A NONLINEAR TEST MODEL FOR FILTERING SLOW-FAST SYSTEMS*

BORIS GERSHGORIN† AND ANDREW MAJDA‡

Abstract. A nonlinear test model for filtering turbulent signals from partial observations of nonlinear slow-fast systems with multiple time scales is developed here. This model is a nonlinear stochastic real triad model with one slow mode, two fast modes, and catalytic nonlinear interaction of the fast modes depending on the slow mode. Despite the nonlinear and non-Gaussian features of the model, exact solution formulas are developed here for the mean and covariance. These formulas are utilized to develop a suite of statistically exact extended Kalman filters for the slow-fast system. Important practical issues such as filter performance with partial observations, which mix the slow and fast modes, model errors through linear filters for the fast modes, and the role of observation frequency and observational noise strength are assessed in unambiguous fashion in the test model by utilizing these exact nonlinear statistics.

Key words. Nonlinear model, slow-fast system, extended Kalman filter

AMS subject classifications. 60H10, 60G35

1. Introduction

Many contemporary problems in science and engineering involve large dimensional turbulent nonlinear systems with multiple time scales, i.e., slow-fast systems. The increasing need for real time predictions, for example, in extended range forecasting of weather and climate, drives the development of improved strategies for data assimilation or filtering [7, 13, 6, 10, 2, 3, 28, 4, 14]. Such filtering algorithms are based on generalizations of the classical Kalman filter [1, 5]. Filtering combines partially observed features of the chaotic turbulent multiscale signal together with a dynamic model to obtain a statistical estimate for the state of the system. The dynamic models for the coupled atmosphere-ocean system are prototype examples of slow-fast systems where the slow modes are advective vortical modes and the fast modes are inertia-gravity waves [29, 9, 21]. Depending on the spatio-temporal scale, one might need only a statistical estimate of the slow modes, as on synoptic scales in the atmosphere [7] or both slow and fast modes such as for squall lines on mesoscales due to the impact of moist convection [17]. In either situation, the noisy partial observations of quantities such as temperature, pressure, and velocity necessarily mix both the slow and fast modes [7, 6, 21]. Furthermore, the dynamical models often suffer from significant model errors due to lack of resolution or inadequate parametrization of physical processes such as clouds, moisture, boundary layers, and topography.

The goal of the present paper is to develop a simple three dimensional nonlinear test model with exactly solvable statistics for slow-fast systems in order to provide unambiguous guidelines for the difficult issues for filtering slow-fast systems from partial observations, as mentioned in the first paragraph. Here, we also study filter performance and model errors in this idealized low-dimensional setting for slow-fast filtering; this study parallels earlier works [11, 26] using low dimensional nonlinear models of Lorenz [19, 20] or their variants [27] without exact statistical solutions to

---

study nonlinear filter performance. However, here we have the additional advantage that with the exactly solvable statistics for the mean and covariance in the nonlinear non-Gaussian slow-fast test model, we can utilize a perfect extended Kalman filter algorithm [1] without additional Monte Carlo errors due to finite ensemble size.

1.1. The nonlinear test model for slow-fast systems.   The nonlinear test model, which we propose here as a prototype for filtering slow-fast systems, is three dimensional with stochastic equations for the real slow mode, u_1, and the complex fast mode, u_2, given by the stochastic triad model,

$$du_1 = (-\gamma_1 u_1 + f_1(t))dt + \sigma_1 dW_1, \tag{1.1}$$

$$du_2 = \left((-\gamma_2 + i\omega_0/\varepsilon + ia_0 u_1)u_2 + f_2(t)\right)dt + \sigma_2 dW_2. \tag{1.2}$$

In equations (1.1) and (1.2), ε is a small parameter that measures the deterministic time scale ratio between the evolution of the fast and slow waves; the damping coefficients, γ_1 and γ_2, and white noise forcings represent the turbulent interaction and energy transfer through nonlinear interaction with other unresolved modes in the triad model [29, 24, 25]; the parameter a_0 measures the nonlinearity in the system, in which energy is always conserved; the terms $f_1(t)$, $f_2(t)$ represent deterministic forcing of the slow and fast modes, respectively. The nonlinear test model in equations (1.1) and (1.2) is motivated directly by both the mathematical theory of slow-fast geophysical flows [9, 21] and high resolution turbulent simulations in slow-fast geophysical regimes [30, 31, 32]; all this work suggests that the slow vortical mode dynamics, u_1 here, and the nonlinear catalytic interactions of the slow modes with the fast modes, u_2 here, are the most important central nonlinear features of slow-fast interaction for geophysical systems. The model in equations (1.1) and (1.2) is the simplest one which retains these nonlinear features.

In the present paper, we set $f_2(t) \equiv 0$ for simplicity, thus we ignore explicit forcing for the fast waves. In section 2, we develop exactly solvable solution formulas for the nonlinear, non-Gaussian statistical solutions for equations (1.1) and (1.2) by generalizing and extending the exact solutions for the Kubo oscillator from statistical physics [18, 23] to non-stationary and cross-correlated dynamics. In section 3, we set up idealized extended Kalman filter algorithms, which utilize the exact solution formulas developed in section 2. In section 4, we study filter performance and model error for the test model in a framework motivated by all the practical issues for filtering slow-fast systems mentioned in the first paragraph. The paper ends with a brief concluding discussion.

2. Exact solutions and exactly solvable statistics in the nonlinear test model

2.1. Model.   As presented in the introduction, consider a system of one slow wave represented by a real-valued function u_1 and two fast waves represented by a complex-valued function u_2. We model this system via a system of coupled stochastic differential equations of the form

$$du_1 = (-\gamma_1 u_1 + f_1(t))dt + \sigma_1 dW_1, \tag{2.1}$$

$$du_2 = (-\gamma_2 + i\omega_0/\varepsilon + ia_0 u_1)u_2 dt + \sigma_2 dW_2, \tag{2.2}$$

where γ_1 and γ_2 are the damping coefficients of the slow and fast waves, respectively, $f_1(t)$ is forcing of the slow wave, σ_1 and σ_2 represent the strength of the noise of

the slow and fast waves, respectively, ε is a small parameter that characterizes the time scale difference between the slow and the fast modes, ω_0 is typical frequency of the fast wave in the units of ε, and a_0 is a nonlinearity parameter. We choose the oscillatory forcing $f_1(t) = A\sin(\omega t)$. The system (1.1)–(1.2) is to be solved with the initial conditions

$$u_1(t_0) = u_{10},$$
$$u_2(t_0) = u_{20},$$

where u_{10} and u_{20} are correlated Gaussian random variables with known parameters: $\langle u_{10} \rangle$, $\langle u_{20} \rangle$, $\mathrm{Var}(u_{10})$, $\mathrm{Var}(u_{20})$, $\mathrm{Cov}(u_{20}, u_{20}^*)$ and $\mathrm{Cov}(u_{20}, u_{10})$. As usual, $\langle \cdot \rangle$ is expectation, $\mathrm{Var}(\cdot)$ is variance, and $\mathrm{Cov}(\cdot, \cdot)$ is covariance.

2.2. Path-wise solution of the model. It is not difficult to develop path-wise solutions of the model equations. Such path-wise solutions provide the signals, which we attempt to filter. First we solve equation (1.1) since it is independent of the fast wave u_2. The slow wave u_1 is easily found using an integrating factor

$$u_1(t) = u_{10}e^{-\gamma_1(t-t_0)} + F_1(t) + \sigma_1 \int_{t_0}^t e^{\gamma_1(s-t)}dW_1(s), \qquad (2.3)$$

where

$$F_1(t) = \int_{t_0}^t f_1(s)e^{-\gamma_1(t-s)}ds$$
$$= \frac{A}{\gamma_1^2 + \omega^2}\left(e^{-\gamma_1(t-t_0)}\left(\omega\cos(\omega t_0) - \gamma_1\sin(\omega t_0)\right) - \omega\cos(\omega t) + \gamma_1\sin(\omega t)\right).$$

Note that $u_1(t)$ is a Gaussian random process, and thus is fully defined by its mean and variance, which will be computed below.

Now, to solve equation (1.2) we treat u_1 as a known function. Using the integrating factor for the equation with time-dependent frequency, we obtain u_2

$$u_2(t) = e^{-\gamma_2(t-t_0)}\psi(t_0,t)u_{20} + \sigma_2 \int_{t_0}^t e^{-\gamma_2(t-s)}\psi(s,t)dW_2(s), \qquad (2.4)$$

where we defined new functions

$$\psi(s,t) = e^{iJ(s,t)},$$
$$J(s,t) = \int_s^t \left(\omega_0/\varepsilon + a_0 u_1(s')\right)ds' = (t-s)\omega_0/\varepsilon + a_0 \int_s^t u_1(s')ds'$$
$$= J_D(s,t) + J_W(s,t) + b(s,t)u_{10}, \qquad (2.5)$$

where the deterministic part of $J(s,t)$ is

$$J_D(s,t) = (t-s)\omega_0/\varepsilon + \frac{Aa_0}{\gamma_1^2 + \omega^2}\left(\frac{\gamma_1}{\omega}\left(\cos(\omega s) - \cos(\omega t)\right) + \left(\sin(\omega s) - \sin(\omega t)\right)\right.$$
$$\left. + \left(e^{-\gamma_1(s-t_0)} - e^{-\gamma_1(t-t_0)}\right)\left(\frac{\omega}{\gamma_1}\cos(\omega t_0) - \sin(\omega t_0)\right)\right),$$

the noisy part of $J(s,t)$ is

$$J_W(s,t)=\sigma_1 a_0 \int_s^t ds' \int_{t_0}^{s'} e^{\gamma_1(s''-s')} dW_1(s''),$$

and the prefactor of u_{10} is

$$b(s,t)=\frac{a_0}{\gamma_1}\left(e^{-\gamma_1(s-t_0)}-e^{-\gamma_1(t-t_0)}\right).$$

Note that $J(s,t)$ linearly depends on u_1 and, therefore, $J(s,t)$ is also a sum of independent Gaussian random fields. We also introduce the following notation for later use

$$\psi_D(s,t)=e^{iJ_D(s,t)},$$
$$\psi_W(s,t)=e^{iJ_W(s,t)}.$$

On the other hand, u_2 is not a Gaussian random variable in the case $a_0 \neq 0$, because the solution formula involves exponentials of Gaussian random variables. Below, we will confirm nongaussianity of u_2 numerically. Nevertheless, we will be able to compute the statistics of u_2 analytically due to the special structure of the solution.

2.3. Invariant measure and choice of parameters.   Here, we obtain the invariant measure of (1.1)–(1.2) without forcing. We separate (1.1)–(1.2) into two parts: the deterministic part,

$$du_1 = 0, \tag{2.6}$$
$$du_2 = i(\omega_0/\varepsilon + a_0 u_1)u_2 dt, \tag{2.7}$$

and the randomly fluctuating and damped part,

$$du_1 = -\gamma_1 u_1 dt + \sigma_1 dW_1, \tag{2.8}$$
$$du_2 = -\gamma_2 u_2 dt + \sigma_2 dW_2. \tag{2.9}$$

Now, we can easily find the invariant measure for both systems. The syste (2.6) and (2.7) has the invariant (Liouville) measure

$$p_{\det}(u_1,u_2)=p_1(u_1)p_2(|u_2|), \tag{2.10}$$

for any probabilistic measures p_1 and p_2. On the other hand, it is well known [8] that the unique invariant measure for the Langevin equation is Gaussian. Therefore, the unique invariant measure for 2.8)–(2.9) is a product of Gaussian measures

$$p_{\text{rand}}(u_1,u_2)=\frac{\sqrt{2\gamma_1\gamma_2}}{\pi\sigma_1\sigma_2}\exp\left(-\frac{\gamma_1 u_1^2}{\sigma_1^2}-\frac{2\gamma_2|u_2|^2}{\sigma_2^2}\right). \tag{2.11}$$

Note that $p_{\text{rand}}(u_1,u_2)$ satisfies equation (2.10) and, therefore, the invariant measure for (1.1)–(1.2) without forcing is given by equation (2.11).

From equation (2.11), we conclude that the average energies for the three modes (which are equal to the corresponding variances) are

$$E_{u_1}=\frac{\sigma_1^2}{2\gamma_1},$$
$$E_{Re[u_2]}=E_{Im[u_2]}=\frac{\sigma_2^2}{4\gamma_2}. \tag{2.12}$$

Now, we can choose parameters of the model in order to control the average energies. For example, for the energy equipartition case, we choose parameters to satisfy

$$E_{u_1} = E_{Re[u_2]} = E_{Im[u_2]}. \tag{2.13}$$

which is the same as

$$\frac{\sigma_1^2}{\gamma_1} = \frac{\sigma_2^2}{2\gamma_2}. \tag{2.14}$$

The decorrelation time of a mode is proportional to the inverse of the damping coefficient. We will consider two different regimes with strong and weak damping.

For the case of strong damping, we choose all the parameters to be of order 1, i.e., $\gamma_1 = 1$, $\sigma_1 = 1$, $\gamma_2 = 1.5$, and $\sigma_2 = 1$. Then, we have $E_{u_1} = 0.5$ and $E_{Re[u_2]} = E_{Im[u_2]} = 0.5625$.

For the case of weak damping, which is the relevant physical regime for slow-fast systems, we choose the decorrelation time of the slow mode to be much longer than the oscillation period of the fast mode. The oscillation period of the slow mode is of the order 1 while the oscillation period of the fast mode is of the order $1/\varepsilon$. We also choose the decorrelation time of the fast mode to be of the same order as the decorrelation time of the slow mode. We take $\varepsilon = 0.1$. Then, the oscillation time of the fast mode is of the order $T_2 = 2\pi/\varepsilon \approx 0.6$. Now we can choose $\gamma_1 = 0.09$ and $\gamma_2 = 0.08$ such that $T_2 \ll 1/\gamma_1$ and $T_2 \ll 1/\gamma_2$. Suppose the average energy of the first mode is $E_{u_1} = 1$. Then, for the energy equipartition case, we have $\sigma_1 = \sqrt{\gamma_1} = 0.3$ and $\sigma_2 = \sqrt{2\gamma_2} = 0.4$. The remaining parameters are chosen to be the same for both strong and weak damping: $\omega_0 = 1$, $a_0 = 1$, $A = 1$, and $\omega = 1$.

2.4. Numerical test of the analytical path-wise solutions.   Here, we confirm that the numerical solution of equations (1.1) and (1.2) converges to the analytical solution given by equations (2.3) and (2.4). We use the standard Euler-Maruyama method for solving equations (1.1) and (1.2) numerically [12]. Since the diffusion is constant here, the Euler-Maruyama method coincides with the Milstein method, which gives first order strong convergence. We denote u_j^n to be a numerical approximation of $u_j(t_n)$ for $j = \{1,2\}$ at time grid point $t_n = hn$, where h is a time step. Then, the numerical scheme becomes

$$u_1^{n+1} = u_1^n + h(-\gamma_1 u_1^n + f_1(t_n)) + \sigma_1 \Delta W_1^n, \tag{2.15}$$

$$u_2^{n+1} = u_2^n + h\left(-\gamma_2 + i\omega_0/\varepsilon + ia_0 u_1^n\right)u_2^n + \sigma_2 \Delta W_2^n, \tag{2.16}$$

where ΔW_1^n are independent real Gaussian random variables with mean 0 and variance h, and ΔW_2^n are independent complex Gaussian random variables with mean 0 and variance $h/\sqrt{2}$ of both real and imaginary parts. Note that even though we have an exact analytical formula for u_2, we still need to evaluate the stochastic integral in equation (2.4). We perform this evaluation numerically and, therefore, the solution u_2 strictly speaking becomes *semi*-analytical. However, we use a very fine time step of $h = 10^{-3}$ (which is the same as for the numerical solution) and, therefore, evaluate the integral in equation (2.4) with high precision. The convergence study of the numerical solution that we present below confirms this approximation.

In Figure 2.1, we demonstrate both analytical solution (equations (2.3), (2.4)) and numerical approximation (equations (2.15), (2.16)). We note that there is an excellent correspondence between them. Moreover, we observe that for the case of
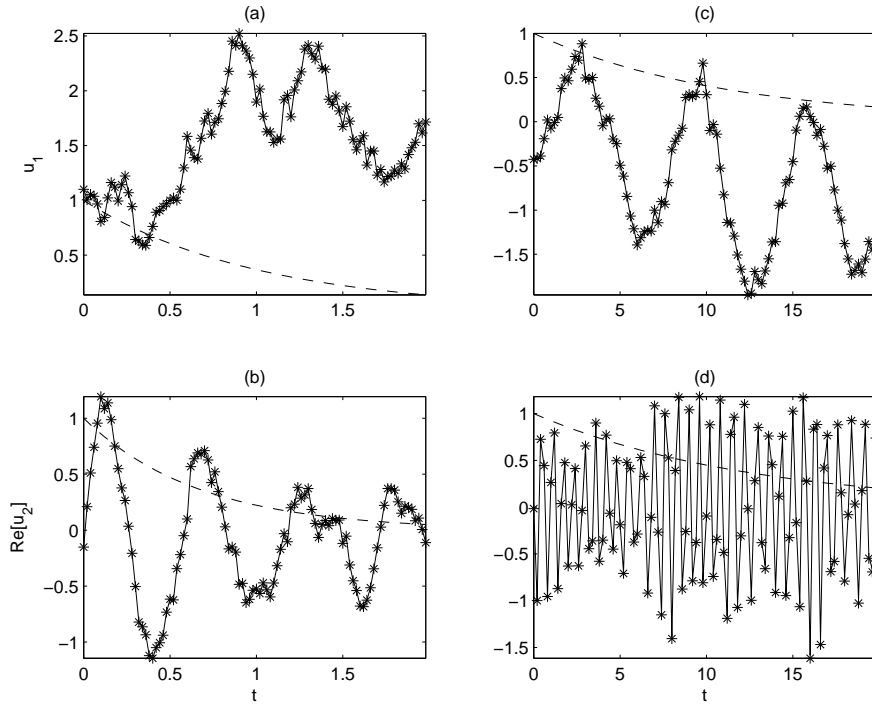
FIG. 2.1. *The analytical solutions (solid line) are computed via equations (2.3)–(2.4) and the numerical solutions (asterisks) are computed via equations (2.15)–(2.16). Panels (a) and (b) show u_1 and $Re[u_2]$ for the case of strong damping. Panels (c) and (d) show u_1 and $Re[u_2]$ for the case of weak damping. The dashed line represents the exponential decay $e^{-\gamma_j t}$ with the corresponding γ_j for $j = \{1,2\}$. Note the different scales of x-axis due to different rates of damping.*

strong damping, u_2 makes only a few oscillations within its typical decorrelation time ($\sim 1/\gamma_2$), whereas for the case of weak damping, u_2 makes many oscillations until it decorrelates. We have confirmed numerically that the Milstein method that we used has the first order convergence with further refinement in h. Next, we compute closed formulas for the mean and covariance of u_1 and u_2, which are needed in an extended Kalman filter for the nonlinear system.

