Data-Driven Learning of Reduced-order Dynamics for a Parametrized Shallow Water Equation

Süleyman Yıldız,[*] Pawan Goyal,[†] Peter Benner,[‡] Bülent Karasözen[§]

Abstract

This paper discusses a non-intrusive data-driven model order reduction method that learns low-dimensional dynamical models for a parametrized shallow water equation. We consider the shallow water equation in non-traditional form (NTSWE). We focus on learning low-dimensional models in a non-intrusive way. That means, we assume not to have access to a discretized form of the NTSWE in any form. Instead, we have snapshots that are obtained using a black-box solver. Consequently, we aim at learning reduced-order models only from the snapshots. Precisely, a reduced-order model is learnt by solving an appropriate least-squares optimization problem in a low-dimensional subspace. Furthermore, we discuss computational challenges that particularly arise from the optimization problem being ill-conditioned. Moreover, we extend the non-intrusive model order reduction framework to a parametric case where we make use of the parameter dependency at the level of the partial differential equation. We illustrate the efficiency of the proposed non-intrusive method to construct reduced-order models for NTSWE and compare it with an intrusive method (proper orthogonal decomposition). We furthermore discuss the predictive capabilities of both models outside the range of the training data.

Keywords: Shallow water equation, scientific machine learning, data-driven modeling, model order reduction, operator inference.

1 Introduction

Shallow water equations (SWE) are a popular set of hyperbolic PDEs with the capability of describing geophysical wave phenomena, e.g., the Kelvin and Rossby waves in the atmosphere and in the oceans. They are frequently used in geophysical flow prediction [11], investigation of baroclinic instability [8, 37], and planetary flows [38]. In this paper, we study a model order reduction (MOR) technique for SWE. MOR techniques allow us to construct low-dimensional models or reduced-order models (ROMs) for a large-scale dynamical system. We refer to the books [5, 27] for an overview of the available techniques. These ROMs are computationally efficient and accurate, and are worthy when a full order model (FOM) needs to be simulated multiple-times for different parameter settings. Additionally, ROMs are even more valuable in the case of SWE, when the interest lies in simulating the model for a very long time horizon. MOR problems for SWE have been intensively studied in the literature, see, e.g., [6, 7, 15, 21, 22, 29].

Most MOR techniques are intrusive in nature. This means that these methods require access to the large-scale semi-discretized FOM, preferably in a matrix-vector form. Moreover, ROMs are typically constructed by projecting the high-fidelity FOM onto a low-dimensional subspace using appropriate projection matrices. The proper orthogonal decomposition (POD) is arguably one of the most popular methods that can be seen as a data-driven intrusive method. POD is data-driven in the sense that we require training data that are usually the solution trajectories for given inputs, initial conditions, and parameters. By taking the singular value decomposition (SVD) of the training data, we determine a low-dimensional subspace, where the most important system dynamics reside. The efficiency of the ROM is based on the separation of the offline cost for evaluating the FOMs and the online cost for evaluating the ROMs.

One of the major drawbacks of intrusive methods is that they require access to the FOM. However, for a complex dynamical process, it is a challenging task to obtain an explicit discretized FOM. It is even intractable if the process is simulated using proprietary software. Therefore, in this work, we are interested in a non-intrusive approach to construct ROMs, where we do not have access to a discretized FOM. We rather have only simulation data, potentially obtained using proprietary software, corresponding to the FOM.

Building a model using only the simulation data directly fits the philosophy of machine learning and neural networks. Using neural networks, a large class of functions [19] can be approximated. These methods aim at constructing an input-output mapping based on data. They learn a model based on the training data that neither requires explicit access to the high-fidelity model operators nor any additional information about the process. However, the amount of data required to learn the model accurately imposes a burden in the context of large-scale PDE simulations [35]. Moreover, some ideas from compressive sensing have been used to learn the operators of a FOM from a large library of candidate functions [28].

---

[*]Institute of Applied Mathematics, Middle East Technical University, 06800 Ankara, Turkey yildiz.suleyman@metu.edu.tr

[†]Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstraße 1, D-39106 Magdeburg, Germany goyalp@mpi-magdeburg.mpg.de

[‡]Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstraße 1, D-39106 Magdeburg, Germany benner@mpi-magdeburg.mpg.de

[§]Institute of Applied Mathematics & Department of Mathematics, Middle East Technical University, 06800 Ankara, Turkey bulent@metu.edu.tr

However, the success of the method heavily depends on the built library, and we generally need to perform computations in the full-order system dimension, thus making the method very challenging in large-scale settings. In recent years, the operator inference (OpInf) framework to construct ROMs has gained much attention. The framework utilizes the knowledge of nonlinear terms at the PDE level. In this framework, the operators defining the ROM can be learnt by formulating an optimization problem, without necessitating the discretized operators of the PDEs. Such a scheme was first investigated in [24] for polynomial nonlinearities. The methodology was later extended to a class of nonlinear systems that can be written as a polynomial or quadratic-bilinear (QB) system by introducing new state variables in [26, 25]. Recently, the authors in [3] have extended the approach to nonlinear systems in which the structure of the nonlinearities is preserved while learning ROMs from data.

In this work, we discuss an application of the OpInf framework [24] to the parameterized NTSWE. OpInf is also investigated in [24] for parametric cases, where the ROMs are constructed at each training parameter via interpolation. However, in this work, we discuss an OpInf framework for the parametric case, where we make use of the known parametric dependency at the PDE level. In the case of a large amount of data, the optimization problem that yields the reduced operators is generally a discrete ill-posed least-squares problem. To mitigate this issue, a regularized least-squares optimization problem is proposed in [24]. In this paper, we discuss alternative approaches such as Tikhonov regularization, truncated SVD and truncated QR.

The remaining structure of the paper is as follows. In Section 2, the NTSWE is briefly described. In Section 3, we discuss the OpInf method to infer reduced operators from data and present its extension to the parametric case. Furthermore, we investigate computational issues related to the optimization problem that learns the reduced operators. In Section 4, we present numerical experiments, where ROMs for (parametric) NTSWEs models are inferred directly from data. The inferred ROMs are compared with ROMs obtained from the intrusive POD method. We show that non-intrusive ROMs outperform in most instances, particularly, in the prediction outside the training data. In Section 5, we provide concluding remarks.

2 Shallow Water Equation

In most ocean and atmosphere models, the Coriolis force only depends on the component of the planetary rotation vector. It is locally normal to the geopotential surfaces, which is called traditional approximation (TA) [14]. The TA is applicable when the horizontal length scales of rotational geophysical flows are much larger than the vertical length scales [16]. However, many atmospheric and oceanographic phenomena are substantially influenced by the non-traditional components of the Coriolis force [32], such as deep convection [23], Ekman spirals [20], and internal waves [17]. The NTSWE [13, 30, 33] includes the non-traditional components of the Coriolis force with the bottom topography. The non-dimensional NTSWE is governed by the following PDE system [33]:

$$\frac{\partial \tilde{\mathbf{u}}}{\partial t} + qh\mathbf{z} \times \mathbf{u} + \nabla\Phi = 0, \tag{1a}$$

$$\frac{\partial h}{\partial t} + \nabla \cdot (h\mathbf{u}) = 0, \tag{1b}$$

where $\tilde{\mathbf{u}} =: (\tilde{u}, \tilde{v})$ is the canonical velocity, $\mathbf{u} =: (u, v)$ is the particle velocity, h is the height field, q is the potential vorticity defined as $q =: \frac{1}{h}\left(\Omega^z + \tilde{v}_x - \tilde{u}_y\right)$, and Φ is the Bernoulli potential, given by

$$\Phi = \frac{1}{2}\left(u^2 + v^2\right) + h_b + h + \frac{1}{2}\delta h\left(\Omega^x v - \Omega^y u\right). \tag{2}$$

The particle velocities are given in terms of the canonical velocities as

$$u = \tilde{u} - \delta\Omega^y\left(h_b + \frac{1}{2}h\right), \qquad v = \tilde{v} + \delta\Omega^x\left(h_b + \frac{1}{2}h\right), \tag{3}$$

where h_b is the bottom topography, $\delta := L/R_d$ is the so-called non-traditional parameter with L being the layer thickness scale, and R_d is the Rossby deformation radius [30, 13]. Ω^x and Ω^y are the x and y components of the angular velocity vector Ω, respectively, and x and y denote horizontal distances within a constant geopotential surface. The orientation of the x and y axes are considered arbitrary with respect to the North. The NTSWE (1) describes inviscid fluid that flows over the bottom topography at $z = h_b(x, y)$ in a frame rotating with angular velocity vector $\boldsymbol{\Omega} = (\Omega^x, \Omega^y, \Omega^z)$. Both Ω^x and Ω^y depend on x and y axes but not on z. The dimensionless angular velocity vector can be given as [31]:

$$\Omega^x = \cos\theta\sin\phi, \quad \Omega^y = \cos\theta\cos\phi, \quad \text{and} \quad \Omega^z = \sin\theta, \tag{4}$$

where θ is the angle corresponding to the latitude, and ϕ is the angle determining the orientation between the x-axis and the eastward direction. In this paper, we set the x-axis of the rotation vector to zero, implying that it is aligned to the East. Moreover, we consider the layer thickness scale for the ocean: $L = 1000\text{m}$, the deformation radius $R_d \approx 6.88\text{km}$, with no bottom topography $h_b = 0$ and the non-traditional parameter $\delta = 0.145$. In Figure1, the components of the angular velocity vector are shown for the latitude θ and $\phi = 0$.
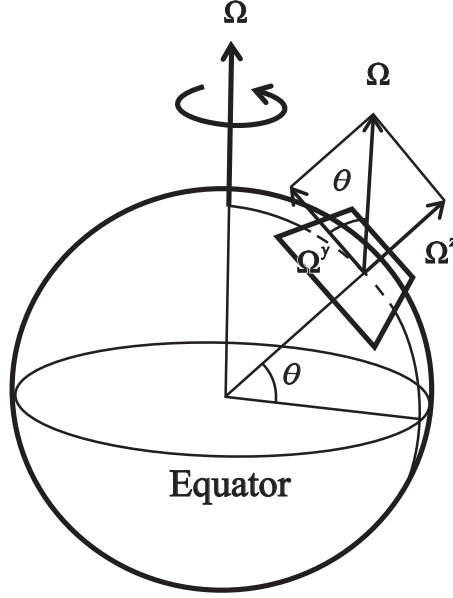
Figure 1: Outline of the components of the angular velocity vector Ω with respect to the latitude angle θ as $\Omega^x = 0$.

3 Learning Parameterized ROMs of Shallow Water Equations

In this section, we study an approach to learn ROMs for NTSWE from data, e.g., obtained from proprietary software, or real-world measurements. In Subsection 3.1, we begin our discussion with the quadratic form of the parametric NTSWE. We furthermore discuss the construction of ROMs via an intrusive POD method. Subsection 3.2 presents an operator inference approach to learn the reduced parametric operators from (simulation) data, where we make use of the knowledge of the parametric dependency at the PDE level. Moreover, we discuss computational aspects for constructing the reduced operators in Subsection 3.3.

