

Non-intrusive reduced order modelling with least squares fitting on a sparse grid

Z. Lin¹, D. Xiao^{2,3,*},†, F. Fang², C. C. Pain² and Ionel M. Navon⁴

¹*School of Mathematical Sciences, Zhejiang University, No. 38 Zheda Road, Hangzhou 310027, China*

²*Novel Reservoir Modelling and Simulation Group, Department of Earth Science and Engineering, Imperial College London, Prince Consort Road, London SW7 2AZ, UK*

³*China University of Geosciences, Wuhan 430074, China*

⁴*Department of Scientific Computing, Florida State University, Tallahassee, FL 32306-4120, USA*

SUMMARY

This paper presents a non-intrusive reduced order model for general, dynamic partial differential equations. Based upon proper orthogonal decomposition (POD) and Smolyak sparse grid collocation, the method first projects the unknowns with full space and time coordinates onto a reduced POD basis. Then we introduce a new least squares fitting procedure to approximate the dynamical transition of the POD coefficients between subsequent time steps, taking only a set of full model solution snapshots as the training data during the construction. Thus, neither the physical details nor further numerical simulations of the original PDE model are required by this methodology, and the level of non-intrusiveness is improved compared with existing reduced order models. Furthermore, we take adaptive measures to address the instability issue arising from reduced order iterations of the POD coefficients. This model can be applied to a wide range of physical and engineering scenarios, and we test it on a couple of problems in fluid dynamics. It is demonstrated that this reduced order approach captures the dominant features of the high-fidelity models with reasonable accuracy while the computation complexity is reduced by several orders of magnitude. Copyright © 2016 John Wiley & Sons, Ltd.

Received 14 December 2015; Revised 27 May 2016; Accepted 1 June 2016

KEY WORDS: non-intrusive; least squares fitting; POD; Smolyak sparse grid

1. INTRODUCTION

In many areas of science and engineering, iterative computations and data assimilations for large-scale dynamical systems are often required to understand, to predict and/or to control various phenomena. Straightforward simulations of such problems may be very inefficient. Extensive resources are exhausted to produce intermediate results that have little significance owing to different sources of parameter variations, errors and noise. Therefore, reduced order models (ROMs) have become prevalent, thanks to their potential to achieve major speedup for standard numerical procedures [1, 2]. Their applicability relies on the presumption that the predominant physical mechanisms operate on a much lower-dimensional space. A particular class of implementations is based on the proper orthogonal decomposition (POD) method whose variants have been successfully applied in

*Correspondence to: Dunhui Xiao, Novel Reservoir Modelling and Simulation Group, Department of Earth Science and Engineering, Imperial College London, Prince Consort Road, London, UK.

†E-mail: dh.xiao@imperial.ac.uk

a wide spectrum of research fields: air pollution dispersion [3], shallow water equations [4, 5], convective flows [6], ocean modelling [5, 7–11], 4-D variational data assimilation [12–14], neutron transport [15], fluid and structure interaction [16, 17], inversion [18], multiphase flows in porous media [19] and molecular simulation [20].

Traditionally, an ROM implies re-deriving a simpler physical model with additional restrictions such as homogenisation and parametrisation. Furthermore, substantial rewriting of the numerical schemes for the physical model is often required. Such a ROM is therefore termed ‘intrusive’ and has been suffering from instability and nonlinear inefficiency problems [9, 10, 21–26]. In addition, the dependency of the source code results in that the ROM is difficult to construct and implement [27].

To circumvent or avoid these issues, researchers found another way of solving the reduced order system: non-intrusive method. The non-intrusive method became popular in recent years because it is less dependent of the complex original dynamic system and is therefore easy to implement even when the source code is not available. A number of non-intrusive reduced order modelling (NIROM) methods have been proposed. Walton *et al.* and Xiao *et al.* proposed an NIROM using radial basis function (RBF) method and POD [28, 29]. Noack [30] and Noori [31] used a neural network to construct an NIROM. Xiao *et al.* [32] used a Smolyak sparse grid collocation method in which multi-dimensional, vector-valued Smolyak functions are used to replace the original differential equations and to evolve the state variables. This approach not only avoids the access to the high-fidelity code but also circumvents the so-called curse of dimensionality, namely, the exponential growth in computational complexity as the dimension of the problem increases.

In this paper, we present a new non-intrusive reduced order model method based on least square fitting and Smolyak sparse grid. The advantage of this method is achieved by its use of the sparse grid method to lower the order of polynomial fit. As we know, high-order polynomials may cause poorer fittings because they oscillate between sample data points. This combination of Smolyak sparse grid and least square fitting induces a great potential in problems with high number of sample data points (many thousands of data points). Instead of traversing each nodal point in a Smolyak sparse grid with its coordinates as the input to a full model simulation and then interpolating with the collective output, we compute the coefficients of the Smolyak polynomials by solving a least squares fitting problem whose data are recycled from the generation of the POD basis. We first properly decompose the temporally equidistant snapshots of the solution variables generated by the CFD software Fluidity [33]. Then we assemble the input data for the least square fitting by projecting the snapshots onto the reduced basis. Finally, a linear system of normal equations is solved producing the desired fit defined by its function values on a sparse grid. The fit, a finite dimensional map for POD coefficients, then serves as an approximation to the infinite dimensional evolution of physical variables. Essentially, the physics and numerics of the original full model are transparent in constructing the ROM, and it can be readily made into a universal black box that is compatible with arbitrary POD framework.

The structure of the paper is as follows. Section 2 reviews the general methodology of POD followed by previous results on non-intrusive ROMs with the introduction of a Smolyak sparse grid. Section 3 derives the least squares fitting problem and its solution for the non-intrusive modelling. Section 4 presents the stabilisation techniques we develop to address the instability issue we encounter with the direct application of the least squares approximation. Section 5 demonstrates the method’s capabilities by solving two test problems in fluid dynamics. Finally, in Section 6, a summary consisting of the conclusions and discussions of future work is presented.

2. PROPER ORTHOGONAL DECOMPOSITION-BASED REDUCED ORDER MODELLING AND SMOLYAK SPARSE GRIDS

We consider a physical, dynamical system such as the Navier–Stokes equations that takes the general form

$$\frac{\partial \Psi}{\partial t} = F(\Psi, \mathbf{x}, t) \quad (1)$$

along with appropriate initial and boundary conditions. Here, the unknown, vector-valued functional $\Psi(\mathbf{x}, t)$ is defined on a spatial region $\mathbf{x} \in \Omega \subset \mathbb{R}^3$ and on the time interval $t \in [0, T]$. The prescribed, external forcing, F , is in general a complex, integro-differential operator derived from some physical laws.

2.1. Reduction via proper orthogonal decomposition

The solution Ψ to the master equation (1) lives in an infinite-dimensional functional space, and the first step of our reduced order modelling is to project it onto a finite-dimensional subspace via POD. This procedure generates a set of basis functions that is constructed from a collection of snapshots that are taken at a number of time instances of the full model simulation. This basis serves as the axis of a coordinate system that allows one to represent and reconstruct functions in physical space with coefficient vectors.

Without loss of generality, we use the Navier–Stokes equations to illustrate the methodology from now on. In this case, the unknown Ψ consists of three velocity components, (u_x, u_y, u_z) , and a pressure component, p . Given a discrete mesh of \mathcal{N} nodes, the s^{th} snapshot of Ψ then contains four \mathcal{N} -vectors, denoted as $\Psi_s^x, \Psi_s^y, \Psi_s^z$ and Ψ_s^p , respectively. Then we assemble four separate $\mathcal{N} \times S$ matrices as

$$\Psi^x = (\Psi_1^x, \Psi_2^x, \dots, \Psi_S^x), \quad \Psi^y = (\Psi_1^y, \Psi_2^y, \dots, \Psi_S^y) \quad (2)$$

and so on for the other two components, where S is the number of snapshots available. For the sake of simplicity, we will omit the superscripts denoting unknown components from here on because the following procedure is applied to each of the four matrices in an independent and identical manner.

