



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Non-Gaussian Test Models for Prediction and State Estimation with Model Errors

Citation for published version:

Branicki, M, Chen, N & Majda, AJ 2013, 'Non-Gaussian Test Models for Prediction and State Estimation with Model Errors' Chinese Annals of Mathematics. Series B, vol 34, no. 1, pp. 29-64. DOI: 10.1007/s11401-012-0759-3

Digital Object Identifier (DOI):

[10.1007/s11401-012-0759-3](https://doi.org/10.1007/s11401-012-0759-3)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Chinese Annals of Mathematics. Series B

Publisher Rights Statement:

The final publication is available at Springer via <http://dx.doi.org/10.1007/s11401-012-0759-3>.

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Non-Gaussian Test Models for Prediction and State Estimation with Model Errors

M. Branicki*, N. Chen, and A.J. Majda

Department of Mathematics and Center for Atmosphere Ocean Science,
Courant Institute of Mathematical Sciences
New York University, New York, USA

Abstract

Turbulent dynamical systems involve dynamics with both a large dimensional phase space and a large number of positive Lyapunov exponents. Such systems are ubiquitous in applications in contemporary science and engineering where statistical ensemble prediction and real-time filtering/state estimation are needed despite the underlying complexity of the system. Statistically exactly solvable test models have a crucial role to provide firm mathematical underpinning or new algorithms for vastly more complex scientific phenomena. Here a class of statistically exactly solvable non-Gaussian test models are introduced where a generalized Feynman-Kac formulation reduces the exact behavior of conditional statistical moments to the solution of inhomogeneous Fokker-Planck equations modified by linear lower order coupling and source terms. This procedure is applied to a test model with hidden instabilities and combined with information theory to address two important issues in contemporary statistical prediction of turbulent dynamical systems: coarse-grained ensemble prediction in a perfect model and improving long range forecasting in imperfect models. The models discussed here should be useful for many other applications and algorithms for real time prediction and state estimation.

1 Introduction

Turbulent dynamical systems involve dynamics with both a large dimensional phase space and a large number of positive Lyapunov exponents. Such extremely complex systems are ubiquitous in many disciplines of contemporary science and engineering such as climate-atmosphere-ocean science, neural science, material science, and engineering turbulence. Topics of wide contemporary interest involve statistical ensemble prediction [31] and real time state estimation/filtering [34] for the extremely complex systems while coping with the fundamental limitations of model error and the curse of small ensemble size [22].

An important role of mathematics in applied sciences is to develop simpler exactly or easily solvable test models with unambiguous mathematical features which nevertheless capture crucial features of vastly more complex systems in science and engineering. Such models provide firm underpinning for both advancing scientific understanding and developing new numerical or statistical understanding. One of the authors has been developing this approach with various collaborators over the past few years for paradigm problems for turbulent dynamical systems. For example, simple statistically exactly solvable test models have been developed for slow-fast systems [12, 13], turbulent tracers [21, 14, 2, 33] and as stochastic parameterization algorithms for real-time filtering of turbulent dynamical systems with judicious model error [9, 8, 15, 35, 34]. Such models have been utilized as unambiguous test models for improving

*Corresponding author: branicki@cims.nyu.edu

prediction with imperfect models in climate science through empirical information theory [10, 28, 29, 30, 5] and for testing algorithms for uncertainty quantification [5, 4, 25].

Here, we study non-Gaussian statistics in a class of test models which are statistically exactly solvable through a generalized Feynman-Kac formula [16, 21] which reduces the exact behavior of conditional statistical moments to the solution of inhomogeneous Fokker-Planck equations modified by linear lower-order coupling terms and source terms. This exact procedure is developed in section 2 below and involves only marginal averaging and integration by parts. In section 3 elementary test models are introduced where the general procedure from section 2 can be evaluated through elementary numerical solutions of the coupled generalized Fokker-Planck equations (CGFPE). Section 4 contains a brief introduction to the use of information theory to quantify model error in a framework adapted to the present context. Section 5 contains two applications of the material in section 3 and 4 to statistical ensemble forecasting: the first application involves coarse-grained ensemble prediction in a perfect model with hidden instabilities; the second application involves the use of imperfect models for long range forecasting.

2 Test Models with Exactly Solvable Conditional Moments

Here, we consider a special class of test models and illustrate the evolution of the exact conditional statistical moments can be calculated through the solution of coupled generalized Fokker-Planck equations (CGFPE). Our elementary derivation follows the philosophy of generalized Feynman-Kac framework [16, 21] although we do not know any specific reference for the general principle developed below.

Consider a vector $\mathbf{u} \in \mathbb{R}^M$ partitioned into components $\mathbf{u} = (\mathbf{u}_I, \mathbf{u}_{II})$ with $\mathbf{u}_I \in \mathbb{R}^{M_I}$, $\mathbf{u}_{II} \in \mathbb{R}^{M_{II}}$, and $M = M_I + M_{II}$. We focus on the special class of test models given by the system of (Itô) SDE's,

$$\begin{aligned} d\mathbf{u}_I &= F_I(\mathbf{u}_I, t) dt + \sigma_I(\mathbf{u}_I, t) dW_I(t), \\ d\mathbf{u}_{II} &= \left(F_{II}(\mathbf{u}_I, t) + \Gamma(\mathbf{u}_I, t) \mathbf{u}_{II} \right) dt + \sigma_{II}(\mathbf{u}_I, t) dW_I(t) + \sigma_{II,A}(\mathbf{u}_I, t) dW_{II,A}(t) \\ &\quad + \left(\sigma_{II,0} + \sigma_{II,M}(\mathbf{u}_I, t) \mathbf{u}_{II} \right) dW_{II,M}, \end{aligned} \quad (1)$$

where W_I is an M_I -dimensional Wiener process and $W_{II,A}, W_{II,0}, W_{II,M}$ are independent M_{II} -dimensional Wiener processes. Note that the dynamics of \mathbf{u}_I is arbitrary while the dynamics of \mathbf{u}_{II} is quasilinear, i.e., linear in \mathbf{u}_{II} in both the drift and noise but with general nonlinear coefficients depending on \mathbf{u}_I . Also, note that the noise for \mathbf{u}_I and \mathbf{u}_{II} can be correlated through W_I appearing in both the equation for \mathbf{u}_I and \mathbf{u}_{II} . All of the nonlinear test models for slow-fast systems [12, 13], turbulent tracers [21, 14, 2, 33] and exactly solvable stochastic parameterized filters [9, 8, 15, 35, 34] have the structural form as in (1). Such systems are known to have exactly solvable non-Gaussian statistics for filters where \mathbf{u}_I is observed conditionally over a time interval [1, 20]. Below, we derive explicit closed equations for the evolution of conditional moments of \mathbf{u}_2 through CGFPE.

The Fokker-Planck equation for the probability density $p(\mathbf{u}_I, \mathbf{u}_{II}, t)$ associated with (1) is given by [7, 36]

$$p_t = -\nabla_I \cdot (F_I p) - \nabla_{II} \cdot \left((F_{II} + \Gamma \mathbf{u}_{II}) p \right) + \frac{1}{2} \nabla \cdot \nabla (Q p) + \frac{1}{2} \nabla_I \cdot \nabla_I (Q_A p) + \frac{1}{2} \nabla_{II} \cdot \nabla_{II} (Q_M p), \quad (2)$$

where $\nabla = (\nabla_I, \nabla_{II})$, and

$$Q = (\sigma_I, \sigma_{II}) \otimes (\sigma_I^T, \sigma_{II}^T), \quad Q_A = \sigma_{II,A} \otimes \sigma_{II,A}^T, \quad Q_M = (\sigma_{II,0} + \sigma_{II,M} \mathbf{u}_{II}) \otimes (\sigma_{II,0}^T + \mathbf{u}_{II}^T \sigma_{II,M}^T). \quad (3)$$

We are interested in developing exact statistical approximations for $p(\mathbf{u}_I, \mathbf{u}_{II}, t)$ which, by Bayes theorem, can be written as

$$p(\mathbf{u}_I, \mathbf{u}_{II}, t) = p(\mathbf{u}_{II} | \mathbf{u}_I, t) \pi(\mathbf{u}_I, t) \quad (4)$$

where $\pi(\mathbf{u}_1, t)$ is the marginal distribution

$$\pi(\mathbf{u}_1, t) \equiv \int p(\mathbf{u}_1, \mathbf{u}_{\text{II}}, t) d\mathbf{u}_{\text{II}}. \quad (5)$$

We first integrate (2) with respect to \mathbf{u}_{II} and use the divergence theorem to calculate that the marginal density, $\pi(\mathbf{u}_1, t)$, satisfies the Fokker-Planck equation

$$\pi_t = \mathcal{L}_{\text{FP},\text{I}} \pi, \quad (6)$$

with

$$\mathcal{L}_{\text{FP},\text{I}} \pi = -\nabla_{\text{I}} \cdot (F_{\text{I}} \pi) + \frac{1}{2} \nabla_{\text{I}} \cdot \nabla_{\text{I}} (Q_{\text{I}} \pi), \quad Q_{\text{I}} = \sigma_{\text{I}} \otimes \sigma_{\text{I}}^T. \quad (7)$$

Next, we derive the closed system of Coupled Generalized Fokker-Planck Equations (CGFPE) for the conditional moments

$$\mathcal{M}_{\alpha}(\mathbf{u}_1, t) \equiv \int \mathbf{u}_{\text{II}}^{\alpha} p(\mathbf{u}_1, \mathbf{u}_{\text{II}}, t) d\mathbf{u}_{\text{II}} = \pi(\mathbf{u}_1, t) \int \mathbf{u}_{\text{II}}^{\alpha} p(\mathbf{u}_{\text{II}}|\mathbf{u}_1, t) d\mathbf{u}_{\text{II}}. \quad (8)$$

Note that $\mathcal{M}_0(\mathbf{u}_1, t) = \pi(\mathbf{u}_1, t)$ is just the marginal density of (1) in \mathbf{u}_1 . Here and below, we use the standard multi-index notation $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_{M_{\text{II}}}) \in \mathbb{R}^{M_{\text{II}}}$ with

$$\mathbf{u}_{\text{II}}^{\alpha} \equiv (\mathbf{u}_{\text{II}})_1^{\alpha_1} (\mathbf{u}_{\text{II}})_2^{\alpha_2} \dots (\mathbf{u}_{\text{II}})_{M_{\text{II}}}^{\alpha_{M_{\text{II}}}}. \quad (9)$$

We have the following general principles for computing the vector $\mathcal{M}_{\alpha}(\mathbf{u}_1, t) \equiv (\mathcal{M}_{\alpha}(\mathbf{u}_1, t))$, $|\alpha| = N$, of conditional moments of order N :

Proposition 2.1 (Generalized Feynman-Kac formula) *The vector $\mathcal{M}_N(\mathbf{u}_1, t)$ of conditional moments of order N associated with the probability density of (1) satisfies the CGFPE*

$$\begin{aligned} \frac{\partial \mathcal{M}_N(\mathbf{u}_1, t)}{\partial t} &= \mathcal{L}_{\text{FP}} \mathcal{M}_N(\mathbf{u}_1, t) + \mathcal{L}_N(\mathbf{u}_1, t) \mathcal{M}_N(\mathbf{u}_1, t) \\ &\quad + \mathcal{F}_N(\mathbf{u}_1, \mathcal{M}_{N-1}(\mathbf{u}_1, t), \nabla_{\text{I}} \mathcal{M}_{N-1}(\mathbf{u}_1, t), \mathcal{M}_{N-2}(\mathbf{u}_1, t)), \end{aligned} \quad (10)$$

with the convention $\mathcal{M}_{-2} = \mathcal{M}_{-1} = 0$ where \mathcal{F}_N is an explicit linear function with coefficients depending on \mathbf{u}_1 of the lower order moments; \mathcal{L}_N is an $N \times N$ Feynman-Kac matrix potential which is an explicit linear function with coefficients depending on \mathbf{u}_1 of the quantities

$$\Gamma(\mathbf{u}_1, t), \quad Q_{\text{II},\text{M}} = \sigma_{\text{II},\text{M}} \otimes \sigma_{\text{II},\text{M}}^T, \quad (11)$$

which vanishes when both $\Gamma = 0$, $Q_{\text{II},\text{M}} = 0$.

The proof below immediately yields explicit formulas for \mathcal{L}_N and \mathcal{F}_N in any concrete application (see section 3 below) but a general notation for these coefficients would be tedious and unnecessary to develop here. The advantage of CGFPE in (10) is that high resolution numerical integrators can be developed for (10) to find these statistics provided M_{I} is low-dimensional or has special algebraic structure (see section 3).

The sketch of the proof below emphasizes the main contributions to the operator \mathcal{L}_N in (10). As in the derivation of (6), we first multiply the Fokker-Planck equation (2) by $\mathbf{u}_{\text{II}}^{\alpha}$ and integrate with respect to \mathbf{u}_{II} to obtain

$$\begin{aligned} \frac{\partial \mathcal{M}_N(\mathbf{u}_1, t)}{\partial t} &= \mathcal{L}_{\text{FP}} \mathcal{M}_N(\mathbf{u}_1, t) - \int \mathbf{u}_{\text{II}}^{\alpha} \cdot \nabla_{\text{II}} (\Gamma(\mathbf{u}_1, t) \mathbf{u}_{\text{II}} p) d\mathbf{u}_{\text{II}} \\ &\quad + \frac{1}{2} \int \mathbf{u}_{\text{II}}^{\alpha} \cdot \nabla_{\text{II}} \cdot \nabla_{\text{II}} (\sigma_{\text{II},\text{M}} \mathbf{u}_{\text{II}} \otimes \mathbf{u}_{\text{II}}^T \sigma_{\text{II},\text{M}}^T p) d\mathbf{u}_{\text{II}} + \dots, \end{aligned} \quad (12)$$

where “+ . . .” denotes all the remaining terms which define the recursive source term \mathcal{F}_N . We simplify (12) using integration by parts of the two last terms on the right hand side; namely

$$- \int \mathbf{u}_{\text{II}}^\alpha \cdot \nabla_{\text{II}} \left(\Gamma(\mathbf{u}_i, t) \mathbf{u}_{\text{II}} p \right) d\mathbf{u}_{\text{II}} = \int \nabla_{\text{II}} \mathbf{u}_{\text{II}}^\alpha \cdot \left(\Gamma(\mathbf{u}_i, t) \mathbf{u}_{\text{II}} \right) p d\mathbf{u}_{\text{II}} = \mathcal{L}_N^{(1,2)} \mathcal{M}_N(\mathbf{u}_i, t), \quad (13)$$

and

$$\begin{aligned} \frac{1}{2} \int \mathbf{u}_{\text{II}}^\alpha \cdot \nabla_{\text{II}} \cdot \nabla_{\text{II}} \left(\sigma_{\text{II},M} \mathbf{u}_{\text{II}} \otimes \mathbf{u}_{\text{II}}^T \sigma_{\text{II},M}^T p \right) d\mathbf{u}_{\text{II}} = \\ \frac{1}{2} \int \nabla_{\text{II}} \cdot \nabla_{\text{II}} \mathbf{u}_{\text{II}}^\alpha \cdot \left(\sigma_{\text{II},M} \mathbf{u}_{\text{II}} \otimes \mathbf{u}_{\text{II}}^T \sigma_{\text{II},M}^T \right) p d\mathbf{u}_{\text{II}} = \mathcal{L}_N^{(2,2)} \mathcal{M}_N(\mathbf{u}_i, t) \end{aligned} \quad (14)$$

so that $\mathcal{L}_N = \mathcal{L}_N^{(1,2)} + \mathcal{L}_N^{(2,2)}$ in (10). The remaining terms in “+ . . .” are computed explicitly by similar integration by parts to define \mathcal{F}_N . The correlated noise terms in (1) involving W_i which defines the noise Q in (2) determine the dependence on $\nabla \mathcal{M}_{N-1}(\mathbf{u}_i, t)$ in \mathcal{F}_N since they have the typical form

$$- \int \mathbf{u}_{\text{II}}^\alpha \nabla_i \cdot \nabla_{\text{II}} \left(\sigma_i \sigma_{\text{II}}^T p \right) d\mathbf{u}_{\text{II}} = \nabla_i \cdot \int \nabla_{\text{II}} \mathbf{u}_{\text{II}}^\alpha \left(\sigma_i \sigma_{\text{II}}^T p \right) d\mathbf{u}_{\text{II}} = \mathcal{F}_N \left(\mathbf{u}_i, \nabla_i \mathcal{M}_{N-1}(\mathbf{u}_i, t) \right). \quad (15)$$

It is worth pointing out that \mathcal{F}_N depends only on the point-wise values of $\mathcal{M}_{N-1}(\mathbf{u}_i, t)$, $\mathcal{M}_{N-2}(\mathbf{u}_i, t)$ if there are non-correlated noise interactions and $\sigma_{\text{II}} = 0$.

3 Application of the Conditional Moment PDE’s to a Non-Gaussian Test Model

Here, we develop the simplest non-Gaussian test model where we can explicitly evaluate non-trivial statistical features utilizing the coupled system of PDE’s in (10) from §2 for the conditional moments $\mathcal{M}_\alpha(\mathbf{u}_i, t)$. We then derive and validate a numerical procedure for accurate numerical solution of the closed system of equations in (10) for the conditional moments in several stringent test problems. This explicit solution procedure is applied in §5 to understand the role of coarse-graining and non-Gaussian statistics with model error in ensemble predictions.

