - K_{k,\infty}g)|$. In our simulations, we see that the backward Euler is a better scheme since $|F_k(1 - K_{k,\infty}g)| \approx 0$ for all k (not shown) while these factors for the trapezoidal method increase as functions of wave number (panel g,h). Without the information criteria, both the backward Euler and trapezoidal methods with standard finite difference noise discretization violate practical controllability criteria despite having filter stability (see circles in panels a, b, e and f); thus the off-line test criteria predict poor filter performance. As in Section 3, for unstable Euler, the off-line criteria predict that this method just trusts the observations.

5.1.2. Numerical simulations of filter performance

We check the actual filter performance with the prediction of the off-line testing shown earlier. In particular, we also compare results from filtering in the Fourier domain with filtering in real space. The filtering in the Fourier domain consists of an independent scalar filter (as in Section 3) for each Fourier mode. For the rest of this paper, we call this filter the Fourier Domain Kalman Filter (FDKF). The real domain filter that we choose for comparison is the Ensemble Transform Kalman Filter (ETKF) of Bishop et al. [5]. The reason why we choose this ensemble filter is because it is easily implemented [16] and for large ensemble size. Note that for the numerical scheme experiments with ETKF, the wave equation in (5) is integrated using an upwind scheme from (53) in real domain.

For our numerical simulations, we generate the true trajectory by evolving an initial state that looks like a Gaussian hump, denoted as $\{\hat{u}_{k,0},\ k = 1, \ldots, 2N + 1\}$ in Fourier space, with the standard exact large time step discretization of (10) as in Section 3 for $L = 100$ steps with time step $\Delta t = 50 = T_{\text{corr}}/2$. We simulate each observation by simply adding uncorrelated Gaussian random variables with mean 0 and variance $\hat{r}^{\text{o}} = r^{\text{o}}/(2N + 1)$ to the true solution at each observation time Δt. In the physical space, this reflects observations with variance r^{o}. As in the previous off-line testing, the observation noise is chosen to be $r^{\text{o}} = 1000$. We initiate each numerical simulation with randomly chosen initial states $u_{j,m|m}$ (or $\hat{u}_{k,m|m}$ in Fourier space).

In Figs. 9 and 10, we plot the RMS errors as functions of time for FDKF for ensemble of size $K = 100$ (first row), the RMS errors as functions of time for ETKF, also for $K = 100$ (second row), the RMS errors as functions of ensemble size for FDKF (third row), and the ensemble error variances as functions of ensemble size for FDKF (fourth row). For the panels in the first two rows, the RMS errors are averaged over space only. For the panels in the third row, the RMS errors are averaged over space and time (from $L_{\text{o}} = 50$ to $L = 100$). The ensemble error variance is defined as

$$\text{error variance} = \text{Var}(u^k_{j,m|m} - u_{j,m}), \tag{56}$$

where the variance is taken over each ensemble member, space and time (also between $L_{\text{o}} = 50$ and $L = 100$).

From our simulations, we first notice that the forward Euler simply trusts the observations, as suggested by the off-line testing, regardless of the spectra (this case mimics Regime B in Section 3 where $r^{\text{o}} < E$, see Figs. 3 and 4). We also notice similar behavior as in the scalar case for Regime C with $\omega > \gamma$ (see Section 3), that is, the backward Euler with time discretized noise $r_{h,k}$ over damps the system and hence the prior forecast states deviate too far away from the noisy observations. For the trapezoidal scheme with time discretized noise, the filter is marginally unstable since $|F_{h,k}(1 - K_{h,k,\infty}g)| \approx 1$ (see Fig. 8). The information criteria improve the stability of both filters as predicted by the off-line testing with statistical accuracy nearly comparable to that of the truth filter with all the different spectra and resolutions. Thus, the off-line test criteria successfully predict all features of the filter performance.

Our numerical simulations suggest that both schemes (FDKF and ETKF) gives comparable performance in term of errors (e.g. compare also the filtered solutions of FDKF and ETKF in Figs. 11 and 12, respectively). The ETKF is not robust in the sense that for an ensemble size of less than 50, the filter diverges (results not shown), while the FDKF with smaller ensemble size (even with only one realization) is performing as well as

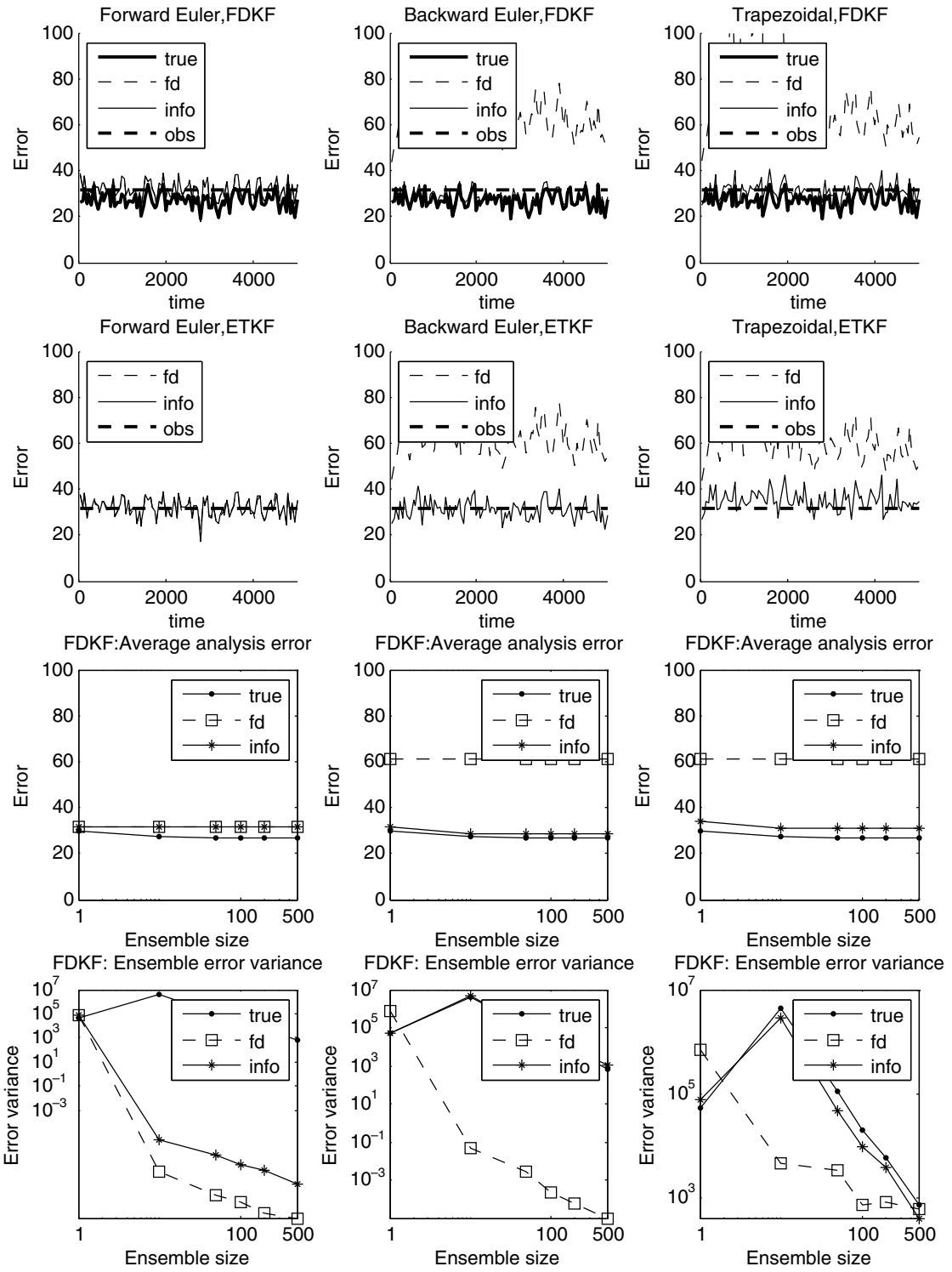

Fig. 9. Uniform damping, $E_k = 100$ with $\Delta t = 50$ and $N = 20$: RMS errors as functions of time for FDKF with ensemble size $K = 100$ (first row), second row for ETKF also with $K = 100$, RMS errors as functions of ensemble size for FDKF (third row), and ensemble error variances as functions of ensemble size for FDKF (fourth row). The panels in the first column depict simulations with forward Euler, second column with backward Euler, and third column with trapezoidal. In each panel, 'true' indicates the true filter, 'fd' denotes the finite difference approximate filter, 'info' denotes the approximate filter with information criterion noise variance, and 'obs' denotes the observation error.

