

STRATEGIES FOR REDUCED-ORDER MODELS FOR PREDICTING THE STATISTICAL RESPONSES AND UNCERTAINTY QUANTIFICATION IN COMPLEX TURBULENT DYNAMICAL SYSTEMS*

ANDREW J. MAJDA[†] AND DI QI[†]

Abstract. Turbulent dynamical systems characterized by both a high-dimensional phase space and a large number of instabilities are ubiquitous among many complex systems in science and engineering including climate, material, and neural science. The existence of a strange attractor in the turbulent systems containing a large number of positive Lyapunov exponents results in a rapid growth of small uncertainties from imperfect modeling equations or perturbations in initial values, requiring naturally a probabilistic characterization for the evolution of the turbulent system. Uncertainty quantification (UQ) in turbulent dynamical systems is a grand challenge where the goal is to obtain statistical estimates such as the change in mean and variance for key physical quantities in their nonlinear responses to changes in external forcing parameters or uncertain initial data. In the development of a proper UQ scheme for systems of high or infinite dimensionality with instabilities, significant model errors compared with the true natural signal are always unavoidable due to both the imperfect understanding of the underlying physical processes and the limited computational resources available through direct Monte-Carlo integration. One central issue in contemporary research is the development of a systematic methodology that can recover the crucial features of the natural system in statistical equilibrium (*model fidelity*) and improve the imperfect model prediction skill in response to various external perturbations (*model sensitivity*).

A general mathematical framework to construct statistically accurate reduced-order models that have skill in capturing the statistical variability in the principal directions with largest energy of a general class of damped and forced complex turbulent dynamical systems is discussed here. There are generally three stages in the modeling strategy, imperfect model selection; calibration of the imperfect model in a training phase; and prediction of the responses with UQ to a wide class of forcing and perturbation scenarios. The methods are developed under a universal class of turbulent dynamical systems with quadratic nonlinearity that is representative in many applications in applied mathematics and engineering. Several mathematical ideas will be introduced to improve the prediction skill of the imperfect reduced-order models. Most importantly, *empirical information theory* and *statistical linear response theory* are applied in the training phase for calibrating model errors to achieve optimal imperfect model parameters; and *total statistical energy dynamics* are introduced to improve the model sensitivity in the prediction phase especially when strong external perturbations are exerted. The validity of general framework of reduced-order models is demonstrated on instructive stochastic triad models. Recent applications to two-layer baroclinic turbulence in the atmosphere and ocean with combinations of turbulent jets and vortices are also surveyed. The uncertainty quantification and statistical response for these complex models are accurately captured by the reduced-order models with only 2×10^2 modes in a highly turbulent system with 1×10^5 degrees of freedom. Less than 0.15% of the total spectral modes are needed in the reduced-order models.

Key words. reduced-order methods, statistical response, uncertainty quantification, anisotropic turbulence

AMS subject classifications. 76F55, 60H30, 86A32

Turbulent dynamical systems characterized by both a high-dimensional phase space and a large number of instabilities are ubiquitous among many complex systems in science and engineering [51, 54, 92, 75]. The existence of a strange attractor [93] in turbulent systems containing a large number of positive Lyapunov exponents results in a rapid growth of small uncertainties from imperfect modeling equations or perturbations in initial values, requiring naturally a probabilistic characterization for the evolution of the turbulent system. Uncertainty quantification (UQ) in turbulent dynamical systems is a grand challenge where the goal is to obtain statistical estimates such as the change in mean and variance for key physical quantities in their nonlinear responses to changes in external forcing parameters or uncertain initial data. One problem of practical significance in contemporary science is using UQ to understand the complexity of anisotropic turbulent processes over a wide range of spatio-temporal scales in engineering shear turbulence [35, 91, 26] as well as climate atmosphere ocean science [84, 92, 51]. This is especially important from a practical viewpoint because energy often flows intermittently from the smaller scales to affect the largest scales in such anisotropic turbulent flows.

In the development of a proper UQ scheme for systems of high or infinite dimensionality with instabilities, the analysis and prediction of phenomena often occur through complex dynamical equations that have significant model errors compared with the true natural signal. The imperfect model errors are always unavoidable due to both the imperfect understanding of the underlying physical processes and the limited computational resources needed for repeated Monte-Carlo simulations in each different scenario of high dimensional systems. Clearly, it is important both to improve the imperfect model's capabilities to recover crucial features of the natural system and also to accurately model the sensitivities in the natural system to changes in external or internal parameters. These efforts are hampered by the fact that the actual dynamics of the natural system are unknown. Important examples with major societal impact involve the Earth's climate and climate change where climate sensitivities are studied through a suite of imperfect comprehensive computer models ([22, 80, 57], and references therein); other examples include imperfect mesoscopic models in materials science [13, 39] and neural science [81] when compared

*Submitted to the editors 12/9/2016.

Funding: This work was funded by the Office of Naval Research through MURI N00014-16-1-2161 and DARPA through W911NF-15-1-0636

[†]Department of Mathematics and Center for Atmosphere and Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, NY 10012 (qidi@cims.nyu.edu, jonjon@cims.nyu.edu).

with actual observed behavior in these complex nonlinear systems.

Recently, information theory has been utilized in different ways to systematically improve model fidelity and sensitivity [58, 57], to quantify the role of coarse-grained initial states in long-range forecasting [28, 29], and to make an empirical link between model fidelity and forecasting skill [19, 20]. Imperfect models for complex systems are constrained by their capability to reproduce certain statistics in a training phase where the natural system has been observed; for example, this training phase in climate science is roughly the 60-year dataset of extensive observations of the Earth's climate system. For long-range forecasting, it is natural to guarantee statistical equilibrium fidelity for an imperfect model, and a framework using information theory is a natural way to achieve this in an unbiased fashion [58, 57, 28, 29, 20]. First, equilibrium statistical fidelity for an imperfect model depends on the choice of coarse-grained variables utilized [58, 57]; second, equilibrium model fidelity is a necessary but not sufficient condition to guarantee long-range forecasting skill [59, 29]. For example, Section 2.6 of [48] extensively discusses three very different strongly mixing chaotic dynamical models with 40 variables and with the same Gaussian equilibrium measure, the TBH, K-Z, and IL-96 models.

One significant application of UQ through empirical information theory is quantifying uncertainty in climate change science [57, 54]. The climate is an extremely complex coupled system involving multiple physical processes for the atmosphere, ocean, and land over a wide range of spatial scales from millimeters to thousands of kilometers and time scales from minutes to decades or centuries [22, 74]. Climate change science focuses on predicting the coarse-grained planetary scale long time changes in the climate system due to either changes in external forcing or internal variability such as the impact of increased carbon dioxide [31, 29]. For several decades the predictions of climate change science have been carried out with some skill through comprehensive computational atmospheric and oceanic simulation (AOS) models [22, 74, 80], which are designed to mimic the complex physical spatio-temporal patterns in nature. Such AOS models either through lack of resolution due to current computing power or through inadequate observation of nature necessarily parameterize the impact of many features of the climate system such as clouds, mesoscale and submesoscale ocean eddies, sea ice cover, etc. Thus, there are intrinsic model errors in the AOS models for the climate system and the effect of such model errors on predicting the coarse-grained large scale long time quantities is of interest. One central scientific issue in contemporary climate change science is the development of a systematic methodology that can recover the crucial features of the natural system in statistical equilibrium/climate (*model fidelity*) and improve the imperfect model prediction skill in response to various external perturbations like climate change and mitigation scenarios (*model sensitivity*) [2, 5, 20, 59, 54].

Here we discuss a general mathematical framework to construct statistically accurate reduced-order models that have the skill in capturing the statistical variability in the principal directions with largest energy of a general class of damped and forced complex turbulent dynamical systems. Low-order truncation methods is especially important for UQ with practical impact since *the curse of ensemble size* [6, 54] forbids to run Monte-Carlo simulations for all possible uncertain forcing scenarios in order to do attribution studies. Thus reduced-order models (ROM) are needed on a low-dimensional subspace where key physical significant quantities are characterized by the degrees of freedom that carry the largest energy or variance. In general, there are three stages in the modeling procedure, *imperfect model selection*; *calibration of the imperfect model* in a training phase using only data in the low-order perfect statistics; and *prediction of the responses* to a wide class of forcing and perturbation scenarios. The methods are developed for a universal class of turbulent dynamical systems with quadratic nonlinearity that is representative in many applications in applied mathematics and engineering [56, 51, 54]. Several mathematical ideas will be introduced to help improve the prediction skill of the imperfect reduced-order models. Most importantly, *empirical information theory* [48] and *statistical linear response theory* [59] are applied in the training phase for calibrating model errors to achieve optimal imperfect model parameters; and *total statistical energy dynamics* [53] are introduced to improve the model sensitivity in the prediction phase especially when strong external perturbations are exerted. The validity of general framework of reduced-order models has been verified by testing the methods on a series of representative turbulent dynamical models ranging from the 40-dimensional Lorenz '96 model [63], one-layer barotropic turbulence with topography [78], and finally the two-layer baroclinic turbulence with internal instability [79].

In the following parts of the paper, we first display the general formulation of the turbulent system with quadratic nonlinearity, and its statistical moment dynamical equations in Section 1. The skill and limitation of many previous low-order modeling ideas are also discussed. Theoretical toolkits that are useful for the development of reduced-order models are introduced in Section 2, where a general strategy to improve imperfect model sensitivity is described using empirical information theory and a general total statistical energy dynamics. Section 3 discusses the construction of reduced-order models in detail under this general framework with these various theoretical tools. In Section 4, we illustrate all these procedures and algorithms for the reduced-order models for some simple but instructive systems of triad stochastic equations with several novel features. In Section 5, we give examples of the skill of the procedures and algorithms on two-layer baroclinic models for both atmosphere

101 and ocean regimes with turbulent jets and vortices with roughly 1×10^5 degrees of freedom and direct and inverse turbulent
 102 cascades. In these very tough regimes, the reduced-order strategies show skill in capturing the response to changes in external
 103 forcing using only 200 modes, less than 0.15% of the modes in the original system.

104 **1. General Formulation of Turbulent Dynamical Systems with Nonlinearity.** One representative feature in many tur-
 105 bulent dynamical systems from nature is the quadratic energy conserving nonlinear interaction that transfers energy from the
 106 unstable modes to stable ones where the energy is dissipated resulting in a statistical steady state in equilibrium. We consider
 107 the following abstract formulation of the turbulent dynamical systems about state variables $\mathbf{u} \in \mathbb{R}^N$ in a high-dimensional phase
 108 space

$$109 \quad (1.1) \quad \frac{d\mathbf{u}}{dt} = (\mathcal{L} + \mathcal{D})\mathbf{u} + B(\mathbf{u}, \mathbf{u}) + \mathbf{F}(t) + \sigma(t) \dot{\mathbf{W}}(t; \omega).$$

110 On the right hand side of the above equation (1.1), the first two components, $(\mathcal{L} + \mathcal{D})\mathbf{u}$, represent linear dispersion and
 111 dissipation effects so that

$$112 \quad (1.2a) \quad \mathcal{L}^* = -\mathcal{L}, \text{ skew-symmetric}; \quad \mathcal{D}^* = \mathcal{D} < 0, \text{ negative-definite},$$

113 where the superscript star ‘*’ represents conjugate transpose of the matrix. The nonlinear effect in the dynamical system is
 114 introduced through a quadratic form, $B(\mathbf{u}, \mathbf{u})$, about the state variables \mathbf{u} that conserves energy when linear operators and all
 115 forcing in (1.1) are ignored, such that,

$$116 \quad (1.2b) \quad \mathbf{u} \cdot B(\mathbf{u}, \mathbf{u}) = \sum_{j=1}^N u_j B_j(\mathbf{u}, \mathbf{u}) \equiv 0, \quad \text{Energy Conservation},$$

117 where the dot on the left hand side denotes the inner product under a proper metric according to the conserved quantity [53, 54].
 118 Besides, the system is forced by external forcing effects that are decomposed into a deterministic component, $\mathbf{F}(t)$, and a
 119 stochastic component usually represented by a Gaussian random process, $\sigma(t) \dot{\mathbf{W}}(t; \omega)$. It needs to be noticed that $\mathbf{F}(t)$ might
 120 be inhomogeneous and introduce anisotropic structure into the system, and $\sigma(t) \dot{\mathbf{W}}(t; \omega)$ might further alter the energy structure
 121 in the fluctuation modes.

122 Many complex turbulent dynamical systems can be categorized into this abstract mathematical structure in (1.1) satisfying
 123 the properties (1.2a) and (1.2b), including the (truncated) Navier-Stokes equation [76] as well as basic geophysical models for
 124 the atmosphere, ocean, and the climate systems with rotation, stratification, and topography [84, 51, 54]. The main goal of the
 125 remainder of this paper is to provide a survey about the development of a consistent mathematical framework for systems like
 126 (1.1) and illustrate emerging applications of turbulent dynamical systems with model error and the curse of ensemble size.

127 **1.1. Exact statistical moment equations for the abstract formulation.** We use a finite-dimensional representation of
 128 the stochastic field \mathbf{u} consisting of a fixed-in-time, N -dimensional, orthonormal basis $\{\mathbf{e}_i\}_{i=1}^N$

$$129 \quad (1.3) \quad \mathbf{u}(t) = \bar{\mathbf{u}}(t) + \sum_{i=1}^N Z_i(t; \omega) \mathbf{e}_i,$$

130 where $\bar{\mathbf{u}}(t) = \langle \mathbf{u}(t) \rangle$ represents the ensemble average of the model state variable response (we use angled bracket to represent
 131 ensemble average), i.e. the mean field, and $Z_i(t; \omega)$ are stochastic coefficients measuring the fluctuation processes along the
 132 direction \mathbf{e}_i .

133 By taking the statistical (ensemble) average over the original equation (1.1) and using the mean-fluctuation decomposition
 134 (1.3), the **evolution equation of the mean state** $\bar{\mathbf{u}} = \langle \mathbf{u} \rangle$ is given by the following dynamical equation

$$135 \quad (1.4) \quad \frac{d\bar{\mathbf{u}}}{dt} = (L + D)\bar{\mathbf{u}} + B(\bar{\mathbf{u}}, \bar{\mathbf{u}}) + \sum_{i,j} R_{ij} B(\mathbf{e}_i, \mathbf{e}_j) + \mathbf{F},$$

136 with $R = \langle \mathbf{Z}\mathbf{Z}^* \rangle$ the second-order covariance matrix of the stochastic coefficients $\mathbf{Z} = \{Z_i\}_{i=1}^N$. The term $B(\bar{\mathbf{u}}, \bar{\mathbf{u}})$ represents the
 137 nonlinear interactions between the mean state, and $R_{ij} B(\mathbf{e}_i, \mathbf{e}_j)$ is the higher-order feedbacks from the fluctuation modes to the
 138 mean state dynamics. Moreover the random fluctuation component of the solution, $\mathbf{u}' = \sum_i Z_i(t; \omega) \mathbf{e}_i$ satisfies

$$139 \quad \frac{d\mathbf{u}'}{dt} = (L + D)\mathbf{u}' + B(\bar{\mathbf{u}}, \mathbf{u}') + B(\mathbf{u}', \bar{\mathbf{u}}) + B(\mathbf{u}', \mathbf{u}') - \langle B(\mathbf{u}', \mathbf{u}') \rangle + \sigma(t) \dot{\mathbf{W}}(t; \omega).$$

140 By projecting the above equation to each orthonormal basis element \mathbf{e}_i we obtain

$$141 \quad \frac{dZ_i}{dt} = Z_j [(L+D)\mathbf{e}_j + B(\bar{\mathbf{u}}, \mathbf{e}_j) + B(\mathbf{e}_j, \bar{\mathbf{u}})] \cdot \mathbf{e}_i + [B(\mathbf{u}', \mathbf{u}') - \langle B(\mathbf{u}', \mathbf{u}') \rangle] \cdot \mathbf{e}_i + \sigma(t) \dot{\mathbf{W}}(t; \omega) \cdot \mathbf{e}_i.$$

142 From the last equation we directly obtain the exact **evolution equation of the covariance matrix** $R = \langle \mathbf{Z}\mathbf{Z}^* \rangle$ by multiplying
143 \mathbf{Z}_j^* on both sides of the equation and taking ensemble statistical average

$$144 \quad (1.5) \quad \frac{dR}{dt} = L_v(\bar{\mathbf{u}})R + RL_v^*(\bar{\mathbf{u}}) + Q_F + Q_\sigma,$$

145 where we have:

146 **i)** the linear dynamical operator $L_v(\bar{\mathbf{u}})$ expresses energy transfers between the mean field and the stochastic modes (effect due
147 to B), as well as energy dissipation (effect due to \mathcal{D}) and non-normal dynamics (effect due to \mathcal{L})

$$148 \quad (1.6a) \quad \{L_v\}_{ij} = [(\mathcal{L} + \mathcal{D})\mathbf{e}_j + B(\bar{\mathbf{u}}, \mathbf{e}_j) + B(\mathbf{e}_j, \bar{\mathbf{u}})] \cdot \mathbf{e}_i;$$

149 **ii)** the positive definite operator Q_σ expresses energy transfer due to the external stochastic forcing

$$150 \quad (1.6b) \quad \{Q_\sigma\}_{ij} = \sum_k (\mathbf{e}_i \cdot \boldsymbol{\sigma}_k) (\boldsymbol{\sigma}_k \cdot \mathbf{e}_j);$$

151 **iii)** as well as the energy flux Q_F expresses nonlinear energy transfer between different modes due to non-Gaussian statistics
152 (or nonlinear terms) modeled through third-order moments

$$153 \quad (1.6c) \quad \{Q_F\}_{ij} = \sum_{m,n} \langle Z_m Z_n Z_j \rangle B(\mathbf{e}_m, \mathbf{e}_n) \cdot \mathbf{e}_i + \langle Z_m Z_n Z_i \rangle B(\mathbf{e}_m, \mathbf{e}_n) \cdot \mathbf{e}_j.$$

154 One important property to notice is that the energy conservation property of the quadratic operator B is inherited in the
155 statistical equations by the matrix Q_F since

$$156 \quad (1.6d) \quad \text{tr}(Q_F) = 2 \sum_i \sum_{m,n} \langle Z_m Z_n Z_i \rangle B(\mathbf{e}_m, \mathbf{e}_n) \cdot \mathbf{e}_i = 2B(\mathbf{u}', \mathbf{u}') \cdot \mathbf{u}' \equiv 0.$$

157 The above exact statistical equations for the state of the mean (1.4) and covariance matrix (1.5) will be the starting point for the
158 developments in the reduced-order models on UQ methods.

159 Note that the statistical dynamics for the mean (1.4) and covariance (1.5) are still not closed due to the inclusion of third-
160 order moments through the nonlinear interactions in Q_F in (1.6c). The basic idea in the general development of reduced-order
161 schemes concerns about proper approximation about this energy flux term Q_F in a simple and efficient manner so that the
162 energy mechanism can be modeled properly in the reduced-order schemes [44, 84, 89].

163 **1.1.1. Low-order truncation methods for UQ and their limitations.** Next we briefly discuss some popular low-order
164 truncation methods for closing the statistical equations (1.4) and (1.5) and their limitations. Low-order truncation models for UQ
165 include projection of the dynamics on leading order empirical orthogonal functions (EOFs) [36], truncated polynomial chaos
166 (PC) expansions [37, 41, 72], and dynamically orthogonal (DO) truncations [85, 86]. Then ideas about closing the low-order
167 truncated system within the resolved modes need to be proposed. A pioneering statistical prediction strategy [23, 24] overcomes
168 the curse of ensemble size for moderate size turbulent dynamical systems by simply neglecting the third-order moments by
169 setting $Q_F \equiv 0$ in the covariance equations (1.5). This *Gaussian closure method* has been applied to short time statistical
170 prediction for truncated geophysical models like the one-layer geophysical models in (1.9c) with some success [24, 90]. Based
171 on the similar idea of neglecting third-order moments, the *eddy-damped quasi-normal Markovian approximation* (EDM) [84,
172 44] is another approximation to the moment hierarchy (1.4) and (1.5) that closes the second moments with (inconsistent)
173 Gaussian approximation in the higher order equations. With a much larger *eddy-damped* parameters, the EDM equations are
174 realizable in a stochastic model.

175 Moreover concise mathematical models and analysis reveal fundamental limitations in truncated EOF expansions [3, 18],
176 PC expansions [9, 56], and DO truncations [87], due to different manifestations of the fact that in many turbulent dynamical
177 systems, modes that carry small variance on average can have important, highly intermittent dynamical effects on the large
178 variance modes. Furthermore, the large dimension of the active variables in turbulent dynamical systems makes direct UQ by

179 large ensemble Monte-Carlo simulations impossible in the foreseeable future while once again, concise mathematical models
 180 [56] point to the limitations of using moderately large yet statistically too small ensemble sizes. Other important methods for
 181 UQ involve the linear statistical response to change in external forcing or initial data through the fluctuation dissipation theorem
 182 (FDT) which only requires the measurement of suitable time correlations in the unperturbed system [1, 31, 32, 33, 61]. Despite
 183 some significant success with this approach for turbulent dynamical systems [1, 31, 32, 61], the method is hampered by the
 184 need to measure suitable approximations to the exact correlations for long time series as well as the fundamental limitation to
 185 parameter regimes with a linear statistical response. All the limitations above imply the need of a more careful treatment for
 186 the higher-order statistics in Q_F in the exact equations for mean and covariance (1.4) and (1.5).

187 **1.2. The overall prediction strategy for the development of reduced-order statistical models.** Before preceding to the
 188 details about developing the reduced-order statistical model framework, we illustrate the basic ideas in the modeling process
 189 as a general overview. Overall, this can serve as a generic procedure where rigorous mathematical theories and various com-
 190 putational strategies are combined to get a crucial improvement for understanding turbulent dynamical systems. In general, we
 191 can decompose the reduced-order statistical modeling strategy into three stages: i) imperfect model selection according to the
 192 complexity of the problem; ii) model calibration in the training phase using equilibrium data; and iii) model prediction with the
 193 optimized model parameters for various responses to external perturbations. The overall prediction strategy is summarized in a
 194 diagram in Figure 1.1. The basic procedure for developing statistical models illustrates a representative example where various
 195 mathematical theories and numerical methods interact and are combined for achieving a better understanding about the natural
 196 system.

197 **1.2.1. Ergodicity and non-trivial invariant measure for the true turbulent dynamical systems.** In the first place,
 198 the best reduced-order approximation strategy can only be achieved through a good understanding about the true turbulent
 199 dynamical system. Several important mathematical theories are especially useful for characterizing the statistical structure of
 200 the turbulent system. Under special damping and random noise forms without the deterministic forcing $\mathbf{F} \equiv 0$, a Gaussian
 201 invariant measure can be generated in the statistical steady state, whereas this Gaussian distribution from equilibrium statistical
 202 mechanics can be only derived from special damping and noise terms [54, 78]. One more generalized situation with importance
 203 in many realistic applications is when no stochastic forcing in the damped and forced dynamical system (1.1), so that the
 204 deterministic system with $\sigma \equiv 0$ has non-trivial long-time dynamics. The uncertainty in such deterministic systems is measured
 205 by the unstable sub-phase space with a number of positive Lyapunov exponents, thus a nontrivial global attractor is generated
 206 through the strong interaction and exchange of energy [82]. This scenario is similar to the Sinai-Ruelle-Bowen (SRB) measure
 207 problem [93, 83]. In that case, a unique distinguished invariant measure p_{eq} , the SRB measure, is the one selected by the
 208 vanishing noise limit with appropriate assumptions on the system and noise. This distinguished invariant measure forms up a
 209 stationary statistical solution p_{eq} in equilibrium, so that

$$210 \quad (1.7) \quad p_{\text{eq}} \left((\Phi^t)^{-1}(\Omega) \right) = p_{\text{eq}}(\Omega), \quad \text{for any } t > 0 \text{ and } \Omega \subset \mathbb{R}^N,$$

211 with $\Phi^t : \mathbb{R}^N \rightarrow \mathbb{R}^N$ as the flow map. This invariant measure (1.7) provides a mechanism for explaining how local instability on
 212 attractors can produce coherent statistics for orbits starting from large sets in the basin. The statistical ensemble behaviour in
 213 equilibrium such as the mean state and covariance can be deduced by taking averages with respect to the invariant measure.

214 Ergodicity is then one important property for the turbulent dynamical system with uncertainty, and means that there exists
 215 a unique invariant measure in statistical equilibrium which attracts all statistical initial data. Geometric ergodicity for finite
 216 dimensional Galerkin truncation models (for example, the two or three dimensional Navier-Stokes equations) with minimal
 217 stochastic forcing is an important research topic [21, 54, 65]. With proper ergodicity assumption about the abstract system (1.1)
 218 and rigorously justified for the system with minimal stochastic forcing [54, 65], the statistical expectation of any functionals
 219 about the state variables can be calculated through averaging the time-series in steady state, that is,

$$220 \quad (1.8) \quad \langle g(\mathbf{u}) \rangle = \int_{\mathbb{R}^N} g(\mathbf{u}) p_{\text{eq}}(\mathbf{u}) d\mathbf{u} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{t_0}^{t_0+T} g[\mathbf{u}(t)] dt,$$

221 where g is any functional about the state variables \mathbf{u} , and p_{eq} is the invariant measure (1.7) in statistical equilibrium. Taking
 222 the ensemble averages from the first equality of (1.8) is usually an extremely challenging problem, while the average along a
 223 trajectory over a long time as the right hand side of (1.8) forms a more practical approach. Ergodicity is crucial in this prediction
 224 strategy for achieving accurate perfect model statistics, and will be assumed throughout the following discussions.

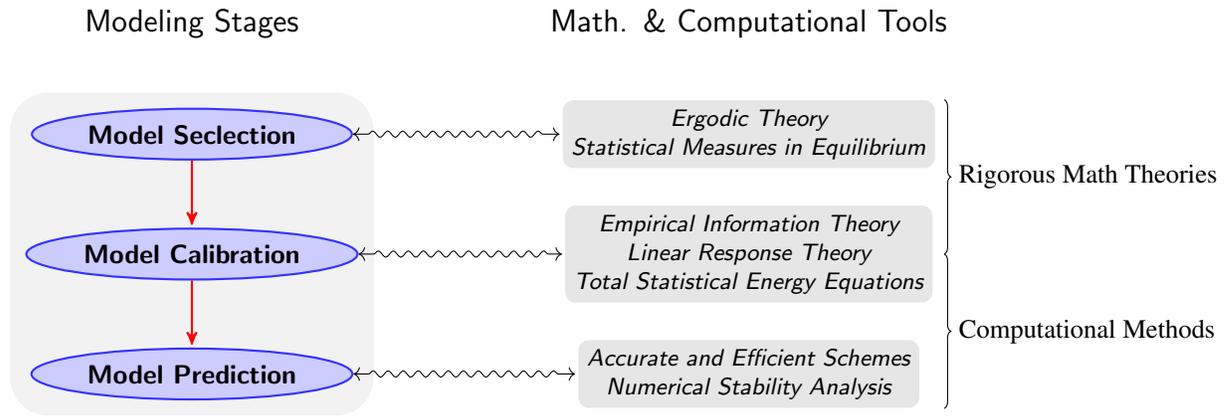


FIG. 1.1. The general strategy for the development of reduced-order statistical models. Three sequential stages are required to carry out the reduced-order statistical model, and rigorous mathematical theories are combined with numerical analysis to calibrate model errors and improve the imperfect model prediction skill.

