Uncertainty quantification of turbulent systems via physically consistent and data-informed reduced-order models

A. Charalampopoulos (Α. Χαραλαμπόπουλος) and T. Sapsis (Θ. Σάψης)
Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
(*Electronic mail: alexchar@mit.edu.)
(Dated: 15 June 2022)

This work presents a data-driven, energy-conserving closure method for the coarse-scale evolution of the mean and covariance of turbulent systems. Spatio-temporally non-local neural networks are employed for calculating the impact of non-Gaussian effects to the low-order statistics of dynamical systems with an energy-preserving quadratic nonlinearity. This property, which characterizes the advection term of turbulent flows, is encoded via an appropriate physical constraint in the training process of the data-informed closure. This condition is essential for the stability and accuracy of the simulations as it appropriately captures the energy transfers between unstable and stable modes of the system. The numerical scheme is implemented for a variety of turbulent systems, with prominent forward and inverse energy cascades. These problems include prototypical models such as an unstable triad-system and the Lorentz-96 system, as well as more complex models: the 2-layer quasi-geostrophic flows and incompressible, anisotropic jets where passive inertial tracers are being advected on. Training data are obtained through high-fidelity direct numerical simulations. In all cases, the hybrid scheme displays its ability to accurately capture the energy spectrum and high-order statistics of the systems under discussion. The generalizability properties of the trained closure models in all the test cases are explored, using out of sample realizations of the systems. The presented method is compared with existing first-order closure schemes, where only the mean equation is evolved. This comparison showcases that correctly evolving the covariance of the system outperforms first-order schemes in accuracy, at the expense of increased computational cost.

I. INTRODUCTION

The defining property of turbulent dynamical systems is the existence of multiple and persistent sources of instability. While from a physical perspective such systems are usually viewed as deterministic in nature, uncertainty manifests into their study via an abundance of mechanisms. Fundamental model assumptions like model structure, constitutive laws, geometry, as well as initial and boundary conditions, may be only approximations of the truth, and as a result carry uncertainty into the implementation. Furthermore, spatio-temporal discretization errors push the model predictions away from the underlying mathematical model, allowing only for bounds between the two solutions. Additionally, other input parameters of the model may exhibit randomness. This effect can be either an intrinsic property of that quantity (i.e. aleatoric uncertainty) or stem from an inability to accurately measure its otherwise deterministic value, due to computational and experimental limitations. Hence, all these effects need to be accounted for in the study of unstable complex systems.

Despite the many sources of randomness, uncertainty quantification (UQ) attempts can be separated into the following categories: (i) propagation of stochasticity from input parameters, via the model, to the output values, (ii) parametrization of input-parameter uncertainty via backwards-propagation of the model output uncertainty. Both kinds of UQ studies have been applied to all fields where nonlinear multiscale systems are observed. UQ is used for performance prediction of integrated circuits with thousands of sub-micrometer parts. It is accounted for in assessing the structural integrity of large structures with variability in their material properties2,3 and in imperfect neuroscopic models in material science4.5. It is used in combustion to describe complex kinetic mechanisms6. Furthermore, in the last 50 years, UQ tools have been widely adopted by researchers interested in turbulent flows. They have been used to model permeability of porous media in multiphase flows7,8. UQ is used to account for randomness in microfluidic applications9 and thermal problems10,11. It is used in engineering applications with shear turbulence12. In addition, it has become a major research interest in climate studies13–15 and climate change in particular16.

The most straightforward UQ approach is the well-known Monte Carlo (MC) method17–19. Techniques inspired by the MC method, perform numerous deterministic simulations of the system for sampled conditions, and proceed to perform a-posteriori statistical analysis on the numerical results20. Yet, repeated simulations of turbulent models are still prohibitively expensive both to generate and store21. Hence, MC techniques find limited applications usually only on low-dimensional systems. As a result, the need for efficient uncertainty propagation schemes arises for more complex and computationally expensive systems.

To reduce the computational cost of UQ, many approaches project the initial system on a low-dimensional subspace of pre-selected modes22. The first such approaches derived reduced-order models based on an energy-based proper orthogonal decomposition (POD)23–26. A similar concept is used in deriving reduced-order models via balanced POD based on linear theory27,28. Orthogonal decompositions are also used to derive dynamically orthogonal field equations29,30. Finite-series representations of randomness are also used in truncated polynomial chaos (PC) expansions31–33. While all these models have found success in weakly chaotic regimes, they suffer from the fact that in turbulent systems, non-energetic modes can intermittently have signif-
icant impact on energetic modes\textsuperscript{34–37}. As a result, despite the low computational cost of truncated-series methods, such approaches are antithetical to the nature of turbulence. Non-modal UQ approaches, include the utilization of the fluctuation-dissipation theorem (FDT)\textsuperscript{38–44} and the modified quasilinear Gaussian closure (MQG) scheme\textsuperscript{45}, which utilizes a second-order moment framework and models the impact of non-Gaussian statistics via incorporating statistical steady-state information appropriately.

It is therefore clear that for reduced-order models to accurately predict that statistical properties of the reference turbulent systems, a physically consistent modeling of the intermittent energy transfers between energetic modes needs to be employed. Hence, tools that allow for the approximation of complicated operators are of interest. Apart from recent conventional basis expansions\textsuperscript{46–48} and parametrizations\textsuperscript{22,49}, deep neural networks have recently seen success in reduced-order modelling of turbulent systems both in a deterministic and in a statistical framework\textsuperscript{50–58}. Despite their lack of closed-form expressions for the closure terms, neural networks have reliably approximated many intricate operators in nonlinear dynamics. In addition, neural networks can seamlessly incorporate spatio-temporal non-locality in their predictions, a property that suits many reduced-order models. Utilizing temporal delays in a closure scheme is theoretically justified via Takens’ embedding theorem\textsuperscript{59}, which states that under the constraint of observing a limited number of the state variables of a system, in principle, we can still obtain the attractor of the full system by using delay embedding of the observed state variables (i.e. a non-Markovian approach). Such methods have seen success in reduced-order modelling of highly non-linear dynamical systems\textsuperscript{57,58,60–63}.

