Available online at www.sciencedirect.com
ScienceDirect
Journal of Computational Physics 227 (2008) 5304–5341
JOURNAL OF COMPUTATIONAL PHYSICS
www.elsevier.com/locate/jcp

Mathematical strategies for filtering complex systems: Regularly spaced sparse observations

J. Harlim *, A.J. Majda

Department of Mathematics and Center for Atmosphere and Ocean Science, Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, NY 10012, United States

Received 13 September 2007; received in revised form 18 January 2008; accepted 28 January 2008
Available online 8 February 2008

Abstract

Real time filtering of noisy turbulent signals through sparse observations on a regularly spaced mesh is a notoriously difficult and important prototype filtering problem. Simpler off-line test criteria are proposed here as guidelines for filter performance for these stiff multi-scale filtering problems in the context of linear stochastic partial differential equations with turbulent solutions. Filtering turbulent solutions of the stochastically forced dissipative advection equation through sparse observations is developed as a stringent test bed for filter performance with sparse regular observations. The standard ensemble transform Kalman filter (ETKF) has poor skill on the test bed and even suffers from filter divergence, surprisingly, at observable times with resonant mean forcing and a decaying energy spectrum in the partially observed signal. Systematic alternative filtering strategies are developed here including the Fourier Domain Kalman Filter (FDKF) and various reduced filters called Strongly Damped Approximate Filter (SDAF), Variance Strongly Damped Approximate Filter (VSDAF), and Reduced Fourier Domain Kalman Filter (RFDKF) which operate only on the primary Fourier modes associated with the sparse observation mesh while nevertheless, incorporating into the approximate filter various features of the interaction with the remaining modes. It is shown below that these much cheaper alternative filters have significant skill on the test bed of turbulent solutions which exceeds ETKF and in various regimes often exceeds FDKF, provided that the approximate filters are guided by the off-line test criteria. The skill of the various approximate filters depends on the energy spectrum of the turbulent signal and the observation time relative to the decorrelation time of the turbulence at a given spatial scale in a precise fashion elucidated here.

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

1. Introduction

An important emerging scientific issue is the real time filtering through observations of noisy turbulent signals for complex dynamical systems with many degrees of freedom as well as the statistical accuracy of various

DOI of original article: 10.1016/j.jcp.2007.12.016.
* Corresponding author.
E-mail address: jharlim@cims.nyu.edu (J. Harlim).

0021-9991/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcp.2008.01.049

filtering strategies in this context. These problems typically have many spatio-temporal scales, rough turbulent energy spectra in the solutions near the mesh scale, and a very large dimensional state space yet real time predictions are needed [24,27]. There is an inherently difficult practical issue, "the curse of ensemble size", due to the large computational overhead in generating individual ensemble members through the forward dynamical operator [18]. These issues become especially subtle when there are sparse observations of the chaotic turbulent signal. Various reduced order filtering strategies [28,13,31,2,3,7,10,11,29,21,16] which are based on the classical Kalman filtering algorithm [1,8,22] have been developed with some success for filtering large scale instabilities in a chaotic system.

Here, we focus on the issue of filtering turbulent solutions through sparse observations on a regularly spaced mesh which is a subset of the discretization mesh (see Fig. 1). In certain situation, this is a notoriously difficult test problem where for example, in the chaotic 40-dimensional L-96 model [23] with sparse regular observations, the ensemble Kalman filter (in this paper, we implement ETKF [5] with large ensemble size) exhibits filter divergence [29] with any variance inflation parameter when the dynamical system is fully turbulent and the observations are infrequent. The context for the study here is the filtering through sparse regular observations of turbulent solutions of a linear constant coefficient partial differential equation with stochastic forcing [6,14,25]. We do this both to gain fundamental theoretical insight into filter performance in a non-trivial idealized setting and to develop alternative filtering strategies to cope with the difficult issues mentioned in the first paragraph for both linear [6] and nonlinear problems [17].

The main objectives here are the following:

I. *ETKF filter performance*: To document the filter performance of ETKF on a non-trivial suite of test problems for turbulent signals with sparse regular observations. Below we find limited skill for ETKF on these problems and even filter divergence both due to the failure of observability [1,8,9,22] and surprisingly in regimes of resonant forcing where observability is satisfied.

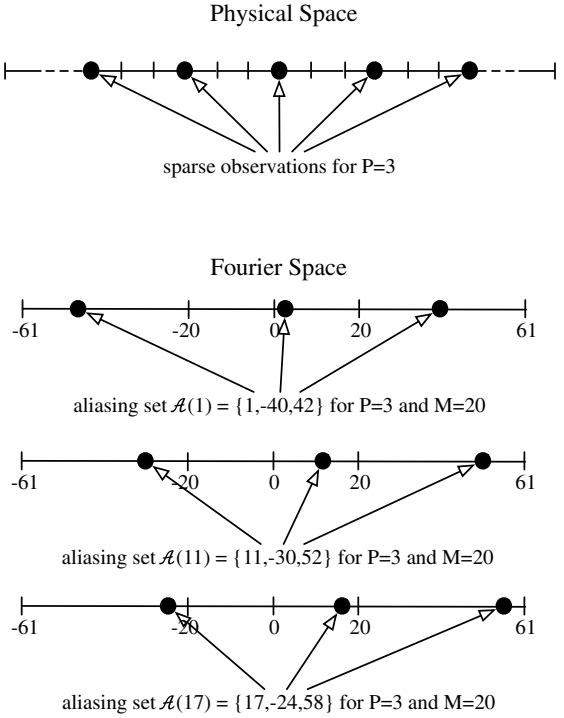

Fig. 1. The top figure describes an example of physical space with regularly spaced sparse observations for $P = 3$. The bottom figures describe examples of three aliasing sets (used in Section 4), i.e., $\mathcal{A}(1), \mathcal{A}(11)$, and $\mathcal{A}(17)$, corresponding to the above physical space $P = 3$ for $M = 20$.

II. *Off-line test criteria*: To develop simple off-line mathematical test criteria [25,6] as guidelines for filtering the extremely stiff multiple space-time scale problems which arise in filtering turbulent signals through sparse regular observations. Theorem 3 from [25] provides the theoretical basis for doing this.
III. *Alternative filters*: To develop more stable alternative reduced Fourier domain filtering strategies which practically filter only the primary "resolved" Fourier modes with dimension equal to the sparse observation set yet automatically incorporate crucial features of the remaining "unresolved" modes in the filter algorithm. Such algorithms simultaneously can be stable and always satisfy observability condition, and statistically accurate with small ensemble size for estimating the evolution of the mean state only as opposed to many ensemble filters that utilizes the ensemble to propagate the covariance in appropriate regimes of turbulent spectrum for filtering the turbulent signals.

With the above discussion, one might anticipate a very rich source of new phenomena in developing criteria and new algorithms for filtering turbulent signals with sparse regular observations. This is true and these are the issues studied in the present paper; next we provide a brief summary of the basic strategy and results.

The context of filtering sparse regularly spaced observations involves the discretization of the stochastic PDE on a fine mesh with only partial observations of the true signal on a coarse regular mesh which is a regular subset of the discretization mesh (see Fig. 1, Section 2, and [25] for precise definitions). We utilize the off-line mathematical test criterion in Theorem 3 of [25] for sparse regular observations in a parallel fashion as in [6] for the situation with plentiful observations; the important difference here is that the canonical off-line test problem is a filtering system coupling various aliased Fourier modes through a single observation rather than the much simpler scalar test problems developed in [6]; the dimension of this canonical problem is the ratio of the number of discrete regular mesh points to the number of discrete regular sparse observations. We set up this test problem in Section 2 and establish there that new issues arise in the test problems regarding violation of the observability condition for linear filters [1,8]. In fact, this violation of observability suggests that the exact perfect model cannot be utilized to filter the truth signal generated from this model. By perfect model, we mean that the true signal and the filter model are exactly the same. This subtle phenomenon leads to the development of alternative systematic approximate Fourier domain filtering strategies which are introduced and developed in Section 3 in the context of the canonical Fourier test problem for sparse observations: The Fourier Domain Kalman Filter (FDKF), the Strongly Damped Approximate Filters (SDAF, VSDAF), and the Reduced Fourier Domain Kalman Filter (RFDKF). The straightforward FDKF algorithm is the most complex filtering procedure while SDAF, VSDAF, and RFDKF are cheaper filters that operate on the lower dimensional state space associated with the primary Fourier modes of the sparse observations: RFDKF is a filter akin to the one used for Fourier methods with non-smooth initial data [26,19] where only the coarse mesh variables are filtered while the true dynamics is utilized on the remaining fine mesh variables. The SDAF algorithm is also novel and introduced in Section 3; it is a "memoryless" filter on a subset of "unresolved" modes coupled with a FDKF on a subset of "resolved" modes with a list of attractive features for filtering turbulent signals developed in Section 3; the SDAF algorithm combines a systematic analytic "turbulence closure" model with FDKF on a reduced phase space while VSDAF is an important variant. All these filtering strategies are tested on the canonical Fourier test problem for sparse observations in Section 4 for a wide range of observation times, mean forcings, correlation times, and turbulent spectra to establish theoretical guidelines for filter performance for turbulent signals for the PDE cases, studied in Sections 5 and 6.

For the canonical Fourier test problem, the approximate filters satisfy observability criteria and often have performance comparable to and even superseding the standard Kalman filter (FDKF) in various important regimes. As in [6], turbulent solutions of the stochastically forced dissipative advection equation are utilized throughout the paper both for numerical experiments presented in Sections 5 and 6 and theoretical illustration and/or motivation in Sections 2–4. With the theoretical guidelines from Section 4, the FDKF, VSDAF, SDAF, and RFDKF algorithms are tested on the discretized PDE's with sparse regular observations and compared to a spatial domain ensemble Kalman filter ETKF [5] in Sections 5 and 6. In particular, the standard ETKF algorithm has strong filter divergence with resonant mean forcing even when observability is satisfied while the SDAF, VSDAF, and RFDKF approximations have significant filter skill in this very difficult test bed depending on the nature of the turbulent signal and the observation time. Both Sections 5 and 6 end with a comprehensive summary and discussion of filter performance for sparse regular observations for turbulent

signals for the stochastically forced dissipative advection equation with selective and small uniform damping, respectively. Section 7 of the paper provides a concluding summary with a perspective on future developments.

2. Filtering discretizations of stochastic PDE's with sparse regular observations

2.1. The perfect model signals

The signals which will be filtered through regularly spaced sparse observation by various algorithms, the perfect truth signals, are determined by solutions of the real valued scalar stochastically forced PDE:

$$\begin{aligned} &\frac{\partial u(x,t)}{\partial t} = \mathcal{P}\left(\frac{\partial}{\partial x}\right)u(x,t) - \gamma\left(\frac{\partial}{\partial x}\right)u(x,t) + \overline{F}(x,t) + \sigma(x)\dot{W}(t), \\ &u(x,0) = u_0(x). \end{aligned} \tag{1}$$

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 $\overline{F}(x,t)$ is a known deterministic forcing and the initial data u_0 is a Gaussian random field with non-zero mean and covariance. The problem in (1) is non-dimensionalized to a 2π-periodic domain so that continuous and discrete Fourier series can be utilized in analyzing (1) 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

$$\begin{aligned} &\mathcal{P}\left(\frac{\partial}{\partial x}\right)e^{ikx} = \tilde{p}(ik)e^{ikx} \\ &\gamma\left(\frac{\partial}{\partial x}\right)e^{ikx} = \gamma(ik)e^{ikx}. \end{aligned} \tag{2}$$

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

$$\tilde{p}(ik) = i\omega_k \tag{3}$$

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

$$\gamma(ik) > 0 \quad \text{for all } k \neq 0. \tag{4}$$

In geophysical applications, it is natural to have a climatological distribution and as discussed below, (3), (4) are needed in order to guarantee this.

The general solution of (1) is defined through Fourier series. The 2π-periodic solution of (1) is expanded in Fourier series

$$u(x,t) = \sum_{k=-\infty}^{\infty} \hat{u}_k(t)e^{ikx}, \quad \hat{u}_{-k} = \hat{u}_k^*, \tag{5}$$

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

$$d\hat{u}_k = [\tilde{p}(ik) - \gamma(ik)]\hat{u}_k(t)dt + \widehat{F}_k(t)dt + \tilde{\sigma}_k\, dW_k, \quad \hat{u}_k(0) = \hat{u}_{0,k}. \tag{6}$$

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 symbol $\tilde{p}(ik)$ of the underlying differential equation is purely imaginary, i.e., a wave-like system, while the symbol $\gamma(ik)$ for the damping is always real and positive, i.e., $\gamma(ik) > 0$, the statistical equilibrium distribution for (6) is a Gaussian with variance, E_k, defining the climatological energy spectrum given by

$$E_k = \frac{\tilde{\sigma}_k^2}{2\gamma_k}, \quad 1 \leqslant k < +\infty. \tag{7}$$

Mathematically, one needs to require $\sum E_k < \infty$ to define the stochastic solution of (1) correctly with a similar requirement on the Gaussian initial data in $u_0(x)$. The damping coefficient, $\gamma_k = \gamma(ik)$ roughly defines the correlation time [6], γ_k^{-1}, while $\omega_k = -i\tilde{p}(ik)$ defines the oscillation frequency at wave number k. Clearly, γ_k^{-1} measures the memory in the signal being filtered. As discussed in [25], the noise in (6) and (7) represents the turbulent fluctuations on the mesh scale for both unresolved and resolved features of the nonlinear dynamics ([24,27] and references there) with a given energy spectrum E_k and decorrelation time γ_k^{-1} at each wave number. 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 [24,27,25]. As in [6], here we consider a general family of power law energy spectrum,

$$E_k = E_o|k|^{-\beta}, \tag{8}$$

where β is a fixed exponent. The case with $\beta = 0$ defines an extremely rough equipartition of energy spectrum while $\beta = 5/3$ defines the familiar Kolmogorov spectrum which generates a fractal random field. Note that strictly speaking, $\sum E_k < \infty$ only if $\beta > 1$; however, the case where $\beta < 1$ with a much rougher spatial random field can arise quite often in practice as occurring on all families of realistic resolved meshes where the smallest available mesh is still much larger than the turbulent dissipation scale; under these circumstances and in many standard discretization approaches, we truncate (1) to the smallest mesh size. In this paper, we only discuss and present results for a single space dimension to avoid cumbersome notation, however, the theory below also applies in several variables.

2.2. 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,

$$(f, \bar{g})_h = \frac{h}{2\pi}\sum_{j=0}^{2N} f_j g_j, \tag{9}$$

where $\bar{g}$ denotes the conjugate of g. The discrete Fourier coefficients, $\hat{f}_k$ are defined by $\hat{f}_k = (f, e^{ikx_j})_h$ for $|k| \leqslant N$, with the well-known properties

$$f_j = \sum_{|k|\leqslant N} \hat{f}_k e^{ikx_j}, \quad \hat{f}_{-k} = \hat{f}_k^*, \quad (f,f)_h = \sum_{|k|\leqslant N} |\hat{f}_k|^2. \tag{10}$$

For a standard finite difference approximation to (1) without any random noise with a time step Δt, the solution is expressed in standard fashion [30] in terms of the amplification factor, $F_{h,k}$, for integers k with $|k| \leqslant N$. With the finite Fourier expansion

$$u^h = \sum_{|k|\leqslant N} \hat{u}_k^h e^{ikx}, \tag{11}$$

the general discrete approximation of (1) is given at the observation times $m\Delta t$ through its Fourier coefficients $\hat{u}_k^h$ by the block diagonal operation,

$$\hat{u}_{k,m+1|m}^h = F_{h,k}\hat{u}_{k,m|m}^h + \overline{F}_{k,m} + \sigma_{h,k,m+1}. \tag{12}$$

In (12) the zero mean complex Gaussian noises, $\sigma_{h,k,m+1}$, are uncorrelated in time and their second moment averages satisfy

$$\langle \sigma_{h,k,m+1}\sigma_{h,k',m+1}^* \rangle = \delta_{k+k'} r_{h,k}, \quad |k|, |k'| \leqslant N \tag{13}$$

with $r_{h,k}$ the variance at wave number k and δ_j the Kronecker delta function. The consistent notation $F_k = F_{h,k} = e^{(\tilde{p}(ik)-\gamma(ik))\Delta t}$ is utilized for the amplification factor of the exact solution operator in (1). In this important special case, the complex Gaussian system noise $\sigma_{h,k} = \sigma_k$ is defined through the exact solution of the stochastic equation in (6) from time step $m\Delta t$ to time step $(m+1)\Delta t$; thus the real and imaginary part of σ_k are zero mean independent random variables with variance

$$r_k = E_k(1 - |F_k|^2), \quad \text{where} \quad F_k = \mathrm{e}^{(\tilde{p}(\mathrm{i}k) - \gamma(\mathrm{i}k))\Delta t}. \tag{14}$$

The external forces $\overline{F}_{k,m}$ are prescribed through a compatible discretization of (6) and are assumed to be given.

2.3. Filtering sparse regular observations

Here, we define the set of regularly spaced sparse observations where the noisy measurements of the truth signal from (1) are taken and filtered through the discrete dynamics from 2.2. Suppose the observations of the truth signal defined in (1) are taken at $2M+1$ grid points which are regularly spaced, i.e., $\tilde{x}_j = j\tilde{h},\ j = 0, 1, \ldots, 2M$ with $(2M+1)\tilde{h} = 2\pi$. When $M < N$ where $(2N+1)h = 2\pi$ and h denotes the mesh spacing for the finite difference approximation, we have sparse regular observations since there are fewer observations than discrete mesh points. Here, for simplicity in exposition, we assume that M and N are related by

$$N = M + \widetilde{P}(2M+1) \quad \text{for any fixed } \widetilde{P} = 1, 2, 3, \ldots$$

so that with $P = 2\widetilde{P} + 1$ there are $P(2M+1)$ regular spaced mesh points and $2M+1$ regular spaced observation points. Thus, P defines the ratio of the total discrete mesh points available for the discretization to the number of sparse regular observation locations in (12) (see Fig. 1 for a visual demonstration).

With the sequence of observation times $m\Delta t$ and the discrete forward operator from (11) and (12), defining $\bar{u}^h_{m+1|m}(x)$, we attempt to get the best estimate of the truth signal $u(x, m\Delta t)$, defined through (1) by utilizing the discrete forward operator for u^h and the noisy sparse regular observations of u at $\tilde{x}_j = j\tilde{h}$, i.e.,

$$u(j\tilde{h}, m\Delta t) = \bar{u}^h_{m|m}(j\tilde{h}) + \sigma^{\mathrm{o}}_{j,m}, \quad j = 0, 1, 2, \ldots, 2M. \tag{15}$$

In (15), the observational noises, $\sigma^{\mathrm{o}}_{j,m}$, are assumed to be zero mean Gaussian random variables which are independent for each location $\tilde{x}_j$ with variance r^{o}. Thus, the filtering algorithm evolves the Gaussian statistical state estimate from time $m\Delta t$ through the mean and covariance first by the dynamics in (11) and (12) to create the prior Gaussian distribution, $\bar{u}^h_{m+1|m}$, which then is constrained by the noisy observations in (15) to produce the state estimate $\bar{u}^h_{m+1|m+1}$ defining the filtering approximation; this is the quantity which is the main interest for us here. When r_k and F_k are given by the exact solution operator for (1) as in (14), this is clearly filtering the truth signal in the perfect model scenario. We will see below, that devising accurate filtering procedures for the rough turbulent solutions studied here presents an interesting challenge even in the perfect model scenario.