225 **1.2.2. Model selection, model calibration, and prediction with optimized imperfect model.** The ergodic theory and
 226 invariant measure enable us to get access to the model equilibrium statistical structures in steady state. Still the major goal
 227 in this investigation is to find the model sensitivity in response to various external perturbations. Especially for the turbulent
 228 dynamical systems with instability like (1.1), nonlinearity forms the key mechanism in the complex chaotic behaviour, and even
 229 small perturbations may drive the system away from its original equilibrium state. Furthermore, strong non-Gaussianity due
 230 to the strange attractor from the SRB measure is another characteristic feature in these turbulent systems with non-Gaussian
 231 measures even in equilibrium. The reduced-order statistical modeling procedure aims at capturing these nonlinear non-Gaussian
 232 statistical responses in the principal directions in the system in an accurate and efficient way.

233 As illustrated in Figure 1.1 for the general strategy, the modeling procedure begins with the model selection stage where
 234 proper approximation method is adopted through a careful analysis about the statistical theories. Specifically in the reduced-
 235 order models to be developed here, usually additional damping and random noise corrections are introduced for the unresolved
 236 higher-order statistics. The equilibrium invariant measure and ergodic theory [65, 89] can help determine the optimal Galerkin
 237 truncation wavenumber for the reduced-order model and the proper basis that can cover the most important directions in the
 238 system. Especially, non-Gaussian statistics in the unperturbed equilibrium state would also become important and require
 239 careful consideration in the model calibration.

240 The model calibration procedure is usually carried out in a training phase before the prediction, so that the optimal imperfect
 241 model parameters can be achieved through a careful calibration about the true higher-order statistics. The ideal way is to find
 242 a unified systematic strategy where various external perturbations can be predicted from the same set of optimal parameters
 243 through this training phase. To achieve this, various statistical theories and numerical strategies need to be blended together in a
 244 judicious fashion. Most importantly, we need to consider the linear statistical response theory to calibrate the model responses
 245 in mean and variances [48, 52, 33]; and use empirical information theory [58, 59, 51] to get a balanced measure for the error in
 246 the leading order moments. In the final model prediction stage, the optimized imperfect model parameters are applied for the
 247 forecast of various model responses to perturbations. In the construction about numerical models, numerical issues also need
 248 be taken into account to make sure numerical stability and accuracy. Especially, proper schemes with accuracy order consistent
 249 with the reduced model approximation error should be proposed to ensure optimal performance.

250 **1.3. Low-order models illustrating model selection, calibration, and prediction in UQ.** Here we provide a brief dis-
 251 cussion of some instructive quantitative and qualitative low-order models where the above strategy for improved prediction
 252 and UQ is displayed. The test models as in nature often exhibit intermittency [26, 76] where some components of a turbulent
 253 dynamical system have low amplitude phases followed by irregular large amplitude bursts of extreme events. Intermittency is
 254 an important physical phenomena. Exactly solvable test models as a test bed for the prediction and UQ strategy [54, 48, 57]
 255 including information barriers are discussed extensively in models ranging from linear stochastic models to nonlinear mod-
 256 els with intermittency in the research expository article [56] as well as in [7, 8]. Some more sophisticated applications are
 257 mentioned next in Section 1.4.

258 Turbulent diffusion in exactly solvable models is a rich source of highly nontrivial spatiotemporal multi-scale models to test
 259 the strategies in *empirical information theory* and *kicked statistical response theory* in a more complex setting [27, 58, 59, 60].
 260 Even though these models have no positive Lyapunov exponents, they have been shown rigorously to exhibit intermittency and
 261 extreme events [66]. Calibration strategies for imperfect models using information theory have been developed recently to yield
 262 statistical accurate prediction of these extreme events by imperfect inexpensive linear stochastic models for the velocity field
 263 [77]. This topic merits much more attention by other modern applied mathematicians [70, 71].

264 **1.3.1. Nonlinear regression models for time series.** A central issue in contemporary science is the development of
 265 data driven statistical dynamical models for the time series of a partial set of observed variables which arise from suitable
 266 observations from nature (see [17] and references therein); examples are multi-level linear autoregressive models as well as *ad*
 267 *hoc* quadratic nonlinear regression models. It has been established recently [67] that *ad hoc* quadratic multi-level regression
 268 models can have finite time blow up of statistical solutions and pathological behavior of their invariant measure even though
 269 they match the data with high precision. A new class of physics-constrained multi-level nonlinear regression models was
 270 developed which involve both memory effects in time as well as physics-constrained energy conserving nonlinear interactions
 271 [34, 62], which completely avoid the above pathological behavior with full mathematical rigor.

272 A striking application of these ideas combined with information calibration to the predictability limits of tropical in-
 273 traseasonal variability such as the Madden-Julian oscillation (MJO) and the monsoon has been developed in a series of papers
 274 [16, 15, 14]. They yield an interesting class of low-order turbulent dynamical systems with extreme events and intermittency.
 275 The nonlinear low-order stochastic model (see Section 4.2 of [54]) has been shown to have significant skill for determining
 276 the predictability limits of the large-scale cloud patterns of the boreal winter MJO [16] and the summer monsoon [14]. It is an
 277 interesting open problem to rigorously describe the intermittency and other mathematical features in these low-order turbulent
 278 dynamical systems.

279 **1.4. Examples of complex turbulent dynamical systems.** Here we list some typical prototype models of complex turbu-
 280 lent dynamical systems with the structure in (1.1). These qualitative and quantitative models with increasing complexity form
 281 a desirable set of testing models for prediction, UQ, and state estimation [54]. We will finally test the reduced-order modeling
 282 strategies on all these typical models as a thorough discussion about the effectiveness and limitations of the model reduction
 283 ideas including a complete new treatment for the triad example.

284 (A) *The triad system with quadratic energy transfer.* The triad model [50, 54] is the elementary building block of complex tur-
 285 bulent systems with energy conserving nonlinear interactions. It is a 3-dimensional ODE system with inhomogeneous
 286 damping and both deterministic and stochastic forcing terms

$$\begin{aligned}
 \frac{du_1}{dt} &= L_2 u_3 - L_3 u_2 - d_1 u_1 + B_1 u_2 u_3 + F_1 + \sigma_1 \dot{W}_1, \\
 \frac{du_2}{dt} &= L_3 u_1 - L_1 u_3 - d_2 u_2 + B_2 u_3 u_1 + F_2 + \sigma_2 \dot{W}_2, \\
 \frac{du_3}{dt} &= L_1 u_2 - L_2 u_1 - d_3 u_3 + B_3 u_1 u_2 + F_3 + \sigma_3 \dot{W}_3.
 \end{aligned}$$

287 (1.9a)

288 The triad system is an instructive test model for the reduced-order strategies. A self-contained pedagogical discussion
 289 about the triad system is shown in Section 4.

290 (B) *40-dimensional Lorenz '96 model.* The Lorenz '96 model [46, 63, 54, 51] is a 40-dimensional turbulent dynamics defined
 291 with periodic boundary condition which mimics weather waves of the mid-latitude atmosphere. Various representative
 292 statistical features can be generated by changing the external forcing values in F

$$\frac{du_j}{dt} = (u_{j+1} - u_{j-2}) u_{j-1} - d_j u_j + F_j, \quad u_0 = u_J, \quad j = 0, \dots, J-1, \quad J = 40.$$

293 (1.9b)

294 See [63] for the detailed reduced-order modeling strategy.

295 (C) *One-layer barotropic model with topography.* The one-layer barotropic system [51, 54, 78] is a basic and simple geo-
 296 physical model for the atmosphere or ocean with the essential geophysical effects of rotation, topography, and both

297 deterministic and random forcing.

$$\begin{aligned}
 & \frac{\partial q}{\partial t} + \nabla^\perp \boldsymbol{\psi} \cdot \nabla q + U \frac{\partial q}{\partial x} = -\mathcal{D}(\Delta) \boldsymbol{\psi} + \mathcal{F}(\mathbf{x}, t) + \Sigma(\mathbf{x}) \dot{W}(t), \\
 (1.9c) \quad & \frac{dU}{dt} + \int \frac{\partial h}{\partial x} \boldsymbol{\psi}(t) = -\mathcal{D}_0 U + \mathcal{F}_0(t) + \Sigma_0 \dot{W}_0(t), \\
 & q = \Delta \boldsymbol{\psi} + h + \beta y.
 \end{aligned}$$

299 See [78] for the detailed reduced-order modeling strategy.

300 **(D)** *Two-layer quasi-geostrophic model with baroclinic instability.* The two-layer quasi-geostrophic model with baroclinic
 301 instability in a two-dimensional periodic domain [84, 92, 79] is one fully nonlinear fluid model, and is quite capable
 302 in capturing the essential physics of the relevant internal variability despite its relatively simple dynamical structure.

$$\begin{aligned}
 (1.9d) \quad & \frac{\partial q_\boldsymbol{\psi}}{\partial t} + J(\boldsymbol{\psi}, q_\boldsymbol{\psi}) + J(\boldsymbol{\tau}, q_\boldsymbol{\tau}) + \beta \frac{\partial \boldsymbol{\psi}}{\partial x} + U \frac{\partial}{\partial x} \Delta \boldsymbol{\tau} = -\frac{\kappa}{2} \Delta(\boldsymbol{\psi} - \boldsymbol{\tau}) - \nu \Delta^s q_\boldsymbol{\psi} + \mathcal{F}_\boldsymbol{\psi}(\mathbf{x}, t), \\
 & \frac{\partial q_\boldsymbol{\tau}}{\partial t} + J(\boldsymbol{\psi}, q_\boldsymbol{\tau}) + J(\boldsymbol{\tau}, q_\boldsymbol{\psi}) + \beta \frac{\partial \boldsymbol{\tau}}{\partial x} + U \frac{\partial}{\partial x} (\Delta \boldsymbol{\psi} + k_d^2 \boldsymbol{\psi}) = \frac{\kappa}{2} \Delta(\boldsymbol{\psi} - \boldsymbol{\tau}) - \nu \Delta^s q_\boldsymbol{\tau} + \mathcal{F}_\boldsymbol{\tau}(\mathbf{x}, t).
 \end{aligned}$$

304 See [79] and discussions in Section 5 for the reduced-order modeling strategy.

305

306 **2. Statistical Theory Toolkits for Improving Model Prediction Skill.** In this section we introduce the general theoretical
 307 toolkits that are useful for capturing the key statistical features in turbulent systems like (1.1) and improving imperfect model
 308 prediction skill. Despite the complex model statistical responses in each component as the turbulent dynamical system gets
 309 perturbed, there exists a simple and exact statistical energy conservation principle for the total statistical energy of the system
 310 describing the overall (inhomogeneous) statistical structure in the system through a simple scalar dynamical equation [53, 54].
 311 The theory is briefly described in Section 2.1. Then the construction about the imperfect reduced-order models concerns about
 312 the consistency in equilibrium (climate fidelity) and the responses to perturbations (model sensitivity). Equilibrium statistical
 313 fidelity should be guaranteed in the first place so that the reduced-order model will converge to the true unperturbed equilibrium
 314 statistics. To further calibrate the detailed model sensitivity to perturbations in each statistical component, the linear response
 315 theory can offer useful quantities to measure for quantifying the crucial statistics in the model structure. Combining with the
 316 relative entropy under empirical information theory, a general information-theoretical framework can be proposed to tune the
 317 imperfect model parameters in a training phase, thus optimal model parameters can be used for model prediction in various
 318 dynamical regimes. We will describe the basic statistical theories in this section.

319 **2.1. A statistical energy conservation principle.** Despite the fact that the exact equations for the statistical mean (1.4)
 320 and the covariance fluctuations (1.5) are not closed equations, there is suitable statistical symmetry so that the energy of the
 321 mean plus the trace of the covariance matrix satisfies an energy conservation principle even with general deterministic and
 322 random forcing. Here we briefly introduce the theory developed in [53, 54] about a total statistical energy dynamics for the
 323 abstract system (1.1). This total statistical energy offers a general description about the total responses in the perturbed system
 324 and will be shown useful for the construction of reduced-order models.

325 Consider the statistical mean energy, $\bar{E} = \frac{1}{2} |\bar{\mathbf{u}}|^2 = \frac{1}{2} \bar{\mathbf{u}} \cdot \bar{\mathbf{u}}$, and the statistical fluctuation energy, $E' = \frac{1}{2} \langle \mathbf{u}' \cdot \mathbf{u}' \rangle = \frac{1}{2} \text{tr} R$.
 326 Assume the following symmetries involving the nonlinear interaction operator B under the orthonormal basis $\{\mathbf{e}_i\}$:

327 **A)** The self interactions vanish in the quadratic interaction,

$$(2.1a) \quad B(\mathbf{e}_i, \mathbf{e}_i) \equiv 0, \quad 1 \leq i \leq N;$$

329 **B)** The dyad interaction coefficients vanish through the symmetry,

$$(2.1b) \quad \mathbf{e}_i \cdot [B(\mathbf{e}_j, \mathbf{e}_i) + B(\mathbf{e}_i, \mathbf{e}_j)] = 0, \quad \text{for any } i, j.$$

331 Therefore the *detailed triad symmetry* guarantees that the nonlinear interaction $B(\mathbf{u}, \mathbf{u})$ will not alter the total statistical energy
 332 structure in the system (though the state of the mean and covariance may both changed due to the nonlinear term in (1.4) and
 333 (1.5)). So we have the following theorem [53, 54]:

334 THEOREM 2.1. (**Statistical Energy Conservation Principle**) Under the structural assumptions (2.1a), (2.1b) on the basis
 335 \mathbf{e}_i , for any turbulent dynamical systems in (1.1), the total statistical energy, $E = \bar{E} + E' = \frac{1}{2}\bar{\mathbf{u}} \cdot \bar{\mathbf{u}} + \frac{1}{2}\text{tr}R$, satisfies

$$336 \quad (2.2) \quad \frac{dE}{dt} = \bar{\mathbf{u}} \cdot D\bar{\mathbf{u}} + \bar{\mathbf{u}} \cdot \mathbf{F} + \text{tr}(DR) + \frac{1}{2}\text{tr}Q_\sigma,$$

337 where R satisfies the exact covariance equation in (1.5). Matrix Q_σ expresses energy transfer due to external stochastic forcing,
 338 and is a diagonal matrix with entries, $Q_{\sigma,kk} = |\sigma_k|^2$.

339 For most practical dynamical systems, for example, the systems we have illustrated in (1.9a)-(1.9d), the symmetries in (2.1)
 340 are usually satisfied. Also a generalization allowing both dyads and triads in the statistical energy conservation principle is in
 341 [54]. Thus the statistical energy conservation principle can always be applied. Notice that especially under the homogeneous
 342 dissipation case, $D = -dI$, the right hand side of the statistical energy equation (2.2) will become a linear damping term for
 343 the total energy, $-dE$, plus the deterministic forcing applying on the mean state and the stochastic forcing contribution. This
 344 implies that the total energy structure (and thus the total variance in all the modes) can be determined from the statistical mean
 345 state by solving the scalar equation above.

346 **2.2. Statistical equilibrium fidelity in approximation models.** Here we consider the statistical energy of the dynamical
 347 system in each individual (spectral) mode. Statistical equilibrium fidelity concerns the convergence to the true equilibrium
 348 statistics in statistical steady state in the reduced-order models. Recall the true second-order statistical equation (1.5) about the
 349 covariance matrix

$$350 \quad \frac{dR}{dt} = L_\nu R + RL_\nu^* + Q_F + Q_\sigma.$$

351 The most difficult and expensive part in solving the above system comes from evaluating the nonlinear flux term Q_F where
 352 higher order statistics are involved, that is,

$$353 \quad \{Q_F\}_{ij} = \sum_{m,n} \langle Z_m Z_n Z_j \rangle B(\mathbf{e}_m, \mathbf{e}_n) \cdot \mathbf{e}_i + \langle Z_m Z_n Z_i \rangle B(\mathbf{e}_m, \mathbf{e}_n) \cdot \mathbf{e}_j.$$

354 Note that the third-order moments always include triad interactions of modes $\{Z_m, Z_n, Z_j\}$ between different scales, where
 355 nonlinear energy forward-cascade and backward-cascade along the energy spectrum can be induced. Thus the central issue in
 356 developing closure models becomes to find proper approximation about the nonlinear flux term $Q_F^M \sim Q_F$ which can offer a
 357 statistically consistent estimation. First of all, it is important to remember the conservation of the total nonlinear flux $\text{tr}Q_F \equiv 0$
 358 from (1.6d). This equality implies that the nonlinear interactions will not introduce additional energy source or sink into the
 359 system. Thus the same constraint should be maintained in designing the approximation models, $\text{tr}Q_F^M = 0$. Consideration about
 360 accuracy and computational efficiency should be balanced in determining the explicit form of Q_F^M in the implementation of
 361 reduced methods. Here we first display some theoretical principles about the equilibrium nonlinear flux $Q_{F,\text{eq}}$ that can be used
 362 as guidelines for determining the values in Q_F^M .

363 **2.2.1. Calibration about higher-order statistics in full phase space.** In the prediction of model responses it is most
 364 important to find the variability along each principal direction. In general, the nonlinear flux Q_F illustrates the nonlinear energy
 365 transfer between modes with different scales. In fact, we can decompose the matrix $Q_F = Q_F^+ + Q_F^-$ by singular value decom-
 366 position into a positive-definite and negative-definite component. The positive definite part Q_F^+ illustrates the additional energy
 367 that is injected into this mode from other scales, while the negative definite part Q_F^- shows the extraction of energy through
 368 nonlinear transfer to other scales. Thus the accurate approximation about the nonlinear flux $Q_{F,ij}$ in each (spectral) component
 369 becomes important. On the other hand, this approximation requires the calibration about the third-order moments $\langle Z_m Z_n Z_j \rangle$
 370 and $\langle Z_m Z_n Z_i \rangle$, and will always include the interactions between the (resolved) large-scale modes and (unresolved) smaller-scale
 371 fluctuations. Direct simulation would require ensemble averages for the third-order moments, where large numerical errors and
 372 high computational loads are almost unavoidable.

373 Instead, from the statistical dynamics for the covariance equation (1.5) in statistical steady state, the temporal derivative on
 374 the left hand side vanishes, $\frac{d}{dt}R_{\text{eq}} \equiv 0$, thus the equilibrium solution $(\bar{\mathbf{u}}_{\text{eq}}, R_{\text{eq}})$ necessarily satisfies the steady state equation

$$375 \quad 0 = L_\nu (\bar{\mathbf{u}}_{\text{eq}}) R_{\text{eq}} + R_{\text{eq}} L_\nu^* (\bar{\mathbf{u}}_{\text{eq}}) + Q_{F,\text{eq}} + Q_\sigma,$$

376 where $Q_{F,\text{eq}}$ includes the third-order moments evaluated at the statistical steady state. Therefore we can get the measurements
 377 about equilibrium third-order nonlinear flux through the lower order steady state solution of the mean, $\bar{\mathbf{u}}_{\text{eq}}$, and the covariance,

378 R_{eq} , so that

$$379 \quad (2.3) \quad Q_{F,\text{eq}} = -L_{\nu}(\bar{\mathbf{u}}_{\text{eq}})R_{\text{eq}} - R_{\text{eq}}L_{\nu}^*(\bar{\mathbf{u}}_{\text{eq}}) - Q_{\sigma}.$$

380 The quasi-linear operator $L_{\nu}(\bar{\mathbf{u}}_{\text{eq}})$ is defined through (1.6a) containing the interactions between mean state. Especially, the
 381 non-trivial third moments play a crucial dynamical role in the statistical closure models. As an example in the case without
 382 random forcing $Q_{\sigma} \equiv 0$, the necessary and sufficient condition for a non-Gaussian statistical steady state [54] requires that

$$383 \quad L_{\nu}(\bar{\mathbf{u}}_{\text{eq}})R_{\text{eq}} + R_{\text{eq}}L_{\nu}^*(\bar{\mathbf{u}}_{\text{eq}}) \neq 0,$$

384 so the above matrix has non-zero entries. This is an important constraint that needs to be considered first in the construction
 385 about reduced-order models in the next section.

386 **2.2.2. Calibration of higher-order statistics in the reduced subspace.** Despite the above exact model calibration for
 387 higher-order statistics (2.3) using equilibrium mean and covariance, in many realistic problems, resolving the entire covariance
 388 matrix $R \in \mathbb{C}^{N \times N}$ of order $O(N^2)$ is still expensive and unnecessary especially for high dimensional systems $N \gg 1$ with strong
 389 interactions between small and large scales. Often the key physical significant quantities are characterized by the degrees of
 390 freedom which carry the largest energy (or variance). Thus, for most cases we are mostly interested in the model variability in
 391 a low-dimensional subspace along the principal directions spanned by the subsatial basis

$$392 \quad P = [\mathbf{v}_1, \dots, \mathbf{v}_s], \quad s \ll N.$$

393 One simplest proposal to get the low-order basis $\{\mathbf{v}_j\}$ is through the leading order EOFs or energy based proper orthogonal
 394 decomposition [36, 3]. The reduced-order third-order nonlinear flux can be calculated through a more efficient way using only
 395 the mean state, $\bar{\mathbf{u}}_{\text{eq}}$, and covariance in the subspace of interest, $C = P^*RP \in \mathbb{C}^{s \times s}$. By projecting the original nonlinear flux
 396 formulation (2.3) onto the subspace, we have the reduced-order formulation

$$397 \quad (2.4) \quad Q_{F,\text{eq}}^{\text{red}} \equiv P^*Q_{F,\text{eq}}P = -L_{\nu}^{\text{red}}(\bar{\mathbf{u}}_{\text{eq}})C_{\text{eq}} - C_{\text{eq}}L_{\nu}^{\text{red}*}(\bar{\mathbf{u}}_{\text{eq}}) - Q_{\sigma}^{\text{red}},$$

398 where the reduced-order quasi-linear operator and reduced-order noise can also be calculated efficiently only using information
 399 in the subspace with resolved leading order statistics

$$400 \quad L_{\nu,i j}^{\text{red}} \equiv \{P^*L_{\nu}P\}_{ij} = [(\mathcal{L} + \mathcal{D})\mathbf{v}_j + B(\bar{\mathbf{u}}, \mathbf{v}_j) + B(\mathbf{v}_j, \bar{\mathbf{u}})] \cdot \mathbf{v}_i, \quad Q_{\sigma}^{\text{red}} = P^*Q_{\sigma}P.$$

401 Thus even though $Q_{F,\text{eq}}^{\text{red}}$ may still include many third moments between the low-wavenumber resolved modes and high-wavenumber
 402 modes that are not calculated explicitly in the reduced-order equations only for C , we can still achieve the equilibrium nonlin-
 403 ear flux constrained in the resolved subspace of interest by using only the mean and covariances $(\bar{\mathbf{u}}_{\text{eq}}, C_{\text{eq}})$ along the resolved
 404 directions $\{\mathbf{v}_1, \dots, \mathbf{v}_s\}$.

405 In general, the first two moments in equilibrium can be achieved through the ergodicity (1.8) by averaging the variables
 406 of interest along one solution trajectory, thus we can get the calibration about the third-order moment feedbacks in the second-
 407 order dynamics by solving the equation (2.3) or (2.4). Besides, we also find one necessary condition for confirming equilibrium
 408 fidelity for the reduced-order models for the construction of nonlinear flux term, so that consistent nonlinear flux Q_F is guaran-
 409 teed in the final steady state

$$410 \quad (2.5) \quad Q_F^M \rightarrow Q_{F,\text{eq}}, \quad \text{as } t \rightarrow \infty.$$

411 Actually, the idea of estimating the higher-order statistics through low-order moments has been exploited for several specific
 412 models in [89, 88, 63]. The equilibrium statistics from (2.4) can efficiently calibrate the model nonlinear energy transfer
 413 mechanism along each resolved principal direction. However, as external perturbations are exerted, nonlinear responses will
 414 take place with large deviation from the original equilibrium statistical data calculated in $Q_{F,\text{eq}}$. Next we will discuss the strategy
 415 to calibrate the model sensitivity to perturbations in a unified way.

416 **2.3. Linear response theory and kicked responses.** The linear response theory as well as fluctuation-dissipation theorem
 417 (FDT) offers a convenient way to get leading-order statistical linear approximation about model responses to perturbations

418 [12, 48, 68, 59]. Consider the general unperturbed system (1.1), $\delta \mathbf{F} = \mathbf{0}$, with invariant measure $p_{\text{eq}}(\mathbf{u})$, and an external forcing
419 perturbation in separation with temporal and spatial variables,

$$420 \quad \delta \mathbf{F}(\mathbf{u}, t) = \mathbf{w}(\mathbf{u}) \delta f(t).$$

421 Therefore the resulting perturbed probability density p^δ can be asymptotically expanded as the equilibrium and the fluctuation
422 correction [48]

$$423 \quad (2.6) \quad p^\delta(t) = p_{\text{eq}} + \delta p'(t), \quad \int_{\mathbb{R}^N} p_{\text{eq}}(\mathbf{u}) d\mathbf{u} = 1, \quad \int_{\mathbb{R}^N} \delta p'(\mathbf{u}) d\mathbf{u} = 0.$$

424 The equilibrium statistics and leading-order correction to the perturbation of some functional about the state variable $A(\mathbf{u})$
425 can be formulated as an asymptotic expansion, $\langle A(\mathbf{u}) \rangle = \langle A(\mathbf{u}) \rangle_{\text{eq}} + \delta \langle A(\mathbf{u}) \rangle(t) + O(\delta^2)$ according to the measure (2.6) with
426 $\langle A(\mathbf{u}) \rangle_{\text{eq}} = \int A(\mathbf{u}) p_{\text{eq}}(\mathbf{u})$ the expectation of A according to equilibrium distribution p_{eq} , while $\delta \langle A(\mathbf{u}) \rangle = \int A(\mathbf{u}) \delta p'(\mathbf{u})$
427 according to $\delta p'$. Therefore we get the leading order responses from

$$428 \quad (2.7) \quad \langle A(\mathbf{u}) \rangle_{\text{eq}} = \int_{\mathbb{R}^N} A(\mathbf{u}) p_{\text{eq}}(\mathbf{u}) d\mathbf{u}, \quad \delta \langle A(\mathbf{u}) \rangle(t) = \int_0^t \mathcal{R}_A(t-s) \delta f(s) ds.$$

429 Above the pointed-bracket denotes the statistical average under the solution from Fokker-Planck equation. $\mathcal{R}_A(t)$ is the *linear*
430 *response operator* corresponding to the functional A , which is calculated through correlation functions in the unperturbed
431 statistical equilibrium (climate) only

$$432 \quad (2.8) \quad \mathcal{R}_A(t) = \langle A[\mathbf{u}(t)] B[\mathbf{u}(0)] \rangle_{\text{eq}}, \quad B(\mathbf{u}) = -\frac{\text{div}_{\mathbf{u}}(\mathbf{w} p_{\text{eq}})}{p_{\text{eq}}}.$$

433 The noise in the equations is not needed for FDT to be valid, but is required to generate the smooth equilibrium measure p_{eq}
434 for the linear response operator \mathcal{R}_A . There is even a rigorous proof of the validity of FDT in this context [33]. Note that even
435 though in general the linear response operator is difficult to calculate considering the complicated and unaccessible equilibrium
436 distribution, a variety of Gaussian approximations for p_{eq} and improved algorithms have been developed for response via FDT
437 [43, 48, 61, 52]. FDT can have high skill for the mean response and some skill for the variance response for a wide variety of
438 turbulent dynamical systems [55, 1, 61, 2, 47, 32].

