

A new efficient parameter estimation algorithm for high-dimensional complex nonlinear turbulent dynamical systems with partial observations

Nan Chen

Department of Mathematics, University of Wisconsin-Madison, Madison, Wisconsin, USA

Andrew J. Majda

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

Abstract

Parameter estimation for high-dimensional complex nonlinear turbulent dynamical systems with only partial observations is an important and practical issue. However, most of the existing parameter estimation algorithms are computationally expensive in the presence of a large number of state variables or parameters. In this article, a new efficient algorithm is developed for estimating parameters in high-dimensional nonlinear turbulent dynamical systems with conditional Gaussian structures. This algorithm exploits the closed analytical form of the conditional statistics to recover the unobserved trajectories in an optimal and deterministic way, which facilitates the calculation of the likelihood function and circumvents the computationally expensive data augmentation approach in sampling the unobserved trajectories as widely used in the literature. Such an efficient method of recovering the unobserved trajectories is then incorporated into a simple Markov chain Monte Carlo (MCMC) algorithm to estimate parameters in complex dynamical system using only a short period of training data.

*Corresponding author: Nan Chen

Email addresses: chennan@math.wisc.edu (Nan Chen), jonjon@cims.nyu.edu (Andrew J. Majda)

Next, two effective strategies are developed and incorporated into the algorithm that facilitates an efficient estimation of the parameters in high-dimensional systems. The first strategy involves a judicious block decomposition of the state variables such that the original problem is divided into several subproblems coupled in a specific way that allows an extremely cheap parallel computation for the parameter estimation. The second strategy exploits statistical symmetry for a further reduction of the computational cost when the system is statistically homogeneous. The new parameter estimation algorithm is applied to a two-layer Lorenz 96 model with 80 state variables and 162 parameters and the model mimics the realistic features of atmosphere wave propagations and excitable media. The efficient algorithm results in an accurate estimation of the parameters, which further allows a skillful prediction by the model with estimated parameters. Other simple nonlinear models are also used to illustrate the features of the new algorithm.

Keywords: conditional Gaussian nonlinear models, closed analytical formulae, MCMC, high-dimensional non-Gaussian systems, block decomposition

2010 MSC: 65C40, 68Q17, 85-08, 60G20

1. Introduction

Complex nonlinear turbulent dynamical systems are ubiquitous in geophysics, engineering, neuroscience and material science [1, 2, 3, 4, 5, 6, 7, 8]. Key features of these complex nonlinear systems are multiscale dynamics, high-dimensional phase space, nonlinear energy transfers, highly non-Gaussian probability density functions (PDFs), intermittent instability, random internal and external forcing as well as extreme events. The prerequisite of understanding and predicting these complex nonlinear turbulent systems is an accurate estimation of the model parameters given observations. In many practical situations, due to the lack of physical understanding and the inadequate resolution in the measurement, these complex nonlinear systems typically involve unresolved variables or unknown hidden processes that have no direct observations

[9, 10]. Nevertheless, these unobserved processes play important roles in transferring nonlinear energy and influencing the variables in the resolved scales. On the other hand, even the observations for the large- or resolved-scale physical variables can sometimes be very sparse and there is little available observations in certain areas such as the deep ocean [11, 12]. Therefore, developing efficient parameter estimation algorithms using partial observations becomes important for understanding and predicting these complex nonlinear turbulent systems, especially in high dimensions.

Various linear or nonlinear optimization methods have been proposed for estimating parameters in turbulent systems [14, 15]. Unfortunately, most of these methods will either be trapped into a local optimal solution or are too time-consuming to be applied to high dimensional systems. Regarding the model parameters as augmented state variables, algorithms based on particle or ensemble Kalman filters were designed for parameter estimation [16, 17, 18, 19, 20]. These sequential methods are widely used in practice and provide some successful results. However, these methods sometimes have extremely slow convergence that requires a large amount of training data and the optimality of the solutions is often not guaranteed due to the ignorance of the higher order moments in the ensemble Kalman filters. Another practically useful approach for estimating parameters in complex turbulent systems is to apply random sampling techniques, such as the Markov chain Monte Carlo (MCMC) algorithms [21, 22, 23, 24], and find the optimal solutions based on Bayesian inference. In the presence of partial observations, MCMC algorithms are often combined with data augmentation [25, 26], which samples the trajectories associated with the unresolved variables, to facilitate the calculation of the likelihood function. Such a combination of the MCMC algorithms with data augmentation can usually result in the global optimal solutions due to the random search technique. Yet, sampling the unobserved trajectories using data augmentation is computationally expensive, which prevents such an approach from being applied to high-dimensional complex turbulent dynamical systems.

In this article, a new efficient algorithm is developed for estimating param-

eters in complex turbulent dynamical systems with conditional Gaussian structures [27]. Decomposing the state variables $\mathbf{u}$ into two groups $\mathbf{u} = (\mathbf{u}_\mathbf{I}, \mathbf{u}_\mathbf{II})$ with $\mathbf{u}_\mathbf{I} \in R^{N_\mathbf{I}}$ and $\mathbf{u}_\mathbf{II} \in R^{N_\mathbf{II}}$, the conditional Gaussian systems are characterized by the fact that once a single trajectory of $\mathbf{u}_\mathbf{I}(s \leq t)$ is given, $\mathbf{u}_\mathbf{II}(t)$ conditioned on $\mathbf{u}_\mathbf{I}(s \leq t)$ becomes a Gaussian process. Despite the conditional Gaussianity, the coupled systems remain highly nonlinear and is able to capture strong non-Gaussian features such as skewed or fat-tailed distributions as observed in nature [27]. Many complex turbulent dynamical systems belong to this conditional Gaussian model family, such as the noisy versions of the Lorenz models, the Boussinesq equations with noise and quite a few stochastically coupled reaction-diffusion models in neuroscience and ecology. A gallery of examples of conditional Gaussian systems can be found in [28]. One of the desirable features of such conditional Gaussian system is that it allows closed analytical formulae for solving the conditional distribution $p(\mathbf{u}_\mathbf{II}(t) | \mathbf{u}_\mathbf{I}(s \leq t))$ [29]. Applications of the conditional Gaussian systems to strongly nonlinear systems include predicting the intermittent time-series of the Madden-Julian oscillation (MJO) and monsoon intraseasonal variabilities [30, 31, 32], filtering the stochastic skeleton model for the MJO [33], and recovering the turbulent ocean flows with noisy observations from Lagrangian tracers [34, 35, 36]. Other studies that also fit into the conditional Gaussian framework includes the cheap exactly solvable forecast models in dynamic stochastic superresolution of sparsely observed turbulent systems [37, 38], stochastic superparameterization for geophysical turbulence [39], physics constrained nonlinear regression models [40, 41] and blended particle filters for large-dimensional chaotic systems [42].

The conditional Gaussian framework provides an efficient way of estimating parameters in complex turbulent dynamical systems with only partial observations $\mathbf{u}_\mathbf{I}$. In fact, the analytically solvable conditional statistics allows a deterministic and computationally efficient approach to recover the trajectories associated with the unobserved variables $\mathbf{u}_\mathbf{II}$, which are also optimal based on the Bayesian inference. This deterministic method circumvents the expensive and time-consuming random sampling of the trajectories associated with the

75 unobserved variables in the infinite dimensional space using data augmentation
and thus greatly enhances the computational efficiency. With these recovered
unobserved trajectories, the MCMC technique can then be applied to sample
the parameters. The well-established MCMC theory guarantees that the esti-
mated parameters are globally optimal. In addition, a short training period is
80 typically sufficient in this new parameter estimation algorithm and it is therefore
practically useful.

Next, many complex turbulent dynamical systems in nature have a large
dimension and contain quite a few parameters. In such a scenario, the direct
application of both the MCMC algorithms for sampling parameters and the con-
85 ditional Gaussian framework for recovering the unobserved trajectories can be
computationally expensive. To overcome this difficulty, two effective strategies
are developed and incorporated into the above parameter estimation algorithm
(and the new version is named as the improved algorithm). These strategies
are developed according to the salient features of many complex systems with
90 multiscale structures [39], multilevel dynamics [43] or state-dependent parame-
terizations [37]. Here, the first strategy involves a judicious block decomposition
of the state variables such that the original problem is divided into several sub-
problems [44]. These subproblems are coupled in a specific way that allows
an extremely efficient parallel computation for the parameter estimation due
95 to the small size of each individual subproblem. The second strategy exploits
statistical symmetry for a further reduction of the computational cost when the
system is statistically homogeneous. A two-layer Lorenz 96 model [45, 43, 44]
that mimics the realistic features of atmosphere wave propagations and excitable
media is used to test the parameter estimation skill with the improved algorithm.
100 m. This model contains 80 state variables and 162 parameters. The estimated
parameters of the improved algorithm are accurate even with a short training
period and the algorithm is computationally efficient.

The rest of the article is organized as follows. Section 2 introduces the con-
ditional Gaussian nonlinear turbulent dynamical systems. A quick review of the
105 MCMC algorithms and data augmentation is included in Section 3. The new

efficient parameter estimation algorithm involving using the conditional Gaussian statistics for recovering the unobserved trajectories and using the MCMC technique for sampling parameters is shown in Section 4, which is followed by Section 5 that includes the application of the new parameter estimation algorithm to three test examples with strong nonlinear and non-Gaussian features. The improved algorithm by incorporating the judicious block decomposition and statistical symmetry into the basic version is described in Section 6. The parameter estimation of a two-layer Lorenz 96 model using the improved algorithm is shown in Section 7. The article is concluded in Section 8.

2. Conditional Gaussian nonlinear turbulent dynamical systems

2.1. Conditional Gaussian system

The conditional Gaussian systems have the following abstract form [27],

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

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

where $\mathbf{u}_I$ usually represents the observed variables and $\mathbf{u}_{II}$ represents the unobserved ones. Both $\mathbf{u}_I$ and $\mathbf{u}_{II}$ are multidimensional. In (1), $\mathbf{A}_0, \mathbf{A}_1, \mathbf{a}_0, \mathbf{a}_1, \Sigma_I$ and Σ_{II} are vectors and matrices that depend only on time t and the state variables $\mathbf{u}_I$, and $\mathbf{W}_I(t)$ and $\mathbf{W}_{II}(t)$ are independent Wiener processes. In the coupled system (1), 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,

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

Despite the conditional Gaussianity, the coupled system (1) remains highly nonlinear and is able to capture the non-Gaussian features as in nature. Many complex turbulent dynamical systems belong to this conditional Gaussian model family, such as the noisy versions of the Lorenz models, the Boussinesq equations with noise and quite a few stochastically coupled reaction-diffusion models in neuroscience and ecology. A gallery of examples of conditional Gaussian systems can be found in [28].

2.2. Closed analytic form of the conditional Gaussian statistics

One of the important features of the conditional Gaussian system (1) is that the conditional Gaussian distribution $p(\mathbf{u}_{\mathbf{II}}(t)|\mathbf{u}_{\mathbf{I}}(s \leq t))$ in (2) has closed analytic form [29],

$$d\bar{\mathbf{u}}_{\mathbf{II}}(t) = [\mathbf{a}_0(t, \mathbf{u}_{\mathbf{I}}) + \mathbf{a}_1(t, \mathbf{u}_{\mathbf{I}})\bar{\mathbf{u}}_{\mathbf{II}}]dt + (\mathbf{R}_{\mathbf{II}}\mathbf{A}_1^*(t, \mathbf{u}_{\mathbf{I}}))(\Sigma_{\mathbf{I}}\Sigma_{\mathbf{I}}^*)^{-1}(t, \mathbf{u}_{\mathbf{I}}) \times \\ [d\mathbf{u}_{\mathbf{I}} - (\mathbf{A}_0(t, \mathbf{u}_{\mathbf{I}}) + \mathbf{A}_1(t, \mathbf{u}_{\mathbf{I}})\bar{\mathbf{u}}_{\mathbf{II}})dt], \quad (3a)$$

$$d\mathbf{R}_{\mathbf{II}}(t) = \left\{ \mathbf{a}_1(t, \mathbf{u}_{\mathbf{I}})\mathbf{R}_{\mathbf{II}} + \mathbf{R}_{\mathbf{II}}\mathbf{a}_1^*(t, \mathbf{u}_{\mathbf{I}}) + (\Sigma_{\mathbf{II}}\Sigma_{\mathbf{II}}^*)(t, \mathbf{u}_{\mathbf{I}}) \right. \\ \left. - (\mathbf{R}_{\mathbf{II}}\mathbf{A}_1^*(t, \mathbf{u}_{\mathbf{I}}))(\Sigma_{\mathbf{I}}\Sigma_{\mathbf{I}}^*)^{-1}(t, \mathbf{u}_{\mathbf{I}})(\mathbf{R}_{\mathbf{II}}\mathbf{A}_1^*(t, \mathbf{u}_{\mathbf{I}}))^* \right\} dt, \quad (3b)$$

125 which can be solved in an exact and efficient way. The recovery of the conditional statistics in (3) is in an optimal way based on the Bayesian inference [29]. In fact, the formulae in (3) correspond to the optimal filters of the conditional Gaussian system (1) that combines the observational information in the processes of $\mathbf{u}_{\mathbf{I}}$ and the dynamical information in the processes of $\mathbf{u}_{\mathbf{II}}$. The closed analytical
130 formulae in (3) have been widely applied for state estimation, filtering, data assimilation and prediction [28]. Note that the classical Kalman-Bucy filter [46, 47, 48, 49] is the simplest and special filtering (or data assimilation) example within the conditional Gaussian framework.

3. A quick review of the MCMC algorithm with data augmentation

The Markov Chain Monte Carlo (MCMC) method [21, 22, 23, 24] has been applied for sampling from a probability distribution. It is also widely used for parameter estimation in stochastic systems given observations. The basic idea of the MCMC algorithm in parameter estimation is via the random sampling and the Bayesian approach. Assume $\boldsymbol{\theta}$ contains all the model parameters and $\mathbf{u}$ includes all the observational variables. The goal of the MCMC method is to explore the conditional distribution (or the so-called posterior distribution) of $\boldsymbol{\theta}$ given $\mathbf{u}$ by constructing a Markov chain. Using the Bayesian approach, the parameter estimation has the following form,

$$p(\boldsymbol{\theta}|\mathbf{u}) \propto p(\boldsymbol{\theta})p(\mathbf{u}|\boldsymbol{\theta}), \quad (4)$$

135 where $p(\boldsymbol{\theta})$ is the prior distribution while $p(\mathbf{u}|\boldsymbol{\theta})$ is the likelihood function. Clearly, if the likelihood function is known, then constructing the Markov chain becomes straightforward. The basic procedure of using the MCMC for parameter estimation is as follows.

