Beating the Curse of Dimension with Accurate Statistics for the Fokker-Planck Equation in Complex Turbulent Systems

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

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

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 new 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 new algorithms are applied to a 1000-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 new 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

The Fokker-Planck equation is a partial differential equation (PDE) that governs the time evolution of the probability density function (PDF) of a complex system with noise (1, 2). For a general nonlinear dynamical system,

$$d\mathbf{u} = \mathbf{F}(\mathbf{u}, t)dt + \mathbf{\Sigma}(\mathbf{u}, t)d\mathbf{W}, \qquad [1]$$

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

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

with $\mathbf{Q} = \mathbf{\Sigma}\mathbf{\Sigma}^T$. 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 $\mathbf{u}$ 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 have 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 modelling 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 (9) involve a hybrid strategy that requires only a small

Significance Statement

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 new 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 new algorithms is illustrated for highly non-Gaussian systems in excitable media and geophysical turbulence.

www.pnas.org/cgi/doi/10.1073/pnas.XXXXXXXXXX

PNAS | October 13, 2017 | vol. XXX | no. XX | 1–6

number of samples. In these algorithms, a conditional Gaussian mixture in the high-dimensional subspace of $\mathbf{u_{II}}$ via an extremely efficient parametric method is combined with a judicious Gaussian kernel density estimation in the remaining low-dimensional subspace of $\mathbf{u_I}$. Particularly, the parametric method provides closed analytical formulae for determining the conditional Gaussian distributions in the high-dimensional subspace of $\mathbf{u_{II}}$ and is therefore computationally efficient and accurate. It has been shown in a stringent set of numerical tests (9) that with an order of $L \sim O(100)$ samples the mixture distribution has a significant skill in capturing both the statistically steady state and the transient behavior with fat tails of non-Gaussian PDFs in up to 6 dimensions. Rigorous analysis (19) indicates that L does not increase exponentially as the dimension of the high-dimensional subspace of $\mathbf{u_{II}}$ in order 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 (9) (hereafter, basic algorithms) that enable the new 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 $\mathbf{u_{II}}$ 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 $\mathbf{u} = \cup_{k=1}^{K} \mathbf{u}_k$ with each $\mathbf{u}_k = (\mathbf{u}_{\mathbf{I},k}, \mathbf{u}_{\mathbf{II},k}) \in \mathbb{R}^{N_{\mathbf{I},k}+N_{\mathbf{II},k}}$ is developed. Then the high-dimensional conditional covariance matrix of $\mathbf{u_{II}}$ conditioned on $\mathbf{u_I}$ becomes a block diagonal matrix under the conditions that the nonlinear terms on the right hand side of $\mathbf{u}_k$ in Eq. (1) consist of any nonlinear interactions between different components of $\mathbf{u_I}$ but only those between $\mathbf{u}_{\mathbf{II},k}$ and nonlinear functions of $\mathbf{u}_{\mathbf{I},k}$. 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 $\mathbf{u}_{\mathbf{II},k}$ conditioned on $\mathbf{u_I}$, has no interactions with that of $\mathbf{u}_{\mathbf{II},k'}$ for all $k' \neq k$. This allows an extremely efficient parallel computation due to the small size of each individual block. On the other hand, the conditional means of different $\mathbf{u}_{\mathbf{II},k}$ 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 $\mathbf{u}_k$ are identical with each other. The statistical symmetry is often satisfied when the underlying dynamics Eq. (1) represents 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 $\mathbf{u}_k$ becomes $L' = LK$ and therefore a much smaller L is needed to reach the same level of accuracy as in the situation without utilizing 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 1000-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 (20, 23, 24), which is widely used as a testbed for data assimilation and parametrization 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 follows (10, 25):

$$d\mathbf{u_I} = [\mathbf{A}_0(t, \mathbf{u_I}) + \mathbf{A}_1(t, \mathbf{u_I})\mathbf{u_{II}}]dt + \mathbf{\Sigma_I}(t, \mathbf{u_I})d\mathbf{W_I}(t), \quad [3a]$$