This work aims to expand on the MQG approach for uncertainty quantification, presented in\textsuperscript{45,64–66}. The approach studies turbulent systems with a quadratic and energy-preserving nonlinearity. Similar to the approach of traditional MQG, a second-order statistical framework is employed for this study. This framework allows for a computationally cheap reduced-order model. Uncertainty is modelled under a parametric Bayesian point of view. In previous MQG approaches, the nonlinear energy transfers, that result in non-Gaussian effects, are modelled via a quasi-linear approach, where the closure is tuned to specific steady-state statistics. The current method replaces this assumption with spatio-temporally nonlocal neural networks. The scheme is constructed so that it a-priori respects the energy transfers dictated by the nonlinear term. This is achieved by using steady-state statistics of flow realizations that are incorporated during the training process. These physical constraints allow for correct prediction of mode-to-mode energy transfers during simulations. The utilization of deep neural networks allow for a richer representation of nonlinear energy transfers and higher accuracy, not only for the mean but also for the second-order statistics of the flow. Past values of these features are also employed as inputs for the turbulent closures in a causal manner. We finally employ imitation learning\textsuperscript{67} to improve the stability properties of the computed closures. A related approach, aiming to machine learn second-order closures, has been presented recently\textsuperscript{68}; this effort does not encode the underlying physical constraints in the training process and it has not been assessed in turbulent fluid flows.

The outline of the paper is as follows. Section II formulates the general non-linear problem of interest. The appropriate physical constraint, arising from the form of the system is derived. Furthermore, the data-driven parametrization, as well as the objective function used during training are discussed. Next, the numerical investigation of the formulated scheme is included in section III. The closure scheme is assessed in a multitude of turbulent models. In all cases, the results are presented for out of sample realizations of the models (i.e. not including in the training process) to showcase the generalizability of the closure scheme. The role of the physical constraint on the stability properties and accuracy of the coarse-scale equations are examined. The accuracy of the closure scheme is measured via its ability to capture the correct energy transfers between modes and thus reach the appropriate statistical equilibrium. This goal is quantified differently in each test case, but numerical comparisons include energy of the mean, total variance of the system, energy spectra and heatflux spectra whenever they are of interest. Finally, section IV includes the major conclusions drawn from the numerical investigation of the previous section.

Expanding on the numerical investigation, the first application is the triad system, a toy-problem used to study energy transfers between a large-scale variable and two fluctuating modes. This system is of particular interest due to its intrinsically unstable mean, rendering it challenging to approximate despite having only 3 degrees of freedom. Next, the method is applied to the Lorenz-96 system, a prototypical model for studying turbulence due to baroclinic instability in midlatitude atmospheric flows. Results are presented both for constant and time-varying excitation. In both cases the data-informed scheme compares favourably with the results derived by Monte Carlo simulations. The scheme is then applied to a 2-layer, quasi-geostrophic (QG) model. First, results are presented for the statistics of high-latitude oceanic turbulence. The scheme once again is able to capture the correct energy and heatflux spectra of the full physical model. Afterwards, results on the statistics of mid-latitude oceanic and high-latitude atmospheric flows are also presented. The final test-case involves an inhomogeneous system, where the closure scheme is applied on unimodal and bimodal turbulent jets over which inertial passive tracers are being transported.

II. PROBLEM FORMULATION

The proposed methodology combines a second-order statistical formulation with neural networks that are trained under appropriate physical constraints. This framework produces accurate uncertainty quantification predictions for nonlinear dynamical problems. Let $\\mathbf{u}$ be a field describing the state of the system. The evolution equation of $\\mathbf{u}$ has the following general form

$$\frac{d\\mathbf{u}}{dt} = A\\mathbf{u} + B(\\mathbf{u},t) + F(t) + \dot{W}(t;\\omega)\sigma_k(t),$$  \hspace{1cm} (1)
where $A$ is a linear operator, $F$ denotes a deterministic external forcing and $W_t \sigma_k$ corresponds to a stochastic forcing with white noise characteristics. The operator $B$ is assumed to be quadratic and energy-preserving, i.e.

$$B(u, u) \cdot u = 0.$$  \hfill (2)

This restrictive definition of $B$ is valid for many important problems in fluid mechanics, retaining the physical relevance of the formulation. For example, $B$ can be viewed as the advection term of turbulent flows, a class of problems that has historically attracted the attention of reduced-order modelling literature.

Using well-known linear algebra results\(^6\), the linear operator $A$ can be decomposed as

$$A = \frac{1}{2}(A - A^T) + \frac{1}{2}(A + A^T) = L + D,$$ \hfill (3)

where $L$ is a skew-symmetric linear operator and $D$ is a symmetric linear operator. Throughout this work $D$ will also be assumed to be negative definite, which implies that it corresponds to a linear dissipative process. The quantity of interest $u$ is analyzed using the finite-dimensional expansion

$$u = \bar{u} + u' = \bar{u} + \sum_{i=1}^{N} Z_i(t; \omega) v_i,$$ \hfill (4)

where $v_i$ form a prescribed orthonormal basis, while $Z_i$ are zero-mean, time-dependent random functions. The symbol $\omega$ denotes the random argument and the mean field $\bar{u}$ can be interpreted as an ensemble average, $\mathbb{E}[u] = \bar{u}$, with respect to $\omega$. Using the above representation, the original dynamical equation can be re-written as

$$\frac{d \bar{u}}{dt} + \frac{d u'}{dt} = (L + D) \bar{u} + (L + D) u' + B(\bar{u}, u) + B(u', \bar{u}) + B(u', u') + F + \bar{W}_k(t; \omega) \sigma_k(t).$$ \hfill (5)