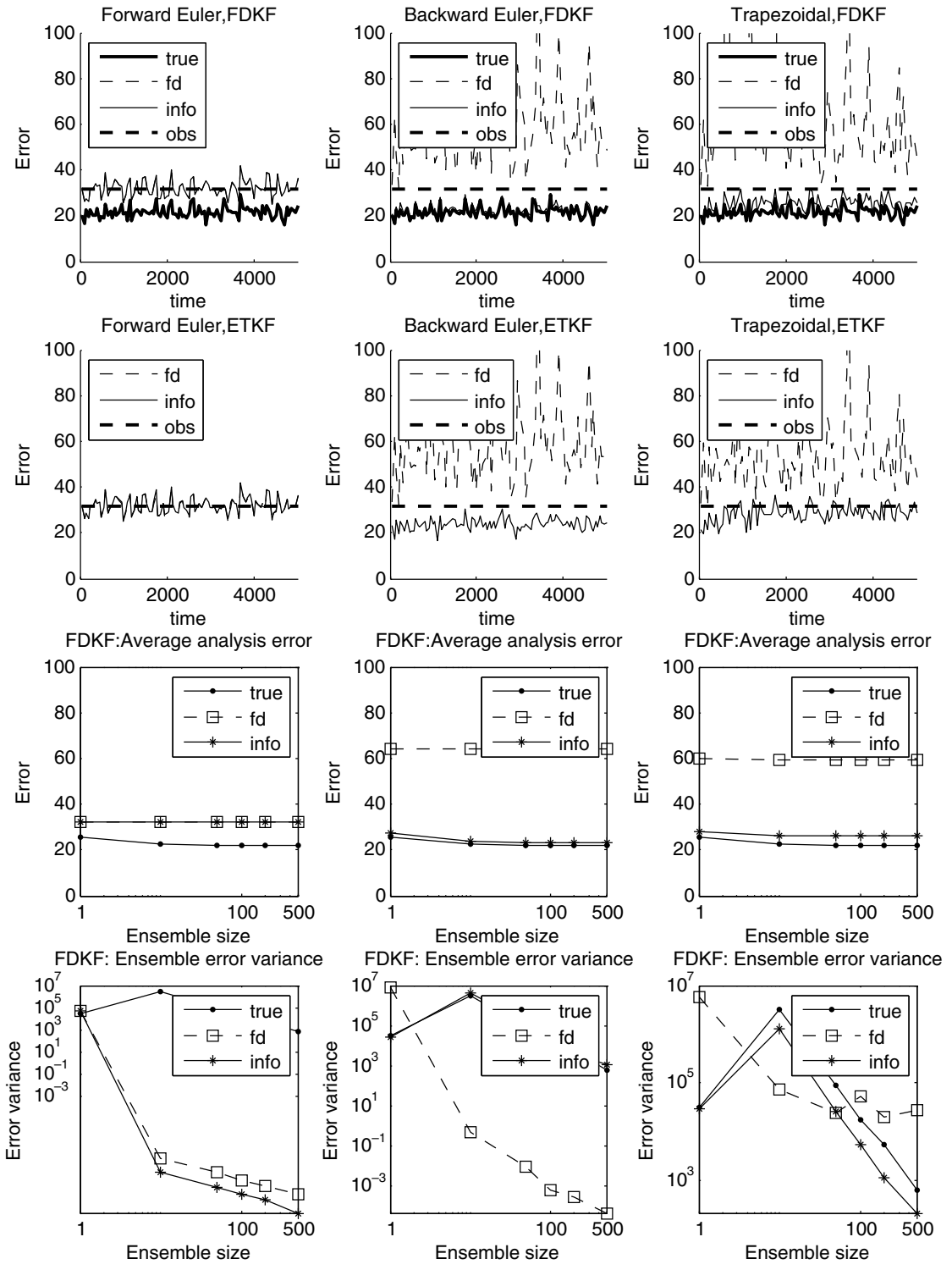

Fig. 10. Uniform damping, $E_k \sim k^{-5/3}$ with $\Delta t = 50$ and $N = 20$: RMS errors as functions of time for FDKF with ensemble size $K = 100$ (first row), second row for ETKF also with $K = 100$, RMS errors as functions of ensemble size for FDKF (third row), and ensemble error variances as functions of ensemble size for FDKF (fourth row). The panels in the first column depict simulations with forward Euler, second column with backward Euler, and third column with trapezoidal. In each panel, 'true' indicates the true filter, 'fd' denotes the finite difference approximate filter, 'info' denotes the approximate filter with information criterion noise variance, and 'obs' denotes the observation error.