$$d\mathbf{u_{II}} = [\mathbf{a}_0(t, \mathbf{u_I}) + \mathbf{a}_1(t, \mathbf{u_I})\mathbf{u_{II}}]dt + \mathbf{\Sigma_{II}}(t, \mathbf{u_I})d\mathbf{W_{II}}(t), \quad [3b]$$

where $\mathbf{u} = (\mathbf{u_I}, \mathbf{u_{II}})$ with both $\mathbf{u_I} \in R^{N_{\mathbf{I}}}$ and $\mathbf{u_{II}} \in R^{N_{\mathbf{II}}}$ being multidimensional state variables. In Eq. (3), $\mathbf{A}_0, \mathbf{A}_1, \mathbf{a}_0, \mathbf{a}_1, \mathbf{\Sigma_I}$ and $\mathbf{\Sigma_{II}}$ 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. The systems in Eq. (3) are named as conditional Gaussian systems due to the fact that once $\mathbf{u_I}(s)$ for $s \leq t$ is given, $\mathbf{u_{II}}(t)$ conditioned on $\mathbf{u_I}(s)$ becomes a Gaussian process with mean $\bar{\mathbf{u}}_{\mathbf{II}}(t)$ and covariance $\mathbf{R_{II}}(t)$, i.e.,

$$p(\mathbf{u_{II}}(t)|\mathbf{u_I}(s \leq t)) \sim \mathcal{N}(\bar{\mathbf{u}}_{\mathbf{II}}(t), \mathbf{R_{II}}(t)). \quad [4]$$

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),

$$d\bar{\mathbf{u}}_{\mathbf{II}}(t) = [\mathbf{a}_0(t, \mathbf{u_I}) + \mathbf{a}_1(t, \mathbf{u_I})\bar{\mathbf{u}}_{\mathbf{II}}]dt + (\mathbf{R_{II}}\mathbf{A}_1^*(t, \mathbf{u_I})) \times$$
$$(\mathbf{\Sigma_I}\mathbf{\Sigma_I}^*)^{-1}(t, \mathbf{u_I})[d\mathbf{u_I} - (\mathbf{A}_0(t, \mathbf{u_I}) + \mathbf{A}_1(t, \mathbf{u_I})\bar{\mathbf{u}}_{\mathbf{II}})dt], \quad [5a]$$
$$d\mathbf{R_{II}}(t) = \{\mathbf{a}_1(t, \mathbf{u_I})\mathbf{R_{II}} + \mathbf{R_{II}}\mathbf{a}_1^*(t, \mathbf{u_I}) + (\mathbf{\Sigma_{II}}\mathbf{\Sigma_{II}}^*)(t, \mathbf{u_I})$$
$$-\mathbf{R_{II}}\mathbf{A}_1^*(t, \mathbf{u_I})(\mathbf{\Sigma_I}\mathbf{\Sigma_I}^*)^{-1}(t, \mathbf{u_I})(\mathbf{R_{II}}\mathbf{A}_1^*(t, \mathbf{u_I}))^*\}dt. \quad [5b]$$

Basic algorithms. Here, we summarize the basic efficient statistically accurate algorithms developed in (9). First, we generate L independent trajectories of the variables $\mathbf{u_I}$, namely $\mathbf{u_I}^1(s \leq t), \ldots, \mathbf{u_I}^L(s \leq t)$. Then, different strategies are used to deal with $\mathbf{u_I}$ and $\mathbf{u_{II}}$. The PDF of $\mathbf{u_{II}}$ is estimated via a parametric method that exploits the closed form of the conditional Gaussian statistics in Eq. (5),

$$p(\mathbf{u_{II}}(t)) = \lim_{L \to \infty} \frac{1}{L} \sum_{i=1}^{L} p(\mathbf{u_{II}}(t)|\mathbf{u_I}^i(s \leq t)). \quad [6]$$

