Computational and Applied Mathematics (2025) 44:308
https://doi.org/10.1007/s40314-025-03273-0

Residual-based data-driven variational multiscale reduced order models for parameter-dependent problems

Birgul Koc[1] · Samuele Rubino[2] · Tomás Chacón Rebollo[2] · Traian Iliescu[3]

Received: 5 January 2025 / Revised: 22 April 2025 / Accepted: 16 May 2025
© The Author(s) 2025

Abstract
In this paper, we propose a novel residual-based data-driven closure strategy for reduced order models (ROMs) of under-resolved, convection-dominated problems. The new ROM closure model is constructed in a variational multiscale (VMS) framework by using the available full order model data and a model form ansatz that depends on the ROM residual. We emphasize that this closure modeling strategy is fundamentally different from the current data-driven ROM closures, which generally depend on the ROM coefficients. We investigate the new residual-based data-driven VMS ROM closure strategy in the numerical simulation of three test problems: (i) a one-dimensional parameter-dependent advection-diffusion problem; (ii) a two-dimensional time-dependent advection-diffusion-reaction problem with a small diffusion coefficient ($\varepsilon = 1e{-}4$); and (iii) a two-dimensional flow past a cylinder at Reynolds number $Re = 1000$. Our numerical investigation shows that the new residual-based data-driven VMS-ROM is more accurate than the standard coefficient-based data-driven VMS-ROM.

Keywords Reduced order models · Variational multiscale · Data-driven modeling · Residual

Mathematics Subject Classification 65M60 · 76-04

1 Introduction

Reduced order models (ROMs) have been instrumental in significantly reducing the computational cost of full order models (FOMs) (e.g., finite element or finite volume methods)

✉ Birgul Koc
bkoc@us.es

Samuele Rubino
samuele@us.es

Tomás Chacón Rebollo
chacon@us.es

Traian Iliescu
iliescu@vt.edu

[1] Departamento EDAN, Universidad de Sevilla, Sevilla, Spain

[2] Departamento EDAN & IMUS, Universidad de Sevilla, Sevilla, Spain

[3] Department of Mathematics, Virginia Tech, Blacksburg, USA

Published online: 04 June 2025

in applications that require repeated model runs, such as, design and control, uncertainty quantification, inverse problems, and data assimilation. However, in convection-dominated problems (e.g., turbulent flows), standard ROMs generally yield inaccurate results. Indeed, to ensure a low computational cost, relatively low-dimensional ROMs are generally used in practice. Convection-dominated problems, however, usually require a large number of ROM basis functions in order to accurately represent the underlying dynamics. Thus, standard, low-dimensional ROMs usually yield spurious numerical oscillations that significantly degrade the solution accuracy. To alleviate this inaccurate behavior, the standard ROMs are generally equipped with (i) numerical stabilizations of different types (e.g., projection-based (Azaïez et al. 2021; Chacón Rebollo et al. 2022; Novo and Rubino 2021), subspace rotation (Balajewicz et al. 2016), variational multiscale (Bergmann et al. 2009; Iliescu and Wang 2013, 2014; Reyes and Codina 2020), streamline-upwind Petrov-Galerkin (Parish et al. 2020), filter-based (Girfoglio et al. 2021), and physically constraints (Sanderse 2020)); or (ii) ROM closures, which are terms that are added to the standard ROM to model the effect of the unresolved ROM scales. There are several types of ROM closure strategies, which are reviewed in Ahmed et al. (2021).

The variational multiscale (VMS) ROM closures are a popular class of ROM closure strategies. The VMS-ROM closures leverage the VMS framework (Hughes et al. 1998), which has been extensively used at the FOM level (see Ahmed et al. (2017) for a review). Specifically, the physical space is first decomposed in the ROM space (i.e., the space used to construct the ROM) and the space of unresolved scales. Then, the VMS-ROM equations are obtained by projecting the underlying equations onto the ROM space. We emphasize that the VMS-ROM includes both standard ROM operators (that depend on the ROM space), as well as a ROM closure term, i.e., a ROM operator that depends on the space of unresolved scales. To obtain a practical, self-contained ROM, this closure term in the VMS-ROM needs to be modeled by using exclusively the ROM space. This, in a nutshell, is the celebrated ROM closure problem.

Several successful VMS-ROM closure strategies have been proposed over the years. A review of these approaches is performed in Section IV.A.5 of Ahmed et al. (2021). Next, we outline several VMS-ROM closure models. In Bergmann et al. (2009), a residual-based VMS model was proposed as a ROM stabilization strategy. In Reyes and Codina (2020), a two-scale VMS-ROM equipped with time-dependent orthogonal sub-grid scales was developed. In Iliescu and Wang (2013), the authors proposed a VMS-ROM closure that includes an artificial viscosity added only to the small resolved scales of the gradient. The numerical tests in Iliescu and Wang (2013) showed the increased numerical stability and accuracy of the VMS-ROM over the standard G-ROM and illustrated the theoretical convergence rates. In particular, a problem displaying shock-like phenomena was considered (a 2D traveling wave) at a moderate Péclet number ($\nu = 10^{-4}$). In Iliescu and Wang (2014), the VMS-ROM was extended and studied for the incompressible Navier-Stokes equations. Recent VMS-ROM developments can be found in, e.g., Eroglu et al. (2017); Parish and Duraisamy (2017); Roop (2013); Stabile et al. (2019); Tello et al. (2019).

A paradigm shift in the development of VMS-ROM closures occurred in Mou et al. (2021) (see also Xie et al. (2018) for relevant work), where the classical, physical modeling used to develop VMS-ROM closures was replaced with data-driven modeling. Instead of using traditional arguments (e.g., eddy viscosity), the novel data-driven VMS-ROM (d2-VMS-ROM) proposed in Mou et al. (2021) was constructed by leveraging available data. Specifically, the d2-VMS-ROM was built by first postulating a model form for the closure term (i.e., a linear or quadratic model), and then solving a least squares problem to determine the model parameters that yielded the closest fit between the model form and the available data. The

original least squares formulation in the d2-VMS-ROM was replaced with a machine learning strategy in Ahmed et al. (2023) (see Xie et al. (2020) for relevant work). The d2-VMS-ROM has been extended in different directions (e.g., adding physical constraints (Mohebujjaman et al. 2019), providing mathematical support (Koc et al. 2022), and developing a stochastic framework for efficient data assimilation (Mou et al. 2023)), and has been successfully used in challenging numerical simulations (e.g., from the quasigeostrophic equations (Mou et al. 2020) to the turbulent channel flow (Mou 2021)).

Despite its significant achievements, the d2-VMS-ROM has been exclusively used in its coefficient-based form. Specifically, the data-driven closure term in the d2-VMS-ROM depends exclusively on the ROM coefficients. In this paper, we propose a fundamentally different d2-VMS-ROM strategy, in which the ROM closure term is a function of the ROM residual. The main advantage of the new residual-based ROM closure term is that it is *consistent*: As the ROM dimension increases, the residual decreases, and thus the ROM closure term decreases as well. This behavior is consistent with the physical role of the ROM closure term: As the ROM dimension increases, more physical scales are resolved by the d2-VMS-ROM, and thus the role of the ROM closure term decreases. We emphasize that, in contrast to the new residual-based d2-VMS-ROM, the current coefficient-based d2-VMS-ROM are not consistent (i.e., as the ROM dimension increases, the coefficient-based ROM closures do not necessarily decrease). In this paper, we perform a numerical investigation of the new residual-based d2-VMS-ROM and show that it is more accurate and efficient than the classical coefficient-based d2-VMS-ROM.

The outline of the paper is as follows: In Sect. 2, we briefly outline the standard Galerkin ROM for the incompresible Navier-Stokes equations. In Sect. 3, we describe the small-scale to large-scale decomposition that underpins the VMS-ROM framework. In Sect. 4, we introduce two strategies for the construction of the novel residual-based d2-VMS-ROM, and outline the standard coefficient-based d2-VMS-ROM.

In Sect. 5, we perform a numerical investigation of the new residual-based d2-VMS-ROMs in the simulation of three test problems: (i) a one-dimensional parameter-dependent advection-diffusion equation; (ii) a two-dimensional time-dependent advection-diffusion-reaction with small viscosity; and (iii) a two-dimensional flow past a circular cylinder at Reynolds number $Re = 1000$. To assess the performance of the new residual-based d2-VMS-ROMs, we compare them with the standard Galerkin ROM, the classical coefficient-based d2-VMS-ROM, and an ideal VMS-ROM, in which the closure term is computed from the FOM data. As benchmark for our numerical investigation, we use the FOM data. In Sect. 6, we conclude the paper with a short summary and future research directions.

2 Galerkin reduced order model (G-ROM)

This section provides a brief overview of the standard Galerkin ROM (G-ROM) strategy, which is one of the most common types of ROMs for fluid flows (Hesthaven et al. 2016; Noack et al. 2011; Quarteroni et al. 2015). As a mathematical model, we use the incompressible Navier-Stokes equations (NSE) (1)–(2):

$$\frac{\partial \boldsymbol{u}}{\partial t} - Re^{-1}\Delta \boldsymbol{u} + \boldsymbol{u}\cdot\nabla \boldsymbol{u} + \nabla p = \boldsymbol{f}\,, \tag{1}$$

$$\nabla\cdot\boldsymbol{u} = 0\,, \tag{2}$$

where $\boldsymbol{u}$ is the velocity, p the pressure, $\boldsymbol{f}$ the force, and Re the Reynolds number. For clarity of presentation, we use homogeneous Dirichlet boundary conditions and zero force, i.e., $\boldsymbol{f} = 0$.

In Algorithm 1, we outline the construction of G-ROM, which is carried out using the velocity field. In the ROM framework, we employ a divergence-free velocity basis (utilizing Scott-Vogelius elements in the finite element setting). Consequently, the pressure term is omitted in the G-ROM. For ROMs that include the pressure approximation, see, e.g., Ballarin et al. (2015); Bergmann et al. (2009); Chacón Rebollo et al. (2022); Hesthaven et al. (2016); Noack et al. (2005); Novo and Rubino (2021); Quarteroni et al. (2015); Reyes and Codina (2020); Stabile and Rozza (2018).

Algorithm 1 Galerkin ROM (G-ROM)

1: Use available FOM data to construct dominant modes by using the proper orthogonal decomposition (POD), $\{\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_L\}$, $L \ll d$ (where d is the dimension of the input dataset), which correspond to the largest relative kinetic energy content and represent the dominant spatial structures of the given test problem;

2: Construct a ROM velocity approximation:

$$\boldsymbol{u}_L = \sum_{j=1}^{L} (\boldsymbol{a}_L)_j \, \boldsymbol{\varphi}_j, \tag{3}$$

as a linear combination of ROM basis functions $\boldsymbol{\varphi}_j$ with ROM coefficients $(\boldsymbol{a}_L)_j$;

3: Replace $\boldsymbol{u}$ in the given test problem with the ROM solution $\boldsymbol{u}_L$ given in (3);

4: Use the Galerkin projection, which projects the system obtained in step 3 onto the ROM velocity space $\boldsymbol{X}^L$ spanned by $\{\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_L\}$.

By using Algorithm 1 for the NSE equations (1)–(2), we obtain the following G-ROM:

$$\frac{d\boldsymbol{a}_L}{dt} = \boldsymbol{A}_{LL}\, \boldsymbol{a}_L + \boldsymbol{a}_L{}^{\top}\, \boldsymbol{B}_{LLL}\, \boldsymbol{a}_L, \tag{4}$$

where $\boldsymbol{A}_{LL}$ is an $L \times L$ matrix with entries $\boldsymbol{A}_{ij} := -Re^{-1}(\nabla\boldsymbol{\varphi}_i, \nabla\boldsymbol{\varphi}_j)$ and $\boldsymbol{B}_{LLL}$ is an $L \times L \times L$ tensor with entries $\boldsymbol{B}_{ijk} := -(\boldsymbol{\varphi}_i, \boldsymbol{\varphi}_j \cdot \nabla\boldsymbol{\varphi}_k)$. The G-ROM (4) is an L-dimensional system of ODEs that can be used for time intervals and/or parameters different from those used in the training regime (i.e., in the construction of the G-ROM).

3 Variational multiscale reduced order model (VMS-ROM)

In this section, we construct the VMS-ROM framework, which will be used in the next sections to build the d2-VMS-ROMs.