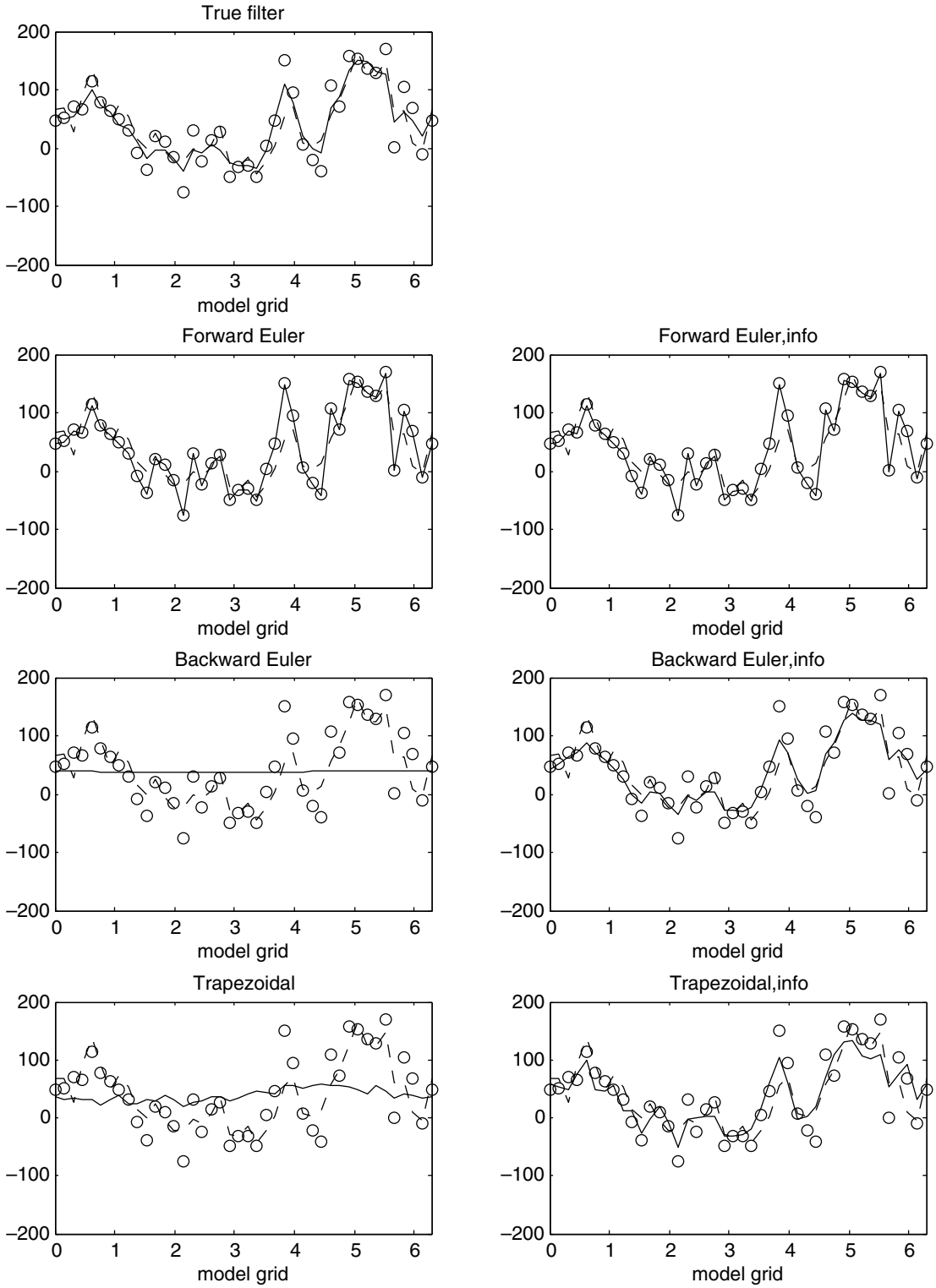

Fig. 11. Snapshots of filtered solutions with FDKF as functions of model space after $L = 100$ assimilation cycles for uniform damping, $E_k \sim k^{-5/3}$ with $\Delta t = 50$ and $N = 20$. In each panel, we show the filtered solution (solid), the true signal (dashes), and the observations (circle).

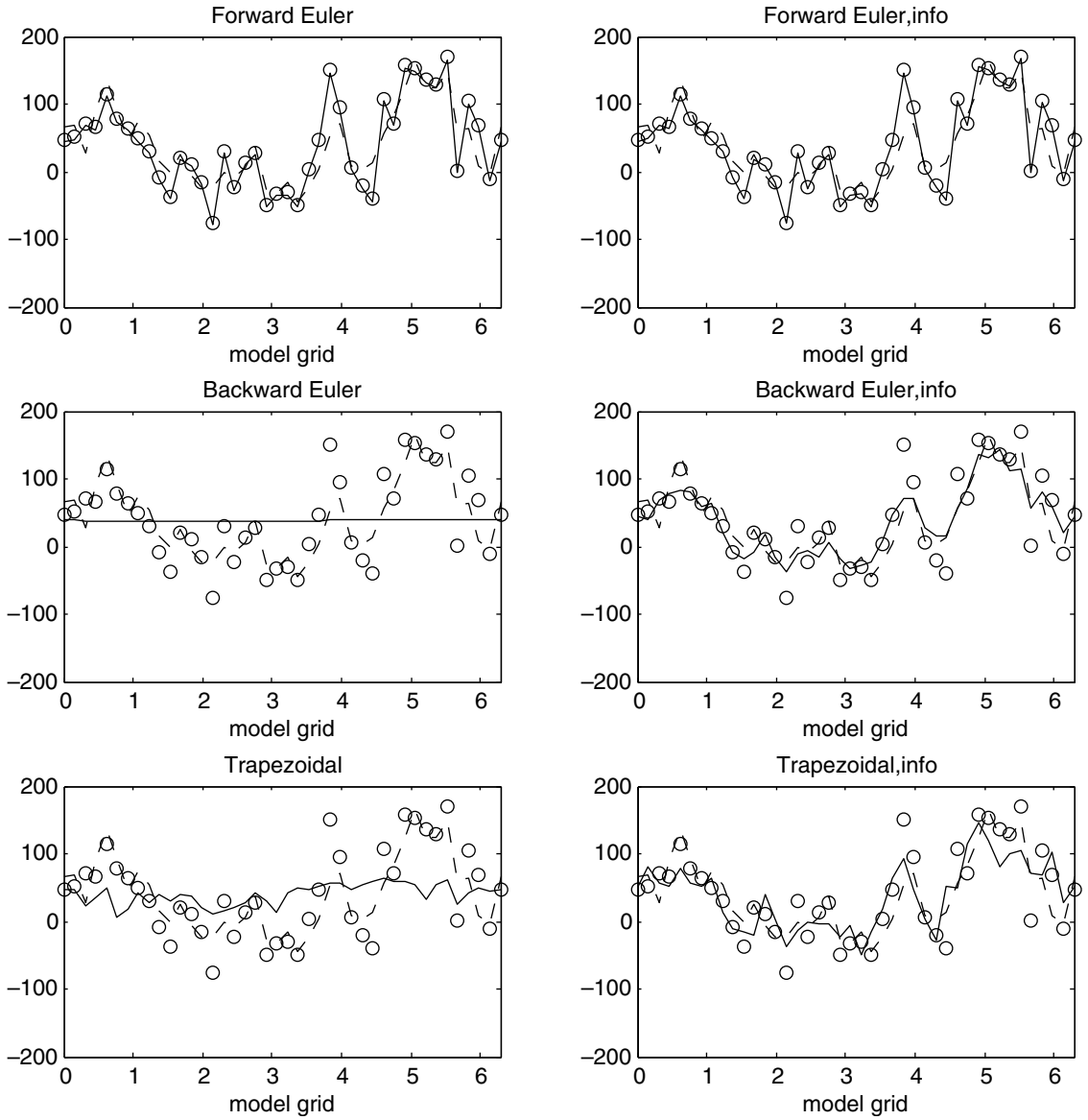

Fig. 12. Snapshots of filtered solutions with ETKF as functions of model space after $L = 100$ assimilation cycles for uniform damping, $E_k \sim k^{-5/3}$ with $\Delta t = 50$ and $N = 20$. In each panel, we show the filtered solution (solid), the true signal (dashes), and the observations (circle).

that with larger ensemble size in terms of RMS errors. In Figs. 9 and 10 (fourth row), we notice that the forward Euler has the smallest ensemble error variance since as in Section 3, the Kalman gain is 1 and all ensemble members trust the observations. In contrast, the over-damped backward Euler with time discretized noise has Kalman gain 0 (see Fig. 8). In this situation, all ensemble members trust the dynamics and therefore it is obvious that the ensemble error variances are smaller than the those of the truth model. From the comparisons between two schemes, the FDKF is preferable since it is a less expensive scheme (the ETKF involves computations on matrices of size $K \times N$ while FDKF are composed of independent scalar filters), it is independent of tunable parameters (recall that ETKF depends on variance inflation), and more importantly, its filtered solutions are as skillful as ETKFs with FDKF neglects the correlation between different wavenumbers, whereas ETKF accounts these correlations.