2.5. Mean and variance of u_1.   We start with u_1. By taking the average of equation (2.3), we obtain

$$\langle u_1 \rangle = \langle u_{10} \rangle e^{-\gamma_1(t-t_0)}$$
$$+ \frac{A}{\gamma_1^2 + \omega^2} \left(e^{-\gamma_1(t-t_0)} \left(\omega \cos(\omega t_0) - \gamma_1 \sin(\omega t_0) \right) - \omega \cos(\omega t) + \gamma_1 \sin(\omega t) \right),$$

(2.17)

where we used the fact that the mean of the stochastic integral is always zero [12]. Furthermore, the variance of u_1 becomes

$$\mathrm{Var}(u_1) = \langle (u_1 - \langle u_1 \rangle)^2 \rangle = \mathrm{Var}(u_{10}) e^{-2\gamma_1(t-t_0)} + \frac{\sigma_1^2}{2\gamma_1}(1 - e^{-2\gamma_1(t-t_0)}). \quad (2.18)$$

In order to compute the variance we used the Ito isometry formula [12]

$$\left\langle \left(\int g(t) dW(t) \right)^2 \right\rangle = \int g^2(t) dt,$$

for any deterministic $g(t)$.

2.6. Mean and covariance of u_2.   After averaging equation (2.4), we have

$$\langle u_2 \rangle = e^{-\gamma_2(t-t_0)} \langle \psi(t_0,t) \ u_{20} \rangle. \tag{2.19}$$

For simplicity of notation we drop parameters in the functions $J(s,t)$, $J_D(s,t)$, $J_W(s,t)$, $b(s,t)$, $\psi(s,t)$, $\psi_D(s,t)$, $\psi_W(s,t)$ for $s=t_0$, and denote them as J, J_D, J_W, b, ψ, ψ_D, ψ_W, respectively. Using the assumption that the noise $W_1(t)$ is independent of the initial conditions u_{10} and u_{20}, we obtain

$$\langle u_2 \rangle = e^{-\gamma_2(t-t_0)} \psi_D \langle \psi_W \rangle \langle u_{20} \exp(ibu_{10}) \rangle. \tag{2.20}$$

The averages in the right hand side of equation (2.20) can be computed via the characteristic function of a Gaussian random variable. For any probability distribution, we define a characteristic function as

$$\phi_{\mathbf{v}}(\mathbf{s}) = \langle \exp(i\mathbf{s}^T \mathbf{v}) \rangle, \tag{2.21}$$

where $\mathbf{s} \in \mathbb{C}^d$ and d is the number of dimensions. For a Gaussian distribution, the characteristic function is known [12] to have the following form:

$$\phi_{\mathbf{v}}(\mathbf{s}) = \exp\left(i\mathbf{s}^T \langle \mathbf{v} \rangle - \frac{1}{2}\mathbf{s}^T \Sigma \mathbf{s} \right), \tag{2.22}$$

where Σ is the covariance matrix. In particular, for $\langle \psi_W \rangle$ we obtain

$$\langle \psi_W \rangle = \langle e^{iJ_W} \rangle = \exp\left(i\langle J_W \rangle - \frac{1}{2}\mathrm{Var}(J_W) \right) = \exp\left(-\frac{1}{2}\mathrm{Var}(J_W) \right). \tag{2.23}$$

Computation of $\mathrm{Var}(J_W)$ is straightforward since J_W is Gaussian, and the result is

$$\mathrm{Var}(J_W) = -\frac{\sigma_1^2 a_0^2}{2\gamma_1^3} \left(3 - 4e^{-\gamma_1(t-t_0)} + e^{-2\gamma_1(t-t_0)} - 2\gamma_1(t-t_0) \right). \tag{2.24}$$

Next, we compute $\langle u_{20} \exp(ibu_{10}) \rangle$. Here, it is convenient to use the triad real-valued representation of $(u_1(t), u_2(t))$,

$$x = u_1, \tag{2.25}$$
$$y = Re[u_2], \tag{2.26}$$
$$z = Im[u_2]. \tag{2.27}$$

Then we just need to find $\langle y_0 \exp(ibx_0) \rangle$ and $\langle z_0 \exp(ibx_0) \rangle$, and afterwards combine them using $u_{20} = y_0 + iz_0$ (where the zero subscript refers to the initial values at $t=t_0$) to obtain the second average in the right hand side of equation (2.19). Consider a three-dimensional vector $\mathbf{v} = (x_0, y_0, z_0)$ and the corresponding characteristic function given by equation (2.22). By its definition, the characteristic function is a Fourier transform of the corresponding probability density function (pdf)

$$\phi_{\mathbf{v}}(\mathbf{s}) = \frac{1}{(2\pi)^3} \int \exp(i\mathbf{s}^T \mathbf{v}) g(\mathbf{v}) d\mathbf{v}. \tag{2.28}$$

Next, we use the basic property of Fourier transform, i.e., multiplication by a variable in physical space (e.g., y_0) corresponds to differentiation over the dual variable (e.g., s_2). We have

$$\frac{\partial \phi_{\mathbf{v}}(\mathbf{s})}{\partial s_2} = \frac{1}{(2\pi)^3} \int i y_0 \exp(i\mathbf{s}^T \mathbf{v}) g(\mathbf{v}) d\mathbf{v} = i \langle y_0 \exp(i\mathbf{s}^T \mathbf{v}) \rangle. \tag{2.29}$$

Therefore, we obtain

$$\langle y_0 \exp(ibx_0) \rangle = -i \frac{\partial \phi_{\mathbf{v}}(\mathbf{s})}{\partial s_2} \bigg|_{\mathbf{s}=(b,0,0)^T}. \tag{2.30}$$

Similarly, we find

$$\langle z_0 \exp(ibx_0) \rangle = -i \frac{\partial \phi_{\mathbf{v}}(\mathbf{s})}{\partial s_3} \bigg|_{\mathbf{s}=(b,0,0)^T}.$$

Using the particular form of a Gaussian pdf for $\mathbf{v}$, we find that

$$\frac{\partial \phi_{\mathbf{v}}(\mathbf{s})}{\partial s_2} = \Big(i\langle y_0 \rangle - \mathrm{Var}(y_0)s_2 - \mathrm{Cov}(x_0, y_0)s_1 - \mathrm{Cov}(y_0, z_0)s_3 \Big) \phi_{\mathbf{v}}(\mathbf{s}),$$

$$\frac{\partial \phi_{\mathbf{v}}(\mathbf{s})}{\partial s_3} = \Big(i\langle z_0 \rangle - \mathrm{Var}(z_0)s_3 - \mathrm{Cov}(x_0, z_0)s_1 - \mathrm{Cov}(y_0, z_0)s_2 \Big) \phi_{\mathbf{v}}(\mathbf{s}).$$

After evaluating the partial derivatives at $\mathbf{s} = (b,0,0)^T$, we obtain

$$\langle y_0 \exp(ibx_0) \rangle = \Big(\langle y_0 \rangle + i\mathrm{Cov}(x_0, y_0)b \Big) \exp\Big(ib\langle x_0 \rangle - \frac{1}{2}\mathrm{Var}(x_0)b^2 \Big), \tag{2.31}$$

$$\langle z_0 \exp(ibx_0) \rangle = \Big(\langle z_0 \rangle + i\mathrm{Cov}(x_0, z_0)b \Big) \exp\Big(ib\langle x_0 \rangle - \frac{1}{2}\mathrm{Var}(x_0)b^2 \Big). \tag{2.32}$$

Combining equations (2.31) and (2.32) yields

$$\langle u_{20} \exp(ibu_{10}) \rangle = \Big(\langle u_{20} \rangle + i\mathrm{Cov}(u_{20}, u_{10})b \Big) \exp\Big(ib\langle u_{10} \rangle - \frac{1}{2}b^2 \mathrm{Var}(u_{10}) \Big). \tag{2.33}$$

Therefore, we have found all the components of the right hand side of equation (2.20) using the initial data. In a similar manner, we find the variance of u_2, cross-covariance of u_2 with u_2^* and of u_2 with u_1. We have

$$\mathrm{Var}(u_2) = e^{-2\gamma_2(t-t_0)} \Big(\mathrm{Var}(u_{20}) + |\langle u_{20} \rangle|^2 - |\langle u_{20}\psi \rangle|^2 \Big) + \frac{\sigma_2^2}{2\gamma_2}\Big(1 - e^{-2\gamma_2(t-t_0)} \Big), \tag{2.34}$$

where

$$|\langle u_{20}\psi \rangle|^2 = e^{-\mathrm{Var}(J_W) - b^2 \mathrm{Var}(u_{10})} |\langle u_{20} \rangle + ib\mathrm{Cov}(u_{20}, u_{10})|^2. \tag{2.35}$$

In Appendix A, we obtain the cross-covariances in a way similar to computing equation (2.33):

$$\mathrm{Cov}(u_2, u_1) = \psi_D \langle \psi_W \rangle e^{-(\gamma_1 + \gamma_2)(t-t_0)} \Big(\langle u_{10}u_{20} \exp(ibu_{10}) \rangle - \langle u_{10} \rangle \langle u_{20} \exp(ibu_{10}) \rangle \Big)$$

$$+ \frac{i}{2}\psi_D \langle \psi_W \rangle e^{-\gamma_2(t-t_0)} \frac{1}{a_0} \frac{\partial}{\partial t} \Big(\mathrm{Var}(J_W(t_0, t)) \Big) \langle u_{20} \exp(ibu_{10}) \rangle, \tag{2.36}$$

$$\mathrm{Cov}(u_2, u_2^*) = \exp(-2\gamma_2(t-t_0) + 2iJ_D - 2\sigma_{J_W}^2) \langle u_{20}^2 \exp(i2bu_{10}) \rangle - \langle u_2 \rangle^2, \tag{2.37}$$

where $\langle u_{10}u_{20}\exp(ibu_{10})\rangle$ and $\langle u_{20}^2\exp(i2bu_{10})\rangle$ are given by equations (A.8) and (A.9).

2.7. Testing the analytical formulas via Monte Carlo simulations. It is very instructive to provide a visual comparison of the analytical formulas for the various statistics of u_1 and u_2 that we derived above with the numerically averaged values using Monte Carlo simulations. We used an ensemble of $M = 10^4$ members in Monte Carlo averaging. As an object of study, let us choose the mean of u_2.
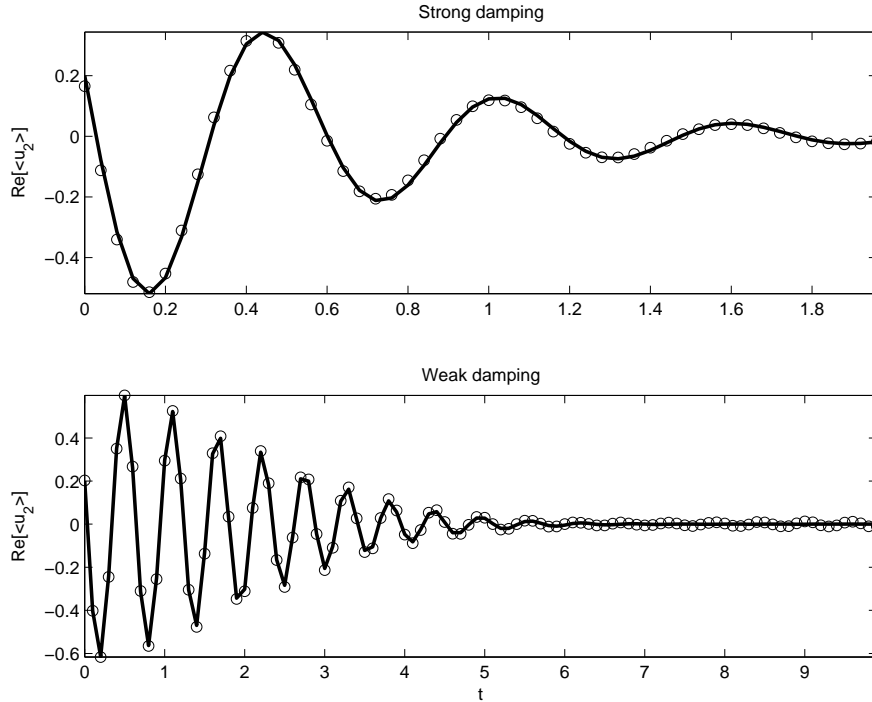

FIG. 2.2. *The solid line corresponds to $\langle u_2 \rangle$, computed via equation (2.20). The circles correspond to Monte Carlo averaging of an ensemble of solutions u_2, each computed via equation (2.4). Note the different scales of x-axis due to different rates of damping.*

In Figure 2.2, the solid line represents $Re[\langle u_2 \rangle]$ that we computed via equation (2.20). The circles, on the other hand, were obtained using Monte Carlo averaging of an ensemble of trajectories that were computed via equation (2.4). The upper and lower panels correspond to the case of strong and weak damping, correspondingly. The plots for $Im[\langle u_2 \rangle]$ is not displayed here because they are similar to the plots of $Re[\langle u_2 \rangle]$. We observe excellent agreement between the analytically obtained mean of u_2 and the result of the averaging using Monte Carlo simulation. Next, in Figure 2.3 we demonstrate time evolution of cross-covariance between u_2 and u_2^*. Here, we also have very good agreement between the analytical prediction and Monte Carlo averaging.

Next, we present the comparison of the analytical formula with the results of Monte Carlo averaging for the correlator $\langle u_2 u_1 \rangle$ shown in Figure 2.4. Note that here we again have excellent agreement.
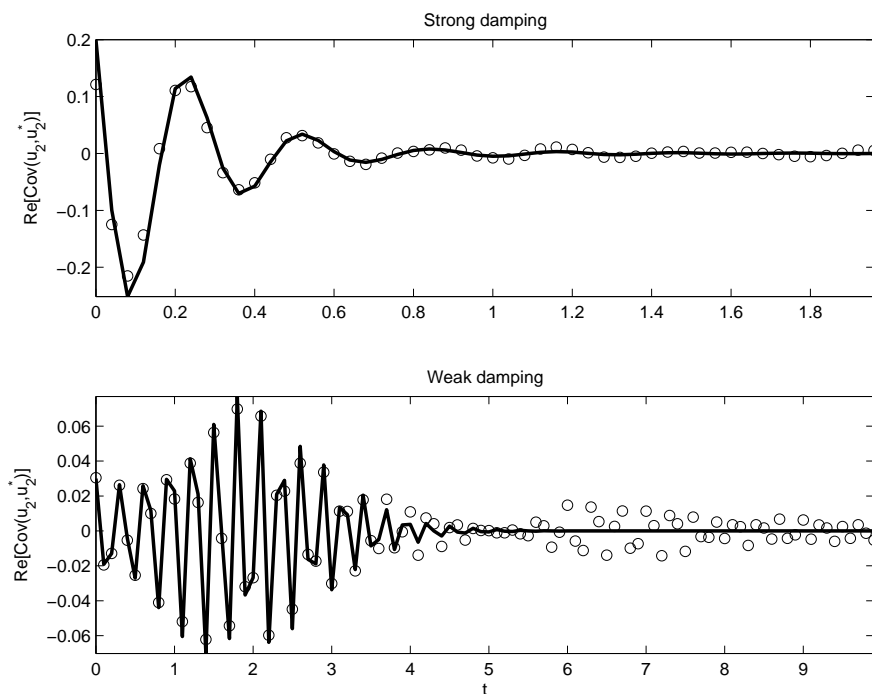
FIG. 2.3. *The solid line corresponds to* $\mathrm{Cov}(u_2,u_2^*)$, *computed via equation (2.37). The circles correspond to Monte Carlo averaging of an ensemble of solutions u_2, each computed via equation (2.4). Note the different scales of x-axis due to different rates of damping.*