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

$$p\big(\mathbf{u_I}(t)\big) = \lim_{L \to \infty} \frac{1}{L} \sum_{i=1}^{L} K_{\mathbf{H}}\Big(\mathbf{u_I}(t) - \mathbf{u_I}^i(t)\Big), \qquad [7]$$

where $K_{\mathbf{H}}(\cdot)$ is a Gaussian kernel centered at each sample point $\mathbf{u_I}^i(t)$ with covariance given by the bandwidth matrix $\mathbf{H}(t)$. The kernel density estimation algorithm here involves a "solve-the-equation plug-in" approach for optimizing the bandwidth (26) that works for any non-Gaussian PDFs. Finally, combining Eq. (6) and Eq. (7), a hybrid method is applied to solve the joint PDF of $\mathbf{u_I}$ and $\mathbf{u_{II}}$ through a Gaussian mixture,

$$p(\mathbf{u_I}(t), \mathbf{u_{II}}(t))$$
$$= \lim_{L \to \infty} \frac{1}{L} \sum_{i=1}^{L} \Big(K_{\mathbf{H}}(\mathbf{u_I}(t) - \mathbf{u_I}^i(t)) \cdot p(\mathbf{u_{II}}(t) | \mathbf{u_I}^i(s \le t)) \Big). \quad [8]$$

Practically, $L \sim O(100)$ is sufficient for the hybrid method to solve the joint PDF with $N_{\mathbf{I}} \le 3$ and $N_{\mathbf{II}} \sim 10$. Since L is small, the trajectories $\mathbf{u_I}^1(s \le t), \dots, \mathbf{u_I}^L(s \le t)$ can be obtained by running a Monte Carlo simulation for the coupled system Eq. (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 new 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

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

where $1 \le k \le K$, $N_{\mathbf{I}} = \sum_{k=1}^{K} N_{\mathbf{I},k}$ and $N_{\mathbf{II}} = \sum_{k=1}^{K} N_{\mathbf{II},k}$. Correspondingly, the full dynamics in 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 $\boldsymbol{\Sigma}_{\mathbf{I}}$ and $\boldsymbol{\Sigma}_{\mathbf{II}}$ 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.
Condition 1: In the dynamics of each $\mathbf{u}_k$ in Eq. (3), the terms $\mathbf{A}_{0,k}$ and $\mathbf{a}_{0,k}$ can depend on all the components of $\mathbf{u_I}$ while the terms $\mathbf{A}_{1,k}$ and $\mathbf{a}_{1,k}$ are only functions of $\mathbf{u_{I},}_k$, namely,

$$\begin{aligned} \mathbf{A}_{0,k} &:= \mathbf{A}_{0,k}(t, \mathbf{u_I}), & \mathbf{a}_{0,k} &:= \mathbf{a}_{0,k}(t, \mathbf{u_I}), \\ \mathbf{A}_{1,k} &:= \mathbf{A}_{1,k}(t, \mathbf{u_{I},}_k), & \mathbf{a}_{1,k} &:= \mathbf{a}_{1,k}(t, \mathbf{u_{I},}_k). \end{aligned} \qquad [9]$$

In addition, only $\mathbf{u_{II},}_k$ interacts with $\mathbf{A}_{1,k}$ and $\mathbf{a}_{1,k}$ on the right hand side of the dynamics of $\mathbf{u}_k$.
Condition 2: The initial values of $(\mathbf{u_{I},}_k, \mathbf{u_{II},}_k)$ and $(\mathbf{u_{I},}_{k'}, \mathbf{u_{II},}_{k'})$ with $k \ne k'$ are independent with each other.

These two conditions are not artificial and they are actually the salient features of many complex systems with multiscale

structures (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 $\mathbf{u_{II},}_k$ conditioned on $\mathbf{u_I}$ is given by,

$$\begin{aligned} d\mathbf{R_{II},}_k(t) = \Big\{ &\mathbf{a}_{1,k}\mathbf{R_{II},}_k + \mathbf{R_{II},}_k\mathbf{a}_{1,k}^* + (\boldsymbol{\Sigma}_{\mathbf{II},k}\boldsymbol{\Sigma}_{\mathbf{II},k}^*) \\ &- (\mathbf{R_{II},}_k\mathbf{A}_{1,k}^*)(\boldsymbol{\Sigma}_{\mathbf{I},k}\boldsymbol{\Sigma}_{\mathbf{I},k}^*)^{-1}(\mathbf{R_{II},}_k\mathbf{A}_{1,k}^*)^* \Big\} dt, \end{aligned}$$

