Filtering Nonlinear Turbulent Dynamical Systems through Conditional Gaussian Statistics

Nan Chen*

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

Andrew J. Majda

Department of Mathematics and Center for Atmosphere Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, USA and Center for Prototype Climate Modeling, NYU Abu Dhabi, Saadiyat Island, Abu Dhabi, UAE.

* *Corresponding author address:* Department of Mathematics and Center for Atmosphere Ocean Science, Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, NY 10012, USA.

E-mail: chennan@cims.nyu.edu

ABSTRACT

In this paper, a conditional Gaussian framework for filtering complex turbulent systems is introduced. Despite the conditional Gaussianity, such systems are nevertheless highly nonlinear and are able to capture the non-Gaussian features of nature. The special structure of the filter allows closed analytical formulae for updating the posterior states and is thus computationally efficient. An information-theoretic framework is developed to assess the model error in the filter estimates. Three types of applications in filtering conditional Gaussian turbulent systems with model error are illustrated. First, dyad models are utilized to illustrate that ignoring the energy-conserving nonlinear interactions in designing filters leads to significant model errors in filtering turbulent signals from nature. Then a triad (noisy Lorenz 63) model is adopted to understand the model error due to noise inflation and underdispersion. It is also utilized as a test model to demonstrate the efficiency of a novel algorithm, which exploits the conditional Gaussian structure, to recover the time-dependent probability density functions associated with the unobserved variables. Furthermore, regarding model parameters as augmented state variables, the filtering framework is applied to the study of parameter estimation with detailed mathematical analysis. A new approach with judicious model error in the equations associated with the augmented state variables is proposed, which greatly enhances the efficiency in estimating model parameters. Other examples of this framework include recovering random compressible flows from noisy Lagrangian tracers, filtering the stochastic skeleton model of the Madden-Julian oscillation (MJO) and initialization of the unobserved variables in predicting the MJO/Monsoon indices.

1. Introduction

Turbulent dynamical systems are ubiquitous in many disciplines of contemporary science and engineering (Hinze and Hinze 1959; Townsend 1980; Frisch 1995; Majda and Wang 2006; Vallis 2006; Salmon 1998). They are characterized by both a large dimensional phase space and a large dimensional space of instability with positive Lyapunov exponents. These linear instabilities are mitigated by energy-conserving nonlinear interactions, yielding physical constraints (Majda and Harlim 2013; Sapsis and Majda 2013b; Majda and Harlim 2012; Harlim et al. 2014), which transfer energy to the linear stable modes where it is dissipated resulting in a statistical steady state. Both understanding complex turbulent systems and improving initializations for prediction require filtering for an accurate estimation of full state variables from noisy partial observations. Since the filtering skill for turbulent signals from nature is often limited by errors due to utilizing an imperfect forecast model, coping with model errors is of wide contemporary interest (Majda and Harlim 2012; Majda 2012).

Many turbulent dynamics are summarized as conditional Gaussian systems (Majda and Harlim 2012; Majda 2003; Kalnay 2003; Majda and Gershgorin 2013; Majda et al. 1999). Despite the conditional Gaussianity, such systems nevertheless can be highly nonlinear and able to capture the non-Gaussian features of nature (Berner and Branstator 2007; Neelin et al. 2010). In this paper, we introduce a general conditional Gaussian framework for continuous-time filtering. The conditional Gaussianity means that once the trajectories of the observational variables are given, the dynamics of the unobserved variables conditioned on these highly nonlinear observed trajectories become Gaussian processes. One of the desirable features of such conditional Gaussian filter is that it allows closed analytical formulae for updating the posterior states associated with the unobserved variables (Liptser and Shiryaev 2001) and is thus computationally efficient.

Recently, the conditional Gaussian nonlinear filter was adopted for filtering the stochastic skeleton model for the Madden-Julian oscillation (MJO) (Chen and Majda 2015a), where equatorial waves and moisture were filtered given the observations of the highly intermittent envelope of convective activity. Another application of this exact and accurate nonlinear filter involves filtering turbulent flow fields utilizing observations from noisy Lagrangian tracer trajectories (Chen et al. 2014c, 2015), where an information barrier was shown as increasing the number of tracers (Chen et al. 2014c) and a multiscale filtering strategy was studied for the system with coupled slow vortical modes and fast gravity waves (Chen et al. 2015). In addition, a family of low-order physics-constrained nonlinear stochastic models with intermittent instability and unobserved variables, which belong to the conditional Gaussian family, was proposed for predicting the MJO and the monsoon indices (Chen et al. 2014b; Chen and Majda 2015d,c). The effective filtering scheme was adopted for the on-line initialization of the unobserved variables that facilitates ensemble prediction algorithm. Other applications that fit into the conditional Gaussian framework includes the cheap exactly solvable forecast models in dynamic stochastic superresolution of sparsely observed turbulent systems (Branicki and Majda 2013; Keating et al. 2012), stochastic superparameterization for geophysical turbulence (Majda and Grooms 2014), and blended particle filters for large-dimensional chaotic systems (Majda et al. 2014; Qi and Majda 2015) that captures non-Gaussian features in an adaptively evolving low-dimensional subspace through particles interacting with conditional Gaussian statistics on the remaining phase space.

In this paper, we illustrate three types of applications of the conditional Gaussian filtering framework, where the effect of model error is extensively studied. In addition to the traditional path-wise measures, an information-theoretic framework (Branicki and Majda 2014; Branicki et al. 2013; Majda and Branicki 2012; Majda and Wang 2006) is adopted to assess the lack of information and model error in filtering these turbulent systems.

The first application involves utilizing dyad models (Majda and Lee 2014; Majda 2015) to study the effect of model error due to the ignorance of energy-conserving nonlinear interactions in forecast models in filtering turbulent signals from nature. Such model error exists in many ad-hoc quadratic multi-level regression models (Kravtsov et al. 2005; Kondrashov et al. 2005; Wikle and Hooten 2010; Cressie and Wikle 2011) that are utilized as data-driven statistical models for time series of partial observations of nature. However, these models were shown to suffer from finite-time blow up of statistical solutions (Majda and Yuan 2012; Majda and Harlim 2013). To understand the effect of such model error in filtering, a physics-constrained dyad model (Majda and Lee 2014) is adopted to generate the turbulent signals of nature while a stochastic parameterized model without energy-conserving nonlinearities (Majda and Harlim 2012) is adopted as the imperfect filter. The skill of this stochastic parameterized filter is studied in different dynamical regimes and the lack of information in the filter estimates is compared with that using the perfect filter. Meanwhile, the role of observability (Gajic and Lelic 1996; Majda and Harlim 2012) is explored and its necessity in filtering turbulent systems is emphasized.

The second application of the conditional Gaussian framework is to filter a family of triad models (Majda 2003; Majda et al. 1999, 2001, 2002b), which include the noisy Lorenz 63 (L-63) model (Lorenz 1963). The goal here is to explore the effect of model error due to noise inflation and underdispersion in designing filters. The motivation of studying such kind of model error comes from the fact that many models for turbulence are underdispersed since they have too much dissipation (Palmer 2001) due to inadequate resolution and deterministic parameterization of unresolved features. On the other hand, suitably inflating the noise in imperfect forecast models are widely adopted to reduce the lack of information (Anderson 2001; Kalnay 2003; Majda and Harlim 2012) and also to suppress catastrophic filter divergence (Harlim and Majda 2010a; Tong et al. 2015). Besides filtering a single trajectory, recovering the full probability density function (PDF) associated with the unobserved variables given an ensemble of observational trajectories is also of particular interest. Combining the ensembles of the analytically solvable conditional Gaussian distribution associated with filtering each unobserved single trajectory, an effective conditional Gaussian ensemble mixture approach is proposed to approximate the time-dependent PDF associated with the unobserved variables. In this fashion, an efficient algorithm can be generated for systems with a large number of the unobserved variables, compared with applying a direct Monte Carlo method which is extremely slow and expensive due to the "curse of dimensionality" (Majda and Harlim 2012; Daum and Huang 2003).

Parameter estimation in turbulent systems is an important issue and this is the third topic within the conditional Gaussian filtering framework of this paper. Regarding the model parameters as augmented state variables, algorithms based on particle or ensemble Kalman filters were designed for parameter estimation (Dee 1995; Smedstad and O'Brien 1991; Van Der Merwe et al. 2001; Plett 2004; Wenzel et al. 2006; Campillo and Rossi 2009; Harlim et al. 2014; Salamon and Feyen 2009). Although many successful results utilizing these algorithms were obtained, very little mathematical analysis was provided for exploring the convergence rate and understanding the potential limitation of such algorithms. Guidelines for enhancing the efficiency of the algorithms are desirable since a short training period is preferred in many real-world applications. In the conditional Gaussian framework, the closed analytic form of the posterior state estimations facilitates the analysis of both the error and the uncertainty in the estimated parameters for a wide family of models, where detailed mathematical justifications are accessible. Here, focus is on the parameter estimation skill dependence on different factors of the model as well as the observability. In some applications that certain prior information of the parameters is available (Yeh 1986; Iglesias et al. 2014). Yet, none of the existing filtering-based parameter estimation approaches emphasizes exploiting such prior information in improving the algorithms. In this paper, stochastic parameterized equations (Majda and Harlim 2012), involving the prior knowledge of the parameters, are incorporated into the filtering algorithm as the underlying processes of the augmented state variables. This improved algorithm greatly enhances the convergence rate at the cost of only introducing a small model error and it is particularly useful when the system loses practical observability.

The remainder of this paper is as follows. The general framework of the conditional Gaussian nonlinear systems is introduced in Section 2. In Section 3, an information-theoretic framework for assessing the model error in filtering is proposed. The information measures compensate the insufficiency of the path-wise ones in measuring the lack of information in the filtered solutions. Section 4 deals with dyad models, where focus is on model error in filtering due to the lack of respecting the underlying physical dynamics of the partially observed system. In Section 5, a general family of triad model is proposed and the noisy L-63 model is adopted as a test model for understanding the model error in noise inflation and underdispersion. In the same section, the conditional Gaussian ensemble mixture for approximating the PDF associated with unobserved variables is introduced and the model error in filtering the PDF utilizing imperfect models is studied. Section 6 involves parameter estimation, where the skill of estimating both additive and multiplicative parameters is illustrated with detailed mathematical analysis. The comparison of utilizing direct method and stochastic parameterized equations approach is shown for estimating parameters in both linear and nonlinear systems. Summary conclusions are included in Section 7.

2. Conditional Gaussian nonlinear systems

The conditional Gaussian systems have the following abstract form,

$$d\mathbf{u_I} = [\mathbf{A}_0(t,\mathbf{u_I}) + \mathbf{A}_1(t,\mathbf{u_I})\mathbf{u_{II}}]dt + \boldsymbol{\Sigma}_\mathbf{I}(t,\mathbf{u_I})d\mathbf{W_I}(t), \tag{1a}$$

$$d\mathbf{u_{II}} = [\mathbf{a}_0(t,\mathbf{u_I}) + \mathbf{a}_1(t,\mathbf{u_I})\mathbf{u_{II}}]dt + \boldsymbol{\Sigma}_\mathbf{II}(t,\mathbf{u_I})d\mathbf{W_{II}}(t), \tag{1b}$$

where $\mathbf{u_I}(t)$ and $\mathbf{u_{II}}(t)$ are vector state variables, $\mathbf{A}_0, \mathbf{A}_1, \mathbf{a}_0, \mathbf{a}_1, \boldsymbol{\Sigma}_\mathbf{I}$ and $\boldsymbol{\Sigma}_\mathbf{II}$ are vectors and matrices that depend only on time t and state variables $\mathbf{u_I}$, and $\mathbf{W_I}(t)$ and $\mathbf{W_{II}}(t)$ are independent Wiener processes. Once $\mathbf{u_I}(s)$ for $s \leq t$ is given, $\mathbf{u_{II}}(t)$ conditioned on $\mathbf{u_I}(s)$ becomes a Gaussian process with mean $\bar{\mathbf{u}}_\mathbf{II}(t)$ and covariance $\mathbf{R_{II}}(t)$, i.e.,

$$p\big(\mathbf{u_{II}}(t)|\mathbf{u_I}(s \leq t)\big) \sim \mathcal{N}(\bar{\mathbf{u}}_\mathbf{II}(t), \mathbf{R_{II}}(t)). \tag{2}$$

Despite the conditional Gaussianity, the coupled system (1) remains highly nonlinear and is able to capture the non-Gaussian features such as skewed or fat-tailed distributions as observed in nature (Berner and Branstator 2007; Neelin et al. 2010).

One of the desirable features of the conditional Gaussian system (1) is that the conditional distribution in (2) has the following closed analytic form (Liptser and Shiryaev 2001),

$$\begin{aligned}
d\bar{\mathbf{u}}_\mathbf{II}(t) =& [\mathbf{a}_0(t,\mathbf{u_I}) + \mathbf{a}_1(t,\mathbf{u_I})\bar{\mathbf{u}}_\mathbf{II}]dt + (\mathbf{R_{II}}\mathbf{A}_1^*(t,\mathbf{u_I}))(\boldsymbol{\Sigma}_\mathbf{I}\boldsymbol{\Sigma}_\mathbf{I}^*)^{-1}(t,\mathbf{u_I}) \times \\
& \qquad [d\mathbf{u_I} - (\mathbf{A}_0(t,\mathbf{u_I}) + \mathbf{A}_1(t,\mathbf{u_I})\bar{\mathbf{u}}_\mathbf{II})dt], \\
d\mathbf{R_{II}}(t) =& \{\mathbf{a}_1(t,\mathbf{u_I})\mathbf{R_{II}} + \mathbf{R_{II}}\mathbf{a}_1^*(t,\mathbf{u_I}) + (\boldsymbol{\Sigma}_\mathbf{II}\boldsymbol{\Sigma}_\mathbf{II}^*)(t,\mathbf{u_I}) \\
& \qquad -(\mathbf{R_{II}}\mathbf{A}_1^*(t,\mathbf{u_I}))(\boldsymbol{\Sigma}_\mathbf{I}\boldsymbol{\Sigma}_\mathbf{I}^*)^{-1}(t,\mathbf{u_I})(\mathbf{R_{II}}\mathbf{A}_1^*(t,\mathbf{u_I}))^*\}\, dt.
\end{aligned} \tag{3}$$

The exact and accurate solutions in (3) provide a general framework for studying continuous-time filtering and uncertainty quantification of the conditional Gaussian system (1). In filtering the turbulent system (1), if $\mathbf{u_I}(s \leq t)$ is the observed process, then the posterior states of the unobserved process $\mathbf{u_{II}}(t)$ in (2) are updated following the analytic formulae (3) associated with the nonlinear filter (1).

3. An information-theoretic framework for assessing the model error

Assume $\mathbf{u}_t$ is the true signal and $\mathbf{u}_t^{filter}$ is the filtered solution. The traditional measures for assessing the filtering skill in the i-th dimension of $\mathbf{u}_t$ and $\mathbf{u}_t^{filter}$ are the root-mean-square (RMS) error and anomaly pattern correlation (Hyndman and Koehler 2006; Kalnay 2003; Majda and Harlim 2012),

$$\begin{aligned}\text{RMS error} &= \sqrt{\mathbb{E}\left[(u_t - u_t^{filter})^2\right]},\\ \text{Pattern Correlation} &= \frac{\mathbb{E}\Big[\big(u_t - \mathbb{E}[u_t]\big)\big(u_t^{filter} - \mathbb{E}[u_t^{filter}]\big)\Big]}{\sqrt{\mathbb{E}\Big[\big(u_t - \mathbb{E}[u_t]\big)^2\Big]\cdot\mathbb{E}\Big[\big(u_t^{filter} - \mathbb{E}[u_t^{filter}]\big)^2\Big]}},\end{aligned} \tag{4}$$

where u_t and u_t^{filter} represent the i-th dimension of the vector fields $\mathbf{u}_t$ and $\mathbf{u}_t^{filter}$, respectively.

Despite their wide applications in assessing filtering and prediction skill, these path-wise measures fail to assess the lack of information in the filter estimates and the predicted states (Branicki and Majda 2014; Chen and Majda 2015d). As shown in (Chen and Majda 2015d), two predicted trajectories with completely different amplitudes can have the same RMS error and anomaly pattern correlation. Undoubtedly, the solution having comparable amplitude as the truth is more skillful than the one with strongly underestimated amplitude, which misses all the extreme events (Majda et al. 2010b; Majda and Harlim 2012; Majda and Branicki 2012) that are important for the turbulent systems. Different from the indistinguishable skill utilizing the path-wise measurements, an information-theoretic framework including the measurement of the lack of information succeeds in discriminating the prediction skill of the two solutions.

In (Branicki and Majda 2014), a systematic information-theoretic approach was developed to quantify the statistical accuracy of Kalman filters with model error and the optimality of the imperfect Kalman filters in terms of three information measures was presented. Another application of information theory is illustrated in (Branicki and Majda 2015) for improving imperfect predictions via multi-model ensemble forecasts.

Following the general information-theoretic framework in (Branicki and Majda 2014; Chen and Majda 2015d), we consider three information measures:

The Shannon entropy $S(\mathbf{U}_t)$ of the residual $\mathbf{U}_t = \mathbf{u}_t - \mathbf{u}_t^{filter}$ is given by (Majda and Wang 2006; Abramov and Majda 2004)

$$\mathscr{S}(\mathbf{U}_t) := -\int p(\mathbf{U}_t) \ln p(\mathbf{U}_t) d\mathbf{U}_t. \tag{5}$$

The relative entropy $\mathscr{P}(\pi, \pi^{filter})$ of the PDF π^{filter} associated with $\mathbf{u}_t^{filter}$ compared with the truth π is given by (Majda et al. 2005; Majda and Wang 2006; Majda and Branicki 2012; Majda et al. 2002a),

$$\mathscr{P}(\pi, \pi^{filter}) := \int \pi(\mathbf{u}) \ln \frac{\pi(\mathbf{u})}{\pi^{filter}(\mathbf{u})} d\mathbf{u}. \tag{6}$$

The mutual information $\mathscr{M}(\mathbf{u}_t, \mathbf{u}_t^{filter})$ between the true signal $\mathbf{u}_t$ and the filtered one $\mathbf{u}_t^{filter}$ is given by the symmetric formula (MacKay 2003; Branicki and Majda 2014),

$$\mathscr{M}(\mathbf{u}_t, \mathbf{u}_t^{filter}) := \int\int p(\mathbf{u}_t, \mathbf{u}_t^{filter}) \ln \frac{p(\mathbf{u}_t, \mathbf{u}_t^{filter})}{\pi(\mathbf{u}_t)\pi^{filter}(\mathbf{u}_t^{filter})} d\mathbf{u}_t d\mathbf{u}_t^{filter}. \tag{7}$$

Each one of the three measures provides different information about the filtering skill. The mutual information $\mathscr{M}(\mathbf{u}_t, \mathbf{u}_t^{filter})$ measures the dependence between $\mathbf{u}_t$ and $\mathbf{u}_t^{filter}$. The Shannon entropy of the residual $\mathscr{S}(\mathbf{U}_t)$ measures the uncertainty in the filtered solution $\mathbf{u}_t^{filter}$ compared with the truth $\mathbf{u}_t$. These two information measures are the surrogates for the anomaly pattern correlation and RMS error in the path-wise sense, respectively (Branicki and Majda 2014). Particularly, if both the truth $\mathbf{u}_t$ and the filtered solution $\mathbf{u}_t^{filter}$ are Gaussian distributed, then the asymptotic anomaly pattern correlation and RMS error can be expressed in analytic forms by the mutual information and the Shannon's entropy. The relative entropy $\mathscr{P}(\pi, \pi^{filter})$ quantifies the lack of information in the statistics of the filtered solution $\mathbf{u}_t^{filter}$ relative to that of the truth $\mathbf{u}_t$ (Majda and Gershgorin 2010; Majda and Branicki 2012). Therefore, it is an indicator of assessing the disparity in the amplitudes and spread between $\mathbf{u}_t^{filter}$ and $\mathbf{u}_t$. Importantly, the relative entropy is able to quantify the ability of capturing the extreme events (Chen et al. 2014b; Chen and Majda 2015d; Branicki and Majda 2014), corresponding to the tails of a distribution, in the filtered solutions. The relative entropy is often interpreted as a 'distance' between the two probability densities but it is not a true metric. It is non-negative with $\mathscr{P} = 0$ only when $\pi = \pi^{filter}$ and it is invariant under nonlinear changes of variables.

Due to the importance of measuring the lack of information in the filtered solutions, the relative entropy is included in assessing the filtering skill throughout this work. Along with the relative entropy, we nevertheless show the anomaly pattern correlation and the RMS error instead of the mutual information and the Shannon's entropy since the readers are more familiar with these traditional path-wise measures. Yet, it is important to bear in mind that the mutual information and the Shannon's entropy are the surrogates of the path-wise measures in the information-theoretic framework.

In the study of filtering the unobserved single trajectories in Section 4 and 5a, the posterior mean estimation is chosen as the filter estimate $\mathbf{u}_t^{filter}$. Both the path-wise filtering skill in $\mathbf{u}_t^{filter}$ using (4) and the lack of information in the time-averaged PDF of $\mathbf{u}_t^{filter}$ related to that of the truth $\mathbf{u}_t$ via the relative entropy (6) are assessed. In measuring the lack of information in the recovered time-dependent PDF in Section 5b and 5c, the relative entropy (6) in the recovered PDF π^{filter} related to the truth π at each time instant is computed, where π^{filter} is obtained from conditional Gaussian ensemble mixture approach.

