scientific reports

OPEN

Variational multiscale reinforcement learning for discovering reduced order closure models of nonlinear spatiotemporal transport systems

Omer San[1]✉, Suraj Pawar[1] & Adil Rasheed[2,3]

A central challenge in the computational modeling and simulation of a multitude of science applications is to achieve robust and accurate closures for their coarse-grained representations due to underlying highly nonlinear multiscale interactions. These closure models are common in many nonlinear spatiotemporal systems to account for losses due to reduced order representations, including many transport phenomena in fluids. Previous data-driven closure modeling efforts have mostly focused on supervised learning approaches using high fidelity simulation data. On the other hand, reinforcement learning (RL) is a powerful yet relatively uncharted method in spatiotemporally extended systems. In this study, we put forth a modular dynamic closure modeling and discovery framework to stabilize the Galerkin projection based reduced order models that may arise in many nonlinear spatiotemporal dynamical systems with quadratic nonlinearity. However, a key element in creating a robust RL agent is to introduce a feasible reward function, which can be constituted of any difference metrics between the RL model and high fidelity simulation data. First, we introduce a multi-modal RL to discover mode-dependant closure policies that utilize the high fidelity data in rewarding our RL agent. We then formulate a variational multiscale RL (VMRL) approach to discover closure models without requiring access to the high fidelity data in designing the reward function. Specifically, our chief innovation is to leverage variational multiscale formalism to quantify the difference between modal interactions in Galerkin systems. Our results in simulating the viscous Burgers equation indicate that the proposed VMRL method leads to robust and accurate closure parameterizations, and it may potentially be used to discover scale-aware closure models for complex dynamical systems.

Reduced order models (ROMs) often refer to simplifications of high-fidelity models that capture dominant system dynamics using minimal computational resources. Over the past decades, we have witnessed an ever increasing number of reduced order modeling approaches and their enormous impact on fluid dynamics research[1–4]. A chief motivation behind these approaches being introduced is to be able to use ROMs in multi-query applications such as control and optimization[5,6].

Broadly speaking, closure modeling in fluid flow simulations refers to parameterizing the interactions between high-fidelity and coarse-grained descriptions. Although projection based ROMs have been utilized extensively in many fluid dynamics applications[1–4], they might yield inaccurate results when they are used in the under-resolved regime, i.e., when the number of modes is not large enough to capture the parameterized or transient dynamics of the underlying system[7]. Prior studies have suggested that closure models are efficacious in decreasing such modal truncation errors[4]. In fact, ROM closures can be viewed as correction or residual terms that are added to classical ROMs in order to model the effect of the discarded ROM modes in under-resolved simulations.

Consequently, an emerging thrust in modern ROM closure development efforts is to incorporate machine learning (ML) models[4,8–12]. The last decade has seen the growth of data-driven modeling technologies (e.g., deep

[1]School of Mechanical and Aerospace Engineering, Oklahoma State University, Stillwater, OK 74078, USA. [2]Department of Engineering Cybernetics, Norwegian University of Science and Technology, 7465 Trondheim, Norway. [3]Department of Mathematics and Cybernetics, SINTEF Digital, 7034 Trondheim, Norway. ✉email: osan@okstate.edu

neural networks). So far, a substantial body of closure modeling works has focused on supervised learning[9,13,14]. A detailed discussion on these models can be found in a recent survey[4]. In principle, the problem of building a data-driven closure model from the multimodal datasets can be posed as an optimization and ML task. Although supervised learning techniques become more of commodity tools nowadays, reinforcement learning (RL) is relatively an uncharted approach in computational science communities. In contrast to many other data-driven supervised learning based approaches introduced for closures[4], this powerful approach can be formulated to tackle with the closure problem in an automated fashion.

RL often provides a comprehensive iterative computational framework that implies modular goal-directed interactions of an agent with its environment. More recently, Novati et al.[15] demonstrated the power of RL in discovery of turbulence closure models in large-eddy simulation of turbulent flows. The RL framework being introduced in the turbulence modeling context is quite comprehensive, and might apply equally well to other coarse-grained reduced order modeling approaches. In relevant works[16,17], the feasibility of using RL to learn optimal ROM closures has been discussed. We also refer to a recent review[18] for the theory and application of RL approaches in fluid mechanics.

A fundamental question in RL is how to construct robust state, action and reward definitions relevant to the underlying problem. In this study, we put forth and examine the scale-aware RL mechanisms that automate closure modeling of projection based ROMs. Specifically, we first introduce a multi-modal RL (MMRL) approach, which discovers the mode dependant policies to stabilize the evolution of the truncated Galerkin system. One of the main objectives of our study is therefore to demonstrate how physical insights play a key role in designing an effective RL environment that converges to a robust control policy for learning and parameterizing the closure terms.

In fact, a key element in forging a robust RL agent is to introduce an appropriate reward function, which can be, in principle, constituted of any difference metrics between the RL model and high fidelity simulation data. In this study, instead, we explore how RL can be formulated using a variational multiscale approach to discover closure models without requiring access to the high fidelity data in designing the reward function. The general concept of the variational multiscale framework has been first introduced in finite element community[19–23], and its underpinning idea has been later adopted by ROM modellers[24–29]. In our study, we put forth a variational multiscale RL (VMRL) approach by leveraging this multiscale concept that introduces a natural hierarchy to quantify the difference between modal interactions in the Galerkin projection based ROM systems. Specifically, our work addresses the following questions:

- How can RL be used to discover reliable closure models in reduced order models of transport equations?
- Which parameterization processes lead to improved closure approaches that may reduce uncertainty in the evolution of projection coefficients of the Galerkin ROM systems?
- How does a mode-dependent closure formulation affect overall predictive performance?
- What are the design considerations for formulating a feasible reward function that does not require access to the supervised training data?

Therefore, the main goal of this paper is to address these questions in the context of closure model discovery for complex nonlinear spatiotemporal systems.

Methods

Reduced order modeling.

To illustrate the proposed approaches, we focus on the viscous Burgers equation, a generic partial differential equation that represents broad nonlinear transport phenomena in fluid dynamics, which is given as[4]

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = \nu\frac{\partial^2 u}{\partial x^2}, \tag{1}$$

where u refers to velocity, and ν is the kinematic viscosity (i.e., $\nu = 1/\mathrm{Re}$ in dimensionless form, where Re refers to the Reynolds number). From a model reduction perspective, we highlight that Eq. (1) possesses the hallmarks of general nonlinear multidimensional advection-diffusion problems[30]. Defining our spatiotemporal domain, $x \in [0, 1]$ and $t \in [0, 1]$, the viscous Burgers equation admits an analytical solution in the form of[30,31]

$$u(x,t) = \frac{\frac{x}{t+1}}{1 + \sqrt{\frac{t+1}{t_0}}\exp(\frac{x^2}{4\nu(t+1)})}, \tag{2}$$

where $t_0 = \exp(\frac{1}{8\nu})$. This closed form expression is used to generate snapshot data for our forthcoming model order reduction analysis. The database is constituted with Eq. (2) using $N = 1024$ spatial collocation points at each snapshot. We set $\nu = 0.001$ (i.e., $Re = 1000$) to generate our training data and our training database consists of 500 snapshots from $t = 0$ to $t = 1$. In considering such spatiotemporal system, we often use a modal decomposition to the field $u(x, t)$

$$u(x,t) = \sum_{k=1}^{R} \alpha_k(t)\psi_k(x), \tag{3}$$

where $\alpha_k(t)$ and $\psi_k(x)$ refer to the kth modal coefficient and kth proper orthogonal decomposition (POD) basis function, respectively. Figure 1 shows the the first eight most energetic POD basis functions utilized in this