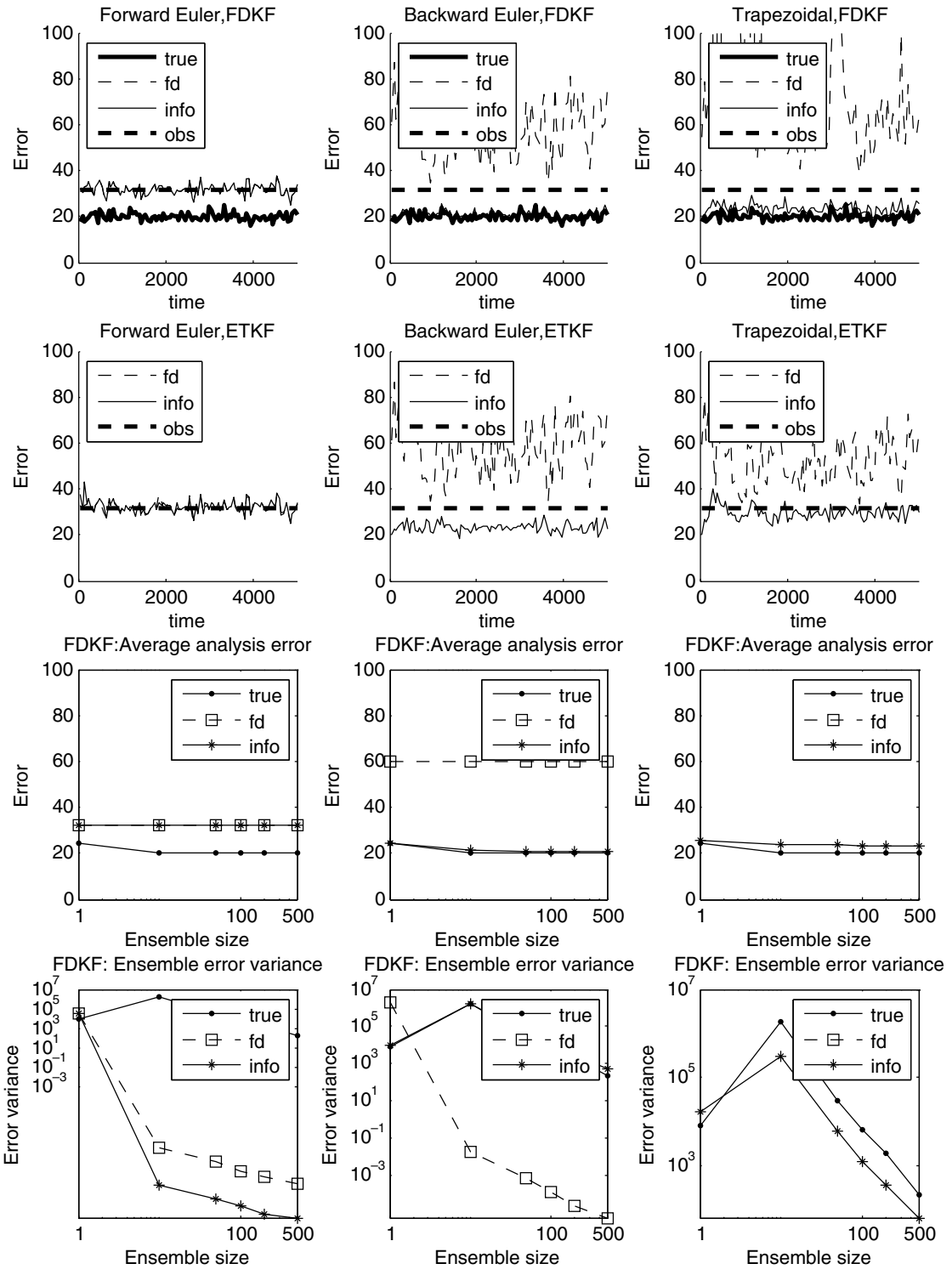

Fig. 13. Uniform damping, $E_k \sim k^{-5/3}$ with $\Delta t = 50$ and $N = 40$: RMS errors as functions of time for FDKF with ensemble size $K = 100$ (first row), second row for ETKF also with $K = 100$, RMS errors as functions of ensemble size for FDKF (third row), and ensemble error variances as functions of ensemble size for FDKF (fourth row). The panels in the first column depict simulations with forward Euler, second column with backward Euler, and third column with trapezoidal. In each panel, 'true' indicates the true filter, 'fd' denotes the finite difference approximate filter, 'info' denotes the approximate filter with information criterion noise variance, and 'obs' denotes the observation error.

Now we check the filter performance for variations of model resolution $N = 20$, 40 and 80 (see Figs. 13 and 14). In a standard extended Kalman filter, the higher the model resolution, the larger size the error covariance matrix. In principle, the basic idea of an ensemble Kalman filter, including ETKF is to approximate the error covariance matrix by the sample covariance from many realizations. Thus, the higher model resolution in ETKF requires larger ensemble size (see [15, Chapter 3]). In our ETKF simulations, we use $K = 200$ for $N = 80$ but $K = 100$ for $N = 20$ and 40. On the other hand, we see that the FDKF (including those of the time discretized schemes) are not sensitive at all to the variations of resolutions, especially small ensemble size performs as well as large ensemble size (see Figs. 13 and 14).

From our numerical tests, we learn that the stable implicit schemes (backward Euler and trapezoidal) are the best filters provided that their system noises are chosen to satisfy the information criteria. Between these two schemes, the backward Euler performs as well as the true filter. All these tendencies are fully predicted by off-line testing. In particular the superior performance of backward Euler over the trapezoidal method for the decaying spectrum can be traced to the decaying stability factor for large spatial wave numbers in Fig. 8. The spectacular failure of the filtering performance of backward Euler and trapezoidal with standard time discretized noise is again predicted by the off-line Kalman gain which predicts full reliance on the highly inaccurate discrete dynamics without observation input.

Practically, both implicit filters in Fourier space are computationally inexpensive with such a giant time step and thus one can afford large ensemble size. However, we also found that a large ensemble size is not necessary. As we have seen earlier, even one realization is often acceptable. Moreover, the scalar Fourier domain filter is not sensitive to the variations of model resolution and independent of tunable parameters. The omission of accounting for the correlation between different modes in our scalar filtering is also found to be insignificant. In contrast, an ensemble Kalman filter that mimics the extended Kalman filter suggests that more realizations are needed when the model resolution is increased for filter convergence; the ensemble Kalman filter also depends on the variance inflation coefficient.

5.2. Filtering turbulent signals with selective damping

Here we consider (5) with $d = 0$, and $\mu > 0$, so that there is selective damping at larger spatial wave numbers. In Section 5.1, we tested the filter in a highly energetic turbulent field, while here, we will perform the filtering with a less energetic system. That is, we choose the equipartition spectrum $E_k = 1$ and less turbulent energy spectrum $E_k = k^{-5/3}$. In our testing, we fix the advective coefficient $c = 1$, the diffusivity coefficient $\mu = 0.1$ (the same parameter as in [14]). The observation noise variance $r^o = 60$ (in Fourier space, this corresponds to noise with variance $\hat{r}^o = 60/(2N+1)$, which is larger than the energy spectra E_k for all k when $N = 20$) and the observation time $\Delta t = 1$. Here, the observation time Δt is chosen to be larger than the correlation times of all wave numbers except for wave numbers 1, 2 and 3.

In this section, we will focus on the filter performance for resolution $N = 20$. To illustrate the generality of the theoretical guidelines, for this simple advection-diffusion equation we consider a simple central difference approximation. In Fourier space, the spatial derivative is given by

$$\frac{\partial}{\partial x}u(x_j,\cdot) \approx \frac{u_{j+1}-u_{j-1}}{2h} \to \frac{\mathrm{i}}{h}\sin(kh), \tag{57}$$

$$\frac{\partial^2}{\partial x^2}u(x_j,\cdot) \approx \frac{u_{j+1}-2u_j+u_{j-1}}{h^2} \to -\frac{4}{h^2}\sin^2\left(\frac{kh}{2}\right). \tag{58}$$

Thus, the approximate scheme for the advection–diffusion equation has

$$\lambda_k = -\frac{4\mu}{h^2}\sin^2\left(\frac{kh}{2}\right) - \mathrm{i}\frac{c}{h}\sin(kh). \tag{59}$$

As in Sections 3 and 5.1, we perform our testing with three time discretization strategies: forward Euler, backward Euler, and trapezoidal. To each approximate filter, we check the role of information criteria. With observation time $\Delta t = 1$, the forward Euler is an unstable numerical predictor ($|F_{h,k}| > 1$) while the two other implicit schemes are always stable with only weak damping for the trapezoidal method.