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

Next, the structures of $\mathbf{A}_{0,k}$ and $\mathbf{a}_{0,k}$ in Eq. (9) allow the coupling among all the K groups of variables in the conditional mean according to Eq. (5a). The evolution of $\bar{\mathbf{u}}_{\mathbf{II},k}$, namely the conditional mean of $\mathbf{u_{II},}_k$ conditioned on $\mathbf{u_I}$, is given by

$$\begin{aligned} d\bar{\mathbf{u}}_{\mathbf{II},k}(t) = &[\mathbf{a}_{0,k} + \mathbf{a}_{1,k}\bar{\mathbf{u}}_{\mathbf{II},k}]dt + \mathbf{R_{II},}_k\mathbf{A}_{1,k}^*(\boldsymbol{\Sigma}_{\mathbf{I},k}\boldsymbol{\Sigma}_{\mathbf{I},k}^*)^{-1} \times \\ &[d\mathbf{u_I},_k - (\mathbf{A}_{0,k}(t, \mathbf{u_I}) + \mathbf{A}_{1,k}\bar{\mathbf{u}}_{\mathbf{II},k})dt]. \end{aligned}$$

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

$$p\big(\mathbf{u_{I},}_k(t), \mathbf{u_{II},}_k(t)\big) = p\big(\mathbf{u_{I},}_{k'}(t), \mathbf{u_{II},}_{k'}(t)\big), \quad \text{for all } t, k \text{ and } k',$$

including the initial conditions. The statistical symmetry is often satisfied when the underlying dynamical system represents a discrete approximation of some PDEs in a periodic domain with nonlinear advection, diffusion and homogeneous external forcing (21, 27).

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

A stochastic coupled FitzHugh-Nagumo (FHN) model

The efficient statistically accurate algorithms developed in this article work for a wide class of models in excitable media (4), including different versions of the famous FitzHugh-Nagumo (FHN) model that describes activation and deactivation dynamics of spiking neurons. Here, the algorithms are applied to a stochastic coupled FHN model with N elements (4, 22),

$$\epsilon \frac{du_i}{dt} = u_i - \frac{1}{3}u_i^3 - v_i + \sqrt{\epsilon}\delta_1 \dot{W}_{u_i} + d_u(u_{i+1} + u_{i-1} - 2u_i),$$
$$i = 1, \dots, N, \qquad [10a]$$

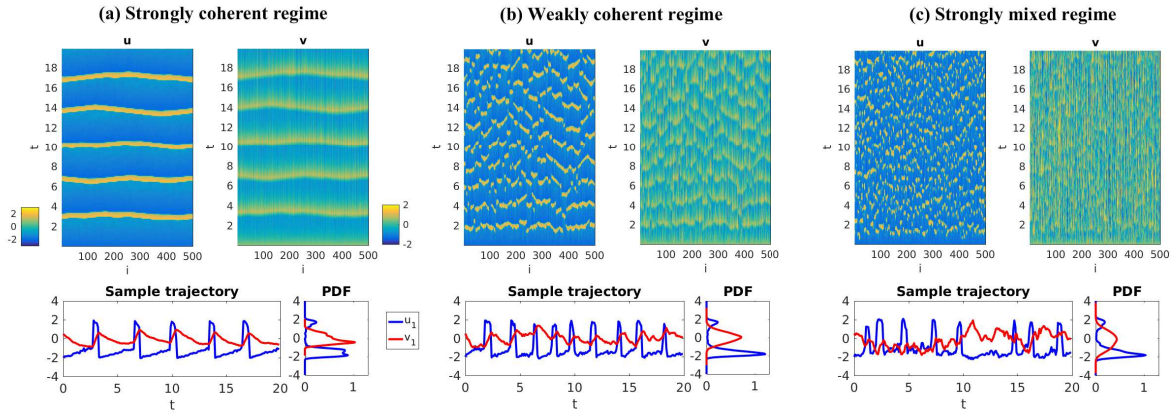$$\frac{dv_i}{dt} = u_i + a + \delta_2 \dot{W}_{v_i}, \qquad [10b]$$