Clearly, the simplest models to consider with the structure as in (1) have $M_i = M_{\text{II}} = 1$ so that the recursion formulas in (10) involve scalar fields and the CGFPE are integrated in a single spatial dimension. For \mathbf{u}_i we choose the general nonlinear scalar Itô SDE

$$du_i = F_i(u_i, t)dt + \sigma_i(u_i, t)dW_i, \quad (16)$$

while for u_{II} we utilize the quasi-linear equation

$$du_{\text{II}} = \left(-u_i u_{\text{II}} + f(t) \right) dt + \sigma_{\text{II}} dW_{\text{II}}, \quad (17)$$

where $f(t)$ does not depend on u_i , and the noise σ_{II} is constant. Note that u_i enters in (17) as a multiplicative coefficient and fluctuations in u_i can introduce growth and intermittent instabilities with highly non-Gaussian behavior even when u_i in (16) has a positive mean [3, 5, 25]. The stochastic models for u_i in (16) will vary from linear stochastic models (a special case of the SPEKF models for filtering [9, 34, 35, 25]) to cubic nonlinear models with additive and multiplicative noise [25]. For systems with dynamics as in (16)-(17), the closed equations for the conditional moments \mathcal{M}_α in (8) become

$$\frac{\partial}{\partial t} \mathcal{M}_N(u_i, t) = \mathcal{L}_{\text{FP}} \mathcal{M}_N(u_i, t) - N u_i \mathcal{M}_N(u_i, t) + N f(t) \mathcal{M}_{N-1}(u_i, t) + \frac{1}{2} N(N-1) \sigma_{\text{II}}^2 \mathcal{M}_{N-2}(u_i, t), \quad (18)$$

where $N = 0, 1, \dots, N_{\text{max}}$ and $\mathcal{M}_{-2} = \mathcal{M}_{-1} = 0$. Such models illustrate a wide range of intermittent non-Gaussian behavior mimicking that in vastly more complex systems [22]. These simple yet revealing models will be used in §5 to study various new aspects of model error in ensemble predictions for non-Gaussian turbulent systems.

3.1 Validation of a Numerical Method for Solving the CGFPE

Determination of the time evolution of the conditional moments \mathcal{M}_α in (8) requires an accurate numerical procedure for solving the inhomogeneous system of coupled Fokker-Planck equations (CGFPE) in (10). The algorithms discussed below apply to the case when $u_I \in \mathbb{R}$ (i.e., $M_I = 1$) which is sufficient for our purposes and leads to many new insights on model error in imperfect ensemble predictions of turbulent systems with positive Lyapunov exponents, as discussed in §5. Similar to the case of the homogeneous Fokker-Planck equation, solving the inhomogeneous CGFPE system (10) for $M_I \geq 3$ poses a formidable challenge which, conveniently, is unnecessary here.

Here, the coupled system in (10) is solved using the third-order temporal discretization through the backward differentiation formulas (e.g., [17]) and the second-order spatial discretization via the finite volume method [19] (see Appendix A for details). The performance of the numerical procedure for solving CGFPE in one spatial dimension (i.e., $u_I \in \mathbb{R}$ in (10)) is tested in the following widely varying dynamical configurations:

- (i) Dynamics with time-invariant statistics on the attractor/equilibrium with
 - Nearly Gaussian marginal equilibrium PDFs in u_{II} and linear Gaussian dynamics for u_I in (16),
 - Fat-tailed marginal equilibrium PDFs in u_{II} and linear Gaussian dynamics for u_I in (16),
 - Highly non-Gaussian marginal equilibrium PDFs in u_{II} and cubic dynamics for u_I in (16) with highly skewed equilibrium PDFs.
- (ii) Dynamics with time-periodic statistics on the attractor with time-periodic regime switching between nearly Gaussian and highly skewed regimes with cubic dynamics for u_I in (16) and highly non-Gaussian dynamics of u_{II} in (17).

Below, we introduce the relevant test models in §3.1.1 and provide evidence for good accuracy of the developed technique in §3.1.2, as well as its advantages over direct Monte Carlo sampling.

3.1.1 Non-Gaussian test models for validating CGFPE

Here, we consider two non-Gaussian models with intermittent instabilities and with the structure as in (16)-(17) where we adopt the following notation

$$u_I = \gamma, \quad u_{II} = u.$$

The first model is a simplified version of the SPEKF model developed originally for filtering turbulent systems with stochastically parameterized unresolved variables [9, 8, 15, 35, 34] and given by

$$\begin{aligned} a) \quad d\gamma &= \left(-d_\gamma(\gamma - \hat{\gamma}) + f_\gamma(t) \right) dt + \sigma_\gamma dW_\gamma \\ b) \quad du &= \left(-\gamma u + f_u(t) \right) dt + \sigma_u dW_u. \end{aligned} \tag{19}$$

Note that despite the Gaussian dynamics of the damping fluctuations γ , the dynamics of u in (19) can be highly non-Gaussian with intermittently positive Lyapunov exponents even when the equilibrium mean, $\hat{\gamma}$, is positive [3, 5, 4, 25]. The system (19) possesses a wide range of turbulent dynamical regimes ranging from highly non-Gaussian dynamics with intermittency and fat-tailed marginal PDFs for u to laminar regimes with nearly Gaussian statistics; a detailed discussion of properties of this system can be found in [3, 5]. In the numerical tests discussed in the next section we examine the accuracy of the numerical algorithm for solving CGFPE in the dynamical regime characterized by a highly intermittent marginal dynamics in u associated with fat-tailed marginal equilibrium PDFs for u (see figure 1 for examples of such dynamics).

The second model we examine has a cubic nonlinearity in the dynamics of the damping fluctuations, γ , and is given by

$$\begin{aligned} a) \quad d\gamma &= \left[-a\gamma + b\gamma^2 - c\gamma^3 + f_\gamma(t) \right] dt + (A - B\gamma)dW_C + \sigma_\gamma dW_\gamma, \\ b) \quad du &= \left(-\gamma u + f_u(t) \right) dt + \sigma_u dW_u, \end{aligned} \tag{20}$$

The above nonlinear model for γ with correlated additive and multiplicative noise W_C and exactly solvable equilibrium statistics was first derived in [26] as a normal form for a single low-frequency variable in climate models where the noise correlations arise through advection of the large scales by the small scales and simultaneously strong cubic damping. The nonlinear dynamics of γ has many interesting features which were studied in detail elsewhere [25]. Here, we consider a more complex problem where the dynamics of u in (20a) is coupled with γ through the quadratic nonlinearity. In the numerical tests below we focus on the particularly interesting regime where the damping fluctuations γ exhibit regime switching despite unimodality of the associated equilibrium statistics (see figure 1 for an example). This configuration represents the simplest possible test model for the analogous behavior occurring in comprehensive climate models [27, 23]. Another important configuration of (20) tested below with relevance to atmospheric/climate dynamics corresponds to time-periodic transitions in γ between a highly skewed and a nearly Gaussian phases in γ with the dynamics in u remaining highly non-Gaussian throughout the evolution (see figure 2 for an illustration of such dynamics).

The above two non-Gaussian models are utilized below to validate the accuracy of our numerical method for solving the CGFPE system (10); this framework is then used to analyze model error in imperfect predictions of turbulent non-Gaussian systems in §5.

3.1.2 Numerical tests

Here, we use the test models introduced in the previous section to analyze the performance of the numerical scheme for solving the CGFPE system (10) in one-spatial dimension. In order to assess the accuracy of the algorithm, we consider following two types of relative error in the conditional moments: the point-wise relative error in the N -th conditional moment

$$\epsilon_N(\gamma, t) = \left| \frac{\mathcal{M}_N^{\text{CGFPE}}(\gamma, t) - \mathcal{M}_N^{\text{ref}}(\gamma, t)}{\mathcal{M}_N^{\text{ref}}(\gamma, t)} \right|, \tag{21}$$

and the L^2 relative error for each fixed time

$$\epsilon_N(t) = \frac{\|\mathcal{M}_N^{\text{CGFPE}}(\gamma, t) - \mathcal{M}_N^{\text{ref}}(\gamma, t)\|_{L^2}}{\|\mathcal{M}_N^{\text{ref}}(\gamma, t)\|_{L^2}}. \tag{22}$$

The reference values for the conditional moments, $\mathcal{M}_N^{\text{ref}}$, in the above formulas are obtained from either the analytical solutions (in the case of system (19) through the formulas derived in [9]), or via the Monte Carlo estimates. The conditional moments are normalized in the standard fashion, with the conditional

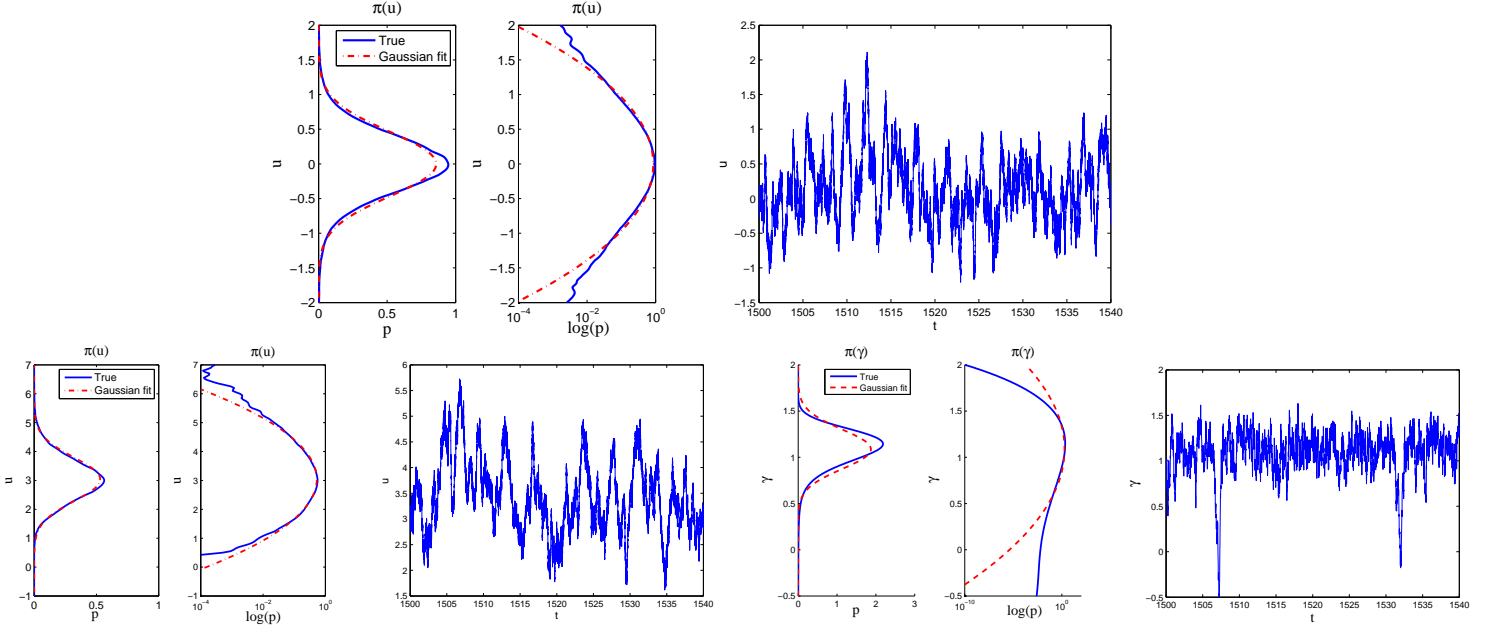


Figure 1: (Top) Marginal statistics, $p_{eq}(u)$, and a path-wise solution $u(t)$ on the attractor of the system (19) in the non-Gaussian regime with invariant measure characterized by intermittent transient instabilities and fat-tailed marginal PDFs (dynamics of γ is Gaussian in this model). (Bottom) Marginal statistics, $p_{eq}(\gamma)$, $p_{eq}(u)$, and a path-wise solutions, $\gamma(t)$, $u(t)$, on the attractor of the system (20) in the non-Gaussian regime with regime switching in the path-wise dynamics despite a unimodal, skewed marginal PDF in γ .

mean, variance, skewness and kurtosis given by

$$\widetilde{\mathcal{M}}_0(\gamma, t) = \mathcal{M}_0(\gamma, t), \quad \widetilde{\mathcal{M}}_1(\gamma, t) = \mathcal{M}_1(\gamma, t), \quad (23)$$

$$\widetilde{\mathcal{M}}_2(\gamma, t) = \int (u(t) - \mathcal{M}_1(\gamma, t))^2 p(u, \gamma, t) du = \mathcal{M}_2(\gamma, t) - \mathcal{M}_1^2(\gamma, t), \quad (24)$$

$$\begin{aligned} \widetilde{\mathcal{M}}_3(\gamma, t) &= \frac{1}{\widetilde{\mathcal{M}}_2^{3/2}(\gamma, t)} \int (u(t) - \mathcal{M}_1(\gamma, t))^3 p(u, \gamma, t) du \\ &= \frac{\mathcal{M}_3(\gamma, t) - 3\mathcal{M}_1(\gamma, t)\mathcal{M}_2(\gamma, t) + 2\mathcal{M}_1^3(\gamma, t)}{\widetilde{\mathcal{M}}_2^{3/2}(\gamma, t)}, \end{aligned} \quad (25)$$

$$\begin{aligned} \widetilde{\mathcal{M}}_4(\gamma, t) &= \frac{1}{\widetilde{\mathcal{M}}_2^2(\gamma, t)} \int (u(t) - \mathcal{M}_1(\gamma, t))^4 p(u, \gamma, t) du \\ &= \frac{\mathcal{M}_4(\gamma, t) - 4\mathcal{M}_1(\gamma, t)\mathcal{M}_3(\gamma, t) + 6\mathcal{M}_1^2(\gamma, t)\mathcal{M}_2(\gamma, t) - 3\mathcal{M}_1^4(\gamma, t)}{\widetilde{\mathcal{M}}_2^2(\gamma, t)}. \end{aligned} \quad (26)$$

The L^2 errors for the two test models discussed in the previous section and parameters as specified below are listed in Tables 1-3; note that the errors in the conditional moments do not exceed 6% for the wide range of dynamical regimes considered. Moreover, comparison of the results in Tables 1-2 shows that the numerical algorithm developed here is more efficient and accurate than the Monte Carlo estimates, even when a relatively large sample size ($\sim 10^7$) is used in the MC simulations.

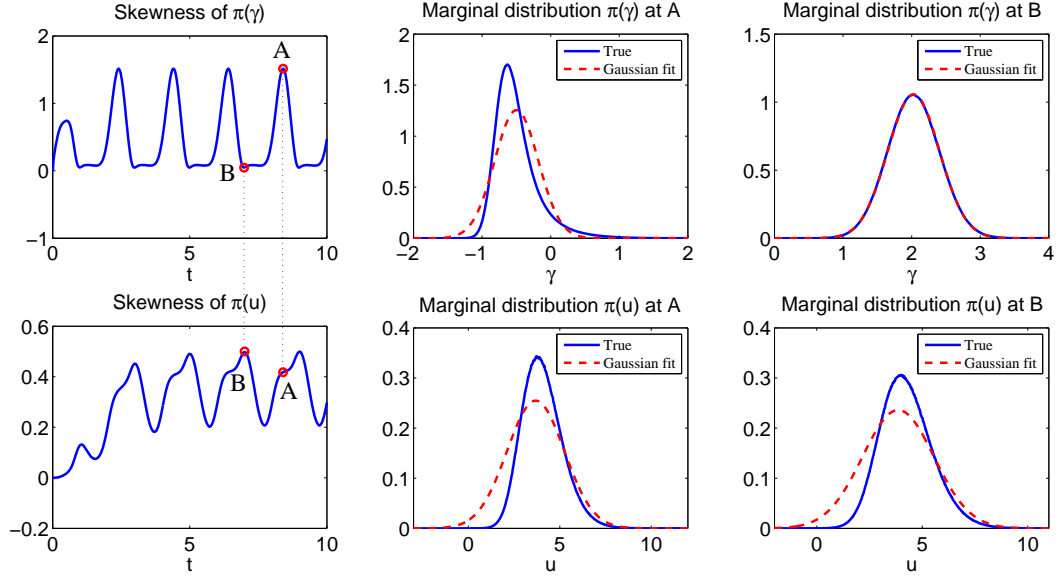


Figure 2: (Top) Time-periodic evolution of the skewness of the marginal dynamics of γ in the non-Gaussian system (20) with cubic nonlinearity in γ in the configuration where γ cycles between a highly skewed (top middle) and a nearly Gaussian (top right) phases. The phases of high/low skewness in the marginal statistics of γ are correlated with those in the marginal statistics of u ; note, however, that the dynamics of u remains highly non-Gaussian throughout the evolution. The snapshots of the marginal PDFs in u on the bottom are shown for the times indicated on the top panel.

In figure 3 we illustrate the performance of the algorithm for computing the conditional moments in (10) associated with the conditional equilibrium density $p_{eq}(u|\gamma)$ for the system (19). The system parameters, $(d_\gamma, \sigma_\gamma, \hat{\gamma}, \sigma_u)$, in (19) are chosen to represent the non-Gaussian dynamics in the regime with intermittent instabilities and a fat-tailed marginal equilibrium PDF in u ; in particular, we choose

$$\sigma_\gamma = 10, \quad d_\gamma = 10, \quad \hat{\gamma} = 3, \quad f_u = f_\gamma = 0;$$

see figure 1 for an example of the corresponding dynamics.

In figures 5-6 we illustrate the performance of our algorithm for computing the conditional moments, $\mathcal{M}_\alpha(\gamma, t)$, of u in the system (20) with the cubic nonlinearity in γ which is coupled multiplicatively to the dynamics in u . Here, we consider two distinct configurations. For constant forcing we choose the parameters in (20) in such a way that γ displays regime switching with the unimodal, highly skewed marginal equilibrium PDF for γ , while the marginal dynamics of u is highly non-Gaussian and second-order stable; this dynamical configuration can be achieved by setting, for example,

$$a = 1, \quad b = 1, \quad c = 1, \quad A = 0.5, \quad B = -2, \quad \sigma_\gamma = 1, \quad \sigma_u = 1, \quad f_u = 1, \quad f_\gamma = 3;$$

see figure 1 for an illustration of such a dynamics. For time-periodic forcing, when the dynamics in γ cycles between highly skewed and a nearly Gaussian phases while u remains highly non-Gaussian, we set

$$a = 1, \quad b = 1, \quad c = 1, \quad A = 0.5, \quad B = -0.5, \quad \sigma_\gamma = 0.5, \quad \sigma_u = 1, \quad f_u = -0.5,$$

with the time-periodic forcing in γ given by $f_\gamma(t) = 6.5 \sin(\pi t - \pi/2) + 2.5$.

