



OPEN ACCESS

EDITED BY

Traian Iliescu,
Virginia Tech, United States

REVIEWED BY

Marin I. Marin,
Transilvania University of Braşov,
Romania
Changhong Mou,
University of Wisconsin-Madison,
United States
Zhu Wang,
University of South Carolina,
United States

*CORRESPONDENCE

Ping-Hsuan Tsai,
pht2@illinois.edu
Paul Fischer,
fischerp@illinois.edu

SPECIALTY SECTION

This article was submitted to Statistical and Computational Physics, a section of the journal Frontiers in Physics

RECEIVED 24 March 2022

ACCEPTED 07 July 2022

PUBLISHED 07 September 2022

CITATION

Tsai P-H and Fischer P (2022),
Parametric model-order-reduction
development for unsteady convection.
Front. Phys. 10:903169.
doi: 10.3389/fphy.2022.903169

COPYRIGHT

© 2022 Tsai and Fischer. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Parametric model-order-reduction development for unsteady convection

Ping-Hsuan Tsai^{1*} and Paul Fischer^{1,2*}

¹Department of Computer Science, University of Illinois, Urbana-Champaign, Champaign, IL, United States, ²Department of Mechanical Science and Engineering, University of Illinois, Urbana-Champaign, Champaign, IL, United States

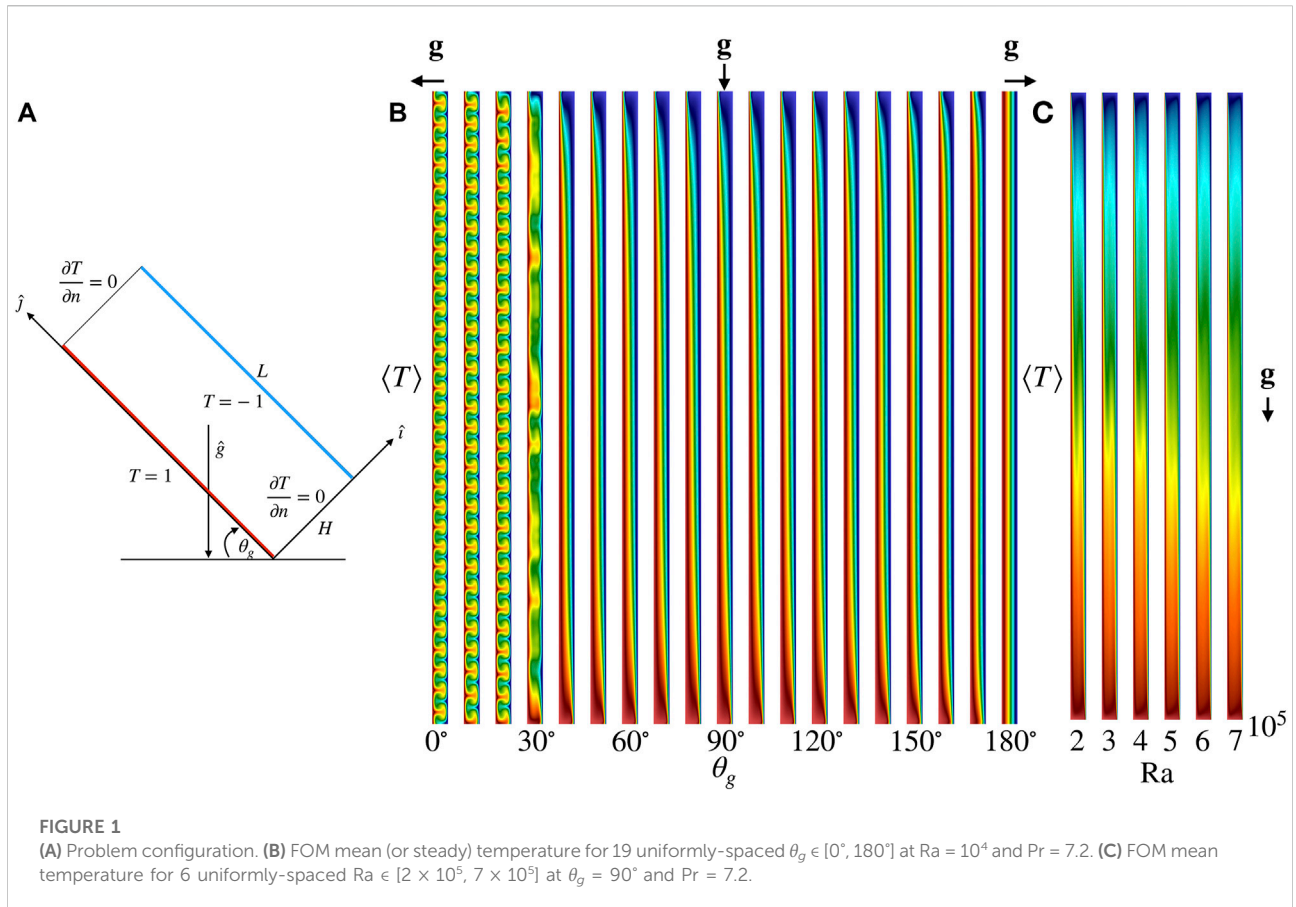
A time-averaged error indicator with POD-*h*Greedy is developed to drive parametric model order reduction (pMOR) for 2D unsteady natural convection in a high-aspect ratio slot parameterized with the Prandtl number, Rayleigh number, and slot angle with respect to the gravity. The error indicator is extended to accommodate the energy equation and Leray regularization. Despite being two-dimensional and laminar, the target flow regime presents several challenges: 1) there is a bifurcation in the angle parameter space; 2) the solution can be multivalued, even at steady state; and 3) the solution exhibits spatio-temporal chaos at several points in the parameter space. The authors explore several reduced-order models (ROMs) and demonstrate that Leray-regularized Galerkin ROMs provide a robust solution approach for this class of flows. They further demonstrate that error-indicated pMOR can efficiently predict several QOIs, such as mean flow, mean Nusselt number and mean turbulent kinetic energy, even in the presence of a bifurcation. Finally, they show that spatio-temporal chaos can lead to lack of reproducibility in both the full-order model and the reduced-order model and that the variance in the full-order model provides a lower bound on the pMOR error in these cases.

KEYWORDS

reduced basis method, model-order reduction (MOR), spatio-temporal chaos, a posteriori error estimate, proper orthogonal decomposition (POD), stabilization, leray regularization, incompressible Navier-Stokes equations

1 Introduction

Fluid-thermal analysis via direct numerical (DNS), large-eddy (LES), and even unsteady Reynolds-averaged Navier-Stokes (uRANS) have become tractable in geometries of ever increasing complexity due to advances in high-performance computing and modern algorithms over the past several decades. Despite these advances, when it comes to *routine analysis and design of thermal hydraulic systems*, which requires running hundreds of cases, the cost remains prohibitive. To overcome this



issue, a rapid turn-around tool for engineering query is required; *parametric model order reduction* (pMOR) is one of the promising approaches.

The main idea of pMOR is to reduce the computational burden by employing reduced-order models (ROMs) built on data from full-order models (FOMs) such as DNS, LES, or RANS-based simulations. A fundamental requirement in this case is to determine how well these approaches can reproduce the flow dynamics with same input parameters as the originating FOM, which is known as the *reproduction problem*. For turbulent flows, FOM can require $\mathcal{N} = 10^7 - 10^{11}$ degrees-of-freedom, while ROMs could potentially represent the flow dynamics which govern the behavior of quantities of interest (QOIs) with only $N \approx 10^1 - 10^2$ basis functions. Most often, the basis is obtained from a proper orthogonal decomposition (POD) of the FOM flow snapshots (i.e., the velocity and temperature fields). The basis is then used in conjunction with the governing partial differential equation to form a low-dimensional dynamical system that, ideally, captures the principal features of the underlying flows [1–8].

As noted in [9], a successful pMOR for *unsteady* flows must be able to address 1) the *reproduction problem* and 2) the *parametric problem*. In the reproduction problem, the ROM

and FOM are evaluated at the same $p_j^* \in \mathcal{P}_{\text{anchor}}$ and the ability of the ROM to recover the QOIs at p_j^* is examined. In the parametric problem, ROMs are built from a small number of FOMs that are generated over a set of *anchor* points in the parameter space, $\mathcal{P}_{\text{anchor}} = \{p_1^*, \dots, p_m^*\}$, and the ability of ROMs to predict the QOIs at $p \notin \mathcal{P}_{\text{anchor}}$ is examined. An *error indicator* to assess the ROM's fidelity at any given $p \in \mathcal{P}_{\text{train}}$ is usually required to efficiently construct $\mathcal{P}_{\text{anchor}}$. *Error-indicated pMOR* has the potential to be an important engineering analysis tool [10, 11].

In this work we explore the pMOR process for a surprisingly challenging 2D natural convection problem that serves as a surrogate for more difficult 3D buoyancy-driven flows encountered in a variety of mechanical and nuclear engineering applications. The geometry is the tilted slot configuration of Figure 1 and the governing equations are the unsteady Boussinesq equations. The problem is characterized by four parameters, the Rayleigh number, Ra , the Prandtl number, Pr , the slot angle with respect to gravity, θ_g , and the aspect-ratio, Γ . However, for small aspect ratios, $\Gamma \leq 8$, we find the flow is rather simple. Hence in this work, we focus on the more challenging case of $\Gamma = 40$ with (Pr, Ra, θ_g) as the parameter space.

Figure 1A illustrates the problem configuration. The aspect ratio is defined as $\Gamma = L/H$, where H is the width of the slot and L is the height of the slot. We take $H = 1$ and $L = 40$ throughout the study. With $\theta_g = 0^\circ$ the flow corresponds to standard Rayleigh-Bénard convection, 90° corresponds to vertical slot convection, and $\theta_g = 180^\circ$ leads to a pure conduction solution with the cold side on the bottom of the horizontal slot. The Rayleigh and Prandtl numbers are

$$\text{Ra} = \frac{\rho\beta\Delta TH^3 g}{\nu\alpha}, \quad \text{Pr} = \frac{\nu}{\alpha} \quad (1)$$

where ρ is the fluid density, β is the thermal expansion coefficient, $\Delta T = 2$ is the wall-to-wall temperature difference, H is the gap width, g is the gravitational acceleration, ν is the kinematic viscosity, and α is the thermal diffusivity. Figure 1B illustrates representative mean (or steady) temperature fields for 19 uniformly-spaced $\theta_g \in [0^\circ, 180^\circ]$ at $\text{Ra} = 10^4$ and $\text{Pr} = 7.2$. Figure 1C shows the mean temperature field for 6 uniformly-spaced $\text{Ra} \in [2 \times 10^5, 7 \times 10^5]$ at $\theta_g = 90^\circ$ and $\text{Pr} = 7.2$. The configuration has many interesting applications. For example, it represents energy-efficient double-glazed windows, in which the sealed air gap between the two panes acts as an added layer of insulation. Finding the optimum angle θ_g that enhances the heat transfer is an important question. Convection in a tilted fluid layer is also of meteorological and oceanographic interest. More information on the impact of θ_g and Γ on heat transport and flow organization for this configuration can be found in [12].

There has been significant work on pMOR development for the *steady* Boussinesq equations, including rigorous error estimation [10, 11, 13–15]. For pMOR, the steady problem is easier than the unsteady problem for several reasons: 1) rigorous error estimates are usually achievable, 2) there is often a well defined attractor, and 3) no temporal instability needs to be considered. Once the problem becomes unsteady many open research issues remain. To our knowledge, there are few pMOR works addressing the unsteady parameterized Boussinesq equations. In [16], the authors develop rigorous a posteriori error bounds applied to a 2D Rayleigh-Bénard problem parameterized with Gr and θ_g . However, due to exponential instability in time, the rigor is not for very high Gr and large final times. In [17], the authors overcame the high Gr issue by considering a space-time formulation which enabled effective long-time certification of a reduced basis approximation of noncoercive PDEs. However, the approach is limited due to large offline computational effort since only one snapshot is generated from one FOM solve due to the formulation. For example, to cover the parameter space, 125 FOMs are solved during the offline in their case.

Fick *et al.* [9] developed a POD-*h*Greedy pMOR to study challenging incompressible flow using a time-averaged error indicator. The authors showed that the error indicator is highly-correlated with the error in mean flow prediction and

can be efficiently computed through an offline/online strategy. We view the methodology as having high potential for routine analysis and design of turbulent flows that are characteristic of thermal hydraulic systems. Hence, we explore that approach here by extending the time-averaged error indicator to accommodate the energy equation. In previous work [18] on ROM stabilization and turbulent thermal transport problems, we investigated the performance of pMOR with constrained stabilization [9], and Leray regularization [19]. Here, we extend the error indicator to support Leray regularization. For each approach, we assess the performance through the mean flow and QOIs including, mean Nusselt number (Nu), standard deviation in Nu, mean temperature fluctuation and mean turbulent kinetic energy (TKE)¹.

Even though our 2D model problem generates only laminar flows, pMOR is quite challenging in this application for several reasons: 1) there is bifurcation in θ_g parameter space; 2) the solution can be multivalued, even at steady state; and 3) the solution exhibits spatio-temporal chaos at several points in the parameter space. As our initial efforts happened to be focused in one of the spatio-temporal chaos regimes, we decided to map out a larger space to identify where pMOR could succeed, where it would have difficulty, and where it *a priori* could not succeed. Table 1 reflects a broad range of flow regimes identified inside the high-aspect ratio slot from hundreds of FOM simulations conducted at multiple Ra , θ_g with $\text{Pr} = 0.07, 0.71, 7.2$. We categorize the flow into six types: 1) motionless, 2) steady, 3) periodic, 4) quasi-periodic, 5) chaotic and 6) spatio-temporal chaotic. We identify the flow regimes by examining the (mean) solution field and the energy and Nu histories. Such analysis can readily distinguish the motionless, steady and periodic flow cases. Even though the energy and Nu analysis seem to be a reliable way to distinguish the quasi-periodic and chaotic flow, it is only a heuristic—a more rigorous analysis is through computing the power spectrum of Nu or energy [20]. Tools such as Lyapunov exponent and fractal dimension are probably the most widely used diagnostic for chaotic systems [21, 22]. The first five types of flow have consistent mean flow in differing time windows, each averaged over 500 convective time units (CTUs). We define a flow to be spatio-temporal chaotic [23] if its mean solution is not consistent in at least three different time windows. This type of flow has strong irregularities in both space and time and has been observed in Rayleigh Bénard convection and in other complex dynamical systems [24]. To characterize spatio-temporal chaos, one

¹ Technically, since these flows are not turbulent the TKE should be referred to as *velocity variance*. Because the is more widely used and the mathematical formulation is the same in either case, we prefer to use the more widely recognized appellation, TKE.

TABLE 1 Distribution of six flow types with Ra and θ_g at Pr = 0.07, 0.71, 7.2.

flow \ Pr	Pr = 0.07	Pr = 0.71	Pr = 7.2
motionless	Ra < 1.1×10^3 , $\theta_g = 0^\circ$	Ra < 8.75×10^2 , $\theta_g = 0^\circ$	Ra < 9×10^2 , $\theta_g = 0^\circ$
steady	Ra = 10^3 , $\theta_g = 90^\circ$ Ra = 10^4 , $\theta_g \in [0^\circ, 40^\circ] \cup [170^\circ, 180^\circ]$	Ra = 10^3 , $\theta_g \in [0^\circ, 180^\circ]$ Ra = 10^4 , $\theta_g \in [0^\circ, 30^\circ] \cup [70^\circ, 80^\circ] \cup [120^\circ, 180^\circ]$ Ra = 1.5×10^4 , $\theta_g = 90^\circ$	Ra = 10^3 , $\theta_g \in [0^\circ, 180^\circ]$ Ra = 10^4 , $\theta_g \in [0^\circ, 180^\circ] \setminus \{30^\circ\}$ Ra $\in [2 \times 10^4, 10^5]$, $\theta_g = 90^\circ$ Ra = 8×10^4 , $\theta_g \in [80^\circ, 180^\circ]$ Ra = 3×10^5 , $\theta_g \in [130^\circ, 180^\circ]$
periodic	Ra $\in [5 \times 10^3, 7 \times 10^3]$, $\theta_g = 90^\circ$	Ra = 10^4 , 1.75×10^4 , $\theta_g = 90^\circ$	N/A
quasi-periodic	Ra = 10^4 , $\theta_g \in [50^\circ, 110^\circ]$ Ra = 8×10^3 , 9×10^3 , 2×10^4 , $\theta_g = 90^\circ$	Ra $\in [1.8 \times 10^4, 2 \times 10^4]$, $\theta_g = 90^\circ$	N/A
chaotic	Ra = 10^4 , $\theta_g \in [120^\circ, 160^\circ]$ Ra $\in [6 \times 10^4, 1.5 \times 10^5]$, $\theta_g = 90^\circ$	N/A	Ra = 8×10^4 , $\theta_g \in [40^\circ, 70^\circ]$ Ra = 3×10^5 , $\theta_g \in [50^\circ, 120^\circ]$ Ra $\in [2 \times 10^5, 8 \times 10^5]$, $\theta_g = 90^\circ$
spatio-temporal chaotic	Ra $\in [3 \times 10^4, 5 \times 10^4]$, $\theta_g = 90^\circ$	Ra = 10^4 , $\theta_g \in [40^\circ, 60^\circ] \cup [100^\circ, 110^\circ]$ Ra = 1.25×10^4 , $\theta_g = 90^\circ$ Ra $\in [2.05 \times 10^4, 3 \times 10^5]$, $\theta_g = 90^\circ$	Ra = 10^4 , $\theta_g = 30^\circ$ Ra = 8×10^4 , $\theta_g \in [0^\circ, 30^\circ]$ Ra = 3×10^5 , $\theta_g \in [0^\circ, 40^\circ]$ Ra $\in [2 \times 10^5, 8 \times 10^5]$, $\theta_g = 0^\circ, 10^\circ$

could also consider Lyapunov exponents at each grid point. A detail analysis of spatio-temporal chaos is beyond the scope of this paper. A comprehensive review on this topic can be found in [25].

In the present work, we start with solution reproduction problem, which represents the first step towards the development of a ROM for the parametric problem. We study the reproduction capability of the FOM, ROM and ROM with stabilization for the six types of flow reported in Table 1. We report only the cases of chaotic and spatio-temporal chaotic flow. Notice that it has been reported that ROM can often capture the first five types of flow accurately, in some cases, with the need of stabilization methods. However, for the spatio-temporal chaotic flow, to our knowledge, it has not been studied. We believe this is the first work investigate ROM's reproducibility of spatio-temporal chaotic flow.