Next, we construct a mean-zero snapshot matrix $\tilde{\Psi}$ with columns

$$\tilde{\Psi}_k = \Psi_k - \frac{1}{S} \sum_{k=1}^S \Psi_k, \quad k = 1, 2, \dots, S \quad (3)$$

to which we then apply a singular value decomposition (SVD) as

$$\tilde{\Psi} = U \Sigma V^T. \quad (4)$$

Here, the matrices $U \in R^{\mathcal{N} \times \mathcal{N}}$ and $V \in R^{S \times S}$ consist of the orthonormal eigenvectors for $\tilde{\Psi} \tilde{\Psi}^T$ and $\tilde{\Psi}^T \tilde{\Psi}$, respectively, and Σ is a diagonal matrix of size $\mathcal{N} \times S$ whose nonzero (positive) entries, arranged in a decreasing order, are the singular values of $\tilde{\Psi}$, and we denote them as $\lambda_k, k = 1, 2, \dots, S$. Then the POD of $\tilde{\Psi}$, namely, a reduced order, orthonormal set of basis functions $\{\phi_k\}_{k=1}^S$ is readily retrieved from the column vectors of U as

$$\phi_k = U_k, \quad k = 1, 2, \dots, S. \quad (5)$$

Moreover, this basis can be optimised by keeping only the first P members that correspond to the largest P singular values, respectively. Note that these vectors are optimal in the sense that no other rank- P set of basis vectors can be closer to the snapshot matrix $\tilde{\Psi}$ measured in the Frobenius norm by the Eckart–Young theorem. This further reduction is especially accurate when there is a ‘scale separation’ among the singular values at λ_P , namely,

$$\frac{\sum_{i=P+1}^S \lambda_i^2}{\sum_{i=1}^S \lambda_i^2} \ll 1 \quad (6)$$

With an optimal POD basis $\{\phi_k\}_{k=1}^P$, any variable ψ on an \mathcal{N} -node mesh can be represented by

$$\psi = \bar{\psi} + \sum_{j=1}^P \alpha_j \phi_j, \quad (7)$$

where $\alpha_j \equiv \alpha_j(t)$ denote the time-varying coefficients of the POD expansion and $\bar{\psi}$ is the time-independent mean of the ensemble of snapshots for the variable ψ , which is analogous to the average

in (3). This series expansion transforms the full physical space to the reduced order space and vice versa.

A standard Galerkin procedure can then be applied for the series expansion (7). Substituting the series into the master equation (1) and following by a first-order finite difference discretisation in time, we finally arrive at an ROM, namely, an iteration scheme for the reduced order POD coefficient at arbitrary time step. Denote the coefficient vector at the n^{th} time step as

$$\boldsymbol{\alpha}^n = (\alpha_1(n), \alpha_2^n, \dots, \alpha_P^n) \tag{8}$$

and the iteration formula is then given by

$$\alpha_k^{n+1} = f_k(\boldsymbol{\alpha}^n), \quad n = 1, 2, \dots \tag{9}$$

plus the starting values α_k^0 for $k = 1, 2, \dots, P$. The central subject in reduced order modelling is therefore to find an efficient way to estimate the $\mathbb{R}^P \rightarrow \mathbb{R}$ functions f_k so one can compute the POD coefficients at arbitrary times and reconstruct the physical variables from them using the expansion (7).

2.2. Non-intrusive reduced order modelling on a Smolyak sparse grid

Xiao *et al.* [32] developed a non-intrusive algorithm to interpolate the transitional functions f_k' s in (9) on a Smolyak sparse grid. In particular, instead of the standard, intrusive approach that requires a major re-derivation of the full model, they approximated each f_k with a multidimensional polynomial interpolant. Additionally, to avoid the curse of dimensionality, which is the exponential growth of interpolation points entailing the increase of the POD basis size P , a Smolyak sparse grid was introduced whose number of nodes is only a polynomial function of its dimension size.

The Smolyak interpolant $\hat{f}_k^{P,\mu}(\mathbf{x})$ for any P -dimensional point $\mathbf{x} = (x_1, x_2, \dots, x_P)$ with approximation level μ can be expanded as a weighted sum of tensor product operators

$$\hat{f}_k^{P,\mu}(\mathbf{x}) = \sum_{|\mathbf{i}|=\max(P, \mu+1)}^{P+\mu} (-1)^{P+\mu-|\mathbf{i}|} \times \binom{P-1}{P+\mu-|\mathbf{i}|} \times (U^{i_1} \otimes \dots \otimes U^{i_P}) \left(\hat{f}_k^{P,\mu} \right) (\mathbf{x}) \tag{10}$$

where the summation index $\mathbf{i} = (i_1, \dots, i_P) \in \mathbb{N}^P$, $|\mathbf{i}| = \sum_{n=1}^P |i_n|$ traverses through the μ different levels of the sparse grid in each dimension and the operator

$$(U^{i_1} \otimes \dots \otimes U^{i_P}) \left(\hat{f}_k^{P,\mu} \right) (\mathbf{x}) = \sum_{j_1=1}^{M(i_1)} \dots \sum_{j_P=1}^{M(i_P)} \left(\hat{f}_k^{P,\mu}(x_{j_1}^{i_1}, \dots, x_{j_P}^{i_P}) \times \prod_{\ell=1}^P \psi_{j_\ell}^{i_\ell}(x_\ell) \right) \tag{11}$$

is also a multidimensional sum weighted by the nodal values of the interpolant. Here, the i_ℓ^{th} level 1D grid in the ℓ^{th} dimension, $\ell = 1, 2, \dots, P$, has the size

$$M(i_\ell) = \begin{cases} 2^{i_\ell} + 1, & i_\ell = 1, 2, \dots, \mu; \\ 1, & i_\ell = 0, \end{cases} \tag{12}$$

and its j_ℓ^{th} node is denoted by $x_{j_\ell}^{i_\ell}$. Finally, the last term on the right-hand side of (11) is the product of one-dimensional, basis interpolating polynomials. For example, in the standard Lagrange form,

$$\psi_j^i(x) = \prod_{\substack{n=1 \\ n \neq j}}^{M(i)} \frac{x - x_n^i}{x_j^i - x_n^i}. \tag{13}$$

It is worth noting that the sparse grid, as the collection of all nodes, can be written as a union of μ nested product grids

$$\mathcal{H}^{P,\mu} = \bigcup_{|\mathbf{i}|=\max(P, \mu+1)}^{P+\mu} \mathcal{G}_i = \bigcup_{|\mathbf{i}|=\max(P, \mu+1)}^{P+\mu} \left(\bigotimes_{\ell=1}^P \left\{ x_j^{i_\ell} \right\}_{j=1}^{M(i_\ell)} \right). \tag{14}$$

The interpolation conditions simply imply

$$\hat{f}_k^{P,\mu}(\mathbf{x}_j^i) = f_k(\mathbf{x}_j^i), \quad k = 1, 2, 3, \dots, P. \quad (15)$$

with which Xiao *et al.* applied the following algorithm to obtain a non-intrusive reduced order model:

- (1) Perform a POD given snapshots of the solution to equation (1);
- (2) Generate a Smolyak sparse grid that covers the POD projections of solution snapshots onto the reduced order space;
- (3) For every node on the sparse grid, reconstruct an initial condition in physical space using the transform (7);
- (4) Evolve the full model for a short time, Δt , with the reconstructed initial conditions and transform the results back to the reduced coefficient space with (7);
- (5) Find the Smolyak interpolants for the ROM (9) using the formulas (10), (11) and (15) with the nodal data computed in the previous step.

3. LEAST SQUARES FITTING ON A SMOLYAK SPARSE GRID

The POD-based non-intrusive procedure circumvents the need to manipulate the full model (1) or its numerical implementation and results in impressive computational economy with the use of Smolyak sparse grids. In this section, we will propose an alternative way to approximate $f_k(\mathbf{x})$, $k = 1, 2, \dots, P$ using a least squares fitting approach that is independent from the full model except for an input set of solution snapshots. Effectively, Steps 3 through 5 of the algorithm outlined at the end of Section 2.2 will be replaced, and the ROM will take form of a least squares fit instead of a Smolyak interpolant.

Now, we seek the optimal fit, denoted by $F_k^{P,\mu}$, $k = 1, \dots, P$, from the family of admissible functions, Ω , namely, all $\mathbb{R}^P \rightarrow \mathbb{R}$ polynomials defined on the same Smolyak sparse grid as in Section 2.2. The least squares condition for data fitting demands that $F_k^{P,\mu}$ minimises the standard L^2 distance in P -dimensional space, that is,

$$\sum_{n=1}^{S-1} \|F_k^{P,\mu}(\boldsymbol{\alpha}^n) - \alpha_k^{n+1}\|_2^2 = \min_{g \in \Omega} \sum_{n=1}^{S-1} \|g(\boldsymbol{\alpha}^n) - \alpha_k^{n+1}\|_2^2. \quad (16)$$

Here, the data set

$$\left\{ \left(\boldsymbol{\alpha}^n = (\alpha_1^n, \dots, \alpha_P^n), \alpha_k^{n+1} = f_k(\boldsymbol{\alpha}^n) \right), n = 1, 2, \dots, S-1 \right\} \quad (17)$$

contains duple pairs of POD projections of the solution at two subsequent time steps calculated by the full model, and they are assumed to be connected by the transition relation (9).