Based on the results summarized in figures 3-6 and Tables 1-3, we make the following points:

	\mathcal{M}_0	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_4
System (19): Nearly Gaussian reg.	0.0031		0.0241		0.0494
System (19): Fat algebraic tail reg.	0.0225		0.0202		0.0593
System (20): High skewness reg.	0.0179	0.0181	0.0183	0.0196	0.0236

Table 1: Relative errors, ϵ_N (22), in the conditional moments \mathcal{M}_0 – \mathcal{M}_4 (23)–(26) at equilibrium for the two test models (19) and (20) with the reference input obtained from Monte Carlo estimates from 10^7 runs.

	\mathcal{M}_0	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_4
System (19): Nearly Gaussian reg.	2.1520×10^{-5}	0		0	
System (19): Fat algebraic tail reg.	3.6825×10^{-6}	0		0	
System (20): High skewness reg.	0.0018				

Table 2: Relative errors (22) in the conditional moments \mathcal{M}_0 , \mathcal{M}_1 , and \mathcal{M}_3 at equilibrium for the two test models (19) and (20) with the reference input obtained from analytical solutions.

	\mathcal{M}_0	\mathcal{M}_1	\mathcal{M}_2	\mathcal{M}_3	\mathcal{M}_4
$t_* = 7.00$	0.0185	0.0199	0.0218	0.0242	0.0271
$t_* = 8.40$	0.0299	0.0337	0.0400	0.0447	0.0561
$t_* = 7.70$	0.0309	0.0316	0.0321	0.0327	0.0332
$t_* = 9.00$	0.0182	0.0196	0.0229	0.0275	0.0330

Table 3: Relative error in time-periodic conditional moments \mathcal{M}_0 – \mathcal{M}_4 (23)–(26) for the test model (20) in the regime with transitions (see figure 1) between highly skewed and nearly Gaussian marginal densities $\pi_{att}(\gamma)$; the reference input obtained from Monte Carlo estimates from 10^7 runs.

- The numerical algorithm for solving the coupled system (10) in the CGFPE framework with $u_1 \in \mathbb{R}$ provides robust and accurate estimates for the conditional moments (8).
- The discrepancies between the estimates obtained from (10) and direct Monte Carlo estimates with large sample size ($\sim 10^7$) are below 6% for both time-periodic and time-invariant attractor statistics.
- The largest discrepancies in the normalized conditional moments obtained from CGFPE and Monte Carlo estimates the in the normalized moments occur in tail regions where the corresponding probability densities are very small.
- The developed algorithm for solving the CGFPE system (10) is more efficient and more accurate than the Monte Carlo estimates with relatively large sample sizes ($\sim 10^7$).

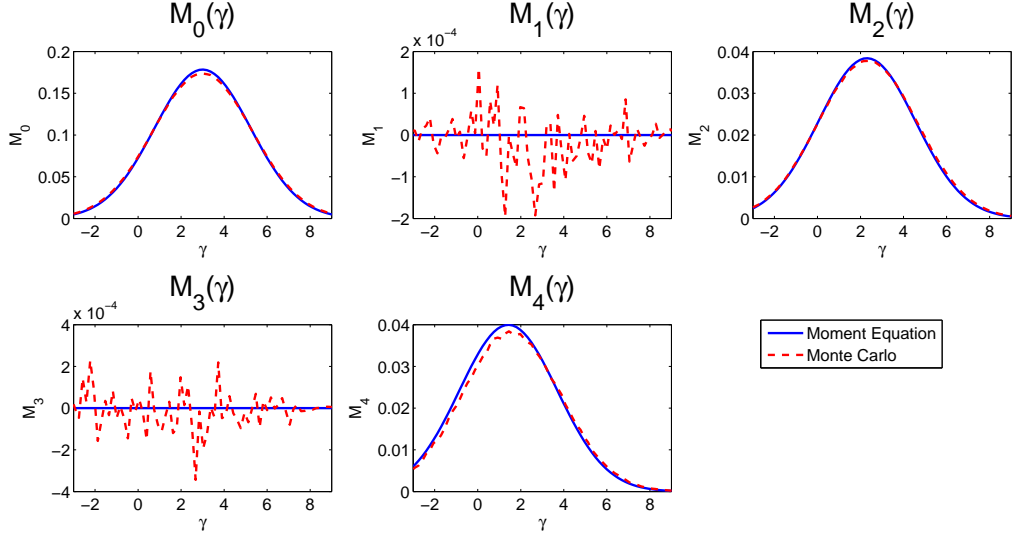


Figure 3: **Equilibrium conditional statistics of the system (19) with Gaussian damping fluctuations, intermittent instabilities and fat-tailed marginal PDF in u .** Unnormalized conditional moments, $\mathcal{M}_0(\gamma) - \mathcal{M}_4(\gamma)$, (8) of u at equilibrium of the two-dimensional non-Gaussian turbulent system (19) with intermittent instabilities due to Gaussian damping fluctuations; the results of CGFPE (10) and Monte Carlo estimates from 10^7 runs are compared. In the dynamical regime shown the marginal equilibrium PDF, $p_{eq}(u)$, is symmetric and fat-tailed due to these intermittent instabilities (see figure 1). Note the errors in the Monte Carlo estimates in the odd moments.

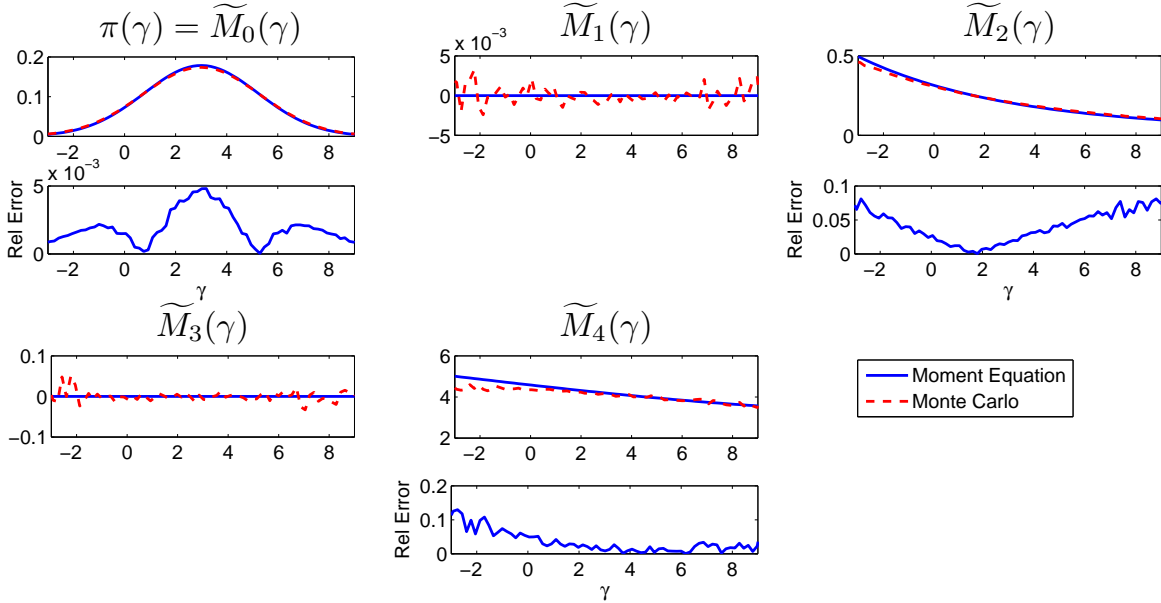


Figure 4: Same as in figure 3 above but for centered, normalized conditional moments, $\widetilde{\mathcal{M}}_0(\gamma) - \widetilde{\mathcal{M}}_4(\gamma)$, which correspond to marginal density $\pi(\gamma)$, the conditional mean, variance, skewness and kurtosis given by (23)-(26).

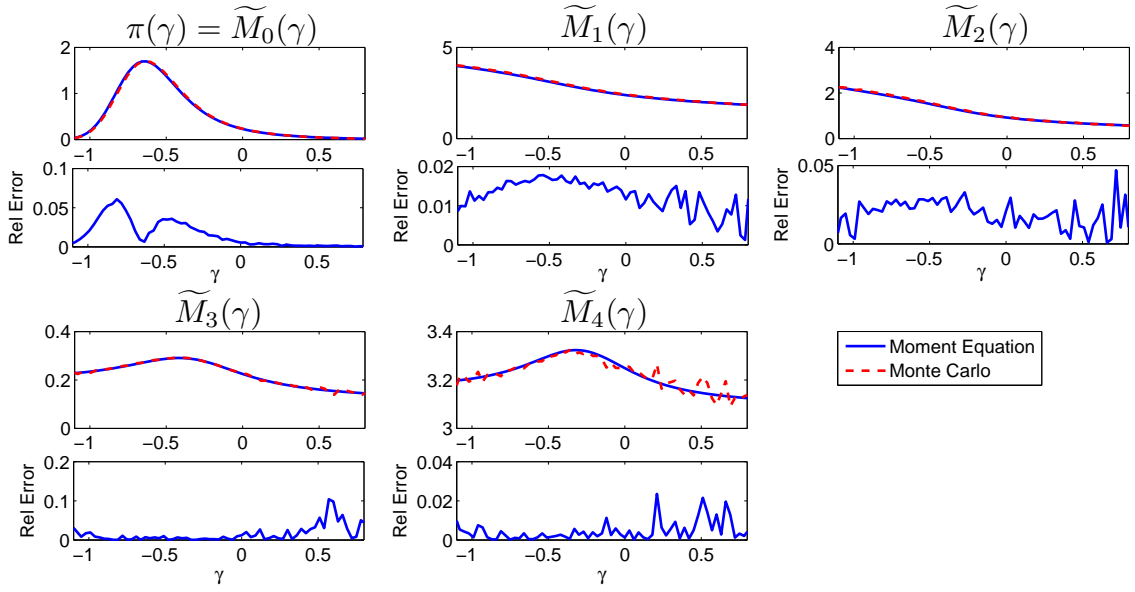


Figure 5: **Snapshot of time-periodic conditional statistics on the attractor of the system (20); cubic nonlinearity in damping fluctuations, highly skewed PDF phase.** Normalized conditional moments, $\widetilde{M}_0(\gamma, t_*) - \widetilde{M}_4(\gamma, t_*)$, (23)-(26) of u on the time-periodic attractor of the two-dimensional non-Gaussian turbulent system (20) with cubic dynamics of damping fluctuations γ ; the results obtained via CGFPE (10) and Monte Carlo simulations with 10^7 sample runs are compared at time $t_* = 8.4$ which corresponds to the highly non-Gaussian phase with highly skewed marginal PDFs, $\pi_{att}(u, t_*)$, $\pi_{att}(\gamma, t_*)$ (see figure 2). The normalized conditional moments are the conditional mean, variance skewness and kurtosis.

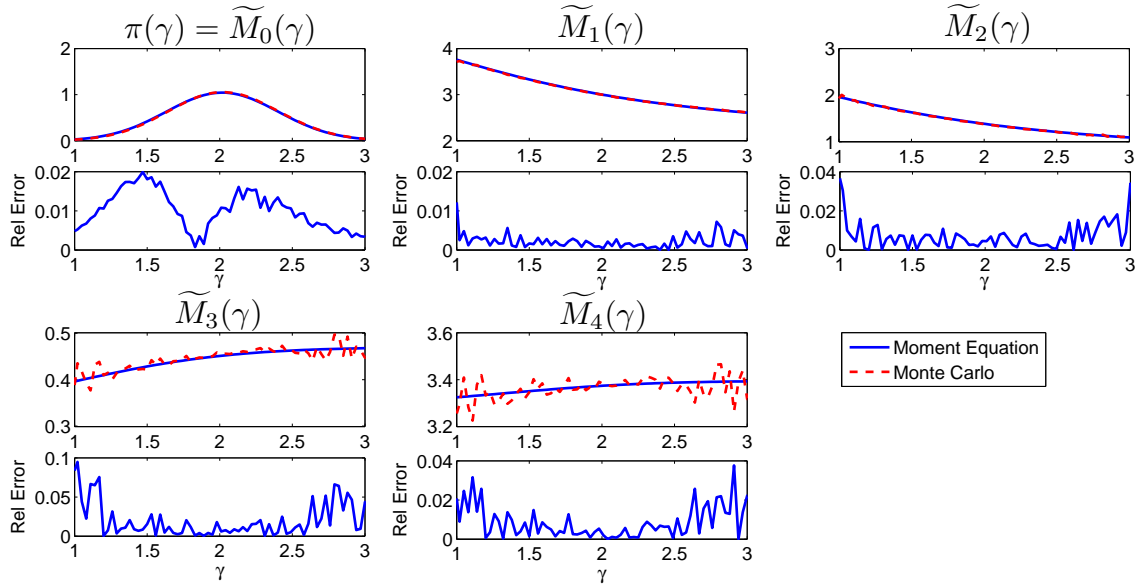


Figure 6: Same as figure 5 above but showing the normalized conditional moments, $\widetilde{M}_1(\gamma, t_*) - \widetilde{M}_4(\gamma, t_*)$ (i.e., the conditional mean, variance, skewness and kurtosis), at $t_* = 7$ which corresponds to the nearly Gaussian phase in γ but a highly skewed marginal $\pi_{att}(u, t_*)$ (see also figure 2).

4 Quantifying Model Error Through Empirical Information Theory

As discussed extensively recently [28, 29, 30, 11, 25, 5], a very natural way to quantify model error in statistical solutions of complex systems is through the relative entropy, $\mathcal{P}(p, q) \geq 0$ for two probability measures, p, q , given by

$$\mathcal{P}(p, q) = \int p \ln \frac{p}{q} = -\mathcal{S}(p) - \int p \ln q, \quad (27)$$

where

$$\mathcal{S}(p) = - \int p \ln p \quad (28)$$

is the Shannon entropy of the probability measure p . The relative entropy, $\mathcal{P}(p, q)$, measures the lack of information in q about the probability measure p . If p is the perfect density and p_M , $M \in \mathcal{M}$ is a class of probability densities, then M_1 is a better model than M_2 provided that

$$\mathcal{P}(p, p_{M_1}) < \mathcal{P}(p, p_{M_2}), \quad (29)$$

and the best model $M^* \in \mathcal{M}$ satisfies

$$\mathcal{P}(p, p_{M^*}) = \min_{M \in \mathcal{M}} \mathcal{P}(p, p_M). \quad (30)$$

There are extensive applications of information theory to improve imperfect models in climate science developed recently [28, 29, 30, 11, 25, 5]; the interested reader can see these references. The goal here is to develop and illustrate this information theory perspective on model error for direct application to estimate model error for the setup developed above in sections 2, 3; these formulas are utilized in §5 below.

We consider a probability density for the perfect model $p(\mathbf{u}_I, \mathbf{u}_{II})$ which can be written by Bayes theorem as

$$p(\mathbf{u}_I, \mathbf{u}_{II}) = p(\mathbf{u}_{II}|\mathbf{u}_I)\pi(\mathbf{u}_I), \quad (31)$$

where, here and below, $\pi(\mathbf{u}_I)$ is the marginal

$$\pi(\mathbf{u}_I) = \int p(\mathbf{u}_I, \mathbf{u}_{II}) d\mathbf{u}_{II}. \quad (32)$$

From the CGFPE procedure developed in sections 2, 3, we have exact expressions for the conditional moments up to some order L for $p(\mathbf{u}_{II}|\mathbf{u}_I)$ evolving in time already, this is a source of information loss through coarse graining of $p(\mathbf{u}_I, \mathbf{u}_{II})$. To quantify this information loss by measuring only the conditional moments up to order L , let

$$p_L(\mathbf{u}_I, \mathbf{u}_{II}) = p_L(\mathbf{u}_{II}|\mathbf{u}_I)\pi(\mathbf{u}_I) \quad (33)$$

where for each value \mathbf{u}_I the conditional density $p_L(\mathbf{u}_{II}|\mathbf{u}_I)$ satisfies the maximum entropy (least biased) criterion [31, 24, 32]

$$\mathcal{S}(p_L(\mathbf{u}_{II}|\mathbf{u}_I)) = \max_{\pi_L \in \mathcal{L}} \mathcal{S}(\pi_L(\mathbf{u}_{II})), \quad (34)$$

where \mathcal{L} is a class of marginal densities π_L with identical moments up to order L , i.e.,

$$\int \mathbf{u}_{II}^\alpha \pi_L(\mathbf{u}_{II}) d\mathbf{u}_{II} = \int \mathbf{u}_{II}^\alpha p_L(\mathbf{u}_I, \mathbf{u}_{II}) d\mathbf{u}_{II} = \int \mathbf{u}_{II}^\alpha p(\mathbf{u}_I, \mathbf{u}_{II}) d\mathbf{u}_{II}, \quad |\alpha| \leq L. \quad (35)$$

Below and in section 5, we will always apply the variational problem in (34) for $L = 2$ which guarantees that $p_L(\mathbf{u}_{II}|\mathbf{u}_I)$ is a Gaussian density with specified conditional mean and variance. In general, for L even and $L > 2$, it is a subtle issue as to whether the solution of the variational problem (34) exists [34] but here

we tacitly assume this. We remark here that highly non-Gaussian densities can have Gaussian conditional densities like $p_L(\mathbf{u}_{\text{II}}|\mathbf{u}_I)$ as discussed in §5. Natural imperfect densities with model error have the form

$$p_L^M(\mathbf{u}_I, \mathbf{u}_{\text{II}}) = p_L^M(\mathbf{u}_{\text{II}}|\mathbf{u}_I)\pi^M(\mathbf{u}_I). \quad (36)$$

The simplest model with model error is a Gaussian density, $p_G(\mathbf{u}_I, \mathbf{u}_{\text{II}})$, which is defined by its mean and variance; the standard regression formula for Gaussian densities [7] automatically guarantees that the form in (36) applies with $L = 2$ in this important case.

Another important way to generate an imperfect model with the form (36) is to have a different model [22, 25] for the stochastic dynamics of \mathbf{u}_I than that in (1) and compute the conditional moments up to order L in the approximate model through CGFPE so that the model approximations automatically have the form (36) (see section 5 below).

