Rigorous Analysis for Efficient Statistically Accurate Algorithms for Solving Fokker-Planck Equations in Large Dimensions*

Nan Chen†, Andrew J. Majda‡, and Xin T. Tong§

Abstract. This article presents a rigorous analysis for efficient statistically accurate algorithms for solving the Fokker-Planck equations associated with high-dimensional nonlinear turbulent dynamical systems with conditional Gaussian structures. Despite the conditional Gaussianity, these nonlinear systems contain many strong non-Gaussian features such as intermittency and fat-tailed probability density functions (PDFs). The algorithms involve a hybrid strategy that requires only a small number of samples L to capture both the transient and the equilibrium non-Gaussian PDFs with high accuracy. Here, 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. Rigorous analysis shows that the mean integrated squared error in the recovered PDFs in the high-dimensional subspace is bounded by the inverse square root of the determinant of the conditional covariance, where the conditional covariance is completely determined by the underlying dynamics and is independent of L. This is fundamentally different from a direct application of kernel methods to solve the full PDF, where L needs to increase exponentially with the dimension of the system and the bandwidth shrinks. A detailed comparison between different methods justifies that the efficient statistically accurate algorithms are able to overcome the curse of dimensionality. It is also shown with mathematical rigour that these algorithms are robust in long time provided that the system is controllable and stochastically stable. Particularly, dynamical systems with energy-conserving quadratic nonlinearity as in many geophysical and engineering turbulence are proved to have these properties.

Key words. Fokker-Planck equation, high-dimensional non-Gaussian PDFs, hybrid strategy, small sample size, long time persistence

AMS subject classifications. 35Q84, 76F55, 65C05, 37C75, 93B05

1. Introduction.

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 [26, 65]. Many complex dynamical systems in geophysical and engineering turbulence, neuroscience and excitable media have large dimensions and strong nonlinearities, the associated PDFs of which are highly non-Gaussian with intermittency and extreme events [41, 38]. Predicting the rare and extreme events [15, 19, 29, 63, 61, 20, 73], quantifying the uncertainty in the presence of intermittent instabilities [47, 6, 30, 5] and characterizing other

*Submitted to the editors DATE.

Funding: 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. N.C. is supported as a postdoctoral fellow through A.J.M's ONR MURI Grant. X.T.T is supported by NUS grant R-146-000-226-133.

†Department of Mathematics and Center for Atmosphere Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, NY, USA (chennan@cims.nyu.edu).

‡Department of Mathematics and Center for Atmosphere Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, NY, USA and Center for Prototype Climate Modeling, New York University Abu Dhabi, Saadiyat Island, Abu Dhabi, UAE. (jonjon@cims.nyu.edu).

§Department of Mathematics, National University of Singapore, Singapore (mattxin@nus.edu.sg).

non-Gaussian features [62, 32] all require solving high-dimensional Fokker-Planck equations with strong non-Gaussian features.

Since there is no general closed-form solution for the Fokker-Planck equation, various numerical and approximate approaches have been developed to solve the evolution of the PDF $p(\mathbf{u}, t)$, where $\mathbf{u}$ consists of the state variables and t is the time. However, traditional numerical methods such as finite element and finite difference as well as the direct Monte Carlo simulations of the underlying dynamics all suffer from the curse of dimensionality [66, 22, 64, 35, 70]. Furthermore, even in the low-dimensional scenarios, substantial computational cost is already required for an accurate estimation of the fat tails of the highly intermittent non-Gaussian PDFs. On the other hand, different methods for solving the partial or the approximate solutions of $p(\mathbf{u}, t)$ have been proposed for special dynamical systems. For example, asymptotic expansion with truncations provides good approximate PDFs associated with the slow varying variables in non-Gaussian systems with multiscale features [26, 55, 56, 44]. Splitting methods [23, 24], orthogonal functions and tensor decompositions [75, 71, 65] are able to provide reasonably good estimations of the steady state PDFs. If the systems are weakly nonlinear with additive noise, then equivalent linearization method [69, 3] is also frequently used for solving approximate solutions.

In recent work by two of the authors [14], efficient statistically accurate algorithms have been developed for solving the Fokker-Planck equation associated with high-dimensional nonlinear turbulent dynamical systems with conditional Gaussian structures [11]. Decomposing the state variables $\mathbf{u}$ into two groups $\mathbf{u} = (\mathbf{u_I}, \mathbf{u_{II}})$ with $\mathbf{u_I} \in R^{N_\mathbf{I}}$ and $\mathbf{u_{II}} \in R^{N_\mathbf{II}}$. The conditional Gaussian systems are characterized by the fact that once a single trajectory of $\mathbf{u_I}(s \leq t)$ is given, $\mathbf{u_{II}}(t)$ conditioned on $\mathbf{u_I}(s \leq t)$ becomes a Gaussian process. Despite the conditional Gaussian structure, the coupled system of $\mathbf{u_I}$ and $\mathbf{u_{II}}$ is highly nonlinear and it is able to capture many strong non-Gaussian features such as intermittency and fat-tailed PDFs that are commonly seen in nature [11]. Note that in most turbulent dynamical systems, the observed variables $\mathbf{u_I}$ represent large scale or resolved variables, which usually have only a small dimension, while the dimension of the unresolved or unobserved variables $\mathbf{u_{II}}$ can be very large [53, 41]. Applications of the conditional Gaussian framework to highly nonlinear turbulent dynamical systems include modelling and predicting the highly intermittent and non-Gaussian times series of the Madden-Julian oscillation (MJO) and monsoon [15, 10, 9], filtering the stochastic skeleton model for the MJO [12], and state estimation of the turbulent ocean flows from noisy Lagrangian tracers [16, 17, 13]. Other studies that also fit into the conditional Gaussian framework includes the dynamic stochastic superresolution of sparsely observed turbulent systems using cheap exactly solvable forecast models [7, 34], stochastic superparameterization for geophysical turbulent flows [50], physics constrained nonlinear regression models [52, 31], stochastic parameterized extended Kalman filter [28, 27, 6, 8, 36] and blended particle filters for high-dimensional chaotic systems [54].

The efficient statistically accurate algorithms [14] 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 $\mathbf{u_{II}}$ via an extremely efficient parametric method is combined with a judicious Gaussian kernel density estimation in the low-dimensional subspace of $\mathbf{u_I}$. In particular, the conditional Gaussian distributions in the high-dimensional subspace are solved via closed analytical formulae and are therefore computationally efficient and accurate. The full non-Gaussian joint PDF of the system is then given by a Gaussian mixture. One remarkable feature of these efficient hybrid algorithms is that each conditional Gaussian distribution is able to cover a significant portion of the high-dimensional PDF. This guarantees the sufficiency of using only a small number of samples, which overcomes the curse of dimensionality. It has been shown in a stringent set of numerical tests [14] that with an order of $O(100)$ samples the mixture distribution has a significant skill in capturing both the statistically steady state and the transient behavior with fat tails of the high-dimensional non-Gaussian PDFs in up to 6 dimensions while an order of $O(10^6)$ samples is required in the Monte Carlo simulation to reach the same accuracy. In [14], the restriction to 6 dimension of the hybrid method is not essential but was utilized to allow comprehensive validation of the statistics in the truth model with an instructive simple model.

This article serves as a rigorous analysis for these efficient statistically accurate algorithms. The main focus here is the accuracy of the recovered PDFs in terms of the sample size L as well as its dependence on different factors, in particular the dimension of the state variables and the time span. Throughout the article, the mean integrated square error (MISE) is used to quantify the accuracy.

Our first result Theorem 3.1 reveals that the MISE in the recovered high-dimensional PDFs associated with the unresolved variables $\mathbf{u}_{\mathbf{II}}$ is bounded by $\mathbb{E}(\det(\mathbf{R}_{\mathbf{II}})^{-1/2})$, where $\mathbf{R}_{\mathbf{II}}$ is the conditional covariance of $\mathbf{u}_{\mathbf{II}}$ given the trajectory of $\mathbf{u}_{\mathbf{I}}$. Notably, $\mathbf{R}_{\mathbf{II}}$ is completely determined by the underlying dynamical systems and has no dependence on the sample size L. In contrast, if a direct kernel density method is applied to recover the PDF of $\mathbf{u}_{\mathbf{II}}$, then the bandwidth of the kernel H is scaled as the reciprocal of L to a certain power in order to minimize the MISE and the resulting MISE is proportional to $L^{-1/N_{\mathbf{II}}}$, which means L has to increase exponentially with $N_{\mathbf{II}}$ to guarantee the accuracy in the solution. This indicates the curse of dimensionality in the direct kernel density estimation and other smoothed versions of Monte Carlo methods. Such a notorious issue is overcome by the efficient statistically accurate algorithms due to the independence between $\mathbf{R}_{\mathbf{II}}$ and L in the high-dimensional subspace of $\mathbf{u}_{\mathbf{II}}$. Another significant feature of the efficient statistically accurate algorithms is their long term persistence, which is affirmed by Theorem 3.7 in a rigorous way provided that the joint process $(\mathbf{u}_{\mathbf{I}}, \mathbf{u}_{\mathbf{II}})$ is controllable and stochastically stable. Theorem 3.7 also supplies a lower bound of $\mathbf{R}_{\mathbf{II}}$ using the controllability condition. In addition, Proposition 3.8 demonstrates that dynamical systems with energy conserving quadratic nonlinear interactions as in most geophysical and engineering turbulence [41] automatically satisfy all the conditions for the long time persistence, which justifies the skillful performance of the efficient statistically accurate algorithms in the numerical tests reported in [14]. Further validations of the controllability and other theoretical conditions in the algorithms are demonstrated in the numerical simulations at the end of this article.

The remaining of this article is organized as follows. The high-dimensional nonlinear turbulent dynamical systems with conditional Gaussian structures are summarized in section 2, which is followed by a brief review of the efficient statistically accurate algorithms in [14] for solving the PDFs of such kind of systems. The main theoretical results are shown in section 3, where the proofs are included in section 4 and Supplementary Material. In section 5, numerical tests on a nonlinear triad model and its modified versions are used to validate the theoretical results. Conclusion and discussions are given in section 6.

2. Review of the efficient statistically accurate algorithms for solving the PDFs of nonlinear dynamical systems with conditional Gaussian structures.

2.1. High-dimensional conditional Gaussian models with nonlinear and intermittent dynamical features .

The general framework of high-dimensional conditional Gaussian models is given as follows [39, 11]:

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

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

where the state variables are $\mathbf{u} = (\mathbf{u}_{\mathbf{I}}, \mathbf{u}_{\mathbf{II}})$ with both $\mathbf{u}_{\mathbf{I}} \in R^{N_{\mathbf{I}}}$ and $\mathbf{u}_{\mathbf{II}} \in R^{N_{\mathbf{II}}}$ being multidimensional variables. In (1), $\mathbf{A}_0, \mathbf{A}_1, \mathbf{a}_0, \mathbf{a}_1, \mathbf{\Sigma}_{\mathbf{I}}$ and $\mathbf{\Sigma}_{\mathbf{II}}$ are vectors and matrices that are functions of time t and the state variables $\mathbf{u}_{\mathbf{I}}$, and $\mathbf{W}_{\mathbf{I}}(t)$ and $\mathbf{W}_{\mathbf{II}}(t)$ are independent Wiener processes. Here the noise coefficient matrix $\mathbf{\Sigma}_{\mathbf{I}}$ is non-degenerated in order to guarantee the observability while there is no special requirement for $\mathbf{\Sigma}_{\mathbf{II}}$. The dynamics (1) are named as conditional Gaussian systems due to the fact that once a single trajectory $\mathbf{u}_{\mathbf{I}}(s)$ for $s \le t$ is given, $\mathbf{u}_{\mathbf{II}}(t)$ conditioned on $\mathbf{u}_{\mathbf{I}}(s)$ becomes a Gaussian process with mean $\bar{\mathbf{u}}_{\mathbf{II}}(t)$ and covariance $\mathbf{R}_{\mathbf{II}}(t)$, i.e.,

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

Despite the conditional Gaussianity, the coupled system (1) remains highly nonlinear and is able to capture the strong non-Gaussian features as observed in nature [11]. One of the desirable properties of the conditional Gaussian system (1) is that the conditional distribution in (2) has the following closed analytical form [39],

$$\begin{aligned} d\bar{\mathbf{u}}_{\mathbf{II}}(t) =& [\mathbf{a}_0(t,\mathbf{u}_{\mathbf{I}}) + \mathbf{a}_1(t,\mathbf{u}_{\mathbf{I}})\bar{\mathbf{u}}_{\mathbf{II}}]dt + (\mathbf{R}_{\mathbf{II}}\mathbf{A}_1^*(t,\mathbf{u}_{\mathbf{I}}))(\mathbf{\Sigma}_{\mathbf{I}}\mathbf{\Sigma}_{\mathbf{I}}^*)^{-1}(t,\mathbf{u}_{\mathbf{I}})\times \\ & \qquad [d\mathbf{u}_{\mathbf{I}} - (\mathbf{A}_0(t,\mathbf{u}_{\mathbf{I}}) + \mathbf{A}_1(t,\mathbf{u}_{\mathbf{I}})\bar{\mathbf{u}}_{\mathbf{II}})dt], \\ d\mathbf{R}_{\mathbf{II}}(t) =& \{\mathbf{a}_1(t,\mathbf{u}_{\mathbf{I}})\mathbf{R}_{\mathbf{II}} + \mathbf{R}_{\mathbf{II}}\mathbf{a}_1^*(t,\mathbf{u}_{\mathbf{I}}) + (\mathbf{\Sigma}_{\mathbf{II}}\mathbf{\Sigma}_{\mathbf{II}}^*)(t,\mathbf{u}_{\mathbf{I}}) \\ & \qquad -(\mathbf{R}_{\mathbf{II}}\mathbf{A}_1^*(t,\mathbf{u}_{\mathbf{I}}))(\mathbf{\Sigma}_{\mathbf{I}}\mathbf{\Sigma}_{\mathbf{I}}^*)^{-1}(t,\mathbf{u}_{\mathbf{I}})(\mathbf{R}_{\mathbf{II}}\mathbf{A}_1^*(t,\mathbf{u}_{\mathbf{I}}))^*\big\}\, dt. \end{aligned} \tag{3}$$

In most geophysical and engineering turbulent dynamical systems, the nonlinear terms such as the nonlinear advection have quadratic forms and these quadratic nonlinear interactions conserve energy [31, 46, 52, 41, 55, 56]. The nonlinear interactions allow energy transfer between different scales that induces intermittent instabilities in the turbulent dynamical systems. Such instabilities are then mitigated by energy-conserving quadratic nonlinear interactions that transfer energy back to the linearly stable modes where it is dissipated, resulting in a statistical steady state. Note that the nonlinear turbulent systems without the energy-conserving nonlinear interactions may suffer from non-physical finite-time blow up of statistical solutions and pathological behavior of the related invariant measure [58]. Mathematically, the turbulent dynamical systems with energy-conserving quadratic nonlinear interactions have the following abstract forms:

$$d\mathbf{u} = \big[-\Lambda\mathbf{u} + \mathbf{B}(\mathbf{u},\mathbf{u}) + \mathbf{F}(t)\big]dt + \mathbf{\Sigma}(t,\mathbf{u})d\mathbf{W}(t), \tag{4}$$

where $-\Lambda = \mathbf{L} + \mathbf{D}$. Here, $\mathbf{L}$ is a skew-symmetric linear operator that can represent the β effect of Earth's curvature and topography, while $\mathbf{D}$ is a negative definite symmetric operator representing dissipative processes such as surface drag, radiative damping and viscosity, etc [67, 72, 45, 74]. The quadratic operator $\mathbf{B}(\mathbf{u},\mathbf{u})$ conserves energy by itself so that it satisfies the following:

$$\mathbf{u}\cdot\mathbf{B}(\mathbf{u},\mathbf{u}) = 0. \tag{5}$$

Notably, a rich class of turbulent models with energy-conserving quadratic nonlinear interactions in (4) belongs to the conditional Gaussian systems (1), including the noisy version of Lorenz 63 model [40], the reduced stochastic climate model [49, 42], the nonlinear triad model mimicking structural features of low-frequency variability of GCMs with non-Gaussian features [48], the modified conceptual dynamical model for turbulence [53], and the two-layer Lorenz 96 model [37]. See [14] and its appendix for a general framework of conditional Gaussian systems with energy-conserving nonlinear interactions as well as concrete examples.

