Lessons in uncertainty quantification for turbulent dynamical systems

LESSONS IN UNCERTAINTY QUANTIFICATION FOR TURBULENT DYNAMICAL SYSTEMS

Andrew J. Majda and Michal Branicki

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

Abstract. The modus operandi of modern applied mathematics in developing very recent mathematical strategies for uncertainty quantification in partially observed high-dimensional turbulent dynamical systems is emphasized here. The approach involves the synergy of rigorous mathematical guidelines with a suite of physically relevant and progressively more complex test models which are mathematically tractable while possessing such important features as the two-way coupling between the resolved dynamics and the turbulent fluxes, intermittency and positive Lyapunov exponents, eddy diffusivity parameterization and turbulent spectra. A large number of new theoretical and computational phenomena which arise in the emerging statistical-stochastic framework for quantifying and mitigating model error in imperfect predictions, such as the existence of information barriers to model improvement, are developed and reviewed here with the intention to introduce mathematicians, applied mathematicians, and scientists to these remarkable emerging topics with increasing practical importance.

1. **Introduction.** The 'inevitable reality' when it comes to predicting the dynamical behavior of turbulent, high-dimensional systems from nature is that the employed mathematical and numerical models need to properly account for propagation of uncertainty arising due to the limited understanding and partial observations of the true dynamics. Uncertainty Quantification (UQ) is undoubtedly an important issue in many physics, engineering, neural science and geoscience applications where complex nonlinear interactions between the resolved and unresolved processes, as well as various parametric uncertainties, need to be properly treated in order to reliably estimate the uncertain evolution of the true system on some subset of coarse-grained variables. In UQ for dynamical systems, one necessarily measures evolving estimates for the means, variances, etc. of principal quantities of interest. Thus, the mathematical framework of statistical solutions of complex dynamical systems is a starting point for UQ. The development of techniques for UQ in high-dimensional systems, attempting to estimate the evolution of uncertainty and minimize the model error in order to improve predictions, has been in its infancy until recently. There is much current activity in disparate areas of mathematics, statistics, engineering and computer science leading to ideas and techniques which are relevant for UQ in dynamical systems; examples range from Bayesian hierarchical space-time models (e.g., [121]) to variational approximations (e.g., [118, 53]),

---

to statistical-stochastic hybrids [80, 81], to Polynomial Chaos Expansions (e.g., [36, 125, 123, 63, 43, 98]) and traditional Monte Carlo methods.

However, despite the rapid improvements in numerical algorithms and availability of computing power, the fundamental mathematical problems for UQ in high-dimensional multi-scale systems remain unsolved. One well known obstacle in high-dimensional applications is the 'curse of dimensionality' which renders various UQ or data assimilation techniques based on brute-force sampling of the underlying state space unrealistic in such configurations. A more fundamental issue, however, arises when dealing with natural systems which are only partially observable yet are vastly more complicated than any conceivable model. A very useful and timely example is the the Earth's climate which is undoubtedly an extremely complex system coupling physical processes for the atmosphere, ocean, and land over a wide range of spatial and temporal scales (e.g., [21, 100]). The dynamical equations for the actual climate system are obviously unknown. Moreover, all that is available from the true climate are some coarse-grained observations of functions such as mean or variance of temperature or tracer greenhouse gases, or the large scale horizontal winds. Thus, a fundamental practical difficulty in estimating sensitivity of the climate system to external or internal perturbations lies in predicting the coarse-grained response of an extremely complex high-dimensional system from partial observations of its present equilibrium combined with imperfect models.

It is clear that UQ techniques for partially observed high-dimensional turbulent dynamical systems require a synergistic framework capable of quantifying the model error and systematic model improvement leading to reliable predictions for a subset of coarse-grained variables of the perfect system. In this context basic questions and new issues arise such as the following:

(A) How to measure, in an unbiased fashion, the statistical accuracy of a given imperfect model for reproducing the equilibrium statistics of the true system?
(B) How to make the best possible estimate of response of the true system to changes in external or internal parameters by utilizing the imperfect knowledge available of the present equilibrium?
(C) Given uncertain initial conditions, how to make the best possible predictions of the true system dynamics at short times, and at medium-long time ranges when both the initial data response and the asymptotic system properties are important?
(D) How do coarse-grained measurements of different functionals of the true equilibrium on the subset variables which are available for measurements affect the assessments in (A)-(C)? What are the weights which should be assigned these functionals to improve the performance of the imperfect models? Which new functionals of the true equilibrium should be observed in order to improve the above assessments?

The problem of quantifying and mitigating model error in coarse-grained imperfect models of complex high-dimensional systems is an important and a challenging one from both the practical and theoretical viewpoint. This task is especially difficult in turbulent systems in which energy often flows intermittently from the smaller unresolved or marginally resolved scales to impact much larger and longer spatio-temporal scales of motion of interest [72]. Atmospheric sciences, and meteorology in particular, have long been at the forefront of numerical and theoretical developments in modeling and forecasting complex turbulent systems. While the initial focus of

work on numerical general circulation modeling was concerned with the short-term behavior of the atmosphere (i.e., the "weather"), it was soon realized that the investigation of long-term trends (i.e., the "climate") could also be attacked using a blend of computational techniques and reduced models.

On the other hand, the medium range forecasts, where both the initial conditions and the asymptotic system characteristics are important, present a substantially more challenging task and improvements in this area lag behind the other two regimes. These are the regimes of large contemporary interest in climate atmosphere ocean science where extended range forecasting including the behavior of extreme events within a season or the effects of global warming over the next ten to fifty years have major societal importance worldwide. In fact, these difficulties were recognized over five decades ago by John von Neumann in the context of numerical weather prediction who famously stated "The approach [...] is to try first short-range forecasts, then long-range forecasts of those properties of the circulation that can perpetuate themselves over arbitrarily long periods, and only finally to attempt to forecast medium-long time periods ..." [101].

Recently, a stochastic-statistical framework for a systematic improvement of imperfect models and linking the statistical equilibrium fidelity of imperfect models with their sensitivity for capturing the forced response of the perfect system was proposed in [79, 80, 81, 32]. This newly emerging viewpoint blends detailed, physics-constrained dynamical modeling, stochastic parameterization and purely statistical analysis by combining empirical information theory with an appropriate fluctuation dissipation theorem, and it has at least two mathematically desirable features: (i) the approach is based on a skill measure given by the relative entropy which, unlike other metrics in UQ, is unbiased and invariant under the general change of variables [62, 56, 84, 87], and (ii) the optimization principles based on the relative entropy systematically minimize the loss of information in the imperfect model which generally does not imply minimizing the error in individual moments of the associated probability densities like the mean or covariance; this is particularly important for UQ in high-dimensional systems.

The main goal of this research expository paper is to describe these recent and ongoing developments emphasizing the remarkable new mathematical and physical phenomena that emerge from the modern applied mathematics modus operandi applied to uncertainty quantification in partially observable high-dimensional dynamical systems. The use of these ideas in applied mathematics and numerical analysis for quantifying model error, as well as in climate change science and various engineering applications for improving imperfect predictions, is only beginning but already the wealth of new important concepts and approaches warrants a detailed treatment. Here, we adopt a systematic framework for discussing these issues based on a suite of physically relevant and progressively more complex test models which are mathematically tractable while possessing such important features as the two-way coupling between the resolved dynamics and stochasticity due to the interactions with unresolved processes, intermittency and positive Lyapunov exponents, eddy diffusivity parameterization and overdamped turbulent spectra. The two recurring themes in the following discussion are:

- The utility of the unified stochastic-statistical framework for UQ and improvement of imperfect predictions with model error in turbulent dynamical systems with intermittency, positive Lyapunov exponents and hidden instabilities;

- The existence of information barriers to model improvement in classes of imperfect models, their hallmarks and consequences for UQ in turbulent dynamical systems.

The mathematical toolkit utilized below includes empirical information theory, fluctuation-dissipation theorems and systematic physics-constrained, statistical-stochastic modelling for large-dimensional turbulent dynamical systems. While the nomenclature of this paper is biased towards the climate science applications, there are many immediately obvious analogies throughout the text to problems involving high-dimensional nonlinear and nonautonomous stochastic dynamical systems with non-trivial attractors in neural science, material science, or engineering.

The important points discussed and illustrated throughout the paper for UQ in turbulent high-dimensional systems with intermittency include the following:

- The information-theoretic optimization advocated here can dramatically improve predictive performance and sensitivity of imperfect models (see also [79, 80, 81, 32]).
- Statistical equilibrium fidelity of imperfect models on the coarse-grained subset of resolved variables is necessary but not sufficient for high predictive skill and sensitivity to perturbations (see also [81]).
- There exist explicit information barriers to model improvement within a given class of imperfect models beyond which the loss of information can be only reduced by expanding the class of models to allow for more degrees of freedom (see also [73, 32]).
- The information-theoretic optimization of imperfect models requires tuning the marginal probability densities for suitable coarse-grained variables with those of the perfect system on the unperturbed attractor. In the simplest Gaussian framework such a procedure implies simultaneous tuning of means and covariances variances [80, 81].
- The sensitivity of imperfect models to capturing the effects of perturbations of the perfect system attractor can be tested via algorithms exploiting a suitable fluctuation dissipation theorem and experiments in the training phase in the unperturbed climate (see also [81]).
- Nonlinear, non-Gaussian models can have long memory of initial conditions, including the initial conditions for the unresolved processes. Linear Gaussian imperfect models cannot reproduce the response in the variance to forcing perturbations in climate change scenarios. Consequently, the change in variability due to perturbations of the system's attractor remain undetected by linear imperfect models of a nonlinear system (see also [82, 80, 81]).
- Intermittency and the presence of fat-tailed PDFs arising through the complex interactions with the unresolved processes leads to fundamental limitations in traditional UQ techniques, such as Polynomial Chaos Expansions and Monte Carlo simulations [10].

The authors hope that this article inspires other mathematicians, scientists, and engineers to explore the use of modern applied mathematics in developing strategies for UQ in reduced models of high-dimensional turbulent dynamical systems. The plan for the remainder of the paper is the following one. In Section 2, we discuss the general principles of information theory in the context of improving the statistical fidelity of imperfect models, as well as the origins of information barriers (§2.2) and the important link [79, 80, 81] between the statistical equilibrium

fidelity of imperfect models and their sensitivity (§2.3). In Section 3 we introduce the simplest linear Gaussian models with hidden dynamics which are used to illustrate the utility of the information-theoretic framework for model optimization of §2 and transparent information barriers for short and intermediate range forecasting. Other central issues arising in the context of imperfect coarse-grained predictions are developed in this elementary and instructive context. In Section §4 we discuss UQ via the statistical-stochastic framework of §2 in two physically relevant nonlinear and non-Gaussian test models with hidden dynamics and intermittency; these mathematically tractable models for the dynamics of a single mode in a turbulent signal allow for examining the effects of model error and information barriers due to moment closure approximations and dimensional reduction in an unambiguous setting. In section 5 we discuss fundamental limitations of truncated Polynomial Chaos Expansions [10] as a method for UQ in systems with intermittency and fat-tailed PDFs; two examples based on the models discussed in §4 are used there to show the failure of such methods in the simplest possible setting. Section §6 discusses UQ based on the information-theoretic framework of §2 in the complex case of spatially extended systems with turbulent energy spectra and non-trivial interactions between the mean dynamics and the turbulent fluxes. Finally, Section 7 discusses a number of instructive examples of purely numerical artifacts which might lead to severe bias in UQ techniques; these range from fake fat-tailed numerical estimates for the marginal PDFs of the true statistics to situations where unbounded growth in time of the statistical moments remains undetected by various UQ techniques. Concluding remarks and final comments on a number of topics not discussed in this article are presented in §8.

Below is the table of contents for the remainder of the article:

2. **Statistical dynamics of turbulent dynamical systems and imperfect model improvement via empirical information theory.** Successful predictions in complex nonlinear systems encountered in contemporary climate change, atmosphere-ocean, or engineering applications are difficult since the actual dynamics is a turbulent large-dimensional system with positive Lyapunov exponents and intermittency on essentially all spatio-temporal scales. Moreover, the true system is only partially observable based on some coarse-grained statistical measurements of functionals on the system's attractor; here, we assume that the true system has a global attractor in the general nonautonomous setting (i.e., the global time forward and pullback attractors exist and coincide; see, e.g., [4, 87] for a detailed discussion of these issues). Applications of empirical information theory and fluctuation dissipation theory to systematically improve imperfect model predictions have been addressed at length in the context of climate change applications in [79, 80, 81, 32]. Here, we highlight the most important concepts which are necessary for the following exposition, including the information-theoretic framework for improving imperfect model fidelity to the unperturbed attractor, as well as the important link between the attractor fidelity and sensitivity. As already mentioned, the goal of such a stochastic-statistical framework is to estimate sensitivity of the true system to external forcing and/or initial data response and predict with imperfect models the coarse-grained response of an extremely complex high-dimensional system from partial observations of the present unperturbed attractor/climate.

Consider here for concreteness the Earth's climate system. While the actual equations governing climate dynamics on earth are unknown, it is natural to assume that these dynamics are Markovian, i.e., the future state depends only on the present state, on a suitably large space of variables $\boldsymbol{v} \in I\!R^P$, $P \gg 1$. Thus, it is reasonable to assume that the perfect dynamical system for the climate is given by

$$\dot{\boldsymbol{v}} = \boldsymbol{f}(\boldsymbol{v}, t) + \sigma(\boldsymbol{v})\dot{W}(t), \tag{1}$$

where σ is a $P \times K$ noise matrix and $\dot{W}(t) \in I\!\!R^K$ is K-dimensional white noise defined for $t \in I\!\!R$. The use of statistical descriptions for the climate system dates back to early predictability studies for simplified atmosphere models [66, 67, 68, 22].

The high-dimensionality of such problems necessitates the development of reduced imperfect models; these are given by a known dynamical system

$$\dot{\boldsymbol{v}}_{\mathrm{M}} = \boldsymbol{f}^{\mathrm{M}}(\boldsymbol{v}_{\mathrm{M}}, t) + \sigma^{\mathrm{M}}(\boldsymbol{v}_{\mathrm{M}})\dot{W}(t), \tag{2}$$

which has a similar structure to (1) but its phase space, $I\!\!R^M$, is typically completely different from that of the perfect system with $M \ll P$; the perfect system and its model share, however, the coarse-grained variables $\boldsymbol{u} \in I\!\!R^N$ where $N \leqslant P - M$. Throughout the following analysis we are interested in characterizing the statistical departures of the imperfect model dynamics relative to the perfect model on the subspace of the coarse-grained, resolved variables $\boldsymbol{u}$.

The evolution of the probability densities $p(\boldsymbol{v})$ and $p(\boldsymbol{v}_{\mathrm{M}})$ associated with the perfect and imperfect models satisfy appropriate Fokker-Planck equations (FPE) which we omit here for brevity. However, it is important to note that solving FPE in geophysical or engineering applications involving turbulent flows is unrealistic, despite the linearity of the problem, due to a massive number of relevant degrees of freedom in the dynamics. Therefore, a systematic procedure for constructing and validating reduced models on the subspace of relevant coarse-grained variables is of great importance in many contemporary AOS and engineering applications.

The assessment of the predictive performance of an imperfect stochastic model requires appropriate metrics for determining how well the marginal density $\pi(\boldsymbol{u})$ of the perfect system on the common coarse-grained variables $\boldsymbol{u}$ is approximated by the imperfect model marginal density $\pi^{\mathrm{M}}(\boldsymbol{u})$. The natural way [62, 87] to measure the lack of information in one probability density, here the marginal $\pi^{\mathrm{M}}(\boldsymbol{u})$, compared with the other, here $\pi(\boldsymbol{u})$, is through the relative entropy, $\mathcal{P}(\pi, \pi^{\mathrm{M}})$, given by

$$\mathcal{P}(\pi, \pi^{\mathrm{M}}) = \int \pi \ln \frac{\pi}{\pi^{\mathrm{M}}}, \tag{3}$$

with the integration assumed over the whole subspace of the common coarse grained variables $\boldsymbol{u} \in I\!\!R^M$. Despite the lack of symmetry in its arguments, the relative entropy $\mathcal{P}$ provides an attractive framework for assessing model error in high-dimensional turbulent systems, and in particular in AOS applications [56, 84, 17, 19, 2, 79, 12, 117, 37, 38, 80, 81], due to its two 'distance-like' properties:

- (i) $\mathcal{P}(\pi, \pi^{\mathrm{M}})$ is always positive unless $\pi = \pi^{\mathrm{M}}$ [62, 87],
- (ii) it is invariant under any invertible change of variables [84, 87].

Thus, the relative entropy (3) provides the following useful diagnostic definitions in the context of uncertainty quantification in high-dimensional turbulent systems:

Definition 2.1. (**Model error**) characterizes the lack of information in the marginal probability density of the imperfect model, π^M, relative to the true marginal density, π, on the coarse-grained variables $\boldsymbol{u}$ [79]. For given Gaussian statistical initial conditions with mean and variance $\boldsymbol{u}_0, R_0$ so that $\pi_{t_0} \equiv \pi(\boldsymbol{u}, t|\bar{\boldsymbol{u}}_0, R_0)$, $\pi_{t_0}^{\mathrm{M}} \equiv \pi^{\mathrm{M}}(\boldsymbol{u}, t|\bar{\boldsymbol{u}}_0, R_0)$, the model error is

$$\mathcal{E}(t \,|\, \bar{\boldsymbol{u}}_0, R_0) = \mathcal{P}(\pi_{t_0}, \pi_{t_0}^{\mathrm{M}}). \tag{4}$$

Definition 2.2. (**Internal prediction skill**) of a perfect/imperfect model quantifies the role of initial conditions in the forecast of a future state of a system [56, 37, 38] and it represents the gain of information beyond the perfect/imperfect climate (i.e.,

the probability density on the attractor). The prediction skill for the perfect model and the internal prediction skill for the imperfect model are

$$\mathcal{S}_K\big(t\,|\,\bar{\boldsymbol{u}}_0, R_0\big) = \mathcal{P}(\pi_{t_0}, \pi_{att}), \qquad \mathcal{S}_K^M\big(t\,|\,\bar{\boldsymbol{u}}_0, R_0\big) = \mathcal{P}(\pi_{t_0}^M, \pi_{att}^M), \tag{5}$$

where $\pi_{t_0} \equiv \pi(\boldsymbol{u}, t | \bar{\boldsymbol{u}}_0, R_0)$ is the marginal density on the coarse-grained variables $\boldsymbol{u}$ conditioned on the initial data and $\pi_{att}(\boldsymbol{u}, t) \equiv \lim_{t_0 \to -\infty} \pi_{t_0}$ is the climatological marginal density of the perfect model; analogous definitions hold for the imperfect model.

Definition 2.3. (**Model sensitivity** [79]) quantifies the gain of information in the marginal probability density of the perfect/imperfect system on the coarse-grained variables $\boldsymbol{u}$ in response to external perturbations from its climate (the probability density on the attractor); it is expressed via the relative entropy as

$$\mathcal{S}_E\big(t\,|\,\bar{\boldsymbol{u}}_0, R_0\big) = \mathcal{P}(\pi_\delta, \pi_{att}), \qquad \mathcal{S}_E^M\big(t\,|\,\bar{\boldsymbol{u}}_0, R_0\big) = \mathcal{P}(\pi_\delta^M, \pi_{att}^M) \tag{6}$$

where the climatological marginal densities π_{att}, π_{att}^M are defined as in Definition 2.2 and π_δ, π_δ^M are the perfect/imperfect model marginal densities perturbed from the original attractor and conditioned on the particular perturbation.

Note that most forced dissipative high-dimensional turbulent dynamical systems have marginal densities for coarse-grained variables on the attractor that are smooth despite the dissipative fractal pieces of the attractor so Definitions 2.1-2.3 make sense (see [75] for many examples).

The most practical setup for utilizing the framework of empirical information theory in AOS applications arises when both the measurements of the perfect system and its imperfect model involve only the mean and covariance of the resolved variables $\boldsymbol{u}$ so that π_L is Gaussian with climate mean $\bar{\boldsymbol{u}}$ and covariance R, whereas π^M is Gaussian with model mean $\bar{\boldsymbol{u}}^M$ and covariance R^M. In this case, $\mathcal{P}(\pi_L, \pi^M)$ has an explicit formula [56, 87] which is extensively used throughout this study

$$\mathcal{P}(\pi, \pi^M) = \left[\tfrac{1}{2}(\bar{\boldsymbol{u}} - \bar{\boldsymbol{u}}^M)R_M^{-1}(\bar{\boldsymbol{u}} - \bar{\boldsymbol{u}}^M)\right] + \tfrac{1}{2}\left[\mathrm{tr}[RR_M^{-1}] - N - \ln\det[R\,R_M^{-1}]\right]. \tag{7}$$

The first term in brackets in (7) is the signal, reflecting the model error in the mean but weighted by the inverse of the model variance, R_M, whereas the second term in brackets, the dispersion, involves only the model error covariance ratio, $R\,R_M^{-1}$. The signal and dispersion terms in (7) are individually invariant under any (linear) change of variables which maps Gaussian distributions to Gaussians; this property is very important for unbiased model calibration.

2.1. **Systematically improving imperfect models through empirical information theory.** As argued earlier in [80, 81, 73], the framework of empirical information theory provides a convenient and unambiguous way for characterizing the model error and for improving the performance of imperfect models, as well as for assessment of the model forced response via appropriate tests in the unperturbed statistical equilibrium and appropriate fluctuation dissipation theorems [81, 11].

Consider first a class of imperfect models, $\mathcal{M}$; the best imperfect model on the coarse-grained variables $\boldsymbol{u}$ is characterized by the marginal density π^{M_*}, $M_* \in \mathcal{M}$, so that the perfect model with the marginal density π has the smallest additional information beyond the imperfect model density [79, 80, 81, 73], i.e.,

$$\mathcal{P}(\pi, \pi^{M_*}) = \min_{M \in \mathcal{M}} \mathcal{P}(\pi, \pi^M). \tag{8}$$

An important issue to contend with in any realistic scenario is the fact that the perfect model density, π, in (8) is not known and only its best unbiased estimate, π_L, based on L measurements $\bar{\boldsymbol{E}}_L$ of the perfect system during the training phase is available. The following general principle [75, 79] facilitates the practical calculation of (8)

$$\mathcal{P}(\pi, \pi_{L'}^{\mathrm{M}}) = \mathcal{P}(\pi, \pi_L) + \mathcal{P}(\pi_L, \pi_{L'}^{\mathrm{M}}), \quad L' \leqslant L, \tag{9}$$

where L' denotes the number of measurements of the perfect system. The first term $\mathcal{P}(\pi, \pi_L)$ in (9) represents the lack of information in the least biased density based on the L measurements π_L relative to the true density π and it cannot be reduced unless more measurements are incorporated. As shown in [75], the information barrier in (9) is given by

$$\mathcal{P}(\pi, \pi_L) = \mathcal{S}(\pi_L) - \mathcal{S}(\pi), \tag{10}$$

where

$$\mathcal{S}(\tilde{\pi}) = -\int \tilde{\pi} \ln \tilde{\pi}. \tag{11}$$

is the entropy or uncertainty in the density $\tilde{\pi}$; we will return to this important concept in §2.2.

Note also that the second term, $\mathcal{P}(\pi_L, \pi_{L'}^{\mathrm{M}})$, in (9) exactly gives the lack of information in the "coarse-grained" model probability density associated with fewer constraints, $\pi_{L'}^{\mathrm{M}}$ relative to the least biased L-measurement estimate of the true density. Consequently, the optimization principle (8) can be computed by replacing the unknown π by the hypothetically known π_L so that the optimal model within the considered class of imperfect models satisfies

$$\mathcal{P}(\pi_L, \pi_{L'}^{\mathrm{M}*}) = \min_{\mathrm{M}\in\mathcal{M}} \mathcal{P}(\pi_L, \pi_{L'}^{\mathrm{M}}), \qquad L' \leqslant L. \tag{12}$$

Definition 2.4. (***Statistical equilibrium or Climate fidelity***) [79, 80, 20] of an imperfect model consistent with the L measurements of the coarse-grained variables $\boldsymbol{u}$ arises when

$$\mathcal{P}(\pi_L, \pi_{L'}^{\mathrm{M}*}) \ll 1, \tag{13}$$

with perfect climate fidelity for $\mathcal{P}(\pi_L, \pi_{L'}^{\mathrm{M}*}) \equiv 0$.

Remarks: (i) in an effort to minimize unnecessary technicalities in the following exposition, we frequently abuse the terminology and do not distinguish between statistical equilibrium in the autonomous case and time-dependent statistics on the attractor in the nonautonous case; see, e.g., [4, 87], for a detailed discussion of these issues, (ii) for notational simplicity we skip the subscripts L and L' in (8) and (12) in the following discussion with the tacit assumption that the number of the coarse-grained measurements of the imperfect model does not exceed the number of analogous measurements of the perfect system (i.e., $L' \leqslant L$).

It turns out, as illustrated in the subsequent sections, that even when the time-averaged climate fidelity is achieved, i.e.,

$$\overline{\mathcal{P}(\pi, \pi^{\mathrm{M}*})} \equiv \frac{1}{T} \int_0^T \mathcal{P}(\pi, \pi^{\mathrm{M}*}) \ll 1, \tag{14}$$

with T the period on the attractor of the perfect system, a significant improvement in the model prediction skill and sensitivity can be achieved. Climate fidelity of imperfect models is a necessary but not a sufficient condition for good model sensitivity (see [80, 81] for details), as illustrated in the following sections based on a suite

of increasingly complex models. Below, we provide a simple example of this important link which is described in more detail, exploiting the fluctuation-dissipation theorem in [81, 11].

2.2. **Information barriers.** The notion of an information barrier will prove to be an important concept in the following discussion and it requires special attention. Intuitively, information barriers arise in classes of imperfect models which fundamentally limit the skill of any model in the class to reproduce the correct statistical behavior from nature. As introduced earlier in (9) and (10) the first instance of information barrier was associated with the intrinsic lack of information $\mathcal{P}(\pi, \pi_L)$ in the least biased density π_L based on L measurements of the true density π. The intrinsic lack of information (10) cannot be reduced within the class of the least biased L-measurement densities and, consequently, only a part of the total lack of information is available for optimization in this configuration through (12). The information barrier (10) can only be overcome by expanding the framework to incorporate more measurements of the true density. Note also that additional information barriers can exist in the given L-measurement framework if one considers a subclass of the coarse grained model densities, say $\pi_{L'}^{\mathrm{M}\alpha} \subset \pi_{L'}^{\mathrm{M}}$, $L' \leqslant L$, which does not contain M* minimizing (12); this case will be illustrated in §4 and §6.

Another situation in which the information barriers arise naturally is associated with the sensitivity of imperfect models with climate fidelity to internal or external perturbations. As shown in [81], the model error between the true perturbed density and the model density can be expressed in such a case as

$$\mathcal{P}(\pi_\delta, \pi_\delta^{\mathrm{M}}) = \mathcal{P}(\pi_\delta, \pi_{L,\delta}) + \frac{1}{2} \int \pi_L^{-1} (\delta\pi_L - \delta\pi^{\mathrm{M}})^2 + \mathcal{O}(\delta^3), \tag{15}$$

where $\delta\pi_L, \delta\pi^{\mathrm{M}}$ denote the perturbations of the respective densities and $\mathcal{P}(\pi_\delta, \pi_{L,\delta})$ represents the information barrier which cannot be overcome unless more measurements of the perfect systems are taken into consideration. More details regarding the last formula (15) are given in the next section based on a simple example in the Gaussian framework.

2.3. **The link between statistical equilibrium fidelity and sensitivity of imperfect models.** The crucial issue in climate change science as well as many other areas involving high-dimensional turbulent systems regards estimating the accuracy with which the statistical evolution of an imperfect model predicts the perturbed coarse-grained statistics of the perfect model. The framework developed in [80, 81] directly links the uncertainty of the imperfect model forecast in response to external perturbations with its fidelity to the unperturbed climate; this link can be further exploited to asses the sensitivity of imperfect models to forcing perturbations based on the fluctuation dissipation theorem and appropriate tests carried out in the training phase [81, 32, 11] (see also remarks in §8).

In order to understand this link, assume that the perfect system or the imperfect model or both are perturbed so that $\pi_\delta(\boldsymbol{u}, t)$ the unknown perfect distribution, $\pi_{\delta,L}(\boldsymbol{u}, t)$, its least-biased density based on L measurements, and $\pi_\delta^{\mathrm{M}}(\boldsymbol{u}, t)$ the model distribution all vary smoothly with the parameter δ, i.e.,

$$\pi_\delta(\boldsymbol{u}) = \pi(\boldsymbol{u}) + \delta\pi(\boldsymbol{u}), \qquad \int \delta\pi(\boldsymbol{u})\mathrm{d}\boldsymbol{u} = 0, \tag{16}$$

with analogous expressions for the perturbed imperfect marginal density π_δ^{M} in terms of π^{M} and $\delta\pi^{\mathrm{M}}$; the explicit time dependence in (16) was skipped for clarity. For

stochastic dynamical systems rigorous theorems guarantee this smooth dependence under minimal hypothesis [41]. It is instructive to consider here an elementary but informative example assuming the Gaussian framework based on the measurements of the means and covariances which are assumed to be strictly diagonal for simplicity. Then, assuming climate fidelity of the imperfect models, the leading order Taylor expansion in the (small) parameter δ of the error $\mathcal{P}(\pi_\delta, \pi_\delta^\mathrm{M})$ of the imperfect predictions (see Definition 2.1) leads to

$$
\begin{aligned}
\mathcal{P}(\pi_\delta, \pi_\delta^\mathrm{M}) = {} & \mathcal{S}(\pi_{\delta,G}) - \mathcal{S}(\pi_\delta) \\
& + \frac{1}{2} \sum_{|k| \le N} \left(\delta \bar{u}_k - \delta \bar{u}_k^\mathrm{M} \right) R_k^{-1} \left(\delta \bar{u}_k - \delta \bar{u}_k^\mathrm{M} \right) \\
& + \frac{1}{4} \sum_{|k| \le N} R_k^{-2} \left(\delta R_k - \delta R_k^\mathrm{M} \right)^2 + O(\delta^3).
\end{aligned}
\tag{17}
$$

The entropy difference, $\mathcal{S}(\pi_{\delta,G}) - \mathcal{S}(\pi_\delta)$, in (17) corresponds to the intrinsic error due to measuring only the mean and covariance of the perfect system. This entropy difference cannot be reduced within the class of Gaussian models and represents an *information barrier* to improving the imperfect model sensitivity; as already signaled in §2.2, we will return to this important concept repeatedly throughout the paper. The first (second) summation in (17) is the signal (dispersion) contribution to the model error (see (7)). For more general results, including the discussion of the link between climate fidelity and sensitivity of imperfect models via the fluctuation dissipation theorem, see [80, 32].

Given a class of imperfect models $\mathcal{M}$, the error (17) in predicting the true system's response to external perturbations is minimized for the model which is the most consistent with the unperturbed climate, i.e., the model $\mathrm{M}_* \in \mathcal{M}$ satisfying (8) or (12). In fact, climate consistency of an imperfect model is a necessary but not a sufficient condition for its predictive skill, as shown in [80], where simple yet instructive examples reveal the possibility of intrinsic barriers to improving model sensitivity even with perfect climate fidelity. The only way to overcome such barriers is by extending the class of imperfect models to account for more degrees of freedom. In the following sections, we will present further examples of such situations in more complex non-Gaussian models.

3. UQ and forecasting with model error in linear systems with hidden dynamics.
Here, we introduce a simple two-dimensional linear stochastic model which is very useful for elucidating many important effects of model error in short and medium range imperfect predictions on a common subset of resolved variables. This is the simplest exactly solvable model which allows for studying these features. An important source of model error in dynamical forecasts of complex high-dimensional turbulent systems using imperfect models on the coarse-grained subset of resolved variables is introduced due to neglecting, or misrepresenting, the interactions with unresolved processes which are effectively hidden from the coarse-grained imperfect models.

Here, we utilize the simplest possible two-dimensional linear framework and assume that one of the two dynamical variables is observed/resolved, and the second variable is hidden from observations while nontrivially affecting the resolved dynamics. We will also study imperfect models with correctly prescribed statistical

properties of the marginal equilibrium dynamics for the observed variable and quantify their short and intermediate range predictive skill. In particular, we focus on the role of initial conditions in short range imperfect forecasts, and the subtle interplay between the initial data and the asymptotic system characteristics in the medium range forecasts for the resolved dynamics.

Important issues illustrated in this section are:
- There exist many imperfect models with climate fidelity on the resolved variables which have very different short and medium range prediction skill;
- Models with high skill for short range predictions are not necessarily those with high skill for intermediate or long range predictions; moreover, models with unphysical negative damping in the resolved dynamics can outperform models with positive damping for short range predictions;
- In the linear Gaussian framework tuning imperfect models to reproduce the marginal climatological distribution at all time lags guarantees the correct response to any time dependent external forcing. Models with unphysical parameter values and inferior initial data response may have high skill in the forced response, exceeding that of imperfect models which are tuned to reproduce equilibrium statistics in the marginal climate;
- There exist explicit information barriers for model improvement which require extending the class of imperfect models in order to achieve further improvement;
- There exist important limitations for estimating the prediction skill based on the the superensemble framework and mutual information.

As will be shown below and in the following sections, this simple linear Gaussian framework with hidden/unresolved dynamics provides a very revealing testbed for analyzing and understanding many important issues associated with short, and medium range predictions of much more complex nonlinear natural systems where the partial observability of the true system and the subtle interplay between the initial conditions and model error are much more difficult to capture. Problems of this nature arise even in the relatively simple framework of multilevel linear regression models which attempt to tune imperfect model parameters based on the marginal dynamics of the resolved variables (see [89] for an extensive treatment); some remarks regarding this issue are presented in §8.

3.1. **The linear system for the perfect model.** Consider a situation when the resolved/observed dynamics $x(t)$ is given by the marginal dynamics of the following 2×2 linear stochastic system which serves as the perfect model

$$\mathrm{d}\begin{pmatrix} x \\ y \end{pmatrix} = \left[\hat{L} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} F(t) \\ 0 \end{pmatrix} \right] \mathrm{d}t + \begin{pmatrix} 0 \\ \sigma \end{pmatrix} \mathrm{d}W(t). \tag{18}$$

The system matrix $\hat{L}$ and its eigenvalues $\lambda_{1,2}$ are, respectively,

$$\hat{L} = \begin{pmatrix} a & 1 \\ q & A \end{pmatrix}, \qquad \lambda_{1,2} = \frac{1}{2}\left(a + A \pm \sqrt{(a-A)^2 + 4q} \right). \tag{19}$$

There are four coefficients (a, q, A, σ) in (18) where a is the damping in the resolved dynamics $x(t)$, and q is the damping in the unresolved dynamics $y(t)$ with A the coupling parameter in the unresolved dynamics. Clearly, in the above formulation the resolved dynamics $x(t)$ is affected by the hidden process $y(t)$. We assume that

the deterministic forcing $F(t)$ acts only in the resolved subspace x and the stochastic noise generated by the scalar Wiener process $W(t)$, $t \in \mathbb{R}$, $W(0) = 0$ is present only in the unresolved subspace with the variable y. Since the system (18) is linear with additive noise, it can be easily shown that for constant forcing F it has a Gaussian invariant measure provided that

$$a + A = \lambda_1 + \lambda_2 < 0, \qquad aA - q = \lambda_1 \lambda_2 > 0. \tag{20}$$

The above stability conditions guarantee that the eigenvalues of $\hat{L}$ have negative real parts so that the equilibrium mean of (18) is given by

$$\bar{x}_{att} = -\frac{AF}{aA - q}, \qquad \bar{y}_{att} = \frac{qF}{aA - q}. \tag{21}$$

The autocovariance at equilibrium, $\mathcal{C}_{att}(\tau) = \langle \boldsymbol{x}(t+\tau) \otimes \boldsymbol{x}^T(t) \rangle_{att}$, depends only on the lag τ regardless of the nature of the forcing and it is given by

$$\mathcal{C}_{att}(\tau) = \lim_{t_0 \to -\infty} \mathcal{C}(\tau, t, t_0) = \Sigma\, e^{\hat{L}^T \tau}, \tag{22}$$

where the covariance $\Sigma \equiv \mathcal{C}_{att}(\tau = 0, t)$ on the attractor of (18) is given by

$$\Sigma \equiv = \begin{pmatrix} 1 & -a \\ -a & aA - q + a^2 \end{pmatrix} \frac{\sigma^2}{2(a + A)(q - aA)}. \tag{23}$$

Extensions to the nonautonomous case with time-periodic forcing are trivially accomplished provided that the stability conditions (20) are satisfied so that there exists a time-periodic Gaussian measure on the attractor (defined in the nonautonomous sense; see, e.g., [87, 4]) with the time-periodic mean, $\bar{\boldsymbol{x}}_{att}(t) \equiv \lim_{t_0 \to -\infty} \bar{\boldsymbol{x}}(t, t_0)$, and autocovariance (22).

It is important to realize that even for the class of simple linear systems (18) the dynamics during the approach to the same attractor can be quite different depending on the coefficients $\{a, q, A, \sigma\}$. There exist three distinct regimes of stable dynamics of the linear Gaussian system (18) with different properties of the statistical evolution of the 2×2 system (18):

PURELY REAL EIGENVALUES $\lambda_{1,2}$:
- *Normal mean dynamics:* In this regime the two eigenvectors of the system matrix $\hat{L}$ with purely real eigenvalues $\lambda_{1,2}$ are nearly orthogonal. Provided that $\Lambda \equiv \lambda_1/\lambda_2 \neq 1$, this configuration occurs for

$$a \approx \frac{1}{2}\left((\Lambda + 1)\lambda_1 \pm \sqrt{(\Lambda + 1)^2 \lambda_1^2 - 4(1 + \Lambda\lambda_1^2)} \right).$$

- *Non-normal mean dynamics:* For large eigenvalue ratio $\Lambda \gg 1$ and weak damping, $a \ll \Lambda$, in the resolved dynamics the angle between the eigenvectors is sensitive to variations of a but only one of the eigenvectors significantly changes its orientation. For $\Lambda \sim 1$ corresponding to $(a - A)^2 + 4q \approx 0$, the eigenvectors are nearly collinear and the direction of both eigenvectors depends strongly on the damping a in the resolved dynamics.

COMPLEX-CONJUGATE EIGENVALUES $\lambda_{1,2}$:
- *Mean dynamics with complex-conjugate eigenvalues.* This configuration occurs for $(a - A)^2 + 4q < 0$ when the complex conjugate eigenvalues satisfy the stability condition (20).

We will show in the next section that the marginal statistical properties for the resolved variable $x(t)$ on the attractor can be reproduced by many different imperfect models. The crucial observation is that some of these models may be characterized by a normal mean dynamics while others will display a high sensitivity to model parameters and the statistical initial conditions due to non-normal character of their dynamics.

3.2. **Imperfect models.** One obvious source of model error in imperfect predictions of natural high-dimensional systems stems from the fact the imperfect models can be only tuned to the true dynamics on a subset of resolved, coarse-grained variables on the attractor of the true system. Here, we introduce two classes of imperfect models of (18) which we use to illustrate the following important facts:

- There exist families of imperfect models of the linear system (18) with the same statistical characteristics of the resolved dynamics on the attractor (see §3.2.1).
- Different imperfect models with the same attractor/climate fidelity (13) on the subspace of resolved variables can have significantly different skill for short and medium range prediction as well as the forced response (see §3.4).

The imperfect models described below will be subsequently used in combination with the stochastic-statistical framework of §2 to illustrate a number of important issues arising in imperfect predictions of vastly more complex nonlinear natural systems, such as the interplay between forced response and initial data response and limitations of the superensemble framework. The issues related to predictive performance of these imperfect models and their forced response sensitivity are discussed in the subsequent sections.

2×2 LINEAR MODELS WITH MODEL ERROR

The imperfect models in this class have the same structure as the perfect system (18) but with different coefficients $(a_\mathrm{M}, A_\mathrm{M}, q_\mathrm{M}, \sigma_\mathrm{M})$ and forcing $F_\mathrm{M}(t)$. The pathwise solutions and statistics of the imperfect 2×2 models have the same structure as those for the perfect system (18), as shown in Appendix A.1.

Here, the model error is introduced by the incorrect coefficients $(a_\mathrm{M}, A_\mathrm{M}, q_\mathrm{M}, \sigma_\mathrm{M})$ which are tuned so that the marginal statistics of the resolved dynamics, $x^\mathrm{M}(t)$, in the model and the truth, $x(t)$, coincide on the attractor (see §3.2.1).

MEAN STOCHASTIC MODEL (MSM) FOR THE RESOLVED DYNAMICS

These models are the simplest examples representing under-resolution of a perfect model through a lower dimensional dynamical system. Here, these imperfect models represent the simplest possible reduced models of the perfect system (18) where the interactions between the unresolved dynamics $y(t)$ and the resolved dynamics $x(t)$ are represented by the white noise. Consequently, the resolved dynamics $x(t)$ in (18) is modelled by the scalar Mean Stochastic Model (MSm) given by the linear SDE (e.g., [28])

$$\mathrm{d}\, x^{\mathrm{MSM}}(t) = \big(-\gamma_{\mathrm{MSM}}\, x^{\mathrm{MSM}}(t) + F_{\mathrm{MSM}}(t) \big)\mathrm{d}t + \sigma_{\mathrm{MSM}}\mathrm{d}W(t), \tag{24}$$

where $\gamma_{\mathrm{MSM}} > 0$ is the damping, F_{MSM} denotes the deterministic forcing, and σ_{MSM} is the amplitude of the white noise forcing defined on the whole real line. The pathwise solutions and statistics of the Gaussian process satisfying (24) are the standard

solutions for the Ornstein-Uhlenbeck process and are given in Appendix A.2. The dynamics of x^{MSM} is stable for $\gamma_{\text{MSM}} > 0$ so that the statistics on the attractor are

$$\bar{x}_{att}^{\text{MSM}}(t) = \lim_{t_0 \to -\infty} \int_{t_0}^{t} e^{-\gamma_{\text{MSM}}(t-s)} F_{\text{MSM}}(s) \mathrm{d}s, \quad \mathcal{C}_{att}^{\text{MSM}}(\tau) = \Sigma^{\text{MSM}} e^{-\gamma_{\text{MSM}}\tau}, \quad \Sigma^{\text{MSM}} = \frac{\sigma_{\text{MSM}}^2}{2\gamma_{\text{MSM}}}. \tag{25}$$

The model error in MSm arises due to the dimensional reduction and representing the interactions with the unresolved processes through white noise forcing. It is important to note that MSm cannot reproduce the marginal autocovariance, $(\mathcal{C}_{att})_{11}(\tau)$, of the perfect system (18) at all lags and it cannot be tuned to reproduce the marginal equilibrium statistics; this affects the skill for predicting the forced response (see also Proposition 3 and §3.3).

Finally, we mention that another class of imperfect models of (18) can be obtained via a multilevel linear least-squares regression of the model coefficients $(a_{\text{M}}, q_{\text{M}}, A_{\text{M}}, \sigma_{\text{M}})$ and the forcing F_{M} from long time series of the marginal dynamics $x(t)$. The limitations of the multilevel approach for constructing two-dimensional models and the advantages of physics-constrained scalar models were discussed at length in [89]; some remarks regarding this approach are relegated to §8.

3.2.1. *Families of linear models with the same marginal statistics on the attractor.* Here, we show that there exist families of imperfect linear models of the system (18) with perfect climate fidelity (cf. Definition 2.4). Moreover, there also exist families of imperfect models with perfect climate fidelity and correct marginal two-point statistics at all time lags. This non-uniqueness will be reflected in different short and medium range forecast skill of imperfect models with climate fidelity and different forced response properties. The discussion below builds on the results reported in [89] in the unforced case; more details are given in Appendix A.3 which outlines straightforward extensions to the time-periodic configuration.

The two definitions below distinguish between two degrees of statistical consistency on the attractor of a stochastic system:

Definition 3.1 (***Marginal two-point equilibrium statistics***). Consider the solution $(x(t,t_0), y(t,t_0))$ of the linear stochastic system (18) satisfying (20) and with a Gaussian invariant measure at equilibrium with mean (21), and autocovariance (22). Then, the marginal mean and autocovariance

$$\bar{x}_{att}, \quad \left(\mathcal{C}_{att}(\tau)\right)_{11}, \quad \tau \geqslant 0, \tag{26}$$

is called the marginal two-point statistics on the attractor for the process $x(t)$.

Definition 3.2 (***Marginal climatology*** *(marginal one-point equilibrium statistics)*). Consider the solution $(x(t,t_0), y(t,t_0))$ of the system (18) with Gaussian invariant measure at equilibrium with mean (21), and autocovariance (22). The marginal equilibrium mean and variance

$$\bar{x}_{att}, \quad \Sigma_{11} = \left(C_{att}(\tau = 0)\right)_{11}, \tag{27}$$

are called the marginal climatology for the Gaussian process $x(t)$.

Remarks: (i) Based on definitions (3.1) and (3.2) it is clear that linear systems with the same marginal two-point statistics on the attractor are contained within the family of systems with the same marginal climatology. (ii) The above definitions

can be trivially extended to the nonautonomous case with a time-periodic attractor [87, 4] provided that the stability conditions conditions (20) are satisfied.

It can be easily verified with the help of straightforward algebra and Definitions 3.1-3.2 that there exist families of imperfect models of (18) with same equilibrium marginal statistics, or the same climatology as that of the resolved variable $x(t)$ satisfying the perfect system (18); we summarize these results below:

Proposition 1. *[Family of imperfect 2×2 linear models with the same marginal two-point equilibrium statistics for $x(t)$]: Consider the linear 2×2 system (18) satisfying the stability conditions (20) with constant forcing and coefficients (a, q, A, σ, F). Then, the marginal two-point equilibrium statistics of (18) is controlled by three parameters*

$$\left\{ \lambda_{1,2}, \, \sigma, \, AF \right\}. \tag{28}$$

Since there are five independent coefficients in the (18) and four constraints (28), there exists a one-parameter family of 2×2 linear models (18) and coefficients

$$\left(a_{\mathrm{M}}(w), q_{\mathrm{M}}(w), A_{\mathrm{M}}(w), \sigma_{\mathrm{M}}(w), F_{\mathrm{M}}(w) \right), \, w \in I\!\!R,$$

with the same marginal two-point equilibrium statistics for $x(t)$.

