Predicting Observed and Hidden Extreme Events in Complex Nonlinear Dynamical Systems with Partial Observations and Short Training Time Series

Nan Chen[1, a)] and Andrew J. Majda[2, b)]

[1)] *Department of Mathematics, University of Wisconsin-Madison*

[2)] *Department of Mathematics and Center for Atmosphere Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, NY, USA, and Center for Prototype Climate Modeling, New York University Abu Dhabi, Saadiyat Island, Abu Dhabi, UAE*

(Dated: 30 July 2019)

Extreme events appear in many complex nonlinear dynamical systems. Predicting extreme events has important scientific significance and large societal impacts. In this paper, a new mathematical framework of building suitable nonlinear approximate models is developed, which aims at predicting both the observed and hidden extreme events in complex nonlinear dynamical systems for short-, medium- and long-range forecasting using only short and partially observed training time series. Different from many ad-hoc data-driven regression models, these new nonlinear models take into account physically motivated processes and physics constraints. They also allow efficient and accurate algorithms for parameter estimation, data assimilation and prediction. Cheap stochastic parameterizations, judicious linear feedback control and suitable noise inflation strategies are incorporated into the new nonlinear modeling framework, which provide accurate predictions of both the observed and hidden extreme events as well as the strongly non-Gaussian statistics in various highly intermittent nonlinear dyad and triad models, including the Lorenz 63 model. Then a stochastic mode reduction strategy is applied to a 21-dimensional nonlinear paradigm model for topographic mean flow interaction. The resulting 5-dimensional physics-constrained nonlinear approximate model is able to accurately predict the extreme events and the regime switching between zonally blocked and unblocked flow patterns. Finally, incorporating judicious linear stochastic processes into a simple nonlinear approximate model succeeds in learning certain complicated nonlinear effects of a 6-dimensional low-order Charney-DeVore model with strong chaotic and regime switching behavior. The simple nonlinear approximate model then allows accurate online state estimation and the short- and medium-range forecasting of extreme events.

PACS numbers: 02.50.Ey, 05.45.-a, 05.45.Tp, 92.60.Aa

[a] chennan@math.wisc.edu; Corresponding author

[b] jonjon@cims.nyu.edu

Extreme events appear in many complex nonlinear dynamical systems. These extreme events are associated with the sudden changes of states in the underlying complex systems and the occurrence of extreme events often results in large social impact. Therefore, predicting extreme events has both scientific significance and practical implications. However, the big challanges of prdicting the extreme events in complex nonlinear systems include the lack of understanding of physics, the huge computational cost in running the complex models and data assimilation, as well as the availability of only short and partially observed training data. In this paper, a new mathematical framework of building suitable nonlinear approximate models is developed, which aims at predicting both the observed and hidden extreme events in complex nonlinear dynamical systems for short-, medium- and long-range forecasting using only short and partially observed training time series. This framework also allows efficient and accurate data assimilation, parameter estimation and prediction algorithms. Different effective and practical strategies are incorporated into the framework to develop suitable approximate models for predicting extreme events and other non-Gaussian features in various complex turbulent dynamical systems.

I. INTRODUCTION

Extreme events appear in many complex nonlinear dynamical systems in geoscience, engineering, excitable media, neural science and material science[1–8]. Examples include oceanic rogue waves[9,12], extreme weather and climate patterns[10,11] such as blocking events and turbulent tracers[13–15], and bursting neurons[16]. These extreme events are associated with the sudden changes of states in the underlying complex systems and the occurrence of extreme events often results in large social impact. Therefore, predicting extreme events has both scientific significance and practical implications.

However, predicting the extreme events in complex nonlinear systems is quite challenging. First, nature or the perfect model is never known in practice. Model error due to the lack of the understanding of physics may prevent the skillful predictions of the extreme events[1,17–19]. Second, even if the perfect model is known, the underlying nonlinear dynamics of nature can be extremely complicated with strong non-Gaussian characteristics, multiscale features and high dimensionality[10,20,21]. Thus, running the perfect model is usually computationally unaffordable for real-time prediction. On the other hand, despite that coarse-graining the numerical resolutions improves the computational efficiency, such a numerical approximation often results in missing the key nonlinear interactions between different temporal and spatial scales and brings about large errors, especially for extreme events. Third, it is important to notice that only partial and noisy observations are available in many practical situations[22–24], which implies the states of the unobserved variables have to be estimated via online data assimilation algorithms. Unfortunately, the existing data assimilation algorithms for general complex nonlinear dynamical systems are either quite expensive (e.g., particle filter) or involving intrinsic approximate errors due to the coarse-grained statistics (e.g., ensemble Kalman filter)[21,25–28]. The assimilated states from the latter may also contain large biases due to the fact that high order moments are important contributors to the extreme events. Finally, the actual climate signal is often measured through time series. However, since the high-resolution satellites and other refined measurements were not widely developed until recent times, the available useful training data is very limited with about only 50 years in many real applications. Thus, predicting extreme events using short and partially observed training time series is another remarkably challenging task.

For the reasons given above, developing suitable approximate models for predicting extreme events is crucial in practice. These approximate models aim at capturing the key nonlinear dynamical and non-Gaussian statistical features of nature. They also need to be computationally tractable and allow efficient algorithms for online data assimilation, parameter estimation and prediction. There have been some recent progress in the extreme events prediction. For example, a new statistical dynamical model was developed to predict extreme events and anomalous features in shallow water waves[12]. A suite of reduced-order stochastic models was built, which succeeds in predicting the extreme events in complex geophysical flows[29] and their long-term non-Gaussian features[30] as well as forecasting the associated statistical responses and quantifying the uncertainty[31]. In addition, mode decomposition techniques were applied for probing the most unstable modes and building low-order models for extreme events prediction[32,33].

In this paper, a new mathematical framework of building suitable nonlinear approximate models is developed, which aims at predicting both the observed and hidden extreme events in complex nonlinear dynamical systems using only short and partially observed training time series. The models belonging to this mathematical framework are highly nonlinear and are able to capture many key non-Gaussian characteristics as observed in nature[34]. Unlike traditional regression and other ad hoc models with prescribed basis functions or structures, this framework contains a rich class of statistical dynamical models and is amenable to a wide range of applications. One important feature of this nonlinear modeling framework is that physically motivated processes and physics constraints[35,36] can be incorporated into the models, which is fundamentally different from many purely data-driven statistical models that have no clear physical meanings. Such a trait not only enables the models to take into account both the dynamical and statistical information but also allows using only a short training time series for model calibration. The latter is due to the (partially) identified dynamical structures from some physics reasoning and physics constraints. Another key advantage of this new framework is that despite the intrinsic nonlinearity, it allows closed analytic formulae for assimilating the states of the unobserved variables[37,38], which is computationally efficient and accurate. This provides an extremely useful and practical approach for predicting extreme events and other non-Gaussian features in complex nonlinear dynamical systems.

Short-, medium- and long-range forecasting of extreme events all have practical significance[10,39–41]. The efficient data assimilation scheme associated with the nonlinear models within the above framework provides an accurate estimation of the initial values, which play a crucial role in improving the short-term prediction skill. On the other hand, the focus of the long-term prediction is on the statistics, which is calculated by making use of a long trajectory together with the ergodic property of many complex turbulent systems[1]. In particular, reproducing the statistical equilibrium non-Gaussian probability density function (PDF) with fat tails is a good evidence of the successful prediction of extreme events, where the extreme events and intermittency are the main contributors to the fat tails. The medium-range forecast aims at recovering the transition behavior of the underlying dynamics. A skillful medium-range prediction requires both an accurate estimation of the initial values and a suitable description of the time evolution of the approximate model, and is often a challenging task. Finally, certain internal or external perturbations are able to kick the model variables outside the attractor. Therefore, predicting the time evolution of the extreme events that start from a state outside the attractor also has practical importance. It is worth remarking that many purely data-driven or machine learning methods fail to predict extreme events even though most of those methods show high skill in fitting the observed time series. For example, as has been pointed out in some recent work[42,43], even one of the most advanced neural networks with long short-term memory[44] and the Gaussian process regression[45] suffer from a finite time blowup issue when they are applied for predicting extreme events. Such a pathological behavior can only be overcome by using hybrid strategies that combine these methods with suitable models[42]. Note that these purely data-driven methods often demand tremendous training data[46,47], which is not practical in many scientific scenarios where only short training time series are available. In addition, without suitable models, predicting extreme events in the unobserved processes becomes extremely difficult.

This paper aims at incorporating practical strategies into the development of suitable approximate models for predicting both the observed and hidden extreme events. These approximate models belong to the new nonlinear modeling framework, which allows an efficient and accurate data assimilation scheme and only short and partially observed time series are needed for model calibration. The first effective strategy is to adopt simple stochastic parameterizations for approximating complicated hidden processes. Despite the simple forms, the judicious applications of these stochastic parameterizations are nevertheless able to capture the nonlinear interactions between the observed and hidden variables and predict the associated extreme events. Such an idea has been successfully applied to the stochastic parameterized extended Kalman filter (SPEKF) forecast models[48,49], dynamic stochastic superresolution of sparsely observed turbulent systems[50,51] and stochastic superparameterization for geophysical turbulence[52]. The second strategy here is motivated from control theory, which involves incorporating simple feedback control terms into the approximate models for model simplification. This simple feedback control strategy succeeds in capturing the key nonlinear statistical interactions as well as the causal effects between the observed and hidden variables, which are essential to accurately predicting the extreme events in the hidden processes. Note that predicting the hidden extreme events is typically a great challenge given only partial observations. The third strategy makes use of the stochastic mode reduction technique[53–56], which allows a significant dimension reduction in the approximate models for many multiscale turbulent dynamical systems while the reduced order models retain the crucial nonlinear and non-Gaussian features. Applying this strategy, the nonlinear effects of the unresolved fast modes in the motion of the resolved variables are represented by effective damping and stochastic forcing. The resulting approximate models naturally belong to the new nonlinear modeling framework that allows extremely efficient data assimilation and prediction schemes. These approximate models also preserve physics-constrained properties. Another extremely useful strategy is to incorporate simple stochastic processes with additive noise and memory into the approximate models, which aim at effectively describing certain complicated nonlinear components that are hard to deal with in strongly nonlinear and chaotic dynamical systems. Due to the unique feature of the new nonlinear modeling framework, it allows an efficient and accurate way of using simple stochastic processes to learn these complex nonlinear components on the fly, which greatly facilitates the short- and medium-range forecasts of both the observed and hidden extreme events. Other approaches of building approximate predictive models that can be incorporated into the new nonlinear framework developed here involve using the noise inflation technique to effectively characterize the contributions from small-scale variables and fast-wave averaging of the variables with rapid decaying[57].

The rest of the paper is organized as follows. Section II describes the new nonlinear mathematical framework for developing suitable approximate models. Section III contains the efficient and accurate data assimilation, parameter estimation and prediction algorithms. Both the path-wise and information measurements in quantifying the prediction skill are also included in this section. Section IV illustrates the skill of predicting intermittent extreme events using cheap stochastic parameterizations with significant model error. Section V makes use of a nonlinear energy-conserving dyad model to show the success of applying the simple feedback control strategy in facilitating the prediction of the hidden extreme events. The effect of noise inflation in approximate models for predicting extreme events is illustrated based on the chaotic Lorenz 63 model in Section VI. Section VII starts with a 21-dimensional nonlinear topographic mean flow interaction model with regime switching. Stochastic mode reduction strategy is applied in a suitable way to develop an approximate nonlinear model with only 5 dimensions, which is nevertheless able to predict the observed and hidden extreme events as well as the regime switching between zonally blocked and unblocked flow patterns with high accuracy. In Section VIII, it is shown that incorporating judicious linear stochastic processes into a simple nonlinear approximate model succeeds in learning certain complicated nonlinear effects of a 6-dimensional low-order chaotic Charney-DeVore model with strong chaotic and regime switching behavior. The resulting nonlinear approximate model allows accurate online state estimation and the short- and medium-range forecasting of extreme events. The paper is concluded in Section IX.

II. A NONLINEAR MATHEMATICAL MODELING FRAMEWORK WITH SOLVABLE CONDITIONAL STATISTICS

A nonlinear mathematical modeling framework is established in this section, which will be used to the development of suitable approximate models for predicting extreme events. The general form of the nonlinear models within this framework is the following[38],

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

$$d\mathbf{u}_{\mathbf{II}} = [\mathbf{a}_0(t, \mathbf{u}_{\mathbf{I}}) + \mathbf{a}_1(t, \mathbf{u}_{\mathbf{I}})\mathbf{u}_{\mathbf{II}}]dt + \mathbf{\Sigma}_{\mathbf{II}}(t, \mathbf{u}_{\mathbf{I}})d\mathbf{W}_{\mathbf{II}}(t), \tag{1b}$$

where the state variables are written in the form $\mathbf{u} = (\mathbf{u}_{\mathbf{I}}, \mathbf{u}_{\mathbf{II}})$ with both $\mathbf{u}_{\mathbf{I}} \in R^{N_{\mathbf{I}}}$ and $\mathbf{u}_{\mathbf{II}} \in R^{N_{\mathbf{II}}}$ being multidimensional variables. In (1), $\mathbf{A}_0, \mathbf{A}_1, \mathbf{a}_0, \mathbf{a}_1, \mathbf{\Sigma}_{\mathbf{I}}$ and $\mathbf{\Sigma}_{\mathbf{II}}$ are vectors and matrices that depend only on time t and the state variables $\mathbf{u}_{\mathbf{I}}$, and $\mathbf{W}_{\mathbf{I}}(t)$ and $\mathbf{W}_{\mathbf{II}}(t)$ are independent Wiener processes. The systems in (1) are named as conditional Gaussian systems due to the fact that once $\mathbf{u}_{\mathbf{I}}(s)$ for $s \leq t$ is given, $\mathbf{u}_{\mathbf{II}}(t)$ conditioned on $\mathbf{u}_{\mathbf{I}}(s)$ becomes a Gaussian process with mean $\bar{\mathbf{u}}_{\mathbf{II}}(t)$ and covariance $\mathbf{R}_{\mathbf{II}}(t)$, i.e.,

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

Despite the conditional Gaussianity, the coupled system (1) remains highly nonlinear and is able to capture the non-Gaussian features as in nature. This conditional Gaussian nonlinear modeling framework includes many physics-constrained nonlinear stochastic models[35,36], large-scale dynamical models in turbulence, fluids and geophysical flows, as well as stochastically coupled reaction-diffusion models in neuroscience and ecology. See a recent work[34] for a gallery of examples of the conditional Gaussian systems. Applications of the conditional Gaussian systems to strongly nonlinear systems include developing low-order nonlinear stochastic models for predicting the intermittent time series of the Madden-Julian oscillation (MJO) and the monsoon intraseasonal variabilities[58–60], filtering the stochastic skeleton model for the MJO[61], and recovering the turbulent ocean flows with noisy observations from Lagrangian tracers[62–64]. Other studies that also fit into the conditional Gaussian framework includes the cheap exactly solvable forecast models in dynamic stochastic superresolution of sparsely observed turbulent systems[50,51], stochastic superparameterization for geophysical turbulence[52] and blended particle filters for large-dimensional chaotic systems[65].

One important feature of the above conditional Gaussian nonlinear framework is that the conditional Gaussian distribution $p\big(\mathbf{u_{II}}(t)|\mathbf{u_I}(s \le t)\big)$ in (2) has closed analytic form[37],

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

It is natural to assume $\mathbf{u_I}$ contains the observed variables while $\mathbf{u_{II}}$ is a collection of the unobserved ones. Therefore, the analytically solvable conditional statistics in (3) allows an extremely efficient and accurate way of estimating the hidden states given the observations, known as the data assimilation, which facilitates predictions. Note that in the data assimilation language the conditional mean and conditional covariance in (3) are also known as the posterior mean and posterior covariance. In addition, the conditional Gaussian nonlinear modeling framework (1) and its closed analytical form of the conditional statistics (3) offer a statistical efficient and accurate way of solving the time evolution of the associated Fokker-Planck equation in high dimensions[66–68], which also provides a powerful tool for carrying out ensemble forecasts.

III. DATA ASSIMILATION, PREDICTION, AND THE QUANTIFICATION OF PREDICTION SKILL

A. Data assimilation of the unobserved variables

Data assimilation (also known as state estimation or filtering)[21,25–28], a procedure of estimating the states of the unobserved variables, is the precondition of predicting complex dynamical systems. In fact, data assimilation of the unobserved variables can also be regarded as the online "prediction" of these variables due to the fact that the recovered states of the unobserved variables are given by combining the information in the dynamics with the values of the observed variables.

Data assimilation of the unobserved variables plays an important role in short- and medium-range forecasts. This is because the ensemble prediction algorithm requires running the model forward with the given initial values for all the state variables. Since there is no direct observations of the hidden or unresolved variables, assimilating their initial states becomes a necessity part of the ensemble forecast. In practice, the data assimilation is often required in an "online" form in the sense that the states of the unobserved variables need to be estimated at each time instant as time evolves. Therefore, developing an efficient and accurate data assimilation method is a crucial first step for predicting nonlinear complex dynamical systems and the associated extreme events. However, the classical Kalman filter or its continuous form Kalman-Bucy filter[69–71] works only for linear models. On the other hand, for assimilating general complex nonlinear dynamical systems, the particle filter is quite expensive and contains sampling error while the ensemble Kalman filter takes into account only the first two moments which may end up with large biases for assimilating extreme events.

The conditional Gaussian nonlinear modeling framework in Section II provides an efficient way of estimating the states of the highly non-Gaussian hidden variables $\mathbf{u_{II}}(t)$ in the complex nonlinear dynamical systems given the observations up to the current time $\mathbf{u_I}(s \leq t)$. The closed analytic formula in (3) avoids numerical and sampling errors, and it results in an extremely efficient and accurate way of computing the optimal states of $\mathbf{u_{II}}(t)$.

B. Short-, medium- and long-range forecasting

Prediction problems have been described by Lorenz as falling into two categories[72,73]. Problems that depend on the initial condition, such as short- to medium-range weather forecasting, are described as "predictions of the first kind", while problems for predicting the longer-term climatology, are referred to as "predictions of the second kind".

For short- and medium-range forecasts, the system starts from an initial time t_0, where the initial values of the unobserved variables are determined by data assimilation. Then an ensemble prediction algorithm is applied by running the model forward up to a given time t_1. Typically, t_1 is not quite far from t_0 and therefore the system has not completely lost its memory of the initial values. Therefore, a good state estimation of the initial values via data assimilation plays an important role in providing an accurate short- and medium-range forecasting skill. The ensemble mean, which is the average value of all the ensemble members, is often used as a predictor for the evolution of the trajectories and the ensemble spread measures the uncertainty in the ensemble mean forecasts. The difference between short- and medium-range forecasts is that the prediction skill at very short lead time largely depends on the accuracy of the initial values while both the dynamical structures and the initial values will be essential in predicting the model transition behavior in the medium range. Capturing the time evolution of the large bursts in intermittent time series with small uncertainty is the goal of short- and medium-range forecasts of the extreme events.

On the other hand, for the long-range forecast where t_1 is much larger than t_0, the system will lose its memory from the initial time and arrives at the statistical equilibrium state. In such a scenario, the ensemble mean prediction provides no information beyond the mean of the statistical equilibrium state. Therefore, the aim of the long-range forecast is to predict the statistical behavior. In particular, reproducing the statistical equilibrium non-Gaussian probability density function (PDF) with fat tails is a good evidence of successfully predicting the extreme events, where the extreme events and intermittency are the main contributors to the fat tails.

Note that different models may have the same characteristics for the long-term statistics but they often have significantly different skill for short and medium range prediction as well as the forced response. In a recent paper[74], several instructive examples using both a simple linear 2×2 system and more complicated nonlinear models unambiguously illustrate such a feature of predicting complex turbulent dynamical systems. It is also shown in the paper[74] that in the presence of model error, developing suitable approximate models that are skillful in one of the short-, medium- or long-range forecasting is already a quite difficult task. In many cases, there exists an information barrier[75] that prevents the approximate models predicting the exact statistics and capturing the perfect response.

It is worthwhile to mention that a grand challenge in contemporary climate, atmosphere, and ocean science is to understand and predict intraseasonal variability for time scales from 30 to 60 days, which is longer than standard weather time scales of at most a week and much shorter than the yearly time scales of short-term climate. Therefore, it belongs to the medium-range forecasts. Von Neumann[76] called such problems at the intersection of weather and climate the greatest challenge in future meteorology[10,77]. The Indian-Asian monsoon and the MJO[10,78–80] are the most significant intraseasonal variability occurs in the tropical areas. Notably, it is shown in the recent work[58–60] that the nonlinear modeling framework in Section II facilitates the development of effective low-order nonlinear stochastic models for predicting the intermittent time series of the MJO and the monsoon as well as extending the predictability of these intraseasonal variabilities.

C. Prediction of the dynamical evolution towards to the attractor

The short-, medium-, and long-range forecasts discussed above typically assume the initial values lie in the statistical equilibrium states. On the other hand, certain internal or external perturbations are able to kick the model variables outside the attractor. Therefore, predicting the time evolution of the extreme events that start from a state outside the attractor and its returning path to the statistical equilibrium state is another important issue. Since most approximate models are calibrated using the training data from the attractor, there is no guarantee that these approximate models are automatically able to predict the relaxation towards the attractor. This results in a great challenge of predicting the extreme events starting outside the attractor. Some instructive studies of the prediction and linear response skill with the initial condition being off the attractor can be found in a recent paper[74].

D. Calibration of the model through parameter estimation

One important issue before applying approximate models for predicting extreme events is the model calibration through parameter estimation. The method adopted here follows the algorithm in a recent work[81]. The main difficulty in estimating the parameters in general nonlinear systems with only partial observations is that the closed form of the likelihood function is typically unavailable. Therefore, data augmentation of trajectories associated with the hidden variables is often applied[82–85], which then allows using the Markov Chain Monte Carlo (MCMC) methods to sample the parameters and the hidden trajectories in an alternative way for parameter estimation. Yet, since the hidden trajectories lie in an infinitely dimensional space (or finite but large dimensional with the discrete approximation), the data augmentation can be quite slow in many applications.

Here, in light of the closed analytic formulae (3) of the conditional Gaussian nonlinear approximate models (1), data assimilation can be incorporated into a classical MCMC algorithm to circumvent the most expensive part of the parameter estimation algorithm, namely sampling the unobserved trajectories using data augmentation. Specifically, in each iteration step k of the MCMC, we make use of the observed trajectories $\mathbf{u_I}$ and the current updated parameters $\boldsymbol{\theta}^{(k)}$ to recover the unobserved trajectories of $\mathbf{u_{II}}$, namely $\mathbf{u_{II}}^{mis,(k)}$ via data assimilation (3) in a deterministic and optimal way. Then $\mathbf{u_{II}}^{mis,(k)}$, $\mathbf{u_I}$ and $\boldsymbol{\theta}^{(k)}$ are used together to compute the likelihood function

$$p(\mathbf{u_I}|\boldsymbol{\theta}^{(k)}) = p(\mathbf{u_I}|\boldsymbol{\theta}^{(k)}; \mathbf{u_{II}}^{mis,(k)}),$$

which will be used in the MCMC algorithm for updating the parameters in the $k+1$ iteration step. For a complete description of the algorithm, see[81] for details. The slight difference of the algorithm applied here compared to the original version in[81] is that an adaptive MCMC procedure[86] for choosing the proposal function is applied.

E. Quantifying the prediction skill

1. *Path-wise measurements*

The root-mean-square error (RMSE) and the pattern correlation (Corr).

The root-mean-square error (RMSE) and the pattern correlation (Corr) are the two path-wise measurements that have been widely applied to quantify the prediction skill[21,87–91]. Denote u_i the true signal and $\hat{u}_i$ the prediction estimate, where $i = 1, \ldots, n$ is an index in time. These measurements are given by

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (\hat{u}_i - u_i)^2}{n}}, \tag{4}$$

$$\text{Corr} = \frac{\sum_{i=1}^n (\hat{u}_i - \overline{\hat{u}}_i)(u_i - \overline{u}_i)}{\sqrt{\sum_{i=1}^n (\hat{u}_i - \overline{\hat{u}}_i)^2}\sqrt{\sum_{i=1}^n (u_i - \overline{u}_i)^2}}, \tag{5}$$

where $\overline{\hat{u}}_i$ and $\overline{u}_i$ denote the mean of $\hat{u}_i$ and u_i respectively.

In practice, the trajectory of the ensemble mean is often used as $\hat{u}_i$ in measuring the RMSE and Corr. These two path-wise measurements are intuitive and easy to be applied. Typically, a prediction is said to be skillful if the RMSE is below one standard deviation of the true signal and the Corr is above the threshold Corr $= 0.5$.

Yet, we have to point out several potential issues in these measurements. First, since only the ensemble mean is used as the predictor, the predicted uncertainty which involves the ensemble spread (or the confidence interval of the ensemble mean prediction) is not involved in these path-wise measurements. Second, these path-wise measurements fail to quantify the skill of the long-term forecast, since the ensemble mean simply becomes the equilibrium mean state of the system. In addition, both the RMSE and Corr take into account only the information up to the second order statistics. Thus, they may lead to biased conclusions for predicting extreme events and non-Gaussian features. Nevertheless, due to the simple form, these path-wise measurements can still be applied to provide some useful information for the short- and medium-range forecasts.

The temporal autocorrelation function (ACF).

Autocorrelation is the correlation of a signal with a delayed copy of itself, as a function of delay. For a zero mean and stationary random process u, the autocorrelation function (ACF) can be calculated as

$$\mathrm{ACF}(t) = \lim_{T\to\infty}\frac{1}{T}\int_0^T \frac{u(t+\tau)u^*(\tau)}{\mathrm{Var}(u)}d\tau, \tag{6}$$

where $\cdot^*$ denotes the complex conjugate. The ACF has been widely used to measure the system memory. It also plays an important role in improving the linear response via the fluctuation-dissipation theorem[31,92]. If the perfect model and the approximate model share the similar ACFs, then the two systems usually have a similar dynamical behavior at least up to the second order statistics. However, for nonlinear and chaotic systems, high order statistics may play an important roles for extreme events. Therefore, the ACF can only be regarded as a crude indicator of the overall predictability of the underlying system. As a remark, the information theory is able to provide a rigorous and practical way to quantify the error in the two ACFs associated with the perfect and approximate models by making use of their spectral representations. See[30,93] for details.

2. Information measurements

Information theory provides a natural way to quantify the prediction skill and model error by measuring the lack of information. Different from the path-wise measurements, the information measurements assess the statistical behavior of the systems. The lack of information in one probability density q compared with another p is through the relative entropy $\mathcal{P}(p,q)$[2,94–98],

$$\mathcal{P}(p,q) = \int p\log\left(\frac{p}{q}\right), \tag{7}$$

which is also known as the Kullback-Leibler divergence or information divergence[99–101]. Despite the lack of symmetry, the relative entropy has two attractive features. First, $\mathcal{P}(p, q) \geq 0$ with equality if and only if $p = q$. Second, $\mathcal{P}(p, q)$ is invariant under general nonlinear changes of variables.

Long-term prediction.

The long range forecast using the approximate model aims at capturing the non-Gaussian statistical equilibrium states of the truth, especially the fat tails that correspond to the extreme events. This is very different from the short- or medium-range forecasts, where the path-wise measurements of the ensemble mean are informative. In fact, the path-wise measurements completely fail to quantify the long range forecasting skill. Information theory, on the other hand, provides a natural quantification of the statistical prediction skill in the approximate model, which is given by

$$\mathcal{E}_{eq} = \mathcal{P}(p_{eq}, p^M_{eq}), \tag{8}$$

where p_{eq} and p^M_{eq} are the equilibrium PDFs of the perfect model and the approximate model, respectively. The information measurement in (8) is able to quantify the skill of the approximate model in capturing both the majority of the events represented by the mode of the PDF and the intermittent extreme events in the PDF tails. Note that minimizing the information score in (8) is also known as capturing the model fidelity[94] using approximate models.

