J Nonlinear Sci manuscript No.
(will be inserted by the editor)

Improving Prediction Skill of Imperfect Turbulent Models through Statistical Response and Information Theory

Andrew J. Majda, and Di Qi

Received: date / Accepted: date

Abstract Turbulent dynamical systems with a large phase space and a high degree of instabilities are ubiquitous in climate science and engineering applications. Statistical uncertainty quantification (UQ) to the response to the change in forcing or uncertain initial data in such complex turbulent systems requires the use of imperfect models due to both the lack of physical understanding and the overwhelming computational demands of Monte Carlo simulation with a large dimensional phase space. Thus, the systematic development of reduced low order imperfect statistical models for UQ in turbulent dynamical systems is a grand challenge. This paper applies a recent mathematical strategy for calibrating imperfect models in a training phase and accurately predicting the response by combining information theory and linear statistical response theory in a systematic fashion. A systematic hierarchy of simple statistical imperfect closure schemes for UQ for these problems are designed and tested which are built through new local and global statistical energy conservation principles combined with statistical equilibrium fidelity. The forty mode Lorenz 96 (L-96) model which mimics forced baroclinic turbulence is utilized as a test bed for the calibration and predicting phases for the hierarchy of computationally cheap imperfect closure models both in the full phase space and in a reduced three dimensional subspace containing the most energetic modes. In all of phase spaces, the nonlinear response of the true model is captured accurately for the mean and variance by the systematic closure model while alternative methods based on the fluctuation dissipation theorem alone are much less accurate. For reduced order model for UQ in the three dimensional subspace for L-96, the systematic low order imperfect closure models coupled with the training strategy provide the highest predictive skill over other existing methods for general forced response yet have simple design principles based on a statistical global energy equation. The systematic imperfect closure models and the calibration strategies for UQ for the L-96 model serve as a new template for similar strategies for UQ with model error in vastly more complex realistic turbulent dynamical systems.

Keywords turbulent systems · low-order statistical closure models · linear response theory · information metric

1 Introduction

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. The Earth's climate is a perfect example of such an extremely complex and only partially known system coupling physical processes for the atmosphere, ocean, and land over a wide range of spatial and temporal scales (Majda (2003); Emanuel et al (2005); Neelin et al (2006)). 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 initial values, requiring naturally a probabilistic characterization for the evolution of the system. Uncertainty

Andrew J. Majda
Department of Mathematics and Center for Atmosphere and Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, NY 10012
E-mail: jonjon@cims.nyu.edu

Di Qi
Department of Mathematics and Center for Atmosphere and Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, NY 10012
Tel.: (+1)347-735-0576
E-mail: qidi@cims.nyu.edu

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 response 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 at hand. One central issue in contemporary climate science is the development of a systematic methodology that can recover the crucial features of the natural system in equilibrium/climate (model fidelity), and then improve the imperfect model prediction skill in response to external perturbations like the climate change (model sensitivity). A conceptual framework intermediate between detailed dynamical physical modeling and purely statistical analysis based on empirical information theory has been proposed (Majda and Gershgorin (2011a, 2010); Gershgorin and Majda (2012)) to address imperfect model fidelity and sensitivity problems. This information-theoretic framework based on the relative entropy, which offers an unbiased and invariant measure for model distributions (Kullback and Leibler (1951); Majda et al (2002, 2005)), has been utilized to systematically improve model fidelity and sensitivity (Majda and Gershgorin (2010, 2011a)), and to make an empirical link between model fidelity and forecasting skill (DelSole (2005); DelSole and Shukla (2010)). However equilibrium statistical fidelity is a necessary but not sufficient condition for imperfect model predictive skill. Simple examples (Majda and Gershgorin (2011a); Majda and Branicki (2012); Majda et al (2005)) with perfect model equilibrium fidelity can demonstrate the large intrinsic barrier for capturing the correct sensitivity to perturbed dynamics under imperfect models. In Majda and Gershgorin (2011b), a direct link by utilizing fluctuation dissipation theorem (FDT) for complex systems together with the framework of empirical information theory for improving imperfect models is developed. The FDT utilizes the linear response theory (Leith (1975); Majda et al (2005, 2010a); Hairer and Majda (2010)) to predict linear first order responses to perturbations, requiring only measurements in the unperturbed system. The potential validity of the FDT for idealized climate models with various approximations and numerical procedures has been widely studied both in developing theory (Majda et al (2010b); Majda and Wang (2006); Bell (1980)) and algorithms (Abramov and Majda (2007, 2008, 2009); Gritsun and Branstator (2007); Gritsun et al (2008); Carnevale et al (1991)). Despite some success in complex systems, this method is hampered by the fundamental limitation to parameter regimes with linear statistical response. Furthermore one important practical problem is that FDT needs to be applied on a reduced subspace for realistic complex systems (Majda et al (2010c)) where various approximations are utilized with reasonable accuracy for the mean but deteriorating skill for the variance (Bell (1980); Majda et al (2005, 2010c)). The information barrier with climate consistent models and skill in linear FDT predictions are displayed on simple models with intermittent instabilities (Branicki and Majda (2012); Majda et al (2010c)) through various moment closure approximations for systems with model error. Thus, new strategies for imperfect low order models on subspaces are important and are a main theme of the present research (Sapsis and Majda (2013c)).

In this paper, we investigate and develop systematic strategies for improving the imperfect model prediction skill for complex turbulent dynamical systems by employing ideas in both the information-theoretic framework and linear response theory mentioned above. Following the direct link between the linear response and empirical information theory demonstrated in Majda and Gershgorin (2011b) for models with equilibrium fidelity, it is shown that they can be seamlessly combined into a precise systematic framework to improve imperfect model sensitivity through measuring the information error of the *linear response operator* in the training phase with unperturbed statistics. Without dependence on the specific perturbation form, this framework in general can be applied to any turbulent dynamic systems with model error for predicting any perturbed responses. Instead of seeking for practical implementations of high-dimensional realistic systems such as atmospheric global circulation models (GCMs), we begin with a hierarchy of simpler second-order models (with only statistical mean and covariance involved) mimicking increasingly complex features of the vastly more complex true natural systems. The simplified, mathematical tractable models offer more controllable scenarios for concentrating on the core mathematical mechanisms in turbulent systems, and for developing systematic methodology for model improvement. We focus on a generic class of turbulent dynamical systems characterized by quadratic nonlinearity associated with conservative energy-transfer mechanisms. Specifically, the 40 dimensional Lorenz 96 (L-96) system first introduced in Lorenz (1996) is taken as one typical testbed for the illustration of this class of dynamical systems. With the help of the simplified structure in the L-96 system, exact statistical dynamical equations for the first two order moments are derived for both the perfect system and imperfect systems with model error. Important statistical features then are analyzed for a deeper understanding of the strengths and limitations of second-order schemes, and the potential information barrier with imperfect consideration of the entire energy spectra is also discussed for constructing better closure schemes. Based on the understanding about energy transfer mechanism as well as the potential information barrier, several second-order closure methods with increasing accuracy and complexity are then proposed. Despite the unavoidable model error from replacing higher order statistics by the low order closure approximations, we display

the uniform ability of the proposed information-response framework to improve imperfect model prediction skill with various types of external perturbations.

One very important further issue arises when the large dimension of the active variables in turbulent dynamical systems makes the UQ scheme computationally impossible to resolve statistics in the entire phase space. This is called the 'curse of dimensionality' that has been investigated from various points of view (Bengtsson et al (2008); Majda and Branicki (2012)). Reduced order truncation models for UQ (Sapsis and Majda (2013c)) have been developed focusing on the key physical quantities including statistics in leading order empirical orthogonal functions (EOFs). However, as discussed in Section 3, the energy transfers between unresolved modes (or between resolved mode and unresolved mode) play an important role in stabilizing the nonlinear system. Several additional systematic corrections according to the previous analysis of the statistical dynamics will be suggested for more careful quantification about the unresolved statistics in the low-order closure models. And we generalize the imperfect model optimization framework for tuning parameters to the reduced subspatial case by minimizing the information error in the resolved low-dimensional subspace of interest. Again we test the reduced order methods on the 40 dimensional L-96 testbed with only 2 resolved modes, and compare the improvement from the correction ideas.

In the following part of this paper, first in Section 2 we will introduce several theoretical toolkits including the information theory and the linear response theory which are important foundations for the development of the methods. Then the statistical features of the nonlinear dynamical systems with quadratic terms are investigated in Section 3. Without loss of generality, we use the L-96 system with 40 grid points as a standard test bed to illustrate the statistical dynamics. Knowing the statistical structures of the true dynamical system, imperfect models with statistical closure strategies are then proposed in Section 4 and 5. Parameters of the imperfect models need to be tuned in a training phase to guarantee statistical equilibrium fidelity and optimal responses to external perturbations under the information measure. Statistical closure models in full phase space are described in Section 4. And further for genuinely high dimensional systems, resolving the statistics in the entire phase space is computational forbidden and unnecessary. Reduced order models are then developed in Section 5 aiming at capturing only the most important statistics in the primary directions. Performances of these closure models in response to various external forcing perturbations are compared for both full spatial models and reduced order models using the L-96 system. Conclusions and future works are summarized in Section 6.

2 Theories for improving imperfect model prediction skill

Before investigating the specific statistical features of the turbulent dynamical systems, we propose one systematical framework for improving and optimizing systems with model error in general. Usually among one class of proposed imperfect models with several parameters, the problems include selecting the optimal model with statistical consistency in the equilibrium steady state, and more importantly with accurate sensitivity to different external perturbations. Particularly, we would like one unified strategy that can improve the model performance uniformly for all kinds of perturbations rather than an impractical process to tune the parameters individually for each specific case. The information theory offers a least biased measure for quantifying the error between the imperfect model prediction and the truth; and the linear response theory gives an important tool relating the model responses to stationary state statistics of the dynamical system. We will describe the basic ideas for these useful mathematical tools in the first place. Then with the help of these theories (Majda and Gershgorin (2011a,b)), one systematical process to tune model parameters in a training phase to possibly achieve the optimal model with sensitivity to all kinds of perturbations is discussed.

2.1 Empirical information theory

A natural way to measure the lack of information in one probability density from the imperfect model, π^M, compared with the true probability density, π, is through the *relative entropy* or *information distance* (Kullback and Leibler (1951); Majda et al (2002)), given by

$$\mathscr{P}\left(\pi,\pi^M\right)=\int \pi \log \frac{\pi}{\pi^M}. \tag{1}$$

Despite the lack of symmetry in its arguments, the relative entropy, $\mathscr{P}\left(\pi,\pi^M\right)$ provides an attractive framework for assessing model error like a metric. Importantly, the following two crucial features are satisfied: (i) $\mathscr{P}\left(\pi,\pi^M\right)\geq 0$, and the equality holds if and only if $\pi=\pi^M$; and (ii) it is invariant under any invertible change of variables. The most practical setup for utilizing the framework of empirical information theory arises when only the Gaussian statistics of the distributions are considered. By only comparing the first two moments of the density functions, we get the following fact (Majda et al (2002)):

Fact 1: If the density functions π, π^M involve only the first two moments, that is, $\pi \sim \mathscr{N}(\bar{\mathbf{u}}, R)$ and $\pi^M \sim \mathscr{N}(\bar{\mathbf{u}}_M, R_M)$, the relative entropy in (1) has the explicit formula

$$\mathscr{P}\left(\pi, \pi^M\right) = \frac{1}{2}\left(\bar{\mathbf{u}} - \bar{\mathbf{u}}_M\right)^T R_M^{-1}\left(\bar{\mathbf{u}} - \bar{\mathbf{u}}_M\right) + \frac{1}{2}\left(\mathrm{tr}\left(RR_M^{-1}\right) - J - \log\det\left(RR_M^{-1}\right)\right). \tag{2}$$

The first term on the right hand side of (2) is called the *signal*, reflecting the model error in the mean but weighted by the inverse of the model variance R_M, whereas the second term is the *dispersion*, which involves only the model error covariance ratio RR_M^{-1}, measuring the differences in the covariance matrix.

Remark. Sometimes it might be useful to only measure the information distance $\mathscr{P}\left(\pi\,|_{V_s}, \pi^M\,|_{V_s}\right)$ between the marginal distributions inside a subspace V_s (say, when we want to check the uncertainty along the principal EOFs). If we are only interested in the relative entropy in a subspace spanned by basis $P = [\mathbf{v}_1, \cdots, \mathbf{v}_s]$, we only need replace the covariance matrix R by the reduced order covariance $R_s = P^T R P$ in the above formula (2).

2.2 Kicked response theory

Assume the perfect model of a dynamical system is

$$\mathbf{u}_t = \mathbf{f}(\mathbf{u}), \tag{3}$$

with equilibrium measure assumed as π_{eq}. We are interested in the model's response to an external spatially uniform perturbation $\delta\mathbf{f}' = \delta\mathbf{w}(\mathbf{u}) f'(t)$ added to the system in the form of

$$\mathbf{f}^\delta = \mathbf{f} + \delta\mathbf{f}'. \tag{4}$$

Therefore the resulting perturbed probability density π^δ can be asymptotically expanded as

$$\pi^\delta(t) = \pi_{\mathrm{eq}} + \delta\pi'(t), \qquad \int \delta\pi' = 0. \tag{5}$$

And the corresponding statistics of some functional $A(\mathbf{u})$ under this perturbed density function can be expressed as

$$\mathbb{E}^\delta A(\mathbf{u}) = E_{\mathrm{eq}}(A) + \delta E_A', \tag{6}$$

where $E_{\mathrm{eq}}(A) = \int A(\mathbf{u})\pi_{\mathrm{eq}}(\mathbf{u})$ is the expectation of A according to equilibrium distribution π_{eq}, while $\delta E_A' = \int A(\mathbf{u})\delta\pi'(\mathbf{u})$ is according to $\delta\pi'$. Linear response theory states the following fact (Majda et al (2005)):

Fact 2: If δ is small enough, the leading order correction to the statistics in (6) in response to the spatially uniform perturbation in (4) becomes

$$\delta E_A' = \int_0^t \mathscr{R}_A(t-s)\,\delta f'(s)\,ds + O\left(\delta^2\right), \tag{7}$$

where $\mathscr{R}_A(t)$ is called the *linear response operator* that is only related to the statistics in the unperturbed equilibrium state.

Generally the linear response operator can be computed through correlation functions in the unperturbed stationary climate through the Fluctuation Dissipation Theorem (FDT) (Majda et al (2005))

$$\mathscr{R}_A(t) = \langle A(\mathbf{u}(t)) B(\mathbf{u}(0))\rangle, \quad B(\mathbf{u}) = -\frac{\mathrm{div}_{\mathbf{u}}\left(\mathbf{w}\pi_{\mathrm{eq}}\right)}{\pi_{\mathrm{eq}}}. \tag{8}$$

The problem in calculating the leading order response using (7) and (8) is that the equilibrium distribution π_{eq} is expensive to calculate for general systems. A variety of Gaussian approximations for π_{eq} (Leith (1975); Gritsun and Branstator (2007); Gritsun et al (2008); Bell (1980); Carnevale et al (1991)) and improved algorithms (Abramov and Majda (2007, 2008, 2009); Bell (1980); Majda et al (2010c); Majda and Wang (2006); Majda et al (2005)) have been developed for response via FDT and a rigorous proof of its validity is available (Hairer and Majda (2010)). FDT can have high skill for the mean response and some skill for the variance response for a wide variety of turbulent dynamical systems (Gritsun and Branstator (2007); Gritsun et al (2008); Abramov and Majda (2007, 2008, 2009); Majda et al (2010b,c, 2005)). One strategy to approximate the linear response operator which avoids direct evaluation of π_{eq} through the FDT formula is through the *kicked response* of an unperturbed system to a perturbation $\delta\mathbf{u}$ of the initial state from the equilibrium measure, that is,

$$\pi\,|_{t=0} = \pi_{\mathrm{eq}}(\mathbf{u} - \delta\mathbf{u}) = \pi_{\mathrm{eq}} - \delta\mathbf{u}\cdot\nabla\pi_{\mathrm{eq}} + O\left(\delta^2\right). \tag{9}$$

One important advantage of adopting this kicked response strategy is that higher order statistics due to nonlinear dynamics will not be ignored (compared with other linearized strategy using only Gaussian statistics Majda et al (2010a)). The kicked response theory gives the following fact (Majda et al (2005); Majda and Gershgorin (2011b)) for calculating the linear response operator:

Fact 3: For δ small enough, the linear response operator $\mathscr{R}_A(t)$ can be calculated by solving the unperturbed system (3) with a perturbed initial distribution in (9). Therefore, the linear response operator can be achieved through

$$\delta\mathscr{R}_A(t) \equiv \delta\mathbf{u}\cdot\mathscr{R}_A = \int A(\mathbf{u})\,\delta\pi' + O\left(\delta^2\right). \tag{10}$$

Here $\delta\pi'$ is the resulting leading order expansion of the transient density function from unperturbed dynamics using initial value perturbation. The Monte Carlo algorithm to approximate (10) is sketched in Appendix B.

2.3 Dynamical calibration strategy for model improvement

As hinted by Fact 2 and Fact 3, the prediction skill of imperfect models can be improved by comparing the information distance through the linear response operator with the true model. The following fact offers a convenient way to measure the lack of information in the perturbed imperfect model requiring only knowledge of linear responses for the mean and variance $\delta\bar{\mathbf{u}} \equiv \delta\mathscr{R}_{\mathbf{u}}, \delta R \equiv \delta\mathscr{R}_{(\mathbf{u}-\bar{\mathbf{u}})^2}$. For this result, it is important to tune the imperfect model to satisfy equilibrium model fidelity (Majda and Gershgorin (2011a,b)) in the first place, $\mathscr{P}\left(\pi_G, \pi_G^M\right) = 0$.

Fact 4: Under assumptions with diagonal covariance matrices $R = \mathrm{diag}(R_k)$ and equilibrium model fidelity $\mathscr{P}\left(\pi_G, \pi_G^M\right) = 0$, the relative entropy in (2) between perturbed model density π_δ^M and the true perturbed density π_δ with small perturbation δ can be expanded componentwisely as

$$\begin{aligned}\mathscr{P}\left(\pi_\delta, \pi_\delta^M\right) = {} & \mathscr{S}\left(\pi_{G,\delta}\right) - \mathscr{S}\left(\pi_\delta\right) \\ & + \frac{1}{2}\sum_k \left(\delta\bar{u}_k - \delta\bar{u}_{M,k}\right) R_k^{-1}\left(\delta\bar{u}_k - \delta\bar{u}_{M,k}\right) \\ & + \frac{1}{4}\sum_k R_k^{-2}\left(\delta R_k - \delta R_{M,k}\right)^2 + O\left(\delta^3\right).\end{aligned} \tag{11}$$

Here in the first line $\mathscr{S}\left(\pi_{G,\delta}\right) - \mathscr{S}\left(\pi_\delta\right)$ is the intrinsic error from Gaussian approximation of the system. R_k is the equilibrium 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. Proof of this result can be found in Majda and Gershgorin (2011b); Majda et al (2002).

The above facts about empirical information theory and linear response theory together provide a convenient and unambiguous way of improving the performance of imperfect models in terms of increasing their model sensitivity regardless of the specific form of external perturbations $\delta\mathbf{f}'$. The formula (7) in Fact 2 as well as (6) illustrates that the skill of an imperfect model in predicting forced changes to perturbations with general external forcing is directly linked to the model's skill in estimating the linear response operators $\mathscr{R}_A$ for the mean and variances (that is, use the functional $A = \mathbf{u}, (\mathbf{u}-\bar{\mathbf{u}})^2$) in a suitably weighted fashion as dictated by information theory (11). This offers us useful hints of training imperfect models for optimal responses for the mean and variance in a universal sense. From the linear response theory in Section 2.2, it shows that the system's responses to various external perturbations can be approximated by a convolution with the linear response operator $\mathscr{R}_A$ (which is only related to the statistics in the unperturbed equilibrium statistics). It is reasonable to claim that an imperfect model with precise prediction of this linear response operator should possess uniformly good sensitivity to different kinds of perturbations. On the other hand, the response operator can be calculated easily by the transient state distribution density function using the kicked response formula as in (10). Considering all these good features of the linear response operator, information barrier due to model sensitivity to perturbations can be overcome by minimizing the information error in the imperfect model kicked response distribution relative to the true response from observation data (Majda and Gershgorin (2011b)).

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

$$\left\|\mathscr{P}\left(\pi_\delta, \pi_\delta^{M^*}\right)\right\|_{L^1([0,T])} = \min_{M\in\mathscr{M}} \left\|\mathscr{P}\left(\pi_\delta, \pi_\delta^M\right)\right\|_{L^1([0,T])}, \tag{12}$$

where π_δ^M can be achieved through a kicked response procedure (10) in the training phase compared with the actual observed data π_δ in nature, and the information distance between perturbed responses $\mathscr{P}\left(\pi_\delta, \pi_\delta^M\right)$ can be calculated with ease through the expansion formula (11). The information distance $\mathscr{P}\left(\pi_\delta(t), \pi_\delta^M(t)\right)$ is measured

at each time instant, so the entire error is averaged under the L^1-norm inside a proper time window $[0,T]$. Some low dimensional examples of this procedure for turbulent systems can be found in Branicki et al (2013); Branicki and Majda (2012); Majda and Branicki (2012).

3 L-96 system as a test bed and its statistical dynamics

Now we consider turbulent dynamical systems with state variables $\mathbf{u} \in \mathbb{R}^N$ in the abstract form with quadratic nonlinearity as

$$\mathbf{u}_t = \mathbf{L}\mathbf{u} + \mathbf{B}(\mathbf{u},\mathbf{u}) + \mathbf{F}, \tag{13}$$

where the linear operator $\mathbf{L}$ represents the linear effects in the system, $\mathbf{F}$ is the external forcing term (which can either be deterministic or include random effects), and $\mathbf{B}(\mathbf{u},\mathbf{u})$ is a bilinear form representing quadratic nonlinear interactions which conserves energy $\mathbf{u}\cdot\mathbf{B}(\mathbf{u},\mathbf{u}) = 0$. The structure in (13) for turbulent dynamical systems can be found in many applications in geosciences and other areas of engineering (Majda (2003); Majda et al (2005); Majda and Wang (2006); Branicki et al (2013)).