It is worthwhile remarking that although most of the focus of this paper is on assessing the lack of information in the path-wise sense, the conditional Gaussian framework (1)–(3) also provides a general framework for quantifying the uncertainty using imperfect models in ensemble prediction. Assume the joint distributions regarding $\mathbf{u_I}$ and $\mathbf{u_{II}}$ in (1) for perfect and imperfect models are given by

$$p(\mathbf{u_I},\mathbf{u_{II}}) = p(\mathbf{u_{II}}|\mathbf{u_I})\pi(\mathbf{u_I}), \qquad p^M(\mathbf{u_I},\mathbf{u_{II}}) = p^M_L(\mathbf{u_{II}}|\mathbf{u_I})\pi^M(\mathbf{u_I}),$$

where due to the incomplete knowledge or the coarse-grained effect the distribution p^M associated with the imperfect model is assumed to be formed only by the conditional moments up to L. According to (Branicki et al. 2013), the lack of information in the imperfect model related to the perfect one is given by

$$\mathscr{P}(p(\mathbf{u_I},\mathbf{u_{II}}), p^M_L(\mathbf{u_I},\mathbf{u_{II}})) = \mathscr{P}(p(\mathbf{u_I},\mathbf{u_{II}}), p_L(\mathbf{u_I},\mathbf{u_{II}})) + \mathscr{P}(p_L(\mathbf{u_I},\mathbf{u_{II}}), p^M_L(\mathbf{u_I},\mathbf{u_{II}})), \qquad (8)$$

where p_L is the PDF reconstructed using the L moments of the perfect model. The first term on the right hand side of (8) is called the intrinsic barrier, which measures the lack of information in the perfect model due to the coarse-grained effect from the insufficient measurement, and the second term is the model error using the imperfect model. Direct calculation shows that

$$\text{Intrinsic barrier} = \int \pi(\mathbf{u_I})\left(\mathscr{S}(p_L(\mathbf{u_{II}})) - \mathscr{S}(p(\mathbf{u_{II}}))\right), \qquad (9)$$

$$\text{Model error} = \mathscr{P}(\pi(\mathbf{u_I}), \pi^M(\mathbf{u_I})) + \int \pi^M(\mathbf{u_I})\mathscr{P}(p_L(\mathbf{u_{II}}|\mathbf{u_I}), p^M_L(\mathbf{u_{II}}|\mathbf{u_I}))d\mathbf{u_I}. \qquad (10)$$

In the conditional Gaussian framework, $L=2$ and the relative entropy for the conditional Gaussian distributions in (10) is assessed in light of the closed analytic formulae (3) for both the distributions. Note that in filtering complex turbulent systems, if the observations in the imperfect filter $\pi^M(\mathbf{u_I})$ are assumed to be the same as $\pi(\mathbf{u_I})$ in the perfect filter, then the lack of information in the imperfect filter related to the perfect one is simply assessed by

$$\mathscr{E}(t) = \mathscr{P}(p_L(\mathbf{u_{II}}(t)|\mathbf{u_I}(s)), p^M_L(\mathbf{u_{II}}(t)|\mathbf{u_I}(s))), \qquad 0 \le s \le t. \qquad (11)$$

The information measurement in (11) provides a guideline in designing practical imperfect filters. An example of applying (11) to assess the information model error in different imperfect filters is shown in (Chen and Majda 2015b) for filtering a turbulent flow field using noisy Lagrangian tracers.

4. Dyad models

Many turbulent dynamical systems involve dyad and triad interactions (Majda 2015; Majda and Lee 2014; Majda et al. 2009). These nontrivial nonlinear interactions between large-scale mean flow and turbulent fluctuations generate intermittent instability while the total energy from the nonlinear interactions is conserved. In this and the next sections, we study the filtering skill of dyad and triad models, where the effect of different model errors is explored.

In this section, we utilize dyad models to understand the effect of model error due to the ignorance of energy-conserving nonlinear interactions in forecast models in filtering turbulent signals from nature. As discussed in Section 1, such model error exists in many ad-hoc quadratic multi-level regression models (Kravtsov et al. 2005; Kondrashov et al. 2005; Wikle and Hooten 2010; Cressie and Wikle 2011) for fitting and predicting time series of partial observations of nature, which were shown to suffer from finite-time blow up of statistical solutions and also have pathological behavior of the related invariant measure (Majda and Yuan 2012; Majda and Harlim 2013). Recently, a new class of physics-constrained nonlinear regression models were developed (Majda and Harlim 2013) and the application of these physics-constrained models in ensemble Kalman filtering is shown in (Harlim et al. 2014) together with other recent applications to prediction (Chen et al. 2014b; Chen and Majda 2015c,d).

The general form of the dyad models is described in (Majda 2015; Majda and Lee 2014). Here we focus on the following dyad model,

$$du = (-d_{uu}u + \gamma uv + F_u)dt + \sigma_u dW_u, \tag{12a}$$

$$dv = (-d_{vv}v - \gamma u^2)dt + \sigma_v dW_v. \tag{12b}$$

In (12), u is regarded as representing one of the resolved modes in a turbulent signal, which interacts with the unresolved mode v through quadratic nonlinearities. The conserved energy in the quadratic nonlinear terms in (12) is seen by

$$(u,v)\cdot\begin{pmatrix}\gamma u v\\ -\gamma u^2\end{pmatrix} = 0.$$

Below, the physics-constrained dyad model (12) is utilized to generate true signals of nature. The goal here is to filter the unobserved process v given one single realization of the observed process u. In addition to adopting the perfect filter (12), an imperfect filter with no energy-conserving nonlinear interactions is studied for comparison. In this imperfect filter, the nonlinear feedback $-\gamma u^2$ in v is dropped and the result is a stochastic parameterized filter (Majda and Harlim 2012),

$$du = (-d_{uu}u + \gamma uv + F_u)dt + \sigma_u dW_u, \tag{13a}$$

$$dv = -d_{vv}^M(v - \bar{v}^M)dt + \sigma_v^M dW_v. \tag{13b}$$

In the stochastic parameterized filter (13), the parameters in the resolved variable u are assumed to be the same as nature (12). We further assume the statistics of the unobserved variable v of nature (12) are available. Thus, the parameters $d_{vv}^M, \bar{v}^M$ and σ_v^M in the unresolved process v are calibrated (Harlim and Majda 2008, 2010b; Branicki et al. 2013) by matching the mean, variance and decorrelation time of those in (12). Note that both (12) and (13) belong to the conditional Gaussian framework (1) by denoting $\mathbf{u_I} = u$ and $\mathbf{u_{II}} = v$.

One important issue in filtering is observability (Gajic and Lelic 1996; Majda and Harlim 2012). The coupled system (12) is said to lose its observability if the observed process u provides no information in determining the unobserved variable v. Intuitively, this corresponds to $u = 0$ in (12), in which case v disappears in the observed process u. The rigorous definition of the observability is included in Appendix A. To understand the role of observability in filtering, we consider the following two dynamics regimes,

$$\begin{aligned} &\text{(A)} \qquad d_{uu} = 1, \qquad d_{vv} = 1, \qquad \gamma = 1.5, \qquad \text{and} \qquad F_u = 1. \\ &\text{(B)} \qquad d_{uu} = 1, \qquad d_{vv} = 1, \qquad \gamma = 1.5, \qquad \text{and} \qquad F_u = 0. \end{aligned} \tag{14}$$

The fixed point associated with the deterministic part of (12) is given respectively by

$$\begin{aligned} &\text{(A)} \qquad u_c = 0.5741, \qquad v_c = -0.4945, \\ &\text{(B)} \qquad u_c = 0, \qquad\qquad v_c = 0. \end{aligned}$$

It is clear that in dynamical regime (B) the system (12) loses practical observability when the solution is around the fixed point. As shown in Figure 2, both models are able to generate intermittency with suitable choices of the observational noise σ_u and the system noise σ_v.

Below, the true signals are generated from the dyad model (12) with different observational noise σ_u and system noise σ_v. The filtering skill scores utilizing both the physics-constrained perfect filter (12) and the stochastic parameterized imperfect filter (13) are shown in Figure 1. The first two rows show the RMS error and pattern correlation in the posterior mean estimation of v and the third row illustrates the information model error $\mathscr{P}(\pi, \pi^{filter})$ in the time-averaged PDF of the posterior mean estimation π^{filter} related to that of the truth π. Here, if the model error is larger than $\mathscr{P} = 5$, which is already significant, then the same color as $\mathscr{P} = 5$ is utilized for representation in Figure 1.

The skill scores in dynamical regime (A) are shown in column (a) and (b) of Figure 1. The physics-constrained perfect filter (12) has high filtering skill when $\sigma_u/\sigma_v \ll 1$ and $\sigma_u/\sigma_v \geq 3$.

As contrast, the stochastic parameterized filter (13) is skillful only when $\sigma_u/\sigma_v \ll 1$, in which case the filter estimation of v is mostly determined from the observation process u with small observational noise and therefore the two filters are expected to have comparable high skill when the system has observability. Panel (a) of Figure 2 compares the posterior mean estimations across time with $\sigma_u = 0.2$ and $\sigma_v = 2$. Clearly, both filters succeed in filtering v provided that the practical observability is satisfied, i.e. u not approaching zero. On the other hand, as shown in panel (b) of Figure 2 with $\sigma_u = 2$ and $\sigma_v = 0.2$, the energy-conserving perfect filter (12) filters v almost perfectly while the stochastic parameterized filter (13) has no skill. In fact, $\sigma_v \ll \sigma_u$ implies that the filter trusts more towards the dynamics of v and the amplitude of energy feedback $-\gamma u^2$ is much larger than the stochastic forcing in v. Thus, the process of v is largely driven by the nonlinear energy feedback $-\gamma u^2$ in (12b). However, the stochastic parameterized filter (13) has no such mechanism and therefore the posterior mean estimation is simply around the maximum likelihood state of v, i.e., the mean $\bar{v}^M$. Importantly, without the nonlinear energy feedback term $-\gamma u^2$, the information model error $\mathscr{P}(\pi, \pi^{filter})$ utilizing the imperfect filter (13) remains huge unless $\sigma_u/\sigma_v \ll 1$.

Next, we study the filtering skill in dynamical regime (B), at the fixed point of which the system has no observability. Comparing with regime (A), significant deterioration of the filtering skill is found when $\sigma_u \ll \sigma_v \leq 1$ in the truth, where the trajectory of u is around the fixed point $u_c = 0$ implying no practical observability. See panel (a) of Figure 3. With the increase of σ_v, more positive values of v are reached, which correspond to the increase of intermittent phases of u with large bursts. Clearly, the observability is regained at these intermittent phases and thus an improved skill in filtering is found. See panel (b) of Figure 3.

To conclude, the energy-conserving nonlinear feedback plays a significant role in filtering the dyad model (12), especially with large observational noise σ_u. Despite comparable RMS errors, the imperfect stochastic parameterized filter (13) without energy-conserving nonlinearities leads to a much larger information model error $\mathscr{P}(\pi, \pi^{filter})$ than the energy-conserving perfect filter (12) for $\sigma_u/\sigma_v \not\ll 1$. In addition, the observability becomes quite important when the system noise σ_v is moderate and the observational noise σ_u is small. An increase of σ_v enhances the intermittency that improves the filtering skill.

5. Triad models

The nonlinear coupling in triad systems is generic of nonlinear coupling between any three modes in larger systems with quadratic nonlinearities. Here, we introduce the general form of the triad models that belongs to the conditional Gaussian framework (1),

$$\begin{aligned} du_I &= (L_{11}u_I + L_{12}u_{II} + F_1)\,dt + \sigma_1\,dW_I, \\ d\vec{u}_{II} &= (L_{22}\vec{u}_{II} + L_{21}u_I + \Omega\vec{u}_{II} + F_2)\,dt + \sigma_2\,dW_{II}, \end{aligned} \tag{15}$$

where $u_I = u_1$ and $u_{II} = (u_2, u_3)^T$ and the coefficients $L_{11}, L_{12}, L_{21}, L_{22}$ and Ω are functions of only the observed variable. In (15), either u_I or u_{II} can be regarded as the observed variable and correspondingly the other one becomes the unresolved variable that requires filtering. The triad model (15) has wide applications in atmosphere and ocean science. One example is the stochastic mode reduction model (also known as MTV model) (Majda et al. 2003, 1999, 2002b, 2001), which includes both a wave-mean flow triad model and a climate scattering triad model for barotropic equations (Majda et al. 2001). Another example of (15) involves the slow-fast waves in the coupled atmosphere-ocean system (Majda and Harlim 2012), where one slow vortical mode interacts with two fast gravity modes with the same Fourier wavenumber.

With the following choice of the matrices and vectors in (15),

$$u_I = x, \qquad u_{II} = (y,z)^T, \qquad L_{11} = -\sigma, \qquad L_{12} = (\sigma, 0), \qquad L_{21} = (\rho x, 0)^T, \qquad \sigma_1 = \sigma_x,$$

$$L_{22} = \begin{pmatrix} -1 & \\ & -\beta \end{pmatrix}, \qquad \Omega = \begin{pmatrix} 0 & -x \\ x & 0 \end{pmatrix}, \qquad \sigma_2 = \begin{pmatrix} \sigma_y & \\ & \sigma_z \end{pmatrix},$$

the triad model (15) becomes the noisy Lorenz 63 (L-63) model (Lorenz 1963),

$$\begin{aligned} dx &= \sigma(y-x)dt + \sigma_x dW_x, \\ dy &= \big(x(\rho - z) - y\big)dt + \sigma_y dW_y, \\ dz &= (xy - \beta z)dt + \sigma_z dW_z. \end{aligned} \tag{16}$$

As is known, adopting the following parameters

$$\rho = 28, \qquad \sigma = 10, \qquad \beta = 8/3, \tag{17}$$

the deterministic version of (16) has chaotic solutions, where the trajectory of the system has a butterfly profile at the attractor. Such a feature is preserved in the appearance of small or moderate noise in (16). See Figure 4 for the trajectories of (16) with $\sigma_x = \sigma_y = \sigma_z = 0, 5$ and 10. Note that the noisy L-63 model possesses the property of energy-conserving nonlinear interactions.

The noisy L-63 model (16) equipped with the parameters (17) is utilized as a test model in this section. Below, we first study filtering the unresolved trajectories given one realization of the noisy observations. Then an efficient conditional Gaussian ensemble mixture approach is designed to approximate the time-dependent PDF associated with the unresolved variables, which requires only a small ensemble of the observational trajectories. In both studies, the effect of model error due to noise inflation and underdispersion is studied. The underdispersion occurs in many models for turbulence since they have too much dissipation (Palmer 2001) due to inadequate resolution and deterministic parameterization of unresolved features while noise inflation is adopted in many imperfect forecast models to reduce the lack of information (Anderson 2001; Kalnay 2003; Majda and Harlim 2012) and suppress the catastrophic filter divergence (Harlim and Majda 2010a; Tong et al. 2015).

a. Model error in filtering the unresolved processes

We explore filtering the unresolved single trajectories in L-63 model utilizing imperfect filters, where model error comes from the observational and system noise σ_x, σ_y and σ_z. Here, the noisy L-63 model (16) is adopted to generate true signals. The model utilized for filtering differs from (16) by the noise amplitudes

$$\begin{aligned} dx &= \sigma(y-x)dt + \sigma_x^M dW_x, \\ dy &= \big(x(\rho - z) - y\big)dt + \sigma_y^M dW_y, \\ dz &= (xy - \beta z)dt + \sigma_z^M dW_z. \end{aligned} \tag{18}$$

Note that although the system noise in (18) can be arbitrary, the observational noise amplitude must be non-zero to avoid the singularity in solving the posterior estimations (3).

1) Filtering the deterministic L-63 system utilizing the imperfect forecast model with noise

The first test involves the situation that the true signal is generated from the L-63 model which has no stochastic noise, i.e, $\sigma_x = \sigma_y = \sigma_z = 0$ in (16),

$$\begin{aligned} dx &= \sigma(y-x)dt, \\ dy &= \big(x(\rho - z) - y\big)dt, \\ dz &= (xy - \beta z)dt. \end{aligned} \tag{19}$$

The filtering skill utilizing the imperfect model (18) with nonzero noise is studied. This demonstrates the role of noise inflation in the forecast model. In the situation of filtering x with observations from y and z, we have the following results.

Proposition 1 *Assume the true signal is generated from the system* (19)*. In the situation of filtering x with observations from y and z, the posterior variance R_t of x_t and the error in the posterior mean $\|\mu_t - x_t\|^2$ utilizing the forecast model* (18) *with nonzero σ_y^M and σ_z^M, are bounded by*

$$R_t \leq R_0 e^{-2\sigma t} + \left(\sigma_x^M\right)^2 \frac{1-e^{-2\sigma t}}{2\sigma}, \tag{20a}$$

$$\|\mu_t - x_t\|^2 \leq \|\mu_0 - x_0\|^2 e^{-\sigma t}, \tag{20b}$$

where μ_0 and R_0 are the mean and uncertainty of variable x at initial time.

The detailed derivations of Proposition 1 is shown in Appendix B. The results in (20) imply that the error in the posterior mean estimation decays to zero in an exponentially fast rate regardless of the noise level σ_x^M, σ_y^M and σ_z^M in the imperfect filter (18). The uncertainty after the initial period is essentially bounded by the system noise variance $(\sigma_x^M)^2$ over the known parameter (2σ). This indicates if the system noise σ_x^M is zero in the imperfect forecast model (18), then the posterior estimation will converge to the truth with an uncertainty that decays exponentially to zero. Panel (a) in Figure 5 validates Proposition 1, where the statistics are averaged across time $t \in [5, 50]$. The nearly zero RMS error and nearly one pattern correlation reveal that the posterior mean converges to the truth. The posterior variance increases as the observational noise σ_x^M increases.

The qualitative conclusions are the same in the situation of observing x and filtering y and z. See panel (b) in Figure 5. The uncertainty in filtering z is larger than that in filtering y, since y is directly related in the observational process x in (18). The trajectories as a function of time are shown in panel (c) and (d), with nonzero and zero system noise $\sigma_y^M = \sigma_z^M$, respectively. In both cases the posterior mean converges to the truth. If the system noise is nonzero, then the posterior variance for both y and z remains nonzero but is bounded.

These results indicate that noise inflation in the imperfect forecast model brings no error regarding the posterior mean estimation and a bounded posterior uncertainty after a short relaxation time provided that the signal is generated from the system with no stochastic noise.

2) Filtering the noisy L-63 system utilizing the imperfect forecast model with no system noise

Next, we reverse the setup in the previous subsection. We assume the true signal is generated from the noisy L-63 model (16) but the imperfect forecast model (18) contains no system noise. This illustrates the effect of utilizing an underdispersive imperfect forecast model in filtering. Note that the observational noise in (18) must be nonzero to avoid the singularity in solving the posterior states in (3). Since the two situations that observing either x or y and z lead to qualitatively the same results, we focus on the situation that only x is observed. Thus, the imperfect filter has the following form,

$$
\begin{aligned}
dx &= \sigma(y-x)dt + \sigma_x^M dW_x, \\
dy &= \big(x(\rho - z) - y\big)dt, \\
dz &= (xy - \beta z)dt.
\end{aligned}
\tag{21}
$$

Below, we assume the observational noise σ_x^M in (21) is the same as σ_x in the model (16) that generates the true signal. Then model error in filtering simply comes from the ignorance of the system noise σ_y and σ_z in (16). Column (a)-(c) of Figure 6 show the dependence of the statistics on the system noise σ_y and σ_z in (16), where we set $\sigma_y = \sigma_z$ for simplicity. Clearly, with the increase of σ_y and σ_z, the filtering skill regarding the RMS error and the pattern correlation in the posterior mean of both y and z becomes worse while these posterior states are quite certain, both of which indicate the negative effect of underdispersion in the imperfect forecast model. In addition, the model error $\mathscr{P}(\pi, \pi^{filter})$ (6) increases as a function of the system noise σ_y and σ_z in (16) and is larger in filtering variable z than y. The comparable statistics in the three columns of Figure 6 reveal that increasing the observational noise σ_x in the true model (16) has little effect in filtering the unresolved variables provided that the observational noise σ_x^M in the imperfect forecast model (21) equals σ_x.

In column (d) of Figure 6, we show the trajectories with $\sigma_x = 1$ and $\sigma_y = \sigma_z = 10$. Therefore a severe underdispersion exists in the imperfect forecast model (21). A larger skewness is found in time-averaged PDF of the filter estimation of z than that of the truth, which explains the model error.

3) Filtering the noisy L-63 system utilizing the imperfect L-63 forecast model with different noise amplitudes

Finally, we study the general situation that both the system that generates the true signal (16) and the imperfect filter (18) contain non-zero noise. Again, we illustrate the situation with observing x and filtering y and z. The other case has the same qualitative results.

Figure 7 shows the filtering skill utilizing the imperfect filter (18), where the model error comes from either the observational noise σ_x^M or system noise σ_y^M, σ_z^M and the noise levels in the truth σ_x, σ_y and σ_z are set to equal with each other. In Column (a), (c) and (e), the noise level in the true dynamics (16) is gradually increased $\sigma_x = \sigma_y = \sigma_z = 1, 5$ and 10. In the imperfect filter, the system noise σ_y^M and σ_z^M are taken to be the same as σ_y and σ_z and the filtering skill with different observational noise σ_x^M is studied. Clearly, inflating the observational noise σ_x^M in the imperfect forecast model (18) leads to only small model errors (column (a), and column (c) with $\sigma_x^M > 5$). On the other hand, underestimating σ_x^M corresponds to a rapid increase of the RMS error and a quick decrease of the pattern correlation (column (e)). At the same time, the posterior variance R_t in the underdispersion case becomes smaller, implying these inaccurate estimations are quite certain. It is worthwhile noting that the model error $\mathscr{P}(\pi, \pi^{filter})$ in the time-averaged PDF of the posterior mean estimation associated with variable z shoots up in the underdispersive case (column (e)), indicating a significant lack of information in the filter estimates. Similar conclusions are found in with imperfect system noise levels σ_y^M and σ_z^M. Large errors are found when σ_y^M and σ_z^M are underdispersed (column (f)) while noise inflation has little negative effect on the model error (column (b) and (d)).