1. Generate an initial guess of the parameters $\boldsymbol{\theta}^{(0)}$.
- 140 2. At each iteration step k , generate a candidate $\boldsymbol{\theta}^{(k)}$ for the next sample by picking up from the distribution (known as the proposal function) $Q(\boldsymbol{\theta}^{(k)}|\boldsymbol{\theta}^{(k-1)})$.
3. Then calculate the acceptance ratio

$$\alpha = \frac{Q(\boldsymbol{\theta}^{(k-1)}|\boldsymbol{\theta}^{(k)})p(\boldsymbol{\theta}^{(k)})p(\mathbf{u}|\boldsymbol{\theta}^{(k)})}{Q(\boldsymbol{\theta}^{(k)}|\boldsymbol{\theta}^{(k-1)})p(\boldsymbol{\theta}^{(k-1)})p(\mathbf{u}|\boldsymbol{\theta}^{(k-1)})}$$

4. Accept the new candidate $\boldsymbol{\theta}^{(k)}$ if $\alpha > 1$. Otherwise, accept the new candidate with a certain probability. If the new candidate is rejected, 145 then set $\boldsymbol{\theta}^{(k)} = \boldsymbol{\theta}^{(k-1)}$.

In Step 2, one simple and practical algorithm is the Metropolis algorithm [50], where Q is a symmetric function, namely $Q(\boldsymbol{\theta}^{(k)}|\boldsymbol{\theta}^{(k-1)}) = Q(\boldsymbol{\theta}^{(k-1)}|\boldsymbol{\theta}^{(k)})$. Therefore, a Gaussian proposal is a natural choice. In addition, in Step 4, a classical way is to generate a uniform random number μ on $[0, 1]$. If $\mu < \alpha$, 150 then the candidate is accepted.

However, in most of the realistic situations, only partial observations are available. In other words, only the signals of $\mathbf{u}_I$ in (1) are observed while the trajectories of $\mathbf{u}_{II}$ are completely unavailable. In such a scenario, the likelihood function typically has no closed form due to the lack of information of the trajectories of $\mathbf{u}_{II}$. Data augmentation [25, 26] is a common approach to sample the missing trajectories (or paths) of $\mathbf{u}_{II}$, which are then incorporated in computing the likelihood function. The Bayesian framework (4) is modified accordingly,

$$\begin{aligned} p(\boldsymbol{\theta}, \mathbf{u}_{II}^{mis}|\mathbf{u}_I) &\propto p(\boldsymbol{\theta})p(\mathbf{u}_I, \mathbf{u}_{II}^{mis}|\boldsymbol{\theta}) \\ &\propto p(\boldsymbol{\theta})p(\mathbf{u}_{II}^{mis}|\boldsymbol{\theta})p(\mathbf{u}_I|\mathbf{u}_{II}^{mis}, \boldsymbol{\theta}), \end{aligned} \tag{5}$$

where $\mathbf{u}_{\mathbf{I}}$ again represents the observed variables while $\mathbf{u}_{\mathbf{I}}^{mis}$ contains the missing path of the unobserved variables. However, sampling the missing paths of $\mathbf{u}_{\mathbf{I}}$ itself is a difficult and time-consuming task since the dimension of the missing path is infinity (or at least quite large using discrete approximations). A typical
155 strategy in sampling the missing path is to divide the entire missing paths into a large number of blocks, each containing a short interval, and then sample the missing paths in different small intervals alternatively [51, 52, 53]. Meanwhile, Gaussian approximations in these small intervals are often used for simplicity [54]. Another approach is to apply particle filter to estimate the parameters and
160 states, which aims at estimating and improving the parameters dynamically once new observations are available [55, 56, 57]. These methods work quite well for low-dimensional cases but the high computational cost in the particle methods impedes them to work efficiently for large-dimensional systems. In addition, Gaussian approximation may also be inaccurate for strongly nonlinear systems.

165 **4. A new efficient algorithm for parameter estimation of conditional Gaussian systems**

As discussed in Section 3, sampling the infinite dimensional trajectories of the unobserved variables $\mathbf{u}_{\mathbf{I}}$ (or large dimensional ones in the discrete form) itself is computationally expensive. Below, a new parameter estimation algo-
170 rithm is developed. This new algorithm makes use of the closed form of the conditional Gaussian statistics (3) to recover the unobserved trajectories of $\mathbf{u}_{\mathbf{I}}$ in a deterministic and optimal way, which is then combined with the MCMC algorithm to efficiently sample and estimate the model parameters.

175 *4.1. An efficient, deterministic and optimal approach to recover the unobserved trajectories*

Recall that the conditional Gaussian systems (1) allow closed analytic formulae for the conditional distribution (3). This property can be incorporated into the MCMC algorithm in Section 3 to circumvent the most expensive part

of the parameter estimation algorithm, namely sampling the unobserved trajectories using data augmentation. Specifically, in each iteration step, we make use of the observed trajectories $\mathbf{u}_I$ and the current updated parameters $\boldsymbol{\theta}^{(k)}$ to recover the unobserved trajectories of $\mathbf{u}_{II}$, namely $\mathbf{u}_{II}^{mis,(k)}$, which are given by the paths of the conditional Gaussian mean (3a). In fact, since the mode of a Gaussian distribution equals its mean, the trajectories of the conditional mean are the optimal recovery of $\mathbf{u}_{II}$ given the observations $\mathbf{u}_I$ based on the Bayesian inference. While the traditional data augmentation approaches aim at sampling the unobserved trajectories, in the conditional Gaussian framework once $\mathbf{u}_I$ and $\boldsymbol{\theta}^{(k)}$ are given, the unobserved trajectories $\mathbf{u}_{II}^{mis,(k)}$ are computed via the cheap and explicit formulae (3a) in a deterministic and optimal way. Note that the conditional mean depends on the conditional covariance and therefore the information of the latter is implicitly included. Finally, $\mathbf{u}_{II}^{mis,(k)}$, $\mathbf{u}_I$ and $\boldsymbol{\theta}^{(k)}$ are used together to compute the likelihood function in (4)

$$p(\mathbf{u}_I|\boldsymbol{\theta}^{(k)}) = p(\mathbf{u}_I|\boldsymbol{\theta}^{(k)}; \mathbf{u}_{II}^{mis,(k)}).$$

4.2. The algorithm

Now, we summarize the new MCMC algorithm involving using the conditional Gaussian statistics for recovering the unobserved trajectories. Again, assume $\mathbf{u}_I$ and $\mathbf{u}_{II}$ are the observed and unobserved variables, respectively. We denote $\boldsymbol{\mu}$ the conditional mean of $\mathbf{u}_{II}$ given the trajectories of $\mathbf{u}_I$. We use $\boldsymbol{\theta}^{(k)}$ to represent the parameters in the k -th step. The algorithm is as follows:

1. Generate an initial guess of the parameters $\boldsymbol{\theta}^{(0)}$. Apply the closed analytical formulae for the conditional Gaussian statistics (3) to recover the corresponding unobserved trajectories $\boldsymbol{\mu}^{(0)}$. Compute the likelihood function $p(\mathbf{u}|\boldsymbol{\theta}^{(0)}) = p(\mathbf{u}|\boldsymbol{\theta}^{(0)}; \boldsymbol{\mu}^{(0)})$.
2. For $k = 1, 2, \dots$
 - (a) At each iteration step k , generate a candidate $\boldsymbol{\theta}^{(k)}$ for the next sample by picking from a given proposal distribution $Q(\boldsymbol{\theta}^{(k)}|\boldsymbol{\theta}^{(k-1)})$, which is Gaussian here. Under the new parameter candidate, apply the

- 190 closed analytical formulae for the conditional Gaussian statistics (3)
 and obtain the corresponding unobserved trajectories $\boldsymbol{\mu}^{(k)}$.
- (b) Calculate the acceptance ratio $\alpha = p(\boldsymbol{\theta}^{(k)})p(\mathbf{u}|\boldsymbol{\theta}^{(k)})/p(\boldsymbol{\theta}^{(k-1)})p(\mathbf{u}|\boldsymbol{\theta}^{(k-1)})$,
 where $p(\mathbf{u}|\boldsymbol{\theta}^{(k)}) = p(\mathbf{u}|\boldsymbol{\theta}^{(k)}; \boldsymbol{\mu}^{(k)})$.
- (c) Accept the new candidate $\boldsymbol{\theta}^{(k)}$ and $\boldsymbol{\mu}^{(k)}$ if $\alpha > 1$. Otherwise, ac-
 195 cept the new candidate with probability $1/(1 - \beta \log \alpha)$. If the new
 candidate is rejected, then set $\boldsymbol{\theta}^{(k)} = \boldsymbol{\theta}^{(k-1)}$, $\boldsymbol{\mu}^{(k)} = \boldsymbol{\mu}^{(k-1)}$.
3. Ignoring the burn-in period for $k < k_{start}$, average over the sampled pa-
 rameters from $k = k_{start}$ to $k = k_{end}$ to obtain the estimated values of the
 parameters.

200 4.3. Technique details

(a). Prior distributions.

The prior distributions of the parameters $\boldsymbol{\theta}$ need to be prescribed in order to ap-
 ply the Bayesian method in (4). Here we always assume we have almost no prior
 knowledge about the parameters. Thus, we set the prior distribution of most of
 205 the parameters to be a uniform distribution in a large interval $[-10^6, 10^6]$. For
 diffusion coefficients, we adopt a uniform distribution in $[10^{-6}, 10^6]$ for each of
 them since by definition they are positive.

(b). Computing the likelihood function.

Although the observations are continuous, we compute and store the data of $\mathbf{u}_{\mathbf{I}}$
 210 in a discrete way in the numerical simulations with a small time step Δt . Here
 Euler-Maruyama scheme is used in the numerical simulation. The likelihood at
 each time step is computed as follows. Starting from the observational value
 of $\mathbf{u}_{\mathbf{I}}$ at the time instant t , say $\mathbf{u}_{\mathbf{I}}^{obs}(t)$, and computing the distribution of $\mathbf{u}_{\mathbf{I}}$
 one time step forward using Euler-Maruyama scheme, the result at $t + \Delta t$ is a
 215 Gaussian distribution $p_{pred}(\mathbf{u}_{\mathbf{I}}(t + \Delta t))$, where the evolutions of the mean and
 covariance are solved through analytical formulae. Then plugging the observa-
 tional data at time $t + \Delta t$, i.e., $\mathbf{u}_{\mathbf{I}}^{obs}(t + \Delta t)$, into $p_{pred}(\mathbf{u}_{\mathbf{I}}(t + \Delta t))$ leads to
 the likelihood at time $t + \Delta t$. The full likelihood function is the product of the
 likelihood components at different time instants.

220 (c). *The setup of MCMC in sampling parameters.*

In the algorithm presented above, only a simple and basic MCMC sampling technique is adopted for updating the parameters. Here, instead of adopting an adaptive selection, the proposal functions are prescribed at the beginning. We slightly tune the parameter β in Step 2(c) such that the acceptance ratio
225 is around 40% to 50%. To avoid numerical issue, we also take a logarithm likelihood instead of likelihood itself in Step 2(c). Note that other choices can be used in accepting new candidate. Yet, the emphasis here is to use efficient conditional Gaussian solver (3) to replace the expensive data augmentation and therefore only this basic MCMC algorithm is used. Incorporating more advanced
230 MCMC algorithms is left as a future work.

(d). *Estimating the diffusion coefficients in the dynamics of $\mathbf{u}_I$.*

While the standard MCMC algorithm is able to easily estimate the diffusion coefficients in the dynamics of $\mathbf{u}_I$, these diffusion coefficients can be estimated in a more efficient way here due to the fact that continuous observations are
235 available. In fact, with continuous observations, the quadratic variation of the observed signal can be obtained (see Appendix), which can be used to estimate the diffusion coefficients in the observed process $\mathbf{u}_I$. For more complex systems, the diffusion coefficients can also be easily estimated by applying a crude MCMC algorithm using a simplified drift term and run the MCMC algorithm for a small
240 number of iterations before estimating other parameters.

5. Test examples of nonlinear turbulent dynamical systems

In this section, the parameter estimation algorithm developed in Section 4 is tested on three nonlinear turbulent dynamical models. All the tests are conducted on a desktop with Ubuntu 18.04.1 LTS system using MATLAB 2018a
245 (academic version). The computational time of all the test examples in this section is just a few seconds.

5.1. A noisy version of the Lorenz 63 model

The first test model is a noisy version of the Lorenz 63 (L-63) model [58],

$$dx = \sigma(y - x)dt + \sigma_x dW_x, \quad (6a)$$

$$dy = (x(\rho - z) - y)dt + \sigma_y dW_y, \quad (6b)$$

$$dz = (xy - \beta z)dt + \sigma_z dW_z, \quad (6c)$$

The deterministic L-63 model ($\sigma_x = \sigma_y = \sigma_z = 0$ in (6)) was proposed by Lorenz in 1963 [58]. It is a simplified mathematical model for atmospheric convection.

250 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
255 parameters proportional to the Prandtl number, Rayleigh number, and certain physical dimensions of the layer itself [59]. The L-63 model is also widely used as simplified models for lasers, dynamos, thermosyphons, electric circuits, chemical reactions and forward osmosis [60, 61, 62, 63, 64, 65, 66]. the noisy version of the L-63 includes more turbulent and small-scale features and their interactions
260 with the three large scale variables while it retains the characteristics in the original L-63. The noisy L-63 model is a conditional Gaussian system (1) with $\mathbf{u}_I = x$ and $\mathbf{u}_{II} = (y, z)^T$.

In order to test the parameter estimation skill, the following parameters are used to generate the true signal,

$$\sigma_x = \sigma_y = \sigma_z = 5, \quad \sigma = 10, \quad \rho = 28, \quad \beta = 8/3. \quad (7)$$

The three parameters σ , ρ and β are the classical choices that result in a butterfly profile of the L-63 model. The three noise coefficients provide moderate noise
265 such that the dynamics has some small scale features with nonlinearity and multiplicative noise while retaining the rough butterfly profile. A simulation of the L-63 model (6) with these true parameters in (7) is shown in Figure 1.

