Beating the curse of dimension with accurate statistics for the Fokker–Planck equation in complex turbulent systems

Nan Chen*a,b,1 and Andrew J. Majda*a,b,c

*aDepartment of Mathematics, Courant Institute of Mathematical Sciences, New York University, New York, NY 10012; bCenter for Atmosphere Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, NY 10012; and cCenter for Prototype Climate Modeling, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

Contributed by Andrew J. Majda, October 13, 2017 (sent for review September 28, 2017; reviewed by Joseph Tribbia and Xiaoming Wang)

Solving the Fokker–Planck equation for high-dimensional complex dynamical systems is an important issue. Recently, the authors developed efficient statistically accurate algorithms for solving the Fokker–Planck equations associated with high-dimensional nonlinear turbulent dynamical systems with conditional Gaussian structures, which contain many strong non-Gaussian features such as intermittency and fat-tailed probability density functions (PDFs). The algorithms involve a hybrid strategy with a small number of samples $L$, where a conditional Gaussian mixture in a high-dimensional subspace via an extremely efficient parametric method is combined with a judicious Gaussian kernel density estimation in the remaining low-dimensional subspace. In this article, two effective strategies are developed and incorporated into these algorithms. The first strategy involves a judicious block decomposition of the conditional covariance matrix such that the evolutions of different blocks have no interactions, which allows an extremely efficient parallel computation due to the small size of each individual block. The second strategy exploits statistical symmetry for a further reduction of $L$. The resulting algorithms can efficiently solve the Fokker–Planck equation with strongly non-Gaussian PDFs in much higher dimensions even with orders in the millions and thus beat the curse of dimension. The algorithms are applied to a 1,000-dimensional stochastic coupled FitzHugh–Nagumo model for excitable media. An accurate recovery of both the transient and equilibrium non-Gaussian PDFs requires only $L = 1$ samples! In addition, the block decomposition facilitates the algorithms to efficiently capture the distinct non-Gaussian features at different locations in a 240-dimensional two-layer inhomogeneous Lorenz 96 model, using only $L = 500$ samples.

high-dimensional non-Gaussian PDFs | hybrid strategy | block decomposition | statistical symmetry | small sample size

T he 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 (1, 2). For a general nonlinear dynamical system,

$$du = F(u, t)dt + \Sigma(u, t)dW,$$  

[1]

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

$$\frac{\partial}{\partial t} p(u, t) = -\nabla_u (F(u, t)p(u, t)) + \frac{1}{2} \nabla_u \nabla_u (Q(u, t)p(u, t)), \quad p|_{t=t_0} = p_0(u),$$  

[2]

with $Q = \Sigma \Sigma^T$. In many complex dynamical systems, such as geophysical and engineering turbulence, neuroscience, and excitable media, the solution of the Fokker–Planck equation in Eq. 2 involves strong non-Gaussian features with intermittency and extreme events (3–5). In addition, the dimension of $u$ in these complex systems is typically very large, representing a variety of variability in different temporal and spatial scales (3, 6). Therefore, solving the high-dimensional Fokker–Planck equation for both the steady-state and transient phases with non-Gaussian features is an important issue. However, traditional numerical methods such as finite element and finite difference as well as the direct Monte Carlo simulations of Eq. 1 all suffer from the curse of dimension (7, 8).

Recently, the authors developed efficient statistically accurate algorithms for solving the Fokker–Planck equation associated with high-dimensional nonlinear turbulent dynamical systems with conditional Gaussian structures (9). These conditional Gaussian nonlinear dynamical systems capture many strong non-Gaussian features such as intermittency and fat-tailed PDFs (10). Applications of the conditional Gaussian framework include modeling and predicting the highly intermittent time series of the Madden–Julian oscillation and monsoon (11–13), state estimation of the turbulent ocean flows from noisy Lagrangian tracers (14–16), dynamic stochastic superresolution of sparsely observed turbulent systems (17), and stochastic superparameterization for geophysical turbulent flows (18), etc. The efficient statistically accurate algorithms in ref. 9 involve a hybrid strategy that requires only a small number of samples. In these algorithms, a conditional Gaussian mixture in the high-dimensional subspace of $u_0$ via an extremely efficient parametric method is combined with a judicious Gaussian kernel density estimation in the remaining low-dimensional subspace of $u_1$. Particularly,

Significance

Solving the Fokker–Planck equation for high-dimensional complex dynamical systems is an important issue. Effective strategies are developed and incorporated into efficient statistically accurate algorithms for solving the Fokker–Planck equations associated with a rich class of high-dimensional nonlinear turbulent dynamical systems with strong non-Gaussian features. These effective strategies exploit a judicious block decomposition of high-dimensional conditional covariance matrices and statistical symmetry to facilitate an extremely efficient parallel computation and a significant reduction of sample numbers. The resulting algorithms can efficiently solve the Fokker–Planck equation in much higher dimensions even with orders in the millions and thus beat the curse of dimension. Skillful behavior of the algorithms is illustrated for highly non-Gaussian systems in excitable media and geophysical turbulence.