2.8. Nongaussianity of u_2.   Here, we present the evidence of nongaussianity of the fast wave u_2 with the periodic forcing $f_1(t) = A\sin(\omega t)$. Due to the nonlinear structure of the governing equation (1.2), it is natural to expect that u_2 is not Gaussian. However, as we will see below, depending on the regime of the system we can observe both strongly nongaussian and almost Gaussian statistics of u_2. Note that, as we discussed earlier, the invariant measure of our system is always Gaussian. Therefore, nongaussianity can only appear in the transient. In order to detect the deviation from Gaussian statistics, we measure skewness and kurtosis excess. The skewness and kurtosis are defined as the third and the fourth normalized central moments, correspondingly. For a random variable ξ, the skewness is

$$\text{skewness} = \frac{\left\langle \left(\xi - \langle\xi\rangle\right)^3 \right\rangle}{\mathrm{Var}(\xi)^{3/2}}, \tag{2.38}$$

and kurtosis is

$$\text{kurtosis} = \frac{\left\langle \left(\xi - \langle\xi\rangle\right)^4 \right\rangle}{\mathrm{Var}(\xi)^2}. \tag{2.39}$$
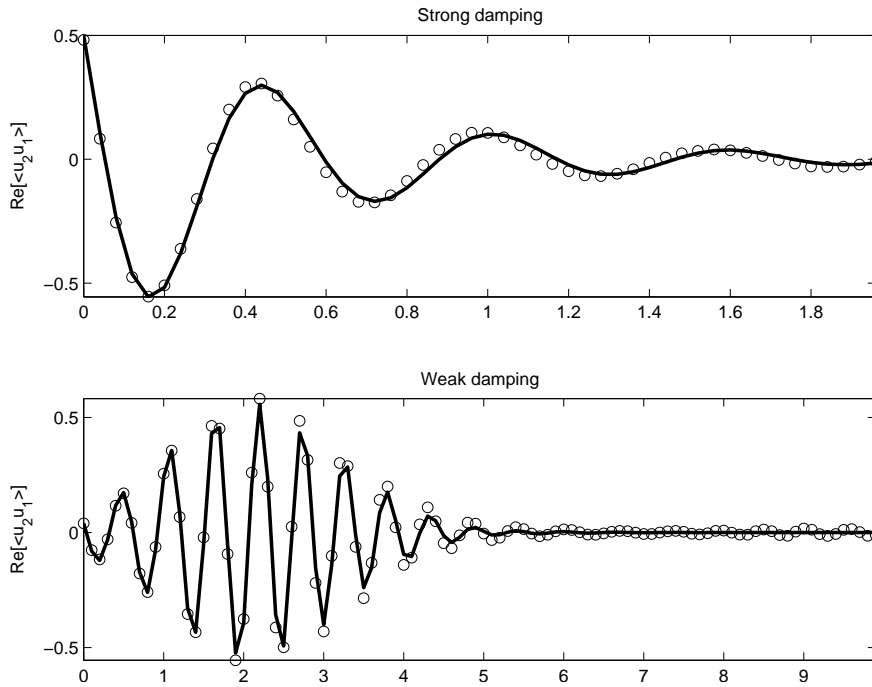
Fig. 2.4.   *The solid line corresponds to* $\langle u_2 u_1 \rangle$, *computed via equations (2.36), (2.17) and (2.20). The circles correspond to Monte Carlo averaging of an ensemble of solutions* u_1, *each computed via equation (2.3), and* u_2, *each computed via equation (2.4). Note the different scales of x-axis due to different rates of damping.*

For Gaussian ξ, we obtain skewness $= 0$ and kurtosis $= 3$. In general, the skewness and the kurtosis excess, which is equal to

$$\text{kurtosis excess} = \text{kurtosis} - 3, \qquad (2.40)$$

measure the nongaussian properties of a given random variable.

In Figures 2.5 and 2.6, we demonstrate the time evolution of the skewness and kurtosis excess, respectively, for both the weakly damped and the strongly damped cases. Note that in the strongly damped case (Figures 2.5 and 2.6 upper panels), both skewness and kurtosis excess have values close the Gaussian ones. In the strongly damped case, the transient time is short and the statistics of the system do not deviate much from their Gaussian values. On the other hand, in the weakly damped regime (Figures 2.5 and 2.6 lower panels) we observe strong nongaussianity in both skewness and kurtosis excess. In section 2.3, we have shown that invariant measure for the system (1.1) and (1.2) with $f_1(t) \equiv 0$ is Gaussian and Figures 2.5 and 2.6 confirm that statistics converge to their Gaussian values after the decorrelation time even for oscillatory $f_1(t)$. We note that a sufficiently large ensemble should be used in Monte Carlo simulations for higher moments in order to obtain a precise result. In our simulation, we used the ensemble of $M = 10^4$ members, which is large enough to get an accurate qualitative picture.

Finally, we end this section with some comments. In principle, analytic expressions for the higher order moments can be computed in a similar fashion as for the

second order statistics; however, the explicit formulas become extremely lengthy. Also, the perceptive reader will note that we could use non-Gaussian initial data in the exact solution for the mean and covariance provided that we know the characteristic function of this random variable explicitly.
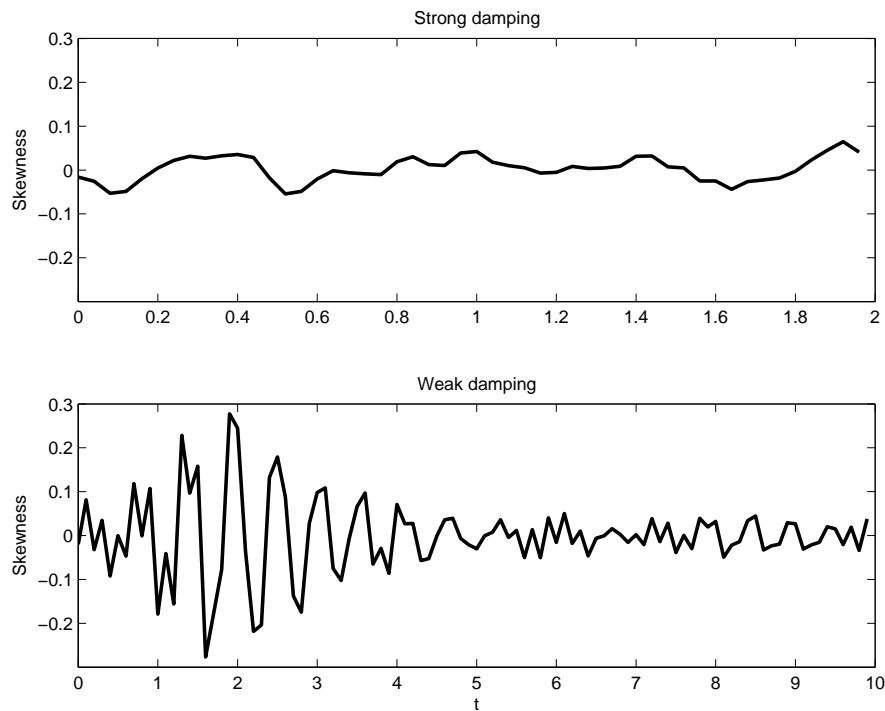

FIG. 2.5. *Skewness evolution for the strongly damped (upper panel) and the weakly damped (lower panel) cases. The weakly damped system is more nongaussian than strongly damped. Note the different scales of x-axis due to different rates of damping.*

3. Filtering algorithms for the test model

3.1. Extended Kalman filter.   Here, we briefly introduce the extended Kalman filter algorithm for the test model (1.1) and (1.2). Suppose that at time $t_m = m\Delta t$, where $m \geq 0$ is a time step index and Δt is the observation time step, the truth signal is denoted as $\mathbf{u}_m$, which is a realization of (u_1, u_2) computed via equations (2.3) and (2.4). However, we assume that $\mathbf{u}_m$ is unknown and instead we are given some linear transformation of $\mathbf{u}_m$ distorted by some Gaussian noise

$$\mathbf{v}_m = G\mathbf{u}_m + \sigma_m^0, \tag{3.1}$$

where $\mathbf{v}_m$ is called the observation, G is a rectangular matrix of the size $q \times 3$ with the number of observations $q = \{1, 2, 3\}$, $\mathbf{u} = (x, y, z)^T \equiv (u_1, Re[u_2], Im[u_2])^T$ and σ_m^0 is the observation noise. The observation noise is assumed to be unbiased (mean-zero) with covariance matrix R^0 of the size $q \times q$. The goal of filtering is to find the filtered signal $\mathbf{u}^f$, which is as close as possible to the original truth signal $\mathbf{u}$. The information that can be used in filtering is limited to
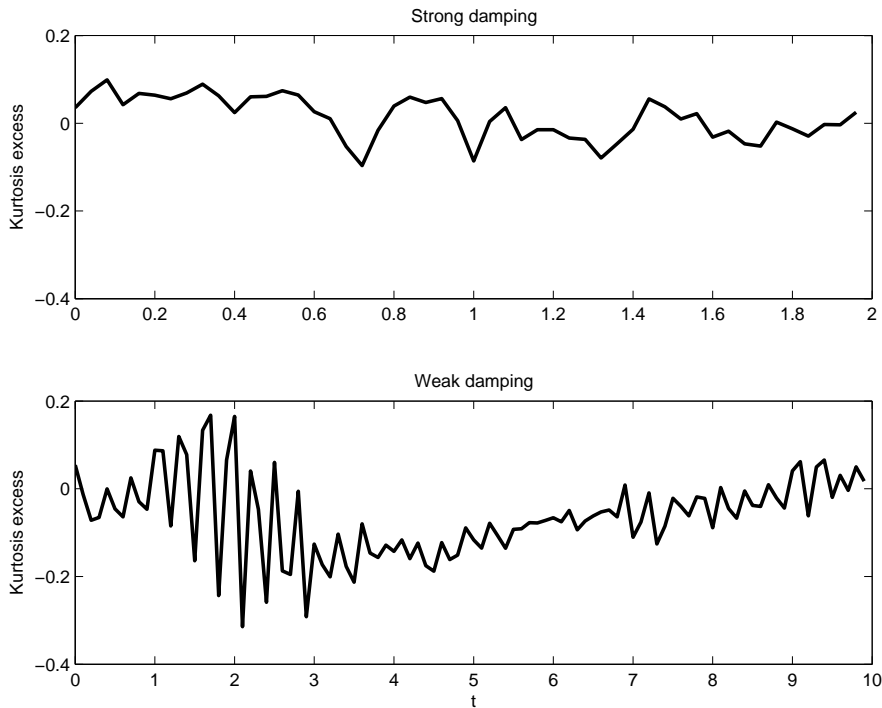
FIG. 2.6. *Kurtosis excess evolution for the strongly damped (upper panel) and the weakly damped (lower panel) cases. The weakly damped system is more nongaussian than strongly damped. Note the different scales of x-axis due to different rates of damping.*

- the model for dynamical evolution of $\mathbf{u}_m$,
- the matrix G,
- and the mean and covariance of the Gaussian noise σ_m^0.

In the case when the evolution of $\mathbf{u}_m$ is described by a linear equation, the best approximation to the truth signal in the least squares sense is given via the Kalman filter algorithm [1, 5]. The Kalman filter consists of two steps: (i) forecasting using the dynamics and (ii) correction using observations. If we assume Gaussian initial conditions, then due to linear dynamics the solution stays linear for all times and can be fully described by the mean and covariance matrix. Denote the mean and covariance of the filtered signal at time t_m as $\langle \mathbf{u} \rangle_{m|m}$ and $\Gamma_{m|m}$, respectively. Then the forecasting step gives us the following, so called prior, values of the mean and covariance at the next time step t_{m+1}

$$\langle \mathbf{u} \rangle_{m|m} \rightarrow \langle \mathbf{u} \rangle_{m+1|m},$$
$$\Gamma_{m|m} \rightarrow \Gamma_{m+1|m}. \tag{3.2}$$

Note that $\langle \mathbf{u} \rangle_{m+1|m}$ and $\Gamma_{m+1|m}$ depend solely on the prior information up to time t_m. In order to utilize the observations $\mathbf{v}_{m+1}$ at time t_{m+1}, the least squares correction

method is used, which yields the posterior values of mean and covariance

$$
\begin{aligned}
\langle \mathbf{u} \rangle_{m+1|m+1} &= \langle \mathbf{u} \rangle_{m+1|m} + K_{m+1}(\langle \mathbf{v} \rangle_{m+1} - G \langle \mathbf{u} \rangle_{m+1|m}), \\
\Gamma_{m+1|m+1} &= (I_3 - K_{m+1}G)\Gamma_{m+1|m}, \\
K_{m+1} &= \Gamma_{m+1|m} G^T (G \Gamma_{m+1|m} G^T + R^0)^{-1},
\end{aligned}
\tag{3.3}
$$

where K_{m+1} is a Kalman gain matrix of the size $3 \times q$ and I_3 is the identity 3×3 matrix. The posterior distribution is the Gaussian distribution with the mean and covariance given in equation (3.3). We note that Kalman gain tells us how much weight the filter puts on the observations vs prior forecast.

In the more general case, when the dynamics is given by nonlinear equations, the procedure described in equations (3.2) and (3.3) is called Extended Kalman Filter (EKF). Due to nonlinearity, the gaussianity of the signal can be lost and this filter may not be optimal anymore. One of the purposes of this paper is to investigate the skill of the EKF using our test model. The advantage of studying the test model (1.1) and (1.2) is in the fact that exact analytical formulas can be used for the mean and covariance (see section 2) in order to make the prior forecast (3.2).

In the EKF, we use exact evolution equations for the mean and, therefore, in the simulation we use the mean value of the observation $\langle \mathbf{v} \rangle_m$ in the first equation of (3.3). The effect of the observation noise size is accounted for in computing the Kalman gain matrix K_{m+1}.

3.2. Structure of observations.   In a typical slow-fast system such as the shallow water equations, observation of pressure, temperature, and velocity automatically mixes the fast and slow components and can corrupt the filtering of the slow component. Below, we introduce prototype observations with these features. We will consider three different types of observations with the corresponding observation matrices G and covariances R^0

- 1 observation

$$
G = \left(1 \ \tfrac{1}{\sqrt{2}} \ \tfrac{1}{\sqrt{2}} \right),
$$

$$
v_1 = x + \frac{1}{\sqrt{2}}(y+z) + \sigma_1^0,
$$

$$
R^0 = 2r^0,
$$

- 2 observations

$$
G = \begin{pmatrix} 1 & \tfrac{1}{\sqrt{2}} & \tfrac{1}{\sqrt{2}} \\ 1 & \tfrac{1}{\sqrt{2}} & -\tfrac{1}{\sqrt{2}} \end{pmatrix},
$$

$$
\begin{cases} v_1 = x + \tfrac{1}{\sqrt{2}}(y+z) + \sigma_1^0, \\ v_2 = x + \tfrac{1}{\sqrt{2}}(y-z) + \sigma_2^0, \end{cases}
$$

$$
R^0 = \begin{pmatrix} 2r^0 & 0 \\ 0 & 2r^0 \end{pmatrix},
$$

- 3 observations

$$G = \begin{pmatrix} 1 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 1 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ 1 & 0 & 0 \end{pmatrix},$$

$$\begin{cases} v_1 = x + \frac{1}{\sqrt{2}}(y + z) + \sigma_1^0, \\ v_2 = x + \frac{1}{\sqrt{2}}(y - z) + \sigma_2^0, \\ v_3 = x + \sigma_3^0, \end{cases}$$

$$R^0 = \begin{pmatrix} 2r^0 & 0 & 0 \\ 0 & 2r^0 & 0 \\ 0 & 0 & r^0 \end{pmatrix}.$$

Here, we assumed that the observation noise components σ_1^0, σ_2^0, and σ_3^0 are independent mean-zero Gaussian with variances $2r^0$, $2r^0$, and r^0, respectively.

3.3. Generating the truth signal. We generate the truth signal using the following procedure. We consider any random initial data as discussed in section 2 and obtain a realization of the trajectory (u_1, u_2) using the exact solutions given by equations (2.3) and (2.4). Note that u_1 can be easily computed since its random part is Gaussian with known statistics. However, u_2 is not Gaussian and its random part depends on the evolution of u_1. Therefore, even if we need the true trajectory only at discrete times with a time step Δt, we still need to compute u_1 with a much finer resolution with time step h. This fine trajectory of u_1 is then used to compute u_2.

3.4. Linear filter with model error. Now, suppose that the prior forecast $\mathbf{u}_{m+1|m}$ is made using the linearized version of the analytical equations that we obtained in section 2. We set $a_0 = 0$ in equation (1.2) and thus we introduce model error. In this case, the linearization is made at the climatological mean state. As we have seen from the invariant measure, in the long run the correlation between the slow wave u_1 and fast wave u_2 vanishes and so does the effect of nonlinear coupling through a_0. On the other hand, for shorter observation time steps Δt, the nongaussianity as an effect of nonlinearity can be sufficiently strong (see section 2.8), and the model error can be rather large. The advantage of using a linear model as an approximation of the true dynamics is for practical application. In real physical problems, the true dynamics of the model are often unknown and ensemble approximations to the Kalman filter are very expensive for a large dimensional system. Thus, the performance of the linear filter in the nonlinear test model for the slow-fast system is interesting for several reasons [15, 16]. Note that the truth signal is always produced via the nonlinear version of equations (1.1) and (1.2) with $a_0 \neq 0$. Therefore, if we use the linear approximation to the original system, we may not obtain the optimal filtered signal due to model error. Below, we will compare the error in filtering the test problem using the perfect model assumption (forecast via first two moments for the nonlinear equation) and the linear dynamics equations with model error. In our test model, linearization only affects the fast mode u_2. Substituting $a_0 = 0$ into equation (2.20) yields the following linear Equ.

$$\langle u_2 \rangle = \exp\left((-\gamma_2 + i\omega_0/\varepsilon)\Delta t \right) \langle u_{20} \rangle. \tag{3.4}$$

Therefore, the forecast is made according to

$$\langle \mathbf{u} \rangle_{m+1|m} = B \langle \mathbf{u} \rangle_{m|m} + C, \tag{3.5}$$

where

$$B = \begin{pmatrix} e^{-\gamma_1 \Delta t} & 0 & 0 \\ 0 & e^{-\gamma_2 \Delta t} \cos(\alpha) & -e^{-\gamma_2 \Delta t} \sin(\alpha) \\ 0 & e^{-\gamma_2 \Delta t} \sin(\alpha) & e^{-\gamma_2 \Delta t} \cos(\alpha) \end{pmatrix}, \tag{3.6}$$

with

$$\alpha = \Delta t \omega_0 / \varepsilon,$$

and

$$C = \begin{pmatrix} F_1(t) \\ 0 \\ 0 \end{pmatrix}. \tag{3.7}$$

equation (3.5) is used as a prior forecast for the mean. Similarly by substituting $a_0 = 0$ into equations (2.34), (2.36), and (2.37) we obtain the prior covariance of the linearized model.