Proposition 2 (*Family of imperfect 2×2 linear models with the same marginal climatology for $x(t)$).* *Consider the linear system (18) satisfying the stability conditions (20) with constant forcing and coefficients (a, q, A, σ, F). Then, the marginal climatology for $x(t)$ is determined by*

$$\left\{ \frac{\sigma^2}{(a+A)(aA-q)}, \, \frac{AF}{aA-q} \right\}. \tag{29}$$

Since there are five coefficients in the system (18) and two constraints (29), there exists a three-parameter family of linear models with structure (18) and coefficients

$$\left(a_{\mathrm{M}}(\boldsymbol{w}), q_{\mathrm{M}}(\boldsymbol{w}), A_{\mathrm{M}}(\boldsymbol{w}), \sigma_{\mathrm{M}}(\boldsymbol{w}), F_{\mathrm{M}}(\boldsymbol{w}) \right), \, \boldsymbol{w} \in I\!\!R^3,$$

with the same marginal climatology for $x(t)$.

Proposition 3 (*MSMs with correct marginal climatology*)*.* *Consider the Mean Stochastic Model (24) with constant forcing and coefficients $\left\{ \gamma_{\mathrm{MSM}}, \sigma_{\mathrm{MSM}}, F_{\mathrm{MSM}} \right\}$. Provided that $\gamma_{\mathrm{MSM}} > 0$, the climatology of (24) is controlled by two parameters*

$$\left\{ \frac{\sigma_{\mathrm{MSM}}^2}{2\gamma_{\mathrm{MSM}}}, \, \frac{F_{\mathrm{MSM}}}{\gamma_{\mathrm{MSM}}} \right\}. \tag{30}$$

The process $x_{\mathrm{MSM}}(t)$ satisfying (24) has the same climatology as $x(t)$ satisfying the perfect system (18) provided that

$$\sigma_{\mathrm{MSM}}^2 = -\gamma_{\mathrm{MSM}} \frac{\sigma^2}{(a+A)(aA-q)}, \quad F_{\mathrm{MSM}} = -\gamma_{\mathrm{MSM}} \frac{AF}{aA-q}, \tag{31}$$

where a, A, q, σ are the coefficients of the perfect system (18) with constant forcing F. Since there are three coefficients in the system (24) with constant forcing and two constraints (31), there is a one-parameter family of MSm's with coefficients

$$\left\{ \gamma_{\mathrm{MSM}}(w), \, \sigma_{\mathrm{MSM}}(w), \, F_{\mathrm{MSM}}(w) \right\}, \quad w \in I\!\!R,$$

with the same climatology as that of the perfect system (18). In general, MSm cannot reproduce the marginal two-point equilibrium statistics of the perfect system (18).

The proofs of the above propositions are elementary and rely on algebraic manipulations of the expressions for the marginal second-order statistics at equilibrium of the systems (18) and (24); we omit them here for brevity. Time-averaged extensions of the above propositions to the nonautonomous case with time-periodic statistics on the attractor can be easily implemented but they are unnecessary for the following discussion.

3.3. **Imperfect model optimization and information barriers.** In many practical situations, such as actual experiments in climate science, it is important to understand the response of the true system to changes in external forcing, δF, and to construct imperfect models whose response captures the features of the true system's response on the resolved coarse-grained variables. Here, following [80, 73], we discuss the first example of imperfect model optimization via the information-theoretic principles of §2 which shows unambiguously the existence of information barriers in the simplest yet revealing case; here, the true system is given by the linear Gaussian system (18) and the imperfect models have perfect unperturbed climate fidelity (13). It can be also easily shown that in this linear Gaussian framework only the imperfect models with correct marginal equilibrium statistics at all time lags (cf. Definition 3.1) have the correct forced response at all times.

Consider the imperfect models of (18) discussed in §3.2 with perfect climate fidelity (13) on the resolved subspace x. In the linear Gaussian framework this is easily achieved by matching the equilibrium means and variances, i.e,

$$\bar{x}_{att}^{\mathrm{M}} = \bar{x}_{att}^{\mathrm{MSM}} = \bar{x}_{att}, \quad \text{and} \quad E_{\mathrm{M}} = E_{\mathrm{MSM}} = E \equiv (\Sigma_{att})_{11}, \tag{32}$$

where $\bar{x}_{att}^{\mathrm{M}}$, $\bar{x}_{att}^{\mathrm{MSM}}$ are the marginal imperfect model means and E_{M}, E_{MSM} are the corresponding marginal variances

$$E = -\frac{\sigma^2}{2(a+A)(aA-q)}, \quad E_M = -\frac{\sigma_{\mathrm{M}}^2}{2(a_{\mathrm{M}}+A_{\mathrm{M}})(a_{\mathrm{M}}A_{\mathrm{M}}-q_{\mathrm{M}})}, \quad E_{\mathrm{MSM}} = \frac{\sigma_{\mathrm{MSM}}^2}{2\gamma_{\mathrm{MSM}}}. \tag{33}$$

The asymptotic, infinite-time response of the perfect system to forcing perturbations from the statistically steady state can be obtained by replacing F in (18) by $F + \delta F$ while the same experiment in the imperfect models for (18) involves replacing F_{M} by $F_{\mathrm{M}} + \delta F$. Since the considered models are linear and Gaussian, the only change in the expected response to forcing perturbations is through the change in mean

$$a) \quad \delta x = -\frac{A}{aA-q}\delta F, \qquad b) \quad \delta x^{\mathrm{M}} = -\frac{A_{\mathrm{M}}}{a_{\mathrm{M}}A_{\mathrm{M}}-q_{\mathrm{M}}}\delta F, \qquad c) \quad \delta x^{\mathrm{MSM}} = \frac{1}{\gamma_{\mathrm{MSM}}}\delta F, \tag{34}$$

while the variance of x, x^{M} and x^{MSM} for the perfect and imperfect models remains unchanged with the values E_{M} and E_{MSM} tuned to E as in (53).

Now, assume that the perfect linear system (18) with parameters (a, q, A, σ) and constant forcing F satisfies the stability conditions (20), with $A > 0$. As shown in [80, 73] in such a case no MSm (24) with $\gamma_{\mathrm{M}} > 0$ can match the sensitivity of the perfect system (18) since the perfect and imperfect model sensitivity are always anti-correlated; this is easy to see by noticing that for $A > 0$, $\mathrm{sign}(\delta x) = -\mathrm{sign}(\delta F)$

in (34a) but for all MSm models $\text{sign}(\delta x_{\text{MSM}}) = \text{sign}(\delta F)$ in (34c). The formula in (17) applies exactly to these models with perfect climate fidelity with

$$\mathcal{P}(\pi_\delta, \pi_\delta^{\text{MSM}}) = \frac{1}{2} E^{-1} \left| -\frac{A}{aA - q} - \frac{1}{\gamma_{\text{MSM}}} \right|^2 |\delta F|^2. \tag{35}$$

In this situation with $A > 0$ there is an intrinsic barrier to sensitivity improvement for the MSm models (24), since any attempt to minimize the information theoretic model error in the sensitivity through the general principle in (8) is futile because no finite minimum over γ_{MSM} of (35) is achieved and necessarily $\gamma_{\text{MSM}} \to \infty$ in the approach to this minimum value. This information barrier can only be overcome by enlarging the class of models beyond (24) by introducing more degrees of freedom in the model; in this simple framework expanding the class of imperfect models to include the 2×2 linear models allows for a trivial minimization of (17) to achieve the perfect forced response.

On the other hand, if the perfect system satisfies (20) with $A < 0$, then using (35) to optimize the sensitivity of MSm leads to

$$\gamma_{\text{MSM}}^* = -(aA - q)/A, \quad A < 0. \tag{36}$$

Note that while for $A < 0$ in (18) the MSm with perfect climate fidelity and γ_{MSM}^* also captures the infinite time forced response of the perfect model on the resolved variable, it does not correctly reproduce the forced response at finite times; this can be easily seen using the standard 'fluctuation-dissipation' formulas for the mean linear response of a stochastic system to external perturbations (e.g., [75, 87, 88, 11]) which for the response in the mean are given by

$$\delta \bar{x}(t) = \int_{t_0}^{t} \mathcal{R}_{\bar{x}}(t - s) \delta F(s) \mathrm{d}s, \qquad \mathcal{R}_{\bar{x}}(\tau) = \left\langle x(\tau)(x(0) - \bar{x}) \right\rangle_{att}, \tag{37}$$

where $\langle \cdot \rangle_{att}$ denotes the ensemble average at the unperturbed equilibrium; the response in the variance is identically zero for linear Gaussian systems. Thus, the correct forced response at all times in the Gaussian framework requires correct marginal autocorrelations at all lags τ which cannot be achieved for the MSm (see Proposition 3). On the other hand, the imperfect 2×2 models with correct two-point marginal equilibrium statistics have the correct forced response from Proposition 1.

3.4. **Prediction skill of imperfect models with perfect statistical equilibrium fidelity.** The simple linear framework of two-dimensional Gaussian models with hidden dynamics provides a useful testbed for analyzing and understanding many important issues associated with short, and medium range predictions of much more complex nonlinear natural systems where the partial observability of the true system and the subtle interplay between the initial conditions and model error are much more difficult to capture. Here, we use the linear Gaussian system (18) and its imperfect models to illustrate the following issues:

- Initial data response at short and intermediate ranges for the imperfect models with climate fidelity;
- Limitations of superensemble framework and mutual information for estimating prediction skill.

As shown in §2, the prediction skill and model error for imperfect models can be quantified using the information-theoretical metrics given in Definitions 2.1-2.3. Note first that for imperfect models with perfect statistical equilibrium fidelity

(see (13) or (14)) and statistical initial conditions $\bar{\boldsymbol{u}}_0 \equiv (\bar{x}_0, \bar{y}_0)$, R_0, the following relationships hold

$$\text{Asymptotic model error}: \quad \lim_{t \to \infty} \mathcal{P}(\pi_{t_0}(t|\bar{\boldsymbol{u}}_0, R_0), \pi_{t_0}^{\text{M}}(t|\bar{\boldsymbol{u}}_0, R_0)) = 0,$$

$$\text{Asymptotic internal prediction skill}: \quad \lim_{t \to \infty} \mathcal{P}(\pi_{t_0}(t|\bar{\boldsymbol{u}}_0, R_0), \pi_{att}^{\text{M}}) = 0.$$

Unsurprisingly, this implies that all imperfect models with perfect climate fidelity are equivalent for long range forecasts where the sensitive dependence on the initial conditions vanishes. However, as already discussed in §3.2.1 imperfect models with climate fidelity can have very different short and medium range prediction skill due to different dynamical behavior during the approach to equilibrium from a given initial condition. Here, the sensitivity of imperfect predictions for the resolved variable $x(t)$ to the unresolved initial conditions $\bar{y}_0$ is especially interesting since the uncertainties in these initial conditions are impossible to control in practice.

In order to mimic realistic forecast situations with uncertainties in initial conditions as well as the model error, we construct ensembles of perfect/imperfect forecasts with initial conditions normally distributed around the initial ensemble mean, $\bar{\boldsymbol{u}}_0 = (\bar{x}_0, \bar{y}_0)$, and with covariance R_0 of the initial conditions in the ensemble smaller (in a suitable norm) than the climate variance at t_0. We use the analytical formulas for the second-order statistics of the perfect system and the imperfect models (see Appendix A) and represent every ensemble member by its mean and covariance.

The examples discussed below are based on a suite of imperfect models introduced in §3.2 with either the correct marginal two-point statistics or correct marginal climatology (see Definitions 3.1-3.2 in §3.2.1). Given the perfect model (18) with coefficients (a, q, A, σ) and constant forcing F, we consider three imperfect models tuned to have the correct marginal two-point equilibrium statistics and two models with the correct marginal climatology (see Definitions 3.1, 3.2 and Proportions 1-3); the cumbersome formulas for the coefficients $a_{\text{M}}, A_{\text{M}}, q_{\text{M}}, \sigma_{\text{M}}, F_{\text{M}}$ incorporating the above constraints are relegated to Appendix A.3. The models in the suite are given by:

MODELS WITH CORRECT MARGINAL TWO-POINTS STATISTICS ON ATTRACTOR

(i) 2×2 model with overestimated damping in $x^{\text{M}}(t)$: $a_{\text{M}} = 1.5a$ and A_{M}, q_{M}, σ_{M}, F_{M} given by (104),

(ii) 2×2 model with underestimated damping in $x^{\text{M}}(t)$: $a_{\text{M}} = a/1.5$ and A_{M}, q_{M}, σ_{M}, F_{M} given by (104),

(iii) 2×2 model with negative damping in $x^{\text{M}}(t)$: $a_{\text{M}} = -0.1a$ and A_{M}, q_{M}, σ_{M}, F_{M} given by (104).

MODELS WITH CORRECT MARGINAL CLIMATOLOGY (single-point statistics on attractor)

(iv) 2×2 model with $a_{\text{M}} = 1.1a$, and $q_{\text{M}} = 0.7q$, $A_{\text{M}} = 0.65A$ obtained from (105),

(v) MSm with correct decorrelation time for the resolved variable: $\gamma_{\text{MSM}} = -(aA - q)/(a + A)$ and σ_{MSM}, F_{MSM} obtained from (106).

Note here that the model (iii) incorrectly assumes negative damping in the resolved dynamics while reproducing the correct two-point marginal equilibrium statistics and has, therefore, the correct response for all forcing perturbations. The last two models (iv) and (v) only have the correct marginal climatology (cf Definition 3.2) and they lack the correct forced response. Despite these obvious differences, the

choice of the best model depends largely on the forecast lead time, as discussed below.

Figures 1, 2, 3, 4 illustrate the model error (4) and internal prediction skill (5) of the probabilistic forecasts from the imperfect models (i)-(v) in six distinct configurations of the perfect model: (1) nearly normal dynamics with the fast direction perpendicular to the resolved subspace (figure 1), (2) nearly normal dynamics with the fast direction parallel to the resolved subspace (figure 2), (3) non-normal dynamics with non-degenerate eigenspaces (figure 3), (4) non-normal dynamics with nearly degenerate eigenspaces (figure 4), (5-6) dynamics with complex eigenvalues (figure 5 with small real parts, and figure 6, with comparable real and imaginary parts). Note that in of all these regimes the nature of the dynamics in the imperfect models (i)-(v) can differ drastically from that of the perfect system. High predictive skill of a given imperfect model is achieved for a small model error $\mathcal{E}$ and high potential predictive skill $\mathcal{S}$K. The main points regarding the predictive skill of the optimized imperfect models (i)-(v), illustrated in the figures 1, 2, 3, 4, are:

- The choice of the best imperfect model for initial data response depends on the nature of the true dynamics, the range of initial conditions, and the forecast range. Moreover, the model sensitivity to initial conditions for the unresolved dynamics is important.
- When the perfect system dynamics is nearly normal with the strongest attracting direction along the unresolved subspace y, imperfect models with (at least) correct marginal climatology (Definition 3.2) have good overall predictive skill (figure 1).
- The imperfect model (iv) tends to have the best overall skill in our test suite and good short range prediction skill in highly oscillatory systems (see figure 5); note that this model only reproduces the correct marginal climatology and it does not have the correct forced response (see §3.3).
- The predictive skill of MSm, representing the simplest reduced model, is generally poor and is highly dependent on the initial conditions and the perfect system geometry.
- Imperfect models with unphysical negative damping but correct marginal two-point statistics (and forced response) can outperform models with positive damping for short range forecasts (see model (iii) in the bottom panels of figures 1, 3, 6) or even medium ranges (model (iii) in bottom panels of figures 4, 5). This model usually has better skill than MSm for predictions with all lead times.
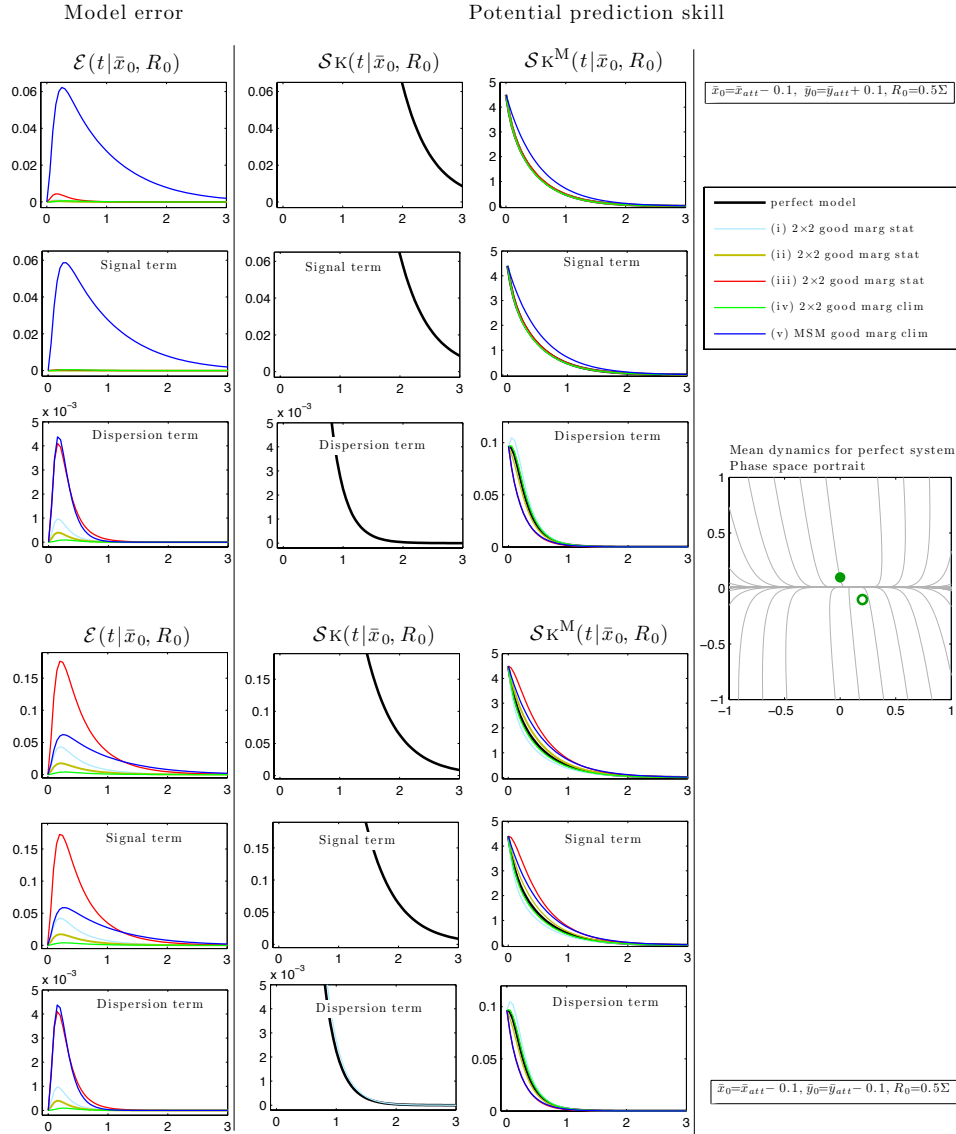
FIGURE 1. Model error (4) and potential prediction skill (5) for forecasts using the imperfect 2×2 linear systems (18) and the MSMs (24) with statistical equilibrium fidelity (13) (models (i)-(iii) also have the correct two-point equilibrium statistics; see Def. 3.1). The perfect system **dynamics is nearly normal** with parameters ($a = -1, \lambda_1 = -1, \lambda_2 = -10, \sigma = 0.5, F = 0.1$). The two examples are for different statistical initial conditions $\bar{x}_0, \bar{y}_0, R_0$.
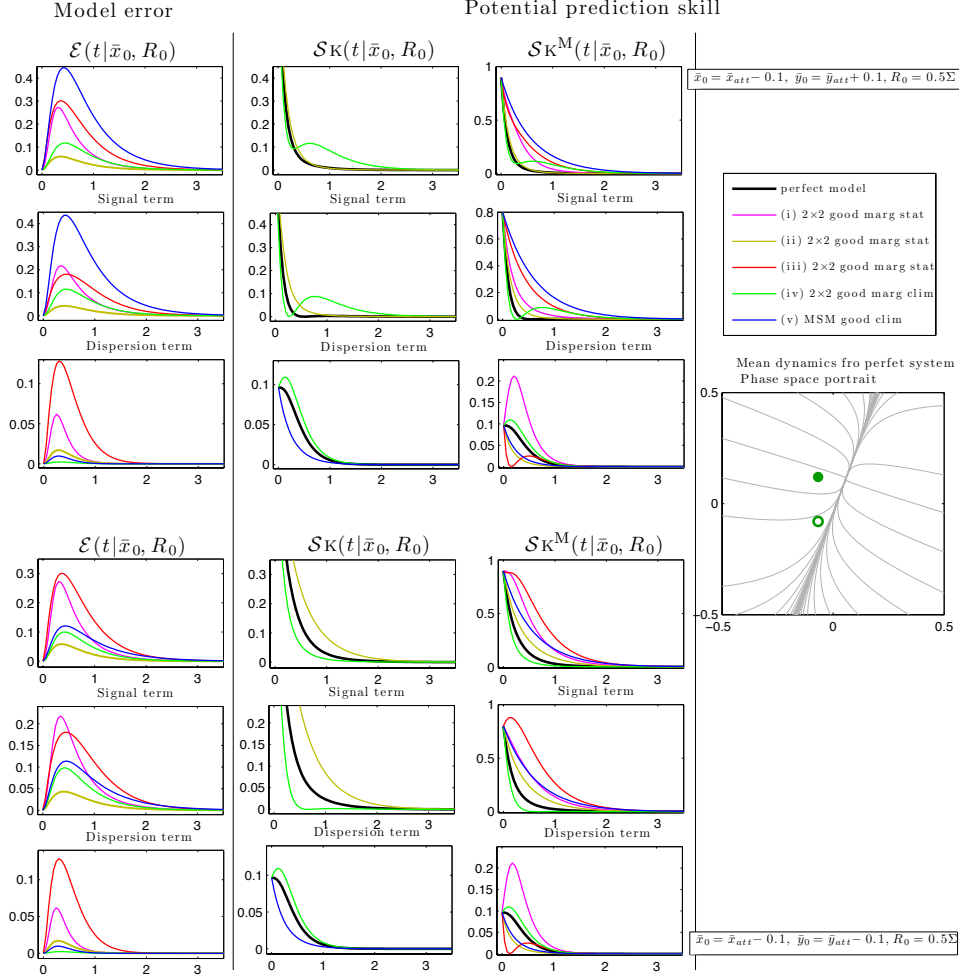
FIGURE 2. Model error (4) and potential prediction skill (5) for forecasts using the imperfect 2×2 linear systems (18) and the MSm (24) with statistical equilibrium fidelity (13) (models (i)-(iii) also have the correct two-point equilibrium statistics; see Def. 3.1). The perfect system **dynamics is nearly normal with strongest attracting direction nearly parallel to** x with coefficients ($a = -3.7, \lambda_1 = -1, \lambda_2 = -4, \sigma = 0.5, F = 0.1$). The two examples are for different statistical initial conditions $\bar{x}_0, \bar{y}_0, R_0$.
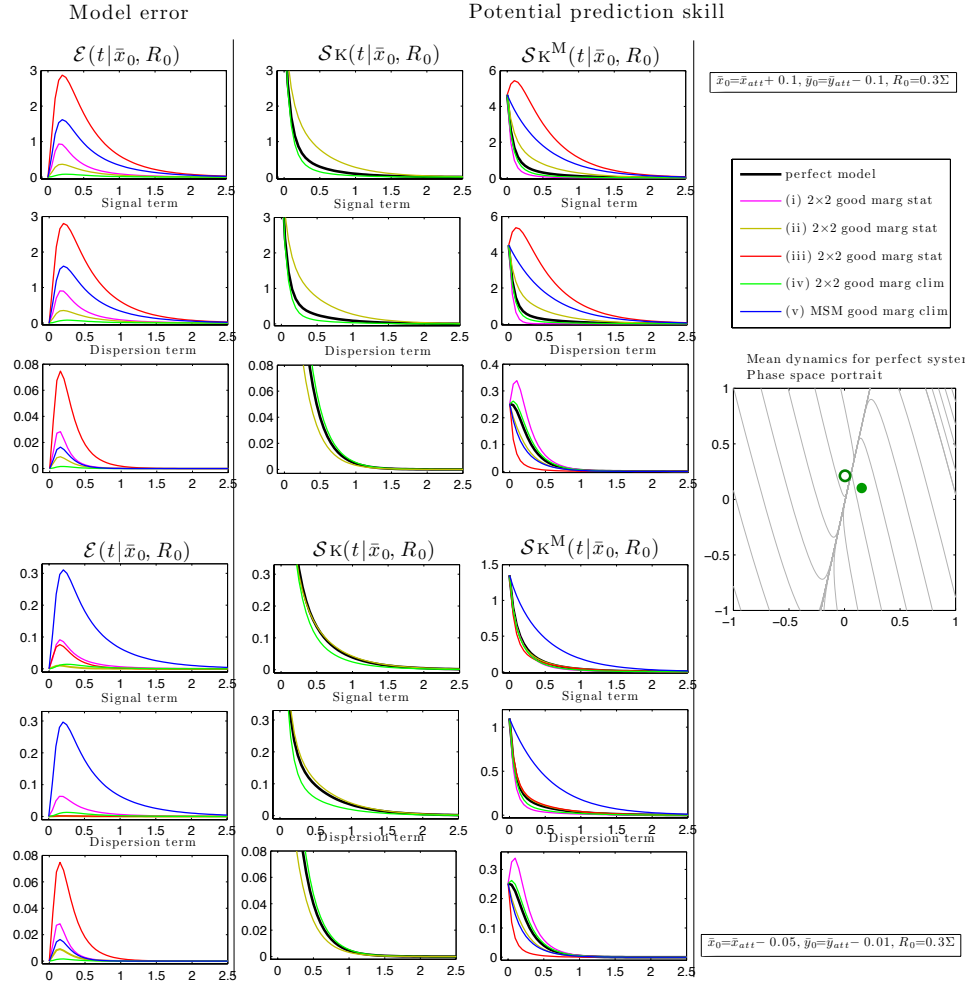
FIGURE 3. Model error (4) and potential prediction skill (5) for forecasts using the imperfect 2×2 linear systems (18) and the MSM (24) with statistical equilibrium fidelity (13) (models (i)-(iii) also have the correct two-point equilibrium statistics; see Def. 3.1). The perfect system **dynamics is non-normal** with parameters ($a = -5.5, \lambda_1 = -1, \lambda_2 = -10, \sigma = 0.5, F = 0.1$). The two examples are shown for different statistical initial conditions $\bar{x}_0, \bar{y}_0, R_0$.
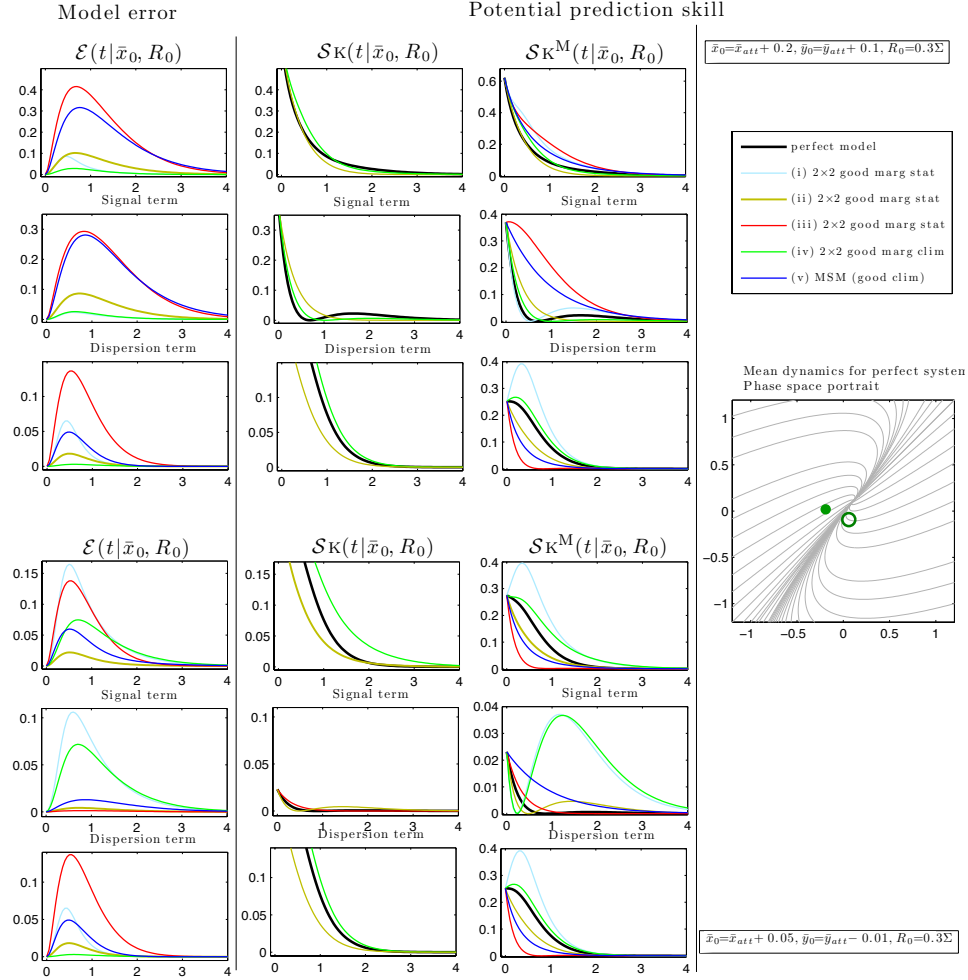
Figure 4. Model error (4) and potential prediction skill (5) for forecasts using the imperfect 2×2 linear systems (18) and the MSm (24) with statistical equilibrium fidelity (13) (models (i)-(iii) also have the correct two-point equilibrium statistics; see Def. 3.1). The perfect system **dynamics is nearly normal** with parameters ($a = -2, \lambda_1 = -1, \lambda_2 = -1.1, \sigma = 0.5, F = 0.1$). The two examples are shown for different statistical initial conditions $\bar{x}_0, \bar{y}_0, R_0$.

Model error $\qquad$ $\mathcal{S}_{\mathrm{K}}(t|\bar{x}_0, R_0)$ $\qquad$ Internal prediction skill
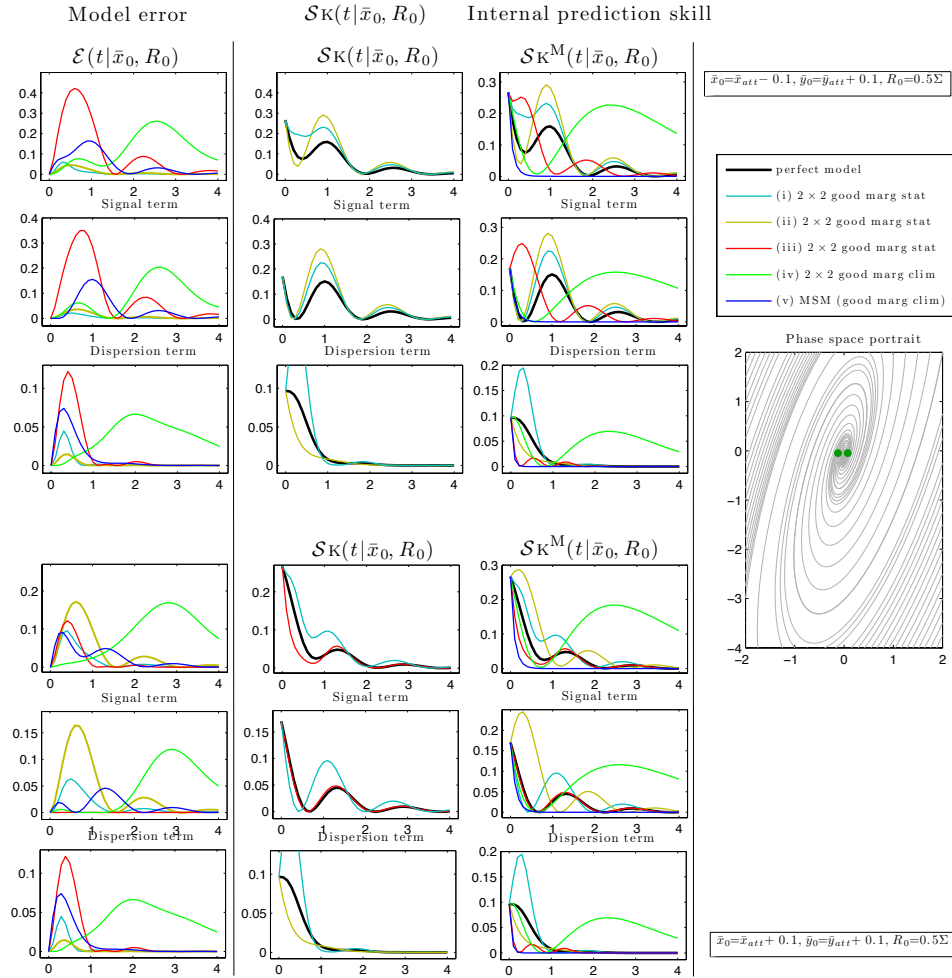

FIGURE 5. Model error (4) and potential prediction skill (5) for forecasts using the imperfect 2×2 linear systems (18) and the MSm (24) with statistical equilibrium fidelity (13) (models (i)-(iii) also have the correct two-point equilibrium statistics; see Def. 3.1). The perfect system and the imperfect models (except (iv)), have complex-conjugate eigenvalues $\lambda_{1,2} = \alpha + \beta i, \ \alpha \ll \beta$; the imperfect model (iv) has purely real eigenvalues and correct marginal climatology and parameters ($a = -2, \lambda_{1,2} = -0.5 \pm 2i, \sigma = 0.5, F = 0.1$). The two examples are shown for different statistical initial conditions $\bar{x}_0, \bar{y}_0, R_0$.
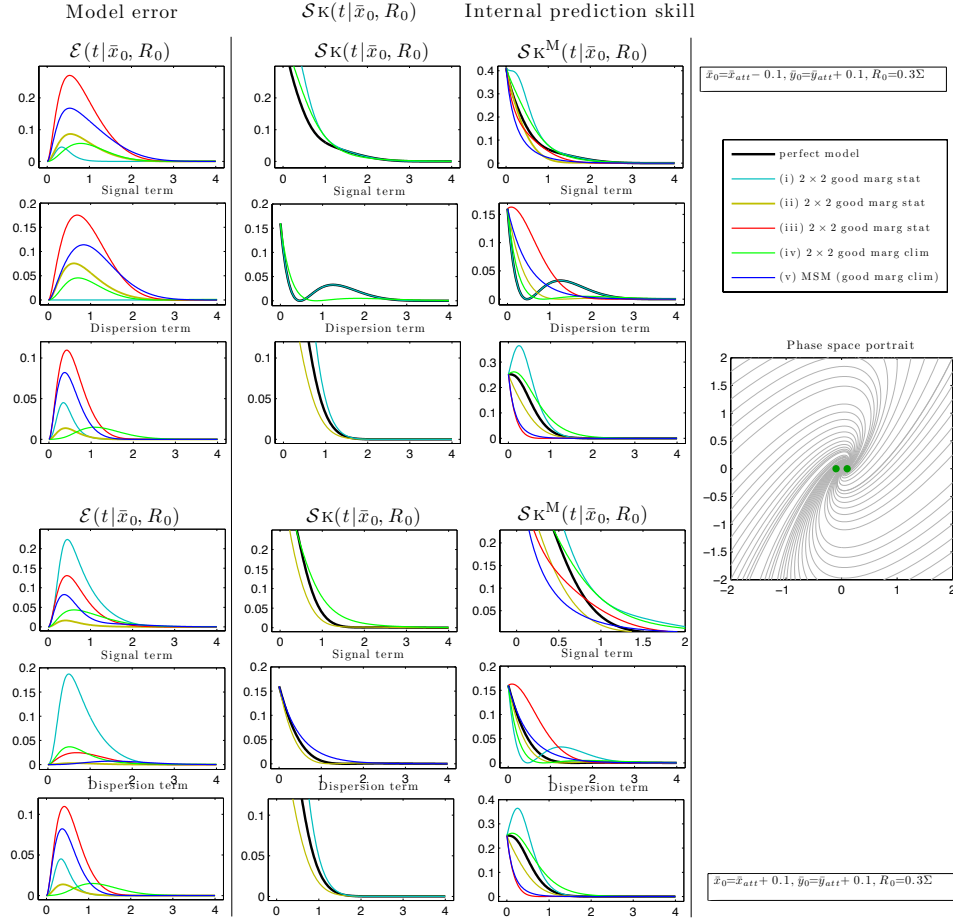
FIGURE 6. Model error (4) and potential prediction skill (5) for forecasts using the imperfect 2×2 linear systems (18) and the MSM (24) with statistical equilibrium fidelity (13) (models (i)-(iii) also have the correct two-point equilibrium statistics; see Def. 3.1). The perfect system and the imperfect models (except (iv)), have complex-conjugate eigenvalues $\boldsymbol{\lambda_{1,2} = \alpha + \beta i, \ \alpha \sim \beta}$; the imperfect model (iv) has purely real eigenvalues and correct marginal climatology with parameters ($a = -2, \lambda_{1,2} = -1 \pm 1i, \sigma = 0.5, F = 0.1$). The two examples are shown for different statistical initial conditions $\bar{x}_0, \bar{y}_0, R_0$.

4. **Elementary Non-Gaussian test models with intermittency.** Here, we discuss two mathematically tractable nonlinear and non-Gaussian models with intermittency which are very useful for analyzing and understanding many nontrivial aspects of UQ and predictive skill in more complex systems; the discussion of UQ in the spatially extended systems presented in §6 builds on these simple models by combining them in an appropriate fashion in the Fourier domain.

The first system with intermittency [30, 29, 94, 9] discussed below (§4.1) combines turbulent dynamics of a complex scalar with unresolved processes affecting the resolved dynamics through colored noise fluctuations in the damping and colored noise fluctuations in the forcing. The second system [76], discussed in §4.2, is a one-dimensional reduced climate model for low-frequency variability with cubic nonlinearities and correlated additive and multiplicative white noise forcing; this model displays regime switching despite a unimodal PDF. Both of these models are very useful for studying various imperfect Gaussian or non-Gaussian approximations obtained either via the direct simplification of the true dynamics (e.g., linearization of the underlying dynamics, dimensional reduction) and/or moment closure approximations of the turbulent fluxes. This stochastic approach allows for analyzing many properties which are relevant for UQ and prediction in high-dimensional turbulent systems in a greatly simplified one-mode setting.

Important points discussed here include:

- Effects of various types of model error on the prediction skill in nonlinear systems with non-trivial mean-turbulent flux interactions;
- Utility of the information-theoretic framework for systematic model improvement in the presence of intermittency and positive Lyapunov exponents;
- Existence of information barriers for improving the sensitivity response of imperfect models.
- False multiple equilibria in the imperfect models;
- Physically consistent white noise limits of complex dynamical systems as test models with judicious model error for UQ and filtering/data assimilation.

4.1. **Simple test model with hidden dynamics and intermittency.** Consider a single Fourier mode $u(t)$ of a turbulent signal modeled by the following stochastic system (see [30, 29, 9, 11])

$$
\begin{aligned}
(a) \quad & \mathrm{d}u(t) = \big[(-\gamma(t)+i\omega)u(t)+b(t)+F(t)\big]\mathrm{d}t+\sigma_u\mathrm{d}W_u(t), \\
(b) \quad & \mathrm{d}b(t) = (-d_b+i\omega_b)b(t)\mathrm{d}t+\sigma_b\mathrm{d}W_b(t), \qquad\qquad (38) \\
(c) \quad & \mathrm{d}\gamma(t) = -d_\gamma(\gamma(t)-\hat{\gamma})\mathrm{d}t+\sigma_\gamma\mathrm{d}W_\gamma(t),
\end{aligned}
$$

where W_u, W_b are independent complex Wiener processes and W_γ is a real Wiener process defined for all time $t \in \mathbb{R}$. There are eight parameters in the system (38): two damping parameters d_b, $d_\gamma > 0$, two oscillation frequencies ω and ω_b, the stationary mean of the damping $\hat{\gamma}$, and noise amplitudes $\sigma_u, \sigma_b, \sigma_\gamma > 0$; $F(t)$ is a deterministic forcing.

Here, we regard $u(t)$ as one of the resolved modes in a turbulent signal where the nonlinear mode-interaction terms are replaced by a stochastic drag $\gamma(t)$ and an additive noise term $b(t)$, as is often done in turbulence models [75, 18]. The nonlinear system (38), introduced first in [30] for filtering multiscale turbulent signals with hidden instabilities has a rich dynamics mimicking turbulent signals in various
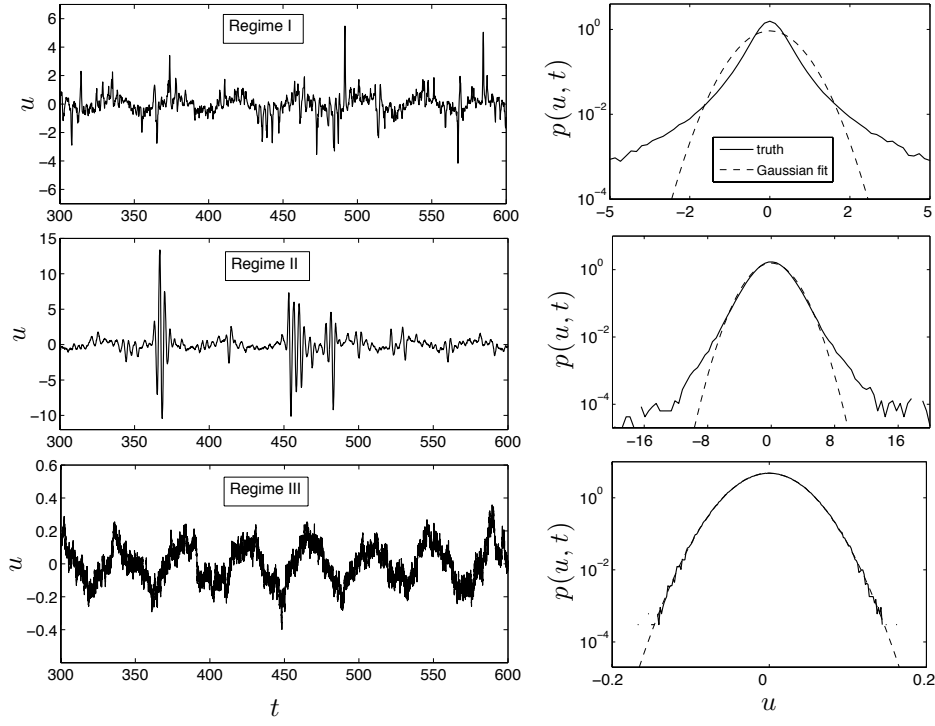
FIGURE 7. **Dynamical regimes of the elementary model with intermittency (38)**. Schematic illustration of path-wise dynamics (left) and time-periodic statistics (at fixed time) on the attractor (right) in two intermittent and one laminar regime of the system (38); these dynamical regimes are discussed in §4.1.

regimes of the turbulent spectrum, including intermittently positive finite-time Lyapunov exponents, as discussed below (see figure 7 and [9, 11]). Moreover, due to the particular structure of the nonlinearity in (38), exact path-wise solutions and exact second-order statistics of this non-Gaussian system can be obtained analytically, as discussed in [33, 34, 30]. The mathematical tractability of this model and its rich dynamical behavior provides a perfect testbed for analyzing effects of errors due to various moment closure approximations and dimensional reduction in a suite of imperfect models introduced in §4.1.1.

The physically relevant dynamical regimes of (38) satisfying the mean-stability condition (see [9])

$$\chi = -\hat{\gamma} + \frac{\sigma_\gamma^2}{2d_\gamma^2} < 0, \tag{39}$$

are (see also figure 7):

(I) *Regime of plentiful, short-lasting transient instabilities in the resolved component $u(t)$ with fat-tailed marginal equilibrium PDF.* This is a regime characteristic of the turbulent energy transfer range and is associated with rapidly decorrelating damping fluctuations $\gamma(t)$;

typical parameter values are $\sigma_\gamma, d_\gamma \gg 1$, $\sigma_\gamma/d_\gamma \sim \mathcal{O}(1)$ and positive $\hat{\gamma} \sim \mathcal{O}(1)$.

(II) *Regime of intermittent large-amplitude bursts of instability in $u(t)$ with fat-tailed marginal equilibrium PDF.* This regime is characteristic of turbulent modes in the dissipative range and it occurs for small σ_γ, d_γ, with $\sigma_\gamma/d_\gamma \sim \mathcal{O}(1)$ and $\hat{\gamma} \sim \mathcal{O}(1)$.

(III) *'Laminar' regime with nearly Gaussian equilibrium PDF.* This regime is characteristic of the laminar modes in the turbulent spectrum. Here, fluctuations in $u(t)$ decorrelate rapidly compared to the damping fluctuations $\gamma(t)$ and the transient instabilities in $u(t)$ occur very rarely. This type of dynamics occurs in (38) for $\sigma_\gamma^2/2d_\gamma^2 \gg 1$, $\sigma_\gamma \sim \mathcal{O}(1)$ and $\hat{\gamma} \gg 1$.

4.1.1. *Imperfect models, information-theoretic optimization and information barriers.* We describe here four classes of imperfect models derived from (38) which introduce model errors through various moment closure approximations or a dimensional reduction of the system (38); two of these models are nonlinear with 'Gaussianized' statistics while the simplest model, the MSm discussed already in §3, is both linear and Gaussian. The effects of model errors introduced by these approximations and ways of mitigating these errors are discussed in §4.1.2 where we apply the information-theoretic framework discussed in §2 to highlight the following important points:

- The information-theoretic optimization (§2.1) can dramatically improve predictive performance and sensitivity of imperfect models. In particular, for systems with too much dissipation a simple inflation of the stochastic forcing in order to optimize the time-averaged climate fidelity (14) leads to a significant point-wise reduction of model error and greatly improved prediction skill in models with consistent moment closure approximations.
- Climate fidelity (cf. (13), (14)) of imperfect models on the coarse-grained subset of resolved variables is necessary but not sufficient for high skill in forced response predictions. Climate fidelity requires the optimization of the whole model density and not its individual moments.
- There exist information barriers to model improvements within a given class of imperfect models.
- Nonlinear, non-Gaussian models can have long memory of initial conditions, including the memory of initial conditions for the unresolved processes.
- Linear Gaussian models cannot reproduce the response in the variance to forcing perturbations in nonlinear systems [82].