Author contributions: N.C. and A.J.M. designed research, performed research, and wrote the paper.

Reviewers: J.T., National Center for Atmospheric Research (NCAR); and X.W., Florida State University.

The authors declare no conflict of interest.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

1To whom correspondence may be addressed. Email: jonjon@cims.nyu.edu or chennan@cims.nyu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1717071114/DCSupplemental.
the parametric method provides closed analytical formulas for determining the conditional Gaussian distributions in the high-dimensional subspace of $u_0$ and is therefore computationally efficient and accurate. It has been shown in a stringent set of numerical tests (9) that with an order of $L \sim O(100)$ in $x$, the mixture distribution has significant skill in capturing both the statistically steady state and the transient behavior with fat tails of non-Gaussian PDFs in up to six dimensions. Rigorous analysis (19) indicates that $L$ does not increase exponentially as the dimension of the high-dimensional subspace of $u_0$ to maintain a given level of accuracy, which is fundamentally different from Monte Carlo methods.

In this article, two effective strategies are developed and incorporated into the algorithms in ref. 9 (hereafter, basic algorithms) that enable the expanded algorithms to efficiently solve the Fokker–Planck equation in much higher dimensions, even with orders in the millions. In fact, the major computational cost in the basic algorithms for systems with a large dimension of state variables $u_0$ comes from solving the time evolution of the conditional covariance. To overcome this difficulty, an effective strategy involving a decomposition of the state variables into different groups $u = \bigcup_{k} u_k$, with each $u_k \in \mathbb{R}^{N_k}$ is developed. Then the high-dimensional conditional covariance matrix of $u_0$ conditioned on $u_k$ becomes a block diagonal matrix under the conditions that the nonlinear terms on the right-hand side of $u_k$ in Eq. 1 consist of any nonlinear interactions between different components of $u_0$ but only those between $u_{0,k}$ and nonlinear functions of $u_{0,k}$. Note that such conditions are not artificial and they are actually salient features of many complex dynamical systems with multiscale structures (18), multilevel dynamics (20), or state-dependent parameterizations (17). One important characteristic of the resulting covariance matrix is that the evolution of the $k$th block, representing the conditional covariance of $u_{0,k}$ conditioned on $u_k$, has no interactions with that of $u_{0,k}$ for all $k \neq k'$. This allows for an extremely efficient parallel computation due to the small size of each individual block. On the other hand, the conditional means of different $u_{0,k}$ are all coupled in their evolutions.

The second effective strategy exploits the statistical symmetry if the dynamical system in Eq. 1 is statistically homogeneous; namely, the statistics of different $u_k$ are identical with each other. The statistical symmetry is often satisfied when the underlying dynamics in Eq. 1 represent a discrete approximation of some PDEs in a periodic domain with nonlinear advection, diffusion, and homogeneous external forcing. Examples include a rich class of models in geophysical turbulence and excitable media (4, 21). In light of statistical symmetry, the number of samples $L$ in the algorithms can be greatly reduced. In fact, the effective sample size of each $u_k$ becomes $L' = L K$ and therefore a much smaller $L$ is needed to reach the same level of accuracy as in the situation without using statistical symmetry.

These effective strategies are incorporated into the basic algorithms and then applied to two highly non-Gaussian dynamical systems. The first model is a 1,000-dimensional stochastic coupled FitzHugh–Nagumo (FHN) model for excitable media with extreme events (4, 22). It describes activation and deactivation dynamics of spiking neurons and has scale-invariant features. The block decomposition leads to solving the evolution of 500 individual covariance matrices and the statistical symmetry allows an accurate estimation of both the transient and the steady-state PDFs using only $L=1$ samples! The second model is the so-called two-layer Lorenz 96 (L96) model in geophysical turbulence (23, 24), which is widely used as a testbed for data assimilation and parameterization in numerical weather forecasting. Inhomogeneous damping and coupling are adopted in the model with $240$ state variables that mimic the atmosphere motion along a latitude circle with different dissipation over land and sea. Despite the absence of statistical symmetry, the block decomposition facilitates the algorithms to efficiently capture the distinct non-Gaussian features at different locations, using only $L = 500$ samples.