Fig. 1. [Stochastic coupled FHN model]. 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 subfigure shows the spatial-temporal evolutions of u and v (top) as well as the time series and equilibrium PDFs at $i = 1$. Due to the homogenous property, the PDFs at different grid points are the same.

where u_i and v_i are activator and inhibitor variables and they belong to $\mathbf{u_I}$ and $\mathbf{u_{II}}$ 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 time scale 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 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 δ_1, δ_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 δ_2 and d_u are chosen for the three regimes. In column (a), 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_1 and v_1 have nearly regular oscillations with large bursts in u_i around every 3 units. The associated statistical equilibrium PDF of u_1 is bimodal while that of v_1 is skewed. With an increase of the noise $\delta_2 = 0.4$ and a decrease of the diffusion $d_u = 0.5$ (column (b)), the coherent patterns becomes much weaker and only quasi-regular periods are found in the time series of u_1. The associated PDF of v_1 turns into symmetric and slightly sub-Gaussian. With a further increase of the noise to $\delta_2 = 0.8$ (column (c)), the spatial-temporal pattern becomes strongly mixed and the time series is more irregular. Correspondingly, the PDF of v_1 becomes nearly Gaussian with a large variance.

It is shown in *Supporting Information* that the FHN model Eq. (10) is scale invariant in all the three regimes. The scale invariant structure means that the spatial-temporal structures in any given scale has little change 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 *Supporting Information*.

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 (column (b) in Fig. 1). Due to the statistical symmetry, the effective sample size is $L' = NL = 500L$, where L is the number of the 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_1 and v_1 are shown in Fig. 2, where the black dots mark three transient phases and the statistical equilibrium phase with different non-Gaussian features. Fig. 3 shows the skill of solving the one-point statistics at these phases. First, the PDFs of u_1 at $t = 1.6, 1.9$ and 25 with different bimodal features are all accurately recovered. Next, $p(u_1)$ at $t = 2.7$ has a tiny second peak around $u_1 = 1$ 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_1 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 $p(v_1)$ at $t = 1.6$ and the sub-Gaussian $p(v_1)$ at $t = 1.9, 2.7$ and 25 are all recovered with high accuracy by the algorithm. In addition, the recovered joint PDFs $p(u_1, v_1)$ at different phases all resemble the truth.

Fig. 4 illustrates the recovered two-point statistics at $t = 1.6$. Rows (a) and (b) show the PDFs $p(u_1, u_2)$ and $p(u_1, u_{20})$ and those for v, respectively. The variables u_1 and u_2 have a strong correlation while u_1 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 $p(v_1, v_2)$
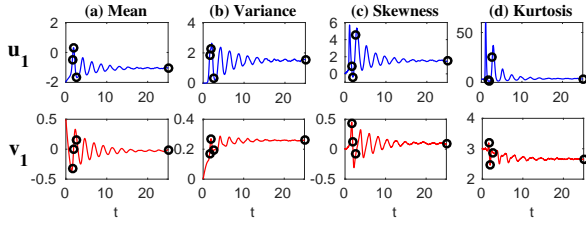

Fig. 2. [Stochastic coupled FHN model]. Time evolution of 1D marginal statistics associated with u_1 and v_1. Due to the homogeneous property, these statistics at different grid points remain the same. The black dots in each panel mark the time instants that the recovered PDFs from the efficient statistically accurate algorithms are compared with the truth in Figs. 3 and 4. At the statistical equilibrium phase, the skewness and kurtosis of u_1 are 1.5 and 3.7, respectively.
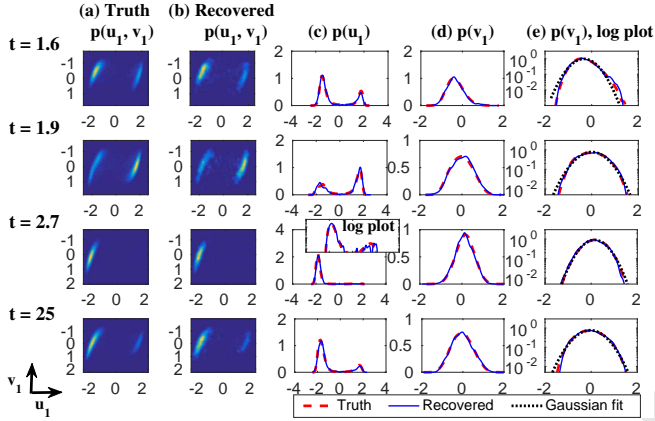