Here is a precise way to quantify the model error in an imperfect model in the present setup:

Proposition 4.1 *Given the perfect model distribution, $p(\mathbf{u}_I, \mathbf{u}_{\text{II}})$, its conditional approximation, $p_L(\mathbf{u}_I, \mathbf{u}_{\text{II}})$ in (33) and the imperfect model density, $p_L^M(\mathbf{u}_I, \mathbf{u}_{\text{II}})$ defined in (36), we have*

$$\mathcal{P}\left(p(\mathbf{u}_I, \mathbf{u}_{\text{II}}), p_L^M(\mathbf{u}_I, \mathbf{u}_{\text{II}})\right) = \mathcal{P}\left(p(\mathbf{u}_I, \mathbf{u}_{\text{II}}), p_L(\mathbf{u}_I, \mathbf{u}_{\text{II}})\right) + \mathcal{P}\left(p_L(\mathbf{u}_I, \mathbf{u}_{\text{II}}), p_L^M(\mathbf{u}_I, \mathbf{u}_{\text{II}})\right) \quad (37)$$

where

$$\begin{aligned} 0 \leq \mathcal{P}\left(p(\mathbf{u}_I, \mathbf{u}_{\text{II}}), p_L(\mathbf{u}_I, \mathbf{u}_{\text{II}})\right) &= \int \pi(\mathbf{u}_I) \left[\mathcal{S}(p_L(\mathbf{u}_{\text{II}}|\mathbf{u}_I)) - \mathcal{S}(p(\mathbf{u}_{\text{II}}|\mathbf{u}_I)) \right] d\mathbf{u}_I \\ &= \int \pi(\mathbf{u}_I) \mathcal{P}(p(\mathbf{u}_{\text{II}}|\mathbf{u}_I), p_L(\mathbf{u}_{\text{II}}|\mathbf{u}_I)) d\mathbf{u}_I, \end{aligned} \quad (38)$$

and

$$\begin{aligned} 0 \leq \mathcal{P}\left(p_L(\mathbf{u}_I, \mathbf{u}_{\text{II}}), p_L^M(\mathbf{u}_I, \mathbf{u}_{\text{II}})\right) &= \mathcal{P}\left(\pi(\mathbf{u}_I), \pi^M(\mathbf{u}_I)\right) + \int \pi(\mathbf{u}_I) \left[\mathcal{S}(p_L^M(\mathbf{u}_{\text{II}}|\mathbf{u}_I)) - \mathcal{S}(p_L(\mathbf{u}_{\text{II}}|\mathbf{u}_I)) \right] d\mathbf{u}_I \\ &= \mathcal{P}\left(\pi(\mathbf{u}_I), \pi^M(\mathbf{u}_I)\right) + \int \pi(\mathbf{u}_I) \mathcal{P}(p(\mathbf{u}_{\text{II}}|\mathbf{u}_I), p_L^M(\mathbf{u}_{\text{II}}|\mathbf{u}_I)) d\mathbf{u}_I. \end{aligned} \quad (39)$$

In particular, $\mathcal{P}\left(p(\mathbf{u}_I, \mathbf{u}_{\text{II}}), p_L(\mathbf{u}_I, \mathbf{u}_{\text{II}})\right)$ quantifies an intrinsic information barrier [29, 30, 11, 25, 5] for all imperfect model densities with the form as in (36).

The proof of Proposition (4.1) is by direct calculation utilizing the general identity [6]

$$\mathcal{P}\left(p_L(\mathbf{u}_I, \mathbf{u}_{\text{II}}), p_L^M(\mathbf{u}_I, \mathbf{u}_{\text{II}})\right) = \mathcal{P}\left(\pi(\mathbf{u}_I), \pi^M(\mathbf{u}_I)\right) + \int \pi(\mathbf{u}_I) \mathcal{P}\left(p(\mathbf{u}_{\text{II}}|\mathbf{u}_I), p_L^M(\mathbf{u}_{\text{II}}|\mathbf{u}_I)\right) d\mathbf{u}_I, \quad (40)$$

which is easily verified by the reader. Next, for each \mathbf{u}_I , use the general identity for least biased densities which follows from the max-entropy principle in (34) (see [24, Chapter 2])

$$\mathcal{P}\left(p(\mathbf{u}_I, \mathbf{u}_{\text{II}}), p_L^M(\mathbf{u}_I, \mathbf{u}_{\text{II}})\right) = \mathcal{P}\left(p(\mathbf{u}_I, \mathbf{u}_{\text{II}}), p_L^M(\mathbf{u}_I, \mathbf{u}_{\text{II}})\right) + \mathcal{P}\left(p(\mathbf{u}_I, \mathbf{u}_{\text{II}}), p_L(\mathbf{u}_I, \mathbf{u}_{\text{II}})\right), \quad (41)$$

and insert this in (40). Next, computing $\mathcal{P}\left(p(\mathbf{u}_I, \mathbf{u}_{\text{II}}), p_L(\mathbf{u}_I, \mathbf{u}_{\text{II}})\right)$ and $\mathcal{P}\left(p_L(\mathbf{u}_I, \mathbf{u}_{\text{II}}), p_L^M(\mathbf{u}_I, \mathbf{u}_{\text{II}})\right)$ by the formula in (40) once again, with simple algebra we arrive at the required formulas in (37)-(39).

5 Non-Gaussian Test Models for Statistical Prediction with Model Error

Here, we apply the material developed in sections 3 and 4 with $L = 2$ to gain new insight into statistical predictions with the effects of coarse graining and model error in non-Gaussian setting. In the first part of this section, we consider the effect of model error through coarse graining the statistics in a perfect model setting [18] for short, medium, and long range forecasting. In the second part of this section, we consider the effect of model error in the dynamics of \mathbf{u}_t [25] on the long range forecasting skill. The error in both the full probability density as well as the marginal densities in \mathbf{u}_t is considered.

5.1 Choice of initial statistical conditions

As already mentioned in §4, we are particularly interested in assessing the model error due to various coarse-grainings of the perfect statistics; these model errors arise naturally when either deriving the approximate least-biased conditional densities through estimating the conditional moments in the CGFPE framework of §2, or when deriving the Gaussian estimators of non-Gaussian densities. The effects of initial conditions are clearly important in the short and medium range prediction, for both the perfect and the coarse-grained statistics, and the choice of a representative set of statistical initial conditions requires some care.

In the following sections we consider the least-biased conditionally Gaussian estimators (i.e., $L = 2$ in §4) of the true statistics $p(u, \gamma, t)$, leading to the non-Gaussian densities $p_2(u, \gamma, t)$, as well as fully Gaussian approximations $p_G(u, \gamma, t)$ of the true non-Gaussian statistics $p(u, \gamma, t)$. Therefore, in order to compare the effects of coarse-graining the structure of the PDFs in a standardized setting, we consider the initial joint densities with identical second-order moments, i.e., any two initial densities, \tilde{p}_i, \tilde{p}_j , satisfy

$$\int u^\alpha \gamma^\beta \tilde{p}_i(u, \gamma) du d\gamma = \int u^\alpha \gamma^\beta \tilde{p}_j(u, \gamma) du d\gamma, \quad 0 \leq \alpha + \beta \leq 2. \quad (42)$$

Here, for simplicity we choose the initial densities with uncorrelated variables,

$$\tilde{p}_i(u, \gamma) = \tilde{\pi}_i(u) \tilde{\pi}_i(\gamma),$$

where the marginal densities $\tilde{\pi}_i(u), \tilde{\pi}_i(\gamma)$ are given by the mixtures of simple densities (see Appendix B for more details). This procedure is sufficient for the present purposes and reduces the complexity of exposition. Analogous procedure can be used to generate PDFs with correlated variables by, for example, changing the coordinate frame; such a step might be necessary when studying the model error in filtering problems.

The following set of non-Gaussian initial conditions, shown in figure 7 and constructed in the way described above, is used in the suite of tests discussed next (see also Appendix B):

- (1) $\tilde{p}_1(u, \gamma)$: Nearly Gaussian PDF with the Gaussian marginal in u being and a weakly sub-Gaussian marginal in γ .
- (2) $\tilde{p}_2(u, \gamma)$: PDF with a bimodal marginal in u and a weakly skewed marginal in γ .
- (3) $\tilde{p}_3(u, \gamma)$: Multimodal PDF with a bimodal marginal in u and a tri-modal marginal in γ .
- (4) $\tilde{p}_4(u, \gamma)$: PDF with a highly skewed marginal in u and a bimodal marginal in γ .
- (5) $\tilde{p}_5(u, \gamma)$: PDF with a weakly skewed marginal in u and a highly skewed marginal in γ .
- (6) $\tilde{p}_6(u, \gamma)$: Multimodal PDF with a Gaussian marginal in u and a tri-modal marginal in γ .
- (7) $\tilde{p}_7(u, \gamma)$: Multimodal PDF with a bimodal marginal in u and a Gaussian marginal γ .

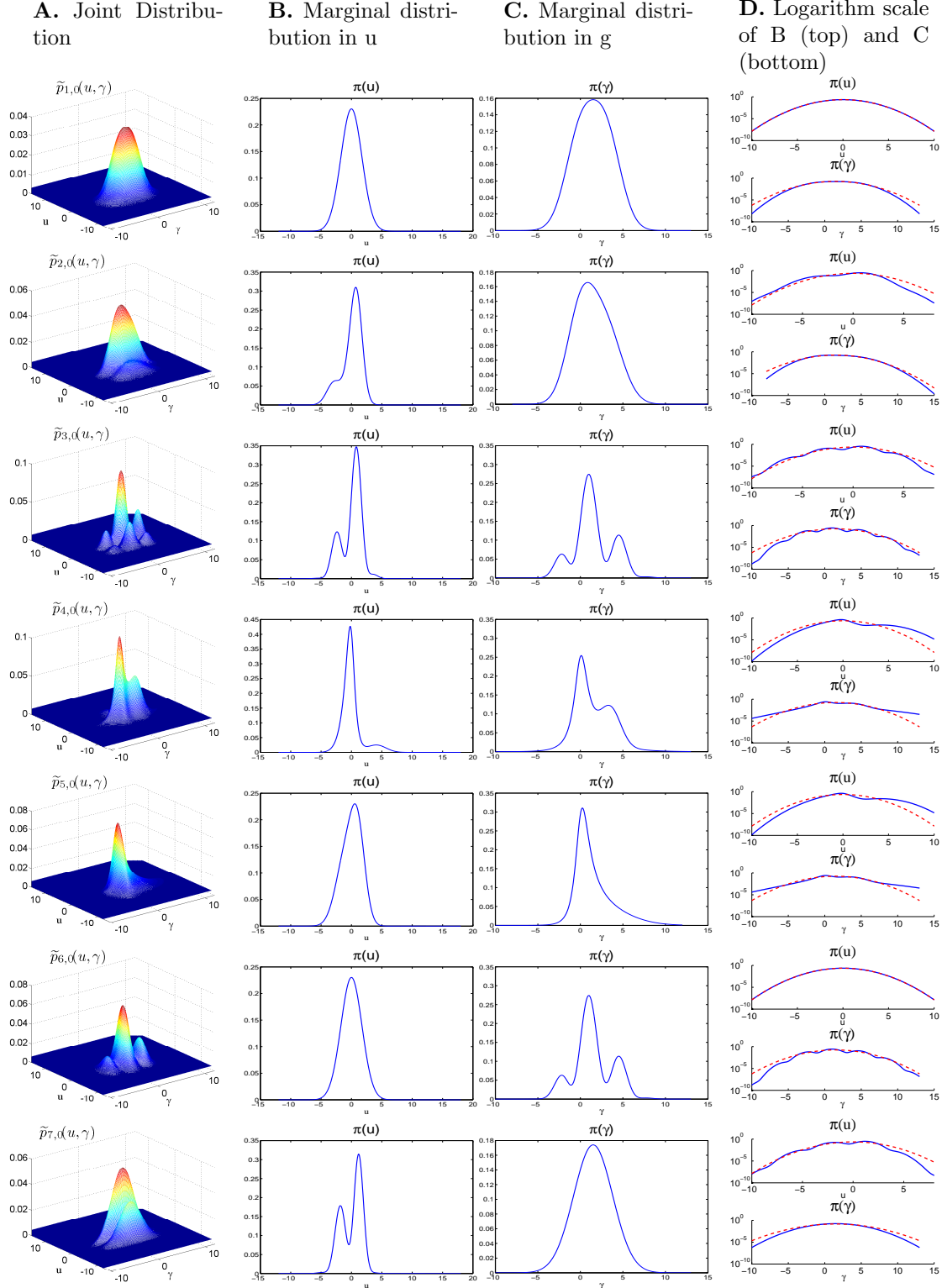


Figure 7: Set of seven non-Gaussian initial conditions with identical second-order statistics used in the tests in figures 9-20; see §5.1 and Appendix B for more details.

5.2 Ensemble prediction with model error due to coarse-graining the perfect dynamics

Here, we consider the dynamics of the same non-Gaussian system (19) with intermittent instabilities as in §3.1.1 which has the general structure as in (16)-(17). The wide range of interesting turbulent dynamical regimes [3, 5, 4, 25] makes this statistically exactly solvable system an unambiguous testbed for studying the effects of model error introduced through various coarse-grainings of the perfect density $p(u, \gamma, t)$ as discussed in §4. In this section, following the methodology introduced in §4, we focus on the model error arising from two particular coarse-grainings of the perfect model density $p(u, \gamma, t)$:

- $p_2(u, \gamma, t)$: non-Gaussian density obtained through the least-biased conditionally Gaussian approximation of the true conditional densities such that the true density, $p(u, \gamma, t)$, and the coarse-grained density, $p_2(u, \gamma, t)$, have the same first two conditional moments, i.e., for each fixed γ and t , we set

$$\mathcal{S}(p_2(u|\gamma, t)) = \max_{\mathcal{M}_{N,2}=\mathcal{M}_N} \mathcal{S}(q(u)),$$

where

$$\mathcal{M}_N = \int u^N p(u|\gamma, t) du, \quad \mathcal{M}_{N,2} = \int u^N q(u) du, \quad 0 \leq n \leq 2.$$

Note that, despite the Gaussian approximations for the conditional densities $p_2(u|\gamma, t)$, the coarse-grained joint and marginal densities $p_2(u, \gamma, t) = p_2(u|\gamma, t)\pi(\gamma, t)$, $\pi_2(u, t) = \int p_2(u, \gamma, t) d\gamma$ can be highly non-Gaussian.

- $p_G(u, \gamma, t)$: Gaussian approximation of the joint density $p(u, \gamma, t)$. The error in the Gaussian estimators, $p_G(u, \gamma, t)$ and $\pi_G(u, t) = \int p_G(u, \gamma, t) d\gamma$, arises from the least-biased approximation of the true non-Gaussian density $p(u, \gamma, t)$, which for each fixed t maximizes the entropy

$$\mathcal{S}(p_G(u, \gamma, t)) = \max_{\mathcal{M}_{i,j,G}=\mathcal{M}_{i,j}} \mathcal{S}(q(u, \gamma)),$$

subject to the following moments constraints

$$\mathcal{M}_{i,j} = \int u^i \gamma^j p(u, \gamma, t) du d\gamma, \quad \mathcal{M}_{i,j,G} = \int u^i \gamma^j q(u, \gamma) du d\gamma, \quad 0 \leq i + j \leq 2.$$

In the above set-up the conditional approximations, p_2 and π_2 , represent the best possible (least-biased) estimates for the true joint and marginal densities given the first two conditional moments. Thus, the errors $\mathcal{P}(p, p_2)$ and $\mathcal{P}(\pi, \pi_2)$ represent the intrinsic information barriers which cannot be overcome by models based utilizing two-moment approximations of the true densities (see Proposition 4.1 in §4).

In figures 9-11 we show the evolution of model error (37) due to different coarse-grainings in p_2 and p_G in the following three dynamical regimes of the system (19) with Gaussian damping fluctuations (see also figure 8):

Regime I (figure 11): Regime with plentiful, short-lasting transient instabilities in the resolved component $u(t)$ with fat-tailed marginal equilibrium densities $\pi(u)$; here, the parameters used in (19) are

$$\hat{\gamma} = 2, \sigma_\gamma = d_\gamma = 10, \sigma_u = 1, f_u = 0.$$

Regime II (figure 10): Regime with intermittent large-amplitude bursts of instability in $u(t)$ with fat-tailed marginal equilibrium densities $\pi(u)$; here, the parameters used in (19) are

$$\hat{\gamma} = 2, \sigma_\gamma = d_\gamma = 2, \sigma_u = 1, f_u = 0.$$

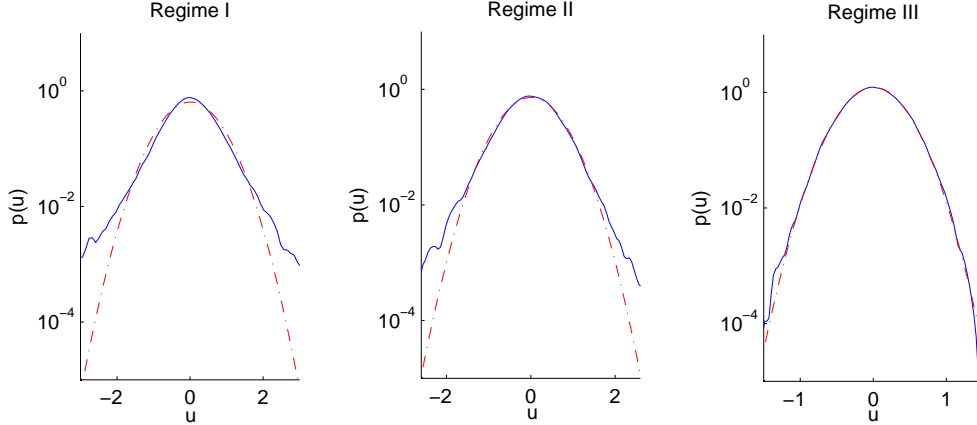


Figure 8: Three dynamical regimes of the non-Gaussian system (19) characterized by different equilibrium marginal densities $\pi_{eq}(u)$ used for studying the model error in coarse-grained densities in §5.2 (see figures 9–11). Regimes I and II of (19) are characterized by intermittent dynamics of u due to transient instabilities induced by damping fluctuations γ .