Even though this system (13) may only be driven by a deterministic forcing $\mathbf{F}$, uncertainties can still be introduced by different degrees of internal instabilities as well as errors from the initial condition. To quantify these uncertainties, we are interested in resolving the statistical features of this dynamical system, especially the first two order moments. By a proper choice of basis $\{\mathbf{v}_k\}$, the state variables $\mathbf{u}(t)$ can be decomposed into the statistical mean state and the fluctuations along each direction $\mathbf{v}_k$ as

$$\mathbf{u}(t) = \bar{\mathbf{u}}(t) + \sum_k Z_k(t)\,\mathbf{v}_k. \tag{14}$$

Here $\bar{\mathbf{u}}(t)$ can be viewed as the statistical ensemble mean at each time instant, while Z_k represents the (complex) stochastic coefficient measuring the uncertainty in each direction $\mathbf{v}_k$. The dynamical equations for the statistical mean and covariance matrix can be derived as (Sapsis and Majda (2013d))

$$A) \qquad \frac{d\bar{\mathbf{u}}}{dt} = \mathbf{L}(t)\,\bar{\mathbf{u}} + \mathbf{B}(\bar{\mathbf{u}},\bar{\mathbf{u}}) + \sum_{i,j} R_{ij}\mathbf{B}(\mathbf{v}_i,\mathbf{v}_j) + \mathbf{F}(t), \tag{15a}$$

$$B) \qquad \frac{dR}{dt} = L_v R + R L_v^* + Q_F + Q_\sigma. \tag{15b}$$

with

$$R_{ij}(t) = \left\langle Z_i Z_j^*(t)\right\rangle,$$

$$L_{v,ij} = (\mathbf{L}(t)\,\mathbf{v}_j + \mathbf{B}(\bar{\mathbf{u}},\mathbf{v}_j) + \mathbf{B}(\mathbf{v}_j,\bar{\mathbf{u}}))\cdot\mathbf{v}_i^*,$$

$$Q_{F,ij} = \sum_{m,n}\left\langle Z_m Z_n^* Z_j\right\rangle \mathbf{B}(\mathbf{v}_m,\mathbf{v}_n)^*\cdot\mathbf{v}_i + \langle Z_m^* Z_n Z_i^*\rangle\,\mathbf{B}(\mathbf{v}_m,\mathbf{v}_n)\cdot\mathbf{v}_j^*,$$

and Q_σ (possibly 0) from the random component of $\mathbf{F}$. The quasilinear operator L_v describes the linear effects (due to the factor $\mathbf{L}$) and the energy transfers between the mean $\bar{\mathbf{u}}$ and each mode $\mathbf{v}_k$ (due to $\mathbf{B}$). Instabilities come from the directions with positive eigenvalues (that is, positive Lyapunov coefficients) of L_v. The increase in energy due to the unstable modes is balanced by the nonlinear energy transfer from the nonlinear flux term Q_F. Importantly, the term Q_F satisfies $\mathrm{tr}Q_F = 0$ as a consequence of energy conservation for the nonlinear term $\mathbf{u}\cdot\mathbf{B}(\mathbf{u},\mathbf{u}) = 0$, and acts as a dissipative mechanism for the unstable modes and external noise for the stable modes bringing all of them into a statistical equilibrium state (Sapsis and Majda (2013a,b,d,c)). Note that third order moments $\left\langle Z_m Z_n^* Z_j\right\rangle$ are included in Q_F coming from the quadratic interactions between modes $\mathbf{B}(\mathbf{v}_m,\mathbf{v}_n)$, where we should take the conjugate of the second coefficient component with complex cases included. And to make the matrix unitary $Q_F^* = Q_F$, the conjugate part is added in the formula.

This structure in (15a) and (15b) is ubiquitous in a variety of turbulent systems, and in general the state variable $\mathbf{u}$ lives in a high dimensional phase space roughly of order 10,000 or larger. However a genuinely high dimensional system containing all kinds of complex structures becomes too difficult to analyze. Instead in the first place, we would rather focus on the dominant core dynamical features, therefore it is useful to begin with simpler systems focusing on the key mechanism. For illustration, the Lorenz 96 (L-96) system is the simplest but nevertheless representative paradigm of a complex turbulent dynamical system possessing properties found in realistic turbulent systems such as, energy-preserving advection, damping and forcing, a large number of persistent instabilities, and strong nonlinear energy transfers between modes. In the following part of this paper, we will focus on this simplified system and check the model improvement strategies based on this system.

3.1 Homogeneous statistical dynamics of L-96 system

3.1.1 General moment equations for L-96 system

The L-96 system is a 40-dimensional dynamical system with state variables $\mathbf{u} = (u_0, u_1, ..., u_{J-1})^{\mathrm{T}}$ such that

$$\frac{du_j}{dt} = \left(u_{j+1} - u_{j-2}\right) u_{j-1} - d_j(t)\, u_j + F_j(t), \quad j = 0, 1, ..., J-1,\ J = 40. \tag{16}$$

Periodic boundary condition $u_J = u_0$ is applied. The model is designed to mimic baroclinic turbulence in the midlatitude atmosphere with the effects of energy conserving nonlinear advection and dissipation. By changing the amplitude of the external forcing F_j, the system shows a wide range of different dynamical regimes ranging from weakly chaotic, strongly chaotic, to finally full turbulence with varying statistics, which makes it a desirable test model. More detailed discussion about the L-96 system in various dynamical regimes can be found in (Abramov and Majda (2007); Lorenz (1996); Majda and Wang (2006)).

To compare with the abstract form (13) above, we can write the linear operator for L-96 system as

$$\mathbf{L}(t) = \mathrm{diag}\left(-d_0(t), \cdots, -d_{J-1}(t)\right),$$

and define the quadratic form as

$$\mathbf{B}(\mathbf{u}, \mathbf{v}) = \left\{u^*_{j-1}\left(v_{j+1} - v_{j-2}\right)\right\}_{j=0}^{J-1}.$$

As we will see in the next part with homogeneous assumption, it is convenient to express the equations in Fourier domain. Choose the Fourier basis as $\{\mathbf{v}_k\}_{k=-J/2+1}^{J/2}$ with

$$\mathbf{v}_k = \left\{\frac{1}{\sqrt{J}} e^{2\pi \imath k \frac{j}{J}}\right\}_{j=0}^{J-1},$$

and the decomposition as in (14) can be expressed under the Fourier basis

$$\mathbf{u}(t) = \bar{\mathbf{u}}(t) + \sum_{k=-J/2+1}^{J/2} Z_k(t)\, \mathbf{v}_k, \qquad Z_{-k} = Z_k^*.$$

First, the dynamical equation for the mean state can be achieved through taking ensemble average over both sides of (16). It needs to be noticed that second order moments $r_{mn} = \langle Z_m Z_n^* \rangle$ will come into the mean dynamics due to the quadratic form $\mathbf{B}$

$$\frac{d\bar{u}_j}{dt} = -d_j(t)\, \bar{u}_j + \bar{u}_{j-1}\left(\bar{u}_{j+1} - \bar{u}_{j-2}\right) + \frac{1}{J}\sum_{m,n} r_{mn} e^{-2\pi \imath m \frac{j-1}{J}} \left(e^{2\pi \imath n \frac{j+1}{J}} - e^{-2\pi \imath n \frac{j-2}{J}}\right) + F_j(t). \tag{17}$$

Then, subtracting equation (17) from the original dynamics (16), and projecting the resulting fluctuation parts to each orthonormal Fourier basis $\mathbf{v}_k$, we can get the dynamical equations for each stochastic coefficient Z_k. Finally the dynamics for the second order moments r_{ij} can be calculated by multiplying Z_k on both sides of the coefficient equation and taking expectations, therefore

$$\frac{dr_{ij}}{dt} = \left(L_v R + L_v^* R\right)_{ij} + Q_{F,ij}, \tag{18}$$

with $R = (r_{ij})$. The quasilinear interaction L_v and nonlinear flux term Q_F for L-96 can be derived in the form with δ_{ij} the standard Kronecker delta function

$$\begin{aligned} L_{v,ij} = &-d_j(t)\, \delta_{ij} + \frac{1}{J}\left(e^{2\pi \imath \frac{j}{J}} - e^{-2\pi \imath \frac{2j}{J}}\right) \sum_m \bar{u}_{m-1} e^{2\pi \imath m \frac{j-i}{J}} \\ &+ \frac{1}{J} e^{2\pi \imath \frac{j}{J}} \sum_m \left(\bar{u}_{m+1} - \bar{u}_{m-2}\right) e^{-2\pi \imath \frac{j+i}{J}}. \end{aligned}$$

$$\begin{aligned} Q_{F,ij} = &\frac{1}{\sqrt{J}} \sum_{m,n} \left\langle Z_m Z_n^* Z_j \right\rangle e^{-2\pi \imath \frac{m}{J}} \left(e^{-2\pi \imath \frac{n}{J}} - e^{2\pi \imath \frac{2n}{J}}\right) \delta_{n-m,i} \\ &+ \frac{1}{\sqrt{J}} \sum_{m,n} \left\langle Z_m^* Z_n Z_i^* \right\rangle e^{2\pi \imath \frac{m}{J}} \left(e^{2\pi \imath \frac{n}{J}} - e^{-2\pi \imath \frac{2n}{J}}\right) \delta_{n-m,j}. \end{aligned}$$

We neglect the tedious calculations for these dynamical equations and put the details in Appendix A.

Remark. 1. Under the discrete Fourier basis $\{\mathbf{v}_k\}_{k=-J/2+1}^{J/2}$, we have $J-2$ complex modes which are conjugate in pairs, $\mathbf{v}_{-k}=\mathbf{v}_k^*$; and the other two modes $\mathbf{v}_0$, $\mathbf{v}_{J/2}$ contain only real parts. Here and after we will always use the convention of J even with one more positive Fourier mode $k=J/2$ in the spectra.

2. Note that the equations (17), (18) are still not closed since third order moments are included in the dynamics for the covariance in the term $Q_{F,ij}$. Further if we go on calculating the dynamical equations for the third order moments, forth order moments will again enter the equations. As a result the equations will never be closed under this process by calculating the dynamical equations for each order moments.

3. The quasilinear operator L_v also includes the nonlinear effects due to the nonlinear interactions between the mean state $\bar{\mathbf{u}}$ and the Fourier modes $\mathbf{v}_j$ besides all the linear damping and dissipation. For systems whose nonlinearity largely comes from these kinds of interactions, quasilinear closure models only including L_v in the covariance equations (18) can perform well with good prediction skill (Branicki and Majda (2012); Majda and Branicki (2012)).

3.1.2 Moment equations with homogeneous assumption

Still the mean and covariance equations (17) and (18) derived above are cumbersome for both analyzing the statistical properties and applying for closure models. As a further simplification, we assume uniform damping and forcing for the L-96 system (16). That is, the damping and forcing terms $d(t)$ and $F(t)$ are assumed only functions of time and stay the same value at different grid points j. With this assumption and noting that the quadratic term $\left(u_{j+1}-u_{j-2}\right)u_{j-1}$ is also translation invariant in space, homogeneous solutions can be generated in this case. Specifically, the statistics of state variables $\mathbf{u}$ become homogeneous meaning invariant under translation over the entire duration of the process

$$\langle u_{i_1}u_{i_2}\cdots u_{i_n}\rangle=\left\langle u_{i_1+l}u_{i_2+l}\cdots u_{i_n+l}\right\rangle,\quad \forall\, l\in\mathbb{N},$$

with periodic boundary condition $u_J=u_0$ and spatially homogeneous initial conditions assumed. Under the homogeneous assumption, Fourier basis becomes eigenfunctions (or the EOFs) for the L-96 operator, and the first three moments can be further simplified as (derivation in Appendix A)

$$\bar{\mathbf{u}}(t)=\bar{u}(t)(1,1,...,1)^T, \tag{19a}$$

$$R(t)=\mathrm{diag}\left(r_{-J/2+1}(t),\cdots,r_0(t),...,r_{J/2}(t)\right), \tag{19b}$$

$$\left\langle Z_iZ_jZ_k\right\rangle\neq 0,\quad \text{only if } i+j+k=0. \tag{19c}$$

That is, we only need to focus on the scalar mean $\bar{u}$ and the variances r_k along each eigen-direction $\mathbf{v}_k$.

Under these properties, the stochastic coefficients Z_k from (14) satisfy a simplified dynamics

$$\frac{dZ_k}{dt}=-d(t)Z_k+\left(e^{2\pi i\frac{k}{J}}-e^{-2\pi i\frac{2k}{J}}\right)\bar{u}Z_k+\frac{1}{\sqrt{J}}\sum_{m=-J/2+1}^{J/2}Z_{k+m}Z_m^*\left(e^{2\pi i\frac{2m+k}{J}}-e^{-2\pi i\frac{m+2k}{J}}\right). \tag{20}$$

Following the same procedure as before, moment equations in (17) and (18) for the mean and covariance matrix can be then simplified under homogeneous assumption simplifications in (19a)-(19c). Thus we arrive at the major equations of interest of this paper

Exact Low Order Moment Equations

$$\frac{d\bar{u}(t)}{dt}=-d(t)\bar{u}(t)+\frac{1}{J}\sum_{k=-J/2+1}^{J/2}r_k(t)\Gamma_k+F(t), \tag{21a}$$

$$\frac{dr_k(t)}{dt}=2\left[-\Gamma_k\bar{u}(t)-d(t)\right]r_k(t)+Q_{F,kk},\quad k=0,1,...,J/2. \tag{21b}$$

Note that we denote $\Gamma_k=\cos\frac{4\pi k}{J}-\cos\frac{2\pi k}{J}$, $r_{-k}=\left\langle Z_{-k}Z_{-k}^*\right\rangle=\left\langle Z_kZ_k^*\right\rangle=r_k$, and the nonlinear flux Q_F becomes diagonal

$$Q_{F,kk'}=\frac{2}{\sqrt{J}}\sum_m\mathfrak{Re}\left\{\langle Z_mZ_{-m-k}Z_k\rangle\left(e^{-2\pi i\frac{2m+k}{J}}-e^{2\pi i\frac{m+2k}{J}}\right)\right\}\delta_{kk'},$$

with energy conservation $\mathrm{tr}Q_F=0$.

Remark. Even though we derive the moment equations above from the L-96 system under homogeneous assumption for the sake of analysis and will focus on them in the following discussions, the equations (21a) and (21b) are actually quite representative and are easy to be extended to general nonlinear systems with conservative quadratic forms.

3.2 Properties of single point statistics and statistical energy conservation

Now we can focus on the simplified moment equations (21a) and (21b) and investigate the statistical properties inside this system. Note that still these moment equations above are not closed and not easy to solve directly. The most difficult and expensive part in solving the above system for the mean and covariance comes from evaluating the nonlinear flux term Q_F where higher order statistics are involved. Thus the central issue in developing closure models becomes to find proper approximation about the nonlinear flux term $Q_F^M \sim Q_F$ which can offer a statistically consistent estimation. Consideration about accuracy and computational efficiency should be balanced in determining the explicit form of Q_F^M, and we leave the detailed discussions about choosing and testing this term for numerical simulations in the next section.

Here in the first place, we check some statistical properties of the perfect and imperfect models, and as we will see in the next section they can serve as the guideline for the designing of proper approximation methods. Of particular interests in both theory and application, the statistical mean and variance at each individual grid point play an important role as key statistical quantities to predict. In this section, we focus on this *single point mean and variance* of the system which ignore the cross-correlation between different grid points. With homogeneous assumption of the system as described above, the moments at each grid point become translation invariance. Therefore the single point mean $\bar{u}_{1\text{pt}}$ and variance $R_{1\text{pt}}$ can be defined by averaging each Fourier mode, that is,

$$\bar{u}_{1\text{pt}} = \frac{1}{J}\sum_{j=0}^{J-1} u_j = \bar{u}, \quad R_{1\text{pt}} = \frac{1}{J}\sum_{k=-J/2+1}^{J/2} r_k = \frac{1}{J}\text{tr}R. \tag{22}$$

Furthermore, as we will see in the following discussion, the conservation of energy in the nonlinear flux Q_F plays an important role in model prediction for single point statistics. Thus we define

$$\text{symmetry of nonlinear energy} \Leftrightarrow \text{tr}Q_F = 0. \tag{23}$$

With all the above definitions, we can claim the single point statistics consistency: for any closure Q_F^M satisfying dynamics with the same statistical symmetry in (23), the equilibrium consistency for single point statistics $\bar{u}_{M,\infty} = \bar{u}_\infty, \text{tr}R_{M,\infty} = \text{tr}R_\infty$ can be arranged by only tuning for the mean state in equilibrium. Furthermore, the model response of the single point variance $\delta\text{tr}R_M$ to perturbations can be predicted accurately if we predict the response of the mean $\delta\bar{u}_M$ accurately.

The claim can be seen by simple manipulations of equations (21a) and (21b). By multiplying $\bar{u}$ on both sides of the mean equation (21a), we get

$$\frac{d\bar{u}^2}{dt} = -2d\bar{u}^2 + 2\bar{u}F + \frac{2}{J}\sum_k \Gamma_k r_k \bar{u}.$$

And by summing up all the modes in the variance equation (21b)

$$\frac{d\text{tr}R}{dt} = 2\left(-\sum_k \Gamma_k r_k \bar{u}\right) - 2d\text{tr}R + \text{tr}Q_F.$$

It is convenient to define the *statistical energy* including both mean and total variance as

$$E(t) = \frac{J}{2}\bar{u}^2 + \frac{1}{2}\text{tr}R = \frac{J}{2}\left(\bar{u}_{1\text{pt}}^2 + R_{1\text{pt}}\right). \tag{24}$$

With this definition the corresponding dynamical equation for the statistical energy E of the true system can be easily derived as

$$\frac{dE}{dt} = -2dE + JF\bar{u} + \frac{1}{2}\text{tr}Q_F = -2dE + JF\bar{u}, \tag{25}$$

with symmetry of nonlinear energy conservation (23) assumed. Correspondingly, if we have some imperfect model with approximated nonlinear flux Q_F^M in (21b), with a similar process we can derive E_M from the imperfect model as

$$\frac{dE_M}{dt} = -2dE_M + JF\bar{u}_M + \frac{1}{2}\text{tr}Q_F^M. \tag{26}$$

First considering the statistical stationary state, the left hand side of (25) and (26) vanishes and under a weaker constraint for imperfect nonlinear energy conservation in equilibrium $\text{tr}Q_{F,\infty}^M = 0$ we have

$$E_\infty = -\frac{JF}{2d}\bar{u}_\infty, \quad E_{M,\infty} = -\frac{JF}{2d}\bar{u}_{M,\infty}.$$

It is direct conclusion that we can achieve equilibrium consistency for single point variance $\mathrm{tr}R_{M,\infty} = \mathrm{tr}R_\infty$ once that we have the consistency for the mean $\bar{u}_{M,\infty} = \bar{u}_\infty$. Next considering the model responses to perturbations, we can write the solution for (25) and (26) formally as

$$\begin{aligned} E(t) &= E_0 \exp\left(-2\int_{t_0}^t d(s)\,ds\right) + J\int_{t_0}^t \exp\left(-2\int_s^t d(\tau)\,d\tau\right) F(s)\,\bar{u}(s)\,ds, \\ E_M(t) &= E_{M,0}\exp\left(-2\int_{t_0}^t d(s)\,ds\right) + J\int_{t_0}^t \exp\left(-2\int_s^t d(\tau)\,d\tau\right) F(s)\,\bar{u}_M(s)\,ds \\ &\quad + \frac{1}{2}\int_{t_0}^t \exp\left(-2\int_s^t d(\tau)\,d\tau\right) \mathrm{tr}Q_F^M(s)\,ds. \end{aligned}$$

The last part in the right hand side of E_M comes from the error in the approximation for nonlinear flux Q_F^M. Use the previous asymptotic expansion with perturbation, and assume the same equilibrium statistics and initial distribution

$$E_\delta = E_\infty + \delta E, \quad E_{\delta,M} = E_\infty + \delta E_M, \quad E_0 = E_{M,0}.$$

Therefore the error in the response to perturbations can be estimated as

$$\delta E - \delta E_M = J\int_{t_0}^t \exp\left(-2\int_s^t d(\tau)\,d\tau\right) F(s)\,(\delta\bar{u} - \delta\bar{u}_M)(s)\,ds - \frac{1}{2}\int_{t_0}^t \exp\left(-2\int_s^t d(\tau)\,d\tau\right)\mathrm{tr}Q_F^M(s)\,ds.$$

$$\|\delta E - \delta E_M\| \le \tilde{C}_0\|\delta\bar{u} - \delta\bar{u}_M\| + C_1\left\|\mathrm{tr}Q_F^M\right\|. \tag{27}$$

Also through the definition of the statistical energy $E = \frac{J}{2}\bar{u}^2 + \frac{1}{2}\mathrm{tr}R$, we have

$$\|\delta E\| \ge \|\delta\mathrm{tr}R\| - J\|\bar{u}\|\,\|\delta\bar{u}\|. \tag{28}$$

We get the error estimation for the single point variance through the error from the mean and nonlinear flux by combining (27) and (28)

$$\|\delta\mathrm{tr}R - \delta\mathrm{tr}R_M\| \le C_0\|\delta\bar{u} - \delta\bar{u}_M\| + C_1\left\|\mathrm{tr}Q_F^M\right\|, \tag{29}$$

with C_0, C_1 constants. Usually by construction of imperfect models we need to require at least $\left\|\mathrm{tr}Q_F^M\right\| \sim O(\delta) \ll 1$, the error in the statistical variance response can be then controlled by the error in the mean with a good approximation for the nonlinear flux term. In summary, the following proposition is achieved:

Proposition 1. *Consider a system with homogeneous statistical solution and an imperfect closure model with flux Q_F^M satisfying symmetry of nonlinear energy conservation in equilibrium $\mathrm{tr}Q_{F,\infty}^M = \mathrm{tr}Q_{F,\infty} = 0$. Then (i) Statistical equilibrium fidelity for one point statistics (22) of the variance $\mathrm{tr}R_{M,\infty} = \mathrm{tr}R_\infty$ is satisfied if consistent mean state $\bar{u}_{M,\infty} = \bar{u}_\infty$ is achieved at equilibrium; (ii) Furthermore, with equilibrium consistency for the mean and variance, the error in the response for single point variance $\left\|\mathrm{tr}R_{\delta,M} - \mathrm{tr}R_\delta\right\|$ can be bounded by the error from the response for the mean $\left\|\bar{u}_{\delta,M} - \bar{u}_\delta\right\|$, given an accurate approximation for the nonlinear flux $\left\|\mathrm{tr}Q_F^M\right\| \sim 0$.*

Remark. Proposition 1 states that we can achieve agreement in both equilibrium fidelity and responses to perturbations in single point statistics (and the total variance of the system) through tuning the models only for the statistical mean state. And also we can see from (29) that the model approximation for the nonlinear flux with energy conservation $\left\|\mathrm{tr}Q_F^M\right\| \sim 0$ plays an important role in model sensitivity to perturbations. In Section 4 we will illustrate these crucial aspects in designing approximation models with a hierarchy of closure methods.

