



Improving filtering and prediction of spatially extended turbulent systems with model errors through stochastic parameter estimation

B. Gershgorin^a, J. Harlim^{b,*}, A.J. Majda^a

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

^bDepartment of Mathematics, North Carolina State University, NC 27695, United States

ARTICLE INFO

Article history:

Received 13 April 2009

Received in revised form 9 September 2009

Accepted 18 September 2009

Available online 29 September 2009

Keywords:

Stochastic parameter estimation

Kalman filter

Filtering turbulence

Data assimilation

Model error

Predictability

ABSTRACT

The filtering and predictive skill for turbulent signals is often limited by the lack of information about the true dynamics of the system and by our inability to resolve the assumed dynamics with sufficiently high resolution using the current computing power. The standard approach is to use a simple yet rich family of constant parameters to account for model errors through parameterization. This approach can have significant skill by fitting the parameters to some statistical feature of the true signal; however in the context of real-time prediction, such a strategy performs poorly when intermittent transitions to instability occur. Alternatively, we need a set of dynamic parameters. One strategy for estimating parameters on the fly is a stochastic parameter estimation through partial observations of the true signal. In this paper, we extend our newly developed stochastic parameter estimation strategy, the Stochastic Parameterization Extended Kalman Filter (SPEKF), to filtering sparsely observed spatially extended turbulent systems which exhibit abrupt stability transition from time to time despite a stable average behavior. For our primary numerical example, we consider a turbulent system of externally forced barotropic Rossby waves with instability introduced through intermittent negative damping. We find high filtering skill of SPEKF applied to this toy model even in the case of very sparse observations (with only 15 out of the 105 grid points observed) and with unspecified external forcing and damping. Additive and multiplicative bias corrections are used to learn the unknown features of the true dynamics from observations. We also present a comprehensive study of predictive skill in the one-mode context including the robustness toward variation of stochastic parameters, imperfect initial conditions and finite ensemble effect. Furthermore, the proposed stochastic parameter estimation scheme applied to the same spatially extended Rossby wave system demonstrates high predictive skill, comparable with the skill of the perfect model for a duration of many eddy turnover times especially in the unstable regime.

© 2009 Elsevier Inc. All rights reserved.

1. Introduction

Filtering is the process of obtaining the best statistical estimate of a physical system from partial observations of the true signal. In many contemporary applications in science and engineering, real-time filtering of a turbulent signal involving many degrees of freedom is needed to make accurate predictions of the future state. Important practical examples involve

DOI of original article: [10.1016/j.jcp.2009.08.019](https://doi.org/10.1016/j.jcp.2009.08.019)

* Corresponding author.

E-mail address: jharlim@ncsu.edu (J. Harlim).

the real-time filtering and prediction of weather and climate as well as the spread of hazardous plumes or pollutants. A major difficulty in accurate filtering and prediction of noisy turbulent signals with many degrees of freedom is model error [35]: the fact that the true signal is processed through an imperfect model where important physical processes are parameterized due to inadequate numerical resolution or incomplete physical understanding. Under these circumstances, it is natural to devise strategies for parameter estimation to cope with model errors to improve both filtering and prediction skill [32,2–4,9,8,6,17].

Recently, we proposed the use of the test models with stochastic parameter estimation [14], where certain parameters of the system such as damping and external forcing can be learned from the observations. This approach, called the Stochastic Parameterization Extended Kalman Filter (SPEKF) in [14], has been shown to be effective when applied to a single mode [14]. In this paper, we extend this stochastic parameterization strategy to filtering sparsely observed spatially extended systems with a decaying (Kolmogorov type) turbulent spectrum [19] and intrinsic model errors as in [14]. According to results in [19] for the perfectly specified filter case, we can obtain accurate filtered solutions by filtering only the observed modes with a reduced filter when the energy spectrum decays as a function of wavenumber. Our new approach here is to unify ideas from these two papers [19,14], that is, to apply SPEKF only to the observed modes and let the estimates of the remaining unobserved and least energetic modes to be unfiltered and propagated in time with the climatological model.

In this paper, we will test our newly developed strategy on a nontrivial toy model for turbulent barotropic Rossby waves in a one-dimensional periodic domain with a time periodic external forcing. We design this model such that it exhibits instability that mimics the baroclinic instabilities in the midlatitude atmosphere at random times despite a stable long time average behavior. In this example, the model errors are introduced through our lack of information about the onset time and duration of instability regimes. Moreover, we also introduce a second source of model error by purposely not specifying the external forcing. In our numerical experiments with this example, we resolve this model with 105 equally spaced grid points in a periodic domain and consider rather sparse observations at 15 equally spaced grid points. Through extensive numerical studies, we will find that our new stochastic parametrization strategy, especially the one that combines both multiplicative and additive bias corrections, produces a significantly improved and robust filtering skill as was shown earlier in the one-mode context in [14]. We will also find significant improvement in predictive skill with the combined model relative to the mean model.

The paper is organized as follows. In Section 2, we describe idealized spatially extended turbulent systems: this includes the simplest stochastic models for turbulent signals, the turbulent Rossby waves problem as an example, and the two-state Markov process as an underlying mechanism for stability regime transitions. In Section 3, we discuss various strategies for filtering with model errors, including the standard approaches such as the mean stochastic model (MSM) and the standard online bias correction strategy using extended Kalman filter [22,1,7] which motivates the Stochastic Parameterization Extended Kalman Filter (SPEKF) in our recent work [14]. In this section, we also review results of SPEKF in the one-mode context [14] and then specify a strategy for implementing SPEKF for spatially extended systems with sparse observations. In Section 4, we present the numerical results for filtering, including the correctly specified and unspecified forcing and the filter skill robustness throughout variations of parameters. In Section 5, we extensively study the predictive skill on one Fourier mode in a “super-ensemble” setup. In particular, we try to understand the effect of errors from initial conditions and a finite ensemble size. Consequently, we show results on the full SPDE. We close both Sections 4 and 5 with short summaries. We end the paper with concluding discussions in Section 6. Detailed calculations of correlation function and Kalman filter formulas are reported in [Appendixes A and B](#).

2. Idealized spatially extended turbulent systems

The simplest models for representing turbulent fluctuations involve replacing nonlinear interaction by additional linear damping and stochastic white noise forcing in time which incorporate the observed climatological spectrum and turnover time for the turbulent field [10,28,30].

As in the standard classical numerical analysis test criteria for finite difference schemes [33,29], we start with a linearized complex $s \times s$ PDE at a constant coefficient background, $\vec{u}_t = \mathcal{P}(\partial_x)\vec{u} + \vec{f}(x, t)$. Here $\vec{f}(x, t)$ is a known deterministic forcing term. In accordance with the above approximations, additional damping $-\gamma(\partial_x)\vec{u}$ and spatially correlated noise

$$\sigma(x)\dot{W}(t) \equiv \sum_{k=-\infty}^{\infty} \sigma_k \dot{W}_k(t) e^{ikx}, \quad (1)$$

where $W_k(t)$ are independent complex Wiener processes for each $k \geq 0$ and the independent real and imaginary parts have the same variance $1/2$ and $\sigma_{-k} = \sigma_k$ and $W_{-k}(t) = W_k^*(t)$, are added to the PDE to represent the small scale unresolved turbulent motions resulting in the basic frozen coefficient canonical PDE. For simplicity in notation here we discuss a scalar field in a single space real variable but everything generalizes to a matrix system of stochastic PDEs in several space dimensions.

2.1. The simplest stochastic models for turbulent signals

With the above motivation, we consider solutions of the real valued scalar stochastic partial differential equation (SPDE):

$$\begin{aligned}\frac{\partial u(x, t)}{\partial t} &= \mathcal{P}\left(\frac{\partial}{\partial x}\right)u(x, t) - \gamma\left(\frac{\partial}{\partial x}\right)u(x, t) + f(x, t) + \sigma(x)\dot{W}(t), \\ u(x, 0) &= u_0(x),\end{aligned}\quad (2)$$

as the simplest model for turbulent signals. Here the initial data u_0 is a Gaussian random field with nonzero covariance. As in the usual finite difference linear stability analysis, we non-dimensionalize this problem in a 2π -periodic domain so that a solution of (2) is given by the infinite Fourier series

$$u(x, t) = \sum_{k=-\infty}^{\infty} \hat{u}_k(t)e^{ikx}, \quad \hat{u}_{-k}(t) = \hat{u}_k^*(t), \quad (3)$$

where $\hat{u}_k(t)$ for $k > 0$ can be utilized in analyzing (2). For finitely discretized system, we use the finite sum of (3).

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 wavenumber k by

$$\mathcal{P}\left(\frac{\partial}{\partial x}\right)e^{ikx} = \tilde{p}(ik)e^{ikx}, \quad (4)$$

$$\gamma\left(\frac{\partial}{\partial x}\right)e^{ikx} = \gamma(ik)e^{ikx}. \quad (5)$$

Substituting (1), (3)–(5) into initial value problem in (2), we obtain a system of uncoupled forced Langevin equations on each Fourier mode

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

The term $\hat{f}_k(t)$ is the Fourier coefficient of the deterministic forcing $f(x, t)$.

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

$$\tilde{p}(ik) = i\omega_k, \quad (7)$$

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

$$\gamma(ik) > 0 \quad \text{for all } k \neq 0. \quad (8)$$

Under these assumptions, Gaussian equilibrium distribution for (6) exists and provided $f(x, t) = 0$ this statistical equilibrium distribution has mean zero and variance E_k , defining the climatological energy spectrum

$$E_k = \frac{\sigma_k^2}{2\gamma(ik)}, \quad 1 \leq k < +\infty. \quad (9)$$

Mathematically, one requires $\sum E_k < \infty$ to define the stochastic solution of (2) correctly with a similar requirement on the Gaussian initial data in $u_0(x)$. However, there is genuine physical interest in situations with an even rougher turbulent spectrum such as white noise where E_k is constant. In these cases, we truncate the sum in (3) to a large finite sum.

Another quantity that is typically measured in a turbulent system is the eddy turnover time, which is the time it takes for an eddy to decorrelate in the equilibrium statistical state. Mathematically, the eddy turnover time at a given wavenumber is defined as the integral on the positive half line (from 0 to ∞) of the absolute value of the correlation function

$$R_k(\tau) \equiv \langle (\hat{u}_k(t) - \tilde{u}_k)(\hat{u}_k(t + \tau) - \tilde{u}_k)^* \rangle = E_k e^{-\gamma(ik)\tau} e^{-i\omega_k\tau}, \quad (10)$$

which is the absolute damping time, $1/\gamma(ik)$, when normalized by E_k . In [14], we called this quantity the decorrelation time since we only studied one Fourier mode in that paper. In this paper, we refer to this quantity as the damping time since we do not want to confuse the reader with the physical space decorrelation time defined below. The detailed calculation of the correlation function, $R_k(\tau)$, is based on the exact solution of SDE in (6) and is given in Appendix A.

2.2. Example: a simple model for turbulent Rossby waves

In this paper, we consider barotropic Rossby waves with phase varying only in the north–south direction in a one-dimensional periodic domain, with dispersion relation [27,31] given by

$$\omega_k = \frac{\beta}{k}. \quad (11)$$

On the planetary scale, the midlatitude beta plane approximates the effect of rotation with

$$\beta = \frac{f}{a} = \frac{2\Omega \cos(\theta)}{a}, \quad (12)$$

where $a = 6.37 \times 10^6$ m is the mean radius of the earth and $\Omega = 7.292 \times 10^{-5}$ rad/s is the earth angular speed of rotation [21]. In our model, we consider a periodic domain of length 2π so that the radius has a unit length $a = 1$. Converting the time

into days, we find that the parameter β at latitude $\theta = 45^\circ$ is $\beta = 2\Omega \cos(45^\circ) \text{ s}^{-1} = 8.91 \text{ day}^{-1}$. Thus the natural frequency for this model is given by $\omega_k = 8.91/k$ so the lowest wavenumber, $k = 1$, has an oscillation period of roughly $2\pi/\omega_1 = 17 \text{ h}$.

It is also natural to assume uniform damping

$$\gamma(ik) = \bar{d} > 0, \tag{13}$$

representing the bottom drag (Ekman) friction. It is known from observations that on scales of order of thousands of kilometers these waves have a k^{-3} energy spectrum [26] so that

$$E_k = k^{-3}. \tag{14}$$

Tuning the model to have 3 day prediction like the weather [25], we choose a damping strength $\bar{d} = 1.5$ such that the physical space correlation function

$$\mathcal{R}(\tau) = \frac{\sum_{k=-\infty}^{\infty} \mathbf{R}_k(\tau)}{\sum_{k=-\infty}^{\infty} E_k} = e^{-\bar{d}\tau} \frac{\sum_{k=-\infty}^{\infty} E_k e^{-i\omega_k \tau}}{\sum_{k=-\infty}^{\infty} E_k} \tag{15}$$

decays after 3 days. Fig. 1 shows the physical space correlation function $\mathcal{R}(\tau)$ as a function of the lag τ for a finite sum of $N = 52$ discrete Fourier modes. We also use the equilibrium energy spectrum in (9) to calibrate the system noise strength $\sigma_k = \sqrt{2\bar{d}/k^3}$ at each wavenumber.

We also consider the following external periodic forcing

$$\hat{f}_k(t) = \begin{cases} 0.5 e^{0.15it}, & \text{if } 0 < k \leq 5, \\ 0.5 \left(\frac{6}{k}\right)^3 e^{0.15it}, & \text{if } k > 5, \\ \hat{f}_{-k}^*, & \text{if } k < 0, \end{cases} \tag{16}$$

which mostly acts on the lower modes of the system. The frequency is chosen to represent a long wave background forcing with period of $2\pi/0.15 = 41.88$ days, which is much slower than the 3 day decorrelation time. In the numerical simulations below, we will also consider random forcing.

2.3. Instability transitions with a two-state Markov process

Now, we describe the procedure for generating a signal that exhibits transitions between stable and unstable regimes. In the unstable regime, we mimic the baroclinic instability that governs weather wave patterns in the midlatitude atmosphere.

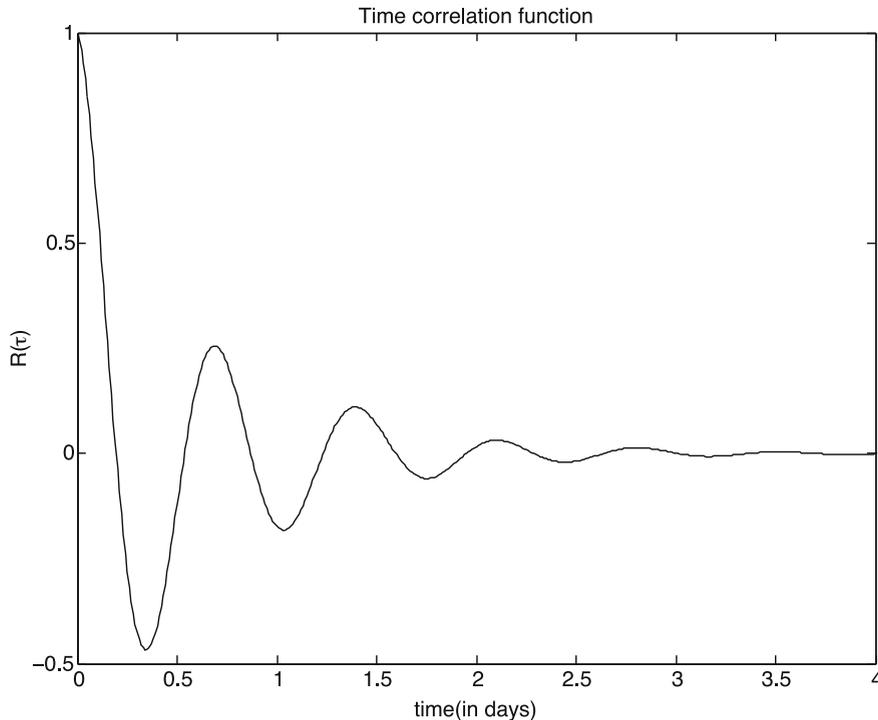


Fig. 1. Physical space temporal correlation function of $u(x, t)$, computed using $\mathcal{R}(\tau)$ in (15), decorrelates after 3 days.

