Conditional Gaussian Systems for Multiscale Nonlinear Stochastic Systems: Prediction, State Estimation and Uncertainty Quantification

Nan Chen [1,*] and Andrew J. Majda [1,2]

[1] Department of Mathematics and Center for Atmosphere Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, NY 10012, USA; jonjon@cims.nyu.edu

[2] Center for Prototype Climate Modeling, New York University Abu Dhabi, Saadiyat Island, Abu Dhabi 129188, UAE

[*] Correspondence: nan.chen@nyu.edu or chennan@cims.nyu.edu

check for updates

Abstract: A conditional Gaussian framework for understanding and predicting complex multiscale nonlinear stochastic systems is developed. Despite the conditional Gaussianity, such systems are nevertheless highly nonlinear and are able to capture the non-Gaussian features of nature. The special structure of the system allows closed analytical formulae for solving the conditional statistics and is thus computationally efficient. A rich gallery of examples of conditional Gaussian systems are illustrated here, which includes data-driven physics-constrained nonlinear stochastic models, stochastically coupled reaction–diffusion models in neuroscience and ecology, and large-scale dynamical models in turbulence, fluids and geophysical flows. Making use of the conditional Gaussian structure, efficient statistically accurate algorithms involving a novel hybrid strategy for different subspaces, a judicious block decomposition and statistical symmetry are developed for solving the Fokker–Planck equation in large dimensions. The conditional Gaussian framework is also applied to develop extremely cheap multiscale data assimilation schemes, such as the stochastic superparameterization, which use particle filters to capture the non-Gaussian statistics on the large-scale part whose dimension is small whereas the statistics of the small-scale part are conditional Gaussian given the large-scale part. Other topics of the conditional Gaussian systems studied here include designing new parameter estimation schemes and understanding model errors.

Keywords: conditional Gaussian systems; multiscale nonlinear stochastic systems; physics-constrained nonlinear stochastic models; stochastically coupled reaction–diffusion models; conditional Gaussian mixture; superparameterization; conformation theory; model error; hybrid strategy; parameter estimation

Contents

1. Introduction

　　Multiscale nonlinear dynamical systems are ubiquitous in different areas, including geoscience, engineering, neural science and material science [1–8]. They are characterized by a large dimensional phase space and a large dimension of strong instabilities which transfer energy throughout the system. Many non-Gaussian characteristics, such as extreme and rare events, intermittency and fat-tailed probability density functions (PDFs), are often observed in these multiscale nonlinear dynamical systems [1,9,10]. Key mathematical issues are their basic mathematical structural properties and qualitative features [2,3,11,12], short- and long-range forecasting [13–15], uncertainty quantification (UQ) [16–18], and state estimation, filtering or data assimilation [15,19–22]. Due to the lack of physical understanding or the inadequate resolution in the models because of the current computing power, coping with the inevitable model errors that arise in approximating such complex systems becomes a necessary and crucial issue in dealing with these multiscale nonlinear dynamical systems [16,23–26]. Understanding and predicting complex multiscale turbulent dynamical systems involve the blending of rigorous mathematical theory, qualitative and quantitative modelling, and novel numerical procedures [2,27].

　　Despite the fully nonlinearity in many multiscale turbulent dynamical systems and the non-Gaussian features in both the marginal and joint PDFs, these systems have conditional Gaussian structures [1,15,28,29]. Writing the state variables as $\mathbf{u} = (\mathbf{u_I}, \mathbf{u_{II}})$, the conditional Gaussianity means that once the trajectories of $\mathbf{u_I}$ are given, the dynamics of $\mathbf{u_{II}}$ conditioned on these highly nonlinear observed trajectories become Gaussian processes. In this article, we develop a general conditional Gaussian framework for mathematical modeling, prediction, state estimation and UQ. One of the desirable features of such conditional Gaussian system is that it allows closed analytical formulae for updating the conditional statistics of $\mathbf{u_{II}}$ given the trajectories of $\mathbf{u_I}$ [28,30]. This facilitates the development of both rigorous mathematical theories and efficient numerical algorithms for these complex multiscale turbulent dynamical systems.

　　With the great progress in developing and improving the observational networks, a vast amount of observational data is now available. An important scientific issue is to make use of these observational data to build data-driven models that advance the understanding of the underlying physical processes and improve the prediction skill. Physics-constrained nonlinear stochastic models are the recent development of data-driven statistical-dynamical models for the time series of a partial subset of observed variables, which arise from observations of nature or from an extremely complex physical model [31,32]. The physics-constrained nonlinear stochastic modeling framework builds on early works for single layer models without memory effects, which uses physical analytic properties to constrain data-driven methods. They succeed in overcoming both the finite-time blowup and the lack of physical meaning issues in various ad hoc multi-layer regression models [31,33]. Many of the

physics-constrained nonlinear stochastic models belong to the conditional Gaussian framework, including the noisy version of the famous Lorenz 63 and 84 models as well as a two-layer Lorenz 96 model [34–36]. In addition, two important conceptual models for turbulent dynamical systems [37,38] and a conceptual model of the coupled atmosphere and ocean [39] also fit perfectly into the physics-constrained nonlinear stochastic modeling framework with conditional Gaussian structures. These models are extremely useful for testing various new multiscale data assimilation and prediction schemes [1,38,40]. Other physics-constrained nonlinear stochastic models with conditional Gaussian structures include a low-order model of Charney–DeVore flows [29] and a paradigm model for topographic mean flow interaction [32]. Notably, the physics-constrained nonlinear stochastic models combining with the conditional Gaussian framework have been successfully applied for the real-time prediction and data assimilation of the Madden–Julian oscillation (MJO) and monsoon [41–44], which are the dominant modes of intraseasonal variabilities in nature.

In many multiscale turbulent dynamical systems, there is a natural time scale separation between different variables. Therefore, the MTV strategy [45–48] (named after Majda, Timofeyev and Vanden-Eijnden) can be applied to these systems for stochastic mode reduction. Using the MTV strategy, the equations of motion for the unresolved fast modes are modified by representing the nonlinear self-interactions terms between unresolved modes by damping and stochastic terms. The resulting systems then naturally belong to the conditional Gaussian framework and they also preserve physics-constrained properties. Examples include a simple stochastic model with key features of atmospheric low-frequency variability [49] and a simple prototype nonlinear stochastic model with triad interactions that mimics structural features of low-frequency variability of global circulation models (GCM) [50] .

In addition to the data-driven physics-constrained nonlinear stochastic models, many large-scale dynamical models in turbulence, fluids and geophysical flows also have conditional Gaussian structures. These stochastic partial differential equation (SPDE) models include the stochastic skeleton model for the MJO [51] and a coupled El Niño dynamical model capturing the El Niño diversity [52], where nonlinearity interacts with state-dependent (multiplicative) noise and many salient non-Gaussian features such as intermittency and fat-tailed PDFs are observed in these systems. The stochastically forced Boussinesq equation [11] for the Rayleigh–Bénard convection [53,54] and strongly stratified flows in geophysics [55] as well as its generalized version, namely the Darcy–Brinkman–Oberbeck–Boussinesq system [56] for the convection phenomena in porous media, also belong to the conditional Gaussian framework. Likewise, the multiscale shallow water equation [11] with coupled geostrophically balanced (GB) and gravity modes fits into the conditional Gaussian model family. On the other hand, quite a few stochastically coupled reaction–diffusion models in neuroscience and ecology illustrate conditional Gaussian structures as well. Examples include the stochastically coupled FitzHugh–Nagumo (FHN) models [57], the predator–prey models [58], a stochastically coupled Susceptible-Infectious-Recovered (SIR) epidemic model [59] and a nutrient-limited model for avascular cancer growth [60]. Furthermore, the conditional Gaussian framework can be applied to model the coupled observation-filtering systems for the state estimation of turbulent ocean flows using Lagrangian tracers [61–63]. The framework can also be utilized to develop cheap and effective stochastic parameterizations for turbulent dynamical systems [64,65].

One of the key issues in studying the complex multiscale nonlinear turbulent dynamical systems is to solve the time evolution of the associated PDFs, which is extremely useful in ensemble prediction, data assimilation as well as understanding the intermittency and extreme events. The time evolution of the PDFs associated with the underlying turbulent dynamical systems is described by the so-called Fokker–Planck equation [66,67]. Due to the complexity of many multiscale turbulent dynamical systems, the dimension of the systems can be quite large. However, solving the Fokker–Planck equation in high dimensions is a well-known notoriously difficult problem. Traditional numerical methods such as finite element and finite difference as well as the direct Monte Carlo simulations all suffer from the curse of dimension [68,69]. Nevertheless, for the conditional Gaussian systems,

efficient statistically accurate algorithms can be developed for solving the Fokker–Planck equation in high dimensions and thus beat the curse of dimension. The algorithms involve a hybrid strategy that requires only a small number of ensembles [38]. Specifically, a conditional Gaussian mixture in a high-dimensional subspace via an extremely efficient parametric method is combined with a judicious non-parametric Gaussian kernel density estimation in the remaining low-dimensional subspace. The parametric method provides closed analytical formulae for determining the conditional Gaussian distributions in the high-dimensional subspace and is therefore computationally efficient and accurate. The full non-Gaussian PDF of the system is then given by a Gaussian mixture [38]. For even larger dimensional systems, a judicious block decomposition and statistical symmetry are further applied that facilitate an extremely efficient parallel computation and a significant reduction of sample numbers [36]. These algorithms are applied here to the statistical prediction of a stochastically coupled FHN model with 1500 dimensions and an inhomogeneous two-layer Lorenz 96 model with 240 dimensions. Significant prediction skill shows the advantages of these algorithms in terms of both accuracy and efficiency.

The conditional Gaussian framework also provides a power tool for the multiscale data assimilation. In fact, data assimilation of turbulent signals is an important challenging problem because of the extremely complicated large dimension of the signals and incomplete partial noisy observations that usually mix the large scale mean flow and small scale fluctuations. Due to the limited computing power, it is desirable to use multiscale forecast models which are cheap and fast to mitigate the curse of dimensionality in turbulent systems. An extremely cheap multiscale data assimilation scheme, called stochastic superparameterization [70–73], resolves the large-scale mean flow on a coarse grid and replaces the nonlinear small-scale eddy terms by quasilinear stochastic processes on formally infinite embedded domains where the stochastic processes are Gaussian conditional to the large scale mean flow. The key ingredient of such a multiscale data assimilation method is the systematic use of conditional Gaussian mixtures which make the methods efficient by filtering a subspace whose dimension is smaller than the full state. This conditional Gaussian closure approximation results in a seamless algorithm without using the high resolution space grid for the small scales and is much cheaper than the conventional superparameterization, with significant success in difficult test problems [71,72,74] including the Majda-McLaughlin-Tabak (MMT) model [71,75] and ocean turbulence [76–78]. The method uses particle filters [20,79] or ensemble filters on the large scale part [75,76] whose dimension is small enough so that the non-Gaussian statistics of the large scale part can be calculated from a particle filter, whereas the statistics of the small scale part are conditionally Gaussian given the large scale part. This framework is not restricted to superparameterization as the forecast model and other cheap forecast models can also be employed. In fact, another multiscale filtering algorithm with quasilinear Gaussian dynamically orthogonality method as the forecast method in an adaptively evolving low dimensional subspace has been developed [80]. The multiscale filtering also provides a mathematical framework for representation errors, which are due to the contribution of unresolved scales [81,82] in the observations. Other theoretic and applied studies of the conditional Gaussian framework include effective parameter estimation, model reduction and the understanding of various model errors using information theory.

The remaining of the article is organized as follows. Section 2 provides an overview of data, model and data-driven modeling framework as well as efficient data assimilation and prediction strategies with solvable conditional statistics. Section 3 summarizes the general mathematical structure of nonlinear conditional Gaussian systems, the physics-constrained nonlinear stochastic models and the application of the MTV strategy to the conditional Gaussian systems. Then, a gallery of examples of conditional Gaussian systems is shown in Section 4. Section 5 involves the effective statistically accurate algorithms that beat the curse of dimension for Fokker–Planck equation for conditional Gaussian systems together with their applications to statistical prediction. The topics related to the parameter estimation, model error and multiscale data assimilation utilizing the conditional Gaussian framework are shown in Section 6. The article is concluded in Section 7.

2. Overview: Data vs. Model, Data-Driven Modeling Framework, and Efficient Data Assimilation and Prediction Strategies with Solvable Conditional Statistics

A central contemporary issue for complex turbulent systems is to use data to improve scientific understanding of the underlying physics, make real-time predictions, and quantify the uncertainty in long range forecasting [1,27]. Recently, with the great progress in developing and improving the observational networks, vast amount of observational data are now available. Many purely data-driven statistical methods (regression, clustering and analog forecast etc.) [83–87] are thus developed and have become popular approaches in attempting to understand and predict nature. Despite the success in a crude understanding of nature in terms of explaining and forecasting some of the coarse-grained variables in the largest scale to some extent, these purely data-driven statistical methods usually cannot discover the nonlinear and intermittent nature in the underlying complex dynamics. They often fail to reveal the non-trivial interactions between physical processes in different scales either. In addition, these purely data-driven statistical methods typically require a large training dataset in order to obtain complete and unbiased information from nature, which is however infeasible in many areas, including climate, atmosphere and ocean science. In fact, satellite observations are only available for a few decades [88,89] which are far from enough in understanding decadal or interannual variabilities containing rare and extreme events. Note that many complex turbulent systems in nature are in high dimensions, and therefore most purely data driven statistical methods are extremely expensive to use.

Therefore, combining model with data becomes necessary in understanding and predicting nature. Suitable models involve important physical mechanisms and they can be used for real-time prediction by incorporating only a small amount of data. In many simple but natural scenarios, some low-dimensional physical variables are observed and low-order nonlinear stochastic models are preferred for describing their behavior. Using data outcome to fit a quadratic regression model [90,91] is a data-driven modeling strategy which outweighs the linear regression models and allows nonlinearity and memory in time. However, there is no physical information in such quadratic regression models. In fact, it has been shown in [33] via rigorous mathematical theory that such ad hoc nonlinear regression strategies can exhibit finite time blow-up and pathological invariant measures even though they fit the data with high precision. To include physical information into the models, systematic physics-constrained multi-level quadratic regression models have been developed [31,32]. These models avoid pathological behavior by incorporating physical constraints with nonlinear energy-conserving principles developed in the earlier stochastic climate modeling strategy [45]. Meanwhile, these physics-constrained models allow memory in effects. Although the number of the parameters in physics-constrained nonlinear models can be large, the models are in general robust with respect to the perturbation of the parameters around their optimal values [41]. This is crucial in practice because it requires only a crude estimation of the parameters for the model, which greatly reduces the computational cost for searching in high-dimensional parameter space. A striking real-world application of these physics-constrained nonlinear models is to assess the predictability limits of time series associated with the tropical intraseasonal variability such as the the Madden–Julian oscillation (MJO) and monsoon. It has been shown in [41–43] that these physics-constrained nonlinear models are able to reach the predictability limits of the large-scale MJO and monsoon and improves prediction skill using the low-order models. Notably, the physics-constrained nonlinear stochastic models require only a short training period [42,44] because the model development automatically involves a large portion of the information of nature. This is extremely important since practically only a limited observational data is available. In fact, comparable and even slightly more skillful prediction results have been found using the physics-constrained nonlinear model compared with those using non-parametric methods in predicting the MJO and monsoon intermittent time series [92,93], but the prediction using the physics-constrained nonlinear model adopted a much shorter training period [44].

Another important reason that entails the model in understanding and predicting nature is due to the fundamental limitations in measuring the observations. In fact, only partial observations are available in most applications and observational noise is inevitable. For example, the sea surface

temperature is easy to measure, but the temperature in the deep ocean is hard to estimate. In addition, although the surface wind in the entire earth can be measured by satellites, the measured signals usually contain large noise. Therefore, models are combined with the available observations for the state estimation of the unobserved variables as well as reducing the noise in the observed ones. This is known as data assimilation or filtering [15,20,22]. Note that most of the complex multiscale dynamical systems involve strong nonlinearity and extreme events. Thus, the classical Kalman filter that works only for linear models [94] fails to capture the nonlinear and non-Gaussian features in nature. Although the so-called particle filter [95,96] is able to recover all the non-Gaussian statistics, it is computationally expensive and can only be applied to low-dimensional scenarios, which are typically not the case in most applications. In fact, even a global weather model with very coarse grid points at every 200 km already has about 10^6 state variables. Other direct strategies in handling large dimensional systems, such as the ensemble Kalman filters [97,98], can also be problematic. For example, they are not immune from "catastrophic filter divergence" (diverge beyond machine infinity) when observations are sparse, even when the true signal is a dissipative system [99,100]. Therefore, designing new efficient strategies for data assimilation and prediction that are accurate and can overcome the curse of dimension is crucial in studying complex multiscale nonlinear dynamical systems with only noisy partial observations [20,27,65,101,102]. Since both the data assimilation and prediction involve running forward the dynamical models (known as the forecast models), the development of new strategies entails the understanding and utilization of model structures.

Due to the complexity in many complex multiscale turbulent dynamical systems, developing cheap and effective approximate models which nevertheless capture the main characteristics of the underlying dynamics becomes necessary in data assimilation, prediction and the understanding of nature. In particular, novel stochastic parameterization schemes play a significant role in reducing the model complexity while retaining the key features of various multiscale turbulent dynamical systems. These key features include the statistical feedback from small to large scales, accurate dynamical and statistical behavior in large scale variables, and the main effect of the large-scale variables on the statistical evolution of small scale processes. Then, efficient hybrid strategy can be developed for dealing with large and small scale variables, respectively.

A simple yet practically useful strategy in filtering nonlinear and intermittent signals is via the so-called stochastic parameterized extended Kalman filter (SPEKF) type of the forecast model [64,65]. The idea of the SKEPF model is that the small or unresolved scale variables are stochastically parameterized by cheap linear and Gaussian processes, representing stochastic damping, stochastic phase and stochastic forcing. Despite the model error in using such Gaussian approximations for the unresolved nonlinear dynamics, these Gaussian processes succeed in providing accurate statistical feedback from the unresolved scales to the resolved ones and thus the intermittency and non-Gaussian features as observed in the resolved variables can be accurately recovered. The statistics in the SPEKF model can also be solved with exact and analytic formulae, which allow an accurate and efficient estimation of the model states. The SPEKF type of model has been used for filtering mutiscale turbulent dynamical systems [20], stochastic superresolution [103] and filtering Navier–Stokes equations with model error [104]. A detailed description of the SPEKF model can be found in Section 4.5.1. Notably, the SPEKF model involves a hybrid approach, where certain cheap and statistically solvable Gaussian approximations are used to describe the statistical features of the unresolved (or small scale) variables in a large dimensional subspace while the highly nonlinear dynamical features of the resolved (or large scale) variables as well as their interactions with the unresolved variables are retained in a relatively low dimensional subspace. Such a strategy can be used as a guideline for designing suitable approximate forecast models for various complex multiscale turbulent dynamical systems.

For multiscale data assimilation of many realistic and complex turbulent dynamical systems, an extremely cheap scheme, called stochastic superparameterization [70–73], has been developed, which makes use of a similar hybrid strategy as discussed above but with a more refined setup. In the stochastic superparameterization, the large-scale mean flow is resolved on a coarse grid and

the nonlinear small-scale eddy terms are replaced by quasilinear stochastic processes on formally infinite embedded domains where these stochastic processes are Gaussian conditional to the large scale mean flow. The key ingredient of such a multiscale data assimilation method is the systematic use of conditional Gaussian mixtures which make the methods efficient by filtering a subspace whose dimension is smaller than the full state. This conditional Gaussian closure approximation results in a seamless algorithm without using the high resolution space grid for the small scales and is thus extremely efficient. The method uses particle filters [20,79] or ensemble filters on the large scale part [75,76] whose dimension is small enough so that the non-Gaussian statistics of the large scale part can be calculated from a particle filter, whereas the statistics of the small scale part are conditionally Gaussian given the large scale part. See Section 6.3 for more details. Note that such ideas can also be applied to other cheap forecast models in addition to the superparameterization. For example, another multiscale filtering algorithm with quasilinear Gaussian dynamically orthogonality method as the forecast method in an adaptively evolving low dimensional subspace has been developed in [80].

Due to the significance in model approximation, data assimilation and prediction, it is necessary to develop a general framework for the complex multiscale nonlinear turbulent dynamical systems with conditional Gaussian structures. The key idea of such conditional Gaussian systems is that a hybrid strategy can be applied to deal with the state variables $\mathbf{u} = (\mathbf{u_I}, \mathbf{u_{II}})$, where particle methods are used to solve the non-Gaussian statistics in the relatively low dimensional subspace associated with $\mathbf{u_I}$ and extremely cheap algorithms with closed analytical formulae are adopted to solve the conditional Gaussian statistics of $\mathbf{u_{II}}$ conditioned on $\mathbf{u_I}$ in the remaining high dimensional subspace. The Gaussianity in the conditional statistics are usually quite accurate as approximations since the small or unresolved scales with the given (or fixed) large-scale variables usually represent fluctuations in multiscale systems and the statistics are close to Gaussian. Nevertheless, the marginal statistics of the small-scale variables $\mathbf{u_{II}}$ themselves can be highly non-Gaussian, which is one of the salient features as in nature [37]. Notably, developing such computationally efficient models that explores conditional statistics also involves the mutual feedback between large scale and small scale variables. Thus, the full system is completely nonlinear and allows physics-constrained nonlinear interactions. The general conditional Gaussian nonlinear modeling framework provides a powerful tool in multiscale data assimilation, statistical prediction, solving high-dimensional PDFs as well as parameter estimation, causality analysis and understanding model errors.

3. A Summary of the General Mathematical Structure of Nonlinear Conditional Gaussian Systems

3.1. Conditional Gaussian Nonlinear Dynamical Systems

The conditional Gaussian systems have the following abstract form [28]:

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

$$d\mathbf{u_{II}} = [\mathbf{a}_0(t, \mathbf{u_I}) + \mathbf{a}_1(t, \mathbf{u_I})\mathbf{u_{II}}]dt + \mathbf{\Sigma_{II}}(t, \mathbf{u_I})d\mathbf{W_{II}}(t). \tag{2}$$

Once $\mathbf{u_I}(s)$ for $s \leq t$ is given, $\mathbf{u_{II}}(t)$ conditioned on $\mathbf{u_I}(s)$ becomes a Gaussian process:

$$p\big(\mathbf{u_{II}}(t)|\mathbf{u_I}(s \leq t)\big) \sim \mathcal{N}(\bar{\mathbf{u}}_{II}(t), \mathbf{R_{II}}(t)). \tag{3}$$

Despite the conditional Gaussianity, the coupled system (1) and (2) remains highly nonlinear and is able to capture the non-Gaussian features as in nature. The conditional Gaussian distribution $p\big(\mathbf{u_{II}}(t)|\mathbf{u_I}(s \leq t)\big)$ in (3) has closed analytic form [30]

$$d\bar{\mathbf{u}}_{II}(t) = [\mathbf{a}_0(t, \mathbf{u_I}) + \mathbf{a}_1(t, \mathbf{u_I})\bar{\mathbf{u}}_{II}]dt + (\mathbf{R_{II}}\mathbf{A}_1^*(t, \mathbf{u_I}))(\mathbf{\Sigma_I}\mathbf{\Sigma_I^*})^{-1}(t, \mathbf{u_I}) \times \\ [d\mathbf{u_I} - (\mathbf{A}_0(t, \mathbf{u_I}) + \mathbf{A}_1(t, \mathbf{u_I})\bar{\mathbf{u}}_{II})dt], \tag{4}$$

$$d\mathbf{R_{II}}(t) = \Big\{ \mathbf{a}_1(t, \mathbf{u_I})\mathbf{R_{II}} + \mathbf{R_{II}}\mathbf{a}_1^*(t, \mathbf{u_I}) + (\Sigma_{II}\Sigma_{II}^*)(t, \mathbf{u_I})$$
$$- (\mathbf{R_{II}}\mathbf{A}_1^*(t, \mathbf{u_I}))(\Sigma_{I}\Sigma_{I}^*)^{-1}(t, \mathbf{u_I})(\mathbf{R_{II}}\mathbf{A}_1^*(t, \mathbf{u_I}))^* \Big\}dt, \tag{5}$$

which can be solved in an exact and efficient way. The conditional Gaussian framework (1)-(2)-(5) is useful in the parameter estimation, data assimilation, prediction and uncertainty quantification of complex turbulent dynamical systems as will be discussed throughout this article.

In this article, many conditional Gaussian systems with complex turbulent dynamical structures will be studied. For the convenience of the reader, we will always use a blue color to represent the variables that belong to $\mathbf{u_I}$ and use a magenta color to denote those for $\mathbf{u_{II}}$.

3.2. Kalman–Bucy Filter: The Simplest and Special Data Assimilation Example within Conditional Gaussian Framework

A special case of the general conditional Gaussian framework (1) and (2) is the so-called Kalman–Bucy filter [105–108]. The Kalman–Bucy filter is a continuous time version of the Kalman filter [21,94] and it deals with the linear coupled systems:

$$d\mathbf{u_I} = [\mathbf{A}_0(t) + \mathbf{A}_1(t)\mathbf{u_{II}}]dt + \Sigma_I(t)d\mathbf{W_I}(t), \tag{6}$$
$$d\mathbf{u_{II}} = [\mathbf{a}_0(t) + \mathbf{a}_1(t)\mathbf{u_{II}}]dt + \Sigma_{II}(t)d\mathbf{W_{II}}(t). \tag{7}$$

In (6) and (7), all the vectors and matrices $\mathbf{A}_0, \mathbf{A}_1, \mathbf{a}_0, \mathbf{a}_1, \Sigma_I$ and Σ_{II} are functions of only t and they have no dependence on $\mathbf{u_I}$ in order to guarantee the linearity in the coupled system. In the Kalman–Bucy filter, (7) is the underlying dynamics and (6) is the observational process. The observation cannot change the underlying dynamics and therefore no $\mathbf{u_I}$ appears in (7).

The filter estimate (also known as the posterior distribution) is the conditional distribution of $\mathbf{u_{II}}(t)$ given the observation $\mathbf{u_I}(s \leq t)$, i.e., $p(\mathbf{u_{II}}(t)|\mathbf{u_I}(s \leq t))$. In light of (4) and (5), the mean and variance of $p(\mathbf{u_{II}}|\mathbf{u_I})$ has the following explicit expressions:

$$d\overline{\mathbf{u}}_{\mathbf{II}}(t) = [\mathbf{a}_0(t) + \mathbf{a}_1(t)\overline{\mathbf{u}}_{\mathbf{II}}]dt + (\mathbf{R_{II}}\mathbf{A}_1^*(t))(\Sigma_I\Sigma_I^*)^{-1}(t)\left[d\mathbf{u_I} - (\mathbf{A}_0(t) + \mathbf{A}_1(t)\overline{\mathbf{u}}_{\mathbf{II}})dt\right],$$
$$d\mathbf{R_{II}}(t) = \Big\{\mathbf{a}_1(t)\mathbf{R_{II}} + \mathbf{R_{II}}\mathbf{a}_1^*(t) + (\Sigma_{II}\Sigma_{II}^*)(t) - (\mathbf{R_{II}}\mathbf{A}_1^*(t))(\Sigma_I\Sigma_I^*)^{-1}(t)(\mathbf{R_{II}}\mathbf{A}_1^*(t))^*\Big\}dt.$$

Corresponding to (4) and (5), Chapter 6 of Bensoussan's book [109] includes rigorous mathematical derivations of the exact solutions of the Kalman–Bucy filter and some other more general conditional Gaussian filters. It is also pointed out in [109] that all these filters belong to the general conditional Gaussian filtering framework in (1)-(2)-(4)-(5) introduced in [110], which is an early version of the book authored by Liptser and Shiryaev [30].

3.3. Physics-Constrained Nonlinear Models with Conditional Gaussian Statistics

Physics-constrained nonlinear stochastic models are the recent development of data driven statistical-dynamical models for the time series of a partial subset of observed variables, which arise from observations of nature or from an extremely complex physical model [31,32]. The physics-constrained nonlinear stochastic modeling framework builds on early works for single layer models without memory effects, which uses physical analytic properties to constrain data driven methods. These physics-constrained nonlinear stochastic models succeed in overcoming both the finite-time blowup and the lack of physical meaning issues in various ad hoc multi-layer regression models [31,33].

The physics-constrained nonlinear stochastic models contain energy-conserving quadratic nonlinear interactions [1,31,32], namely

$$d\mathbf{u} = \left[(\mathbf{L} + \mathbf{D})\mathbf{u} + \mathbf{B}(\mathbf{u}, \mathbf{u}) + \mathbf{F}(t)\right]dt + \Sigma(t, \mathbf{u})d\mathbf{W}(t),$$
$$\text{with} \quad \mathbf{u} \cdot \mathbf{B}(\mathbf{u}, \mathbf{u}) = 0, \tag{8}$$

where $\mathbf{u} = (\mathbf{u_I}, \mathbf{u_{II}})$ and the dimensions of $\mathbf{u_I}$ and $\mathbf{u_{II}}$ are N_I and N_{II}, respectively. In (8), $\mathbf{L} + \mathbf{D}$ is a linear operator representing dissipation and dispersion. Here, $\mathbf{L}$ is skew symmetric representing dispersion and $\mathbf{D}$ is a negative definite symmetric operator representing dissipative process such as surface drag, radiative damping, viscosity, etc. $B(\mathbf{u}, \mathbf{u})$ is a bilinear term and it satisfies energy conserving property with $\mathbf{u} \cdot B(\mathbf{u}, \mathbf{u}) = 0$.

Many of the physics-constrained nonlinear stochastic models belong to the conditional Gaussian framework. The goal here is to derive a general class of conditional Gaussian physics-constrained nonlinear stochastic models. To this end, we rewrite the Equation (8) in the following way:

$$
\begin{aligned}
d\mathbf{u_I} &= \left(\mathbf{L_{I,1}}\mathbf{u_I} + \mathbf{L_{I,2}}\mathbf{u_{II}} + \mathbf{B_I}(\mathbf{u}, \mathbf{u}) + \mathbf{F_I} \right) dt + \mathbf{\Sigma_I}(\mathbf{u_I}) d\mathbf{W_I}, \\
d\mathbf{u_{II}} &= \left(\mathbf{L_{II,1}}\mathbf{u_I} + \mathbf{L_{II,2}}\mathbf{u_{II}} + \mathbf{B_{II}}(\mathbf{u}, \mathbf{u}) + \mathbf{F_{II}} \right) dt + \mathbf{\Sigma_{II}}(\mathbf{u_I}) d\mathbf{W_{II}},
\end{aligned}
\tag{9}
$$

where the explicit dependence of the coefficients on time t has been omitted for notation simplicity. In (9), $\mathbf{L_{I,1}}\mathbf{u_I}$, $\mathbf{L_{I,2}}\mathbf{u_{II}}$, $\mathbf{L_{II,1}}\mathbf{u_I}$ and $\mathbf{L_{II,2}}\mathbf{u_{II}}$ correspond to the the linear term $\mathbf{L} + \mathbf{D}$ in (8) while $\mathbf{B_I}(\mathbf{u}, \mathbf{u})$ and $\mathbf{B_{II}}(\mathbf{u}, \mathbf{u})$ represent the nonlinear terms in the processes associated with the variables in (1) and those in (2), respectively. Since the conditional Gaussian systems do not allow quadratic nonlinear interactions between $\mathbf{u_{II}}$ and itself, both $\mathbf{B_I}(\mathbf{u}, \mathbf{u})$ and $\mathbf{B_{II}}(\mathbf{u}, \mathbf{u})$ can be written down as follows:

$$
\begin{aligned}
\mathbf{B_I}(\mathbf{u}, \mathbf{u}) &= \mathbf{B_{I,1}}(\mathbf{u_I}, \mathbf{u_I}) + \mathbf{B_{I,2}}(\mathbf{u_I}, \mathbf{u_{II}}), \\
\mathbf{B_{II}}(\mathbf{u}, \mathbf{u}) &= \mathbf{B_{II,1}}(\mathbf{u_I}, \mathbf{u_I}) + \mathbf{B_{II,2}}(\mathbf{u_I}, \mathbf{u_{II}}),
\end{aligned}
\tag{10}
$$

where $\mathbf{B_{\cdot,1}}(\mathbf{u_I}, \mathbf{u_I})$ stands for the quadratic terms involving only $\mathbf{u_I}$ and $\mathbf{B_{\cdot,2}}(\mathbf{u_I}, \mathbf{u_{II}})$ represents the quadratic interactions between $\mathbf{u_I}$ and $\mathbf{u_{II}}$. Given the nonlinear terms in (10), the energy-conserving quadratic nonlinearity in (8) implies

$$
\mathbf{u_I} \cdot \left(\mathbf{B_{I,1}}(\mathbf{u_I}, \mathbf{u_I}) + \mathbf{B_{I,2}}(\mathbf{u_I}, \mathbf{u_{II}}) \right) + \mathbf{u_{II}} \cdot \left(\mathbf{B_{II,1}}(\mathbf{u_I}, \mathbf{u_I}) + \mathbf{B_{II,2}}(\mathbf{u_I}, \mathbf{u_{II}}) \right) = 0.
\tag{11}
$$

Inserting (10) into (9) yields the conditional Gaussian systems with energy-conserving quadratic nonlinear interactions,

$$
d\mathbf{u_I} = \left(\mathbf{B_{I,1}}(\mathbf{u_I}, \mathbf{u_I}) + \mathbf{B_{I,2}}(\mathbf{u_I}, \mathbf{u_{II}}) + \mathbf{L_{I,1}}\mathbf{u_I} + \mathbf{L_{I,2}}\mathbf{u_{II}} + \mathbf{F_I} \right) dt + \mathbf{\Sigma_I}(\mathbf{u_I}) d\mathbf{W_I},
\tag{12}
$$

$$
d\mathbf{u_{II}} = \left(\mathbf{B_{II,1}}(\mathbf{u_I}, \mathbf{u_I}) + \mathbf{B_{II,2}}(\mathbf{u_I}, \mathbf{u_{II}}) + \mathbf{L_{II,1}}\mathbf{u_I} + \mathbf{L_{II,2}}\mathbf{u_{II}} + \mathbf{F_{II}} \right) dt + \mathbf{\Sigma_{II}}(\mathbf{u_I}) d\mathbf{W_{II}},
\tag{13}
$$

Now, we explore the detailed forms of the energy-conserving nonlinear terms in (12) and (13). We start with $\mathbf{B_{II,2}}(\mathbf{u_I}, \mathbf{u_{II}})$, which can be written as

$$
\mathbf{B_{II,2}}(\mathbf{u_I}, \mathbf{u_{II}}) = \mathbf{S_{II}}(\mathbf{u_I})\mathbf{u_{II}}, \qquad \text{with} \qquad \mathbf{S_{II}}(\mathbf{u_I}) = \sum_{j=1}^{N_I} S_{II,j} u_{I,j},
\tag{14}
$$

where each $S_{II,j}$ is a skew-symmetric matrix with $S_{II,j}^T = -S_{II,j}$ and $u_{I,j}$ is the j-th entry of $\mathbf{u_I}$. The energy-conserving property is easily seen by multiplying $\mathbf{u_{II}}$ to $\mathbf{B_{II,2}}(\mathbf{u_I}, \mathbf{u_{II}})$ in (14),

$$
\mathbf{u_{II}} \cdot \mathbf{B_{II,2}}(\mathbf{u_I}, \mathbf{u_{II}}) = \mathbf{u_{II}} \cdot \mathbf{S}(\mathbf{u_I})\mathbf{u_{II}} = \sum_{j=1}^{N_I} u_{I,j} \cdot \left(\mathbf{u_{II}} \cdot S_{II,j}\mathbf{u_{II}} \right) = 0.
$$

Due to the skew-symmetric property of $S_{II,j}$. In fact, $\mathbf{B_{II,2}}(\mathbf{u_I}, \mathbf{u_{II}})$ usually represents the internal oscillation with non-constant oscillation frequency that depends on $\mathbf{u_I}$.

Next, $\mathbf{B_{I,2}}(\mathbf{u_I}, \mathbf{u_{II}})$ contains three components,

$$
\mathbf{B_{I,2}}(\mathbf{u_I}, \mathbf{u_{II}}) = \mathbf{B_{I,2}^1}(\mathbf{u_I}, \mathbf{u_{II}}) + \mathbf{B_{I,2}^2}(\mathbf{u_I}, \mathbf{u_{II}}) + \mathbf{B_{I,2}^3}(\mathbf{u_I}, \mathbf{u_{II}}).
\tag{15}
$$

One of the components in (15), say $\mathbf{B_{I,2}^1}(\mathbf{u_I}, \mathbf{u_{II}})$, has its own energy conservation, i.e.,

$$\mathbf{u_I} \cdot \mathbf{B_{I,2}^1}(\mathbf{u_I}, \mathbf{u_{II}}) = 0.$$

Here, $\mathbf{B_{I,2}^1}(\mathbf{u_I}, \mathbf{u_{II}}) = \mathbf{S_I}(\mathbf{u_I})\mathbf{u_{II}}$ and therefore

$$\mathbf{u_I} \cdot \mathbf{S_I}(\mathbf{u_I})\mathbf{u_{II}} = 0, \tag{16}$$

where each column of $\mathbf{S_I}(\mathbf{u_I})$ is given by

$$\mathbf{S_{I,j}}(\mathbf{u_I}) = S_{I,j}\mathbf{u_I}, \tag{17}$$

with $S_{I,j}$ being a skew-symmetric matrix. Thus, with (17) in hand, (16) becomes

$$\sum_{j=1}^{N_{II}} \left(\mathbf{u_I} \cdot S_{I,j} \cdot \mathbf{u_I} \right) \mathbf{u_{II,j}} = 0,$$

where $\mathbf{u_{II,j}}$ is the j-th entry of $\mathbf{u_{II}}$.

The other two components of $\mathbf{B_{I,2}}(\mathbf{u_I}, \mathbf{u_{II}})$ in (12) involve the interactions with $\mathbf{B_{II,1}}(\mathbf{u_I}, \mathbf{u_I}) = \mathbf{B_{II,1}^2}(\mathbf{u_I}, \mathbf{u_I}) + \mathbf{B_{II,1}^3}(\mathbf{u_I}, \mathbf{u_I})$ in (13). On one hand, the energy-conserving property in the following two terms is obvious,

$$\mathbf{B_{I,2}^2}(\mathbf{u_I}, \mathbf{u_{II}}) = \sum_{j=1}^{N_I} \Gamma_j \mathbf{u_{I,j}} \mathbf{u_{II}}, \tag{18}$$

$$\mathbf{B_{II,1}^2}(\mathbf{u_I}, \mathbf{u_I}) = - \sum_{j=1}^{N_I} \Gamma_j^T \mathbf{u_I^2}, \tag{19}$$

where each Γ_j is a $N_I \times N_{II}$ matrix, $\mathbf{u_{I,j}}$ is the j-th entry of $\mathbf{u_I}$ and $\mathbf{u_I^2}$ is a vector of size $N_I \times 1$ with the j-th entry being $\mathbf{u_{I,j}^2}$. On the other hand, the remaining two terms $\mathbf{B_{I,2}^3}(\mathbf{u_I}, \mathbf{u_{II}})$ and $\mathbf{B_{II,1}^3}(\mathbf{u_I}, \mathbf{u_I})$ are similar to those in (18)–(20) but deal with the cross-interactions between different components of $\mathbf{u_I}$ such as replacing $\mathbf{u_I^2}$ by $\mathbf{u_{I,j_1}} \mathbf{u_{I,j_2}}$ in (20). To this end, we define the following

$$\mathbf{G}(\mathbf{u_I}) = \sum_{j=1}^{N_I} G_j \mathbf{u_{I,j}}, \tag{20}$$

which satisfies

$$\mathbf{u_I} \cdot \mathbf{G}(\mathbf{u_I})\mathbf{u_{II}} - \mathbf{u_{II}} \cdot \mathbf{G}^T(\mathbf{u_I})\mathbf{u_I} = 0. \tag{21}$$

In fact, (18)–(21) are important for generating the intermittent instability, where $\mathbf{u_{II}}$ plays the role of both damping and anti-damping for the dynamics of $\mathbf{u_I}$.

Finally, $\mathbf{B_{I,1}}(\mathbf{u_I}, \mathbf{u_I})$ involves any iterations between $\mathbf{u_I}$ and itself that satisfies

$$\mathbf{u_I} \cdot \mathbf{B_{I,1}}(\mathbf{u_I}, \mathbf{u_I}) = 0. \tag{22}$$

Therefore, with (14)–(22) in hand, the conditional Gaussian system with energy-conserving quadratic nonlinear interactions (12) and (13) has the following form:

$$d\mathbf{u_I} = \left(\mathbf{B_{I,1}}(\mathbf{u_I}, \mathbf{u_I}) + \sum_{j=1}^{N_I} \Gamma_j u_{I,j}\mathbf{u_{II}} + \mathbf{S_I}(\mathbf{u_I})\mathbf{u_{II}} + \mathbf{G}(\mathbf{u_I})\mathbf{u_{II}} + \mathbf{L_{I,1}}\mathbf{u_I} + \mathbf{L_{I,2}}\mathbf{u_{II}} + \mathbf{F_I} \right)dt + \mathbf{\Sigma_I}(\mathbf{u_I})d\mathbf{W_I}, \tag{23}$$

$$d\mathbf{u_{II}} = \left(\mathbf{S_{II}}(\mathbf{u_I})\mathbf{u_{II}} - \sum_{j=1}^{N_I} \Gamma_j^T \mathbf{u_I^2} + \mathbf{L_{II,1}}\mathbf{u_I} + \mathbf{L_{II,2}}\mathbf{u_{II}} - \mathbf{G}^T(\mathbf{u_I})\mathbf{u_I} + \mathbf{F_{II}} \right)dt + \mathbf{\Sigma_{II}}(\mathbf{u_I})d\mathbf{W_{II}}. \tag{24}$$

3.4. Multiscale Conditional Gaussian with MTV Stochastic Modeling Strategy

Let's start with a general nonlinear deterministic model with quadratic nonlinearity,

$$d\mathbf{u} = \big[(\mathbf{L} + \mathbf{D})\mathbf{u} + \mathbf{B}(\mathbf{u}, \mathbf{u}) + \mathbf{F}(t)\big]dt, \tag{25}$$

where the notations of vectors and matrices are the same as in Section 3.3. In (25), the state variables are again decomposed into $\mathbf{u} = (\mathbf{u_I}, \mathbf{u_{II}})$. Here, $(\mathbf{u_I}, \mathbf{u_{II}})$ has multiscale features, where $\mathbf{u_I}$ denotes the resolved variables that evolve slowly in time (e.g., climate variables) while $\mathbf{u_{II}}$ are unresolved or unobserved fast variables (e.g., weather variables). The system (25) can be written down into more detailed forms:

$$\begin{aligned}
d\mathbf{u_I} = \Big[&(\mathbf{L}_{11} + \mathbf{D}_{11})\mathbf{u_I} + (\mathbf{L}_{12} + \mathbf{D}_{12})\mathbf{u_{II}} + \mathbf{B}^1_{11}(\mathbf{u_I}, \mathbf{u_I}) \\
&+ \mathbf{B}^1_{12}(\mathbf{u_I}, \mathbf{u_{II}}) + \mathbf{B}^1_{22}(\mathbf{u_{II}}, \mathbf{u_{II}}) + \mathbf{F}_1(t)\Big]dt,
\end{aligned} \tag{26}$$

$$\begin{aligned}
d\mathbf{u_{II}} = \Big[&(\mathbf{L}_{22} + \mathbf{D}_{22})\mathbf{u_{II}} + (\mathbf{L}_{21} + \mathbf{D}_{21})\mathbf{u_I} + \mathbf{B}^2_{11}(\mathbf{u_I}, \mathbf{u_I}) \\
&+ \mathbf{B}^2_{12}(\mathbf{u_I}, \mathbf{u_{II}}) + \mathbf{B}^2_{22}(\mathbf{u_{II}}, \mathbf{u_{II}}) + \mathbf{F}_2(t)\Big]dt.
\end{aligned} \tag{27}$$

Different from (12) and (13), the nonlinear interaction between fast scale variables themselves $\mathbf{B}^1_{22}(\mathbf{u_{II}}, \mathbf{u_{II}})$ and $\mathbf{B}^2_{22}(\mathbf{u_{II}}, \mathbf{u_{II}})$ are also included in (26) and (27). To make the mutiscale system (26) and (27) fit into the conditional Gaussian framework, two modifications are needed. First, if we link (26) and (27) with the general conditional Gaussian framework (1) and (2), the quadratic terms involving the interactions between $\mathbf{u_{II}}$ and itself, namely $\mathbf{B}^1_{22}(\mathbf{u_{II}}, \mathbf{u_{II}})$ and $\mathbf{B}^2_{22}(\mathbf{u_{II}}, \mathbf{u_{II}})$, are not allowed there. Second, stochastic noise is required at least to the system of $\mathbf{u_I}$.

