Oral presentation | Reduced order models

Reduced order models-II

Mon. Jul 15, 2024 2:00 PM - 4:00 PM Room C

[2-C-03] Parametric Reduced Order Model built from RBF-FD meshless simulations of flow and temperature fields in a 3D pipe with wavy surfaces

Fausto Dicech[1], Riccardo Zamolo[1], *Lucia Parussini[1] (1. University of Trieste)

Keywords: Reduced Order Model, meshless, parametric analysis

Twelfth International Conference on
Computational Fluid Dynamics (ICCFD12),
Kobe, Japan, July 14-19, 2024

Parametric Reduced Order Model built from RBF-FD meshless simulations of flow and temperature fields in a 3D pipe with wavy surfaces

F. Dicech*,**, R. Zamolo* and L. Parussini*
Corresponding author: lparussini@units.it
* Department of Engineering and Architecture, University of Trieste, Via Valerio 6/1, Trieste, 34127, IT.
** Development Centre, Optimad S.r.l., Località Padriciano 99, Trieste, 34149, IT.

Abstract: The building of a POD-based ROM (Proper Orthogonal Decomposition based Reduced Order Model) representation from RBF-FD (Radial Basis Function-generated Finite Differences) meshless simulations is proposed, thus introducing a novel and original approach to the analysis of parametric problems. The RBF-FD meshless method is most suitable when solving thermofluid dynamic problems on parametric domains, because it can handle complex geometries and large deformations without the need for mesh, or grid, or tessellation generation and refinement. A simple distribution of nodes over the domain is only needed, and the convergence rate of RBF-FD methods can be easily increased. Nevertheless, a reliable exploration of the parameter space still requires many simulations in order to capture the behaviour of the analysed system. So, reduced order modelling methods can be used to significantly speed up the analysis, aiming to accurately describe the physical process with a relatively small number of DOFs (Degrees of Freedom). In particular, we want to compare the capability of a low-fidelity approximation of the problem and a ROM built from a few high-fidelity simulations to represent the parametric solution. The intent is to understand how to exploit the two models to achieve the best multi-fidelity ROM representation of the parametric problem. The approach is applied to the parametric analysis of flow and temperature fields in a 3D pipe with wavy surfaces, considering both geometric and physical parameters.

Keywords: Meshless Methods, Multi-fidelity, Reduced Order Model.

1 Introduction

For various engineering applications, such as fluid dynamics and heat transfer parametric simulations, structural mechanics analysis, design optimization problems and uncertainty quantification, computational efficiency and accuracy are both critical. Such applications can benefit from the combination of POD-based ROM and RBF-FD meshless methods, which can be a powerful tool for performing efficient parametric analysis of complex systems.

ROMs [1] aim to capture the essential dynamics of a system with a smaller set of degrees of freedom compared to the full-order model and this can significantly reduce computational time, especially for problems with high dimensionality or complex geometries. In contrast to intrusive ROM methods, that modify the original governing equations of the system, non-intrusive POD-based ROM [2] operates entirely on the data generated by the FOM (Full-Order Model) [3, 4] and it is used here for extracting the FOM key features.

The data needed to generate the ROM model are obtained by RBF-FD (Radial Basis Function-generated Finite Differences) meshless method [5, 6], thus proposing a very original combination of the POD-based ROM and RBF-FD meshless methods. The RBF-FD meshless method is a numerical approach for solving partial differential equations (PDEs) which uses nodes to represent the problem domain. Basis functions centered at each node are then constructed using radial basis functions. These functions depend only on the distance between a point and the node, making the method highly adaptable to complex geometries and eliminating the need for mesh generation, which can be time-consuming especially for complex geometries. This means that combining the RBF-FD meshless method with POD-based ROM allows for faster exploration of different design parameters without regenerating the mesh for each case.

Mainly with the purpose of illustrating its potential, the combined approach is here applied to the parametric analysis of flow and temperature fields in a 3D annular pipe with wavy surfaces, considering both geometric and physical parameters. These annular wavy pipes are of interest for enhancing heat exchange in various applications and they are used in the domains of compact heat exchangers, nuclear reactors, microfluidic devices, etc. For this specific application, in particular, it was investigated how the error of the POD-based ROM is affected by the level of fidelity of the RBF-FD meshless solutions and the number of solutions used to train the ROM. Indeed, the convergence rate of RBF-FD meshless methods can be easily increased by increasing the number of nodes [7], but at the expense of higher computational cost. Therefore it becomes interesting to understand how to best exploit low-fidelity (LF) and high-fidelity (HF) solutions, eventually in a Multi-Fidelity ROM (MF-ROM) representation [8, 9] of the parametric problem.

MF-ROMs combine the strengths of multi-fidelity modeling and reduced-order modeling to overcome limitations faced by each technique individually. MF-ROMs leverage both HF and LF models. HF data captures complex behavior, while LF data provides a computationally cheap way to explore a wider parameter space. In general, by combining them, MF-ROMs can achieve higher accuracy compared to using just ROMs. In fact, MF-ROMs utilize inexpensive LF simulations for most calculations, reserving expensive HF simulations for critical points where high accuracy is essential, and this significantly reduces the overall computational burden. However, developing MF-ROMs involves building both HF and LF models and then creating a framework to combine their outputs. This can be more complex than using a single-fidelity ROM. Therefore, it is relevant to understand whether there is a real advantage in using MF-ROMs.

Let us point out the limited availability of established MF-ROM methodologies. MF-ROMs are a relatively new research area compared to single-fidelity ROMs, and in the literature various approaches can be found. In [10] the main idea is to use POD to interpolate the difference between the field solution obtained on two different grids, a coarse one and a fine one, and an approximation of the fine grid solution is obtained summing a coarse grid solution and the POD approximation. Similarly, [11] proposed Gappy POD, and in [12] a multi-fidelity non intrusive POD methodology is presented, appending the POD basis with low-fidelity modes projected on the complementary space spanned by the high-fidelity data available. In [8, 9] kriging meta-models to approximate the POD coefficients are extended to a multi-fidelity context. [13] introduced a non-intrusive reduced basis method for parametrized nonlinear partial differential equations, where multi-fidelity Gaussian process regression is employed to approximate the combination coefficients of the reduced basis. [14] used an artificial neural network to map the relationship between the POD coefficients from different fidelity solutions, and [15] showed a comparison of utilizing kriging and artificial neural network to map this relationship. In [16] a manifold alignment method is presented to fuse inconsistent fields from HF and LF simulations by individually projecting their solution onto a common shared latent space, using cokriging model for the prediction of high-fidelity mode coefficients. [17] proposed a multi-fidelity multi-step Bayesian neural network, [18] multi-fidelity convolutional auto-encoders, [19] multi-fidelity deep neural networks for Bayesian model updating, [20] transfer learning for field reconstruction based on multi-fidelity solutions. Recently, [21] has proposed a multi-fidelity surrogate based on Galerkin-POD model, interpreting closure problem, to compensate for the contribution of the truncated scales onto the resolved ones, as a multi-fidelity problem and using a multi-fidelity deep operator network (DeepONet) framework to address it.

While various approaches are being developed, established and well-tested methodologies may not yet be widely available for all types of problems.

In the context of this work, POD-based ROM is adopted using NARGP (Nonlinear AutoRegressive multi-fidelity Gaussian Process regression) model for the prediction of mode coefficients. In particular, differences with a single-fidelity GP-based (Gaussian Process-based) ROM will be presented.

The paper is organized as follows: in Section 2 a description of the problem is given, the methodology is presented in Section 3 and its implementation in Section 4. In Section 5 the results are shown and discussed. Finally conclusions are given in Section 6.

2 Problem Statement

The approach proposed here is aimed at solving the general parametric BVP (Boundary Value Problem):

$$\begin{cases} \mathcal{L}\left(u, \boldsymbol{x}, \boldsymbol{y}\right) = f\left(\boldsymbol{x}, \boldsymbol{y}\right) & \text{in } \Omega\left(\boldsymbol{y}\right) \\ \mathcal{B}\left(u, \boldsymbol{x}, \boldsymbol{y}\right) = g\left(\boldsymbol{x}, \boldsymbol{y}\right) & \text{on } \partial\Omega\left(\boldsymbol{y}\right) \end{cases} \tag{1}$$

where $\mathcal{L}$ is a linear differential operator and $\mathcal{B}$ is the boundary operator, f and g are functions depending on the location in the space domain and the p parameters, where $\boldsymbol{x} \in \Omega(\boldsymbol{y}) \subset \mathbb{R}^d$ is the location in d dimensions, $\boldsymbol{y} \in \Pi \subset \mathbb{R}^p$ are the parameters, u is the solution. Eventually the approach can be trivially extended to IBVPs (Initial Boundary Value Problems).

The specific problem solved here is the incompressible, laminar and steady-state forced convection problem in an annular 3D pipe with wavy surfaces with fluid density ρ, kinematic viscosity ν and thermal diffusivity α. The thermo-fluid problem is described by the following conservation equations of mass, momentum and energy, respectively

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

$$(\boldsymbol{v} \cdot \nabla)\boldsymbol{v} = -\nabla p + \frac{1}{Re}\nabla^2 \boldsymbol{v} \tag{3}$$

$$\boldsymbol{v} \cdot \nabla T = \frac{1}{Re \cdot Pr}\nabla^2 T \tag{4}$$

where length, velocity $\boldsymbol{v} = \{v_1, v_2, v_3\}$, pressure p and temperature T are made nondimensional by taking L, U_0, ρU_0^2 and ΔT as reference quantities, respectively. $Re = U_0 L/\nu$ is the flow Reynolds number, $Pr = \nu/\alpha = 0.71$ is the fluid Prandtl number.

Exploiting the double symmetry of the considered pipe [22], the domain Ω, where the thermo-fluid problem is solved, is, in cylindrical coordinates, such that $(r, \theta, z) \in \Omega$ if:

$$r_{min} \leq r \leq r_{max} \tag{5}$$

$$0 \leq \theta \leq \frac{\pi}{2} \tag{6}$$

$$0 \leq z \leq H \tag{7}$$

with $r_{min} = R_1 + A_1\delta(\theta, z)$, $r_{max} = R_2 + A_2\delta(\theta, z)$, being R_1, R_2, H the internal radius, the external radius and the height of the undeformed cylinder respectively, and $\delta(\theta, z) = (\cos(2\theta - 1))(\cos(2\pi z/H) - 1)$ the wavy function.

The following boundary conditions are enforced: at the inlet ($z = 0$) $v_1 = v_2 = 0$ and $v_3 = T = u_z(r)$ with $u_z(r)$ as defined in [22]; at the outlet ($z = H$) $p = 0$ and $\partial T/\partial n = 0$; on the wavy surfaces ($r = r_{min}, r = r_{max}$) $\boldsymbol{v} = \boldsymbol{0}$ and $T = 0$ and conditions of symmetry for $\theta = 0$ and $\theta = \pi/2$.