2.2. The efficient statistically accurate algorithms for solving the PDFs of the conditional Gaussian systems. Assume the dimension $N_{\mathbf{I}}$ of the observed variables is low, while the dimension $N_{\mathbf{II}}$ of the unobserved variables can be high. This is the typical scenario in most turbulent dynamical systems, where the low-dimensional variables $\mathbf{u}_{\mathbf{I}}$ represent large scales or resolved variables while the high-dimensional ones $\mathbf{u}_{\mathbf{II}}$ stand for the unresolved and unobserved variables [53, 41].

Below, we summarize the procedures of the efficient statistical algorithms developed in [14]. First, we generate L independent trajectories from the stochastic dynamical systems (1). In fact, the only information that is required for these algorithms is L independent trajectories of the observed variables, namely $\mathbf{u}_{\mathbf{I}}^1(s\le t),\dots,\mathbf{u}_{\mathbf{I}}^L(s\le t)$. Then, different strategies are used to deal with the observed variables $\mathbf{u}_{\mathbf{I}}$ and unobserved variables $\mathbf{u}_{\mathbf{II}}$, respectively. The PDF of $\mathbf{u}_{\mathbf{II}}$ is estimated via a parametric method that exploits the closed form of the conditional Gaussian posterior statistics (3),

$$p(\mathbf{u}_{\mathbf{II}}(t)) = \lim_{L\to\infty}\frac{1}{L}\sum_{i=1}^{L} p(\mathbf{u}_{\mathbf{II}}(t)|\mathbf{u}_{\mathbf{I}}^i(s\le t)). \tag{6}$$

Note that the limit $L\to\infty$ in (6) (as well as (7) and (9) 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 $\mathbf{u}_{\mathbf{I}}$,

$$p\big(\mathbf{u}_{\mathbf{I}}(t)\big) = \lim_{L\to\infty}\frac{1}{L}\sum_{i=1}^{L} K_{\mathbf{H}}\Big(\mathbf{u}_{\mathbf{I}}(t)-\mathbf{u}_{\mathbf{I}}^i(t)\Big), \tag{7}$$

where $\mathbf{H}=\mathbf{H}(t)$ is the bandwidth matrix, and $K_{\mathbf{H}}(\cdot)$ is a Gaussian kernel centered at each sample point with covariance $\mathbf{H}(t)$,

$$K_{\mathbf{H}}\Big(\mathbf{u}_{\mathbf{I}}(t)-\mathbf{u}_{\mathbf{I}}^i(t)\Big) \sim \mathcal{N}\Big(\mathbf{u}_{\mathbf{I}}^i(t),\mathbf{H}(t)\Big). \tag{8}$$

Below, we simply use $\mathbf{H}$ to represent the bandwidth at time t for the notation simplicity.

The kernel density estimation algorithm here involves a "solve-the-equation plug-in" approach for optimizing the bandwidth, the idea of which was originally proposed in [4]. The solve-the-equation approach does not impose any requirement for the profile of the underlying PDF. Therefore, it works for the non-Gaussian cases and the computational cost comes from numerically solving a scalar high order algebraic equation for the optimal bandwidth in order to minimize the asymptotic mean integrated squared error (AMISE) in the estimator. Furthermore, we adopt a diagonal matrix for $\mathbf{H}$. This greatly reduces the computational costs while remains the results with reasonable accuracy. Note that in the limit $L \to \infty$, the kernel density method is simply the Monte Carlo simulation, where the bandwidth shrinks to zero.

Finally, with (6) and (7) in hand, a hybrid method is applied to solve the joint PDF of $\mathbf{u_I}$ and $\mathbf{u_{II}}$ through a Gaussian mixture,

$$p(\mathbf{u_I}(t), \mathbf{u_{II}}(t)) = \lim_{L\to\infty} \frac{1}{L} \sum_{i=1}^{L} \Big(K_{\mathbf{H}}(\mathbf{u_I}(t) - \mathbf{u_I}^i(t)) \cdot p(\mathbf{u_{II}}(t)|\mathbf{u_I}^i(s \le t)) \Big). \tag{9}$$

One important features of these algorithms is that the solutions of both the two marginal distributions in (6) and (7) and the joint distribution in (9) are consistent with those of solving the Fokker-Planck equation for $p(\mathbf{u_{II}}(t)), p(\mathbf{u_I}(t))$ and $p(\mathbf{u_I}(t), \mathbf{u_{II}}(t))$, respectively.

Practically, $L \sim O(100)$ is sufficient for the efficient hybrid method (9) to solve the joint PDF with $N_\mathbf{I} \le 3$ and $N_\mathbf{II} \sim 10$ while an order of $O(10^6)$ samples is required for solving the joint PDF using classical Monte Carlo methods to reach the same accuracy for a 6 dimensional turbulent system [14]. Since L is only of order $O(100)$, the L independent trajectories $\mathbf{u_I}^1(s \le t), \ldots, \mathbf{u_I}^L(s \le t)$ can be obtained by running a Monte Carlo simulation for the coupled system (1) with L samples, which is computationally affordable. In addition, the closed form of the L conditional distributions in (6) can be computed in a parallel way due to their independence, which further reduces the computational cost. See [14] for more details.

3. Main theoretical results.

The rigorous analysis of the efficient statistically accurate algorithms involving the hybrid strategy (9) is studied in this section. For comparison, the theoretical results by applying the kernel density estimation method to the full system (1) is also illustrated. Note that the kernel density estimation is essentially the Monte Carlo simulation when L is large and therefore it suffers from the curse of dimensionality. Such comparison facilitates the understanding of the advantages of the efficient algorithm (9) in recovering the high-dimensional subspace of $\mathbf{u_{II}}$ using only a small number of samples. Below, $p_t(\mathbf{u_I}, \mathbf{u_{II}})$ represents the true PDF while $\tilde{p}_t(\mathbf{u_I}, \mathbf{u_{II}})$ and $\hat{p}_t(\mathbf{u_I}, \mathbf{u_{II}})$ stand for the recovered PDFs based on the pure kernel density estimation and the efficient hybrid method (9), respectively.

Kernel density estimation for the joint PDF.

$$\tilde{p}_t(\mathbf{u_I}, \mathbf{u_{II}}) = \frac{1}{L} \sum_{i=1}^{L} K_H((\mathbf{u_I}, \mathbf{u_{II}}) - (\mathbf{u_I}^i(t), \mathbf{u_{II}}^i(t))), \tag{10}$$

$$\text{with} \quad K_H(\mathbf{u_I}, \mathbf{u_{II}}) = (2\pi H)^{-\frac{N_\mathbf{I}+N_\mathbf{II}}{2}} \exp\left(-\frac{1}{2H} \sum_{i=1}^{N_\mathbf{I}} c_i^2 \mathbf{u}_{\mathbf{I},i}^2 - \frac{1}{2H} \sum_{i=1}^{N_\mathbf{II}} c_{i+N_\mathbf{I}}^2 \mathbf{u}_{\mathbf{II},i}^2 \right). \tag{11}$$

Hybrid method — kernel density estimation for $\mathbf{u_I}$ *and conditional Gaussian mixture for* $\mathbf{u_{II}}$.

$$\hat{p}_t(\mathbf{u_I}, \mathbf{u_{II}}) = \frac{1}{L}\sum_{i=1}^{L} K_H(\mathbf{u_I} - \mathbf{u_I}^i(t))p(\mathbf{u_{II}}|\mathbf{u_I}^i(s \le t)). \tag{12}$$

$$\text{with} \quad K_H(\mathbf{u_I}) = (2\pi H)^{-\frac{N_\mathbf{I}}{2}} \exp\left(-\frac{1}{2H}\sum_{i=1}^{N_\mathbf{I}} c_i^2 \mathbf{u}_{\mathbf{I},i}^2\right). \tag{13}$$

In (11) and (13), we let $\mathbf{H} = H\mathbf{C}$ as in (9). The scalar H is the scale of the bandwidth [68, 76, 77, 4] and c_i^2 are the diagonal terms of $\mathbf{C}$ such that $c_i^2 H$ represents the bandwidth in one direction. In the following, we mostly concern the performance of $\tilde{p}_t$ and $\hat{p}_t$ when L is large.

One standard metric to measure the performance of a density estimator is the mean integrated squared error (MISE). The MISE of the hybrid method, for example, is the average L^2 distance to the true density:

$$\text{MISE} = \mathbb{E}\int |p_t(\mathbf{u_I}, \mathbf{u_{II}}) - \hat{p}_t(\mathbf{u_I}, \mathbf{u_{II}})|^2 d\mathbf{u_I} d\mathbf{u_{II}}.$$

Note that $\hat{p}_t$ relies on the realization of the samples and therefore it is natural to take the expectation of the distance.

Applying the Bias-Variance decomposition [25] to the MISE yields

$$\text{MISE} = \underbrace{\mathbb{E}\int |\hat{p}_t(\mathbf{u_I}, \mathbf{u_{II}}) - \bar{p}_t(\mathbf{u_I}, \mathbf{u_{II}})|^2 d\mathbf{u_I} d\mathbf{u_{II}}}_{\text{Bias}} + \underbrace{\int |p_t(\mathbf{u_I}, \mathbf{u_{II}}) - \bar{p}_t(\mathbf{u_I}, \mathbf{u_{II}})|^2 d\mathbf{u_I} d\mathbf{u_{II}}}_{\text{Variance}}, \tag{14}$$

where $\bar{p}_t := \mathbb{E}\hat{p}_t$. The variance part comes from the sampling error of the method and the bias part comes from the usage of the kernel method. See (28) for a direct proof of this decomposition.

The MISE and its decomposition (14) will be used to understand the performance of the two density estimation methods in (10) and (12), where the scenarios with a large number of samples and a large dimension of the variables $N_\mathbf{II}$ are of particular interest. Main results are presented below and the rigorous proofs of these results are shown in section 4. Note that despite quite a few studies of kernel density estimation, especially in the asymptotic limit, exist in literature [68, 76, 77, 33, 4], no analysis has been established for the hybrid method (12). Moreover, the results here are all non-asymptotic, and therefore they hold for arbitrary choice of bandwidth parameters. This is important in practice, as the bandwidth matrix $\mathbf{H}(t)$ may change with t.

3.1. MISE of the hybrid method.

The main result of our analysis is the following:

Theorem 3.1. *The two parts of MISE in* (14) *for the hybrid method* (12) *are bounded:*

(15)
$$\hat{p}_t \ \textit{Variance} \leq \frac{1}{L}\mathbb{E}\left(\prod_{i=1}^{N_{\mathbf{I}}}(\pi H c_i^2)\,det(\pi \mathbf{R}_{\mathbf{II}}(t))\right)^{-\frac{1}{2}},$$
$$\hat{p}_t \ \textit{Bias} \leq \frac{1+\delta}{4}H^2 J\left(\sum_{i=1}^{N_{\mathbf{I}}} c_i^2 \partial^2_{\mathbf{u}^2_{\mathbf{I},i}} p_t(\mathbf{u_I},\mathbf{u_{II}})\right) + \frac{1+\delta^{-1}}{2}M^2H^3\left(\sum_{i=1}^{N_{\mathbf{I}}} c_i^2\right)^3 J(M(\mathbf{u_I},\mathbf{u_{II}})).$$

Here δ is any fixed strictly positive number. $\mathbb{E}$ is the statistical average. $J(f(\mathbf{u_I},\mathbf{u_{II}}))$ denotes the integral $\int f^2(\mathbf{u_I},\mathbf{u_{II}})d\mathbf{u_I}d\mathbf{u_{II}}$. The function $M(\mathbf{u_I},\mathbf{u_{II}})$ is an upper bound of the third order directional derivative of p_t in the direction of $\mathbf{u_I}$ around $(\mathbf{u_I},\mathbf{u_{II}})$. That is, we assume

$$\left|\frac{d^3}{ds^3}p_t(\mathbf{u_I}+s\mathbf{v},\mathbf{u_{II}})\right| \leq M(\mathbf{u_I},\mathbf{u_{II}}), \quad \textit{for all } \mathbf{v}\in\mathbb{R}^{N_{\mathbf{I}}}, |\mathbf{v}|\leq 1. \tag{16}$$

In a practical scenario, as the sample size L increases, bandwidth H can decrease, so that both the variance and bias terms decrease to zero. By taking δ close to zero and ignoring the higher order term in the bias upper bound, we recover an upper bound similar to the asymptotic MISE (AMISE) in [76, Eqn. (2.6)], except that our method also consists a random component of $\mathbf{R_{II}}(t)$:

$$\text{AMISE} \leq \frac{1}{L}\mathbb{E}\left(\prod_{i=1}^{N_{\mathbf{I}}}(\pi H c_i^2)\det(\pi \mathbf{R}_{\mathbf{II}}(t))\right)^{-\frac{1}{2}} + \frac{1}{4}H^2 J\left(\sum_{i=1}^{N_{\mathbf{I}}} c_i^2 \partial^2_{\mathbf{u}^2_{\mathbf{I},i}} p_t(\mathbf{u_I},\mathbf{u_{II}})\right), \tag{17}$$

where the two terms on the right hand side represents the variance and bias, respectively. It is natural to equate the order of these two terms, that is letting $LH^{-\frac{1}{2}N_{\mathbf{I}}} \sim O(H^2)$. This leads to the common choice of the bandwidth [33]

$$H \sim O\left(L^{-\frac{2}{4+N_{\mathbf{I}}}}\right) \quad \text{and consequentially} \quad \text{MISE} \sim O\left(L^{-\frac{4}{4+N_{\mathbf{I}}}}\right). \tag{18}$$

Notably, the variance part of MISE in (17) depends on $\mathbf{u_{II}}$ through $\mathbb{E}\sqrt{\det(\pi\mathbf{R_{II}}(t))}^{-1}$, which indicates that the hybrid method in (12) performs better with a larger $\mathbf{R_{II}}(t)$. This is consistent with the intuition that a large $\mathbf{R_{II}}(t)$ corresponds to a conditional distribution $\mathcal{N}(\bar{\mathbf{u}}_{\mathbf{II}}(t),\mathbf{R_{II}}(t))$ with a wide band that is able to recover a sufficient portion of the PDF.

