# PREDICTING FAT-TAILED INTERMITTENT PROBABILITY DISTRIBUTIONS IN PASSIVE SCALAR TURBULENCE WITH IMPERFECT MODELS THROUGH EMPIRICAL INFORMATION THEORY[*]

DI QI[†] AND ANDREW J. MAJDA[‡]

**Abstract.** Modelling extreme events is a central issue in climate science and engineering. The capacity of imperfect models to capture intermittent behavior with fat-tailed probability distributions of a passive scalar field advected by turbulent flow systems is investigated here. We consider the effects with complicated flow systems including strong nonlinear and non-Gaussian interactions, and construct much simpler and cheaper imperfect models with model error to capture the crucial statistical features in the stationary tracer field. The Lorenz '96 (L-96) system is utilized as a test model to generate the turbulent advection flow field. Tracer statistics under this L-96 flow field are analyzed both theoretically and numerically, and strong intermittent fat tails can be observed in different dynamical regimes of the flow system with distinct statistical features. The complexity and large computational expense in resolving the true advection flow require the introduction of simpler and more tractable imperfect models which still maintain the ability to capture the key intermittent features in the tracer field. The simplest linear stochastic models containing no positive Lyapunov exponents are proposed here to approximate the tracer advected by the original L-96 system with large degrees of internal instabilities. It is demonstrated that the prediction skill of this imperfect linear model can be greatly improved through fitting the autocorrelation functions with empirical information theory. A systematic framework of measuring the autocorrelation function under spectral representation with the help of empirical information theory is developed, and the optimal model parameters under this unbiased information measure can be achieved easily in a training phase before running the predictions. This imperfect model using optimal parameters achieved through the information-theoretic framework is tested in a variety of dynamical regimes of the L-96 system. Uniformly high skill of the optimal model is displayed in accurately capturing the crucial tracer statistical features in a stationary statistical steady state, especially in getting accurate intermittent fat tails in tracer density distributions. This information framework for tuning autocorrelation functions can be further generalized to more complicated turbulent models and should have many applications.

**Key words.** Turbulent diffusion, passive scalar field, intermittency, Gaussian velocity model, information metric.

**AMS subject classifications.** 76F25, 62B10, 86A32.

## 1. Introduction

Turbulent diffusion models of passive tracers have numerous applications in geophysical science and engineering. These applications range from the spread of contaminants or hazardous plumes in environmental science, to the behavior of anthropogenic and natural tracers in climate change science, and many related areas [8,9,28,33]. The scalar field $T(\mathbf{x},t)$ describes the concentration of the passive tracer immersed in the fluid which is carried with the local fluid velocity but which does not itself significantly influence the dynamics of the fluid. We will consider this issue in the context of the evolution of this scalar field through the joint effect of turbulent advection, diffusion, and usually uniform damping

$$\frac{\partial T}{\partial t} + \mathbf{v} \cdot \nabla T = -d_T T + \kappa \Delta T. \tag{1.1}$$

The tracer field is passively advected by a velocity field $\mathbf{v}(x,y,t)$ which could be the solution of the Navier–Stokes equation with high Reynolds number, and it will typically also have some intrinsic self-diffusivity $\kappa$ due to microscopic Brownian motion and a linear damping $d_T$ due to the friction. One key feature of great interest in the tracer turbulent model (1.1) is the existence of intermittency, which can be observed in atmosphere observation data [33], laboratory experiments [40], and numerical simulations of idealized models [4, 25, 27, 28], resulting in fat tails in the density distribution functions and random large spikes during time evolution of the passive tracers. It is challenging to understand the true mechanisms behind the intermittent phenomena due to the involvement of many complex factors in the real advection flow system $\mathbf{v}$. So it is useful and necessary to develop simplified models which still have the ability to create intermittency and fat tails in distribution.

A special form of the velocity field $\mathbf{v}$ which is a superposition of a spatially uniform but possibly temporally fluctuating cross-sweep in the $x$-direction, and a random shear flow (with fluctuations possibly in both time and spatial $x$-direction) in the $y$-direction

$$\mathbf{v}(x,y,t) = (U(t), v(x,t)), \tag{1.2}$$

has been proposed by Majda *et al.* [25, 28] and tested on simple mathematical models [11, 12, 14, 24]. Assume the existence of a background mean gradient for the tracer varying in only the $y$-variable and a tracer fluctuation component dependent with only the $x$-variable

$$T(\mathbf{x},t) = T'(x) + \alpha y. \tag{1.3}$$

Together with the tracer dynamics (1.1) and the simplified flow field (1.2), the fluctuation part of the tracer $T'$ satisfies the following dynamics

$$\frac{\partial T'}{\partial t} + U(t)\frac{\partial T'}{\partial x} = -\alpha v(x,t) - d_T T' + \kappa \frac{\partial^2 T'}{\partial x^2}. \tag{1.4}$$

Despite their simplicity, the model (1.4) in random shear flow with a mean sweep can capture and preserve key features for various inertia range statistics for turbulent diffusion [1, 2, 4, 13, 32] including intermittency. Even for roughly Gaussian velocity fields $\mathbf{v}$ in (1.2) as observed in turbulent flows, the linear scalar field can experience rare but very large fluctuations in amplitude, and its statistics can depart significantly from Gaussianity displaying fatter tails representing the intermittency [6, 16–18, 39]. Explicit formulations about the statistical solutions of the tracers have been derived in [25] under this simplified flow system, and a rigorous mathematical proof about the intermittent fat tails in tracer distributions has been achieved recently in [32].

Although important and instructive in investigating the tracer intermittency advected by a background flow $\mathbf{v}$ with Gaussian random functions, nonlinear and non-Gaussian features in the flow components are unavoidable and ubiquitous especially in realistic turbulent flows. In this paper, we investigate the effects from a nonlinear advection flow on the steady state passive tracer intermittency, and especially the errors and performances of various imperfect approximation models are tested in a variety of turbulent regimes. As a representative testbed, we take the 40-dimensional Lorenz 96 (L-96) system first introduced in [20] to simulate the background flow. The L-96 system is designed to mimic baroclinic turbulence in the midlatitude atmosphere with the effects of energy conserving nonlinear advection and dissipation, displaying a wide range of distinct dynamical regimes from Gaussian to extremely non-Gaussian statistics,

which makes it a desirable test model to begin with. With the nonlinear L-96 system as the background advection system, we analyze and compute the tracer energy spectra and distribution functions under a variety of distinct statistical regimes. Transition between nearly Gaussian behavior and fat tailed highly intermittent distributions can be observed as the tracer parameter and flow structure vary. Additionally the high dimensional turbulent system including an irreducible attractor in real applications adds extra difficulties in resolving the exact high-order statistics in the system from a wide range of scales. Therefore in constructing practical models for the turbulent systems, proper Galerkin truncation with a high wave-number cutoff is always introduced to the true system and only the statistics in a finite number of large-scale modes are calculated. It is important to notice the conflict between the small truncation size which may add severe model error, and the limited ensemble size which adds an impractical computational load. The inherent difficulties in truncation size and ensemble size set unavoidable obstacles in resolving the exact flow field with accuracy. We show and discuss these difficulties also under the L-96 example.

The difficulties in resolving the complex advection flow raise the questions about whether we can propose simpler imperfect models to approximate the advection flow field $\mathbf{v}$, and what is the prediction skill for tracer statistics given this imperfect field with model error. Due to the central quantity of interest is the tracer statistics rather than the advection field, the replacement of $\mathbf{v}$ by a simpler imperfect flow field is practical and advantageous in real applications. Here we begin with the simplest possible approximation model with only Gaussian linear dynamics in the advection flow, and check its ability in capturing tracer statistical structures. The Gaussian velocity stochastic models display attractive characteristics that are much more controllable with explicit solutions and require much smaller ensemble size to achieve accuracy in statistics. On the other hand, it is also challenging by applying the Gaussian linear models with no positive Lyapunov exponents to estimate the non-Gaussian flow field including various degrees of internal instabilities. Therefore a systematic procedure in calibrating the imperfect model parameters is required. It turns out through theoretical analysis that the imperfect model prediction skill about the autocorrelation function plays an important role in determining the structure of tracer statistics. An information-theoretic framework [21,22,24,30] is then proposed to train the imperfect model parameters in a training phase so that the model predicted stationary process can possess the least biased estimation in energy and autocorrelation function compared with the truth. Through experiments with the L-96 system in various regimes, it proves that the performance of these linear stochastic models can be greatly improved through this proposed tuning strategy under a proper information metric.

The following part of this paper is summarized as follows: Section 2 discusses some theoretical approximations and numerical experiments about the true model under the exact L-96 system. The passive tracer statistics including the steady state energy spectra and probability distributions are analyzed and computed. Also the errors from pure Galerkin truncation model and the effects with ensemble size are shown and discussed. We turn to the issue about using imperfect approximation models to estimate the true tracer distributions in Section 3. Simple linear stochastic systems are proposed. Two different strategies to estimate the model parameters are applied. Importantly, we show that the optimal model parameters through the information framework can be achieved by adopting the spectral representation of the stationary random field and autocorrelation function. In Section 4, the improvement in models' skill of capturing intermittency through this information-theoretic framework is tested using the L-96

model under various dynamical regimes. We conclude this paper with simple discussion about the direction of future research in Section 5.

## 2. Passive tracer statistics with nonlinear advection flow

We begin with some discussion and analysis about the exact statistical features about the scalar tracer fields advected by the flow generated by the L-96 system. It is interesting to see that strong non-Gaussian distributions with fat tails can be generated in the tracer modes despite the sub-Gaussian statistics in the L-96 flow. Besides, it is also useful to check the possible errors when Galerkin truncation and limited number of particles are applied in resolving the complex turbulent system, which is actually common strategies that are used in realistic applications.

**2.1. L-96 system as the advection flow.**    We consider the situation when the passive tracer field is advected by a nonlinear non-Gaussian flow field generated from the Lorenz '96 (L-96) system. The L-96 system [20] is a 40-dimensional dynamical system with state variables $\mathbf{u} = (u_0, u_1, \ldots, u_{J-1})^T$ such that

$$\frac{du_j}{dt} = (u_{j+1} - u_{j-2}) u_{j-1} - d(t) u_j + F(t), \quad j = 0, 1, \ldots, J-1, \, J = 40. \quad (2.1)$$

Periodic boundary condition $u_J = u_0$ is applied to simulate flows at a constant latitude circle.   Nonlinearity comes from the bilinear quadratic form $B_j(\mathbf{u}, \mathbf{u}) = (u_{j+1} - u_{j-2}) u_{j-1}$ as the first term on the right-hand side of (2.1), which conserves energy through $\mathbf{u} \cdot \mathbf{B}(\mathbf{u}, \mathbf{u}) = 0$ (note that this property is crucial in the energy analysis below and will be exploited further in the tracer statistics). A linear damping term and a deterministic forcing part are also applied at each grid point in the system. Especially, by changing the amplitude of the external forcing $F$, the L-96 system displays a wide range of different dynamical regimes ranging from weakly chaotic ($F = 5$), strongly chaotic ($F = 8$), to finally full turbulence ($F = 16$) with varying statistics, which makes it a desirable test model to simulate the turbulent flow in the first place (see next subsection for various illustrations about the statistical features of L-96 system in different dynamical regimes). Due to a variety of interior instabilities (which can be illustrated by the number of positive Lyapunov exponents) as well as the uncertainty in the initial condition, it is proper to use a random field $\mathbf{u}$ rather than one deterministic trajectory to describe the time evolution of this system.

The advection flow field $\mathbf{v} = (U(t), v(x, t))$ then is constructed from the L-96 system solution. Note that the system is homogeneous and transition invariant along each grid point, so standard Fourier basis $\mathbf{e}_k = \left\{ e^{2\pi i k j / J} \right\}_{j=0}^{J-1}$ naturally becomes the empirical orthogonal functions (EOFs) of the system [30]. The state variables of the system can be decomposed under Fourier basis as

$$\mathbf{u}(x, t) = \bar{\mathbf{u}}(t) + \sum_{k=-J/2+1}^{J/2} \hat{u}_k(t) \mathbf{e}_k(x), \quad \langle \hat{u}_k \rangle = 0, \quad \hat{u}_{-k} = \hat{u}_k^*. \quad (2.2)$$

Here $\langle \cdot \rangle$ can be viewed as the ensemble average. We construct the passive tracer fields (1.1) nonlinearly advected by the flow generated through the L-96 system. The gradient cross-sweeping component $U(t)$ is from the mean state with randomness from zero mode, while the shearing component $v(x_j, t)$ simulated by the flow fluctuation modes with varying values at each grid point. Therefore, the advection flow field with proper nonlinear dynamics can be defined from the components in L-96 system as

$$U(t) = \bar{u}(t) + \hat{u}_0(t), \qquad v(x_j, t) = \sum_{k \neq 0} \hat{u}_k(t) e^{2\pi i k x_j}. \quad (2.3)$$

In the following parts of the paper, we will focus on the statistical features of the scalar tracer field in stationary steady state. To make sure the system converges to the final stationary state, that is, $\bar{u}(t) \rightarrow \bar{u}_{\infty}, r_k(t) = \left\langle |\hat{u}_k|^2(t) \right\rangle \rightarrow r_{k,\infty}$ as $t \rightarrow \infty$, we consider the simplified dynamics of (2.1) with constant damping and forcing terms $d \equiv d(t), F \equiv F(t)$. Before investigating the imperfect model approximations of the true system, we would like to investigate the general mathematical properties for both the advection flow $\mathbf{v}$ and the passive tracer field $T$ in the first place. Hopefully, this can reveal the contributions of each mode in the steady state flow system for the non-Gaussianity and the possible sources of fat tail structures in the tracer probability distributions.

**2.1.1. Advection flow in steady state.** The exact dynamical equations for each mode in the shearing flow $\hat{u}_k$ and the mean gradient $U$ can be derived from the L-96 system (2.1) as

$$\frac{dU}{dt} = -dU(t) + \sum_{k \neq 0} \Gamma_k |\hat{u}_k|^2(t) + F, \tag{2.4}$$

$$\begin{aligned}
\frac{d\hat{u}_k}{dt} &= -d\hat{u}_k + \left( e^{2\pi i \frac{k}{J}} - e^{-2\pi i \frac{2k}{J}} \right) U(t) \hat{u}_k \\
&\quad + \sum_{m \neq 0} \hat{u}_{k+m} \hat{u}_m^* \left( e^{2\pi i \frac{2m+k}{J}} - e^{-2\pi i \frac{m+2k}{J}} \right), \quad k = 1, \ldots, J/2, \tag{2.5}
\end{aligned}$$

where we define the energy transfer rate $\Gamma_k = \cos \frac{4\pi k}{J} - \cos \frac{2\pi k}{J}$. The derivation of the above formulas is from the standard procedure for systems with uncertainties by projecting the original equation (2.1) onto each EOF mode, and exploiting the properties in homogeneous solutions. We neglect the detailed steps and the standard solutions for general systems can be found in [10, 30]. The cross-sweep field $U$ is forced by the combined effects from each fluctuation mode $\sum_{k \neq 0} \Gamma_k |\hat{u}_k|^2$, and conversely the shearing flow is advected by the mean drift through the second term in the first line in (2.5). One important property of the above system is the conservation of the quadratic form $\mathbf{u} \cdot \mathbf{B}(\mathbf{u}, \mathbf{u}) = \mathbf{0}$, therefore the last part in (2.5) will vanish as we calculate the dynamics for the total fluctuations $\sum_k |\hat{u}_k|^2$. Specifically, by multiplication of $\hat{u}_k$ on both sides of (2.5) and taking summations about all the modes, the total energy in the shearing flow follows the dynamics

$$\frac{d}{dt} \sum_k |\hat{u}_k|^2 = -2d \sum_k |\hat{u}_k|^2 - 2\bar{u} \sum_k \Gamma_k |\hat{u}_k|^2.$$

Similarly, by multiplication of $U$ on both side of (2.4), the corresponding energy in the cross-sweep becomes

$$\frac{d}{dt} U^2 = -2dU^2 + 2U \sum_k \Gamma_k |\hat{u}_k|^2(t) + 2UF.$$

Considering the two above formulas together, it is convenient to define the *chaotic energy* $\mathcal{E} = U^2 + \sum_{k \neq 0} |\hat{u}_k|^2$ as a combination of mean and fluctuation energy in the flow $(U, v)$. The dynamical equation for the chaotic energy $\mathcal{E}$ can be derived as

$$\frac{d\mathcal{E}}{dt} = -2d\mathcal{E} + 2\hat{u}_0 \sum_{k \neq 0} \Gamma_k r_k + 2UF. \tag{2.6}$$

Note that the defined chaotic energy $\mathcal{E}$ is still a random process with uncertainties contained in each mode $\hat{u}_k, k \in \{0, 1, \ldots\}$. And if we further define the statistical energy of the system as the ensemble average over each realization of the chaotic energy $E = \langle \mathcal{E} \rangle = \bar{u}^2 + \sum_k r_k$, we come back to the same energy equation as derived in [30]

$$\frac{dE}{dt} = -2dE + 2\bar{u}F. \tag{2.7}$$

The dynamical equation (2.6) illustrates a crucial property that the uncertainty in $\mathcal{E}$ can be uniquely determined by the zero mode since the only source of randomness come from $\hat{u}_0$ on the right-hand side of (2.6). To check this property in detail, we can calculate the formal solution of the chaotic statistical energy equation (2.6).