3.4.1. Observability.   Another advantage of using linearized equations is that it then becomes possible to strictly address the issue of observability [1, 5]. Let us study the observability of the system for the case of observation type 1. In this case, the observability matrix is defined by

$$\mathbf{O} = \begin{bmatrix} G \\ GB \\ GB^2 \end{bmatrix}$$

$$= \begin{pmatrix} 1 & 1/\sqrt{2} & 1/\sqrt{2} \\ e^{-\gamma_1 \Delta t} & \frac{e^{-\gamma_2 \Delta t}}{\sqrt{2}} \Big(\cos(\alpha) + \sin(\alpha) \Big) & \frac{e^{-\gamma_2 \Delta t}}{\sqrt{2}} \Big(\cos(\alpha) - \sin(\alpha) \Big) \\ e^{-2\gamma_1 \Delta t} & \frac{e^{-2\gamma_2 \Delta t}}{\sqrt{2}} \Big(\cos(2\alpha) + \sin(2\alpha) \Big) & \frac{e^{-2\gamma_2 \Delta t}}{\sqrt{2}} \Big(\cos(2\alpha) - \sin(2\alpha) \Big) \end{pmatrix}, \tag{3.8}$$

This is a 3×3 matrix and in order for the system (3.5) to be fully observable, matrix $\mathbf{O}$ should have rank 3. It is easy to conclude from equation (3.8) that whenever we have $\sin(\alpha) = 0$ or equivalently

$$\Delta t = 2\pi l \varepsilon / \omega_0, \tag{3.9}$$

for any integer l, the two last columns become equal and, therefore, matrix $\mathbf{O}$ becomes singular. This results in losing observability. Let us find the determinant of $\mathbf{O}$

$$\det(\mathbf{O}) = -e^{-\gamma_2 \Delta t} \sin(\alpha) \Big(\big[e^{-\gamma_2 \Delta t} - e^{-\gamma_1 \Delta t} \cos(\alpha) \big]^2 + e^{-2\gamma_1 \Delta t} \big[1 - \cos^2(\alpha) \big] \Big). \tag{3.10}$$

Therefore, in equation (3.9) we have all the values Δt for which observability is lost. However, practically we still can lack observability even if the matrix $\mathbf{O}$ has rank 3, but is close to being singular, which means that $\det(\mathbf{O})$ is close to zero.

The analysis of observability of the linearized model will also be useful for the nonlinear regime because linear dynamics still plays an important role in the system evolution in the nonlinear regime. In the nonlinear case, we also might expect deteriorating filter performance around the values of Δt described by equation (3.9).

4. Filter performance

4.1. Details of numerical simulation of filtering. We turn to the discussion of the results of the numerical simulation of EKF on our test model (1.1) and (1.2). Here, we describe the parameters that were used in the numerical simulations. The procedure of generating the truth signal was given in section 2. In order to test the filter performance, we compare the truth signal $\mathbf{u}_m$ with the filtered mean $\langle \mathbf{u} \rangle_{m|m}$. We also estimate the role of prior forecasts and observations by studying the prior mean $\langle \mathbf{u} \rangle_{m|m-1}$. In order to measure how well the filter reproduces the truth signal (we refer to this property as filter skill), we use root mean square error (RMSE) and cross correlation (XC) between the truth signal and posterior (and sometimes between truth and prior) signals. RMSE is defined via

$$\text{RMSE}(\mathbf{z} - \mathbf{w}) = \sqrt{\frac{1}{N} \sum_{j=1}^{N} |z_j - w_j|^2}, \tag{4.1}$$

where $\mathbf{z}$ and $\mathbf{w}$ are the complex vectors to be compared and N is the length of each vector. Looking ahead at Figures 4.1-4.4, where we show individual trajectories that are being filtered, the typical amplitude of $\mathbf{u}$ is below 2 in magnitude so the RMSE roughly is twice as big as the normalized percentage error in the study below. For real vectors $\mathbf{x}$ and $\mathbf{y}$, the XC is defined by

$$\text{XC}(\mathbf{x}, \mathbf{y}) = \frac{\sum_{j=1}^{N} x_j y_j}{\sqrt{\sum_{j=1}^{N} x_j^2 \sum_{j=1}^{N} y_j^2}}. \tag{4.2}$$

For the complex valued vectors $\mathbf{z}$ and $\mathbf{w}$, the cross correlation is computed via

$$\text{XC}(\mathbf{z}, \mathbf{w}) = \frac{1}{2} \Big(\text{XC}(Re[\mathbf{z}], Re[\mathbf{w}]) + \text{XC}(Im[\mathbf{z}], Im[\mathbf{w}]) \Big). \tag{4.3}$$

Note that if two signals are close to each other, the RMSE of their difference is then close to zero and their XC is close to one. On the other hand, for two different signals the RMSE diverges from zero and in principle is unbounded and XC approaches zero.

In all the simulations we use $N = 10^4$ data filtering cycles with the given observation time step Δt. The type of observation, observation variance r^0, and observation time step Δt will be varied and specified in each particular situation. Here, we will only discuss the weak damping case since it has more physical importance for filtering slow-fast systems.

4.2. Examples of filtering individual trajectories. Here, we demonstrate how the EKF works on individual trajectories. We chose two values of the observation time step: $\Delta t = 0.13$ and $\Delta t = 1.43$. The first one is considerably smaller than the typical oscillation period $T_2 = 2\pi\omega_0/\varepsilon \approx 0.63$ of the fast mode u_2 and the second one is larger than T_2. We also chose two values of the observation variance $r^0 = 0.1$ and $r^0 = 2.0$. The first value of r^0 is chosen to be smaller than the average energy of each mode $E = 1$ and the second value is chosen to be larger than $E = 1$. Therefore, we have four pairs $(\Delta t, r^0)$ for which we have studied the performance of the EKF. We have used observations of type 1 in this numerical experiment so that slow and fast modes are mixed in the single observation.

The four panels in Figure 4.1 correspond to the four pairs of parameters $(\Delta t, r^0)$. In each panel in Figure 4.1, we demonstrate a segment of the truth trajectory u_1

$r^0 \backslash \Delta t$	0.13	1.43
0.1	0.988 (0.982)	0.936 (0.888)
2.0	0.951 (0.945)	0.855 (0.823)

TABLE 4.1. *EKF performance on u_1: cross correlations XC_p (between truth and posterior) and XC_f (in the parenthesis, between truth and prior). The segments of the corresponding trajectories are shown in Figure 4.1.*

together with the prior forecast and the posterior signal. In Table 4.1, we also show the cross correlation XC_f between the truth signal and prior forecast and cross correlation XC_p between the truth signal and posterior signal. Table 4.1 is made for the same set of parameters as Figure 4.1, and the cross correlations were measured along the trajectories of $n = 10^4$ observation time steps Δt. Note that for all four cases we have $XC_f < XC_p$, which means that the correction step of the EKF improves the forecast using observations. Comparing the columns of Table 4.1, we observe that the skill of the EKF decreases significantly when we increase Δt. On the other hand, comparing the rows of Table 4.1, we also observe that the skill of the EKF decreases significantly when we increase r^0.
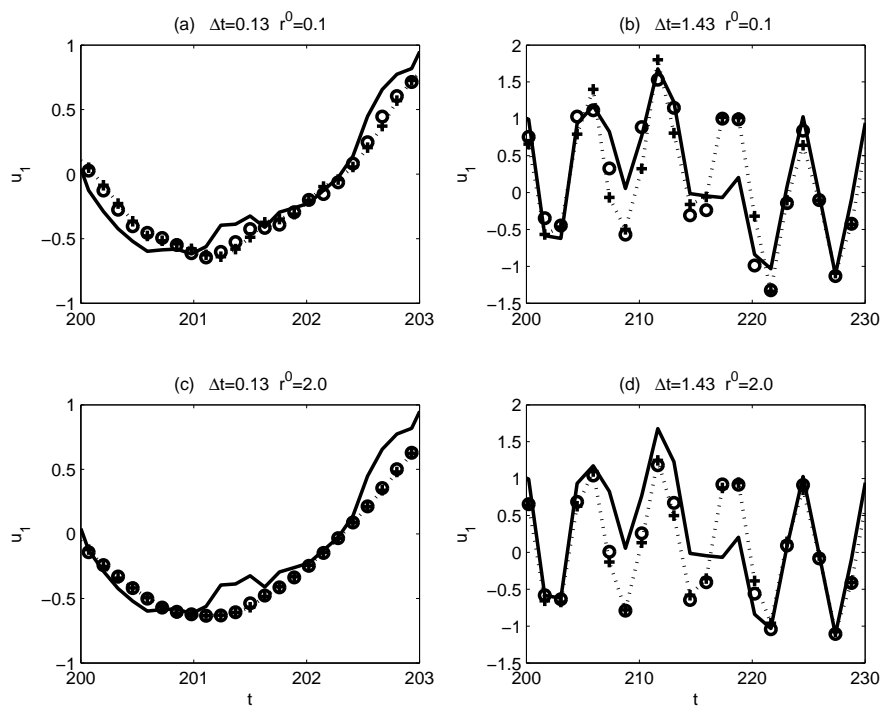

FIG. 4.1. *The truth signal u_1 is shown with the solid line, the prior forecast is shown with pluses connected with a dotted line, and the posterior signal is shown with circles. The values of Δt and r^0 are shown on top of each panel. The corresponding cross correlations XC_f and XC_p are given in Table 4.1. Note the different time scales for different Δt.*

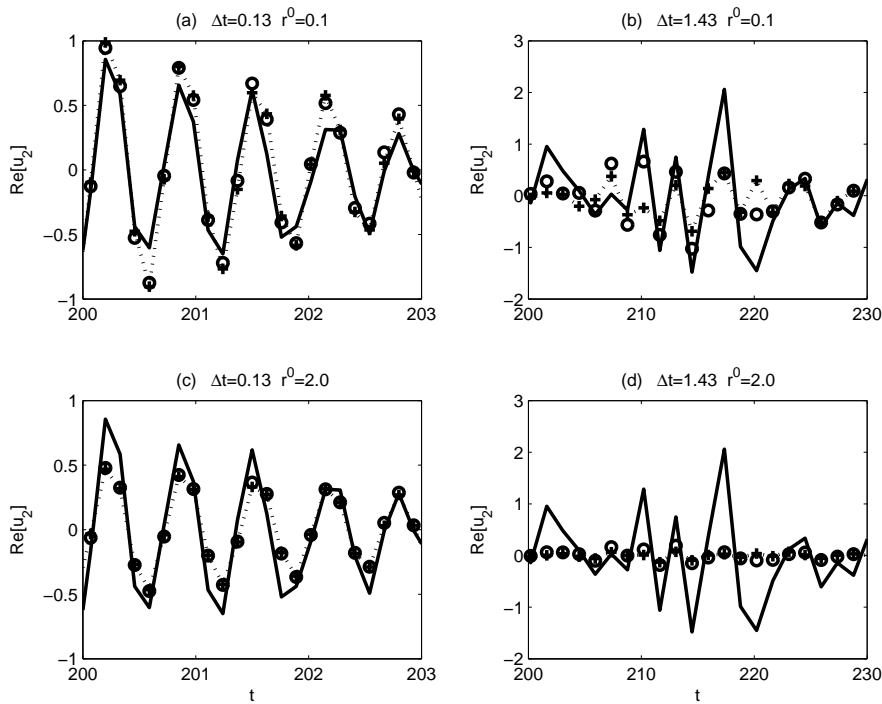Figure 4.2 shows the evolution of u_2 together with the prior forecast and posterior

FIG. 4.2. *The truth signal u_2 is shown with the solid line, the prior forecast is shown with pluses connected with a dotted line, and the posterior signal is shown with circles. The values of Δt and r^0 are shown on top of each panel. The corresponding cross correlations XC_f and XC_p are given in Table 4.2. Note the different time scales for different Δt.*

$r^0 \backslash \Delta t$	0.13	1.43
0.1	0.967 (0.955)	0.740 (0.544)
2.0	0.911 (0.897)	0.622 (0.357)

TABLE 4.2. *EKF performance on u_2: cross correlations XC_p (between truth and posterior) and XC_f (in the parenthesis, between truth and prior). The segments of the corresponding trajectories are shown in Figure 4.2.*

signal.

Table 4.2 shows the cross correlations XC_p and XC_f that correspond to the trajectories of u_2 (shown partially in Figure 4.2). Here, we also note the decrease in the skill of the EKF if larger Δt or r^0 are used. However, we should point out that if the time step Δt becomes significantly larger than the oscillation time T_2 (in our case $\Delta t = 1.43$ and $T_2 \approx 0.63$) we observe very poor filter skill in u_2 but not in u_1, which is still filtered quite well (compare $XC_p = 0.855$ for u_1 in Table 4.1, with $XC_p = 0.622$ for u_2 in Table 4.2 for $\Delta t = 1.43$ and $r^0 = 2.0$).

In Secs. 4.5 and 4.6, we present a thorough study of how the EKF performance depends on the observation time step Δt, observation variance r^0, type of observations, and the model that is used for the forecast.

$r^0 \backslash \Delta t$	0.13	1.43
0.1	0.988 (0.985)	0.936 (0.921)
2.0	0.951 (0.950)	0.855 (0.856)

TABLE 4.3. *EKF performance on u_1: cross correlations XC_p (between truth and posterior without model error, i.e., $a_0=1$) and XC_{me} (in the parenthesis, between truth and posterior with model error, i.e., $a_0=0$). The segments of the corresponding trajectories are shown in Figure 4.3.*

$r^0 \backslash \Delta t$	0.13	1.43
0.1	0.967 (0.864)	0.740 (0.543)
2.0	0.911 (0.618)	0.622 (0.436)

TABLE 4.4. *EKF performance on u_2: cross correlations XC_p (between truth and posterior without model error, i.e., $a_0=1$) and XC_{me} (in the parenthesis, between truth and posterior with model error, i.e., $a_0=0$). The segments of the corresponding trajectories are shown in Figure 4.4.*

4.3. Filtering with model error.   Here, we discuss how the filter performance changes if we use the linearized model with $a_0=0$ as a prior forecast instead of the exact nonlinear model with $a_0 \neq 0$. In Figures 4.3 and 4.4, we show segments of the trajectories of the truth signal together with two filtered signals, where one was obtained using the exact model, and another one — using the linearized model. Here, we had the same set of parameters that we used for generating Figures 4.1 and 4.2. Similarly, we constructed Tables 4.3 and 4.4 with the values of the cross-correlation XC_p already discussed above and cross correlation XC_{me} between the truth signal and the posterior signal obtained via the linearized model, i.e., with model error.

We note that although by setting $a_0=0$ we seem to make u_1 and u_2 decoupled, they still stay coupled through the mixed observations of type 1 or 2. Observations of type 3 make mode u_1 independent from mode u_2 throughout the filtering process with model error, i.e., with $a_0=0$. By comparing the filtered signals with the truth trajectories, we conclude that model error is quite small in the filtering of u_1 (Table 4.3) regardless of the size of r^0 or Δt. Observations of type 1 were used here. On the other hand, using the linearized model for the prior forecast drastically affects the performance of filtering u_2 — from Table 4.4, we conclude that $XC_p > XC_{me}$ for u_2. Below, we will study how model error depends on the type of observations, observation variance, and observation time step.