The remainder of this article includes a detailed description of these effective strategies and their application to the stochastic coupled FHN and two-layer L96 models in solving the highly non-Gaussian PDFs at both the transient and statistical equilibrium phases, using a small number of samples.

### Algorithms

#### Conditional Gaussian Framework

The general framework of conditional Gaussian models is given as (10, 25)

$$
\frac{d u_i}{d t} = [A_0(t, u_i) + A_1(t, u_i)u_i] dt + \Sigma_1(t, u_i) dW_i(t). \tag{3a}
$$

$$
\frac{d u_H}{d t} = [a_0(t, u_H) + a_1(t, u_H)u_H] dt + \Sigma_1(t, u_H) dW_H(t). \tag{3b}
$$

where $u = (u_0, u_H)$ with both $u_0 \in R^{N_0}$ and $u_H \in R^{N_H}$ being multidimensional state variables. In Eq. 3, $A_0, A_1, a_0, a_1, \Sigma_1$ and $\Sigma_2$ are vectors and matrices that depend only on time $t$ and the state variables $u_0$ and $W_i(t)$ and $W_H(t)$ are independent Wiener processes. The systems in Eq. 3 are named as conditional Gaussian systems due to the fact that once $u_0(s)$ for $s \leq t$ is given, $u_0(t)$ conditioned on $u_0(s)$ becomes a Gaussian process with mean $\mu(t)$ and covariance $R(t)$, i.e.,

$$
\mathbb{P}(u_0(t) | u_0(s) \leq t) \sim N(\mu(t), R(t)). \tag{4}
$$

Despite the conditional Gaussianity, the coupled system in Eq. 3 remains highly nonlinear and is able to capture many non-Gaussian features as observed in nature (10). One of the desirable features of Eq. 3 is that the conditional distribution in Eq. 4 has the following closed analytical form (25):

$$
\frac{d u_H(t)}{d t} = [a_0(t, u_H) + a_1(t, u_H)u_H] dt + \left( \Sigma_1(t) \Sigma_1^T(t) \right) ^{-1/2} \left( [a_0(t, u_H) + A_1(t, u_H)u_H] dt - R(t) \right) \cdot \frac{d W_I(t)}{d t}, \tag{5a}
$$

$$
\frac{d R(t)}{d t} = \left( [a_0(t, u_H) + A_1(t, u_H)u_H] dt \right) \cdot \left( \Sigma_1(t) \Sigma_1^T(t) \right) ^{-1/2} \left( [a_0(t, u_H) + A_1(t, u_H)u_H] dt - R(t) \right) \cdot \frac{d W_I(t)}{d t}. \tag{5b}
$$

### Basic Algorithms

Here, we summarize the basic efficient statistically accurate algorithms developed in ref. 9. First, we generate $L$ independent trajectories of the variables $u_0$, namely $u_0(1), \ldots, u_0(L)$. Then, different strategies are used to deal with $u_0$ and $u_H$. The PDF of $u_0$ is estimated via a parametric method that exploits the closed form of the conditional Gaussian statistics in Eq. 5,

$$
\mathbb{P}(u_0(t)) = \lim_{L \to \infty} \frac{1}{L} \sum_{i=1}^{L} \mathbb{P}(u_0(t) | u_0(s) \leq t). \tag{6}
$$

Note that the limit $L \to \infty$ in Eq. 6 (as well as Eqs. 7 and 8 below) is taken to illustrate the statistical intuition, while the estimator is the nonasymptotic version. On the other hand, a Gaussian kernel density estimation method is used for solving the PDF of the observed variables $u_0$,

$$
\mathbb{P}(u_0(t)) = \lim_{L \to \infty} \frac{1}{L} \sum_{i=1}^{L} \mathbb{K}_H \left( u_0(t) - u_0(t) \right), \tag{7}
$$

where $\mathbb{K}_H(\cdot)$ is a Gaussian kernel centered at each sample point $u_0(t)$ with covariance given by the bandwidth matrix $H(t)$. The kernel density estimation algorithm here involves a “solve-the-equation plug-in” approach for optimizing the bandwidth (26) for the work function $t$ PDFs. Finally, combining Eqs. 6 and 7, a hybrid method is applied to solve the joint PDF of $u_0$ and $u_H$ through a Gaussian mixture,