The short- and medium-range forecasts.

The information theory can also be applied to quantify the short- and medium-range forecasting skill. The fundamental difference between the information measurements and the path-wise ones is that the information measurements are able to take into account the predicted uncertainty. Denote p_t and p^M_t the PDFs of the time-dependent perfect and approximate models starting from the same initial time. Similar to (8), an information metric for quantifying the predicted model error as a function of time can be defined as

$$\mathcal{E}_t = \mathcal{P}(p_t, p^M_t). \tag{9}$$

A suitable approximate model is expected to have a small model error throughout the time.

The information measurements can also be used to assess the predictability, also known as the internal prediction skill, of both the perfect and approximate models using the following matric[74,102,103],

$$\mathcal{D}_t = \mathcal{P}(p_t, p_{eq}), \qquad \text{and} \qquad \mathcal{D}_t^M = \mathcal{P}(p_t^M, p_{eq}^M). \tag{10}$$

Clearly, the measurement in (10) quantifies the information provided by the initial conditions about the future state of the system beyond the prior knowledge available through equilibrium statistics. Obviously, both $\mathcal{D}_t$ and $\mathcal{D}_t^M$ will decay to zero eventually. Therefore, the measurement in (10) can be regarded as an analog to the ACF but it takes into account the entire predicted PDF rather than simply the path associated with the ensemble mean prediction.

IV. A SIMPLE MODEL WITH HIGHLY NON-GAUSSIAN BEHAVIOR IN THE HIDDEN PROCESS

Stochastic parameterizations are widely used in developing approximate models for complex dynamical systems with partial observations[1,104–106]. The idea of applying stochastic parameterizations is to use simple stochastic processes to describe the complicated dynamics of the unobserved or unresolved scales such that the overall computational cost of the approximate models is greatly reduced. One important practical issue is to develop suitable stochastic parameterizations for the hidden processes such that the intermittent features are captured and the approximate models with the stochastic parameterizations are able to accurately predict the extreme events in the observed variables.

The goal of this section is to test the skill of a simple and efficient stochastic parameterization strategy in predicting intermittent non-Gaussian features and extreme events based on a low-order highly non-Gaussian test model given only a short period of training data with partial observations.

A. The perfect and approximate models

The perfect model.

The perfect test model here is given by a two-dimensional system where only a short trajectory of one variable u is observed. The model reads,

$$du = \Big(- \gamma u + F_u \Big) dt + \sigma_u dW_u, \tag{11a}$$

$$d\gamma = (a_\gamma \gamma + b_\gamma \gamma^2 + c_\gamma \gamma^3 + f_\gamma) dt + (A_\gamma + B_\gamma \gamma) dW_{\gamma,1} + \sigma_\gamma dW_{\gamma,2}. \tag{11b}$$

In this model, the variable γ acts as a stochastic damping in the equation of u and the averaged value of γ over time needs to be positive to guarantee the mean stability of u[107]. Once the sign of γ switches from positive values to negative values, γ becomes anti-damping and it leads to the intermittent events in u. On the other hand, γ is driven by a cubic nonlinear equation with correlated additive and multiplicative noise. This cubic model is a canonical model for low frequency atmospheric variability[108,109]. This one-dimensional, normal form has been applied in a regression strategy for data from a prototype atmosphere and ocean model to build one-dimensional stochastic models for low-frequency patterns such as the North Atlantic Oscillation and the leading principal component that has features of the Arctic Oscillation. Given the non-Gaussian features and the potential physical explanations, the low-order model (11) becomes a useful testbed for developing suitable stochastic parameterization strategies of the hidden process that allows skillful prediction of the extreme events in the observed variable.

The following parameters are taken in the perfect model,

$$\begin{gathered} F_u = 0.3, \qquad \sigma_u = 0.1, \qquad a_\gamma = -\frac{3}{8}, \qquad b_\gamma = 1, \qquad c_\gamma = -\frac{1}{2}, \\ A_\gamma = 0, \qquad B_\gamma = \frac{1}{2\sqrt{2}}, \qquad f_\gamma = 0.1, \qquad \sigma_\gamma = \frac{1}{2\sqrt{2}}. \end{gathered} \tag{12}$$

With these parameters, the model trajectories together with the equilibrium PDFs and ACFs are shown in Panels (a)–(c) of Figure 1. Note that the time series in Panel (a) only contains a length of 500 time units but the PDFs and ACFs in Panels (b)–(c) are computed based on the model simulation with a length of $10,000$ units in order to minimize the sampling bias in showing these statistics.

In this dynamical regime, the time series of γ shows a stochastic switching behavior. Roughly speaking, γ has two statistical states. The averaged value in one state is slightly negative, corresponding to the intermittent phase of u, while another state of γ is positive, corresponding to the quiescent phase of u. The PDF of u, due to the intermittent extreme events, is highly skewed with an one-sided fat tail. On the other hand, the PDF of γ shows a bimodal behavior, which is also significantly non-Gaussian. The ACFs indicate that overall u has a longer memory than γ.

The approximate model.

The perfect model (11) here can be regarded as a paradigm model in many real applications, where the hidden variables are driven by some unknown complicated processes that interact with the observed variables in a highly nonlinear way. From a practical point of view, it is important to develop a simple and computationally tractable approximate model which is nevertheless able to capture the key nonlinear feedback from the unobserved variable γ to the observed variable u. The approximate model is expected to predict the extreme events of the observed process u.

One commonly used reduced order modeling strategy is to adopt a mean stochastic model (MSM) for the observed process u. The MSM makes use of the averaged value of γ as the damping term and the resulting system is

$$du = (-\hat{\gamma}u + F_u)dt + \sigma_u dW_u. \tag{13}$$

Since the mean stability is guaranteed in the original system, the constant $\hat{\gamma}$ is positive. Thus, the MSM is a linear model with Gaussian statistics. It has been shown in[22,107] that the MSM is unable to capture the short-term rapid increment of the intermittent trajectory of u due to the lack of intermittent instability mechanism. Such a Gaussian model also fails to predict the long-term non-Gaussian PDF with skewness and fat tails.

Here, a new approximate model is developed using the stochastic parameterized equation technique[48,49], the idea of which has been applied to the extended Kalman filters (known as the SPEKF-type model) and other prediction and data assimilation forecast models. The approximate model has the following form,

$$du = (-\gamma u + F_u)dt + \sigma_u dW_u, \tag{14a}$$

$$d\gamma = -d_\gamma(\gamma - \hat{\gamma})dt + \sigma_\gamma dW_\gamma. \tag{14b}$$

In (14), the nonlinear process γ with correlated additive and multiplicative noise in (11b) has been simplified to a linear process with only Gaussian additive noise. Nevertheless, the variable γ remains switching between positive and negative phases, representing damping and anti-damping effects as a feedback to u. Therefore, the variable γ is still able to trigger intermittent extreme events in u. One important feature is that the approximate model (14) belongs to the conditional Gaussian nonlinear framework as was described in Section II, which allows the effective algorithm (3) to solve the conditional statistics of the hidden variable γ given the observations from u. This greatly facilitates the data assimilation and predictions.

B. Parameter estimation

Before applying the approximate model for prediction, the parameters $(F_u, \sigma_u, d_\gamma, \hat{\gamma}, \sigma_\gamma)$ in the approximate model (14) need to be estimated. The training time series only involves the observed variable u and the training data has only a short period with 500 units as shown in Panel (a) of Figure 1. Applying the parameter estimation algorithm described in Section III D, the results are shown in Figure 2. The trace plots associated with all the parameters clearly indicate the convergence towards certain values with small uncertainties. Notably, the estimation values of the two parameters σ_u and F_u in the observed process are almost the same as the ones in the perfect model. The averaged values of the trace plots from iteration $k = 5000$ to iteration $k = 10000$ are utilized as the estimated parameters in the approximate model for prediction:

$$d_\gamma = 0.2545, \qquad \hat{\gamma} = 1.121, \qquad F_u = 0.2489, \qquad \sigma_\gamma = 0.4362, \qquad \sigma_u = 0.1008. \tag{15}$$

C. Long-term prediction

With the estimated parameters, we begin with studying the long-term prediction using the approximate model (14). As an analogy to Panels (a)–(c) in Figure 1 for the perfect model, Panels (d)–(f) in Figure 1 show the trajectories, PDFs and ACFs of the approximate model. Note that Panel (d) is simply a free run of the model. Therefore, there is no point-to-point correspondence between the trajectories shown in Panels (a) and (d) for the perfect and approximate models. Nevertheless, it is easy to see that the trajectories from the perfect and approximate models are qualitatively similar to each other, indicating the skill of the approximate model in capturing the long-term dynamical and statistical behavior. Next, to understand the quantitative similarity between the two models, the equilibrium PDFs and the ACFs are compared.

Panels (c) and (f) show that both the ACFs of u and γ associated with the two models are very similar to each other, indicating the success of the approximate model in capturing the temporal information of the perfect system. On the other hand, as shown in Panels (b) and (e), the PDF of u is also perfectly recovered by the approximate model, where using the information distance (7) the difference between the PDFs associated with the approximate and perfect models $\mathcal{P}(p_{eq}(u), p^M_{eq}(u)) = 0.0345$ is a negligible value. The PDF of γ is not perfectly recovered because the approximate model uses only a linear system with additive noise for γ, which fails to capture the non-Gaussian PDF of γ. This is known as the information barrier[22]. Nevertheless, the PDF of γ associated the approximate model is nearly exact the same as the Gaussian fit of the bimodal distribution associated with the truth. This in fact implies that the approximate model has reached its predictability limit in predicting the long-term statistics of the hidden variable γ.

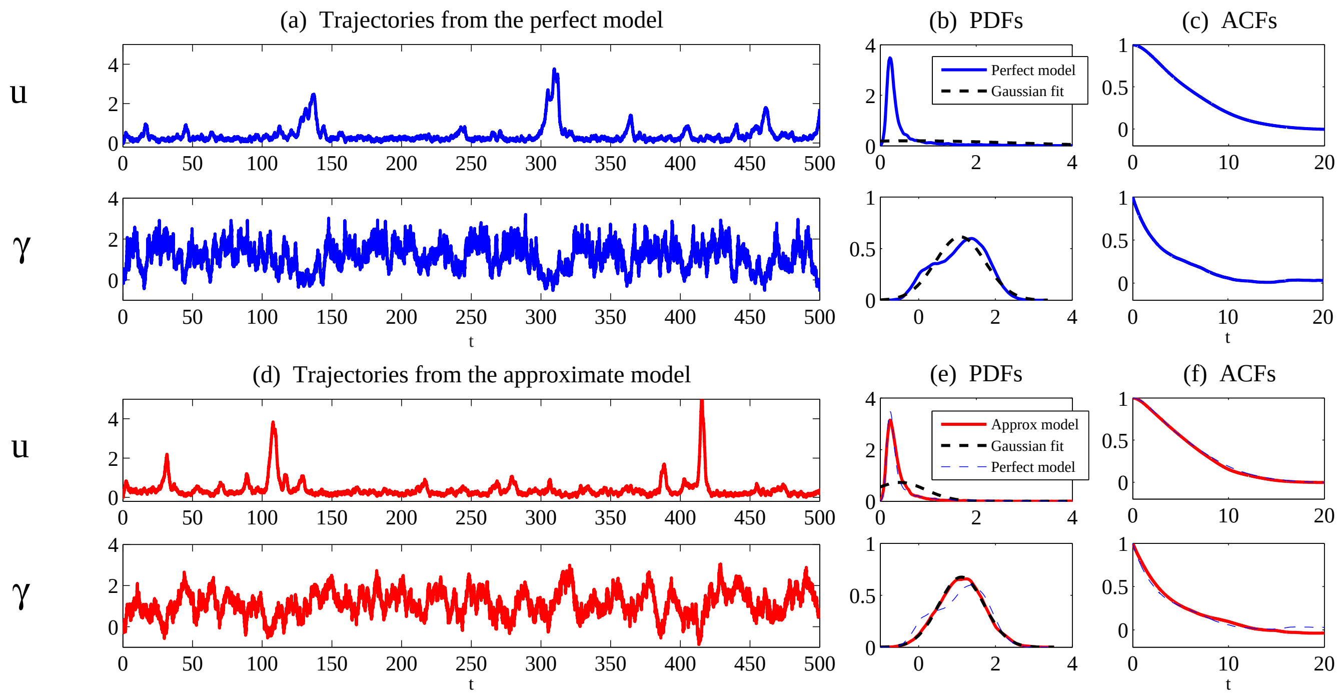

FIG. 1. Model trajectories, PDFs and ACFs of the perfect model (11) with parameters in (12). Top: u ; Bottom: γ. Note that the PDFs and ACFs are computed based on the model simulation with a length of $10,000$ units. But in panels (a) and (c) only time series with a length of 500 time units are shown. The black curves in the PDFs show the Gaussian fits.

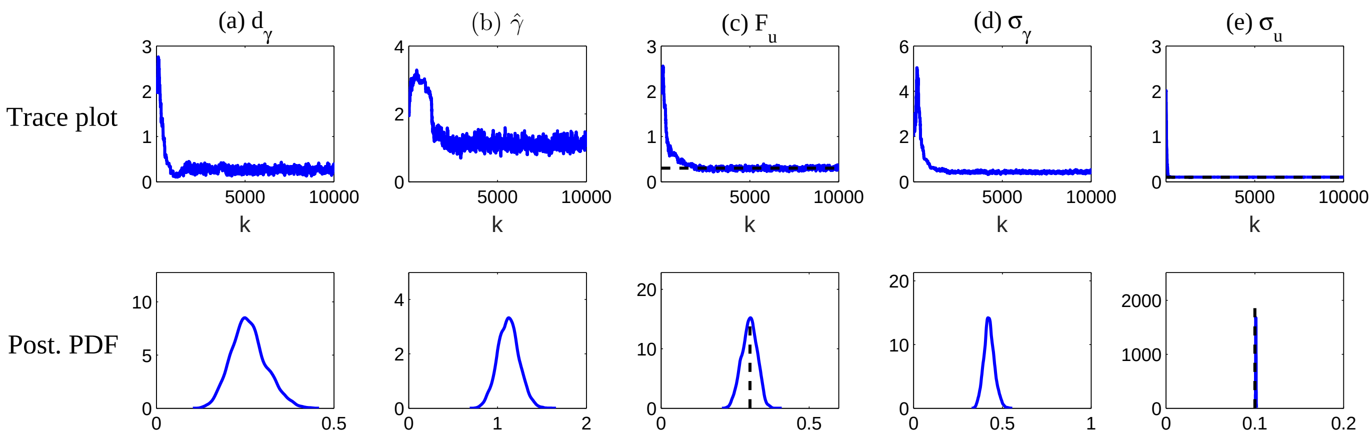

FIG. 2. Parameter estimation of the approximate model (14). Top: trace plot. Bottom: posterior PDFs of the parameters from the trace plot taking the values from $k = 5000$ to $k = 10000$. The black lines show the values of σ_u and F_u in the perfect model (11), serving as reference values.

D. Data assimilation

One key feature of the approximate model (14) is that it belongs to the conditional Gaussian model family (1), which allows using the closed analytic formulae (3) to solve the the conditional distribution $p(\gamma|u)$ for assimilating the unobserved variable γ. Note that the perfect model (11) is not a conditional Gaussian system and expensive particle methods have to be used in order to assimilate the unobserved variable γ even in this two-dimensional system. Therefore, the approximate model (14) is much more computationally efficient for state estimation, data assimilation and prediction.

Figure 3 shows the data assimilation results using the approximate model (14) as the forecast model. It is clear that the γ values associated with the intermittent phase of u are recovered with both high accuracy and low uncertainty. The accurate recovery of the hidden variable γ at the intermittent phase of u indicates its potential for predicting the extreme events. On the other hand, assimilating the γ states corresponding to the quiescent phase of u are recovered with high uncertainty. The posterior mean also fails to track the fluctuations in the true trajectory. This is not surprising since the quiescent phases of u have weak amplitudes and therefore the noise-to-signal ratio is large. In fact, as long as the hidden variable γ stays positive, playing the role as a damping, it has very weak influence on the dynamics u at the quiescent phases. The assimilated values and uncertainties of γ accurately reflect these features.

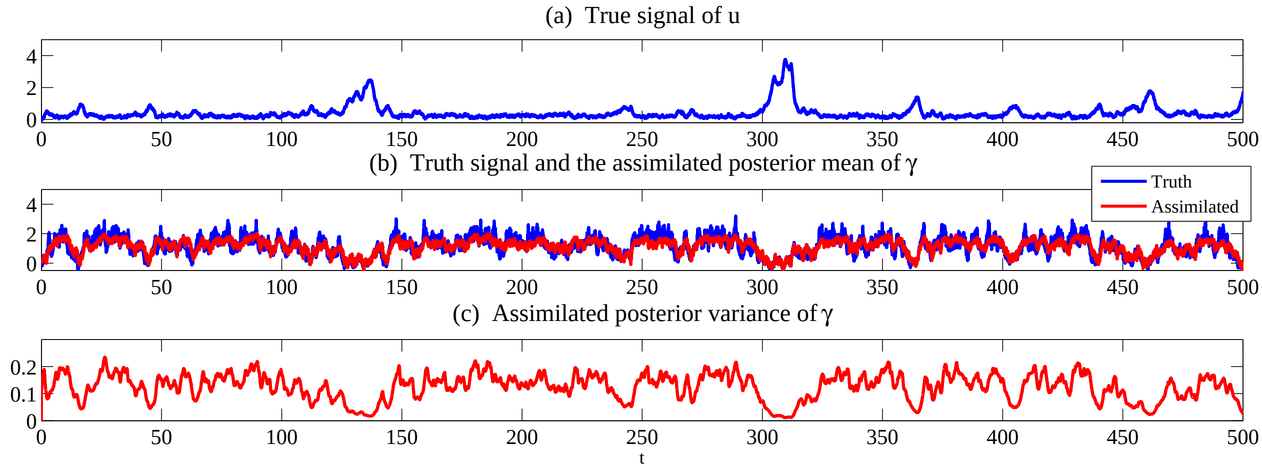

FIG. 3. Data assimilation of the hidden variable γ using the approximate model (14) as the forecast model. The true signal of the observed variable (panel (a)) is generated from (11). Panel (b) shows the true signal of γ from (11) and the assimilated (filtered) posterior mean of γ using the approximate forecast model (14). Panel (c) shows the posterior variance. The black dashed boxes indicates the events that will be studied for short-term prediction in the next subsections.

E. Short- and medium- range forecasts

To study the short- and medium-range forecast, we first show the RMSE and the Corr between the predicted time series and the truth as a function of lead time. Here the ensemble mean is used as the predicted time series. As illustrated in Panels (a)–(b) of Figure 4, the approximate model with the assimilated initial conditions has an overall comparable prediction skill as the perfect model prediction with the perfect initial conditions. The only main difference lies in the very short term for predicting γ, where the prediction using the approximate model with the assimilated initial conditions has a larger error. This is due to the large uncertainty in the assimilated initial conditions at the quiescent phases. In fact, if we adopt the approximate model as the forecast model but use the perfect initial conditions (green curves), then the prediction skill is almost the same as using the perfect model prediction. Note that the overall skillful prediction of u lasts up to 5 units while that of γ is around 2 units.

Panels (c)–(d) and (e)–(f) of Figure 4 show the lead time prediction at $0.2, 0.5$ and 1.5 units using the approximate model with the assimilated initial conditions and the perfect model with the perfect initial conditions, respectively. The prediction of u, especially the extreme events, is quite accurate at all the three lead times for both the models. The prediction of the negative phase of γ is also nearly perfect. The only difference between the two models lies in predicting the positive phases of γ, where the approximate model cannot provide an accurate prediction even at a very short lead time. This is due to the error and the uncertainty in the assimilated initial conditions as was discussed above. On the other hand, while the perfect model is able to predict the positive phase of γ (corresponding to the quiescent phases of u) in a very short term, it is interesting to see that even with the perfect model and perfect initial conditions, some significant errors already appear in predicting the positive phases of γ at a lead time 0.5. At a lead time 1.5, the perfect model essentially gives the same results as the approximate model, where an accurate prediction is found in both u and the negative phase of γ while the model is not very skillful in predicting the positive phase of γ. These facts indicate that when γ is positive it only has a weak influence on u and therefore the system has an intrinsic weak dependence of γ.

To conclude, the approximate model has almost the same short- and medium-range forecasting skill as the perfect model, especially in predicting the extreme events in u and the corresponding triggering phases in γ.

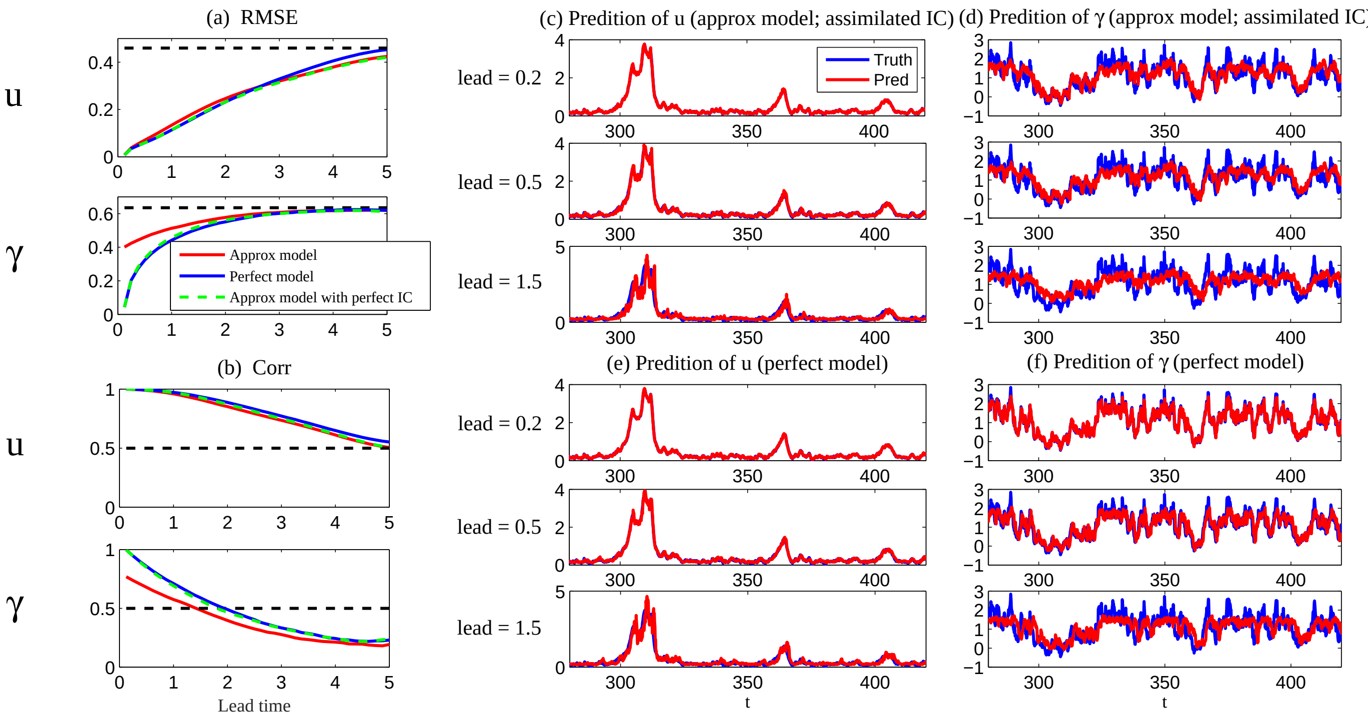

FIG. 4. Short- and medium-range forecasts. Panels (a)–(b): RMS error and pattern correlation between the predicted time series and the truth as a function of lead time. Red: prediction using the approximate model (14), where the initial values of γ are obtained by data assimilation. Dashed blue: prediction using the perfect model (11) with perfect initial conditions. Dashed green: prediction using the approximate model (14) but with perfect initial conditions. Panels (c)–(d): Ensemble mean prediction using the approximate model with assimilated initial condition (IC) at lead times $0.2, 0.5$ and 1.5. The blue curves show the truth while the red ones show the prediction. Panels (e)–(f): Similar to (c)–(d) but using the perfect model and perfect initial condition.

F. Prediction with an initial value starting outside the attractor

Finally, we study the prediction skill of the approximate model if the initial value is outside the attractor (the statistical equilibrium state). In Figure 5, we consider the situations where either the initial value of u or that of γ is outside the attractor. It is clear that when $u(0)$ is outside the attractor while γ stays in the attractor (Panels (a) and (c)), the trajectory of u releases to the attractor in a similar fashion using both the approximate model and the perfect model. This is because there is no approximation in the observed process u and the time evolution of γ at the attractor has already been shown to be accurately described using the approximate model. On the other hand, if γ starts from a value that is outside the attractor (Panels (b) and (d)), then the approximate model in capturing the relaxation of γ towards the attractor may contain errors. In fact, when γ starts from a large value as shown in Panel (b), the cubic damping plays an important role in strongly pushing the system towards the attractor. Starting from a large value of γ, the impact of the cubic damping is much stronger than that at the attractor and therefore the approximate model with a linear damping in γ fails to capture this feature. Nevertheless, if u stays on the attractor, then as long as γ is positive it has only a weak influence on the observed variable u. Therefore, the overall dynamics of u can still be described quite well using the approximate model.

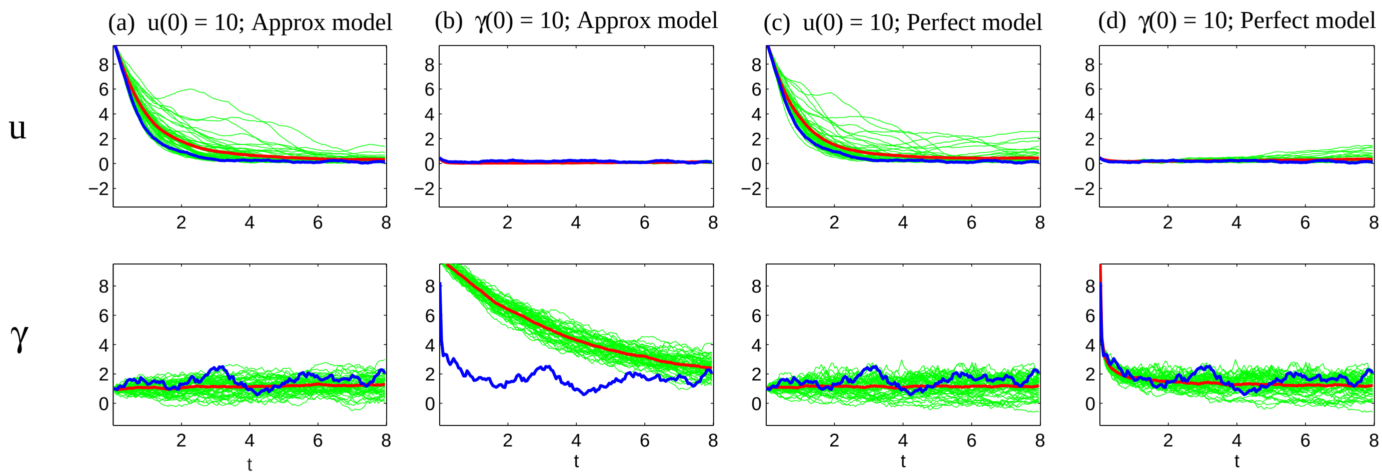

FIG. 5. Ensemble predictions with the initial values starting outside the attractor. Panels (a) and (c): $u(0) = 10$ starts from a value that is off the attractor. $\gamma(0) = 1$ is inside the attractor. Panels (b) and (d): $u(0) = 0.5$ starts from a value that is inside the attractor. $\gamma(0) = 10$ is off the attractor. Here Panels (a)–(b) show the results using the approximate model while Panels (c)–(d) show those using the perfect model. In all the panels, blue curves show the truth and red curves show the ensemble mean which is the average value of 50 ensembles showing in green color.