3.1 Exploiting the quadratic form of the parameterized NTSWE

The success of the OpInf approach [24, 25] lies in exploiting the structure at the PDE level. The OpInf approach aims at determining a ROM, without having access to the FOMs in a matrix-vector form. It consists of setting up an optimization problem for determining the reduced operators by taking advantage of the underlying structure of the FOM but at the PDE level.

The NTSWE (1) can be explicitly written in terms of the canonical velocities by taking into account that the x-axis of the angular velocity vector is aligned to the East and setting $\phi = 0$ as

$$\frac{\partial}{\partial t}\tilde{u} = -h_x + \sin\theta\,\tilde{v} - \tilde{u}\tilde{u}_x - \tilde{v}\tilde{u}_y + \delta\cos\theta(h\tilde{u})_x - \frac{3}{8}\left(\delta\cos\theta\right)^2(h^2)_x, \tag{5a}$$

$$\frac{\partial}{\partial t}\tilde{v} = -h_y + \sin\theta\,\tilde{u} + \frac{1}{2}\delta\sin\theta\cos\theta\,h - \tilde{u}\tilde{v}_x - \tilde{v}\tilde{v}_y + \delta\cos\theta\left((h\tilde{u})_y + \frac{1}{2}h\left(\tilde{v}_x - \tilde{u}_y\right)\right) - \frac{3}{8}\left(\delta\cos\theta\right)^2(h^2)_y, \tag{5b}$$

$$\frac{\partial}{\partial t}h = -(h\tilde{u})_x - (h\tilde{v})_y + \frac{1}{2}\delta\cos\theta(h^2)_x. \tag{5c}$$

3.2 Operator inference approach to learning parameterized reduced-operators

Let us consider a parameter vector $\mu \in \mathcal{D} \subset \mathbb{R}^d$, the state vector $\mathbf{w} : [0, T] \times \mathcal{D} \to \mathbb{R}^N$ with N degrees of freedom, and the time $t \in [0, T] \subset \mathbb{R}$. The NTSWE model (5) includes linear terms and quadratic polynomial nonlinearities that can be exploited to create quadratic ROMs. Hence, we consider the following linear-quadratic ODE system:

$$\dot{\mathbf{w}}(t; \mu) = \mathbf{A}(\mu)\mathbf{w}(t; \mu) + \mathbf{H}(\mu)(\mathbf{w}(t; \mu) \otimes \mathbf{w}(t; \mu)), \tag{6}$$

where $\mathbf{A}(\mu) \in \mathbb{R}^{N \times N}$ corresponds to the linear terms, $\mathbf{H}(\mu) \in \mathbb{R}^{N \times N^2}$ to the quadratic term. We allow the initial condition to depend on the parameter μ as well, i.e., $\mathbf{w}(0, \mu) = \mathbf{w}_0(\mu)$.

Our primary goal is to construct a reduced-order parametrized model that captures the important dynamics of the high-dimensional model (6) for a given parameter range as follows:

$$\dot{\hat{\mathbf{w}}}(t; \mu) = \widehat{\mathbf{A}}(\mu)\widehat{\mathbf{w}}(t; \mu) + \widehat{\mathbf{H}}(\mu)(\widehat{\mathbf{w}}(t; \mu) \otimes \widehat{\mathbf{w}}(t; \mu)), \tag{7}$$

3

where $\widehat{\mathbf{A}}(\mu) \in \mathbb{R}^{r \times r}$ and $\widehat{\mathbf{H}}(\mu) \in \mathbb{R}^{r \times r^2}$ with $r \ll N$. If the FOM is available in explicit form, e.g., in matrix-vector form, then intrusive MOR techniques can be applied, such as POD [4, 2] and interpolation-based methods [1]. Assuming an explicit form of the FOM is available, the ROM can be constructed with the projection matrix $\mathbf{V} \in \mathbb{R}^{N \times r}$ so that $\mathbf{w}(t, \mu) \approx \mathbf{V}\widehat{\mathbf{w}}(t, \mu)$, for all $t \geq 0$ and $\mu \in \mathcal{D}$ obtained by POD. Then, reduced-order matrices of the system (7) can be computed as follows:

$$\widehat{\mathbf{A}}(\mu) = \mathbf{V}^T \mathbf{A}(\mu) \mathbf{V} \in \mathbb{R}^{r \times r}, \quad \widehat{\mathbf{H}}(\mu) = \mathbf{V}^T \mathbf{H}(\mu)(\mathbf{V} \otimes \mathbf{V}). \tag{8}$$

Furthermore, assume that the system matrices in (6) depend affinely on functions of the parameter μ:

$$\mathbf{A}(\mu) = \alpha_1(\mu)\mathbf{A}_1 + \cdots + \alpha_{\mathbf{n_a}}(\mu)\mathbf{A}_{\mathbf{n_a}}, \tag{9a}$$

$$\mathbf{H}(\mu) = \eta_1(\mu)\mathbf{H}_1 + \cdots + \eta_{\mathbf{n_h}}(\mu)\mathbf{H}_{\mathbf{n_h}}, \tag{9b}$$

where $\mathbf{A}_i \in \mathbb{R}^{N \times N}$, $\mathbf{H}_j \in \mathbb{R}^{N \times N^2}$ are constant matrices, and $\alpha_i(\mu), \eta_j(\mu) : \mathbb{R}^d \to \mathbb{R}$ are smooth functions of the parameter μ. In this case, the reduced-matrices in (8) can be precomputed, e.g., $\widehat{\mathbf{A}}(\mu) = \alpha_1(\mu)\widehat{\mathbf{A}}_1 + \cdots + \alpha_{\mathbf{n_a}}(\mu)\widehat{\mathbf{A}}_{\mathbf{n_a}}$, where $\widehat{\mathbf{A}}_i = \mathbf{V}^T \mathbf{A}_i \mathbf{V}$, $i \in \{1, \ldots, \mathbf{n_a}\}$.

However, as discussed earlier, it is not easy or almost impossible to obtain the FOM in an explicit matrix form, from proprietary software. Therefore, our primary interest lies in constructing reduced-order operators without having access to the FOM, but rather having access only to simulation data and some knowledge at the PDE level. With this aim, we collect simulation data for a training parameter set, $\mu_i \in \mathcal{D}$ for $i = 1, \ldots, M$. Thus, let us define a global snapshot matrix:

$$\mathbf{S}_\mu = [\mathbf{S}(\mu_1), \ldots, \mathbf{S}(\mu_M)], \quad \mathbf{S}(\mu_i) = [\mathbf{w}(t_1; \mu_i), \mathbf{w}(t_2; \mu_i), \ldots, \mathbf{w}(t_K; \mu_i)] \in \mathbb{R}^{N \times K}, \tag{10}$$

where $\mathbf{w}(t_j, \mu_i)$ denotes the value at time t_j for the parameter μ_i. The projection matrix $\mathbf{V}$ is determined by the SVD of the snapshot matrix

$$\mathbf{S}_\mu = \mathbf{V}_\mu \Sigma_\mu \mathbf{U}_\mu^T, \tag{11}$$

where $\mathbf{V}_\mu \in \mathbb{R}^{N \times M \cdot K}, \Sigma_\mu \in \mathbb{R}^{M \cdot K \times M \cdot K}, \mathbf{U}_\mu \in \mathbb{R}^{M \cdot K \times M \cdot K}$, and $\mathbf{V}$ is given then by the first r columns of $\mathbf{V}_\mu$. In order to determine reduced operators by employing an OpInf approach, we first project the snapshot matrix $\mathbf{S}_\mu$ onto the dominant subspace spanned by $\mathbf{V}$, yielding the reduced snapshot matrix:

$$\widehat{\mathbf{S}}_\mu := \mathbf{V}^T \mathbf{S}_\mu = \left[\widehat{\mathbf{S}}(\mu_1), \ldots, \widehat{\mathbf{S}}(\mu_M)\right], \tag{12}$$

where

$$\widehat{\mathbf{S}}(\mu_i) = [\widehat{\mathbf{w}}(t_1; \mu_i), \widehat{\mathbf{w}}(t_2; \mu_i), \ldots, \widehat{\mathbf{w}}(t_K; \mu_i)] \in \mathbb{R}^{r \times K}$$

in which $\widehat{\mathbf{w}}(t_j, \mu_i) := \mathbf{V}^T \mathbf{w}(t_j, \mu_i)$. Furthermore, let us define

$$\dot{\widehat{\mathbf{S}}}_\mu = \left[\dot{\widehat{\mathbf{S}}}(\mu_1), \ldots, \dot{\widehat{\mathbf{S}}}(\mu_M)\right], \tag{13}$$

where $\dot{\widehat{\mathbf{S}}}(\mu_i)$ can either be determined using the right-hand side of (6)–if accessible–followed by projecting using $\mathbf{V}$, or can be approximated using $\widehat{\mathbf{S}}(\mu_i)$ by employing a time-derivative approximation scheme, see, e.g., [24]. Subsequently, the reduced operators of the reduced parametric model (7) are determined by solving the following least-squares problem:

$$\min_{\widehat{\mathbf{A}}_i \in \mathbb{R}^{r \times r}, \widehat{\mathbf{H}}_j \in \mathbb{R}^{r \times r^2}} \sum_{k=1}^{M} \left\| -\dot{\widehat{\mathbf{S}}}(\mu_k)^T + \sum_{i=1}^{\mathbf{n_a}} \left(\alpha_i(\mu_k)\widehat{\mathbf{S}}(\mu_k)^T \widehat{\mathbf{A}}_i^T\right) + \sum_{i=1}^{\mathbf{n_h}} \left(\eta_i(\mu_k)\left(\widehat{\mathbf{S}}(\mu_k)\hat{\otimes}\widehat{\mathbf{S}}(\mu_k)\right)^T \widehat{\mathbf{H}}_i^T\right) \right\|_F^2 \tag{14}$$

where $\hat{\otimes}$ denotes the column-wise Kronecker product. It can be noted that the optimization problem (14) does not involve any explicit knowledge of the FOM, it only involves simulation data projected onto the dominant POD subspace. Moreover, we can rewrite the optimization problem (14) in standard form as follows:

$$\min_{\mathcal{X} \in \mathbb{R}^{r \times \mathbf{n_a}r + \mathbf{n_h}r^2}} \sum_{k=1}^{M} \left\| \mathcal{A}(\mu_k)\mathcal{X}^T - \dot{\widehat{\mathbf{S}}}(\mu_k)^T \right\|_F^2, \tag{15}$$

where

$$\mathcal{X} = \left[\widehat{\mathbf{A}}_1, \ldots, \widehat{\mathbf{A}}_{\mathbf{n_a}}, \widehat{\mathbf{H}}_1, \ldots, \widehat{\mathbf{H}}_{\mathbf{n_h}}\right], \text{ and}$$

$$\mathcal{A}(\mu_k) = \left[[\alpha_1(\mu_k), \ldots, \alpha_{\mathbf{n_a}}(\mu_k)] \otimes \widehat{\mathbf{S}}(\mu_k)^T, [\eta_1(\mu_k), \ldots, \eta_{\mathbf{n_h}}] \otimes \left(\widehat{\mathbf{S}}(\mu_k)\hat{\otimes}\widehat{\mathbf{S}}(\mu_k)\right)^T\right].$$

3.3 Computational Aspects

In this section, we discuss computational aspects of the OpInf approach (15). Solving the least-squares problem (15) can be a computationally challenging task because (a) the problem can be highly ill-conditioned, and (b) its computational cost grows quadratically with the order r of the reduced system and linearly with the number of snapshots. The computational cost of the optimization problem (15) can be reduced by decoupling of the least-squares problem. In the remainder, we discuss techniques for the conditioning. In the OpInf framework, inferred operators are solutions of the potentially discrete ill-posed least-squares problem (15), where ill-conditioning may be due to nearly linearly dependent columns of the snapshot matrix. When the distance to a matrix with linearly dependent columns decreases, the condition number of the data matrix in the least-squares problem increases. Hence, the least-squares problem arising from the OpInf framework needs a suitable regularization method. There exist different ways to deal with this issue.