Numerically, we allow wavenumbers 3–5 (typical modes where baroclinic instability occurs [31]) to switch between stable and unstable regimes. Mathematically, the stable (unstable) regime is characterized by positive (negative) damping and corresponds to the loss (gain) of barotropic energy. Simultaneously, the first two modes are weakly (strongly) damped when the next three modes are in the stable (unstable) regime. This mechanism models energy flux from the intermediate waves to the largest waves in the stable regime and vice versa in the unstable regime.

As in [14], we consider an exponentially distributed waiting time between these two regimes. That is, the waiting time the system spends in the stable regime before it switches to the unstable regime is a random variable T_{st} with cumulative distribution function given by

$$P(T_{st} < \tau) = 1 - e^{-\nu\tau}, \quad (17)$$

where ν is a rate of change from stable to unstable regime. Similarly, for the random time T_{un} that the system spends in the unstable regime, we have

$$P(T_{un} < \tau) = 1 - e^{-\mu\tau}, \quad (18)$$

where μ is a rate of change from unstable to stable regime. In our numerical model, we set $\nu = 0.1$ and $\mu = 0.2$ such that the system spends on average 10 days in the stable regime and 5 days in the unstable regime. We choose the damping strength of each mode as in Table 1 such that the average damping is constant,

$$\bar{d} = \frac{\nu d^- + \mu d^+}{\nu + \mu} = \frac{\nu d_w + \mu d_s}{\nu + \mu} = 1.5. \quad (19)$$

In Fig. 2(a), we show a space–time plot of the solutions of the stochastic toy model for the Rossby waves described in Section 2.2 together with the instability transitions (the “switching” SPDE) described above, solved with time step 0.001 and discretized with a total of $2N + 1 = 105$ grid points in a 2π periodic domain. In Fig. 2(b), we also show the stability regime, i.e., the damping strength of modes 3–5, which is typically unknown in reality. In this snapshot, we see a strong coherent westward wind burst which begins 2 days after the unstable transition. The fact that the occurrence of this westward wind burst is not exactly right after regime transition makes this turbulent signal an extremely hard test problem. In the remainder of this paper, we refer to this turbulent signal as the true signal that we want to filter and predict.

3. Filtering with model errors through stochastic parameter estimation

In the perfect model scenario, we can filter and predict the true signal (see Fig. 2) with the “perfectly specified filter”. In this paper, we consider the Fourier domain Kalman filter (the reader unfamiliar with the Kalman filter can consult Appendix B below), described in [5,19], with prior statistics generated through solving the statistics of SDE in (6) exactly, assuming that the exact time series of the damping coefficients of the switching modes are known. In Section 4, we will present solutions with this perfectly specified filter as a benchmark.

In reality, however, we do not know in what regime the signal is and when a transition is going to occur. Moreover, we do not even know which modes exhibit instability. This incomplete knowledge introduces what is known as “model error”. Another source of model error we will also study comes through unspecified external forcing $f(x, t)$.

3.1. Standard approaches

In reality, what is available is the long time average statistics of the true signal based on various measurements from the past events, such as the average damping time $1/\bar{d}$ and the equilibrium variance $E_k = \sigma_k^2/2\bar{d}$ (in turbulence theory, this quantity is also called the energy spectrum). In this section, we describe briefly the two standard approaches for filtering:

3.1.1. Mean stochastic model (MSM)

The simplest commonly used filter model [18,20] is the mean stochastic model (MSM), which is a model that is based on the two equilibrium statistical quantities, energy spectrum and damping time; MSM is exactly the climatological stochastic model (CSM) in [18,20]. The mean stochastic model (MSM) solves

$$\frac{d\hat{u}_k(t)}{dt} = (-\bar{d} + i\omega_k)\hat{u}_k(t) + \sigma_k \dot{W}_k(t) + \tilde{f}_k(t), \quad (20)$$

Table 1
Damping strength.

Modes	Stable	Unstable
1, 2	$d_w = 1.3$	$d_s = 1.6$
3, 4, 5	$d^+ = 2.25$	$d^- = -0.04$
≥ 6	$\bar{d} = 1.5$	$\bar{d} = 1.5$

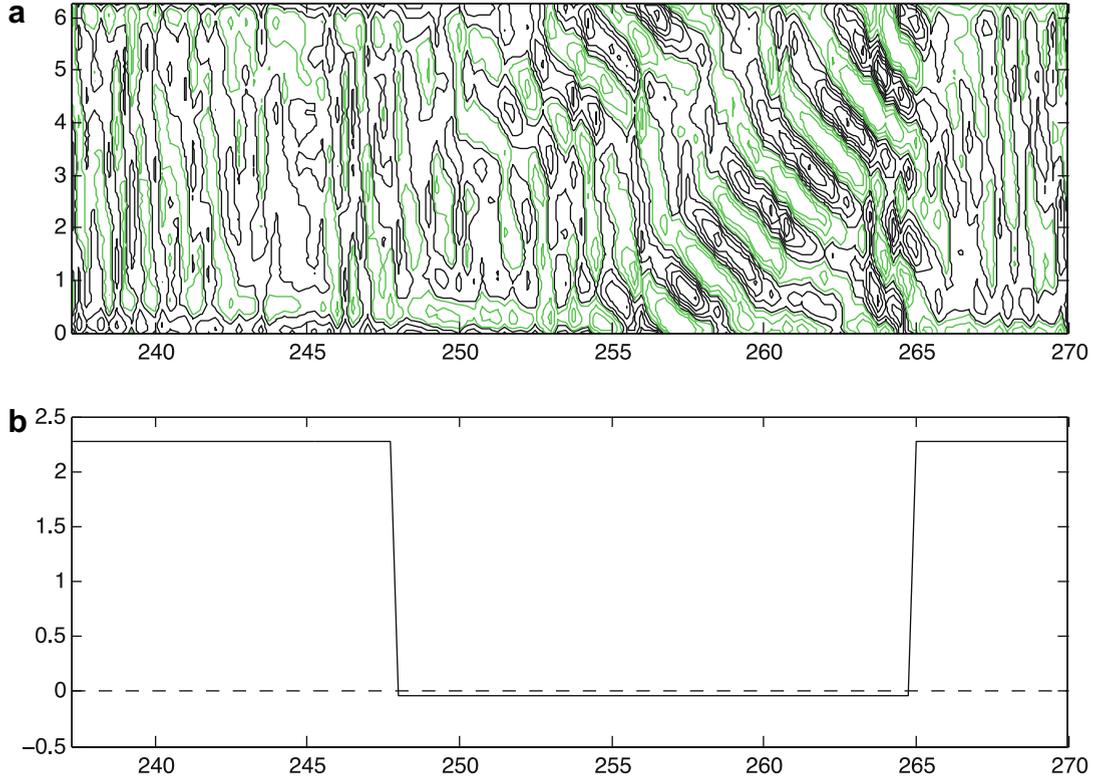


Fig. 2. The true signal $u(x, t)$ for period of time between $t = 240$ and 270 . (a) Space–time plot with contour interval 2 unit/day, dark contour for positive value and grey contour for negative value. (b) Damping $\gamma(t)$ for the solution $u(x, t)$ shown in panel (a).

for prediction and solves its first and second ordered statistics for filtering. Here $\bar{d} = 1.5$ is an average damping constant for all modes; model errors are introduced through unknown time dependent true positive and negative damping $\gamma(ik)$ in the switching modes (as discussed in Section 2.3) and $\hat{f}_k(t)$ can be even an incorrectly specified forcing.

3.1.2. Online bias estimation

A more sophisticated approach to filtering is through updating the unknown damping and forcing coefficients on the fly [32,2–4,9,8,6,17,11,12]. In general, given any nonlinear dynamical system depending on unknown parameters, λ ,

$$\frac{du}{dt} = F(u, t, \lambda), \tag{21}$$

one augments the state variable u by parameters λ and adjoins an approximate dynamical equation for the parameters

$$\frac{d\lambda}{dt} = g(\lambda). \tag{22}$$

The right hand side of (22) is often chosen on an ad-hoc basis as $g(\lambda) \equiv 0$ or white noise forcing with a small variance [11,12].

The partial observations of the true signal are often processed by an Extended Kalman Filter (EKF, see [22,1,7]) applied to the augmented system in (21) and (22) where the parameters λ are estimated adaptively from these partial observations. Note that even if the original model in (21) is linear, it readily can have nonlinear dependence on the parameters λ through (22) so typically an EKF involving the linear tangent approximation is needed for parameter estimation in this standard case. Some recent applications of these and similar ideas to complex nonlinear dynamical systems can be found in [32,2,3,9,8].

3.2. The Stochastic Parameterization Extended Kalman Filters (SPEKF)

In this section, we review the stochastic parametrization strategy “Stochastic Parameterization Extended Kalman Filter” (SPEKF) recently introduced by the authors [14]. Although our stochastic parameter estimation strategy is motivated by the augmentation approach discussed earlier in Section 3.1.2, it differs from the other online approaches since it utilizes nonlinear exactly solvable statistics that includes both the additive and multiplicative bias corrections. Therefore, no linear tangent approximation is needed in SPEKF.

The SPEKF appends the stochastic model for evolution of state variable $\hat{u}_k(t)$ with *combined* additive, $b_k(t)$, and multiplicative, $\gamma_k(t)$, bias correction terms:

$$\begin{aligned} d\hat{u}_k(t) &= ((-\gamma_k(t) + i\omega_k)\hat{u}_k(t) + b_k(t) + \tilde{f}_k(t))dt + \sigma_k dW_k(t), \\ db_k(t) &= (-\gamma_{b,k} + i\omega_{b,k})(b_k(t) - \hat{b}_k)dt + \sigma_{b,k} dW_{b,k}(t), \\ d\gamma_k(t) &= -d_{\gamma,k}(\gamma_k(t) - \hat{\gamma}_k)dt + \sigma_{\gamma,k} dW_{\gamma,k}(t), \end{aligned} \quad (23)$$

for improving filtering and prediction with model errors. Here, stochastic parameters $\gamma_{b,k}$ and $d_{\gamma,k}$ represent the damping and parameters $\sigma_{b,k}$ and $\sigma_{\gamma,k}$ represent the strength of the white noise forcing of the additive and multiplicative bias correction terms, respectively. The stationary mean bias correction values of $b_k(t)$ and $\gamma_k(t)$ are given by \hat{b}_k and $\hat{\gamma}_k$, correspondingly; and the frequency of the additive noise is denoted as $\omega_{b,k}$. As in our earlier study in [14], we choose the stationary mean bias correction values to be $\hat{b}_k = 0$ since this is the statistical long time average for periodic mean zero function and the stationary damping to be $\hat{\gamma}_k = \bar{d}$, which is the average damping of the original system. Note that $W_{\gamma,k}(t)$ is a Wiener process while $W_k(t)$ and $W_{b,k}(t)$ are complex valued Wiener processes with their real and imaginary parts being independent Wiener processes.

It is important to realize that the parameters of the state variable $\hat{u}_k(t)$ come from the characteristics of the physical system, which is modeled by the first equation in (23). On the other hand, stochastic parameters of $b_k(t)$ and $\gamma_k(t)$, $\{\gamma_{b,k}, \omega_{b,k}, \sigma_{b,k}, d_{\gamma,k}, \sigma_{\gamma,k}\}$, are introduced in the model and, in principle, cannot be directly obtained from the characteristics of the physical system. At first glance, this approach looks very counter intuitive in the sense that it introduces five stochastic parameters to parameterize b_k and γ_k on each Fourier mode alone. However, our comprehensive numerical study [14] on the fifth mode of (2) showed that there exists a robust parameter set for high filtering skill beyond the skill of MSM and in certain situations as good as the perfectly specified filter. In the numerical simulations in Section 4, we will find that if we use the same five stochastic parameters for each wavenumber, the high filtering skill on the fifth mode of (2) as reported in [14] translates to the full switching SPDE by choosing a parameter set that belongs to the robust set of the fifth mode as in [14] independent of wavenumber. Therefore, we avoid having to choose $5N$ parameters (assuming there are five stochastic parameters on each mode and a total of N wavenumbers) but only 5 for the whole SPDE. For predictive skill, we first perform a comprehensive numerical study for the fifth mode alone in Section 5 to determine whether there is a robust stochastic parameter set. Then, we will proceed the same way as in the study for filtering when such robust set exists for the whole SPDE.

As in [14], we also consider two special cases of the combined model (23): the *additive model* when we have only the additive bias correction

$$\begin{aligned} d\hat{u}_k(t) &= ((-\bar{d} + i\omega_k)\hat{u}_k(t) + b_k(t) + \tilde{f}_k(t))dt + \sigma_k dW_k(t), \\ db_k(t) &= (-\gamma_{b,k} + i\omega_{b,k})b_k(t)dt + \sigma_{b,k} dW_{b,k}(t), \end{aligned} \quad (24)$$

and the *multiplicative model* when we have only the multiplicative bias correction

$$\begin{aligned} d\hat{u}_k(t) &= ((-\gamma_k(t) + i\omega_k)\hat{u}_k(t) + \tilde{f}_k(t))dt + \sigma_k dW_k(t), \\ d\gamma_k(t) &= -d_{\gamma,k}(\gamma_k(t) - \bar{d})dt + \sigma_{\gamma,k} dW_{\gamma,k}(t), \end{aligned} \quad (25)$$

where \bar{d} is the mean value of the damping.

For the true signal, we only have noisy observations of $\hat{u}_{k,m} \equiv \hat{u}_k(t_m)$ at discrete times t_m but not of $b_k(t)$ and $\gamma_k(t)$ since the last two variables are parameters artificially introduced in the model. Therefore, the observation $\hat{v}_{k,m}$ at discrete time t_m is modeled by

$$\hat{v}_{k,m} = \hat{u}_{k,m} + \hat{\sigma}_{k,m}^o, \quad (26)$$

where $\hat{\sigma}_{k,m}^o$ is a Gaussian noise with mean zero and variance \hat{r}^o . The SPEKF uses the exact mean and second moment to solve the nonlinear filtering problem (23) and (26). As in [14], we refer to this approach as SPEKF-C. Alternatively, we also have SPEKF-A that solves the linear filtering problem (24) and (26) and SPEKF-M for nonlinear problem (25) and (26). Note that in each of these three strategies, the prior mean and covariance are solved analytically using the calculus tricks introduced in [15,16] for filtering slow-fast systems. For detailed computation of these statistics, one can consult [14].

To conclude this section, we review and summarize some of the important results for filtering the one-mode case in [14], which include: (1) SPEKF-C is the method of choice for filtering with model errors since its high filtering skill (nearly as good as the perfect model in many regimes) is robust and the least sensitive to variations of stochastic parameters, observation time and observation error variance; (2) the simpler strategies, SPEKF-A and SPEKF-M, clearly beat MSM in most parameter regimes and sometimes their skill is as good or even slightly supersedes the skill of SPEKF-C. The multiplicative method is better than the combined model when forcing is specified correctly, this is because the additive bias term, $b(t)$, introduces sampling error. However, SPEKF-A and SPEKF-M are not as robust as SPEKF-C towards the changes in the stochastic parameters; (3) we also note that the advantage of SPEKFs over MSM in addressing the model errors is more significant when regime transitions occur more often; (4) when the external forcing is unspecified or incorrectly specified, the additive bias correction term, $b_k(t)$, in SPEKF-C and SPEKF-A learns the unknown part of the forcing and the combined and additive models

still perform very well with RMS errors comparable to the RMS error of the perfectly specified filter. This high skill was not found in SPEKF-M when forcing is unspecified or incorrectly specified.