The formulas (10) through (13) demonstrate how any polynomial function $g(\mathbf{x})$ on the sparse grid $\mathcal{H}^{P,\mu}$ is uniquely specified by its values on grid nodes. Further, we reorganise the series (10) and (11) into a single sum as

$$g(\mathbf{x}) = \sum_{\mathbf{x}^s \in \mathcal{H}^{P,\mu}} g(\mathbf{x}^s) w_{\mathbf{x}^s}(\mathbf{x}) \quad (18)$$

where $w_{\mathbf{x}^s}$ is the weighting function associated with the node \mathbf{x}^s . Then we can conveniently define a quadratic cost function by

$$J_k[g] = \sum_{n=1}^{S-1} [g(\boldsymbol{\alpha}^n) - \alpha_k^{n+1}]^2 = \sum_{n=1}^{S-1} \left[\sum_{\mathbf{x}^s \in \mathcal{H}^{P,\mu}} g(\mathbf{x}^s) w_{\mathbf{x}^s}(\boldsymbol{\alpha}^n) - \alpha_k^{n+1} \right]^2. \quad (19)$$

Now setting all the partial derivatives with respect to the nodal values, $\{g(\mathbf{x}^s), \mathbf{x}^s \in \mathcal{H}^{P,\mu}\}$, to be zero leads to a linear system, namely, the *normal equations*, whose solution consists of the nodal values of the optimal fit $F_k^{P,\mu}(\mathbf{x})$. That is,

$$\frac{1}{2} \frac{\partial J_k}{\partial g(\mathbf{x}^*)} \Big|_{g=F_k^{P,\mu}} = \sum_{\mathbf{x}^s \in \mathcal{H}^{P,\mu}} F_k^{P,\mu}(\mathbf{x}^s) \times \left(\sum_{n=1}^{S-1} w_{\mathbf{x}^*}(\alpha^n) w_{\mathbf{x}^s}(\alpha^n) \right) - \sum_{n=1}^{S-1} w_{\mathbf{x}^*}(\alpha^n) \alpha_k^{n+1} = 0 \tag{20}$$

for any node \mathbf{x}^* on the grid $\mathcal{H}^{P,\mu}$. Equivalently, in matrix notation,

$$\begin{bmatrix} \langle w_{\mathbf{x}^1}, w_{\mathbf{x}^1} \rangle_s & \langle w_{\mathbf{x}^1}, w_{\mathbf{x}^2} \rangle_s & \langle w_{\mathbf{x}^1}, w_{\mathbf{x}^3} \rangle_s & \cdots \\ \langle w_{\mathbf{x}^2}, w_{\mathbf{x}^1} \rangle_s & \langle w_{\mathbf{x}^2}, w_{\mathbf{x}^2} \rangle_s & \langle w_{\mathbf{x}^2}, w_{\mathbf{x}^3} \rangle_s & \cdots \\ \langle w_{\mathbf{x}^3}, w_{\mathbf{x}^1} \rangle_s & \langle w_{\mathbf{x}^3}, w_{\mathbf{x}^2} \rangle_s & \langle w_{\mathbf{x}^3}, w_{\mathbf{x}^3} \rangle_s & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \cdot \begin{pmatrix} F_k^{P,\mu}(\mathbf{x}^1) \\ F_k^{P,\mu}(\mathbf{x}^2) \\ F_k^{P,\mu}(\mathbf{x}^3) \\ \vdots \end{pmatrix} = \begin{pmatrix} \langle w_{\mathbf{x}^1}, \mathbb{P}_k^+ \rangle_s \\ \langle w_{\mathbf{x}^2}, \mathbb{P}_k^+ \rangle_s \\ \langle w_{\mathbf{x}^3}, \mathbb{P}_k^+ \rangle_s \\ \vdots \end{pmatrix} \tag{21}$$

in which $\{\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3, \dots\}$ is an arbitrary ordering of the nodes of $\mathcal{H}^{P,\mu}$, the inner product

$$\langle f, g \rangle_s = \sum_{n=1}^{S-1} f(\alpha^n) g(\alpha^n) \tag{22}$$

and the projector \mathbb{P}_k^+ returns the k^{th} component of α^{n+1} . By solving the aforementioned linear system for $k = 1, 2, \dots, P$, we obtain the least square fitting approximations to the ROM (9). It is worth noting that the coefficient matrix in (21) is clearly a Gram matrix that can be decomposed as $\mathbf{W}^T \cdot \mathbf{W}$ where

$$\mathbf{W} = \{W_{ij}\} = \{w_{\mathbf{x}^j}(\alpha^i)\}. \tag{23}$$

When the number of sparse grid nodes is less than that of the snapshots, the matrix \mathbf{W} is rank-full because the weighting functions associated with different nodes are linearly independent to each other. Additionally, because

$$\text{rank}(\mathbf{W}^T \cdot \mathbf{W}) = \text{rank}(\mathbf{W}) \tag{24}$$

the linear system (21) is in fact symmetric positive-definite and can then be solved efficiently with standard numerical techniques such as QR factorisation and successive over-relaxation.

Now, we see the significant advantage of the fitting approach in that the ROM is generated by linear solvers instead of full model simulations as described in Section 2.2, because no exact solution to the master equation (1) is required by the least squares formulation. This separation of the ROM from detailed derivation and simulation of the physics results in an improved level of non-intrusiveness as well as in an enormous gain in computational economy.

4. ITERATION STABILISATION WITH COEFFICIENT DAMPING

As we found out in preliminary results, straightforward iterations of formula (9) using its least squares approximation reconstructed with the solution to the system (21) suffer from instability issues that often arise in reduced order modelling. In particular, depending on specific problems and parameters, some POD coefficients diverge to infinity after a random number of iterations when we use the reduced order modelling to evolve the dynamical system. Aubry *et al.* first considered the issue and proposed a stabilisation procedure with a base flow variation by Reynolds stress [34]. Further, Noack *et al.* devised and analysed a least-order approach in a series of efforts that significantly improves the accuracy of simulation with a mean-field correction term in the Karhunen–Loéve decomposition of the dynamical system [21, 22, 35, 36].

Alternatively, we adopt a simple and effective remedy by supplementing the least squares fitting algorithm with a *prediction–correction* step. This is analogous to the eddy viscosity models in order

to dampen unrealistic high-frequency oscillations, while here, we introduce the damping in the decomposed space directly and intentionally to avoid the manipulation and the knowledge of the full model. Essentially, this ‘viscosity’ is numerical, and its physical implication depends on the specifics of the dynamics and is therefore beyond the scope of this paper.

The prediction–correction procedure operates as follows:

Once we detect that an output component of the model (9) exceeds the constraints imposed by the Smolyak grid that covers the data set, we enforce a damping to this component. That is, with the tilde denoting the fitting approximations and the starting value $\tilde{\alpha}^1 = \alpha^1$, one iteration of this ROM consists of two parts:

$$\text{Prediction:} \quad \hat{\alpha}_k^{n+1} = F_k^{P,\mu}(\tilde{\alpha}^n); \quad (25a)$$

$$\text{Correction:} \quad \tilde{\alpha}_k^{n+1} = \begin{cases} \hat{\alpha}_k^{n+1}, & \alpha_k^{\min} < \hat{\alpha}_k^{n+1} < \alpha_k^{\max}; \\ \delta \alpha_k^{\min}, & \hat{\alpha}_k^{n+1} \leq \alpha_k^{\min}; \\ \delta \alpha_k^{\max}, & \hat{\alpha}_k^{n+1} \geq \alpha_k^{\max} \end{cases} \quad (25b)$$

for $n = 1, 2, \dots, k = 1, 2, \dots, P$, and the admissible range of $\tilde{\alpha}_k^{n+1}$ is pre-determined from the data as

$$\alpha_k^{\min} = \min_{i=1}^S \alpha_k^i, \quad \alpha_k^{\max} = \max_{i=1}^S \alpha_k^i. \quad (26)$$

Here, we have introduced a variable damping parameter, $\delta \in (0, 1]$, and we will see in the numerical examples that an empirical choice of $\delta \in (0.8, 1.0]$ yields results with reasonable accuracy for different problems under a wide range of parameter settings and that the approach exhibits considerable robustness and general applicability.