Here, the problem depends on three parameters $\boldsymbol{y} = (Re, A_1, A_2) \in \Pi \subset \mathbb{R}^3$: the Reynolds number $Re \in [100, 500]$, the amplitude of the bump on the duct internal wall $A_1 \in [-0.0125, 0.0125]$ and the amplitude of the bump on the duct external wall $A_2 \in [-0.0125, 0.0125]$. $R_1 = 0.4$, $R_2 = 0.6$, $H = 1$ are constant.

As an example, in Figures 1 and 2, HF velocity and temperature fields are shown, respectively, for two different combinations of parameter values.

(a) $Re = 296$, $A_1 = 0.0113$, $A_2 = -0.0101$

(b) $Re = 357$, $A_1 = -0.0118$, $A_2 = 0.0109$

Figure 1: High-fidelity velocity magnitude fields varying both geometric and physical parameters on RBF-FD meshless internal nodes.

(a) $Re = 296$, $A_1 = 0.0113$, $A_2 = -0.0101$ (b) $Re = 357$, $A_1 = -0.0118$, $A_2 = 0.0109$

Figure 2: High-fidelity temperature fields varying both geometric and physical parameters on RBF-FD meshless internal nodes.

These solutions are obtained by the RBF-FD meshless method described below.

The parametric representation of velocity and temperature fields will be given, building a POD-based ROM from a finite set of RBF-FD meshless solutions.

3 Methodology

3.1 RBF-FD meshless discretization

In this section the RBF-FD method is briefly described. This section is divided into three subsections, following the corresponding building blocks of the RBF-FD approach: node generation, RBF-FD collocation and solution procedure.

3.1.1 Node generation

Let us consider the 3D pipe domain Ω, its boundary $\partial\Omega$ and a given spacing function $s(\boldsymbol{x})$ defining the desired, non-constant separation distance between the nodes. The employed spacing function is tuned in order to accurately capture the viscous and thermal boundary layers at the walls. The node generation phase can be divided into the following tasks (see [23, 24] for further details):

- generation of an initial node distribution within Ω that satisfies $s(\boldsymbol{x})$ on average. This distribution is not boundary conforming, i.e., nodes are not placed onto $\partial\Omega$.
- iterative refinement of the initial node distribution through a node-repel procedure, where the nodes are moved according to the mutual repulsion forces arising between neighbouring nodes. Boundary conformity is enforced at each refinement iteration by projection onto the nearest boundary point.

The resulting node distribution is composed by a set $\{\boldsymbol{x}_i\}_{i=1}^{N}$ of N nodes: N_I internal nodes in Ω and N_B boundary nodes on $\partial\Omega$, where $N_I + N_B = N$. Examples of such node distributions are shown in Figures 1 and 2 in terms of HF velocity and temperature fields at the nodes, and also in Figure 3.

3.1.2 RBF-FD collocation method

Referring to the general non parametric BVP problem, the employed RBF-FD collocation method is based on the Hermite RBF interpolation [25], where a suitable Radial Basis Function (RBF) local expansion u^h interpolates function values at the internal nodes and satisfies boundary conditions at the boundary nodes. RBFs are written as

$$\varphi_i(\boldsymbol{x}) = \varphi\big(||\boldsymbol{x} - \boldsymbol{x}_i||_2\big) \tag{8}$$

where φ is the basic function; in this work the MultiQuadric basic function has been chosen: $\varphi(r) = \sqrt{1 + (\epsilon r)^2}$ with shape factor $\epsilon = 0.5/s(\boldsymbol{x})$.

By considering a generic point $\boldsymbol{x}$ together with the associated stencil, i.e., the set $\{\boldsymbol{x}_i\}_{i=1}^m$ of m nearest nodes to $\boldsymbol{x}$, the local RBF expansion is written as follows

$$u^h(\boldsymbol{x}) = \sum_{i=1}^{m_I} \alpha_i \varphi_i(\boldsymbol{x}) + \sum_{i=m_I+1}^{m} \alpha_i \mathcal{B}\varphi_i(\boldsymbol{x}) + \sum_{j=1}^{q} \beta_j p_j(\boldsymbol{x}) \tag{9}$$

where $\mathcal{B}$ is the boundary operator associated to the generic boundary condition $\mathcal{B}u = g$, while $\{p_j\}_{j=1}^q$ is the canonical basis of the multivariate polynomial space $\mathbb{P}_P^3$ of total degree P in 3 dimensions. In this paper, $P = 3$ and $m = 40$ stencil nodes are chosen. The chosen order for the nodes in each stencil is such that the first $m_I \leq m$ are internal nodes, while the remaining are boundary nodes. The weights $\boldsymbol{\alpha} = \{\alpha_i\}_{i=1}^m$ and $\boldsymbol{\beta} = \{\beta_j\}_{j=1}^q$ are uniquely determined by writing the following local interpolation conditions at the stencil nodes:

$$u^h(\boldsymbol{x}_i) = u(\boldsymbol{x}_i), \quad i = 1, \dots, m_I \tag{10}$$

$$\mathcal{B}u^h(\boldsymbol{x}_i) = g(\boldsymbol{x}_i), \quad i = m_I + 1, \dots, m \tag{11}$$

and additional orthogonality conditions

$$\sum_{i=1}^{m} \alpha_i p_j(\boldsymbol{x}_i) = 0, \quad j = 1, \dots, q \tag{12}$$

are then enforced in order to ensure a unique solution [26]. By writing the above equations in matrix form, the following linear system is obtained:

$$\boldsymbol{M} \begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{pmatrix} = \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{g} \\ \boldsymbol{0} \end{pmatrix} \tag{13}$$

where $\boldsymbol{u} = \{u(x_i)\}_{i=1}^{m_I}$ and $\boldsymbol{g} = \{g(\boldsymbol{x}_i)\}_{i=m_I+1}^{m}$.

Given the following generic partial differential equation (PDE)

$$\mathcal{L}u = f \tag{14}$$

with $\mathcal{L}$ a linear differential operator, $\mathcal{L}$ can be applied to the RBF expansion, Equation (9), leading to the following FD (Finite Difference) formula

$$\mathcal{L}u^h(\boldsymbol{x}) = \boldsymbol{l}(\boldsymbol{x})^T \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{g} \\ \boldsymbol{0} \end{pmatrix} \tag{15}$$

where the vector of FD coefficients $\boldsymbol{l}(\boldsymbol{x})$ is obtained as the solution of the following linear system

$$\boldsymbol{M}^T \boldsymbol{l}(\boldsymbol{x}) = \Psi(\boldsymbol{x}) \tag{16}$$

In the previous equation, $\Psi(\boldsymbol{x}) = \{\mathcal{L}\psi_i(\boldsymbol{x})\}_{i=1}^{m+q}$ is the vector obtained by applying the differential operator $\mathcal{L}$ to each basis function $\psi_i(\boldsymbol{x})$ of the RBF expansion, Equation (9), i.e.

$$\psi_i(\boldsymbol{x}) = \begin{cases} \varphi_i, & \text{if } i \leq m_I \\ \mathcal{B}\varphi_i, & \text{if } m_I < i \leq m \\ p_{i-m} & \text{otherwise} \end{cases} \tag{17}$$

The FD formula expressed by Equation (15) is then employed to discretize the left-hand-side term in Equation (14) at a generic point $\boldsymbol{x}$: the resulting approximated equation is then made valid for each of the N_I internal nodes of the node distribution. By appropriately arranging the corresponding N_I equations, the following large and sparse linear system is obtained

$$\boldsymbol{L}\boldsymbol{u}_I = \boldsymbol{f}_I \tag{18}$$

which represents the discretized version of the initial PDE, Equation (14), and whose solution $\boldsymbol{u}_I$ rep-

resents the vector of the approximated solution values at the internal nodes. The interested reader is referred to [27, 28, 5, 29, 30] for the detailed description of the method.

3.1.3 Solution procedure

In order to solve the steady-state flow satisfying the Equations (2)-(4), the time dependent conservation equations

$$\nabla \cdot \boldsymbol{v} = 0, \tag{19}$$

$$\frac{\partial \boldsymbol{v}}{\partial t} + (\boldsymbol{v} \cdot \nabla)\boldsymbol{v} = -\nabla p + \frac{1}{Re}\nabla^2 \boldsymbol{v} \tag{20}$$

$$\frac{\partial T}{\partial t} + \boldsymbol{v} \cdot \nabla T = \frac{1}{Re \cdot Pr}\nabla^2 T \tag{21}$$

are actually discretized by RBF-FD collocation method and solved iteratively until steady-state convergence, assuming the viscous solution as the solution at initial time. Equations (19)-(21) are solved through a projection approach [31] with a three-level Gear scheme for the time discretization. A tentative velocity $\boldsymbol{v}^*$ is obtained from the linearized nondimensional momentum equation

$$\frac{3\boldsymbol{v}^* - 4\boldsymbol{v}^n + \boldsymbol{v}^{n-1}}{2\Delta t} + (\boldsymbol{v}^n \cdot \nabla)\,\boldsymbol{v}^* = -\nabla p^n + \frac{1}{Re}\nabla^2 \boldsymbol{v}^* \tag{22}$$

where n is the time level and Δt is the timestep size, chosen to be $\Delta t = 0.1$ for all the presented computations. Then $\boldsymbol{v}^*$ is forced to satisfy the mass conservation, Equation (2), by means of the following irrotational correction

$$\boldsymbol{v}^{n+1} = \boldsymbol{v}^* - \nabla \phi \tag{23}$$

which leads to the following Poisson equation in the auxiliary variable ϕ

$$\nabla^2 \phi = \nabla \cdot \boldsymbol{v}^* \tag{24}$$

The pressure is then updated as $p^{n+1} = p^n + \frac{\phi}{\Delta t}$ and the temperature T^{n+1} is computed from the time discretized version of the energy Equation (4):

$$\frac{3T^{n+1} - 4T^n + T^{n-1}}{2\Delta t} + \boldsymbol{v}^{n+1}\nabla T^{n+1} = \frac{1}{Re \cdot Pr}\nabla^2 T^{n+1} \tag{25}$$

Equations (22), (24) and (25) are discretized in space by means of the RBF-FD collocation approach presented in §3.1.2 and solved with BiCGSTAB iterative solver with an incomplete lower-upper preconditioner using a relative residuals tolerance of 10^{-8}.

Since steady-state solutions are sought, steady-state convergence is declared when

$$\frac{\|f^{n+1} - f^n\|}{\Delta t \|f^{n+1}\|} < 5 \cdot 10^{-6} \tag{26}$$

for each scalar field $f = v_1, v_2, v_3, p, T$.

3.2 Reduced order model

The solution of a FOM, usually requiring the solution of parameterized PDEs, can be very demanding. ROMs aim to be an inexpensive FOM approximation, being especially suited for many-query and real-time applications [32]. To do so, ROMs require a dimensionality reduction often achieved through projection methods such as Proper Orthogonal Decomposition (POD). When the projection accesses the problem governing equations, the ROM is called intrusive, and therefore, the ROM requires manipulation of the PDEs' source code. On the other hand, non-intrusive ROMs are purely data-driven, needing only a finite number of results from the FOM. Because of their differences, intrusive approaches are more adherent to the physics of the problem, while non-intrusive ones result in being more flexible. In the context of this work, we used non-intrusive POD-based ROMs. To approximate the FOM, the ROM computes a linear combination of basis functions resulting from a collection of FOM solutions.