3.3. Implementation of SPEKF on sparsely observed spatially extended systems

In practice, observations may not be always available at every grid point when models are resolved with higher resolution and in many situation sparse observations are available at random locations. In this paper, we consider a situation where sparse observations are available at every P grid points as in [19,20]. Strategies for randomly located observations were reported elsewhere by one of the authors [29].

Suppose the observations of the true signal defined in (2) are taken at $2M + 1$ grid points which are regularly spaced, i.e., at $\tilde{x}_j = j\tilde{h}$, $j = 0, 1, \dots, 2M$ 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 \tilde{x}_j coincides with x_{q+jP} for some $q \in \mathbb{Z}$, and

$$P = \frac{2N + 1}{2M + 1}, \tag{27}$$

defines the ratio of the total number of mesh points available to the number of sparse regular observation locations in (2). Therefore, the model for sparse observation is given by

$$v(\tilde{x}_j, t_m) = u(\tilde{x}_j, t_m) + \sigma_m^o, \quad \tilde{x}_j = j\tilde{h}, \quad j = 0, 1, \dots, 2M, \tag{28}$$

where σ_m^o is a Gaussian noise with mean zero and variance r^o . In our numerical simulation, we will consider rather sparse observations with $M = 7$ and $N = 52$, that is, 15 observations uniformly distributed at every $P = 7$ grid points of a total of 105 model grid points. We take the observations at every $\Delta t = t_{m+1} - t_m = 0.25$ time units which correspond to 6 h in physical time. In each Fourier mode, this observation time corresponds to roughly one third of the damping time scale $1/\bar{d} = 0.66$. However, our result is not restricted to this specific observation time, e.g. see [14] for a complete study of the fifth mode of (2) with longer observation times.

The sparse observation model in (28) can be represented in Fourier space as follows:

$$\hat{v}_{\ell,m} = \sum_{k \in \mathcal{A}(\ell)} \hat{u}_{k,m} + \hat{\sigma}_m^o, \quad |\ell| \leq M, \tag{29}$$

where $\hat{\sigma}_m^o$ is a Gaussian noise with mean zero and variance $r^o = r^o/(2M + 1)$ and

$$\mathcal{A}(\ell) = \{k | k = \ell + (2M + 1)q, |k| \leq N, q \in \mathbb{Z}\}, \tag{30}$$

is the aliasing set of wavenumber ℓ , which has P components (see [29,19,20] for details). In earlier work [19], two of the authors demonstrated high filtering skill in the perfectly specified filter context with no stability transitions (described in Section 2.3) by reducing the filtering problem (2) and (28) to M decoupled filters, each consisting of P -dimensional diagonal Langevin equation (6) with scalar observation model (29). Furthermore, they showed that further filter reduction to independent scalar filters can still produce high filtering skill when the energy spectrum is decaying (i.e., $E_k = k^{-3}$ in our case).

One type of filter reduction [19] is to require the aliased modes in each aliasing set, $k \in \mathcal{A}(\ell)$, $k \neq \ell$, to fully trust the dynamics since the observed (resolved or primary) mode ℓ is the most energetic mode. This approximation, called the “Reduced Fourier Domain Kalman Filter” (RFDKF), models the observation in the following fashion

$$\hat{v}'_{\ell,m} \equiv \hat{v}_{\ell,m} - \sum_{k \in \mathcal{A}(\ell), k \neq \ell} \hat{u}_{k,m} = \hat{u}_{\ell,m} + \hat{\sigma}_m^o, \tag{31}$$

such that it is in the form of (26). In (31), observation \hat{v}_ℓ is adjusted by the summation of the aliased modes. These aliased modes are propagated forward with the mean model in (20) without taking into account any observations.

In this paper, we implement SPEKFs (discussed in Section 3.2) with observation model RFDKF in (31). Numerically, we only filter M resolved modes (which are wavenumbers 1–7 in our numerical example) with SPEKF and propagate the remaining modes, $k = 8, \dots, 52$, with the mean model in (20) without taking into account any observations. For diagnostic purposes, we also implement the perfectly specified filter and MSM (discussed in the beginning of Section 3 and 3.1.1) with the observation model in (31).

4. Numerical results for filtering

In our numerical simulations, we consider solutions of the form (3) with the finite number of modes $N = 52$. This corresponds to $2N + 1$ discrete points on the 2π periodic interval. The true signal as shown in Fig. 2 is generated for a total of 500 days, which corresponds to $T = 2000$ assimilation cycles with observation time $\Delta t = 0.25$. Observations are generated at uniformly distributed locations as described earlier in (28) with error noise variance $r^o = 0.3$ in the correctly specified forcing case and $r^o = 0.5$ in the incorrectly specified forcing case. Both noise variances in the k -space, $r^o/(2M + 1)$, exceed

the equilibrium energy spectrum $E_k = k^{-3}$ of the unforced system for $k > 3$. When the turbulent signal is externally forced (with (16) in our example), its energy spectrum is given as follows:

$$E_k = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_0^T |\hat{u}_k(s)|^2 ds. \tag{32}$$

In Fig. 3(b), we find that the above choice of observation error variance, indeed, exceeds the energy spectrum (32) for wavenumbers $k > 7$. In this figure, the energy spectrum (32) is approximated by averaging over a finite amount of time $T = 450$ days after a transient period of 50 days. For the remainder of this paper, the term “energy spectrum” refers to the spectrum of the forced system defined in (32), and not to that of the unforced case, k^{-3} . We will discuss results with different noise level in Section 4.3.

We quantify the performance by measuring the Root-Mean-Square (RMS) difference between the true signal, $u^t(x_j, t_m)$, and the filtered solution, $\bar{u}(x_j, t_m)$,

$$\mathcal{E} = \sqrt{\frac{1}{\tilde{M}(2N+1)} \sum_{m=1}^{\tilde{M}} \sum_{j=1}^{2N+1} |\bar{u}(x_j, t_m) - u^t(x_j, t_m)|^2}, \tag{33}$$

where $t_{\tilde{M}} = T$. Below, we also qualitatively compare the energy spectra generated from the true signal with those generated through various posterior filtered solutions.

For SPEKFs, we choose stochastic parameters that belong to the robust set of parameters based on the comprehensive study in [14]; they are $\{\gamma_{b,k} = 0.1\bar{d}, \omega_{b,k} = \omega_k, \sigma_{b,k} = 4\sigma_5 = 0.6, d_{\gamma,k} = 0.1\bar{d}, \sigma_{\gamma,k} = 4\sigma_5 = 0.6\}$ for filtered wavenumbers $k = 1, \dots, 7$. In Section 4.3, we will check the robustness of the filter skill when these parameters are changed.

4.1. Correctly specified forcing

In Fig. 3(a), we compare the RMS errors of the nine most energetic modes. The perfectly specified filter provides the smallest error as expected since the perfect model utilizes the exact dynamics. The multiplicative model produces a filtered solution with error just slightly larger than the error of the perfectly specified filter. Note that the error of the multiplicative model is still very low for both stable and unstable regions of the spectrum (as described in Table 1 instability occurs in modes 3–5). Next, the combined and additive models produce the filtered solutions with errors that are almost the same as the errors of the perfectly specified filter for modes 1 and 2 and then deviate from the corresponding values of the perfectly specified filter error for higher modes. And finally, the MSM has a very large error for the unstable modes 3–5. On modes higher than 5, MSM is the perfectly specified filter since the true damping of these modes is exactly equal to the average damping of the system $\bar{d} = 1.5$. One important bulk quantity to recover from filtering is the energy spectrum (32) of the true signal. In Fig. 3(b), we see that the energy spectrum of the posterior state of MSM underestimates the true energy spectrum of the unstable modes. On the other hand the multiplicative and combined models produce filtered solutions with energy spectrum very close to the true energy spectrum for all seven filtered modes. The energy spectrum of the additive model is also close to the true energy spectrum with just slight deviations from it.

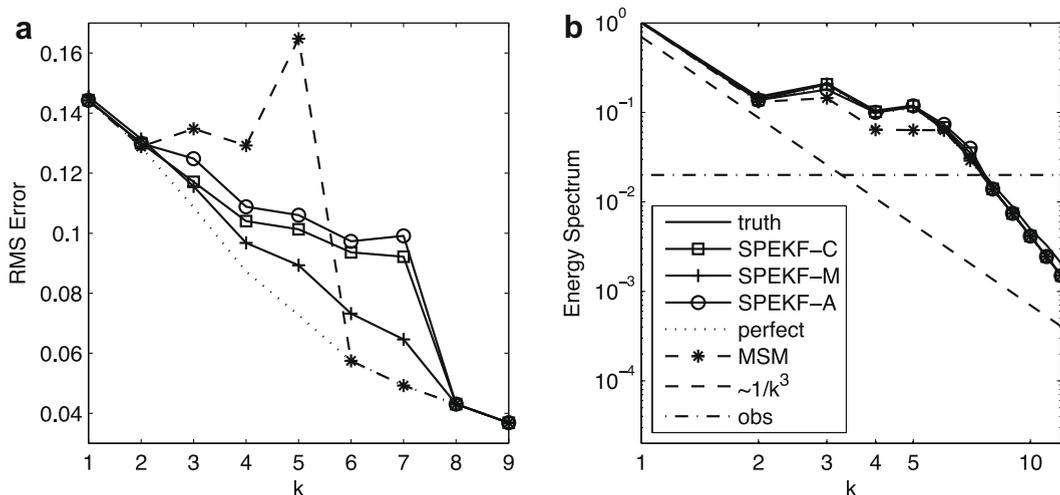


Fig. 3. Filtering with correctly specified forcing: (a) RMS errors as functions of wavenumbers. (b) Energy spectrum as a function of wavenumbers.

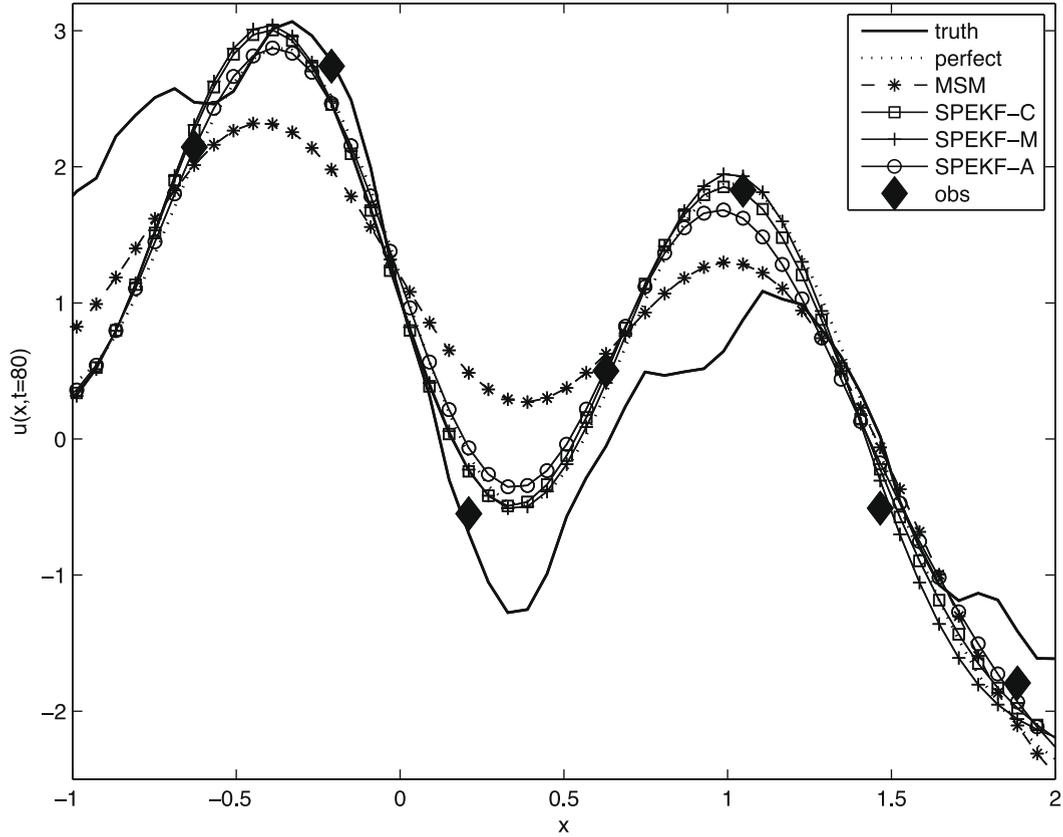


Fig. 4. Filtering with correctly specified forcing: snapshot of the true trajectory in the unstable regime at time $t = 80$ together with filtered signals as a function of space. We only show the region with $x \in [-1, 2]$ for clarity of presentation.

In Fig. 4, we show a snapshot of the true trajectory in the unstable regime at $t = 80$ together with the posterior values of various filtered solutions. There, we observe that in the unstable regime all SPEKFs produce filtered solutions that are much closer to the true signal relative to that of the MSM. This is explained by the fact that the multiplicative bias correction $\gamma_k(t)$ and additive bias correction $b_k(t)$ help the SPEKFs to recover the true dynamics better than via the MSM with averaged damping, especially in the unstable modes 3–5, which we will discuss next.

In Fig. 5, we show the time series of the additive and multiplicative bias corrections, $b_k(t)$ and $\gamma_k(t)$, respectively, for wavenumbers $k = 3, 4, 5$ with intermittent unstable damping. In the panels for $\gamma_k(t)$, we also show the true value of the damping as discussed in Section 2.3. We note that $\gamma_k(t)$ of both the combined and multiplicative models follow the trajectory of the true damping. However, the multiplicative model produces a better estimate of the true damping relative to the combined model in all three unstable modes. As a result, we expect the multiplicative model to produce a better approximation to the true signal $u(x, t)$, which is confirmed in Table 2. On the other hand, the panels of Fig. 5 that correspond to the additive bias correction $b_k(t)$ show that $b_k(t)$ do not deviate much from zero, except for the times when the damping is unstable. In the unstable regime, both the combined and additive models use the additive bias correction to recover the true signal. However, since the model error is multiplicative for the correctly specified forcing, we expect the multiplicative model to show the best performance among the three SPEKFs. This result is also confirmed in Table 2, in which we present the RMS errors in real space, computed using the error formula in (33). In the same table, we also report results for the unforced case ($\hat{f}_k = 0$ both for the true signal and the filter) run with slightly smaller $r^o = 0.2$ (which is greater than k^{-3} for wavenumbers $k > 4$); here we find similar conclusions as in the forced case discussed above: the multiplicative model is the method of choice.