By taking the expectation of the above equation, we derive the dynamical equation for the average state $\bar{u}$, i.e.

$$\frac{d \bar{u}}{dt} = (L + D) \bar{u} + B(\bar{u}, \bar{u}) + \sum_{i=1}^{N} \sum_{j=1}^{N} Z_i Z_j B(v_i, v_j) + F.$$ \hfill (6)

For a second-order statistical framework, an evolution equation for the perturbations $u' = u - \bar{u}$ is derived:

$$\frac{d u'}{dt} = (L + D) u' + B(\bar{u}, u') + B(u', \bar{u}) + B(u', u') + F + \sum_{i=1}^{N} \sum_{j=1}^{N} [Z_i Z_j - R_{ij}] B(v_i, v_j) + W_k \sigma_k.$$ \hfill (7)

The projection of the evolution equation for $u'$ onto a basis function $v_n$, yields

$$\frac{d Z_n}{d t} = + \sum_{i=1}^{N} Z_i [A v_i + B(\bar{u}, v_i) + B(v_i, \bar{u})] \cdot v_n + \sum_{i=1}^{N} \sum_{j=1}^{N} [Z_i Z_j - R_{ij}] B(v_i, v_j) \cdot v_n + W_k \sigma_k \cdot v_n.$$ \hfill (8)

By multiplying the above equation with $Z_m$ and taking the ensemble average, we have

$$\frac{d Z_m}{d t} = + \sum_{i=1}^{N} R_{im} [A v_i + B(\bar{u}, v_i) + B(v_i, \bar{u})] \cdot v_n + \sum_{i=1}^{N} \sum_{j=1}^{N} Z_i Z_j B(v_i, v_j) \cdot v_n + Z_m W_k \sigma_k \cdot v_n,$$ \hfill (9)

since

$$R_{ij} Z_m = R_{ij} Z_m = 0.$$ \hfill (10)

Hence, the evolution of the elements of the covariance matrix is dictated by the equation

$$\frac{d R_{mn}}{d t} = \sum_{i=1}^{N} R_{im} [A v_i + B(\bar{u}, v_i) + B(v_i, \bar{u})] \cdot v_n + \sum_{i=1}^{N} \sum_{j=1}^{N} Z_i Z_j B(v_i, v_j) \cdot v_n + \sum_{i=1}^{N} \sum_{j=1}^{N} Z_i Z_j Z_m B(v_i, v_j) \cdot v_m + (\sigma_k \cdot v_m) (\sigma_k \cdot v_n).$$ \hfill (11)

The evolution equation for the covariance of the system can then be written as

$$\frac{d R}{d t} = \mathcal{L} R + R \mathcal{L}^* + \mathcal{Q} + \mathcal{Z},$$ \hfill (12)

where $\mathcal{L}$ is a linear operator expressing dissipation and energy transfers between the mean field and the stochastic modes

$$\mathcal{L}_{ij} = [(L + D) v_j + B(\bar{u}, v_j) + B(v_j, \bar{u})] \cdot v_i,$$ \hfill (13)

with $\mathcal{L}^*$ being its transpose. The operator $(\mathcal{Q})_{ij} = (\sigma_k \cdot v_j) (\sigma_k \cdot v_i)$ models effects due to the stochastic external forcing and $\mathcal{Z}$ corresponds to third-order effects that express energy fluxes between different stochastic modes

$$\mathcal{Q}_{mn} = \sum_{i=1}^{N} \sum_{j=1}^{N} Z_i Z_j Z_m B(v_i, v_j) \cdot v_n + \sum_{i=1}^{N} \sum_{j=1}^{N} Z_i Z_j Z_m B(v_i, v_j) \cdot v_m.$$ \hfill (14)

In more detail, $\mathcal{L}$ includes the effects of the linear operators on each mode, as well as energy transfers between the mean and each mode via the energy-preserving nonlinear operator. All these effects can be studied and understood under the framework of second-order statistics. It is the mode-to-mode nonlinear energy transfers that require knowledge of higher moments for their estimation. In order to close the system for the covariance and the mean, a model for the third-order terms $\mathcal{Z}$ appearing in Eq. (11) is required. To this end, neural networks are utilized to parameterize them. The architecture and constraints used are presented in the following subsections.
A. Physical constraints related to nonlinear energy transfers

Before describing the details of the data-driven model for \( \mathcal{Q} \), we note that an important feature of the presented closure scheme is the requirement to satisfy certain physical principles, which characterize turbulent systems. More specifically, modes that carry small energy or variance can still have important, dynamical effects on the large variance modes through nonlinear energy transfers. A schematic of this property can be seen in Fig. 1. The external force excites the mean, which then transfers part of this energy to the unstable modes. Since, these modes are unstable, the linear operator \( \mathcal{L} \) can only increase them in amplitude. Hence, the only way for the system to reach a statistical steady state is due to i) dissipation, and ii) the nonlinear terms \( \mathcal{Q} \) transfer energy from the unstable modes to the stable ones.

\[
\text{Re}[\mathcal{L}_i(u)] > 0 \quad \text{and} \quad \text{Re}[\mathcal{L}_j(u)] < 0
\]

where \( \text{Re} \) denotes the real part of a complex number and \( \mathcal{L}_i \) and \( \mathcal{L}_j \) are linear operators.

\[
B(u', u') \cdot u' = 0
\]

or in terms of the closure term\(^{45}\)

\[
\text{Tr}[\mathcal{Q}] = 0.
\]

This constraint is enforced during training to improve the numerical stability of the predictions. It is emphasized that correctly capturing this constraint is necessary for the purpose of correctly modeling the energy exchanges between stable and unstable modes and thus capturing the correct statistical steady state of the system under discussion.