3.2.1 Proper Orthogonal Decomposition

POD, also known as Karhunen-Loeve Expansion, identifies a set of orthogonal basis functions that maximize the energy norm associated with the base.

Referring to the general parametric BVP problem in Equation (1), given a field $\mathbf{u}_i \in \mathbb{R}^N$, called snapshot, obtained as a solution from the FOM, evaluated at N fixed locations $\{\boldsymbol{x}_k\}_{k=1}^N$ for a parameter $\boldsymbol{y}_i$, the collection $\mathbf{S} = \{\mathbf{u}_1, ..., \mathbf{u}_n\}$ is the snapshot matrix. Each element of $\mathbf{S}$ is a different solution of the FOM, here associated with a distinct parameter in $\{\boldsymbol{y}_i\}_{i=1}^n$ from the parameterized PDEs the FOM is solving. Consecutively, through an eigenvalue problem from the snapshot matrix $\mathbf{S}$, we compute a set of POD modes $\boldsymbol{\Xi} = \{\boldsymbol{\xi}_j\}_{j=1,...,n_{modes}}$, with $\boldsymbol{\xi}_j \in \mathbb{R}^N$. The modes will be a set of orthonormal basis functions, often obtained with techniques such as SVD (Singular Value Decomposition).

$$\boldsymbol{\Xi} \xleftarrow{SVD} \mathbf{S}$$

and for all the elements $\boldsymbol{\xi}_j$ is true that

$$\mathbf{S}\mathbf{S}^T\boldsymbol{\xi}_j = \lambda_j\boldsymbol{\xi}_j. \tag{27}$$

$\lambda_j \in \mathbb{R}$ is the singular value associated to the $\boldsymbol{\xi}_j$ mode $\forall\, j = 1, ..., n_{modes}$, and $\lambda_j > \lambda_{j+1}\, \forall\, j = 1, ..., n_{modes} - 1$. Since most of the information is relative to the greater singular values, the set $\boldsymbol{\Xi}$ is often truncated to the first r modes, having $r < n_{modes} \leq n$ without sacrificing the representability of the basis functions. A possible criterion to set r is to evaluate the POD energy value ε defined as

$$\varepsilon(r) = \frac{\sum_{j=1}^r \lambda_j}{\sum_{j=1}^n \lambda_j}. \tag{28}$$

then choose r when $\varepsilon(r) \leq \varepsilon_{threshold}$, where $\varepsilon_{threshold}$ is user-defined.

Through the reduced base $\boldsymbol{\Xi}$ POD modes, it is possible to encode each snapshot $\mathbf{u}_i \subset \mathbf{S}$, or any snapshot obtained from the FOM, by projection. A vector $\mathbf{a} \in \mathbb{R}^r$ can represent a single snapshot $\mathbf{u}$, where the components of $\mathbf{a}$ are called POD coefficients and are defined as

$$a_j = \mathbf{u} \cdot \boldsymbol{\xi}_j \quad \forall\, j = 1, ..., r \tag{29}$$

In general, the projection causes a loss of information, nevertheless, since $r \ll N$ the encoding $\mathbf{a}$ is a very low-dimensional representation of $\mathbf{u}$. We can retrieve an approximation of $\mathbf{u}$ through a linear combination of the POD modes $\boldsymbol{\Xi}$ with the respective POD coefficients $\mathbf{a}$

$$\mathbf{u} \sim \mathbf{u}^* = \sum_{j=1}^r a_j\boldsymbol{\xi}_j \tag{30}$$

From Equation (31) it is possible to evaluate the error introduced with the projection measuring the error field of a single snapshot, often using the L-2 norm in $\mathbb{R}^N$

$$e_{prj} = \frac{\|\mathbf{u} - \mathbf{u}^*\|}{\|\mathbf{u}\|} \tag{31}$$

3.2.2 Gaussian Process regression models

For a parametric problem, the single snapshot is associated to the vector parameter $\boldsymbol{y} \in \Pi \subset \mathbb{R}^p$, where p is the number of parameters, so that $\mathbf{u}_i = \mathbf{u}(\boldsymbol{y}_i)$. Accordingly, the POD coefficient for the j-th mode, as in Equation (29), becomes parameter dependent $a_j = a_j(\boldsymbol{y})$. Given a set of snapshots to perform the POD, they define a set of known POD coefficients for each mode. Using these coefficients to train a map

$$\hat{a}_j(\boldsymbol{y}) : \boldsymbol{y} \longrightarrow a_j \quad \forall j = 1, ..., r \tag{32}$$

in combination with the POD modes $\boldsymbol{\xi}_j$ allow to explore an approximation of any solution $\mathbf{u}$ given any value of $\boldsymbol{y}$. To do so, the map can be obtained through different regression models, such as GPs.

In more general terms, a GP regression model aim to surrogate the output of a system by the means of a GP [33]. Given a finite number of known parameter-observation couples $(\boldsymbol{y}, a_j)$, with $\boldsymbol{y} \in \mathbb{R}^p$ being the parameter and $a_j \in \mathbb{R}$ the output, it is possible to fit a GP to predict the output for unknown parameter configurations. Since a GP is a collection of random variables, any finite number of which has a joint Gaussian distribution, it is fully defined by a mean function $m(\boldsymbol{y})$ and a covariance function $k(\boldsymbol{y}, \boldsymbol{y}')$, or kernel, with the notation

$$GP = \mathcal{N}\left(m(\boldsymbol{y}), k(\boldsymbol{y}, \boldsymbol{y}')\right)$$

where $\mathcal{N}$ denotes a normal distribution. It will be assumed $\hat{a}_j(\boldsymbol{y}) = m_{a_j}(\boldsymbol{y})$ to model $a_j(\boldsymbol{y})$, with $\sigma^2_{a_j}(\boldsymbol{y})$ the uncertainty of the model $\hat{a}_j(\boldsymbol{y})$.

Kriging is a GP regression model that approximates an output a_j, considered as a realisation of a multivariate Gaussian process A_j, solving a MLE (Maximum Likelihood Estimation) problem on the model's hyperparameters $\boldsymbol{\theta}$. In particular, the hyperparameters completely characterizes both the GP's mean and kernel functions. Given $\boldsymbol{\theta}$ and a set of observations $\mathbf{a}_j$, realisations of the Gaussian vector $\mathbf{A}_j$, associated with a set of parameters $\mathbf{Y} = \{\boldsymbol{y}_i\}_{i=1}^n$, it is possible to compute the likelihood function $l(\mathbf{Y}, \mathbf{a}_j; \boldsymbol{\theta})$, which measures the probability that the sample $\mathbf{a}_j$ can be obtained from the $\mathcal{N}\left(m(\mathbf{Y}; \boldsymbol{\theta}), k(\mathbf{Y}, \mathbf{Y}'; \boldsymbol{\theta})\right)$. The MLE becomes an optimisation problem to find the optimal hyperparameters $\boldsymbol{\theta}_{opt}$ so that

$$\boldsymbol{\theta}_{opt} = \arg\max_{\boldsymbol{\theta}} l(\mathbf{Y}, \mathbf{a}_j; \boldsymbol{\theta})$$

In general, there are no assumptions regarding the structure of $m(\boldsymbol{y})$, and can be considered independent from $\boldsymbol{\theta}$. On the other hand, the kernel $k(\boldsymbol{y}, \boldsymbol{y}'; \boldsymbol{\theta})$ is the main contributor to the MLE problem. Some of the GP kernel formulations are the Squared Exponential, the Matern 3/2, or the Matern 5/2.

$$k_{SE}(\boldsymbol{y}, \boldsymbol{y}') = \sigma^2 \exp\left(-\frac{1}{2}\frac{\|\boldsymbol{y} - \boldsymbol{y}'\|^2}{l^2}\right) \tag{33}$$

$$k_{Mat32}(\boldsymbol{y}, \boldsymbol{y}') = \sigma^2 \left(1 + \frac{\sqrt{3}\|\boldsymbol{y} - \boldsymbol{y}'\|}{l}\right) \exp\left(\frac{\sqrt{3}\|\boldsymbol{y} - \boldsymbol{y}'\|}{l}\right) \tag{34}$$

$$k_{Mat52}(\boldsymbol{y}, \boldsymbol{y}') = \sigma^2 \left(1 + \frac{\sqrt{5}\|\boldsymbol{y} - \boldsymbol{y}'\|}{l} + \frac{5\|\boldsymbol{y} - \boldsymbol{y}'\|^2}{3l^2}\right) \exp\left(\frac{\sqrt{5}\|\boldsymbol{y} - \boldsymbol{y}'\|}{l}\right) \tag{35}$$

where σ^2, l are examples of kernel hyperparameters, called variance and lenghtscale. Equations (33) - (35) are isotropic representations of the kernels, while with automatic relevance determination [33] it is possible to obtain anisotropic kernels in $\mathbb{R}^p$, having p lengthscale values instead of just 1.

To predict any unknown output A_j^*, knowing the optimal values of the hyperparameters $\boldsymbol{\theta}_{opt}$, the prior distribution of a the random vector $\left[\mathbf{A}_j, A_j^*\right]$ belonging to the GP will be

$$\begin{bmatrix}\mathbf{A}_j \\ A_j^*\end{bmatrix} = \mathcal{N}\left(\mathbf{0}, \begin{bmatrix}k(\mathbf{Y}, \mathbf{Y}) & k(\mathbf{Y}, \mathbf{Y}^*) \\ k(\mathbf{Y}^*, \mathbf{Y}) & k(\mathbf{Y}^*, \mathbf{Y}^*)\end{bmatrix}\right) \tag{36}$$

assuming without loss of generality a 0 prior mean. Since the variables in $\left[\mathbf{A}_j, A_j^*\right]$ are joint normal, the posterior distribution is known

$$\left(A_j^* \,|\mathbf{Y}^*, \mathbf{Y}, \mathbf{A}_j\right) = \mathcal{N}(m_{a_j}(\boldsymbol{y}), \sigma^2_{a_j}(\boldsymbol{y})) \tag{37}$$

with

$$m_{a_j}(\boldsymbol{y}) = k(\mathbf{Y}^*, \mathbf{Y}) k(\mathbf{Y}, \mathbf{Y})^{-1} Y \tag{38}$$

$$\sigma^2_{a_j}(\boldsymbol{y}) = k(\mathbf{Y}^*, \mathbf{Y}^*) - k(\mathbf{Y}, \mathbf{Y}^*) k(\mathbf{Y}, \mathbf{Y})^{-1} k(\mathbf{Y}^*, \mathbf{Y}) \tag{39}$$

where $\mathbf{Y}^*$ is the matrix of parameters associated to the unknowns A_j^*, and with $|$ we refer to the conditional probability. Subsequently, the mean of the posterior distribution in Equation (37) approximates

the values A_j^*, and the associated covariance matrix principal diagonal provides the variances of the predictions, allowing the prediction's confidence interval quantification.

Since all outputs $\mathbf{a}_j$ come from the same system, the approximation obtained with a Kriging is considered single-fidelity. Conversely, if the $\mathbf{a}_j$ were obtained from heterogeneous systems, there would have been the necessity to use a multi-fidelity model.