4.2. Unspecified forcing

Here, we consider a true signal with forcing given by

$$\hat{f}_k(t) = A_{f,k} \exp(i(\omega_{f,k}t + \phi_{f,k})), \tag{34}$$

for $k = 1, \dots, 7$ with amplitude $A_{f,k}$, frequency $\omega_{f,k}$ and phase $\phi_{f,k}$ drawn randomly from uniform distributions,

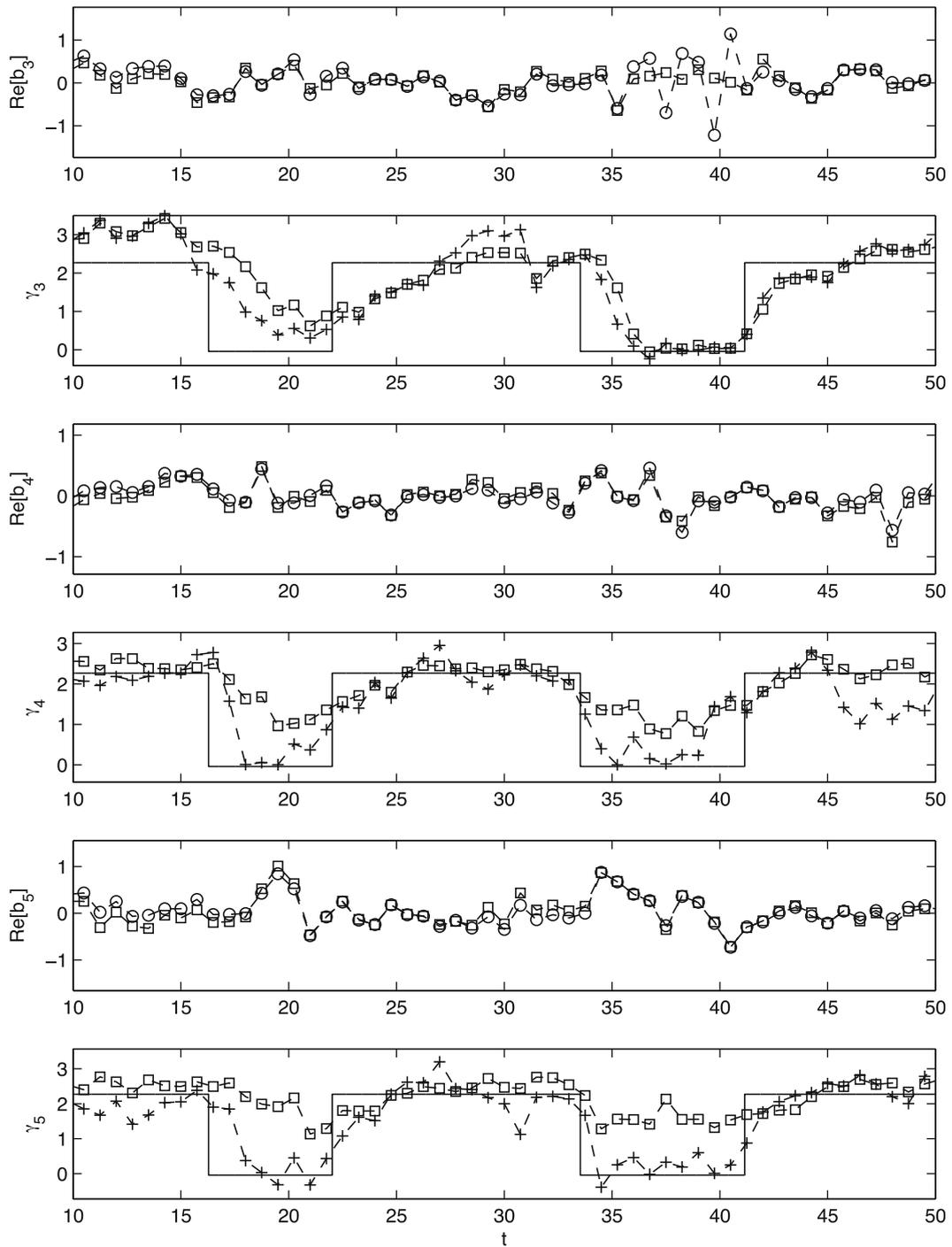


Fig. 5. Filtering with correctly specified forcing: additive bias correction, $b_k(t)$, and multiplicative bias correction, $\gamma_k(t)$, for SPEKF-C (squares), SPEKF-A (circles) and SPEKF-M (pluses) for the modes $k = 3, 4, 5$ with unstable damping. The solid line shows the true damping $\gamma(t)$.

$$\begin{aligned}
 A_{f,k} &\sim U(0.6, 1), \\
 \omega_{f,k} &\sim U(0.1, 0.4), \\
 \phi_{f,k} &\sim U(0, 2\pi), \\
 \hat{f}_k &= \hat{f}_{-k}^*,
 \end{aligned}$$

Table 2

Spatial RMS errors for simulations with unforced case, correctly forced case and incorrectly forced case.

Forcing	Unforced case	Correct forcing	Incorrect forcing
r^o	0.2	0.3	0.5
RMSE of perfect filter	0.346	0.391	0.454
RMSE of MSM	0.394	0.477	0.728
RMSE of $MSM_{f=0}$	–	–	1.169
RMSE of SPEKF-C	0.380	0.444	0.588
RMSE of SPEKF-M	0.359	0.418	0.787
RMSE of SPEKF-A	0.398	0.457	0.601

and unforced, $\hat{f}_k(t) = 0$, for modes $k > 7$. However, we do not specify this true forcing to the filter model, i.e., we use $\tilde{f}_k = 0$ for all modes.

In Fig. 6, we compare the RMS errors and the energy spectra of the true signal and of the filtered solutions mode by mode. In terms of RMS errors, SPEKF-C is the best strategy followed by SPEKF-A and both RMS errors are smaller than that of the MSM with perfect forcing in the unstable modes 3–5. The high filtering skill with SPEKF-M in the perfect forcing case deteriorates as the forcing is incorrectly specified; in this case, the absence of additive bias correction term, b_k , in the multiplicative model degrades the filtering skill significantly. In terms of energy spectrum, we find that the first 7 most energetic modes are filtered very well by the perfectly specified filter as well as by the combined and additive models if we compare with the spectrum of the true signal. The multiplicative model does not provide a good approximation of the energy spectrum for the wavenumbers $k > 3$. The MSM with correctly specified forcing misses the unstable modes 3–5, while the rest of the modes are filtered well. The MSM with unspecified forcing only filters the first two modes well, while missing the rest of the modes. Similar conclusions hold when spatial RMS error, reported in Table 2, is used for the performance measure.

4.3. Robustness and sensitivity to stochastic parameters and observation error variances

In this section, we study how sensitive the proposed SPEKFs are to variations of stochastic parameters and how the skill of the filters vary for different values of observation error variance, r^o , for a fixed set of stochastic parameters. In this study, we keep all but one of the parameters fixed and vary the remaining parameter in a broad range of values. The fixed parameters are $\{d_{\gamma,k} = 0.1d, \gamma_{b,k} = 0.1d, \sigma_{\gamma,k} = 4\sigma_5, \sigma_{b,k} = 4\sigma_5, \omega_{b,k} = 1\}$. It is very important to realize that we use the same set of stochastic parameters for all the switching modes, that is, wavenumbers $|k| \leq 7$. These modes have different energies and different correlations in time. Therefore, using the same set of stochastic parameters for a number of modes is a tough test for the robustness of the SPEKFs.

In Fig. 7, we demonstrate the dependence of the spatially averaged RMS errors of the various filters on the stochastic parameters and observation variance for the correctly specified forcing case. We note that both SPEKF-M and SPEKF-C are robust to the variations of d_γ/d and σ_γ but SPEKF-M has smaller RMS errors relative to SPEKF-C (see panels (a) and (b) in Fig. 7). The robustness in SPEKF-C toward these two parameters is very similar to the robustness of the fifth mode studied in [14]. For the SPEKF-M, the sensitivity toward large damping and small noise on the fifth mode study [14] disappears when all modes are accounted for and this is because the errors in the non-switching modes are smaller. The sensitivity of the fil-

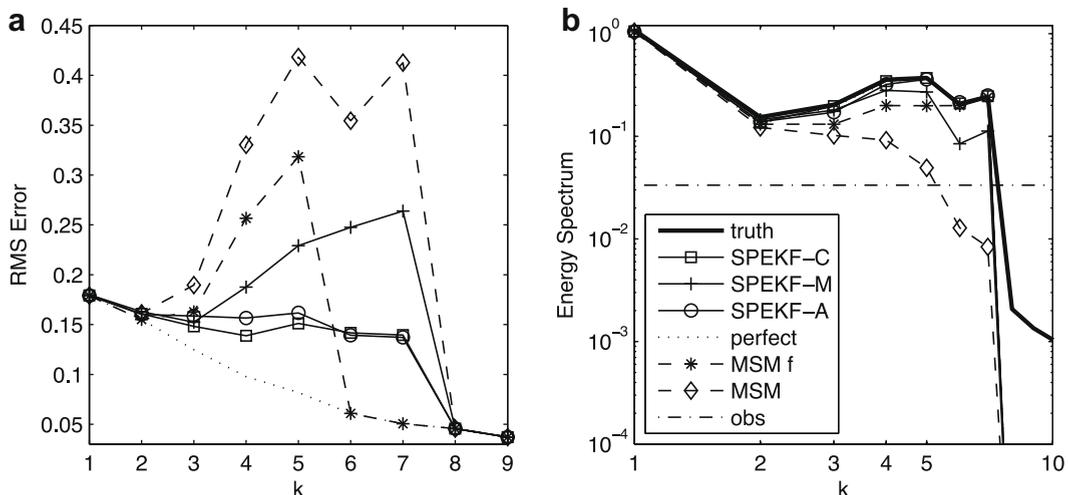


Fig. 6. Filtering with unspecified forcing: (a) RMS errors as functions of wavenumbers. (b) Energy spectrum as a function of wavenumbers.

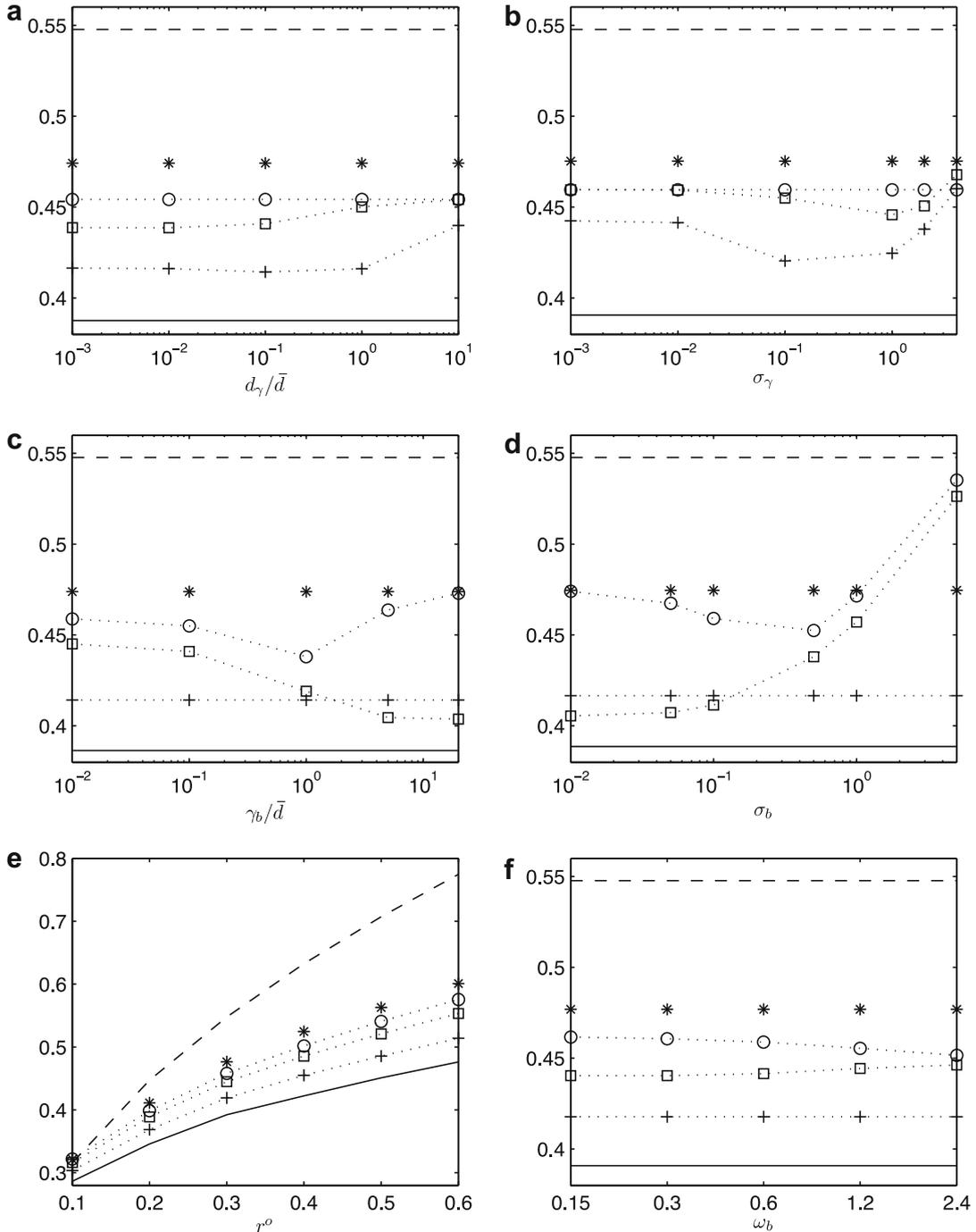


Fig. 7. Filtering with correctly specified forcing: spatially averaged RMS errors of the perfectly specified filter (solid line), CSM (asterisks), SPEKF-C (squares), SPEKF-M (pluses), SPEKF-A (circles) and observation error (dashes). The fixed parameters had the values $d_\gamma/\bar{d} = 0.1$, $\gamma_b/\bar{d} = 0.1$, $\sigma_\gamma = 0.6$, $\sigma_b = 0.6$, $\omega_b = 1$, $r^o = 0.3$.

tering skill toward variations of the additive stochastic parameters, γ_b , ω_b , σ_b reflects the one-mode study on the fifth mode (see panels (c) and (e) in Fig. 7); SPEKF-C is robust toward variations both in the damping and frequency phase. For noise strength σ_b , SPEKF-C produces rather high errors when the magnitude of σ_b approaches the observation error; this result exactly reflects what we found in the one-mode study in [14], except there the observation error is rather small. From Fig. 7(e), we conclude that the RMS errors of all the filters increase as functions of r^o . Moreover, the multiplicative model (SPEKF-M) performs better than the combined model, SPEKF-C, which in turn is better than the additive model, SPEKF-A.