3.2. Comparison between the two density estimators. Theorem 3.1 already reveals the advantage of the hybrid method (12) over the the direct kernel density method (10). For a qualitative comparison of the two methods, we can view the latter as a trivial application of the hybrid method by taking $\mathbf{u}'_{\mathbf{I}} = (\mathbf{u_I},\mathbf{u_{II}})$ and $\mathbf{u}'_{\mathbf{II}} = \emptyset$, and therefore $\mathbf{u}'_{\mathbf{II}}$ is trivially linear conditioned on $\mathbf{u}'_{\mathbf{I}}$. A direct application of Theorem 3.1 leads to

(19)
$$\tilde{p}_t \ \text{Variance} \leq \frac{1}{L}\mathbb{E}\left(\prod_{i=1}^{N_{\mathbf{I}}+N_{\mathbf{II}}} \pi H c_i^2\right)^{-\frac{1}{2}},$$
$$\tilde{p}_t \ \text{Bias} \leq \frac{(1+\delta)H^2}{4}J\left(\sum_{i=1}^{N_{\mathbf{I}}} c_i^2\partial^2_{\mathbf{u}^2_{\mathbf{I},i}}p_t + \sum_{i=1}^{N_{\mathbf{II}}} c^2_{i+N_{\mathbf{I}}}\partial^2_{\mathbf{u}^2_{\mathbf{II},i}}p_t\right) + \frac{(1+\delta^{-1})H^3}{2}\left(\sum_{i=1}^{N_{\mathbf{I}}} c_i^2\right)^3 J(\widetilde{M}).$$

where $\widetilde{M} \geq M$ is the upper bound for third order directional derivative in $\mathbb{R}^{N_{\mathbf{I}}+N_{\mathbf{II}}}$ of p_t. Similar results in the asymptotic setting can be found in [76].

If we use the same bandwidth H and sample size L in both method, Comparing (19) with (15), we find that

$$\tilde{p}_t \text{ Bias bound} \geq \hat{p}_t \text{ Bias bound},$$

and moreover

$$\frac{\tilde{p}_t \text{ Variance bound}}{\hat{p}_t \text{ Variance bound}} = \frac{H^{-\frac{N_{\mathbf{II}}}{2}} \prod_{i=1}^{N_{\mathbf{II}}} c_{i+N_{\mathbf{I}}}}{\mathbb{E}\sqrt{\det(\mathbf{R}_{\mathbf{II}}(t))}^{-1}}.$$

Practically, a large L is chosen to guarantee the accuracy of the recovered PDFs, which corresponds to a small bandwidth H. Then the variance part of the direct kernel method is several magnitudes larger than that of hybrid method, especially when the dimension $N_{\mathbf{II}}$ is high.

As discussed above, one would optimize the choice of H such that the two quantities in (19) are of the same order, which leads to the scaling $H \sim O\left(L^{-\frac{2}{4+N_{\mathbf{I}}+N_{\mathbf{II}}}}\right)$, and also the overall MISE $\sim O\left(L^{-\frac{4}{4+N_{\mathbf{I}}+N_{\mathbf{II}}}}\right)$, However, This is much worse than the MISE associated with the conditional Gaussian method (18) when $N_{\mathbf{II}}$ is large. Alternatively, if one wants the performance of the direct kernel method to be the same as the conditional Gaussian one (18), then the sample size needs to increase to $\widetilde{L} = L^{\frac{4+N_{\mathbf{I}}+N_{\mathbf{II}}}{4+N_{\mathbf{I}}}}$, which can be many magnitudes larger than L.

In conclusion, direct application of the kernel method suffers from the curse of dimensionality. This is due to the fact that the variance scales with the bandwidth as $H^{-\frac{N_{\mathbf{I}}+N_{\mathbf{II}}}{2}}$, and therefore one needs to increase sample size exponentially with the dimension in order to have a small bandwidth that guarantees the accuracy of the recovered PDFs. However, when H is small, the kernel density method approximates the standard Monte Carlo simulation, which suffers from the curse of dimensionality. On the other hand, the hybrid method resolves this issue by estimating the $\mathbf{u}_{\mathbf{II}}$ part using a parametric method where the bandwidth (or the covariance) does not depend on L. Therefore, the performance of the hybrid method (12) can be much superior than the direct kernel method (10) when $N_{\mathbf{II}}$ is large.

3.3. Marginal distribution of $\mathbf{u}_{\mathbf{II}}(t)$. There are scenarios where the focus is only on estimating the density of $\mathbf{u}_{\mathbf{II}}(t)$. Again, both methods can be applied here. The direct kernel method (10) results in the estimation of the marginal density

(20)

$$\tilde{p}_t(\mathbf{u}_{\mathbf{II}}) := \frac{1}{L}\sum_{i=1}^{L} K_H(\mathbf{u}_{\mathbf{II}} - \mathbf{u}_{\mathbf{II}}^i(t)), \quad K_H(\mathbf{u}_{\mathbf{II}}) = (2\pi H)^{-\frac{N_{\mathbf{II}}}{2}} \exp\left(-\frac{1}{2H}\sum_{i=1}^{N_{\mathbf{II}}} c_{i+N_{\mathbf{I}}}^2 \mathbf{u}_{\mathbf{II},i}^2\right).$$

On the other hand, the hybrid method (12) simply becomes a conditional Gaussian mixture method which contains no kernel density estimation

$$\hat{p}_t(\mathbf{u}_{\mathbf{II}}) := \frac{1}{L}\sum_{i=1}^{L} p(\mathbf{u}_{\mathbf{II}}|\mathbf{u}_{\mathbf{I}}^i(s \leq t)). \tag{21}$$

It is straightforward to check these density estimators are the marginal PDFs of the joint distributions in (10) and (12).

Since there is no kernel involved for the conditional Gaussian method in (21), the MISE has a simple bound without the bias part:

Proposition 3.2. *The marginal MISE of the conditional Gaussian estimator in* (21) *is bounded as*

$$\hat{p}_t \ \mathit{MISE} \leq \frac{1}{L}\mathbb{E}\left(det(\pi \mathbf{R}_{\mathbf{II}}(t))\right)^{-\frac{1}{2}}. \tag{22}$$

Following the derivation of (19), the MISE of the direct kernel method in (20) is given by

$$\tilde{p}_t \text{ MISE} \leq \frac{1}{L}\mathbb{E}\left(\prod_{i=1}^{N_{\mathbf{II}}}\pi H c^2_{i+N_{\mathbf{I}}}\right)^{-\frac{1}{2}} + \frac{(1+\delta)H^2}{4}J\left(\sum_{i=1}^{N_{\mathbf{II}}} c^2_{i+N_{\mathbf{I}}}\partial^2_{\mathbf{u}^2_{\mathbf{II},i}} p_t\right) + \frac{(1+\delta^{-1})H^3}{2}\left(\sum_{i=1}^{N_{\mathbf{I}}} c_i^2\right)^3 J(\widetilde{M}).$$

With the optimal choice $H \sim O\left(L^{-\frac{2}{4+N_{\mathbf{II}}}}\right)$, the direct kernel method MISE $\sim O\left(L^{-\frac{4}{4+N_{\mathbf{II}}}}\right)$. The hybrid method with the conditional Gaussian mixture is clearly superior for marginal density estimation, as its MISE (22) is essentially $O(L^{-1})$, and the bandwidth H has no dependence on L.

3.4. Fixed subspace. In many scenarios, only a part of $\mathbf{u}_{\mathbf{II}}$ is of practical interest. To this end, we consider here $\mathbf{u}^P_{\mathbf{II}} = \mathbf{P}\mathbf{u}_{\mathbf{II}}$, where $\mathbf{P} : \mathbb{R}^{N_{\mathbf{II}}} \mapsto \mathbb{R}^{N^P_{\mathbf{II}}}$ maps $\mathbf{u}_{\mathbf{II}}$ onto a lower dimensional subspace. Below, we study the estimation of the density $p^P_t(\mathbf{u}_{\mathbf{I}}, \mathbf{u}^P_{\mathbf{II}})$ of $(\mathbf{u}_{\mathbf{I}}(t), \mathbf{u}^P_{\mathbf{II}}(t))$ using the hybrid method.

It is straightforward to show the conditional distribution of $\mathbf{u}^P_{\mathbf{II}}(t)$ given $\mathbf{u}_{\mathbf{I}}(s\leq t)$ follows the Gaussian density $p(\mathbf{u}^P_{\mathbf{II}}|\mathbf{u}_{\mathbf{I}}(s\leq t))$ of the following form

$$\det(2\pi\mathbf{P}\mathbf{R}_{\mathbf{II}}(t)\mathbf{P}^*)^{-\frac{1}{2}}\exp\left(-\tfrac{1}{2}(\mathbf{u}^P_{\mathbf{II}} - \mathbf{P}\bar{\mathbf{u}}_{\mathbf{II}}(t))^*[\mathbf{P}\mathbf{R}_{\mathbf{II}}(t)\mathbf{P}^*]^{-1}(\mathbf{u}^P_{\mathbf{II}} - \mathbf{P}\bar{\mathbf{u}}_{\mathbf{II}}(t))\right).$$

The density of $(\mathbf{u}_{\mathbf{I}}(t), \mathbf{u}^P_{\mathbf{II}}(t))$ can be estimated by

$$\hat{p}^P_t(\mathbf{u}_{\mathbf{I}}, \mathbf{u}^P_{\mathbf{II}}) = \frac{1}{L}\sum_{i=1}^{L} K_H(\mathbf{u}_{\mathbf{I}} - \mathbf{u}^i_{\mathbf{I}}(t))p(\mathbf{u}^P_{\mathbf{II}}|\mathbf{u}^i_{\mathbf{I}}(s\leq t)).$$

Following Theorem 3.1, we can show that

Corollary 3.3. *Under the same assumption as in Theorem 3.1, the MISE decomposition of $\hat{p}^P_t$ has the following two bounds*

$$\hat{p}^P_t \ \mathit{Variance} \leq \frac{1}{L}\mathbb{E}\left(\prod_{i=1}^{N_{\mathbf{I}}}(\pi H c_i^2)\, det(\pi\mathbf{P}\mathbf{R}_{\mathbf{II}}(t)\mathbf{P}^*)\right)^{-\frac{1}{2}},$$

$$\hat{p}^P_t \ \mathit{Bias} \leq \frac{1+\delta}{4}H^2 J\left(\sum_{i=1}^{N_{\mathbf{I}}} c_i^2\partial^2_{\mathbf{u}^2_{\mathbf{I},i}} p^P_t(\mathbf{u}_{\mathbf{I}}, \mathbf{u}^P_{\mathbf{II}})\right) + \frac{1+\delta^{-1}}{2}H^3\left(\sum_{i=1}^{N_{\mathbf{I}}} c_i^2\right)^3 J(M^P(\mathbf{u}_{\mathbf{I}}, \mathbf{u}^P_{\mathbf{II}})),$$

where M^P is a upper bound of third order derivative of p_t^P in $\mathbf{u_I}$, as in (16).
Notably, the variance term depends only on $\mathbb{E}\sqrt{\det(\pi \mathbf{P}\mathbf{R_{II}}(t)\mathbf{P}^*)}^{-1}$, where $\mathbf{P}\mathbf{R_{II}}(t)\mathbf{P}^*$ is a $N_{\mathbf{II}}^P \times N_{\mathbf{II}}^P$ matrix that is independent of the components complementary to $\mathbf{u}_{\mathbf{II}}^P(t)$. In other words, the performance of the hybrid estimator on a certain part of the components is independent of the other components. This is particularly useful when $N_{\mathbf{II}}^P$ is small. Note that such a property also holds for the direct kernel method but in practice the kernel method works only for the case when $N_{\mathbf{II}}$ is small.

3.5. Controllability and a lower bound of $\mathrm{R_{II}}$. According to Theorem 3.1, $\mathbf{R_{II}}(t)$ controls the sampling variance term in the MISE. Therefore, it is desirable to derive a lower bound for $\mathbf{R_{II}}(t)$. Note that in the conditional Gaussian system (1), $\mathbf{u_I}$ can be interpreted as an observation of $\mathbf{u_{II}}$, and $p(\mathbf{u_{II}}|\mathbf{u_I}(s\le t))$ is essentially the optimal Kalman filter with covariance $\mathbf{R_{II}}(t)$. Therefore, a lower bound of $\mathbf{R_{II}}(t)$ can be guaranteed by the controllability of the associated signal-observation system. In short, the controllability condition ensures the noise in the system is regular enough such that the optimal filter is not accurate to a singular degree in any component. More discussions on the controllability of Kalman filters can be found in [18, 21, 51]. A recent work [2] has summarized some of the major results in this area. It is noteworthy that since the term $\mathbf{a}_1$ depends on realization of $\mathbf{u_I}$, both the controllability condition and the lower bounds rely on the realization of $\mathbf{u_I}$.