V. A DYAD MODEL WITH ENERGY-CONSERVING NONLINEAR INTERACTION

The nonlinear test model in the previous section involves only an one-way influence from γ to u. Yet, in many applications, the observed variables and the unobserved ones have mutual interactions, which are also often though energy-conserving nonlinear terms[35,36]. Therefore, it is important to understand different strategies in building approximate models to predict the extreme events and other non-Gaussian behavior in such kind of the systems. In this section, a simple but judicious feedback control strategy is adopted to facilitate the prediction of the hidden extreme events in an energy-conserving nonlinear dyad model.

A. The models

The perfect model.

Consider a nonlinear dyad model with energy-conserving nonlinear interaction,

$$\begin{aligned} dv &= \Big(- d_v v - c u^2 \Big) dt + \sigma_v dW_v, \\ du &= \Big((-d_u + cv) u + F_u \Big) dt + \sigma_u dW_u. \end{aligned} \tag{16}$$

Again only partial observations are available in this nonlinear dyad model, where v is the observed variable while u is unobserved. This low-order nonlinear model can be regarded as a toy model of complex turbulent flows. For example, v can be treated as one of the Fourier modes associated with the large-scale observed variables while u is associated with the hidden mechanism that drives v. If u represents unresolved or small-scale variables, then its statistics can be highly non-Gaussian. Here, v plays the role of the stochastic damping in the process of u such that intermittent extreme events appear in the trajectory of u. Note that this model is very different from the SPEKF-type of the model described in the previous section. In fact, in the dyad model (16), the variable u also provides a nonlinear feedback to v via $-cu^2$ such that the total energy in the nonlinear terms of the coupled system is conserved, which is known as the physics constraint[35,36].

Below, the nonlinear dyad model (16) is used as the perfect model. The focus of this section is to predict the extreme events in the unobserved process u. To this end, the following parameters are taken in the nonlinear dyad model (16),

$$F_u = 1, \qquad d_u = 0.8, \qquad d_v = 0.8, \qquad \sigma_u = 0.2, \qquad \sigma_v = 2, \qquad c = 1.2. \qquad (17)$$

As shown in Panels (a)–(c) of Figure 6, the nearly Gaussian observed variable v switches between positive and negative states, which leads to the intermittency in the hidden process u. The non-zero forcing $F_u = 1$ makes the signal of u stay almost within the positive values and the PDF of u is skewed with an one-side fat tail. Note that the amplitude of this forcing term provides different dynamical behavior of the model. In the last part of this section, the prediction skill in different dynamical regimes with various values of F_u will be reported.

The approximate model.

Again, a suitable approximate model is able to predict the extreme events and other important non-Gaussian features of the perfect model. Meanwhile, the approximate model is expected to be computationally efficient for data assimilation and prediction. Due to the closed analytic formulae of the conditional Gaussian models in assimilating the unobserved variables, we now aim at developing a suitable approximate model that belongs to the conditional Gaussian framework. Note that by observing v, the perfect model (16) is not a conditional Gaussian nonlinear system. One starting idea for building an approximate model is to apply a bare truncation strategy, which ignores the quadratic feedback term $-u^2$ in the process of v in (16). This is actually a commonly used strategy in developing approximate models for many complicated systems in practice, where some nonlinear terms are dropped. However, this strategy does not work for studying the extreme events with partial observations. In fact, without this feedback term, the variable u is completely decoupled from the process of v. In other words, given only the observations in v, the processes and the parameters of u are not even identifiable. What is more, using the same parameters as in (17), such an approximate model suffers from a finite-time blowup of the signals[35,110].

The failure of the bare truncation model is due to the complete ignorance of the nonlinear feedback term from u to v. This nonlinear feedback not only provides the observability of u in the v process but also offers the important causal effects between the two processes. Therefore, a suitable approximate model is supposed to take into account such an interaction between the two processes. To this end, the following approximate model is adopted,

$$\begin{aligned} dv &= \Big(- d_v v - cu\Big)dt + \sigma_v dW_v, \\ du &= \Big((-d_u + cv)u + F_u\Big)dt + \sigma_u dW_u. \end{aligned} \tag{18}$$

This approximate model uses a linear feedback $-cu$ to approximate the nonlinear interaction $-cu^2$ in the original dyad model. This simplification can be regarded as using a linear control term to retain the mutual dependence of u and v. It also allows the approximate model to belong to the conditional Gaussian framework that facilitates efficient data assimilation and prediction algorithms.

B. Parameter estimation

For the parameter estimation of the approximate model (18), we make use of a short training data of v with only 500 time units as shown in Panel (a) of Figure 6. The parameter estimation algorithm is run for $K = 15000$ steps and the averaged values from the trace plots between $k = 5000$ to $k = 15000$ is used as the estimated parameters,

$$\begin{gathered} d_v = 0.9234, \qquad d_u = 0.6672, \qquad c = 1.8249, \qquad F_u = 0.6041, \\ \sigma_u = 0.0527, \qquad \sigma_v = 2.0203. \end{gathered} \tag{19}$$

It is useful to compare the estimated parameter values in the approximate model (19) with those in the perfect model (17). This helps understand the dynamical properties of the approximate model.

The feedback parameter c in the approximate model (17) is increased. This is due to the fact that cu^2 in the perfect model is replaced by cu in the approximate model while the amplitude of u in the perfect model is often larger than 1 especially in the intermittent phases. Therefore, the coefficient c has to be increased in order to retain the amplitude of the feedback from u to v. On the other hand, according to the second equation in (18), due to the increase of c, the amplitude of u will increase as well especially for the intermittent phase. Therefore, the forcing F_u in the approximate model is decreased in order to retain the amplitude of the observed variable u as in the perfect model.

C. Long-term prediction

With the estimated parameters in hand, we first compare the long range forecasts between the perfect dyad model (16) and the approximate model with the linear feedback (18).

In Panels (d)–(f) of Figure 6, the trajectories, the PDFs and the ACFs associated with the approximate model are shown, where for a fair comparison of the time series, the same random number seeds are used. The recovered trajectory of the observed variable v using the approximate model with the linear feedback term almost perfectly matches that of the truth (with Corr = 0.998 and RMSE = 0.011).

Now let us focus on the hidden intermittent variable u. Comparing the second and the fourth rows of Figure 6, it is clear that the approximate model with the linear feedback (18) is skillful in generating the intermittent extreme events in u. In fact, the pattern correlation between the two time series in these two rows is 0.93, which also indicates that the approximate model is able to capture the timing of the occurrence of extreme events. Yet, there are two main errors in the approximate model. First, the amplitudes of the intermittent events seem to be slightly overestimated. This is easy to understand because in order to reach the same observed trajectory v, the linear feedback requires a larger u in the approximate model than the quadratic nonlinear feedback in the perfect model. Second, the quiescent events also seem to be slightly underestimated in the approximate model. This results in the fact that the peak of the associated PDF is closer to zero than that of the true signal. The model error

$$\mathcal{E}_{eq} = \mathcal{P}(p_{eq}, p_{eq}^M) = 0.46, \tag{20}$$

which, although is non-negligible, comes largely from the quiescent events. The PDFs of u associated with both the perfect and approximate models are skewed with an one-sided fat tail. Therefore, the long range behavior of the approximate model in capturing the information in the tail that corresponds to the extreme events remains similar to that of the perfect model.

Another conclusion drawn from Figure 6 is that the ACFs of u and v associated with both the perfect model and the approximate model are very similar to each other, decaying to zero after one time unit.

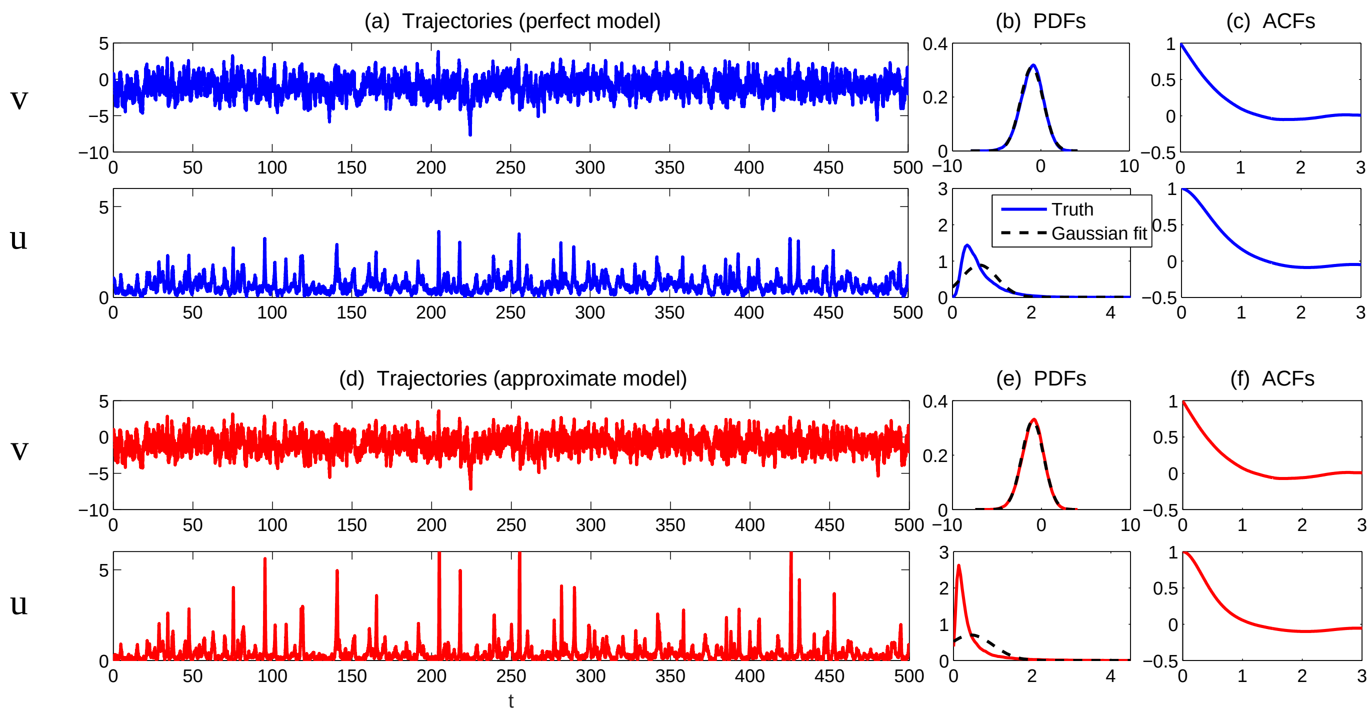

FIG. 6. Panels (a)–(c): Time series, PDFs and ACFs of the dyad model (16) with the parameters in (17). Panels (d)–(f): those of the approximate model (17) with the estimated parameters in (19).

D. Data assimilation, Short- and medium-range forecasts

Given the observation in v, the assimilated u is shown in Figure 7. Overall, the assimilated signal of the hidden variable u and the truth have a very good match in terms of the patterns. Yet, due to the intrinsic model error as discussed above, the quiescent and intermittent phases are slightly underestimated and overestimated, respectively.

Panels (a)–(b) of Figure 8 show the RMSE and the Corr between the true signal and the ensemble mean predictions as a function of lead time. Except at the very short lead time, where the data assimilation results in some uncertainties in the initial values, the approximate model essentially gives the same prediction skill as the perfect model in terms of the RMSE and the Corr. This indicates the overall skillful prediction using the approximate model. Note that since our focus is the extreme events in the hidden process, some extra information beyond the RMSE and Corr needs to be explored. In Panels (c)–(d), a comparison of the medium range forecasts and the forecast PDFs at lead time $t = 0.6$ is shown. It is clear that the approximate model is more skillful in capturing the extreme events and the fat tail of the predicted PDF than the perfect model. This is not surprising. In fact, it is well known that the amplitude of the ensemble mean prediction decays as time evolves. On the other hand, the slight overestimation of the amplitude of u in the approximate model compensates the underestimation of the amplitudes in the ensemble mean forecast, which is crucial in predicting extreme events at the medium range.

Figure 9 shows four case studies of the time evolution of the predicted PDFs starting from different initial phases. The predictions of v using both the perfect model and the approximate model are overall similar to each other. Note that more ensemble members in the prediction using the approximate model are actually able to forecast the extreme events than the perfect model. This feature is quite useful for medium-range forecast, especially when the starting time is an onset phase of the extreme events in u (Cases 1 and 2). In addition, even with some errors in the initial condition due to the data assimilation (Case 4), the approximate model is still able to capture the time evolution of extreme events.

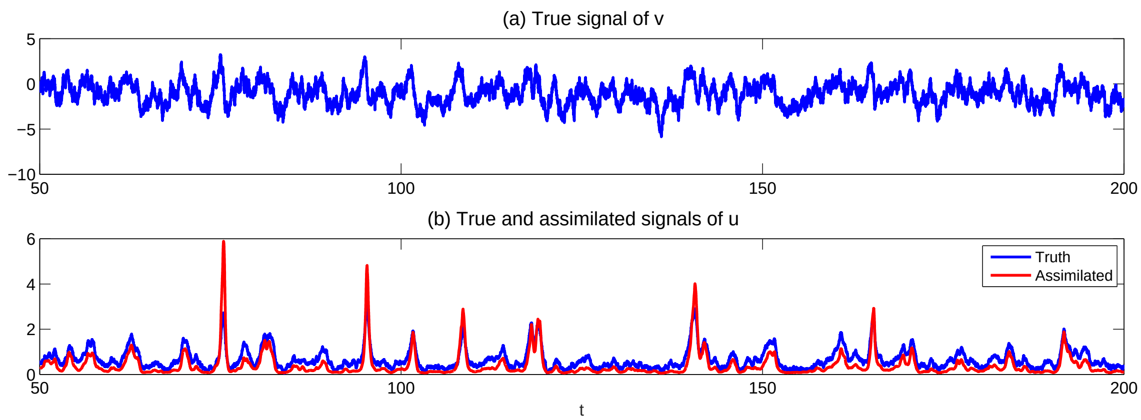

FIG. 7. Panel (a): the true signal of v from the perfect dyad model (16). Panel (b): the true signal of u from the perfect dyad model (16) and assimilated signals of u using the approximate model (18).

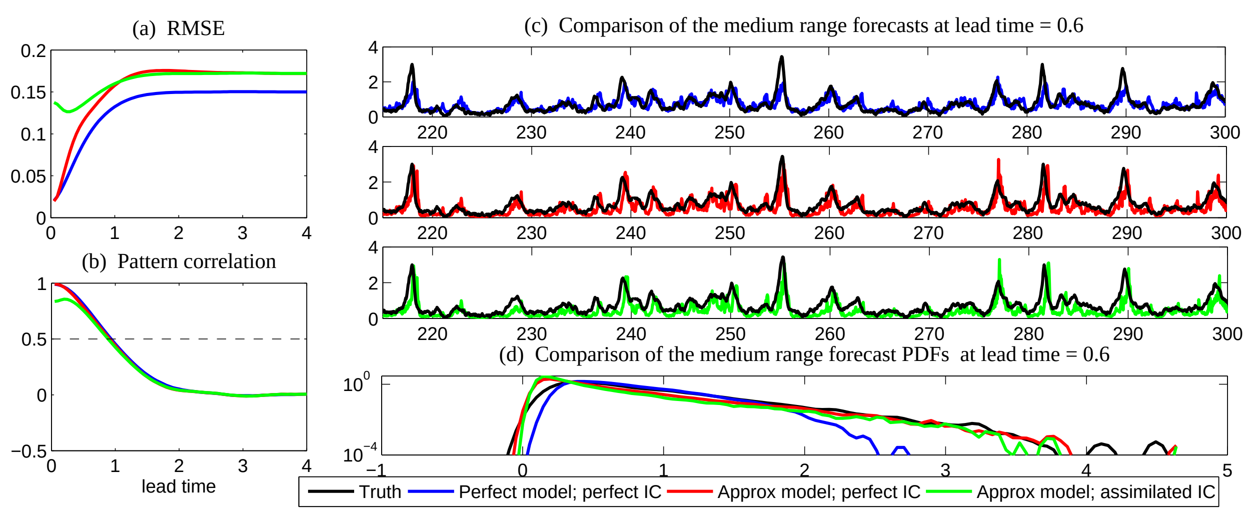

FIG. 8. Short- and medium-range forecasts based on the ensemble mean forecast. Panel (a)–(b): RMSE and Corr between the true signal and the prediction ones as a function of lead time. Blue: perfect model prediction with the perfect initial condition. Red: approximate model prediction with perfect initial condition. Green: approximate model prediction with assimilated initial conditions. Panel (c): comparison of the medium range forecasts at lead time $t = 0.6$. Panel (d): comparison of the medium range forecast PDFs at lead time $t = 0.6$.

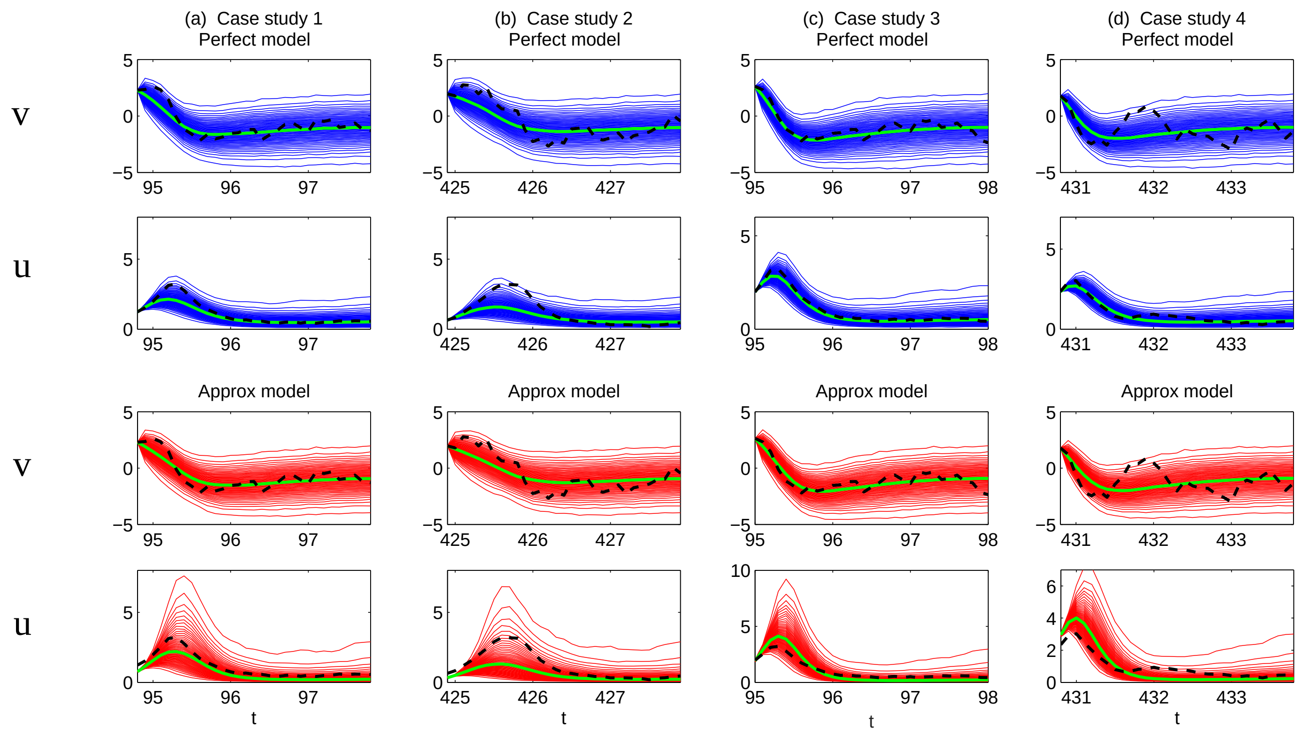

FIG. 9. Case studies. Time evolution of the predicted PDFs starting from different initial phases. Each PDF is shown with 50 thin curves, which represent the 1st, 3nd, 5th, . . . , 97th and 99th percentiles of the of the PDF. The green curve represents the mode of the PDF since the PDF is non-Gaussian. The black dashed curve is the true signal. Columns (a)–(b): starting from an onset phase of extreme events. Columns (c)–(d): starting from a mature phase of extreme events.

E. Prediction with an initial value starting outside the attractor

In this subsection, we compare the predictions when the initial values are outside the attractor. According to Figure 6, the attractor of u contains values that are positive but almost always stay below $u = 4$. Thus, we consider the following two situations: A). the hidden variable u starting from a negative value, and B). the hidden variable u starting from a large positive value.

A. The hidden variable u starting from a negative value.

In Panels (a)–(f) of Figure 10, we show the prediction where u starts from a negative value $u(0) < 0$. Here, v always starts from its equilibrium mean value $v(0) = -0.9584$. In Panels (a)–(b), $u(0) = -0.2$ is slightly negative. The approximate model behaves in a similar way as the perfect model, where after a short term, the trajectory will arrive at the attractor. However, when $u(0) = -0.5$ as shown in Panels (c)–(d), some of the ensemble members in the approximate model blows up in finite time (around $t = 1.5$). See the second row of Panel (d). Such a behavior becomes even worse when $u(0)$ is decreased to $u(0) = -0.8$ as shown in Panels (e)–(f), where quite a few ensemble members blow up in a short finite time (around $t = 0.5$ to $t = 1.5$). Panel (g) of Figure 10 shows the percentage of the events that blow up as a function of the initial value $u(0)$. As expected, with the decrease of $u(0)$, the number of blowup events increases.

Now we look at both the perfect and approximate models (16) and (18) to understand the mechanism that leads to such a finite time blowup issue in the approximate model. First, when u and v are at the attractor, u stays in positive values. When the amplitude of u increases due to a negative value of v, both the linear and nonlinear feedback in (16) and (18) will push v back to a negative value and the consequence is that v will strongly damp u and decreases the amplitude of u. However, when u is negative, the nonlinear feedback $-cu^2$ and the linear feedback $-cu$ will play completely different roles since $-cu^2 < 0$ while $-cu > 0$. The dynamical property of the perfect dyad model (16) remains unchanged. But the blowup issue appears in the approximate model (18). In fact, once u is negative, the linear feedback will make v become positive. As a result, the positive anti-damping of v will further increase the amplitude of u, which makes u blow up in a short time. When the initial value $u(0)$ has a small amplitude (e.g., $u(0) = -0.2$), the forcing $F_u = 1 > 0$ may be able to overcome the anti-damping in the short term and push the solution to the attractor.

But if the amplitude of $u(0)$ is large, then the role of F_u is weaker than the anti-damping from v, and the solution has a much higher chance to blow up.

B. the hidden variable u starting from a large positive value.

Now we let the hidden variable u start from a large positive value and study how the solution adjusts to the attractor. See Panels (h)–(k) in Figure 10.

First, with a moderately large initial condition $u(0) = 5$ as shown in Panels (h)–(i), the hidden variable u using the approximate model releases to the attractor in almost the same way as that using the perfect model. The trajectories of v are slightly different, but since v is always very negative, the strong damping of v makes the trajectories of u in the two models have very similar behavior.

Next, we increase the initial condition to $u(0) = 10$. Then we first notice a more significant different in the predicted trajectory of v, where in a short term $t = 0.2$ the true trajectory and the perfect model prediction can reach $v = -8$ while the approximate model only allows $v = -3$. This is due to the model error in the feedback terms. In fact, when u is large, $-cu^2$ in the perfect model will be much larger than $-cu$ in the approximate model. This leads to the large error in v. As a result, the damping in the approximate model then becomes much weaker compared with the perfect model. Therefore, u releases slower in the approximate model (see the second row of Panel (k)). Notably, the ensemble prediction in the second row of Panel (k) seems not to be too far from the truth (black dashed curve). But the truth is outside 99 percentile of the prediction (the most bottom red curve) when t is between $t = 0.1$ and $t = 0.5$.

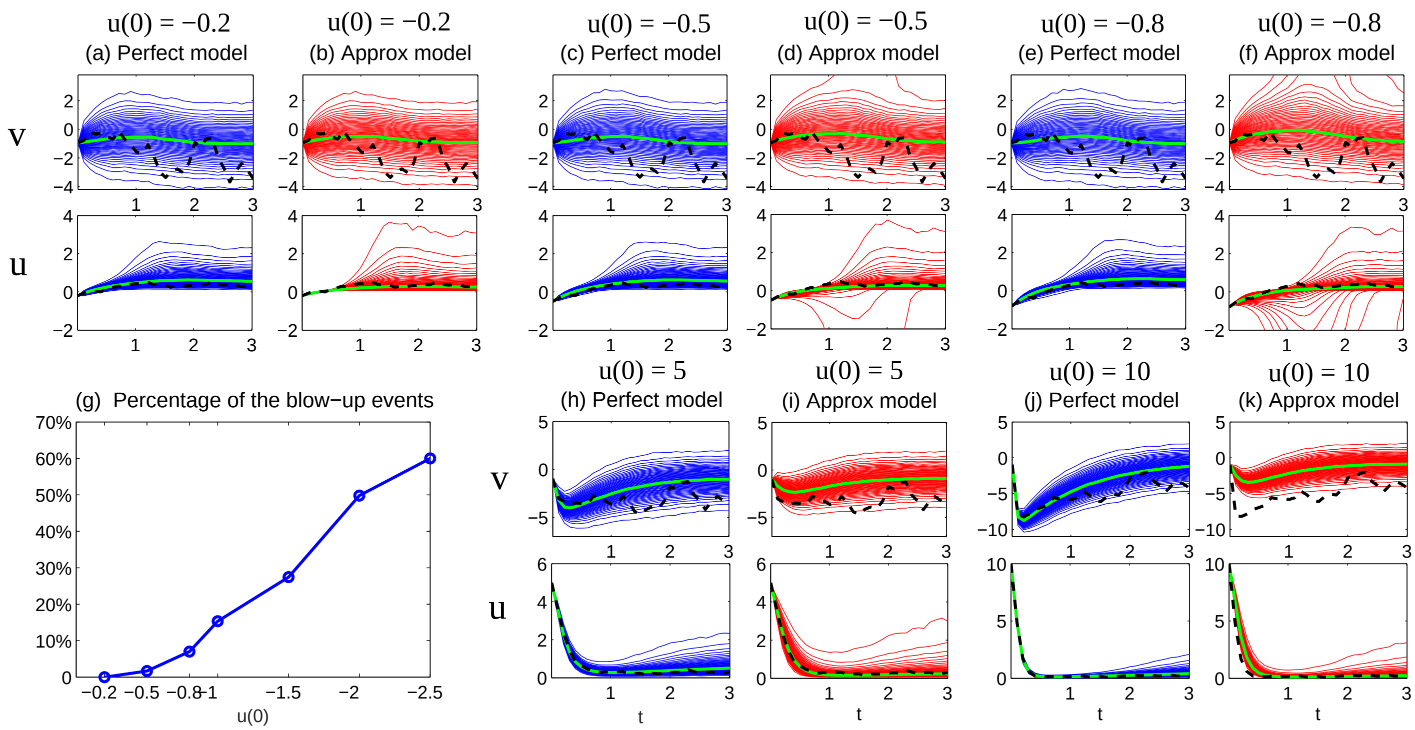

FIG. 10. Prediction of the dyad model with initial values being outside the attractor. Panels (a)–(f) and (h)–(k): Time evolution of the predicted PDFs starting from different initial phases. Each PDF is shown with 50 thin curves, which represent the 1st, 3nd, 5th, . . . , 97th and 99th percentiles of the of the PDF. The green curve represents the mode of the PDF since the PDF is non-Gaussian. The black dashed curve is the true signal. In Panels (a)–(f), the hidden variable u starting from a negative value. In panels (h)–(k), the hidden variable u starting from a large positive value. Panel (g) shows the percentage of the events that blow up as a function of the initial value $u(0)$.