First, we note that when all the available ROM modes are used to create a ROM solution, the ROM approximation becomes

$$\boldsymbol{u}_d = \sum_{j=1}^{d} (\boldsymbol{a}_d)_j \, \boldsymbol{\varphi}_j. \tag{5}$$

In this case, $\boldsymbol{u}_d$ is the most accurate ROM approximation of the FOM solution with the given data in the POD sense (i.e., from the energetic point of view (Volkwein 2013)).

For laminar flows, using a few ($L \ll d$) ROM basis functions is enough to capture the main dynamics of the given problem, i.e., we are in the *resolved regime*. In this regime, a

low-dimensional ROM solution $\boldsymbol{u}_L$, with small $L \ll d$, yields an accurate approximation of the FOM solution.

However, for turbulent flows, the low-dimensional ROM solution (3) with small $L \ll d$ is not an accurate approximation of the FOM solution, i.e., we are in the *under-resolved regime*. To increase the accuracy of the L-dimensional ROM solution (3), we generally have two options: (i) increase the G-ROM dimension, L, or (ii) add numerical stabilization or a low-dimensional closure term to the G-ROM. In this paper, we aim to increase the numerical accuracy without significantly increasing the computational cost. Thus, we choose the second option. Next, we explain what ROM closure modeling is (see Ahmed et al. (2021) for a review) and how it is performed in a VMS setting.

The orthogonality of the ROM basis functions (which is intrinsic to the POD framework) allows us to decompose the ROM space as follows: $\boldsymbol{X}^d = \boldsymbol{X}^L \oplus \boldsymbol{X}^S$, where $\boldsymbol{X}^d := \text{span}\{\boldsymbol{\varphi}_1, ..., \boldsymbol{\varphi}_d\}$, $\boldsymbol{X}^L := \text{span}\{\boldsymbol{\varphi}_1, ..., \boldsymbol{\varphi}_L\}$, and $\boldsymbol{X}^S := \text{span}\{\boldsymbol{\varphi}_{L+1}, ..., \boldsymbol{\varphi}_d\}$. By using this decomposition, we define the large-scale and sub-scale solutions of the most accurate (in the POD sense) ROM solution, $\boldsymbol{u}_d$, as follows:

$$\boldsymbol{u}_L := \sum_{j=1}^{L} (\boldsymbol{a}_L)_j \, \boldsymbol{\varphi}_j, \tag{6a}$$

$$\boldsymbol{u}_S := \sum_{j=L+1}^{d} (\boldsymbol{a}_S)_j \, \boldsymbol{\varphi}_j. \tag{6b}$$

Next, we note that the most accurate ROM approximation in (5), $\boldsymbol{u}_d$, solves the following d-dimensional weak form of NSE equations (1)–(2)

$$\mathcal{D}_t(\boldsymbol{v}_d, \boldsymbol{u}_d) + a(\boldsymbol{v}_d, \boldsymbol{u}_d) + b(\boldsymbol{v}_d, \boldsymbol{u}_d, \boldsymbol{u}_d) = 0 \quad \forall \, \boldsymbol{v}_d \in \boldsymbol{X}^d, \tag{7}$$

where the bilinear forms are $\mathcal{D}_t(\boldsymbol{v}_d, \boldsymbol{u}_d) := (\boldsymbol{v}_d, \partial_t \boldsymbol{u}_d)$ and $a(\boldsymbol{v}_d, \boldsymbol{u}_d) := Re^{-1}(\nabla \boldsymbol{v}_d, \nabla \boldsymbol{u}_d)$, and the nonlinear form is $b(\boldsymbol{v}_d, \boldsymbol{u}_d, \boldsymbol{u}_d) := (\boldsymbol{v}_d, \boldsymbol{u}_d \cdot \nabla \boldsymbol{u}_d)$. Furthermore, $(\cdot, \cdot)$ represents the inner product in $\mathcal{L}^2(\Omega)$. By using the VMS method and choosing $\boldsymbol{v}_L = \boldsymbol{\varphi}_L \in \boldsymbol{X}^L$ and $\boldsymbol{v}_S = \boldsymbol{\varphi}_S \in \boldsymbol{X}^S$ ($\boldsymbol{v}_d = \boldsymbol{v}_L + \boldsymbol{v}_S$), we can decompose (7) into two problems as follows:

$$\mathcal{D}_t(\boldsymbol{\varphi}_L, \boldsymbol{u}_L + \boldsymbol{u}_S) + a(\boldsymbol{\varphi}_L, \boldsymbol{u}_L + \boldsymbol{u}_S) + b(\boldsymbol{\varphi}_L, \boldsymbol{u}_L + \boldsymbol{u}_S, \boldsymbol{u}_L + \boldsymbol{u}_S) = 0 \tag{8a}$$

$$\mathcal{D}_t(\boldsymbol{\varphi}_S, \boldsymbol{u}_L + \boldsymbol{u}_S) + a(\boldsymbol{\varphi}_S, \boldsymbol{u}_L + \boldsymbol{u}_S) + b(\boldsymbol{\varphi}_S, \boldsymbol{u}_L + \boldsymbol{u}_S, \boldsymbol{u}_L + \boldsymbol{u}_S) = 0. \tag{8b}$$

The matrix-vector forms of (8a)–(8b) are as follows:

$$\frac{d\boldsymbol{a}_L}{dt} = \boldsymbol{A}_{LL}\,\boldsymbol{a}_L + \boldsymbol{A}_{LS}\,\boldsymbol{a}_S + \boldsymbol{a}_L^\top \boldsymbol{B}_{LLL}\boldsymbol{a}_L + \boldsymbol{a}_L^\top \boldsymbol{B}_{LLS}\boldsymbol{a}_S + \boldsymbol{a}_S^\top \boldsymbol{B}_{LSL}\boldsymbol{a}_L + \boldsymbol{a}_S^\top \boldsymbol{B}_{LSS}\boldsymbol{a}_S, \tag{9a}$$

$$\frac{d\boldsymbol{a}_S}{dt} = \boldsymbol{A}_{SS}\,\boldsymbol{a}_S + \boldsymbol{A}_{SL}\,\boldsymbol{a}_L + \boldsymbol{a}_L^\top \boldsymbol{B}_{SLL}\boldsymbol{a}_L + \boldsymbol{a}_L^\top \boldsymbol{B}_{SLS}\boldsymbol{a}_S + \boldsymbol{a}_S^\top \boldsymbol{B}_{SSL}\boldsymbol{a}_L + \boldsymbol{a}_S^\top \boldsymbol{B}_{SSS}\boldsymbol{a}_S. \tag{9b}$$

The VMS-ROM idea can be explained by using the matrix formulation in (9a)–(9b). First, we note that this matrix form was obtained by using the variational formulation in (8a)–(8b). Second, we note that the matrix formulation in (9a)–(9b) is a multiscale formulation since $\boldsymbol{a}_L$ and $\boldsymbol{a}_S$ correspond to the large and small scales in the system, respectively. Thus, the matrix formulation in (9a)–(9b) is truly a variational multiscale ROM formulation.

The VMS-ROM rationale is that, since $L \ll d$, $\boldsymbol{a}_L$ can be computed efficiently. In contrast, since $L \ll S$, we should try to avoid the expensive computation of $\boldsymbol{a}_S$. However, the challenge is that the equations for $\boldsymbol{a}_L$ and $\boldsymbol{a}_S$ in (9a)–(9b) are coupled.

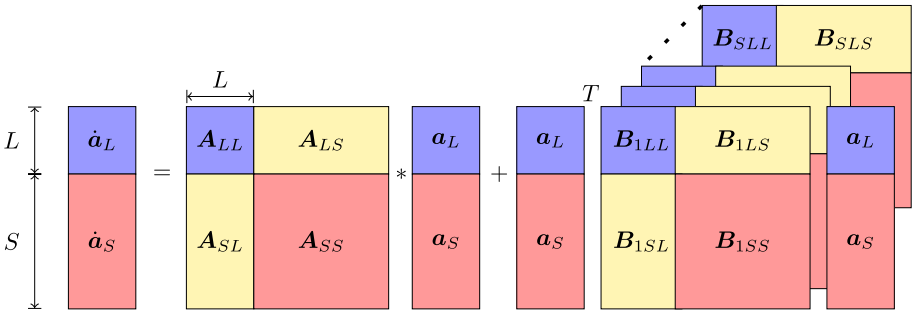

Fig. 1 Diagram of the algebraic form of the coupled system (9a)–(9b)

The VMS-ROM strategy centers on two simple ideas, which are illustrated in the schematic in Fig. 1: The first idea is that VMS-ROMs reduce the large system of equations (9a)–(9b) to a low-dimensional system of equations for $\boldsymbol{a}_L$. The second idea is that, to obtain an accurate approximation for $\boldsymbol{a}_L$, the effect of $\boldsymbol{a}_S$ needs to be modeled (i.e., the closure problem needs to be addressed). In the next section, we present two new data-driven strategies for modeling the effect of $\boldsymbol{a}_S$ (Sect. 4.1–4.2).

4 Data-driven variational multiscale ROM (d2-VMS-ROM)

In this section, we explain how we build the low-dimensional closure term in d2-VMS-ROM (10), which aims to increase the numerical accuracy by modeling the effect of the sub-scales:

$$\frac{d\boldsymbol{a}_L}{dt} = \boldsymbol{A}_{LL}\,\boldsymbol{a_L} + \boldsymbol{a_L}^{\top}\,\boldsymbol{B}_{LLL}\,\boldsymbol{a_L} + (\text{Closure-Term}). \tag{10}$$

In the following sections, we present two fundamentally different types of Closure-Term. In Sect. 4.1–4.2, we propose novel residual-based d2-VMS-ROMs in which the sub-scale information in the closure term is modeled by using the large-scale ROM residual, $Res_S(\boldsymbol{a}_L)$. Specfically, in Sect. 4.1 we propose R1-ROM which is associated with a single Closure-Term, and in Sect. 4.2 we propose R2-ROM which leverages two Closure-Terms to better control sub-scale effects. In Sect. 4.3, we describe the classical coefficient-based d2-VMS-ROM, which has recently been proposed (Mou et al. 2021). In this model, the sub-scale information in the Closure-Term is modeled by using the large-scale ROM coefficient vector, $\boldsymbol{a}_L$. Our goal in the numerical investigation in Sect. 5 is to show that the new residual-based d2-VMS-ROMs (R1-ROM and R2-ROM) are more accurate and efficient than the standard coefficient-based d2-VMS-ROM.

To build the d2-VMS-ROM (10), we first note that, to ensure that it is an efficient, L-dimensional model, its closure term should be modeled using only the large-scale ROM coefficient, $\boldsymbol{a}_L$. The next step in the d2-VMS-ROM construction is to postulate a model form for the Closure-Term, denoted Ansatz($\boldsymbol{a}_L$). To find the Ansatz-Operators that yield the most accurate results, we use data-driven modeling. Specifically, in the offline phase, we solve a least squares problem that minimizes the difference between Ansatz($\boldsymbol{a}_L$) and the true sub-scale term, denoted Sub-Scale-Term($\boldsymbol{a}_L, \boldsymbol{a}_S$), evaluated with the FOM sub-scale and

large-scale coefficient vectors:

$$\min_{\text{Ansatz-Operators}} \sum_{k=1}^{M} \left\| \text{Sub-Scale-Term}(\boldsymbol{a}_L^k, \boldsymbol{a}_S^k) \; - \; \text{Ansatz}(\boldsymbol{a}_L^k) \right\|_{\mathcal{L}^2}^2, \tag{11}$$

where M represents the number of snapshots. Then, in the online phase, the d2-VMS-ROM with the Ansatz computed from (11) is used for time intervals and/or parameters different from those used in the training stage.

We note that alternative data-driven strategies for ROM closure have been used in, e.g., Maulik et al. (2020); Prakash and Zhang (2024), and reviewed in Sanderse et al. (2024).

4.1 Residual-based d2-VMS-ROM with one ansatz (R1-ROM)

In this section, we introduce the first residual-based d2-VMS-ROM, R1-ROM.

In Algorithm 2, we outline the construction of the R1-Closure-Term.

Algorithm 2 Residual-Based Closure Term with One Ansatz (R1-Closure-Term)

1: Introduce the Sub-Scale-Term, which needs to be modeled, and the form of Ansatz, which appears in the Closure-Term of R1-ROM (14):

$$\begin{aligned} \text{Sub-Scale-Term} &= \boldsymbol{a}_S \\ &\approx \text{Ansatz} = \widetilde{\boldsymbol{A}}\, \text{Res}_S(\boldsymbol{a}_L) + \text{Res}_S(\boldsymbol{a}_L)^\top\, \widetilde{\boldsymbol{B}}\, \text{Res}_S(\boldsymbol{a}_L), \end{aligned} \tag{12}$$

where the large-scale ROM residual, derived from (9b), has the form $\text{Res}_S(\boldsymbol{a}_L) := \boldsymbol{A}_{SL}\, \boldsymbol{a}_L + \boldsymbol{a}_L^\top\, \boldsymbol{B}_{SLL}\, \boldsymbol{a}_L$.