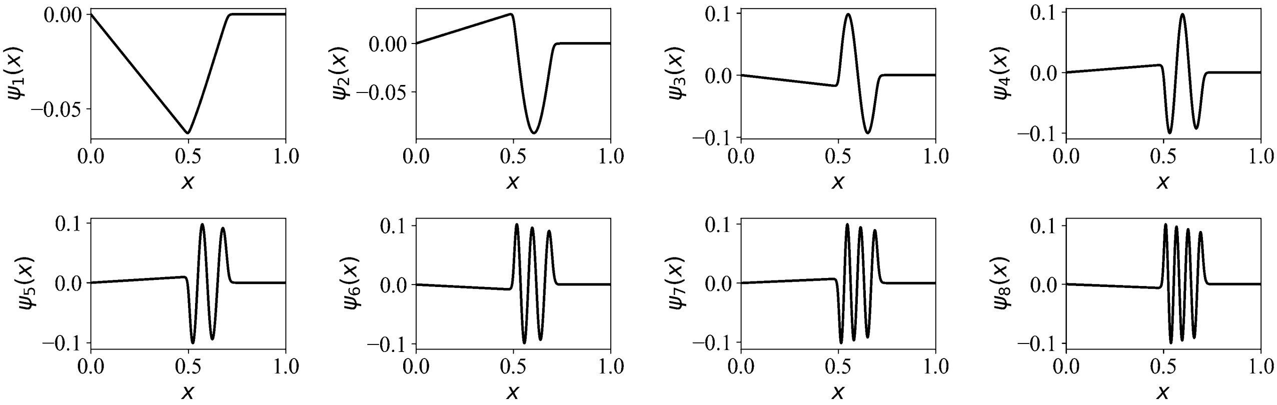

Figure 1. Illustration of the first eight POD basis functions generated using a total of 500 data snapshots at $Re = 1000$.

study. We note that, without losing generality, one can also use Fourier harmonics[9] or randomized orthogonal functions[32] to forge a set of basis functions. Here, we note that access to a set of previously recorded data snapshots is necessary in order to generate the POD basis functions. The proposed framework, however, is not necessarily dependent on the POD procedure and is easily adaptable to different base functions. In other words, let's take into account the following layers when formulating our problem: (i) create basis functions, (ii) develop a projection-based reduced order model, (iii) establish an ansatz/model for closures, and (iv) use reinforcement learning to learn the parameters of the ansatz/model. We stress that the data has only been used in step (i) when we construct bases here since we concentrate on the POD-based model reduction strategy. However, our approach can be used effectively with the use of different bases, such as Fourier basis functions.

Once a set of spatial orthonormal modes (i.e., for $k = 1, 2, \ldots, R$) is obtained from snapshot data, we then apply the Galerkin projection (GP) to obtain a dynamical system that evolves in the latent space of $\alpha_k(t)$. The resulting ROM, denoted as GP in this study, becomes

$$\frac{d\alpha_k}{dt} = \sum_{i=1}^{R} \mathfrak{L}_k^i \alpha_i + \sum_{i=1}^{R}\sum_{j=1}^{R} \mathfrak{N}_k^{ij} \alpha_i \alpha_j, \quad \forall\, k = 1, \ldots, R \tag{4}$$

where

$$\mathfrak{L}_k^i = \left(\nu \frac{\partial^2 \psi_i(x)}{\partial x^2}, \psi_k(x) \right), \tag{5}$$

$$\mathfrak{N}_k^{ij} = \left(-\psi_i(x) \frac{\partial \psi_j(x)}{\partial x}, \psi_k(x) \right), \tag{6}$$

where the notion of $(\cdot,\cdot)$ represents the standard inner product. We note that these tensorial coefficients in GP model only depend on spatial modes, which are often precomputed from the available snapshot data when designing projection based ROMs.

Closure modeling. We first illustrate the underlying closure modeling concept for a prototype demonstration as shown in Fig. 2. To formulate our ROM closure problem, we modify Eq. (4) by adding a functional form of distributed control. This control is often referred to as eddy viscosity approach that has strong roots in large eddy simulations to model or compensate the residual effects of the truncated scales[33–36]. Therefore, the modified ROM becomes

$$\frac{d\alpha_k}{dt} = \sum_{i=1}^{R} \mathfrak{L}_k^i \alpha_i + \sum_{i=1}^{R} \tilde{\mathfrak{L}}_k^i \alpha_i + \sum_{i=1}^{R}\sum_{j=1}^{R} \mathfrak{N}_k^{ij} \alpha_i \alpha_j, \quad \forall\, k = 1, \ldots, R \tag{7}$$

where the proposed closure term can be parameterized by defining an eddy viscosity coefficient η as follows

$$\tilde{\mathfrak{L}}_k^i = \left(\eta \frac{\partial^2 \psi_i(x)}{\partial x^2}, \psi_k(x) \right). \tag{8}$$

Several techniques have been introduced to improve the accuracy of closure parameterizations, including definition of a nonlinear eddy viscosity model[37] or dynamic closure models[38,39] that allow varying eddy viscosity in time (i.e., $\eta(t) \leftarrow \eta$). In this paper, we first formulate an RL environment and design an agent to discover this eddy viscosity parameter $\eta(t)$. We call this approach linear-mode RL (LMRL).

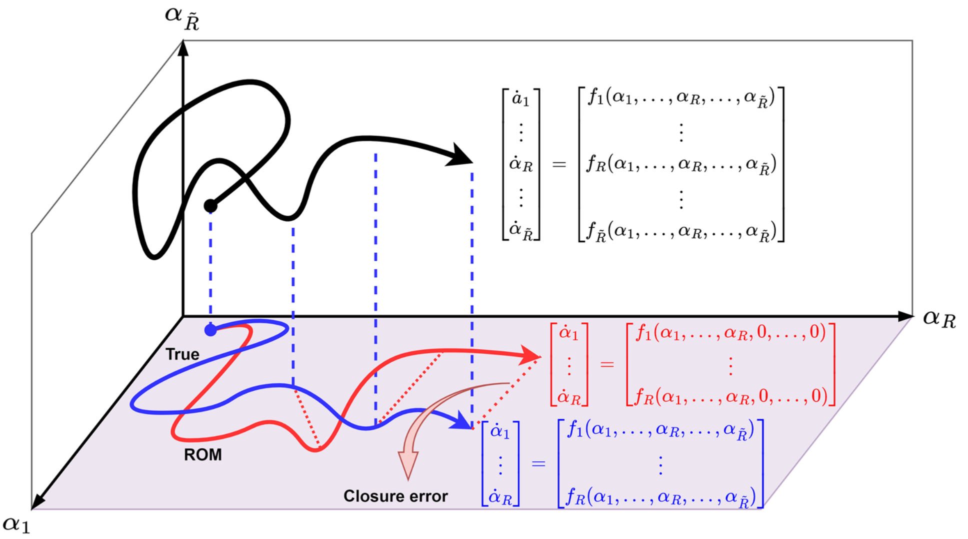
$\alpha_{\tilde{R}}$
$\begin{bmatrix}\dot{\alpha}_1\\ \vdots \\ \dot{\alpha}_R \\ \vdots \\ \dot{\alpha}_{\tilde{R}}\end{bmatrix} = \begin{bmatrix} f_1(\alpha_1,\dots,\alpha_R,\dots,\alpha_{\tilde{R}}) \\ \vdots \\ f_R(\alpha_1,\dots,\alpha_R,\dots,\alpha_{\tilde{R}}) \\ \vdots \\ f_{\tilde{R}}(\alpha_1,\dots,\alpha_R,\dots,\alpha_{\tilde{R}})\end{bmatrix}$
α_R
True
$\begin{bmatrix}\dot{\alpha}_1\\ \vdots \\ \dot{\alpha}_R\end{bmatrix} = \begin{bmatrix} f_1(\alpha_1,\dots,\alpha_R,0,\dots,0) \\ \vdots \\ f_R(\alpha_1,\dots,\alpha_R,0,\dots,0)\end{bmatrix}$
ROM
$\begin{bmatrix}\dot{\alpha}_1\\ \vdots \\ \dot{\alpha}_R\end{bmatrix} = \begin{bmatrix} f_1(\alpha_1,\dots,\alpha_R,\dots,\alpha_{\tilde{R}}) \\ \vdots \\ f_R(\alpha_1,\dots,\alpha_R,\dots,\alpha_{\tilde{R}})\end{bmatrix}$
Closure error
α_1