As stated above, we want to focus on the model statistics in stationary steady state. Then solve the system (2.6) from time $t_0$ to $t$ through Duhamel's formula and next let the initial time $t_0 \to -\infty$, the stationary solution for the chaotic energy can be achieved

$$
\begin{aligned}
\mathcal{E} = U^2 + \sum_{k \neq 0} |\hat{u}_k|^2 &\to 2 \sum_k \Gamma_k r_{k,\infty} \int_{-\infty}^t e^{-2d(t-s)} \hat{u}_0(s)\, ds + 2F \int_{-\infty}^t e^{-2d(t-s)} U(s)\, ds \\
&= 2(d\bar{u}_\infty - F) \int_{-\infty}^t e^{-2d(t-s)} \hat{u}_0(s)\, ds + 2F \int_{-\infty}^t e^{-2d(t-s)} U(s)\, ds \\
&= 2d\bar{u}_\infty \int_{-\infty}^t e^{-2d(t-s)} \hat{u}_0(s)\, ds + \frac{\bar{u}_\infty F}{d}.
\end{aligned}
\tag{2.8}
$$

The first equality in the second row just applies the statistical relation $\sum_{k \neq 0} \Gamma_k r_{k,\infty} = d\bar{u}_\infty - F$ between the mean and variances in steady state, and the second equality separates $U = \bar{u} + \hat{u}_0$. Explicitly from the stationary steady state solution, it can be concluded that the total uncertainties including all the higher order non-Gaussian statistics due to the nonlinear dynamics in the shearing flow $\sum_{k \neq 0} |\hat{u}_k|^2$ (or the one-point statistics in the shearing flow) can be solely determined by the randomness in the zero mode $\hat{u}_0$. We summarize the results from above analysis as the following proposition.

PROPOSITION 2.1.    *Define the chaotic statistical energy of the advection flow system (2.4) and (2.5) of* $\mathbf{v} = (U(t), v(x,t))$ *from the L-96 system as* $\mathcal{E} = U^2 + \sum_{k \neq 0} |\hat{u}_k|^2$. *The dynamical evolution of the chaotic statistical energy can be described by (2.6) as*

$$\frac{d\mathcal{E}}{dt} = -2d\mathcal{E} + 2\hat{u}_0 \sum_{k \neq 0} \Gamma_k r_k + 2UF,$$

*with uncertainty only from the random cross-sweeping part in zero mode* $U = \bar{u} + \hat{u}_0$. *And in the stationary state with uniform damping* $d$ *and forcing term* $F$, *the random field of chaotic statistical energy* $\mathcal{E}$ *approaches the statistical steady state (2.8)*

$$\mathcal{E} = U^2 + \sum_{k \neq 0} |\hat{u}_k|^2 \to 2d\bar{u}_\infty \int_{-\infty}^t e^{-2d(t-s)} \hat{u}_0(s)\, ds + \frac{\bar{u}_\infty F}{d},$$

*almost surely as* $t_0 \to -\infty$. *The randomness and non-Gaussianity in the one-point statistics of the shearing flow* $v(x,t)$ *then is determined by the probability distribution of the sweeping part* $U$ *in stationary steady state.*

**2.1.2. Tracer structure in steady state.**     Now we turn to the advection-diffusion equation of the scalar tracer (1.4) under the Fourier expansion $T' = \sum_k \hat{T}_k e^{2\pi i k x}$ passively advected by the nonlinear flow **v** as described above in (2.4) and (2.5). Under the Fourier representation, the dynamical equation for each Fourier component of the tracer can be written componentwise as [25]

$$\frac{d\hat{T}_k}{dt} = (-\gamma_{T_k} + i\omega_{T_k})\hat{T}_k - \alpha \hat{u}_k, \qquad k = 1, 2, \ldots, J/2, \tag{2.9}$$

with $\gamma_{T_k} = \kappa k^2 + d_T$, and $\omega_{T_k} = -U(t)k$. The solution for the tracer component in each Fourier mode can be achieved through the standard procedure with the help of the Green's function

$$\hat{T}_k = -\alpha \int_{t_0}^t G_k(s,t)\hat{u}_k(s)\,ds, \tag{2.10}$$

where the Green's function is defined as

$$G_k(s,t) = \exp(-\gamma_{T_k}(t-s) - ikJ(s,t)), \qquad J(s,t) = \int_s^t U(\tau)\,d\tau. \tag{2.11}$$

Note that under this construction, the energy in the zero mode of the tracer $\hat{T}_0$ goes to zero quickly, as a result in steady state then we can just focus on the energy in each fluctuation mode $k > 0$. Explicit statistical solutions of (2.9) together with the corresponding probability density functions with fat tails under linear Gaussian flow field $(U, v)$ have been investigated extensively in detail from different directions [4, 25, 32]. Whereas when it come to the case including non-Gaussian features in the advection flow, the situation becomes more sophisticated. First, explicit solutions for each flow mode $\hat{u}_k$ as well as the sweeping component $U$ are unavailable for further analysis. Second, the effects from higher order moments (about $U$, $v$, and their cross-correlations) need to be taken into account in the analysis, which could play quite significant roles in the final tracer distributions. Furthermore, resolving the exact nonlinear statistical characteristics of each mode $\hat{u}_k$ with accuracy is expensive or computational forbidden even under numerical approaches. So in the following sections, we would like to investigate the accuracy of various methods with Galerkin truncations and different kinds of imperfect Gaussian stochastic models. Of course with the cut-off of higher order statistics in these imperfect approximations, model errors will be introduced into final estimates of the tracer statistics in $\hat{T}_k$ and more careful manipulations are required. In this section, we first check the contributions of each mode in determining the stationary one-point statistics of the tracer $T$ and estimate proper bounds for this random field, which can serve as a guideline for the designing of imperfect approximation models in the following sections.

In the *one-point statistics* of the scalar field $T(x) = \sum_k \hat{T}_k e^{2\pi i k x}$, we focus on the statistics at each grid point $x_j$ and ignore the covariances between different points $x_i, x_j, i \neq j$. Due to the translation invariant property of the flow field **v**, one-point statistics at different points share the same distribution. Specifically, conditional on each realization of the horizontal sweep $U_{s<t}$ up to time $t$ and using the expression in

(2.10), the one-point energy in variance follows:

$$\mathbb{E}\left(T^2\left(x,t\right)|U_{s<t}\right)=\mathbb{E}\left(\sum_{k,l}\hat{T}_k\hat{T}_l^*e^{2\pi i(k-l)x}\,|\,U_{s<t}\right)$$

$$=\sum_k\mathbb{E}\left(\left|\hat{T}_k\right|^2|U_{s<t}\right)=\mathbb{E}\left(\sum_k\left|\hat{T}_k\right|^2|U_{s<t}\right).$$

Therefore we consider the distributions of $T_{1\mathrm{pt}}^2\sim\sum_k\left|\hat{T}_k\right|^2$ ('$\sim$' defined as mean-square equivalence) to analyze the statistical properties and non-Gaussian features for the one-point statistics of $T$. From the pathwise solution of the tracer component in (2.10), we can write the steady state fluctuation in each component as

$$\left|\hat{T}_k\right|^2=\alpha^2\left|\int_{-\infty}^t\int_{-\infty}^t G_k\left(s,t\right)G_k\left(s',t\right)\hat{u}_k\left(s\right)\hat{u}_k^*\left(s'\right)dsds'\right|$$

$$=\alpha^2\left|\int_{-\infty}^t\int_{-\infty}^t\exp\left(-\gamma_{T_k}\left(2t-s-s'\right)\right)\exp\left(-ik\int_s^{s'}U\left(\tau\right)d\tau\right)\hat{u}_k\left(s\right)\hat{u}_k^*\left(s'\right)dsds'\right|$$

$$\leq\alpha^2\int_{-\infty}^t\int_{-\infty}^t\left|\exp\left(-\gamma_{T_k}\left(2t-s-s'\right)\right)\exp\left(-ik\int_s^{s'}U\left(\tau\right)d\tau\right)\right|\left|\hat{u}_k\left(s\right)\hat{u}_k^*\left(s'\right)\right|dsds'$$

$$\leq C\int_{-\infty}^t\exp\left(-\gamma_{T_k}\left(t-s\right)\right)\int_0^\infty\exp\left(-\gamma_{T_k}r\right)|\hat{u}_k\left(s\right)\hat{u}_k^*\left(s+r\right)|drds\sim\tilde{C}\int_0^\infty|\hat{u}_k\left(0\right)\hat{u}_k^*\left(r\right)|dr.$$

$$(2.12)$$

The last line above just uses a change of variable $r=s'-s$, and in stationary state $|\hat{u}_k\left(s\right)\hat{u}_k^*\left(s+r\right)|$ and $|\hat{u}_k\left(0\right)\hat{u}_k^*\left(r\right)|$ are equivalent in distribution. This implies that the statistics in $\hat{T}_k$ is non-local in time involving the lagged-covariance of $\hat{u}_k$ (or the auto-correlation function) besides the energy in each mode. The above inequality holds for each realization of $\hat{u}_k$ and $\hat{T}_k$. Note that the inner integrand is related to the *absolute decorrelation time* defined as

$$T_d=\int_0^\infty\frac{\mathbb{E}\left|\hat{u}_k\left(0\right)\hat{u}_k^*\left(r\right)\right|}{r_k}dr. \qquad (2.13)$$

So it is reasonable to *assume* that there exists a constant $C'>0$ so that the random process is bound by

$$\int_0^\infty\exp\left(-\gamma_{T_k}r\right)|\hat{u}_k\left(s\right)\hat{u}_k^*\left(s+r\right)|dr\leq C'\left|\hat{u}_k\left(s\right)\right|^2T_d, \qquad (2.14)$$

almost surely. Combining the pathwise estimation (2.12) and the bound (2.14), we get the estimate for tracer fluctuation in each mode

$$\left|\hat{T}_k\right|^2\leq CT_d\int_{-\infty}^t\exp\left(-\gamma_{T_k}\left(t-s\right)\right)\left|\hat{u}_k\left(s\right)\right|^2ds, \qquad a.s.$$

Now we come to the one-point statistics of the tracer $\sum_k\left|\hat{T}_k\right|^2$, and note from Proposition 1 that stationary steady state statistics of $\sum_k|\hat{u}_k|^2$ is only a function of $U$. Therefore, the one-point statistics of the tracer are bounded from above only dependent on the fluctuations in the mean sweep $U$ as

$$\sum_k\left|\hat{T}_k\right|^2\leq CT_d\int_{-\infty}^t e^{-d_T(t-s)}\sum_k|\hat{u}_k\left(s\right)|^2ds$$

$$= C_1 \bar{u}_\infty T_d \int_{-\infty}^t e^{-d_T(t-s)} \int_{-\infty}^s e^{-2d(s-s')} \hat{u}_0(s') ds' ds + C_2(U) T_d, \qquad (2.15)$$

almost surely. In [32], estimated bounds for intermittency about the one-point fluctuations in $T$ through simpler models are investigated. Finally we summarize the one-point tracer statistics we achieved by the following proposition.

PROPOSITION 2.2 (Tracer variance statistics). *Consider the passive tracer field (2.9) advected by nonlinear flow (2.4) and (2.5) from the L-96 system. Then in the stationary steady state, the fluctuations $T_{\text{1pt}}^2 \sim \sum_k \left|\hat{T}_k\right|^2$ of the one-point statistics of the tracer can be bounded by the statistics in the sweeping flow $U = \bar{u} + \hat{u}_0$. That is,*

$$\sum_k \left|\hat{T}_k\right|^2 \leq C_1 \bar{u}_\infty T_d \int_{-\infty}^t e^{-d_T(t-s')} \int_{-\infty}^{s'} e^{-2d(s'-s)} \hat{u}_0(s) ds ds' + C_2(U) T_d, \quad a.s.$$

*if the estimation in (2.14)*

$$\int_0^\infty \exp(-\gamma_{T_k} r) |\hat{u}_k(s) \hat{u}_k^*(s+r)| dr \leq C' |\hat{u}_k(s)|^2 T_d,$$

*holds almost surely with some positive constant $C'$.*

**2.2. Intermittency in tracer probability density functions.** In the previous subsection, we estimate the possible bounds for the single-point statistics of the passive tracer distributions in a statistically steady state with the help of the energy equation. Still it is necessary to check whether intermittency can be generated in the tracer probability density functions in both physical domain and the spectral modes under the advection flow generated by the L-96 system. As we have described previously, L-96 offers an desirable testbed in the first stage with varying dynamical regimes for both Gaussian and non-Gaussian statistics. Specifically, we choose three representative statistical regimes with $F = 5$ (weakly chaotic), $F = 8$ (strongly chaotic), and $F = 16$ (fully turbulent). Strong non-Gaussianity with skewed distributions is displayed in $F = 5$ regime, while the statistics appear increasingly like Gaussian in more turbulent regimes $F = 8, 16$ (see [31, 34] for more discussions about the model statistics). To capture the statistical characteristics in the true flow with accuracy, we run a Monte-Carlo simulation with large enough ensemble size of 10000. In all these dynamical regimes, we choose fixed mean gradient $\alpha = 2$ and diffusivity $\kappa = 0.001$, and the damping rate is chosen as $d_T = 1, 5, 10$ correspondingly for each dynamical regime. Then, we observe the corresponding tracer variance spectra and the intermittencies in tracer PDFs given distinct statistical structures in the advection flow $\mathbf{v}$.

In figures 2.1, 2.2, and 2.3, we compare the tracer statistics in these three representative dynamical regimes $F = 5, 8, 16$ in comparison with the flow statistics. The steady state energy spectra in Fourier domain for both the flow and tracer are compared in the first row with the representative probability density functions following. We also draw a Gaussian distribution with the same variance to emphasize the non-Gaussianity in each case. For the strongly non-Gaussian regime $F = 5$, there is one dominant mode $k = 7$ which contains most of the energy in the system. Therefore this mode is also dominant in the tracer spectrum. The distribution for the horizontal cross-sweep $U$ (represented by $k = 0$ mode) is strongly skewed. Correspondingly, strong non-Gaussian features with fat tails are generated in the tracer distributions especially for the dominant mode. In regimes $F = 8$ and $F = 16$, the statistics in the flow become increasingly

more Gaussian-like (or sub-Gaussian) companied with more uniform energy spectra. Still obvious intermittencies are observed in the tracer PDFs with fat tails in both physical domain and the principal modes in spectral domain despite the near-Gaussian advection flow in $\hat{u}_0$ (note this distinct statistical feature in comparison with the regime $F=5$). In the spectral domain, the first two modes $\hat{T}_1, \hat{T}_2$ perform like Gaussian while the most energetic ones $\hat{T}_7, \hat{T}_8, \hat{T}_9$ display strong intermittencies in the PDFs. And in the one-point statistics in the physical domain, the flow distributions are all sub-Gaussian while fat tails are generated in the tracer distributions for all the cases.

Finally, we check the change in the tracer spectra and PDFs as the tracer damping rate $d_T$ changes. We choose the typical regime with $F=8$ and change the damping rate in the range $d_T=0, 0.1, 0.5, 5, 20$. The results for the steady state energy spectra and the PDFs in one-point statistics in physical domain as well as the most energetic spectral mode are compared in Figure 2.4. As expected, the tracer modes are more energetic and the spectrum is dominant by the first few low frequency modes as $d_T \ll 1$, while the energy reduces and the spectrum becomes similar to that of the advection flow as $d_T \gg 1$. Intermittency is generated with moderate damping rates and the PDFs becomes Gaussian or sub-Gaussian in the two limits $d_T \to 0$ and $d_T \to \infty$.