Now assume the trajectory of x in (1) is the only available observation. In other words, there is no direct observation for y and z . Our goal here is to estimate the parameters in the noisy L-63 model with such partial observations. As described in the Section 4.3, the noise coefficient σ_x is estimated by the quadratic variation formula. Due to the continuous observations, the diffusion coefficient σ_u can be estimated perfectly. To estimate the other parameters, we first assign some initial guesses for these parameters in the MCMC algorithm,

$$\sigma_y^{(0)} = \sigma_z^{(0)} = 10, \quad \sigma^{(0)} = 15, \quad \rho^{(0)} = 42, \quad \beta^{(0)} = 14/3. \quad (8)$$

Note that these initial guesses are far from the truth (7), where the errors in the initial guesses for all the parameters are more than 50%. The proposal functions of the updates are given as follows,

$$\begin{aligned} \sigma_y^{(k)} &= \sigma_y^{(k-1)} + 0.1X_1, & \sigma_z^{(k)} &= \sigma_z^{(k-1)} + 0.1X_2, \\ \sigma^{(k)} &= \sigma^{(k-1)} + 0.2X_3, & \rho^{(k)} &= \rho^{(k-1)} + 0.3X_4, & \beta^{(k)} &= \beta^{(k-1)} + 0.05X_5, \end{aligned}$$

where $X_i, i = 1, \dots, 5$ are random numbers generated from independent standard Gaussian distributions. Note that the proposal density has a larger variance for those variables which themselves are larger.

The parameter estimation results are shown in Figure 2, where the marginal posterior distribution for each parameter is formed by collecting the points in the corresponding trace plot from $k = 1000$ to $k = 10000$. It is clear from the trace plots that the large biases in the initial guesses of the parameters are eliminated quickly. The estimation of the parameters are overall quite accurate in the sense that the averaged value of the trace plot for most of the parameters is almost the same as the truth and the uncertainty in the marginal posterior distribution is typically small. Among all the estimated parameters, the one that has the largest error is ρ . Nevertheless, the averaged estimation value 30.4 is only 8% larger than the truth $\rho = 28$. Note that it is hard to make a conclusion by simply looking at the relative error that whether the parameter estimation algorithm is skillful. In fact, the deviation in the estimated parameters is probably due to the use of a short training period, which has only 100

units here. A more reasonable approach of measuring the parameter estimation
285 skill is to compare the dynamical and the statistical behavior of the noisy L-96
model using the estimated parameters and using the true ones. To this end, we
plug these estimated parameters into the noisy L-63 model and run the model
simulation, the results of which are shown in Figure 3 and they are named as
model predictions (with estimated parameters). It is clear that the x, y and z
290 trajectories in Figure 3 are all similar to those in Figure 1. Note that these
predictions are all free runs of the model. Due to the stochasticity and chaotic
nature, we do not expect a point-to-point correspondence in the trajectories in
these two figures. Nevertheless, both the dynamical and the statistical features
represented by the temporal autocorrelation functions (ACFs) and the probab-
295 ity density functions (PDFs) for the truth and the prediction are very similar to
each other. Thus, we reach the conclusion that these parameters are estimated
with high accuracy.

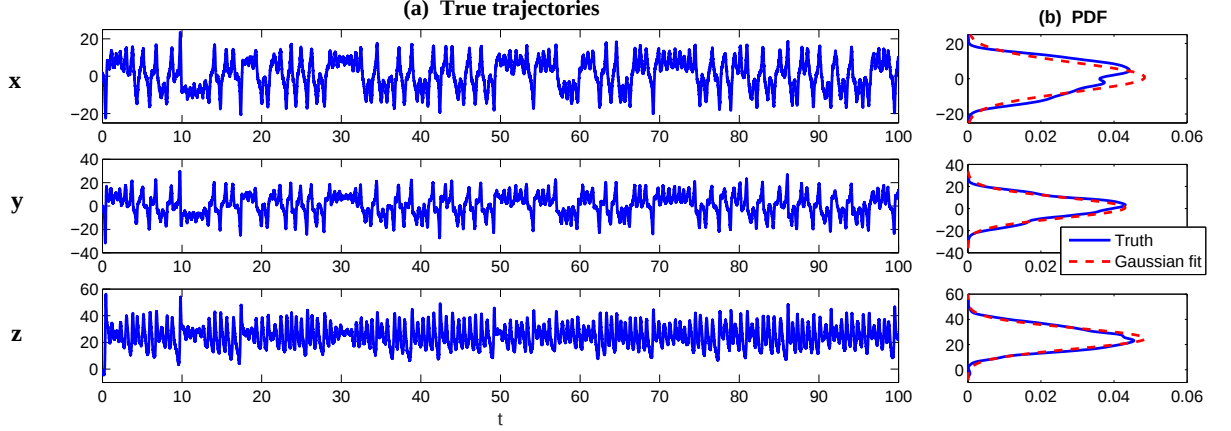


Figure 1: Simulations of the L-63 model (6) with true parameters in (7). Panel (a): Trajectories of x , y and z . Panel (b): The associated probability density functions (PDFs).

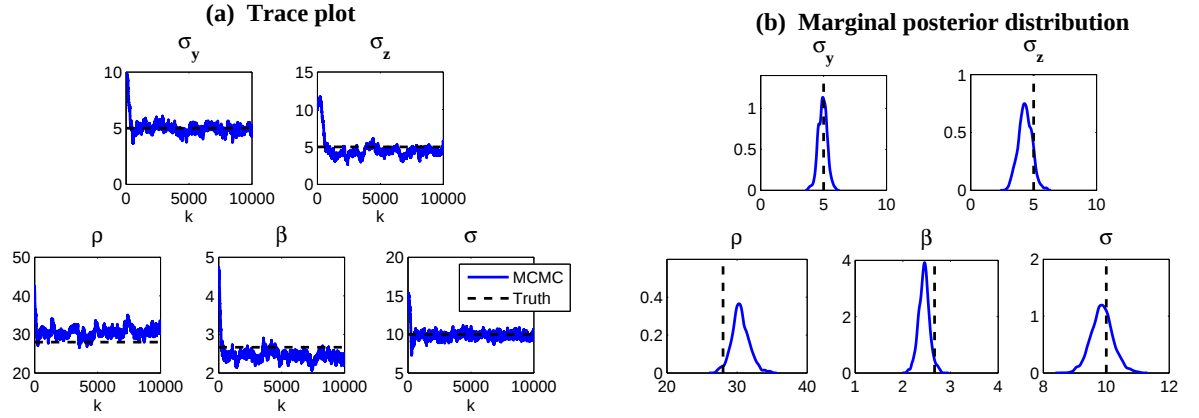


Figure 2: The parameter estimation skill of the noisy L-63 model. Panel (a): Trace plots. Here k is the iteration step. Panel (b): Marginal posterior distributions formed by collecting the points in the corresponding trace plot from $k = 1000$ to $k = 10000$. The black dashed line in each subplot shows the true parameter values in (7).

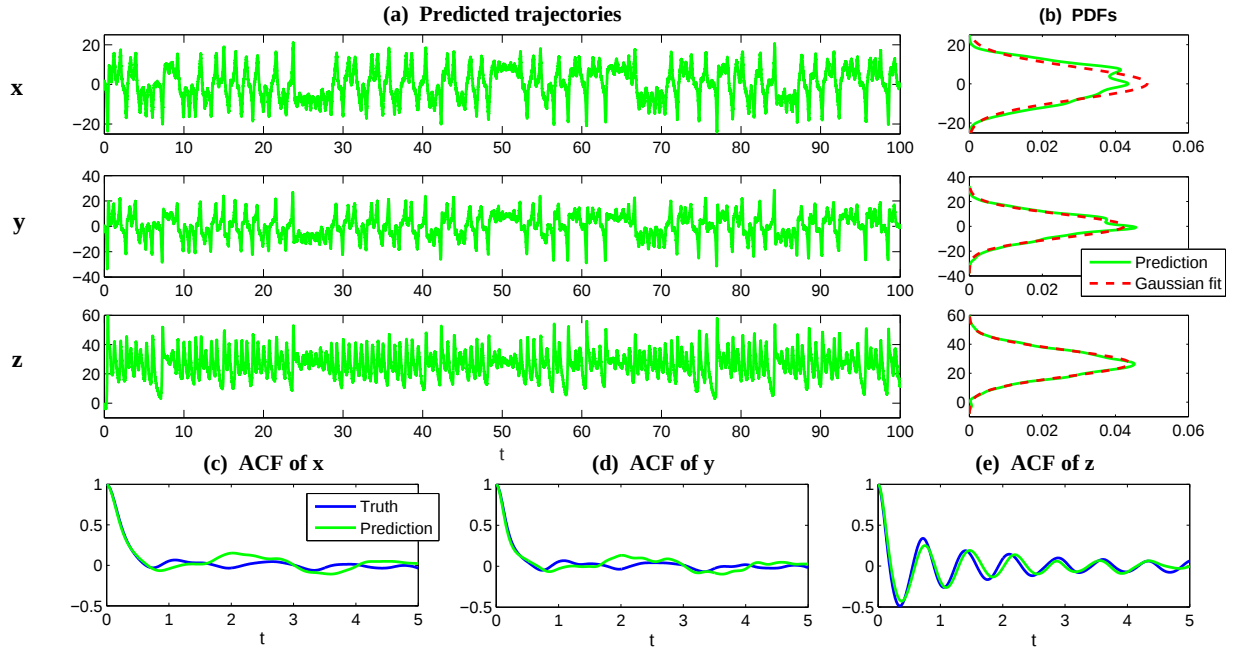


Figure 3: Predictions of the L-63 model (6) with estimated parameters in Figure 2 (averaged from $k = 1000$ to $k = 10000$). Panel (a): Trajectories of x , y and z . Panel (b): The associated PDFs. Panels (c)–(e): Comparison of the autocorrelation functions (ACFs) in the prediction (green) with those of the truth (blue) from the signal in Figure 1.

5.2. The SPEKF model with multiplicative noise

The parameter estimation is often the prerequisite of data assimilation and prediction. Therefore, the second test model of the parameter estimation algorithm developed in Section 4 is the so-called stochastic parameterized extended Kalman filter (SPEKF) model [67, 68, 69], which has been used for filtering and predicting complex turbulent dynamical systems [70, 9, 1]. Here we focus on a simplified version, namely the SPEKF-M model ('M' stands for 'multiplicative'), which nevertheless contains a multiplicative noise and is able to generate strongly non-Gaussian features,

$$\begin{aligned} du &= (-\gamma u + F)dt + \sigma_u dW_u, \\ d\gamma &= -d_\gamma(\gamma - \hat{\gamma})dt + \sigma_\gamma dW_\gamma. \end{aligned} \tag{9}$$

The SPEKF-M model (9) belongs to the conditional Gaussian framework with $\mathbf{u_I} = u$ and $\mathbf{u_{II}} = \gamma$, where u is the observed variable while the trajectory of γ is hidden from observations. The variable γ interacts with u in a multiplicative way and it plays the role of a stochastic damping for the process of u . Note that the full SPEKF model [67, 68, 69] includes a stochastic phase and a stochastic forcing in addition to this stochastic damping. In addition to filtering and predicting intermittent signals from nature in the presence of model error [71, 72, 73, 74], other important applications of the SPEKF for complex spatial-extended systems include stochastic dynamical superresolution [37] and effective filters for Navier-Stokes equation [75].

Below, we test the parameter estimation algorithm in the SPEKF-M model (9) in two different dynamical regimes: a strong intermittent regime and a moderate intermittent regime. The true parameters are given as follows:

Strong intermittent regime :

$$\sigma_u = 0.5, \quad F = 1, \quad \sigma_\gamma = 1.2, \quad d_\gamma = 1, \quad \hat{\gamma} = 1, \tag{10}$$

Moderate intermittent regime :

$$\sigma_u = 0.5, \quad F = 1, \quad \sigma_\gamma = 1.2, \quad d_\gamma = 1, \quad \hat{\gamma} = 2.$$

The only difference in the true parameters is the mean damping $\hat{\gamma}$. In fact, with

the increase of $\hat{\gamma}$, the system becomes more stable and thus the intermittency becomes weaker. The simulations of the system using these true parameters are given in Figure 4, where intermittency and extreme events can both be seen in the trajectories of u and the associated PDFs are non-Gaussian fat-tailed. Note that in the parameter estimation tests here, only the trajectory of u in each regime is observed and there is no observation of γ . The trajectories of u in Figure 4 are used as the input of the algorithm. Again, the estimation of the diffusion coefficient σ_u in the observed process u is perfectly estimated via the quadratic variation formula. The initial guesses of the other parameters in the algorithm are given as follows:

Strong intermittent regime :

$$\text{Initial values : } F^{(0)} = 0, \quad \sigma_\gamma^{(0)} = 2, \quad d_\gamma^{(0)} = 0.2, \quad \hat{\gamma}^{(0)} = 2,$$

Moderate intermittent regime :

$$\text{Initial values : } F^{(0)} = 0, \quad \sigma_\gamma^{(0)} = 2, \quad d_\gamma^{(0)} = 0.2, \quad \hat{\gamma}^{(0)} = 1. \quad (11)$$

These initial guesses imply that the system at the initial state is unforced and the process of γ has a large uncertainty. In the parameter estimation algorithm, the proposal of the updates are given as follows,

$$\begin{aligned} \sigma_\gamma^{(k)} &= \sigma_\gamma^{(k-1)} + 0.05X_1, & d_\gamma^{(k)} &= d_\gamma^{(k-1)} + 0.05X_2, \\ \hat{\gamma}^{(k)} &= \hat{\gamma}^{(k-1)} + 0.05X_3, & F^{(k)} &= F^{(k-1)} + 0.05X_4, \end{aligned}$$

where $X_1, \dots, X_4$ are random numbers sampled from independent standard

310 Gaussian distributions.

The parameter estimation results in these two regimes are shown in Figure 5. Clearly, in both regimes, the initial errors decays quite fast and the algorithm provides good estimations of the parameters.