$$
\mathbb{P}(u_0(t), u_H(t)) = \lim_{L \to \infty} \frac{1}{L} \sum_{i=1}^{L} \left( \mathbb{K}_H(u_0(t) - u_0(t)) \cdot \mathbb{P}(u_0(t) | u_0(s) \leq t) \right). \tag{8}
$$
Practically, \( L \sim O(100) \) is sufficient for the hybrid method to solve the joint PDF with \( N_{\text{f}} \leq 3 \) and \( N_{\text{d}} \sim 10 \). Since \( L \) is small, the trajectories \( u_{i}^{k}(s \leq t) \), \( \ldots \), \( u_{i}^{k}(s \leq t) \) can be obtained by running a Monte Carlo simulation for the coupled system [3], which is computationally affordable. In addition, the closed form of the \( L \) conditional distributions in Eq. 6 can be solved in a parallel way due to their independence (9), which further reduces the computational cost.

Beating the Curse of Dimension with Block Decomposition. The basic algorithms succeed in solving the Fokker–Planck equation with \( O(10) \) state variables. Now we develop an effective strategy with block decomposition and incorporate it into the basic algorithms. The expanded algorithms can efficiently solve the Fokker–Planck equation in much higher dimensions even with orders in the millions and beat the curse of dimension.

Consider the following decomposition of state variables

\[
\mathbf{u}_{k} = (\mathbf{u}_{1,k}, \mathbf{u}_{2,k}) \quad \text{with} \quad \mathbf{u}_{1,k} \in \mathbb{R}^{N_{1,k}} \quad \text{and} \quad \mathbf{u}_{2,k} \in \mathbb{R}^{N_{2,k}},
\]

where \( 1 \leq k \leq K \), \( N_{\text{f}} = \sum_{k=1}^{K} N_{1,k} \), and \( N_{\text{d}} = \sum_{k=1}^{K} N_{2,k} \). Correspondingly, the full dynamics in Eq. 3 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, we assume both \( \Sigma_{t} \) and \( \Sigma_{d} \) are diagonal for notation simplicity.

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

\textbf{Condition 1:} In the dynamics of each \( \mathbf{u}_{k} \) in Eq. 3, the terms \( A_{0,k} \) and \( a_{0,k} \) depend on all of the components of \( \mathbf{u}_{k} \) while the terms \( A_{1,k} \) and \( a_{1,k} \) are only functions of \( \mathbf{u}_{1,k} \); namely,

\[
\begin{align*}
A_{0,k} & := A_{0,k}(t, \mathbf{u}_{k}), \quad a_{0,k} := a_{0,k}(t, \mathbf{u}_{k}), \\
A_{1,k} & := A_{1,k}(t, \mathbf{u}_{1,k}), \quad a_{1,k} := a_{1,k}(t, \mathbf{u}_{1,k}).
\end{align*}
\]

In addition, only \( \mathbf{u}_{2,k} \) interacts with \( A_{1,k} \) and \( a_{1,k} \) on the right-hand side of the dynamics of \( \mathbf{u}_{1,k} \).

\textbf{Condition 2:} The initial values of \( (\mathbf{u}_{1,k}, \mathbf{u}_{2,k}) \) and \( (\mathbf{u}_{1,k'}, \mathbf{u}_{2,k'}) \) with \( k \neq k' \) are independent from each other.

\textbf{Conditions 1 and 2} are not artificial and they are actually the salient features of many complex systems with multiscale structures (18), multilevel dynamics (20), or state-dependent parameterizations (17). Under these two conditions, the conditional covariance matrix becomes block diagonal, which can be easily verified according to Eq. 5b. The evolution of the conditional covariance of \( \mathbf{u}_{1,k} \) conditioned on \( \mathbf{u}_{1} \) is given by

\[
\begin{align*}
d\mathbf{R}_{1,k}(t) &= \left\{ a_{1,k}(t, \mathbf{R}_{1,k} + \mathbf{R}_{1,k} A_{1,k}^{*} + (\Sigma_{1,k}^{*} \Sigma_{1,k}^{-1} (\mathbf{R}_{1,k} A_{1,k}^{*}))^{-1} \right\} dt, \\
&= \left[ \mathbf{R}_{1,k} - \mathbf{R}_{1,k} A_{1,k}^{*} (\Sigma_{1,k}^{*} \Sigma_{1,k}^{-1} (\mathbf{R}_{1,k} A_{1,k}^{*}))^{-1} \left[ \mathbf{R}_{1,k} - \mathbf{R}_{1,k} A_{1,k}^{*} \right] \right] dt.
\end{align*}
\]

where \( \mathbf{R}_{1,k} \) is diagonal, which can be computed efficiently with block decomposition.