**2.3. Model errors from Galerkin truncation and ensemble size.** In the above examples with the simple L-96 system as the background flow, it is observed that strongly non-Gaussian features with fat tails can be generated in the passive tracer distribution functions even under a near Gaussian advection flow. Further to cases about high or infinite dimensional turbulent flows in realistic applications, it becomes impossible to resolve the detailed statistics in small scale dynamics considering both the limited available computational resources and the insufficient understanding of the turbulent systems. Due to these practical restrictions, a high wave-number truncation is usually applied ignoring the effects from the small scale modes with high frequency and small amount of energy. However it is noted in [29, 38] that the *inverse energy cascade* from the neglected small scales to large scales also plays an crucial role in the dynamical structures in many situations. Furthermore for practical implementation, only limited small ensemble size is affordable for resolving the statistical structures along each direction especially in the case of large dimensional models. The so-called 'curse of dimensionality' is a well-known obstacle for the development of statistical models with accuracy in capturing non-Gaussian features [3, 23]. Using the 40-dimensional L-96 system as an example, we illustrate the difficulties about the possible errors from the truncation of high wave-number modes and the effects from small ensemble size. It can help us get a more thorough understanding about the statistical structures of the turbulent systems and help construct the much simpler linear models in the next section. It is also useful for offering useful hints about modeling the realistic turbulent systems.

**2.3.1. Errors from truncation of non-energetic modes.** As discussed above, to effectively reduce the computational costs in resolving the entire spectrum of high dimensional turbulent systems, one direct strategy is to apply the standard Galerkin truncation method ignoring all the high frequency modes leaving only the large-scale modes in consideration. Deficiencies in ignoring these less energetic modes have been noticed and discussed in [36, 37]. To investigate the possible errors introduced in this truncation and seek proper way to improve the imperfect model performance, we investigate the possible errors from this simple Galerkin truncation method using the finite dimensional L-96 system. Use the model (2.1) with $J=40$ as the perfect system with full spectrum of Fourier modes $N=J/2$. Then introduce the imperfect
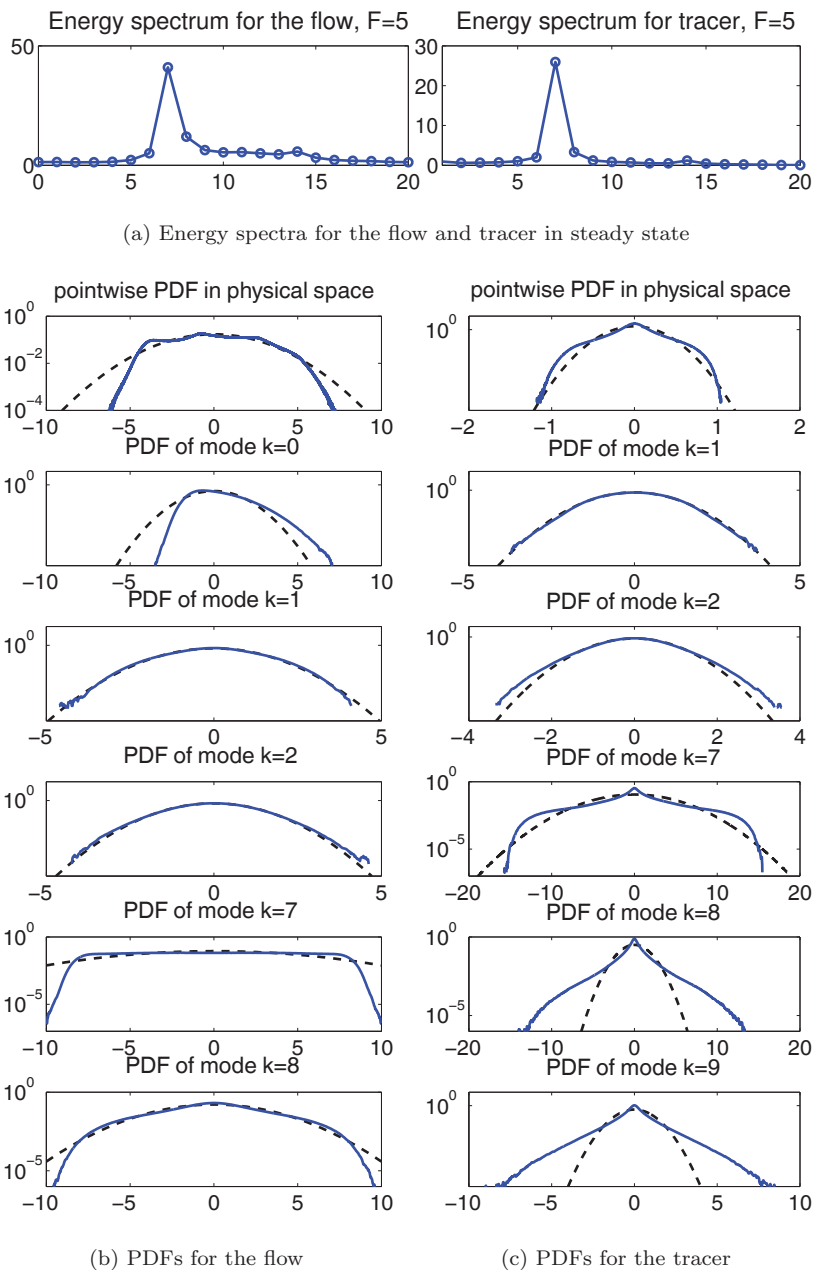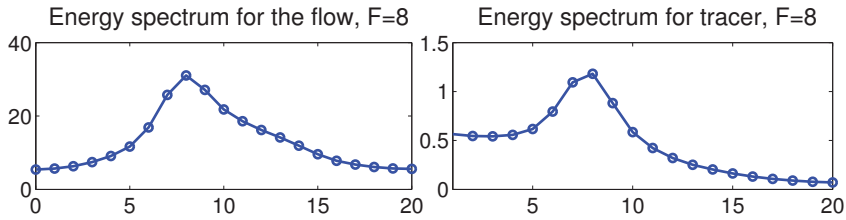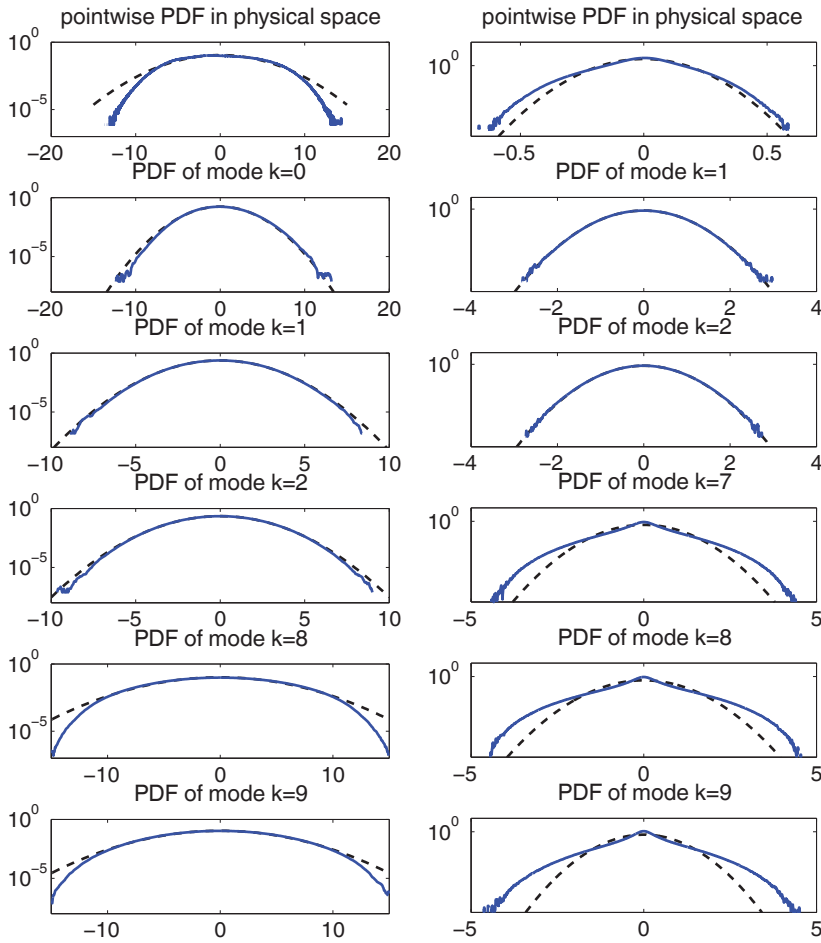
(a) Energy spectra for the flow and tracer in steady state



(b) PDFs for the flow

(c) PDFs for the tracer

Fig. 2.1: *Steady state statistics for the background flow* **u** *from the L-96 system (left column) and the passive tracer $T$ (right column) in weakly chaotic regime $F = 5$ with tracer parameters $d_T = 1, \alpha = 2, \kappa = 0.001$. The first row shows the steady state energy spectra under each Fourier mode for both* **u** *and $T$. The following rows display the PDFs for both pointwise values in physical domain as well as the typical modes in spectral domain. The steady state distribution functions are compared with corresponding Gaussian distributions with the same variance in dashed lines. Note that the zero mode $\hat{u}_0$ is used as the zonal sweep for the tracer dynamics, and the first two modes $\hat{u}_1, \hat{u}_2$ and the most energetic modes $\hat{u}_7, \hat{u}_8, \hat{u}_9$ always have the most important roles in the tracer statistics.*
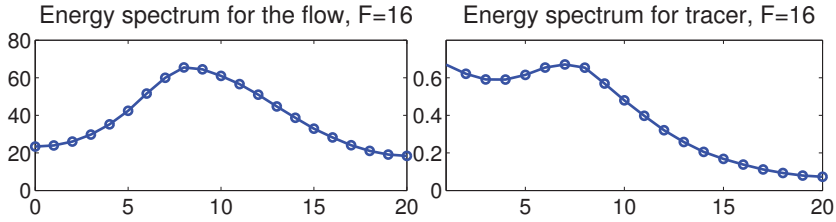
(a) Energy spectra for the flow and tracer in steady state



(b) PDFs for the flow

(c) PDFs for the tracer

Fig. 2.2: *Steady state statistics for the background flow* **u** *from L-96 system (left column) and the passive tracer* $T$ *(right column) in strongly chaotic regime* $F = 8$ *with tracer parameters* $d_T = 5, \alpha = 2, \kappa = 0.001$. *The first row shows the steady state energy spectra under each Fourier mode for both* **u** *and* $T$. *The following rows display the PDFs for both pointwise values in physical domain as well as the typical modes in spectral domain. The steady state distribution functions are compared with corresponding Gaussian distributions with the same variance in dashed lines. Note that the zero mode* $\hat{u}_0$ *is used as the zonal sweep for the tracer dynamics, and the first two modes* $\hat{u}_1, \hat{u}_2$ *and the most energetic modes* $\hat{u}_7, \hat{u}_8, \hat{u}_9$ *always have the most important roles in the tracer statistics.*

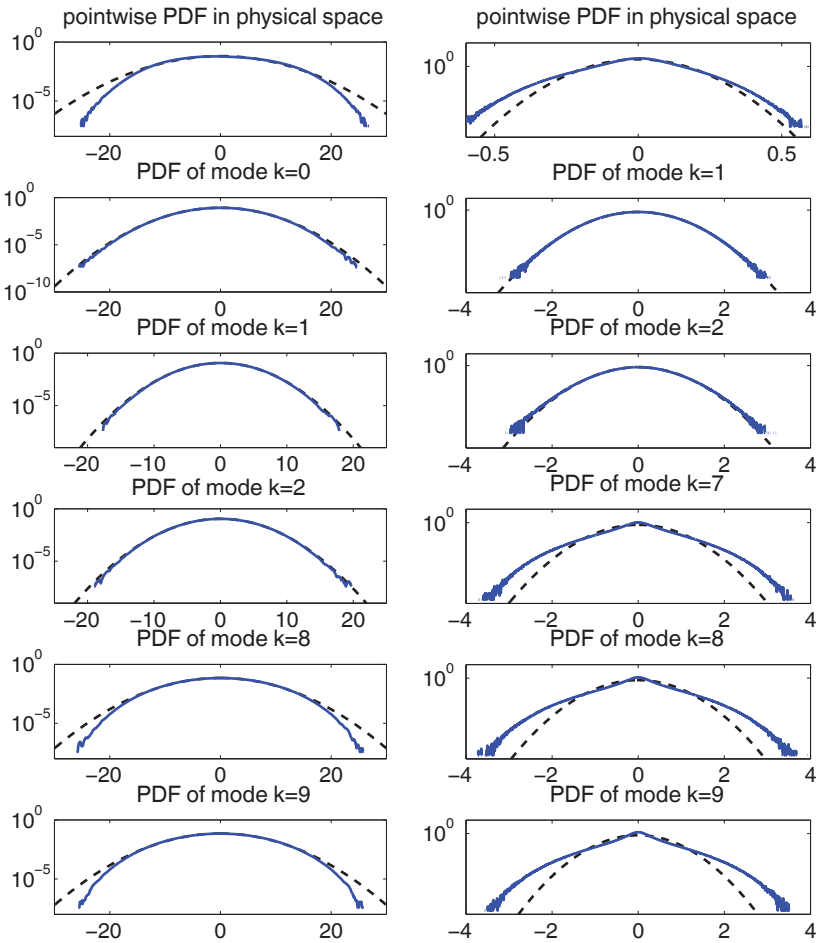(a) Energy spectra for the flow and tracer in steady state



(b) PDFs for the flow

(c) PDFs for the tracer

Fig. 2.3: *Steady state statistics for the background flow* $\mathbf{u}$ *from L-96 system (left column) and the passive tracer* $T$ *(right column) in turbulent regime* $F = 16$ *and tracer parameters* $d_T = 10, \alpha = 2, \kappa = 0.001$. *The first row shows the steady state energy spectra under each Fourier mode for both* $\mathbf{u}$ *and* $T$. *The following rows display the PDFs for both pointwise values in physical domain as well as the typical modes in spectral domain. The steady state distribution functions are compared with corresponding Gaussian distributions with the same variance in dashed lines. Note that the zero mode* $\hat{u}_0$ *is used as the zonal sweep for the tracer dynamics, and the first two modes* $\hat{u}_1, \hat{u}_2$ *and the most energetic modes* $\hat{u}_7, \hat{u}_8, \hat{u}_9$ *always have the most important roles in the tracer statistics.*
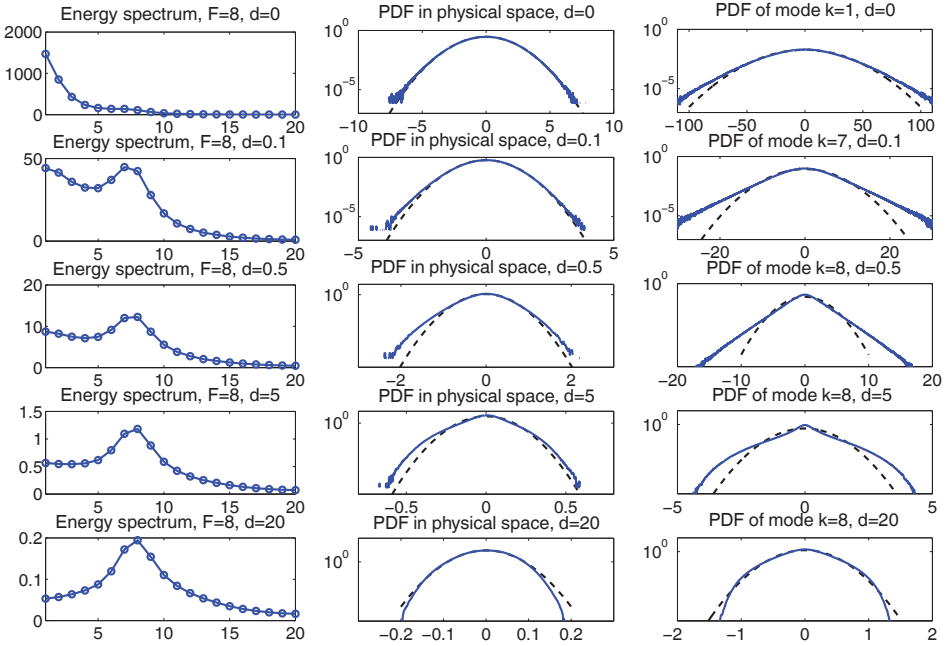
Fig. 2.4: *Comparison of tracer statistics with different damping rate in regime $F = 8$. The parameters are set to be the same as the cases before while the tracer damping rates are changing in the range $d_T = 0, 0.1, 0.5, 5, 20$. The steady state tracer energy spectra for each case as while as the PDFs in the physical domain and the most energetic Fourier mode are compared. The same as before, tracer distributions are compared with Gaussian with the same variance in dashed lines.*

truncated model by ignoring the high Fourier modes with wave-number larger than $K$ ($K < N$). With this high frequency truncation for the relatively simple system, we can test and analyze the performance with this imperfect model and the amplitude of errors introduced by truncation. In the following parts of the paper, we will always refer $K$ as the resolved number of modes and $N$ as the total number of modes.