A suitable and widely used candidate for this task is Tikhonov regularization [36]. Tikhonov regularization filters the small singular values to reduce the amplification effect on the least-squares algorithm. The quality of learning via Tikhonov regularization depends on the L-curve [18]. Using the L-curve information, the stability of the learning algorithm can be improved [34]. However, the computation of the efficient Tikhonov parameter is costly for large problems [9]. Although one can argue that the problem (15) is in a low dimension, it still can be of a large scale when the number of snapshots or/and the number of training parameters are large. The Tikhonov regularization applied to (15) can be written in a compact form as follows:

$$\min_{x_i \in \mathbb{R}^{\mathbf{n_a} r + \mathbf{n_h} r^2}} \|\mathcal{A}_\mu x_i - s_i\|_2^2 + \lambda \|x_i\|_2^2, \quad i = 1, \ldots, r, \tag{16}$$

where x_i are the columns of $\mathcal{X}^T$, s_i are the columns of $\hat{\dot{\mathbf{S}}}_\mu$ and $\mathcal{A}_\mu = \left[\mathcal{A}^T(\mu_1), \ldots, \mathcal{A}^T(\mu_M) \right]^T$. A heuristic approach to deal with the conditioning of the data matrix is proposed in [24] in which a subset of the data is considered, by taking the data in a regular interval, e.g., every 10th time-step. In [24], it is shown that this can alleviate the ill-conditioning problem to some extend in some cases. However, the choice of the interval should be done in such a way that the important snapshots are not missed, thus the choice of the interval plays a key role. This problem can be referred to as a heuristic column subset selection problem (CSSP). The CSSP seeks to find a subset of the most linearly independent columns of a matrix which gives the best information in the matrix.

The truncated QR method (tQR) with the minimum norm solution [10] finds a suitable subset for the CSSP, which can be used to find an accurate solution of the rank-deficient least-squares problems. For the tQR method, we use the QR decomposition of the data matrix with column pivoting (QR-CP). A major advantage of the tQR algorithm is that it allows us to monitor the linearly dependent columns via QR-CP. Thus, it also allows us to improve the condition number of the data matrix by selecting linearly independent vectors from the data matrix. The QR-CP algorithm naturally finds a subset for the CSSP problem via a permutation matrix. One alternative to this approach is the truncated SVD (tSVD) algorithm. The tSVD is also an attractive method with its best rank-k approximation. The tSVD method and tQR method generally give very close solutions. Nevertheless, the calculation of the tSVD is more expensive than the tQR algorithm.

For simplicity, suppose $\mathcal{A}_\mu \in \mathbb{R}^{m \times n}$ is exactly rank deficient with rank$(\mathcal{A}_\mu) = p$. Then, there always exists a QR-CP factorization of $\mathcal{A}_\mu$ of the form

$$\mathcal{A}_\mu \Pi = QR, \tag{17}$$

where $\Pi \in \mathbb{R}^{n \times n}$ is a permutation matrix, $Q \in \mathbb{R}^{m \times m}$ is an orthogonal matrix, and R is an upper triangular matrix of the form

$$R = \begin{pmatrix} R_1 & R_2 \\ 0 & 0 \end{pmatrix} \tag{18}$$

and $R_1 \in \mathbb{R}^{p \times p}$ is upper triangular with rank$(R_1) = p$. The diagonal entries of R in (17) satisfy $|R_{ii}| \geq |R_{jj}|$ with $j > i$ so that the effective rank of $\mathcal{A}$ can be determined as the smallest integer p such that

$$|R_{p+1,p+1}| < \texttt{tol} \cdot |R_{11}|,$$

where $\texttt{tol}$ can be considered as the tolerance for the linear dependency of the columns of data matrix $\mathcal{A}_\mu$. Hence, this approach gives us a flexibility of adjustment on the condition number of the data matrix $\mathcal{A}_\mu$. The tolerance of tQR $\texttt{tol}$ can also be determined by the L-curve [12]. Among all solutions of the optimization problem (15), we take the unique minimum L_2 norm solution. The minimum norm solution of (15) by tQR method can be obtained as in [10]. In this paper, we study an QR-CP based regularization and compare with the Tikhonov regularization in the next section.

4 Numerical results

In this section, we demonstrate the performance of the OpInf approach for two numerical test problems and compare it with the intrusive POD method. We also study the prediction capabilities of the OpInf approach for the parametric and non-parametric NTSWE. Furthermore, we examine the non-intrusive methods with the Tikhonov regularization (16) for the penalty parameter $\lambda = 0.01$ and tQR method with the tolerance 10^{-6}. To determine the regularization parameters of

Tikhonov regularization and tQR, we used L-curve criteria. As a first test example, we consider the propagation of the inertia-gravity waves by Coriolis force, known as geostrophic adjustment [33]. In the second example, we investigate the shear instability in the form of a roll-up of an unstable shear layer, known as barotropic instability [33].

The NTSWE (5) is semi-discretized in space by replacing the first-order spatial derivatives with central finite-differences. The resulting system is a quadratic semi-discrete system that depends on the parameter $\mu = \theta$:

$$\dot{\mathbf{w}}(t; \mu) = \mathbf{A}(\mu)\mathbf{w}(t; \mu) + \mathbf{H}(\mu)(\mathbf{w}(t; \mu) \otimes \mathbf{w}(t; \mu)), \tag{19}$$

where $\mathbf{A}(\mu) \in \mathbb{R}^{N \times N}$ corresponds to the linear terms, $\mathbf{H}(\mu) \in \mathbb{R}^{N \times N^2}$ to the quadratic term, $\mu \in \mathcal{D} \subset \mathbb{R}^d$, the state vector $\mathbf{w} : [0, T] \times \mathcal{D} \to \mathbb{R}^N$ with N degrees of freedom, and the time $t \in [0, T] \subset \mathbb{R}$. The operators $\mathbf{A}(\mu)$ and $\mathbf{H}(\mu)$ in (9) have affine parameter dependence (9) with respect to the parameter μ as follows:

$$\alpha_1(\mu) = 1, \quad \alpha_2(\mu) = \Omega^z = \sin(\mu), \quad \alpha_3(\mu) = \Omega^z \Omega^y = \sin(\mu)\cos(\mu), \tag{20a}$$

$$\eta_1(\mu) = 1, \quad \eta_2(\mu) = \Omega^y = \cos(\mu), \quad \eta_3(\mu) = (\Omega^y)^2 = (\cos(\mu))^2. \tag{20b}$$

We consider the NTSWE under periodic boundary conditions, and assume there are no forcing or input terms in (19). All the models are simulated using the function `ode15s` in MATLAB® with both the relative and absolute error tolerances set to 10^{-8}. In all numerical examples, the spatial domain is discretized with 101×101 equidistant grid points. The snapshots are sampled at equidistant time instances with the time step $\Delta t = 0.1$. The accuracy of the ROMs is measured using the relative error in the Frobenius norm

$$\mathcal{E} = \frac{||\mathbf{S}_{\text{FOM}} - \mathbf{V}\mathbf{S}_{\text{ROM}}||_F}{||\mathbf{S}_{\text{FOM}}||_F}, \tag{21}$$

where $\mathbf{S}_{\text{FOM}} \in \mathbb{R}^{N \times K}$ is the snapshot matrix of the FOM and $\mathbf{S}_{\text{ROM}} \in \mathbb{R}^{r \times K}$ is the snapshot matrix of either the non-intrusive or the intrusive ROMs. In the parametric case, the relative error is computed with snapshot matrices by concatenating the trajectories for parameter samples. A typical approach to determine the reduced dimension r is done through the projection error as follows:

$$\mathcal{E}_{\text{proj}} = \frac{||\mathbf{S}_{\text{FOM}} - \mathbf{V}\mathbf{V}^T\mathbf{S}_{\text{FOM}}||_F}{||\mathbf{S}_{\text{FOM}}||_F}. \tag{22}$$

4.1 Single-layer geostrophic adjustment

The initial conditions are prescribed in the form of a motionless layer with an upward bulge of the height field in a periodic domain $[-5, 5] \times [-5, 5]$:

$$h(x, y, 0) = 1 + \frac{1}{2} \exp\left[-\left(\frac{4x}{5}\right)^2 - \left(\frac{4y}{5}\right)^2\right], \qquad u(x, y, 0) = 0, \qquad v(x, y, 0) = 0.$$

The inertia-gravity waves propagate after the collapse of the initial symmetric peak with respect to the axes. Nonlinear interactions create shorter waves, propagating around the domain, and the interactions construct more complicated patterns [33].

4.1.1 Non-parametric case

We consider the NTSWE for a fixed parameter $\mu = \frac{\pi}{4}$. The snapshots corresponding to the FOM are collected by solving (1) in the time domain $[0, T]$ with $T = 60$. The discrete state vectors are concatenated into $\mathbf{w} = [\mathbf{u}, \mathbf{v}, \mathbf{h}]^T \in \mathbb{R}^{30000}$, leading to a training set of the size $\mathbf{S}_{\text{FOM}} \in \mathbb{R}^{30000 \times 601}$. The inferred non-intrusive ROM and the intrusive ROM are then used to predict the height field outside of the training set at time $T = 80$.

The decay of the leading 300 normalized singular values of the snapshot matrix $\mathbf{S}_{\text{FOM}}$ is shown in Figure 2. The slow decay of the normalized singular values indicates the difficulty of obtaining accurate ROMs for a small number of POD modes, which is a common problem for hyperbolic PDEs, e.g., like the NTSWE.
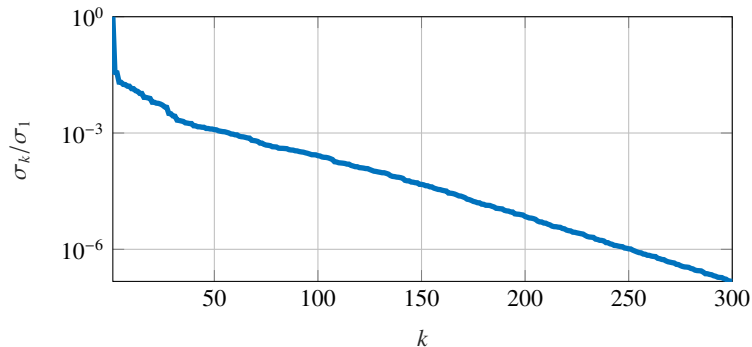

Figure 2: Single-layer geostrophic adjustment: Normalized singular values.