The pMOR development is broken into several parametric problems. We start with a problem at Ra = 10^3 and Pr = 7.2 in which the solution is steady for all θ_g but nonetheless exhibits a bifurcation at $\theta_g = 20^\circ$. We find either h - or p -Greedy with the error indicator based on the dual norm of the residual is able to drive pMOR successfully³. We next consider two sets of parametric problems: 1) problem

parameterized with θ_g at higher Ra. 2) Problem parameterized with Ra. In the first set, similar to the steady case, a bifurcation is observed and the solution space is a blend of steady and unsteady solutions. In the second set, no bifurcation is observed and the solutions are all unsteady.

By proceeding in this manner, we are able to isolate several difficulties and eventually come up with an important observation for the pMOR: Accurate prediction (<10%) with pMOR is achievable if the solution in the parametric space is either only chaotic or the spatio-temporal chaos is not significant, regardless a bifurcation exists or not. Once the spatio-temporal chaos becomes significant, the performance of the pMOR deteriorates and the maximum errors of the mean flow and QOIs are dominated by the flow chaos.

The paper is organized as follows. In Section 2, we introduce the model problem and governing equations. In Section 2.1, we introduce the Galerkin formulation for the FOM. The ROM, as well as Leray regularization is introduced in 2.2. In Section 3, we consider the solution reproduction problem and assess the numerical performance. The parametric problem is discussed starting from Section 4. We first introduce POD- h Greedy algorithms in Section 4.1 with some remarks on applying POD- p Greedy to this model problem at the end of the section. We then introduce the time-averaged error indicator with thermal extension in Section 4.2. A straight-forward integration with Leray regularization is also shown in the same section. In Sections 4.3 and 4.4, we present the pMOR results with θ_g variation and Ra variation. In Section 5, we discuss the spatio-temporal chaos and multiple states issues found in this model problem. Finally, we conclude the paper in Section 6.

² The ROM coefficients in the steady problems are typically found through a Newton minimization over the POD approximation space [10, 11].

³ The pMOR greedy strategy uses the maximal indicated error among the parametric training set to select the next anchor point. p -Greedy combines basis functions from FOMs at different anchor points to form an enriched approximation space; h -greedy builds an independent ROM for each anchor point Ngoc Cuong et al. [10].

2 A parametrized natural convection problem

We start with the Boussinesq equations for buoyancy-driven flow [26],

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} + \nabla p = \nu \nabla^2 \mathbf{u} + T \mathbf{g}(\theta_g), \quad \nabla \cdot \mathbf{u} = 0, \quad (2)$$

$$\frac{\partial T}{\partial t} + (\mathbf{u} \cdot \nabla) T = \alpha \nabla^2 T, \quad (3)$$

where \mathbf{u} is the velocity, p is the pressure, T is the temperature and $\mathbf{g}(\theta_g)$ is the unit vector represents the direction of the buoyancy force and it is defined by $\mathbf{g}(\theta_g) = \cos(\theta_g)\hat{i} + \sin(\theta_g)\hat{j}$, with θ_g the angle of the slot with respect to the gravity. The velocity boundary conditions are no-slip ($\mathbf{u} = 0$). The temperature boundary conditions are no-flux (insulated) on the top and bottom, heated ($T = 1$) on the left wall, and cooled ($T = -1$) on the right wall. The initial conditions are $\mathbf{u} = 0$ and $T = 0$.

In our non-dimensional setting, we set $\nu = (\text{Pr}/\text{Ra})^{1/2}$ and $\alpha = (\text{Pr}/\text{Ra})^{-1/2}$. The Rayleigh number, $\text{Ra} = \rho\beta g H^3 \Delta T / (\nu\alpha)$, represents the ratio of buoyancy force to thermal and momentum diffusive force. The Prandtl number $\text{Pr} = \nu/\alpha$, reflects the relative importance of momentum diffusivity compared to thermal diffusivity. With this nondimensionalization the characteristic velocity is $U_c = \sqrt{\beta g H \Delta T}$, the characteristic length is the slot width H , and the reference time is $t_r = H/U_c$. The temperature is made dimensionless by subtracting the temperature on the right wall and scaling with $\Delta T = 2$. We note that U_c is sometimes referred to as the “free-fall” velocity, indicating that one might expect $\|\mathbf{u}\| \approx 1$, with only a weak dependence on Ra . While that expectation is realized for $\theta_g = 0$, we in fact see much larger velocities ($\|\mathbf{u}\| \approx 40$) for $\theta_g = 90^\circ$ because the domain height $L = 40H$ in that case.

For unsteady problems, the QOIs are the mean flow, mean Nu, standard deviation in Nu, mean TKE and mean temperature fluctuation. The symbol $\langle \cdot \rangle$ is used to indicate a time-averaged quantity. The mean velocity and temperature field are defined as:

$$\langle \mathbf{u} \rangle = \frac{1}{J - J_0} \sum_{j=J_0+1}^J \mathbf{u}(t^j), \quad \langle T \rangle = \frac{1}{J - J_0} \sum_{j=J_0+1}^J T(t^j), \quad (4)$$

with $t^j = j\Delta t$ and Δt being the time step. The selection of J_0 is based on when the solution reaches it statistically steady state. The mean quantities are then averaged over 500 CTUs, with the time scale defined above. The instantaneous Nusselt number is defined as

$$\text{Nu}(t) = \frac{q_w''}{k(\Delta T)/H}, \quad (5)$$

with $q_w'' = -\int_{\partial\Omega_h} k \nabla T \cdot \hat{\mathbf{n}} dS$ being the integrated heat flux on the heated wall, $\partial\Omega_h$. The mean Nu and its standard deviation are then defined as

$$\langle \text{Nu} \rangle = \frac{1}{J - J_0} \sum_{j=J_0+1}^J \text{Nu}(t^j),$$

$$\text{Std}(\text{Nu}) = \sqrt{\frac{1}{J - J_0} \sum_{j=J_0+1}^J (\text{Nu}(t^j) - \langle \text{Nu} \rangle)^2}. \quad (6)$$

The mean TKE and mean temperature fluctuation are defined as

$$\langle \text{TKE} \rangle = \frac{1}{2(J - J_0)} \sum_{j=J_0+1}^J \|\mathbf{u}(t^j) - \langle \mathbf{u} \rangle\|_{L^2}^2,$$

$$\langle T_{\text{fluc}} \rangle = \frac{1}{J - J_0} \sum_{j=J_0+1}^J \|T(t^j) - \langle T \rangle\|_{L^2}^2. \quad (7)$$

For steady problems, the QOIs are simply the steady solutions to Eqs. 2 and 3 and the corresponding Nu using Eq. 5.

2.1 Galerkin formulation for the full-order model

The FOM is constructed through the spectral element method (SEM) and the P_q - P_{q-2} velocity-pressure coupling [27], where the velocity is represented as a tensor-product Lagrange polynomial of degree q in the reference element $\hat{\Omega} := [-1, 1]^2$ while the pressure is of degree $q - 2$. The solution in $\Omega = \bigcup_e \Omega^e$ consists of local representations of \mathbf{u} , p , and T that are mapped from $\hat{\Omega}$ to Ω^e for each element, $e = 1, \dots, E$. In the current FOMs for the slot problem we use $E = 516$ elements (an array of 6×86 in the $H \times L$ directions), of order $q = 9$, for a total of $\mathcal{N} \approx 42000$ grid points. The FOM simulations are performed using the open-source code Nek5000 [28].

For any $\mathbf{u}(\mathbf{x}, t)$, we have a corresponding vector of basis coefficients $\underline{\mathbf{u}} = [\mathbf{u}_1 \dots \mathbf{u}_{\mathcal{N}}]^T$ such that

$$\mathbf{u}(\mathbf{x}, t) = \sum_{j=1}^{\mathcal{N}} \mathbf{u}_j(t) \phi_j(\mathbf{x}) \in \mathbf{X}_0^{\mathcal{N}} \subset \{\mathcal{H}_0^1\}, \quad (8)$$

with $\phi_j(\mathbf{x})$ the underlying spectral element basis functions spanning the FOM approximation space, $\mathbf{X}_0^{\mathcal{N}}$. Because the SEM is nodal-based, each $\mathbf{u}_j(t)$ represents the two velocity components at grid point \mathbf{x}_j in the spectral element mesh at time t . Similarly, the temperature is given by

$$T(\mathbf{x}, t) = \sum_{j=1}^{\mathcal{N}} T_j(t) \phi_j(\mathbf{x}) \in X_0^{\mathcal{N}} \subset \mathcal{H}_0^1. \quad (9)$$

Here, \mathcal{H}^1 is the set of square-integrable functions on Ω whose gradient is also square-integrable and $X^{\mathcal{N}} \subset \mathcal{H}^1$ is the finite dimensional SEM approximation space spanned by $\{\phi_j(\mathbf{x})\}$. \mathcal{H}_0^1 is the set of functions in \mathcal{H}^1 that vanish wherever Dirichlet conditions associated with Eq. 3 are applied on the domain boundary $\partial\Omega$ and \mathcal{H}_b^1 is the set of functions in \mathcal{H}^1 that satisfy the prescribed Dirichlet conditions for temperature. Bold-face

indicates that the space is spanned by vector-valued functions having d components ($d = 2$ or 3) and, in the case of $\mathbf{X}_0^N \subset \{\mathcal{H}\}_0^1$, that the functions vanish where Dirichlet conditions are applied for Eq. 2. The pressure p is in $Y^N \subset \mathcal{L}^2(\Omega)$, which is the space of piecewise continuous functions on Ω such that $\int_{\Omega} p^2 d\mathbf{x} < \infty$. For convenience, we denote $\bar{\mathbf{Z}}^N = (\mathbf{X}^N, Y^N, X^N)$ as the collection of the relevant finite-dimensional spaces and will add a subscript 0 or b where required to explicitly indicate the imposed boundary conditions.

Both the FOM and ROM are cast within the same Galerkin framework. To begin, we introduce several inner products for elements in the FOM space, $\bar{\mathbf{Z}}^N$. For any pair of scalar fields $(p, q) \in \mathcal{L}^2(\Omega)$ and d -dimensional vector fields, $\mathbf{v}(\mathbf{x}) = [v_1(\mathbf{x}) \dots v_d(\mathbf{x})]$, $\mathbf{u}(\mathbf{x}) = [u_1(\mathbf{x}) \dots u_d(\mathbf{x})]$ whose components are also in \mathcal{L}^2 , let

$$(q, p) := \int_{\Omega} q p d\mathbf{x}, \quad (\mathbf{v}, \mathbf{u}) := \int_{\Omega} (v_1 u_1 + \dots + v_d u_d) d\mathbf{x}. \quad (10)$$

Further, for $S, T \in X^N$ and $\mathbf{v}, \mathbf{u} \in \mathbf{X}^N$, let

$$a(S, T) := (\nabla S, \nabla T), \quad a(\mathbf{v}, \mathbf{u}) := (\nabla \mathbf{v}, \nabla \mathbf{u}), \quad (11)$$

$$c(S, \mathbf{u}, T) := (S, \mathbf{u} \cdot \nabla T), \quad c(\mathbf{v}, \mathbf{u}, \mathbf{w}) := (\mathbf{v}, \mathbf{u} \cdot \nabla \mathbf{w}). \quad (12)$$

For the FOM, we consider the (semi-discrete) weak form of Eqs. 2 and 3 [29], Find $(\bar{\mathbf{u}}, p, \bar{T}) \in \bar{\mathbf{Z}}_b^N$ such that, for all $(\mathbf{v}, q, S) \in \bar{\mathbf{Z}}_0^N$,

$$\left(\mathbf{v}, \frac{\partial \bar{\mathbf{u}}}{\partial t} \right) + \nu a(\mathbf{v}, \bar{\mathbf{u}}) - (\nabla \cdot \mathbf{v}, p) = -c(\mathbf{v}, \bar{\mathbf{u}}, \bar{\mathbf{u}}) + (\mathbf{v}, \mathbf{g}(\theta_g) \bar{T}), \quad (13)$$

$$-(q, \nabla \cdot \bar{\mathbf{u}}) = 0, \quad (14)$$

$$\left(S, \frac{\partial \bar{T}}{\partial t} \right) + \alpha a(S, \bar{T}) = -c(S, \bar{\mathbf{u}}, \bar{T}). \quad (15)$$

Here, we have introduced $\bar{\mathbf{u}} = \mathbf{u} + \mathbf{u}_0(\mathbf{x})$ and $\bar{T} = T + T_0(\mathbf{x})$ as functions that have been augmented by (potentially trivial) lifting functions, \mathbf{u}_0 and T_0 , which are functions of space only. If these functions satisfy the (time-independent) boundary conditions, then one can account for inhomogeneous boundary conditions by moving them to the right-hand side. In the case of the ROM, the lifting functions can also provide an initial approximation to the solution. In the sequel, our principal unknowns will be \mathbf{u} and T .

Following [30], we consider a semi-implicit scheme BDFk/EXTk to discretize Eqs. 13–15 in time; k th-order backward differencing (BDFk) is used for the time-derivative term, k th-order extrapolation (EXTk) is used for the advection and buoyancy terms and implicit treatment on the dissipation terms. As discussed in [30], $k = 3$ is used to ensure the imaginary eigenvalues associated with skew-symmetric advection operator are within the stability region of the

BDFk/EXTk time-stepper. Denoting the solution at time $t^n = \Delta t \cdot n$ as $(\bar{\mathbf{u}}^n, p^n, \bar{T}^n)$, the full discretization of the FOM reads Find $(\mathbf{u}^n, p^n, T^n) \in \bar{\mathbf{Z}}_0^N$ such that, for all $(\mathbf{v}, q, S) \in \bar{\mathbf{Z}}_0^N$,

$$\frac{\beta_0}{\Delta t} (\mathbf{v}, \mathbf{u}^n) + \nu a(\mathbf{v}, \mathbf{u}^n) - (\nabla \cdot \mathbf{v}, p^n) = (\mathbf{v}, \mathbf{f}^n), \quad (16)$$

$$-(q, \nabla \cdot \mathbf{u}^n) = (q, \nabla \cdot \mathbf{u}_0), \quad (17)$$

$$\frac{\beta_0}{\Delta t} (S, T^n) + \alpha a(S, T^n) = (S, Q^n). \quad (18)$$

Equations 16–18 represent a linear unsteady Stokes plus unsteady heat equation to be solved at each time-step t^n . The inhomogeneous terms comprise the BDF, advection, buoyancy and lifting terms

$$(\mathbf{v}, \mathbf{f}^n) := - \sum_{s=1}^k \left[\frac{\beta_s}{\Delta t} (\mathbf{v}, \mathbf{u}^{n-s}) + \alpha_s (c(\mathbf{v}, \bar{\mathbf{u}}^{n-s}, \bar{\mathbf{u}}^{n-s}) - (\mathbf{v}, \mathbf{g}(\theta_g) \bar{T}^{n-s})) \right] - \nu a(\mathbf{v}, \mathbf{u}_0), \quad (19)$$

$$(S, Q^n) := - \sum_{s=1}^k \left[\frac{\beta_s}{\Delta t} (S, T^{n-s}) + \alpha_s c(S, \bar{\mathbf{u}}^{n-s}, \bar{T}^{n-s}) \right] - \alpha a(S, T_0). \quad (20)$$

Here, the β_s s and α_s s are the respective s th-order BDF and extrapolation coefficients for the BDFs/EXTs time-stepper [30]. Note that the right-hand side of Eq. 17 will be zero if \mathbf{u}_0 is divergence free or at least satisfies the weak divergence-free condition Eq. 14.

Under the assumption that $\nabla \cdot \mathbf{u}_0 = 0$, the compact matrix form [27, 31, 32] for Eqs. 16–20 is

$$\begin{bmatrix} \mathbf{H} & -\mathbf{D}^T \\ -\mathbf{D} & 0 \end{bmatrix} \begin{pmatrix} \underline{\mathbf{u}}^n \\ \underline{p}^n \end{pmatrix} = \begin{pmatrix} \hat{\mathbf{f}}(\bar{\mathbf{u}}^n, \bar{T}^n; \theta_g) \\ 0 \end{pmatrix}, \quad (21)$$

$$H_{\alpha} \underline{T}^n = \hat{\mathbf{Q}}(\bar{\mathbf{u}}^n, \bar{T}^n). \quad (22)$$

Here, $\underline{\mathbf{u}}^n$, \underline{p}^n , and \underline{T}^n are the vectors of spectral element basis coefficients. The corresponding block matrices are

$$\mathbf{H} = \begin{bmatrix} H_v & \\ & H_p \end{bmatrix}, \quad \mathbf{D} = [D_1 \ D_2], \quad (23)$$

with $H_v = \frac{\beta_0}{\Delta t} M + \nu A$ and $H_{\alpha} = \frac{\beta_0}{\Delta t} M + \alpha A$, with matrices M and A defined below. The velocity data vectors are $\hat{\mathbf{f}}(\bar{\mathbf{u}}^n, \bar{T}^n; \theta_g) = [\hat{f}_1 \ \hat{f}_2]^T$, with

$$\begin{aligned} \hat{f}_m &:= - \sum_{s=1}^k \left[\frac{\beta_s}{\Delta t} M \underline{u}_m^{n-s} + \alpha_s (C(\bar{\mathbf{u}}^{n-s}) \underline{u}_m^{n-s} - g_m M \bar{T}^{n-s}) \right] \\ &\quad - \nu A \underline{u}_{m,0}, \\ m &= 1, 2. \end{aligned} \quad (24)$$

where $g_1 = \cos(\theta_g)$ and $g_2 = \sin(\theta_g)$ represent the parametric forcing. The thermal load in Eq. 22 is

$$\hat{\mathbf{Q}}(\bar{\mathbf{u}}^n, \bar{T}^n) := - \sum_{s=1}^k \left[\frac{\beta_s}{\Delta t} M \bar{T}^{n-s} + \alpha_s C(\bar{\mathbf{u}}^{n-s}) \bar{T}^{n-s} \right] - \alpha A T_0. \quad (25)$$

Entries of the respective stiffness, mass, convection, and gradient matrices are

$$A_{ij} = \int_{\Omega} \nabla \phi_i \nabla \phi_j \, d\mathbf{x}, \quad (26)$$

$$M_{ij} = \int_{\Omega} \phi_i \phi_j \, d\mathbf{x}, \quad (27)$$

$$C_{ij}(\mathbf{w}) = \int_{\Omega} \phi_i \cdot (\mathbf{w} \cdot \nabla) \phi_j \, d\mathbf{x}, \quad (28)$$

$$D_{m,ij} = \int_{\Omega} \psi_i \frac{\partial \phi_j}{\partial x_m} \, d\mathbf{x}, \quad m = 1, 2. \quad (29)$$

Note that $\{\phi_i(\mathbf{x})\}$ forms the spectral element velocity/temperature basis while $\{\psi_i(\mathbf{x})\}$ constitutes the pressure basis.