3.3 Information barrier in single point statistics

Proposition 1 above shows that we can achieve both equilibrium consistency and sensitivity in responses in single point statistics by tuning at most one parameter of the imperfect model. However, pointwise statistics by only considering the variance at each grid point, and ignoring the correlations between different grids may not be sufficient for accurate model predictions. In this section, we display that single point simplification is not enough for desirable model performance by measuring the information barrier (Majda and Gershgorin (2011a,b); Majda and Branicki (2012)) with this simplification despite the fact that practitioners in climate science have proposed such a strategy (DelSole and Shukla (2010)).

Here in this section, we generalize the system a little to a vector field $\mathbf{u}_j \in \mathbb{R}^n$ at each grid point rather than a simpler scalar field as in L-96. So we use R_j rather than the previous r_j to represent the covariance matrix at each grid point, and ignore the cross-covariance between different grids in the single point simplification. Let the density function from the true model be $\pi(\mathbf{u})$ as before, and consider imperfect models where we only measure pointwise

marginal distribution $\pi^M_{1\text{pt}}(\mathbf{u}_j) \equiv \pi^M_{1\text{pt},j}$ at each grid point $j = 0, 1, \cdots, J-1$. Construct the probability density function with only single point statistics from the marginal distribution as $\pi^M_{1\text{pt}} = \prod_{j=0}^{J-1} \pi^M_{1\text{pt},j}$. By Proposition 4.1 of Majda et al (2002), the information distance between the truth and imperfect model prediction has the form

$$\mathscr{P}\left(\pi, \pi^M_{1\text{pt}}\right) = \left[\mathscr{S}(\pi_G) - \mathscr{S}(\pi)\right] + \mathscr{P}\left(\pi_G, \prod_{j=0}^{J-1} \pi^G_{1\text{pt},j}\right) + \sum_{j=0}^{J-1} \mathscr{P}\left(\pi^G_{1\text{pt},j}, \pi^M_{1\text{pt},j}\right), \tag{30}$$

with $\pi^G_{1\text{pt},j} = \mathscr{N}(\bar{\mathbf{u}}_j, R_j)$. The first part on the right hand side of (30) is the intrinsic information barrier in Gaussian approximation. And the third part with homogeneous assumption of the system can be written as

$$\sum_{j=0}^{J-1} \mathscr{P}\left(\pi^G_{1\text{pt},j}, \pi^M_{1\text{pt},j}\right) = J\mathscr{P}\left(\pi^G_{1\text{pt}}, \pi^M_{1\text{pt}}\right).$$

Proposition 1 tells that the mean and single point statistics can be approximated with accuracy in the imperfect models. Therefore this part will also vanish (or at least be minimized) since only $\bar{u}$ and tr$R$ are included in $\pi_{1\text{pt}}$. The error from single point approximation (and ignoring the cross-covariance) then comes only from the information barrier in marginal approximation as shown in the second part on the right hand side of (30). To get the lack of information in the marginal distribution, it is useful to introduce the random field representations for both the Gaussian approximation and the uncorrelated spatial random field with correct one-point statistics (Yaglom (2004))

$$\mathbf{u}^G = \bar{\mathbf{u}} + \sum_{k=-J/2+1}^{J/2} R_k^{1/2} e^{ikx_j} \hat{\mathbf{W}}_k, \tag{31}$$

$$\mathbf{u}^G_{1\text{pt}} = \bar{\mathbf{u}} + \sum_{k=-J/2+1}^{J/2} \left(\frac{\sum_{j=0}^{J-1} R_j}{J}\right)^{1/2} e^{ikx_j} \hat{\mathbf{W}}_k. \tag{32}$$

Note here we assume both Gaussian distributions for $\mathbf{u}^G$ and $\mathbf{u}^G_{1\text{pt}}$ with the same pointwise mean $\bar{\mathbf{u}}$ in (31) and (32), so only the dispersion part in (2) is non-zero when comparing the information distance. Using the notation $R_{1\text{pt}} = \frac{\sum_{j=0}^{J-1} R_j}{J}$, the information barrier due to single point statistics simplification becomes

$$\begin{aligned}
\mathscr{P}\left(\pi_G, \prod_{j=0}^{J-1} \pi^G_{1\text{pt},j}\right) &= \sum_{k=-J/2+1}^{J/2} \left[-\log\det\left(R_k R_{1\text{pt}}^{-1}\right) + \text{tr}\left(R_k R_{1\text{pt}}^{-1} - I\right)\right] \\
&= -\sum_{k=-J/2+1}^{J/2} \log\det\left(R_k R_{1\text{pt}}^{-1}\right) + \text{tr}\left[\sum_{k=-J/2+1}^{J/2} \left(R_k R_{1\text{pt}}^{-1} - I\right)\right] \\
&= -\log\left(\prod_{k=-J/2+1}^{J/2} \frac{\det R_k}{\det R_{1\text{pt}}}\right) \\
&= J\log\left[\frac{\det\left(\sum_{j=0}^{J-1} R_j/J\right)}{\left(\prod_{j=0}^{J-1} \det R_j\right)^{1/J}}\right].
\end{aligned}$$

The second equality just applies the definition of $R_{1\text{pt}}$ so that $\sum_{k=-J/2+1}^{J/2} \left(R_k R_{1\text{pt}}^{-1} - I\right) = 0$. Then the following proposition can be achieved via this decomposition:

Proposition 2. *The information barrier between the Gaussian random field (31) and the uncorrelated one point statistics random field (32) is given by*

$$\mathscr{P}\left(\pi_G, \prod_{j=0}^{J-1} \pi^G_{1\text{pt}}\right) = J\log\left[\frac{\det\left(\sum_{j=0}^{J-1} R_j/J\right)}{\left(\prod_{j=0}^{J-1} \det R_j\right)^{1/J}}\right]. \tag{33}$$

Remark. Jensen's inequality guarantees the non-negativeness of (33) knowing that $\log\det(\cdot)$ is concave as an independent check. The information distance in (30) vanishes only for the case with equipartition of energy, *i.e.* $r_j = \frac{\text{tr}R}{J}$. For general situations, always we need to face situations like $\det R \ll \text{tr}R$. This will end up with large information barrier when only one point statistics are considered in the imperfect model.

Finally in order to offer a clearer illustration about the possible information barrier from Proposition 2, we estimate the lower bound of the error in (33). Returning to the scalar case, note that the ratio on the right hand side of (33) is between the arithmetic mean $A=\sum_j r_j/J=R_{1\text{pt}}$ and geometric mean $G=\left(\prod_j r_j\right)^{1/J}$ of a positive sequence $r_0,\cdots,r_{J-1}$. A nice lower bound for the difference between the arithmetic and geometric mean can be estimated (Tung (1975)) as

$$0\le J^{-1}\left(\sigma_{\max}-\sigma_{\min}\right)^2\le A-G, \tag{34}$$

where $\sigma_{\max}^2=\max\{r_k\}$, $\sigma_{\min}^2=\min\{r_k\}$ are the largest and smallest variances. Therefore we have the estimation for the ratio

$$\frac{A}{G}-1\ge\frac{\left(\sigma_{\max}-\sigma_{\min}\right)^2}{JG}\ge\frac{\left(\sigma_{\max}-\sigma_{\min}\right)^2}{JA}.$$

Substituting this into the information distance in (33) yields the bound

$$\mathscr{P}\left(\pi_G,\prod_{j=0}^{J-1}\pi_{1\text{pt}}^G\right)=J\log\frac{A}{G}\ge\log\left(1+\frac{\left(\sigma_{\max}-\sigma_{\min}\right)^2}{JA}\right)^J\to R_{1\text{pt}}^{-1}\left(\sigma_{\max}-\sigma_{\min}\right)^2,\quad \text{as}J\to\infty.$$

This shows that for high dimensional systems the information barrier due to single point approximation can be measured between the largest and smallest variance in the energy spectra

$$\mathscr{P}\left(\pi_G,\prod_{j=0}^{J-1}\pi_{1\text{pt}}^G\right)\sim O\left(\left(\sigma_{\max}-\sigma_{\min}\right)^2\right). \tag{35}$$

In general, this gap between the largest and smallest variance (or standard deviation) could become quite large considering the common rapidly decaying energy spectra in turbulent systems. This information barrier can only be broken by introducing more careful calibration about the dynamics in each eigen-direction of the system individually. See Figure 2 for an example.

4 Statistical closure methods in full phase space

In this section, we begin to develop statistical closure models for uncertainty quantification with consideration in both accuracy and computation efficiency according to the statistical properties discussed previously. As shown in Section 3.3, large information barrier in (33) may still exist if we only tune the statistical mean for consistency in single point statistics. To overcome such barriers using the closure models, more precise calibration about the nonlinear flux Q_F accounting for more degrees of freedom must be used. Starting with the simplest possible closure ideas, our goal here is to compare the advantages and limitations of different levels of imperfect models and check how the theories from Section 2 and 3 can help with improving the model prediction skill, especially the model sensitivity to various perturbations.

We may consider the closure ideas by taking another look at the dynamics for stochastic coefficients in (20)

$$\frac{dZ_k}{dt}=-d\left(t\right)Z_k+\left(e^{2\pi i\frac{k}{J}}-e^{-2\pi i\frac{2k}{J}}\right)\bar{u}Z_k+\frac{1}{\sqrt{J}}\sum_{m=-J/2+1}^{J/2}Z_{k+m}Z_m^*\left(e^{2\pi i\frac{2m+k}{J}}-e^{-2\pi i\frac{m+2k}{J}}\right).$$

Major nonlinearity comes from the last term above representing interactions between different modes. The basic idea here is to model the effect of the nonlinear energy transfers on each mode by adding additional damping $d_{M,k}$ balancing the linearly unstable character of these modes, and adding additional (white) stochastic excitation which will model the energy received by the stable modes (Sapsis and Majda (2013d)). We want to constrain ourselves to second order models considering computational expense, hence the additional parts $d_{M,k},\sigma_{M,k}$ only include statistics up to second order moments. Specifically we replace this nonlinear term by

$$\frac{1}{\sqrt{J}}\sum_{m=-J/2+1}^{J/2}Z_{k+m}Z_m^*\left(e^{2\pi i\frac{2m+k}{J}}-e^{-2\pi i\frac{m+2k}{J}}\right)\to-d_{M,k}\left(R\right)Z_k+\sigma_{M,k}\left(R\right)\dot{W}_k,$$

with $R=\left\langle Z_kZ_k^*\right\rangle$. Accordingly, the mean and variance dynamics for the (imperfect) closure method can be derived with the same form as (21a) and (21b)

$$\frac{d\bar{u}_M\left(t\right)}{dt}=-d\left(t\right)\bar{u}_M\left(t\right)+\frac{1}{J}\sum_{k=-J/2+1}^{J/2}r_{M,k}\left(t\right)\Gamma_k+F\left(t\right), \tag{36a}$$

$$\frac{dr_{M,k}\left(t\right)}{dt}=2\left[-\Gamma_k\bar{u}_M\left(t\right)-d\left(t\right)\right]r_{M,k}\left(t\right)+Q_{F,kk}^M,\quad k=0,1,...,J/2. \tag{36b}$$

only with the nonlinear flux Q_F replaced by

$$Q^M_{F,kk} = Q^M_{F-,kk} + Q^M_{F+,kk} = -2d_{M,k}(R)\, r_{M,k} + \sigma^2_{M,k}(R). \tag{37}$$

Here $Q^M_{F-} = -2d_{M,k}(R)\, r_{M,k}$ represents the additional damping to stabilize the unstable modes with positive Lyapunov coefficients, while $Q^M_{F+} = \sigma^2_{M,k}(R)$ is the additional noise to compensate for the overdamped modes. Now the problem is converted to finding expressions for $d_{M,k}$ and $\sigma^2_{M,k}$. Following by gradually adding more detailed characterization about the statistical dynamical model we display the general procedure of constructing a hierarchy of the closure methods step by step.

4.1 Hierarchy of closure models

We denote the equilibrium states for the mean and variance with unperturbed uniform forcing as $\bar{u}_\infty \equiv \langle \bar{u} \rangle$, $r_{j,\infty} \equiv \langle r_j \rangle$. And with a little abuse of notation, let $d = \langle d(t) \rangle$, and $F = \langle F(t) \rangle$. Following step by step, we bring in more and more considerations in characterizing the uncertainties in each mode. Finally three different sets of closure methods with increasing complexity and accuracy in prediction skill will be proposed, illustrating one important statistical feature in each category.

1. *Quasilinear Gaussian closure model*: The simplest approximation for the closure methods (Epstein (1969)) at the first stage should be simply neglecting the nonlinear part entirely. That is, to set

$$d_{M,k}(R) \equiv 0, \quad \sigma^2_{M,k}(R) \equiv 0, \quad Q^{\mathrm{QG}}_F \equiv 0. \tag{38}$$

   Obviously this crude approximation will not work well due to the cutoff of the energy flow when strong nonlinear interactions between modes occur. Actually, it can be proved (Sapsis and Majda (2013d)) that for L-96 system with QG closure model, in final equilibrium state there exists only one active mode with critical wavenumber

$$k_{\mathrm{cr}} = \mathrm{argmax}_k \left(-\Gamma_k \bar{u}_{\mathrm{cr}} - d\right).$$

   And the critical mean and the only non-zero variance for $k = k_{\mathrm{cr}}$ always converge to the only critical point no matter what value F takes

$$\bar{u}_{\mathrm{cr}} = -\max_k \left(d/\Gamma_k\right), \qquad r_{k_{\mathrm{cr}}} = \left(\bar{u}_{\mathrm{cr}} - F\right)/\Gamma_{k_{\mathrm{cr}}}.$$

   Such closures are only useful in the weakly nonlinear case.
2. *Models with consistent equilibrium single point statistics*: Here we want to construct the simplest closure model with consistent equilibrium single point statistics (22). So the direct way is to choose constant damping and noise term at most scaled with the total variance. We propose two possible choices for (37) below.
   Gaussian closure 1 (GC1-1pt): let

$$d_{M,k}(R) = d_M \equiv \text{const.}, \quad \sigma^2_{M,k}(R) = \sigma^2_M \equiv \text{const.}, \quad Q^{\mathrm{GC1}}_F = -\left(d_M R + R d_M\right) + \sigma^2_M I; \tag{39}$$

   Gaussian closure 2 (GC2-1pt): let

$$d_{M,k}(R) = \varepsilon_M \frac{J}{2} \frac{(\mathrm{tr}R)^{1/2}}{(\mathrm{tr}R_\infty)^{3/2}} \equiv \varepsilon_M \bar{d}, \quad \sigma^2_{M,k}(R) = \varepsilon_M \frac{(\mathrm{tr}R)^{3/2}}{(\mathrm{tr}R_\infty)^{3/2}}, \quad Q^{\mathrm{GC2}}_F = -\varepsilon_M \left(\bar{d}R + R\bar{d}\right) + \varepsilon_M \frac{(\mathrm{tr}R)^{3/2}}{(\mathrm{tr}R_\infty)^{3/2}} I. \tag{40}$$

   GC1-1pt is the familiar strategy of adding constant damping and white noise forcing to represent nonlinear interaction (Majda et al (2005)). In GC2-1pt, the term multiplying dissipation scales with $(\mathrm{tr}R)^{1/2}$ while the term multiplying noise scales with $(\mathrm{tr}R)^{3/2}$; these are dimensionally correct surrogates for the quadratic nonlinear terms.
   Note that GC1-1pt includes parameters $\left(d_M, \sigma^2_M\right)$ and the nonlinear energy $\mathrm{tr}Q^{\mathrm{GC1}}_F = -2d_M \mathrm{tr}R_M + J\sigma^2_M$ may not be conserved, while GC2-1pt has one parameter ε_M and nonlinear energy conservation is enforced by construction $\mathrm{tr}Q^{\mathrm{GC2}}_F = 0$. Single point statistics consistency can be fulfilled through tuning the control parameters according to Proposition 1. Still as discussed in Section 3.3, it is not sufficient with only single point statistics consistency for desirable model forecasts (see Figure 2 for the inherent information barrier of these approaches). Thus for better approximation, we need to consider changing damping and noise amplitudes for each mode.

3. *Models with consistent equilibrium statistics for each mode*: Next we improve the previous closure methods to ensure equilibrium statistical consistency in each mode. Simply this can be achieved through changing the damping rate for each mode according to the stationary state statistics. Specifically for the above GC1 and GC2 in (39) and (40), the models can be improved by a slight modification in the damping rates along each mode.
GC1:
$$d_{M,k}(R) = d_{M,k}, \quad \sigma_{M,k}^2(R) = \sigma_M^2, \quad d_{M,k} = [-\Gamma_k \bar{u}_\infty - d] + \sigma_M^2 / 2 r_{k,\infty}; \tag{41}$$

GC2:
$$d_{M,k}(R) = \varepsilon_{1,k}\bar{d}, \qquad \sigma_{M,k}^2(R) = \varepsilon_M \frac{(\mathrm{tr}R)^{3/2}}{(\mathrm{tr}R_\infty)^{3/2}},$$
$$\varepsilon_{1,k} = \frac{2[-\Gamma_k \bar{u}_\infty - d] r_{k,\infty} + \varepsilon_M}{J r_{k,\infty}/\mathrm{tr}R_\infty}, \quad \bar{d} = \frac{J}{2}\frac{(\mathrm{tr}R)^{1/2}}{(\mathrm{tr}R_\infty)^{3/2}}. \tag{42}$$

Above $d_{M,k}$ or $\varepsilon_{1,k}$ is chosen so that the system in (36) converges to the same equilibrium mean $\bar{u}_\infty$ and variance $r_{k,\infty}$ as the true model, therefore the values of these parameters ensuring equilibrium consistency can be calculated by finding the steady state solutions of (36) through simple algebraic manipulations (see Appendix C). Still in (41) and (42) the damping and noise are chosen empirically (depending on the one additional parameter σ_M^2 or ε_M) without consideration about the true dynamical features in each mode. A more sophisticated strategy with slightly more complexity in computation is to introduce the damping and noise judiciously according to the linearized dynamics (that is, the operator L_v in (15b)). Then climate consistency for each mode can be satisfied automatically. That is the modified Gaussian closure model (MQG) introduced in Sapsis and Majda (2013d). We can also include this model into our category as
MQG:
$$d_M(R) = \frac{f(R)}{f(R_\infty)} N_\infty, \quad \sigma_M^2(R) = -\frac{\mathrm{tr}Q_{F-}^{\mathrm{MQG}}}{\mathrm{tr}Q_{F+,\infty}^{\mathrm{MQG}}}\left[(\Gamma_k \bar{u}_\infty + d) r_{k,\infty}\delta_{I_+} + q_s\right], \tag{43}$$

with
$$N_{\infty,kk} = [\Gamma_k \bar{u}_\infty + d]\,\delta_{I_-} - \frac{1}{2} q_s r_{k,\infty}^{-1}.$$

Above I_- represents the unstable modes with $\Gamma_k \bar{u}_\infty + d > 0$ while I_+ is the stable ones with $\Gamma_k \bar{u}_\infty + d \le 0$. We usually choose $f(R) = \sqrt{\mathrm{tr}R}$, and $q_s = d_s \lambda_{\max}(Q_{F,\infty})$ ($\lambda_{\max}$ the largest eigenvalue of $Q_{F,\infty}$) as one additional tuning parameter to control the model responses.

4. *Improve forecast skill with kicked response operator*: The above methods (41), (42), and (43) construct closure models with consistent equilibrium statistics. Still equilibrium fidelity of imperfect models is a necessary but not sufficient condition for model prediction skill with many examples (Majda and Gershgorin (2011a,b); Majda and Branicki (2012)). In order to get precise forecasts for various forced responses, it is also crucial to seek models that can correctly reflect the system's 'memory' to its previous states. The linear response operator $\mathscr{R}_A$ represents the lagged-covariance of certain functions (and thus can describe the 'memory' of the system to previous states). Also note that for all the methods above, there is still one more free parameter (σ_M^2, ε_M, or q_s) for us to control the model performance. Adopting the general strategy suggested in Section 2.3, we can improve model sensitivity through tuning imperfect models in a training phase before the prediction step. Thus the optimal model parameter can be selected through minimizing the information distance in the linear response operators in (12) between the imperfect closure model and the truth.