Figure 8 illustrates the posterior mean estimation as a function of time compared with the truth in two underdispersive situations. In the case that the observational noise σ_x^M is underestimated (panel (a)-(c)), the filtered trajectories of both y and z are quite noisy. In addition, the mean of the PDF associated with the filtered variable z has a positive bias, which explains the large model error in column (e) of Figure 7. Looking at the filtered trajectory of z in panel (c), the filtered solution misses many negative extreme events such as those around time $t = 7.5, 14.5$ and 18. At these time instants, the corresponding values of the observed variable x are all near zero, which implies the system losses practical observability (See Appendix A for details). In fact, when $x = 0$ the process of z is completely decoupled from x and y in (18) and observing x plays no role in filtering z. On the other hand, as shown in panel (d)-(f), even though the model errors $\mathscr{P}(\pi, \pi^{filter})$ in the unresolved variables y and z are small with the underestimated system noise σ_y^M and σ_z^M, the RMS error in the filter estimation remains significant.

Therefore, we conclude that underdispersion in the imperfect filter deteriorates the filtering skill while noise inflation within certain range introduces little error.

b. Recovering the time-dependent PDF of the unresolved variables utilizing conditional Gaussian mixture

One important issue in uncertainty quantification for turbulent systems is to recover the time-dependent PDF associated with the unobserved processes. In a typical scenario, the phase space of the unobserved variables is quite large while that of the observed ones remains moderate or small. The classical approaches involve solving the Fokker-Planck equation or adopting Monte Carlo simulation, both of which are quite expensive with the increase of the dimension, known as the “curse of dimensionality” (Majda and Harlim 2012; Daum and Huang 2003). For conditional Gaussian systems, the PDF associated with the unobserved processes can be approximated by an efficient conditional Gaussian ensemble mixture with high accuracy, where only a small ensemble of observed trajectories is needed due to its relatively low dimension and is thus computationally affordable. Note that the idea here is similar to that of the blended method for filtering high dimensional turbulent systems (Majda et al. 2014; Qi and Majda 2015; Slivinski et al. 2015; Sapsis and Majda 2013a).

Below, we provide a general framework of utilizing conditional Gaussian mixtures in approximating the time-dependent PDF associated with the unobserved processes. Although the test examples of this approach below are based on the 3D noisy L-63 system, this method can be easily generalized to systems with a large number of unobserved variables. This section deals with the situation with no model error. In Section 5c, the skill of recovering the PDF in the appearance of the model error due to noise inflation or underdispersion is explored.

Let us recall the observed variables $\mathbf{u_I}$ and the unobserved variables $\mathbf{u_{II}}$ in the conditional Gaussian system (1). Their joint distribution is denoted by

$$p(\mathbf{u_I},\mathbf{u_{II}}) = p(\mathbf{u_I})p(\mathbf{u_{II}}|\mathbf{u_I}).$$

Assume we have L independent observational trajectories $\mathbf{u}_\mathbf{I}^1, \ldots, \mathbf{u}_\mathbf{I}^L$ and therefore they are equally weighted. The marginal distribution of $\mathbf{u}_\mathbf{I}$ is approximated by

$$p(\mathbf{u}_\mathbf{I}) \approx \frac{1}{L}\sum_{i=1}^{L} \delta\left(\mathbf{u}_\mathbf{I} - \mathbf{u}_\mathbf{I}^i\right). \tag{22}$$

The marginal distribution of $\mathbf{u}_\mathbf{II}$ at time t is expressed by

$$\begin{aligned} p(\mathbf{u}_\mathbf{II}) &= \int p(\mathbf{u}_\mathbf{I}, \mathbf{u}_\mathbf{II}) d\mathbf{u}_\mathbf{I} = \int p(\mathbf{u}_\mathbf{I}) p(\mathbf{u}_\mathbf{II}|\mathbf{u}_\mathbf{I}) d\mathbf{u}_\mathbf{I} \\ &\approx \frac{1}{L}\sum_{i=1}^{L} p(\mathbf{u}_\mathbf{II}|\mathbf{u}_\mathbf{I}^i), \end{aligned} \tag{23}$$

where for each observation $\mathbf{u}_\mathbf{I}^i$, according to the analytically closed form (3),

$$p\big(\mathbf{u}_\mathbf{II}(t)|\mathbf{u}_\mathbf{I}^i(s \leq t)\big) \sim \mathcal{N}(\bar{\mathbf{u}}_\mathbf{II}^i(t), \mathbf{R}_\mathbf{II}^i(t)). \tag{24}$$

Thus, the PDF associated with the unobserved variable $\mathbf{u}_\mathbf{II}$ is approximated utilizing (23) and (24). Note that in many practical issues associated with turbulent systems, the dimension of the observed variables is much lower than that of the unobserved ones. Thus, only a small number of L is needed in approximating the low-dimensional marginal distribution $p(\mathbf{u}_\mathbf{I})$ in (22) to recover the marginal distribution $p(\mathbf{u}_\mathbf{II})$ associated with the unobserved process with this conditional Gaussian ensemble mixture approach.

Now we utilize the noisy L-63 model (16) as a test model for the conditional Gaussian ensemble mixture idea in (23). Here we assume x is the observed process while y and z are the unobserved ones. The tests with different observational noise σ_x and system noise σ_y and σ_z ranging from 1 to 10 reach similar qualitative conclusions and thus we only show the situation where $\sigma_x = \sigma_y = \sigma_z = 5$. The initial distribution is assumed to be Gaussian with mean $(x_0, y_0, z_0) = (1.51, -1.53, 25.46)$ following (Majda and Harlim 2012) and a small covariance $\mathrm{R}_0 = 0.1 I_3$.

Figure 9 shows the recovery of the first four central moments, i.e., mean, variance, skewness and kurtosis, associated with the unobserved variable z with different L. For comparison, we also show the results by adopting Monte Carlo simulation with a large ensemble number $N = 50,000$, which is regarded as the truth. Even with $L = 20$ in (23), the short-term transitions in the mean, variance and skewness are captured quite well. With $L = 100$ ensembles, the leading four moments have already been recovered with high accuracy. If $L = 500$ ensembles are adopted, then these statistics are recovered almost perfectly. The same results are found in variable y and thus they are omitted here.

In panel (a) of Figure 10, we show the model error (6) utilizing the conditional Gaussian ensemble mixture in recovering the marginal PDF of y at a short-term transition time $t = 0.46$, where the skewness arrives at its maximum. Clearly, the model error decays as L and it is already negligible with $L = 100$. The comparison of the marginal PDFs is shown in panel (e), which validates the results in panel (a). Panel (b) and (f) are the analogy to panel (a) and (e) for recovering the marginal PDF of z at the most skewed transition phase $t = 0.35$. Panel (c) and (d) show the model error dependence of L at the essentially statistical equilibrium phase ($t = 10$). Again, $L = 100$ is a sufficient number for approximating the marginal PDFs with high skill.

We have also tested the model error dependence on the ensemble number N utilizing Monte Carlo simulations. To reach a comparable skill with $L = 100$ utilizing conditional Gaussian ensemble mixture, the ensemble size utilizing Monte Carlo simulation is around $N = 5000$, which is much larger than L. Note that N increases dramatically with the dimension of the unobserved processes.

c. Recovering the time-dependent PDF of the unresolved variables with model error

Now we study recovering the time-dependent PDF of the unresolved variables in noisy L-63 model with model error. The true signal associated with the observed variable x is generated from model (16) and the imperfect model with model error in observational and system noise (18) is utilized to recover the unobserved PDFs. Below, the effect of the model error due to both noise inflation and underdispersion are explored.

First, we take $\sigma_x = \sigma_y = \sigma_z = 2$ in noisy L-63 model (16) to generate the true signal while the imperfect model for recovering the hidden PDFs (18) are equipped with noise $\sigma_x^M = \sigma_y^M = \sigma_z^M = 5$. Therefore, the noise is inflated in the imperfect forecast model. See Figure 11. The recovered statistics associated with y are quite accurate utilizing the conditional Gaussian mixture approach (23) with $L = 100$. On the other hand, there is an information barrier in the recovered PDF of z at a short-term transition time $t = 0.38$ due to the underestimation of the skewness. See column (c) and (e). Despite the failure of capturing this non-Gaussian feature at the short transition time, the time-dependent mean and variance of z are recovered with high accuracy with $L = 100$ and the equilibrium marginal distributions (column (f) and (g)) associated with both y and z are estimated with almost no model error.

Next, we take $\sigma_x = \sigma_y = \sigma_z = 10$ in noisy L-63 model (16) to generate the true signal while the imperfect model for recovering the hidden PDFs (18) are equipped with noise $\sigma_x^M = \sigma_y^M = \sigma_z^M = 5$. Therefore, model error comes from underdispersion in the imperfect forecast model. As shown in column (b) of Figure 12, the marginal variance of both y and z is always underestimated. Therefore, the recovered marginal PDFs have smaller spreads than the truth, especially at a short-term transition phase $t = 0.30$ for variable z (column (e)). Moreover, even at the essentially statistical equilibrium state $t = 5$, obvious errors are found in the tails of the recovered PDFs (column (g)), which implies the probability of the extreme events is severely underestimated.

6. Parameter estimation

One of the important issues in many scientific and engineering areas is to estimate model parameters given noisy observations. Classical ways of estimating parameters includes maximum likelihood (Snijders 2011; Sowell 1992), Bayesian inference (Bretthorst 2013; Golightly and Wilkinson 2008; Chen et al. 2014a) and least square methods (Marquardt 1963). One promising way for the real-time estimation of the parameters in turbulent systems is via filtering/data assimilation, in which the parameters are regarded as augmented state variables. Here, we study the skill of estimating the parameters in the dynamics that have the following general form,

$$d\mathbf{U} = (\mathbf{A}_0(t,\mathbf{U}) + \mathbf{A}_1(t,\mathbf{U})\boldsymbol{\Gamma}^*)\, dt + \boldsymbol{\Sigma}_U(\mathbf{U})\, d\mathbf{W}_U, \tag{25}$$

where $\mathbf{U} = (u_1,\ldots,u_m)^T$ are the observed state variables and $\boldsymbol{\Gamma}^* = (\gamma_1^*,\ldots,\gamma_n^*)^T$ are the parameters to be estimated that are assumed to be constants. We also assume the system contains random noise, the amplitude of which $\boldsymbol{\Sigma}_U(\mathbf{U})$ is known. Evidently, the function $\mathbf{A}_0(t,\mathbf{U}) + \mathbf{A}_1(t,\mathbf{U})\boldsymbol{\Gamma}^*$ can consist of polynomials or trigonometric polynomials, where the coefficient of each monomial is to be estimated. Note that both the dyad and triad systems in (12) and (15) belong to the model family (25) provided that all the state variables are observed.

Since these parameters $\boldsymbol{\Gamma}^*$ are constants, it is natural to augment the system (25) by an n-dimensional trivial equations for $\boldsymbol{\Gamma}^*$ (Harlim et al. 2014; Smedstad and O'Brien 1991; Van Der Merwe et al. 2001; Plett 2004; Wenzel et al. 2006). This forms the framework of *parameter estimation with direct approach*,

$$d\mathbf{U} = (\mathbf{A}_0(t,\mathbf{U}) + \mathbf{A}_1(t,\mathbf{U})\boldsymbol{\Gamma})\, dt + \boldsymbol{\Sigma}_U(\mathbf{U})\, d\mathbf{W}_U, \tag{26a}$$

$$d\boldsymbol{\Gamma} = 0. \tag{26b}$$

Throughout this section, $\boldsymbol{\Gamma}^*$ (with asterisk) always represents the true value of the parameters while $\boldsymbol{\Gamma}$ stands for the variables in the parameter estimation framework.

In some applications, prior information about the possible range of the parameters is available. To incorporate such information into the parameter estimation framework, we augment the system (25) by a group of stochastic equations of $\boldsymbol{\Gamma}$ (Majda and Harlim 2012), where the equilibrium distributions of these stochastic processes represent the prior information for the range of the parameters $\mathbf{\Gamma}$. This framework is called *parameter estimation with stochastic parameterized equations*,

$$d\mathbf{U} = (\mathbf{A}_0(t,\mathbf{U}) + \mathbf{A}_1(t,\mathbf{U})\mathbf{\Gamma})\, dt + \mathbf{\Sigma}_U(\mathbf{U})\, d\mathbf{W}_U, \tag{27a}$$

$$d\mathbf{\Gamma} = (\mathbf{a}_0 + \mathbf{a}_1\mathbf{\Gamma})\, dt + \mathbf{\Sigma}_\Gamma\, d\mathbf{W}_\Gamma. \tag{27b}$$

Given an initial value $\mu_{0,i}$ and an initial uncertainty $R_{0,i}$ of each component of $\mathbf{\Gamma}$, both the augmented systems (26) and (27) belong to the conditional Gaussian framework (1)–(3), where $\mathbf{u_I} = \mathbf{U}$ and $\mathbf{u_{II}} = \mathbf{\Gamma}$. Therefore, the time evolution of $\mathbf{\Gamma}$ is solved via closed analytic formulae.

Below we aim at studying the dependence of the error $\mu_{t,i} - \gamma_i^*$ and uncertainty $R_{t,i}$ on different factors, such as the noise in the system, the initial uncertainty and the model structure, utilizing both the framework in (26) and (27). The important role of observability in parameter estimation will be emphasized. In addition, the difference of the skill in parameter estimation in linear and nonlinear problems will be explored. The detailed derivations associated with the propositions shown below are all included in Appendix C.

a. Estimating one additive parameter in a linear scalar model

We start with estimating one additive parameter γ^* in the following linear scalar model,

$$du = (A_0 u + A_1 \gamma^*)dt + \sigma_u dW_u. \tag{28}$$

Given the initial guess μ_0 of the parameter γ^* with initial uncertainty R_0, the simple structure of model (28) allows the analytic expression of the error $\mu_t - \gamma^*$ in the posterior mean estimation and the posterior uncertainty R_t as a function of time.

We start with estimating the additive parameter γ^* in (28) within the framework utilizing direct approach (26),

$$du = (A_0 u + A_1 \gamma)dt + \sigma_u dW_u, \tag{29a}$$

$$d\gamma = 0. \tag{29b}$$

Proposition 2 *In estimating the additive parameter γ^* in* (28) *within the framework utilizing direct approach* (29)*, the posterior variance R_t and the error in the posterior mean $\mu_t - \gamma^*$ have the following closed analytical form,*

$$R_t = \frac{R_0}{1 + A_1^2 \sigma_u^{-2} R_0 t}, \tag{30a}$$

$$\mu_t - \gamma^* = \frac{\mu_0 - \gamma^*}{1 + A_1^2 \sigma_u^{-2} R_0 t} + \frac{A_1 \sigma_u^{-1} R_0}{1 + A_1^2 \sigma_u^{-2} R_0 t} \int_0^t dW_u(s). \tag{30b}$$

According to (30), both the posterior uncertainty R_t and the deterministic part of the error in posterior mean converge to zero asymptotically at an algebraic rate of time t^{-1}. The second term on its right hand side of (30b) represents the stochastic fluctuation of the error that comes from the system noise. The variance of this fluctuation at time t is given by

$$\text{var}(\mu_t - \gamma^*) = \frac{(A_1 \sigma_u^{-1} R_0)^2 t}{(1 + A_1^2 \sigma_u^{-2} R_0 t)^2},$$

the asymptotic convergence rate of which is t^{-1} as well.

It is clear from (30) that decreasing the noise σ_u and increasing the prefactor A_1 helps accelerate the reduction of both the error and the uncertainty for long-term behavior. In fact, a nearly zero A_1 implies the system losses practical observability, which corresponds to a slow convergence rate. On the other hand, although increasing the initial uncertainty R_0 enhances the convergence rate of the deterministic part of μ_t, it has no effect on the long-term behavior of reducing either the uncertainty R_t and the error in the fluctuation part of $\mu_t - \gamma^*$. In addition, a large R_0 leads to a large error in the fluctuation part of $\mu_t - \gamma^*$ at the initial period.

Next, we study estimating γ^* in (28) within the framework utilizing the stochastic parameterized equations in (27). To this end, we form the augmented system,

$$du = (A_0 u + A_1\gamma)dt + \sigma_u dW_u, \tag{31a}$$

$$d\gamma = (a_0 - a_1\gamma)dt + \sigma_\gamma dW_\gamma, \tag{31b}$$

where the equilibrium distribution of γ in (31b) is Gaussian with mean $\bar{\gamma} = a_0/a_1$ and variance $\mathrm{var}(\gamma) = \sigma_\gamma^2/(2a_1)$.

Proposition 3 *In estimating the additive parameter γ^* in* (28) *within the framework utilizing stochastic parameterized equations* (31)*, the posterior variance R_t and the error in the posterior mean $\mu_t - \gamma^*$ have the following closed analytical form,*

$$R_t = r_2 + \frac{r_1 - r_2}{1 - \left(\frac{R_0 - r_1}{R_0 - r_2}\right)\cdot \exp\left(-A_1^2\sigma_u^{-2}(r_1 - r_2)t\right)}, \tag{32a}$$

$$\begin{aligned}\mu_t - \gamma^* \approx &(\mu_0 - \gamma^*)e^{-(a_1 + R_{eq}A_1^2\sigma_u^{-2})t} + \frac{1 - e^{-(a_1 + R_{eq}A_1^2\sigma_u^{-2})t}}{a_1 + R_{eq}A_1^2\sigma_u^{-2}}(a_0 - a_1\gamma^*)\\ &+ R_{eq}A_1\sigma_u^{-1}\int_0^t e^{-(a_1 + R_{eq}A_1^2\sigma_u^{-2})(t-s)}dW_u(s),\end{aligned} \tag{32b}$$

where R_0 is assumed to be larger than r_1 in (32a) *and $r_1 > 0 > r_2$ are the two roots of the algebraic equation*

$$-A_1^2\sigma_u^{-2}R_t^2 - 2a_1R_t + \sigma_\gamma^2 = 0.$$

In (32b)*, the variance R_t is replaced by its equilibrium value R_{eq} for the conciseness of the expression due to its exponentially fast convergence rate.*

Unlike (30b) where the error in the posterior mean estimation converges to zero eventually, the error utilizing the stochastic parameterized equation (32b) converges to

$$|\mu_t - \gamma^*|_{eq} = \frac{|a_0 - a_1\gamma^*|}{a_1 + R_{eq}A_1^2\sigma_u^{-2}},$$

which is nonzero unless the mean of the stochastic parameterized equation (31b), i.e., $\gamma = -a_0/a_1$, equals the true value of the parameter γ^*. Similarly, the posterior uncertainty converges to a nonzero value r_1 unless the right hand side of (31b) disappears.

Yet, comparing the formulae in (30) and (32), it is obvious that the parameter estimation framework utilizing stochastic parameterized equations (31) leads to an exponential convergence rate for both the reduction of the posterior uncertainty and the error in the posterior mean, which implies a much shorter training data is needed in the framework (31). The convergence rate is controlled by the tuning factors in the stochastic parameterized equations. Thus, with a suitable choice of (31b), the convergence rate is greatly improved at the cost of only introducing a small bias in parameter estimation.

b. Estimating one multiplicative parameter in a linear model

Many applications require estimating parameters that appear as the multiplicative factors of the state variables. Here we study a simple situation where only one multiplicative parameter γ^* appears in the dynamics. Consider the following system,

$$du = (A_0 - \gamma^* u)dt + \sigma_u dW_u, \tag{33}$$

where we assume the parameter $\gamma^* > 0$ to guarantee the mean stability of the system. Given the initial guess μ_0 of the parameter γ^* with initial uncertainty R_0, the analytic expressions of the error $\mu_t - \gamma^*$ and the uncertainty R_t are still available in the framework utilizing direct approach (26). There is no simple closed expression for the error estimation in the framework utilizing stochastic parameterized equations (27) but numerical results will be provided for comparing the skill utilizing the two approaches.