When forcing is incorrectly specified (see Section 4.2), the situation changes considerably. In this case, the skill of SPEKF-C and SPEKF-A is very robust with respect to variations of multiplicative stochastic parameters, d_γ , σ_γ (see panels (a) and (b) in Fig. 8). On the other hand, SPEKF-M is not skillful at all as we found and discussed in Section 4.2. When the additive stochastic parameters are chosen such that γ_b is small and σ_b is rather large, then the skill of the combined and additive models is good (see Fig. 8(c) and (d)). There is one particular regime when both the combined and additive models fail, that is, the re-

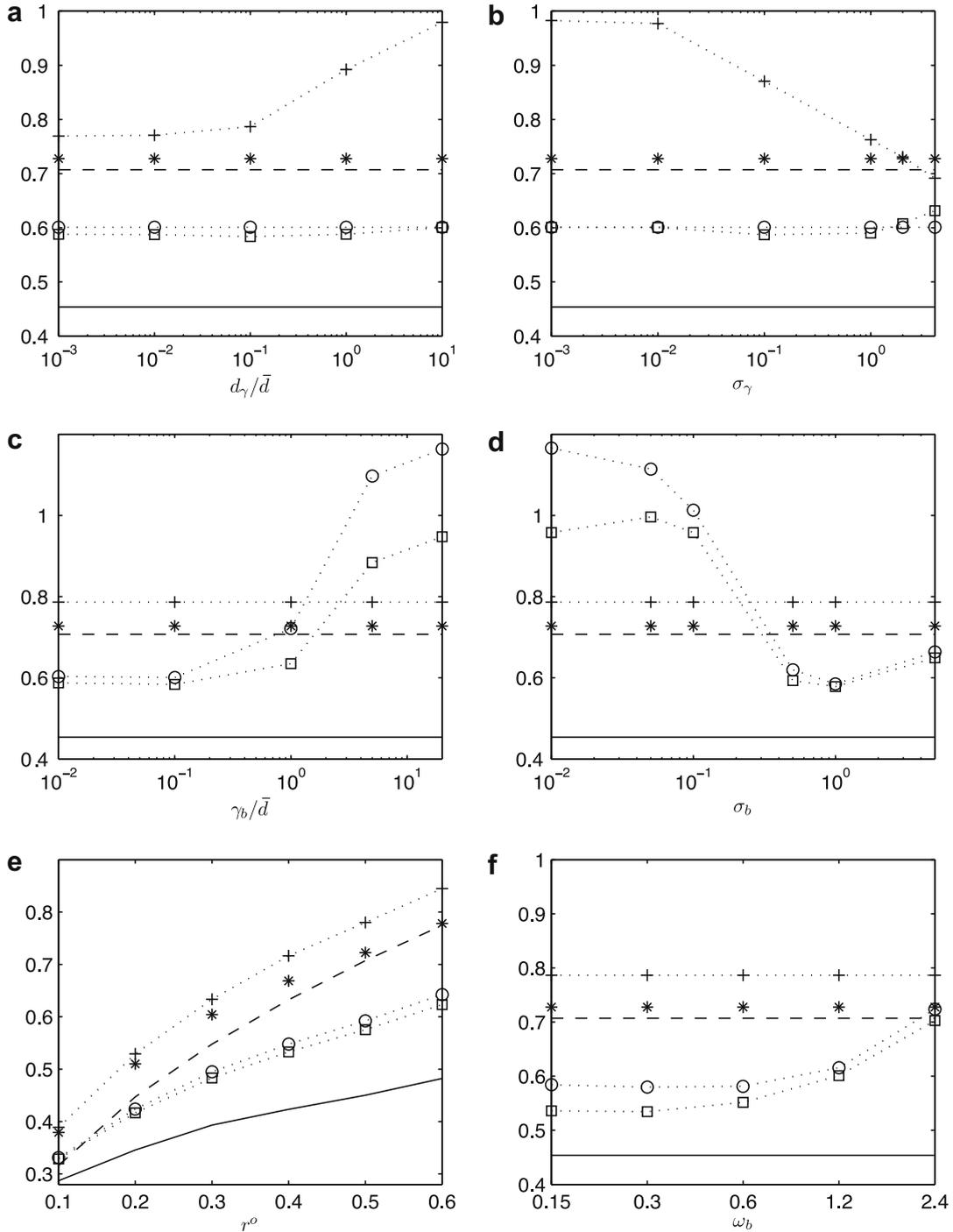


Fig. 8. Filtering with unspecified forcing: spatially averaged RMS errors of the perfectly specified filter (solid line), CSM (asterisks), SPEKF-C (squares), SPEKF-M (pluses), SPEKF-A (circles) and observation error (dashes). The fixed parameters had the values $d_\gamma/\bar{d} = 0.1$, $\gamma_b/\bar{d} = 0.1$, $\sigma_\gamma = 0.6$, $\sigma_b = 0.6$, $\omega_b = 1$, $r^o = 0.5$.

gime when the additive noise is too small. In this regime we naturally expect the combined and additive model to perform poorly since the additive bias correction is needed to recover the unspecified forcing. Next, the performance of the combined and additive models is quite robust to the changes of ω_b unless this parameters takes values of order 10 or larger. (see Fig. 8(f)). So we conclude that in our model the frequency of the additive bias correction $b_k(t)$ should not exceed the maximum frequency of the system which is equal $\omega_k = 8.91$ for $k = 1$. In Fig. 8(e), we demonstrate the performance of the filters as a function of observation variance r^o . Here we see that the combined model is the best for the whole range of observation variance. As we already pointed out earlier in this Section, it is very important to realize that the same set of stochastic parameters is used for all wavenumbers with $|k| < 7$ even though initially this set was chosen for the fifth mode of our model as in [14]. Here, we have shown that this set of stochastic parameters belongs to a broad range of robust stochastic parameters which produce similar high skill filtered solutions with SPEKFs, and, therefore, one should not tune the stochastic parameters for a specific mode.

4.4. Summary

As an important test, the posterior energy spectra from both the SPEKF-M and SPEKF-C recover the energy spectrum of the true signal with high accuracy while MSM has large errors for both correctly and incorrectly specified forcing (see Figs. 3 and 6). In the correctly specified forcing case, the multiplicative model (SPEKF-M) is better than the combined model since the model errors are completely introduced by the multiplicative noise and hence the additional correction with additive term is redundant. However, we can always choose an additive parameter set for the combined model such that SPEKF-C behaves exactly like SPEKF-M. As a method of choice, we strongly recommend the combined model (SPEKF-C), which is a very effective and robust filtering strategy when neither the true damping nor the true forcing are specified. The multiplicative and additive bias corrections learn the damping and the forcing from the observations and significantly improve the filtering skill of the SPEKF-C when compared with the skill of the MSM for both correctly and incorrectly specified forcing.

5. Predictive skill

In this section, we extensively study the potential predictive skill improvement of using our additive, multiplicative and combined models in (24), (25) and (23), respectively, over the mean model MSM in (20). Our goal is to first understand the predictive skill in a one-mode context, especially to understand the effect of errors in initial conditions and a finite ensemble of realizations. We will check whether there are robust parameter regimes for stochastic parameter estimation with high predictive skill and finally whether we can translate the results in this one-mode study to the spatially extended system for the switching SPDE as discussed earlier.

5.1. One-mode study: perfect initial conditions

To achieve our goals, we first consider a “super-ensemble” setup, that is, noiseless initial conditions and forecasts with exact statistics. In particular, we consider solutions of

$$\frac{du(t)}{dt} = (-\gamma(t) + \omega i)u(t) + f(t), \quad (35)$$

as the perfect initial conditions. Following [14], we fix the frequency $\omega = 8.91/k$ with $k = 5$ to represent the fifth mode of our switching SPDE and a periodic forcing $f(t) = \exp(0.15it)$. We allow the damping $\gamma(t)$ to switch between stable, $d^+ = 2.27$, and unstable, $d^- = -0.04$, regimes with random switching time as described in Section 2.3 with average damping strength of $\bar{d} = 1.5$. In our numerical simulations, we consider 10,000 initial conditions, (u, b, γ) with $b = 0$, by sampling solutions of (35) at every 2.5 time units. To generate noiseless forecasts in the super-ensemble setup, we use the exact mean and covariance of each model [14] with zero initial covariances. In our numerical experiments, we propagate these statistics for 2.5 time units, which is about four times the average damping time of the true signal $1/\bar{d} = 0.67$. With this forecast time, 2.5 units, the 10,000 initial conditions are practically independent.

To quantify the predictive skill, we compute the following average RMS error

$$\text{RMS}(\tau) = \frac{1}{|\mathcal{B}|} \sum_{t_i \in \mathcal{B}} \sqrt{|\bar{u}^m(t_i + \tau|t_i) - u^t(t_i + \tau)|^2}. \quad (36)$$

In (36), this RMS error is averaged over predictions with initial times t_i that belong to set \mathcal{B} . In (36), we denote $|\mathcal{B}|$ as the number of components in subset \mathcal{B} , $\bar{u}^m(t_i + \tau|t_i)$ as the mean forecast with any model (including the additive, multiplicative, combined and MSM) starting at time t_i for lead time τ , and $u^t(t_i + \tau)$ as a solution of 35 at time $t_i + \tau$. We will consider different sets of initial condition, including: (a) the whole initial conditions, W , (b) those in the stable regime, S , (c) in the unstable regime, U , (d) initial conditions at the instances just after a transition from stable to unstable, A_s and (e) initial conditions at the instances just after a transition from stable to unstable, A_u . In our particular noiseless true signal, we have $|W| = 9989$ omitting the last 2.75 time units from the 10,000 initial conditions, $|S| = 6692$, $|U| = 3296$, $|A_u| = 127$ and $|A_s| = 157$.

Theoretically, there is no difference between the multiplicative and combined models in this setup since $b = 0$ and all the initial covariances are zero. On the other hand, the additive model is nothing more than the MSM since both use constant damping \bar{d} with, again, $b = 0$. We find that the predictive skill is not very sensitive to variations of noise strength, σ_γ , but it cannot be too large. When forcing is specified correctly, we obtain better skill with weaker damping, $d_\gamma/\bar{d} = 0.01$ or 0.1 . When the damping is too strong, the predictive skill of the combined (multiplicative) model becomes similar to the predictive skill of MSM. In Fig. 9, we show the RMS errors as functions of lead time τ for the correct forcing case, where we use

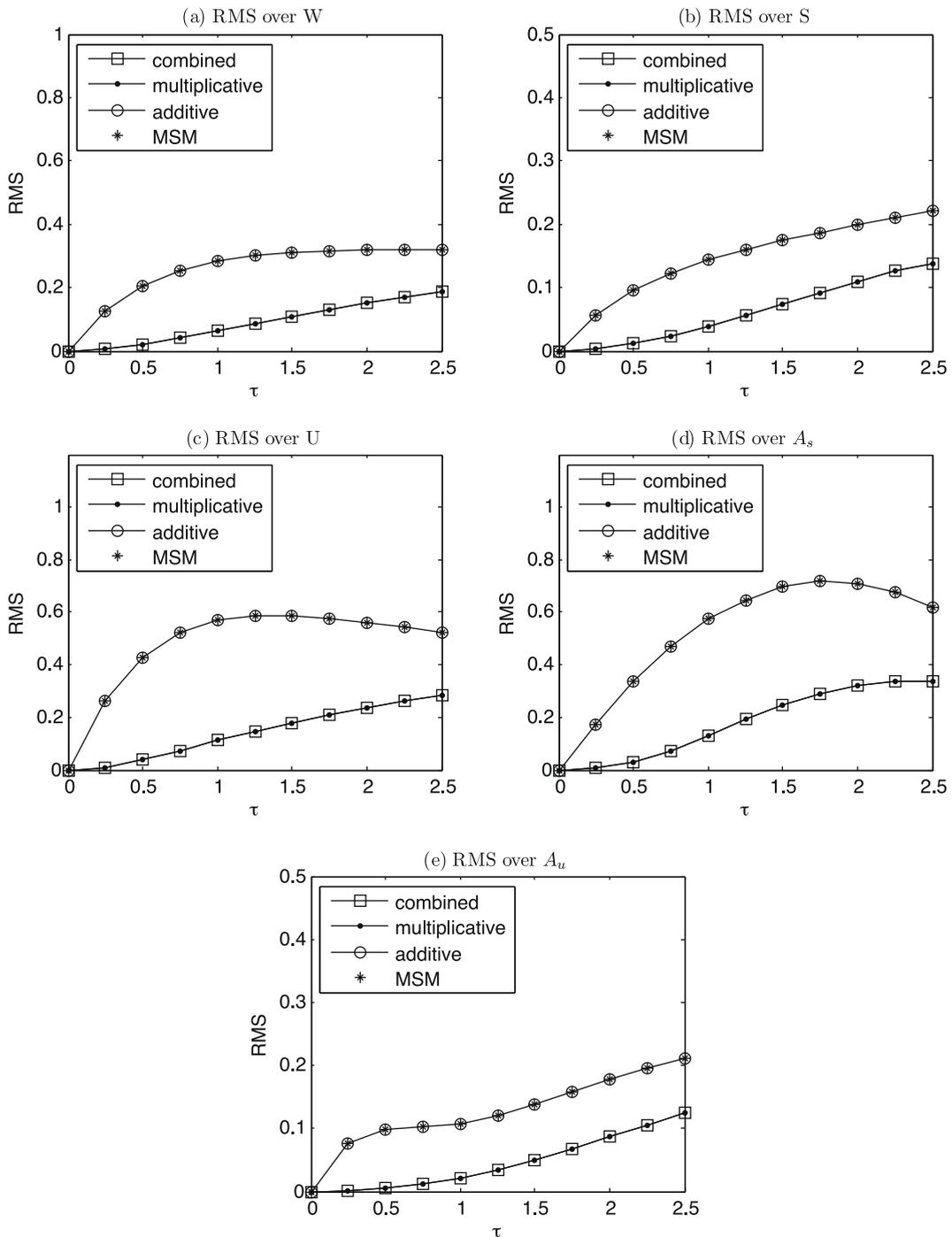


Fig. 9. Predictive skill of the one-mode ODE with perfect initial conditions: average RMS errors over W , S , U , A_s , A_u as functions of lead time for correctly specified forcing. Compare with Fig. 10 for the case of unspecified forcing.

stochastic parameters $d_\gamma = 0.01\bar{d}$, $\sigma_b = \sigma_\gamma = 0.1\sigma$, $\omega_b = \omega$, and an arbitrary γ_b . Here, we see a significant forecast improvement with the combined (multiplicative) model relative to the MSM (additive) model, especially in subsets U and A_s . On the other hand, when forcing is unspecified (in our numerical simulation, we set $\tilde{f}(t) = 0$ for our filter model), we find that stronger damping d_γ is necessary yet the advantage of the combined (multiplicative) model over the MSM (additive) is significantly reduced (see Fig. 10).

In Fig. 11, we also show the effect of finite ensemble size for the correctly specified forcing as in Fig. 9. Here, the mean forecast is generated by solving (23)–(25) using an ensemble size of $K = 10^n$, $n = 1, 2, 3, 4$ and averaging over these ensem-

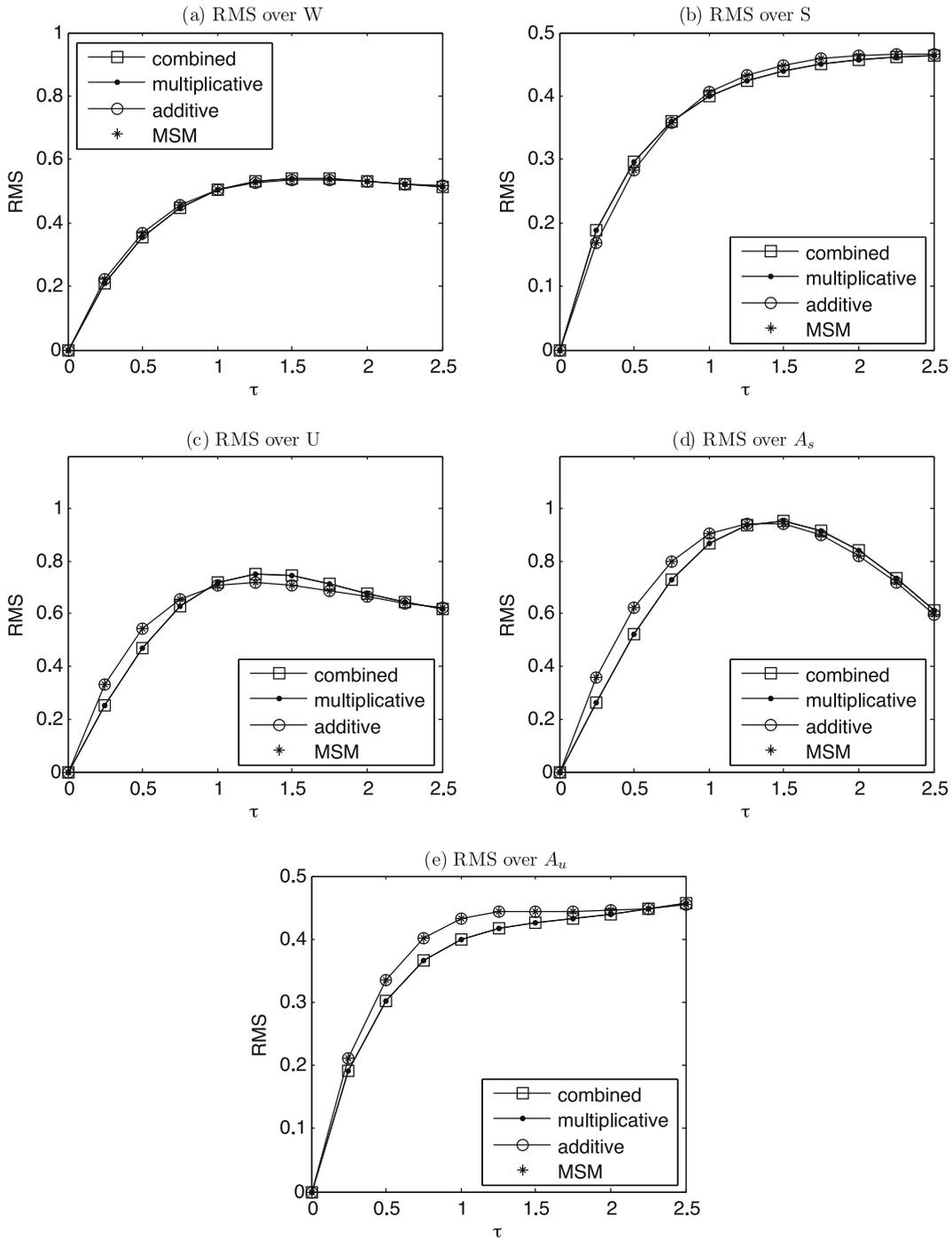


Fig. 10. Predictive skill of the one-mode ODE with perfect initial conditions: average RMS errors over W , S , U , A_s , A_u as functions of lead time for unspecified forcing. Compare with Fig. 9 for the case of correctly specified forcing.