Without the external forcing, $\overline{F}(x,t)$, the truth signal in (1) always decays exponentially in time to the climatological equilibrium defined through (6) and (7) (see [12]) and in particular the mean state exponentially decays to zero. The external forcing $\widehat{F}_k(t)$ in (6) both has physical importance and prevents the mean state from decaying to zero; this provides another stringent test for filter performance in time. Here, we utilize low wave number time periodic forcing for (6) with

$$\widehat{F}_k(t) = \begin{cases} A\mathrm{e}^{\mathrm{i}\omega_{\mathrm{o}}(k)t}, & \text{if } k \leqslant M \\ 0, & \text{if } k > M \end{cases} \tag{16}$$

for prescribed $\omega_{\mathrm{o}}(k)$. By calculating the exact solution of (6), the time discrete forcing in the perfect filtering model induced by $\widehat{F}_k(t)$ in (16) is given by

$$\overline{F}_{k,m} = \begin{cases} \frac{A\mathrm{e}^{\mathrm{i}\omega_{\mathrm{o}} m\Delta t}}{\gamma_k + \mathrm{i}(\omega_{\mathrm{o}} - \omega_k)}\left(\mathrm{e}^{\mathrm{i}\omega_{\mathrm{o}}\Delta t} - \mathrm{e}^{(-\gamma_k + \mathrm{i}\omega_k)\Delta t}\right), & \text{if } k \leqslant M \\ 0, & \text{if } k > M \end{cases} \tag{17}$$

with $\mathrm{i}\omega_k = \tilde{p}(\mathrm{i}k)$ and $\gamma_k = \gamma(\mathrm{i}k)$.

2.4. The canonical Fourier filter for sparse regular observations

Given a discrete Fourier wave number, ℓ, $|\ell| \leqslant M$, associated with the sparse observation mesh $j\tilde{h},\ j = 0, 1, 2, \ldots, M$, the aliasing set $\mathcal{A}(\ell)$ on the fine discretization mesh is the set of P wave numbers given by

$$\mathcal{A}(\ell) = \{k | k = \ell + (2M+1)q, |k| \leqslant M + \widetilde{P}(2M+1)\}, \tag{18}$$

where $q = 0, \pm 1, \pm 2, \pm 3, \ldots$, and from 2.3, $P = 2\widetilde{P} + 1$. In Theorem 3 of [25], it is established under natural hypotheses that the filtering problem defined by (11), (12), and (18) is equivalent to the following family of complex P-dimensional Fourier domain filtering problems for the amplitudes at wave numbers in each aliasing set, $\mathcal{A}(\ell)$.

$$\begin{aligned} &\text{(A)} \quad \hat{u}^h_{k,m+1|m} = F_{h,k}\hat{u}^h_{k,m|m} + \overline{F}_{k,m} + \sigma_{h,k,m+1} \quad \text{for } k \in \mathcal{A}(\ell), \\ &\text{(B)} \quad \sum_{k\in\mathcal{A}(\ell)} \hat{u}_k((m+1)\Delta t) = \sum_{k\in\mathcal{A}(\ell)} \overline{\hat{u}}^h_{k,m+1|m+1} + \hat{\sigma}^{\mathrm{o}}_{\ell,m+1}, \end{aligned} \tag{19}$$

where for each ℓ, $0 \leqslant \ell \leqslant M$, the observational noise is a zero mean complex Gaussian random variable independent for different $\ell \geqslant 0$ with variance

$$\hat{r}^{\mathrm{o}} = (2M+1)^{-1} r^{\mathrm{o}} \tag{20}$$

with r^{o} the physical space observation noise variance. The important fact here is that the analysis of filtering for the reduced problem in (19) provides mathematically rigorous guidelines for filtering the original problem posed in 2.1–2.3 akin to the scalar test problems utilized in [6]; here, the sparse regular observations introduce the new feature of coupling among Fourier modes in the aliasing set in the canonical model. Since the original filter problem has only real coefficients, the reduced problem in (19) only needs to be studied or calculated for $0 \leqslant \ell \leqslant M$. Given a wave number ℓ, it is convenient to build the complex P-dimensional column vector

$$\vec{\hat{u}}(\ell) = \begin{pmatrix} \hat{u}^h_\ell \\ \hat{u}^h_{\ell+(2M+1)q_2} \\ \vdots \\ \hat{u}^h_{\ell+(2M+1)q_P} \end{pmatrix} \equiv \begin{pmatrix} \hat{u}^h_{k_1} \\ \hat{u}^h_{k_2} \\ \vdots \\ \hat{u}^h_{k_P} \end{pmatrix}, \tag{21}$$

where the various components in the aliasing set $\mathcal{A}(\ell)$ are indexed by the increasing magnitude of $k_j = \ell + (2M+1)q_j$, $j = 1, 2, \ldots, P$; the primary wave number associated with $\hat{u}^h_\ell$ is the first component of (21) and obviously has a special role. If we introduce the P-vector $\vec{g}_P = (1, \ldots, 1)^{\mathrm{T}}$, with the notation in (21) we rewrite (19)B) as given by

$$\vec{g}_P \cdot \vec{\hat{u}}((m+1)\Delta t) = \vec{g}_P \cdot \vec{\overline{\hat{u}}}^h_{m+1|m+1} + \hat{\sigma}^{\mathrm{o}}_{\ell,m+1}. \tag{22}$$

The standard simplest tests for stable filter performance in the canonical model are those for observability and controllability [1,8]. Since the noises $\sigma_{h,k,m+1}$ in (19)(A) have non-zero variance, controllability is always satisfied but there are still subtle issues of practical controllability elucidated in [6]; as noted already in [9], the observability condition can be very subtle for discretizations of PDE's. Observability for the canonical model in (19) involves the invertibility of the $P \times P$ matrix,

$$\mathcal{O}_p = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ F_{h,k_1} & F_{h,k_2} & \ldots & F_{h,k_P} \\ \vdots & \vdots & & \vdots \\ F^P_{h,k_1} & F^P_{h,k_2} & \ldots & F^P_{h,k_P} \end{pmatrix} \tag{23}$$

with

$$\det \mathcal{O}_p = \prod_{i<j} (F_{h,k_i} - F_{h,k_j}). \tag{24}$$

It is important to test observability in the perfect model scenario, where

$$F_{h,k} = \mathrm{e}^{\tilde{p}(\mathrm{i}k)\Delta t - \gamma(\mathrm{i}k)\Delta t}. \tag{25}$$

If there is a selective damping so that

the dissipation $\gamma(ik)$ strictly increases with k for $k > 0$,
then observability is always satisfied for canonical model(19) since det $\mathcal{O}_p \neq 0$. (26)

On the other hand, if there is uniform damping, as often occurs with Ekmann friction or radiative cooling in geophysical models [24,27], so that

$$\gamma(ik) = d \tag{27}$$

then observability is readily violated for appropriate observation times Δt, where det $\mathcal{O}_p = 0$. With (24), (25) and (27) we calculate that for uniform damping,

$$\det \mathcal{O}_p = e^{-(\binom{P}{2}C)d\Delta t} \prod_{k_i \neq k_j \in \mathcal{A}(\ell)} e^{\tilde{p}(ik_i)\Delta t} - e^{\tilde{p}(ik_j)\Delta t}, \tag{28}$$

where $\binom{P}{2}C \equiv P!((P-2)!2!)^{-1}$. Thus, from (28), observability fails in the perfect model with uniform damping at the observation time Δt, provided that, for some ℓ with $|\ell| < M$ and some q_j with $k_j = \ell + (2M+1)q_j$ belonging to the aliasing set, $\mathcal{A}(\ell)$, there is an integer Q so that

$$(\tilde{p}(ik_i) - \tilde{p}(ik_j))\Delta t = i(\omega_{k_i} - \omega_{k_j})\Delta t = Q2\pi i, \tag{29}$$

where $\tilde{p}(ik_j) = i\omega_{k_j}$. As we demonstrate next, the observability test often fails for wave-like systems with uniform damping. One important issue treated in this paper is the significance of the failure of the classical observability condition and its role in actual filter performance for the turbulent filtering problems with sparse regular observations described in 2.3.

The theoretical result in (26) provides one possible computational filtering strategy to avoid the failure of observability for the truth model in the uniformly damped case, i.e.,

filter the uniformly damped turbulent signal generated from (27) with a modified model with
selective damping $\tilde{\gamma}(ik)$ satisfying (26) and use $F_{h,k} = e^{\tilde{p}(ik)\Delta t - \tilde{\gamma}(ik)\Delta t}$ in the computational filter. (30)

Next, we demonstrate all of these phenomena in a test bed of models utilized throughout the remainder of the paper.

2.5. The stochastically forced dissipative advection equation

Here, we illustrate the issues in 2.4 for filtering solutions of the general damped, stochastically forced advection equation

$$\frac{\partial u(x,t)}{\partial t} = -c\frac{\partial u(x,t)}{\partial x} - d u(x,t) + \mu\frac{\partial^2 u(x,t)}{\partial x^2} + \overline{F}(x,t) + \sigma(x)\dot{W}(t). \tag{31}$$

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

$$\gamma(ik) = d + \mu k^2. \tag{32}$$

For a fixed $\mu > 0$ we have selective damping so that according to (26) observability is always satisfied. On the other hand, for $d > 0$ and $\mu \equiv 0$, we have uniform damping and the observability condition fails if (29) is satisfied, that is,

$$\Delta t = \frac{Q2\pi}{cj(2M+1)} \tag{33}$$

for any positive integer Q and fixed j with $1 \leqslant |j| \leqslant P$. Note that for the uniformly damped advection equation observability fails simultaneously for all wave numbers ℓ with $|\ell| \leqslant M$ and simultaneously for many pairs of wave numbers in a given aliasing set $\mathcal{A}(\ell)$. Thus, filtering turbulent solutions of (31) with small uniform damping at observation times satisfying (33) is a severe computational test problem discussed later in Section 6.

Other stringent test problems are achieved by utilizing *resonant mean forcing* in the truth signal in (16) so that

$$\omega_o(k) = \omega_k \quad \text{for all } k, |k| \leqslant M. \tag{34}$$

This is a severe test for filter performance utilized in the remainder of the paper especially when the observability condition is violated. In this paper, we also utilize non-resonant forcing with $\omega_o(k) = \omega_k + 1/2$ in (16).

2.6. Standard parameters

The standard values in the remainder of this paper are $M = 20, P = 3$ so that there are observations of the turbulent signal from (31) at 41 sparse regular observation points with 123 discrete mesh points for the approximate filter. In Section 5, we study the filter performance for a case where the true signals are solutions of the advection–diffusion equation with small diffusion coefficient,

$$\mu = 10^{-2} \quad \text{and} \quad d = 0. \tag{35}$$

In contrast, in Section 6 we analyze and study the filtering performance where the true signals are solutions of the uniformly damped advection equation with small damping

$$d = 10^{-2} \quad \text{and} \quad \mu = 0. \tag{36}$$

In testing the strategy outlined in (30) for filtering the true signal from the uniformly damped model with (36), we use $\mu = 10^{-2}$ while retaining $d = 10^{-2}$ so that according to (32), the filter model is selectively damped. The filter performance is also analyzed for power law turbulent energy spectra in the truth model with $E_0 = 1$ and the exponent β varying for $\beta = 5/3, 1/2, 0$, as successively more stringent tests.

3. Approximate filters in the Fourier domain

With the notation in (21) and (22), the canonical Fourier domain filtering problem for regularly spaced sparse observations from (19) is written concisely for the complex P-vector $\hat{u}^h$ as given by

$$\begin{aligned} &(\mathrm{A}) \quad \hat{u}^h_{m+1|m} = F_h \hat{u}^h_{m|m} + \overline{F}_m + \sigma_{h,m+1}, \\ &(\mathrm{B}) \quad \bar{v}_{m+1} = \vec{g}_P \cdot \overline{\hat{u}}^h_{m+1|m+1} + \hat{\sigma}^o_{m+1} \end{aligned} \tag{37}$$

where

$$\begin{aligned} &(\mathrm{A}) \quad \vec{g}_P = (1, 1, \ldots, 1)^{\mathrm{T}} \in \mathbb{R}^P, \\ &(\mathrm{B}) \quad \bar{v}_{m+1} = \vec{g}_P \cdot \overline{\hat{u}}((m+1)\Delta t) \end{aligned} \tag{38}$$

so that $\bar{v}_{m+1}$ is determined from the aliased components of the truth signal defined in Section 2.1. In (37)(A), F_h is the diagonal $P \times P$ complex matrix with entries determined by (19)(A).

The Fourier Domain Kalman Filter (FDKF) is simply the standard Kalman filter algorithm [1,8] applied to all the disjoint aliasing sets $\mathcal{A}(\ell)$ for all $0 \leqslant \ell \leqslant M$. In standard fashion, the $P \times P$ complex covariance matrix $R_{m|m}$ is updated in two steps: first, via the dynamics in (37)(A)

$$R_{m+1|m} = F_h R_{m|m} F_h^* + R_h, \tag{39}$$

where R_h is the diagonal matrix with entries $r_{h,k}$; then $R_{m+1|m}$ is updated to $R_{m+1|m+1}$ via the single observation in (37)(B) so that

$$R_{m+1|m+1} = R_{m+1|m} - \Lambda(R_{m+1|m}, \hat{r}^o) R_{m+1|m} (\vec{g}_P \otimes \vec{g}_P^{\mathrm{T}}) R_{m+1|m} \tag{40}$$

with the scalar factor $\Lambda(R_{m+1|m}, \hat{r}^o)$ given by

$$\Lambda(R_{m+1|m}, \hat{r}^o) = (\vec{g}_P \cdot R_{m+1|m} \vec{g}_P + \hat{r}^o)^{-1}. \tag{41}$$

In (40), $\vec{g}_P \otimes \vec{g}_P^{\mathrm{T}}$ is the rank one $P \times P$ matrix given by $(\vec{g}_P \otimes \vec{g}_P^{\mathrm{T}})\vec{w} = \vec{g}_P(\vec{g}_P \cdot \vec{w})$. The update of the mean, $\overline{\hat{u}}^h_{m+1|m+1}$, in the filtering process [1,8] is given by

$$\overline{\hat{u}}^h_{m+1|m+1} = \vec{K}_{m+1}\bar{v}_{m+1} + \overline{\hat{u}}^h_{m+1|m} - \vec{K}_{m+1}\vec{g}_P \cdot \overline{\hat{u}}^h_{m+1|m}, \tag{42}$$

where the provisional update of the mean $\overline{\hat{u}}^h_{m+1|m}$ is determined by the dynamics in (37)(A),

$$\overline{\hat{u}}^h_{m+1|m} = F_h \overline{\hat{u}}^h_{m|m} + \overline{F}_m. \tag{43}$$

The P-vector $\vec{K}_{m+1}$ is the Kalman gain vector [8]

$$\vec{K}_{m+1} = \Lambda(R_{m+1|m}, \hat{r}^{\mathrm{o}}) R_{m+1|m}\vec{g}_P. \tag{44}$$

The FDKF algorithm is already a reduced filter and much less expensive and more stable than implementing the Kalman filter or the ensemble Kalman filter such as ETKF [5] directly on the spatial domain (see Sections 5 and 6). Nevertheless, there are regimes of turbulent spectra and dynamics in the basic truth signal from (2.1) where this filter can be ill-conditioned and less accurate; thus, it is desirable to develop simpler approximations to FDKF as both alternative computational filters and for theoretical purposes. This is the topic of the remainder of this section.

3.1. The Strongly Damped Approximate Filters (SDAF, VSDAF)

The SDAF filter is motivated by the following situation which readily arises in practice with selective damping (see Sections 4 and 5). For a given P, assume that there are two separate groups in the dynamics (19)(A) with

$$\begin{aligned} &\text{(A)} \quad \text{Moderate damping}: \quad |F_{h,k_i}| = \mathcal{O}(1), \quad 1 \leqslant i \leqslant P_{\mathrm{o}} \\ &\text{(B)} \quad \text{Strong damping}: \quad |F_{h,k_i}| = \mathcal{O}(\epsilon) \ll 1, \quad P_{\mathrm{o}} + 1 \leqslant i \leqslant P. \end{aligned} \tag{45}$$

The assumption in (45)(B) means that there is a strong damping in a subset of the aliased modes; in this situation, the dynamic covariance update in (39) can be inaccurate because the cross-covariances between the first P_{o} components and the last strongly damped $P - P_{\mathrm{o}}$ components involve successive multiplications by large and small numbers to get order one quantities. The SDAF algorithm eliminates this potential numerical difficulty. It approximates the covariance matrix, $R_{m+1|m}$, by the block diagonal covariance matrix

$$R_{m+1|m} = \left(\begin{array}{c|c} R_{m+1|m,P_o} & 0 \\ \hline 0 & \begin{array}{ccc} r_{h,k_{P_o+1}} & & \\ & r_{h,k_{P_o+2}} & \\ & & \ddots \\ & & r_{h,k_P} \end{array} \end{array}\right) \tag{46}$$

This is the covariance matrix that results from replacing the dynamics in (19)(A) on the strongly damped modes, $P_{\mathrm{o}} + 1 \leqslant i \leqslant P$, by the "memoryless dynamics"

$$\hat{u}^h_{k_i,m+1|m} = \overline{F}_{k_i,m} + \sigma_{h,k_i}, \quad P_{\mathrm{o}} + 1 \leqslant i \leqslant P. \tag{47}$$

When the situation in (45) is satisfied, it is intuitively clear that the approximation in (47) is a reasonable one. With this approximation in (46), the filter scaling factor Λ in (41) becomes

$$\Lambda(R_{m+1|m}, \hat{r}^{\mathrm{o}}) = \left(\vec{g}_{P_{\mathrm{o}}} \cdot R_{m+1|m,P_{\mathrm{o}}}\vec{g}_{P_{\mathrm{o}}} + \sum_{i=P_{\mathrm{o}}+1}^{P} r_{h,k_i} + \hat{r}^{\mathrm{o}}\right)^{-1} \equiv \Lambda(R_{m+1|m,P_{\mathrm{o}}}, \hat{r}_{P_{\mathrm{o}}}) \tag{48}$$

with $\hat{r}_{P_o}$ the augmented noise

$$\hat{r}_{P_o} = \sum_{i=P_o+1}^{P} r_{h,k_i} + \hat{r}^o. \tag{49}$$

With the approximation in (45), the Kalman gain splits into contributions from moderately damped and memoryless (or strongly damped) modes and the gain vector is given by $\vec{K}_{m+1} = (\vec{K}_I, \vec{K}_{II})^T$ where

$$\begin{aligned} \vec{K}_I &= \Lambda(R_{m+1|m,P_o}, \hat{r}_{P_o}) R_{m+1|m,P_o} \vec{g}_{P_o}, \\ \vec{K}_{II} &= \Lambda(R_{m+1|m,P_o}, \hat{r}_{P_o}) r_{h,k_i}, \quad P_o + 1 \leqslant i \leqslant P. \end{aligned} \tag{50}$$

Furthermore, the $P_o \times P_o$ block of the covariance matrix $R_{m+1|m+1,P_o}$ is updated by utilizing the same formula in (40) applied to the P_o dimensional filtering problem. The important feature of the SDAF algorithm with the memoryless dynamics in (47) is that the entries of the covariance, $R_{m+1|m+1}$, corresponding to the memoryless modes never need to be computed. In the SDAF algorithm, the means of the memoryless modes are updated by (47) so

$$\overline{\hat{u}}^h_{k_i,m+1|m} = \overline{F}_{k_i,m}, \quad P_o + 1 \leqslant i \leqslant P. \tag{51}$$

This completes the construction of the SDAF algorithm since the mean state is updated through the general formula in (42) by utilizing (50) for the Kalman gain vector.

There is an important variant of the SDAF algorithm which needs to be mentioned here. Namely, one can use the memoryless approximation in (47) to update the covariance matrix in (46) but utilize the complete dynamics in (43) to update the mean state and calculate $\overline{\hat{u}}^h_{m+1|m}$; the mean state update $\overline{\hat{u}}^h_{m+1|m+1}$ is then achieved through the same formulas in (50) and (42). We call this filtering algorithm, the Variance Strong Damping Approximate Filter (VSDAF), since the strong damping is applied only for calculating the covariance matrix. Unlike SDAF, the VSDAF algorithm allows a weight for the dynamics of the memoryless modes in the filter.

Advantages of the Strong Damping Algorithms: First, the strong damping algorithms (both SDAF and VSDAF) only require actual Kalman filtering on a reduced subspace of P_o dimensions yet automatically yield an update of the mean state of the system on all the variables including all of the P–P_o memoryless variables. Furthermore, while we motivated these algorithms under the hypotheses in (45), the strong damping algorithms can be applied in any situation as a filtering approximation, even in circumstances where (45) is not satisfied (see Sections 4–6). Another attractive feature of these algorithms is that the subtle features of violation of observability, discussed earlier in Sections 2.4 and 2.5 only need to be analyzed for the reduced P_o dimensional filter; in particular, if $P_o = 1$, the strong damping algorithms always satisfy observability on the reduced subspace. Thus, the strong damping filters are an alternative computational strategy to the one proposed earlier in (30) with artificial selective damping as a remedy for failure of observability. The filter performance of both strategies is compared in subsequent sections.