Denote the set of resolved Fourier modes as $V_K = \{\mathbf{e}_k | k \in \Lambda_K\}$, with index set $\Lambda_K = \{|k| \le K | K < N\}$. The truncated expansion (2.2) of the state variable under the Fourier basis inside $\Lambda_K$ becomes

$$\mathbf{u}(x,t) = \bar{\mathbf{u}}(t) + \sum_{k \in \Lambda_K} \hat{u}_k(t) \mathbf{e}_k(x).$$

Therefore the corresponding background flow (2.4) and (2.5) for the cross-sweep $U(t)$ and shearing $v(x,t)$ in *Galerkin truncated dynamics* becomes

$$\frac{dU}{dt} = -dU(t) + \sum_{k \in \Lambda_K} \Gamma_k |\hat{u}_k|^2(t) + F, \tag{2.16}$$

$$\begin{aligned}
\frac{d\hat{u}_k}{dt} = &-d\hat{u}_k + \left( e^{2\pi i \frac{k}{J}} - e^{-2\pi i \frac{2k}{J}} \right) U(t) \hat{u}_k \\
&+ \sum_{k+m, k \in \Lambda_K - \{0\}} \hat{u}_{k+m} \hat{u}_m^* \left( e^{2\pi i \frac{2m+k}{J}} - e^{-2\pi i \frac{m+2k}{J}} \right), \quad k = 1, \dots, K,
\end{aligned} \tag{2.17}$$

with only the first $K$ Fourier modes under consideration in the interaction term between modes. In Figure 2.5, we compare the model predicted mean $\bar{u}_M$ and single point variance $\mathrm{tr}R_M/J$ under different truncation wave-number $K = 10, 15, 19$ (note that the most energetic modes are always contained in $\Lambda_K$ within these truncation numbers) together with the truth $K = N = 20$. Large model errors appear in all these truncated models even with only one Fourier mode $k = 20$ neglected. Further by checking the steady state energy spectra captured by these truncated models, the prediction skill in the variances of each mode also appears quite poor even with $K = 19$ out of the total 20 modes resolved. The reason for the large errors through the truncation models can be explained by the crucial role in the higher order interactions between the large scale and small scale modes. In the L-96 system, nonlinearity mostly comes from the interaction between modes represented by the second line in equation (2.5). Whereas under the truncated models, only the low frequency mode interactions are included in the dynamics as shown in (2.17). By checking the third order moments in Figure 2.6, it appears obviously that the most important contribution of the interactions happens between modes $\hat{u}_7, \hat{u}_8$ with $\hat{u}_{20}$, which is neglected in the truncated model (2.17). This becomes the inherent obstacle that we always need to keep in mind when the truncated methods are applied to true systems.
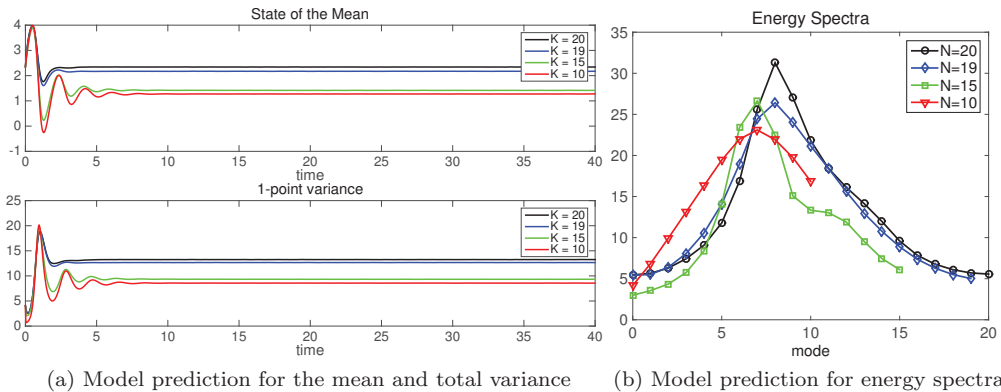


(a) Model prediction for the mean and total variance    (b) Model prediction for energy spectra

Fig. 2.5: *Model prediction of the mean state $\bar{u}_M$ (left upper row) and averaged energy $\mathrm{tr}R_M/J$ (or 1-point statistic on each grid point, left lower row) from Galerkin truncated L-96 models in regime $F = 8$ with number of positive Fourier modes $K = 19, 15, 10$, compared with the true model with total number of modes $K = 20$. The steady state spectra achieved through these truncated models are compared with the state variables $\hat{u}_k$ from the true L-96 system (right).*

REMARK 2.1.    Third order moments in Figure 2.6 show the possible errors we may face in Galerkin truncation models. However rather than the nonlocal third order moments in the spectral domain, the moments in physical domain (shown in the second row) are quite localized with major values near the center. Therefore one possible correction for the truncation method is to get the higher order moments locally near the grid points in the physical domain and then transfer them to the spectral domain to update the system. See [36,37] for some applications about this idea, and we will also investigate this further in future researches.

**2.3.2. Errors from insufficient ensemble size.**    Here we check the other issue about the model performance with changing ensemble size for Monte-Carlo method. It is useful to estimate the error from insufficient number of particles especially when the high dimensionality of the system makes it impossible to run large ensemble members.
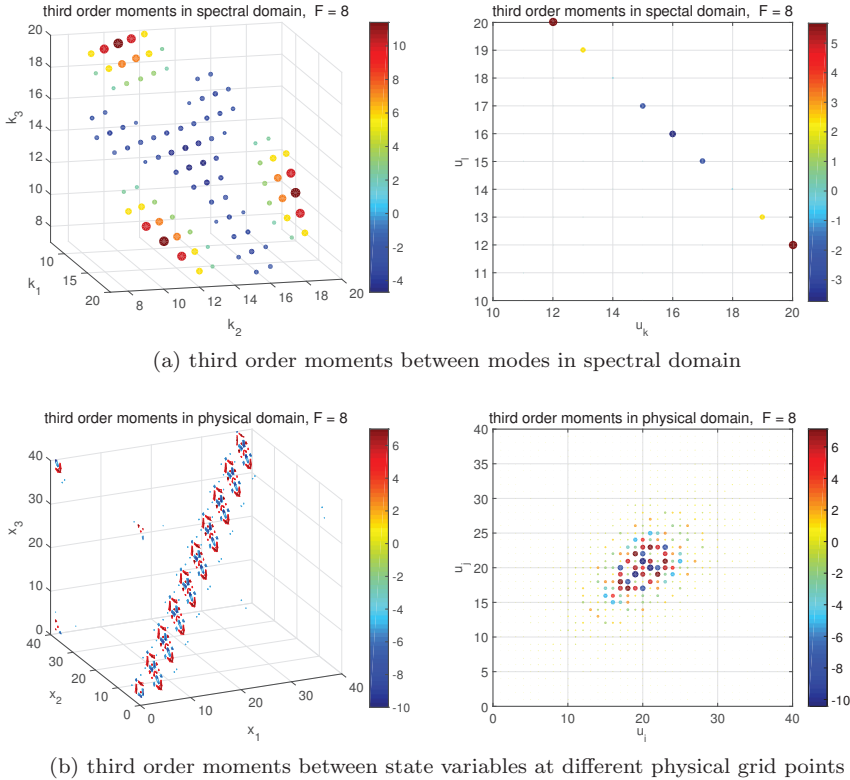
(a) third order moments between modes in spectral domain



(b) third order moments between state variables at different physical grid points

Fig. 2.6: *Third order moments of the flow variable* **u** *in steady state in both spectral and physical domain in regime* $F = 8$. *The left panels show the 3D scatter plot for values of* $\langle \hat{u}'_m \hat{u}'_n \hat{u}'_k \rangle$ *and* $\langle u_i u_j u_l \rangle$; *and the right panels display one cross-section of the third moments of* $\langle \hat{u}'_m \hat{u}'_n \hat{u}'_8 \rangle$ *and* $\langle u_i u_j u_1 \rangle$. *The size and color of the dots represent the values of third order moments and only large values are plotted.*

| MC | 50 | 100 | 500 | $1 \times 10^3$ | $5 \times 10^3$ | $1 \times 10^4$ | $1 \times 10^5$ |
|---|---|---|---|---|---|---|---|
| $\bar{u}_M$ | 2.3952 | 2.3921 | 2.3939 | 2.3966 | 2.3958 | 2.3961 | 2.3958 |
| $\mathrm{std}\,(\bar{u}_M)$ | 0.0453 | 0.0311 | 0.0156 | 0.0120 | 0.0043 | 0.0041 | 0.0010 |

Table 2.1: *Time averaged mean from models with varying ensemble size, and the standard deviation for the time series of the ensemble mean.*
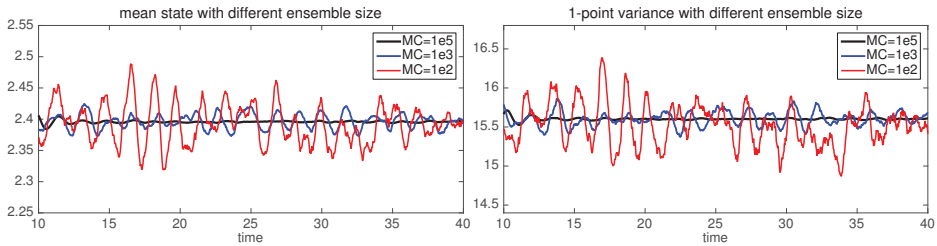


Fig. 2.7: *Model predictions with different ensemble size* $\mathrm{MC} = 1 \times 10^5, 1 \times 10^3, 1 \times 10^2$.

From central limit theorem (CLT), the i.i.d. Monte-Carlo particles $u_i$ with mean $\mu$ and variance $\sigma^2$ tend to have a Gaussian distribution as the particle number $N$ increases $\sum_{i=1}^{N} u_i - N\mu / \sqrt{N\sigma^2} \to \mathbb{N}(0,1)$. Therefore we can estimate the uncertainty of the statistical mean from Monte-Carlo simulation at one time instant as

$$\sqrt{\operatorname{var}(\bar{u})} = \sqrt{\operatorname{var}\left(\frac{1}{N}\sum_{i=1}^{N} u_i\right)} \sim \frac{\sigma}{\sqrt{N}}. \tag{2.18}$$

This shows the extremely slow convergence rate and the high computational requirement as the scale of the problem increases. Table 2.1 and Figure 2.7 below show the model prediction for the mean and total variance with varying ensemble size. We can see that the time averaged mean and variance can always offer good approximation for the truth. On the other hand, the uncertainty in the ensemble prediction at one time instant keeps increasing as the number of particles used decrease. The uncertainty for this ensemble mean at each time step can be estimated through CLT in order $O\left(N^{-1/2}\right)$.

## 3. Gaussian velocity stochastic models for passive tracer statistics

As we have displayed in the previous Section 2.3, the accuracy in the steady state passive tracer statistics is limited by the modeling and computation skill of the complex background advection flow $\mathbf{v}$. Several difficulties cannot be bypassed if we directly go with the true flow system with nonlinearity. First, simple Galerkin truncation of high-frequency wave-numbers in the dynamical equations may introduce large errors to the flow system due to strong nonlinear interactions between the (truncated) small scale and large scale modes. Second, even with a low dimensional Galerkin truncation model, large ensemble size may still be required to resolve the flow if non-Gaussian features and intermittencies are important and of interest. On the other hand, returning to our original problem, the central issue of major interest is the turbulent fluctuation and statistical structure of the passive tracer $T$ rather than the background flow field $\mathbf{v}$. Considering both sides of the problem, the question that is worth asking is whether we can predict the crucial features (such as, intermittency) in steady state tracer statistics advected and forced by nonlinear non-Gaussian background flow $\mathbf{v}$ using simpler imperfect models with error for the background dynamical field.

In this section, we begin with the simplest approximation about the advection flow with imperfect models using linear stochastic dynamics along each spectral mode from the Ornstein–Uhlenbeck process [13, 14, 25, 27], and check whether we can predict the steady state tracer statistics using this simplified Gaussian model for the background flow. The advantages of this choice are obvious: i) The dynamics and statistical structure become much more tractable with explicit solutions that enable us to design the model and tune parameters with ease. Thus the truncation error from the Galerkin method can be converted into consideration in the linear model parameters; ii) the computational difficulty and cost are also greatly reduced considering the simple and controllable structure of this model. Therefore, much smaller ensemble size is needed to get the exact statistical distributions from the linear models, or even we can get the exact analytical solution for the linear system. However, the difficulty that comes with this simplification is the accuracy of the prediction skill for the tracer distributions using only Gaussian velocity models. It is important to guarantee that the intermittency (or the fat tails in the PDFs) in the tracer can still be captured under this simplified methods. To achieve this, one systematic framework using information theory is proposed to train the imperfect models so that the optimal parameters with information consistent statistics can be reached. Following we first introduce the Gaussian velocity stochastic

model and present the exact solution for this model. Then the idea for tuning model parameters under information metric is introduced.

**3.1. Gaussian velocity stochastic models as the background flow.**    We model the advection flow dynamics under each Fourier mode (2.4) and (2.5) through simple Gaussian stochastic models. The state variables **u** is decomposed into mean $\bar{u}$ and fluctuation terms for each mode $\hat{u}_k$ (and $\hat{u}_0$ measures the variance in $\bar{u}$), with $\langle \hat{u}_k \rangle = 0, k = 0, \ldots, J/2$ as in [13, 14, 27]. The correlations and interactions between different modes are totally neglected in this coarse approximation model, thus each mode is assumed to be independent with Gaussian linear dynamics from an Ornstein–Uhlenbeck process. Also we can apply the Galerkin truncation strategy by resolving only the first $K$ leading modes in Fourier space. The truncated linear stochastic models for each mode can be written as

$$\frac{d\hat{u}_k^M}{dt} = (-\gamma_{u_k} + i\omega_{u_k})\hat{u}_k^M + \sigma_{u_k}\dot{W}_k, \quad |k| \le K, \tag{3.1}$$

with $\gamma_{u_k}, \omega_{u_k}, \sigma_{u_k}$ as parameters to be determined, together with the dynamics for the mean

$$\frac{d\bar{u}^M}{dt} = -d\bar{u}^M + \sum_{k \in \Lambda_K} \Gamma_k r_k^M + \sum_{k \in \Lambda_K^c} \Gamma_k r_{k,\infty} + \hat{F}, \tag{3.2}$$

with $\Gamma_k = \cos \frac{4\pi k}{J} - \cos \frac{2\pi k}{J}$. Here $\Lambda_K = \{|k| \le K | K \le J/2\}$ representing the first $K$ resolved Fourier modes with $J/2$ the total number of spectral modes in the system. And $\Lambda_K^c$ is the set of unresolved modes and steady state statistics $r_{j,\infty}$ are used here as a correction (possibly $\Lambda_K^c$ can be an empty set if we apply the full spectrum with every mode resolved). Under this approximation model (3.1) and (3.2), the background flow $\mathbf{v}^M = \left(U^M(t), v^M(x_j, t)\right)$ can be constructed as before for the mean cross-sweep $U^M$ and the shearing flow $v^M$ in the tracer model (1.1)

$$U^M(t) = \bar{u}^M(t) + \hat{u}_0^M(t), \qquad v^M(x_j, t) = \sum_{k \ne 0} \hat{u}_k^M(t) e^{2\pi i x_j}.$$

Now the problem is converted to finding systematic strategies of assigning values to three undetermined coefficients $\gamma_{u_k}, \omega_{u_k}, \sigma_{u_k}$ so that the tracer structure (intermittency) can be reconstructed from this imperfect model. They should be chosen in an unambiguous way according to the true steady state statistics of the system (which is available from observations). In comparison with the original equation for each mode described in (2.5), the linear Gaussian approximation of L-96 system replaces the nonlinear interaction part in the second line of (2.5) by linear damping and rotation together with a white noise

$$\sum_{m \ne 0} \hat{u}_{k+m}\hat{u}_m^* \left(e^{2\pi i \frac{2m+k}{J}} - e^{-2\pi i \frac{m+2k}{J}}\right) \sim (-\gamma_{u_k} + i\omega_{u_k})\hat{u}_k^M + \sigma_{u_k}\dot{W}_k.$$

The white noise $\sigma_{u_k}\dot{W}_k$ is added to each Fourier mode in order to make sure that the system converges to the consistent equilibrium steady state spectra. $\gamma_{u_k}$ represents the damping that neutralizes the additional energy from the white noise. The imaginary component $\omega_{u_k}$ is the additional degree of freedom for tuning the autocorrelation function (or in other words, to control the 'memory' of this mode of its previous history). Note that the quasi-linear part with $U(t)$ in the first line of the formula (2.5) is also

included in the coefficients $\gamma_{u_k}, \omega_{u_k}$. The exact statistical solutions together with the passive tracer under this linear flow are solved and discussed in details in [23, 25], and the intermittencies in the tails of the tracer distributions are investigated theoretically with computational tests in [32]. It is discovered that even under this linear flow field with Gaussian statistics, intermittency with fat-tailed distributions can be generated in the steady state tracer distributions [25, 32]. Still here in our situation with a nonlinear true advection field with strong non-Gaussianity like in the L-96 system (2.1), the challenge is whether we can still capture the correct structure in the tracer spectra and density functions, especially for the intermittency, under these imperfect linear models. Therefore, judicious choice of the model parameters needs to be investigated.