2.2 Galerkin formulation for the reduced-order model

Within the Galerkin framework of the preceding section it is relatively straightforward to develop a ROM. One defines a set of functions $\zeta_j(\mathbf{x}) \in \mathbf{X}^N \subset \mathbf{X}^N$, $\theta_j(\mathbf{x}) \in X^N \subset X^N$ such that the coarse-space (ROM) solution, is expressed as

$$\bar{\mathbf{u}}_c(\mathbf{x}) = \sum_{j=0}^N \zeta_j(\mathbf{x}) u_{c,j}, \quad \bar{T}_c(\mathbf{x}) = \sum_{j=0}^N \theta_j(\mathbf{x}) T_{c,j}. \quad (30)$$

For the ROM, we insert the expansions Eq. 30 into Eqs. 13–15 and require equality for all (\mathbf{v}, S) in \mathbf{Z}_0^N . In order to set the boundary conditions, we have augmented the trial (approximation) spaces \mathbf{X}^N and X^N with the lifting function $\zeta_0 = \mathbf{u}_0$ and $\theta_0 = T_0$. The corresponding test spaces, $\mathbf{X}_0^N := \{\zeta_j\}_{j=1}^N$ and $X_0^N := \{\theta_j\}_{j=1}^N$, satisfying homogeneous boundary conditions, as is standard for Galerkin formulation. The coarse space $\mathbf{Z}_0^N := (\mathbf{X}_0^N, X_0^N)$ is typically based on a linear combination of full spectral element solutions of Eqs. 21 and 22, such as snapshots at certain time-points, t^n , or solutions at various parametric values. Under these conditions and with a carefully chosen \mathbf{u}_0 , \mathbf{X}^N is a set of velocity-space functions that are (weakly) divergence-free and the pressure terms drop out of Eqs. 13 and 14. We also note that ζ_j and θ_j are *modal*, not nodal, basis functions. In this work, we consider proper orthogonal decomposition (POD) to construct the reduced-basis. The N -dimensional POD-space is the space that minimizes the averaged distance between the snapshot set and the N -dimensional subspace of the snapshots set in the H^1 semi-norm. Further details of the POD basis selection are provided in [18] and references therein.

The matrix form for the ROM is readily derived by constructing a pair of rectangular basis matrices, \mathbf{B} and B , having entries

$$\mathbf{B}_{ij} = \zeta_j(\mathbf{x}_i), \quad B_{ij} = \theta_j(\mathbf{x}_i), \quad (31)$$

where the \mathbf{x}_i s are the spectral element nodal points. The coarse-system matrices are $H_{c,\nu} = \mathbf{B}^T \mathbf{H} \mathbf{B}$ and $H_{c,\alpha} = B^T H_{\alpha} B$ and the governing system for the ROM becomes

$$H_{c,\nu} \hat{\mathbf{u}}_c^n = \mathbf{B}^T \hat{\mathbf{f}}(\bar{\mathbf{u}}_c^n, \bar{T}_c^n; \theta_g), \quad H_{c,\alpha} \hat{T}_c^n = B^T \hat{Q}(\bar{\mathbf{u}}_c^n, \bar{T}_c^n). \quad (32)$$

We refer Eq. 32 as Galerkin ROM (G-ROM) throughout the paper. The ROM coefficient vectors, $\hat{\mathbf{u}}_c^n = [u_{c,1}^n \dots u_{c,N}^n]^T$, $\hat{T}_c^n = [T_{c,1}^n \dots T_{c,N}^n]^T$, are determined by solving the $2N \times N$ linear systems. Note that the coefficients for the lifting functions are prescribed: $u_{c,0} = T_{c,0} = 1$. The initial coefficients for the ROM are obtained by projecting the initial condition onto the coarse space \mathbf{Z}_0^N with the H^1 semi-norm,

$$\hat{\mathbf{u}}_c^0 = \mathbf{B}^T \mathbf{A} \mathbf{u}^0, \quad \hat{T}_c^0 = B^T A T^0, \quad (33)$$

which follows from the fact that the columns of \mathbf{B} and B are, respectively, \mathbf{A} - and A -orthonormal, where $\mathbf{A} = \text{block-diagonal}(A)$. To recover the spectral element representation, we simply prolong the N -length vectors $\hat{\mathbf{u}}_c^n$ and \hat{T}_c^n with the set of basis functions and add it with the lifting functions \mathbf{u}_0 and T_0

$$\bar{\mathbf{u}}_c^n = \mathbf{B} \hat{\mathbf{u}}_c^n + \mathbf{u}_0, \quad \bar{T}_c^n = B \hat{T}_c^n + T_0. \quad (34)$$

The functional representations, $\bar{\mathbf{u}}_c^n(\mathbf{x})$ and $\bar{T}_c^n(\mathbf{x})$, are then obtained from Eqs. 8 and 9.

Next we consider the Galerkin ROM with Leray regularization using a spatial filter, following ideas presented in [19]. The approach simply requires regularizing the *advecting* field in the Navier-Stokes equations and energy equation through a low-pass filter function F (i.e., $\bar{\mathbf{u}}_{\text{filtered}} = F(\bar{\mathbf{u}})$). As noted in [33, 34], a small amount of regularization is sufficient to make gains in proving existence and uniqueness of the solution to the Navier-Stokes equations. Thus, Leray regularization is of interest both from a numerical (and physical) stabilization viewpoint and from a theoretical perspective.

The formulation of G-ROM with Leray regularization is shown in Eq. 35 and the only difference comparing to G-ROM Eq. 32 are the velocity data and thermal load.

$$H_{c,\nu} \hat{\mathbf{u}}_c^n = \mathbf{B}^T \hat{\mathbf{f}}_{\text{filtered}}(\bar{\mathbf{u}}_c^n, \bar{T}_c^n; \theta_g), \quad H_{c,\alpha} \hat{T}_c^n = B^T \hat{Q}_{\text{filtered}}(\bar{\mathbf{u}}_c^n, \bar{T}_c^n), \quad (35)$$

where $\hat{\mathbf{f}}_{\text{filtered}}(\bar{\mathbf{u}}_c^n, \bar{T}_c^n; \theta_g) = [\hat{f}_1 \hat{f}_2]^T_{\text{filtered}}$. We use the subscript filtered to denote the advecting field in the velocity data and thermal load are being filtered,

$$\hat{f}_{\text{filtered},m} = - \sum_{s=1}^k \left[\frac{\beta_s}{\Delta t} M \mathbf{u}_m^{n-s} + \alpha_s (C(\bar{\mathbf{u}}_{\text{filtered}}^{n-s}) \bar{\mathbf{u}}_m^{n-s} - g_m M \bar{T}_m^{n-s}) \right] - \nu A \mathbf{u}_{m,0}, \quad m = 1, 2. \quad (36)$$

and

$$\hat{\mathbf{Q}}_{\text{filtered}}(\bar{\mathbf{u}}^n, \bar{T}^n) = -\sum_{s=1}^k \left[\frac{\beta_s}{\Delta t} M \bar{T}^{n-s} + \alpha_s C(\bar{\mathbf{u}}_{\text{filtered}}^{n-s}) \bar{T}^{n-s} \right] - \alpha A \bar{T}_0. \quad (37)$$

In this work, we will be focusing on a PDE- (or differential-) based filter, which is characterized by a filter width, δ [35]. Following [19, 36], such filters are developed in a POD context as follows: Find $\bar{\mathbf{u}}_{\text{c,filter}} \in \mathbf{X}^N$ such that

$$((I - \delta^2 \nabla^2) \bar{\mathbf{u}}_{\text{c,filter}}^n, \zeta_j) = (\bar{\mathbf{u}}_c, \zeta_j), \quad \forall j = 1, \dots, N. \quad (38)$$

Besides the differential filter, one could also consider a more economic spatial filter, namely, a POD-projection (Proj) filter as discussed in [19]. In this case, one simply truncates the higher POD mode contributions when constructing $\bar{\mathbf{u}}_{\text{c,filtered}}^n$, just as one would do in a Fourier reconstruction [18].

Besides G-ROM and LDF-ROM, we also consider the constrained-evolution stabilization introduced in [9]. The idea behind this approach is to use information from the snapshot set to establish *a priori* limits on the ROM coefficients $\hat{\mathbf{u}}_c$ by replacing Eq. 32 with a constrained minimization problem. At each time-step, the coefficients satisfy

$$\begin{aligned} \hat{\mathbf{u}}_c^n &= \arg \min_{\substack{\hat{\mathbf{u}}_c \in \mathbb{R}^N \\ m_j \leq u_j^n \leq M_j}} \frac{1}{2} \|H_{c,y} \hat{\mathbf{u}}_c - \mathbf{B}^T \hat{\mathbf{f}}(\bar{\mathbf{u}}_c^n, \bar{T}_c^n; \theta_g)\|_{H^{-1}}^2, \\ &\text{s.t. } m_j \leq u_j^n \leq M_j. \end{aligned} \quad (39)$$

where the constraints m_j and M_j on the basis coefficients $u_{c,j}^n$, $j = 1, \dots, N$ are derived from the observation snapshot set. A constrained minimization problem for the thermal ROM coefficients, \hat{T}_c , is derived similarly. We denote Eq. 39 as C-ROM. Further implementation details can be found in [9, 18].

3 The solution reproduction problem

In this section, we consider the solution reproduction problems for $Ra = 2 \times 10^5$, 7×10^5 at $Pr = 7.2$ and $\theta_g = 90^\circ$, where the solutions are chaotic. We assess the performance of G-ROM, C-ROM and LDF-ROM introduced in Section 2.1 through the accuracy of the mean field and the QOIs. The mean field is computed by averaging the POD coefficients and reconstructing with the rectangular basis matrices \mathbf{B} and B . The QOIs are the mean Nu and Std(Nu), which are estimated through Eq. 6, and the mean TKE and mean temperature fluctuation, estimated through Eq. 7. The quantities are averaged over 500 CTUs.

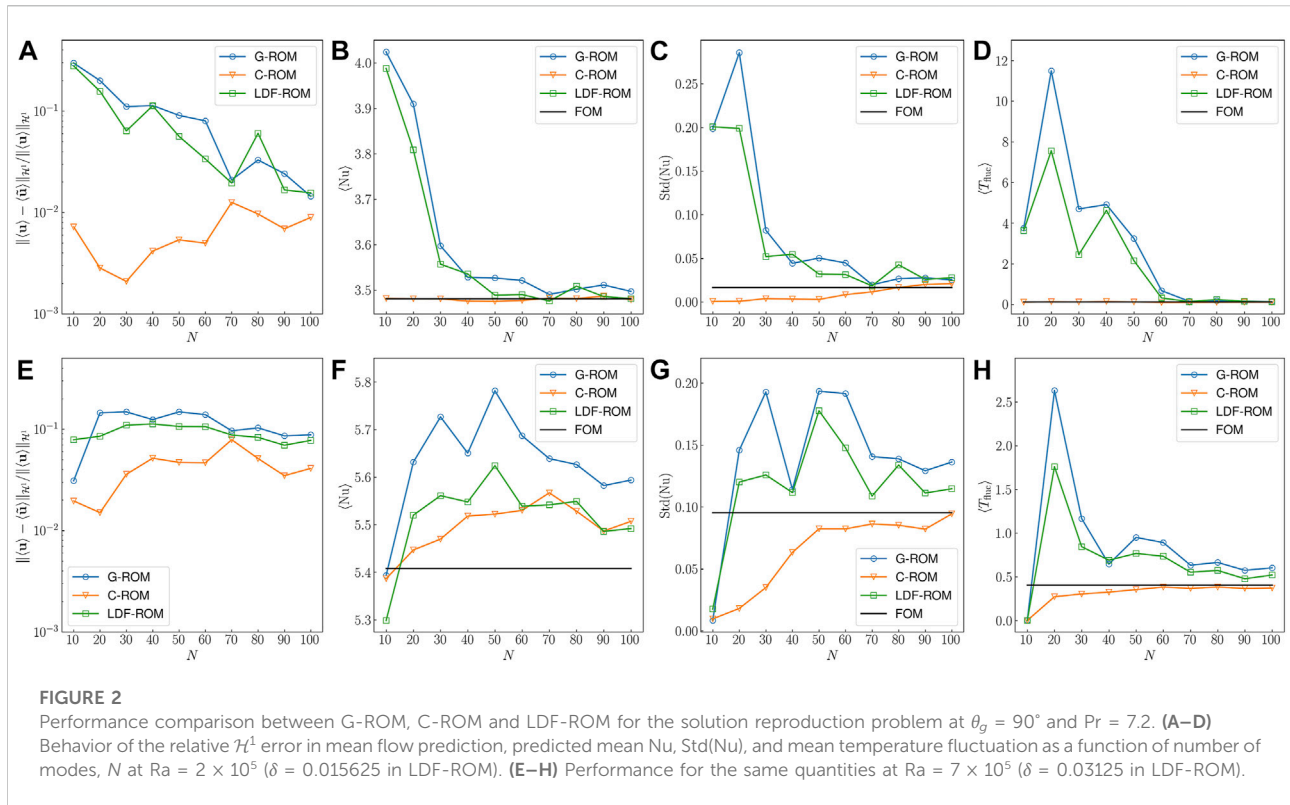
Although of limited practical interest, the solution reproduction problem is an important step towards the development of a MOR procedure for the parametric problem. The reproduction results for spatio-temporal chaotic flow are presented and discussed in Section 5.1.

3.1 Numerical results

Results for $Ra = 2 \times 10^5$ are shown in Figures 2A–D. The performances of G-ROM, C-ROM and LDF-ROM are indicated by blue, orange, green solid line. In LDF-ROM, the radius of the differential filter is $\delta = 0.015625$. Figure 2A shows the behavior of the relative \mathcal{H}^1 error in the mean flow prediction versus N . We observe less than 1% error in C-ROM with small N . The error in G-ROM and LDF-ROM decreases as N increases and eventually reaches 1% with $N = 100$. Similar trends for the mean Nu, Std(Nu) and mean temperature fluctuation are observed in Figures 2B–D. The FOM data is denoted as black solid line. For mean Nu prediction, we observe around 0.1% error in C-ROM for almost all N . The error in G-ROM and LDF-ROM decreases as N increases and LDF-ROM has error around 0.01% with $N = 100$. For Std(Nu) and mean temperature fluctuation prediction, we observe convergence in both QOIs with G-ROM and LDF-ROM. For C-ROM, the predictions are in good agreement with the FOM data for all values of N . Behaviors of the same quantities for $Ra = 7 \times 10^5$ are shown in Figures 2E–H. In LDF-ROM, the radius of the differential filter is $\delta = 0.03125$. We still observe convergence in the mean flow and QOIs but it is much slower because the flow is more chaotic. The error in the mean flow is around 10% in G-ROM and the prediction in QOIs are over-estimated and less accurate than for the other methods. With LDF-ROM, the predictions are slightly better. C-ROM is still the most effective and has 5% error in mean flow prediction. Although the prediction in mean Nu is only as good as the LDF-ROM, the Std(Nu) and mean temperature fluctuation are bounded from above for all values of N and converge to the correct value as N increases.

Note that the differential filter radius δ selected for the two Ra yields the best accuracy in mean flow among the five differential filter radius $\delta = 0.25, 0.125, 0.0625, 0.03125, 0.015625$. Besides, we find $\delta = 0.015625$ yield the best results at smaller Ra and as Ra increases, results with $\delta = 0.03125$ becomes better and are comparable with $\delta = 0.015625$ at $Ra = 7 \times 10^5$. The tendency is reasonable since the flow is more chaotic as Ra increases, therefore one should expect a larger δ to stabilize the flow.

From the results, we observe the mean flow and QOIs converge with N for all ROMs. With higher Ra , convergence in those quantities is much slower and a larger N is required to reach to the same accuracy as in the lower Ra case. Note that, because of the $O(N^3)$ online costs, requiring a large value of N for convergence might require off-line resources for timely simulation, which would greatly diminish the intrinsic advantage of the ROM/pMOR framework. This potentiality highlights the importance of stabilization methods. Indeed, we find that C-ROM is able to predict the mean flow and QOIs with a better accuracy with smaller N . On the other hand, although LDF-ROM is not as effective as C-ROM, and only slightly better



than G-ROM, it will play a key role in the pMOR presented in Section 4, especially in the parametric problem with θ_g variation.

4 The parametric problem

We turn now to study the performance of pMOR for the slot problem with G-ROM, C-ROM, and LDF-ROM. Two sets of parameterization are considered. With $Pr = 7.2$ fixed, we seek to estimate the solution and QOIs of Eqs. 13 and 15 for: 1) $\theta_g \in \mathcal{P} = [0^\circ, 180^\circ]$ at multiple Ra , and 2) $Ra \in \mathcal{P} = [2 \times 10^5, 7 \times 10^5]$ with $\theta_g = 90^\circ$. Throughout, we take the lifting functions to be the zero velocity field and the heat conduction solution.

For efficient selection of the pMOR anchor points we consider the POD-*h*Greedy algorithm proposed in [9] which combines POD in time with Greedy in parameter. The term Greedy refers to the optimization strategy of basing anchor point selection on the training point that exhibits the largest value in the error estimate. Error-indicated selection of the anchor points reduces the number of FOM solves and is thus critical for the feasibility of pMOR. Here, the error indicator corresponds to the dual norm of the residual associated with the time-averaged momentum and energy equations.

The section is organized as follows. In Section 4.1 we present the POD-*h*Greedy algorithm. In Section 4.2, we extend the time-averaged error indicator introduced in [9] to accommodate the

energy equation and Leray regularization. Finally, in Sections 4.3 and 4.4, we present the numerical results.

4.1 Proper orthogonal decomposition-*h*Greedy algorithm

In this section, we present the POD-*h*Greedy algorithm for the construction of the reduced spaces $\{\mathbf{X}_{0,\ell}^N, X_{0,\ell}^N\}_{\ell=1}^L$, and the partition $\{\mathcal{I}_\ell\}_{\ell=1}^L$ of \mathcal{P} , based on the results of L full-order simulations associated with the parameters p_1^*, \dots, p_L^* . The algorithm is similar to the one in [9] but with extensions for thermal fields.