Therefore we reach three different ideas for the closure methods with increasing complexity, namely, GC1, GC2, and MQG in abbreviation. As we have discussed in Section 3.2, it is useful to check the energy conservation properties of these schemes which can imply the prediction skill of the models in advance. Proposition 1 states that good performance of the closure schemes can be expected only if the nonlinear energy is conserved sufficiently well, that is, $\left\|\mathrm{tr}Q_F^M\right\| \ll 1$. From the construction of these models, GC1 may not guarantee the symmetry of nonlinear energy, while GC2 is constructed with better nonlinear energy conservation property. And symmetry of nonlinear energy is always enforced for MQG by the scaling factor in (43). On the other hand, increasing computational costs are required in GC1, GC2, and MQG. To be specific, we can calculate $\mathrm{tr}Q_F^M$ for each scheme explicitly, then
$$\mathrm{tr}Q_F^{\mathrm{GC1}} = -\sigma_M^2 \sum_k \frac{\delta r_k}{r_{k,\infty}}, \quad \mathrm{tr}Q_F^{\mathrm{GC2}} = \varepsilon\left(\frac{\mathrm{tr}R}{\mathrm{tr}R_\infty}\right)^{1/2}\left(\sum_k \frac{\delta r_k}{r_{k,\infty}} - J\frac{\delta \mathrm{tr}R}{\mathrm{tr}R_\infty}\right), \quad \mathrm{tr}Q_F^{\mathrm{MQG}} = 0. \tag{44}$$

The error for GC1 from the nonconservative nonlinear interactive could be large as the perturbation increases, while GC2 offers much better conservation property due to the cancellation in the second parenthesis. MQG conserves nonlinear energy exactly, and as an expense requires more detailed calibrations in each mode. These properties will be further checked in the following parts with numerical tests.

4.2 Algorithm for full space models

Here we summarize the closure methods resolving the entire phase space of the system.

Algorithm. *(Statistical closure methods for full space models)*

- Set up the dynamical equations for the first two moments with proper model nonlinear flux closure scheme Q_F^M in (37)

$$\boxed{\begin{cases} \frac{d\bar{u}}{dt} &= -d(t)\,\bar{u}(t) + \frac{1}{J}\sum_{k=-J/2+1}^{J/2} r_k(t)\,\Gamma_k + F(t), \\ \frac{dr_k}{dt} &= 2\left[-\Gamma_k\bar{u}(t) - d(t)\right] r_k(t) + Q_{F,kk}^M, \quad k = 0,1,...,J/2; \end{cases}} \tag{45}$$

- *Calibration step*:
  - Decide the closure form to use and calculate the equilibrium consistent parameters according to equilibrium statistics as in (41), (42), or (43) for GC1, GC2, or MQG respectively;
  - Calculate kicked response operators of the model through (9) and (10), and train parameters to achieve the optimal linear response (12) for imperfect models under the information metric using the first order relative entropy expansion (11);
- *Prediction step*:
  - Make predictions for the statistics of the dynamical system in response to different kinds of external forcing perturbations using the optimal parameters.

4.3 Forecast skill of the closure models for forced responses

Now we display the imperfect model prediction skill applying the closure methods described in the algorithm above for the homogeneous L-96 testbed in (16)

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

The statistics are resolved under Fourier spectral domain where the system is diagonalized under this homogeneous setup. We are mostly interested in checking the imperfect models' sensitivity in capturing responses as the system is perturbed away from the steady state with homogeneous external forcing perturbation $F(t) = \bar{F} + \delta F(t)$. First we will illustrate several important issues as discussed in the previous sections through simple examples. Then the imperfect model prediction skill as well as the improvement through the information-response framework will be compared through checking the models' ability to capture the responses to several different types of perturbed external forcing terms.

4.3.1 Information barrier and calibration in training phase

Comparing FDT response with dynamical model prediction It might have been noticed that the system's responses to perturbations can be approximated through the fluctuation-dissipation theorem (FDT) using the formula in (7) as well as the kicked response (10) without really running the dynamical model. However, FDT together with the linear response theory just calculates the responses in first order of the perturbation δ. It can only be viewed as a linear expansion of the predictions and will deviate from the truth as the perturbations becomes large. In Figure 1, we compare the prediction skill between FDT linear theory with the imperfect dynamical model. The prediction skills for stationary mean and variance are checked with constant perturbations in the forcing term $\delta F \in [-2,2]$ added to the unperturbed regime $F = 8$. In FDT prediction, the response operator $\mathscr{R}$ is calculated from the perfect model, so it should be quite accurate with little error (note that in practical applications, this accurate response operator might always be unavailable). And we use the GC2 model with optimal parameter from the training phase as an example for the dynamical imperfect models. As observed in the results, the FDT results appear just linear and deviate from the (nonlinear) truth as the perturbation amplitude δF increases even though with accurate estimation of the linear response operator and constant perturbation. On the other hand, the dynamical imperfect model using the closure method offers more precise prediction for the nonlinear responses for both the mean and variance. The nonlinear and large deviations from the the equilibrium statistics with perturbations also imply that estimation in the unresolved variables using stationary statistics may not be accurate. We will discuss this in more detail in the next section for reduced order models.

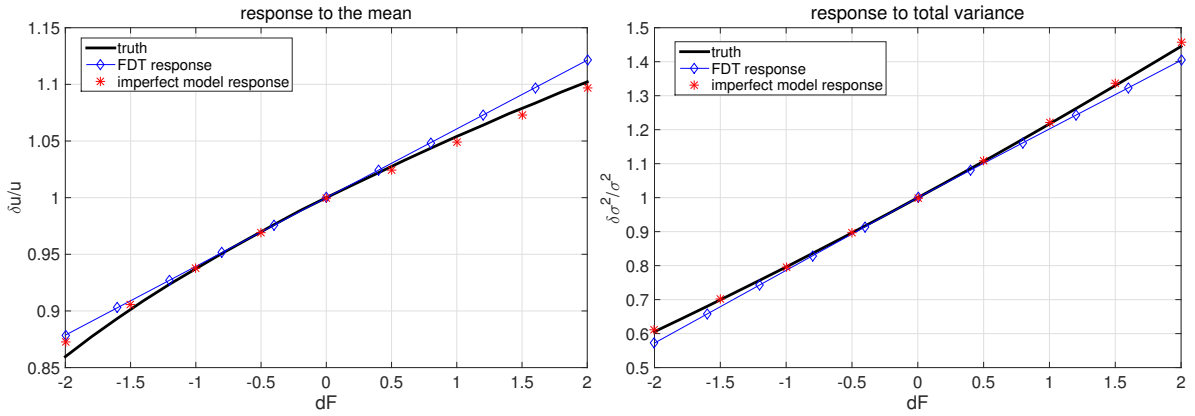


Fig. 1: Comparison between linear FDT responses to constant forcing perturbations and the model predictions (using GC2 model with optimal parameter here as an example). The ratios between the perturbed responses and the equilibrium value for the mean $\delta\bar{u}/\bar{u}_\infty$ and total variance $\delta \mathrm{tr}R/\mathrm{tr}R_\infty$ are plotted as a function of the constant perturbed external forcing δF under the unperturbed regime $F = 8$.

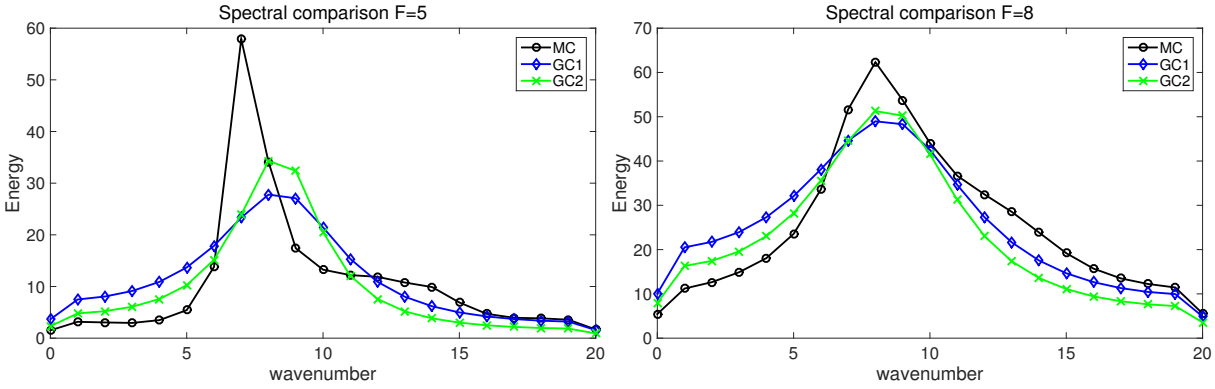
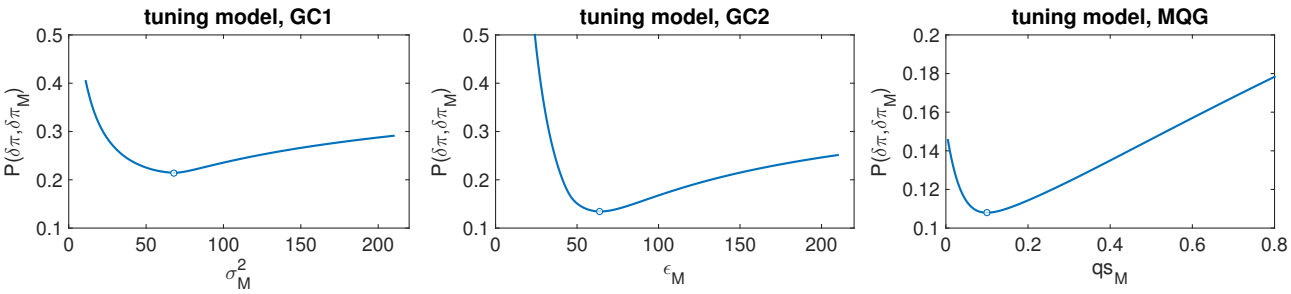


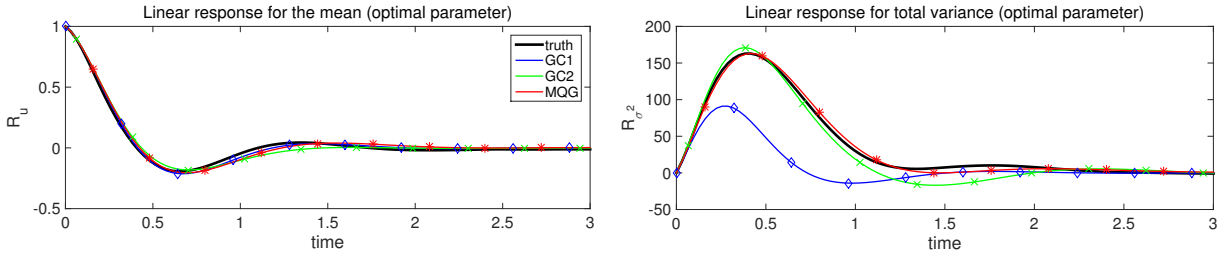
Fig. 2: Information barriers for imperfect closure models GC1-1pt and GC2-1pt with only consistent equilibrium single point statistics as constructed in (39) and (40). The steady state variances under Fourier basis from the two imperfect model results (with consistency in single point statistics $\mathrm{tr}R = \mathrm{tr}R_M$) are compared with the truth from Monte-Carlo simulations in two typical dynamical regimes $F = 5, 8$.

Information barrier with single point statistics We have displayed the information barrier when only the single point statistics consistency is considered in Section 3.3. For a further illustration using numerical experiments, we check the errors using simple closure models GC1-1pt and GC2-1pt as in (39) and (40) with only equilibrium single point consistency guaranteed. Figure 2 shows the stationary state spectra in spectral domain from GC1-1pt and GC2-1pt with the optimal parameters according to (12) compared with the true spectrum using Monte-Carlo simulation. In this coarser version of GC1-1pt and GC2-1pt model, we only make sure the consistency in total variance in equilibrium by enforcing symmetry of nonlinear energy $\mathrm{tr}Q^M_{F,\infty} = 0$. Still large errors (thus large information barrier for these models) exist for each individual mode for both dynamical regimes $F = 5$ (weakly chaotic) and $F = 8$ (strongly chaotic), consistent with what we have calculated from (35) for single point statistics. This example shows the necessity of improving the models further for consistency in each mode using (41) and (42). It is interesting to remark that as the forcing F increases for $F \gg 1$, the spectrum of L-96 approaches equipartition (Majda et al (2005); Majda and Wang (2006)), so the information barrier of the single point statistics in (33) is reduced in this regime.

Improving imperfect model skill in training phase Finally we turn to the models (41), (42), or (43) with consistent equilibrium fidelity by construction. As suggested in the algorithm, the optimal parameter σ_M^2, ε_M, or q_s accordingly is achieved through searching for the minimized information distance in response operators under the information measure (12). The response operators for the mean and variance can be calculated through the unperturbed dynamics $F = \bar{F}$ with perturbed initial value $\mathbf{u}_0 = \bar{\mathbf{u}}_0 - \delta\mathbf{u}'$ using the kicked response strategy as described



(a) Information errors in response operator for different models



(b) Linear response operators for the mean and total variance with optimal parameter

Fig. 3: Tuning imperfect model parameters in the training phase with full dimensional models. The first row shows the time-averaged information distances in linear response operators for the three closure models GC1, GC2, and MQG as the model parameter changes (the point where the information error is minimized is marked with a circle). The second row compares the optimized imperfect model response operators for the mean and total variance with the truth for all these three imperfect models.

in (9) and (10) (Check Appendix B for details about calculating the kicked response operators numerically). In Figure 3, we plot the total information errors for each method with changing parameters in the first row, and the optimal model predicted response operators for the mean and total variance are displayed in the second row. From the errors, the optimized information errors for GC2 and MQG are smaller than GC1. This is consistent with our discussion in (44) considering the symmetry in nonlinear energy of each method. The same can be observed from the plots for the response operators. For GC2 and MQG, good agreements for the mean state always imply good fitting for the total variance, while large errors in the total variance with GC1 appear even though the mean state is fit well.

4.3.2 Testing imperfect model prediction skill with different forced perturbations

We have achieved the optimal model parameters by tuning response operators in the training phase with the help of information theory. As we have shown in Section 2 and 3, this optimal model can minimize the information barrier in model predictions and offer uniform performance in response to various perturbations. To validate this point, we compare and check the model improvement in prediction skill according to various forcing perturbations. Particularly here, we choose four different perturbed external forcing forms representing distinct dynamical features. In Figure 4, the four different external forcing terms that will be tested are plotted. The first two are the ramp-type perturbations of the external forcing driving the system smoothly from equilibrium to a perturbed state with higher or lower energy. This could be viewed as a simple model mimicking a climate change scenario. Next considering the simulation about a seasonal cycle, we would also like to check the case with periodic perturbations. And finally, the case with random white noise forcing is applied to test the models' ability for random effects. All perturbations δF are of an amplitude (or variance) of 10% of the equilibrium value $F = 8$.

Figure 5-8 compare three imperfect model performances under the four different forcing perturbations. To illustrate the improvement in prediction skill through this information-response framework, the model predictions with optimal parameters from the training phase are displayed together with another prediction result using non-optimal parameter by fitting the mean only in the training phase. We can regard this imperfect model as a familiar sophisticated version of the strategy from climate science when only the mean is tuned. In the beginning two rows, we show the model outputs for the mean and total variance with closure methods GC1, GC2, MQG compared with the truth from Monte-Carlo simulation. As expected, the model prediction skill increases as more and more detailed calibration about the nonlinear flux are proposed from GC1 to GC2, MQG. For a clearer comparison of the three models, we can check the information distance in signal part (for the mean state) and dispersion part

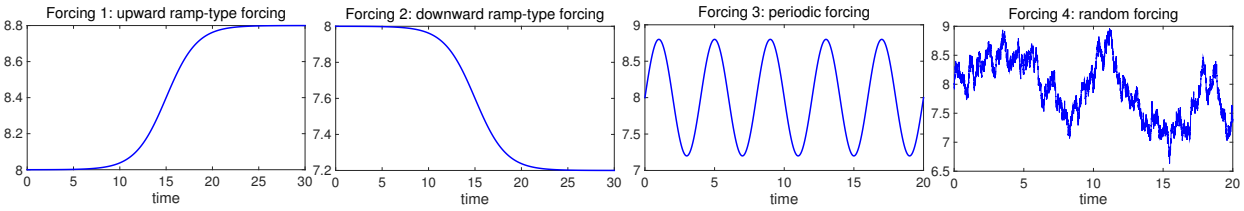


Fig. 4: Different perturbed forcing terms δF used to test the model prediction skill under unperturbed test regime $F = 8$.

(for variances) separately in the following two rows. Logarithmic coordinate in y-direction is adopted for clarity in distinguishing the three model errors. In the signal part, the error in the mean can be minimized to small amount for all three models under optimal parameter; while for dispersion part, MQG and GC2 have much better prediction for the prediction in variance compared with GC1. And MQG performs slightly better than the GC2 model. This can be explained through the analysis in Proposition 1 as well as the nonlinear energy conservation property summarized in (44). With exact symmetry in nonlinear energy $\mathrm{tr}Q_F^{\mathrm{MQG}} = 0$, MQG can give exact prediction for the variances once the accurate prediction for the mean is achieved. And GC2 only includes small errors in the nonlinear flux term (importantly through the scaling factor in the additional damping and excitation), thus offers nearly the same accuracy in prediction. However, for GC1 without the nonlinear energy conservation, good prediction in the mean state cannot be a guarantee for the precision in variance predictions. Besides, GC1 is also a good example illustrating that one model with equilibrium fidelity may lack the ability in model sensitivity, proving that equilibrium fidelity is only necessary but not sufficient for model prediction (Majda and Gershgorin (2011a,b); Majda and Branicki (2012)).

Finally we check the model improvement in prediction skill in the last row. We compare the full information errors for these three models separately with optimal parameter and non-optimal one for comparison. Clearly, the information barrier can be reduced uniformly with optimal parameter for all these three closure methods and all these four perturbation scenarios regardless of any specific perturbation forcing applied, proving the uniform improvement through tuning the linear response operators in the training phase. Further MQG model can achieve the best and largest minimization in error as the optimal parameter is applied, and GC2 model also displays excellent improvement. Again the improvement from GC1 model is the smallest and limited. This is due to its inherent model error with crude estimation about the closure form. But on the other hand, it also needs to be noticed that more computational complexities are included for MQG and GC2 models in order to achieve the more precise calibration.

5 Low order models in a reduced subspace

The closure methods above will become impractical when it comes to really high dimensional turbulent dynamical systems, for example, climate systems with dimensionality of order 10^3. In this situation, the covariance matrix R will be too huge that it becomes computational intractable to evolve the entire matrix in time. Therefore one alternative practical strategy is to develop reduced order methods that only explicitly calculate variances in a low-dimensional subspace spanned by primary empirical orthogonal functions (EOFs) $\{\mathbf{v}_0, \cdots, \mathbf{v}_s\}$ with $s \ll J$ (J the dimensionality of the system), see for example Sapsis and Majda (2013c). The corresponding reduced order representation of the state variables under these resolved basis becomes $\mathbf{u} = \bar{\mathbf{u}} + \sum_{k=0}^{s} Z_k \mathbf{v}_k$. To see the possibility of achieving this, first note that the dynamical equations for variances (21b) in each mode $r_k = \langle Z_k Z_k^* \rangle$ are rather independent with each other according to the previous closure strategies with higher order interactions replaced. Thus it is realizable to restrict the variance equations inside the chosen subspace. Actually following the same strategy by replacing the high order interaction terms by proper damping and noise as in Section 4, the equivalent counterpart of the closure models can be formulated as a low-order stochastic system

$$\frac{d\bar{u}_M}{dt} = -d(t)\,\bar{u}_M(t) + \frac{1}{J}\sum_{|k|\le s} r_{M,k}(t)\,\Gamma_k + F(t) + G_\infty, \tag{46}$$

$$\frac{dZ_k}{dt} = -d(t)\,Z_k + \left(e^{2\pi i \frac{k}{J}} - e^{-2\pi i \frac{2k}{J}}\right)\bar{u}_M Z_k - d_{M,k}(R)\,Z_k + \sigma_{M,k}(R)\,\dot{W}_k, \quad k = 0, 1, ..., s, \tag{47}$$

with $R = \mathrm{diag}\left\{\langle Z_k Z_k^* \rangle\right\}$ for $k = -J/2+1, \cdots, J/2$, the full covariance matrix. The mean dynamics is the same as the closure models with an additional correction term G_∞ to compensate the unresolved modes (see the following

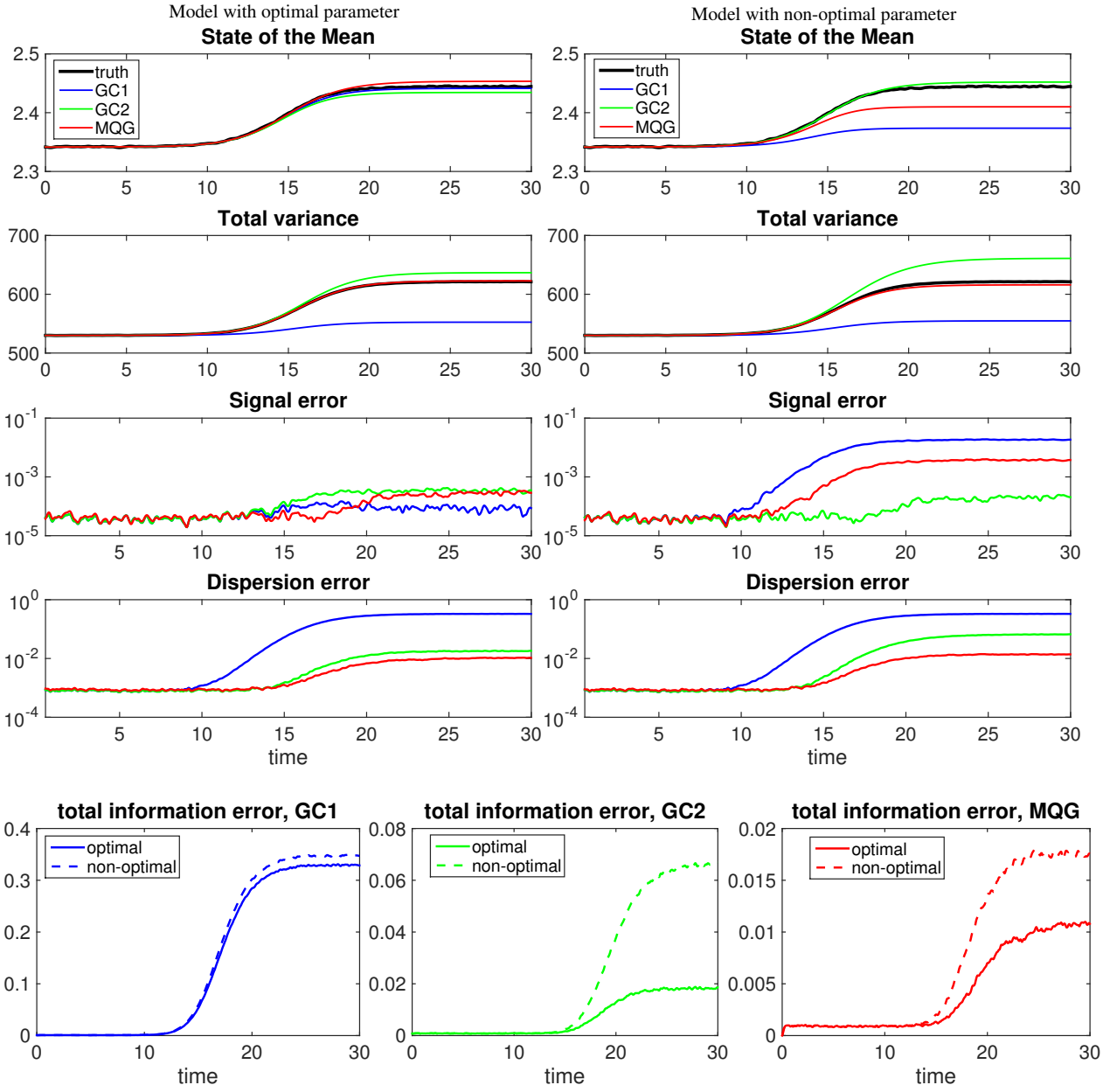


Fig. 5: Model predictions with upward ramp-type forcing. Predicted mean and total variance for the closure methods GC1 (blue), GC2 (green), and MQG (red) are compared with the truth (black) from Monte-Carlo simulation in the first two rows. The information errors in signal part and dispersion part separately for these three models are followed in the next two rows (note that we use logarithmic coordinate in y-direction for distinguishability between models). Results with optimal parameter (left column) and one nonoptimal case by fitting only the mean in the training phase (right column) are compared. Finally to display the improvement from training the response operators for each model, we show the total information errors for each model with optimal and nonoptimal parameters in the last row.