The above issues, alongside those discussed in the context of the linear Gaussian framework in §3, will be important in the spatially extended framework discussed in §6; the examples discussed here provide an unambiguous illustration of analogous phenomena occurring in vastly more complex natural systems where they are much more difficult to capture.

Note first that, with the state vector $\boldsymbol{v} = (u,\, b,\, \gamma)^T$ of the system (38) with one resolved component u and two hidden components (b, γ), the deterministic part in (38) can be written as

$$\boldsymbol{f}(\boldsymbol{v}, t) = \hat{L}(t)\boldsymbol{v} + \mathbf{B}(\boldsymbol{v}, \boldsymbol{v}, t) + \mathbf{F}(t), \tag{40}$$

with $\hat{L}$ a linear operator, $\mathbf{B}$ a bilinear function, and $\mathbf{F}$ a spatially uniform term representing generalized deterministic forcing; the exact form of these terms can be

easily obtained by comparing (40) with (38) and (1). In what follows we will skip the explicit dependence on time in $\boldsymbol{f}$ in order to simplify the notation.

It can be easily shown (e.g., [52, 9]) that by adopting an analogue of the averaged Reynolds decomposition of the state vector, $\boldsymbol{v} = \bar{\boldsymbol{v}} + \boldsymbol{v}'$, such that $\overline{\boldsymbol{v}'} = 0$ and $\overline{\bar{v}_i v_j'} = 0$, the evolution of the mean $\bar{\boldsymbol{v}}$ and covariance $R \equiv \overline{\boldsymbol{v}' \otimes \boldsymbol{v}'^T}$ for the process $\boldsymbol{v}$ satisfying (38) is given by

$$a) \quad \dot{\bar{\boldsymbol{v}}} = \boldsymbol{f}(\bar{\boldsymbol{v}}) + \overline{\mathbf{B}(\boldsymbol{v}', \boldsymbol{v}')},$$

$$b) \quad \dot{R} = R\,A^T(\bar{\boldsymbol{v}}) + A(\bar{\boldsymbol{v}})R + \Sigma + \overline{\boldsymbol{v}'\mathbf{B}^T(\boldsymbol{v}', \boldsymbol{v}') + \mathbf{B}(\boldsymbol{v}', \boldsymbol{v}')\boldsymbol{v}'^T}, \tag{41}$$

where A is the Jacobian of $\boldsymbol{f}$ at $\bar{\boldsymbol{v}}$, i.e., $A(\bar{\boldsymbol{v}}) \equiv \nabla \boldsymbol{f}(\bar{\boldsymbol{v}})$ and the overbar denotes an ensemble average.

Due to the particular form of the quadratic nonlinearity in the perfect system (38), exact formulas for the second-order statistics and path-wise solutions can be derived [30] without the explicit knowledge of the associated time-dependent probability density. Derivation of imperfect models in this framework utilizes some type of moment closure approximation applied to the turbulent fluxes in (41). Below we describe two nonlinear and two linear imperfect models of the system (38) which introduce model error due to various moment closure approximations applied to (41) and/or due to a dimensional reduction. These are standard ways to introduce model error in vastly more complex turbulent systems with the form (40) (e.g, [22]). The effects of errors introduced by these models on the short, medium and long range prediction skill, as well as ways of mitigating these errors via the information-theoretic framework of §2, are discussed in the following sections. It is important to foreshadow the following discussion and stress that while the first three imperfect models below have Gaussian statistics due to the moment closure approximations, the first two models remain nonlinear in the state variables $\boldsymbol{v} = (u, b, \gamma)$. On the other hand, the last imperfect model is linear and non-Gaussian due to the presence of multiplicative noise which arises through appropriate white noise limits of the true dynamics.

GAUSSIAN CLOSURE MODEL (GCM). This model is nonlinear in the state variables but has a 'Gaussianized' statistics with nontrivial and consistent mean-fluctuation interactions. The quasi-Gaussian closure approximation (e.g., [52]), which is familiar from turbulence theory, implies neglecting the third moment of the true probability density $p(\boldsymbol{v}, t)$ in the equation in (41b) for the covariances. Thus, we assume

$$\overline{\boldsymbol{v}'_M\,\mathbf{B}^T(\boldsymbol{v}'_M, \boldsymbol{v}'_M) + \mathbf{B}(\boldsymbol{v}'_M, \boldsymbol{v}'_M)\,\boldsymbol{v}'_M{}^T} = 0 \tag{42}$$

in (41b) with $\boldsymbol{v}_M = (u_M, b_M, \gamma_M)$, $\boldsymbol{v}_M = \bar{\boldsymbol{v}}_M + \boldsymbol{v}'_M$. The closure (42) results in a fully coupled nonlinear dynamical system for the second-order statistics and it introduces model error due to neglecting the turbulent fluxes in the covariance (41b) of (38). Thus, the GCm closure for a general quadratic system is given by

$$a) \quad \dot{\bar{\boldsymbol{v}}}_M = \boldsymbol{f}(\bar{\boldsymbol{v}}_M) + \overline{\mathbf{B}(\boldsymbol{v}'_M, \boldsymbol{v}'_M)}, \qquad b) \quad \dot{R}_M = R_M\,A^T(\bar{\boldsymbol{v}}_M) + A(\bar{\boldsymbol{v}}_M)\,R_M + \Sigma_M, \tag{43}$$

with $A(\bar{\boldsymbol{v}}_M) = \nabla \boldsymbol{f}(\bar{\boldsymbol{v}}_M)$.

DETERMINISTIC-MEAN MODEL (DMm). This model, similarly to GCm, is nonlinear in the state variables and has a Gaussianized statistics but, apart from neglecting the turbulent third-order flux in the covariance, it also neglects the turbulent backscatter into the mean. This ad-hoc closure corresponds to assuming

$$a) \quad \overline{\boldsymbol{v}'_{\mathrm{M}}\mathbf{B}^T(\boldsymbol{v}'_{\mathrm{M}}, \boldsymbol{v}'_{\mathrm{M}}) + \mathbf{B}(\boldsymbol{v}'_{\mathrm{M}}, \boldsymbol{v}'_{\mathrm{M}})\boldsymbol{v}'_{\mathrm{M}}{}^T} = 0, \qquad b) \quad \overline{\mathbf{B}(\boldsymbol{v}'_{\mathrm{M}}, \boldsymbol{v}'_{\mathrm{M}})} = 0, \qquad (44)$$

in (41). Note that the constraint (44b) is inconsistent with the nontrivial evolution of the covariance in (41b) and effectively implies that $\overline{\boldsymbol{f}(\boldsymbol{v}_{\mathrm{M}})} = \boldsymbol{f}(\bar{\boldsymbol{v}}_{\mathrm{M}})$. Thus, the DMm closure for a general quadratic system is given by

$$a) \quad \dot{\bar{\boldsymbol{v}}}_{\mathrm{M}} = \boldsymbol{f}(\bar{\boldsymbol{v}}_{\mathrm{M}}), \qquad b) \quad \dot{R}_{\mathrm{M}} = R_{\mathrm{M}} A^T(\bar{\boldsymbol{v}}_{\mathrm{M}}) + A(\bar{\boldsymbol{v}}_{\mathrm{M}})R_{\mathrm{M}} + \Sigma_{\mathrm{M}}, \qquad (45)$$

with $A(\bar{\boldsymbol{v}}_{\mathrm{M}}) \equiv \nabla \boldsymbol{f}(\bar{\boldsymbol{v}}_{\mathrm{M}})$.

MEAN STOCHASTIC MODEL (MSm). This model with reduced dimensionality was already discussed in the linear Gaussian framework of §3. MSm uses the same moment closure approximations (44) as DMm but because of its linearization in the state variables it only accounts for the dynamics of the resolved variable $u(t)$ with the mean values of the unresolved variables in (38a), $\overline{\gamma} = \hat{\gamma}_{\mathrm{MSM}}$ and $\overline{b} = 0$. The MSm model utilized here is given by

$$\dot{u}_{\mathrm{MSM}} = \left(-\hat{\gamma}_{\mathrm{MSM}} + i\omega_{\mathrm{MSM}}\right)u_{\mathrm{MSM}} + F_{\mathrm{MSM}}(t) + \sigma_{\mathrm{MSM}}\dot{W}(t). \qquad (46)$$

The mean $\bar{u}_{\mathrm{MSM}}$ and covariance R_{MSM} for this linear model can be computed analytically in a standard fashion leading to

$$\bar{u}_{\mathrm{MSM}}(t) = e^{(-\hat{\gamma}_{\mathrm{MSM}} + i\omega_{\mathrm{MSM}})(t-t_0)}\bar{u}_0 + \int_{t_0}^t e^{(-\hat{\gamma}_{\mathrm{MSM}} + i\omega_{\mathrm{MSM}})(t-s)}F_{\mathrm{MSM}}(s)\mathrm{d}s, \qquad (47)$$

$$R_{\mathrm{MSM}}(t) = e^{\hat{A}(t-t_0)}R_0 e^{\hat{A}^T(t-t_0)} + \int_{t_0}^t e^{\hat{A}(t-s)}\Sigma_{\mathrm{MSM}}e^{\hat{A}^T(t-s)}\mathrm{d}s, \qquad (48)$$

where $\hat{A} \equiv \nabla \boldsymbol{f}_u|_{(\bar{u}_{\mathrm{MSM}}, b=0, \hat{\gamma}_{\mathrm{MSM}})}$, $\Sigma_{\mathrm{MSM}} = \frac{1}{2}\mathrm{diag}[\sigma_{\mathrm{MSM}}^2, \sigma_{\mathrm{MSM}}^2]$, and the covariance $R_{\mathrm{MSM}} = \overline{\tilde{\boldsymbol{u}} \otimes \tilde{\boldsymbol{u}}^T}$ is considered in the real variables $\tilde{\boldsymbol{u}} = \left(\Re e[u_{\mathrm{MSM}} - \bar{u}_{\mathrm{MSM}}], \Im m[u_{\mathrm{MSM}} - \bar{u}_{\mathrm{MSM}}]\right)^T$. Note that the linearity of MSm implies that the covariance R_{MSM} is insensitive to forcing perturbations. Consequences of this fact on improving the model fidelity and sensitivity are discussed in the next section.

WHITE NOISE LIMITS OF THE TEST SYSTEM (38). One alternative to obtaining reduced models via the moment closure approximations or the direct dynamics linearization is to consider appropriate white noise limits of the unresolved (or poorly understood) components of the perfect system. Such a procedure leads to simplified stochastic models with reduced dimensionality and judicious model error. This approach will prove particularly useful and revealing in §7 where we discuss various numerical artifacts in complex spatially extended models.

Here, we illustrate this strategy on the single-mode system (38) by assuming that both the damping fluctuations $\gamma(t)$ and the additive forcing uncertainty $b(t)$ affecting the resolved dynamics $u(t)$ decorrelate very fast, i.e., we take the formal limit in (38)

$$d_\gamma \to \infty, \qquad \sigma_\gamma \to \infty, \qquad \tilde{\sigma}_\gamma \equiv \sigma_\gamma/d_\gamma = const > 0, \qquad (49)$$

$$d_b \to \infty, \qquad \sigma_b \to \infty, \qquad \tilde{\sigma}_b \equiv \sigma_b/d_b = const > 0. \qquad (50)$$

As shown in more detail in Appendix B, the formal limit preserving the non-vanishing correlations $\overline{u(t)\gamma(t)}$ leads to the Stratonovitch SDE (e.g., [28]) for the complex scalar $u(t)$, given by

$$\mathrm{d}u(t) = \big[\big(-\hat{\gamma}+i\hat{\omega}(t)\big)u(t)+F(t)\big]\mathrm{d}t - \tilde{\sigma}_\gamma u(t)\circ\mathrm{d}W_\gamma(t)+\tilde{\sigma}_b\mathrm{d}W_b(t)+\sigma_u\mathrm{d}W_u(t), \quad (51)$$

where W_γ and the components of $W_u = \frac{1}{\sqrt{2}}(W_{1u}+iW_{2u})$ and $W_b = \frac{1}{\sqrt{2}}(W_{1b}+iW_{2b})$ are independent Wiener processes defined for all time $t \in I\!\!R$. We rewrite (111) in Ito form as

$$\mathrm{d}u(t) = \big[\big(-\alpha+i\hat{\omega}(t)\big)u(t)+F(t)\big]\mathrm{d}t - \tilde{\sigma}_\gamma u(t)\mathrm{d}W_\gamma(t)+\tilde{\sigma}_b\mathrm{d}W_b(t)+\sigma_u\mathrm{d}W_u(t), \quad (52)$$

where $\alpha = \hat{\gamma} - \frac{1}{2}\tilde{\sigma}_\gamma^2$ is the damping term in the white noise limit. In contrast to the original system (38) with colored multiplicative and additive noise terms, the marginal equilibrium PDF for the components $u_R \equiv \Re e[u]$, $u_I \equiv \Im m[u]$ for the system (112) in the unforced case $F \equiv 0$ can be obtained analytically (see Appendix B) in the form

$$p_{eq}(u_R) = \frac{N_0}{(\tilde{\sigma}_\gamma^2 u_R^2 + \tilde{\sigma}_b^2 + \sigma_u^2)^\delta}, \quad (53)$$

with $\delta = \frac{1}{2} + (\hat{\gamma} - \frac{1}{2}\tilde{\sigma}_\gamma^2)/\tilde{\sigma}_\gamma^2$. Although this is not shown, the marginal equilibrium PDF (53) for $u(t)$ obtained in the white noise limit of $\gamma(t), b(t)$ provides a very good approximation of the marginal equilibrium statistics of $u(t)$ in regime I (abundant transient instabilities) of the original system (38).

4.1.2. *Using stochastic parameterization to improve prediction skill and forced response of imperfect models.* Here, we illustrate the utility of the information-theoretic framework discussed in §2 to systematically and simultaneously improve the climate fidelity and prediction skill of imperfect models GCm, DMm, and MSm of the system (38) in the presence of intermittency and positive Lyapunov exponents. We illustrate this approach in this nonlinear and non-Gaussian configuration by discussing: (i) model optimization and the existence of information barriers, (ii) forced response of optimized models, and (iii) the prediction skill of optimized imperfect models when both the initial conditions and the forced response are important; for more details see [11].

Note that the moment closures employed in the derivation of GCm and DMm discussed in the previous section, or the dimensionality reduction used in MSm, do not preserve a-priori the statistical attractor of the perfect system. In fact, as was shown in [11], these imperfect models have too much dissipation and underestimate both the equilibrium mean and variance relative to the perfect system (see figure 8). Here, following the methodology of [80, 11], we focus on improving the model fidelity by inflating the stochastic forcing in the resolved dynamics of the imperfect models in order to minimize the annually averaged information content, $\overline{\mathcal{P}(\pi_{att}, \pi_{att}^{\mathrm{M}})}$ as in (14) of §2. In figure 8 we show an example of such an optimization procedure carried out in regime II (see §4.1) of mean-stable dynamics of the perfect system (38); we point out the following:

- For all imperfect models the attractor fidelity is significantly improved by inflating the amplitude of the stochastic forcing $\sigma_u^{M_*}$. The best results can be achieved for GCm which is based on consistent moment closure approximations.
- The total lack of information in the imperfect model attractor dynamics is greatly reduced at all times in the time-periodic setting even though only the
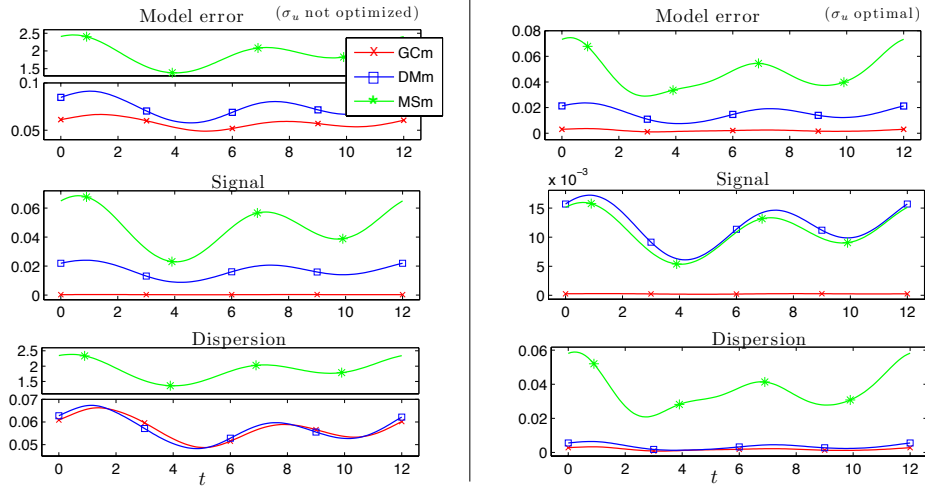
FIGURE 8. **Information-theoretic optimization of imperfect models of the quadratic intermittent system (38)**. Model error in the climate, $\mathcal{P}(\pi_{att}, \pi_{att}^{\text{M}})$, before (left) and after (right) the optimization of the imperfect models of the system (38) by stochastic noise inflation aimed at minimizing the time-averaged relative entropy (14). Note the significant improvement in the optimized imperfect models reflected in the point-wise decrease of the relative entropy. This example is typical of regime II in (38); the parameters used in computations are $\hat{\gamma} = 1.2, \sigma_u = 0.5, \sigma_\gamma = d_\gamma = 0.5$.

time-averaged formula (14) is used in the optimization procedure; this was observed earlier for the Gaussian models in [32].

In the examples discussed below we use the time periodic forcing:

$$F(t) = A_0 + A_1 \cos(\omega t + \phi_1) + A_2 \cos(2\omega t + \phi_2), \tag{54}$$

where $\phi_1 = 1, \phi_2 = 0$ and $\omega = \pi/6$ which ensures that the period $T_0 = 12$ so that it can be interpreted as a year consisting of 12 months; the forcing has two frequencies so that the equilibrium statistics of (38) has a more variable structure.

INFORMATION BARRIERS TO IMPERFECT MODEL IMPROVEMENT

The important issue in model optimization concerns the extent to which the climate/attractor fidelity of the imperfect models can be improved. This issue is particularly important and interesting in situations when the true system has a wide range of dynamical regimes, including turbulent regimes with intermittency. In figure 9, which is complementary to figure 8, we show results of the information-theoretic optimization of the imperfect models of (38) which is achieved by the stochastic noise inflation for different parameters and dynamical regimes of the perfect system (38). As before, the optimal noise amplitude $\sigma_u^{\text{M}*}$ is determined by minimizing the time-averaged lack of information (14) in the imperfect models. The following remarks are based on figure 9:

- There exist barriers to model improvement by noise inflation for DMm and MSm as the hidden, transient instabilities become more abundant (see the transition between regimes II and I of (38) for increasing d_γ in figure 9).
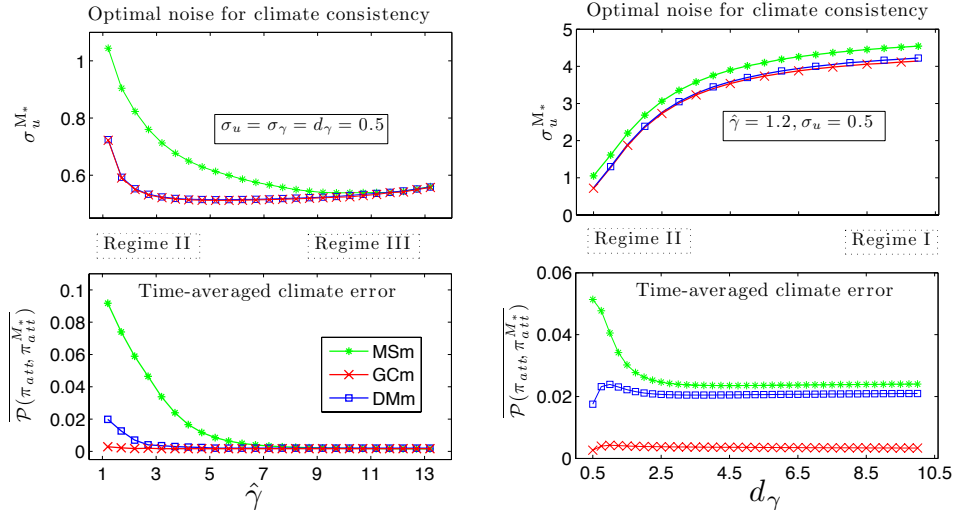
FIGURE 9. **Information-theoretic optimization of imperfect models of the quadratic intermittent system (38)**. Optimal noise amplitude σ_u^{M*} for the time-averaged climate fidelity (14) and the corresponding model error as functions of the mean damping $\hat{\gamma}$ and damping of fluctuations d_γ for the imperfect models GCm, DMm and MSm. Note the barriers to model improvement by stochastic forcing inflation in DMm and MSm for increasing d_γ as the hidden, transient instabilities become more abundant in $u(t)$ on the approach to Regime I (see figure 7).

- For weak mean damping $\hat{\gamma}$, associated with very intermittent dynamics of $u(t)$ and fat-tailed marginal equilibrium PDFs (see Regime II in figure 7), perfect attractor/climate fidelity cannot be achieved by inflating the stochastic forcing
- GCm retains the best climate fidelity throughout the parameter range.

FORCED RESPONSE OF IMPERFECT MODELS WITH OPTIMAL NOISE

Here, the main focus is on elucidating the link between the attractor/climate fidelity of imperfect models and their sensitivity to perturbations $\delta F(t)$ of the external forcing. Consequently, we consider here long lead times so that essentially all knowledge of the initial conditions is lost and only the perturbations to the attractor of the perfect system are important. The sensitivity of the perfect system and its imperfect models to such perturbations (see Definition 2.3) is quantified via the relative entropies $\mathcal{P}(\pi_{\delta F}, \pi_{att})$, $\mathcal{P}(\pi_{\delta F}^{M*}, \pi_{att}^{M*})$; this is illustrated in figure 10 in a typical configuration with intermittent transient instabilities with

$$\delta F(t) = (A_0^\delta - A_0)\frac{\tanh(a(t - t_c)) + \tanh(a\,t_c))}{1 + \tanh(a\,t_c))}, \tag{55}$$

where A_0^δ is the perturbed mean forcing, A_0 is the unperturbed mean forcing, and a controls the time scale of the perturbation centered at time t_c; in the computations are $A_0 = 4, A_0^\delta = 4.4, a = 0.3, t_c = 20$.

The following observations are worth pointing out in this context:

- The sensitivity of GCm and DMm in turbulent regimes with sufficiently long decorrelation time of the damping fluctuations γ is comparable to that of the perfect model despite the underestimated imperfect model response in the
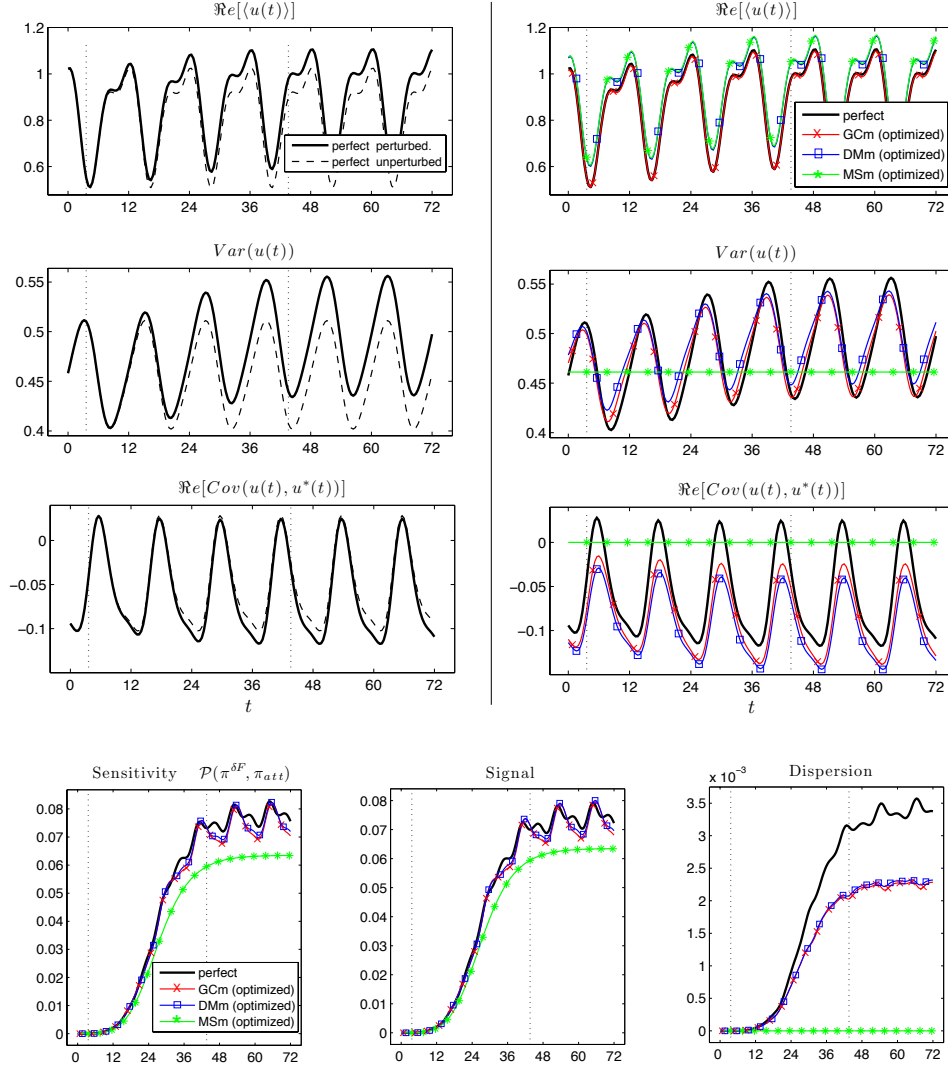
FIGURE 10. **Forced response of the quadratic model with intermittency (38) and the sensitivity of its imperfect models.** (Top left) Response of the resolved component $u(t)$ in the perfect system (38) to ramp-type forcing perturbations δF. (Top right) Analogous response computed from the imperfect models with optimal noise $\sigma_u^{\mathrm{M}*}$ in $u^{\mathrm{M}}(t)$. (Bottom) The sensitivity $\mathcal{P}(\pi_{\delta F}, \pi_{att})$ (6) of the perfect system and its imperfect models to the forcing perturbations δF (55). Note that MSm fails to reproduce the response in the covariance due to its inherent linearization. This example represents a typical situation in regime II (intermittent transient instabilities with large amplitudes in $u(t)$; see figure 7) of the perfect model dynamics (38), the perfect system parameters used in computations are $\hat{\gamma} = 1.2, \sigma_u = 0.5, \sigma_\gamma = d_\gamma = 0.5$.
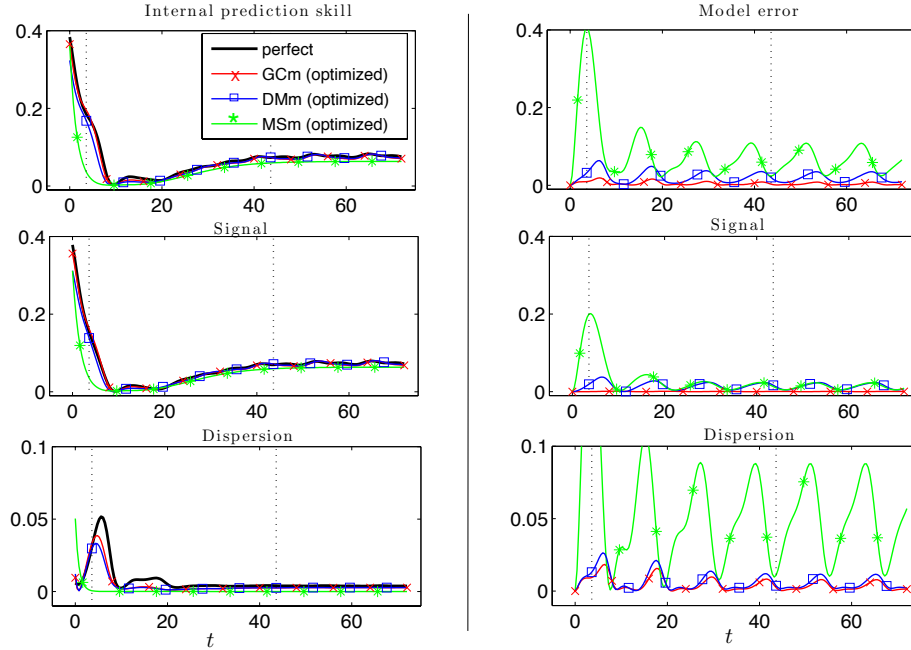
FIGURE 11. **Interplay between the initial data response and the forced response for the quadratic model with intermittency (38)**. (Right) Model error $\mathcal{P}(\pi_{\delta F,t_0}, \pi_{\delta F,t_0}^{\mathrm{M}})$ (4) and (left) the internal prediction skill, $\mathcal{P}(\pi_{\delta F,t_0}, \pi_{att}^{\mathrm{M}})$, for the predictions of the resolved component $u(t)$ using the imperfect models optimized for unperturbed climate fidelity (14). The initial conditions at t_0 are off the attractor and the subsequent climate change is induced by ramp-type forcing perturbations δF (55). Note the high skill of GCm at all ranges, including the short range skill in the dispersion. This situation is typical of regime II (see figure 7) of the perfect model (38) where both GCm and DMm achieve perturbed climate consistency; MSm has no short range skill and only a marginal skill for the long range forecasts. The perfect system parameters used in computations are $\hat{\gamma} = 1.2, \sigma_u = 0.5, \sigma_\gamma = d_\gamma = 0.5$.

dispersion. Unlike DMm and MSm, the sensitivity of GCm for the response in the mean remains very good in all dynamical regimes (not shown); this stems from the consistent moment closure approximations used in GCm.

• The linear Gaussian model MSm completely fails to detect the covariance response to the forcing perturbations. The linearization inherent in the MSm has relatively little effect in regimes with no transient instabilities in the resolved dynamics (e.g., nearly-Gaussian regime III of (38)) when all the discussed models perform well (not shown).

INTERPLAY BETWEEN INITIAL DATA AND FORCED RESPONSE WITH OPTIMAL NOISE

The most difficult forecasting scenarios are those involving the short and medium range imperfect model predictions when both the memory of the initial conditions, model error, and the response due to forcing perturbations are important in the forecasts [12, 117, 37, 38].

In figure 11 we illustrate the predictive skill of the imperfect models with optimal noise in a scenario where the statistical initial condition is away from the attractor and the external forcing is perturbed at some later time during the evolution. As described in §2, the predictive skill of the imperfect models optimized for climate fidelity can be assessed by monitoring the model error, $\mathcal{P}(\pi_{\delta F,t_0}(t), \pi_{\delta F,t_0}^{\mathrm{M}*}(t))$, and the internal prediction skill $\mathcal{P}(\pi_{\delta F,t_0}^{\mathrm{M}*}(t), \pi_{att}^{\mathrm{M}*}(t))$. The following remarks are based on the results illustrated in figure 11:

- GCm with optimal stochastic forcing has good prediction skill in dynamical regimes with intermittent large-amplitude transient instabilities (e.g., regime II of (38)), including the short term initial data response and the long term perturbed climate/attractor consistency. DMm has a good short range prediction skill and MSm has no skill in this regime.
- All optimized imperfect models achieve perturbed attractor consistency in regimes with essentially no transient instabilities like, for example, regime III of (38 (not shown see [11]). GCm and DMm are essentially the same in this regime and have good prediction skill for all lead times. MSm, which completely neglects the dynamics of the unresolved variables has no short range prediction skill due to a poor initial data response.
- In dynamical regimes characterized by abundant short-lasting transient instabilities (e.g., regime I of (38)) GCm has a good skill for predicting the mean dynamics but it fails, like all the other imperfect models, at predicting the covariances (the dispersion part in these cases dominates the total model error).

4.2. **Cubic scalar SDE with regime switching and unimodal PDF.** The systematic development of reduced low-dimensional stochastic models from observations or comprehensive high-dimensional models is an important topic for atmospheric low-frequency variability, climate sensitivity, and improved extended range forecasting; other areas where such issues are important include neural science and engineering applications in situations where low-frequency variability can be efficiently described by just a few large-scale teleconnection patterns dominating the coarse-grained dynamics. Systematic strategies are essential in successfully developing reduced models and the recently developed stochastic mode reduction strategy provides a systematic procedure for the derivation of reduced stochastic models [86, 85, 75, 77, 78]. Systematically reduced stochastic models are attractive for sensitivity studies of the climate and other complex high-dimensional systems because: (i) they are computationally much more efficient than state-of-the-art climate models and have been shown to have comparable prediction skill, and (ii) they allow for a better understanding of the climate system due to the reduced complexity. Here, we consider an important example of a nonlinear scalar reduced model with correlated additive and multiplicative noise arising from advection of the large scales by the small scales and simultaneously strong cubic damping. This normal form for a single low frequency variable possesses an interesting dynamical configuration with regime switching despite unimodality of the associated PDF resembling the situation occurring in comprehensive climate models [77, 74]. Two imperfect models, one Gaussian and one non-Gaussian, are utilized below to highlight the following issues:
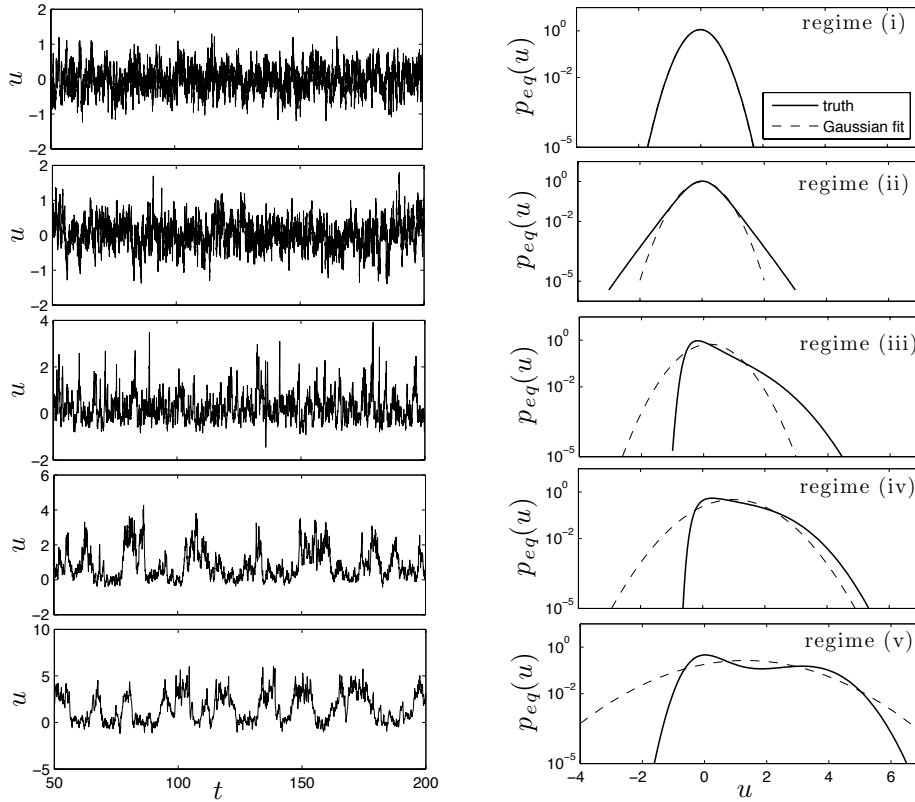
FIGURE 12. Schematic illustration of relevant dynamical regimes of the cubic scalar system (56) with correlated additive and multiplicative noise. (Left) Path wise solutions and corresponding equilibrium PDFs (Right). The information-theoretic optimization of two imperfect models of (56), the nonlinear Gaussian model (GCm) and the linear non-Gaussian model (58), is discussed in 4.2.1.

- Gaussian approximations of non-Gaussian systems can have, in certain circumstances, a smaller model error and better climate fidelity than non-Gaussian imperfect models.
- Moment closure approximations of nonlinear systems can have false multiple equilibria which can be removed by information-theoretic optimization.

The cubic scalar framework utilized here provides the simplest possible illustration of both these issues which are important in complex high-dimensional situations where they are much more difficult to isolate.

The one-dimensional reduced climate model [76] for low-frequency variability with cubic nonlinearities and correlated additive and multiplicative white noise forcing is given by

$$\mathrm{d}u = \big(-au + bu^2 - cu^3 + f(t) \big)\mathrm{d}t + (A - Bu)\mathrm{d}W_{AB}(t) + \sigma_u\mathrm{d}W_u(t), \qquad (56)$$

where a, b, c, A, B, σ_u are arbitrary real constant coefficients, and $W_{AB}(t)$, $W_u(t)$ are independent Wiener processes defined for $t \in I\!\!R$; the correlated additive and multiplicative noise associated with W_{AB} arises through the systematic stochastic mode reduction procedure [76]. Even for time-independent forcing this model possesses a wide range of interesting dynamical regimes which can be deduced from its equilibrium PDF given by (see [76])

$$p_{eq}(u) = \frac{N_0}{((Bu - A)^2 + \sigma_u^2)^{a_1}} \exp\left(d_1 \arctan\left(\frac{Bu - A}{\sigma_u}\right)\right) \exp\left(-\frac{c_1 u^2 + b_1 u}{B^4}\right), \tag{57}$$

where N_0 is a normalizing constant and

$$a_1 = 1 + \frac{3A^2 c + aB^2 - 2AbB - c\sigma^2}{B^4}, \quad b_1 = 2bB^2 - 4cAB, \quad c_1 = cB^2,$$

$$d = \frac{d_1}{\sigma_u} + d_2 \sigma_u, \quad d_1 = 2\frac{A^2 bB - A^3 c - AaB^2 + B^3 f}{B^4}, \quad d_2 = \frac{6cA - 2bB}{B^4}.$$

The dynamical regimes of interest are illustrated in figure 12 and include:

  (i) Nearly Gaussian regime;
  (ii) Regime with symmetric fat-tailed algebraic tails;
  (iii) Regime with highly skewed unimodal PDF and no regime switching;
  (iv) Regime with highly skewed unimodal PDF and regime switching;
  (v) Bimodal regime.

The non-Gaussian cubic system (56) serves as an instructive and revealing model for comparing nonlinear Gaussian and linear non-Gaussian approximations and understanding the origins of information barriers to model improvement in the regimes (i)-(iv) of the cubic test system (56). Here, we consider two imperfect models, one linear non-Gaussian, and one nonlinear Gaussian; both these models are described below:

Non-Gaussian linear model with multiplicative noise

This model is obtained via the linearization of (56) in the state variable u which leads to

$$\mathrm{d}u_{\mathrm{M}} = \big(-a_{\mathrm{M}} u_{\mathrm{M}} + f_{\mathrm{M}}(t)\big)\mathrm{d}t + (A_{\mathrm{M}} - B_{\mathrm{M}} u_{\mathrm{M}})\mathrm{d}W_{AB}(t) + \sigma_u^{\mathrm{M}}\mathrm{d}W(t), \tag{58}$$

where the model coefficients $(a_{\mathrm{M}}, b_{\mathrm{M}}, c_{\mathrm{M}}, B_{\mathrm{M}}, A_{\mathrm{M}}, \sigma_u^{\mathrm{M}})$ are, in principle, different from those of the perfect system (56). For time-independent forcing the equilibrium PDF associated with (58) can be easily derived as

$$p_{eq}^{\mathrm{LIN}}(u_{\mathrm{M}}) = \frac{N_0}{((B_{\mathrm{M}} u_{\mathrm{M}} - A_{\mathrm{M}})^2 + \sigma_u^{\mathrm{M}2})_2^{\delta}} \exp\left(d_2 \arctan\left(\frac{B_{\mathrm{M}} u_{\mathrm{M}} - A_{\mathrm{M}}}{\sigma_u^{\mathrm{M}}}\right)\right), \tag{59}$$

where $\delta_2 = 1 - a_{\mathrm{M}}/B_{\mathrm{M}}^2$ and $d_2 = 2(A_{\mathrm{M}} a_{\mathrm{M}} + B_{\mathrm{M}} f_{\mathrm{M}})/B_{\mathrm{M}}^2 \sigma_{\mathrm{M}}$ (compare this with derivations in Appendix B.2.1).

Gaussian closure model (GCm)

This nonlinear model with Gaussianized statistics is obtained in an analogous way to that discussed in §4.1.1 by consistently neglecting odd moments in the equations for the mean and variance of $u(t)$; the resulting closed system of ODEs for the

approximate second-order statistics is given by

$$a) \quad \dot{\bar{u}}_{\mathrm{M}} \quad = -a_{\mathrm{M}}\bar{u}_{\mathrm{M}} + b_{\mathrm{M}}\bar{u}_{\mathrm{M}}^2 + b_{\mathrm{M}}\overline{u'^2}_{\mathrm{M}} - c_{\mathrm{M}}\bar{u}_{\mathrm{M}}^3 - 3c_{\mathrm{M}}\bar{u}_{\mathrm{M}}\overline{u'^2}_{\mathrm{M}} + f_{\mathrm{M}}(t),$$

$$b) \quad \dot{\overline{u'^2}}_{\mathrm{M}} \quad = 2(-a_{\mathrm{M}} + 2b_{\mathrm{M}}\bar{u}_{\mathrm{M}} - 3c_{\mathrm{M}}\bar{u}_{\mathrm{M}}^2 - 3c_{\mathrm{M}}\overline{u'^2}_{\mathrm{M}}) \qquad\qquad (60)$$

$$+B_{\mathrm{M}}^2[\overline{u'^2}_{\mathrm{M}} + (A_{\mathrm{M}}^2 + B_{\mathrm{M}}^2\bar{u}_{\mathrm{M}}^2 - 2A_{\mathrm{M}}B_{\mathrm{M}}\bar{u}_{\mathrm{M}} + \sigma_u^{\mathrm{M}2})],$$

where the model coefficients $(a_{\mathrm{M}}, b_{\mathrm{M}}, c_{\mathrm{M}}, B_{\mathrm{M}}, A_{\mathrm{M}}, \sigma_u{}^{\mathrm{M}})$ are not necessarily identical to those in the perfect system (56). Here, the time-dependent Gaussian PDF for GCm is given by $p^{\mathrm{GCM}}(t) = \mathcal{N}\left(\bar{u}_{\mathrm{M}}, \overline{u'^2}_{\mathrm{M}}\right)$; it is important to note that GCm for the cubic system (56) can have non-unique invariant measures due to false multiple equilibria in (60). This important new consequence of model error is discussed in more detail in the next section and is illustrated in figures 16 and 15.

4.2.1. *Information barriers for model optimization.* The presence of the correlated multiplicative and additive noise and the cubic nonlinearity in the non-Gaussian system (56) for the real scalar $u(t)$ provides a natural complement of the previous quadratic test model (38) with colored damping fluctuations discussed in §4.1. Here, we use the system (56) and its imperfect models discussed in the previous section to address in a simple unambiguous framework two new issues which are important for UQ in more complex high-dimensional systems: (i) the skill of non-Gaussian linear models and quasi-Gaussian nonlinear models for approximating non-Gaussian dynamics, (ii) existence of false multiple equilibria in the statistics of nonlinear imperfect models. The information-theoretic framework described in §2 is utilized again to illustrate these new issues in the simple one-dimensional setting.