F. Dynamical regimes with different F_u

So far, we have focused on the regime with $F_u = 1$. In this subsection, the role of F_u will be explored and dynamical regimes with different F_u will be studied for predicting the hidden extreme events.

In Panel (a) of Figure 11, the trajectories of u from the perfect model (16) with different F_u are shown. Here, the same random number seeds are used in generating these time series for a fair comparison.

Regime I: $0.7 \le F_u$**.**

When F_u is sufficiently large, the approximate model with the linear feedback (18) is a suitable model for predicting the hidden extreme events.

Regime II: $0 \le F_u < 0.3$**.**

When F_u approaches zero, the intermittent events in u can have both signs. As was discussed in Section V E, when u is negative, the linear feedback $-cu$ in (18) will play a significant different role compared with the nonlinear feedback $-cu^2$ in the perfect model (16). In fact, the linear feedback $-cu$ becomes positive and make v to be positive. Then the anti-damping of v in the process of u leads to the finite time blowup. Therefore, we conclude that using the approximate model (18) with a linear feedback to predict the hidden extreme events in u requires that the forcing F_u in the perfect dyad model cannot be too small. If the forcing F_u in the perfect dyad model is too small, then the approximate model does not have a mechanism to recover the intermittent events in u when u is negative. A new approximate model that has skill in capturing the extreme events with both signs needs to be developed.

Regime III: $0.3 \le F_u < 0.5$**.**

When $F_u \ge 0.3$, the intermittent events in the true signal of u only occur in the positive phase. However, the true trajectory of u still goes below 0 quite frequently (with small amplitudes). Panel (c) of Figure 11 shows the data assimilation of u using the approximate model with the linear feedback (18), where the parameters are re-estimated based on the observed signal of v in $F_u = 0.3$ regime. One important result is that the assimilated state of u can occasionally become quite negative! In fact, as is shown in Panels (b)–(c), before the assimilated u goes to a negative value, the signal of v is large and positive while u is nearly zero. Therefore, when the trajectory of u becomes slightly negative in the true signal, the anti-damping v will amplify the negative phase of u. Since the positive forcing $F_u = 0.3$ here is pretty weak, this forcing is unable to push u back to the attractor with positive values immediately and therefore the assimilated state of u will stay in the negative phase for a while. According to the discussions in Section V E, if the prediction starts with a large negative value of u, then even for a short term, the prediction using the approximate model may suffer from a short-term blowup[110].

Regime IV: $0.5 \leq F_u < 0.7$**.**

Now the data assimilation results using the approximate model (18) provides the state of u that is always positive. Thus, there will be no issue in data assimilation. However, as shown in Panels (d)–(f) of Figure 11, the approximate model can still suffer from a long (but finite) time blow up issue. This is again related to the insufficient strength of F_u. In fact, the trajectory of u still has some chances to become slightly negative and the corresponding values of v at these time instants are usually large. Therefore, the anti-damping v and the forcing F_u in the process of u compete with each other. If the strength of forcing is not strong enough, then for some events, the anti-damping can results in the blowup issue of u[110].

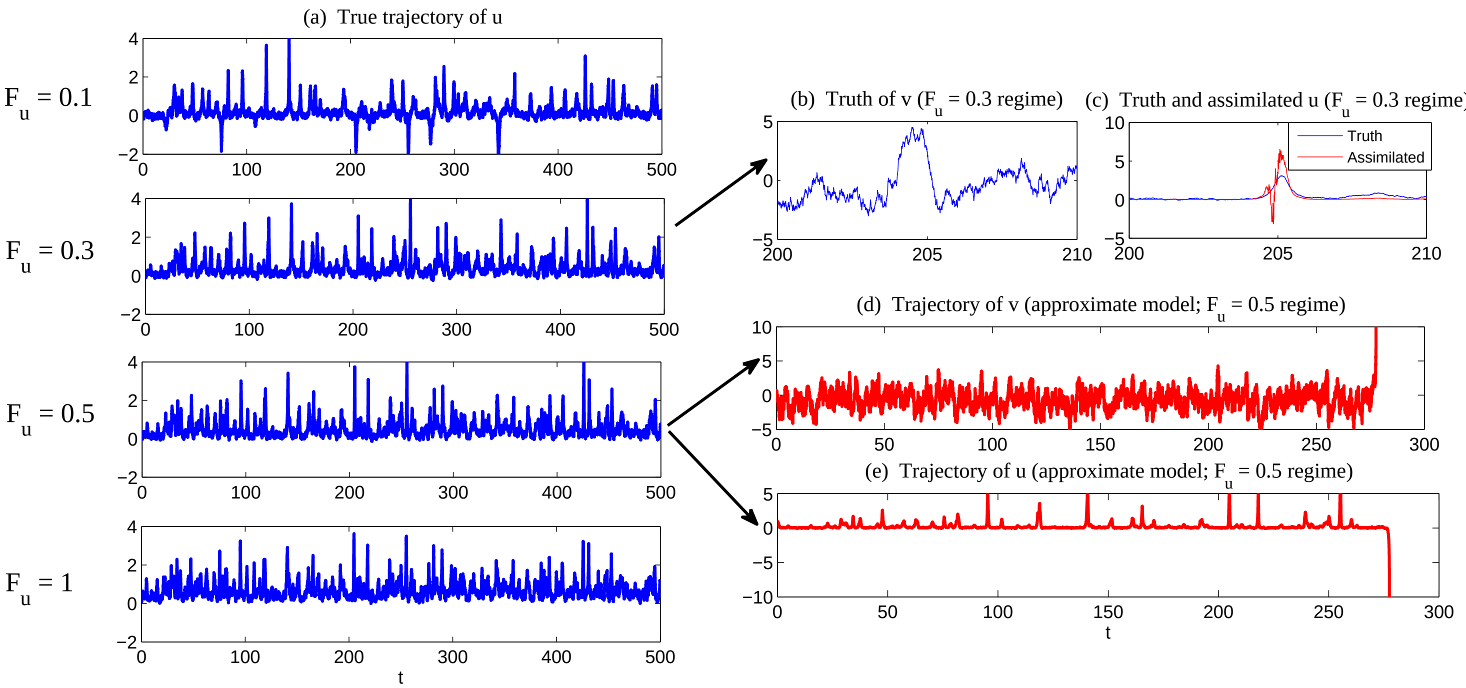

FIG. 11. Dynamical regimes with different F_u. Panel (a): trajectories of u from the perfect model (16) with different F_u are shown. Here, the same random number seeds are used in generating these time series for a fair comparison. Panel (b): True signal of v in $F_u = 0.3$ regime from the perfect model. Panel (d): True signal of u in $F_u = 0.3$ regime from the perfect model (blue) and the assimilated posterior mean using the approximate model. Panels (e) and (f): trajectories of the approximate model with the estimated parameters from the observed true signal of v in $F_u = 0.5$ regime.

VI. THE LORENZ 63 MODEL

In many applications with chaotic or turbulent phenomena, due to the incomplete knowledge of the underlying dynamics, noise inflation is often incorporated into the dynamical processes[21,104,111]. The enhanced noise plays the role of parameterizing small-scale fluctuations, which helps increase the variability of the system and has a potential of improving the data assimilation and prediction skill. Yet, it has not been well understood the effect of noise inflation in the extreme events prediction. Therefore, in this and the next two sections (Section VII and Section VIII), noise inflation will be incorporated into the dynamical systems for testing the ensemble prediction skill of the observed and hidden extreme events as well as other non-Gaussian characteristics. The difference between the studies in these three sections is as follows. In Section VII and Section VIII, the noise inflation will be combined with various effective and practical strategies for developing effective and simple approximate models for improving the prediction of the extreme events resulting from complicated turbulent dynamical systems with regime switching. In this section, the chaotic Lorenz 63 model is used as a testbed to understand the skill of the extreme events predictions, where the inflated noise acts as the only source of the model error.

A. The perfect and approximate models

The model considered in this section is the Lorenz 63 model[72]. It is a simplified mathematical model for atmospheric convection with chaotic behavior. The equations relate the properties of a two-dimensional fluid layer uniformly warmed from below and cooled from above. In particular, the equations describe the rate of change of three quantities with respect to time: x is proportional to the rate of convection, y to the horizontal temperature variation, and z to the vertical temperature variation. The constants σ, ρ, and β are system parameters proportional to the Prandtl number, Rayleigh number, and certain physical dimensions of the layer itself[112]. The Lorenz 63 model is also widely used as simplified models for lasers, dynamos, thermosyphons, electric circuits, chemical reactions and forward osmosis[113–119].

Here, we study a slightly different version of the original Lorenz 63 model by adding a small noise into the x process. We also assume that only a short trajectory of x is observed as the training data while y and z are the unobserved variables. The model reads:

$$dx = \sigma(y - x)dt + \sigma_x dW_x, \tag{21a}$$

$$dy = \big(x(\rho - z) - y\big)dt, \tag{21b}$$

$$dz = (xy - \beta z)dt, \tag{21c}$$

The small noise here can be regarded as the observational or measurement uncertainty. It also helps prevent the singularity in the data assimilation formula in (3), which requires a non-zero noise in the observational process. Nevertheless, with a small noise coefficient, the dynamical behavior of the model in (21) remains almost the same as the original noise-free Lorenz 63 model. Below, we always take $\sigma_x = 1$, which is a sufficiently small value. The other parameters that are used to generate the true signals of (21) are

$$\sigma = 10, \qquad \rho = 28, \qquad \beta = 8/3. \tag{22}$$

These are the classical choices of the Lorenz 63 model. Figure 12 shows the trajectories, PDFs and phase plots of the Lorenz 63 model (21), where the butterfly profile in the phase plots and the chaotic features in the model trajectories are clearly demonstrated. Notably, there are quite a few extreme events that appear in all the three components due to the fact that one of the Lyapunov exponents of the Lorenz 63 system is positive. These extreme events occur when the system states switch between the two branches of the "butterfly wings".

The short trajectory of x in Panel (a) of Figure 12 with only 50 units will be used as the observed training data for the approximate models below.

The approximate models.

Below, we aim at understanding the model error that comes from the noise inflation. To this end, it is natural to propose the following approximate model,

$$\begin{aligned} dx &= \sigma(y - x)dt + \sigma_x dW_x, \\ dy &= \big(x(\rho - z) - y\big)dt + \sigma_y dW_y, \\ dz &= (xy - \beta z)dt + \sigma_z dW_z, \end{aligned} \tag{23}$$

where the noise coefficients σ_y and σ_z are given and fixed empirically, which account for the noise inflation in the hidden variables. Note that here the deterministic parts in the perfect model (21) and the approximate model (23) are the same, which is not always the case in

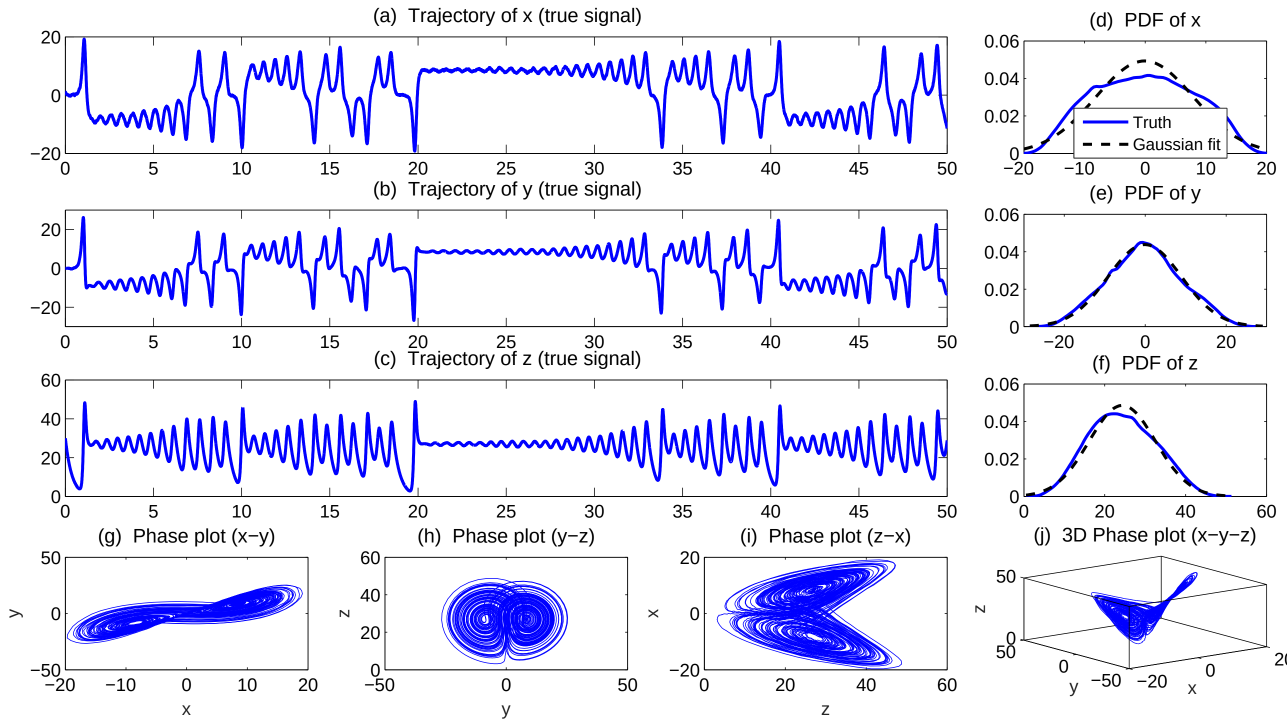

FIG. 12. Trajectories (Panels (a)–(c)), PDFs (Panels (d)–(f)) and the phase plots (Panels (g)–(j)) of the noisy Lorenz 63 model (21) with parameters in (22) and a small noise coefficient $\sigma_x = 1$.

real applications where noise inflation is often used to compensate other model errors and simplifications. Nevertheless, the setup here allows us to understand the model error that comes purely from the noise inflation and its effect on the prediction skill.

The noise coefficient σ_x will be estimated from the parameter estimation algorithm. Note that since σ_x is associated with the quadratic variation of the continuously observed training data, a prescribed value with inflation may lead to pathological behavior of the parameter estimation. Depending on the level of noise inflation, we consider the following three approximate models,

$$
\begin{aligned}
&\text{Approximate model I.} && \text{Small noise inflation:} && \sigma_y = \sigma_z = 1,\\
&\text{Approximate model II.} && \text{Moderate noise inflation:} && \sigma_y = \sigma_z = 3,\\
&\text{Approximate model III.} && \text{Large noise inflation:} && \sigma_y = \sigma_z = 5.
\end{aligned}
\tag{24}
$$

B. Parameter estimation

In the approximate models, there are four parameters to be estimated: ρ, σ, β and σ_x. Here the parameter estimation algorithm as described in Section III D is run up to $K = 15000$ steps and the averaged values of the trace plots from $k = 5000$ to $k = 15000$ are used as the estimated parameters, which are:

$$
\begin{aligned}
&\text{Approx model I:} && \rho = 27.48, && \sigma = 10.34, && \beta = 2.70, && \sigma_x = 1.03,\\
&\text{Approx model II:} && \rho = 31.04, && \sigma = 9.051, && \beta = 2.33, && \sigma_x = 1.06,\\
&\text{Approx model III:} && \rho = 34.17, && \sigma = 7.525, && \beta = 2.20, && \sigma_x = 1.08.
\end{aligned} \tag{25}
$$

Note that due to the model error from noise inflation, the estimated parameters in the approximate models are not exactly the same as those in the perfect model. In particular, with the increase of the noise coefficients σ_y and σ_z, the estimated parameter ρ and σ seem to be more different compared with the one in the perfect model in order to compensate the model error.

C. Data assimilation

Figure 13 shows the data assimilation results using the approximate model (23) with the estimated parameters, where the true signal of the observed variable x is generated using the perfect model (21).

In the approximate model I, due to the small model error in the inflated noise coefficients, the assimilated values of y and z are nearly perfect and the uncertainty reflected by the posterior variance in both variables is small. In the approximate model II, the assimilated values of y are still quite accurate but those of z show some errors where the mean state of z has a slight shift towards the positive value. Such a bias in the assimilated posterior mean state is possibly due to the fact that the noise σ_y leads to the change of the mean value of xy in z process since x and y are highly correlated. On the other hand, x and z are not so closely correlated, and therefore the mean value of xz that contributes to the mean state of y is hardly polluted by the noise. Finally, in the approximate model III, where the inflated noise coefficients are large, there are some non-negligible errors in the assimilated states of z and the associated uncertainty increases as well. Nevertheless, despite such a mean state shift, the overall patterns and amplitudes of z are assimilated quite well.

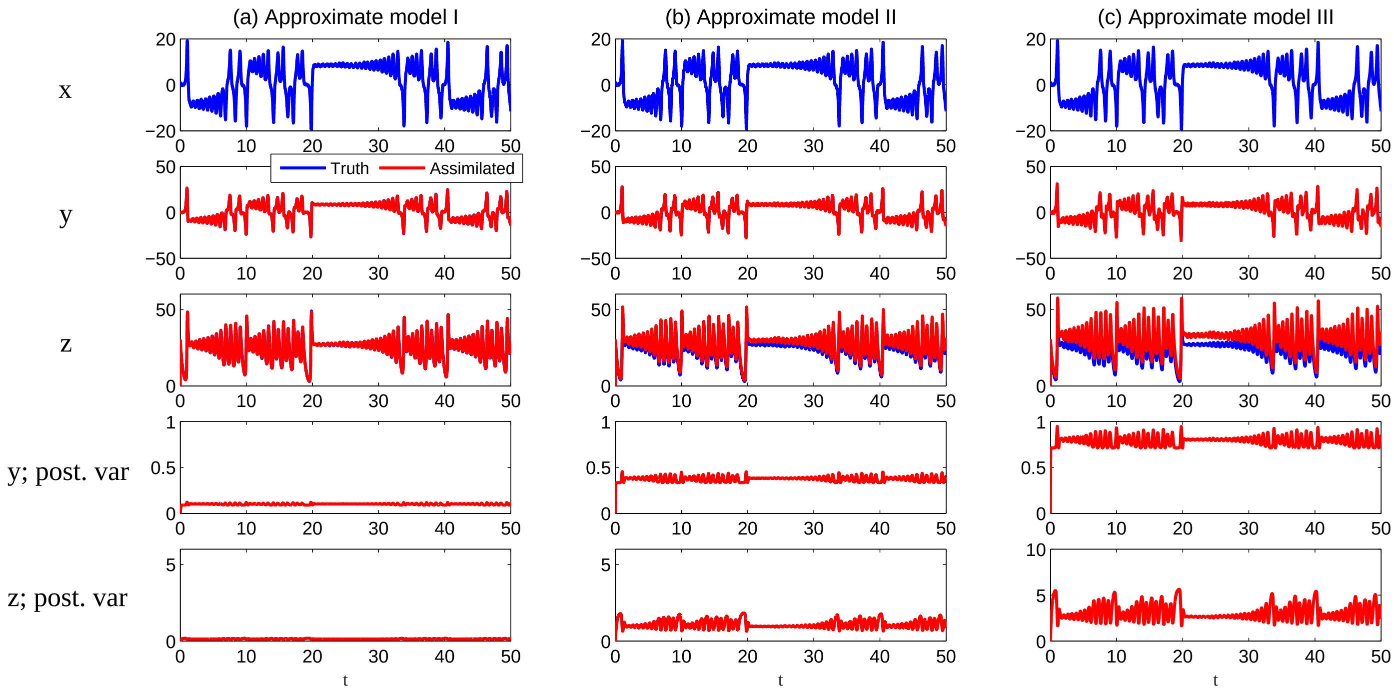

FIG. 13. Data assimilation using the approximate model (23) with different noise inflation levels (24). The first row shows the true trajectory of x. The second and third rows show the true signals of y and z as well as the posterior mean estimations from data assimilation (red). The fourth and fifth rows show the posterior variance of y and z, respectively.

D. Long-range forecast

To quantify the long-range forecast skill, the comparison of the equilibrium PDFs and the ACFs between the perfect model and approximate models is shown in Figure 14.

First, all the three approximate models are able to capture the equilibrium non-Gaussian PDFs of both x and y with high accuracy, where the information model error in the equilibrium PDF $\mathcal{P}(p_{eq}, p_{eq}^M) \leq 0.05$ is tiny even using the approximate model III. For the variable z using the approximate model III, the error is slightly larger $\mathcal{P}(p_{eq}, p_{eq}^M) = 0.28$ but is still acceptable. Such a model error is due to the fact that the PDF associated with z using the approximate model has a mean shift compared with the truth, which has already been seen in the data assimilation results.

Next, the approximate models and the perfect model overall share quite similar ACFs, indicating similar time evolution behavior (at least up to the second order statistics in time). In particular, the direct relaxation of the ACFs of x, y and the oscillated relaxation of that of z are both captured by the approximate models. The only non-negligible difference appears in the ACF of z when the noise inflation level is large, i.e., in approximate model III, where the approximate model has a slightly faster decaying ACF.

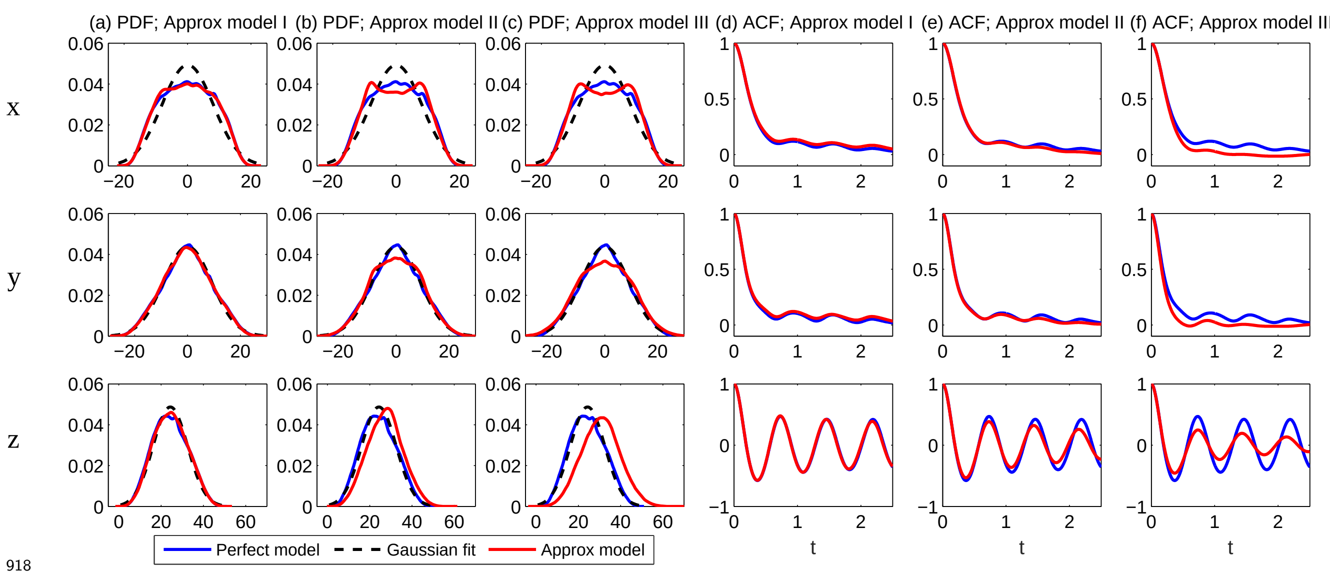

FIG. 14. Comparison of the PDFs and the ACFs of the perfect model and the approximate model.

E. Short- and medium-range forecasts

To study the short- and medium-range forecasts, we show in Figure 15 three skill scores of the predictions as a function of lead time. Two of them, namely the RMSE and the Corr, are the classical path-wise measurements while the third one is an information criterion, that is, the relative entropy (7) between the PDF of the predicted time series and that of the truth. In order to distinguish the errors due to the noise inflation and the initial uncertainty with data assimilation, we show the predictions using the approximate model with either assimilated initial conditions (ICs) or with perfect initial conditions. All the predictions here are based on the ensemble mean, which is the average of 50 ensemble members.

Columns (a)–(b) and Columns (d)–(e) show the RMSE and Corr of the predictions using the approximate models I and II, respectively. These path-wise measurements indicate that the skillful predictions of the approximate models regardless of using perfect or assimilated initial conditions are up to nearly 3 time units. However, the conclusion based on these path-wise measurements can be misleading in this strongly chaotic system. In Columns (g)–(h), the relative entropy has a significant increase as the lead time, especially using the approximate model II. This implies certain non-negligible errors are not captured by the two path-wise measurements. To see such errors, the ensemble mean prediction using the approximate models (green) and the truth (blue) at lead time $t = 1$ are compared in Figure 16. Both the trajectories and the PDFs are shown in order to compare the path-wise and the information measurements. Note that only the Gaussian fits of the PDFs are shown here for the purpose of comparing the variance in the truth and the predicted PDFs which reflects the skill of capturing the amplitudes especially those of the extreme events. In Column (b) of Figure 16, it is shown that although the patterns of the predicted signal are quite consistent with the truth, the amplitudes of all the extreme events are underestimated to some extent. Thus, the predicted PDF has a narrower shape compared with the truth. Such a phenomenon becomes more significant in Column (c) of Figure 16 where the approximate model III is used. At lead time $t = 1$, despite that Corr ≈ 0.8 for x and y and Corr ≈ 0.5 for z remain skillful, the large values of the relative entropy clearly indicate the discrepancy between the predicted PDF and the truth, which is due to the fact that the amplitudes of the extreme evens are severely underestimated. These facts conclude the importance of using the information criterion in quantifying the model error in the PDFs in addition to the path-wise measurements.

Figure 17 includes a case study of the ensemble prediction for short- and medium range forecasts using approximate models I and III starting from $t = 16.5$. First, the ensemble mean (green) using the approximate model I is skillful up to 2.5 units of the lead time while that using the approximate model III has a much shorter skillful prediction. These are consistent with the results shown in Figure 16. Next, the uncertainty of the prediction is reflected in the ensemble spread. It is clear that in the perfect model prediction the ensembles do not spread out until $t = 19$ while those in the approximate models start spreading out around $t = 17$. This is obviously due to the fact that the noise level is higher in the approximate models. Using approximate model I, despite some members give false prediction due to the intrinsic chaotic behavior, most of the ensemble members are still able to follow the true trajectories, which also results in the skillful ensemble mean prediction. However, using the approximate model III, both the large noise inflation and the initial uncertainty due to the data assimilation lead to a quick divergence of the ensembles. The ensemble spread is able to tell the uncertainty but the ensembles reach the attractor much faster than those using the perfect model and therefore the ensemble mean using the approximate model losses its skill.