NARGP (Nonlinear AutoRegressive multi-fidelity Gaussian Process regression) is a multi-fidelity regression model [34] based on GP regression, that allows to overcome the limitations of linear autoregressive schemes such as cokriging models [35, 36].

The NARGP is recursive so that we can consider any number of fidelity levels, and, without loss of generality, we will take into account only two fidelity levels, LF (low-fidelity) and HF (high-fidelity), respectively. The LF and HF models handle the same inputs $\boldsymbol{y}$, but have different outputs $a_{jLF}(\boldsymbol{y})$ and $a_{jHF}(\boldsymbol{y})$, respectively, with $\boldsymbol{y} \in \mathbb{R}^p$. Usually, for the problem of practical interest, the LF information is cheaper and easier to obtain when compared to the HF one.

$$\begin{aligned}&\mathbf{Y}_{LF} = \{\boldsymbol{y}_i\}_{i=1,\dots,n_{LF}}\\&\mathbf{Y}_{HF} = \{\boldsymbol{y}_j\}_{j=1,\dots,n_{HF}} \subset \mathbf{Y}_{LF}\\&n_{LF} >> n_{HF}\end{aligned}$$

Firstly, a single-fidelity GP regression model is necessary to train the NARGP. We call $\hat{a}_{jLF}(\boldsymbol{y}) \sim a_{jLF}(\boldsymbol{y})$ the LF model, which can approximate the LF function for any input parameter $\boldsymbol{y}$. As a result, we can compute the posterior mean of the LF GP for the inputs of the known HF training points, getting

$$\mathbf{m}_{LF}^{HF} = \hat{a}_{jLF}(\mathbf{Y}_{HF}) \tag{40}$$

where here $\mathbf{m}_{LF}$ is the LF GP posterior mean. In the NARGP $\mathbf{m}_{LF}^{HF}$ is exploited to enrich its HF input $\mathbf{XX}_{HF}$, having

$$\mathbf{YY}_{HF} = [\mathbf{Y}_{HF},\, \mathbf{m}_{LF}^{HF}] \in \mathbb{R}^{n_{HF} \times d+1} \tag{41}$$

To make full use of the multi-fidelity input $\mathbf{XX}_{HF}$ at the HF level, an appropriate kernel is mandatory. Products and sums of positive semi-definite functions are again positive semi-definite, consequently the combination of kernel

$$k_{HF}((\boldsymbol{y}, \mathbf{m}_{LF}(\boldsymbol{y})), (\boldsymbol{y}', \mathbf{m}_{LF}(\boldsymbol{y}')) = k_\rho(\boldsymbol{y}, \boldsymbol{y}') \cdot k_f(\mathbf{m}_{LF}(\boldsymbol{y}), \mathbf{m}_{LF}(\boldsymbol{y}')) + k_\delta(\boldsymbol{y}, \boldsymbol{y}') \tag{42}$$

is still a valid kernel. Different input components are managed by various parts of the kernel. Through the kernel's shape optimization achieved in solving a MLE problem, the HF model of the NARGP can infer the interconnection between LF and HF information, without the need for any prior assumption on their relation. However, the complexity of the model increases due to the composite kernel, hence more hyperparameters have to be optimized.

Given a set of unknown input values $\mathbf{Y}^*$, the NARGP needs first the LF GP prediction $\hat{a}_{jLF}(\mathbf{Y}^*)$. Then, the multi-fidelity input for the NARGP's HF level will be $\mathbf{YY}^* = [\mathbf{Y}^*, \hat{a}_{jLF}(\mathbf{Y}^*)]$, allowing the NARGP to approximate the unknown outputs $\mathbf{a}_{jHF}^*$ using the enriched input, and the optimized kernel from Equations (41) and (42) respectively, in the posterior distribution in Equation (37). Since the LF GP output has some uncertainty, represented by its variance, a Monte Carlo integration [34] is required to compute the posterior distribution of the HF NARGP prediction.

In the context of a multi-fidelity ROM, given a decomposition of the snapshots, the projection of the LF snapshots onto the reduced basis can be used as the LF approximation of the HF POD coefficients.

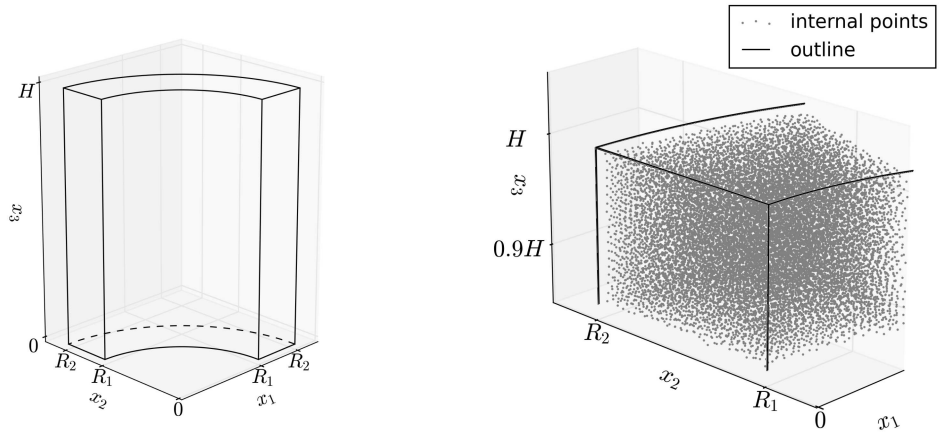
4 Implementation

4.1 DOE generation

The parameter space has been sampled with a Halton sequence, identifying 300 DOEs (Designs of Experiments). The RBF-FD meshless method has been used to simulate each DOE. Both LF and HF solutions have been computed, obtained with $75k$ nodes and $500k$ nodes, respectively.

Some validation sets are drawn from the DOE HF simulations to assess the quality and robustness of the POD-based ROM models, built using an increasing number of LF simulations, HF simulations and their combinations.

POD requires a fixed fluid domain discretization. In particular, all the snapshots should have the same number of points while keeping a coherent spatial allocation. To this end, we decided to use a portion of a reference cylinder Ω_{ref} with a fixed cloud point distribution, as in Figure 3. For this application, we generated a point cloud distribution appropriate for a high-fidelity solution with $600k$ points, $546k$ of which are internal points, and the remaining ones are on the boundary $\partial\Omega_{ref}$.



(a) Outline.

(b) Close up with internal points.

Figure 3: Reference cylinder Ω_{ref}.

Since the DOE includes geometrical deformations, as described in Section 2, the geometry associated with each solution will be different. Therefore, there is a need to map every geometry on the reference cloud point distribution in Figure 3. A displacement is applied to the reference points so they live in the same locus of the respective RBF-FD solution Ω. To do so, we exploited the definition of the computational domain Ω given in Section 2. A function $A(r) : r \longrightarrow A$ that assign an amplitude to the wavy function for any value of $r \in [R_1, R_2]$ can be found with a linear interpolation between the couples (R_1, A_1) and (R_2, A_2). By doing so, any point $\mathbf{P} = (r, \theta, z) \in \Omega_{ref}$ can be mapped in $\mathbf{P'} = (r', \theta', z') \in \Omega$ with a smooth and unique transformation $\mathcal{M}$

$$\mathcal{M}(\mathbf{P}) = \mathbf{P'} = \left\{ \begin{array}{c} r + A(r)g(\theta, z) \\ \theta \\ z \end{array} \right\} \tag{43}$$

consisting in a radial displacement of the reference points.

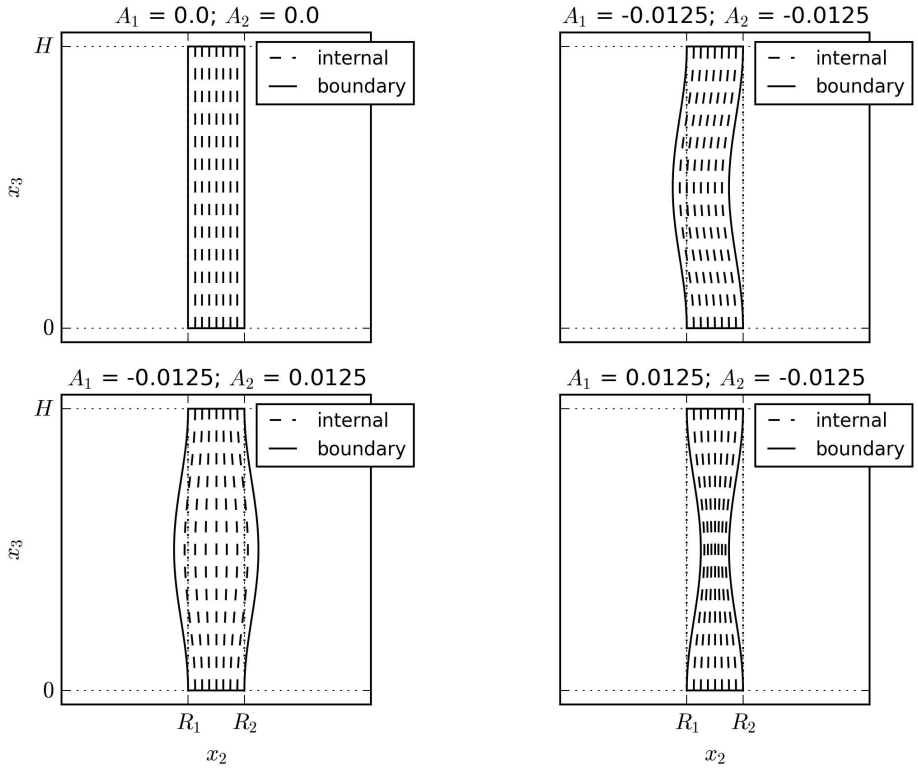


Figure 4: Mapping example for different Ω configurations, slice in the plane x_2, x_3 at $x_1 = 0$. Top left: undeformed; Top right and bottom: extreme A_1, A_2 values from the DOE. Internal and boundary lines on the surface are represented to visualise how the displacement works.

Subsequently it is possible to interpolate the solution field on the reference points in Ω_{ref} with a k-nearest neighbour interpolation [37]. Both LF and HF solution are mapped on the same Ω_{ref} discretization. An example of mapping for different Ω configurations is shown in Figure 4.

4.2 ROM set-up

In this work we want to approximate a total of 3 fields, each of them with a separate ROM. Namely, we have the pressure field p, the temperature field T and the velocity field U. The latter has to be intended as a vector field composed by all three components of the velocity.

Decomposition To perform the POD necessary for the ROMs, an energy value $\varepsilon = 0.9999$ has been used for every field of interest. Remembering the definition in Equation (28), the higher the ε the higher the number of the kept POD modes. For both T and U fields, more basis functions do not improve the representability of the basis. Sensitive improvements can be found for the p when adding more snapshots, however, the lower projection error, defined in Equation (44), did not show any improvement for the ROM reconstructions. This will be better explained in Section 5.