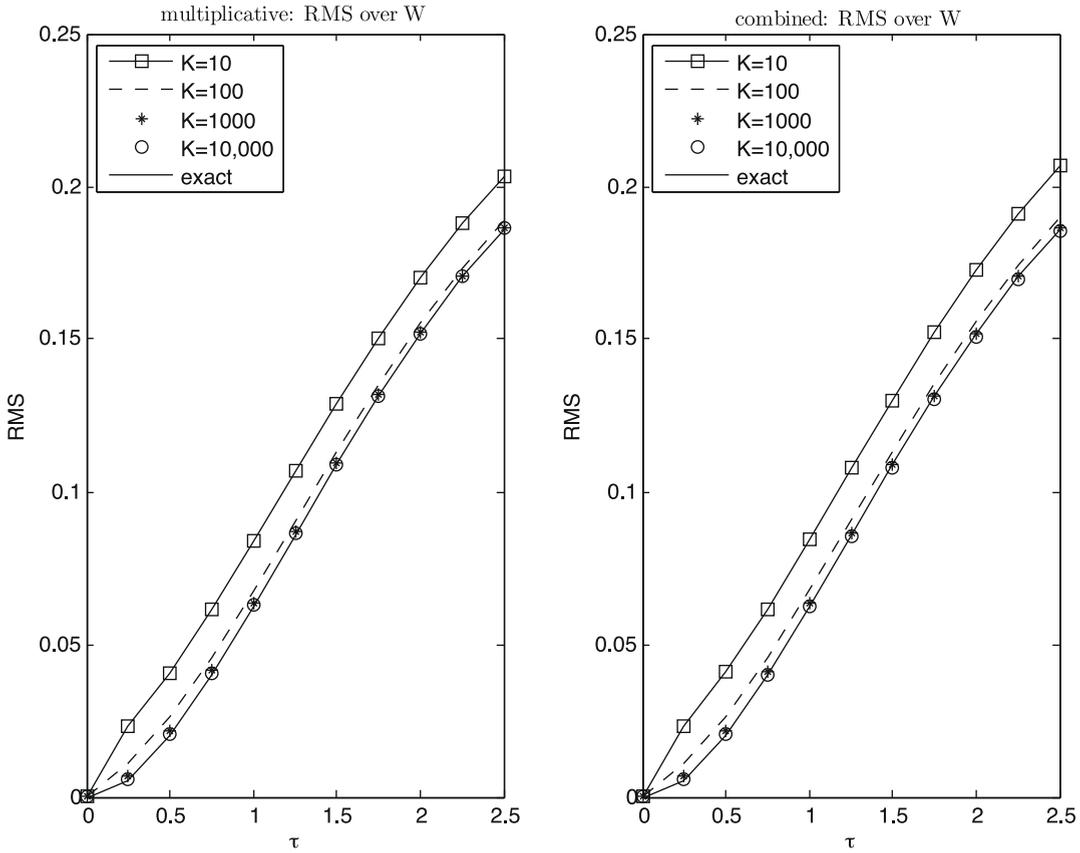


Fig. 11. Predictive skill of the one-mode ODE with perfect initial conditions: average RMS errors over W as functions of lead time for correctly specified forcing with finite ensemble. Multiplicative model (left panel) and combined model (right panel).

bles instead of using exact statistics as computed in [14]. From these experiments, we find tiny forecast degradation when an ensemble of size 10 is used with both the combined and multiplicative models. For the additive model, there is almost no difference between the finite ensemble experiment with $K \leq 10$ and the perfect statistics (result is not reported in Fig. 11).

To conclude this section, we define a quantity for indicating the average predictive skill improvement using the combined model relative to MSM over time interval $0 < t < \tau_n$ by

$$\langle \text{Impr} \rangle_{0 < t < \tau_n} = \frac{1}{n} \sum_{i=1}^n \left(1 - \frac{\text{RMS}_C(\tau_i)}{\text{RMS}_{\text{MSM}}(\tau_i)} \right), \quad \tau_i = i\Delta t, \tag{37}$$

where $\tau_n = \min(n\Delta t, 2.5)$ is the maximum time for the combined model to provide better forecast skill relative to MSM, i.e., $\text{RMS}_C(n\Delta t) < \text{RMS}_{\text{MSM}}(n\Delta t)$. Table 3 shows the average forecast improvement of the combined model relative to MSM for different subsets. We see a significant improvement (70% on average) up to lead time $\tau_n = 2.5$ (four times damping time, 0.67) for the correctly specified forcing. On the other hand, when forcing is unspecified, the improvement reduces significantly; the best performance is about 10% improvement in the unstable regime U up to lead time $\tau_n = 1.25$ (which is still twice the damping time).

Table 3
Average forecast improvement with the combined model over MSM for perfect initial condition.

Subset	Correct forcing $\langle \text{Impr} \rangle_{0 < t < \tau_n}$ (%)	τ_n	Unspecified forcing $\langle \text{Impr} \rangle_{0 < t < \tau_n}$ (%)	τ_n
W	68	>2.5	3	1
S	63	>2.5	14	0.75
U	72	>2.5	11	1.25
A_s	69	>2.5	5	2.25
A_u	69	>2.5	N/A	

5.2. One-mode study: imperfect initial conditions

Now we study the predictive skill when initial conditions are imperfect, specifically when they are produced from filtering through MSM and SPEKFs (discussed in Sections 3.1.1 and 3.2). To perform numerical simulations in the super-ensemble mode, we generate initial conditions by filtering noiseless observations, $v_m = u(t_m)$, of noiseless true signals in (35) with SPEKFs and MSM. In the following numerical experiment, we set $d_y/\bar{d} = 0.01$ for the correctly specified forcing, $d_y = \bar{d}$ for the unspecified forcing, and noise strengths $\sigma_b = \sigma_\gamma = 0.1\sigma$, as in the previous case. As in the above section, we also consider predictions with MSM, additive, multiplicative and combined models.

When initial conditions are produced by SPEKF-M, the only available components of initial conditions are u and γ . Thus, we set $b = 0$ and it becomes clear that the multiplicative and combined models are exactly identical. On the other hand, MSM is identical to the additive model since the damping term γ is not updated in these two models. In Fig. 12, we see the predictive skill improvement with combined (multiplicative) model relative to the MSM (additive) model for the correctly specified forcing case in subsets S and U but not in A_s and A_u . Here, the multiplicative bias correction terms in the combined and multiplicative models significantly improve the prediction in subsets S (on average, 46% improvement up to $\tau_n = 2.5$, see Table 4) and U (59% up to $\tau_n = 2$) but they are unable to react quickly to transitions (in A_u , A_s) since the initial damping γ contains errors.

When initial conditions are from SPEKF-A, they do not have a γ component. In this case, we choose the initial γ to be the equilibrium average damping, $\gamma = \bar{d}$, which turns off the advantage of the multiplicative bias correction. Subsequently, the multiplicative model is identical to the MSM and the additive model is identical to the combined model. In our numerical simulation with the correctly specified forcing and additive damping strength $\gamma_b = \bar{d}$, we find that the combined (additive) model has better predictive skill relative to MSM (multiplicative) model in subsets S (10% up to $\tau_n = 0.75$) and U (13% up to $\tau_n = 2.5$), see Fig. 13(c) for RMS averaged over subset U . This predictive skill improvement, however, is very sensitive toward variations of additive damping γ_b . In Fig. 13(a), we show that when $\gamma_b = 0.1\bar{d}$ the predictive skill improvement degrades after $\tau = 1$. On the other hand, when damping is too strong, $\gamma_b = 10\bar{d}$, the combined and additive models behave like MSM (see Fig. 13(e)).

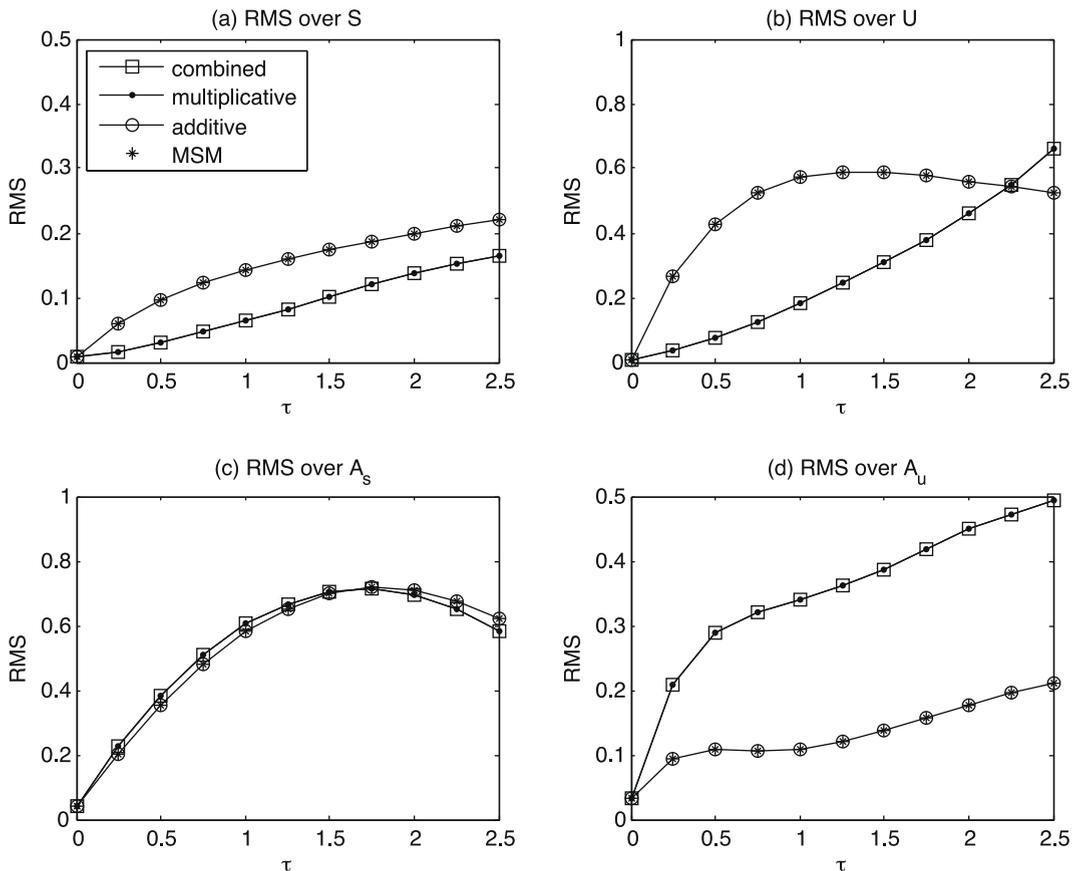


Fig. 12. Predictive skill of the one-mode ODE with imperfect initial conditions from SPEKF-M: average RMS errors over S , U , A_s , A_u as functions of lead time for correctly specified forcing.

Table 4

Average forecast improvement with the combined model over MSM for imperfect initial condition and correctly specified forcing.

Subset	SPEKF-A (Impr) _{0<t<τ_n} (%)	τ _n	SPEKF-M (Impr) _{0<t<τ_n} (%)	τ _n	NEKC-C (Impr) _{0<t<τ_n} (%)	τ _n
W	16	1.25	51	2.25	22	>2.5
S	10	0.75	46	>2.5	17	>2.5
U	13	>2.5	59	2	25	>2.5

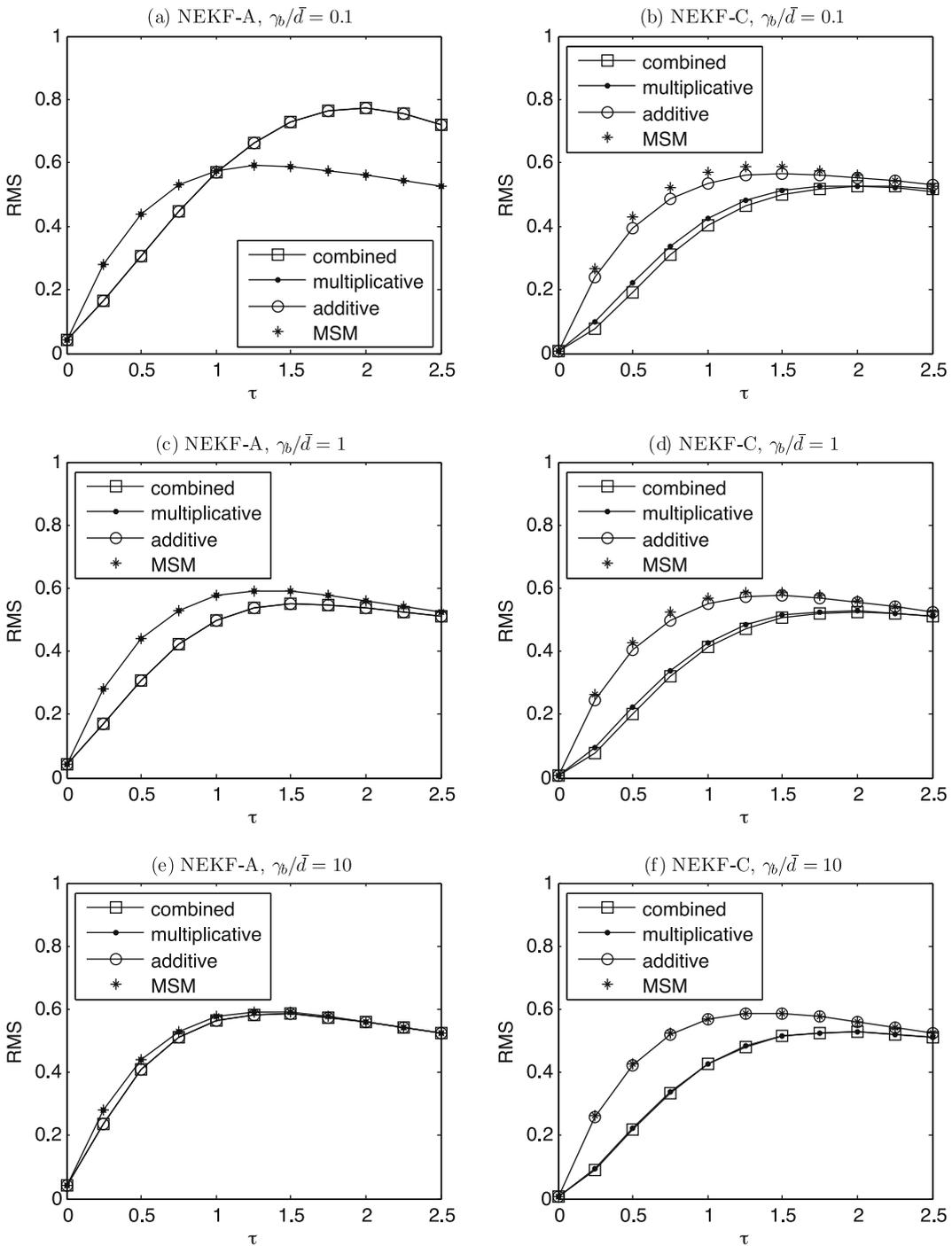


Fig. 13. Predictive skill of the one-mode ODE with imperfect initial conditions: average RMS errors over U as functions of lead time for correctly specified forcing with initial conditions from SPEKF-A (left column) and SPEKF-C (right column). Each row shows results with different γ_b .