Statistical Symmetry. The computational cost in the algorithms developed above can be further reduced if the coupled system Eq. 3 has statistical symmetry; namely,

\[
\begin{align*}
\Sigma_{t} &:= \Sigma_{t}(t, \mathbf{u}_{1,k}), \quad \Sigma_{d} := \Sigma_{d}(t, \mathbf{u}_{1,k}), \\
&= \left\{ a_{1,k}(t, \mathbf{R}_{1,k} + \mathbf{R}_{1,k} A_{1,k}^{*} + (\Sigma_{1,k}^{*} \Sigma_{1,k}^{-1} (\mathbf{R}_{1,k} A_{1,k}^{*}))^{-1} \right\} dt, \\
&= \left[ \mathbf{R}_{1,k} - \mathbf{R}_{1,k} A_{1,k}^{*} (\Sigma_{1,k}^{*} \Sigma_{1,k}^{-1} (\mathbf{R}_{1,k} A_{1,k}^{*}))^{-1} \left[ \mathbf{R}_{1,k} - \mathbf{R}_{1,k} A_{1,k}^{*} \right] \right] dt.
\end{align*}
\]

In Fig. 1, Stochastic coupled FHN model. Shown are three dynamical regimes of the stochastic coupled FHN model in Eq. 10. (A) Strongly coherent regime with \( \delta_{2} = 0.1 \) and \( d_{u} = 10 \). (B) Weakly coherent regime with \( \delta_{2} = 0.4 \) and \( d_{u} = 0.5 \). (C) Strongly mixed regime with \( \delta_{2} = 0.8 \) and \( d_{u} = 0.5 \). Each plot shows the spatial-temporal evolutions of \( u \) and \( v \) (A–C, Top) as well as the time series and equilibrium PDFs at \( i = 1 \) (A–C, Bottom). Due to the homogeneous property, the PDFs at different grid points are the same.

In Fig. 2, Stochastic coupled FHN model. Shown is time evolution of 1D marginal statistics associated with \( u_{1} \) and \( v_{1} \). Due to the homogeneous property, these statistics at different grid points remain the same. The black circles in A–D mark the time instants that the recovered PDFs from the efficient statistically accurate algorithms are compared with the truth from Figs. 3 and 4. At the statistical equilibrium phase, the skewness and kurtosis of \( u_{1} \) are 1.5 and 3.7, respectively.
to drive the system above the threshold level of global stability and triggers limit cycles intermittently. Note that with $N = 1$, the model reduces to the classical FHN model with a single neuron and it contains the model families with both coherence resonance and self-induced stochastic resonance (29). With different choices of the noise strength $\delta_1, \delta_2$ and the diffusion coefficient $d_u$, the system in Eq. 10 exhibits rich dynamical behaviors. Below, we adopt constant parameters and the initial values are $u_i(0) = -2, v_i(0) = 0.5$ for all $i$. Therefore, the model satisfies the statistical symmetry.

Model Behavior in Different Dynamical Regimes. Fig. 1 shows the model behavior in three different regimes with $N = 500$. Here, the noise coefficient $\delta_1 = 0.2$ is fixed while different values of $\delta_2$ and $d_u$ are chosen for the three regimes. In Fig. 1d, the spatial–temporal patterns are highly coherent due to the choice of a weak noise $\delta_2 = 0.1$ and a strong diffusion $d_u = 10$. The time series of both $u_i$ and $v_i$ have nearly regular oscillations with large bursts in $u_i$ around every 3 units. The associated statistical equilibrium PDF of $v_i$ is bimodal while that of $v_i$ is skewed. With an increase of the noise $\delta_2 = 0.4$ and a decrease of the diffusion $d_u = 0.5$ (Fig. 1b), the coherent patterns become much weaker and only quasi-periodic periods are found in the time series of $v_i$. The associated PDF of $v_i$ turns into symmetric and slightly sub-Gaussian. With a further increase of the noise to $\delta_2 = 0.8$ (Fig. 1c), the spatial–temporal pattern becomes strongly mixed and the time series is more irregular. Correspondingly, the PDF of $v_i$ becomes nearly Gaussian with a large variance.