Fig. 3. [Stochastic coupled FHN model]. Comparison of the truth and recovered PDFs at three transient phases $t = 1.6, 1.9$ and 2.7 as well as the statistical equilibrium phase $t = 25$. Columns (a) and (b) show the truth and recovered 2D PDF $p(u_1, v_1)$. Columns (c) and (d) compare the recovered 1D PDFs $p(u_1)$ and $p(v_1)$ (blue) with the truth (red). Column (e) shows the logarithm plots of column (d), where the black dotted curves are the Gaussian fits of the truth. The subpanel above the subplot of $p(u_1)$ at $t = 2.7$ is also in logarithm scale ranging from 5×10^{-3} to 5.

and $p(v_1, v_{20})$ are also recovered with high accuracy.

Two-layer inhomogeneous Lorenz 96 (L96) model

The two-layer Lorenz 96 (L96) model is a conceptual model in geophysical turbulence that is widely used as a testbed for data assimilation and parametrization in numerical weather forecasting (20, 23, 24). The model can be regarded as 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 $v_{i,j}$,

$$\frac{du_i}{dt} = u_{i-1}(u_{i+1} - u_{i-2}) + \sum_{j=1}^{J} \gamma_{i,j} u_i v_{i,j} - \bar{d}_i u_i$$
$$+ F + \sigma_u \dot{W}_{u_i}, \qquad i = 1, \dots, I, \quad [11a]$$
$$\frac{dv_{i,j}}{dt} = -d_{v_{i,j}} v_{i,j} - \gamma_j u_i^2 + \sigma_{i,j} \dot{W}_{v_{i,j}}, \quad j = 1, \dots, J, \quad [11b]$$

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
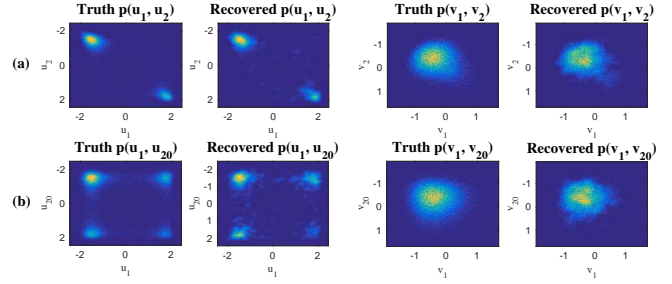

Fig. 4. [Stochastic coupled FHN model]. Comparison of the 2D PDFs with components from two different grid points at $t = 1.6$. Row (a) and (b) show the truth and the recovered $p(u_1, u_2), p(v_1, v_2)$ and $p(u_1, u_{20}), p(v_1, v_{20})$, respectively.

in Eq. (3) with $\mathbf{u_I} = \{u_i\}$ and $\mathbf{u_{II}} = \{v_{i,j}\}$. Below, we take $I = 40$ as in the standard L96 model. Associated with each u_i, there are $J = 5$ variables $v_{i,j}$ representing different scales of fluctuations. Thus, the total number of state variables is 240. A constant forcing $F = 8$ is adopted in Eq. (11a) while both the damping $\bar{d}_i$ and the coupling $\gamma_{i,j}$ are functions in space. Therefore, the model is inhomogeneous. These mimic the situation that the damping and coupling above the ocean is weaker than that above the land since the latter usually has a stronger friction or dissipation. As a result, the large-scale wave patterns over the ocean is more significant. See 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 *Supporting Information* 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 is simulated from Eq. (11) to implement the efficient statistically accurate algorithms. As comparison, a direct Monte Carlo method requires $L_C = 150,000$ samples for each of the 240 variables for an accurate estimation of at least the one-point statistics. This means the total number of samples is around 4×10^7! For an efficient calculation of the truth, we only focus 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 i regarding the first four moments of u_i and each $v_{i,j}$. 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 (See *Supporting Information*). 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. Columns (a) and (b) show the PDFs $p(u_i, v_{i,5})$ at $i = 11$ and 21. These highly non-Gaussian PDFs with distinct features are recovered accurately by the algorithms. On the other hand, column (c) shows the joint PDFs of two smallest scale fluctuation variables $v_{11,5}$ and $v_{13,5}$, which are highly correlated and have strong