4.4. Filtering the linear model and the test of observability.   In Figure 4.5, we present the results of the numerical filtering of the linearized problem with $a_0=0$. In this simulation, the truth signal was also computed via the linearized version of equations (2.3) and (2.4), i.e., with $a_0=0$. Therefore, the Kalman filter produces the optimal filtered signal and there is no model error in this simulation. In Figure 4.5(a), we show the dependency of the RMSE on the observation time step Δt. In Figure 4.5(b), we demonstrate the corresponding dependency of $\det(\mathbf{O})^{-1}$ on Δt. We note the strong correlation between the two plots. The peaks that were predicted by observability analysis (Figure 4.5(b)) are observed at the same locations on the plot of RMSE of u_1 (Figure 4.5(a)). Here, we have taken very small observation variance $r^0=0.002$. Thus, we ensure that the filter tends to trust the observations and, therefore, observability becomes crucial. We note that for the fully observable
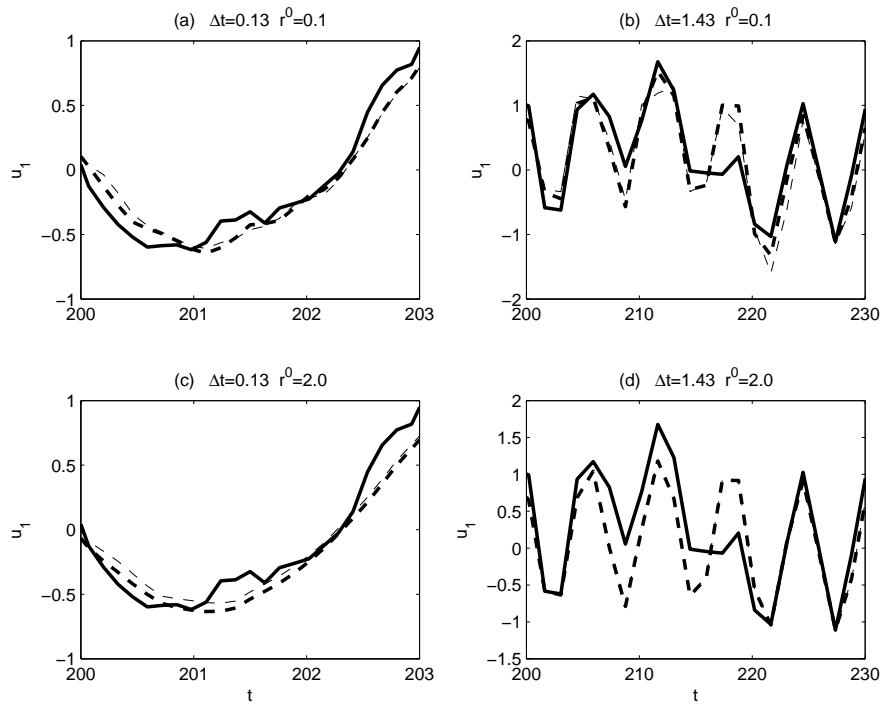
FIG. 4.3. *The truth signal u_1 is shown with the solid line, the posterior signal computed via nonlinear forecast is shown with bold dashed line, and the posterior signal computed via linear forecast is shown with the thin dashed line (note that in panel (d) the two dashed lines almost coincide). The corresponding values of Δt and r^0 are shown on top of each panel. The corresponding cross correlations XC_p and XC_{me} are given in Table 4.3. Note the different time scales for different Δt.*

case, i.e., with observations of type 3, there are no peaks at all. In that case, the observation matrix G always has rank 3, regardless of Δt. However, for observations of types 1 and 2 we clearly see lack of observability around the values of Δt predicted by equation (3.9).

4.5. Filter performance as a function of observation time step. Next, we study how performance of the EKF depends on the observation time step Δt under various conditions such as type of observations, observation variance, and the model used in the prior forecast. As we studied above, for the linear model ($a_0 = 0$) the dependence of the EKF performance on Δt is strongly correlated with observability properties of the EKF. Here, we will see how that result is reflected in filtering the nonlinear model.

In Figure 4.6, we show the RMSE of the filtered solution of u_1 compared with the truth signal of u_1. We have chosen three fixed values of observation variance: $r^0 = 0.002$, $r^0 = 0.256$, $r^0 = 2.048$ (panels (a), (b), and (c) in Figure 4.6). For each r^0, we consider all three types of observations. Moreover, for each type of observation, we used two models for the prior forecast: the perfect model with $a_0 = 1$ (the same a_0 that was used for creating the truth signal), and the linearized model with $a_0 = 0$ (with model error). As a result of this simulation, we draw the following conclusions:

1. The RMSE of u_1 approaches some finite value as Δt increases and r^0 is fixed.
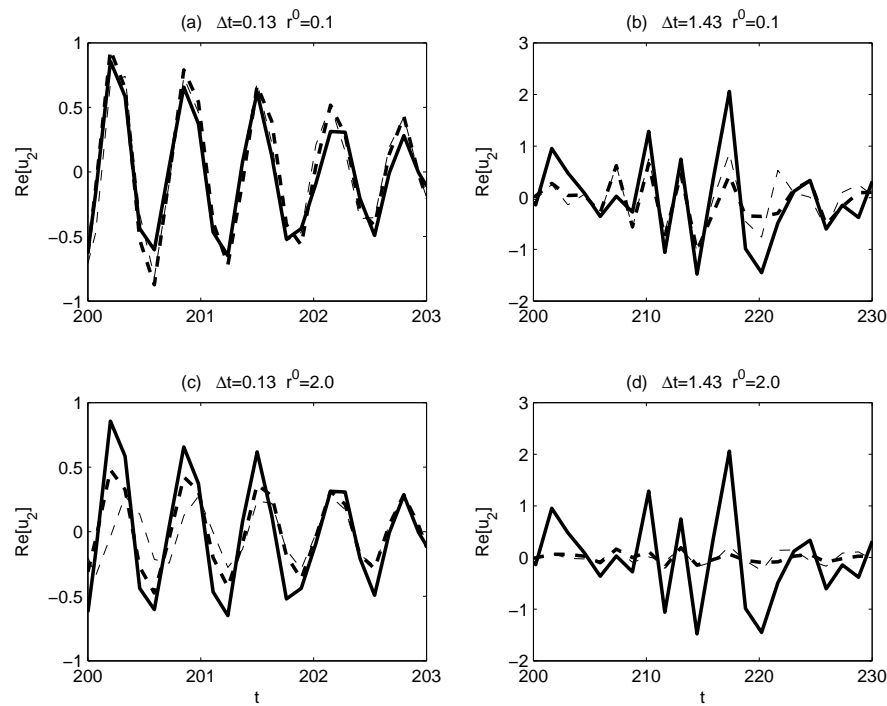
FIG. 4.4. *The truth signal u_2 is shown with the solid line, the posterior signal computed via nonlinear forecast is shown with the bold dashed line, and the posterior signal computed via linear forecast is shown with the thin dashed line (note that in panel (d) the two dashed lines almost coincide). The corresponding values of Δt and r^0 are shown on top of each panel. The corresponding cross correlations XC_p and XC_{me} are given in Table 4.4. Note the different time scales for different Δt.*

2. For observations of types 1 and 2, the EKF performance becomes very poor around the values of Δt, where the linearized system has lack of observability, whereas observations of type 3 make the EKF always fully observable.

3. A larger number of observations leads to better EKF performance regardless of r^0.

4. For observations of type 3, the RMSE is almost negligible for small r^0 and monotonically increases for larger r^0.

5. The model error is much larger with smaller r^0 and it becomes almost negligible for larger r^0 as Δt increases.

6. For larger r^0, the peaks that correspond to poor filter performance are less sharp than for smaller r^0, which is explained by the fact that observability is less important with a larger observation error than with a smaller one since greater weight is given to the dynamics.

We now study the dependence of the EKF performance on mode u_2, which is shown in Figure 4.7. We make the following observations after examining Figure 4.7:

1. The RMSE of u_2 approaches some finite value as Δt increases and r^0 is fixed.
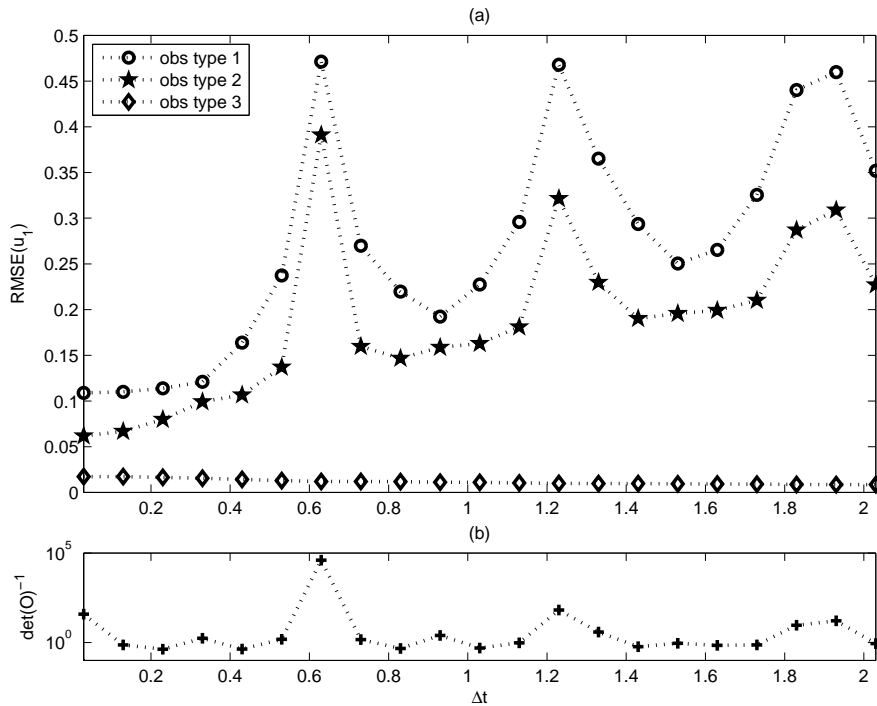
FIG. 4.5. *(a) The RMSE of u_1 for three types of observations and observation variance $r^0 = 0.002$ as a function of Δt, (b) $\det(\mathbf{O})^{-1}$ as a function of Δt (note the logarithmic scale of the y-axis). Weak damping was used.*

2. The RMSE of u_2 is smaller when the number of observations is larger.

3. Observations of type 1 and 2 show lack of observability at the values of Δt predicted by observability analysis (equation (3.9)).

4. Observations of type 1 and 2 have an extra peak at $\Delta t \approx T_2/2$, which also corresponds to lack of observability at this time step (see a small peak at $\Delta t \approx T_2/2$ in Figure 4.5(b)). This peak was absent in the corresponding plot for u_1.

5. For all Δt, the difference among the types of observations becomes smaller as r^0 increases.

6. For all Δt, the model error becomes smaller as r^0 increases.

4.6. Filter performance as a function of observation variance. In this Section, we study how the filter performance depends on r^0 for the fixed observation type and Δt. In Figure 4.8, we demonstrate the RMSE of filtering u_1 as a function of r^0 for three different values of Δt. Let us point out the conclusions that we draw from Figure 4.8:

1. The RMSE is a monotonically growing function of r^0.

2. A larger number of observations lead to better EKF performance regardless of Δt.

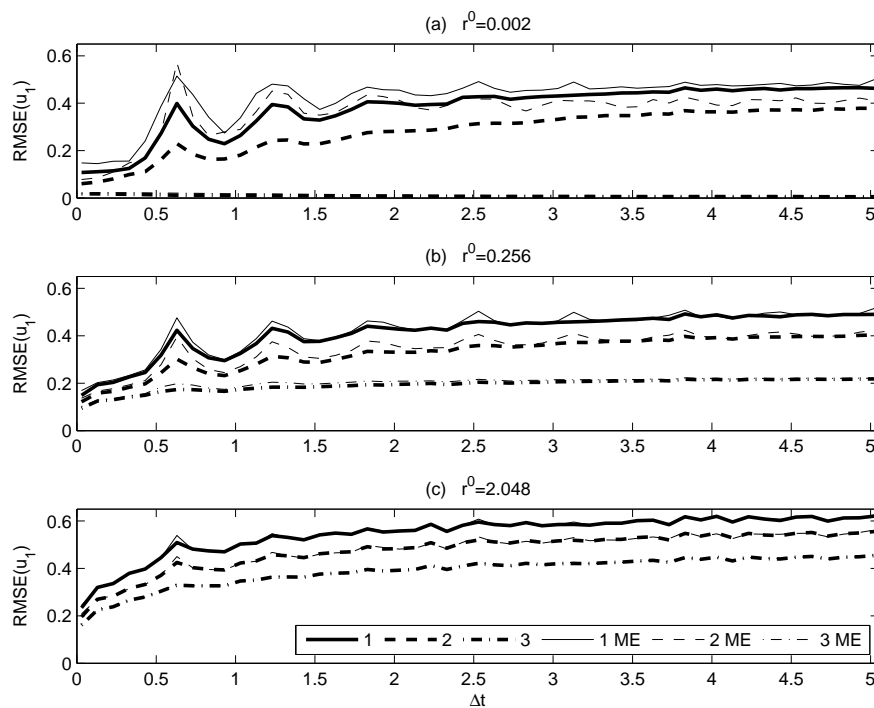3. In the limit $r^0 \to 0$, the RMSE of filtering with observations of types 1 and 2

FIG. 4.6. *The RMSE of u_1 as a function of Δt. Observations of type 1 (solid line), type 2 (dashed line), and type 3 (dashed-dotted line) with the nonlinear model (bold line) and the linearized model with model error (ME, thin line) were used in the simulation. A fixed r^0 (shown on top of the corresponding panel) was used in each simulation.*

converges to some finite positive value, which means that even if the observation error is negligible, the filter still does not reproduce the truth signal because the observations of type 1 and 2 only carry partial information about the truth signal.

4. Observations of type 3 lead to a vanishing error as $r^0 \to 0$.

5. As r^0 grows, the difference in the RMSE for different observation types decreases. This is explained by the fact that EKF chooses to trust the dynamics more when the observation error variance is large and, hence, the influence of the observations on the filtered signal diminishes.

6. The RMSE for filtering with model error ($a_0 = 0$) increases as $r^0 \to 0$ for $\Delta t = 0.63$ and observations of types 1 and 2 (thin solid and dashed lines in Figure 4.8(b)). For other values of Δt (for which the corresponding linear system is observable), the RMSE of filtering with model error is a monotonically decreasing function of r^0.

7. For observations of types 1 and 2, model error due to linearization (difference between the thick and thin solid (observations of type 1) and dashed (observations of type 2) curves) is increasing as $r^0 \to 0$. This indicates that the model error increases as r^0 approaches 0. We will have a more thorough study of the model error in section 4.7.
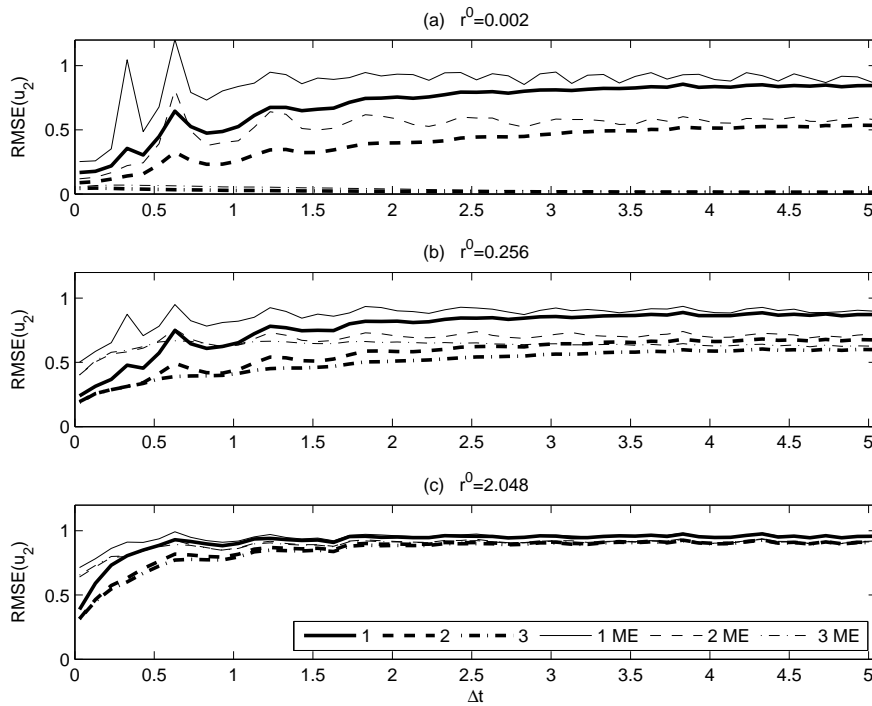
FIG. 4.7. *The RMSE of u_2 as a function of Δt. Observations of type 1 (solid line), type 2 (dashed line), and type 3 (dashed-dotted line) with the nonlinear model (bold line) and the linearized model with model error (ME, thin line) were used in the simulation. A fixed r^0 (shown on top of the corresponding panel) was used in each simulation.*

8. The model error is very small with observations of type 3 (thick and thin dashed-dotted lines in Figure 4.8 almost coincide.)

Next, we note that the RMSE of u_2 as a function of r^0, shown in Figure 4.9, behaves similarly to the RMSE of u_1. However, the error in filtering the fast mode u_2 is larger because the faster mode has more irregular time dynamics, which is harder to capture from the observations with finite step Δt. Moreover, the linearization ($a_0 = 0$) introduces a larger model error in the dynamics of u_2, since for u_2 the dynamical equation explicitly depends on a_0, unlike for u_1, where the dependence on a_0 only implicitly comes from observations.

4.7. Statistics of mean model error and individual realizations. Mean model error statistics are a standard part of offline testing for Kalman filter performance in linear models [1]. How useful are mean model error statistics as a qualitative guideline for filter performance in nonlinear slow-fast systems? The exact solution formulas for statistics in the test model allow us to address these issues here. Let us now study the mean model error due to linearization. In section 3.4, we estimated model error by following individual trajectories for a given observation type and the set of parameters Δt and r^0, and then by measuring averaged model error along these trajectories.