GP regression For both the GP and NARGP regression models of the POD coefficients we used squared-exponential kernels, as in Equation (33), with automatic relevance determination [33] to model the anisotropy along different dimensions. The same snapshots used for the decomposition were used to obtain the training POD coefficients. We supposed output observations to be noisy, so we assumed a Gaussian noise standard deviation hyperparameter in the order of $10^{-2}/10^{-3}$ of the normalized output. To improve the hyperparameters optimization robustness, we used a low-memory BFGS algorithm with 15 random restarts to solve the MLE for each POD coefficient model.

Multi-fidelity ROM For what concerns the multi-fidelity ROM, both low- and high-fidelity snapshots were mixed at the decomposition step. This has been done since it allowed us to obtain a reduced basis with better representability, especially for U and T fields. This can be observed in Figure 5, where the singular values are compared for HF, LF and multi-fidelity POD from 20 HF snapshots, 20 LF snapshots and 20 HF plus 150 LF snapshots, respectively. The larger the singular value, the more important the corresponding dimension is for capturing the variability of the data. Here HF and LF ROMs have similar singular values, while the multi-fidelity model counts many more modes. This difference arises from the low- and high-fidelity snapshot mixing computing the POD. This consideration, related to the particular case, can be extended in general. Out of curiosity, in Appendix A the HF POD modes related to the singular values in Figure 5 are shown.

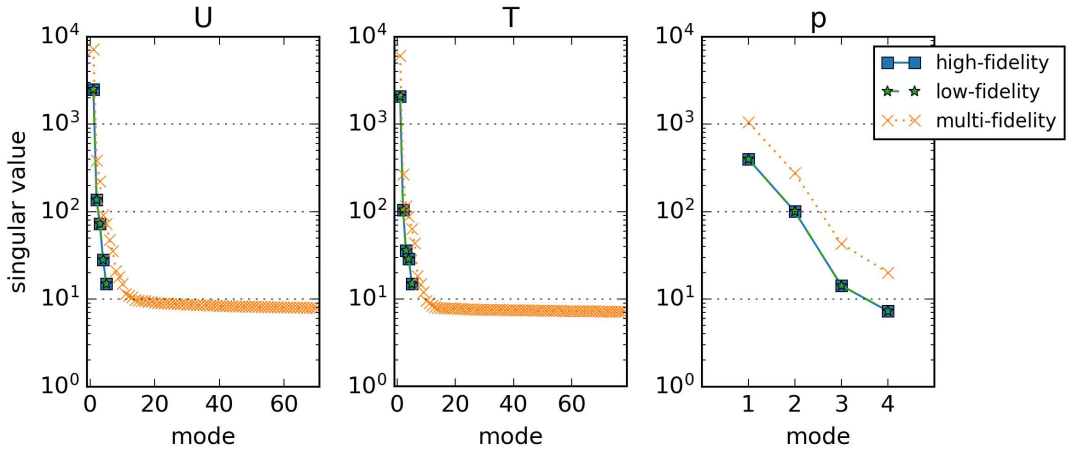


Figure 5: Singular values for HF, LF and multi-fidelity POD from 20 HF snapshots, 20 LF snapshots and 20 HF plus 150 LF snapshots, respectively.

We choose to infer the POD coefficients with the multi-fidelity model for those modes that showed a Pearson's R absolute value between LF and HF representation of the POD coefficients above 0.5, with a p-value of less than 5%. The remaining POD coefficients were approximated with single-fidelity HF GP regression models. Indeed, this choice permits to easily automate the decision regarding the multi-fidelity approximation, while allowing to use the multi-fidelity approximation also with weak linear correlations. This does not represent a problem since the NARGP is designed to handle non-linear relationships between data with different fidelities. An intuition regarding the motivation to use multi-fidelity approaches can be found in Figure 9, where the first modes show a good linear correlation between LF and HF POD coefficients.

5 Results

The error metrics were evaluated on 60 HF reserved validation snapshots from the tail of the Halton sequence used in the DOE creation. Consequently, the validation snapshots evenly cover the parameter space, avoiding overlap with any training snapshot. In the validation database, there are both designs for which the ROM approximations can be considered as interpolations and extrapolations since some of them will stay outside the convex hull defined by the training snapshots in the input parameter space.

The projection error measures the decomposition capabilities to represent unseen snapshots with the given basis. Intuitively, it can be considered the best possible approximation that the ROM can provide. Given a set of validation snapshots $\{\mathbf{u}_i^{test}\}_i$, the mean projection error is

$$Err_{prj} = \frac{1}{n_{test}} \sum_{i=1}^{n_{test}} \frac{\left\| \left(\sum_{j=1}^{n_{modes}} (\mathbf{u}_i^{test} \cdot \boldsymbol{\xi}_j) \boldsymbol{\xi}_j \right) - \mathbf{u}_i^{test} \right\|_2}{\|\mathbf{u}_i^{test}\|_2} \tag{44}$$

On the other hand, the reconstruction error measures how much the ROM can replicate the validation snapshot. Similarly to Equation (44), it is defined as

$$Err_{recon} = \frac{1}{n_{test}} \sum_{i=1}^{n_{test}} \frac{\left\| \left(\sum_{j=1}^{n_{modes}} \hat{a}_j(\boldsymbol{y}_i^{test}) \boldsymbol{\xi}_j \right) - \mathbf{u}_i^{test} \right\|_2}{\left\| \mathbf{u}_i^{test} \right\|_2} \tag{45}$$

where $\hat{a}_j$ represent the approximation of the POD coefficient as in Equation (32).

Figure 6 shows the projection error for the HF ROM for the temperature, velocity, and pressure fields as the POD energy value changes. On the basis of these results, the choice to use an energy value $\varepsilon = 0.9999$ for POD modes truncation used in following analyses.

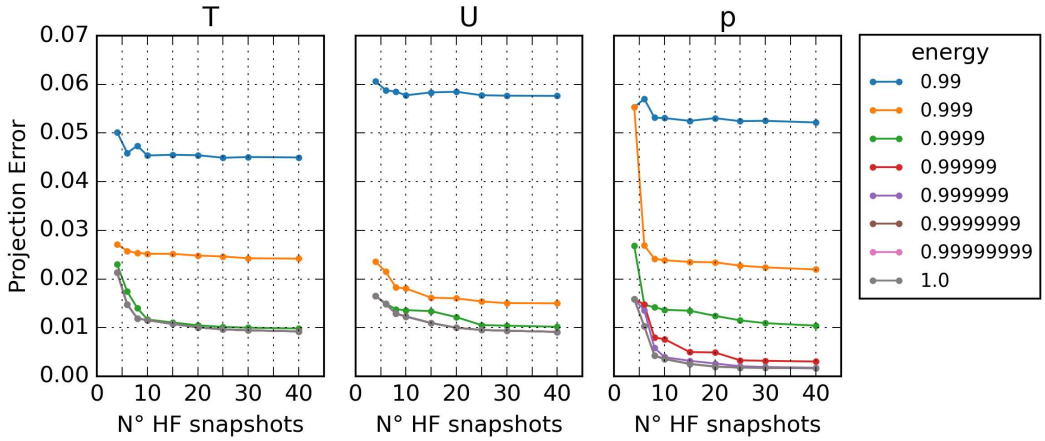
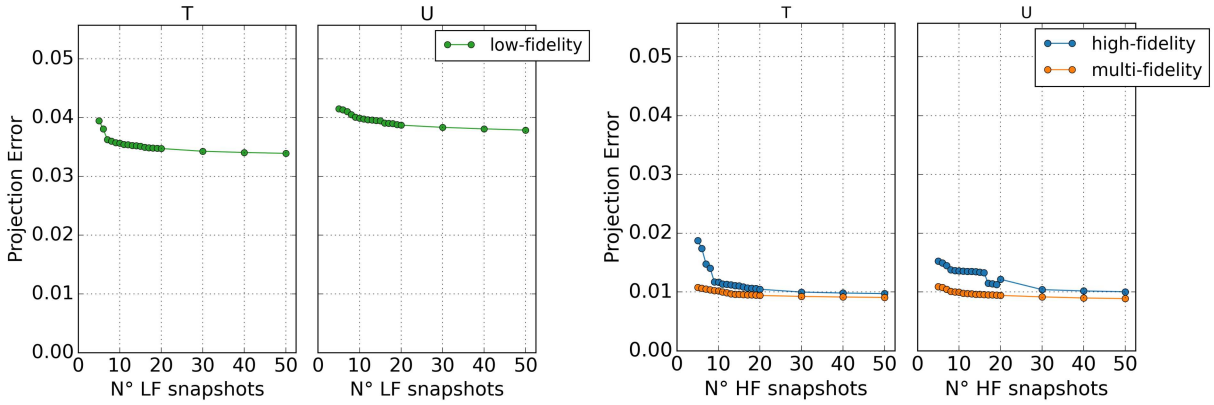


Figure 6: Projection error for a single-fidelity HF ROM, given different energy values.

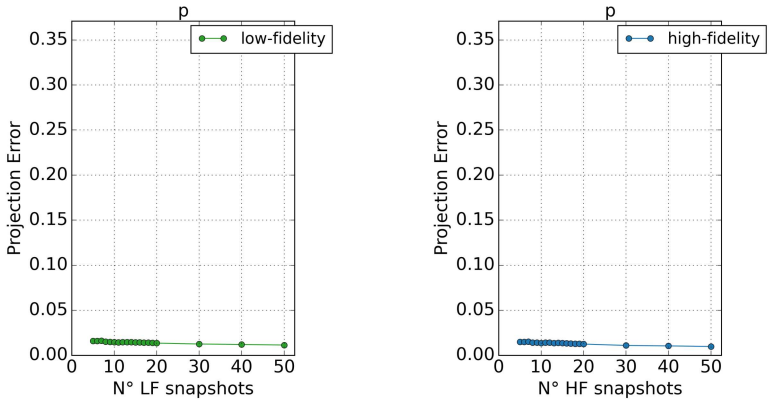
In Figures 7 and 8, there are, respectively, the projection error increasing the LF snapshots number for the LF ROM and the projection error increasing the HF snapshots number for both the HF and MF ROMs. For MF ROMs, the same 150 LF snapshots were always used increasing the number of HF snapshots, so that these are always nested to LF snapshots. For temperature and velocity fields, the high- and multi-fidelity decompositions better represent the validation snapshots compared to the LF ones, with a difference of around 3% in the projection error. For this application, adding more than 30 HF snapshots does not lower the projection error. Unlike the temperature and velocity fields, the LF and HF pressure fields are very similar, as shown in Figure 15, so there are no sensitive differences between the LF and HF projection errors.



(a) LF ROM.

(b) HF and multi-fidelity ROM.