Notice that this filter only affects how the ROM substitutes the full model, while the derivation of the model itself, and its non-intrusive features, remained unaltered. There are various alternative options that may achieve regularity such as limiting the scope of fitting data (e.g. finite impulse response filter [37]) and putting more weights on the data with small perturbations/gradients (e.g. weighted least squares [38]). However, for the sake of simplicity, we elect to postpone the exploration towards these directions to future work.

5. NUMERICAL EXAMPLES

In this section, we demonstrate the performance of the POD-based, non-intrusive reduced order scheme with least squares fitting by modelling a gyre flow and the flow past a cylinder that produces a von Kármán vortex street with two different Reynolds numbers. For both test cases, the solutions from the fidelity full model, which is simulated at the platform of Fluidity, serve as the reference and exact values for model comparisons as well as the snapshots for the POD basis generation. The error analysis was carried out with the same metrics used by Xiao *et al.* [32], namely, the root mean square error (RMSE) and correlation coefficient on the \mathcal{N} -node finite element, physical mesh. For example, the RMSE at the n^{th} time step is defined by

$$\text{RMSE}^n = \frac{\|\psi_{\text{NIROM}}^n - \psi_f^n\|_2}{\sqrt{\mathcal{N}}}. \quad (27)$$

where ψ_{ROM}^n and ψ_f^n are the nodal snapshots of the NIROM solution and of the full solution on the full mesh, respectively, and $\|\cdot\|_2$ is the standard vector two-norm. In all simulations, the ROM operates on a one-level Smolyak sparse grid, that is, $\mu = 1$ in equation (16), as increasing the number of grid level is found to provide negligible accuracy gain, while the growing grid size demands doubling or quadrupling the snapshot data to guarantee the well-posedness of the linear system (21).

5.1. Problem 1: gyre flow

We first test the least squares NIROM (LSROM) on the simulation of a shallow-water gyre flow. The geophysical scenario depicts a fluid circulating within a rectangular domain of the size $L_x \times L_y = 1000 \times 1000$ km. The flow is driven by the unidirectional surface wind stress prescribed by

$$\tau_y = \tau_0 \cos(\pi y/L_y) \quad \text{and} \quad \tau_x = 0 \quad (28)$$

where the maximum zonal stress is $\tau_0 = 0.1 \text{ N m}^{-1}$. In the full model, we address the Coriolis effect with the beta-plane approximation by setting $\beta = 1.8 \times 10^{-11} \text{ m}^{-1} \text{ s}^{-1}$, and we adopt the baseline fluid density $\rho_0 = 10^3 \text{ kg m}^{-3}$.

In this case, the high-fidelity full model was generated on a 2823-node finite element mesh for a duration of $T = 36.5$ days using a time step of $\Delta t = 0.365$ days. Accordingly, we record 100 snapshots of the solution from which only three POD basis functions for each variable (velocities u , v and pressure p) are extracted using the procedure described in Section 2.1. Then with the least square fitting method outlined in Section 3, we compute the nodal values of the transition functions (9) on the Smolyak grid determined by Step 2 at the end of Section 2.2. And finally, we iterate the ROM for each time step and reconstruct the LSROM-approximated solution on the physical mesh with expansion formula (7).

The comparison between the full model and the LSROM approximation is illustrated in Figure 1, in which we plotted the velocity magnitudes at two time instances, $t = 18$ days (the 50th time step) and $t = 36$ days (the 98th time step), as well as the difference between the full model and LSROM at these times. It is clear that the dominant features of the flow field are accurately resolved by the ROM.

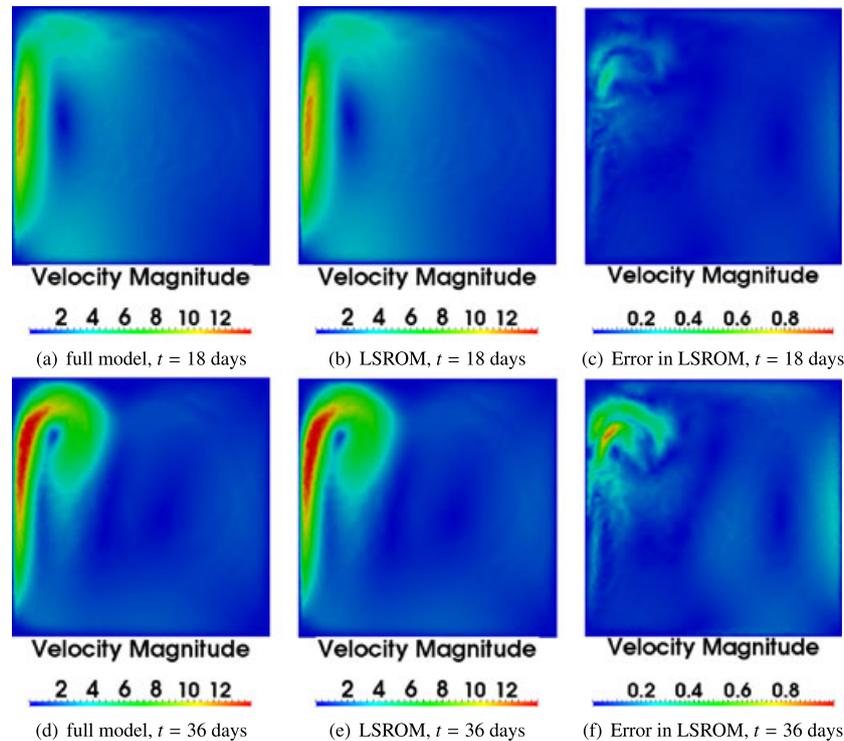


Figure 1. The gyre flow computed at time instances 18 (left) and 36 days (right). The panels (a) and (d) are full model simulations. The panels (b) and (e) are LSROM approximations with three POD basis function for each variable. The panels (c) and (f) are the differences between the full model and LSROM. (a) full model, $t = 18$ days; (b) LSROM, $t = 18$ days; (c) error in LSROM, $t = 18$ days; (d) full model, $t = 36$ days; (e) LSROM, $t = 36$ days; (f) error in LSROM, $t = 36$ days. LSROM, least squares non-intrusive reduced order model.

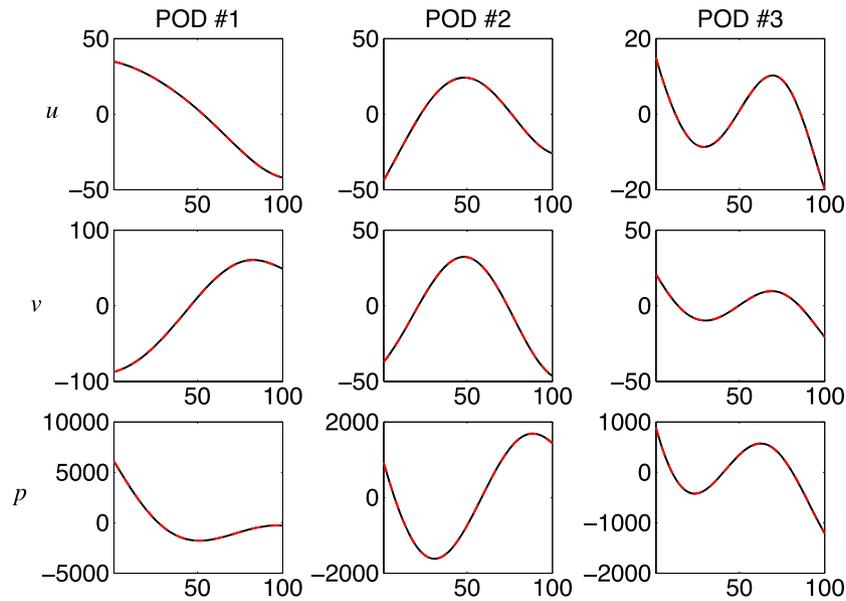


Figure 2. Gyre flow: the evolutions of POD coefficients for 3×3 POD basis functions. In each of the figures on top, the horizontal axis is the time step, and the vertical axis is the value of the POD coefficients; the black solid line is the projected POD coefficient of the full model solution, and the red dashed line is its least squares non-intrusive reduced order model approximation. In these figures, the two curves are indistinguishable from each other.