Chen *et al.*

PNAS | **October 13, 2017** | vol. XXX | no. XX | **5**

non-Gaussian features. Column (d) shows the joint PDFs $p(u_{11}, u_{13})$, 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.
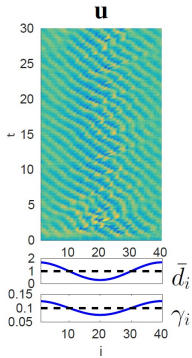

Fig. 5. Spatial-temporal evolution of the observed variables $u_i, i = 1, \ldots 40$ in the two-layer L-96 model Eq. (11) with $F = 8$. The profiles of damping $\bar{d}_i$ and coupling coefficients γ_i as a function of i are shown below the spatial-temporal evolution of u_i in solid blue curves and the black dashed lines represent the spatial averaged values. Here, $\bar{d}_i = 1 + 0.7\cos(2\pi i/J)$ and $\gamma_{i,j} = \gamma_i = 0.1 + 0.025\cos(2\pi i/J)$.
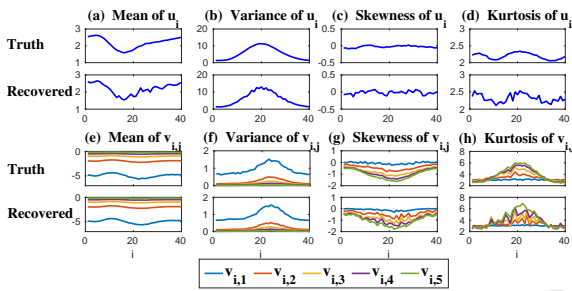

Fig. 6. [Two-layer inhomogeneous L-96 model]. Comparison of the truth and the recovered first four moments of u_i and $v_{i,j}$ for $j = 1, \ldots, 5$ as a function of i.
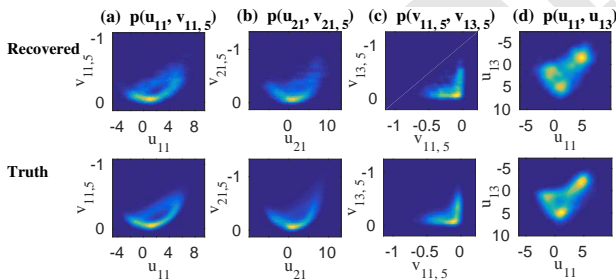

Fig. 7. [Two-layer inhomogeneous L-96 model]. Comparison of 2D joint PDFs at a fixed grid point (Panels (a)-(b)) and between two different grid points (Panels (c)-(d)).

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 (9). The resulting new 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 the illustration purpose, the algorithms can actually provide an accurate estimation of the full joint PDF of $\mathbf{u_{II}}$ using a small number of samples. This is because the sample size in these algorithms does not grow exponentially as the dimension of $\mathbf{u_{II}}$, which is fundamentally different from Monte Carlo methods. See (19) for a theoretical justification. The algorithms developed here are extremely useful in understanding the causality as well as improving the parameterizations and predictions of high-dimensional complex turbulent dynamical systems with non-Gaussian features.