To begin, we introduce the discretized parameter space $\mathcal{P}_{\text{train}} = \{p_i\}_{i=1}^{n_{\text{train}}}$, the integers L which fix the maximum number of offline solves, the integer $n_{\text{cand}} < L$, which is the number of ROM evaluations performed online for a given value of the parameters, and an error indicator Δ . The error indicator takes as input sequences $\{\bar{\mathbf{u}}_c^n\}_{n=0}^J$ and $\{\bar{T}_c^n\}_{n=0}^J$ and the value of the parameter, p^* , and returns an estimate of the error in the prediction of the mean flow. We formally present the indicator in Section 4.2.

Algorithm 1 presents the computational procedure for both offline and online stage. The offline procedure starts with an anchor point that could either be selected randomly from the training space $\mathcal{P}_{\text{train}}$ or be user specified. At each iteration ℓ , a

FOM simulation at the anchor point p_ℓ^* is conducted and returns a set of snapshots. The snapshot set is then processed through POD and returns the first N orthonormalized POD modes. The value of N is determined by reproduction problem at the anchor point. The ROM and the error indicator, Δ_ℓ , are then built with the reduced spaces $\mathbf{X}_{0,\ell}^N$ and $X_{0,\ell}^N$. The coefficients and the error estimates are then computed for each $p \in \mathcal{P}_{\text{train}}$ and the next anchor point is identified as the parameter that has the maximum value in the current (including previous) error estimate. The procedure starts again with the new selected anchor point. If the error indicator is sufficiently small over all points in $\mathcal{P}_{\text{train}}$ or the procedure reaches the maximum number of FOM solves L , the offline stage terminates.

Given the ROM/anchor point data $(\mathbf{X}_{0,\ell}^N, X_{0,\ell}^N, p_\ell^*)$ for $\ell = 1, \dots, L$, and error indicator, Δ , the h Greedy online stage starts with finding the n_{cand} candidate anchor points nearest to the test parameter p . The ROMs associated to the candidate anchor points are then used to compute the coefficients and error estimate at p . The coefficients are then returned based on the ROM that has smallest error estimate. The POD- h Greedy approach is analogous to h -refinement in the finite element method in that the POD bases are not shared between anchor points. Convergence is therefore expected to be linear in the distance from the nearest anchor point.

Offline stage: $\{(\mathbf{X}_{0,\ell}^N, X_{0,\ell}^N)_{\ell=1}^L\} = \text{Offline}(\mathcal{P}_{\text{train}}, L, \Delta)$.
Inputs: $\mathcal{P}_{\text{train}} = \{p_i\}_{i=1}^{n_{\text{train}}}$ = discretized parameter space, L = maximum number of offline solves, Δ = error indicator defined in (44).
Output: $\{(\mathbf{X}_{0,\ell}^N, X_{0,\ell}^N, p_\ell^*)\}_{\ell=1}^L$ = reduced space/anchor point pairs.
1: $p_1^* = \text{smallest } p \text{ in } \mathcal{P}_{\text{train}}$.
2: **for** $\ell = 1, \dots, L$ **do**
3: $\{(\mathbf{u}^k(p_\ell^*), T^k(p_\ell^*))_{k=1}^K\}$ = DNS-solver (p_ℓ^*).
4: $\{(\zeta_{n,\ell}, \theta_{n,\ell})_{n=1}^N\}$ = POD ($\{(\mathbf{u}^k(p_\ell^*), T^k(p_\ell^*))_{k=1}^K\}$).
5: Define $\mathbf{X}_{0,\ell}^N = \text{span}\{\zeta_{n,\ell}\}_{n=1}^N$, $X_{0,\ell}^N = \text{span}\{\theta_{n,\ell}\}_{n=1}^N$, build the ROM operators and error indicator Δ_ℓ .
6: **for** $j = 1, \dots, n_{\text{train}}$ **do**
7: $\{(\mathbf{u}_c^j(p_i), T_c^j(p_i))_{j=0}^J\}$ = ROM-solver ($p_i, \mathbf{X}_{0,\ell}^N, X_{0,\ell}^N$).
8: Compute the weak dual norm Δ_ℓ at p_i .
9: **end for**
10: $p_{\ell+1}^* = \arg \max_{p \in \mathcal{P}_{\text{train}}} \min\{\Delta_\ell(p)\}_{\ell=1}^L$.
11: **end for**
Online stage: $\{(\mathbf{u}_c^j, T_c^j)\}_j = \text{Online}(\{(\mathbf{X}_{0,\ell}^N, X_{0,\ell}^N, p_\ell^*)\}_{\ell=1}^L, \Delta, n_{\text{cand}}, p)$.
Inputs: $\{(\mathbf{X}_{0,\ell}^N, X_{0,\ell}^N, p_\ell^*)\}_{\ell=1}^L$ = reduced space/anchor point pairs, n_{cand} = online ROM evaluations, p = input parameter.
Output: $\{(\mathbf{u}_c^j, T_c^j)\}_j$ = solution estimate.
1: Find the n_{cand} candidate anchors nearest to p : $p_1^*, \dots, p_{n_{\text{cand}}}^*$.
2: **for** $i = 1, \dots, n_{\text{cand}}$ **do**
3: $\{(\mathbf{u}_{c,(i)}^j, T_{c,(i)}^j)_{j=0}^J\}$ = ROM-solver ($p_i^*, \mathbf{X}_{0,(i)}^N, X_{0,(i)}^N$).
4: Compute the error estimate $\Delta_{(i)}(p)$.
5: **end for**
6: Return $\{(\mathbf{u}_c^j, T_c^j)\}_j = (\mathbf{u}_{c,(i^*)}^j, T_{c,(i^*)}^j)_j$, where i^* is the minimizer of $\{\Delta_{(i)}(p)\}_i$.

Algorithm 1. POD- h Greedy algorithm for the construction of $\{\mathbf{X}_{0,\ell}^N, X_{0,\ell}^N, \mathcal{I}_\ell\}$.

Another strategy is the POD- p Greedy algorithm, following the definitions of [37], as first proposed in [38] and analyzed in [39]. The algorithm combines data from different parameters to generate a single reduced basis set that covers the entire parameter space \mathcal{P} . The procedure is similar to Algorithm 1 but with few differences:

1) The reduced bases are shared between anchor points. POD is still used to construct the new basis but the collected snapshot set is projected onto the orthogonal complement of the existing basis.

2) In the online/training stage, only one ROM is used instead of a set of ROMs and there is no need to check for the nearest anchor points.
3) The anchor point is selected based on the single error estimate Δ in current iteration, rather than the individual estimates for each ROM.

Although it has a better convergence rate than POD- h Greedy, POD- p Greedy can easily fail for unsteady problems. Combining modes at different anchor points, especially ones whose solution exhibits different physics, can easily lead to instability and deteriorate the performance, as noted in [9]. Moreover, stabilizations that work for POD- h Greedy can fail in the POD- p Greedy approach. For example, in C-ROM, it is not clear how to construct the constraints for the combined basis. A naive approach is to apply POD to all the snapshots at anchor points. However, this approach is inefficient and can be limited by the computer storage requirements during the offline phase. Leray regularization with the projection filter (i.e., trivially truncated basis set for the advector) is also limited since the combined basis is no longer ordered in a Fourier-like, energy-decaying, sequence. To address this, one could apply POD to all the snapshots that have been collected but this approach is again limited by the storage and therefore not practical. An alternative is to consider DF filter, denoted as LDF-ROM here. Once the radius δ is specified, it will filter right amount of energy in each basis.

4.2 A time-averaged error indicator

In this section, we extend the time-averaged error indicator proposed in [9] to accommodate the energy equation and Leray regularization. The error indicator is based on the dual norm of the discrete time-averaged residual. Given the ROM solution sequence $\{\bar{\mathbf{u}}_c^n\}_{n=J_0+1}^J$ and $\{\bar{T}_c^n\}_{n=J_0+1}^J$ and the parameters of interest $\underline{p} = (\nu, \theta_g, \alpha)$, the discrete time-averaged residual for velocity and temperature are defined as:⁴

$$\langle R_u \rangle (\{\bar{\mathbf{u}}_c^n\}_{n=J_0+1}^J, \mathbf{v}; \nu, \theta_g) = \frac{1}{J - J_0} \sum_{n=J_0+1}^J r_u(\bar{\mathbf{u}}_c^n, \mathbf{v}; \nu, \theta_g), \quad \forall \mathbf{v} \in \mathbf{V}_{\text{div}}, \quad (40)$$

$$\langle R_T \rangle (\{\bar{T}_c^n\}_{n=J_0+1}^J, S; \alpha) = \frac{1}{J - J_0} \sum_{n=J_0+1}^J r_T(\bar{T}_c^n, S; \alpha), \quad \forall S \in X_0^N, \quad (41)$$

⁴ In [16], the velocity and temperature residuals are coupled because the velocity-temperature solutions are obtained through a coupled Newton's method. Here, we do not couple the residuals because we solve Eqs. 21 and 22 separately.

where $r_u(\bar{\mathbf{u}}_c^n, \mathbf{v}; \nu, \theta_g) \in \mathbf{V}'_{\text{div}}$ (dual space of \mathbf{V}_{div}) and $r_T(\bar{T}_c^n, S; \alpha) \in X_0^N$ (dual space of X_0^N) are the residual associated with Eq. 32 at time t^n and defined as

$$r_u(\bar{\mathbf{u}}_c^n, \mathbf{v}; \nu, \theta_g) = \sum_{s=1}^3 \alpha_s \left[(\mathbf{v}, \mathbf{g}(\theta_g) \bar{T}_c^{n-s}) - c(\mathbf{v}, \bar{\mathbf{u}}_c^{n-s}, \bar{\mathbf{u}}_c^{n-s}) \right] - \sum_{s=0}^3 \frac{\beta_s}{\Delta t} (\mathbf{v}, \bar{\mathbf{u}}_c^{n-s}) - \nu a(\mathbf{v}, \bar{\mathbf{u}}_c^n), \tag{42}$$

$$r_T(\bar{T}_c^n, S; \alpha) = - \sum_{s=1}^3 \alpha_s c(S, \bar{\mathbf{u}}_c^{n-s}, \bar{T}_c^{n-s}) - \sum_{s=0}^3 \frac{\beta_s}{\Delta t} (S, T_c^{n-s}) - \alpha(S, \bar{T}_c^n). \tag{43}$$

Note for simplicity, we assume only BDF3/EXT3 is used for time discretization in Eqs. 42 and 43. Besides, the residual is defined over $\{\mathcal{V}\}'_{\text{div}} := \{\mathbf{v} | \mathbf{v} \in \mathbf{X}_0^N, \nabla \cdot \mathbf{v} = 0\}$ and X_0^N , rather than $\mathbf{X}_0^N \subset \mathcal{H}_0^1$ and X_0^N , because we measure our reduced-basis error relative to the FOM.

We define the *time-averaged error indicator*, $\Delta: \otimes_{n=J_0}^J \mathbf{X}^N \times X^N \times \mathcal{P} \rightarrow \mathbb{R}_+$, as follows:

$$\Delta(\{\bar{\mathbf{u}}_c\}_{n=J_0}^J, \{\bar{T}_c\}_{n=J_0}^J; \underline{\mathbf{p}}) := \sqrt{\| \langle R_u \rangle(\{\bar{\mathbf{u}}_c\}_{n=J_0}^J, \cdot; \nu, \theta_g) \|_{\mathbf{V}'_{\text{div}}}^2 + \| \langle R_T \rangle(\{\bar{T}_c\}_{n=J_0}^J, \cdot; \alpha) \|_{X_0^N}^2}. \tag{44}$$

The residuals Eqs. 42 and 43 can be further expressed in the matrix-vector form since the spaces $\{\mathcal{V}\}'_{\text{div}}$ and X_0^N are finite dimensional,

$$r_u(\bar{\mathbf{u}}_c^n, \mathbf{v}; \nu, \theta_g) = \underline{\mathbf{v}}^T \underline{\mathbf{r}}_u^n = \underline{\mathbf{v}}^T \hat{\mathbf{f}}(\bar{\mathbf{u}}_c^n, \bar{T}_c^n; \theta_g) - \underline{\mathbf{v}}^T \mathbf{H} \mathbf{B} \hat{\underline{\mathbf{u}}}_c^n, \quad \forall \underline{\mathbf{v}} \in \mathbb{R}^{2N}, \tag{45}$$

$$r_T(\bar{T}_c^n, S; \alpha) = \underline{S}^T \underline{\mathbf{r}}_T^n = \underline{S}^T \hat{\underline{\mathbf{Q}}}(\bar{\mathbf{u}}_c^n, \bar{T}_c^n) - \underline{S}^T H_\alpha B \hat{\underline{\mathbf{T}}}_c^n, \quad \forall \underline{S} \in \mathbb{R}^N. \tag{46}$$

The matrix-vector version of the discrete time-averaged residual Eqs. 40 and 41 is then expressed as

$$\langle R_u \rangle(\{\bar{\mathbf{u}}_c\}_{n=J_0}^J, \mathbf{v}; \nu, \theta_g) = \underline{\mathbf{v}}^T \underline{\mathbf{R}}_u = \underline{\mathbf{v}}^T \left(\frac{1}{J-J_0} \sum_{n=J_0+1}^J \hat{\mathbf{f}}(\bar{\mathbf{u}}_c^n, \bar{T}_c^n; \theta_g) - \mathbf{H} \mathbf{B} \hat{\underline{\mathbf{u}}}_c^n \right), \tag{47}$$

$$\langle R_T \rangle(\{\bar{T}_c\}_{n=J_0}^J, S; \alpha) = \underline{S}^T \underline{\mathbf{R}}_T = \underline{S}^T \left(\frac{1}{J-J_0} \sum_{n=J_0+1}^J \hat{\underline{\mathbf{Q}}}(\bar{\mathbf{u}}_c^n, \bar{T}_c^n) - H_\alpha B \hat{\underline{\mathbf{T}}}_c^n \right), \tag{48}$$

$\forall \underline{\mathbf{v}} \in \mathbb{R}^{2N}$ and $\forall \underline{S} \in \mathbb{R}^N$. The norm of the residual is closely related to the error and it is tempting to use $\| \underline{\mathbf{R}}_u \|_2$ and $\| \underline{\mathbf{R}}_T \|_2$ to estimate the error. However, this is not correct since $\langle R_u \rangle(\{\bar{\mathbf{u}}_c\}_{n=J_0}^J, \cdot; \nu, \theta_g): \mathbf{V}_{\text{div}} \rightarrow \mathbb{R}$ and $\langle R_T \rangle(\{\bar{T}_c\}_{n=J_0}^J, \cdot; \alpha): X_0^N \rightarrow \mathbb{R}$ are bounded *linear* functionals whose size is appropriately measured through the dual norm:

$$\| \langle R_u \rangle(\{\bar{\mathbf{u}}_c\}_{n=J_0}^J, \cdot; \nu, \theta_g) \|_{\{\mathcal{V}\}'_{\text{div}}} = \sup_{\mathbf{v} \in \mathbf{V}_{\text{div}}} \frac{\langle R_u \rangle(\{\bar{\mathbf{u}}_c\}_{n=J_0}^J, \mathbf{v}; \nu, \theta_g)}{\| \mathbf{v} \|_{\mathbf{V}_{\text{div}}}}, \tag{49}$$

$$\| \langle R_T \rangle(\{\bar{T}_c\}_{n=J_0}^J, \cdot; \alpha) \|_{X_0^N} = \sup_{S \in X_0^N} \frac{\langle R_T \rangle(\{\bar{T}_c\}_{n=J_0}^J, S; \alpha)}{\| S \|_{X_0^N}}. \tag{50}$$

Thanks to the Riesz representation theorem, there exist a unique $\langle \hat{\mathbf{R}}_u \rangle \in \mathbf{V}_{\text{div}}$ and $\langle \hat{R}_T \rangle \in X_0^N$ such that

$$\langle \hat{\mathbf{R}}_u \rangle, \mathbf{v} \Big|_{\mathbf{V}_{\text{div}}} = \langle R_u \rangle(\{\bar{\mathbf{u}}_c\}_{n=J_0}^J, \mathbf{v}; \nu, \theta_g), \quad \forall \mathbf{v} \in \mathbf{V}_{\text{div}}, \tag{51}$$

$$\langle \hat{R}_T \rangle, S \Big|_{X_0^N} = \langle R_T \rangle(\{\bar{T}_c\}_{n=J_0}^J, S; \alpha), \quad \forall S \in X_0^N. \tag{52}$$

It thus follows that

$$\| \langle R_u \rangle(\{\bar{\mathbf{u}}_c\}_{n=J_0}^J, \cdot; \nu, \theta_g) \|_{\mathbf{V}'_{\text{div}}} = \| \langle \hat{\mathbf{R}}_u \rangle \|_{\mathbf{V}_{\text{div}}}, \tag{53}$$

$$\| \langle R_T \rangle(\{\bar{T}_c\}_{n=J_0}^J, \cdot; \alpha) \|_{X_0^N} = \| \langle \hat{R}_T \rangle \|_{X_0^N}. \tag{54}$$

Equations 51 and 52 allows one to compute the Riesz representers $\langle \hat{\mathbf{R}}_u \rangle$ and $\langle \hat{R}_T \rangle$ and Eqs. 53 and 54 allows one to evaluate the dual norm of the residual through Riesz representation without computing the supremum.

In practice, determination of the Riesz representers, $\langle \hat{\mathbf{R}}_u \rangle$ and $\langle \hat{R}_T \rangle$, is relatively straightforward because the coarse (i.e., ROM) and truth (FOM) representations live in finite-dimensional spaces, meaning that there is a direct linear-algebra problem to be solved for the Riesz representers. Expanding Eqs. 51 and 52, we have the corresponding linear algebra statement,

$$\begin{bmatrix} \mathbf{A} & -\mathbf{D}^T \\ -\mathbf{D} & 0 \end{bmatrix} \begin{pmatrix} \langle \hat{\mathbf{R}}_u \rangle \\ \underline{\mathbf{p}} \end{pmatrix} = \begin{pmatrix} \underline{\mathbf{R}}_u \\ 0 \end{pmatrix}, \tag{55}$$

$$\mathbf{A} \langle \hat{R}_T \rangle = \underline{\mathbf{R}}_T. \tag{56}$$

Here, \mathbf{A} corresponds to \mathbf{H} introduced in Eq. 23 with $\beta_0 = 0$ and $\nu = 1$. We remark that the essential difference between the velocity and temperature representers is that the former satisfies the divergence-free constraint by virtue of the 2×2 block system in Eq. 55. Evaluation of the error indicator Δ entails solving Eqs. 55 and 56, computing the corresponding \mathcal{H}^1 norms of the outputs, and ultimately using these results in Eq. 44.