**3.1.1. Mean stochastic model for the background advection flow.** One of the simplest and most direction way to estimate the undetermined coefficients $\gamma_{u_k}, \omega_{u_k}, \sigma_{u_k}$ is through the mean stochastic model (MSM) introduced in [26] and Chapter 12 of [27]. Noticing that the random flow field is a Gaussian process, each random coefficient $\hat{u}_k$ can be characterized by the energy and autocorrelation function defined as follows:

$$E_k \equiv \mathrm{var}\left(\hat{u}_k\left(t\right)\right) = \left\langle \left| \hat{u}_k\left(t\right) - \langle \hat{u}_k \rangle \right|^2 \right\rangle, \tag{3.3}$$

$$\mathcal{R}_k\left(t\right) \equiv \frac{\left\langle \left( \hat{u}_k\left(\tau\right) - \langle \hat{u}_k \rangle \right) \left( \hat{u}_k\left(\tau+t\right) - \langle \hat{u}_k \rangle \right)^* \right\rangle}{\mathrm{var}\left(\hat{u}_k\left(\tau\right)\right)}. \tag{3.4}$$

Here $\langle \cdot \rangle$ can be viewed as ensemble averages. If we integrate the autocorrelation function (3.4) in time, the resulting quantity is called the *decorrelation time*,

$$T_k + i\theta_k = \int_0^\infty \mathcal{R}\left(t\right) dt, \tag{3.5}$$

which quantifies the memory of the system. For clarification, we decompose the decorrelation time (3.5) into the real part and imaginary part. The mean stochastic model determines the model parameters $\{\gamma_{u_k}, \omega_{u_k}, \sigma_{u_k}\}$ using the characterizing energy $E_k$ and decorrelation time $T_k + i\theta_k$. The true values can be achieved through observation data, while the imperfect model predictions of $E_k, T_k, \theta_k$ for the linear system (3.1) can be calculated exactly as (this is through the standard process of solving the stochastic system, detailed deviations can be found in [27] or [10])

$$E_k = \frac{\sigma_{u_k}^2}{2\gamma_{u_k}}, \qquad T_k + i\theta_k = \frac{\gamma_{u_k} + i\omega_{u_k}}{\gamma_{u_k}^2 + \omega_{u_k}^2}.$$

Solving the corresponding values in this equation, we get the values for MSM parameters

$$\gamma_{u_k}^{\mathrm{MSM}} = \frac{T_k}{T_k^2 + \theta_k^2}, \quad \omega_{u_k}^{\mathrm{MSM}} = \frac{\theta_k}{T_k^2 + \theta_k^2}, \quad \sigma_{u_k}^{2,\mathrm{MSM}} = \frac{2T_k E_k}{T_k^2 + \theta_k^2}. \tag{3.6}$$

Despite the simplicity in this mean stochastic model, we can obtain reasonably skillful prediction in uncertainty quantification as well as filtering under this strategy for turbulent systems [26, 27]. However MSM still suffers several shortcomings when strong nonlinearity takes place in the system. Most importantly, the decorrelation time $T_k + i\theta_k$ involves only the time-integrated effects in each mode. This works well when the system is strongly mixing within a nearly Gaussian regime. Whereas when non-Gaussian features become crucial in the system, the pointwise decaying process of the

entire autocorrelation function $\mathcal{R}(t)$ becomes important and we need take into account the temporal performance of the autocorrelation in the linear model approximation (for example, when the autocorrelation function becomes strongly oscillatory, see the examples in Section 4). We will illustrate this with the L-96 system in the next section.

**3.1.2. Explicit statistical solutions under Gaussian linear approximations.** One advantage of the linear stochastic approximation (3.1) for the flow system is that it makes it possible to calculate the exact statistical solutions for each mode in the system. Hopefully, this can offer us hints about the true structure of the tracer statistics. Following we derive the exact formulas for the tracer statistics advected by the linear model in two steps with increasingly stronger assumption about each mode. These formulas can be viewed as special cases among the results derived in [25] under more generalized framework. First, the pathwise solution for the tracer is achieved in (2.10)

$$\hat{T}_k = -\alpha \int_{t_0}^t \exp\left(-\gamma_{T_k}(t-s) - ikJ(s,t)\right)\hat{u}_k(s)\,ds,$$

with $J(s,t) = \int_s^t U(\tau)\,d\tau$. So the mean of each mode decays to zero in steady state $\langle \hat{T}_k \rangle \to 0$, as $t_0 \to -\infty$. For the second order moments, the lagged covariance of the flow mode $\hat{u}_k$ enters so that in steady state $t_0 \to -\infty$, the equation can be first simplified as

$$\left\langle \left|\hat{T}_k(t)\right|^2 \right\rangle = \alpha^2 \int_{-\infty}^t ds \int_{-\infty}^t ds'\, e^{-\gamma_{T_k}(t-s)-\gamma_{T_k}(t-s')} \left\langle e^{-ik\int_s^{s'} U(\tau)d\tau}\hat{u}_k(s)\hat{u}_k^*(s') \right\rangle$$

$$= \alpha^2 \int_{-\infty}^t ds \int_{-\infty}^{t-s} dr\, e^{-2\gamma_{T_k}t+\gamma_{T_k}(r+2s)} \left\langle e^{-ik\int_s^{s+r} U(\tau)d\tau}\hat{u}_k(s)\hat{u}_k^*(s+r) \right\rangle.$$

The second line of the above equation is a simple change of variable $r = s' - s$. If we consider the stationary joint distribution of $\hat{u}_0, \hat{u}_k$ in steady state which is invariant under time shifting, the expectation above inside $\langle \cdot \rangle$ is always valid in domain $[0,r]$, that is,

$$\left\langle e^{-ik\int_s^{s+r} \hat{u}_0(\tau)d\tau}\hat{u}_k(s)\hat{u}_k^*(s+r) \right\rangle = \left\langle e^{-ik\int_0^r \hat{u}_0(\tau)d\tau}\hat{u}_k(0)\hat{u}_k^*(r) \right\rangle.$$

Further under the linear stochastic model, $\hat{u}_0$ and $\hat{u}_k$ are decoupled and the explicit form for the autocorrelation $\langle \hat{u}_k^M(0)\hat{u}_k^{M*}(r) \rangle \sim \exp\left(-(\gamma_{u_k} - i\omega_{u_k})r\right)$ can also be calculated. Therefore, we can get the following two estimations for the energy spectra of $\hat{T}_k$ under increasingly stronger assumptions.

PROPOSITION 3.1. *Estimations for tracer energy spectra:*
- Assume that $\hat{u}_0$ is independent with other modes $\hat{u}_k$, and the zero mode $\hat{u}_0$ has pure Gaussian distribution. *This can be viewed as a first order expansion of the exact formula by expanding the expectation in $\langle \cdot \rangle$. Therefore, we get*

$$\left\langle \left|\hat{T}_k\right|^2 \right\rangle = \frac{\alpha^2}{\gamma_{T_k}} \mathfrak{Re} \int_0^\infty e^{-\gamma_{T_k}r} e^{-ik\bar{u}r} \left\langle e^{-ik\int_0^r \hat{u}_0(\tau)d\tau} \right\rangle \langle \hat{u}_k(0)\hat{u}_k^*(r) \rangle\, dr + \text{h.o.t.}$$

$$= \frac{\alpha^2}{\gamma_{T_k}} \mathfrak{Re} \int_0^\infty e^{-\gamma_{T_k}r} e^{-ik\bar{u}r} e^{-\frac{1}{2}k^2 \text{var} J_0(r)} \langle \hat{u}_k(0)\hat{u}_k^*(r) \rangle\, dr. \qquad (3.7)$$

*Here we further assume the mean is time independent in stationary steady state*
$\langle u \rangle (t) = \bar{u}$ *for simplicity. And denote* $J_0(r) = -\int_0^r \hat{u}_0(\tau) d\tau$, *which is still Gaussian as a finite integration of a Gaussian process* $\hat{u}_0$ *by assumption. Then with a simple application of the property of the characteristic function, we have* $\langle e^{-ikJ_0} \rangle = e^{-\frac{1}{2}k^2 \mathrm{var} J_0}$, *and it is easy to calculate*

$$\mathrm{var} J_0(r) = 2\mathrm{var}\hat{u}_0 \left[ \frac{r}{\gamma_{u_0}} + \frac{1}{\gamma_{u_0}^2} \left( e^{-\gamma_{u_0} r} - 1 \right) \right].$$

*Here* $\gamma_{u_0}$ *is the equivalent damping rate for zero mode* $\hat{u}_0$ *if a Gaussian process is assigned to this variable.*

- Assume further that each mode $\hat{u}_k$ is Gaussian following the linear stochastic model (3.1). *Under this assumption, the explicit form of the autocorrelation in (3.7) can also be calculated. The steady state energy spectrum can be further simplified as*

$$\left\langle \left| \hat{T}_k \right|^2 \right\rangle = \frac{\alpha^2}{\gamma_{T_k}} \mathfrak{Re} \int_0^\infty e^{-(\gamma_{T_k} + \gamma_{u_k} - i\omega_{u_k} - ik\bar{u})r} e^{-\frac{1}{2}k^2 \mathrm{var} J_0(r)} dr. \tag{3.8}$$

The two above approximation formulas (3.7) and (3.8) for steady state tracer spectra are useful for understanding the contribution of each part in the final energy spectra by comparing with the truth, then estimating the intrinsic barriers from adopting Gaussian velocity models. By comparing (3.7) with the true spectra, the difference illustrates the error from ignoring the cross-correlation and higher order moments between modes $\hat{u}_0$ and $\hat{u}_k$ if imperfect models with independent modes are applied. By comparing (3.7) and (3.8) (note that in (3.7), we can still apply the true autocorrelation function from the true system in the formula, whereas (3.8) uses the autocorrelation predicted from a Gaussian process), the errors by using a Gaussian process to approximate the original autocorrelation function can be displayed. It shows the potential improvement through fitting the autocorrelation function as we will discuss in the next part.

**3.2. Spectral Information criterion for improving imperfect model prediction skill.** Results in both (2.15) and (3.7) imply that the accurate estimate with the autocorrelation function $\langle \hat{u}_k(0) \hat{u}_k^*(r) \rangle$ plays a crucial role in model predictions of both steady state PDFs and the energy spectrum of the tracer. The true structure of this autocorrelation is determined by the nonlinear interactions between modes in the true system (2.5), but still we can try to access the optimal fitting of the autocorrelation structure through our linear stochastic model. The MSM with integrated form of the decorrelation time is sometimes insufficient for pointwise agreement with the true autocorrelation (see examples under the L-96 system in Section 4.1). This leads to the requirement for tuning the imperfect model parameters under more detailed calibration about the autocorrelation function. As an unbiased measure between the two probability measure, information theory [21, 22] can offer a balanced estimation for the distance between the truth and imperfect model prediction. Here applying information theory for the autocorrelation functions, we try to develop one systematic strategy that can optimize the linear stochastic model with information consistent prediction skill in the autocorrelation function.

Consider $u(t)$ as a mean zero ($\langle u \rangle = 0$) stationary process in the statistical steady state and focus on its second order statistics. For simplicity, in this subsection we neglect the hat '$\hat{\ }$' and subscript $k$ in the modes $\hat{u}_k$. Consider the goal to construct a linear stochastic model (3.1) with state variable $u^M(t)$ to approximate the truth $u(t)$ with

optimal agreement in the first two order of moments. By comparing the information distance between the two random processes $u$ and $u^M$, the optimal parameter $\sigma$ can be found with consistent equilibrium energy $E$ in (3.3) for the model. $\sigma$ then can be chosen by fitting the climatology spectrum for the variances $E$ as $\sigma = \sqrt{2\gamma E}$. Next including the entire second order statistics of the stationary processes, the random field $u$ is further characterized by the autocorrelation function, measuring how fast the system can forget its previous information. Assuming the process is ergodic, the true autocorrelation function (3.4) with exact dynamics can be calculated as a time average

$$\mathcal{R}(t) = \lim_{T \to \infty} \frac{1}{T} \int_0^T \frac{u(t+\tau)u^*(\tau)}{\left\langle |u|^2 \right\rangle} d\tau. \tag{3.9}$$

The autocorrelation function of corresponding linear model (3.1) can be calculated in explicit form making use of the simple dynamical structure

$$\mathcal{R}^M(t) = e^{(-\gamma_u + i\omega_u)t}. \tag{3.10}$$

Therefore the task now is about how to find the information consistent model with parameters $\{\gamma, \omega\}$ ($\gamma > 0$) which offer the optimal approximation of the true autocorrelation function $\mathcal{R}(t)$ in stationary state.

### 3.2.1. Formal motivation of the spectral entropy information criterion.
Now consider the information metric between the two distribution functions of the random processes $u(t)$ and $u^M(t)$ in stationary states. The *information distance* or *relative entropy* [14,19,24] offers a natural way of measuring the lack of information in one probability density from the imperfect model, $\pi^M$, compared with the true probability density, $\pi$, given by

$$\mathcal{P}\left(\pi, \pi^M\right) = \int \pi \log \frac{\pi}{\pi^M}. \tag{3.11}$$

Despite the lack of symmetry in its arguments, the relative entropy, $\mathcal{P}\left(\pi, \pi^M\right)$ provides an attractive framework for assessing model error like a metric. Importantly, the following two crucial features are satisfied: (i) $\mathcal{P}\left(\pi, \pi^M\right) \geq 0$, and the equality holds if and only if $\pi = \pi^M$; and (ii) it is invariant under any invertible change of variables. The difficulty here is that the information distance in (3.11) only measures the difference between the distributions of the two stationary processes $u(t)$ and $u^M(t)$ *at fixed time instant*, whereas the lagged-in-time covariance $\left\langle u(0)u(t)^* \right\rangle$ (or the memory in the state variables) cannot be included in this metric. So only tuning $u^M$ under the information metric compared with $u$ in the stationary phase offers no improvement for the autocorrelation function. On the other hand, from the definition, the autocorrelation function $\mathcal{R}(t)$ may oscillate with negative values, thus it becomes improper to directly substitute $\mathcal{R}(t)$ into the formula (3.11) by replacing $\pi$ to measure the distance.