Figure 2. A schematic overview of the closure modeling in a hypothetical three-dimensional latent space. When we truncate a model spanned in a higher dimensional state space (i.e., $\tilde{R} = 3$ in the figure, a three-dimensional model spanned in $\alpha_1, \alpha_2, \alpha_3$) to a lower dimensional latent space (i.e., $R = 2$ in the figure, a reduced order two-dimensional model spanned only in α_1, α_2), a closure error will be introduced due to the underlying nonlinear interactions. The main goal in closure modeling is to find a parameterized model, which is only function of resolved modal coefficients (i.e., $\alpha_1, \dots, \alpha_R$) to minimize this closure error. Therefore, in this paper we formulate an RL problem to discover this closure model using the state variables (i.e., resolved modal coefficients).

In their seminal work, Östh et al.[40] further enhanced the closure theory emphasizing the modal eddy viscosity concept. The roots of such mode-dependent correction go back to the work of Rempfer and Fasel[41] in order to provide improved closure models. These multi-modal closures can be specified as

$$\tilde{\mathfrak{L}}^i_k = \left(\eta_k \frac{\partial^2 \psi_i(x)}{\partial x^2}, \psi_k(x)\right). \tag{9}$$

where η_k refers to the kth modal eddy viscosity coefficient. In our current work, we formulate an RL framework to learn $\eta_k(t)$, and call this approach as multi-modal RL (MMRL). Although the proposed closure problem can be formulated using more traditional adjoint based[37] or sensitivity based approaches[42], our chief motivation in this study is to explore the feasiblity of RL workflows for the ROM closure problems. More specifically, in this paper we aim to introduce a variational multiscale RL (VMRL) approach by formulating a new procedure to forge a reward function that does not require access to the training data. Our approach therefore facilitates new RL workflows since RL enhanced computational frameworks might play an integral role in designing many end-to-end data-driven approaches for broader optimization and control problems.

Deep reinforcement learning.

In our context, deep RL presents a modular computational framework to learn $\eta_k(t)$ in Eq. (9). Here, we briefly describe the formulation of the RL problem and the proximal policy optimization (PPO) algorithm[43]. In RL, at each time step t, the agent observes some representation of the state of the system, $s_t \in \mathscr{S}$, and based on this observation selects an action, $a_t \in \mathscr{A}$. The agent receives the reward, $r_t \in \mathscr{R}$ as a consequence of the action and the environment enters in a new state s_{t+1}. Therefore, the interaction of an agent with the environment gives rise to a trajectory as follows

$$\tau = \{s_0, a_0, r_0, s_1, a_1, r_1, \dots\}. \tag{10}$$

The goal of the RL is to find an optimal strategy for the agent that will maximize the expected discounted reward over the trajectory τ and can be written mathematically as follows

$$\mathfrak{R}(\tau) = \sum_{t=0}^{T} \gamma^t r_t, \tag{11}$$

where γ is a parameter called discount rate that lies between [0, 1], and T is the horizon of the trajectory. The discount rate determines how much weightage to be assigned to the long-term reward compared to an immediate reward.

In RL, the agent's decision making strategy is characterized by a policy $\pi(s, a) \in \Pi$. The RL agent is trained to find a policy to optimize the expected return when starting in the state s at time step t and is called as state-value function. We can write the state-value function as follows

$$V^{\pi}(s) = \mathbb{E}_{\pi}\left[\sum_{k=0}^{\infty} \gamma^{k} r_{t+k} | s_t = s, \pi\right]. \tag{12}$$

Similarly, the expected return starting in a state s, taking an action a, and thereafter following a policy π is called as the action-value function and can be written as

$$Q^{\pi}(s,a) = \mathbb{E}_{\pi}\left[\sum_{k=0}^{\infty} \gamma^{k} r_{t+k} | s_t = s, a_t = a, \pi\right]. \tag{13}$$

We also define an advantage function that can be considered as an another version of action-value function with lower variance by taking the state-value function as the baseline. The advantage function can be written as

$$A^{\pi}(s,a) = Q^{\pi}(s,a) - V^{\pi}(s). \tag{14}$$

We use $\pi_w(\cdot)$ to denote that the policy is parameterized by $w \in \mathbb{R}^d$ (i.e., the weights and biases of the neural network in deep RL). The agent is trained with an objective function defined as[44]

$$J(w) \doteq V^{\pi_w}(s_0), \tag{15}$$

where an episode starts in some particular state s_0, and V^{π_w} is the value function for the policy π_w. The policy parameters w are updated by estimating the gradient of an objective function and plugging it into a gradient ascent algorithm as follows

$$w \leftarrow w + \beta \nabla_w J(w), \tag{16}$$

where β is the learning rate of the optimization algorithm. The gradient of an objective function can be computed using the policy gradient theorem[45] as follows

$$\nabla_w V^{\pi_w}(s_0) = \mathbb{E}_{\pi_w}\left[\nabla_w\left(\log\ \pi_w(s,a)\right) Q^{\pi_w}(s,a)\right]. \tag{17}$$

The accurate calculation of empirical expectation in Eq. (17) requires large number of samples. Furthermore, the performance of policy gradient methods is highly sensitive to the learning rate leading to difficulty in obtaining stable and steady improvement. The PPO algorithm introduces a clipped surrogate objective function[43] to avoid excessive update in policy parameters in a simplified way as follows

$$J^{\text{clip}}(w) = \mathbb{E}\left[\min(r_t(w)A^{\pi_w}(s,a), \text{clip}(r_t(w), 1-\varepsilon, 1+\varepsilon)A^{\pi_w}(s,a))\right], \tag{18}$$

where $r_t(w)$ denotes the probability ratio between new and old policies as given below

$$r_t(w) = \frac{\pi_{w+\Delta w}(s,a)}{\pi_w(s,a)}. \tag{19}$$

The ε in Eq. (18) is a hyperparameter that controls how much new policy can deviate from the old. This is done through a function $\text{clip}(r_t(w), 1-\varepsilon, 1+\varepsilon)$ that enforces the ratio between new and old policy ($r_t(w)$) to stay between the limit $[1-\varepsilon, 1+\varepsilon]$.