The augmented system utilizing direct approach (26) has the following form,

$$
\begin{aligned}
&du = (A_0 - \gamma u)dt + \sigma_u dW_u, \\
&d\Gamma = 0.
\end{aligned}
\tag{34}
$$

Proposition 4 *In estimating the multiplicative parameter γ^* in* (33) *within the parameter estimation framework utilizing direct approach* (34)*, the posterior variance R_t and the error in the posterior mean $\mu_t - \gamma^*$ have the following closed analytical form,*

$$
R_t = \frac{R_0}{1 + R_0\sigma_u^{-2}\int_0^t u^2(s)ds}, \tag{35a}
$$

$$
\mu_t - \gamma^* = \frac{\mu_0 - \gamma^*}{1 + R_0\sigma_u^{-2}\int_0^t u^2(s)ds} - \frac{R_0\sigma_u^{-1}}{1 + R_0\sigma_u^{-2}\int_0^t u^2(s)ds}\int_0^t u(s)dW_u(s). \tag{35b}
$$

The long-term behavior of (35) *can be further simplified. Apply the Reynold's decomposition*

$$
u(t) = \bar{u}(t) + u'(t) \qquad \textit{with} \qquad \overline{u'} = 0 \;\; \textit{and} \;\; \overline{u'\bar{u}} = 0, \tag{36}
$$

where $\bar{u}(t)$ represents the ensemble mean of a random variable u at a fixed time t. Thus,

$$
\int_0^t u^2(s)ds = \int_0^t (\bar{u}(s))^2 ds + \int_0^t (u'(s))^2 ds. \tag{37}
$$

Utilizing ergodicity, the two integrals on the right hand of (37) *are given by*

$$
\begin{aligned}
&\lim_{t\to\infty}\frac{1}{t}\int_0^t (\bar{u}(s))^2 ds = \int_{-\infty}^{\infty} (\bar{u}(s))^2 p_{eq}(u)dU = \frac{A_0}{\gamma^*}, \\
&\lim_{t\to\infty}\frac{1}{t}\int_0^t (u'(s))^2 ds = \int_{-\infty}^{\infty} (u')^2 p_{eq}(u)dU = \frac{\sigma_u^2}{2\gamma^*},
\end{aligned}
\tag{38}
$$

respectively, where $p_{eq}(u)$ is the equilibrium Gaussian distribution associated with the system (33)*. Thus, the long-term behavior of* (35) *simplifies to*

$$
R_t \approx \frac{R_0}{1 + R_0\sigma_u^{-2}A_0^2(\gamma^*)^{-2}t + R_0(2\gamma^*)^{-1}t}. \tag{39a}
$$

$$
\begin{aligned}
\mu_t - \gamma^* \approx\; & \frac{\mu_0 - \gamma^*}{1 + R_0\sigma_u^{-2}A_0^2(\gamma^*)^{-2}t + R_0(2\gamma^*)^{-1}t} \\
& - \frac{R_0\sigma_u^{-1}}{1 + R_0\sigma_u^{-2}A_0^2(\gamma^*)^{-2}t + R_0(2\gamma^*)^{-1}t}\int_0^t u(s)dW_u(s).
\end{aligned}
\tag{39b}
$$

Similar to the situation in estimating one additive parameter in (30), the convergence of both the error and uncertainty in (39) is at an algebraic rate t^{-1}. However, the convergence strongly depends on the prefactor A_0. When A_0 is zero, the denominator of the terms on the right hand side of (39) becomes $(1+R_0(2\gamma^*)^{-1}t)$, which is independent of the noise amplitude σ_u. On the other hand, when A_0 is highly non-zero, decreasing the noise level σ_u accelerates the convergence. In fact, a nearly zero A_0 implies that the mean state of u is nearly zero and the system (34) has no practical observability. With a small noise, losing practical observability implies a much slower convergence.

Alternately, the augmented system utilizing stochastic parameterized equations (27) has the form,

$$du = (A_0 - \gamma u)dt + \sigma_u dW_u, \tag{40a}$$

$$d\gamma = (a_0 - a_1\gamma)dt + \sigma_\gamma dW_\gamma. \tag{40b}$$

Since there is no simple closed formulae for the error and uncertainty in the posterior estimation, we show the numerical results utilizing the equations in (3) for estimating γ^* utilizing (40) and compare with those from (34).

In Figure 13, we show the posterior mean μ_t and the posterior uncertainty R_t in estimating the multiplicative parameter γ^* in (33) utilizing both the direct approach (34) and the stochastic parameterized equation (40). Here, the truth is $\gamma^* = 5$. The constant factor A_0 in (33) is set to be $A_0 = 0$ such that the system has no practical observability. Different noise σ_u and initial uncertainty R_0 are chosen. To introduce an initial error, the initial value of γ in both (34) and (40) is set to be $\gamma_0 = 2$. When estimating γ utilizing stochastic parameterized equation (40), the ratio $a_0/a_1 = 5.5$ is assumed such that there exists a bias in the equilibrium mean in (40b) and the equilibrium variance $\sigma_\gamma^2/(2a_1) = 2$ is also fixed. Thus, there is only one freedom a_1 in (40b), the inverse of which is the decorrelation time.

We first look at the parameter estimation skill utilizing the direct approach (34). Since $A_0 = 0$, the system has no practical observability and the convergence rate has no dependence on σ_u according to (39), which is validated by panels (a)-(c) in Figure 13. Clearly, the posterior uncertainty R_t goes to zero but the error in the posterior mean $|\mu_t - \gamma^*|$ is still above 0.5 even after $t = 100$ nondimensional units. When the initial uncertainty decreases from $R_0 = 0.5$ (panel (a)) to $R_0 = 0.1$ (panel (d)), the convergence becomes slower as expected from (39). On the other hand, the convergence utilizing the stochastic parameterized equation (40b) (panels (e)-(h)) is much faster and it is almost unchanged by reducing the initial uncertainty (panel (h)). Although the equilibrium mean of the stochastic parameterized equation (40b) has a bias of 0.5 unit in γ, with the help of observations the averaged posterior mean at the equilibrium differs from the truth by only 0.1 to 0.2 unit. In addition, the posterior mean estimation is quite robust with respect to the choice of the coefficients a_0, a_1 and σ_γ in the stochastic parameterized equations (40b) as seen in panels (e)-(g). Yet, overestimating (panel (e)) and underestimating (panel (g)) a_1 lead to the increase of fluctuations and the decrease of convergence, respectively. The optimal choice here is $a_1 = 0.005$ as shown in panel (f).

We have so far focused on the parameter estimation skill in the appearance of one observational trajectory. In some applications, repeated experiments are available and therefore it is worthwhile studying the parameter estimation skill given an ensemble of independent observations. Assume the number of the independent observed trajectory is L. Corresponding to (34), the parameter estimation framework utilizing direct approach is given by

$$\begin{aligned} d\mathbf{u} &= (\mathbf{A}_0 - \gamma\mathbf{u})dt + \Sigma_u d\mathbf{W}_u, \\ d\Gamma &= 0, \end{aligned} \tag{41}$$

where $\mathbf{u}$ is a $1\times L$ column vector, representing L independent observations. All the entries in the $1\times L$ column vector $\mathbf{A}_0$ are equal to A_0. Both Σ_u and $\mathbf{W}_u$ are $L\times L$ diagonal matrices, where each diagonal entry of Σ_u is σ_u.

Proposition 5 *In estimating the multiplicative parameter γ^* in* (33) *within the parameter estimation framework utilizing direct approach* (41) *with L independent observed trajectories, the posterior variance R_t and the error in the posterior mean $\mu_t - \gamma^*$ have the following closed analytical form,*

$$R_t = \frac{R_0}{1 + LR_0\sigma_u^{-2}\int_0^t u^2(s)ds}, \tag{42a}$$

$$\mu_t - \gamma^* = \frac{\mu_0 - \gamma^*}{1 + LR_0\sigma_u^{-2}\int_0^t u^2(s)ds} - \frac{R_0\sigma_u^{-1}}{1 + LR_0\sigma_u^{-2}\int_0^t u^2(s)ds}\int_0^t u(s)dW_u(s). \tag{42b}$$

Comparing (35) and (42), the asymptotic convergence with L independent trajectories within the direct approach framework is enhanced by a multiplier L in front of t. Thus, increasing the number of independent observations accelerates the convergence but the convergence rate remains algebraic.

c. Estimating parameters in cubic nonlinear models

From now on, we study the parameter estimation issue in nonlinear models. Our focus is on a model with cubic nonlinearity,

$$du = (a^*u + b^*u^2 - c^*u^3 + f^*)dt + \sigma_u dW_u, \tag{43}$$

where $c^* > 0$ to guarantee the mean stability. The cubic model (43) is a special case of the normal form for reduced stochastic climate model (Majda et al. 2009) and it is utilized as a test model for fluctuation-dissipation theorems in (Majda et al. 2010a). The goal is to estimate the four parameters $\mathbf{\Gamma} = (a^*, b^*, c^*, f^*)$.

To understand the underlying difference of estimating parameters in nonlinear and linear dynamics, we begin with a simplified version of (43),

$$du = (A_0 - \gamma^* u^3)dt + \sigma_u dW_u, \tag{44}$$

where the analytic formulae of the posterior uncertainty and the error in the posterior mean are available in the framework utilizing the direct approach (26),

$$du = (A_0 - \gamma u^3)dt + \sigma_u dW_u, \tag{45a}$$

$$d\gamma = 0. \tag{45b}$$

Proposition 6 *For any odd k, the framework utilizing direct approach* (26) *to estimate the parameter γ^* in*

$$du = (A_0 - \gamma^* u^k)dt + \sigma_u dW_u$$

is given by

$$du = (A_0 - \gamma u^k)dt + \sigma_u dW_u, \tag{46a}$$

$$d\gamma = 0. \tag{46b}$$

The posterior variance R_t and the error in the posterior mean $\mu_t - \gamma^*$ associated with system (46) *have the following closed analytical form,*

$$\begin{aligned} R_t &= \frac{R_0}{1 + R_0 \sigma_u^{-2} \int_0^t u^{2k}(s)ds}, \\ \mu_t - \gamma^* &= \frac{\mu_0 - \gamma^*}{1 + R_0 \sigma_u^{-2} \int_0^t u^{2k}(s)ds} - \frac{R_0 \sigma_u^{-1}}{1 + R_0 \sigma_u^{-2} \int_0^t u^{2k}(s)ds} \int_0^t u(s) dW_u(s). \end{aligned} \tag{47}$$

Applying Reynold's decomposition (36)*, the integral $\int_0^t u^{2k}(s)ds$ can be rewritten as*

$$\int_0^t u^{2k}(s)ds = \int_0^t (\bar{u}(s) + u'(s))^{2k} ds = \int_0^t \sum_{m=0}^{2k} \binom{2k}{m} \bar{u}^m \cdot (u'(s))^{2k-m} ds. \tag{48}$$

Regarding the cubic model in (45)*, the index k in* (46) *and* (48) *is set to be $k = 3$. Further consider the situation with $A_0 = 0$, which implies the system losses practical observability with $\bar{u} = 0$ at the equilibrium. Clearly, the only non-zero term on the right hand side of* (48) *at a long-term range is $\int_0^t (u'(s))^6 ds$. Since $\bar{u} = 0$, we utilize u to replace u' for notation simplicity. In light of the ergodicity of u,*

$$\int_{-\infty}^{\infty} u^6 p_{eq}(u) du = \lim_{t\to\infty} \frac{1}{t} \int_0^t u^6(s) ds,$$

where the analytic expression of the equilibrium PDF $p_{eq}(u)$ is given by (Majda et al. 2009),

$$p_{eq}(u) = N_0 \exp\left(\frac{2}{\sigma_u^2}\left(-\frac{\gamma^*}{4}u^4\right)\right).$$

Direct calculation shows that

$$\int_{-\infty}^{\infty} U^6 p_{eq}(U) dU = \left(\frac{2}{\gamma^*}\right)^{\frac{3}{2}} \sigma_U^3 \left(\Gamma\left(\frac{1}{4}\right)\right)^{-1} \Gamma\left(\frac{7}{4}\right), \tag{49}$$

where Γ is the Gamma function (Abramowitz et al. 1965). Therefore, the posterior variance R_t and the error in the posterior mean $\mu_t - \gamma^*$ for the long-term behavior utilizing the direct approach (45) *with $A_0 = 0$ have the following closed analytical form,*

$$R_t \approx \frac{R_0}{1 + \tilde{c} R_0 \sigma_u t}, \tag{50a}$$

$$\mu_t - \gamma^* \approx \frac{\mu_0 - \gamma^*}{1 + \tilde{c} R_0 \sigma_u t} - \frac{R_0 \sigma_u^{-1}}{1 + \tilde{c} R_0 \sigma_u t} \int_0^t u(s) dW_u(s), \tag{50b}$$

where the constant $\tilde{c} = (2/\gamma^*)^{3/2}\Gamma(7/4)/\Gamma(1/4)$.

We compare the results in (50) of cubic nonlinear system (44) with those in (39) of the linear system (33). The most significant difference is the role of the noise σ_u. In the linear model, without practical observability, i.e, $A_0 = 0$, the convergence rate has no dependence on σ_u. On the other hand, in the cubic nonlinear model, increasing the noise σ_u accelerates the convergence! This seems to be counterintuitive. However, the cubic nonlinearity, serving as the cubic damping in (44), indicates that the state variable u is trapped to the region around its attractor $u = 0$ more severely than that in the linear model. Since the system has no practical observability around $u = 0$, an enhanced σ_u is preferred for increasing the amplitude of u and thus improves the parameter estimation skill.

Now we focus on the full cubic system (43) and estimate the four parameters (a^*, b^*, c^*, f^*) in different dynamical regimes. Phase portrait analysis indicates that the deterministic part of (43) can have either 1) one stable equilibrium or 2) two stable equilibria and one unstable equilibrium. Here we fix the parameter $b^* = -4$ and $c^* = 4$ and consider the free parameters a^* and f^*. The phase space (a^*, f^*) is divided into two separate regions with different dynamical behaviors, where the dividing curve between these two regimes can be written down analytically (Majda et al. 2009),

$$f^* = -\frac{a^* b^*}{3c^*} - \frac{2(b^*)^3}{27(c^*)^2} \pm 2c^* \left(\frac{a^*}{3c^*} + \frac{(b^*)^2}{9(c^*)^2} \right)^{3/2}.$$

See Figure 14.

Below, we study the parameter estimation skill within the framework (26) utilizing the direct approach in three dynamical regimes as shown in Figure 14, where regime I $(f^* = 2)$ and regime III $(f^* = 10)$ correspond to one and three equilibria in phase portrait, respectively, and regime II $(f^* = 4.5)$ has one equilibrium but the parameter values are near the dividing curve. Here a moderate noise $\sigma_u = \sqrt{2}$ is chosen. Panel (a) and (b) in Figure 15 show the observed trajectories and equilibrium PDFs of u for the three regimes. The bimodal and nearly Gaussian PDFs for regime I and III are due to the number of stable equilibria. The PDF for regime II is skewed where the one-sided extreme events in the trajectory increase the probability at the tail of the PDF. In the parameter estimation framework (26), the initial value μ_0 of each parameter is chosen to be 2 units smaller than the truth and the initial uncertainty is set to be $R_0 = 5$.

The posterior mean μ_t and posterior uncertainty R_t associated with the parameter c corresponding to three regimes are shown in panel (c) and (d) of Figure 15. A rapid convergence in both posterior mean and posterior uncertainty is found in regime I, where the two distinct states in the trajectory of u clearly indicate the dynamical behavior. As contrast, the convergence of the posterior uncertainty in regime III is quite slow and the posterior mean remains far from the truth even after $t = 500$ nondimensional unit. Such unskillful behavior is due to the fact that the dynamical structure is hard to be recovered from the noisy trajectory with short memory. An interesting phenomenon is found in the regime II. The convergence remains slow at short- and medium-lead times while a sudden uncertainty reduction occurs around $t = 135$, at which time an extreme event occurs in u. Such extreme events, despite having small probability, are important in conveying information of the dynamical structure.

It is worthwhile remarking that if the noise σ_u is too small in Regime I, then the trajectory of u will be trapped in one attractor, which leads to an extremely slow convergence of the posterior uncertainty and a significant error in the posterior mean estimation. Thus, a moderately large noise helps enhance the parameter estimation skill in the model with cubic nonlinearity, which is consistent with the conclusions from the special case in (50).

Finally, to overcome the slow convergence of parameter estimation utilizing the direct approach (26) in the Regime III, we turn to the framework utilizing stochastic parameterized equation (27), which is given by

$$du = (au + bu^2 - cu^3 + f)dt + \sigma_u dW_u, \tag{51a}$$

$$d\gamma = (a_0 - a_1\gamma)dt + \sigma_\gamma dW_\gamma, \qquad \gamma \text{ stands for } a, b, c \text{ or } f. \tag{51b}$$

For each parameter, we set the mean of the stochastic parameterized equations (51b) a_0/a_1 to be 0.5 units larger than the truth, representing model error, and the equilibrium variance is assumed to be $\sigma_\gamma^2/(2a_1) = 2$. The damping coefficient is set to be $a_1 = 0.01$ for all the four parameters. The comparison of estimating the four parameters utilizing the direct approach (26) and stochastic parameterized equation (27) is shown in Figure 16. Estimating the parameters utilizing the stochastic parameterized equations has a much faster convergence and the model error in the stochastic parameterized equation is alleviated with observations.

7. Summary conclusions

In this paper, we study filtering the nonlinear turbulent dynamical system (1) through conditional Gaussian statistics. The special structure of the system allows closed analytic form for the updates of the posterior states (Section 2). Information measures (Section 3) are adopted for assessing the model error and lack of information in filtering.

The role of energy-conserving nonlinear interactions in filtering the turbulent systems is studied in Section 4 based on a dyad model (12). The lack of information in the stochastic parameterized filter (13) is large and the energy-conserving nonlinear feedback is found to be particularly important when the stochastic noise amplitude σ_u in the observed process is not negligible. The observability plays a key role with moderate σ_v and small σ_u in generating the true signal. Intermittency increases the signal to noise ratio, which helps improve the filtering skill.

The model error in the stochastic forcing amplitudes is studied in Section 5 where the L-63 model (a triad model) is adopted as a test model. Both mathematical analysis (Proposition (1)) and numerical experiments (Figure 7) support that noise inflation leads to little error in filtering the unobserved trajectory while significant model errors are found in the imperfect filter due to underdispersion (Figure 6, 7 and 8). An efficient conditional Gaussian ensemble mixture method (23) is proposed in approximating the time-dependent PDF of the unobserved processes, which requires only small ensembles (Figure 10) and can be generalized to systems with a large number of unobserved variables. Again, noise inflation in the imperfect model leads to only a small model error (Figure 11) while underdispersion results in an obvious gap in estimating the PDF, where a severe underestimation of the variance implies the failure of capturing extreme events (Figure 12).

The conditional Gaussian framework also allows systematical study of parameter estimation skill, where the parameters are regarded as the augmented state variables. The convergence rate of the estimated parameters depends largely on the observability. Without practical observability, a slow convergence rate is found utilizing the direct parameter estimation approach (26) (Proposition 2, Proposition 4). On the other hand, a suitable choice of the stochastic parameterized equations for the augmented state variables (27) leads to an exponentially fast convergence rate at the cost of only introducing a small error (Proposition 3, Figure 13). In estimating parameters in a cubic nonlinear system, the convergence rate varies in different dynamical regimes utilizing the direct approach (26). The solutions converge to the truth very quickly in a bimodal regime while an extremely slow convergence is found in the nearly Gaussian regime (Figure 15). Adopting the stochastic parameterized equations (27) again improves the skill of parameter estimation significantly (Figure 16).

Developing a systematic framework for optimizing the stochastic parameterized equations will be useful for estimating parameters in more complex systems. The information-theoretic framework described in Section 3 is a good candidate for this optimization, which remains as a future work. Other future works involve designing computationally affordable filters based on (3), following the guidelines provided in this work, for more complex turbulent systems. Noticeably, the conditional Gaussian framework (1)–(3) is also quite useful in studying the ensemble prediction skill and quantifying the uncertainty with model error.

Acknowledgments. This research of A.J.M is partially supported by the Office of Naval Research grant ONR MURI N00014-12-1-0912. N.C. is supported as a graduate research assistant on this grant. A.J.M also gratefully acknowledges the generous support of the Center for Prototype Climate Modeling of NYU Abu Dhabi Research Institute grant.

APPENDIX A

Observability of continuous systems

Observability plays an important role in filtering the hidden variables from observations. Let's consider the linearized coupled observation-filtering system,

$$\dot{\mathbf{u}} = A\mathbf{v} + B\mathbf{u}, \tag{A1}$$

$$\dot{\mathbf{v}} = C\mathbf{v} + D\mathbf{u}, \tag{A2}$$

where $\mathbf{u}$ and $\mathbf{v}$ are the observational and filtering processes, respectively.

The observability (Gajic and Lelic 1996) of system (A1)-(A2) can be derived as follows. Taking one more derivative with respect to (A1), with the help of (A2), yields

$$\begin{aligned}\ddot{\mathbf{u}} &= A\dot{\mathbf{v}} + B\dot{\mathbf{u}} \\ &= A(C\mathbf{v} + D\mathbf{u}) + B(A\mathbf{v} + B\mathbf{u}) \\ &= (AC + BA)\mathbf{v} + (AD + B^2)\mathbf{u}.\end{aligned} \tag{A3}$$

Similar argument applies for higher order derivative of $\mathbf{u}$. Therefore, the augmented system is given by

$$\begin{aligned}\begin{pmatrix} \dot{\mathbf{u}} \\ \ddot{\mathbf{u}} \\ \vdots \end{pmatrix} &= \begin{pmatrix} A \\ AC + BA \\ \vdots \end{pmatrix} \mathbf{v} + \begin{pmatrix} B \\ AD + B^2 \\ \vdots \end{pmatrix} \mathbf{u} \\ &:= \mathscr{O}\mathbf{v} + \mathscr{F}\mathbf{u}.\end{aligned} \tag{A4}$$

A system is said to be observable if, for any possible sequence of the state (unobserved variable) $\mathbf{v}(s), (s \leq t)$ and control quantities A, B, C and D, the current state $\mathbf{v}(t)$ can be determined using only the observations $\mathbf{u}(s), (s \leq t)$. Therefore, the condition of the observability is that the rank of matrix $\mathscr{O}$ equals the dimension of $\mathbf{v}$. In practice, due to the noise and numerical errors, the system is said to have no practical observability if the matrix $\mathscr{O}$ is nearly singular.

A.1. Observability of the dyad model (12)**.**

Let's linearize both u and v around the mean states $\bar{u}$ and $\bar{v}$,

$$u = \bar{u} + u', \qquad v = \bar{v} + v'.$$

The associated equations of (12) for the perturbed variables u' and v' are given by

$$\begin{aligned} du' &= -d_{uu}u' + \gamma\bar{v}u' + \gamma\bar{u}v', \\ dv' &= -d_{vv}v' - 2\gamma\bar{u}u'. \end{aligned} \tag{A5}$$

Since v' in (A5) is a scalar, the observability matrix $\mathscr{O}$ in (A4) becomes $\mathscr{O} = \gamma\bar{u}$. Clearly, the dyad system (12) losses its observability when $\bar{u} = 0$. This implies the unobserved variable v is decoupled from the observational process. In dynamical regime (B) with $F_u = 0$, the fixed point is $u_c = 0$, around which the system has no practical observability.