Regime III (figure 9): Regime with nearly Gaussian marginal equilibrium density $\pi(u)$; here, the parameters used in (19) are

$$\hat{\gamma} = 7, \sigma_\gamma = d_\gamma = 1, \sigma_u = 1, f_u = 0$$

In each regime the model error in the ensemble predictions is examined for the set of seven different initial densities introduced in §5.1 and figure 7 with identical second order-statistics. The evolution of the true density, $p(u, \gamma, t)$, is estimated via Monte Carlo simulations with 10^7 samples, while the coarse-grained joint densities p_2, p_G , and their marginals, π_2, π_G , are computed according to the moment-constrained maximum entropy principle in (34) using the conditional moments computed from the CGFPE procedure (10).

The top row in figures 9–11 shows the evolution of model error in the Gaussian estimators, $p_G(u, \gamma, t)$, $\pi_G(u, t)$, of the true density. The intrinsic information barrier in the Gaussian approximation (see Proposition 4.1), represented by the lack of information in the least-biased density, p_2 , based on two conditional moment constraints is shown for each regime in the middle row. It can be seen in figures 9–11 that the common feature of the model error evolution in all the examined regimes of (19) is the presence of a large error at the intermediate lead times. The source of this phenomenon is illustrated in figure 12 in regime III of (19) with a nearly Gaussian attractor statistics; the large error arises from the presence of a robust transient phase of fat-tailed dynamics in the system (19) which is poorly captured by the coarse grained statistics.

Below, we summarize the results illustrated in figures 9–12 with the focus on the model error in the Gaussian approximations $p_G(u, \gamma, t)$ and $\pi_G(u, t)$:

- For both the Gaussian estimators $p_G(u, \gamma)$, $\pi_G(u)$ and the conditionally Gaussian estimators $p_2(u, \gamma)$, $\pi_2(u)$, there exists a phase of large model error at intermediate lead times. This phase exists in all the examined regimes of (19) irrespective of the initial conditions and it arises due to a transient highly non-Gaussian fat-tailed dynamical phase in (19) which the Gaussian estimators fail to capture.
- The trends in the model error evolution for the joint and the marginal densities are similar; this is to be expected based on Proposition 4.1.

- The contributions to the model error in the Gaussian estimators $p_G(u, \gamma)$, $\pi_G(u)$ from the intrinsic information barrier, $\mathcal{P}(p, p_2)$ (see Proposition 4.1), and from the error $\mathcal{P}(p_2, p_G)$ due to the fully Gaussian vs conditionally Gaussian approximations depends on the dynamical regime.
 - The effects of the intrinsic information barrier are the most pronounced in the non-Gaussian regime I of (19) with abundant transient instabilities in u (see figure 11 and 8); in this regime the information barrier dominates the total model error. In the nearly Gaussian regime the intrinsic information barrier is negligible except at short times due to the errors in coarse-graining the highly-non-Gaussian initial conditions; see figures 9 and 7.
 - In the highly non-Gaussian regime I with abundant instabilities and fat-tailed equilibrium PDF (figure 8) the differences in model error between different initial conditions quickly become irrelevant; the intrinsic information barrier dominates the model error and there is a significant error for long range predictions in both the joint and the marginal coarse-grained densities.
 - In the non-Gaussian regime II of (19) with large amplitude intermittent instabilities the intrinsic information barrier dominates the error in the Gaussian estimators at short ranges. At intermediate lead times the error due to the fully Gaussian vs conditionally Gaussian approximations exceeds the intrinsic barrier. The error at long lead times is significantly smaller than in regime I with comparable contributions from $\mathcal{P}(p, p_2)$ and $\mathcal{P}(p_2, p_G)$.
 - In the nearly Gaussian regime III of (19) the intrinsic information barrier in the Gaussian estimators error is small and dominated by the errors in coarse-graining the non-Gaussian initial conditions.
- The intrinsic information barriers in the joint density, $\mathcal{P}(p, p_2)$, and in the marginal density, $\mathcal{P}(\pi, \pi_2)$, are comparable throughout the evolution and almost identical at short lead times.

5.3 Ensemble prediction with model error due to imperfect dynamics

Here, we focus on the model error which arises through common approximations associated with ensemble prediction: (i) errors due to imperfect/simplified dynamics and (ii) errors due to coarse-graining the statistics of the perfect system which is used for tuning the imperfect models. While the above two approximations are often simultaneously present in applications and are generally difficult to disentangle, it is important to understand the effects of these two contributions in a controlled environment which is developed below.

Similar to the framework used in the previous sections, we consider the dynamics with the structure as in the test model (16)-(17) where the non-Gaussian perfect system, as in (20), is given by

$$\begin{aligned}
a) \quad d\gamma &= \left[-a\gamma + b\gamma^2 - c\gamma^3 + f_\gamma(t) \right] dt + (A - B\gamma)dW_C + \sigma_\gamma dW_\gamma, \\
b) \quad du &= \left(-\gamma u + f_u(t) \right) dt + \sigma_u dW_u,
\end{aligned} \tag{43}$$

with cubic nonlinearity in the damping fluctuations γ . The imperfect non-Gaussian model introduces errors by assuming Gaussian dynamics in the damping fluctuations, as in (19),

$$\begin{aligned}
a) \quad d\gamma^M &= \left(-d_\gamma^M(\gamma^M - \hat{\gamma}^M) + f_\gamma^M(t) \right) dt + \sigma_\gamma^M dW_\gamma^M, \\
b) \quad du^M &= \left(-\gamma^M u^M + f_u^M(t) \right) dt + \sigma_u^M dW_u^M.
\end{aligned} \tag{44}$$

The imperfect model (44) is optimized by tuning its marginal attractor statistics, in either u^M or γ^M depending on the context, to reproduce the respective true marginal statistics. This is a prototype problem for a number of important issues; two topical examples are:

- Reduced models with a subset of unresolved variables (here γ^M) whose statistics is tuned for statistical fidelity in the resolved variables (here u^M),
- Simplification of parts of dynamics in complex multi-component models such as the coupled atmosphere-ocean-land models in climate science; in the present toy-model setting γ can be regarded as the atmospheric forcing of the ocean dynamics u .

In order to illustrate the framework developed in §2-4, we compare the model error arising in the optimized imperfect statistics, $p^{M*}(u, \gamma, t)$ or $\pi^{M*}(u, t)$, associated with (44) with the model error in $p_2(u, \gamma, t)$ or $\pi_2(u, t)$ due to the Gaussian coarse-graining of the conditional densities $p(u|\gamma, t)$ of the perfect system (43) using the CGFPE framework of §2.

In particular, we show that small model error can be achieved at medium and long lead times for imperfect predictions of the marginal dynamics $\pi^{M*}(u)$ using models with tuned unresolved dynamics γ models despite a large model error in the joint density $p^{M*}(u, \gamma)$.

5.3.1 Ensemble predictions with imperfect dynamics and time-independent statistics on the attractor

Here, we consider the perfect system (44) and its model (44) with invariant measures at their respective equilibria; this configuration is achieved by assuming constant forcing $f_\gamma=0.8220$, $f_u=-0.5$, $f_\gamma^M=0$, $f_u^M=-0.5$ in both (43) and (44). We first examine the effects of model error associated with two distinct ways of optimizing the imperfect model (44):

- (I) **Tuning the marginal equilibrium statistics of the damping fluctuations γ^M in (44) for fidelity to the true statistics of γ in (43).**

In order to tune the mean and variance of γ^M to coincide with the true moments, we simply set

$$\hat{\gamma}^M = \langle \gamma \rangle_{eq}, \quad \sigma_{\gamma^M}^2 / (2 d_\gamma^M) = Var_{eq}(\gamma), \quad (45)$$

which leads to a one-parameter family of models in (44) with correct marginal equilibrium density in γ^M . Below, we choose the damping, d_γ^M , in (44) as the free parameter and study the dependence of model error in the class of models satisfying (45) and parameterized by the damping/decorrelation time in γ^M (see figure 13). Note that only one model in this family can match both the equilibrium density, $\pi(\gamma)$, and the decorrelation time, $\tau_\gamma = \int Corr_\gamma(\tau) d\tau$, of the true damping fluctuations in (43); for such a model we have, in addition to (45),

$$\tau_\gamma^M = 1/d_\gamma^M = \tau_\gamma. \quad (46)$$

Examples of prediction error in models (44) optimized for equilibrium fidelity in γ^M but different dampings d_γ^{M*} are shown in figure 13 for the two-state unimodal regime of (43) (see figure 1). We highlight two important observations here:

- Underdamped models (19) optimized for equilibrium fidelity in the damping fluctuations γ^M have the smallest error for medium range forecasts (all models are comparable for long range forecasts). These results are similar to those reported recently in [25] where the short and medium range predictive skill of linear models with optimized marginal statistics of the unresolved dynamics was shown to often exceed the skill of models with correct marginal statistics and decorrelation time.
- Despite the striking reduction in model error intermediate lead times achieved through underdamping the unresolved dynamics in (19), caution is needed when tuning imperfect models for short range forecasts or forced response prediction where the damping, in both the resolved and unresolved dynamics, is relevant for correct system response (e.g., [25]).

(II) **Tuning the marginal equilibrium statistics of the damping fluctuations γ^M in (44) for fidelity to the true statistics of u in (43).**

This case corresponds to the situation in which we construct a simplified model of a system with unresolved degrees of freedom (here γ); these stochastically ‘superparameterized’ unresolved dynamics are then tuned to correctly reproduce the statistical features of the resolved dynamics (here u).

We consider this optimization in the Gaussian framework and optimize the imperfect model (44) by tuning the dynamics of the damping fluctuations γ^M in order to minimize the lack of information in the imperfect marginal density for the resolved variable, i.e., the optimal imperfect model satisfies

$$\mathcal{P}\left(\pi_G(u), \pi_G^{M*}(u)\right) = \min_{d_\gamma^M, \sigma_\gamma^M, \hat{\gamma}^M} \mathcal{P}\left(\pi_G(u), \pi_G^M(u)\right), \quad (47)$$

where π_G and π_G^M are the Gaussian estimators of the respective marginal densities associated with (43) and (44). With the conditional moments of u in the perfect system (43), $M_1(\gamma)$ and $M_2(\gamma)$, obtained by solving (10) in the CGFPE framework in §2, the mean and variance of $p_G(u)$ are given by

$$\bar{u} = \int \mathcal{M}_1(\gamma) d\gamma, \quad R_u = \int \mathcal{M}_2(\gamma) d\gamma - \bar{u}^2. \quad (48)$$

Analogous expressions hold for the mean and variance of $p_G^M(u)$ which are used in the optimization (47).

The two types of model optimization are compared in figure 14 for the two-state unimodal regime of (43) (see figure 1); both procedures yield comparably good results at long lead times when the model error in the marginal densities in $\pi^{M*}(u, t)$ is considered. Unsurprisingly, optimizing the marginal dynamics of u^M by tuning the dynamics of γ^M generally leads to a smaller model error for short and medium range predictions but the type of the optimization largely depends on the applications.

In figures 15-18 we illustrate the evolution of model error in the imperfect statistical prediction of (43) which is optimized according to procedure (I) above. Two non-Gaussian regimes of the true system (43) and illustrated in figure 1 are used to analyze the error in imperfect predictions with optimized models in (44).

5.3.2 Ensemble predictions with imperfect dynamics and time-periodic statistics on the attractor

We finish the analysis by considering the dynamics of the perfect system (44) and of its model (44) with time-periodic statistics on the attractor. We focus on the highly non-Gaussian regime of the perfect system (43) with the cubic nonlinearity in the damping fluctuations periodic transitions between the nearly Gaussian and highly skewed marginal density in the damping fluctuations γ which are induced by the simple time-periodic forcing

$$f_\gamma(t) = f_{\gamma,0} + f_{\gamma,1} \sin(\omega t + \phi);$$

this regime was previously used in §3.1.1 to validate the CGFPE framework (see figure 2). Similarly to the configurations studied with time-independent equilibrium statistics in the previous section, we are interested in the differences between the model error arising in the optimized imperfect dynamics, $p^{M*}(u, \gamma, t)$, $\pi^{M*}(u)$, and the error due to the coarse-graining the perfect statistics in the densities $p_2(u, \gamma, t)$, $\pi_2(u)$ obtained through the Gaussian approximations of the conditionals $p(u|\gamma, t)$.

The issue of tuning the marginal attractor statistics of the damping fluctuations γ^M in the imperfect model (44) requires more care than in the case with time-independent equilibrium statistics; this is due to the presence of an intrinsic information barrier (see §4 or [25, 5]) when tuning the statistics of the Gaussian damping fluctuations γ^M in (44) to the true statistics of (43) in γ . Similar to the time-independent case, we aim at tuning the marginal attractor statistics in γ^M for best fidelity to the true marginal statistics in

γ . Here, however, there exists an information barrier associated with the fact that the attractor variance of the Gaussian fluctuations γ^M is always constant regardless of the forcing $f_\gamma^M(t)$. One way to optimize the imperfect statistics of γ is to tune its decorrelation time, and time-averaged mean and variance on attractor to reproduce the true time-averaged quantities. However, such an approach is clearly insensitive to phase variations of the respective statistical moments. Here, instead, we optimize the imperfect model by first tuning the decorrelation times of γ^M and γ and then minimizing the period-averaged relative entropy between the marginal densities for the damping fluctuations, i.e., the optimized model (44) satisfies

$$\overline{\mathcal{P}(\pi_G(\gamma, t), \pi_G^{M*}(\gamma, t))} = \min_{\sigma_\gamma^M, \{f_\gamma^M\}} \overline{\mathcal{P}(\pi_G(\gamma, t), \pi_G^M(\gamma, t))}, \quad (49)$$

where the overbar denotes the temporal average over one period and $\{f_\gamma^M\}$ denotes a set of parameters in the forcing f_γ^M in (44); in the examples below we assume the form of the forcing f_γ^M with the same time dependence as the true one, i.e.,

$$f_\gamma^M(t) = f_{\gamma,0}^M + f_{\gamma,1}^M \sin(\omega^M t + \phi^M), \quad \text{with} \quad \omega^M = \omega, \quad \phi^M = \phi,$$

so that the optimization in (49) is carried out over a three parameter space $\{\sigma_\gamma^M, f_{\gamma,0}, f_{\gamma,1}\}$ (optimization in the phase and frequency are often crucial and interesting but we skip the discussion for the sake of brevity).

In figures 19-20 we show the model error for the coarse-grained joint and marginal densities p_2, π_2 and compare them with the model error in the joint and marginal densities associated with the optimized imperfect model (44). Here, the parameters used in (43) are

$$\begin{aligned} a = 1, \quad b = 1, \quad c = 1, \quad A = 0.5, \quad B = -0.5, \quad \sigma = 0.5, \quad \sigma_u = 1, \\ f_u = -0.5, \quad f_{\gamma,0} = 2.5, \quad f_{\gamma,1} = 6.5, \quad \omega = \pi, \quad \phi = -\pi/2. \end{aligned} \quad (50)$$

In figure 19 the decorrelation time, $\tau^{M*} = 1/d_\gamma^{M*}$, of the damping fluctuations γ^M is the same as the one in the true dynamics while the results shown in figure 20 illustrate the dependence of model error in the optimized imperfect model on the decorrelation time (see also figure 13 for the configuration with time-independent equilibrium statistics).

The following points summarize the results of §5.3.1, and §5.3.2:

- Small model error can be achieved at medium and long lead times for imperfect predictions of the marginal dynamics $\pi^{M*}(u)$ using models with tuned unresolved dynamics γ models despite a large model error in the joint density $p^{M*}(u, \gamma)$; figures 13, 14, 15, 17, 19, 20.
- The error in the coarse-grained densities $p_2(u, \gamma, t)$, $\pi_2(u, t)$ is much smaller than that in the optimized models with imperfect dynamics with $p^{M*}(u, \gamma, t)$, $\pi^{M*}(u, t)$; figures 15-18.
- The largest error in the optimized models (19) is associated with the presence of transient multimodal phases which cannot be captured by the imperfect models in the class (19); figures 15-19.
- At long lead times the model error in the joint densities, $\mathcal{P}(p(u, \gamma, t), p^{M*}(u, \gamma, t))$, is largely insensitive to the variation of the damping d_γ^{M*} (figure 20).
- The model error in the marginal densities $\pi^{M*}(u, t)$ of the optimized models has non-trivial dependence on the decorrelation time $1/d_\gamma^{M*}$ of the damping fluctuations; the overall trend is that underdamped imperfect models have smaller error in the marginals $\pi^{M*}(u, t)$ for constant or slow forcing, while the overdamped imperfect models are better for strongly varying forcing (see figures 13, 20 for two extreme cases).