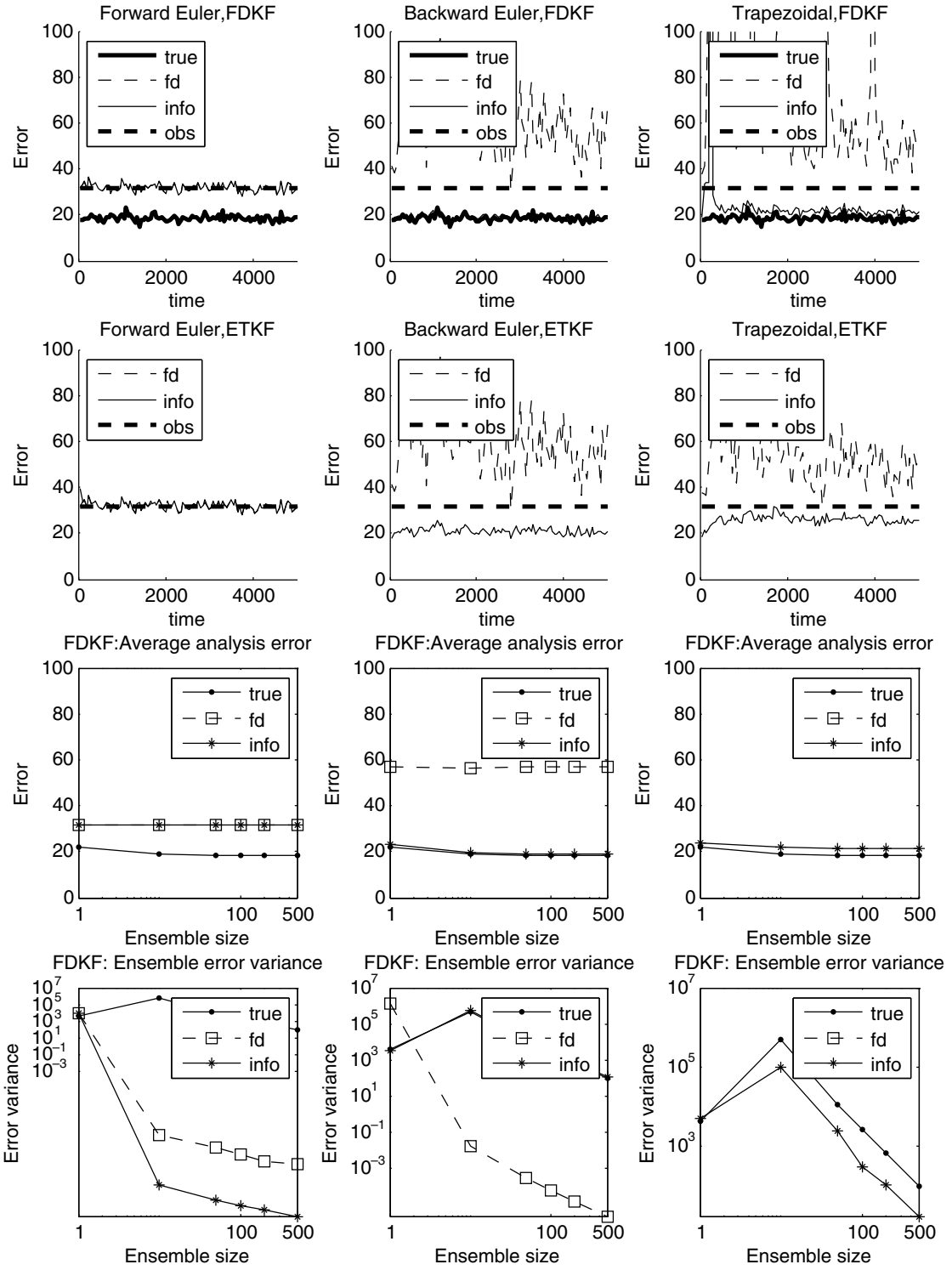

Fig. 14. Uniform damping, $E_k \sim k^{-5/3}$ with $\Delta t = 50$ and $N = 80$: RMS errors as functions of time for FDKF with ensemble size $K = 200$ (first row), second row for ETKF also with $K = 200$, RMS errors as functions of ensemble size for FDKF (third row), and ensemble error variances as functions of ensemble size for FDKF (fourth row). The panels in the first column depict simulations with forward Euler, second column with backward Euler, and third column with trapezoidal. In each panel, 'true' indicates the true filter, 'fd' denotes the finite difference approximate filter, 'info' denotes the approximate filter with information criterion noise variance, and 'obs' denotes the observation error.

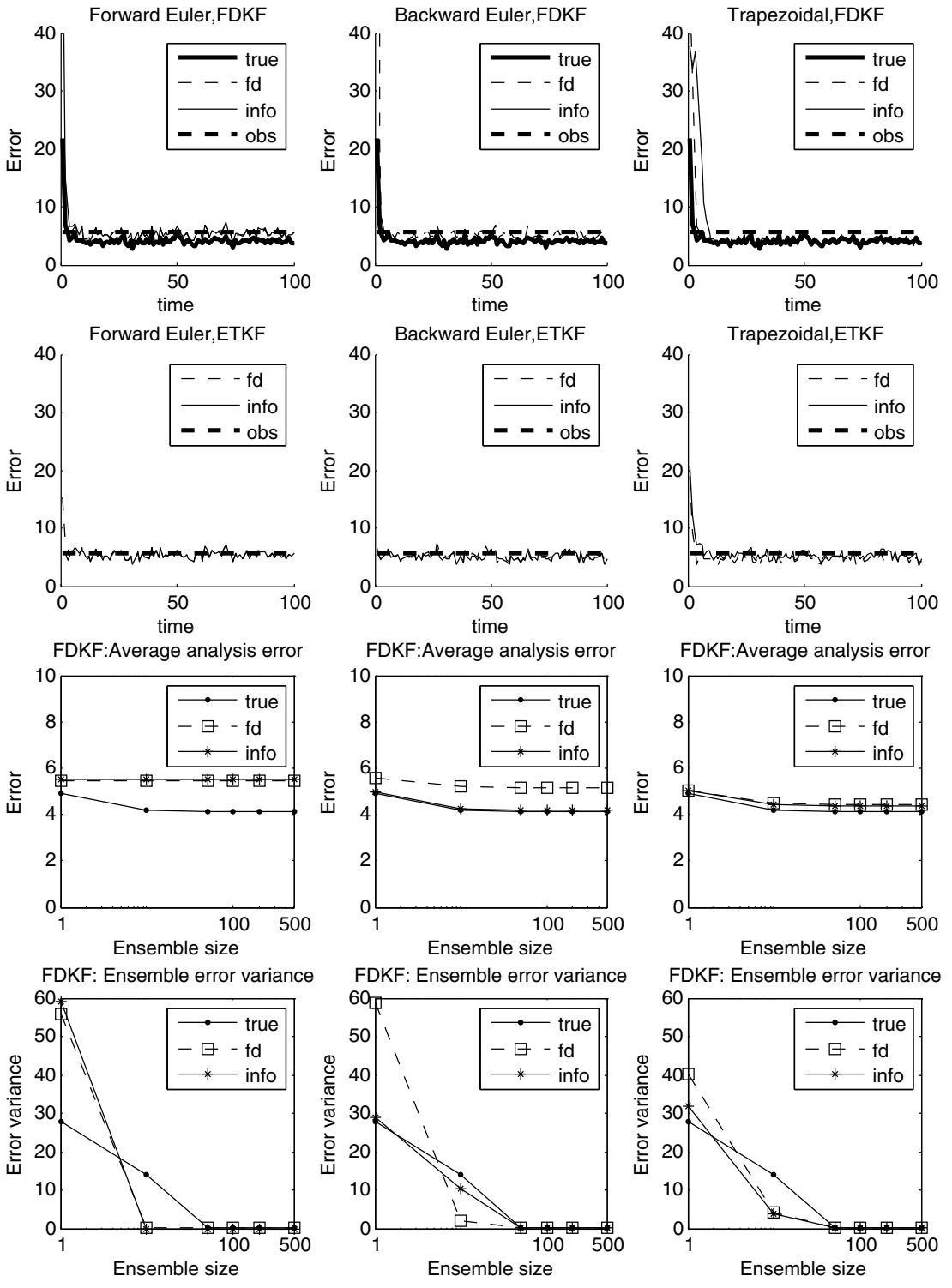

Fig. 15. Selective damping, $E_k = 1$ with $\Delta t = 0.5$ and $N = 20$: RMS errors as functions of time for FDKF with ensemble size $K = 100$ (first row), second row for ETKF also with $K = 100$, RMS errors as functions of ensemble size for FDKF (third row), and ensemble error variances as functions of ensemble size for FDKF (fourth row). The panels in the first column depict simulations with forward Euler, second column with backward Euler, and third column with trapezoidal. In each panel, 'true' indicates the true filter, 'fd' denotes the finite difference approximate filter, 'info' denotes the approximate filter with information criterion noise variance, and 'obs' denotes the observation error.