Ultimately, we are more interested to the case where initial conditions are generated from SPEKF-C since this is the method of choice for filtering as pointed out in Section 4.4. Here, we find that not only does the predictive skill of the combined and multiplicative models increase over the additive model and MSM persist in S and U as observed earlier with perfect initial conditions and correct forcing (see Fig. 9), but it is also robust toward variations of γ_b (see panels (b), (d) and (f) in Fig. 13). In Table 4, we see that the average improvement (computed for the case $\gamma_b = \bar{d}$) is about 17% in S and 25% in U , both up to lead time $\tau_n = 2.5$. Notice that this average improvement (see Fig. 13(d)) is less than that with initial conditions from SPEKF-M (see Fig. 12(b)) because here the initial b contains error and $b \neq 0$. The failure in subsets A_s and A_u is unavoidable for similar reasons discussed earlier.

The MSM does not provide initial values for either multiplicative or additive bias corrections. The trivial choice is to impose the equilibrium initial conditions for the biases, $b = 0$ and $\gamma = \bar{d}$, which simultaneously turns off the dynamics for correcting both additive and multiplicative biases. In Table 5, we report the average improvement of the combined model relative to MSM for various initial conditions for unspecified forcing; the predictive skill improvement is significantly reduced.

5.3. Numerical results for the switching SPDE

In this section, we quantify the predictive skill for the spatially extended switching SPDE in (2). In particular, we consider the filtered solutions in Section 4 as initial conditions. Different from above simulations, these initial conditions are not generated in a super-ensemble setting. The initial conditions from the perfectly specified filter, SPEKF-M, and MSM, do not provide the b component, and we set $b = 0$ as in earlier numerical experiments for one-mode in Section 5.2. In the absence of γ component, we use the true damping value for initial conditions from the perfectly specified filter and the equilibrium damping \bar{d} for initial conditions from SPEKF-A and MSM.

For predictions with additive, multiplicative and combined models, we choose stochastic parameters from the robust parameter set of the fifth mode, as discussed in the one-mode study in Sections 5.1 and 5.2, which are, $\{\gamma_{b,k} = \bar{d}, \omega_{b,k} = \omega_k, \sigma_{b,k} = 0.1\sigma, d_\gamma = 0.01\bar{d}, \sigma_\gamma = 0.1\sigma\}$ for all wavenumbers k . As opposed to the filtering problem (see Section 4), here we propagate not only the observed modes ($k = 1, \dots, 7$), but all modes because we assume no knowledge about the observations for the prediction problem. Obviously, the non-switching modes will have equivalent skill with any of the three prediction models above and MSM. In this section, we also consider predictions with perfect model by utilizing the true γ in addition to the three prediction models discussed above and MSM. Note that we did not show predictions with perfect model in the one-mode study in Sections 5.1 and 5.2 because the error in the initial conditions is nearly zero in the super-ensemble setup and the error growth in the unstable regime U with rate $d = 0.04$ becomes negligible up to lead time $\tau = 2.5$. This, however, is not true in a finite realization setting as we will see shortly.

In Fig. 14, we show the RMS errors (averaged over spatial domain and subsets W, U) as functions of lead time τ for initial conditions from the perfectly specified filter, SPEKF-C, and SPEKF-M, for the correctly specified forcing case. When initial conditions are from the perfectly specified filter, we find that the predictive skill of the combined model supersedes MSM by 17% on average in U up to lead time $\tau_n = 2.5$ and is almost as good as the predictive skill of the perfect model up to lead time $\tau_n = 1.25$. This is almost two times the damping time 0.67 of each Fourier mode and nearly one half of the 3 days physical space decorrelation time. On the other hand, when initial conditions are from SPEKF-C and SPEKF-M, the improvement of predictive skill with combined model over MSM decreases to only 11% in U up to lead time $\tau_n = 2.5$ in the former case and $\tau_n = 1.75$ in the latter case, but both predictions are still close to that of the perfect model up to lead time $\tau_n = 0.75$. When initial conditions are from SPEKF-A or MSM, there is no improvement over MSM using any of the models we proposed. This is because the multiplicative bias correction term is constant, $\gamma(t) = \bar{d}$, for every t as we discussed earlier. This also explains the identical predictive skill of modes higher than 7 as shown in Fig. 15; the initial conditions in these modes are not affected by any observations and they are only solutions of MSM.

We also find the results in Sections 5.1 and 5.2 of the one-mode study translate fully to the SPDE. That is, for initial conditions from the perfectly specified filter and from SPEKF-M, MSM is identical to the additive model and the multiplicative model is identical to the combined model. On the other hand, when initial conditions are from SPEKF-C, no methods are identical but since their differences are only in a few switching modes (see Fig. 15), their spatially averaged RMS errors are hardly distinguishable by sight (see second row in Fig. 14). From both the spatial and the mode-by-mode RMS errors, shown in Figs. 14 and 15, we conclude that the advantage of the combined and multiplicative models over MSM and additive model in the switching modes (3–5) translates to the spatially extended switching PDE.

In Fig. 15, the RMS errors of the combined and multiplicative models are slightly larger than those of the MSM and additive model in the strongly damped modes 1–2 and stable modes 6–7 when the system is in the unstable regime, U . In this

Table 5
Average forecast improvement with the combined model over MSM for imperfect initial condition and unspecified forcing.

Subset	SPEKF-A (Impr) _{0<t<τ_n} (%)	τ _n	SPEKF-M (Impr) _{0<t<τ_n} (%)	τ _n	NEKC-C (Impr) _{0<t<τ_n} (%)	τ _n
W	11	0.5	3	0.75	3	0.75
U	9	0.75	12	0.75	12	0.75

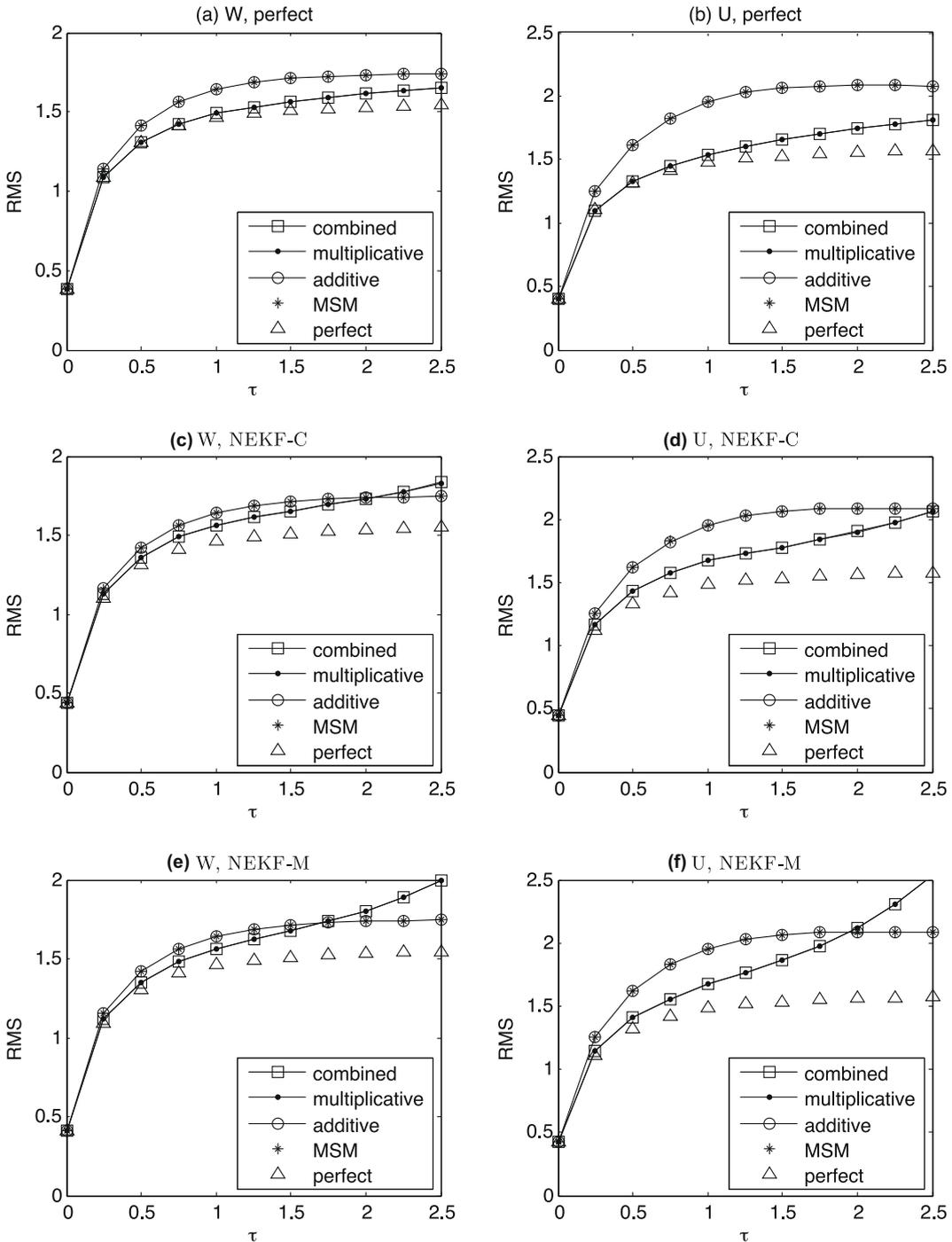


Fig. 14. Predictive skill of the switching PDE: average RMS errors over W and U as functions of lead time for correctly specified forcing with initial conditions from perfectly specified filter (first row), SPEKF-C (second row) and SPEKF-M (third row).

case, the multiplicative bias correction term, γ , tries to readjust the damping strength based on errors in initial conditions. On the other hand, in both the MSM and additive model, the damping is static and is set to the equilibrium damping, $\bar{d} = 1.5$, which is exactly the correct damping for modes 6–7 and close to the damping $d_s = 1.6$ in modes 1–2.

In the stable regime S , the predictive improvement is insignificant ($< 1\%$). We do not show any results with unspecified forcing because they were not very encouraging even in the one-mode study conducted earlier (see Table 5). The high filtering skill with incorrectly specified forcing, studied in Section 4.2, does not translate to predictive skill because the model errors become too large when the lead time τ is much longer than the observation time $\Delta t = 0.25$.

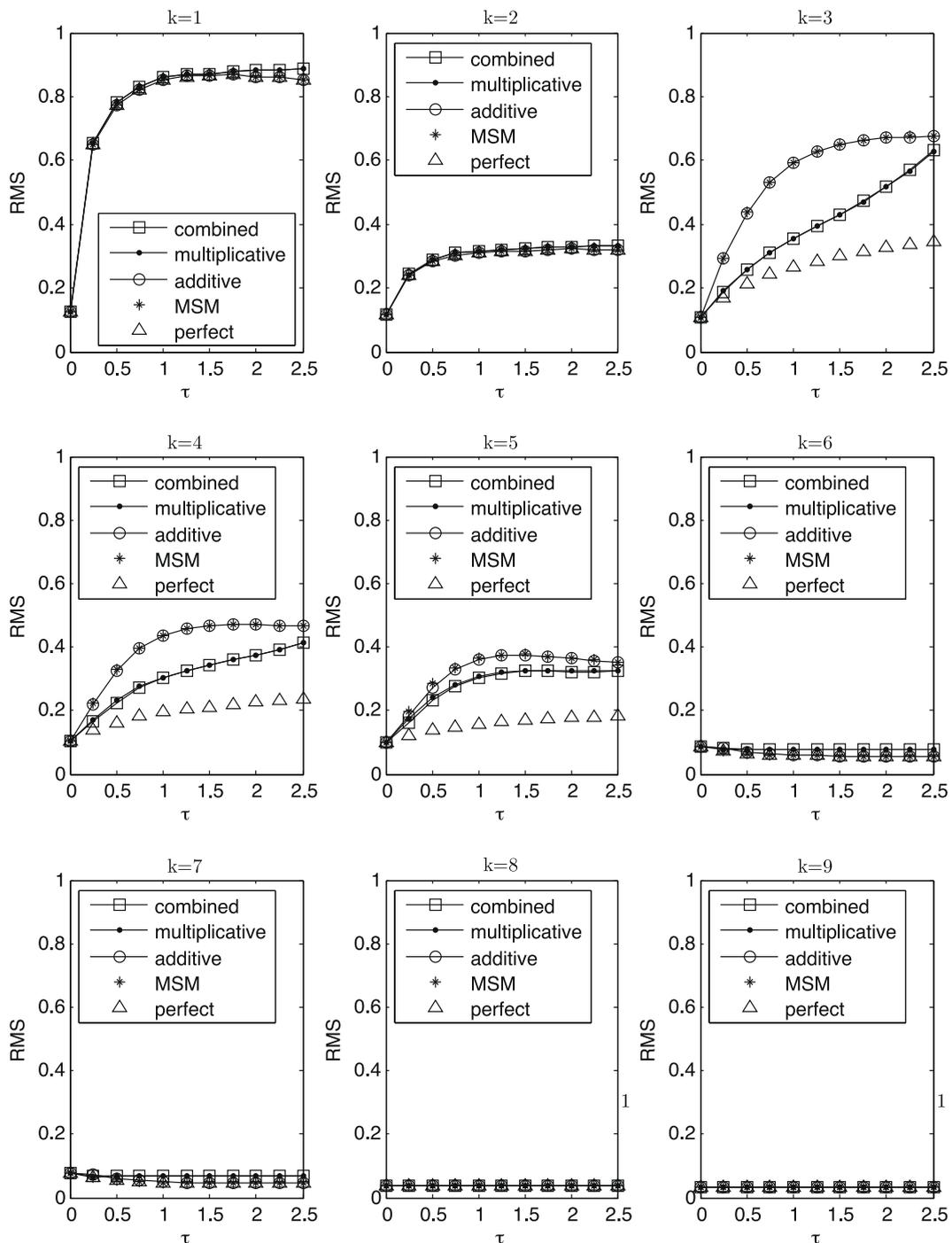


Fig. 15. Predictive skill of the switching PDE: average RMS errors over U as functions of lead time for wavenumbers $k = 1, \dots, 9$ for correctly specified forcing with initial conditions from SPEKF-C.