To fill in these gaps, the most natural way is to apply idea from the MTV strategy to the multiscale system (26) and (27). The MTV strategy [45–48], named after Majda, Timofeyev and Vanden-Eijnden, is a systematic mathematical strategy for stochastic mode reduction. The MTV strategy is a two-step procedure:

1. The equations of motion for the unresolved fast modes are modified by representing the nonlinear self-interactions terms between unresolved modes by stochastic terms.
2. The equations of motion for the unresolved fast modes are eliminated using the standard projection technique for stochastic differential equations.

In fact, we only need to take the first step in the MTV strategy here. Using ϵ to represent the time scale separation between $\mathbf{u_I}$ and $\mathbf{u_{II}}$, the terms with quadratic nonlinearity of $\mathbf{u_{II}}$ and itself are approximated by

$$\begin{aligned}
\mathbf{B}^1_{22}(\mathbf{u_{II}}, \mathbf{u_{II}}) &\approx -\frac{\mathbf{\Gamma}_1}{\epsilon}\mathbf{u_{II}} + \frac{\mathbf{\Sigma_I}}{\sqrt{\epsilon}}\mathbf{W_I}, \\
\mathbf{B}^2_{22}(\mathbf{u_{II}}, \mathbf{u_{II}}) &\approx -\frac{\mathbf{\Gamma}_2}{\epsilon}\mathbf{u_{II}} + \frac{\mathbf{\Sigma_{II}}}{\sqrt{\epsilon}}\mathbf{W_{II}}.
\end{aligned} \tag{28}$$

That is, we replace the quadratic nonlinearity of $\mathbf{u_{II}}$ by a linear damping and stochastic noise, where $\mathbf{\Gamma}_1$, $\mathbf{\Gamma}_2$, $\mathbf{\Sigma_I}$ and $\mathbf{\Sigma_{II}}$ are positive definite symmetric matrices. The motivation of (28) is that the nonlinear self-interacting terms of fast variables $\mathbf{u_{II}}$ are responsible for the chaotic sensitive dependence on small perturbations and do not require a more detailed description if their effect on the coarse-grained dynamics for the climate variables alone is the main objective. On the other hand, the quadratic nonlinear interactions between $\mathbf{u_I}$ and $\mathbf{u_{II}}$ are retained.

Note that in the second step of the MTV strategy, standard techniques of averaging and adiabatic elimination of fast modes in stochastic equations are applied, assuming that $\epsilon \ll 1$. Here, we assume ϵ is a finite small value and thus do not apply the second step of the MTV strategy.

Now, inserting (28) into (26) and (27) yields

$$d\mathbf{u_I} = \Big[(\mathbf{L_{11}} + \mathbf{D_{11}})\mathbf{u_I} + (\mathbf{L'_{12}} + \mathbf{D_{12}})\mathbf{u_{II}} + \mathbf{B}^1_{11}(\mathbf{u_I}, \mathbf{u_I}) + \mathbf{B}^1_{12}(\mathbf{u_I}, \mathbf{u_{II}}) + \mathbf{F_1}(t)\Big]dt + \mathbf{\Sigma'_I}d\mathbf{W_I}(t), \quad (29)$$

$$d\mathbf{u_{II}} = \Big[(\mathbf{L'_{22}} + \mathbf{D_{22}})\mathbf{u_{II}} + (\mathbf{L_{21}} + \mathbf{D_{21}})\mathbf{u_I} + \mathbf{B}^2_{11}(\mathbf{u_I}, \mathbf{u_I}) + \mathbf{B}^2_{12}(\mathbf{u_I}, \mathbf{u_{II}}) + \mathbf{F_2}(t)\Big]dt + \mathbf{\Sigma'_{II}}\mathbf{W_{II}}(t), \quad (30)$$

where $\mathbf{L'_{12}} = \mathbf{L_{12}} - \mathbf{\Gamma_1}/\epsilon$, $\mathbf{L'_{22}} = \mathbf{L_{22}} - \mathbf{\Gamma_2}/\epsilon$, $\mathbf{\Sigma'_I} = \mathbf{\Sigma_I}/\sqrt{\epsilon}$ and $\mathbf{\Sigma'_{II}} = \mathbf{\Sigma_{II}}/\sqrt{\epsilon}$. Clearly, (29) and (30) belongs to the conditional Gaussian framework. Notably, if the nonlinear terms in (29) and (30) satisfy the conditions in Section 3.3, then (29) and (30) becomes a physics-constrained nonlinear model.

4. A Gallery of Examples of Conditional Gaussian Systems

This section includes various conditional Gaussian complex nonlinear dynamical systems and their applications.

4.1. Physics-Constrained Nonlinear Low-Order Stochastic Models

4.1.1. The Noisy Versions of Lorenz Models

Lorenz proposed three famous models in 1963, 1984 and 1996. These models are widely used as simplified models to mimic the turbulent and chaotic behavior and as test models for data assimilation and prediction in atmosphere and ocean science. By incorporating noise and other small-scale parameterizations, these models all belong to the conditional Gaussian family.

1. The Noisy Lorenz 63 (L-63) Model

$$
\begin{aligned}
dx &= \sigma(y - x)dt + \sigma_x dW_x, \\
dy &= \big(x(\rho - z) - y\big)dt + \sigma_y dW_y, \\
dz &= (xy - \beta z)dt + \sigma_z dW_z.
\end{aligned}
\tag{31}
$$

The deterministic L-63 model ($\sigma_x = \sigma_y = \sigma_z = 0$ in (31)) is proposed by Lorenz in 1963 [34]. It is a simplified mathematical model for atmospheric convection. 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 [111]. The L-63 model is also widely used as simplified models for lasers, dynamos, thermosyphons, brushless direct current DC motors, electric circuits, chemical reactions and forward osmosis [112–118]. the noisy version of the L-63 includes more turbulent and small-scale features and their interactions with the three large scale variables while retains the characteristics in the original L-63. The noisy L-63 model is a conditional Gaussian system (1) and (2) with $\mathbf{u_I} = x$ and $\mathbf{u_{II}} = (y, z)^T$. It also belongs to the conditional Gaussian family with $\mathbf{u_I} = (y, z)^T$ and $\mathbf{u_{II}} = x$.

In Figure 1, we show some sample trajectories of the noisy L-63 model (31) with the typical values that Lorenz used [34],

$$\rho = 28, \qquad \sigma = 10, \qquad \beta = 8/3, \tag{32}$$

Together with a moderate noise level

$$\sigma_x = \sigma_y = \sigma_z = 5. \tag{33}$$

In addition to the chaotic behavior and the butterfly profile, the small fluctuations due to the noise are also clearly observed in these trajectories.
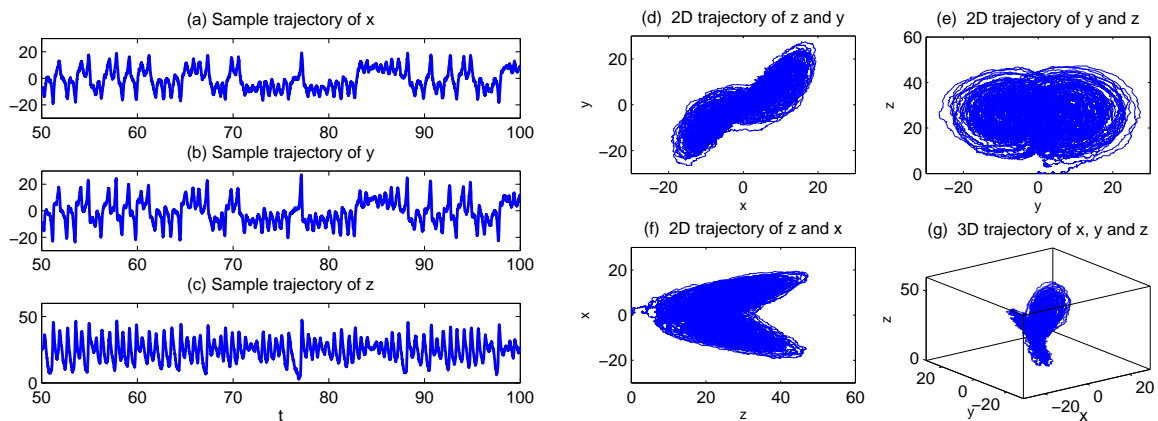
Figure 1. Sample trajectories of the noisy L-63 model (31) with parameters $\rho = 28$, $\sigma = 10$, $\beta = 8/3$, $\sigma_x = \sigma_y = \sigma_z = 5$. (**a–c**) 1D trajectories of x, y and z, respectively; (**d–f**) 2D trajectories of (x, y), (y, z) and (z, x); (**g**) 3D trajectory of (x, y, z).

2. The Noisy Lorenz 96 (L-96) and Two-Layer L-96 Models

The original L-96 model with noise is given by

$$du_i = \left(u_{i-1}(u_{i+1} - u_{i-2}) - \bar{d}_i u_i + F\right) dt + \sigma_u dW_{u_i}. \tag{34}$$

The model can be regarded as a coarse discretization of atmospheric flow on a latitude circle with complicated wave-like and chaotic behavior. It schematically describes the interaction between small-scale fluctuations with larger-scale motions. However, the noisy L-96 model in (34) usually does not have the conditional Gaussian structure unless a careful selection of a subset of u_i for $\mathbf{u_{II}}$. Nevertheless, some two-layer L-96 models do belong to conditional Gaussian framework. These two-layer L96 models are conceptual models in geophysical turbulence that includes the interaction between small-scale fluctuations in the second layer with the larger-scale motions. They are widely used as a testbed for data assimilation and parameterization in numerical weather forecasting [36,40,119,120]. One natural choice of the two-layer L-96 model is the following [36]:

$$du_i = \left(u_{i-1}(u_{i+1} - u_{i-2}) + \sum_{j=1}^{J} \gamma_{i,j} u_i v_{i,j} - \bar{d}_i u_i + F\right) dt + \sigma_u dW_{u_i}, \qquad i = 1, \ldots, I,$$
$$dv_{i,j} = \left(-d_{v_{i,j}} v_{i,j} - \gamma_j u_i^2\right) dt + \sigma_{i,j} dW_{v_{i,j}}, \qquad j = 1, \ldots, J. \tag{35}$$

Linking the general conditional Gaussian framework (1) and (2) with the two-layer L-96 model in (35), $\mathbf{u_I} = \{u_i\}$ represents the large-scale motions while $\mathbf{u_{II}} = \{v_{i,j}\}$ involves the small-scale variables.

In Figures 2–4, numerical simulations of the two-layer L-96 model in (35) with $I = 40$ and $J = 5$ are shown. Here, the damping and coupling coefficients are both functions of space with

$$\bar{d}_i = 1 + 0.7 \cos(2\pi i/J), \qquad \text{and} \qquad \gamma_{i,j} := \gamma_i = 0.1 + 0.25 \cos(2\pi i/J). \tag{36}$$

These mimic the situation that the damping and coupling above the ocean are weaker than those above the land since the latter usually have stronger friction or dissipation. Therefore the model is inhomogeneous and the large-scale wave patterns over the ocean are more significant, where a westward propagation is clearly seen in all these figures (Panel (a)). The other parameters are as follows:

$$\sigma_u = 1, \qquad \sigma_{v_{i,1}} = 0.5, \qquad \sigma_{v_{i,2}} = 0.2, \qquad \sigma_{v_{i,3}} = \sigma_{v_{i,4}} = \sigma_{v_{i,5}} = 0.1.$$
$$d_{v_{i,1}} = 0.2, \qquad d_{v_{i,2}} = 0.5, \qquad d_{v_{i,3}} = 1, \qquad d_{v_{i,4}} = 2, \qquad d_{v_{i,5}} = 5, \tag{37}$$

which imply that the damping time is shorter in the smaller scales. The model in (35) has many desirable properties as in more complicated turbulent systems. Particularly, the smaller scales are more intermittent (Panel (b)) with stronger fat tails in PDFs. Different constant forcing $F = 5, 8$ and 16 are adopted in Figures 2–4, which result in various chaotic behavior for the system. With the forcing increase, the oscillation patterns in u become more regular while the small scale variables at each fixed grid point show more turbulent behavior.

Figure 2. Simulations of the two-layer L-96 model in (35) with inhomogeneous parameters (36). Here, $F = 5$. (**a**) spatio-temporal evolution of the large-scale variable u_i; (**b**) time series of u_i and $v_{i,j}$ at $i = 21$; from top to bottom: larger to smaller scales with increasing intermittency.
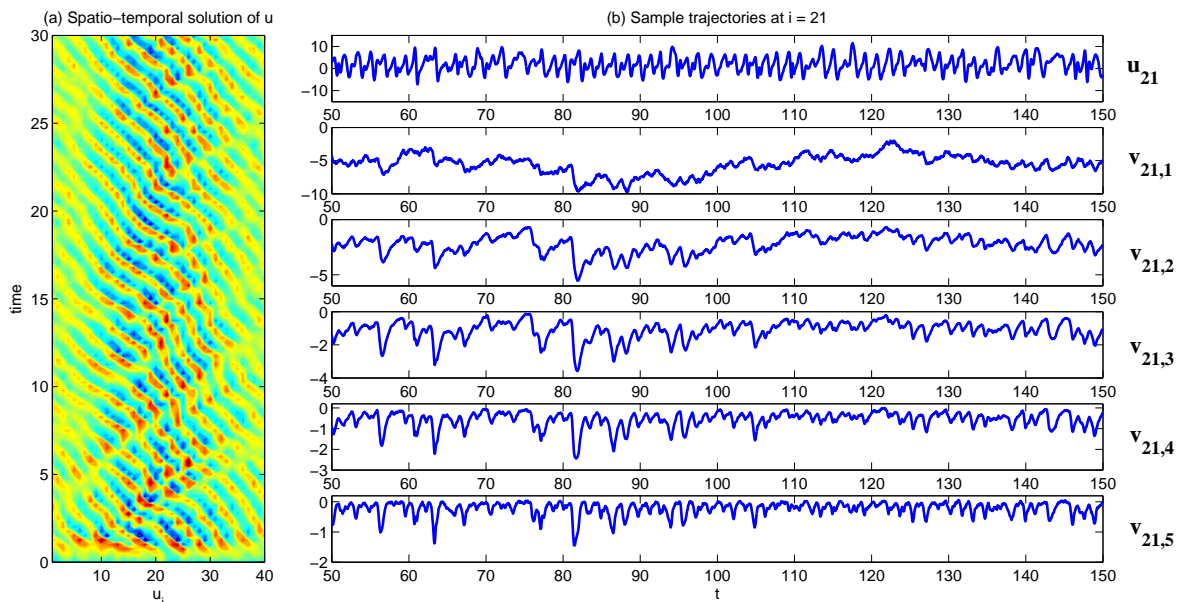
Figure 3. Simulations of the two-layer L-96 model in (35) with inhomogeneous parameters (36). Here, $F = 8$. (**a**) spatio-temporal evolution of the large-scale variable u_i; (**b**) time series of u_i and $v_{i,j}$ at $i = 21$; from top to bottom: larger to smaller scales with increasing intermittency.
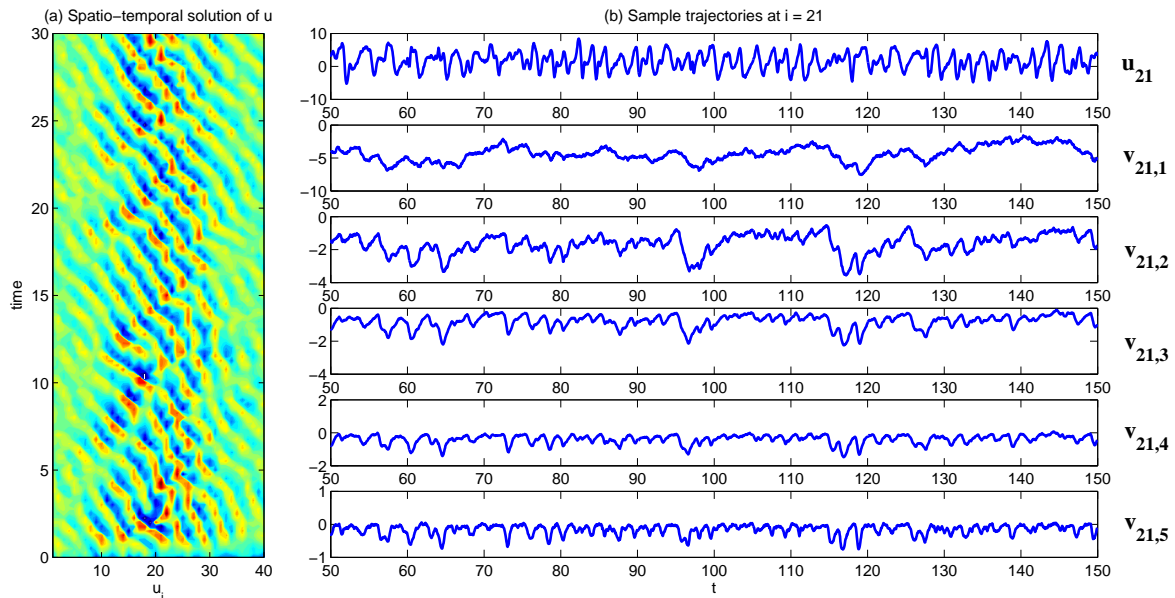
Figure 4. Simulations of the two-layer L-96 model in (35) with inhomogeneous parameters (36). Here, $F = 16$. (**a**) spatio-temporal evolution of the large-scale variable u_i; (**b**) time series of u_i and $v_{i,j}$ at $i = 21$; from top to bottom: larger to smaller scales with increasing intermittency.
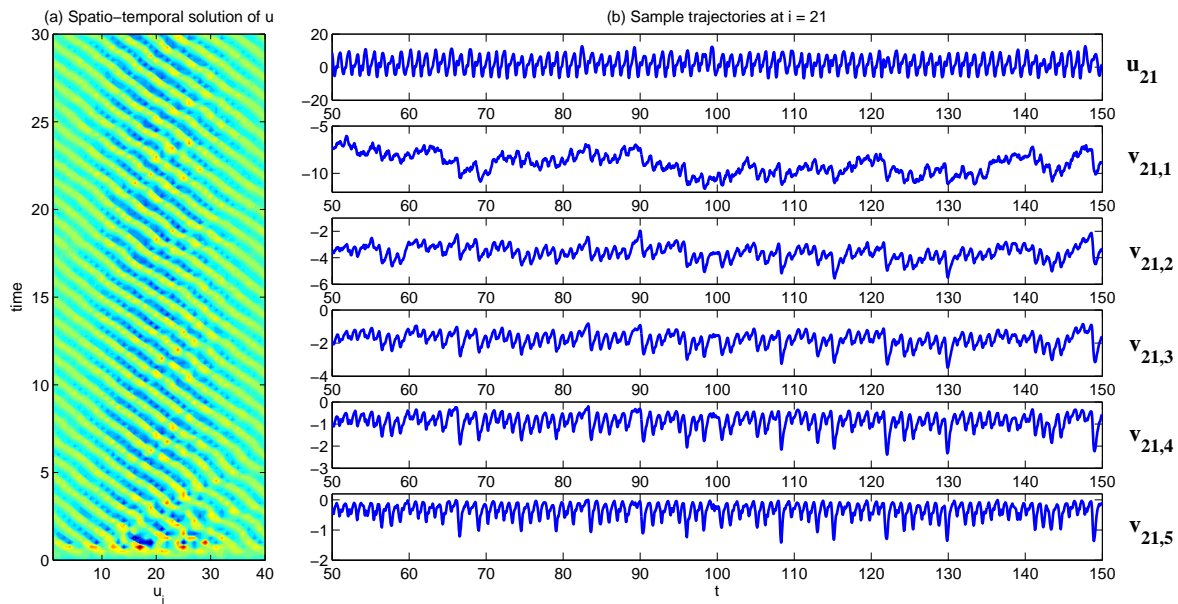
3. The Noisy Lorenz 84 (L-84) Model

The noisy L-84 model is an extremely simple analogue of the global atmospheric circulation [121,122], which has the following form [35,123]:

$$
\begin{aligned}
dx &= \big(-(y^2 + z^2) - a(x - f)\big)dt + \sigma_x dW_x, \\
dy &= \big(-bxz + xy - y + g\big)dt + \sigma_y dW_y, \\
dz &= \big(bxy + xz - z\big)dt + \sigma_z dW_z.
\end{aligned}
\tag{38}
$$

In (38), the zonal flow x represents the intensity of the mid-latitude westerly wind current (or the zonally averaged meridional temperature gradient, according to thermal wind balance), and a wave component exists with y and z representing the cosine and sine phases of a chain of vortices superimposed on the zonal flow. Relative to the zonal flow, the wave variables are scaled so that $x^2 + y^2 + z^2$ is the total scaled energy (kinetic plus potential plus internal). Note that these equations can be derived as a Galerkin truncation of the two-layer quasigeostrophic potential vorticity equations in a channel.

In the L-84 model (38), the vortices are linearly damped by viscous and thermal processes. The damping time defines the time unit and $a < 1$ is a Prandtl number. The terms xy and xz represent the amplification of the wave by interaction with the zonal flow. This occurs at the expense of the westerly current: the wave transports heat poleward, thus reducing the temperature gradient, at a rate proportional to the square of the amplitudes, as indicated by the term $-(y^2 + z^2)$. The terms $-bxz$ and bxy represent the westward (if $x > 0$) displacement of the wave by the zonal current, and $b > 1$ allows the displacement to overcome the amplification. The zonal flow is driven by the external force af which is proportional to the contrast between solar heating at low and high latitudes. A secondary forcing g affects the wave, it mimics the contrasting thermal properties of the underlying surface of zonally alternating oceans and continents. When $g = 0$ and $f < 1$, the system has a single steady solution $x = f, y = z = 0$, representing a steady Hadley circulation. This zonal flow becomes unstable for $f > 1$, forming steadily progressing vortices. For $g > 0$, the system clearly shows chaotic behavior.

Linking (38) to the general conditional Gaussian framework (1) and (2), it is clear that the zonal wind current $\mathbf{u_I} = x$ is the variable for state estimation or filtering given the two phases of the large-scale vortices $\mathbf{u_{II}} = (y, z)$.

In Figure 5, we show the sample trajectories of the system with the following parameters:

$$a = \frac{1}{4}, \qquad b = 4, \qquad f = 8, \qquad g = 1, \tag{39}$$

which were Lorenz adopted [35]. Small noise $\sigma_x = \sigma_y = \sigma_z - 0.1$ is also added to the system. It is clear that y and z are quite chaotic and they appear as a pair (Panels (b,c,f)). On the other hand, x is less turbulent and occurs in a relatively slower time scale. It plays an important role in changing both the phase and amplitude of y and z.
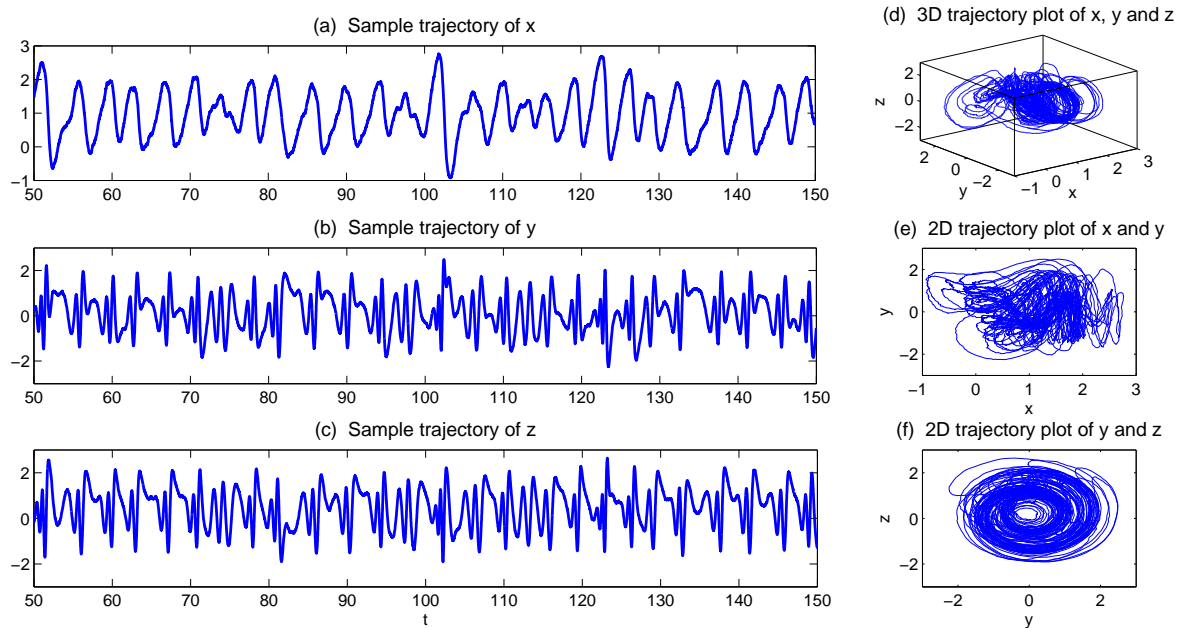
Figure 5. Simulations of the L-84 model (38) with parameters in (39). (**a–c**) sample trajectories of x, y and z, respectively; (**d**) 3D trajectory of (x, y, z); (**e**,**f**) 2D trajectories of (x, y) and (y, z).

4.1.2. Nonlinear Stochastic Models for Predicting Intermittent MJO and Monsoon Indices

Assessing the predictability limits of time series associated with the tropical intraseasonal variability such as the the Madden–Julian oscillation (MJO) and monsoon [41–43] is an important topic. They yield an interesting class of low-order turbulent dynamical systems with extreme events and intermittency. For example, Panels (c,d) in Figure 6 show the MJO time series [41], measured by outgoing longwave radiation (OLR; a proxy for convective activity) from satellite data [88]. These time series are obtained by applying a new nonlinear data analysis technique, namely nonlinear Laplacian spectrum analysis [124], to these observational data. To describe and predict such intermittent time series, the following model is developed:

$$\begin{aligned}
du_1 &= \left(-d_u(t)\, u_1 + \gamma\, v\, u_1 - \omega\, u_2\right) dt + \sigma_u\, dW_{u_1}, \\
du_2 &= \left(-d_u(t)\, u_2 + \gamma\, v\, u_2 + \omega\, u_1\right) dt + \sigma_u\, dW_{u_2}, \\
dv &= \left(-d_v\, v - \gamma\, (u_1^2 + u_2^2)\right) dt + \sigma_v\, dW_v, \\
d\omega &= \left(-d_\omega \omega + \hat{\omega}\right) dt + \sigma_\omega\, dW_\omega,
\end{aligned} \tag{40}$$

with

$$d_u(t) = d_{u0} + d_{u1} \sin(\omega_f t + \phi). \tag{41}$$

Such a model is derived based on the physics-constrained nonlinear data-driven strategy [31,32]. In this model, $\mathbf{u_I} = (u_1, u_2)$ are the two-dimensional indices obtained from observational data while $\mathbf{u_{II}} = (v, \omega)$ are the two hidden unobserved variables representing other important underlying physical processes that interact with the observational ones. The coupled system is a conditional Gaussian one, which plays an important role here since there is no direct observational data for the two hidden processes v and ω. In fact, in order to obtain the initial values of v and ω for ensemble forecast, the data assimilation formula in (4) and (5) is used given the observational trajectories of u_1 and u_2. The parameters here are estimated via information theory. With the calibrated parameters, the sample trajectories as shown in Panels (a,b) capture all the important features as in the MJO indices from observations. In addition, the non-Gaussian PDFs (Panels (e,f)) and the correlation and cross-correlation functions (Panels (g,h)) from the model nearly perfectly match those associated with the observations. In [41], significant prediction skill of these MJO indices using the physics-constrained nonlinear stochastic model (6) was shown. The prediction based on ensemble mean can have skill even up to 40 days. In addition, the ensemble spread accurately quantify the forecast uncertainty in both short and long terms. In light of a twin experiment, it was also revealed in [41] that the model in (6) is able to reach the predictability limit of the large-scale cloud patterns of the MJO.
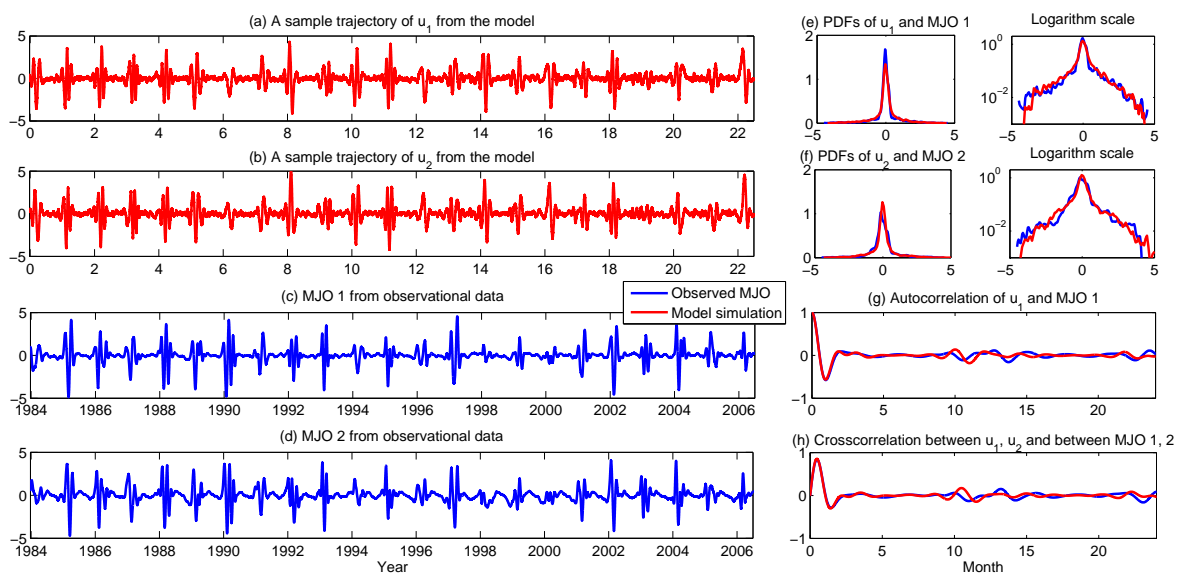
Figure 6. Using physics-constrained nonlinear low-order stochastic model (6) to capture the key features of the observed MJO indices. (**a,b**) sample trajectories from the model (6); (**c,d**) MJO indices from observations; (**e,f**) comparison of the PDFs in both linear and logarithm scales; (**g,h**) comparison of the autocorrelation and cross-correlation functions.

4.1.3. A Simple Stochastic Model with Key Features of Atmospheric Low-Frequency Variability

This simple stochastic climate model [49,125] presented below is set-up in such a way that it features many of the important dynamical features of comprehensive global circulation models (GCMs) but with many fewer degree of freedom. Such simple toy models allow the efficient exploration of the whole parameter space that is impossible to conduct with GCMs:

$$dx_1 = \left(-x_2(L_{12} + a_1 x_1 + a_2 x_2) + d_1 x_1 + F_1 + L_{13}y_1 + b_{123}x_2 y_1\right)dt + \sigma_{x_1}dW_{x_1},$$

$$dx_2 = \left(+x_1(L_{12} + a_1 x_1 + a_2 x_2) + d_2 x_2 + F_2 + L_{24}y_2 + b_{213}x_1 y_1\right)dt + \sigma_{x_2}dW_{x_1},$$

$$dy_1 = \left(-L_{13}x_1 + b_{312}x_1 x_2 + F_3 - \frac{\gamma_1}{\epsilon}y_1\right)dt + \frac{\sigma_{y_1}}{\sqrt{\epsilon}}dW_{y_1}.$$

$$dy_2 = \left(-L_{24}x_2 + F_4 - \frac{\gamma_2}{\epsilon}y_2\right)dt + \frac{\sigma_{y_2}}{\sqrt{\epsilon}}dW_{y_2},$$

$$(42)$$

where $b_{123} + b_{213} + b_{312} = 0$. It contains a quadratic nonlinear part that conserves energy as well as a linear operator. Therefore, this model belongs to physics-constrained nonlinear stochastic model family. The linear operator includes a skew-symmetric part that mimics the Coriolis effect and topographic Rossby wave propagation, and a negative definite symmetric part that is formally similar to the dissipation such as the surface drag and radiative damping. The two variables x_1 and x_2 can be regarded as climate variables while the other two variables y_1 and y_2 become weather variables that occur in a much faster time scale when ϵ is small. In fact, the MTV strategy as described in Section 3.4 is applied to y_1 and y_2 that introduce this ϵ together with stochastic noise and damping terms. The coupling in different variables is through both linear and nonlinear terms, where the nonlinear coupling through b_{ijk} produces multiplicative noise. Note that when $\epsilon \to 0$, applying an explicit stochastic mode reduction results in a two-dimensional system for the climate variables [45,46]. Clearly, the 4D stochastic climate model (42) is a conditional Gaussian system with $\mathbf{u_I} = (x_1, x_2)^T$ and $\mathbf{u_{II}} = (y_1, y_2)^T$.

In Figure 7, sample trajectories and the associated PDFs in two dynamical regimes are shown. The two regimes differ by the scale separation ϵ with $\epsilon = 1$ (Regime I) and $\epsilon = 0.1$ (Regime II), respectively. The other parameters are the same in the two regimes:

$$L_{12} = 1, \quad L_{13} = 0.5, \quad L_{24} = 0.5, \quad a_1 = 2, \quad a_2 = 1, \quad d_1 = -1, \quad d_2 = -0.4,$$
$$\sigma_1 = 0.5, \quad \sigma_2 = 2, \quad \sigma_3 = 0.5, \quad \sigma_4 = 1, \quad b_{123} = 1.5, \quad b_{213} = 1.5, \quad b_{312} = -3, \quad (43)$$
$$F_1 = F_2 = F_3 = F_4 = 0.$$

It is obvious that with $\epsilon = 1$, all the four variables lie in the same time scale. Both "climate variable" x_1 and "weather variable" y_1 can have intermittent behavior with non-Gaussian PDFs. On the other hand, with $\epsilon = 0.1$, a clear scale separation can be seen in the time series, where y_1 and y_2 occur in a much faster time scale than x_1 and x_2. Since the memory time due to the strong damping becomes much shorter in y_1 and y_2, the associated PDFs for these "weather variables" become Gaussian.

4.1.4. A Nonlinear Triad Model with Multiscale Features

The following nonlinear triad system is a simple prototype nonlinear stochastic model that mimics structural features of low-frequency variability of GCMs with non-Gaussian features [50] and it was used to test the skill for reduced nonlinear stochastic models for the fluctuation dissipation theorem [126]:

$$du_1 = \left(-\gamma_1 u_1 + L_{12}u_2 + L_{13}u_3 + Iu_1 u_2 + F(t)\right)dt + \sigma_1 dW_1, \tag{44}$$

$$du_2 = \left(-L_{12}u_1 - \frac{\gamma_2}{\epsilon}u_2 + L_{23}u_3 - Iu_1^2\right)dt + \frac{\sigma_2}{\epsilon^{1/2}}dW_2, \tag{45}$$

$$du_3 = \left(-L_{13}u_1 - L_{23}u_2 - \frac{\gamma_3}{\epsilon}u_3\right)dt + \frac{\sigma_3}{\epsilon^{1/2}}dW_3. \tag{46}$$

The triad model (44)–(46) involves a quadratic nonlinear interaction between u_1 and u_2 with energy-conserving property that induces intermittent instability. On the other hand, the coupling between u_2 and u_3 is linear and is through the skew-symmetric term with coefficient $-L_{23}$, which represents an oscillation structure of u_2 and u_3. Particularly, when L_{23} is large, fast oscillations become dominant for u_2 and u_3 while the overall evolution of u_1 can still be slow provided that the

feedback from u_2 and u_3 is damped quickly. Such multiscale structure appears in the turbulent ocean flows described for example by shallow water equation, where u_1 stands for the geostrophically balanced part while u_2 and u_3 mimics the fast oscillations due to the gravity waves [11,122]. The large-scale forcing $F(t)$ represents the external time-periodic input to the system, such as the seasonal effects or decadal oscillations in a long time scale [121,127]. In addition, the scaling factor ϵ plays the same role as in the 4D stochastic climate model (42) that allows a difference in the memory of the three variables. Note that the MTV strategy as described in Section 3.4 is applied to u_2 and u_3 that introduces the factor ϵ. The nonlinear triad model in (44)–(46) belongs to conditional Gaussian model family with $\mathbf{u_I} = (x_1, x_2)^T$ and $\mathbf{u_{II}} = (y_1, y_2)^T$.
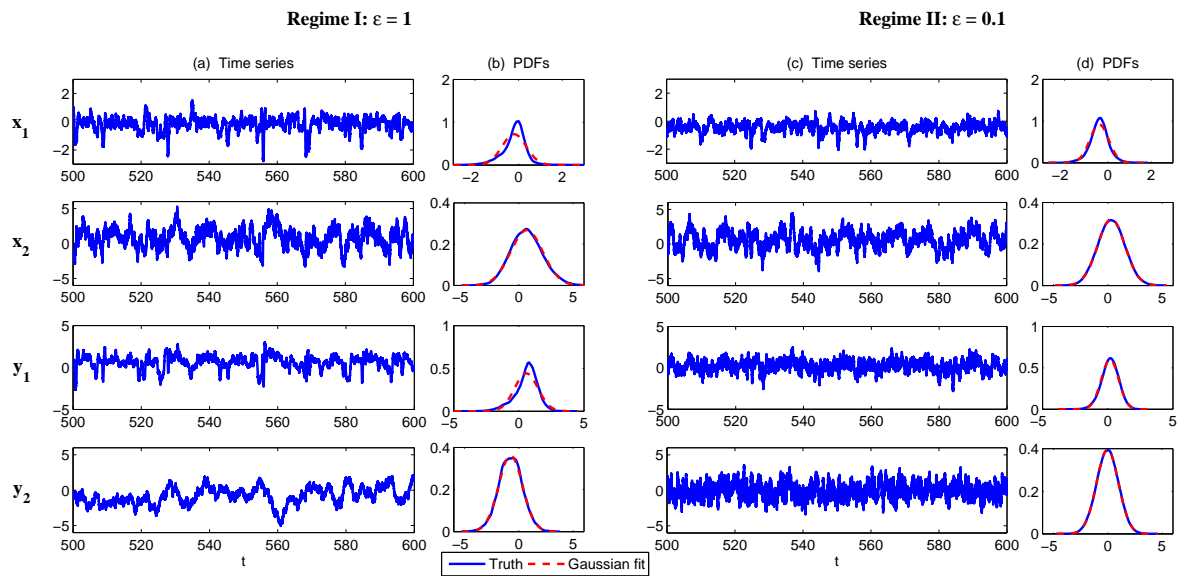
Figure 7. Simulations of the simple stochastic climate model (42) with parameters given by (43). (**Left**) Regime I with $\epsilon = 1$. (**Right**) Regime II with $\epsilon = 0.1$.

To understand the dynamical behavior of the nonlinear triad model (44)–(46), we show the model simulations in Figure 8 for the following two regimes:

$$
\begin{aligned}
\text{Regime I}: \quad & \epsilon = 1.0, \quad I = 1, \quad \sigma_1 = 0.5, \quad L_{12} = 0.4, \quad L_{13} = 0.4, \quad L_{23} = 0, \\
& \gamma_1 = 2.5, \quad \gamma_2 = 0.4, \quad \gamma_3 = 0.4, \quad \sigma_2 = 1.2, \quad \sigma_3 = 0.8, \quad F = 2, \\
\text{Regime II}: \quad & \epsilon = 0.1, \quad I = 5, \quad \sigma_1 = 0.5, \quad L_{12} = 1.0, \quad L_{13} = 1.0, \quad L_{23} = 10, \\
& \gamma_1 = 2.0, \quad \gamma_2 = 0.1, \quad \gamma_3 = 0.1, \quad \sigma_2 = 1.2, \quad \sigma_3 = 0.8, \quad F = 2.
\end{aligned}
\tag{47}
$$

In Regime I, $\epsilon = 1$ and therefore u_1, u_2 and u_3 occur in the same time scale. Since a large noise coefficient $\sigma_1 = 2.5$ is adopted in the dynamics of u_1, the PDF of u_1 is nearly Gaussian and the variance is large. The latter leads to a skewed PDF of u_2 due to the nonlinear feedback term $-Iu_1^2$, where the extreme events are mostly towards the negative side of u_2 due to the negative sign in front of the nonlinear term. As a consequence, the u_3 also has a skewed PDF but the skewness is positive. On the other hand, in Regime I, where $\epsilon = 0.1$, u_2 and u_3 both have a short decorrelation time and the associated PDFs are nearly Gaussian. Nevertheless, the slow variable u_1 is highly non-Gaussian due to the direct nonlinear interaction between u_2 in which u_2 plays the role of stochastic damping and it results in the intermittent instability in the signal of u_1.
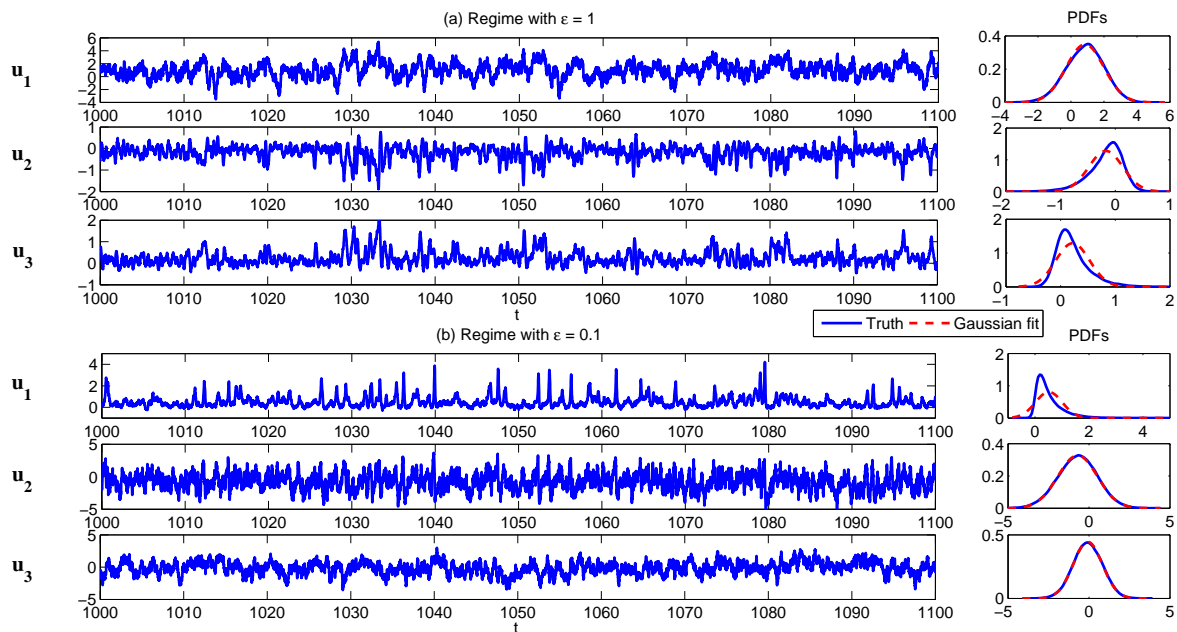
Figure 8. Model simulation of the triad model (44)–(46) with parameters in (47). (**a**) Regime I with $\epsilon = 1$; (**b**) Regime II with $\epsilon = 0.1$.