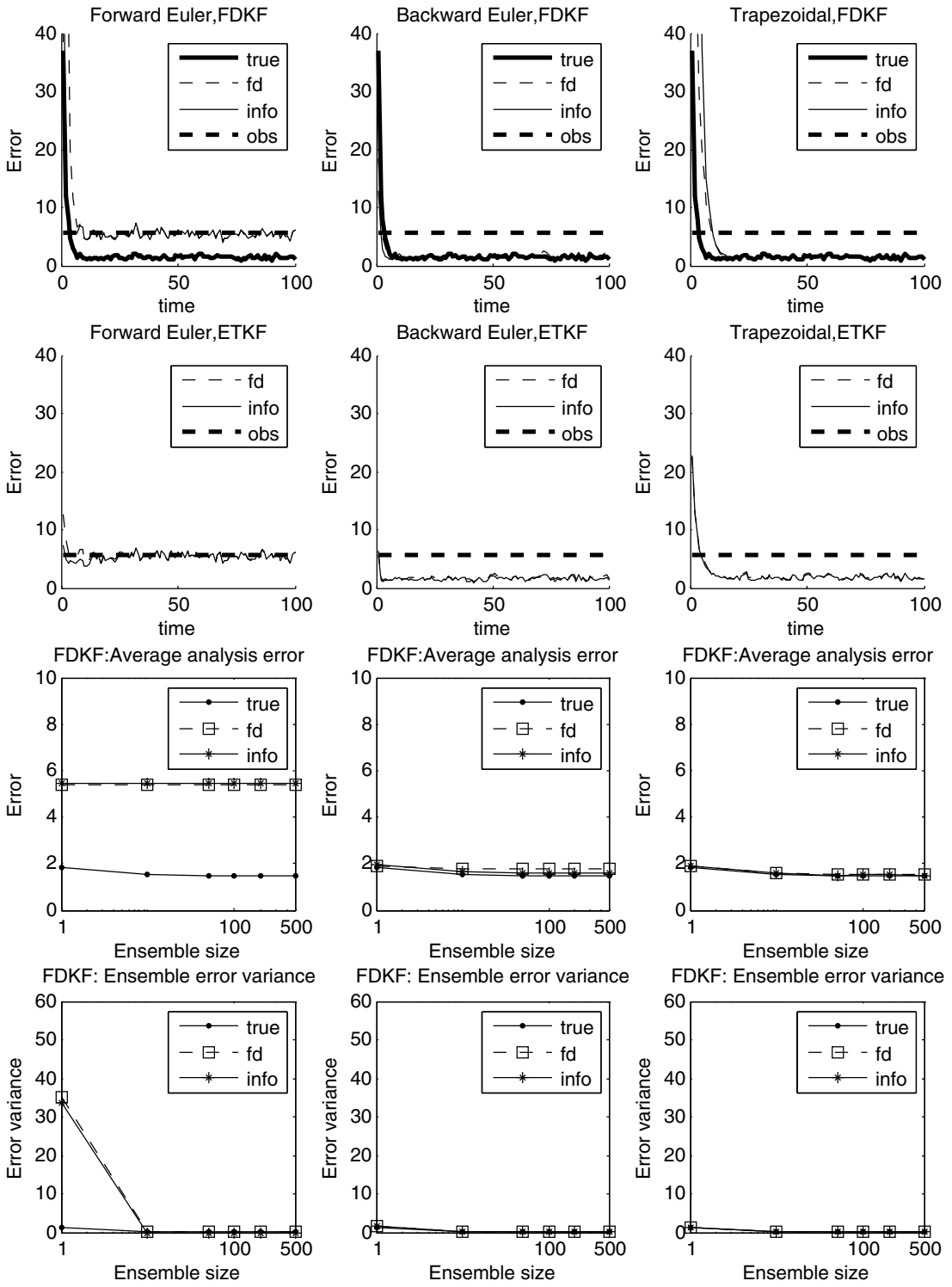

Fig. 16. Selective damping, $E_k = k^{-5/3}$ with $\Delta t = 0.5$ and $N = 20$: RMS errors as functions of time for FDKF with ensemble size $K = 100$ (first row), second row for ETKF also with $K = 100$, RMS errors as functions of ensemble size for FDKF (third row), and ensemble error variances as functions of ensemble size for FDKF (fourth row). The panels in the first column depict simulations with forward Euler, second column with backward Euler, and third column with trapezoidal. In each panel, 'true' indicates the true filter, 'fd' denotes the finite difference approximate filter, 'info' denotes the approximate filter with information criterion noise variance, and 'obs' denotes the observation error.

The predictions of the off-line test criteria in this regime indicate that both implicit schemes with either discretizated noise or augmented noise by information criteria satisfy the practical controllability criteria as well as strong filter stability although the trapezoidal method filter is less stable. As earlier, the off-line criteria for the unstable explicit Euler scheme predict that the filter trusts the observations. Figures demonstrating all of these facts are omitted for brevity.

In Figs. 15 and 16 (first row), we compare the RMS errors of FDKF for constant energy spectrum and relatively smooth spectrum, respectively, for ensemble size $K = 100$. Here, we see that the true filter performs better for energy spectrum $E_k = k^{-5/3}$. The role of the spectra here is completely explained by Theorem 1 in Section 4, i.e., the filter trusts the observations more for equipartition energy spectrum and trusts the dynamics more for $E_k = k^{-5/3}$ (which is also confirmed by the off-line testing which we omit). As predicted in the off-line testing, forward Euler trusts the observations while both implicit schemes are comparable to the true model regardless of how the system noises are chosen. In this regime, the information criteria improve both implicit filters insignificantly. From our simulations, we also notice that the trapezoidal scheme converges slower than backward Euler (see the panels in the first row of Fig. 16), which is clearly reflected by the fact that the filter stability function, $|F_{h,k}(1 - K_{h,k,\infty})|$ of backward Euler is smaller than that of the trapezoidal method.

Our numerical simulations suggest that both schemes (FDKF and ETKF) perform comparably in term of errors (e.g. compare also the filtered solutions of FDKF and ETKF in Figs. 15 and 16, respectively). However, one should note that ETKF is sensitive to variance inflation coefficient and ensemble size. For this experiment, we fixed the variance inflation at 10% and we show results only with ensemble size $K = 100$ since the filter diverges with $K \leqslant 50$. The FDKF performs remarkably well throughout the variations of ensemble size in terms of RMS errors (see the panels in the third row of Figs. 15 and 16) even with a single realization. In each simulation, the ensemble error variance (see the panels in the fourth row of Figs. 15 and 16) decreases as a function of ensemble size.

6. Concluding discussion

This paper develops new theoretical guidelines and illustrates their potential applicability for real-time filtering turbulent signals in complex systems such as those arising for weather prediction and climate change. These issues are studied here in the simplest context of plentiful spatial observations, i.e., the number of observations equals the number of mesh points for a scalar field although these observations can be infrequent in time compared to the local correlation time of the turbulent signal. Such a situation can occur practically for regional weather prediction models in populated areas in developed countries. On the theoretical side, this is the simplest context to develop and analyze radical filtering strategies with large model errors which can have filtering skill while avoiding the "curse of ensemble size". Diverse results have been developed throughout the paper so it is useful to summarize them along with their potential significance here.