In Figure 3 , we demonstrate the L-curves for the basis sizes $r = 20$, tQR tolerances $\texttt{tol} = [10^{-4}, 10^{-5}, \ldots, 10^{-10}]$ and Tikhonov parameters $\lambda = [10^1, 10^0, \ldots, 10^{-7},]$. The vertical axis shows the squared norm of the learned operators, and the horizontal axis shows the least-squares residual. The computation of the L-curves for each dimension is costly therefore the Tikhonov regularization parameter λ and the tQR tolerance $\texttt{tol}$ are chosen close to the upper part of the corners at the L-curves for $r = 20$ which produce stable solutions from dimension $r = 20$ up to dimension $r = 75$. The chosen tQR tolerance and Tikhonov parameter λ are shown as blue dot in Figure 3.

| (a) Tikhonov | (b) tQR |
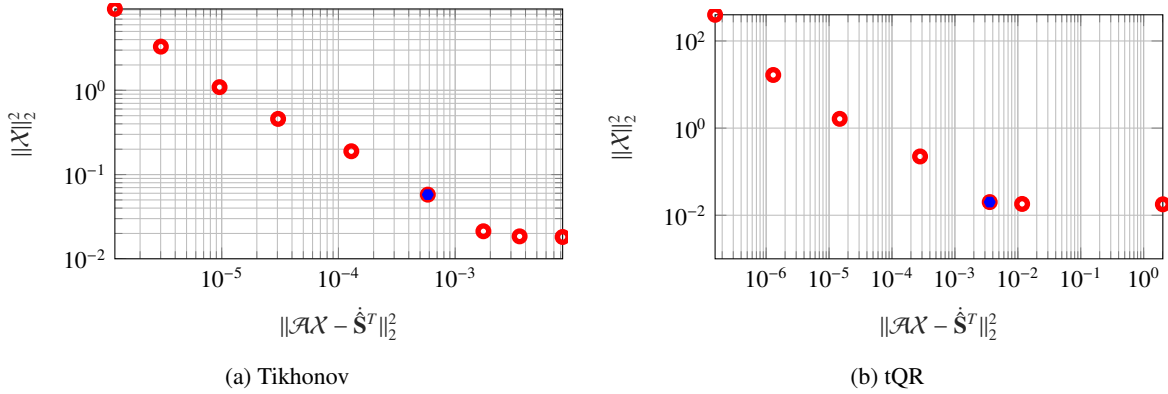
Figure 3: L-curves: Tikhonov (left) , tQR (right).

Next, we build a POD-projection matrix $\mathbf{V} \in \mathbb{R}^{30000 \times r}$ with $r = 75$, which leads to a projection error (22) $\mathcal{E}_{\text{proj}} = 2.07 \cdot 10^{-3}$. Then, we construct ROMs using the intrusive POD and non-intrusive OpInf methods. The non-intrusive methods regularized with both Tikhonov and tQR-based regularizers for comparison. To examine the accuracy of these ROMs, we perform time-domain simulations and compare them with the FOM. In Table 1, the accuracy of the intrusive and non-intrusive ROMs are compared using the FOM-ROM error (21). Table 1 indicates that the non-intrusive method yields a better ROM as compared to the intrusive POD method, and for this example, both regularizations apparently yield a similar result.

Method	POD	Non-intrusive (Tikhonov regularizer)	Non-intrusive (tQR)
$\mathcal{E}$	$3.27 \cdot 10^{-3}$	$2.07 \cdot 10^{-3}$	$2.07 \cdot 10^{-3}$

Table 1: Single-layer geostrophic adjustment: Comparison of intrusive and non-intrusive ROMs.

Moreover, we study the quality of the ROMs as the ROM order increases. For this, we compare the relative error $\mathcal{E}$ (21) obtained using intrusive and non-intrusive methods in Figure 4. We observe that the relative error (21) using the intrusive ROM does not decrease as smoothly as in the case of the non-intrusive case. Furthermore, we observe that both regularizers perform equally well for all orders of the ROMs. However, the penalty parameter of the Tikhonov regularization has to be determined by the L-curve, which requires the SVD computation of the data matrix [18]. Thus, the Tikhonov regularization, combined with L-curve information, is more expensive than the tQR.
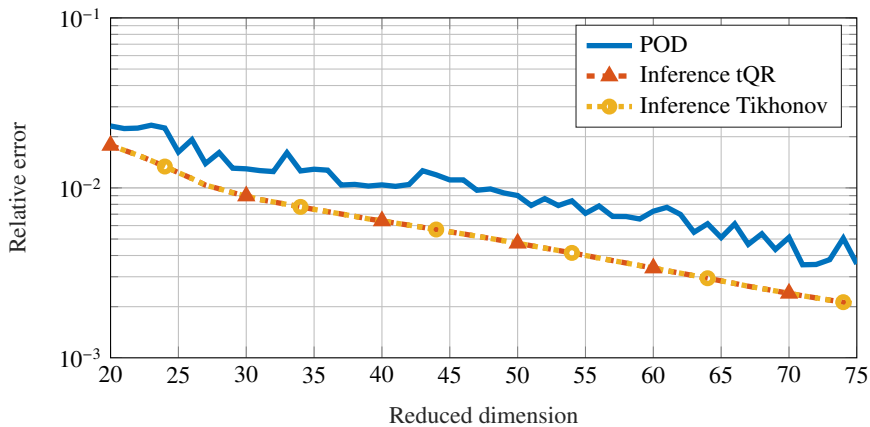

Figure 4: Single-layer geostrophic adjustment: Relative ROMs errors.

In the rest of this paper, for time-domain simulations and prediction, we show the results for the OpInf with tQR since both the Tikhonov and tQR-based methods yield comparable results with respect to the projection error (22). The height field of the FOM and ROMs of order $r = 75$ at time $T = 60$ are shown in Figure 5. These figures show that both ROMs are very close to the FOM solutions, with the non-intrusive ROM being slightly more accurate.

Furthermore, we discuss the prediction capabilities of the obtained ROMs of order $r = 75$. For this, we predict the height field at time $T = 80$ in Figure 6. Note that we have trained the model using the data up to $T = 60$. It indicates that the height field can be predicted with significantly higher accuracy using the non-intrusive ROM as compared to the intrusive POD model.
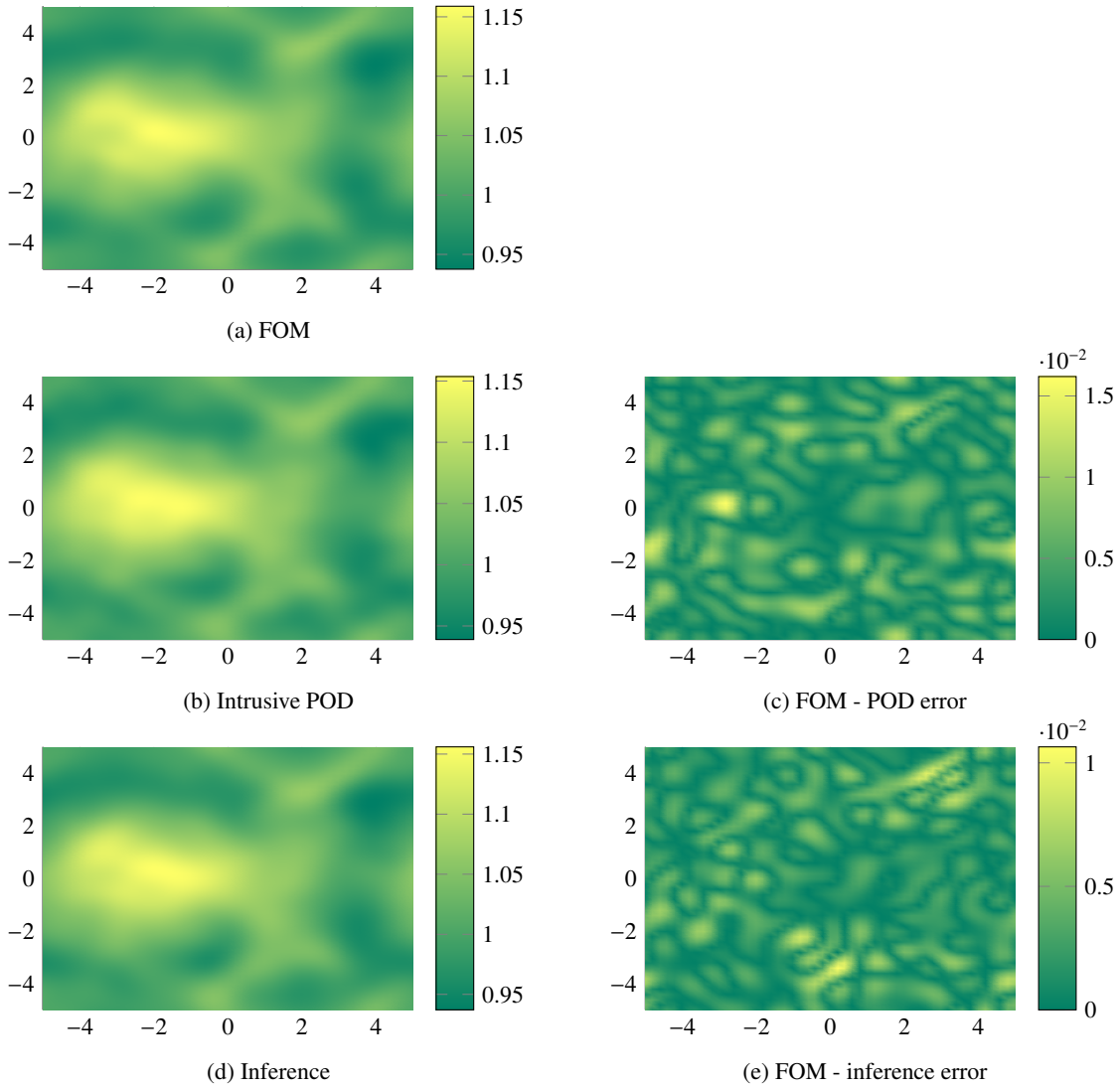

(a) FOM

(b) Intrusive POD

(c) FOM - POD error

(d) Inference

(e) FOM - inference error

Figure 5: Single-layer geostrophic adjustment: Comparison of the height field obtained using the FOM and ROMs of order $r = 75$ at time $T = 60$.