B. Data-driven parametrization of the closure terms

While the dynamical systems under study are Markovian and spatially-local, i.e. the evolution of \( u \) in a specific location and time instant depends only on the current time instant and the current neighborhood, this is not the case for the reduced-order averaged version of these equations. In particular, the evolution of the mean and covariance equations one typically does not have access to the full-state information required to fully describe the evolution of the system (in this case third-order moments).

As has been showcased by recently proposed data-driven schemes for dynamical systems\(^{56,1} \), Takens embedding theorem can be used to enhance the accuracy of predictions\(^{59} \). The theorem states that even if only a limited number of the state variables of a system are observed, in principle, one can still obtain the attractor of the full system by using delay embeddings of the observed state variables. To this end, the closure terms are parametrized with non-local in time (but still causal) representations, based on Temporal Convolutional Networks\(^{72} \) (TCN) and Long Short-Term Memory networks\(^{73} \) (LSTM). These ML architectures have been employed successfully in previous work\(^{37} \) on computing turbulent closures just for the mean equation (6).

In terms of spatial information, the entire mean field and the covariance are utilized as input for the data-informed closure scheme. As a result, the closure terms are modeled in the following form:

\[
\mathbb{D}_{n,m} \left[ \Theta; \xi(t) \right] = \sum_{i=1}^{N} \sum_{j=1}^{N} \mathbb{Z}_{i} \mathbb{Z}_{j} \mathbb{B}(v_{i}, v_{j}) \cdot v_{n} + \sum_{i=1}^{N} \sum_{j=1}^{N} \mathbb{Z}_{i} \mathbb{Z}_{j} \mathbb{B}(v_{i}, v_{j}) \cdot v_{m},
\]

where \( \xi = \{u, R\} \) are 2nd order statistics of the system, and \( \mathcal{X}(t) \) denotes the history of the statistics state backwards from time \( t \), i.e. \( \mathcal{X}(t) = \{t; t - \tau_1, ..., t - \tau_2, ..., t - \tau_N\} \). The entire mean field \( \bar{u} \) and covariance \( \mathbb{R} \) are used as input for the neural network. The vector \( \Theta \) denotes the hyperparameters of the neural network while the temporal neighborhood, \( \mathcal{X}(t) \), is selected such that if further increased, the training error does not significantly reduce any more. The number of points in time that have to be considered depends on the temporal discretization of the domain.

C. Objective function for training

In terms of the training process itself, the input and output data are normalized as typically suggested\(^{44} \). The loss function for this problem is chosen to be the single-step prediction mean square error superimposed with the physical constraint. This can be formulated as

\[
L(\Theta) = \frac{1}{T} \int_{T} \sum_{n} \sum_{m} \left| \mathbb{D}_{n,m} \left[ \Theta; \xi \right] - \mathcal{Q}_{n,m} \right|^{2} dt + \lambda \int_{T} \text{Tr}[\mathcal{Q}] dt,
\]

where \( \lambda \) is a weight that measures the relative importance between the data and the physical constraint, typically chosen to be 1.
III. APPLICATIONS

A. Triad system

The first application involves a three-dimensional dynamical system that consists of 3 Langevin equations coupled via quadratic and energy-preserving nonlinearities. This triad system acts as a simple surrogate model for barotropic instability. It can be viewed as the result of projection of the fluid equations to three modes, one corresponding to the mean flow and the other two corresponding to wave perturbations. Under these assumptions, the system has the form

\[
\begin{bmatrix}
\frac{d\xi_1}{dt} \\
\frac{d\xi_2}{dt} \\
\frac{d\xi_3}{dt}
\end{bmatrix} = \left[\begin{array}{ccc}
-\gamma_1 & 0 & 0 \\
0 & -\gamma_2 & 0 \\
0 & 0 & -\gamma_3
\end{array}\right]
\begin{bmatrix}
\xi_1 \\
\xi_2 \\
\xi_3
\end{bmatrix}
+ \left[\begin{array}{ccc}
\lambda_{11} & \lambda_{12} & \lambda_{13} \\
-\lambda_{12} & 0 & \lambda_{13} \\
-\lambda_{13} & -\lambda_{23} & 0
\end{array}\right]
\begin{bmatrix}
\xi_1 \\
\xi_2 \\
\xi_3
\end{bmatrix}
+ \left[\begin{array}{ccc}
\beta_1 u_2 u_3 \\
\beta_2 u_1 u_3 \\
\beta_3 u_1 u_2
\end{array}\right]
+ \left[\begin{array}{c}
\sigma_1 dW_1 \\
\sigma_2 dW_2 \\
\sigma_3 dW_3
\end{array}\right],
\]

where \( D \) and \( L \) are respectively the symmetric and skew-symmetric component of the linear operator, \( F_1, F_2, F_3 \) correspond to deterministic external excitation and \( dW_1, dW_2, dW_3 \) are independent white-noise processes. It is also assumed that

\[
\beta_1 + \beta_2 + \beta_3 = 0,
\]

to ensure that the quadratic term is energy-preserving. Interaction of this reduced-order model with other modes of the full system is modeled via the white-noise terms and linear dissipation, a standard practice in many stochastic climate models.