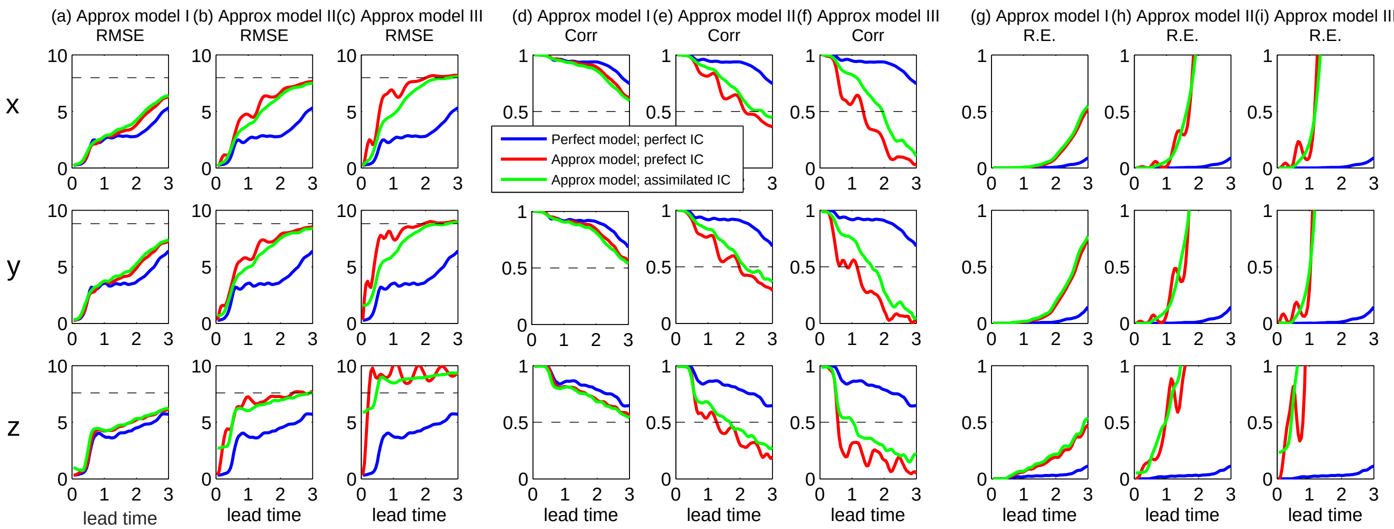

FIG. 15. RMSE (Panels (a)–(c)), Corr (Panels (d)–(f)) and relative entropy (R.E.; Panels (g)–(i)) as a function of lead time for short- and medium-range forecasts using the perfect model (21) (blue) and the three approximate models (23)–(24) with perfect initial conditions (red) and assimilated initial conditions (green). The prediction here is based on the ensemble mean. The dashed black lines in the RMSE panels show one standard deviation of the true signal and those in the Corr panels show the Corr $= 0.5$ threshold.

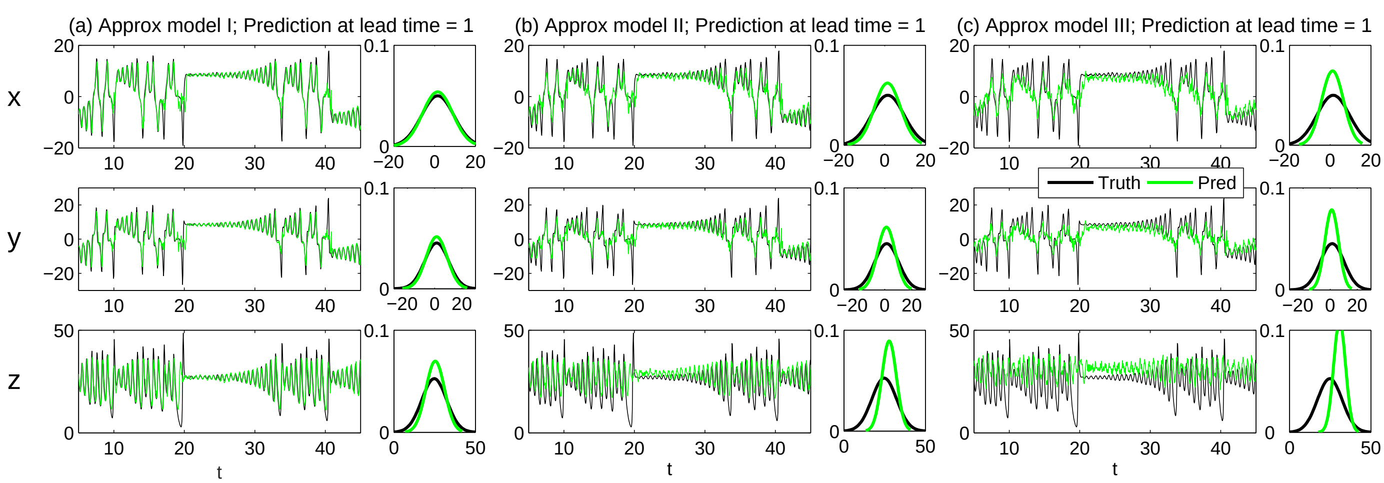

FIG. 16. Comparison of the ensemble mean prediction using the approximate models and the assimilated initial conditions (green) with the truth (blue) at lead time $t = 1$. In each panel, both the trajectories and the Gaussian fits of the PDFs are shown.

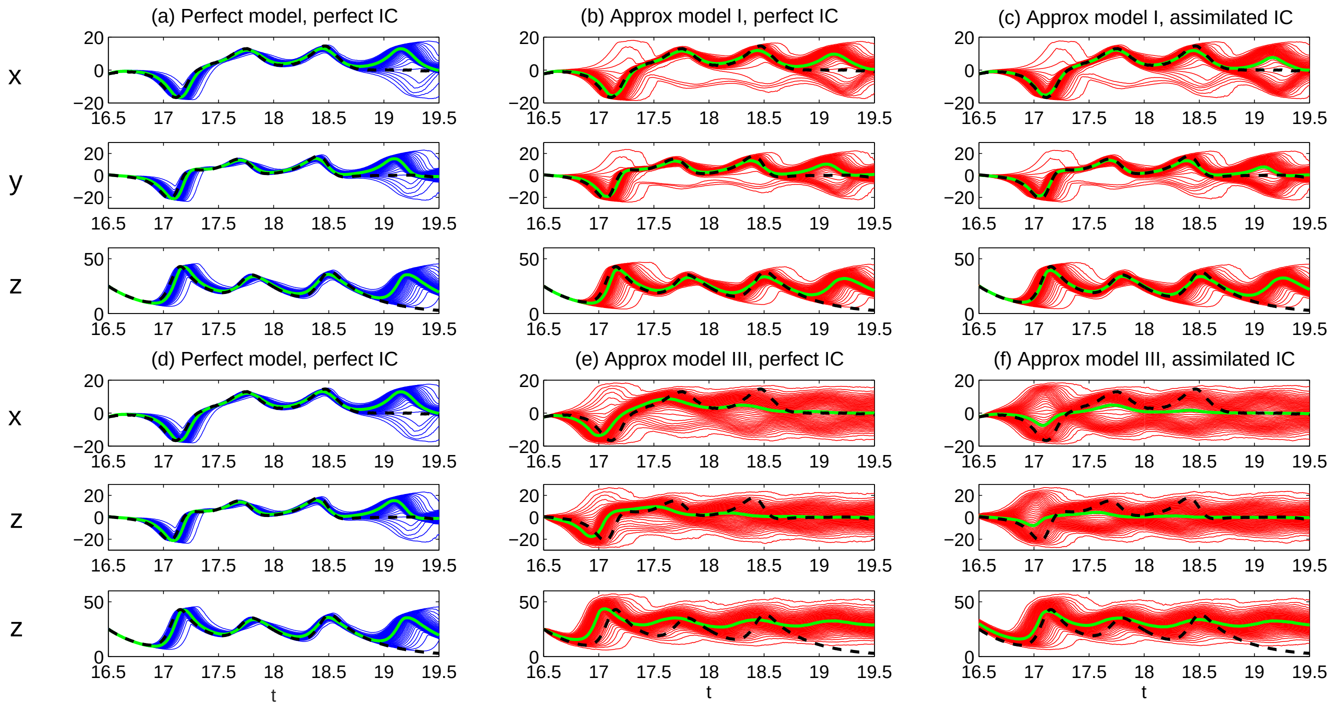

FIG. 17. Case studies of the ensemble forecasts. Panels (a)–(c): a case study using the approximate model I. Panels (d)–(f): the same case study using the approximate model III. Each subpanel shows the time evolution of the prediction, as represented by the time dependent PDF of the ensemble forecast. Note that each PDF is shown with 50 thin curves, which represent the 1st, 3nd, 5th, . . . , 97th and 99th percentiles of the of the PDF. The green curve represents the mode of the PDF since the PDF is non-Gaussian. The black dashed curve is the true signal.

F. Prediction with an initial value starting outside the attractor

Panels (a)–(f) and Panels (g)–(l) in Figure 18 show the prediction where the initial values of the observed variables $x(0) = 150$ and those of the unobserved ones $y(0) = z(0) = 150$ are outside the attractor, respectively.

The skill of capturing the transition behavior of the approximate models depends on the model error in the noise inflation. The approximate model I behaves almost the same as the perfect model due to its small noise inflation. The approximate model II is able to capture the transition behavior in short and medium ranges if the initial values of y and z are off the attractor. However, it fails to predict the two hidden variables after a very short period if the initial value of x is off the attractor. On the other hand, the approximate model III, which has the largest noise inflation coefficients, only has skillful prediction for a short period no matter which variable starts from a value that is outside the attractor.

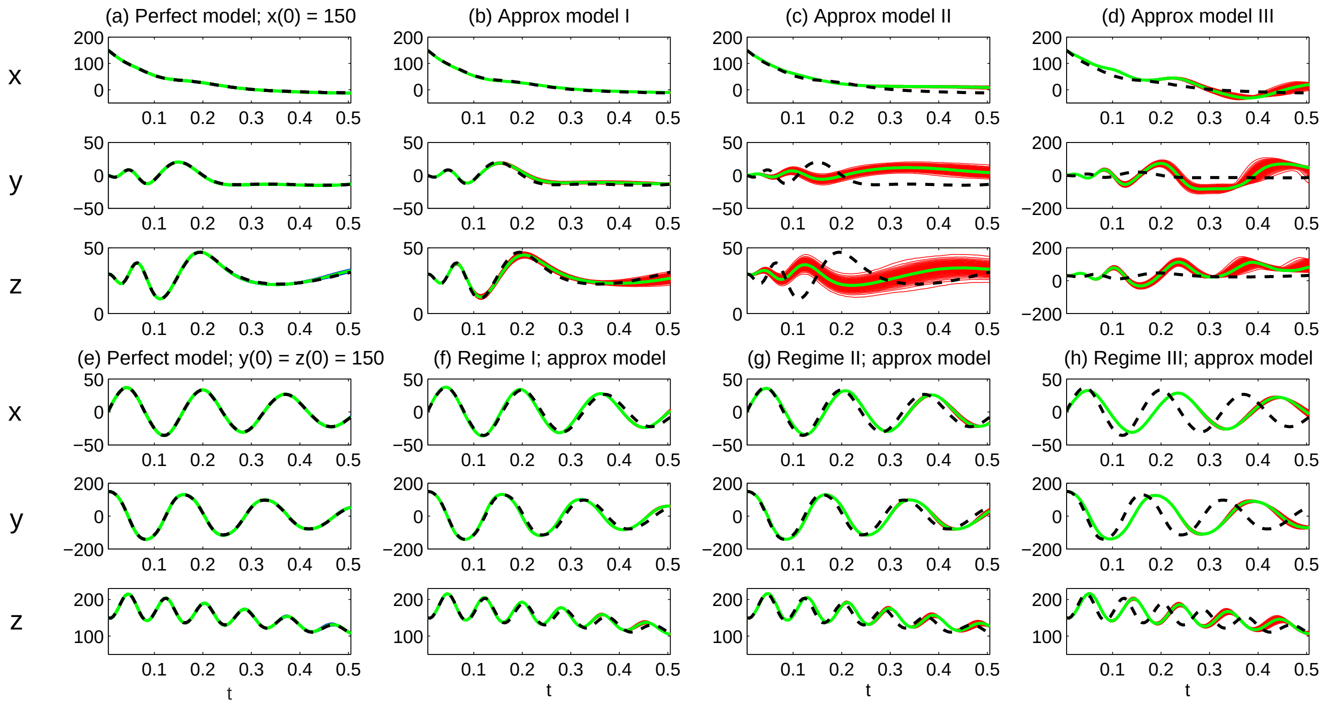

FIG. 18. Prediction with an initial value starting outside the attractor. Panels (a)–(d): Prediction where x starts at $x = 150$, which is a value that is off the attractor. Panels (e)–(h): Prediction where both y and z start at $y = z = 150$, which are values that are off the attractor. Each PDF is shown with 50 thin curves (blue for the perfect model and red for the approximate model), which represent the 1st, 3nd, 5th, …, 97th and 99th percentiles of the of the PDF. The green curve represents the mode of the PDF since the PDF is non-Gaussian. The black dashed curve is the true signal.

VII. A PARADIGM MODELS FOR TOPOGRAPHIC MEAN FLOW INTERACTION WITH REGIME SWITCHING BEHAVIOR

Regime switching between multiple metastable states is a key feature in many nonlinear turbulent dynamical systems[120–122]. One example is the atmospheric flow regimes, which represent the recurrence of certain flow structures despite the intrinsic chaotic behavior of the underlying system. The existence of persistent or recurrent weather patterns[123] with blockings is one of the most pronounced illustrations of synoptic-scale circulation regimes[124,125] while different circulation regimes and their switching were also found in planetary-scale patterns[126,127]. The metastable states have their unique dynamical behavior and the regime switching often triggers extreme events and other important nonlinear phenomena. Notably, the regimes can appear even though the observed data have a nearly Gaussian probability distribution[122,128,129]. Due to the highly complex nature of these regimes and their switching behavior as well as only the availability of the partial observations, it is important to develop suitable approximate models for capturing both the dynamical and statistical features of the regime switching and for predicting the associated extreme events. In this section, we concentrate on the development of nonlinear low-order models to achieve the above tasks, where the topographic effect is regarded as the result of random structures from either atmosphere or ocean in intermediate and small scale.

A. The perfect model

Consider the barotropic quasi-geostrophic equations[2],

$$\frac{\partial q}{\partial t} + \nabla^{\perp}\psi \cdot \nabla q + u(t)\frac{\partial q}{\partial x} + \beta\frac{\partial \psi}{\partial x} = 0, \tag{26a}$$

$$q = \Delta\psi + h, \tag{26b}$$

$$\frac{du}{dt} = -\!\!\!\!\!\!\int h\frac{\partial \psi}{\partial x}. \tag{26c}$$

This is an ideal model to study the complex nonlinear interaction of the large-scale and the small-scale flow and the role of the topography. The model exhibits a regime switching behavior with blocked and unblocked zonal flow structure despite that the associated PDF of the zonal flow has only a single modal. The study of this model for understanding its mathematical properties, developing reduced order models and uncertainty quantification can be found in a series of papers[2,36,53,130,131]. In particular, rigorous statistical bounds in quantifying the uncertainty for the ensemble prediction of barotropic flow over topography has been shown in a recent paper[132].

In this model, the small-scale flow is given in terms of the stream function ψ, and q is the small-scale potential vorticity. The large-scale velocity field only has the zonal component $u(t)$, and the topography is given by the function $h = h(x, y)$. The parameter $\beta > 0$ is the contribution from the beta-plane effect. Both the small-scale potential vorticity q and the small-scale stream function ψ, as well as the topography h, are assumed to be 2π-periodic functions in both variables x and y with zero average. The large-scale velocity $u(t)$ is strongly coupled with the small-scale flow through equation (26c), where the bar across the integral sign indicates that the integral has been normalized by the area of the domain of integration.

Below, we consider a special situation to the full nonlinear system, which inherits the nonlinear coupling of the small-scale flow with the large-scale mean flow via topographic stress. The model is named as the *layered topographic equations*. Here the topography is layered in the fixed direction $\vec{l} = (l_x, l_y)$. We assume that both ψ and q only depend on $\xi = \vec{l} \cdot \vec{x}$ with $\vec{x} = (x, y)$. One key feature of the layered topographic equations is that the small-scale nonlinear term in (26a), $\nabla\psi \cdot \nabla^{\perp} q$, is identically zero. Nevertheless, the nonlinear coupling due to topographic stress remains and is responsible for much of the complex behavior. Without loss of generality we can always rescale the system with $l_x \neq 0$ to align to a special case with $\vec{l} = (1, 0)$.

In such a situation, the Fourier expansion of ψ and h are given by

$$\begin{aligned} \psi(x, y, t) &= \sum_{k \neq 0} \psi_k(t) e^{ik\vec{l}\cdot\vec{x}}, \\ h(x, y) &= \sum_{k \neq 0} h_k e^{ik\vec{l}\cdot\vec{x}}, \end{aligned} \tag{27}$$

where we have assumed that the topography has zero mean with respect to spatial average, that is $h_0 = 0$. Substituting the ansatz (27) into (26) and adding stochastic forcing and damping, we arrive at the layered topographic equations in Fourier form,

$$\begin{aligned} \frac{d\psi_k}{dt} &= -d_k \psi_k + ikl_x \left(\frac{\beta}{k^2 |\vec{l}|^2} - u \right) \psi_k + i \frac{kl_x}{k^2 |\vec{l}|^2} h_k u + \sigma_k \dot{W}_k, \\ \frac{du}{dt} &= -d_u u - il_x \sum_{k \neq 0} k h_k \psi_k^* + \sigma_u \dot{W}_u, \end{aligned} \tag{28}$$

where * denotes the complex conjugate. In (28), $\psi_k, k = 1, 2, \ldots, \Lambda$ are the stream functions and u is the large-scale zonal velocity.

In the study here, we adopt $\Lambda = 10$ and therefore in total there are 21 modes in the model (28), where 1 mode u represents the large-scale zonal flow. The other 20 modes are for the small-scale stream functions with $k = \pm1, \ldots, \pm10$, which based on the layered topographic functions determine the meridional flows. We assign the following function for the topography,

$$h(x) = H_1\Big(\cos(x) + \sin(x)\Big) + H_2\Big(\cos(2x) + \sin(2x)\Big) - \frac{i}{2}\sum_{3\le k\le\Lambda}\frac{e^{i(kx+\theta_k)}}{k^p} + c.c., \qquad (29)$$

where H_1 and H_2 are associated with the leading two Fourier modes $k = \pm1, \pm2$ while the remaining part in (29) represents the amplitudes of the topography for other Fourier modes. Here θ_k are random phase and p is a power that controls the effects of the small-scale topography. The topography plays an important role in altering the stream functions. With a simple manipulation, it is easy to show that the topographic functions associated with the first two Fourier modes are

$$h_1 = H_1/2 - H_1/2i, \qquad \text{and} \qquad h_2 = H_2/2 - H_2/2i. \qquad (30)$$

The other h_k can also be written explicitly using (29). The following parameters are adopted in the study here. The beta-plane effect is $\beta = 2$. The coefficients of the topography are $H_1 = 1$ and $H_2 = 1/2$. The damping coefficients are chosen as

$$d_k = d_u = 0.0125, \qquad (31)$$

which represents a time scale of relaxation time of roughly 80 time units. Such a choice allows a relatively slow (but not infinitely slow) mixing of the system. With different choices of the stochastic noise, the system can also have fast mixing rate. Finally, the stochastic noise coefficients are chosen as follows,

$$\sigma_u = \sigma_1 = \sigma_2 = \frac{1}{20\sqrt{2}}, \qquad \sigma_k = \frac{1}{20\sqrt{2}}\frac{1}{k^p}, \qquad \text{for } p = 3, \ldots, \Lambda. \qquad (32)$$

Dynamical regimes.

Two dynamical regimes will be studied below, which correspond to different values of p with $p = 1$ and $p = 0.5$. Note that the "dynamical regimes" here should not be confused with the "regime switching". The latter stands for the switching of the model variables between different values or states within a given dynamical regime.

Figure 19 shows the time series of the zonal flow u, its associated PDFs and ACFs as well as the accumulated energy in the small-scale stream functions. Here, the accumulated energy $E[\psi_{1:s}]$ is defined as

$$E[\psi_{1:s}] = \sum_{k=1}^{s} |\psi_k|^2. \tag{33}$$

In the regime with $p = 1$, the trajectory of the zonal velocity u switches between roughly two different states and it stays in each state for a while before switching to the other (see Panel (a)). One state with positive u corresponding to the eastward zonal flow contain extreme events. Despite the nearly "two-state" time series, the associated PDF of u is single modal and is slightly skewed where a single (positive) side fat tail correspond to extreme events for the eastward zonal flow. Note that the regime (state) switching behavior with such a single modal distribution has been systematically studied in[122]. Despite the single modal distribution, the ACF is highly different from a Gaussian model with exponential decay. In fact, the ACF here first experiences a sharp decrease to ACF = 0.5 and then it decays slowly with almost a linear decaying rate to zero. The total decaying time is about 60 time units. On the other hand, regarding the small-scale stream functions ψ_k, the leading two modes contain about 84% of the total energy. The ACF associated with ψ_1 has a strong oscillation with a long memory while that associated with ψ_2 only has a weak oscillation. For modes ψ_k with $k \geq 3$, the ACFs decay quite fast.

Next, in $p = 0.5$ regime, the trajectory of the zonal velocity u has a relatively strong mixing rate. The direction of the zonal velocity alternates between eastward and westward quite frequently. Despite the Gaussian statistics, the dynamical regime is still chaotic. The ACF associated with u now behaves in a very different way, where it oscillates and decays quickly to zero. The leading two modes of the small-scale stream functions ψ_1 and ψ_2 contain about 61% of the total energy, and the ACFs associated with ψ_k with $k \geq 3$ now decay more slowly compared with those in $p = 1$ regime.

Notably, in both regimes, the total flow field alternatives between zonally blocked and unblocked patterns as shown in Panels (j)–(k) and (u)–(v). Recall in (27) the topographic effect is imposed on the layered modes with $\vec{l} = (1, 0)$. This implies that the contributions of all the small-scale stream functions ψ_k are on the meridional flows while the zonal flow is driven by the large-scale zonal mode u. As a consequence, when the total flow field is zonally blocked, the large scale zonal velocity $u = 0$ and the total energy lies in the small-scale stream functions (see Panels (j) and (u)). Similarly, when the zonal flow becomes dominant, its kinetic energy accounts for a large portion of the total energy (see Panels (k) and (v)). Therefore, the regime switching not only alters the flow patterns but also adjusts the energy contributions in the total flow field.

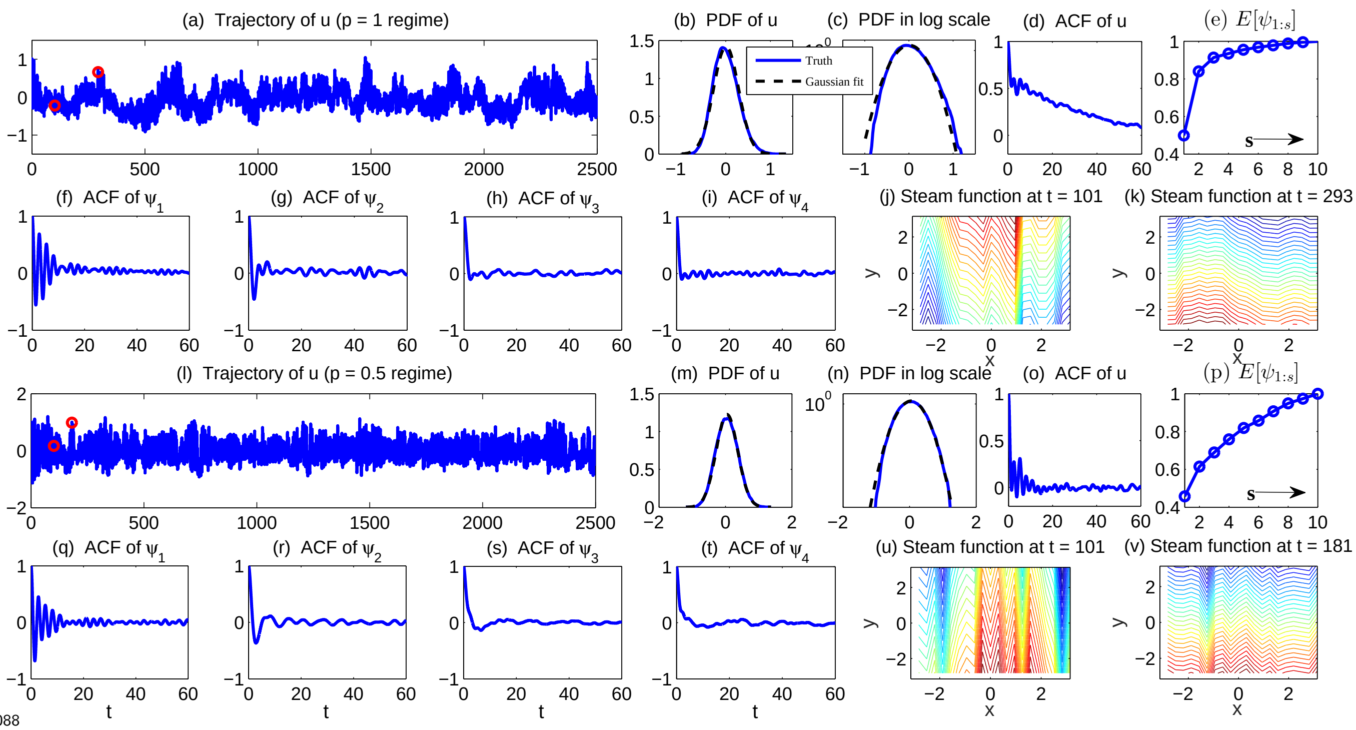

FIG. 19. Dynamical regimes of the layered topographic model (28). Panels (a)–(k): regime with $p = 1$. Panels (a)–(d) show the time series, PDF, PDF in logarithm scale and the ACF of u. Panel (e) shows the accumulated energy $E[\psi_{1:s}]$ defined in (33). Panels (f)–(i) show the ACFs of the first four stream functions. Panels (j)–(k) show the total streamline at two different time instants marked in red dots in Panel (a). At these two time instants, the model shows the blocked and unblocked zonal flow structure, respectively. Panels (l)–(v) are similar to (a)–(k) but for regime with $p = 0.5$.

B. The approximate model

Recall that the perfect model (28) has a 21-degree of freedom. The approximate model developed here has a much simpler form, which includes only the zonal flow u and the leading two Fourier wavenumbers (with $k = \pm 1$ and ± 2).

The motivation of such a choice of the approximate model comes from the rapid decay of the ACFs associated with the small-scale stream functions. In fact, as shown in Figure 19, the stream functions ψ_k with $k = 3, \ldots, 10$ decorrelate very fast while ψ_1 has a much longer relaxation time and ψ_2 also has some memory. Therefore, it is natural to retain the dynamics of the leading two modes and incorporate the effects of the small- and fast-scale modes using extra damping and stochastic forcing in the approximate model. This follows the basic idea of the stochastic mode reduction strategy[53–56], though the manipulation here is less sophisticated. It is also important to notice that the extra stochastic noise added into the approximate model is crucial since the energy in modes ψ_k for $k = 3, \ldots, 10$ as shown in Panels (e) and (p) of Figure 19 is non-negligible. Without these extra stochastic noise, the total variance will be underestimated, which will severely affect the prediction skill of the extreme events in the system.