section for detail); and the dynamical equations for the stochastic coefficients in the above model only solve the dominant modes (and as we will see in the numerical tests, it is possible to have only 1 or 2 resolved modes in our test case). Note that the constructed stochastic system is nonlocal due to the inclusion of the global variance R in the artificial damping $d_M(R)$ and noise $\sigma_M(R)$. Through proper choice of the parameters according to GC1 (41), GC2 (42), or MQG (43) as before, these stochastic system should converge to the same first two order statistics with the moment closure model. This stochastic system sets up the foundation for constructing reduced order moment closure methods. Still several problems need to be taken into account for the above model reduction process: i) How to ensure climate consistency and optimal linear response as before in the reduced model; ii) How to include

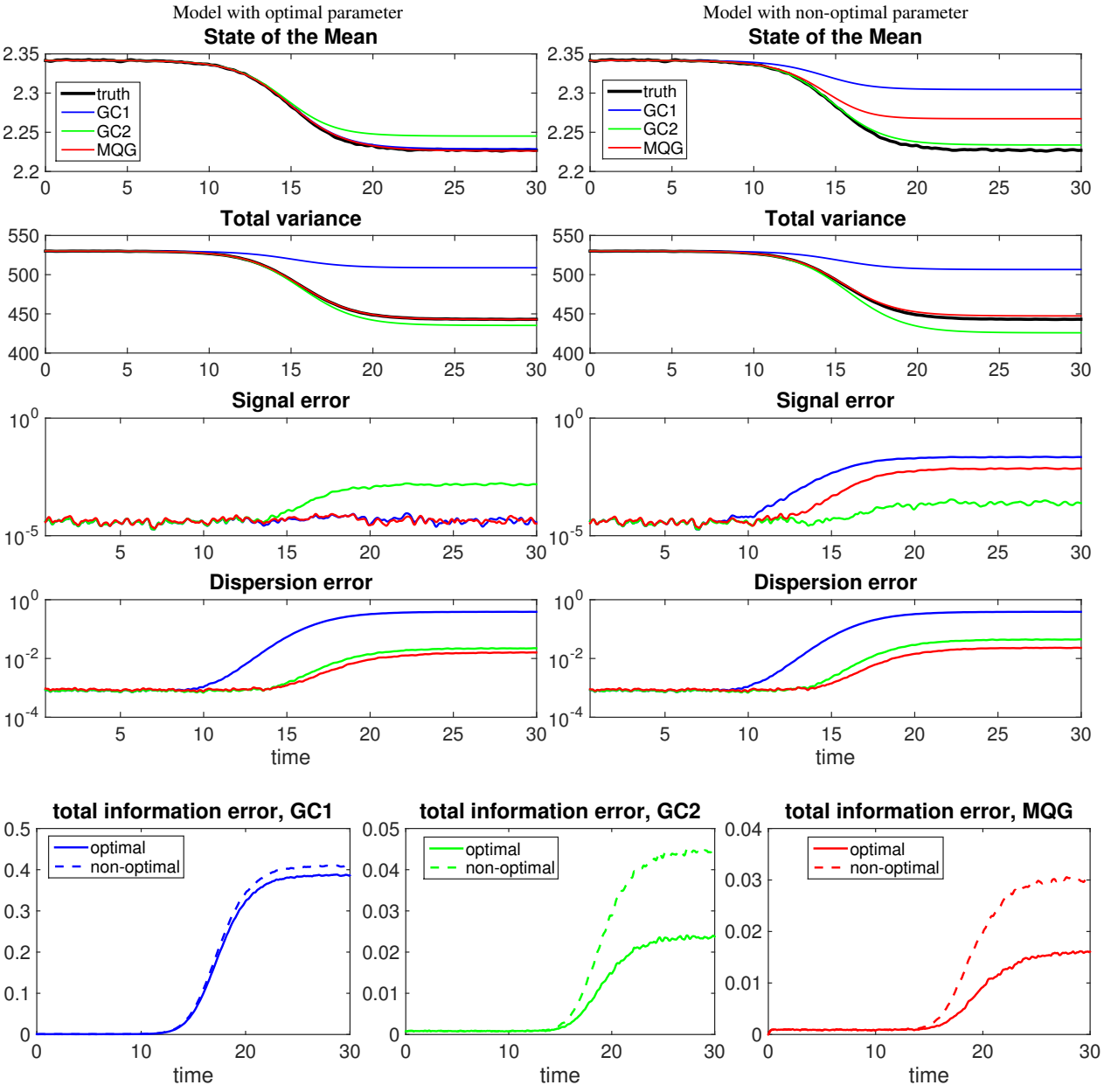


Fig. 6: Model predictions with downward ramp-type forcing. Predicted mean and total variance for the closure methods GC1 (blue), GC2 (green), and MQG (red) are compared with the truth (black) from Monte-Carlo simulation in the first two rows. The information errors in signal part and dispersion part separately for these three models are followed in the next two rows (note that we use logarithmic coordinate in y-direction for distinguishability between models). Results with optimal parameter (left column) and one nonoptimal case by fitting only the mean in the training phase (right column) are compared. Finally to display the improvement from training the response operators for each model, we show the total information errors for each model with optimal and nonoptimal parameters in the last row.

the nonlocal scale factor (which always includes the total energy tr$R$) in the nonlinear flux approximation Q_F^M if only a subspatial variances are resolved; iii) How to get the unresolved parts of $\sum_{k=-J/2+1}^{J/2} r_k(t)\Gamma_k$ in the mean dynamics.

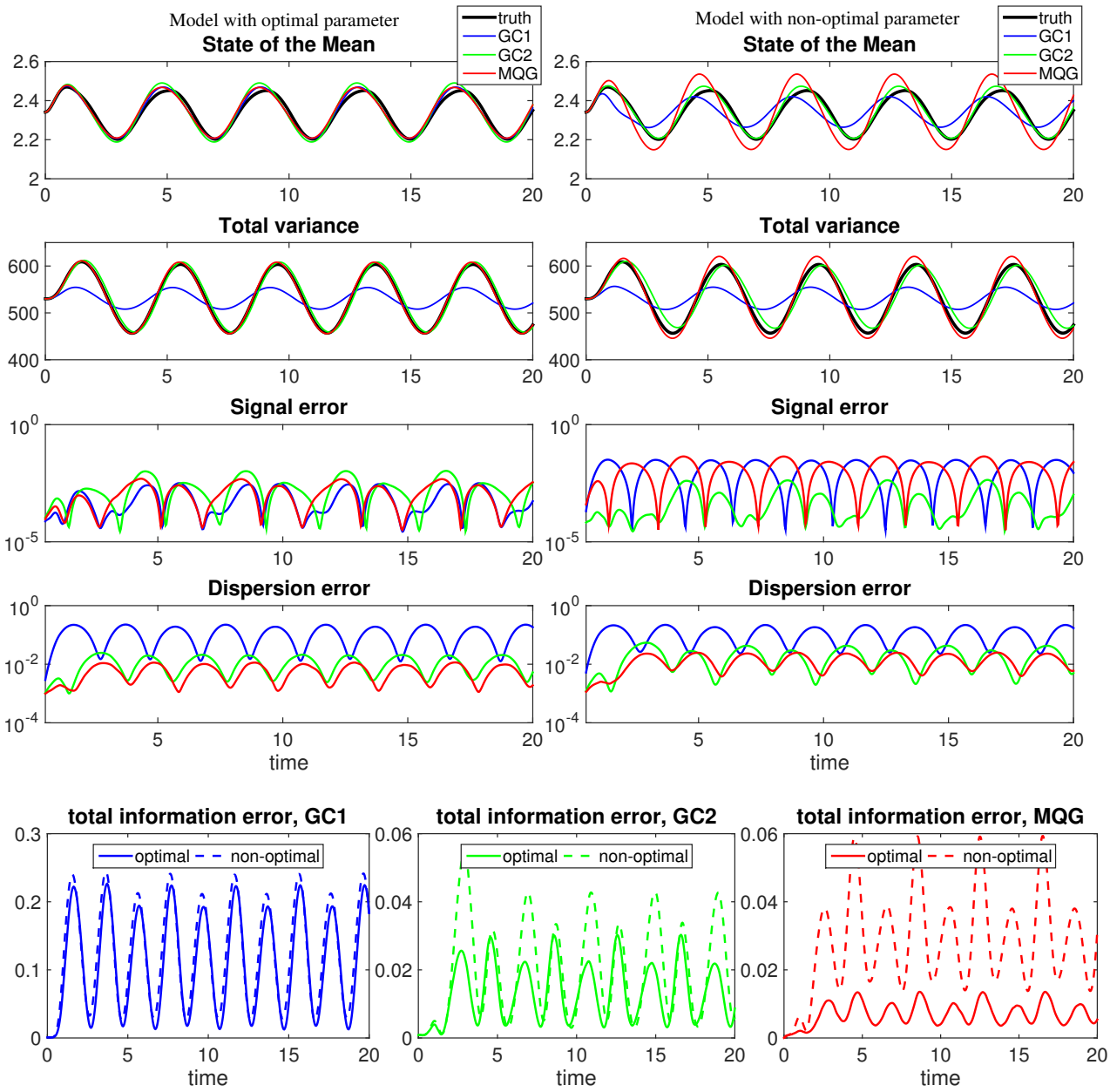


Fig. 7: Model predictions with periodic forcing. Predicted mean and total variance for the closure methods GC1 (blue), GC2 (green), and MQG (red) are compared with the truth (black) from Monte-Carlo simulation in the first two rows. The information errors in signal part and dispersion part separately for these three models are followed in the next two rows (note that we use logarithmic coordinate in y-direction for distinguishability between models). Results with optimal parameter (left column) and one nonoptimal case by fitting only the mean in the training phase (right column) are compared. Finally to display the improvement from training the response operators for each model, we show the total information errors for each model with optimal and nonoptimal parameters in the last row.

5.1 Reduced order models

Here we address the problems raised above for reduced order models with only a small portion of the dynamical system resolved. We begin with the above low-order stochastic model (46) and (47) for consistent moment closure methods. The resolved basis $\{\mathbf{v}_0, \cdots, \mathbf{v}_s\}$ can be chosen from the equilibrium EOFs with largest variances (and the number of resolved basis s is chosen rather empirically depending on the uncertainty of the system). That is, we run the original closure model (36a) and (36b) but under constraint in the subdimensional phase space of

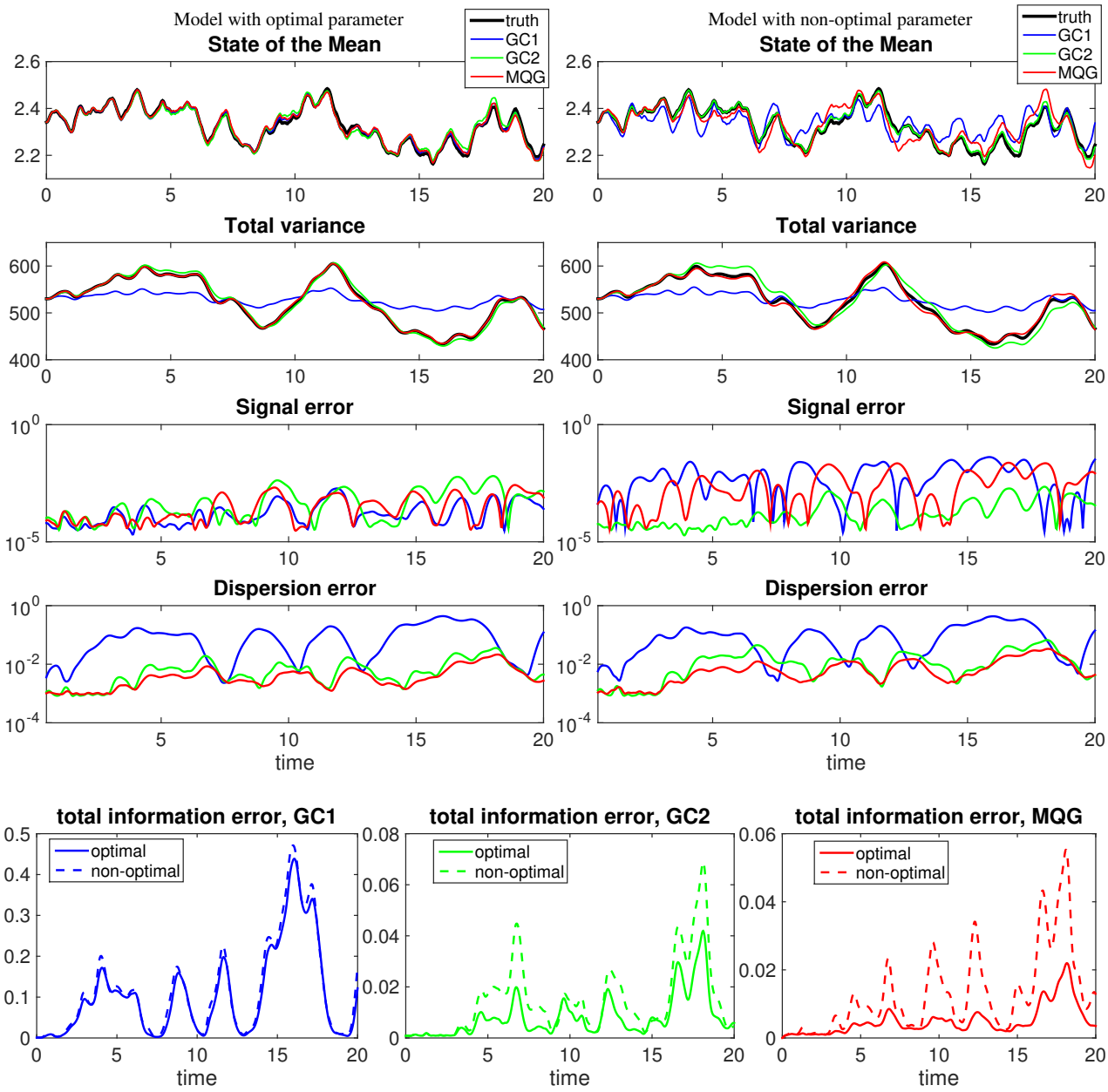


Fig. 8: Model predictions with random forcing. Predicted mean and total variance for the closure methods GC1 (blue), GC2 (green), and MQG (red) are compared with the truth (black) from Monte-Carlo simulation in the first two rows. The information errors in signal part and dispersion part separately for these three models are followed in the next two rows (note that we use logarithmic coordinate in y-direction for distinguishability between models). Results with optimal parameter (left column) and one nonoptimal case by fitting only the mean in the training phase (right column) are compared. Finally to display the improvement from training the response operators for each model, we show the total information errors for each model with optimal and nonoptimal parameters in the last row.

dimensionality s so that the computation expense can be affordable

$$\frac{d\bar{u}_M}{dt} = -d(t)\bar{u}_M(t) + \frac{1}{J}\sum_{|k|\le s} r_{M,k}(t)\Gamma_k + F(t) + G_\infty, \tag{48}$$

$$\frac{dr_{M,k}}{dt} = 2\left[-\Gamma_k\bar{u}_M(t) - d(t)\right] r_{M,k}(t) + Q^M_{F,kk}(\mathrm{tr}R_s), \quad k = 0,1,\ldots,s. \tag{49}$$

The major differences in this reduced model include two points. The unresolved variances required in the mean dynamics (48) is amended by an additional forcing term represented by G_∞ that balances the contribution from

the truncated modes (Sapsis and Majda (2013c)). This additional part can be determined through equilibrium data from statistical steady state

$$G_\infty = 1/J \sum_{|k|>s} r_{k,\infty}\Gamma_k = d\bar{u}_\infty - 1/J \sum_{|k|\le s} r_{k,\infty}\Gamma_k - F.$$

And the (linear) scaling factor tr$R$ required in the variance dynamics (49) in Q_F^M is replaced by the avalible resolved part tr$R_s = \sum_{k=1}^{s} r_k$. We will refer this as the *original model* in the following simulations.

Still as shown in the plot for responses in Figure 1, the changes in variance as the perturbation amplitude varies are rather significant and nonlinear. As a result, using only unperturbed statistics to estimate the unresolved variances $r_{k,\infty}, |k| > s$ may not be accurate and desirable for both the mean and variance dynamics. Keeping all these shortcomings in mind, we propose the following further corrections to the reduced order methods, and refer the resulting model as the *corrected model*.

- **Ensuring climate consistency and optimal linear response in the reduced subspace**:
  We have shown in Fact 1 and 4 in Section 2 that the information distance can be calculated within a subspace, and the kicked response operator can be calculated individually for each mode as in Fact 3. So the models can be tuned in the same way as the previous full space case but minimizing the information distance instead in the low-dimensional subspace in the training phase.
- **Getting total variance of the system through energy equations**:
  In closure schemes (42) and (43) one important scaling factor which always includes the total variance tr$R$ of the system is utilized for the artificial damping and noise. This can be solved with efficiency by introducing one additional scalar equation as described in (25)

$$\frac{dE}{dt} = -2d(t)E + JF(t)\bar{u}.$$

  Then tr$R$ can be achieved by solving $E = \frac{J}{2}\bar{u}^2 + \frac{1}{2}\text{tr}R$.
- **Corrections for mean dynamics**:
  In the mean dynamical equation, it requires one term $\sum_{k=-J/2+1}^{J/2} r_k(t)\Gamma_k$ containing variances for each mode. For reduced models, we can only have access to the resolved leading variances $r_1, \cdots, r_s$. To estimate the values for unresolved modes $r_{k,\text{un}}$, additional correction G_∞ using steady state information $r_{k,\infty}$ is added to equation (48). We can further improve it from the equilibrium statistics by adding finer first-order corrections in response to the perturbation by making use of the linear response operator (7)

$$r_{k,\text{un}} \sim r_{k,\infty} + \delta r_k' = r_{k,\infty} + \int_0^t \mathscr{R}_{r_k}(t-s)\,\delta F'(s)\,ds, \quad k > s \tag{50}$$

Finally we can summarize the reduced order model algorithm as follows

Algorithm. *(Statistical closure methods for reduced order models)*

- Set up the dynamical equations for the first two moments in the reduced order subspace which is spanned by the leading order EOFs $\{\mathbf{v}_1, \cdots, \mathbf{v}_s\}$, $s \ll J$, together with the scalar energy equation for total statistical energy E

$$\boxed{\begin{cases} \frac{d\bar{u}}{dt} &= -d(t)\bar{u}(t) + \frac{1}{J}\sum_{|k|\le s} r_k(t)\Gamma_k + F(t) + G_\infty + \delta G, \\ \frac{dr_k}{dt} &= 2\left[-\Gamma_k\bar{u}(t) - d(t)\right] r_k(t) + Q_{F,kk}^M(\text{tr}R), \quad k = 0,1,...,s, \\ \frac{dE}{dt} &= -2d(t)E + J\bar{u}F(t). \end{cases}} \tag{51}$$

  - Calculate the total variance for the parameter in nonlinear flux approximation Q_F^M from the entire statistical energy formula

$$\text{tr}R = 2E(t) - J\bar{u}^2.$$

  - Calculate the unresolved variances in the mean dynamics from the equilibrium statistics $r_{k,\infty}$ and linear response (50)

$$G \equiv G_\infty + \delta G = G_\infty + \frac{1}{J}\sum_{|k|>s} \delta r_k' \Gamma_k, \tag{52}$$

    with $G_\infty = d\bar{u}_\infty - \frac{1}{J}\sum_{|k|\le s} r_{k,\infty}\Gamma_k - F$.
- *Calibration step*:
  - Decide the closure form and calculate the equilibrium consistent parameters according to equilibrium statistics as in (41), (42), or (43) for the reduced order GC1, GC2, or MQG model respectively;

  - Get the kicked response of the model from (9) and (10), and train the parameters to achieve optimal linear response (12) for imperfect models under the information metric using (11) in the reduced subspace with dimensionality s;
- *Prediction step*:
  - Make predictions for the statistics in principal directions of the dynamical system in response to different kinds of external forcing perturbations using the optimal parameters.

Remark. 1. The reduced order MQG model (ROMQG) needs some more sophisticated calibrations about the nonlinear fluxes in the unresolved directions. We neglect these specific techniques and detailed explanations can be found in Sapsis and Majda (2013c).