A.2. Observability of the L-63 model (16)**.**

Again, we linearize x, y and z around their mean states $\bar{x}$, $\bar{y}$ and $\bar{z}$,

$$x = \bar{x} + x', \qquad y = \bar{y} + y', \qquad z = \bar{z} + z',$$

The associated equations of L-63 model (16) for the perturbed variables are given by

$$\begin{aligned} dx' &= \sigma(y' - x')dt, \\ dy' &= \big(x'(\rho - \bar{z}) - \bar{x}z' - y'\big)dt, \\ dz' &= \big(\bar{x}y' + x'\bar{y} - \beta z'\big)dt. \end{aligned} \tag{A6}$$

If x is the observed variables and y and z are the filtering variables, then corresponding to (A1)–(A2), $\mathbf{u} = x'$, $\mathbf{v} = (y', z')^T$, and

$$A = (\sigma, 0), \qquad B = -\sigma, \qquad C = \begin{pmatrix} -1 & -\bar{x} \\ \bar{x} & -\beta \end{pmatrix}, \qquad D = \begin{pmatrix} \rho - \bar{z} \\ \bar{y} \end{pmatrix}.$$

According to (A4), the observability matrix is given by

$$\mathscr{O} = \begin{pmatrix} A \\ AC + BA \end{pmatrix} = \begin{pmatrix} \sigma & 0 \\ -\sigma - \sigma^2 & -\sigma\bar{x} \end{pmatrix}$$

Since $\sigma = 10$ is given and is non-zero, the system loses observability when $\bar{x} = 0$. It is also clear that when the system loses observability, the second column of the observability matrix $\mathscr{O}$ becomes zero and therefore observations provides no information in filtering the variable z.

On the other hand, if the observed variables are y and z and the filtering variable is x, direct calculations show that the system has no observability when $\bar{y} = 0$ and $\bar{z} = \rho$.

APPENDIX B

Detailed derivation of Proposition 1 in Section 5a of triad models

In the situation of observing y and z while filtering x, the posterior variance utilizing the forecast model (18) is given by, according to (3),

$$\begin{aligned} dR_t &= \left[-2\sigma R_t + (\sigma_x^M)^2 - R_t(\rho - z, y) \begin{pmatrix} (\sigma_y^M)^{-2} & \\ & (\sigma_z^M)^{-2} \end{pmatrix} \begin{pmatrix} \rho - z \\ y \end{pmatrix} R_t \right] dt \\ &= \left[-2\sigma R_t + (\sigma_x^M)^2 - \left((\rho - z)^2 (\sigma_y^M)^{-2} + y^2 (\sigma_z^M)^{-2} \right) R_t^2 \right] dt. \end{aligned} \tag{B1}$$

Note that the covariance matrix R_t remains non-negative in (B1). Clearly,

$$(\rho - z)^2 (\sigma_y^M)^{-2} + y^2 (\sigma_z^M)^{-2} \geq 0,$$

and therefore

$$\left((\rho - z)^2(\sigma_y^M)^{-2} + y^2(\sigma_z^M)^{-2}\right) R_t \geq 0.$$

If we formally write (B1) as

$$dR_t = -2\tilde{\sigma} R_t + (\sigma_x^M)^2, \tag{B2}$$

where

$$\tilde{\sigma} = \sigma + \left((\rho - z)^2(\sigma_y^M)^{-2} + y^2(\sigma_z^M)^{-2}\right) \frac{R_t}{2} > \sigma,$$

then it is obvious that the convergence rate of the posterior covariance to the equilibrium is faster than $\exp(-2\sigma t)$. Actually, the solution of R_t is bounded by

$$R_t \leq R_0 e^{-2\sigma t} + (\sigma_x^M)^2 \frac{1 - e^{-2\sigma t}}{2\sigma}, \tag{B3}$$

where the right hand side of (B3) is the solution of the following equation

$$dR_t = -2\sigma R_t + (\sigma_x^M)^2.$$

If the system noise σ_x^M in the filter model (18) is zero, then the posterior variance converges to zero in the exponential rate.

The posterior mean evolution can be written down explicitly

$$d\mu_t = (\sigma y - \sigma \mu_t)dt + R_t(\rho - z, y) \begin{pmatrix} (\sigma_y^M)^{-2} & \\ & (\sigma_z^M)^{-2} \end{pmatrix} \left[\begin{pmatrix} dy_t \\ dz_t \end{pmatrix} - \left(\begin{pmatrix} -y \\ -\beta z \end{pmatrix} + \begin{pmatrix} \rho - z \\ y \end{pmatrix} \mu_t \right) dt \right]. \tag{B4}$$

Recall y and z equation in the perfect model (19),

$$\begin{pmatrix} dy_t \\ dz_t \end{pmatrix} = \left(\begin{pmatrix} -y \\ -\beta z \end{pmatrix} + \begin{pmatrix} \rho - z \\ y \end{pmatrix} x_t \right) dt. \tag{B5}$$

Therefore, inserting (B5) into (B4) leads to

$$
\begin{aligned}
d\mu_t &= (\sigma y - \sigma\mu_t)dt + R_t(\rho - z, y)\begin{pmatrix} (\sigma_y^M)^{-2} & \\ & (\sigma_z^M)^{-2} \end{pmatrix}\begin{pmatrix} \rho - z \\ y \end{pmatrix}(x_t - \mu_t)dt \\
&= \sigma(y - \mu_t)dt - R_t\left((\rho - z)^2(\sigma_y^M)^{-2} + y^2(\sigma_z^M)^{-2}\right)(\mu_t - x_t)dt.
\end{aligned} \tag{B6}
$$

In addition, note the x equation of the perfect model (19) is given by,

$$
dx_t = \sigma(y - x_t)dt. \tag{B7}
$$

Subtracting (B7) from (B6) leads to

$$
d(\mu_t - x_t) = -\sigma(\mu_t - x_t)dt - R_t\left((\rho - z)^2(\sigma_y^M)^{-2} + y^2(\sigma_z^M)^{-2}\right)(\mu_t - x_t)dt. \tag{B8}
$$

The error equation $\varepsilon = \|\mu_t - x_t\|^2$ is given by

$$
d\varepsilon = -\left(\sigma + R_t\left((\rho - z)^2(\sigma_y^M)^{-2} + y^2(\sigma_z^M)^{-2}\right)\right)\varepsilon dt. \tag{B9}
$$

Since both $(\rho - z)^2(\sigma_y^M)^{-2} + y^2(\sigma_z^M)^{-2}$ are R_t are non-negative, the error is bounded by

$$
\varepsilon \leq \varepsilon_0 e^{-\sigma t}, \tag{B10}
$$

which decays to zero in an exponential rate.

APPENDIX C

Detailed derivations for Proposition 2–4 in Section 6a and 6b of parameter estimation in linear models

C.1. Detailed derivations of Proposition 2.

We aim at estimating the additive parameter γ^* in the linear system,

$$
du = (A_0 u + A_1\gamma^*)dt + \sigma_u dW_u. \tag{C1}
$$

The augmented system for estimating γ^* utilizing direct approach (26) is given by

$$du = (A_0 u + A_1 \gamma)dt + \sigma_u dW_u, \tag{C2a}$$

$$d\gamma = 0, \tag{C2b}$$

where the initial guess μ_0 and the initial uncertainty R_0 are assigned. In light of (3), the evolutions of posterior mean and posterior variance of γ have the following form

$$d\mu_t = R_t A_1 \cdot \sigma_u^{-2} \cdot [du - (A_0 u + A_1 \mu_t)dt], \tag{C3a}$$

$$dR_t = -A_1^2 \sigma_u^{-2} R_t^2 dt. \tag{C3b}$$

The solution of R_t in (C3b) is reached by separation of variables,

$$R_t = \frac{R_0}{1 + A_1^2 \sigma_u^{-2} R_0 t}. \tag{C4}$$

To calculate the error in the posterior mean μ_t compared with the constant truth γ^*, we first rewrite (C3b) as

$$d(\mu_t - \gamma^*) = R_t A_1 \cdot \sigma_u^{-2} \cdot [du - (A_0 u + A_1 \mu_t)dt]. \tag{C5}$$

Since u in (C5) is from the true observations, we insert (C1) into (C5),

$$\begin{aligned} d(\mu_t - \gamma^*) &= R_t A_1 \cdot \sigma_u^{-2} \cdot [(A_0 u + A_1 \gamma^*)dt + \sigma_u dW_u - (A_0 u + A_1 \mu_t)dt], \\ &= -R_t A_1^2 \sigma_u^{-2} (\mu_t - \gamma^*)dt + R_t A_1 \sigma_u^{-1} dW_u. \end{aligned} \tag{C6}$$

With the expression of the variance R_t in (C4), we have

$$d(\mu_t - \gamma^*) = -\frac{R_0 A_1^2 \sigma_u^{-2}}{1 + A_1^2 \sigma_u^{-2} R_0 t}(\mu_t - \gamma^*)dt + \frac{R_0 A_1 \sigma_u^{-1}}{1 + A_1^2 \sigma_u^{-2} R_0 t} dW_u. \tag{C7}$$

For the simplicity of notation, we define

$$y := \mu_t - \gamma^*, \qquad \tilde{a} := A_1^2 \sigma_u^{-2} R_0, \qquad \text{and} \qquad \tilde{b} := R_0 A_1 \sigma_u^{-1}.$$

Then (C7) becomes

$$dy = -\frac{\tilde{a}}{1 + \tilde{a}t} y dt + \frac{\tilde{b}}{1 + \tilde{a}t} dW_u. \tag{C8}$$

Applying the method of integrating factor, we have

$$\begin{aligned} y &= y_0 e^{-\int_0^t \frac{\tilde{a}}{1+\tilde{a}s}ds} + e^{-\int_0^t \frac{\tilde{a}}{1+\tilde{a}s}ds} \int_0^t \frac{\tilde{b}}{1+\tilde{a}s} e^{\int_0^s \frac{\tilde{a}}{1+\tilde{a}v}dv} dW_u(s), \\ &= y_0 e^{-\ln \frac{t+\tilde{a}^{-1}}{\tilde{a}^{-1}}} + e^{-\ln \frac{t+\tilde{a}^{-1}}{\tilde{a}^{-1}}} \int_0^t \frac{\tilde{b}\tilde{a}^{-1}}{s+\tilde{a}^{-1}} e^{\ln \frac{s+\tilde{a}^{-1}}{\tilde{a}^{-1}}} dW_u(s), \\ &= y_0 \frac{\tilde{a}^{-1}}{t+\tilde{a}^{-1}} + \frac{\tilde{a}^{-1}}{t+\tilde{a}^{-1}} \int_0^t \frac{\tilde{b}\tilde{a}^{-1}}{s+\tilde{a}^{-1}} \frac{s+\tilde{a}^{-1}}{\tilde{a}^{-1}} dW_u(s), \\ &= y_0 \frac{\tilde{a}^{-1}}{t+\tilde{a}^{-1}} + \frac{\tilde{a}^{-1}\tilde{b}}{t+\tilde{a}^{-1}} \int_0^t dW_u(s). \end{aligned} \tag{C9}$$

Changing back to the original notations leads to

$$\mu_t - \gamma^* = \frac{\mu_0 - \gamma^*}{1 + A_1^2 \sigma_u^{-2} R_0 t} + \frac{A_1 \sigma_u^{-1} R_0}{1 + A_1^2 \sigma_u^{-2} R_0 t} \int_0^t dW_u(s). \tag{C10}$$

C.2. Detailed derivations of Proposition 3.

Now we estimate the additive parameter γ^* in (C1) utilizing stochastic parameterized equation method (27),

$$du = (A_0 u + A_1 \gamma)dt + \sigma_u dW_u, \tag{C11a}$$

$$d\gamma = (a_0 - a_1 \gamma)dt + \sigma_\gamma dW_\gamma, \tag{C11b}$$

where $a_1 > 0$ is to guarantee the mean stability of (C11b). The evolutions of the posterior mean and variance of γ have the closed form, according to (3),

$$d\mu_t = (a_0 - a_1 \mu_t)dt + R_t A_1 \sigma_u^{-2}[du - (A_0 u + A_1 \mu_t)dt], \tag{C12a}$$

$$dR_t = \left(2a_1 R_t + \sigma_\gamma^2 - A_1^2 \sigma_u^{-2} R_t^2\right) dt. \tag{C12b}$$

Clearly, for $a_1 > 0$ and $\sigma_\gamma \neq 0$, the algebraic equation

$$-A_1^2 \sigma_U^{-2} R_t^2 - 2a_1 R_t + \sigma_\Gamma^2 = 0 \tag{C13}$$

always having two real roots r_1, r_2 with different signs. Let's assume $r_1 > 0 > r_2$ and initial value $R_0 > r_1$. Utilizing separation of variables, the posterior variance R_t is solved,

$$R_t = r_2 + \frac{r_1 - r_2}{1 - \left(\frac{R_0 - r_1}{R_0 - r_2}\right) \cdot \exp\left(-A_1^2 \sigma_u^{-2}(r_1 - r_2)t\right)}. \tag{C14}$$

This implies R_t will converge to the equilibrium state in an exponential way.

To solve the error in the posterior mean, we use the equilibrium variance R_{eq} to replace R_t in (C12a) due to the fact that R_t converges exponentially fast to R_{eq}. The qualitative conclusion does not change if we keep R_t in (C12a) but the expression will becomes extremely complicated. Again, noticing the true value γ^* is a constant and making use of the true dynamics (C1), the error in the posterior mean (C12a) becomes

$$\begin{aligned} d(\mu_t-\gamma^*) &= (a_0-a_1\mu_t)dt + R_{eq}A_1\sigma_u^{-2}[dU-(A_0U+A_1\mu_t)dt] \\ &= (a_0-a_1(\mu_t-\gamma^*)-a_1\gamma^*)\,dt + R_{eq}A_1\sigma_u^{-2}[(A_0u+A_1\gamma^*)dt+\sigma_u dW_u-(A_0u+A_1\mu_t)dt] \\ &= -\left(a_1+R_{eq}A_1^2\sigma_u^{-2}\right)(\mu_t-\gamma^*)dt + (a_0-a_1\gamma^*)dt + R_{eq}A_1\sigma_u^{-1}dW_u. \end{aligned} \tag{C15}$$

Utilizing integrating factor method, we arrive at the solution

$$\begin{aligned} \mu_t-\gamma^* =& (\mu_0-\gamma^*)e^{-(a_1+R_{eq}A_1^2\sigma_u^{-2})t} + \frac{1-e^{-(a_1+R_{eq}A_1^2\sigma_u^{-2})t}}{a_1+R_{eq}A_1^2\sigma_u^{-2}}(a_0-a_1\gamma^*) \\ &+ R_{eq}A_1\sigma_u^{-1}\int_0^t e^{-(a_1+R_{eq}A_1^2\sigma_u^{-2})(t-s)}ds \end{aligned} \tag{C16}$$

C.3. Detailed derivations of Proposition 4.

Now we estimate the multiplicative parameter γ^* in the linear system

$$du = (A_0-\gamma^* u)dt + \sigma_u dW_u. \tag{C17}$$

The augmented system by utilizing the direct method (26) yields

$$du = (A_0-\gamma u)dt + \sigma_u dW_u, \tag{C18a}$$

$$d\gamma = 0. \tag{C18b}$$

The evolutions of mean and variance of γ are given by, according to (3),

$$d\mu_t = -uR_t\cdot\sigma_u^{-2}\cdot[du-(A_0-u\mu_t)dt], \tag{C19a}$$

$$dR_t = -u^2\sigma_u^{-2}R_t^2dt. \tag{C19b}$$

In light of the method of separation of variables, equation (C19b) leads to the solution for the posterior variance,

$$R_t = \frac{R_0}{1 + R_0 \sigma_u^{-2} \int_0^t u^2(s) ds}. \tag{C20}$$

To solve the error in the posterior mean, we make use of (C17), (C19a) and (C20),

$$\begin{aligned} d(\mu_t - \gamma^*) &= -R_t u^2 \sigma_u^{-2}(\mu_t - \gamma^*)dt - R_t U \sigma_u^{-1} dW_u. \\ &= -\frac{R_0 \sigma_u^{-2} U^2}{1 + R_0 \sigma_u^{-2} \int_0^t u^2(s) ds}(\mu_t - \gamma^*)dt - \frac{R_0 \sigma_u^{-1} U}{1 + R_0 \sigma_u^{-2} \int_0^t u^2(s) ds} dW_u \end{aligned} \tag{C21}$$

For the simplicity of notation, we again define

$$y := \mu_t - \gamma^*, \qquad \tilde{a} := R_0 \sigma_u^{-2}, \qquad \tilde{b} := R_0 \sigma_u^{-1}, \qquad \text{and} \qquad F_u(t) := \int_0^t u^2(s) ds.$$

Then (C21) becomes

$$dy = -\frac{\tilde{a} u^2(t)}{1 + \tilde{a} F_u(t)} y dt - \frac{\tilde{b} u(t)}{1 + \tilde{a} F_u(t)} dW_u. \tag{C22}$$

The solution of (C22) is given by

$$y = y_0 e^{-\int_0^t \frac{\tilde{a} u^2(s)}{1 + \tilde{a} F_u(s)} ds} + e^{-\int_0^t \frac{\tilde{a} u^2(s)}{1 + \tilde{a} F_u(s)} ds} \int_0^t -\frac{\tilde{b} u(s)}{1 + \tilde{a} F_u(s)} e^{\int_0^s \frac{\tilde{a} u^2(v)}{1 + \tilde{a} F_u(v)} dv} dW_u(s) \tag{C23}$$

Note that

$$\int_0^t \frac{\tilde{a} u^2(s)}{1 + \tilde{a} F_u(s)} ds = \ln(1 + \tilde{a} F_u(t)).$$

Therefore, (C23) reduces to

$$y = \frac{y_0}{1 + \tilde{a} F_u(t)} + \frac{1}{1 + \tilde{a} F_u(t)} \int_0^t -\frac{\tilde{b} u(s)}{1 + \tilde{a} F_u(s)} (1 + \tilde{a} F_u(s)) dW_u(s). \tag{C24}$$

Changing back to the original notations, (C24) leads to

$$\mu_t - \gamma^* = \frac{\mu_0 - \gamma^*}{1 + R_0 \sigma_U^{-2} \int_0^t u^2(s) ds} - \frac{R_0 \sigma_u^{-1}}{1 + R_0 \sigma_u^{-2} \int_0^t u^2(s) ds} \int_0^t u(s) dW_u(s). \tag{C25}$$

APPENDIX D

Detailed derivations of Proposition 6 in Section 6c of estimating the multiplicative parameter in the special cubic system utilizing direct approach

The derivation of (47) in Proposition 6 follows those in (C20) and (C25). Here, we derive (49).

Recall the analytic expression of the equilibrium PDF $p_{eq}(u)$ is given by (Majda et al. 2009),

$$p_{eq}(u) = N_0 \exp\left(\frac{2}{\sigma_u^2}\left(-\frac{\gamma^*}{4}u^4\right)\right).$$

The integral factor N_0 is given by

$$N_0^{-1} = \int_{-\infty}^{\infty} e^{-\frac{\gamma^* u^4}{2\sigma_u^2}}\, du = 2\int_0^{\infty} e^{-\frac{\gamma^* u^4}{2\sigma_u^2}}\, du = \frac{1}{2}\int_0^{\infty} u^{-3} e^{-\frac{\gamma^* u^4}{2\sigma_u^2}}\, du^4. \tag{D1}$$

Let

$$x = \frac{\gamma^*}{2\sigma_u^2}u^4 \tag{D2}$$

and correspondingly

$$u = \left(\frac{2\sigma_u^2}{\gamma^*}x\right)^{\frac{1}{4}}. \tag{D3}$$

Inserting (D3) into (D1) results in

$$N_0^{-1} = \int_0^{\infty} \left(\frac{2\sigma_u^2}{\gamma^*}\right)^{-\frac{3}{4}} x^{-\frac{3}{4}} e^{-x} \frac{2\sigma_u^2}{\gamma^*} dx = \frac{1}{2}\left(\frac{2\sigma_u^2}{\gamma^*}\right)^{\frac{1}{4}} \int_0^{\infty} x^{-\frac{3}{4}} e^{-x} dx. \tag{D4}$$

Recall the definition of Γ function (Abramowitz et al. 1965)

$$\Gamma(s) = \int_0^{\infty} x^{s-1} e^{-x} dx. \tag{D5}$$

Then (D4) becomes

$$N_0^{-1} = \frac{1}{2}\left(\frac{2\sigma_u^2}{\gamma^*}\right)^{\frac{1}{4}} \Gamma\left(\frac{1}{4}\right) = 2^{-\frac{3}{4}}(\gamma^*)^{-\frac{1}{4}}\sigma_u^{\frac{1}{2}}\Gamma\left(\frac{1}{4}\right).$$

This leads to

$$N_0 = 2^{\frac{3}{4}}(\gamma^*)^{\frac{1}{4}}\sigma_u^{-\frac{1}{2}}\left(\Gamma\left(\frac{1}{4}\right)\right)^{-1}. \tag{D6}$$

With the formula of in N_0 (D6), we are able to solve $\int_{-\infty}^{\infty} u^6 p_{eq}(u)du$,

$$\begin{aligned}
\int_{-\infty}^{\infty} u^6 p_{eq}(u)du &= 2N_0 \int_0^{\infty} u^6 e^{-\frac{\gamma^*}{2\sigma_u^2}u^4} du \\
&= \frac{1}{2}N_0 \int_0^{\infty} u^3 e^{-\frac{\gamma^*}{2\sigma_u^2}u^4} du^4 \\
&= \frac{1}{2}N_0 \int_0^{\infty} \left(\frac{2\sigma_u^2}{\gamma^*}x\right)^{\frac{3}{4}} e^{-x}\frac{2\sigma_u^2}{\gamma^*}dx \qquad \text{using (D3)} \\
&= \frac{1}{2}N_0 \left(\frac{2\sigma_u^2}{\gamma^*}\right)^{\frac{7}{4}} \int_0^{\infty} x^{\frac{3}{4}} e^{-x}dx \\
&= 2^{\frac{3}{4}}(\gamma^*)^{-\frac{7}{4}}\sigma_u^{\frac{7}{2}}\Gamma\left(\frac{7}{4}\right)N_0 \\
&= 2^{\frac{3}{4}}(\gamma^*)^{-\frac{7}{4}}\sigma_u^{\frac{7}{2}}2^{\frac{3}{4}}(\gamma^*)^{\frac{1}{4}}\sigma_u^{-\frac{1}{2}}\left(\Gamma\left(\frac{1}{4}\right)\right)^{-1}\Gamma\left(\frac{7}{4}\right) \qquad \text{using (D6)} \\
&= 2^{\frac{3}{2}}(\gamma^*)^{-\frac{3}{2}}\sigma_u^3\left(\Gamma\left(\frac{1}{4}\right)\right)^{-1}\Gamma\left(\frac{7}{4}\right).
\end{aligned}$$

References

Abramov, R. V., and A. J. Majda, 2004: Quantifying uncertainty for non-Gaussian ensembles in complex systems. *SIAM Journal on Scientific Computing*, **26 (2)**, 411–447.