Another important attractive feature of the SDAF becomes apparent when we recall the nature of the turbulent truth signals described in Section 2.1 which we are attempting to filter. We claim that

> the SDAF algorithm automatically provides a reasonable consistent estimate for the energy in the memoryless modes for any turbulent spectrum in the truth signal. (52)

To support the claim in (52), we use the formula in (38)(B) for the mean signal from the truth model together with (50) and (42) to calculate that the strong damping algorithm produces the following estimate for the mean of the memoryless modes (assuming zero external forcing in (47)),

$$\text{(A)} \quad \overline{\hat{u}}^h_{k_i,m+1|m+1} = \Gamma_i \sum_{k \in \mathcal{A}(\ell)} \hat{u}_k((m+1)\Delta t), \quad P_o \leqslant i \leqslant P \tag{53}$$

with

$$\text{(B)} \quad \Gamma_i = r_{h,k_i} \Lambda(R_{m+1|m,P_o}, \hat{r}_{P_o}).$$

Note that in general, the weights satisfy $\sum \Gamma_i < 1$. Consider for illustration purposes, the difficult case of an equipartition spectrum where all the r_{h,k_i} are equal; then all the Γ_i are also identical and the SDAF

approximation automatically distributes the compressed truth signal, $\sum_{k\in\mathcal{A}(\ell)}\hat{u}_k$, equally among the memoryless modes, which is intuitively consistent with equipartition of energy. Similarly, for decaying spectrum, the SDAF approximation automatically provides less weight from the compressed truth signal in the memoryless modes with less turbulent energy in a weighted fashion roughly consistent with this spectrum.

3.2. The reduced Fourier domain Kalman filter

The RFDKF approximation is based on the intuitive idea that for a sufficiently rapid decay in the turbulent spectrum of the truth signal, the primary mode contains the most energy so only this mode should be actively filtered. Thus, RFDKF always trusts the dynamics in (37) for all the aliased modes $2 \leqslant i \leqslant P$ yielding a Kalman gain vector with the form

$$\vec{K}_{m+1} = (K_{m+1}, 0, 0, \ldots, 0)^{\mathrm{T}} \tag{54}$$

while the primary mode, $\hat{u}^h_{k_1}$, is filtered by the one-dimensional algorithm,

$$\begin{aligned} &\text{(A)} \quad \hat{u}^h_{k_1,m+1|m} = F_{h,k_1}\hat{u}^h_{k_1,m|m} + \overline{F}_{k_1,m} + \sigma_{h,k_1,m+1},\\ &\text{(B)} \quad \bar{v}_{m+1} - \sum_{i=2}^{P}\bar{\hat{u}}_{k_i,m+1|m+1} = \bar{\hat{u}}_{k_1,m+1|m+1} + \hat{\sigma}^{\mathrm{o}}_{m+1} \end{aligned} \tag{55}$$

This type of reduced filtering algorithm was introduced earlier for theoretical purposes in a different context in [14]. Observability is always satisfied for RFDKF provide $0 < |F^h_1|$.

The computational advantage of RFDKF is apparent since only a one-dimensional Kalman filter is needed and the main input of the modes which are not observed by the sparse observations is to increase the resolution of the actual dynamics. In the rest of this paper, we refer to the one-dimensional primary mode as the resolved mode since the single observation is available and the trivially updated aliased modes as the unresolved modes since the observations are not available here. To simplify the jargon in the special case of the strong damping approximations with $P_{\mathrm{o}} = 1$ (as discussed in Section 3.1), we also refer to the moderately damped mode as the resolved mode and the memoryless modes as the unresolved modes.

3.3. Comparison of approximate filter algorithms

Compared with RFDKF which only trusts the dynamics for the unresolved (or aliased) modes, the SDAF approximation with $P_{\mathrm{o}} = 1$ has an opposite design principle of "memoryless dynamics" but yields a non-trivial estimate for the mean on the unresolved modes based on the observations. For the special case with $P_{\mathrm{o}} = 1$, the VSDAF uses the explicit Kalman gain vector in (53) to weight the unresolved dynamics and the effect of sparse observations in the filter dynamics as a natural blending between SDAF and RFDKF. Both the VSDAF with $P_{\mathrm{o}} = 1$ and RFDKF algorithms require detailed Kalman filtering on the primary mode alone and satisfy observability. Furthermore, the VSDAF algorithm with $P_{\mathrm{o}} = 1$ involves only modest additional processing through the explicit formulas in (53) beyond the RFDKF algorithm yet produces a reasonable state estimate for the unresolved Fourier modes. The filter performance of SDAF, VSDAF, and RFDKF as well as FDKF is studied in Section 4 for the canonical filter test problem; the results in Section 4 provide guidelines for filtering performance of these algorithms on the important PDE test problems in Sections 5 and 6.

4. Filter performance on the canonical test model

We follow the strategy utilized in [6,25] where the filtering behavior of discrete approximations were studied in various stiff parameter regimes for a canonical complex scalar model as practical guidelines for filter performance for PDE's with plentiful observations. Here, we study the filter performance of the various

approximations developed in Section 3 on the canonical test model for sparse regularly spaced observations in (19).

4.1. Selective damped signal

In this section, we discuss the filter performance on a canonical P-dimensional problem where the truth signal is a solution of a selective damped process. In particular, we generate the true signal by solving the advection–diffusion equation (31) in Fourier space with standard parameters $d = 0$ and $\mu = 10^{-2}$ as in (35) and for our canonical problem, we check the filtering results on aliasing sets $\mathcal{A}(1) = \{1, -40, 42\}$ and $\mathcal{A}(11) = \{11, -30, 52\}$ (see Fig. 1) where $P = 3$ and $M = 20$. Specifically, the true initial state is a solution of the advection–diffusion after 100 steps and the filter initial condition is randomly drawn from a Gaussian random variable with the true state as the mean and equilibrium kinetic energy as the variance. The data assimilation is then run for 10,000 steps with observation time step Δt.

For our numerical test, we choose three different observation times $\Delta t = 10^{-3},\ 0.1$, and 0.9. In Fig. 2, we see that these observation times include different dynamical phases compared to the decorrelation time $T_{corr} = \gamma_k^{-1}$. The short observation time $\Delta t = 10^{-3}$ yields highly correlated dynamics in each wave number, while with the long observation time $\Delta t = 0.9$, the dynamics in wave numbers higher than 10 are totally uncorrelated with their initial conditions. Additional stringent tests on this problem involve three different control signals:

4.1.(A) Decaying mean $A = 0$ in (16).
4.1.(B) Non-zero resonant periodic forcing on observed modes: $A = 0.1, \omega_o = \omega_k$ for all $|k| \leqslant M$.
4.1.(C) Non-zero non-resonant periodic forcing on observed modes: $A = 0.1, \omega_o = \omega_k + 1/2$ for all $|k| \leqslant M$.

The observations are simulated following (37)(B) and (38). We check the robustness of the filter toward variations of observation noise variance $\hat{r}^o$ from 0.05 to 3. Since the maximum equilibrium kinetic energy spectrum is 1, the values $\hat{r}^o = 0.05, 0.5, \ldots$ represent noise variance of sizes 5%, 50%, ... of the climatological energy spectrum E_k defined in (7), respectively. In each regime, we compare four strategies: FDKF, SDAF,

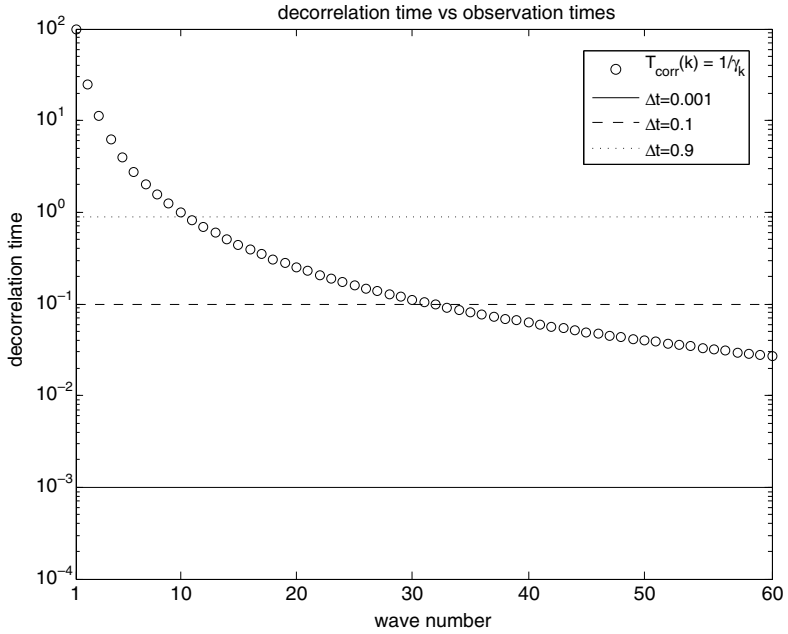

Fig. 2. Decorrelation time of a selective damped signal from the advection–diffusion equation ($\gamma_k = \mu k^2$ for $\mu = 0.01$) compared with the observation times, both plotted as functions of model wave number k.

VSDAF, and RFDKF. Finally, we note that for both strong damping strategies (SDAF and VSDAF), we consider only $P_o = 1$ since we want the filter to satisfy the observability condition, as in RFDKF.

We perform simulations in the aliasing sets, $\mathcal{A}(1)$ and $\mathcal{A}(11)$, and utilize the average RMS error (spatio-temporally averaged RMS difference between the mean posterior state $\overline{\hat{u}}^h_{m|m}$ and the true state $\hat{u}(m\Delta t)$) as the performance indicator. These are reported in Fig. 3 for $\mathcal{A}(11)$. We find the following general trends: when the energy spectrum is rough with equipartition spectrum $E_k = 1$ and the observation time is short $\Delta t = 10^{-3}$, the best strategy is to use the full FDKF. However, the VSDAF is reasonably skillful; the errors are comparable to that of the full FDKF. As the observation time is increased $\Delta t \geqslant 0.1$, the strong damping SDAF is comparable to the first two methods. When the turbulent spectrum is $E_k = k^{-1/2}$ we still see similar trends except that the advantages of FDKF over VSDAF/SDAF and even over the reduced filter RFDKF becomes less important. As the turbulent spectrum becomes smoother $E_k = k^{-5/3}$, we can hardly distinguish the performance between all the four strategies we compared; hence, one may choose either RFDKF or SDAF with $P_o = 1$ for computational efficiency. In Fig. 3, as a representative of the overall trends, we show results for aliasing set $\mathcal{A}(11)$ and observation times $\Delta t = 0.001, 0.1, 0.9$ for various energy spectra. In each panel, we plot the average RMS errors of each filtering strategy as functions of observation errors $\sqrt{\hat{r}^o}$. Since the general filter performances mentioned earlier are quite robust toward variation of $\hat{r}^o$ and the various different mean states 4.1.(A), (B), and (C) in the truth signal, in Fig. 3 we only show the results for non-resonant forcing in 4.1.C).

The poor performance of RFDKF for the short observation time, small observational noise, and an equipartition spectrum is due to the failure of practical controllability; this phenomenon already occurred in [6] with time discretized noise. Recall that controllability implies non-zero system noise variance; as we mentioned in Section 2.4, each mode is controllable for this diagonal Fourier model. However, some modes may have

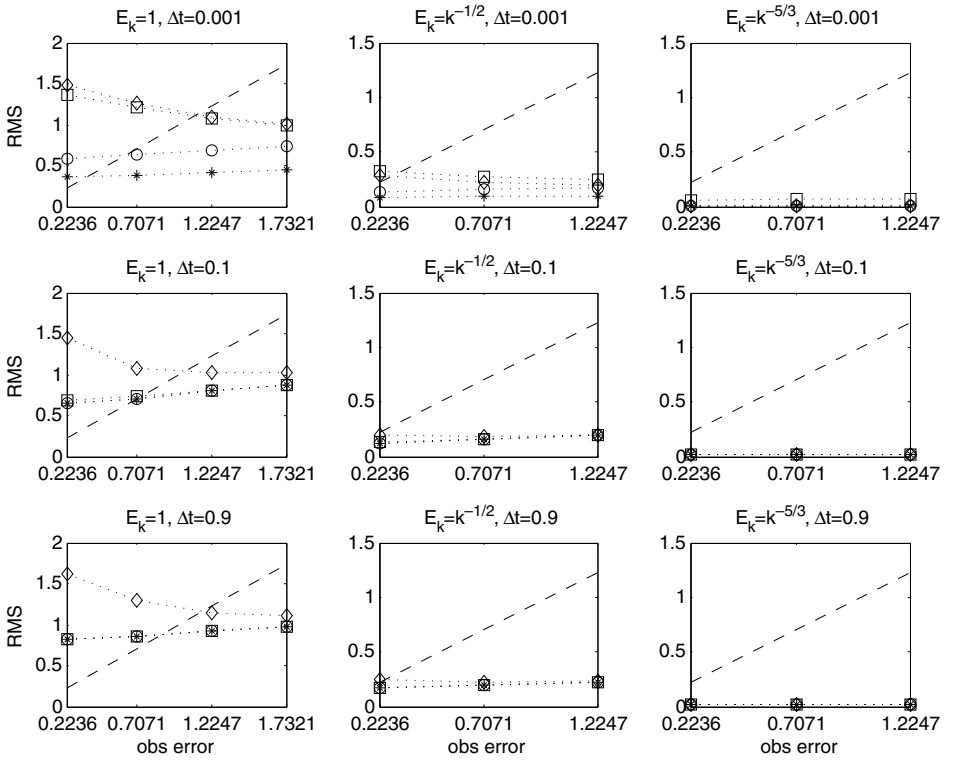

Fig. 3. Selective damped signal: RMS errors as functions of observation error $\sqrt{\hat{r}^o}$ for non-resonant periodic forcing and aliasing set $\mathcal{A}(11)$. The first column for $E_k = 1$, the second column for $E_k = k^{-1/2}$, and the third column for $E_k = k^{-5/3}$. The first row describes performance for observation time $\Delta t = 0.001$, the second row for $\Delta t = 0.1$, and the third row for $\Delta t = 0.9$. In each panel, FDKF is denoted by '*', SDAF in square, VSDAF in circle, RFDKF in diamond, and the dashes denote the observation error size.

very small $\mathcal{O}(10^{-5})$ non-zero system noise variance, so practical controllability is violated and this violation induces an unskillful or non-convergent filtered solution whenever $|F_{h,k_j}| \approx 1$ [1]. With a very short observation time $\Delta t = 0.001$, the system noise variances for the unresolved modes for $E_k = 1$ are much larger than that of the primary mode

$$r_{h,k_j} \gg r_{h,k_1} = r_{h,\ell}, \quad P_o + 1 = 2 \leqslant j \leqslant P. \tag{56}$$

Thus, the reduced filter consists of dynamics with $|F_{h,k_1}| \approx 1$ and the smallest system noise variance and clearly it is a less controllable one-dimensional filter compared to the full P-dimensional filter. The very high skill of VSDAF, even for $\Delta t = 10^{-3}$, occurs because it employs the full dynamics in propagating $\hat{u}^h_{m|m}$ that contains non-zero Kalman gain components in the unresolved modes to the prior state $\hat{u}^h_{m+1|m}$. Now, when the observation time Δt is larger, SDAF becomes comparable to both FDKF and VSDAF since $|F_{h,k_j}| \approx 0$ for the unresolved modes $P_o + 1 \leqslant j \leqslant P$ and hence the strong damping approximation in the mean state becomes appropriate. On the other hand, whenever $E_k = k^{-5/3}$,

$$r_{h,k_j} \lessapprox r_{h,k_1} = r_{h,\ell}, \quad P_o + 1 = 2 \leqslant j \leqslant P, \tag{57}$$

which suggest that the reduced filters and the full filter have compatible controllability condition. Hence, it is obvious that these filters perform comparably.

4.2. Uniformly damped signal

The basic experimental set-up in this section is similar to that in Section 4.1 except that the true signal is a Fourier domain solution of the weakly damped advection equation (31) with standard parameters $d = 0.01$ and $\mu = 0$ as in (36). We now look at the P-dimensional filtering problem for aliasing sets $\mathcal{A}(1) = \{1, -40, 42\}$ and $\mathcal{A}(17) = \{17, -24, 58\}$ (see Fig. 1 for illustration). For this model, we check the filtering performance for three different observation times: a non-observable time $\Delta t = 2\pi/(2M+1) = 0.1532$ that satisfies (33) for $Q = 1$ and $j = 1$, an observable time $\Delta t = 0.5$, and a giant observable time step $\Delta t = 50$ with very infrequent observation which is a half of the damping time scale $T_{\text{corr}} = 1/\gamma_k = 1/d = 100$.

From the numerical experiments, we now see the effects of the different mean signals 4.1.(A), (B), and (C) toward the filtering performance. In Fig. 4, we present the filter performance for aliasing set $\mathcal{A}(17)$ and non-observable time $\Delta t = 2\pi/(2M+1)$. Without resonant periodic forcing, the filtering skill is RMS-wise very similar to the selective damped case. The full FDKF is the best filter for equipartition spectrum $E_k = 1$. For smoother energy spectra with $E_k = k^{-1/2}, k^{-5/3}$, all four strategies are comparable. When a resonant periodic forcing is added to the primary mode, the observability violation in the full filter severely degrades the filtering performance; RMS-wise, we see that the errors in FDKF are on average about 80 for equipartition spectrum, 60 for $k^{-1/2}$, and 40 for $k^{-5/3}$. We also note (without presenting the results) that on aliasing set $\mathcal{A}(1)$, the filter performances are qualitatively very similar to the decaying mean case or non-resonant periodic forcing case. The presence of a low frequency resonant periodic forcing in the aliasing set for small wave numbers ℓ in the primary mode does not degrade the filtering skill on this still rather long time interval of assimilation.