In Section 3, we illustrated the fashion in which a recent theory [22] for filtering complex systems with turbulent energy spectra and plentiful observations can be developed into practical off-line test criteria through explicit formulas for filtering noisy turbulent signals. The off-line criteria include explicit formulas for the asymptotic Kalman gain, strategies to avoid filter divergence using information criteria, and assessment of asymptotic filter stability. All of these off-line criteria were presented in the context of discrete filters for a complex scalar equation test problem as guaranteed by the theory in [22]; however, one needs to develop criteria and methods for discrete filtering of the scalar test problem with large model error in stiff parameter regimes in order to gain insight into filtering rough turbulent signals from PDE's even with plentiful observations. This has been done in detail in Section 3 with surprising new phenomena. The off-line criteria and elementary numerical experiments show that in various regimes of parameters, the natural discretized noise with an implicit scheme can violate practical controllability and yield completely inaccurate filtering for the signal even though the filter is asymptotically strongly stable as for the backward Euler method; the information criteria restore practical controllability as a natural way to account for model error and avoid filter divergence. Also the first order accurate backward Euler filter with augmented noise is often a more statistically accurate filter than the second order trapezoidal method because it has better asymptotic filter stability. Stable filtering for the unstable Euler method occurs because the filter trusts observations.

The off-line guidelines for filter performance developed in Section 3 were tested extensively in Section 5 for the forced dissipative advection equation with very rough turbulent spectra, uniform damping, and less frequent observations in Section 5.1 and for smoother spectra with selective damping and more frequent observations in Section 5.2. First, it was established that the off-line test criteria for filter performance from Section 3 are applicable to these more difficult test cases to predict and understand successful and unsuccessful filter performance. In particular the information criteria for the system noise accounts for enough of the discrete model error to restore accurate filter performance for the implicit methods as comparable to the perfect model as shown in Section 5.1 in this difficult test bed. A second goal achieved in Section 5 is to illustrate that the off-line guidelines for the diagonal Fourier domain filter apply to the extended Kalman filter in physical space which generates additional errors which do not respect the Fourier diagonal structure.

Significantly, it is also demonstrated in Sections 3 and 5 that accurate statistical filtering with the implicit schemes and the information criteria with large time steps can be achieved with extremely small ensemble sizes, even as small as a single member to address "the curse of ensemble size". Futhermore, as discussed in Section 5, the Fourier space filtering methods require no adjustable parameters and much smaller ensemble size when compared with standard ensemble Kalman filters which involve adjustable parameters such as variance inflation. Recently, two of the authors have applied the diagonal stochastic Fourier filters to the Lorenz 96 model from atmosphere science [21,23,24] and compared their performance directly to the non-linear EKTF for the perfect model in that context as an extremely stringent test problem for this approach [17]; they found that FDKF supersedes ETKF in a fully turbulent regime.

In another direction, explicit rigorous mathematical criteria were developed in Section 4 to provide guidelines to address important questions for operational models: If plentiful observations are available on a range of spatial mesh sizes, what is gained in filter performance by increasing the resolution of the operational model? How does this depend on the nature of the turbulent spectrum in the signal being filtered? Finally, two of the authors have utilized the theory in [22] combined with the same overall strategy utilized in this paper to develop guidelines for filtering with sparse regular observations where several new phenomena beyond this work occur [6].

Acknowledgements

The research of Andrew J. Majda is partially supported by NSF Grant DMS-0456713, ONR Grant N0014-05-1-0164, and the Defense Advanced Research Projects Agency Grant N00014-07-1-0750, while Emilio Castronovo is supported as a postdoctoral fellow through the first two grants and John Harlim is supported as a postdoctoral fellow through the last two grants.

Appendix A. Time discretization

In this appendix, the derivation of the system noise for (1) is presented for the explicit and implicit schemes. For explicit forward Euler, we substitute

$$\frac{u(t+\Delta t)-u(t)}{\Delta t} \tag{A.1}$$

for $\mathrm{d}u(t)/\mathrm{d}t$ in (1) and recalling that $\tilde{\sigma}\dot{W}(t) \sim \tilde{\sigma}\Delta t^{-\frac{1}{2}}\mathcal{N}(0,1)$, with $\mathcal{N}(0,1)$ a normal random variable with mean zero and variance one, we obtain

$$u(t+\Delta t) = (1+\lambda\Delta t)u(t) + \tilde{\sigma}\Delta t^{\frac{1}{2}}\mathcal{N}(0,1). \tag{A.2}$$

Thus, the model noise variance is given by $\tilde{\sigma}^2\Delta t$.

For implicit backward Euler, we have

$$u(t+\Delta t) - u(t) = \lambda\Delta t u(t+\Delta t) + \tilde{\sigma}\Delta t^{\frac{1}{2}}\mathcal{N}(0,1), \tag{A.3}$$

solving for $u(t+\Delta t)$, we obtain

$$u(t+\Delta t) = (1-\lambda\Delta t)^{-1}u(t) + (1-\lambda\Delta t)^{-1}\tilde{\sigma}\Delta t^{\frac{1}{2}}\mathcal{N}(0,1) \tag{A.4}$$

with system variance $r_h = |1-\lambda\Delta t|^{-2}\tilde{\sigma}^2\Delta t$.

For symmetric trapezoidal method, we have

$$u(t+\Delta t) - u(t) = \frac{\lambda\Delta t}{2}(u(t+\Delta t)+u(t)) + \tilde{\sigma}\Delta t^{\frac{1}{2}}\mathcal{N}(0,1) \tag{A.5}$$

which results in

$$u(t+\Delta t) = \frac{1+\frac{\lambda\Delta t}{2}}{1-\frac{\lambda\Delta t}{2}}u(t) + \left(1-\frac{\lambda\Delta t}{2}\right)^{-1}\tilde{\sigma}\Delta t^{\frac{1}{2}}\mathcal{N}(0,1) \tag{A.6}$$

with variance $r_h = \left|1-\frac{\lambda\Delta t}{2}\right|^{-2}\tilde{\sigma}^2\Delta t$.

Appendix B. Limiting Kalman gain

Consider the complex system variable $u_{m|m}$, satisfying the evolution–observation system

$$u_{m+1|m} = Fu_{m|m} + \sigma_{m+1}, \tag{B.1}$$

$$v_{m+1} = g\bar{u}_{m|m} + \sigma^{\mathrm{o}} \tag{B.2}$$

with $r_m = \langle|\sigma_m|^2\rangle$ and $r^{\mathrm{o}} = \langle|\sigma^{\mathrm{o}}|^2\rangle$. Then the limiting Kalman filter is given by

$$\bar{u}_{m|m} = u_{m|m-1} + K_\infty(v_m - g\bar{u}_{m|m-1}), \tag{B.3}$$

where $\bar{u}_{m/m}$ is the mean of $u_{m/m}$ and K_∞ is the limiting Kalman gain and it is obtained by

$$K_\infty = \frac{r_\infty g}{g^2 r_\infty + r^{\mathrm{o}}} \tag{B.4}$$

with r_∞ the limiting variance. In order to obtain an analytical expressions for the limiting quantities we consider $u = a + \mathrm{i}b$, $v = x + \mathrm{i}y$; the symbol F and the noise satisfy

$$F = A + \mathrm{i}B, \tag{B.5}$$

$$\sigma^{\mathrm{o}} = \sigma^{\mathrm{o},r} + \mathrm{i}\sigma^{\mathrm{o},i} \quad \text{and} \quad \sigma = \sigma^r + \mathrm{i}\sigma^i. \tag{B.6}$$