For the simplicity of notation, we make a change of variables,

$$\psi_1 = \frac{1}{2\sqrt{2}}\Big((v_2 - v_1) - (v_2 + v_1)i\Big), \qquad \text{and} \qquad \psi_2 = \frac{1}{2\sqrt{2}}\Big((v_4 - v_3) - (v_4 + v_3)i\Big). \tag{34}$$

and therefore the 5-mode approximate model is given by,

$$\begin{aligned}
\frac{du}{dt} &= \omega_1 v_1 + 2\omega_3 v_3 - d_u u + \sigma_u \dot{W}_u,\\
\frac{dv_1}{dt} &= -\beta v_2 + v_2 u - 2\omega_1 u - d_{v_1} v_1 + \sigma_1 \dot{W}_1,\\
\frac{dv_2}{dt} &= \beta v_1 - v_1 u - d_{v_2} v_2 + \sigma_2 \dot{W}_2,\\
\frac{dv_3}{dt} &= -\frac{\beta}{2} v_4 + 2 v_4 u - \omega_3 u - d_{v_3} v_3 + \sigma_3 \dot{W}_3,\\
\frac{dv_4}{dt} &= \frac{\beta}{2} v_3 - 2 v_3 u - d_{v_4} v_4 + \sigma_4 \dot{W}_4,
\end{aligned} \tag{35}$$

where $\omega_1 = H_1/\sqrt{2}$ and $\omega_3 = H_2/\sqrt{2}$. In (35), all the variables u, v_1, v_2, v_3 and v_4 are real. The damping and stochastic forcing here are different from the perfect model since they now also include some effects from the smaller scale modes of the perfect model that are ignored here. Notably, the approximate model (35) satisfies the physics constraint, where the total energy in the nonlinear terms is conserved[35,36].

C. Parameter estimation, data assimilation and long-term prediction skill

The system in (35) is a nonlinear system. In practice, the observational data of the leading a few stream functions can be obtained. Therefore, we assume here the observational time series of ψ_1 and ψ_2 are available. As in many real applications of atmosphere and ocean, the observational training data is very limited. Here only the short period as shown in Panel (a) or Panel (l) of Figure 19 is used for model calibration. On the other hand, we assume that there is no observations for the zonal flow u. Recall that u plays an important role in transferring energy with the small-scale stream functions in a nonlinear way and altering the system between zonally blocked and unblocked patterns. Thus, for predicting the extreme events in the system, assimilating the unobserved zonal flow u becomes necessary. Note that despite the intrinsic nonlinearity in the coupled system (35), the system belongs to the conditional Gaussian framework as was discussed in Section II, which allows an efficient way of implementing parameter estimation and data assimilation.

Parameter estimation.

Applying the parameter estimation algorithm described in Section III D, we arrive at the following estimated parameters in the approximate model (35),

$$
\begin{aligned}
\text{Regime } p = 1: \quad & d_u = 0.0132, && d_v = 0.0187, && \sigma_u = 0.0515, && \sigma_v = 0.0501,\\
& \omega_1 = 0.7035, && \omega_3 = 0.3508, && \beta = 1.9954, &&\\
\text{Regime } p = 0.5: \quad & d_u = 0.1417, && d_v = 0.0205, && \sigma_u = 0.1450, && \sigma_v = 0.0504,\\
& \omega_1 = 0.6712, && \omega_3 = 0.3485, && \beta = 1.9963, &&
\end{aligned}
\tag{36}
$$

where we have assumed all the damping coefficients in the v_i equations are the same and all equal to d_v. Similar assumption is used for the stochastic forcing coefficients in the v_i equations which all equal to σ_v.

For the estimated parameters, those with clear physical meanings, for example β, ω_1 and ω_3, are quite close to the truth. The other parameters, mainly the stochastic forcing and damping coefficients, are different from those in the perfect model. Note in particular that the noise coefficients in the approximate model are larger than those in the perfect model. Such a judicious model error with noise inflation compensates the error in the approximate model due to the ignorance of the small-scale stream functions ψ_k from $k = \pm 3$ to ± 10.

Data assimilation.

Using the approximate model (35) as the forecast model for data assimilation of the zonal flow u, the assimilated values are almost the same as the truth (figures not shown here) with a pattern correlation between the truth and the posterior mean states being 0.98 and 0.95 in $p = 1$ and $p = 0.5$ regimes, respectively. In addition, the amplitudes of the assimilated states and the truth are also comparable with each other, implying the success of assimilating the extreme events. These results indicate the skill of using the approximate model in real-time state estimation of the unobserved process and accurately recovering the overall flow structure.

Long-range forecast.

Figure 20 shows the long-term forecast results. Panels (a)–(b) present model trajectories of ψ_1, ψ_2 and u simulated from the perfect model (28) and the approximate model (35) in $p = 1$ regime. These are simply a free run of each model and therefore we do not expect point-to-point correspondence between the two simulations due to the randomness. Nevertheless, these trajectories indicate that the qualitative features from both the models are similar. In particular, the approximate model succeeds in recovering the regime switching behavior in u. In Panels (c)–(d), the ACFs and PDFs associated with both the models are illustrated. The approximate model is quite skillful in capturing the strong oscillation, weak oscillation and the slowly but non-exponential decay in the ACFs associated with ψ_1, ψ_2 and u respectively. The approximate model also succeeds in recovering the PDFs of all the three variables, especially the variance which is important for predicting the extreme events in short- and medium-range, as will be discussed in the next subsection. Similar conclusions can be made in $p = 0.5$ regime. The only slight error lies in tracing the fast decay ACF of u in the approximate model. But the equilibrium PDFs and the ACFs associated with ψ_1 and ψ_2 are recovered with high accuracy.

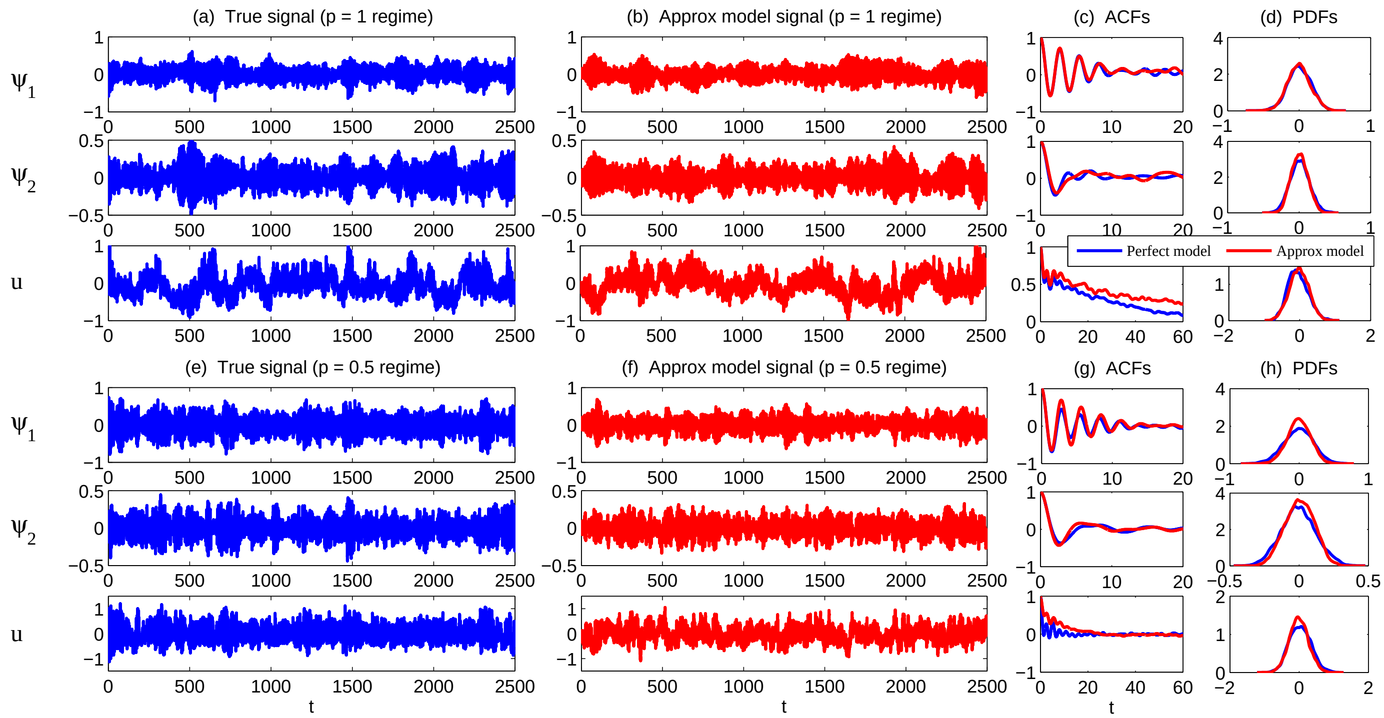

FIG. 20. Long-range forecasts of the layered topographic model. Panels (a)–(d): trajectories of the perfect model, trajectories of the approximate model, ACFs and PDFs in $p = 1$ regime. Panels (e)–(h): similar but for $p = 0.5$ regime.

D. Short- and medium-range forecast

With the approximate model and the assimilated initial conditions of u, the ensemble forecast is applied to study the short- and medium-range forecasts.

Figure 21 shows the RMSE and Corr in the ensemble mean forecast as a function of the lead time. As comparison, the prediction using the perfect model is also included (blue). The approximate model has essentially the same skill as the perfect model in predicting all the three variables ψ_1, ψ_2 and u. The useful prediction based on these path-wise measurements as well as the information criterion for comparing the predicted amplitudes (not shown here) in $p = 1$ regime is about 5 units for all the three variables and that in $p = 0.5$ regime is $3.5, 2.5$ and 1 units for ψ_1, ψ_2 and u, respectively. Figure 22 shows the predicted trajectories at lead time $1, 2$ and 3 units. The prediction of the extreme events up to 3 lead time units in $p = 1$ is quite accurate in terms of both the predicted patterns and the amplitudes. The $p = 0.5$ regime has a shorter range of useful predictions, but the overall skill up to 1 unit for both quiescent and extreme events are significant.

Some case studies are included in Figure 23. In Panels (a)–(c), the ensemble prediction starts from $t = 300, 1390$ and 1460, respectively, and each prediction is run for 30 units forward. Although the overall skillful prediction in $p = 1$ regime as shown in Figure 21 is 5 units, the three events in Panels (a)–(c) of Figure 23 indicate that the useful prediction depends on the initial phase and the follow-up structure of the signal. Despite the intrinsic chaotic behavior, the useful prediction in case study 1 reaches 12 units, where all the extreme events within this time interval are captured accurately by the approximate model. On the other hand, the prediction in case study 2 is completely unskillful due to the fact that u has no internal oscillation structure for this particular event while the long-term trend cannot be captured by the ensemble mean forecast. Case study 3 shows a skillful prediction up to 6 units where again the extreme events within this time interval are captured with high accuracy.

Panels (d)–(e) in Figure 23 compare the predicted stream functions using the approximate model with the truth in the case study 1 from Panel (a) at lead times 1.5 and 6.3, where the truth is generated from the perfect 21-mode model. The true values of the large-scale zonal flow at these two time instants are $u = 0.727$ and $u = -0.03$, respectively. The approximate model is quite skillful in predicting the overall flow patterns. In particular, the predictions succeed in capturing the regime switching phenomenon with a zonally unblocked structure at $t = 301.5$ and a zonally blocked structure at $t = 306.3$. There are some small errors in the prediction. For example, in Panel (d) the true signal at $x = 0.8$ has a sudden increase in the meridional velocity while it is missed in the prediction using the 5-mode approximate model. This meridional velocity is actually triggered by the modes $\psi_k, k = 3, 4, \ldots$, which are not included approximate model. Therefore, even if u, ψ_1 and ψ_2 are predicted almost perfectly, which is the case here, there can be a small intrinsic barrier in recovering the original field because of the simplification of the model by dropping the smaller scale modes.

Similar conclusions are reached for $p = 0.5$ regime, as can be see in Panels (f)–(j) in Figure 23, despite that the useful prediction becomes shorter due to the more intrinsic turbulent behavior in this regime.

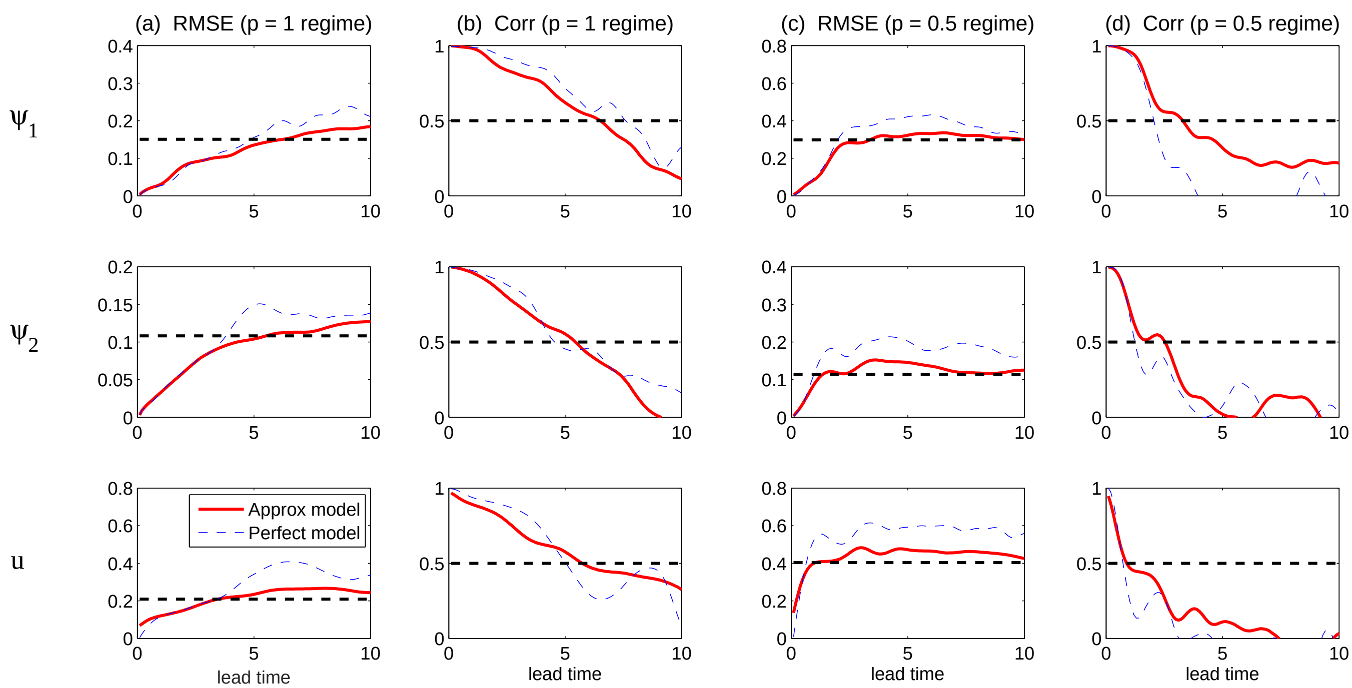

FIG. 21. Short- and medium-range forecasts of the layered topographic model. Panels (a)–(b) show the RMSE and Corr as a function of the lead time in $p = 1$ regime. Panels (c)–(d) show those in $p = 0.5$ regime. The black dashed lines in the RMSE panels show the standard deviation of the true signal and those in the Corr panels show the Corr$= 0.5$ threshold.

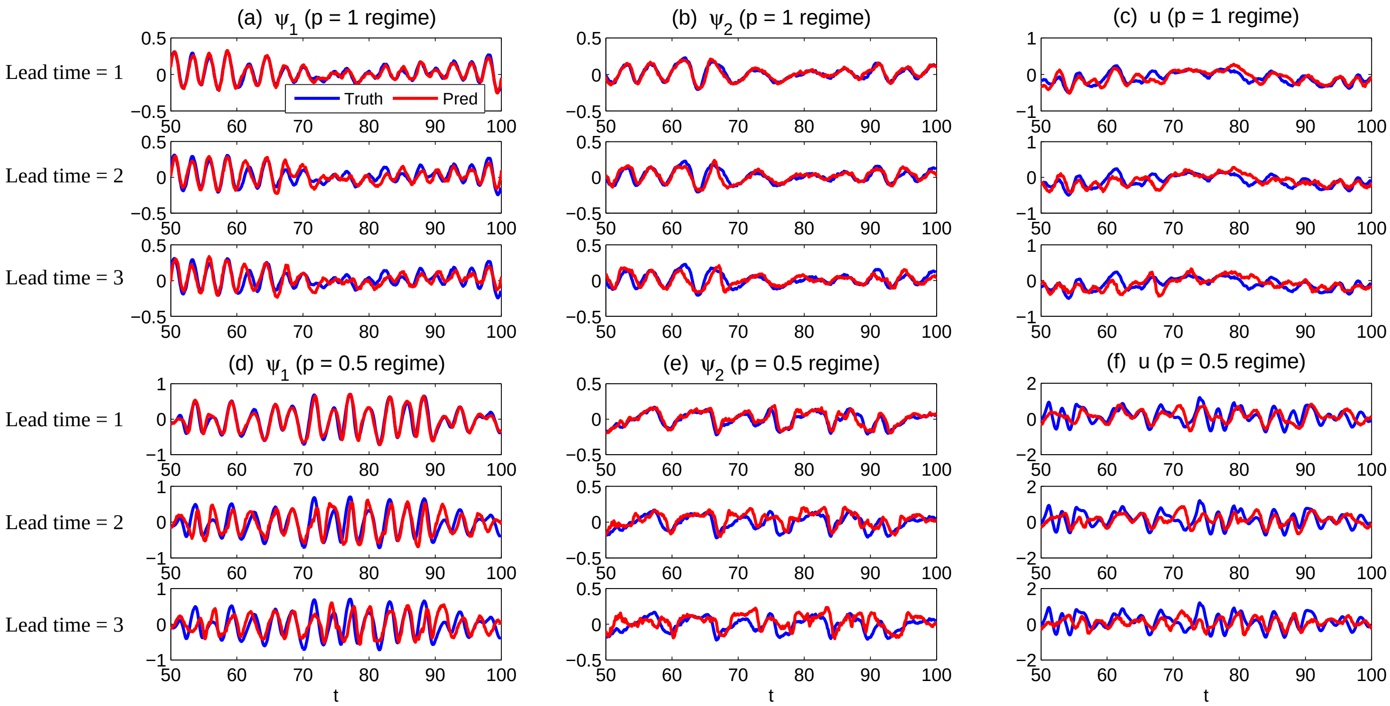

FIG. 22. The truth and the ensemble mean prediction at three different lead times $1, 2$ and 3 using the approximate model with assimilated initial conditions of u. Panels (a)–(c): $p = 1$ regime. Panels (d)–(f): $p = 0.5$ regime.

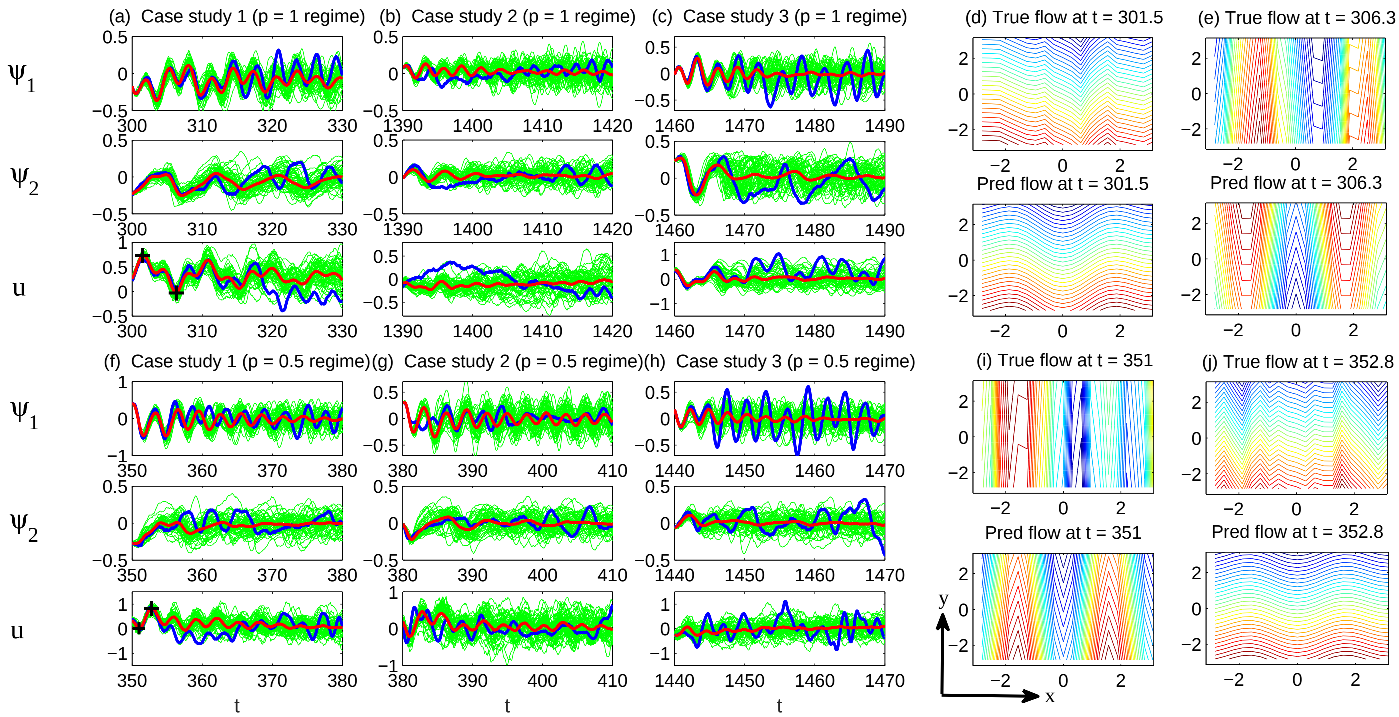

FIG. 23. Case studies of the ensemble prediction. Panels (a)–(e): $p = 1$ regime. Panels (f)–(j): $p = 0.5$ regime. In each subpanel of (a)–(c) and (f)–(h) the blue curve shows the truth and the red one shows the ensemble mean forecasts which is averaged over 50 ensemble members in green. Panels (d)–(e) compare the truth and the predicted overall streamlines in $p = 1$ regime at $t = 301.5$ and $t = 306.3$ (marked in black '+' in Panel (a)), where the starting time is $t = 300$. Panels (i)–(j) compare the truth and the predicted overall streamlines in $p = 0.5$ regime at $t = 351$ and $t = 352.8$ (marked in black '+' in Panel (f)), where the starting time is $t = 350$.

E. Prediction with an initial value starting outside the attractor

Finally, we study the prediction skill of the approximate model, which starts from a value that is outside the attractor. In the three columns of Figure 24, we show the ensemble predictions in $p = 1$ regime by letting the initial value of u, ψ_1 and ψ_2 be outside the attractor, respectively. It is not difficult to tell that the true trajectories starting from values outside the attractor behave in a very different way from those inside the attractor.

When $u(0)$ is outside the attractor (Column (a)), the ensemble mean prediction using the approximate model is accurate up to 3 units. The ensemble spread is very skillful in capturing the envelope of the true signals as time evolves. When $\psi_1(0)$ is outside the attractor (Column (b)). The useful ensemble mean prediction using the approximate model is about 10 units. The extreme events within this 10-unit interval in the zonal velocity are accurately captured. Again, the ensemble spread clearly and accurately indicates the amplitudes in the true signal. When $\psi_2(0)$ is outside the attractor (Column (c)). The skillful ensemble mean prediction using the approximate model extends to 20 units! Note that within the first 10 units, the ensemble spread is very narrow, indicating the high confidence in the ensemble mean prediction, including all the extreme events.

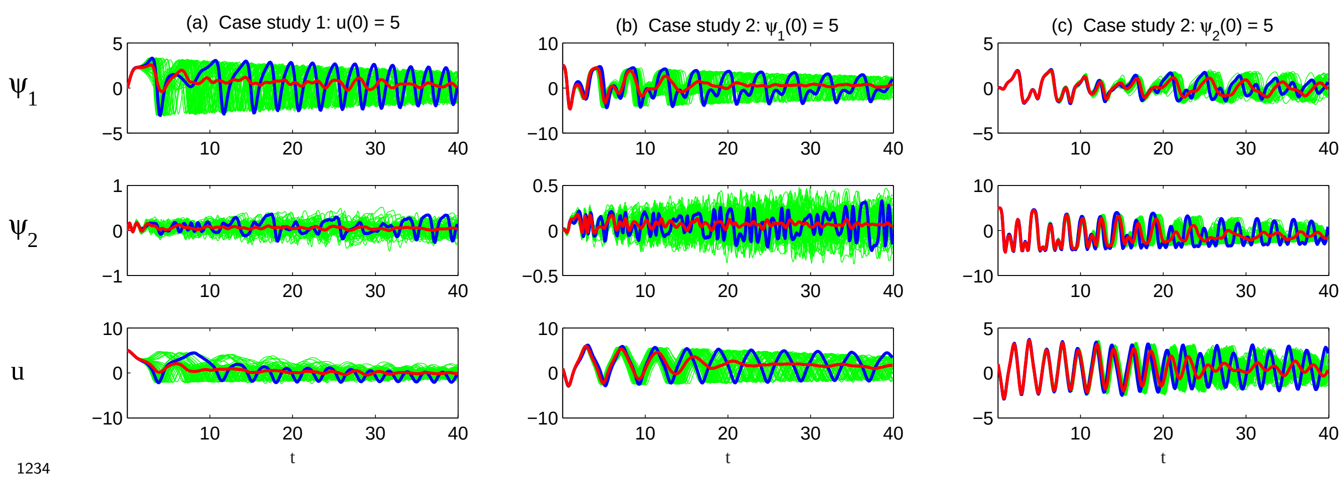

FIG. 24. Prediction using the approximate model starting from a point that is outside the attractor. Column (a) shows that when generating the true signal from the prefect model, the initial value of ψ_1 is outside the attractor. Columns (b) and (c) show that the initial value of ϕ_2 and u are outside the attractor respectively. Again, the blue curve is the true and the red curve is the ensemble mean prediction with the 50 ensemble members shown in green. Here $p = 1$.

VIII. A 6-DIMENSIONAL LOW-ORDER MODEL MIMICKING THE CHARNEY-DEVORE (CDV) MODEL

A. The perfect model and its properties

Charney and DeVore (CDV) made an fundamental contribution for the regime switching behavior of the atmosphere[120]. In this section, a 6-dimensional low-order model that mimics the dynamical behavior of the CDV model is used as the perfect model. Despite the regime switching behavior, this model has distinct mathematical structures and physical mechanisms compared with the one studied in the previous section. It also possesses some unique features, as will be discussed at the end of this subsection, that provide a very tough test for predicting the extreme events and the transition behavior. The goal here is to design suitable and efficient strategies of developing an approximate model that is able to predict the extreme events and other non-Gaussian features in such a model.