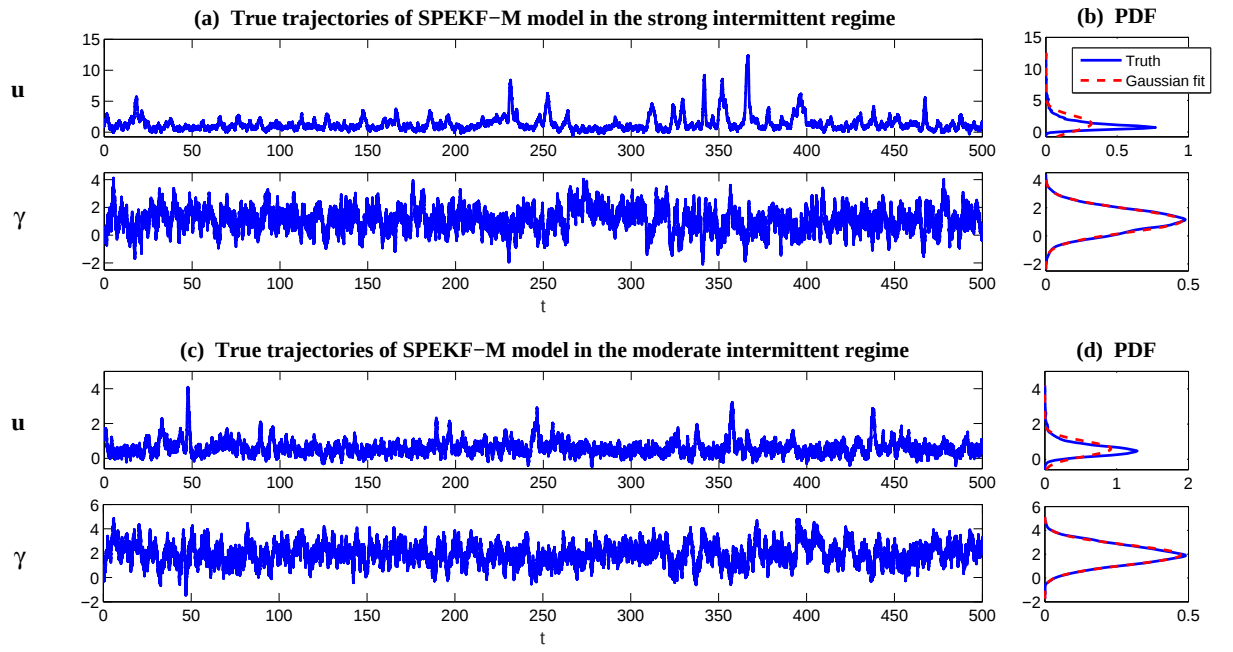


Figure 4: Simulation of the SPEKF-M model with true parameters in (10). Panels (a)–(b): model trajectories and the associated PDFs in strong intermittent regime. Panels (a)–(b): model trajectories and the associated PDFs in strong intermittent regime.

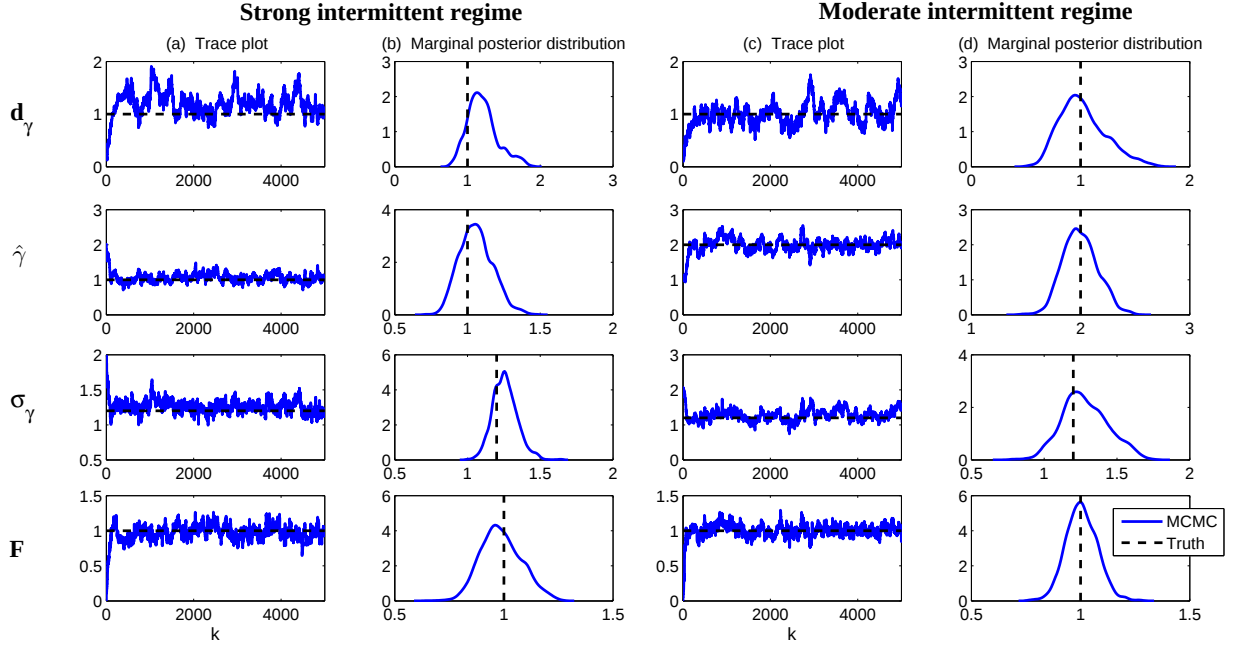


Figure 5: Parameter estimation skill of the SPEKF-M model. Panels (a) and (c): Trace plots. Here k is the iteration step. Panels (b) and (d): Marginal posterior distributions formed by collecting the points in the corresponding trace plot from $k = 1000$ to $k = 5000$. Panels (a)–(b) show the results in the strong intermittent regime while Panels (c)–(d) show those in the moderate intermittent regime. Here k is the iteration step. The black dashed line in each subplot shows the true parameter values in (10).

5.3. A physics-constrained dyad model

The last test example in this section is a physics-constrained dyad model [27],

$$\begin{aligned} du &= \left(-(cv + d_{uu})u + F \right) dt + \sigma_u dW_u, \\ dv &= \left(-d_{vv}v + cu^2 \right) dt + \sigma_v dW_v. \end{aligned} \quad (12)$$

315 The dyad model in (12) has energy-conserving nonlinear interaction [40, 41], which is an important feature of many turbulent dynamical system. The energy-conserving nonlinearity can be easily seen by multiplying u to the first equation and v to the second equation. Summing up the resulting equations gives the cancellation of the nonlinear terms. The difference between this physics-constrained
320 dyad model (12) and the SPEKF-M model (9) is that the latter only contains a one-way feedback from γ to u while the former has a two-way nonlinear interaction between u and v . Connecting the dyad model (12) with the conditional Gaussian framework (1), it is clear that the observational variable is $\mathbf{u}_\mathbf{I} = u$ and the unobserved one is $\mathbf{u}_\mathbf{II} = v$.

The following parameters are used as the true parameters in the dyad model:

$$d_{uu} = 0.8, \quad d_{vv} = 0.8, \quad c = 1.2, \quad \sigma_v = 2, \quad \sigma_u = 0.2, \quad F = 0.5. \quad (13)$$

325 The model simulations with these true parameters are shown in Figure 6. Intermittent instability is again observed in u . Compared with the SPEKF model, the strong intermittency here is more frequent. As in the tests of the SPEKF model, here only the signal of u is observed. The trajectory of u shown in Figure 6 is used as the input of the parameter algorithm. As in the previous
330 examples, the estimation of the diffusion coefficient σ_u in the observed process u is estimated via the quadratic variation formula.

Again, the initial guesses of the parameters are completely different from the truth,

$$d_{uu}^{(0)} = 3, \quad d_{vv}^{(0)} = 3, \quad c^{(0)} = 2.5, \quad \sigma_v^{(0)} = 0.5, \quad F^{(0)} = 2. \quad (14)$$

The proposal functions of the updates are given as follows,

$$\begin{aligned} d_{uu}^{(k)} &= d_{uu}^{(k-1)} + 0.05X_1, & d_{vv}^{(k)} &= d_{vv}^{(k-1)} + 0.05X_2, \\ c^{(k)} &= c^{(k-1)} + 0.05X_3, & \sigma_v^{(k)} &= \sigma_v^{(k-1)} + 0.05X_4, & F^{(k)} &= F^{(k-1)} + 0.05X_5, \end{aligned}$$

where $X_1, \dots, X_5$ are random numbers sampled from independent standard Gaussian distributions.

The parameter estimation results are shown in Figure 7. All the parameters are estimated with high skill. In addition, using the averaged value in the trace plots from $k = 1000$ to $k = 10000$ as the estimated parameters and plugging them into the physics-constrained dyad model (12), the model simulations provide essentially the same prediction skill in terms of the PDFs and the autocorrelation functions as in the noisy L-63 model. We omit these results here.

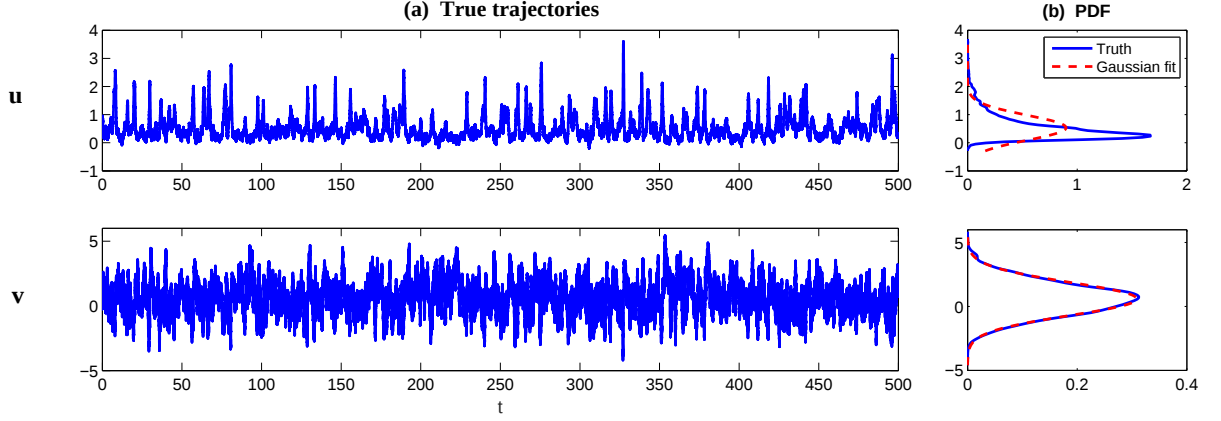


Figure 6: Simulations of the physics-constrained dyad model (12) with true parameters in (13). Panel (a): Trajectories of u and v . Panel (b): The associated PDFs.

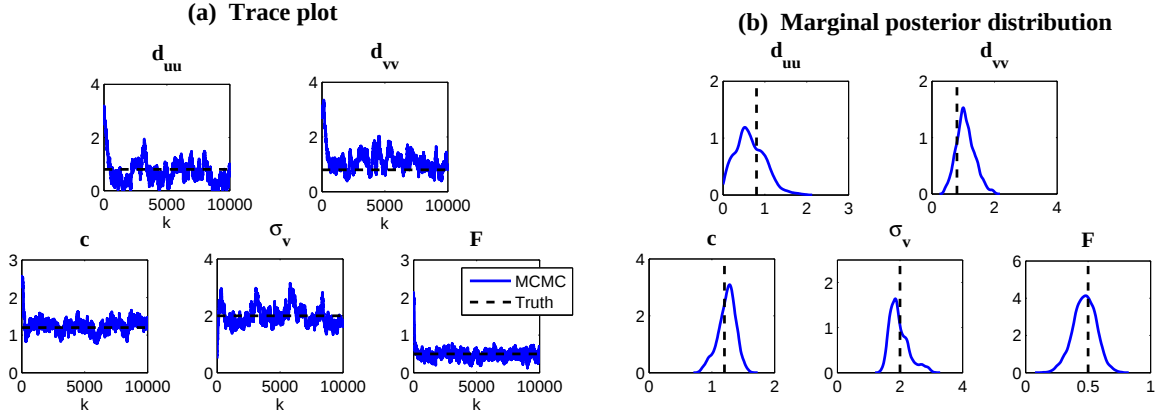


Figure 7: The parameter estimation skill of the physics-constrained dyad model (12). Panel (a): Trace plots. Here k is the iteration step. Panel (b): Marginal posterior distributions formed by collecting the points in the corresponding trace plot from $k = 1000$ to $k = 10000$. The black dashed line in each subplot shows the true parameter values in (13).

6. Improved algorithm for systems with large dimensions

It has been shown in the previous sections that the new algorithm explores the conditional Gaussian property to recover the trajectories of the unobserved variables in an efficient way, which facilitates the calculation of the likelihood function and allows the MCMC algorithm to efficiently sample the parameters. However, many complex turbulent dynamical systems in nature often have a large dimension and contain quite a few parameters. In such a scenario, direct application of both the MCMC algorithms for sampling parameters and the conditional Gaussian framework for recovering the unobserved trajectories can be computationally expensive. Therefore, a straightforward extension of the algorithm developed in Section 4 to high-dimensional turbulent dynamical systems may not be practical. Nevertheless, we can explore the dynamical features of many complex turbulent dynamical systems to improve the algorithm.

6.1. Block decomposition and divide and conquer

Many complex systems with multiscale structures [39], multilevel dynamics [43] or state-dependent parameterizations [37] have the following block decomposition features. The state variables can be divided into different groups $\mathbf{u}_k = (\mathbf{u}_{\mathbf{I},k}, \mathbf{u}_{\mathbf{II},k}) \in (\mathbb{R}^{N_{\mathbf{I},k}}, \mathbb{R}^{N_{\mathbf{II},k}})$, $k = 1, \dots, K$. In the dynamics of each $\mathbf{u}_{\mathbf{I},k}$ and $\mathbf{u}_{\mathbf{II},k}$ in (1), the terms $\mathbf{A}_{0,k}$ and $\mathbf{a}_{0,k}$ depend on all the components of $\mathbf{u}_{\mathbf{I}}$ while the terms $\mathbf{A}_{1,k}$ and $\mathbf{a}_{1,k}$ are only functions of $\mathbf{u}_{\mathbf{I},k}$, namely,

$$\begin{aligned} \mathbf{A}_{0,k} &:= \mathbf{A}_{0,k}(t, \mathbf{u}_{\mathbf{I}}), & \mathbf{a}_{0,k} &:= \mathbf{a}_{0,k}(t, \mathbf{u}_{\mathbf{I}}), \\ \mathbf{A}_{1,k} &:= \mathbf{A}_{1,k}(t, \mathbf{u}_{\mathbf{I},k}), & \mathbf{a}_{1,k} &:= \mathbf{a}_{1,k}(t, \mathbf{u}_{\mathbf{I},k}). \end{aligned} \quad (15)$$

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 $\mathbf{u}_{\mathbf{I},k}$ and $\mathbf{u}_{\mathbf{II},k}$, respectively. Therefore, the equation of each $\mathbf{u}_k = (\mathbf{u}_{\mathbf{I},k}, \mathbf{u}_{\mathbf{II},k})$ becomes

$$d\mathbf{u}_{\mathbf{I},k} = [\mathbf{A}_0(t, \mathbf{u}_{\mathbf{I}}) + \mathbf{A}_1(t, \mathbf{u}_{\mathbf{I},k})\mathbf{u}_{\mathbf{II},k}]dt + \boldsymbol{\Sigma}_{\mathbf{I}}(t, \mathbf{u}_{\mathbf{I},k})d\mathbf{W}_{\mathbf{I}}(t), \quad (16a)$$

$$d\mathbf{u}_{\mathbf{II},k} = [\mathbf{a}_0(t, \mathbf{u}_{\mathbf{I}}) + \mathbf{a}_1(t, \mathbf{u}_{\mathbf{I},k})\mathbf{u}_{\mathbf{II},k}]dt + \boldsymbol{\Sigma}_{\mathbf{II}}(t, \mathbf{u}_{\mathbf{I},k})d\mathbf{W}_{\mathbf{II}}(t). \quad (16b)$$

355 In addition, 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.

Notably, the models with block decomposition structures as described above have also been widely used in designing forecast models in filtering and predicting complex dynamical systems using dynamic stochastic superresolution of
 360 sparsely observed turbulent systems [37, 38] and stochastic superparameterization [39],

For the systems with such block decomposition features, divide and conquer can be applied to both the the MCMC algorithm for sampling parameters and the conditional Gaussian framework for recovering the unobserved trajectories.