In order to generalize the desirable information-theoretic framework to include autocorrelation functions, we expand the random processes $u$ and $u^M$ using *the theory of spectral representation of stationary random fields* discussed in [41] (we put a more detailed description of the spectral theory about stationary random field in Appendix A). It is proved by Khinchin's formula [41], that if the autocorrelation function $\mathcal{R}(t)$ is smooth and rapid-decay (which is the typical property of the autocorrelation function for most systems), we can find a non-negative function $E(\lambda) \geq 0$ so that

$$\mathcal{R}(t) = \int_{-\infty}^{\infty} e^{i\lambda t} dF(\lambda), \tag{3.12}$$

with $dF(\lambda) = E(\lambda)d\lambda$ a non-decreasing function. Therefore we can construct the spectral representation of the stationary process of $u(t)$ as

$$u(t) = \int_{-\infty}^{\infty} e^{i\lambda t} \hat{Z}(d\lambda). \tag{3.13}$$

The exact spectral random measure $\hat{Z}(d\lambda)$ has independent increments whose energy spectrum can be represented by $E(\lambda)$ or $dF(\lambda)$

$$dF(\lambda) = E(\lambda)d\lambda = \mathbb{E}\left|\hat{Z}(d\lambda)\right|^2. \tag{3.14}$$

Applying the theory for spectra representation of stationary processes, we find the one-to-one correspondence between the autocorrelation function $\mathcal{R}(t)$ and non-negative energy spectra $E(\lambda)$, together with the spectral representation $\hat{Z}(d\lambda)$ of the process $u(t)$. Consider the approximation of this random process with only second order statistics by a lattice random field with spacing $\Delta\lambda$. By independence, the true increment $\hat{Z}(\Delta\lambda_j) = \hat{Z}(\lambda_j + \Delta\lambda) - \hat{Z}(\lambda_j)$ has the second order Gaussian probability density function approximation

$$\hat{Z}(\Delta\lambda) \sim p_G(x;\lambda)\Delta\lambda = \mathcal{N}(0, E(\lambda)\Delta\lambda); \tag{3.15}$$

and the corresponding spectral representation from the Gaussian linear model (3.1) also has the (exact) density function

$$\hat{Z}^M(\Delta\lambda) \sim p_G^M(x;\lambda)\Delta\lambda = \mathcal{N}\left(0, E^M(\lambda)\Delta\lambda\right), \tag{3.16}$$

where $\mathcal{N}(m,\sigma^2)$ denotes a Gaussian random variable with mean $m$ and variance $\sigma^2$. Since the spectral measure has independent increment, we approximate the true and linear model Gaussian random fields by

$$p_G = \prod_j \mathcal{N}(0, E(\lambda_j)\Delta\lambda), \quad p_G^M = \prod_j \mathcal{N}\left(0, E^M(\lambda_j)\Delta\lambda\right).$$

Then the normalized relative entropy between these two Gaussian fields becomes

$$\mathcal{P}\left(p_G, p_G^M\right) = \sum_j \mathcal{P}\left(p_G(x;\lambda_j), p_G^M(x;\lambda_j)\right)\Delta\lambda$$
$$\rightarrow \int_{-\infty}^{\infty} \mathcal{P}\left(p_G(x;\lambda), p_G^M(x;\lambda)\right)d\lambda, \text{ as } \Delta\lambda \rightarrow 0. \tag{3.17}$$

Therefore, given spectral densities, $E(\lambda)$ and $E^M(\lambda)$, the spectral relative entropy is given by

$$\mathcal{P}\left(p_G, p_G^M\right) = \mathcal{P}\left(E(\lambda), E^M(\lambda)\right) \triangleq \int_{-\infty}^{\infty} \mathcal{P}\left(p_G(x;\lambda), p_G^M(x;\lambda)\right)d\lambda. \tag{3.18}$$

We abuse the notation above using the spectra $E(\lambda)$ to denote density functions. Since $E$ and $E^M$ are variances for the spectral random variables, it is well-defined of the last part of the above formula (3.18) using the information distance formula (3.11). Through measuring the information distance in the spectral coefficients $\hat{Z}(\lambda)$, we get the lack of information in the autocorrelation function $\mathcal{R}(t)$ from the model.

Recalling the dispersion and signal decomposition of two Gaussian measures [14,21] yields

$$\mathcal{P}\left(p_G\left(\lambda\right),p_G^M\left(\lambda\right)\right)=\mathcal{D}\left(\frac{E\left(\lambda\right)}{E^M\left(\lambda\right)}\right),$$

with $\mathcal{D}\left(x\right)\equiv-\log x+x-I$. Thus we have

$$\mathcal{P}\left(E\left(\lambda\right),E^M\left(\lambda\right)\right)=\int\mathcal{D}\left(\frac{E\left(\lambda\right)}{E^M\left(\lambda\right)}\right)d\lambda.$$

Let the set of parameters $\boldsymbol{\theta}=\{\gamma,\omega\},\gamma>0$ for the linear stochastic model (3.1). Minimum relative entropy criterion implies the process of minimizing the lack of spectral information distance by picking the optimal parameter set $\boldsymbol{\theta}_*$ for the imperfect Gaussian model so that

$$\mathcal{P}\left(E\left(\lambda\right),E^M\left(\lambda,\boldsymbol{\theta}_*\right)\right)=\min_{\boldsymbol{\theta}}\mathcal{P}\left(E\left(\lambda\right),E^M\left(\lambda,\boldsymbol{\theta}\right)\right). \tag{3.19}$$

To summarize, to include the autocorrelation function of a stationary random process into the original information-theoretic framework for tuning model parameters, we transfer the autocorrelation function into the spectral domain with spectral random field $\hat{Z}(\lambda)$. Then the information consistent solution with optimal agreement in the autocorrelation function can be achieved by minimizing the information distance in the spectral density in (3.19). Under this systematic procedure, we can then find the model parameters $\boldsymbol{\theta}=\{\gamma,\omega\}$ $(\gamma>0)$ in the linear model by minimizing the spectral information distance $\mathcal{P}\left(E\left(\lambda\right),E^M\left(\lambda,\boldsymbol{\theta}\right)\right)$ in a training as shown next that guarantees high quality approximation of the autocorrelation functions.

**3.2.2. Relation between spectral information criterion and direct estimations for the autocorrelation functions.**    With the general procedure to tune the model autocorrelation, one last step is to justify the validity of the tuning process using (3.19). Specifically, we need to show that the error in the autocorrelation between the true model results and the linear stochastic model prediction $\left\|\mathcal{R}\left(t\right)-\mathcal{R}^M\left(t\right)\right\|$ can be minimized if the information distance between the spectral densities $\mathcal{P}\left(E\left(\lambda\right),E^M\left(\lambda\right)\right)$ are optimized. The following proposition offers the validation for the optimization process (for the proof, see Appendix A).

PROPOSITION 3.2 (Fitting autocorrelation function using spectral representation of random field).    *The optimal distribution of a stationary random process $u^M$ which gives the least biased approximation of the autocorrelation function $\mathcal{R}(t)$ from the true state variable $u$ can be achieved through minimizing the information distance between the energy spectra $E\left(\lambda\right)$ and $E^M\left(\lambda\right)$ of the two processes according to (3.18). The error in autocorrelation functions $\left\|\mathcal{R}\left(t\right)-\mathcal{R}^M\left(t\right)\right\|$ of two stationary random processes $u\left(t\right)$ and $u^M\left(t\right)$ is bounded by the information distance of their energy spectra $E$ and $E^M$, that is,*
*A)*

$$\max_{-\infty<t<\infty}\left|\mathcal{R}\left(t\right)-\mathcal{R}^M\left(t\right)\right|\leq2\sqrt{3}\left(\int\left(E^2\left(\lambda\right)+E_M^2\left(\lambda\right)\right)d\lambda\right)^{1/2}\mathcal{P}\left(E\left(\lambda\right),E^M\left(\lambda\right)\right)^{1/2}; \tag{3.20}$$

*B)*

$$\int\left|\mathcal{R}\left(t\right)-\mathcal{R}^M\left(t\right)\right|^2dt\leq12\max_{-\infty<t<\infty}\left|E^2\left(\lambda\right)+E_M^2\left(\lambda\right)\right|\mathcal{P}\left(E\left(\lambda\right),E^M\left(\lambda\right)\right). \tag{3.21}$$

REMARK 3.1.

(1) We only consider simple scalar process $u(t)$ in the above method for fitting autocorrelation function. In fact, the above procedure is quite easy to get generalized to a vector random field $\mathbf{u}(t)$. In that case, we need first decompose the vector system into each independent eigendirection of the stationary covariance matrix of $\mathbf{u}$ (especially in the special case for homogeneous system, (spatial) Fourier basis will become the eigenbasis to diagonalize the system with ease). Then we can carry out the above spectral analysis process individually for each mode along the eigendirection.

(2) The mean stochastic model (MSM) by fitting the decorrelation time can be viewed as a special case of the procedure above. If we further assume that we propose a Gaussian model in the white noise limit, that is, the state variable forgets its previous information immediately, then $\mathcal{R}^M$ and $E^M$ become

$$\begin{cases} \mathcal{R}^M(t) \neq 0, & t = 0, \\ \mathcal{R}^M(t) = 0, & t \neq 0, \end{cases} \quad \text{and} \quad E^M(\lambda) = \text{const.} = \int \mathcal{R}^M(t)\, dt.$$

By minimizing the information distance in (3.18), the integrated decorrelation time is fitted as in MSM.

(3) In the above process, we construct a Gaussian random process to fit the true autocorrelation function. Furthermore, we may even use this random measure to construct non-Gaussian processes when non-Gaussianity becomes an important issue. Without any change, the formula (3.18) is still valid for the spectral measures from any other approximation models. Then we may even propose some more sophisticated nonlinear models to approximate the autocorrelation function with higher accuracy. This implies the potential of this framework to much more complicated systems.

## 4. Numerical experiments for capturing intermittency with Gaussian stochastic models

Finally in this section, we check the theories and strategies developed in the previous sections through numerical simulations to show their skill in predicting the fluctuations and intermittencies in the tracer statistics. The 40-dimensional L-96 system (2.1) is used as the truth to simulate the background advection flow, and the linear stochastic models (3.1) and (3.2) with model error are tested as the imperfect approximation models in the skill in capturing true tracer statistics. The models are checked in the same three dynamical regimes as before with $F = 5$ (weakly chaotic), $F = 8$ (strongly chaotic), and $F = 16$ (fully turbulent). Note that $F = 5$ becomes an extremely difficult regime here considering only Gaussian statistics are included in the imperfect flow system. Through the previous discussions about the effects in tracer damping strength $d_T$ in Section 2.2, we choose relatively small damping rate $d_T = 0.1$ for all these cases so the flow spectrum would not be fully dominant. And the same fixed mean gradient $\alpha = 2$ and diffusivity $\kappa = 0.001$ are taken in the tracer equations. For resolving the statistics in the flow dynamics, as we have discussed, much smaller ensemble size with 1000 particles (in comparison with the much larger ensemble size 10000 in the original true system; and in experience with running the models smaller particle number also works fine for resolving the state variables, here relatively larger number of particles are taken to achieve smoother PDFs in presentation) is valid to get the distributions in each Fourier mode $\hat{u}_k$. It needs to be emphasized again that there is no positive Lyapunov exponents in the imperfect linear stochastic models while the true L-96 system includes all kinds of internal instabilities, making it a quite challenging task in capturing the tracer intermittency with accuracy.
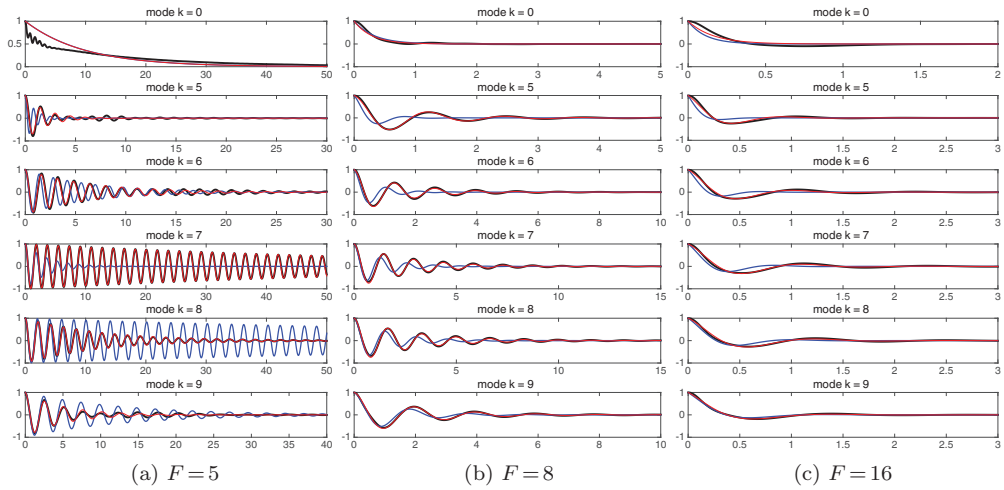
Fig. 4.1: *Fitting of the autocorrelation functions for representative Fourier modes $k = 0$ and $5 \leq |k| \leq 9$ of the flow state variables $\hat{u}_k$ in three dynamical regimes $F = 5, 8, 16$ (only real parts are shown). The true autocorrelation function from the true system is plotted in thick black lines, and the results from MSM is in blue lines, while the optimal model results from tuning parameters in spectral density functions are shown in red lines. Note that black and red lines are largely overlapped together.*
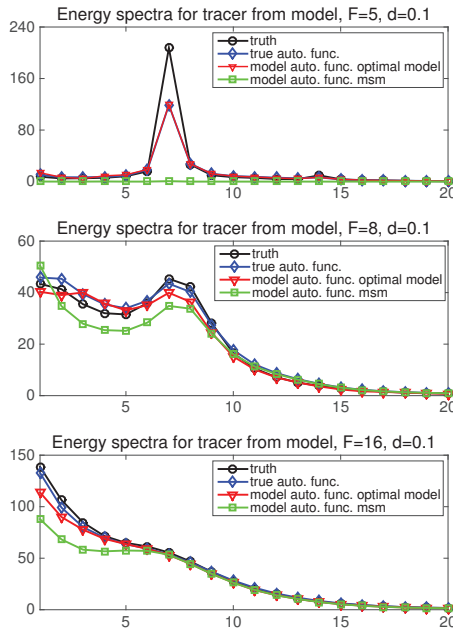


Fig. 4.2: *Steady state tracer energy spectra in regimes $F = 5, 8, 16$. The truth from Monte-Carlo simulations with large ensemble size is shown in black lines with circles. The estimation (3.7) using true autocorrelation function and the estimation (3.8) with optimal linear stochastic model parameters are compared in blue with diamonds and red with triangles. The MSM results are plotted in green with squares.*
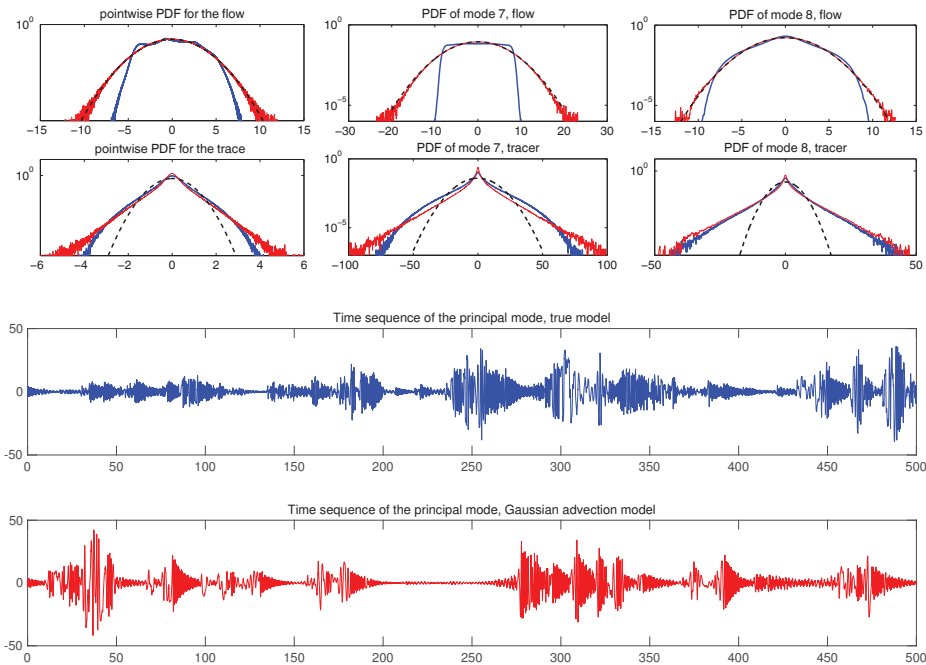
Fig. 4.3: *Comparison of probability density functions captured by the linear stochastic model with optimally tuned parameters (in red) and the truth (in blue) in weakly chaotic regime $F = 5$ with tracer parameters $d_T = 0.1, \alpha = 2, \kappa = 0.001$. The corresponding Gaussian distributions with the same variance are plotted in black dashed lines for comparison. We compare the PDFs of pointwise statistics in the physical domain as well as the two most energetic Fourier modes. The first row is the PDFs in the background flow, and the second row is the corresponding tracer PDFs. The last two rows compare one realization of the time sequence for the tracer principal mode for the true model (blue/dark) and the optimal Gaussian velocity model (red/light).*

**4.1. Fitting autocorrelation functions under information measures.** In the first place, we check the fitting results for the autocorrelation functions from the mean stochastic model (MSM) through (3.6) and from tuning the parameters through the information-theoretic framework (3.18). In Figure 4.1, we compare the autocorrelation functions predicted by MSM and model with optimal parameters, together with the truth with large Monte-Carlo particles. Results for representative wave-numbers $k = 0$ (mode for the horizontal sweep) and $5 \leq |k| \leq 9$ (most energetic modes in the shearing flow) are plotted. Large deficiencies appear for MSM, especially in the weakly mixing strongly non-Gaussian regime $F = 5$ due to the fact that cancellations in the integrated decorrelation time in (3.5Then,make it insufficient to characterize the fast oscillatory modes such as $\hat{u}_7$ and $\hat{u}_8$. As the forcing increases to $F = 8$ and $F = 16$, the system becomes more mixing with Gaussian-like statistics. Errors from MSM become smaller, but still with quite distinctive difference from the truth. Whereas great improvement with uniformly good agreement with the truth in all regimes can be achieved through the information-theoretic framework using optimal parameters for the linear models.