2: Solve the least squares problem (11) to obtain the S-dimensional (the dimension of the sub-scale) Ansatz-Operators $\widetilde{\boldsymbol{A}}$ and $\widetilde{\boldsymbol{B}}$, using the Sub-Scale-Term and Ansatz defined in (12).

3: Substitute all $\boldsymbol{a}_S$ terms in the large-scale equation (9a) with Ansatz from (12) to derive the R1-Closure-Term (13):

$$\begin{aligned} \text{R1-Closure-Term} \approx\; & \boldsymbol{A}_{LS}\, \text{Ansatz} + \boldsymbol{a}_L^\top \boldsymbol{B}_{LLS}\, \text{Ansatz} \\ & + \text{Ansatz}^\top \boldsymbol{B}_{LSL} \boldsymbol{a}_L + \text{Ansatz}^\top \boldsymbol{B}_{LSS}\, \text{Ansatz}. \end{aligned} \tag{13}$$

Replacing the Closure-Term in the d2-VMS-ROM (10) with the R1-Closure-Term (13) from Algorithm 2, we derive the following R1-ROM formulation:

$$\frac{d\boldsymbol{a}_L}{dt} = \boldsymbol{A}_{LL}\, \boldsymbol{a}_L + \boldsymbol{a}_L^\top\, \boldsymbol{B}_{LLL}\, \boldsymbol{a}_L + (\text{R1-Closure-Term}). \tag{14}$$

We highlight that the large-scale ROM residual, $\text{Res}_S(\boldsymbol{a}_L)$, in (12) does not include a time-derivative term due to the orthogonality of the POD modes in the $\mathcal{L}^2$ inner product.

The idea behind the R1-ROM strategy is to avoid solving the expensive, high-dimensional equation (9b). Instead, we only leverage the information in (9b) (i.e., the fact that $\boldsymbol{a}_S$ depends on the residual $Res_S(\boldsymbol{a}_L)$) to model the sub-scales in (9a).

We emphasize that the closure term in the residual-based R1-ROM (14) is consistent since it depends on the large-scale ROM residual, $Res_S(\boldsymbol{a}_L)$. In contrast, the closure term of the standard coefficient-based C-ROM (which is presented in Sect. 4.3) is not consistent since it depends on the large-scale trajectory, $\boldsymbol{a}_L$.

4.2 Residual-based d2-VMS-ROM with two ansatzes (R2-ROM)

In this section, we introduce the second residual-based d2-VMS-ROM, R2-ROM. The main differences between the R1-ROM and R2-ROM are the following: (i) the R2-ROM uses two ansatzes whereas the R1-ROM uses only one ansatz, which is a common ansatz, and (ii) because the R2-ROM uses more ansatzes, it has more information related to the sub-scales.

In Algorithm 3, we outline the construction of R2-ROM.

Algorithm 3 Residual-Based Closure Term with Two Ansatzes (R2-Closure-Term)

1: Introduce a first sub-scale term Sub-Scale-Term1, which needs to be modeled, and the form of Ansatz1. The Sub-Scale-Term1 and its corresponding Ansatz1 are modeled similarly to R1-ROM (see (12)) to approximate the sub-scale ROM coefficient, i.e., $\boldsymbol{a}_S$ in (9a):

$$\begin{aligned} \text{Sub-Scale-Term1} &= \boldsymbol{a}_S \\ &\approx \text{Ansatz1} = \widetilde{\boldsymbol{A}}_1 \,\text{Res}_S(\boldsymbol{a}_L) + \text{Res}_S(\boldsymbol{a}_L)^\top \widetilde{\boldsymbol{B}}_1 \,\text{Res}_S(\boldsymbol{a}_L). \end{aligned} \tag{15}$$

2: Solve the least squares problem (11) to compute the first set of Ansatz operators, $\widetilde{\boldsymbol{A}}_1$ and $\widetilde{\boldsymbol{B}}_1$, which are of dimension S (the sub-scale dimension), utilizing the Sub-Scale-Term1 and Ansatz1 as defined in (15);

3: Replace all instances of $\boldsymbol{a}_S$ in the large-scale equation (9a) with Ansatz1 from (15);

4: Introduce a second sub-scale term (Sub-Scale-Term2) and the corresponding Ansatz2 to enhance numerical accuracy:

$$\begin{aligned} \text{Sub-Scale-Term2} &:= \text{Res}_L(\text{Ansatz1}) = \boldsymbol{A}_{LS}\,(\text{Ansatz1}) + (\text{Ansatz1})^\top \boldsymbol{B}_{LSS}\,(\text{Ansatz1}) \\ &\approx \text{Ansatz2} = \widetilde{\boldsymbol{A}}_2 \,\text{Res}_L(\boldsymbol{a}_S) + (\text{Res}_L(\boldsymbol{a}_S))^\top \widetilde{\boldsymbol{B}}_2 \,\text{Res}_L(\boldsymbol{a}_S); \end{aligned} \tag{16}$$

5: Solve the least squares problem (11) to obtain the second set of Ansatz operators, $\widetilde{\boldsymbol{A}}_2$ and $\widetilde{\boldsymbol{B}}_2$, of dimension L (the large-scale dimension), based on Sub-Scale-Term2 and Ansatz2, as defined in (16);

6: Replace the term $\text{Res}_L(\boldsymbol{a}_S)$ in the large-scale equation (9a) with Ansatz2 from (16), and substitute all remaining instances of $\boldsymbol{a}_S$ in the large-scale equation with Ansatz1 from (15), thereby deriving the R2-closure term:

$$\begin{aligned} \text{R2-Closure-Term} \approx\ & \widetilde{\boldsymbol{A}}_2 \,\text{Res}_L(\text{Ansatz1}) + (\text{Res}_L(\text{Ansatz1}))^\top \widetilde{\boldsymbol{B}}_2 \,\text{Res}_L(\text{Ansatz1}) \\ & + \boldsymbol{a}_L^\top \boldsymbol{B}_{LLS}\,\text{Ansatz1} + \text{Ansatz1}^\top \boldsymbol{B}_{LSL}\,\boldsymbol{a}_L. \end{aligned} \tag{17}$$

Replacing the Closure-Term in the d2-VMS-ROM (10) with the R2-Closure-Term (17) from Algorithm 3, we derive the following R2-ROM formulation:

$$\frac{d\boldsymbol{a}_L}{dt} = \boldsymbol{A}_{LL}\,\boldsymbol{a}_L + \boldsymbol{a}_L^\top \boldsymbol{B}_{LLL}\,\boldsymbol{a}_L + (\text{R2-Closure-Term}). \tag{18}$$

We emphasize that both the large-scale ROM residual, $\text{Res}_S(\boldsymbol{a}_L)$ in (15), and the sub-scale ROM residual, $\text{Res}_L(\boldsymbol{a}_S)$, in (16) do not contain a time-derivative term. This is due to the orthogonality of the POD modes in the $\mathcal{L}^2$ inner product. Furthermore, the inclusion of a second ansatz allows us to refine the approximation of $\text{Res}_L(\text{Ansatz1})$ using sub-scale FOM data.

The expectation is that the residual-based R2-ROM (18) could yield more accurate results than the residual-based R1-ROM (14) since the R2-Closure-Term (17) gradually models the sub-scale and has more sub-scale information than the R1-Closure-Term (13). However, in numerical simulations, since the R2-Closure-Term (17) is more complex, it could yield inaccurate results.

4.3 Coefficient-based d2-VMS-ROM (C-ROM)

In this section, we outline the coefficient-based C-ROM strategy (Mou et al. 2021).

In Algorithm 4, we outline the construction of C-Closure-Term.

Algorithm 4 Coefficient-Based Closure Term with One Ansatz (C-Closure-Term)

1: Introduce the sub-scale term Sub-Scale-Term, which needs to be modeled, and the form of Ansatz, which appears in the Closure-Term of C-ROM (21):

$$\begin{aligned} \text{Sub-Scale-Term} &= \boldsymbol{A}_{LS}\,\boldsymbol{a}_S + \boldsymbol{a}_L^\top\,\boldsymbol{B}_{LLS}\,\boldsymbol{a}_S + \boldsymbol{a}_S^\top\,\boldsymbol{B}_{LSL}\,\boldsymbol{a}_L + \boldsymbol{a}_S^\top\,\boldsymbol{B}_{LSS}\,\boldsymbol{a}_S \\ &\approx \text{Ansatz} = *\widetilde{\boldsymbol{A}}\,\boldsymbol{a}_L + \boldsymbol{a}_L^\top\,\widetilde{\boldsymbol{B}}\,\boldsymbol{a}_L, \end{aligned} \tag{19}$$

where Sub-Scale-Term is derived from (9a).

2: Solve the least squares problem (11) to compute the Ansatz-Operators $\widetilde{\boldsymbol{A}}$ and $\widetilde{\boldsymbol{B}}$, which are of dimension L (the large-scale dimension), utilizing the Sub-Scale-Term and Ansatz defined in (19).

3: Substitute Sub-Scale-Term in the large-scale equation (9a) with Ansatz from (19) to derive the C-Closure-Term (20):

$$\text{C-Closure-Term} \approx \widetilde{\boldsymbol{A}}\,\boldsymbol{a}_L + \boldsymbol{a}_L^\top\,\widetilde{\boldsymbol{B}}\,\boldsymbol{a}_L. \tag{20}$$

Replacing the Closure-Term in the d2-VMS-ROM (10) with the C-Closure-Term (20) from Algorithm 4, we derive the following C-ROM formulation:

$$\frac{d\boldsymbol{a}_L}{dt} = \boldsymbol{A}_{LL}\,\boldsymbol{a}_L + \boldsymbol{a}_L^\top\,\boldsymbol{B}_{LLL}\,\boldsymbol{a}_L + (\text{C-Closure-Term}). \tag{21}$$

4.4 Ideal variational multiscale ROM (I-ROM)

In this section, we outline the ideal ROM (I-ROM), which is used as a benchmark model to discuss the effect of Closure-Term in (10). We emphasize that I-ROM is a purely theoretical model, used to assess the accuracy of ROM closure models. Specifically, the closure term in I-ROM is computed directly from both the large-scale and sub-scale FOM data:

$$\text{I-Closure-Term} = \boldsymbol{A}_{LS}\,\boldsymbol{a}_S + \boldsymbol{a}_L^\top\,\boldsymbol{B}_{LLS}\,\boldsymbol{a}_S + \boldsymbol{a}_S^\top\,\boldsymbol{B}_{LSL}\,\boldsymbol{a}_L + \boldsymbol{a}_S^\top\,\boldsymbol{B}_{LSS}\,\boldsymbol{a}_S. \tag{22}$$

Thus, the I-ROM closure term cannot be used in practical settings, where FOM data is not available. Replacing the Closure-Term in the d2-VMS-ROM (10) with the I-Closure-Term (22), we obtain the I-ROM:

$$\frac{\boldsymbol{a}_L}{dt} = \boldsymbol{A}_{LL}\,\boldsymbol{a}_L + \boldsymbol{a}_L^\top\,\boldsymbol{B}_{LLL}\,\boldsymbol{a}_L + (\text{I-Closure-Term}). \tag{23}$$

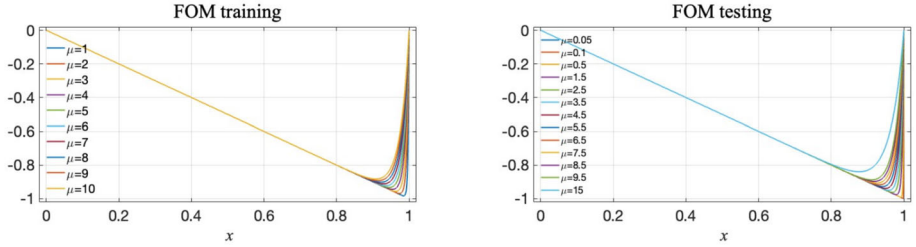

Fig. 2 Training and testing FOM solutions of (24) with different values of the parameter μ

5 Numerical results

In this section, we perform a numerical investigation of the new residual-based d2-VMS-ROMs, i.e., R1-ROM and R2-ROM, the classical d2-VMS-ROM, i.e., C-ROM, and the standard G-ROM. Specifically, we investigate whether the R1-ROM and R2-ROM are more accurate than the C-ROM. As a benchmark for our numerical comparison, we use the I-ROM. We also consider three test cases: (i) a 1D parameter-dependent advection-diffusion equation in Sect. 5.1; (ii) a 2D time-dependent advection-diffusion-reaction equation with a traveling wave solution and viscosity 1e−4 in Sect. 5.2; and (iii) a 2D flow past a circular cylinder with Reynolds number $Re = 1000$ in Sect. 5.3.