365 6.1.1. Applying the block decomposition to solving the conditional Gaussian statistics

Under the condition (16), the conditional covariance matrix becomes block diagonal, which can be easily verified according to (3b). In fact, the evolution of the conditional covariance of $\mathbf{u}_{\mathbf{II},k}$ conditioned on $\mathbf{u}_{\mathbf{I}}$ is given by,

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

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 $k = 1, \dots, K$ can be solved in a parallel way and the
 370 computation is extremely efficient due to the small size of each individual block. This facilitates the algorithm to efficiently solve the covariance matrix in large dimensions.

Next, the structures of $\mathbf{A}_{0,k}$ and $\mathbf{a}_{0,k}$ in (15) allow the coupling among all the K groups of variables in the conditional mean according to (3a). 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

$$d\bar{\mathbf{u}}_{\mathbf{II},k}(t) = [\mathbf{a}_{0,k} + \mathbf{a}_{1,k}\bar{\mathbf{u}}_{\mathbf{II},k}]dt + \mathbf{R}_{\mathbf{II},k}\mathbf{A}_{1,k}^*(\boldsymbol{\Sigma}_{\mathbf{I},k}\boldsymbol{\Sigma}_{\mathbf{I},k}^*)^{-1}[d\mathbf{u}_{\mathbf{I},k} - (\mathbf{A}_{0,k} + \mathbf{A}_{1,k}\bar{\mathbf{u}}_{\mathbf{II},k})dt]. \quad (17)$$

Notably, each component of $\bar{\mathbf{u}}_{\mathbf{II},k}$ in (17) can be solved individually. In fact, the trajectories of $\mathbf{u}_{\mathbf{I}}$ from the observations can simply be regarded as the external
 375 input or forcing of solving the time evolution of $\bar{\mathbf{u}}_{\mathbf{II},k}$ in (17) via functions $\mathbf{a}_{0,k}$ and $\mathbf{A}_{0,k}$. Therefore, despite the fact that the full observation $\mathbf{u}_{\mathbf{I}}$ appears in the equation of each $\bar{\mathbf{u}}_{\mathbf{II},k}$, the equations of $\bar{\mathbf{u}}_{\mathbf{II},k}$ and $\bar{\mathbf{u}}_{\mathbf{II},k'}$ for $k \neq k'$ can still be solved independently since the right hand side of each $\bar{\mathbf{u}}_{\mathbf{II},k}$ does not depend on $\bar{\mathbf{u}}_{\mathbf{II},k'}$ and $\mathbf{R}_{\mathbf{II},k'}$ for any $k' \neq k$. This implies that the unobserved
 390 trajectories can be recovered in an independent and parallel way with the block decomposition.

6.1.2. Computing the likelihood function associated with each block – divide and conquer

According to (16a), it is clear that the likelihood function of $\mathbf{u}_{\mathbf{I},k}$ depends
 385 only on 1) the k -th component of $\mathbf{u}_{\mathbf{II}}$, namely $\mathbf{u}_{\mathbf{II},k}$, and 2) all the observed variables $\mathbf{u}_{\mathbf{I}}$. Note that although $\mathbf{u}_{\mathbf{I},k}$ interacts with other observed variables $\mathbf{u}_{\mathbf{I},k'}$ for all $k' \neq k$ through the full coupled model, the trajectories of all the $\mathbf{u}_{\mathbf{I},k'}$ in the equation of $\mathbf{u}_{\mathbf{I},k}$ are actually given from the observations and are not affected by the update of the parameters in the MCMC algorithm. In fact,
 390 the trajectories of all $\mathbf{u}_{\mathbf{I},k'}$ with $k' \neq k$ can be just regarded as the known input forcing in the equations of $\mathbf{u}_{\mathbf{I},k}$ and $\mathbf{u}_{\mathbf{II},k}$. This crucial property allows us to estimate different groups of the parameters, namely the parameters $\boldsymbol{\theta}_k$ in the equations $(\mathbf{u}_{\mathbf{I},k}, \mathbf{u}_{\mathbf{II},k})$ and $\boldsymbol{\theta}'_k$ in $(\mathbf{u}_{\mathbf{I},k'}, \mathbf{u}_{\mathbf{II},k'})$, independently since the corresponding likelihood functions are independent with each other.

395 A schematic illustration of the idea discussed above is shown in Figure 8.

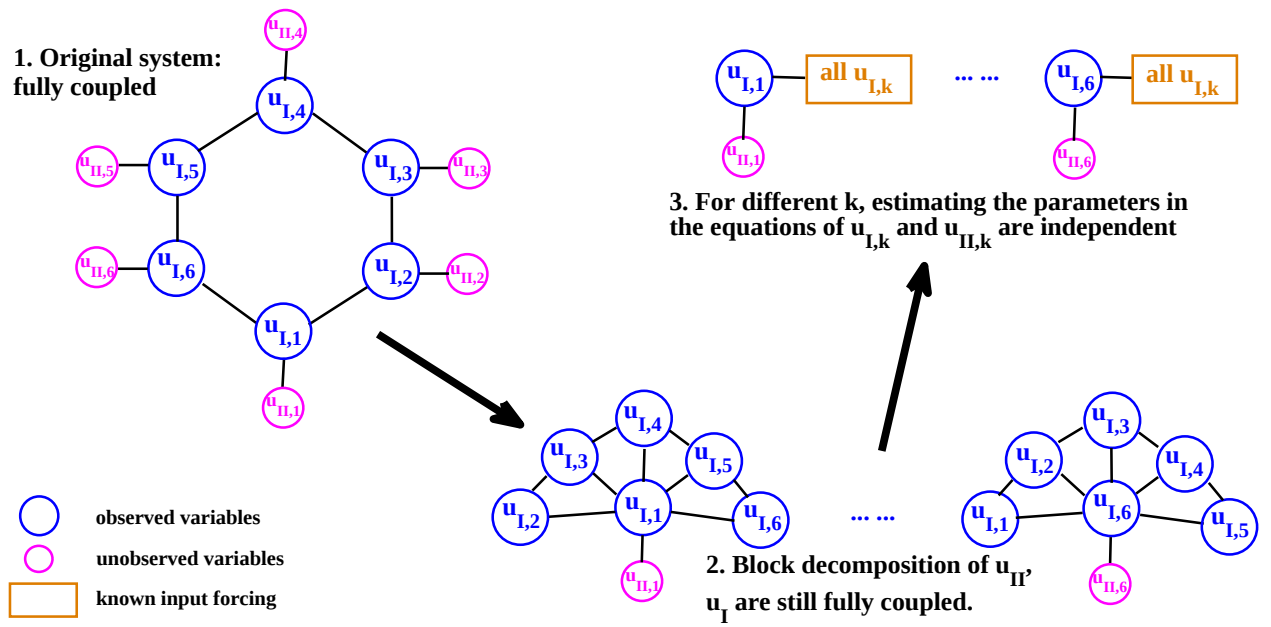


Figure 8: A schematic illustration of block decomposition in computing the conditional statistics and in computing the likelihood functions described in Section 6.1.

6.2. Statistical symmetry

In many applications, the underlying dynamical system represents a discrete approximation of some translation invariant PDEs in a periodic domain with nonlinear advection, diffusion and homogeneous external forcing [9, 2]. In such a scenario, the model usually have statistical symmetry property, namely

$$p(\mathbf{u}_{\mathbf{I},k}(t), \mathbf{u}_{\mathbf{II},k}(t)) = p(\mathbf{u}_{\mathbf{I},k'}(t), \mathbf{u}_{\mathbf{II},k'}(t)), \quad \text{for all } k \text{ and } k'. \quad (18)$$

namely, the statistical features for variables with different k are the same. The computational cost in the algorithms developed above can be further reduced if the coupled system (1) has statistical symmetry [44].

400 In our parameter estimation algorithm, if the statistical symmetry is satisfied, then all the parameters are constants. Therefore, we only need to deal with one block described in Section 6.1 to obtain the parameter values. Alternatively, we can estimate parameters using the information from a small number of blocks and then taking the average values to eliminate the sampling bias in the
405 presence of extremely short training data.

7. Test examples of nonlinear turbulent dynamical systems in large dimensions

Here, we apply the improved algorithm to a two-layer Lorenz 96 (L-96) model. The two-layer L96 model is a conceptual model in geophysical turbulence that is widely used as a testbed for data assimilation and parameterization in numerical weather forecasting [45, 43, 44]. 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. In the model presented here, large-scale motions are denoted by variables u_i , which are coupled to small-scale variables $v_{i,j}$ [44]:

$$\begin{aligned} du_i &= \left(u_{i-1}(u_{i+1} - u_{i-2}) + \gamma_i u_i v_i - \bar{d}_i u_i + F \right) dt + \sigma_u dW_{u_i}, \\ dv_i &= \left(-d_{v_i} v_i - \gamma_i u_i^2 \right) dt + \sigma_{v,i} dW_{v_i}, \end{aligned} \quad (19)$$

with $i = 1, \dots, I$ and periodic boundary conditions in u_i . One important feature of (19) is that the nonlinear interaction between u_i and v_i conserves energy, as
410 observed in nature. The two-layer L-96 model (19) belongs to the conditional Gaussian framework with $\mathbf{u_I} = \{u_i\}$ and $\mathbf{u_{II}} = \{v_i\}$. In (19), the parameters F and σ_u are usually assumed to be constants while the other parameters $\gamma_i, d_{v_i}, \sigma_{v_i}$ and $\bar{d}_i$ can vary as a function of i . If these four parameters take different values at different grid points, then the system is inhomogeneous. If
415 $\gamma_i \equiv \gamma, d_{v_i} \equiv d_v, \sigma_{v_i} \equiv \sigma_v$ and $\bar{d}_i \equiv \bar{d}$, then the system is homogeneous provided that the initial values of the system is also homogeneous.

Applying the block decomposition described in Section 6.1 to the two-layer L-96 model (19) results in I blocks. Each block i contains two equations, namely $\mathbf{u_{I},i} = u_i$ and $\mathbf{u_{II},i} = v_i$ with $i = 1, \dots, I$. Note that although a constant F
420 appears in the two-layer L-96 model (19) even in the inhomogeneous case, F is regarded as a function of i in the parameter estimation algorithm. In other words, F_i is used in the equation of u_i corresponding to the i -th block. The estimated values of F_i in different blocks are expected to be the same.

Below, we take $I = 40$ in (19) and the improved algorithm is applied to
425 both homogeneous and inhomogeneous cases. With $I = 40$, the total number
of state variables are 80. The total number of parameters in the homogeneous
and inhomogeneous cases is 6 and $4I + 2 = 162$, respectively.

Again, all the tests are conducted on a desktop with Ubuntu 18.04.1 LTS
system using MATLAB 2018a (academic version). The computational time of
430 each block $\{\mathbf{u}_{\mathbf{I},i}, \mathbf{u}_{\mathbf{II},i}\}$ is just a few seconds. All the $I = 40$ blocks (in the
inhomogeneous case) can be run in a parallel way.

7.1. Inhomogeneous case

First, we consider the inhomogeneous Lorenz 96 model. The following pa-
rameters are used as the true parameters:

$$\begin{aligned} \bar{d}_i &= -3 - \cos(2\pi i/I), & \gamma_i &= -0.8 - 0.2 \cos(2\pi i/I), & d_{v,i} &= 0.8 - 0.4 \cos(2\pi i/I), \\ \sigma_{v,i} &= 1 + 0.5 \cos(2\pi i/I), & F &= 8, & \sigma_u &= 0.2. \end{aligned} \tag{20}$$

Note that $F = 8$ in the L-96 model corresponds to a strongly chaotic regime
if the second layer is not considered, which distinguish itself from the weakly
435 chaotic regime ($F = 5$) and fully turbulent regime ($F = 16$).

The spatiotemporal patterns are shown in Figure 9, together with the time
series at three different locations $i = 1, 11$ and 21 . It is clear that the spatial
patterns have distinct structures in different locations. At the grid points where
 i is close to 1 or 40, the signal is almost quiescent. At the grid points where
440 i is close to 20, the signal is quite active. Significant wave trains are observed
between grids $i = 10$ and $i = 30$, which propagate westwards and mimic the
atmosphere Rossby waves. Note that the spatiotemporal patterns from $t = 5$ to
 $t = 100$ in Figure 9 are used as the training data in the parameter estimation
algorithm. This mimics most of the realistic situation that the training data is
445 limited and the training period is short.