Systematic model optimization and information barriers

Figure 13 illustrates the results of information-theoretic optimization of the imperfect nonlinear Gaussian (60) and linear non-Gaussian (58) models of the cubic system (56) discussed in the previous section; here, the optimization (§2) is carried out in various dynamical regimes of the system (56) by a simple additive stochastic noise inflation to achieve climate fidelity (13) through (12). The coefficients in the perfect system (56) used in the computations are

	a	b	c	A	B	σ_u	f
Regime (i)	3	0	4	0.01	-1	1	0.01
Regime (ii)	2.2	1.5	1	1	-1	1	2
Regime (iii)	3	-1.5	0.5	0.5	-1	1	0.1
Regime (iv)	4	2	1	1	-1	1	0.1
Regime (v)	2	-2	0.4	0.1	-0.5	1	0.1

The results in figure 13 show the optimal noise values $\sigma_u^{\mathrm{M}*}$ and the minimum model error at equilibrium, $\mathcal{P}(\pi_{att}, \pi_{att}^{\mathrm{M}*})$, as a function of the true additive noise level σ_u in the perfect system (56); the results for the Gaussian and non-Gaussian imperfect models are superimposed in each of the examined dynamical regimes (see figure 12) for easy comparison. The information barriers for the linear non-Gaussian model are prominent in the fat-tailed regime (ii) and the bimodal regime (v), while GCm has information barriers in the highly skewed regimes (iii)-(iv) where the linear non-Gaussian model performs marginally better. The relative entropies are computed numerically based on the analytical expressions for the equilibrium PDFs, since the Gaussian framework (7) is inappropriate in the highly skewed regimes.
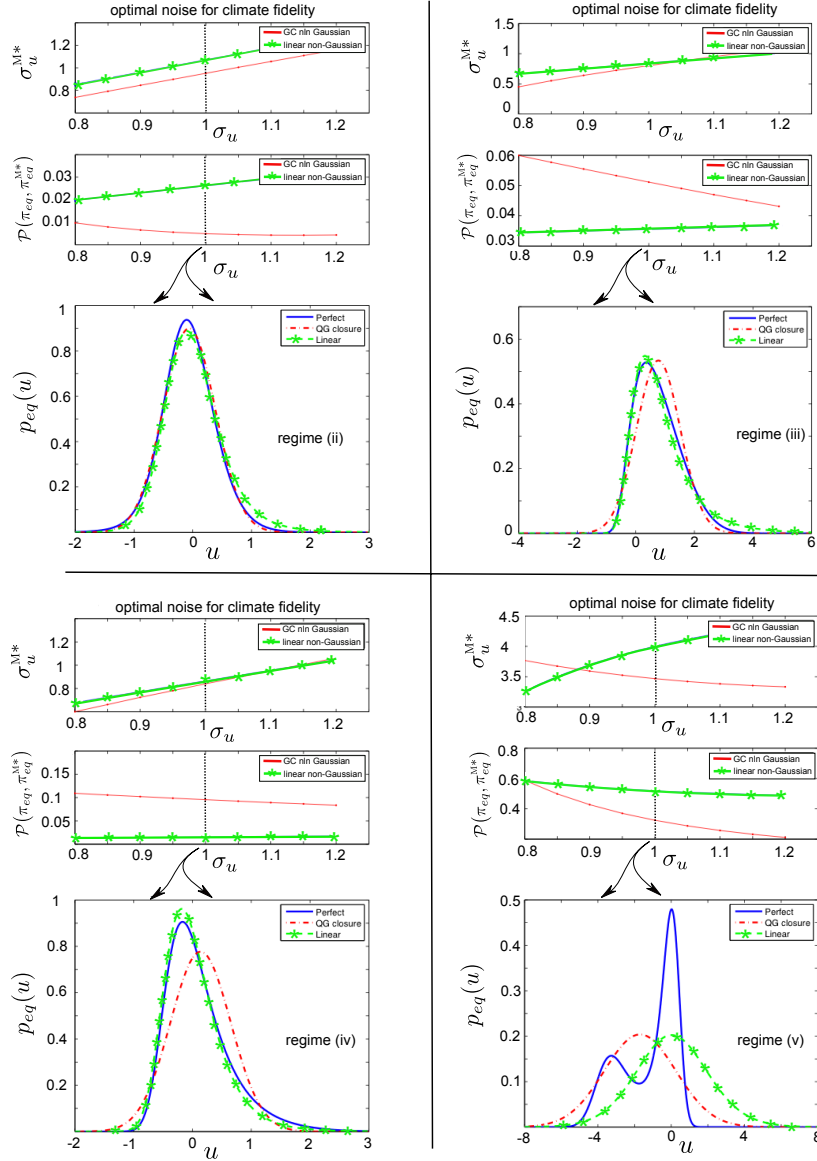
FIGURE 13. **Information-theoretic optimization of imperfect models of the cubic system (56)**. The four insets corresponding to different regimes of (56) show (from top to bottom) the optimal additive noise amplitude $\sigma_u^{\mathrm{M}*}$ for the attractor/climate fidelity (13) obtained via (12), the minimum model error at equilibrium, and the PDFs for the optimized imperfect models for $\sigma_u^{\mathrm{M}*}$ corresponding to $\sigma_u = 1$. Note the information barriers to the improvement of the linear non-Gaussian model in regimes (ii) and (v), and the barriers for GCm in the highly skewed regimes (iii), (iv).

Figure 14 illustrates the attractor/climate fidelity of the two imperfect models with time-periodic forcing in the phase space spanned by the mean and variance;
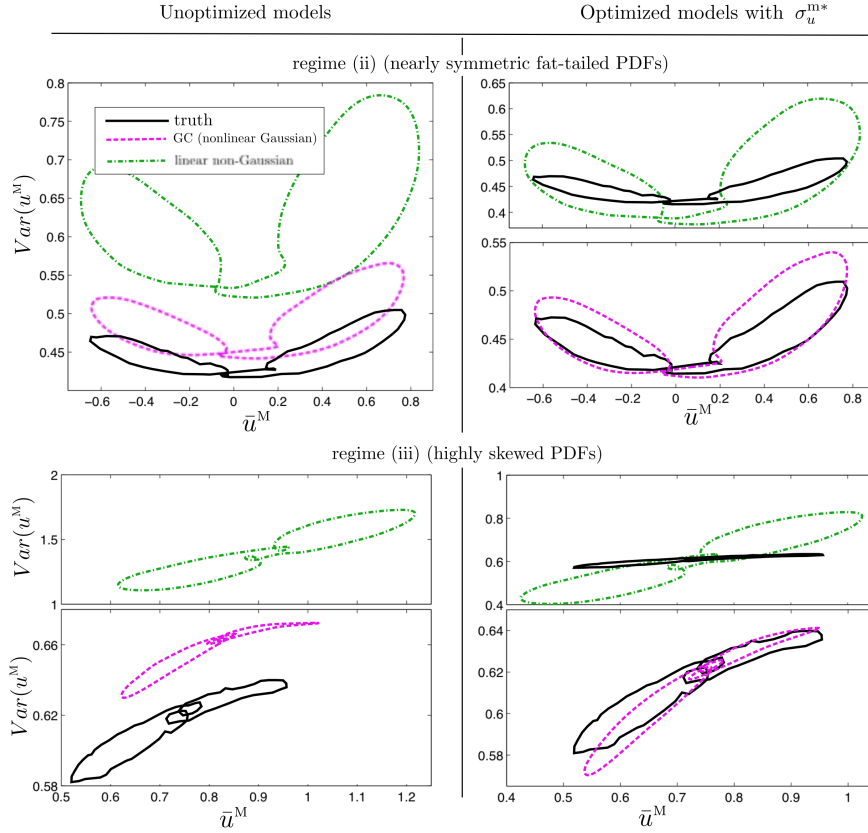
FIGURE 14. **Attractor fidelity of the imperfect models of the cubic system (56) with time-peroidic forcing.** Limit cycles in the second-order statistics of the system (56) and the imperfect models before and after the information-theoretic optimization (cf. §2). In the fat-tailed PDF regime (ii) the optimization is carried out via the additive noise inflation σ_u^{M}, while in the highly-skewed PDF regime (iii) the additive noise and the mean forcing are optimized for the minimum model error.

Forcing: $f(t)=f_0+f_1\sin(t+1)+f_2\sin(3t)$. *True parameters in reg. (ii)*: a=1.5, b=0.01, c=0.1, A=0, B=−1, σ_u=1, f_0=0.1, f_1=1, f_2=0.7, *Optimal model parameters*: GCm $\sigma_u^{\mathrm{M}*}$=0.96, lin. non-Gaussian $\sigma_u^{\mathrm{M}*}$=$A^{\mathrm{M}*}$=0.58. *Parameters in reg. (iii)*: a=2.2, b=1.5, c=1, A=1, B=−1, σ_u=1, f_0=2, f_1=.5, f_2=0.35, *Optimal model parameters*: GCm $\sigma_u^{\mathrm{M}*}$=0.85, $f_0^{\mathrm{M}*}$=1.7, lin. non-Gaussian $\sigma_u^{\mathrm{M}*}$=0.86, $f_0^{\mathrm{M}*}$=1.6.

the two insets, corresponding to the situation before and after the optimal noise inflation, show the limit cycles in the time-periodic second-order statistics for the perfect system and the imperfect models. The two examples shown correspond to the fat-tailed PDF regime (ii) and the highly skewed PDF regime (iii). In both cases GCm has a significantly better attractor/climate fidelity for the first two statistical moments even before the optimal noise inflation, although in the highly skewed PDF regimes the model error in the full PDF of GCm often exceeds that of the linear non-Gaussian model (see figure 13); in the nearly Gaussian regime (i) both models perform similarly well.
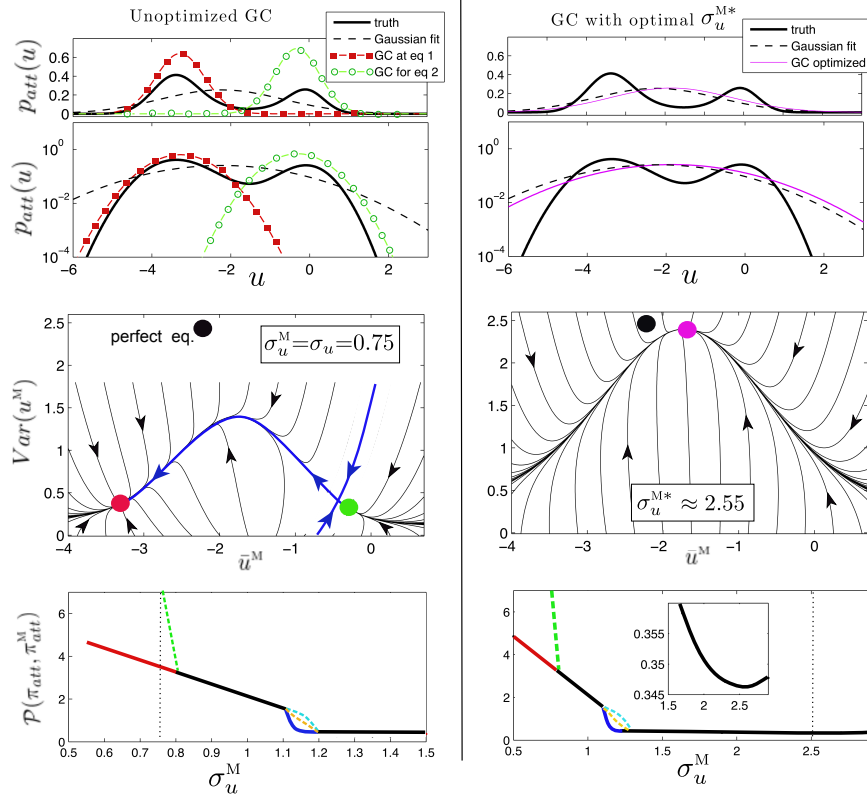
FIGURE 15. **Information-theoretic optimization of the Gaussian closure model of the cubic system (56) in the presence of multiple equilibria in the statistics of GCm**. The configuration shown corresponds to the bimodal PDF regime of the perfect system (56) and the two stable equilibria of GCm are shown in the left column. The information-theoretic optimization of GCm by the additive noise inflation leads to a configuration with a single equilibrium in GCm (right column) whose mean and variance (magenta dot) is close to the mean and variance of the perfect system (black dot). Note the existence of configurations with three stable equilibria in the statistics of GCm for $1.1 \leqslant \sigma_u^{\mathrm{M}} \leqslant 1.2$. *Perfect system parameters*: $a = 2$, $b = -2$, $c = 0.4$, $A = 0.6$, $B = -0.5$, $\sigma_u = 0.75$, $f = 0.1$.

Based on these instructive examples illustrated in figures 13-14, we make the following points:

- GCm has, in general, a better attractor/climate fidelity, including the time-dependent configurations with nontrivial attractors; GCm can also be better optimized for climate fidelity by additive stochastic noise inflation except for highly skewed regimes where the linear non-Gaussian model performs better.
- There exist barriers to imperfect model improvement which can be easily detected within the information-theoretic framework (§2).
- The information barriers for the linear non-Gaussian model are prominent in the fat-tailed regime (ii) and the bimodal regime (v), while GCm has information barriers in the highly skewed regimes (iii), (iv) (figure 13).
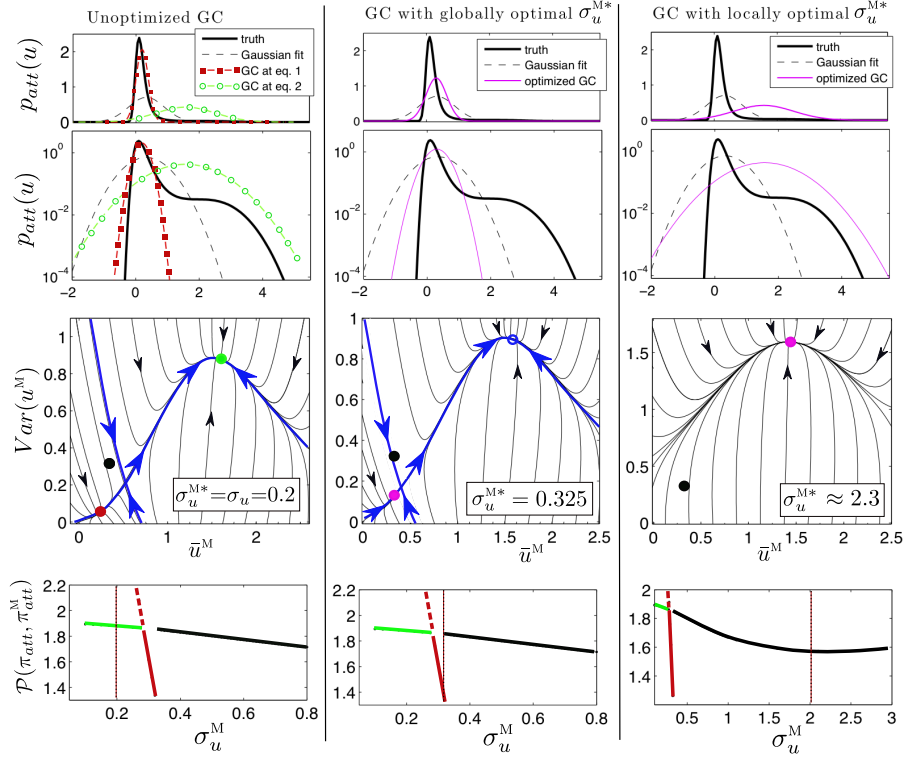
FIGURE 16. **Information-theoretic optimization of the Gaussian closure model of the cubic system (56) in the presence of multiple equilibria in the statistics of GCm**. The configuration shown corresponds to the highly-skewed PDF regime of the perfect system (56) and the two stable equilibria of GCm are shown in the left column. In this case the information-theoretic optimization of GCm by the additive noise inflation has two solutions: (1) global minimum (middle column) corresponding to the two-equilibria configuration in the statistics of GCm right before the disappearance of the optimal equilibrium in a saddle-node bifurcation, (2) local minimum (right column); note that the mean and variance of GCm in this case (magenta dot) is not close to the mean and variance (black dot) of the perfect system (56). *Perfect system parameters*: $a = 1.85$, $b = 2$, $c = 0.5$, $A = 0.1$, $B = -0.5$, $\sigma_u = 0.2$, $f = 0.23$.

FALSE MULTIPLE EQUILIBRIA IN IMPERFECT MODELS

Existence of false equilibria in imperfect models can have profound consequences for UQ in complex high-dimensional turbulent systems where such artifacts might not be easily identifiable. In figures 16 and 15 we show two distinct examples of such artifacts.

First, in figure 15 we show the configuration with two stable and one unstable equilibria in the Gaussian statistics of GCm when the perfect system (56) has a bimodal PDF (regime v). The information-theoretic optimization of §2 via the stochastic noise inflation lads to a one-equilibrium configuration for the statistics of GCm; note that the equilibrium mean and variance in the optimized GCm is
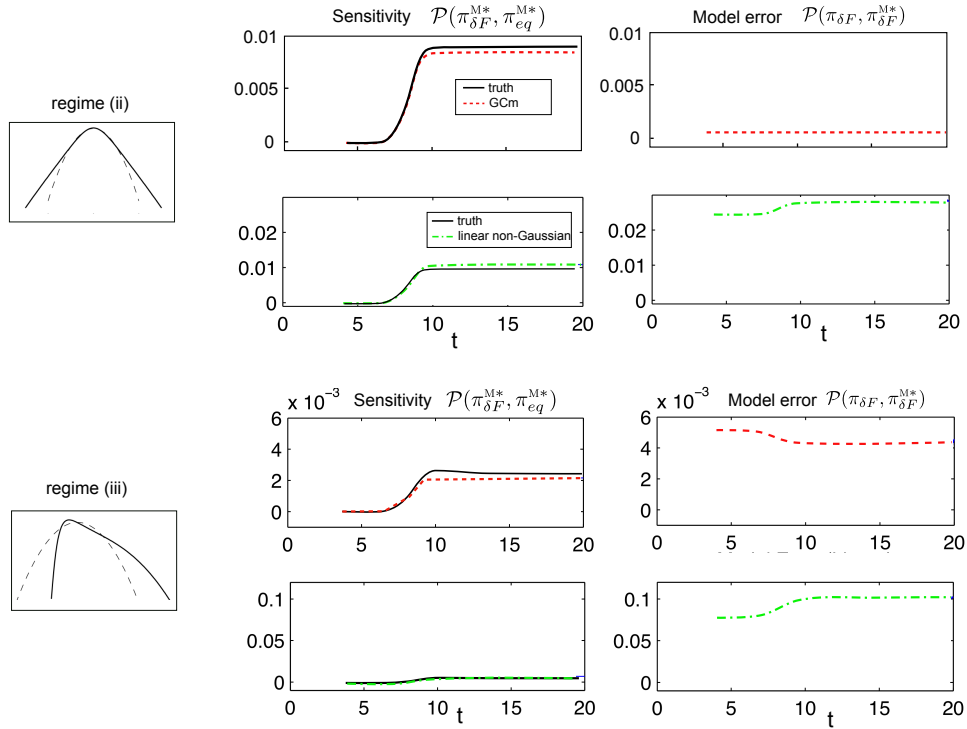
FIGURE 17. **Forced response of the cubic system (56) and the sensitivity** of the optimized imperfect nonlinear Gaussian and the linear non-Gaussian imperfect models to ramp-type perturbations δF (61) of the forcing. The results of tests in two dynamical regimes are shown: (left) the sensitivity (6) and model error (4) for the two imperfect models in the regime (ii) with fat-tailed algebraic tails, (right) the sensitivity and model error for the two imperfect models in the regime (iii) with highly skewed equilibrium PDFs (see figure 12). Note the poor skill of the linear non-Gaussian model, this is particularly important in the regime with highly skewed PDFs where the linear non-Gaussian model has better attractor/climate fidelity (figure 13).

close to the mean and variance of the perfect system (see black and magenta dots in figure 15).

A more complicated situation is illustrated in figure 16 showing the configuration with two stable equilibria in the Gaussian statistics of GCm when the perfect system (56) has a highly-skewed PDF (regime (iii)). In this case the globally optimal amplitude σ_u^{M} of the additive stochastic noise in GCm corresponds to the two-equilibrium configuration right before one of the stable equilibria disappears in the saddle-node bifurcation; note that the fat tail of the perfect model PDF is captured to some extent. The locally optimal configuration with one equilibrium corresponds to much larger noise values as shown in the right column of figure 16; note that in this case the single remaining equilibrium mean and variance of the optimized GCm is quite different from that of the perfect model (see black and magenta dots

in figure 16) with the maximum in the PDF of GCm close to the in the trough in the perfect model PDF.

This new phenomenon, arising as a consequence of model error induced by symmetrization of the model PDF, deserves the following comments:

- While the simple Gaussian imperfect models of nonlinear non-Gaussian systems can have high skill and climate fidelity when appropriately optimized, there exist dynamical regimes, usually associated with multimodality, where such imperfect unoptimized models display false multiple equilibria;
- Information-theoretic optimization (§2) is capable of mitigating these errors and improving the predictive skill even for the simple Gaussian closure models.

FORCED RESPONSE OF IMPERFECT MODELS WITH OPTIMAL NOISE

Here, we consider the sensitivity of optimized linear non-Gaussian (58) and nonlinear Gaussian (60) imperfect models to perturbations $\delta F(t)$ of the external forcing

$$\delta F(t) = \begin{cases} \frac{t-T_1}{T_2-T_1}\Delta F & \text{for} \quad T_1 \leqslant t \leqslant T_2, \\ 0 & \text{for} \quad t \leqslant T_1, t \geqslant T_2. \end{cases} \tag{61}$$

Consequently, we consider here long lead times so that essentially all knowledge of the initial conditions is lost and only the perturbations to the attractor of the perfect system are important. The imperfect models have the optimal additive stochastic forcing for attractort/climate fidelity according to (12). As before, the sensitivity (see Definition 2.3) of the perfect system and the imperfect models to such perturbations is quantified via the relative entropies, $\mathcal{P}(\pi_{\delta F}, \pi_{att})$, $\mathcal{P}(\pi_{\delta F}^{\text{M*}}, \pi_{att}^{\text{M*}})$. We illustrate the sensitivity of the optimized imperfect models in figure 17 in two intermittent regimes of the perfect cubic system (56); the left panel shows the sensitivity of the two imperfect models in the regime (ii) with symmetric fat-tailed PDFs while the right panel shows the results an analogous experiment in the regime (iii) with highly skewed PDF; for the sake of brevity we do not show the results in the nearly Gaussian regime (i), where the sensitivity is good for both models, and in the bimodal regime (v) where both models have poor sensitivity.

The following observations are worth pointing out in this context (see also figure 17):

- The optimized GCm with attractor/climate fidelity has good sensitivity in the nearly Gaussian and the fat algebraic tail regimes while the optimized linear non-Gaussian model (58) has an acceptable skill only in the nearly Gaussian regime.
- Despite the better attractor/climate fidelity in the highly skewed regimes (iii), (iv), the linear non-Gaussian model has a much worse forced response than GCm.
- In regimes where the sensitivity of both optimized imperfect models is poor, including the configuration (iv) with regime switching and the unimodal PDF (see figure 12), GCm still performs better than the linear non-Gaussian model (58).
- Unsurprisingly, the sensitivity of both optimized imperfect models is poor in the bimodal regime of the perfect model (not shown). In the nearly Gaussian regime (v) the sensitivity of both imperfect optimized models is good, as expected.

5. **Fundamental limitations of polynomial chaos expansions for UQ in systems with intermittency.** The theory of homogeneous chaos of Wiener [96], proposed initially for use in the statistical theory of turbulence [97] and widely discarded some thirty years later (e.g. [103, 15, 16]), has recently regained its popularity as a method for quantifying propagation of uncertainty in nonlinear dynamical systems. Practical implementations of the Wiener Chaos framework for UQ, commonly referred to as Polynomial Chaos Expansions (PCE), rely on a truncated spectral expansion of the system variables via a (truncated) Galerkin-type projection onto a space spanned by a fixed orthonormal polynomial basis in random variables (see, e.g., [36, 122, 125, 124, 123, 63, 43] for but a few examples). The PCE framework has many attractive features which are potentially well suited for numerical computations including the desirable separation of the random and deterministic components of the dynamics and 'realizability' of the resulting approximations as probability densities.

Here, following [10] we utilize a simple unambiguous test model, related to the system (38) with intermittency discussed already in §4.1, to show that methods exploiting truncated spectral expansions have fundamental limitations for UQ in systems with intermittency or parametric uncertainty in the damping and fat-tailed probability densities. These shortcoming are important even in the absence of explicit time dependence, or nontrivial attractors which is a well known problem for PCE. The model used to illustrate these limitations is given by a scalar with uncertain damping

$$\dot{u}(t) = -(\hat{\gamma} + \gamma(t))u(t) + f(t), \tag{62}$$

where $\hat{\gamma}$ is the mean damping and $\gamma(t)$ represents the damping fluctuations which are given by either a time-independent random variable in the parametric uncertainty case, or by a Gaussian Ornstein-Uhlenbeck process in the intermittent case, similarly to (38).

Intermittency and fat-tailed probability densities are abundant in the inertial and dissipation range of stochastic turbulence models (e.g., [19]) and we show that in such important dynamical regimes PCE performs, at best, similarly to the simple Gaussian moment closure technique utilized earlier in a different context for UQ within a framework of Empirical Information Theory [11]. The limitations of such truncated spectral expansions arise from:

- The finite truncation of the spectral expansions of either the solution process;
- Non-uniform convergence in time of PCE which leads to a rapidly increasing number of terms that are necessary to get a good approximate representation of the random process as time evolves;
- Constant flux of randomness due to white noise forcing and fundamental problems with capturing this flux via finite truncations of the spectral representation of the associated Wiener process.
- Slow decay of the PCE coefficients in the presence of intermittency which hampers implementation of sparse truncation methods widely used in elliptic problems or in low Reynolds number flows.

As discussed in detail in [10], the above issues lead to a number of fundamental limitations of the PCE approximations in systems with fat-tailed PDFs and intermittency; we illustrate these in §5.2 after a brief description of the PCE framework.

5.1. **Truncated polynomial chaos expansions.** Polynomial Chaos Expansion (PCE) of a solution u of an SDE (e.g., [102]) or SPDE (e.g., [108]) separates the deterministic effects from the randomness arising due to the nonlinear interactions with unresolved processes. This is achieved by expanding the system variables via a Galerkin-type projection of the model variables onto the space spanned by the fixed orthonormal polynomial basis in random variables. In the case of the so-called Wiener-Hermite expansion, the orthonormal polynomials are given by infinite products of normalized Hermite polynomials

$$T_\alpha(\boldsymbol{\xi}) = \prod_{i=1}^{\infty} H_{\alpha_i}(\xi_i), \tag{63}$$

where the multi-indices $\alpha \equiv \{\alpha_1, \alpha_2, \dots\}$ with integer entries satisfy $\sum_{i=1}^{\infty} \alpha_i < \infty$; the polynomials (63) are orthonormal with respect to the Gaussian measure generated by iid normal random variables $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots)$, $\xi_i \sim \mathcal{N}(0, 1)$, $\langle \xi_i \xi_j \rangle = \delta_{ij}$, which are used to represent the stochasticity of the solution process.

According to the Cameron-Martin theorem [13], any process with $\langle |u(t, \boldsymbol{\xi})|^2 \rangle < \infty$ has the following expansion in terms of the orthonormal Wick polynomials T_α (see Appendix C for more details),

$$u(t, \boldsymbol{\xi}) = \sum_\alpha u_\alpha(t) T_\alpha(\boldsymbol{\xi}), \quad u_\alpha(t) = \langle u(t, \boldsymbol{\xi}) T_\alpha(\boldsymbol{\xi}) \rangle. \tag{64}$$

While the convergence of the infinite expansion (64) is guaranteed by the Cameron-Martin theorem, it is however, not uniform in time and the expansion coefficients are related by an infinite-dimensional set of ODEs in an infinite number of unknowns. Thus, the doubly infinite expansion (64) is useless in applications unless a suitable truncation is employed. The need for finite truncations of these expansions required in applications leads to important limitations for UQ in systems with intermittency, as discussed in the following section. The simplest such truncation is obtained by retaining only the first K random Gaussian variables ξ_i and the Wick polynomials up to order N, leading to the truncated expression

$$u_{K,N}(t, \boldsymbol{\xi}) = \sum_{|\alpha| \leqslant N} u_\alpha(t) \prod_{i=1}^{K} H_{\alpha_i}(\xi_i), \tag{65}$$

so that the resulting approximation has altogether $\sum_{n=0}^{N} \binom{K+n-1}{n}$ terms; clearly the number of coefficients in the truncated expansion grows rapidly with increasing K and/or N. In order to reduce the number of expansion coefficients and the computational overhead, Luo [69, 70] and Hou et al. [49] proposed a sparse truncation method for reducing the number of PCE coefficients at a given truncation order. Similar sparse truncation methods were proposed for elliptic problems in [114, 25]. We will show in the next section (see also [10]) that such truncation methods are not adequate for turbulent systems with energy transfer on the attractor and/or intermittency due to a slow decay of amplitudes of the expansion coefficients.

5.2. **Intermittency, fat-tailed PDFs and shortcomings of truncated PCE.** We illustrate the limitations of the truncated PCE approximations in systems with fat-tailed PDFs and intermittency based on two instructive examples. The first example corresponds to stochastic dynamics of a scalar variable with the parametric uncertainty in the damping due to time-independent Gaussian fluctuations. This dynamics is described by an ODE with random initial values of the damping and
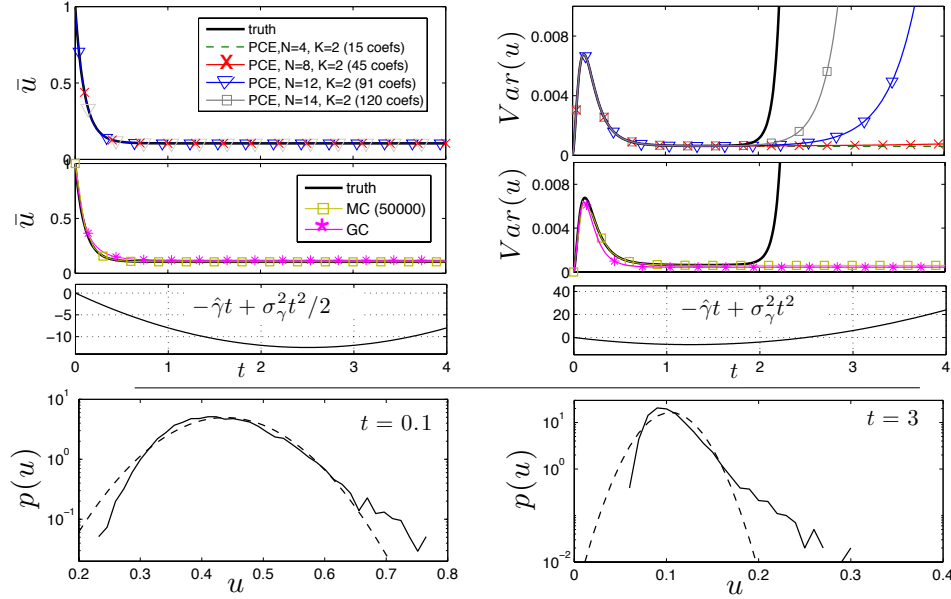
FIGURE 18. **Truncated PCE approximations for parametric uncertainty in damping fluctuations (66)**. Truncated PCE approximation of the second-order statistics for dynamics of u in (66) with parametric uncertainty. Here, the damping fluctuations are represented by a Gaussian random variable $\xi \sim \mathcal{N}(0,1)$. The bottom-line benchmark is given by the Gaussian moment closure (GC). Note that both GC and the Monte Carlo simulations (with 50000 runs) fail to detect the onset of unbounded growth of the variance (see also Appendix C). *System parameters: $\hat{\gamma}=10, \sigma_\gamma=2$, Initial conditions: $\langle u(0) \rangle=1, Var(u(0))=0$, Forcing: $f(t)=1$.*

can be easily shown to have fat-tailed PDFs (see figure 18 and [10]). In the second example the intermittent dynamics of $u(t)$ is driven by a purely deterministic forcing and stochastic damping fluctuations given by an OU process (figure 19). In both cases the truncated PCE approximations fail to capture the intermittency which results in grossly underestimated approximations of low-order statistics (see [10] for more details).

PCE APPROXIMATION OF DYNAMICS WITH UNCERTAIN DAMPING AND FAT-TAILED PDFs

This simple configuration is frequently used for testing PCE techniques for systems with simple parametric uncertainty in the damping fluctuations γ of (62) which leads to fat-tailed PDFs (see figure 18); this simple system is given by

$$\dot{u}(t) = -(\hat{\gamma} + \sigma_\gamma \xi)u(t) + f(t), \qquad \xi \sim \mathcal{N}(0,1). \tag{66}$$

Contrary to the second case discussed below, the statistically exactly solvable dynamics in this configuration is non-mixing and the autocorrelation of the damping fluctuations is constant and equal to σ_γ^2. Moreover, it can be easily shown that the mean and variance grow unboundedly in time, despite a possible initial metastable phase (see [10] for details).

The Cameron-Martin theorem [13] rigorously guarantees that the solutions of (66) admit the PCE expansion (64) on any finite time interval and the truncated equations for the PCE coefficients of the system (62) are

$$\dot{u}_\alpha = -\hat{\gamma}(t)u_\alpha - \sum_{p\in\mathcal{J}_{K,N}} \sum_{\substack{0\leqslant\beta\leqslant\alpha \\ \beta+p=\alpha^1}} C(\alpha,\beta,p)u_{\alpha-\beta+p}\gamma_{\beta+p} + f(t)\delta_{\alpha\mathbf{0}}, \qquad (67)$$

where $\alpha^1 = (1,0,0,\dots)$ and the PCE coefficients for the time-independent damping fluctuations are given by $\gamma_{\alpha^1} = \sigma_\gamma$ and $\gamma_\alpha = 0$ for $\alpha \neq \alpha^1$. In deriving (67) we utilized the orthonormality of the Wick polynomials T_α together with the important identity which allows for expressing the nonlinear terms in (62) via infinite expansions in the Wick polynomials (see [95, 69, 49, 70] and Theorem 2 of Appendix C)

$$u\,\gamma = \sum_{\alpha\in\mathcal{J}} \Big(\sum_{p\in\mathcal{J}} \sum_{0\leqslant\beta\leqslant\alpha} C(\alpha,\beta,p)u_{\alpha-\beta+p}\gamma_{\beta+p} \Big) T_\alpha. \qquad (68)$$

Here, we additionally require that

$$\langle u(t_0)\rangle = 1, \quad Var\big(u(t_0)\big) = 0, \quad \langle u(t_0)\gamma\rangle = 0, \qquad (69)$$

which leads to the following initial conditions in (67):

$$u_{\alpha^0}(t_0) = 1, \quad \text{and} \quad u_\alpha(t_0) = 0 \quad \text{for} \quad \alpha \neq \alpha^0, \qquad (70)$$

where $\alpha^0 = (0,0,\dots)$.

Examples of truncated PCE approximations for this system are shown in figure 18 which we use to illustrate the following shortcomings of PCE in this simple configuration with parametric uncertainty:

- Despite a good performance of truncated PCE approximations for sufficiently short times, even very high-order truncations fail to reproduce the inevitable unbounded growth of the statistics at later times; see figure 18. Thus, the good accuracy of the PCE approximations at short times can be misleading.
- The unbounded growth of the statistics remains essentially undetected below certain truncation order which is parameter dependent but often high (in figure 18 the threshold is approx. $N = 9$); this shortcoming is also associated with the Gaussian Closure or Monte Carlo simulations.

PCE APPROXIMATIONS IN SYSTEMS WITH INTERMITTENT TRANSIENT INSTABILITIES

Here, the intermittency in the dynamics of u is induced by the stochastic damping fluctuations γ given by an Ornstein-Uhlenbeck process; the corresponding system is given by

$$\begin{aligned} a) \quad \dot{u}(t) &= -(\hat{\gamma} + \gamma(t))u(t) + f(t), \\ b) \quad \dot{\gamma}(t) &= -d_\gamma\gamma + \sigma_\gamma\dot{W}_\gamma(t). \end{aligned} \qquad (71)$$

where $\hat{\gamma}$, d_γ re the mean damping parameters, f is deterministic forcing, and $\sigma_\gamma\dot{W}_\gamma(t)$ represents stochastic white noise forcing defined for all time with amplitude σ_γ.

Based on the Cameron-Martin theorem [13] the solutions of (66) admit the PCE expansion (64) with the truncated equations for the PCE coefficients of the system
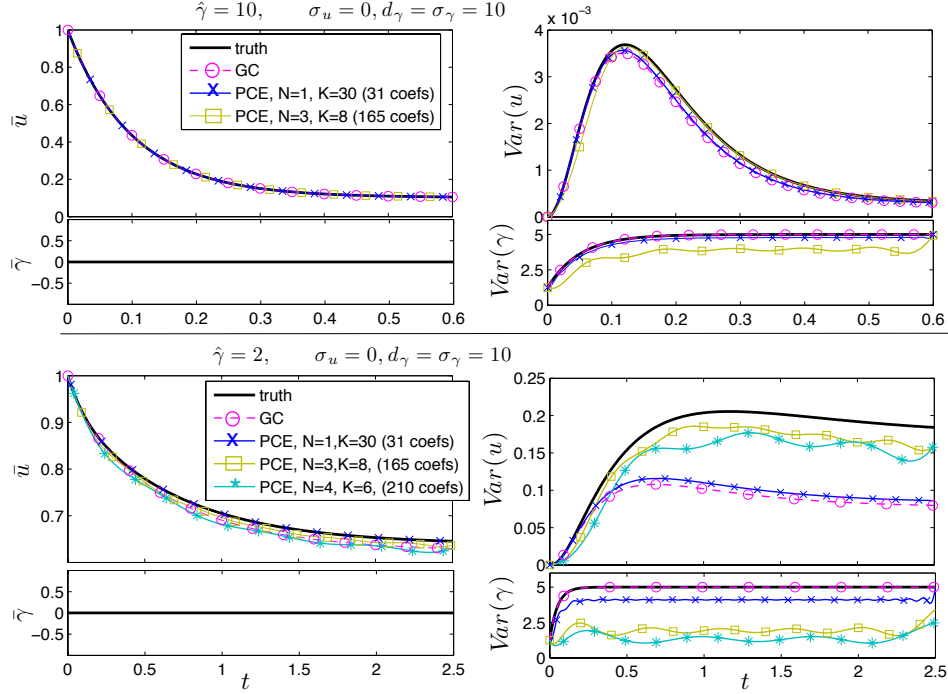
FIGURE 19. **Truncated PCE approximations of intermittent dynamics of a scalar $u(t)$ (71) with stochastic damping fluctuations** and no white noise in the resolved component. The PCE approximation is illustrated for two different damping parameters $\hat{\gamma}$ corresponding to nearly Gaussian dynamics (top) and intermittent dynamics (bottom) with a fat-tailed equilibrium PDF. The bottom line benchmark given by the Gaussian moment closure (GC). *Initial conditions:* $\langle u_0 \rangle = 1, Var(u_0) = 0, Var(\gamma_0) = \frac{1}{2} Var_{eq}(\gamma), \langle u_0 \gamma_0 \rangle = 0$ *Forcing:* $f(t) = 0$.

(71) given by

$$
\begin{aligned}
a) \quad \dot{u}_\alpha &= -\hat{\gamma} u_\alpha - \sum_{p \in \mathcal{J}_{K,N}} \sum_{0 \leqslant \beta \leqslant \alpha} C(\alpha, \beta, p) u_{\alpha - \beta + p} \gamma_{\beta + p} + f(t) \delta_{\alpha \mathbf{0}}, \\
b) \quad \dot{\gamma}_\alpha &= -d_\gamma \gamma_\alpha + \sigma_\gamma \sum_{i=1}^{K} I_{\{\alpha_j = \delta_{ij}\}} m_i(t),
\end{aligned}
\tag{72}
$$

where for fixed i, $I_{\alpha_j = \delta_{ij}}$ equals 1 if $\alpha_j = \delta_{ij}$ and equals zero otherwise; the orthonormal basis functions $m_i(t)$ are used in the truncated spectral representation of the Wiener process $W_\gamma(t)$ (see [49, 69, 10] and the Appendix C). In deriving (72) we utilized the orthonormality of the Wick polynomials T_α and the identity (68). The white-noise forcing in the above truncated deterministic system of ODEs for the coefficients u_α, γ_α is represented by the K basis functions $m_i(t)$. Note that truncating (68) results in neglecting, at least partially, the effects of small scale fluxes on the 'resolved' dynamics $u(t)$ which is particularly important in the presence of intermittency and non-vanishing energy transfer between Fourier modes of u at equilibrium.

Here, we assume that

$$\langle u_0 \rangle = 1, \quad Var(u_0) = 0, \quad Var(\gamma_0) = \frac{1}{2}Var_{eq}(\gamma), \quad \langle u_0\gamma_0 \rangle = 0, \qquad (73)$$

and we also require that the white noise $\dot{W}_\gamma$ is uncorrelated with the initial conditions, i.e.,

$$\langle u_0\dot{W}(t_0) \rangle = 0, \quad \langle \gamma_0\dot{W}(t_0) \rangle = 0. \qquad (74)$$

The constraints (73)-(74) lead to the following initial conditions

$$u_{\alpha^0} = 1, \quad u_{\alpha^1} = 0, \quad u_{\alpha^2} = \sqrt{\sigma_\gamma^2/4d_\gamma}, \qquad u_\alpha = 0 \quad \text{for} \quad \alpha \neq \alpha^0, \alpha^2, \qquad (75)$$

and the truncated representation of the Wiener process

$$W_\gamma(t) = \sum_{i=3}^{K} \xi_i \int_{t_0}^{t} m_{i-2}(\tau)\mathrm{d}\tau. \qquad (76)$$

Limitations of the truncated PCE method for solving white noise driven dynamics were noted in [49] where reliable integration of the uncertain dynamics required an unreasonably large dimension of the random vector $\boldsymbol{\xi}$, except for sufficiently short times. The examples of PCE approximations of the second-order statistics of the system (71) show that the performance of this method is diminished further for UQ in systems with intermittency driven by unresolved, white noise processes. The following remarks are in place based on the examples shown in figure 19:

- Finite truncations of the PCE approximations of the dynamics driven by white noise require a compromise between the order of truncation N and the number K of random variables used in the spectral approximation of the Wiener process.
- Truncated PCE approximation performs well in nearly Gaussian regimes with strong mean damping provided that the white noise driving the evolution of damping fluctuations γ is sufficiently well resolved; however, in such cases the Gaussian closure performs similarly well and is more efficient.
- PCE performs poorly in intermittent regimes, except for short times (figure 19 with weak mean damping $\hat{\gamma}$), when both the high order of approximation and a good spectral resolution of the the white noise forcing are required (see also [49]). In the intermittent example shown in figure 19 long integration is needed when the statistics of u is of interest and the number of expansion coefficients necessary for accurate approximation grows to unacceptable levels.
- Sparse truncation methods in the intermittent regimes driven by white noise are hampered by the slow decay of expansion coefficients.
- For low order approximation ($N \sim 1$) with sufficiently large number of random variables $K \sim 50$ only short-time PCE approximation acceptable
- For higher order truncations $N \sim 4$ longer time approximation improves but there are insufficient number of random variables, realistically $K \sim 6$, to obtain a decent approximation at either short or long times

6. **Turbulent diffusion: New exactly solvable test models for UQ in complex high-dimensional systems.** One of the important paradigm models for the behavior of turbulent systems [7, 71] and the associated uncertainty quantification involves a passive tracer $T(\boldsymbol{x},t)$ which is advected by a velocity field $\boldsymbol{v}(\boldsymbol{x},t)$ with dynamics given by

$$\partial_t T + \boldsymbol{v} \cdot \nabla T = \kappa\Delta T, \qquad (77)$$

where $\kappa > 0$ is molecular diffusion and the velocity field $\boldsymbol{v}$ is incompressible, $\nabla \cdot \boldsymbol{v} = 0$. When $\boldsymbol{v}(\boldsymbol{x}, t)$ is a turbulent velocity field, the statistical properties of solutions of (79) such as their large scale effective diffusivity, energy spectrum, and probability density function (PDF) are all important in applications. These range from, for example, the spread of pollutants or hazardous plumes in environmental science to the behavior of anthropogenic and natural tracers in climate change science [26, 27, 99], to detailed mixing properties in engineering problems such as non-premixed turbulent combustion [107, 105, 8]; some more specific important topics include:

- Transitions between Gaussian and fat tailed highly intermittent PDFs for the tracer in laboratory experiments as the Peclet number varies with a mean background gradient for the tracer [40, 50, 51],
- Fat tail PDFs for anthropogenic and natural tracers with highly intermittent exponential tails in observations of the present climate [99],
- The nature of the sustained turbulent spectrum for scalar variance with a background gradient for the tracer [116],
- Eddy diffusivity approximations for tracers in engineering and climate change science [26, 27, 79].

For turbulent random velocity fields, the passive tracer models in (79) also serve as simpler prototype test problems for closure theories for the Navier Stokes equations since (79) is a linear equation but is statistically nonlinear [58, 59, 5, 6, 90, 91, 60, 61, 115, 71].

Recent advances addressing the above issues [80, 92] and complementing the earlier work mentioned above utilized elementary stochastic models for turbulent diffusion of tracers with a background mean gradient

$$T(\boldsymbol{x}, t) = T'(\boldsymbol{x}, t) + \alpha \boldsymbol{x}, \tag{78}$$

in (79) and which led to new physically important intermittent regime with complex statistical features mimicking crucial aspects of laboratory experiments and atmospheric observations. These statistical features include the transition to intermittent scalar probability density functions with fat exponential tails as certain variances of the advecting mean velocity are increased while satisfying important physical constraints, and exact formulas for tracer eddy diffusivity which is nonlocal in space and time, as well as exact formulas and simple numerics for the tracer variance spectrum in a statistical steady state [92]. The recent use of such simple models with complex statistics as unambiguous test models for central contemporary issues in both climate change science and the real time filtering of turbulent tracers from sparse noisy observations and other important applications [35].

In this section we utilize a suite of Gaussian and non-Gaussian turbulent tracer models to discuss and illustrate a number of issues encountered in many other turbulent high-dimensional spatially extended systems; similarly to the previous sections, the emphasis is on the role of intermittency with nontrivial mean-turbulent flux interactions in UQ applications, and methods for mitigating imperfect model deficiencies through the use of the information-theoretic framework discussed in §2. The approach we assume for the imperfect models is motivated by the contemporary AOS numerical models which have too much dissipation in the model velocity field and employ various parameterizations of the unresolved processes, including eddy diffusivity parameterization. Thus, the imperfect models, which are discussed in

more detail below, have the following general structure

$$\partial_t T^{\mathrm{M}} + \boldsymbol{v}^{\mathrm{M}} \cdot \nabla T^{\mathrm{M}} = \kappa_{\mathrm{M}} \Delta T^{\mathrm{M}} + \sigma(x,t) \dot{W}(t), \qquad (79)$$

where $\boldsymbol{v}^{\mathrm{M}}$ is the coarse-grained model velocity, κ_{M} is the model eddy diffusivity for the tracer, and $\sigma(x,t)\dot{W}(t)$, $t \in I\!\!R$, is the judiciously introduced stochastic forcing which can be spatially correlated; we will show based on detailed examples how the stochastic forcing can improve the climate fidelity and prediction skill for various coarse-grainings in the imperfect models with too much dissipation in this spatially extended case.

Important new issues for UQ discussed here (in addition to those discussed in §3 and §4 for the single Fourier mode) are:

- The interplay between coarse-graining and uncertainty in imperfect models,
- Consequences of overdamped/underdamped turbulent velocity spectra on the predictive skill of spatially extended models,
- Mean-wave interactions, eddy diffusivity parameterization, and mitigation of model error via stochastic forcing,
- Information-theoretic optimization and information barriers in spatially extended systems.

6.1. **New exactly solvable test models for turbulent diffusion with hidden instabilities and intermittency.** Here, we utilize instructive models for turbulent diffusion of a scalar tracer with nontrivial eddy-diffusivity, variance spectrum, and intermittent non-Gaussian statistics like tracers in the atmosphere [26] to provide a highly nontrivial demonstration of improving the fidelity of imperfect models through stochastic forcing. The two classes of statistically exactly solvable models introduced below build on the models discussed earlier for the single Fourier mode in §4; the resulting spatially extended models generalize all the models discussed earlier in [79, 35, 80, 81, 92, 32] by incorporating additional intermittency due to (i) transient instabilities in the dynamics of the shear flow in the turbulent velocity field driving the tracer dynamics and/or (ii) intermittent fluctuations in the concentration of a reactive scalar with large scale instabilities.