5.1 One-dimensional parameter-dependent advection-diffusion

In this section, we consider the one-dimensional parameter-dependent advection-diffusion (AD) equation:

$$\begin{cases} -\mu\, u_{xx} + c\, u_x = f \quad \text{for } x \in [0, 1], \\ u(0) = u(1) = 0, \end{cases} \tag{24}$$

where f is the forcing term, c the advection, u the variable of interest, and μ the diffusion coefficient.

We present numerical results for the parameter-dependent AD problem (24) with the following exact solution, which was used in Chácon Rebollo and Dia (2015):

$$u(x, \mu) = \frac{\exp(cx/\mu) - 1}{\exp(c/\mu) - 1} - x. \tag{25}$$

The forcing term is $f = -c = -400$.

Snapshot Generation

We generate FOM results, $\boldsymbol{u}_h$, using a linear finite element (FE) spatial discretization with mesh size $h = 1/4096$ and three sets of diffusion coefficients: (i) $\mu^{training} \in [1, 10]$ with $\Delta\mu = 1$, (ii) $\mu^{testing}_{interpolation} = \{1.5, 2.5, 3.5, 4.5, 5.5, 6.5, 7.5, 8.5, 9.5\}$, and (iii) $\mu^{testing}_{extrapolation} = \{0.05, 0.1, 0.5, 15\}$. We label the first FOM set of results as *training FOM data*, and the second and third FOM sets of results as *testing FOM data*.

In Fig. 2, we plot the training and testing FOM data. One can observe that when the viscosity is decreasing, the bump moves to the right boundary, which makes the problem harder.

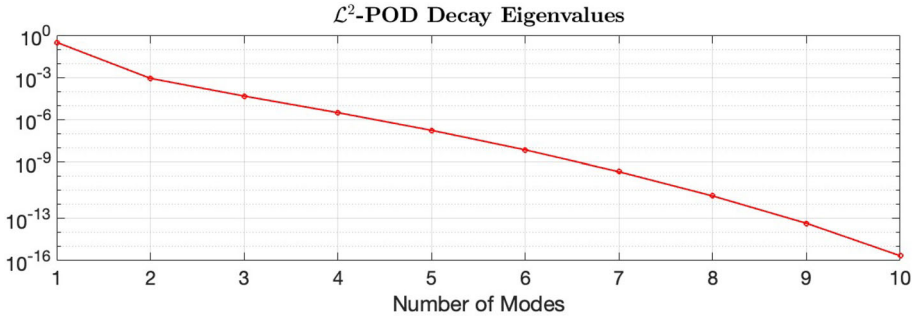

Fig. 3 AD equation; $\mathcal{L}^2$-POD decay of eigenvalues

ROM Construction and Testing

We generate the ROM basis functions and operators by using the training FOM data corresponding to the values $\mu^{training}$. To build the ROM basis, we use the POD (Holmes et al. 1996; Volkwein 2013). In Fig. 3, we plot the decay of eigenvalues of the ROM basis. To train the Ansatz Operators in (11), which are used to construct the Closure-Term in (10), we use the training FOM data.

To quantify the quality of the Ansatz Operators of the d2-VMS-ROMs in (11), we compute the average $\mathcal{L}^2$ FOM consistency error (26):

$$\mathcal{E}_{cons} := \frac{1}{M}\sum_{k=1}^{M}\left\| \text{Sub-Scale-Term}(\boldsymbol{a}_L^{\text{FOM},k}, \boldsymbol{a}_S^{\text{FOM},k}) - \text{Ansatz}(\boldsymbol{a}_L^{\text{FOM},k})\right\|_{\mathcal{L}^2}. \tag{26}$$

In Sect. 5.1.1, we test all the ROMs on testing data that is inside the training range but are different from the training parameters, i.e., for $\mu^{testing}_{interpolation} = \{1.5, ..., 9.5\}$. Likewise, in Sect. 5.1.2, we test all the ROMs on testing data that is outside the training range, i.e., for $\mu^{testing}_{extrapolation} = \{0.05, 0.1, 0.5, 15\}$. Thus, we are testing all the ROMs in the *predictive regime*.

To compare the numerical accuracy of the d2-VMS-ROMs, I-ROM, and G-ROM defined in Sect. 4, we use the average relative $\mathcal{L}^2$ projection error:

$$\mathcal{E}_{\mathcal{L}^2 proj} = \frac{1}{M}\sum_{k=1}^{M}\frac{\left\| \boldsymbol{u}_L(\mu_k^{testing}) - \sum_{i=1}^{L}\left(\boldsymbol{u}_h(\mu_k^{testing}), \boldsymbol{\varphi}_i\right)\boldsymbol{\varphi}_i\right\|_{\mathcal{L}^2}}{\left\| \sum_{i=1}^{L}\left(\boldsymbol{u}_h(\mu_k^{testing}), \boldsymbol{\varphi}_i\right)\boldsymbol{\varphi}_i\right\|_{\mathcal{L}^2}}, \tag{27}$$

where the testing diffusion coefficients can be chosen from $\mu^{testing}_{interpolation}$, $\mu^{testing}_{extrapolation}$, or a combination of them , and M corresponds to the number of the total testing parameters.

In Table 1, we list the FOM consistency error (26) of C-ROM, R1-ROM, and R2-ROM to validate the quality of the Ansatz Operators of (11) using the FOM data corresponding to the training parameters. We observe that the FOM consistency errors of R1-ROM and R2-ROM (for the first ansatz) are almost the same since they are constructed in the same way, and gradually decrease as L increases. Furthermore, the FOM consistency errors of R1-ROM and R2-ROM (for the first ansatz) are orders of magnitude lower than the consistency error of C-ROM.

Table 1 AD equation; average $\mathcal{L}^2$ FOM consistency error (26) of C-ROM, R1-ROM, and R2-ROM for various L values (number of POD modes)

L	C-FOM-Consistency	R1-FOM-Consistency	R2-FOM-Consistency	
			Ansatz1	Ansatz2
1	1.76e+01	2.07e-03	2.07e-03	8.99e+00
2	3.30e+01	1.97e-04	1.97e-04	2.93e+00
3	9.28e+00	4.43e-05	4.43e-05	5.54e+00
4	2.32e+00	7.49e-06	7.49e-06	2.93e+00
5	5.67e-01	3.02e-06	3.01e-06	1.35e+00
6	1.15e-01	3.31e-07	3.31e-07	3.92e-01
7	1.56e-02	7.28e-08	7.28e-08	8.77e-02
8	1.31e-03	1.54e-08	1.54e-08	3.70e-06
9	5.73e-10	1.62e-09	1.62e-09	1.83e-03
10	0	0	0	0

Table 2 AD equation; interpolation testing parameters; average relative $\mathcal{L}^2$ projection error (27) for G-ROM, I-ROM, C-ROM, R1-ROM, and R2-ROM for various L values (number of POD modes)

L	G-ROM	I-ROM	C-ROM	R1-ROM	R2-ROM
1	5.34e-01	1.92e-05	1.06e-03	1.50e-04	1.50e-04
2	1.39e-01	2.41e-05	4.83e-04	5.45e-06	5.44e-06
3	3.96e-02	3.54e-05	1.27e-04	1.82e-06	1.83e-06
4	1.11e-02	4.66e-05	3.05e-05	6.90e-07	7.26e-07
5	3.21e-03	5.15e-05	8.26e-06	1.63e-06	1.67e-06
6	1.01e-03	5.76e-05	2.85e-06	6.57e-07	9.01e-07
7	4.01e-04	6.13e-05	1.28e-06	2.96e-07	3.44e-07

5.1.1 Testing parameters in interpolation regime

Table 2 lists the average relative $\mathcal{L}^2$ projection error (27) for G-ROM, I-ROM, C-ROM, R1-ROM, and R2-ROM for various L values (number of POD modes) for $\mu^{testing}_{interpolation}$ parameters.

In this setting, where interpolation testing parameters are selected, the closure terms in d2-VMS-ROMs have a beneficial impact on their accuracy.

We observe that the new residual-based models R1-ROM and R2-ROM yield much higher accuracy than the classical coefficient-based C-ROM and the G-ROM. For example, with $L = 2, 3, 4$, R1-ROM and R2-ROM are nearly two orders of magnitude more accurate than C-ROM, and approximately four orders of magnitude more accurate than G-ROM. Moreover, we observe that for higher values of L (i.e., the dimension of the large-scale space), R1-ROM is slightly more accurate than R2-ROM.

Furthermore, Table 1 suggests that the smoother decay rate of the FOM consistency error in R1-ROM and R2-ROM results in lower average relative R1-ROM and R2-ROM errors in Table 2.

In Fig. 4, we plot the relative $\mathcal{L}^2$ ROM errors for the testing parameter $\mu^{testing}_{interpolation}$. The plots confirm the conclusions observed in Table 2.

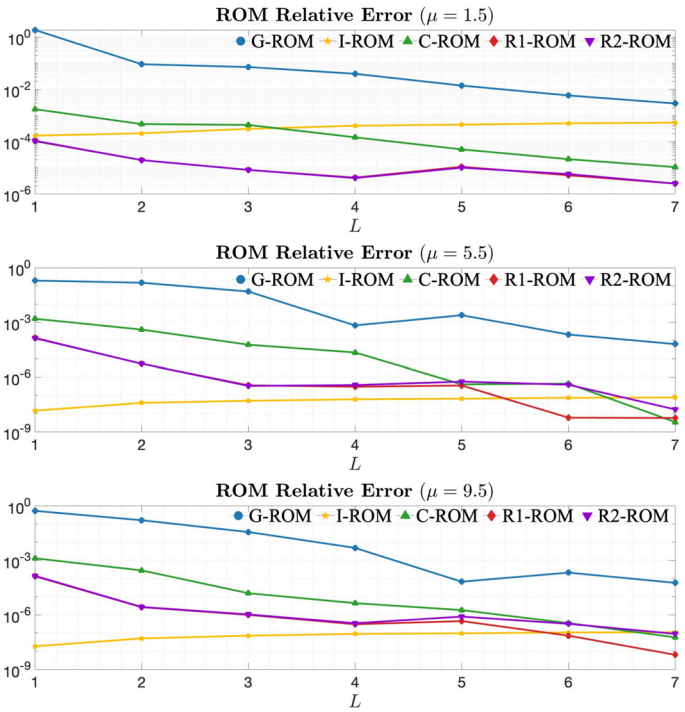

Fig. 4 AD equation; the relative $\mathcal{L}^2$ ROM error for various interpolation testing parameters for $L = 1, 2, 3, 4, 5, 6, 7$

Table 3 AD equation; extrapolation testing parameters; average relative $\mathcal{L}^2$ projection error (27) for G-ROM, I-ROM, C-ROM, R1-ROM, and R2-ROM for various L values (number of POD modes)

L	G-ROM	I-ROM	C-ROM	R1-ROM	R2-ROM
1	3.40e+01	2.34e+00	7.16e-03	1.98e-03	1.98e-03
2	7.33e-01	1.00e-01	1.01e-02	5.49e-04	5.50e-04
3	5.74e+00	1.99e+00	8.81e-03	5.42e-04	5.45e-04
4	5.36e-01	2.20e-01	8.13e-03	5.85e-04	6.06e-04
5	3.35e+00	1.98e+00	7.69e-03	1.77e-03	1.94e-03
6	4.59e-01	2.99e-01	7.24e-03	1.44e-03	1.32e-03
7	2.59e+00	2.01e+00	6.85e-03	1.13e-03	1.09e-03

5.1.2 Testing parameters in extrapolation regime

Table 3 lists the average relative $\mathcal{L}^2$ projection error (27) for G-ROM, I-ROM, C-ROM, R1-ROM, and R2-ROM for various L values (number of POD modes) for testing parameters $\mu_{extrapolation}^{testing}$.