It is shown in SI Appendix that the FHN model Eq. 10 is scale invariant in all three regimes. The scale-invariant structure means that the spatial–temporal structures in any given scale change little as the number of spatial grid points $N$ increases. Mutual information (30) is used to quantify the dependence of different variables with strongly non-Gaussian features, the advantage of which over pattern correlation is also clearly illustrated in SI Appendix.

Recovering the PDFs at Both Transient and Equilibrium Phases

Exploiting Statistical Symmetry. Now we apply the efficient statistically accurate algorithms to solve the PDFs associated with the stochastic coupled FHN system in Eq. 10. Here we focus on the weakly coherent regime (Fig. 1b). Due to the statistical symmetry, the effective sample size is $L' = NL = 500L$, where $L$ is the number of repeated simulations of the systems. Below, we simply take $L = 1$ in the efficient statistically accurate algorithms, which is extremely cheap. For comparison, 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 $L_C = NL_C = 150,000$.

The time evolutions of the first four moments associated with $u_i$ and $v_i$ are shown in Fig. 2, where the black circles mark three transient phases and the statistical equilibrium phase with

\begin{align}
\frac{du_i}{dt} &= u_i - \frac{1}{3}u_i^3 + v_i + \sqrt{\epsilon_1} W_{1i} + d_u (u_{i+1} + u_{i-1} - 2u_i), \\
\frac{dv_i}{dt} &= u_i + a + \delta_2 W_{2i},
\end{align}

where $u_i$ and $v_i$ are the activator and inhibitor variables and they belong to $u_{i,1}$ and $u_{i,0}$ in the conditional Gaussian framework, respectively, such that the conditions of the block decomposition strategy are satisfied. Periodic boundary conditions are imposed on $u_i$ variables. In Eq. 10, the timescale ratio $\epsilon = 0.01 \ll 1$ leads to a slow–fast structure of the model. The parameter $a = 1.05 > 1$ such that the system has a global attractor in the absence of noise and diffusion (28). The random noise is able
different non-Gaussian features. Fig. 3 shows the skill of solving the one-point statistics at these phases. First, the PDFs of \( u_i \) at \( t = 1.6, 1.9 \) and 25 with different bimodal features are all accurately recovered. Next, \( p(u_i) \) at \( t = 2.7 \) has a tiny second peak around \( u_i = 1 \) since only about 1.5% of the events at this time instant have large bursts and these rare events contribute to a significant kurtosis of \( u_i \) at \( t = 2.7 \). Nevertheless, the efficient statistically accurate algorithm succeeds in capturing the appearance of this weak peak and the recovered PDF almost overlaps with the truth. On the other hand, the skewed \( p(v_i) \) at \( t = 1.6 \) and the sub-Gaussian \( p(v_i) \) at \( t = 1.9, 2.7 \), and 25 are all recovered with high accuracy by the algorithm. In addition, the recovered joint PDFs \( p(u_i, v_i) \) at different phases all resemble the truth.

Fig. 4 illustrates the recovered two-point statistics at \( t = 1.6 \). Fig. 4A and B shows the PDFs \( p(u_1, u_2) \) and \( p(u_1, u_{20}) \) and those for \( v_i \), respectively. The variables \( u_i \) and \( u_j \) have a strong correlation while \( u_i \) and \( u_{20} \) are nearly uncorrelated. These features as well as the highly non-Gaussian PDFs with multiple peaks are both captured by the recovered PDFs from the efficient statistically accurate algorithms with only \( L = 1 \). Likewise, the skewed and nearly Gaussian joint PDFs \( p(v_1, v_2) \) and \( p(v_1, v_{20}) \) are also recovered with high accuracy.