4.1.5. Conceptual Models for Turbulent Dynamical Systems

Understanding the complexity of anisotropic turbulence processes over a wide range of spatiotemporal scales in engineering shear turbulence [128–130] as well as climate atmosphere ocean science [73,121,122] is a grand challenge of contemporary science. This is especially important from a practical viewpoint because energy often flows intermittently from the smaller scales to affect the largest scales in such anisotropic turbulent flows. The typical features of such anisotropic turbulent flows are the following: (A) The large-scale mean flow is usually chaotic but more predictable than the smaller-scale fluctuations. The overall single point PDF of the flow field is nearly Gaussian whereas the mean flow pdf is sub-Gaussian, in other words, with less extreme variability than a Gaussian random variable; (B) There are nontrivial nonlinear interactions between the large-scale mean flow and the smaller-scale fluctuations which conserve energy; (C) There is a wide range of spatial scales for the fluctuations with features where the large-scale components of the fluctuations contain more energy than the smaller-scale components. Furthermore, these large-scale fluctuating components decorrelate faster in time than the mean-flow fluctuations on the largest scales, whereas the smaller-scale fluctuating components decorrelate faster in time than the larger-scale fluctuating components; (D) The PDFs of the larger-scale fluctuating components of the turbulent field are nearly Gaussian, whereas the smaller-scale fluctuating components are intermittent and have fat-tailed PDFs, in other words, a much higher probability of extreme events than a Gaussian distribution.

Denote u the largest variable and $\{v_k\}$ a family of small-scale variables. One can think of u as the large-scale spatial average of the turbulent dynamics at a single grid point in a more complex system and $\{v_k\}$ as the turbulent fluctuations at the grid point. To add a sense of spatial scale, one can also regard v_k as the amplitude of the k-th Fourier cosine mode evaluated at a grid point. A hallmark of turbulence is that the large scales can destabilize the smaller scales in the turbulent fluctuations intermittently and this increased small-scale energy can impact the large scales; this key feature is captured in the conceptual models. With the above discussion, here are the simplest models with all these features, the conceptual dynamical models for turbulence [37]:

$$du = \left(-d_u u + \gamma \sum_{k=1}^{K} v_k^2 + F\right) dt,$$

$$dv_k = \left(-d_{v_k} v_k - \gamma u v_k\right) dt + \sigma_{v_k} dW_{v_k}, \qquad 1 \le k \le K. \tag{48}$$

Now, let us take $K = 5$ and use the following parameters which have been tested in [37] that represent the features of turbulent flows:

$$d_u = 0.01, \qquad d_{v_k} = 1 + 0.02k^2, \qquad \frac{\sigma_{v_k}^2}{d_{v_k}} = \frac{0.004}{(1+k)^{5/3}}, \qquad \gamma = 1.5, \qquad F = -0.15. \tag{49}$$

The sample trajectories and the associated PDFs are shown in Figure 9. The largest scale u is nearly Gaussian while more intermittent behavior is observed in the smaller scale variables. Here, the nonlinear interaction plays an important role in generating these turbulent features and the total energy in the nonlinear terms is conserved. Thus, all the features (A)–(D) above are addressed in this model. It is also clear that the conceptual turbulent model (48) is a conditional Gaussian system with $\mathbf{u_I} = (v_1, \ldots, v_5)^T$, $\mathbf{u_{II}} = u$.
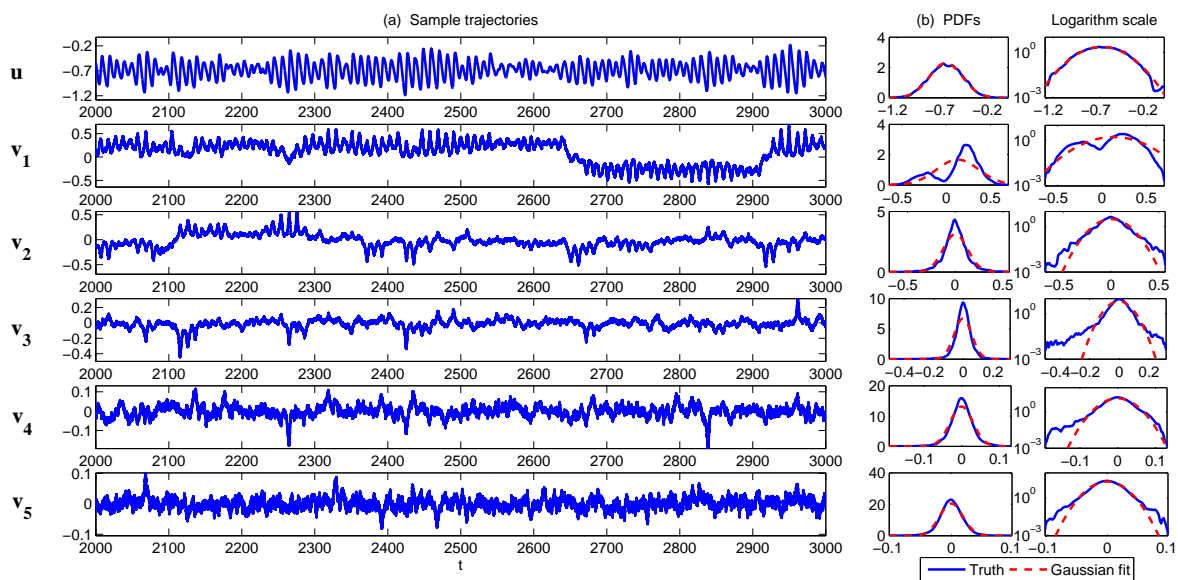
Figure 9. Sample trajectories (**a**) and the associated PDFs (**b**) of the conceptual turbulent model (48) with parameters in (49).

In [36], a modified version of the conceptual model was developed:

$$du = \left(-d_u u + \sum_{k}^{K} \gamma_k u v_k + F\right) dt + \sigma_u dW_u, \tag{50}$$

$$dv_k = \left(-d_{v_k} v_k - \gamma_k u^2\right) dt + \sigma_{v_k} dW_{v_k}, \qquad k = 1, \ldots, K. \tag{51}$$

This modified conceptual turbulent model again fits into the conditional Gaussian framework, where $\mathbf{u_I} = u$ includes the largest scale variable and $\mathbf{u_{II}} = (v_1, \ldots, v_K)^T$ represents small-scale ones. The modified conceptual turbulent model (50) and (51) inherits many important features from the dynamics in (48). For example, with suitable choices of parameters, the large-scale observed variable u is nearly Gaussian while small-scale variables v_k becomes more intermittent with the increasing of k. In addition, the small-scale turbulent flows provide feedback to large scales through the nonlinear

coupling with energy-conserving property. An example is shown in Figure 10 with the following choice of parameters:

$$d_u = 0.1, \quad F = 0.5, \quad \sigma_u = 2, \quad \gamma_k = 0.25,$$
$$d_{v_k} = \{0.2, 0.5, 1, 2, 5\}, \quad \sigma_{v_k} = \{0.5, 0.2, 0.1, 0.1, 0.1\}, \quad k = 1, \ldots, 5. \tag{52}$$
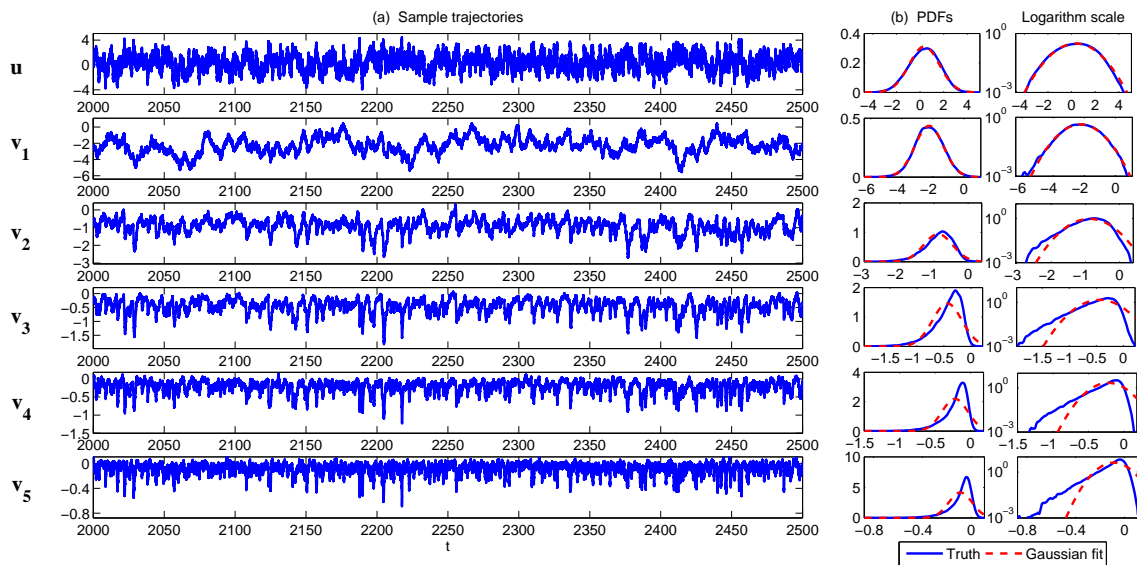
Figure 10. Sample trajectories (**a**) and the associated PDFs (**b**) of the modified conceptual turbulent model (50) and (51) with parameters in (52).

4.1.6. A Conceptual Model of the Coupled Atmosphere and Ocean

In [39], the signatures of feedback between the atmosphere and the ocean are studied with a simple coupled model, which can be used to exam the effects of oceanic variability and seasonality.

The atmosphere component is the Lorenz 1984 model (38) discussed in Section 4.1.1, except that the forcing has an explicit form

$$f(t) = f_0 + f_1 \cos \omega t,$$

which represents seasonal effect. Therefore, the atmosphere model reads

$$
\begin{aligned}
dx &= \big(-(y^2 + z^2) - a(x - f(t))\big)dt + \sigma_x dW_x, \\
dy &= \big(-bxz + xy - y + g\big)dt + \sigma_y dW_y, \\
dz &= \big(bxy + xz - z\big)dt + \sigma_z dW_z.
\end{aligned}
\tag{53}
$$

Again, x represents the amplitude of the zonally averaged meridional temperature gradient while y and z denote the amplitudes of the cosine and sine longitudinal phases of a chain of large-scale waves. The poleward heat transport is achieved by the eddies at a rate proportional to $y^2 + z^2$, and this heat transport reduces the zonally averaged temperature gradient. The term $f(t)$ represents the zonally averaged forcing due to the equator-pole difference in solar heating and it varies on a seasonal timescale.

On the other hand, the oceanic module simulates the wind-driven circulation in a basin that occupies a fraction r of the longitudinal extent of the atmosphere. Its dynamics are described by a set of four ordinary differential equations, namely,

$$
\begin{aligned}
dp &= -(\psi_r^2 + \psi_i^2)p\,dt, \\
dq &= 0, \\
d\psi_r &= (-\sigma\psi_r - \Omega\psi_i)dt, \\
d\psi_i &= (\Omega\psi_r - \sigma\psi_i)dt.
\end{aligned}
\tag{54}
$$

Here, p represents the zonally averaged meridional temperature gradient at the sea surface, while q represents the basin-averaged sea surface temperature. The poleward heat transport is achieved by a large-scale flow, at a rate proportional to $\psi_r^2 + \psi_i^2$. The average temperature Q is conserved in the absence of any coupling with the atmosphere. The transport is represented by two phases of the streamfunction, ψ_r and ψ_i. The streamfunction undergoes damped oscillations with a period, $2\pi/\Omega$, of 5.3 y and a decay time, σ^{-1}, of 17 y. This damped oscillation is the only source of internal variability in the ocean and is due to the intrinsic decadal variability of the wind-driven circulation. Note that the equations for the two phases of the streamfunction can be derived as a Galerkin truncation of the one-and-a-half-layer quasigeostrophic potential vorticity equation for long linear Rossby waves. It is suggested in [131,132] that basin modes with decadal frequencies can be excited by stochastic atmospheric forcing and represent a resonant response of the ocean. This model essentially assumes that the intrinsic decadal variability of the ocean wind-driven circulation is described by one such mode. For weak flow, the wind-driven gyres reduce the zonally averaged north-south temperature gradient in a basin.

The feedback between the ocean and the atmosphere are constructed so as to conserve total heat in the air–sea exchange. The coupled atmosphere-ocean model reads:

$$
\begin{aligned}
dx &= \left(-(y^2 + z^2) - a(x - f(t)) + rm(p - x - \gamma)\right)dt + \sigma_x dW_x, \\
dy &= \left(-bxz + xy - y + g + rm(q - y)\right)dt + \sigma_y dW_y, \\
dz &= \left(bxy + xz - z\right)dt + \sigma_z dW_z, \\
dp &= \left(-np + mc^{-1}(x - p + \gamma)\right)dt + \sigma_p dW_p, \\
dq &= mc^{-1}(y - q) + \sigma_q dW_q, \\
d\psi_r &= \left(-\sigma\psi_r - \Omega\psi_i + \alpha_r x + \beta_r y\right)dt + \sigma_{\phi_r} dW_{\phi_r}, \\
d\psi_i &= \left(\Omega\psi_r - \sigma\psi_i + \alpha_i x + \beta_i y\right)dt + \sigma_{\phi_i} dW_{\phi_i},
\end{aligned}
\tag{55}
$$

where the terms with underlines represent the coupling between atmosphere and ocean. In (55), the term $\psi_r^2 + \psi_i^2$ on the right-hand side of the evolution of p is parameterized by a constant n such that the ocean part become a linear model. Therefore, (55) includes a nonlinear atmosphere model and a linear ocean model. The coupling is through linear terms. Thus, treating the atmosphere variables as $\mathbf{u_I} = (x, y, z)$ and the ocean variables as $\mathbf{u_{II}} = (p, q, \psi_r, \psi_i)$, the coupled system (55) belongs to conditional Gaussian framework. It is also obvious that the coupled model belongs to the physics-constrained regression model family.

In (55), the air–sea heat fluxes are proportional to the difference between the oceanic and the atmospheric temperature: these are the terms $m(p - x)$ and $m(q - y)$. The bulk transfer coefficient, m, is assumed to be constant. In the atmospheric model, this term needs to be multiplied by the fraction of earth covered by ocean, r. In the oceanic model, the air–sea flux is divided by c, which is the ratio of the vertically integrated heat capacities of the atmosphere and the ocean. The constant γ represents the fraction of solar radiation that is absorbed directly by the ocean. There is no heat exchange between the atmospheric standing wave z and the ocean because z represents the sine phase of the longitudinal eddies and has zero zonal mean across the ocean. A feedback between z and the ocean would appear if we added an equation for the longitudinal temperature gradient of the ocean. The effect of the wind stress acting on the ocean is represented as a linear forcing proportional to x and y in the equations for the streamfunction, ψ. The coupling constants α_r, α_i, β_r, and β_i are chosen to produce realistic values for the oceanic heat transport. Detailed discussions of the parameter choices and model behaviors in different dynamical regimes are shown in [39].

4.1.7. A Low-Order Model of Charney–DeVore Flows

The concept of multiple equilibria in a severely truncated "low-order" image of the atmospheric circulation was proposed by Charney and DeVore [133] (the CdV model). The simplest CdV model describes a barotropic zonally unbounded flow over a sinusoidal topography in a zonal channel with quasigeostrophic dynamics. The vorticity balance of such a flow

$$\frac{\partial}{\partial t}\nabla^2\Psi + \mathbf{u}\cdot\nabla\left[\nabla^2\Psi + \beta y + \frac{f_0 b}{H}\right] = R\nabla^2(\Psi^* - \Psi) \tag{56}$$

needs an additional constraint to determine the boundary values of the streamfunction Ψ on the channel walls. The vorticity concept has eliminated the pressure field and its reconstruction in a multiconnected domain requires the validity of the momentum balance, integrated over the whole domain,

$$\frac{\partial U}{\partial t} = R(U^* - U) + \frac{f_0}{H}\left\langle b\frac{\partial\Psi}{\partial x}\right\rangle. \tag{57}$$

Here, U is the zonally and meridionally averaged zonal velocity and $R\nabla^2\Psi^* = -R\partial U^*/\partial y$ is the vorticity and RU^* the zonal momentum imparted into the system.

The depth of the fluid is $H - b$ and the topography height b is taken sinusoidal, $b = b_0\cos Kx\sin Ky$ with $K = 2\pi n/L$, where L is the length and $L/2$ the width of the channel. A heavily truncated expansion

$$\Psi = -Uy + \frac{1}{K}[A\cos Kx + B\sin Kx]\sin Ky$$

represents the flow in terms of the zonal mean U and a wave component with sine and cosine amplitudes A and B. It yields the low-order model [29,133]:

$$
\begin{aligned}
dU &= \left(R(U^* - U) + \frac{1}{4}\delta B\right)dt + \sigma_U dW_U,\\
dA &= \left(-KB(U - c_R) - RA\right)dt + \sigma_A dW_A,\\
dB &= \left(KA(U - c_R) - \frac{1}{2}\delta U - RB\right)dt + \sigma_B dW_B,
\end{aligned}
\tag{58}
$$

where $c_R = \beta/2K^2$ is the barotropic Rossby wave speed and $\delta = f_0 b_0/H$. Apparently, (58) belongs to the conditional Gaussian framework when the zonal mean flow is treated as $\mathbf{u_I} = U$ while the two wave components belong to $\mathbf{u_{II}} = (A, B)^T$. This model also belongs to physics-constrained nonlinear modeling family.

Without the stochastic noise, three equilibria are found if U^* is well above c_R. The three possible steady states can be classified according to the size of the mean flow U compared to the wave amplitudes: (a) The high zonal index regime is frictionally controlled, the flow is intense and the wave amplitude is low; (b) The low zonal index regime is controlled by form stress, the mean flow is weak and the wave is intense; (c) The intermediate state is transitional, it is actually unstable to perturbations. This "form stress instability" works obviously when the slope of the resonance curve is below the one associated with friction, i.e., $\partial(RU - \frac{1}{4}\delta B[U])/\partial U > 0$, so that a perturbation must run away from the steady state.

4.1.8. A Paradigm Model for Topographic Mean Flow Interaction

A barotropic quasi-geostrophic equation with a large scale zonal mean flow can be used to solve topographic mean flow interaction. The full equations are given as follows [2]:

$$\frac{\partial q}{\partial t} = -\nabla^{\perp}\psi \cdot \nabla q - u(t)\frac{\partial q}{\partial x} - \beta\frac{\partial \psi}{\partial x},$$
$$q = \Delta\psi + h, \tag{59}$$
$$\frac{du}{dt} = \frac{1}{4\pi^2}\int h\frac{\partial \psi}{\partial x}.$$

Here, q and ψ are the small-scale potential vorticity and stream function, respectively. The large scale zonal mean flow is characterized by $u(t)$ and the topography is defined by function $h = h(x, y)$. The parameter β is associated with the β-plane approximation to the Coriolis force. The domain considered here is a periodic box $[0, 2\pi]^2$.

Now, we construct a set of special solutions to (59), which inherit the nonlinear coupling of the small-scale flow with the large-scale mean flow via topographic stress. Consider the following Fourier decomposition:

$$\psi(\vec{x}, t) = \sum_{k \neq 0}\psi_k(t)e^{ik\vec{l}\cdot\vec{x}}, \qquad h(\vec{x}) = \sum_{k \neq 0}h_k e^{ik\vec{l}\cdot\vec{x}}, \tag{60}$$

where $\vec{l} = (l_x, l_y)$ and $\vec{x} = (x, y)$. By assuming $h_0 = 0$ and inserting (60) into (59), we arrive at the layered topographic equations in Fourier form:

$$\frac{d\psi_k}{dt} = ikl_x\left(\frac{\beta}{k^2|\vec{l}|^2} - u\right)\psi_k + i\frac{kl_x}{k^2|\vec{l}|^2}h_k u,$$
$$\frac{du}{dt} = -il_x\sum_{k \neq 0}kh_k\psi_k^*,$$

where $\psi_k^* = \psi_{-k}$ and $h_k^* = h_{-k}$. Consider a finite Fourier truncation. By adding extra damping and stochastic noise for compensating the information truncated in the finite Fourier decomposition in ψ_k,

$$d\psi_k = \left(-d_k\psi_k + ikl_x\left(\frac{\beta}{k^2|\vec{l}|^2} - u\right)\psi_k + i\frac{kl_x}{k^2|\vec{l}|^2}h_k u\right)dt + \sigma_k dW_k, \qquad |k| \leq K,$$
$$du = \left(-il_x\sum_{k \neq 0}kh_k\psi_k^*\right)dt.$$

We arrive at a conditional Gaussian system with $\mathbf{u_I} = \{\psi_k\}$ and $\mathbf{u_{II}} = u$.

In [2], very rich chaotic dynamics are reached with two-layer topographic models. Here as a concrete example, we consider a single layer topography in the zonal direction $\vec{l} = (1, 0)$ [32],

$$\psi(x, t) = \psi_1(t)e^{ix} + \psi_{-1}(t)e^{-ix},$$
$$h(x) = H\Big(\cos(x) + \sin(x)\Big) = h_1 e^{-ix} + h_{-1}e^{-ix},$$

where $h_1 = H/2 - H/2i$ and H denotes the topography amplitude. Choose

$$\psi_1(t) = \frac{1}{2}\big(b(t) - ia(t)\big) \qquad \text{and} \qquad \psi_{-1} = \psi_1^*$$

and rotate the variables (a, b) counterclockwise by $45°$ to coordinate (v_1, v_2). We arrive at

$$du = \omega v_1 dt + \sigma_u dW_u,$$
$$dv_1 = \big(-2\omega u - \beta v_2 + uv_2\big)dt, \tag{61}$$
$$dv_2 = (\beta v_1 - uv_1)dt,$$

where $\omega = H/\sqrt{2}$. This model is similar to the Charney and Devore model for nonlinear regime behavior without dissipation and forcing. Then, with additional damping and noise in v_1 and v_2 approximating the interaction with the truncated Rossby wave modes, we have the following system:

$$
\begin{aligned}
du &= \omega v_1 dt, \\
dv_1 &= \left(-d_{v_1} v_1 - 2\omega u - \beta v_2 + uv_2 \right) dt + \sigma_{v_1} dW_{v_1}, \\
dv_2 &= (-d_{v_2} v_2 + \beta v_1 - uv_1) dt + \sigma_{v_2} dW_{v_2}.
\end{aligned}
\tag{62}
$$

Linear stability is satisfied for v_1, v_2 while there is only neutral stability of u. The system in (62) satisfies the conditional Gaussian framework where $\mathbf{u_I} = (v_1, v_2)^T$ and $\mathbf{u_{II}} = u$. Notably, this model also belongs to the physics-constrained nonlinear model family. One interesting property of this model is that, if the invariant measure exists, then, despite the nonlinear terms, the invariant measure is Gaussian. The validation of this argument can be easily reached following [31]. A numerical illustration is shown Figure 11 with the following parameters:

$$
\omega = 1, \qquad d_{v_1} = d_{v_2} = 1, \qquad \beta = 0.5, \qquad \sigma_{v_1} = 0.2, \qquad \sigma_{v_2} = 0.2.
\tag{63}
$$

Note that, if stochastic noise is also added in the evolution equation of u, then the system (62) also belongs to the conditional Gaussian model family with $\mathbf{u_I} = u$ and $\mathbf{u_{II}} = (v_1, v_2)^T$.
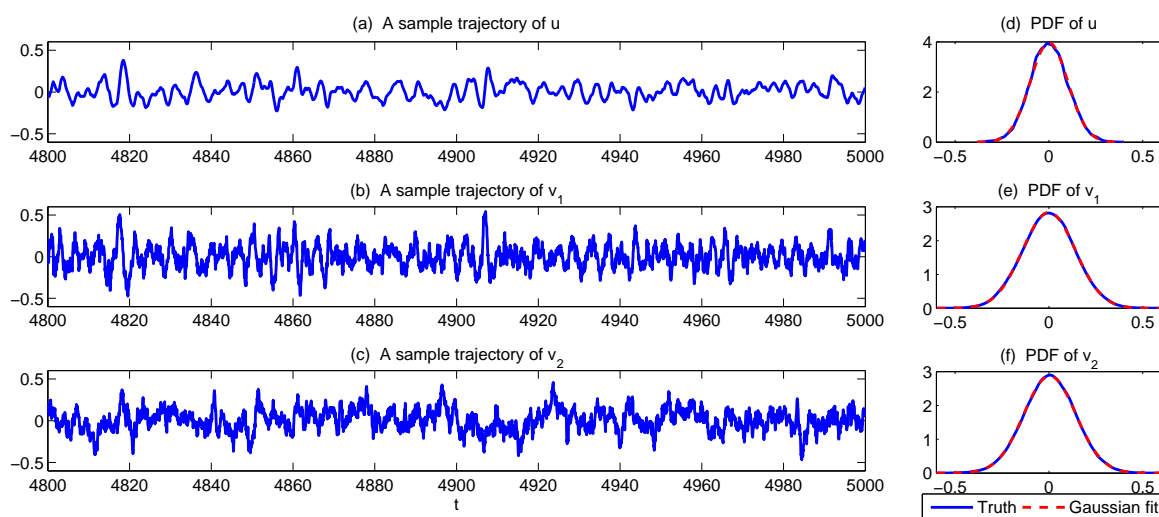
Figure 11. Sample trajectories (**a**–**c**) and the associated PDFs (**d**–**f**) of the single layer topography model (62) with parameters in (63). Note that the invariant measure is Gaussian.

4.2. Stochastically Coupled Reaction–Diffusion Models in Neuroscience and Ecology

4.2.1. Stochastically Coupled FitzHugh–Nagumo (FHN) Models

The FitzHugh–Nagumo model (FHN) is a prototype of an excitable system, which describes the activation and deactivation dynamics of a spiking neuron [57]. Stochastic versions of the FHN model with the notion of noise-induced limit cycles were widely studied and applied in the context of stochastic resonance [134–137]. Furthermore, its spatially extended version has also attracted much attention as a noisy excitable medium [138–141].

One common representation of the stochastic FHN model is given by

$$
\begin{aligned}
\epsilon du &= \left(f_1(u) + c_1 v \right) dt + \sqrt{\epsilon} \delta_1 dW_u, \\
dv &= \left(f_2(u) + c_2 v + s(t) \right) dt + \delta_2 dW_v,
\end{aligned}
\tag{64}
$$

where the time scale ratio ϵ is much smaller than one (e.g., $\epsilon \approx 10^{-2}$), implying that $u(t)$ is the fast and $v(t)$ is the slow variable. The coupled FHN system in (64) is obviously a conditional Gaussian system with $\mathbf{u_I} = u$ and $\mathbf{u_{II}} = v$. The nonlinear function $f_1(u)$ is one of the nullclines of the deterministic system. A common choice for this function is

$$f_1(u) = u - cu^3, \tag{65}$$

where the parameter c is either 1 or 1/3. On the other hand, $f_2(u)$ is usually a linear function of u. In (64), $s(t)$ is an external source and it can be a time-dependent function. In the following, we set $s(t)$ to be a constant external forcing for simplicity. Diffusion term is typically imposed in the dynamics of u. Thus, with these choices, a simple stochastically coupled spatial-extended FHN model is given by

$$
\begin{aligned}
\epsilon du &= \left(d_u \nabla^2 u + u - \frac{1}{3}u^3 - v\right) dt + \sqrt{\epsilon}\delta_1 dW_u, \\
dv &= (u + a) dt + \delta_2 dW_v.
\end{aligned}
\tag{66}
$$

With δ_1 and δ_2, the model in (66) contains the model families with both coherence resonance and self-induced stochastic resonance [142]. Applying a finite difference discretization to the diffusion term in (66), we arrive at the FHN model in the lattice form

$$
\begin{aligned}
\epsilon du_i &= \left(d_u(u_{i+1} + u_{i-1} - 2u_i) + u_i - \frac{1}{3}u_i^3 - v_i\right) dt + \sqrt{\epsilon}\delta_1 dW_{u_i}, \\
dv_i &= (u_i + a) dt + \delta_2 dW_{v_i}, \qquad i = 1, \ldots, N.
\end{aligned}
\tag{67}
$$

Note that the parameter $a > 1$ is required in order to guarantee that the system has a global attractor in the absence of noise and diffusion. The random noise is able to drive the system above the threshold level of global stability and triggers limit cycles intermittently. The model behavior of (67) in various dynamical regimes has been studied in [36]. The model can show both strongly coherent and irregular features as well as scale-invariant structure with different choices of noise and diffusion coefficients.

There are several other modified versions of the FHN model that are widely used in applications. One that appears in the thermodynamic limit of an infinitely large ensemble is the so-called globally coupled FHN model

$$
\begin{aligned}
\epsilon du_i &= \left(d_u(u_{i+1} + u_{i-1} - 2u_i) + u_i - \frac{1}{3}u_i^3 + m(\bar{u} - u_i) - v_i\right) dt + \sqrt{\epsilon}\delta_1 dW_{u_i}, \\
dv_i &= (u_i + a) dt + \delta_2 dW_{v_i}, \qquad i = 1, \ldots, N,
\end{aligned}
\tag{68}
$$

where $\bar{u} = \frac{1}{N}\sum_{i=1}^{N} u_i$. Different from the model in (67) where each u_i is only directly coupled to its two nearest neighbors u_{i-1} and u_{i+1}, each u_i in the globally coupled FHN model (68) is directly affected by all $u_j, j = 1, \ldots, N$. In [57], various closure methods are developed to solve the time evolution of the statistics associated with the globally coupled FHN model (68).

Another important modification of the FHN model is to include the colored noise into the system as suggested by [57,143]. For example, the constant a on the right-hand side of (67) can be replaced by an OU process [143] that allows the memory of the additive noise. Below, by introducing a stochastic coefficient γ in front of the linear term u, a stochastically coupled FHN model with multiplicative noise is developed. The model reads

$$
\begin{aligned}
\epsilon du_i &= \left(d_u(u_{i+1} + u_{i-1} - 2u_i) + u_i - \frac{1}{3}u_i^3 - v_i\right) dt + \sqrt{\epsilon}\delta_1 dW_{u_i}, \\
dv_i &= (\gamma_i u_i + a) dt + \delta_2 dW_{v_i}, \\
d\gamma_i &= -d_{\gamma_i}(\gamma_i - \hat{\gamma}_i) dt + \sigma_{\gamma_i} dW_{\gamma_i}, \qquad i = 1, \ldots, N.
\end{aligned}
\tag{69}
$$

The stochastically coupled FHN model with multiplicative noise belongs to the conditional Gaussian framework with $\mathbf{u_I} = (u_1, \ldots, u_N)^T$ and $\mathbf{u_{II}} = (v_1, \gamma_1, \ldots, v_N, \gamma_N)^T$.

To provide intuition about the dynamical behavior of the stochastically coupled FHN model with multiplicative noise (69), we show the time series of the model with $n = 1$, namely there is no diffusion term and the coupled model is three-dimensions. The three-dimensional model with no diffusion reads,

$$
\begin{aligned}
\epsilon du &= \left(u - \frac{1}{3}u^3 - v \right) dt + \sqrt{\epsilon}\delta_1 dW_u, \\
dv &= \left(\gamma u + a \right) dt + \delta_2 dW_v, \\
d\gamma &= -d_\gamma(\gamma - \hat{\gamma})dt + \sigma_\gamma dW_\gamma,
\end{aligned}
\tag{70}
$$

with the following parameters

$$
\epsilon = 0.01, \qquad \delta_1 = 0.2, \qquad \delta_2 = 0.4, \qquad a = 1.05, \qquad d_\gamma = 1, \qquad \hat{\gamma} = 1.
\tag{71}
$$

Figure 12 shows the model simulations with different σ_γ. Note that, when $\sigma_\gamma = 0$, the model (70) reduces to a two-dimensional model with $\gamma \equiv \hat{\gamma}$. From Figure 12, it is clear that u is always intermittent. On the other hand, v has only a small variation when γ is nearly a constant while more extreme events in v are observed when σ_γ increases (see the PDF of v). Note that the extreme events in v are strongly correlated with the quiescent phase of u according to the phase diagram. These extreme events do not affect the regular ones that form a closed loop with the signal of u. With the increase of σ_γ, the periods of u becomes more irregular. This can be seen in (d), which shows the distribution of the time interval T between successive oscillations in u. With a large σ_γ, this distribution not only has a large variance but also shows a fat tail.
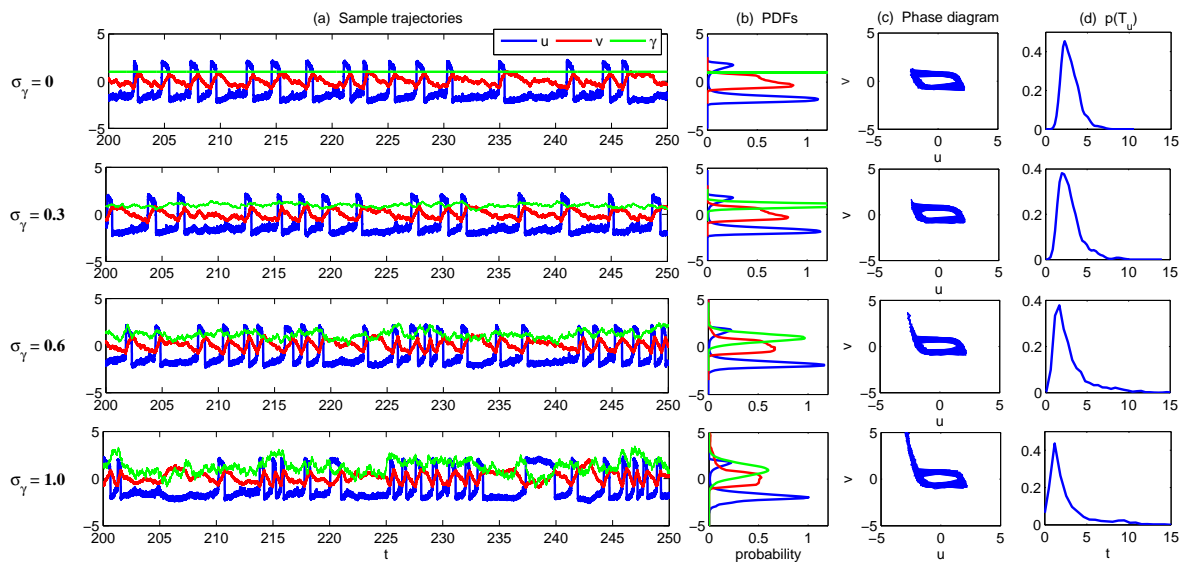
Figure 12. Simulations of the three-dimensional stochastically coupled FHN model with multiplicative noise (70). Different values of noise coefficient σ_γ are used here. (**a**) sample trajectories of u (blue), v (red) and γ (green); (**b**) the associated PDFs; (**c**) phase diagram of (u, v); (**d**) distribution of the time interval T between successive oscillations in u.

Finally, Figure 13 shows the model simulation in the spatial-extended system (69), where $n = 500$. The same parameters as in (71) are taken and $d_{\gamma_i} = d_\gamma = 1$ and $\hat{\gamma}_i = \hat{\gamma} = 1$ for all i. Homogeneous initial conditions $u_i(0) = -2$ and $v_i(0) = 0.5$ are adopted for all $i = 1, \ldots, N$. The four rows show the simulation with different σ_γ, where σ_γ is the noise coefficient at all the grid points, namely $\sigma_{\gamma_i} = \sigma_\gamma$ for

all *i*. With the increase of σ_γ, the spatial structure of u becomes less coherent and more disorganized, which is consistent with the temporal structure as shown in Figure 12. In addition, more extreme events can be observed in the field of v due to the increase of the multiplicative noise.
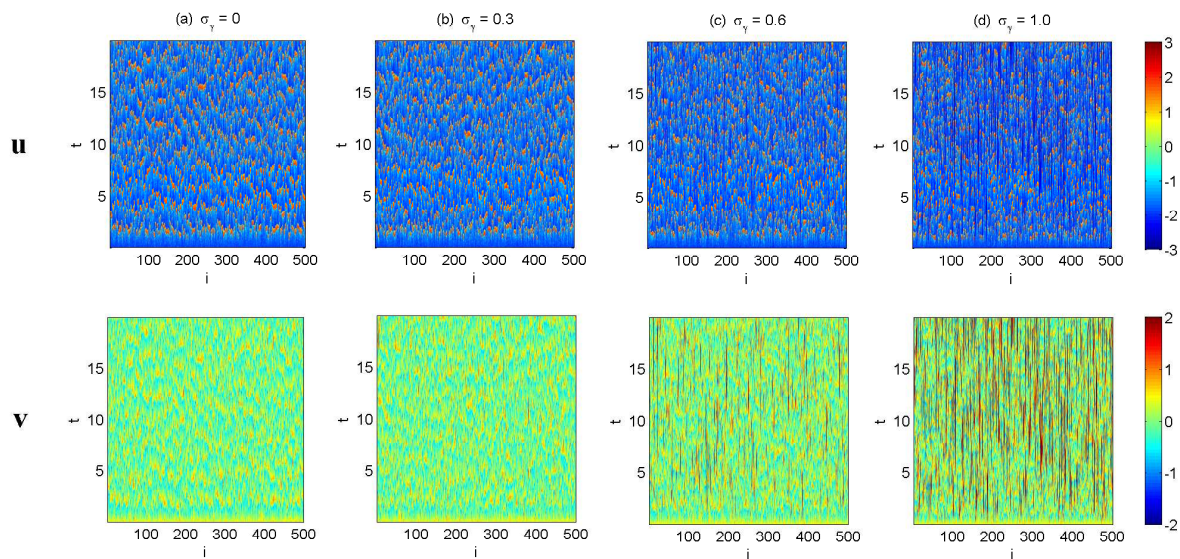
Figure 13. Simulations of the stochastically coupled FHN model with multiplicative noise system with spatial-extended structure (69). The same parameters as in (71) are taken and $d_{\gamma_i} = d_\gamma = 1$ and $\hat\gamma_i = \hat\gamma = 1$ for all *i*. Homogeneous initial conditions $u_i(0) = -2$ and $v_i(0) = 0.5$ are adopted for all $i = 1, \ldots, N$. (**a**–**d**) show the simulation with different $\sigma_\gamma = 0, 0.3, 0.6$ and 1.0, where σ_γ is the noise coefficient at all the grid points, namely $\sigma_{\gamma_i} = \sigma_\gamma$ for all *i*.

4.2.2. The Predator–Prey Models

The functioning of a prey–predator community can be described by a reaction–diffusion system. The standard deterministic predator–prey models have the following form [58]:

$$
\begin{aligned}
\frac{du}{dt} &= D\nabla^2 u + \frac{\alpha}{b} u(1-u) - \gamma \frac{u}{u+h} v, \\
\frac{dv}{dt} &= D\nabla^2 v + \kappa\gamma \frac{u}{u+h} v - \mu v,
\end{aligned}
\tag{72}
$$

where u and v represent predator and prey, respectively. Here, α, b, H, and γ are constants: α is the maximal growth rate of the prey, b is the carrying capacity for the prey population, and H is the half-saturation abundance of prey. Introducing dimensionless variables

$$
\tilde{u} = u/b, \qquad \tilde{v} = v\gamma/(\alpha b), \qquad \tilde{t} = \alpha t, \qquad \tilde{\mathbf{r}} = \mathbf{r}(\alpha/D)^{1/2}
$$

and, using the dimensionless parameters,

$$
h = H/b, \qquad m = \mu/\alpha, \qquad k = \kappa\gamma/\alpha.
$$

The non-dimensional system (by dropping the primes) becomes

$$
\begin{aligned}
\frac{du}{dt} &= \nabla^2 u + u(1-u) - \frac{u}{u+h} v, \\
\frac{dv}{dt} &= \nabla^2 v + k\frac{u}{u+h} v - mv.
\end{aligned}
\tag{73}
$$

Note that, in the predator–prey system, both u and v must be positive.

Clearly, the deterministic model is highly idealized and therefore stochastic noise is added into the system [144,145]. In order to keep the positive constraints for the variables u and v, multiplicative noise is added to the system. One natural choice of the noise is the following:

$$\begin{aligned}
\frac{du}{dt} &= \nabla^2 u + u(1-u) - \frac{u}{u+h} v + f(u)\dot{W}_u, \\
\frac{dv}{dt} &= \nabla^2 v + k\frac{u}{u+h} v - mv.
\end{aligned} \tag{74}$$

Here, $f(u)$ is a function of u and its value vanishes at $u = 0$ to guarantee the positivity of the system. The most straightforward choice of $f(u)$ is $f(u) \propto u$, which leads to small noise when the signal of u is small and large noise when the amplitude of u is large. Another common choice of this multiplicative noise is $f(u) \propto u/\sqrt{1+u^2}$ such that the noise level remains nearly constant when u is larger than a certain threshold. Applying a spatial discretization to the diffusion terms, the stochastic coupled predator–prey in (74) belongs to the conditional Gaussian framework with $\mathbf{u_I} = u$ and $\mathbf{u_{II}} = v$. Note that when $f(u) \approx 0$, the conditional Gaussian estimation (4) and (5) will become singular due to the term $(f(u)f(u)^*)^{-1}$. Nevertheless, the limit cycle in the dynamics will not allow the solution to be trapped at $u = 0$. Assessing this issue using rigorous analysis is an interesting future work.

Figure 14 shows sample trajectories and phase portraits under the simplest setup simulation of (74) without diffusion terms $\nabla^2 u$ and $\nabla^2 v$. The parameters are given as follows:

$$k = 2, \qquad r = 0.4, \qquad m = 0.8, \qquad h = 0.4. \tag{75}$$

Panels (a,b) show the simulations without stochastic noise while Panels (c,d) show the results with a multiplicative noise $f(u) = 0.05u/\sqrt{1+u^2}$. The intermittent behavior is found in both u and v.
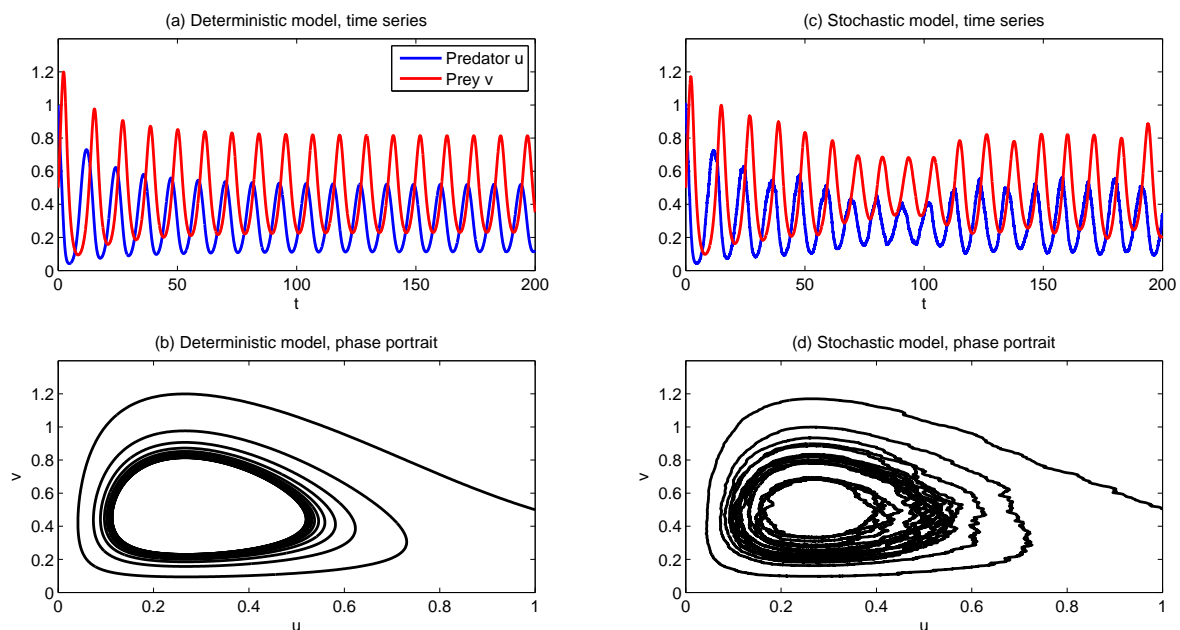
Figure 14. Sample trajectories and phase portraits under the simplest setup simulation of (74) without diffusion terms $\nabla^2 u$ and $\nabla^2 v$. The parameters are given in (75). (**a,b**) simulations without stochastic noise; (**c,d**) simulations with a multiplicative noise $f(u) = 0.05u/\sqrt{1+u^2}$.