Variational multiscale approach. Here we present a two-scale variational multiscale formulation as depicted in Fig. 3, which utilize two orthogonal spaces, $\boldsymbol{X}_A$ and $\boldsymbol{X}_B$. Since the POD basis is orthonormal by construction, we can build these two orthogonal spaces in a natural, straightforward way: $\boldsymbol{X}_A := \text{span}\{\psi_1, \psi_2, \ldots, \psi_R\}$, which represents the resolved ROM scales, and $\boldsymbol{X}_B := \text{span}\{\psi_{R+1}, \psi_{R+2}, \ldots, \psi_{\tilde{R}}\}$, which represents the test scales (i.e., unresolved ROM scales). Following Eq. (4), next we use the ROM approximation of u in the space $\boldsymbol{X}_A \oplus \boldsymbol{X}_B$,

$$u(x,t) = \sum_{k=1}^{R} \alpha_k(t)\psi_k(x) + \sum_{k=R+1}^{\tilde{R}} \alpha_k(t)\psi_k(x), \tag{20}$$

where the first term in the right hand side of Eq. (20) represents the resolved ROM components of u, and the second term represents the unresolved test scales. Plugging the ROM approximation of u in Eq. (1), projecting it onto $\boldsymbol{X}_A$, and using ROM basis orthogonality, we obtain

$$\frac{d\alpha_k}{dt} = \sum_{i=1}^{\tilde{R}} \mathfrak{L}_k^i \alpha_i + \sum_{i=1}^{\tilde{R}} \sum_{j=1}^{\tilde{R}} \mathfrak{N}_k^{ij} \alpha_i \alpha_j, \quad \forall\, k = 1, \ldots, R. \tag{21}$$

To describe the hierarchical structure of the ROM basis, we make use of the variational framework. Therefore, the scales are naturally divided into two groups in the first stage using ROM projection: (i) resolved scales and (ii) unresolved scales. The phrases describing the relationships between the two categories of scales are expressly identified in the second stage. The novelty of the proposed framework is demonstrate in the third step, where we build our reinforcement learning based closure models for the interaction between the two types of scales. We also highlight that the choice of cut-off scale R and the test scale $\tilde{R}$ might have impact on the results of the approach. Moreover, in POD-based reduced order models, raising R typically results in an increase for the

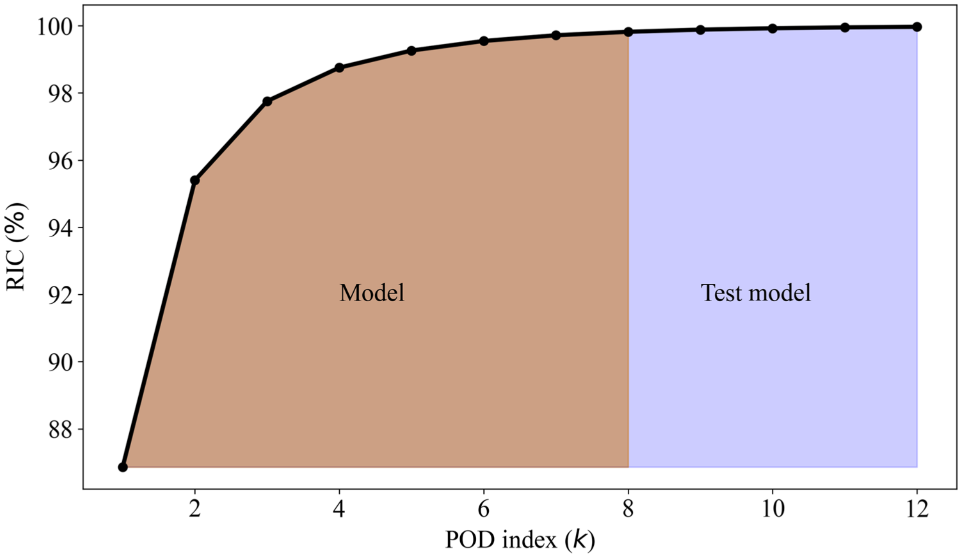

Figure 3. Relative information content (RIC) values as a function of the POD index. Test scales,which are used to build our reward function, represent the contribution of the under-resolved ROM scales.

accuracy[4], whereas $\tilde{R}$ can be taken as a very modest value for an efficient learning performance. Furthermore, there is an analogy between our approach and large eddy simulations where a test filter scale has been traditional used to estimate the coefficients of the dynamic models[39,46].

In summary, our RL environment consists of three model definition for the evolution of the state variables $\alpha_1, \alpha_2, \ldots, \alpha_R$: (i) base ROM given by Eq. (4), (ii) improved ROM by the closure model given by Eq. (7), and (iii) test model given by Eq. (21). Our key hypothesis relies on the fact that the proposed closure model is accurate and representative if the difference between states obtained by Eqs. (7) and 21 is minimized. Therefore, the reward function in our RL framework can now be easily constructed by exploiting the difference between these resolved and test scale modal coefficients. More precisely, let's define the following states at time t: $s_t^{base} := \{\alpha_1, \alpha_1, \ldots, \alpha_R\}$ as the solution of Eq. (4), $s_t^{ROM} := \{\alpha_1, \alpha_1, \ldots, \alpha_R\}$ as the solution of our ROM given by Eq. (7), and $s_t^{test} := \{\alpha_1, \alpha_1, \ldots, \alpha_R\}$ as the solution of Eq. (21). Then we can reward our closure policy according the following definition of the binary reward function:

$$r_t = \begin{cases} +10, & \text{if } \sigma ||s_t^{base} - s_t^{ROM}|| < ||s_t^{base} - s_t^{test}|| \\ -10, & \text{otherwise} \end{cases} \tag{22}$$

where $\sigma > 1$ is a scaling factor that can be chosen between 1 and 2 in practice. In our calculations, we set $\sigma = 1.6$. We note that this binary definition of reward function eliminates the need for access to the snapshot data as we will detail further in our results section. The selection of ±10 in our binary definition is arbitrary since the RL workflows are designed to maximize the sum of the reward over each episodic experiment.

Results

Figure 4 displays the complete deep RL framework for the MMRL approach where the agent observes the POD modal coefficients as the state of the system and takes the action of selecting modal eddy viscosity coefficients. Due to the modal truncation, the effect of unresolved scales on resolved scales is not captured and there is a discrepancy between the true modal coefficients and the ROM modal coefficients. The goal of the agent is to minimize this difference, and therefore, the l_2 norm of the deviation between true and ROM modal coefficients is used as the reward function. Since we are maximizing the reward, the negative of the l_2 norm is first assigned as a reward at each time step that is given by

$$r_t = -||s_t^{ROM} - (\hat{u}, \psi_k)|| \tag{23}$$

where $\hat{u} := u(x, t)$ refers to snapshot data at time step t. The choice of the reward function can have a significant effect on the performance of the agent[15] and needs to be carefully designed depending upon the problem. The performance of the MMRL approach is compared against the linear-modal RL (LMRL) approach where the agent selects only a scalar value of eddy viscosity amplitude as an action and linear viscosity kernel[35] is utilized to assign the modal eddy viscosity coefficients. Specifically, in LMRL approach we have $\eta_k(t) = \eta_e(t)(k/R)$, where η_e is the eddy viscosity amplitude selected by the agent as an action.