439 **2.3.1. Calculate linear response operators through initial kicked responses.** The problem in calculating the leading
440 order response using (2.8) is that the equilibrium distribution p_{eq} is expensive to calculate for general systems with non-
441 Gaussian features in a high dimensional phase space. One strategy to approximate the linear response operator which avoids
442 direct evaluation of p_{eq} through the FDT formula but still includes important non-Gaussian statistics is through the *kicked*
443 *response* of an unperturbed system to a perturbation $\delta \mathbf{u}$ of the initial state from the equilibrium measure, that is, to set the initial
444 distribution with the same variance but a perturbation in the mean state

$$445 \quad (2.9) \quad p|_{t=0} = p_{\text{eq}}(\mathbf{u} - \delta \mathbf{u}) = p_{\text{eq}} - \delta \mathbf{u} \cdot \nabla p_{\text{eq}} + O(\delta^2).$$

446 One important advantage of adopting this kicked response strategy is that higher-order statistics due to nonlinear dynamics
447 will not be ignored (compared with the other linearized strategy using only Gaussian statistics [55]). Then the kicked response
448 theory gives the following proposition [48, 58] for calculating the linear response operator:

449 **PROPOSITION 2.2.** *For δ small enough, the linear response operator $\mathcal{R}_A(t)$ can be calculated by solving the unperturbed*
450 *system (1.1) with a perturbed initial distribution in (2.9). Therefore, the linear response operator can be achieved through*

$$451 \quad (2.10) \quad \delta \mathcal{R}_A(t) \equiv \delta \mathbf{u} \cdot \mathcal{R}_A = \int A(\mathbf{u}) \delta p' + O(\delta^2).$$

452 Here $\delta p'$ is the resulting leading order expansion of the transient density function from unperturbed dynamics using initial
453 value perturbation. From the formula in (2.10), the response operators for the mean and variance can be achieved from the
454 perturbation part of the probability density $\delta p'$. And this density function can also be used to measure the information distance
455 between the truth and imperfect models in the training phase.

456 The proof of the above Proposition 2.2 is a direct application of *Duhamel's principle* to the corresponding Fokker-Planck
 457 equation with forcing perturbations [48]. Thus the variability in the external forcing can be transferred to the perturbations in
 458 initial values. More importantly, the kicked response formulation (2.10) with initial mean state perturbation (2.9) is independent
 459 of the specific perturbation forms. Thus the operator \mathcal{R}_A describes the inherent dynamical mechanisms of the system. We
 460 summarize the practical strategies to calculate the kicked response operators for the mean and variance from (2.10) in Appendix
 461 A.

462 2.4. Empirical information theory for measuring imperfect model errors.

463 **2.4.1. Empirical information theory for leading order statistics.** The empirical information theory [38, 48] builds the
 464 least biased probability measure consistent with the leading order measurements of the true perfect system. Information theory
 465 is often used in statistical science for imperfect model selection [11]. A natural way to measure the lack of information in one
 466 probability density from the imperfect model, p^M , compared with the true probability density, p , is through the *relative entropy*
 467 or *information distance* [42, 48], given by

$$468 \quad (2.11) \quad \mathcal{P}(p, p^M) = \int p \log \frac{p}{p^M}.$$

469 Despite the lack of symmetry in its arguments (that is, $\mathcal{P}(p_1, p_2) \neq \mathcal{P}(p_2, p_1)$ in general), the relative entropy, $\mathcal{P}(p, p^M)$
 470 provides an attractive framework for assessing model error like a probabilistic metric. Importantly, the following two crucial
 471 features are satisfied in the relative entropy: (i) $\mathcal{P}(p, p^M) \geq 0$, and the equality holds if and only if $p = p^M$; and (ii) it is invariant
 472 under any invertible change of variables. The most practical setup for utilizing the framework of empirical information theory
 473 arises when only the Gaussian statistics of the distributions are considered. By only comparing the first two moments of the
 474 density functions, we get the following fact [49]:

475 **PROPOSITION 2.3.** *If the probability density functions p, p^M contain only the first two moments, that is, $p \sim \mathcal{N}(\bar{\mathbf{u}}, R)$ and*
 476 *$p^M \sim \mathcal{N}(\bar{\mathbf{u}}_M, R_M)$, the relative entropy in (2.11) has the explicit formula*

$$477 \quad (2.12) \quad \mathcal{P}(p, p^M) = \frac{1}{2} (\bar{\mathbf{u}} - \bar{\mathbf{u}}_M)^T R_M^{-1} (\bar{\mathbf{u}} - \bar{\mathbf{u}}_M) + \frac{1}{2} (\text{tr}(RR_M^{-1}) - N - \log \det(RR_M^{-1})).$$

478 *The first term on the right hand side of (2.12) is called the signal, reflecting the model error in the mean but weighted by the*
 479 *inverse of the model variance R_M ; whereas the second term is the dispersion, involving only the model error covariance ratio*
 480 *RR_M^{-1} , measuring the differences in the covariance matrices.*

481 Above usually we will use p to denote the probability distribution of the perfect model, which is actually unknown. Neverthe-
 482 less, we can construct the measure of the perfect model p_L using L measurements of the true system. Consider the imperfect
 483 model prediction with its associated probability density p_L^M , the definition of relative entropy (2.11) facilitates the practical
 484 calculation [40, 58, 59, 54, 56]

$$485 \quad \mathcal{P}(p, p^M) = \mathcal{P}(p, p_L) + \mathcal{P}(p_L, p_L^M)$$

$$486 \quad = [\mathcal{S}(p_L) - \mathcal{S}(p)] + \mathcal{P}(p_L, p_L^M).$$

487 The entropy difference $\mathcal{S}(p_L) - \mathcal{S}(p)$ precisely measures an intrinsic error from L measurements of the perfect system, and
 488 this is a simple example of an information barrier for any imperfect model based on L measurements for calibration. With the
 489 measurements L representing the first two moments, the Gaussian approximation (2.12) can be used to estimate the information
 490 error $\mathcal{P}(p_L, p_L^M)$ considering only the first L statistical measurements (in practice, it is usually the measurements about the
 491 statistical mean and covariance).

492 **2.4.2. Climate information barrier in single point statistics in homogeneous systems.** Here as one example, we briefly
 493 illustrate the inherent information barrier in special *homogeneous systems* like the L-96 model in (1.9b) (see [63]) with *uniform*
 494 *damping and forcing* using the above relative entropy metric. Of particular interest in both theory and applications, the statistical
 495 mean and variance at each individual grid point [51, 63, 64, 20] play an important role as key statistical quantities to predict.
 496 In climate science, these might be the mean and variance of the surface temperature at every grid point. The single point mean
 497 $\bar{u}_{1\text{pt}}$ and single point variance $r_{1\text{pt}}$ can be defined by averaging each grid component with presumed homogeneity, that is,

$$498 \quad (2.13) \quad \bar{u}_{1\text{pt}} = \frac{1}{N} \sum_{j=1}^N \bar{u}_j, \quad r_{1\text{pt}} = \frac{1}{N} \text{tr}R.$$

499 For simplicity in representation, we assume homogeneous damping, $D = -dI$, and forcing, $F = fI$, in the energy dynamics
500 (2.2), thus the total statistical energy equations for the true model E and reduced-order model approximation E^M become

$$\begin{aligned} 501 \quad & \frac{dE}{dt} = -2dE + f\bar{u}_{1\text{pt}} + \frac{1}{2}\text{tr}Q\sigma, \\ 502 \quad & \frac{dE^M}{dt} = -2dE^M + f\bar{u}_{1\text{pt}}^M + \frac{1}{2}\text{tr}Q_F^M + \frac{1}{2}\text{tr}Q\sigma. \end{aligned}$$

503 Above the statistical energy can be defined through the single point statistics as

$$504 \quad E = \frac{N}{2} (\bar{u}_{1\text{pt}}^2 + r_{1\text{pt}}), \quad E^M = \frac{N}{2} \left((\bar{u}_{1\text{pt}}^M)^2 + r_{1\text{pt}}^M \right),$$

505 where we assume homogeneity in the first two moments. The last part on the right hand side of E^M equation comes from the
506 error in the approximation for nonlinear flux Q_F^M . Taking the difference of the above two equations for E and E^M and using
507 Gronwall's inequality gives the error in the total statistical energy $\delta E = E - E^M$

$$508 \quad \|\delta E\| \leq \tilde{C}_0 \|\delta \bar{u}_{1\text{pt}}\| + C_1 \|\text{tr}Q_F^M\|,$$

509 and the definition about the statistical energy offers the estimation

$$510 \quad \|\delta E\| \geq \|\delta r_{1\text{pt}}\| - N \|\bar{u}_{1\text{pt}}\| \|\delta \bar{u}_{1\text{pt}}\|.$$

511 The error estimation for the single point variance through the error from the mean and nonlinear flux by combining the above
512 two inequalities

$$513 \quad (2.14) \quad \|\delta r_{1\text{pt}}\| \leq C_0 \|\delta \bar{u}_{1\text{pt}}\| + C_1 \|\text{tr}Q_F^M\|,$$

514 with C_0, C_1 constants. The inequality in (2.14) illustrates that the error in the second-order statistics $\delta r_{1\text{pt}}$ can be controlled
515 by the error in the first-order mean with a good approximation for the nonlinear flux term Q_F^M . Similar special results for the
516 40-dimensional L-96 model are described in [63].

517 With the help of the relative entropy, we can first illustrate the inherent information barrier with the single-point statistics
518 approximation (2.13). It will be shown even with consistent single-point statistics in $(\bar{u}_{1\text{pt}}, r_{1\text{pt}})$, large errors may still appear
519 due to the lack of consideration in the covariances between different modes. Through the definition in (2.11) (and referring to
520 Proposition 4.1 of [49]), the relative entropy between the truth p and imperfect model single-point statistics $p_{1\text{pt}}^M$ has the form

$$521 \quad (2.15) \quad \mathcal{P}(p, p_{1\text{pt}}^M) = [\mathcal{S}(p_G) - \mathcal{S}(p)] + \mathcal{P}(p_G, p_{1\text{pt}}^G) + \mathcal{P}(p_{1\text{pt}}^G, p_{1\text{pt}}^M).$$

522 Above, p_G is the Gaussian fit for the original probability distribution p with the same mean and covariance from the truth;
523 $p_{1\text{pt}}^G = \mathcal{N}(\bar{u}_{1\text{pt}}, r_{1\text{pt}})$ is the single-point approximation for the true system, and $p_{1\text{pt}}^M = \mathcal{N}(\bar{u}_M, r_M)$ is the reduced-order model
524 prediction from the imperfect model with consistent single-point statistics. The first part on the right hand side of (2.15) is the
525 intrinsic information barrier in Gaussian approximation. And the third part with homogeneous assumption of the system will
526 vanish (or at least be minimized) due to the single-point statistics fidelity from (2.14). The error from single point approximation
527 (and ignoring the cross-covariance) then comes only from the information barrier in marginal approximation $\mathcal{P}(p_G, p_{1\text{pt}}^G)$ as
528 shown in the second part on the right hand side of (2.15). Simple calculation using the formula (2.12) and Jensen's inequality
529 [63] yields the estimation for the information barrier in single-point approximation

$$530 \quad (2.16) \quad \mathcal{P}(p_G, p_{1\text{pt}}^G) = N \log \left[\frac{\left(\sum_{j=1}^N r_j \right) / N}{\left(\prod_{j=1}^N r_j \right)^{1/N}} \right] \sim r_{1\text{pt}}^{-1} (\sigma_{\max} - \sigma_{\min})^2,$$

531 where r_j is the variance in the spectral modes, and $\sigma_{\max}^2 = \max \{r_j\}$, $\sigma_{\min}^2 = \min \{r_j\}$ are the largest and smallest variance. In
532 Figure 2.1, we demonstrate this information barrier for imperfect models with exact one-point statistics $(\bar{u}_{1\text{pt}}, r_{1\text{pt}})$ consistency
533 for the L-96 model with $F = 5, 8$. Large errors in the statistical steady state spectra (thus information barrier for these models)

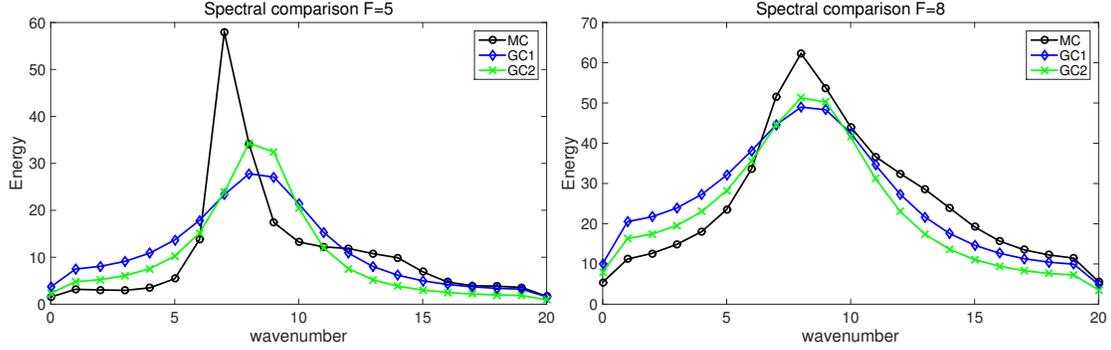


FIG. 2.1. Information barriers for imperfect closure models with only consistent equilibrium single point statistics $(\bar{u}_{1\text{pt}}, r_{1\text{pt}})$ in the L-96 system. The steady state variances under Fourier basis from the two imperfect model results are compared with the truth from Monte-Carlo simulations in two typical dynamical regimes $F = 5$ and $F = 8$.

534 exist for each individual mode for both dynamical regimes $F = 5$ (weakly chaotic) and $F = 8$ (strongly chaotic), consistent with
 535 what we have calculated from (2.16) for single point statistics.

536 The barrier from (2.16) could become significant considering the gap between the largest and smallest variances due to
 537 common decaying energy spectra in turbulent systems. See [63] for a detailed example for the L-96 model. This information
 538 barrier can only be overcome by introducing more careful calibration about the dynamics in each eigen-direction of the system
 539 individually.

540 **2.4.3. Dynamical calibration for imperfect model improvement.** The prediction skill of imperfect models can be im-
 541 proved by comparing the information distance through the linear response operator with the true model. The following fact
 542 offers a convenient way to measure the lack of information in the perturbed imperfect model requiring only knowledge of linear
 543 responses for the mean and variance $\delta\bar{\mathbf{u}} \equiv \delta\mathcal{R}_{\mathbf{u}}$, $\delta R \equiv \delta\mathcal{R}_{(\mathbf{u}-\bar{\mathbf{u}})^2}$. For this result, it is important to tune the imperfect model to
 544 satisfy equilibrium model fidelity,

$$545 \quad \mathcal{P}(p_G(\mathbf{u}), p_G^M(\mathbf{u})) = 0,$$

546 in the first place. Statistical equilibrium fidelity is a natural necessary condition to tune the mean and variance of the imperfect
 547 model to match those of the perfect model; it is far from a sufficient condition [54, 58, 59]. Using simplified assumptions
 548 with block-diagonal covariance matrices $R = \text{diag}(R_k)$ and equilibrium model fidelity $\mathcal{P}(p_G, p_G^M) = 0$, the relative entropy in
 549 (2.11) between the true perturbed density p_δ and the perturbed model density p_δ^M with small perturbation δ can be expanded
 550 componentwisely as the following proposition:

551 **PROPOSITION 2.4.** Under assumptions with block-diagonal covariance matrices $R = \text{diag}(R_k)$ and equilibrium model
 552 fidelity $\mathcal{P}(p_G, p_G^M) = 0$, the relative entropy in (2.12) between perturbed model density p_δ^M and the true perturbed density p_δ
 553 with small perturbation δ can be expanded componentwisely as

$$554 \quad \mathcal{P}(p_\delta, p_\delta^M) = \mathcal{S}(p_{G,\delta}) - \mathcal{S}(p_\delta) \\
 555 \quad \quad \quad + \frac{1}{2} \sum_k (\delta\bar{u}_k - \delta\bar{u}_{M,k}) R_k^{-1} (\delta\bar{u}_k - \delta\bar{u}_{M,k}) \\
 556 \quad (2.17) \quad \quad \quad + \frac{1}{4} \sum_k R_k^{-2} (\delta R_k - \delta R_{M,k})^2 + O(\delta^3).$$

557 Here in the first line $\mathcal{S}(p_{G,\delta}) - \mathcal{S}(p_\delta)$ is the intrinsic error from Gaussian approximation of the system. R_k is the equilibrium
 558 variance in k -th component, and $\delta\bar{u}_k$ and δR_k are the linear response operators for the mean and variance in k -th component.

559 Detailed derivation about this result is shown in [59]. The inherent information error from the first row of (2.17) is due to the
 560 measurement in only first two order of moments, and is independent of the specific imperfect model structures. As a result, this
 561 component, $\mathcal{S}(p_{G,\delta}) - \mathcal{S}(p_\delta)$, can be viewed as a constant and does not need to be calculated in the optimization procedure.

562 The second row of the information distance (2.17) illustrates the *signal error* from the estimation about the mean responses,
 563 while the third row is the *dispersion error* for the errors from the variance responses.

564

565 The above Proposition 2.4 about empirical information theory and linear response theory together provides a convenient and
 566 unambiguous way of improving the performance of imperfect models in terms of increasing their model sensitivity regardless of
 567 the specific form of external perturbations $\delta \mathbf{f}'$. The formula (2.10) in Proposition 2.2 as well as (2.7) illustrates that the skill of
 568 an imperfect model in predicting forced changes to perturbations with general external forcing is directly linked to the model's
 569 skill in estimating the linear response operators \mathcal{R}_A for the mean and variances (that is, use the functional $A = \mathbf{u}, (\mathbf{u} - \bar{\mathbf{u}})^2$ in
 570 calculating the linear response operators) in a suitably weighted fashion as dictated by information theory (2.17). This offers us
 571 useful hints of training imperfect models for optimal responses for the mean and variance in a universal sense. From the linear
 572 response theory, it shows that the system's responses to various external perturbations can be approximated by a convolution
 573 with the linear response operator \mathcal{R}_A (which is only related to the statistics in the unperturbed equilibrium steady state). It is
 574 reasonable to claim that an imperfect model with precise prediction of this linear response operator should possess uniformly
 575 good sensitivity to different kinds of perturbations. On the other hand, the response operator can be calculated easily by the
 576 transient state distribution density function using the kicked response formula as in (2.10). Considering all these good features
 577 of the linear response operator, the information barrier due to model sensitivity to perturbations can be overcome by minimizing
 578 the information error in the imperfect model kicked response distribution relative to the true response from observation data.

579 To summarize, consider a class of imperfect models, \mathcal{M} . The optimal model $M^* \in \mathcal{M}$ that ensures best information
 580 consistent responses to various kinds of perturbations is characterized with the smallest additional information in the linear
 581 response operator \mathcal{R}_A among all the imperfect models, such that

$$582 \quad (2.18) \quad \left\| \mathcal{P} \left(p_\delta, p_\delta^{M^*} \right) \right\|_{L^1([0, T])} = \min_{M \in \mathcal{M}} \left\| \mathcal{P} \left(p_\delta, p_\delta^M \right) \right\|_{L^1([0, T])},$$

583 where p_δ^M can be achieved through a kicked response procedure (2.10) in the training phase compared with the actual observed
 584 data p_δ in nature, and the information distance between perturbed responses $\mathcal{P} \left(p_\delta, p_\delta^M \right)$ can be calculated with ease through
 585 the expansion formula (2.17). The information distance $\mathcal{P} \left(p_\delta(t), p_\delta^M(t) \right)$ is measured at each time instant, so the entire error
 586 is averaged under the L^1 -norm inside a proper time window $[0, T]$ before the linear response function decays back to zero.

587 **3. Reduced-Order Statistical Models for the Turbulent Systems.** Previously in Section 2, the general idea about find-
 588 ing the optimal imperfect model is proposed according to the statistical theories and information distance metric. And we have
 589 shown the basic theoretical tools that can help construct the reduced-order statistical approximations and illustrate the infor-
 590 mation barriers due to these approximations. Then it is important to construct the explicit forms of the reduced-order models
 591 according to the exact dynamics for the mean and covariance in (1.4) and (1.5). Generally the statistical model for the leading
 592 two moments can be formulated in the full phase space as

$$593 \quad (3.1a) \quad \frac{d\bar{\mathbf{u}}_M}{dt} = (L + D) \bar{\mathbf{u}}_M + B(\bar{\mathbf{u}}_M, \bar{\mathbf{u}}_M) + R_{M,ij} B(\mathbf{e}_i, \mathbf{e}_j) + \mathbf{F},$$

$$594 \quad (3.1b) \quad \frac{dR_M}{dt} = L_v(\bar{\mathbf{u}}_M) R_M + R_M L_v^*(\bar{\mathbf{u}}_M) + Q_F^M + Q_\sigma,$$

595 where $\bar{\mathbf{u}}_M \in \mathbb{R}^N$ is the model approximated mean, and R_M is the $N \times N$ full order covariance matrix about the fluctuation
 596 state variable $\mathbf{u}' \in \mathbb{R}^N$. Comparing with the original statistical dynamics (1.4) and (1.5), the most expensive but crucial part
 597 comes from the nonlinear flux term Q_F in (1.6c) where important third-order moments are included representing the nonlinear
 598 interactions between different modes. Therefore the key issue in this section is to construct a judicious estimation about this
 599 nonlinear interaction term Q_F^M in the statistical closure models. Here the basic idea is to start with the simplest possible
 600 imperfect model and compare the advantages and limitations of different levels of imperfect models due to different degrees of
 601 approximation and model calibration, and finally check how the theories from previous sections can help with improving the
 602 model prediction skill, especially the model sensitivity to various perturbations.

603 **3.1. A hierarchy of statistical reduced-order modeling ideas based on stochastic models.** We may consider the statis-
 604 tical closure ideas by taking another look at the dynamics for stochastic coefficients

$$605 \quad \frac{dZ_i}{dt} = Z_j [(L + D) \mathbf{e}_j + B(\bar{\mathbf{u}}, \mathbf{e}_j) + B(\mathbf{e}_j, \bar{\mathbf{u}})] \cdot \mathbf{e}_i + B(\mathbf{u}', \mathbf{u}') \cdot \mathbf{e}_i + \sigma(t) \mathbf{W}(t; \omega) \cdot \mathbf{e}_i.$$

606 Major nonlinearity comes from the term above representing interactions between different fluctuation modes $B(\mathbf{u}', \mathbf{u}') \cdot \mathbf{e}_i$. The
 607 first idea here is to model the effect of the nonlinear energy transfers on each mode by adding additional damping $d_{M,i}$ balancing
 608 the linearly unstable character of these modes, and adding additional (white) stochastic excitation with standard deviation $\sigma_{M,i}$
 609 which will model the energy received by the stable modes. We want to constrain ourselves to second order models concentrating
 610 on the mean and variance and maintaining the computational expense in a low level, hence the additional parts $d_{M,i}, \sigma_{M,i}$ only
 611 include statistics up to second order moments. Specifically we replace this high-order nonlinear term by

$$612 \quad B(\mathbf{u}', \mathbf{u}') \cdot \mathbf{e}_i \equiv \sum_{m,n} Z_m Z_n B(\mathbf{e}_m, \mathbf{e}_n) \cdot \mathbf{e}_i \rightarrow -d_{M,i}(\text{tr}R) Z_i + \sigma_{M,i}(\text{tr}R) \dot{W}_i,$$

613 with $\text{tr}R = \sum_j \langle Z_j Z_j^* \rangle$ for measuring the total energy (variance) structure in the system. Corresponding to the statistical equa-
 614 tions, the nonlinear flux Q_F representing the higher-order interactions is replaced by

$$615 \quad (3.2) \quad Q_F^M = Q_{F-}^M + Q_{F+}^M = -D_M(R) R_M - R_M D_M^*(R) + \Sigma_M(R).$$

616 In (3.2), (D_M, Σ_M) are $N \times N$ matrices that replace the original nonlinear unstable and stable effects from the original dynamics.
 617 Here $Q_{F-}^M = -D_M(R) R_M - R_M D_M^*(R)$ represents the additional damping effect to stabilize the unstable modes with positive
 618 Lyapunov coefficients, while $Q_{F+}^M = \Sigma_{M,k}(R)$ is the positive-definite additional noise to compensate for the overdamped modes.
 619 Now the problem is converted to finding expressions for D_M and Σ_M . In the following by gradually adding more detailed
 620 characterization about the statistical dynamical model we display the general procedure of constructing a hierarchy of the
 621 closure methods step by step. Below is a review about several model closure ideas [54, 88, 63, 78] with increasing complexity:

622 1. *Quasilinear Gaussian closure model*: The simplest approximation for the closure methods at the first stage should be
 623 simply neglecting the nonlinear part entirely [23, 25, 90]. That is, set

$$624 \quad (3.3) \quad D_M(R) \equiv 0, \quad \Sigma_M(R) \equiv 0, \quad Q_F^{\text{QG}} \equiv 0.$$

625 Thus the nonlinear energy transfer mechanism will be entirely neglected in this Gaussian closure model. This is
 626 the similar idea in the *eddy-damped Markovian model* where the moment hierarchy is closed at the level of second
 627 moments with Gaussian assumption and a much larger *eddy-damped* parameter is introduced to replace the molecular
 628 viscosity (see Chapter 5 of [84] and [44] for details). Obviously this crude Gaussian approximation will not work well
 629 in general due to the cutoff of the energy flow when strong nonlinear interactions between modes occur. Actually, the
 630 deficiency of this crude approximation have been shown under the L-96 framework, and in final equilibrium state there
 631 exists only one active mode with critical wavenumber [89, 63]. Such closures are only useful in the weakly nonlinear
 632 case where the quasi-linear effects are dominant.

633 2. *Models with consistent equilibrium statistics*: Next the strategy is to construct the simplest closure model with con-
 634 sistent equilibrium statistics. So the direct way is to choose constant damping and noise term at most scaled with the
 635 total variance. We propose two possible choices as in [63] for the damping and noise in (3.2) below.

636 **Gaussian closure 1 (GC1)**: let

$$637 \quad (3.4) \quad D_M(R) = \varepsilon_M I_N \equiv \text{const.}, \quad \Sigma_M(R) = \sigma_M^2 I_N \equiv \text{const.}, \quad Q_F^{\text{GC1}} = -(\varepsilon_M R + R \varepsilon_M) + \sigma_M^2 I_N;$$

638 **Gaussian closure 2 (GC2)**: let
 639 (3.5)

$$640 \quad D_M(R) = \varepsilon_M \left(\frac{\text{tr}R}{\text{tr}R_{\text{eq}}} \right)^{1/2} I_N, \quad \Sigma_M(R) = \sigma_M^2 \left(\frac{\text{tr}R}{\text{tr}R_{\text{eq}}} \right)^{3/2} I_N, \quad Q_F^{\text{GC2}} = - \left(\frac{\text{tr}R}{\text{tr}R_{\text{eq}}} \right)^{1/2} (\varepsilon_M R + R \varepsilon_M) + \sigma_M^2 \left(\frac{\text{tr}R}{\text{tr}R_{\text{eq}}} \right)^{3/2} I_N.$$

641 Above only two scalar model parameters $(\varepsilon_M, \sigma_M)$ are introduced, and I_N represents the $N \times N$ identity matrix. GC1
 642 is the familiar strategy of adding constant damping and white noise forcing to represent nonlinear interaction; GC2
 643 scales with the total variance $\text{tr}R$ (or total statistical energy) so that the model sensitivity can be further improved as
 644 the system is perturbed. From both GC1 and GC2, we introduce uniform additional damping rate for each spectral
 645 mode controlled by a single scalar parameter ε_M ; while the additional noise with variance σ_M^2 is added to make sure
 646 climate fidelity in equilibrium (we leave the detailed discussion for climate fidelity in Section 3.2.1).