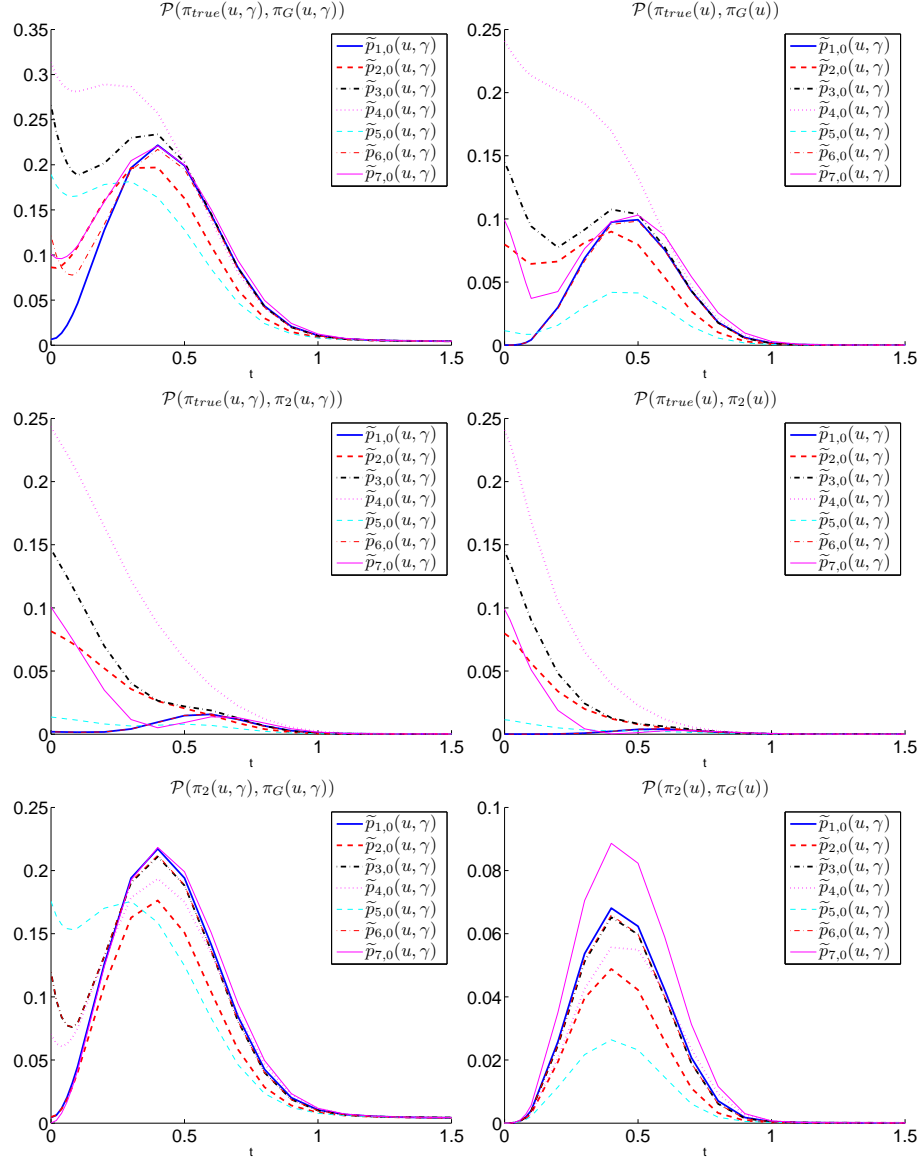


Figure 9: **Model error due to coarse-graining the perfect dynamics of the system (19) in the nearly Gaussian regime** (Regime III in figure 8). (Top two rows) Evolution of model error (37) due to different coarse-grainings of the perfect dynamics in the system (19) with Gaussian damping fluctuations; the non-Gaussian joint and marginal densities, p_2 , π_2 , are obtained through the Gaussian coarse-graining of the conditional statistics $p(u|\gamma)$ (see §3,4), while p_G , π_G are the joint and the marginal density of the Gaussian estimators (see §4). The information barrier (bottom row) equals $\mathcal{P}(p, p_G) - \mathcal{P}(p_2, p_G)$ (see (37)). The respective statistical initial conditions, all with the same second-order moments, are described in §5.1 and shown in figure 7.

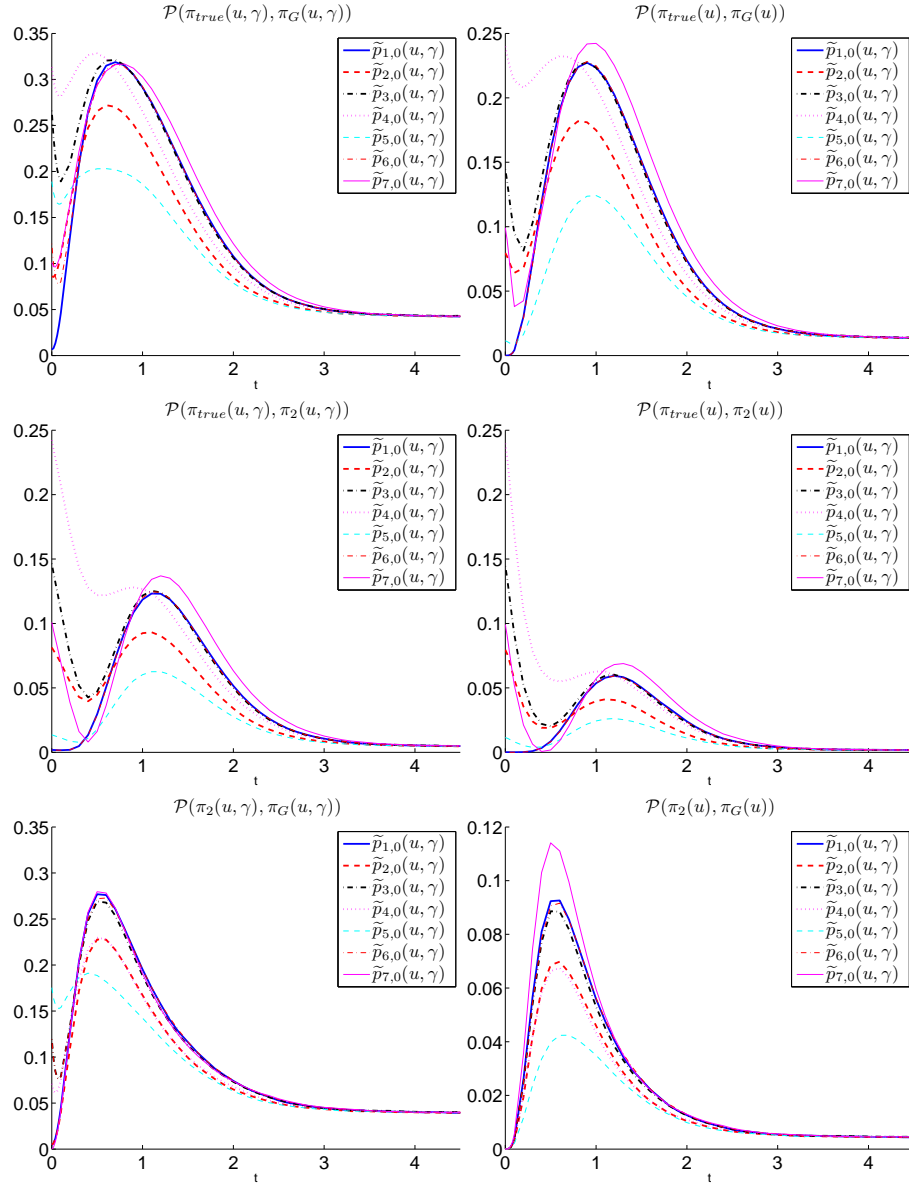


Figure 10: **Model error due to coarse-graining perfect dynamics; system (19) in regime with intermittent large amplitude instabilities.** (Top two rows) Evolution of model error (37) due to different coarse-grainings of the perfect dynamics in the system (19) with Gaussian damping fluctuations; the non-Gaussian joint and marginal densities, p_2, π_2 , are obtained through the Gaussian coarse-graining of the conditional statistics $p(u|\gamma)$ (see §3,4), while p_G, π_G are the joint and the marginal density of the Gaussian estimators (see §4). The information barrier (bottom row) equals $\mathcal{P}(p, p_G) - \mathcal{P}(p_2, p_G)$ (see §37). The respective statistical initial conditions, all with the same second-order moments, are described in §5.1 and shown in figure 7.

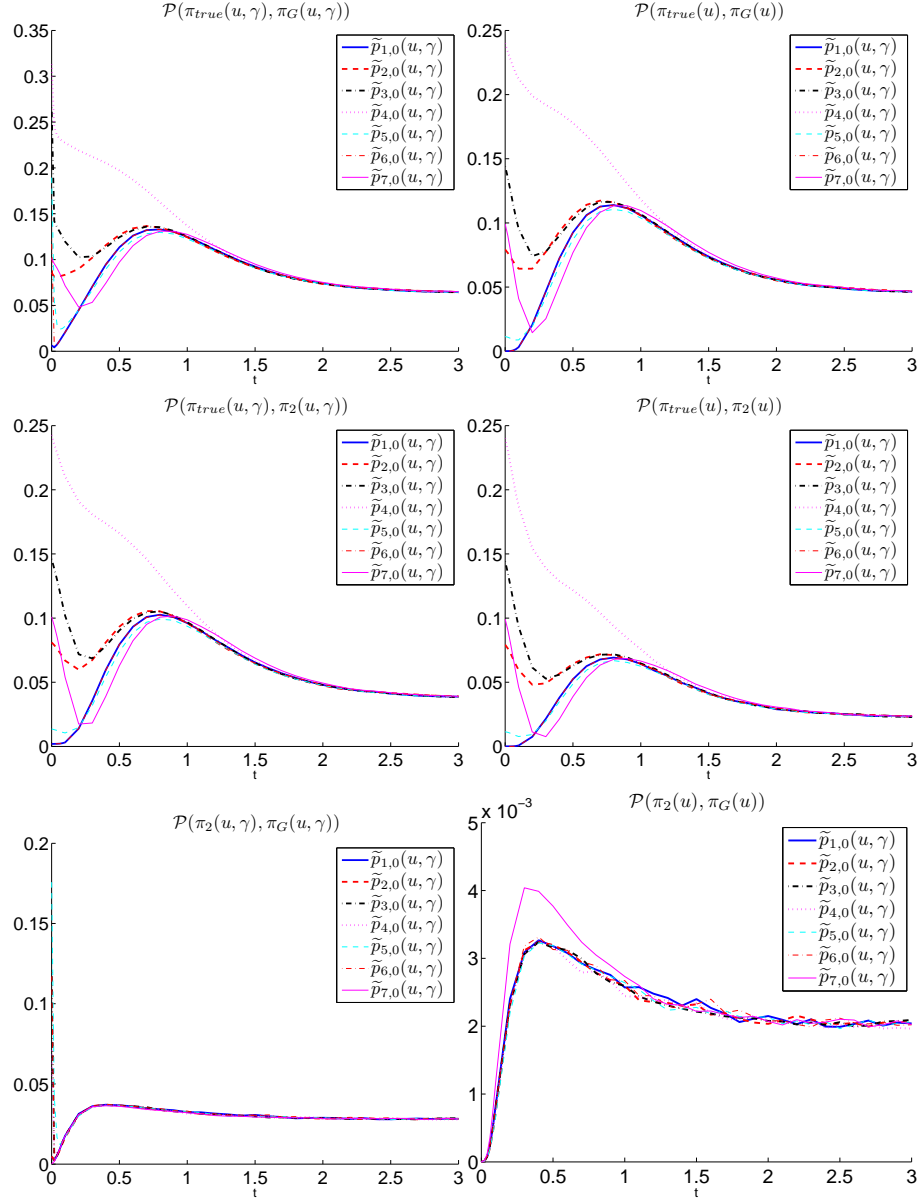


Figure 11: **Model error due to coarse-graining perfect dynamics; system (19) in regime with abundant transient instabilities.** (Top two rows) Evolution of model error (37) due to different coarse-grainings of the perfect dynamics in the system (19) with Gaussian damping fluctuations; the non-Gaussian joint and marginal densities, p_2, π_2 , are obtained through the Gaussian coarse-graining of the conditional statistics $p(u|\gamma)$ (see §3,4), while p_G, π_G are the joint and the marginal density of the Gaussian estimators (see §4). The information barrier (bottom row) equals $\mathcal{P}(p, p_G) - \mathcal{P}(p_2, p_G)$ (see §37). The respective statistical initial conditions, all with the same second-order moments, are described in §5.1 and shown in figure 7.

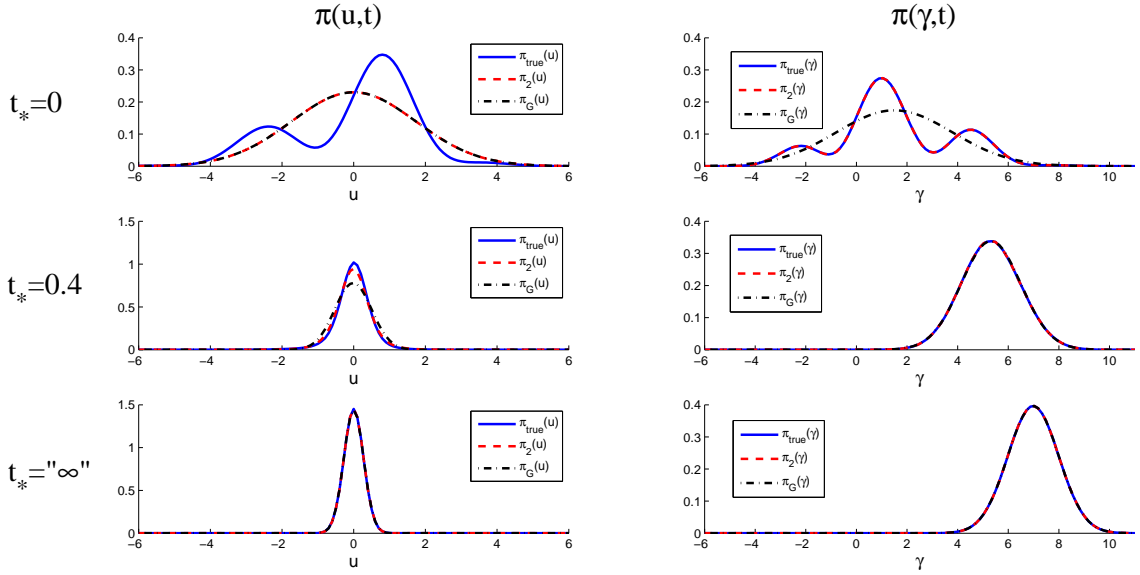


Figure 12: Three distinct stages in the statistical evolution of the system (19) illustrated for the regime with nearly Gaussian dynamics and highly non-Gaussian multimodal initial statistical conditions $\tilde{p}_3(u, \gamma)$ (figure 7); these three stages exist regardless of the dynamical regime of (19) and the form of the initial conditions (not shown). (Top) The initial configuration projected on the marginal densities at $t_* = 0$, (Middle) The fat-tailed phase in the marginal $\pi(u, t)$ corresponding to the large error phase in the coarse grained models (see figures 9-11), (Bottom) Equilibrium marginal statistics on the attractor in the regime with nearly Gaussian statistics (see Regime III in figure 8).

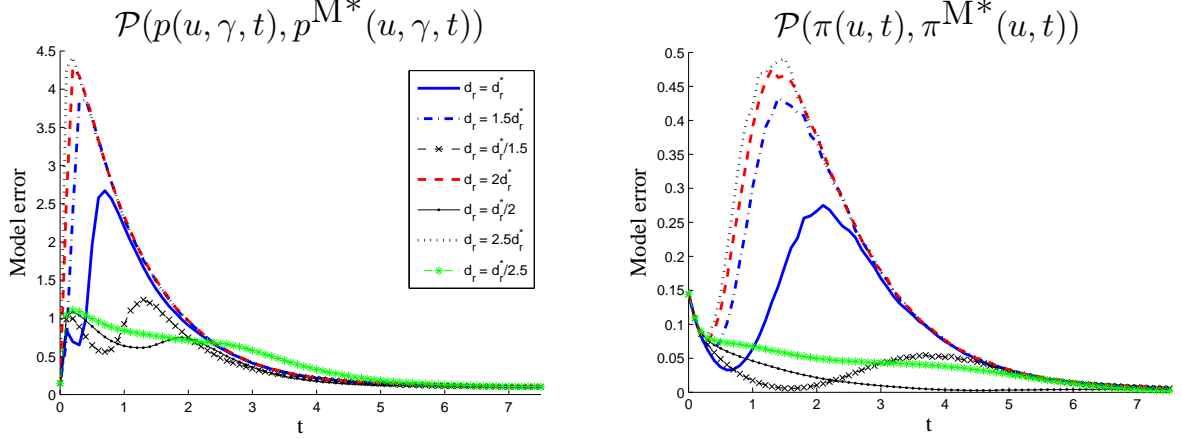


Figure 13: **Ensemble prediction of (43) with imperfect models (44); dependence of model error on the decorrelation time in the imperfect model.** Model error (37) via the relative entropy for the imperfect prediction of the system (43) using imperfect models in (44) with correct climatology in γ^M but different decorrelation times of the damping fluctuations. Note that in this case underdamped imperfect models have the best medium range prediction skill. The results shown were obtained in the skewed two-state unimodal regime (figure 1) of (43), starting from the the statistical initial conditions $\tilde{p}_1(u, \gamma)$ (see §5.1 and figure 7).

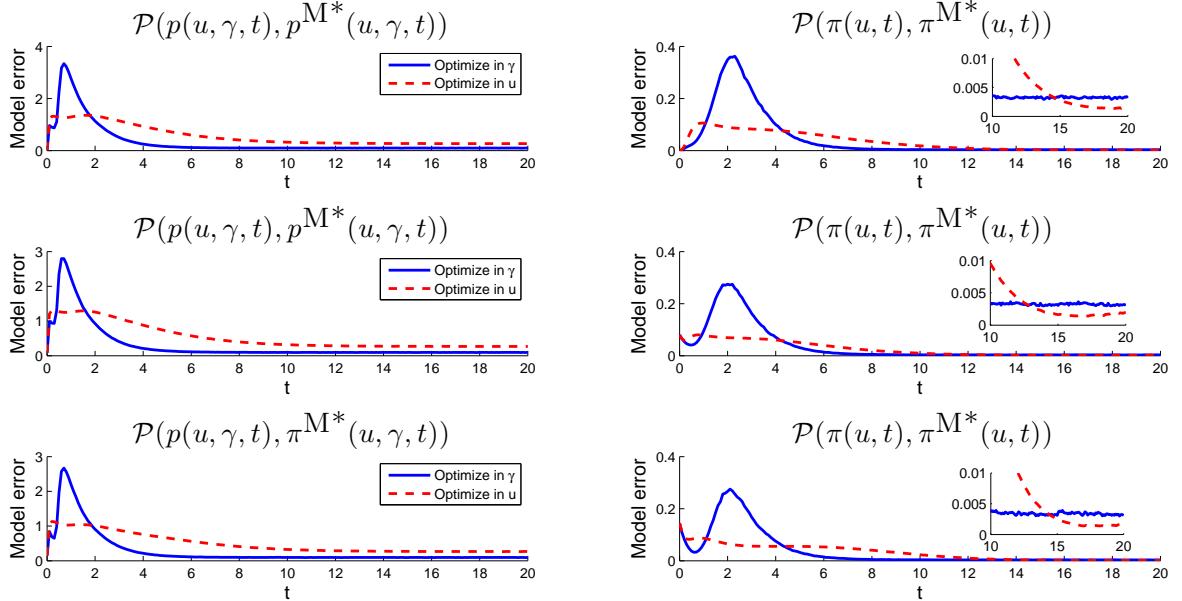


Figure 14: **Ensemble prediction of (43) with imperfect models (44); comparison of model error for different types of model optimization.** Evolution of model error (37) via the relative entropy for imperfect models (44) where the imperfect dynamics of the damping fluctuations, γ^M , is (I) either tuned to correctly reproduce the marginal equilibrium statistics of γ , or (II) γ^M is tuned to correctly reproduce the marginal equilibrium statistics of u in (43). The results shown were obtained for the perfect dynamics in (43) in regime with skewed unimodal statistics and two-state switching in the path-wise dynamics (figure 1), and for three different statistical initial conditions: (top) initial density $\tilde{p}_1(u, \gamma)$, (middle) initial density $\tilde{p}_2(u, \gamma)$, (bottom) initial density $\tilde{p}_3(u, \gamma)$; see also figure 7 and §5.1.