In Figure 15, we show the simulations of the spatial-extended system in (74), where we take 30 and 90 grid points in x and y directions, respectively. Spatially periodic boundary conditions are used here. In all three of the simulations in Columns (a–c), the initial values for both u and v are the same. Column (a) shows the model simulation of (74) with diffusion terms $\nabla^2 u$ and $\nabla^2 v$, but the

noise coefficient is zero, namely $f(u) = 0$. Therefore, the model is deterministic and spatial patterns can be seen in all the time instants. In Column (b), we ignore the diffusion terms $\nabla^2 u = \nabla^2 v = 0$ and thus the system is spatially decoupled. Nevertheless, a stochastic noise $f(u) = 0.05u/\sqrt{1 + u^2}$ is added to the system. The consequence is that the initial correlated structure will be removed by the noise and, at $t = 5$, the spatial structures are purely noisy. Finally, in Column (c), the simulations of the model with both the diffusion terms $\nabla^2 u$, $\nabla^2 v$ and the stochastic noise $f(u) = 0.05u/\sqrt{1 + u^2}$ are shown. Although after $t = 5$, the initial structure completely disappears, the diffusion terms correlate the nearby grid points and spatial structures are clearly observed in these simulations. Due to the stochastic noise, the patterns are polluted and are more noisy than those shown in Column (a). In addition, the structure in v is more clear since the noise is directly imposed only on u.
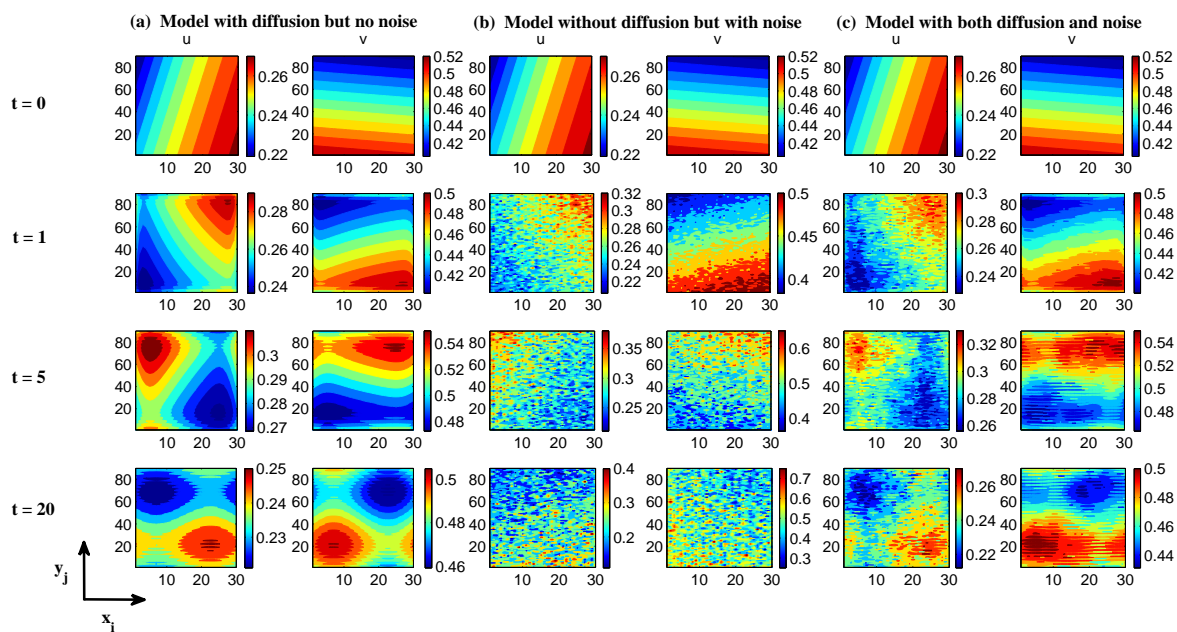
Figure 15. Snapshots at $t = 0, t = 1, t = 5$ and $t = 20$ of the spatial-extended system (74). Spatially periodic boundary conditions are used here. The parameters are given in (75). In all the three simulations in (**a**–**c**), the initial values for both u and v are the same. (**a**) model simulation of (74) with diffusion terms $\nabla^2 u$ and $\nabla^2 v$ but the noise coefficient $f(u) = 0$; (**b**) model simulation of (74) without diffusion terms $\nabla^2 u = \nabla^2 v = 0$. Therefore, the system is spatially decoupled. However, a stochastic noise $f(u) = 0.05u/\sqrt{1 + u^2}$ is added to the system; (**c**) simulations with both the diffusion terms $\nabla^2 u$ and $\nabla^2 v$ and the stochastic noise $f(u) = 0.05u/\sqrt{1 + u^2}$.

4.2.3. A Stochastically Coupled SIR Epidemic Model

The SIR model is one of the simplest compartmental models for epidemiology, and many models are derivations of this basic form [59,146]. The model consists of three compartments: "S" for the number susceptible, "I" for the number of infectious, and "R" for the number recovered (or immune). Each member of the population typically progresses from susceptible to infectious to recovered, namely

$$\text{susceptible} \quad \longrightarrow \quad \text{infectious} \quad \longrightarrow \quad \text{recovered.} \tag{76}$$

This model is reasonably predictive for infectious diseases which are transmitted from human to human, and where recovery confers lasting resistance, such as measles, mumps and rubella.

The classical SIR model is as follows:

$$
\begin{aligned}
dS &= (-\beta SI - \mu_1 S + b)dt, \\
dI &= (\beta SI - \mu_2 I - \alpha I)dt, \\
dR &= (\alpha I - \mu_3 R)dt,
\end{aligned}
\tag{77}
$$

where the total population size has been normalized to one and the influx of the susceptible comes from a constant recruitment rate b. The death rate for the S, I and R class is, respectively, given by μ_1, μ_2 and μ_3. Biologically, it is natural to assume that $\mu_1 < \min\{\mu_2, \mu_3\}$. The standard incidence of disease is denoted by βSI, where β is the constant effective contact rate, which is the average number of contacts of the infectious per unit time. The recovery rate of the infectious is denoted by α such that $1/\alpha$ is the mean time of infection.

When the distribution of the distinct classes is in different spatial locations, the diffusion terms should be taken into consideration and random noise can also be added. Thus, an extended version of the above SIR system (77) can be described as the following [147–149]:

$$
\begin{aligned}
dS &= (\nabla^2 S - \beta SI - \mu_1 S + b)dt + \sigma(S)dW_S, \\
dI &= (\nabla^2 I + \beta SI - \mu_2 I - \alpha I)dt, \\
dR &= (\nabla^2 R + \alpha I - \mu_3 R)dt,
\end{aligned}
\tag{78}
$$

where the noise is multiplicative in order to guarantee the positivity of the three variables. Clearly, the SIR model in (78) is a conditional Gaussian system with $\mathbf{u_I} = S$ and $\mathbf{u_{II}} = (I, R)^T$. It can be used to estimate and predict the number of both infectious and recovered given those susceptible. Note that the SIS model (the model with only S and I variables) [150] is a special case of SIR model and it naturally belongs to the conditional Gaussian framework.

4.2.4. A Nutrient-Limited Model for Avascular Cancer Growth

Here, we present a nutrient-limited model for avascular cancer growth [60], where the cell actions (division, migration, and death) are locally controlled by the nutrient concentration field. Consider a single nutrient field described by the diffusion equation:

$$
dN = (D_N \nabla^2 N - \gamma N \sigma_n - \lambda \gamma N \sigma_c)dt
\tag{79}
$$

in which γ and λ are the nutrient consumption rates of normal and cancer cells, respectively. The domain is the tissue which is represented by a square lattice of size $(L+1) \times (L+1)$ and lattice constant Δ. On the other hand, the growth factor (GF) concentration obeys the diffusion equation

$$
dG = (D_G \nabla^2 G - k^2 G + \Gamma \sigma_c N(G_M - G))dt,
\tag{80}
$$

which includes the natural degradation of GFs, also imposing a characteristic length $\sim 1/k$ for GFs diffusion, and a production term increasing linearly with the local nutrient concentration up to a saturation value G_M. Therefore, we are assuming that the release of GFs involves complex metabolic processes supported by nutrient consumption. The boundary conditions satisfied by the GF concentration field is $G(\vec{x}, t) = 0$ at a large distance $(d > 2/k)$ from the tumor border. Define the non-dimensional variables:

$$
t' = \frac{D_N t}{\Delta^2}, \qquad \vec{x}' = \frac{\vec{x}'}{\Delta}, \qquad N' = \frac{N}{K_0}, \qquad G' = \frac{G}{G_M},
$$

$$
\alpha = \left(\frac{\gamma \Delta^2}{D_N}\right)^{1/2}, \qquad k' = k\left(\frac{\Delta^2}{D_N}\right)^{1/2}, \qquad \Gamma' = \frac{\Gamma \Delta^2}{D_N}, \qquad D = \frac{D_G}{D_N},
$$

with these new variables (dropping the primes) and stochastic noise, the system (79) and (80) becomes

$$dN = (\nabla^2 N - \alpha^2 N \sigma_n - \lambda \alpha^2 N \sigma_c) dt + \sigma_N dW_N,$$
$$dG = (D\nabla^2 G - k^2 G + \Gamma \sigma_c N(1 - G)) dt + \sigma_G dW_G. \tag{81}$$

The coupled system (81) is a conditional Gaussian system with the observations given by the GF concentration $\mathbf{u_I} = G$ and the state estimation for the nutrient field $\mathbf{u_{II}} = N$.

4.3. Large-Scale Dynamical Models in Turbulence, Fluids and Geophysical Flows

4.3.1. The MJO Stochastic Skeleton Model

The Madden–Julian oscillation (MJO) is the dominant mode of variability in the tropical atmosphere on intraseasonal time scales and planetary spatial scales [151–153]. It affects both tropical and extratropical weather and climate. It can also possibly trigger and modify the El Niño-Southern Oscillation [154–156]. Understanding and predicting the MJO is a central problem in contemporary meteorology with large societal impacts.

In [51], a stochastic skeleton model was developed that recovers robustly the most fundamental MJO features: (1) a slow eastward speed of roughly 5 m/s; (2) a peculiar dispersion relation with $d\omega/dk \approx 0$; (3) a horizontal quadrupole vortex structure; (4) the intermittent generation of MJO events; and (5) the organization of MJO events into wave trains with growth and demise, as seen in nature. In fact, the first three features are already covered by the deterministic version of the skeleton model [157,158]. The last two ones are significantly captured by the stochastic version [51]. Using a method based on theoretical waves structures, the MJO skeleton from the model is identified in the observational data [159]. In addition, the stochastic skeleton model is capable of reproducing observed MJO statistics such as the average duration of MJO events and the overall MJO activity [160].

The MJO stochastic skeleton model is given as follows [51]:

$$\begin{aligned}
u_t - yv - \theta_x &= 0, \\
yu - \theta_y &= 0, \\
\theta_t - u_x - v_y &= \bar{H}a - s^\theta, \\
q_t + \bar{Q}(u_x + v_y) &= -\bar{H}a + s^q, \\
a_t &= \text{stochastic birth-death process,}
\end{aligned} \tag{82}$$

and the expectation of the convective activity a satisfies

$$a_t = \Gamma q a, \tag{83}$$

where $\cdot_t$, $\cdot_x$ and $\cdot_y$ denote the derivatives with respect to time t zonal (east-west) coordinate x and meridional (north-south) coordinate y. In (82), u, v and θ are the zonal velocity, meridional velocity and the potential temperature, respectively, and a is the envelope of convective activity. The fourth equation describes the evolution of low-level moisture q. All variables are anomalies from a radiative-convective equilibrium, except a. The skeleton model contains a minimal number of parameters: $\bar{Q}$ is the background vertical moisture gradient, Γ is a proportionality constant. $\bar{H}$ is irrelevant to the dynamics but allows us to define a cooling/drying rate $\bar{H}a$ for the system in dimensional units. s^θ and s^q are external sources of cooling and moistening that need to be prescribed in the system. Notably, the planetary envelope $a \geq 0$ in particular is a collective representation of the convection/wave activity occurring at sub-planetary scale, the details of which are unresolved. A key part of the $q - a$ interaction is the assumption that q influences the tendency of a $a_t = \Gamma q a$, where $\Gamma > 0$ is a constant.

Next, the system (82) is projected onto the first Hermite function in the meridional direction such that $a(x, y, t) = A(x, t)\phi_0$, $q = Q\phi_0$, $s^q = S^q \phi_0$, $s^\theta = S^\theta \phi_0$, where $\phi_0(y) = \sqrt{2}(4\pi)^{-1/4} \exp(-y^2/2)$. Such a meridional heating structure is known to excite only Kelvin waves and the first symmetric equatorial Rossby waves [11,161]. The resulting meridionally truncated equations are

$$K_t + K_x = (S^\theta - \bar{H}A)/2, \tag{84}$$

$$R_t - R_x/3 = (S^\theta - \bar{H}A)/3, \tag{85}$$

$$Q_t + \bar{Q}(K_x - R_x/3) = (\bar{H}A - S^q)(\bar{Q}/6 - 1), \tag{86}$$

$$A_t = \text{stochastic birth-death process.} \tag{87}$$

The expectation of the convective activity A satisfies

$$A_t = \Gamma Q A. \tag{88}$$

Figure 16 shows one model simulation of different fields (Panels (a–d)) as well as the MJO patterns (Panel (e)) with realistic warm pool background heating and moisture sources [51]. The MJO patterns are calculated by the combination of different fields using eigen-decomposition [51]. All the features mentioned at the beginning of this subsection can be clearly seen in Panel (e).
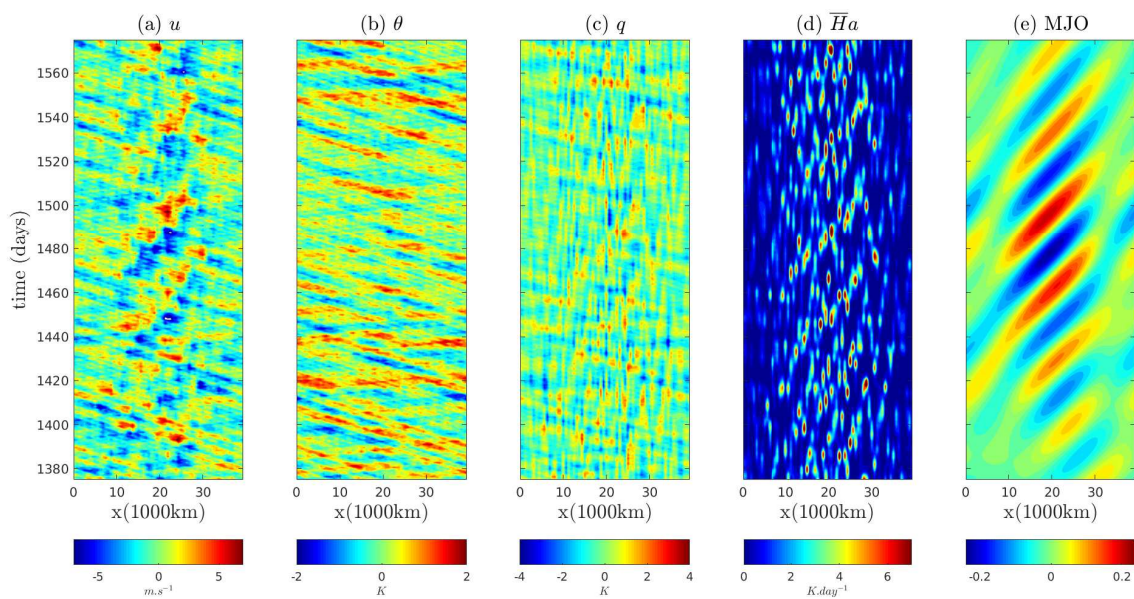
Figure 16. (**a**) Velocity field u; (**b**) temperature θ; (**c**) moisture q; (**d**) convective activity $\bar{H}a$ and (**e**) the MJO patterns in the MJO stochastic skeleton model (84)–(87). The x-axis is the zonal region that extends over the entire equator. The y-axis is time.

To provide intuition of forming a conditional Gaussian system from the original MJO skeleton model, we start with a simplified version of the full spatially extended system in (84)–(87) and (88), namely the stochastic skeleton single-column model,

$$\frac{dK}{dt} = (S^\theta - \bar{H}A)/2, \tag{89}$$

$$\frac{dR}{dt} = (S^\theta - \bar{H}A)/3, \tag{90}$$

$$\frac{dQ}{dt} = (\bar{H}A - S^q)(\bar{Q}/6 - 1), \tag{91}$$

$$\frac{dA}{dt} = \text{stochastic birth-death process,} \tag{92}$$

where the expectation of the convective activity A satisfies

$$\frac{dA}{dt} = \Gamma Q A. \tag{93}$$

In the single-column model, the spatial derivative disappears and the coupled system becomes a four-dimensional stochastic ODEs. In the original stochastic skeleton model [51], the process of the convective activity A is driven by a stochastic birth-death process, which provides a clear physical interpretation of the evolution of the convective activity. Note that in the continuous limit of the stochastic birth-death process (the small jump between the two nearby states $\Delta a \to 0$), (92) converges to a continuous SDE

$$dA = \Gamma Q A dt + \sqrt{\lambda A} dW^A, \tag{94}$$

where W_A is white noise. See [162] for more details. It is clear that (94) involves a multiplicative noise that guarantees the non-negativity of A. In addition, the expectation of the convective activity A converges to (93). Replacing (92) by (94), we arrive at the following system:

$$dA = \Gamma Q A dt + \sqrt{\lambda A} dW^A, \tag{95}$$

$$dK = \frac{1}{2}\left(S^\theta - \bar{H}A\right) dt, \tag{96}$$

$$dR = \frac{1}{3}\left(S^\theta - \bar{H}A\right) dt, \tag{97}$$

$$dQ = \left(1 - \frac{\bar{Q}}{6}\right)(S^q - \bar{H}A) dt, \tag{98}$$

which is a conditional Gaussian system with $\mathbf{u_I} = A$ and $\mathbf{u_{II}} = (K, R, Q)$. Note that, in the content of data assimilation and prediction, judicious model errors [20,28] are often added to the system. These judicious model errors include extra damping and stochastic noise that balance the intrinsic model error and the observational error as well as improve the model capability in capturing the uncertainty. With these judicious model errors, the system (95)–(98) can be modified as

$$dA = \Gamma Q A dt + \sqrt{\lambda A} dW^A, \tag{99}$$

$$dK = \left(-\bar{d}^K K + \frac{1}{2}\left(S^\theta - \bar{H}A\right)\right) dt + \sigma^K dW^K, \tag{100}$$

$$dR = \left(-\bar{d}^R R + \frac{1}{3}\left(S^\theta - \bar{H}A\right)\right) dt + \sigma^R dW^R, \tag{101}$$

$$dQ = \left(-\bar{d}^Q Q + \left(1 - \frac{\bar{Q}}{6}\right)(S^q - \bar{H}A)\right) dt + \sigma^Q dW^Q. \tag{102}$$

Now let us consider the spatially extended system (84)–(87). Making use of the finite Fourier decomposition of the state variables

$$\mathcal{F}(x_j, t) = \sum_{-M+1 \le k \le M} \hat{\mathcal{F}}_k(t) e^{2\pi i k x_j / L} = \sum_{-M+1 \le k \le M} \hat{\mathcal{F}}_k(t) e^{i l_k x_j}, \qquad \mathcal{F} = \{A, K, R, Q\}$$

and plugging them into the truncated system (84)–(87), we reach the following stochastic system:

$$\begin{aligned}
\frac{d\hat{A}_k}{dt} &= \Gamma \sum_{-M+1 \le s \le M} \hat{Q}_s \hat{A}_{k-s} + \sqrt{\lambda_k \hat{A}_k} \dot{W}_k^A, \\
\frac{d\hat{K}_k}{dt} &= (-i l_k - \bar{d}_k^K)\hat{K}_k + \frac{1}{2}\left(\hat{S}_k^\theta - \bar{H}\hat{A}_k\right) + \sigma_k^K \dot{W}_k^K, \\
\frac{d\hat{R}_k}{dt} &= (i l_k / 3 - \bar{d}_k^R)\hat{R}_k + \frac{1}{3}\left(\hat{S}_k^\theta - \bar{H}\hat{A}_k\right) + \sigma_k^R \dot{W}_k^R, \\
\frac{d\hat{Q}_k}{dt} &= -\bar{d}_k^Q \hat{Q}_k - \bar{Q} i l_k \left(\hat{K}_k - \frac{\hat{R}_k}{3}\right) + \left(1 - \frac{\bar{Q}}{6}\right)\left(\hat{S}_k^q - \bar{H}\hat{A}_k\right) + \sigma_k^Q \dot{W}_k^Q, \\
&\quad -M + 1 \le k \le M,
\end{aligned} \tag{103}$$

where judicious model errors with extra damping and stochastic noise are added in these equations. In addition to improving the skill in data assimilation and prediction, these extra damping and stochastic noise also compensate the small scale information that has been truncated in the finite Fourier decomposition.

It is clear that the nonlinear model (103) is a conditional Gaussian system with

$$\mathbf{u_I} = \{\hat{A}_k\}, \quad \text{and} \quad \mathbf{u_{II}} = (\{\hat{K}_k\}, \{\hat{R}_k\}, \{\hat{Q}_k\}).$$

Linking (103) to the original system (82), u, v, θ and q belong to $\mathbf{u_{II}}$ and the wave activity a is the observed variable $\mathbf{u_I}$. Note that there are several further studies of the stochastic skeleton model. By involving additional off equatorial components of convective heating and adding a simple seasonal cycle in a warm pool background [163], the skeleton model succeeds in reproducing the meridionally asymmetric intraseasonal events and a seasonal modulation of intraseasonal variability. Another extended version of the skeleton model involves the refined vertical structure of the MJO in nature [164]. All of these modified versions belong to the conditional Gaussian framework. Note that a rigorous proof has shown that the stochastic skeleton model has the geometric ergodicity property [165].

4.3.2. A Coupled El Niño Model Capturing Observed El Niño Diversity

The El Niño Southern Oscillation (ENSO) has significant impact on global climate and seasonal prediction [166–168]. It is the most prominent year-to-year climate variation on earth, with dramatic ecological and social impacts. The traditional El Niño consists of alternating periods of anomalously warm El Niño conditions and cold La Niña conditions in the equatorial Pacific every two to seven years, with considerable irregularity in amplitude, duration, temporal evolution and spatial structure of these events. In recent decades, a different type of El Niño has been frequently observed [169–172], which is called the central Pacific (CP) El Niño. Different from the traditional El Niño where warm sea surface temperature (SST) occurs in the eastern Pacific, the CP El Niño is characterized by warm SST anomalies confined to the central Pacific, flanked by colder waters to both east and west. Understanding and predicting El Niño diversity has significant scientific and social impacts [173,174].

In [52], a simple modeling framework has been developed that automatically captures the statistical diversity of ENSO. The starting model is a deterministic, linear and stable model that includes the coupled atmosphere, ocean and SST [175]. Then, key factors are added to the system in a systematic way. First, a stochastic parameterization of the wind bursts including both westerly and easterly winds is coupled to the starting model that succeeds in capturing both the eastern Pacific El Niño events and the statistical features in the eastern Pacific [175]. Secondly, a simple nonlinear zonal advection with no ad hoc parameterization of the background SST gradient and a mean easterly trade wind anomaly representing the multidecadal acceleration of the trade wind are both incorporated into the coupled model that enable anomalous warm SST in the central Pacific [176]. Then, a three-state stochastic Markov jump process is utilized to drive the wind burst activity that depends on the strength of the western Pacific warm pool in a simple and effective fashion [52]. It allows the coupled model to simulate the quasi-regular moderate traditional eastern Pacific El Niño, the super El Niño, the CP El Niño as well as the La Niña with realistic features [177]. In particular, the model succeeds in capturing and predicting different super El Niño events [178], including both the directly formed (similar to 1997–1998) and delayed events (similar to 2014–2016). An improved version of the model is also able to recover the season synchronization [179]. The coupled model is as follows:
Atmosphere

$$\begin{aligned} -yv - \partial_x \theta &= 0, \\ yu - \partial_y \theta &= 0, \\ -(\partial_x u + \partial_y v) &= E_q / (1 - \overline{Q}). \end{aligned} \tag{104}$$

Ocean

$$\partial_\tau U - c_1 Y V + c_1 \partial_x H = c_1 \tau_x,$$
$$YU + \partial_Y H = 0, \tag{105}$$
$$\partial_\tau H + c_1(\partial_x U + \partial_Y V) = 0.$$

SST

$$\partial_\tau T + \mu \partial_x(UT) = -c_1 \zeta E_q + c_1 \eta H, \tag{106}$$

Coupling:

$$E_q = \alpha_q T, \qquad \tau_x = \gamma(u + u_p). \tag{107}$$

Here, x is zonal direction and τ is interannual time, while y and Y are meridional direction in the atmosphere and ocean, respectively. The u, v are zonal and meridional winds, θ is potential temperature, U, V are zonal and meridional currents, H is thermocline depth, T is SST, E_q is latent heating, and τ_x is zonal wind stress. All variables are anomalies from an equilibrium state, and are non-dimensional. The coefficient c_1 is a non-dimensional ratio of time scales, which is of order $O(1)$. The term u_p describes stochastic wind burst activity. The atmosphere extends over the entire equatorial belt $0 \le x \le L_A$ with periodic boundary conditions, while the Pacific ocean extends over $0 \le x \le L_O$ with reflection boundary conditions for the ocean model and zero normal derivative at the boundaries for the SST model. The wind bursts and easterly mean trade wind are parameterized as

$$u_p = a_p(\tau) s_p(x) \phi_0(y),$$
$$\frac{da_p}{d\tau} = -d_p(a_p - \hat{a}_p) + \sigma_p(T_W)\dot{W}(\tau), \tag{108}$$

where the noise $\sigma_p(T_W)$ is a state-dependent noise coefficient given by a three-state Markov jump process. In addition, $s_p(x)$ and $\phi_0(y)$ are the prescribed zonal and meridional bases.

Applying the meridional truncation to the coupled system (104)–(108), the coupled model reduces to a set of PDEs that depends only on time t and zonal variables x. These equations describe the zonal propagation of atmosphere and ocean Kelvin and Rossby waves [52]. Figure 17 shows a model simulation. The El Niño diversity is clearly captured, where a traditional eastern Pacific ENSO cycle ($t = 181$–182), a series of CP El Niño ($t = 192$–197) and a super El Niño ($t = 199$) are all observed. Similar to the MJO skeleton model in Section 4.3.1, a suitable set of zonal bases are chosen and the coupled ENSO model becomes a large dimension of coupled ODE system. By adding judicious model errors with extra damping and stochastic forcing (as in the MJO skeleton model), the resulting system belongs to the conditional Gaussian framework, where $\mathbf{u_I}$ includes SST and $\mathbf{u_{II}}$ contains all the variables in atmosphere and ocean.

4.3.3. The Boussinesq Equation

The Boussinesq approximation [11] is a way to solve nonisothermal flow without having to solve for the full compressible formulation of the Navier–Stokes equations. It assumes that variations in density have no effect on the flow field, except that they give rise to buoyancy forces. The Boussinesq equation is derived when the Boussinesq approximation is applied to the full Navier–Stokes equation:

$$\nabla \cdot \mathbf{u} = 0,$$
$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} = -\frac{1}{\rho_0}\nabla p + \nu \nabla^2 \mathbf{u} - g\alpha T + \mathbf{F_u}, \tag{109}$$
$$\frac{\partial T}{\partial t} + \mathbf{u} \cdot \nabla T = \kappa \nabla^2 T + F_T,$$

where T is the temperature, $\mathbf{u}$ is the velocity fields, ρ_0 is the reference density, p is the pressure, g is the acceleration due to gravity, $\mathbf{F_u}$ and F_T are the external forcing and κ is the diffusion coefficient. Note that the forcing terms $\mathbf{F_u}$ and F_T can involve both deterministic and stochastic forcing. The three equations in (109) are the continuity equation, the momentum equation and the thermal equation, respectively. The Boussinesq equation has a wide application, including modeling the Rayleigh–Bénard convection [53,54] and describing strongly stratified flows as in geophysics [55].
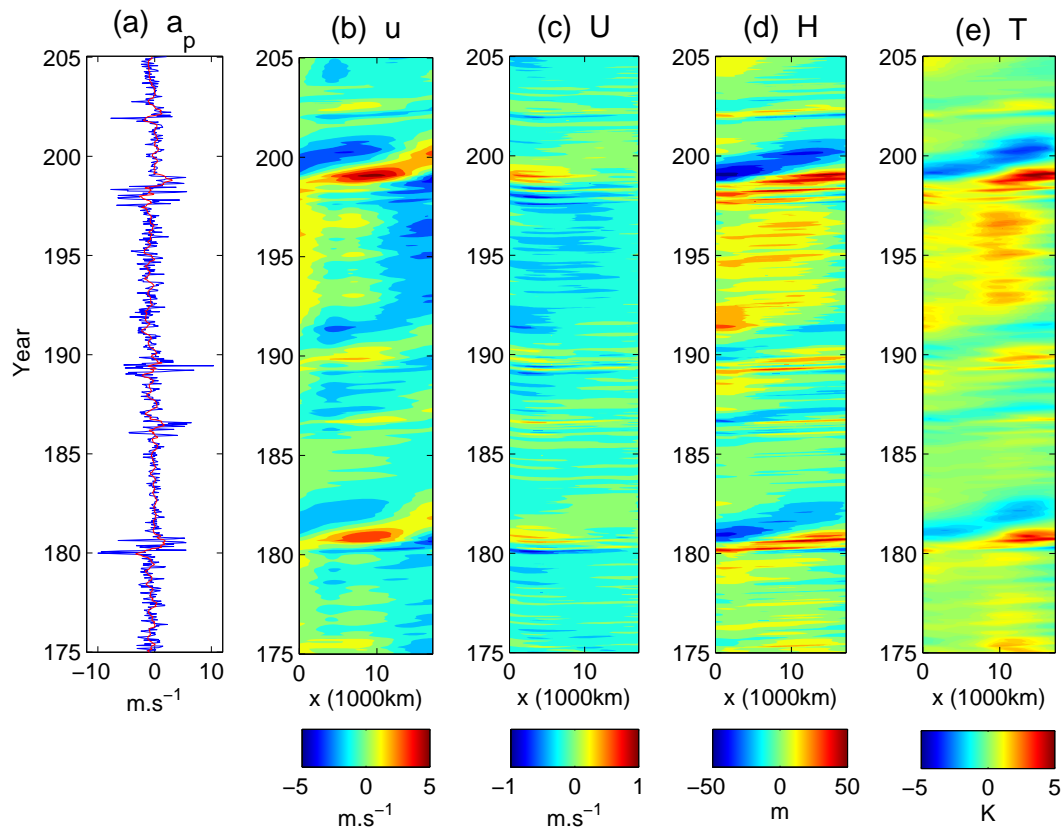
Figure 17. Model simulations of the coupled ENSO model (104)–(108). The *x*-axis is the zonal region of the equatorial Pacific. The *y*-axis is time. (**a**) time series of the stochastic wind burst amplitude a_p (blue) and its 90-day running mean; (**b**–**e**) atmosphere wind u, ocean current U, thermocline depth H and SST T.

Here for simplicity, let us assume the boundary conditions are periodic. The framework below can be easily extended to the Boussinesq system with more general boundary conditions. Similar to the general procedure as described in Section 4.3.1, applying a finite Fourier truncation to (109) and adding judicious model errors with extra damping and noise if needed, the resulting system consists of a large dimension of stochastic ODEs. Here, the continuity equation with divergence free condition provides the eigen-directions of the each Fourier mode associated with the velocity field [11], namely,

$$\mathbf{u}(\mathbf{x}, t) = \sum_{0 < |\mathbf{k}| \le \mathbf{K}} \hat{u}(t) e^{i\mathbf{k}\mathbf{x}} \mathbf{r_k},$$

where $\mathbf{r_k}$ is the eigen-direction associated with the wavenumber $\mathbf{k}$. Note that the resulting system remains highly nonlinear due to the quadratic coupling between different Fourier modes in the advection $\mathbf{u} \cdot \nabla \mathbf{u}$ and $\mathbf{u} \cdot \nabla T$. The straightforward calculation shows that the system belongs to the conditional Gaussian framework with $\mathbf{u_I} = \mathbf{u}$ and $\mathbf{u_{II}} = T$, where the right-hand side should be

understood as the Fourier modes. Such a setup allows the state estimation of the temperature T given the noisy velocity field $\mathbf{u}$ using the closed form (4) and (5) in the conditional Gaussian framework.

In many applications, it is easier to observe the temperature T while the velocity field $\mathbf{u}$ is required to be estimated. However, if we choose $\mathbf{u_I} = \mathbf{T}$ and $\mathbf{u_{II}} = \mathbf{u}$, then (109) is not a conditional Gaussian system. To satisfy the requirement of the conditional Gaussian framework, the equation of the velocity field has to be linearized. Nevertheless, in the situation with large Pradlt number [54,180,181], which occurs quite often in applications (such as Rayleigh–Bénard convection), the nonlinear term $\mathbf{u} \cdot \nabla \mathbf{u}$ can be dropped and the resulting system becomes

$$\nabla \cdot \mathbf{u} = 0,$$
$$\frac{\partial \mathbf{u}}{\partial t} = -\frac{1}{\rho_0} \nabla p + \nu \nabla^2 \mathbf{u} - g\alpha T + \mathbf{F_u}, \tag{110}$$
$$\frac{\partial T}{\partial t} + \mathbf{u} \cdot \nabla T = \kappa \nabla^2 T + F_T.$$

It is important to note that the coupled system remains nonlinear due to the nonlinear coupling $\mathbf{u} \cdot \nabla T$. One interesting issue is to understand the model error by dropping the term $\mathbf{u} \cdot \nabla \mathbf{u}$ in filtering the resulting conditional Gaussian systems.

4.3.4. Darcy–Brinkman–Oberbeck–Boussinesq System—Convection Phenomena in Porous Media

The problem considered in this subsection is the convection phenomena in porous media which is relevant to a variety of science and engineering problems ranging from geothermal energy transport to fiberglass insulation. Consider a Rayleigh–Bénard like problem: convection in a porous media region, Ω, bounded by two parallel planes saturated with fluids. The bottom plate is kept at temperature T_2 and the top plate is kept at temperature T_1 with $T_2 > T_1$. One of the famous models is the following Darcy–Brinkman–Oberbeck–Boussinesq system in the non-dimensional form [56]:

$$\nabla \cdot \mathbf{v} = 0,$$
$$\gamma_a \left(\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{v} \right) + \mathbf{v} - \widetilde{Da}\Delta \mathbf{v} + \nabla p = Ra_D \mathbf{k} T + \mathbf{F_v}, \tag{111}$$
$$\frac{\partial T}{\partial t} + \mathbf{v} \cdot \nabla T = \Delta T + F_T,$$

where $\mathbf{k}$ is the unit normal vector directed upward, $\mathbf{v}$ is the non-dimensional seepage velocity, p is the non-dimensional kinematic pressure, T is the non-dimensional temperature. As in (109), $\mathbf{F_v}$ and F_T here are external forcing, which involve both deterministic and stochastic parts. The parameters in the system are given by the Prandtl–Darcy number, γ_a^{-1}, which is defined as $\gamma_a^{-1} = (\nu h^2)/(\kappa K)$ where ν is the kinematic viscosity of the fluid, h is the distance between the top and bottom plates, κ is the thermal diffusivity and K is the permeability of the fluid; the Brinkman–Darcy number, $\widetilde{Da}$, which is given by $\widetilde{Da} = (\mu_{eff}K)/(\nu h^2)$, where ν_{eff} is the effective kinematic viscosity of the porous media; and the Rayleigh–Darcy number, Ra_D, which takes the form $Ra_D = (g\gamma(T_2 - T_1)Kh)/(\nu\kappa)$, where g is the gravitational acceleration constant, α is the thermal expansion coefficient. The parameter γ_a is also called the non-dimensional acceleration coefficient.

As in the Boussinesq equation, by applying finite Fourier expansion and adding extra damping and noise to the Darcy–Brinkman–Oberbeck–Boussinesq model (111), a conditional Gaussian system is reached with $\mathbf{u_I} = \mathbf{v}$ and $\mathbf{u_{II}} = T$. Again, the right-hand side should be understood as Fourier coefficients.

On the other hand, applying linearization and dropping one of the nonlinear terms $(\mathbf{v} \cdot \nabla)\mathbf{v}$ in (111), the temperature T can be viewed as the observed variable to filter the seepage velocity $\mathbf{v}$,

$$\mathbf{u_I} = T \qquad \text{and} \qquad \mathbf{u_{II}} = \mathbf{v}.$$

The resulting system is given by

$$\nabla \cdot \mathbf{v} = 0,$$
$$\gamma_a \frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} - \widetilde{Da}\Delta\mathbf{v} + \nabla p = Ra_D\mathbf{k}T + \mathbf{F_v}, \tag{112}$$
$$\frac{\partial T}{\partial t} + \mathbf{v} \cdot \nabla T = \Delta T + F_T.$$

4.3.5. The Rotating Shallow Water Equations

The rotating shallow water equation is an appropriate approximation for atmospheric and oceanic motions in the midlatitudes with relatively large length and time scales [11]. In such regimes, the effects of the earth's rotation are important when the fluid motions evolve on a time scale that is comparable or longer than the time scale of rotation. Note that the rotation shallow water equations do not include the effects of density stratification as observed in many other phenomena.

The rotating shallow water equation is as follows:

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} + f\mathbf{u}^\perp + g\nabla h = \mathbf{F_u},$$
$$\frac{\partial h}{\partial t} + \mathbf{u} \cdot \nabla h + (H+h)\nabla \cdot \mathbf{u} = F_h, \tag{113}$$

where $\mathbf{u}$ is the two-dimensional velocity field in the (x,y)-plane, h is the geophysical height. On the right-hand side of (113), $\mathbf{F_u}$ and F_h are forcing terms that include both deterministic and stochastic forcing. As discussed in the previous subsections, choosing $\mathbf{u_I} = \mathbf{u}$ and $\mathbf{u_{II}} = h$ and applying the procedure in Section 4.3.1 with finite Fourier expansion and judicious model error, a conditional Gaussian system is formed.

Alternatively, in some applications, the height variable h can be measured and the velocity field $\mathbf{u}$ is required to be filtered or estimated. Thus, it is natural to set

$$\mathbf{u_I} = h \qquad \text{and} \qquad \mathbf{u_{II}} = \mathbf{u}$$

and then making use of data assimilation to estimate the velocity field. To allow such a problem to fit into conditional Gaussian framework, one of the nonlinear terms $\mathbf{u} \cdot \nabla \mathbf{u}$ is dropped and the resulting system is given by

$$\frac{\partial h}{\partial t} + \mathbf{u} \cdot \nabla h + (H+h)\nabla \cdot \mathbf{u} = F_h,$$
$$\frac{\partial \mathbf{u}}{\partial t} + f\mathbf{u}^\perp + g\nabla h = \mathbf{F_u}. \tag{114}$$

The equations in (114) are then put into the framework in Section 4.3.1 that forms a conditional Gaussian system.

As a side remark, linearizing the shallow water Equation (113) and writing into characteristic forms of the resulting equations provide two types of modes:

1. Geostrophically balanced (GB) modes: $\omega_{\vec{k},B} = 0$; incompressible.

2. Gravity modes: $\omega_{\vec{k},\pm} = \pm\epsilon^{-1}\sqrt{\delta|\vec{k}|^2 + 1}$; compressible.

Let us denote $\hat{v}_{\vec{k},B}$ and $\hat{v}_{\vec{k},\pm}$ as the random coefficients of GB and gravity modes for each Fourier wavenumber, respectively. Following the idea in [20], a coupled system that describes the interactions between the GB and gravity modes are given by

$$d\hat{v}_{\vec{k},B} = (-d_B \hat{v}_{\vec{k},B} + f_{\vec{k},B}(t))dt + \sigma_{\vec{k},B} dW_{\vec{k},B}(t),$$

$$d\hat{v}_{\vec{k},\pm} = \Big((-d_g + i\omega_{\vec{k},\pm} + i\hat{v}_{\vec{k},B})\hat{v}_{\vec{k},\pm} + f_{\vec{k},\pm}(t) \Big)dt + \sigma_{\vec{k},\pm} dW_{\vec{k},\pm}(t). \tag{115}$$

Collecting the equations for different Fourier wavenumbers forms a coupled systems for the GB and gravity modes. Note that the GB modes are in slow time scale while the gravity modes occur in a fast time scale. Denote

$$\mathbf{u_I} = \{\hat{v}_{\vec{k},B}\} \qquad \text{and} \qquad \mathbf{u_{II}} = \{\hat{v}_{\vec{k},\pm}\}.$$

The coupled system (115) including all the Fourier wavenumbers *k* belong to the conditional Gaussian framework.

4.4. Coupled Observation-Filtering Systems for Filtering Turbulent Ocean Flows Using Lagrangian Tracers

Lagrangian tracers are drifters and floaters that collect real-time information of fluid flows [182]. Filtering (or data assimilation) of the turbulent ocean flows using Lagrangian tracers provides a more accurate state estimation and reduces the uncertainty [183–185]. In [61–63], a first rigorous mathematical framework was developed and it was applied to study (1) the information barrier in filtering [61]; (2) the filtering skill of multiscale turbulent flows [62] and (3) the model error in various cheap and practical reduced filters with both linear and strongly nonlinear underlying flows [63].

Consider a *d* dimensional random flow modeled by a finite number of Fourier modes with random amplitudes in periodic domain $(0, 2\pi]^d$,

$$\vec{v}(\vec{x}, t) = \sum_{\vec{k} \in \mathbf{K}} \hat{v}_{\vec{k}}(t) \cdot e^{i\vec{k} \cdot \vec{x}} \cdot \vec{r}_{\vec{k}}.$$

Each $\hat{v}_k(t)$ follows an OU process,

$$d\hat{v}_{\vec{k}}(t) = -d_{\vec{k}} \hat{v}_{\vec{k}}(t)dt + f_{\vec{k}}(t)dt + \sigma_{\vec{k}} dW_{\vec{k}}^v(t). \tag{116}$$

The observations are given by the trajectories of *L* noisy Lagrangian tracers,

$$d\vec{x}_l(t) = \vec{v}(\vec{x}_l(t), t)dt + \sigma_x dW_l^x(t)$$
$$= \sum_{\vec{k} \in \mathbf{K}} \hat{v}_{\vec{k}}(t) \cdot e^{i\vec{k} \cdot \vec{x}_l(t)} \cdot \vec{r}_{\vec{k}} dt + \sigma_x dW_l^x(t), \quad l = 1, \dots, L,$$

which is highly nonlinear due to the appearance of the $\vec{x}_l(t)$ in the exponential term.

Collecting all the Fourier components of the velocity into a vector and including all the tracer displacements into another yields

$$\mathbf{U} = (\hat{v}_1, \dots \hat{v}_\mathbf{K})^\mathbb{T}, \qquad \mathbf{X} = (x_{1,x}, x_{1,y}, \dots, x_{L,x}, x_{L,y})^\mathbb{T}. \tag{117}$$

Then, the coupled observation-filtering system is formed in a concise way,

$$\text{Observations:} \qquad d\mathbf{X} = \mathbf{P}_X(\mathbf{X})\mathbf{U}dt + \Sigma_x dW_X,$$
$$\text{Underlying flow:} \qquad d\mathbf{U} = -\Gamma\mathbf{U}dt + \mathbf{F}(t)dt + \Sigma_u dW_u.$$

This is obviously a conditional Gaussian system with $\mathbf{u_I} = \mathbf{X}$ and $\mathbf{u_{II}} = \mathbf{U}$.

In [62], multiscale turbulent flows were studied. To include both the (slow) geophysically balanced (GB) modes and (fast) gravity modes, the simple flow field in (116) was replaced by

$$d\hat{v}_{\vec{k},B} = (-d_B \hat{v}_{\vec{k},B} + f_{\vec{k},B}(t))dt + \sigma_{\vec{k},B} dW_{\vec{k},B},$$

$$d\hat{v}_{\vec{k},\pm} = \Big((-d_g + i\omega_{\vec{k},\pm})\hat{v}_{\vec{k},\pm} + f_{\vec{k},\pm}(t) \Big)dt + \sigma_{\vec{k},\pm} dW_{\vec{k},\pm}, \tag{118}$$

where $\omega_{\vec{k},\pm} \propto \pm\epsilon^{-1}$ with ϵ being a Rossby number [11,121].