The statistical model closure Q_F^M is used to approximate the third-order moments in the true dynamics, thus the exponents of the total energy $\text{tr}R$ in GC2 should be consistent in scaling dimension. In the positive-definite part $Q_{F,+}^M$, it calibrates the rate of energy injected into the spectral mode due to nonlinear effect in the order $|u'|^3$. The factor scales with the total energy with exponent $3/2$ so that the corrections keep consistent with the third-order moment approximations; In the negative damping rate $Q_{F,-}^M$, the scaling function is used to characterize the amount of energy that flows out the spectral mode due to nonlinear interactions. Scaling factor with a square-root of the total energy with exponent $1/2$ is applied for this damping rate multiplying the variance in order $|u'|^2$ to make it consistent in scaling dimension with third moments.

3. *Modified quasi-Gaussian closure with equilibrium statistics*: In this modified quasi-Gaussian closure model originally proposed in [89, 88], we exploit more about the true nonlinear energy transfer mechanism from the equilibrium statistical information. Thus the additional damping and noise proposed like before are calibrated through the equilibrium nonlinear flux by letting

$$(3.6) \quad D_M(R) = -N_{M,\text{eq}}, \quad \Sigma_M(R) = Q_{F,\text{eq}}^+, \quad Q_F^{\text{MQG}} = -(N_M R + R N_M^*) + Q_F^+;$$

$N_{M,\text{eq}}$ is the effective damping from equilibrium, and $Q_{F,\text{eq}}^+$ is the effective noise from the positive-definite component. Unperturbed equilibrium statistics in the nonlinear flux $Q_{F,\text{eq}}$ are used to calibrate the higher-order moments as additional energy sink and source. The true equilibrium higher-order flux can be calculated without error from first and second order moments in $(\bar{\mathbf{u}}_{\text{eq}}, R_{\text{eq}})$ from the unperturbed true dynamics (1.5) in steady state following the steady state statistical solution relation (2.3) as discussed in Section 2.2

$$(3.7) \quad Q_{F,\text{eq}} = Q_{F,\text{eq}}^- + Q_{F,\text{eq}}^+ = -L_v(\bar{\mathbf{u}}_{\text{eq}})R_{\text{eq}} - R_{\text{eq}}L_v^*(\bar{\mathbf{u}}_{\text{eq}}) - Q_\sigma, \quad N_{M,\text{eq}} = \frac{1}{2}Q_{F,\text{eq}}^-R_{\text{eq}}^{-1}.$$

$Q_{F,\text{eq}}^-$, $Q_{F,\text{eq}}^+$ are the negative and positive definite components in the unperturbed equilibrium nonlinear flux $Q_{F,\text{eq}}$. Since exact model statistics are used in the imperfect model approximations, the true mechanism in the nonlinear energy transfer can be modeled under this first correction form. This is the similar idea used for measuring higher-order interactions in [88], where more sophisticated and expensive calibrations are required to make that model work there.

3.2. A reduced-order statistical energy model with optimal consistency and sensitivity. The above closure model ideas, especially (3.4), (3.5), and (3.6), have advantages of their own. Models in (3.4) and (3.5) are simple and efficient to construct with consistent equilibrium consistency, while (3.6) involves the true information about the higher-order statistics in equilibrium so that the energy mechanism can be characterized well. The validity of these approaches has been tested and compared from several papers [89, 88, 63] using the simplified triad model and L-96 model. Still when it comes to the more complicated and realistic flow systems like the quasi-geostrophic equations, more detailed calibration for model consistency and sensitivity is required to achieve the optimal performance. A preferred approach for the nonlinear flux Q_F^M combining both the detailed model energy mechanism and control over model sensitivity is proposed in the form

$$(3.8) \quad Q_F^M = Q_F^{M,-} + Q_F^{M,+} = f_1(R) \left[-(N_{M,\text{eq}} + d_M I_N) R_M \right] + f_2(R) \left[Q_{F,\text{eq}}^+ + \Sigma_M \right].$$

The closure form (3.8) consists of three indispensable components:

- i) *Higher-order corrections from equilibrium statistics*: In the first part of the correction using the damping and noise operator as $(N_{M,\text{eq}}, Q_{F,\text{eq}}^+)$, unperturbed equilibrium statistics in the nonlinear flux $Q_{F,\text{eq}}$ are used to calibrate the higher-order moments as additional energy sink and source following the procedure in (3.6). Therefore the equilibrium statistics can be guaranteed to be consistent with the truth, and the true energy mechanism can be restored;
- ii) *Additional damping and noise to model changes in nonlinear flux*: The above corrections in step i) by using equilibrium information for nonlinear flux is found to be insufficient for accurate prediction in the reduced-order methods since the scheme is only marginally stable and the energy transferring mechanism may change with large deviation from the equilibrium case when external perturbations are applied. Thus we also introduce the additional damping and noise (d_M, Σ_M) as from (3.4). d_M is just a constant scalar parameter to add uniform dissipation on each mode, and Σ_M is the further correction as an additional energy source to maintain climate fidelity;
- iii) *Statistical energy scaling to improve model sensitivity*: Still note that these additional parameters are added regardless of the true nonlinear perturbed energy mechanism where only unperturbed equilibrium statistics are used. To capture

693 the responses to a specific perturbation forcing, it is better to make the imperfect model parameters change adaptively
 694 according to the total energy structure. Considering this, the additional damping and noise corrections are scaled with
 695 factors $f_1(R)$, $f_2(R)$ related with the total statistical variance $\text{tr}R$ as

$$696 \quad (3.9) \quad f_1(R) = \left(\frac{\text{tr}R}{\text{tr}R_{\text{eq}}} \right)^{1/2}, \quad f_2(R) = \left(\frac{\text{tr}R}{\text{tr}R_{\text{eq}}} \right)^{3/2}.$$

697 Note that in the full model formulation (3.1a) and (3.1b) with the entire covariance matrix R resolved, the total variance
 698 structure $\text{tr}R$ is easy to achieve. However in the low-order models with only the variances in the principal modes resolved
 699 explicitly as will be discussed in the following subsection, $\text{tr}R$ is generally not available directly. This is where the statistical
 700 energy dynamics (2.2) can play an important role and help the development of reduced order models. Besides, a further skew-
 701 symmetric correction for dispersion effects in addition to the scalar model damping d_M in the reduced-order models might be
 702 useful in some situations as the following remark.

703 *Remark 3.1.* In the additional damping correction in (3.8), only a scalar damping parameter d_M is considered. A little more
 704 detailed calibration about the nonlinear exchange of energy is to also introduce an imaginary skew-symmetric operator $i\Omega_M$
 705 applied on the covariance R_M , that is,

$$706 \quad Q_F^{M,-} = (-d_M I_N + i\Omega_M) R_M$$

707 This term will not alter the entire energy structure of the system due to skew symmetry but can offer correction for the dispersion
 708 relation in this imperfect model. However, we may have the additional difficulty in fitting the N by N parameter matrix in the
 709 general case. In practical applications, instead we can exploit the physical structure of the specific model and introduce only
 710 one additional dispersion parameter $i\omega_M$; see [78] and [79] for two examples of adding the dispersion correction to effectively
 711 improve model prediction skill under the barotropic and baroclinic models.

712 Next we discuss the detailed calibrations about the nonlinear flux approximations. Two steps of model calibration should
 713 be considered as from the general framework described in Section 1.2: i) *the equilibrium consistency* that the reduced model
 714 must converge to the true equilibrium statistical state as no perturbations are added; ii) *model sensitivity* by blending statistical
 715 response and information theory so that the imperfect model can capture the responses to various kinds of perturbations as the
 716 system is perturbed. The construction in (3.6) guarantees equilibrium consistency using the true equilibrium model nonlinear
 717 flux structure. On the other hand, to improve model sensitivity, the linear response operators with information distance metric
 718 are used to find optimal parameters from the correction part in (3.4) or (3.5).

719 **3.2.1. Equilibrium statistical fidelity through the additional damping and noise.** In designing the reduced-order mod-
 720 els, equilibrium fidelity for consistent statistics should be guaranteed in the first place in the unperturbed climate. That is, the
 721 same final unperturbed statistical equilibrium R_{eq} should be recovered from the closure models R_M in each component. Com-
 722 paring the true statistical equation (1.5) with the reduced-order model (3.1b), time derivatives about the statistics on the left
 723 hand sides vanish in statistical steady state, thus climate consistency can be achieved if we have exact recovery of the estimation
 724 in the nonlinear flux term. Specifically, it requires that the model nonlinear flux correction term (3.8) converges to the truth,
 725 $Q_M \rightarrow Q_{F,\text{eq}}$, when no external perturbation is added. Under this condition in steady state the closure model covariance equation
 726 (3.1b) goes to the true unperturbed statistics, the equilibrium statistical relation (2.3) implies the relation

$$727 \quad 0 = L_v(\bar{\mathbf{u}}_{\text{eq}}) R_{M,\text{eq}} + R_{M,\text{eq}} L_v^*(\bar{\mathbf{u}}_{\text{eq}}) + Q_{F,\text{eq}}^M + Q_\sigma \rightarrow R_{M,\text{eq}} = R_{\text{eq}}.$$

728 In construction the first component $(N_{M,\text{eq}}, Q_{F,\text{eq}}^+)$ comes from the true equilibrium statistics, and in equilibrium state it will
 729 guarantee the consistency with the truth that

$$730 \quad - (N_{M,\text{eq}} R_{\text{eq}} + R_{\text{eq}} N_{M,\text{eq}}^*) + Q_{F,\text{eq}}^+ = Q_{F,\text{eq}}.$$

731 This part will be automatically equal to the true nonlinear flux in equilibrium. On the other hand climate consistency re-
 732 quires that the second component correction due to the parameters $(d_M I_N, \Sigma_M)$ adds no additional energy source or sink in the
 733 unperturbed system, and no further correction in the scaling functionals. That is, we need Σ_M to satisfy

$$734 \quad (3.10) \quad \Sigma_M = \frac{1}{2} d_M R_{\text{eq}}, \quad f_1(\text{tr}R_{\text{eq}}) = 1, \quad f_2(\text{tr}R_{\text{eq}}) = 1.$$

735 Note again f_1 and f_2 in (3.10) calibrate the model sensitivity to perturbations according to the total energy structure $\text{tr}R$. Thus
 736 it is natural to assume no additional correction in the unperturbed case.

737 By choosing parameters according to (3.10), the climate consistency for the imperfect reduced-order models in (3.1b) in
 738 the unperturbed equilibrium is guaranteed. In addition, we still leave one controlling parameter d_M for the freedom to tune the
 739 imperfect model performance, considering that climate consistency is only the necessary but not sufficient condition for good
 740 model prediction [59].

741 **3.2.2. Model calibration blending statistical response and information theory.** The above methods (3.4), (3.5), (3.6),
 742 as well as (3.8) construct statistical approximation models with consistent equilibrium statistics. Still equilibrium fidelity
 743 of imperfect models is a necessary but not sufficient condition for model prediction skill with many examples [54, 59, 63].
 744 In order to get precise forecasts for various forced responses, it is also crucial to seek models that can correctly reflect the
 745 system's 'memory' to its previous states. From Section 2.2, it shows that the linear response operator \mathcal{R}_A represents the lagged-
 746 covariance of certain functions (and thus can describe the 'memory' of the system to previous states). We try to find a unified
 747 way to achieve the optimal model parameters d_M such that the imperfect models can maintain high performance for various
 748 kinds of external perturbations. Adopting the general strategy suggested in Section 2.3, we can improve model sensitivity
 749 through tuning imperfect models in a training phase before the prediction step. Thus the optimal model parameter can be
 750 selected through minimizing the information distance in the linear response operators in (2.18) between the imperfect closure
 751 model and the truth.

752 *Information-theoretical framework to measure the linear responses in the training phase.* In this training phase, we try
 753 to find the optimal model parameters d_M by comparing the linear response operators from the true system and imperfect
 754 approximation model. The true model linear response operator and the reduced-order model response operator can be calculated
 755 from (2.8), following the procedure from the kicked response strategy with detailed procedure shown in Appendix A. The
 756 distance between these two operators can be calculated through the information metric (2.17) which offers an unbiased and
 757 invariant measure for model distributions

$$758 \quad \mathcal{D}(p_\delta, p_\delta^M) = \frac{1}{2} \sum_k (\delta \bar{u}_k - \delta \bar{u}_{M,k}) R_k^{-1} (\delta \bar{u}_k - \delta \bar{u}_{M,k})$$

$$759 \quad + \frac{1}{4} \sum_k R_k^{-2} (\delta R_k - \delta R_{M,k})^2 + O(\delta^3).$$

760 The first row above is the signal error due to the estimation about the mean; and the second row is the dispersion error
 761 for calibrating the linear responses in the first two order of moments, δR_k . The intrinsic error due to second-order closure
 762 $\mathcal{S}(p_{G,\delta}) - \mathcal{S}(p_\delta)$ is independent of the specific forms of the reduced-order models and is not included in this metric. The
 763 optimization principle in (2.18) is then performed over the parameter d_M .

764 **3.2.3. Comparisons with stochastic modeling about the mean and fluctuations and realizability.** To achieve a better
 765 understanding about the statistical models, it is useful to compare the reduced-order statistical energy model (3.1a) and (3.1b)
 766 with its stochastic correspondences. In the stochastic formulation, we consider the separation with a deterministic mean state
 767 and the stochastic fluctuations

$$768 \quad \mathbf{u}_M = \bar{\mathbf{u}}_M(t) + \sum_j Z_j(t) \mathbf{e}_j,$$

769 where $\bar{\mathbf{u}}_M = \langle \mathbf{u}_M \rangle$ is the statistical mean state following the same dynamical mean equation as before together with the stochastic
 770 dynamics for the fluctuation modes

$$771 \quad (3.11a) \quad \frac{d\bar{\mathbf{u}}_M}{dt} = (L + D) \bar{\mathbf{u}}_M + B(\bar{\mathbf{u}}_M, \bar{\mathbf{u}}_M) + \langle Z_i Z_j^* \rangle B(\mathbf{e}_i, \mathbf{e}_j) + \mathbf{F},$$

$$772 \quad (3.11b) \quad \frac{dZ_i}{dt} = Z_j [(L + D) \mathbf{e}_j + B(\bar{\mathbf{u}}_M, \mathbf{e}_j) + B(\mathbf{e}_j, \bar{\mathbf{u}}_M)] \cdot \mathbf{e}_i - d_{M,i} (\langle \mathbf{Z}\mathbf{Z}^* \rangle) Z_i + \Sigma_{M,ij} (\langle \mathbf{Z}\mathbf{Z}^* \rangle) \dot{W}_{ij}.$$

773 Above the effective damping and noise ($d_M I_N, \Sigma_M$) are added in the same way as constructed in (3.8). The mean dynamics
 774 (3.11a) get the small scale feedbacks from the nonlinear statistical interaction $\langle Z_i Z_j^* \rangle B(\mathbf{e}_i, \mathbf{e}_j)$, while the fluctuation stochastic
 775 dynamics are linked with the mean state through the quasilinear interactions. By direct comparison with the statistical equations
 776 (3.1a) and (3.1b), we see that the mean equation is identical while the equation for the stochastic fluctuations differs in the

777 nonlinear term. The constructed set of closed stochastic equations is a representative of a new class of stochastic systems where
 778 the evolution of each stochastic realization depends on the global statistics, i.e. on the collective or statistical behavior of all the
 779 realizations due to $\langle \mathbf{Z}\mathbf{Z}^* \rangle$. In particular, the associated formal Fokker-Planck equation becomes nonlinear and nonlocal. This
 780 guarantees realizability of the reduced-order models. These novel stochastic equations deserve further mathematical study as a
 781 complex version of McKean-Vlasov systems [69].

782 **3.3. Reduced-order statistical model for principal modes.** Here for genuinely high-dimensional systems, the computa-
 783 tional cost for the full covariance matrix R_M is still unaffordable even with the first two moment closure [6, 56]; for example,
 784 climate systems usually have the dimensionality at least of order 10^3 . On the other hand, in many situations, we are mostly
 785 interested in the variability in the statistics of the first most energetic principal directions. Therefore one alternative practical
 786 strategy is to develop reduced-order methods that only calculate variances in a low-dimensional subspace spanned by primary
 787 EOFs $\{\mathbf{v}_1, \dots, \mathbf{v}_s\}$ with $s \ll N$ (N the dimensionality of the full system). The corresponding reduced-order representation of
 788 the state variables under these resolved basis becomes $\mathbf{u} = \bar{\mathbf{u}} + \sum_{k=1}^s Y_k \mathbf{v}_k$. To see the possibility of achieving this, first note
 789 that the dynamical equations for variances (3.1b) in each mode $r_k = \langle Y_k Y_k^* \rangle$ are rather independent with each other according
 790 to the previous closure strategies with higher-order interactions replaced by additional damping and noise terms. Thus it is
 791 realizable to restrict the variance equations inside the chosen subspace. Actually following the same strategy by replacing the
 792 high-order interaction terms by proper damping and noise, the equivalent counterpart of the closure models can be formulated
 793 as a low-order stochastic system

$$794 \quad (3.12a) \quad \frac{d\bar{\mathbf{u}}_M}{dt} = (L+D)\bar{\mathbf{u}}_M + B(\bar{\mathbf{u}}_M, \bar{\mathbf{u}}_M) + \sum_{i,j \leq s} C_{M,ij} B(\mathbf{v}_i, \mathbf{v}_j) + \mathbf{F} + \mathbf{G},$$

$$795 \quad (3.12b) \quad \frac{dC_M}{dt} = L_v^{\text{red}} C_M + C_M L_v^{\text{red}*} + Q_{F,M}^{\text{red}} + Q_\sigma^{\text{red}}, \quad C_M \in \mathbb{C}^{s \times s}.$$

796 The mean dynamics (3.12a) is the same as the previous closure model (3.1a) with an additional correction term \mathbf{G} to compensate
 797 the unresolved modes. C_M is the reduced-order $s \times s$ covariance matrix where only the leading primary modes are resolved, that
 798 is,

$$799 \quad C_M = P^* R_M P,$$

800 where $P = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_s] \in \mathbb{C}^{N \times s}$ projects the modes to the subspace.

801 Through proper choice of the parameters according to GC1 (3.4), GC2 (3.5), MQG (3.6), or the blended method (3.8) as
 802 before but concentrating on the resolved subspace, these reduced system should converge to the same first two order statistics
 803 with the moment closure model. Still several new problems need to be taken care of for the above model reduction process: i)
 804 How to ensure correct modeling about the true statistics in the mean dynamics (3.12a) due to the many unresolved directions in
 805 the covariance C_M ; ii) How to include the nonlocal scale factor (which always includes the total energy $\text{tr}R = \sum_{k=1}^N r_k$) in (3.9)
 806 in the nonlinear flux approximation $Q_{F,M}^{\text{red}}$ if only subspace variances are resolved. Here in the general strategy to do this, we
 807 follow the ideas in [63, 78, 79].

808 **3.3.1. Correction for the mean dynamics.** Still the simplest way of estimating the unresolved parts in the mean dynamics
 809 is through the statistical equilibrium information \mathbf{G}_{eq} . The value of the additional forcing \mathbf{G}_{eq} is determined using statistical
 810 steady state information for the covariance C_{eq} and the mean $\bar{\mathbf{u}}_{\text{eq}}$. In particular we have the equilibrium equation through the
 811 steady state mean dynamics where $\frac{d}{dt} \bar{\mathbf{u}}_{\text{eq}} \equiv 0$

$$812 \quad (3.13) \quad \mathbf{G}_{\text{eq}} = -(L+D)\bar{\mathbf{u}}_{\text{eq}} + B(\bar{\mathbf{u}}_{\text{eq}}, \bar{\mathbf{u}}_{\text{eq}}) - \sum_{i,j \leq s} C_{\text{eq},ij} B(\mathbf{v}_i, \mathbf{v}_j) - \mathbf{F}_{\text{eq}}.$$

813 Similar as in the estimation about the nonlinear flux, the mean dynamics correction term can also be scaled with the total
 814 variance in the system, so that,

$$815 \quad (3.14) \quad \mathbf{G} = \frac{\text{tr}R}{\text{tr}R_{\text{eq}}} \mathbf{G}_{\text{eq}}.$$

816 In this way, the mean dynamics (3.12a) become consistent in the statistical equilibrium state, and the corrections \mathbf{G} can change
 817 sensitively according to the total energy structure through $\text{tr}R$.

818 *Remark 3.2.* One further optional correction for the unresolved modes in the reduced-order mean equations is to make
 819 further use of the linear response theory predictions. To estimate the values for unresolved modes, we can improve it from the
 820 equilibrium statistics by introducing finer approximation making use of the linear response operator (2.7)

$$821 \quad (3.15) \quad r_{k,\text{un}} \sim r_{k,\text{eq}} + \delta r'_k = r_{k,\text{eq}} + \int_0^t \mathcal{R}_{r_k}(t-s) \delta F^l(s) ds, \quad k > s.$$

822 Therefore these first-order predictions for the unresolved variances $r_{k,\text{un}}$ can also be used in (3.12a) for estimating the unresolved
 823 modes. This idea is applied to the L-96 system with improvements shown in [63].

824 **3.3.2. Correction through total statistical energy.** In the reduced-order covariance dynamics (3.12b) using the closure
 825 form (3.8), two additional scaling factors, f_1, f_2 , are introduced to further quantify the nonlinear energy flux in and out the
 826 spectral modes due to the nonlinear interactions. We propose the dynamical corrections with the total statistical energy $\text{tr}R$
 827 as in the forms (3.9). This total energy correction introduces global information into each spectral mode so the nonlinear
 828 energy transfer can be better characterized in the imperfect model, while solving only one additional scalar equation is the only
 829 additional cost in computation. The scaling factor from $\text{tr}R$ introduces nonlinear global effect into the additional damping and
 830 noise corrections in each mode. This can be solved efficiently by introducing one additional scalar equation as described in
 831 (2.2)

$$832 \quad \frac{dE}{dt} = \bar{\mathbf{u}} \cdot D\bar{\mathbf{u}} + \bar{\mathbf{u}} \cdot \mathbf{F} + \text{tr}(DR) + \frac{1}{2}\text{tr}Q\sigma.$$

833 Then $\text{tr}R$ can be achieved by solving $E = \frac{1}{2}\bar{\mathbf{u}}^2 + \frac{1}{2}\text{tr}R$.

834 Especially in uniform damping case, $D = -dI$, the above statistical energy equation can be simplified as

$$835 \quad (3.16) \quad \frac{dE}{dt} = -dE + \bar{\mathbf{u}} \cdot \mathbf{F} + \frac{1}{2}\text{tr}Q\sigma.$$

836 Note that on the right hand side of (3.16), only mean state information (which can be fully resolved in the reduced mean
 837 dynamics in (3.12a)) needs to be calculated to get the total statistical energy E . In this way, the total second order moments $\text{tr}R$
 838 can be entirely determined only through the first order mean $\bar{\mathbf{u}}$ and the scalar statistical energy equation (3.16).

839 The final issue in the reduced-order model construction is about the tuning process in the training phase. Still the same
 840 kicked-response strategy (2.10) can be applied to the reduced-order formulation (3.12). Importantly, the relative entropy ex-
 841 pansion for the responses in (2.17) is decomposed into each component. Thus it can be directly applied to the reduced-order
 842 case by calculating only the signal and dispersion error in the resolved subspace. Therefore through the same procedure as the
 843 previous case, we can find the optimal model parameter in the training phase for the reduced-order model, and then apply the
 844 optimal model for prediction with various forcing perturbations.

845 **3.4. Summary of the *Reduced-Order Statistical Energy Closure* algorithm.** We summarize the low-dimensional reduced-
 846 order statistical closure algorithm with calibrations from total statistical energy and linear response theory. The general reduced-
 847 order model algorithm is split into the separated steps of a *training phase* and a *prediction phase* after a proper *imperfect model*
 848 *selection* step according to the problem. The training phase is used to improve model sensitivity by tuning the imperfect model
 849 parameter using only unperturbed equilibrium statistics for the linear response operator. Then the optimal parameter can be
 850 applied for predicting model responses to different kinds of external perturbations. Note that in the calibration step in the
 851 algorithm, only the unperturbed statistics in equilibrium are required. Thus this offers the optimal model parameters that are
 852 ideally valid for all kinds of specific forcing perturbation forms. With the help of the linear response operator we are able to
 853 find a unified way to tune the imperfect model parameters and avoid the exhausting and impractical process to tune the models
 854 each time with different kinds of perturbations.

855 **4. Reduced-Order Statistical Models Applied to a Suite of Stochastic Triad Models.** In this section, we illustrate
 856 the performance of the reduced-order formulations by considering a simple but nevertheless instructive model, namely the
 857 triad system with stochastic forcing [30, 50, 54, 89]. The triad systems where three modes interact through quadratic energy-
 858 conserving nonlinear interactions form the building block for more general complex turbulent flow, thus provide a nice simple
 859 test case for the mode elimination strategy in the first stage. The nonlinear interaction in triad systems is generic of nonlinear
 860 coupling between any three modes in larger systems with quadratic nonlinearity. For a three-dimensional system about state

Algorithm 1 Reduced-order statistical closure model for general turbulent systems

- Model selection stage:
 - Decide the low-dimensional subspace spanned by orthonormal basis $\{\mathbf{v}_k\}_{k=-s}^s$ covering the directions with largest variances (energy) among the spectrum.
 - Set up statistical dynamical equations (3.12) by Galerkin projecting the original equations to the resolved subspace for modes with wavenumbers $1 \leq |k| \leq s$, as well as the statistical energy equation (3.16) to get the total statistical energy E_M in the system.
- Model calibration stage:
 - Construct low-order approximation of the nonlinear flux Q_F^M in the statistical equations using the statistical energy closure proposed in (3.8) consistent with the equilibrium first two moments;
 - Compute the true linear response operator from the unperturbed equilibrium statistics, and calculate the imperfect model predicted linear response operator from the kicked response strategy;
 - Determine the imperfect model parameter value through minimizing the information distance in (2.17) and (2.18) between linear response operators from true equilibrium statistics and imperfect model approximation.
- Model prediction stage:
 - Use the optimally tuned parameter achieved from the previous step in the reduced-order model to get statistical responses of the state variables of interest in principal directions with all kinds of specific external perturbations.