The goal of this section is to examine the triad system as a prototypical model for turbulent anisotropic flows. A property of such flows is that the mean exhibits persistent instabilities along certain directions in phase space. Under the assumption that \( u_1 \) corresponds to the mean-flow variable, this instability can be added to the triad system via a negative \( \gamma \) value. Linear dissipation is set as \( \gamma_1 = -0.4 \) and \( \gamma_2 = \gamma_3 = 2 \), following a previously presented setup. The skew-symmetric part is set as \( \lambda_{12} = 0.03, \lambda_{13} = 0.06 \) and \( \lambda_{23} = -0.09 \). The coefficients of the nonlinear term are set as \( \beta_1 = 2 \) and \( \beta_2 = \beta_3 = -1 \). For the first test-case constant deterministic external forcing is tested. To ensure that the perturbation variables \( u_2, u_3 \) are energetic (and thus remove energy from the mean), the deterministic forcing is set to \( F_1 = 0, F_2 = -1 \) and \( F_3 = 1 \). Finally, the white noise amplitudes are tuned so as to ensure the system achieves a statistical steady state. For this to happen the amplitudes are chosen as \( \sigma_1 = 0.25 \) and \( \sigma_2 = \sigma_3 = 0.79 \). Initial conditions are sampled from the distribution

\[
\begin{bmatrix}
\xi_1(0) \\
\xi_2(0) \\
\xi_3(0)
\end{bmatrix} \sim \left[\begin{array}{c}
\mathcal{N}(1,0.5) \\
\mathcal{N}(0,5,0.2) \\
\mathcal{N}(-0.5,0.1)
\end{array}\right].
\]

For reference data, \( 10^5 \) samples of this system are computed (each time using a new sample of both the white noise forcing and the initial conditions). Reference statistics of this dynamical systems are obtained from this MC simulation. All simulations are carried out with time-step \( dt = 0.01 \) from \( t = 0 \) till \( t = 25 \). The system is integrated in time using the forward Euler method. With respect to the reduced-order model, a single-layer LSTM neural network is used during training.

The statistical properties of the system as well as a comparison with the low-order statistical predictions of the proposed data-informed method are depicted in figure 2. Subplots (a) and (d) establish that \( u_1 \) is indeed the most energetic mode of the system and that the system achieves a statistical steady state. Furthermore, as seen from the marginal on the \( u_2 - u_3 \) plane on subplot (f), the system exhibits strong anisotropic behaviour in this plane. Yet, the marginals on the \( u_1 - u_2 \) and \( u_1 - u_3 \) planes are Gaussian. The reduced-order model is able to accurately predict the statistical equilibrium of the system. In addition, it adequately captures the intermittent dynamical evolution of the mean and variance of the system. The good agreement with the MC data is observed despite the fact that the pdf of the system is strongly non-Gaussian in the \( u_3 - u_2 \) space and the mean variable \( u_1 \) is by construction linearly unstable.

The next test-case involves periodic forcing while the other parameters of the system remain the same as before. In more detail, the deterministic part of the forcing is now set to

\[
F_0 = 0, \quad F_2 = -1 + 0.5 \sin \frac{\pi t}{2}, \quad F_3 = 1 + 0.2 \cos \frac{\pi t}{2}.
\]

In addition, the parameters of the stochastic forcing are set to

\[
\sigma_1 = 0.25 - \frac{1.3^2}{10} \sin^2 \frac{\pi t}{2}, \quad \sigma_2 = 0.79 - \frac{2}{5} 0.79 \cdot 1.3^2 \sin^2 \frac{\pi t}{2}.
\]

MC results for the system as well as a comparison with the results of the data-informed closure scheme (trained with the previous set of forcing parameters) are shown in figure 3(a, d). Again, despite the strongly non-Gaussian nature of the problem (fig. 3 (b, e, f)) the closure scheme is able to accurately predict the first- and second-order statistics throughout the duration of the simulation.

B. L96 model

As a next step, a higher-dimensional problem, with a large number of positive Lyapunov exponents, is considered. Specifically, the model under discussion is governed by the equation

\[
\frac{du_i}{dt} = u_{i-1}(u_{i+1} - u_{i-2}) - u_i + F,
\]

for \( N = 40 \) and \( F \) deterministic external forcing. This system acts as a prototypical model to study baroclinic instability in midlatitude flows. The dynamics comprise of an energy-conserving quadratic nonlinearity, a linear dissipation term...
FIG. 2: Comparison between MC (solid line) and reduced-order (dashed-line) results for triad system with constant forcing. (a) Evolution of mean of the variables of the system; (b) Evolution of off-diagonal components of covariance matrix; (c) Evolution of real part of eigenvalues of the linear operator $L_v$; (d) Evolution of total system variance and variance of each mode; (e) Contour of the steady-state system pdf for $u|_{f(u)=10^{-5}}$; (f) Marginal pdfs of the system at steady-state.

FIG. 3: Comparison between MC (solid line) and reduced-order (dashed-line) results for triad system with periodic forcing. (a) Evolution of mean of the variables of the system; (b) Evolution of off-diagonal components of covariance matrix; (c) Evolution of real part of eigenvalues of the linear operator $L_v$; (d) Evolution of total system variance and variance of each mode; (e) Contour of pdf at $t = 25$ for $u|_{f(u)=10^{-5}}$; (f) Marginal pdfs of the system at $t = 25$. 
and external forcing. The forcing term is assumed to be spatially homogeneous, which as a result implies that the system is translationally invariant in space. This property implies that Fourier modes can be used for the basis expansion and that the covariance matrix can be assumed to be diagonal. One can observe that for different magnitudes of forcing, the number of unstable modes changes (from 0 up to 11). Hence, this is model is an excellent case to assess the presented method.

To illustrate numerically the UQ properties of the proposed closure scheme, the model is trained and tested on an aperiodic forcing generated by the Ornstein-Uhlenbeck process

$$dF = -\frac{1}{\tau_F} (F - \bar{F}) dt + \sigma_F dW,$$  

where $\bar{F}$ is the mean value around which the process oscillates and $dW$ models white noise. Here we chose as modes $v_n$ the Fourier modes. Also, we do not perform any covariance reduction, i.e., we model the full covariance matrix.