While use of the direct approach requires access to the FOM machinery in order to generate an error indicator, we note that such access is readily available during the pMOR training/construction phase. The advantage of this approach is that the number of Stokes/Poisson solves scales as the number of points in the training space, which is typically less than N^2 . The other is through the offline/online computational decomposition which takes the advantage of the affine decomposition and expands the residual. By expanding the residuals $\langle R_u \rangle$ and $\langle R_T \rangle$, $2(N+1)^2 + 6(N+1)$ linear functionals are derived, where $2(N+1)^2$ is due to the convection term in the Navier-Stokes and energy equations. Applying the Riesz representation theorem to each linear functional, we end up solving $2(N+1)^2 + 6(N+1)$ Riesz representers, where $(N+1)^2 + 4(N+1)$ of them are solved through Stokes problems and $(N+1)^2 + 2(N+1)$ of them are solved through Poisson problems. Note that

the Riesz representers must be stored in order to accomplish the decomposition and each is a vector of size \mathcal{N} since it lives in the FOM space. For example, if $N = 60$, one would have to store at least 7,200 vectors of size \mathcal{N} which can be prohibitive, even for large multicore workstations. Even if there is no storage limitation, the offline cost is quite high when N is large as it scales quadratically. Once it is done, the online cost is $\mathcal{O}(N^2J + N^4)$, where $\mathcal{O}(N^2J)$ is to solve Eq. 32 and $\mathcal{O}(N^4)$ is required to compute the error estimate. Further details of the decomposition are provided in [9].

4.2.1 Time-averaged error indicator with Leray regularization

The integration of the time-averaged error indicator Δ with Leray regularization is rather straightforward. Recall the difference between G-ROM Eq. 32 and LDF-ROM Eq. 35 is simply the advecting field being filtered. Hence, the residuals $r_u(\bar{\mathbf{u}}_c^n, \mathbf{v}; \nu, \theta_g)$ and $r_T(\bar{T}_c^n, S; \alpha)$ for all $n = J_0 + 1, \dots, J$ are simply modified with the filtered advecting field,

$$r_u(\bar{\mathbf{u}}_c^n, \mathbf{v}; \nu, \theta_g) = \sum_{s=1}^3 \alpha_s \left[(\mathbf{v}, \mathbf{g}(\theta_g) \bar{T}_c^{n-s}) - c(\mathbf{v}, \bar{\mathbf{u}}_{c,\text{filtered}}^{n-s}, \bar{\mathbf{u}}_c^{n-s}) \right] - \sum_{s=0}^3 \frac{\beta_s}{\Delta t} (\mathbf{v}, \mathbf{u}_c^{n-s}) - \nu a(\mathbf{v}, \bar{\mathbf{u}}_c^n), \quad (57)$$

$$r_T(\bar{T}_c^n, S; \alpha) = - \sum_{s=1}^3 \alpha_s c(S, \bar{\mathbf{u}}_{c,\text{filtered}}^{n-s}, \bar{T}_c^{n-s}) - \sum_{s=0}^3 \frac{\beta_s}{\Delta t} (S, T_c^{n-s}) - \alpha a(S, \bar{T}_c^n), \quad (58)$$

for all $\mathbf{v} \in \mathbf{V}_{\text{div}}$ and $S \in X_0^{\mathcal{N}}$. The corresponding time-averaged error indicator Δ is then defined based on the modified residuals Eqs. 57 and 58.

4.3 Parametric model order reduction results: θ_g variation

In this section, we consider the parametric problem parameterized with θ_g at $\text{Pr} = 7.2$. The problem has three characteristics: bifurcation, spatio-temporal chaos over a certain range of θ_g , and a solution manifold that is a blend of steady and unsteady solutions. To identify the major pMOR challenges for this case, three values of Ra are considered:

- 1) $\text{Ra} = 1 \times 10^4$ where the FOM is steady except at $\theta_g = 30^\circ$.
- 2) $\text{Ra} = 8 \times 10^4$ where the FOM is unsteady for $\theta_g \in [0^\circ, 70^\circ]$ and steady for $\theta_g \in [80^\circ, 180^\circ]$.
- 3) $\text{Ra} = 3 \times 10^5$ where the FOM is unsteady for $\theta_g \in [0^\circ, 120^\circ]$ and steady for $\theta_g \in [130^\circ, 180^\circ]$.

In each case, the ROM is constructed through Algorithm 1. In order to assess performance, we generate FOM data for $\theta_g = 0^\circ, 10^\circ, \dots, 180^\circ$ ($n_{\text{train}} = 19$ datapoints). The FOM solution is obtained by solving Eqs. 2 and 3. For parameters where the problem is steady, the solution and the Nu are collected after the

solution difference between ten time steps is less than 10^{-6} . For unsteady problems, the mean flow, mean Nu , $\text{Std}(\text{Nu})$, mean TKE and mean temperature fluctuation are averaged over 500 CTUs after the solution has reached a statistically steady state.

Although not shown here, we remark that at $\text{Ra} = 1 \times 10^3$, the problem is steady with a bifurcation at $\theta_g = 20^\circ$. In this case, either h - or p -Greedy with residual dual-norm base error indicator accurately estimates the solution and QOIs over the parameter space $\theta_g \in [0^\circ, 180^\circ]$.

4.3.1 $\text{Ra} = 1 \times 10^4$

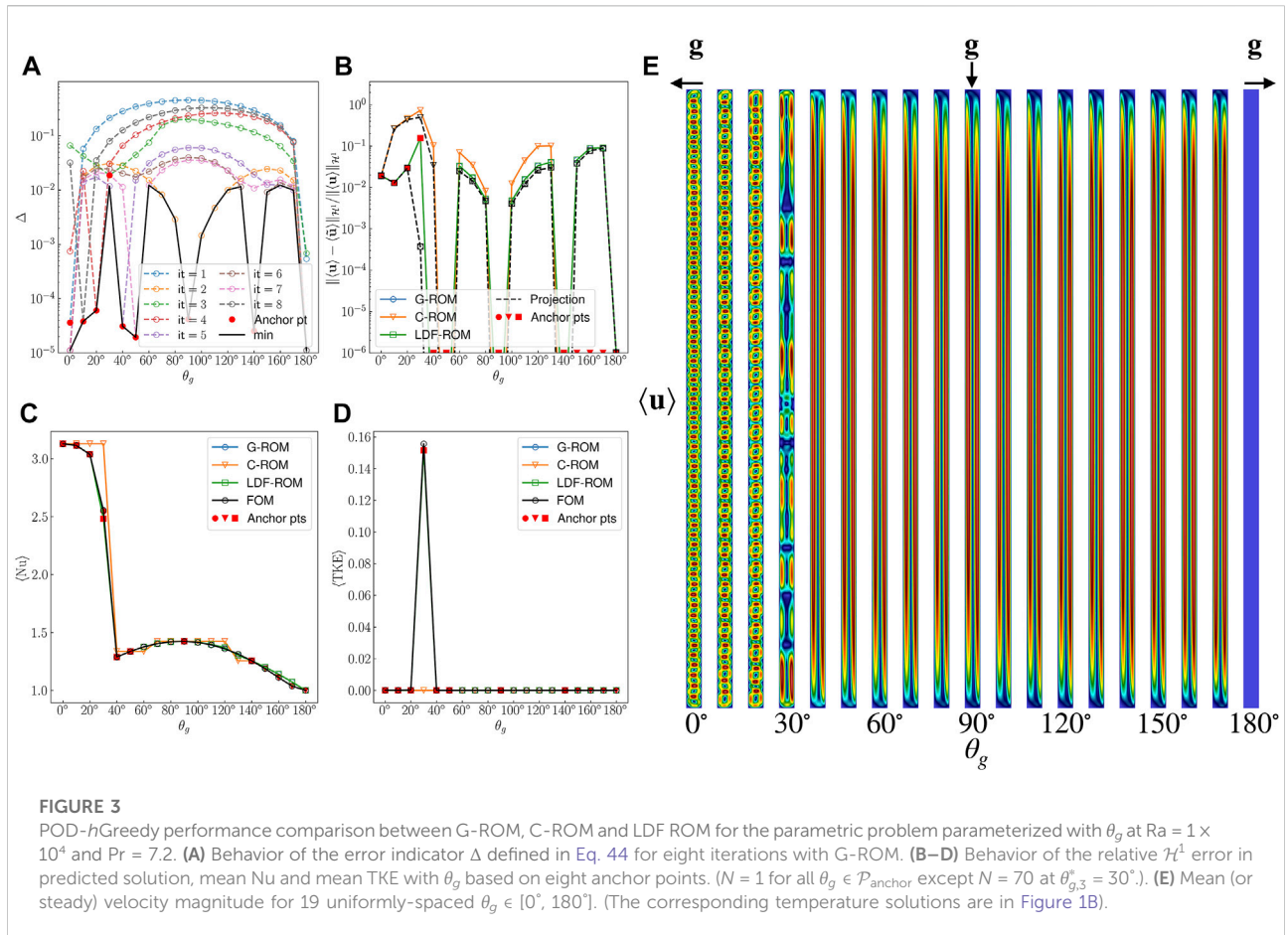
To examine the feasibility of the pMOR (Algorithm 1) in the unsteady case, we begin with $\text{Ra} = 1 \times 10^4$, in which only one unsteady solution is introduced at $\theta_g = 30^\circ$.

Figure 3E shows the steady (or mean) velocity magnitude for 19 uniformly-space training points $\theta_g = 0^\circ, 10^\circ, \dots, 180^\circ$. The corresponding temperature distributions are in Figure 1B. At $\theta_g = 180^\circ$, we observe no flow and the temperature is simply the conduction solution. As θ_g decreases, we observe slot convection and then about $\theta_g = 40^\circ$ there is a bifurcation to the wavy flow and rolls in the velocity. Moreover, we observe spatial-temporal chaos at $\theta_g = 30^\circ$. Figures 3A–D show the results of the application of Algorithm 1 for the construction of the G-ROM, C-ROM and LDF-ROM for the pMOR. The algorithm starts with $\theta_{g,1}^* = 0^\circ$ and is performed with $L = 8$ iterations.

Figure 3A demonstrates the selection process of anchor points (denoted by red circles) for the G-ROM case. We briefly walk through the process: At the first iteration, the error estimate $\Delta_1(\theta_g)$ for $\theta_g \in \mathcal{P}_{\text{train}}$ is computed (blue dashed line). With the largest error estimate, $\theta_g = 90^\circ$ is then identified as the second anchor point. The third anchor point is then selected from $\theta_g \in \mathcal{P}_{\text{train}}$ which maximizes the error estimate $\Delta_{1,2}(\theta_g) := \min\{\Delta_1(\theta_g), \Delta_2(\theta_g)\}$ over $\mathcal{P}_{\text{train}}$. (We reiterate that minimizing over the individual error estimates is a property of the h -Greedy process—there is not a single unifying error estimate as is the case for p -Greedy.) The process continues until the error estimate reaches the desired tolerance or the number of offline solves reaches its maximum. The black solid line denotes the minimum of all error estimates computed up to current iteration. In this case, it represents $\min\{\Delta_1(\theta_g), \dots, \Delta_8(\theta_g)\}$. Note that the error estimate at $\theta_g = 180^\circ$ in each model $\Delta_\ell(\theta_g = 180^\circ)$ is small due to the choice of lift function.

From Figure 3A, we observe the error estimate is small at anchor points where the problem is steady. On the other hand, although the error estimate $\Delta_3(\theta_g)$ (green line) is small at $\theta_{g,3}^* = 30^\circ$ compared to other points in $\mathcal{P}_{\text{train}}$, it is not as small as the estimate at other anchor points. Because of the unsteadiness, it can't reach the same magnitude as in the steady cases.

Following this procedure for the other cases, we present models results for the G-ROM, C-ROM and LDF-ROM cases, denoted respectively blue, orange, green solid lines in Figures 3B–D. The behavior of the relative \mathcal{H}^1 error in the predicted



solution is shown in Figure 3B. The corresponding Galerkin projection is denoted as the black dashed line. We found that G-ROM and LDF-ROM have a similar performance: the error in the solution is nearly identical to the Galerkin projection for cases where the problem is steady, including those that are not in the $\mathcal{P}_{\text{anchor}}$. The maximum error is at $\theta_g = 30^\circ$ where the problem is unsteady. Both methods have around 15% error in the mean flow. Note that at $\theta_g = 30^\circ$, the number of modes N is carefully selected since the ROM diverges after certain N due to the spatio-temporal chaos (15% error with $N = 70$, 17% error with $N = 80$ and 23% error with $N = 90$). On the other hand, the mean solution prediction made by C-ROM has 73% in maximum error and in order for C-ROM to reach same accuracy as in G-ROM and LDF-ROM, two more iterations are required. Already, with this modest $Ra = 1 \times 10^4$, we find C-ROM is less efficient than G-ROM and LDF-ROM. The pMOR behavior for mean Nu and mean TKE are shown along with the FOM results in Figures 3C,D. Again, G-ROM and LDF-ROM are able to make accurate predictions while C-ROM has maximum 22% error in mean Nu and in particular is unable to capture the peak in mean TKE at $\theta_g = 30^\circ$.

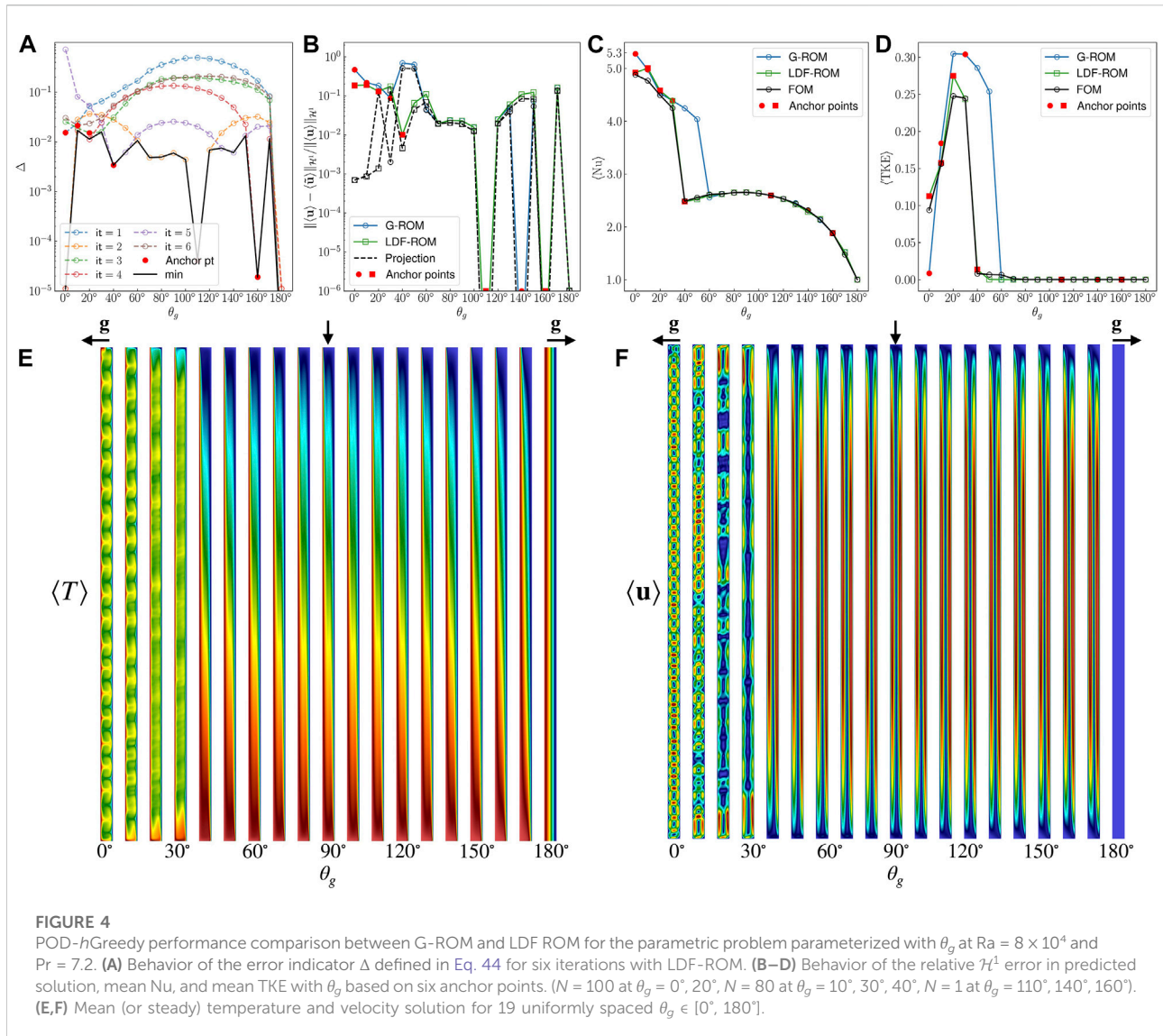
Before closing this section, we highlight some observations with respect to the solution manifolds.

- 1) The solutions at $\theta_g \in [0^\circ, 30^\circ]$ are Rayleigh-Bénard with differing numbers of rolls, analogous to orthogonal sine and cosine functions at different wave numbers. Therefore, there is little hope in reproducing the solution except at selected anchor points. QOI's such as mean Nu, however, are less sensitive to precise mean flow fields and are therefore more tractable.
- 2) At $\theta_g = 170^\circ$ and $\theta_g = 180^\circ$ the thermal metrics are not too different despite the $\mathcal{O}(1)$ difference in velocity solutions.

The first issue is resolved by the error indicator picking $\theta_g \in [0^\circ, 30^\circ]$ as anchor points. The second issue can be a source of error as Ra increases. With $\theta_g = 160^\circ$ as an anchor point and the solution at $\theta_g = 180^\circ$ as the lift function, the error at $\theta_g = 170^\circ$ is 9% for $Ra = 1 \times 10^4$, 16% for $Ra = 8 \times 10^4$, and 19% for $Ra = 3 \times 10^5$.