(a) FOM

(b) Intrusive POD

(c) FOM - POD error

(d) Inference

(e) FOM - inference error
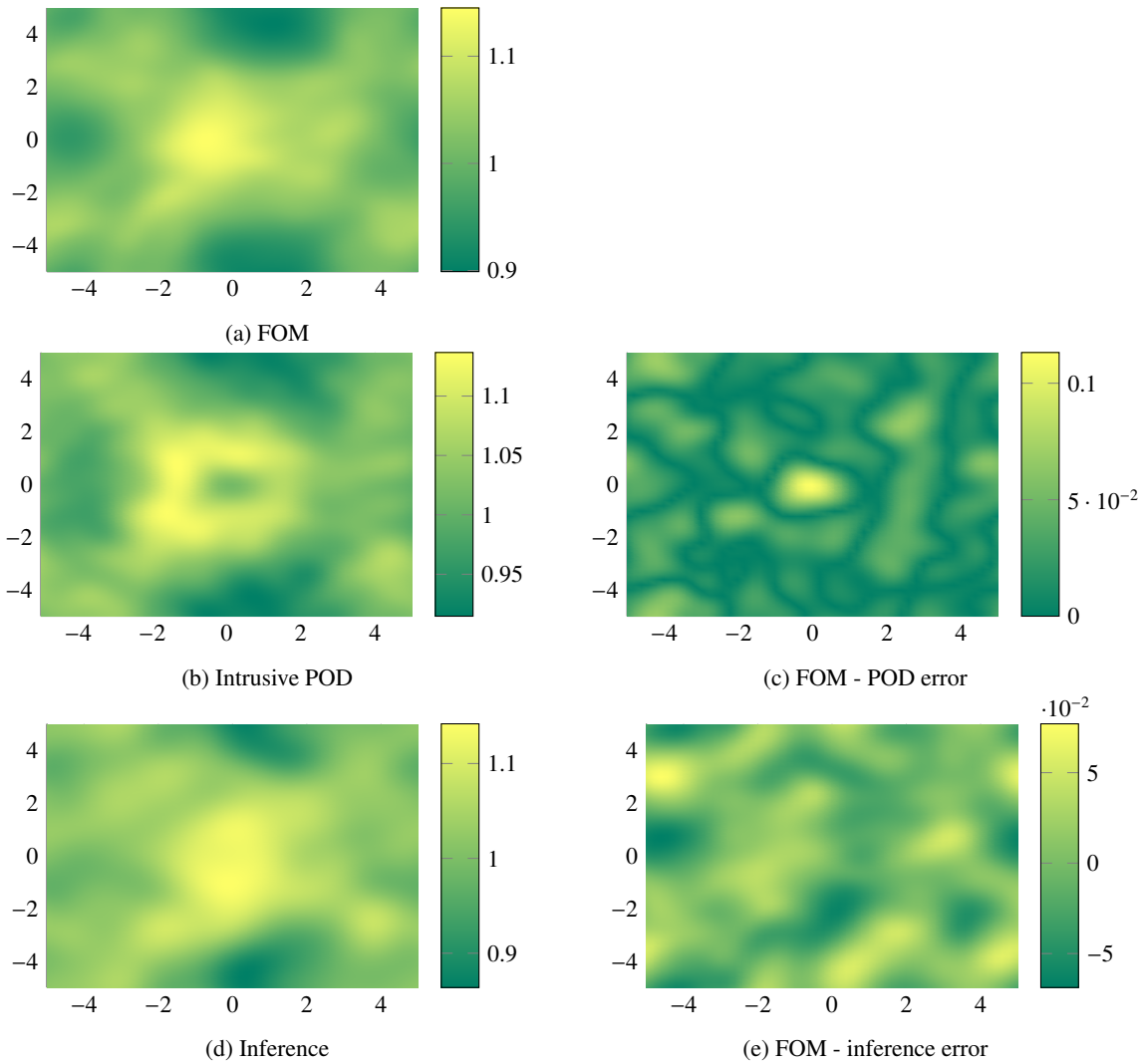
Figure 6: Single-layer geostrophic adjustment: Prediction of the height field obtained using the FOM and ROMs of order $r = 75$ at time $T = 80$.

4.1.2 Parametric Case

In this example, we set the parameter domain of NTSWE as $\mathcal{D} = \left[\frac{\pi}{6}, \frac{\pi}{3}\right] \subset \mathbb{R}$. We consider the time domain $[0, T] \subset R$ with final time $T = 10$. The trajectories are generated with $M = 5$ equidistantly distributed parameters $\mu_1, \mu_2, \ldots, \mu_5 \in \mathcal{D}$. The concatenated snapshot matrix $\mathbf{S}_{\text{FOM}} \in \mathbb{R}^{30000 \times 505}$ is constructed as in (10).

The normalized singular values in Figure 7 decay similar to the non-parametric case in Figure 2. To determine the accuracy of the non-intrusive ROMs, we compute the projection error (22) in the training set, which is $\mathcal{E}_{\text{proj}} = 2.10 \cdot 10^{-4}$ for $r = 75$. In Table 2, the relative errors (21) are very close to the projection error (22) for the ROMs of order $r = 75$, which indicates that non-intrusive ROM solutions of both methods have the same level of accuracy. For the parametric case, we present only the results of the non-intrusive approach; however, we observe a similar behavior as in the non-parametric case when non-intrusive and intrusive methods are compared. We also demonstrate the relative errors for the non-intrusive ROMs of orders $r = 25$ to $r = 75$ over the training set in Figure 8. Again, we observe the same behavior as for the non-intrusive model; when the order of the ROM increases, the relative errors in the training set decrease.
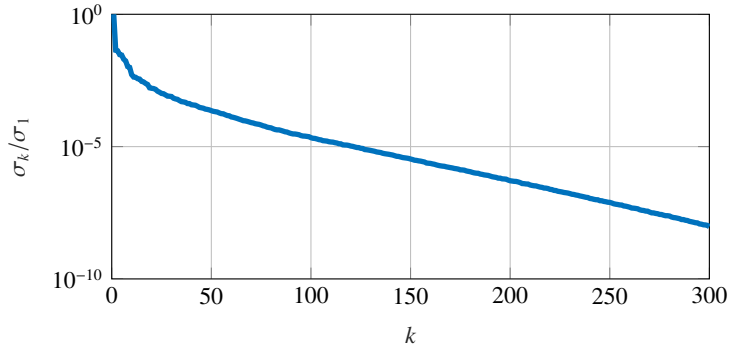
Figure 7: Single-layer geostrophic adjustment: Normalized singular values.

Method	Non-intrusive (Tikhonov regularizer)	Non-intrusive (tQR)
$\mathcal{E}$	$2.10 \cdot 10^{-4}$	$3.32 \cdot 10^{-4}$

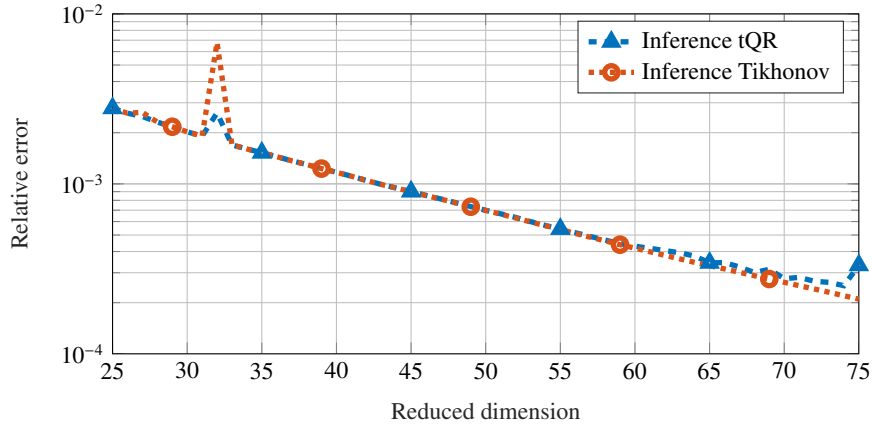Table 2: Single-layer geostrophic adjustment: Comparison of non-intrusive ROMs.

Figure 8: Single-layer geostrophic adjustment: Relative ROMs errors.

Next, we examine the performance of the parametric non-intrusive ROM on the test set, which consists of the parameters at the midpoint of two successive training parameters. In Figure 9, we demonstrate the relative errors (21) of the non-intrusive ROM of orders $r = 25$ and $r = 75$ for both test and training parameters. Figure 9 shows that the accuracy of the non-intrusive ROM increases when the reduced dimension increases for training parameters as well as for test parameters.
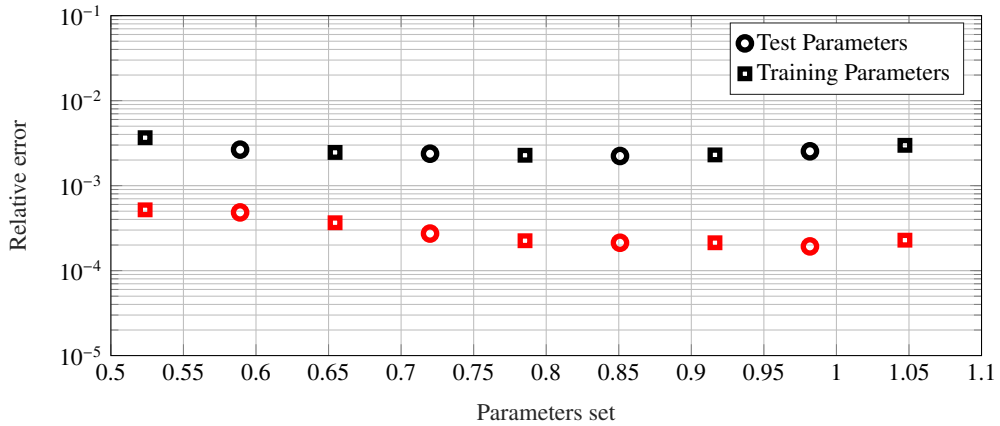

Figure 9: Single-layer geostrophic adjustment: Relative error for testing and training parameters; (square): training set, (circle): testing set. (black): reduced dimension $r = 25$, (red): reduced dimension $r = 75$.

Finally, we show the height field at time $T = 10$ for the parameter $\mu = \frac{5\pi}{24}$ and the corresponding absolute errors in Figure 10 for the non-intrusive ROM of order $r = 75$ which shows that the non-intrusive ROM captures the dynamics of the FOM very well.