In [63], a nonlinear coupling between GB and gravity modes as described in (115) were further included for generating the true signal while the filter adopts (118). This allows the understanding of model errors based on a novel information theoretic framework [63].

4.5. Other Low-Order Models for Filtering and Prediction

4.5.1. Stochastic Parameterized Extended Kalman Filter Model

The stochastic parameterized extended Kalman filter (SPEKF) model [65,101] was introduced to filter and predict the highly nonlinear and intermittent turbulent signals as observed in nature:

$$
\begin{aligned}
du &= \left((-\gamma + i\omega)u + F(t) + b\right)dt + \sigma_u dW_u, \\
d\gamma &= -d_\gamma(\gamma - \hat{\gamma})dt + \sigma_\gamma dW_\gamma, \\
d\omega &= -d_\omega(\omega - \hat{\omega})dt + \sigma_\omega dW_\omega, \\
db &= -d_b(b - \hat{b})dt + \sigma_b dW_b.
\end{aligned}
\tag{119}
$$

In the SPEKF model (119), the process $u(t)$ is driven by the stochastic damping $\gamma(t)$, stochastic phase $\omega(t)$ and stochastic forcing correction $b(t)$, all of which are specified as Ornstein–Uhlenbeck (OU) processes [66]. In (119), $F(t)$ is a deterministic large-scale forcing, representing external effects such as the seasonal cycle. Physically, the variable $u(t)$ in (119) represents one of the resolved modes (i.e., observable) in the turbulent signal, while $\gamma(t), \omega(t)$ and $b(t)$ are hidden processes. In particular, $\gamma(t), \omega(t)$ and $b(t)$ are surrogates for the nonlinear interaction between $u(t)$ and other unobserved modes in the perfect model. The SPEKF model naturally belongs to conditional Gaussian filtering framework with the observed variable $\mathbf{u_I} = u$ and filtering variables $\mathbf{u_{II}} = (\gamma, \omega, b)^T$.

The nonlinear SPEKF system was first introduced in [65,101] for filtering multiscale turbulent signals with hidden instabilities and has been used for filtering and prediction in the presence of model error [1,20]. In addition to filtering and predicting intermittent signals from nature in the presence of model error [101,186–188], other important applications of using SPEKF to filter complex spatial-extended systems include stochastic dynamical superresolution [103] and effective filters for Navier–Stokes equations [104]. It has been shown that the SPEKF model has much higher skill than classical Kalman filters using the so-called mean stochastic model (MSM)

$$
du = \left((-\hat{\gamma} + i\hat{\omega})u + F(t) + \hat{b}\right)dt + \sigma_u dW_u
\tag{120}
$$

to capture the irregularity and intermittency in nature. Here, $\hat{\gamma}, \hat{\omega}$ and $\hat{b}$ in (120) are the constant mean states of the damping, phase and forcing and therefore the MSM is a linear model with Gaussian statistics.

To provide more intuition, we show in Figure 18 the model simulation of SPEKF together with the MSM for comparison. The following parameters are adopted here

$$
\begin{aligned}
&F(t) = 2\exp(0.2t), \quad \sigma_u = 1, \quad d_\gamma = 0.8, \quad \sigma_\gamma = 2, \quad \hat{\gamma} = 2.5, \\
&d_\omega = 0.6, \quad \sigma_\omega = 1, \quad \hat{\omega} = 2, \quad d_b = 0.8, \quad \sigma_b = 1, \quad \hat{b} = 0.
\end{aligned}
\tag{121}
$$

In Panel (a), the simulation of the MSM is shown. As expected, the signal of u has a large-scale time-periodic behavior due to the time-periodic forcing $F(t)$. On top of the large-scale behavior, small fluctuations with nearly a constant oscillation is found, which is due to the constant phase $\hat{\omega}$. There is no intermittency since the damping $\hat{\gamma} = 2.5 > 0$ always stabilizes the system.

On the other hand, with stochastic damping, stochastic phase and stochastic forcing, the variable $u(t)$ has a rich dynamical behavior with intermittency and random oscillations. See Panels (b,c). In fact, when $\gamma(t)$ becomes negative, the anti-damping results in an exponential increase of the signal of u

and generates intermittency. In addition, the stochastic phase $\omega(t)$ leads to different phase oscillation frequency. For example, $\omega(t)$ at Event (a) is much smaller than that at Event (c), and therefore the oscillation at the phase of Event (c) is much faster. Furthermore, the stochastic forcing $b(t)$ also modifies the signal of $u(t)$. One example is shown in Event (b), where a large amplitude of $b(t)$ forces the signal $u(t)$ to go towards a positive value in the local area.
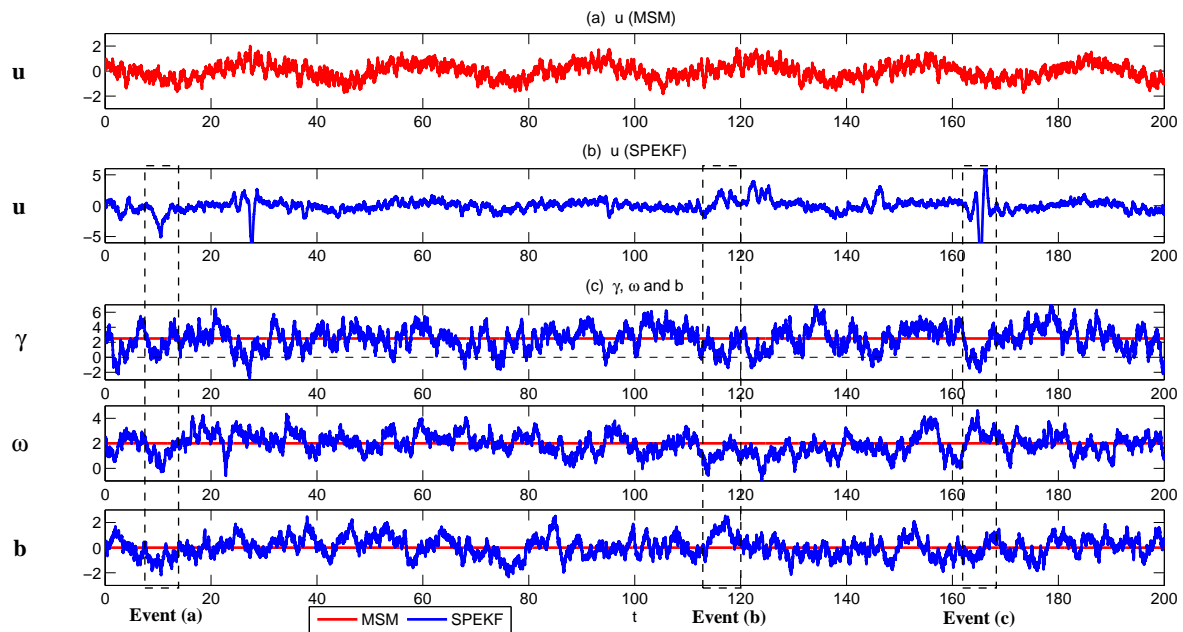

Figure 18. Simulations of the SPEKF model and MSM. (**a**) the observed variable u from MSM (120); (**b**) the observed variable u from SPEKF model; (**c**) hidden variables. The parameters are given in (121). The red curves are for MSM and blue ones are for SPEKF. The black dashed line in the γ process indicates the threshold of zero. When γ is below this threshold, γ becomes anti-damping and the associated signal of u tends to have exponential growth.

4.5.2. An Idealized Surface Wind Model

The following idealized surface wind model is a version of the model for stochastic surface momentum budget [189]. Here, tendencies in the horizontal wind vector (u, v) averaged over a layer of thickness h are modelled as resulting from imbalances between the surface turbulent momentum flux (expressed using a bulk drag law with drag coefficient c_d, taken to be constant) and the ageostrophic difference between the pressure gradient and Coriolis forces:

$$
\begin{aligned}
\frac{du}{dt} &= \langle \Pi_u \rangle - \frac{c_d}{h}(u^2 + v^2)^{1/2}u + \eta_u + \sigma_u \dot{W}_u, \\
\frac{dv}{dt} &= -\frac{c_d}{h}(u^2 + v^2)^{1/2}v + \eta_v + \sigma_v \dot{W}_v, \\
\frac{d\eta_u}{dt} &= -\frac{1}{\tau}\eta_u + \frac{\sigma}{\tau}\dot{W}_1, \\
\frac{d\eta_v}{dt} &= -\frac{1}{\tau}\eta_v + \frac{\sigma}{\tau}\dot{W}_2.
\end{aligned}
$$

The model assumes that tendencies associated with horizontal momentum advection are negligible and that the "large-scale" ageostrophic residual between the pressure gradient and Coriolis forces can be expressed as a mean $(\langle \Pi_u \rangle, 0)$ and fluctuations (η_u, η_v).

It is important to note that simply applying the white noise approximation for u and v (i.e., $\eta_u = \eta_v = 0$) has several limitations [190]. In particular, the white noise model requires unrealistically large values of h to produce serial dependence similar to that observed, and it is

unable to account for the strong anisotropy in the autocorrelation structure of the wind components. Therefore, independent Ornstein–Uhlenbeck processes [66] (η_u, η_v) are included in the coupled system.

With extra additive noise with amplitudes σ_u, σ_v, the coupled system belongs to conditional Gaussian framework with $\mathbf{u_I} = (u, v)^T$ and $\mathbf{u_{II}} = (\eta_u, \eta_v)^T$. The conditional Gaussian framework can be applied to filter the signal of (η_u, η_v), the filtered signal of which provides a clear view of the importance of such parameterization against simple white noise source.

5. Algorithms Which Beat the Curse of Dimension for Fokker–Planck Equation for Conditional Gaussian Systems: Application to Statistical Prediction

The Fokker–Planck equation is a partial differential equation (PDE) that governs the time evolution of the probability density function (PDF) of a complex system with noise [66,67]. For a general nonlinear dynamical system,

$$d\mathbf{u} = \mathbf{F}(\mathbf{u}, t)dt + \mathbf{\Sigma}(\mathbf{u}, t)d\mathbf{W}, \tag{122}$$

with state variables $\mathbf{u} \in \mathbb{R}^N$, noise matrix $\mathbf{\Sigma} \in \mathbb{R}^{N \times K}$ and white noise $\mathbf{W} \in \mathbb{R}^K$, the associated Fokker–Planck equation is given by

$$\frac{\partial}{\partial t} p(\mathbf{u}, t) = -\nabla_{\mathbf{u}} \big(\mathbf{F}(\mathbf{u}, t) p(\mathbf{u}, t) \big) + \tfrac{1}{2} \nabla_{\mathbf{u}} \cdot \nabla_{\mathbf{u}} (\mathbf{Q}(\mathbf{u}, t) p(\mathbf{u}, t)),$$
$$p_t \big|_{t=t_0} = p_0(\mathbf{u}), \tag{123}$$

with $\mathbf{Q} = \mathbf{\Sigma}\mathbf{\Sigma}^T$. In many complex dynamical systems, including geophysical and engineering turbulence, neuroscience and excitable media, the solution of the Fokker–Planck equation in (123) involves strong non-Gaussian features with intermittency and extreme events [1,57,191]. In addition, the dimension of $\mathbf{u}$ in these complex systems is typically very large, representing a variety of variability in different temporal and spatial scales [1,121] (see also the examples in Section 4). Therefore, solving the high-dimensional Fokker–Planck equation for both the steady state and transient phases with non-Gaussian features is an important topic. However, traditional numerical methods such as finite element and finite difference as well as the direct Monte Carlo simulations of (122) all suffer from the curse of dimension [68,69]. Nevertheless, for the conditional Gaussian systems (1) and (2), efficient statistically accurate algorithms can be developed for solving the Fokker–Planck equation in high dimensions and thus beat the curse of dimension. Since the conditional Gaussian system is able to capture many salient features of the turbulent behavior, such algorithms are quite useful in uncertainty quantification, data assimilation and statistical prediction of turbulent phenomena in nature.

5.1. The Basic Algorithm with a Hybrid Strategy

Here, we state the basic efficient statistically accurate algorithms developed in [38]. The only underlying assumption here is that the dimension of $\mathbf{u_I}$ is low while the dimension of $\mathbf{u_{II}}$ can be arbitrarily high.

First, we generate L independent trajectories of the variables $\mathbf{u_I}$, namely $\mathbf{u_I}^1(s \leq t), \dots, \mathbf{u_I}^L(s \leq t)$, where L is a small number. This first step can be done by running a Monte Carlo simulation for the full system, which is computationally affordable with a small L. Then, different strategies are adopted to deal with the parts associated with $\mathbf{u_I}$ and $\mathbf{u_{II}}$, respectively. The PDF of $\mathbf{u_{II}}$ is estimated via a parametric method that exploits the closed form of the conditional Gaussian statistics in (4) and (5),

$$p(\mathbf{u_{II}}(t)) = \lim_{L \to \infty} \frac{1}{L} \sum_{i=1}^{L} p(\mathbf{u_{II}}(t) | \mathbf{u_I}^i(s \leq t)). \tag{124}$$

See [38] for the details of the derivation of (124). Note that the limit $L \to \infty$ in (124) (as well as (125) and (126) below) is taken to illustrate the statistical intuition, while the estimator is the non-asymptotic

version. On the other hand, due to the underlying assumption of the low dimensionality of $\mathbf{u_I}$, a Gaussian kernel density estimation method is used for solving the PDF of the observed variables $\mathbf{u_I}$,

$$p\big(\mathbf{u_I}(t)\big) = \lim_{L \to \infty} \frac{1}{L} \sum_{i=1}^{L} K_{\mathbf{H}}\Big(\mathbf{u_I}(t) - \mathbf{u}_{\mathbf{I}}^{i}(t)\Big), \tag{125}$$

where $K_{\mathbf{H}}(\cdot)$ is a Gaussian kernel centered at each sample point $\mathbf{u}_{\mathbf{I}}^{i}(t)$ with covariance given by the bandwidth matrix $\mathbf{H}(t)$. The kernel density estimation algorithm here involves a "solve-the-equation plug-in" approach for optimizing the bandwidth [192] that works for any non-Gaussian PDFs. Finally, combining (124) and (125), a hybrid method is applied to solve the joint PDF of $\mathbf{u_I}$ and $\mathbf{u_{II}}$ through a Gaussian mixture,

$$p(\mathbf{u_I}(t), \mathbf{u_{II}}(t)) = \lim_{L \to \infty} \frac{1}{L} \sum_{i=1}^{L} \Big(K_{\mathbf{H}}(\mathbf{u_I}(t) - \mathbf{u}_{\mathbf{I}}^{i}(t)) \cdot p(\mathbf{u_{II}}(t)|\mathbf{u}_{\mathbf{I}}^{i}(s \leq t)) \Big). \tag{126}$$

Practically, $L \sim O(100)$ is sufficient for the hybrid method to solve the joint PDF with $N_{\mathbf{I}} \leq 3$ and $N_{\mathbf{II}} \sim O(10)$. See [38] for the illustration of various concrete examples. Note that the closed form of the L conditional distributions in (124) can be solved in a parallel way due to their independence [38], which further reduces the computational cost. Rigorous analysis [193] shows that the hybrid algorithm (126) requires a much less number of samples as compared with traditional Monte Carlo method especially when the dimension of $\mathbf{u_{II}}$ is large.

5.2. Beating the Curse of Dimension with Block Decomposition

The basic algorithm in (126) succeeds in solving the Fokker–Planck equation with $O(10)$ state variables. However, many complex turbulent dynamical systems in nature involve variables that have a much higher dimension (for example, those in Section 4.3). In such a situation, the update of the conditional covariance matrix becomes extremely expensive since the number of entries in the covariance matrix is N^2, where $n = N_{\mathbf{I}} + N_{\mathbf{II}}$ is the total dimension of the variables. Therefore, new strategies are required to be incorporated into the basic algorithm in (126) in order to beat the curse of dimension. In this subsection, we develop an effective strategy with block decomposition [36], which combines with the basic algorithm and can efficiently solve the Fokker–Planck equation in much higher dimensions even with orders in the millions.

Consider the following decomposition of state variables:

$$\mathbf{u}_k = (\mathbf{u}_{\mathbf{I},k}, \mathbf{u}_{\mathbf{II},k}) \quad \text{with} \quad \mathbf{u}_{\mathbf{I},k} \in \mathbb{R}^{N_{\mathbf{I}},k} \text{ and } \mathbf{u}_{\mathbf{II},k} \in \mathbb{R}^{N_{\mathbf{II}},k},$$

where $1 \leq k \leq K$, $N_{\mathbf{I}} = \sum_{k=1}^{K} N_{\mathbf{I},k}$ and $N_{\mathbf{II}} = \sum_{k=1}^{K} N_{\mathbf{II},k}$. Correspondingly, the full dynamics in (1)–(2) are also decomposed into K groups, where the variables on the left-hand side of the k-th group are $\mathbf{u}_k$. In addition, for notation simplicity, we assume both $\mathbf{\Sigma_I}$ and $\mathbf{\Sigma_{II}}$ are diagonal and thus the noise coefficient matrices associated with the equations of $\mathbf{u}_{\mathbf{I},k}$ and $\mathbf{u}_{\mathbf{II},k}$ are $\mathbf{\Sigma}_{\mathbf{I},k}$ and $\mathbf{\Sigma}_{\mathbf{II},k}$, respectively.

To develop efficient statistically accurate algorithms that beat the curse of dimension, the following two conditions are imposed on the coupled system.

Condition 1: In the dynamics of each $\mathbf{u}_k$ in (1) and (2), the terms $\mathbf{A}_{0,k}$ and $\mathbf{a}_{0,k}$ can depend on all the components of $\mathbf{u_I}$ while the terms $\mathbf{A}_{1,k}$ and $\mathbf{a}_{1,k}$ are only functions of $\mathbf{u}_{\mathbf{I},k}$, namely,

$$\begin{aligned}
\mathbf{A}_{0,k} &= \mathbf{A}_{0,k}(t, \mathbf{u_I}), & \mathbf{a}_{0,k} &= \mathbf{a}_{0,k}(t, \mathbf{u_I}), \\
\mathbf{A}_{1,k} &= \mathbf{A}_{1,k}(t, \mathbf{u}_{\mathbf{I},k}), & \mathbf{a}_{1,k} &= \mathbf{a}_{1,k}(t, \mathbf{u}_{\mathbf{I},k}).
\end{aligned} \tag{127}$$

In addition, only $\mathbf{u}_{\mathbf{II},k}$ interacts with $\mathbf{A}_{1,k}$ and $\mathbf{a}_{1,k}$ on the right-hand side of the dynamics of $\mathbf{u}_k$. Therefore, the equation of each $\mathbf{u_I} = (\mathbf{u}_{\mathbf{I},k}, \mathbf{u}_{\mathbf{II},k})$ becomes

$$d\mathbf{u}_{\mathbf{I},k} = [\mathbf{A}_0(t, \mathbf{u}_{\mathbf{I}}) + \mathbf{A}_1(t, \mathbf{u}_{\mathbf{I},k})\mathbf{u}_{\mathbf{II},k}]dt + \mathbf{\Sigma}_{\mathbf{I}}(t, \mathbf{u}_{\mathbf{I},k})d\mathbf{W}_{\mathbf{I}}(t), \tag{128}$$

$$d\mathbf{u}_{\mathbf{II},k} = [\mathbf{a}_0(t, \mathbf{u}_{\mathbf{I}}) + \mathbf{a}_1(t, \mathbf{u}_{\mathbf{I},k})\mathbf{u}_{\mathbf{II},k}]dt + \mathbf{\Sigma}_{\mathbf{II}}(t, \mathbf{u}_{\mathbf{I},k})d\mathbf{W}_{\mathbf{II}}(t). \tag{129}$$

Note that in (128) and (129) each $\mathbf{u}_k$ is fully coupled with other $\mathbf{u}_{k'}$ for all $k' \neq k$ through $\mathbf{A}_0(t, \mathbf{u}_{\mathbf{I}})$ and $\mathbf{a}_0(t, \mathbf{u}_{\mathbf{I}})$. There is no trivial decoupling between different state variables.

Condition 2: The initial values of $(\mathbf{u}_{\mathbf{I},k}, \mathbf{u}_{\mathbf{II},k})$ and $(\mathbf{u}_{\mathbf{I},k'}, \mathbf{u}_{\mathbf{II},k'})$ with $k \neq k'$ are independent with each other.

These two conditions are not artificial and they are actually the salient features of many complex systems with multiscale structures [73], multilevel dynamics [119] or state-dependent parameterizations [103]. Under these two conditions, the conditional covariance matrix becomes block diagonal, which can be easily verified according to (5). The evolution of the conditional covariance of $\mathbf{u}_{\mathbf{II},k}$ conditioned on $\mathbf{u}_{\mathbf{I}}$ is given by:

$$d\mathbf{R}_{\mathbf{II},k}(t) = \left\{ \mathbf{a}_{1,k}\mathbf{R}_{\mathbf{II},k} + \mathbf{R}_{\mathbf{II},k}\mathbf{a}_{1,k}^* + (\mathbf{\Sigma}_{\mathbf{II},k}\mathbf{\Sigma}_{\mathbf{II},k}^*) - (\mathbf{R}_{\mathbf{II},k}\mathbf{A}_{1,k}^*)(\mathbf{\Sigma}_{\mathbf{I},k}\mathbf{\Sigma}_{\mathbf{I},k}^*)^{-1}(\mathbf{R}_{\mathbf{II},k}\mathbf{A}_{1,k}^*)^* \right\} dt,$$

which has no interaction with that of $\mathbf{R}_{\mathbf{II},k'}$ for all $k' \neq k$ since $\mathbf{A}_0$ and $\mathbf{a}_0$ do not enter into the evolution of the conditional covariance. Notably, the evolutions of different $\mathbf{R}_{\mathbf{II},k}$ with $k = 1, \ldots, K$ can be solved in a parallel way and the computation is extremely efficient due to the small size of each individual block. This facilitates the algorithms to efficiently solve the Fokker–Planck equation in large dimensions.

Next, the structures of $\mathbf{A}_{0,k}$ and $\mathbf{a}_{0,k}$ in (127) allow the coupling among all the K groups of variables in the conditional mean according to (4). The evolution of $\bar{\mathbf{u}}_{\mathbf{II},k}$, namely the conditional mean of $\mathbf{u}_{\mathbf{II},k}$ conditioned on $\mathbf{u}_{\mathbf{I}}$, is given by

$$d\bar{\mathbf{u}}_{\mathbf{II},k}(t) = [\mathbf{a}_{0,k} + \mathbf{a}_{1,k}\bar{\mathbf{u}}_{\mathbf{II},k}]dt + \mathbf{R}_{\mathbf{II},k}\mathbf{A}_{1,k}^*(\mathbf{\Sigma}_{\mathbf{I},k}\mathbf{\Sigma}_{\mathbf{I},k}^*)^{-1}[d\mathbf{u}_{\mathbf{I},k} - (\mathbf{A}_{0,k}(t, \mathbf{u}_{\mathbf{I}}) + \mathbf{A}_{1,k}\bar{\mathbf{u}}_{\mathbf{II},k})dt].$$

Finally, let's use concrete examples to illustrate the reduction in the computational cost with the block decomposition.

The first example is the two-layer L-96 model in (35). A typical choice of the number of grid points for the large scale variables u_i ($i = 1, \ldots, I$) is $I = 40$ and that of the equations for the small scale variables $v_{i,j}$ ($j = 1, \ldots, J$) is $J = 5$. Thus, the total number of the small-scale variables $v_{i,j}$ is $I \times J = 200$ and the size of conditional covariance is $n_{total} = (I \times J)^2 = 40,000$. Note that the two-layer L-96 model in (35) can be decomposed into the form in (128) and (129). Each block contains one element of u_k and five elements of $v_{k,j}$ with $j = 1, \ldots, 5$ and k fixed, and the number of the blocks is $K = I = 40$. Therefore, the conditional covariance matrix associated with each block is of size $5 \times 5 = 25$ and the total number of the elements that need to be updated is $n_{reduced} = 25 \times 40 = 1,000$, which is only 2.5% compared with the total elements in the full conditional covariance matrix!

Another example is the stochastically coupled FHN models. In fact, the block decomposition can be applied to all the versions (67)–(69) of the stochastically coupled FHN models as discussed in Section 4.2.1. Here, we illustrate the one in (69) with multiplicative noise. Let's take $n = 500$. Since the conditional statistics involves the information for both v_i and γ_i with $i = 1, \ldots, 500$, the total size of the conditional covariance is $n_{total} = (2 \times 500)^2 = 1,000,000$. On the other hand, with the block decomposition that involves 500 blocks, where each block is of size $2 \times 2 = 4$ and contains the information of only one v_i and one γ_i, the total number of the elements that need to be updated is $n_{reduced} = 500 \times 4 = 2000$. Clearly, the computational cost in the conditional covariance update with the block decomposition is only $n_{reduced}/n_{total} = 0.5\%$ compared with a full update! In Section 5.5, the statistical prediction of both the two-layer L-96 model and the stochastically coupled FHN model with multiplicative noise using the efficient statistically accurate algorithms will be demonstrated.

5.3. Statistical Symmetry

As was discussed in the previous two subsections, the hybrid strategy and the block decomposition provide an extremely efficient way to solve the high-dimensional Fokker–Planck equation associated with the conditional Gaussian systems. In fact, the computational cost in the algorithms developed above can be further reduced if the coupled system (1) and (2) has statistical symmetry [36]:

$$p\big(\mathbf{u}_{\mathbf{I},k}(t), \mathbf{u}_{\mathbf{II},k}(t)\big) = p\big(\mathbf{u}_{\mathbf{I},k'}(t), \mathbf{u}_{\mathbf{II},k'}(t)\big), \quad \text{for all } k \text{ and } k'. \tag{130}$$

Namely, the statistical features for variables with different k are the same. The statistical symmetry is often satisfied when the underlying dynamical system represents a discrete approximation of some PDEs in a periodic domain with nonlinear advection, diffusion and homogeneous external forcing [2,20].

With the statistical symmetry, collecting the conditional Gaussian ensembles $\mathcal{N}(\bar{\mathbf{u}}_{\mathbf{II},k}(t), \mathbf{R}_{\mathbf{II},k}(t))$ for a specific k in K different simulations is equivalent to collecting that for all k with $1 \le k \le K$ in a single simulation. This also applies to $\mathcal{N}(\mathbf{u}_{\mathbf{I}}^i(t), \mathbf{H}(t))$ that are associated with $\mathbf{u}_{\mathbf{I}}$. Therefore, the statistical symmetry implies that the effective sample size is $L' = KL$, where K is the number of the group variables that are statistically symmetric and L is the number of different simulations of the coupled systems via Monte Carlo. If K is large, then a much smaller L is needed to reach the same accuracy as in the situation without utilizing statistical symmetry, which greatly reduces the computational cost.

Below, we discuss the details of constructing the joint PDFs obtaining from the algorithms with statistical symmetry. First, we adopt the one-dimensional case for the convenience of illustration. The method can be easily extended to systems with multivariables and multidimensions. Denote $\bar{u}_1, \bar{u}_2, \ldots, \bar{u}_K$ the mean values of the Gaussian ensembles at different grid points and the associated variance are $R_1, \ldots, R_K$. For simplicity, we only take one full run of the system. Therefore, the total number of Gaussian ensembles is K. Clearly, the 1D PDFs $p(u_i)$ at different grid points are the same and are given by

$$p(u_i) = \lim_{k \to \infty} \frac{1}{K} \sum_{k=1}^{K} \mathcal{N}(\bar{u}_k, R_k), i = 1, \ldots, K. \tag{131}$$

The limit is taken for statistical intuition while a finite and small K is adopted in practice.

Now, we discuss the construction of the joint PDFs. We use the 2D joint PDF $p(u_1, u_2)$ as an illustration. The joint PDF is a Gaussian mixture with K Gaussian ensembles, where the mean of each 2D Gaussian ensemble is

$$\boldsymbol{\mu}_1 = \begin{pmatrix} \bar{u}_1 \\ \bar{u}_2 \end{pmatrix}, \quad \boldsymbol{\mu}_1 = \begin{pmatrix} \bar{u}_2 \\ \bar{u}_3 \end{pmatrix} \quad \ldots \quad \boldsymbol{\mu}_{K-1} = \begin{pmatrix} \bar{u}_{K-1} \\ \bar{u}_K \end{pmatrix}, \quad \boldsymbol{\mu}_K = \begin{pmatrix} \bar{u}_K \\ \bar{u}_1 \end{pmatrix}, \tag{132}$$

and the covariance matrix is given by

$$\mathbf{R}_1 = \begin{pmatrix} R_1 & \\ & R_2 \end{pmatrix}, \quad \mathbf{R}_2 = \begin{pmatrix} R_2 & \\ & R_3 \end{pmatrix} \quad \ldots \quad \mathbf{R}_{K-1} = \begin{pmatrix} R_{K-1} & \\ & R_K \end{pmatrix}, \quad \mathbf{R}_K = \begin{pmatrix} R_K & \\ & R_1 \end{pmatrix}. \tag{133}$$

It is clear from the construction of the ensemble mean in (132) that the subscript in the second component equals that of the first component plus one. That is, the first component of each $\boldsymbol{\mu}_k, k = 1, \ldots, K$ is treated as u_1 due to the statistical symmetry and the second component is treated as the corresponding u_2 associated with each k. The diagonal covariance matrix in (133) comes from the fact that each sample point is independent with each other. This is also true and more obvious for the block diagonal conditional covariance. Notably, the diagonal covariance matrix of each ensemble does not mean that the correlation between u_1 and u_2 is completely ignored. The correlation is reflected regarding how the points of ensemble means $\boldsymbol{\mu}_k, k = 1, \ldots, K$ are distributed.

5.4. Quantifying the Model Error Using Information Theory

Before we apply the efficient statistically accurate algorithms to statistical prediction, we introduce a measurement for quantifying errors. The natural way to quantify the error in the predicted PDF related to the truth is through an information measure, namely the relative entropy (or Kullback–Leibler divergence) [2,187,194–197]. The relative entropy is defined as

$$\mathcal{P}(p(\mathbf{u}), p^M(\mathbf{u})) = \int p(\mathbf{u}) \ln \frac{p(\mathbf{u})}{p^M(\mathbf{u})} d\mathbf{u}, \tag{134}$$

where $p(\mathbf{u})$ is the true PDF and $p^M(\mathbf{u})$ is the predicted one from the efficient statistically accurate algorithms. This asymmetric functional on probability densities $\mathcal{P}(p, p^M) \geq 0$ measures lack of information in p^M compared with p and has many attractive features. First, $\mathcal{P}(p, p^M) \geq 0$ with equality if and only if $p = p^M$. Secondly, $\mathcal{P}(p, p^M)$ is invariant under general nonlinear changes of variables. Notably, the relative entropy is a good indicator of quantifying the difference in the tails of the two PDFs, which is particularly crucial in the turbulent dynamical systems with intermittency and extreme events. On the other hand, the traditional ways of quantifying the errors, such as the relative error $\|p - p^M\|/\|p\|$, usually underestimate the lack of information in the PDF tails.

5.5. Applications to Statistical Prediction

Now, we apply the efficient statistically accurate algorithms developed in the previous subsections to statistical prediction. Two examples of nonlinear complex turbulent dynamical systems in high dimensions are shown below.

The first example is the stochastically coupled FHN model in (69) with multiplicative noise (Section 4.2.1) with $n = 500$,

$$\begin{aligned}
\epsilon du_i &= \left(d_u(u_{i+1} + u_{i-1} - 2u_i) + u_i - \frac{1}{3}u_i^3 - v_i \right) dt + \sqrt{\epsilon}\delta_1 dW_{u_i}, \\
dv_i &= (\gamma_i u_i + a) dt + \delta_2 dW_{v_i}, \\
d\gamma_i &= -d_{\gamma_i}(\gamma_i - \hat{\gamma}_i) dt + \sigma_{\gamma_i} dW_{\gamma_i}, \qquad i = 1, \ldots, N.
\end{aligned} \tag{135}$$

The parameters are given by (71).

The second example is the two-layer L-96 model in (35) (Section 4.1.1) with $I = 40$ and $J = 5$. The parameters are given by (36) and (37):

$$\begin{aligned}
du_i &= \left(u_{i-1}(u_{i+1} - u_{i-2}) + \sum_{j=1}^{J} \gamma_{i,j} u_i v_{i,j} - \bar{d}_i u_i + F \right) dt + \sigma_u dW_{u_i}, \qquad i = 1, \ldots, I, \\
dv_{i,j} &= \left(-d_{v_{i,j}} v_{i,j} - \gamma_j u_i^2 \right) dt + \sigma_{i,j} dW_{v_{i,j}}, \qquad j = 1, \ldots, J.
\end{aligned} \tag{136}$$

Both the stochastically coupled FHN model (135) and the two-layer L-96 model (136) satisfy the model structure as described in Section 5.2 such that the block decomposition applies. Figure 19 provides a schematic illustration of the coupling between different variables in these models (for illustration purposes, $I = n = 6$ is used in the figure). The multiscale and layered structures can be easily observed. It is also clear that all the variables are coupled to each other with no trivial decoupling. Note that the stochastically coupled FHN model (135) with the given parameters here also satisfies the statistical symmetry while the two-layer L-96 model (136) is inhomogeneous.
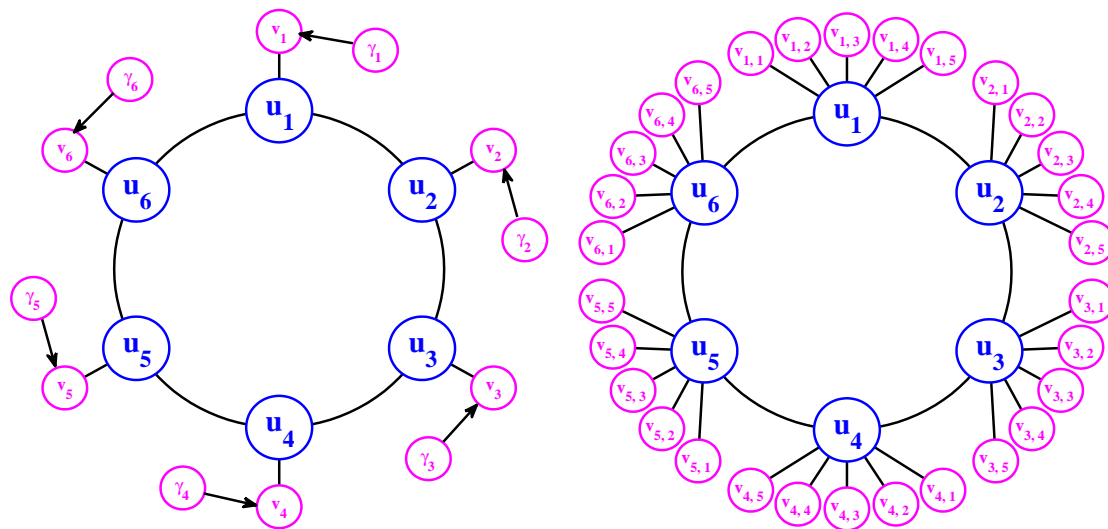
Figure 19. Schematic illustration of the coupling between different variables in the stochastically coupled FHN model (135) (**left**) and the two-layer L-96 model (136) (**right**). For illustration purposes, $I = n = 6$ is used in this figure while $I = 40$ and $n = 500$ in the model simulations. All the variables are coupled to each other with no trivial decoupling.

5.5.1. Application to the Stochastically Coupled FHN Model with Multiplicative Noise Using Statistical Symmetry

Here, we illustrate the statistical prediction of the stochastically coupled FHN model with multiplicative noise (135), where the number of grid points in space is $n = 500$. The noise coefficient in the γ_i process is $\sigma_{\gamma_i} = 0.6$ for all i. The other parameters are the same as those in (71) and the noise coefficients are the same at different grid points i, namely

$$\epsilon = 0.01, \qquad \delta_1 = 0.2, \qquad \delta_2 = 0.4, \qquad a = 1.05, \qquad d_{\gamma_i} = 1, \qquad \hat{\gamma}_i = \hat{\gamma} = 1, \qquad \sigma_{\gamma_i} = \sigma_\gamma = 0.6. \qquad (137)$$

We also adopt homogeneous initial conditions $u_i(0) = -2$ and $v_i(0) = 0.5$ for all $i = 1, \ldots, N$. Therefore, the model satisfies the statistical symmetry. The model simulations were shown in Figures 12 and 13, where intermittent events are observed in both the fields of u and v.

Figure 20 shows the time evolution of the first four moments associated with u_1 and v_1. Note that due to the statistical symmetry, the evolutions of these moments for different u_i and v_i are the same as u_1 and v_1. The dot at $t = 0.68$ indicates the time instant that u arrives at its most non-Gaussian phase while $t = 4.2$ is a non-Gaussian phase, where u is nearly the statistical equilibrium while v is still transient. In Figures 21 and 22, the statistical prediction of the 1D marginal and 2D joint PDFs are shown at these two time instants. At $t = 0.68$, most of the probability density of u_1 is concentrated around $u_1 = -1.8$. However, there is a small but non-zero probability around $u_1 = 2$ (see the subpanel with the logarithm scale), which leads to large skewness and kurtosis of the statistics of u_1 as observed in Figure 20. The efficient statistically accurate algorithms succeed in predicting the PDFs at this transient phase and are able to accurately recover the statistics in predicting such extreme events. A similar performance is found in Figure 22, where the bimodal PDF of u_1, the highly skewed PDF of v_1 and the Gaussian one of γ_1 are all predicted with high accuracy. The fully non-Gaussian joint PDFs $p(u_1, v_1)$ and $p(v_1, \gamma_1)$ are also predicted with almost no bias. These results indicate the skillful behavior of the algorithms developed above.

Note that the total dimension of the stochastically coupled FHN model with multiplicative noise (135) is $3N = 1500$. Due to the statistical symmetry, the effective sample size in the statistical prediction as shown in Figures 21 and 22 is $L' = NL = 500L$, where L is the number of repeated simulations of the systems. In fact, the simulations in the efficient statistically accurate algorithms

are achieved with only $L = 1$. This means we only run the model (135) once and apply the efficient statistically accurate algorithms that provide the accurate results in these figures. Therefore, the statistical prediction here is extremely cheap! On the other hand, we take $L_C = 300$ in Monte Carlo simulations and again use statistical symmetry to generate the true PDFs and therefore the effective sample size in Monte Carlo simulation is $L_C' = NLC = 150,000$.
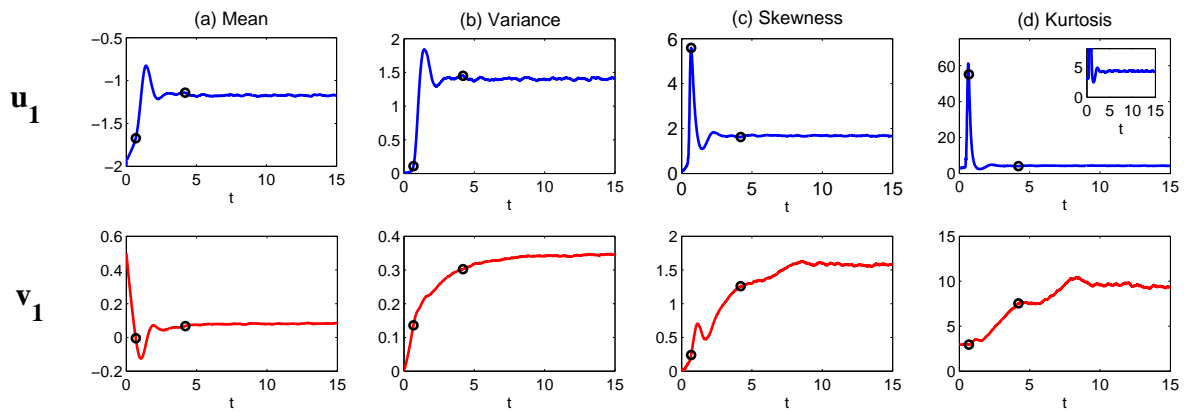

Figure 20. Time evolution of the first four moments ((**a**) mean; (**b**) variance; (**c**) skewness and (**d**) kurtosis) associated with u_1 and v_1 of the stochastically coupled FHN model with multiplicative noise (135), where $n = 500$. Due to the statistical symmetry, the evolutions of these moments for different u_i and v_i are the same as u_1 and v_1. The dots at $t = 0.68$ and $t = 4.2$ indicate the time instants that the statistical prediction using the efficient statistically accurate algorithms are tested. Note that $t = 0.68$ corresponds to the time instant that u arrives at its most non-Gaussian phase while $t = 4.2$ is a non-Gaussian phase where u is nearly the statistical equilibrium while v is still transient.
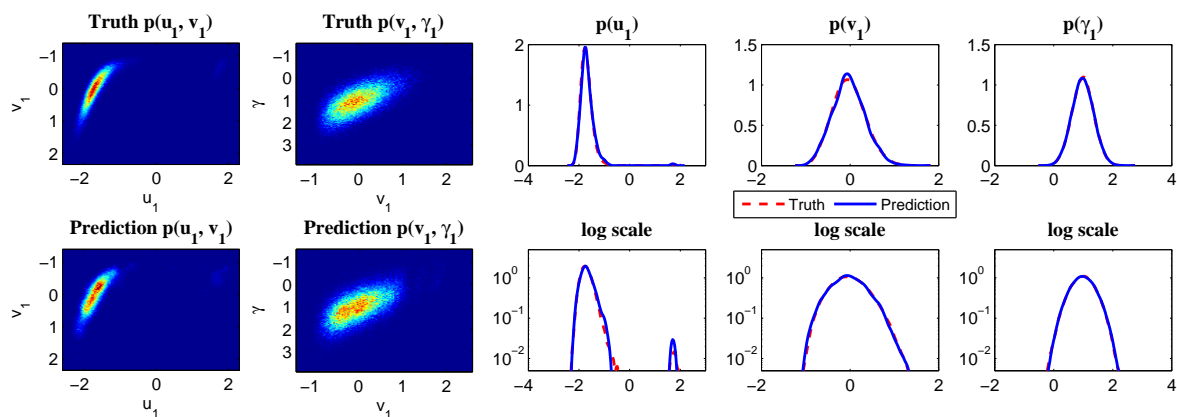

Figure 21. Statistical prediction of the 1D and 2D PDFs of the stochastically coupled FHN model with multiplicative noise (135) at a transit phase $t = 0.68$, where $n = 500$. Note that there is a small but non-zero probability around $u_1 = 2$ (see the subpanel with the logarithm scale).

5.5.2. Application to the Two-Layer L-96 Model with Inhomogeneous Spatial Structures

Now we apply the efficient statistically accurate algorithms to predict the two-layer L-96 model in (136). Here, $I = 40$ and $J = 5$. The parameters are given by (36) and (37). The model behavior with different forcing F is shown in Figures 2–4.

Although the two-layer inhomogeneous L-96 model in (136) has no statistical symmetry, the model structure nevertheless allows the effective block decomposition. Below, $L = 500$ trajectories of each variable u_i are simulated from (136) to implement the efficient statistically accurate algorithms. As a comparison, a direct Monte Carlo method requires $L_C = 150,000$ samples for each of the 240 variables

for an accurate estimation of at least the one-point statistics. This means the total number of samples is around 4×10^7! For an efficient calculation of the truth, we focus only on the statistical equilibrium state here, but the algorithms are not restricted to the equilibrium statistics. The true PDFs are calculated using the Monte Carlo samples over a long time series in light of the ergodicity while the recovered PDFs from the efficient statistically accurate algorithms are computed at $t = 25$.
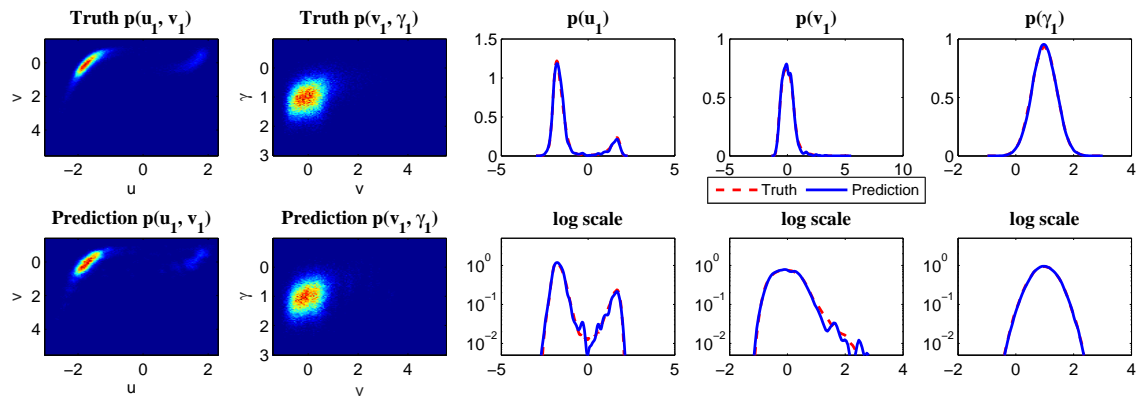

Figure 22. Statistical prediction of the 1D and 2D PDFs of the stochastically coupled FHN model with multiplicative noise (135) at a transit phase $t = 4.2$, where $n = 500$.