We observe that the new residual-based models R1-ROM and R2-ROM yield much higher accuracy than the classical coefficient-based C-ROM, the I-ROM, and the G-ROM. For example, for $L = 2, 3, 4$, R1-ROM and R2-ROM are more than an order of magnitude more accurate than C-ROM, and three orders of magnitude more accurate than I-ROM and

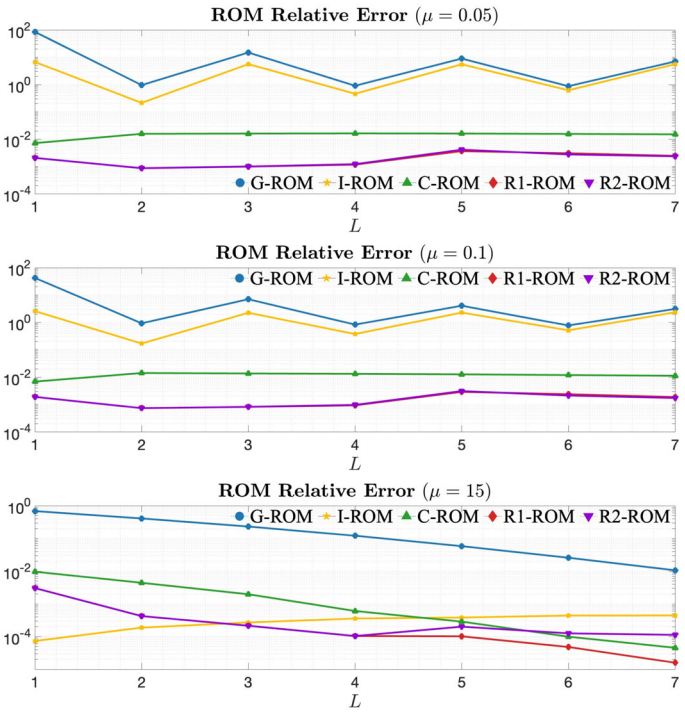

Fig. 5 AD equation; the relative $\mathcal{L}^2$ ROM error for various extrapolation testing parameters for $L = 1, 2, 3, 4, 5, 6, 7$

G-ROM. Furthermore, we observe that for higher values of L (i.e., the dimension of the large-scale space), the R1-ROM is more accurate than the R2-ROM.

We note that it is not surprising that I-ROM yields worse results than d2-VMS-ROMs since the I-Closure-Term (22) in (23) is computed using the testing FOM data (which have different bump locations and are not related to the training FOM data, which are used to build the ROM basis, ROM operators, and Ansatz Operators in Closure-Term of the d2-VMS-ROMs).

In Fig. 5, we plot the relative $\mathcal{L}^2$ ROM errors for testing parameters $\mu^{testing}_{extrapolation}$. The plots confirm the conclusions observed in Table 3.

5.2 Two-dimensional time-dependent advection-diffusion-reaction

In this section, we consider the two-dimensional time-dependent advection-diffusion-reaction (ADR) equation:

$$\begin{cases} u_t - \varepsilon \Delta u + \boldsymbol{b} \cdot \nabla u + cu = f, & \text{in } (0, T] \times \Omega, \\ u = 0 \text{ on } [0, T] \times \partial\Omega, \\ u(0, \boldsymbol{x}) = u_0(\boldsymbol{x}) \quad \text{in } \Omega, \end{cases} \tag{28}$$

where ε is the diffusion coefficient, $\boldsymbol{b}$ the convection velocity field, c the coefficient of the reaction term, and u the variable of interest.

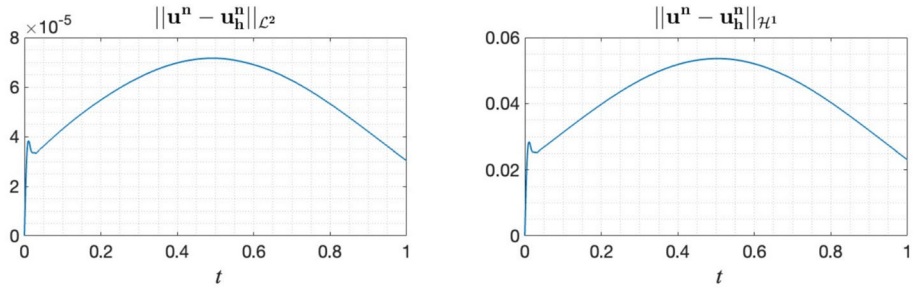

Fig. 6 ADR equation; the absolute $\mathcal{L}^2$ and $\mathcal{H}^1$ errors between the analytical solution and the FOM solution

We present numerical results for the time-dependent ADR problem (28) with the following exact solution, which is a traveling wave:

$$u(x, y, t) = 0.5 \sin(\pi x) \sin(\pi y)\Big(1 + \tanh\big(\frac{x + y - t - 0.5}{\nu}\big)\Big), \tag{29}$$

where ν is a parameter that is generally picked as the square root of ε.

Snapshot Generation

We use the following setting to obtain the FOM data: $\Omega = (0, 1)^2, [0, T] = [0, 1], \varepsilon = 1\text{e}{-4}$, $\boldsymbol{b} = [\cos(\pi/3), \sin(\pi/3)], c = 1$, and $\nu = 4\text{e}{-2}$. We generate FOM results using a quadratic FE spatial discretization with mesh size $h = 1/100$, and a Crank-Nicolson time discretization with time step size $\Delta t = 1\text{e}{-3}$. The viscosity parameter, ε, and the parameter ν that controls the steepness of the bump are the fixed parameters.

ROM Construction and Testing

We present the numerical results for the fixed parameters $\varepsilon = 1\text{e}{-4}$ and $\nu = 4\text{e}{-2}$, and the *predictive regime* in which the training and testing intervals are disjoint. To choose the training and testing intervals, in Fig. 6, we plot the $\mathcal{L}^2$ and $\mathcal{H}^1$ errors between the exact and FE solutions on the time interval [0, 1]. Based on these error plots, we choose the training time interval as [0, 0.6], and the extrapolation testing time interval as [0.6, 1]. We build the POD basis, the ROM operators, and the Ansatz Operators in (11) that are used to construct the Closure-Term in (10) in the training time interval, and test the ROMs in the extrapolation testing time interval. In Fig. 7, we plot the decay of eigenvalues of the ROM basis.

To compare the numerical accuracy of the d2-VMS-ROMs, I-ROM, and G-ROM defined in Sect. 4, we use the average relative $\mathcal{L}^2$ ROM projection error:

$$\mathcal{E}_{\mathcal{L}^2 proj} = \frac{1}{M} \sum_{k=1}^{M} \frac{\Big\| \boldsymbol{u}_L(t_k) - \sum_{i=1}^{L} \Big(\boldsymbol{u}_h(t_k), \boldsymbol{\varphi}_i\Big) \boldsymbol{\varphi}_i \Big\|_{\mathcal{L}^2}}{\Big\| \sum_{i=1}^{L} \Big(\boldsymbol{u}_h(t_k), \boldsymbol{\varphi}_i\Big) \boldsymbol{\varphi}_i \Big\|_{\mathcal{L}^2}}. \tag{30}$$

In Table 4, we list the average relative $\mathcal{L}^2$ ROM projection errors (30) for different ROM dimensions, L, in the extrapolation predictive regime (i.e., where the testing interval extends beyond the training interval), making the problem more challenging. Our results indicate that, when a sufficient number of POD modes are included, R2-ROM achieves the highest accuracy among the d2-VMS-ROMs, with the exception of the case $L = 2$, when R1-ROM is more accurate.

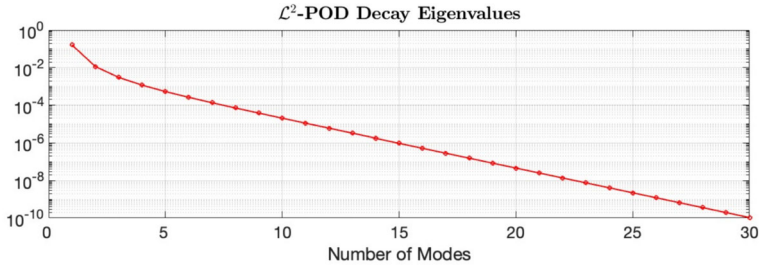

Fig. 7 ADR equation; $\mathcal{L}^2$-POD decay of eigenvalues

Table 4 ADR equation; average relative $\mathcal{L}^2$-projection error (30) for G-ROM, I-ROM, C-ROM, R1-ROM, and R2-ROM for various L values (number of POD modes)

L	G-ROM	I-ROM	C-ROM	R1-ROM	R2-ROM
2	2.14e+00	3.87e-01	1.75e+00	1.60e+00	1.81e+00
4	2.19e+00	4.01e-01	2.00e+00	1.73e+00	1.10e+00
6	2.13e+00	4.59e-01	2.02e+00	1.68e+00	5.25e-01
8	2.06e+00	6.02e-01	1.99e+00	1.66e+00	5.38e-01
10	1.99e+00	9.25e-01	1.95e+00	1.76e+00	5.17e-01

In Fig. 8, we plot the relative ROM error in time for $L = 4, 6$, and 10. We observe that the relative R2-ROM error recovers the I-ROM as L increases. Furthermore, for $L = 10$, the relative R2-ROM error is lower than the relative I-ROM error.

Finally, in Fig. 9, we plot the FOM and all ROM solutions at the final time, $T = 1$, for $L = 8$. We observe that the C-ROM does not diminish the oscillatory bump, whereas the R1-ROM tries to decrease the magnitude of the bump. The R2-ROM decreases the oscillations the most, but it also produces another oscillatory bump.

5.3 Two-dimensional flow past a cylinder

In this section, we consider a 2D channel flow past a circular cylinder at Reynolds number $Re = 1000$. As criteria in our numerical investigation, we use the average $\mathcal{L}^2$ ROM errors, kinetic energy, vortex shedding frequency, and Pareto plots.

Computational Setting

As a mathematical model, we use the NSE (1)–(2). The computational domain is a 2.2×0.41 rectangular channel with a radius $= 0.05$ cylinder, centered at $(0.2, 0.2)$; see Fig. 10.

We prescribe no-slip boundary conditions on the walls and cylinder, and the following inflow and outflow profiles (John 2004; Mohebujjaman et al. 2019, 2017; Rebholz and Xiao 2017):

$$u_1(0, y, t) = u_1(2.2, y, t) = \frac{6}{0.41^2} y(0.41 - y), \tag{31}$$

$$u_2(0, y, t) = u_2(2.2, y, t) = 0, \tag{32}$$

where $\boldsymbol{u} = \langle u_1, u_2 \rangle$. There is no forcing and the flow starts from rest.

Snapshot Generation

For the spatial discretization, we use the pointwise divergence-free, LBB stable (P_2, P_1^{disc}) Scott-Vogelius FE pair on a barycenter refined regular triangular mesh of the barycenter (John

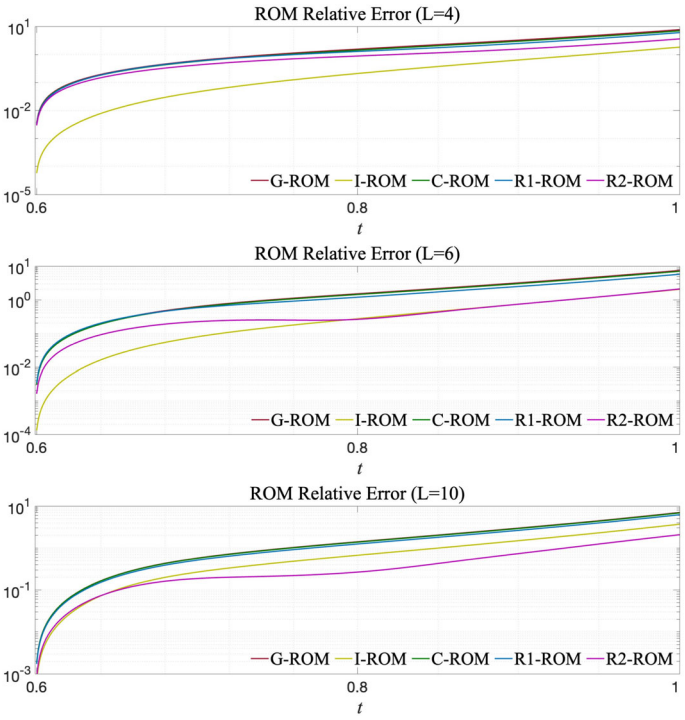

Fig. 8 ADR equation; relative ROM error over the extrapolation testing time interval for various L values

et al. 2016). The mesh provides $103K$ (102962) velocity and $76K$ (76725) pressure degrees of freedom. We utilize the commonly used linearized BDF2 temporal discretization and a time step size $\Delta t = 0.002$ for both the FOM and the ROM time discretizations. In the first time step, we use a backward Euler scheme so that we have two initial time-step solutions, as required for the BDF2 scheme.