We can thus rewrite (B.1) and (B.2) as

$$\begin{pmatrix} a_{m+1|m} \\ b_{m+1|m} \end{pmatrix} = \begin{pmatrix} A & -B \\ B & A \end{pmatrix}\begin{pmatrix} a_{m|m} \\ b_{m|m} \end{pmatrix} + \begin{pmatrix} \sigma^r_{m+1} \\ \sigma^i_{m+1} \end{pmatrix}, \tag{B.7}$$

$$\begin{pmatrix} x_{m+1} \\ y_{m+1} \end{pmatrix} = g\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \bar{a}_{m|m} \\ \bar{b}_{m|m} \end{pmatrix} + \begin{pmatrix} \sigma^{\mathrm{o},r} \\ \sigma^{\mathrm{o},i} \end{pmatrix}. \tag{B.8}$$

The Kalman filter asymptotic covariance is a fixed point T of the following

$$T = F(T - TG^{\mathrm{T}}(GTG^{\mathrm{T}} + R^0))^{-1}F^{\mathrm{T}} + R \tag{B.9}$$

for general matrices

$$F = \begin{pmatrix} A & B \\ -B & A \end{pmatrix}, \quad G = g\mathcal{I}, \quad R = r\mathcal{I}, \quad \text{and} \quad R^{\mathrm{o}} = r^{\mathrm{o}}\mathcal{I}. \tag{B.10}$$

We seek a symmetric T of the form

$$T = \begin{pmatrix} \alpha & \gamma \\ \gamma & \beta \end{pmatrix}. \tag{B.11}$$

By explicitly multiplying out Eq. (B.9), we find that

$$\alpha = \beta = [g^4 r(\alpha\beta - \gamma^2) + g^2(r(\alpha+\beta) + (A^2+B^2)(\alpha\beta-\gamma^2))(r^o) + (r + A^2\alpha + B^2\beta + 2AB\gamma)(r^o)^2](g^4(\alpha\beta-\gamma^2) + g^2(\alpha+\beta)r^o + (r^o)^2)^{-1}, \tag{B.12}$$

$$\gamma = \frac{AB(-\alpha+\beta) + (A^2\gamma - B^2\gamma)(r^o)^2}{g^4(\alpha\beta-\gamma^2) + g^2(\alpha+\beta)r^o + (r^o)^2}. \tag{B.13}$$

Notice also that $\gamma = 0$ solves (B.13). Now if we assume $\alpha = \beta$ and $\gamma = 0$, we can obtain a solution for (B.12). Furthermore, if we consider the system to be *observable* and *controllable*, then there is only one fixed point to Eq. (B.9) and thus $r_\infty = \alpha$. Substituting r_∞ in (B.4), we obtain Eq. (30).

References

[1] B.D. Anderson, J.B. Moore, Optimal Filtering, Prentice-Hall, Englewood Cliffs, NJ, 1979.
[2] J.L. Anderson, An ensemble adjustment Kalman filter for data assimilation, Monthly Weather Review 129 (12) (2001) 2884–2903.
[3] J.L. Anderson, A local least squares framework for ensemble filtering, Monthly Weather Review 131 (4) (2003) 634–642.
[4] L.M. Berliner, R.F. Milliff, C.K. Wikle, Bayesian hierarchical modeling of air–sea interaction, Journal of Geophysical Research (Oceans) 108 (C4) (2003) 3104–3120.
[5] C.H. Bishop, B.J. Etherton, S.J. Majumdar, Adaptive sampling with the ensemble transform Kalman filter. Part I: The theoretical aspects, Monthly Weather Review 129 (2001) 420–436.
[6] J. Harlim, A.J. Majda, Mathematical strategies for filtering complex systems: regularly spaced sparse observations, J. Comp. Phys., doi:10.1016/j.jcp.2008.01.049.
[7] A.J. Chorin, P. Krause, Dimensional reduction for a Bayesian filter, Proceedings of the National Academy of Sciences 101 (42) (2004) 15013–15017.
[8] C.K. Chui, G. Chen, Kalman Filtering, Springer, New York, 1999.
[9] S.E. Cohn, D. Dee, Observability of discretized partial differential equations, SIAM Journal on Numerical Analysis 25 (3) (1988) 586–617.
[10] B.F. Farrell, P.J. Ioannou, State estimation using a reduced-order Kalman filter, Journal of the Atmospheric Sciences 58 (23) (2001) 3666–3680.
[11] B.F. Farrell, P.J. Ioannou, Distributed forcing of forecast and assimilation error systems, Journal of the Atmospheric Sciences 62 (2) (2005) 460–475.
[12] C.W. Gardiner, Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences, Springer-Verlag, New York, 1997.
[13] M. Ghil, P. Malanotte-Rizzoli, Data assimilation in meteorology and oceanography, Advances in Geophysics 33 (1991) 141–266.
[14] M.J. Grote, A.J. Majda, Stable time filtering of strongly unstable spatially extended systems, Proceedings of the National Academy of Sciences 103 (20) (2006) 7548–7553.
[15] J. Harlim, Errors in the initial conditions for numerical weather prediction: a study of error growth patterns and error reduction with ensemble filtering, Ph.D. Thesis, University of Maryland, 2006.
[16] J. Harlim, B.R. Hunt, A non-Gaussian ensemble filter for assimilating infrequent noisy observations, Tellus 59A (2) (2007) 225–237.
[17] J. Harlim, A.J. Majda, Filtering nonlinear dynamical systems with linear stochastic models, Nonlinearity, submitted for publication.
[18] K. Haven, A.J. Majda, R.V. Abramov, Quantifying predictability through information theory: small sample estimation in a non-Gaussian framework, Journal of Computational Physics 206 (1) (2005) 334–362.
[19] A.H. Jazwinski, Stochastic Processes and Filtering Theory, Academic Press, New York, 1970.
[20] J.P. Kaipio, E. Somersalo, Statistical and Computational Inverse Problems, Springer, New York, 2005.
[21] E.N. Lorenz, Predictability – a problem partly solved, in: Proceedings of the Seminar on Predictability, vol. 1, ECMWF, Reading, Berkshire, UK, 1995, pp. 1–18.
[22] A.J. Majda, M.J. Grote, Explicit off-line criteria for stable accurate time filtering of strongly unstable spatially extended systems, Proceedings of the National Academy of Sciences 104 (4) (2007) 1124–1129.
[23] A.J. Majda, R.V. Abramov, M.J. Grote, Information Theory and Stochastics for Multiscale Nonlinear Systems, American Mathematical Society, 2005.
[24] A.J. Majda, X. Wang, Nonlinear Dynamics and Statistical Theories for Basic Geophysical Flows, Cambridge University Press, 2006.
[25] R.N. Miller, E.F. Carter, S.T. Blue, Data assimilation into nonlinear stochastic models, Tellus A 51 (2) (1999) 167–194.
[26] E. Ott, B.R. Hunt, I. Szunyogh, A.V. Zimin, E.J. Kostelich, M. Corazza, E. Kalnay, D.J. Patil, J.A. Yorke, A local ensemble kalman filter for atmospheric data assimilation, Tellus A 56 (5) (2004) 415–428.
[27] R.D. Richtmyer, K.W. Morton, Difference Methods for Initial Value Problems, Wiley, 1967.
[28] A. Simmons, The Control of Gravity Waves in Data Assimilation, 1999. <http://www.ecmwf.int/newsevents/training/rcourse_notes/DATA_ASSIMILATION/GRAV-WAVE_CONTROL/Grav-Wave_control9.html>.
[29] J.J. Tribbia, D.P. Baumhefner, Estimates of the predictability of low-frequency variability with a spectral general circulation model, Journal of Atmospheric Science 45 (1988) 2306–2317.
[30] R. Todling, M. Ghil, Tracking atmospheric instabilities with the Kalman Filter. Part 1: Methodology and one-layer results, Monthly Weather Review 122 (1) (1994) 183–204.