We first estimate the diffusion coefficient σ_u by running a crude model using
a small number of iterations. Due to the continuous observations, the diffusion

coefficient σ_u can be estimated perfectly. Then in applying the improved algorithm discussed in Section 6.1 to estimate other parameters, we first assign initial guesses of the parameters,

$$\bar{d}_i^{\{0\}} = -2, \quad \gamma_i^{\{0\}} = 0, \quad d_{v,i}^{\{0\}} = 2, \quad \sigma_{v,i}^{\{0\}} = 2, \quad F^{\{0\}} = 5, \quad (21)$$

for all $i = 1, \dots, I$. These initial guesses actually represent a homogeneous flow, a decoupled system with $\gamma_i^{\{0\}} = 0$, and a weakly chaotic flow ($F^{\{0\}} = 5$). Therefore, large biases exist in the initial guesses. The proposal functions are given as follows

$$\begin{aligned} \bar{d}_i^{(k)} &= \bar{d}_i^{(k-1)} + 0.025X_1, & \gamma_i^{(k)} &= \gamma_i^{(k-1)} + 0.025X_2, \\ d_{v,i}^{(k)} &= d_{v,i}^{(k-1)} + 0.025X_3, & \sigma_{v,i}^{(k)} &= \sigma_{v,i}^{(k-1)} + 0.025X_4, \\ F_i^{(k)} &= F_i^{(k-1)} + 0.25X_5, \end{aligned} \quad (22)$$

where $X_1, \dots, X_5$ are random numbers sampled from independent standard Gaussian distributions. Note that as discussed at the beginning of this section, F is regarded as a function of i in the parameter estimation algorithm, even though the truth of F is independent of i . Figure 10 compares the true parameters (20), the initial guesses of the parameters (21) and the estimated parameters, where the estimated parameters are computed by averaging over the trace plot from $k = 10000$ to $k = 30000$. Figure 11 shows the trace plots and marginal posterior distributions at $i = 20$ (active phase) and $i = 35$ (quiescent phase). From Figure 10, it is clear that the parameter estimation algorithm is overall skillful. The parameters are estimated quite accurate in the active phases while some errors are observed in the quiescent phases. In addition, the parameter $\bar{d}_i$ seems to be the most difficult parameter to estimate. In fact, in the quiescent phase, the signal is weak, which implies the signal to noise ratio becomes small. Therefore, the parameters are hardly identified from the relatively noisy signals due to the lose of observability [9, 76, 77]. Among all the parameters, $\bar{d}_i$ suffers from this issue the most. In fact, the total damping of u_i is given by $\gamma_i v_i - \bar{d}_i$. In the quiescent phase, such as $i = 1$ shown in Figure 9, $\gamma_i v_i$ provides a strong damping, which then overwhelms the effect from $\bar{d}_i$.

This means the parameter $\bar{d}_i$ in the quiescent phase plays a very weak role in
 465 changing the dynamical behavior. In other words, the signal is quite robust
 with respect to the variation of $\bar{d}_i$ near its optimal values and the identification
 of the original $\bar{d}_i$ has a large uncertainty.

The above discussions indicate that even if some of the estimated parameters
 seem to be quite different from the true values, the model equipped with these
 470 estimated parameters are nevertheless able to recover both the dynamical and
 statistical features of the original model (with the true parameters). In Figure
 12, we show the model prediction with the estimated parameters. Comparing
 with the true signal in Figure 9, it is clear that the predicted spatiotemporal
 patterns are essentially the same as the truth in terms of the overall structures,
 475 amplitudes and intermittent features. Figure 13 further compares the model
 statistics generated from the model simulation using the true parameters and
 the estimated parameters based on the data shown in Figures 9 and 12. It
 is clear that the first four moments of u_i and the autocorrelation functions at
 different grid points of u_i associated with the two simulations are quite similar
 480 to each other. These facts confirm that the model with estimated parameters
 are able to generate the same dynamics as that with the perfect parameters.
 The errors in some parameters have little effect on capturing the dynamical and
 statistical features of the truth. In other words, the model is robust with respect
 to these parameters and there is no need to estimate these parameters precisely.
 485 Note that the errors in the estimated parameters very likely come from the
 sampling error in the finite (and short) training data as shown in Figures 9,
 which has only 95 time units. In many practical issues, the data in the training
 period is limited. Thus, it is unreasonable to expect a perfect estimation of all
 the parameters in the presence of such a sampling error, especially for those
 490 with respect to which the model is robust. Validating the model dynamical
 and statistical features is a reasonable strategy for quantifying the parameter
 estimation skill.

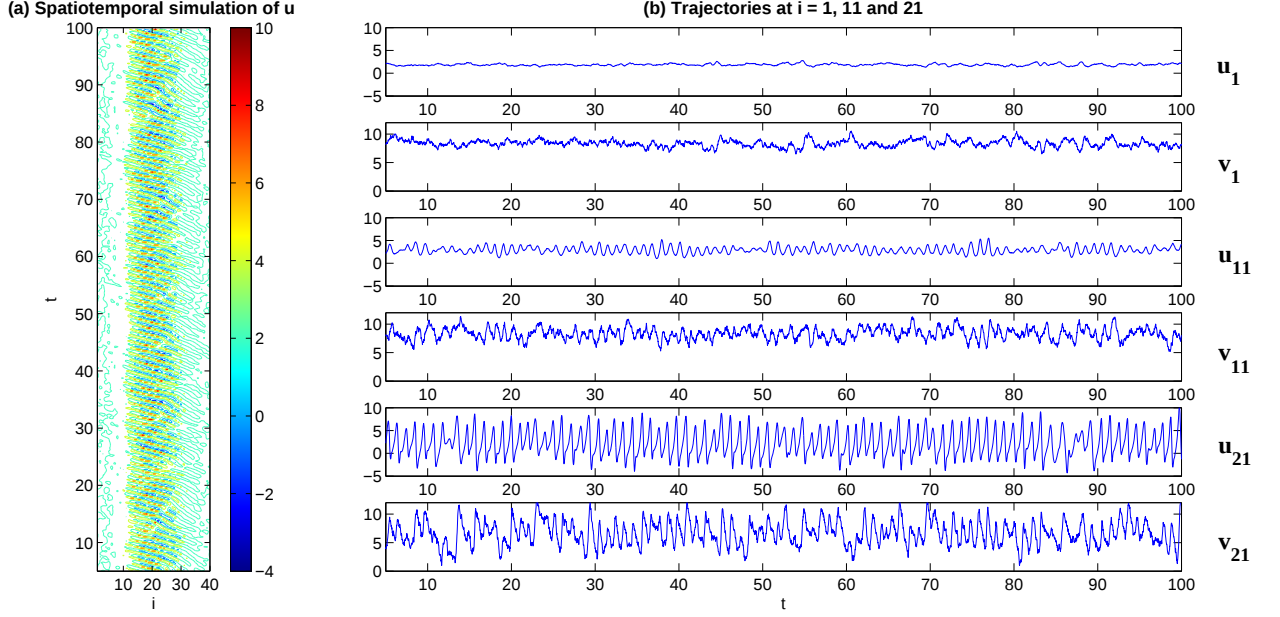


Figure 9: Spatiotemporal patterns (Panel (a)) and time series (Panel (b)) of the inhomogeneous two-layer Lorenz 96 model with parameters in (20).

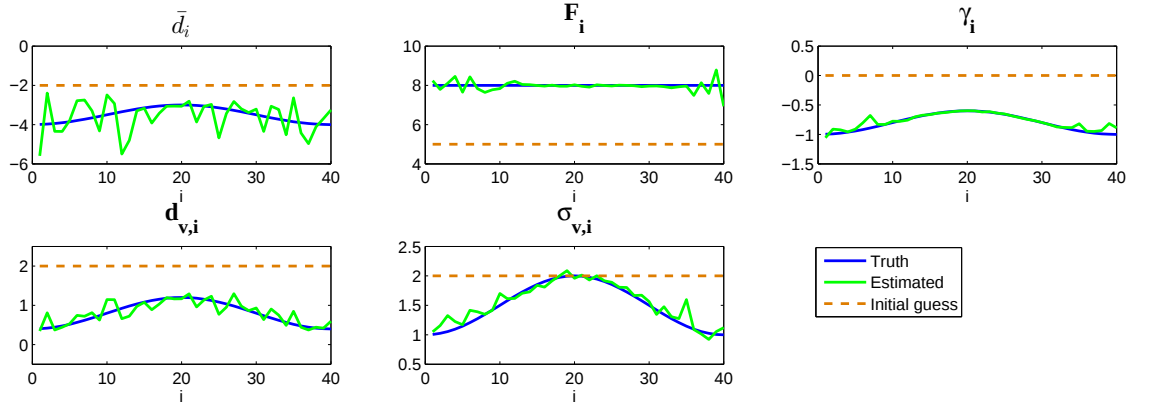


Figure 10: Comparison of the truth, the initial guess and the estimated parameters of the inhomogeneous two-layer Lorenz 96 model. The horizontal axis denotes the spatial locations. The estimated parameters are computed by averaging the trace plots from $k = 10000$ to $k = 30000$.

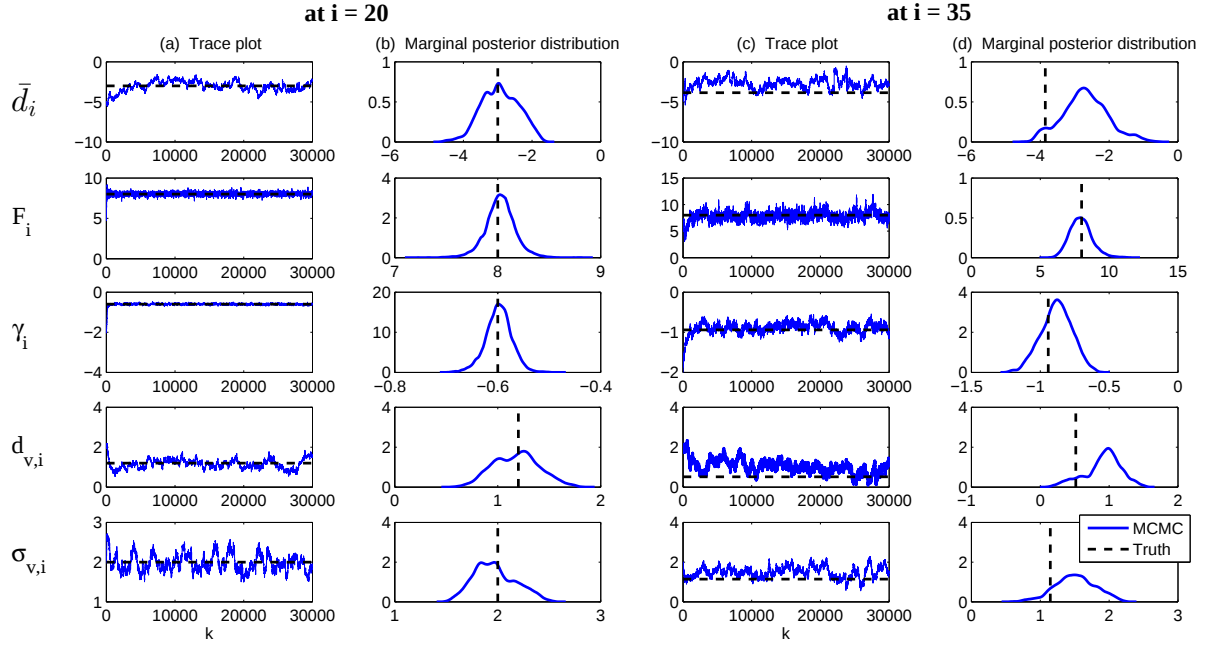


Figure 11: The trace plots and marginal posterior distributions at $i = 20$ (active phase; Panels (a)–(b)) and $i = 35$ (quiescent phase; Panels (c)–(d)) of the inhomogeneous two-layer Lorenz 96 model. The marginal posterior distributions are formed by collecting the points in the corresponding trace plot from $k = 10000$ to $k = 30000$.

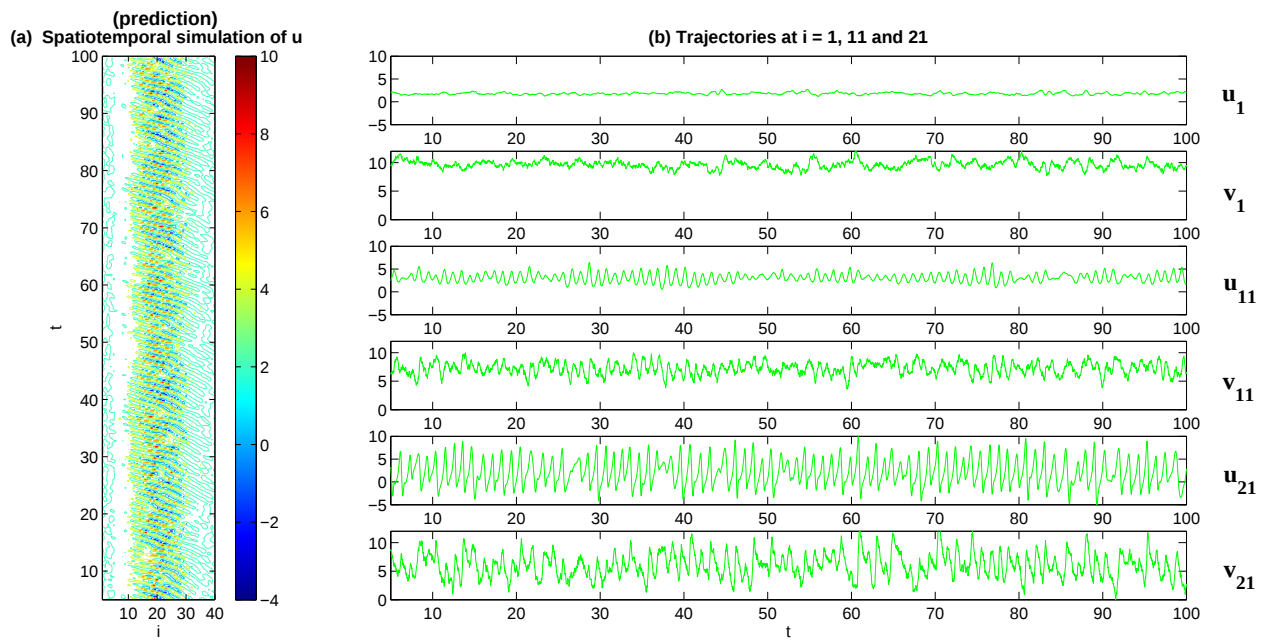


Figure 12: The predicted spatiotemporal patterns (Panel (a)) and time series (Panel (b)) of the inhomogeneous two-layer Lorenz 96 model with the estimated parameters from Figure 10.