In Figure 4.2, we compare the steady state tracer energy spectra in dynamical regimes $F = 5, 8, 16$. We compare the results from approximation formula (3.7) with true autocorrelation function and from approximation formula (3.8) with optimal linear stochastic model parameters by minimizing information distance, together with the MSM results using parameters in (3.6). Note that (3.7) can be viewed as the upper
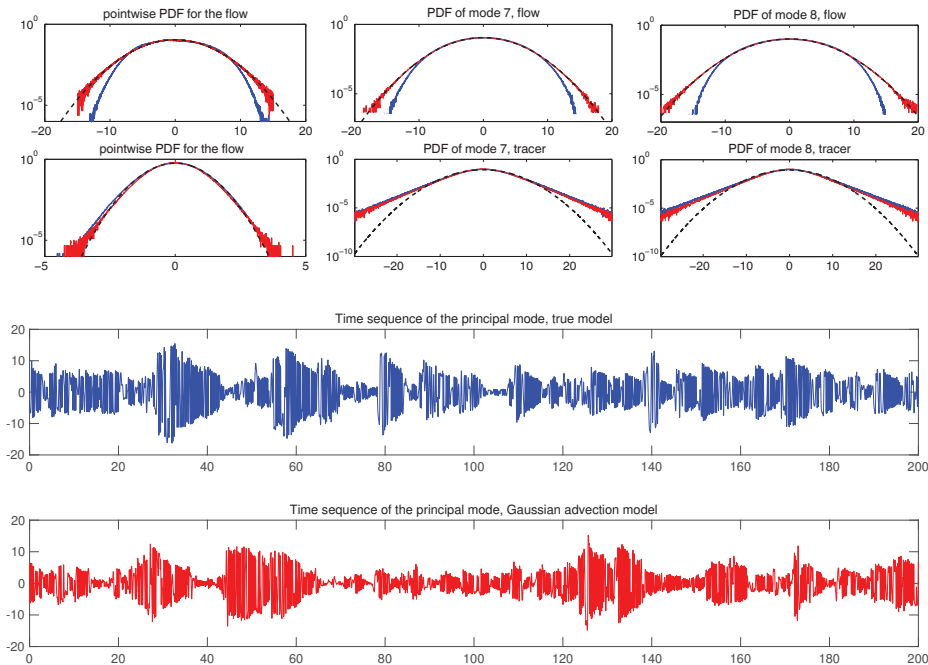
Fig. 4.4: *Comparison of probability density functions captured by the linear stochastic model with optimally tuned parameters (in red) and the truth (in blue) in strongly chaotic regime $F = 8$ with tracer parameters $d_T = 0.1, \alpha = 2, \kappa = 0.001$. The corresponding Gaussian distributions with the same variance are plotted in black dashed lines for comparison. We compare the PDFs of pointwise statistics in the physical domain as well as the two most energetic Fourier modes. The first row is the PDFs in the background flow, and the second row is the corresponding tracer PDFs. The last two rows compare one realization of the time sequence for the tracer principal mode for the true model (blue/dark) and the optimal Gaussian velocity model (red/light).*

bound for using linear models with independent modes, while (3.8) gives similar results with that through numerical simulations of the linear stochastic model using optimal model parameters. In strongly non-Gaussian regime $F = 5$, there exists one dominant mode which contains most of the energy. The major error for the approximation models comes from the non-Gaussianity in the skewed zero mode $\hat{u}_0$ (see PDFs in Figure 2.1). So there exists a barrier in the leading mode by using the linear models whereas the optimal parameter result (in red) works comparably well with the result using the true autocorrelation (in blue). MSM loses its ability in capturing the spectrum fully in this case. As the model becomes more turbulent with $F = 8$ and then $F = 16$, still the linear model with optimal parameter offers desirable estimation with the true energy spectra and has little difference with the results with true autocorrelation function. For all the cases, MSM lacks the accuracy in getting the true model spectral structure.

**4.2. Skill of Gaussian velocity stochastic models in capturing intermittency.** In this final part, we check the Gaussian velocity stochastic models' skill in capturing the fat tails and intermittencies in tracer probability density functions. In figures 4.3, 4.4, and 4.5, we compare the steady state tracer PDFs in both physical and spectral domains for the three representative dynamical regimes $F = 5, 8, 16$. The first rows of the figures display the distributions for the advection flow state variables $u$ in
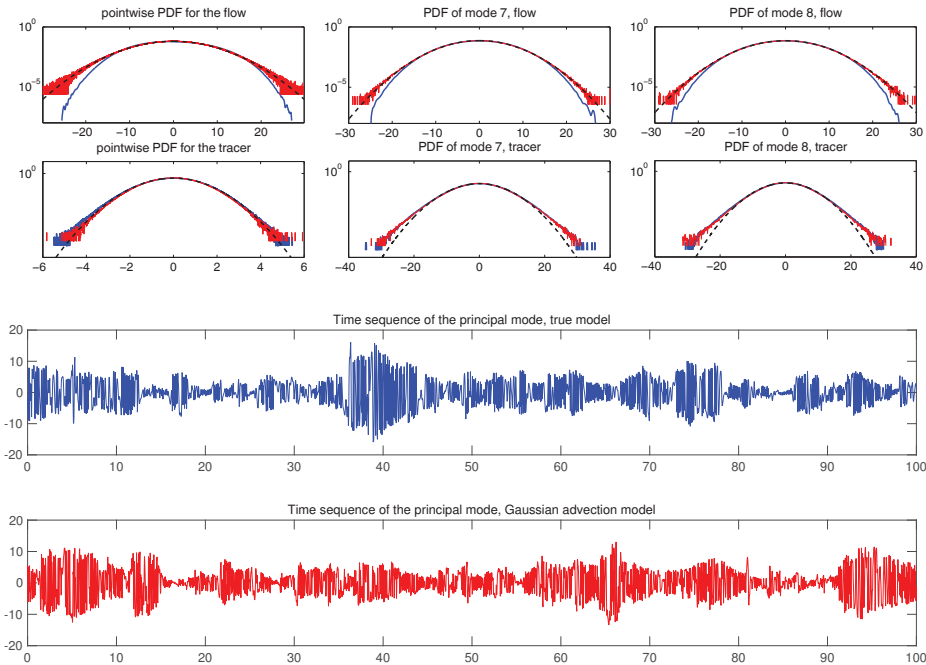
Fig. 4.5: *Comparison of probability density functions captured by the linear stochastic model with optimally tuned parameters (in red) and the truth (in blue) in full turbulent regime $F = 16$ with tracer parameters $d_T = 0.1, \alpha = 2, \kappa = 0.001$. The corresponding Gaussian distributions with the same variance are plotted in black dashed lines for comparison. We compare the PDFs of pointwise statistics in the physical domain as well as the two most energetic Fourier modes. The first row is the PDFs in the background flow, and the second row is the corresponding tracer PDFs. The last two rows compare one realization of the time sequence for the tracer principal mode for the true model (blue/dark) and the optimal Gaussian velocity model (red/light).*

both pointwise PDFs in physical domain and the first two leading Fourier modes in spectral domain. The linear model results for velocity fields in red lines can only offer Gaussian approximations for the state variables in both physical and spectral regimes even though strongly non-Gaussian statistics appear in the true system shown by blue lines. Nevertheless, even under this Gaussian fittings for the flow state variables, the Gaussian velocity stochastic models using optimal parameters through the information-theoretic framework display high skill in capturing intermittency in the tracer statistics. By comparing the imperfect model predictions in the tracer density functions with the truth in the second rows, good agreements especially with accurate approximations in fat tails are achieved in all three dynamical regimes with distinct statistics. Particularly, note that the strongly non-Gaussian case $F = 5$ contains skewed PDFs in the modes of state variables, making it an extremely difficult test case for the Gaussian flow approximation, but the optimal linear models keep the good skill in capturing the key structure in the tracer distributions in both single-point statistics in physical domain and the leading Fourier modes. In addition, we plot one realization of the tracer trajectories in principal mode under both true model and the Gaussian velocity model for these regimes. Strong intermittencies appear in all three dynamical regimes, and it can be observed that the imperfect model under Gaussian advection flow keeps the high skill in capturing important patterns and intermittency in the time sequences. Considering

the simplicity and controllability of the linear stochastic models, this displays the strong skill and obvious advantage in applying tractable and information consistent models to capture nontrivial important statistical features in the passive tracer fields.

## 5. Concluding discussion

In this paper, we focused on the the intermittency in turbulent diffusion described by fat tails in the probability density functions or large spikes randomly occurring in time sequence of passive tracer systems. To investigate the effects of a nonlinear advection field with non-Gaussian features for tracer distributions, the 40-dimensional L-96 system is taken as the test model to simulate the background flow that passively drives the tracer. Among a wide variety of dynamical regimes ranging from strongly non-Gaussian to near Gaussian statistics with strong mixing, intermittency with fat tails in the tracer PDFs can all be generated under this simple setting-up. Another important issue of great interest is the skill of imperfect models in capturing the crucial structures in tracer statistics. This strategy of adopting simple advection flow models has practical significance especially in situations when realistic high dimensional applications are required. We test the simplest linear stochastic models with pure Gaussian statistics as one imperfect approximation to the L-96 system with quite complex dynamical features. Under a systematic manipulation about the model parameters by fitting autocorrelation functions using information theory in a training phase, the achieved optimal Gaussian velocity stochastic model displays uniformly high skill in capturing both the steady state energy spectra and the density distribution functions especially with accuracy in the tails. Furthermore, it needs to be emphasized that the linear stochastic models applied here are only served as a simple illustrative example, and the information framework for tuning autocorrelation functions can be applied to more generalized models.

In our test cases in this paper, we always apply uniform forcing terms in time in the flow fields and focus on the tracer statistics in stationary steady state. On the other hand, model sensitivity to external perturbations is another important issue for quantifying the model prediction skill [14, 30]. In the next stage, it is interesting to investigate the tracer field responses to various external forcing perturbations; and check whether the information-theoretic framework for imperfect models still maintains its skill in predicting fat tails in response to perturbations. Besides all these simpler versions of test models, it is also important to move on to more realistic dynamical systems, and check the skill of this systematic framework in capturing passive tracer statistics in practical applications.

## Appendix A. Spectral representation of the stationary random fields and the information criterion.

**A.1. Spectral representation of general stationary random fields.** For a stationary process $u(t)$, by the definition in (3.9) its autocorrelation function $\mathcal{R}(t)$ is positive-definite in time, that is,

$$\sum_{i,j=1}^{N} \mathcal{R}(t_i - t_j) a_i a_j^* \geq 0, \quad \forall \{a_1, \ldots, a_N\} \in \mathbb{C}^N.$$

Thus it is proved by Khinchin's formula [41] that the autocorrelation function has a spectral representation

$$\mathcal{R}(t) = \int_{-\infty}^{\infty} e^{i\lambda t} dF(\lambda), \tag{A.1}$$

with $F(\lambda)$ a non-decreasing function. The function $F(\lambda)$ is called the *spectral distribution function* of the stationary process $u(t)$. According to the formula (A.1), the spectral distribution function $F(\lambda)$ of the process $u(t)$ can be determined from its autocorrelation function $\mathcal{R}(t)$ via Fourier transform. Conversely, we can calculate the autocorrelation function easily from the corresponding spectral distribution function. By further imposing that $\mathcal{R}(t)$ is smooth and rapid-decaying (which is valid for most systems), we can find non-negative function $E(\lambda) \geq 0$ so that

$$dF(\lambda) = E(\lambda)\,d\lambda,$$

where $E(\lambda)$ is usually called the *spectral density function* of the stationary process $u(t)$. Therefore we can construct the spectral representation of the stationary random field $u(t)$ as

$$u(t) = \int_{-\infty}^{\infty} e^{i\lambda t} \hat{Z}(d\lambda), \tag{A.2}$$

with $\hat{Z}(d\lambda)$ a random measure which can be defined in a similar form of Fourier transform as

$$\hat{Z}(d\lambda) = \lim_{T \to \infty} \frac{d\lambda}{2T} \int_{-T}^{T} e^{-i\lambda t} u(t)\,dt.$$

The spectral representation (A.2) of the stationary random field decomposes the process into mutually independent spectral increments $\hat{Z}(d\lambda) \equiv \hat{Z}(\lambda + d\lambda) - \hat{Z}(\lambda)$, therefore offers us the flexibility of analyzing the statistics at each individual frequency $\lambda$. By combining (A.1) and (A.2) we can find that $E(\lambda)$ or $dF(\lambda)$ can be viewed as the energy spectrum of the random field $u$ which measures the variance in $\hat{Z}(d\lambda)$

$$dF(\lambda) = E(\lambda)\,d\lambda = \mathbb{E}\left| \hat{Z}(d\lambda) \right|^2. \tag{A.3}$$

Note here that in general $\hat{Z}(d\lambda)$, which depends on the statistics of the stationary process $u(t)$, is not necessarily a Gaussian random variable and may include higher order statistical information. But here we only concern fitting this process with consistent statistics up to second order, so a simple Gaussian approximation $u_G(t)$ with statistics up to second order moments is considered.

**A.2. Spectral information criterion for measuring the autocorrelation functions.** As we have discussed in the main text, we want to approximate the (possibly non-Gaussian) stationary process $u(t)$ using the linear stochastic model (3.1) of state variable $u^M(t)$ with only Gaussian statistics. Since we only concentrate on the second order moment at the present case, we can approximate the true spectral random measure $\hat{Z}(d\lambda)$ with a Gaussian random field $\hat{Z}_G(d\lambda)$ with consistent second order moment, $\mathbb{E}\left|\hat{Z}(d\lambda)\right|^2 = \mathbb{E}\left|\hat{Z}_G(d\lambda)\right|^2$. Correspondingly, the imperfect linear model approximation $u^M(t)$ can be naturally expressed under a Wiener random measure $d\hat{W}(\lambda)$. That is,

$$u_G(t) = \int_{-\infty}^{\infty} e^{i\lambda t} \hat{Z}_G(d\lambda) = \int_{-\infty}^{\infty} e^{i\lambda t} E^{1/2}(\lambda)\,d\hat{W}(\lambda), \tag{A.4}$$

$$u^M(t) = \int_{-\infty}^{\infty} e^{i\lambda t} \hat{Z}_M(d\lambda) = \int_{-\infty}^{\infty} e^{i\lambda t} E_M^{1/2}(\lambda)\,d\hat{W}(\lambda). \tag{A.5}$$

Above $E(\lambda)$ is the true spectral energy, while $E_M(\lambda) = \frac{2\gamma}{\gamma^2 + (\lambda - \omega)^2}$ is the explicit form for the density of autocorrelation function for the linear model from (A.3). Instead of measuring the information distance between the stationary random fields $u(t)$ and $u^M(t)$ at fixed time $t$, a more favorable way is to compare the distributions between the true spectral random measure $\hat{Z}(d\lambda)$ and the imperfect model approximation field $\hat{Z}_M(d\lambda)$ which can include the information about the autocorrelation functions (thus include the 'memory' of the system).