In Figures 23–25, we show the statistical prediction in all the three regimes with $F = 5, 8$ and 16, respectively, using the efficient statistically accurate algorithm and compare them with the truth. Here, we only show the results at $i = 11$. Qualitative similar results are found at other grid points. In these figures, the diagonal subpanels here show the 1D marginal PDFs of u_{11} and $v_{11,1}, \ldots, v_{11,5}$, where the blue one is the prediction and the red one is the truth. The (k_i, k_j)-subpanel with $k_i > k_j$ (below the diagonal panel) shows the true 2D PDF using a large number of Monte Carlo samples (red colormap) while the one with $k_i < k_j$ (above the diagonal panels) shows the predicted one using the efficient statistically accurate algorithm (blue colormap). The (k_i, k_j)-panel is compared with the (j, i)-panel. Note that for the simplicity of comparison the labels u_{11} and $v_{11,1}, \ldots, v_{11,5}$ on the bottom and left of the (i, j)-panel correspond to the x-axis and y-axis of the truth and the y-axis and x-axis of the predicted PDFs. On the other hand, Figure 26 shows several 2D joint PDFs of the large scale $p(u_{i_1}, u_{i_2})$ with different i_1 and i_2. It is clear that all the strong non-Gaussian features, including highly skewed, multimodal and fat-tailed distribution, are accurately predicted by the statistically accurate algorithm. Finally, Figure 27 shows the error in the predicted PDF as a function of the sample points L compared with the truth, where the error is computed via the relative entropy (134). In fact, with $L \sim O(100)$ samples in the efficient statistically accurate algorithms, the statistical prediction has already been accurate enough. Note that the small non-zero asymptotic values as a function of L in the error curves is due to the numerical error in computing the integral of the relative entropy (134) as well as the sampling in the Monte Carlo simulations.

It is worthwhile pointing out that although only the predicted one-point and two-point statistics are shown in both the examples here, the algorithms can actually provide an accurate estimation of the full joint PDF of $\mathbf{u}_{II}$, using a small number of samples. This is because the sample size in these algorithms does not grow exponentially as the dimension of $\mathbf{u}_{II}$, which is fundamentally different from Monte Carlo methods. See [193] for a theoretical justification.
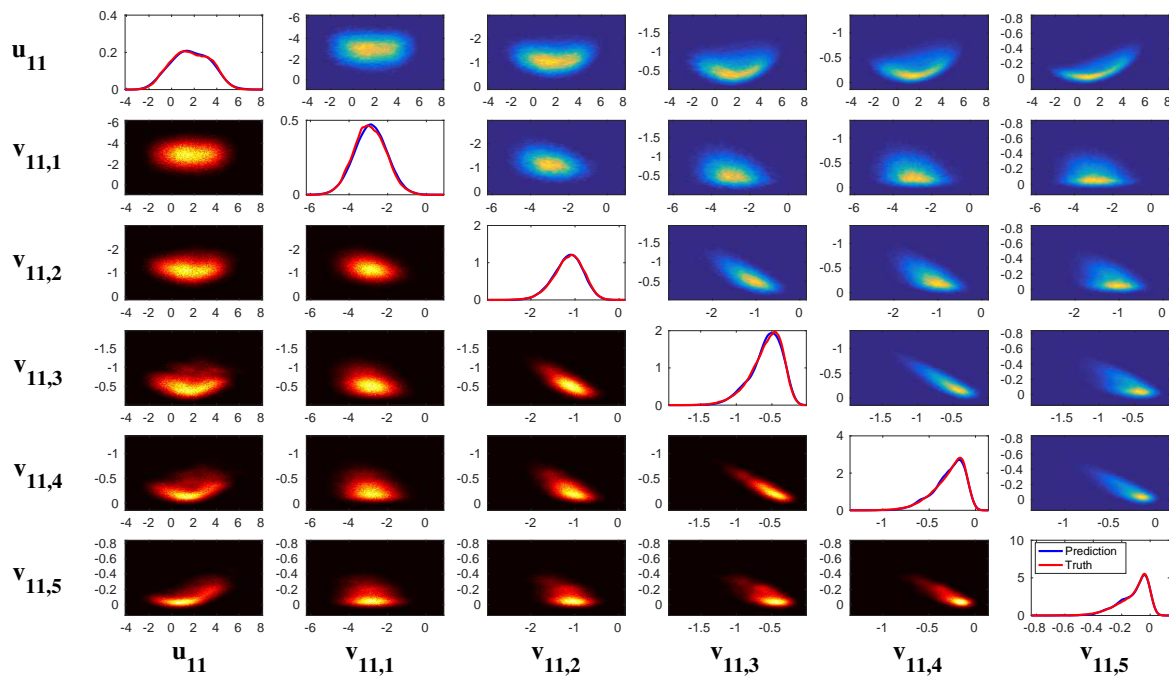
Figure 23. Inhomogeneous two-layer L-96 model (35) with $F = 5$. The other parameters are given in (36) and (37). Comparison of the 1D and 2D true and recovered PDFs at $i = 11$. The diagonal subpanels here show the 1D marginal PDFs of u_{11} and $v_{11,1}, \ldots, v_{11,5}$, where the blue one is the prediction and the red one is the truth. In this figure with 6×6 subpanels, the (k_i, k_j)-subpanel with $k_i > k_j$ (below the diagonal panel) shows the true 2D PDF using a large number of Monte Carlo samples (red colormap) while the one with $k_i < k_j$ (above the diagonal panels) shows the predicted one using the efficient statistically accurate algorithm (blue colormap). The (k_i, k_j)-panel is compared with the (j, i)-panel. Note that for the simplicity of comparison the labels u_{11} and $v_{11,1}, \ldots, v_{11,5}$ on the bottom and left of the (i, j)-panel correspond to the x-axis and y-axis of the truth and the y-axis and x-axis of the predicted PDFs.

Figure 24. Inhomogeneous two-layer L-96 model (35), similar to Figure 23 but with $F = 8$.

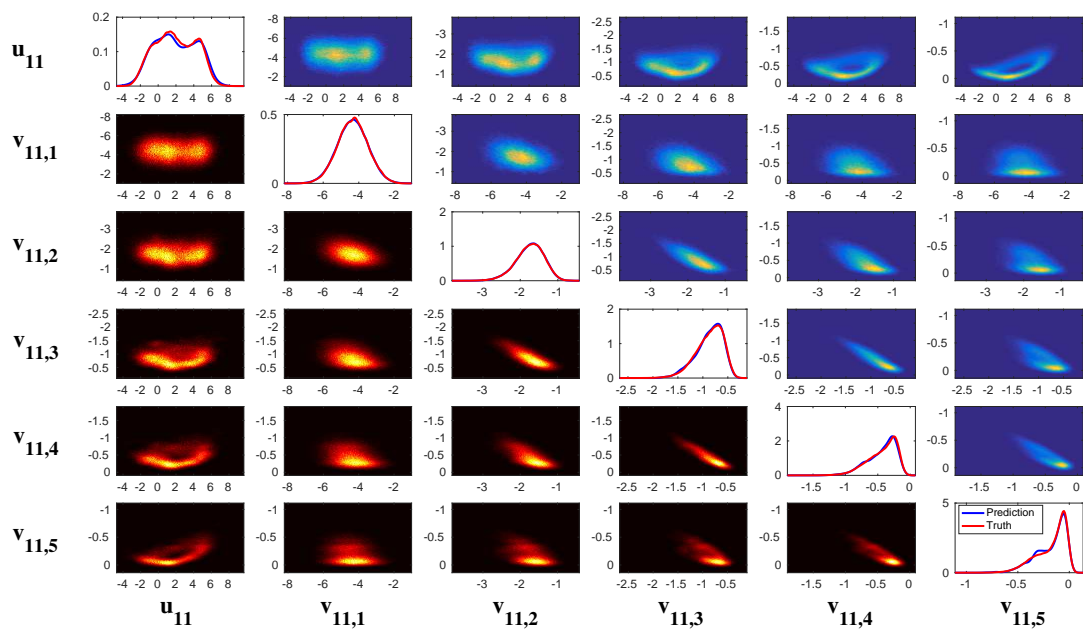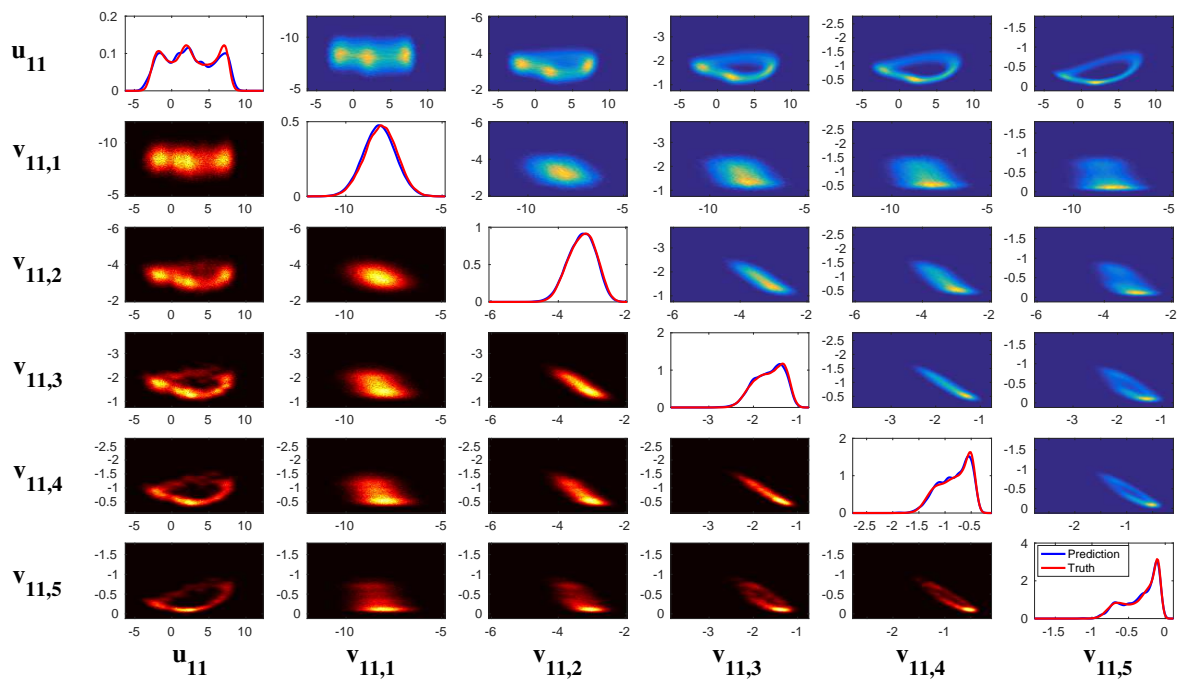**Figure 25.** Inhomogeneous two-layer L-96 model (35), similar to Figure 23 but with *F* = 16.
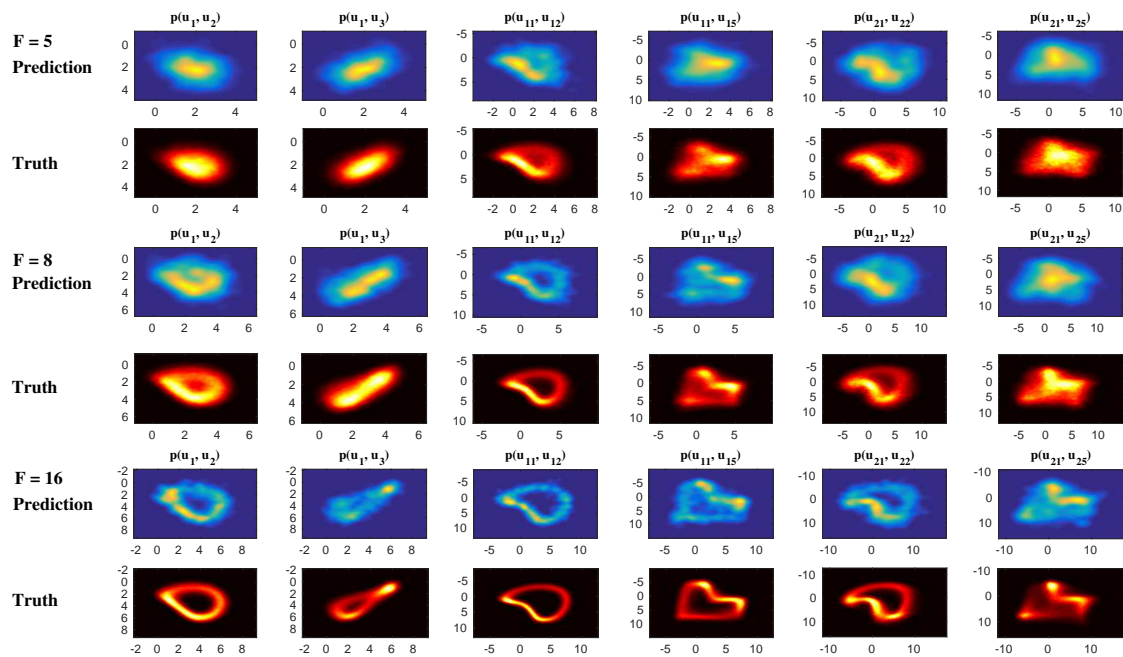

Figure 26. Inhomogeneous two-layer L-96 model (35) with $F = 5$ (**top**); $F = 8$ (**middle**) and $F = 16$ (**bottom**). The other parameters are given in (36) and (37). Comparison of the joint PDFs of the large scale $p(u_{i_1}, u_{i_2})$ with different i_1 and i_2.
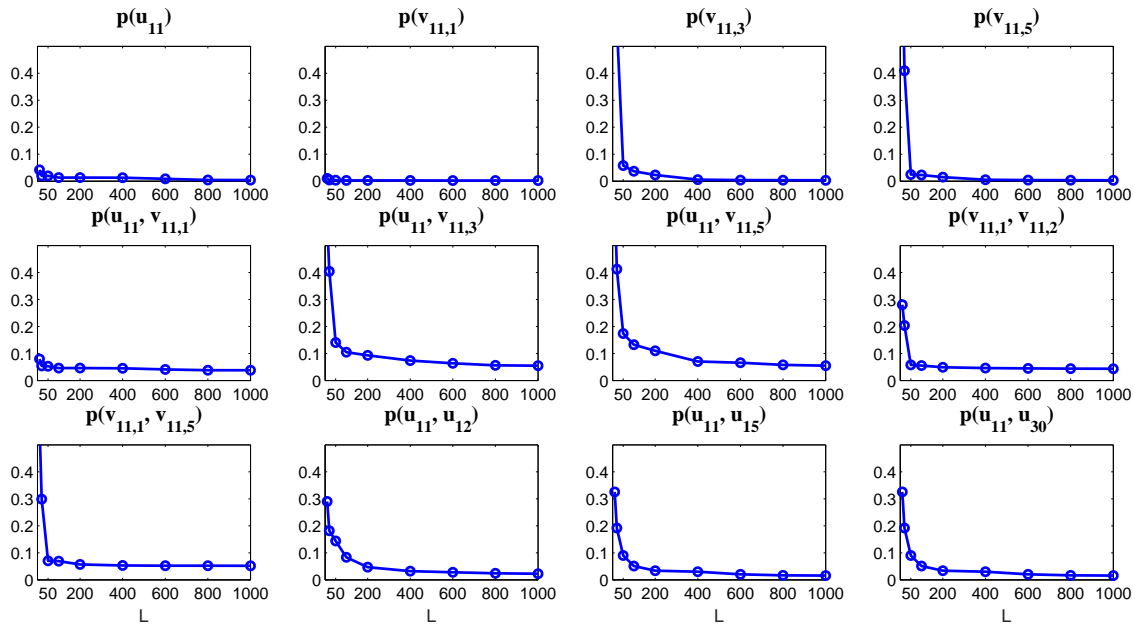
Figure 27. Inhomogeneous two-layer L-96 model (35) with $F = 5$. Error in the predicted PDF as a function of the sample points L compared with the truth. The error is computed via the relative entropy (134).

6. Multiscale Data Assimilation, Particle Filters, Conditional Gaussian Systems and Information Theory for Model Errors

6.1. Parameter Estimation

The conditional Gaussian system (1) and (2) and its closed analytic formula for solving the conditional statistics (4) and (5) also provide a framework for parameter estimation. In fact, $\mathbf{u_{II}}$ can be written as

$$\mathbf{u_{II}} = (\widetilde{\mathbf{u}}_{\mathbf{II}}, \boldsymbol{\Lambda}),$$

where $\mathbf{u_{II}}$ in $\mathbb{R}^{\widetilde{N}_2}$ is physical process variables and $\boldsymbol{\Lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_p) \in \mathbb{R}^{N_{2,p}}$ denotes the model parameters. Here, $N_2 = \widetilde{N}_2 + N_{2,p}$. Rewriting the conditional Gaussian system (1) and (2) in terms of $\mathbf{u_{II}} = (\widetilde{\mathbf{u}}_{\mathbf{II}}, \boldsymbol{\Lambda})$ yields

$$d\mathbf{u_I} = [\mathbf{A}_0(t, \mathbf{u_I}) + \mathbf{A}_1(t, \mathbf{u_I})\widetilde{\mathbf{u}}_{\mathbf{II}} + \mathbf{A}_{1,\lambda}(t, \mathbf{u_I})\boldsymbol{\Lambda}]dt + \boldsymbol{\Sigma}_{\mathbf{I}}(t, \mathbf{u_I})d\mathbf{W_I}(t), \tag{138}$$

$$d\mathbf{u_{II}} = [\mathbf{a}_0(t, \mathbf{u_I}) + \mathbf{a}_1(t, \mathbf{u_I})\widetilde{\mathbf{u}}_{\mathbf{II}} + \mathbf{a}_{1,\lambda}(t, \mathbf{u_I})\boldsymbol{\Lambda}]dt + \boldsymbol{\Sigma}_{\mathbf{II}}(t, \mathbf{u_I})d\mathbf{W_{II}}(t). \tag{139}$$

The Equation (138) and (139) includes both the dynamics of $\widetilde{\mathbf{u}}_{\mathbf{II}}$ and those of $\boldsymbol{\Lambda}$. Both the physical process variables $\widetilde{\mathbf{u}}_{\mathbf{II}}$ and the parameters $\boldsymbol{\Lambda}$ are coupled with any highly nonlinear functions of $\mathbf{u_I}$. Nevertheless, any monomial involving both $\widetilde{\mathbf{u}}_{\mathbf{II}}$ and $\boldsymbol{\Lambda}$ is not allowed since otherwise the conditional Gaussian structure will break. Notably, although $\widetilde{\mathbf{u}}_{\mathbf{II}}$ and $\boldsymbol{\Lambda}$ are named as model states and parameters, they can also be understood as stochastic parameterization and simplification of complex physical process via bulk average, respectively. Therefore, $\widetilde{\mathbf{u}}_{\mathbf{II}}$ and $\boldsymbol{\Lambda}$ often share the same role in providing extra information of $\mathbf{u_I}$ that leads to various non-Gaussian characteristics. This is a typical feature in multiscale complex turbulent dynamical systems.

Below, we provide different frameworks for parameter estimation. Let's temporally ignore $\widetilde{\mathbf{u}}_{\mathbf{II}}$ for notation simplicity and the dynamics is given by

$$d\mathbf{u_I} = [\mathbf{A}_0(t, \mathbf{u_I}) + \mathbf{A}_{1,\lambda}(t, \mathbf{u_I})\boldsymbol{\Lambda}^*]dt + \boldsymbol{\Sigma}_{\mathbf{I}}(t, \mathbf{u_I})d\mathbf{W_I}(t), \tag{140}$$

where $\boldsymbol{\Lambda}^*$ is the true parameter.

6.1.1. Direct Parameter Estimation Algorithm

Since $\mathbf{\Lambda}$ are constant parameters, it is natural to augment the dynamics with the following relationship,

$$d\mathbf{u_I} = [\mathbf{A}_0(t, \mathbf{u_I}) + \mathbf{A}_{1,\lambda}(t, \mathbf{u_I})\mathbf{\Lambda}]dt + \mathbf{\Sigma_I}(t, \mathbf{u_I})d\mathbf{W_I}(t), \qquad (141)$$

$$d\mathbf{\Lambda} = 0, \qquad (142)$$

where an initial uncertainty for the parameter $\mathbf{\Lambda}$ is assigned. According to (4) and (5), the time evolutions of the mean $\bar{\mathbf{u}}_{\mathbf{II}}$ and covariance $\mathbf{R_{II}}$ of the estimate of $\mathbf{\Lambda}$ are given by

$$d\bar{\mathbf{u}}_{\mathbf{II}}(t) = (\mathbf{R_{II}}\mathbf{A}_1^*(t, \mathbf{u_I}))(\mathbf{\Sigma_I}\mathbf{\Sigma_I^*})^{-1}(t, \mathbf{u_I}) \left[d\mathbf{u_I} - (\mathbf{A}_0(t, \mathbf{u_I}) + \mathbf{A}_1(t, \mathbf{u_I})\bar{\mathbf{u}}_{\mathbf{II}})dt \right], \qquad (143)$$

$$d\mathbf{R_{II}}(t) = -(\mathbf{R_{II}}\mathbf{A}_1^*(t, \mathbf{u_I}))(\mathbf{\Sigma_I}\mathbf{\Sigma_I^*})^{-1}(t, \mathbf{u_I})(\mathbf{R_{II}}\mathbf{A}_1^*(t, \mathbf{u_I}))^* dt. \qquad (144)$$

The formula in (144) indicates that $\mathbf{R_{II}} = 0$ is a solution, plugging which into (143) results in $\bar{\mathbf{u}}_{\mathbf{II}} = \mathbf{\Lambda}^*$. This means, by knowing the perfect model, the estimated parameters in (141) and (142) and (143) and (144) under certain conditions will converge to the truth.

As a simple test example, we consider estimating the three parameters σ, ρ and γ in the noisy L-63 model (31) with $\rho = 28, \sigma = 10, \beta = 8/3$. Putting into the framework (141) and (142), the augmented system becomes

$$\begin{aligned}
dx &= \sigma(y - x)dt + \sigma_x dW_x, \\
dy &= \big(x(\rho - z) - y\big)dt + \sigma_y dW_y, \\
dz &= (xy - \beta z)dt + \sigma_z dW_z, \\
d\sigma &= 0, \\
d\rho &= 0, \\
d\beta &= 0,
\end{aligned} \qquad (145)$$

where $\mathbf{u_I} = (x, y, z)^T$ and $\mathbf{\Lambda} = (\sigma, \rho, \beta)^T$. Figure 28 shows the parameter estimation skill with $\sigma_x = \sigma_y = \sigma_z = 5$. Here, the initial guess of all the three parameters are $\rho = \sigma = \beta = 0$ and the initial uncertainty is a 3×3 identity matrix. It is clear that the estimated parameter converge quickly to the truth and the corresponding uncertainty in the estimated parameters goes to zero. In Figure 29, the parameter estimation skill with different noise level is compared. When the noise level σ_x, σ_y and σ_z in the observed signals become larger, the convergence rate becomes slower.

In [28], systematic studies of both one-dimensional linear and nonlinear systems have revealed that the convergence rate depends on different factors, such as the observability, the signal-to-noise ratio and the initial uncertainty. In particular, both theoretical analysis and numerical simulations in [28] showed that the convergence rate can be extremely slow and sometimes with an undesirable initial guess the estimation even converges to a wrong solution for some highly nonlinear systems. Thus, alternative parameter estimation scheme is required.
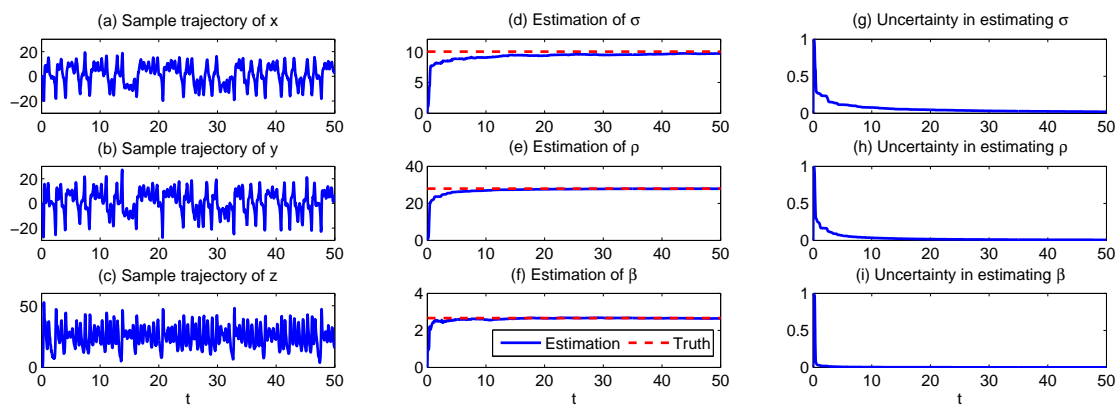
Figure 28. Parameter estimation of the noisy L-63 system using the direct approach (145) with $\sigma_x = \sigma_y = \sigma_z = 5$. (**a**–**c**) sample trajectories of x, y, z which are used as observations; (**d**–**f**) time evolution of the estimated parameters and the truth (red); (**g**–**i**) the uncertainty evolution of each estimated parameter.
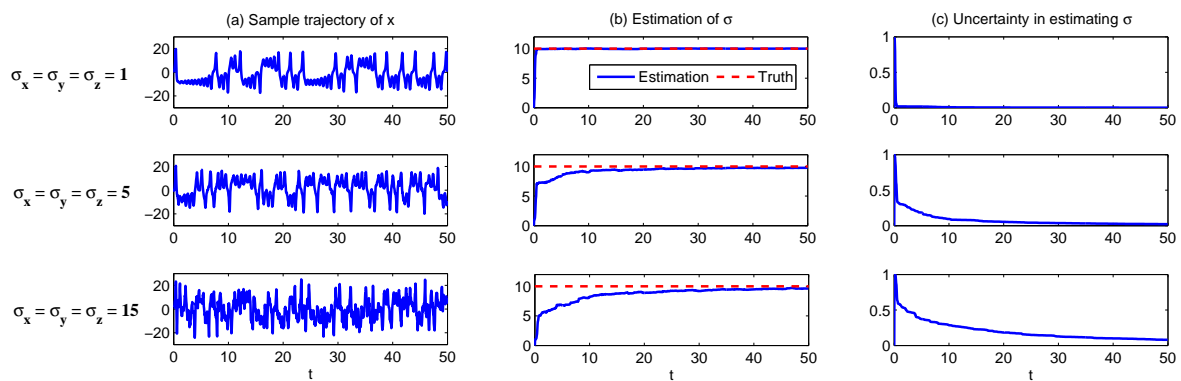

Figure 29. Parameter estimation of the noisy L-63 system using the direct approach (145) with $\sigma_x = \sigma_y = \sigma_z = 1$, 5 and 15 for the three rows. For illustration purposes, only the trajectory of x and the estimation of σ is shown in all the three cases. (**a**) sample trajectory of x; (**b**,**c**) mean estimation of σ and the associated uncertainty.

6.1.2. Parameter Estimation Using Stochastic Parameterized Equations

Instead of augmenting the equation in a trivial but natural way for the parameters as shown in (142), a new approach of the augmented system can be formed in the following way [28]:

$$d\mathbf{u_I} = [\mathbf{A}_0(t, \mathbf{u_I}) + \mathbf{A}_{1,\lambda}(t, \mathbf{u_I})\mathbf{\Lambda}]dt + \mathbf{\Sigma_I}(t, \mathbf{u_I})d\mathbf{W_I}(t), \tag{146}$$

$$d\mathbf{\Lambda} = [\mathbf{c}_1\mathbf{\Lambda} + \mathbf{c}_2]dt + \sigma_{\mathbf{\Lambda}}d\mathbf{W}_{\mathbf{\Lambda}}(t). \tag{147}$$

Here, $\mathbf{c}_1$ is a negative-definite diagonal matrix, $\mathbf{c}_2$ is a constant vector and $\sigma_{\mathbf{\Lambda}}$ is a diagonal noise matrix. Different from (142), a stochastic parameterized equation is used in (147) to describe the time evolution of each component of the parameter $\mathbf{\Lambda}_i$. This approach is motivated from the stochastic parameterized extended Kalman filter model [64,65] as discussed in Section 4.5.1. The stochastic parameterized equations in (147) serve as the prior information of the parameter estimation. Although certain model error will be introduced in the stochastic parameterized equations due to the appearance of $\mathbf{c}_1$, $\mathbf{c}_2$ and $\sigma_{\mathbf{\Lambda}}$, it has shown in [28] that the convergence rate will be greatly accelerated. In fact, in linear models, rigorous analysis reveals that the convergence rate using stochastic parameterized Equation (146) and (147) is exponential while that using the direct method (146) and (147) is only algebraic.

Now, we apply the parameter estimation using stochastic parameterized Equation (146) and (147) for the noisy L-63 model. The augmented system reads,

$$
\begin{aligned}
dx &= \sigma(y - x)dt + \sigma_x dW_x, \\
dy &= \big(x(\rho - z) - y\big)dt + \sigma_y dW_y, \\
dz &= (xy - \beta z)dt + \sigma_z dW_z, \\
d\sigma &= -d_\sigma(\sigma - \hat\sigma) + \sigma_\sigma dW_\sigma, \\
d\rho &= -d_\rho(\rho - \hat\rho) + \sigma_\rho dW_\rho, \\
d\beta &= -d_\beta(\beta - \hat\beta) + \sigma_\beta dW_\beta.
\end{aligned}
\tag{148}
$$

The same initial values are taken as those in the direct method. Recall in Figure 29 that the system with a large observational noise $\sigma_x = \sigma_y = \sigma_z = 15$ leads to a slow convergence rate. Below, we focus on this case and use the parameter estimation scheme (148) to improve the result. The parameters in the stochastic parameterized equation are chosen as follows:

$$
\begin{aligned}
\hat\sigma = 10 + 2 = 12, \qquad \hat\rho &= 28 + 5.6 = 33.6, \qquad \hat\beta = 8/3 + 1.6/3 = 3.2, \\
d_\sigma = d_\rho = d_\beta = 0.5, \qquad \sigma_\sigma &= 2, \qquad \sigma_\rho = 5.6, \qquad \sigma_\beta = 1.6/3.
\end{aligned}
\tag{149}
$$

Here, we introduce 20% errors in all the mean states $\hat\sigma, \hat\rho$ and $\hat\beta$. The variance of the stochastic parameters is given by $\sigma_\sigma^2/(2d_\sigma), \sigma_\rho^2/(2d_\rho)$ and $\sigma_\beta^2/(2d_\beta)$, respectively. Therefore, the truth is located at one standard deviation (square root of the variance) of the mean state. The decorrelation time of all the process is $1/d_\sigma = 1/d_\rho = 1/d_\beta = 2$ time units.

Figure 30 shows the results of parameter estimation using the method with stochastic parameterized Equations (146) and (147). Despite the error in $\hat\sigma, \hat\rho$ and $\hat\beta$ as the prior information, the estimation of the parameters is quite accurate. In fact, by using the averaged value over the time interval $t \in [10, 50]$ as the estimation, we compare the estimated parameters with the truth:

$$
\begin{aligned}
\text{Estimation:} \qquad \sigma = 10.9, \qquad \rho &= 29.1, \qquad \beta = 2.88, \\
\text{Truth:} \qquad \sigma = 10.0, \qquad \rho &= 28.0, \qquad \beta = 2.67.
\end{aligned}
\tag{150}
$$
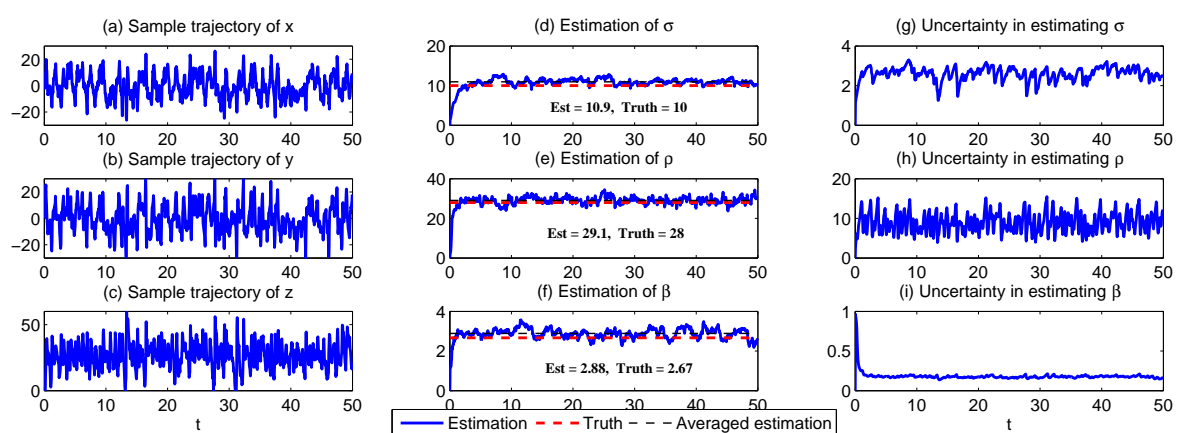
Figure 30. Parameter estimation of the noisy L-63 system using the method with stochastic parameterized Equations (146) and (147) with $\sigma_x = \sigma_y = \sigma_z = 15$. (**a–c**) sample trajectories of x, y, z which are used as observations; (**d–f**) time evolution of the estimated parameters and the truth (red); the thin black line shows the average of the estimation from $t = 10$ to $t = 50$; (**g**)–(**i**) the uncertainty evolution of each estimated parameter.

The relative error in the three estimated parameters compared with the truth is 9.0%, 3.9% and 7.8%, respectively, all of which are much less than 20% in $\hat{\sigma}, \hat{\rho}$ and $\hat{\beta}$. This is because the conditional Gaussian framework (1) and (2) automatically makes use of the information in both the observations and the stochastic parameterized equations. The combination of these two components reduces the error even if the prior knowledge in the stochastic parameterized equations is biased (as in most of the real-world applications). Notably, the convergence of the estimated parameters as shown in Figure 30 is much faster than the direct method. This is one of the key features of the approach using stochastic parameterized equations and is practically quite useful since only limited observational data is known in many applications.

It is also important to note that the perfect model is unknown in most realistic situations. Therefore, dealing with model error in the parameter estimation is a crucial topic. The judicious model error in the approach using stochastic parameterized equations allows parameter uncertainties and provides more robust results compared with the direct method. Thus, the approach using stochastic parameterized equations has a wider application in real-world problems. More comparison between the two methods developed here can be found in [28].

6.1.3. Estimating Parameters in the Unresolved Processes

As discussed in Sections 4.5.1 and 6.1.2, stochastic parameterizations are widely used in describing turbulent signals as observed in nature. One important step is to estimate the parameters in the equations associated with these stochastic parameterizations (e.g, $\sigma_\gamma, \hat{\gamma}, d_\gamma, \sigma_\omega, \hat{\omega}, d_\omega, \sigma_b, \hat{b}$ and d_b in (119)). Note that there are typically no direct observations of these stochastic parameterized processes since they usually represent unresolved variables or they may not have clear physical correspondence. However, they play important roles in describing the underlying complex turbulent dynamical system. An accurate estimation of these parameters is crucial. See examples in Sections 4.5.1 and 6.1.2 and those in [65,101].

Traditional methods in estimating these parameters include Markov Chain Monte Carlo (MCMC) [198–201], maximum likelihood estimation [202,203] and the ensemble Kalman filter [204,205]. Note that, if both the state variables γ, ω, b and the associated 9 parameters in these three equations are treated as unobserved variables $\mathbf{u}_{\mathrm{II}}$, then the augmented system does not belong to the conditional Gaussian model family. Thus, the closed form in (4) and (5) cannot be applied in the two methods discussed in Sections 6.1.1 and 6.1.2 and ensemble Kalman filter has to be adopted in these methods, which is computationally expensive.

Nevertheless, a judicious application of the conditional Gaussian framework still allows an efficient parameter estimation algorithm for the parameters in the stochastic parameterized equations. Below, we discuss the idea using the so-called SPEKF-M model, which is a simplified version of the SPEKF model (119) and contains only the multiplicative noise process γ (and "M" stands for multiplicative). The model reads

$$
\begin{aligned}
du &= \Big(-\gamma u + F(t) \Big) dt + \sigma_u dW_u, \\
d\gamma &= -d_\gamma (\gamma - \hat{\gamma}) dt + \sigma_\gamma dW_\gamma.
\end{aligned}
\tag{151}
$$

The goal here is to estimate the three parameters σ_γ, d_γ and $\hat{\gamma}$ by observing only one sample trajectory of u. First, choose an arbitrary initial estimation of the three parameters $\sigma_\gamma^{(0)}, d_\gamma^{(0)}$ and $\hat{\gamma}^{(0)}$. Then, run the conditional Gaussian filter (4) and (5), which gives a time series of the conditional mean $\overline{\gamma}^{(0)}(t)$ and another time series of the conditional variance $R^{(0)}(t)$ of $\gamma(t)$ conditioned on the trajectory of u. With the $\overline{\gamma}^{(0)}(t)$ and $R^{(0)}(t)$, applying the same argument as that in (124) but changing the L Gaussian distributions associated with L different trajectories by those at L different time instants, the distribution $p(\gamma^{(0)})$ can be obtained via a Gaussian mixture. Here, $p(\gamma^{(0)})$ is also a Gaussian distribution due to its dynamical structure in (151). This Gaussian distribution provides two quantities—the mean $\mu_{\gamma,eq}^{(0)}$ and the variance $R_{\gamma,eq}^{(0)}$. In addition, the sample autocorrelation function

$\tau_{\gamma,eq}^{(0)}$ of the conditional mean $\overline{\gamma}^{(0)}(t)$ can be easily computed. Therefore, these three quantities are used to update the iteration of the three parameters σ_γ, d_γ and $\hat{\gamma}$ with the relation

$$\frac{(\sigma_\gamma^{(1)})^2}{2d_\gamma^{(1)}} = R_\gamma^{(0)}, \qquad \hat{\gamma}^{(1)} = \mu_{\gamma,eq}^{(0)}, \qquad d_\gamma^{(1)} = \frac{1}{\tau_{\gamma,eq}^{(0)}}. \tag{152}$$

Then, run the conditional Gaussian filter (4) and (5) again using the same observed time series $u(t)$ but with the updated parameters $\sigma_\gamma^{(1)}, d_\gamma^{(1)}$ and $\hat{\gamma}^{(1)}$. Repeating the above procedure leads to another update of the parameters $\sigma_\gamma^{(2)}, d_\gamma^{(2)}$ and $\hat{\gamma}^{(2)}$. Continue this process until the updates of the parameters converge. This fixed point iteration results in an efficient parameter estimation algorithm.

As a simple test example, the following parameters are adopted to generate the true signal in the SPEKF-M model (151),

$$\sigma_\gamma = 0.5, \qquad d_\gamma = 0.5, \qquad \hat{\gamma} = 1, \qquad \sigma_u = 0.5, \qquad F = 2. \tag{153}$$

A pair of sample trajectories $(u(t), \gamma(t))$ is shown in Panel (a) of Figure 31 and the corresponding PDFs are shown in Panel (b). Clearly, these parameters result in a highly intermittent time series of $u(t)$ with a non-Gaussian PDF with an one-side fat tail. Panel (c) shows the conditional mean and conditional variance of γ conditioned on the trajectory of u with the perfect parameters. The conditional mean does not equal the truth, but the conditional mean $\mu(t)$ at the intermittent phases of u are close to the truth of γ with a small conditional variance. On the other hand, at quiescent phases, $\mu(t)$ differs from the truth while the uncertainty (conditional variance) in the estimation is large.

Now, we apply the algorithm above to estimate the three parameters σ_γ, d_γ and $\hat{\gamma}$ assuming the other two parameters $\sigma_u = 0.5$ and $F = 2$ are known. The trajectory of u (with the length of 1000 time units) in Panel (a) of Figure 31 is used as the input of the algorithm. The iteration curve of the algorithm is shown in Figure 32, where the initial values $\sigma_\gamma^{(0)} = 0.1, d_\gamma^{(0)} = 1$ and $\hat{\gamma}^{(0)} = 2.5$ are far from the truth. It is clear that, after five iterations, the update of the parameters converges to fixed points. The parameter $\hat{\gamma}$ converges exactly to the truth while both d_γ and σ_γ are slightly larger than the true values. Nevertheless, the variance of γ using the estimated parameters, namely $\sigma_\gamma^2/(2d_\gamma) = 0.61^2/(2 \cdot 0.66) = 0.28$, is close to that of the truth, which is $0.5^2/(2 \cdot 0.5) = 0.25$. The slightly overestimate of the variance compensates the slightly underestimate of the damping time $1/d_\gamma$ and therefore the model with the estimated parameters is able to generate essentially the same intermittent signals as in the observed period of u in Figure 31. Note that since the input of the time series of u is only of finite length (and is actually pretty short here), the estimated parameters reflect the features in the observed signal, which may be slightly different from the true underlying dynamics unless an infinitely long time series is used as the input. Next, the skill of the estimated parameters is examined via the model statistics. In Figure 33, both the trajectories and the associated PDFs of u and γ using the true parameters and the estimated parameters are shown. Due to the high skill in estimating both the mean and variance of γ, the PDFs of γ^M and γ are quite close to each other with only a slight overestimation of the variance in γ^M. The decorrelation time of the trajectory of γ^M is slightly shorter than that of γ as discussed above. Nevertheless, γ^M and γ provide quite similar feedback to u^M and u. Therefore, the statistical features in u^M, including the decorrelation time, mean, variance and highly non-Gaussian PDF, are almost the same as those in u.
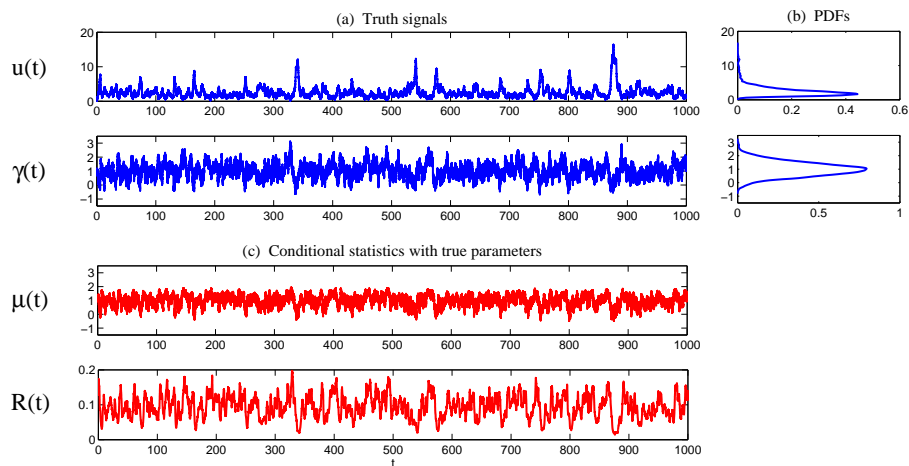
Figure 31. (**a**) a pair of sample trajectories $(u(t), \gamma(t))$ of the SPEKF-M model (151) with parameters in (153); (**b**) the corresponding PDFs. Note that the PDF of u is high non-Gaussian with an one-side fat tail; (**c**) the conditional mean $\mu(t)$ and conditional variance $R(t)$ using the model with perfect parameters.