Training occurs with a single stochastic forcing realization. The reference statistics are derived by averaging over a distribution of initial conditions. To derive the required training data $10^4$ direct numerical simulations are averaged. These statistics allow for the calculation of the time-evolution of $\mathbb{E}[u]$ and $\mathbb{E}[Z_k^2] = \mathbb{E}[(u - \mathbb{E}[u]) \cdot v_n]^2$. The neural network is then trained using the derived data and under the constraint of Eq. (16), which is equivalent to the constrain that the quadratic operator should be energy conserving. A single-layer LSTM neural network with 150-timesteps as time-history is used. Testing of the closure scheme is carried out for 4 out-of-sample random realizations of the forcing. Only 10 time-steps are used as initial input for the neural network. Results are presented in Figure 4. For all realizations, we have excellent agreement between the spectrum predicted from the data-driven closure and the MC simulation. This is the case despite the strongly transient nature of the excitation that pushes the system away from its statistical steady state.

C. Multiphase flow

The next test case is an anisotropic multiphase flow setup. It involves the advection of bubbles, which are assumed to be passive inertial tracers, over an incompressible fluid flow. The fluid flow is governed by the Navier-Stokes equations in dimensionless form:

$$\frac{Du}{Dt} = -\nabla p + \frac{1}{Re} \Delta u + \nu \Delta^{-2} u + F,$$

$$\nabla \cdot u = 0,$$

where $u$ is the velocity field of the fluid, $p$ its pressure, $Re$ is the Reynolds number of the flow, $D/Dt$ is the material derivative operator and $F$ denotes an external forcing term. Parameter $\nu$ is a hyperviscosity coefficient aiming to remove energy from large scales and maintain the flow in a turbulent regime.

One can follow a similar process for the advection equation governing the motion of small inertia particles their Lagrangian velocity, $v$, is a small perturbation of the underlying fluid velocity field $\bar{u}$:

$$v = u + \epsilon \left(\frac{3R}{2} - 1\right) \frac{Du}{Dt} + O(\epsilon^2),$$

where

$$\epsilon = \frac{St}{R} \ll 1, \quad R = \frac{2\rho_f}{\rho_f + 2\rho_p}.$$

The parameter $\epsilon$ represents the importance of inertial effects, while $St$ is the bubble Stokes number, measuring the ratio between the characteristic timescales of the bubbles over that of the flow. Parameter $R$ is a density ratio with $\rho_f$ and $\rho_p$ being the density of bubble and the fluid respectively. By introducing $\rho$ as the concentration of tracers at a particular point, the following transport equation can be derived for the bubble flow:

$$\partial_t \rho + \nabla \cdot (\rho \bar{u}) = v_2 \Delta^4 \rho.$$

The right-hand-side of the transport equation represents a hyperviscosity term. The hyperviscosity parameter is tuned so as to remove energy from scales close to the resolution limit of the numerical simulations.

Utilizing the series expansion presented here for the variables of the problem:

$$u = \bar{u} + \sum_k Z_{k,u} v_{k,u}, \quad \rho = \bar{\rho} + \sum_k Z_{k,\rho} v_{k,\rho},$$

one can develop evolution equations for the mean and the variance for both the fluid flow and the bubble flow. For the particular problem, the data-informed closure scheme is trained on data from unimodal jet flows (details of the jet configuration have been presented previously $^{27}$). The Reynolds numbers used in training are $Re \in \{650, 750, 850\}$. The reference simulations are carried out on a $256 \times 256$ grid with doubly periodic boundary conditions on a $[0, 2\pi]^2$ rectangular domain. All flows are evolved until they reach a statistical equilibrium. Since MC simulations are prohibitively expensive for this problem, time-averaging is used for the derivation of the energy spectra in the statistical equilibrium.

The derived closure is first tested on unimodal jets in the range $Re \in \{500, 1000\}$. For the mean-field model (eq. (6)) a coarse resolution of $32 \times 32$ is employed. For the covariance of both the flow field and the density, Fourier modes are utilized. Specifically, modes with wavenumber $|k| \leq 48$ are considered in the covariance evolution (eq. (11)) that is complemented with the ML closures.

Figure 5 presents the space-time-averaged mean-square error between the $x - y$ locally averaged DNS flow field, $\bar{u}^*$, and the coarse scale model, $\bar{u}$:

$$|\bar{u}^* - \bar{u}|_2^2 = \int_0^{2\pi} \int_0^{2\pi} (\bar{u}^* - \bar{u})^2 dx \, dt.$$
FIG. 4: Comparison of ML uncertainty quantification scheme with exact statistics produced by the Monte-Carlo method. Results are shown for different dynamical regimes of the aperiodic forcing parameter $F$ generated as an Ornstein-Uhlenbeck process. The colorplots present the evolution of the exact and approximated spectrum. We also present the energy of the mean and the trace of the covariance over time. At the last row we show the steady state spectrum (exact and approximate).
proved stability, albeit the additional computational cost. The numerical results of the presented scheme are compared with first-order closures where the ML correction is applied only on the mean equation (6) and the covariance is not modeled. For the 1st-order closure both TCN and LSTM architectures were assessed. The constrained versions, cTCN and cLSTM, correspond to closures where, in addition to the $L^2$-error of the closure terms, the physical constraint related to the energy-preserving property of the advection term is included (eq. (18)). As expected the second-order closure scheme outperforms the corresponding first-order schemes, even using a coarser resolution. The inclusion of the constraint significantly improves the accuracy of the predicted statistics for all cases. Furthermore, to showcase the importance of using the constraint derived in eq. (16), a version of the 2nd-order closure scheme with $\lambda = 0$ is also presented. As it can be seen, numerical results are improved when including the constraint.

As an additional test-case, the closure scheme is tested on bimodal jets with $\text{Re} \in \{500, 1000\}$ (details of the bimodal jet configuration have been presented previously). The closure is trained on the same data set as before, i.e. on unimodal jets with Reynolds number $\text{Re} \in \{650, 750, 850\}$. Figure 6 presents the space-time-averaged mean-square error between the $x-y$ locally averaged DNS flow field. In this case, including the constraint during training not only improves numerical accuracy of the results but allows us to avoid numerical instabilities for Reynolds numbers outside the training data-set.