2. In the correction term for the mean (52), we may not need to calculate the linear response for all the unresolved modes (which may still be expensive for large systems). Only adding the linear corrections $\delta r_k'$ up to a large enough wavenumber K should be enough for practical implementation. We will illustrate this with simple example in the final part of this section.

5.2 Forecast skill of the reduced models for forced responses

Now we check the prediction skill of the reduced order models in capturing uncertainties in principal directions using the L-96 testbed. We will only take a three dimensional subspace (compared with the 40 dimensional model) under the zero base mode $\mathbf{v}_0 = 1/\sqrt{J}(1,\cdots,1)^T$, and the most energetic Fourier mode $\mathbf{v}_1$ (including real and complex part) with largest variance in equilibrium spectrum. We are interested in checking whether these correction strategies for the reduced methods can actually improve the model prediction skill. Particularly, for the reduced GC1 model (41), no scaling factors but constant damping and excitation forms are used in the closure scheme Q_F^{GC1} (so we can only improve the model from the mean equation and the energy equation cannot help in this model), while GC2 and MQG are a good examples to display the combined improvement from both corrections in mean and covariance equations. We will see below that a calibrated GC2 model by the training method will be easy to design systematically and yet will be superior to ROMQG from Sapsis and Majda (2013c).

5.2.1 Calibration in training phase for optimal response

As before in the full space case, we need to first tune the reduced model parameters in a training phase for optimal responses. In Figure 9, we show the full information distance in the resolved subspace $\{\mathbf{v}_0, \mathbf{v}_1\}$ for the three reduced models GC1, GC2, MQG, together with their errors in the signal part and dispersion part separately. Improvements from adopting the corrected method with energy and linear response corrections can be observed for all these three strategies by comparing with the results with the original method. Furthermore larger improvements can be achieved through GC2 and MQG model due to the energy correction in the variance equations, while for the less accurate GC1 model the improvement is limited due to the lack of precision in this closure strategy. Observing the errors in signal and dispersion part separately for the original model in the second row, it can be found that large inherent information barrier (especially for the mean prediction) exists for improving the model prediction skill no matter how well we tune the parameters in the training phase. On the other hand, for the corrected model results shown in the third row, the information barrier can be overcome with the signal error effectively reduced. We also compare the optimal response operators for the mean and resolved variances in Figure 10 for both original model and the corrected one. Better agreement with the truth can be observed for both parts by using the corrected model.

5.2.2 Testing imperfect model prediction skill in principal directions

Again we check the reduced models' ability in capturing the most important statistics in response to various external forcing perturbations. The setting-ups for the L-96 system is the same as before in the full model case but we will only run the models in the subspace spanned by $\{\mathbf{v}_0, \mathbf{v}_1\}$. The same four different types of perturbation forms are checked as before in Figure 4. Improvement for all the four cases in predicting the statistical mean and variance in the leading mode $\mathbf{v}_1$ through the corrected model can be observed in Figure 11-14. The results from reduced order GC1, GC2, and MQG are shown separately in the first three rows and the following three rows display the difference in full information distance as well as the errors in signal part and dispersion part separately for these three methods. As expected, GC1 always gives the smallest improvement for all cases since we can do nothing about improving its variance equations. Especially observe that there is nearly no improvement for the dispersion error in the reduced GC1 model, compared with the large improvement from GC2 and MQG results (This emphasizes the crucial role of energy equation correction in (51) by comparing GC1 and GC2). Improvements in GC2 and MQG model are larger with better agreement in both mean and principal variance predictions. Surprisingly, GC2

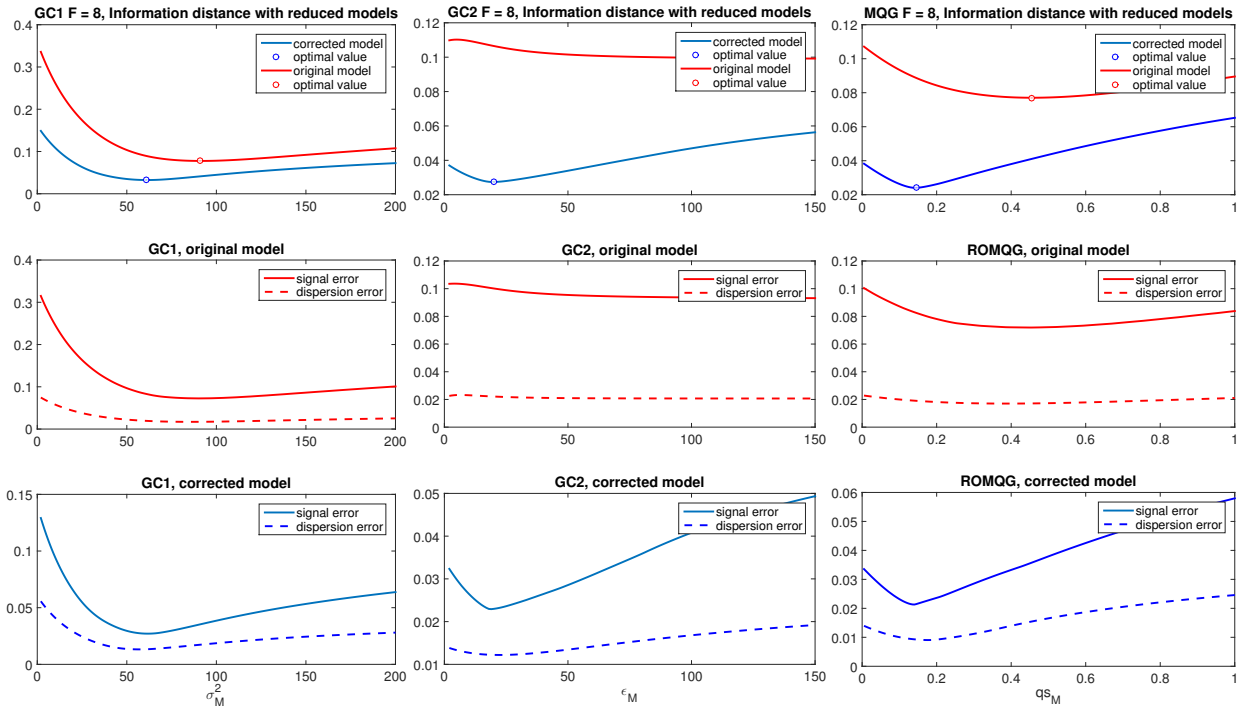


Fig. 9: Tuning imperfect model parameters in the training phase within a reduced subspace. The time-averaged information distances for reduced order GC1 (first column), GC2 (second column), MQG (third column) are compared inside the subspace spanned by the base mode $\mathbf{v}_0$ and most energetic mode $\mathbf{v}_1$. The improvement from the correction strategies for each method is shown by comparing the the original reduced order schemes. The information error for the signal part and dispersion part are also compared individually for each model in the second and third row.

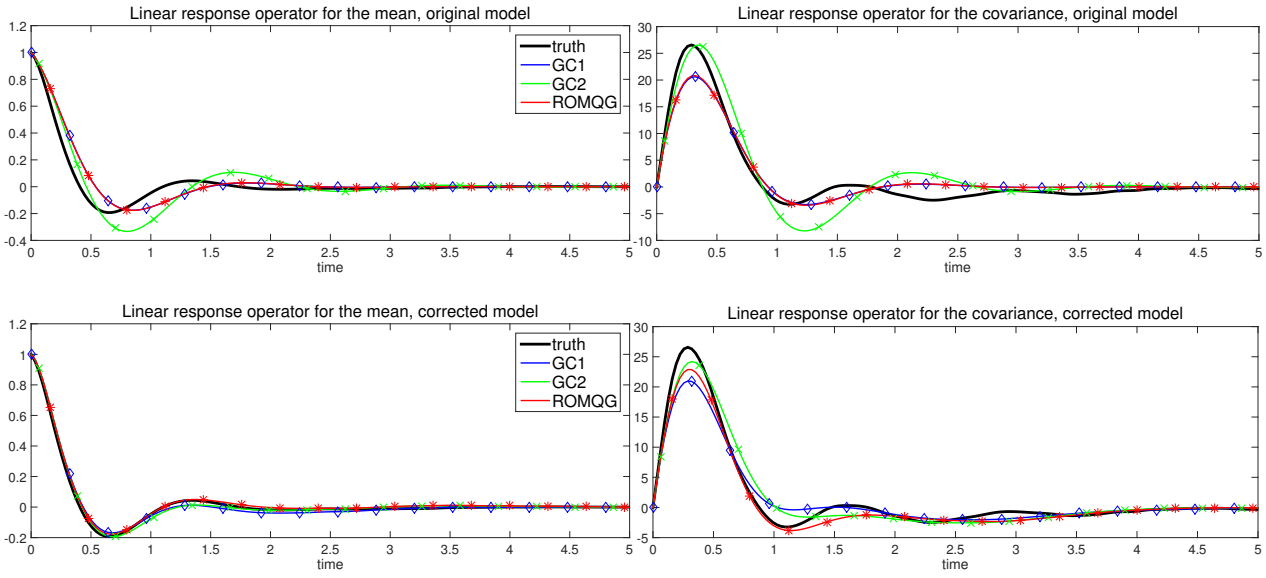
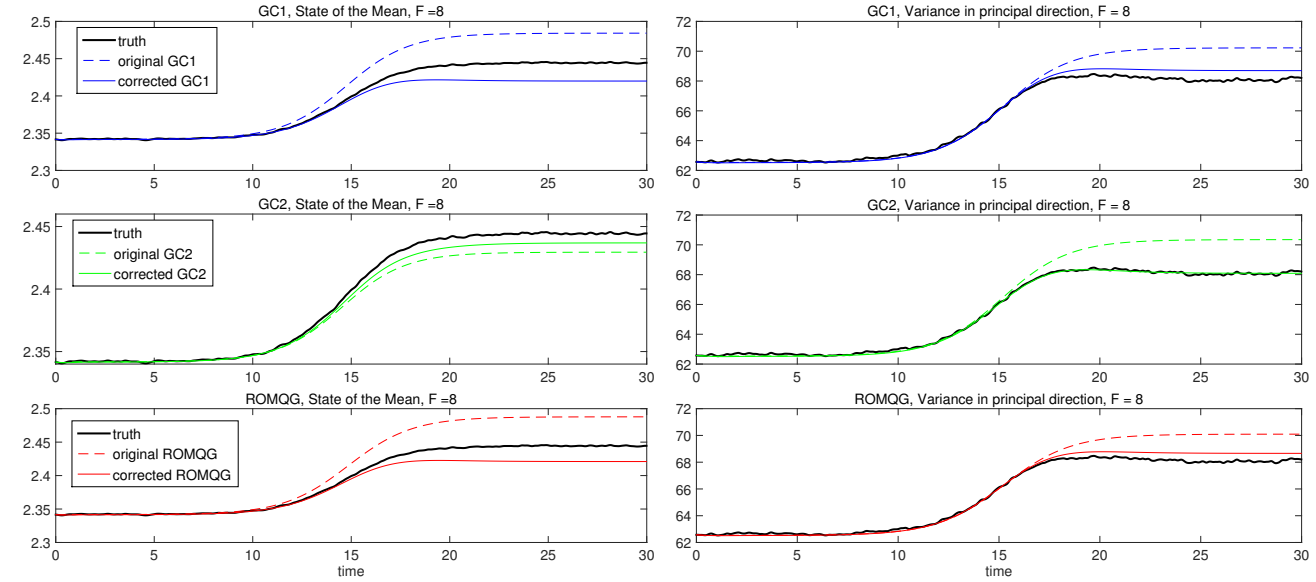
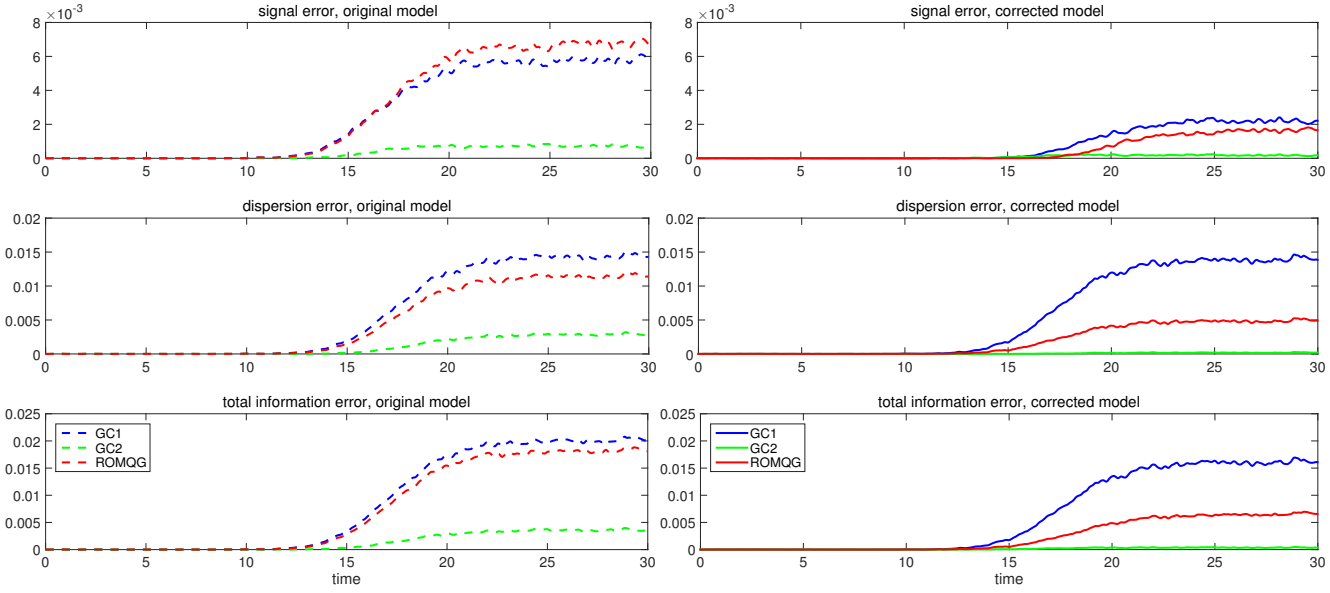


Fig. 10: Comparison of the optimal linear response operators for the mean (first column) and variance (second column) in $\mathbf{v}_0, \mathbf{v}_1$ using closure different closure models. The improvements by adopting the correction strategies can be seen by comparing the optimal fitting for and original method (first row) and the corrected method (second row).



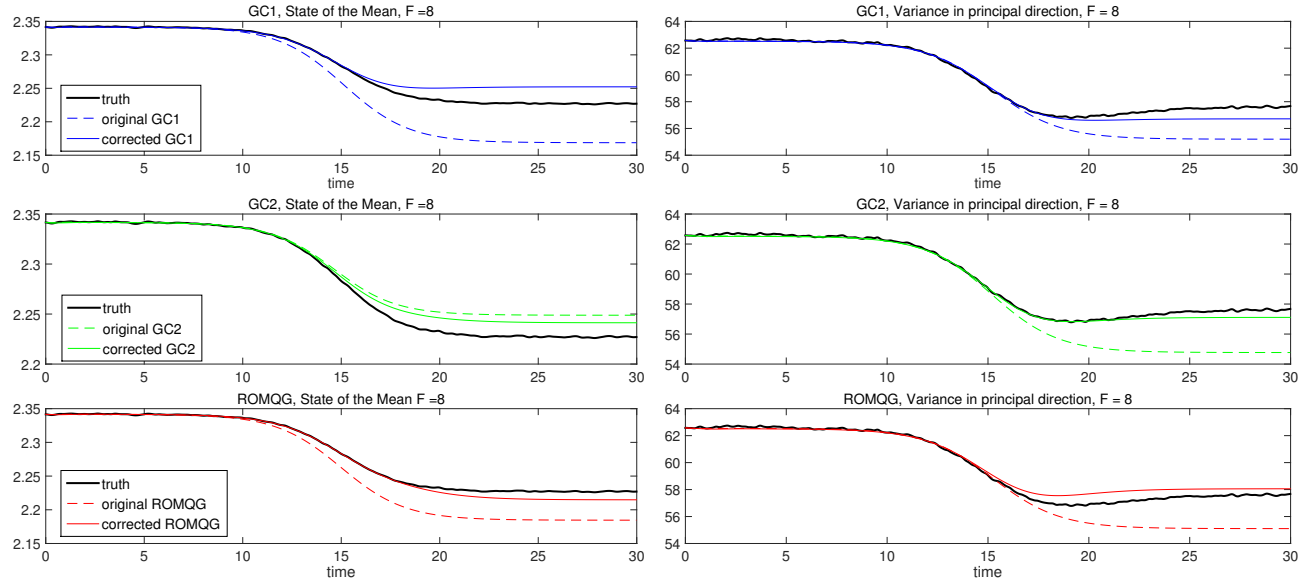
(a) Model predictions for the mean and variance



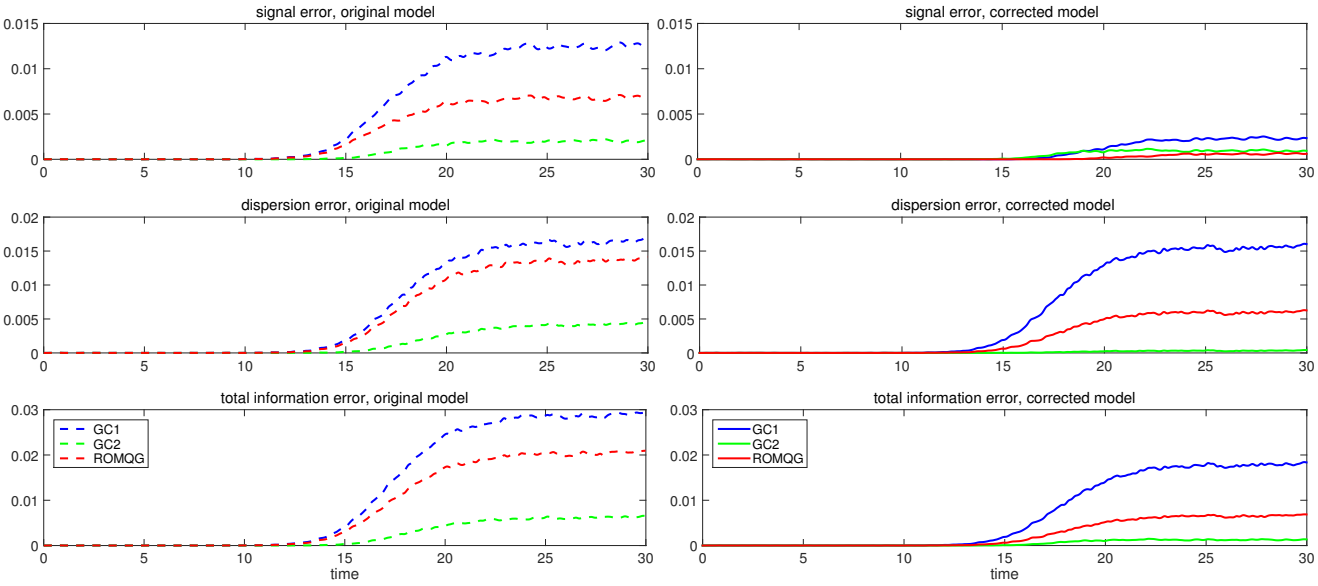
(b) Information errors

Fig. 11: Reduced model prediction for upward ramp-type forcing. Only the subspace spanned by the base mode and most energetic mode $\{\mathbf{v}_0, \mathbf{v}_1\}$ is resolved under the 40-dimensional L-96 model. The first three rows compare mean and variance predictions from the three reduced order closure methods GC1 (blue), GC2 (green), MQG (red) separately. The improvement from corrected model (solid) in comparison with the original method (dashed) is displayed. For further illustration for the improvement, the next three rows show the total information error measured in the resolved subspace as well as the signal and dispersion part for all three closure methods using original reduced order strategy (left, dashed) and the corrected strategy with energy and linear response corrections (right, solid).

can even offer better prediction in the reduced order case than ROMQG model considering that it's also cheaper in computation. Finally it is useful to emphasize that even though the prediction skill in GC1 is relatively poor among all the three cases, it can still serve as an effective strategy when the requirement for the accuracy is not that high, especially considering that it is the cheapest and easiest model to construct for all the situations.