In our context, a standard way to characterize this notion is the following assumption:

Assumption 3.4. *Let $\mathcal{E}_{s,t}$ be the matrix flow generated by $\mathbf{a}_1$:*

$$\frac{d}{dt}\mathcal{E}_{s,t} = \mathbf{a}_1(t, \mathbf{u_I}(t))\mathcal{E}_{s,t}, \quad \mathcal{E}_{s,s} = I_{N_{\mathbf{II}}}.$$

Suppose there are constants $v > 0, m \ge 0$ and $D_c \ge 1$ such that for any $t \ge v$ and $s \in [t-v, t]$,

$$D_c^{-1} I_{N_{\mathbf{II}}} \preceq \mathcal{E}_{s,t}\mathcal{E}_{s,t}^* \preceq D_c I_{N_{\mathbf{II}}}, \quad \sigma_{\mathbf{II},-}^2 I_{N_{\mathbf{II}}} \preceq \mathbf{\Sigma_{II}^*\Sigma_{II}} \preceq \sigma_{\mathbf{II},+}^2 I_{N_{\mathbf{II}}},$$

$$\mathbf{A}_1^*(t, \mathbf{u_I}(t))[\mathbf{\Sigma_I\Sigma_I^*}]^{-1}\mathbf{A}_1(t, \mathbf{u_I}(t)) \preceq D_c(|\mathbf{u_I}(t)|^{2m} + 1) I_{N_{\mathbf{II}}}.$$

Throughout this paper, for two real symmetric matrices A and B, we use $A \preceq B$ to indicate that $B - A$ is a positive semi-definite matrix.
While $\mathbf{A}_1^*(\mathbf{\Sigma_I\Sigma_I^*})^{-1}\mathbf{A}_1$ actually concerns of observability, this bound is very mild. Thus, we still call Assumption 3.4 the controllability condition.

Proposition 3.5. *Suppose $N_{\mathbf{II}} \ge 2$, and the controllability condition, Assumption 3.4 holds, then for any $t \ge v$, $\mathbf{R_{II}}(t) \succeq h_{t,v}^{-1}(\mathbf{u_I}) I_{N_{\mathbf{II}}}$, where*

$$h_{t,v}(\mathbf{u_I}) := v^2\sigma_{\mathbf{II},+}^2\sigma_{\mathbf{II},-}^{-2}D_c^6\left(v + \int_{t-v}^t |\mathbf{u_I}(r)|^{2m}dr\right) + v^{-1}D_c\sigma_{\mathbf{II},-}^{-2}.$$

In particular there are constants D_1 and D_2 such that

$$\mathbb{E}\sqrt{det\mathbf{R_{II}}(t)}^{-1} \le D_1 + D_2\int_{t-v}^t \mathbb{E}|\mathbf{u_I}(r)|^{mN_{\mathbf{II}}}dr.$$

The dependence of $\mathbf{R_{II}}(t)$ on $\mathbf{u_I}(s)|_{t-v\le s\le t}$ comes from the observational term $\mathbf{A}_1$. As is seen from (3), if $\mathbf{A}_1^*(\mathbf{\Sigma_I}\mathbf{\Sigma_I}^*)^{-1}\mathbf{A}_1$ is large, $\mathbf{R_{II}}(t)$ has a large quadratic damping, which can bring it to a very low level.

In symmetry, an upper bound can be derived if a lower bound of $\mathbf{A}_1^*(\mathbf{\Sigma_I}\mathbf{\Sigma_I}^*)^{-1}\mathbf{A}_1$ is assumed. Furthermore, one can show that the Riccati flow of $\mathbf{R_{II}}(t)$ is contractive, so its dependence on $\mathbf{R_{II}}(0)$ is diminishing. Since these results are not directly related to the performance of the hybrid estimator, we put them in the supplementary material along with the verification of Proposition 3.5.

3.6. Long time performance. The simulation of $(\mathbf{u}_\mathbf{I}^i(t), \mathbf{u}_\mathbf{II}^i(t))$ can be maintained continuously, and the conditional Gaussian density estimator (12) can be applied for an online estimation. One important question to ask is whether the performance, and in particular the MISE, degenerates with time. If this is the case, additional samples are needed to reinforce the estimation, which is however usually difficult to carry out in practice. In this subsection, we show that the conditional Gaussian density estimator has a long time stable performance, as long as the joint process $(\mathbf{u_I}, \mathbf{u_{II}})$ is stable and ergodic.

In stochastic analysis, the stability and ergodicity of a process can be guaranteed by energy dissipation and non-degenerate stochastic forcing. For our purpose, we can assume the energy is dissipative, while the noise is elliptic [57].

Assumption 3.6. *Suppose $\mathbf{\Sigma_I}$ and $\mathbf{\Sigma_{II}}$ are full rank, and the energy is dissipative with a rate $\rho>0$ and a constant D_e*

$$\mathbf{u_I}\cdot(\mathbf{A}_0+\mathbf{A}_1\mathbf{u_{II}})+\mathbf{u_{II}}\cdot(\mathbf{a}_0+\mathbf{a}_1\mathbf{u_{II}})\le -\rho(|\mathbf{u_I}|^2+|\mathbf{u_{II}}|^2)+D_e. \tag{23}$$

Theorem 3.7. *Under Assumption 3.6, the following hold.*
1) The joint density p_t converges geometrically to an ergodic measure p_∞ with a rate $c>0$. In particular, there is a constant D_0 so that

$$\int\left|\frac{p_t}{p_\infty}(\mathbf{u_I},\mathbf{u_{II}})-1\right|^2 p_\infty(\mathbf{u_I},\mathbf{u_{II}})d\mathbf{u_I}d\mathbf{u_{II}}\le D_0e^{-ct}\langle|\mathbf{u}|^2+1,p_0\rangle\left\|\frac{p_0}{p_\infty}-1\right\|_\infty^2. \tag{24}$$

Here $\langle|\mathbf{u}|^2+1,p_0\rangle$ denotes the quantity $\int(|\mathbf{u_I}|^2+|\mathbf{u_{II}}|^2+1)p_0(\mathbf{u_I},\mathbf{u_{II}})d\mathbf{u_I}d\mathbf{u_{II}}$, and $\|f\|_\infty$ denotes the supremum $\|f\|_\infty=\sup_{\mathbf{u_I},\mathbf{u_{II}}}|f(\mathbf{u_I},\mathbf{u_{II}})|$.
2) Suppose Assumption 3.4 also holds, then for any $t>0$ and $\delta>0$, $N_\mathbf{II}\ge 2$, the two parts of the MISE using the hybrid method are bounded by

$$\hat p_t\ \textit{Variance}\le\frac{D_{m,N_\mathbf{II},v}}{L\pi^{\frac{N_\mathbf{I}+N_\mathbf{II}}{2}}H^{\frac{N_\mathbf{I}}{2}}\prod_{i=1}^{N_\mathbf{I}}c_i}\left(\exp(-\tfrac12\rho mN_\mathbf{II}t)\mathbb{E}|\mathbf{u}(0)|^{mN_\mathbf{II}}+D_{m,N_\mathbf{II},v}\right),$$

$$\begin{aligned}\hat p_t\ \textit{Bias}\le&\frac{(1+\delta)^2}{4}H^2J\left(\sum_{i=1}^{N_\mathbf{I}}c_i^2\partial^2_{\mathbf{u}^2_{\mathbf{I},i}}p_\infty(\mathbf{u_I},\mathbf{u_{II}})\right)+\frac{(1+\delta)^2}{2\delta}H^3\left(\sum_{i=1}^{N_\mathbf{I}}c_i^2\right)^3J(M_\infty(\mathbf{u_I},\mathbf{u_{II}}))\\&+8(1+\delta^{-1})D_0e^{-ct}\langle|\mathbf{u}|^2+1,p_0\rangle\left\|\tfrac{p_0}{p_\infty}-1\right\|_\infty^2\|p_\infty\|_\infty,\end{aligned}$$

where $D_{m,N_\mathbf{II},v}$ is a constant independent of L and H, and M_∞ is a bound for the third order $\mathbf{u_I}$-directional derivative of p_∞ as in (16).

In particular, when $t \to \infty$, we have

$$\limsup_{t\to\infty} \text{MISE} \le \frac{D^2_{m,N_{\mathbf{II}},v}}{L\pi^{\frac{N_{\mathbf{I}}+N_{\mathbf{II}}}{2}} H^{\frac{N_{\mathbf{I}}}{2}} \prod_{i=1}^{N_{\mathbf{I}}} c_i} + \frac{(1+\delta)^2}{4} H^2 J\left(\sum_{i=1}^{N_{\mathbf{I}}} c_i^2 \partial^2_{\mathbf{u}^2_{\mathbf{I},i}} p_\infty(\mathbf{u}_{\mathbf{I}}, \mathbf{u}_{\mathbf{II}})\right)$$
$$+ \frac{(1+\delta)^2}{2\delta} H^3 \left(\sum_{i=1}^{N_{\mathbf{I}}} c_i^2\right)^3 J(M_\infty(\mathbf{u}_{\mathbf{I}}, \mathbf{u}_{\mathbf{II}})).$$

This leads to the same bandwidth and MISE scaling with L, namely:

$$H \sim O\left(L^{-\frac{2}{4+N_{\mathbf{I}}}}\right) \quad \text{and} \quad \text{MISE} \sim O\left(L^{-\frac{4}{4+N_{\mathbf{I}}}}\right).$$

The proof strategy of Theorem 3.7 is straightforward. The first part is simply corollaries of [60, 59, 1]. To reach a bound on the variance part in 2), it suffices to have a lower bound on $\mathbb{E}\sqrt{\det \mathbf{R}_{\mathbf{II}}(t)}^{-1}$. This can be achieved by Proposition 3.5 and an energy dissipation argument. For the bias term, we use the Poincaré inequality (24) to approximate it with the bias term at equilibrium.

3.7. Conditional Gaussian turbulent dynamical systems with energy-conserving quadratic nonlinearity.

Recall the turbulence model $\mathbf{u}$ with quadratic energy conserving nonlinear interactions (4)–(5)

$$d\mathbf{u} = -\Lambda \mathbf{u} dt + \mathbf{B}(\mathbf{u}, \mathbf{u})dt + \mathbf{F}dt + \mathbf{\Sigma} d\mathbf{W}_t.$$

The linear damping part provides a uniform dissipation, so for some $\lambda_- > 0$,

$$\mathbf{u} \cdot \Lambda \mathbf{u} \ge \lambda_- |\mathbf{u}|^2,$$

and the nonlinearity term $\mathbf{B}$ is quadratic and conserves energy.

In our conditional Gaussian setup, we can decompose the dynamics into the form below

$$\begin{aligned} d\mathbf{u}_{\mathbf{I}} &= (-\Lambda_{\mathbf{I},0}\mathbf{u}_{\mathbf{I}} + \mathbf{B}_{\mathbf{I},0}(\mathbf{u}_{\mathbf{I}}, \mathbf{u}_{\mathbf{I}}) + \mathbf{F}_{\mathbf{I}})dt + (-\Lambda_{\mathbf{I},1} + \mathbf{B}_{\mathbf{I},1}(\mathbf{u}_{\mathbf{I}}))\mathbf{u}_{\mathbf{II}} dt + \mathbf{\Sigma}_{\mathbf{I}} d\mathbf{W}_{\mathbf{I}}, \\ d\mathbf{u}_{\mathbf{II}} &= (-\Lambda_{\mathbf{II},0}\mathbf{u}_{\mathbf{I}} + \mathbf{B}_{\mathbf{II},0}(\mathbf{u}_{\mathbf{I}}, \mathbf{u}_{\mathbf{I}}) + \mathbf{F}_{\mathbf{II}})dt + (-\Lambda_{\mathbf{II},1} + \mathbf{B}_{\mathbf{II},1}(\mathbf{u}_{\mathbf{I}}))\mathbf{u}_{\mathbf{II}} dt + \mathbf{\Sigma}_{\mathbf{II}} d\mathbf{W}_{\mathbf{II}}. \end{aligned} \tag{25}$$

The quantities in the brackets naturally correspond to $\mathbf{A}_0, \mathbf{A}_1, \mathbf{a}_0$ and $\mathbf{a}_1$ respectively.

For the damping term Λ, we assume there are constants $0 < \lambda_- \le \lambda_+$,

$$\lambda_- I_{N_{\mathbf{I}}+N_{\mathbf{II}}} \preceq \begin{bmatrix} \Lambda_{\mathbf{I},0} & \Lambda_{\mathbf{I},1} \\ \Lambda_{\mathbf{II},0} & \Lambda_{\mathbf{II},1} \end{bmatrix} \preceq \lambda_+ I_{N_{\mathbf{I}}+N_{\mathbf{II}}}. \tag{26}$$

The energy conservation condition, $\mathbf{u} \cdot \mathbf{B}(\mathbf{u}, \mathbf{u}) = 0$, requires that

$$\mathbf{u}_{\mathbf{I}} \cdot \mathbf{B}_{\mathbf{I},0}(\mathbf{u}_{\mathbf{I}}, \mathbf{u}_{\mathbf{I}}) = 0, \quad \mathbf{u}_{\mathbf{II}} \cdot \mathbf{B}_{\mathbf{II},1}(\mathbf{u}_{\mathbf{I}})\mathbf{u}_{\mathbf{II}} = 0, \quad \mathbf{u}_{\mathbf{I}} \cdot \mathbf{B}_{\mathbf{I},1}(\mathbf{u}_{\mathbf{I}})\mathbf{u}_{\mathbf{II}} + \mathbf{u}_{\mathbf{II}} \cdot \mathbf{B}_{\mathbf{II},0}(\mathbf{u}_{\mathbf{I}}, \mathbf{u}_{\mathbf{I}}) = 0. \tag{27}$$

See the Appendix of [14] for details.