(a) FOM

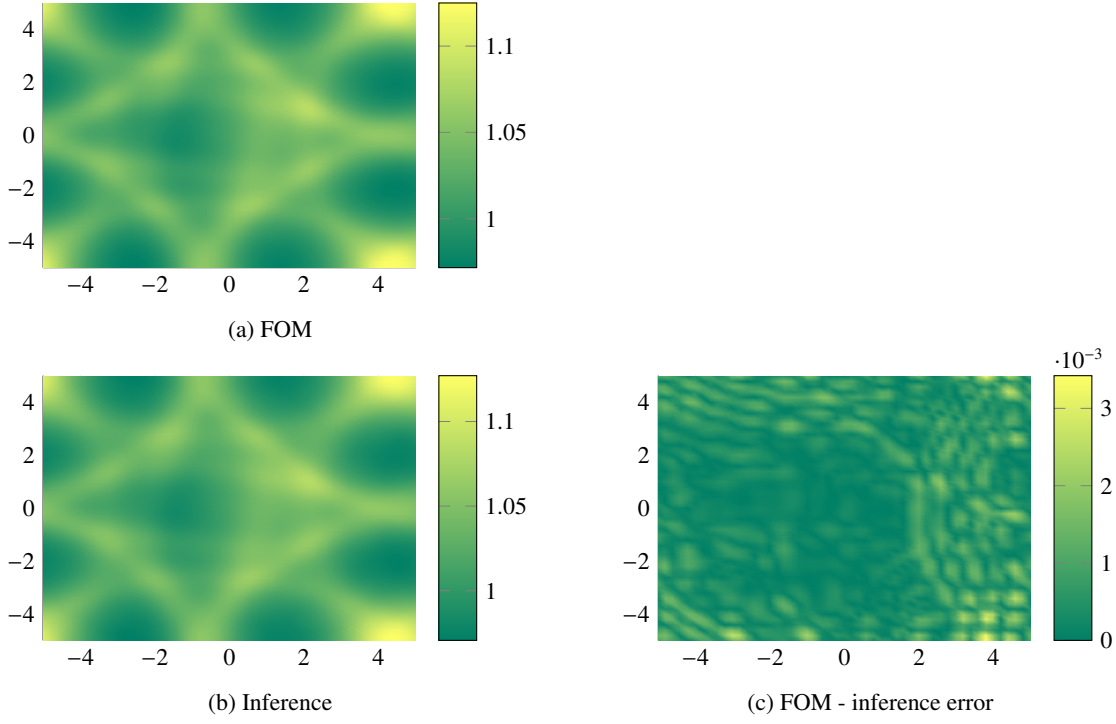(b) Inference

(c) FOM - inference error

Figure 10: Single-layer geostrophic adjustment: Comparison of the height field for the parameter $\mu = \frac{5\pi}{24}$ obtained using the FOM and non-intrusive ROM of order $r = 75$ at time $T = 10$.

4.2 Single-layer shear instability

The initial conditions for the second test example on the periodic domain $[0, 10] \times [0, 10]$ are given as:

$$h(x, y, 0) = 1 + \Delta h \sin\left\{\frac{2\pi}{L}\left[y - \Delta y \sin\left(\frac{2\pi x}{L}\right)\right]\right\},$$

$$u(x, y, 0) = -\frac{2\pi \Delta h}{\Omega^z L} \cos\left\{\frac{2\pi}{L}\left[y - \Delta y \sin\left(\frac{2\pi x}{L}\right)\right]\right\},$$

$$v(x, y, 0) = -\frac{4\pi^2 \Delta h \Delta y}{\Omega^z L^2} \cos\left\{\frac{2\pi}{L}\left[y - \Delta y \sin\left(\frac{2\pi x}{L}\right)\right]\right\} \cos\left(\frac{2\pi x}{L}\right),$$

where $\Delta h = 0.2$, $\Delta y = 0.5$ and the dimensionless spatial domain length $L = 10$. This test example illustrates the roll-up of an unstable shear layer [33].

4.2.1 Non-parametric case

The performance of the OpInf method is shown in terms of learning the vorticity dynamics for the parameter $\mu = \frac{\pi}{4}$. We collect the snapshots as the discrete state vectors concatenated into $\mathbf{w} = [\mathbf{u}, \mathbf{v}, \mathbf{h}]^T \in \mathbb{R}^{30000}$. We perform simulations of the FOM (19) in the time domain $[0, 60]$, which yields the training set of the size $\mathbf{S}_{\text{FOM}} \in \mathbb{R}^{30000 \times 601}$. The quality of the ROMs in terms of relative errors (21) are shown in Figure 11 , which shows that the relative errors of the non-intrusive ROMs decrease smoothly with decreasing singular values in Figure 12, whereas the non-intrusive ROM errors decrease non-smoothly.
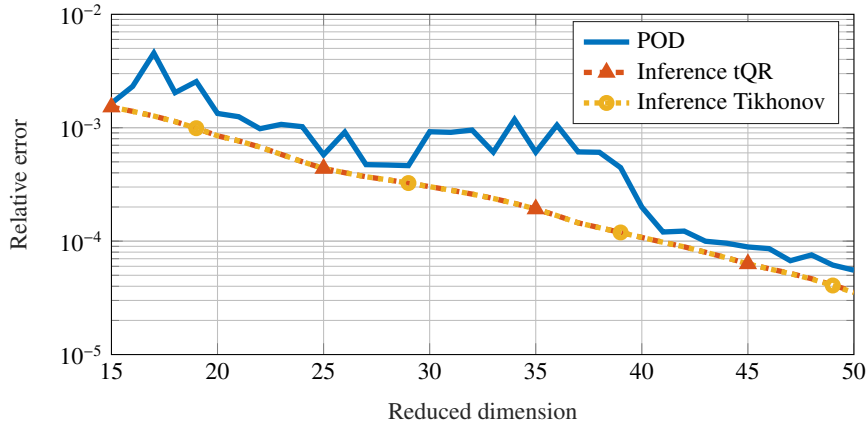
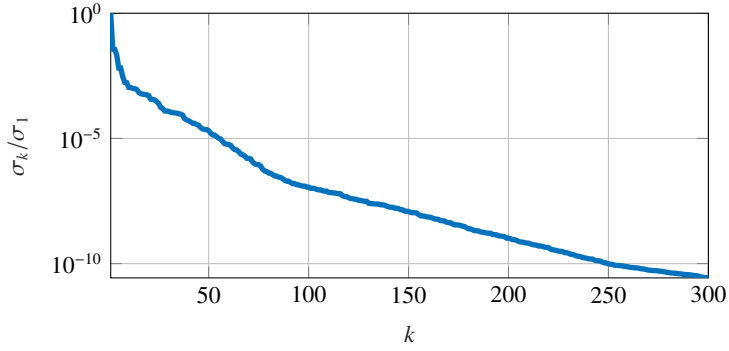Figure 11: Single-layer shear instability: Relative ROMs errors.

Figure 12: Single-layer shear instability: Normalized singular values

The projection error (22) for $r = 50$ yields, $\mathcal{E}_{\mathrm{proj}} = 3.52 \cdot 10^{-5}$. Next, we compare the projection error with relative errors (21) of ROMs of order $r = 50$ in Table 3, which again indicates that the non-intrusive ROMs are more accurate.

Method	POD	Non-intrusive (Tikhonov regularizer)	Non-intrusive (tQR)
$\mathcal{E}$	$5.54 \cdot 10^{-5}$	$3.52 \cdot 10^{-5}$	$3.58 \cdot 10^{-5}$

Table 3: Single-layer shear instability: Comparison of intrusive and non-intrusive ROMs.

The potential vorticities of the FOM and ROMs of order $r = 50$ as well as corresponding absolute error are shown in Figure 13, where both the FOM and ROMs share similar roll-up behavior in the vorticity dynamics. In Figures 13c, 13e, we observe that the accuracy of intrusive and non-intrusive ROMs of order $r = 50$ is similar for the vorticity dynamics.

(a) FOM

(b) Intrusive POD

(c) FOM - POD error

(d) Inference
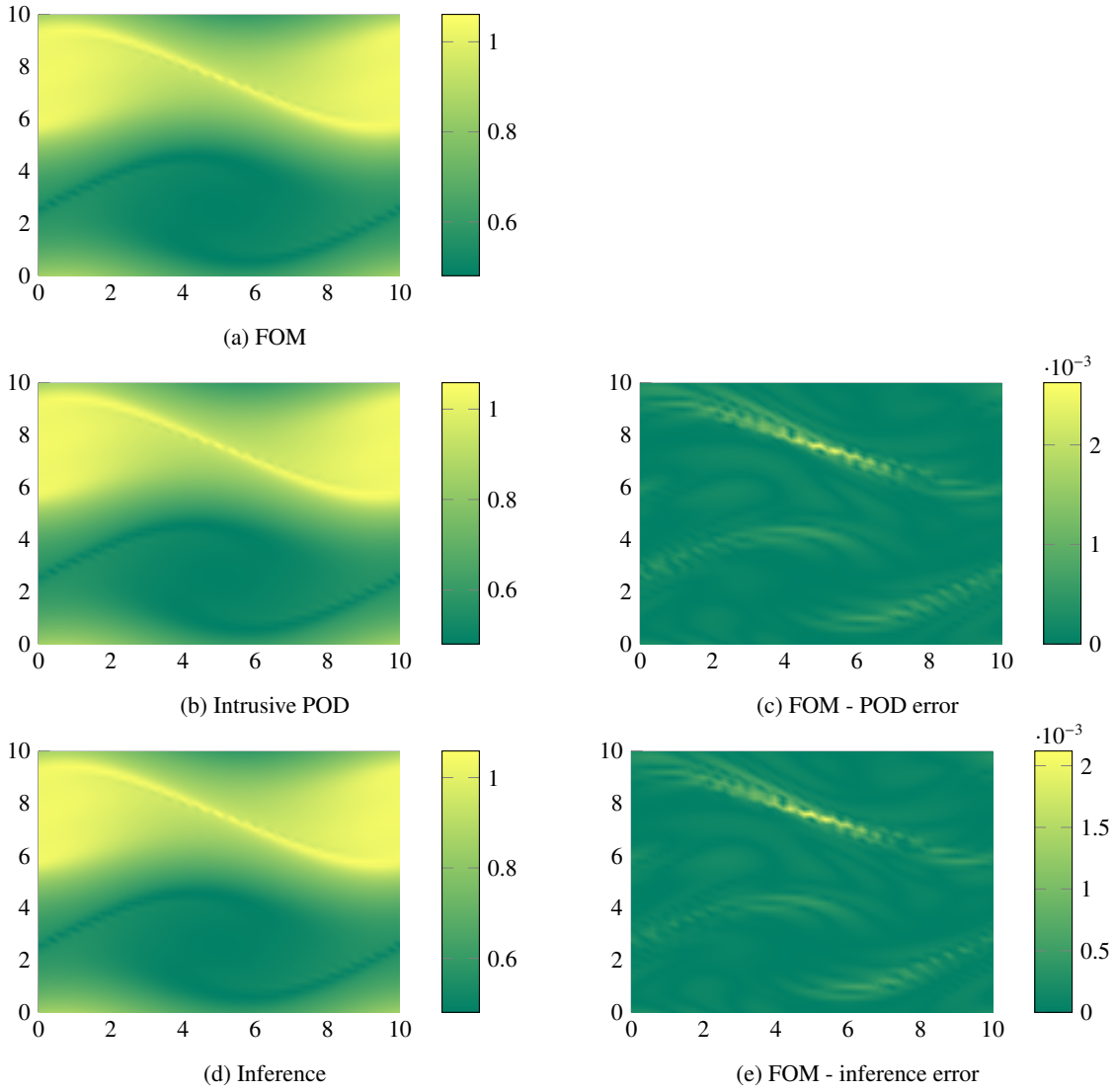
(e) FOM - inference error

Figure 13: Single-layer shear instability: Comparison of the potential vorticity field obtained using the FOM and ROMs of order $r = 50$ at time $T = 60$.