(a) Model predictions for the mean and variance

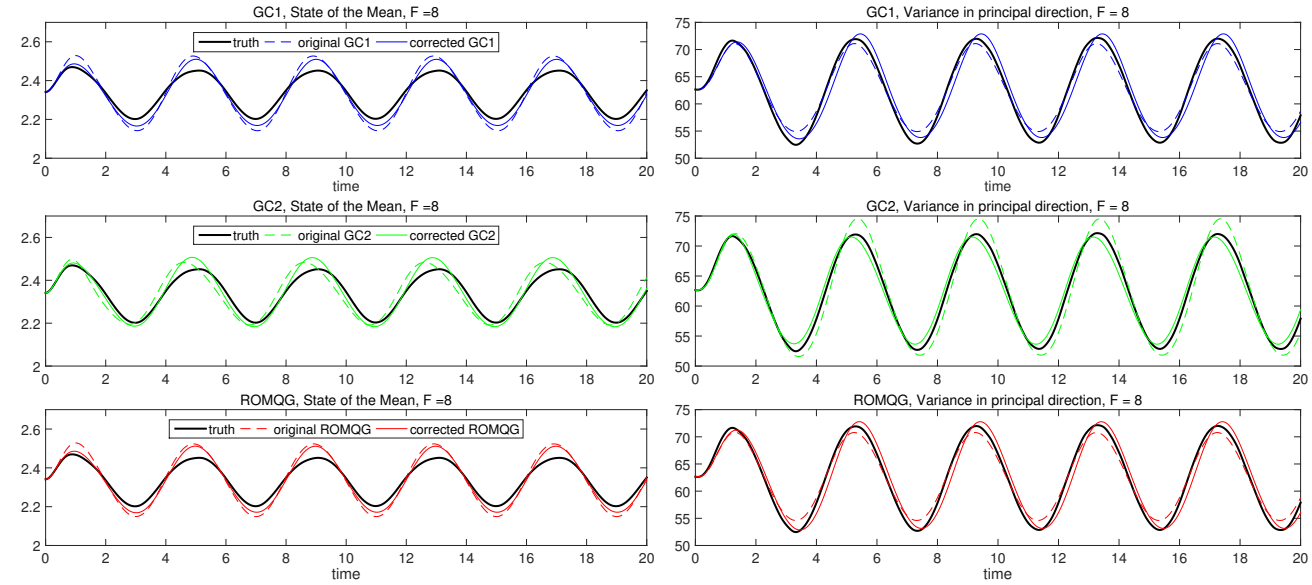


(b) Information errors

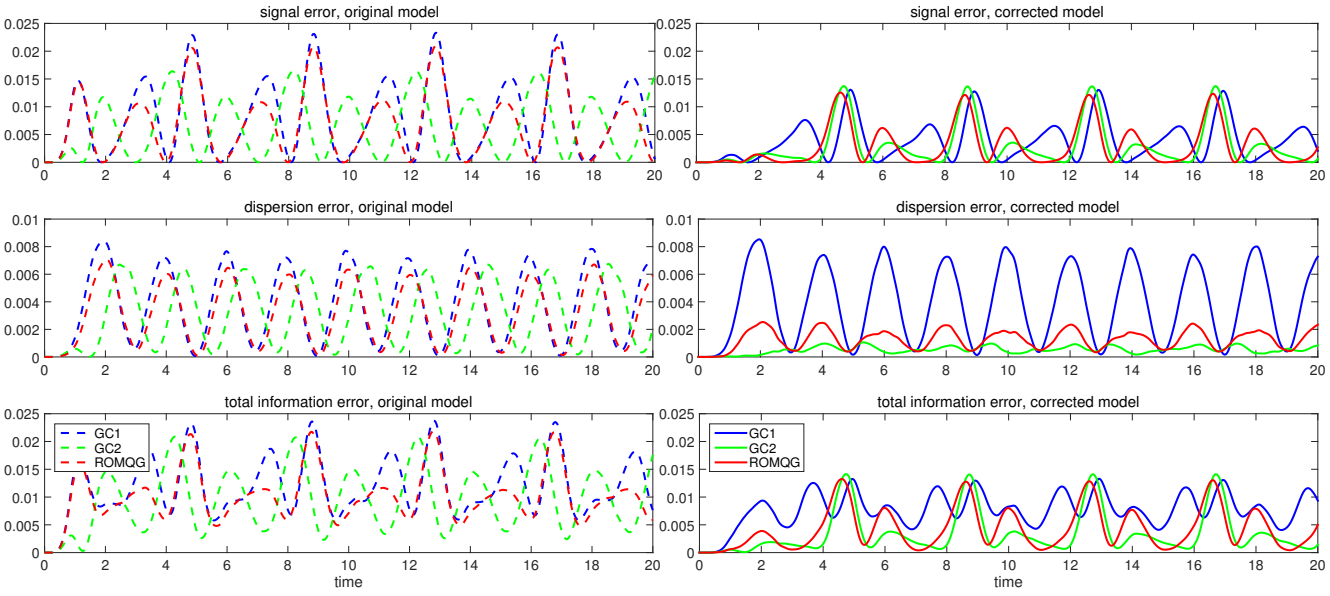
Fig. 12: Reduced model prediction for downward ramp-type forcing. Only the subspace spanned by the base mode and most energetic mode $\{\mathbf{v}_0, \mathbf{v}_1\}$ is resolved under the 40-dimensional L-96 model. The first three rows compare mean and variance predictions from the three reduced order closure methods GC1 (blue), GC2 (green), MQG (red) separately. The improvement from corrected model (solid) in comparison with the original method (dashed) is displayed. For further illustration for the improvement, the next three rows show the total information error measured in the resolved subspace as well as the signal and dispersion part for all three closure methods using original reduced order strategy (left, dashed) and the corrected strategy with energy and linear response corrections (right, solid).

5.2.3 Mean correction for high dimensional systems

As a final comment, we discuss a little more about the case when realistic high dimensional dynamical systems are applied. As pointed out in Remark 2 in Section 5.1, the dimensionality of the unresolved subspace could become so large that even calculating all the unresolved linear response variances becomes impractical. In that case, a further simplification is just to replace parts of the unresolved variances using the linear response correction, and leave the rest with small energy for original steady state ideas with total energy correction. Therefore, in the most



(a) Model predictions for the mean and variance



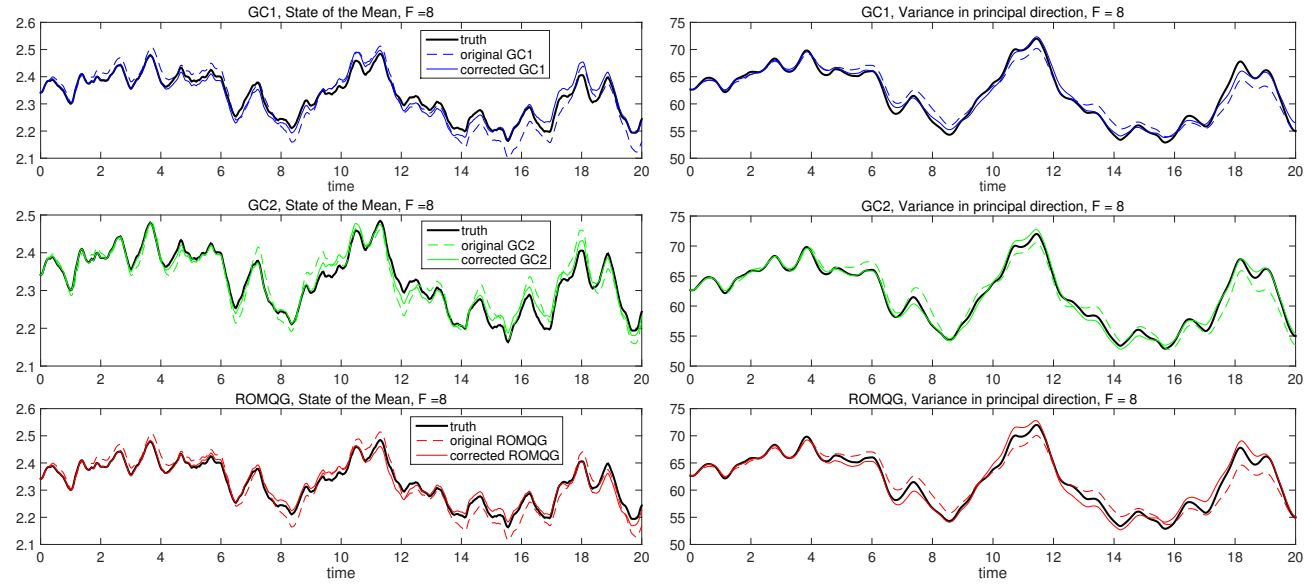
(b) Information errors

Fig. 13: Reduced model prediction for periodic forcing. Only the subspace spanned by the base mode and most energetic mode $\{\mathbf{v}_0, \mathbf{v}_1\}$ is resolved under the 40-dimensional L-96 model. The first three rows compare mean and variance predictions from the three reduced order closure methods GC1 (blue), GC2 (green), MQG (red) separately. The improvement from corrected model (solid) in comparison with the original method (dashed) is displayed. For further illustration for the improvement, the next three rows show the total information error measured in the resolved subspace as well as the signal and dispersion part for all three closure methods using original reduced order strategy (left, dashed) and the corrected strategy with energy and linear response corrections (right, solid).

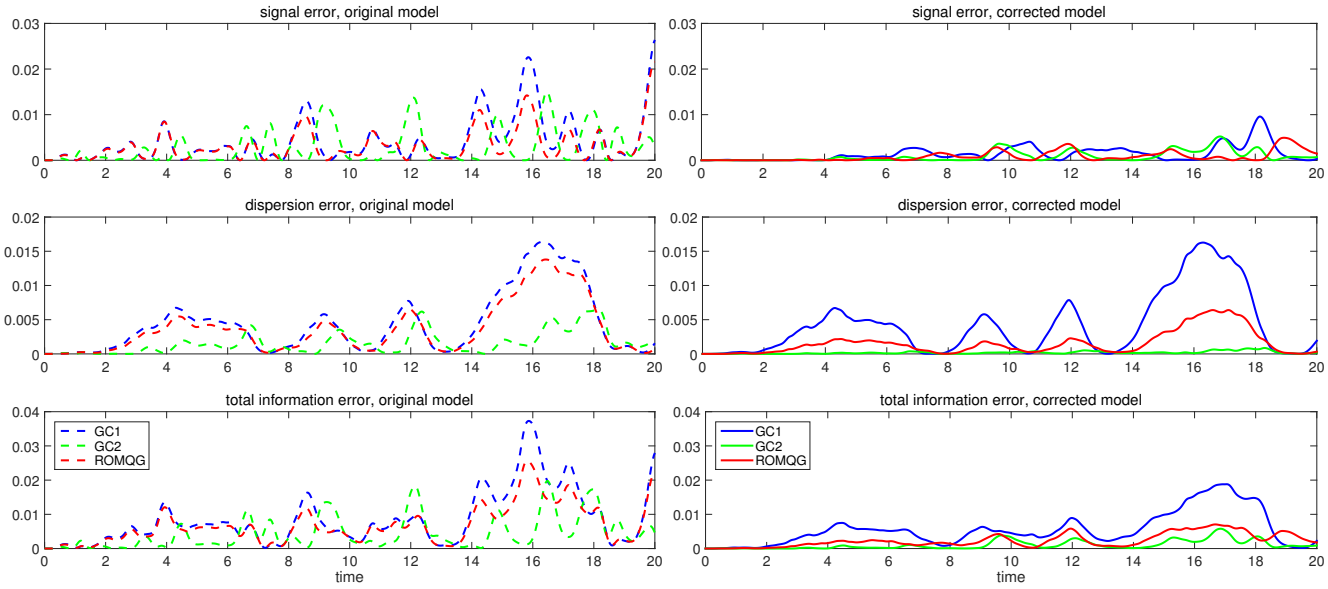
complicated case, the mean correction term G may take the form

$$G \equiv G_\infty + \delta G_K = \frac{1}{J}\sum_{k=s+1}^{K}\left(r_{k,\infty}+\delta r_k\right)\Gamma_k + \frac{1}{J}\sum_{k=K+1}^{J} r_{k,\infty}\Gamma_k \frac{f(R)}{f(R_\infty)}. \tag{53}$$

In this formula, the mean correction for unresolved modes δG is further decomposed into two parts. The first part up to a proper wavenumber cutoff K with relatively more importance in variance is still corrected using the linear response theory, while the rest of the less important modes returns to the original approximation with only steady state information with only a total variance scale factor included.



(a) Model predictions for the mean and variance



(b) Information errors

Fig. 14: Reduced model prediction for random forcing. Only the subspace spanned by the base mode and most energetic mode $\{\mathbf{v}_0, \mathbf{v}_1\}$ is resolved under the 40-dimensional L-96 model. The first three rows compare mean and variance predictions from the three reduced order closure methods GC1 (blue), GC2 (green), MQG (red) separately. The improvement from corrected model (solid) in comparison with the original method (dashed) is displayed. For further illustration for the improvement, the next three rows show the total information error measured in the resolved subspace as well as the signal and dispersion part for all three closure methods using original reduced order strategy (left, dashed) and the corrected strategy with energy and linear response corrections (right, solid).

As one illustrative example, Figure 15 compares the skills of the reduced models with truncated number of linear response corrections K in (53). We use the previous case with upward ramp-type forcing, and use GC2 model with the same optimal parameter from the training phase. Therefore, it can be observed that little degeneracy for the prediction skill appears with linear corrections for the first 5 most energetic unresolved modes compared with the results with corrections for all unresolved modes. This shows the possibility that we need only correct the variances of the most important ones for applications in the really high dimensional models. Also for comparison, we show the results of corrected model retaining G_∞ but with $K = 0$ (and with correction from total energy equation for tr$R$) and the original model even without the energy correction. Poorer prediction skills of these models are

displayed illustrating the indispensable roles of these correction strategies (for both mean correction δG and energy correction tr$R$) in model improvements.

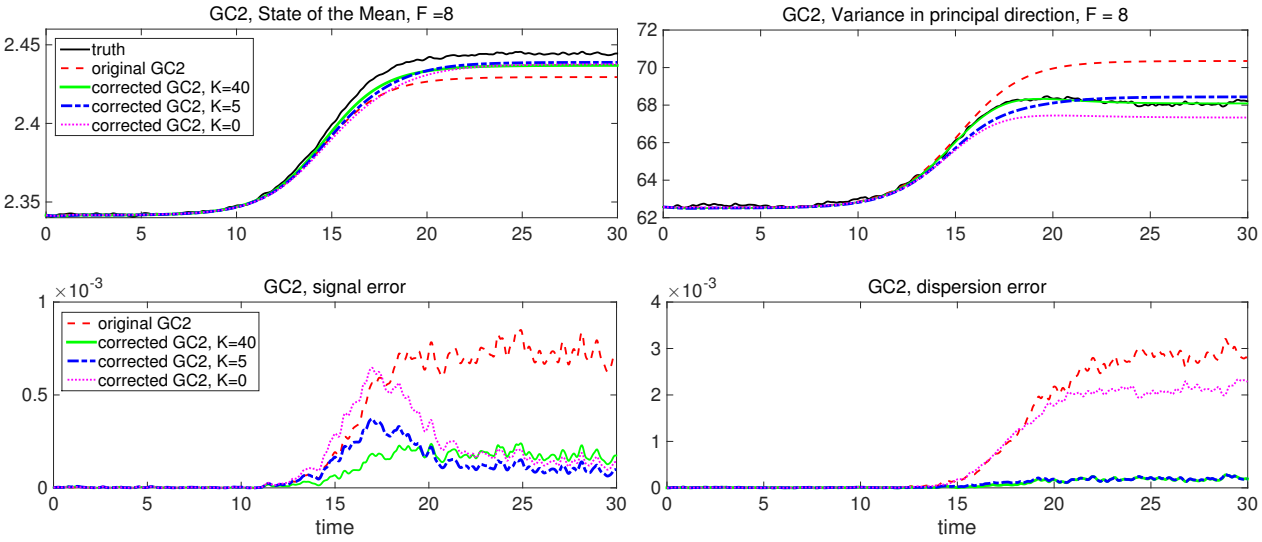


Fig. 15: Comparison of prediction skills with linear corrections for only first 5 unresolved modes in the mean dynamics. The upward ramp-type forcing and GC2 model with optimal parameter are used as the test case here. The corrected model results with linear correction for the mean dynamics using all the unresolved modes $K = 40$ as before are shown in green lines, while the further simplified model results with linear correction only for the first five most important modes $K = 5$ among all the unresolved modes are shown in blue dotted dashed lines, together with the original model (dashed red lines) and the model without mean correction $K = 0$ (dotted magenta lines). There is little degeneracy appearing in the further simplification method with fewer unresolved modes $K = 5$ using linear correction. This shows the possibility that we need only correct the variances of the most important ones for applications in the really high dimensional models.

6 Conclusion and future work

Imperfect models with statistical equilibrium fidelity may still suffer inherent information barrier for capturing correct responses to external perturbations in turbulent dynamical systems. We display a generic systematic framework combining the merits from linear response theory and empirical information theory (Majda and Gershgorin (2011a,b); Branicki and Majda (2012)) to improve the imperfect model sensitivity to various perturbation forms. The advantage is that the optimal parameters can be achieved in a training phase only according to the unperturbed statistics, therefore the model parameters don't need to be tuned each time for every specific perturbation form which is rather impractical for realistic predictions for changing external perturbation terms. Here to get deeper understanding about the central characterization about the nonlinear turbulent systems, we begin with models with simple structures and focus on second-order closure schemes in the first place. The improvement in the second-order closure models is checked under the homogeneous 40-dimensional L-96 system, which offers a desirable testbed with tractable statistical features and conservative nonlinear interaction. We further consider proper reduced order strategies which have improved ability in capturing the most important statistics in the leading EOFs when the dimensionality of the systems increases. Corrections to the unresolved statistics are proposed according to the statistical features of the dynamical system, and the same optimization framework is applied to the reduced order models to get the optimal response in the resolved subspace. Several important points can be concluded from the theoretical analysis and numerical tests using the L-96 testbed:

- The second-order statistical closure models outperform the linear FDT predictions for capturing responses to external perturbations, especially in regimes with larger perturbations and stronger nonlinearities. This shows the importance and necessity of adopting dynamical methods in getting correct model sensitivity for UQ predictions;
- Single point statistics which concentrate on the statistics on each grid point and ignore cross-correlations between points are useful quantities for illustrating the basic statistical features of the turbulent systems. Imperfect

statistical closure models with symmetry in nonlinear energy can predict accurate single point variance once accurate prediction for the mean is achieved. This is an important result showing that higher order moments can be determined by the lower order approximations and offers important guideline for designing imperfect closure schemes;
- Still accurate single point statistics prediction is not sufficient for the imperfect models to break information barriers (Majda and Gershgorin (2011a,b); Majda and Branicki (2012)). An imperfect model with perfect recovery of the single point statistics may still display large errors in the predictions for statistics in each EOF mode;
- The information-response framework by tuning imperfect models in the training phase for optimal model response operators shows promising skill in breaking information barriers in model sensitivity to perturbations. Imperfect model prediction skill can be improved uniformly regardless of the specific perturbation form applied;
- It is important for practical applications that the information-response framework can also be applied systematically to reduced order models which focus on capturing the uncertainties in the dominant modes. Importantly, the prediction skill in reduced order methods can be improved effectively by introducing corrections from a simple scalar energy equation and linear corrections for the unresolved modes.

Within this paper, we concentrate on the specific L-96 system in order to have a cleaner dynamical core containing only essential structure for analysis and construction of models. The L-96 model is quite representative for a wide variety of systems with large number of instabilities and conservative nonlinear operators. It is useful to move forward to realistic high dimensional systems like the general circulation models (GCMs) and check the effectiveness of this framework and the model reduction strategies. Blended reduced order schemes (Sapsis and Majda (2013a,b)) have been developed that offer a promising direction for capturing principal higher order statistics with efficiency. Furthermore, in many areas of applications, it is useful to incorporate partial observation data from the natural system with the imperfect model prediction for combined improvement. That is, we need to develop proper filtering or data assimilation strategies for models with high dimensional phase space and strong nonlinearity. So it is also interesting to combine the UQ schemes developed here with the low order blended filtering schemes we have developed previously in Majda et al (2014); Qi and Majda (2015). This is a promising direction for real time filtering with the skill in accurate capturing of the nonlinear non-Gaussian features of the high dimensional turbulent systems.

Acknowledgements The research of Andrew Majda is partially supported by Office of Naval Research grant ONR MURI N00014-12-1-0912. Di Qi is supported as a graduate research assistant on this grant.

References

Abramov RV, Majda AJ (2007) Blended response algorithms for linear fluctuation-dissipation for complex nonlinear dynamical systems. Nonlinearity 20(12):2793

Abramov RV, Majda AJ (2008) New approximations and tests of linear fluctuation-response for chaotic nonlinear forced-dissipative dynamical systems. Journal of Nonlinear Science 18(3):303–341

Abramov RV, Majda AJ (2009) A new algorithm for low-frequency climate response. Journal of the Atmospheric Sciences 66(2):286–309

Bell TL (1980) Climate sensitivity from fluctuation dissipation: Some simple model tests. Journal of the Atmospheric Sciences 37(8):1700–1707

Bengtsson T, Bickel P, Li B, et al (2008) Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems. In: Probability and statistics: Essays in honor of David A. Freedman, Institute of Mathematical Statistics, pp 316–334

Branicki M, Majda AJ (2012) Quantifying uncertainty for predictions with model error in non-gaussian systems with intermittency. Nonlinearity 25(9):2543

Branicki M, Chen N, Majda AJ (2013) Non-gaussian test models for prediction and state estimation with model errors. Chinese Annals of Mathematics, Series B 34(1):29–64

Carnevale G, Falcioni M, Isola S, Purini R, Vulpiani A (1991) Fluctuation-response relations in systems with chaotic behavior. Physics of Fluids A: Fluid Dynamics (1989-1993) 3(9):2247–2254

DelSole T (2005) Predictability and information theory. part ii: Imperfect forecasts. Journal of the atmospheric sciences 62(9):3368–3381

DelSole T, Shukla J (2010) Model fidelity versus skill in seasonal forecasting. Journal of Climate 23(18):4794–4806

Emanuel K, Wyngaard J, McWilliams J, Randall D, Yung Y (2005) Improving the scientific foundation for atmosphere-land-ocean simulations. Natl Acad Press, Washington DC p 72

Epstein ES (1969) Stochastic dynamic prediction1. Tellus 21(6):739–759

Gershgorin B, Majda AJ (2012) Quantifying uncertainty for climate change and long-range forecasting scenarios with model errors. part i: Gaussian models. Journal of Climate 25(13):4523–4548

Gritsun A, Branstator G (2007) Climate response using a three-dimensional operator based on the fluctuation-dissipation theorem. Journal of the atmospheric sciences 64(7):2558–2575

Gritsun A, Branstator G, Majda A (2008) Climate response of linear and quadratic functionals using the fluctuation-dissipation theorem. Journal of the Atmospheric Sciences 65(9):2824–2841

Hairer M, Majda AJ (2010) A simple framework to justify linear response theory. Nonlinearity 23(4):909

Kullback S, Leibler RA (1951) On information and sufficiency. The annals of mathematical statistics pp 79–86