In Fig. 5, the trajectory of the mean reward is shown for training an RL agent with the LMRL and MMRL approaches. The agent is trained for Reynolds number $Re = 1000$. It can be seen that the maximum reward attained with the MMRL approach is almost twice the magnitude of the reward achieved with the LMRL approach. Figure 6 depicts the evolution of selected POD modal coefficients for $Re = 1500$. The prediction from both LMRL and MMRL approaches is in better agreement with the true projection modal coefficients compared

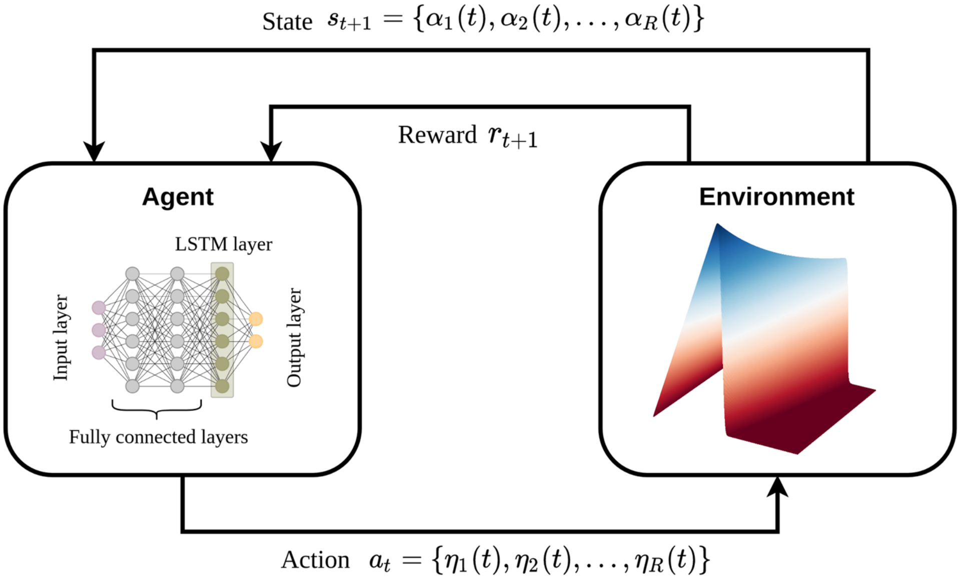
State $s_{t+1} = \{\alpha_1(t), \alpha_2(t), \ldots, \alpha_R(t)\}$
Reward r_{t+1}
Agent
LSTM layer
Input layer
Output layer
Fully connected layers
Environment
Action $a_t = \{\eta_1(t), \eta_2(t), \ldots, \eta_R(t)\}$

Figure 4. Schematic of the deep RL framework for MMRL ROM closure approach. The RL agent observes modal coefficients and selects model eddy viscosity coefficient as an action.

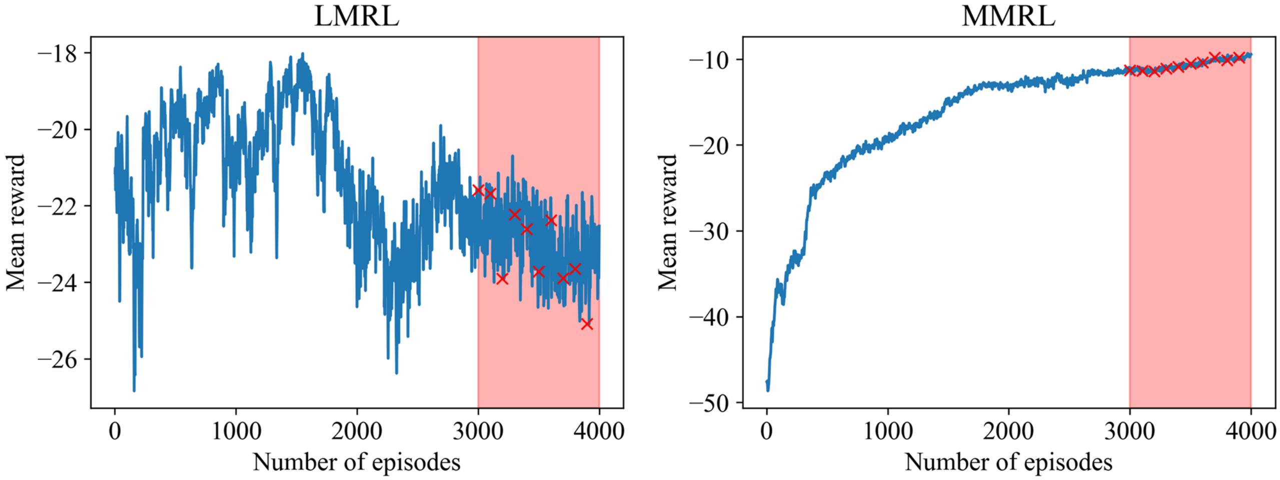

Figure 5. Evolution of the moving average mean reward for LMRL (left) and MMRL (right) approaches. The models used for testing are indicated by red symbols.

to GP with the prediction from MMRL being more accurate compared to LMRL. However, we highlight that both LMRL and MMRL approaches utilize the reward function given by Eq. (23), which requires access to the true snapshot data. We highlight that in our evaluation, both mean and two-standard deviation (i.e., 95% confidence interval) of 10 different RL models are shown (e.g., see red symbols in Fig. 5 for those models).

On the other hand, Fig. 7 illustrates the trajectory of the mean reward for training an RL agent with the VMRL approach that utilizes a binary reward function given by Eq. (22). Figure 8 shows a comparison between MMRL and VMRL approaches at $Re = 1500$. We highlight that both approaches utilizes multi-modal action space (i.e., discovering $\eta_k(t)$ for $k = 1, 2, \ldots, R$). Figure 8 clearly demonstrates that the VMRL approach obtains an accurate policy without requiring access to the true labeled data in defining the reward function. This key aspect of the proposed VMRL approach paves the way of designing novel RL workflows exploiting the modal interaction between resolved and test scales.

The spatiotemporal velocity field with different ROMs is shown in Fig. 9. It should be noted that we can at the most recover the true projection of the full order model (FOM) solution. This true reduced order representation (ROR) is also shown in Fig. 9. In our analogy given in Fig. 2, the blue curve represents the ROR, the red curve represents the GP model. For quantitative analysis, the root mean squared error (RMSE) for different ROM approaches at different Reynolds numbers (different from the training setting at $Re = 1000$) is reported in Table 1. The RMSE for LMRL, MMRL, and VMRL approaches is significantly smaller compared the the GP model. We also observe that the VMRL approach provides marginally more accurate solution than the LMRL and MMRL approaches at higher Reynolds number. The trends for the LMRL and MMRL techniques indicate that