Next, we present a geophysical interpretation of these models although versions with engineering applications can be easily set-up [35]. Both of these systems have a zonal (east-west) turbulent jet, $U(t)$, a family of planetary and synoptic scale waves with north-south velocity $v(x,t)$ with x, a spatially periodic variable representing a fixed mid-latitude circle in the east-west direction, and tracer fluctuations $T'(x,t)$ with a north-south mean gradient α; the governing equations for these two models can be written compactly as

$$a) \quad \frac{\partial U}{\partial t} = -d_U U + f(t) + \sigma_U \dot{W}_U(t),$$

$$b) \quad \frac{\partial v}{\partial t} = P\left(\frac{\partial}{\partial x}, U(t)\right) v + f_v(t) + b_v(x,t) + \sigma_v(x) \dot{W}_v(t), \qquad (80)$$

$$c) \quad \frac{\partial T'}{\partial t} = Q\left(\frac{\partial}{\partial x}, U(t)\right) T' - \alpha v(x,t) + \kappa \frac{\partial^2 T}{\partial x^2},$$

where $f(t)$, $f_v(x,t)$ are known deterministic functions, $b_v(x,t)$ represents a stochastic finitely correlated uncertainty in the forcing of the one-dimensional shear v, while $\dot{W}_U$, $\dot{W}_v$ represent two independent white noise fluctuations defined for all

time. The zonal jet $U(t) = \bar{U}(t) + U'(t)$, where $\bar{U}(t) > 0$ is the climatological periodic mean which is strongly eastward while the random fluctuations, $U'(t)$, have a standard deviation consistent with such eastward dynamical behavior. The above system is solved by Fourier series with independent scalar complex variable versions of (80) for each different wave number k; in Fourier space the pseudo-differential operators P and Q have the symbols $\hat{P}_k = -\gamma_{v_k} + i\omega_{v_k}$ and $\hat{Q}_k = -\gamma_{T_k} + i\omega_{T_k}$ respectively, with $\gamma_{v_k}, \gamma_{T_k}, \omega_{v_k}, \omega_{T_k}$ known functions of time which can be given by stochastic processes, as discussed below.

NON-GAUSSIAN MODELS WITH INTERMITTENT INSTABILITIES IN THE VELOCITY FIELD

This non-Gaussian statistically exactly solvable system for the turbulent tracer with the mean gradient α is driven by a correlated turbulent velocity field with intermittency due to transient instabilities in the turbulent shear $v(x,t)$. This model is obtained by adopting

$$\hat{P}_k = -\gamma_{v_k}(t) + i\omega_{v_k}(t), \quad \text{and} \quad \hat{Q}_k = -d_{T_k} + i\omega_{T_k}(t),$$

for the symbols of P and Q in (80) with each $\gamma_{v_k}(t)$ given by a Gaussian stochastic process and $\omega_{v_k}, \omega_{T_k}$ prescribed functions of wavenumber k. For the passive tracer mode T'_k in correlated velocity field with turbulent Rossby waves in the shear, we have

$$\omega_{v_k}(t) = a_k U(t) + b_k, \qquad \omega_{T_k}(t) = -kU(t), \tag{81}$$

with the values of the real coefficients a_k, b_k depending on the physical situation considered, as in [92]. In particular, for uncorrelated baroclinic Rossby waves $a_k \equiv 0$ and $b_k \propto k/(k^2 + F_s)$ where and F_s is the stratification; this configuration is considered in one of the examples below.

The equations for the k-th Fourier mode of the tracer fluctuations T'_k and the shear flow v_k with a spatially uniform zonal flow U are

$$
\begin{array}{lll}
a) & \dot{T}'_k(t) & = (-d_{T_k} + i\omega_{T_k}(t))T'_k(t) - \alpha v_k(t), \\[2ex]
b) & \dot{U}(t) & = -d_U U(t) + f_U(t) + \sigma_U \dot{W}_U(t), \\[2ex]
c) & \dot{v}_k(t) & = (-d_{v_k} - \gamma_{v_k}(t) + i\omega_{v_k}(t))v_k(t) + b_{v_k}(t) + f_{v_k}(t) + \sigma_{v_k}\dot{W}_{v_k}(t), \\[2ex]
d) & \dot{\gamma}_{v_k}(t) & = -d_{\gamma_{v_k}}\gamma_{v_k}(t) + \sigma_{\gamma_{v_k}}\dot{W}_{\gamma_{v_k}}(t), \\[2ex]
e) & \dot{b}_{v_k}(t) & = (-d_{b_{v_k}} + i\omega_{b_{v_k}})b_{v_k}(t) + \sigma_{b_{v_k}}\dot{W}_{b_{v_k}}(t),
\end{array}
\tag{82}
$$

where γ_{v_k} and b_{v_k} are, respectively, the damping fluctuations and the forcing fluctuations in the k-th mode of the turbulent shear flow v_k. This statistically exactly solvable system (see Appendix D for the analytical solutions) reduces to the model studied in [79, 35, 80, 81, 92, 32] for $\gamma_{v_k} \equiv 0, b_{v_k} \equiv 0$, i.e., in the absence of stochastic fluctuations in both the damping and forcing in the dynamics of the Fourier modes of the v_k.

There exist two important and distinct sources of tracer intermittency in this model, both of them occur for a strongly eastward mean jet $\bar{U}$. First, similarly to the simpler model with no transient instabilities in the velocity, i.e., $\gamma_{v_k} \equiv 0$, the intermittent tracer behavior occurs for a turbulent mean jet $U(t) \neq \bar{U}(t)$, and finitely correlated in time turbulent shear waves v in the presence of the large scale

tracer gradient α (see [80] for details). The second source of intermittency arises due to the transient instabilities in the shear waves v for nontrivial stochastic solutions in (82) and a sufficiently weak mean damping d_{v_k} (see [9, 11]). Combination of these two sources of intermittency with exact statistical solvability makes the system (82) a very attractive test model for UQ and data assimilation with judicious model error in realistic situations with the hidden instabilities due to γ_{v_k} mimicking, for example, geophysical boundary layers.

NON-GAUSSIAN MODELS FOR REACTIVE TRACER WITH INTERMITTENT INSTABILITIES IN THE VELOCITY FIELD AND IN THE TRACER

This non-Gaussian, statistically exactly solvable model for a reactive turbulent tracer with intermittent large scale tracer instabilities and a correlated velocity field with intermittent instabilities in the turbulent shear flow is obtained by setting

$$\hat{P}_k = -\gamma_{v_k}(t) + i\omega_{v_k}(t) \quad \text{and} \quad \hat{Q}_k = -d_{T_k} - r\gamma(t) + i\omega_{T_k}(t),$$

for the symbols of P and Q in (80) with the dispersion relations ω_{v_k}, ω_{T_k} given by the general formulas (81). The Fourier modes of the turbulent tracer and velocity in this model satisfy

$$
\begin{aligned}
a) \quad & \dot{T}'_k(t) && = (-d_{T_k} - r\gamma(t) + i\omega_{T_k}(t))T'_k(t) - \alpha v_k(t), \\[4pt]
b) \quad & \dot{\gamma}(t) && = (-d_\gamma + i\omega_\gamma)\gamma(t) + \sigma_\gamma \dot{W}_\gamma(t), \\[8pt]
c) \quad & \dot{U}(t) && = -d_U U(t) + f_U(t) + \sigma_U \dot{W}_U(t), \\[4pt]
d) \quad & \dot{v}_k(t) && = (-d_{v_k} - \gamma_{v_k}(t) + i\omega_{v_k}(t))v_k(t) + b_{v_k}(t) + f_{v_k}(t) + \sigma_{v_k}\dot{W}_{v_k}(t), \\[4pt]
e) \quad & \dot{\gamma}_{v_k}(t) && = -d_{\gamma_{v_k}}\gamma_{v_k}(t) + \sigma_{\gamma_{v_k}}\dot{W}_{\gamma_{v_k}}(t), \\[4pt]
f) \quad & \dot{b}_{v_k}(t) && = (-d_{b_{v_k}} + i\omega_{b_{v_k}})b_{v_k}(t) + \sigma_{b_{v_k}}\dot{W}_{b_{v_k}}(t),
\end{aligned}
$$

$$\tag{83}$$

where γ_{v_k} and b_{v_k} are, respectively, the damping fluctuations and the forcing fluctuations in the k-th mode of the turbulent shear flow v_k, and $\gamma(t)$ represents the spatially uniform damping fluctuations in the k-th mode of the tracer fluctuations T_k.

In addition to the two physically important sources of intermittency outlined in the previous model, the system (83) contains a third independent source of tracer intermittency due to the large scale reactive events in the tracer occurring at the rate r in (83a). The statistical exact solvability of this model (see Appendix D) makes it potentially very useful in realistic atmospheric applications with externally driven intermittent large scale tracer depletion and production.

Below we provide a highly nontrivial demonstration of improving the fidelity of imperfect models through information-theoretic optimization by inflating the stochastic forcing in the imperfect models.

6.2. Optimization of coarse-grained models for improved prediction skill and forced response. For spatially extended systems, the issue of coarse-graining becomes important in the context of model climate fidelity and sensitivity. Here, we discuss how coarse-graining is reflected in the quantification of model error using information metrics discussed in §2. This issue is compounded further by the
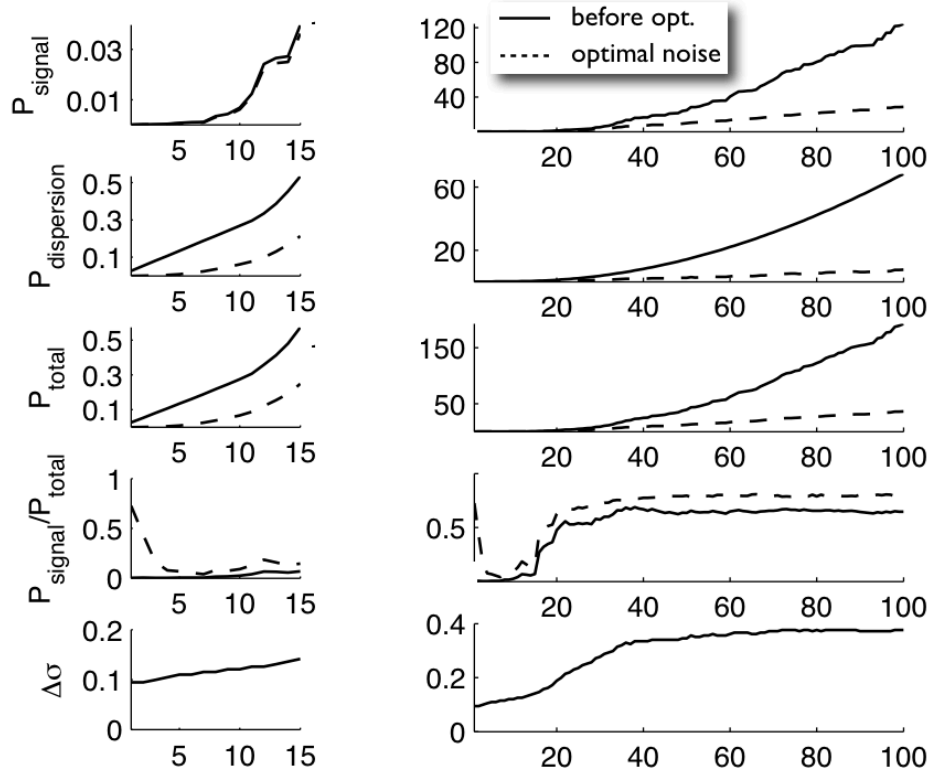
FIGURE 20. **Information-theoretic optimization of turbulent Gaussian Rossby waves**. Rows 1-4: Relative entropy of the imperfect model with too much damping before optimizing noise (solid line), $\mathcal{P}(\pi, \pi^{\mathrm{M}})$, and after optimizing noise (dashed line), $\mathcal{P}(\pi, \pi^{\mathrm{M}*})$, as a function of coarse-graining. Left side corresponds to coarse-graining up to 15 modes, right side shows coarse-grainings up to 100 modes. Note different scales of y-axis on the left and right sides. Row 5: Additional noise $\Delta\sigma$ as a function of coarse-graining obtained by minimizing the lack of information.

presence of overdamped velocity spectra and eddy diffusivity parameterization in the imperfect models discussed below.

The first example [32] deals only with the turbulent shear flow which is assumed linear and Gaussian and we show that the uncertainty in the imperfect models grows with the number of Fourier modes considered (i.e., the coarse-graining); the information-theoretic optimization (see §2) by stochastic noise inflation greatly reduces the uncertainty for various coarse-grainings. The second example discusses optimization the Gaussian tracer model [32, 81] which can be obtained from the more general model (82) as briefly explained below. In this more complex case, involving the effects of coarse-graining, eddy diffusivity and overdamped velocity spectra, a significant gain can be achieved using the stochastic-statistical optimization advocated in §2.

Optimization of coarse-grained Gaussian models for turbulent velocity

Here, as a simple illustration, we consider a linear Gaussian model for the turbulent shear flow given for each mode v_k by (82c) with $\gamma_{v_k} = \beta_{v_k} \equiv 0$, constant damping and noise amplitude d_{v_k}, σ_{v_k} and constant forcing f_{v_k}. The perfect model has $K = 100$ modes and the energy spectrum

$$E_k = \frac{\sigma_{v_k}^2}{2d_{v_k}} = \begin{cases} \mathcal{E}, & \text{for} \quad k \leqslant K_0 \\ \mathcal{E}\left(\frac{k}{K_0}\right)^{-\theta} \end{cases}, \tag{84}$$

with $\mathcal{E} = 1, K_0 = 10$ and $\theta = 3$ which is appropriate for large scale turbulence in the atmosphere; the forcing is chosen to be roughly twice the standard deviation of v_k, i.e., $|f_{v_k}|^2 = 2\sigma_{v_k}(d_{v_k}^2 + \omega_{v_k}^2)/d_{v_k}$.

Most models for turbulence in spatially extended systems have too much dissipation (Palmer 2001) due to inadequate resolution and deterministic parameterization of unresolved features involving wave-mean flow interaction and stochastic backscatter. We mimic this overdamping in the imperfect models for the turbulent velocity by imposing $E_k > E_k^{\mathrm{M}}$ through $\mathcal{E}^{\mathrm{M}} = 0.8$ and $\theta^{\mathrm{M}} = 3.5$ in (84).

The information-theoretic optimization discussed in §2 of the overdamped model for the turbulent shear can be carried out by inflating the stochastic forcing in order to maximize the climate fidelity (8); the results before and after such an optimization procedure are shown in figure 20 where, for simplicity, the noise amplitude is the same for all modes (see [32] for details). The first four rows illustrate the uncertainty content in the signal and dispersion components of the total uncertainty for different coarse-grainings, and the bottom row shows the difference between the optimal and true noise amplitudes for different coarse-grainings.

We note the following facts based on the example shown in figure 20:

- The uncertainty in unoptimized models grows with increasing resolution.
- The model improvement is significant in both the signal and dispersion components of the total uncertainty even when the noise amplitude $\sigma_k = \sigma$ is assumed the same for all modes.
- The difference between the true and optimal noise amplitudes increases with the number of modes in the coarse graining.

Optimization of coarse-grained Gaussian models for turbulent Gaussian or non-Gaussian tracer

This example is based on the model for the turbulent tracer discussed in [81, 32] which can be obtained from (82) by disregarding the transient instabilities in the uncorrelated velocity field $(U(t), v(x,t))$, i.e., we set $\gamma_{v_k} = b_{v_k} \equiv 0$ in (82); moreover we set $\omega_{v_k} = \beta k/(k^2 + F_s)$ corresponding to the dispersion relation of baroclinic Rossby waves where β is the north-south gradient of rotation, and F_s is the stratification. The perfect system is non-Gaussian for $\sigma_U \neq 0$ in (82a) and it is Gaussian otherwise.

Here, the imperfect models are Gaussian with the same dynamics for the zonal jet and Rossby waves and coarse-grained components $(U^{\mathrm{M}}(t), v^{\mathrm{M}}(x,t))$ and the tracer equation is given by

$$\partial_t T'^{\mathrm{M}} + \bar{U}^{\mathrm{M}}(t)\partial_x T'^{\mathrm{M}} = -\alpha v^{\mathrm{M}}(x,t) + \kappa_{\mathrm{M}}\partial_x^2 T'^{\mathrm{M}} + \sigma_T \dot{W}(x,t), \tag{85}$$
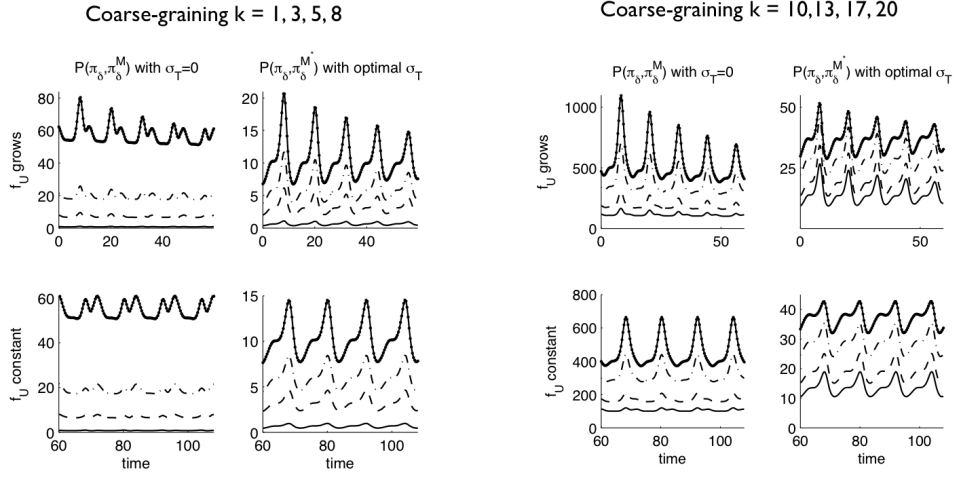
FIGURE 21. **Sensitivity and information-theoretic optimization of Gaussian models for a Gaussian turbulent tracer (82) with $\sigma_U = 0$.** The top panels show the minimum lack of information before and after the optimal noise inflation in the models (85) at different coarse-grainings during the growth phase of the forcing perturbartion. The bottom panel shows the minimum lack of information for the corresponding coarse-grainings for the new climate.

where $\kappa_{\mathrm{M}} = \kappa + \kappa_{\mathrm{M}}^{eddy}$ with $\kappa_{\mathrm{M}}^{eddy}$ the eddy diffusivity coefficient, often utilized for parameterization of unresolved turbulence in climate science [26, 27] while $\dot{W}(x,t)$ denotes space-time white noise forcing with variance σ_T. Contrary to the standard deterministic parameterizations in climate science with $\sigma_T = 0$, we consider here an extended class of stochastic models in (85) with $\sigma_T \neq 0$ to be improved by stochastic forcing inflation [104]. We also assume that the eddy diffusivity is given by $\kappa_{\mathrm{M}}^{eddy} = \theta\kappa_{\mathrm{M}}^{*}$ with $0 \leqslant \theta \leqslant 1$ where $\kappa_{\mathrm{M}}^{*} = \sigma^2/2d_\gamma$ is the exact value of the eddy diffusivity in the white noise limit of the jet fluctuations fluctuations $U'(t)$ the perfect system [80, 81, 92]. Here, we focus on model error for time-periodic statistical steady states for the tracer which are coarse-grained to a given number of spatial Fourier modes; the larger the number of these modes, the larger the demands on the imperfect model to represent smaller scale behavior accurately. Thus, the natural information metric we utilize here for model fidelity consists of (12) utilized spatially but averaged over the period on the attractor similarly to (14).

The most important aspects of UQ in spatially extended systems, in addition to those discussed for the single Fourier mode in the previous sections, are illustrated in figures 21 and 22 (see [80, 81, 92] for a more detailed treatment of these issues).

Figure 21 shows an example of improvements in the forced response of Gaussian imperfect models (85) of the Gaussian tracer by optimizing the unperturbed climate fidelity (14) via inflation of the stochastic forcing; see [32] for details. Here, both the perfect system and the imperfect models are Gaussian with no transient instabilities in the shear dynamics and no fluctuations in the zonal jet, i.e., $\gamma_{v_k} \equiv 0$ and $\sigma_U \equiv 0$. Consequently, the imperfect Gaussian models for the tracer are given by (85) with $\kappa_{\mathrm{M}}^{eddy} \equiv 0$ due to the absence of the mean-wave interaction in the Gaussian perfect model for the tracer. The external perturbations are imposed by increasing the
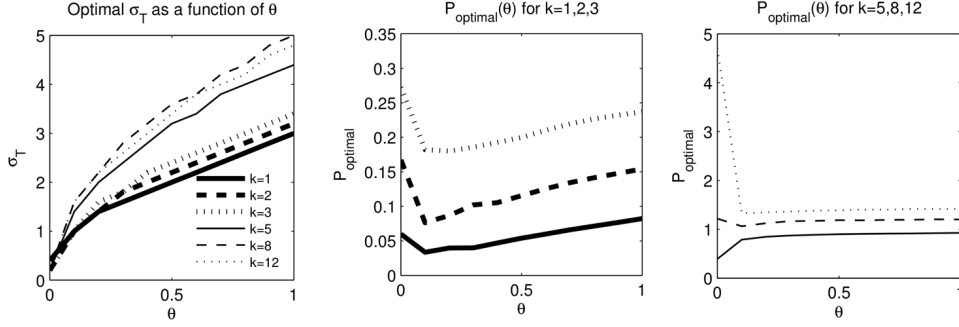
FIGURE 22. **Information-theoretic optimization of Gaussian models (85) for turbulent non-Gaussian tracer (82) with $\sigma_U \neq 0$.** (Left) optimal noise amplitude in (85) for different eddy diffusivity parameterizations $\kappa_{\mathrm{M}}^{eddy} = \theta \kappa_{\mathrm{M}}^*$ with κ_{M}^* the eddy diffusivity in the white noise limit of the zonal jet fluctuations (see text). (Middle and Right) Minimal lack of information at different coarse-grainings in the climate for the class of imperfect models (85) with optimal stochastic forcing. Note the information barriers (§2.2) for the coarse-grainings $k = 5, 8, 12$.

strength of the mean jet $\bar{U}$ through a perturbation $\delta f_U(t)$ of the forcing which is of the ramp type and increases linearly over a finite time interval to reach a new constant value at the end of the process. However, similar to the previously discussed example for the turbulent Gaussian shear flow, the imperfect Gaussian tracer models are driven by overdamped Gaussian velocity fields (here with $\mathcal{E} = 0.7, \theta = 4, K_0 = 10$, in (84) and the model diffusivity $\kappa_{\mathrm{M}} = 0.6$ instead of the true $\kappa = 0.001$.) The measure of the predictive skill, $\mathcal{P}(\pi_\delta, \pi_\delta^{\mathrm{M}})$, is shown in figure 21 next to the corresponding values of uncertainty for the original imperfect model. For the coarse-grainings of $K = 1, 3$, and 5 modes, the improvement is roughly four fold, while for higher coarse-grainings of up to $K = 20$ modes, a nearly 20-fold improvement is achieved by the simple optimal stochastic noise inflation.

Figure 22 illustrates the case of non-Gaussian tracer dynamics with $\sigma_U \neq 0$ in (82); we show several facets of the systematic information-theoretic improvement over the deterministic models with $\sigma_T \equiv 0$ in (85) obtained using the imperfect Gaussian tracer models with optimal stochastic forcing $\sigma_T^* \neq 0$ according to (14). First, the optimal value σ_T^* increases as the deterministic eddy diffusivity increases from zero to κ_{M}^* in the white noise limit of the jet fluctuations; secondly, there is a significant gain of information by utilizing the optimal stochastic imperfect models compared with the deterministic ones and this information gain is necessarily in the dispersion of the tracer. Finally, the right panel in figure 22 illustrates existence of information barriers for optimizing the class of models (85) by noise inflation; optimal value with the smallest information discrepancy occurs for $\kappa_{\mathrm{M}}^{eddy} = 0.1\kappa_{\mathrm{M}}^*$ and this is especially significant for large scale coarse-grainings like $k = 1, 2, 3$ but for finer scale coarse-grainings like $k = 5, 8, 12$ clear information barriers appear for $\kappa_{\mathrm{M}}^{eddy}$ regardless of its value.

We make the following points in this context based on the examples shown in figures 21, 22:

- The predictive skill of the overdamped imperfect models with optimal stochastic noise is greatly improved for various coarse grainings even for simple noise inflation with the same amplitude across all modes and in the time-averaged framework (14).
- The uncertainty in unoptimized imperfect models grows with increasing resolution (see also figure 20).
- Optimal noise σ_T^* depends on the coarse-graining, and the time interval used for averaging; σ_T^* increases as the deterministic eddy diffusivity increases from zero to the white noise limit κ_{M}^*.
- In systems with nontrivial mean-wave interactions, the optimal noise σ_T^* depends on the number of components used for minimizing the relative entropy in (8) or (14), e.g., the tracer or the tracer and the velocity (see [32] for details).
- There exist information barriers to improving imperfect models of spatially extended systems (figure 22). In this complex setting the nature of such barriers depends on the deterministic parameterizations and it is scale dependent, often allowing for uncertainty reduction at the coarse scales while hampering improvements at smaller scales.

7. **Numerical artifacts in UQ with long time integration.** Incorrect or inappropriate numerical algorithms can introduce significant numerical errors into UQ algorithms and bias both the effects of model error and techniques for mitigating this error, such as those discussed in the previous sections. The main difficulty in dealing with numerical artifacts in UQ applications is that even in low-dimensional configurations they often lead to plausible results without unphysical singularities, instabilities, etc., and their consequences can be much more elusive than in the purely deterministic framework. Nevertheless, detection and proper handling of such numerical artifacts is crucial for reliable UQ procedures and imperfect model optimization. Here, we illustrate these issues based on the models already discussed in the previous sections and we discuss a number of examples of numerical artifacts which are particularly misleading for long time integration of complex nonlinear systems with unresolved degrees of freedom. The examples used below are:

- Fake intermittency in numerical PDFs;
- Divergence of PCE approximations for long time integrations;
- Failure of MC with large sample due to rare events.

The time scales for the 'long time integration' vary depending on the problem; however, in all cases 'long time' means comparable with or exceeding the time scales associated either with a transient adjustment onto the attractor or time scales associated with some transient dynamics in the absence of an attractor.

Fake intermittency in numerical PDFs for stochastic systems:
In §6 we showed, following [92], how the presence of the mean gradient in the dynamics of a passive tracer can lead to intermittent PDF with fat tails provided that the velocity field driving the tracer dynamics is turbulent but with a finitely correlated in time Rossby waves shearing the turbulent zonal jet. However, it can be shown (see Appendix D.3) that in the white noise limit of (82) with rapidly
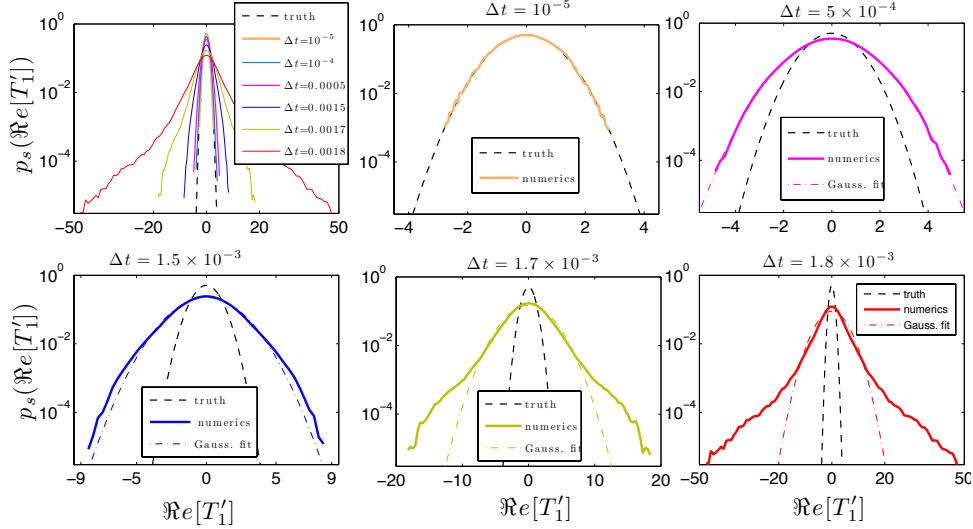
FIGURE 23. **Numerical artifacts in long time integration of a stochastic system with a Gaussian invariant measure.** Comparison of equilibrium PDFs for the tracer mode T_1 computed for the system (86) using Euler scheme with different time steps Δt. Inadequate time step in the numerical integration scheme results in erroneous equilibrium PDFs; for the system parameters used here the PDFs have fat tails for $\Delta t \gtrsim 1.2 \times 10^{-3}$.

System parameters: $f_U = 5, d_U = 0.5, \sigma_U = 2$ $f_{T_k} = 0, d_{T_k} = 0.101, \alpha\eta_{v_k} = 0.5$.

decorrelating shear waves v that the tracer dynamics is given by

$$
\begin{aligned}
a) \quad \mathrm{d}T_k'(t) &= \big[(-d_{T_k} - ikU(t))T_k'(t) - \alpha\bar{f}_{v_k}(t)\big]\mathrm{d}t - \alpha\eta_{v_k}\mathrm{d}W_{v_k}(t), \\
b) \quad \mathrm{d}U(t) &= \big[-d_U U(t) + f_U(t)\big]\mathrm{d}t + \sigma_U \mathrm{d}W_U(t),
\end{aligned}
\tag{86}
$$

where $\bar{f}_{v_k}$ is the forcing in the dynamics of v_k in the white noise limit, $\eta_{v_k} = \sigma_{v_k}/d_{v_k}$, $\omega_{v_k}(t) = a_k U(t) + b_k$, $d_{T_k} = d_T + \kappa k^2$; other parameters are defined as in (82). It can be shown analytically (see Appendix D.3) that when $\bar{f}_{v_k} = 0$ all tracer modes have Gaussian statistics even in the presence of mean gradient $\alpha \neq 0$ and turbulent zonal jet, i.e., for $\sigma_U \neq 0$. We use these analytical results to show that numerical estimates of these Gaussian PDFs can lead to fat-tailed PDFs if the integration time step is too large.

In figure 23 we show various fake fat-tailed PDFs obtained from numerical approximations of (86) with too large time-steps and for the same Fourier mode with $k = 1$; the forward Euler method was used in the computations but it is well known (e.g., [57]) that both the Euler and Milstein methods have the same weak order of convergence, i.e., they have the same order of accuracy of approximating the statistics of the solutions. The first inset in figure 23 shows a superposition of all the erroneous PDFs and the subsequent insets compare the true Gaussian PDF with the numerical approximations obtained with different integration time steps. Note that for time steps too small for the presence of fake fat tails, one can obtain nearly Gaussian PDFs with incorrect variance; such errors might be difficult to spot even if the Gaussian nature of the true PDF is known.
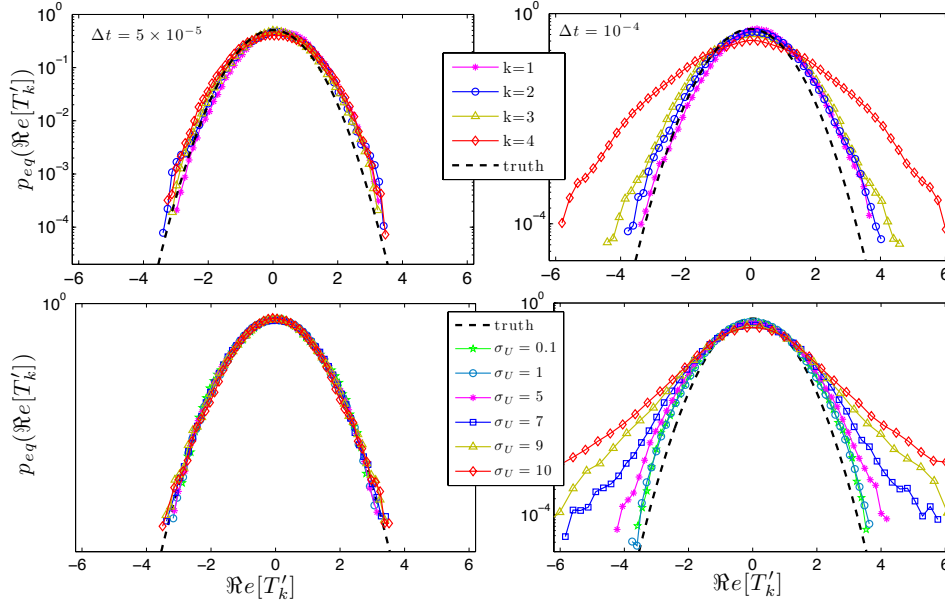
FIGURE 24. **Numerical artifacts in long time integration of the stochastic system with a Gaussian invariant measure**. (Top) Comparison of equilibrium PDFs computed for the system (86) using the explicit Euler scheme for two different time steps Δt and four different wavenumbers $k = 1, ..., 4$. The forcing in $U(t)$ and T_k, as well as the noise amplitudes and the damping d_U are kept constant and independent of k. (Bottom) Comparison of the equilibrium PDFs computed for the system (86) for two different time steps Δt and noise amplitudes σ_U.
System parameters: $f_U = 5, d_U = 0.5, f_{T_k} = 0, d_{T_k} = 0.101, \alpha\eta_{v_k} = 0.5$; $\sigma_U = 2$ (top) and varied (bottom).

Figure 24 illustrates the appearance of fake fat tails in numerically estimated PDFs for increasing wave number of the tracer (top) and increasing noise amplitude σ_U in the jet (bottom). The two insets in the top row show results of computations with two different integration time steps using the forward Euler method and for different wave numbers k. For sufficiently small time step (relative to the system parameters) the numerical approximations remain nearly Gaussian throughout the considered wavenumbers $k = 1, \ldots, 4$; however, this situation changes dramatically for larger time steps when approximating the PDFs for high wave numbers, as can be seen in the inset computed with $\Delta t = 10^{-4}$. Similar fat-tailed artifacts occur when increasing the noise amplitude, σ_U, in $U(t)$ as shown in the npottom two insets of figure 24.

The following points deserve highlighting based on the above examples:

- Insufficiently small time step (relative to the system parameters) in the numerical integration scheme, be it forward Euler or Milstein, leads to erroneous fat-tailed PDFs (figure 23). The variance is first affected by the errors leading to nearly Gaussian PDFs with incorrect variance.

- The time-step value yielding accurate approximations for the statistics of turbulent modes at small wave numbers may be insufficiently small for larger wave numbers in the spectrum (figure 24).

Note that the above artifacts might be difficult to detect in high-dimensional nonlinear systems since they are not associated with any spectacular divergences, singularities, etc. There is a great need for mathematical theory and the systematic assessment of numerical algorithms which capture the long-time statistical dynamics of turbulent dynamical systems with high accuracy. Recent work in this important area was carried out in [119, 120] for the example of a turbulent dynamical system arising in the large Prandtl number limit of classical Rayleigh-Benard convection; this work serves as a model for further research.

Divergence of PCE approximations for long time integrations of stochastic systems.

These limitations of the Polynomial Chaos Expansion approximation were already discussed in §5 and at length in [10] and are associated with the finite order truncations in systems with fat-tailed PDFs and intermittency; the most important of these artifacts are:

- Failure of truncated PCE approximations to detect unbounded growth of statistics in systems with parametric uncertainty in the damping (see figure 18 and [10]).
- Failure of truncated PCE approximations in systems with nontrivial attractors (e.g., [106, 64, 3]).
- Divergence of truncated PCE in systems driven by white noise, except at sufficiently short times (figure 19 and [10]).

Similarly to the issues raised in the previous paragraph, the artifacts arising in the context of the truncated PCE approximations are not associated with spectacular blow-up of the approximations; rather, they lead to underestimated statistical moments which are usually more difficult to detect in complex configurations.

Failure of Monte Carlo simulations with insufficiently large ensemble size due to rare events.

This shortcoming of the MC was already illustrated in figure 18 of §5 where, similarly to the PCE approximations and to the Gaussian moment closure, MC completely failed to detect the unbounded growth in low order statistics despite large sample sizes for this toy problem with 50000 realizations. The problems associated with MC simulations and large scale phenomena triggered by rare events are well known (e.g. [54]) but it is important to stress their role in creating various artifacts in UQ applications.

8. **Concluding remarks.** Here, we reviewed the recent developments and discussed a range of new important mathematical issues arising in the emerging stochastic-statistical framework to quantifying uncertainty and mitigating model error in imperfect predictions of partially observed complex high-dimensional turbulent dynamical systems. The systematic discussion of issues in imperfect model predictions for uncertain initial data at short and medium-long ranges, as well as the forced response to system perturbations, was accomplished based on a suite of physically relevant and progressively more complex but mathematically tractable test models. These instructive test models possessed such important features as

the two-way coupling between the resolved dynamics and the stochasticity due to the interactions with unresolved processes, intermittency and positive Lyapunov exponents, eddy diffusivity parameterization and overdamped turbulent spectra. In particular, the following two themes were emphasized throughout the article:

- The utility of the unified stochastic-statistical framework for UQ and improvement of imperfect predictions with model error in turbulent dynamical systems with intermittency, positive Lyapunov exponents and hidden instabilities;
- Existence of information barriers to model improvement, their hallmarks and consequences for UQ in turbulent dynamical systems.

Other topics developed here were concerned with the role and mitigation of model error due to coarse-graining and dimensional reduction, moment closure approximations and associated turbulent flux parameterization, the effects of memory of initial conditions in producing medium and long range forecasts and the sensitivity of imperfect models to external perturbations. The mathematical toolkit utilized empirical information theory, fluctuation-dissipation theorems and systematic physics-constrained, statistical-stochastic modelling for large-dimensional turbulent dynamical systems. New mathematical and computational phenomena in addressing these problems have been discussed throughout this article. Many theoretical issues remain in UQ and prediction of turbulent dynamical systems and the authors hope that this review inspires mathematicians, applied mathematicians, and scientists to contribute to this increasingly important contemporary science discipline.

Finally, we mention several important topics for mathematical research directly connected with this expository article and not discussed in detail.

(1) Data assimilation and filtering with judicious model error for turbulent systems

A key feature of turbulence in inertial and energy dissipation regimes is bursts of energy across multiple scales with intermittent instabilities and effectively stochastic forcing due to complicated interactions between the turbulent modes; the resulting dynamics is highly non-Gaussian with associated fat-tailed PDFs. Stochastic Parameterization Extended Kalman Filters (SPEKF) have been introduced and analyzed recently [30, 29, 94, 93] as computationally cheap algorithms which make judicious model errors while retaining high filtering skill for complex turbulent signals [42, 55, 9, 94]. For example, aliasing is usually viewed as a bad feature of numerical algorithms; in the present context, judicious use of aliasing yields stochastic superresolution [94, 55, 93]. The basis for the SPEKF algorithms is the system (38), used earlier in a different context, for the complex scalar partially observed turbulent mode $u(t)$ (the reader can think of a Fourier amplitude of turbulence at a given spatial wavenumber) coupled with stochastic additive forcing and multiplicative damping/instability coefficients, $\gamma(t)$, $b(t)$, which are learned on the fly from the observed turbulent signal. The advantage of the system (38) is that is has non-Gaussian dynamics but nevertheless exactly solvable first and second-order statistics, so they are readily implemented practically in a filtering algorithm. The rich dynamical behavior of (38), as already illustrated in figure 7, can be utilized to test the filter performance of a wide variety of Gaussian filter approximations [9]. See the recent book [93] for an elementary introduction to these topics.

(2) Fluctuation-Dissipation Theorems for high-dimensional turbulent systems

There is a great need for further developments exploiting the appropriate fluctuation dissipation theorems for linking statistical equilibrium fidelity and sensitivity of imperfect models of high dimensional turbulent systems within the framework of information theory for model error and sensitivity. Moreover, much more rigorous work, in addition to [41], should be done for time-periodic systems, general multiplicative noise, and rigorous FDT representation formulas. The recent papers [1, 79, 74, 80, 81, 82, 11] and, in particular, [81, 88, 31] contain much of the formal research program and demonstrate it on exactly solvable test models. Some recent results [81, 11] open up enticing prospects for developing techniques for improving imperfect model sensitivity based on specific tests carried out in the unperturbed climate and using the kicked response FDT.

The fundamental concepts justifying the linear response theory for forced dissipative stochastic dynamical systems through the fluctuation dissipation theorem have been substantially refined since the pioneering work of Leith [65] and are well known in the literature (see [41] and the references above). Despite this fact, minor modifications of this general concept appear intermittently in the literature in various guises. One recent example proposed in [14] is the method for predicting the low-frequency variability of El Nino-Southern Oscillation (ENSO) based on a particular noise sampling technique applied to the reduced model obtained by fitting a quadratic model to the ENSO data through a multilevel linear regression. We note that such an approach reduces to the classical FDT framework in the limit of taking infinitely many sample paths of the suggested noise 'snippets', provided that the reduced ad hoc quadratic model does not yield unphysical solutions with a finite-time blow-up (see [89] and (4) below). The authors in [14] seem to be unaware of this connection to earlier work [1, 88, 82, 31] yet they empirically take this limit.

(3) Blended Dynamic Subspace techniques for UQ in turbulent systems

Development of computationally efficient and accurate methods for propagation of uncertainty in turbulent systems is difficult since the complex intermittent interactions between the resolved and unresolved processes are important across a wide range of spatial and temporal scales with nontrivial energy transfer on the attractor. Due to this strong, time-dependent, two-way coupling between the mean dynamics and stochasticity, most attempts aimed at reducing the underlying complexity of the problem by separating the mean evolution from the dominant dimensionality of the uncertain subspace associated with the most energetic uncertain unresolved processes usually end in failure.

Contrary to the truncated Polynomial Chaos Expansions which retain a finite dimensional 'uncertain subspace' spanned by a fixed basis set, the methods exploiting dynamically orthogonal (DO) field equations applied to spatially extended systems with uncertainty attempt to capture the evolution of the most energetic stochastic modes in a dynamically consistent fashion [109, 110]. Nevertheless, this improved method suffers from deterioration of UQ skill in time for turbulent dynamical systems and even non-normal linear systems [112]. New recently developed methods by T. Sapsis and one of the authors, which combine the DO framework with an approximate but statistically consistent mechanism for turbulent energy transfer between the reduced uncertain subspace and the under-resolved complementary dynamics

offers a very promising way forward. However, this work is reported elsewhere [112, 111, 113].

(4) PHYSICS CONSTRAINED DATA-DRIVEN STATISTICAL-STOCHASTIC MODELS

Development of data-driven reduced stochastic-statistical models of turbulent dynamical systems for uncertainty quantification and long range forecasting is an extremely important issue. Standard linear regression models can have some skill but they suffer from inherent mathematical limitations and intrinsic barriers in skill [82]. Ad-hoc nonlinear regression models can exhibit improved skill in a training time series (see references in [89]) but can suffer unphysical finite-time blow-up of statistical solutions, as well as unphysical pathology in their invariant measure [89]. There are rigorous proofs [126] that physics-constrained stochastic mode reduction models which are Markovian have the physically correct asymptotic behavior for their invariant measure for low-frequency variability; however, further generalizations are required to include non-Markovian memory effects which are crucial in many applications. There is a recent new approach to designing such methods [83]. There is a wide array of data-driven clustering algorithms [77, 46, 47, 44, 23, 24, 45, 48] to develop multiple regime Markov models for use in prediction. Giannakis and one of the authors [37, 38, 39] apply empirical information theory to assess the skill of coarse-grained partitions of phase space and reduced Markov models for long-range prediction. The methods of Horenko [46, 47, 44, 45, 48, 39] are especially promising in this context but need further physical constraints to be more useful for long range forecasting. This is an exciting area for future interdisciplinary research.

Appendix A. Linear Gaussian models. The linear Gaussian framework for UQ was discussed in details in §3. Here, we provide more details on various derivations from that section and discuss some generalizations of the previously discussed results.

A.1. **Path-wise solutions and statistics of the $2{\times}2$ linear systems.** The solutions of the linear system (18) can be easily found in the form

$$\begin{pmatrix} x(t, t_0) \\ y(t, t_0) \end{pmatrix} = e^{\hat{L}(t-t_0)} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} + \int_{t_0}^{t} e^{\hat{L}(t-s)} \begin{pmatrix} F(s) \\ 0 \end{pmatrix} \mathrm{d}s + \sigma \int_{t_0}^{t} e^{\hat{L}(t-s)} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \mathrm{d}W(s),$$

(87)

where the Wiener process, $W(s)$, $W(0) = 0$, is defined on the whole real line, and the operator $e^{\hat{L}t}$ is given by

$$e^{\hat{L}t} = \frac{1}{\lambda_1 - \lambda_2} \begin{pmatrix} (\lambda_1 - a)e^{\lambda_2 t} - (\lambda_2 - a)e^{\lambda_1 t} & e^{\lambda_1 t} - e^{\lambda_2 t} \\ q\left(e^{\lambda_1 t} - e^{\lambda_2 t}\right) & (\lambda_1 - a)e^{\lambda_1 t} - (\lambda_2 - a)e^{\lambda_2 t} \end{pmatrix},$$

(88)

and the eigenvalues of $\hat{L}$ are given by

$$\lambda_{1,2} = \frac{1}{2}\left(a + A \pm \sqrt{(a-A)^2 + 4q}\right). \tag{89}$$

Clearly, the conditions (20) imply that the eigenvalues (89) have negative real parts and the system converges to a statistical attractor (see, e.g. [4, 87], for a detailed definition of the attractor in the general nonautonomous case).