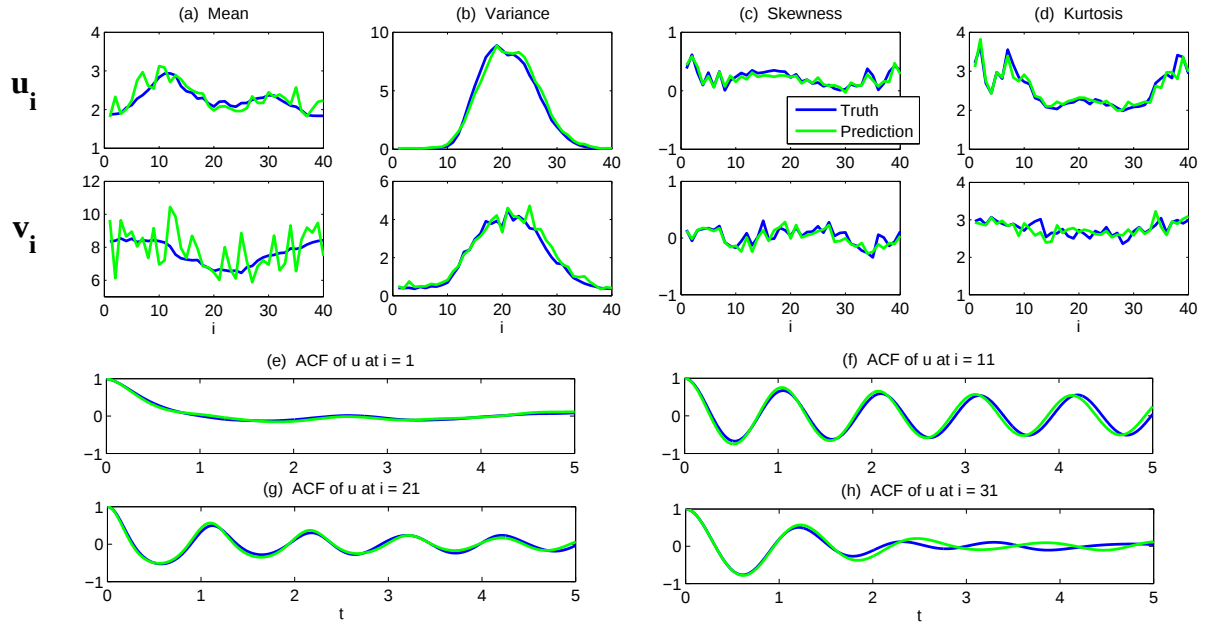


Figure 13: Comparison of the first four moments (Panels (a)–(d)) and the autocorrelation functions (Panels (e)–(h)) of the inhomogeneous L-96 model with the true parameters and the estimated parameters.

7.2. Homogeneous case

Finally, we apply the parameter estimation algorithm involving both the statistical symmetry and the block decomposition to the homogeneous two-layer L-96 model (19). The following parameters are used to generate the true signal,

$$\bar{d} = -3.5, \quad \gamma = -0.8, \quad d_v = 0.8, \quad \sigma_v = 1.5, \quad F = 8, \quad \sigma_u = 0.2. \quad (23)$$

These parameters result in a unique spatiotemporal structure as shown in Figure 14. In fact, as in the inhomogeneous case (Figure 9, wave trains propagate westwards are clearly seen in Figure 14. In addition to these individual wave trains, their envelopes actually propagate eastwards. This behavior in a different model mimics the realistic features of the Madden-Julian Oscillation which is the dominant intraseasonal variability in the equatorial domain [78, 79]. It is also clear that the statistical features at different grid points are the same and therefore the model is homogeneous and satisfies the statistical symmetry.

The same proposal functions as in (22) is used in this homogeneous situation. The initial guess of the parameters are

$$\begin{aligned} \bar{d}_i^{\{0\}} &= -3.5 - 0.5 \cos(2\pi i/I), & \gamma_i^{\{0\}} &= -0.8 - 0.2 \cos(2\pi i/I), \\ d_{v,i}^{\{0\}} &= 0.8 - 0.4 \cos(2\pi i/I), & \sigma_{v,i}^{\{0\}} &= 1.5 - 0.5 \cos(2\pi i/I), & F^{\{0\}} &= 5, \end{aligned} \quad (24)$$

where the forcing F is chosen such that the initial state is in a weakly chaotic regime while the initial guesses of all the other parameters represent an inhomogeneous flow, which is far from the truth.

Figure 15 shows the trace plots and the marginal posterior distributions of the estimated parameters, where the estimation is based on solving the likelihood function only on the first grid point. These parameters are overall estimated quite accurately. The only exception is the damping $\bar{d}$, which takes quite a few iteration steps to converge around the truth. Nevertheless, as in the inhomogeneous case, the model is robust with respect to this damping parameter around its optimal value. In fact, we have tested the model even with

$\bar{d} = -2$, and the model simulation is still able to capture the key dynamical and statistical features in the truth. Therefore, the estimated parameters can be used in the original system for understanding and predicting the complex
515 signal from nature.

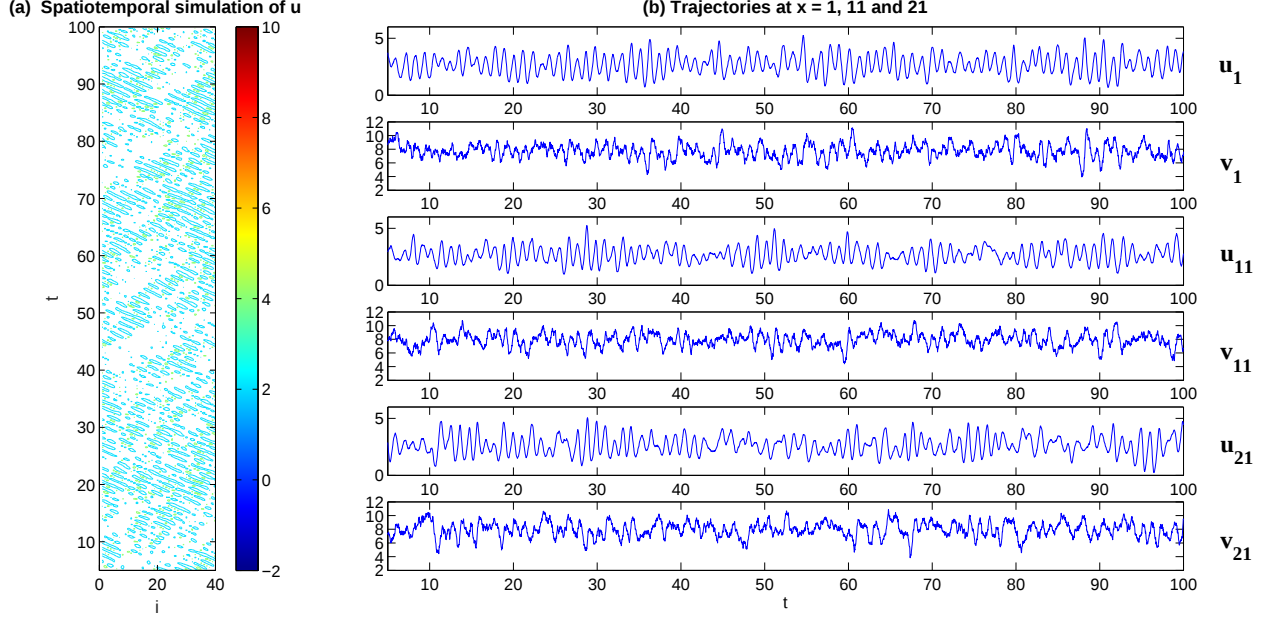


Figure 14: Spatiotemporal patterns (Panel (a)) and time series (Panels (b)) of the homogeneous two-layer Lorenz 96 model (19) with parameters in (23).

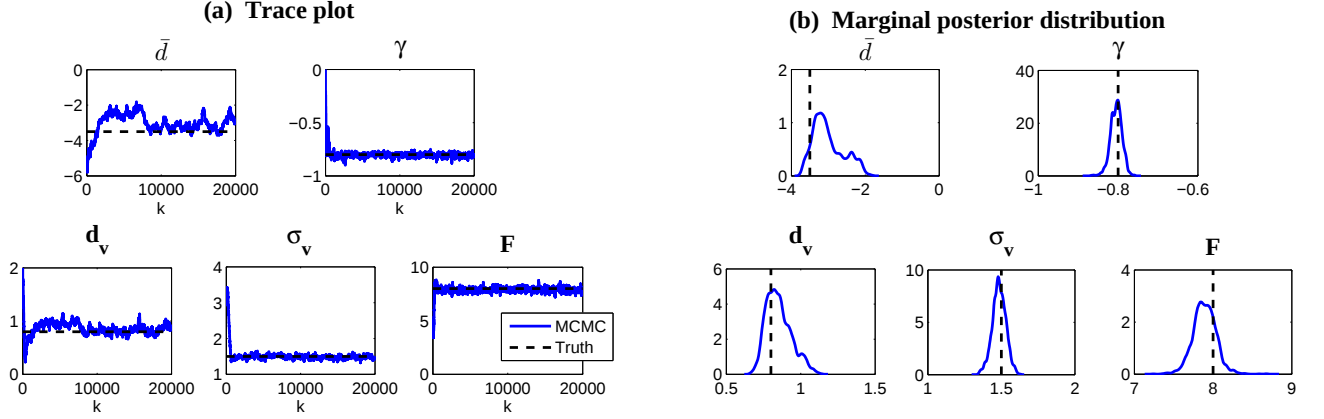


Figure 15: Trace plots (Panel (a)) and marginal posterior distributions (Panel (b)) of the parameter estimation of the two-layer Lorenz 96 model with parameters in (23). The marginal posterior distributions are formed by collecting the points in the corresponding trace plot from $k = 5000$ to $k = 20000$.

8. Conclusion

In this article, a new efficient parameter estimation algorithm is developed, which applies to the conditional Gaussian nonlinear systems in large dimensions with only partial observations of $\mathbf{u}_I$. Note that despite the conditional Gaussian-
520 ity, the coupled systems remain highly nonlinear and is able to capture strong non-Gaussian features such as skewed or fat-tailed distributions as observed in nature.

The algorithm exploits the analytically solvable conditional statistics in the conditional Gaussian framework to recover the trajectories associated with the
525 unobserved variables $\mathbf{u}_{II}$ (Section 4). This deterministic and optimal method in recovering the unobserved trajectories circumvents the expensive and time-consuming random sampling of the trajectories associated with the unobserved variables in the infinite dimensional space using the traditional data augmentation approaches and thus greatly enhances the computational efficiency. With
530 these recovered unobserved trajectories, the MCMC technique can then be applied to sample the parameters. Notably, only a short training period is sufficient in this new parameter estimation algorithm and it is therefore practically useful. The computational cost also increases only linearly as a function of the length of the training period. Test examples based on nonlinear and non-Gaussian low-
535 order models with intermittency and extreme events show both the parameter estimation skill and the computational efficiency of the new algorithm (Section 5).

In order to estimate the parameters in high-dimensional systems with a large number of parameters, two effective strategies are further developed and
540 incorporated into the above parameter estimation algorithm (Section 6). The first strategy involves a judicious block decomposition of the state variables such that the original problem is divided into several subproblems, which allow an extremely efficient parallel computation for the parameter estimation. The second strategy exploits statistical symmetry for a further reduction of the
545 computational cost when the system is statistically homogeneous. A two-layer

Lorenz 96 model (with 80 state variables and 162 parameters) that mimics the realistic features of atmosphere wave propagations is adopted to illustrate the skillful behavior of such an improved parameter estimation algorithm using only a short training period (Section 7).

550 Since this article aims at emphasizing the new approach of exploiting the conditional statistics to recover the unobserved trajectories, only the basic MCMC method is adopted in the current version of the algorithm. Incorporating adaptive MCMC techniques or non-symmetric proposal functions are natural extensions to further improve the current parameter estimation framework. Se-
 555 quential methods in recovering the unobserved trajectories and computing the likelihood function can also be included into the framework. In addition, the closed analytical formulae of the conditional Gaussian statistics and the well-established theories of various MCMC techniques also allow using rigorous mathematical analysis to study the accuracy and the convergence of this parameter
 560 estimation framework.

Note that all the numerical tests shown in this article are perfect model twin experiments. Yet, most parameter estimation problems in practice contain model errors. In addition, due to the measurement and representation errors [13], the observations may also be imperfect and noisy. Thus, estimating pa-
 565 rameters in various imperfect models is of practical importance. This parameter estimation framework allows using imperfect models for parameter estimation since the only information needed from the truth is the observational signals. Therefore, a future work is to understand the effect of model error and representation error in affecting the parameter estimation skill, which also has potentials
 570 in providing guidelines for developing reasonable imperfect models in practice.

The parameter estimation framework developed here has a wide application. In fact, many practical approaches for filtering and predicting complex turbulent systems involve hybrid strategies in developing the forecast models, where the conditional Gaussian processes are used in modeling small-scale features
 575 via stochastic parameterizations, superparameterization and dynamical super-resolution etc. Examples in real-world applications include filtering sparsely

observed geophysical flows [73], improving the filtering and prediction skill of atmosphere and ocean flows using stochastic superparameterizations [80, 81] and multiscale data assimilation using hybrid models [82, 83]. The parameter
580 estimation framework developed here can be applied to these problems for model calibration and improve the understanding and prediction of these complex dynamical systems.

Acknowledgement

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

Appendix A. Calculating the diffusion coefficient in the observed process via its quadratic variation.

590

Here, we summarize the results on calculating of the diffusion coefficient in the observed process via its quadratic variation [84]. The quadratic variation of a stochastic integral process $[H \cdot Y]$ with $(H \cdot Y)_t = \int_0^t H_s Y_s$ is defined as

$$[H \cdot Y]_t = \int_0^t H_s^2 d[Y]_s. \quad (\text{A.1})$$

Proposition Appendix A.1. *Consider a general stochastic differential equation (SDE)*

$$du = \mu(t, u)dt + \sigma(t, u)dW_t. \quad (\text{A.2})$$

Its quadratic variation is given by

$$[u]_t = \int_0^t \sigma^2(s, u_s)ds. \quad (\text{A.3})$$

Proof. The equivalent integral form of the general SDE in (A.2) is given by,

$$u = u_0 + \int_0^t \mu(s, u)ds + \int_0^t \sigma(s, u)dW_s. \quad (\text{A.4})$$

Using the definition in (A.1), we have

$$[u]_t = [\mu \cdot t]_t + [\sigma \cdot W]_t = \int_0^t \mu_s^2 d[s]_s + \int_0^t \sigma_s^2 d[W]_s = \int_0^s \sigma^2(s, u_s) ds. \quad (\text{A.5})$$

where $[t]_t = 0$ since all continuous processes of finite variation have zero quadratic variation, and $[W]_t = t$. $\square$

When $\sigma(s, u) \equiv \sigma_u$ is a constant, Proposition (A.1) provides a simple way to compute the diffusion coefficient σ , which is

$$[u]_t = t\sigma^2, \quad (\text{A.6})$$

or equivalently

$$\sigma = \sqrt{\frac{[u]_t}{t}}. \quad (\text{A.7})$$

References

- [1] A. J. Majda, Introduction to turbulent dynamical systems in complex systems, Springer, 2016.
- [2] A. Majda, X. Wang, Nonlinear dynamics and statistical theories for basic geophysical flows, Cambridge University Press, 2006.
- [3] S. H. Strogatz, Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering, CRC Press, 2018.
- [4] D. Baleanu, J. A. T. Machado, A. C. Luo, Fractional dynamics and control, Springer Science & Business Media, 2011.
- [5] T. Deisboeck, J. Y. Kresh, Complex systems science in biomedicine, Springer Science & Business Media, 2007.
- [6] J. Stelling, A. Kremling, M. Ginkel, K. Bettenbrock, E. Gilles, of Book: Foundations of systems biology, MIT press, 2001.
- [7] S. A. Sheard, A. Mostashari, Principles of complex systems for systems engineering, Systems Engineering 12 (4) (2009) 295–311.