**Two-Layer Inhomogeneous L96 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 (20, 23, 24). 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 \( v_{i,j} \),

\[
\frac{du_i}{dt} = u_{i-1}(u_{i+1} - u_i - 2) + \sum_{j=1}^{J} \gamma_{i,j} u_i v_{i,j} - \tilde{d}_i u_i + F + \sigma_i \tilde{W}_{u,i}, \quad i = 1, \ldots, I, \quad [11a]
\]

\[
\frac{dv_{i,j}}{dt} = -d_{i,j} v_{i,j} - \gamma_{i,j} u_i^2 + \sigma_{i,j} \tilde{W}_{v_{i,j}}, \quad j = 1, \ldots, J, \quad [11b]
\]

with periodic boundary conditions in \( u_i \). One important feature of Eq. 11 is that the nonlinear interaction between \( u_i \) and \( v_{i,j} \) conserves energy, as observed in nature (3). The two-layer L96...
model belongs to the conditional Gaussian framework in Eq. 3 with \( u_i = \{ u_i \} \) and \( u_{ij} = \{ v_{ij} \} \). Below, we take \( I = 40 \) as in the standard L96 model. Associated with each \( u_i \) there are \( J = 5 \) variables \( v_{ij} \) representing different scales of fluctuations. Thus, the total number of state variables is 240. A constant forcing \( F = 8 \) is imposed in Eq. 11 to both the damping \( d_i \) and the coupling \( g_{ij} \) are functions in space. Therefore, the model is inhomogeneous. 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. As a result, the large-scale wave patterns over the ocean are more significant (Fig. 5). The model in Eq. 11 has many desirable properties as in more complicated turbulent systems. Particularly, the smaller scales are more intermittent with stronger fat tails in PDFs. See SI Appendix for more details.

Although the two-layer inhomogeneous L96 model in Eq. 11 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 Eq. 11 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 \( 4 \times 10^7 \). 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 \( t = 25 \).

Fig. 6 shows the one-point statistics as a function of \( t \) regarding the first four moments of \( u_i \) and each \( v_{ij} \). The mean and variance resulting from the efficient statistically accurate algorithms of all variables are highly consistent with the truth. For the skewness and kurtosis, despite small fluctuations in the recovered statistics, the pattern correlations between the curves associated with the truth and the recovered statistics are significant unless both are nearly flat (SI Appendix). These results indicate the success of the efficient statistically accurate algorithms in capturing the inhomogeneous behavior of the model. Fig. 7 demonstrates the skill of recovering different 2D joint PDFs. Figs. 7A and B shows the PDFs \( p(u_i, v_{ij}) \) at \( i = 11 \) and 21. These highly non-Gaussian PDFs with distinct features are recovered accurately by the algorithms. On the other hand, Fig. 7C shows the joint PDFs of the two smallest-scale fluctuation variables \( v_{11,5} \) and \( v_{13,3,5} \), which are highly correlated and have strong non-Gaussian features. Fig. 7D shows the joint PDFs \( p(u_{11}, v_{13}) \), where the two components also have a strong correlation. The efficient statistically accurate algorithms succeed in solving all these joint PDFs with only \( L = 500 \) samples.

**Concluding Discussion**

Effective strategies involving block decomposition of the conditional covariance matrix and statistical symmetry are developed and incorporated into the efficient statistically accurate algorithms in ref. 9. The resulting expanded algorithms are able to efficiently solve the Fokker–Planck equation in much higher dimensions even with orders in the millions and thus beat the curse of dimension. Applications of these effective strategies to both the stochastic coupled FHN and the two-layer inhomogeneous L96 models illustrate the efficiency and accuracy.

It is worthwhile pointing out that although only the recovered one-point and two-point statistics are shown in this article for illustration purposes, the algorithms can actually provide an accurate estimation of the full joint PDF of \( u_i \) using a small number of samples. This is because the sample size in these algorithms does not grow exponentially as the dimension of \( u_i \) which is fundamentally different from Monte Carlo methods. See ref. 19 for a theoretical justification. The algorithms developed here are extremely useful in understanding the causal- as well as improving the parameterizations and predictions of high-dimensional complex turbulent dynamical systems with non-Gaussian features.

**ACKNOWLEDGMENTS.** The research of A.J.M. is partially supported by the Office of Naval Research (ONR) Multidisciplinary University Research Initiative (MURI) Grant N0001416-1-2161 and the New York University Abu Dhabi Research Institute. N.C. is supported as a postdoctoral fellow through A.J.M.’s ONR MURI Grant.