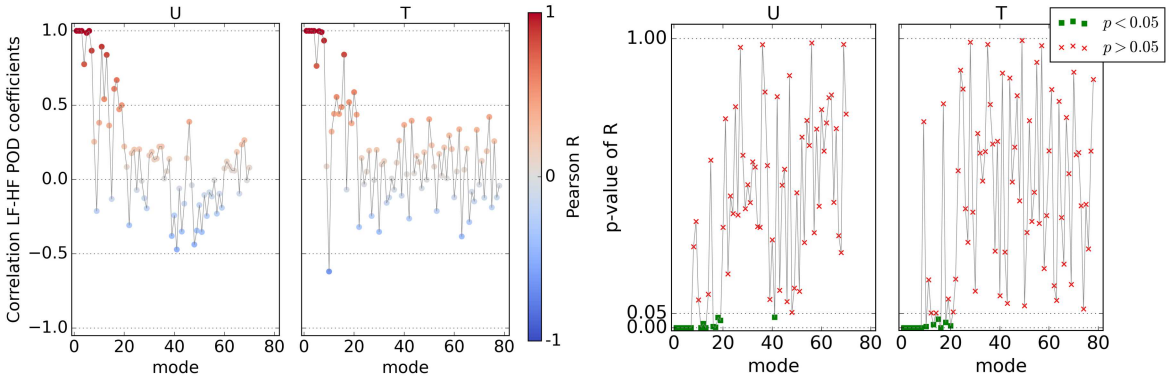
Figure 7: Temperature and velocity LF projection error on the validation snapshots, LF ROM vs HF/multi-fidelity ROM.



(a) LF ROM. (b) HF and multi-fidelity ROM.

Figure 8: Pressure LF projection error on the validation snapshots, LF ROM vs HF/multi-fidelity ROM.

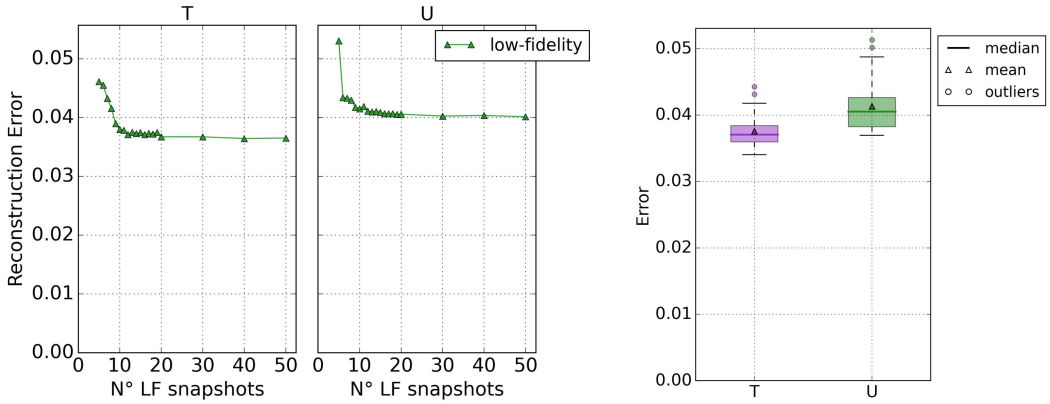
Now, before comparing the reconstruction error of LF, HF and MF ROMs, let's make a brief digression to understand how the multi-fidelity ROM models were constructed. When the HF and LF snapshots are mapped accordingly onto the same reference discretization, they can be projected on the same POD modes. Given a mixed-fidelity POD for each mode, we can find a LF and HF representation of the POD coefficient. Therefore, we can evaluate the linear correlation between different fidelities POD coefficients, as in Figure 9, with the purpose of understanding whether we can exploit it. We used Pearson's R test to measure the correlation, and the first modes showed a strong relationship. This is expected since most of the predominant physics is generally related to macro-scale behaviors, and the first POD modes often describe low-frequency components of the solution, and LF simulations can capture only such type of information. On the other hand, lower energy multi-fidelity modes do not show any correlation between fidelities. This implies that to infer the POD coefficients the multi-fidelity model is useful only when LF and HF POD coefficients show some correlation. This is the reason why the remaining POD coefficients were approximated with single-fidelity HF GP regression models.



(a) Pearson's R correlation coefficient. (b) P-value.

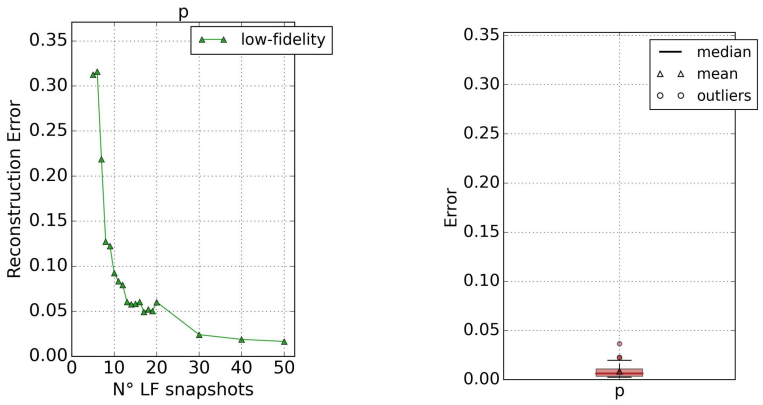
Figure 9: Correlation analysis with Pearson's R test between low- and high- fidelity representation of the POD coefficients; multi-fidelity ROM.

Let us now analyse the reconstruction errors. In Figures 10 and 11, the LF ROM reconstruction errors are compared to direct HF solution approximations with LF solutions. In general, with 20 LF temperature or velocity snapshots we reach almost the same accuracy of the LF solutions. For the pressure, more LF snapshots are needed to obtain similar results.



(a) LF ROM (b) LF solutions.

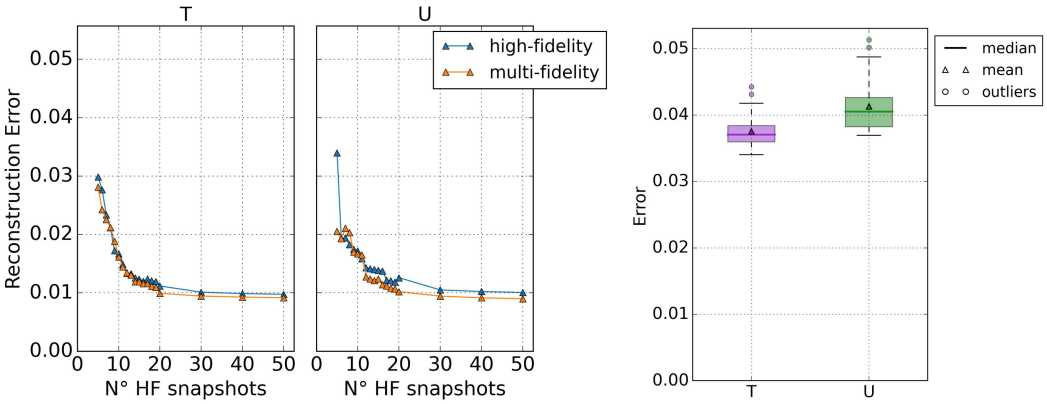
Figure 10: Temperature and velocity LF approximations of HF validation snapshots; (a) LF ROM mean reconstruction error, (b) box-plot of LF solution errors.



(a) LF ROM (b) LF solutions.

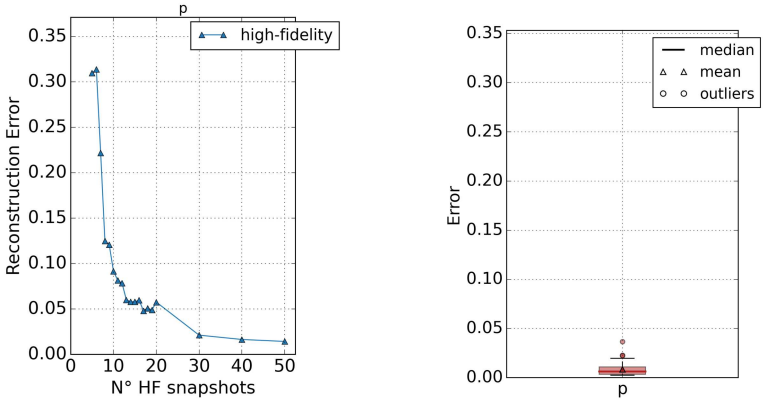
Figure 11: Pressure LF approximations of HF validation snapshots; (a) LF ROM mean reconstruction error, (b) box-plot of LF solution errors.

Similarly, Figures 12 and 13, present the high- and multi-fidelity reconstruction error. Due to the two ROMs having close performances for what concerns the pressure field, we omit the multi-fidelity approach. Indeed, we can see an improvement in adopting the high- and multi-fidelity ROMs of the temperature and velocity fields instead of using the LF solution. With 20 or more HF snapshots, the two ROMs outperform the LF solver accuracy, reducing the error from $3.5 - 5\%$ to even below 1%. On the other hand, the pressure ROM is almost equivalent to the LF pressure solutions only with 50 or more HF snapshots. With fewer HF snapshots, the reconstruction errors are higher than the velocity and temperature field by an order of magnitude. Together with low projection errors, this suggests that the POD coefficients have a higher complexity. Therefore, more information is required to approximate them.



(a) HF and multi-fidelity ROM (b) LF solutions.

Figure 12: Temperature and velocity approximations of HF validation snapshots; (a) high- and multi-fidelity ROM mean reconstruction error, (b) ox-plot of LF solution errors. 150 LF snapshots have been added for the multi-fidelity approach.



(a) HF ROM (b) LF solutions.

Figure 13: Pressure approximations of HF validation snapshots; (a) HF ROM mean reconstruction error, (b) box-plot of LF solution errors.

Figure 14 shows a close-up on the errors of the high- and multi-fidelity temperature and velocity ROMs trained with 10 to 20 HF snapshots. Even if the performances are very close between the two models, it is possible to notice two main features. Firstly, the multi-fidelity ROM performs better than the HF ROM when expected to do so, namely in HF poor information conditions. Secondly, the multi-fidelity velocity prediction shows enough improvement to bring the reconstruction error below the HF ROM projection error, which, for non-intrusive ROMs, represents the lower limit for the reconstruction error. This behavior is possible only thanks to the LF and HF snapshots mixing in the multi-fidelity ROM. As shown in Figure 6, the HF ROM can not reduce the projection error below a limit value, even using more modes.

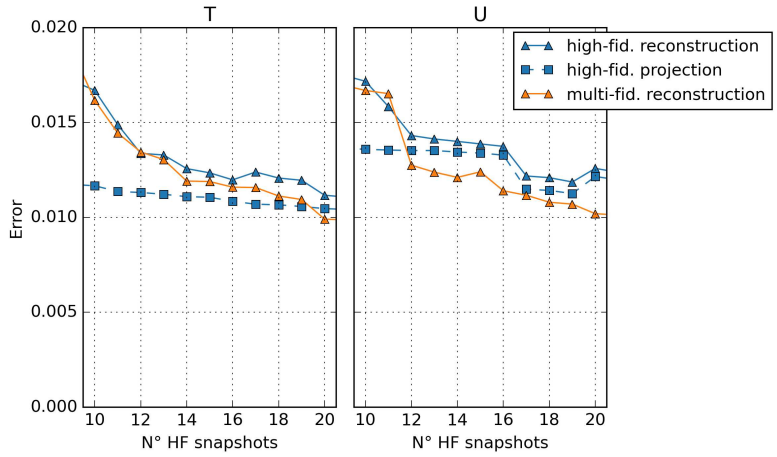


Figure 14: Close up on high- and multi-fidelity ROMs between 10 and 20 HF trainign snapshots.