Abramowitz, M., I. A. Stegun, and Coauthors, 1965: *Handbook of mathematical functions*, Vol. 1046. Dover New York.

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Monthly weather review*, **129 (12)**, 2884–2903.

Berner, J., and G. Branstator, 2007: Linear and nonlinear signatures in the planetary wave dynamics of an AGCM: Probability density functions. *Journal of the atmospheric sciences*, **64 (1)**, 117–136.

Branicki, M., N. Chen, and A. J. Majda, 2013: Non-Gaussian test models for prediction and state estimation with model errors. *Chinese Annals of Mathematics, Series B*, **34 (1)**, 29–64.

Branicki, M., and A. Majda, 2013: Dynamic stochastic superresolution of sparsely observed turbulent systems. *Journal of Computational Physics*, **241**, 333–363.

Branicki, M., and A. Majda, 2014: Quantifying Bayesian filter performance for turbulent dynamical systems through information theory. *Comm. Math. Sci*, **12 (5)**, 901–978.

Branicki, M., and A. J. Majda, 2015: An information-theoretic framework for improving imperfect predictions via Multi-Model Ensemble forecasts. *Journal of Nonlinear Science*, **25 (3)**, 489–538.

Bretthorst, G. L., 2013: *Bayesian spectrum analysis and parameter estimation*, Vol. 48. Springer Science & Business Media.

Campillo, F., and V. Rossi, 2009: Convolution particle filter for parameter estimation in general state-space models. *Aerospace and Electronic Systems, IEEE Transactions on*, **45 (3)**, 1063–1072.

Chen, N., D. Giannakis, R. Herbei, and A. J. Majda, 2014a: An MCMC algorithm for parameter estimation in signals with hidden intermittent instability. *SIAM/ASA Journal on Uncertainty Quantification*, **2 (1)**, 647–669.

Chen, N., and A. J. Majda, 2015a: Filtering the stochastic skeleton model for the Madden-Julian oscillation. *Monthly Weather Review*, accepted.

Chen, N., and A. J. Majda, 2015b: Model error in filtering random compressible flows utilizing noisy Lagrangian tracers. *Monthly Weather Review*, submitted.

Chen, N., and A. J. Majda, 2015c: Predicting the cloud patterns for the boreal summer intraseasonal oscillation through a low-order stochastic model. *Mathematics of Climate and Weather Forecasting*, **1 (1)**, 1–20.

Chen, N., and A. J. Majda, 2015d: Predicting the real-time multivariate Madden-Julian oscillation index through a low-order nonlinear stochastic model. *Monthly Weather Review*, **143 (6)**, 2148–2169.

Chen, N., A. J. Majda, and D. Giannakis, 2014b: Predicting the cloud patterns of the Madden-Julian oscillation through a low-order nonlinear stochastic model. *Geophysical Research Letters*, **41 (15)**, 5612–5619.

Chen, N., A. J. Majda, and X. T. Tong, 2014c: Information barriers for noisy Lagrangian tracers in filtering random incompressible flows. *Nonlinearity*, **27 (9)**, 2133.

Chen, N., A. J. Majda, and X. T. Tong, 2015: Noisy Lagrangian tracers for filtering random rotating compressible flows. *Journal of Nonlinear Science*, **25 (3)**, 451–488.

Cressie, N., and C. K. Wikle, 2011: *Statistics for spatio-temporal data*. John Wiley & Sons.

Daum, F., and J. Huang, 2003: Curse of dimensionality and particle filters. *Aerospace Conference, 2003. Proceedings. 2003 IEEE*, IEEE, Vol. 4, 4_1979–4_1993.

Dee, D. P., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Monthly weather review*, **123 (4)**, 1128–1145.

Frisch, U., 1995: *Turbulence*. Cambridge university press.

Gajic, Z., and M. Lelic, 1996: *Modern control systems engineering*. Prentice-Hall, Inc.

Golightly, A., and D. J. Wilkinson, 2008: Bayesian inference for nonlinear multivariate diffusion models observed with error. *Computational Statistics & Data Analysis*, **52 (3)**, 1674–1693.

Harlim, J., A. Mahdi, and A. J. Majda, 2014: An ensemble Kalman filter for statistical estimation of physics constrained nonlinear regression models. *Journal of Computational Physics*, **257**, 782–812.

Harlim, J., and A. Majda, 2008: Filtering nonlinear dynamical systems with linear stochastic models. *Nonlinearity*, **21 (6)**, 1281.

Harlim, J., and A. J. Majda, 2010a: Catastrophic filter divergence in filtering nonlinear dissipative systems. *Communications in Mathematical Sciences*, **8 (1)**, 27–43.

Harlim, J., and A. J. Majda, 2010b: Filtering turbulent sparsely observed geophysical flows. *Monthly Weather Review*, **138 (4)**, 1050–1083.

Hinze, J. O., and J. Hinze, 1959: *Turbulence: An introduction to its mechanism and theory*, Vol. 162. McGraw-Hill New York.

Hyndman, R. J., and A. B. Koehler, 2006: Another look at measures of forecast accuracy. *International Journal of Forecasting*, **22 (4)**, 679–688.

Iglesias, M. A., K. Lin, and A. M. Stuart, 2014: Well-posed Bayesian geometric inverse problems arising in subsurface flow. *Inverse Problems*, **30 (11)**, 114 001.

Kalnay, E., 2003: *Atmospheric modeling, data assimilation, and predictability*. Cambridge university press.

Keating, S. R., A. J. Majda, and K. S. Smith, 2012: New methods for estimating ocean eddy heat transport using satellite altimetry. *Monthly Weather Review*, **140 (5)**, 1703–1722.

Kondrashov, D., S. Kravtsov, A. W. Robertson, and M. Ghil, 2005: A hierarchy of data-based ENSO models. *Journal of climate*, **18 (21)**, 4425–4444.

Kravtsov, S., D. Kondrashov, and M. Ghil, 2005: Multilevel regression modeling of nonlinear processes: Derivation and applications to climatic variability. *Journal of Climate*, **18 (21)**, 4404–4424.

Liptser, R. S., and A. N. Shiryaev, 2001: *Statistics of Random Processes II: II. Applications*, Vol. 2. Springer.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *Journal of the atmospheric sciences*, **20 (2)**, 130–141.

MacKay, D. J., 2003: *Information theory, inference and learning algorithms*. Cambridge university press.

Majda, A., 2003: *Introduction to PDEs and Waves for the Atmosphere and Ocean*, Vol. 9. American Mathematical Soc.

Majda, A., R. V. Abramov, and M. J. Grote, 2005: *Information theory and stochastics for multiscale nonlinear systems*, Vol. 25. American Mathematical Soc.

Majda, A., R. Kleeman, and D. Cai, 2002a: A mathematical framework for quantifying predictability through relative entropy. *Methods and Applications of Analysis*, **9 (3)**, 425–444.

Majda, A., I. Timofeyev, and E. Vanden-Eijnden, 2002b: A priori tests of a stochastic mode reduction strategy. *Physica D: Nonlinear Phenomena*, **170 (3)**, 206–252.

Majda, A., and X. Wang, 2006: *Nonlinear dynamics and statistical theories for basic geophysical flows*. Cambridge University Press.

Majda, A. J., 2012: Challenges in climate science and contemporary applied mathematics. *Communications on Pure and Applied Mathematics*, **65 (7)**, 920–948.

Majda, A. J., 2015: Statistical energy conservation principle for inhomogeneous turbulent dynamical systems. *Proceedings of the National Academy of Sciences*, **112 (29)**, 8937–8941.

Majda, A. J., R. Abramov, and B. Gershgorin, 2010a: High skill in low-frequency climate response through fluctuation dissipation theorems despite structural instability. *Proceedings of the National Academy of Sciences*, **107 (2)**, 581–586.

Majda, A. J., and M. Branicki, 2012: Lessons in uncertainty quantification for turbulent dynamical systems. *Discrete Cont. Dyn. Systems*, **32 (9)**.

Majda, A. J., C. Franzke, and D. Crommelin, 2009: Normal forms for reduced stochastic climate models. *Proceedings of the National Academy of Sciences*, **106 (10)**, 3649–3653.

Majda, A. J., and B. Gershgorin, 2010: Quantifying uncertainty in climate change science through empirical information theory. *Proceedings of the National Academy of Sciences*, **107 (34)**, 14 958–14 963.

Majda, A. J., and B. Gershgorin, 2013: Elementary models for turbulent diffusion with complex physical features: eddy diffusivity, spectrum and intermittency. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, **371 (1982)**, 20120 184.

Majda, A. J., and I. Grooms, 2014: New perspectives on superparameterization for geophysical turbulence. *Journal of Computational Physics*, **271**, 60–77.

Majda, A. J., and J. Harlim, 2012: *Filtering Complex Turbulent Systems*. Cambridge University Press.

Majda, A. J., and J. Harlim, 2013: Physics constrained nonlinear regression models for time series. *Nonlinearity*, **26 (1)**, 201–271.

Majda, A. J., J. Harlim, and B. Gershgorin, 2010b: Mathematical strategies for filtering turbulent dynamical systems. *Discrete and Continuous Dynamical Systems*, **27 (2)**, 441–486.

Majda, A. J., and Y. Lee, 2014: Conceptual dynamical models for turbulence. *Proceedings of the National Academy of Sciences*, **111 (18)**, 6548–6553.

Majda, A. J., D. Qi, and T. P. Sapsis, 2014: Blended particle filters for large-dimensional chaotic dynamical systems. *Proceedings of the National Academy of Sciences*, **111 (21)**, 7511–7516.

Majda, A. J., I. Timofeyev, and E. V. Eijnden, 1999: Models for stochastic climate prediction. *Proceedings of the National Academy of Sciences*, **96 (26)**, 14 687–14 691.

Majda, A. J., I. Timofeyev, and E. Vanden Eijnden, 2001: A mathematical framework for stochastic climate models. *Communications on Pure and Applied Mathematics*, **54 (8)**, 891–974.

Majda, A. J., I. Timofeyev, and E. Vanden-Eijnden, 2003: Systematic strategies for stochastic mode reduction in climate. *Journal of the Atmospheric Sciences*, **60 (14)**, 1705–1722.

Majda, A. J., and Y. Yuan, 2012: Fundamental limitations of ad hoc linear and quadratic multi-level regression models for physical systems. *Discrete and Continuous Dynamical Systems B*, **17 (4)**, 1333–1363.

Marquardt, D. W., 1963: An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial & Applied Mathematics*, **11 (2)**, 431–441.

Neelin, J. D., B. R. Lintner, B. Tian, Q. Li, L. Zhang, P. K. Patra, M. T. Chahine, and S. N. Stechmann, 2010: Long tails in deep columns of natural and anthropogenic tropospheric tracers. *Geophysical Research Letters*, **37 (5)**.

Palmer, T. N., 2001: A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. *Quarterly Journal of the Royal Meteorological Society*, **127 (572)**, 279–304.

Plett, G. L., 2004: Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs: Part 3. State and parameter estimation. *Journal of power sources*, **134 (2)**, 277–292.

Qi, D., and A. J. Majda, 2015: Blended particle methods with adaptive subspaces for filtering turbulent dynamical systems. *Physica D: Nonlinear Phenomena*, **298**, 21–41.

Salamon, P., and L. Feyen, 2009: Assessing parameter, precipitation, and predictive uncertainty in a distributed hydrological model using sequential data assimilation with the particle filter. *Journal of Hydrology*, **376 (3)**, 428–442.

Salmon, R., 1998: *Lectures on geophysical fluid dynamics*. Oxford University Press.

Sapsis, T. P., and A. J. Majda, 2013a: Blending modified Gaussian closure and non-Gaussian reduced subspace methods for turbulent dynamical systems. *Journal of Nonlinear Science*, **23 (6)**, 1039–1071.

Sapsis, T. P., and A. J. Majda, 2013b: Statistically accurate low-order models for uncertainty quantification in turbulent dynamical systems. *Proceedings of the National Academy of Sciences*, **110 (34)**, 13 705–13 710.

Slivinski, L., E. Spiller, A. Apte, and B. Sandstede, 2015: A hybrid particle–ensemble Kalman filter for Lagrangian data assimilation. *Monthly Weather Review*, **143 (1)**, 195–211.

Smedstad, O. M., and J. J. O'Brien, 1991: Variational data assimilation and parameter estimation in an equatorial Pacific Ocean model. *Progress in Oceanography*, **26 (2)**, 179–241.

Snijders, T. A., 2011: *Multilevel analysis*. Springer.

Sowell, F., 1992: Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *Journal of Econometrics*, **53 (1)**, 165–188.

Tong, X. T., A. J. Majda, and D. Kelly, 2015: Nonlinear stability of the ensemble Kalman filter with adaptive covariance inflation. *Comm. Math. Sci.*, accepted.

Townsend, A. A., 1980: *The structure of turbulent shear flow*. Cambridge university press.

Vallis, G. K., 2006: *Atmospheric and oceanic fluid dynamics: Fundamentals and large-scale circulation*. Cambridge University Press.

Van Der Merwe, R., E. Wan, and Coauthors, 2001: The square-root unscented Kalman filter for state and parameter-estimation. *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*, IEEE, Vol. 6, 3461–3464.

Wenzel, T. A., K. Burnham, M. Blundell, and R. Williams, 2006: Dual extended Kalman filter for vehicle state and parameter estimation. *Vehicle System Dynamics*, **44 (2)**, 153–171.

Wikle, C. K., and M. B. Hooten, 2010: A general science-based framework for dynamical spatio-temporal models. *Test*, **19 (3)**, 417–451.

Yeh, W. W.-G., 1986: Review of parameter identification procedures in groundwater hydrology: The inverse problem. *Water Resources Research*, **22 (2)**, 95–108.

LIST OF FIGURES

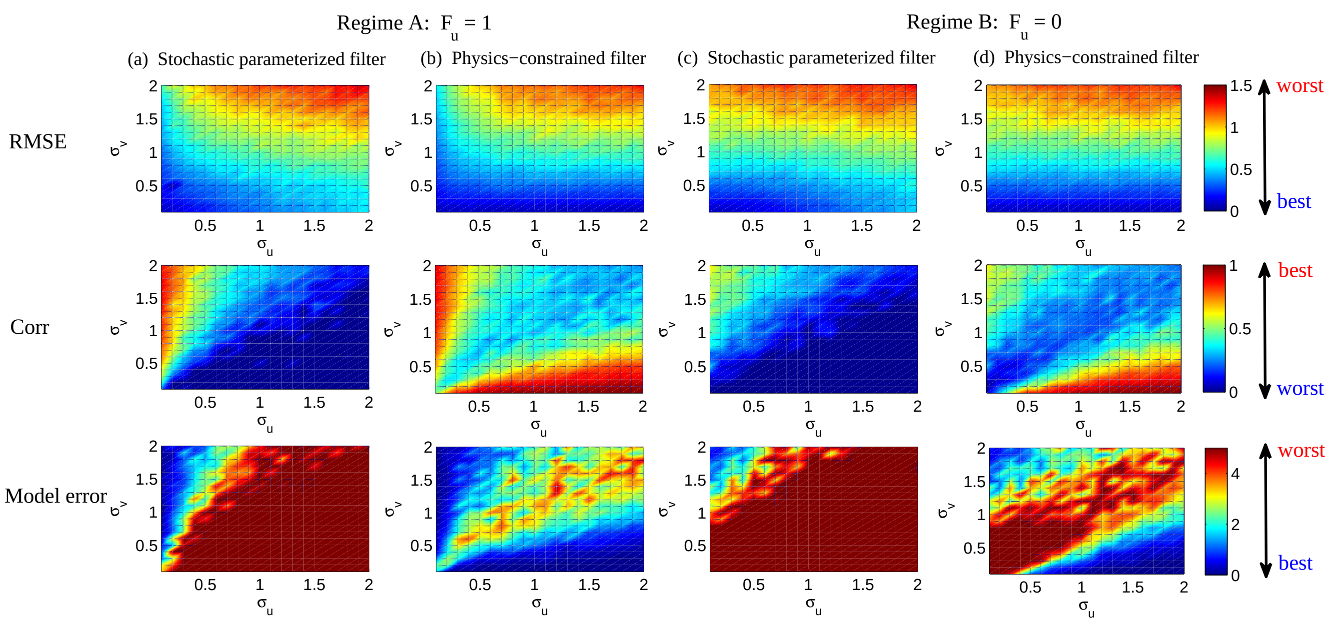


FIG. 1. Skill scores for filtering the unresolved process v using physics-constrained (perfect) filter (12) and stochastic parameterized (imperfect) filter (13) as a function of σ_u and σ_v in generating the truth. Column (a) and (b) show the skill scores in dynamics regime (A) while column (c) and (d) show those in dynamics regime (B). The first and second rows show the RMS error and pattern correlation in the filtered solution compared with the truth. The third row show the model error $\mathscr{P}(\pi,\pi^{filter})$ (6) in the time-averaged PDF of the posterior mean estimation π^{filter} compared with that of the truth π. The parameters $d_{vv}^M, \bar{v}^M$ and σ_v^M in the stochastic parameterized filter (13) is calibrated by matching the statistics with those of nature (12).

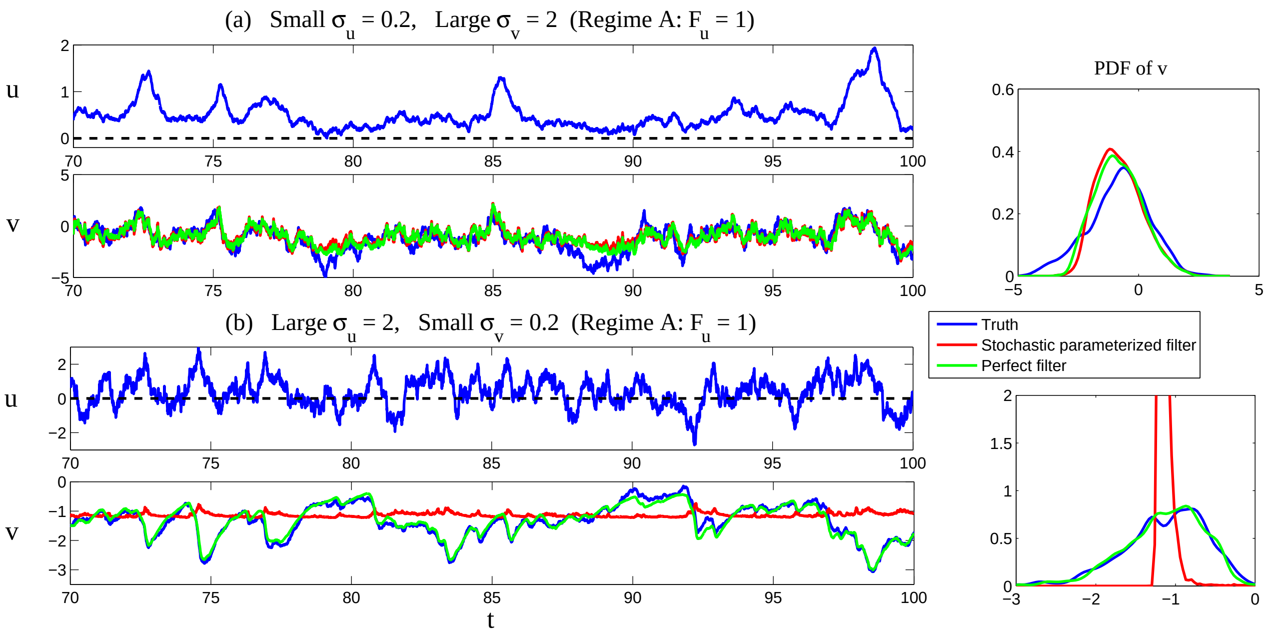


FIG. 2. (Dynamical regime A: with practical observability at the attractor). Comparison of the posterior mean estimation of v across time using physics-constrained (perfect) filter (12) and stochastic parameterized (imperfect) filter (13). The time-averaged PDFs of the filtered solutions compared with the truth are also illustrated. Panel (a) show the situation with small observational noise $\sigma_u = 0.2$ and large system noise $\sigma_v = 2$. Panel (b) shows the situation with large observational noise $\sigma_u = 2$ and small system noise $\sigma_v = 0.2$. The parameters $d_{vv}^M, \bar{v}^M$ and σ_v^M in the stochastic parameterized filter (13) is calibrated by matching the statistics with those of nature (12).