The mean dynamics for the linear Gaussian system (18) can be easily found by taking the ensemble average of (87) which leads to

$$\begin{pmatrix} \bar{x}(t,t_0) \\ \bar{y}(t,t_0) \end{pmatrix} = e^{\hat{L}(t-t_0)} \begin{pmatrix} \bar{x}_0 \\ \bar{y}_0 \end{pmatrix} + \int_{t_0}^{t} e^{\hat{L}(t-s)} \begin{pmatrix} F(s) \\ 0 \end{pmatrix} \mathrm{d}s, \tag{90}$$

where $\bar{x}_0$, $\bar{y}_0$ denote the mean initial conditions at t_0. The mean on the attractor can be obtained by taking the limit $t_0 \to -\infty$ in (90) which exists for any bounded forcing $F(t)$ provided that the stability conditions (20) hold. It can be easily seen with the help of (88) that for time-periodic forcing $F(t) = f^0 + f^1 \sin\boldsymbol{\omega} t$ the time averaged mean on the attractor is given by

$$\begin{aligned} a) \quad & \bar{x}_{att}^T \equiv \lim_{t_0 \to -\infty} \frac{1}{T} \int_{t_0}^{t_0+T} \bar{x}(t,t_0)\mathrm{d}t \quad = -\frac{Af^0}{aA-q}, \\[2mm] b) \quad & \bar{y}_{att}^T \equiv \lim_{t_0 \to -\infty} \frac{1}{T} \int_{t_0}^{t_0+T} \bar{y}(t,t_0)\mathrm{d}t \quad = \frac{qf^0}{aA-q}, \end{aligned} \tag{91}$$

where $T = 2\pi/\omega$ in the time-periodic case ($f^1 \neq 0$); for constant forcing taking an arbitrary non-zero T reduces (91) to (21) in §3.

The covariance matrix of the process $\boldsymbol{x}(t,t_0)$ satisfying (18) is given by

$$R(t,t_0) = e^{\hat{L}(t-t_0)} R_0 e^{\hat{L}^T(t-t_0)} + \int_{t_0}^{t} e^{\hat{L}(t-s)} \begin{pmatrix} 0 & 0 \\ 0 & \sigma^2 \end{pmatrix} e^{\hat{L}^T(t-s)}. \tag{92}$$

so that, using (92) and (88), the the covariance Σ on attractor is

$$\Sigma \equiv \lim_{t_0 \to -\infty} R(t,t_0) = \begin{pmatrix} 1 & -a \\ -a & aA-q+a^2 \end{pmatrix} \frac{\sigma^2}{-2(a+A)(aA-q)}. \tag{93}$$

and the autocovariance on the attractor is given by

$$\mathcal{C}_{att}(\tau,t) = \lim_{t_0 \to -\infty} \left\langle \begin{pmatrix} x(t,t_0)-\langle x(t,t_0)\rangle \\ y(t,t_0)-\langle y(t,t_0)\rangle \end{pmatrix} \begin{pmatrix} x(t+\tau,t_0)-\langle x(t+\tau,t_0)\rangle \\ y(t+\tau,t_0)-\langle y(t+\tau,t_0)\rangle \end{pmatrix}^T \right\rangle = \Sigma e^{\hat{L}^T \tau}. \tag{94}$$

Note that in this linear Gaussian framework the autocovariance on the attractor is independent of the forcing $F(t)$.

A.2. Path-wise solutions and statistics of the MSm model.

The solutions satisfying the Mean Stochastic Model (24) are given by the Ornstein-Uhlenbeck process with sample paths

$$x^{\mathrm{MSM}}(t,t_0) = e^{-\gamma_{\mathrm{MSM}}(t-t_0)} x_0^{\mathrm{MSM}} + \int_{t_0}^{t} e^{-\gamma_{\mathrm{MSM}}(t-s)} F_{\mathrm{MSM}}(s)\mathrm{d}s + \sigma_{\mathrm{MSM}} \int_{t_0}^{t} e^{-\gamma_{\mathrm{MSM}}(t-s)} \mathrm{d}W(s), \tag{95}$$

where x_{MSM}^0 is the initial condition at $t = t_0$. Consequently, the statistics of this Gaussian process is fully determined by

$$\bar{x}^{\text{MSM}}(t, t_0) = e^{-\gamma_{\text{MSM}}(t-t_0)}\bar{x}_0^{\text{MSM}} + \int_{t_0}^t e^{-\gamma_{\text{MSM}}(t-s)}F_{\text{MSM}}(s)\mathrm{d}s, \tag{96}$$

$$R^{\text{MSM}}(t, t_0) \equiv Var\left(x^{\text{MSM}}(t, t_0)\right) = e^{-2\gamma_{\text{MSM}}(t-t_0)}Var\left(x_0^{\text{MSM}}\right) + \frac{\sigma_{\text{MSM}}^2}{2\gamma_{\text{MSM}}}\left(1 - e^{-2\gamma_{\text{MSM}}(t-t_0)}\right), \tag{97}$$

so that for $\gamma_{\text{MSM}} > 0$ the statistics on the attractor in the general nonautonomous case (see, e.g., [87, 4]) is

$$\bar{x}_{att}^{\text{MSM}}(t) = \lim_{t_0 \to -\infty}\int_{t_0}^t e^{-\gamma_{\text{MSM}}(t-s)}F_{\text{MSM}}(s)\mathrm{d}s, \tag{98}$$

$$\mathcal{C}_{att}^{\text{MSM}}(\tau) = \lim_{t_0 \to -\infty}\left\langle(x^{\text{MSM}}(t, t_0) - \bar{x}^{\text{MSM}}(t, t_0))(x^{\text{MSM}}(t+\tau, t_0) - \bar{x}^{\text{MSM}}(t+\tau, t_0))\right\rangle$$
$$= \frac{\sigma_{\text{MSM}}^2 e^{-\gamma_{\text{MSM}}\tau}}{2\gamma_{\text{MSM}}}. \tag{99}$$

so that the variance on the attractor becomes $R_{att} \equiv \mathcal{C}_{att}^{\text{MSM}}(\tau = 0) = \sigma_{\text{MSM}}^2/2\gamma_{\text{MSM}}$.

Note that $x^{MS}(t)$ cannot reproduce the marginal autocovariance (94) for $x(t)$ since the necessary condition

$$e^{-\gamma_{\text{MSM}}\tau} = \frac{\lambda_1 e^{\lambda_2\tau} - \lambda_2 e^{\lambda_1\tau}}{\lambda_1 - \lambda_2}, \tag{100}$$

cannot be satisfied for all lags τ except for the degenerate case $\lambda_1 = \lambda_2$. A particular choice of γ_{MSM} which guarantees correct decorrelation time for $x^{\text{MSM}}(t)$ is

$$\int_0^\infty e^{-\gamma_{\text{MSM}}\tau}\mathrm{d}\tau = \int_0^\infty \frac{\lambda_1 e^{\lambda_2\tau} - \lambda_2 e^{\lambda_1\tau}}{\lambda_1 - \lambda_2}\mathrm{d}\tau, \tag{101}$$

and leads to

$$\gamma_{\text{MSM}} = -\frac{aA - q}{a + A}. \tag{102}$$

A.3. **Families of imperfect models with correct marginal equilibrium statistics.** The existence of families of imperfect models the linear system (18) with correct marginal two-point equilibrium statistics or climatology was discussed in §3.2.1. Here, we obtain the formulas for the coefficients in the imperfect models within a given family. These relationships are used in §3.4 in order to construct the suite of imperfect models. Inline with the propositions introduced in §3.2.1, we focus mainly on systems with constant forcing; remarks concerning extensions to the nonautonomous case with time-periodic statistics on the attractor (see, e.g., [87, 4] for detailed definitions) are included in the relevant places.

A.3.1. *Distinct families of 2×2 systems.*

(I) **Constant forcing; models with the same marginal two-point statistics at equilibrium.** Consider a linear stochastic model (18) with coefficients (a, q, A, σ) and constant forcing $F = f^0$. According to Prop. 1, any 2×2 linear model (18) with the coefficients $(a_{\mathrm{M}}, q_{\mathrm{M}}, A_{\mathrm{M}}, \sigma_{\mathrm{M}}, f_{\mathrm{M}}^0)$ satisfying

$$a_{\mathrm{M}} + A_{\mathrm{M}} = a + A, \quad a_{\mathrm{M}} A_{\mathrm{M}} - q_{\mathrm{M}} = aA - q, \quad \sigma_{\mathrm{M}} = \sigma, \quad A_{\mathrm{M}} f_{\mathrm{M}}^0 = A f^0, \qquad (103)$$

has the same equilibrium marginal statistics for $x(t)$. Thus, there is a one-parameter family of models with structure (18) with the same marginal statistics for $x(t)$ and $x^{\mathrm{M}}(t)$ provided that the coefficients of the imperfect models satisfy

$$a_{\mathrm{M}} = w, \quad A_{\mathrm{M}} = a + A - w, \quad q_{\mathrm{M}} = q - aA + w(a + A - w),$$
$$f_{\mathrm{M}}^0 = \frac{f^0 A}{a + A - w}, \qquad (104)$$

where w is a free parameter; for $w = a$ this system coincides with (18).

Remark. For time-periodic forcing $F = f^0 + f^{1s} \sin \omega t + f^{1c} \cos \omega t$ in the perfect system (18) the four-parameter family of 2×2 models with correct time-averaged two-point marginal statistics on the attractor and coefficients is given by (104) with three additional free parameters $f_{\mathrm{M}}^{1s}, f_{\mathrm{M}}^{1c}, \omega_{\mathrm{M}}$.

(II) **Constant forcing; identical marginal climatology.** Consider the linear model (18) with coefficients (a, q, A, σ) and forcing $F = f^0$. According to Prop. 2, there is a three-parameter family of models with structure (18) and coefficients $(a_{\mathrm{M}}, q_{\mathrm{M}}, A_{\mathrm{M}}, \sigma_{\mathrm{M}}, f_{\mathrm{M}}^0)$, with the same marginal climatology for $x(t)$ and $x^{\mathrm{M}}(t)$ provided that the coefficients of the imperfect models satisfy:

$$a_{\mathrm{M}} = w_1, \quad q_{\mathrm{M}} = w_2, \quad A_{\mathrm{M}} = w_3, \quad \sigma_{\mathrm{M}} = \sigma \frac{(w_1 + w_3)(w_1 w_3 - w_2)}{(a + A)(aA - q)},$$
$$f_{\mathrm{M}}^0 = f^0 \frac{A(w_1 w_3 - w_2)}{w_3(aA - q)}, \qquad (105)$$

where $\{w_1, w_2, w_3\}$ are the free parameters.

Remark. For time-periodic forcing $F = f^0 + f^{1s} \sin \omega t + f^{1c} \cos \omega t$ in the perfect system (18) the six-parameter family of 2×2 models with correct time-averaged marginal climatology for x^{M} has coefficients satisfying (104) with three additional free parameters $f_{\mathrm{M}}^{1s}, f_{\mathrm{M}}^{1c}, \omega_{\mathrm{M}}$.

A.3.2. *Families of MSMs with the same marginal climatology.* As stated in Proposition (3), the MSm (24) does not have sufficient number of degrees of freedom to match the marginal two-point statistics on the attractor for $x(t)$ satisfying (18). MSm's with the same climatology can be described as follows.

(III) **Constant forcing; correct marginal climatology for $x^{\mathrm{MSM}}(t)$.** Consider the linear model (18) with coefficients (a, q, A, σ) and constant forcing $F = f^0$. According to Proposition (3), there exists a one-parameter family of MSm's (24) with coefficients $(\gamma_{\mathrm{MSM}}, \sigma_{\mathrm{MSM}}, F_{\mathrm{MSM}})$ and the same marginal climatology for x^{MSM} and x; these coefficients satisfy

$$\gamma_{\mathrm{MSM}} = w, \quad \sigma_{\mathrm{MSM}}^2 = -w \frac{\sigma^2}{(a + A)(aA - q)}, \quad F_{\mathrm{MSM}} = -w \frac{A f^0}{aA - q}, \qquad (106)$$

where w is a free parameter. The MSm with correct marginal climatology and correct decorrelation time for x^{MSM} is the one with $w = -(aA - q)/(a + A)$ consistent with (102).

Remark. For time-periodic forcing $F = f^0 + f^{1s}\sin\omega t + f^{1c}\cos\omega t$ in the perfect system (18) the four-parameter family of MSm models with correct time-averaged marginal climatology for x^{MSM} has coefficients given by by (106) with three additional free parameters $f^{1s}_{\text{MSM}}, f^{1c}_{\text{MSM}}, \omega_{\text{MSM}}$.

Appendix B. **White noise limit for elementary systems with intermittency.** Construction of imperfect models by considering special white noise limits was discsussed in §4.1.1. Here, we consider the limiting dynamics of the quadratic complex scalar system (38) assuming that the stochastic fluctuations of the damping $\gamma(t)$ and the forcing fluctuations $b(t)$ decorrelate infinitely fast. Similar procedure for the turbulent tracer is discussed in §7 and Appendix D.3 in order to unambiguously illustrate some numerical artifacts in long time numerical integrations of stochastic systems.

Consider the dynamics of the system (38) in the following limit

$$d_\gamma \to \infty, \qquad \sigma_\gamma \to \infty, \qquad \tilde{\sigma}_\gamma \equiv \sigma_\gamma/d_\gamma = const > 0. \tag{107}$$

$$d_b \to \infty, \qquad \sigma_b \to \infty, \qquad \tilde{\sigma}_b \equiv \sigma_b/d_b = const > 0. \tag{108}$$

so that the absolute value of the autocovariance of $\gamma(t)$ and $b(t)$ formally approach the Dirac delta distributions, i.e,

$$\lim_{\substack{d_\gamma,t\to\infty \\ \sigma_\gamma/d_\gamma=const}} |\langle\gamma(t+\tau)\gamma(t)\rangle| = \frac{\tilde{\sigma}_\gamma}{2}\delta_0(\tau), \qquad \lim_{\substack{d_b\to\infty \\ \sigma_b/d_b=const}} |\langle b(t+\tau)b(t)\rangle| = \frac{\tilde{\sigma}_b}{2}\delta_0(\tau),$$
$$\tag{109}$$

and the processes $\gamma(t)$, $b(t)$ in (38) approach the white noise processes

$$d\gamma(t) \to \tilde{\sigma}_\gamma dW_\gamma(t), \quad db(t) \to \tilde{\sigma}_b dW_b(t), \tag{110}$$

where $W_\gamma(t), W_b(t)$ are two independent Wiener processes defined for all time, as in (38). As usual in physics and engineering, the white noise limit of colored noise leads to the Stratonovich integral (e.g., [28]). This fact is a consequence of requiring non-vanishing correlations $\langle u(t)\gamma(t)\rangle$, $\langle u(t)b(t)\rangle$ before the limit and after the limit. The limiting process leads to the Stratonovitch SDE for the complex scalar $u(t)$,

$$du(t) = \Big[\big(-\hat{\gamma}+i\hat{\omega}(t)\big)u(t)+F(t)\Big]dt-\tilde{\sigma}_\gamma u(t)\circ dW_\gamma(t)+\tilde{\sigma}_b dW_b(t)+\sigma_u dW_u(t), \tag{111}$$

where W_γ and the components of W_b and W_u are all independent Wiener processes. We rewrite (111) in the Ito form as

$$du(t) = \Big[\big(-\alpha+i\hat{\omega}(t)\big)u(t)+F(t)\Big]dt-\tilde{\sigma}_\gamma u(t)dW_\gamma(t)+\tilde{\sigma}_b dW_b(t)+\sigma_u dW_u(t), \tag{112}$$

where $\alpha = \hat{\gamma} - \frac{1}{2}\tilde{\sigma}_\gamma^2$ is the damping term in the white noise limit.

B.1. **Invariant measure.** Here, we illustrate the procedure for deriving the invariant measure associated with (112) in a simplified case when $F \equiv 0$. First, it is convenient to rewrite (112) by means of the orthogonal transformation $\tilde{u}(t) = e^{-i\int_{t_0}^{t} \hat{\omega}(s)\mathrm{d}s} u(t)$ which leads to

$$\mathrm{d}\tilde{u}(t) = -\alpha\tilde{u}(t)\mathrm{d}t - \tilde{\sigma}_\gamma \tilde{u}(t)\mathrm{d}W_\gamma(t) + \tilde{\sigma}_b e^{-i\int_{t_0}^{t} \hat{\omega}(s)\mathrm{d}s}\mathrm{d}W_b(t) + \sigma_u e^{-i\int_{t_0}^{t} \hat{\omega}(s)\mathrm{d}s}\mathrm{d}W_u(t),$$
(113)

followed by a subseqeuent transformation to the polar coordinates using

$$\mu + i\phi = \ln\tilde{u},$$
(114)

where ϕ is the polar angle and $\mu \equiv \ln\varrho$. Standard manipulations using (113), (114) and the Ito calculus lead to the following system with the same statistics as (113) and (112)

$$a) \quad \mathrm{d}\varrho = \left[-\alpha\varrho + \frac{\tilde{\sigma}_u^2}{2\varrho} \right]\mathrm{d}t - \tilde{\sigma}_\gamma\varrho\,\mathrm{d}W_\gamma + \tilde{\sigma}_b\mathrm{d}W_b^\varrho + \tilde{\sigma}_u\mathrm{d}W_u^\varrho,$$

$$b) \quad \mathrm{d}\phi = \frac{1}{\varrho}\left(\tilde{\sigma}_u\mathrm{d}W_u^\phi + \tilde{\sigma}_b\mathrm{d}W_b^\phi\right),$$
(115)

where $\tilde{\sigma}_u \equiv \sigma_u/\sqrt{2}$ the new noise terms are defined via the orthogonal transformations

$$\left[\begin{array}{c} \mathrm{d}W_{u,b}^\varrho(t) \\ \mathrm{d}W_{u,b}^\phi(t) \end{array} \right] = \left(\begin{array}{cc} \cos(\phi+\Omega) & \sin(\phi+\Omega) \\ -\sin(\phi+\Omega) & \cos(\phi+\Omega) \end{array} \right) \left[\begin{array}{c} \mathrm{d}W_{u,b}^1(t) \\ \mathrm{d}W_{u,b}^2(t) \end{array} \right],$$
(116)

where $W_{u,b}^1 \equiv \Re e[W_{u,b}]$, $W_{u,b}^2 \equiv \Im m[W_{u,b}]$ and $\Omega(t) \equiv \int \hat{\omega}(s)\mathrm{d}s$.

The Fokker-Planck equation for the probability density p associated with the process satisfying (115) is given, in polar coordinates, by

$$\partial_t p(\varrho,\phi,t) = -\partial_\varrho \left[\left(-(\hat{\gamma}-\tfrac{1}{2}\tilde{\sigma}_\gamma^2)\varrho + \frac{\tilde{\sigma}_u^2}{2\varrho} \right) p - \frac{1}{2}\partial_\varrho \left[(\tilde{\sigma}_\gamma^2\varrho^2 + \tilde{\sigma}_b^2 + \tilde{\sigma}_u^2)\,p \right] \right] + \frac{\tilde{\sigma}_u^2 + \tilde{\sigma}_b^2}{2\varrho^2}\partial_\phi^2 p,$$
(117)

The stationary solutions of the Fokker-Planck equation satisfying the detailed balance conditions (see e.g. [28]) lead to the following two equations

$$a) \left(-(\hat{\gamma}-\tfrac{1}{2}\tilde{\sigma}_\gamma^2)\varrho + \frac{\tilde{\sigma}_u^2}{2\varrho} \right) p - \frac{1}{2}\partial_\varrho \left[(\tilde{\sigma}_\gamma^2\varrho^2 + \tilde{\sigma}_b^2 + \tilde{\sigma}_u^2)\,p \right] = 0, \qquad b) \frac{\sigma_u^2 + \tilde{\sigma}_b^2}{2\varrho^2}\partial_\phi p = 0.$$
(118)

Equation (118b) implies that $p_{eq} = p_{eq}(\varrho)$ which implies that (118a) is solved by

$$p_{eq}(\varrho) = \frac{N_0\varrho}{(\tilde{\sigma}_\gamma^2\varrho^2 + \tilde{\sigma}_b^2 + \tilde{\sigma}_u^2)^\delta}, \qquad \delta = \frac{3}{2} + \frac{\hat{\gamma} - \tfrac{1}{2}\tilde{\sigma}_\gamma^2}{\tilde{\sigma}_\gamma^2}.$$
(119)

Finally, the marginal probability distributions for $u_1 \equiv \Re e[u]$ is given by

$$p_{eq}(u_1) = N_0 \int_{-\infty}^{\infty} \frac{\mathrm{d}\xi}{(\xi^2 + \tilde{\sigma}_\gamma^2 u_1^2 + \tilde{\sigma}_b^2 + \tilde{\sigma}_u^2)^\delta}.$$
(120)

For an arbitrary $\delta > 0$ the integral in (120) can be expressed through the Hypergeometric function whose properties can be used to simplify the above expression

$$\boxed{p_{eq}(u_1) = \frac{N_0}{(\tilde{\sigma}_\gamma^2 u_1^2 + \tilde{\sigma}_u^2)^\kappa},} \qquad \kappa = \frac{1}{2} + \frac{\hat{\gamma} - \tfrac{1}{2}\tilde{\sigma}_\gamma^2}{\tilde{\sigma}_\gamma^2}.$$
(121)

However, a simpler although incomplete justification of the above fact can be obtained by evaluating the integral in (120) at $\hat{\gamma} = m\tilde{\sigma}_\gamma^2$, m a positive integer, so that $\delta = 1 + m$ and standard solutions apply.

Appendix C. **Truncated polynomial chaos expansions.** Here, we augment the discussion of the truncated Polynomial Chaos Expansions (PCE) in §64 with a few additional technical details; all of the information below is well known in the PCE literature.

The following result, often referred to as the Cameron-Martin theorem, forms the foundation of Wiener Chaos theory

Theorem 1 (Cameron-Martin [13]**)** *Assume that for fixed x and $s \leqslant t$, $u(x,s)$ is a functional of the Wiener process W on the interval $[0,\ s]$ with $\langle |u(x,s)|^2 \rangle < \infty$, then $u(x,s)$ has the following Fourier-Hermite expansion:*

$$u(x,s) = \sum_{\alpha \in \mathcal{J}} u_\alpha(x,s) T_\alpha, \quad u_\alpha(x,s) = \langle u(x,s) T_\alpha \rangle, \tag{122}$$

where T_α are the Wick polynomials defined via the Hermite polynomials $H_{\alpha_i}, \alpha_i \in \mathbb{N}$ by

$$T_\alpha(\boldsymbol{\xi}) = \prod_{i=1}^{\infty} H_{\alpha_i}(\xi_i). \tag{123}$$

The multiindex α in (123) with a finite number of nonzero components is defined as

$$\mathcal{J} = \left\{ \alpha = (\alpha_i,\ i \geqslant 1 \mid \alpha_i \in \{0, 1, 2, \dots\},\ |\alpha| = \sum_{i=1}^{\infty} \alpha_i < \infty) \right\}, \tag{124}$$

so that the product in the right hand side of (123) has a finite number of factors and it is well defined. Wick polynomials are defined over a in infinite dimensional space with variables $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots)$, where $\langle \xi_i \rangle = 0$, $\langle \xi_i \xi_j \rangle = \delta_{ij}$, and they form a complete orthonormal basis in L^2 on the probability space with respect to the Gaussian measure generated by $\boldsymbol{\xi}$. In particular

$$\langle T_\alpha T_\beta \rangle = \delta_{\alpha\beta}, \quad \langle T_0 \rangle = 1, \quad \langle T_\alpha \rangle = 0 \ \text{when} \ \alpha \neq 0. \tag{125}$$

The order of the Wick polynomial T_α is defined as $|\alpha| = \sum \alpha_i$. Furthermore, the first two statistical moments of $u(x,s)$ are given by:

$$\langle u(x,s) \rangle = u_0(x,s), \tag{126}$$

and

$$\langle u^2(x,s) \rangle = \sum_{\alpha \in \mathcal{J}} |u_\alpha(x,s)|^2. \tag{127}$$

The following theorem due to [95] is very useful for our analysis since it allows for exact treatment of nonlinearities

Theorem 2. *Suppose u, γ have Wiener Chaos expansions*

$$u = \sum_{\alpha \in \mathcal{J}} u_\alpha T_\alpha(\xi), \qquad \gamma = \sum_{\beta \in \mathcal{J}} \gamma_\alpha T_\beta(\xi). \tag{128}$$

If $\langle |uv|^2 \rangle < \infty$, then the product uv has the Wiener Chaos expansion

$$uv = \sum_{\theta \in \mathcal{J}} \left(\sum_{p \in \mathcal{J}} \sum_{0 \leqslant \beta \leqslant \theta} C(\theta, \beta, p) u_{\theta - \beta + p} \gamma_{\beta + p} \right) T_\theta(\xi), \tag{129}$$

where

$$C(\theta, \beta, p) = \left[\begin{pmatrix} \theta \\ \beta \end{pmatrix} \begin{pmatrix} \beta + p \\ p \end{pmatrix} \begin{pmatrix} \theta - \beta + p \\ p \end{pmatrix} \right]^{\frac{1}{2}}. \tag{130}$$

The operations on multi-indices are defined as

$$\alpha = \beta \quad \Leftrightarrow \quad \alpha_i = \beta_i \ \forall \, i, \tag{131}$$

$$\alpha \leqslant \beta \quad \Leftrightarrow \quad \alpha_i \leqslant \beta_i \ \forall \, i, \tag{132}$$

and

$$\alpha! = \prod_i \alpha_i! \tag{133}$$

The doubly infinite expansion (122) is useless in applications unless a suitable truncation is introduced through the truncated index set

$$\mathcal{J}_{K,N} = \left\{ \alpha = (\alpha_i, \ i \geqslant 1 \mid \alpha_i \in \{0, 1, 2, \dots, K\}, \ |\alpha| = \sum_{i=1}^{K} \alpha_i \leqslant N) \right\}, \tag{134}$$

so that the truncated PCE in (122) becomes

$$u_{K,N}(x, t, \xi) = \sum_{\alpha \in \mathcal{J}_{K,N}} u_\alpha(x, t) T_\alpha = \sum_{|\alpha| \leqslant N} u_\alpha(x, t) \prod_{i=1}^{K} H_{\alpha_i}(\xi_i). \tag{135}$$

The resulting approximation has altogether $\sum_{n=0}^{N} \begin{pmatrix} K + n - 1 \\ n \end{pmatrix}$ terms. Various sparse truncation methods have been proposed in this context [49, 114, 25]; such methods are problematic in systems with intermittency [10].

C.1. **Spectral representation of the white noise.** when considering systems driven by white noise a suitable spectral representation of the associated Wiener process is required in the PCE approximations. Here, we follow here the approach of [69, 70] and [49].

Consider any $T > 0$ and orthonormal basis $\{m_i(s), i = 1, 2, \dots\}$ in $L^2([0, T])$, for example trigonometric functions

$$m_1(s) = \frac{1}{\sqrt{T}}, \quad \text{and} \quad m_i(s) = \sqrt{\frac{2}{T}} \cos \frac{(i-1)\pi s}{T}, \ \text{for} \ i = 2, 3, \dots. \tag{136}$$

It is easy to see that the random variables defined by

$$\xi_i = \int_0^T m_i(s) \mathrm{d}W(s), \qquad i = 1, 2, \dots, \tag{137}$$

are i.i.d. standard Gaussian random variables, $\xi_i \sim \mathcal{N}(0, 1)$, so that the Wiener process can be represented by the following Fourier-Hermite expansion [49, 69, 70]

$$W(s) = \sum_{i=1}^{\infty} \xi_i \int_0^s m_i(\tau) \mathrm{d}W(\tau). \tag{138}$$

The expansion (138) converges in the mean-square sense

$$\left\langle \left[W(s) - \sum_{i=1}^{N} \xi_i \int_0^s m_i(\tau) \mathrm{d}W(\tau) \right]^2 \right\rangle \underset{N\to\infty}{\to} 0. \tag{139}$$

Appendix D. **Analytical formulas for the statistics of the turbulent tracer models.** Two new classes of non-Gaussian statistically exactly solvable test models for the turbulent diffusion of a tracer were introduced and discussed in §6. Below, we present the analytical formulas for the second-order statistics in these two systems.

First, the statistics of the turbulent velocity field with transient instabilities in the shear waves is derived; this velocity field drives both models (82) and (83) in §6. Next, the statistics of the tracer driven by the turbulent velocity field is derived. The general formulas for the non-Gaussian reactive tracer statistics are derived first and the statistics for the passive tracer are obtained by neglecting the binary reaction term in the general expressions.

D.1. **Statistics for the the velocity field in with intermittent transient instability.** The dynamics of the two-dimensional turbulent velocity field $(U(t), v(x,t))$ with intermittent transient instabilities driving the two turbulent tracer models is given by (82b-e) or (83c-f). The second-order statistics is derived systematically below.

PATH-WISE SOLUTIONS FOR THE CROSS-SWEEP $U(t)$

The path-wise solutions for the cross sweep $U(t) = \bar{U}(t) + U'(t)$ in (82b) and (83c) are given by

$$\bar{U}(t) = \int_{t_0}^t G_U(s,t) f_U(s) \mathrm{d}s, \tag{140}$$

$$U'(t) = G_U(t_0, t) U_0 + \sigma_U \int_{t_0}^t G_U(s,t) \mathrm{d}W_U(s), \tag{141}$$

where we use the short-hand notation for the initial condition $U_0 \equiv U(t = t_0)$ and for the Green's function associated with $U(t)$

$$G_U(s,t) = e^{-\gamma_U(t-s)}. \tag{142}$$

The initial conditions U_0 in the zonal jet are assumed Gaussian so that $U(t)$ is Gaussian.

PATH-WISE SOLUTIONS FOR THE SHEAR $v(x,t)$

The solution for each Fourier mode of the shear flow in the general case in (82) and (83) (correlated $U(t)$ and $v_k(t)$, in the presence of transient instabilities) has the form

$$v_k(t) = G_{v_k}(t_0, t) v_{k,0} + \int_{t_0}^t G_{v_k}(s,t)(b_{v_k}(s) + f_{v_k}(s)) \mathrm{d}s + \sigma_{v_k} \int_{t_0}^t G_{v_k}(s,t) \mathrm{d}W_{v_k}(s), \tag{143}$$

where $v_{k,0} \equiv v_k(t = t_0)$ and the Greens function associated with v_k is given by

$$G_{v_k}(s,t) = e^{-d_{v_k}(t-s)-I_k(s,t)+iJ_k(s,t)}, \tag{144}$$

$$I_k(s,t) = \int_s^t \gamma_{v_k}(s')\mathrm{d}s', \qquad J_k(s,t) = \int_s^t \omega_{v_k}(s')\mathrm{d}s'. \tag{145}$$

where d_{v_k} denotes the time-independent, deterministic part of the damping in (82c) and (83d).

The solutions for the damping fluctuations γ_{v_k} and the forcing fluctuations b_{v_k} in the shear velocity (82c) and (83d) are given by (Gaussian) Ornsteein Uhlenbeck processes and can be written in the same format as the solutions for the turbulent zonal jet U (see (140), (141)) but with the Greens functions

$$G_{b_{v_k}}(s,t) = e^{(-d_{b_{v_k}}+i\omega_{b_{v_k}})(t-s)}, \tag{146}$$

$$G_{\gamma_{v_k}}(s,t) = e^{-d_{\gamma_{v_k}}(t-s)}. \tag{147}$$

These solutions are needed for rewriting (145) in a way that simplifies the computations of the second-order statistics discussed below:

$$I_k(s,t) = \int_s^t \gamma_{v_k}(s')\mathrm{d}s' = M_{0,k}(s,t)\gamma_{v_k,0} + M_{W,k}(s,t), \tag{148}$$

$$M_{0,k}(s,t) = -\frac{1}{d_{\gamma_{v_k}}}\left(G_{\gamma_{v_k}}(t_0,t) - G_{\gamma_{v_k}}(t_0,s)\right), \tag{149}$$

$$M_{W,k}(s,t) = \sigma_{\gamma_{v_k}} \int_s^t \int_{t_0}^{s'} G_{\gamma_{v_k}}(r',s')\mathrm{d}W_{\gamma_{v_k}}(r')\mathrm{d}s', \tag{150}$$

$$J_k(s,t) = \int_s^t \omega_{v_k}(s')\mathrm{d}s' = a_k L(s,t) + b_k(t-s), \tag{151}$$

$$L(s,t) = \int_{t_0}^t U(s')\mathrm{d}s' = L_0(s,t)U_0 + L_f(s,t) + L_W(s,t), \tag{152}$$

$$L_0(s,t) = -\frac{1}{\gamma_U}\left(G_U(t_0,t) - G_U(t_0,s)\right), \tag{153}$$

$$L_f(s,t) = \int_s^t \int_{t_0}^{s'} G_U(r',s')f_U(r')\mathrm{d}r'\mathrm{d}s', \tag{154}$$

$$L_W(s,t) = \sigma_U \int_s^t \int_{t_0}^{s'} G_U(r',s')\mathrm{d}W_U(r')\mathrm{d}s'. \tag{155}$$

Second order statistics of the velocity field

The statistics of the Gaussian cross-sweep $U(t)$ can be easily found from the pathwise solutions (140)-(141) and it is given by

$$\langle U(t)\rangle = G_U(t_0,t)\langle U_0\rangle + \int_{t_0}^t G_U(s,t)f_U(s)\mathrm{d}s, \tag{156}$$

$$Var(U(t)) = G_U^2(t_0,t)Var(U_0) + \frac{\sigma_U^2}{2\gamma_U}\left(1 - G_U^2(t_0,t)\right). \tag{157}$$

The statistics of the damping fluctuations and the additive bias correction are easily obtained as

$$\langle \gamma_{v_k}(t) \rangle = G_{\gamma_{v_k}}(t_0, t) \langle \gamma_{v_k,0} \rangle, \tag{158}$$

$$\langle b_{v_k}(t) \rangle = G_{b_{v_k}}(t_0, t) \langle \gamma_{b_k,0} \rangle, \tag{159}$$

$$Var(\gamma_{v_k}(t)) = G_{\gamma_{v_k}}^2(t_0, t) Var(\gamma_{v_k,0}) + \frac{\sigma_{\gamma_{v_k}}^2}{2d_{\gamma_{v_k}}} \left(1 - G_{\gamma_{v_k}}^2(t_0, t)\right), \tag{160}$$

$$Var(b_{v_k}(t)) = |G_{b_{v_k}}(t_0, t)|^2 Var(b_{v_k,0}) + \frac{\sigma_{b_{v_k}}^2}{2d_{b_{v_k}}} \left(1 - |G_{b_{v_k}}(t_0, t)|^2\right), \tag{161}$$

$$Cov(b_{v_k}(t), b_{v_j}(t)) = G_{b_{v_k}}(t_0, t) G_{b_{v_j}}^*(t_0, t) Cov(b_{v_k,0}, b_{v_j,0}), \tag{162}$$

$$Cov(b_{v_k}(t), b_{v_j}^*(t)) = G_{b_{v_k}}(t_0, t) G_{b_{v_j}}(t_0, t) Cov(b_{v_k,0}, b_{v_j,0}^*), \tag{163}$$

$$Cov(b_{v_k}(t), \gamma_{v_j}(t)) = G_{b_{v_k}}(t_0, t) G_{\gamma_{v_j}}(t_0, t) Cov(b_{v_k,0}, \gamma_{v_j,0}) \tag{164}$$

$$Cov(b_{v_k}(t), U(t)) = G_{b_{v_k}}(t_0, t) G_U(t_0, t) Cov(b_{v_k,0}, U_0). \tag{165}$$

The mean shear $\langle v_k(t) \rangle$ can be expressed as

$$\langle v_k(t) \rangle = \langle G_{v_k}(t_0, t) v_{k,0} \rangle + \int_{t_0}^t \langle G_{v_k}(s, t) b_{v_k}(s) \rangle ds + \int_{t_0}^t \langle G_{v_k}(s, t) \rangle f_{v_k}(s) ds. \tag{166}$$

In order to evaluate the terms in (166), we use the fact that for any two complex Gaussian variables z and ξ the following identity holds

$$\langle z e^{\xi} \rangle = (\langle z \rangle + Cov(z, \xi^*)) e^{\langle \xi \rangle + \frac{1}{2} Cov(\xi, \xi^*)}, \tag{167}$$

where $Cov(z, \xi) = \langle z \xi^* \rangle - \langle z \rangle \langle \xi^* \rangle$ and $Cov(z, \xi^*) = \langle z \xi \rangle - \langle z \rangle \langle \xi \rangle$. Using (167), we obtain the following (for $s \leqslant t$)

$$\langle G_{v_k}(s, t) \rangle = e^{-d_{v_k}(t-s) - \langle I_{v_k}(s,t) \rangle + i \langle J_{v_k}(s,t) \rangle} \times$$
$$\times e^{\frac{1}{2} Var(I_{v_k}(s,t)) - \frac{1}{2} Var(J_{v_k}(s,t)) - i Cov(I_{v_k}(s,t), J_{v_k}(s,t))} \tag{168}$$

$$\langle G_{v_k}(t_0, t) v_{k,0} \rangle = \Big(\langle v_{k,0} \rangle - Cov(v_{k,0}, I_{v_k}(t_0, t))$$
$$+ i Cov(v_{k,0}, J_{v_k}(t_0, t)) \Big) \langle G_{v_k}(t_0, t) \rangle \tag{169}$$

$$\langle G_{v_k}(s, t) b_{v_k}(s) \rangle = \Big(\langle b_{v_k}(s) \rangle - Cov(b_{v_k}(s), I_{v_k}(s, t))$$
$$+ i Cov(b_{v_k}(s), J_{v_k}(s, t)) \Big) \langle G_{v_k}(s, t) \rangle \tag{170}$$

$$\langle I_{v_k}(s, t) \rangle = M_{0,k}(s, t) \langle \gamma_{v_k,0} \rangle, \tag{171}$$

$$Var(I_{v_k}(s, t)) = M_{0,k}^2(s, t) Var(\gamma_{v_k,0}) + Var(M_{W,k}(s, t)) \tag{172}$$

$$Var(M_{W,k}(s, t)) = \frac{k^2 \sigma_{\gamma_{v_k}}^2}{d_{\gamma_{v_k}}^3} \Big(d_{\gamma_{v_k}}(t - s) + e^{-d_{\gamma_{v_k}}(t-s)} + G_{\gamma_{v_k}}(2t_0, s + t)$$
$$- 1 - \tfrac{1}{2} \left(G_{\gamma_{v_k}}^2(t_0, t) + G_{\gamma_{v_k}}^2(t_0, s) \right) \Big) \tag{173}$$

$$\langle J_{v_k}(s,t)\rangle = a_k\langle L(s,t)\rangle + b_k(t-s) \tag{174}$$

$$Var(J_{v_k}(s,t)) = a_k^2 Var(L(s,t)) \tag{175}$$

$$\langle L(s,t)\rangle = L_f(s,t) + L_0(s,t)\langle U_0\rangle \tag{176}$$

$$Var(L(s,t)) = L_0^2(s,t)Var(U_0) + Var(L_W(s,t)) \tag{177}$$

$$Var(L_W(s,t)) = \frac{\sigma_U^2}{\gamma_U^3}\Big(\gamma_U(t-s) + e^{-\gamma_U(t-s)} + G_U(2t_0, s+t)$$
$$-1 - \tfrac{1}{2}\left(G_U^2(t_0,t) + G_U^2(t_0,s)\right)\Big) \tag{178}$$

$$Cov(I_{v_k}(s,t), J_{v_j}(r,t)) = M_{0,k}(s,t)L_0(r,t)Cov(\gamma_{v_k,0}, U_0) \tag{179}$$

$$Cov(I_{v_k}(s,t), I_{v_j}(r,t)) = M_{0,k}(s,t)M_{0,j}(r,t)Cov(\gamma_{v_k,0}, \gamma_{v_j,0}) + \delta_{kj}\delta_{sr}Var(M_{W,k}(s,t)) \tag{180}$$

$$Cov(J_{v_k}(s,t), J_{v_j}(r,t)) = a_k a_j\Big(L_0(s,t)L_0(r,t)Var(U_0) + \delta_{jk}\delta_{sr}Var(L_W(s,t))\Big) \tag{181}$$

$$Cov(b_{v_j}(r), I_{v_k}(s,t)) = \frac{1}{d_{\gamma_{v_k}}}\left(e^{-d_{\gamma_{v_k}}(s-t_0)} - e^{-d_{\gamma_{v_k}}(t-t_0)}\right) \times$$
$$\times\, e^{(-d_{b_{v_j}} + i\omega_{b_{v_j}})(r-t_0)}Cov(b_{v_j,0}, \gamma_{v_k,0}) \tag{182}$$

$$Cov(b_{v_j}(r), J_{v_k}(s,t)) = \frac{a_k}{d_U}\left(e^{-d_U(s-t_0)} - e^{-d_U(t-t_0)}\right) \times$$
$$\times\, e^{(-d_{b_{v_j}} + i\omega_{b_{v_j}})(r-t_0)}Cov(b_{v_j,0}, U_0) \tag{183}$$

The remaining correlations can be computed as follows:

$$\boxed{Var(v_k)} = Cov(v_k, v_k) = \langle|v_k(t)|^2\rangle - |\langle v_k(t)\rangle|^2$$

We rewrite $\langle|v_k(t)|^2\rangle$ as

$$\langle|v_k(t)|^2\rangle = \langle|A|^2\rangle + \langle|B|^2\rangle + \langle|C|^2\rangle + \Re e[\langle A^*B\rangle] \tag{184}$$

where

$$\langle|A|^2\rangle = \langle|G_{v_k}(t_0,t)|^2|v_{k,0}|^2\rangle \tag{185}$$

$$\langle|B|^2\rangle = \int_{t_0}^t\int_{t_0}^t\langle G_{v_k}(s,t)G_{v_k}^*(r,t)\big(b_{v_k}(s) + f_{v_k}(s)\big)\big(b_{v_k}^*(r) + f_{v_k}^*(r)\big)\rangle\mathrm{d}s\mathrm{d}r \tag{186}$$

$$\langle|C|^2\rangle = \sigma_{v_k}^2\int_{t_0}^t\langle|G_{v_k}(s,t)|^2\rangle\mathrm{d}s \tag{187}$$

$$\langle A^*B\rangle = 2\int_{t_0}^{t}\langle G_{v_k}(s,t)G_{v_k}^*(t_0,t)\big(b_{v_k}(s)+f_{v_k}(s)\big)v_{k,0}^*\rangle\mathrm{d}s \tag{188}$$