Proposition 3.8. *For the stochastic flow with energy conserving quadratic nonlinearity* (25)*, assume that* (26) *and* (27) *hold, and* $\mathbf{\Sigma_I}$ *and* $\mathbf{\Sigma_{II}}$ *are of full rank. We have the following results:*
1). Assumption 3.6 *holds with* $\rho = \frac{1}{2}\lambda_-$ *and* $D_e = \frac{1}{2\lambda_-}(|\mathbf{F_I}|^2 + |\mathbf{F_{II}}|^2)$.
2). Assumption 3.4 *holds with* $v = 1, m = 1$ *and*

$$D_c = \max\left\{1, \frac{2\lambda_+\sigma_{\mathbf{II},-}^{-2}}{1-\exp(-2\lambda_+)}, \frac{\sigma_{\mathbf{II},+}^2}{2\lambda_-}, 2\lambda_+^2\sigma_{\mathbf{I},-}^{-2}, 2\lambda_B^2\sigma_{\mathbf{I},-}^{-2}, \exp(2\lambda_+)\right\},$$

where the constants are chosen such that $|\mathbf{B}_{\mathbf{II},1}(\mathbf{u_I})| \leq \lambda_B|\mathbf{u_I}|$ *and*

$$\sigma_{\mathbf{I},-}^2 I_{N_\mathbf{I}} \preceq \mathbf{\Sigma_I\Sigma_I^*}, \quad \sigma_{\mathbf{II},-}^2 I_{N_\mathbf{II}} \preceq \mathbf{\Sigma_{II}\Sigma_{II}^*} \preceq \sigma_{\mathbf{II},+}^2 I_{N_\mathbf{II}}.$$

The proof of Proposition 3.8 is shown in SM4. The energy conservation property plays an essential role in verifying the system stability, and

4. Proofs.

4.1. Finite time MISE.

Proof of Theorem 3.1. Denote the one sample path density function:

$$\hat{p}_i(\mathbf{u_I}, \mathbf{u_{II}}) := K_H(\mathbf{u_I} - \mathbf{u_I}^i(t))p(\mathbf{u_{II}}|\mathbf{u_I}^i(s \leq t)),$$

such that the recovered PDF is given by $\hat{p}_t(x,y) = \frac{1}{L}\sum_{i=1}^L \hat{p}_i(x,y)$. Consider its average

$$\bar{p}_t(\mathbf{u_I}, \mathbf{u_{II}}) = \mathbb{E}K_H(\mathbf{u_I} - \mathbf{u_I}(t))p(\mathbf{u_{II}}|\mathbf{u_I}^i(s \leq t)) = \mathbb{E}\hat{p}_t(\mathbf{u_I}, \mathbf{u_{II}}).$$

The true density can be written as $p_t(\mathbf{u_I}, \mathbf{u_{II}}) = \mathbb{E}\delta_{\mathbf{u_I}^i(t)}(\mathbf{u_I})p(\mathbf{u_{II}}|\mathbf{u_I}^i(s \leq t))$, since for any test function f, the following holds

$$\begin{aligned}
&\int p_t(\mathbf{u_I}, \mathbf{u_{II}})f(\mathbf{u_I}, \mathbf{u_{II}})d\mathbf{u_I}d\mathbf{u_{II}} = \mathbb{E}f(\mathbf{u_I}^i(t), \mathbf{u_{II}}^i(t)) \\
&\quad = \mathbb{E}\mathbb{E}(f(\mathbf{u_I}^i(t), \mathbf{u_{II}}^i(t))|\mathbf{u_I}^i(s \leq t)) = \mathbb{E}\int f(\mathbf{u_I}, \mathbf{u_{II}})\delta_{\mathbf{u_I}^i(t)}(\mathbf{u_I})p(\mathbf{u_{II}}|\mathbf{u_I}^i(s \leq t))d\mathbf{u_I}d\mathbf{u_{II}}.
\end{aligned}$$

This gives the following result

$$\begin{aligned}
\bar{p}_t(\mathbf{u_I}, \mathbf{u_{II}}) &= \mathbb{E}K_H(\mathbf{u_I} - \mathbf{u_I}(t))p(\mathbf{u_{II}}|\mathbf{u_I}^i(s \leq t)) \\
&= \mathbb{E}\int d\mathbf{u}_\mathbf{I}' K_H(\mathbf{u_I} - \mathbf{u}_\mathbf{I}')\delta_{\mathbf{u_I}^i(t)}(\mathbf{u}_\mathbf{I}')p(\mathbf{u_{II}}|\mathbf{u_I}^i(s \leq t)) \\
&= \int d\mathbf{u}_\mathbf{I}' K_H(\mathbf{u_I} - \mathbf{u}_\mathbf{I}')p_t(\mathbf{u}_\mathbf{I}', \mathbf{u_{II}}) =: K_H * p_t(\mathbf{u_I}, \mathbf{u_{II}}),
\end{aligned}$$

where $*$ denotes the convolution. The Variance-Bias decomposition of the MISE can be made:

$$\begin{aligned}
&\mathbb{E}\int |\hat{p}_t(\mathbf{u_I}, \mathbf{u_{II}}) - p_t(\mathbf{u_I}, \mathbf{u_{II}})|^2 d\mathbf{u_I}d\mathbf{u_{II}} \\
&= \int \mathbb{E}|\hat{p}_t(\mathbf{u_I}, \mathbf{u_{II}}) - \bar{p}_t(\mathbf{u_I}, \mathbf{u_{II}})|^2 d\mathbf{u_I}d\mathbf{u_{II}} + \int |\bar{p}_t(\mathbf{u_I}, \mathbf{u_{II}}) - p_t(\mathbf{u_I}, \mathbf{u_{II}})|^2 d\mathbf{u_I}d\mathbf{u_{II}} \\
&= \int \mathrm{var}\ \hat{p}_t(\mathbf{u_I}, \mathbf{u_{II}})d\mathbf{u_I}d\mathbf{u_{II}} + \int |\bar{p}_t(\mathbf{u_I}, \mathbf{u_{II}}) - p_t(\mathbf{u_I}, \mathbf{u_{II}})|^2 d\mathbf{u_I}d\mathbf{u_{II}}. \qquad (28)
\end{aligned}$$

Since $\bar{p}_t = p_t * K_H$, so

$$|\bar{p}_t(\mathbf{u}_\mathbf{I}, \mathbf{u}_\mathbf{II}) - p_t(\mathbf{u}_\mathbf{I}, \mathbf{u}_\mathbf{II})| = \left| \int K_H(\mathbf{u}_\mathbf{I} - \mathbf{u}'_\mathbf{I})(p_t(\mathbf{u}'_\mathbf{I}, \mathbf{u}_\mathbf{II}) - p_t(\mathbf{u}_\mathbf{I}, \mathbf{u}_\mathbf{II}))d\mathbf{u}'_\mathbf{I} \right|.$$

In Lemma SM1.2, a Taylor expansion on $(p_t(\mathbf{u}'_\mathbf{I}, \mathbf{u}_\mathbf{II}) - p_t(\mathbf{u}_\mathbf{I}, \mathbf{u}_\mathbf{II}))$ leads to the following upper bound for the bias part:

$$\frac{1+\delta}{4} H^2 J \left(\sum_{i=1}^{N_\mathbf{I}} c_i^2 \partial^2_{\mathbf{u}^2_{\mathbf{I},i}} p_t(\mathbf{u}_\mathbf{I}, \mathbf{u}_\mathbf{II}) \right) + \frac{1+\delta^{-1}}{2} M^2 H^3 \left(\sum_{i=1}^{N_\mathbf{I}} c_i^2 \right)^3 J(M(\mathbf{u}_\mathbf{I}, \mathbf{u}_\mathbf{II})), \quad \forall \delta > 0.$$

Moreover, in light of the relation $\hat{p}_t(\mathbf{u}_\mathbf{I}, \mathbf{u}_\mathbf{II}) = \frac{1}{L}\sum_{i=1}^L \hat{p}_i(\mathbf{u}_\mathbf{I}, \mathbf{u}_\mathbf{II})$ and the independence of the density samples $\hat{p}_i$, we have

$$\begin{aligned}\int \text{var } \hat{p}_t(\mathbf{u}_\mathbf{I}, \mathbf{u}_\mathbf{II}) d\mathbf{u}_\mathbf{I} d\mathbf{u}_\mathbf{II} &= \frac{1}{L} \int \text{var } \hat{p}_i(\mathbf{u}_\mathbf{I}, \mathbf{u}_\mathbf{II}) d\mathbf{u}_\mathbf{I} d\mathbf{u}_\mathbf{II} \\ &\le \frac{1}{L} \int \mathbb{E}|\hat{p}_i(\mathbf{u}_\mathbf{I}, \mathbf{u}_\mathbf{II})|^2 d\mathbf{u}_\mathbf{I} d\mathbf{u}_\mathbf{II} = \frac{1}{L}\mathbb{E} \int |\hat{p}_i(\mathbf{u}_\mathbf{I}, \mathbf{u}_\mathbf{II})|^2 d\mathbf{u}_\mathbf{I} d\mathbf{u}_\mathbf{II}.\end{aligned}$$

Note that each $\hat{p}_i(x, y)$ is a Gaussian density with mean $(\mathbf{u}^i_\mathbf{I}(t), \bar{\mathbf{u}}_\mathbf{II}(t))$ and a block diagonal covariance, where the blocks are given by $H\mathbf{C}$ and $\mathbf{R}_\mathbf{II}(t)$, respectively. In Lemma SM1.1, a straightforward computation of the L^2 norm of a Gaussian density shows that

$$\int |\hat{p}_i(\mathbf{u}_\mathbf{I}, \mathbf{u}_\mathbf{II})|^2 d\mathbf{u}_\mathbf{I} d\mathbf{u}_\mathbf{II} = \frac{1}{\sqrt{\prod_{i=1}^{N_\mathbf{I}} (\pi H c_i^2) \det(\pi \mathbf{R}_\mathbf{II}(t))}}.$$

This leads to the bound of the MISE. ■

Proof of Proposition 3.2. Denote $\hat{p}_i(\mathbf{u}_\mathbf{II}) = p(\mathbf{u}_\mathbf{II}|\mathbf{u}^i_\mathbf{I}(s \le t))$, then following the same proof as in Theorem 3.1, we have $p_t(\mathbf{u}_\mathbf{II}) = \mathbb{E}\hat{p}_i(\mathbf{u}_\mathbf{II})$ and $\hat{p}_t(\mathbf{u}_\mathbf{II}) = \frac{1}{L}\sum_{i=1}^L \hat{p}_i(\mathbf{u}_\mathbf{II})$. Thus,

$$\begin{aligned}\int |p_t(\mathbf{u}_\mathbf{II}) - \hat{p}_t(\mathbf{u}_\mathbf{II})|^2 d\mathbf{u}_\mathbf{II} = \int \text{var } \hat{p}_t(\mathbf{u}_\mathbf{II}) d\mathbf{u}_\mathbf{II} &= \frac{1}{L} \int \text{var } \hat{p}_i(\mathbf{u}_\mathbf{II}) d\mathbf{u}_\mathbf{II} \\ &\le \frac{1}{L} \int \mathbb{E}|\hat{p}_i(\mathbf{u}_\mathbf{II})|^2 d\mathbf{u}_\mathbf{II} = \frac{1}{L}\mathbb{E}\frac{1}{\sqrt{\det(\pi \mathbf{R}_\mathbf{II}(t))}}.\end{aligned}$$ ■

Proof of Corollary 3.3. The proof is identical to the one of Theorem 3.1, as long as one replaces the densities involving $\mathbf{u}_\mathbf{II}$ to the version for $\mathbf{u}^P_\mathbf{II}$. Therefore it is omitted here. ■

4.2. Long time result.

Proof of Theorem 3.7. Part 1): The geometric ergodicity, i.e. the following L^1 convergence,

$$\int |p_t(\mathbf{u}_\mathbf{I}, \mathbf{u}_\mathbf{II}) - p_\infty(\mathbf{u}_\mathbf{I}, \mathbf{u}_\mathbf{II})| d\mathbf{u}_\mathbf{I} d\mathbf{u}_\mathbf{II} \le D_0 e^{-ct} \langle |\mathbf{u}|^2 + 1, p_0 \rangle,$$

is a direct result that comes from the framework of [60, 59]. Its equivalence to the Poincaré type of inequality (24) is a result by [1]. We will try to verify the conditions needed in [1].

We claim that $V(\mathbf{u_I},\mathbf{u_{II}}) = |\mathbf{u_I}|^2 + |\mathbf{u_{II}}|^2 + 1$ is a Lyapunov function of Definition 1.1 in [1]. Apply the generator $\mathcal{L}$ of the diffusion process

$$\begin{aligned}\mathcal{L}V &= 2\mathbf{u_I}\cdot(\mathbf{A}_0 + \mathbf{A}_1\mathbf{u_{II}}) + 2\mathbf{u_{II}}\cdot(\mathbf{a}_0 + \mathbf{a}_1\mathbf{u_{II}}) + \mathrm{tr}(\mathbf{\Sigma_I}\mathbf{\Sigma_I^*} + \mathbf{\Sigma_{II}}\mathbf{\Sigma_{II}^*})\\ &\leq -2\rho V + (2\rho + 2D_e + \mathrm{tr}(\mathbf{\Sigma_I}\mathbf{\Sigma_I^*} + \mathbf{\Sigma_{II}}\mathbf{\Sigma_{II}^*})) \leq -\rho V + b\mathbb{1}_{\mathcal{U}},\end{aligned}$$

where $b = 2\rho + 2D_e + \mathrm{tr}(\mathbf{\Sigma_I}\mathbf{\Sigma_I^*} + \mathbf{\Sigma_{II}}\mathbf{\Sigma_{II}^*})$, and $\mathcal{U} = \{V(\mathbf{u_I},\mathbf{u_{II}}) \leq b\}$. The fact that $\mathcal{U}$, and actually any compact subset, is a petite set can be verified by the same proof of Lemma 3.4 in [59], since we assume $\mathbf{\Sigma_I}$ and $\mathbf{\Sigma_{II}}$ are full rank. The fact the stochastic process is irreducible can also be verified using the same argument. More details on these arguments are provided in [57] for more general conditions.

Therefore, applying theorem 1.2 of [1] leads to the L^1 convergence above. Theorem 2.1 also applies with $f(\mathbf{u_I},\mathbf{u_{II}}) = \frac{p_0}{p_\infty}(\mathbf{u_I},\mathbf{u_{II}})$, which gives (24).

Part 2): We again decompose the MISE into (28).

$$\mathrm{MISE} = \int \mathrm{var}\ \hat{p}_t(\mathbf{u_I},\mathbf{u_{II}})d\mathbf{u_I}d\mathbf{u_{II}} + \int |\bar{p}_t(\mathbf{u_I},\mathbf{u_{II}}) - p_t(\mathbf{u_I},\mathbf{u_{II}})|^2 d\mathbf{u_I}d\mathbf{u_{II}}.$$

Following the proof of Theorem 3.1, we have the variance part

$$\int \mathrm{var}\hat{p}_t(\mathbf{u_I},\mathbf{u_{II}})d\mathbf{u_I}d\mathbf{u_{II}} \leq \mathbb{E}\frac{1}{L\sqrt{\prod_{i=1}^{N_\mathbf{I}}(\pi H c_i^2)\det(\pi \mathbf{R_{II}}(t))}}.$$

Proposition 3.5 leads to $\mathbb{E}\frac{1}{\sqrt{\det(\mathbf{R_{II}}(t))}} \leq D_1 + D_2\int_{t-v}^t \mathbb{E}|\mathbf{u_I}(r)|^{mN_\mathbf{II}}dr$. To provide a bound for $\mathbb{E}|\mathbf{u_I}(t)|^{mN_\mathbf{II}}$, we verify that any fixed moment $|\mathbf{u}|^{2n} = (|\mathbf{u_I}|^2 + |\mathbf{u_{II}}|^2)^n$ is also dissipative. Applying the generator of the diffusion process yields

$$\begin{aligned}\mathcal{L}|\mathbf{u}(t)|^{2n} &= 2n|\mathbf{u}|^{2(n-1)}(\mathbf{u_I}\cdot(\mathbf{A}_0 + \mathbf{A}_1\mathbf{u_{II}}) + \mathbf{u_{II}}\cdot(\mathbf{a}_0 + \mathbf{a}_1\mathbf{u_{II}}))\\ &\quad + n\mathrm{tr}(\mathbf{\Sigma}^*(|\mathbf{u}|^{2(n-1)}I + 2(n-1)|\mathbf{u}|^{2(n-2)}\mathbf{u}\mathbf{u}^*)\mathbf{\Sigma})\\ &\leq -2n\rho|\mathbf{u}|^{2n} + 2nD_e|\mathbf{u}|^{2(n-1)} + 2n^2\mathrm{tr}(\mathbf{\Sigma_I}\mathbf{\Sigma_I^*} + \mathbf{\Sigma_{II}}\mathbf{\Sigma_{II}^*}))|\mathbf{u}|^{2(n-1)} \leq -n\rho|\mathbf{u}|^{2n} + D_{n,\mathbf{\Sigma}},\end{aligned}$$

where $\mathbf{\Sigma} = [\mathbf{\Sigma_I^*},\mathbf{\Sigma_{II}^*}]^*$ and the constant $D_{n,\mathbf{\Sigma}}$ exists because of Young's inequality.