4.3.2 $Ra = 8 \times 10^4$

Figure 4 shows pMOR results analogous to Figure 3 for the case $Ra = 8 \times 10^4$. Here, we consider only the G-ROM and

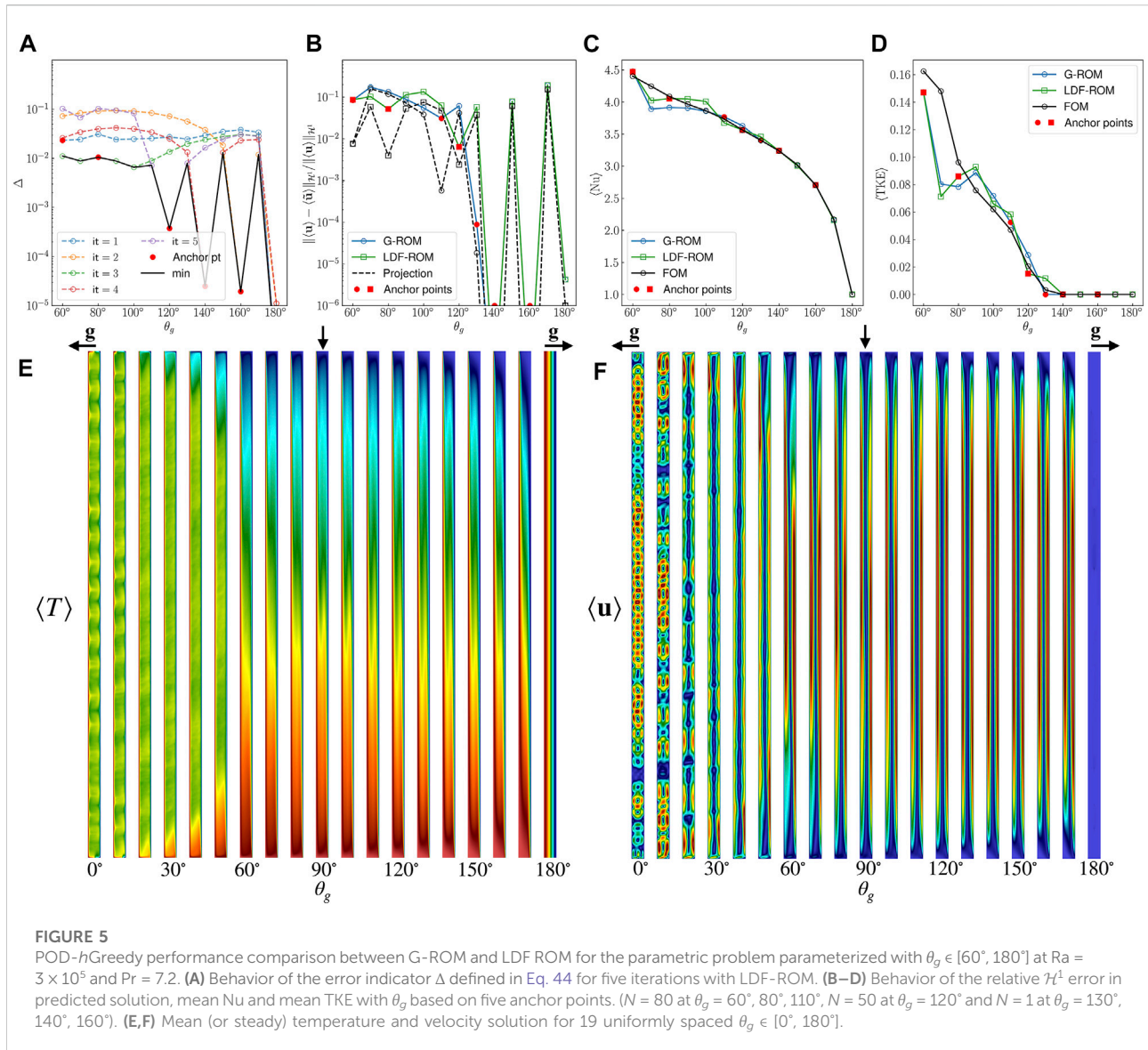


LDF-ROM. The algorithm starts with $\theta_{g,1}^* = 0^\circ$ and terminates at $L = 6$ iterations. Figures 4E,F show the FOM mean (or steady) temperature and velocity solution at the training points. (In an actual pMOR, these FOMs would of course not be computed *a priori*.) With this Ra, the bifurcation occurs at $\theta_g = 40^\circ$. Moreover, we observe spatio-temporal chaos for $\theta_g \in [0^\circ, 30^\circ]$ with the lower values being more chaotic.

The anchor-point selection process with LDF-ROM is demonstrated in Figure 4A. Starting with $\theta_{g,1}^* = 0^\circ$ the peak error in first iteration is at 110° , which is chosen to be $\theta_{g,2}^*$, and so on. Again, we find the error indicator is small at anchor points where the problem is steady and that it is larger where it is unsteady ($\theta_g \in [0^\circ, 60^\circ]$). Nonetheless, the error indicator is still able to identify where solution changes rapidly and select most of the anchor points in the region $[0^\circ, 40^\circ]$.

The behaviors of the relative \mathcal{H}^1 error in the predicted solution with θ_g using G-ROM and LDF-ROM are shown in Figure 4B. For $\theta_g \in [80^\circ, 180^\circ]$, where the solution is steady, we find the estimation is almost identical to the Galerkin projection in both models. On the other hand, for $\theta_g \in [0^\circ, 70^\circ]$, where the solution is unsteady, we find the error at anchor points $\theta_g = 0^\circ, \theta_g = 10^\circ$ is large due to the spatio-temporal chaos. The maximum error is around 19% at $\theta_g = 10^\circ$ with LDF-ROM while 69% at $\theta_g = 40^\circ$ with G-ROM. Although the maximum error in G-ROM can be reduced by further iterations of the algorithm, the error will eventually be dominated by the high reproduction error arising from spatio-temporal chaos at $\theta_g = 0^\circ$ and 10° .

The behavior for mean Nu and mean TKE are shown in Figures 4C,D for G-ROM and LDF-ROM. Despite large errors in the mean flow prediction at $\theta_g = 0^\circ, 10^\circ$, the LDF-ROM is able to predict mean Nu with a maximum error around 5% whereas

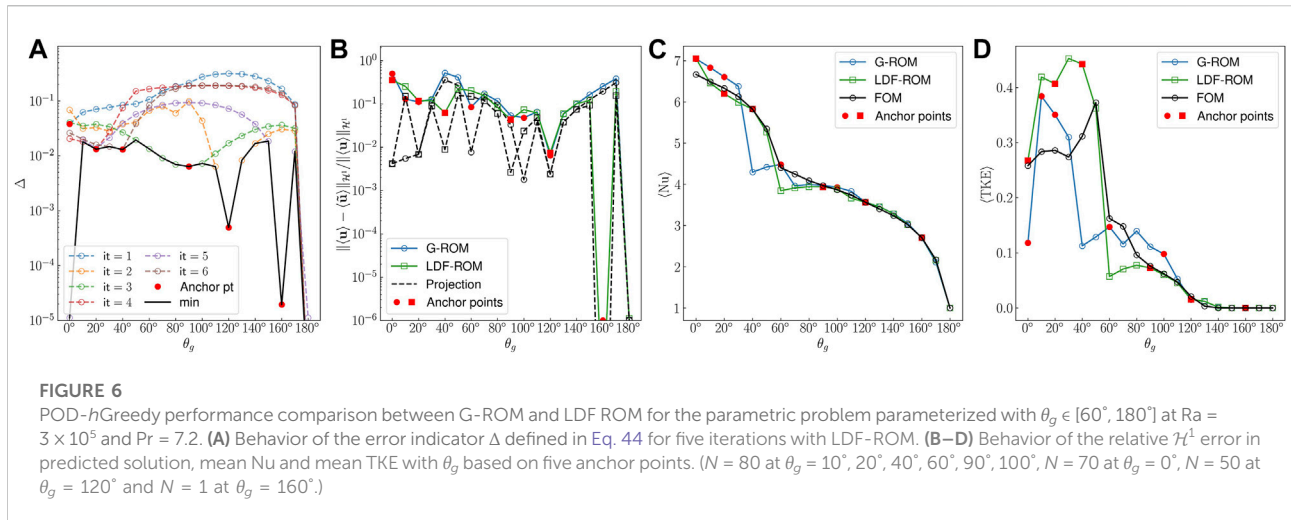


G-ROM has maximum error around 28%. In addition, LDF-ROM is able to more accurately predict mean TKE than G-ROM.

4.3.3 $Ra = 3 \times 10^5$

For $(Pr, Ra) = (7.2, 3 \times 10^5)$ the flow is quite chaotic (similar to what is found for $Pr = 0.71$ at lower Ra). **Figures 5E, F** show the mean (or steady) temperature and velocity solution at $\theta_g = 0^\circ, 10^\circ, \dots, 180^\circ$ (19 datapoints). This time, the bifurcation occurs at $\theta_g = 60^\circ$. We use this elevated Rayleigh-number case to explore the behavior of the h -Greedy pMOR convergence by considering application of the algorithm to two different training sets, $\mathcal{P}_1 = [60^\circ, 70^\circ, \dots, 180^\circ]$ and $\mathcal{P}_2 = [0^\circ, 10^\circ, \dots, 180^\circ]$. The set \mathcal{P}_1 excludes the spatio-temporal chaotic regime while \mathcal{P}_2 spans the full range of flow phenomena.

The anchor point selection process for \mathcal{P}_1 with LDF-ROM is demonstrated in **Figure 5A**, starting with $\theta_{g,1}^* = 60^\circ$ and proceeding for $L = 5$ iterations. Again, we observe that the error estimate at the anchor points, $\theta_{g,1}^* = 60^\circ$, $\theta_{g,3}^* = 80^\circ$ and $\theta_{g,5}^* = 120^\circ$ are larger than other anchor points because of unsteadiness. The behavior of the relative \mathcal{H}^1 error in predicted solution is shown in **Figure 5B**. For $\theta_g \in [130^\circ, 180^\circ]$, where the solution is steady, the ROM estimates at the anchor points are almost identical to the Galerkin projection. For $\theta_g \in [60^\circ, 120^\circ]$, where the solution is unsteady, the errors at the anchor points ($\theta_g = 60^\circ, 80^\circ, 110^\circ, 120^\circ$) are less than 10%. However, because of the irregular solution manifold, there is a 20% maximum error at $\theta_g = 170^\circ$, despite the ROM being based on the nearby $\theta_g = 160^\circ$ anchor point.



The behavior of the mean Nu and mean TKE are shown in Figures 5C,D. The maximum error in the predicted mean Nu is around 5% with LDF-ROM and 8% with G-ROM. The mean TKE estimation is also reasonable but is underestimated at $\theta_g = 70^\circ$.

Next we examine the same problem configuration but with the full parameter space \mathcal{P}_2 . The problem now includes spatio-temporal chaos for $\theta_g \in [0^\circ, 40^\circ]$ with the lower values being more chaotic. Figure 6 show the results of the application of Algorithm 1 for the construction of the LDF-ROM for the parametric problem with \mathcal{P}_2 . The algorithm starts with $\theta_{g,1}^* = 0^\circ$ and is performed with $L = 6$ iterations. The anchor point selection process is demonstrated in Figure 6A. We observe the same issue as in the previous cases, where unsteadiness leads to larger error estimates than with the steady regimes.

The relative \mathcal{H}^1 error in predicted solution is shown in Figure 6B. Again we find the estimation is almost identical to the Galerkin projection for $\theta_g \in [130^\circ, 180^\circ]$ where the solution is steady. For $\theta_g \in [0^\circ, 120^\circ]$ where the solution is unsteady, the errors at anchor points $\theta_g = 20^\circ, 40^\circ, 90^\circ, 120^\circ$ are less than 10% but 35% at $\theta_g = 0^\circ$, which corresponds to “simple” Rayleigh-Bénard convection. Note that 35% is the error after carefully chosen N and spatial radius δ in Leray filtering. The corresponding mean Nu and mean TKE behavior are shown in Figures 6C,D. The maximum error in the predicted mean Nu is around 12%. For mean TKE, the estimation for $\theta_g \in [60^\circ, 180^\circ]$ is acceptable, while it is overestimated for $\theta_g \in [0^\circ, 50^\circ]$.

We are also aware that in some applications, the $\text{Std}(\text{Nu})$ could be considered as QOI. However, comparing to the mean Nu and mean TKE, we find $\text{Std}(\text{Nu})$ is in general a more challenging QOI. Figure 7 shows the predicted $\text{Std}(\text{Nu})$ in the three Ra cases. We observe accurate prediction in $Ra = 1 \times 10^4$ case. However, unlike the mean TKE, the $\text{Std}(\text{Nu})$ soon becomes intractable with $Ra = 8 \times 10^4$ even with Leray regularization and is even worse in $Ra = 3 \times 10^5$.

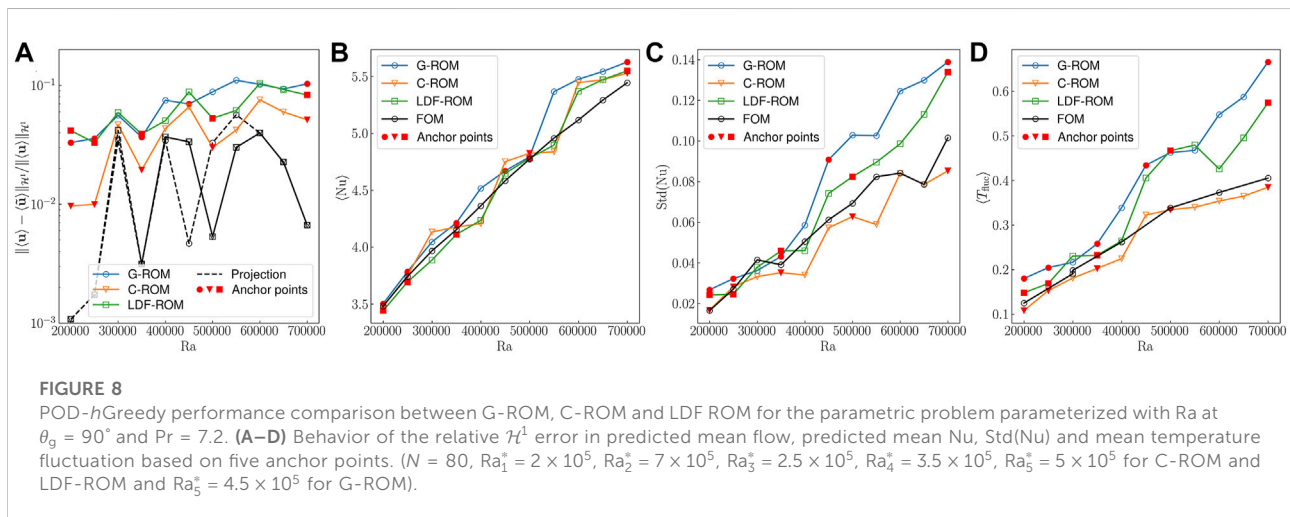
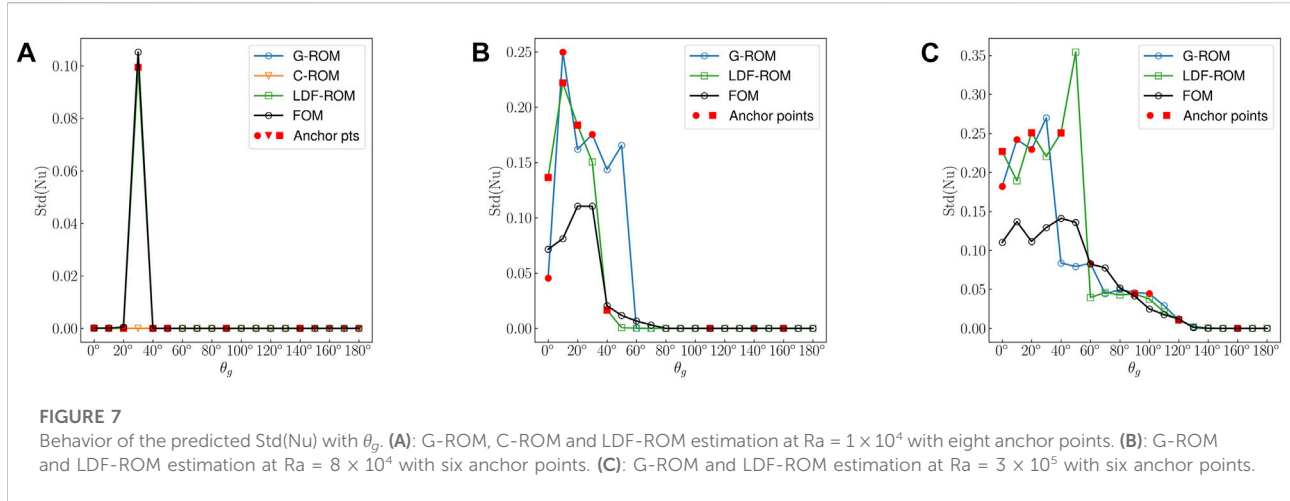
4.4 Parametric model order reduction results: Ra variation

In this section, we consider the slot problem at $\theta_g = 90^\circ$ and $Pr = 7.2$ with the parametric space defined by $Ra \in \mathcal{P} = [2 \times 10^5, 7 \times 10^5]$. Unlike the problem with θ_g variation, all solutions are unsteady and there is no parametric bifurcation. In order to assess performance, we generate FOM data, including mean flow, mean Nu, $\text{Std}(\text{Nu})$, mean TKE and mean temperature fluctuation, for $Ra = 2 \times 10^5, 2.5 \times 10^5, \dots, 7 \times 10^5$ ($n_{\text{train}} = 11$ datapoints). The quantities are averaged over 500 CTUs once the solution reaches the statistically steady state.

Figure 8 shows the results of the application of Algorithm 1 for the pMOR using G-ROM, C-ROM and LDF-ROM. The solid line denotes the performance of the reduced model which minimizes the error indicator, and thus is selected by the Greedy procedure (cf. Algorithm 1, $n_{\text{cand}} = 2$). Anchor points are denoted as red circle while FOM data is denoted as black solid line. The algorithm starts with $Ra_1^* = 2 \times 10^5$ and is performed with $L = 5$ iterations. The number of POD basis N with anchor points are listed in the figure caption.

Figure 8A shows the behavior of the relative \mathcal{H}^1 error in mean flow prediction with Ra. First, we observe the errors at the anchor points are less than 10% with C-ROM and LDF-ROM while G-ROM has 10% error at $Ra_2^* = 7 \times 10^5$. The maximum error is roughly 11% in G-ROM, 10% in LDF-ROM and 8% in C-ROM. Comparing with the Galerkin projection error (denoted by the black dashed line), the pMOR accuracy is seen to be quite satisfactory throughout \mathcal{P} .