This 6-dimensional low-order model is obtained by a Galerkin projection and truncation of the barotropic vorticity equation on a β-plane channel[133,134]. The barotropic vorticity equation is the following,

$$\frac{\partial}{\partial t}\nabla^2\psi = -J\big(\psi, \nabla^2\psi + f + \gamma h\big) - C\nabla^2(\psi - \psi^*), \tag{37}$$

where the domain of longitude and latitude (x, y) are given by $[0, 2\pi]\times[0, \pi b]$. The parameter $b = 2B/L$ determines the ratio between the dimensional zonal length L and the meridional width B of the channel. The stream function ψ is periodic in x. The meridional boundaries $y = 0$ and $y = \pi$ have the conditions $\partial\psi/\partial x = 0$ and $\int_0^{2\pi}(\partial\psi/\partial y)dx = 0$. The Coriolis parameter f generates the beta effect in model. Orography enters with h, the orographic height, and is scaled with γ. J is the Jacobi operator and the damping coefficient C is the newtonian relaxation to the streamfunction profile ψ^*, which represents the forcing associated with the two zonal modes as will be discussed shortly. Next, the barotropic vorticity equation (37) is projected on a set of basis functions which are eigenfunctions of the Laplace operator ∇^2,

$$\phi_{0m}(y) = \sqrt{2}\cos(my/b), \qquad \phi_{nm}(x, y) = \sqrt{2}e^{inx}\sin(my/b),$$

The 6-dimensional model is obtained by truncating the expansion of the stream function and the topographic height after $|n| = 1$ and $m = 2$. Then the time-dependent complex variables of the stream functions $\psi_{01}, \psi_{02}, \psi_{\pm 11}, \psi_{\pm 12}$ are transformed to real variables:

$$
\begin{aligned}
x_1 &= \frac{1}{b}\psi_{01}, \qquad & x_2 &= \frac{1}{b\sqrt{2}}(\psi_{11} + \psi_{-11}), \qquad & x_3 &= \frac{i}{b\sqrt{2}}(\psi_{11} - \psi_{-11}), \\
x_4 &= \frac{1}{b}\psi_{02}, \qquad & x_5 &= \frac{1}{b\sqrt{2}}(\psi_{12} + \psi_{-12}), \qquad & x_6 &= \frac{i}{b\sqrt{2}}(\psi_{12} - \psi_{-12}),
\end{aligned}
$$

while the topography h is chosen to have only the $(1,1)$ wave profile,

$$h(x,y) = \cos(x)\sin(y/b).$$

These manipulations lead to a 6-dimensional ODE model, where x_1, x_4 represent the zonal flow, x_2, x_3 are the topographic Rossby waves and x_5, x_6 are the Rossby waves.

In the study here, extra small noise is added to this model, which allows some effects from the small-scale modes to enter into this low-order model. The noisy version of the 6-dimensional CDV model reads,

$$
\begin{aligned}
dx_1 &= \Big(\gamma_1^* x_3 - C(x_1 - x_1^*)\Big)dt + \sigma_1 dW_1, \\
dx_4 &= \Big(\gamma_2^* x_6 - C(x_4 - x_4^*) + \epsilon(x_2 x_6 - x_3 x_5)\Big)dt + \sigma_4 dW_4, \\
dx_2 &= \Big(-(\alpha_1 x_1 - \beta_1)x_3 - C x_2 - \delta_1 x_4 x_6\Big)dt + \sigma_2 dW_2, \\
dx_3 &= \Big((\alpha_1 x_1 - \beta_1)x_2 - \gamma_1 x_1 - C x_3 + \delta_1 x_4 x_5\Big)dt + \sigma_3 dW_3, \\
dx_5 &= \Big(-(\alpha_2 x_1 - \beta_2)x_6 - C x_5 - \delta_2 x_4 x_3\Big)dt + \sigma_5 dW_5, \\
dx_6 &= \Big((\alpha_2 x_1 - \beta_2)x_5 - \gamma_2 x_4 - C x_6 + \delta_2 x_4 x_2\Big)dt + \sigma_6 dW_6.
\end{aligned}
\tag{38}
$$

Here the terms multiplied by α_i model the advection of the waves by the zonal flow. The β_i terms are due to the Coriolis force; the γ terms are generated by the topography. The C terms are the Newtonian damping to the zonal profile $x^* = (x_1^*, 0, 0, x_4^*, 0, 0)$. The $\delta-$ and $\epsilon-$terms describe the nonlinear triad interaction between the zonal $(0,2)$ mode and the $(1,1)$ and $(1,2)$ waves. This triad is responsible for the possibility of barotropic instability of the $(0,2)$ mode. Note that the model is scaled such that 1 time unit in the model roughly corresponds to 1 day.

Following[133,134], the following parameter values are taken: $C = 0.1$, corresponding to a damping time of 10 days; $\beta = 1.25$, corresponding to a channel centered around a latitude of 45^o; $b = 0.5$, the north-south extent of the channel is 25% of its east-west extent; and $x_1^* = 0.95$ and $x_4^* = -0.76095$. These parameters allow a combination of topographic and barotropic instabilities. The noise coefficients added here are $\sigma_1 = ... = \sigma_6 = 0.005$. Such a choice of the noise coefficients allow the dynamical behavior of this stochastic model to remain similar to its deterministic version as was studied in[133,134].

Note that x_1 and x_4 associated with ψ_{01} and ψ_{02} describe the zonal flows, and the forcing x_1^* and x_4^* are only imposed on these modes. Therefore, it is natural to assume x_1 and x_4 are the observed variables while x_2, x_3, x_5 and x_6 are unobserved. The goal is to predict the extreme events in the system given only short trajectories of x_1 and x_4.

Model properties.

Panels (a)–(c) of Figure 25 show the chaotic trajectories, non-Gaussian PDFs and the ACFs of the model (38). It is easy to tell from the model trajectories that this model has multiple equilibria, which is confirmed by the phase plot (x_1, x_4) in Panel (d) (two stable equilibria; top left and bottom right). The spatial patterns associated with these two equilibria are quite different with each other, as shown in Panels (e) and (f) for two sample events corresponding to the time instants marked in red in Panel (a) that lie near these two equilibria. The streamlines shown in Panel (e) corresponds to an equilibria with largely zonal character with strong zonal jets while that in Panel (f) is dominated by topographically effects with vortices and meander jets. When the blocking events happen, x_1 reaches its maximum while x_4 lies in its minimum value.

Panel (b) shows the equilibrium PDFs of all the 6 model variables. The profiles of these PDFs are quite different: x_1 and x_3 have weakly bimodal distributions; x_2, x_4 and x_6 are highly skewed with an one-sided fat tail towards the negative side; and x_5 is skewed with a fat tail towards the positive side. Panel (c) illustrates the ACFs, which imply multiple decorrelation time scales of the system, where x_1, x_3 and x_4 has a much longer memory than x_2, x_5 and x_6.

One very interesting and important feature of this model (or more precisely its deterministic version) is that projecting this 6-dimensional model to its leading 5 Empirical Orthogonal Functions (EOFs) explains 99.5% of the variance. However, such a 5-dimensional projected dynamics completely misses the dynamical features in the original model, where the multiple equilibria disappears and the 5-dimensional model cannot reproduce regime transitions[134]. Therefore, this 6-dimensional model provides a very useful and tough testbed for developing suitable approximate models in predicting the transition behavior and extreme events in highly chaotic systems.

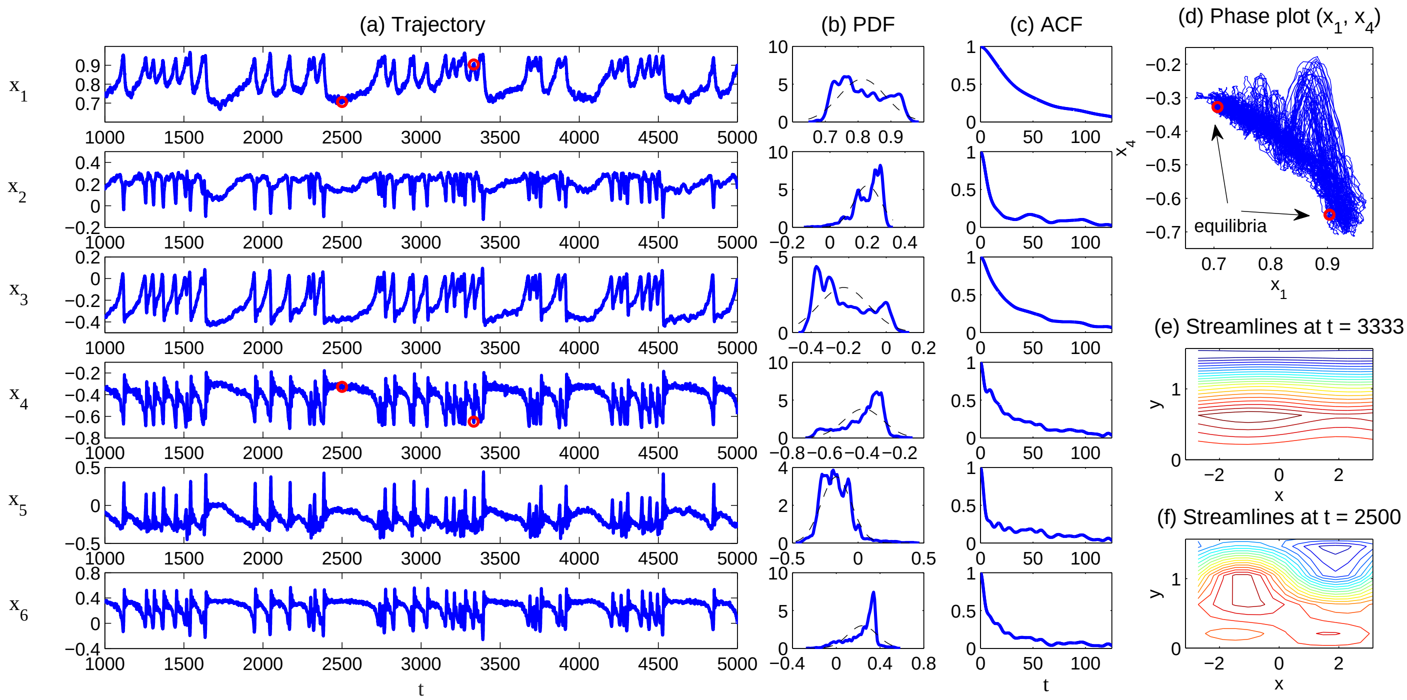

FIG. 25. Panels (a)–(c): Model trajectories, PDFs and ACFs of the 6-D CDV model (38). The black dashed lines in column (b) show the Gaussian fits of the PDFs. Panel (d): Phase plot of (x_1, x_4). Panels (e)–(f): Streamlines at $t = 3333$ and $t = 2500$, corresponding to the time instants marked in red dots in Panel (a).

B. The approximate model

Our goal here is to develop a suitable approximate model for describing and predicting the key features of the 6-dimensional low-order CDV model (38). Recall that the conditional Gaussian nonlinear models in Section II allow an efficient and accurate data assimilation algorithm, which facilitates effective predictions. Therefore, it is natural to develop a suitable approximate model that belongs to the conditional Gaussian nonlinear modeling framework. Note that by observing x_1 and x_4, the 6-dimensional CDV model (38) is not a conditional Gaussian model due to the nonlinear coupling term $\epsilon(x_2x_6 - x_3x_5)$. In fact, the topographic Rossby waves x_2, x_3 and the Rossby waves x_5, x_6 interact with each other through the above nonlinear coupled term. Only in the absence of the Rossby waves x_5, x_6, the coupled system $x_1, \ldots, x_4$ is a conditional Gaussian system. Therefore, suitable strategies need to be developed to cope with this nonlinear term in the approximate model.

Strategy 1: A bare truncation model.

The simplest way to deal with this nonlinear term is to build a bare truncation model, where the nonlinear term $\epsilon(x_2x_6 - x_3x_5)$ is completely dropped. However, this bare truncation model suffers from finite time blowup issue. In fact, the blowup occurs very quickly and even for a very short lead time (much shorter than the decorrelation time), the predicted values have a large chance to go to infinity.

Strategy 2: A nonlinear approximate model with linear feedback terms.

Another straightforward idea is to replace the quadratic term $\epsilon(x_2x_6 - x_3x_5)$ by a combination of four linear terms $c_1x_2 + c_2x_6 - c_3x_3 - c_4x_5$. This approximation is better than the bare truncation model in the sense that the system will not blow up in a very short term. However, the predicted trajectories from this model still have a high probability to blow up in a finite time. In addition, the skillful prediction only lasts for very short time even if the predicted amplitude remains finite within that range.

Strategy 3: An approximate model with a stochastic forcing term.

Instead of using a deterministic and linear way to parameterize the nonlinear quadratic term, a new strategy is developed here, which involves using a simple stochastic forcing process b_1 to describe the effect of the quadratic term $\epsilon(x_2x_6 - x_3x_5)$. The approximate model reads,

$$
\begin{aligned}
dx_1 &= \Big(\gamma_1^* x_3 - C(x_1 - x_1^*)\Big)dt + \sigma_1 dW_1,\\
dx_4 &= \Big(\gamma_2^* x_6 - C(x_4 - x_4^*)\underline{+\, b_1}\Big)dt + \sigma_4 dW_4,\\
dx_2 &= \Big(-(\alpha_1 x_1 - \beta_1)x_3 - Cx_2 - \delta_1 x_4 x_6\Big)dt + \sigma_2 dW_2,\\
dx_3 &= \Big((\alpha_1 x_1 - \beta_1)x_2 - \gamma_1 x_1 - Cx_3 + \delta_1 x_4 x_5\Big)dt + \sigma_3 dW_3,\\
dx_5 &= \Big(-(\alpha_2 x_1 - \beta_2)x_6 - Cx_5 - \delta_2 x_4 x_3\Big)dt + \sigma_5 dW_5,\\
dx_6 &= \Big((\alpha_2 x_1 - \beta_2)x_5 - \gamma_2 x_4 - Cx_6 + \delta_2 x_4 x_2\Big)dt + \sigma_6 dW_6,\\
db_1 &= \Big(-d_b b_1 + \sigma_b b_2 + f_b\Big)dt + \sigma_b dW_b,\\
db_2 &= \Big(-d_b b_2 - \sigma_b b_1 + f_b\Big)dt + \sigma_b dW_b.
\end{aligned}
\tag{39}
$$

This is motivated by the SPEKF model[48,49], where a stochastic forcing is able to automatically learn the missing information on the fly via online data assimilation. Here, we adopt the simplest possible choice — a stochastic forcing b_1 driven by a Gaussian process. Note that two new processes b_1 and b_2 are actually incorporated into the approximate model. They together form a linear stochastic oscillator while only b_1 gives feedback to the x_4 process. The reason to impose an "oscillated" forcing is that all the variables x_i have chaotic oscillator structures and so does the nonlinear term $\epsilon(x_2x_6 - x_3x_5)$. As will be seen below, with this cheap stochastic strategy, the approximate model is able to avoid finite time blowup issue and it provides surprisingly skillful predictions in both short and medium ranges. Notably, treating b_1 and b_2 as the extra unobserved variables, the resulting 8-dimensional nonlinear system in (39) is a conditional Gaussian nonlinear model, where only x_1 and x_2 have observations. The estimated parameters are given by:

$$
F_b = 0.0081, \qquad \omega_b = 0.6815, \qquad d_b = 0.1339, \qquad \sigma_b = 0.01326. \tag{40}
$$

Below the focus will be on the data assimilation and prediction skill using the approximate model in (39).

C. Data assimilation

Since the approximate model with the stochastic forcing (39) is a conditional Gaussian system, the data assimilation algorithm (3) provides an efficient state estimation of both the unobserved variables x_2, x_3, x_5, x_6 and the stochastic forcing b_1, b_2, which are shown in Figure 26 (in red color). As comparison, we also show the truth of the unobserved variables x_2, x_3, x_5, x_6 (in blue color). It is clear that the data assimilation with the help of such a stochastic forcing term provides very accurate estimation of the hidden variables x_2, x_3, x_5, x_6, where the pattern correlation of the assimilated and the true signals is higher than 0.95 for all the variables.

Another striking result is presented in Panels (c) and (d) of Figure 26, where a comparison between the assimilated state of the stochastic forcing b_1 and the nonlinear term $\epsilon(x_2x_6 - x_3x_5)$ computed from the perfect model is illustrated. It is clear that the stochastic forcing b_1 almost perfectly recovers the nonlinear feedback, especially at the time instants that the nonlinear feedback is intermittent. This is a very important feature because it guarantees that the stochastic forcing is able to, at least for a short term, play the role of the nonlinear feedback term in prediction.

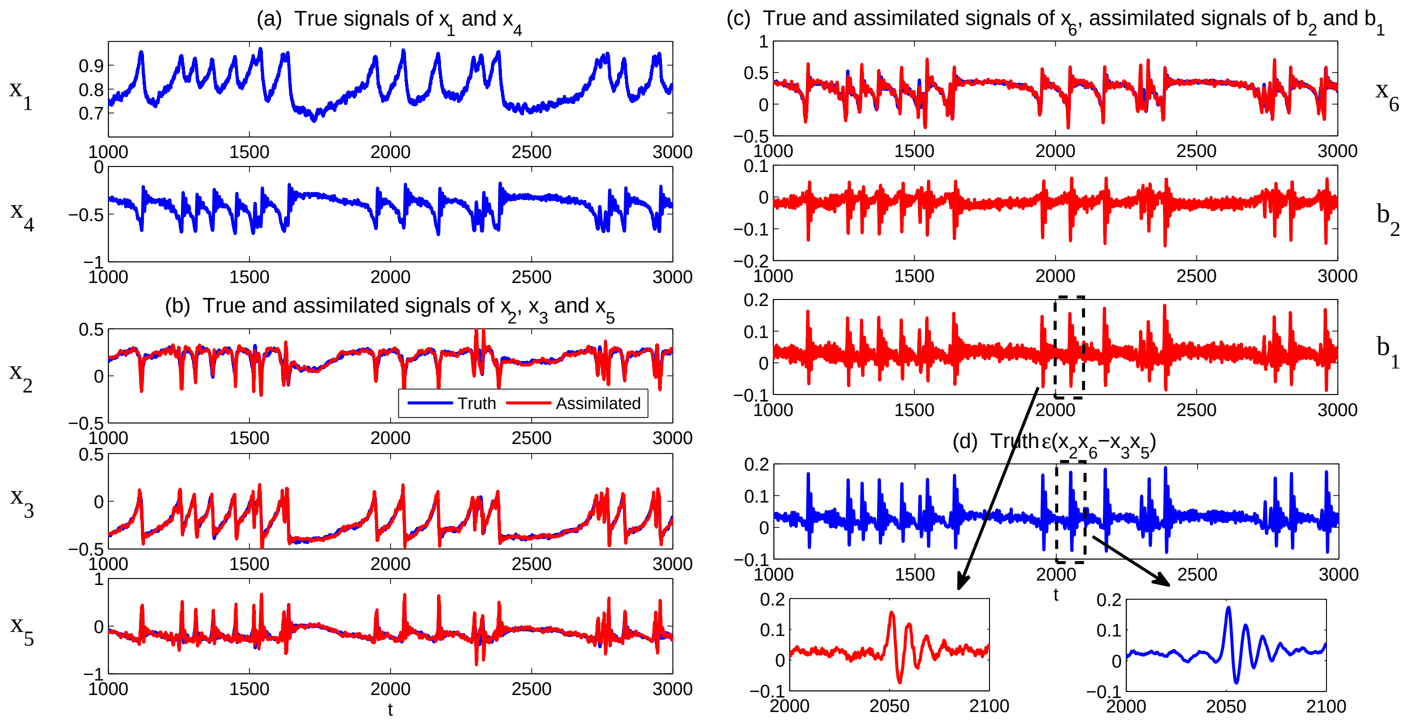

FIG. 26. Data assimilation of the 6-D CDV model using the approximate model (39). Panel (a): the true signals of x_1 and x_4. Panels (b)–(c) the true and the assimilated posterior mean of x_2, x_3, x_5 and x_6, and the assimilated stochastic forcing b_2 and b_1. Panel (d): the true value of the nonlinear term $\epsilon(x_2x_6 - x_3x_5)$. A zoomed-in period of b_1 and the nonlinear term is also shown for comparison.

D. Predictions

Short- and medium-range forecasts.

Our focus now is on the short- and medium-range forecasts. In Figure 27, the prediction skill in terms of the RMSE and Corr as a function of lead time is presented. The blue curves show the predictions using the perfect model with the perfect initial conditions; the red curves show those using the approximate model (39) with the perfect initial conditions; and the green curves show those using the approximate model (39) and the assimilated initial conditions. The ensemble mean is used here for computing the RMSE and Corr.

Despite that the prediction using the approximate model (39) is less skillful than that using the perfect model as the increase of lead time, it is clear that the useful prediction for all the variables using the approximate model (39) is still at least 8 units. For some variables such as x_3 the useful prediction is 16 days and for x_1 it is much longer. In addition, the approximate model (39) using the assimilated initial conditions has nearly the same prediction skill as that using the perfect initial conditions, which verifies the accuracy in the assimilated states. These results imply that the approximate model is a suitable model for both short- and medium-range forecasts of such an extremely tough test model.

Figures 28–29 include two case studies of the prediction tests. The ensemble mean prediction shown in Figure 28 is extremely accurate for both short and medium ranges. On the other hand, although the ensemble mean prediction in Figure 29 has a slight phase shift, which results in the deterioration of the pattern correlation, the overall prediction using the approximate model remains skillful. From these figures, it is clear that most of the extreme events take around 8 units to develop from the onset phase to the peak, which is within the skillful prediction range of the approximate model (39). Therefore, the approximate model is able to predict the entire development phase of the extreme events. On the other hand, starting from the peak of an extreme event, the approximate model succeeds in predicting the returning path to the quiescent state. In addition to the ensemble mean, the ensemble envelope also plays an important role in the prediction here. In fact, despite a slight phase shift in the ensemble mean prediction in Figure 29, the ensemble envelope clearly predicts the correct overall time evolution trends of the truth. Admittedly, the ensemble spread using the approximate model with the assimilated initial condition is larger than the prediction with the perfect initial condition, which is mainly due to the initial uncertainty in assimilating the hidden variables and the uncertainty introduced from the stochastic forcing. Nevertheless, the correct trends are still unambiguously predicted by the ensemble members in both short and medium terms.

Long range forecast.

The approximate model (39) fails to reproduce the same long-term equilibrium PDFs as the truth. This is not surprising since the stochastic forcing for the long range forecast loses its memory of the initial condition and essentially becomes a constant. Its contribution to the system is then quite different from the original nonlinear term $\epsilon(x_2x_5 - x_3x_6)$, which evolves in time. Note that the study in the previous work[74] has already illustrated that in the presence of model error it is extremely difficult to develop suitable approximate models that are able to simultaneously have both short and long range forecast skill. Nevertheless, the approximate model (39) is still able to provide some useful information for the long-range forecasts. First, the approximate model (39) avoids finite-time blowup issue and its equilibrium PDFs contain non-Gaussian features, which already outweighs many other approximation strategies, such as bare truncation and linear approximations, for describing strongly chaotic systems. Second, the ACFs of x_1, x_2 and x_3 from the approximate model (39) are quite similar to the truth and the errors in the ACFs of x_4, x_5 and x_6 are also only moderate. These features in the ACFs together with the accurate data assimilation results actually guarantee the skillful short- and medium-range forecasts.

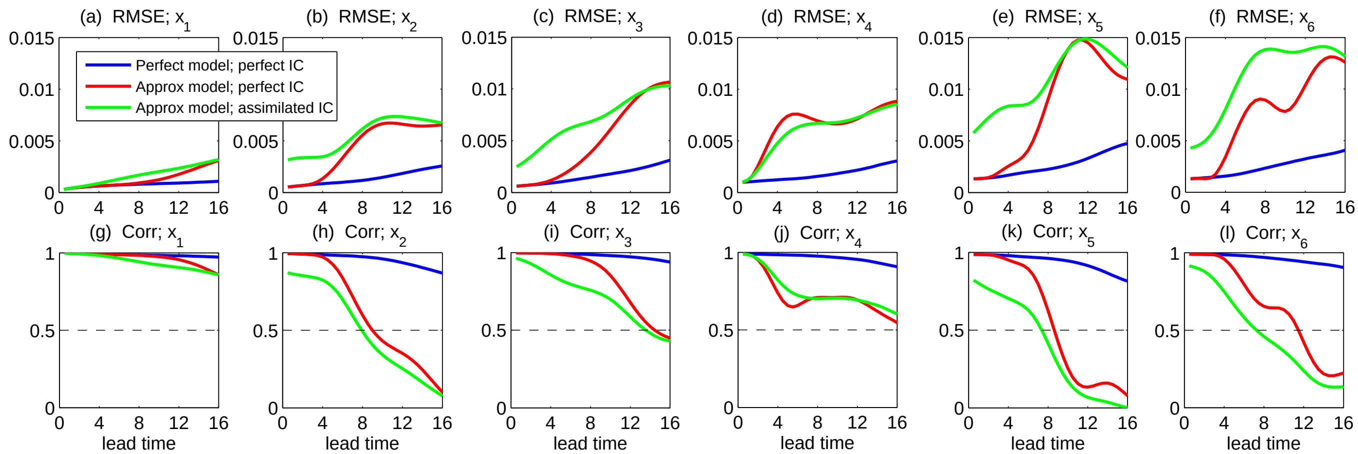

FIG. 27. Short- and medium-range forecasts using the perfect model and the approximate model (39). Top and bottom rows show the RMSE and Corr and a function of lead time. The blue curves show the predictions using the perfect model with the perfect initial conditions; the red curves show those using the approximate model (39) model with the perfect initial conditions; and the green curves show those using the approximate model (39) and the assimilated initial conditions. The ensemble mean is used here for computing the RMSE and Corr with the truth.

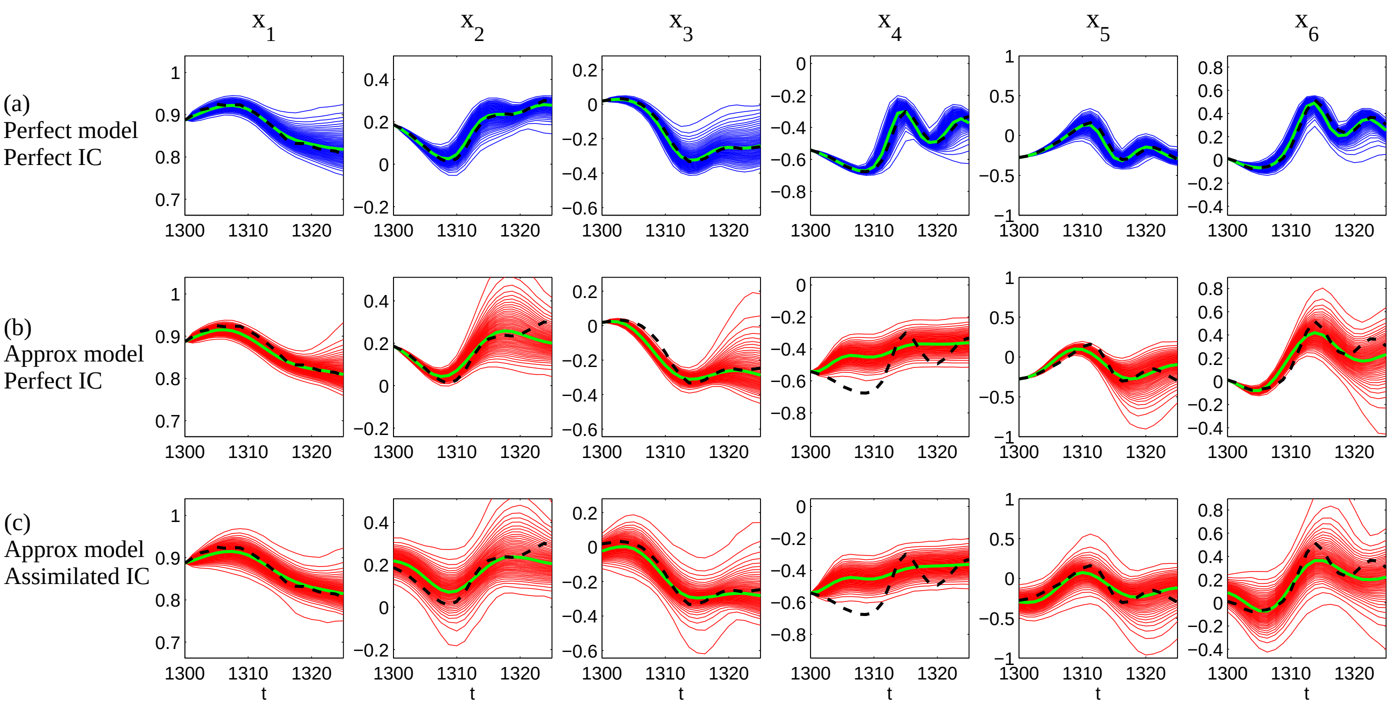

FIG. 28. Case study. Prediction starting from $t = 1300$. Note that each PDF is shown with 50 thin curves (blue for the perfect model and red for the approximate model), which represent the 1st, 3nd, 5th, ..., 97th and 99th percentiles of the of the PDF. The green curve represents the mode of the PDF since the PDF is non-Gaussian. The black dashed curve is the true signal.