1. Gardiner CW (1985) *Stochastic methods.* (Springer-Verlag, Berlin–Heidelberg–New York–Tokyo).
2. Risken H (1989) The Fokker-Planck equation. Methods of solution and applications, vol. 18 of. *Springer Series in Synergetics*.
3. Majda A (2016) *Introduction to turbulent dynamical systems in complex systems, Frontiers in Applied Dynamical Systems: Reviews and Tutorials 5.* (Springer).
4. Lindner B, García-Ojalvo J, Neiman A, Schimansky-Geier L (2004) Effects of noise in excitable systems. *Physics reports* 392(6):321–424.
5. Cousins W, Sapsis TP (2016) Reduced-order precursors of rare events in unidirectional nonlinear water waves. *Journal of Fluid Mechanics* 790:368–388.
6. Vallis GK (2017) *Atmospheric and oceanic fluid dynamics.* (Cambridge University Press).
7. Pichler L, Masud A, Bergman LA (2013) Numerical solution of the Fokker–Planck equation by finite difference and finite element methods – a comparative study in *Computational Methods in Stochastic Dynamics*. (Springer), pp. 69–85.
8. Robert CP (2004) *Monte Carlo methods.* (Wiley Online Library).
9. Chen N, Majda AJ (2017) Efficient statistically accurate algorithms for solving Fokker-Planck equations in large dimensions. *Journal of Computational Physics*. In press.
10. Chen N, Majda AJ (2016) Filtering nonlinear turbulent dynamical systems through conditional Gaussian statistics. *Monthly Weather Review* 144(12):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. *Geophysical Research Letters* 41(15):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 Review* 143(6):2148–2169.
13. Chen N, Majda AJ (2015) Predicting the cloud patterns for the boreal summer intraseasonal oscillation through a low-order stochastic model. *Mathematics of Climate and Weather Forecasting* 1(1):1–20.
14. Chen N, Majda AJ, Tong XT (2014) Information barriers for noisy Lagrangian tracers in filtering random incompressible flows. *Nonlinearity* 27(9):2133.
15. Chen N, Majda AJ, Tong XT (2015) Noisy Lagrangian tracers for filtering random rotating compressible flows. *Journal of Nonlinear Science* 25(3):451–488.
16. Chen N, Majda AJ (2016) Model error in filtering random compressible flows utilizing noisy Lagrangian tracers. *Monthly Weather Review* 144(11):4037–4061.
17. Branicki M, Majda AJ (2013) Dynamic stochastic superresolution of sparsely observed turbulent systems. *Journal of Computational Physics* 241:333–363.
18. Majda AJ, Grooms I (2014) New perspectives on superparameterization for geophysical turbulence. *Journal of Computational Physics* 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/ASA Journal of Uncertainty Quantification*. In press.
20. Wilks DS (2005) Effects of stochastic parametrizations in the lorenz'96 system. *Quarterly Journal of the Royal Meteorological Society* 131(606):389–407.
21. Majda A, Wang X (2006) *Nonlinear dynamics and statistical theories for basic geophysical flows.* (Cambridge University Press).
22. Muratov CB, Vanden-Eijnden E, Weinan E (2007) Noise can play an organizing role for the recurrent dynamics in excitable media. *Proceedings of the National Academy of Sciences* 104(3):702–707.
23. Lee Y, Majda AJ (2017) Multiscale data assimilation and prediction using clustered particle filters. *Journal of Computational Physics*. (In press).
24. Arnold H, Moroz I, Palmer T (2013) Stochastic parametrizations and model uncertainty in the lorenz'96 system. *Phil. Trans. R. Soc. A* 371(1991):20110479.
25. Liptser RS, Shiryaev AN (2001) *Statistics of Random Processes II: II. Applications.* (Springer) Vol. 2.
26. Botev ZI, Grotowski JF, Kroese DP, , et al. (2010) Kernel density estimation via diffusion. *The Annals of Statistics* 38(5):2916–2957.
27. Majda AJ, Harlim J (2012) *Filtering complex turbulent systems.* (Cambridge University Press).
28. Guckenheimer J, Holmes PJ (2013) *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields.* (Springer Science & Business Media) Vol. 42.
29. DeVille RL, Vanden-Eijnden E, Muratov CB (2005) Two distinct mechanisms of coherence in randomly perturbed dynamical systems. *Physical Review E* 72(3):031105.
30. Branicki M, Majda A (2014) Quantifying bayesian filter performance for turbulent dynamical systems through information theory. *Commun. Math. Sci* 12(5):901–978.