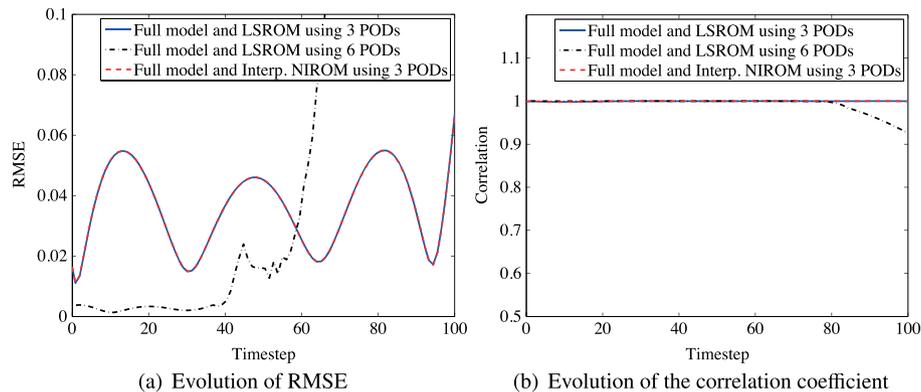


Figure 3. Gyre flow: the error diagnostics of the LSROM approximation with the reference of the interpolation-based NIROM. (a) Evolution of RMSE and (b) evolution of the correlation coefficient. NIROM, non-intrusive reduced order model; LSROM, least squares non-intrusive reduced order model.

On the other hand, we plot the evolution of the POD coefficients in the reduced space in Figure 2. Here, the three POD coefficients for each of the three variables (a total of nine) are plotted for all times in which the full model projections are virtually indistinguishable from the NIROM approximations. POD 1 is the coefficient associated with the leading POD basis function for each variable, or equivalently, α_1 in the series expansion (7) and so on. This figure also confirms that new ROM is a good approximation to the full model.

Quantitatively, Figure 3 displays evolutions of the RMSE and of the correlation coefficient between the full model and the non-intrusive model, respectively. Here, we plotted the results for two gyre simulations of LSROM, one with three POD basis functions and the other with six, and the result by the interpolation-based NIROM with three POD basis functions. Together, they further verify that the ROM approximation achieves an overall good agreement with the full model, especially for the LSROM with only three POD bases for which the results are essentially equiv-

Table I. Gyre flow: comparison of the normalised CPU consumption by different models at each time step.

Model	Assembling and solving	Projecting	Interpolation	Fitting	Total
Full model	1.0070	0.0000	0.0000	0.0000	1.0070
Interp. NIROM	0.0000	0.0040	0.0020	0.0000	0.0060
LSROM	0.0000	0.0040	0.0000	0.0001	0.0041

NIROM, non-intrusive reduced order model; LSROM, least squares non-intrusive reduced order mode.

alent to the interpolation-based NIROM [32] under the same setting. It should be noted that in these results, the damping parameter for stabilisation $\delta = 1.0$ was applied according to Section 4, and this choice is optimal in the sense that the temporal maximum of the associated RMSE is minimised.

However, we notice an intriguing phenomenon: we are able to increase the accuracy, indicated by smaller RMSE and larger correlation coefficient, with a doubled number of POD bases (from three to six). However, the improvement only lasts until the 57th time step, after which the RMSE soars and the correlation coefficient diverges from unity significantly after the 80th time step. This demonstrates another aspect of the stability issue we encountered as more POD modes potentially promote instability. An intuition suggests that the fitting nature of the model necessitates more snapshot data (e.g. more frequent sampling or equivalently, smaller Δt) to maintain a certain accuracy level as we also see in exploratory simulations. Because of the scope of this paper, we elect to study the detailed statistical relationship between the sizes of the POD decomposition and of the snapshot matrix in this LSROM framework to future work. For example, one may determine *a priori* the optimal number of POD bases used given a fixed number of snapshots and vice versa [39].

Furthermore, we compare the computation costs by different models in Table I where we list the CPU times required by different stages of different models at each time step. We observed that although the previous interpolation-based NIROM and the new fitting-based LSROM share the POD projection costs and both methods achieve great speed-ups in comparison with the full model, the cost of the fitting step in LSROM is negligible compared with that of the interpolations in Xiao *et al.* Because the fitting solves a model-independent linear system (21) and the interpolations always run the full model for a short period of time, it is to be expected that the LSROM with improved non-intrusiveness will possess an even bigger advantage for more complicated scenarios.

5.2. Problem 2: Flow past a cylinder

The second case in which we test the newly proposed LSROM method is also a classical example in computational fluid dynamics: a viscous inlet flow passes through a rectangular channel section that contains a cylinder and the oscillating downstream flow shed a street of von Kármán vortices. The simulation domain here is 2×0.4 in non-dimensional units, and a cylinder of radius 0.12 is located at the point (0.2, 0.2). The uniform upstream flow enters the domain from the left edge with unit speed rightward and is slightly compressible. Typical outflow condition is applied at the right edge, while the fluid is required to have no slip and no outward flow at the upper and lower edges of the channel. Finally, Dirichlet boundary conditions are enforced at the cylinder wall. We run the model under two settings of Reynolds number, $Re = 400$ and $Re = 3600$, to verify the general applicability of the method.

First, we run the high-fidelity full model with a prescribed Reynolds number $Re = 400$ on a 3213-node finite element mesh. Then we record the u , v and p solution variables between $t \in [2, 3]$ at regularly spaced time instances with $\Delta t = 0.01$, and a total of 100 snapshots are obtained. Subsequently, we generated six POD basis functions for each of the three variables before conducting the LSROM approximation procedure.

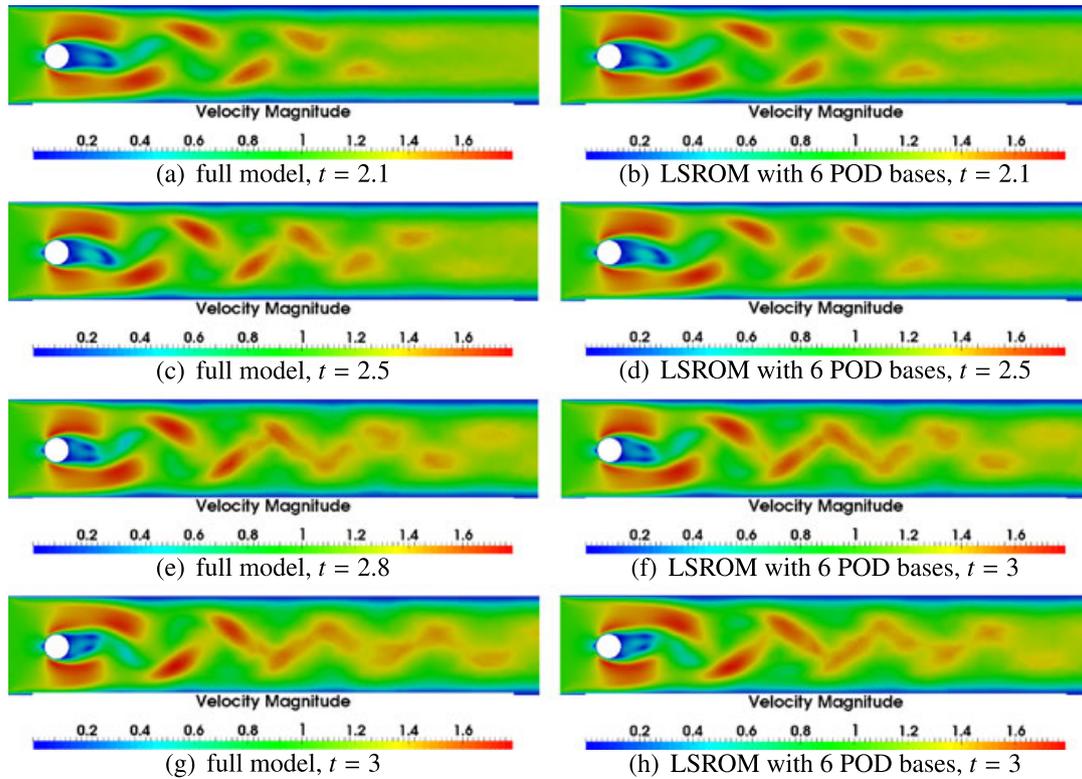


Figure 4. Flow past a cylinder at $Re = 400$. The figures on top compare the full model (figures (a), (c), (e) and (g)) and the LSROM with six POD basis functions (figures (b), (d), (f) and (h)) at $t = 2.1, 2.5, 2.8$ and 3.0 . (a) full model, $t = 2.1$; (b) LSROM with six POD bases, $t = 2.1$; (c) full model, $t = 2.5$; (d) LSROM with six POD bases, $t = 2.5$; (e) full model, $t = 2.8$; (f) LSROM with six POD bases, $t = 3$; (g) full model, $t = 3$; (h) LSROM with six POD bases, $t = 3$. LSROM, least squares non-intrusive reduced order model.