Now, we consider averaging over all possible ensembles of the random signal trajectories as well as the random noise in observations. This way, we will obtain a fully
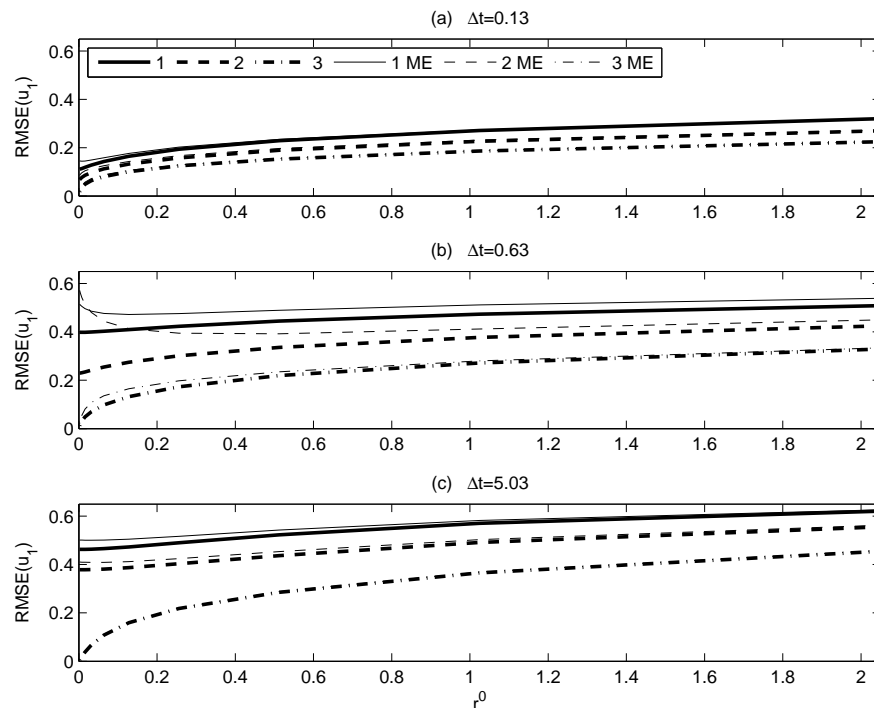
FIG. 4.8. *The RMSE of u_1 as a function of r^0. Observations of type 1 (solid line), type 2 (dashed line), and type 3 (dashed-dotted line) with the nonlinear model (bold line) and the linearized model with model error (ME, thin line) were used in the simulation. A fixed Δt (shown on top of the corresponding panel) was used in each simulation.*

deterministic system for estimating mean model error. The solution of this system requires much less computational work than resolving a long individual trajectory and then filtering it. As a result of the mean model error analysis, we expect to obtain some indications of the model error along the individual trajectories.

Next, we describe the algorithm of measuring the mean model error. Note that since we have the exact analytical formulas for the time evolution of the mean $\langle \mathbf{u} \rangle$ of the model in a nonlinear regime ($a_0 \neq 0$), we can compute it, and, therefore, obtain the truth signal for the mean. Moreover, we use the same formulas for $\langle \mathbf{u} \rangle$ together with the formulas for the evolution of the covariance but in the linearized form ($a_0 = 0$) to produce a prior forecast.

$$\bar{\mathbf{u}}_{m|m} \xrightarrow{a_0=0} \bar{\mathbf{u}}_{m+1|m}, \tag{4.4}$$

where we use the notation $\bar{\mathbf{u}}$ instead of $\langle \mathbf{u} \rangle$ to distinguish between the filtering of the mean signal and filtering of the individual random trajectories. The posterior forecast in the mean model error estimation is made via the following reasoning. The observations $\mathbf{v}_m = G\mathbf{u}_m$ have mean zero, therefore after averaging over all possible
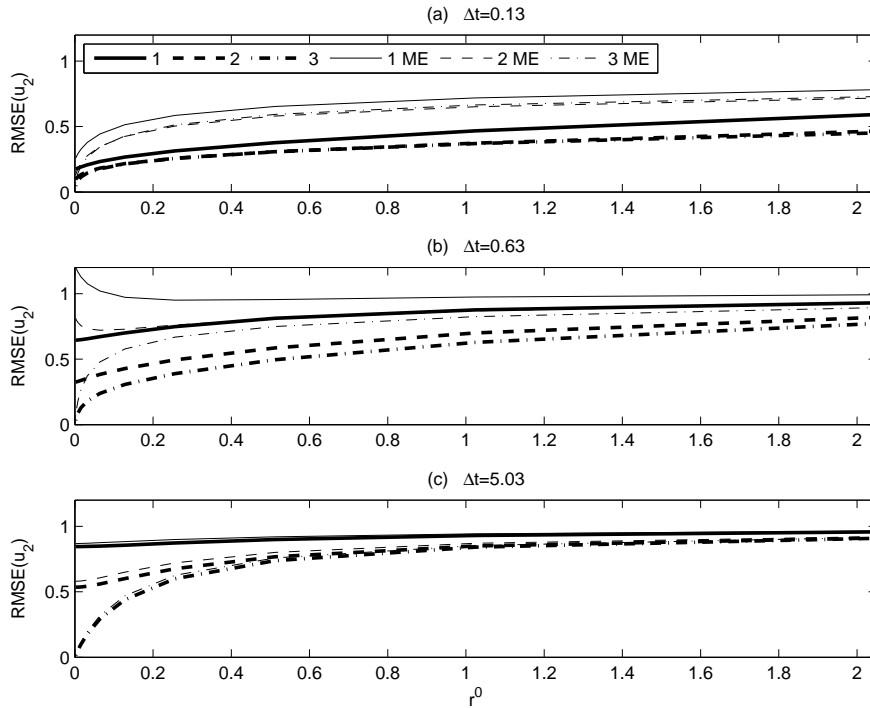
FIG. 4.9. *The RMSE of u_2 as a function of Δt. Observations of type 1 (solid line), type 2 (dashed line), and type 3 (dashed-dotted line) with the nonlinear model (bold line) and the linearized model with model error (ME, thin line) were used in the simulation. A fixed Δt (shown on top of the corresponding panel) was used in each simulation.*

trajectories we have $\bar{\mathbf{v}}_m = G\bar{\mathbf{u}}_m$. The correction step of the filter becomes

$$\bar{\mathbf{u}}_{m+1|m+1} = \bar{\mathbf{u}}_{m+1|m} + K_{m+1}G(\bar{\mathbf{u}}_{m+1} - \bar{\mathbf{u}}_{m+1|m}), \tag{4.5}$$

$$\Gamma_{m+1|m+1} = (I_3 - K_{m+1}G)\Gamma_{m+1|m}, \tag{4.6}$$

$$K_{m+1} = \Gamma_{m+1|m}G^T(G\Gamma_{m+1|m}G^T + R^0)^{-1}. \tag{4.7}$$

Note that the so-called off-line statistics Γ and K are computed by the same formulas as for the individual trajectories, and only the update of the mean is different. It is necessary to stress here that even though the Kalman filter algorithm for computing the mean model error seems to be the same as for computing the filtered signal of individual trajectories, in the mean model error analysis we do averaging prior to filtering while previously, in the filtering the individual trajectories, we first performed filtering and then did averaging.

In Figure 4.10, we demonstrate the time evolution of the mean (panels (a) and (b)) and the covariance (panels (c)-(f)) of $(\bar{u}_1, \bar{u}_2)$ and their corresponding filtered values. Observations of type 1 were used. From Figure 4.10(a), we see that the error in filtering $\bar{u}_1$ is very small. On the other hand, the error in filtering $\bar{u}_2$ is rather large within the decorrelation time $(\sim 1/\gamma_1 \approx 1/\gamma_2)$ until both the truth signal $\bar{u}_{2,m}$ and the filtered signal $\bar{u}_{2,m|m}$ converge to zero. The mean model error is characterized by the measure of the difference between $\bar{u}_m$ and $\bar{u}_{m|m}$, e.g., the cross covariance between
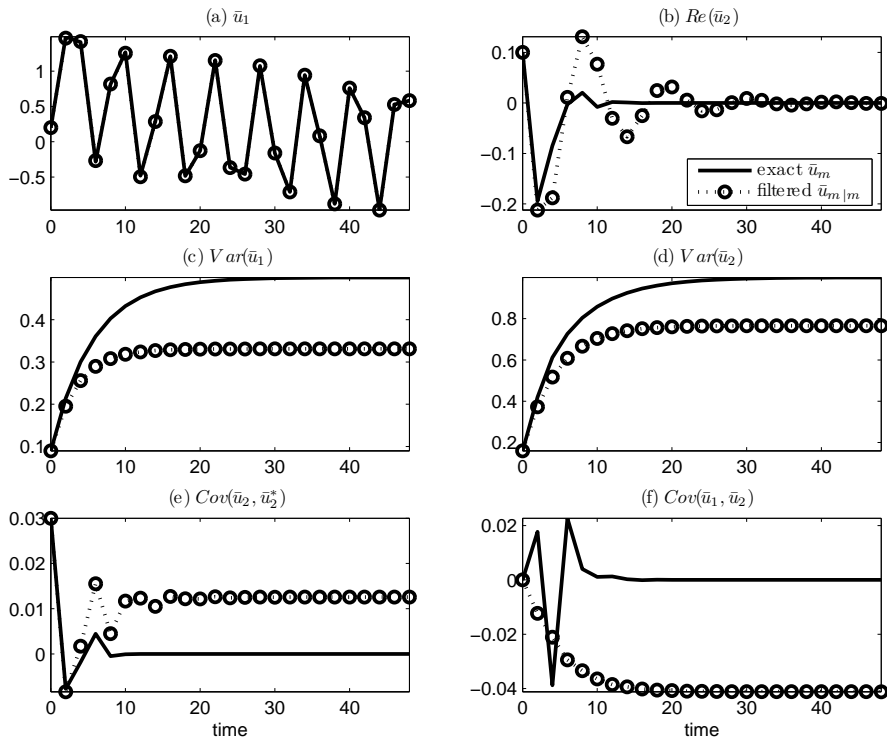
FIG. 4.10. *Mean model error. The solid line shows the truth values and the circles show the filtered values.* $r^0 = 2.048$, $\Delta t = 2$, *and observations of type 1 were used.*

the two of them. The various components of the truth covariance and posterior covariance reach constant values within the decorrelation time. The difference between the corresponding components of the truth and posterior covariance can be another characteristic of the mean model error.

Next, we study what information about the mean model error is relevant to estimating the model error in filtering individual trajectories, and what information can be misleading. In Figure 4.11, we show the dependence of mean model error of the slow mode u_1 on the observation time step Δt for three fixed values observation variance r^0. As a measure of the mean model error, we have chosen the cross-correlation XC_1^M between $\bar{u}_{1,m}$ and $\bar{u}_{1,m|m}$ along the typical decorrelation time (chosen to be $\tau = 30$ from studying Figure 4.10, as it appears to be the typical relaxation time). On the other hand, in Figure 4.12, we demonstrate the cross correlation XC_1 between the two signals $\langle u_{1,m|m} \rangle$ filtered from the same individual trajectory: one computed via the nonlinear dynamics ($a_0 \neq 0$) and the other one computed via the linearized dynamics ($a_0 = 0$). Moreover, in Figures 4.11 and 4.14, we show similar dependencies of XC_2^M and XC_2 on Δt for the fast mode u_2. It is very instructive to compare Figure 4.11 with Figure 4.12 and Figure 4.13 with Fig 4.14.

Below, we present the ideas about the model error that we can predict from the mean model error analysis, i.e., the similarities between XC_1^M and XC_1 and between XC_2^M and XC_2:

    1. The model error is large at the values of Δt that were predicted by observ-

ability analysis.

2. The model error of filtering with observations of type 3 is much less that with observations of type 1 and 2.

3. For larger r^0, model error is less dependent on the type of observations.

4. Disregarding local oscillations, the model error for u_2 first increases as Δt increases and then decreases. This does not mean that the filter skill improves as Δt increases, it just means that model error due to linearization becomes less important.

5. For larger r^0, observations of types 2 and 3 give almost same model error, while observations of type 1 lead to larger model error.

Now, we point out the properties of the model error along individual trajectories that are not predicted by the mean model error analysis

1. For smaller r^0, observations of type 2 have larger model error than observations of type 1.

2. The measure of the model error is not predicted well quantitatively, although the qualitative behavior is similar.

3. The dependence on Δt appears to be quite smooth (especially for observations of types 2 and 3), while the mean model error analysis predicts it to be more oscillatory.
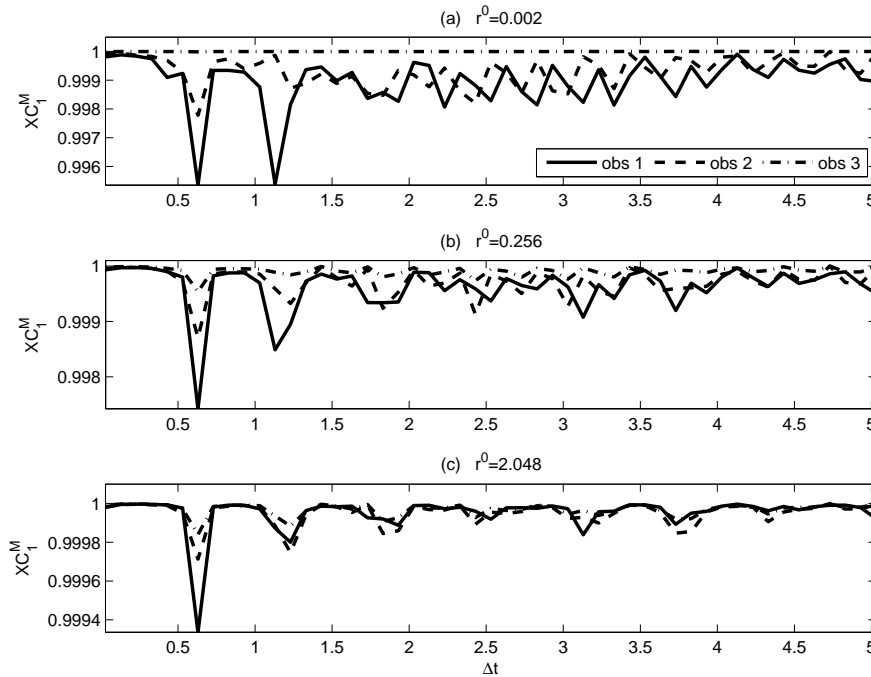

FIG. 4.11. *Mean model error of u_1. Observations of type 1 (solid line), type 2 (dashed line), and type 3 (dashed-dotted line) are shown.*

What is particularly striking here is the strong correlation of the mean model error for the fast mode u_2, from Figure 4.13, with observations of types 1 and 2 (the
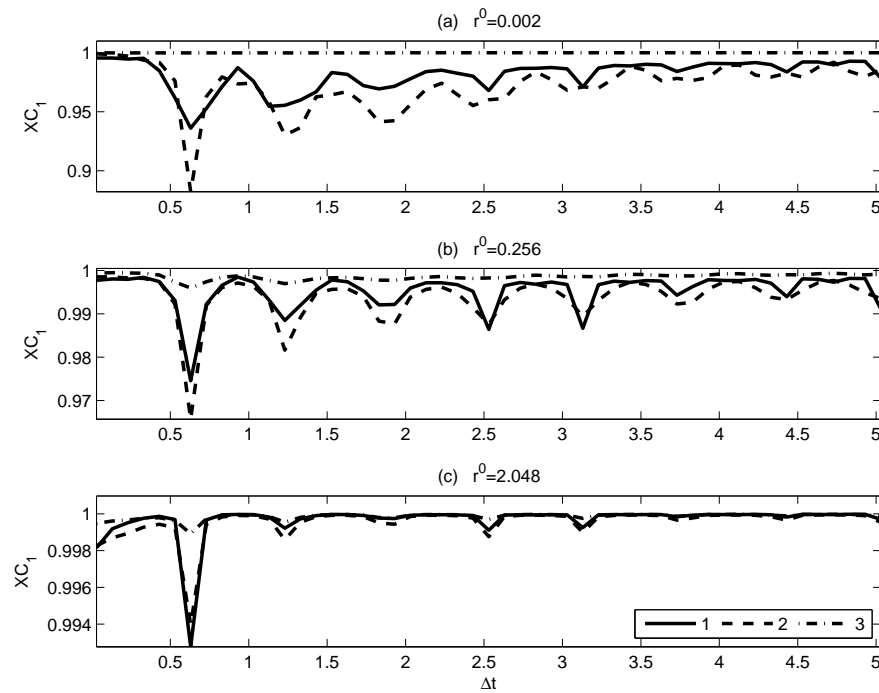
FIG. 4.12. *Model error of u_1. Observations of type 1 (solid line), type 2 (dashed line), and type 3 (dashed-dotted line) are shown.*

two larger observational noises) with the non-Gaussianity in the truth signal displayed in Figures 2.5 and 2.6. With larger observational noise, the Kalman filter trusts the Gaussian dynamics of the linearized model significantly, so there are larger mean model errors on the faster mode, even with three observations here.

Next, we study if the dependence of the mean model error on r^0 can provide any guidelines about the model error of the individual trajectories. We compare Figure 4.15 with Figure 4.16 for u_1 and Figure 4.17 with Figure 4.18 for u_2. Here, we outline the properties of the model error that follow from the mean model error analysis.

1. The model error of u_1 decreases as r^0 increases except for very small r^0 and observations of type 3, where the model error vanishes as $r^0 \to 0$.
2. Observations of type 3 have the smallest model error in u_1.
3. The model error of u_2 grows as r^0 increases (note opposite behavior for u_1).
4. The model error becomes independent of the observation type as both Δt and r^0 increase.