Figure 8B shows the behavior of the predicted mean Nu with Ra. At the anchor points, we observe good agreement between ROMs and FOM and that stabilization does improve its accuracy. The maximum relative error is roughly 8% in G-ROM, 6.5% in C-ROM and 5% in LDF-ROM. Figures 8C,D show the behavior



of the predicted Std(Nu) and mean temperature fluctuation. In both QOIs, we find C-ROM outperforms the other two models. At $Ra = 7 \times 10^5$, LDF-ROM is only slightly better than G-ROM.

This parametric space is in general more tractable than those involving variation in θ_g . This outcome might be anticipated by observing the mean temperature fields shown in Figure 1C, which suggests that the solution manifold with respect to Ra is quite smooth. This is also reflected in the QOIs, for example, the mean Nu, Std(Nu) and mean temperature fluctuation behave almost linearly as Ra increases.

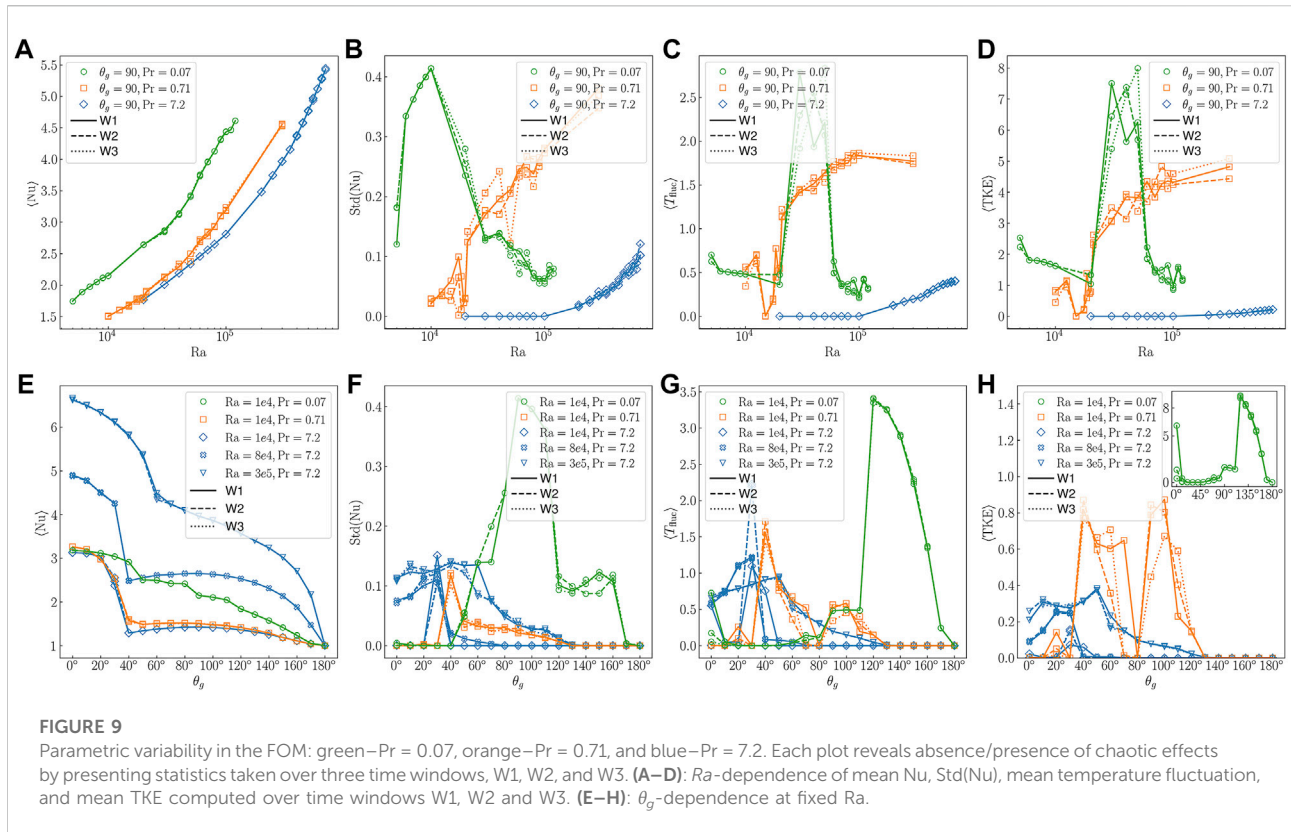
5 Discussion

In this section, we investigate some of the flow behaviors exhibited by the FOM to better understand how they influence the relative performance of the ROMs. We note that we cannot,

in general, expect a ROM to be able to predict FOM behavior if the flow itself is not predictable. Thus, variability in the FOM provides an anticipated lower bound on ROM performance for the reproduction problem.

5.1 Spatio-temporal chaos

As pointed out in the introduction, we classify a flow to be spatio-temporal chaotic by examining its consistency in mean flow over various time windows. (We use this simple metric here for convenience—we have also examined the flow fields and the time traces of multiple QOIs.) Here, we explore how lack of consistency influences four QOIs, mean Nu, Std(Nu), mean temperature fluctuation, and mean TKE, at three successive time windows, W1, W2 and W3. These quantities are used to indicate the variability in the FOM. As with the



preceding cases, we consider averaging times of 500 CTUs for each of the three windows. The starting time for W1 differs with given parameters as some cases take a longer time to reach a statistically steady state. For example, for Pr = 0.71 at Ra = 1.8×10^4 and $\theta_g = 90^\circ$, the flow is chaotic until 6,000 CTUs and then becomes periodic.

Figures 9A–D show the behavior of the four QOIs with Ra at three Pr for $\theta_g = 90^\circ$. Pr = 0.07 is denoted as green line, Pr = 0.71 is denoted as orange line, while Pr = 7.2 is denoted as blue line. Window W1 is denoted by a solid line, W2 by a dashed line, and W3 by a dotted line.

From Figures 9A–D we can see the following:

- 1) For Pr = 0.07 the QOIs are fairly consistent except for $Ra \in [3 \times 10^4, 5 \times 10^4]$.
- 2) For Pr = 0.71, we find large variability in Std(Nu), mean temperature fluctuation and mean TKE for $Ra > 2 \times 10^4$.
- 3) Pr = 7.2 exhibits the least variability.

For Ra where we find that the QOI variability is high, we have also examined the mean flow at multiple time windows and found that those are also inconsistent. In all cases, the mean Nu is quite repeatable.

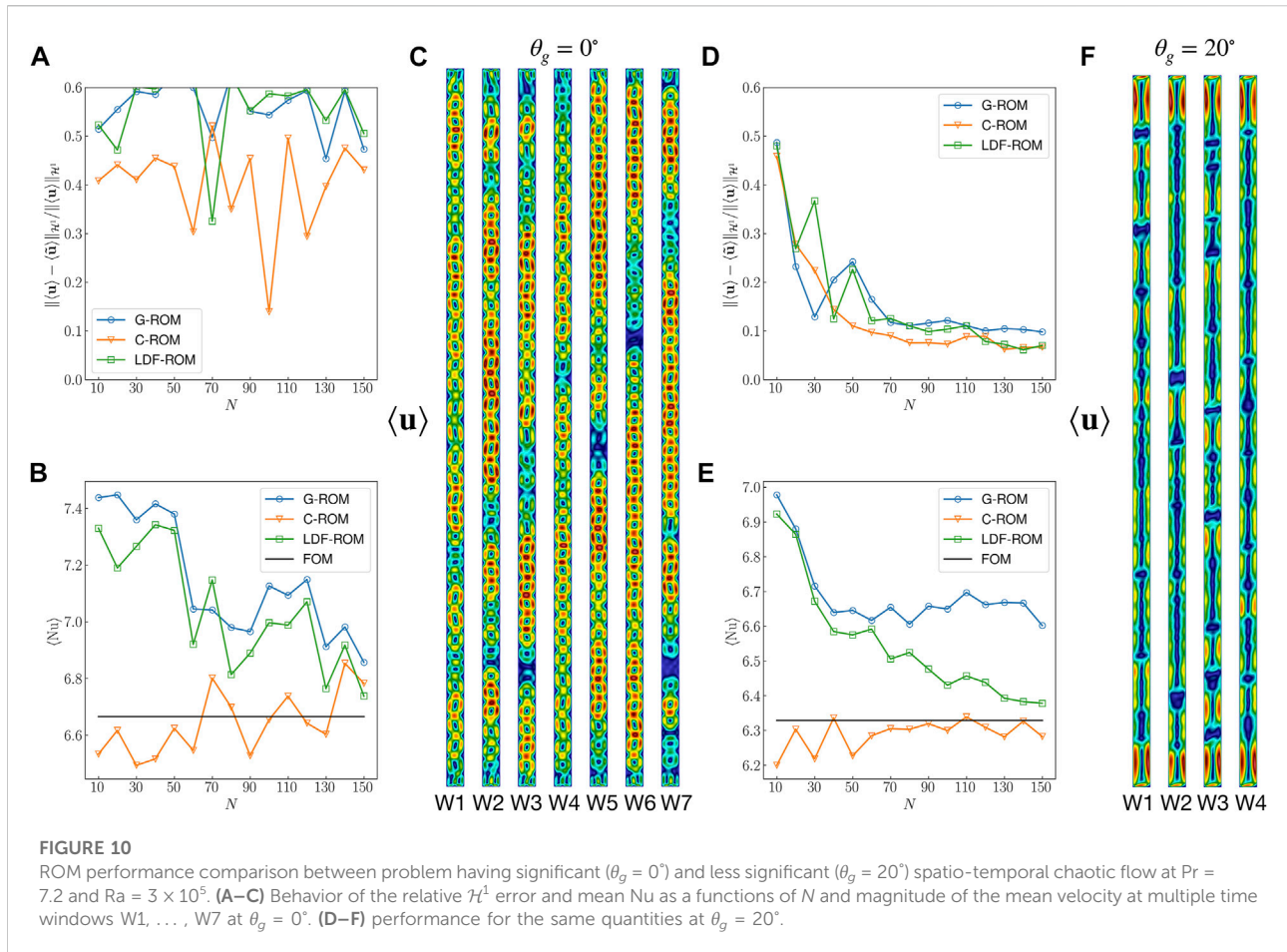
Figures 9E–H show the behavior of the same QOIs as a function of θ_g . For Pr = 7.2, we consider three different values of

Ra. We find for most of the θ_g , the variability is small except for small θ_g where we also report spatio-temporal chaotic flow. For Pr = 0.07, we find the QOIs has small variability with $Ra = 1 \times 10^4$. However, for Pr = 0.71, we find large variability, especially in the mean TKE. Not only it has spatio-temporal chaotic flow (for example $\theta_g = 100^\circ$), but also the solution manifold is not smooth. By varying $\theta_g = 80^\circ$ to $\theta_g = 100^\circ$, the solution changes from steady to periodic then spatio-temporal chaotic.

From Figures 9E–H we observe the following:

- 1) For Pr = 0.07, Std(Nu) exhibits up to 50% variability (e.g., at $\theta_g = 70^\circ$) while $\langle T_{\text{fluc}} \rangle$ and $\langle \text{TKE} \rangle$ have orders-of-magnitude relative variability at $\theta_g = 0^\circ$.
- 2) For Pr = 0.71, $\langle T_{\text{fluc}} \rangle$ and $\langle \text{TKE} \rangle$ exhibit significant variability for $\theta_g \in [60^\circ, 110^\circ]$.
- 3) For Pr = 7.2, the most notable variation is at $\theta_g = 30^\circ$ for Std(Nu), $\langle T_{\text{fluc}} \rangle$, and $\langle \text{TKE} \rangle$ at $Ra = 10^4$. Remarkably, the higher Rayleigh number cases do not exhibit as much variance.

As in Figures 9A–D, the mean Nu is seen to be a repeatable QOI. It is worth noting the real challenge and sensitivity of this class of problems is illustrated in Figure 9H. Here, we observe for the (Pr, Ra) = (0.71, 10^4) case that the flow alternates from steady

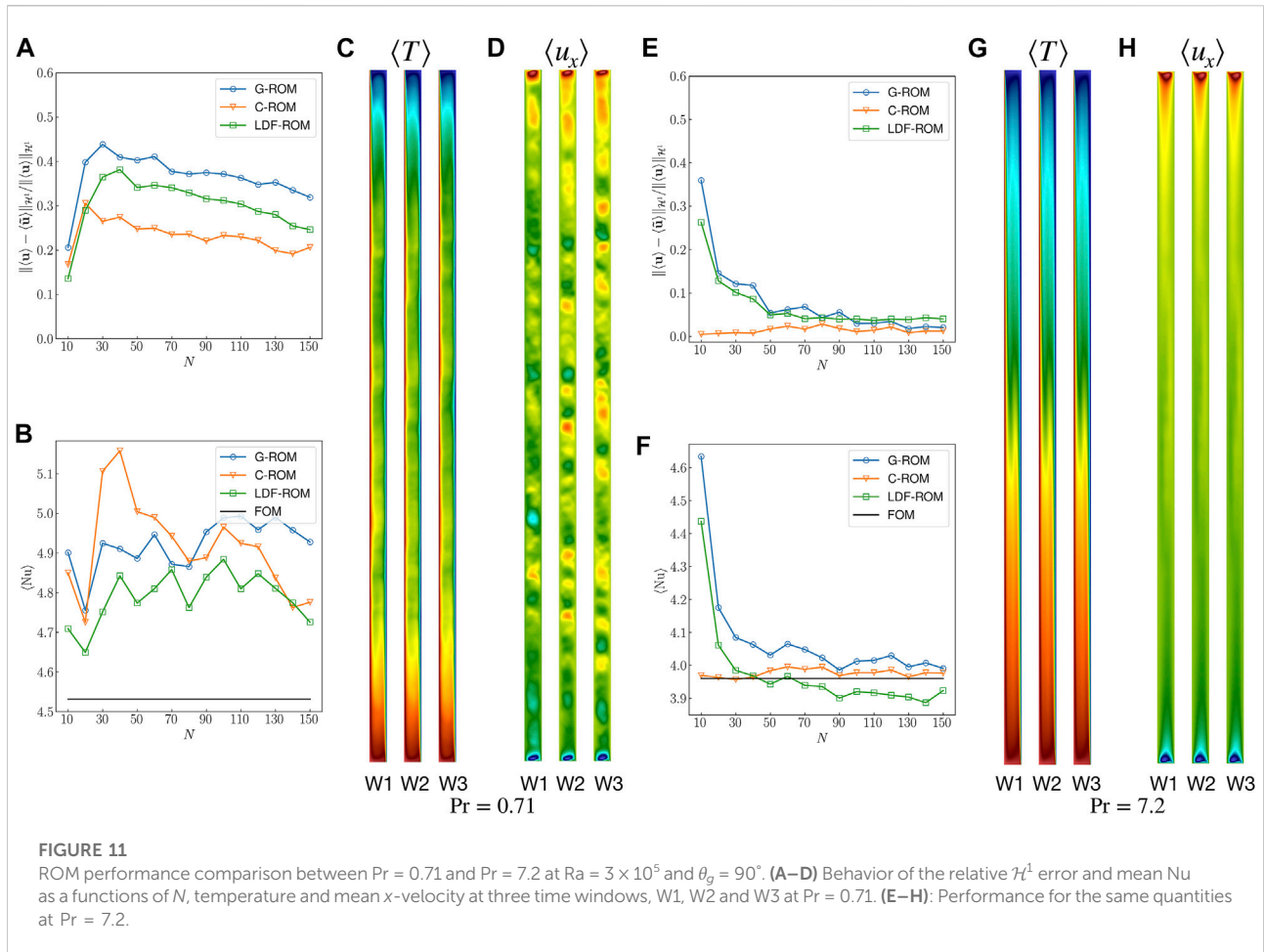


to unsteady at multiple points along the one-dimensional θ_g parameter space as indicated by several distinct zeros in the TKE. In Figure 10, we further explore the influence of spatio-temporal chaos by examining the mean-flow distributions and ROM performance for $(Pr, Ra) = (7.2, 3.5 \times 10^4)$ at $\theta_g = 0^\circ$ and $\theta_g = 20^\circ$. Figures 10A–C show the behavior of the \mathcal{H}^1 error and mean Nu predictions as a function of N , along with the mean-velocity magnitude distributions over seven time windows for $\theta_g = 0^\circ$. Figures 10D–F show the same quantities but with $\theta_g = 20^\circ$ and only three time windows. In Figure 10C, we observe that for $\theta_g = 0^\circ$ the number of rolls in the mean velocity field changes across different time windows. Similar changes are observed, to a lesser extent, at $\theta_g = 20^\circ$. Hence, both solutions are categorized as spatio-temporal chaotic flow, but the $\theta_g = 0^\circ$ case is more significant. Comparing the mean flow error and mean Nu, we observe that the ROM convergence for the reproduction problem is slower (or nonexistent) at $\theta_g = 0^\circ$, while the convergence behavior is more favorable at $\theta_g = 20^\circ$.

We have also computed the relative error between FOM mean flows across seven time windows for the two values of θ_g . The maximum relative \mathcal{H}^1 error is 34% for $\theta_g = 0^\circ$ and 10% for

$\theta_g = 20^\circ$. These FOM discrepancies can be considered as a bound on the predictive capabilities of the ROM. Indeed, the values of 34 and 10% are consistent with the lower bounds realized in Figures 10A,D.

In Figure 11, we examine the influence of Prandtl number by comparing results for $Pr = 0.71$ and 7.2 at $(Ra, \theta_g) = (3 \times 10^5, 90^\circ)$. Figures 11A–D show the convergence behavior for the \mathcal{H}^1 error and mean Nu as well as mean temperature and x -velocity fields at three time windows for $Pr = 0.71$, while Figures 11E–H show the same quantities for $Pr = 7.2$. From the mean fields, we observe that the number of rolls and its position changes with time window at $Pr = 0.71$, while minimal variance is observed at $Pr = 7.2$. Hence the solution at $Pr = 0.71$ is considered to be spatio-temporal chaotic while only chaotic at $Pr = 7.2$. Comparing the behavior of the relative \mathcal{H}^1 error in the mean flow and mean Nu, we observe convergence issues in the ROM at $Pr = 0.71$, while the same metrics converge at $Pr = 7.2$. We further compute the relative variance between two FOM mean flows across seven time windows for the two considered θ_g . The maximum relative \mathcal{H}^1 error is 14% for



$Pr = 0.71$ and only 1% for $Pr = 7.2$. This again explains why the approximation errors are different between the two cases. Considering these variance levels as a lower bound for the ROM, we can view 20% error in C-ROM as acceptable. On the other hand, the approximation error for $Pr = 7.2$ is able to reach below 10%.