861 variables $\mathbf{u} = (u_1, u_2, u_3)^T \in \mathbb{R}^3$ with a quadratic part that is energy preserving, the triad system possesses the general form

$$862 \quad (4.1a) \quad \frac{du_1}{dt} = L_2 u_3 - L_3 u_2 - d_1 u_1 + B_1 u_2 u_3 + F_1 + \sigma_1 \dot{W}_1,$$

$$863 \quad (4.1b) \quad \frac{du_2}{dt} = L_3 u_1 - L_1 u_3 - d_2 u_2 + B_2 u_3 u_1 + F_2 + \sigma_2 \dot{W}_2,$$

$$864 \quad (4.1c) \quad \frac{du_3}{dt} = L_1 u_2 - L_2 u_1 - d_3 u_3 + B_3 u_1 u_2 + F_3 + \sigma_3 \dot{W}_3.$$

865 The triad system (4.1) is easy to summarize in the original abstract formulation (1.1), where

$$866 \quad L = \begin{bmatrix} 0 & -L_3 & L_2 \\ L_3 & 0 & -L_1 \\ -L_2 & L_1 & 0 \end{bmatrix}, \quad D = \begin{bmatrix} -d_1 & & \\ & -d_2 & \\ & & -d_3 \end{bmatrix},$$

867 are the skew-symmetric and dissipation operator in (1.2a) representing respectively the Coriolis forcing and dissipation; and
868 $B(\mathbf{u}, \mathbf{u})$ satisfies

$$869 \quad B(\mathbf{u}, \mathbf{u}) = (B_1 u_2 u_3, B_2 u_3 u_1, B_3 u_1 u_2)^T, \quad B_1 + B_2 + B_3 = 0,$$

870 which forms the nonlinear triad coupling that satisfies the energy conservation in (1.2b). The triad system can form the building
871 block of complex turbulent dynamical systems since it can be viewed as a three-dimensional Galerkin truncation of many
872 general dynamics. One celebrated example is the famous Lorenz model [45] that can be viewed as a special case of this
873 procedure. An interpretation of these low-order models with atmospheric problems and geoscience is illustrated in [30]. Though
874 simple in appearance of this triad system, complex and interesting statistical features can be generated through changing the
875 model parameters.

876 **4.1. Statistical properties for the triad system.** First we can check the general statistical properties described in Section
877 1.2 with the triad system. Typically we would like to investigate the evolution of a smooth probability density function $p(\mathbf{u}, t)$
878 due to the internal and external stochasticity. Associated with the triad equations (4.1), the statistical solution satisfies the
879 Fokker-Planck equation

$$880 \quad (4.2) \quad \frac{\partial p}{\partial t} = -(B(\mathbf{u}, \mathbf{u}) + (L + D)\mathbf{u} + \mathbf{F}) \cdot \nabla_{\mathbf{u}} p + \sum_{i=1}^3 \left(d_i p + \frac{1}{2} \sigma_i^2 \partial_{u_i}^2 p \right),$$

$$p(\mathbf{u}, t) |_{t=0} = p_0(\mathbf{u}).$$

881 Above we assume the forcing terms are only dependent on time, $\mathbf{F} \equiv \mathbf{F}(t)$, $\sigma_i \equiv \sigma_i(t)$. While the original triad system (4.1) is
 882 nonlinear, the statistical dynamics in (4.2) are linear equation for smooth functions p . The Fokker-Planck equation will reduce
 883 to the *Liouville equation* in the case of zero stochastic noise $\sigma \equiv 0$. However the explicit solution of the Fokker-Planck equation
 884 (4.2) is still difficult to get directly even with the triad model.

885 **4.1.1. Equilibrium invariant measure with equipartition of energy.** In general the explicit solutions for the Fokker-
 886 Planck equation above (4.2) is difficult to achieve due to the nonlinear interactions in the triad system. Still under special
 887 arrangement about the damping and noise coefficients, one special solution of a Gaussian invariant measure, p_{eq} , can be reached
 888 in the equilibrium. In the absence of the deterministic forcing, $\mathbf{F} = \mathbf{0}$, assume the damping operator d_i and random noise forcing
 889 σ_i satisfy the following relation in each component

$$890 \quad (4.3) \quad \sigma_{\text{eq}}^2 = \frac{\sigma_1^2}{2d_1} = \frac{\sigma_2^2}{2d_2} = \frac{\sigma_3^2}{2d_3}.$$

891 Therefore, a Gaussian invariant measure as defined in (1.7) can be found with equipartition of energy in each component, that
 892 is,

$$893 \quad (4.4) \quad p_{\text{eq}} = C^{-1} \exp\left(-\frac{1}{2} \sigma_{\text{eq}}^{-1} \mathbf{u} \cdot \mathbf{u}\right).$$

894 Above σ_{eq}^2 is the equilibrium variance in the Gaussian invariant distribution p_{eq} that controls the variability in each mode. To
 895 see this, we can substitute the invariant measure (4.4) back into the Fokker-Planck equation (4.2). It is a special case from the
 896 Theorem in [54], and detailed energy mechanism and stability for the triad system can be found in [50].

897 In the general case with deterministic external forcing and inhomogeneous structure, energy is injected into the modes and
 898 transferred to each other due to the nonlinear quadratic interaction through more complicated mechanism, thus strong nonlinear
 899 non-Gaussian statistics with energy cascade and internal instabilities can be generated.

900 **4.1.2. A link with quasi-geostrophic turbulence.** The triad model (4.1) is the building block of complex turbulent dy-
 901 namical systems since a three-dimensional Galerkin truncation of many complex turbulent dynamics possesses the energy-
 902 conserving nonlinearity as in (1.1). The random forcing together with the damping term represents the inhomogeneous effect
 903 of the interaction with other modes in a turbulent dynamical system that are not resolved in the three dimensional subspace.
 904 Stochastic triad models are qualitative models for a wide variety of turbulent phenomena regarding energy exchange and cas-
 905 cades and supply important intuition for many effects. They also provide elementary test models with subtle features for
 906 prediction, UQ, and state estimation [50, 64, 87].

907 As a simple illustration about the link to more complex turbulent systems, we can consider the quasi-geostrophic (QG)
 908 potential vorticity equation with no external forcing and dissipation

$$909 \quad \frac{\partial q}{\partial t} + \nabla^\perp \psi \cdot \nabla q = 0, \quad q = \nabla^2 \psi.$$

910 We have the barotropic triads of three barotropic components, $\psi_{\mathbf{k}}, \psi_{\mathbf{m}}, \psi_{\mathbf{n}}$, obeying the selecting rule $\mathbf{k} + \mathbf{m} + \mathbf{n} = \mathbf{0}$. Consider
 911 an initial condition in which only these three components of a particular triad are excited, then these three modes will only
 912 interact with each other while no other modes will get excited due to the particular triad relations as the system evolves in time.
 913 By projecting the above equation to the active triad modes, we get the dynamical equations for the selected modes

$$914 \quad (4.5) \quad \frac{d\psi_{\mathbf{k}}}{dt} + A_{\mathbf{k}\mathbf{m}\mathbf{n}} \psi_{\mathbf{m}} \psi_{\mathbf{n}} = 0, \quad \mathbf{k} + \mathbf{m} + \mathbf{n} = \mathbf{0},$$

915 where $A_{\mathbf{k}\mathbf{m}\mathbf{n}} = \frac{|\mathbf{n}|^2}{|\mathbf{k}|^2} \mathbf{m}^\perp \cdot \mathbf{n}$ is the triad interaction coefficient with the detailed symmetry $A_{\mathbf{k}\mathbf{m}\mathbf{n}} + A_{\mathbf{m}\mathbf{n}\mathbf{k}} + A_{\mathbf{n}\mathbf{k}\mathbf{m}} = 0$, showing the
 916 conservation of kinetic energy,

$$917 \quad \frac{d}{dt} \left(|\mathbf{k}|^2 |\psi_{\mathbf{k}}|^2 + |\mathbf{m}|^2 |\psi_{\mathbf{m}}|^2 + |\mathbf{n}|^2 |\psi_{\mathbf{n}}|^2 \right) = 0.$$

918 The typical forward and backward cascades of energy and enstrophy in turbulent flow are characterized by the triad interactions
 919 between the three models. Hence from the above discussion, in the two-dimensional QG turbulence, the nonlinear energy
 920 transfer is exactly governed by the barotropic triads the same as (4.1) in the nonlinear interaction part.

921 **4.2. Typical dynamical regimes in the triad system.** Though simple in appearance, the triad system (4.1) has represen-
 922 tative statistical features including energy cascade between modes and internal instabilities that can be created in this simple
 923 set-up. A fundamental factor in the triad system is the internal instabilities that make the mean unstable over various directions
 924 in phase space as is typical for anisotropic fully turbulent systems. Elementary intuition about energy transfer in such models
 925 can be gained by looking at the special situation with $L = D = F = \sigma \equiv 0$ so that there are only the nonlinear interactions in
 926 (4.1). We examine the linear stability of the fixed point, $\bar{\mathbf{u}} = (\bar{u}_1, 0, 0)^T$. Elementary calculations show that the perturbation δu_1
 927 satisfies $\frac{d\delta u_1}{dt} = 0$ while the perturbations $\delta u_2, \delta u_3$ satisfy the second-order equations

$$\frac{d^2}{dt^2} \delta u_2 = B_2 B_3 \bar{u}_1^2 \delta u_2, \quad \frac{d^2}{dt^2} \delta u_3 = B_2 B_3 \bar{u}_1^2 \delta u_3,$$

929 so that

930 there is instability with $B_2 B_3 > 0$ and
 931 (4.6) the energy of $\delta u_2, \delta u_3$ grows provided B_1 has
 932 the opposite sign of B_2 and B_3 with $B_1 + B_2 + B_3 = 0$.

933 The elementary analysis in (4.6) suggests that we can expect a flow or cascade of energy from u_1 to u_2 and u_3 where it is
 934 dissipated provided the interaction coefficient B_1 has the opposite sign from B_2 and B_3 .

935 Then energy cascades can be induced from the strongly forced unstable energetic mode to the stable less energetic modes
 936 with stronger damping effects. Particularly, we can generate distinct statistical features from Gaussian to highly skewed non-
 937 Gaussian PDFs in the following dynamical regimes:

- 938 • *Regime I: Equipartition of energy.* Set the equipartition of energy in stationary steady state. That is, $\frac{\sigma_1^2}{2d_1} = \frac{\sigma_2^2}{2d_2} = \frac{\sigma_3^2}{2d_3} =$
 939 σ_{eq}^2 . Gaussian distributions as in (4.4), $p_{\text{eq}} \sim \exp(-\frac{1}{2}\sigma_{\text{eq}}^{-2}\mathbf{u} \cdot \mathbf{u})$, will be reached under this set-up in the equilibrium
 940 state. The parameters are chosen as $d_1 = 0.2, d_2 = 0.1, d_3 = 0.1$, and $B_1 = 1, B_2 = -0.6, B_3 = -0.4$. Skew-symmetric
 941 interactions are added as $L_1 = 3, L_2 = 2, L_3 = -1$, and there is no deterministic forcing $F_1 = F_2 = F_3 = 0$ added for the
 942 unperturbed equilibrium;
- 943 • *Regime II: Nonlinear regime with forward energy cascade.* Consider the system with one weakly damped strongly
 944 forced mode and two other strongly damped weakly forced modes, that is, $d_1 = 1, d_2 = 2, d_3 = 2$, and $\sigma_1^2 = 10, \sigma_2^2 =$
 945 $0.01, \sigma_3^2 = 0.01$. The nonlinear coupling is taken as $B_1 = 2, B_2 = B_3 = -1$. The skew-symmetric interaction is set to
 946 be zero, $L_1 = L_2 = L_3 = 0$, and there is no deterministic forcing $F_1 = F_2 = F_3 = 0$ added in unperturbed equilibrium.
 947 In this case, the first mode is strongly forced by the random forcing while the other two modes are much less energetic.
 948 The values in the nonlinear coupling coefficients make sure that the additional energy injected in u_1 cascades to the
 949 other two less energetic modes u_2, u_3 , and then gets dissipated by the strong damping;
- 950 • *Regime III: Nonlinear regime with dual energy cascade.* Use the same damping and noise forcing parameters as in
 951 the energy cascade case, that is, $d_1 = 1, d_2 = 2, d_3 = 2$, and $\sigma_1^2 = 10, \sigma_2^2 = 0.01, \sigma_3^2 = 0.01$. The nonlinear coupling
 952 coefficients are also taken the same values as before, $B_1 = 2, B_2 = B_3 = -1$. Skew-symmetric interactions are added in
 953 this regime as $L_1 = 0.09, L_2 = 0.06, L_3 = -0.03$ to enhance the interactions between modes. Deterministic forcings are
 954 applied in modes u_2, u_3 as $F_1 = 0, F_2 = -1, F_3 = 1$. In addition to the forward energy cascade from u_1 to u_2, u_3 as in the
 955 previous case, the additional forcing introduces energy sources in modes u_2, u_3 and leads to backward energy cascade
 956 from modes u_2, u_3 back to u_1 . The linear skew-symmetric operator further alters the equilibrium energy structure in the
 957 system, resulting in skewed probability distribution functions in the steady state. This regime is especially interesting
 958 because strong internal instability can be generated here.

959 The first test case is the simplest but nevertheless representative with equipartition of energy. The higher-order moment effects
 960 are relatively small in this equipartition energy case, while the dynamics are dominantly Gaussian with zero cross-covariances
 961 as the system evolves in time. The second test case above is also relatively simple without the skew-symmetric interaction
 962 between modes and most energy will accumulate in the first dominant mode. Nevertheless important third-order interactions
 963 will take place in this case, and large errors will be introduced if the cross-covariances are ignored without care. In the third
 964 test case the non-zero forcing in the unperturbed system creates skewed equilibrium distributions with important third-order
 965 moments. Also the non-zero linear skew-symmetric interaction terms add extra emphasis on the cross-covariances. These
 966 induce stronger interactions and energy cascades between the triad modes.

967 **4.2.1. Numerical results about the the unperturbed triad system in equilibrium.** The true statistical features of the
 968 triad system in the above dynamical regimes are resolved through direct Monte-Carlo simulations. We run an ensemble of
 969 $N = 10000$ particles, which shall be enough for capturing the statistics in a three-dimensional phase space. Forward Euler
 970 scheme with small time step is used to integrate the system in time due to its simplicity. The stochastic forcing is simulated
 971 through the standard *Euler-Maruyama* scheme. The initial ensemble is chosen from a standard Gaussian random sampling.

972 For more details about the model statistics in these regimes, Figure 4.1 displays the probability distributions in these three
 973 test regimes. In mode u_1 , Gaussian (or quasi-Gaussian) distributions can always be observed in all the three regimes. Also
 974 the equipartition of energy structure can be observed in the first regime as predicted from analysis in the previous sections.
 975 Marginal PDFs are consistent with the Gaussian fits from theory and the joint 2-dimensional distributions are also in Gaussian
 976 structure. On the other hand, in the forward energy cascade case, fat tails can be observed in the marginal PDFs in both u_2, u_3
 977 as well as the star-shaped joint-distribution, showing the strong nonlinear effects in the modes. Note that the non-Gaussianity in
 978 u_2, u_3 can affect the final structure in the dominant mode u_1 despite its near-Gaussian marginal distribution. Furthermore, in the
 979 third test regime with dual energy cascades, besides the fat tails, skewed PDFs appear in both modes u_2, u_3 due to the non-zero
 980 deterministic forcing applied on them. The non-Gaussianity can be further confirmed by the joint distributions where strong
 981 skewness is shown. Also note that the mean states are not centered in zero in this case. These illustrate important third-order
 982 moments in this case for accurate predictions. Below we concentrate on this toughest regime.

983 **4.3. Reduced-order statistical models for the triad system.** In the development about reduced-order models, we focus
 984 on the accurate estimation about the first two moments, that is, the statistical mean state and covariance matrix. The major
 985 interest is to check the models' skill in capturing sensitivity in response to external perturbations besides equilibrium consistency
 986 in the models. Considering the relatively simple structure in the triad dynamics, we focus on the GC1 and GC2 formulation in
 987 (3.4) and (3.5) respectively in the following test cases and the parameter calibration strategy proposed in (3.8). In the model
 988 reduction procedure, we first consider the *fully resolved model*, where the 3-dimensional mean and 3×3 covariance matrix are
 989 resolved entirely; and then the *diagonal model*, where only the mean and diagonal variances are calculated explicitly. In the
 990 final step, a severely *reduced-order model* is introduced where only the variance in the principal mode is resolved.

991 **4.3.1. Reduced model formulation for the triad system (Model selection).** Applying exactly the modeling procedure in
 992 Section 3 to the triad system, the two-moment closure schemes replace the higher order moments with additional damping and
 993 excitement containing only first two order of moments. The first approach is to run the full system with mean and covariance
 994 matrix where both diagonal variances and off-diagonal components are resolved for $\bar{\mathbf{u}} \in \mathbb{R}^3$ and $R \in \mathbb{R}^{3 \times 3}$

$$995 \quad (4.7a) \quad \frac{d\bar{\mathbf{u}}}{dt} = L\bar{\mathbf{u}} + B(\bar{\mathbf{u}}, \bar{\mathbf{u}}) + \sum_{i,j} R_{ij} B(\mathbf{e}_i, \mathbf{e}_j) + \mathbf{F},$$

$$996 \quad (4.7b) \quad \frac{dR}{dt} = L_v(\bar{\mathbf{u}})R + RL_v^*(\bar{\mathbf{u}}) + Q_{GC} + Q_\sigma.$$

997 As a test for reduced order methods, we also check the models' skill by ignoring the cross-covariances. Especially for the triad
 998 system, the off-diagonal cross-covariances play a crucial role in the computation of mean and variance responses, while on
 999 the other hand, are difficult and expensive to estimate with accuracy. Then the unresolved covariances are approximated from
 1000 steady state information.

1001 • **Fully Resolved Model:** In this fully resolved model, the entire mean and covariance matrix are calculated through the
 1002 dynamical equations (4.7). The nonlinear flux term is approximated by additional damping and additional noise as in
 1003 (3.4) and (3.5)

$$1004 \quad Q_{GC} = -(D_M R + R D_M^*) + Q_M.$$

1005 • **Diagonal Model ignoring cross-covariances:** In this diagonal model, the cross-covariances between modes are ig-
 1006 nored to improve model efficiency. So the covariance matrix is replaced by diagonal matrix $R = \text{diag}(r_k)$. The
 1007 correction in Q_{GC} for higher-order moments calibration is kept the same as the full model for GC1 or GC2 respec-
 1008 tively. To correct the error due to the neglect of cross-covariances, we add the cross-covariance correction using only
 1009 steady state information

$$1010 \quad r_{ij} = \frac{\text{tr}R}{\text{tr}R_{\text{eq}}} \langle u'_i u'_j \rangle_{\text{eq}}.$$

1011 The diagonal model can be much more efficient compared with the fully resolved model with a model reduction from
 1012 $O(N^2)$ to $O(N)$.

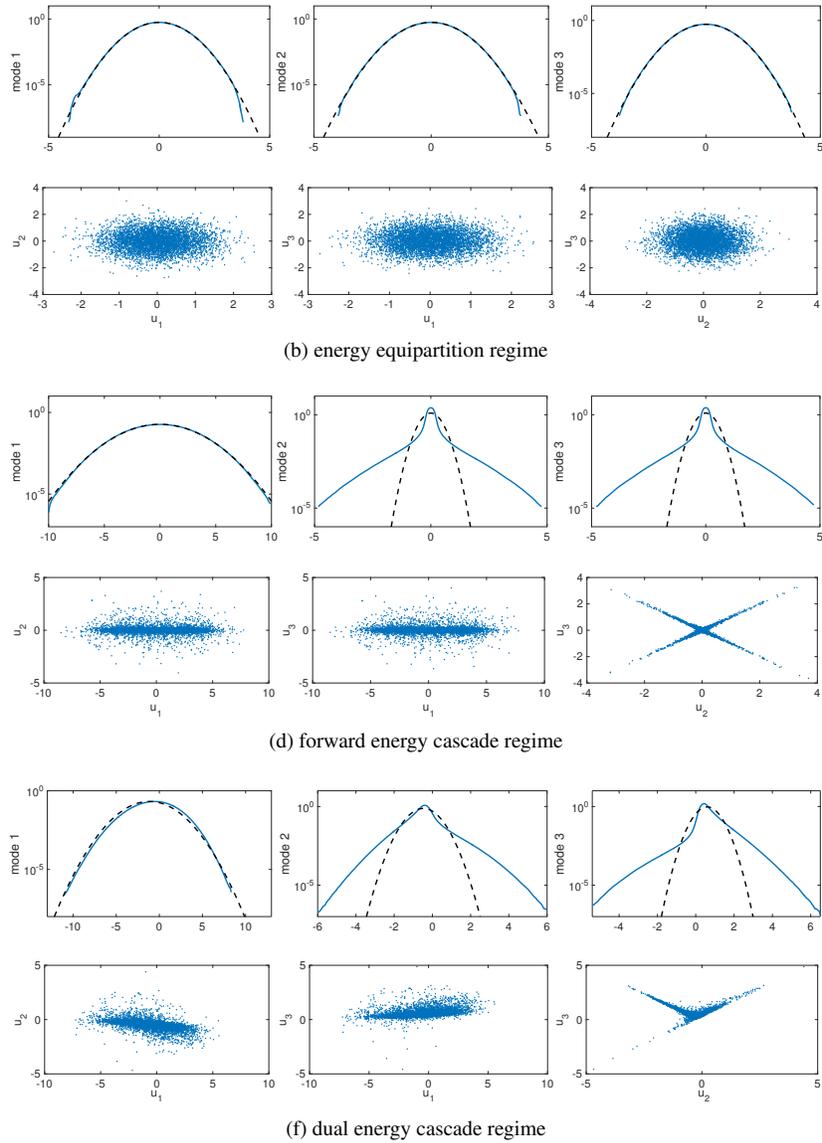


FIG. 4.1. Marginal PDFs in three modes u_1, u_2, u_3 and the joint distributions in scatter plots in three regimes with distinct statistics. Gaussian density functions with the same variance are shown in dashed black lines.

1013 Through construction using the unperturbed equilibrium information, climate consistency is guaranteed in both GC1 and GC2
 1014 model. In GC1, only constant damping and noise are added to approximate the unresolved higher order moments, while
 1015 these terms are further corrected with the total variance in GC2. The exponents in the scaling factors are designed to make
 1016 them consistent in dimension with the estimated third-order statistics. Actually, these scaling factor becomes quite crucial in
 1017 capturing model responses to perturbations and model performances in transient state.

1018 As a further reduction in the model we consider the two-moment closure methods in reduced order subspace. In this case,
 1019 we only resolve the variance in the first mode u_1 , and calculate the mean dynamics for all three modes. Then *the reduced order*

1020 *model* can be expressed as

$$1021 \quad (4.8a) \quad \frac{d\bar{\mathbf{u}}}{dt} = L\bar{\mathbf{u}} + B(\bar{\mathbf{u}}, \bar{\mathbf{u}}) + \sum_{i,j \in \Lambda} C_{ij}^{\text{red}} B(\mathbf{e}_i, \mathbf{e}_j) + \mathbf{F} + \mathbf{G},$$

$$1022 \quad (4.8b) \quad \frac{dC^{\text{red}}}{dt} = L_v^{\text{red}}(\bar{\mathbf{u}})C^{\text{red}} + C^{\text{red}}L_v^{\text{red}*}(\bar{\mathbf{u}}) + Q_{GC}^{\text{red}} + Q_{\sigma}^{\text{red}}.$$

1023 $\Lambda = \{1\}$ is the index set that includes the resolved modes. Especially with only the leading variance resolved, $C^{\text{red}} = r_1 = \langle u_1^2 \rangle$
 1024 is only the variance in mode u_1 . The construction about the reduced-order parameters can follow the general strategy in (3.12a)
 1025 and (3.12b) exactly.

1026 • **Reduced Model with only the principal variance resolved:** In the variance equation (4.8b), the formulation is similar
 1027 as before and we only need to constrain the covariance matrix C^{red} in the resolved subspace with the first mode
 1028 u_1 resolved. The additional damping and noise can be applied in the same way as the reduced-model correction
 1029 as discussed in (3.12). In the mean dynamics, the unresolved second-order moments in $R_{ij}B(\mathbf{e}_i, \mathbf{e}_j)$ are corrected
 1030 following (3.13)

$$1031 \quad G = \frac{\text{tr}R}{\text{tr}R_{\text{eq}}} \sum_{i,j \in \Lambda^c} R_{ij}B(\mathbf{e}_i, \mathbf{e}_j).$$

1032 The total variance for model sensitivity correction can also be achieved through the approximated statistical energy
 1033 dynamics as in (3.16)

$$1034 \quad \frac{dE^{\text{est}}}{dt} = -d_{\text{eff}}E^{\text{est}} + \sum_{i=1}^3 \left(F_i \bar{u}_i + \frac{1}{2} \sigma_i^2 \right).$$

1035 Note that we have calculated the mean in each mode, then the total variance can be calculated through the total
 1036 statistical energy $\text{tr}R = 2E - \sum_{i=1}^3 \bar{u}_i^2$. One additional difficulty in the inhomogeneous case is that different damping
 1037 rates d_j are applied to different modes. So we introduce the effective damping rate d_{eff} through the statistical steady
 1038 state information as

$$1039 \quad d_{\text{eff}} = \frac{\sum_{i=1}^3 (F_i \bar{u}_{i,\text{eq}} + \frac{1}{2} \sigma_i^2)}{E_{\text{eq}}}.$$

1040 We also list the explicit statistical mean and variance dynamical equations in Appendix B, as well as the explicit statistical
 1041 energy dynamics to achieve the total variance of the system when only the leading mode is resolved.

1042 **4.3.2. Model consistency and tuning parameters in the training phase (*Model calibration*).** In the model calibration
 1043 step, we check the imperfect models' consistency with climatology when no perturbation is applied. As noticed before, the
 1044 imperfect models' skill in capturing the right statistics in transient state is crucial for the convergence to the right fixed point in
 1045 the energy equation. As an example, Figure 4.2 compares the model performances in regimes with dual energy cascade. It can
 1046 be observed that the statistical steady state mean and variances are recovered with accuracy due to the climate consistent choice
 1047 of parameters through (3.4) and (3.5), while GC1 lacks the ability to accurately capture the transient state in the beginning due
 1048 to its lack of sensitivity.

1049 We illustrate the tuning process for optimal model parameters in the training phase in Figure 4.3. Again we use the dual
 1050 energy cascade regime as an example. Since both the deterministic and stochastic forcing might be perturbed in the external
 1051 forcing, we consider a kicked response in the initial value according to the perturbation form described in (4.9) and (4.10). In
 1052 the first row, the information errors with changing model parameter values are shown in total relative entropy from (2.17) as
 1053 well as the errors in signal and dispersion component. Note that the errors in GC2 model results stay uniformly small in the
 1054 entire parameter regime, showing the robustness of the method; on the other hand, GC1 model displays larger information error
 1055 no matter how well we tune the model parameter. This illustrates the inherent information barrier in the closure schemes if
 1056 we do not consider proper statistical energy scaling in the model damping and noise terms. In the second row, we show the
 1057 approximation about the linear response operators in the mean and variance using optimal parameters. The transient structures
 1058 can be captured with accuracy in GC2 model, while GC1 lacks the skill due to the insufficient calibration in the higher-order
 1059 nonlinear flux approximation.