On the other hand, we note that there are a few facts about the model error that are not reflected in mean model error analysis

1. For small $\Delta t = 0.13$, observations of type 1 show better filtering skill than observations of type 2 for r^0 greater than some value (compare the solid and dashed lines in Figure 4.16(a)) ($r^0 > 0.7$.)
2. For bigger $\Delta t = 0.63$, observations of type 1 show better filtering skill than observations of type 2 for r^0 smaller than some value (compare the solid and
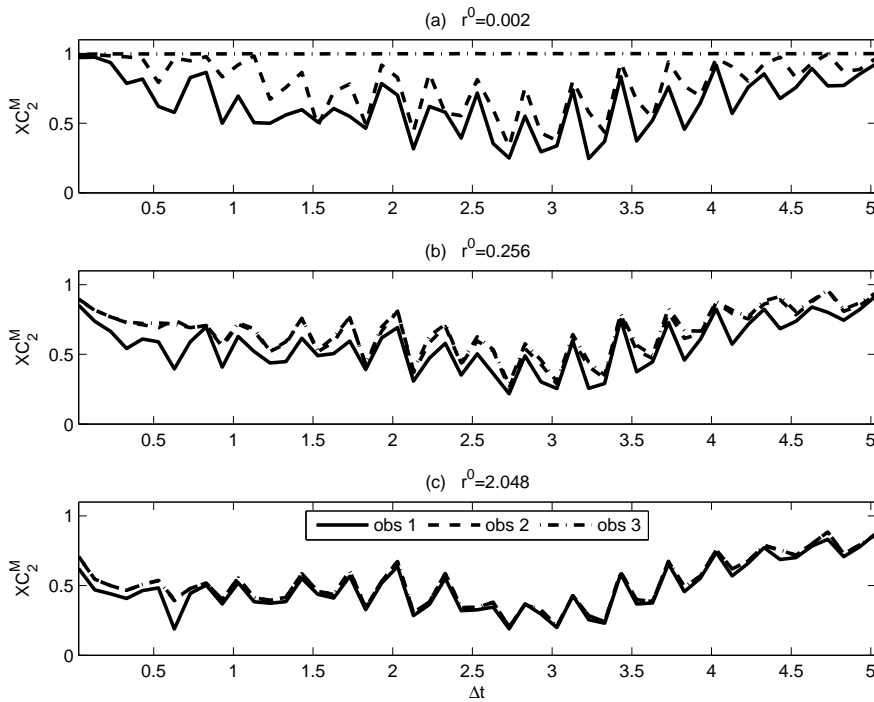
FIG. 4.13. *Mean model error of u_2. Observations of type 1 (solid line), type 2 (dashed line), and type 3 (dashed-dotted line) are shown.*

dashed lines in Figure 4.16(b)) ($r^0 < 0.8$)

3. The mean model error analysis only provides qualitative estimates about the model error.

5. Concluding discussion

In the introduction, we motivated the need to develop a simple nonlinear test model for filtering slow-fast systems and introduced a nonlinear three dimensional real stochastic triad model as the simplest prototype model for slow-fast systems. In section 2, we established analytic non-Gaussian nonlinear statistical properties of the model including exact formulas for the propagating mean and covariance. An exact nonlinear extended Kalman filter (EKF) for the slow-fast system and a linear Kalman filter with a non-trivial model error on the fast dynamics through linearization at the climitological mean state were introduced in section 3. Various aspects of filter performance for these two filters were discussed in section 4. While there were detailed summaries in section 4, it is useful to summarize several features of the filter performance.

First, the partial observations of type 1 and 2 were designed to mix both slow and fast modes as occurs in many practical applications. The theoretical analysis of observability in section 3.4.1 lead us to predict deteriorating filter skill in the vicinity of the non-observable times for observations of type 1 and 2 predicted by equation (3.9); this prediction is confirmed for both the linear and nonlinear filters in Figures 4.5-4.7 for small and moderate observational noise. On the other hand, for observation
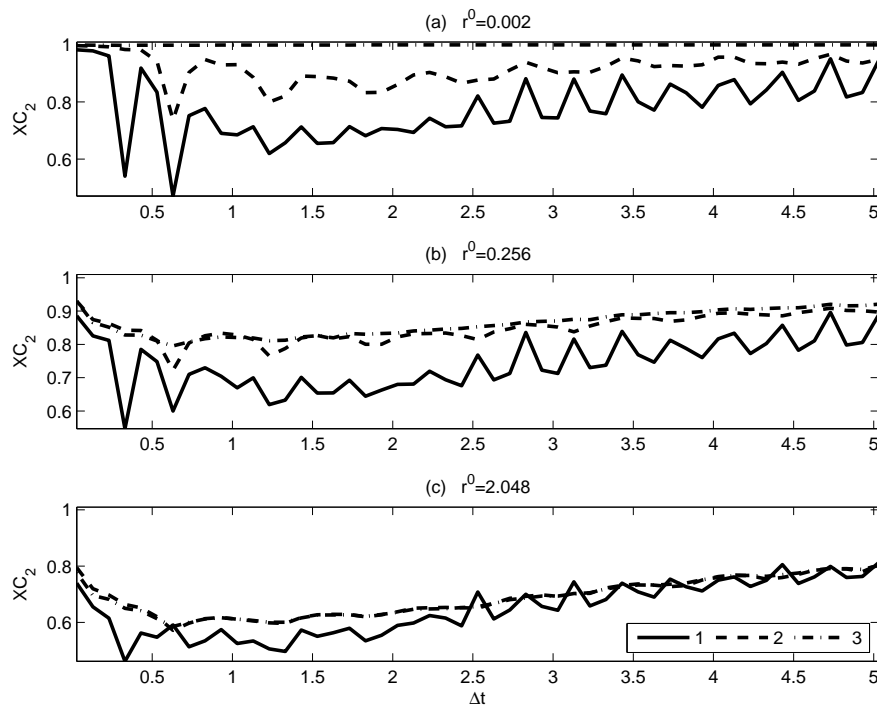
FIG. 4.14. *Model error of u_2. Observations of type 1 (solid line), type 2 (dashed line), and type 3 (dashed-dotted line) are shown.*

time steps Δt, which are away from the non-observable ones and either shorter or longer than the fast oscillation period, as reported in section 4.2, the nonlinear EKF always has significant skill on the slow mode u_1 for observations of type 1, which mix the slow and fast modes, and retains non-trivial skill for the fast modes (Tables 4.1 and 4.2); on the other hand, the linear filter with model error retains the significant filter skill of the nonlinear model for the slow mode (Table 4.3) but the filter skill for the fast mode deteriorates significantly (Table 4.4). The slow mode in the nonlinear test model has an exact linear slow manifold [27] and the filter skill with model error on the slow mode reflects this fact; nevertheless, these results here suggest interesting possibilities for filtering slow-fast systems by linear systems with model error [15, 16] when only estimates of the slow dynamics are required in an application. Finally, in section 4.7, we compared the exact nonlinear analysis for filtering skill through offline (super)ensemble mean model error in the test model with the actual filter performance for individual realizations discussed throughout the paper, since such offline tests can provide important guidelines for filter performance [1, 22, 4, 14]. It is established in section 4.7 that the offline mean model error analysis for the nonlinear test model provides very good qualitative guidelines for the actual filter performance on individual signals with some discrepancies discussed in detail there.

In future work, the authors plan to utilize the nonlinear test model with non-trivial forcing of the fast waves, $f_2(t) \neq 0$ in equation (1.2), to develop filters with significant skill for both the slow mode and the slow envelope of the fast modes. As noted earlier in the introduction, this is an important problem for filter skill in
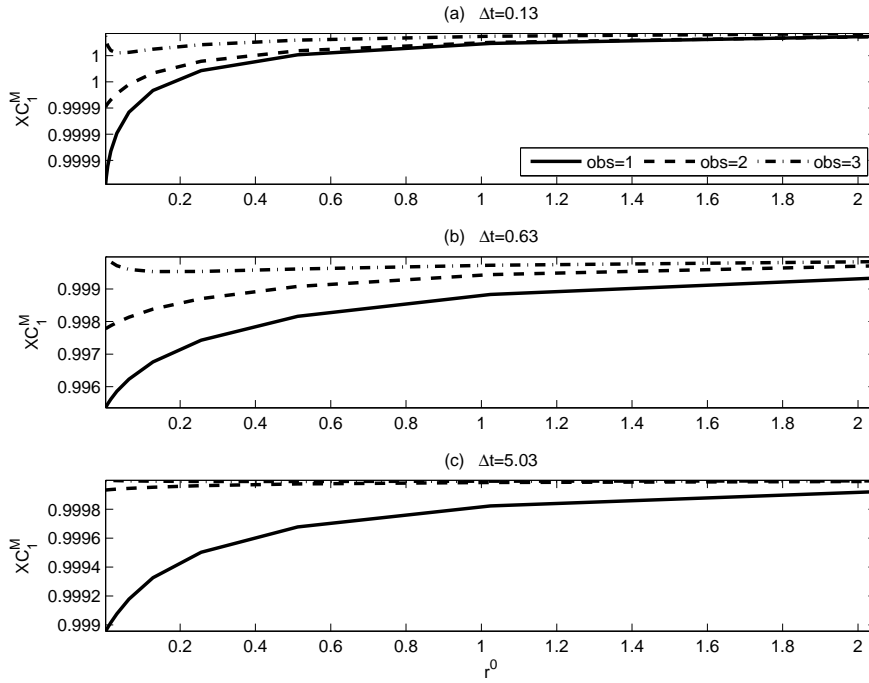
FIG. 4.15. *Mean model error of u_1. Observations of type 1 (solid line), type 2 (dashed line), and type 3 (dashed-dotted line) are shown.*

contemporary meteorology where on mesoscales, slow and fast modes are coupled nonlinearly through moist convection [17].

Appendix A. Computation of cross-covariances. In this Appendix, we compute the cross-covariances $\mathrm{Cov}(u_2, u_2^*)$ and $\mathrm{Cov}(u_2, u_1)$. For $\mathrm{Cov}(u_2, u_2^*)$ we find

$$\mathrm{Cov}(u_2, u_2^*) = \left\langle (u_2 - \langle u_2 \rangle)(u_2 - \langle u_2 \rangle) \right\rangle = \langle u_2^2 \rangle - \langle u_2 \rangle^2 = e^{-2\gamma_2(t-t_0)} \langle \psi^2 u_{20}^2 \rangle - \langle u_2 \rangle^2,$$

where

$$\langle \psi^2 u_{20}^2 \rangle = \psi_D^2 \langle \psi_W^2 \rangle \langle u_{20}^2 \exp(i2bu_{10}) \rangle. \tag{A.1}$$
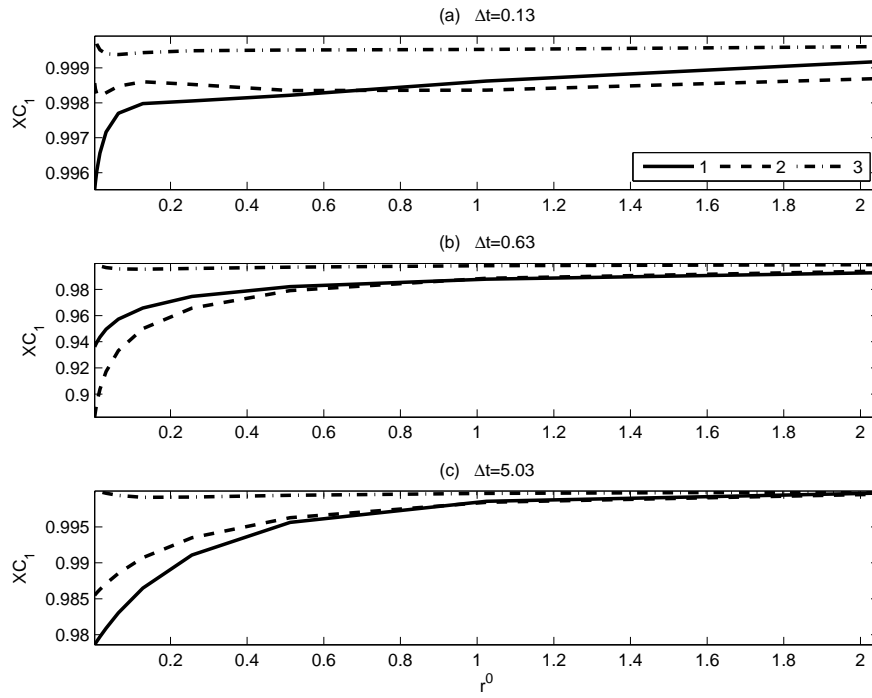
FIG. 4.16. *Model error of u_1. Observations of type 1 (solid line), type 2 (dashed line), and type 3 (dashed-dotted line) are shown.*

Next, for $\mathrm{Cov}(u_2, u_1)$ we obtain

$$\mathrm{Cov}(u_2, u_1) = \Big\langle (u_2 - \langle u_2 \rangle)(u_1 - \langle u_1 \rangle) \Big\rangle = \Big\langle \Big(e^{-\gamma_2(t-t_0)} [\psi(t_0, t) u_{20} - \langle \psi(t_0, t) u_{20} \rangle]$$

$$+ \sigma_2 \int_{t_0}^t e^{-\gamma_2(t-s)} \psi(s, t) dW_2(s) \Big) \Big([u_{10} - \langle u_{10} \rangle] e^{-\gamma_1(t-t_0)} + \sigma_1 \int_{t_0}^t e^{-\gamma_1(t-s)} dW_1(s) \Big) \Big\rangle$$

$$= e^{-(\gamma_1+\gamma_2)(t-t_0)} [\langle u_{10} u_{20} \psi(t_0, t) \rangle - \langle u_{10} \rangle \langle u_{20} \psi(t_0, t) \rangle]$$

$$+ \sigma_1 e^{-\gamma_2(t-t_0)} \Big\langle \int_{t_0}^t e^{-\gamma_1(t-s)} dW_1(s) u_{20} \psi(t_0, t) \Big\rangle$$

In order to compute the last term, we note that

$$\frac{\partial}{\partial t} \Big\langle \exp(i J_W(t_0, t)) \Big\rangle$$

$$= \sigma_1 a_0 \frac{\partial}{\partial t} \Big\langle \exp\Big(i \int_{t_0}^t ds \int_{t_0}^s e^{-\gamma_1(s-s')} dW_1(s') \Big) \Big\rangle$$

$$= i \sigma_1 a_0 \Big\langle \exp(i J_W(t_0, t)) \int_{t_0}^t e^{-\gamma_1(t-s')} dW_1(s') \Big\rangle.$$

On the other hand, we obtain

$$\frac{\partial}{\partial t} \Big\langle \exp(i J_W(t_0, t)) \Big\rangle = -\frac{1}{2} \exp\Big(\frac{1}{2} \mathrm{Var}(J_W(t_0, t)) \Big) \frac{\partial}{\partial t} \Big(\mathrm{Var}(J_W(t_0, t)) \Big).$$

(a) $\Delta t = 0.13$
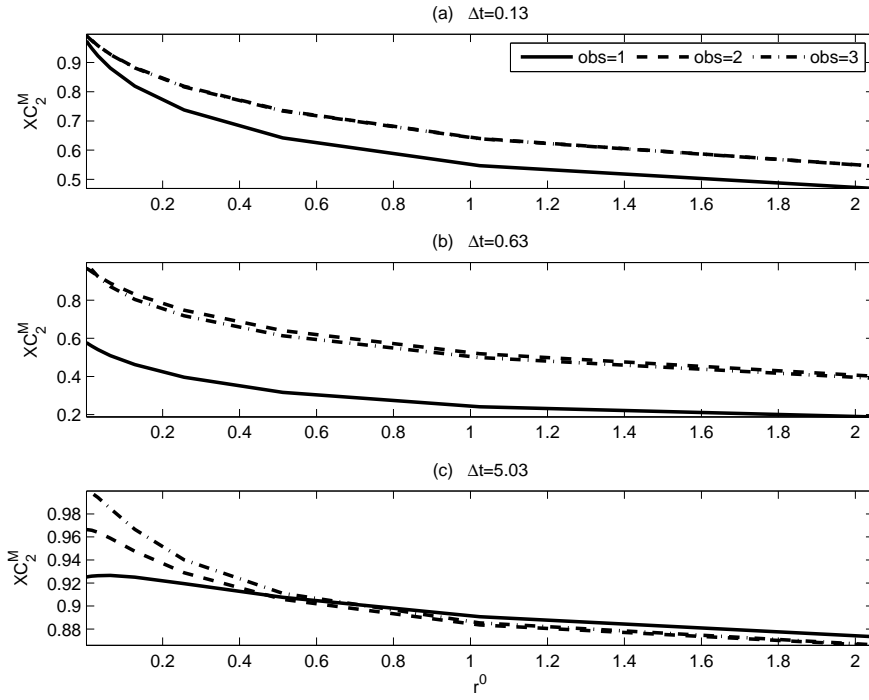
(b) $\Delta t = 0.63$

(c) $\Delta t = 5.03$

FIG. 4.17. *Mean model error of u_2. Observations of type 1 (solid line), type 2 (dashed line), and type 3 (dashed-dotted line) are shown.*

Finally, we obtain

$$\mathrm{Cov}(u_2, u_1) = \psi_D \langle \psi_W \rangle e^{-(\gamma_1 + \gamma_2)(t - t_0)} \Big(\langle u_{10} u_{20} \exp(ibu_{10}) \rangle - \langle u_{10} \rangle \langle u_{20} \exp(ibu_{10}) \rangle \Big)$$

$$+ \frac{i}{2} \psi_D \langle \psi_W \rangle e^{-\gamma_2(t - t_0)} \frac{1}{a_0} \frac{\partial}{\partial t} \Big(\mathrm{Var}(J_W(t_0, t)) \Big) \langle u_{20} \exp(ibu_{10}) \rangle. \qquad \text{(A.2)}$$