The presence of resonant periodic forcing also degrades VSDAF in the equipartition spectrum for small observation noise size (5% of the equilibrium spectrum) and non-observable time $\Delta t = 0.1532$ (see Fig. 4). To understand this fact, let us consider the following scaling analysis. When $E_k = 1$, according to formula (14),

$$r_{h,k_j} \sim 10^{-2}, \quad P_o + 1 \leqslant j \leqslant P.$$

Hence, the amplitude of the unresolved components of the Kalman gain are

$$\vec{K}_{\text{II}} \sim 10^{-1}$$

since each component in (48) and (50) is on the order of 10^{-2}. Now suppose that the signal in the resolved mode is on the order of 10^1 while the unresolved modes are on the order of 10^0; this situation reflects the case with a resonant periodic forcing signal. Thus, the observation is on the order of $v \sim 10^1$ since it is generated via (37) and (38), and therefore, the updated mean state is shifted by

$$\vec{K}_{\text{II}} v \sim 10^0, \tag{58}$$

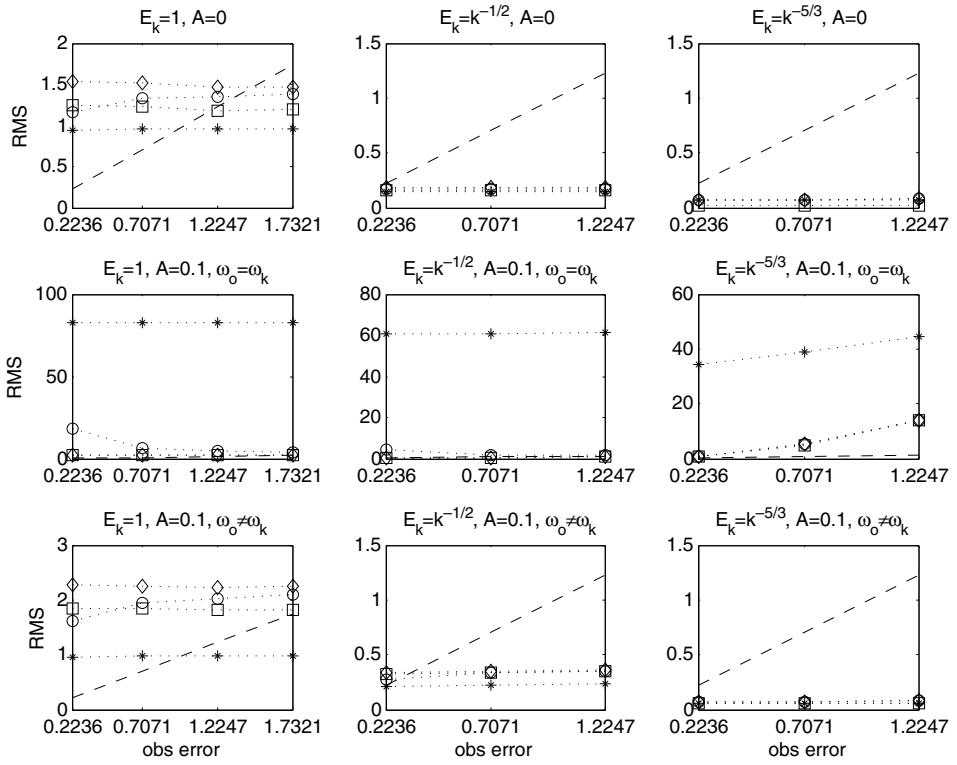

Fig. 4. Uniformly damped signal: RMS errors as functions of observation error $\sqrt{\hat{r}^o}$ for non-observable time $\Delta t = 0.1532$ and aliasing set $\mathcal{A}(17)$. The first column for $E_k = 1$, the second column for $E_k = k^{-1/2}$, and the third column for $E_k = k^{-5/3}$. The first row describes performance for decaying mean signal 4.1.(A), the second row for resonant periodic forcing 4.1.(B), and the third row for non-resonant periodic forcing. In each panel, FDKF is denoted by '*', SDAF in square, VSDAF in circle, RFDKF in diamond, and the dashes denote the observation error size.

which is relatively large for the unresolved modes. Now, since the memory in the VSDAF is given by a uniformly damped filter, the dynamical operator has an amplitude $|F_{h,k_j}| \approx 1$ in the unresolved directions. Hence this large drift is carried into each data assimilation cycle and it is the main contributor for the less accurate filter performance in VSDAF in this regime.

On the other hand, if all of the modes are of the same order 10^0, then the observations also have the same order $v \sim 10^0$. In this situation, the additional Kalman weights in the unresolved components should improve the filtering skill because the $\vec{K}_{II} v$ is in appropriate order for all modes and it is intuitively clear that an updated state should be weighted more toward the observations when the observations are relatively accurate $\hat{r}^o = 0.05$. In Fig. 4, in fact, we see a slight advantage of VSDAF over SDAF with $E_k = 1$ and $\hat{r}^o$ small for a decaying mean signal and a non-resonant periodic forcing. Back to the resonant periodic forcing case; when $\hat{r}^o$ is increased, the unresolved components of the Kalman gain $\vec{K}_{II}$ become small and meaningless. In contrast to (58), the updated mean state is shifted by a much smaller $\vec{K}_{II} v$ so that the memory in VSDAF does not degrade the filtering skill. In SDAF, the Kalman weight $\vec{K}_{II}$ is never used in (42) for the mean update on the unresolved modes because SDAF is a memoryless filter there. This argument also holds for justifying the skillful solution in the RFDKF since the Kalman gain for the unresolved components are zero.

Now, let us focus our interest on the regime with smooth spectrum $E_k = k^{-5/3}$ where the presence of resonant periodic forcing degrades the filter significantly even when observability is satisfied in the RFDKF and SDAF. A similar trend is also observed for the other observable times $\Delta t = 0.5, 50$. In this regime, the system noise variance for the resolved mode is on the order of 10^{-5} while $|F_{h,k_i}| \approx 1$ for $1 \leqslant i \leqslant P$, so practical controllability is already an issue even for a one-dimensional filtering problem. Furthermore, this failure is

magnified when the observation noise variance is large since the filter tends to trust the dynamics, i.e., $\vec{K}_{k_j,\infty} \approx 0$, which then implies marginal stability factor $|F_h(\mathcal{I} - \vec{K}_\infty \vec{g}_P^{\mathrm{T}})| \approx 1$ and this definitely degrades the filter convergence [1,6]. In Section 6, we will go back into this issue for small $\hat{r}^o$ and show that it is in fact possible to obtain a much more skillful filtered solution when practical controllability is reinstalled; this was done systematically in [6] for discretizations with model error through an information criterion for the setting with plentiful observations.

4.2.1. Modified Model Using the Selective Damped filter

Now, let us consider the strategy described in (30): that is, we make a model error and modify the model slightly by adding a selective damping through a diffusive coefficient $\mu = 10^{-2}$ for the filter model. As it was mentioned before, this strategy is motivated by the fact that the selective damped model is always observable (see (26)). In Fig. 5, we show numerical results of filtering with FDKF and RFDKF for the modified filter and compare it to SDAF and VSDAF for the unmodified filter with $\mu = 0$ which also satisfy observability.

From our numerical simulations, we find that when the main signal is either a decaying mean or a non-resonant periodic forcing, the modified full filter is comparable to the unmodified FDKF (compare Fig. 4 for unmodified filter and Fig. 5 for modified filter in the same aliasing set $\mathcal{A}(17)$ and non-observable time $\Delta t = 0.1532$) for equipartition energy spectrum. For smoother spectrum $E_k = k^{-5/3}$, there is a slight advantage of using the modified filter compared to strong damping strategy, however, it is not significant.

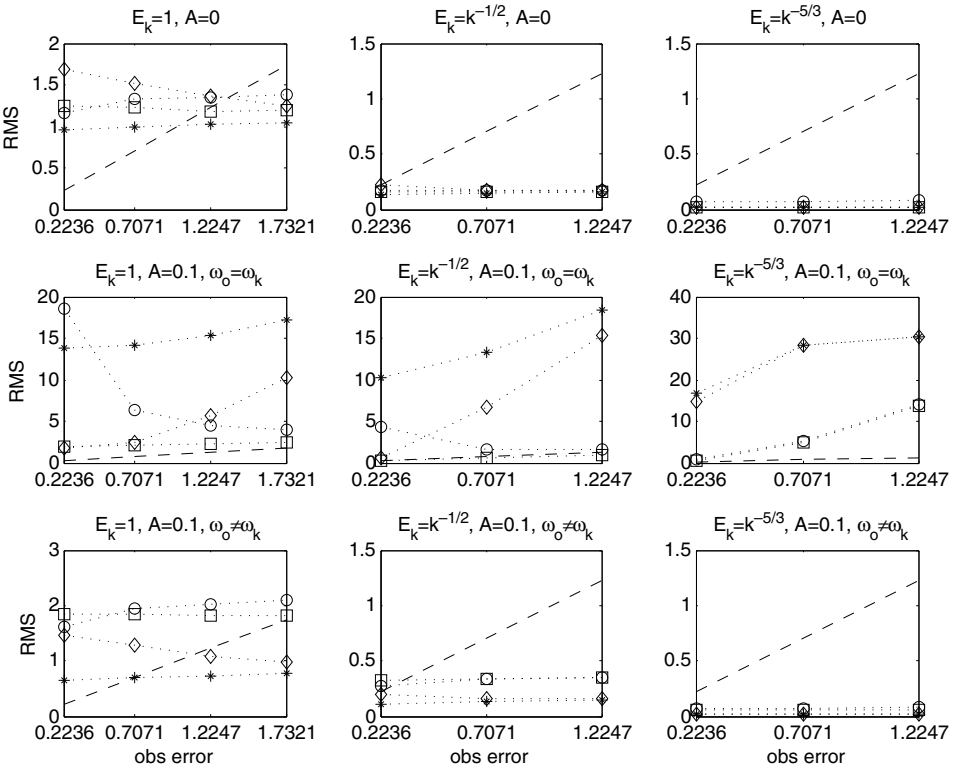

Fig. 5. Filtering uniformly damped signal with modified filters: RMS errors as functions of observation error $\sqrt{\hat{r}^o}$ for non-observable time $\Delta t = 0.1532$ and aliasing set $\mathcal{A}(17)$. The first column for $E_k = 1$, the second column for $E_k = k^{-1/2}$, and the third column for $E_k = k^{-5/3}$. The first row describes performance for decaying mean signal 4.1.(A), the second row for resonant periodic forcing 4.1.(B), and the third row for non-resonant periodic forcing. In each panel, modified FDKF is denoted by '*', SDAF in square, VSDAF in circle, modified RFDKF in diamond, and the dashes denote the observation error size.

In the presence of resonant periodic forcing, this approach completely fails because the signal to noise ratio is large and the additional damping introduces significant model errors. Additionally, when the resonant periodic forcing accumulates in time, the mean model error from the difference between the two mean periodic forcing solution $\overline{F}_{k,m}(\gamma_k)$ and $\overline{F}_{k,m}(\tilde{\gamma}_k)$ (see (17)), where $\tilde{\gamma}_k = \gamma_k + \mu k^2$, becomes important.

4.3. Summary on the canonical test model

By simply using the RMS error as the performance measure, we conclude that in general we always have robust skillful results in filtering the selective damped signal with the full filter FDKF; however, the cheaper reduced filter VSDAF performs comparably including the difficult case of equipartition spectrum where RFDKF performs poorly. On the smoother spectrum, the cheaper reduced filters have comparable skill to FDKF.

When the uniformly damped signal is considered, an additional complexity is introduced via the violation of observability for FKDF. In general, we find that when the signal to noise ratio is small (e.g., with decaying mean or non-resonant periodic forcing), FDKF is the best method for equipartition spectrum although SDAF has significant skill; all the cheaper reduced filters give comparable performance to FDKF for smoother spectrum since the signal to noise ratio for smoother spectrum is slightly larger than the signal to noise ratio for equipartition spectrum, especially in the high wave numbers. In this regime, it is also advantageous to use a modified model as suggested through (30) with additional damping term $\mu = 10^{-2}$ although the improvement is very slightly. When the signal to noise ratio is large (e.g., with resonant periodic forcing), the full filter FDKF fails completely since the observability condition is practically violated. The filter with the additional damping term with $\mu = 10^{-2}$ also fails even when the observability condition is reinstalled since the model errors are too large when the signal amplitude is large. In this situation, SDAF and RFDKF produce the best filter performance since they both satisfy the observability condition and do not have model errors on the primary mode.

In the following two sections, we will use these results as theoretical guidelines for filtering the spatio-temporal PDE systems with sparse observations. Note that in that context, we will consider an additional measure beyond the average RMS error for assessing the filter performance, the spatial pattern correlation of the filtered solution and the actual turbulent signal.

5. Filter performance on the stochastically driven advection–diffusion equation

In this section, we discuss the selective damped filtering problem considered in Section 4.1 on the PDE level; in particular, we combine all of the P-dimensional filtering problems for aliasing sets $\mathcal{A}(\ell)$, $\ell = 0, 1, \ldots, M$, and study the filter performance in the spatial domain. We adopt the same standard parameters for an advection–diffusion equation and the same experimental design set-up as in Section 4.1 except that now we only run the assimilation for $T = 1000$ steps. We will only show results for the smallest $\hat{r}^{\mathrm{o}} = 0.05$ since this corresponds to observation noise variance $r^{\mathrm{o}} = 2.05$ in the spatial domain while the signals are on the order of 10^0–10^1. For larger r^{o}, we find that the filter is mostly weighted to the dynamics so that it is quite often that the average RMS errors are small (smaller than the observation noise size) but the filtered solution may not be as accurate. We will only show results for equipartition spectrum $E_k = 1$ and smooth spectrum $E_k = k^{-5/3}$ since the results with intermediate spectrum $E_k = k^{-1/2}$ behave in between the two former spectra we consider.

To measure the filtering performance on spatially extended systems for sparse observations, we cannot rely solely on the RMS error and compare it to the observation noise size, especially when the signal to noise ratio is large. In the rest of this paper, we will consider one additional measure for filter performance: the spatial correlation between the mean corrected state $\bar{u}^h_{m|m}(jh)$ and the true signal $u(jh, m\Delta t)$, for $j = 0, 1, \ldots, 2N$. The spatial correlation between discrete functions is their discrete inner product divided by the product of their discrete mean square norms; this is a measure of how closely the spatial graphs of the two functions resemble each other with a correlation at the maximum value one measuring perfect coincidence of the graphs with a positive scale factor [24,25]. Thus, an accurate filtered solution is one with small RMS error and high temporal average spatial correlation.

5.1. Ensemble transform Kalman filter

The extended Kalman filter is not practically useful when the model state space is on the order of 10^6–10^7, which is often the case in an operational global weather model, since its covariance update involves multiplication of a matrix of size $10^6 \times 10^6$ (or $10^7 \times 10^7$) in each data assimilation cycle. In order to assess the performance of our cheap filtering strategies, we consider an operationally plausible scheme, Ensemble Transform Kalman Filter (ETKF), which is introduced in [5] and advocated in [21] with spatial local analysis for computational efficiency and to avoid spurious correlation between two distance away grid points, as a benchmark. In our Fourier domain filters, we basically ignore the correlation between the Fourier coefficients in different aliasing sets and there are no obvious relations between a particular grid location in the physical space with particular set of Fourier coefficients. For a fair comparison, therefore, we only implement ETKF globally as in [5] with ensemble update as in [32,15,16].

In our experiments, we test ETKF with various ensemble sizes $K = 100, 150, 300$, and 500 if necessary whereas the model state space is $2N + 1 = 123$. For small ensemble size and a nonlinear model, it is often found that the uncertainty is underestimated [20,4,33]. A simple remedy is to use a multiplicative covariance inflation. For ETKF, we implement the multiplicative covariance inflation as prescribed in [15]. In our experiments, we compare and vary the covariance inflation r from 0% to 40% even when the model is linear and the ensemble sizes are larger than the model state space. In Table 1, we show average RMS errors of extensive runs in a regime ($E_k = k^{-5/3}$, $\Delta t = 0.1$, and decaying mean) where ETKF is expected to excel. We find that the RMS errors are slightly below the observation noise $\sqrt{r^o} = 1.43$ for experiments with ensemble sizes 100 and 150, however the highest average spatial correlation is only 0.55 and it is achieved by the experiment with the lowest RMS error ($r = 40\%$ and $K = 100$). Furthermore, as shown in Fig. 10, there can be filter divergence for ETKF with resonant periodic forcing even for the $k^{-5/3}$ spectrum and for selectively damped advection–diffusion equation. In the following section, we will see that the Fourier domain strategies never diverge in this regime and always produce more skillful results compared to ETKF. For rough spectrum $E_k = 1$, the situation is worse and ETKF has very little skill even for a decaying mean state and for an ensemble size of $K = 500$. The lowest average RMS errors is roughly 11.28 and the highest average spatial correlation is only 0.38.

5.2. Numerical results

From our numerical experiments, we find that the overall results are parallel to the guideline from the ODE in Section 4.1. For short observation time $\Delta t = 10^{-3}$, FDKF dominates in both spectra we considered but the cheaper VSDAF is a very skillful filter with reasonably comparable performance to the FDKF. That is, their RMS errors are comparable while in terms of temporal average spatial correlation, FDKF is on average about 0.15 better than VSDAF (see Fig. 6 for decaying mean signal and equipartition spectrum $E_k = 1$), and both of them are far superior to RFDKF and SDAF. With such a short observation time, the signal is on the order of 10^0 with or without the resonant periodic forcing so the effect of mean forcing is negligible here. Here, for the first time we see an important facet of filter skill with sparse observations which occurs repeatedly in Section 5 and 6; namely in Fig. 6 the RMS errors for FDKF and VSDAF are above the observation error, yet the average correlation values 0.78 and 0.64 indicate substantial skill for these filters. This is in contrast to the situation with plentiful observations [6] where RMS errors below the observational noise are the signature of filter skill.

Table 1
RMS errors of filtering the stochastically driven advection–diffusion equation with ETKF in regime with smooth spectrum $E_k = k^{-5/3}$, decaying mean, and observable time $\Delta t = 0.1$ for various ensemble size K and covariance inflation r

RMS error	$r = 0\%$	$r = 10\%$	$r = 20\%$	$r = 30\%$	$r = 40\%$
$K = 100$	1.41	1.34	1.31	1.26	1.22
$K = 150$	1.37	1.30	1.29	1.26	1.24
$K = 300$	1.35	1.46	1.54	1.59	–

The maximum temporal average spatial correlation is 0.55 and is achieved by the experiment with lowest RMS errors with $K = 100$ and $r = 40\%$.

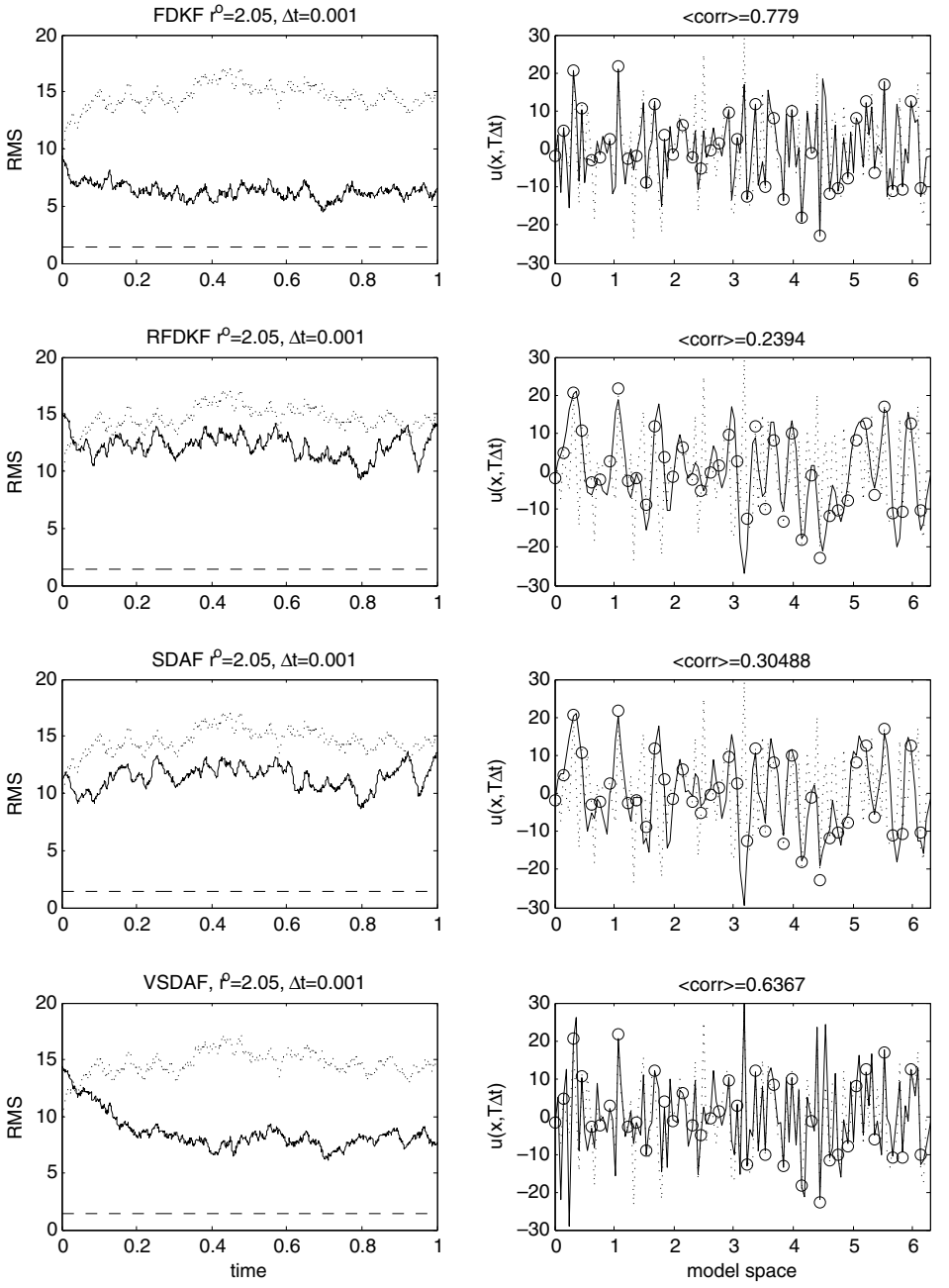

Fig. 6. Advection–diffusion equation for $\Delta t = 10^{-3}$, $E_k = 1$, and decaying mean. The first column shows RMS errors as functions of time for unfiltered solution (dotted line), the filtered solutions (solid), and observation noise size (dashes). The second column shows the snapshots of the filtered solution (corresponds to the filtering strategy stated in the subtitle of each panel in the first column for the same row) as functions of model spatial domain at time 1 (solid), the true state (dotted), and the sparse observations (circle).