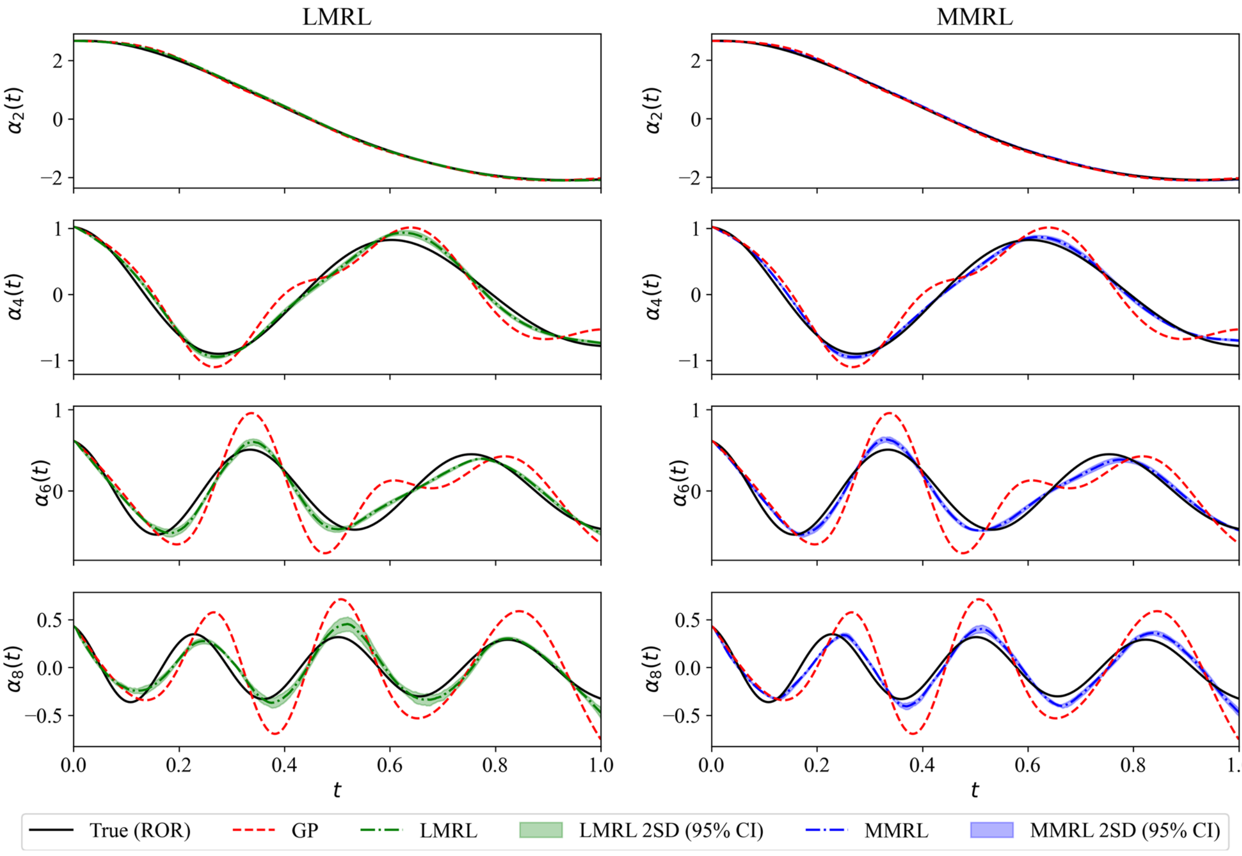

Figure 6. Evolution of the second, fourth, sixth and the last POD modal coefficients at $Re = 1500$ for LMRL (left) and MMRL (right) approaches.

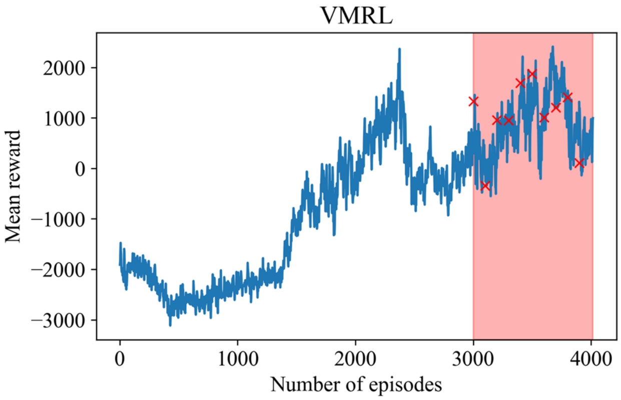

Figure 7. Evolution of the moving average mean reward for VMRL approaches. The models used for testing are indicated by red symbols.

error increases as Reynolds number rises, as expected. However, the error for $Re = 1200$ is slightly bigger for the VMRL technique than it is for $Re = 1500$. This might be related to the underlying modeling hyperparameters, and that somewhat different results might be obtained by using a different RL architecture.

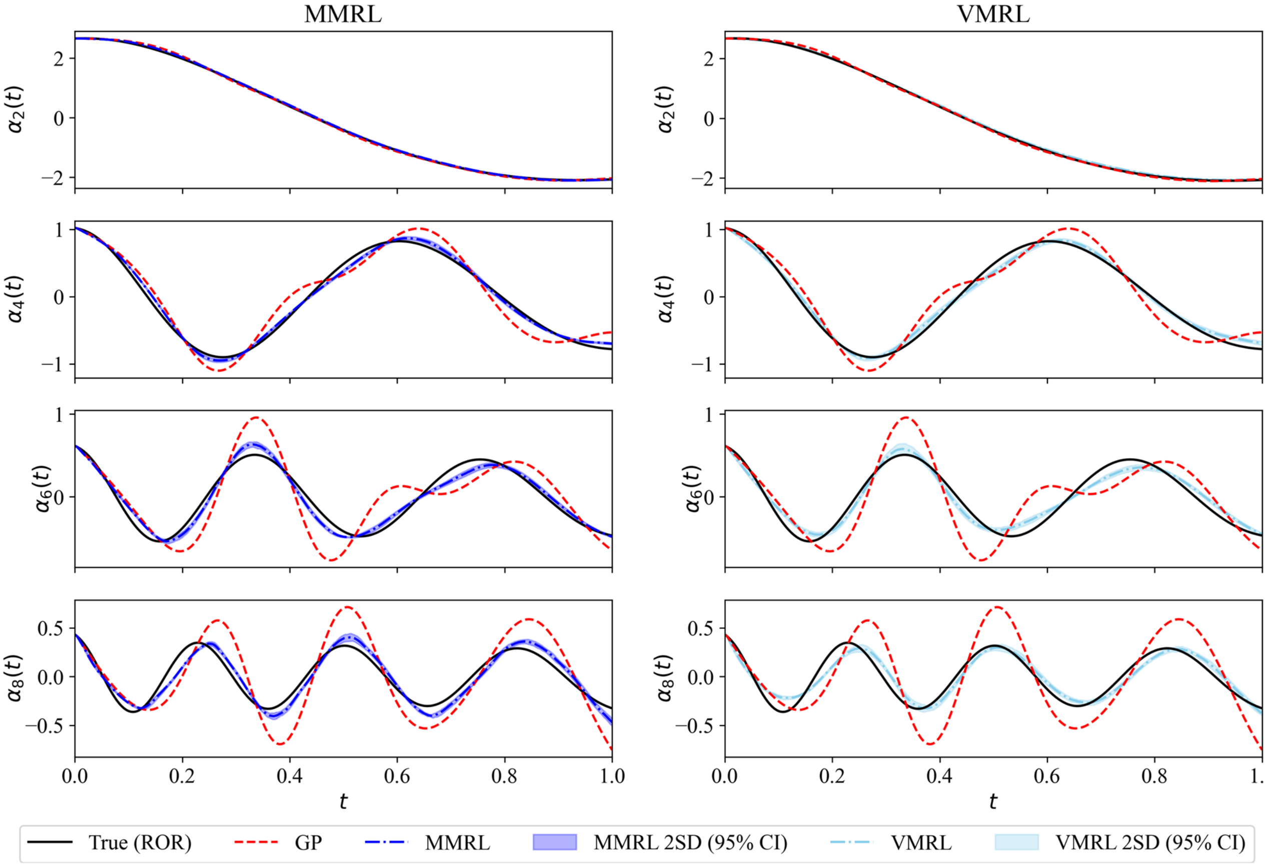

Figure 8. Evolution of the second, fourth, sixth and the last POD modal coefficients at $Re = 1500$ for MMRL (left) and VMRL (right) approaches.

Discussion

This study introduces scale-aware reinforcement learning (RL) framework to automate the discovery of closure models in projection based ROMs. We treat the closure as a control input in the latent space of the ROM and build the parameterized model with a dissipation term. The feasibility of the RL framework is first demonstrated with linear-modal RL (LMRL) where a linear eddy viscosity constraint is utilized for parameterization and with multi-modal RL (MMRL) which finds mode dependant eddy viscosity model coefficient. The agent is incorporated in a reduced-order solver, observes the POD modal coefficients, and accordingly computes the closure term. Here, both RL approaches minimize the discrepancy between the true POD modal coefficients and prediction from closure ROM, and the obtained closure model generalizes to different Reynolds number. We then demonstrate how to formulate an RL framework without requiring access to the true data using the variational multiscale formalism. We find that this variational-multiscale RL (VMRL) is a robust closure discovery framework that utilizes a reward function based on the modal energy transfer effect. Building on the promising results presented in this study to develop ROM closure models using deep RL, our future work will concentrate on incorporating uncertainties associated with observations into account while selecting the action of an agent. Although our report demonstrates the feasibility of the proposed RL framework while considering the Burgers equation, more complex examples need to be examined in order to conclude whether the proposed methods are indeed applicable to simulations of complex phenomena, a subject we intend to investigate in the future. The key benefit of utilizing reinforcement learning in this reduced order modeling context is its modular nature, which allows for the identification and learning of new closure modeling approaches, despite the fact that it is likely one of the most computational time- and resource-intensive disciplines of machine learning.