The above formulas can be evaluated based on the following identity which holds for any Gaussian variables z, w and ξ

$$\langle zwe^{\xi}\rangle = \Big(\langle z\rangle\langle w\rangle + Cov(z,w^*) + \langle z\rangle Cov(w,\xi^*)$$
$$+ \langle w\rangle Cov(z,\xi^*) + Cov(z,\xi^*)Cov(w,\xi^*)\Big)e^{\langle\xi\rangle+\frac{1}{2}Cov(\xi,\xi^*)}. \tag{189}$$

where $Cov(x,y) = \langle x\,y^*\rangle - \langle x\rangle\langle y^*\rangle$ and $Cov(x,y^*) = \langle x\,y\rangle - \langle x\rangle\langle y\rangle$. Using (189) we compute

$$\langle G_{v_k}(s,t)G_{v_j}(r,t)\rangle = e^{-d_{v_k}(t-s)-d_{v_j}(t-r)-\langle I_{v_k}(s,t)\rangle-\langle I_{v_j}(r,t)\rangle+i\langle J_{v_k}(s,t)\rangle+i\langle J_{v_j}(r,t)\rangle}\times$$
$$\times e^{\frac{1}{2}Var(I_{v_k}(s,t))-\frac{1}{2}Var(J_{v_k}(s,t))-iCov(I_k(s,t),J_k(s,t))}$$
$$\times e^{\frac{1}{2}Var(I_{v_j}(r,t))-\frac{1}{2}Var(J_{v_j}(r,t))-iCov(I_j(r,t),J_j(r,t))}$$
$$\times e^{Cov(I_{v_k}(s,t),I_{v_j}(r,t))-Cov(J_{v_k}(s,t),J_{v_j}(r,t))}$$
$$\times e^{-iCov(I_{v_k}(s,t),J_{v_j}(r,t))-iCov(J_{v_k}(s,t),I_{v_j}(r,t))} \tag{190}$$

$$\langle G_{v_k}(s,t)G_{v_j}^*(r,t)\rangle = e^{-d_{v_k}(t-s)-d_{v_j}(t-r)-\langle I_{v_k}(s,t)\rangle-\langle I_{v_j}(r,t)\rangle+i\langle J_{v_k}(s,t)\rangle-i\langle J_{v_j}(r,t)\rangle}\times$$
$$\times e^{\frac{1}{2}Var(I_{v_k}(s,t))-\frac{1}{2}Var(J_{v_k}(s,t))-iCov(I_k(s,t),J_k(s,t))}$$
$$\times e^{\frac{1}{2}Var(I_{v_j}(r,t))-\frac{1}{2}Var(J_{v_j}(r,t))+iCov(I_j(r,t),J_j(r,t))}$$
$$\times e^{Cov(I_{v_k}(s,t),I_{v_j}(r,t))+Cov(J_{v_k}(s,t),J_{v_j}(r,t))}$$
$$\times e^{iCov(I_{v_k}(s,t),J_{v_j}(r,t))-iCov(J_{v_k}(s,t),I_{v_j}(r,t))} \tag{191}$$

so that $\langle|A|^2\rangle, \langle|B|^2\rangle, \langle A^*B\rangle$ in (184) can be expressed as

$$\langle|A|^2\rangle = \langle|G_{v_k}(t_0,t)|^2\rangle\Big[|\langle v_{k,0}\rangle|^2 + Var(v_{k,0})$$
$$- 4\Re e\big[\langle v_{k,0}\rangle Cov(v_{k,0}^*,I_k(t_0,t))\big] + 4|Cov(v_{k,0},I_k(t_0,t))|^2\Big]$$

$$
\begin{aligned}
\langle |B|^2 \rangle = \ & \int_{t_0}^{t} \int_{t_0}^{t} \langle G_{v_k}(s,t) G_{v_k}^*(r,t) \rangle G_{b_{v_k}}(t_0,s) G_{b_{v_k}}^*(t_0,r) \Bigg[|\langle b_{v_k,0} \rangle|^2 + Var[b_{v_k,0}] \\
& - \langle b_{v_k,0}^* \rangle \Big\{ \Big(Cov(b_{v_k,0}, I_k(s,t)) + Cov(b_{v_k,0}, I_k(r,t)) \Big) - i \Big(Cov(b_{v_k,0}, J_k(s,t)) - Cov(b_{v_k,0}, J_k(r,t)) \Big) \Big\} \\
& - \langle b_{v_k,0} \rangle \Big\{ \Big(Cov(b_{v_k,0}^*, I_k(s,t)) + Cov(b_{v_k,0}^*, I_k(r,t)) \Big) - i \Big(Cov(b_{v_k,0}^*, J_k(s,t)) - Cov(b_{v_k,0}^*, J_k(r,t)) \Big) \Big\} \\
& + \Big(Cov[b_{v_k,0}, I_k(s,t)] + Cov[b_{v_k,0}, I_k(r,t)] - i(Cov[b_{v_k,0}, J_k(s,t)] - Cov[b_{v_k,0}, J_k(r,t)]) \Big) \times \\
& \times \Big(Cov[b_{v_k,0}^*, I_k(s,t)] + Cov[b_{v_k,0}^*, I_k(r,t)] - i(Cov[b_{v_k,0}^*, J_k(s,t)] - Cov[b_{v_k,0}^*, J_k(r,t)]) \Big) \Bigg] \mathrm{d}s \mathrm{d}r \\
& + \int_{t_0}^{t} \int_{t_0}^{t} \langle G_{v_k}(s,t) G_{v_k}^*(r,t) \rangle G_{b_{v_k}}^*(t_0,r) f_{v_k}(s) \Bigg[\langle b_{v_k,0}^* \rangle \\
& - \Big(Cov(b_{v_k,0}^*, I_k(s,t)) + Cov(b_{v_k,0}^*, I_k(r,t)) \Big) + i \Big(Cov(b_{v_k,0}^*, J_k(s,t)) - Cov(b_{v_k,0}^*, J_k(r,t)) \Big) \Bigg] \mathrm{d}s \mathrm{d}r \\
& + \int_{t_0}^{t} \int_{t_0}^{t} \langle G_{v_k}(s,t) G_{v_k}^*(r,t) \rangle G_{b_{v_k}}(t_0,s) f_{v_k}^*(r) \Bigg[\langle b_{v_k,0} \rangle \\
& - \Big(Cov(b_{v_k,0}, I_k(s,t)) + Cov(b_{v_k,0}, I_k(r,t)) \Big) + i \Big(Cov(b_{v_k,0}, J_k(s,t)) - Cov(b_{v_k,0}, J_k(r,t)) \Big) \Bigg] \mathrm{d}s \mathrm{d}r \\
& + \int_{t_0}^{t} \int_{t_0}^{t} \langle G_{v_k}(s,t) G_{v_k}^*(r,t) \rangle f_{v_k}(s) f_{v_k}^*(r) \mathrm{d}s \mathrm{d}r
\end{aligned}
$$

$$
\begin{aligned}
\langle A^* B \rangle = \ & 2 \int_{t_0}^{t} \langle G_{v_k}(s,t) G_{v_k}^*(t_0,t) \rangle G_{b_{v_k}}(t_0,s) \Bigg[\langle b_{v_k,0} \rangle \langle v_{k_0}^* \rangle + Cov[b_{v_k,0}, v_{k,0}] \\
& - \langle v_{k,0}^* \rangle \Big\{ \Big(Cov[b_{v_k,0}, I_k(s,t)] + Cov[b_{v_k,0}, I_k(t_0,t)] \Big) - i \Big(Cov[b_{v_k,0}, J_k(s,t)] - Cov[b_{v_k,0}, J_k(t_0,t)] \Big) \Big\} \\
& - \langle b_{v_k,0} \rangle \Big\{ \Big(Cov[v_{k,0}^*, I_k(s,t)] + Cov[v_{k,0}^*, I_k(t_0,t)] \Big) - i \Big(Cov[v_{k,0}^*, J_k(s,t)] - Cov[v_{k,0}^*, J_k(t_0,t)] \Big) \Big\} \\
& + (Cov[b_{v_k,0}, I_k(s,t)] + Cov[b_{v_k,0}, I_k(t_0,t)] - i(Cov[b_{v_k,0}, J_k(s,t)] - Cov[b_{v_k,0}, J_k(t_0,t)])) \times \\
& \times \Big(Cov[v_{k,0}^*, I_k(s,t)] + Cov[v_{k,0}^*, I_k(t_0,t)] - i(Cov[v_{k,0}^*, J_k(s,t)] - Cov[v_{k,0}^*, J_k(t_0,t)]) \Big) \Bigg] \mathrm{d}s \\
& + 2 \int_{t_0}^{t} \langle G_{v_k}(s,t) G_{v_k}^*(t_0,t) \rangle f_{v_k}(s) \Bigg[\langle v_{k,0}^* \rangle \\
& - \Big(Cov[v_{k,0}^*, I_k(s,t)] + Cov[v_{k,0}^*, I_k(t_0,t)] \Big) + i \Big(Cov[v_{k,0}^*, J_k(s,t)] - Cov[v_{k,0}^*, J_k(t_0,t)] \Big) \Bigg] \mathrm{d}s
\end{aligned}
$$

$$\boxed{\boldsymbol{Cov(v_k(t), v_j(t))}} = \langle v_k(t)v_j^*(t)\rangle - \langle v_k(t)\rangle\langle v_j^*(t)\rangle$$

We rewrite $\langle v_k(t)v_j^*(t)\rangle$ as

$$\langle v_k(t)v_j^*(t)\rangle = \langle A^2\rangle + \langle B^2\rangle + \langle AB\rangle, \tag{192}$$

where

$$\langle A^2\rangle = \langle G_{v_k}(t_0,t)G_{v_j}^*(t_0,t)v_{k,0}v_{j,0}^*\rangle \tag{193}$$

$$\langle B^2\rangle = \int_{t_0}^t\int_{t_0}^t \langle G_{v_k}(s,t)G_{v_j}^*(r,t)\big(b_{v_k}(s) + f_{v_k}(s)\big)\big(b_{v_j}^*(r) + f_{v_j}^*(r)\big)\rangle \mathrm{d}s\mathrm{d}r \tag{194}$$

$$\langle AB\rangle = \int_{t_0}^t \langle G_{v_k}(t_0,t)G_{v_j}^*(s,t)\big(b_{v_j}^*(s) + f_{v_j}^*(s)\big)v_{k,0}\rangle\mathrm{d}s$$

$$+ \int_{t_0}^t \langle G_{v_j}^*(t_0,t)G_{v_k}(s,t)\big(b_{v_k}(s) + f_{v_k}(s)\big)v_{j,0}^*\rangle\mathrm{d}s \tag{195}$$

The above formulas can be computed using (189).

$$
\begin{aligned}
\langle A^2 \rangle = \ & \langle G_{v_k}(t_0,t) G_{v_j}^*(t_0,t) \rangle \Big[\langle v_{k,0} \rangle \langle v_{j,0}^* \rangle + Cov[v_{k,0}, v_{j,0}] \\
& - \langle v_{k,0} \rangle \Big\{ \Big(Cov[v_{j,0}^*, I_k(t_0,t)] + Cov[v_{j,0}^*, I_j(t_0,t)] \Big) - i \Big(Cov[v_{j,0}^*, J_k(t_0,t)] - Cov[v_{j,0}^*, J_j(t_0,t)] \Big) \Big\} \\
& - \langle v_{j,0}^* \rangle \Big\{ \Big(Cov[v_{k,0}, I_k(t_0,t)] + Cov[v_{k,0}, I_j(t_0,t)] \Big) - i \Big(Cov[v_{k,0}, J_k(t_0,t)] - Cov[v_{k,0}, J_j(t_0,t)] \Big) \Big\} \\
& + \Big\{ \Big(Cov[v_{j,0}^*, I_k(t_0,t)] + Cov[v_{j,0}^*, I_j(t_0,t)] \Big) - i \Big(Cov[v_{j,0}^*, J_k(t_0,t)] - Cov[v_{j,0}^*, J_j(t_0,t)] \Big) \Big\} \times \\
& \quad \times \Big\{ \Big(Cov[v_{k,0}, I_k(t_0,t)] + Cov[v_{k,0}, I_j(t_0,t)] \Big) - i \Big(Cov[v_{k,0}, J_k(t_0,t)] - Cov[v_{k,0}, J_j(t_0,t)] \Big) \Big\} \Big]
\end{aligned}
$$

$$
\begin{aligned}
\langle B^2 \rangle = \ & \int_{t_0}^t \int_{t_0}^t \langle G_{v_k}(s,t) G_{v_j}^*(r,t) \rangle G_{b_{v_k}}(t_0,s) G_{b_{v_j}}^*(t_0,r) \Big[\langle b_{v_k,0} \rangle \langle b_{v_j,0}^* \rangle + Cov[b_{v_k,0}, b_{v_j,0}] \\
& - \langle b_{v_k,0} \rangle \Big\{ \Big(Cov[b_{v_j,0}^*, I_k(s,t)] + Cov[b_{v_j,0}^*, I_j(r,t)] \Big) - i \Big(Cov[b_{v_j,0}^*, J_k(s,t)] - Cov[b_{v_j,0}^*, J_j(r,t)] \Big) \Big\} \\
& - \langle b_{v_j,0}^* \rangle \Big\{ \Big(Cov[b_{v_k,0}, I_k(s,t)] + Cov[b_{v_k,0}, I_j(r,t)] \Big) - i \Big(Cov[b_{v_k,0}, J_k(s,t)] - Cov[b_{v_k,0}, J_j(r,t)] \Big) \Big\} \\
& + \Big\{ \Big(Cov[b_{v_j,0}^*, I_k(s,t)] + Cov[b_{v_j,0}^*, I_j(r,t)] \Big) - i \Big(Cov[b_{v_j,0}^*, J_k(s,t)] - Cov[b_{v_j,0}^*, J_j(r,t)] \Big) \Big\} \times \\
& \quad \times \Big\{ \Big(Cov[b_{v_k,0}, I_k(s,t)] + Cov[b_{v_k,0}, I_j(r,t)] \Big) - i \Big(Cov[b_{v_k,0}, J_k(s,t)] - Cov[b_{v_k,0}, J_j(r,t)] \Big) \Big\} \Big] \mathrm{d}s \mathrm{d}r \\
\\
& + \int_{t_0}^t \int_{t_0}^t \langle G_{v_k}(s,t) G_{v_j}^*(r,t) \rangle G_{b_{v_j}}^*(t_0,r) f_{v_k}(s) \Big[\langle b_{v_j,0}^* \rangle \\
& \quad - \Big(Cov(b_{v_j,0}^*, I_k(s,t)) + Cov(b_{v_j,0}^*, I_j(r,t)) \Big) + i \Big(Cov(b_{v_j,0}^*, J_k(s,t)) - Cov(b_{v_j,0}^*, J_j(r,t)) \Big) \Big] \mathrm{d}s \mathrm{d}r \\
\\
& + \int_{t_0}^t \int_{t_0}^t \langle G_{v_k}(s,t) G_{v_j}^*(r,t) \rangle G_{b_{v_k}}(t_0,s) f_{v_j}^*(r) \Big[\langle b_{v_k,0} \rangle \\
& \quad - \Big(Cov(b_{v_k,0}, I_k(s,t)) + Cov(b_{v_k,0}, I_j(r,t)) \Big) + i \Big(Cov(b_{v_k,0}, J_k(s,t)) - Cov(b_{v_k,0}, J_j(r,t)) \Big) \Big] \mathrm{d}s \mathrm{d}r \\
\\
& + \int_{t_0}^t \int_{t_0}^t \langle G_{v_k}(s,t) G_{v_j}^*(r,t) \rangle f_{v_k}(s) f_{v_j}^*(r) \mathrm{d}s \mathrm{d}r
\end{aligned}
$$

$$
\begin{aligned}
\langle AB \rangle = \ & \int_{t_0}^{t} \langle G_{v_k}(t_0,t) G_{v_j}^*(s,t) \rangle G_{b_{v_j}}^*(t_0,s) \Big[\langle b_{v_j,0}^* \rangle \langle v_{k,0} \rangle + Cov[v_{k,0}, b_{v_j,0}] \\
& - \langle v_{k,0} \rangle \Big\{ \Big(Cov[b_{v_j,0}^*, I_j(s,t)] + Cov[b_{v_j,0}^*, I_k(t_0,t)] \Big) + i \Big(Cov[b_{v_j,0}^*, J_j(s,t)] - Cov[b_{v_j,0}^*, J_k(t_0,t)] \Big) \Big\} \\
& - \langle b_{v_j,0}^* \rangle \Big\{ \Big(Cov[v_{k,0}, I_j(s,t)] + Cov[v_{k,0}, I_k(t_0,t)] \Big) + i \Big(Cov[v_{k,0}, J_j(s,t)] - Cov[v_{k,0}, J_k(t_0,t)] \Big) \Big\} \\
& + \Big\{ \Big(Cov[b_{v_j,0}^*, I_j(s,t)] + Cov[b_{v_j,0}^*, I_k(t_0,t)] \Big) + i \Big(Cov[b_{v_j,0}^*, J_j(s,t)] - Cov[b_{v_j,0}^*, J_k(t_0,t)] \Big) \Big\} \times \\
& \quad \times \Big\{ \Big(Cov[v_{k,0}, I_j(s,t)] + Cov[v_{k,0}, I_k(t_0,t)] \Big) + i \Big(Cov[v_{k,0}, J_j(s,t)] - Cov[v_{k,0}, J_k(t_0,t)] \Big) \Big\} \Big] \mathrm{d}s \\[2ex]
& + \int_{t_0}^{t} \langle G_{v_k}(t_0,t) G_{v_j}^*(s,t) \rangle f_{v_j}^*(s) \Big[\langle v_{k,0} \rangle \\
& - \Big(Cov(v_{k,0}, I_j(s,t)) + Cov(v_{k,0}, I_k(t_0,t)) - i \Big(Cov(v_{k,0}, J_j(s,t)) - Cov(v_{k,0}, J_k(t_0,t)) \Big) \Big] \mathrm{d}s \\[2ex]
& + \int_{t_0}^{t} \langle G_{v_j}^*(t_0,t) G_{v_k}(s,t) \rangle G_{b_{v_k}}(t_0,s) \Big[\langle b_{v_k,0} \rangle \langle v_{j,0} \rangle + Cov[b_{v_k,0}, v_{j,0}] \\
& - \langle v_{j,0}^* \rangle \Big\{ \Big(Cov[b_{v_k,0}, I_k(s,t)] + Cov[b_{v_k,0}, I_j(t_0,t)] \Big) - i \Big(Cov[b_{v_k,0}, J_k(s,t)] - Cov[b_{v_k,0}, J_j(t_0,t)] \Big) \Big\} \\
& - \langle b_{v_k,0} \rangle \Big\{ \Big(Cov[v_{j,0}, I_k(s,t)] + Cov[v_{j,0}, I_j(t_0,t)] \Big) - i \Big(Cov[v_{j,0}, J_k(s,t)] - Cov[v_{j,0}, J_j(t_0,t)] \Big) \Big\} \\
& + \Big\{ \Big(Cov[b_{v_k,0}, I_k(s,t)] + Cov[b_{v_k,0}, I_j(t_0,t)] \Big) - i \Big(Cov[b_{v_k,0}, J_k(s,t)] - Cov[b_{v_k,0}, J_j(t_0,t)] \Big) \Big\} \times \\
& \quad \times \Big\{ \Big(Cov[v_{j,0}, I_k(s,t)] + Cov[v_{j,0}, I_j(t_0,t)] \Big) - i \Big(Cov[v_{j,0}, J_k(s,t)] - Cov[v_{j,0}, J_j(t_0,t)] \Big) \Big\} \Big] \mathrm{d}s \\[2ex]
& + \int_{t_0}^{t} \langle G_{v_j}^*(t_0,t) G_{v_k}(s,t) \rangle f_{v_k}(s) \Big[\langle v_{j,0}^* \rangle \\
& - \Big(Cov(v_{j,0}^*, I_k(s,t)) + Cov(v_{j,0}^*, I_j(t_0,t)) + i \Big(Cov(v_{j,0}^*, J_k(s,t)) - Cov(v_{j,0}^*, J_j(t_0,t)) \Big) \Big] \mathrm{d}s
\end{aligned}
$$

$$\boxed{Cov(v_k(t), v_k^*(t))} = \langle v_k^2(t)\rangle - \langle v_k(t)\rangle^2$$

We find that

$$\langle v_k^2(t)\rangle = \langle A^2\rangle + \langle B^2\rangle + \langle AB\rangle \tag{196}$$

where

$$\langle A^2\rangle = \langle G_{v_k}^2(t_0,t)v_{k,0}^2\rangle \tag{197}$$

$$\langle B^2\rangle = \int_{t_0}^t \int_{t_0}^t \langle G_{v_k}(s,t)G_{v_k}(r,t)\big(b_{v_k}(s) + f_{v_k}(s)\big)\big(b_{v_k}(r) + f_{v_k}(r)\big)\rangle \mathrm{d}s\mathrm{d}r \tag{198}$$

$$\langle AB\rangle = 2\int_{t_0}^t \langle G_{v_k}(t_0,t)G_{v_k}(s,t)\big(b_{v_k}(s) + f_{v_k}(s)\big)v_{k,0}\rangle \mathrm{d}s \tag{199}$$

The above formulas can be computed using (189).

$$
\begin{aligned}
\langle A^2\rangle = \ & \langle G_{v_k}^2(t_0,t)\rangle\Bigg[\langle v_{k,0}\rangle^2 + Cov[v_{k,0}, v_{k,0}^*] \\[2mm]
& - 4\langle v_{k,0}\rangle\Big(Cov[v_{k,0}, I_k(t_0,t)] - iCov[v_{k,0}, J_k(t_0,t)]\Big) \\[2mm]
& + 4\Big(Cov[v_{k,0}, I_k(t_0,t)] - iCov[v_{k,0}, J_k(t_0,t)]\Big)^2\Bigg]
\end{aligned}
$$

$$
\begin{aligned}
\langle B^2\rangle = \ & \int_{t_0}^t \int_{t_0}^t \langle G_{v_k}(s,t)G_{v_k}(r,t)\rangle G_{b_{v_k}}(t_0,s)G_{b_{v_k}}(t_0,r)\Bigg[\langle b_{v_k,0}\rangle^2 + Cov[b_{v_k,0}, b_{v_k,0}^*] \\[2mm]
& - 2\langle b_{v_k,0}\rangle\Big\{\Big(Cov[b_{v_k,0}, I_k(s,t)] + Cov[b_{v_k,0}, I_k(r,t)]\Big) - i\Big(Cov[b_{v_k,0}, J_k(s,t)] + Cov[b_{v_k,0}, J_k(r,t)]\Big)\Big\} \\[2mm]
& + \Big\{\Big(Cov[b_{v_k,0}, I_k(s,t)] + Cov[b_{v_k,0}, I_k(r,t)]\Big) - i\Big(Cov[b_{v_k,0}, J_k(s,t)] + Cov[b_{v_k,0}, J_k(r,t)]\Big)\Big\}^2\Bigg]\mathrm{d}s\mathrm{d}r \\[2mm]
& + 2\int_{t_0}^t \int_{t_0}^t \langle G_{v_k}(s,t)G_{v_k}(r,t)\rangle G_{b_{v_k}}(t_0,r)f_{v_k}(s)\Bigg[\langle b_{v_k,0}\rangle \\[2mm]
& - \Big(Cov(b_{v_k,0}, I_k(s,t)) + Cov(b_{v_k,0}, I_k(r,t))\Big) + i\Big(Cov(b_{v_k,0}, J_k(s,t)) + Cov(b_{v_k,0}, J_k(r,t))\Big)\Bigg]\mathrm{d}s\mathrm{d}r \\[2mm]
& + \int_{t_0}^t \int_{t_0}^t \langle G_{v_k}(s,t)G_{v_k}(r,t)\rangle f_{v_k}(s)f_{v_k}(r)\mathrm{d}s\mathrm{d}r
\end{aligned}
$$

$$
\begin{aligned}
\langle AB \rangle = \ & 2\int_{t_0}^{t} \langle G_{v_k}(t_0,t)G_{v_k}(s,t)\rangle G_{b_{v_k}}(t_0,s)\Big[\langle b_{v_k,0}\rangle\langle v_{k_0}\rangle + Cov[b_{v_k,0},v_{k,0}^*] \\
& - \langle v_{k,0}\rangle\Big\{\Big(Cov[b_{v_k,0},I_k(s,t)]+Cov[b_{v_k,0},I_k(t_0,t)]\Big) - i\Big(Cov[b_{v_k,0},J_k(s,t)]+Cov[b_{v_k,0},J_k(t_0,t)]\Big)\Big\} \\
& - \langle b_{v_k,0}\rangle\Big\{\Big(Cov[v_{k,0},I_k(s,t)]+Cov[v_{k,0},I_k(t_0,t)]\Big) - i\Big(Cov[v_{k,0},J_k(s,t)]+Cov[v_{k,0},J_k(t_0,t)]\Big)\Big\} \\
& + \Big\{\Big(Cov[b_{v_k,0},I_k(s,t)]+Cov[b_{v_k,0},I_k(t_0,t)]\Big) - i\Big(Cov[b_{v_k,0},J_k(s,t)]+Cov[b_{v_k,0},J_k(t_0,t)]\Big)\Big\}\times \\
& \times \Big\{\Big(Cov[v_{k,0},I_k(s,t)]+Cov[v_{k,0},I_k(t_0,t)]\Big) - i\Big(Cov[v_{k,0},J_k(s,t)]+Cov[v_{k,0},J_k(t_0,t)]\Big)\Big\}\Big]\mathrm{d}s \\[6pt]
& + 2\int_{t_0}^{t}\langle G_{v_k}(t_0,t)G_{v_k}(s,t)\rangle f_{v_k}(s)\Big[\langle v_{k,0}\rangle \\
& - \Big(Cov(v_{k,0},I_k(s,t))+Cov(v_{k,0},I_k(t_0,t))\Big) + i\Big(Cov(v_{k,0},J_k(s,t))+Cov(v_{k,0},J_k(t_0,t))\Big)\Big]\mathrm{d}s
\end{aligned}
$$

$$(200)$$

$$\boxed{Cov(v_k(t),v_j^*(t))} = \langle v_k(t)v_j(t)\rangle - \langle v_k(t)\rangle\langle v_j(t)\rangle$$

We rewrite $\langle v_k(t)v_j(t)\rangle$ as

$$\langle v_k(t)v_j(t)\rangle = \langle A^2\rangle + \langle B^2\rangle + \langle AB\rangle, \tag{201}$$

where

$$\langle A^2\rangle = \langle G_{v_k}(t_0,t)G_{v_j}(t_0,t)v_{k,0}v_{j,0}\rangle \tag{202}$$

$$\langle B^2\rangle = \int_{t_0}^{t}\int_{t_0}^{t}\langle G_{v_k}(s,t)G_{v_j}(r,t)\big(b_{v_k}(s)+f_{v_k}(s)\big)\big(b_{v_j}(r)+f_{v_j}(r)\big)\rangle\mathrm{d}s\mathrm{d}r \tag{203}$$

$$
\begin{aligned}
\langle AB\rangle = \ & \int_{t_0}^{t}\langle G_{v_k}(t_0,t)G_{v_j}(s,t)\big(b_{v_j}(s)+f_{v_j}(s)\big)v_{k,0}\rangle\mathrm{d}s \\
& + \int_{t_0}^{t}\langle G_{v_j}(t_0,t)G_{v_k}(s,t)\big(b_{v_k}(s)+f_{v_k}(s)\big)v_{j,0}\rangle\mathrm{d}s \tag{204}
\end{aligned}
$$

The above formulas can be computed using ([189](#)).

$$
\begin{aligned}
\langle A^2\rangle = \ & \langle G_{v_k}(t_0,t)G_{v_j}(t_0,t)\rangle\Big[\langle v_{k,0}\rangle\langle v_{j,0}\rangle + Cov[v_{k,0},v_{j,0}^*] \\
& - \langle v_{k,0}\rangle\Big\{\Big(Cov[v_{j,0},I_k(t_0,t)]+Cov[v_{j,0},I_j(t_0,t)]\Big) - i\Big(Cov[v_{j,0},J_k(t_0,t)]+Cov[v_{j,0},J_j(t_0,t)]\Big)\Big\} \\
& - \langle v_{j,0}\rangle\Big\{\Big(Cov[v_{k,0},I_k(t_0,t)]+Cov[v_{k,0},I_j(t_0,t)]\Big) - i\Big(Cov[v_{k,0},J_k(t_0,t)]+Cov[v_{k,0},J_j(t_0,t)]\Big)\Big\} \\
& + \Big\{\Big(Cov[v_{j,0},I_k(t_0,t)]+Cov[v_{j,0},I_j(t_0,t)]\Big) - i\Big(Cov[v_{j,0},J_k(t_0,t)]+Cov[v_{j,0},J_j(t_0,t)]\Big)\Big\}\times \\
& \times \Big\{\Big(Cov[v_{k,0},I_k(t_0,t)]+Cov[v_{k,0},I_j(t_0,t)]\Big) - i\Big(Cov[v_{k,0},J_k(t_0,t)]+Cov[v_{k,0},J_j(t_0,t)]\Big)\Big\}\Big]
\end{aligned}
$$

$$
\begin{aligned}
\langle B^2 \rangle = \quad & \int_{t_0}^{t} \int_{t_0}^{t} \langle G_{v_k}(s,t) G_{v_j}(r,t) \rangle G_{b_{v_k}}(t_0,s) G_{b_{v_j}}(t_0,r) \Bigg[\langle b_{v_k,0} \rangle \langle b_{v_j,0} \rangle + Cov[b_{v_k,0}, b_{v_j,0}^*] \\
& - \langle b_{v_k,0} \rangle \Big\{ \Big(Cov[b_{v_j,0}, I_k(s,t)] + Cov[b_{v_j,0}, I_j(r,t)] \Big) - i \Big(Cov[b_{v_j,0}, J_k(s,t)] + Cov[b_{v_j,0}, J_j(r,t)] \Big) \Big\} \\
& - \langle b_{v_j,0} \rangle \Big\{ \Big(Cov[b_{v_k,0}, I_k(s,t)] + Cov[b_{v_k,0}, I_j(r,t)] \Big) - i \Big(Cov[b_{v_k,0}, J_k(s,t)] + Cov[b_{v_k,0}, J_j(r,t)] \Big) \Big\} \\
& + \Big\{ \Big(Cov[b_{v_j,0}, I_k(s,t)] + Cov[b_{v_j,0}, I_j(r,t)] \Big) - i \Big(Cov[b_{v_j,0}, J_k(s,t)] + Cov[b_{v_j,0}, J_j(r,t)] \Big) \Big\} \times \\
& \times \Big\{ \Big(Cov[b_{v_k,0}, I_k(s,t)] + Cov[b_{v_k,0}, I_j(r,t)] \Big) - i \Big(Cov[b_{v_k,0}, J_k(s,t)] + Cov[b_{v_k,0}, J_j(r,t)] \Big) \Big\} \Bigg] \mathrm{d}s \mathrm{d}r \\[2ex]
& + \int_{t_0}^{t} \int_{t_0}^{t} \langle G_{v_k}(s,t) G_{v_j}(r,t) \rangle G_{b_{v_j}}(t_0,r) f_{v_k}(s) \Bigg[\langle b_{v_j,0} \rangle \\[1ex]
& \qquad - \Big(Cov(b_{v_j,0}, I_k(s,t)) + Cov(b_{v_j,0}, I_j(r,t)) \Big) + i \Big(Cov(b_{v_j,0}, J_k(s,t)) + Cov(b_{v_j,0}, J_j(r,t)) \Big) \Bigg] \mathrm{d}s \mathrm{d}r \\[2ex]
& + \int_{t_0}^{t} \int_{t_0}^{t} \langle G_{v_k}(s,t) G_{v_j}(r,t) \rangle G_{b_{v_k}}(t_0,s) f_{v_j}(r) \Bigg[\langle b_{v_k,0} \rangle \\[1ex]
& \qquad - \Big(Cov(b_{v_k,0}, I_k(s,t)) + Cov(b_{v_k,0}, I_j(r,t)) \Big) + i \Big(Cov(b_{v_k,0}, J_k(s,t)) + Cov(b_{v_k,0}, J_j(r,t)) \Big) \Bigg] \mathrm{d}s \mathrm{d}r \\[2ex]
& + \int_{t_0}^{t} \int_{t_0}^{t} \langle G_{v_k}(s,t) G_{v_j}(r,t) \rangle f_{v_k}(s) f_{v_j}(r) \mathrm{d}s \mathrm{d}r
\end{aligned}
$$

$$
\begin{aligned}
\langle AB \rangle = \ & \int_{t_0}^{t} \langle G_{v_k}(t_0,t) G_{v_j}(s,t) \rangle G_{b_{v_j}}(t_0,s) \Big[\langle b_{v_j,0} \rangle \langle v_{k,0} \rangle + Cov[b_{v_j,0}, v_{k,0}^*] \\
& - \langle v_{k,0} \rangle \Big\{ \Big(Cov[b_{v_j,0}, I_j(s,t)] + Cov[b_{v_j,0}, I_k(t_0,t)] \Big) - i \Big(Cov[b_{v_j,0}, J_j(s,t)] + Cov[b_{v_j,0}, J_k(t_0,t)] \Big) \Big\} \\
& - \langle b_{v_j,0} \rangle \Big\{ \Big(Cov[v_{k,0}, I_j(s,t)] + Cov[v_{k,0}, I_k(t_0,t)] \Big) - i \Big(Cov[v_{k,0}, J_j(s,t)] + Cov[v_{k,0}, J_k(t_0,t)] \Big) \Big\} \\
& + \Big\{ \Big(Cov[b_{v_j,0}, I_j(s,t)] + Cov[b_{v_j,0}, I_k(t_0,t)] \Big) - i \Big(Cov[b_{v_j,0}, J_j(s,t)] + Cov[b_{v_j,0}, J_k(t_0,t)] \Big) \Big\} \times \\
& \qquad \times \Big\{ \Big(Cov[v_{k,0}, I_j(s,t)] + Cov[v_{k,0}, I_k(t_0,t)] \Big) - i \Big(Cov[v_{k,0}, J_j(s,t)] + Cov[v_{k,0}, J_k(t_0,t)] \Big) \Big\} \Big] ds \\[2ex]
& + \int_{t_0}^{t} \langle G_{v_k}(t_0,t) G_{v_j}(s,t) \rangle f_{v_j}(s) \Big[\langle v_{k,0} \rangle \\
& - \Big(Cov(v_{k,0}, I_j(s,t)) + Cov(v_{k,0}, I_k(t_0,t)) + i \Big(Cov(v_{k,0}, J_j(s,t)) + Cov(v_{k,0}, J_k(t_0,t)) \Big) \Big] ds \\[2ex]
& + \int_{t_0}^{t} \langle G_{v_j}(t_0,t) G_{v_k}(s,t) \rangle G_{b_{v_k}}(t_0,s) \Big[\langle b_{v_k,0} \rangle \langle v_{j,0} \rangle + Cov[b_{v_k,0}, v_{j,0}^*] \\
& - \langle v_{j,0} \rangle \Big\{ \Big(Cov[b_{v_k,0}, I_k(s,t)] + Cov[b_{v_k,0}, I_j(t_0,t)] \Big) - i \Big(Cov[b_{v_k,0}, J_k(s,t)] + Cov[b_{v_k,0}, J_j(t_0,t)] \Big) \Big\} \\
& - \langle b_{v_k,0} \rangle \Big\{ \Big(Cov[v_{j,0}, I_k(s,t)] + Cov[v_{j,0}, I_j(t_0,t)] \Big) - i \Big(Cov[v_{j,0}, J_k(s,t)] + Cov[v_{j,0}, J_j(t_0,t)] \Big) \Big\} \\
& + \Big\{ \Big(Cov[b_{v_k,0}, I_k(s,t)] + Cov[b_{v_k,0}, I_j(t_0,t)] \Big) - i \Big(Cov[b_{v_k,0}, J_k(s,t)] + Cov[b_{v_k,0}, J_j(t_0,t)] \Big) \Big\} \times \\
& \qquad \times \Big\{ \Big(Cov[v_{j,0}, I_k(s,t)] + Cov[v_{j,0}, I_j(t_0,t)] \Big) - i \Big(Cov[v_{j,0}, J_k(s,t)] + Cov[v_{j,0}, J_j(t_0,t)] \Big) \Big\} \Big] ds \\[2ex]
& + \int_{t_0}^{t} \langle G_{v_j}(t_0,t) G_{v_k}(s,t) \rangle f_{v_k}(s) \Big[\langle v_{j,0} \rangle \\
& - \Big(Cov(v_{j,0}, I_k(s,t)) + Cov(v_{j,0}, I_j(t_0,t)) + i \Big(Cov(v_{j,0}, J_k(s,t)) + Cov(v_{j,0}, J_j(t_0,t)) \Big) \Big] ds
\end{aligned}
$$

$$\boxed{Cov(v_k(t), b_{v_k}^*(t))} = \langle v_k b_{v_k} \rangle - \langle v_k \rangle \langle b_k \rangle,$$

where

$$\langle v_k(t) b_{v_k}(t) \rangle = G_{b_{v_k}}(t_0, t) \langle G_{v_k}(t_0, t) v_{k,0} b_{v_k,0} \rangle$$

$$+ \int_{t_0}^{t} \langle G_{v_k}(s, t)(b_{v_k}(s) + f_{v_k}(s)) b_{v_k}(t) \rangle \mathrm{d}s$$

$$= \langle G_{v_k}(t_0, t) \rangle G_{b_{vk}}(t_0, t) \Bigg[\langle v_{k,0} \rangle \langle b_{v_k,0} \rangle + Cov[v_{k,0}, b_{v_{k,0}}^*]$$

$$- \langle b_{v_k,0} \rangle \Big(Cov[v_{k,0}, I_k(t_0, t)] - iCov[v_{k,0}, J_k(t_0, t)] \Big)$$

$$- \langle v_{k,0} \rangle \Big(Cov[b_{v_k,0}, I_k(t_0, t)] - iCov[b_{v_k,0}, J_k(t_0, t)] \Big)$$

$$+ \Big(Cov[v_{k,0}, I_k(t_0, t)] - iCov[v_{k,0}, J_k(t_0, t)] \Big) \times$$

$$\times \Big(Cov[b_{v_k,0}, I_k(t_0, t)] - iCov[b_{v_k,0}, J_k(t_0, t)] \Big) \Bigg]$$

$$+ \int_{t_0}^{t} \langle G_{v_k}(s, t) \rangle G_{b_{vk}}(t_0, t) G_{b_{vk}}(t_0, s) \Big[\langle b_{v_k,0} \rangle^2 + Cov[b_{v_k,0}, b_{v_k,0}^*]$$

$$- 2\langle b_{v_k,0} \rangle \Big(Cov[b_{v_k,0}, I_k(s, t)] - iCov[b_{v_k,0}, J_k(s, t)] \Big)$$

$$+ \Big(Cov[b_{v_k,0}, I_k(s, t)] - iCov[b_{v_k,0}, J_k(s, t)] \Big)^2 \Big] \mathrm{d}s$$

$$+ \int_{t_0}^{t} \langle G_{v_k}(s, t) \rangle G_{b_{vk}}(t_0, t) f_{v_k}(s) \Big[\langle b_{v_k,0} \rangle$$

$$- Cov[b_{v_k,0}, I_k(s, t)] + iCov[b_{v_k,0}, J_k(s, t)] \Big] \mathrm{d}s$$

$$\boxed{Cov(v_k(t), b_{v_j}^*(t))} = \langle v_k b_{v_j}\rangle - \langle v_k\rangle\langle b_{v_j}\rangle$$

where

$$\langle v_k(t) b_{v_j}(t)\rangle = G_{b_{v_j}}(t_0, t)\langle G_{v_k}(t_0, t) v_{k,0} b_{v_j,0}\rangle$$
$$+ \int_{t_0}^{t} \langle G_{v_k}(s, t)(b_{v_k}(s) + f_{v_k}(s)) b_{v_j}(t)\rangle \mathrm{d}s$$

$$= \langle G_{v_k}(t_0, t)\rangle G_{b_{v_j}}(t_0, t) \Bigg[\langle v_{k,0}\rangle\langle b_{v_j,0}\rangle + Cov[v_{k,0}, b_{v_j,0}^*]$$
$$- \langle b_{v_j,0}\rangle\Big(Cov[v_{k,0}, I_k(t_0, t)] - iCov[v_{k,0}, J_k(t_0, t)]\Big)$$
$$- \langle v_{k,0}\rangle\Big(Cov[b_{v_j,0}, I_k(t_0, t)] - iCov[b_{v_j,0}, J_k(t_0, t)]\Big)$$
$$+ \Big(Cov[b_{v_j,0}, I_k(t_0, t)] - iCov[b_{v_j,0}, J_k(t_0, t)]\Big) \times$$
$$\times \Big(Cov[v_{k,0}, I_k(t_0, t)] - iCov[v_{k,0}, J_k(t_0, t)]\Big)\Bigg]$$

$$+ \int_{t_0}^{t} \langle G_{v_k}(s, t)\rangle G_{b_{v_j}}(t_0, t) G_{b_{vk}}(t_0, s)\Big[\langle b_{v_k,0}\rangle\langle b_{v_j,0}\rangle + Cov[b_{v_k,0}, b_{v_j,0}^*]$$
$$- \langle b_{v_k,0}\rangle\Big(Cov[b_{v_j,0}, I_k(s, t)] - iCov[b_{v_j,0}, J_k(s, t)]\Big)$$
$$- \langle b_{v_j,0}\rangle\Big(Cov[b_{v_k,0}, I_k(s, t)] - iCov[b_{v_k,0}, J_k(s, t)]\Big)$$
$$+ \Big(Cov[b_{v_k,0}, I_k(s, t)] - iCov[b_{v_k,0}, J_k(s, t)]\Big) \times$$
$$\times \Big(Cov[b_{v_j,0}, I_k(s, t)] - iCov[b_{v_j,0}, J_k(s, t)]\Big)\Big]\mathrm{d}s$$

$$+ \int_{t_0}^{t} \langle G_{v_k}(s, t)\rangle G_{b_{v_j}}(t_0, t) f_{v_k}(s)\Big[\langle b_{v_j,0}\rangle$$
$$- Cov[b_{v_j,0}, I_k(s, t)] + iCov[b_{v_j,0}, J_k(s, t)]\Big]\mathrm{d}s$$

$$\boxed{Cov(v_k(t), b_{v_k}(t))} = \langle v_k b_{v_k}^* \rangle - \langle v_k \rangle \langle b_{v_k}^* \rangle$$

where

$$\langle v_k(t) b_{v_k}^*(t) \rangle = \langle G_{v_k}(t_0, t) G_{b_{v_k}}^*(t_0, t) v_{k,0} b_{v_k,0}^* \rangle$$
$$+ \int_{t_0}^t \langle G_{v_k}(s, t)(b_{v_k}(s) + f_{v_k}(s)) b_{v_k}^*(t) \rangle \mathrm{d}s$$

$$= G_{b_{vk}}^*(t_0, t) \langle G_{v_k}(t_0, t) \rangle \Big[\langle v_{k,0} \rangle \langle b_{v_k,0}^* \rangle + Cov[v_{k,0}, b_{v_k,0}]$$
$$- \langle b_{v_k,0}^* \rangle \Big(Cov[v_{k,0}, I_k(t_0, t)] - i Cov[v_{k,0}, J_k(t_0, t)] \Big)$$
$$- \langle v_{k,0} \rangle \Big(Cov[b_{v_k,0}^*, I_k(t_0, t)] - i Cov[b_{v_k,0}^*, J_k(t_0, t)] \Big)$$
$$+ \Big(Cov[v_{k,0}, I_k(t_0, t)] - i Cov[v_{k,0}, J_k(t_0, t)] \Big) \times$$
$$\times \Big(Cov[b_{v_k,0}^*, I_k(t_0, t)] - i Cov[b_{v_k,0}^*, J_k(t_0, t)] \Big) \Big]$$

$$+ \int_{t_0}^t \langle G_{v_k}(s, t) \rangle G_{b_{vk}}^*(t_0, t) G_{b_{vk}}(t_0, s) \Big[|\langle b_{v_k,0} \rangle|^2 + Var(b_{v_k,0})$$
$$- \langle b_{v_k,0}^* \rangle \Big(Cov[b_{v_k,0}, I_k(s, t)] - i Cov[b_{v_k,0}, J_k(s, t)] \Big)$$
$$- \langle b_{v_k,0} \rangle \Big(Cov[b_{v_k,0}^*, I_k(s, t)] - i Cov[b_{v_k,0}^*, J_k(s, t)] \Big)$$
$$+ \Big(Cov[b_{v_k,0}, I_k(s, t)] - i Cov[b_{v_k,0}, J_k(s, t)] \Big) \times$$
$$\times \Big(Cov[b_{v_k,0}^*, I_k(s, t)] - i Cov[b_{v_k,0}^*, J_k(s, t)] \Big) \Big] \mathrm{d}s$$

$$+ \int_{t_0}^t \langle G_{v_k}(s, t) \rangle G_{b_{vk}}^*(t_0, t) f_{v_k}(s) \Big[\langle b_{v_k,0}^* \rangle$$
$$- Cov[b_{v_k,0}^*, I_k(s, t)] + i Cov[b_{v_k,0}^*, J_k(s, t)] \Big] \mathrm{d}s$$

$$\boxed{Cov(v_k(t), b_{v_j}(t))} = \langle v_k b_{v_j}^* \rangle - \langle v_k \rangle \langle b_{v_j}^* \rangle$$

where

$$\langle v_k(t) b_{v_j}^*(t) \rangle = \langle G_{v_k}(t_0,t) G_{b_{v_j}}^*(t_0,t) v_{k,0} b_{v_j,0}^* \rangle$$

$$+ \int_{t_0}^{t} \langle G_{v_k}(s,t)(b_{v_k}(s) + f_{v_k}(s)) b_{v_j}^*(t) \rangle \mathrm{d}s$$

$$= G_{b_{v_j}}^*(t_0,t) \langle G_{v_k}(t_0,t) \rangle \Big[\langle v_{k,0} \rangle \langle b_{v_j,0}^* \rangle + Cov[v_{k,0}, b_{v_j,0}]$$

$$- \langle b_{v_j,0}^* \rangle \Big(Cov[v_{k,0}, I_k(t_0,t)] - iCov[v_{k,0}, J_k(t_0,t)] \Big)$$

$$- \langle v_{k,0} \rangle \Big(Cov[b_{v_j,0}^*, I_k(t_0,t)] - iCov[b_{v_j,0}^*, J_k(t_0,t)] \Big)$$

$$+ \Big(Cov[v_{k,0}, I_k(t_0,t)] - iCov[v_{k,0}, J_k(t_0,t)] \Big) \times$$

$$\times \Big(Cov[b_{v_j,0}^*, I_k(t_0,t)] - iCov[b_{v_j,0}^*, J_k(t_0,t)] \Big) \Big]$$

$$+ \int_{t_0}^{t} \langle G_{v_k}(s,t) \rangle G_{b_{v_j}}^*(t_0,t) G_{b_{v_k}}(t_0,s) \Big[\langle b_{v_k,0} \rangle \langle b_{v_j,0}^* \rangle + Cov(b_{v_k,0}, b_{v_j,0})$$

$$- \langle b_{v_j,0}^* \rangle \Big(Cov[b_{v_k,0}, I_k(s,t)] - iCov[b_{v_k,0}, J_k(s,t)] \Big)$$

$$- \langle b_{v_k,0} \rangle \Big(Cov[b_{v_j,0}^*, I_k(s,t)] - iCov[b_{v_j,0}^*, J_k(s,t)] \Big)$$

$$+ \Big(Cov[b_{v_k,0}, I_k(s,t)] - iCov[b_{v_k,0}, J_k(s,t)] \Big) \times$$

$$\times \Big(Cov[b_{v_j,0}^*, I_k(s,t)] - iCov[b_{v_j,0}^*, J_k(s,t)] \Big) \Big] \mathrm{d}s$$

$$+ \int_{t_0}^{t} \langle G_{v_k}(s,t) \rangle G_{b_{v_j}}^*(t_0,t) f_{v_k}(s) \Big[\langle b_{v_j,0}^* \rangle$$

$$- Cov[b_{v_j,0}^*, I_k(s,t)] + iCov[b_{v_j,0}^*, J_k(s,t)] \Big] \mathrm{d}s$$

$$\boxed{Cov(v_k(t), \gamma_{v_k}(t))} \qquad = \langle v_k \gamma_{v_k} \rangle - \langle v_k \rangle \langle \gamma_{v_k} \rangle,$$

where

$$\langle v_k(t)\gamma_{v_k}(t)\rangle = G_{\gamma_{v_k}}(t_0,t)\langle G_{v_k}(t_0,t)v_{k,0}\gamma_{v_k,0}\rangle + \int_{t_0}^t \langle G_{v_k}(s,t)(b_{v_k}(s) + f_{v_k}(s))\gamma_{v_k}(t)\rangle \mathrm{d}s$$

$$= G_{\gamma_{vk}}(t_0,t)\langle G_{v_k}(t_0,t)\rangle \Big[\langle v_{k,0}\rangle \langle \gamma_{v_k,0}\rangle + Cov[v_{k,0}, \gamma_{v_k,0}]$$

$$- \langle \gamma_{v_k,0}\rangle \Big(Cov[v_{k,0}, I_k(t_0,t)] - iCov[v_{k,0}, J_k(t_0,t)] \Big) - \langle v_{k,0}\rangle \Big(Cov[\gamma_{v_k,0}, I_k(t_0,t)] - iCov[\gamma_{v_k,0}, J_k(t_0,t)] \Big)$$

$$+ \Big(Cov[\gamma_{v_k,0}, I_k(t_0,t)] - iCov[\gamma_{v_k,0}, J_k(t_0,t)] \Big) \times \Big(Cov[v_{k,0}, I_k(t_0,t)] - iCov[v_{k,0}, J_k(t_0,t)] \Big) \Big]$$

$$+ \int_{t_0}^t \langle G_{v_k}(s,t)\rangle G_{b_{vk}}(t_0,s)G_{\gamma_{v_k}}(t_0,t) \Big[\langle b_{v_k,0}\rangle \langle \gamma_{v_k,0}\rangle + Cov[b_{v_k,0}, \gamma_{v_k,0}]$$

$$- \langle b_{v_k,0}\rangle \Big(Cov[\gamma_{v_k,0}, I_k(s,t)] - iCov[\gamma_{v_k,0}, J_k(s,t)] \Big) - \langle \gamma_{v_k,0}\rangle \Big(Cov[b_{v_k,0}, I_k(s,t)] - iCov[b_{v_k,0}, J_k(s,t)] \Big)$$

$$+ \Big(Cov[b_{v_k,0}, I_k(s,t)] - iCov[b_{v_k,0}, J_k(s,t)] \Big) \times \Big(Cov[\gamma_{v_k,0}, I_k(s,t)] - iCov[\gamma_{v_k,0}, J_k(s,t)] \Big) \Big] \mathrm{d}s$$

$$+ \int_{t_0}^t \langle G_{v_k}(s,t)\rangle G_{\gamma_{vk}}(t_0,t)f_{v_k}(s) \Big[\langle \gamma_{v_k,0}\rangle - Cov[\gamma_{v_k,0}, I_k(s,t)] + iCov[\gamma_{v_k,0}, J_k(s,t)] \Big] \mathrm{d}s$$

$$\boxed{Cov(v_k(t), \gamma_{v_j}(t))} \qquad = \langle v_k \gamma_{v_j} \rangle - \langle v_k \rangle \langle \gamma_{v_j} \rangle,$$

where

$$\langle v_k(t)\gamma_{v_j}(t)\rangle = G_{\gamma_{v_j}}(t_0,t)\langle G_{v_k}(t_0,t)v_{k,0}\gamma_{v_j,0}\rangle + \int_{t_0}^t \langle G_{v_k}(s,t)(b_{v_k}(s) + f_{v_k}(s))\gamma_{v_j}(t)\rangle \mathrm{d}s$$

$$= G_{\gamma_{v_j}}(t_0,t)\langle G_{v_k}(t_0,t)\rangle \Big[\langle v_{k,0}\rangle \langle \gamma_{v_j,0}\rangle + Cov[v_{k,0}, \gamma_{v_j,0}]$$

$$- \langle \gamma_{v_j,0}\rangle \Big(Cov[v_{k,0}, I_k(t_0,t)] - iCov[v_{k,0}, J_k(t_0,t)] \Big) - \langle v_{k,0}\rangle \Big(Cov[\gamma_{v_j,0}, I_k(t_0,t)] - iCov[\gamma_{v_j,0}, J_k(t_0,t)] \Big)$$

$$+ \Big(Cov[\gamma_{v_j,0}, I_k(t_0,t)] - iCov[\gamma_{v_j,0}, J_k(t_0,t)] \Big) \times \Big(Cov[v_{k,0}, I_k(t_0,t)] - iCov[v_{k,0}, J_k(t_0,t)] \Big) \Big]$$

$$+ \int_{t_0}^t \langle G_{v_k}(s,t)\rangle G_{b_{vk}}(t_0,s)G_{\gamma_{v_j}}(t_0,t) \Big[\langle b_{v_k,0}\rangle \langle \gamma_{v_j,0}\rangle + Cov[b_{v_k,0}, \gamma_{v_j,0}]$$

$$- \langle b_{v_k,0}\rangle \Big(Cov[\gamma_{v_j,0}, I_k(s,t)] - iCov[\gamma_{v_j,0}, J_k(s,t)] \Big) - \langle \gamma_{v_j,0}\rangle \Big(Cov[b_{v_k,0}, I_k(s,t)] - iCov[b_{v_k,0}, J_k(s,t)] \Big)$$

$$+ \Big(Cov[b_{v_k,0}, I_k(s,t)] - iCov[b_{v_k,0}, J_k(s,t)] \Big) \times \Big(Cov[\gamma_{v_j,0}, I_k(s,t)] - iCov[\gamma_{v_j,0}, J_k(s,t)] \Big) \Big] \mathrm{d}s$$

$$+ \int_{t_0}^t \langle G_{v_k}(s,t)\rangle G_{\gamma_{v_j}}(t_0,t)f_{v_k}(s) \Big[\langle \gamma_{v_j,0}\rangle - Cov[\gamma_{v_j,0}, I_k(s,t)] + iCov[\gamma_{v_j,0}, J_k(s,t)] \Big] \mathrm{d}s$$

$$\boxed{Cov(v_k(t), U(t))} \qquad = \langle v_k U \rangle - \langle v_k \rangle \langle U \rangle,$$

where

$$\begin{aligned}
\langle v_k U \rangle = \ & G_U(t_0, t) \langle G_{v_k}(t_0, t) v_{k,0} U_0 \rangle + \int_{t_0}^t G_U(t_0, t) \langle G_{v_k}(b_{v_k}(s) + f_{v_k}(s)) U_0 \rangle \mathrm{d}s \\
= \ & G_U(t_0, t) \langle G_{v_k}(t_0, t) \rangle \Big[\langle v_{k,0} \rangle \langle U_0 \rangle + Cov[v_{k,0}, U_0] \\
& \qquad - \langle U_0 \rangle Cov[v_{k,0}, I_k(t_0, t)] + i \langle U_0 \rangle Cov[v_{k,0}, J_k(t_0, t)] \\
& \qquad - \langle v_{k,0} \rangle Cov[U_0, I_k(t_0, t)] + i \langle v_{k,0} \rangle Cov[U_0, J_k(t_0, t)] \\
& \qquad + \Big(Cov[U_0, I_k(t_0, t)] - iCov[U_0, J_k(t_0, t)] \Big) \times \\
& \qquad\qquad \times \Big(Cov[v_{k,0}, I_k(t_0, t)] - iCov[v_{k,0}, J_k(t_0, t)] \Big) \Big] \\
& + \int_{t_0}^t \langle G_{v_k}(s, t) \rangle G_U(t_0, t) G_{b_{v_k}}(t_0, s) \Big[\langle b_{v_k, 0} \rangle \langle U_0 \rangle + Cov[b_{v_k, 0}, U_0] \\
& \qquad - \langle b_{v_k, 0} \rangle Cov[U_0, I_k(s, t)] + i \langle b_{v_k, 0} \rangle Cov[U_0, J_k(s, t)] \\
& \qquad - \langle U_0 \rangle Cov[b_{v_k, 0}, I_k(s, t)] + i \langle U_0 \rangle Cov[b_{v_k, 0}, J_k(s, t)] \\
& \qquad + \Big(Cov[b_{v_k, 0}, I_k(s, t)] - iCov[b_{v_k, 0}, J_k(s, t)] \Big) \times \\
& \qquad\qquad \times \Big(Cov[U_0, I_k(s, t)] - iCov[U_0, J_k(s, t)] \Big) \Big] \mathrm{d}s \\
& + \int_{t_0}^t \langle G_{v_k}(s, t) \rangle G_U(t_0, t) f_{v_k}(s) \Big[\langle U_0 \rangle \\
& \qquad - Cov[U_0, I_k(s, t)] + iCov[U_0, J_k(s, t)] \Big] \mathrm{d}s
\end{aligned}$$

D.2. Statistics for the passive tracer fluctuations in the mean gradient driven by a turbulent velocity field.