In Figs. 7 and 8, we show results for $\Delta t = 0.1$ and decaying mean signal for experiments with $E_k = 1$ and $E_k = k^{-5/3}$ spectra, respectively, as benchmarks. Note that for the equipartition spectrum, the unfiltered solution has an average RMS error 15.55 and average spatial correlation almost zero, while in the smooth spectrum $E_k = k^{-5/3}$, the unfiltered solution has RMS error 2.46 and average spatial correlation 0.1. Here, the results, again, reflect those of the ODE case in Section 4.1, that is: FDKF, SDAF, and VSDAF are all superior for $E_k = 1$ with comparable performance and an average spatial correlation about 0.7 while RFDKF has

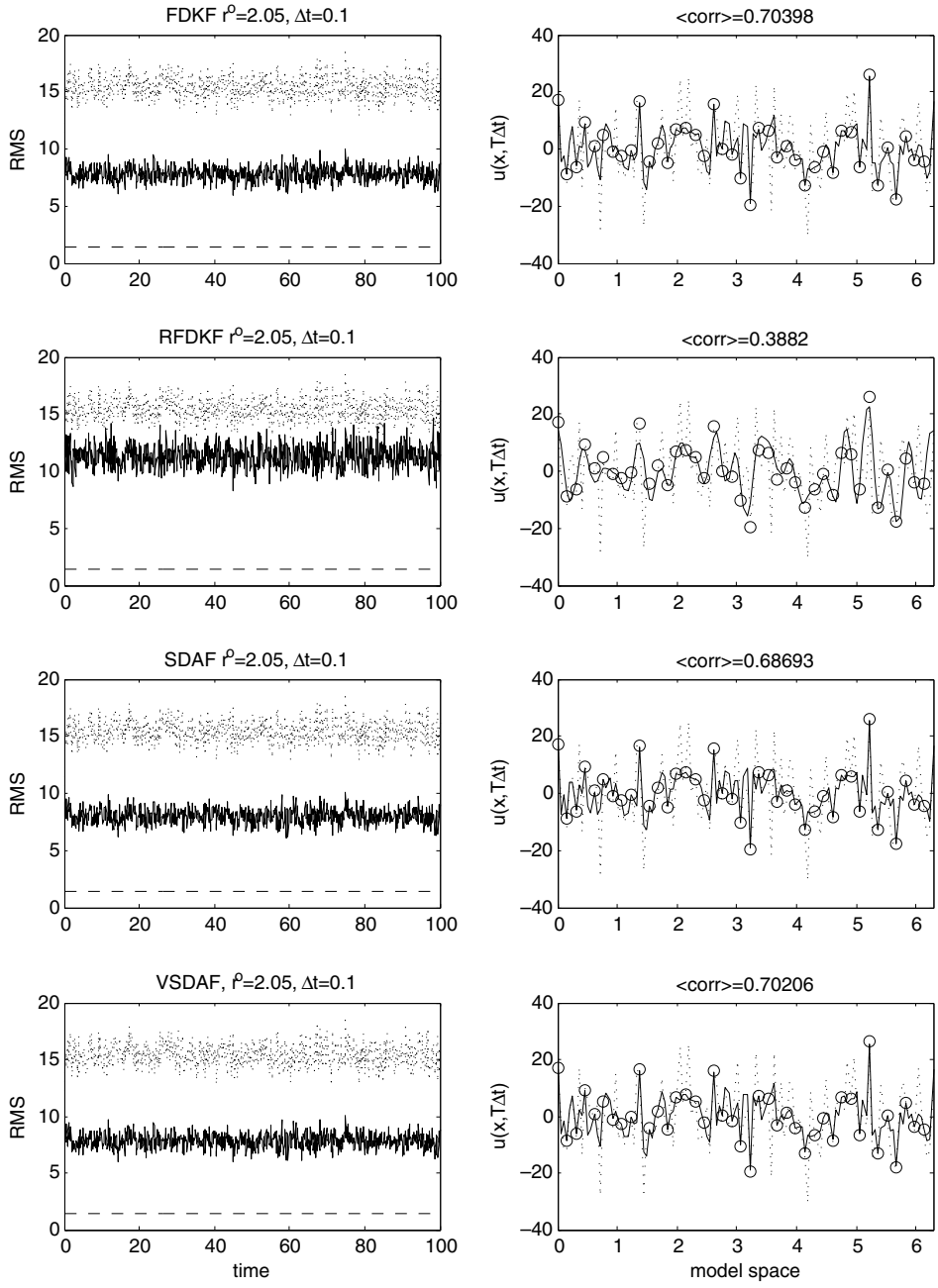

Fig. 7. Advection–diffusion equation for $\Delta t = 0.1$, $E_k = 1$, and decaying mean. The first column shows RMS errors as functions of time for unfiltered solution (dotted), the filtered solutions (solid), and observation noise size (dashes). The second column shows the snapshots of the filtered solution (corresponds to the filtering strategy stated in the subtitle of each panel in the first column for the same row) as functions of model spatial domain at time 100 (solid), the true state (dotted), and the sparse observations (circle).

much lower average correlation, only 0.39, and larger RMS errors. All our schemes are comparable for smoother spectrum $E_k = k^{-5/3}$ with average spatial correlation 0.86 as shown in Fig. 8. The results with the non-resonant periodic forcing signal are very similar to the decaying mean case since in both cases the signal to noise ratios are relatively small and comparable.

Now, we discuss the case where the signal to noise ratio is large that reflects the resonant periodic forcing. Notice that for the rough spectrum $E_k = 1$, the filtered solution for the decaying mean signal has average

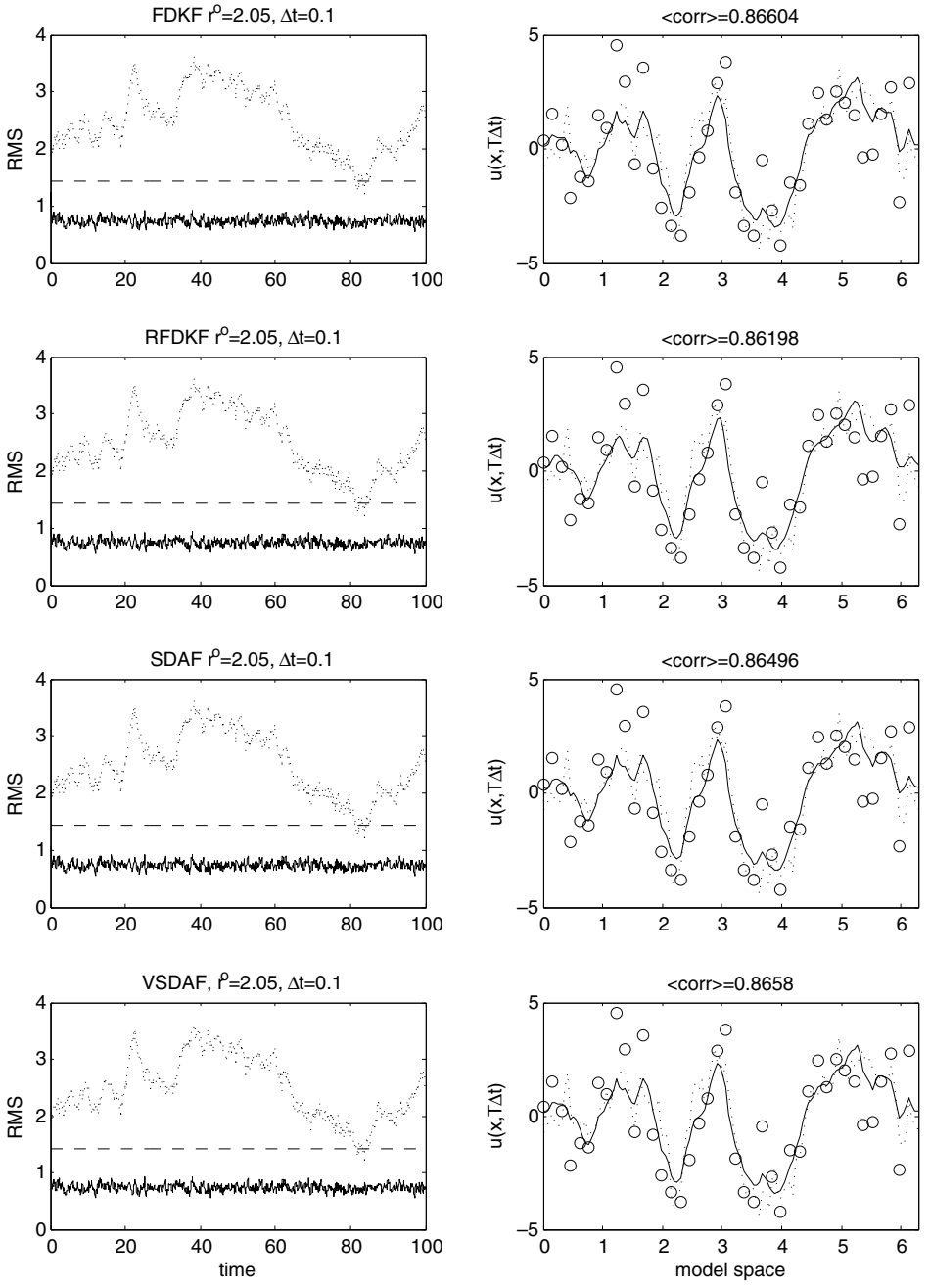

Fig. 8. Advection–diffusion equation for $\Delta t = 0.1$, $E_k = k^{-5/3}$, and decaying mean. The first column shows RMS errors as functions of time for unfiltered solution (dotted), the filtered solutions (solid), and observation noise size (dashes). The second column shows the snapshots of the filtered solution (corresponds to the filtering strategy stated in the subtitle of each panel in the first column for the same row) as functions of model spatial domain at time 100 (solid), the true state (dotted), and the sparse observations (circle).

amplitude of order 10^1 while in the smoother spectrum $E_k = k^{-5/3}$, it is on average on the order of 10^0 (again, see Figs. 7 and 8). Since the resonant periodic forcing for this observation time is on the order of 10^1, it will not affect the amplitude of the solutions in the rough spectrum as much as in the smoother spectrum. In Fig. 9, we show results for $E_k = 1$ with resonant periodic forcing signal. As one can see, the signal amplitude does not increase significantly and the highest average spatial correlation increases from 0.70 to 0.79 compared to

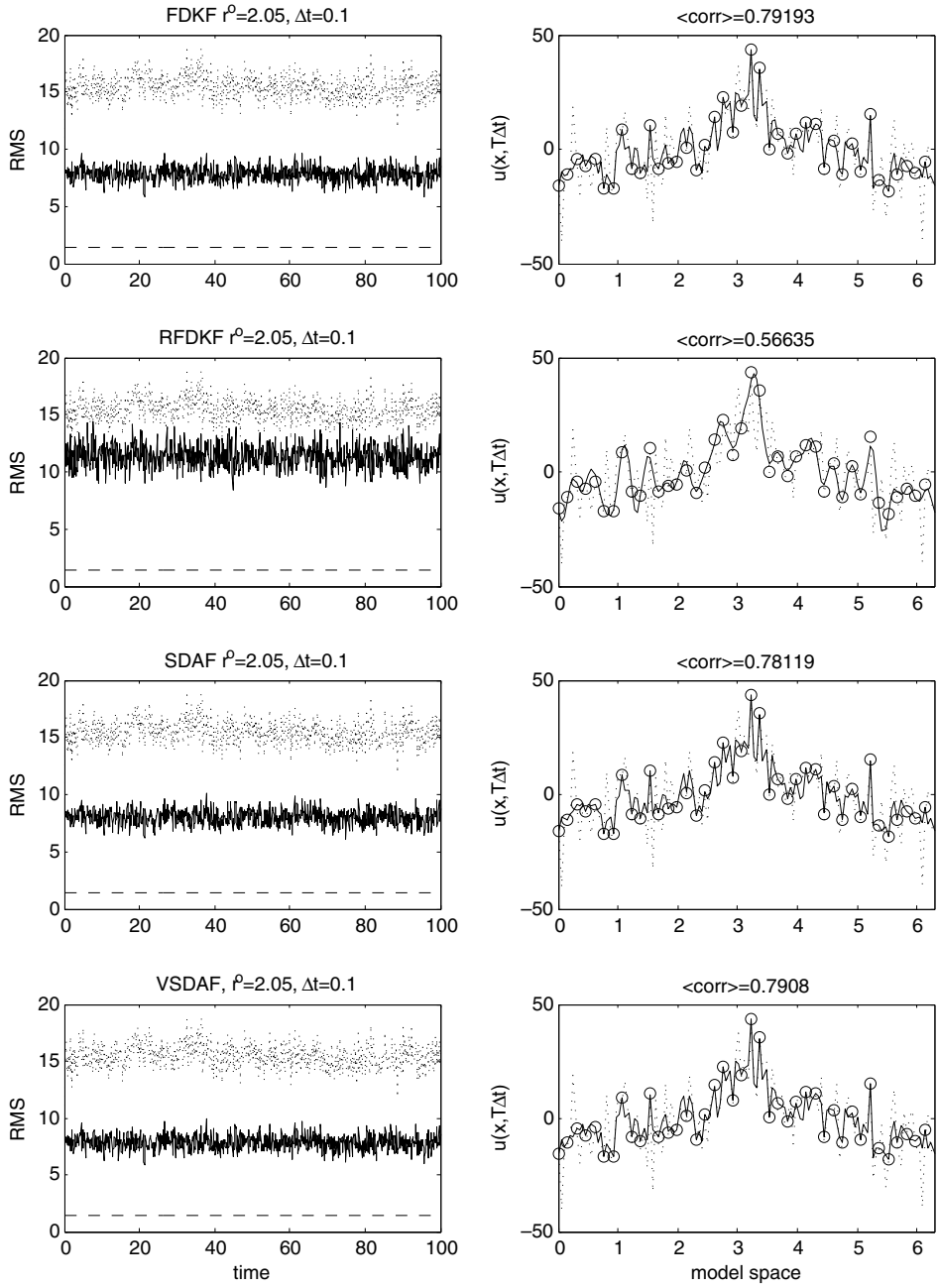

Fig. 9. Advection–diffusion equation for $\Delta t = 0.1$, $E_k = 1$, and resonant periodic forcing. The first column shows RMS errors as functions of time for unfiltered solution (dotted), the filtered solutions (solid), and observation noise size (dashes). The second column shows the snapshots of the filtered solution (corresponds to the filtering strategy stated in the subtitle of each panel in the first column for the same row) as functions of model spatial domain at time 100 (solid), the true state (dotted), and the sparse observations (circle).

the decaying mean case (see Fig. 7). This increase of correlation skill is not so surprising considering that the unfiltered solution average spatial correlation also increases from 0 to 0.3 when a resonant periodic forcing is added but the skill of the RFDKF still is significantly lower. For smoother spectra, the additional resonant periodic forcing increases the signal to noise ratio by one order of magnitude and it makes the four schemes we compare almost indistinguishable. What is intriguing is that when the signal to noise ratio is this large, the

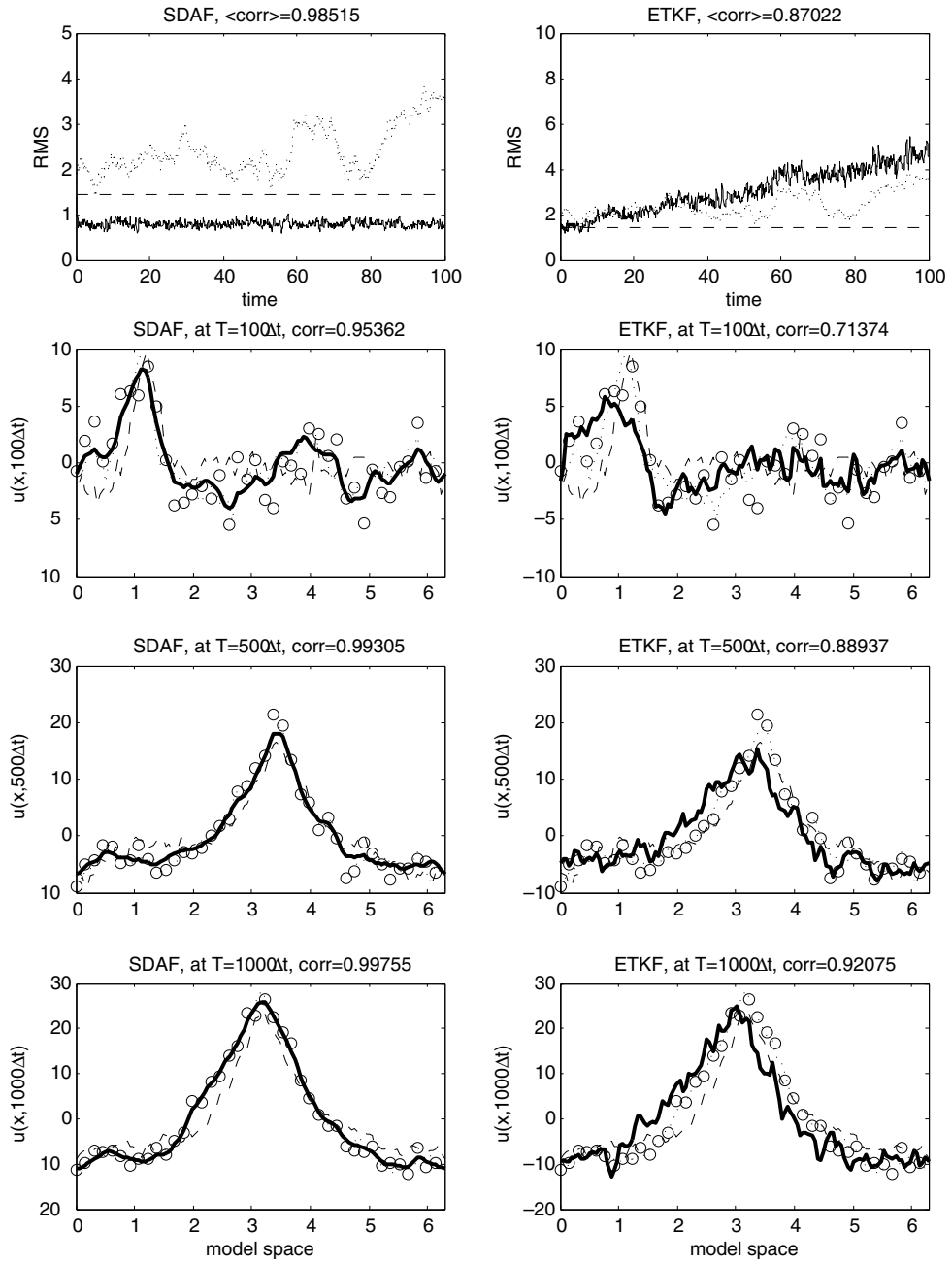

Fig. 10. Advection–diffusion equation for $\Delta t = 0.1$, $E_k = k^{-5/3}$, and resonant periodic forcing. The first column shows the results with SDAF while the second column with ETKF with $K = 150, r = 40\%$. The first row shows the RMS errors as functions of time for no filter in solid line, the unfiltered solutions (dotted), and observation noise size (dashes). The last three rows show the snapshots of the filtered solution (thick solid) as functions of model spatial domain at time 10, 50, and 100 (or after $T = 100, 500$, and 1000 cycles), respectively, and compare it to the unfiltered solution (dashes), the true state (dotted), and the sparse observations (circle).

unfiltered solutions have RMS error 2.4 and have a very high average correlation 0.93. In fact, the correlation skill converges to one as time grows. Hence, the Kalman filter seems to be redundant if one is interested only in the equilibrium state. However, for practical real time prediction problems the accurate transient filtering of the location and amplitude of the large spike has obvious practical importance. Based on this point of view, in Fig. 10, we plot snapshots at transient times (after 100 and 500 assimilation cycles) as well as after 1000

assimilation cycles for SDAF. The numerical results suggest a tremendously high filtering skill with SDAF even after 1000 assimilation cycles (first column of Fig. 10) whereas ETKF diverges with $K = 150, r = 40\%$ (second column of Fig. 10) and the unfiltered solutions (green curves) are nowhere near the true signals.