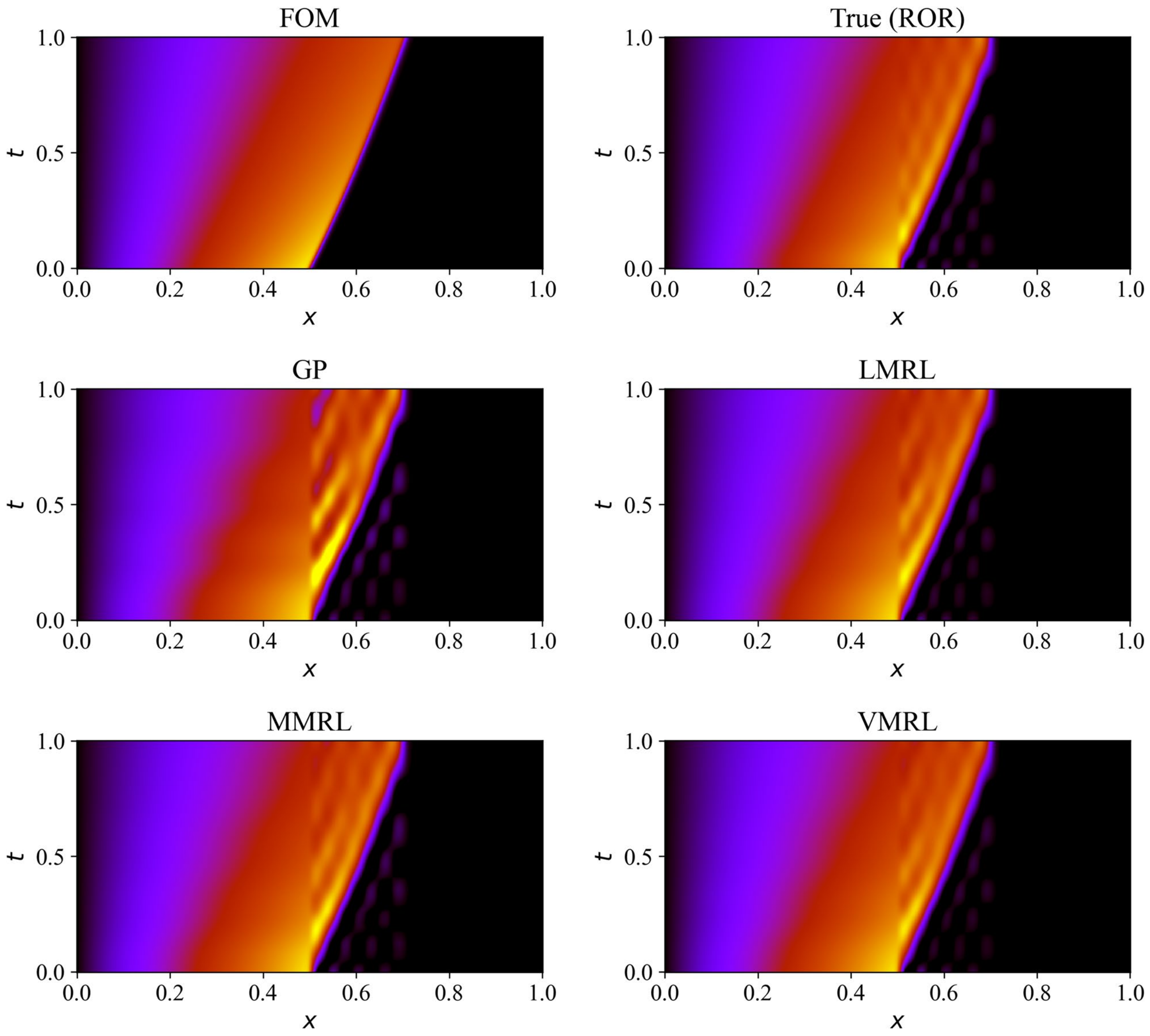

Figure 9. Spatiotemporal visualization of the velocity field at $Re = 1500$ for different modeling approaches.

ROM	Action	RMSE ($Re = 1200$)	RMSE ($Re = 1500$)	RMSE ($Re = 2000$)
GP	–	11.837×10^{-3}	16.217×10^{-3}	22.809×10^{-3}
LMRL	$a_t \in \{\eta_e(t)\}$	4.645×10^{-3}	6.144×10^{-3}	10.092×10^{-3}
MMRL	$a_t \in \{\eta_1(t), \eta_2(t), \ldots, \eta_R(t)\}$	3.111×10^{-3}	5.258×10^{-3}	9.746×10^{-3}
VMRL	$a_t \in \{\eta_1(t), \eta_2(t), \ldots, \eta_R(t)\}$	5.341×10^{-3}	5.262×10^{-3}	8.063×10^{-3}

Table 1. ℓ_2 norm for the deviation of the velocity with respect to its true projection value for $t \in [0, 1]$. We note that both snapshot data generation for POD analysis and RL training are performed at $Re = 1000$. Here LMRL and MMRL models use the reward function defined in Eq. (23) utilizing the true snapshot data, whereas the VMRL model uses the reward function defined in Eq. (22) that utilizes the variational multiscale formalism. We note that computational time of the RL training is about 5.1 s for each episode that has 1000 time steps.

Data availability

The data that supports the findings of this study is available within the article. The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 26 June 2022; Accepted: 17 October 2022
Published online: 26 October 2022