In Figure 14, we demonstrate the prediction capability of the ROMs of order $r = 50$ obtained via intrusive POD and non-intrusive OpInf methods by training them in the time interval $[0, 60]$. We set the final time for the prediction as $T = 80$. Figure 14 shows that the non-intrusive OpInf solutions are more accurate than the intrusive POD solutions.

(a) FOM

(b) POD

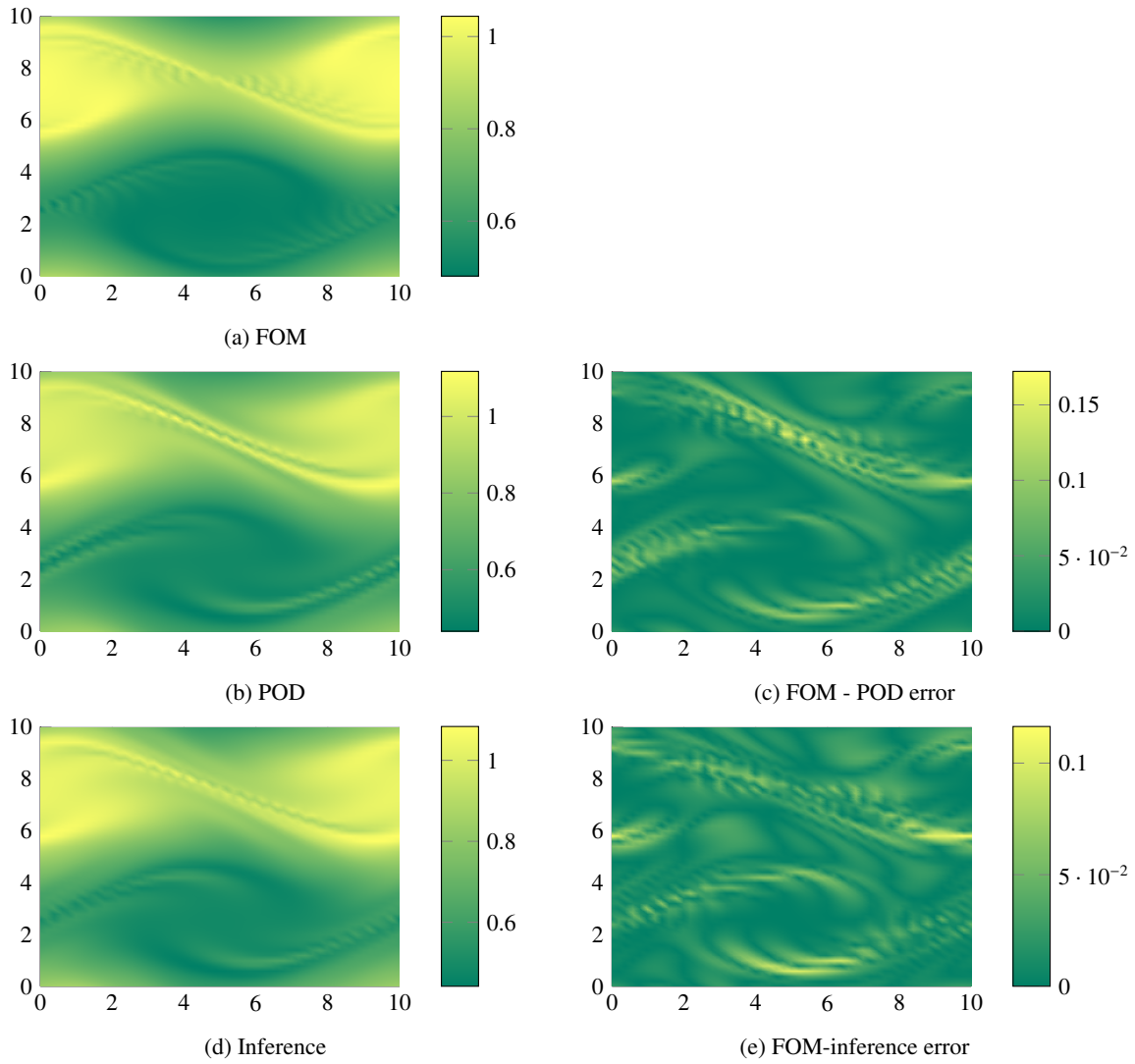(c) FOM - POD error

(d) Inference

(e) FOM-inference error

Figure 14: Single-layer shear instability: Prediction of the potential vorticity field obtained using the FOM and ROMs of order $r = 50$ at time $T = 80$.

4.2.2 Parametric Case

In the last example, we examine the performance of the ROMs for the vorticity dynamics by setting the parameter domain of NTSWE as $\mathcal{D} = \left[\frac{\pi}{6}, \frac{\pi}{3}\right] \subset \mathbb{R}$. In this case, the initial condition is dependent on the angular velocity vector Ω^z. The trajectories in the training set are constructed simulating NTSWE (19) on the time domain $[0, T]$ with final time $T = 30$ and $M = 5$ equidistantly distributed parameters $\mu_1, \mu_2, \ldots, \mu_5 \in \mathcal{D}$. The total size of the concatenated snapshot matrix is $\mathbf{S}_{\text{FOM}} \in \mathbb{R}^{30000 \times 1505}$.

The relative errors (21) of the non-intrusive ROMs of orders 20 to 65 over the training set shown in Figure 15, are smoothly decreasing when the normalized singular values decrease as shown in Figure 16 as in the first test example. The projection error (22) of the FOM for $r = 65$ is $\mathcal{E} = 6.12 \cdot 10^{-5}$. Next, we compare the projection error with the relative error (21) of the non-intrusive ROMs of order $r = 65$ in Table 4 , which shows that both regularizers provide accurate simulations of the same order.
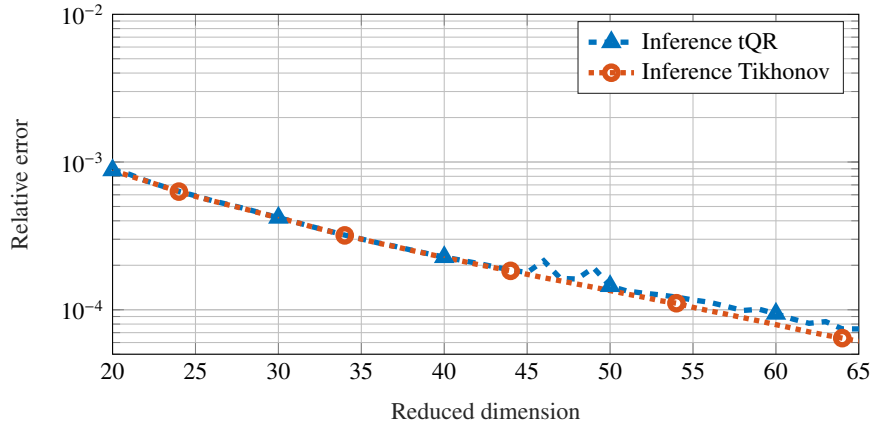
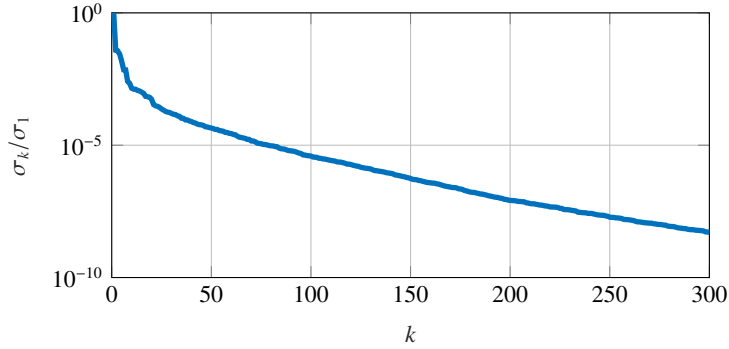Figure 15: Single-layer shear instability: Relative ROMs errors.

Figure 16: Single-layer shear instability: Normalized singular values

Method	Non-intrusive (Tikhonov regularizer)	Non-intrusive (tQR)
$\mathcal{E}$	$6.12 \cdot 10^{-5}$	$7.41 \cdot 10^{-5}$

Table 4: Single-layer shear instability (parametric case): Comparison of non-intrusive ROMs.

We again consider the test set consisting of the parameters at the middle points of two successive training parameters. Figure 17 shows the relative errors for parameters in the test and training sets for non-intrusive ROMs of order 20 and 65. In Figure 17, the inferred solution for the test parameter $\mu = \frac{9\pi}{48}$ is less accurate than at other parameters in the test set for reduced dimension 65. This indicates that the truncation tolerance for the QR-CP method can deteriorate the accuracy of the ROMs.
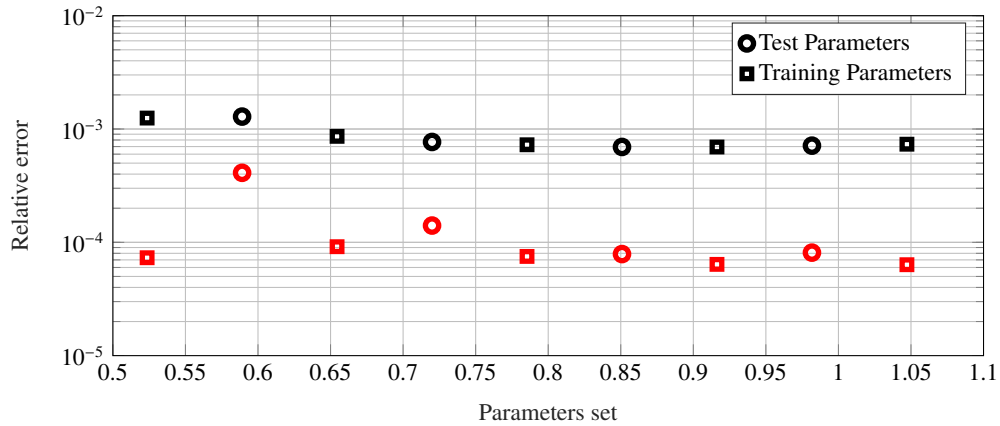

Figure 17: Single-layer shear instability: Relative error for testing and training parameters; (square):training set, (circle):testing set. (black): reduced dimension 20, (red): reduced dimension 65.