When Δt is increased to 0.9 so there are infrequent observations, the system noise variances for equipartition spectrum become large and hence the signal to noise ratio becomes much smaller so that one expects a substantial filter degradation for the decaying mean case and the non-resonant periodic forcing case. In fact, we found that in this regime the spatial correlation decreases to 0.56. For a smoother spectrum, the system noise variances are much smaller for high wave numbers k so the filtering skills for $\Delta t = 0.1$ and $\Delta t = 0.9$ are almost identical since the signal to noise ratio is almost identical even when Δt is larger than 0.1. For resonant periodic forcing, as Δt is increased, the signal becomes larger since $\overline{F}_k \sim \Delta t$.

5.3. The sampling effect

In all of the numerical experiments above, we test the Fourier domain filtering strategies by assuming that they all have perfect realization, in other words, we use formula (43) directly and we assume that there are no sampling effects which may not be true in more complicated nonlinear problems. In this section, we assess the filter performance with finite number of realizations (technically, the mean prior forecast state is computed by taking an ensemble average of $\hat{u}^h_{m+1|m}$ propagated with (37)) and we compare it to the unbiased solution computed directly via (43).

In Fig. 11, we show numerical results for 1 and 10 realizations and compare it to the ideal situation with perfect realization for a regime with $E_k = 1$, decaying mean, and $\Delta t = 0.1$. This regime is considered to be a difficult test since the signal to noise ratio is rather small and the uniform spectrum will ensure a rather high fluctuation in the unresolved directions. From Fig. 11, we see that the simulation with 10 realizations is almost indistinguishable compared to the simulation with perfect realization. The sampling effect is notable when the numerical simulation is employed with one realization, however, the error difference is relatively small and acceptable considering that we only use a single realization. Thus, the sampling effect is negligible in filtering sparse observations with small ensemble size provided the ensemble has about 10 members. This result is robust for all of the alternative filters we test in this paper in all their regimes.

5.4. Summary on the advection–diffusion equation as a test bed

The overall results in this section are directly predicted by simply looking at the corresponding canonical P-dimensional filtering problem as shown in Section 4.1. In particular, the filtering skills are not affected by the different mean forcing for the signals in 4.1.(A), (B), and (C). The filtering skill of the various filtering strategies depends on the observation time Δt, compared to the correlation time at each wave number in Fig. 2 as well as the energy spectrum. For the short observation time $\Delta t = 0.001$ where all Fourier modes in the truth signal are substantially correlated, FDKF has significant skill and VSDAF has nearly comparable skill. For moderate observation times $\Delta t \geqslant .1$ so that some of the high wave number modes in the truth signal are completely decorrelated, for the rough spectrum, $E_k = 1$, FDKF, VSDAF, and SDAF all have comparable significant skill while RFDKF has significantly less skill; for $\Delta t \geqslant 1$ and the smoother spectrum $E_k = k^{-5/3}$, all the filters FDKF, VSDAF, SDAF, and RFDKF have significant comparable skill. Thus, if one is interested in computational efficiency, these results suggest the use of the alternative filters VSDAF, SDAF, and RFDKF in various regimes of observation time and spectrum. As indicated in Fig. 11, all of these results are robust for filtering the mean signal with only 10 ensemble members. Furthermore, the filter skill of these alternative methods in the appropriate regime exceeds that of ETKF implemented in physical space; this is true for ETKF with the optimum variance inflation parameter from Table 1 to give the best filter correlation skill for ETKF with a large ensemble size. Furthermore, as shown in Fig. 10, surprisingly, for resonant periodic mean forcing and the relatively smooth $k^{-5/3}$ spectrum, ETKF with large ensemble size suffers from practical filter divergence while all the alternative filters never exhibit divergence even with a small number of realizations.

We also find that the RMS errors are rather large for $E_k = 1$ compared to those with the smoother spectrum $E_k = k^{-5/3}$ for the filtering strategies we tested including ETKF. This is simply because the model with $E_k = 1$

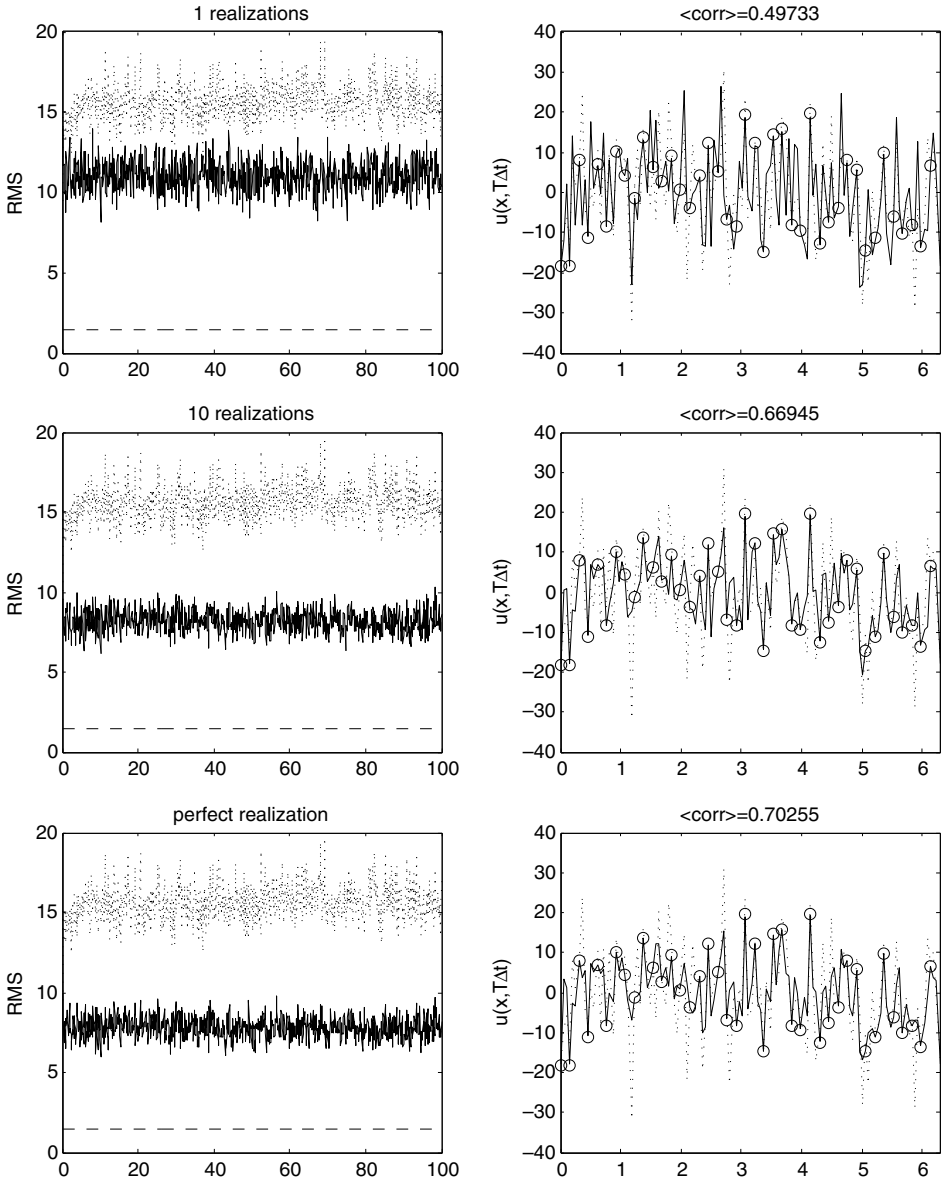

Fig. 11. Advection–diffusion equation for $\Delta t = 0.1$, $E_k = 1$, and decaying mean. The first row shows the RMS errors as functions of time and a snapshot at time 100 for simulation with one realization, the second row for simulation with 10 realizations, and the third row for simulation with perfect realization. The legend in each panel is similar to that in Fig. 9.

has relatively large system noise variances in the unresolved components and hence the signal to noise ratio is relatively small even for the case with resonant periodic forcing case. The difficulties with small signal to noise ratio are obvious since the filtered solutions are dominated by the noise independent of how the filtered solutions are weighted. Similar justification holds when the filtered solutions degrade for large observation time $\Delta t = 0.9$. Furthermore, the Fourier domain filtering strategies are more attractive computationally because the sampling effect is negligible for small ensemble size and they compute the exact covariance directly with the Kalman filter formula as opposed to ETKF which depends heavily on the ensemble size for covariance estimation.

6. Filter performance on the stochastically driven advection equation with weak damping

Parallel to Section 5, we now discuss the filter performance for sparse observations of a turbulent signal generated by the weakly damped advection equation in (31) and (36). One additional feature in this filtering problem is that for an appropriate Δt as in (29), the filtering problem with the perfect model is non-observable. Also with small uniform damping, unlike the situation in Section 5 displayed in Fig. 2, all observation times Δt here are below the correlation time at all spatial wave numbers. These features combine to make this problem an extremely difficult test bed especially with equipartition spectrum. As in Section 5, here, we only run the filter for 1000 assimilation cycles with $\hat{r}^o = 0.05$ and show results for $E_k = 1$ and $E_k = k^{-5/3}$. As in Section 5, we utilize the criteria in Section 4.2 for the canonical test problem to provide guidelines for this extremely stiff problem. We begin the discussion by showing the performance of ETKF as a benchmark in regimes where observability condition is fully satisfied. Subsequently, we discuss the filtering performances with perfect realization for observable times and for a non-observable time. In this section, we omit the discussion of finite ensemble effect since our numerical experiments suggest similar conclusion as in Section 5.3. We then conclude this section by a short summary.

6.1. ETKF performance for an observable time

The filtering problem in this section is designed to be the hardest test bed when the filter is simultaneously non-observable, has rough equipartition kinetic energy spectrum, and is dominated by a resonant periodic forcing signal. Here, we will show that even when these three conditions are not met, ETKF fails to produce a reasonably skillful solution. In our extensive runs with an observable time $\Delta t = 0.5$, decaying mean, and a smooth spectrum $E_k = k^{-5/3}$, the highest temporal average of the spatial correlation is only 0.34. In Table 2, the best average RMS errors from various ensemble sizes and covariance inflation are worse than the observation error $\sqrt{r^o} = \sqrt{(2M+1)\hat{r}^o} = 1.43$.

When $E_k = 1$, the results are even worse; the lowest average RMS error is about 12.65 and the highest average spatial correlation is 0.22. Again, we emphasize that these results are not in the severe regime yet since the observability is not violated and at the same time non-periodic forcing is not considered. In the following two sections, we will show that much better filtered solutions are obtainable with the Fourier domain filtering strategies. In the presence of a resonant periodic forcing we will also see divergence in the full FDKF but this divergence is avoidable when the reduced filter or strong damping approximations are considered or practical controllability is reinstalled in the appropriate context.

6.2. Fourier domain filter performance for observable times

When $\Delta t = 0.5$, the filtered solutions with $E_k = 1$ are very similar to the solutions of filtering the selective damped signals from Section 5 when the observation time Δt is below the correlation time of all wave numbers. In this case, the FDKF and VSDAF are the best filters for any spectrum with average spatial correlation of at least 0.98 and average RMS error of at most 1.5 (slightly above the observation error 1.43 but still far below the unfiltered solutions average RMS error 15.3), see Fig. 12. The three types of signals 4.1.(A), (B), and (C) do not affect the superiority of FDKF and VSDAF over the other strategies. This is a significant improvement over ETKF as discussed in 6.1. Note that when the smoother spectrum $E_k = k^{-5/3}$ is considered, the same conclusion holds for the decaying mean signal with the fixed observation noise.

Table 2
RMS errors of filtering stochastically driven weakly damped advection equation with ETKF in regime with smooth spectrum $E_k = k^{-5/3}$, decaying mean, and observable time $\Delta t = 0.5$ for various ensemble size K and covariance inflation r

RMS error	$r = 0\%$	$r = 10\%$	$r = 20\%$	$r = 30\%$	$r = 40\%$
$K = 100$	1.90	2.09	2.17	2.23	2.33
$K = 150$	2.02	2.01	2.08	2.16	2.31
$K = 300$	2.19	2.47	2.58	2.69	–

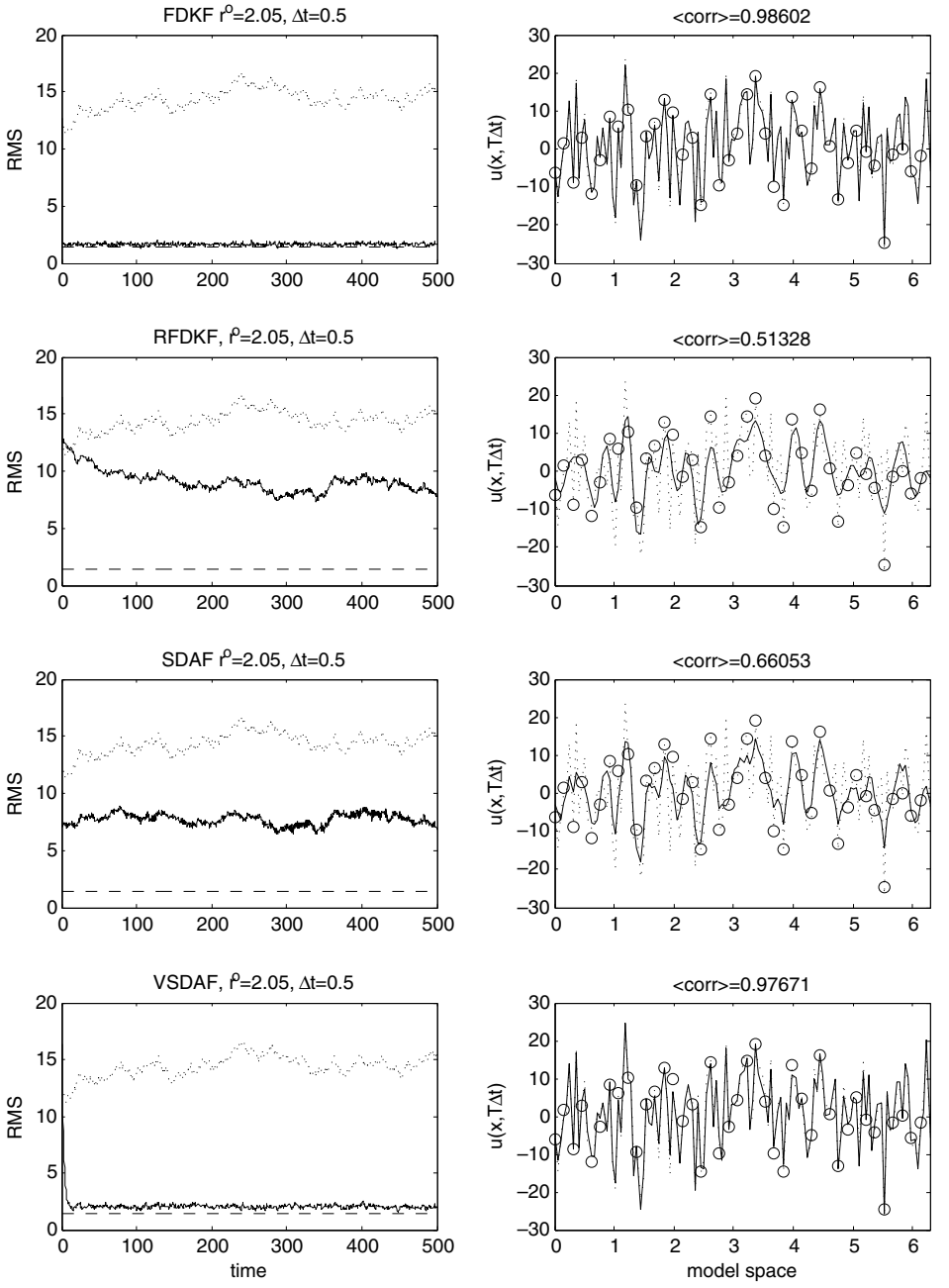

Fig. 12. Weakly damped advection equation for $\Delta t = 0.5$, $E_k = 1$, and decaying mean. The first column shows RMS errors as functions of time for unfiltered solution (dotted), the filtered solutions (solid), and observation noise size (dashes). The second column shows the snapshots of the filtered solution (corresponds to the filtering strategy stated in the subtitle of each panel in the first column for the same row) as functions of model spatial domain (solid), the true state (dotted), and the sparse observations (circle), all at time 500.

However, for $E_k = k^{-5/3}$, results vary for non-resonant and resonant periodic forcing. In the former case, the best results are found by strategies with additional selective damping proposed in (30) from Section 2.4 and discussed earlier at the ODE level in Section 4.2. Here, the average spatial correlation is 0.87 for RFDKF with $\mu = 0.01$ while the unmodified filters (with $\mu = 0$) have average spatial correlation of at most 0.75 (see Fig. 13). In the left panels, notice also that the modified filters (last two rows) have RMS errors slightly lower than those of the unmodified filters (first three rows). The right panels in Fig. 13 also show the improved capability

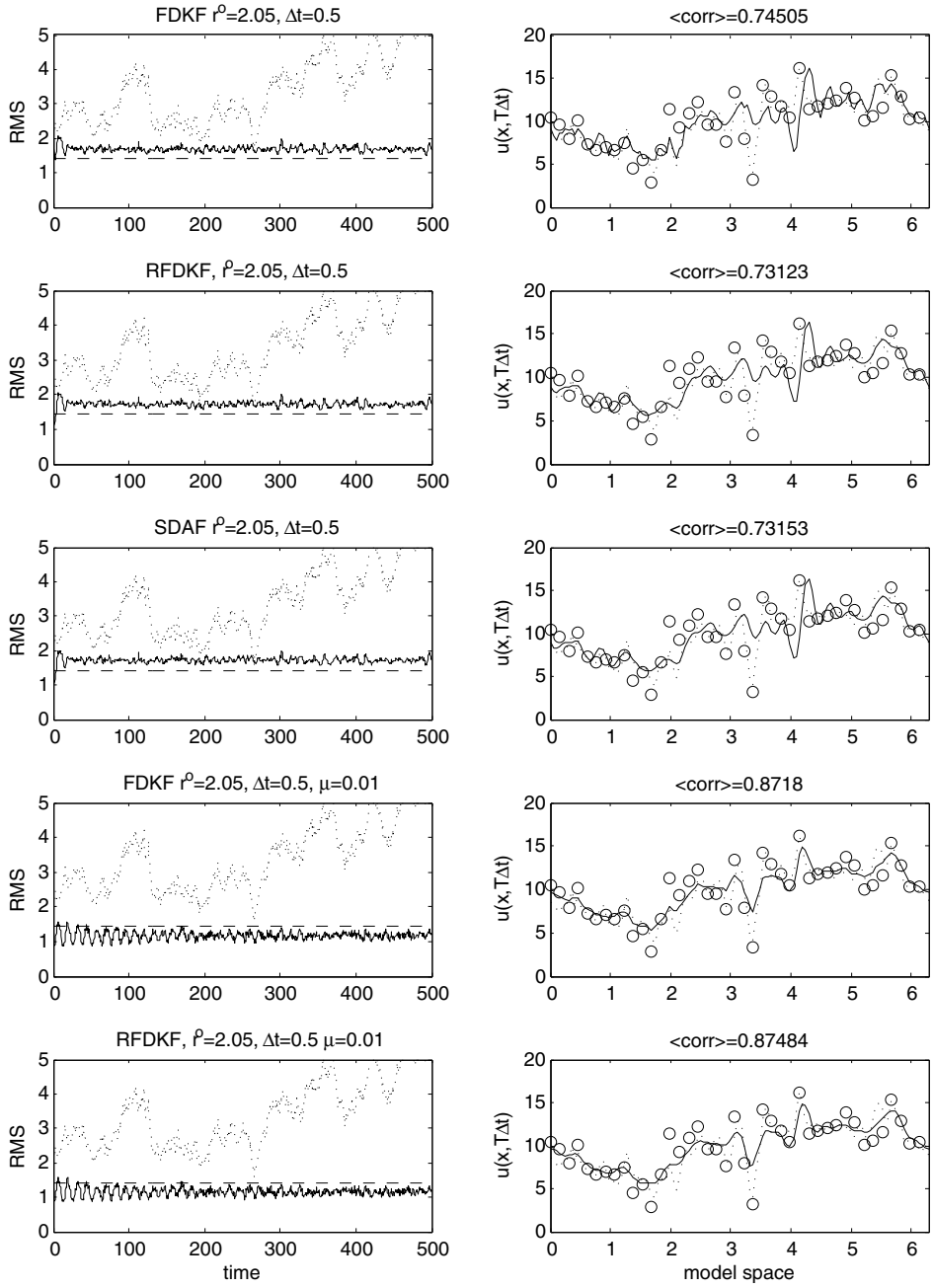

Fig. 13. Weakly damped advection equation for $\Delta t = 0.5$, $E_k = k^{-5/3}$, and non-resonant periodic forcing. The first column shows RMS errors as functions of time for unfiltered solution (dotted), the filtered solutions (solid), and observation noise size (dashes). The second column shows the snapshots of the filtered solution (corresponds to the filtering strategy stated in the subtitle of each panel in the first column for the same row) as functions of model spatial domain (solid), the true state (dotted), and the sparse observations (circle), all at time 500.