References

1. Lucia, D. J., Beran, P. S. & Silva, W. A. Reduced-order modeling: new approaches for computational physics. *Prog. Aerosp. Sci.* **40**, 51–117 (2004).
2. Rowley, C. W. & Dawson, S. T. Model reduction for flow analysis and control. *Annu. Rev. Fluid Mech.* **49**, 387–417 (2017).
3. Taira, K. *et al.* Modal analysis of fluid flows: Applications and outlook. *AIAA J.* **58**, 998–1022 (2020).
4. Ahmed, S. E. *et al.* On closures for reduced order models-a spectrum of first-principle to machine-learned avenues. *Phys. Fluids* **33**, 091301 (2021).
5. Benner, P., Gugercin, S. & Willcox, K. A survey of projection-based model reduction methods for parametric dynamical systems. *SIAM Rev.* **57**, 483–531 (2015).
6. Peherstorfer, B., Willcox, K. & Gunzburger, M. Survey of multifidelity methods in uncertainty propagation, inference, and optimization. *SIAM Rev.* **60**, 550–591 (2018).
7. Snyder, W. *et al.* Reduced order model closures: A brief tutorial. arXiv preprint arXiv:2202.14017 (2022).
8. Milano, M. & Koumoutsakos, P. Neural network modeling for near wall turbulent flow. *J. Comput. Phys.* **182**, 1–26 (2002).
9. San, O. & Maulik, R. Neural network closures for nonlinear model order reduction. *Adv. Comput. Math.* **44**, 1717–1750 (2018).
10. Pawar, S. *et al.* A deep learning enabler for nonintrusive reduced order modeling of fluid flows. *Phys. Fluids* **31**, 085101 (2019).
11. Pan, S. & Duraisamy, K. Data-driven discovery of closure models. *SIAM J. Appl. Dyn. Syst.* **17**, 2381–2413 (2018).
12. Gupta, A. & Lermusiaux, P. F. Neural closure models for dynamical systems. *Proc. R. Soc. A* **477**, 20201004 (2021).
13. San, O. & Maulik, R. Extreme learning machine for reduced order modeling of turbulent geophysical flows. *Phys. Rev. E* **97**, 042322 (2018).
14. Ahmed, S. E., San, O., Rasheed, A. & Iliescu, T. A long short-term memory embedding for hybrid uplifted reduced order models. *Physica D* **409**, 132471 (2020).
15. Novati, G., de Laroussilhe, H. L. & Koumoutsakos, P. Automating turbulence modelling by multi-agent reinforcement learning. *Nat. Mach. Intell.* **3**, 87–96 (2021).
16. Benosman, M., Chakrabarty, A. & Borggaard, J. Reinforcement learning-based model reduction for partial differential equations. *IFAC-PapersOnLine* **53**, 7704–7709 (2020).
17. Benosman, M., Chakrabarty, A. & Borggaard, J. Reinforcement learning-based model reduction for partial differential equations: Application to the burgers equation. In *Handbook of Reinforcement Learning and Control*, 293–317 (Springer, 2021).
18. Garnier, P. *et al.* A review on deep reinforcement learning for fluid mechanics. *Comput. Fluids* **225**, 104973 (2021).
19. Hughes, T. J., Feijóo, G. R., Mazzei, L. & Quincy, J.-B. The variational multiscale method-a paradigm for computational mechanics. *Comput. Methods Appl. Mech. Eng.* **166**, 3–24 (1998).
20. Hughes, T. J., Mazzei, L. & Jansen, K. E. Large eddy simulation and the variational multiscale method. *Comput. Vis. Sci.* **3**, 47–59 (2000).
21. Hughes, T. J., Oberai, A. A. & Mazzei, L. Large eddy simulation of turbulent channel flows by the variational multiscale method. *Phys. Fluids* **13**, 1784–1799 (2001).
22. Codina, R., Badia, S., Baiges, J. & Principe, J. Variational multiscale methods in computational fluid dynamics. Encyclopedia of Computational Mechanics Second Edition 1–28 (2018).
23. John, V. *Finite Element Methods for Incompressible Flow Problems* (Springer, 2016).
24. Stabile, G., Ballarin, F., Zuccarino, G. & Rozza, G. A reduced order variational multiscale approach for turbulent flows. *Adv. Comput. Math.* **45**, 2349–2368 (2019).
25. Reyes, R. & Codina, R. Projection-based reduced order models for flow problems: A variational multiscale approach. *Comput. Methods Appl. Mech. Eng.* **363**, 112844 (2020).
26. Tello, A., Codina, R. & Baiges, J. Fluid structure interaction by means of variational multiscale reduced order models. *Int. J. Numer. Methods Eng.* **121**, 2601–2625 (2020).
27. Mou, C., Koc, B., San, O., Rebholz, L. G. & Iliescu, T. Data-driven variational multiscale reduced order models. *Comput. Methods Appl. Mech. Eng.* **373**, 113470 (2021).
28. Koc, B. et al. Verifiability of the data-driven variational multiscale reduced order model. arXiv preprint arXiv:2108.04982 (2021).
29. Ahmed, S. E., San, O., Rasheed, A., Iliescu, T. & Veneziani, A. Physics guided machine learning for variational multiscale reduced order modeling. arXiv preprint arXiv:2205.12419 (2022).
30. San, O., Maulik, R. & Ahmed, M. An artificial neural network framework for reduced order modeling of transient flows. *Commun. Nonlinear Sci. Numer. Simul.* **77**, 271–287 (2019).
31. Maleewong, M. & Sirisup, S. On-line and off-line POD assisted projective integral for non-linear problems: A case study with Burgers' equation. *Int. J. Math. Comput. Phys. Electr. Comput. Eng.* **5**, 984–992 (2011).
32. Bistrian, D. A., San, O. & Navon, I. M. Digital twin data modelling by randomized orthogonal decomposition and deep learning. arXiv preprint arXiv:2206.08659 (2022).
33. Borggaard, J., Iliescu, T. & Roop, J. P. A bounded artificial viscosity large eddy simulation model. *SIAM J. Numer. Anal.* **47**, 622–645 (2009).
34. Akhtar, I., Wang, Z., Borggaard, J. & Iliescu, T. A new closure strategy for proper orthogonal decomposition reduced-order models. *J. Comput. Nonlinear Dyn.* **7** (2012).
35. San, O. & Iliescu, T. Proper orthogonal decomposition closure models for fluid flows: Burgers equation. *Int. J. Numer. Anal. Model. Ser. B* **5**, 285–305 (2014).
36. San, O. & Iliescu, T. A stabilized proper orthogonal decomposition reduced-order model for large scale quasigeostrophic ocean circulation. *Adv. Comput. Math.* **41**, 1289–1319 (2015).
37. Cordier, L. *et al.* Identification strategies for model-based control. *Exp. Fluids* **54**, 1–21 (2013).
38. Wang, Z., Akhtar, I., Borggaard, J. & Iliescu, T. Proper orthogonal decomposition closure models for turbulent flows: A numerical comparison. *Comput. Methods Appl. Mech. Eng.* **237**, 10–26 (2012).
39. Rahman, S. M., Ahmed, S. E. & San, O. A dynamic closure modeling framework for model order reduction of geophysical flows. *Phys. Fluids* **31**, 046602 (2019).
40. Östh, J., Noack, B. R., Krajnović, S., Barros, D. & Borée, J. On the need for a nonlinear subscale turbulence term in pod models as exemplified for a high-Reynolds-number flow over an ahmed body. *J. Fluid Mech.* **747**, 518–544 (2014).
41. Rempfer, D. & Fasel, H. The dynamics of coherent structures in a flat-plate boundary layer. In *Advances in Turbulence IV*, 73–77 (Springer, 1993).
42. Ahmed, S. E., Bhar, K., San, O. & Rasheed, A. Forward sensitivity approach for estimating eddy viscosity closures in nonlinear model reduction. *Phys. Rev. E* **102**, 043302 (2020).
43. Schulman, J., Wolski, F., Dhariwal, P., Radford, A. & Klimov, O. Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347 (2017).
44. Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction* (MIT press, 2018).

45. Sutton, R. S., McAllester, D. A., Singh, S. P. & Mansour, Y. Policy gradient methods for reinforcement learning with function approximation. In *Advances in Neural Information Processing Systems*, 1057–1063 (2000).
46. Germano, M., Piomelli, U., Moin, P. & Cabot, W. H. A dynamic subgrid-scale eddy viscosity model. *Phys. Fluids A* **3**, 1760–1765 (1991).

Acknowledgements

O.S. gratefully acknowledges the financial support of the National Science Foundation under Award Number DMS-2012255 and the Early Career Research Program (ECRP) support of the U.S. Department of Energy under Award Number DE-SC0019290.

Author contributions

O.S. conceived the proposed variational multiscale reinforcement learning closure modeling approach. O.S. and S.P conducted the numerical experiments, analysed the data, and wrote the initial draft of the manuscript. A.R. provided feedback and contributed in shaping the research, analysis and manuscript. All coauthors reviewed the manuscript, discussed the results, and contributed to the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to O.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.