Apply Dynkin's formula for $e^{\rho n t}|\mathbf{u}(t)|^{2n}$, and combine it with the result above, we have the following Gronswall's inequality

$$\mathbb{E}|\mathbf{u}(t)|^{2n} \leq e^{-\rho n t}\mathbb{E}|\mathbf{u}(0)|^{2n} + \frac{D_{n,\mathbf{\Sigma}}}{n\rho}. \tag{29}$$

To continue, we let $n = mN_\mathbf{II}/2$ in (29) and integrate it in time range $[t-v,t]$,

$$\mathbb{E}\int_{t-v}^t |\mathbf{u_I}(s)|^{mN_\mathbf{II}}ds \leq v\exp(-\tfrac{1}{2}\rho m N_\mathbf{II}(t-v))\mathbb{E}|\mathbf{u}(0)|^{mN_\mathbf{II}} + \frac{2vD_{mN_\mathbf{II}/2,\mathbf{\Sigma}}}{mN_\mathbf{II}\rho}.$$

Consequently, there exists a constant $D_{m,N_\mathbf{II},v}$ such that

$$\int \mathrm{var}\hat{p}_t(\mathbf{u_I},\mathbf{u_{II}})d\mathbf{u_I}d\mathbf{u_{II}} \leq \frac{D_{m,N_\mathbf{II},v}}{L\pi^{\frac{N_\mathbf{I}+N_\mathbf{II}}{2}}H^{\frac{N_\mathbf{I}}{2}}\prod_{i=1}^{N_\mathbf{I}}c_i}\left(\exp(-\tfrac{1}{2}\rho m N_\mathbf{II} t)\mathbb{E}|\mathbf{u}(0)|^{mN_\mathbf{II}} + D_{m,N_\mathbf{II},v}\right).$$

For the bias term, $\int |\bar{p}_t(\mathbf{u_I},\mathbf{u_{II}}) - p_t(\mathbf{u_I},\mathbf{u_{II}})|^2 d\mathbf{u_I} d\mathbf{u_{II}}$, we use the Cauchy Schwartz

$$(a+b+c)^2 \le \left(\frac{1}{1+\delta} + \frac{\delta}{2(1+\delta)} + \frac{\delta}{2(1+\delta)}\right)\left((1+\delta)a^2 + 2(1+\delta^{-1})b^2 + 2(1+\delta^{-1})c^2\right),$$

with

$$a = |p_t(\mathbf{u_I},\mathbf{u_{II}}) - p_\infty(\mathbf{u_I},\mathbf{u_{II}})|,\ b = |\bar{p}_t(\mathbf{u_I},\mathbf{u_{II}}) - \bar{p}_\infty(\mathbf{u_I},\mathbf{u_{II}})|,\ c = |p_\infty(\mathbf{u_I},\mathbf{u_{II}}) - \bar{p}_\infty(\mathbf{u_I},\mathbf{u_{II}})|.$$

Recall that $\bar{p}_\infty = K_H * p_\infty$. Using the same proof as in Theorem 3.1, we have

$$\begin{aligned}&\int |p_\infty(\mathbf{u_I},\mathbf{u_{II}}) - \bar{p}_\infty(\mathbf{u_I},\mathbf{u_{II}})|^2 d\mathbf{u_I} d\mathbf{u_{II}}\\ &\qquad \le \frac{1+\delta}{4} H^2 R\left(\sum_{i=1}^{N_\mathbf{I}} c_i^2 \partial^2_{\mathbf{u}^2_{\mathbf{I},i}} p_\infty(\mathbf{u_I},\mathbf{u_{II}})\right) + \frac{1+\delta^{-1}}{2} H^3 \left(\sum_{i=1}^{N_\mathbf{I}} c_i^2\right)^3.\end{aligned}$$

Then apply (24), we have

$$\begin{aligned}&\int |p_t(\mathbf{u_I},\mathbf{u_{II}}) - p_\infty(\mathbf{u_I},\mathbf{u_{II}})|^2 d\mathbf{u_I} d\mathbf{u_{II}}\\ &\qquad \le \|p_\infty\|_\infty \int |p_t(\mathbf{u_I},\mathbf{u_{II}}) - p_\infty(\mathbf{u_I},\mathbf{u_{II}})|^2 \frac{1}{p_\infty(\mathbf{u_I},\mathbf{u_{II}})} d\mathbf{u_I} d\mathbf{u_{II}}\\ &\qquad \le D_0 e^{-ct}\langle |\mathbf{u}|^2 + 1, p_0\rangle \left\|\tfrac{p_0}{p_\infty} - 1\right\|_\infty^2 \|p_\infty\|_\infty.\end{aligned}$$

Next, recall that $\bar{p}_t(\mathbf{u_I},\mathbf{u_{II}}) = \int K_H(\mathbf{u}'_\mathbf{I}) p_t(\mathbf{u_I} - \mathbf{u}'_\mathbf{I}, \mathbf{u_{II}}) d\mathbf{u}'_\mathbf{I}$. Therefore, by Cauchy Schwartz

$$\begin{aligned}&|\bar{p}_t(\mathbf{u_I},\mathbf{u_{II}}) - \bar{p}_\infty(\mathbf{u_I},\mathbf{u_{II}})|^2\\ &\qquad = \left(\int K_H(\mathbf{u}'_\mathbf{I})(p_t(\mathbf{u_I} - \mathbf{u}'_\mathbf{I}, \mathbf{u_{II}}) - p_\infty(\mathbf{u_I} - \mathbf{u}'_\mathbf{I}, \mathbf{u_{II}})) d\mathbf{u}'_\mathbf{I}\right)^2\\ &\qquad \le \int K_H(\mathbf{u}'_\mathbf{I}) d\mathbf{u}'_\mathbf{I} \int (p_t(\mathbf{u_I} - \mathbf{u}'_\mathbf{I}, \mathbf{u_{II}}) - p_\infty(\mathbf{u_I} - \mathbf{u}'_\mathbf{I}, \mathbf{u_{II}}))^2 K_H(\mathbf{u}'_\mathbf{I}) d\mathbf{u}'_\mathbf{I}\\ &\qquad \le \int (p_t(\mathbf{u_I} - \mathbf{u}'_\mathbf{I}, \mathbf{u_{II}}) - p_\infty(\mathbf{u_I} - \mathbf{u}'_\mathbf{I}, \mathbf{u_{II}}))^2 K_H(\mathbf{u}'_\mathbf{I}) d\mathbf{u}'_\mathbf{I}.\end{aligned}$$

Consequently,

$$\begin{aligned}&\int |\bar{p}_t(\mathbf{u_I},\mathbf{u_{II}}) - \bar{p}_\infty(\mathbf{u_I},\mathbf{u_{II}})|^2 d\mathbf{u_I} d\mathbf{u_{II}}\\ &\qquad \le \int (p_t(\mathbf{u_I} - \mathbf{u}'_\mathbf{I}, \mathbf{u_{II}}) - p_\infty(\mathbf{u_I} - \mathbf{u}'_\mathbf{I}, \mathbf{u_{II}}))^2 K_H(\mathbf{u}'_\mathbf{I}) d\mathbf{u}'_\mathbf{I} d\mathbf{u_I} d\mathbf{u_{II}}\\ &\qquad = \int \left(\int (p_t(\mathbf{u_I} - \mathbf{u}'_\mathbf{I}, \mathbf{u_{II}}) - p_\infty(\mathbf{u_I} - \mathbf{u}'_\mathbf{I}, \mathbf{u_{II}}))^2 d\mathbf{u_I} d\mathbf{u_{II}}\right) K_H(\mathbf{u_I} - \mathbf{u}'_\mathbf{I}) d\mathbf{u}'_\mathbf{I}\\ &\qquad = \int |p_t(\mathbf{u_I},\mathbf{u_{II}}) - p_\infty(\mathbf{u_I},\mathbf{u_{II}})|^2 d\mathbf{u_I} d\mathbf{u_{II}}.\end{aligned}$$

Combining the results finishes the proof. ∎

5. Numerical examples. Below, numerical examples are used to support the theoretical results in section 3. The test model considered here is the following *triad model* [52],

$$\frac{du_1}{dt} = A_1 u_2 u_3, \tag{30a}$$

$$\frac{du_2}{dt} = A_2 u_3 u_1 - d_2 u_2 + \sigma_2 \dot{W}_2, \tag{30b}$$

$$\frac{du_3}{dt} = A_3 u_1 u_2 - d_3 u_3 + \sigma_3 \dot{W}_3, \tag{30c}$$

where $A_1 + A_2 + A_3 = 0$ represents the energy-conserving nonlinear interactions and $d_2 > 0, d_3 > 0$ are the damping terms. Note that there is no damping and dissipation in (30a) but (30) is a hypoelliptic diffusion [57, 59]. Linear stability is satisfied for u_2, u_3 while there is only neutral stability of u_1. Define $E_2 = \sigma_2^2/(2d_2)$ and $E_3 = \sigma_3^2/(2d_3)$. It is straightforward to show that the triad system (30) has a Gaussian invariant measure [43, 52]

$$p_{eq}(u) = C \exp\left(-\frac{1}{2}\left(\frac{u_1^2}{E_1} + \frac{u_2^2}{E_2} + \frac{u_3^2}{E_3}\right)\right),$$

provided that the following condition is satisfied

$$E_1 = -A_1 E_2 E_3 (A_2 E_3 + A_3 E_2)^{-1} > 0. \tag{31}$$

If the condition in (31) is violated, namely $E_1 < 0$, then the variance in u_1 direction will increase unboundedly and there is no invariant measure for the triad system (30).

Below, two dynamical regimes of the triad model (30) are studied, where the corresponding parameters are listed in the Table 1. Particularly, the triad system (30) in Regime I has a Gaussian invariant measure while there is no invariant measure in Regime II due to the fact that $E_1 < 0$. See Figure 1 for the time evolution of the three marginal variances and one realization of each variable and [41] for dynamical introduction about such triad models.

Table 1
Parameters of two dynamical regimes of the triad model (30)

	A_1	A_2	A_3	d_2	d_3	σ_2	σ_3		E_2	E_3	E_1	$\mathrm{Var}(u_1)$
Regime I	-2.5	1	1.5	1	0.5	1	1	$\Longrightarrow$	0.5	1	5/11	Bounded
Regime II	-0.5	-1	1.5	1	0.5	1	1		0.5	1	$-5/3$	Unbounded

Denote $\mathbf{u_I} = (u_2, u_3)^T$ and $\mathbf{u_{II}} = u_1$. The triad system (30) belongs to the conditional Gaussian family (1). Notably, the noise coefficient in $\mathbf{u_{II}}$ is $\mathbf{\Sigma_{II}} = 0$, which implies the system has no controllability. The initial values in the tests below are all given at origin. Here only the hybrid method (9) is tested and the number of samples is always $L = 500$.

Figure 2 shows the recovered PDF at $t = 1$ in Regime I of the triad model. Despite an accurate estimation of the joint PDF of the observed variables $p(u_2, u_3)$ as shown in Panel (e), the recovered PDF of the unobserved variable u_1 in Panel (f) has quite a few noisy fluctuations and the recovered joint PDFs $p(u_1, u_2)$ and $p(u_3, u_1)$ in Panel (d) and (f) are non-smooth in u_1 direction as well. Such pathological behavior results from the loss of controllability of the system, which is consistent with the theoretical discussion in subsection 3.5. In fact, the term $\mathbf{a}_1$ in (1) associated with the triad system (30) is zero. Therefore, according to (3), $\mathbf{\Sigma_{II}} = 0$ implies the posterior variance $\mathbf{R_{II}} = \mathbf{0}$ and the posterior mean $\bar{\mathbf{u}}_{\mathbf{II}}$ simply follows the sampled trajectory of $\mathbf{u_{II}}$. In other words, the posterior states from the algorithm are exactly the Monte Carlo samples, as is validated in Panel (h). The same performance is found in Regime II and thus we omit the figure here.

In order to make the triad system have controllability, a small noise is added to (30a) and the resulting *modified triad system* is given as follows,

$$\frac{du_1}{dt} = A_1 u_2 u_3 + \epsilon \dot{W}_1, \tag{32a}$$

$$\frac{du_2}{dt} = A_2 u_3 u_1 - d_2 u_2 + \sigma_2 \dot{W}_2, \tag{32b}$$

$$\frac{du_3}{dt} = A_3 u_1 u_2 - d_3 u_3 + \sigma_3 \dot{W}_3, \tag{32c}$$

where ϵ is the noise coefficient of u_1 with $\mathbf{\Sigma_{II}} = \epsilon$ in (1). Below we set $\epsilon = 0.1 \ll \sigma_2 = \sigma_3 = 1$. The other parameters in (32) remain the same as those in Table 1.

This extra noise implies the triad system is controllable, which significantly improves the accuracy of the recovered PDFs. See Figure 3 for the results in Regime I at $t = 1$. In particular, Panel (h) of Figure 3 shows that the posterior means are quite different from the Monte Carlo samples and the posterior variances are no longer zero. It is also shown in Figure SM1 that the recovered PDFs at a long time $t = 20$ (i.e., statistically steady state) are very close to the truth with this extra small noise.