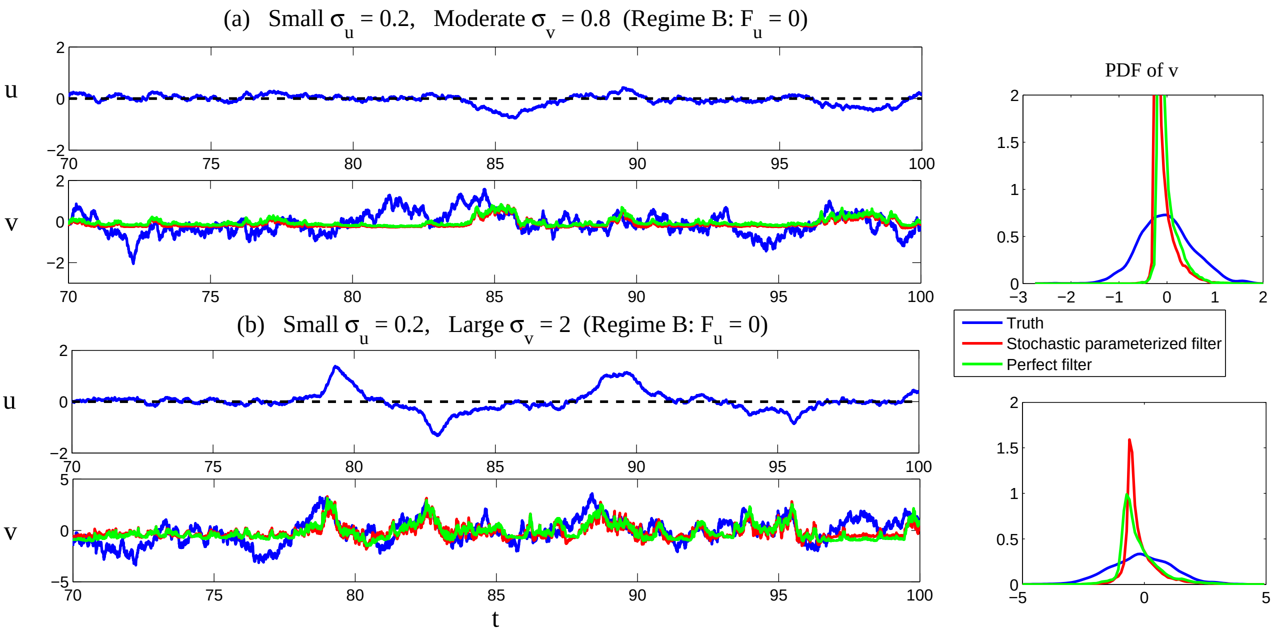


FIG. 3. (Dynamical regime B: without practical observability at the attractor). Similar to Figure 2. Panel (a) show the situation with small observational noise $\sigma_u = 0.2$ and moderate system noise $\sigma_v = 0.8$. Panel (b) shows the situation with small observational noise $\sigma_u = 0.2$ and large system noise $\sigma_v = 2$.

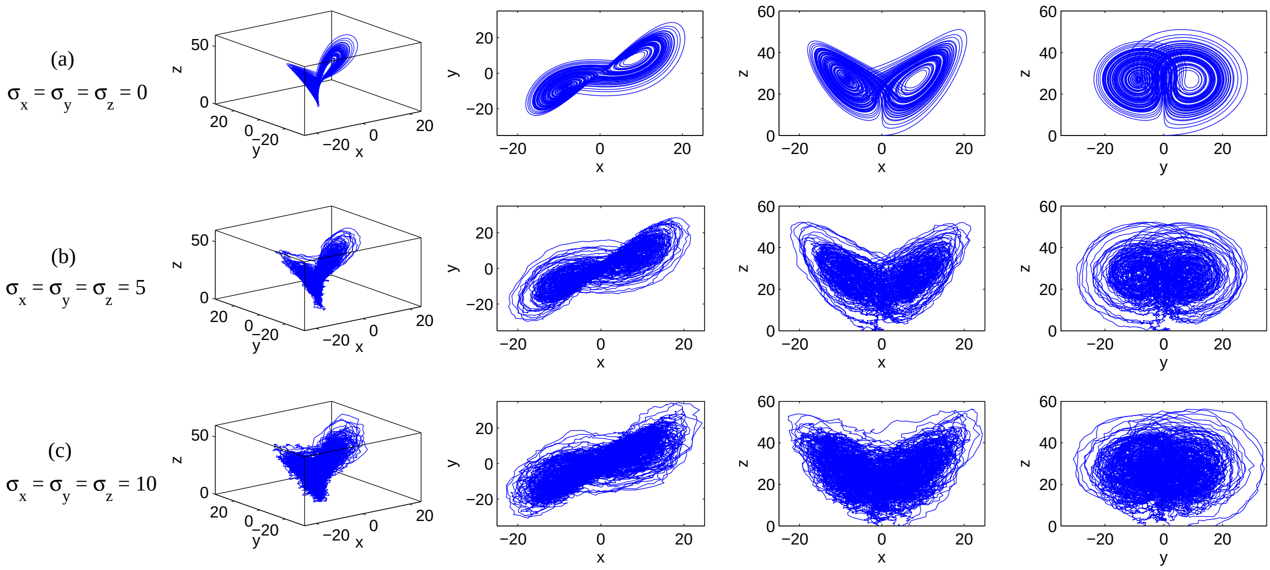


FIG. 4. Trajectories of the noisy L-63 model (16). Row (a): $\sigma_x = \sigma_y = \sigma_z = 0$; Row (b): $\sigma_x = \sigma_y = \sigma_z = 5$; Row (c): $\sigma_x = \sigma_y = \sigma_z = 10$.

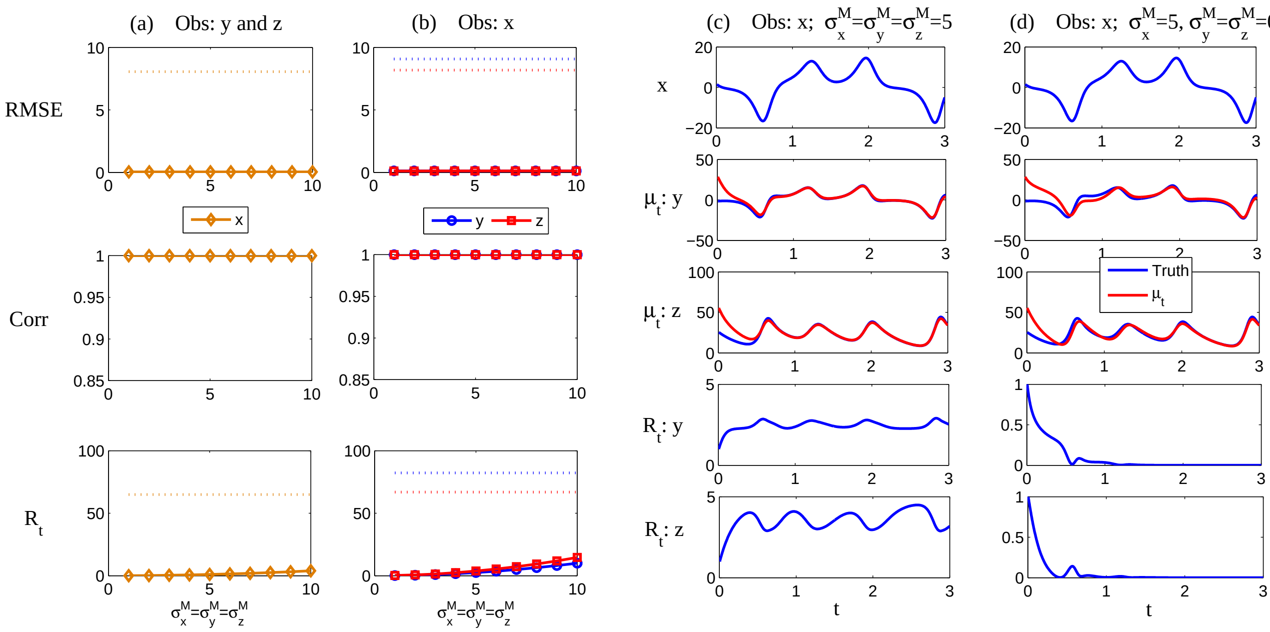


FIG. 5. Filtering the L-63 model (19) utilizing the imperfect noisy L-63 model (18). Column (a): observing y and z and filtering x. The three noise levels σ_x^M, σ_y^M and σ_z^M in the imperfect forecast model (18) are set to be the same, and the RMS error and pattern correlation in the posterior mean estimation compared with the truth as a function of these noise levels are shown in the first two rows, where the statistics are computed across time $t \in [5,50]$ and the dotted line shows the standard deviation of each variable at equilibrium. The third row shows the posterior variance as a function of the noise levels, where the statistics are averaged across time $t \in [5,50]$ and the dotted line shows the equilibrium covariance of each variable. Column (b): observing x and filtering y and z. Column (c): the observation in x across time and the filtering estimators in y and z with $\sigma_x^M = 5$ and nonzero system noise $\sigma_y^M = \sigma_z^M = 5$ in the imperfect forecast model (18). Column (d): the same as Column (c) but with zero system noise $\sigma_y^M = \sigma_z^M = 0$.

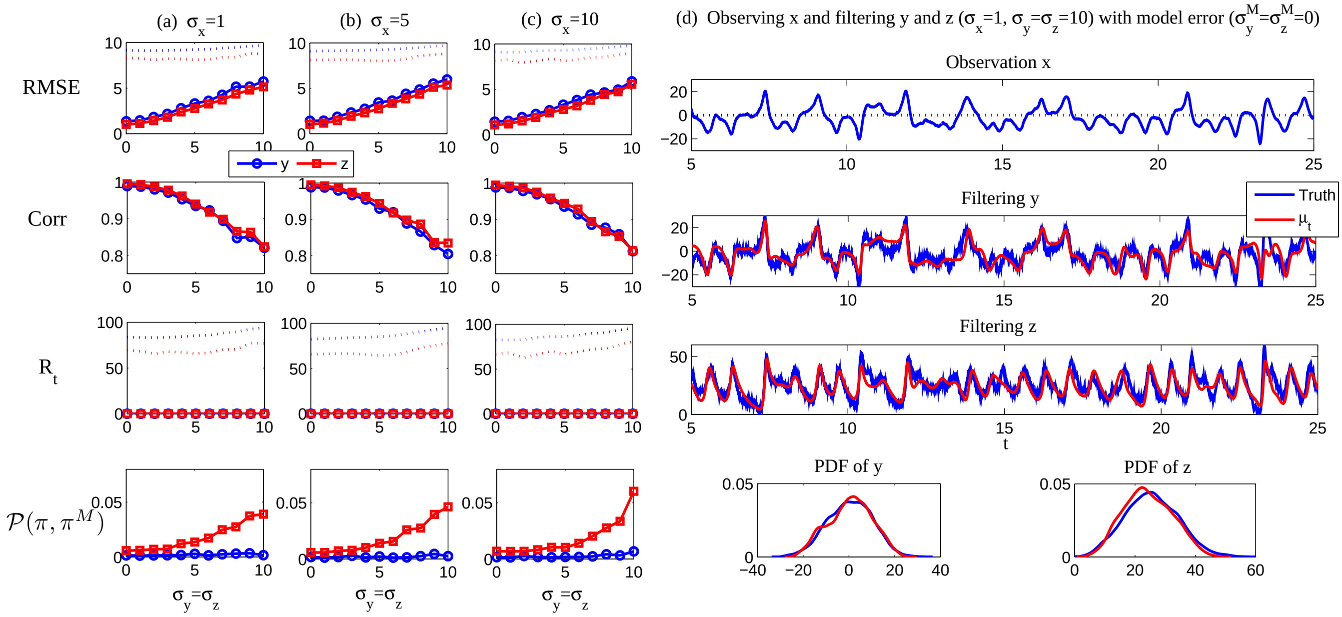


FIG. 6. Filtering the noisy L-63 model (16) utilizing the imperfect forecast L-63 model (21) with no system noise, where x is the observed variable and y and z are the variables for filtering. The observational noise σ_x^M in (21) is set to be the same σ_x in (16). Column (a)-(c) show the RMS error and pattern correlation in the posterior mean compared with the truth, the posterior covariance, and model error (6) with different σ_x as a function of the system noise σ_y and σ_z in the true model (16) where $\sigma_y = \sigma_z$. The dotted line in the first row shows the equilibrium standard deviation and that in the third row shows the equilibrium variance of each variable. The statistics are averaged across time $t \in [5, 50]$. Column (d) show the filtering estimation across time, where $\sigma_x = 1, \sigma_y = \sigma_z = 10$.

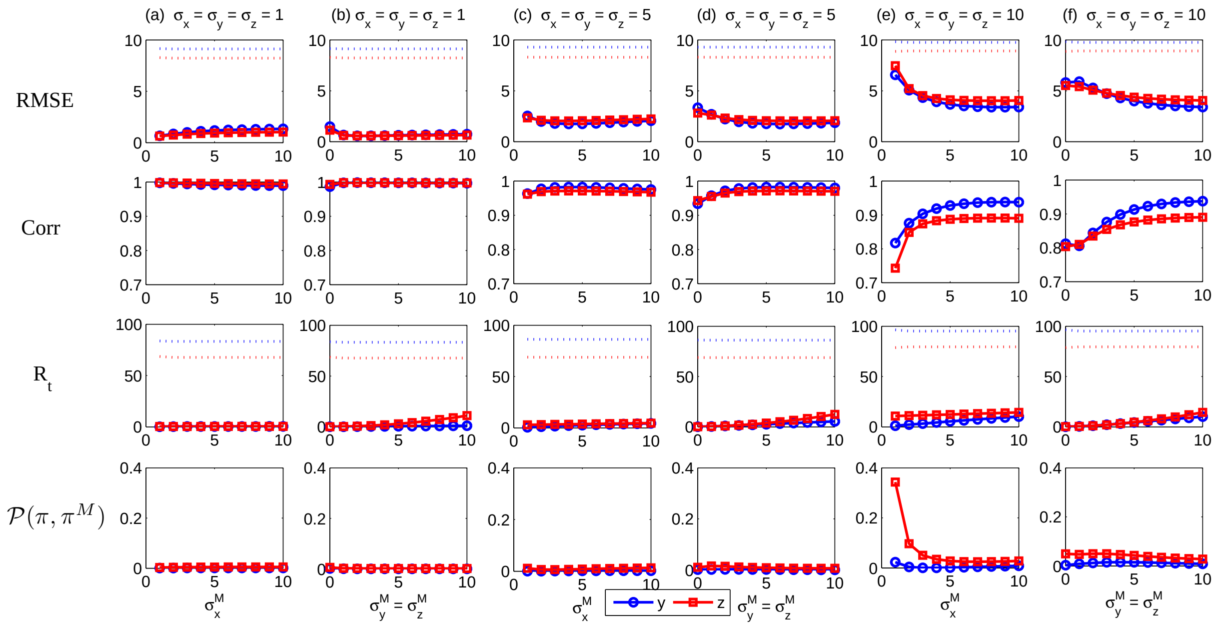


FIG. 7. Filtering the noisy L-63 model (16) utilizing an imperfect forecast model (18) with model error in the noise, where the observational variable is x and the variables for filtering are y and z. Column (a), (c) and (e): filtering skill as a function of the observational noise σ_x^M. Panel (b), (d) and (f): filtering skill as a function of the system noise σ_y^M and σ_z^M, where $\sigma_y^M = \sigma_z^M$. Panel (a) and (b), (c) and (d), and (e) and (f) show small, moderate and large noise $\sigma_x = \sigma_y = \sigma_z = 1, 5$ and 10 in the true system. The dotted line in the first row shows the equilibrium standard deviation and that in the third row shows the equilibrium variance of each variable. The statistics are averaged across time $t \in [5, 50]$.

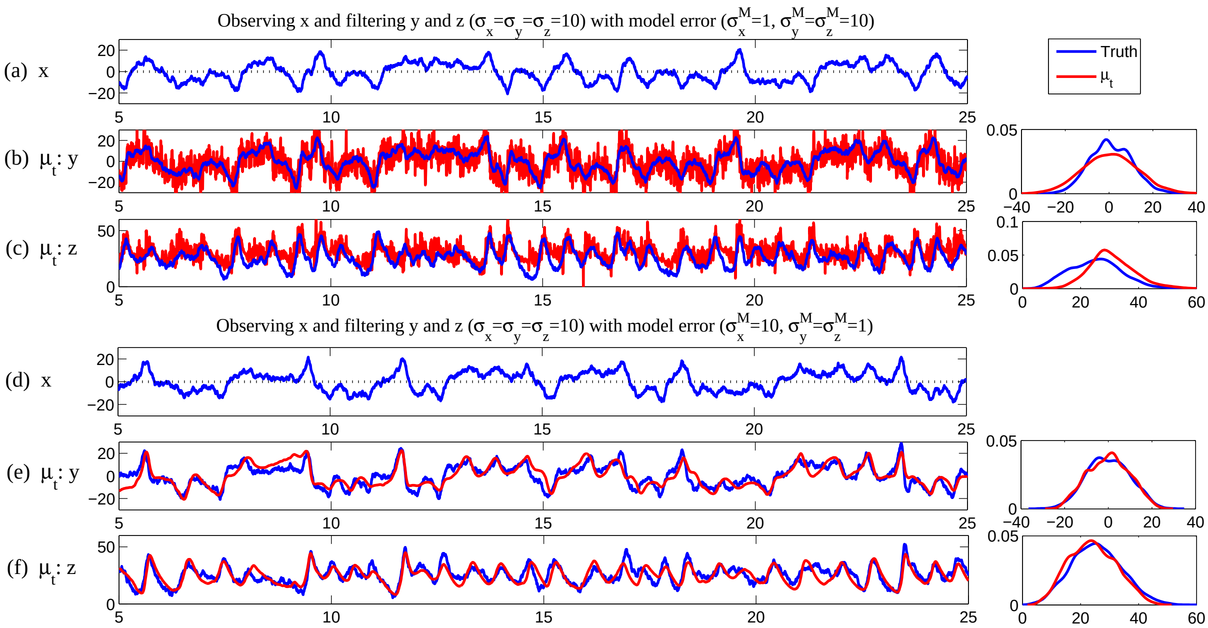


FIG. 8. Comparison of the true signal (blue) and posterior mean estimation (red) in the underdispersion cases, where $\sigma_x = \sigma_y = \sigma_z = 10$ in the model that generates true signal (16). Panel (a)-(c), filtering skill with underdispersed observational noise $\sigma_x^M = 1$. Panel (d)-(f): filtering skill with underdispersed system noise $\sigma_y^M = \sigma_z^M = 1$.

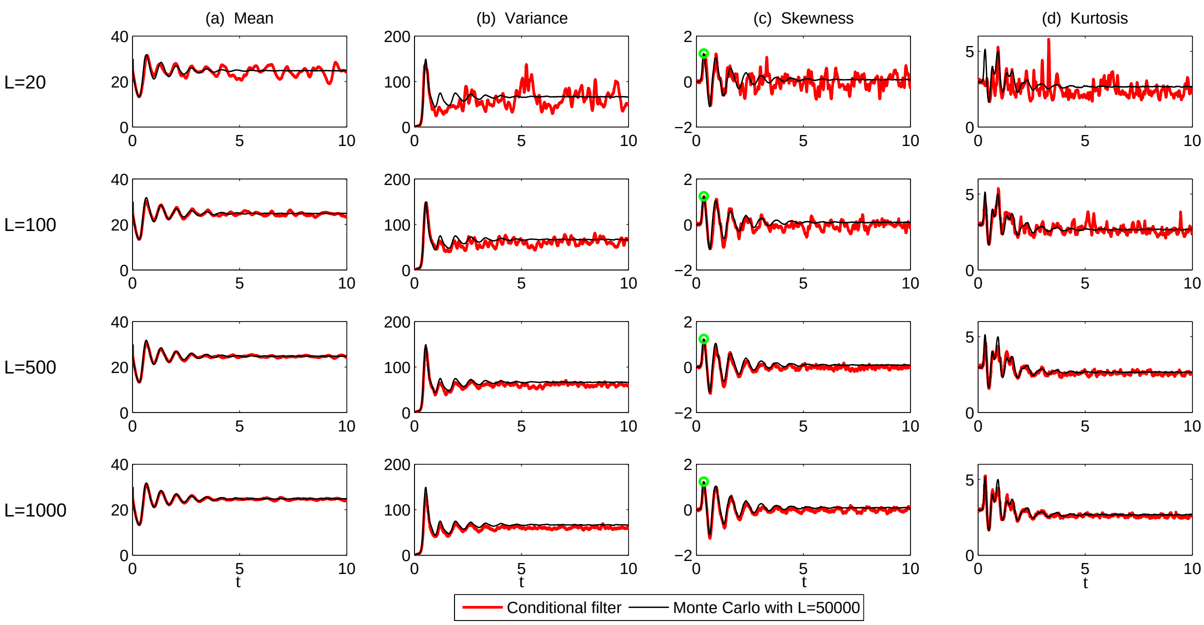


FIG. 9. Recovering of the mean, variance, skewness and kurtosis associated with the marginal PDF associated with the unobserved variable z in noisy L-63 model (16) utilizing the conditional Gaussian ensemble mixture approach (23) with different number of ensembles L in a perfect model setting. As comparison, the recovered statistics utilizing Monte Carlo simulation with $N = 50,000$ ensemble members are also included. The green dot in column (c) indicates the largest skewness in the transition phase, which will be utilized in Figure 10.