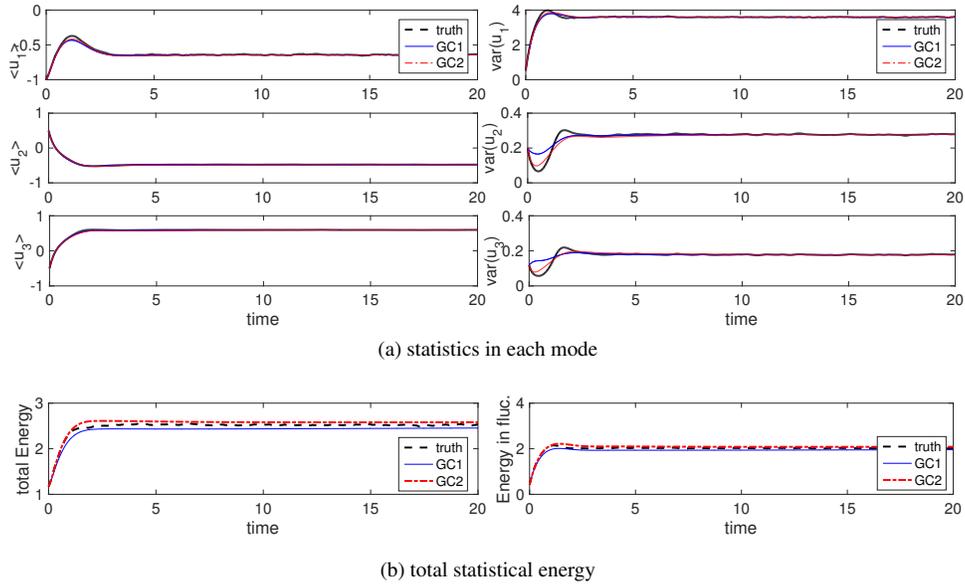


FIG. 4.2. Imperfect model performances in convergence to statistical equilibrium state in dual energy cascade regime. Both GC1 (with constant damping and noise) and GC2 (with correction from total variance) can recover the steady state mean and variance but GC1 lacks the ability to accurately capture the transient state in the beginning due to its lack of sensitivity.

1060 **4.3.3. Model prediction skill to perturbations (*Model prediction*).** Periodic perturbations added in both deterministic
 1061 and random forcing are representative in checking the imperfect models' prediction skill in response to perturbations. The
 1062 perturbations are introduced in the following forms:

- 1063 • *periodic forcing perturbations in the deterministic forcing*: The deterministic forcing perturbation is introduced by a
 1064 periodic addition to the mean forcing $F = \bar{F} + \delta F$. As one typical test case, we add periodic forcing perturbations δF
 1065 to each mode, that is,

$$(4.9) \quad \delta F_i = A_i \sin(\omega t),$$

1066 with ω taking the value $\pi/4$, and A_i measures the perturbation amplitude in each mode.

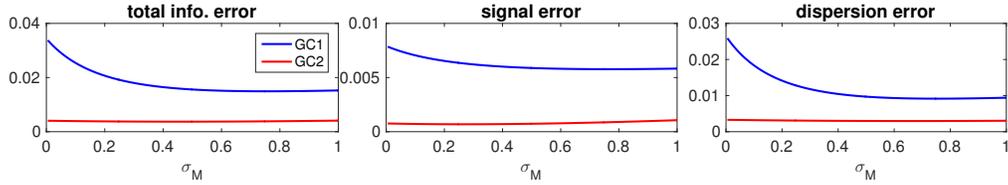
- 1067 • *periodic perturbation in the stochastic random forcing*: Also to add stronger time-dependent effect to the variances,
 1068 we add periodic random forcing to the system by setting
 1069

$$(4.10) \quad \sigma_j = \bar{\sigma}_j + \delta f^2(t) (\sigma_{T_j} - \bar{\sigma}_j),$$

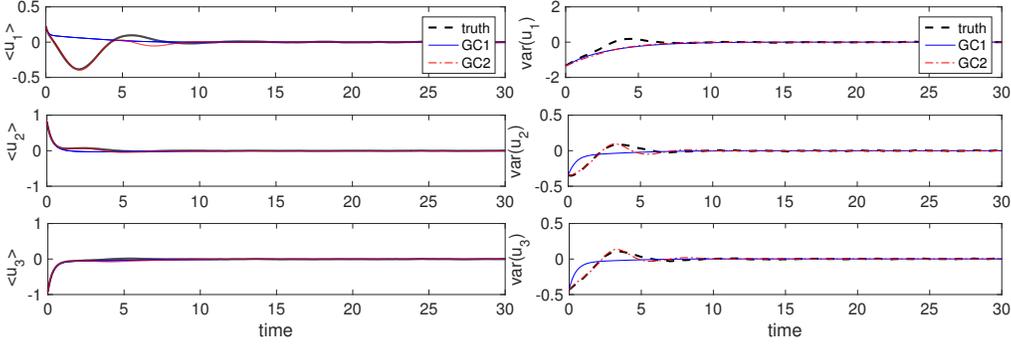
1070 with $\delta f(t) = \sin(\omega t)$ also set to be periodic. Here $\bar{\sigma}_j$ is the mean stochastic forcing amplitude in the unperturbed case
 1071 with $\sigma_{T_j} > \bar{\sigma}_j$ to add perturbations to this unperturbed mean.

1072 We check the model performances in predicting response to the periodic perturbations in both deterministic and stochastic
 1073 components. Here we display the imperfect model prediction skill in the toughest regime with dual energy cascade. Thus
 1074 strong nonlinear coupling is present between the modes. In Figure 4.4, the *fully resolved model* with mean and 3×3 covariance
 1075 matrix is applied to the typical regime in GC1 and GC2 closure methods. In this regime with skewed distributions, higher order
 1076 moments become crucial and need more detailed calibration. As we can see from the results, the deficiency of GC1 method
 1077 appears in this regime due to the inaccurate approximation about the third-order moments. Large deviation takes place in the
 1078 skewed modes u_2, u_3 due to the errors from third-order energy transfer. In contrast, GC2 model maintains the high skill in
 1079 predicting the mean and variances with the more careful calibration about the nonlinear flux through the scaling factor using
 1080 total statistical energy.
 1081

1082 In Figure 4.5, we check the *diagonal models* with only mean and variances in each mode resolved and ignoring the off-
 1083 diagonal cross-covariances. Like the previous case of full model, GC1 loses the skill in predicting the responses in u_2, u_3 due to
 1084 the lack of information in the third-order interactions. The errors in the beginning transient regime drive the statistical equation
 1085 to the wrong state or even blowing up. GC2 keeps the skill in capturing the response structures of both the mean and variances.
 1086 And again most of the error takes place in the variance estimations.



(a) information error with changing parameter values



(b) linear responses with optimal parameter

FIG. 4.3. Illustration about the training phase for tuning optimal imperfect model parameters for GC1 and GC2. The first row is the information errors with changing values of the tuning parameter σ_M . The second row displays the approximation for the linear response operators in GC1 and GC2 using optimal parameters from the tuning process above.

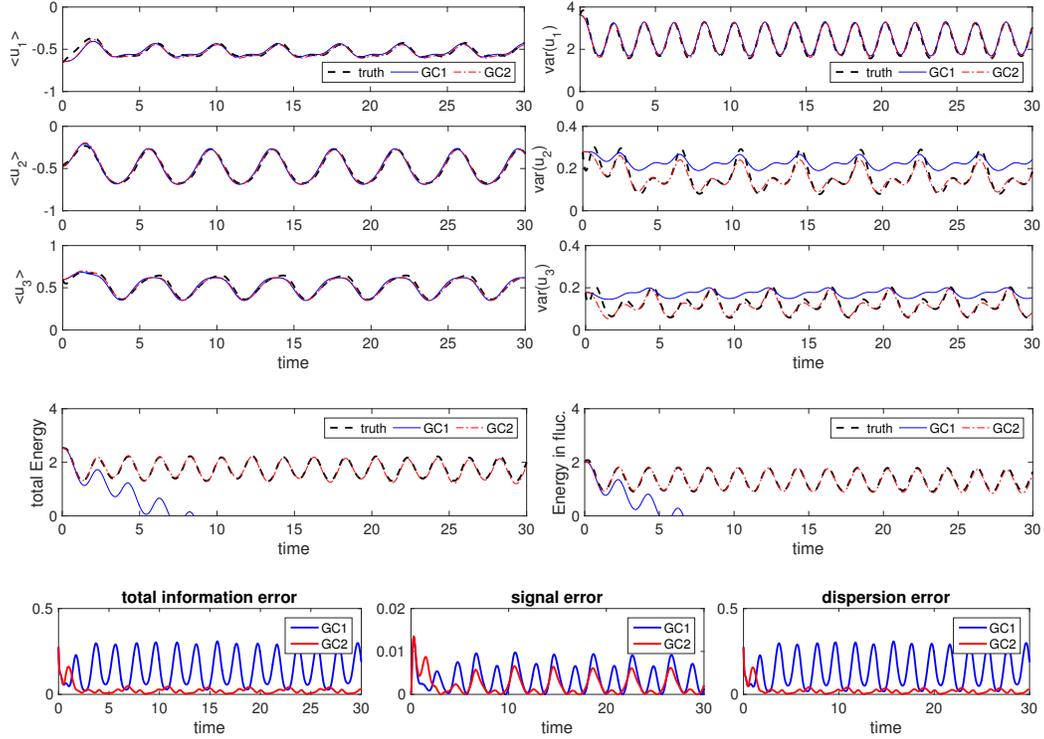


FIG. 4.4. Full GC1 and GC2 model with optimal parameter in approximating responses to periodic external forcing in regimes with dual cascade. The first three rows show the GC1 and GC2 predictions for the statistical mean and variances in each mode together with the truth from MC simulations in thick black lines. The total energy E from the energy equation and the fluctuation energy $E^f = E - \frac{1}{2}\bar{\mathbf{u}}^2$ are compared in the following row. The last line shows the total information error in the imperfect models together with the signal and dispersion components.

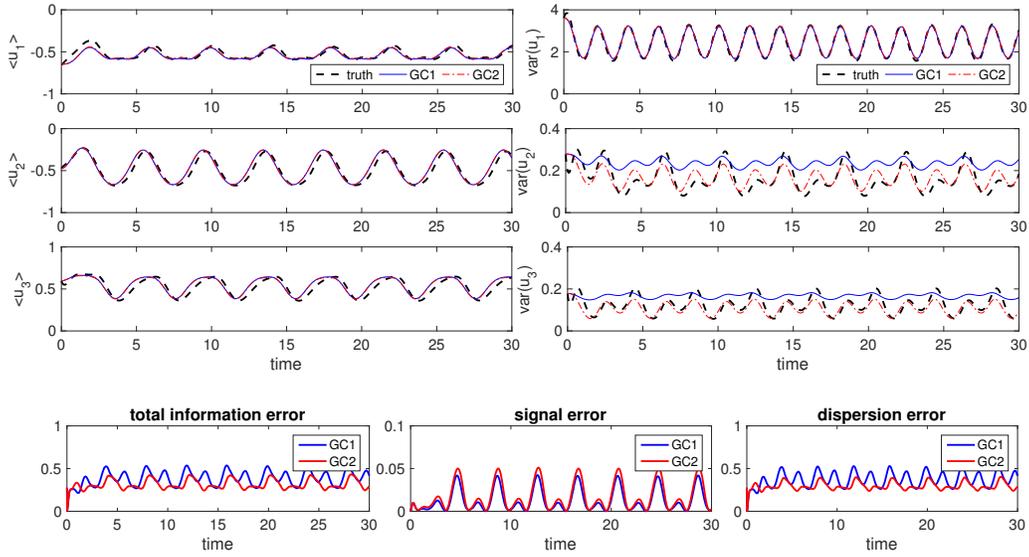


FIG. 4.5. Diagonal GC1 and GC2 model with optimal parameter in approximating responses to periodic external forcing in regimes with dual cascade. The first three rows show the GC1 and GC2 predictions for the statistical mean and variances in each mode together with the truth from MC simulations in thick black lines. The last line shows the total information error in the imperfect models together with the signal and dispersion components.

1087 Testing the feasibility of the *reduced-order strategies* described previous is important for further applications to general
 1088 high dimensional turbulent systems because the triad interactions always represent the nonlinear interactions in different scales
 1089 which usually are ignored in realistic modeling. The same dynamical regime is tested here for the reduced-order models. In this
 1090 case, only the variance in the first mode u_1 is resolved explicitly. Especially for this regime with strong dual energy cascade,
 1091 as we have seen in the true statistical dynamics, strong coupling exists between the high energy mode u_1 and the less energetic
 1092 modes, u_2, u_3 , in both third-order moments and second-order cross-covariances. Thus this becomes a challenging situation
 1093 for the reduced-order models for capturing the responses with accuracy. The model prediction results are shown in Figure 4.6.
 1094 With forward and backward energy cascade, strong nonlinear high-order interactions become crucial here. GC1 loses its skill
 1095 in this case and ends with large errors especially in the mean state u_1 . This is no surprise considering the strong perturbed
 1096 deviation from the equilibrium state in this regime due to the nonlinear energy transfer while GC1 model only uses unperturbed
 1097 equilibrium information. On the other hand, GC2 keeps its skill and can capture the responses in both mean and variance with
 1098 only a single low-order mode resolved.

1099 **4.3.4. Additional results for the triad model with equipartition of energy.** In the final part of this section, we illustrate
 1100 the imperfect model prediction skills in the equipartition of energy regime with Gaussian statistics. In this case, the three
 1101 modes (u_1, u_2, u_3) in the triad model possess same amount of energy in statistical equilibrium state as in (4.3) and (4.4). In
 1102 Figure 4.7 and 4.8, we show the prediction results to the periodic perturbations from the *diagonal model* with off-diagonal
 1103 cross-covariances neglected, and from the *reduced model* with only the variance in the first mode u_1 resolved. Both GC1 and
 1104 GC2 can capture the mean states in all three modes quite accurately, and GC1 and GC2 results have little difference. This is due
 1105 to the relatively simple energy mechanism in this regime with same amount of energy in each mode. On the other hand, in the
 1106 prediction of variances in the diagonal model larger errors appear especially in the modes u_2 and u_3 . This shows the important
 1107 effects of the off-diagonal covariances in predicting second order moments in this equipartition energy regime. The reduced
 1108 model gets good predictions for the variance in the first resolved mode u_1 . With the more expensive full model, the prediction
 1109 for the variances in all three modes will become accurate with larger computational cost.

1110 **5. Reduced-Order Statistical Models Applied to Two-Layer Baroclinic Turbulence.** In this section, we validate the
 1111 performance of the reduced-order models by the more complicated two-layer quasi-geostrophic system with baroclinic instabil-
 1112 ity. It is shown that the baroclinic model is capable in capturing the essential physics of the relevant internal variability despite
 1113 its relatively simple dynamical structure. Two dynamical regimes with typical statistical features are representative in many
 1114 applications [90, 4, 79]. The first one is the fully turbulent flow with homogeneous statistics as a result of internal baroclinic

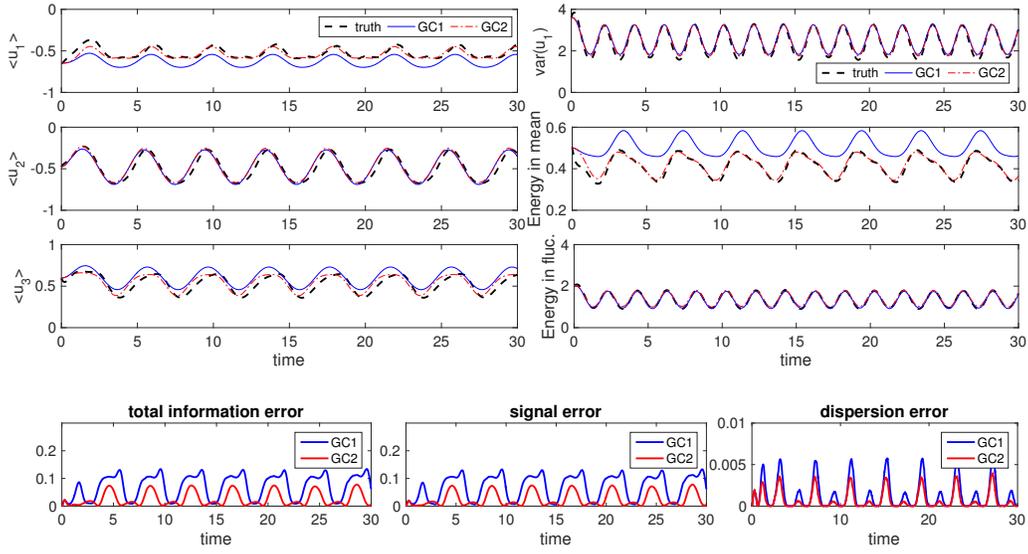


FIG. 4.6. Reduced-order GC1 and GC2 model with the variance in first mode resolved in approximating responses to periodic external forcing in dual energy cascade regime. The first three rows show the GC1 and GC2 predictions for the statistical mean and variances in each mode together with the truth from MC simulations in thick black lines. The last line shows the total information error in the imperfect models together with the signal and dispersion components.

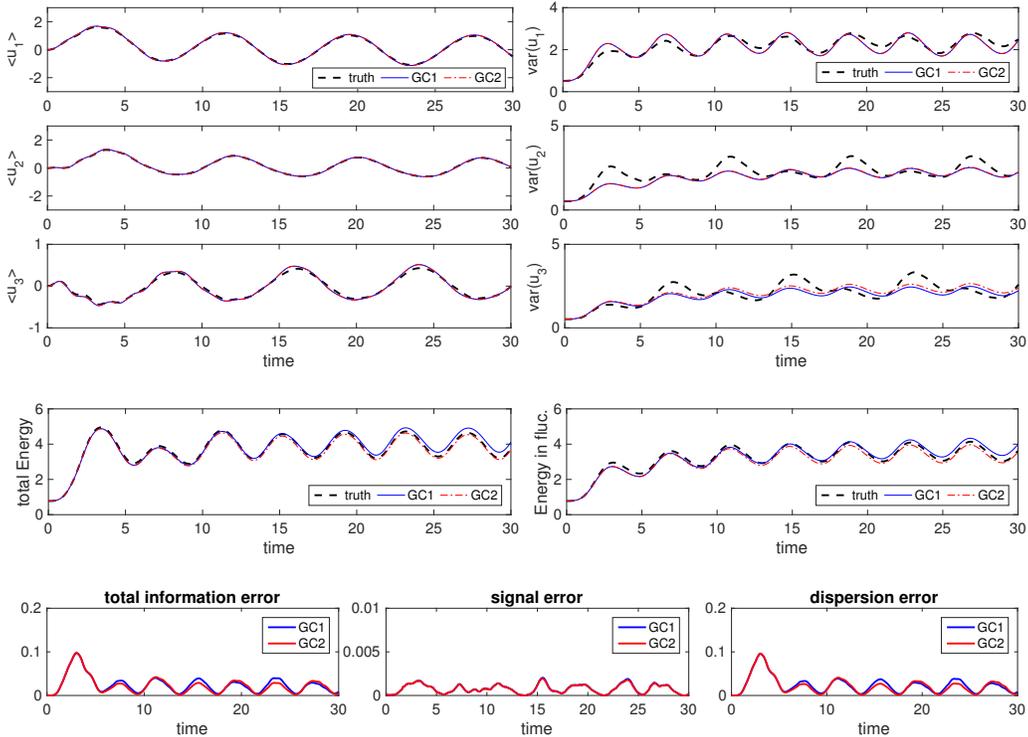


FIG. 4.7. Diagonal GC1 and GC2 model with optimal parameter in approximating responses to periodic external forcing in regimes with equipartition of energy. The first three rows show the GC1 and GC2 predictions for the statistical mean and variances in each mode together with the truth from MC simulations in thick black lines. The total energy E from the energy equation and the fluctuation energy $E' = E - \frac{1}{2}\bar{\mathbf{u}}^2$ are compared in the following row. The last line shows the total information error in the imperfect models together with the signal and dispersion components.

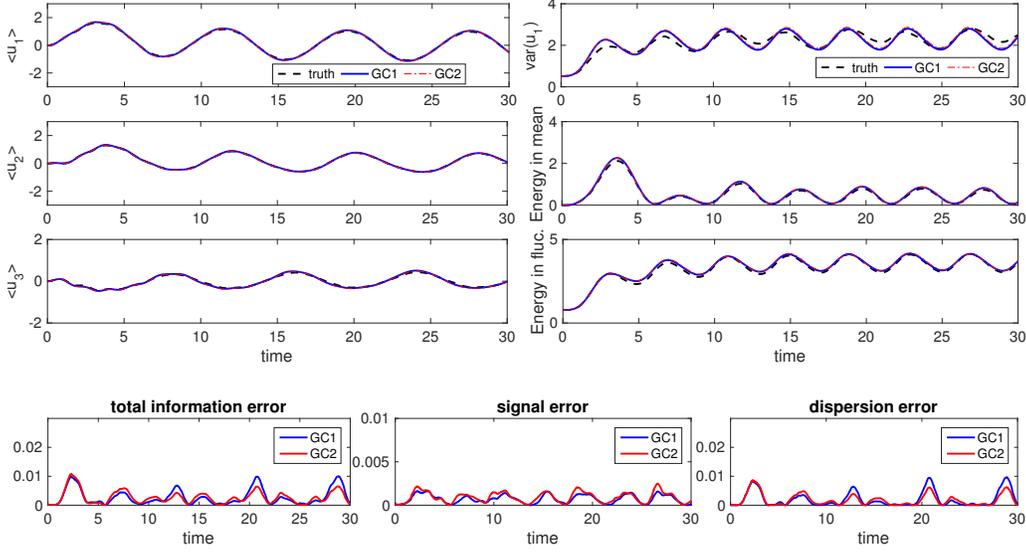


FIG. 4.8. Reduced-order GC1 and GC2 model with the variance in first mode resolved in approximating responses to periodic external forcing in equipartition energy regime.

1115 instability corresponding to the high-latitude ocean and atmosphere; the second one is the anisotropic flow field with strong
 1116 meandering zonal jets as in the low/mid-latitude regime. Detailed discussions and comparisons of the construction and test
 1117 about the reduced-order models in various regimes with representative zonal jets and vortices can be found in [79].

1118 The governing two-layer quasi-geostrophic (QG) equations in a barotropic-baroclinic mode formulation for potential vor-
 1119 ticity anomalies (q_ψ, q_τ) with periodic boundary condition in both x, y directions are [84, 92]

$$\begin{aligned}
 1120 \quad (5.1) \quad & \frac{\partial q_\psi}{\partial t} + J(\psi, q_\psi) + J(\tau, q_\tau) + \beta \frac{\partial \psi}{\partial x} + U \frac{\partial}{\partial x} \Delta \tau = -\frac{\kappa}{2} \Delta(\psi - \tau) - \nu \Delta^s q_\psi + \mathcal{F}_\psi(\mathbf{x}, t), \\
 & \frac{\partial q_\tau}{\partial t} + J(\psi, q_\tau) + J(\tau, q_\psi) + \beta \frac{\partial \tau}{\partial x} + U \frac{\partial}{\partial x} (\Delta \psi + k_d^2 \psi) = \frac{\kappa}{2} \Delta(\psi - \tau) - \nu \Delta^s q_\tau + \mathcal{F}_\tau(\mathbf{x}, t).
 \end{aligned}$$

1121 Above $q_\psi = \Delta \psi$, $q_\tau = \Delta \tau - k_d^2 \tau$ are the *disturbance* potential vorticities in barotropic and baroclinic mode respectively, while
 1122 ψ, τ are the corresponding *disturbance* barotropic and baroclinic stream functions. The barotropic mode ψ can be viewed as the
 1123 vertically averaged effect from the flow, and the baroclinic mode τ is usually related with the thermal effect in heat transport.
 1124 Besides, $J(A, B) = A_x B_y - A_y B_x$ represents the Jacobian operator. $k_d = \sqrt{8}/L_d = (2f_0/NH)^2$ is the baroclinic deformation
 1125 wavenumber corresponding to the Rossby radius of deformation L_d . A large-scale vertical shear $(U, -U)$ with the same strength
 1126 and opposite directions is assumed in the background to induce baroclinic instability. In the dissipation operators on the right
 1127 hand sides of the equations (5.1), besides the hyperviscosity, $\nu \Delta^s q_i$, we only use Ekman friction, $\kappa \Delta \psi_2$, with strength κ on the
 1128 lower layer of the flow.

1129 **5.1. Representative dynamical regimes for the two-layer baroclinic turbulence.** The two-layer quasi-geostrophic sys-
 1130 tem can display various dynamical regimes with distinct statistical features as the parameters are changed. Parameters for high
 1131 and low/mid latitude dynamical regimes are shown in Table 1. In numerical simulations, the true statistics are calculated by a
 1132 pseudo-spectra code by resolving the two-layer equations (5.1) with 128 spectral modes zonally and meridionally, correspond-
 1133 ing to $256 \times 256 \times 2$ grid points in total. In the reduced-order methods, only the large-scale modes $|\mathbf{k}| \leq 10$ are resolved, which
 1134 is about 0.15% of the full model resolution.

1135 In the simulations for the unperturbed system in ocean and atmosphere regimes, Figure 5.1 displays the two-layer flow
 1136 structure in *high-latitude ocean regime*. The first row is the snapshots of the barotropic and baroclinic vorticity. Homogeneous
 1137 structure can be observed in both cases while larger scale structures appear in the baroclinic mode. It is important to notice the
 1138 strong correlation in the coherent structures in the barotropic and baroclinic field, illustrating the strong energy transfer between
 1139 the two modes. The following part shows time-series of energy in barotropic and baroclinic mode, $-f \psi q_\psi$, $-f \tau q_\tau$, as well as

TABLE 1

Model parameters for ocean and atmosphere dynamical regimes in high and low/mid latitude. N is the model resolution, β, k_d are the rotation parameter and the deformation frequency, U is the background mean shear flow, κ is the Ekman drag in the bottom layer. The last three columns display the unstable waveband from linear analysis. (k_{\min}, k_{\max}) shows the range of unstable wavenumbers; σ_{\max} is the largest linear growth rate; and $(k_x, k_y)_{\max}$ is the position of the mode with maximum growth rate.

regime	N	β	k_d	U	κ	(k_{\min}, k_{\max})	σ_{\max}	$(k_x, k_y)_{\max}$
ocean regime, high lat.	256	10	10	1	9	(2.25, 14.61)	0.411	(4, 0)
atmosphere regime, high lat.	256	1	4	0.2	0.2	(1.58, 6.78)	0.099	(2, 0)

(a) high-latitude regime

regime	N	β	k_d	U	κ	(k_{\min}, k_{\max})	σ_{\max}	$(k_x, k_y)_{\max}$
ocean regime, low/mid lat.	256	100	10	1	1	(7.14, 15.63)	0.104	(2, 8)
atmosphere regime, low/mid lat.	256	2.5	4	0.2	0.05	(2.51, 7.06)	0.053	(3, 0)

(b) low/mid-latitude regime

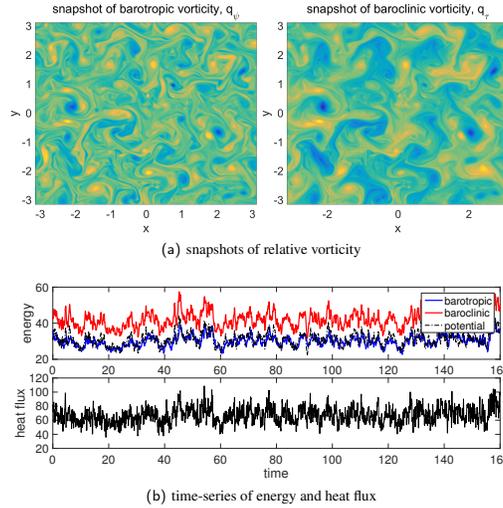


FIG. 5.1. Snapshots of high-latitude ocean regime barotropic and baroclinic vorticity in unperturbed system with no external forcing terms. Time-series of energy in barotropic and baroclinic modes, as well as potential energy, are also compared with the heat flux.