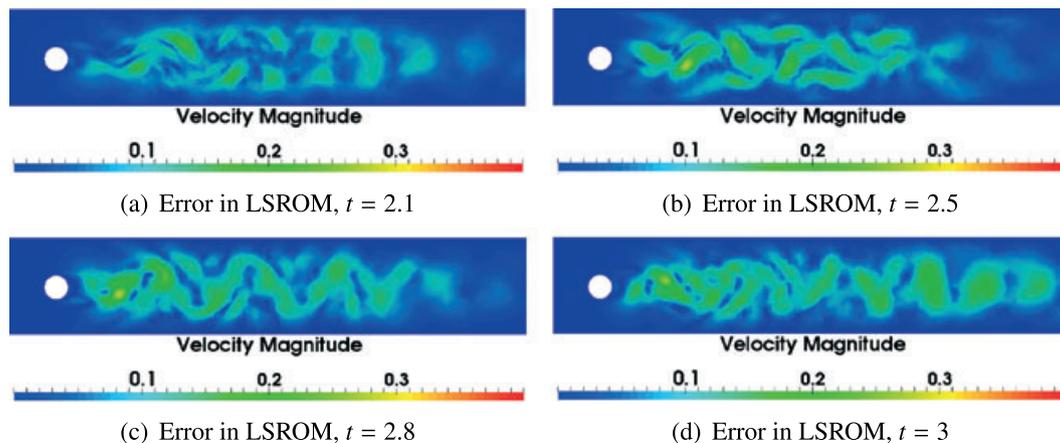


Figure 5. Flow past a cylinder at $Re = 400$: The difference between full model and LSROM with six POD basis functions. $Re = 400$. (a) error in LSROM, $t = 2.1$; (b) error in LSROM, $t = 2.5$; (c) error in LSROM, $t = 2.8$; (d) error in LSROM, $t = 3$. LSROM, least squares non-intrusive reduced order model.

Figures 4 through 6 illustrates how the LSROM resolves the full model in this case. Visual inspection of Figure 4 confirms the satisfactory performance of the LSROM for this problem by reproducing the location, the size and the magnitude of almost all the vortices. Figure 5 displays the spatial distribution of the error in LSROM. Comparing Figure 4 and 5, we see that the point-wise difference is in general an order of magnitude smaller than the exact values. Figure 6 documents

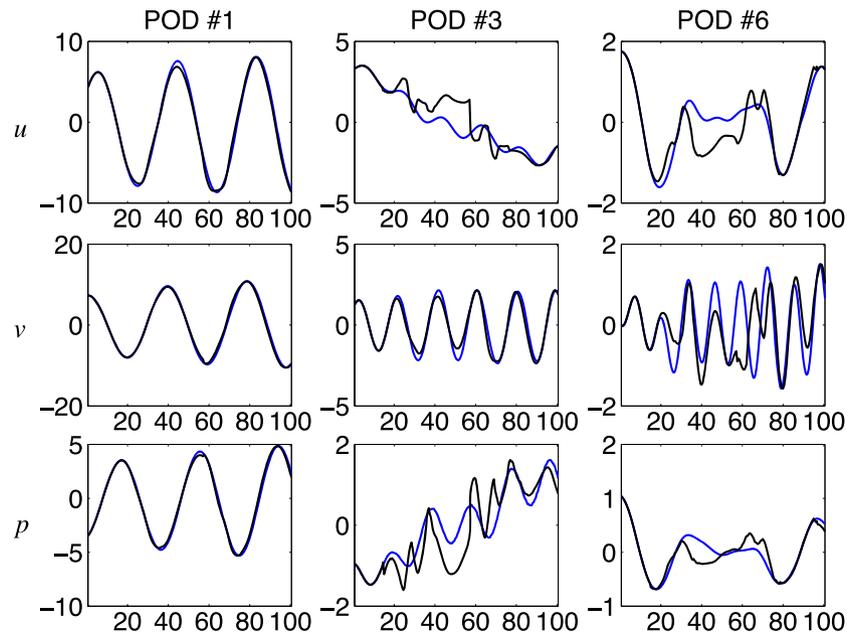


Figure 6. Flow past a cylinder at $Re = 400$: The evolutions of POD coefficients for nine POD basis functions. In each of the figures on top, the horizontal axis is the time step, and the vertical axis is the value of the POD coefficients; the black, solid line is the projected POD coefficient of the full model solution, and the blue dotted line is its least squares non-intrusive reduced order model approximation.

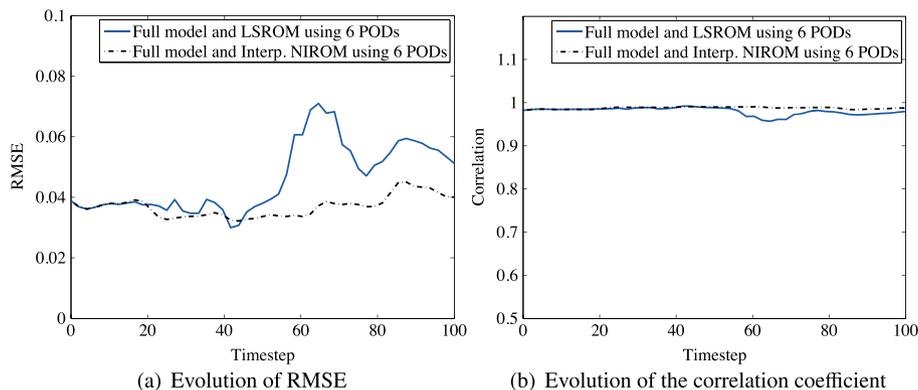


Figure 7. Flow past a cylinder at $Re = 400$: the error diagnostics of LSROM with six POD bases in comparison with the interpolation-based NIROM with the same number of POD bases. (a) Evolution of RMSE. (b) Evolution of the correlation coefficient. LSROM, least squares non-intrusive reduced order model; NIROM, non-intrusive reduced order model.

the evolution of the POD coefficients of the full model and of the LSROM. Additionally, the error diagnostics for this problem are shown in Figure 7 with a comparison with the results obtained by the interpolation-based NIROM using the same number of POD bases. Here, the optimal damping parameter δ for the LSROM is chosen to be 0.94. We see that for this test problem, the interpolation-based NIROM deviates from the LSROM and performs slightly better after the midpoint in the temporal axis, but both methods are accurate approximations.

Next, we increase the level of turbulence of the flow by setting $Re = 3600$, and we keep track of the simulation for a longer duration for $t \in [2, 6]$ with $\Delta t = 0.02$. All the other physical and numerical parameters remain the same. The corresponding results are demonstrated in Figures 8 through 10. Here, the optimal damping parameter δ for the LSROM is chosen to be 0.84. We can see that the LSROM still resolves the main flow structures. However the accuracy is apparently