As already mentioned earlier, we derive here the tracer statistics for the reactive tracer in model (83); the statistics for the passive tracer is obtained from the general formulas by setting the reaction rate $r = 0$.

The path-wise solutions for the k-th tracer mode in (83a) are given by

$$T_k(t) = G_{T_k}(t_0, t) T_{k,0} - \alpha \int_{t_0}^t G_{T_k}(s, t) v_k(s) \mathrm{d}s, \tag{205}$$

where $T_{k,0} = T_k(t = t_0)$ and the Green's function associated with (83a) is given by

$$G_{T_k}(s, t) = e^{-d_{T_k}(t-s) - ikL_U(s,t)}, \tag{206}$$

$$L_U(s, t) = \int_s^t U(s') \mathrm{d}s'. \tag{207}$$

The second-order statistics for the velocity field was derived in the previous section. Here, we show how to compute the remaining statistics including the tracer modes.

The mean for the tracer is given by

$$
\langle T_k(t)\rangle = \langle G_{T_k}(t_0,t)T_{k,0}\rangle - \alpha \int_{t_0}^t \langle G_{T_k}(s,t)G_{v_k}(t_0,s)v_{k,0}\rangle \mathrm{d}s
$$

$$
- \alpha \int_{t_0}^t \int_{t_0}^s \langle G_{T_k}(s,t)G_{v_k}(r,s)(b_{v_k}(r)+f_{v_k}(r))\rangle \mathrm{d}r\mathrm{d}s \quad (208)
$$

where $\langle G_{T_k}(s,t)G_{v_k}(r,s)\rangle$ is a characteristic function of a Gaussian and can be computed in a similar way to G_{v_k} leading to

$$
\langle G_{T_k}(s,t)G_{v_k}(r,s)\rangle = e^{-\tilde{d}_k(r,s,t)-\langle I_{v_k}(r,s)\rangle -i\langle \tilde{J}_k(r,s,t)\rangle} \times
$$

$$
\times\, e^{\frac{1}{2}Var[I_{v_k}(r,s)]-\frac{1}{2}Var[\tilde{J}_k(r,s,t)]-iCov[I_{v_k}(r,s),\tilde{J}_k(r,s,t)]} \quad (209)
$$

where

$$
\tilde{d}_k(r,s,t) = d_{T_k}(t-s) + d_{v_k}(s-r), \quad (210)
$$

$$
\tilde{J}_k(r,s,t) = kL_U(s,t) - J_{v_k}(r,s). \quad (211)
$$

$\boxed{Cov[T_k(t),T_j(t)]}$

Using (205) we find

$$
Cov[T_k(t),T_j(t)] = \langle G_{T_k}(t_0,t)G^*_{T_j}(t_0,t)T_{k,0}T^*_{j,0}\rangle - \langle G_{T_k}(t_0,t)T_{k,0}\rangle\langle G^*_{T_j}(t_0,t)T^*_{j,0}\rangle
$$

$$
+ \alpha^2 \int_{t_0}^t \int_{t_0}^t \Big(\langle G_{T_k}(s,t)G^*_{T_j}(r,t)v_k(s)v^*_j(r)\rangle - \langle G_{T_k}(s,t)v_k(s)\rangle\langle G^*_{T_j}(r,t)v^*_j(r)\rangle \Big)\mathrm{d}s\mathrm{d}r
$$

$$
- \alpha \int_{t_0}^t \Big(\langle G_{T_k}(t_0,t)G^*_{T_j}(r,t)T_{k,0}v^*_j(r)\rangle - \langle G_{T_k}(t_0,t)T_{k,0}\rangle\langle G^*_{T_j}(r,t)v^*_j(r)\rangle \Big)\mathrm{d}r
$$

$$
- \alpha \int_{t_0}^t \Big(\langle G^*_{T_j}(t_0,t)G_{T_k}(r,t)T^*_{j,0}v_k(r)\rangle - \langle G^*_{T_j}(t_0,t)T^*_{j,0}\rangle\langle G_{T_k}(r,t)v_k(r)\rangle \Big)\mathrm{d}r \quad (212)
$$

Note that all terms in (212) have the form $\langle ze^\xi\rangle$ or $\langle zwe^\xi\rangle$ where z,w,ξ are Gaussian variables; consequently, all terms in (212) can be evaluated using the identities (167) and (189).

In the statistically steady state, the mean of the tracer becomes

$$
\langle T_k(t)\rangle = -\alpha \int_{-\infty}^t \int_{-\infty}^s \langle G_{T_k}(s,t)G_{v_k}(r,s)\rangle f_{v_k}(r)\mathrm{d}r\mathrm{d}s \quad (213)
$$

and the tracer spectrum can be obtained from

$$
Cov(T_j(t),T_k(t)) = \alpha^2 \int_{-\infty}^t \int_{-\infty}^t \Big(\langle G_{T_j}(s,t)G^*_{T_k}(s',t)v_j(s)v^*_k(s')\rangle
$$

$$
- \langle G_{T_j}(s,t)v_j(s)\rangle\langle G_{T_k}(s',t)v_k(s')\rangle^* \Big)\mathrm{d}s\mathrm{d}s' \quad (214)
$$

where

$$\langle G_{T_j}(s,t)G^*_{T_k}(s',t)v_j(s)v^*_k(s')\rangle =$$

$$\int_{-\infty}^{s}\int_{-\infty}^{s'}\langle G_{v_j}(r,s)G^*_{v_k}(r',s')G_{T_j}(s,t)G^*_{T_k}(s',t)\rangle f_{v_j}(r)f^*_{v_k}(r')\mathrm{d}r\mathrm{d}r'$$

$$+\sigma^2_{v_k}\delta^j_k\int_{-\infty}^{\min(s,s')}\langle G_{T_k}(s,t)G^*_{T_k}(s',t)G_{v_k}(r,s)G^*_{v_k}(r,s')\rangle\mathrm{d}r' \qquad (215)$$

$\boxed{Cov(T_k(t),U(t))}$

We find that

$$Cov(T_k(t),U(t)) = G_U(t_0,t)\big(\langle G_{T_k}(t_0,t)T_{k,0}U_0\rangle - \langle G_{T_k}(t_0,t)T_{k,0}\rangle\langle U_0\rangle\big)$$

$$-\alpha G_U(t_0,t)\int_{t_0}^{t}\big(\langle G_{T_k}(s,t)v_k(s)U_0\rangle - \langle G_{T_k}(s,t)v_k(s)\rangle\langle U_0\rangle\big)\mathrm{d}s$$

$$(216)$$

$\boxed{Cov(T_k(t),v_j(t))}$

$$Cov(T_k(t),v_j(t)) = \langle G_{T_k}(t_0,t)G^*_{v_j}(t_0,t)T_{k,0}v^*_{j,0}\rangle - \langle G_{T_k}(t_0,t)T_{k,0}\rangle\langle G^*_{v_j}(t_0,t)v^*_{j,0}\rangle$$

$$-\alpha\int_{t_0}^{t}\int_{t_0}^{t}\Big(\langle G_{T_k}(s,t)G^*_{v_j}(r,t)v_k(s)(b^*_{v_j}(r)+f^*_{v_j}(r))\rangle - \langle G_{T_k}(s,t)v_k(s)\rangle\langle G^*_{v_j}(r,t)(b^*_{v_j}(r)+f^*_{v_j}(r))\rangle\Big)\mathrm{d}s\mathrm{d}r$$

$$+\int_{t_0}^{t}\Big(\langle G_{T_k}(t_0,t)G^*_{v_j}(r,t)(b^*_{v_j}(r)+f^*_{v_j}(r))T_{k,0}\rangle - \langle G_{T_k}(t_0,t)T_{k,0}\rangle\langle G^*_{v_j}(r,t)(b^*_{v_j}(r)+f^*_{v_j}(r))\rangle\Big)\mathrm{d}r$$

$$-\alpha\int_{t_0}^{t}\Big(\langle G_{T_k}(s,t)G^*_{v_j}(t_0,t)v_k(s)v^*_{j,0}\rangle - \langle G_{T_k}(s,t)v_k(s)\rangle\langle G^*_{v_j}(t_0,t)v^*_{j,0}\rangle\Big)\mathrm{d}r \quad (217)$$

$\boxed{Cov(T_k(t),v^*_j(t))}$

$$Cov(T_k(t),v^*_j(t)) = \langle G_{T_k}(t_0,t)G_{v_j}(t_0,t)T_{k,0}v_{j,0}\rangle - \langle G_{T_k}(t_0,t)T_{k,0}\rangle\langle G_{v_j}(t_0,t)v_{j,0}\rangle$$

$$-\alpha\int_{t_0}^{t}\int_{t_0}^{t}\big(\langle G_{T_k}(s,t)G_{v_j}(r,t)v_k(s)(b_{v_j}(r)+f_{v_j}(r))\rangle - \langle G_{T_k}(s,t)v_k(s)\rangle\langle G_{v_j}(r,t)(b_{v_j}(r)+f_{v_j}(r))\rangle\big)\mathrm{d}s\mathrm{d}r$$

$$+\int_{t_0}^{t}\big(\langle G_{T_k}(t_0,t)G_{v_j}(r,t)(b_{v_j}(r)+f_{v_j}(r))T_{k,0}\rangle - \langle G_{T_k}(t_0,t)T_{k,0}\rangle\langle G_{v_j}(r,t)(b_{v_j}(r)+f_{v_j}(r))\rangle\big)\mathrm{d}r$$

$$-\alpha\int_{t_0}^{t}\big(\langle G_{T_k}(s,t)G_{v_j}(t_0,t)v_k(s)v_{j,0}\rangle - \langle G_{T_k}(s,t)v_k(s)\rangle\langle G_{v_j}(t_0,t)v_{j,0}\rangle\big)\mathrm{d}r \quad (218)$$

$\boxed{Cov(T_k(t),\gamma_{v_j}(t))}$

$$Cov(T_k(t), \gamma_{v_j}(t)) = G_{\gamma_{v_j}}(t_0, t) \left(\langle G_{T_k}(t_0, t) T_{k,0} \gamma_{v_j,0} \rangle - \langle G_{T_k}(t_0, t) T_{k,0} \rangle \langle \gamma_{v_j,0} \rangle \right)$$

$$- \alpha \, G_{\gamma_{v_j}}(t_0, t) \int_{t_0}^{t} \left(\langle G_{T_k}(s, t) v_k(s) \gamma_{v_j,0} \rangle - \langle G_{T_k}(s, t) v_k(s) \rangle \langle \gamma_{v_j,0} \rangle \right) ds$$

$$(219)$$

$$\boxed{Cov(T_k(t), b_{v_j}(t))}$$

$$Cov(T_k(t), b_{v_j}(t)) = G_{b_{v_j}}^*(t_0, t) \left(\langle G_{T_k}(t_0, t) T_{k,0} b_{v_j,0}^* \rangle - \langle G_{T_k}(t_0, t) T_{k,0} \rangle \langle b_{v_j,0}^* \rangle \right)$$

$$- \alpha \, G_{b_{v_j}}^*(t_0, t) \int_{t_0}^{t} \left(\langle G_{T_k}(s, t) v_k(s) b_{v_j,0}^* \rangle - \langle G_{T_k}(s, t) v_k(s) \rangle \langle b_{v_j,0}^* \rangle \right) ds$$

$$(220)$$

$$\boxed{Cov(T_k(t), b_{v_j}^*(t))}$$

$$Cov(T_k(t), b_{v_j}(t)) = G_{b_{v_j}}(t_0, t) \left(\langle G_{T_k}(t_0, t) T_{k,0} b_{v_j,0} \rangle - \langle G_{T_k}(t_0, t) T_{k,0} \rangle \langle b_{v_j,0} \rangle \right)$$

$$(221)$$

$$- \alpha \, G_{b_{v_j}}(t_0, t) \int_{t_0}^{t} \left(\langle G_{T_k}(s, t) v_k(s) b_{v_j,0} \rangle - \langle G_{T_k}(s, t) v_k(s) \rangle \langle b_{v_j,0} \rangle \right) ds$$

$$(222)$$

As already mentioned earlier all the terms in the above equations can be evaluated using (189).

D.3. **White noise limits of the turbulent Rossby waves in the turbulent tracer dynamics.** Here, we consider the dynamics of the turbulent tracer discussed in §6 in the limit of rapidly decorrelating shear waves; this procedure is similar to that discussed in [92] and in Appendix B.2. The limiting dynamics, described by (230), was used in §7 to illustrate possible numerical facts in long time integrations of stochastic systems.

Consider the simplified model (82) with $b_{v_k} \equiv 0$ for turbulent diffusion of a passive tracer with mean gradient

$$a) \quad \dot{T}_k'(t) \quad = (-d_{T_k} + i\omega_{T_k}(t)) T_k'(t) - \alpha v_k(t),$$

$$b) \quad dU(t) \quad = \left[-d_U U(t) + f_U(t) \right] dt + \sigma_U dW_U(t),$$

$$c) \quad dv_k(t) \quad = \left[(-d_{v_k} - \gamma_{v_k}(t) + i\omega_{v_k}(t)) v_k(t) + f_{v_k}(t) \right] dt + \sigma_{v_k} dW_{v_k}(t),$$

$$d) \quad d\gamma_{v_k}(t) \quad = -d_{\gamma_{v_k}} \gamma_{v_k}(t) dt + \sigma_{\gamma_{v_k}} dW_{\gamma_{v_k}}(t),$$

$$(223)$$

where $\omega_{T_k} = -U(t)k$ and $\omega_{v_k} = a_k U(t) + b_k$ as discussed in §6. Here, we consider a limiting dynamics of (223) when the waves v_k decorrelate very fast and can be effectively considered as white noise.

First, decompose the waves into two parts

$$v_k(t) = \bar{v}_{k|(U,\gamma_{v_k})}(t) + v'_{k|(U,\gamma_{v_k})}(t), \tag{224}$$

where $\bar{v}_{k|(U,\gamma_{v_k})}$ is a conditional mean of the shear flow for given realization of the cross-sweep and the damping fluctuations. The second part, $v'_{k|(U,\gamma_{v_k})}$, in (224) denotes fluctuations of the shear around the conditional mean. From (223) the dynamics of the conditional mean is given by

$$\frac{\mathrm{d}\bar{v}_{k|(U,\gamma_{v_k})}}{\mathrm{d}t} = (-\hat{d}_{v_k} - \gamma_{v_k}(t) + i\omega_{v_k}(t))\bar{v}_{k|(U,\gamma_{v_k})} + f_{v_k}(t), \tag{225}$$

and the fluctuations around the mean satisfy the following Ito SDE

$$\mathrm{d}\bar{v}'_{k|(U,\gamma_{v_k})} = (-\hat{d}_{v_k} - \gamma_{v_k}(t) + i\omega_{v_k}(t))v'_{k|(U,\gamma_{v_k})} + \sigma_{v_k}\mathrm{d}W_{v_k}(t). \tag{226}$$

Following the analogous procedure to that presented in Appendix B.2 and [92], in the limit when

$$d_{v_k} \to \infty, \qquad \sigma_{v_k} \to \infty, \qquad \eta_{v_k} = \sigma_{v_k}/d_{v_k} = const.,$$

the fluctuations $v'_{k|(U,\gamma_{v_k})}$ approach white noise

$$\mathrm{d}v'_{k|(U,\gamma_{v_k})}(t) \to \eta_{v_k}\mathrm{d}W_{v_k}(t). \tag{227}$$

In order to determine the mean $\bar{v}_{k|(U,\gamma_{v_k})}$ in the white noise limit, suppose that the forcing f_{v_k} grows as the mean dissipation $\hat{d}_{v_k}$ in the white noise limit, i.e. $f_{v_k} = d_{v_k}\bar{f}_{v_k}$, where $\bar{f}_k$ is independent of d_{v_k}. Then, using the general solution

$$\bar{v}_{k|(U,\gamma_{v_k})} = G_{v_k}(t_0,t)v_{k|(U,\gamma_{v_k})}(t_0) + \int_{t_0}^{t} G_{v_k}(s,t)f_{v_k}(s)\mathrm{d}s \tag{228}$$

we find that $\bar{v}_{k|(U,\gamma_{v_k})} \to \bar{f}_{v_k}(t)$ for $d_{v_k} \to \infty$. Consequently, the shear velocity (224) in the white noise limit of rapidly decorrelating fluctuations can be written as

$$v_k(t) = \bar{f}_{v_k}(t) + \eta_{v_k}\dot{W}_{v_k}(t), \tag{229}$$

so that the model for the turbulent tracer becomes

$$\begin{aligned} a) \quad \mathrm{d}T'_k(t) &= \left[(-d_{T_k} + i\omega_{T_k}(t))T'_k(t) - \alpha\bar{f}_{v_k}(t)\right]\mathrm{d}t - \alpha\eta_{v_k}\mathrm{d}W_{v_k}(t), \\ b) \quad \mathrm{d}U(t) &= \left[-d_U U(t) + f_U(t)\right]\mathrm{d}t + \sigma_U\mathrm{d}W_U(t). \end{aligned} \tag{230}$$

where $\omega_{T_k}(t) = -kU(t)$, α is the large scale tracer gradient and $\bar{f}_{v_k}$ is a known deterministic function.

The system (230) with the turbulent finitely correlated zonal jet and additive white noise is exactly solvable and the second-order statistics for the system (230) can be found in a similar way to that for the full system (83) by utilizing the identities for $\langle zwe^{\xi}\rangle$ with Gaussian variables z, w, ξ.

INVARIANT MEASURES FOR THE SYSTEM (230)

It is important to note that the system (230) has a Gaussian invariant measure even for the turbulent Gaussian zonal jet $U(t)$. This can be deduced by applying the following transformation to (230)

$$T'_k = e^{i\int_{t_0}^{t}\omega_{T_k}(r)\mathrm{d}r}\tilde{T}_k + \alpha\int_{t_0}^{t}e^{-d_{T_k}(t-s)+i\int_{s}^{t}\omega_{T_k}(r)\mathrm{d}r}\bar{f}_{v_k}(s)\mathrm{d}s, \tag{231}$$

which after applying the rules of Ito calculus leads to

$$d\tilde{T}_k(t) = -d_{T_k}\tilde{T}_k(t)dt - \alpha\eta_{v_k}d\tilde{W}_{v_k}(t), \tag{232}$$

which has complex Gaussian solutions (for Gaussian initial data) with independent components since $\tilde{W}_{v_k}(t) = \frac{1}{2}\left(\tilde{W}_{v_k}^R(t) + i\tilde{W}_{v_k}^I(t)\right)$ has independent components. It is easy to see that T'_k is also Gaussian for arbitrary σ_U provided that either $\alpha = 0$ or $\bar{f}_{v_k} = 0$.

REFERENCES

[1] R. V. Abramov and A. J. Majda, *Blended response algorithms for linear fluctuation-dissipation for complex nonlinear dynamical systems*, Nonlinearity, **20** (2007), 2793–2821.

[2] R. V. Abramov, A. J. Majda and R. Kleeman, *Information theory and predictability for low-frequency variability*, J. Atmos. Sci., **62** (2005), 65–87.

[3] A. Alexanderian, O. Le Maitre, H. Najm, M. Iskandarani and O. Knio, *Multiscale Stochastic Preconditioners in Non-intrusive Spectral Projection*, Journal of Scientific Computing, (2011), 1–35.

[4] L. Arnold, "Random Dynamical Systems," Springer Monographs in Mathematics, Springer-Verlag, Berlin, 1998.

[5] M. Avellaneda and A. J. Majda, *Approximate and exact renormalization theories for a model for turbulent transport*, Phys. Fluids A, **4** (1992), 41–57.

[6] M. Avellaneda and A. J. Majda, *Renormalization theory for eddy diffusivity in turbulent transport*, Phys. Rev. Lett., **68** (1992), 3028–3031.

[7] M. Avellaneda and A. J. Majda, *Simple examples with features of renormalization for turbulent transport*, Phil. Trans. R. Soc. Lond. A, **346** (1994), 205–233.

[8] A. Bourlioux and A. J. Majda, *An elementary model for the validation of flamelet approximations in non-premixed turbulent combustion*, Combust. Theory Model., **4** (2000), 189–210.

[9] M. Branicki, B. Gershgorin and A. J. Majda, *Filtering skill for turbulent signals for a suite of nonlinear and linear kalman filters*, J. Comp. Phys, **231** (2012), 1462–1498.

[10] M. Branicki and A. J. Majda, *Fundamental Limitations of Polynomial Chaos Expansions for Uncertainty Quantification in Systems with Intermittent Instabilities*, Comm. Math. Sci., accepted and in press, 2012.

[11] M. Branicki and A. J. Majda, *Quantifying uncertainty for long range forecasting scenarios with model errors in non-Guassian models with intermittency*, Nonlinearity, submitted, 2012.

[12] G. Branstator and H. Teng, *Two limits of initial-value decadal predictability in a CGCM*, J. Climate, **23** (2010), 6292–6311.

[13] R. H. Cameron and W. T. Martin, *The orthogonal development of non-linear functionals in series of Fourier-Hermite functionals*, Ann. Math. (2), **48** (1947), 385–392.

[14] M. D. Chekroun, D. Kondrashov and M. Ghil, *Predicting stochastic systems by noise sampling, and application to the El Niño-Southern Oscillation*, PNAS, **108** (2011), 11766–11771.

[15] A. J. Chorin, *Gaussian fields and random flow*, J. Fluid Mech., **63** (1974), 21–32.

[16] S. C. Crow and G. H. Canavan, *Relationship between a Wiener-Hermite expansion and an energy cascade*, J. Fluid Mech., **41** (1970), 387–403.

[17] T. DelSole, *Predictability and information theory. I: Measures of predictability*, J. Atmos. Sci., **61** (2004), 2425–2440.

[18] T. DelSole, *Stochastic models of quasigeostrophic turbulence*, Surveys in Geophys., **25** (2004), 107–149.

[19] T. DelSole, *Predictability and information theory. II: Imperfect models*, J. Atmos. Sci., **62** (2005), 3368–3381.

[20] T. DelSole and J. Shukla, *Model fidelity versus skill in seasonal forecasting*, J. Climate, **23** (2010), 4794–4806.

[21] K. A. Emanuel, J. C. Wyngaard, J. C. McWilliams, D. A. Randall and Y. L. Yung, "Improving the Scientific Foundation for Atmosphere-Land Ocean Simulations," Natl. Acad. Press, Washington DC, 2005.

[22] E. S. Epstein, *Stochastic dynamic predictions*, Tellus, **21** (1969), 739–759.

[23] C. Franzke, D. Crommelin, A. Fischer and A. J. Majda, *A hidden Markov model perspective on regimes and metastability in atmospheric flows*, J. Climate, **21** (2008), 1740–1757.

[24] C. Franzke, I. Horenko, A. J. Majda and R. Klein, *Systematic metastable regime identification in an AGCM*, J. Atmos. Sci., **66** (2009), 1997–2012.

[25] P. Frauenfelder, C. Schwab and R. A. Todor, *Finite elements for elliptic problems with stochastic coefficients*, Comput. Methods Appl. Mech. Eng., **194** (2005), 205–228.

[26] D. Frierson, *Robust increases in midlatitude static stability in global warming simulations*, Geophys. Res. Lett., **33** (2006), L24816.

[27] D. Frierson, *Midlatitude static stability in simple and comprehensive general circulation models*, J. Atmos. Sci., **65** (2008), 1049–1062.

[28] C. Gardiner, "Stochastic Methods: A Handbook for the Natural and Social Sciences," Fourth edition, Springer Series in Synergetics, Springer-Verlag, Berlin, 2009.

[29] 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.

[30] B. Gershgorin, J. Harlim and A. J. Majda, *Test models for improving filtering with model errors through stochastic parameter estimation*, J. Comp. Phys., **229** (2010), 1–31.

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

[32] B. Gershgorin and A. J. Majda, *Quantifying uncertainty for climate change and long range forecasting scenarios with model errors. I: Gaussian models,* J. Climate, in press, 2012.

[33] B. Gershgorin and A. J. Majda, *A nonlinear test model for filtering slow-fast systems*, Comm. Math. Sci., **6** (2008), 611–650.

[34] B. Gershgorin and A. J. Majda, *Filtering a nonlinear slow-fast system with strong fast forcing*, Comm. Math. Sci., **8** (2010), 67–92.

[35] 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.

[36] R. G. Ghanem and P. D. Spanos, "Stochastic Finite Elements: A Spectral Approach," Springer-Verlag, New York, 1991.

[37] D. Giannakis and A. J. Majda, *Quantifying the predictive skill in long-range forecasting. I: Coarse-grained predictions in a simple ocean model*, J. Climate, in press, 2011.

[38] D. Giannakis and A. J. Majda, *Quantifying the predictive skill in long-range forecasting. II: Model error in coarse-grained Markov models with application to ocean-circulation regimes*, J. Climate, in press, 2011.

[39] D. Giannakis, A. J. Majda and I. Horenko, *Information theory, model error, and predictive skill of stochastic models for complex nonlinear systems*, Physica D, submitted, 2011.

[40] J. P. Gollub, J. Clarke, M. Gharib, B. Lane and O. N. Mesquita, *Fluctuations and transport in a stirred fluid with a mean gradient*, Phys. Rev. Lett., **67** (1991), 3507–3510.

[41] M. Hairer and A. J. Majda, *A simple framework to justify linear response theory*, Nonlinearity, **23** (2010), 909–922.

[42] J. Harlim and A. J. Majda, *Filtering turbulent sparsely observed geophysical flows*, Mon. Wea. Rev., **138** (2010), 1050–1083.

[43] H. G. Matthies and A. Keese, *Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations*, Comput. Methods Appl. Mech. Eng., **194** (2005), 1295–1331.

[44] I. Horenko, *On clustering of non-stationary meteorological time series*, Dyn. Atmos. Oceans, **49** (2010), 164–187.

[45] I. Horenko, *On robust estimation of low-frequency variability trends in discrete Markovian sequences of atmospheric circulation patterns*, J. Atmos. Sci., **66** (2009), 2059–2072.

[46] I. Horenko, *Finite element approach to clustering of multidimensional time series*, SIAM J. Sci. Comput., **32** (2010), 62–83.

[47] I. Horenko, *On the identification of nonstationary factor models and their application to atmospheric data analysis*, J. Atmos. Sci., **67** (2010), 1559–1574.

[48] I. Horenko, *Nonstationarity in multifactor models of discrete jump processes, memory and application to cloud modeling*, J. Atmos. Sci., 2011.

[49] T. Hou, W. Luo, B. Rozovskii and H.-M. Zhou, *Wiener Chaos expansions and numerical solutions of randomly forced equations of fluid mechanics*, J. Comp. Phys., **216** (2006), 687–706.

[50] Jayesh and Z. Warhaft, *Probability distribution of a passive scalar in grid-generated turbulence*, Phys. Rev. Lett., **67** (1991), 3503–3506.

[51] Jayesh and Z. Warhaft, *Probability distribution, conditional dissipation, and transport of passive temperature fluctuations in grid-generated turbulence*, Phys. Fluids A, **4** (1992), 2292.

[52] A. H. Jazwinski, "Stochastic Processes and Filtering Theory," Academic Press, New York, 1970.

[53] M. I. Jordan, *Graphical models*, Statistical Science, **19** (2004), 140–155.

[54] M. A. Katsoulakis, A. J. Majda and A. Sopasakis, *Intermittency, metastability and coarse graining for coupled deterministic-stochastic lattice systems*, Nonlinearity, **19** (2006), 1021–1047.

[55] S. R. Keating, A. J. Majda and K. S. Smith, *New methods for estimating poleward eddy heat transport using satellite altimetry*, Mon. Wea. Rev., in press, 2011.

[56] R. Kleeman, *Measuring dynamical prediction utility using relative entropy*, J. Atmos. Sci., **59** (2002), 2057–2072.

[57] P. E. Kloeden and E. Platen, "Numerical Solution of Stochastic Differential Equations," Applications of Mathematics (New York), Springer-Verlag, Berlin, 1992.

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

[59] R. H. Kraichnan, *Eddy viscosity and diffusivity: Exact formulas and approximations*, Complex Systems, **1** (1987), 805–820.

[60] R. H. Kraichnan, *Anomalous scaling of a randomly advected passive scalar*, Phys. Rev. Lett., **72** (1994), 1016–1019.

[61] R. H. Kraichnan, V. Yakhot and S. Chen, *Scaling relations for a randomly advected passive scalar field*, Phys. Rev. Lett., **75** (1995), 240–243.

[62] S. Kullback and R. Leibler, *On information and sufficiency*, Ann. Math. Stat., **22** (1951), 79–86.

[63] O. P. Le Maître, O. Knio, H. Najm and R. Ghanem, *A stochastic projection method for fluid flow. I. Basic formulation*, J. Comput. Phys., **173** (2001), 481–511.

[64] O. P. Le Maître, L. Mathelin, O. M. Knio and M. Y. Hussaini, *Asynchronous time integration for polynomial chaos expansion of uncertain periodic dynamics*, DCDS, **28** (2010), 199–226.

[65] C. E. Leith, *Climate response and fluctuation dissipation*, J. Atmospheric Sci., **32** (1975), 2022–2025.

[66] E. N. Lorenz, *Deterministic nonperiodic flow*, J. Atmos. Sci., **20** (1963), 130–141.

[67] E. N. Lorenz, *A study of predictability of a 28-variable atmospheric model*, Tellus, **17** (1968), 321–333.

[68] E. N. Lorenz, *The predictability of a flow which possesses many scales of motion*, Tellus, **21** (1969), 289–307.

[69] W. Luo, "Wiener Chaos Expansion and Numerical Solutions of Stochastic Partial Differential Equations," Ph.D thesis, Caltech, 2006.

[70] W. Luo, "Wiener Chaos Expansion and Numerical Solutions of Stochastic PDE," VDM Verlag, 2010.

[71] A. Majda and P. Kramer, *Simplified models for turbulent diffusion: Theory, numerical modeling, and physical phenomena*, Phys. Reports, **314** (1999), 237–574.

[72] A. J. Majda, *Real world turbulence and modern applied mathematics,* in "Mathematics: Frontiers and Perspectives," American Math. Society, Providence, RI, (2000), 137–151.

[73] A. J. Majda, *Challenges in Climate Science and Contemporary Applied Mathematics*, Comm. Pure Appl. Math, in press, 2011.

[74] 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), 581–586.

[75] A. J. Majda, R. V. Abramov and M. J. Grote, "Information Theory and Stochastics for Multiscale Nonlinear Systems," CRM Monograph Series, **25**, Americal Mathematical Society, Providence, RI, 2005.

[76] A. J. Majda, C. Franzke and D. Crommelin, *Normal forms for reduced stochastic climate models*, Proc. Natl. Acad. Sci. USA, **106** (2009), 3649–3653.

[77] 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. USA, **103** (2006), 8309–8314.

[78] A. J. Majda, C. Franzke and B. Khouider, *An applied mathematics perspective on stochastic modelling for climate*, Phil. Trans. R Soc. Lond. Ser. A Math. Phys. Eng. Sci., **366** (2008), 2429–2455.

[79] A. J. Majda and B. Gershgorin, *Quantifying uncertainty in climage change science through empirical information theory*, Proc. Natl. Acad. Sci., **107** (2010), 14958–14963.

[80] A. J. Majda and B. Gershgorin, *Improving model fidelity and sensitivity for complex systems through empirical information theory*, Proc. Natl. Acad. Sci. USA, **108** (2011), 10044–10049.

[81] A. J. Majda and B. Gershgorin, *Link between statistical equilibrium fidelity and forecasting skill for complex systems with model error*, Proc. Natl. Acad. Sci. USA, **108** (2011), 12599–12604.

[82] A. J. Majda, B. Gershgorin and Y. Yuan, *Low-frequency climate response and fluctuation–Dissipation theorems: Theory and practice*, J. Atmos. Sci., **67** (2010), 1186–1201.

[83] A. J. Majda and J. Harlim, *Physics constrained nonlinear regression models for time series*, Nonlinearity, submitted, 2012.

[84] A. J. Majda, R. Kleeman and D. Cai, *A mathematical framework for predictability through relative entropy*, Methods Appl. Anal., **9** (2002), 425–444.

[85] A. J. Majda, I. I. Timofeyev and E. Vanden Eijnden, *A mathematical framework for stochastic climate models*, Comm. Pure Appl. Math., **54** (2001), 891–974.

[86] A. J. Majda, I. I. Timofeyev and E. Vanden Eijnden, *Systematic strategies for stochastic mode reduction in climate*, J. Atmos. Sci., **60** (2003), 1705–1722.

[87] A. J. Majda and X. Wang, "Nonlinear Dynamics and Statistical Theories for Basic Geophysical Flows," Cambridge University Press, Cambridge, 2006.

[88] A. J. Majda and X. Wang, Linear response theory for statistical ensembles in complex systems with time-periodic forcing, Comm. Math. Sci., **8** (2010), 145–172.

[89] A. J. Majda and Y. Yuan, *Fundamental limitations of Ad hoc linear and quadratic multi-level regression models for physical systems*, Discrete Cont. Dyn. Systems, **17** (2010), 1333–1363.

[90] A. J. Majda, *Explicit inertial range renormalization theory in a model for turbulent diffusion*, J. Statist. Phys., **73** (1993), 515–542.

[91] A. J. Majda, *Random shearing direction models for isotropic turbulent diffusion*, J. Statist. Phys., **25** (1994), 1153–1165.

[92] A. J. Majda and B. Gershgorin, *Elementary models for turbulent diffusion with complex physical features: Eddy diffusivity, spectrum, and intermittency*, Phil. Trans. Roy. Soc., in press, 2011.

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

[94] A.J. Majda, J. Harlim and B. Gershgorin, *Mathematical strategies for filtering turbulent dynamical systems*, DCDS, **27** (2010), 441–486.

[95] R. Mikulevicius and B. Rozovskii, *Stochastic Navier-Stokes equations for turbulence flow*, SIAM J. Math. Anal., **35** (2004), 1250–1310.

[96] N. Wiener, *The Homogeneous Chaos*, Am. J. Math., **60** (1938), 897–936.

[97] N. Wiener, *The use of statistical theory in the study of turbulence*, in "Proc. Fifth Int. Cong. Appl. Mech.," Wiley, New York, (1939), 356.

[98] H. N. Najm, *Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics*, in "Annu. Rev. Fluid Mech.," **41**, Annual Reviews, Palo Alto, CA, (2009), 35–52.

[99] J. Neelin, B. Lintner, B. Tian, Q. Li, L. Zhang, P. Patra, M. Chahine and S. Stechmann, *Long tails in deep columns of natural and anthropogenic tropospheric tracers*, Geophysical Res. Lett., **37** (2010), L05804.

[100] J. D. Neelin, M. Munnich, H. Su, J. E. Meyerson and C. E. Holloway, *Tropical drying trends in global warming models and observations*, Proc Natl Acad Sci USA, **103** (2006), 6110–6115.

[101] J. Von Neumann, *Some remarks on the problem of forecasting climatic fluctuations*, in "Dynamics of Climate," (ed. R. L. Preffer), Pergamon Press, New York, (1960), 9–11.

[102] B. K. Oksendal, "Stochastic Differential Equations: An Introduction with Applications," Springer, 2010.

[103] S. A. Orszag and L. R. Bissonnette, *Dynamical properties of truncated Wiener-Hermite expansions*, Phys. Fluids, **10** (1967), 2603–2613.

[104] T. Palmer, *A nonlinear dynamical perspective on model error: A proposal for nonlocal stochastic dynamic parameterizations in weather and climate prediction models*, Quart. J. Roy. Meteor. Soc., **127** (2001), 279–303.

[105] N. Peters, "Turbulent Combustion," Cambridge Monographs on Mechanics, Cambridge University Press, 2000.

[106] C. L. Pettit and P. S. Beran, *Spectral and multiresolution Wiener expansions of oscillatory stochastic processes*, J. Sound Vib., **294** (2006), 752–779.

[107] S. B. Pope, *The probability approach to the modelling of turbulent reacting flows*, Combust. Flame, **27** (1976), 299–312.

[108] C. Prévôt and M. Röckner, "A Concise Course on Stochastic Partial Differential Equations," Lecture Notes in Mathematics, **1905**, Springer, Berlin, 2007.

[109] T. Sapsis and P. Lermusiaux, *Dynamically orthogonal field equations for continuous stochastic dynamical systems*, Physica D, **238** (2009), 2347–2360.

[110] T. Sapsis and P. Lermusiaux, *Dynamical criteria for the evolution of the stochastic dimensionality in flows with uncertainty*, Physica D, **241** (2012), 60–76.

[111] T. Sapsis and A. J. Majda, *Blended dynamic reduced subspace algorithms for uncertainty quantification in turbulent dynamical systems*, Nonlinearity, in preparation, 2012.

[112] T. Sapsis and A. J. Majda, *Dynamic reduced subspace and quasilinear gaussian methods for uncertainty quantification in turbulent dynamical systems*, Comm. Math. Sci., in preparation, 2012.

[113] T. Sapsis and A. J. Majda, *Efficient blended dynamic reduced subspace algorithms for uncertainty quantification in high-dimensional turbulent dynamical systems*, J. Comp. Phys., in preparation, 2012.

[114] R.-A. Todor and C. Schwab, *Sparse finite elements for elliptic problems with stochastic loading*, Numer. Math., **95** (2003), 707–734.

[115] L. M. Smith and S. L. Woodruff, *Renormalization-group analysis of turbulence*, in "Annu. Rev. Fluid Mech.," Vol. 30, Annual Reviews, Palo Alto, CA, (1998), 275–310.

[116] K. R. Sreenivasan, *The passive scalar spectrum and the Obukhov-Corrsin constant*, Phys. Fluids, **8** (1996), 189–196.

[117] H. Teng and G. Branstator, *Initial-value predictability of prominent modes of North Pacific subsurface temperature in a CGCM*, Climate Dyn., **36** (2010), 1813–1834.

[118] D. M. Titterington, *Bayesian methods for neural networks and related models*, Statistical Science, **19** (2004), 128–139.

[119] X. Wang, *Stationary statistical properties of Rayleigh-Bénard convection at large Prandtl number*, Comm. Pur. Appl. Math., **61** (2008), 789–815.

[120] X. Wang, *Approximation of stationary statistical properties of dissipative dynamical systems: Time discretization*, Math. Comp., **79** (2010), 259–280.

[121] C. K. Wikle and M. B. Hooten, A general science-based framework for dynamical spatiotemporal models, TEST, **19** (2010), 417–451.

[122] D. Xiu and G. E. Karniadakis, *Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos*, Comput. Methods Appl. Mech. Engrg., **191** (2002), 4927–4948.

[123] D. Xiu and G. E. Karniadakis, *Modeling uncertainty in flow simulations via generalized polynomial chaos*, J. Comput. Phys., **187** (2003), 137–167.

[124] D. Xiu, D. Lucor, C. H. Su and G. E. Karniadakis, *Stochastic modeling of flow-structure interactions using generalized polynomial chaos*, J. Fluid Engrg., **124** (2002), 51–59.

[125] G. E. Karnidakis and D. Xiu, *The Wiener-Askey polynomial chaos for stochastic differential equations*, SIAM J. Sci. Comput., **24** (2002), 619–644.

[126] Y. Yuan and A. J. Majda, *Invariant measures and asymptotic bounds for normal forms of stochastic climate models*, Chin. Ann. Math. B, **32** (2010), 343–368.

E-mail address: jonjon@cims.nyu.edu
E-mail address: mibran@gmail.com