In Figure 7 we compare the energy spectra for the fluid and bubble flows as obtained by DNS and the second order closure scheme. We also include a comparison with the first-order closure obtained in previous work. For both the unimodal and bimodal jets we have accurate computation of the energy spectra of the fluid flow and the advected bubble dispersion in the statistical equilibrium. Furthermore, the second order model clearly outperforms the corresponding first-order scheme, indicating that incorporating additional physics through the second order equation is leading to better performance and improved stability, albeit the additional computational cost.

### D. Application to quasi-geostrophic (QG) flows

The last test-case involves a 2-layer quasi-geostrophic (QG) model. The model considered here consists of an advection-diffusion equation for the potential vorticity $q_i$, in each of two immiscible layers with fractional layer thickness $\delta = H_1/H_0$, $1 - \delta = H_2/H_0$, respectively (where $H_0 = H_1 + H_2$) and mean zonal velocities $U_1 > U_2$. The domain under discussion is a square doubly periodic domain with rigid-lid surface boundary conditions. The governing equations for the potential vorticity (PV) of each layer become

$$\partial_t q_1 = -J(q_1, q_1) - \beta \partial_y q_1 - (U_1 - U_2) L_p^2 (1 - \delta) \partial_x q_1 + F_1,$$

and

$$\partial_t q_2 = -J(q_2, q_2) - \beta \partial_x q_2 + (U_1 - U_2) L_p^2 2 \delta \partial_y q_2 + (U_1 - U_2) \partial_y q_2 + F_2 - \nu \nabla^2 q_2,$$

where the field $q_1$ corresponds to the upper-layer PV and $q_2$ to the bottom-layer PV, with $\psi_i$ being the respective stream-functions and $J(a, b) = \partial_i a \partial_j b - \partial_i a \partial_j b$ is the Jacobian operator. The $\beta$ terms arise from the variation of the vertical projection of Coriolis frequency with latitude and the $k^2$ terms result from the imposed shear. The inverse relations are

$$q_1 = \nabla^2 \psi_1 - \frac{f_0^2}{g \Omega_1} (\psi_1 - \psi_2),$$

$$q_2 = \nabla^2 \psi_2 + \frac{f_0^2}{g \Omega_2} (\psi_1 - \psi_2) + f_0 \frac{h_2}{H_2}.$$

The dynamics can also be described in terms of the barotropic and baroclinic modes and their corresponding
FIG. 7: Energy spectra for the fluid flow and bubble flow for (a) a unimodal jet with Re = 800 and (b) a bimodal jet with Re = 800.

The model dynamics can then be rewritten in terms of the barotropic and baroclinic modes. For periodic boundary conditions with a flat bottom, this yields

\[
\begin{align*}
    \partial_t q_l &= -J(q_1, q_l) - J(q_c, q_l) - (1 - \delta) U \nabla^2 q_l - (1 - \delta) \nabla \nabla^2 q_l - U \partial_x \nabla^2 q_l - \beta \partial_x \psi_l - \nabla^4 q_l, \\
    \partial_t q_c &= -J(q_1, q_c) - J(q_c, q_c) - \xi J(q_c, q_c) + \sqrt{\delta(1 - \delta)} \nabla^2(q_l - a^{-1} \psi_c) - \beta \partial_x \psi_c \\
    &- U \partial_x (\nabla^2 \psi_c + \lambda^2 \psi_l + \xi \nabla^2 \psi_c) - \nabla^4 q_c,
\end{align*}
\]

(38)

where

\[
\lambda^2 = k_d^2, \quad \xi = \frac{1 - 2 \delta}{\sqrt{\delta(1 - \delta)}},
\]

(39)

with \( \xi \) expressing the triple interaction coefficient and \( U = \)

\[
\begin{align*}
    q_l &= \delta q_1 + (1 - \delta) q_2 = \nabla^2 \psi_l, \\
    \psi_l &= \delta \psi_1 + (1 - \delta) \psi_2, \\
    q_c &= \sqrt{\delta(1 - \delta)}(q_1 - q_2) = (\nabla^2 - k_d^2) \psi_c, \\
    \psi_c &= \sqrt{\delta(1 - \delta)}(\psi_1 - \psi_2).
\end{align*}
\]

(37)
responds to baroclinic ocean turbulence at high latitudes. A

\[ \sqrt{\delta(1-\delta)}(U_1-U_2) \] is the shear intensity. Three different regimes can be distinguished for this model, designated by values of the model parameters: (i) ‘low latitude’ or weakly supercritical \((\beta \approx k_0^2/2, r = 1)\), (ii) ‘mid latitude’ or moderately supercritical \((\beta \approx k_0^2/4, r = 4)\) and (iii) ‘high latitude’ or strongly supercritical \((\beta \approx 0, r = 16)\). Hence, \(\beta\) decreases as latitude increases, while the bottom friction coefficient \(r\) is increased in order to keep the inverse energy cascade from reaching the resolution limit.

\[ \max_{k} \Re \lambda_i(k) \] normalized over its maximum magnitude, where \(\lambda_i(k)\) are the vertical eigenvalues for each wavenumber; Wavenumber-averaged BT/BC nonlinear energy fluxes.

The results presented here are derived for parameters \(\delta = 0.2, r = 9, \beta = 10\) and \(\lambda = 10\); a set of parameters that corresponds to baroclinic ocean turbulence at high latitudes. A typical snapshot of the \(q_t, q_c\) is shown in Fig. 8. Typical energy properties of such flows are shown in Fig. 9.