1. Gardiner CW (1985) Stochastic Methods (Springer, Berlin).
2. Risken H (1989) The Fokker-Planck equation. Methods of Solution and Applications, Springer Series in Synergetics, ed Haken H (Springer, Berlin), Vol 18.
3. Majda A (2016) Introduction to Turbulent Dynamical Systems in Complex Systems, Frontiers in Applied Dynamical Systems: Reviews and Tutorials 5 (Springer, Cham, Switzerland).
4. Lindner B, Garcia-Ojalvo J, Neiman A, Schimansky-Geier L (2004) Effects of noise in excitable systems. Phys Rep 392:321–424.
5. Cousins W, Sapsis TP (2016) Reduced-order precursors of rare events in unidirectional nonlinear water waves. J Fluid Mech 790:368–388.
6. Vallis GK (2017) Atmospheric and Oceanic Fluid Dynamics (Cambridge Univ Press, Cambridge, UK).
7. Pichler I, Masud A, Bergman LA (2013) Numerical solution of the Fokker-Planck equation by finite difference and finite element methods – A comparative study. Computational Methods in Stochastic Dynamics, eds Papadakis M, Stefanou G, Papadopoulos V (Springer, Dordrecht, The Netherlands), pp 69–85.
8. Robert CP (2004) Monte Carlo Methods (Wiley, Hoboken, NJ).
9. Chen N, Majda AJ (2017) Efficient statistically accurate algorithms for solving Fokker-Planck equations in large dimensions. J Comput Phys, 10.1016/j.jcp.2017.10.022.
10. Chen N, Majda AJ (2016) Filtering nonlinear turbulent dynamical systems through conditional Gaussian statistics. Monthly Weather Rev 144:4885–4917.
11. Chen N, Majda AJ, Giannakis D (2014) Predicting the cloud patterns of the Madden-Julian oscillation through a low-order nonlinear stochastic model. Geophys Res Lett 41:5612–5619.
12. Chen N, Majda AJ (2015) Predicting the real-time multivariate Madden-Julian oscillation index through a low-order nonlinear stochastic model. Monthly Weather Rev 143:2148–2169.
13. Chen N, Majda AJ (2015) Predicting the cloud patterns for the boreal summer intraseasonal oscillation through a low-order stochastic model. Math Clima Weather Forecast 1:1–20.
14. Chen N, Majda AJ, Tong XT (2014) Information barriers for noisy Lagrangian tracers in filtering random incompressible flows. Nonlinearity 27:2133–2163.
15. Chen N, Majda AJ, Tong XT (2015) Noisy Lagrangian tracers for filtering random rotating compressible flows. J Nonlinear Sci 25:451–488.
16. Chen N, Majda AJ (2016) Model error in filtering random compressible flows utilizing noisy Lagrangian tracers. Mon Weather Rev 144:4037–4061.
17. Branicki M, Majda AJ (2013) Dynamic stochastic superresolution of sparsely observed turbulent systems. J Comput Phys 241:333–363.
18. Majda AJ, Grooms I (2014) New perspectives on superparameterization for geophysical turbulence. J Comput Phys 271:60–77.
19. Chen N, Majda AJ, Tong XT (2017) Rigorous analysis for efficient statistically accurate algorithms for solving Fokker-Planck equations in large dimensions. SIAM J Uncertain Quantif, in press.
20. Wilks DS (2005) Effects of stochastic parametrizations in the lorenz’96 system. Q J R Meteorol Soc 131:389–407.
21. Majda A, Wang X (2006) Nonlinear Dynamics and Statistical Theories for Basic Geophysical Flows (Cambridge Univ Press, Cambridge, UK).
22. Muratov CB, Van den Eijnde E, Weinan E (2007) Noise can play an organizing role for the recurrent dynamics in excitable media. Proc Natl Acad Sci USA 104:702–707.
23. Lee Y, Majda AJ (2017) Multiscale data assimilation and prediction using clustered particle filters. J Comput Phys, in press.
24. Arnold H, Moroz I, Palmer T (2013) Stochastic parameterizations and model uncertainty in the lorenz’96 system. Philos Trans R Soc A 371:20110479.
25. Liptser RS, Shiryaev AN (2001) Statistics of Random Processes II: II. Applications (Springer, Berlin), Vol 2.
26. Voss J, et al. (2010) Kernel density estimation via diffusion. Ann Stat 38:2916–2957.
27. Majda AJ, Harlim J (2012) Filtering Complex Turbulent Systems (Cambridge Univ Press, Cambridge, UK).
28. Guckenheimer J, Holmes PJ (2013) Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields (Springer, New York), Vol 42.
29. DeVille RL, Van den Eijnde E, Muratov CB (2005) Two distinct mechanisms of coherence in randomly perturbed dynamical systems. Phys Rev E 72:031105.
30. Branicki M, Majda AJ (2014) Quantifying Bayesian filter performance for turbulent dynamical systems through information theory. Commun Math Sci 12:961–978.