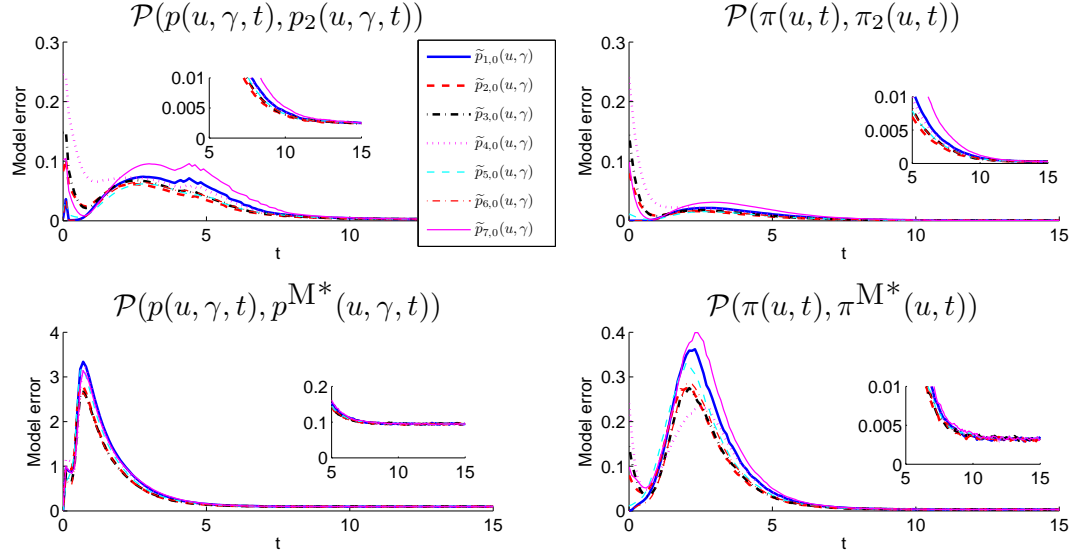


Figure 15: **Ensemble prediction with optimized imperfect dynamics**; perfect system (43) with **skewed unimodal statistics and regime switching**, imperfect model given by (44). Comparison of two types of model error in ensemble predictions: (top row) model error (37) due to coarse-graining the perfect conditional statistics (see §4), and (bottom row) model error due to imperfect dynamics (44) where γ^M is tuned for the correct marginal entropy statistics and correlation time of the damping fluctuations γ in (43). The model error via the relative entropy 37 is shown for the joint densities (left column) and the marginal densities in u (right column). The respective initial conditions are shown in figure 7.

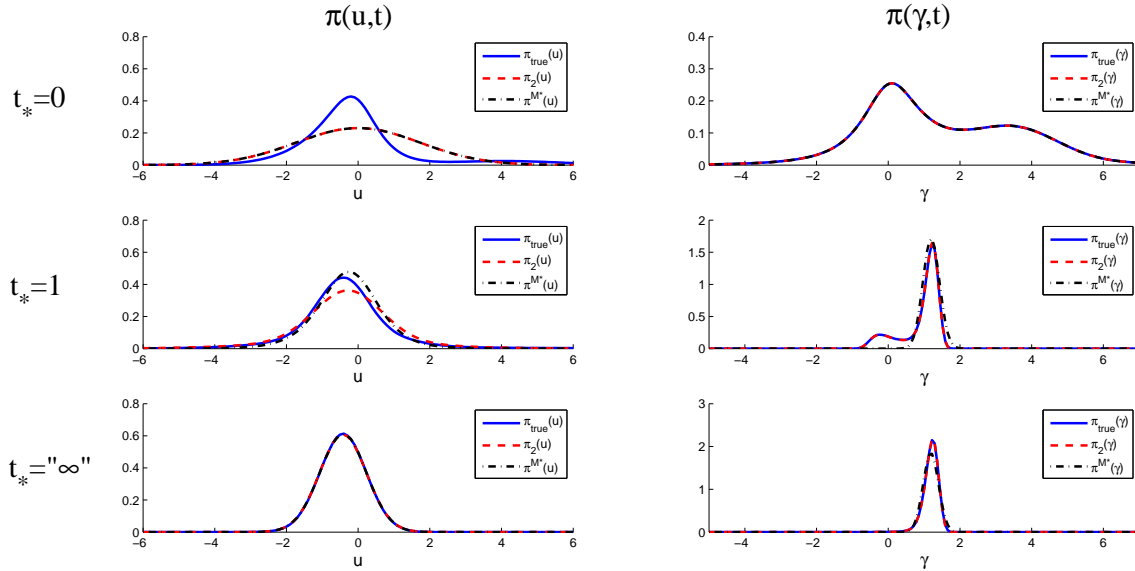


Figure 16: Three distinct stages in the statistical evolution of the system (43) and its imperfect models (44) with different contributions to model error; the example shown corresponds to the evolution from initial condition \tilde{p}_3 (see §5.1) in the regime with time-invariant statistics at equilibrium with unimodal PDF and regime switching (see figure 1). (Top) The initial configuration at $t_* = 0$, (Middle) The fat-tailed phase in the true marginal $\pi(u, t)$ corresponding to the large error phase in the coarse-grained and the Gaussian models (see figure 15), (Bottom) Equilibrium marginal statistics on the attractor with the skewed marginals $\pi(\gamma)$, $\pi(\gamma^M)$ of the damping fluctuations.

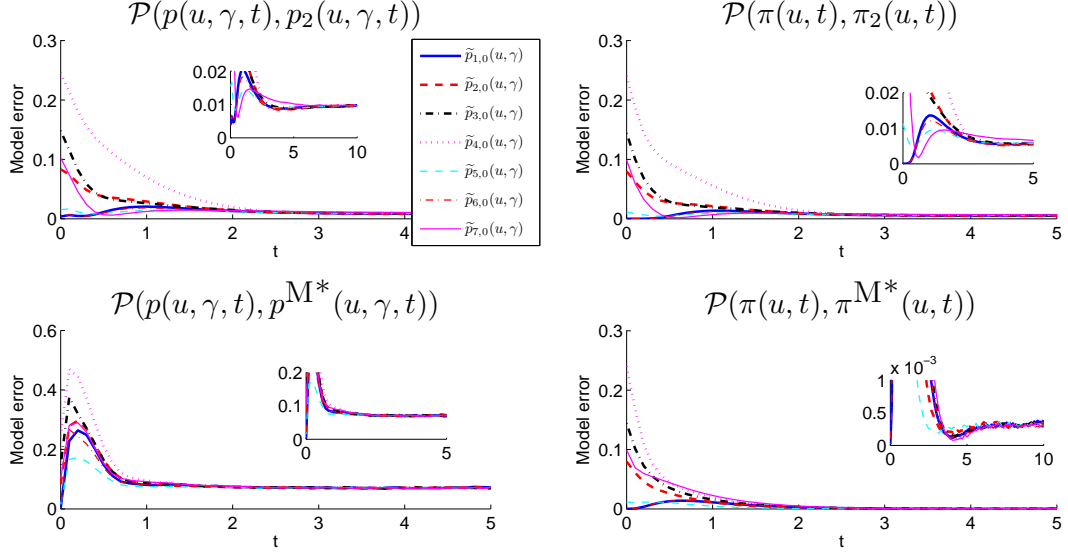


Figure 17: **Ensemble prediction with optimized imperfect dynamics**; perfect system (43) with **fat-tailed statistic**, imperfect model given by (44). Comparison of two types of model error in ensemble predictions: (top row) model error due to coarse-graining the perfect dynamics (43), and (bottom row) model error due to imperfect dynamics (44) where γ^M is tuned for the correct marginal equilibrium statistics and correlation time of the damping fluctuations γ in (43). The model error via the relative entropy (37) is shown for the joint densities (left column) and the marginal densities in u (right column). The respective initial conditions are shown in figure 7.

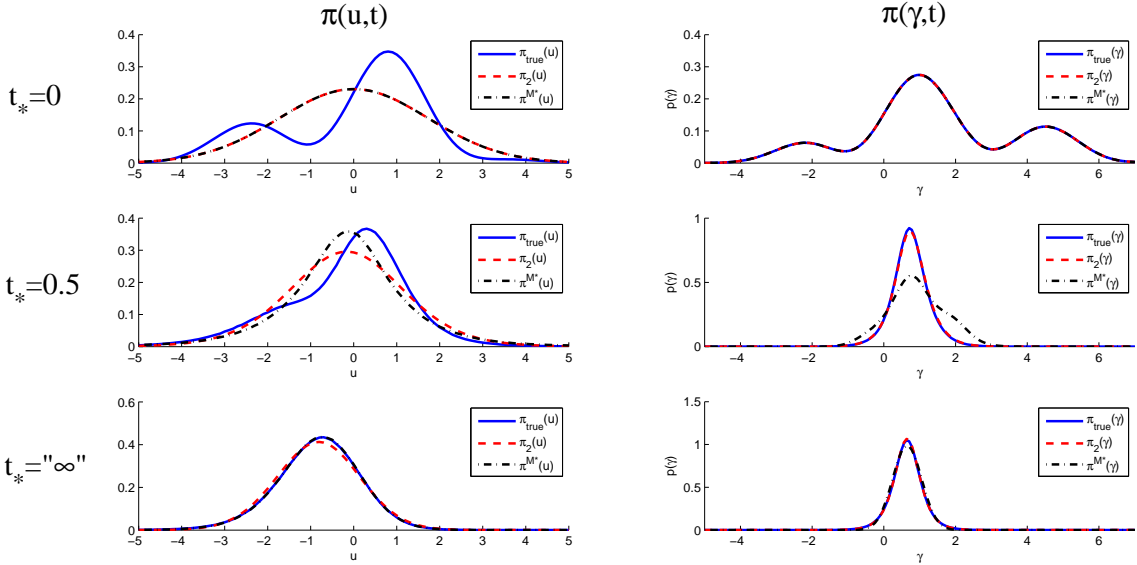


Figure 18: Three distinct stages in the statistical evolution of the perfect system (43) and its imperfect models (44) with different contributions to model error; the example shown corresponds to the evolution from initial condition \tilde{p}_3 (see §5.1) in the regime with time-invariant statistics at equilibrium and fat-tailed PDF (see figure 1). (Top) The initial configuration at $t_* = 0$, (Middle) The fat-tailed phase in the true marginal $\pi(u, t)$ corresponding to the large error phase in the coarse-grained and the Gaussian models (see figure 15), (Bottom) Equilibrium marginal statistics on the attractor with the fat-tailed marginals $\pi(\gamma)$, $\pi(\gamma^M)$ of the damping fluctuations.

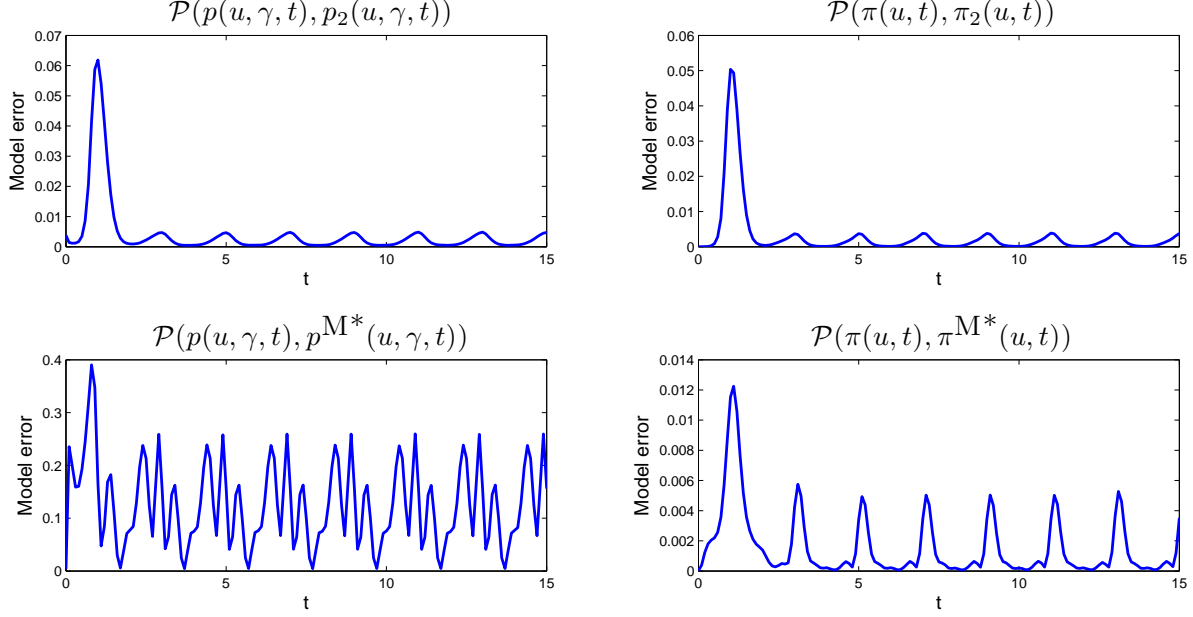


Figure 19: **Model error in imperfect optimized ensemble predictions of non-Gaussian systems with time-periodic statistics**; perfect model (43) with attractor statistics **nearly Gaussian** \longleftrightarrow **high skewness** in γ (see figure 2). Evolution of model error (37) associated with the statistical prediction of (43) in the highly non-Gaussian regime with time-periodic statistics using two non-Gaussian models: (top row) models with coarse-grained perfect conditional density $p_2(u|\gamma)$ (see §4), and (bottom row) models with imperfect dynamics of the damping fluctuations, γ^M , (44) which are optimized by matching the decorrelation time of γ and minimizing the period-averaged relative entropy (see §5.3.2 and §4 for details).

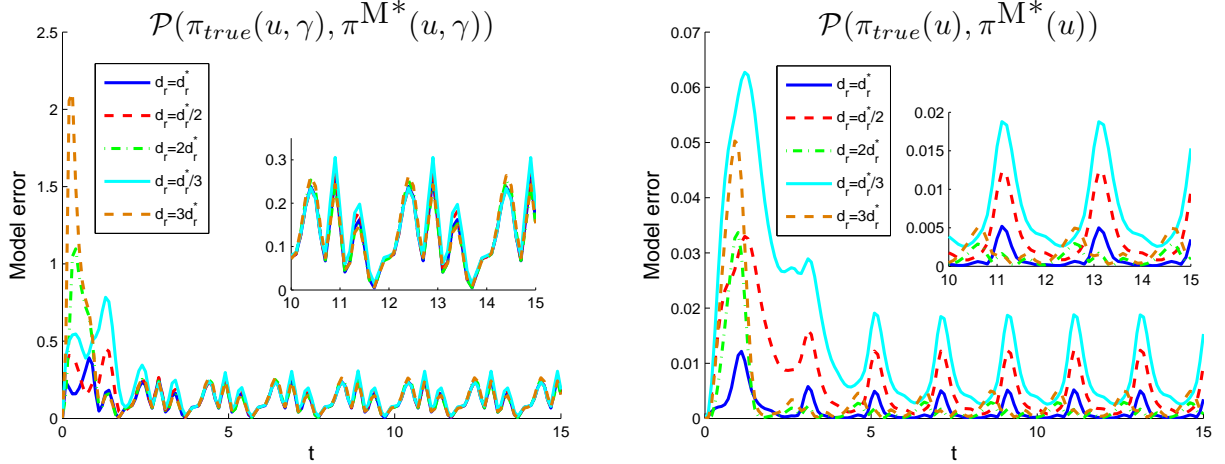


Figure 20: **Dependence of model error on decorrelation time in imperfect optimized ensemble predictions of non-Gaussian systems with time-periodic statistics**; perfect model (43) and its attractor statistics as in figure (19). Evolution of model error (37) for imperfect predictions of the true dynamics (43) using the models (44) with different decorrelation times of damping fluctuations γ^M ; $\tau_\gamma = 1/d_\gamma$ denotes the decorrelation time of γ in the true dynamics (43). For a given decorrelation time $1/d_\gamma^M$, the model (44) is optimized in the remaining parameters by minimizing the period-averaged relative entropy $\overline{\mathcal{P}(p(u, \gamma, t), p^{M^*}(u, \gamma, t))}$ (see §5.3.2 and §4 for details).

6 Concluding discussion

We considered a class of statistically exactly solvable non-Gaussian test models where the generalized Feynman-Kac formulation developed here reduces the exact behavior of conditional statistical moments to the solution of inhomogeneous Fokker-Planck equations modified by linear lower order coupling and source terms. This procedure was applied to a test models with hidden instabilities and combined with information theory to address two important issues in contemporary statistical prediction of turbulent dynamical systems: coarse-grained ensemble prediction in a perfect model and improving long range forecasting in imperfect models. Here, the focus was on studying these model errors in conditionally Gaussian approximations of the highly non-Gaussian test models. In particular, we showed that in many turbulent non-Gaussian dynamical regimes small model error can be achieved for imperfect medium and long range forecasts of the resolved variables using models with appropriately tuned statistics of the unresolved dynamics. The framework developed here, combining the generalized Feynman-Kac approach with information theory, also allows for identifying dynamical regimes with information barriers and/or transient phases in the non-Gaussian dynamics where the imperfect models fail to capture the characteristics of the true dynamics. The techniques and models developed here should be useful for quantifying and mitigating the model error in filtering and prediction in a variety of other contexts. These applications will be developed by the authors in the near future.