5.4. Summary

We conclude this section by recommending the combined model as the method of choice for better predictive skill provided that forcing is correctly specified, especially in the unstable regime, since it is robust toward variations of parameters and imperfect initial conditions. Most importantly, its predictive skill is as good as the predictive skill with the perfect model at least up to the damping time, 0.67, regardless of any errors in initial conditions. This high predictive skill is indeed a sig-

nificant improvement over the MSM up to 2.5 days as shown in Fig. 10, which is almost four times the damping time and is very close to the 3 days decorrelation time of the SPDE. We also find the predictive skill with the combined model is not very sensitive toward finite ensemble size; our simulation suggests that an ensemble of order 10 is good enough.

On the other hand, we find that when the forcing is unspecified, the model errors become too large as the prediction lead time grows beyond the observation time. In the stable regime, there is almost no improvement in predictive skill with the combined model relative to the MSM. When initial conditions are imperfect, both the combined and multiplicative models do not have any ability to detect the transitions, this is apparent from the failure observed in subsets A_s and A_u (right after transitions) in Fig. 12.

6. Concluding discussion

In this paper, we introduce a systematic strategy to filter and predict partially observed spatially extended turbulent systems with model errors. In particular, we blend the filter reduction strategies for sparse observations, introduced in [19], and the stochastic parametrization strategy, introduced in [14]. In our numerical experiments, we test our proposed strategy on a toy model for turbulent Rossby waves in a one-dimensional periodic domain that allows baroclinic instability to occur one third of the times at random time despite a stable long time average statistics. In this switching SPDE, the model errors are introduced through the randomly occurring instabilities and/or the unspecified external forcing.

In the filtering case with or without correct forcing, the high filtering skill with the combined model is found not only by quantifying the RMS errors but also by matching the energy spectra of the filtered and true signals. The posterior estimate of the true energy spectrum with the combined model is very accurate while the posterior with the mean stochastic model (MSM) has large errors for both correctly and incorrectly specified forcing (see Figs. 3 and 6). When forcing is correctly specified, both filtering and predictive skill of the multiplicative model are identical to those of the combined model since all errors are from the multiplicative noise term. In terms of prediction, we find that the combined model is as good as the perfect model up to lead times of at least one damping time, 0.67. This high predictive skill is indeed an improvement over MSM of at least 10% for up to 2.5 days with larger improvement at intermediate times (see Figs. 14 and 15). In the one-mode study, this 10% improvement is boosted to nearly 20–60% (depending on which initial conditions are used, see Table 4) for almost four times the damping time scale 0.67. On the other hand, when the forcing is unspecified, the high filtering skill with the combined model does not persist for prediction since the model errors become too large when the forecast lead time is much larger than the observation times considered for filtering.

Numerically, we studied the robustness of the proposed filtering and prediction strategies towards variations of parameters in the stochastic parameter estimation scheme. We have found that the same set of stochastic parameters can be used for all filtered or predicted modes even though they have different energies and dissipations. We showed numerically that the filtering and predictive skill do not change much when the stochastic parameters are varied in significant broad ranges. We also showed robustness of both the filtering skill toward variations of observation noise variances and the prediction skill toward variations of initial conditions.

Finally, we conclude this paper by advertising our ongoing work on applying this encouraging strategy to a more realistic model for baroclinic instability due to vertical wind shear in a three-dimensional domain [34] and to a three cloud model that represents tropical atmospheric dynamics in a single latitude belt [24].

Acknowledgments

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 grants N00014-07-1-0750 and N00014-08-1-1080. Boris Gershgorin and John Harlim are supported as postdoctoral fellows through the last two agencies.

Appendix A. Time correlation functions

In this appendix, we compute the statistical steady state time correlation function of the k th Fourier mode, $\hat{u}_k(t)$, of the solution of SDE (6) which is given by

$$\hat{u}_k(t) = \hat{u}_k(0)e^{p(ik)t} + \int_0^t \hat{f}_k(s)e^{p(ik)(t-s)} ds + \sigma_k \int_0^t e^{p(ik)(t-s)} dW_k(t), \quad (38)$$

where with $p(ik) = \bar{p}(ik) - \gamma(ik)$ and white noise $dW_k = 2^{-1/2}(dW_{k,1} + idW_{k,2})$ and each $W_{k,j}$ is a Wiener process. Note that we assume zero initial covariance for simplicity. The time correlation function is defined as follows:

$$\begin{aligned}
R_k(t, t') &= \langle (\hat{u}_k(t) - \bar{u}_k)(\hat{u}_k(t') - \bar{u}_k)^* \rangle = \left\langle \sigma_k \int_0^t e^{p(\text{ik})(t-s)} dW_k(s) \left(\sigma_k \int_0^{t'} e^{p(\text{ik})(t'-s')} dW_k(s') \right)^* \right\rangle \\
&= \sigma_k^2 e^{p(\text{ik})(t+p^*(\text{ik})t')} \int_0^t \int_0^{t'} e^{-(p(\text{ik})s+p^*(\text{ik})s')} \frac{1}{2} \sum_{i=1}^2 \langle dW_{k,i}(s) dW_{k,i}(s') \rangle \\
&= \sigma_k^2 e^{p(\text{ik})(t+p^*(\text{ik})t')} \int_0^t \int_0^{t'} e^{-(p(\text{ik})s+p^*(\text{ik})s')} \delta(s-s') ds ds' = \sigma_k^2 e^{p(\text{ik})(t+p^*(\text{ik})t')} \int_0^{t'} e^{-(p(\text{ik})+p^*(\text{ik}))s'} ds' \\
&= \frac{\sigma_k^2}{2\gamma(\text{ik})} e^{p(\text{ik})(t+p^*(\text{ik})t')} (e^{2\gamma(\text{ik})t} - 1) = \frac{\sigma_k^2}{2\gamma(\text{ik})} e^{\gamma(\text{ik})(t-t') - \bar{p}(\text{ik})(t'-t)} (1 - e^{-2\gamma(\text{ik})t}).
\end{aligned} \tag{39}$$

In deriving this, we assume $\sigma_k \in \mathbb{R}$ and we use the standard properties of white noise and the integral of the delta function [22,13]. Let $t' = t + \tau$ and let $t \rightarrow \infty$, and substitute $E_k = \sigma_k^2/2\gamma(\text{ik})$ as in Eq. (9); then we obtain the asymptotic temporal correlation function

$$R_k(\tau) = E_k e^{-\gamma(\text{ik})\tau} e^{-\bar{p}(\text{ik})\tau}. \tag{40}$$

Appendix B. Kalman filter formula

The classical Kalman filter [23,22] is an optimal two-step predictor and corrector method that includes observations at every discrete time $t_{m+1} = t_m + \Delta t$ to adjust the prediction when the filter and observation models are linear and Gaussian. In discrete time setting, the Kalman filter solves the following linear problem

$$u_{m+1} = F_{m+1}u_m + f_{m+1} + \sigma_{m+1}, \tag{41}$$

$$v_{m+1} = Gu_{m+1} + \sigma_{m+1}^o, \tag{42}$$

where in Eq. (41), the true signal $u_m \in \mathbb{C}^N$ is the quantity of interest at time t_m , $F_m \in \mathbb{C}^{N \times N}$ is the discrete linear deterministic operator that maps $u(t)$ forward in time, $f_m \in \mathbb{C}^N$ is an external forcing at time t_m , the N -dimensional complex valued noise $\sigma_m = (\sigma_{m,1} + i\sigma_{m,2})/\sqrt{2}$ is defined with Gaussian white noise components $\{\sigma_{m,i} \in \mathbb{R}^N, i = 1, 2\}$ with mean zero and variance $r_m \in \mathbb{C}^{N \times N}$. Observation $v_m \in \mathbb{C}^M$ is modeled as a transformation of the true signal u_m via linear operator $G \in \mathbb{C}^{M \times N}$ plus a complex Gaussian noise $\sigma_m^o = (\sigma_{m,1}^o + i\sigma_{m,2}^o)/\sqrt{2} \in \mathbb{C}^M$, where each component $\{\sigma_{m,i}^o \in \mathbb{R}^M, i = 1, 2\}$ has zero mean and variance $r^o \in \mathbb{C}^{M \times M}$.

The basic Kalman filter solution to (41) and (42) produces an estimate of the mean and covariance of u_{m+1} prior and posterior to knowing observation v_{m+1} [23,22,1,7]. The prior mean state and covariance are denoted by $\bar{u}_{m+1|m}$ and $r_{m+1|m}$, consecutively, while the posterior mean state and covariance are denoted by $\bar{u}_{m+1|m+1}$ and $r_{m+1|m+1}$, consecutively. These statistics are dynamically updated as follows:

Prior update:

$$\bar{u}_{m+1|m} = \tilde{F}_{m+1}\bar{u}_{m|m} + \tilde{f}_{m+1}, \tag{43}$$

$$r_{m+1|m} = \tilde{F}_{m+1}r_{m|m}\tilde{F}_{m+1}^* + \tilde{r}_{m+1}, \tag{44}$$

Posterior update:

$$\bar{u}_{m+1|m+1} = \bar{u}_{m+1|m} + K_{m+1}(v_{m+1} - G\bar{u}_{m+1|m}), \tag{45}$$

$$r_{m+1|m+1} = (\mathcal{I} - K_{m+1}G)r_{m+1|m}, \tag{46}$$

$$K_{m+1} = r_{m+1|m}G^*(Gr_{m+1|m}G^* + r^o)^{-1}, \tag{47}$$

where the asterisk^{*} in Eqs. (44) and (47) denotes the complex adjoint. The posterior mean update, $\bar{u}_{m+1|m+1}$, in (45) is simply a linear combination between the prior mean state, $\bar{u}_{m+1|m}$, and observation, v_{m+1} , weighted in accordance to the Kalman gain matrix $K_{m+1} \in \mathbb{C}^{N \times M}$. In a perfect model context, $\tilde{F}_m = F_m$, $\tilde{f}_m = f_m$ and $\tilde{r}_m = r_m$.

References

- [1] B. Anderson, J. Moore, Optimal Filtering, Prentice-Hall, Englewood Cliffs, NJ, 1979.
- [2] J. Anderson, An adaptive covariance inflation error correction algorithm for ensemble filters, Tellus A 59 (2007) 210–224.
- [3] S.-J. Baek, B. Hunt, E. Kalnay, E. Ott, I. Szunyogh, Local ensemble Kalman filtering in the presence of model bias, Tellus A 58 (3) (2006) 293–306.
- [4] J. Carton, G. Chepurin, X. Cao, A simple ocean data assimilation analysis of the global upper ocean 1950–1995. Part I: Methodology, J. Phys. Oceanogr. 30 (2000) 294–309.
- [5] E. Castronovo, J. Harlim, A. Majda, Mathematical criteria for filtering complex systems: plentiful observations, J. Comput. Phys. 227 (7) (2008) 3678–3714.
- [6] J. Ching, J. Beck, K. Porter, Bayesian state and parameter estimation of uncertain dynamical systems, Prob. Eng. Mech. 21 (1) (2006) 81–96.
- [7] C. Chui, G. Chen, Kalman Filtering, Springer, New York, 1999.
- [8] D. Dee, A.D. Silva, Data assimilation in the presence of forecast bias, Quart. J. R. Meteorol. Soc. 124 (1998) 269–295.
- [9] D. Dee, R. Todling, Data assimilation in the presence of forecast bias: the GEOS moisture analysis, Mon. Weather Rev. 128 (9) (2000) 3268–3282.
- [10] T. Delsole, Stochastic model of quasigeostrophic turbulence, Surv. Geophys. 25 (2) (2004) 107–149.

- [11] B. Friedland, Treatment of bias in recursive filtering, *IEEE Trans. Automat. Contr.* AC-14 (1969) 359–367.
- [12] B. Friedland, Estimating sudden changes of biases in linear dynamical systems, *IEEE Trans. Automat. Contr.* AC-27 (1982) 237–240.
- [13] C. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences*, Springer-Verlag, New York, 1997.
- [14] B. Gershgorin, J. Harlim, A. Majda, Test models for improving filtering with model errors through stochastic parameter estimation, *J. Comput. Phys.* in press, doi:10.1016/j.jcp.2009.08.019.
- [15] B. Gershgorin, A. Majda, A nonlinear test model for filtering slow–fast systems, *Commun. Math. Sci.* 6 (3) (2008) 611–649.
- [16] B. Gershgorin, A. Majda, Filtering a nonlinear slow–fast system with strong fast forcing, *Commun. Math. Sci.* 7 (1) (2009).
- [17] R. Ghanem, M. Shinozuka, Structural system identification. I: Theory, *J. Eng. Mech., ASCE* 121 (2) (1995) 255–264.
- [18] J. Harlim, A. Majda, Filtering nonlinear dynamical systems with linear stochastic models, *Nonlinearity* 21 (6) (2008) 1281–1306.
- [19] J. Harlim, A. Majda, Mathematical strategies for filtering complex systems: regularly spaced sparse observations, *J. Comput. Phys.* 227 (10) (2008) 5304–5341.
- [20] J. Harlim, A. Majda, Catastrophic filter divergence in filtering nonlinear dissipative systems, *Commun. Math. Sci.* 7 (1) (2009).
- [21] J.R. Holton, *An Introduction to Dynamic Meteorology*, third ed., Academic Press, 1992.
- [22] A. Jazwinski, *Stochastic Processes and Filtering Theory*, Academic, San Diego, CA, 1970.
- [23] R. Kalman, R. Bucy, A new results in linear prediction and filtering theory, *Trans. AMSE J. Basic Eng.* 83D (1961) 95–108.
- [24] B. Khouider, A. Majda, A simple multcloud parametrization for convectively coupled tropical waves. Part II: Nonlinear simulations, *J. Atmos. Sci.* 64 (2007) 381–400.
- [25] E. Lorenz, Predictability – a problem partly solved, in: *Proceedings on Predictability*, held at ECMWF on September 4–8, 1995, 1996.
- [26] A. Majda, Real world turbulence and modern applied mathematics, in: V. Arnold, M. Atiyah, P. Lax, B. Mazur (Eds.), *Mathematics: Frontiers and Perspectives*, American Mathematical Society, 2000.
- [27] A. Majda, *Introduction to PDEs and Waves for the Atmosphere and Ocean*, Courant Lecture Notes in Mathematics, vol. 9, American Mathematical Society, 2003.
- [28] A. Majda, R. Abramov, M. Grote, *Information theory and stochastics for multiscale nonlinear systems*, CRM Monograph Series v. 25, American Mathematical Society, Providence, RI, USA, 2005.
- [29] A. Majda, M. Grote, Explicit off-line criteria for stable accurate time filtering of strongly unstable spatially extended systems, *Proc. Natl. Acad. Sci.* 104 (2007) 1124–1129.
- [30] A. Majda, X. Wang, *Nonlinear Dynamics and Statistical Theories for Basic Geophysical Flows*, Cambridge University Press, UK, 2006.
- [31] J. Pedlosky, *Geophysical Fluid Dynamics*, Springer, New York, 1979.
- [32] E. Pitcher, Application of stochastic dynamic prediction to real data, *J. Atmos. Sci.* 34 (1) (1977) 3–21.
- [33] R. Richtmeyer, K. Morton, *Difference Methods for Initial Value Problems*, Wiley, 1967.
- [34] S. Smith, G. Vallis, The scales and equilibration of midocean eddies: freely evolving flow, *J. Phys. Oceanogr.* 31 (2001) 554–571.
- [35] J. Tribbia, D. Baumhefner, The reliability of improvements in deterministic short-range forecasts in the presence of initial state and modeling deficiencies, *Mon. Weather Rev.* 116 (1988) 2276–2288.