Finally, we show the potential vorticity of the FOM, the non-intrusive ROM of order $r = 65$ and the corresponding

15

absolute error at time $T = 30$ for the parameter $\mu = \frac{5\pi}{24}$ in Figure 18, which shows that the non-intrusive ROM captures the dynamics of the NTSWE accurately.
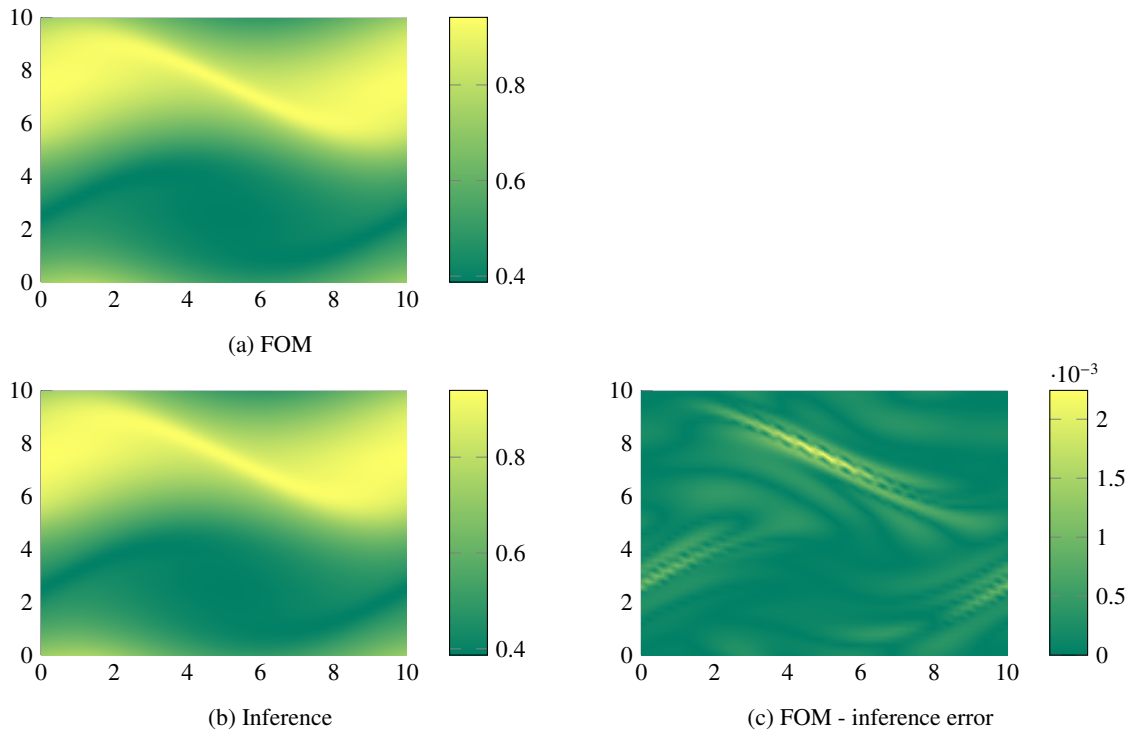

(a) FOM

(b) Inference

(c) FOM - inference error

Figure 18: Single-layer shear instability: Comparison of the potential vorticity field for the parameter $\mu = \frac{5\pi}{24}$ obtained using the FOM and non-intrusive ROM of order $r = 65$ at time $T = 30$.

5 Conclusions

We have constructed data-driven projection-based ROMs of NTSWE by exploiting the structure of the equations. The OpInf framework is used to construct the non-intrusive ROM. Since the least-squares problem of the OpInf method may suffer from ill-conditioning, the solutions are regularized using the QR factorization as an alternative to Tikhonov regularization. The performance of the inferred models is examined in terms of prediction capabilities on two test examples. Numerical results show that the prediction of the learned model of OpInf is more accurate than the intrusive POD.

Acknowledgments

References

[1] P. Benner and P. Goyal. Interpolation-based model order reduction for polynomial parametric systems. arXiv:1904.11891, 2019.

[2] P. Benner, P. Goyal, and S. Gugercin. $\mathcal{H}_2$-quasi-optimal model order reduction for quadratic-bilinear control systems. *SIAM Journal on Matrix Analysis and Applications*, 39(2):983–1032, 2018.

[3] P. Benner, P. Goyal, B. Kramer, P. Peherstorfer, and K. Willcox. Operator inference for non-intrusive model reduction of systems with non-polynomial nonlinear terms. arXiv:2002.09726, 2020.

[4] Peter Benner and Tobias Breiten. Two-sided projection methods for nonlinear model order reduction. *SIAM Journal on Scientific Computing*, 37(2):B239–B260, 2015.

[5] Peter Benner, Albert Cohen, Mario Ohlberger, and Karen Willcox, editors. *Model Reduction and Approximation*, volume 15 of *Computational Science & Engineering*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2017.

[6] D. A. Bistrian and I. M. Navon. An improved algorithm for the shallow water equations model reduction: Dynamic Mode Decomposition vs POD. *International Journal for Numerical Methods in Fluids*, 78(9):552–580, 2015.

[7] Diana Alina Bistrian and Ionel Michael Navon. Randomized dynamic mode decomposition for nonintrusive reduced order modelling. *International Journal for Numerical Methods in Engineering*, 112(1):3–25, 2017.

[8] E Boss, N Paldor, and L Thompson. Stability of a potential vorticity front: from quasi-geostrophy to shallow water. *Journal of Fluid Mechanics*, 315:65–84, 1996.

[9] Daniela Calvetti, Per Christian Hansen, and Lothar Reichel. L-curve curvature bounds via Lanczos bidiagonalization. *Electron. Trans. Numer. Anal*, 14:20–35, 2002.

[10] Tony F Chan and Per Christian Hansen. Some applications of the rank revealing QR factorization. *SIAM Journal on Scientific and Statistical Computing*, 13(3):727–741, 1992.

[11] Colin J Cotter and Jemma Shipton. Mixed finite elements for numerical weather prediction. *Journal of Computational Physics*, 231(21):7076–7091, 2012.

[12] Yann-Hervé De Roeck. Sparse linear algebra and geophysical migration: A review of direct and iterative methods. *Numerical Algorithms*, 29(4):283–322, 2002. Regularization with sparse and structured matrices.

[13] Paul J. Dellar and Rick Salmon. Shallow water equations with a complete Coriolis force and topography. *Physics of Fluids*, 17(10):106601, 2005.

[14] Carl Eckart. Hydrodynamics of oceans and atmospheres, 1960. *Dt. hydrogr. Z.*, 13:197–199, 1960.

[15] Vahid Esfahanian and Khosro Ashrafi. Equation-free/Galerkin-free reduced-order modeling of the shallow water equations based on Proper Orthogonal Decomposition. *Journal of Fluids Engineering*, 131(7):071401–071401–13, 2009.

[16] T Gerkema, JTF Zimmerman, LRM Maas, and H Van Haren. Geophysical and astrophysical fluid dynamics beyond the traditional approximation. *Reviews of Geophysics*, 46(2), 2008.

[17] Theo Gerkema and Victor I Shrira. Near-inertial waves in the ocean: beyond the traditional approximation. *Journal of Fluid Mechanics*, 529:195–219, 2005.

[18] PC Hansen. *The L-curve and its use in the numerical treatment of inverse problems*, volume 5, page 119. WIT press Holland, 2001.

[19] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359 – 366, 1989.

[20] S Leibovich and SK Lele. The influence of the horizontal component of Earth's angular velocity on the instability of the Ekman layer. *Journal of Fluid Mechanics*, 150:41–87, 1985.

[21] Alexander Lozovskiy, Matthew Farthing, and Chris Kees. Evaluation of Galerkin and Petrov-Galerkin model reduction for finite element approximations of the shallow water equations. *Computer Methods in Applied Mechanics and Engineering*, 318:537 – 571, 2017.

[22] Alexander Lozovskiy, Matthew Farthing, Chris Kees, and Eduardo Gildin. POD-based model reduction for stabilized finite element approximations of shallow water flows. *Journal of Computational and Applied Mathematics*, 302:50 – 70, 2016.

[23] John Marshall and Friedrich Schott. Open-ocean convection: Observations, theory, and models. *Reviews of Geophysics*, 37(1):1–64, 1999.

[24] Benjamin Peherstorfer and Karen Willcox. Data-driven operator inference for nonintrusive projection-based model reduction. *Computer Methods in Applied Mechanics and Engineering*, 306:196 – 215, 2016.

[25] E. Qian, B. Kramer, B. Peherstorfer, and K. Willcox. Lift & learn: Physics-informed machine learning for large-scale nonlinear dynamical systems. *Physica D: Nonlinear Phenomena*, 406:132401, 2020.

[26] Elizabeth Qian, Boris Kramer, Alexandre Marques, and Karen Willcox. Transform & learn: A data-driven approach to nonlinear model reduction. In *2019 AIAA Aviation and Aeronautics Forum and Exposition, June 17-21, Dallas, TX*, 2019.

[27] Alfio Quarteroni and Gianluigi Rozza. *Reduced Order Methods for Modeling and Computational Reduction*, volume 9 of *MS&A*. Springer, Milano, 1 edition, 2014.

[28] Samuel. Rudy, Alessandro. Alla, Steven L. Brunton, and J. Nathan. Kutz. Data-driven identification of parametric partial differential equations. *SIAM Journal on Applied Dynamical Systems*, 18(2):643–660, 2019.

[29] Răzvan Ştefănescu, Adrian Sandu, and Ionel M Navon. Comparison of POD reduced order strategies for the nonlinear 2D shallow water equations. *International Journal for Numerical Methods in Fluids*, 76(8):497–521, 2014.

[30] Andrew L. Stewart and Paul J. Dellar. Multilayer shallow water equations with complete Coriolis force. Part 1. Derivation on a non-traditional beta-plane. *Journal of Fluid Mechanics*, 651:387413, 2010.

[31] Andrew L Stewart and Paul J Dellar. Multilayer shallow water equations with complete coriolis force. Part 2. Linear plane waves. *Journal of Fluid Mechanics*, 690:16–50, 2012.

[32] Andrew L Stewart and Paul J Dellar. Multilayer shallow water equations with complete coriolis force. Part 3. Hyperbolicity and stability under shear. *Journal of Fluid Mechanics*, 723:289–317, 2013.

[33] Andrew L. Stewart and Paul J. Dellar. An energy and potential enstrophy conserving numerical scheme for the multilayer shallow water equations with complete Coriolis force. *Journal of Computational Physics*, 313:99 – 120, 2016.

[34] Renee Swischuk, Boris Kramer, Cheng Huang, and Karen Willcox. Learning physics-based reduced-order models for a single-injector combustion process. *AIAA Journal*, 58(6):2658–2672, 2020.

[35] Renee Swischuk, Laura Mainini, Benjamin Peherstorfer, and Karen Willcox. Projection-based model reduction: Formulations for physics-based machine learning. *Computers & Fluids*, 179:704–717, 2019.

[36] Andrey N Tikhonov and Vasilii Iakkovlevich Arsenin. *Solutions of ill-posed problems*, volume 14. Winston, Washington, DC, 1977.

[37] Geoffrey K Vallis. *Atmospheric and Oceanic Fluid Dynamics*. Cambridge University Press, 2017.

[38] Emma S Warneford and Paul J Dellar. Thermal shallow water models of geostrophic turbulence in jovian atmospheres. *Physics of Fluids*, 26(1):016603, 2014.