To better understand these results, Figure 15 compares the HF solutions of a validation with the LF solutions and the HF ROM approximations for each field of interest. The validation design, used as an example, is computed for $Re = 387$, $A_1 = -0.0074$ and $A_2 = 0.0116$. Here higher deviations from the diagonal highlight a poor agreement with the HF solution. The larger LF temperature and velocity solution differences with the HF solutions reconfirm the results of Figure 12. Most of the LF velocity deviation in Figure 15 is around $U = 0$, mainly representing the flow near the wall, conversely to the more accurate HF ROM. It suggests that the HF ROM is a better fit to study problems where there is a need for a better description of the boundary layer than coarse LF solutions. On the other hand, the LF pressure solution is already a good fit for the HF problem, performing better than a high-fidelity ROM trained on 20 HF snapshots.

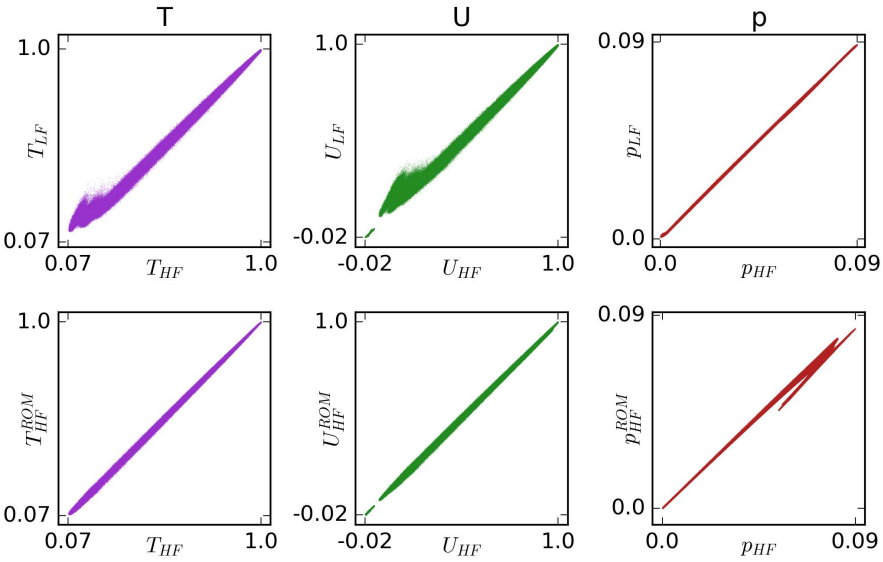
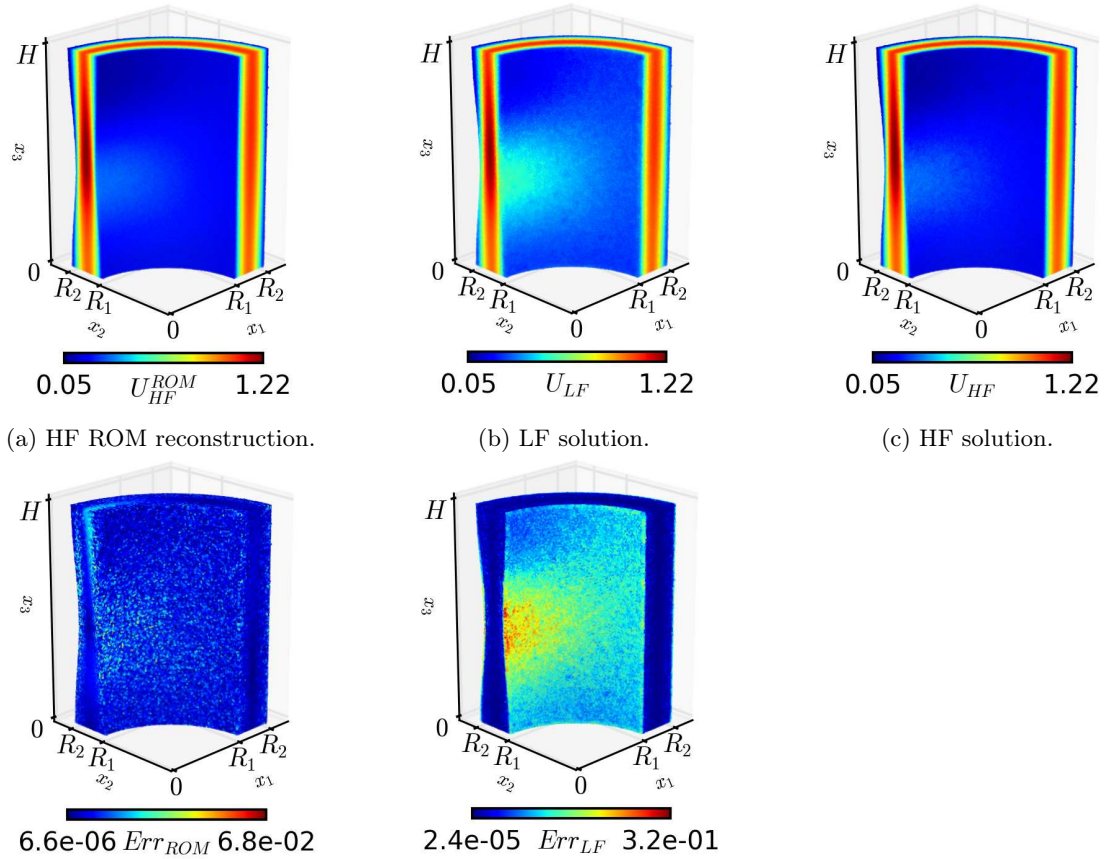


Figure 15: Scatter plots of the temperature, velocity and pressure fields comparing the HF results with the LF ones (first row), and the HF ROM (second row) for an out of sample design.

In Figure 16, again as an example, there is a comparison between the LF solution and the HF ROM reconstruction for the velocity field when approximating the HF solution. The chosen design is out of sample and corresponds to the one with the median error for the HF ROM trained with 20 HF snapshots. This design has $Re = 284$, $A_1 = 0.0062$ and $A_2 = -0.0097$. Figure 16 shows that most of the near-wall solution is well approximated, as previously pointed out. Even if the LF approximation of the bulk is accurate, most of the near-wall solution is poor. Since most of the computational nodes are close to

the wall, and the LF absolute errors are higher than that of the HF ROM, it well explains the better performance of the HF ROM.



(a) HF ROM reconstruction. (b) LF solution. (c) HF solution.

(d) HF ROM absolute error. (e) LF solution absolute error.

Figure 16: Velocity magnitude field with parameters $Re = 284$, $A_1 = 0.0062$ and $A_2 = -0.0097$. The HF ROM is trained with 20 HF snapshots. For reference, the absolute error field has a reconstruction error of 1% in (d) and 5% in (e).

Let us add some final considerations on the computational time demands. Table 1 shows the computational time needed to train a high- and multi-fidelity ROM. All computation are done with a Intel Xeon E5-2630 v3 CPU.

	High-fidelity ROM	Multi-fidelity ROM
Decomposition time $[s]$	7.7	87.0
Regression time $[s]$	4.1	316.9
Number of POD modes	5	70

Table 1: Single and multi-fidelity velocity ROM's computational training time; 20 HF snapshots, plus 150 LF snapshots for the multi-fidelity ROM.

The offline computational time required to obtain the LF and HF solutions can be found in Table 2.

RBF-FD meshless simulation	Low-fidelity	High-fidelity
number of nodes	$75k$	$500k$
wall-time	3 to 6 minutes	20 to 60 minutes

Table 2: Details on RBF-FD meshless simulations.

The snapshot preparation is negligible in time if compared to the CFD simulations. The two most resource-intensive operations are the POD and the training of the POD coefficient regression models.

The HF ROM is computationally cheaper than the multi-fidelity ROM. The main reason stays in the lower number of POD modes, given the same POD energy, in addition to the fact that the NARGP model used for some of the coefficients is more expensive to train than a classical GP. The parallelization of the training could significantly improve the regression's efficiency. Despite this, the training of both ROMs is overall inexpensive since most of the computational time for the ROM is used to perform the CFD simulations.

6 Conclusion and Future Work

In this work, we proposed a non-intrusive ROM originally combined with an RBF-FD meshless method to solve parametric thermofluid dynamic problems, focusing on the geometrical and physical parametrization of a 3D test case. Having different fidelity representations of the solution, namely the HF and LF, we compared the performances of various approaches. Firstly, we showed that an LF ROM can replace the LF solution, however, LF models fail to approximate the HF velocity and temperature. Secondly, the HF ROM can better surrogate velocity and temperature using a limited number of snapshots. The proposed MF-ROM approach showed modest improvements over the HF ROM while increasing the model's complexity. Moreover, the MF-ROM proved that adding LF information allows for lowering the reconstruction error below the single-fidelity projection error. Nonetheless, a ROM represents an approximation of the full-order model, so it is crucial to validate the ROM to avoid propagating errors into the simulation process. Furthermore, the ROM is expected to be accurate only in the proximity of the training data. With this application, we showed that:

- ROMs have to be tailored to specific systems, indeed proving to be worthed only for analyses needing many simulations.
- For a 3D pipe with wavy surfaces, we could accurately explore a broad parameter domain with a limited number of full-order simulations.
- A low snapshot count allowed for the ROM computational cost minimization since the full-order simulations are an order, or more, of magnitude more expensive than the ROM training.

Alongside these positive aspects, the RBF-FD meshless method reduced the complexity of the problems, easing the full-order snapshot evaluation.

Future work will focus on applying the proposed methodologies to more complex test cases to further evaluate and refine the performance. In particular, we aim to improve the capabilities of the multi-fidelity ROM since, for the relative simplicity of the analysed problem, it showed modest but promising improvements over the single-fidelity ROM.

A High Fidelity POD modes

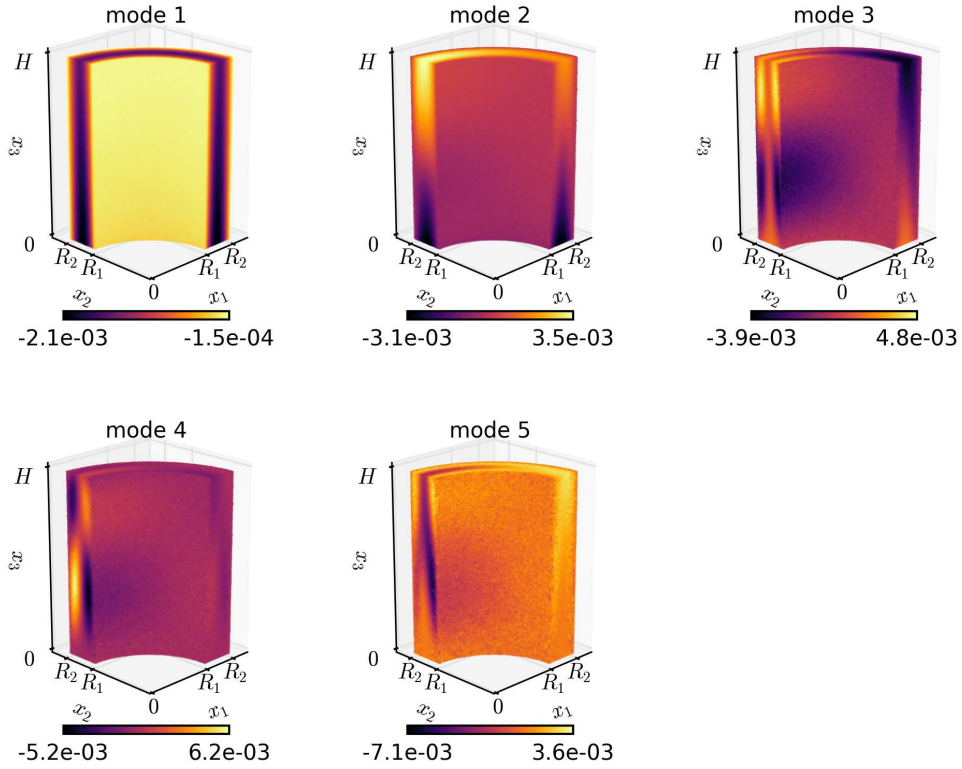


Figure 17: HF POD modes of the temperature field on Ω_{ref}; 20 HF snapshots.