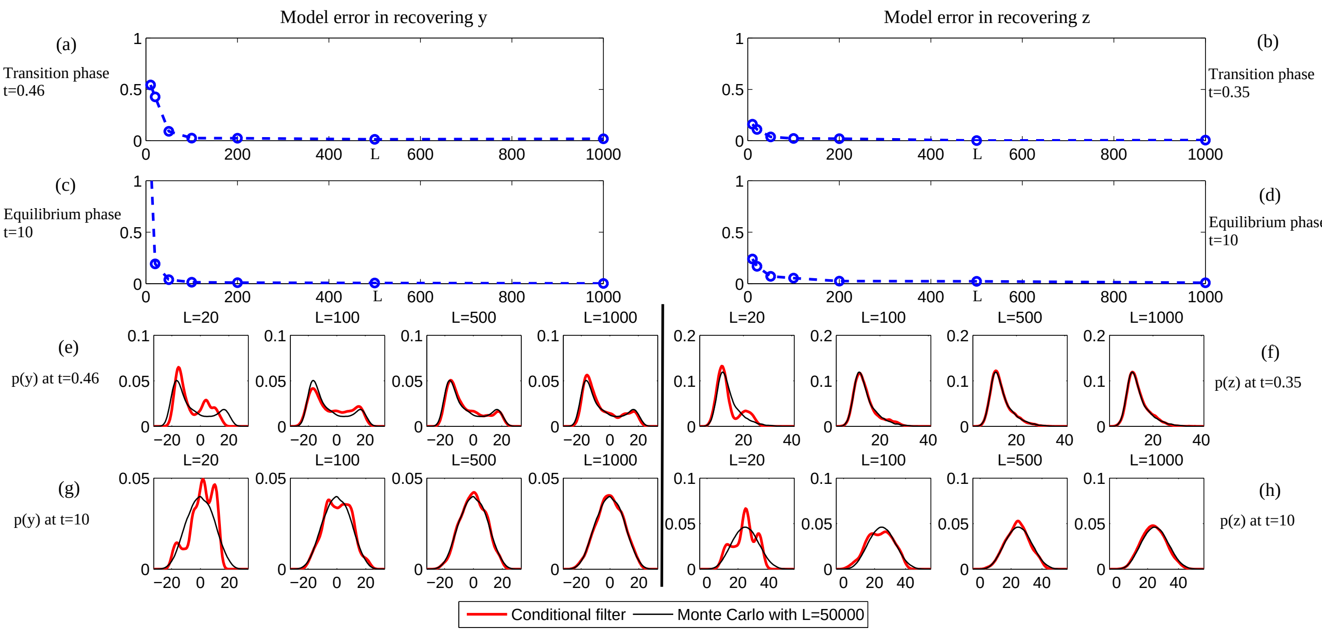
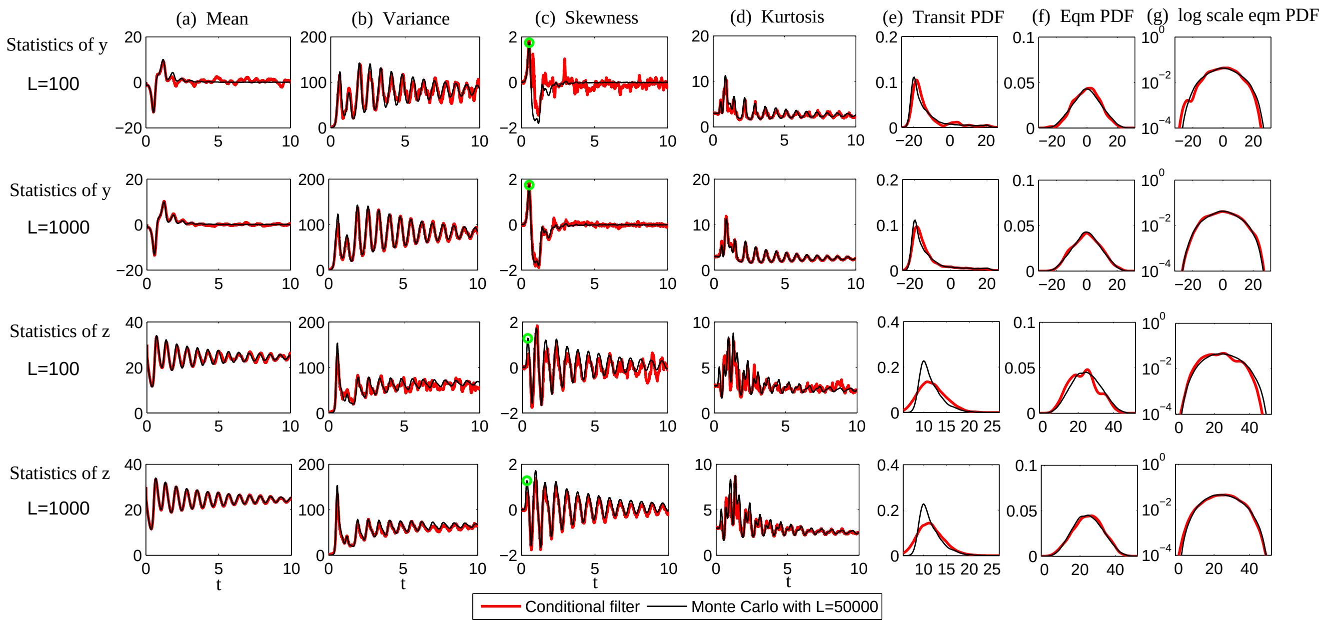


FIG. 10. Model error (6) in recovering the marginal PDF associated with the unobserved variables y and z in noisy L-63 model (16) utilizing the conditional Gaussian ensemble mixture approach (23) with different L in a perfect model setting . Panel (a): model error in the PDF of y as a function of L at a short-term transition phase $t = 0.46$ with largest skewness. Panel (e): comparing the PDFs of y at $t = 0.46$ utilizing conditional Gaussian mixture (23) and Monte Carlo with $N = 50,000$. Panel (c) and (g) are similar to Panel (a) and (e) but a time $t = 10$, at which the system reaches statistical equilibrium state. Panel (b), (d), (f) and (h) are for the marginal distribution associated with the unobserved variable z.



1276 FIG. 11. Recovery of the marginal PDFs associated with the unobserved variables y and z in the presence of
1277 model error from noise inflation. The noisy L-63 model (16) with $\sigma_x = \sigma_y = \sigma_z = 2$ is utilized in generating
1278 the true signal. The imperfect model (18) with $\sigma_x^M = \sigma_y^M = \sigma_z^M = 5$ is adopted for recovering the hidden
1279 PDFs. Column (a)-(d) show the recovered mean, variance, skewness and kurtosis compared with the truth that
1280 is computed by Monte Carlo simulation with $N = 50,000$ samples. Column (e) shows the recovered PDFs at
1281 short-term transition time with maximum skewness, where $t = 0.54$ for y and $t = 0.38$ for z. Column (f) shows
1282 the recovered PDFs at statistical equilibrium state $t = 25$ for both y and z and column (g) shows the PDFs at
1283 $t = 25$ in logarithm scale.

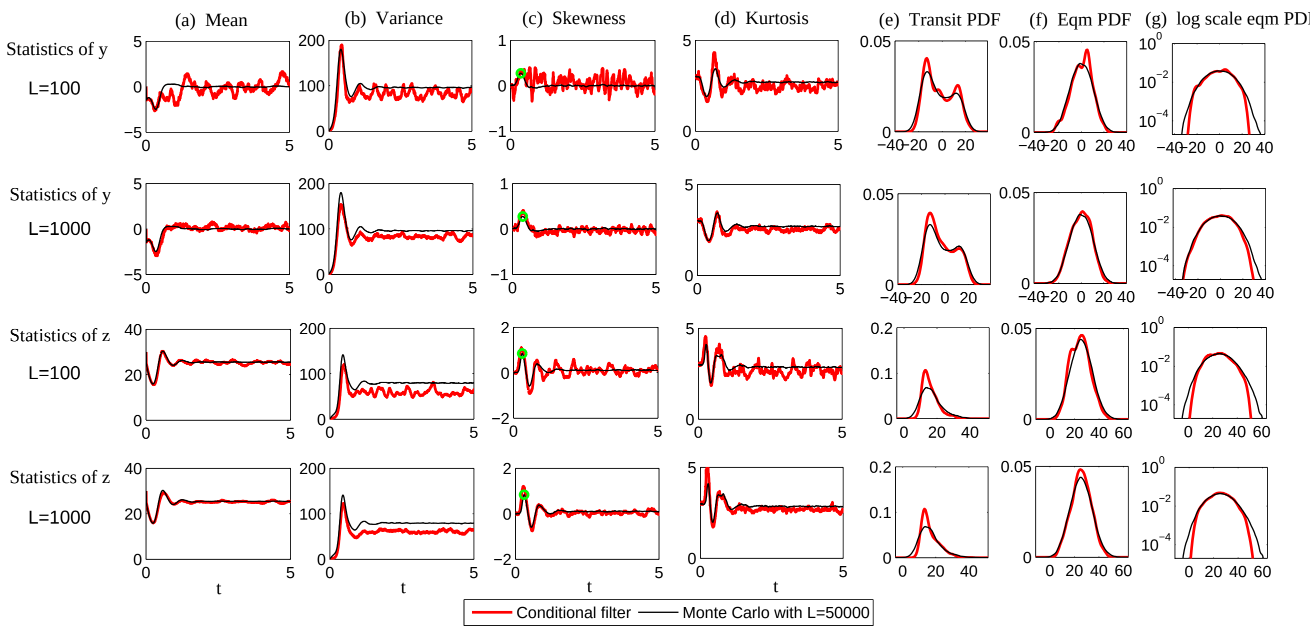


FIG. 12. Recovery of the marginal PDFs associated with the unobserved variables y and z in the presence of model error from underdispersion of noise. The noisy L-63 model (16) with $\sigma_x = \sigma_y = \sigma_z = 10$ is utilized in generating the true signal. The imperfect model (18) with $\sigma_x^M = \sigma_y^M = \sigma_z^M = 5$ is adopted for recovering the hidden PDFs. Column (a)-(d) show the recovered mean, variance, skewness and kurtosis compared with the truth that is computed by Monte Carlo simulation with $N = 50,000$ samples. Column (e) shows the recovered PDFs at short-term transition time with maximum skewness, where $t = 0.36$ for y and $t = 0.30$ for z. Column (f) shows the recovered PDFs at statistical equilibrium state $t = 5$ for both y and z and column (g) shows the PDFs at $t = 5$ in logarithm scale.

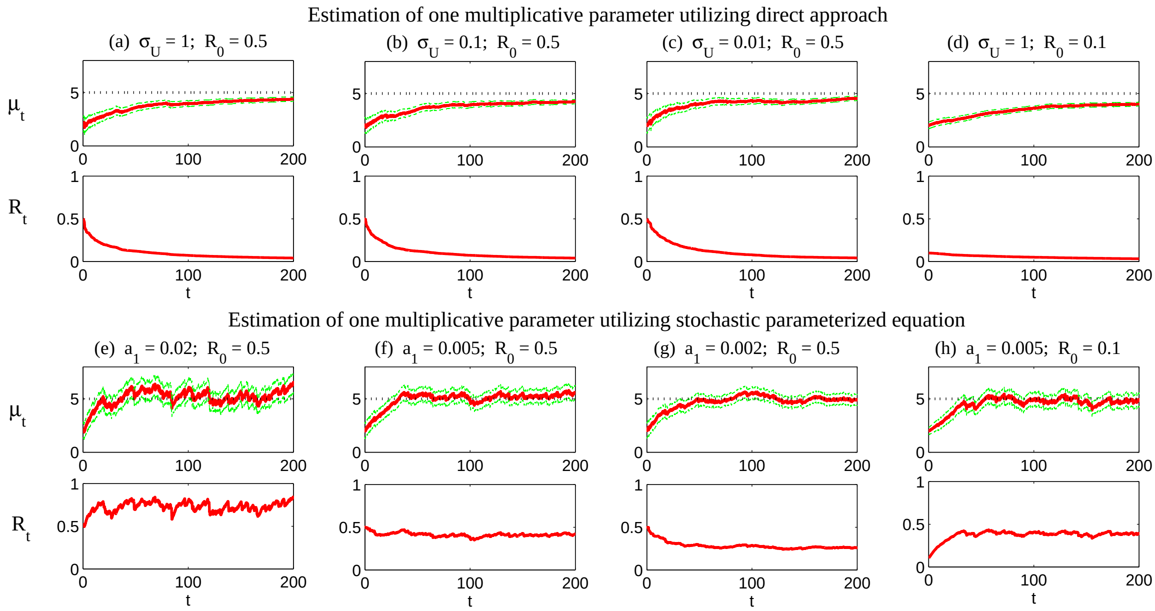
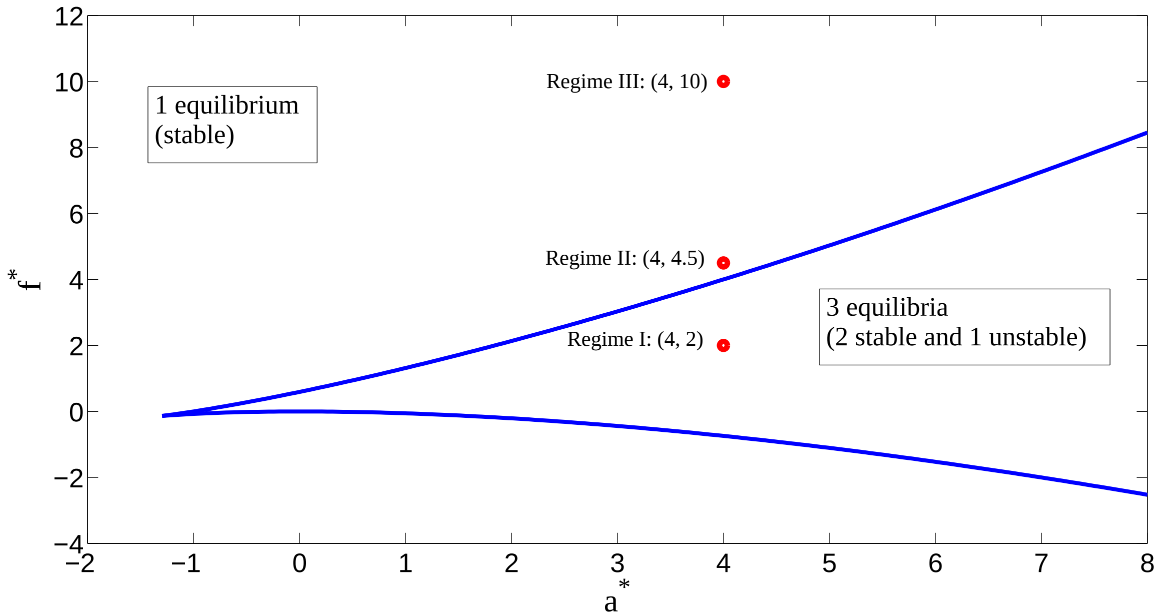


FIG. 13. Comparison of estimating the multiplicative parameter γ^* in (33). *Top* (panel (a)-(d)): estimation skill utilizing direct approach (34). *Bottom* (panel(e)-(h)): estimation skill utilizing stochastic parameterized equation (40). Here, the truth $\gamma^* = 5$ and $A_0 = 0$ are adopted. In the stochastic parameterized equation (40), the equilibrium mean $a_0/a_1 = 5.5$ and equilibrium $\sigma_\gamma^2/(2a_1) = 2$ are fixed. The black dotted line represents the truth $\gamma^* = 5$; the red curve is the posterior mean μ_t or posterior variance R_t; the two green curves around the posterior mean differs from the mean μ_t by one standard deviation, i.e., $\mu_t \pm \sqrt{R_t}$.



1298 FIG. 14. Phase portrait of (a^*, f^*) for the deterministic part of cubic model with $b^* = -4$ and $c^* = 4$ fixed.
1299 The three red dots are the three examples utilized in Figure 15 to study the parameter estimation skill.

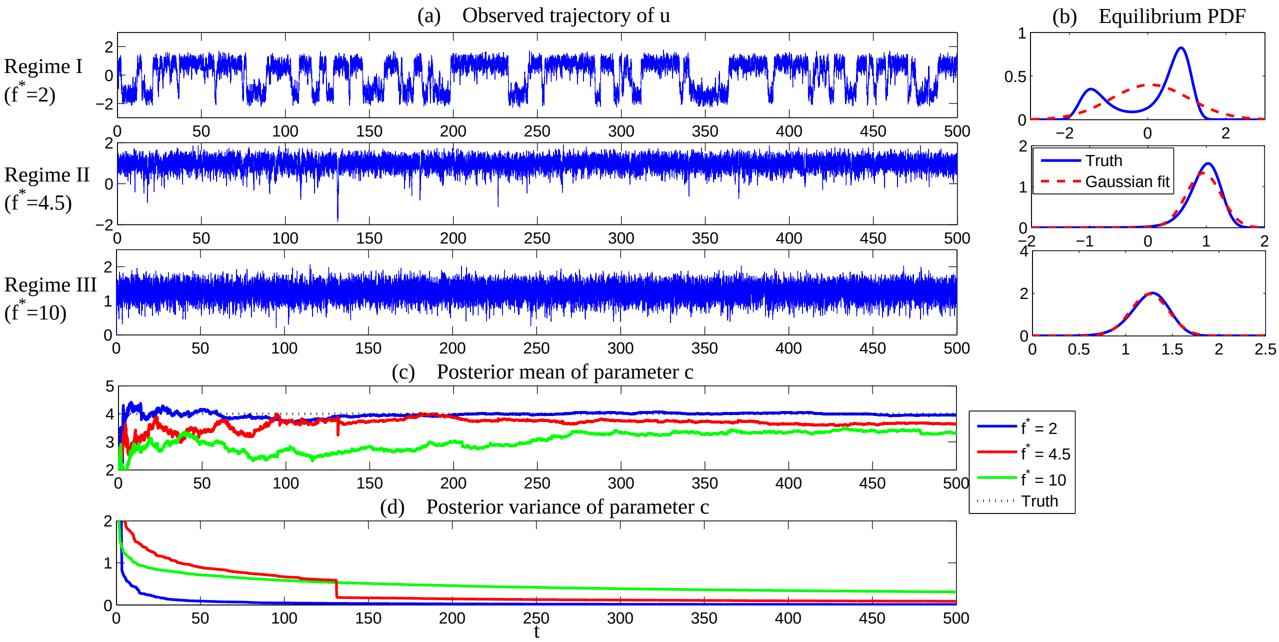


FIG. 15. Parameter estimation of the cubic model (43) in different dynamical regimes. Panel (a) shows the observational trajectories with $f^* = 2, f^* = 4.5$ and $f^* = 10$ as shown in Figure 14. The other parameters are $a^* = 4, b^* = -4, c^* = 4$ and $\sigma_u = \sqrt{2}$. Panel (b) shows the corresponding PDF. Panel (c) and (d) show the estimated parameters and the associated estimation uncertainty of parameter c.

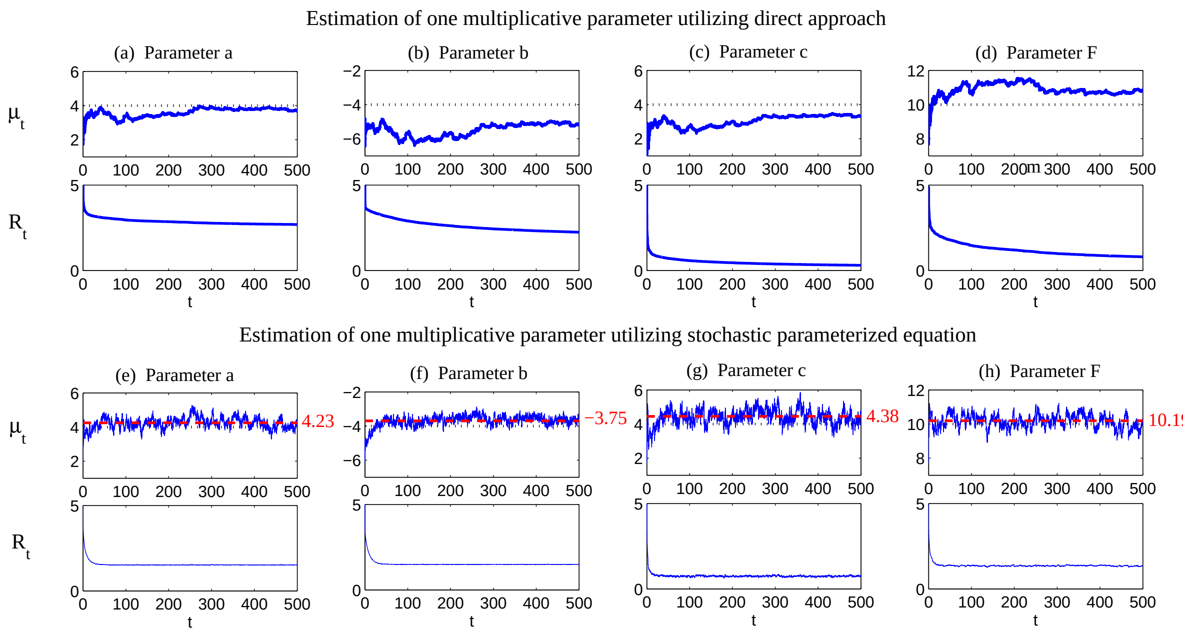


FIG. 16. Parameter estimation of the cubic nonlinear model (43) in Regime III with $(a^*, b^*, c^*, f^*) = (4, -4, 4, 10)$ with direct approach *(top)* and stochastic parameterized equation method *(bottom)*. The black dotted line shows the truth of each parameter and the red dashed line shows the averaged value of the estimation of each parameter utilizing stochastic parameterized equation method at equilibrium. Here the equilibrium mean of stochastic parameterized equation a_0/a_1 has 0.5 unit bias from the truth. The equilibrium variance of each stochastic parameterized equation is $\sigma_\gamma^2/(2a_1) = 2$. The damping coefficient in the stochastic parameterized equation is set to be $a_1 = 0.01$.