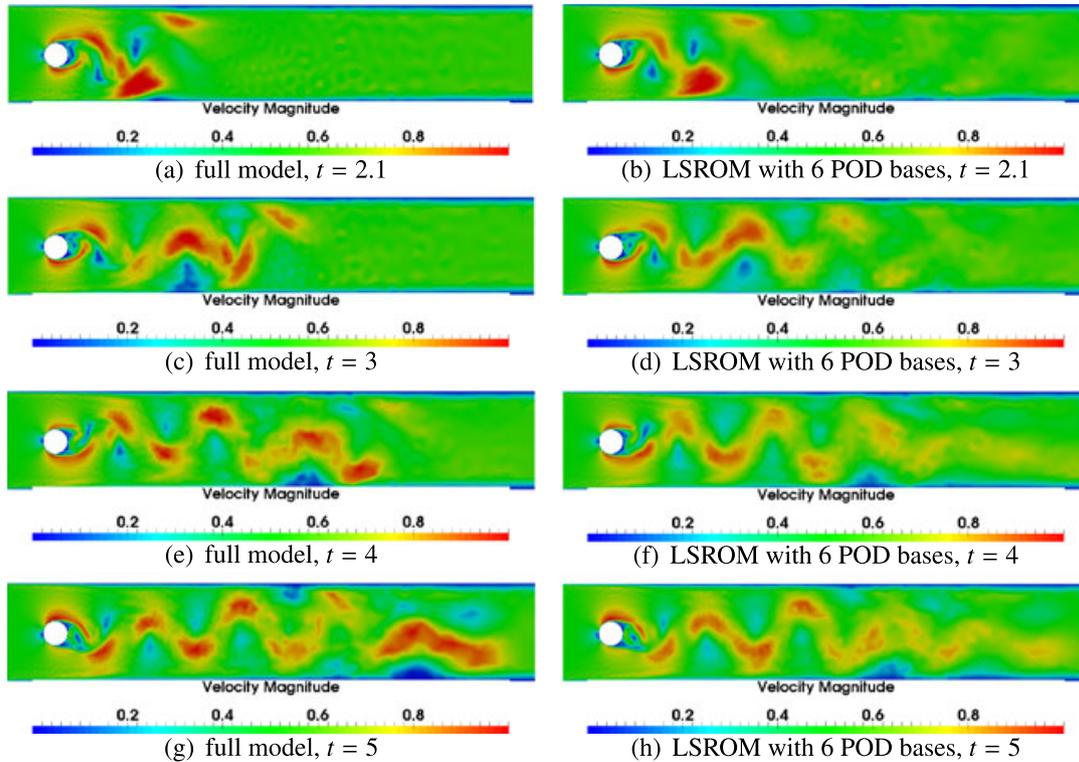


Figure 8. Flow past a cylinder at $Re = 3600$. The figures on top compare the full model (figures (a), (c), (e) and (g)) and the LSROM with six POD basis functions (figures (b), (d), (f) and (h)) at $t = 2.1, 2.5, 2.8$ and 3.0 . (a) full model, $t = 2.1$; (b) LSROM with six POD bases, $t = 2.1$; (c) full model, $t = 3$; (d) LSROM with six POD bases, $t = 3$; (e) full model, $t = 4$; (f) LSROM with six POD bases, $t = 4$; (g) full model, $t = 5$; (h) LSROM with six POD bases, $t = 5$. LSROM, least squares non-intrusive reduced order model

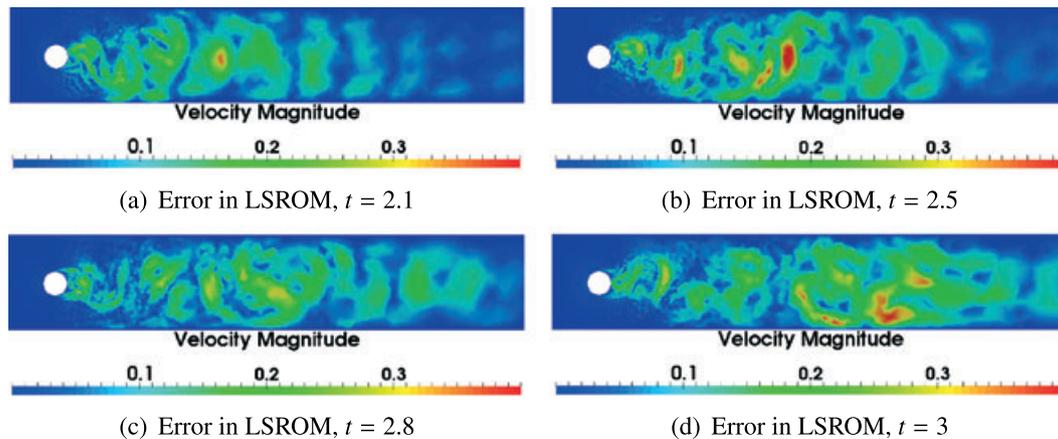


Figure 9. Flow past a cylinder at $Re = 3600$: the error diagnostics of LSROM with six POD bases. (a) error in LSROM, $t = 2.1$; (b) error in LSROM, $t = 2.5$; (c) error in LSROM, $t = 2.8$; (d) error in LSROM, $t = 3$. LSROM, least squares non-intrusive reduced order model.

reduced compared with the previous two cases, namely, the gyre flow and the flow past a cylinder at a lower Reynolds number. This should not be surprising as the spatial and temporal gradients of the variable are significantly magnified in this case and are thus harder to capture. These results can no doubt be improved by tuning the numerical parameters, such as the number of POD bases and of the snapshots used in the fitting, but the overall agreement achieved by this very restricted set of

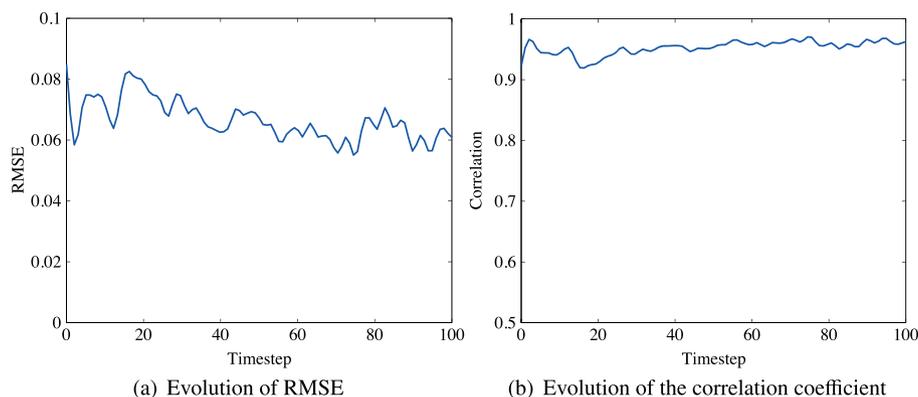


Figure 10. Flow past a cylinder at $Re = 3600$: the error diagnostics of LSROM with six POD bases. (a) Evolution of RMSE; (b) Evolution of the correlation coefficient.

POD basis is still worth reporting. And as we mentioned before, we will postpone the investigation in this aspect to future study.

6. CONCLUSIONS AND DISCUSSIONS

In this paper, a new non-intrusive reduced order model is introduced. The key ingredient here is a least squares fitting procedure to approximate the ROM formulated by standard POD routines that project the full model solution onto a lower finite-dimensional space. The new method also represents the ROM by polynomials defined on a Smolyak sparse grid whose coefficients are now obtained by data fitting. Compared with other intrusive ROMs and previous interpolation-based NIROM, this method only requires a number of solution snapshots from the full model and does not need to modify nor further run the original model in the process of ROM generation. In other words, the detailed physics and numerics of the original model are completely transparent to this LSROM. Another necessary module of the method involves the stabilisation of the reduced order iteration. A numerical damping parameter is chosen to avoid the divergence of the POD coefficients.

The least square NIROM was tested on two CFD test cases against a high-performance fluid solver (Fluidity). In the simulations of a shallow water gyre and of a Kármán vortex street downstream of a cylinder at $Re = 400$, the non-intrusive model resolves the flow structures with great precision, measured by RMSE and correlation coefficients, with only three or six POD basis functions. For the flow past a cylinder at $Re = 3600$, the accuracy of LSROM with six POD bases is decreased, but it still captures the main features of the flow field. Preliminary results show that using more POD bases could increase the accuracy for a limited time, beyond which the instability emerges. Overall, the least squares fitting model with improved non-intrusiveness is a good approximation to the full model with enormous gain in both computational efficiency and modelling economy.

Future work will address the stability issues of the ROM with greater details, such as optimising the data set for the least squares fitting by windowing (constructing the ROM using only a moving and continuous subset of the snapshots such as finite impulse response filtering) or by weighting (weighted least squares) and error analysis to determine the relationship between the optimal number of POD bases used and the number of full model solution snapshots available. Moreover, the LSROM will be applied to more complicated problems in sciences and engineering to fulfil its potential.

ACKNOWLEDGEMENTS

This work was carried out under partial funding from the Imperial College-Zhejiang University Joint Applied Data Science Lab and from the Data Science Institute at the Imperial College London. Authors would like to acknowledge the support of the UK's Natural Environment Research

Council projects (NER/A/S/2003/00595, NE/C52101X/1 and NE/C51829X/1), the Engineering and Physical Sciences Research Council (GR/R60898, EP/I00405X/1 and EP/I002011/1), and the Imperial College High Performance Computing Service. Navon acknowledges the support of NSF/CMG grant ATM-0931198. Xiao acknowledges the support of NSFC grant 11502241 and China post-doctoral science foundation grant (2014M562087). Lin acknowledges the support of NSFC grant 11201419. Pain and Fang are grateful for the support provided by BP Exploration. Pain is grateful for the support of the EPSRC MEMPHIS multi-phase flow programme grant.