First, only consider a small independent spectral increment $\hat{Z}(\Delta\lambda) = \hat{Z}(\lambda + \Delta\lambda) - \hat{Z}(\lambda)$ between the true system $u(t)$ and the imperfect linear model approximation $u^M(t)$. According to the information relation derived in [24], the information distance $\mathcal{P}(p(x;\lambda), p_M(x;\lambda))$ can be decomposed into two parts

$$\mathcal{P}(p(x;\lambda), p_M(x;\lambda)) = \mathcal{P}(p(x;\lambda), p_G(x;\lambda)) + \mathcal{P}(p_G(x;\lambda), p_M(x;\lambda))$$
$$= (\mathcal{S}(p_G(x;\lambda)) - \mathcal{S}(p(x;\lambda))) + \mathcal{P}(p_G(x;\lambda), p_M(x;\lambda)). \quad \text{(A.6)}$$

In (A.6), $\mathcal{S}(p) = -\int p \log p$ is the absolute entropy [21, 24] of the probability measure $p$, and $p$, $p_G$, and $p_M$ are the probability density functions of the spectral increments $\hat{Z}(\Delta\lambda)$, $\hat{Z}_G(\Delta\lambda)$, and $\hat{Z}_M(\Delta\lambda)$ correspondingly

$$\hat{Z}(\Delta\lambda) \sim p(x;\lambda)\Delta\lambda, \quad \hat{Z}_G(\Delta\lambda) \sim p_G(x;\lambda)\Delta\lambda \quad \hat{Z}_M(\Delta\lambda) \sim p_M(x;\lambda)\Delta\lambda.$$

Therefore, the first part $\mathcal{S}(p_G) - \mathcal{S}(p)$ on the right-hand side of (A.6) measures the information error due to the Gaussian restriction. Here we only focus on the prediction skills using Gaussian linear models, thus this error becomes the inherent information barrier for these Gaussian models unless higher order statistics are taken into account. Then the second part $\mathcal{P}(p_G, p_M)$ in (A.6) expresses the information error between the true second order statistics and the imperfect model results. Ideally this part can be minimized by tuning the imperfect model parameters in a training phase.

Above with the help of the spectral decomposition of stationary random process, we decompose the state variables into the spectral form in (A.4) and (A.5) respectively under each Fourier basis $\{e^{i\lambda t}\}$ in time and measure the information distance between one single spectral increment at frequency $\lambda$. Here consider the full spectra of random measures

$$p_G = \prod_{\lambda} p_G(x;\lambda) \sim \left\{ \hat{Z}_G(d\lambda) \right\}_{\lambda=-\infty}^{\infty} = \left\{ E^{1/2}(\lambda) d\hat{W}(\lambda) \right\}_{\lambda=-\infty}^{\infty},$$

and

$$p_M = \prod_{\lambda} p_M(x;\lambda) \sim \left\{ \hat{Z}_M(d\lambda) \right\}_{\lambda=-\infty}^{\infty} = \left\{ E_M^{1/2}(\lambda) d\hat{W}(\lambda) \right\}_{\lambda=-\infty}^{\infty},$$

as two random process about $\lambda \in \mathbb{R}$. Due to the independent increments between $\hat{Z}(d\lambda_1)$ and $\hat{Z}(d\lambda_2)$, the information distance between these two random processes can be expressed as the integration between all the spectral modes

$$\mathcal{P}(p_G, p_M) = \mathcal{P}\left( \prod_{\lambda} p_G(x;\lambda), \prod_{\lambda} p_M(x;\lambda) \right) = \int_{-\infty}^{\infty} d\lambda \mathcal{P}(p_G(x;\lambda), p_M(x;\lambda)). \quad \text{(A.7)}$$

Actually considering two random measure $\hat{Z}_1 = \hat{Z}(\Delta\lambda_1)$ and $\hat{Z}_2 = \hat{Z}(\Delta\lambda_2)$ with $\Delta\lambda_1$ and $\Delta\lambda_2$ having no intersection, the chain rule for relative entropy [7] implies

$$\mathcal{P}(p(\hat{z}_1, \hat{z}_2), p_M(\hat{z}_1, \hat{z}_2)) = \mathcal{P}(p(\hat{z}_1), p_M(\hat{z}_1)) + \mathcal{P}(p(\hat{z}_2 \mid \hat{z}_1), p_M(\hat{z}_2 \mid \hat{z}_1))$$

$$= \mathcal{P}\left(p\left(\hat{z}_1\right), p_M\left(\hat{z}_1\right)\right) + \mathcal{P}\left(p\left(\hat{z}_2\right), p_M\left(\hat{z}_2\right)\right).$$

In the second line of the above equation, the independence between $\hat{Z}_1$ and $\hat{Z}_2$ is from the definition of the random measure. We can summarize the result as the following lemma.

LEMMA A.1.    *The information distance between the random spectra $\left\{\hat{Z}_G\left(d\lambda\right)\right\}$ and $\left\{\hat{Z}_M\left(d\lambda\right)\right\}$ of two stationary random processes $u_G\left(t\right)$ and $u_M\left(t\right)$ equals to the information distance between the two non-negative energy spectra $E\left(\lambda\right)$ and $E_M\left(\lambda\right)$ of the two processes over the spectral domain defined as*

$$\mathcal{P}\left(p_G, p_M\right) = \mathcal{P}\left(E\left(\lambda\right), E_M\left(\lambda\right)\right) = \int d\lambda \mathcal{D}\left(\frac{E\left(\lambda\right)}{E_M\left(\lambda\right)}\right), \tag{A.8}$$

*where $\mathcal{D}\left(x\right) = -\log x + x - I$ is the Gaussian relative entropy with a zero mean state. Thus the stationary random field $u_G\left(t\right)$ can be fitted using a Gaussian random process by minimizing information distance of energy spectrum $E\left(\lambda\right)$.*

**A.3. Proof of Proposition 3.2.**    Finally we prove Proposition 3.2 showing that minimum relative entropy solution from (A.8) gives the best approximation for the autocorrelation function $\mathcal{R}\left(t\right)$. The proof utilizes the Hellinger distance [5,15,35] of two probability densities $p, q$

$$d_H^2\left(p, q\right) = \frac{1}{2} \int \left(\sqrt{p} - \sqrt{q}\right)^2, \tag{A.9}$$

with the relation of Hellinger distance to relative entropy

$$d_H^2\left(p, q\right) \leq \mathcal{P}\left(p, q\right). \tag{A.10}$$

On the other hand, we have the standard formula for the autocorrelations from the Fourier interpretation

$$\mathcal{R}\left(t\right) = \int_{-\infty}^{\infty} e^{i\lambda t} E\left(\lambda\right) d\lambda, \qquad \mathcal{R}^M\left(t\right) = \int_{-\infty}^{\infty} e^{i\lambda t} E^M\left(\lambda\right) d\lambda. \tag{A.11}$$

Below we use a mean zero scalar Gaussian random variable $X \sim p_G\left(x; \lambda\right)$. So the second moment $E$ and fourth moment $\left\langle x^4 \right\rangle$ are defined as

$$E\left(\lambda\right) = \int x^2 p_G\left(x; \lambda\right) dx, \quad \left\langle x^4 \right\rangle\left(\lambda\right) = 3E^2\left(\lambda\right). \tag{A.12}$$

We need the following lemma.

LEMMA A.2.    *For two zero mean Gaussian random variables, $p_G, q_G$, then*

$$\left|\int x^2 p_G - \int x^2 q_G\right| \leq 2\sqrt{3}\left(\left(\int x^2 p_G\right)^2 + \left(\int x^2 q_G\right)^2\right)^{1/2} d_H\left(p_G, q_G\right), \tag{A.13}$$

*with the immediate corollary of (A.13) using the notation in (A.12)*

$$\left|E\left(\lambda\right) - E^M\left(\lambda\right)\right| \leq 2\sqrt{3}\left(E^2\left(\lambda\right) + E_M^2\left(\lambda\right)\right)^{1/2} d_H\left(p_G\left(\lambda\right), p_G^M\left(\lambda\right)\right). \tag{A.14}$$

*Proof.*   By applying Cauchy–Schwarz inequality and (A.12),

$$
\left| \int x^2 p_G - \int x^2 q_G \right| = \left| \int x^2 \left( p_G^{1/2} + q_G^{1/2} \right) \left( p_G^{1/2} - q_G^{1/2} \right) \right|
$$

$$
\leq \sqrt{2} \left( \int x^4 \left( p_G + 2 p_G^{1/2} q_G^{1/2} + q_G \right) \right)^{1/2} d_H \left( p_G, q_G \right)
$$

$$
\leq 2 \left( \int x^4 \left( p_G + q_G \right) \right)^{1/2} d_H \left( p_G, q_G \right)
$$

$$
= 2\sqrt{3} \left( \left( \int x^2 p_G \right)^2 + \left( \int x^2 q_G \right)^2 \right)^{1/2} d_H \left( p_G, q_G \right),
$$

□

Then with this lemma, and its corollary, and the Fourier representation in (A.11) together we have for A) in the proposition

$$
\max_{-\infty < t < \infty} \left| \mathcal{R}(t) - \mathcal{R}^M(t) \right|
$$

$$
\leq \int \left| E(\lambda) - E^M(\lambda) \right| d\lambda
$$

$$
\leq 2\sqrt{3} \int \left( E^2(\lambda) + E_M^2(\lambda) \right)^{1/2} d_H \left( p_G(\lambda), p_G^M(\lambda) \right) d\lambda
$$

$$
\leq 2\sqrt{3} \left( \int \left( E^2(\lambda) + E_M^2(\lambda) \right) d\lambda \right)^{1/2} \left( \int d_H^2 \left( p_G(\lambda), p_G^M(\lambda) \right) d\lambda \right)^{1/2}
$$

$$
\leq 2\sqrt{3} \left( \int \left( E^2(\lambda) + E_M^2(\lambda) \right) d\lambda \right)^{1/2} \mathcal{P} \left( E, E^M \right)^{1/2}.
$$

And for B) in the proposition

$$
\int \left| \mathcal{R}(t) - \mathcal{R}^M(t) \right|^2 dt = \int \left| E(\lambda) - E^M(\lambda) \right|^2 d\lambda
$$

$$
\leq 12 \int \left( E^2(\lambda) + E_M^2(\lambda) \right) d_H^2 \left( p_G(\lambda), p_G^M(\lambda) \right)
$$

$$
\leq 12 \max_{\lambda} \left( E^2(\lambda) + E_M^2(\lambda) \right) \mathcal{P} \left( E, E^M \right).
$$

This finishes the proof of Proposition 3.2.

REFERENCES

[1]  M. Avellaneda and A.J. Majda, *Mathematical models with exact renormalization for turbulent transport*, Commun. Math. Phys., 131, 381–429, 1990.

[2]  M. Avellaneda and A.J. Majda, *Mathematical models with exact renormalization for turbulent transport, ii: Fractal interfaces, non-Gaussian statistics and the sweeping effect*, Commun. Math. Phys., 146, 139–204, 1992.

[3]  T. Bengtsson, P. Bickel, and B. Li, et al., *Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems*, in Probability and statistics: Essays in honor of David A. Freedman, Inst. Math. Stat., 316–334, 2008.

[4] A. Bourlioux and A.J. Majda, *Elementary models with probability distribution function intermittency for passive scalars with a mean gradient*, Physics of Fluids (1994-present), 14, 881–897, 2002.

[5] M. Branicki and A.J. Majda, *Quantifying bayesian filter performance for turbulent dynamical systems through information theory*, Commun. Math. Sci., 12(5), 902–978, 2014.

[6] J.C. Bronski and R.M. McLaughlin, *The problem of moments and the Majda model for scalar intermittency*, Physics Letters A, 265, 257–263, 2000.

[7] I. Csiszar and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Cambridge University Press, 2011.

[8] D.M. Frierson, *Robust increases in midlatitude static stability in simulations of global warming*, Geophysical research letters, 33, 2006.

[9] D.M. Frierson, *Midlatitude static stability in simple and comprehensive general circulation models*, Journal of the Atmospheric Sciences, 65, 1049–1062, 2008.

[10] C. Gardiner, *Stochastic Methods: A handbook for the Natural and Social Sciences Springer Series in Synergetics*, Springer, Berlin, Germany, 2009.

[11] B. Gershgorin, and A. Majda, et al., *A nonlinear test model for filtering slow-fast systems*, Commun. Math. Sci., 6, 611–649, 2008.

[12] B. Gershgorin and A.J. Majda, *A test model for fluctuation–dissipation theorems with time-periodic statistics*, Physica D: Nonlinear Phenomena, 239, 1741–1757, 2010.

[13] B. Gershgorin and A.J. Majda, *Filtering a statistically exactly solvable test model for turbulent tracers from partial observations*, J. Comput. Phys., 230, 1602–1638, 2011.

[14] B. Gershgorin and A.J. Majda, *Quantifying uncertainty for climate change and long-range forecasting scenarios with model errors. Part I: Gaussian models*, Journal of Climate, 25, 4523–4548, 2012.

[15] A.L. Gibbs and F.E. Su, *On choosing and bounding probability metrics*, International Statistical Review, 70, 419–435, 2002.

[16] R.H. Kraichnan, *Small-scale structure of a scalar field convected by turbulence*, Physics of Fluids (1958-1988), 11, 945–953, 1968.

[17] R.H. Kraichnan, *Anomalous scaling of a randomly advected passive scalar*, Physical review letters, 72, 1016–1019, 1994.

[18] P.R. Kramer, A.J. Majda, and E. Vanden-Eijnden, *Closure approximations for passive scalar turbulence: A comparative study on an exactly solvable model with complex features*, J. Stat. Phys., 111, 565–679, 2003.

[19] S. Kullback and R.A. Leibler, *On information and sufficiency*, Annals of Mathematical Statistics, 22(22), 79–86, 1951.

[20] E.N. Lorenz, *Predictability: A Problem Partly Solved*, in Proc. Seminar on Predictability, 1, 1996.

[21] A. Majda, R. Kleeman, and D. Cai, et al., *A mathematical framework for quantifying predictability through relative entropy*, Meth. Appl. Anal., 9, 425–444, 2002.

[22] A.J. Majda, R.V. Abramov, and M.J. Grote, *Information theory and stochastics for multiscale nonlinear systems*, American Math. Soc., 25, 2005.

[23] A.J. Majda and M. Branicki, *Lessons in uncertainty quantification for turbulent dynamical systems*, Discrete Cont. Dyn. Systems, 32(9), 2012.

[24] A.J. Majda and B. Gershgorin, *Quantifying uncertainty in climate change science through empirical information theory*, Proceedings of the National Academy of Sciences, 107, 14958–14963, 2010.

[25] A.J. Majda and B. Gershgorin, *Elementary models for turbulent diffusion with complex physical features: eddy diffusivity, spectrum and intermittency*, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 371, 173–185, 2013.

[26] A.J. Majda, B. Gershgorin, and Y. Yuan, *Low-frequency climate response and fluctuation-dissipation theorems: theory and practice*, Journal of the Atmospheric Sciences, 67, 1186–1201, 2010.

[27] A.J. Majda and J. Harlim, *Filtering Complex Turbulent Systems*, Cambridge University Press, 2012.

[28] A.J. Majda and P.R. Kramer, *Simplified models for turbulent diffusion: theory, numerical modelling, and physical phenomena*, Physics reports, 314, 237–574, 1999.

[29] A.J. Majda, D.W. McLaughlin, and E.G. Tabak, *A one-dimensional model for dispersive wave turbulence*, Journal of Nonlinear Science, 7, 9–44, 1997.

[30] A.J. Majda and D. Qi, *Improving prediction skill of imperfect turbulent models through statistical response and information theory*, J. Nonlinear Science, 1–53, 2015.

[31] 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, 7511–7516, 2014.

[32] A.J. Majda and X.T. Tong, *Intermittency in turbulent diffusion models with a mean gradient*,

Nonlinearity, 28(11), 4171–4208, 2015.

[33] 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.

[34] D. Qi and A.J. Majda, *Blended particle methods with adaptive subspaces for filtering turbulent dynamical systems*, Physica D: Nonlinear Phenomena, 298, 21–41, 2015.

[35] R.D. Reiss, *Approximate Distributions of Order Statistics*, Springer, 1989.

[36] T.P. Sapsis and A.J. Majda, *Blending modified Gaussian closure and non-Gaussian reduced subspace methods for turbulent dynamical systems*, Journal of Nonlinear Science, 23, 1039–1071, 2013.

[37] T.P. Sapsis and A.J. Majda, *A statistically accurate modified quasilinear Gaussian closure for uncertainty quantification in turbulent dynamical systems*, Physica D: Nonlinear Phenomena, 252, 34–45, 2013.

[38] L.M. Smith and F. Waleffe, *Transfer of energy to two-dimensional large scales in forced, rotating three-dimensional turbulence*, Physics of fluids, 11, 1608–1622, 1999.

[39] E. Vanden Eijnden, *Non-Gaussian invariant measures for the Majda model of decaying turbulent transport*, Commun. Pure Appl. Math., 54, 1146–1167, 2001.

[40] Z. Warhaft et al., *Probability distribution, conditional dissipation, and transport of passive temperature fluctuations in grid-generated turbulence*, Physics of Fluids A: Fluid Dynamics (1989-1993), 4, 2292–2307, 1992.

[41] A.M. Yaglom, *An introduction to the theory of stationary random functions*, Courier Corporation, 5(3), 479, 2004.