Leith C (1975) Climate response and fluctuation dissipation. Journal of the Atmospheric Sciences 32(10):2022–2026
Lorenz EN (1996) Predictability: A problem partly solved. In: Proc. Seminar on predictability, vol 1
Majda A (2003) Introduction to PDEs and Waves for the Atmosphere and Ocean, vol 9. American Mathematical Soc.
Majda A, Wang X (2006) Nonlinear dynamics and statistical theories for basic geophysical flows. Cambridge University Press
Majda A, Kleeman R, Cai D (2002) A mathematical framework for quantifying predictability through relative entropy. Methods and Applications of Analysis 9(3):425–444
Majda A, Abramov RV, Grote MJ (2005) Information theory and stochastics for multiscale nonlinear systems, vol 25. American Mathematical Soc.
Majda A, Wang X, et al (2010a) Linear response theory for statistical ensembles in complex systems with time-periodic forcing. Communications in Mathematical Sciences 8(1):145–172
Majda AJ, Branicki M (2012) Lessons in uncertainty quantification for turbulent dynamical systems. Discrete Cont Dyn Systems 32(9)
Majda AJ, Gershgorin B (2010) Quantifying uncertainty in climate change science through empirical information theory. Proceedings of the National Academy of Sciences 107(34):14,958–14,963
Majda AJ, Gershgorin B (2011a) Improving model fidelity and sensitivity for complex systems through empirical information theory. Proceedings of the National Academy of Sciences 108(25):10,044–10,049
Majda AJ, Gershgorin B (2011b) Link between statistical equilibrium fidelity and forecasting skill for complex systems with model error. Proceedings of the National Academy of Sciences 108(31):12,599–12,604
Majda AJ, Abramov R, Gershgorin B (2010b) High skill in low-frequency climate response through fluctuation dissipation theorems despite structural instability. Proceedings of the National Academy of Sciences 107(2):581–586
Majda AJ, Gershgorin B, Yuan Y (2010c) Low-frequency climate response and fluctuation-dissipation theorems: theory and practice. Journal of the Atmospheric Sciences 67(4):1186–1201
Majda AJ, Qi D, Sapsis TP (2014) Blended particle filters for large-dimensional chaotic dynamical systems. Proceedings of the National Academy of Sciences 111(21):7511–7516
Neelin J, Münnich M, Su H, Meyerson J, Holloway C (2006) Tropical drying trends in global warming models and observations. Proceedings of the National Academy of Sciences 103(16):6110–6115
Qi D, Majda AJ (2015) Blended particle methods with adaptive subspaces for filtering turbulent dynamical systems. Physica D: Nonlinear Phenomena 298:21–41
Sapsis TP, Majda AJ (2013a) Blended reduced subspace algorithms for uncertainty quantification of quadratic systems with a stable mean state. Physica D: Nonlinear Phenomena 258:61–76
Sapsis TP, Majda AJ (2013b) Blending modified gaussian closure and non-gaussian reduced subspace methods for turbulent dynamical systems. Journal of Nonlinear Science 23(6):1039–1071
Sapsis TP, Majda AJ (2013c) Statistically accurate low-order models for uncertainty quantification in turbulent dynamical systems. Proceedings of the National Academy of Sciences 110(34):13,705–13,710
Sapsis TP, Majda AJ (2013d) A statistically accurate modified quasilinear gaussian closure for uncertainty quantification in turbulent dynamical systems. Physica D: Nonlinear Phenomena 252:34–45
Tung S (1975) On lower and upper bounds of the difference between the arithmetic and the geometric mean. Mathematics of Computation 29(131):834–836
Yaglom AM (2004) An introduction to the theory of stationary random functions. Courier Corporation

Appendix A: Derivation of the moment equations for L-96 system

Here we display the detailed derivations about the moment equations and the properties under homogeneous assumption in Section 3.1. Under the Fourier representation of the basis $\{\mathbf{v}_j\}$, we can write explicit formulas for each part of the L-96 system (16) as follows

1. The quadratic interaction between Fourier mode $\mathbf{v}_i$ and $\mathbf{v}_j$ can be written explicitly from the definition

$$\begin{aligned}\mathbf{B}(\mathbf{v}_i,\mathbf{v}_j) &= \left\{v_i^{l-1*}\left(v_j^{l+1}-v_j^{l-2}\right)\right\}_{l=0}^{J-1}\\ &= \left\{\frac{1}{J}e^{-2\pi\iota i\frac{l-1}{J}}\left(e^{2\pi\iota j\frac{l+1}{J}}-e^{-2\pi\iota j\frac{l-2}{J}}\right)\right\}_{l=0}^{J-1}\\ &= \frac{1}{\sqrt{J}}e^{2\pi\iota\frac{i+j}{J}}\mathbf{v}_{j-i}-\frac{1}{\sqrt{J}}e^{2\pi\iota\frac{i-2j}{J}}\mathbf{v}_{j-i}\\ &= \frac{1}{\sqrt{J}}e^{2\pi\iota\frac{j}{J}}\left(e^{2\pi\iota\frac{j}{J}}-e^{2\pi\iota\frac{-2j}{J}}\right)\mathbf{v}_{j-i}.\end{aligned}$$

   Particularly, we can find that the nonlinear interaction term vanishes if and only if its second component is only under the zero base mode $\mathbf{v}_0=\frac{1}{\sqrt{J}}(1,\cdots,1)^T$, that is,

$$\mathbf{B}(\mathbf{v}_i,\mathbf{v}_j)=\mathbf{0}\Leftrightarrow e^{2\pi\iota\frac{3j}{J}}=1\Leftrightarrow j=\frac{nJ}{3},\ n\in\mathbb{N},\ J=40\Leftrightarrow j=0.$$

2. Using the expression above, we can calculate the explicit formulas for the second-order interaction form in the mean dynamics

$$\begin{aligned}\sum_{i,j} R_{ij}\mathbf{B}\left(\mathbf{v}_i,\mathbf{v}_j\right) &= \frac{1}{\sqrt{J}}\sum_{i,j} r_{ij}e^{2\pi\iota\frac{i}{J}}\left(e^{2\pi\iota\frac{j}{J}}-e^{2\pi\iota\frac{-2j}{J}}\right)\mathbf{v}_{j-i}\\ &= \left\{\frac{1}{J}\sum_{i,j} r_{ij}e^{-2\pi\iota i\frac{l-1}{J}}\left(e^{2\pi\iota j\frac{l+1}{J}}-e^{-2\pi\iota j\frac{l-2}{J}}\right)\right\}_{l=0}^{J-1}.\end{aligned}$$

Then to calculate the explicit formula for the covariance dynamics, we need to get the linear interaction part L_v first

$$\mathbf{B}\left(\bar{\mathbf{u}},\mathbf{v}_j\right)_l = \frac{1}{\sqrt{J}}\bar{u}_{l-1}\left(e^{2\pi\iota j\frac{l+1}{J}}-e^{2\pi\iota j\frac{l-2}{J}}\right),$$

$$\mathbf{B}\left(\mathbf{v}_j,\bar{\mathbf{u}}\right)_l = \frac{1}{\sqrt{J}}e^{-2\pi\iota j\frac{l-1}{J}}\left(\bar{u}_{l+1}-\bar{u}_{l-2}\right).$$

Therefore by definition, we get

$$\begin{aligned}L_{v,ij} &= \left(\mathbf{L}\left(t\right)\mathbf{v}_j+\mathbf{B}\left(\bar{\mathbf{u}},\mathbf{v}_j\right)+\mathbf{B}\left(\mathbf{v}_j,\bar{\mathbf{u}}\right)\right)\cdot\mathbf{v}_i^*\\ &= -d_j\left(t\right)\delta_{ij}+\frac{1}{J}\left(e^{2\pi\iota\frac{j}{J}}-e^{-2\pi\iota\frac{2j}{J}}\right)\sum_l \bar{u}_{l-1}e^{2\pi\iota l\frac{j-i}{J}}\\ &\quad+\frac{1}{J}e^{2\pi\iota\frac{j}{J}}\sum_l\left(\bar{u}_{l+1}-\bar{u}_{l-2}\right)e^{-2\pi\iota\frac{j+i}{J}}.\end{aligned}$$

where the index l represents the l-th component of a vector.

3. Finally we need to calculate the formula for the nonlinear flux term Q_F in the covariance dynamics. Again use the explicit formula for the nonlinear interaction term between modes $\mathbf{B}\left(\mathbf{v}_i,\mathbf{v}_j\right)$ and the definition of the nonlinear flux, we have

$$\begin{aligned}Q_{F,ij} &= \sum_{m,n}\left\langle Z_mZ_n^*Z_j\right\rangle\mathbf{B}\left(\mathbf{v}_m,\mathbf{v}_n\right)^*\cdot\mathbf{v}_i+\left\langle Z_m^*Z_nZ_i^*\right\rangle\mathbf{B}\left(\mathbf{v}_m,\mathbf{v}_n\right)\cdot\mathbf{v}_j^*\\ &= \frac{1}{\sqrt{J}}\sum_{m,n}\left\langle Z_mZ_n^*Z_j\right\rangle e^{-2\pi\iota\frac{m}{J}}\left(e^{-2\pi\iota\frac{n}{J}}-e^{2\pi\iota\frac{2n}{J}}\right)\delta_{n-m,i}\\ &\quad+\frac{1}{\sqrt{J}}\sum_{m,n}\left\langle Z_m^*Z_nZ_i^*\right\rangle e^{2\pi\iota\frac{m}{J}}\left(e^{2\pi\iota\frac{n}{J}}-e^{-2\pi\iota\frac{2n}{J}}\right)\delta_{n-m,j}.\end{aligned}$$

Therefore, we get the explicit forms for each part of the moment equations for the mean and covariance matrix.

Another important issue is due to the simplification of each order of moments under the homogeneous assumption of the L-96 system. Specifically, it tells that if each order of moments is invariant under shifting in grid points

$$\left\langle u_{i_1}u_{i_2}\cdots u_{i_n}\right\rangle = \left\langle u_{i_1+l}u_{i_2+l}\cdots u_{i_n+l}\right\rangle,\quad\forall l,$$

then the first three moments under the Fourier basis become

$$\begin{aligned}\bar{\mathbf{u}}\left(t\right) &= \bar{u}\left(t\right)\left(1,1,...,1\right)^T,\\ R\left(t\right) &= \mathrm{diag}\left(r_{-J/2+1}\left(t\right),\cdots,r_0\left(t\right),...,r_{J/2}\left(t\right)\right),\\ \left\langle Z_iZ_jZ_k\right\rangle &\neq 0,\quad \text{only if } i+j+k=0.\end{aligned}$$

The first equation for the mean state $\bar{\mathbf{u}}$ is direct from the definition. To get the simplified forms for the second and third moments R and $\left\langle Z_iZ_jZ_k\right\rangle$, first we write the Fourier transform as a change of basis as

$$\begin{aligned}\mathbf{u}' &= V\mathbf{Z},\\ Z_j &= \sum_l \mathbf{v}_j^{l*}u_l'.\end{aligned}$$

where $\mathbf{u}'=\mathbf{u}-\bar{\mathbf{u}}=\left(u_1',u_2',\cdots,u_J'\right)^T$, $\mathbf{Z}=\left(Z_{-J/2+1},Z_{-J/2+2},\cdots,Z_{J/2}\right)^T$ are the coefficients under natural basis and Fourier basis separately, and $V=\left(\mathbf{v}_{-J/2+1},\mathbf{v}_{-J/2+2},\cdots,\mathbf{v}_{J/2}\right)$ is the transformation matrix formed by Fourier basis.

Therefore, the components of the second order moments under the transform invariant property become

$$\begin{aligned} R_{ij} &= \left\langle Z_i Z_j^* \right\rangle = \sum_{m,n} v_i^{m*} \left\langle u_m' u_n' \right\rangle v_j^n \\ &= \sum_n \sum_l \left\langle u_0' u_n' \right\rangle v_i^{l*} v_j^{n+l} \\ &= \sum_n \left\langle u_0' u_n' \right\rangle e^{2\pi\iota \frac{nj}{J}} \sum_l e^{2\pi\iota l \frac{j-i}{J}} \\ &= J \sum_n \left\langle u_0' u_n' \right\rangle e^{2\pi\iota \frac{nj}{J}} \delta_{ij}. \end{aligned}$$

Here the second equality is using the homogeneity in physical space $\langle u_m' u_n' \rangle = \langle u_0' u_{n-m}' \rangle$. Therefore we can see that the covariance matrix becomes diagonal under the homogeneous assumption and each component takes the form

$$r_j = J \sum_n \left\langle u_0' u_n' \right\rangle e^{2\pi\iota \frac{nj}{J}}.$$

In the same way, we can calculate the third moments as

$$\begin{aligned} \left\langle Z_i Z_j Z_k \right\rangle &= \sum_{m,n,s} \left\langle v_i^{m*} u_m' v_j^{n*} u_n' v_k^{s*} u_s' \right\rangle \\ &= \sum_{m,n,s} \left\langle u_m' u_n' u_s' \right\rangle v_i^{m*} v_j^{n*} v_k^{s*} \\ &= \sum_{n,s} \sum_l \left\langle u_0' u_n' u_s' \right\rangle v_i^{l*} v_j^{n+l*} v_k^{s+l*} \\ &= \sum_{n,s} \left\langle u_0' u_n' u_s' \right\rangle e^{-2\pi\iota \frac{nj+ks}{J}} \sum_l e^{2\pi\iota l \frac{i+j+k}{J}} \\ &= J \sum_{n,s} \left\langle u_0' u_n' u_s' \right\rangle e^{-2\pi\iota \frac{nj+ks}{J}} \delta_{i+j+k}. \end{aligned}$$

The same we use the homogeneous property $\langle u_m' u_n' u_s' \rangle = \langle u_0' u_{n-m}' u_{s-m}' \rangle$. Therefore the third order moments can only be non-zeros when the wavenumber satisfies $i+j+k=0$, and the non-zeros terms can be represented as

$$\left\langle Z_{-j-k} Z_j Z_k \right\rangle = J \sum_{n,s} \left\langle u_0' u_n' u_s' \right\rangle e^{-2\pi\iota \frac{nj+ks}{J}}.$$

With all these homogeneous properties, both the linear and nonlinear flux terms L_v and Q_F become diagonal. The detailed calculation is shown as follows. First, each part of the linear interaction term L_v simplifies

$$\sum_{i,j} R_{ij} \mathbf{B}\left(\mathbf{v}_i, \mathbf{v}_j\right) = \sum_i r_i(t) \mathbf{B}\left(\mathbf{v}_i, \mathbf{v}_i\right) = \frac{1}{\sqrt{J}} \sum_i r_i \left(e^{2\pi\iota \frac{2i}{J}} - e^{-2\pi\iota \frac{i}{J}} \right) \mathbf{v}_0,$$

$$\mathbf{B}\left(\mathbf{v}_i, \bar{\mathbf{u}}\right) = \bar{u}(t) \mathbf{B}\left(\mathbf{v}_i, \mathbf{v}_0\right) = \mathbf{0},$$

$$\mathbf{B}\left(\bar{\mathbf{u}}, \mathbf{v}_i\right) = \bar{u}(t) \mathbf{B}\left(\mathbf{v}_0, \mathbf{v}_i\right) = \bar{u}(t) \left(e^{2\pi\iota \frac{i}{J}} - e^{-2\pi\iota \frac{2i}{J}} \right) \mathbf{v}_i,$$

$$L_{v,ij} = -d(t) \delta_{ij} + \bar{u}(t) \left(e^{2\pi\iota \frac{j}{J}} - e^{-2\pi\iota \frac{2j}{J}} \right) \delta_{ij}.$$

Then the simplified form of the nonlinear flux term Q_F can also be calculated

$$\begin{aligned} Q_{F,ij} = & \frac{1}{\sqrt{J}} \sum_{m,n} \left\langle Z_m Z_n^* Z_j \right\rangle e^{-2\pi\iota \frac{m}{J}} \left(e^{-2\pi\iota \frac{n}{J}} - e^{2\pi\iota \frac{2n}{J}} \right) \delta_{n-m,i} \\ & + \frac{1}{\sqrt{J}} \sum_{m,n} \left\langle Z_m^* Z_n Z_i^* \right\rangle e^{2\pi\iota \frac{m}{J}} \left(e^{2\pi\iota \frac{n}{J}} - e^{-2\pi\iota \frac{2n}{J}} \right) \delta_{n-m,j} \\ = & \frac{1}{\sqrt{J}} \sum_{m,n} \left\langle Z_m Z_{-n} Z_j \right\rangle e^{-2\pi\iota \frac{m}{J}} \left(e^{-2\pi\iota \frac{n}{J}} - e^{2\pi\iota \frac{2n}{J}} \right) \delta_{n-m,i} \delta_{m-n+j} \\ & + \frac{1}{\sqrt{J}} \sum_{m,n} \left\langle Z_{-m} Z_n Z_{-i} \right\rangle e^{2\pi\iota \frac{m}{J}} \left(e^{2\pi\iota \frac{n}{J}} - e^{-2\pi\iota \frac{2n}{J}} \right) \delta_{n-m,j} \delta_{-m+n-i} \\ = & \frac{2}{\sqrt{J}} \sum_m \mathfrak{Re} \left\{ \left\langle Z_m Z_{-m-j} Z_j \right\rangle \left(e^{-2\pi\iota \frac{2m+j}{J}} - e^{2\pi\iota \frac{m+2j}{J}} \right) \right\} \delta_{ij}. \end{aligned}$$

Appendix B: Numerical strategies to calculate the kicked response operators

In Section 4, we use the kicked response theory to tune the imperfect model parameters in the training phase. Here we describe the details about calculating the kicked response operators for the mean and variance numerically. From the formula in (10), the response operators for the mean and variance can be achieved from the perturbation part of the probability density $\delta\pi'$. And this density function is also used to measure the information distance between the truth and imperfect model result in the training phase. Below we describe the numerical procedure to get this distribution function $\delta\pi'$ for the true system and the imperfect closure model separately.

- *Kicked response for the true model*: For the true system, we want to achieve the most accurate possible estimation for the response operators both for comparison with the imperfect model results and for calculating the FDT linear prediction. Therefore we use a Monte-Carlo simulation with an ensemble size of $10,000$ particles to capture the response in density. The initial equilibrium ensemble is picked by sampling from a normal distribution with consistent equilibrium mean and variance of the true system. For the kicked response to the mean a constant perturbation with 10 percent of the equilibrium state mean $\delta\mathbf{u} = 0.1\bar{\mathbf{u}}_\infty$ is added to each initial ensemble member (in fact, as observed in numerical experiments, this perturbation amplitude has little effect on the results in the response distribution as long as it's not too large); and the initial variance of the ensemble is kept unchanged. The response distribution $\delta\pi'$ then is achieved by monitoring the decay of the ensemble particles back to equilibrium under unperturbed dynamics and uniformly perturbed initial value (and the length of the time window that we need to monitor depends on the mixing property of the turbulent system). See Bell (1980); Majda et al (2005) for similar version of this alogrithm.
- *Kicked response for the imperfect model*: For the imperfect model, we just need to run the closure equations to get the responses for the mean and variance. In the same way as the true model, the initial mean is taken from the equilibrium distribution and a perturbation with amplitude $\delta\mathbf{u} = 0.1\bar{\mathbf{u}}_\infty$ is added to the mean initial state. The initial value for the variance is taken the same as the equilibrium state value and kept unperturbed. Then using this initial mean and variance, the imperfect model with specific closure strategies is applied to monitor the decay of the mean and variance back to equilibrium.

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

Appendix C: Calculating equilibrium consistent parameters for statistical closure models

The improved statistical closure models discussed in Section 4 require equilibrium fidelity at stationary steady state as $t \to \infty$ as a necessary condition. We proposed the climate consistent parameter values in (41) and (42) correspondingly for GC1 and GC2 models calculated through simple algebraic manipulations. In stationary steady state with uniform damping and forcing terms, each order of moments of the closure system (36) converges to the equilibrium state such that

$$\frac{d\bar{u}_{M,\infty}}{dt} = 0, \quad \frac{dr^M_{k,\infty}}{dt} = 0, \ \forall k.$$

Applying (36) in the corresponding steady states, the equilibrium mean and variance $\bar{u}_{M,\infty}$, $r^M_{k,\infty}$ satisfies

$$\begin{aligned} 0 &= -d\bar{u}_{M,\infty} + \frac{1}{J}\sum_{k=-J/2+1}^{J/2} r^M_{k,\infty}\Gamma_k + F, \\ 0 &= 2\left[-\Gamma_k\bar{u}_{M,\infty} - d\right] r^M_{k,\infty} + Q^M_{F,kk,\infty}, \quad k = 0, 1, ..., J/2, \end{aligned}$$

with nonlinear interaction term in steady state $Q^M_{F,kk,\infty} = -2d_{M,k}(R_\infty)\, r^M_{k,\infty} + \sigma^2_{M,k}(R_\infty)$. A similar procedure like that in Section 3 by summing up the modes in the second equations above and substituting the mean equation can be carried out so that

$$\sum_k Q^M_{F,kk,\infty} = 2\bar{u}_{M,\infty}\sum_k \Gamma_k r^M_{k,\infty} + 2d\mathrm{tr}R^M_\infty = 2J\bar{u}_{M,\infty}\left(d\bar{u}_{M,\infty} - F\right) + 2d\mathrm{tr}R^M_\infty.$$

The right hand side of the above relation between the equilibrium mean and total variance should equal to zero for climate consistent models. Furthermore, in order to enforce consistent equilibrium statistics with the truth along each direction, that is, $\bar{u}_{M,\infty} = \bar{u}_\infty$, $r^M_{k,\infty} = r_{k,\infty}$, one necessary condition requires that

$$2\left[\Gamma_k \bar{u}_{M,\infty} + d\right] r^M_{k,\infty} = -2d_{M,k}\left(R_\infty\right) r^M_{k,\infty} + \sigma^2_{M,k}\left(R_\infty\right),$$

for each mode $k = 0, \cdots, J/2$ from the steady state variance equations. Substituting the corresponding forms of GC1 and GC2 in (41) and (42) into the above equation, the model parameters satisfying equilibrium fidelity with the truth must follow the relations

$$-2d_{M,k} r^M_{k,\infty} + \sigma^2_M = 2\left[\Gamma_k \bar{u}_{M,\infty} + d\right] r^M_{k,\infty},$$

for GC1, and

$$-\varepsilon_{1,k} J \left(\mathrm{tr} R_\infty\right)^{-1} r^M_{k,\infty} + \varepsilon_M = 2\left[\Gamma_k \bar{u}_{M,\infty} + d\right] r^M_{k,\infty},$$

for GC2. By solving the above equations we find the consistent parameters $d_{M,k}$ and $\varepsilon_{1,k}$ for each mode in GC1 and GC2 respectively as shown in (41) and (42).