ROM Construction and Testing

The FOM simulation achieves the statistically steady state after $t = 13$ in the numerical investigation. To build the ROM basis functions and operators, we decided to use 3 time units of FOM data. Thus, we collect FOM snapshots from $t = 13$ to $t = 16$ and label them as training FOM data. In Fig. 11, we plot the decay of eigenvalues of the ROM basis.

To train the Ansatz Operators in (11), which are used to construct the Closure-Term in (10), we use the training FOM data. Due to the periodic behavior of the flow and in order to decrease the computational cost of constructing the Ansatz Operators, we use a half-period of training FOM data, i.e., 68 FOM snapshots, from $t = 13$ to $t = 13.134$. Then we test all ROMs in the extrapolation testing time interval, $t = 16$ to $t = 23$.

In Sect. 5.3.1–5.3.3, we compare the quality of all ROMs based on three criteria: $\mathcal{L}^2$-norm error, kinetic energy error, and vortex shedding frequency, respectively.

In Table 5, we list the average $\mathcal{L}^2$ FOM consistency error (26), computed using a reduced basis of $d = 22$ POD modes. This choice of d, which is substantially lower than the full rank of the system, is used consistently in Sect. 5.3.1 and 5.3.2. From the results in Table 5, we observe that both R1-ROM and R2-ROM (for both ansatzes) have lower FOM consistency errors than C-ROM. Furthermore, the FOM consistency errors of R1-ROM and R2-ROM

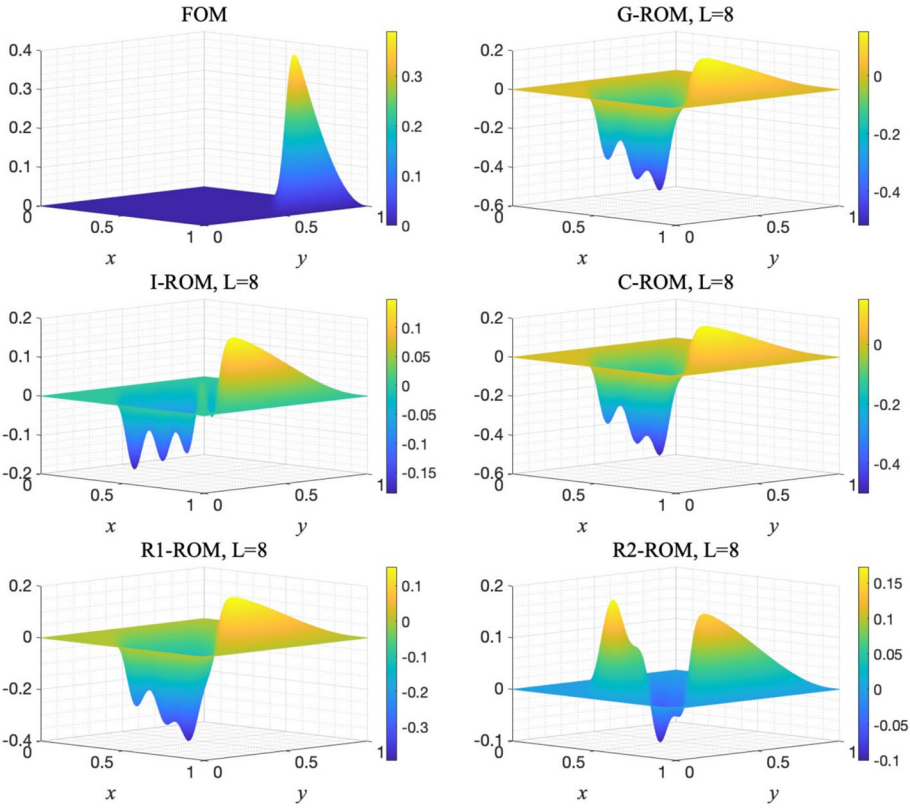

Fig. 9 ADR equation; ROM solutions at the final time, $T = 1$, for $L = 8$

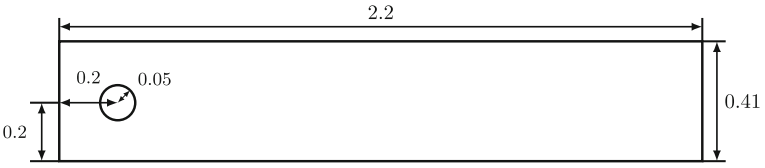

Fig. 10 Geometry of the flow past a circular cylinder numerical experiment

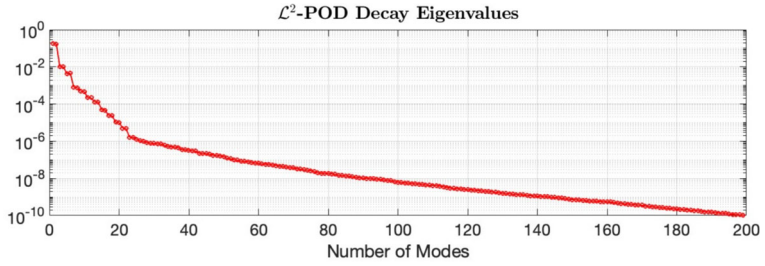

Fig. 11 Flow past a cylinder; $\mathcal{L}^2$-POD decay of eigenvalues

Table 5 Flow past a cylinder; $d = 22$; average $\mathcal{L}^2$ FOM consistency error (26) of C-ROM, R1-ROM, and R2-ROM for various L values (number of POD modes).

L	C-FOM-Consistency	R1-FOM-Consistency	R2-FOM-Consistency	
			Ansatz1	Ansatz2
2	2.28e-01	2.59e-02	2.05e-02	7.37e-02
5	1.49e+00	1.72e-02	2.62e-02	3.61e-03
8	4.01e-01	1.22e-04	6.60e-05	3.03e-05
11	1.35e+00	1.27e-08	1.54e-08	3.63e-09
14	2.46e-01	5.39e-03	4.24e-03	2.25e-05
17	7.29e-01	6.19e-03	6.19e-03	1.03e-04
20	1.43e-01	2.34e-03	2.52e-03	8.35e-04
22	0	0	0	0

Table 6 Flow past a cylinder; $d = 14$; average $\mathcal{L}^2$ FOM consistency error (26) of C-ROM, R1-ROM, and R2-ROM for various L values (number of POD modes).

L	C-FOM-Consistency	R1-FOM-Consistency	R2-FOM-Consistency	
			Ansatz1	Ansatz2
2	2.27e-01	2.29e-02	2.29e-02	7.19e-03
3	2.59e+00	6.39e-04	3.58e-03	1.41e-03
5	1.49e+00	1.63e-02	1.63e-02	1.82e-03
7	7.63e-02	4.28e-02	4.28e-02	6.37e-03
9	2.10e+00	1.50e-02	1.50e-02	4.84e-04
11	1.33e+00	1.77e-02	1.14e-02	6.61e-03
13	1.15e+00	7.90e-03	7.90e-03	2.28e-03
14	0	0	0	0

reach the lowest value at $L = 11$, and then increase again. Since the least squares problem (11) is sensitive based on using the size of the data, to investigate this behavior in Table 5, in Table 6, we also list the average $\mathcal{L}^2$ FOM consistency error (26), calculated using a reduced basis with $d = 14$ POD modes. In Table 6, we observe a similar behavior as that observed in Table 5 but less significant increasing, from $L = 5$.

For the flow past a cylinder flow, due to the complex nonlinear interactions in the Navier-Stokes equations, the consistency error has a more complex behavior than in the linear ADR case. Specifically, the consistency error decreases as the dimension L of the reduced space increases, up to some value $L = L_{opt}$ (depending on the total number of modes taken into account to build the ROM, d), and then it *approximately* stagnates: $L_{opt} = 11$ when $d = 22$, and $L_{opt} = 3$ when $d = 14$.

5.3.1 $\mathcal{L}^2$-norm error

In this section, we compare the numerical accuracy of the d2-VMS-ROMs, I-ROM, and G-ROM defined in Sect. 4 by using the average relative $\mathcal{L}^2$ ROM projection error (30).

Table 7 Flow past a cylinder; average relative $\mathcal{L}^2$-projection error (30) for G-ROM, I-ROM, C-ROM, R1-ROM, and R2-ROM for various L values (number of POD modes).

L	G-ROM	I-ROM	C-ROM	R1-ROM	R2-ROM
2	1.94e+00	1.33e-02	6.96e-01	6.07e-01	3.05e-01
3	1.54e+00	4.80e-03	9.20e-01	1.03e-01	1.13e-01
4	1.19e+00	4.61e-03	3.26e-01	1.95e-01	9.55e-02
5	1.19e+00	2.15e-03	9.51e-01	5.09e-01	4.03e-01
6	5.76e-01	3.09e-02	2.41e-01	7.10e-02	3.09e-02
7	4.91e-01	2.00e-02	4.56e-01	8.60e-02	3.51e-01
8	2.58e-01	4.23e-03	1.53e-01	4.12e-02	6.40e-02

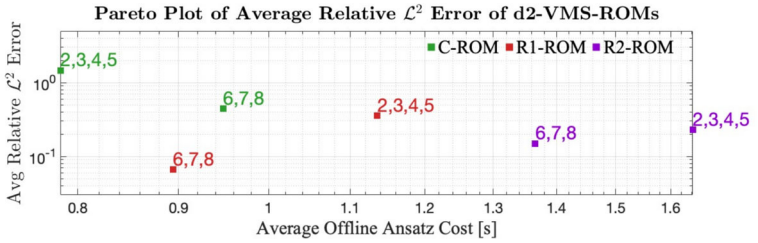

Fig. 12 Flow past a cylinder; Pareto plot of average relative $\mathcal{L}^2$ error of d2-VMS-ROMs, i.e., R1-ROM, R2-ROM, and C-ROM

In Table 7, we list the average relative $\mathcal{L}^2$ ROM errors (27) for different ROM dimensions, L. We observe that both R1-ROM and R2-ROM yield more accurate results than C-ROM and G-ROM. Overall, the R2-ROM is more accurate than the R1-ROM, except for L = 3, 7, 8.

In Fig. 12, we present a Pareto plot of the d2-VMS-ROMs, i.e., R1-ROM and R2-ROM, and C-ROM, averaging the $\mathcal{L}^2$ error and offline ansatz cost over the low ROM dimensions, i.e., L = 2, 3, 4, 5, and the high-ROM dimensions, i.e., L = 6, 7, 8. For the low-ROM dimension, we observe that the R2-ROM yields the most accurate model, although it is the most expensive model. For the high-ROM dimension L = 6, 7, 8, the R1-ROM is the most accurate and least expensive model.

5.3.2 Kinetic energy error

In this section, we compare the numerical accuracy of the d2-VMS-ROMs, I-ROM, and G-ROM defined in Sect. 4 by using the kinetic energy (KE) criterion:

$$E_{kin} := \frac{1}{2}\|\boldsymbol{u}\|^2_{\mathcal{L}^2} = \frac{1}{2}\int_\Omega |\boldsymbol{u}|^2 \, d\Omega. \tag{33}$$

The aim of our numerical investigation is to observe how the kinetic energy of ROMs evolves in the extrapolation time interval, e.g. whether it dissipates, blows up, or stagnates.

$$\mathcal{E}_{KE} = \frac{\sum_{k=1}^{M} | \, E^{FOM}_{kin}(t_k) \, - \, E^{ROM}_{kin}(t_k) \, |}{\sum_{k=1}^{M} | \, E^{FOM}_{kin}(t_k) \, |}. \tag{34}$$

In Table 8, we list the average relative kinetic energy error (34) of the ROMs. We observe that R2-ROM generally yields a lower kinetic energy error than R1-ROM, and a much lower kinetic energy error than C-ROM and G-ROM for low ROM dimensions. For high ROM

Table 8 Flow past a cylinder; average relative kinetic energy error (34) for G-ROM, I-ROM, C-ROM, R1-ROM, and R2-ROM for various L values (number of POD modes)

L	G-ROM	I-ROM	C-ROM	R1-ROM	R2-ROM
2	4.99e-01	5.91e-03	1.40e-02	1.16e-01	4.34e-03
3	1.62e-01	1.35e-03	1.25e-01	4.57e-03	2.48e-02
4	1.08e-01	2.14e-03	2.22e-02	5.79e-02	4.52e-03
5	8.35e-02	9.35e-04	2.40e-01	1.32e-01	9.41e-02
6	2.14e-01	1.94e-03	1.17e-01	2.74e-02	2.73e-03
7	2.06e-01	2.51e-03	2.08e-01	1.44e-02	1.14e-01
8	7.53e-02	1.73e-03	3.47e-02	8.57e-03	1.31e-02

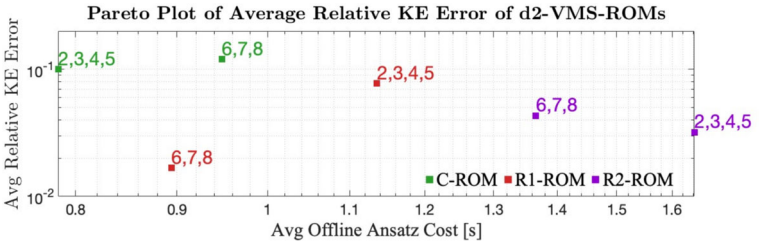

Fig. 13 Flow past a cylinder; Pareto plot of average relative KE error of d2-VMS-ROMs, i.e., R1-ROM, R2-ROM, and C-ROM

dimensions, R1-ROM yields a lower kinetic energy error than R2-ROM, and a much lower kinetic energy error than C-ROM and G-ROM.