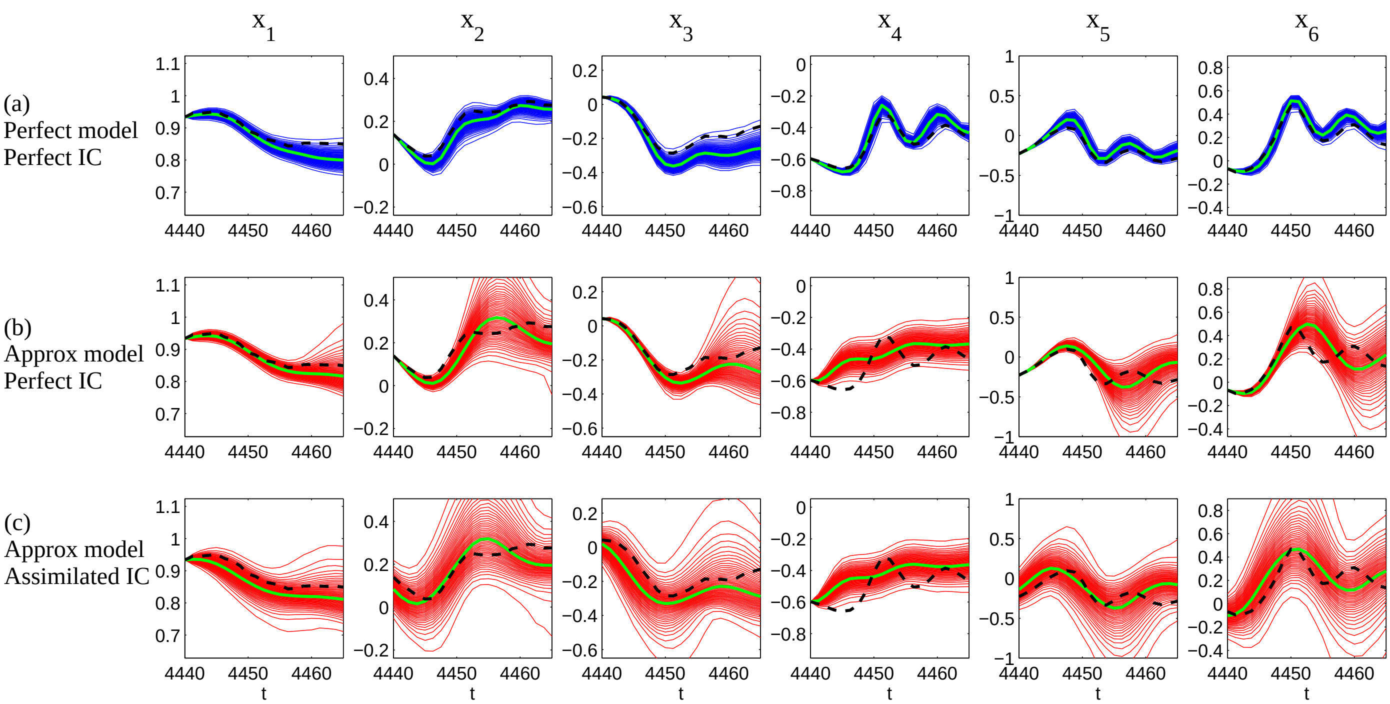

FIG. 29. Case study. Prediction starting from $t = 4440$.

IX. CONCLUSION

Extreme events appear in many complex nonlinear dynamical systems. Predicting extreme events has both scientific significance and practical implications. The main difficulties in predicting the extreme events include the lack of a complete understanding of physics, the unaffordable computational cost of running the complex dynamical systems and the errors in data assimilation or state estimation. Notably, in many practical situations, only partially observed time series are available for model calibration and the training period is often very short. All these facts result in great challenges and lead to the failure of many purely data-driven methods in the extreme events prediction.

In this paper, a new mathematical framework of building suitable nonlinear approximate models is developed, which aims at predicting both the observed and hidden extreme events in complex nonlinear dynamical systems using only short and partially observed training time series. The models belonging to this mathematical framework are highly nonlinear and are able to capture many key non-Gaussian characteristics as observed in nature. Physically motivated processes and physics constraints can be incorporated into the models, which make this framework fundamentally different from many purely data-driven statistical models that have no clear physical meanings. Such a feature also allows using only a short training time series for model calibration. In addition, this modeling framework provides closed analytic formulae for assimilating the states of the unobserved variables, which is computationally efficient and accurate. The details of this modeling framework is shown in Section II. Section III contains the efficient and accurate data assimilation, parameter estimation and prediction algorithms as well as the details of using both the path-wise and information measurements in quantifying the prediction skill. Different effective and practical strategies of developing suitable approximate models for predicting extreme events and other non-Gaussian features in various complex turbulent dynamical systems are illustrated in Section IV to Section VIII.

In Section IV, the skill of applying a cheap stochastic parameterization to approximate the complicated dynamical behavior in the hidden process is explored. This simple and efficient stochastic parameterization is able to recover the nonlinear feedback from the unresolved variable to the observed one. Notably, the nonlinear approximate model with such a cheap stochastic parameterization has nearly the same skill in predicting the extreme events at all short, medium and long ranges. Section V makes use of a nonlinear dyad model to show the success of applying a simple feedback control strategy in the approximate model to facilitate the prediction of the hidden extreme events, which is a great challenge given only partial observations. In Section VI, the Lorenz 63 model is used as a simple test model for predicting extreme events in the intrinsic chaotic models. The goal for testing this model is to understand the model error due to the noise inflation in affecting the extreme events prediction, where the noise inflation is a typical strategy of developing approximate models in many real applications. It is shown that a moderate noise inflation retains the skill of the extreme events prediction at all short, medium and long ranges. Next, regime switching between multiple metastable states is a key feature in many nonlinear turbulent dynamical systems. Section VII starts with a 21-dimensional nonlinear topographic mean flow interaction model with regime switching. A simplified version of the stochastic mode reduction strategy is applied in a suitable way to develop an approximate physics-constrained nonlinear model with only 5 dimensions. The 5-dimensional physics-constrained nonlinear model has a significant skill in predicting both the observed and hidden extreme events as well as other non-Gaussian features, nearly the same as the perfect model prediction. It also succeeds in predicting the regime switching between the zonally blocked and unblocked patterns with high accuracy. In Section VIII, a 6-dimensional low-order Charney-DeVore (CDV) model is used as a testbed for predicting extreme events. This model is highly nonlinear and has strong chaotic features. The leading 5 EOFs contain more than 99.5% of the explained variance but they completely miss the nonlinear dynamical features and the regime switching behavior. Therefore, this 6-dimensional model is an extremely tough test model for predicting the intrinsic nonlinear transitions and extreme events. It is shown that a simple but judicious linear stochastic process with additive noise and memory has a significant skill in learning certain complicated nonlinear effects of this model on the fly. The resulting approximate nonlinear model by incorporating such a simple stochastic process allows efficient and accurate data assimilation. It succeeds in predicting both the observed and hidden extreme events in short and medium terms.

ACKNOWLEDGMENTS

The research of N.C. is supported by the Office of Vice Chancellor for Research and Graduate Education (VCRGE) at University of Wisconsin-Madison. The research of A.J.M. is partially supported by the Office of Naval Research Grant ONR MURI N00014-16-1-2161 and the Center for Prototype Climate Modeling (CPCM) at New York University Abu Dhabi Research Institute. The research of both N.C. and A.J.M is supported by ONR MURI N00014-19-1-2421.

REFERENCES

[1] A. J. Majda, *Introduction to turbulent dynamical systems in complex systems* (Springer, 2016).

[2] A. Majda and X. Wang, *Nonlinear dynamics and statistical theories for basic geophysical flows* (Cambridge University Press, 2006).

[3] S. H. Strogatz, *Nonlinear Dynamics and Chaos: with Applications to Physics, Biology, Chemistry, and Engineering* (CRC Press, 2018).

[4] D. Baleanu, J. A. T. Machado, and A. C. Luo, *Fractional dynamics and control* (Springer Science & Business Media, 2011).

[5] T. Deisboeck and J. Y. Kresh, *Complex systems science in biomedicine* (Springer Science & Business Media, 2007).

[6] J. Stelling, A. Kremling, M. Ginkel, K. Bettenbrock, and E. Gilles, *Foundations of Systems Biology* (MIT press, 2001).

[7] S. A. Sheard and A. Mostashari, "Principles of complex systems for systems engineering," Systems Engineering **12**, 295–311 (2009).

[8] D. C. Wilcox, "Multiscale model for turbulent flows," AIAA journal **26**, 1311–1320 (1988).

[9] E. Pelinovsky, C. Kharif, *et al.*, *Extreme ocean waves* (Springer, 2008).

[10] W. K.-M. Lau and D. E. Waliser, *Intraseasonal variability in the atmosphere-ocean climate system* (Springer Science & Business Media, 2011).

[11] A. J. Clarke, *An introduction to the dynamics of El Niño and the Southern Oscillation* (Elsevier, 2008).

[12] A. J. Majda, M. Moore, and D. Qi, "Statistical dynamical model to predict extreme events

and anomalous features in shallow water waves with abrupt depth change," Proceedings of the National Academy of Sciences **116**, 3982–3987 (2019).

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

[14]A. J. Majda and B. Gershgorin, "Elementary models for turbulent diffusion with complex physical features: eddy diffusivity, spectrum and intermittency," Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **371**, 20120184 (2013).

[15]A. J. Majda and X. T. Tong, "Intermittency in turbulent diffusion models with a mean gradient," Nonlinearity **28**, 4171 (2015).

[16]E. G. Altmann and H. Kantz, "Recurrence time analysis, long-term correlations, and extreme events," Physical Review E **71**, 056106 (2005).

[17]W. L. Oberkampf, S. M. DeLand, B. M. Rutherford, K. V. Diegert, and K. F. Alvin, "Error and uncertainty in modeling and simulation," Reliability Engineering & System Safety **75**, 333–357 (2002).

[18]M. Collins, B. B. Booth, B. Bhaskaran, G. R. Harris, J. M. Murphy, D. M. Sexton, and M. J. Webb, "Climate model errors, feedbacks and forcings: a comparison of perturbed physics and multi-model ensembles," Climate Dynamics **36**, 1737–1766 (2011).

[19]T. Palmer and J. Räisänen, "Quantifying the risk of extreme seasonal precipitation events in a changing climate," Nature **415**, 512 (2002).

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

[21]E. Kalnay, *Atmospheric Modeling, Data Assimilation and Predictability* (Cambridge university press, 2003).

[22]A. Majda and N. Chen, "Model error, information barriers, state estimation and prediction in complex multiscale systems," Entropy **20**, 644 (2018).

[23]I. M. Navon, "Data assimilation for numerical weather prediction: a review," in *Data assimilation for atmospheric, oceanic and hydrologic applications* (Springer, 2009) pp. 21–65.

[24]X. Zou, I. Navon, and F. Le Dimet, "Incomplete observations and control of gravity waves in variational data assimilation," Tellus A: Dynamic Meteorology and Oceanography **44**,

273–296 (1992).

[25]W. Lahoz, B. Khattatov, and R. Ménard, "Data assimilation and information," in *Data Assimilation* (Springer, 2010) pp. 3–12.

[26]A. J. Majda and J. Harlim, *Filtering Complex Turbulent Systems* (Cambridge University Press, 2012).

[27]G. Evensen, *Data Assimilation: the Ensemble Kalman Filter* (Springer Science & Business Media, 2009).

[28]K. Law, A. Stuart, and K. Zygalakis, *Data Assimilation: a Mathematical Introduction*, Vol. 62 (Springer, 2015).

[29]D. Qi and A. J. Majda, "Predicting extreme events for passive scalar turbulence in two-layer baroclinic flows through reduced-order stochastic models," Commun Math Sci **16**, 17–51 (2018).

[30]D. Qi and A. J. Majda, "Predicting fat-tailed intermittent probability distributions in passive scalar turbulence with imperfect models through empirical information theory," Communications in Mathematical Sciences **14**, 1687–1722 (2016).

[31]A. J. Majda and D. Qi, "Strategies for reduced-order models for predicting the statistical responses and uncertainty quantification in complex turbulent dynamical systems," SIAM Review **60**, 491–549 (2018).

[32]M. Farazmand and T. Sapsis, "Extreme events: Mechanisms and prediction," Applied Mechanics Reviews (2018).

[33]T. P. Sapsis, "New perspectives for the prediction and statistical quantification of extreme events in high-dimensional dynamical systems," Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **376**, 20170133 (2018).

[34]N. Chen and A. Majda, "Conditional Gaussian systems for multiscale nonlinear stochastic systems: Prediction, state estimation and uncertainty quantification," Entropy **20**, 509 (2018).

[35]A. J. Majda and J. Harlim, "Physics constrained nonlinear regression models for time series," Nonlinearity **26**, 201 (2012).

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

[37]R. S. Liptser and A. N. Shiryaev, "Statistics of random processes II: Applications," Appl.

Math **6** (2001).

[38]N. Chen and A. J. Majda, "Filtering nonlinear turbulent dynamical systems through conditional Gaussian statistics," Monthly Weather Review **144**, 4885–4917 (2016).

[39]D. Giannakis and A. J. Majda, "Quantifying the predictive skill in long-range forecasting. Part II: Model error in coarse-grained Markov models with application to ocean-circulation regimes," Journal of Climate **25**, 1814–1826 (2012).

[40]J. S. Armstrong, *Long-range forecasting* (Wiley New York ETC., 1985).

[41]H. H. Hendon, B. Liebmann, M. Newman, J. D. Glick, and J. Schemm, "Medium-range forecast errors associated with active episodes of the Madden–Julian oscillation," Monthly Weather Review **128**, 69–86 (2000).

[42]P. R. Vlachas, W. Byeon, Z. Y. Wan, T. P. Sapsis, and P. Koumoutsakos, "Data-driven forecasting of high-dimensional chaotic systems with long short-term memory networks," Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **474**, 20170844 (2018).

[43]Z. Y. Wan and T. P. Sapsis, "Reduced-space Gaussian process regression for data-driven probabilistic forecast of chaotic dynamical systems," Physica D: Nonlinear Phenomena **345**, 40–55 (2017).

[44]S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural computation **9**, 1735–1780 (1997).

[45]K. P. Murphy, *Machine learning: a probabilistic perspective* (MIT press, 2012).

[46]S. Haykin, *Neural networks: a comprehensive foundation* (Prentice Hall PTR, 1994).

[47]B. Scholkopf and A. J. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond* (MIT press, 2001).

[48]B. Gershgorin, J. Harlim, and A. J. Majda, "Improving filtering and prediction of spatially extended turbulent systems with model errors through stochastic parameter estimation," Journal of Computational Physics **229**, 32–57 (2010).

[49]B. Gershgorin, J. Harlim, and A. J. Majda, "Test models for improving filtering with model errors through stochastic parameter estimation," Journal of Computational Physics **229**, 1–31 (2010).

[50]M. Branicki and A. J. Majda, "Dynamic stochastic superresolution of sparsely observed turbulent systems," Journal of Computational Physics **241**, 333–363 (2013).

[51]S. R. Keating, A. J. Majda, and K. S. Smith, "New methods for estimating ocean

eddy heat transport using satellite altimetry," Monthly Weather Review **140**, 1703–1722 (2012).

[52]A. J. Majda and I. Grooms, "New perspectives on superparameterization for geophysical turbulence," Journal of Computational Physics **271**, 60–77 (2014).

[53]A. J. Majda, I. Timofeyev, and E. V. Eijnden, "Models for stochastic climate prediction," Proceedings of the National Academy of Sciences **96**, 14687–14691 (1999).

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

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

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

[57]A. J. Majda and P. Embid, "Averaging over fast gravity waves for geophysical flows with unbalanced initial data," Theoretical and computational fluid dynamics **11**, 155–169 (1998).

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

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

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

[61]N. Chen and A. J. Majda, "Filtering the stochastic skeleton model for the Madden–Julian oscillation," Monthly Weather Review **144**, 501–527 (2016).

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

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

[64]N. Chen and A. J. Majda, "Model error in filtering random compressible flows utilizing

noisy Lagrangian tracers," Monthly Weather Review **144**, 4037–4061 (2016).

[65] A. J. Majda, D. Qi, and T. P. Sapsis, "Blended particle filters for large-dimensional chaotic dynamical systems," Proceedings of the National Academy of Sciences , 201405675 (2014).

[66] N. Chen and A. J. Majda, "Efficient statistically accurate algorithms for the Fokker–Planck equation in large dimensions," Journal of Computational Physics **354**, 242–268 (2018).

[67] N. Chen, A. J. Majda, and X. T. Tong, "Rigorous analysis for efficient statistically accurate algorithms for solving Fokker–Planck equations in large dimensions," SIAM/ASA Journal on Uncertainty Quantification **6**, 1198–1223 (2018).

[68] N. Chen and A. J. Majda, "Beating the curse of dimension with accurate statistics for the Fokker–Planck equation in complex turbulent systems," Proceedings of the National Academy of Sciences **114**, 12864–12869 (2017).

[69] R. E. Kalman, "A new approach to linear filtering and prediction problems," Journal of basic Engineering **82**, 35–45 (1960).

[70] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," Journal of basic engineering **83**, 95–108 (1961).

[71] K. Brammer and G. Siffling, *Kalman-Bucy Filters* (Artech House on Demand, 1989).

[72] E. N. Lorenz, "Deterministic nonperiodic flow," Journal of the atmospheric sciences **20**, 130–141 (1963).

[73] E. N. Lorenz, "Predictability: A problem partly solved," in *Proc. Seminar on predictability*, Vol. 1 (1996).

[74] A. J. Majda and M. Branicki, "Lessons in uncertainty quantification for turbulent dynamical systems," Discrete & Continuous Dynamical Systems-A **32**, 3133–3221 (2012).

[75] A. J. Majda and B. Gershgorin, "Improving model fidelity and sensitivity for complex systems through empirical information theory," Proceedings of the National Academy of Sciences **108**, 10044–10049 (2011).

[76] J. von Neumann, "Some remarks on the problem of forecasting climatic fluctuations," in *Dynamics of climate* (Elsevier, 1960) pp. 9–11.

[77] M. W. Moncrieff, M. A. Shapiro, J. M. Slingo, and F. Molteni, "Collaborative research at the intersection of weather and climate," Bulletin of the World Meteorological Organization **56**, 204–211 (2007).

[78] R. A. Madden and P. R. Julian, "Detection of a 40–50 day oscillation in the zonal wind in the tropical pacific," Journal of the atmospheric sciences **28**, 702–708 (1971).

[79] R. A. Madden and P. R. Julian, "Description of global-scale circulation cells in the tropics with a 40–50 day period," Journal of the atmospheric sciences **29**, 1109–1123 (1972).

[80] A. J. Majda, S. N. Stechmann, S. Chen, R. H. Ogrosky, and S. Thual, *Tropical Intraseasonal Variability and the Stochastic Skeleton Method* (Springer, 2019).

[81] N. Chen and A. J. Majda, "A new efficient parameter estimation algorithm for high-dimensional complex nonlinear turbulent dynamical systems with partial observations," Journal of Computational Physics (2019), accepted.

[82] M. A. Tanner and W. H. Wong, "The calculation of posterior distributions by data augmentation," Journal of the American statistical Association **82**, 528–540 (1987).

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

[84] G. C. Wei and M. A. Tanner, "A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms," Journal of the American statistical Association **85**, 699–704 (1990).

[85] O. Papaspiliopoulos, G. O. Roberts, and O. Stramer, "Data augmentation for diffusions," Journal of Computational and Graphical Statistics **22**, 665–688 (2013).

[86] M. Vihola, "Robust adaptive Metropolis algorithm with coerced acceptance rate," Statistics and Computing **22**, 997–1008 (2012).

[87] K. E. Taylor, "Summarizing multiple aspects of model performance in a single diagram," Journal of Geophysical Research: Atmospheres **106**, 7183–7192 (2001).

[88] P. L. Houtekamer and H. L. Mitchell, "Data assimilation using an ensemble Kalman filter technique," Monthly Weather Review **126**, 796–811 (1998).

[89] P. F. Lermusiaux, "Data assimilation via error subspace statistical estimation. Part I-I: Middle Atlantic Bight shelfbreak front simulations and ESSE validation," Monthly Weather Review **127**, 1408–1432 (1999).

[90] H. H. Hendon, E. Lim, G. Wang, O. Alves, and D. Hudson, "Prospects for predicting two flavors of El Niño," Geophysical Research Letters **36** (2009).

[91] H.-M. Kim, P. J. Webster, and J. A. Curry, "Seasonal prediction skill of ECMWF System 4 and NCEP CFSv2 retrospective forecast for the northern hemisphere winter," Climate

Dynamics **39**, 2957–2973 (2012).

[92] A. J. Majda and D. Qi, "Linear and nonlinear statistical response theories with prototype applications to sensitivity analysis and statistical control of complex turbulent dynamical systems," Chaos: An Interdisciplinary Journal of Nonlinear Science (2019), submitted.

[93] D. Qi and A. J. Majda, "Low-dimensional reduced-order models for statistical response and uncertainty quantification: Two-layer baroclinic turbulence," Journal of the Atmospheric Sciences **73**, 4609–4639 (2016).

[94] A. J. Majda and B. Gershgorin, "Quantifying uncertainty in climate change science through empirical information theory," Proceedings of the National Academy of Sciences **107**, 14958–14963 (2010).

[95] R. Kleeman, "Measuring dynamical prediction utility using relative entropy," Journal of the atmospheric sciences **59**, 2057–2072 (2002).

[96] R. Kleeman, "Information theory and dynamical system predictability," Entropy **13**, 612–649 (2011).

[97] M. Branicki and A. Majda, "Quantifying Bayesian filter performance for turbulent dynamical systems through information theory," Commun. Math. Sci **12**, 901–978 (2014).

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

[99] S. Kullback and R. A. Leibler, "On information and sufficiency," The annals of mathematical statistics **22**, 79–86 (1951).

[100] S. Kullback, "Letter to the editor: The Kullback-Leibler distance," AMERICAN STATISTICIAN (1987).

[101] S. Kullback, "Statistics and information theory," J Wiley Sons, New York (1959).

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

[103] D. Giannakis, A. J. Majda, and I. Horenko, "Information theory, model error, and predictive skill of stochastic models for complex nonlinear systems," Physica D: Nonlinear Phenomena **241**, 1735–1752 (2012).

[104] T. N. Palmer, "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**, 279–304 (2001).

[105] J. Berner, U. Achatz, L. Batté, L. Bengtsson, A. d. l. Cámara, H. M. Christensen, M. Colangeli, D. R. Coleman, D. Crommelin, S. I. Dolaptchiev, *et al.*, "Stochastic parameterization: Toward a new view of weather and climate models," Bulletin of the American Meteorological Society **98**, 565–588 (2017).

[106] J. W.-B. Lin and J. D. Neelin, "Toward stochastic deep convective parameterization in general circulation models," Geophysical research letters **30** (2003).

[107] M. Branicki and A. J. Majda, "Quantifying uncertainty for predictions with model error in non-Gaussian systems with intermittency," Nonlinearity **25**, 2543 (2012).

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

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

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

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

[112] C. Sparrow, *The Lorenz equations: bifurcations, chaos, and strange attractors*, Vol. 41 (Springer Science & Business Media, 2012).

[113] H. Haken, "Analogy between higher instabilities in fluids and lasers," Physics Letters A **53**, 77–78 (1975).

[114] E. Knobloch, "Chaos in the segmented disc dynamo," Physics Letters A **82**, 439–440 (1981).

[115] M. Gorman, P. Widmann, and K. Robbins, "Nonlinear dynamics of a convection loop: a quantitative comparison of experiment with theory," Physica D: Nonlinear Phenomena **19**, 255–267 (1986).

[116] N. Hemati, "Strange attractors in brushless DC motors," IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications **41**, 40–45 (1994).

[117] K. M. Cuomo and A. V. Oppenheim, "Circuit implementation of synchronized chaos with applications to communications," Physical review letters **71**, 65 (1993).

[118] D. Poland, "Cooperative catalysis and chemical chaos: a chemical model for the Lorenz

equations," Physica D: Nonlinear Phenomena **65**, 86–99 (1993).

[119]S. I. Tzenov, "Strange attractors characterizing the osmotic instability," arXiv preprint arXiv:1406.0979 (2014).

[120]J. G. Charney and J. G. DeVore, "Multiple flow equilibria in the atmosphere and blocking," Journal of the atmospheric sciences **36**, 1205–1216 (1979).

[121]A. Wiin-Nielsen, "Steady states and stability properties of a low-order barotropic system with forcing and dissipation," Tellus **31**, 375–386 (1979).

[122]C. Franzke, D. Crommelin, A. Fischer, and A. J. Majda, "A hidden Markov model perspective on regimes and metastability in atmospheric flows," Journal of Climate **21**, 1740–1757 (2008).

[123]F. Baur, "Extended-range weather forecasting," in *Compendium of meteorology* (Springer, 1951) pp. 814–833.

[124]D. F. Rex, "Blocking action in the middle troposphere and its effect upon regional climate: I. An aerological study of blocking action," Tellus **2**, 196–211 (1950).

[125]R. M. Dole and N. D. Gordon, "Persistent anomalies of the extratropical northern hemisphere wintertime circulation: Geographical distribution and regional persistence characteristics," Monthly Weather Review **111**, 1567–1586 (1983).

[126]X. Cheng and J. M. Wallace, "Cluster analysis of the northern hemisphere wintertime 500-hPa height field: Spatial patterns," Journal of the Atmospheric Sciences **50**, 2674–2696 (1993).

[127]M. Kimoto and M. Ghil, "Multiple flow regimes in the northern hemisphere winter. part II: Sectorial regimes and preferred transitions," Journal of the atmospheric sciences **50**, 2645–2673 (1993).

[128]C. Franzke and A. J. Majda, "Low-order stochastic mode reduction for a prototype atmospheric GCM," Journal of the atmospheric sciences **63**, 457–479 (2006).

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

[130]A. J. Majda, C. Franzke, and B. Khouider, "An applied mathematics perspective on stochastic modelling for climate," Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **366**, 2427–2453 (2008).

[131]D. Qi and A. J. Majda, "Low-dimensional reduced-order models for statistical response

and uncertainty quantification: Barotropic turbulence with topography," Physica D: Nonlinear Phenomena **343**, 7–27 (2017).

[132] D. Qi and A. J. Majda, "Rigorous statistical bounds in uncertainty quantification for one-layer turbulent geophysical flows," Journal of Nonlinear Science **28**, 1709–1761 (2018).

[133] D. T. Crommelin, J. Opsteegh, and F. Verhulst, "A mechanism for atmospheric regime behavior," Journal of the atmospheric sciences **61**, 1406–1419 (2004).

[134] D. Crommelin and A. Majda, "Strategies for model reduction: Comparing different optimal bases," Journal of the Atmospheric Sciences **61**, 2206–2217 (2004).