Similarly, Figure 4 shows the recovered PDFs of Regime II with $\epsilon = 0.1$ at $t = 1$, the error in which compared with the truth is negligible. Notably, although the amplitude of u_1 has an unbounded growth in this regime due to the fact that $E_1 < 0$, the recovered PDFs with $\epsilon = 0.1$ at $t = 20$ as illustrated in Figure SM2 remain quite accurate. Next, the performance of the hybrid algorithm at a very long time in this regime is studied. Figure 5 shows the recovered PDFs at $t = 400$. Similar to Figure 2, the noisy fluctuations are found in the recovered PDF of u_1. In fact, direct calculations show that the posterior variance $\mathbf{R_{II}}$ in (3) is bounded from above since the unbounded signal u_1 does not enter into the evolution of $\mathbf{R_{II}}$, which is also validated by the numerical simulation in Panel (h). Since the variance of u_1 increases with time, the percentage of the portion covered by each conditional Gaussian distribution decreases in time, which reduces the skill in the recovered PDFs by the conditional Gaussian mixtures. In Figure SM3, we show that by further imposing a damping in the dynamics of u_1 of the modified triad model (32) in Regime II, the model then satisfies all the conditions in Proposition 3.8 and the resulting model has an invariant measure. In such a scenario, the hybrid algorithm is skillful in both short and long time as is affirmed by Proposition 3.8.

It is also worthwhile pointing out that all the test models in [14], including the noisy version of Lorenz 63 model [40], the stochastic climate model [49, 42], the nonlinear triad model mimicking structural features of low-frequency variability of GCMs with non-Gaussian features [48] and the modified conceptual dynamical model for turbulence [53], all satisfy the conditions in Proposition 3.8. Therefore, the hybrid algorithm (9) is able to solve the PDFs of those models with high accuracy with only a small number of samples.

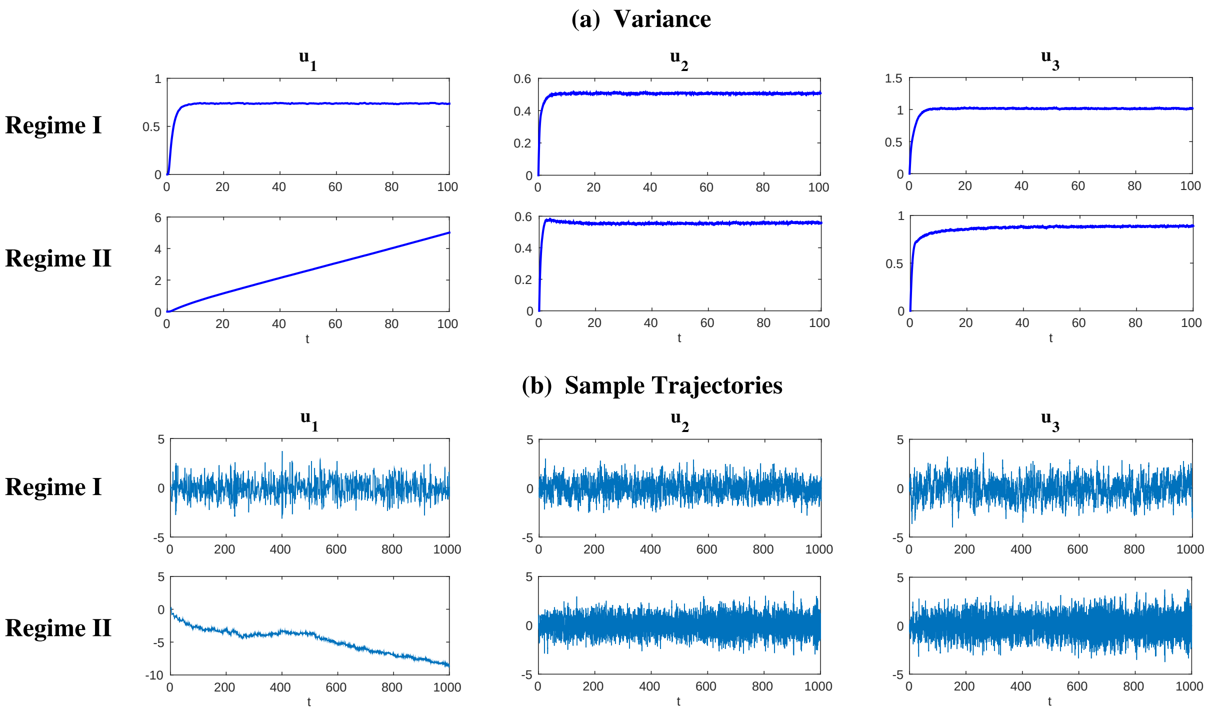

Figure 1. *Triad model (30). (a) Marginal variance as a function of time ($t \in [0, 100]$) in the two dynamical regimes with parameters in Table 1. (b) Sample trajectories up to $t = 1000$ of the two dynamical regimes. Note the unbounded growth of the amplitude of u_1 in Regime II.*

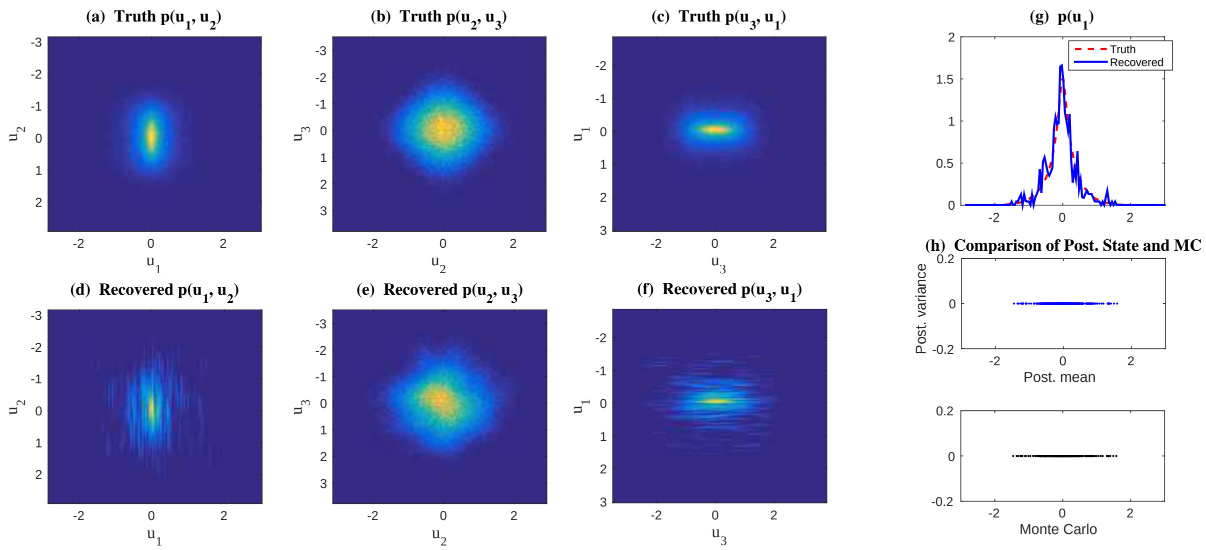

Figure 2. *Triad model (30), Regime I at $t = 1$. (a)-(c) True 2D PDF. (d)-(f) Recovered PDF. (g) True and recovered 1D PDF $p(u_1)$. (h) Top: Posterior mean (x-axis) and posterior variance (y-axis). Bottom: Monte Carlo samples. The total number of samples is $L = 500$.*

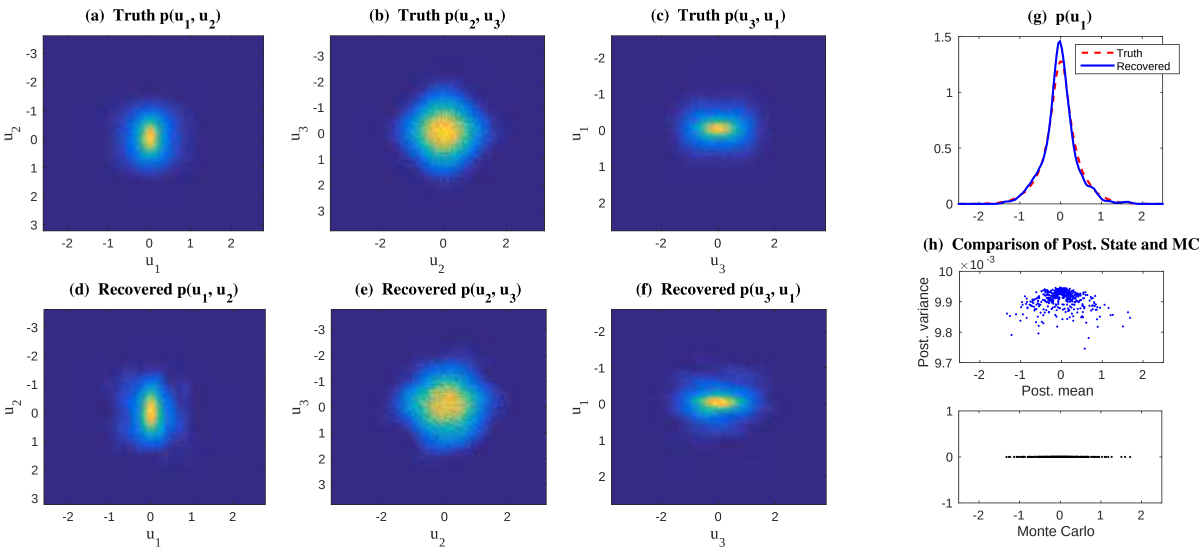

Figure 3. *Modified triad model* (32)*, Regime I at* $t = 1$*. Same captions as in Figure* 2*.*

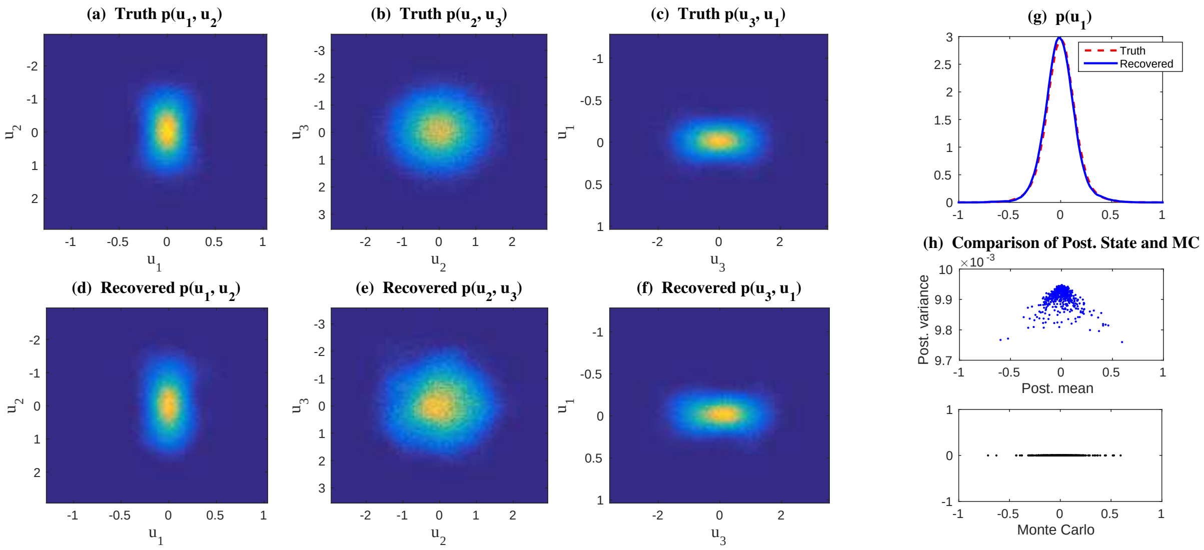

Figure 4. *Modified triad model* (32)*, Regime II at* $t = 1$*. Same captions as in Figure* 2*.*

6. Discussion and Conclusions.

This article presents a rigorous analysis for the efficient statistically accurate algorithms developed in [14], which succeed in solving both the transient and the equilibrium solutions of Fokker-Planck equations associated with high-dimensional nonlinear turbulent dynamical systems with conditional Gaussian structures. Despite the conditional Gaussianity, these nonlinear systems capture many strong non-Gaussian features such as intermittency and fat-tailed PDFs. The algorithms involve a hybrid strategy that

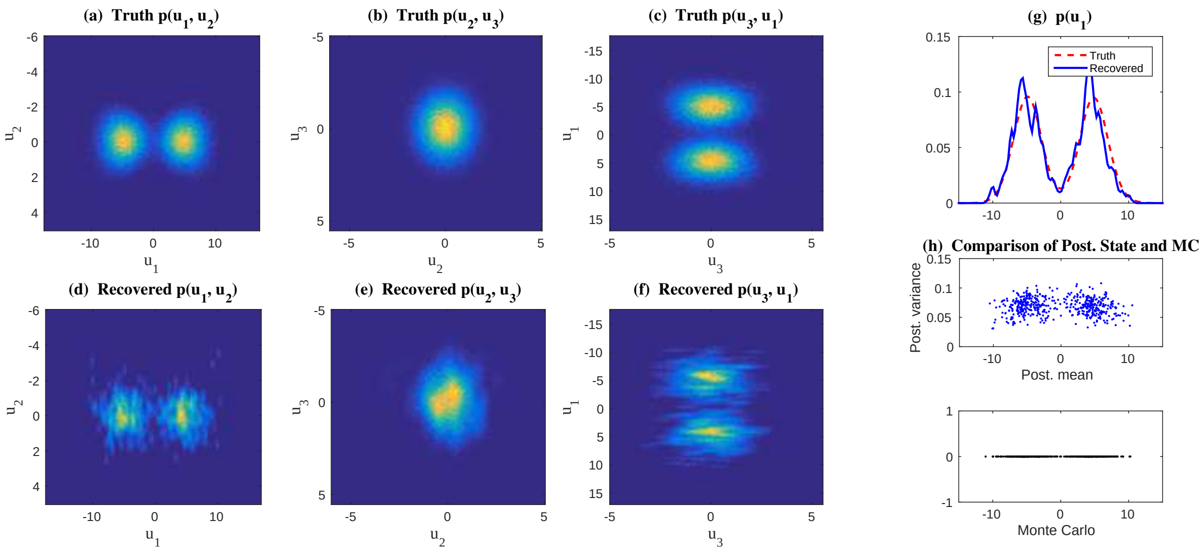

Figure 5. *Modified triad model* (32)*, Regime II at* $t = 400$*. Same captions as in Figure 2.*

requires only a small number of samples L to capture both the transient and the equilibrium non-Gaussian PDFs with high accuracy.