1140 the potential energy, $f k_d^2 \tau^2$, compared with the meridional heat flux, $k_d^2 U f \psi_x \tau$. In Figure 5.2 the results for the two-layer flow
 1141 in *high-latitude atmosphere regime* are compared. One important feature here is the flow field alternating between blocked and
 1142 unblocked regimes. In the stream functions, it can be observed that in the blocked regime, zonal flow is blocked and the field is
 1143 restricted at separated regimes, while in the unblocked regime strong zonal flow can be observed. Strong meridional heat flux
 1144 can be observed in the blocked regime while the flow is in state with lower energy and low heat transfer in the zonal unblocked
 1145 regime.

1146 In mid/low latitude regimes, both the ocean and atmosphere are distinctly inhomogeneous on large scales. The existence of
 1147 large-amplitude meandering zonal jets in these regimes suggests the regional metastable equilibria, while the large-scale forced
 1148 perturbations may lead to regular or irregular fluctuations in some extent. The jet structures are illustrated in more detail in
 1149 Figure 5.3 for the time-series of the zonally average mean flow, $u = -\partial_y \psi$. In this low/mid latitude case, especially for the
 1150 ocean regime, due to the strong zonal jets in wavenumber $k_y = 6$, zonal modes with $k_x = 5, 6$ become active due to the nonlinear
 1151 interactions.

1152 The general steady state statistical structures in the spectral field are shown in Figure 5.4. As implied from the homogeneous

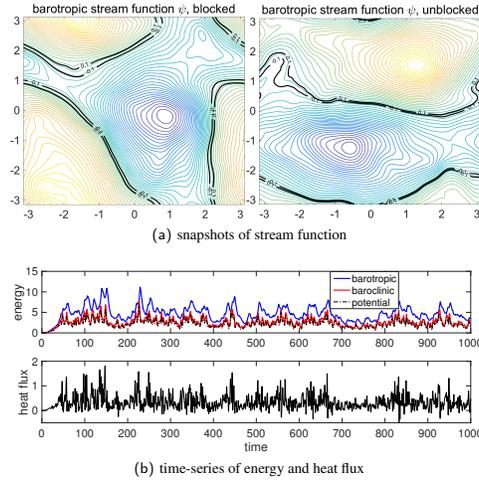


FIG. 5.2. Snapshots of high-latitude atmosphere regime barotropic and baroclinic stream function in unperturbed system with no external forcing terms. Time-series of energy in barotropic and baroclinic modes, as well as potential energy, are also compared with the heat flux.

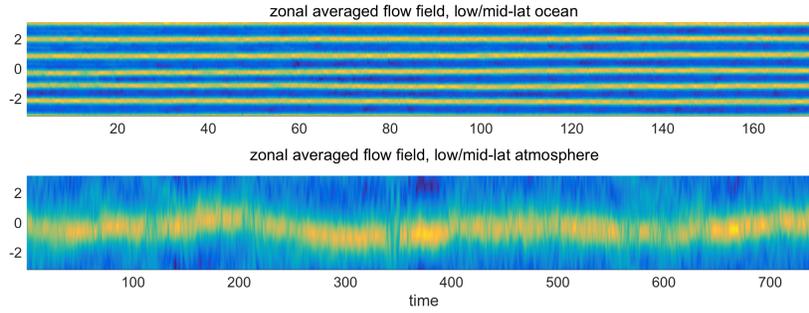


FIG. 5.3. Time-series of zonal averaged mean flow with jets in low/mid-latitude atmosphere and ocean regimes.

1153 statistics in high-latitude, the mean states stay in small values within fluctuation errors in both ocean and atmosphere regimes.
 1154 From the energy spectra, one observation is that the potential energy is dominant in large scales in the baroclinic modes, and
 1155 the kinetic baroclinic energy becomes more important in small scales. For both regimes, we observe wide and energetic spectra
 1156 that exchange energy between different scales, which indicates strong forward and backward energy cascades along the entire
 1157 spectral modes.

1158 **5.2. Predicting quasi-geostrophic statistical responses in reduced-order models.** The quasi-geostrophic response to
 1159 both stochastic and deterministic perturbations is an important subject in understanding the earth's atmospheric and oceanic
 1160 interactions [2, 47]. The same strategy developed in Section 3.3 and applied in the triad system for the first mode in Section 4.3
 1161 can be directly generalized to the statistical modeling of the two-layer QG system here.

1162 **5.2.1. Statistical formulation about the two-layer baroclinic equations.** We formulate the two-layer QG system with
 1163 Galerkin truncation to finite number of spectral modes. In model simulations, consider a set of rescaled normalized quantities
 1164 with a high wavenumber truncation N under standard Fourier basis $\mathbf{e}_{\mathbf{k}} = \exp(i\mathbf{k} \cdot \mathbf{x})$ due to the periodic boundary condition, so
 1165 that

$$1166 \quad (5.2) \quad \begin{aligned} p_{\psi, \mathbf{k}} &= q_{\psi, \mathbf{k}} / |\mathbf{k}| = -|\mathbf{k}| \psi_{\mathbf{k}}, \\ p_{\tau, \mathbf{k}} &= q_{\tau, \mathbf{k}} / \sqrt{|\mathbf{k}|^2 + k_d^2} = -\sqrt{|\mathbf{k}|^2 + k_d^2} \tau_{\mathbf{k}}. \end{aligned}$$

1167 The introduction of this new set of quantities (5.2) offers convenience that the energy inner-product becomes the standard
 1168 Euclidean form. Under the above settings, the rescaled set of equations of (5.1) can be summarized in the abstract form in the

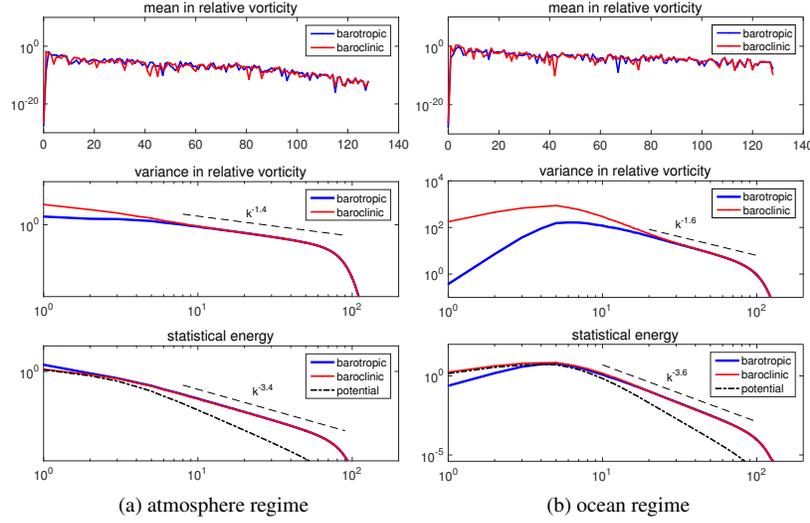


FIG. 5.4. Radial-averaged spectra in mean and second-order moments in both atmosphere and ocean high-latitude regimes. The first row compares the statistical mean states (in logarithmic coordinate). The following two rows show the variances, and statistical energy, in barotropic and baroclinic modes, as well as the potential energy.

1169 truncated subspace $|\mathbf{k}| \leq N$ as in (1.1)

$$1170 \quad (5.3) \quad \frac{d\mathbf{p}_{\mathbf{k}}}{dt} = B_{\mathbf{k}}(\mathbf{p}_{\mathbf{k}}, \mathbf{p}_{\mathbf{k}}) + (\mathcal{L}_{\mathbf{k}} - \mathcal{D}_{\mathbf{k}})\mathbf{p}_{\mathbf{k}} + \mathcal{F}_{\mathbf{k}}, \quad \mathbf{p}_{\mathbf{k}} = (p_{\psi, \mathbf{k}}, p_{\tau, \mathbf{k}})^T,$$

1171 where the linear operators are decomposed into the non-symmetric part $\mathcal{L}_{\mathbf{k}}$ involving β -effect and vertical shear flow U and
 1172 dissipation part $\mathcal{D}_{\mathbf{k}}$, together with the forcing $\mathcal{F}_{\mathbf{k}}$ combining deterministic component and stochastic component compared with
 1173 (5.1). Most importantly, $B(\mathbf{p}, \mathbf{p})$ is the nonlinear interactions that conserve both energy and enstrophy.

1174 The same reduced-order modeling strategy then can be applied to the two-layer model following the algorithm in Section
 1175 3.4. Therefore the true dynamical equations for the statistical moment $R_{\mathbf{k}} = \langle \mathbf{p}_{\mathbf{k}}^* \mathbf{p}_{\mathbf{k}} \rangle$ in the form of a 2×2 matrix containing
 1176 barotropic and baroclinic mode in same wavenumber \mathbf{k} become

$$1177 \quad (5.4) \quad \frac{dR_{\mathbf{k}}}{dt} = (\mathcal{L}_{\mathbf{k}} - \mathcal{D}_{\mathbf{k}})R_{\mathbf{k}} + Q_{F, \mathbf{k}} + Q_{\sigma, \mathbf{k}} + c.c., \quad |\mathbf{k}| \leq N,$$

1178 where $c.c.$ represent the complex completion for the conjugate parts. On the right hand side of the equation, $\mathcal{L}_{\mathbf{k}}$, $\mathcal{D}_{\mathbf{k}}$ represent
 1179 the linear interactions between modes, including β -effect through the rotation of the earth, the effects from the mean shear
 1180 flow U , as well as the dissipations from Ekman drag and hyperviscosity. $Q_{\sigma, \mathbf{k}}$ is the external forcing perturbations represented
 1181 by hypothetical stirring and heating forces. Importantly, the nonlinear flux $Q_{F, \mathbf{k}}$ represents the nonlinear interactions between
 1182 different wavenumbers due to the advection term. Third-order moments with triad modes $\mathbf{m} + \mathbf{n} = \mathbf{k}$ enter the first two order
 1183 moments dynamics representing the nonlinear energy transfer between small and large scales. The nonlinear energy exchange
 1184 mechanism is crucial in the energy budget, and the conservation property is satisfied due to the triad symmetry as $\sum_{\mathbf{k}} \text{tr} Q_{F, \mathbf{k}} = 0$.

1185 **5.2.2. Reduced-order model predictions for responses in various dynamical regimes.** In constructing the reduced-
 1186 order models, the same strategy is applied to the crucial but expensive nonlinear flux term Q_F as in (3.8)

$$1187 \quad Q_{M, \mathbf{k}} = Q_{M, \mathbf{k}}^- + Q_{M, \mathbf{k}}^+ = f_1(E) [-(N_{M, \mathbf{k}, \text{eq}} + d_M)R_{M, \mathbf{k}}] + f_2(E) [Q_{F, \mathbf{k}, \text{eq}}^+ + \sigma_{M, \mathbf{k}}^2].$$

1188 Both equilibrium higher-order statistics and additional corrections are combined, and statistical energy equation is important to
 1189 provide the scaling factor for optimal consistency and sensitivity. See [79] for more details.

1190 In checking the model sensitivity in the homogeneous high-latitude regimes, we introduce the forcing perturbation by
 1191 changing the background jet strength U . Note that the deterministic perturbation about zonal mean flow advection forms a

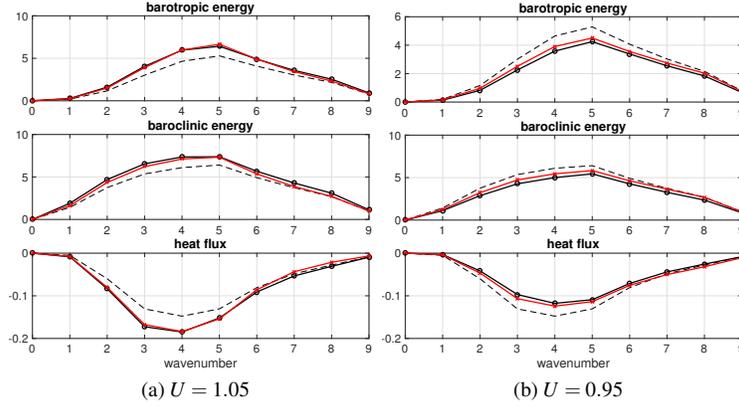


FIG. 5.5. Reduced-order model predictions to mean shear flow perturbation $\delta U = \pm 0.05$ in high-latitude ocean regime. The reduced-order model predictions for the spectra are compared with the truth. Black lines with circles show the perturbed model responses in the barotropic energy, baroclinic energy, and heat flux. The dashed black lines are the unperturbed statistics, and the reduced order model predictions are in red lines.

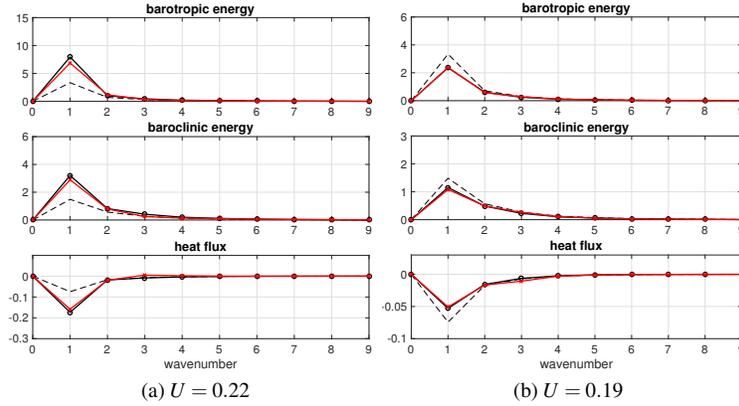


FIG. 5.6. Reduced-order model predictions to mean shear flow perturbation in high-latitude atmosphere regime. The reduced-order model predictions for the spectra are compared with the truth. Black lines with circles show the perturbed model responses in barotropic energy, baroclinic energy, and heat flux. The dashed black lines are the unperturbed statistics. And the reduced order model predictions are in red lines.

1192 difficult test case because the forcing is applied along all wavenumbers with stronger mean-fluctuation interactions involved.
 1193 On the other hand, for the reduced order methods, only the perturbations at the limited resolved modes are quantified. This
 1194 gives the inherent difficulty for applying the reduced order models to this kind of perturbations since we have no knowledge
 1195 of the unresolved modes where large amount of energy is contained. The results with mean flow perturbations $\delta U = \pm 0.05$
 1196 in the ocean regime and perturbations $\delta U = 0.02, -0.01$ in the atmosphere regime are shown in Figure 5.5 and 5.6 separately.
 1197 The perturbation accounts for about 5%-10% of the original shear strength U , and the corresponding responses in both energy
 1198 and heat flux spectra are large due to this global perturbation at every wavenumber and nonlinear energy cascade. In the ocean
 1199 regime, a wide waveband of modes $|\mathbf{k}| = 3, 4, 5, 6$ becomes sensitive to the perturbations; while in the atmosphere regime, the
 1200 first dominant mode $|\mathbf{k}| = 1$ is especially sensitive to even small perturbations. This illustrates the strong nonlinear interactions
 1201 between the high and low wavenumber modes. The reduced-order method displays uniform skill in capturing the sensitive
 1202 responses in the large-scale modes for both positive and negative perturbation cases with only first 10×10 spectral modes
 1203 resolved compared with the 256×256 full resolution model.

1204 In Figure 5.7 and 5.8, we compare the model responses in both low/mid-latitude ocean and atmosphere regimes. In
 1205 this inhomogeneous regime with anisotropic jets, the statistical variables combine the responses in the mean and variance,
 1206 $P_{1,\mathbf{k}}^* P_{2,\mathbf{k}} = \bar{P}_{1,\mathbf{k}}^* \bar{P}_{2,\mathbf{k}} + P_{1,\mathbf{k}}^{\prime*} P_{2,\mathbf{k}}^{\prime}$, to display the total effect from the perturbation. In the ocean regime, the dominant mode with
 1207 largest sensitivity is at wavenumber $|\mathbf{k}| = 6$ due to the zonal jet structure. The sensitivity is captured with accuracy in the

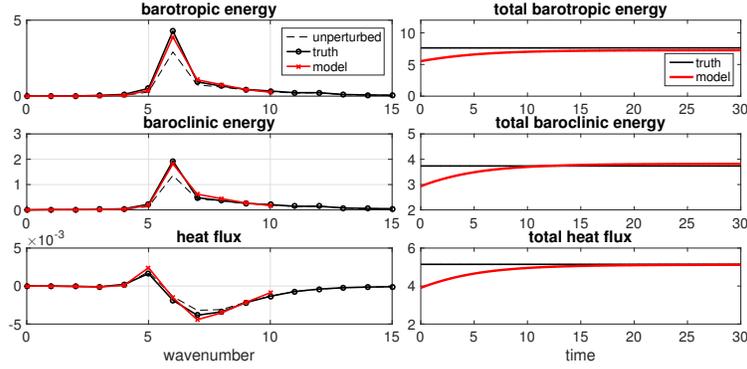


FIG. 5.7. Model responses in low/mid-latitude ocean regime with random forcing perturbation. The left panel shows the spectra for the barotropic and baroclinic energy as well as the heat flux with first ten modes resolved in the reduced-order method. The right panel is the time-series of the total energy and heat flux. The truth is shown in black lines while reduced-order model predictions are in red lines.

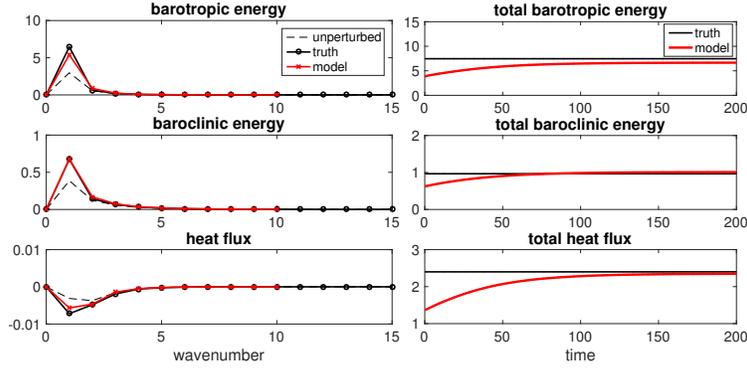


FIG. 5.8. Model responses in low/mid-latitude atmosphere regime with random forcing perturbation. The left panel shows the spectra for the barotropic and baroclinic energy as well as the heat flux with first 10 modes resolved in the reduced-order method. The right panel is the time-series of the total energy and heat flux. The truth is shown in dashed black lines while reduced-order model predictions are in red lines.

1208 reduced-order method. Also we compare the time evolution of the total resolved energy and heat flux. The prediction is
 1209 also good with small error. In the atmosphere regime, $|\mathbf{k}| = 1$ mode gets the largest statistical energy and is most sensitive to
 1210 perturbations. One important feature is the large change in the heat flux in the first two modes, representing the exchange of
 1211 energy in the dominant barotropic and baroclinic mode. Still the responses can be captured with accuracy in each mode in the
 1212 spectra as well as the total energy and heat flux profile with only 10^2 modes resolved. Note that in both cases, the heat flux is
 1213 weak due to the blocking effect from strong zonal jets.

1214 **6. Summary and Some Future Research Directions.** Understanding and improving the predictive skill of imperfect
 1215 models for high-dimensional complex turbulent systems is a formidable and challenging problem and has been investigated
 1216 through multiple approaches with various mathematical theories through the years [84, 29, 44, 63, 88, 47]. Low-order truncation
 1217 methods for statistical prediction can overcome the curse of dimensionality [23, 62] by concentrating on the subspace containing
 1218 largest variability. On the other hand, anisotropic turbulent processes are representative in many engineering and environmental
 1219 fluid flows [35, 26] where energy transports intermittently from the smaller scales to impact the largest scales in these flows.
 1220 Therefore significant model errors always occur due to the high wavenumber truncation in the imperfect model approximations.
 1221 A systematic information-theoretic framework has been shown useful to improve model fidelity and sensitivity [58, 59, 10] for
 1222 complex systems including perturbation formulas and multimodel ensembles that can be utilized to improve model error. In
 1223 many applications to complex systems with model error such as the climate change science [22, 74], it is crucially important to
 1224 provide guidelines to improve the predictive skill of imperfect models for their responses to changes in various external forcing

1225 perturbations.

1226 We discuss the general framework of efficient low-dimensional reduced-order models in this paper for turbulent dynamical
 1227 systems with nonlinearity to capture statistical responses to external perturbations. The validity of the reduced-order modeling
 1228 procedure is displayed via the simplest 3-dimensional triad model which is the building block of general turbulent systems, and
 1229 further on the more complicated two-layer barotropic model with huge model reduction. The computational cost is reduced
 1230 through a systematic approximation about the expensive nonlinear higher-order interactions. Additional damping and noise
 1231 corrections are proposed to replace the third-order moments. Model consistency in unperturbed equilibrium and sensitivity to
 1232 external perturbations are maintained through careful calibration about imperfect model error in a training phase before the pre-
 1233 diction. The model errors are calibrated and reduced effectively through a combination of linear response theory involving only
 1234 unperturbed equilibrium statistics and an information-theoretic framework using information theory. The general framework
 1235 has been tested in detail for a series of dynamical systems with increasing complexity [63, 78, 79, 54].

1236 For future development about the methods, there exist several interesting and promising directions that are worth further
 1237 investigation in the next stage:

- 1238 **A)** Tracking the model fluctuation statistics about the perfect statistical mean state. In many problems for turbulent systems,
 1239 we can assume the statistical mean state is known with reasonable accuracy by averaging along the data trajectory,
 1240 while the statistical fluctuations about the mean state are the quantities of interest [51, 64]. It is useful to consider
 1241 accurate and efficient ways to quantify the model fluctuations involving both uncertainties in the perfect system and
 1242 errors due to imperfect model approximation;
- 1243 **B)** Design of a mitigation control strategy by using novel low-order statistical models. There is need to combine control theory
 1244 with the statistical model reduction strategies for the principal large-scale modes in the turbulent dynamical systems.
 1245 For example, it is interesting to consider the effects of climate change using control and statistical modeling strategies;
- 1246 **C)** Predicting passive scalar turbulence with complex flow field. Besides the turbulent flow field, the dynamics of the passive
 1247 tracer advected by the turbulent flow has many interesting features with practical implications and is worth investigat-
 1248 ing. One important feature in the turbulent tracer field is the appearance of intermittency despite the near-Gaussian
 1249 statistics in the background advection flow. The intermittency in time-series and fat-tails in the passive tracer dis-
 1250 tributions have been observed in nature [73], and have been investigated under a simpler modeling framework both
 1251 theoretically and numerically [60, 66, 77]. It is thus interesting to consider low-order stochastic and statistical model-
 1252 ing about the tracer field advected by complex turbulent flows;
- 1253 **D)** More detailed consideration about the model nonlinearity. In the reduced-order approximation here, the overall strategy
 1254 does not require explicit calculation of the inefficient quadratic forms directly, but instead mimic the statistical sym-
 1255 metry in the nonlinearity in simple and efficient forms. Still it is interesting to check the improvement in the low-order
 1256 models with more detailed approximation.

1257 **Acknowledgment .** This research of the Andrew Majda is partially supported by the Office of Naval Research through
 1258 MURI N00014-16-1-2161 and DARPA through W911NF-15-1-0636. Di Qi is supported as a graduate research assistant on
 1259 these grant.

1260 **Appendix A. Numerical strategies to calculate the kicked response operators.** In the calibration step of the reduced-
 1261 order models, we use the statistical kicked response theory to tune the imperfect model parameters in the training phase. Here
 1262 we describe the details about calculating the kicked response operators for the mean and variance numerically. From the formula
 1263 in (2.10), the response operators for the mean and variance can be achieved from the perturbation part of the probability density
 1264 $\delta p'$. And this density function is also used to measure the information distance between the truth and imperfect model result in
 1265 the training phase. Below we describe the numerical procedure to get this distribution function $\delta p'$ for the true system and the
 1266 imperfect closure model separately.

- 1267 • *Kicked response for the true model:* For the true system, we want to achieve the most accurate possible estimation
 1268 for the response operators both for comparison with the imperfect model results and for calculating the FDT linear
 1269 prediction in (2.7). Therefore we use a Monte-Carlo simulation with an large enough ensemble size to capture the
 1270 response in density. The initial equilibrium ensemble is picked by sampling from a normal distribution with consistent
 1271 equilibrium mean and variance of the true system. For the kicked response to the mean, a constant perturbation with
 1272 ten percent of the equilibrium state mean $\delta \mathbf{u} = 0.1 \bar{\mathbf{u}}_{\text{eq}}$ is added to each initial ensemble member (in fact, as observed
 1273 in numerical experiments, this perturbation amplitude has little effect on the results in the response distribution as long
 1274 as it's not too large); and the initial variance of the ensemble is kept unchanged. The response distribution $\delta p'$ then
 1275 is achieved by monitoring the decay of the ensemble particles back to equilibrium under unperturbed dynamics and

1276 uniformly perturbed initial value (and the length of the time window that we need to monitor depends on the mixing
1277 property of the turbulent system). See [5, 48] for similar version of this algorithm.

1278 • *Kicked response for the imperfect model:* For the imperfect model, we just need to run the closure equations to get the
1279 responses for the mean and variance. In the same way as the true model, the initial mean is taken from the equilibrium
1280 distribution and a perturbation with amplitude $\delta \mathbf{u} = 0.1 \bar{\mathbf{u}}_{\text{eq}}$ is added to the initial mean state. The initial value for
1281 the variance is taken the same as the equilibrium state value and kept unperturbed. Then using this initial mean and
1282 variance, the imperfect model with specific closure strategies is applied to monitor the decay of the mean and variance
1283 back to equilibrium.

1284 One additional important point that requires attention is that even if the unperturbed equilibrium initial conditions are applied,
1285 the system will still deviate from the equilibrium state first and reapproach equilibrium again after some relaxation time. This
1286 is due to the insufficient characterization of the entire distribution of the true system with a Gaussian approximation (note that
1287 nonlinearities are also included in the imperfect closure methods). To eliminate this effect in computing the kicked response in
1288 both the true and imperfect models, we subtract the statistics computed using the unperturbed initial value from the statistics
1289 computed using the perturbed Gaussian initial condition to achieve more accurate characterization of the responses.