REFERENCES

1. Lucia DJ, Beran PS, Silva WA. Reduced-order modeling: new approaches for computational physics. *Progress in Aerospace Sciences* 2004; **40**(1):51–117.
2. Schilders WHA, Van der Vorst HA, Rommes J. *Model Order Reduction: Theory, Research Aspects and Applications*, Vol. 13. Springer-Verlag: Berlin Heidelberg, 2008.
3. Fang F, Zhang T, Pavlidis D, Pain CC, Buchan AG, Navon IM. Reduced order modelling of an unstructured mesh air pollution model and application in 2D/3D urban street canyons. *Atmospheric Environment* 2014; **96**:96–106.
4. Ștefănescu R, Navon I M. POD/DEIM nonlinear model order reduction of an ADI implicit shallow water equations model. *Journal of Computational Physics* 2013; **237**:95–114.
5. Ștefănescu R, Sandu A, Navon IM. Comparison of POD reduced order strategies for the nonlinear 2D shallow water equations. *International Journal for Numerical Methods in Fluids* 2014; **76**(8):497–521.
6. San O, Borggaard J. Principal interval decomposition framework for POD reduced-order modeling of convective Boussinesq flows. *International Journal for Numerical Methods in Fluids* 2015; **78**(1):37–62.
7. Bistrrian DA, Navon IM. An improved algorithm for the shallow water equations model reduction: dynamic mode decomposition vs POD. *International Journal for Numerical Methods in Fluids* 2015; **78**(9):552–580.
8. Fang F, Pain CC, Navon IM, Elsheikh AH, Du J, Xiao D. Non-linear Petrov–Galerkin methods for reduced order hyperbolic equations and discontinuous finite element methods. *Journal of Computational Physics* 2013; **234**:540–559.
9. Xiao D, Fang F, Du J, Pain CC, Navon IM, Buchan AG, Elsheikh AH, Hu G. Nonlinear Petrov–Galerkin methods for reduced order modelling of the Navier–Stokes equations using a mixed finite element pair. *Computer Methods in Applied Mechanics and Engineering* 2013; **255**:147–157.
10. Xiao D, Fang F, Buchan AG, Pain CC, Navon IM, Du J, Hu G. Nonlinear model reduction for the Navier–Stokes equations using residual DEIM method. *Journal of Computational Physics* 2014; **263**:1–18.
11. Xiao D, Fang F, Pain CC, Navon IM. Non-intrusive reduced order 3D free surface modelling. *submitted to Ocean Engineering* 2015.
12. Daescu DN, Navon IM. A dual-weighted approach to order reduction in 4D-Var data assimilation. *Monthly Weather Review* 2008; **136**(3):1026–1041.
13. Cao Y, Navon IM, Luo Z. A reduced order approach to four-dimensional variational data assimilation using proper orthogonal decomposition. *International Journal for Numerical Methods in Fluids* 2007; **53**:1571–1583.
14. Ștefănescu R, Sandu A, Navon IM. POD/DEIM reduced-order strategies for efficient four-dimensional variational data assimilation. *Journal of Computational Physics* 2015; **295**:569–595.
15. Buchan AG, Calloo AA, Goffin MG, Dargaville S, Fang F, Pain CC, Navon IM. A POD reduced order model for resolving angular direction in neutron/photon transport problems. *Journal of Computational Physics* 2015; **296**:138–157.
16. Xiao D, Yang P, Fang F, Xiang J, Pain CC, Navon IM. Non-intrusive reduced order modeling of fluid–structure interactions. *Computer Methods in Applied Mechanics and Engineering* 2016; **303**:35–54.
17. Barone MF, Kalashnikova I, Brake MR, Segalman DJ. Reduced order modeling of fluid/structure interaction. *Sandia National Laboratories Report, SAND No* 2009; **7189**:44–72.
18. Wang M, Dutta D, Kim K, Brigham JC. A computationally efficient approach for inverse material characterization combining gappy POD with direct inversion. *Computer Methods in Applied Mechanics and Engineering* 2015; **286**:373–393.
19. Xiao D, Lin Z, Fang F, Pain CC, Navon IM, Salinas P, Muggeridge A. Non-intrusive reduced order modelling for multiphase porous media flows using Smolyak sparse grids. *International Journal for Numerical Methods in Fluids* 2016. Accepted for publication.
20. Hoang KC, Fu Y, Song JH. An HP-proper orthogonal decomposition moving least squares approach for molecular dynamics simulation. *Computer Methods in Applied Mechanics and Engineering* 2016; **298**:548–575.
21. Schlegel M, Noack BR. On long-term boundedness of Galerkin models. *Journal of Fluid Mechanics* 2015; **765**:325–352.
22. Öst J, Noack BR, Krajnović S, Barros D, Bore J. On the need for a nonlinear subscale turbulence term in POD models as exemplified for a high-Reynolds-number flow over an Ahmed body. *Journal of Fluid Mechanics* 2014; **747**:518–544.
23. Barrault M, Maday Y, Nguyen NC, Patera AT. An empirical interpolation method: application to efficient reduced-basis discretization of partial differential equations. *C. R. Acad. Sci. Paris, Ser* 2004; **339**:667–672.

24. Chaturantabut S, Sorensen DC. Nonlinear model reduction via discrete empirical interpolation. *SIAM Journal of Scientific Computation* 2010; **32**:2737–2764.
25. Rewienski MJ. A trajectory piecewise-linear approach to model order reduction of nonlinear dynamical systems. *Ph.D. Thesis*, 2003.
26. Wang Y, Navon IM, Wang X, Cheng Y. 2D Burgers equation with large Reynolds number using POD/DEIM and calibration. *International Journal for Numerical Methods in Fluids* 2016. DOI: 10.1002/flid.4249.
27. Han C. Blackbox stencil interpolation method for model reduction. *Master's Thesis*, 2012.
28. Walton S, Hassan O, Morgan K. Reduced order modelling for unsteady fluid flow using proper orthogonal decomposition and radial basis functions. *Applied Mathematical Modelling* 2013; **37**(20):8930–8945.
29. Xiao D, Fang F, Pain C, Hu G. Non-intrusive reduced order modelling of the Navier–Stokes equations based on RBF interpolation. *International Journal for Numerical Methods in Fluids* 2015; **79**(11):580–595.
30. Noack BR, Morzynski M, Tadmor G. *Reduced-Order Modelling for Flow Control*, Vol. 528. Springer, 2011.
31. Noori R, Karbassi AR, Ashrafi K, Ardestani M, Mehrdadi N. Development and application of reduced-order neural network model based on proper orthogonal decomposition for BOD5 monitoring: active and online prediction. *Environmental Progress and Sustainable Energy* 2013; **32**(1):120–127.
32. Xiao D, Fang F, Buchan AG, Pain CC, Navon IM, Muggeridge A. Non-intrusive reduced order modelling of the Navier–Stokes equations. *Computer Methods in Applied Mechanics and Engineering* 2015; **293**:522–541.
33. Pain CC, Piggott MD, Goddard AJH, Fang F, Gorman GJ, Marshall DP, Eaton MD, Power PW, De Oliveira CRE. Three-dimensional unstructured mesh ocean modelling. *Ocean Modelling* 2005; **10**(1):5–33.
34. Aubry N, Holmes P, Lumley JL, Stone E. The dynamics of coherent structures in the wall region of a turbulent boundary layer. *Journal of Fluid Mechanics* 1988; **192**:115–173.
35. Noack BR, Afanasiev K, Morzynski M, Tadmor G, Thiele F. A hierarchy of low-dimensional models for the transient and post-transient cylinder wake. *Journal of Fluid Mechanics* 2003; **497**:335–363.
36. Cordier L, Noack BR, Tissot G, Lehnasch G, Delville J, Balajewicz M, Daviller G, Niven RK. Identification strategies for model-based control. *Experiments in Fluids* 2013; **54**(8):1–21.
37. Cetin AE, Gerek ÖN, Yardimci Y. Equiripple FIR filter design by the FFT algorithm. *Signal Processing Magazine, IEEE* 1997; **14**(2):60–64.
38. Strutz T. *Data Fitting and Uncertainty: A Practical Introduction to Weighted Least Squares and Beyond*. Springer Vieweg: Teubner Verlag, 2011.
39. Moosavi A, Ștefănescu R, Sandu A. Efficient construction of local parametric reduced order models using machine learning techniques. *arXiv preprint arXiv:1511.02909* 2015.