- [8] D. C. Wilcox, Multiscale model for turbulent flows, *AIAA journal* 26 (11) (1988) 1311–1320.
- 610 [9] A. J. Majda, J. Harlim, Filtering complex turbulent systems, Cambridge University Press, 2012.
- [10] A. Majda, N. Chen, Model error, information barriers, state estimation and prediction in complex multiscale systems, *Entropy* 20 (9) (2018) 644.
- 615 [11] J. S. Whitaker, G. P. Compo, J.-N. Thépaut, A comparison of variational and ensemble-based data assimilation systems for reanalysis of sparse observations, *Monthly Weather Review* 137 (6) (2009) 1991–1999.
- 620 [12] S. Kouketsu, T. Kawano, S. Masuda, N. Sugiura, Y. Sasaki, T. Toyoda, H. Igarashi, Y. Kawai, K. Katsumata, H. Uchida, et al., Deep ocean heat content changes estimated from observation and reanalysis product and their influence on sea level change, *Journal of Geophysical Research: Oceans* 116 (C3).
- [13] G. Evensen, Data assimilation: the ensemble Kalman filter, Springer Science & Business Media, 2009.
- 625 [14] M. Rodriguez-Fernandez, P. Mendes, J. R. Banga, A hybrid approach for efficient and robust parameter estimation in biochemical pathways, *Biosystems* 83 (2-3) (2006) 248–265.
- [15] K. Schittkowski, Numerical data fitting in dynamical systems: a practical introduction with applications and software, Vol. 77, Springer Science & Business Media, 2013.
- 630 [16] D. P. Dee, On-line estimation of error covariance parameters for atmospheric data assimilation, *Monthly weather review* 123 (4) (1995) 1128–1145.
- [17] O. M. Smedstad, J. J. O’Brien, Variational data assimilation and parameter estimation in an equatorial pacific ocean model, *Progress in Oceanography* 26 (2) (1991) 179–241.

- 635 [18] R. Van Der Merwe, E. A. Wan, The square-root unscented kalman filter for state and parameter-estimation, in: Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on, Vol. 6, IEEE, 2001, pp. 3461–3464.
- [19] T. A. Wenzel, K. Burnham, M. Blundell, R. Williams, Dual extended
640 kalman filter for vehicle state and parameter estimation, *Vehicle System Dynamics* 44 (2) (2006) 153–171.
- [20] G. L. Plett, Extended kalman filtering for battery management systems of lipb-based hev battery packs: Part 3. state and parameter estimation, *Journal of Power sources* 134 (2) (2004) 277–292.
- 645 [21] C. Andrieu, N. De Freitas, A. Doucet, M. I. Jordan, An introduction to mcmc for machine learning, *Machine learning* 50 (1-2) (2003) 5–43.
- [22] W. H. Press, S. A. Teukolsky, B. P. Flannery, W. T. Vetterling, *Numerical Recipes with Source Code CD-ROM 3rd Edition: The Art of Scientific Computing*, Cambridge University Press, 2007.
- 650 [23] M. Richey, The evolution of markov chain monte carlo methods, *The American Mathematical Monthly* 117 (5) (2010) 383–413.
- [24] H. Haario, M. Laine, A. Mira, E. Saksman, Dram: efficient adaptive mcmc, *Statistics and computing* 16 (4) (2006) 339–354.
- [25] M. A. Tanner, W. H. Wong, The calculation of posterior distributions by
655 data augmentation, *Journal of the American statistical Association* 82 (398) (1987) 528–540.
- [26] G. C. Wei, M. A. Tanner, A monte carlo implementation of the em algorithm and the poor man’s data augmentation algorithms, *Journal of the American statistical Association* 85 (411) (1990) 699–704.
- 660 [27] N. Chen, A. J. Majda, Filtering nonlinear turbulent dynamical systems through conditional gaussian statistics, *Monthly Weather Review* 144 (12) (2016) 4885–4917.

- [28] N. Chen, A. Majda, Conditional gaussian systems for multiscale nonlinear stochastic systems: Prediction, state estimation and uncertainty quantification, *Entropy* 20 (2018) 509.
- [29] R. S. Liptser, A. N. Shiryaev, *Statistics of random processes ii: Applications*, *Appl. Math* 6.
- [30] N. Chen, A. J. Majda, D. Giannakis, Predicting the cloud patterns of the madden-julian oscillation through a low-order nonlinear stochastic model, *Geophysical Research Letters* 41 (15) (2014) 5612–5619.
- [31] N. Chen, A. J. Majda, Predicting the real-time multivariate madden–julian oscillation index through a low-order nonlinear stochastic model, *Monthly Weather Review* 143 (6) (2015) 2148–2169.
- [32] N. Chen, A. J. Majda, Predicting the cloud patterns for the boreal summer intraseasonal oscillation through a low-order stochastic model, *Mathematics of Climate and Weather Forecasting* 1 (1) (2015) 1–20.
- [33] N. Chen, A. J. Majda, Filtering the stochastic skeleton model for the madden–julian oscillation, *Monthly Weather Review* 144 (2) (2016) 501–527.
- [34] N. Chen, A. J. Majda, X. T. Tong, Information barriers for noisy lagrangian tracers in filtering random incompressible flows, *Nonlinearity* 27 (9) (2014) 2133.
- [35] N. Chen, A. J. Majda, X. T. Tong, Noisy lagrangian tracers for filtering random rotating compressible flows, *Journal of Nonlinear Science* 25 (3) (2015) 451–488.
- [36] N. Chen, A. J. Majda, Model error in filtering random compressible flows utilizing noisy lagrangian tracers, *Monthly Weather Review* 144 (11) (2016) 4037–4061.

- [37] M. Branicki, A. J. Majda, Dynamic stochastic superresolution of sparsely
690 observed turbulent systems, *Journal of Computational Physics* 241 (2013)
333–363.
- [38] S. R. Keating, A. J. Majda, K. S. Smith, New methods for estimating ocean
eddy heat transport using satellite altimetry, *Monthly Weather Review*
140 (5) (2012) 1703–1722.
- [39] A. J. Majda, I. Grooms, New perspectives on superparameterization for
695 geophysical turbulence, *Journal of Computational Physics* 271 (2014) 60–
77.
- [40] A. J. Majda, J. Harlim, Physics constrained nonlinear regression models
for time series, *Nonlinearity* 26 (1) (2012) 201.
- [41] J. Harlim, A. Mahdi, A. J. Majda, An ensemble kalman filter for statistical
700 estimation of physics constrained nonlinear regression models, *Journal of*
Computational Physics 257 (2014) 782–812.
- [42] A. J. Majda, D. Qi, T. P. Sapsis, Blended particle filters for large-
dimensional chaotic dynamical systems, *Proceedings of the National A-*
705 *cademy of Sciences* (2014) 201405675.
- [43] D. S. Wilks, Effects of stochastic parametrizations in the lorenz’96 system,
Quarterly Journal of the Royal Meteorological Society 131 (606) (2005)
389–407.
- [44] N. Chen, A. J. Majda, Beating the curse of dimension with accurate statis-
710 tics for the fokker–planck equation in complex turbulent systems, *Proceed-*
ings of the National Academy of Sciences 114 (49) (2017) 12864–12869.
- [45] H. Arnold, I. Moroz, T. Palmer, Stochastic parametrizations and model
uncertainty in the lorenz96 system, *Phil. Trans. R. Soc. A* 371 (1991) (2013)
20110479.

- 715 [46] R. E. Kalman, R. S. Bucy, New results in linear filtering and prediction theory, *Journal of basic engineering* 83 (1) (1961) 95–108.
- [47] K. Brammer, G. Siffling, *Kalman-bucy filters*, Artech House on Demand, 1989.
- [48] R. S. Bucy, P. D. Joseph, *Filtering for stochastic processes with applications to guidance*, Vol. 326, American Mathematical Soc., 1987.
- 720 [49] A. H. Jazwinski, *Stochastic processes and filtering theory*, Courier Corporation, 2007.
- [50] H. Haario, E. Saksman, J. Tamminen, et al., An adaptive metropolis algorithm, *Bernoulli* 7 (2) (2001) 223–242.
- 725 [51] S. Chib, F. Nardari, N. Shephard, Analysis of high dimensional multivariate stochastic volatility models, *Journal of Econometrics* 134 (2) (2006) 341–371.
- [52] G. B. Durham, A. R. Gallant, Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes, *Journal of Business & Economic Statistics* 20 (3) (2002) 297–338.
- 730 [53] N. Chen, D. Giannakis, R. Herbei, A. J. Majda, An mcmc algorithm for parameter estimation in signals with hidden intermittent instability, *SIAM/ASA Journal on Uncertainty Quantification* 2 (1) (2014) 647–669.
- [54] A. Golightly, D. J. Wilkinson, Bayesian inference for nonlinear multivariate diffusion models observed with error, *Computational Statistics & Data Analysis* 52 (3) (2008) 1674–1693.
- 735 [55] M. K. Pitt, N. Shephard, Filtering via simulation: Auxiliary particle filters, *Journal of the American statistical association* 94 (446) (1999) 590–599.
- [56] A. Doucet, S. Godsill, C. Andrieu, On sequential monte carlo sampling methods for bayesian filtering, *Statistics and computing* 10 (3) (2000) 197–208.
- 740

- [57] P. Del Moral, A. Doucet, A. Jasra, Sequential monte carlo samplers, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68 (3) (2006) 411–436.
- 745 [58] E. N. Lorenz, Deterministic nonperiodic flow, *Journal of the atmospheric sciences* 20 (2) (1963) 130–141.
- [59] C. Sparrow, *The Lorenz equations: bifurcations, chaos, and strange attractors*, Vol. 41, Springer Science & Business Media, 2012.
- [60] H. Haken, Analogy between higher instabilities in fluids and lasers, *Physics Letters A* 53 (1) (1975) 77–78.
- 750 [61] E. Knobloch, Chaos in the segmented disc dynamo, *Physics Letters A* 82 (9) (1981) 439–440.
- [62] M. Gorman, P. Widmann, K. Robbins, Nonlinear dynamics of a convection loop: a quantitative comparison of experiment with theory, *Physica D: Nonlinear Phenomena* 19 (2) (1986) 255–267.
- 755 [63] N. Hemati, Strange attractors in brushless dc motors, *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications* 41 (1) (1994) 40–45.
- [64] K. M. Cuomo, A. V. Oppenheim, Circuit implementation of synchronized chaos with applications to communications, *Physical review letters* 71 (1) (1993) 65.
- 760 [65] D. Poland, Cooperative catalysis and chemical chaos: a chemical model for the lorenz equations, *Physica D: Nonlinear Phenomena* 65 (1-2) (1993) 86–99.
- 765 [66] S. I. Tzenov, Strange attractors characterizing the osmotic instability, *arXiv preprint arXiv:1406.0979*.

- [67] B. Gershgorin, J. Harlim, A. J. Majda, Test models for improving filtering with model errors through stochastic parameter estimation, *Journal of Computational Physics* 229 (1) (2010) 1–31.
- 770 [68] B. Gershgorin, J. Harlim, A. J. Majda, Improving filtering and prediction of spatially extended turbulent systems with model errors through stochastic parameter estimation, *Journal of Computational Physics* 229 (1) (2010) 32–57.
- [69] A. J. Majda, M. Branicki, Lessons in uncertainty quantification for turbulent dynamical systems, *Discrete & Continuous Dynamical Systems-A* 32 (9) (2012) 3133–3221.
- 775 [70] W. Lee, A. Stuart, Derivation and analysis of simplified filters, *Communications in Mathematical Sciences* 15 (2) (2017) 413–450.
- [71] M. Branicki, B. Gershgorin, A. J. Majda, Filtering skill for turbulent signals for a suite of nonlinear and linear extended kalman filters, *Journal of Computational Physics* 231 (4) (2012) 1462–1498.
- 780 [72] M. Branicki, N. Chen, A. J. Majda, Non-gaussian test models for prediction and state estimation with model errors, *Chinese Annals of Mathematics, Series B* 34 (1) (2013) 29–64.
- [73] J. Harlim, A. J. Majda, Filtering turbulent sparsely observed geophysical flows, *Monthly Weather Review* 138 (4) (2010) 1050–1083.
- 785 [74] A. J. Majda, J. Harlim, B. Gershgorin, Mathematical strategies for filtering turbulent dynamical systems, *Discrete Contin. Dyn. Syst* 27 (2) (2010) 441–486.
- [75] M. Branicki, A. J. Majda, K. Law, Accuracy of some approximate gaussian filters for the navier-stokes equation in the presence of model error, *Multiscale Modeling and Simulation* Submitted.
- 790

- [76] R. Kalman, On the general theory of control systems, IRE Transactions on Automatic Control 4 (3) (1959) 110–110.
- 795 [77] E. D. Sontag, Mathematical control theory: deterministic finite dimensional systems, Vol. 6, Springer Science & Business Media, 2013.
- [78] W. K.-M. Lau, D. E. Waliser, Intraseasonal variability in the atmosphere-ocean climate system, Springer Science & Business Media, 2011.
- [79] C. Zhang, Madden-julian oscillation, Reviews of Geophysics 43 (2).
- 800 [80] J. Harlim, A. J. Majda, Test models for filtering with superparameterization, Multiscale Modeling & Simulation 11 (1) (2013) 282–308.
- [81] I. Grooms, A. J. Majda, K. S. Smith, Stochastic superparameterization in a quasigeostrophic model of the antarctic circumpolar current, Ocean Modelling 85 (2015) 1–15.
- 805 [82] L. Slivinski, E. Spiller, A. Apte, B. Sandstede, A hybrid particle–ensemble kalman filter for lagrangian data assimilation, Monthly Weather Review 143 (1) (2015) 195–211.
- [83] Y. Lee, A. J. Majda, Multiscale methods for data assimilation in turbulent systems, Multiscale Modeling & Simulation 13 (2) (2015) 691–713.
- 810 [84] L. C. G. Rogers, D. Williams, Diffusions, Markov processes and martingales: Volume 2, Itô calculus, Vol. 2, Cambridge university press, 1994.