1290 **Appendix B. Explicit statistical dynamical formulations for the triad system.** We can derive for the triad system (4.1)
1291 the **dynamical equations for the mean state**

$$1292 \quad (\text{B.1a}) \quad \frac{d\bar{u}_1}{dt} = L_2 \bar{u}_3 - L_3 \bar{u}_2 - d_1 \bar{u}_1 + B_1 \left(\bar{u}_2 \bar{u}_3 + \overline{u'_2 u'_3} \right) + F_1,$$

$$1293 \quad (\text{B.1b}) \quad \frac{d\bar{u}_2}{dt} = L_3 \bar{u}_1 - L_1 \bar{u}_3 - d_2 \bar{u}_2 + B_2 \left(\bar{u}_3 \bar{u}_1 + \overline{u'_3 u'_1} \right) + F_2,$$

$$1294 \quad (\text{B.1c}) \quad \frac{d\bar{u}_3}{dt} = L_1 \bar{u}_2 - L_2 \bar{u}_1 - d_3 \bar{u}_3 + B_3 \left(\bar{u}_1 \bar{u}_2 + \overline{u'_1 u'_2} \right) + F_3.$$

1295 On the right hand sides of the above equations (B.1a)-(B.1c), the first parts include the skew-symmetric interactions between
1296 modes as well as the linear damping for the mean. The nonlinear interaction parts enter the mean dynamics both from the
1297 interactions between the mean states, and more importantly from the second-order moments of the fluctuations. Thus the mean
1298 dynamical equations are not closed by themselves due to the inclusion of unresolved higher-order statistics. Also note that in
1299 the second-order moments in the mean equations, diagonal variances won't appear while the cross-diagonal covariances take
1300 place as the role of transferring energy in the mean. Next we consider the dynamics for the fluctuation parts of the state variables

$$1301 \quad (\text{B.2a}) \quad \frac{du'_1}{dt} = L_2 u'_3 - L_3 u'_2 - d_1 u'_1 + B_1 \left(\bar{u}_2 u'_3 + u'_2 \bar{u}_3 + u'_2 u'_3 - \overline{u'_2 u'_3} \right) + \sigma_1 \dot{W}_1,$$

$$1302 \quad (\text{B.2b}) \quad \frac{du'_2}{dt} = L_3 u'_1 - L_1 u'_3 - d_2 u'_2 + B_2 \left(\bar{u}_3 u'_1 + u'_3 \bar{u}_1 + u'_3 u'_1 - \overline{u'_3 u'_1} \right) + \sigma_2 \dot{W}_2,$$

$$1303 \quad (\text{B.2c}) \quad \frac{du'_3}{dt} = L_1 u'_2 - L_2 u'_1 - d_3 u'_3 + B_3 \left(\bar{u}_1 u'_2 + u'_1 \bar{u}_2 + u'_1 u'_2 - \overline{u'_1 u'_2} \right) + \sigma_3 \dot{W}_3.$$

1304 The above equations can be achieved by subtracting the mean equations from the original triad system. Then the dynamics for
1305 higher order moments can be achieved through the fluctuations equations. Importantly, we can get the **dynamical equations**
1306 **for the variances** in each mode

$$1307 \quad (\text{B.3a}) \quad \frac{1}{2} \frac{d\overline{u_1'^2}}{dt} = L_2 \overline{u'_1 u'_3} - L_3 \overline{u'_1 u'_2} - d_1 \overline{u_1'^2} + B_1 \left(\bar{u}_2 \overline{u'_1 u'_3} + \overline{u'_1 u'_2} \bar{u}_3 \right) + B_1 \overline{u'_1 u'_2 u'_3} + \frac{1}{2} \sigma_1^2,$$

$$1308 \quad (\text{B.3b}) \quad \frac{1}{2} \frac{d\overline{u_2'^2}}{dt} = L_3 \overline{u'_1 u'_2} - L_1 \overline{u'_2 u'_3} - d_2 \overline{u_2'^2} + B_2 \left(\bar{u}_1 \overline{u'_2 u'_3} + \overline{u'_1 u'_2} \bar{u}_3 \right) + B_2 \overline{u'_1 u'_2 u'_3} + \frac{1}{2} \sigma_2^2,$$

$$1309 \quad (\text{B.3c}) \quad \frac{1}{2} \frac{d\overline{u_3'^2}}{dt} = L_1 \overline{u'_2 u'_3} - L_2 \overline{u'_1 u'_3} - d_3 \overline{u_3'^2} + B_3 \left(\bar{u}_1 \overline{u'_2 u'_3} + \overline{u'_1 u'_3} \bar{u}_2 \right) + B_3 \overline{u'_1 u'_2 u'_3} + \frac{1}{2} \sigma_3^2.$$

1310 And the **dynamical equations for the cross-covariances** between modes become

$$1311 \quad \frac{d\overline{u'_1 u'_2}}{dt} = L_2 \overline{u'_2 u'_3} - L_3 \overline{u'_2^2} - d_1 \overline{u'_1 u'_2} + B_1 \left(\overline{u_2 u'_2 u'_3} + \overline{u_2^2 \bar{u}_3} \right) + B_1 \overline{u'_2 u'_2 u'_3}$$

$$1312 \quad (B.4a) \quad + L_3 \overline{u'_1^2} - L_1 \overline{u'_1 u'_3} - d_2 \overline{u'_1 u'_2} + B_2 \left(\overline{u_1 u'_1 u'_3} + \overline{u_1^2 \bar{u}_3} \right) + B_2 \overline{u'_1 u'_1 u'_3},$$

$$1313 \quad \frac{d\overline{u'_1 u'_3}}{dt} = L_2 \overline{u'_3^2} - L_3 \overline{u'_2 u'_3} - d_1 \overline{u'_1 u'_3} + B_1 \left(\overline{u_2 u'_3^2} + \overline{u_2 u'_3 \bar{u}_3} \right) + B_1 \overline{u'_2 u'_3 u'_3}$$

$$1314 \quad (B.4b) \quad + L_1 \overline{u'_1 u'_2} - L_2 \overline{u'_1^2} - d_3 \overline{u'_1 u'_3} + B_3 \left(\overline{u_1 u'_1 u'_2} + \overline{u_1^2 \bar{u}_2} \right) + B_3 \overline{u'_1 u'_1 u'_2},$$

$$1315 \quad \frac{d\overline{u'_2 u'_3}}{dt} = L_3 \overline{u'_1 u'_3} - L_1 \overline{u'_3^2} - d_2 \overline{u'_2 u'_3} + B_2 \left(\overline{u_1 u'_3^2} + \overline{u_1 u'_3 \bar{u}_3} \right) + B_2 \overline{u'_1 u'_3 u'_3}$$

$$1316 \quad (B.4c) \quad + L_1 \overline{u'_2^2} - L_2 \overline{u'_1 u'_2} - d_3 \overline{u'_2 u'_3} + B_3 \left(\overline{u_1 u'_2^2} + \overline{u_1 u'_2 \bar{u}_2} \right) + B_3 \overline{u'_1 u'_2 u'_2}.$$

1317 For most situations, it is the diagonal variances in (B.3a)-(B.3c) that we are more interested in, while the off-diagonal covari-
 1318 ances (B.4a)-(B.4c) are less important and expensive to resolve. On the other hand, it is noticed that only the cross-covariance
 1319 terms take place in the central variance dynamics in (B.3a)-(B.3c) for the linear and quasi-linear interaction. In this typical
 1320 case, if we only consider the diagonal model and ignore the off-diagonal terms in the statistical closure dynamics, huge errors
 1321 could be introduced in the variance dynamical equations. Thus in the development of reduced order statistical models, careful
 1322 calibration about the unresolved components becomes crucial in the accuracy of the prediction results.

1323 With the dynamical equations for the mean (B.1) and for the variances in each mode (B.3) and (B.4), we can derive the
 1324 statistical energy dynamics following the general framework proposed in (2.2) of Theorem 2.1 and also [53]. The statistical
 1325 energy can be defined as the combination of the mean energy and the fluctuation energy as

$$1326 \quad E = \frac{1}{2} \sum_{i=1}^3 \left(\overline{u_i^2} + \overline{u_i'^2} \right).$$

1327 In the dynamics for the mean and covariance, the major difficulty in resolving the equations explicitly comes from the complex
 1328 third-order moments in Q_F as well as the covariance interactions through $R_{ij}B(\mathbf{e}_i, \mathbf{e}_j)$. However due to the conservation of
 1329 energy and the detailed triad symmetry in the triad system, nonlinear interactions cancel in the mean and variance equations
 1330 and the **statistical energy equation** becomes

$$1331 \quad (B.5) \quad \frac{dE}{dt} = \frac{d}{dt} \left(\frac{1}{2} \sum_{i=1}^3 \left(\overline{u_i^2} + \overline{u_i'^2} \right) \right) = - \sum_{i=1}^3 d_i \left(\overline{u_i^2} + \overline{u_i'^2} \right) + \sum_{i=1}^3 \left(F_i \bar{u}_i + \frac{1}{2} \sigma_i^2 \right).$$

1332 Therefore the total statistical structure can be calculated from (B.5) without knowing the higher-order moments as well as
 1333 cross-covariances which are in general difficult to resolve exactly without error. Furthermore, considering the special case with
 1334 homogeneous damping in each mode, $d_j \equiv d$, the dissipation term on the right hand side of (B.5) becomes $-dE$. Thus we can
 1335 get the total second-order statistics from only the information in the first-order moments with the help of the statistical energy
 1336 dynamics.

1337

REFERENCES

- 1338 [1] Rafail V Abramov and Andrew J Majda. Blended response algorithms for linear fluctuation-dissipation for complex nonlinear dynamical systems.
 1339 *Nonlinearity*, 20(12):2793–2821, 2007.
- 1340 [2] Rafail V Abramov and Andrew J Majda. Low-frequency climate response of quasigeostrophic wind-driven ocean circulation. *Journal of Physical*
 1341 *Oceanography*, 42(2):243–260, 2012.
- 1342 [3] Nadine Aubry, Wen-Yu Lian, and Edriss S Titi. Preserving symmetries in the proper orthogonal decomposition. *SIAM Journal on Scientific Computing*,
 1343 14(2):483–505, 1993.
- 1344 [4] Nikolaos A Bakas and Petros J Ioannou. Structural stability theory of two-dimensional fluid flow under stochastic forcing. *Journal of Fluid Mechanics*,
 1345 682:332–361, 2011.
- 1346 [5] Thomas L Bell. Climate sensitivity from fluctuation dissipation: Some simple model tests. *Journal of the Atmospheric Sciences*, 37(8):1700–1707,
 1347 1980.
- 1348 [6] Thomas Bengtsson, Peter Bickel, Bo Li, et al. Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems. In *Probability*
 1349 *and statistics: Essays in honor of David A. Freedman*, pages 316–334. Institute of Mathematical Statistics, 2008.

- 1350 [7] Michal Branicki, Nan Chen, and Andrew J Majda. Non-Gaussian test models for prediction and state estimation with model errors. *Chinese Annals of*
1351 *Mathematics, Series B*, 34(1):29–64, 2013.
- 1352 [8] Michal Branicki and Andrew J Majda. Quantifying uncertainty for predictions with model error in non-Gaussian systems with intermittency. *Nonlin-*
1353 *earity*, 25(9):2543–2578, 2012.
- 1354 [9] Michal Branicki and Andrew J Majda. Fundamental limitations of polynomial chaos for uncertainty quantification in systems with intermittent instabil-
1355 ities. *Comm. Math. Sci*, 11(1):55–103, 2013.
- 1356 [10] Michal Branicki and Andrew J Majda. An information-theoretic framework for improving imperfect dynamical predictions via multi-model ensemble
1357 forecasts. *Journal of Nonlinear Science*, 25(3):489–538, 2015.
- 1358 [11] Kenneth P Burnham and David Anderson. Model selection and multi-model inference. *A Pratical informatio-theoric approach*. Springer, 2003.
- 1359 [12] GF Carnevale, M Falcioni, S Isola, R Purini, and Angelo Vulpiani. Fluctuation-response relations in systems with chaotic behavior. *Physics of Fluids*
1360 *A: Fluid Dynamics (1989-1993)*, 3(9):2247–2254, 1991.
- 1361 [13] Abhijit Chatterjee and Dionisios G Vlachos. An overview of spatial microscopic and accelerated kinetic Monte Carlo methods. *Journal of computer-*
1362 *aided materials design*, 14(2):253–308, 2007.
- 1363 [14] N Chen and AJ Majda. Predicting the cloud patterns for the boreal summer intraseasonal oscillation through a low-order stochastic model. *Mathematics*
1364 *of Climate and Weather Forecasting*, 1(1), 2015.
- 1365 [15] N Chen and AJ Majda. Predicting the real-time multivariate madden–julian oscillation index through a low-order nonlinear stochastic model. *Monthly*
1366 *Weather Review*, 143(6):2148–2169, 2015.
- 1367 [16] N Chen, AJ Majda, and D Giannakis. Predicting the cloud patterns of the Madden-Julian oscillation through a low-order nonlinear stochastic model.
1368 *Geophysical Research Letters*, 41:5612–5619, 2014.
- 1369 [17] Noel Cressie and Christopher K Wikle. *Statistics for spatio-temporal data*. John Wiley & Sons, 2015.
- 1370 [18] DT Crommelin and AJ Majda. Strategies for model reduction: comparing different optimal bases. *Journal of the Atmospheric Sciences*, 61(17):2206–
1371 2217, 2004.
- 1372 [19] Timothy DelSole. *Predictability and information theory*, volume 1. Center for Ocean-Land-Atmosphere Studies, 2003.
- 1373 [20] Timothy DelSole and Jagadish Shukla. Model fidelity versus skill in seasonal forecasting. *Journal of Climate*, 23(18):4794–4806, 2010.
- 1374 [21] Weinan E, JC Mattingly, and Ya Sinai. Gibbsian dynamics and ergodicity for the stochastically forced Navier–Stokes equation. *Communications in*
1375 *Mathematical Physics*, 224(1):83–106, 2001.
- 1376 [22] KA Emanuel, JC Wyngaard, JC McWilliams, DA Randall, and YL Yung. Improving the scientific foundation for atmosphere-land-ocean simulations.
1377 *Natl. Acad. Press, Washington DC*, page 72, 2005.
- 1378 [23] Edward S Epstein. Stochastic dynamic prediction. *Tellus*, 21(6):739–759, 1969.
- 1379 [24] Edward S Epstein and Rex J Fleming. Depicting stochastic dynamic forecasts. *Journal of the Atmospheric Sciences*, 28(4):500–511, 1971.
- 1380 [25] Rex J Fleming. On stochastic dynamic prediction. *Mon. Wea. Rev.*, 99:851–872, 1971.
- 1381 [26] Uriel Frisch. *Turbulence: the legacy of AN Kolmogorov*. Cambridge university press, 1995.
- 1382 [27] Boris Gershgorin and Andrew J Majda. Quantifying uncertainty for climate change and long-range forecasting scenarios with model errors. Part I:
1383 Gaussian models. *Journal of Climate*, 25(13):4523–4548, 2012.
- 1384 [28] Dimitrios Giannakis and Andrew J Majda. Quantifying the predictive skill in long-range forecasting. Part I: Coarse-grained predictions in a simple
1385 ocean model. *Journal of Climate*, 25(6):1793–1813, 2012.
- 1386 [29] Dimitrios Giannakis and Andrew J Majda. Quantifying the predictive skill in long-range forecasting. Part II: Model error in coarse-grained Markov
1387 models with application to ocean-circulation regimes. *Journal of Climate*, 25(6):1814–1826, 2012.
- 1388 [30] Alexander Gluhovsky and Ernest Agee. An interpretation of atmospheric low-order models. *Journal of the atmospheric sciences*, 54(6):768–773, 1997.
- 1389 [31] Andrey Gritsun and Grant Branstator. Climate response using a three-dimensional operator based on the fluctuation-dissipation theorem. *Journal of the*
1390 *atmospheric sciences*, 64(7):2558–2575, 2007.
- 1391 [32] Andrey Gritsun, Grant Branstator, and Andrew Majda. Climate response of linear and quadratic functionals using the fluctuation-dissipation theorem.
1392 *Journal of the Atmospheric Sciences*, 65(9):2824–2841, 2008.
- 1393 [33] Martin Hairer and Andrew J Majda. A simple framework to justify linear response theory. *Nonlinearity*, 23(4):909, 2010.
- 1394 [34] John Harlim, Adam Mahdi, and Andrew J Majda. An ensemble Kalman filter for statistical estimation of physics constrained nonlinear regression
1395 models. *Journal of Computational Physics*, 257:782–812, 2014.
- 1396 [35] JO Hinze. *Turbulence: An introduction to its mechanisms and theory*. Mechanical Engineering, 1959.
- 1397 [36] Philip Holmes, John L Lumley, and Gal Berkooz. *Turbulence, coherent structures, dynamical systems and symmetry*. Cambridge university press, 1998.
- 1398 [37] Thomas Y Hou, Wuan Luo, Boris Rozovskii, and Hao-Min Zhou. Wiener chaos expansions and numerical solutions of randomly forced equations of
1399 fluid mechanics. *Journal of Computational Physics*, 216(2):687–706, 2006.
- 1400 [38] Edwin T Jaynes. Information theory and statistical mechanics. *Physical review*, 106(4):620, 1957.
- 1401 [39] Markos A Katsoulakis, Andrew J Majda, and Dionisios G Vlachos. Coarse-grained stochastic processes for microscopic lattice systems. *Proceedings*
1402 *of the National Academy of Sciences*, 100(3):782–787, 2003.
- 1403 [40] Richard Kleeman, Andrew J Majda, and Ilya Timofeyev. Quantifying predictability in a model with statistical features of the atmosphere. *Proceedings*
1404 *of the National Academy of Sciences*, 99(24):15291–15296, 2002.
- 1405 [41] Omar M Knio, Habib N Najm, Roger G Ghanem, et al. A stochastic projection method for fluid flow: I. basic formulation. *Journal of computational*
1406 *Physics*, 173(2):481–511, 2001.
- 1407 [42] Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, pages 79–86, 1951.
- 1408 [43] CE Leith. Climate response and fluctuation dissipation. *Journal of the Atmospheric Sciences*, 32(10):2022–2026, 1975.
- 1409 [44] Marcel Lesieur. *Turbulence in fluids*, volume 40. Springer Science & Business Media, 2012.
- 1410 [45] Edward N Lorenz. Deterministic nonperiodic flow. *Journal of the atmospheric sciences*, 20(2):130–141, 1963.
- 1411 [46] Edward N Lorenz. Predictability: A problem partly solved. In *Proc. Seminar on predictability*, volume 1, 1996.
- 1412 [47] Nicholas J Lutsko, Isaac M Held, and Pablo Zurita-Gotor. Applying the fluctuation–dissipation theorem to a two-layer model of quasigeostrophic
1413 turbulence. *Journal of the Atmospheric Sciences*, 72(8):3161–3177, 2015.
- 1414 [48] Andrew Majda, Rafail V Abramov, and Marcus J Grote. *Information theory and stochastics for multiscale nonlinear systems*, volume 25. American

- 1415 Mathematical Soc., 2005.
- 1416 [49] Andrew Majda, Richard Kleeman, and David Cai. A mathematical framework for quantifying predictability through relative entropy. *Methods and*
1417 *Applications of Analysis*, 9(3):425–444, 2002.
- 1418 [50] Andrew Majda, Ilya Timofeyev, and Eric Vanden-Eijnden. A priori tests of a stochastic mode reduction strategy. *Physica D: Nonlinear Phenomena*,
1419 170(3):206–252, 2002.
- 1420 [51] Andrew Majda and Xiaoming Wang. *Nonlinear dynamics and statistical theories for basic geophysical flows*. Cambridge University Press, 2006.
- 1421 [52] Andrew Majda and Xiaoming Wang. Linear response theory for statistical ensembles in complex systems with time-periodic forcing. *Communications*
1422 *in Mathematical Sciences*, 8(1):145–172, 2010.
- 1423 [53] Andrew J Majda. Statistical energy conservation principle for inhomogeneous turbulent dynamical systems. *Proceedings of the National Academy of*
1424 *Sciences*, 112(29):8937–8941, 2015.
- 1425 [54] Andrew J Majda. *Introduction to turbulent dynamical systems in complex systems*. Springer, 2016.
- 1426 [55] Andrew J Majda, Rafail Abramov, and Boris Gershgorin. High skill in low-frequency climate response through fluctuation dissipation theorems despite
1427 structural instability. *Proceedings of the National Academy of Sciences*, 107(2):581–586, 2010.
- 1428 [56] Andrew J Majda and Michal Branicki. Lessons in uncertainty quantification for turbulent dynamical systems. *Discrete Cont. Dyn. Systems*, 32(9):3133–
1429 3221, 2012.
- 1430 [57] Andrew J Majda and Boris Gershgorin. Quantifying uncertainty in climate change science through empirical information theory. *Proceedings of the*
1431 *National Academy of Sciences*, 107(34):14958–14963, 2010.
- 1432 [58] Andrew J Majda and Boris Gershgorin. Improving model fidelity and sensitivity for complex systems through empirical information theory. *Proceedings*
1433 *of the National Academy of Sciences*, 108(25):10044–10049, 2011.
- 1434 [59] Andrew J Majda and Boris Gershgorin. Link between statistical equilibrium fidelity and forecasting skill for complex systems with model error.
1435 *Proceedings of the National Academy of Sciences*, 108(31):12599–12604, 2011.
- 1436 [60] Andrew J Majda and Boris Gershgorin. Elementary models for turbulent diffusion with complex physical features: eddy diffusivity, spectrum and
1437 intermittency. *Phil. Trans. R. Soc. A*, 371: 20120184, 2013.
- 1438 [61] Andrew J Majda, Ilya Timofeyev, and Yuan Yuan. Low-frequency climate response and fluctuation-dissipation theorems: theory and practice. *Journal*
1439 *of the Atmospheric Sciences*, 67(4):1186–1201, 2010.
- 1440 [62] Andrew J Majda and John Harlim. Physics constrained nonlinear regression models for time series. *Nonlinearity*, 26(1):201–217, 2012.
- 1441 [63] Andrew J Majda and Di Qi. Improving prediction skill of imperfect turbulent models through statistical response and information theory. *Journal of*
1442 *Nonlinear Science*, 26(1):233–285, 2016.
- 1443 [64] Andrew J Majda, Ilya Timofeyev, and Eric Vanden-Eijnden. Systematic strategies for stochastic mode reduction in climate. *Journal of the Atmospheric*
1444 *Sciences*, 60(14):1705–1722, 2003.
- 1445 [65] Andrew J Majda and Xin T Tong. Ergodicity of truncated stochastic Navier-Stokes with deterministic forcing and dispersion. *Journal of Nonlinear*
1446 *Science*, 26(5):1483–1506, 2015.
- 1447 [66] Andrew J Majda and Xin T Tong. Intermittency in turbulent diffusion models with a mean gradient. *Nonlinearity*, 28(11):4171–4208, 2015.
- 1448 [67] Andrew J Majda and Yuan Yuan. Fundamental limitations of ad hoc linear and quadratic multi-level regression models for physical systems. *Discrete*
1449 *and Continuous Dynamical Systems B*, 17(4):1333–1363, 2012.
- 1450 [68] Umberto Marini Bettolo Marconi, Andrea Puglisi, Lamberto Rondoni, and Angelo Vulpiani. Fluctuation–dissipation: response theory in statistical
1451 physics. *Physics reports*, 461(4):111–195, 2008.
- 1452 [69] Henry P McKean. A class of markov processes associated with nonlinear parabolic equations. *Proceedings of the National Academy of Sciences*,
1453 56(6):1907–1911, 1966.
- 1454 [70] Mustafa A Mohamad, Will Cousins, and Themistoklis P Sapsis. A probabilistic decomposition-synthesis method for the quantification of rare events
1455 due to internal instabilities. *arXiv preprint arXiv:1511.02392*, 2015.
- 1456 [71] Mustafa A Mohamad and Themistoklis P Sapsis. Probabilistic response and rare events in mathieu s equation under correlated parametric excitation.
1457 *Ocean Engineering*, 120:289–297, 2016.
- 1458 [72] Habib N Najm. Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics. *Annual Review of Fluid Mechanics*,
1459 41:35–52, 2009.
- 1460 [73] J David Neelin, Benjamin R Lintner, Baijun Tian, Qinbin Li, Li Zhang, Prabir K Patra, Moustafa T Chahine, and Samuel N Stechmann. Long tails in
1461 deep columns of natural and anthropogenic tropospheric tracers. *Geophysical Research Letters*, 37(5), 2010.
- 1462 [74] J David Neelin, Matthias Münnich, Hui Su, Joyce E Meyerson, and Christopher E Holloway. Tropical drying trends in global warming models and
1463 observations. *Proceedings of the National Academy of Sciences*, 103(16):6110–6115, 2006.
- 1464 [75] Dwight Roy Nicholson. *Introduction to plasma theory*. Cambridge Univ Press, 1983.
- 1465 [76] Stephen B Pope. *Turbulent flows*. IOP Publishing, Bristol, 2001.
- 1466 [77] Di Qi and Andrew J Majda. Predicting fat-tailed intermittent probability distributions in passive scalar turbulence with imperfect models through
1467 empirical information theory. *Comm. Math. Sci.*, 14(6):1687–1722, 2015.
- 1468 [78] Di Qi and Andrew J Majda. Low-dimensional reduced-order models for statistical response and uncertainty quantification: barotropic turbulence with
1469 topography. *Physica D: Nonlinear Phenomena*, <http://dx.doi.org/10.1016/j.physd.2016.11.006>, 2016.
- 1470 [79] Di Qi and Andrew J Majda. Low-dimensional reduced-order models for statistical response and uncertainty quantification: two-layer baroclinic turbu-
1471 lence. *Journal of the Atmospheric Sciences*, 73(12):4609–4639, 2016.
- 1472 [80] David A Randall, Richard A Wood, Sandrine Bony, Robert Colman, Thierry Fichefet, John Fyfe, Vladimir Kattsov, Andrew Pitman, Jagadish Shukla,
1473 Jayaraman Srinivasan, et al. Climate models and their evaluation. In *Climate change 2007: The physical science basis. Contribution of Working*
1474 *Group I to the Fourth Assessment Report of the IPCC (FAR)*, pages 589–662. Cambridge University Press, 2007.
- 1475 [81] Aaditya V Rangan, Louis Tao, Gregor Kovacic, and David Cai. Multiscale modeling of the primary visual cortex. *IEEE Engineering in Medicine and*
1476 *Biology Magazine*, 28(3):19–24, 2009.
- 1477 [82] James C Robinson. *Infinite-dimensional dynamical systems: an introduction to dissipative parabolic PDEs and the theory of global attractors*, vol-
1478 ume 28. Cambridge University Press, 2001.
- 1479 [83] David Ruelle. Differentiation of SRB states. *Communications in Mathematical Physics*, 187(1):227–241, 1997.

- 1480 [84] Rick Salmon. *Lectures on geophysical fluid dynamics*. Oxford University Press, 1998.
- 1481 [85] Themistoklis P Sapsis. Attractor local dimensionality, nonlinear energy transfers and finite-time instabilities in unstable dynamical systems with appli-
1482 cations to two-dimensional fluid flows. In *Proc. R. Soc. A*, volume 469, page 20120550. The Royal Society, 2013.
- 1483 [86] Themistoklis P Sapsis and Pierre FJ Lermusiaux. Dynamically orthogonal field equations for continuous stochastic dynamical systems. *Physica D: Nonlinear Phenomena*, 238(23):2347–2360, 2009.
- 1484 [87] Themistoklis P Sapsis and Andrew J Majda. Blended reduced subspace algorithms for uncertainty quantification of quadratic systems with a stable
1485 mean state. *Physica D: Nonlinear Phenomena*, 258:61–76, 2013.
- 1486 [88] Themistoklis P Sapsis and Andrew J Majda. Statistically accurate low-order models for uncertainty quantification in turbulent dynamical systems.
1487 *Proceedings of the National Academy of Sciences*, 110(34):13705–13710, 2013.
- 1488 [89] Themistoklis P Sapsis and Andrew J Majda. A statistically accurate modified quasilinear Gaussian closure for uncertainty quantification in turbulent
1489 dynamical systems. *Physica D: Nonlinear Phenomena*, 252:34–45, 2013.
- 1490 [90] Kaushik Srinivasan and WR Young. Zonostrophic instability. *Journal of the atmospheric sciences*, 69(5):1633–1656, 2012.
- 1491 [91] Albert A Townsend. *The structure of turbulent shear flow*. Cambridge university press, 1980.
- 1492 [92] Geoffrey K Vallis. *Atmospheric and oceanic fluid dynamics: Fundamentals and large-scale circulation*. Cambridge University Press, 2006.
- 1493 [93] Lai-Sang Young. What are SRB measures, and which dynamical systems have them? *Journal of Statistical Physics*, 108(5):733–754, 2002.
- 1494