A more detailed study of this algorithm, including the application to the full SPEKF model (119) and the convergence rate dependence on different model ingredients, will be included in a follow-up work.
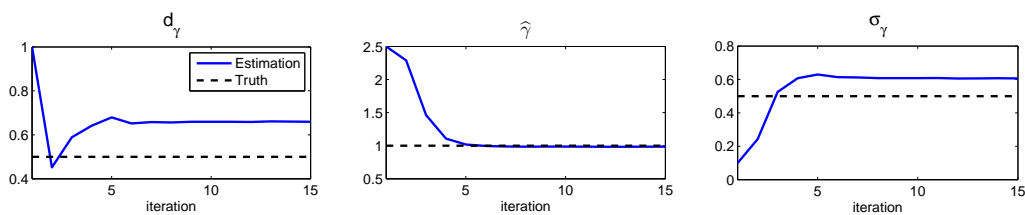

Figure 32. Iteration of the updates of the parameters (blue curve). The initial values $\sigma_\gamma^{(0)} = 0.1, d_\gamma^{(0)} = 1$ and $\hat{\gamma}^{(0)} = 2.5$ are far from the truth (black dashed lines).
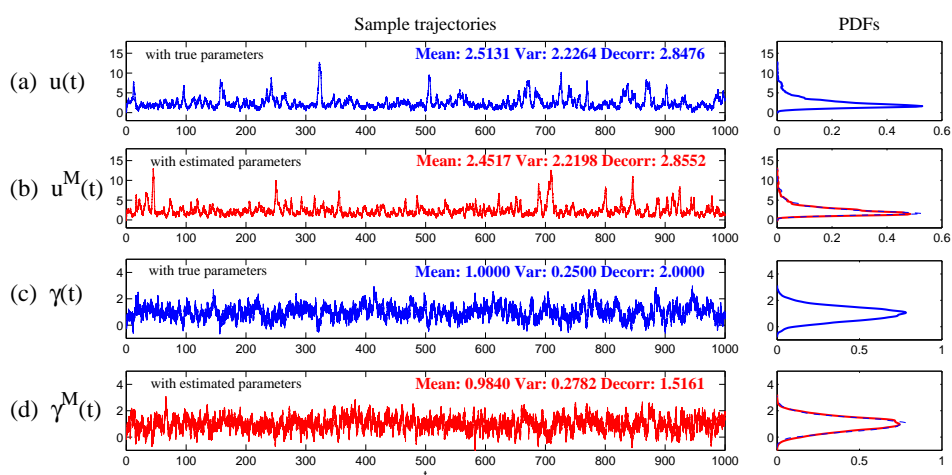

Figure 33. Sample trajectories and the associated PDFs using the true parameters (blue) and the estimated ones (red). (**a**) $u(t)$; (**b**) $u^M(t)$; (**c**) $\gamma(t)$; (**d**) $\gamma^M(t)$. The PDFs are calculated based on a long trajectory with $10,000$ time units but only 1000 time units are shown here. In the panels of the PDFs of u^M and γ^M, those of u and γ are also illustrated (thin dashed blue curves) for comparison.

6.2. Data Assimilation with Physics-Constrained Forecast Models and Information Theory for Quantifying Model Errors

6.2.1. An Information Theoretical Framework for Data Assimilation and Prediction

The following two traditional path-wise measurements are widely used in assessing the skill in data assimilation and prediction [15,206–210]. Denote u_i, $i = 1, \ldots, n$ the true signal and $\hat{u}_i$ the filtering/prediction estimate. These measurements are given by

1. The root-mean-square error (RMSE):

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{u}_i - u_i)^2}{n}}. \tag{154}$$

2. The pattern correlation (PC):

$$\mathrm{PC} = \frac{\sum_{i=1}^{n}(\hat{u}_i - \overline{\hat{u}}_i)(u_i - \overline{u}_i)}{\sqrt{\sum_{i=1}^{n}(\hat{u}_i - \overline{\hat{u}}_i)^2}\sqrt{\sum_{i=1}^{n}(u_i - \overline{u}_i)^2}}, \tag{155}$$

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

While these two path-wise measurements are easy to implement and are able to quantify the data assimilation and prediction skill to some extent, they have fundamental limitations. It has been shown in [42,211] that these two measurements fail to quantify the skill of capturing the extreme events and other non-Gaussian features, which lead to misleading results. In fact, concrete examples even in the Gaussian models [42,211] showed that two different predictions can have the same RMSE and PC, but one is way more skillful than the other in capturing the extreme events.

Due to the fundamental limitations of assessing the data assimilation and prediction skill based only on these two traditional path-wise measurements, various information measurements have been proposed to improve the quantification of the data assimilation and prediction [16,194,195,212–218]. In [218–222], an information measurement called Shannon entropy difference was introduced and was used to assess the data assimilation and prediction skill. The Shannon entropy difference takes into account the estimation of the extreme events and improves the insufficiency in the traditional path-wise measurements. However, relying solely on the Shannon entropy difference in assessing the data assimilation and prediction skill is also misleading. In fact, the Shannon entropy difference fails to capturing the model error in the mean state and it computes the uncertainty of the two distributions separately rather than considering the pointwise difference between the two PDFs.

Due to the fundamental limitations in the two classical path-wise measurement, RMSE and PC, as well as those in the Shannon entropy difference, a new information-theoretic framework [186] has developed to assess the data assimilation/prediction skill. Denote $\pi \equiv \pi(u)$ and $\pi^M \equiv \pi(u^M)$ the PDFs associated with truth u and the data assimilation/prediction estimate u^M, respectively. Denote $p(u, u^M)$ the joint PDF of u and u^M. Let $\mathcal{U} = u - u^M$ be the residual between the truth and the estimate. This information-theoretic framework involves three information measurements:

1. The Shannon entropy residual,

$$\mathcal{S}(\mathcal{U}) = -\int p(\mathcal{U}) \log p(\mathcal{U}). \tag{156}$$

2. The mutual information,

$$\mathcal{M}(\pi, \pi^M) = \iint p(u, u^M) \log\left(\frac{p(u, u^M)}{\pi(u)\pi(u^M)}\right). \tag{157}$$

3.　The relative entropy,

$$\mathcal{R}(\pi, \pi^M) = -\int \pi \log\left(\frac{\pi}{\pi^M}\right). \tag{158}$$

The Shannon entropy residual quantifies the uncertainty in the point-wise difference between u and u^M. It is an information surrogate of the RMSE in the Gaussian framework. The mutual information quantifies the dependence between the two processes. It measures the lack of information in the factorized density $\pi(u)\pi(u^M)$ relative to the joint density $p(u, u^M)$, which follows the identity,

$$\mathcal{M}(\pi, \pi^M) = \mathcal{P}\Big(p(u, u^M), \pi(u)\pi(u^M)\Big). \tag{159}$$

The mutual information is an information surrogate of the PC in the Gaussian framework. On the other hand, the relative entropy quantifies the lack of information in π^M related to π and it is a good indicator of the skill of u^M in capturing the peaks and extreme events of u. It also takes into account the pointwise discrepancy between π and π^M rather than only computing the difference between the uncertainties associated with the two individual PDFs (as in the Shannon entropy difference). Therefore, the combination of these three information measurements is able to capture all the features in assessing the data assimilation/prediction skill and overcomes the shortcomings as discussed in the previous subsection. Note that, when $\pi \sim \mathcal{N}(\overline{u}, R)$ and $\pi^M \sim \mathcal{N}(\overline{u}^M, R_M)$ are both Gaussian, then the above three information measurements have explicit expressions [215].

The information-theoretic framework (156)–(158) is usually defined in the super-ensemble sense [215]. Note that in many practical issues only one realization (trajectory) is available. Nevertheless, the information-theoretic framework can also be used in a pathwise way, where the statistics are computed by collecting all the sample points in the given realization. Some realistic applications of the information-theoretic framework for filtering and prediction can be found in [42,63,215].

6.2.2. Important Roles of Physics-Constrained Forecast Models in Data Assimilation

It has been shown in Section 3.3 that the physics-constrained nonlinear stochastic models are the recent development for data-driven statistical models with partial observations. The physics-constrained nonlinear stochastic models overcome the finite-time blowup and the lack of physical meaning issues in various ad hoc multi-layer regression models [31,32]. Here, our goal is to show that the physics-constrained nonlinear stochastic models also play important role in data assimilation (or filtering). Ignoring the energy-conserving nonlinear interactions in the forecast models will result in large model errors.

The test model below is a simple dyad model, which mimics the interactions between resolved large-scale mean flow and unresolved turbulent fluctuations with intermittent instability [37,50,223]. The dyad model reads

$$du = (-d_u u + \gamma u v + F_u)dt + \sigma_u dW_u, \tag{160}$$

$$dv = (-d_v v - \gamma u^2)dt + \sigma_v dW_v. \tag{161}$$

In (160) and (161), u is regarded as representing one of the resolved modes in a turbulent signal, which interacts with the unresolved mode v through quadratic nonlinearities. The conserved energy in the quadratic nonlinear terms in (160) and (161) is easily seen. Below, the physics-constrained dyad model (160) and (161) is utilized to generate true signals of nature. The goal here is to filter the unobserved process v given one single realization of the observed process u. In addition to adopting the perfect filter (160) and (161), an imperfect filter with no energy-conserving nonlinear interactions is

studied for comparison. In this imperfect filter, the nonlinear feedback $-\gamma u^2$ in v is dropped and the result is a stochastic parameterized filter [20],

$$du = (-d_u u + \gamma u v + F_u)dt + \sigma_u dW_u, \tag{162}$$
$$dv = -d_v^M(v - \bar{v}^M)dt + \sigma_v^M dW_v. \tag{163}$$

In the stochastic parameterized filter (162) and (163), the parameters in the resolved variable u are assumed to be the same as nature (160) and (161). We further assume the statistics of the unobserved variable v of nature (160) and (161) are available. Thus, the parameters $d_v^M, \bar{v}^M$ and σ_v^M in the unresolved process v are calibrated [187,188,224] by matching the mean, variance and decorrelation time of those in (160) and (161). Note that both (160) and (161) and (162) and (163) belong to the conditional Gaussian framework (1) and (2) by denoting $\mathbf{u_I} = u$ and $\mathbf{u_{II}} = v$ and (4) and (5) is used to efficiently calculate the filter estimates.

Note that in (160) and (161), if $F_u = 0$, then the fixed point associated with the deterministic part is $u_c = v_c = 0$. This leads to an important issue in the state estimation of v, namely the observability [207,225]. The coupled system (160) and (161) is said to lose its observability if the observed process u provides no information in determining the unobserved variable v. Intuitively, this corresponds to $u = 0$ in (160) and (161), in which case v disappears in the observed process u. Therefore, if $F_u = 0$, then the filtering skill of v is expected to deteriorate especially with a small σ_u. Below, we consider two different dynamical regimes:

$$
\begin{array}{llllll}
\text{Regime I:} & d_u = 0.8, & d_v = 0.8, & \gamma = 1.2, & \text{and} & F_u = 1, \\
\text{Regime II:} & d_u = 0.8, & d_v = 0.8, & \gamma = 1.2, & \text{and} & F_u = 0.
\end{array}
\tag{164}
$$

In Regime II, there is no practical observability in the quiescent phase (near the fixed point associated with the deterministic model) while in Regime I the forcing drives the signal out of the value with $u_c = 0$.

Figure 34 shows the model error in terms of RMSE, PC and relative entropy as a function of σ_u and σ_v. Both the perfect physics-constrained forecast model (160) and (161) and the stochastic parameterized filter (162) and (163) are used. Here, instead of showing the Shannon entropy residual and the mutual information, we still use the RMSE and PC since most readers are familiar with these two measurements. Nevertheless, the readers should keep in mind that the Shannon entropy residual and the mutual information are more suitable measurements in the non-Gaussian framework. On the other hand, the relative entropy is shown here in assessing the model error.

First, in Columns (a,b) of Figure 34, the filtering skill in Regime I with $F_u = 1$ is illustrated. With a small σ_u, both filters have skillful estimation. This is because when σ_u is small, the filters trust more towards the observational process, which has a large signal to noise ratio and therefore it provides accurate estimates. However, when σ_u is large but σ_v is small in generating the true signal, the stochastic parameterized filter becomes much worse than the perfect filter using physics-constrained forecast model. In fact, a large σ_u leads to large signals in u and it also tells the filter to trust more towards the underlying process of v. This implies the filter estimate of v is then essentially driven by the feedback term $-\gamma u^2$. Since the stochastic parameterized filter has no such feedback mechanism, the error becomes large. See Panel (b) of Figure 35 for an example with path-wise trajectories. It is also important to note that in such a situation the PDF of the filter estimate is completely different from the truth and thus a large model error is found.

Columns (c,d) of Figure 34 show the filter estimates in Regime II with $F_u = 0$. Compared with the results in Regime I, it is clear that when σ_u is small in generating the true signal, the filter estimates become inaccurate. This is in fact due to the observability issue since small σ_u means the signal of u stays near original. This is clearly seen in Panel (a) of Figure 36, where the filter estimate is accurate only at intermittent phases. One interesting phenomenon is that although the filter estimate using the stochastic parameterized filter in Panel (b) of Figure 35 has a smaller RMSE compared with that

in Panel (a) of Figure 36, the relative entropy clearly indicates a much worse filter estimates in the former case since they fail to capture any of the amplitudes. These facts all indicate the importance in including the physics-constrained structure in designing filters especially in the regimes that are dominated by the energy-conserving nonlinear interactions.
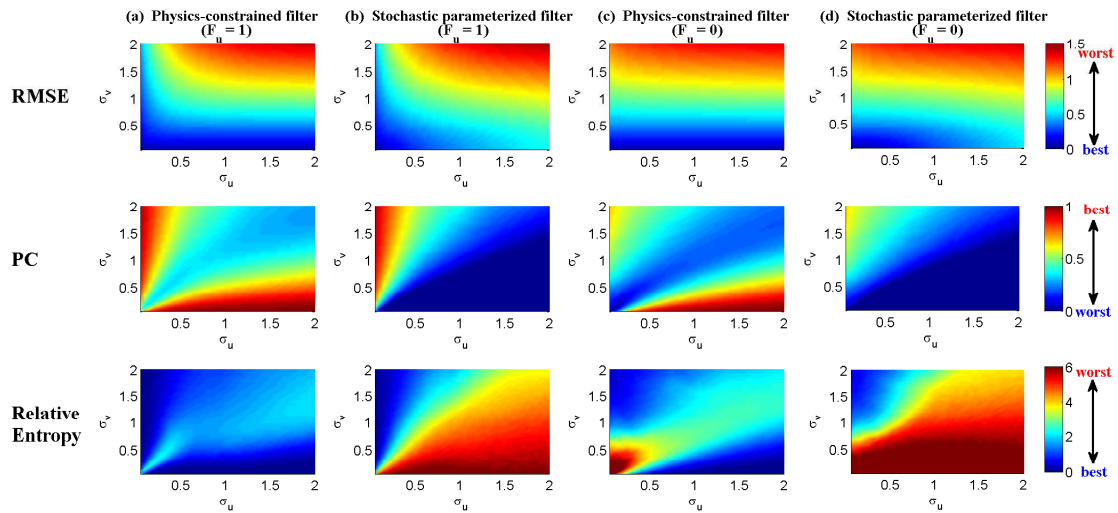

Figure 34. Model error in terms of RMSE, PC and relative entropy as a function of σ_u and σ_v in the true signal. (**a,b**) Regime I with $F_u = 1$; (**c,d**) Regime I with $F_u = 0$; (**a,c**) show the model error using the perfect physics-constrained forecast model (160) and (161) while (**b,d**) uses the stochastic parameterized filter (162) and (163).
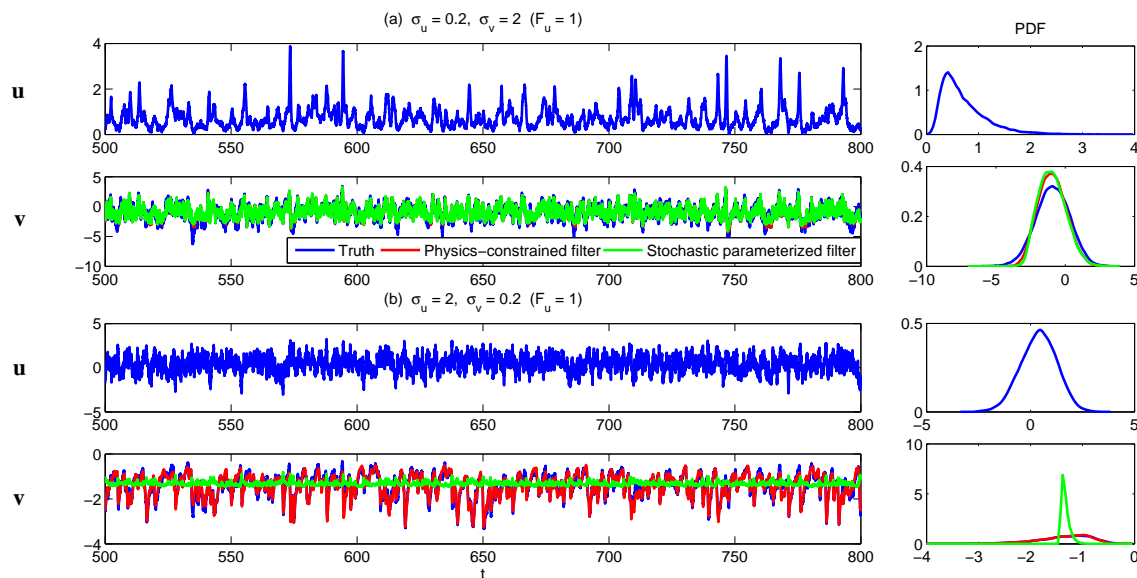

Figure 35. Regime I with $F_u = 1$. Path-wise trajectories and the PDFs of the true signal (blue) and the filter estimation of the physics-constrained forecast model (160) and (161) (red) and the stochastic parameterized filter (162) and (163) (green). (**a**) $\sigma_u = 0.2$ and $\sigma_v = 2$; (**b**) $\sigma_u = 2$ and $\sigma_v = 0.2$.
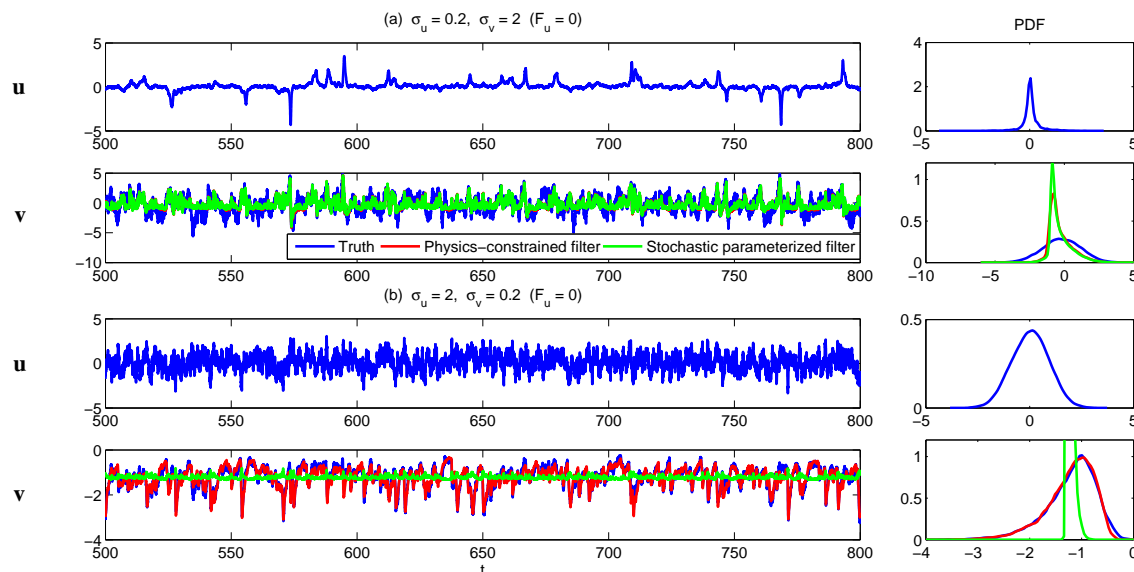
Figure 36. Regime II with $F_u = 0$. Path-wise trajectories and the PDFs of the true signal (blue) and the filter estimation of the physics-constrained forecast model (160) and (161) (red) and the stochastic parameterized filter (162) ans (163) (green). (**a**) $\sigma_u = 0.2$ and $\sigma_v = 2$; (**b**) $\sigma_u = 2$ and $\sigma_v = 0.2$.

Finally, it is important to note that although the stochastic parameterized filter (162) and (163) is not able to recover the signal due to the strong feedback from the physics-constrained nonlinear interactions, the stochastic parameterized filter (162) and (163) is still quite useful in detecting the intermittent phases in turbulent signals. In fact, in Panel (a) of both Figures 35 and 36, the intermittent phases are all accurately recovered by the stochastic parameterized filter. Other works showing the skillful behavior of the stochastic parameterized filter and its advantages over the mean stochastic models can be found in [16,64,65].

6.3. Multiscale Data Assimilation with Particles Interacting with Conditional Gaussian Statistics

6.3.1. A General Description

Data assimilation of turbulent signals is an important challenging problem because of the extremely complicated large dimension of the signals and incomplete partial noisy observations which usually mix the large scale mean flow and small scale fluctuations. See Chapter 7 of [20] for examples of new phenomena due to this multiscale coupling through the observations even for linear systems. Due to the limited computing power, it is desirable to use multi-scale forecast models which are cheap and fast to mitigate the curse of dimensionality in turbulent systems. Thus, model errors from imperfect forecast models are unavoidable in the development of a data assimilation method in turbulence.

Among different methods, conventional superparameterization is a multi-scale algorithm that was originally developed for the purpose of parameterizing unresolved cloud process in tropical atmospheric convection [73,226,227]. This conventional superparameterization resolves the large scale mean flow on a coarse grid in a physical domain while the fluctuating parts are resolved using a fine grid high resolution simulation on periodic domains embedded in the coarse grid. A much cheaper version of superparameterization, called stochastic superparameterization [70–73], replaces the nonlinear eddy terms by quasilinear stochastic processes on formally infinite embedded domains where the stochastic processes are Gaussian conditional to the large scale mean flow. The key ingredient of these multiscale data assimilation methods is the systematic use of conditional Gaussian mixtures which make the methods efficient by filtering a subspace whose dimension is smaller than the full state. This conditional Gaussian closure approximation results in a seamless algorithm without

using the high resolution space grid for the small scales and is much cheaper than the conventional superparameterization, with significant success in difficult test problems [71,72,74] including the MMT model [71,75] and ocean turbulence [76–78].

The key idea of the multiscale data assimilation method is to use conditional Gaussian mixtures [80,228] whose distributions are compatible with superparameterization. The method uses particle filters (see [79] and Chapter 15 of [20]) or ensemble filters on the large scale part [75,76] whose dimension is small enough so that the non-Gaussian statistics of the large scale part can be calculated from a particle filter, whereas the statistics of the small scale part are conditionally Gaussian given the large scale part. This framework is not restricted to superparameterization as the forecast model and other cheap forecast models can also be employed. See [80] for another multiscale filter with quasilinear Gaussian dynamically orthogonality method as the forecast method in an adaptively evolving low dimensional subspace without using superparameterization. Note that data assimilation using superparameterization has already been discussed in [229] with noisy observations of the large scale part of the signal alone. There it was shown that, even in this restricted setting, ignoring the small scale fluctuations even when they are rapidly decaying can completely degrade the filter performance compared with the high skill using superparameterization. Here, in contrast to [229], we consider multiscale data assimilation methods with noisy observations with contributions from both the large and small scale parts of the signal, which is a more difficult problem than observing only the large scale because it requires accurate estimation of statistical information of the small scales [75,76,230]. In addition, mixed observations of the large and small scale parts occur typically in real applications. For example, in geophysical fluid applications, the observed quantities such as temperature, moisture, and the velocity field necessarily mix both the large and small scale parts of the signal [20,231]. Thus, the multiscale filtering also provides a mathematical framework for representation errors, which are due to the contribution of unresolved scales [81,82] in the observations.

6.3.2. Particle Filters with Superparameterization

Superparameterization retains the large scale variables by resolving them on a coarse grid while the effect of the small scales on the large scales is parameterized by approximating the small scales on local or reduced spaces. Stochastic superparameterization discussed in the previous section uses Gaussian closure for the small scales conditional to the large scale variable $\overline{\mathbf{u}}$ with $\overline{\mathbf{u}} \in \mathbb{R}^N$ [70–73]. Thus, we consider a multi-scale filtering algorithm with forecast prior distributions given by the conditional distribution

$$p^f(\mathbf{u}) = p^f(\overline{\mathbf{u}}, \mathbf{u}') = p^f(\overline{\mathbf{u}}) p_G^f(\mathbf{u}'|\overline{\mathbf{u}}), \tag{165}$$

where $p_G^f(\mathbf{u}'|\overline{\mathbf{u}})$ is a Gaussian distribution conditional to $\overline{\mathbf{u}}$

$$p_G^f(\mathbf{u}'|\overline{\mathbf{u}}) = \mathcal{N}(\mathbf{u}'(\overline{\mathbf{u}}), \mathbf{R}'(\overline{\mathbf{u}})). \tag{166}$$

Here, we assume that N_1 is sufficiently small enough that particle filters (see Chapter 15 of [20]) can be applied to the large scales. For a low dimensional space $\overline{\mathbf{u}}$, the marginal distribution of $\overline{\mathbf{u}}$ can be approximated by Q particles

$$p^f(\overline{\mathbf{u}}) = \sum_{j=1}^{Q} p_j^f \delta(\overline{\mathbf{u}} - \overline{\mathbf{u}}_j), \tag{167}$$

where $p_j^f \geq 0$ are particle weights such that $\sum_j p_j^f = 1$. After the forecast step where superparameterization is applied to each particle member, we have the following general form for the prior distribution $p^f(\mathbf{u})$

$$p^f(\mathbf{u}) = p^f(\overline{\mathbf{u}}, \mathbf{u}') = \sum_{j=1}^{Q} p_j^f \delta(\overline{\mathbf{u}} - \overline{\mathbf{u}}_j) p_G^f(\mathbf{u}'^f | \overline{\mathbf{u}}_j)$$

$$= \sum_{j=1}^{Q} p_j^f \delta(\overline{\mathbf{u}} - \overline{\mathbf{u}}_j) \mathcal{N}(\mathbf{u}'(\overline{\mathbf{u}}_j)^f, \mathbf{R}'(\overline{\mathbf{u}}_j)^f), \tag{168}$$

which is a conditional Gaussian mixture distribution where each summand is a Gaussian distribution conditional to $\overline{\mathbf{u}}_j$. The Gaussian mixture has already been used in data assimilation [232–234] but the multi-scale method developed here is different in that conditional Gaussian distributions are applied in the reduced subspace $\mathbf{u}'$ with particle approximations only in the lower dimensional subspace $\overline{\mathbf{u}}$. Thus, the proposed multi-scale data assimilation method can be highly efficient and fast in comparison with conventional data assimilation methods which use the whole space for the filter.

Assume that the prior distribution from the forecast is in the from (168) and that the observations have the following structure:

$$\mathbf{v} = G(\overline{\mathbf{u}}, \mathbf{u}') + \sigma_\theta = \overline{\mathbf{G}}\overline{\mathbf{u}} + \mathbf{G}'(\overline{\mathbf{u}})\mathbf{u}' + \sigma_\theta, \tag{169}$$

where $\mathbf{G}'(\overline{\mathbf{u}}_j)$ has rank M and σ_θ is the observation noise error which is Gaussian. Then, the posterior distribution in the analysis step taking into account the observations (169) is in the form of (168)

$$p^a(\mathbf{u}) = p^a(\overline{\mathbf{u}}, \mathbf{u}') = \sum_{j=1}^{Q} p_j^a \delta(\overline{\mathbf{u}} - \overline{\mathbf{u}}_j) \mathcal{N}(\mathbf{u}'(\overline{\mathbf{u}}_j)^a, \mathbf{R}'(\overline{\mathbf{u}}_j)^a). \tag{170}$$

The new mixture weights are

$$p_j^a = \frac{p_j^f I_j}{\sum_{k=1}^{Q} p_k^f I_k}, \tag{171}$$

where $I_j = \int p(\mathbf{v}|\overline{\mathbf{u}}_j, \mathbf{u}')p(\mathbf{u}'|\overline{\mathbf{u}}_j)d\mathbf{u}'$ and for each particle $\overline{\mathbf{u}}_j$, the posterior mean and variance of $\mathbf{u}'$, $\mathbf{u}'(\overline{\mathbf{u}_j})^a$ and $\mathbf{R}'(\overline{\mathbf{u}_j})^a$, respectively, are

$$\mathbf{u}'(\overline{\mathbf{u}}_j)^a = \mathbf{u}'^f + \mathbf{K}'(\mathbf{v} - \overline{\mathbf{G}}\overline{\mathbf{u}}_j^f - \mathbf{G}'(\overline{\mathbf{u}}_j^f)\mathbf{u}'),$$

$$\mathbf{R}'(\overline{\mathbf{u}}_j)^a = (\mathbf{I} - \mathbf{K}'\mathbf{G}'(\overline{\mathbf{u}}_j^f))\mathbf{R}'(\overline{\mathbf{u}})^f, \tag{172}$$

where the Kalman gain matrix $\mathbf{K}'$ is given by

$$\mathbf{K}' = \mathbf{R}'^f \mathbf{G}'(\overline{\mathbf{u}}_j^f)^T (\mathbf{G}'(\overline{\mathbf{u}}_j^f)\mathbf{R}'^f \mathbf{G}'(\overline{\mathbf{u}}_j^f)^T + \mathbf{r}_\theta)^{-1}. \tag{173}$$

See the supplementary material of [80] for more details.

6.3.3. Clustered Particle Filters and Mutiscale Data Assimilation

Clustered particle filters (CPFs) [40,235] are a new class of particle filters, introduced for high-dimensional dynamical systems such as geophysical systems. The clustered particle filters use relatively few particles compared with the standard particle filter and capture the non-Gaussian features of the true signal, which are typical in complex nonlinear systems. The method is also robust for the difficult regime of high-quality sparse and infrequent observations and does not show any filter divergence in our tests. In the clustered particle filter, coarse-grained localization is implemented through the clustering of state variables and particles are adjusted to stabilize the filter.

One of the key features of the CPF is particle adjustment. Particle adjustment updates the prior particles closer to the observation instead of reweighing the particles when the prior is too far from the observation likelihood. When observations are sparse, unobserved adjacent state

variables must have the same particle weights with the observed variable as they are updated using cross-correlations. For this purpose, the CPF partitions the state variables into nonoverlapping clusters $\{C_l, l = 1, 2, \ldots, N_{obs}\}$, where each cluster boundary is chosen as the midpoint of two adjacent observations, which is easily applicable to irregularly spaced observations. This yields N_{obs} clusters corresponding to N_{obs} observationlocations. Instead of using different weights for each state variable in the localized particle filter, the CPF uses scalar particle weights $\{\omega_{l,k}\}$ for the state variables in the same cluster C_l. For the substate vector $\mathbf{x}_{C_l} = \{x_i | x_i \in C_l\}$ corresponding to cluster C_l, the CPF considers the marginalized PDF,

$$p(\mathbf{x}_{C_l}) = \sum_k^K \omega_{l,k} \delta(\mathbf{x}_{C_l} - \mathbf{x}_{C_l,k})$$

and each observation y_j updates only the marginalized PDF of the corresponding cluster that implements coarse-grained localization. Thus, the assimilation of the full state vector is decomposed into N_{obs} independent assimilation problems for each cluster of a dimension smaller than the full state dimension N_{state}. Note that, in contrast to the localization using a smoothly varying correlation function with a localization radius parameter, the CPF has no adjustable parameter to tune localization. See [235] for more details.

The CPF can also be applied for multiscale particle filtering [40,235]. As a test example, it is shown below the skill of the multiscale cluster particle filter for the wave turbulence model introduced by Majda, McLaughlin and Tabak (MMT) [236,237] as a computationally tractable model of waveturbulence. The model is described by the following one dimensional partial differential equation for a complex scalar ψ:

$$i\partial_t \psi = |\partial_x|^{1/2} \psi - |\psi|^2 \psi + iF + iD\psi. \tag{174}$$

in a periodic domain of length L with large-scale forcing set to $F = 0.0163 \sin(4\pi x/L)$ and dissipation D for both the large and small scales. It has several features of wave turbulence that make it a difficult test problem for data assimilation. The model has a shallow energy spectrum proportional to $k^{-5/6}$ for wavenumber k and an inverse cascade of energy from small to large scales. It also has non-Gaussian extreme event statistics caused by intermittent instability and breaking of solitons. Because the unresolved small scales carry more than two-thirds of the total variance, it is a difficult filtering problem to estimate resolved large scales using mixed observations of the large- and small-scale components.

Here, we compare the filtering results of the ensemble-based multiscale data assimilation method [75] and the multiscale CPF for the MMT model. As the forecast model for both filtering methods, we use the stochastic superparameterization multiscale method [73] as discussed in the previous subsection. The forecast model uses only 128 grid points, whereas the full resolution uses 8192 grid points, which yields about 250 times cheaper computational savings. Because the forecast model has a low computational cost compared with the full-resolution model, the forecast model has significant model errors. Observations of the full-scale variables are available at uniformly distributed 64 grid points (which are extremely sparse compared with the full-resolution 8192 grid points) with an observation error variance corresponding to 3% of the total climatological variance at every time interval of 0.25. The ensemble-based method uses the tuned parameters in [75] (i.e., a short localization radius 1 and 2% covariance inflation). For the hard threshold version CPF, the particle adjustment is triggered if either real or imaginary parts are not in the convex hull of the corresponding predicted observations as we observe both parts of the true signal. Both particle and ensemble-based methods use 129 samples.

The time series of the large-scale estimation RMSE of the ensemble-based filter and clustered multiscale particle filter are shown in Figure 37. The dashed and dotted line is the effective observation error 0.34, which is defined as $\sqrt{\text{observation error variance} + \text{small-scale variance}}$ by treating the small-scale contribution as an additional error (i.e., a representation error). The dashed

line is the climatological error 0.20, which is the standard deviation of the large-scale variables. The ensemble-based method has RMSE smaller than the effective observation error but larger than the climatological error. The CPF, however, shows skillful filter performance with RMSE smaller than the climatological error. The forecast PDFs and forecast error PDFs that show the prediction skill of the method are shown in Figure 37. The CPF has a better forecast PDF fit to the true signal and a narrower peak in the forecast error PDF than the ensemble-based method.
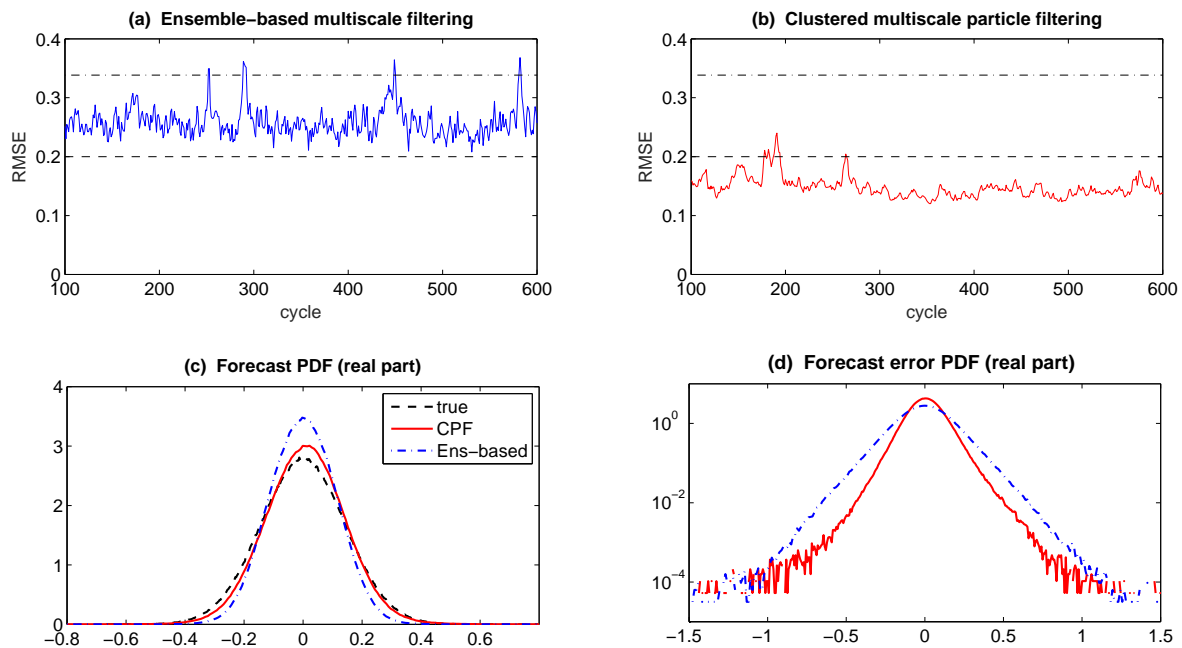

Figure 37. (**a**,**b**) time series of the large-scale estimation RMSE of the ensemble-based multiscale data assimilation method [75] and the CPF with 64 observations for the MMT model. The dashed line is the climatological error 0.20. The dashed and dotted line is the effective observation error 0.34; (**c**,**d**) large-scale forecast PDF and forecast error PDF using 64 observations.

6.3.4. Blended Particle Methods with Adaptive Subspaces for Filtering Turbulent Dynamical Systems

In the multi-scale data assimilation algorithms discussed above based on superparameterization, the subspace of particles defined by $\bar{\mathbf{u}}$ is fixed. An attractive idea is to change the subspace with particles adaptively in time to capture the non-Gaussian features as they change in time. Very accurate filtering algorithms based on these ideas for multi-scale filtering utilizing this adaptive strategy have been developed [80,238]. Nonlinear statistical forecast models such as the modified quasilinear Gaussian [17,239] are implemented in the adaptive algorithm. In particular, the paper [238] also contains many detailed numerical experiments and interesting counterexamples to more naive strategies for multi-scale data assimilation.

6.3.5. Extremely Efficient Multi-Scale Filtering Algorithms: SPEKF and Dynamic Stochastic Superresolution (DSS)

The SPEKF models as discussed in Section 4.5.1 are a class of nonlinear filters which are exact statistical equations for the mean and covariance for nonlinear forecast models that learn hidden parameters "on the fly" from the observed data. The parameters represent adaptive additive and multiplicative bias corrections from model error. They explicitly make judicious model error and utilize conditional Gaussian structure as developed in Section 4.5.1 above. The book [20] contains many examples and successful applications of this method.

Dynamical Stochastic Superresolution (DSS) uses the same idea but in addition exploits the aliased information in the observations to super-resolve a multi-scale turbulent signal [103,240].

Nontrivial applications of DSS including recovering geophysical turbulence from surface satellite observations [240] and filtering "black swans" and dispersive wave turbulence [103] with severe judicious model errors. An interesting mathematical problem is to understand the reasons for the skill of these radical methods. Recent progress in conceptual understanding of these methods for the two dimensional Navier–Stokes equations can be found in [104].

7. Conclusions

Multiscale nonlinear dynamical systems are ubiquitous in different areas, including geoscience, engineering, neural science and material science. In this article, a conditional Gaussian framework is developed and applied to the prediction, the state estimation and the uncertainty quantification of multiscale nonlinear stochastic systems. Despite the conditional Gaussianity, such systems are nevertheless highly nonlinear and are able to capture the non-Gaussian features of nature. The special structure of the system allows closed analytical formulae for solving the conditional statistics and is thus computationally efficient.

In Section 2, an overview of data, model and data-driven modeling framework is presented. Data and models are combined with each other to improve the understanding of nature and promote the filtering and prediction skill. However, solving the high-dimensional complex multiscale nonlinear dynamical systems in a direct way is computationally unaffordable and sometimes even not necessary. Therefore, cheap and effective approaches are required to efficiently solve the systems. Hybrid strategies can be used to greatly reduce the computational cost while they are able to preserve the key feature of the complex systems. In the hybrid strategies, particle methods are combined with analytically solvable conditional Gaussian statistics to deal with highly non-Gaussian characteristics in a relatively low dimensional subspace and the conditional Gaussian features in the remaining subspace, respectively. This indicates the importance of a systematic study of the conditional Gaussian system. Section 3 summarizes the general mathematical structure of nonlinear conditional Gaussian systems, the physics-constrained nonlinear stochastic models and the application of the MTV strategy to the conditional Gaussian systems. To show the wide application of the conditional Gaussian framework, a rich gallery of examples of conditional Gaussian systems are illustrated in Section 4, which includes data-driven physics-constrained nonlinear stochastic models, stochastically coupled reaction–diffusion models in neuroscience and ecology, large-scale dynamical models in turbulence, fluids and geophysical flow, and other models for filtering and predicting complex multiscale turbulent dynamical systems. Section 5 involves the effective statistically accurate algorithms that beat the curse of dimension for Fokker–Planck equation for conditional Gaussian systems. A hybrid strategy is developed where a conditional Gaussian mixture in a high-dimensional subspace via an extremely efficient parametric method is combined with a judicious non-parametric Gaussian kernel density estimation in the remaining low-dimensional subspace. For even larger dimensional systems, a judicious block decomposition and statistical symmetry are further applied that facilitate an extremely efficient parallel computation and a significant reduction of sample numbers. These algorithms are applied to the statistical prediction of a stochastically coupled FHN model with 1500 dimensions and an inhomogeneous two-layer Lorenz 96 model with 240 dimensions. Significant prediction skill shows the advantages of these algorithms in terms of both accuracy and efficiency. In Section 6, the conditional Gaussian framework is applied to develop extremely cheap multiscale data assimilation schemes, such as the stochastic superparameterization, which use particle filters to capture the non-Gaussian statistics on the large-scale part whose dimension is small whereas the statistics of the small-scale part are conditional Gaussian given the large-scale part. Other topics of the conditional Gaussian systems studied here include designing new parameter estimation schemes and understanding model errors using information theory.

The conditional Gaussian framework can also be used to study many other important topics. For example, the closed analytic formulae in the conditional statistics provide an efficient way to understand the causality between different processes in light of the information theory. The representation error is

another important issue that requires a comprehensive study and the conditional Gaussian framework have great potentials to provide both theoretic and applied insights. In addition, model selection, model reduction and more studies on the parameter estimation are all important future works within the conditional Gaussian framework.