of the modified methods in capturing the negative spike in the truth signal. In this case, the model errors (contributed by the additional damping term and the difference between $\overline{F}_{k,m}$ in (17) for $\mu = 0$ and $\mu = 0.01$) are still small enough to produce skillful filtered solution. In the presence of resonant periodic forcing, all of the unmodified filters including RFDKF which fully satisfy the observability condition have an average spatial

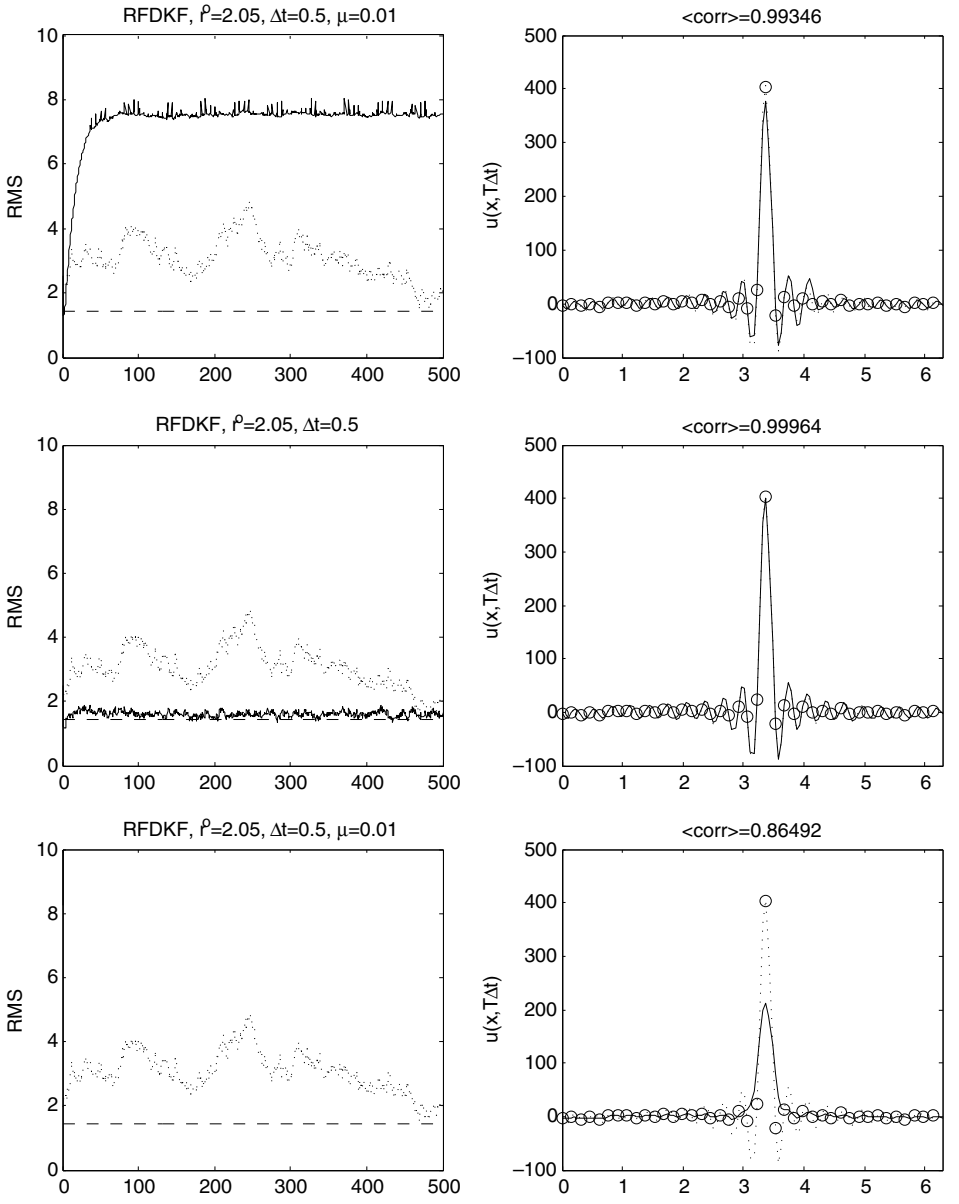

Fig. 14. RFDKF (first row), RFDKF with increased noise (second row), and the modified RFDKF (third row) on the weakly damped advection equation for $\Delta t = 0.5$, $E_k = k^{-5/3}$, and resonant periodic forcing. The first column shows RMS errors as functions of time for unfiltered solution (dotted), the filtered solutions (solid), and observation noise size (dashes). The second column shows the snapshots of the filtered solution (corresponds to the filtering strategy stated in the subtitle of each panel in the first column for the same row) as functions of model spatial domain (solid), the true state (dotted), and the sparse observations (circle), all at time 500.

correlation of 0.99, however, their RMS errors are much larger than the unfiltered solutions. These unsynchronized performance measures are caused by the lack of controllability as discussed in Section 4.2 for the test model. In Fig. 14, we show the RMS errors as functions of time and snapshots of filtered solutions after 1000 assimilation cycles in regime with $\Delta t = 0.5, E_k = k^{-5/3}$, and resonant periodic forcing for RFDKF. The first row is for the unmodified filter with $\mu = 0$, where $r_{k_1} \sim 10^{-4}$ for aliasing sets $\mathcal{A}(\ell)$ with $\ell \geqslant 5$ when computed with (14); the RFDKF filter is unmodified as well in the second row, however, we now increase the system noise variance of mode $k = 5, \ldots, M$ (that is, the first component of aliasing sets $\mathcal{A}(\ell)$ with $\ell \geqslant 5$) by

0.01. Here, we see a significant improvement of the filter with larger system noise variances. This issue has already been pointed out in our earlier work [6] where in there we utilize information criteria for choosing system noise variances. The modified filter with $\mu = 0.01$ (see third row of Fig. 14) introduces large model errors in the presence of resonant periodic forcing and this failure is understood and discussed at the ODE level in Section 4.2. In this large signal to noise regime for filtering, the modified models substantially damp the large localized spike in the solution as could be expected by a diffusive modified model.

When the observation time is large $\Delta t = 50$, the best strategies for each regime are similar to those suggested for $\Delta t = 0.5$, the only difference is that the average spatial correlations for such a long time are significantly smaller. This filter degradation is justified in similar fashion as in the case for $\Delta t = 0.9$ for an advection–diffusion equation, that is, the signal to noise ratio becomes smaller as Δt is increased.

6.3. Fourier domain filter performance for a non-observable time

In this section, we focus on the regime where observability is violated by setting the observation time to be $\Delta t = 2\pi/(2M+1) = 0.1532$. In a regime where the energy spectrum is relatively smooth $E_k = k^{-5/3}$, the observability condition is not needed to produce filtered solutions for FDKF with a decaying mean as shown in Fig. 15 and the results qualitatively resemble those in Section 6.2 for all the other approximate filters; these results mirror Section 4.2 for the test model. For non-resonant periodic mean forcing, at this non-observable time, Fig. 16 shows a degradation in filter performance for FDKF with the lowest correlation skill, only 0.67, and the highest RMS errors; however, there is no evidence for filter divergence for this spectrum for FDKF at this non-observable time. In Fig. 16, we see that the modified filter with $\mu = 0.01$ gives the best solution in the presence of non-resonant periodic forcing with average spatial correlation 0.91 while the unmodified filters are only able to produce solutions with correlation at around 0.77 with SDAF and RFDKF. RMS-wise, the errors of the unmodified filters are on average 1.59 and slightly larger than the modified filters 0.99. This slight advantage of the modified filter has already been noticed earlier in the ODE level in Section 4.2. In the presence of resonant periodic forcing, although it is not reported here in detail, we see the similar filtering failure as in Section 6.2 due to a lack of practical controllability of the filter. Note that in Figs. 15 and 16 we do not show VSDAF because it is not better than SDAF.

Now let us turn into the toughest regime for the test bed with rough spectrum $E_k = 1$, that is, simultaneous occurrence of identical noise strength in the energy spectrum of the turbulent signal and violation of observability. Secondly, we consider both the decaying mean and non-resonant periodic forcing for which the signal to noise ratio is small. In Fig. 17, we show results from simulations for decaying mean signal with FDKF and RFDKF, both for modified and unmodified models, and SDAF. Again, we do not show VSDAF because it is not better than SDAF. We see that the full filter FDKF with and without modification is the better filter although the average correlation is rather low 0.53.

The last regime we consider is the rough spectrum $E_k = 1$ with resonant periodic forcing so that the signal to noise ratio is large. At this unobservable time for resonant periodic forcing with a rough spectrum, we finally see strong filter divergence for FDKF and VSDAF as reported in Fig. 18 with unacceptably large ghost spikes in the filtered solution; the filter divergence in VSDAF was explained in detail in Section 4.2 through scaling analysis. Notice that the filter divergence that occurs in FDKF, VSDAF, and ETKF is avoidable with all the other alternative strategies. In Fig. 18, we see that the better solutions are produced by the reduced filters, especially SDAF with average spatial correlation 0.94 and RMS error 10.96 whereas the unfiltered solutions have average spatial correlation of 0.89 and RMS error 14.14. In this regime, we do not show results with modified filters since the presence of large signal to noise ratio introduces significant model errors due to the augmented diffusion as noted earlier.

6.4. Summary on the weakly damped advection equation as a test bed

In general, the filtering performance of all the Fourier diagonal filtering strategies reflects the guidelines given in Section 4.2 and these strategies supersede ETKF significantly. In the present setting with small uniform damping, the observation time Δt exceeds the correlation time at all wave numbers in the true signal.

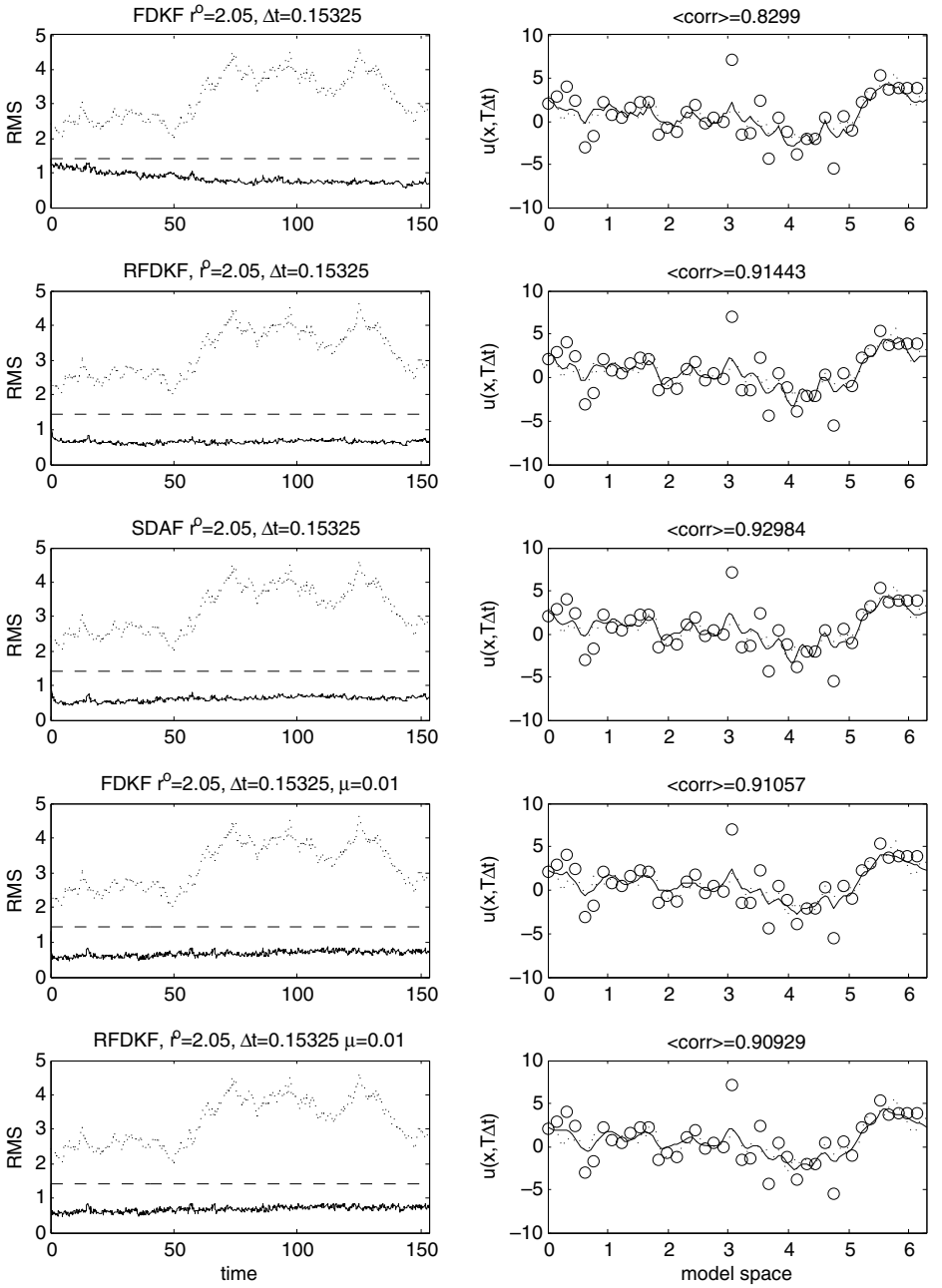

Fig. 15. Weakly damped advection equation for $\Delta t = 2\pi/(2M+1) = 0.1532$, $E_k = k^{-5/3}$, and decaying mean. The first column shows RMS errors as functions of time for unfiltered solution (dotted), the filtered solutions (solid), and observation noise size (dashes). The second column shows the snapshots of the filtered solution (corresponds to the filtering strategy stated in the subtitle of each panel in the first column for the same row) as functions of model spatial domain (solid), the true state (dotted), and the sparse observations (circle), all at time 153.2.

Thus, for the observable case, the best filtering strategies in each regime are similar to those for the advection–diffusion equation as discussed in Section 5; FDKF and VSDAF are the best choices for an equipartition spectrum (Fig. 12) while the other filters SDAF, RFDKF have comparable skill for a decaying spectrum (Fig. 13).

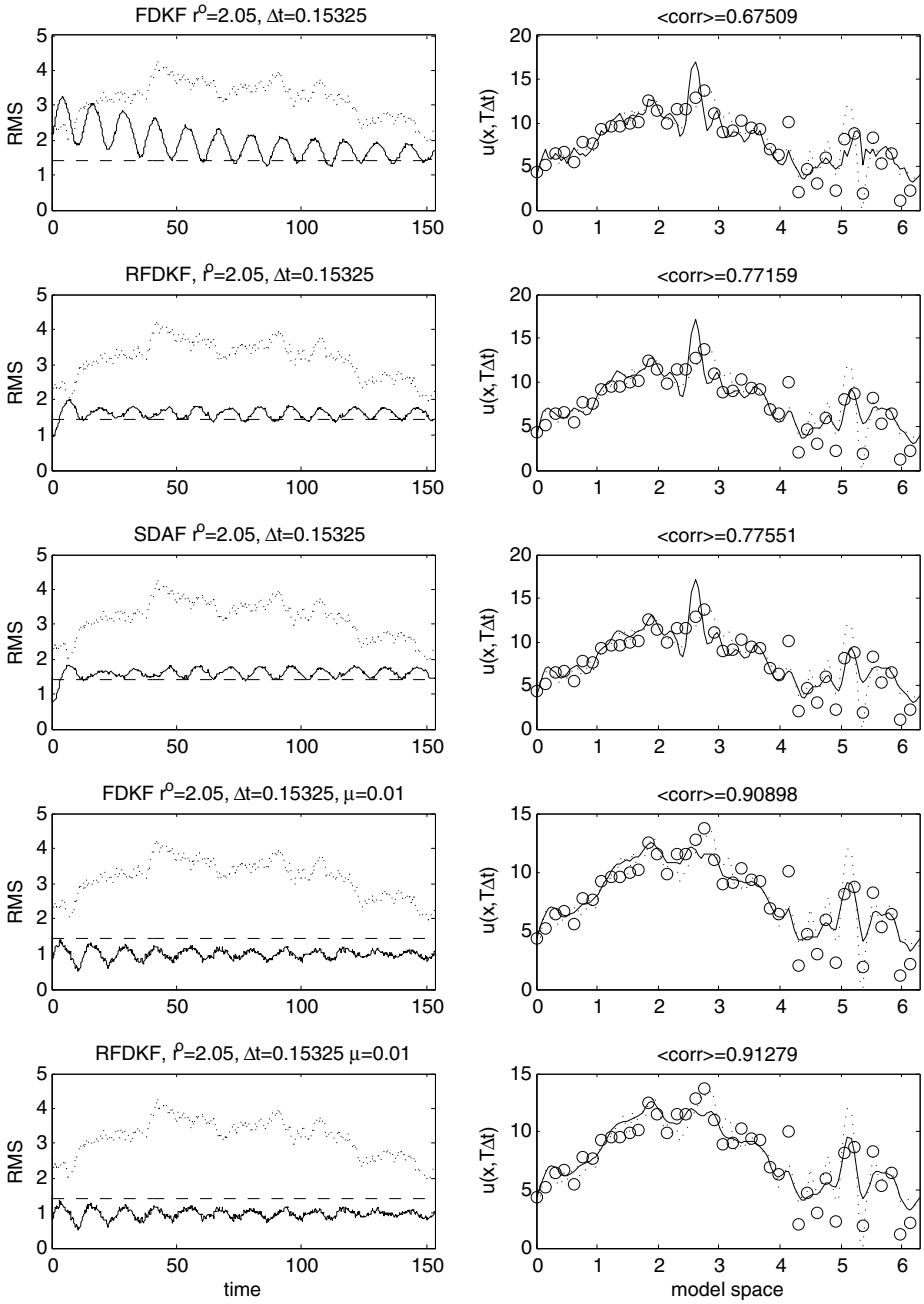

Fig. 16. Weakly damped advection equation for $\Delta t = 2\pi/(2M+1) = 0.1532$, $E_k = k^{-5/3}$, and non-resonant periodic forcing. The first column shows RMS errors as functions of time for unfiltered solution (dotted), the filtered solutions (solid), and observation noise size (dashes). The second column shows the snapshots of the filtered solution (corresponds to the filtering strategy stated in the subtitle of each panel in the first column for the same row) as functions of model spatial domain (solid), the true state (dotted), and the sparse observations (circle), all at time 153.2.