Theorem 3.1 shows that the MISE in the recovered high-dimensional PDFs associated with the unresolved variables $\mathbf{u_{II}}$ is bounded by $\mathbb{E}(\det(\mathbf{R_{II}})^{-1/2})$, where $\mathbf{R_{II}}$ is completely determined by the underlying dynamical systems and it has no dependence on the sample size L. This is fundamentally different from the direct application of the kernel methods to recover the PDF of $\mathbf{u_{II}}$, in which the bandwidth of the kernel H is scaled as a reciprocal of L to a certain power and the resulting MISE is proportional to $L^{-1/N_{\mathbf{II}}}$. This implies the curse of dimensionality in the kernel density estimation and other smoothed Monte Carlo methods due to the fact that L has to increase exponentially as $N_{\mathbf{II}}$ in order to guarantee the accuracy in the solution. As is shown in Theorem 3.1, many fewer samples are needed in the efficient statistically accurate algorithms in order to reach the same accuracy as using the smoothed Monte Carlo methods, especially with a large $N_{\mathbf{II}}$. Theorem 3.7 affirms the long term persistence of the efficient statistically accurate algorithms in a rigorous way under the assumption that the joint process $(\mathbf{u_I}, \mathbf{u_{II}})$ is controllable and stochastically stable. It also provides a lower bound of $\mathbf{R_{II}}$ using the controllability condition. The validations of the controllability and other theoretical conditions in the algorithms are demonstrated in the numerical simulations in section 5. Furthermore, Proposition 3.8 illustrates that the turbulent dynamical systems with quadratic energy conserving nonlinear interactions [41] automatically satisfy all the conditions for the long time persistence. This justifies the skillful performance of the efficient statistically accurate algorithms in the numerical tests reported in [14] and provides important guidelines for future applications.

REFERENCES

[1] D. Bakry, P. Cattiaux, and A. Guillin, *Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré*, Journal of Functional Analysis, 254 (2008), pp. 727–759.
[2] A. N. Bishop and P. Del Moral, *On the stability of Kalman-Bucy diffusion processes*, arXiv preprint arXiv:1610.04686, (2016).
[3] R. C. Booton, *Nonlinear control systems with random inputs*, IRE Transactions on Circuit Theory, 1 (1954), pp. 9–18.
[4] Z. I. Botev, J. F. Grotowski, D. P. Kroese, et al., *Kernel density estimation via diffusion*, The Annals of Statistics, 38 (2010), pp. 2916–2957.
[5] M. Branicki, N. Chen, and A. J. Majda, *Non-Gaussian test models for prediction and state estimation with model errors*, in Partial Differential Equations: Theory, Control and Approximation, Springer, 2014, pp. 99–138.
[6] M. Branicki and A. J. Majda, *Quantifying uncertainty for predictions with model error in non-gaussian systems with intermittency*, Nonlinearity, 25 (2012), p. 2543.
[7] M. Branicki and A. J. Majda, *Dynamic stochastic superresolution of sparsely observed turbulent systems*, Journal of Computational Physics, 241 (2013), pp. 333–363.
[8] N. Chen, D. Giannakis, R. Herbei, and A. J. Majda, *An MCMC algorithm for parameter estimation in signals with hidden intermittent instability*, SIAM/ASA Journal on Uncertainty Quantification, 2 (2014), pp. 647–669.
[9] N. Chen and A. J. Majda, *Predicting the cloud patterns for the boreal summer intraseasonal oscillation through a low-order stochastic model*, Mathematics of Climate and Weather Forecasting, 1 (2015), pp. 1–20.
[10] N. Chen and A. J. Majda, *Predicting the real-time multivariate Madden–Julian oscillation index through a low-order nonlinear stochastic model*, Monthly Weather Review, 143 (2015), pp. 2148–2169.
[11] N. Chen and A. J. Majda, *Filtering nonlinear turbulent dynamical systems through conditional Gaussian statistics*, Monthly Weather Review, 144 (2016), pp. 4885–4917.
[12] N. Chen and A. J. Majda, *Filtering the stochastic skeleton model for the Madden–Julian oscillation*, Monthly Weather Review, 144 (2016), pp. 501–527.
[13] N. Chen and A. J. Majda, *Model error in filtering random compressible flows utilizing noisy Lagrangian tracers*, Monthly Weather Review, 144 (2016), pp. 4037–4061.
[14] N. Chen and A. J. Majda, *Efficient statistically accurate algorithms for solving Fokker-Planck equations in large dimensions*, Journal of Computational Physics, (2017). (Submitted).
[15] N. Chen, A. J. Majda, and D. Giannakis, *Predicting the cloud patterns of the Madden-Julian Oscillation through a low-order nonlinear stochastic model*, Geophysical Research Letters, 41 (2014), pp. 5612–5619.
[16] N. Chen, A. J. Majda, and X. T. Tong, *Information barriers for noisy Lagrangian tracers in filtering random incompressible flows*, Nonlinearity, 27 (2014), p. 2133.
[17] N. Chen, A. J. Majda, and X. T. Tong, *Noisy Lagrangian tracers for filtering random rotating compressible flows*, Journal of Nonlinear Science, 25 (2015), pp. 451–488.
[18] C. K. Chui, G. Chen, et al., *Kalman filtering*, With real time applications, (1999).
[19] W. Cousins and T. P. Sapsis, *Quantification and prediction of extreme events in a one-dimensional nonlinear dispersive wave model*, Physica D: Nonlinear Phenomena, 280 (2014), pp. 48–58.
[20] W. Cousins and T. P. Sapsis, *Reduced-order precursors of rare events in unidirectional nonlinear water waves*, Journal of Fluid Mechanics, 790 (2016), pp. 368–388.
[21] D. Crisan and B. Rozovskii, *The Oxford handbook of nonlinear filtering*, Oxford University Press, 2011.
[22] F. Daum and J. Huang, *Curse of dimensionality and particle filters*, in Aerospace Conference, 2003. Proceedings. 2003 IEEE, vol. 4, IEEE, 2003, pp. 4_1979–4_1993.
[23] G.-K. Er, *Methodology for the solutions of some reduced Fokker-Planck equations in high dimensions*, Annalen der Physik, 523 (2011), pp. 247–258.
[24] G.-K. Er and V. P. Iu, *State-space-split method for some generalized Fokker-Planck-Kolmogorov equations in high dimensions*, Physical Review E, 85 (2012), p. 067701.
[25] J. H. Friedman, *On bias, variance, 0/1loss, and the curse-of-dimensionality*, Data mining and knowledge discovery, 1 (1997), pp. 55–77.
[26] C. W. Gardiner, *Stochastic methods*, Springer-Verlag, Berlin–Heidelberg–New York–Tokyo, 1985.

[27] B. Gershgorin, J. Harlim, and A. J. Majda, *Improving filtering and prediction of spatially extended turbulent systems with model errors through stochastic parameter estimation*, Journal of Computational Physics, 229 (2010), pp. 32–57.
[28] B. Gershgorin, J. Harlim, and A. J. Majda, *Test models for improving filtering with model errors through stochastic parameter estimation*, Journal of Computational Physics, 229 (2010), pp. 1–31.
[29] M. Ghil, P. Yiou, S. Hallegatte, B. Malamud, P. Naveau, A. Soloviev, P. Friederichs, V. Keilis-Borok, D. Kondrashov, V. Kossobokov, et al., *Extreme events: dynamics, statistics and prediction*, Nonlinear Processes in Geophysics, 18 (2011), pp. 295–350.
[30] A. Greco, W. Matthaeus, S. Servidio, P. Chuychai, and P. Dmitruk, *Statistical analysis of discontinuities in solar wind ACE data and comparison with intermittent MHD turbulence*, The Astrophysical Journal Letters, 691 (2009), p. L111.
[31] J. Harlim, A. Mahdi, and A. J. Majda, *An ensemble Kalman filter for statistical estimation of physics constrained nonlinear regression models*, Journal of Computational Physics, 257 (2014), pp. 782–812.
[32] H.-L. Huang and P. Antonelli, *Application of principal component analysis to high-resolution infrared measurement compression and retrieval*, Journal of Applied Meteorology, 40 (2001), pp. 365–388.
[33] M. C. Jones, J. S. Marron, and S. J. Sheather, *A brief survey of bandwidth selection for density estimation*, Journal of the American Statistical Association, 91 (1996), pp. 401–407.
[34] S. R. Keating, A. J. Majda, and K. S. Smith, *New methods for estimating ocean eddy heat transport using satellite altimetry*, Monthly Weather Review, 140 (2012), pp. 1703–1722.
[35] P. Kumar and S. Narayanan, *Solution of Fokker-Planck equation by finite element and finite difference methods for nonlinear systems*, Sadhana, 31 (2006), pp. 445–461.
[36] W. Lee and A. Stuart, *Derivation and analysis of simplified filters*, Communications in Mathematical Sciences, 15 (2017), pp. 413–450.
[37] Y. Lee and A. J. Majda, *Multiscale data assimilation and prediction using clustered particle filters*, Journal of Computational Physics, (2017). (Submitted).
[38] B. Lindner, J. Garcia-Ojalvo, A. Neiman, and L. Schimansky-Geier, *Effects of noise in excitable systems*, Physics reports, 392 (2004), pp. 321–424.
[39] R. S. Liptser and A. N. Shiryaev, *Statistics of Random Processes II: II. Applications*, vol. 2, Springer, 2001.
[40] E. N. Lorenz, *Deterministic nonperiodic flow*, Journal of the atmospheric sciences, 20 (1963), pp. 130–141.
[41] A. Majda, *Introduction to turbulent dynamical systems in complex systems, Frontiers in Applied Dynamical Systems: Reviews and Tutorials 5*, Springer, 2016.
[42] A. Majda, R. V. Abramov, and M. J. Grote, *Information theory and stochastics for multiscale nonlinear systems*, vol. 25, American Mathematical Soc., 2005.
[43] A. Majda, I. Timofeyev, and E. Vanden-Eijnden, *A priori tests of a stochastic mode reduction strategy*, Physica D: Nonlinear Phenomena, 170 (2002), pp. 206–252.
[44] A. Majda, I. Timofeyev, and E. Vanden-Eijnden, *Stochastic models for selected slow variables in large deterministic systems*, Nonlinearity, 19 (2006), p. 769.
[45] A. Majda and X. Wang, *Nonlinear dynamics and statistical theories for basic geophysical flows*, Cambridge University Press, 2006.
[46] A. J. Majda, *Statistical energy conservation principle for inhomogeneous turbulent dynamical systems*, Proceedings of the National Academy of Sciences, 112 (2015), pp. 8937–8941.
[47] A. J. Majda and M. Branicki, *Lessons in uncertainty quantification for turbulent dynamical systems*, Discrete Cont. Dyn. Systems, 32 (2012), pp. 3133–3221.
[48] A. J. Majda, C. Franzke, and D. Crommelin, *Normal forms for reduced stochastic climate models*, Proceedings of the National Academy of Sciences, 106 (2009), pp. 3649–3653.
[49] A. J. Majda, C. Franzke, and B. Khouider, *An applied mathematics perspective on stochastic modelling for climate*, Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 366 (2008), pp. 2427–2453.
[50] A. J. Majda and I. Grooms, *New perspectives on superparameterization for geophysical turbulence*, Journal of Computational Physics, 271 (2014), pp. 60–77.
[51] A. J. Majda and J. Harlim, *Filtering complex turbulent systems*, Cambridge University Press, 2012.
[52] A. J. Majda and J. Harlim, *Physics constrained nonlinear regression models for time series*, Nonlin-

earity, 26 (2012), p. 201.

[53] A. J. Majda and Y. Lee, *Conceptual dynamical models for turbulence*, Proceedings of the National Academy of Sciences, 111 (2014), pp. 6548–6553.

[54] A. J. Majda, D. Qi, and T. P. Sapsis, *Blended particle filters for large-dimensional chaotic dynamical systems*, Proceedings of the National Academy of Sciences, 111 (2014), pp. 7511–7516.

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

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

[57] A. J. Majda and X. T. Tong, *Ergodicity of truncated stochastic Navier-Stokes with deterministic forcing and dispersion*, Journal of Nonlinear Science, 26 (2016), pp. 1483–1506.

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

[59] J. C. Mattingly, A. M. Stuart, and D. J. Higham, *Ergodicity for SDEs and approximations: locally Lipschitz vector fields and degenerate noise*, Stochastic processes and their applications, 101 (2002), pp. 185–232.

[60] S. P. Meyn and R. L. Tweedie, *Markov chains and stochastic stability*, Springer Science & Business Media, 2012.

[61] M. A. Mohamad and T. P. Sapsis, *Probabilistic response and rare events in Mathieu's equation under correlated parametric excitation*, Ocean Engineering, 120 (2016), pp. 289–297.

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

[63] T. Palmer and J. Räisänen, *Quantifying the risk of extreme seasonal precipitation events in a changing climate*, Nature, 415 (2002), pp. 512–514.

[64] L. Pichler, A. Masud, and L. A. Bergman, *Numerical solution of the Fokker–Planck equation by finite difference and finite element methodsa comparative study*, in Computational Methods in Stochastic Dynamics, Springer, 2013, pp. 69–85.

[65] H. Risken, *The Fokker-Planck equation. Methods of solution and applications, vol. 18 of*, Springer Series in Synergetics, (1989).

[66] C. P. Robert, *Monte Carlo methods*, Wiley Online Library, 2004.

[67] R. Salmon, *Lectures on geophysical fluid dynamics*, Oxford University Press, 1998.

[68] B. W. Silverman, *Using kernel density estimates to investigate multimodality*, Journal of the Royal Statistical Society. Series B (Methodological), (1981), pp. 97–99.

[69] L. Socha, *Linearization methods for stochastic dynamic systems*, vol. 730, Springer Science & Business Media, 2007.

[70] B. Spencer and L. Bergman, *On the numerical solution of the Fokker-Planck equation for nonlinear stochastic systems*, Nonlinear Dynamics, 4 (1993), pp. 357–372.

[71] Y. Sun and M. Kumar, *Numerical solution of high dimensional stationary Fokker–Planck equations via tensor decomposition and Chebyshev spectral differentiation*, Computers & Mathematics with Applications, 67 (2014), pp. 1960–1977.

[72] A. F. Thompson and W. R. Young, *Scaling baroclinic eddy fluxes: Vortices and energy balance*, Journal of physical oceanography, 36 (2006), pp. 720–738.

[73] S. Thual, A. J. Majda, N. Chen, and S. N. Stechmann, *Simple stochastic model for El Niño with westerly wind bursts*, Proceedings of the National Academy of Sciences, (2016), p. 201612002.

[74] G. K. Vallis, *Atmospheric and oceanic fluid dynamics*, Cambridge University Press, 2017.

[75] U. von Wagner and W. V. Wedig, *On the calculation of stationary solutions of multi-dimensional Fokker–Planck equations by orthogonal functions*, Nonlinear Dynamics, 21 (2000), pp. 289–306.

[76] M. Wand, *Error analysis for general multtvariate kernel estimators*, Journal of Nonparametric Statistics, 2 (1992), pp. 1–15.

[77] M. P. Wand and M. C. Jones, *Multivariate plug-in bandwidth selection*, Computational Statistics, 9 (1994), pp. 97–116.