References

1. Majda, A.J. *Introduction to Turbulent Dynamical Systems in Complex Systems*; Springer: New York, NY, USA, 2016.
2. Majda, A.; Wang, X. *Nonlinear Dynamics and Statistical Theories for Basic Geophysical Flows*; Cambridge University Press: Cambridge, UK, 2006.
3. Strogatz, S.H. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*; CRC Press: Boca Raton, FL, USA, 2018.
4. Baleanu, D.; Machado, J.A.T.; Luo, A.C. *Fractional Dynamics and Control*; Springer: New York, NY, USA, 2011.
5. Deisboeck, T.; Kresh, J.Y. *Complex Systems Science in Biomedicine*; Springer: Boston, MA, USA, 2007.
6. Stelling, J.; Kremling, A.; Ginkel, M.; Bettenbrock, K.; Gilles, E. *Foundations of Systems Biology*; MIT Press: Cambridge, MA, USA, 2001.
7. Sheard, S.A.; Mostashari, A. Principles of complex systems for systems engineering. *Syst. Eng.* **2009**, *12*, 295–311. [CrossRef]
8. Wilcox, D.C. Multiscale model for turbulent flows. *AIAA J.* **1988**, *26*, 1311–1320. [CrossRef]
9. Mohamad, M.A.; Sapsis, T.P. Probabilistic description of extreme events in intermittently unstable dynamical systems excited by correlated stochastic processes. *SIAM/ASA J. Uncertain. Quantif.* **2015**, *3*, 709–736. [CrossRef]
10. Sornette, D. Probability distributions in complex systems. In *Encyclopedia of Complexity and Systems Science*; Springer: New York, NY, USA, 2009; pp. 7009–7024.
11. Majda, A. *Introduction to PDEs and Waves for the Atmosphere and Ocean*; American Mathematical Society: Providence, RI, USA, 2003; Volume 9.
12. Wiggins, S. *Introduction to Applied Nonlinear Dynamical Systems and Chaos*; Springer: New York, NY, USA, 2003; Volume 2.
13. Vespignani, A. Predicting the behavior of techno-social systems. *Science* **2009**, *325*, 425–428. [CrossRef] [PubMed]
14. Latif, M.; Anderson, D.; Barnett, T.; Cane, M.; Kleeman, R.; Leetmaa, A.; O'Brien, J.; Rosati, A.; Schneider, E. A review of the predictability and prediction of ENSO. *J. Geophys. Res. Oceans* **1998**, *103*, 14375–14393. [CrossRef]
15. Kalnay, E. *Atmospheric Modeling, Data Assimilation and Predictability*; Cambridge University Press: Cambridge, UK, 2003.
16. Majda, A.J.; Branicki, M. Lessons in uncertainty quantification for turbulent dynamical systems. *Discret. Contin. Dyn. Syst. A* **2012**, *32*, 3133–3221.
17. Sapsis, T.P.; Majda, A.J. Statistically accurate low-order models for uncertainty quantification in turbulent dynamical systems. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 13705–13710. [CrossRef] [PubMed]
18. Mignolet, M.P.; Soize, C. Stochastic reduced order models for uncertain geometrically nonlinear dynamical systems. *Comput. Methods Appl. Mech. Eng.* **2008**, *197*, 3951–3963. [CrossRef]
19. Lahoz, W.; Khattatov, B.; Ménard, R. Data assimilation and information. In *Data Assimilation*; Springer: New York, NY, USA, 2010; pp. 3–12.
20. Majda, A.J.; Harlim, J. *Filtering Complex Turbulent Systems*; Cambridge University Press: Cambridge, UK, 2012.
21. Evensen, G. *Data Assimilation: The Ensemble Kalman Filter*; Springer: New York, NY, USA, 2009.

22. Law, K.; Stuart, A.; Zygalakis, K. *Data Assimilation: A Mathematical Introduction*; Springer: New York, NY, USA, 2015; Volume 62.

23. Palmer, T.N. A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. *Q. J. R. Meteorol. Soc.* **2001**, *127*, 279–304. [CrossRef]

24. Orrell, D.; Smith, L.; Barkmeijer, J.; Palmer, T. Model error in weather forecasting. *Nonlinear Process. Geophys.* **2001**, *8*, 357–371. [CrossRef]

25. Hu, X.M.; Zhang, F.; Nielsen-Gammon, J.W. Ensemble-based simultaneous state and parameter estimation for treatment of mesoscale model error: A real-data study. *Geophys. Res. Lett.* **2010**, *37*, L08802. [CrossRef]

26. Benner, P.; Gugercin, S.; Willcox, K. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM Rev.* **2015**, *57*, 483–531. [CrossRef]

27. Majda, A.J. Challenges in climate science and contemporary applied mathematics. *Commun. Pure Appl. Math.* **2012**, *65*, 920–948. [CrossRef]

28. Chen, N.; Majda, A.J. Filtering nonlinear turbulent dynamical systems through conditional Gaussian statistics. *Mon. Weather Rev.* **2016**, *144*, 4885–4917. [CrossRef]

29. Olbers, D. A gallery of simple models from climate physics. In *Stochastic Climate Models*; Springer: Berlin, Germany, 2001; pp. 3–63.

30. Liptser, R.S.; Shiryaev, A.N. Statistics ofRandom Processes II: Applications. In *Applied Mathematics*; Springer: Berlin/Heidelberg, Germany, 2001; Volume 6.

31. Majda, A.J.; Harlim, J. Physics constrained nonlinear regression models for time series. *Nonlinearity* **2012**, *26*, 201. [CrossRef]

32. Harlim, J.; Mahdi, A.; Majda, A.J. An ensemble Kalman filter for statistical estimation of physics constrained nonlinear regression models. *J. Comput. Phys.* **2014**, *257*, 782–812. [CrossRef]

33. Majda, A.J.; Yuan, Y. Fundamental limitations of ad hoc linear and quadratic multi-level regression models for physical systems. *Discret. Contin. Dyn. Syst. B* **2012**, *17*, 1333–1363. [CrossRef]

34. Lorenz, E.N. Deterministic nonperiodic flow. *J. Atmos. Sci.* **1963**, *20*, 130–141. [CrossRef]

35. Lorenz, E.N. Formulation of a low-order model of a moist general circulation. *J. Atmos. Sci.* **1984**, *41*, 1933–1945. [CrossRef]

36. Chen, N.; Majda, A.J. Beating the curse of dimension with accurate statistics for the Fokker–Planck equation in complex turbulent systems. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 12864–12869. [CrossRef] [PubMed]

37. Majda, A.J.; Lee, Y. Conceptual dynamical models for turbulence. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 6548–6553. [CrossRef] [PubMed]

38. Chen, N.; Majda, A.J. Efficient statistically accurate algorithms for the Fokker–Planck equation in large dimensions. *J. Comput. Phys.* **2018**, *354*, 242–268. [CrossRef]

39. Ferrari, R.; Cessi, P. Seasonal synchronization in a chaotic ocean–atmosphere model. *J. Clim.* **2003**, *16*, 875–881. [CrossRef]

40. Lee, Y.; Majda, A. Multiscale data assimilation and prediction using clustered particle filters. *J. Comput. Phys.* **2017**, in press.

41. Chen, N.; Majda, A.J.; Giannakis, D. Predicting the cloud patterns of the Madden–Julian Oscillation through a low-order nonlinear stochastic model. *Geophys. Res. Lett.* **2014**, *41*, 5612–5619. [CrossRef]

42. Chen, N.; Majda, A.J. Predicting the real-time multivariate Madden–Julian oscillation index through a low-order nonlinear stochastic model. *Mon. Weather Rev.* **2015**, *143*, 2148–2169. [CrossRef]

43. Chen, N.; Majda, A.J. Predicting the Cloud Patterns for the Boreal Summer Intraseasonal Oscillation through a Low-Order Stochastic Model. *Math. Clim. Weather Forecast.* **2015**, *1*, 1–20. [CrossRef]

44. Chen, N.; Majda, A.J.; Sabeerali, C.; Ajayamohan, R. Predicting Monsoon Intraseasonal Precipitation using a Low-Order Nonlinear Stochastic Model. *J. Clim.* **2018**, *25*, 4403–4427. [CrossRef]

45. Majda, A.J.; Timofeyev, I.; Eijnden, E.V. Models for stochastic climate prediction. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 14687–14691. [CrossRef] [PubMed]

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

47. Majda, A.; Timofeyev, I.; Vanden-Eijnden, E. A priori tests of a stochastic mode reduction strategy. *Phys. D Nonlinear Phenom.* **2002**, *170*, 206–252. [CrossRef]

48. Majda, A.J.; Timofeyev, I.; Vanden-Eijnden, E. Systematic strategies for stochastic mode reduction in climate. *J. Atmos. Sci.* **2003**, *60*, 1705–1722. [CrossRef]

49. Majda, A.J.; Franzke, C.; Khouider, B. An applied mathematics perspective on stochastic modelling for climate. *Philos. Trans. R. Soc. Lond. A* **2008**, *366*, 2427–2453. [CrossRef] [PubMed]

50. Majda, A.J.; Franzke, C.; Crommelin, D. Normal forms for reduced stochastic climate models. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 3649–3653. [CrossRef] [PubMed]

51. Thual, S.; Majda, A.J.; Stechmann, S.N. A stochastic skeleton model for the MJO. *J. Atmos. Sci.* **2014**, *71*, 697–715. [CrossRef]

52. Chen, N.; Majda, A.J. Simple stochastic dynamical models capturing the statistical diversity of El Niño Southern Oscillation. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 201620766. [CrossRef] [PubMed]

53. Castaing, B.; Gunaratne, G.; Heslot, F.; Kadanoff, L.; Libchaber, A.; Thomae, S.; Wu, X.Z.; Zaleski, S.; Zanetti, G. Scaling of hard thermal turbulence in Rayleigh-Bénard convection. *J. Fluid Mech.* **1989**, *204*, 1–30. [CrossRef]

54. Wang, X. Infinite Prandtl number limit of Rayleigh-Bénard convection. *Commun. Pure Appl. Math.* **2004**, *57*, 1265–1282. [CrossRef]

55. Majda, A.J.; Grote, M.J. Model dynamics and vertical collapse in decaying strongly stratified flows. *Phys. Fluids* **1997**, *9*, 2932–2940. [CrossRef]

56. Kelliher, J.P.; Temam, R.; Wang, X. Boundary layer associated with the Darcy–Brinkman–Boussinesq model for convection in porous media. *Phys. D Nonlinear Phenom.* **2011**, *240*, 619–628. [CrossRef]

57. Lindner, B.; García-Ojalvo, J.; Neiman, A.; Schimansky-Geier, L. Effects of noise in excitable systems. *Phys. Rep.* **2004**, *392*, 321–424. [CrossRef]

58. Medvinsky, A.B.; Petrovskii, S.V.; Tikhonova, I.A.; Malchow, H.; Li, B.L. Spatiotemporal complexity of plankton and fish dynamics. *SIAM Rev.* **2002**, *44*, 311–370. [CrossRef]

59. Shulgin, B.; Stone, L.; Agur, Z. Pulse vaccination strategy in the SIR epidemic model. *Bull. Math. Biol.* **1998**, *60*, 1123–1148. [CrossRef]

60. Ferreira, S., Jr.; Martins, M.; Vilela, M. Reaction-diffusion model for the growth of avascular tumor. *Phys. Rev. E* **2002**, *65*, 021907. [CrossRef] [PubMed]

61. Chen, N.; Majda, A.J.; Tong, X.T. Information barriers for noisy Lagrangian tracers in filtering random incompressible flows. *Nonlinearity* **2014**, *27*, 2133. [CrossRef]

62. Chen, N.; Majda, A.J.; Tong, X.T. Noisy Lagrangian tracers for filtering random rotating compressible flows. *J. Nonlinear Sci.* **2015**, *25*, 451–488. [CrossRef]

63. Chen, N.; Majda, A.J. Model error in filtering random compressible flows utilizing noisy Lagrangian tracers. *Mon. Weather Rev.* **2016**, *144*, 4037–4061. [CrossRef]

64. Gershgorin, B.; Harlim, J.; Majda, A.J. Improving filtering and prediction of spatially extended turbulent systems with model errors through stochastic parameter estimation. *J. Comput. Phys.* **2010**, *229*, 32–57. [CrossRef]

65. Gershgorin, B.; Harlim, J.; Majda, A.J. Test models for improving filtering with model errors through stochastic parameter estimation. *J. Comput. Phys.* **2010**, *229*, 1–31. [CrossRef]

66. Gardiner, C.W. *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*; Springer Series in Synergetics; Springer: Berlin/Heidelberg, Germany, 2004; Volume 13.

67. Risken, H. *The Fokker–Planck Equation—Methods of Solution and Applications*; Springer Series in Synergetics; Springer: Berlin/Heidelberg, Germany, 1989; Volume 18, p. 301.

68. Pichler, L.; Masud, A.; Bergman, L.A. Numerical solution of the Fokker–Planck equation by finite difference and finite element methods-a comparative study. In *Computational Methods in Stochastic Dynamics*; Springer: Dordrecht, The Netherlands, 2013; pp. 69–85.

69. Robert, C.P. *Monte Carlo Methods*; Wiley Online Library: Hoboken, NJ, USA, 2004.

70. Majda, A.J.; Grote, M.J. Mathematical test models for superparametrization in anisotropic turbulence. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 5470–5474. [CrossRef] [PubMed]

71. Grooms, I.; Majda, A.J. Stochastic superparameterization in a one-dimensional model for wave turbulence. *Commun. Math. Sci* **2014**, *12*, 509–525. [CrossRef]

72. Grooms, I.; Majda, A.J. Stochastic superparameterization in quasigeostrophic turbulence. *J. Comput. Phys.* **2014**, *271*, 78–98. [CrossRef]

73. Majda, A.J.; Grooms, I. New perspectives on superparameterization for geophysical turbulence. *J. Comput. Phys.* **2014**, *271*, 60–77. [CrossRef]

74. Grooms, I.; Lee, Y.; Majda, A.J. Numerical schemes for stochastic backscatter in the inverse cascade of quasigeostrophic turbulence. *Multiscale Model. Simul.* **2015**, *13*, 1001–1021. [CrossRef]

75. Grooms, I.; Lee, Y.; Majda, A.J. Ensemble Kalman filters for dynamical systems with unresolved turbulence. *J. Comput. Phys.* **2014**, *273*, 435–452. [CrossRef]

76. Grooms, I.; Lee, Y.; Majda, A.J. Ensemble filtering and low-resolution model error: Covariance inflation, stochastic parameterization, and model numerics. *Mon. Weather Rev.* **2015**, *143*, 3912–3924. [CrossRef]

77. Grooms, I.; Majda, A.J. Efficient stochastic superparameterization for geophysical turbulence. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 4464–4469. [CrossRef] [PubMed]

78. Grooms, I.; Majda, A.J.; Smith, K.S. Stochastic superparameterization in a quasigeostrophic model of the Antarctic Circumpolar Current. *Ocean Model.* **2015**, *85*, 1–15. [CrossRef]

79. Bain, A.; Crisan, D. *Fundamentals of stochastic filtering*; Springer: New York, NY, USA, 2009; Volume 3.

80. Majda, A.J.; Qi, D.; Sapsis, T.P. Blended particle filters for large-dimensional chaotic dynamical systems. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 7511–7516. [CrossRef] [PubMed]

81. Janjić, T.; Cohn, S.E. Treatment of observation error due to unresolved scales in atmospheric data assimilation. *Mon. Weather Rev.* **2006**, *134*, 2900–2915. [CrossRef]

82. Daley, R. Estimating observation error statistics for atmospheric data assimilation. *Ann. Geophys.* **1993**, *11*, 634–647.

83. Qian, C.; Zhou, W.; Fong, S.K.; Leong, K.C. Two approaches for statistical prediction of non-Gaussian climate extremes: A case study of Macao hot extremes during 1912–2012. *J. Clim.* **2015**, *28*, 623–636. [CrossRef]

84. Hjorth, J.U. *Computer Intensive Statistical Methods: Validation, Model Selection, and Bootstrap*; Routledge: London, UK, 2017.

85. Alessandrini, S.; Delle Monache, L.; Sperati, S.; Cervone, G. An analog ensemble for short-term probabilistic solar power forecast. *Appl. Energy* **2015**, *157*, 95–110. [CrossRef]

86. Alessandrini, S.; Delle Monache, L.; Sperati, S.; Nissen, J. A novel application of an analog ensemble for short-term wind power forecasting. *Renew. Energy* **2015**, *76*, 768–781. [CrossRef]

87. Nair, P.; Chakraborty, A.; Varikoden, H.; Francis, P.; Kuttippurath, J. The local and global climate forcings induced inhomogeneity of Indian rainfall. *Sci. Rep.* **2018**, *8*, 6026. [CrossRef] [PubMed]

88. Hodges, K.; Chappell, D.; Robinson, G.; Yang, G. An improved algorithm for generating global window brightness temperatures from multiple satellite infrared imagery. *J. Atmos. Ocean. Technol.* **2000**, *17*, 1296–1312. [CrossRef]

89. Winker, D.M.; Vaughan, M.A.; Omar, A.; Hu, Y.; Powell, K.A.; Liu, Z.; Hunt, W.H.; Young, S.A. Overview of the CALIPSO mission and CALIOP data processing algorithms. *J. Atmos. Ocean. Technol.* **2009**, *26*, 2310–2323. [CrossRef]

90. Kravtsov, S.; Kondrashov, D.; Ghil, M. Multilevel regression modeling of nonlinear processes: Derivation and applications to climatic variability. *J. Clim.* **2005**, *18*, 4404–4424. [CrossRef]

91. Wikle, C.K.; Hooten, M.B. A general science-based framework for dynamical spatio-temporal models. *Test* **2010**, *19*, 417–451. [CrossRef]

92. Alexander, R.; Zhao, Z.; Székely, E.; Giannakis, D. Kernel analog forecasting of tropical intraseasonal oscillations. *J. Atmos. Sci.* **2017**, *74*, 1321–1342. [CrossRef]

93. Harlim, J.; Yang, H. Diffusion Forecasting Model with Basis Functions from QR-Decomposition. *J. Nonlinear Sci.* **2017**, 1–26. [CrossRef]

94. Kalman, R.E. A new approach to linear filtering and prediction problems. *J. Basic Eng.* **1960**, *82*, 35–45. [CrossRef]

95. Arulampalam, M.S.; Maskell, S.; Gordon, N.; Clapp, T. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans. Signal Process.* **2002**, *50*, 174–188. [CrossRef]

96. Ristic, B.; Arulampalam, S.; Gordon, N. *Beyond the Kalman Filter: Particle Filters for Tracking Applications*; Artech House: Norwood, MA, USA, 2003.

97. Tippett, M.K.; Anderson, J.L.; Bishop, C.H.; Hamill, T.M.; Whitaker, J.S. Ensemble square root filters. *Mon. Weather Rev.* **2003**, *131*, 1485–1490. [CrossRef]

98. Anderson, J.L. An ensemble adjustment Kalman filter for data assimilation. *Mon. Weather Rev.* **2001**, *129*, 2884–2903. [CrossRef]

99. Harlim, J.; Majda, A.J. Catastrophic filter divergence in filtering nonlinear dissipative systems. *Commun. Math. Sci.* **2010**, *8*, 27–43. [CrossRef]

100. Tong, X.T.; Majda, A.J.; Kelly, D. Nonlinear stability and ergodicity of ensemble based Kalman filters. *Nonlinearity* **2016**, *29*, 657. [CrossRef]

101. Majda, A.J.; Harlim, J.; Gershgorin, B. Mathematical strategies for filtering turbulent dynamical systems. *Discret. Contin. Dyn. Syst.* **2010**, *27*, 441–486. [CrossRef]

102. Frenkel, Y.; Majda, A.J.; Khouider, B. Using the stochastic multicloud model to improve tropical convective parameterization: A paradigm example. *J. Atmos. Sci.* **2012**, *69*, 1080–1105. [CrossRef]

103. Branicki, M.; Majda, A.J. Dynamic stochastic superresolution of sparsely observed turbulent systems. *J. Comput. Phys.* **2013**, *241*, 333–363. [CrossRef]

104. Branicki, M.; Majda, A.J.; Law, K.J.H. Accuracy of some approximate gaussian filters for the Navier–Stokes equation in the presence of model error. *Multiscale Model. Simul.* **2018**, submitted.

105. Kalman, R.E.; Bucy, R.S. New results in linear filtering and prediction theory. *J. Basic Eng.* **1961**, *83*, 95–108. [CrossRef]

106. Brammer, K.; Siffling, G. *Kalman–Bucy Filters*; Artech House: Norwood, MA, USA, 1989.

107. Bucy, R.S.; Joseph, P.D. *Filtering for Stochastic Processes with Applications to Guidance*; American Mathematical Society: Providence, RI, USA, 1987; Volume 326.

108. Jazwinski, A.H. *Stochastic Processes and Filtering Theory*; Courier Corporation: Washington, DC, USA, 2007.

109. Bensoussan, A. *Stochastic Control of Partially Observable Systems*; Cambridge University Press: Cambridge, UK, 2004.

110. Liptser, R.S.; Shiryaev, A. *Statistics of Random Processes*; Springer: New York, NY, USA, 1977.

111. Sparrow, C. *The Lorenz Equations: Bifurcations, Chaos, and Strange Attractors*; Springer: New York, NY, USA, 2012; Volume 41.

112. Haken, H. Analogy between higher instabilities in fluids and lasers. *Phys. Lett. A* **1975**, *53*, 77–78. [CrossRef]

113. Knobloch, E. Chaos in the segmented disc dynamo. *Phys. Lett. A* **1981**, *82*, 439–440. [CrossRef]

114. Gorman, M.; Widmann, P.; Robbins, K. Nonlinear dynamics of a convection loop: A quantitative comparison of experiment with theory. *Phys. D Nonlinear Phenom.* **1986**, *19*, 255–267. [CrossRef]

115. Hemati, N. Strange attractors in brushless DC motors. *IEEE Trans. Circuits Syst. I* **1994**, *41*, 40–45. [CrossRef]

116. Cuomo, K.M.; Oppenheim, A.V. Circuit implementation of synchronized chaos with applications to communications. *Phys. Rev. Lett.* **1993**, *71*, 65. [CrossRef] [PubMed]

117. Poland, D. Cooperative catalysis and chemical chaos: A chemical model for the Lorenz equations. *Phys. D Nonlinear Phenom.* **1993**, *65*, 86–99. [CrossRef]

118. Tzenov, S.I. Strange attractors characterizing the osmotic instability. *arXiv* **2014**, arXiv:1406.0979.

119. Wilks, D.S. Effects of stochastic parametrizations in the Lorenz'96 system. *Q. J. R. Meteorol. Soc.* **2005**, *131*, 389–407. [CrossRef]

120. Arnold, H.; Moroz, I.; Palmer, T. Stochastic parametrizations and model uncertainty in the Lorenz'96 system. *Phil. Trans. R. Soc. A* **2013**, *371*, 20110479. [CrossRef] [PubMed]

121. Vallis, G.K. *Atmospheric and Oceanic Fluid Dynamics*; Cambridge University Press: Cambridge, UK, 2017.

122. Salmon, R. *Lectures on Geophysical Fluid Dynamics*; Oxford University Press: Oxford, UK, 1998.

123. Lorenz, E.N. Irregularity: A fundamental property of the atmosphere. *Tellus A* **1984**, *36*, 98–110. [CrossRef]

124. Giannakis, D.; Majda, A.J. Nonlinear Laplacian spectral analysis for time series with intermittency and low-frequency variability. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 2222–2227. [CrossRef] [PubMed]

125. Majda, A.; Abramov, R.V.; Grote, M.J. *Information Theory and Stochastics for Multiscale Nonlinear Systems*; American Mathematical Society: Providence, RI, USA, 2005; Volume 25.

126. Majda, A.J.; Gershgorin, B.; Yuan, Y. Low-frequency climate response and fluctuation–dissipation theorems: Theory and practice. *J. Atmos. Sci.* **2010**, *67*, 1186–1201. [CrossRef]

127. Mantua, N.J.; Hare, S.R. The Pacific decadal oscillation. *J. Oceanogr.* **2002**, *58*, 35–44. [CrossRef]

128. Hinze, J. Turbulence: An introduction to its mechanisms and theory. *Mech. Eng.* **1959**, *1396*. [CrossRef]

129. Townsend, A.A. *The Structure of Turbulent Shear Flow*; Cambridge University Press: Cambridge, UK, 1980.

130. Frisch, U. *Turbulence: The Legacy of AN Kolmogorov*; Cambridge University Press: Cambridge, UK, 1995.

131. Cessi, P.; Primeau, F. Dissipative selection of low-frequency modes in a reduced-gravity basin. *J. Phys. Oceanogr.* **2001**, *31*, 127–137. [CrossRef]

132. LaCasce, J.; Pedlosky, J. Baroclinic Rossby waves in irregular basins. *J. Phys. Oceanogr.* **2002**, *32*, 2828–2847. [CrossRef]

133. Charney, J.G.; DeVore, J.G. Multiple flow equilibria in the atmosphere and blocking. *J. Atmos. Sci.* **1979**, *36*, 1205–1216. [CrossRef]

134. Treutlein, H.; Schulten, K. Noise Induced Limit Cycles of the Bonhoeffer-Van der Pol Model of Neural Pulses. *Ber. Bunsenges. Phys. Chem.* **1985**, *89*, 710–718. [CrossRef]

135. Lindner, B.; Schimansky-Geier, L. Coherence and stochastic resonance in a two-state system. *Phys. Rev. E* **2000**, *61*, 6103. [CrossRef]

136. Longtin, A. Stochastic resonance in neuron models. *J. Stat. Phys.* **1993**, *70*, 309–327. [CrossRef]

137. Wiesenfeld, K.; Pierson, D.; Pantazelou, E.; Dames, C.; Moss, F. Stochastic resonance on a circle. *Phys. Rev. Lett.* **1994**, *72*, 2125. [CrossRef] [PubMed]

138. Neiman, A.; Schimansky-Geier, L.; Cornell-Bell, A.; Moss, F. Noise-enhanced phase synchronization in excitable media. *Phys. Rev. Lett.* **1999**, *83*, 4896. [CrossRef]

139. Hempel, H.; Schimansky-Geier, L.; Garcia-Ojalvo, J. Noise-sustained pulsating patterns and global oscillations in subexcitable media. *Phys. Rev. Lett.* **1999**, *82*, 3713. [CrossRef]

140. Hu, B.; Zhou, C. Phase synchronization in coupled nonidentical excitable systems and array-enhanced coherence resonance. *Phys. Rev. E* **2000**, *61*, R1001. [CrossRef]

141. Jung, P.; Cornell-Bell, A.; Madden, K.S.; Moss, F. Noise-induced spiral waves in astrocyte syncytia show evidence of self-organized criticality. *J. Neurophysiol.* **1998**, *79*, 1098–1101. [CrossRef] [PubMed]

142. DeVille, R.L.; Vanden-Eijnden, E.; Muratov, C.B. Two distinct mechanisms of coherence in randomly perturbed dynamical systems. *Phys. Rev. E* **2005**, *72*, 031105. [CrossRef] [PubMed]

143. Casado, J. Noise-induced coherence in an excitable system. *Phys. Lett. A* **1997**, *235*, 489–492. [CrossRef]

144. Sadhu, S.; Kuehn, C. Stochastic mixed-mode oscillations in a three-species predator–prey model. *Chaos* **2018**, *28*, 033606. [CrossRef] [PubMed]

145. Sun, G.Q.; Jin, Z.; Liu, Q.X.; Li, B.L. Rich dynamics in a predator–prey model with both noise and periodic force. *BioSystems* **2010**, *100*, 14–22. [CrossRef] [PubMed]

146. McCluskey, C.C. Complete global stability for an SIR epidemic model with delay-distributed or discrete. *Nonlinear Anal.* **2010**, *11*, 55–59. [CrossRef]

147. Kim, K.I.; Lin, Z.; Zhang, Q. An SIR epidemic model with free boundary. *Nonlinear Anal.* **2013**, *14*, 1992–2001. [CrossRef]

148. Allen, L.J.; Brauer, F.; Van den Driessche, P.; Wu, J. *Mathematical Epidemiology*; Springer: New York, NY, USA, 2008; Volume 1945.

149. Gray, A.; Greenhalgh, D.; Hu, L.; Mao, X.; Pan, J. A stochastic differential equation SIS epidemic model. *SIAM J. Appl. Math.* **2011**, *71*, 876–902. [CrossRef]

150. Zhou, Y.; Liu, H. Stability of periodic solutions for an SIS model with pulse vaccination. *Math. Comput. Model.* **2003**, *38*, 299–308. [CrossRef]

151. Madden, R.A.; Julian, P.R. Observations of the 40–50-day tropical oscillation—A review. *Mon. Weather Rev.* **1994**, *122*, 814–837. [CrossRef]

152. Zhang, C. Madden-julian oscillation. *Rev. Geophys.* **2005**, *43*, 2004RG000158. [CrossRef]

153. Lau, W.K.M.; Waliser, D.E. *Intraseasonal Variability in the Atmosphere-Ocean Climate System*; Springer: Berlin, Germany, 2011.

154. Hendon, H.H.; Wheeler, M.C.; Zhang, C. Seasonal dependence of the MJO–ENSO relationship. *J. Clim.* **2007**, *20*, 531–543. [CrossRef]

155. McPhaden, M.J. Genesis and evolution of the 1997-98 El Niño. *Science* **1999**, *283*, 950–954. [CrossRef] [PubMed]

156. Tang, Y.; Yu, B. MJO and its relationship to ENSO. *J. Geophys. Res. Atmos.* **2008**, *113*, D14106. [CrossRef]

157. Majda, A.J.; Stechmann, S.N. The skeleton of tropical intraseasonal oscillations. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 8417–8422. [CrossRef] [PubMed]

158. Majda, A.J.; Stechmann, S.N. Nonlinear dynamics and regional variations in the MJO skeleton. *J. Atmos. Sci.* **2011**, *68*, 3053–3071. [CrossRef]

159. Stechmann, S.N.; Majda, A.J. Identifying the skeleton of the Madden–Julian oscillation in observational data. *Mon. Weather Rev.* **2015**, *143*, 395–416. [CrossRef]

160. Stachnik, J.P.; Waliser, D.E.; Majda, A.J.; Stechmann, S.N.; Thual, S. Evaluating MJO event initiation and decay in the skeleton model using an RMM-like index. *J. Geophys. Res. Atmos.* **2015**, *120*, 11486–11508. [CrossRef]

161. Biello, J.A.; Majda, A.J. Modulating synoptic scale convective activity and boundary layer dissipation in the IPESD models of the Madden–Julian oscillation. *Dyn. Atmos. Oceans* **2006**, *42*, 152–215. [CrossRef]

162. Chen, N.; Majda, A.J. Filtering the stochastic skeleton model for the Madden–Julian oscillation. *Mon. Weather Rev.* **2016**, *144*, 501–527. [CrossRef]

163. Thual, S.; Majda, A.J.; Stechmann, S.N. Asymmetric intraseasonal events in the stochastic skeleton MJO model with seasonal cycle. *Clim. Dyn.* **2015**, *45*, 603–618. [CrossRef]

164. Thual, S.; Majda, A.J. A skeleton model for the MJO with refined vertical structure. *Clim. Dyn.* **2016**, *46*, 2773–2786. [CrossRef]

165. Majda, A.J.; Tong, X.T. Geometric ergodicity for piecewise contracting processes with applications for tropical stochastic lattice models. *Commun. Pure Appl. Math.* **2016**, *69*, 1110–1153. [CrossRef]

166. Clarke, A.J. *An Introduction to the Dynamics of El Niño and the Southern Oscillation*; Elsevier: Amsterdam, The Netherlands, 2008.

167. Sarachik, E.S.; Cane, M.A. *The El Nino-Southern Oscillation Phenomenon*; Cambridge University Press: Cambridge, UK, 2010.

168. Holton, J.R.; Dmowska, R. *El Niño, La Niña, and the Southern Oscillation*; Academic Press: New York, NY, USA, 1989; Volume 46.

169. Lee, T.; McPhaden, M.J. Increasing intensity of El Niño in the central-equatorial Pacific. *Geophys. Res. Lett.* **2010**, *37*, L14603. [CrossRef]

170. Kao, H.Y.; Yu, J.Y. Contrasting eastern-Pacific and central-Pacific types of ENSO. *J. Clim.* **2009**, *22*, 615–632. [CrossRef]

171. Ashok, K.; Behera, S.K.; Rao, S.A.; Weng, H.; Yamagata, T. El Niño Modoki and its possible teleconnection. *J. Geophys. Res. Oceans* **2007**, *112*, C11. [CrossRef]

172. Kug, J.S.; Jin, F.F.; An, S.I. Two types of El Niño events: Cold tongue El Niño and warm pool El Niño. *J. Clim.* **2009**, *22*, 1499–1515. [CrossRef]

173. Capotondi, A.; Wittenberg, A.T.; Newman, M.; Di Lorenzo, E.; Yu, J.Y.; Braconnot, P.; Cole, J.; Dewitte, B.; Giese, B.; Guilyardi, E.; et al. Understanding ENSO diversity. *Bull. Am. Meteorol. Soc.* **2015**, *96*, 921–938. [CrossRef]

174. Chen, D.; Lian, T.; Fu, C.; Cane, M.A.; Tang, Y.; Murtugudde, R.; Song, X.; Wu, Q.; Zhou, L. Strong influence of westerly wind bursts on El Niño diversity. *Nat. Geosci.* **2015**, *8*, 339. [CrossRef]

175. Thual, S.; Majda, A.J.; Chen, N.; Stechmann, S.N. Simple stochastic model for El Niño with westerly wind bursts. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 10245–10250. [CrossRef] [PubMed]

176. Chen, N.; Majda, A.J. Simple dynamical models capturing the key features of the Central Pacific El Niño. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 11732–11737. [CrossRef] [PubMed]

177. Chen, N.; Majda, A.J.; Thual, S. Observations and mechanisms of a simple stochastic dynamical model capturing El Niño diversity. *J. Clim.* **2018**, *31*, 449–471. [CrossRef]

178. Thual, S.; Majda, A.J.; Chen, N. Mechanisms of the 2014–2016 Delayed Super El Nino Captured by Simple Dynamical Models. *Clim. Dyn.* **2018**, in press. [CrossRef]

179. Thual, S.; Majda, A.; Chen, N. Seasonal synchronization of a simple stochastic dynamical model capturing El Niño diversity. *J. Clim.* **2017**, *30*, 10047–10066. [CrossRef]

180. Wang, X. A Note on Long-Time Behavior of Solutions to the Boussinesq System at Large Prandtl Number. *Contemp. Math.* **2005**, *371*, 315.

181. Constantin, P.; Doering, C.R. Infinite Prandtl number convection. *J. Stat. Phys.* **1999**, *94*, 159–172. [CrossRef]

182. Boffetta, G.; Davoudi, J.; Eckhardt, B.; Schumacher, J. Lagrangian tracers on a surface flow: The role of time correlations. *Phys. Rev. Lett.* **2004**, *93*, 134501. [CrossRef] [PubMed]

183. Ide, K.; Kuznetsov, L.; Jones, C.K. Lagrangian data assimilation for point vortex systems*. *J. Turbul.* **2002**, *3*. [CrossRef]

184. Apte, A.; Jones, C.K.; Stuart, A.; Voss, J. Data assimilation: Mathematical and statistical perspectives. *Int. J. Numerical Methods Fluids* **2008**, *56*, 1033–1046. [CrossRef]

185. Salman, H.; Ide, K.; Jones, C.K. Using flow geometry for drifter deployment in Lagrangian data assimilation. *Tellus A* **2008**, *60*, 321–335. [CrossRef]

186. Branicki, M.; Gershgorin, B.; Majda, A.J. Filtering skill for turbulent signals for a suite of nonlinear and linear extended Kalman filters. *J. Comput. Phys.* **2012**, *231*, 1462–1498. [CrossRef]

187. Branicki, M.; Chen, N.; Majda, A.J. Non-Gaussian test models for prediction and state estimation with model errors. *Chin. Ann. Math. Ser. B* **2013**, *34*, 29–64. [CrossRef]

188. Harlim, J.; Majda, A.J. Filtering turbulent sparsely observed geophysical flows. *Mon. Weather Rev.* **2010**, *138*, 1050–1083. [CrossRef]

189. Monahan, A.H. Temporal Filtering Enhances the Skewness of Sea Surface Winds. *J. Clim.* **2018**. [CrossRef]

190. Monahan, A.H. The temporal autocorrelation structure of sea surface winds. *J. Clim.* **2012**, *25*, 6684–6700. [CrossRef]

191. Cousins, W.; Sapsis, T.P. Reduced-order precursors of rare events in unidirectional nonlinear water waves. *J. Fluid Mech.* **2016**, *790*, 368–388. [CrossRef]

192. Botev, Z.I.; Grotowski, J.F.; Kroese, D.P. Kernel density estimation via diffusion. *Ann. Stat.* **2010**, *38*, 2916–2957. [CrossRef]

193. Chen, N.; Majda, A.J.; Tong, X.T. Rigorous analysis for efficient statistically accurate algorithms for solving Fokker–Planck equations in large dimensions. *arXiv* **2017**, arXiv:1709.05585.

194. Majda, A.J.; Gershgorin, B. Quantifying uncertainty in climate change science through empirical information theory. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 14958–14963. [CrossRef] [PubMed]

195. Majda, A.J.; Gershgorin, B. Link between statistical equilibrium fidelity and forecasting skill for complex systems with model error. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 12599–12604. [CrossRef] [PubMed]

196. Kullback, S.; Leibler, R.A. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [CrossRef]

197. Kullback, S. Letter to the editor: The Kullback-Leibler distance. *Am. Stat.* **1987**, *41*, 340–341.

198. Chen, N.; Giannakis, D.; Herbei, R.; Majda, A.J. An MCMC algorithm for parameter estimation in signals with hidden intermittent instability. *SIAM/ASA J. Uncertain. Quantif.* **2014**, *2*, 647–669. [CrossRef]

199. Andrieu, C.; De Freitas, N.; Doucet, A.; Jordan, M.I. An introduction to MCMC for machine learning. *Mach. Learn.* **2003**, *50*, 5–43. [CrossRef]

200. Beskos, A.; Roberts, G.; Stuart, A.; Voss, J. MCMC methods for diffusion bridges. *Stoch. Dyn.* **2008**, *8*, 319–350. [CrossRef]

201. Golightly, A.; Wilkinson, D.J. Bayesian inference for stochastic kinetic models using a diffusion approximation. *Biometrics* **2005**, *61*, 781–788. [CrossRef] [PubMed]

202. Johansen, S.; Juselius, K. Maximum likelihood estimation and inference on cointegration-with applications to the demand for money. *Oxf. Bull. Econ. Stat.* **1990**, *52*, 169–210. [CrossRef]

203. Bock, R.D.; Aitkin, M. Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika* **1981**, *46*, 443–459. [CrossRef]

204. Evensen, G. The ensemble Kalman filter for combined state and parameter estimation. *IEEE Control Syst.* **2009**, *29*, 83–104. [CrossRef]

205. Annan, J.; Hargreaves, J.; Edwards, N.; Marsh, R. Parameter estimation in an intermediate complexity earth system model using an ensemble Kalman filter. *Ocean Model.* **2005**, *8*, 135–154. [CrossRef]

206. Taylor, K.E. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res. Atmos.* **2001**, *106*, 7183–7192. [CrossRef]

207. Houtekamer, P.L.; Mitchell, H.L. Data assimilation using an ensemble Kalman filter technique. *Mon. Weather Rev.* **1998**, *126*, 796–811. [CrossRef]

208. Lermusiaux, P.F. Data assimilation via error subspace statistical estimation. Part II: Middle Atlantic Bight shelfbreak front simulations and ESSE validation. *Mon. Weather Rev.* **1999**, *127*, 1408–1432. [CrossRef]

209. Hendon, H.H.; Lim, E.; Wang, G.; Alves, O.; Hudson, D. Prospects for predicting two flavors of El Niño. *Geophys. Res. Lett.* **2009**, *36*, L19713. [CrossRef]

210. Kim, H.M.; Webster, P.J.; Curry, J.A. Seasonal prediction skill of ECMWF System 4 and NCEP CFSv2 retrospective forecast for the Northern Hemisphere Winter. *Clim. Dyn.* **2012**, *39*, 2957–2973. [CrossRef]

211. Majda, A.J.; Chen, N. Model Error, Information Barriers, State Estimation and Prediction in Complex Multiscale Systems. *Entropy* **2018**, submitted.

212. Majda, A.J.; Gershgorin, B. Improving model fidelity and sensitivity for complex systems through empirical information theory. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 10044–10049. [CrossRef] [PubMed]

213. Gershgorin, B.; Majda, A.J. Quantifying uncertainty for climate change and long-range forecasting scenarios with model errors—Part I: Gaussian models. *J. Clim.* **2012**, *25*, 4523–4548. [CrossRef]

214. Branicki, M.; Majda, A.J. Quantifying uncertainty for predictions with model error in non-Gaussian systems with intermittency. *Nonlinearity* **2012**, *25*, 2543. [CrossRef]

215. Branicki, M.; Majda, A. Quantifying Bayesian filter performance for turbulent dynamical systems through information theory. *Commun. Math. Sci* **2014**, *12*, 901–978. [CrossRef]

216. Kleeman, R. Information theory and dynamical system predictability. *Entropy* **2011**, *13*, 612–649. [CrossRef]

217. Kleeman, R. Measuring dynamical prediction utility using relative entropy. *J. Atmos. Sci.* **2002**, *59*, 2057–2072. [CrossRef]

218. Majda, A.; Kleeman, R.; Cai, D. A mathematical framework for quantifying predictability through relative entropy. *Methods Appl. Anal.* **2002**, *9*, 425–444.

219. Xu, Q. Measuring information content from observations for data assimilation: Relative entropy versus Shannon entropy difference. *Tellus A* **2007**, *59*, 198–209. [CrossRef]

220. Barato, A.; Seifert, U. Unifying three perspectives on information processing in stochastic thermodynamics. *Phys. Rev. Lett.* **2014**, *112*, 090601. [CrossRef] [PubMed]

221. Kawaguchi, K.; Nakayama, Y. Fluctuation theorem for hidden entropy production. *Phys. Rev. E* **2013**, *88*, 022147. [CrossRef] [PubMed]

222. Pham, D.T. Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Mon. Weather Rev.* **2001**, *129*, 1194–1207. [CrossRef]

223. Majda, A.J. Statistical energy conservation principle for inhomogeneous turbulent dynamical systems. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 8937–8941. [CrossRef] [PubMed]

224. Harlim, J.; Majda, A. Filtering nonlinear dynamical systems with linear stochastic models. *Nonlinearity* **2008**, *21*, 1281. [CrossRef]

225. Gajic, Z.; Lelic, M. *Modern Control Systems Engineering*; Prentice-Hall, Inc.: Upper Saddle River, NJ, USA, 1996.

226. Grabowski, W.W.; Smolarkiewicz, P.K. CRCP: A cloud resolving convection parameterization for modeling the tropical convecting atmosphere. *Phys. D Nonlinear Phenom.* **1999**, *133*, 171–178. [CrossRef]

227. Grabowski, W.W. An improved framework for superparameterization. *J. Atmos. Sci.* **2004**, *61*, 1940–1952. [CrossRef]

228. Doucet, A.; Godsill, S.; Andrieu, C. On sequential Monte Carlo sampling methods for Bayesian filtering. *Stat. Comput.* **2000**, *10*, 197–208. [CrossRef]

229. Harlim, J.; Majda, A.J. Test models for filtering with superparameterization. *Multiscale Model. Simul.* **2013**, *11*, 282–308. [CrossRef]

230. Lee, Y.; Majda, A.J. Multiscale methods for data assimilation in turbulent systems. *Multiscale Model. Simul.* **2015**, *13*, 691–713. [CrossRef]

231. Daley, R. *Atmospheric Data Analysis, Cambridge Atmospheric and Space Science Series*; Cambridge University Press: Cambridge, UK, 1991; Volume 6966, p. 25.

232. Sorenson, H.W.; Alspach, D.L. Recursive Bayesian estimation using Gaussian sums. *Automatica* **1971**, *7*, 465–479. [CrossRef]

233. Hoteit, I.; Luo, X.; Pham, D.T. Particle Kalman filtering: A nonlinear Bayesian framework for ensemble Kalman filters. *Mon. Weather Rev.* **2012**, *140*, 528–542. [CrossRef]

234. Alspach, D.L.; Samant, V.S.; Sorenson, H.W. *Practical Control Algorithms for Nonlinear Stochastic Systems and Investigations of Nonlinear Filters*; Technical Report for Statistics and Probability: San Diego, CA, USA, July 1980.

235. Lee, Y.; Majda, A.J. State estimation and prediction using clustered particle filters. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 14609–14614. [CrossRef] [PubMed]

236. Majda, A.; McLaughlin, D.; Tabak, E. A one-dimensional model for dispersive wave turbulence. *J. Nonlinear Sci.* **1997**, *7*, 9–44. [CrossRef]

237. Cai, D.; Majda, A.J.; McLaughlin, D.W.; Tabak, E.G. Dispersive wave turbulence in one dimension. *Phys. D Nonlinear Phenom.* **2001**, *152*, 551–572. [CrossRef]

238. Qi, D.; Majda, A.J. Blended particle methods with adaptive subspaces for filtering turbulent dynamical systems. *Phys. D Nonlinear Phenom.* **2015**, *298*, 21–41. [CrossRef]

239. Sapsis, T.P.; Majda, A.J. Blended reduced subspace algorithms for uncertainty quantification of quadratic systems with a stable mean state. *Phys. D Nonlinear Phenom.* **2013**, *258*, 61–76. [CrossRef]

240. Keating, S.R.; Majda, A.J.; Smith, K.S. New methods for estimating ocean eddy heat transport using satellite altimetry. *Mon. Weather Rev.* **2012**, *140*, 1703–1722. [CrossRef]