In Fig. 13, we present the Pareto plot of the d2-VMS-ROMs, i.e., R1-ROM, R2-ROM, and C-ROM, averaging the average relative kinetic energy and offline ansatz cost over the low ROM dimensions, i.e., L $= 2, 3, 4, 5$, and the high ROM dimensions, i.e, L $= 6, 7, 8$. For the low L values, we observe that R2-ROM yields the most accurate results, although its computational cost is the highest. On the other hand, for the high L values, the R1-ROM is the most accurate and its computational cost is the lowest. The conclusion of the kinetic energy Pareto plot in Fig. 13 is consistent with the conclusion of the average $\mathcal{L}^2$ Pareto plot in Fig. 12.

In Fig. 14, we plot the kinetic energy (33) of the FOM projection, G-ROM, I-ROM, C-ROM, R1-ROM, and R2-ROM for the ROM dimension values L $= 2, 4, 6, 8$ over the extrapolation time interval. These plots support the results of Table 8.

5.3.3 Vortex shedding frequency matching between FOM and d2-VMS-ROMs

In this section, we compute the average vortex shedding frequency f_s of d2-VMS-ROMs and FOM based on the vortex shedding period $\mathcal{T}_s$, which is defined as follows:

$$\mathcal{T}_s = \frac{1}{N_s}\sum_{k=1}^{N_s}(t_s(k+1) - t_s(k)), \tag{35}$$

where $t_s(k)$ denotes the time instances corresponding to successive peaks in the kinetic energy within the extrapolation testing time interval [18, 23]. These peaks are used to estimate the dominant vortex shedding cycle of the respective models.

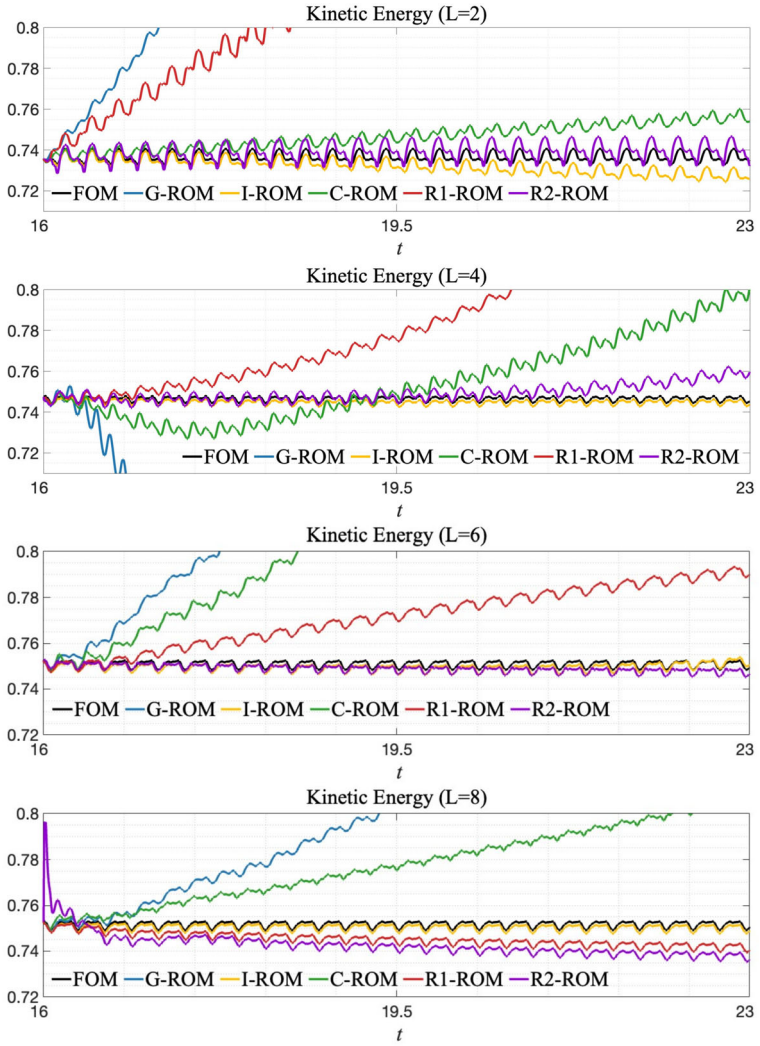

Fig. 14 Flow past a cylinder; kinetic energy of FOM projection, G-ROM, I-ROM, C-ROM, R1-ROM, and R2-ROM for L = 2, 4, 6, 8

In Table 9, we list the vortex shedding frequency, i.e., $f_s = 1/\mathcal{T}_s$, of the FOM and d2-VMS-ROMs, i.e., C-ROM, R1-ROM, and R2-ROM, for various L values, i.e., $L = 2, 4, 6, 8$. Overall, the results in Table 9 show that R2-ROM yields the most accurate predictions of the Strouhal number, especially for low L values.

In Table 10, we list the relative vortex shedding frequency errors of d2-VMS-ROMs:

$$\mathcal{E}_s = \frac{|f_s^{FOM} - f_s^{ROM}|}{f_s^{FOM}}. \tag{36}$$

Based on the relative errors in vortex shedding frequency listed in Table 10, R2-ROM demonstrates superior accuracy in capturing the vortex shedding period of the FOM among the d2-VMS-ROM variants.

Table 9 Flow past a cylinder; vortex shedding frequency, i.e., f_s for FOM, C-ROM, R1-ROM, and R2-ROM for various L values (number of POD modes)

L	FOM	C-ROM	R1-ROM	R2-ROM
2	3.77e+00	3.74e+00	3.83e+00	3.79e+00
4	3.77e+00	3.79e+00	3.73e+00	3.78e+00
6	3.77e+00	3.82e+00	3.78e+00	3.77e+00
8	3.77e+00	3.78e+00	3.77e+00	3.77e+00

Table 10 Flow past a cylinder; relative vortex shedding frequency error (36) for C-ROM, R1-ROM, and R2-ROM for various L values (number of POD modes)

L	C-ROM	R1-ROM	R2-ROM
2	9.14e-03	1.40e-02	3.79e-03
4	4.21e-03	1.04e-02	4.19e-04
6	1.88e-02	8.39e-04	4.19e-04
8	1.26e-03	0	0

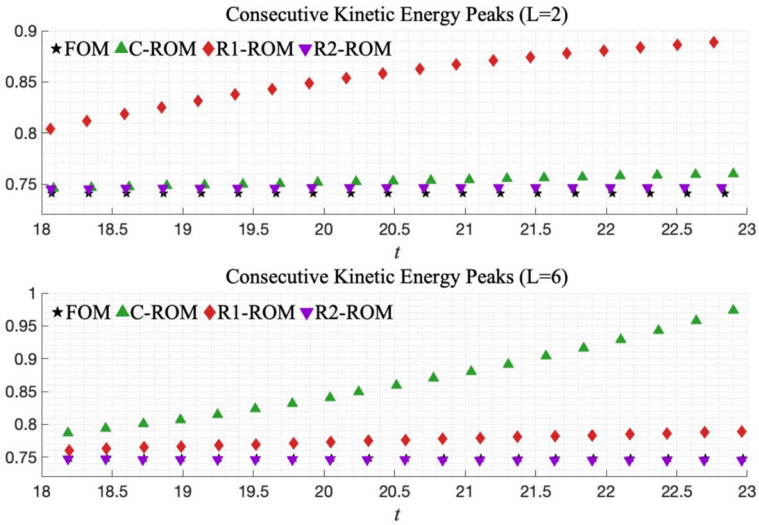

Fig. 15 Flow past a cylinder; vortex shedding periods of d2-VMS-ROMs, i.e., R1-ROM, R2-ROM, and C-ROM for $L = 2, 6$

In Fig. 15, we present the sequence of kinetic energy peaks for the FOM and the d2-VMS-ROMs for $L = 2$ and $L = 6$, thus visualizing the vortex shedding periods exhibited by each model.

For $L = 6$, in Fig. 15, we observe that both R1-ROM and R2-ROM more accurately recover the vortex shedding periods of the FOM than C-ROM. Furthermore, the amplitudes of the kinetic energy peaks in C-ROM are noticeably higher than those in R1-ROM and R2-ROM. These observations are consistent with the lower ROM errors (see Table 7) and kinetic energy errors (see Table 8) exhibited by R1-ROM and R2-ROM.

For $L = 2$, R2-ROM demonstrates the best match to FOM in terms of both vortex shedding frequency and amplitude of kinetic energy peaks. Although the vortex shedding frequencies of C-ROM and R1-ROM are similar, R1-ROM significantly overestimates the peak amplitudes compared to FOM.

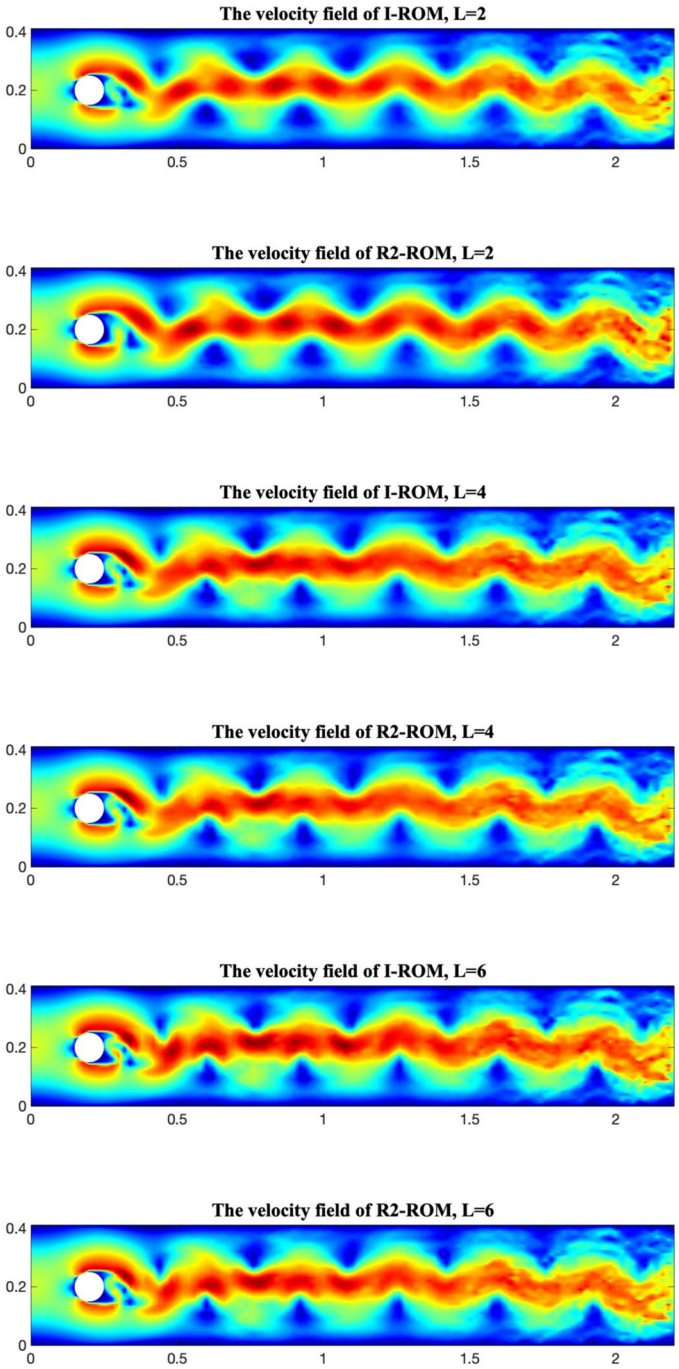

Fig. 16 Flow past a cylinder; the velocity of the I-ROM and most accurate d2-VMS-ROM, i.e., R2-ROM, for $L = 2, 4, 6$ at $T = 23$

Finally, in Fig. 16, we plot the velocity field of I-ROM and the most accurate d2-VMS-ROMs, i.e., R2-ROM, for $L = 2, 4, 6$ at the final time, $T = 23$. We observe that R2-ROM recovers the main features of the flow with good accuracy.