Note that $\langle u_{20} \exp(ibu_{10}) \rangle$ is given by equation (2.33). In order to find averages $\langle u_{20}^2 \exp(i2bu_{10}) \rangle$ and $\langle u_{10} u_{20} \exp(ibu_{10}) \rangle$, we use the real representation given by equations (2.25)–(2.27), which reduces the problem to computing the real-valued averages $\langle x_0 y_0 \exp(ibx_0) \rangle$, $\langle x_0 z_0 \exp(ibx_0) \rangle$, $\langle y_0^2 \exp(i2bx_0) \rangle$, $\langle z_0^2 \exp(i2bx_0) \rangle$, $\langle y_0 z_0 \exp(i2bx_0) \rangle$. Using the definition of the characteristic function (2.21), we find
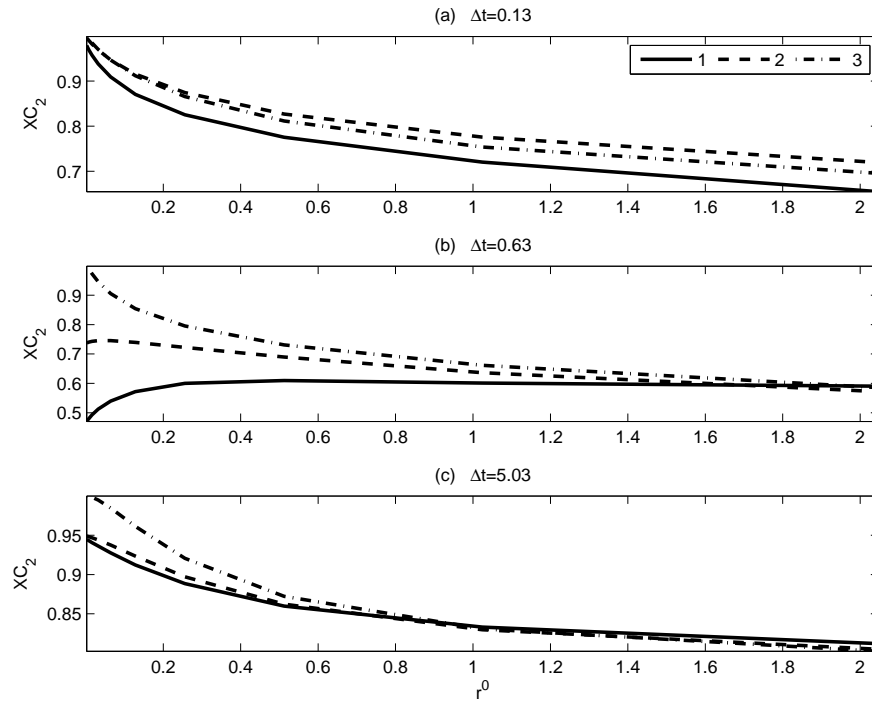
FIG. 4.18. *Model error of u_2. Observations of type 1 (solid line), type 2 (dashed line), and type 3 (dashed-dotted line) are shown.*

$$\langle x_0 y_0 \exp(ibx_0) \rangle = -\left.\frac{\partial^2 \phi_v(\mathbf{s})}{\partial s_1 \partial s_2}\right|_{\mathbf{s}=(b,0,0)^T},$$

$$\langle x_0 z_0 \exp(ibx_0) \rangle = -\left.\frac{\partial^2 \phi_v(\mathbf{s})}{\partial s_1 \partial s_3}\right|_{\mathbf{s}=(b,0,0)^T},$$

$$\langle y_0^2 \exp(i2bx_0) \rangle = -\left.\frac{\partial^2 \phi_v(\mathbf{s})}{\partial s_2^2}\right|_{\mathbf{s}=(2b,0,0)^T},$$

$$\langle z_0^2 \exp(i2bx_0) \rangle = -\left.\frac{\partial^2 \phi_v(\mathbf{s})}{\partial s_3^2}\right|_{\mathbf{s}=(2b,0,0)^T},$$

$$\langle y_0 z_0 \exp(i2bx_0) \rangle = -\left.\frac{\partial^2 \phi_v(\mathbf{s})}{\partial s_2 \partial s_3}\right|_{\mathbf{s}=(2b,0,0)^T}.$$

For a Gaussian vector $\mathbf{v} = (x_0, y_0, z_0)^T$ and any complex vector $\mathbf{s} = (s_1, s_2, s_3)^T$, we have a characteristic function in the form

$$\phi_{\mathbf{v}}(\mathbf{s}) = \int \exp(i\mathbf{s}^T \mathbf{v}) g_{\langle \mathbf{v} \rangle, \Sigma}(\mathbf{v}) d\mathbf{v}, \tag{A.3}$$

where mean is given by

$$\langle \mathbf{v} \rangle = (\langle x_0 \rangle, \langle y_0 \rangle, \langle z_0 \rangle)^T, \tag{A.4}$$

and covariance matrix is given by

$$\Sigma = \begin{pmatrix} \mathrm{Var}(x_0) & \mathrm{Cov}(x_0, y_0) & \mathrm{Cov}(x_0, z_0) \\ \mathrm{Cov}(x_0, y_0) & \mathrm{Var}(y_0) & \mathrm{Cov}(y_0, z_0) \\ \mathrm{Cov}(x_0, z_0) & \mathrm{Cov}(y_0, z_0) & \mathrm{Var}(z_0) \end{pmatrix}.$$

If we expand the inner products, we obtain

$$\phi_{\mathbf{v}}(\mathbf{s}) = \exp\left(i(s_1 \langle x_0 \rangle + s_2 \langle y_0 \rangle + s_3 \langle z_0 \rangle) - \frac{1}{2}\left(\mathrm{Var}(x_0)s_1^2 + \mathrm{Var}(y_0)s_2^2 + \mathrm{Var}(z_0)s_3^2 \right) \right)$$
$$\times \exp\left(-\mathrm{Cov}(x_0, y_0)s_1 s_2 - \mathrm{Cov}(x_0, z_0)s_1 s_3 - \mathrm{Cov}(y_0, z_0)s_2 s_3 \right). \tag{A.5}$$

After evaluating the partial derivatives of the characteristic function at the prescribed points, we find the following equations

$$\frac{\partial^2 \phi_v(\mathbf{s})}{\partial s_1 \partial s_2} = -\mathrm{Cov}(x_0, y_0)\phi_v(\mathbf{s}) + \left(i\langle x_0 \rangle - \mathrm{Var}(x_0)s_1 - \mathrm{Cov}(x_0, y_0)s_2 - \mathrm{Cov}(x_0, z_0)s_3\right)$$
$$\times \left(i\langle y_0 \rangle - \mathrm{Var}(y_0)s_2 - \mathrm{Cov}(x_0, y_0)s_1 - \mathrm{Cov}(y_0, z_0)s_3\right)\phi_v(\mathbf{s}),$$

$$\frac{\partial^2 \phi_v(\mathbf{s})}{\partial s_1 \partial s_3} = -\mathrm{Cov}(x_0, z_0)\phi_v(\mathbf{s}) + \left(i\langle x_0 \rangle - \mathrm{Var}(x_0)s_1 - \mathrm{Cov}(x_0, y_0)s_2 - \mathrm{Cov}(x_0, z_0)s_3\right)$$
$$\times \left(i\langle z_0 \rangle - \mathrm{Var}(z_0)s_3 - \mathrm{Cov}(x_0, z_0)s_1 - \mathrm{Cov}(y_0, z_0)s_2\right)\phi_v(\mathbf{s}),$$

$$\frac{\partial^2 \phi_v(\mathbf{s})}{\partial s_2^2} = \left(-\mathrm{Var}(y_0) + \left(i\langle y_0 \rangle - \mathrm{Var}(y_0)s_2 - \mathrm{Cov}(x_0, y_0)s_1 - \mathrm{Cov}(y_0, z_0)s_3\right)^2\right)\phi_v(\mathbf{s}),$$

$$\frac{\partial^2 \phi_v(\mathbf{s})}{\partial s_3^2} = \left(-\mathrm{Var}(z_0) + \left(i\langle z_0 \rangle - \mathrm{Var}(z_0)s_3 - \mathrm{Cov}(x_0, z_0)s_1 - \mathrm{Cov}(y_0, z_0)s_2\right)^2\right)\phi_v(\mathbf{s}),$$

$$\frac{\partial^2 \phi_v(\mathbf{s})}{\partial s_2 \partial s_3} = -\mathrm{Cov}(y_0, z_0)\phi_v(\mathbf{s}) + \left(i\langle y_0 \rangle - \mathrm{Var}(y_0)s_2 - \mathrm{Cov}(x_0, y_0)s_1 - \mathrm{Cov}(y_0, z_0)s_3\right)$$
$$\times \left(i\langle z_0 \rangle - \mathrm{Var}(z_0)s_3 - \mathrm{Cov}(x_0, z_0)s_1 - \mathrm{Cov}(y_0, z_0)s_2\right)\phi_v(\mathbf{s}),$$

Finally, we obtain the following averages

$$\langle x_0 y_0 \exp(ibx_0) \rangle = \left(\mathrm{Cov}(x_0, y_0) - \left(i\langle x_0 \rangle - \mathrm{Var}(x_0)b\right)\left(i\langle y_0 \rangle - \mathrm{Cov}(x_0, y_0)b\right)\right)$$
$$\times \exp\left(ib\langle x_0 \rangle - \frac{1}{2}\mathrm{Var}(x_0)b^2\right),$$

$$\langle x_0 z_0 \exp(ibx_0) \rangle = \left(\mathrm{Cov}(x_0, z_0) - \left(i\langle x_0 \rangle - \mathrm{Var}(x_0)b\right)\left(i\langle z_0 \rangle - \mathrm{Cov}(x_0, z_0)b\right)\right)$$
$$\times \exp\left(ib\langle x_0 \rangle - \frac{1}{2}\mathrm{Var}(x_0)b^2\right),$$

$$\langle y_0^2 \exp(i2bx_0) \rangle = \left(\mathrm{Var}(y_0) - \left(i\langle y_0 \rangle - \mathrm{Cov}(x_0, y_0)2b\right)^2\right)$$
$$\times \exp\left(i2b\langle x_0 \rangle - 2\mathrm{Var}(x_0)b^2\right),$$

$$\langle z_0^2 \exp(i2bx_0) \rangle = \left(\mathrm{Var}(z_0) - \left(i\langle z_0 \rangle - \mathrm{Cov}(x_0, z_0)2b\right)^2\right)$$
$$\times \exp\left(i2b\langle x_0 \rangle - 2\mathrm{Var}(x_0)b^2\right),$$

$$\langle y_0 z_0 \exp(i2bx_0) \rangle = \left(\mathrm{Cov}(y_0, z_0) - \left(i\langle y_0 \rangle - \mathrm{Cov}(x_0, y_0)2b\right)\left(i\langle z_0 \rangle - \mathrm{Cov}(x_0, z_0)2b\right)\right)$$
$$\times \exp\left(i2b\langle x_0 \rangle - 2\mathrm{Var}(x_0)b^2\right),$$

Next, we use the following properties of the covariances

$$\mathrm{Cov}(u_2, u_2^*) = \mathrm{Var}(y) - \mathrm{Var}(z) + 2i\mathrm{Cov}(y, z), \tag{A.6}$$
$$\mathrm{Cov}(u_2, u_1) = \mathrm{Cov}(x, y) + i\mathrm{Cov}(x, z), \tag{A.7}$$

to obtain

$$\langle u_{10} u_{20} \exp(ibu_{10})\rangle$$

$$= \Bigg(\langle u_{10}\rangle\langle u_{20}\rangle + \mathrm{Cov}(u_{20}, u_{10}) + ib\Big(\langle u_{10}\rangle\mathrm{Cov}(u_{20}, u_{10}) + \langle u_{20}\rangle\mathrm{Var}(u_{10})\Big)$$

$$- b^2\mathrm{Var}(u_{10})\mathrm{Cov}(u_{20}, u_{10})\Bigg) \exp\left(ib\langle u_{10}\rangle - \frac{1}{2}b^2\mathrm{Var}(u_{10})\right), \qquad (A.8)$$

$$\langle u_{20}^2 \exp(i2bu_{10})\rangle$$

$$= \Big(\mathrm{Cov}(u_{20}, u_{20}^*) + \langle u_{20}\rangle^2 + 4bi\langle u_{20}\rangle\mathrm{Cov}(u_{20}, u_{10}) - 4b^2\mathrm{Cov}(u_{20}, u_{10})^2\Big)$$

$$\times \exp\big(i2b\langle u_{10}\rangle - 2b^2\mathrm{Var}(u_{10})\big). \qquad (A.9)$$

REFERENCES

[1] B.D.O. Anderson and J.B. Moore, *Optimal Filtering*, Prentice-Hall Englewood Cliffs, NJ, 1979.

[2] J.L. Anderson, *An ensemble adjustment Kalman filter for data assimilation*, Mon. Wea. Rev., 129(12), 2884–2903, 2001.

[3] J.L. Anderson, *A local least squares framework for ensemble filtering*, Mon. Wea. Rev., 131(4), 634–642, 2003.

[4] E. Castronovo, J. Harlim and A. Majda, *Mathematical criteria for filtering complex systems: plentiful observations*, J. Comp. Phys., 227(7), 3678–3714, 2008.

[5] C. Chui and G. Chen, *Kalman Filtering*, Springer New York, 1999.

[6] S.E. Cohn and S.F. Parrish, *The behavior of forecast error covariances for a Kalman filter in two dimensions*, Mon. Wea. Rev., 119, 1757–1785, 1991.

[7] R. Daley, *Atmospheric Data Analysis*, Cambridge University Press, 457, 1991.

[8] R. Durrett, *Stochastic Calculus: A Practical Introduction*, CRC-Press LLC, 1996.

[9] P.F. Embid and A.J. Majda, *Low Froude number limiting dynamics for stably stratified flow with small or finite Rossby number*, Geophys. Astrophys. Fluid Dyn., 87, 1–50, 1998.

[10] G. Evensen, *Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics*, J. Geophys. Res., 99, 10143–10162, 1994.

[11] G. Evensen, *Advanced data assimilation for strongly nonlinear dynamics*, Mon. Wea. Rev., 125 1342–1354, 1997.

[12] C.W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry, and Natural Sciences*, Springer-Verlag, New York, 1997.

[13] M. Ghil and P. Malanotte-Rizzoli, *Data assimilation in meteorology and oceanography*, Adv. in Geophys., 33, 141–266, 1991.

[14] J. Harlim and A. Majda, *Mathematical strategies for filtering complex systems: regularly spaced sparse observations*, J. of Comp. Phys, doi:10.1016/j.jcp2008.01.049, 2008.

[15] J. Harlim and A. Majda, *Filtering nonlinear dynamical systems with linear stochastic models*, Nonlinearity, submitted, 2007.

[16] J. Harlim and A. Majda, *Catastrophic filter divergence in filtering nonlinear dissipative systems*, Commun. Math. Sci., to appear, 2008.

[17] R. Klein and A. Majda, *Systematic multiscale models for deep convection on mesoscales*, Theo. Comp. Fluid Dyn., 20, 525–551, 2006.

[18] R. Kubo, *Stochastic Liouville equations*, J. Math. Phys., 4(2), 174–183, 1963.

[19] E.N. Lorenz, *Deterministic non-periodic flow*, J. Atmos. Sci., 20, 130–141, 1963.

[20] E.N. Lorenz, *On the existence of a slow manifold*, J. Atmos. Sci., 43, 1547–1557, 1986.

[21] A. Majda, *Introduction to PDEs and Waves for the Atmosphere and Ocean, Volume 9 of Courant Lecture Notes in Mathematics*, AIMS/CIMS, New York, 2003.

[22] A.J. Majda and M. Grote, *Explicit off-line criteria for stable accurate time filtering of strongly unstable spatially extended systems*, Proc. Nat. Acad. Sci., 104, 1124–1129, 2007.

[23] A.J. Majda and P.R. Kramer, *Simplified models for turbulent diffusion: theory, numerical modelling and physical phenomena*, Phys. Rep., 314(4–5), 237–574, 1999.

[24] A.J. Majda, I. Timofeyev and E. Vanden-Eijnden, *A mathematical framework for stochastic climate models*, Commun. Pure Appl. Math, 54, 891–974, 2001.

[25] A.J. Majda, I. Timofeyev and E. Vanden-Eijnden, *A priori tests of a stochastic mode reduction strategy*, Physica D, 170, 206–252, 2002.

[26] R.N. Miller, E.F. Carter and S.T. Blue, *Data assimilation into nonlinear stochastic models*, Tellus A, 51(2), 167–194, 1999.

[27] L.J. Neef, S.M. Polavarapu and T.G Shepherd, *Four-dimensional data assimilation and balanced dynamics*, J. Atmos. Sci., 63, 1840–1858, 2006.

[28] E. Ott, B. Hunt, I. Szunyogh, A. Zimin, E. Kostelich, M. Corrazza, E. Kalnay and J. Yorke, *A local ensemble Kalman filter for atmospheric data assimilation*, Tellus A, 56, 415–428, 2004.

[29] R. Salmon, *Lectures On Geophysical Fluid Dynamics*, Oxford university Press, 378, 1998.

[30] L.M. Smith, *Numerical study of two-dimensional stratified turbulence. Advances in wave interaction and turbulence*, Contemp. Math., 283, 2001.

[31] L.M. Smith and F. Waleffe, *Generation of slow large scales in forced rotating stratified turbulence*, J. Fluid Mech., 451, 145–168, 2002.

[32] M.L. Waite and P. Bartello, *Stratified turbulence dominated by vortical motion*, J. Fluid Mech., 517, 281–308, 2004.