For the implementation of the presented method, the potential vorticity is expanded to

\[ q_j = \sum_{i=1}^{N} Z_i(t) v_i. \] (40)

Due to the periodicity of the domain, Fourier modes are used as a basis.

To test the performance of the closure scheme, we first train the model for \(U = 1.00\) and to assess its performance we test it for different mean velocity, chosen within the interval \(U \in [0.75, 1.25]\). For training, validation and testing the flow is solved assuming doubly periodic lateral boundary conditions and a \(256 \times 256\) discretization. For the coarse-scale simulations a discretization \(48 \times 48\) is utilized for the mean flow, while for the covariance we include all Fourier modes with wavenumbers \(|k| \leq 16\). To assess the performance of the closure scheme, the first metric used will be the energy of the system and its spectrum

\[ E_{\text{total}} = E^t + E^c = \sum_k \left[ |k|^2 |\hat{\psi}_{k,l}|^2 + (|k|^2 + \lambda^2) |\hat{\kappa}_{k,l}|^2 \right], \] (41)

where \(E^t\) and \(E^c\) are the energy carried by the barotropic and baroclinic modes respectively. Parameters \(k\) and \(l\) signify the zonal and meridional component of wavenumber, respectively. Furthermore, the heatflux and its spectrum will also be used as metrics

\[ H_{l} = \frac{\lambda}{U^2} \sum_k \frac{i |k| g_{\omega,k,l} q^2_{l,k,l}}{|k|^2 (|k|^2 + \lambda^2)}. \] (42)

In Figure 10 the total mean energy and heatflux are shown for different values of \(U \in [0.75, 1.25]\); we present the results from the coarse resolution solver with the ML closures.
and compare these with the DNS. We also show the radially-averaged energy and heat flux spectra and note the favorable comparison between the DNS results and the data-informed closure-scheme. For a more detailed comparison, the energy and heat flux spectra for the case $U = 0.95$ are also computed. Results are presented in Fig. 11, where the total normalized energy spectrum, the heat flux and nonlinear energy fluxes are compared with DNS calculations. In Fig. 12 the energy components carried by the barotropic and baroclinic modes respectively are also compared with DNS results. In all cases, the coarse-scale simulation is able to accurately capture the equilibrium statistics of the flow.

Furthermore, in Fig. 13 the marginal pdfs for the leading modes $\Psi_{(1,0)}, \Psi_{(0,1)}$ of the barotropic streamfunction $\Psi = (\psi_1 + \psi_2)/2$ are presented. Training takes place for flows with parameters $(\beta, r) = (0.8, 0.2), (1.5, 0.1), (2.5, 0.1)$. We compare the closure schemes for two cases: the mid-latitude case that corresponds to $k_d = 4, \beta = 2, r = 0.1$, and the high-latitude case that corresponds to $k_d = 4, \beta = 1, r = 0.2$.

We also include a comparison with the first-order closure scheme. The first-order closure results use the same resolution as the presented closure, with a TCN-based architecture and a constraint to the loss function that enforces the energy-preserving property of the quadratic nonlinearity of the system. We observe that the second-order closure captures very accurately the non-Gaussian structure of the pdf. This is not
FIG. 11: Comparison between DNS simulations and ML model for $U = 0.95$. Results show total energy spectrum and heat flux. The black dashed line is the $-10\%\max_k |H_k|$ contour of the heat flux field.

FIG. 12: Comparison of barotropic and baroclinic energy between spectral code and ML-closure scheme for $U = 0.95$. The case for the first-order closure which typically underestimates the variance but also misses the bimodal character of the mode $\tilde{\psi}_{(0,1)}$. Similarly, in Fig. 14, the pdf of the top and bottom layer streamfunctions are shown. Again, the proposed 2nd-order closure outperforms the 1st-order closure scheme. Discrepancies in the tails of the pdf may be due to unresolved high wavenumbers, especially since the large-scale modes seem to be well approximated by the 2nd-order scheme. Fi-
nally, it is noted that training the second-order scheme with the same hyperparameters but setting $\lambda = 0$ yields a closure that becomes unstable as the flow evolves, highlighting the importance of the constraint in numerical stability.

IV. CONCLUSIONS

We have formulated and assessed a data-informed turbulence-closure scheme that respects the underlying conservation properties of nonlinear advection. The method employs a second-order framework for the uncertainty quantification of nonlinear and turbulent dynamical systems. We first applied our approach to prototypical problems of nonlinear
dynamics, like the unstable triad system and the Lorenz-96 model. In both cases, the data-informed approach produced results in good agreement with reference MC simulations. Furthermore, the method was applied to more realistic turbulent flows, involving anisotropic multiphase flows and a 2-layer quasigeostrophic model. The obtained results demonstrated the improvement of applying the closure at the second-order level, as opposed to mean-flow closures, but also how the results are improved by encoding the energy conservation property of the nonlinear terms in the training process. In addition, we illustrated that the ML closure of the covariance equation allows for accurate modeling of highly non-trivial non-Gaussian statistics that govern the response of low wavenumbers in the QG model. Future work involves the application of these ideas to 3D turbulence as well as the detailed statistical modeling of spatio-temporally intermittent phenomena.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ACKNOWLEDGEMENTS

This paper has been supported through ONR-MURI Grant No. N00014-17-1-2676 and AFOSR-MURI Grant No. FA9550-21-1-0058 and the DARPA Grant No. HR00112290029.

14B. Salmon, Lectures on geophysical fluid dynamics (Oxford University Press, 1998).
15A. Majda and X. Wang, Nonlinear dynamics and statistical theories for basic geophysical flows (Cambridge University Press, 2006).
36P. Sapis, A. Vakakis, and L. Bergman, “Targeted energy transfer between dynamical components due to essential nonlinearities: A stochastic-per


74S. Shalev-Shwartz and S. Ben-David, Understanding machine learning: From theory to algorithms (Cambridge university press, 2014).