6 Conclusions and outlook

In this paper, we proposed a novel residual-based data-driven ROM closure model for under-resolved, convection-dominated problems. The new data-driven ROM closure model was constructed by leveraging the VMS framework and the available data. To build the new data-driven VMS-ROM (d2-VMS-ROM), we first postulated a closure model form ansatz that depends on the ROM residual, and then we solved a least squares problem to find the ansatz parameters that yield the closest fit between the ansatz and the FOM data. We also considered two types of residual-based ansatzes, which yielded two types of d2-VMS-ROMs, denoted R1-ROM and R2-ROM. The main novelty of the proposed residual-based d2-VMS-ROMs is that they depend on the ROM residual instead of the ROM coefficients (which is the standard approach in current data-driven ROM closures). To assess the novel residual-based d2-VMS-ROMs, we compared them with the standard coefficient-based d2-VMS-ROM, denoted C-ROM (Mou et al. 2021; Xie et al. 2018). Finally, for comparison purposes, we also investigated a standard G-ROM in which neither stabilization nor closure was used. We investigated the new residual-based d2-VMS ROMs, i.e., R1-ROM and R2-ROM, as well as the standard C-ROM and G-ROM, in the numerical simulation of three test problems: (i) a one-dimensional parameter-dependent advection-diffusion problem (Section 5.1); (ii) a two-dimensional time-dependent advection-diffusion-reaction problem with a small diffusion coefficient $\varepsilon = 1e-4$ (Section 5.2); and (iii) a two-dimensional flow past a cylinder at Reynolds number $Re = 1000$ (Section 5.3).

Our numerical investigation yielded the following conclusions: The novel residual-based d2-VMS-ROMs, R1-ROM and R2-ROM, were significantly more accurate than the standard coefficient-based d2-VMS-ROM, C-ROM. Furthermore, R2-ROM generally yielded slightly more accurate results than R1-ROM, but there were cases in which R1-ROM was more accurate. Finally, all the d2-VMS-ROMs (i.e., R1-ROM, R2-ROM, and C-ROM) were significantly more accurate than the standard G-ROM. Since R1-ROM has a simpler formulation than R2-ROM and is easier to construct, and since the R1-ROM and R2-ROM accuracies are similar, R1-ROM appears preferable to R2-ROM in practice.

There are several research directions that can be pursued next. Probably the most important is the extension of the new residual-based d2-VMS-ROM framework to more complex convection-dominated problems, such as under-resolved turbulent flows. Another important research direction is providing mathematical support for the new residual-based d2-VMS-ROM. For example, we plan to prove the residual-based d2-VMS-ROM's verifiability, i.e., to show that when the closure model error decreases, the ROM error decreases at the same rate. The first step in this direction has been taken in Koc et al. (2022), where we proved the verifiability of the standard coefficient-based d2-VMS-ROM, C-ROM.

Acknowledgements The first author is partially supported by Project PID2021-123153OB-C21 funded by MCIN/AEI/10.13039/501100011033/FEDER, UE and Juan de la Cierva 2022 with project number 2023/1061. The second and third authors are funded by Project PID2021-123153OB-C21 funded by MCIN/AEI/10.13039/501100011033/FEDER, UE. The fourth author is funded by ARIA MSCA-RISE EU Grant 872442 and National Science Foundation grant DMS-2012253.

Author Contributions We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who meet the authorship criteria but are not listed. We further confirm that all have approved the order of authors listed in the manuscript of us. We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing, we confirm that we have followed the regulations of our institutions concerning intellectual property. We understand that the Corresponding Author is the sole contact for the Editorial process (including Editorial Manager and direct communications with the office). The corresponding author is responsible for communicating with the other authors about progress, submissions of revisions, and final approval of proofs. We confirm that we have provided a current, correct email address accessible by the Corresponding Author.

Funding Funding for open access publishing: Universidad de Sevilla/CBUA. This work has been supported by National Science Foundation grant DMS-2012253 and Project PID2021-123153OB-C21 funded by MCIN/AEI/10.13039/501100011033/FEDER, UE.

Data Availability Data will be made available on request.

Declarations

Competing Interests We wish to confirm that there are no known conflicts of interest associated with this publication and that there has been no significant financial support for this work that could have influenced its outcome. Furthermore, the authors have no relevant financial or nonfinancial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Ahmed N, Chacón Rebollo T, John V, Rubino S (2017) A review of variational multiscale methods for the simulation of turbulent incompressible flows. Arch Comput Methods Eng 24(1):115–164. https://doi.org/10.1007/s11831-015-9161-0

Ahmed SE, Pawar S, San O, Rasheed A, Iliescu T, Noack BR (2021) On closures for reduced order models — a spectrum of first-principle to machine-learned avenues. Phys Fluids 33(9):091301

Ahmed SE, San O, Rasheed A, Iliescu T, Veneziani A (2023) Physics guided machine learning for variational multiscale reduced order modeling. SIAM J Sci Comput 45(3):283–313

Azaïez M, Chacón Rebollo T, Rubino S (2021) A cure for instabilities due to advection-dominance in POD solution to advection-diffusion-reaction equations. J Comput Phys 425:109916. https://doi.org/10.1016/j.jcp.2020.109916

Balajewicz MJ, Tezaur I, Dowell EH (2016) Minimal subspace rotation on the Stiefel manifold for stabilization and enhancement of projection-based reduced order models for the compressible Navier-Stokes equations. J Comput Phys 321:224–241

Ballarin F, Manzoni A, Quarteroni A, Rozza G (2015) Supremizer stabilization of POD-Galerkin approximation of parametrized steady incompressible Navier-Stokes equations. Int J Numer Meth Engng 102:1136–1161

Bergmann M, Bruneau CH, Iollo A (2009) Enablers for robust POD models. J Comput Phys 228(2):516–538

Chácon Rebollo T, Dia BM (2015) A variational multi-scale method with spectral approximation of the subscales: Application to the 1D advection-diffusion equations. Comput Meth Appl Mech Engrg 285:406–426

Chacón Rebollo T, Rubino S, Oulghelou M, Allery C (2022) Error analysis of a residual-based stabilization-motivated POD-ROM for incompressible flows. Comput Meth Appl Mech Engrg 401:115627

Eroglu FG, Kaya S, Rebholz LG (2017) A modular regularized variational multiscale proper orthogonal decomposition for incompressible flows. Comput Meth Appl Mech Engrg 325:350–368
Girfoglio M, Quaini A, Rozza G (2021) A POD-Galerkin reduced order model for a LES filtering approach. J Comput Phys 436:110260
Hesthaven JS, Rozza G, Stamm B (2016) Certified reduced basis methods for parametrized partial differential equations. Springer, Cham
Holmes P, Lumley JL, Berkooz G (1996) Turbulence, Coherent structures. Dynamical Systems and Symmetry. Cambridge University Press, Cambridge
Hughes TJR, Feijóo GR, Mazzei L, Quincy J-B (1998) The variational multiscale method–a paradigm for computational mechanics. Comput Methods Appl Mech Engrg 166(1–2):3–24. https://doi.org/10.1016/S0045-7825(98)00079-6
Iliescu T, Wang Z (2013) Variational multiscale proper orthogonal decomposition: Convection-dominated convection-diffusion-reaction equations. Math Comput 82(283):1357–1378
Iliescu T, Wang Z (2014) Variational multiscale proper orthogonal decomposition: Navier-Stokes equations. Num Meth P.D.E.s 30(2):641–663
John V (2004) Reference values for drag and lift of a two dimensional time-dependent flow around a cylinder. Int J Num Meth Fluids 44:777–788
John V, Linke A, Merdon C, Neilan M, Rebholz LG (2016) On the divergence constraint in mixed finite element methods for incompressible flows. SIAM Rev
Koc B, Mou C, Liu H, Wang Z, Rozza G, Iliescu T (2022) Verifiability of the data-driven variational multiscale reduced order model. J Sci Comput 93(2):54
Maulik R, Mohan A, Lusch B, Madireddy S, Balaprakash P, Livescu D (2020) Time-series learning of latent-space dynamics for reduced-order model closure. Phys D 405:132368
Mohebujjaman M, Rebholz LG, Xie X, Iliescu T (2017) Energy balance and mass conservation in reduced order models of fluid flows. J Comput Phys 346:262–277
Mohebujjaman M, Rebholz LG, Iliescu T (2019) Physically-constrained data-driven correction for reduced order modeling of fluid flows. Int J Num Meth Fluids 89(3):103–122
Mou C (2021) Data-driven variational multiscale reduced order modeling of turbulent flows. PhD thesis, Virginia Tech
Mou C, Liu H, Wells DR, Iliescu T (2020) Data-driven correction reduced order models for the quasi-geostrophic equations: A numerical investigation. Int J Comput Fluid Dyn 34(2):147–159
Mou C, Koc B, San O, Rebholz LG, Iliescu T (2021) Data-driven variational multiscale reduced order models. Comput Meth Appl Mech Engrg 373:113470
Mou C, Chen N, Iliescu T (2023) An efficient data-driven multiscale stochastic reduced order modeling framework for complex systems. J Comp Phys 493:112450
Noack BR, Papas P, Monkewitz PA (2005) The need for a pressure-term representation in empirical Galerkin models of incompressible shear flows. J Fluid Mech 523:339–365
Noack BR, Morzynski M, Tadmor G (2011) Reduced-Order modelling for flow control, vol 528. Springer, Vienna
Novo J, Rubino S (2021) Error analysis of proper orthogonal decomposition stabilized methods for incompressible flows. SIAM J Numer Anal 59(1):334–369. https://doi.org/10.1137/20M1341866
Parish EJ, Duraisamy K (2017) A unified framework for multiscale modeling using the Mori-Zwanzig formalism and the variational multiscale method. arXiv preprint, http://arxiv.org/abs/1712.09669
Parish EJ, Wentland C, Duraisamy K (2020) The adjoint Petrov-Galerkin method for non-linear model reduction. Comput Meth Appl Mech Engrg 365:112991
Prakash A, Zhang YJ (2024) Projection-based reduced order modeling and data-driven artificial viscosity closures for incompressible fluid flows. Comput Meth Appl Mech Engrg 425:116930
Quarteroni A, Manzoni A, Negri F (2015) Reduced basis methods for partial differential equations: An introduction, vol 92. Springer, Cham
Rebholz L, Xiao M (2017) Improved accuracy in algebraic splitting methods for Navier-Stokes equations. SIAM J Sci Comput 39(4):1489–1513
Reyes R, Codina R (2020) Projection-based reduced order models for flow problems: A variational multiscale approach. Comput Meth Appl Mech Engrg 363:112844
Roop JP (2013) A proper-orthogonal decomposition variational multiscale approximation method for a generalized Oseen problem. Adv Numer Anal 2013(1):974284
Sanderse B (2020) Non-linearly stable reduced-order models for incompressible flow with energy-conserving finite volume methods. J Comput Phys 421:109736
Sanderse B, Stinis P, Maulik R, Ahmed SE (2024) Scientific machine learning for closure models in multiscale problems: a review. arXiv preprint arXiv:2403.02913

Stabile G, Rozza G (2018) Finite volume POD-Galerkin stabilised reduced order methods for the parametrised incompressible Navier-Stokes equations. Comput & Fluids 173:273–284

Stabile G, Ballarin F, Zuccarino G, Rozza G (2019) A reduced order variational multiscale approach for turbulent flows. Adv Comput Math, 1–20

Tello A, Codina R, Baiges J (2019) Fluid structure interaction by means of variational multiscale reduced order models. Int J Num Meth Eng

Volkwein S (2013) Proper orthogonal decomposition: Theory and reduced-order modelling. Lecture Notes, University of Konstanz. http://www.math.uni-konstanz.de/numerik/personen/volkwein/teaching/POD-Book.pdf

Xie X, Mohebujjaman M, Rebholz LG, Iliescu T (2018) Data-driven filtered reduced order modeling of fluid flows. SIAM J Sci Comput 40(3):834–857

Xie X, Webster C, Iliescu T (2020) Closure learning for nonlinear model reduction using deep residual neural network. Fluids 5(1):39

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