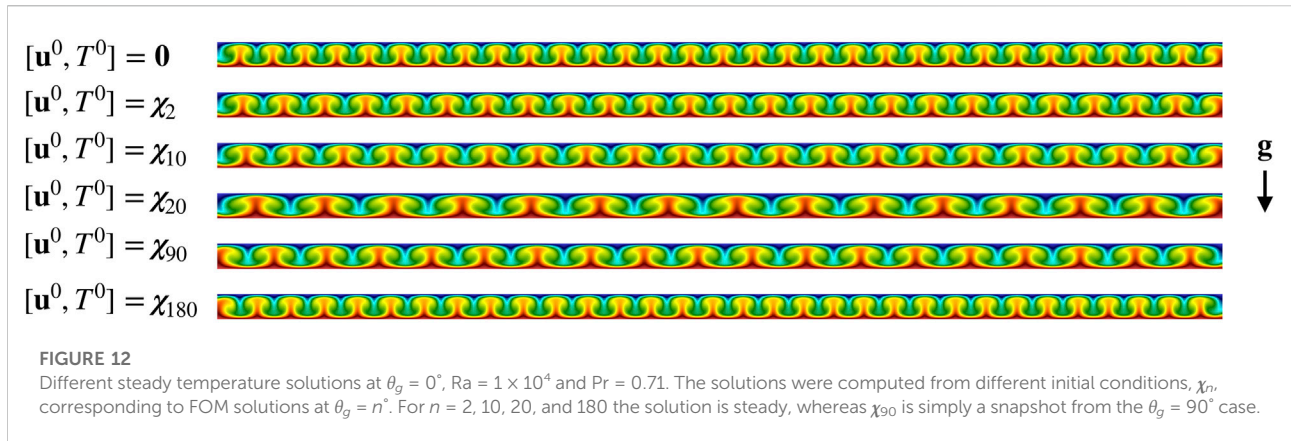
In summary, the results of this section show that convergence issues and variations in the QOIs in the ROM can have a high correlation with the flow being spatio-temporal chaotic. From Table 1 and Figure 9, we see that $Pr = 0.71$ has a more complicated solution manifold found with the other two Prandtl numbers and also exhibits spatio-temporal chaos at a relatively small Rayleigh number, $Ra = 10^4$.

5.2 Multiple steady-state solutions

In this section, we report the existence of multiple steady-state solutions observed for the case $(\theta_g, Ra) = (0^\circ, 10^4)$. The variations are characterized by different numbers of

recirculation rolls, which are induced by using different initial conditions. Figure 12 shows steady temperature solutions generated by starting with steady solutions, χ_n from other values of $\theta_g = n^\circ$, save for the χ_{90} case, which corresponds to a single snapshot of the unsteady flow/temperature field at $\theta_g = 90^\circ$. Multiple steady states are also observed for this Prandtl number at $\theta_g = 10^\circ$ and 20° and have been reported by other authors as well [12, 14, 40].

For solution reproduction, the multiplicity of the solutions is not an issue as long as the ROM uses the same initial condition as the FOM. However, for parametric problem, these multiple states could easily lead to an incorrect (or at least, unexpected or unverifiable) conclusion. For example, if the ROM anchor at $\theta_g = 10^\circ$ is used to approximate the solution at $\theta_g = 0^\circ$. With the initial condition at $\theta_g = 10^\circ$, the approximate solution will be the third temperature solution shown in Figure 12. However, if one collects the FOM data at θ_g with zero initial condition, one will consider the first temperature solution as the truth solution. As we could consider those roll solutions as sine and



cosine functions, the first and third solution are nearly orthogonal; their difference is $\mathcal{O}(1)$ and one would thus conclude that the pMOR had failed when it in fact had generated a valid solution.

5.3 Discussion summary

We have noted in Table 1 the broad range of flow regimes encountered for the tilted slot problem and in this section have illuminated a correlation between the flow states and predictive power of the MOR/pMOR framework. The cases with spatio-temporal chaos are generally the most challenging for model-order reduction and the pMOR errors are found to be (approximately) bounded from below by the variance observed in successive FOM simulations performed at the same parametric point. The development of the pMOR thus needs to be performed with care.

Two parameterizations were considered: 1) θ_g -variation, where a bifurcation exists and solution space is a blend of unsteady and steady solutions, and 2) Ra -variation, where no bifurcation exists and one finds only unsteady solutions. In the θ_g -variation problem, accurate prediction in mean flow and other QOIs by the pMOR was demonstrated in the $Ra = 1 \times 10^4$ and $Ra = 8 \times 10^4$ cases. In high Ra cases, acceptable prediction of Nu is achieved with LDF-ROM but a small mean-flow error is not realizable because of spatio-temporal chaos.

The results also indicate that the LDF-ROM is a better candidate for parametric problems with bifurcation than C-ROM. This observation is new, yet consistent with the results of [9], where the authors show that C-ROM is effective for parametric problems that do not have a bifurcation. For the parametric problem parameterized with Ra , without spatio-temporal chaos, we find that pMOR with any of the

three methods, G-ROM, C-ROM, or LDF-ROM, is able to predict the mean flow quite well. In this case, C-ROM is the most accurate in mean flow prediction and other targeted QOIs. This result is not surprising given that the solution manifold does not have a bifurcation. Lastly, we remark that $\text{Std}(Nu)$ is generally the most challenging QOI of those explored here.

From the results, we are able to make an important observation. For parametric problems where pMOR is successful (e.g., errors $< 10\%$), the solution is either only chaotic (e.g., Ra variation with $\theta_g = 90^\circ$) or the solution does not have significant spatio-temporal chaos (e.g., θ_g variation with $Ra = 1 \times 10^4$, 8×10^4). Once the spatio-temporal chaos becomes significant, the predictive power of pMOR deteriorates and the maximum errors are dominated by variance in the truth solution.

Although not shown here, we have also applied POD- p Greedy to this problem. In the parametric problem parameterized with θ_g , it works only in the steady case $Ra = 1 \times 10^3$. Once the unsteady solution emerges, for example at $Ra = 1 \times 10^4$, combining modes associated with different values of θ_g leads to an unstable ROM even with the Leray regularization. Although no rigorous proof is given, we hypothesize that the issue is due to the bifurcation in solution behavior. This point was also suggested in [9], which empirically showed that combining modes associated with qualitatively different behaviors might lead to poor prediction. By contrast, when the current problem is parameterized with Ra we find that POD- p Greedy is more efficient than the h -refinement approach.

6 Conclusion

In this paper an error-indicated pMOR is applied to a 2D unsteady natural convection in a tilted high-aspect ratio slot.

We first considered the solution reproduction problem (non-predictive case) to demonstrate the convergence of the ROMs and the effectiveness of the stabilization methods. We next addressed the parametric problem (predictive case) to validate the error indicator and, more broadly, the stabilized POD- h Greedy procedures. Principal contributions include, 1) extension of the error indicator proposed in [9] to buoyancy-driven flows; 2) demonstration that Leray-regularized Galerkin ROMs provide a robust solution approach for this class of flows; 3) identification of spatio-temporal chaos as a source of irreproducibility in both the FOM and the ROM and that the variance in the FOM provides a lower bound on the pMOR error in these cases; 4) observation that accurate prediction ($<10\%$) with pMOR is achievable if the solution in the parametric space is either only chaotic or the spatio-temporal chaos is not significant, regardless of whether a bifurcation exists or not. Once the spatio-temporal chaos becomes significant, the performance of the pMOR deteriorates and the maximum errors of the mean flow and QOIs are dominated by the flow chaos.

We also highlight a number of challenges that are particularly relevant for buoyancy-driven flows and which should be taken into consideration in the design of pMOR strategies for 3D buoyancy driven turbulent flow. First, one needs to be aware of potential convergence issues for the mean flow and other QOI predictions when the FOM exhibits large-scale spatio-temporal chaos. Second, it is difficult to combine modes associated with different flow regimes, especially for the p Greedy case. Third, even relatively simple (e.g., steady) flows can exhibit multiple states at a given parameter. And fourth, there are large offline costs both in terms of computational time and required storage for error indicator and $O(N^4)$ costs for online-only error indicators⁵.

We outline potential next steps in pMOR development for this class of problems.

- 1) *Extension to higher dimensional parameter space.* In this work, we considered only one-dimensional parameter space since the pMOR behavior needed to be carefully diagnosed; however higher dimensional parameter spaces are more interesting for engineering applications.
- 2) *hp-Greedy with a bifurcation detection technique.* Although we find LDF-ROM is more efficient than C-ROM for parametric problems that have a bifurcation, the h -refinement strategy considered in this paper might require an infeasible number of offline simulations as the

dimension of the parameter space increases. To tackle complex parametrizations, more advanced sampling strategies that combine h - and p -refinement [37], potentially with bifurcation detection, could be beneficial [42].

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding authors.

Author contributions

P-HT derived the formulation and error indicator, performed the numerical simulations, collected the data, analyzed and interpreted the results. PF supervised the project, provided critical feedback and helped on interpreted the results. Both authors prepared the manuscript.

Funding

This research is supported by the DOE Office of Nuclear Energy under the Nuclear Energy University Program (Proj. No. DE_NE0008780). Simulations were performed at the DOE Office of Science User Facility ALCF (Argonne Leadership Computing Facility).

Acknowledgments

We thank Anthony T. Patera (MIT) for his valuable comments, many contributions, insights and guidelines.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

⁵ The $O(N^4)$ costs, which arise from the rank-3 advection tensor, might be mitigated by an $O(N^2)$ approximation to the advection operator, such as suggested in [41].

References

1. Grepl MA, Patera AT. A posteriori error bounds for reduced-basis approximations of parametrized parabolic partial differential equations. *ESAIM: Math Model Numer Anal* (2005) 39:157–81. doi:10.1051/m2an:2005006
2. Rozza G, Huynh DBP, Patera AT. Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations. *Arch Comput Methods Eng* (2008) 15:229–75. doi:10.1007/s11831-008-9019-9
3. Merzari E, Ninokata H, Mahmood A, Rohde M. Proper orthogonal decomposition of the flow in geometries containing a narrow gap. *Theor Comput Fluid Dyn* (2009) 23:333–51. doi:10.1007/s00162-009-0152-3
4. Merzari E, Pointer WD, Fischer P. A POD-based solver for the advection-diffusion equation. *Fluids Eng Division Summer Meet* (2011) 44403:1139–47.
5. Quarteroni A, Rozza G, Manzoni A. Certified reduced basis approximation for parametrized partial differential equations and applications. *J Math Ind* (2011) 1:3. doi:10.1186/2190-5983-1-3
6. Wang Z, Akhtar I, Borggaard J, Iliescu T. Proper orthogonal decomposition closure models for turbulent flows: A numerical comparison. *Computer Methods Appl Mech Eng* (2012) 237:10–26. doi:10.1016/j.cma.2012.04.015
7. Kaneko K, Fischer P. Augmented reduced order models for turbulence. *Front Phys* (Forthcoming 2022).
8. Quarteroni A, Manzoni A, Negri F. *Reduced basis methods for partial differential equations: An introduction*. Switzerland: Springer (2015).
9. Fick L, Maday Y, Patera AT, Taddei T. A stabilized POD model for turbulent flows over a range of Reynolds numbers: Optimal parameter sampling and constrained projection. *J Comput Phys* (2018) 371:214–43. doi:10.1016/j.jcp.2018.05.027
10. Ngoc Cuong N, Veroy K, Patera AT. Certified real-time solution of parametrized partial differential equations. In: *Handbook of materials modeling*. Dordrecht, Netherlands: Springer (2005). p. 1529–64.
11. Veroy K, Patera AT. Certified real-time solution of the parametrized steady incompressible Navier–Stokes equations: Rigorous reduced-basis a posteriori error bounds. *Int J Numer Methods Fluids* (2005) 47:773–88. doi:10.1002/flid.867
12. Wang Q, Wan ZH, Yan R, Sun DJ. Multiple states and heat transfer in two-dimensional tilted convection with large aspect ratios. *Phys Rev Fluids* (2018) 3:113503. doi:10.1103/physrevfluids.3.113503
13. Deparis S. Reduced basis error bound computation of parameter-dependent Navier–Stokes equations by the natural norm approach. *SIAM J Numer Anal* (2008) 46:2039–67. doi:10.1137/060674181
14. Deparis S, Rozza G. Reduced basis method for multi-parameter-dependent steady Navier–Stokes equations: Applications to natural convection in a cavity. *J Comput Phys* (2009) 228:4359–78. doi:10.1016/j.jcp.2009.03.008
15. Ballarin F, Rebollo TC, Ávila ED, Mármol MG, Rozza G. Certified Reduced Basis VMS-Smagorinsky model for natural convection flow in a cavity with variable height. *Comput Mathematics Appl* (2020) 80:973–89. doi:10.1016/j.camwa.2020.05.013
16. Knezevic DJ, Nguyen NC, Patera AT. Reduced basis approximation and a posteriori error estimation for the parametrized unsteady Boussinesq equations. *Math Models Methods Appl Sci* (2011) 21:1415–42. doi:10.1142/s0218202511005441
17. Yano M. A space-time Petrov–Galerkin certified reduced basis method: Application to the Boussinesq equations. *SIAM J Sci Comput* (2014) 36:A232–66. doi:10.1137/120903300
18. Kaneko K, Tsai PH, Fischer P. Towards model order reduction for fluid-thermal analysis. *Nucl Eng Des* (2020) 370:110866. doi:10.1016/j.nucengdes.2020.110866
19. Wells D, Wang Z, Xie X, Iliescu T. An evolve-then-filter regularized reduced order model for convection-dominated flows. *Int J Numer Methods Fluids* (2017) 84:598–615. doi:10.1002/flid.4363
20. Karimi A, Paul MR. Quantifying spatiotemporal chaos in Rayleigh–Bénard convection. *Phys Rev E* (2012) 85:046201. doi:10.1103/physreve.85.046201
21. Wolf A, Swift JB, Swinney HL, Vastano JA. Determining Lyapunov exponents from a time series. *Physica D: Nonlinear Phenomena* (1985) 16:285–317. doi:10.1016/0167-2789(85)90011-9
22. Goldhirsch I, Sulem PL, Orszag SA. Stability and Lyapunov stability of dynamical systems: A differential approach and a numerical method. *Physica D: Nonlinear Phenomena* (1987) 27:311–37. doi:10.1016/0167-2789(87)90034-0
23. Cross M, Hohenberg P. Spatiotemporal chaos. *Science* (1994) 263:1569–70. doi:10.1126/science.263.5153.1569
24. Egolf DA, Melnikov IV, Pesch W, Ecke RE. Mechanisms of extensive spatiotemporal chaos in Rayleigh–Bénard convection. *Nature* (2000) 404:733–6. doi:10.1038/35008013
25. Cross MC, Hohenberg PC. Pattern formation outside of equilibrium. *Rev Mod Phys* (1993) 65:851–1112. doi:10.1103/revmodphys.65.851
26. Busse F. Non-linear properties of thermal convection. *Rep Prog Phys* (1978) 41:1929–67. doi:10.1088/0034-4885/41/12/003
27. Maday Y, Patera AT, Ronquist EM. A well-posed optimal spectral element approximation for the Stokes problem. *Tech Rep* (1987) 1987.
28. Fischer PF, Lottes JW, Kerkemeier SG. *Nek5000 web page* (2008).
29. Quarteroni A, Valli A. Numerical approximation of partial differential equations. In: *Springer series in computational mathematics*. Berlin: Springer (1994).
30. Fischer P, Schmitt M, Tomboulides A. Recent developments in spectral element simulations of moving-domain problems. In: *Recent progress and modern challenges in applied mathematics, modeling and computational science*. Berlin: Springer (2017). p. 213–44.
31. Rønquist EM, Patera AT. A Legendre spectral element method for the Stefan problem. *Int J Numer Methods Eng* (1987) 24:2273–99. doi:10.1002/nme.1620241204
32. Maday Y, Patera AT, Rønquist EM. *The PN × PN–2 method for the approximation of the Stokes problem*. Paris: Laboratoire d'Analyse Numérique (1992).
33. Guermont JL, Oden JT, Prudhomme S. Mathematical perspectives on large eddy simulation models for turbulent flows. *J Math Fluid Mech* (2004) 6:194–248. doi:10.1007/s00021-003-0091-5
34. Guermont JL, Pasquetti R, Popov B. Entropy viscosity method for nonlinear conservation laws. *J Comput Phys* (2011) 230:4248–67. doi:10.1016/j.jcp.2010.11.043
35. Mullen J. *Development of a parallel spectral element based large eddy simulation model for the flow of incompressible fluids in complex geometries*. Ph.D. thesis. Providence, Rhode Island: Brown University (1999).
36. Sabetghadam F, Jafarpour A. α regularization of the POD–Galerkin dynamical systems of the Kuramoto–Sivashinsky equation. *Appl Mathematics Comput* (2012) 218:6012–26. doi:10.1016/j.amc.2011.11.083
37. Eftang JL, Knezevic DJ, Patera AT. An hp certified reduced basis method for parametrized parabolic partial differential equations. *Math Computer Model Dynamical Syst* (2011) 17:395–422. doi:10.1080/13873954.2011.547670
38. Haasdonk B, Ohlberger M. Reduced basis method for finite volume approximations of parametrized linear evolution equations. *ESAIM: Math Model Numer Anal* (2008) 42:277–302. doi:10.1051/m2an:2008001
39. Haasdonk B. Convergence rates of the POD–Greedy method. *ESAIM: Math Model Numer Anal* (2013) 47:859–73. doi:10.1051/m2an/2012045
40. Gelfgat AY, Bar-Yoseph P, Yarin A. Stability of multiple steady states of convection in laterally heated cavities. *J Fluid Mech* (1999) 388:315–34. doi:10.1017/s0022112099004796
41. Barrault M, Maday Y, Nguyen NC, Patera AT. An empirical interpolation method: Application to efficient reduced-basis discretization of partial differential equations. *Comptes Rendus Mathématique* (2004) 339:667–72. doi:10.1016/j.crma.2004.08.006
42. Pichi F. *Reduced order models for parametric bifurcation problems in nonlinear PDEs*. Ph.D. thesis. Trieste, Italy: Scuola Internazionale Superiore di Studi Avanzati (2020).