Acknowledgements

The research of A. Majda is partially supported by ONR grants, ONR DRI N0014-10-1-0554, and the DOD-MURI award “*Physics Constrained Stochastic-Statistical Models for Extended Range Environmental Prediction*”. M. Branicki is supported by the ONR DRI grant as a postdoctoral fellow. N. Chen is funded as a graduate research assistant on the MURI award.

A Numerical scheme for solving the CGFPE system (10)

Here, we outline the numerical method for solving the CGFPE system in (10) in one spatial dimension. This is achieved by combining the third-order Backward Differentiation Formulas (e.g., [17]) with the method of lines [19] and the second-order, finite-volume representation for (10).

Recall that the CGFPE system consists of a hierarchy of inhomogeneous Fokker-Planck equations for the conditional moments $\mathcal{M}_N(\gamma, t)$ with the forcing terms depending linearly on $\mathcal{M}_N(\gamma, t)$ and inhomogeneities depending linearly on $\mathcal{M}_{N-i}(\gamma, t)$, $i > 1$. Thus, due to the form of (10) and linearity of the forcing and inhomogeneities, we outline here the present algorithm applied to the homogeneous Fokker-Planck part of (10), written in the conservative form

$$\frac{\partial \pi}{\partial t} = -\frac{\partial}{\partial \gamma} \left[\left(F - \frac{1}{2} G_\gamma \right) \pi - \frac{1}{2} G \pi_\gamma \right], \quad (51)$$

where $\pi(\gamma, t) = \int p(u, \gamma, t) du$ and $G(\gamma, t) = \tilde{\sigma}^2(\gamma, t)$. Given the spatial grid with nodes γ_i , $i = 1, \dots, N$, and uniform spacing $\Delta\gamma$, and the approximation

$$Q_i(t) \equiv \frac{1}{\Delta\gamma} \int_{\gamma_{i-1/2}}^{\gamma_{i+1/2}} \pi(\gamma, t) d\gamma, \quad (52)$$

we discretize (51) in space through the second-order finite volume formula as

$$\begin{aligned} \frac{dQ_i}{dt} = & -\frac{1}{\Delta\gamma} \left[\left(F - \frac{1}{2}G_\gamma \right)_{i+1/2} \left(\frac{9}{16}Q_i + \frac{9}{16}Q_{i+1} - \frac{1}{16}Q_{i-1} - \frac{1}{16}Q_{i+2} \right) \right. \\ & - \left. \left(F - \frac{1}{2}G_\gamma \right)_{i-1/2} \left(\frac{9}{16}Q_{i-1} + \frac{9}{16}Q_i - \frac{1}{16}Q_{i-2} - \frac{1}{16}Q_{i+1} \right) \right] \\ & + \frac{1}{2} \frac{1}{\Delta\gamma} \left[G_{i+1/2} \left(-\frac{9}{8}Q_i + \frac{9}{8}Q_{i+1} + \frac{1}{24}Q_{i-1} - \frac{1}{24}Q_{i+2} \right) \right. \\ & \left. - G_{i-1/2} \left(-\frac{9}{8}Q_{i-1} + \frac{9}{8}Q_i + \frac{1}{24}Q_{i-2} - \frac{1}{24}Q_{i+1} \right) \right]. \end{aligned} \quad (53)$$

The above expression is obtained by seeking higher order interpolants for $Q_{i+1/2}^{n+1}$ in the standard finite-volume formulation

$$\begin{aligned} \frac{dQ_i}{dt} = & -\frac{1}{\Delta\gamma} \left[\left(F - \frac{1}{2}G_\gamma \right)_{i+1/2} Q_{i+1/2}^{n+1} - \left(F - \frac{1}{2}G_\gamma \right)_{i-1/2} Q_{i-1/2}^{n+1} \right] \\ & + \frac{1}{2\Delta\gamma} \left[G_{i+1/2} Q_{i+1/2}^{n+1} - G_{i-1/2} Q_{i-1/2}^{n+1} \right]. \end{aligned} \quad (54)$$

The second order approximations for $Q_{i+1/2}^{n+1}$ are obtained by determining the coefficients a, b, c, d in the expansion

$$\tilde{Q}_{i+1/2} = a Q_i + b Q_{i+1} + c Q_{i-1} + d Q_{i+2},$$

such that $\tilde{Q}_{i+1/2} - Q_{i+1/2}$ is of order $O((\Delta\gamma)^3)$.

The time discretization of (51) or (10) is obtained using the three-step backward differentiation formula (BDF3) [17], which belongs to the family of linear multistep methods. In particular, (51) is discretized in time as follows

$$Q^{n+3} - \frac{18}{11}Q^{n+2} + \frac{9}{11}Q^{n+1} - \frac{2}{11}Q^n = \frac{6}{11}\Delta t f(Q^{n+3}). \quad (55)$$

The above implicit formulation can be solved explicitly due to the linearity of (51) where we have

$$f(Q^{n+3}) = \begin{cases} MQ^{n+3} & \text{for solving } \mathcal{M}_0, \\ MQ^{n+3} + f_{Q_3} & \text{for solving } \mathcal{M}_i \text{ with } i > 1. \end{cases}$$

Thus, (55) can be rewritten as

$$Q^{n+3} = \begin{cases} (I - \frac{6}{11}\Delta t M)^{-1} (\frac{18}{11}Q^{n+2} - \frac{9}{11}Q^{n+1} + \frac{2}{11}Q^n) & \text{for solving } \mathcal{M}_0, \\ (I - \frac{6}{11}\Delta t M)^{-1} (\frac{18}{11}Q^{n+2} - \frac{9}{11}Q^{n+1} + \frac{2}{11}Q^n + \frac{6}{11}\Delta t f_{Q_3}) & \text{for solving } \mathcal{M}_i \text{ with } i > 1. \end{cases}$$

The (local) accuracy of the temporal discretization is $\mathcal{O}((\Delta t)^3)$. Analogous discretization is implemented for solving the inhomogeneous system (10).

B Expressions for the initial densities

Here, we list the formulas used for generating the initial densities $\tilde{p}_i(u, \gamma)$ introduced in §5.1. Recall that we chose the initial densities with uncorrelated variables,

$$\tilde{p}_i(\gamma, u) = \tilde{\pi}_i(\gamma) \tilde{\pi}_i(u),$$

where the marginal densities, $\tilde{\pi}_i(\gamma)$, $\tilde{\pi}_i(u)$, are given by the mixtures

$$\tilde{\pi}_i(\gamma) \propto \sum_n R_n(\gamma), \quad \tilde{\pi}_i(u) \propto \sum_n Q_n(u),$$

with identical first and second moments chosen as

$$\langle u \rangle = 0, \quad \langle u^2 \rangle = 3, \quad \langle \gamma \rangle = 1.5, \quad \langle \gamma^2 \rangle = 7.5 \quad \text{and} \quad \langle \gamma u \rangle = 0.$$

In particular, the seven initial densities in §5.1 with the same joint second-order statistics are obtained as follows:

(1) Joint density

$$\tilde{p}_1(u, \gamma) = \frac{1}{2} \left(R_1(\gamma) + R_2(\gamma) \right) Q_1(u),$$

where

$$R_i(\gamma) \propto \exp \left(-\frac{(\gamma - \bar{\gamma}_i)^2}{2\sigma_i^\gamma} \right), \quad Q_1(u) \propto \exp \left(-\frac{(u - \bar{u}_1)^2}{2\sigma_1^u} \right).$$

(2) Joint density

$$\tilde{p}_2(u, \gamma) = \frac{1}{4} \left(R_1(\gamma) + R_2(\gamma) \right) \left(Q_1(u) + Q_2(u) \right),$$

where

$$R_i(\gamma) \propto \exp \left(-\frac{(\gamma - \bar{\gamma}_i)^2}{2\sigma_i^\gamma} \right), \quad Q_i(u) \propto \exp \left(-\frac{(u - \bar{u}_i)^2}{2\sigma_i^u} \right) (2 + \sin(u)).$$

(3) Joint density

$$\tilde{p}_3(u, \gamma) = \frac{1}{4} \left(R_1(\gamma) + R_2(\gamma) \right) \left(Q_1(u) + Q_2(u) \right),$$

where

$$R_i(\gamma) \propto \exp \left(-\frac{(\gamma - \bar{\gamma}_i)^2}{2\sigma_i^\gamma} \right) \left(\frac{3}{2} + \sin\left(\frac{\pi\gamma}{2}\right) \right), \quad Q_i(u) \propto \exp \left(-\frac{(u - \bar{u}_i)^2}{2\sigma_i^u} \right) \left(\frac{3}{2} + \sin\left(\frac{\pi u}{2}\right) \right).$$

(4) Joint density

$$\tilde{p}_4(u, \gamma) = \frac{1}{4} \left(R_1(\gamma) + R_2(\gamma) \right) \left(Q_1(u) + Q_2(u) \right),$$

where

$$R_i(\gamma) \propto \exp \left(-\frac{(\gamma - \bar{\gamma}_i)^2}{2\sigma_i^\gamma} \right) \frac{1}{\gamma^2 + 1}, \quad Q_i(u) \propto \exp \left(-\frac{(u - \bar{u}_i)^2}{2\sigma_i^u} \right) \frac{1}{u^2 + 1}.$$

(5) Joint density

$$\tilde{p}_5(u, \gamma) = \frac{1}{2} R_1(\gamma) \left(Q_1(u) + Q_2(u) \right),$$

where

$$R_i(\gamma) \propto \exp \left(-\frac{(\gamma - \bar{\gamma}_i)^2}{2\sigma_i^\gamma} \right) \frac{1}{\gamma^2 + 1}, \quad Q_i(u) \propto \exp \left(-\frac{(u - \bar{u}_i)^2}{2\sigma_i^u} \right).$$

(6) Joint density

$$\tilde{p}_6(u, \gamma) = \frac{1}{2} \left(R_1(\gamma) + R_2(\gamma) \right) Q_1(u),$$

where

$$R_i(\gamma) \propto \exp \left(-\frac{(\gamma - \bar{\gamma}_i)^2}{2\sigma_i^\gamma} \right) \left(\frac{3}{2} + \sin\left(\frac{\pi\gamma}{2}\right) \right), \quad Q_1(u) \propto \exp \left(-\frac{(u - \bar{u}_1)^2}{2\sigma_1^u} \right).$$

(7) Joint density

$$\tilde{p}_7(u, \gamma) = \frac{1}{2} R_1(\gamma) \left(Q_1(u) + Q_2(u) \right),$$

where

$$R_1(\gamma) \propto \exp \left(-\frac{(\gamma - \bar{\gamma}_1)^2}{2\sigma_1^\gamma} \right), \quad Q_i(u) \propto \exp \left(-\frac{(u - \bar{u}_i)^2}{2\sigma_i^u} \right) \left(\frac{3}{2} + \sin\left(\frac{\pi u}{2} - 1\right) \right).$$

The parameters used in (1)-(7) are listed in the table below:

	$\bar{\gamma}_1$	$\bar{\gamma}_2$	σ_1^γ	σ_2^γ	\bar{u}_1	\bar{u}_2	σ_1^u	σ_2^u
(1)	0.0000	3.0000	3.0000	3.0000	0.0000		3.0000	
(2)	0.0506	2.9494	2.6492	3.6492	-1.1667	0.5291	2.7234	1.3088
(3)	0.0167	2.9055	2.7649	3.6316	-0.9653	0.8210	2.7216	1.7763
(4)	4.0209	4.1964	21.9235	1.2482	-1.0970	5.0522	2.2703	2.0612
(5)	5.2632		11.1937		1.0204	-1.0203	1.4575	2.4603
(6)	0.0163	2.9064	2.7691	3.6204	0.0000	5.0000	3.0000	2.0000
(7)	1.5000		5.2500		-0.7417	-0.9372	1.3784	2.4465

References

- [1] A. Bensoussan, *Stochastic control of partially observable systems*, Cambridge University Press, 1992.
- [2] A. Bourlioux and A.J. Majda., *An elementary model for the validation of flamelet approximations in non-premixed turbulent combustion.*, *Combust. Theory Modell.* **4** (2000), no. 2, 189–210.
- [3] M. Branicki, B. Gershgorin, and A.J. Majda, *Filtering skill for turbulent signals for a suite of non-linear and linear kalman filters*, *J. Comp. Phys* **231** (2012), 1462–1498.
- [4] M. Branicki and A. J. Majda, *Fundamental Limitations of Polynomial Chaos for Uncertainty Quantification in Systems with Intermittent Instabilities*, *Comm. Math. Sci.* **11** (2012), no. 1, in press.
- [5] M. Branicki and A.J. Majda, *Quantifying uncertainty for predictions with model error in non-Gaussian models with intermittency*, *Nonlinearity* **25** (2012), 2543–2578.
- [6] T. A. Cover and J. A. Thomas, *Elements of information theory*, 2 ed., Wiley-Interscience, Hoboken, 2006.
- [7] C. Gardiner, *Stochastic methods: A handbook for the natural and social sciences*, 4 ed., Springer Series in Synergetics, Springer, Berlin, 2010.
- [8] B. Gershgorin, J. Harlim, and A.J. Majda, *Improving filtering and prediction of spatially extended turbulent systems with model errors through stochastic parameter estimation*, *J. Comp. Phys.* **229** (2010), 32–57.
- [9] ———, *Test models for improving filtering with model errors through stochastic parameter estimation*, *J. Comp. Phys.* **229** (2010), 1–31.

- [10] B. Gershgorin and A. J. Majda, *A test model for fluctuation–dissipation theorems with time-periodic statistics*, *Physica D* **239** (2010), 1741–1757.
- [11] ———, *Quantifying uncertainty for climate change and long range forecasting scenarios with model errors. Part I: Gaussian models.*, *J. Climate* **25** (2012), 4523–4548.
- [12] B Gershgorin and A.J. Majda, *A nonlinear test model for filtering slow-fast systems*, *Comm. Math. Sci.* **6** (2008), 611–649.
- [13] B. Gershgorin and A.J. Majda, *Filtering a nonlinear slow-fast system with strong fast forcing*, *Comm. Math. Sci.* **8** (2009), 67–92.
- [14] B. Gershgorin and A.J. Majda, *Filtering a statistically exactly solvable test model for turbulent tracers from partial observations*, *J. Comp. Phys.* **230** (2011), 1602–1638.
- [15] J. Harlim and A.J. Majda, *Filtering turbulent sparsely observed geophysical flows.*, *Mon. Wea. Rev.* **138** (2010), no. 4, 1050–1083.
- [16] R. Hersh, *Random evolutions: a survey of results and problems*, *Rocky Mtn. J. of Math.* **4** (1974), no. 3, 443–477.
- [17] A. Iserles, *A First Course in the Numerical Analysis of Differential Equations*, Cambridge University Press, 1996.
- [18] R. Kleeman, *Information theory and dynamical system predictability*, *Entropy* **13** (2011), 612–649.
- [19] R. LeVeque, *Numerical Methods for Conservation Laws*, ETH Lectures in Mathematics Series, Birkhauser-Verlag, 1990.
- [20] R.S. Liptser and A.N. Shiryaev, *Statistics of random process*, 2 ed., Springer-Verlag, New York, 2001.
- [21] A. Majda and P. Kramer., *Simplified models for turbulent diffusion: Theory, numerical modeling, and physical phenomena*, *Phys. Reports* **314** (1999), no. 4, 237–257.
- [22] A. J. Majda, *Challenges in Climate Science and Contemporary Applied Mathematics*, *Comm. Pure Appl. Math.* **65** (2012), no. 7, 920–948.
- [23] A. J. Majda, R. Abramov, and B. Gershgorin, *High skill in low frequency climate response through fluctuation dissipation theorems despite structural instability*, *Proc. Natl. Acad. Sci. USA* **107** (2010), no. 2, 581–586.
- [24] A. J. Majda, R. V. Abramov, and M. J. Grote, *Information theory and stochastics for multiscale nonlinear systems*, CRM Monograph Series, vol. 25, Americal Mathematical Society, Providence, 2005.
- [25] A. J. Majda and M. Branicki, *Lessons in Uncertainty Quantification for Turbulent Dynamical Systems*, *Discrete Cont. Dyn. Systems* **32** (2012), no. 9, 3133–3231.
- [26] A. J. Majda, C. Franzke, and D. Crommelin, *Normal forms for reduced stochastic climate models*, *Proc. Natl. Acad. Sci.* **106** (2009), no. 10, 3649–3653.
- [27] A. J. Majda, C. Franzke, A. Fischer, and D. T. Crommelin, *Distinct metastable atmospheric regimes despite nearly Gaussian statistics: A paradigm model*, *Proc. Natl. Acad. Sci.* **103** (2006), no. 22, 8309–8314.
- [28] A. J. Majda and B. Gershgorin, *Quantifying uncertainty in climage change science through empirical information theory*, *Proc. Natl. Acad. Sci.* **107** (2010), no. 34, 14958–14963.

- [29] ———, *Improving model fidelity and sensitivity for complex systems through empirical information theory*, Proc. Natl. Acad. Sci. **108** (2011), no. 31, 10044–10049, 10.1073/pnas.1105174108.
- [30] ———, *Link between statistical equilibrium fidelity and forecasting skill for complex systems with model error*, Proc. Natl. Acad. Sci. **108** (2011), no. 31, 12599–12604, 10.1073/pnas.1108132108.
- [31] A. J. Majda, R. Kleeman, and D. Cai, *A mathematical framework for predictability through relative entropy*, Methods Appl. Anal. **9** (2002), no. 3, 425–444.
- [32] A. J. Majda and X. Wang, *Nonlinear dynamics and statistical theories for basic geophysical flows*, Cambridge University Press, Cambridge, 2006.
- [33] A.J. Majda and B. Gershgorin, *Elementary models for turbulent diffusion with complex physical features: eddy diffusivity, spectrum, and intermittency*, Phil. Trans. Roy. Soc (2011), In press.
- [34] A.J. Majda and J. Harlim, *Filtering complex turbulent systems*, Cambridge University Press, 2012.
- [35] A.J. Majda, J. Harlim, and B. Gershgorin, *Mathematical strategies for filtering turbulent dynamical systems*, Discrete Cont. Dyn. Systems **27** (2010), 441–486.
- [36] H. Risken, *The Fokker-Planck Equation: Methods of Solutions and Applications*, 2 ed., Series in Synergetics, Springer, 1996.