From the numerical results, we conclude that there are three new features in this filtering problem. The first feature is that when the signal to noise ratio is small, regardless to whether the filter model satisfies the observability condition or not, the modified filter with additional damping as suggested by (30) has superior filter skill; especially when the mean signal is driven by a non-resonant periodic forcing. In contrast, when

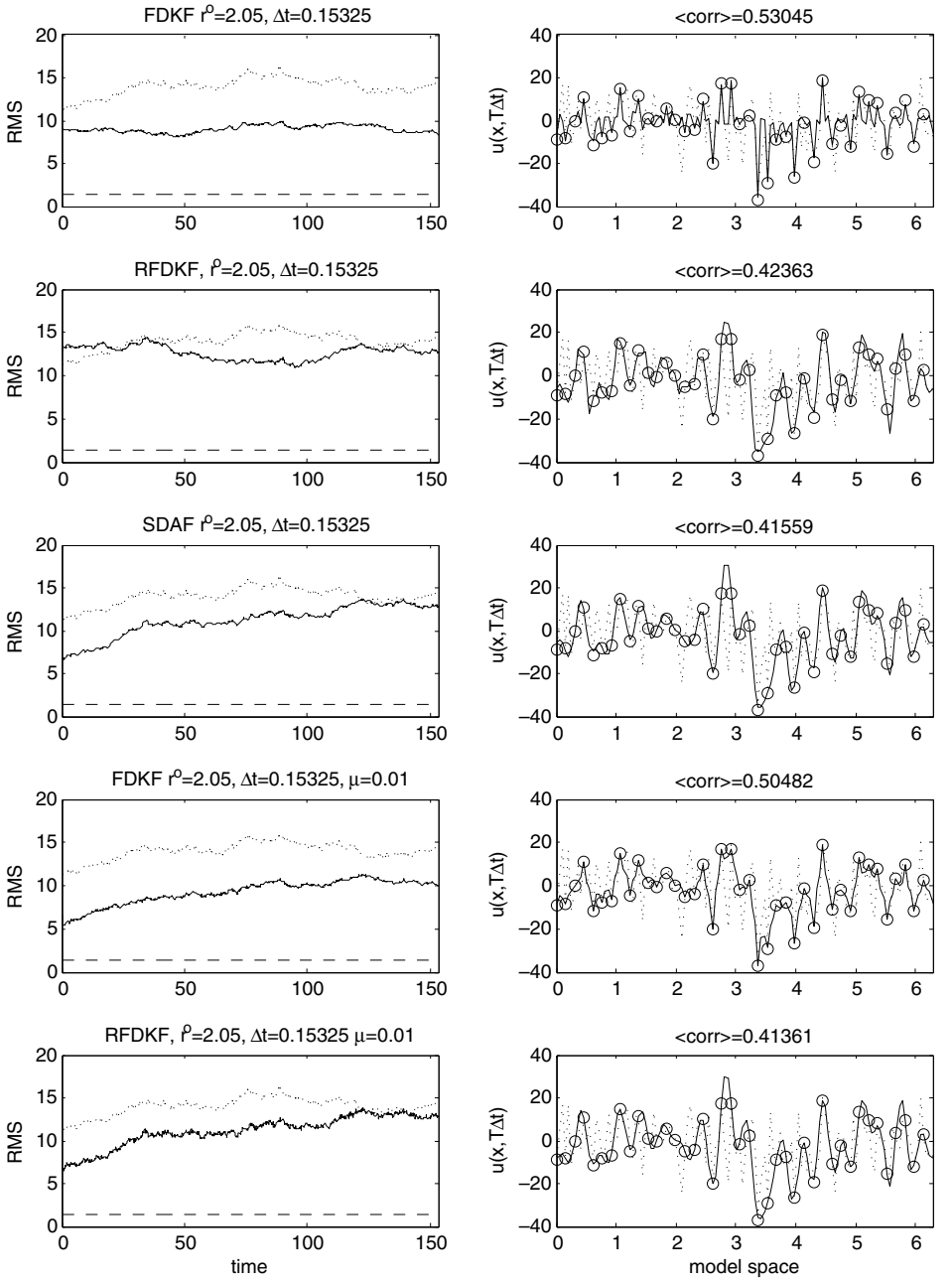

Fig. 17. Weakly damped advection equation for $\Delta t = 2\pi/(2M+1) = 0.1532$, $E_k = 1$, and decaying mean. The first column shows RMS errors as functions of time for unfiltered solution (dotted), the filtered solutions (solid), and observation noise size (dashes). The second column shows the snapshots of the filtered solution (corresponds to the filtering strategy stated in the subtitle of each panel in the first column for the same row) as functions of model spatial domain (solid), the true state (dotted), and the sparse observations (circle), all at time 153.2.

the signal to noise ratio is large, the modified filter is a bad choice since its additional damping introduces significant model errors. The second feature occurs in a specific regime where the energy spectrum is smooth and the signal to noise ratio is large for both observable and non-observable times. In this exceptional regime, the filter is not practically controllable so that the signals are not synchronized. However, we can improve the

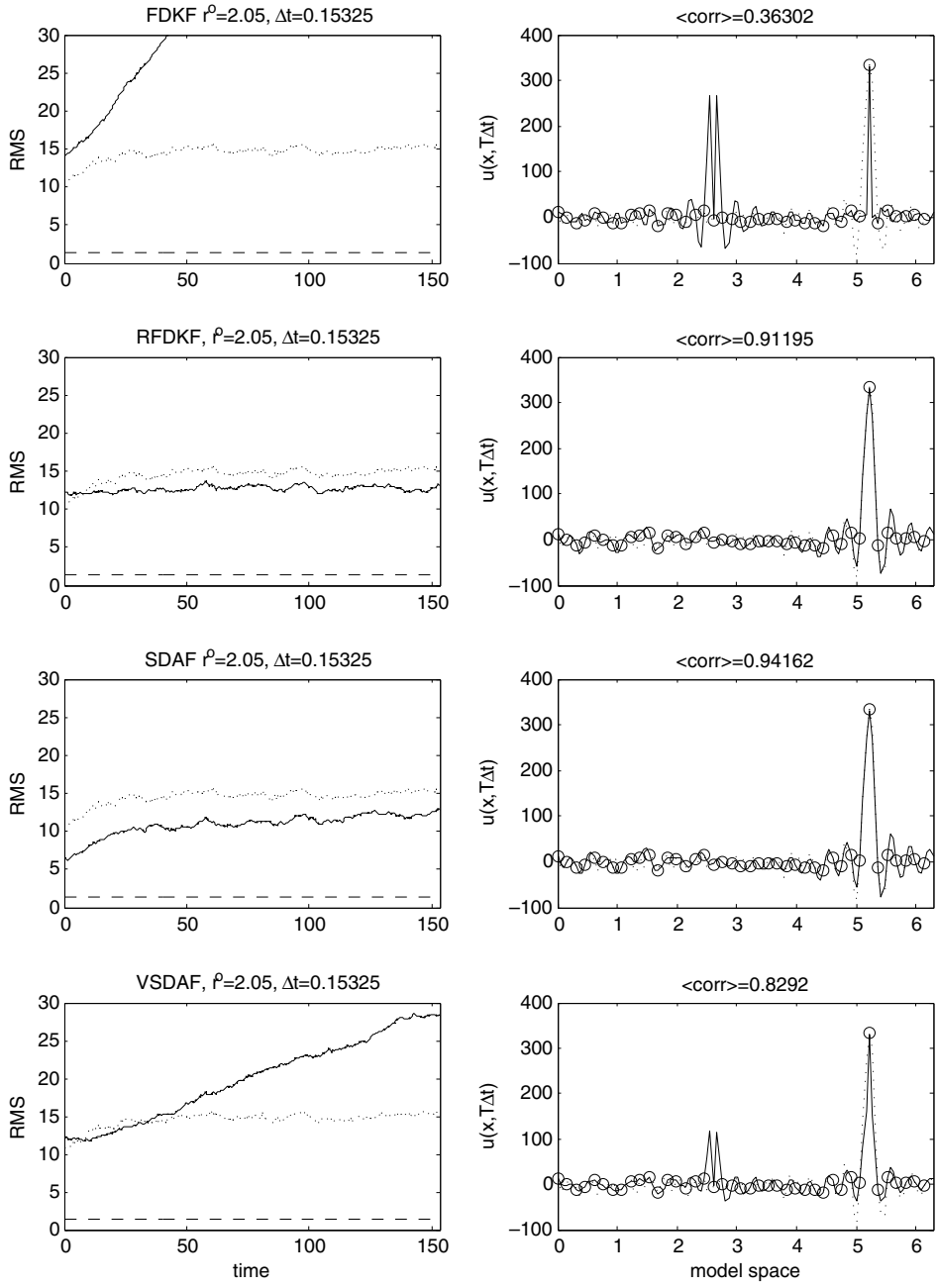

Fig. 18. Weakly damped advection equation for $\Delta t = 2\pi/(2M+1) = 0.1532$, $E_k = 1$, and resonant periodic forcing. The first column shows RMS errors as functions of time for unfiltered solution (dotted), the filtered solutions (solid), and observation noise size (dashes). The second column shows the snapshots of the filtered solution (corresponds to the filtering strategy stated in the subtitle of each panel in the first column for the same row) as functions of model spatial domain (solid), the true state (dotted), and the sparse observations (circle), all at time 153.2.

filtering skill significantly by increasing the system noise variance as discussed earlier in Section 4 and in [6]. The third feature occurs when observability is not satisfied and the energy spectrum is rough, $E_k = 1$, then SDAF and RFDKF are the methods of choice. Finally for the unobservable time, FDKF showed no evidence of filter divergence with a decaying mean state, strong filter divergence with a resonant periodic mean forcing,

and deteriorated filter skill below both SDAF and RFDKF without divergence for non-resonant periodic mean forcing.

7. Concluding discussion

In the above sections, we considered the filtering of noisy turbulent signals generated by a linear stochastic PDE which is observed through sparse regularly spaced observations. Consistent with the objectives in I, II, III, from Section 1, we have checked the filter performance of ETKF in Sections 5 and 6, developed theoretical guidelines for filter performance on these stiff multi-scale problems utilizing a suite of simpler canonical test models in Section 4 following the strategy in [25,6], and developed more stable and computationally efficient alternative filters in the Fourier domain in Secion 3, FDKF, SDAF, VSDAF, and RFDKF. As computational algorithms SDAF, VSDAF, and RFDKF operate only on the primary Fourier modes for the sparse observations, 41 modes rather than 123 modes in Sections 5 and 6. A test bed of challenging problems for filtering noisy turbulent signals for the stochastically forced dissipative advection equation with decaying mean state as well as resonant and non-resonant periodic mean forcing have been developed in Sections 2, 5, and 6; in this test bed, there are regimes of turbulent solutions with the rougher equipartition spectrum, $E_k = 1$, as well as the smoother $k^{-5/3}$ spectrum and turbulent signals to be filtered through sparse regular observations with low and moderate signal to noise ratio (decaying mean and non-resonant forcing) as well as high signal to noise ratio (resonant forcing). Such a test bed should be useful in future studies for filtering turbulent signals. Section 5 involves the test case of the stochastically forced advection–diffusion equation which has selective damping while Section 6 involves the more difficult test bed of the advection equation with small uniform damping. Performance of the various filtering strategies was assessed both through RMS errors and the average spatial pattern correlation [24,27] with the truth signal; this latter measure of filter skill is especially important for filtering sparse observations since the RMS error at all mesh points can often exceed the observation noise error but the filter nevertheless has high skill reflected by a high pattern correlation; examples of this behavior occur throughout Sections 5 and 6.

There are detailed summaries at the end of Sections 4–6 for filter performance skill on the suite of test problems as well as detailed discussion and comparison of the approximate filter strategies at the end of Section 3. Here, we simply note some of the salient points. First, the filtering performance on the PDE's in Sections 5 and 6 strongly reflects the guidelines from the filtering study on the simpler canonical ODE test problem in Section 4 in suitable stiff regimes. Secondly, the physical domain ETKF has poor filter skill on the entire test bed in both Sections 5 and 6 with large RMS error (Tables 1 and 2), low correlation skill, and surprisingly, strong filter divergence with a decaying spectrum and selective damping for the turbulent solutions of the advection–diffusion equation in Section 5 with resonant mean periodic forcing. All of the Fourier domain filters studied here have significant skill beyond ETKF in appropriate regimes. The skill of these approximate filters depends on the observation time compared to the decorrelation time of the turbulent signal on the unresolved modes (see Fig. 2), the energy spectrum, and also in Section 6 whether the observation time satisfies the classical observability criterion (Sections 2.4 and 2.5). For Section 5 and observable times in Section 6, if the observation time is below the correlation time of all unresolved modes, FDKF and the cheaper VSDAF have very high filtering skill for any spectrum; as the observation time increases SDAF also has comparable skill for all spectra but RFDKF has poor skill for the equipartition spectrum and comparable skill for the decaying $k^{-5/3}$ spectrum. The effect on filter performance at an unobservable time for FDKF is interesting: as reported in Section 6, there is high filter skill for the decaying mean state, strong filter divergence for resonant forcing, and deteriorated filter skill compared to SDAF and RFDKF for non-resonant periodic forcing. The approximate filters SDAF and RFDKF were designed to be always observable in Section 3; on the other hand, VSDAF also suffers from weaker filter divergence with resonant mean forcing in Section 6 and a theoretical explanation is given in Section 4.2.

The alternative regularization proposed in (30) by adding artificial viscosity to the uniformly damped advection equation to generate a modified model for filtering was also tested in Section 6. For turbulent signals with low or moderate signal to noise ratio, such as the decaying mean state or non-resonant forcing, these modified filters had the highest skill; however, for a large signal to noise ratio with resonant mean forcing the performance of SDAF was superior to the modified filters due to their large model error; the modified

methods introduce too much viscosity in the attempt to filter a localized large amplitude signal. On the other hand, no attempt was made in Section 6 to optimize the artificial viscosity to minimize model error and this is a topic worthwhile to pursue in the future.

Another practical source of filtering failure (which occurs even with RFDKF) is due to a lack of controllability when the spectrum decays smoothly. However, in Sections 4 and 6, we show that a significant filter improvement is achieved when controllability is improved with larger system noise variance; this strategy mimics that of [6]. The approximate filters also deal with the "curse of ensemble size" in this idealized setting; as shown in Section 5, only 10 ensemble members are needed to achieve the idealized filter performance.

Finally, with the encouraging results developed here and in [17], in future work we plan to test the Fourier domain, SDAF, VSDAF, and RFDKF algorithms developed here, in the context of the nonlinear chaotic L-96 model in the severe environment of sparse regular observations where ETKF often suffers from filter divergence [29]. Other future directions are to implement the Fourier domain filtering strategies in a different basis such as the spherical harmonics in a global weather model and to develop filtering strategies for randomly located sparse observations since typically the measurements, e.g., from buoys, satellite, or weather balloons, are not taken at the model grid points.

Acknowledgements

The research of A.J. Majda is partially supported by the National Science Foundation Grant DMS-0456713, the Office of Naval Research Grant N00014-05-1-0164, and the Defense Advanced Research Projects Agency Grant N00014-07-1-0750. John Harlim is supported as a postdoctoral fellow through the last two agencies.

References

[1] B. Anderson, J. Moore, Optimal Filtering, Prentice-Hall, Englewood Cliffs, NJ, 1979.
[2] J. Anderson, An ensemble adjustment Kalman filter for data assimilation, Monthly Weather Review 129 (2001) 2884–2903.
[3] J. Anderson, A local least squares framework for ensemble filtering, Monthly Weather Review 131 (4) (2003) 634–642.
[4] J. Anderson, S. Anderson, A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts, Monthly Weather Review 127 (1999) 2741–2758.
[5] C. Bishop, B. Etherton, S. Majumdar, Adaptive sampling with the ensemble transform Kalman filter part I: the theoretical aspects, Monthly Weather Review 129 (2001) 420–436.
[6] E. Castronovo, J. Harlim, A. Majda, Mathematical test criteria for filtering complex systems: plentiful observations, Journal of Computational Physics 227 (7) (2008) 3678–3714.
[7] A. Chorin, P. Krause, Dimensional reduction for a bayesian filter, Proceedings of the National Academy of Sciences 101 (42) (2004) 15013–15017.
[8] C. Chui, G. Chen, Kalman Filtering, Springer, New York, 1999.
[9] S. Cohn, D. Dee, Observability of discretized partial differential equations, SIAM Journal on Numerical Analysis 25 (3) (1988) 586–617.
[10] B. Farrell, P. Ioannou, State estimation using a reduced-order Kalman filter, Journal of the Atmospheric Sciences 58 (23) (2001) 3666–3680.
[11] B. Farrell, P. Ioannou, Distributed forcing of forecast and assimilation error systems, Journal of the Atmospheric Sciences 62 (2) (2005) 460–475.
[12] C. Gardiner, Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences, Springer-Verlag, New York, 1997.
[13] M. Ghil, P. Malanotte-Rizolli, Data assimilation in meteorology and oceanography, Advances in Geophysics 33 (1991) 141–266.
[14] M. Grote, A. Majda, Stable time filtering of strongly unstable spatially extended systems, Proceedings of the National Academy of Sciences 103 (2006) 7548–7553.
[15] J. Harlim, B. Hunt, A non-Gaussian ensemble filter for assimilating infrequent noisy observations, Tellus A 59 (2) (2007) 225–237.
[16] J. Harlim, B. Hunt, Four-dimensional local ensemble transform Kalman filter: numerical experiments with a global circulation model, Tellus A 59 (5) (2007) 731–748.
[17] J. Harlim, A. Majda, Filtering nonlinear dynamical systems with linear stochastic models, Nonlinearity, submitted for publication.
[18] K. Haven, A. Majda, R. Abramov, Quantifying predictability through information theory: small sample estimation in a non-gaussian framework, Journal of Computational Physics 206 (2005) 334–362.
[19] J. Hesthaven, S. Gottlieb, D. Gottlieb, Spectral Methods for Time-dependent Problems, Cambridge Monographs on Applied and Computational Mathematics, vol. 21, Cambridge University Press, Cambridge, UK, 2007.
[20] P. Houtekamer, H. Mitchell, Data assimilation using an ensemble Kalman filter technique, Monthly Weather Review 126 (1998) 796–811.

[21] B. Hunt, E. Kostelich, I. Szunyogh, Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter, Physica D 230 (2007) 112–126.
[22] J. Kaipio, E. Somersalo, Statistical and Computational Inverse Problems, Springer, New York, 2005.
[23] E. Lorenz, Predictability – a problem partly solved, in: Proceedings on predictability, held at ECMWF on 4–8 September 1995, 1996.
[24] A. Majda, R. Abramov, M. Grote, Information Theory and Stochastics for Multiscale Nonlinear Systems, CRM Monograph Series, vol. 25, American Mathematical Society, Providence, RI, USA, 2005.
[25] A. Majda, M. Grote, Explicit off-line criteria for stable accurate time filtering of strongly unstable spatially extended systems, Proceedings of the National Academy of Sciences 104 (2007) 1124–1129.
[26] A. Majda, J. McDonough, S. Osher, The Fourier method for nonsmooth initial data, Mathematics of Computation 32 (144) (1978) 1041–1081.
[27] A. Majda, X. Wang, Nonlinear Dynamics and Statistical Theories for Basic Geophysical Flows, Cambridge University Press, UK, 2006.
[28] R. Miller, E. Carter, S. Blue, Data assimilation into nonlinear stochastic models, Tellus A 51 (2) (1999) 167–194.
[29] E. Ott, B. Hunt, I. Szunyogh, A. Zimin, E. Kostelich, M. Corrazza, E. Kalnay, J. Yorke, A local ensemble Kalman filter for atmospheric data assimilation, Tellus A 56 (2004) 415–428.
[30] R. Richtmeyer, K. Morton, Difference Methods for Initial Value Problems, Wiley, 1967.
[31] R. Todling, M. Ghil, Tracking atmospheric instabilities with the kalman filter. Part I: methodology and one-layer results, Monthly Weather Review 122 (1) (1994) 183–204.
[32] X. Wang, C. Bishop, S. Julier, Which is better, an ensemble of positive-negative pairs or a centered spherical simplex ensemble? Monthly Weather Review 132 (7) (2004) 1590–1605.
[33] J. Whitaker, T. Hamill, Ensemble data assimilation without perturbed observations, Monthly Weather Review 130 (2002) 1913–1924.