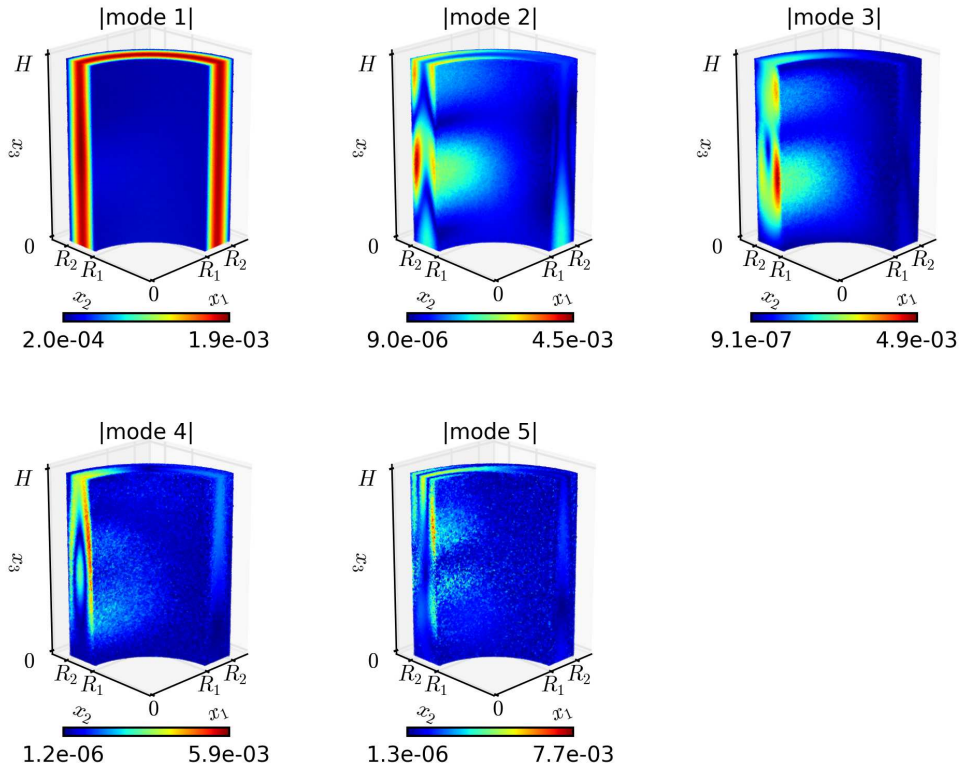


Figure 18: HF POD modes magnitude fields of the velocity on Ω_{ref}; 20 HF snapshots.

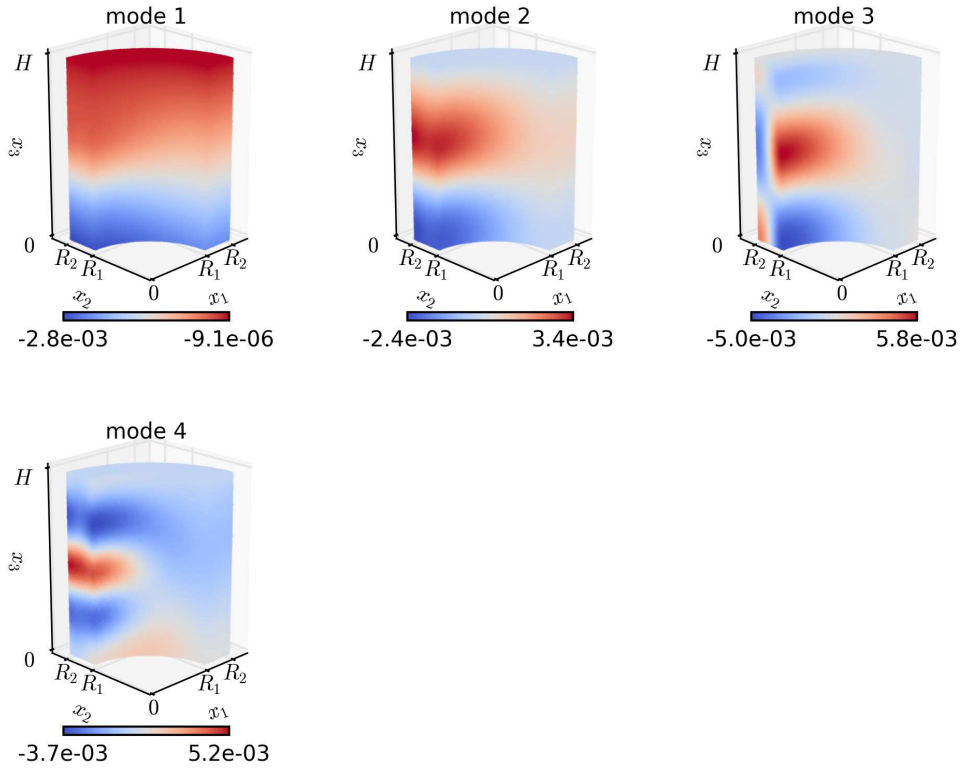


Figure 19: HF POD modes of the pressure field on Ω_{ref}; 20 HF snapshots.

References

[1] E. Dowell. Reduced-order modeling: a personal journey. *Nonlinear Dynamics*, 111:9699–9720, 2023.

[2] S. L. Brunton and J. N. Kutz. *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. Cambridge University Press, 2019.

[3] K. Angshu, Z. Tanisha, A. Tasnim, M. H. U. Khan, M. H. O. R. Molla, M. I. H. Udoy, M. Chowdhury, and A. R. Choudhuri. Review on reduced order modeling and its application in the digital twinning industry. In *AIAA SCITECH 2023 Forum*, page 1206, 2023.

[4] G. Padula, M. Girfoglio, and G. Rozza. A brief review of reduced order models using intrusive and non-intrusive techniques, 2024.

[5] B. Fornberg and N. Flyer. *A Primer on Radial Basis Functions with Applications to the Geosciences*. Society for Industrial and Applied Mathematics, 2015.

[6] R. Zamolo and L. Parussini. Analysis of geometric uncertainties in CFD problems solved by RBF-FD meshless method. *J. Comput. Phys.*, 421:109730, 2020.

[7] M. Jančič, J. Slak, and G. Kosec. Monomial Augmentation Guidelines for RBF-FD from Accuracy Versus Computational Time Perspective. *J. Sci. Comput.*, 87(9), 2021.

[8] A. Bertram, C. Othmer, and R. Zimmermann. Towards real-time vehicle aerodynamic design via multi-fidelity data-driven reduced order modeling. In *2018 AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, page 0916, 2018.

[9] X. Wang, J. Kou, and W. Zhang. Multi-fidelity surrogate reduced-order modeling of steady flow estimation. *Int. J. Numer. Methods Fluids*, 92(12):1826–1844, 2020.

[10] B. Malouin, J.-Y. Trépanier, and M. Gariépy. Interpolation of transonic flows using a proper orthogonal decomposition method. *International Journal of Aerospace Engineering*, 2013:928904, 2013.

[11] D. Toal. On the potential of a multi-fidelity g-pod based approach for optimization and uncertainty quantification. volume Volume 2B: Turbomachinery of *Turbo Expo: Power for Land, Sea, and Air*, 2014.

[12] T. Benamara, P. Breitkopf, I. Lepot, C. Sainvitu, and P. Villon. Multi-fidelity pod surrogate-assisted optimization: Concept and aero-design study. *Structural and Multidisciplinary Optimization*, 56(6):1387–1412, 2017.

[13] M. Kast, M. Guo, and J. S. Hesthaven. A non-intrusive multifidelity method for the reduced order modeling of nonlinear problems. *Computer Methods in Applied Mechanics and Engineering*, 364:112947, 2020.
[14] J. Yu and J. S. Hesthaven. Flowfield reconstruction method using artificial neural network. *AIAA Journal*, 57(2):482–498, 2019.
[15] H. Kang, Z. Tian, G. Chen, L. Li, and T. Chu. Investigation on the nonintrusive multi-fidelity reduced-order modeling for PWR rod bundles. *Nuclear Engineering and Technology*, 54(5):1825–1834, 2022.
[16] C. Perron, D. Sarojini, D. Rajaram, J. Corman, and D. Mavris. Manifold alignment-based multi-fidelity reduced-order modeling applied to structural analysis. *Struct. Multidiscip. Optim.*, 65(8), 2022.
[17] X. Meng, H. Babaee, and G.E. Karniadakis. Multi-fidelity bayesian neural networks: Algorithms and applications. *J. Comput. Phys.*, 438, 2019.
[18] L. Partin, G. Geraci, A. A. Rushdi, M. S. Eldred, and D. E. Schiavazzi. Multifidelity data fusion in convolutional encoder/decoder networks. *Journal of Computational Physics*, 472:111666, 2023.
[19] S. Mariani M. Torzoni, A. Manzoni. A deep neural network, multi-fidelity surrogate model approach for Bayesian model updating in SHM. In *European Workshop on Structural Health Monitoring*. Springer, 2023.
[20] D. H. Song and D. M. Tartakovsky. Transfer learning on multifidelity data. *Journal of Machine Learning for Modeling and Computing*, 3(1):31–47, 2022.
[21] S. E. Ahmed and P. Stinis. A multifidelity deep operator network approach to closure for multiscale systems, 2023.
[22] R. Zamolo, D. Miotti, and E. Nobile. Analysis of geometric uncertainties in 3D thermo-fluid problems solved by RBF-FD meshless method. *J. Phys. Conf. Ser.*, 2685(1):012006, 2024.
[23] R. Zamolo and E. Nobile. Two algorithms for fast 2D node generation: Application to RBF meshless discretization of diffusion problems and image halftoning. *Comput. Math. Appl.*, 75(12):4305–4321, 2018.
[24] R. Zamolo and E. Nobile. Node generation in complex 3D domains for heat conduction problems solved by RBF-FD meshless method. *J. Phys. Conf. Ser.*, 2116(1):012020, 2021.
[25] D. Stevens, H. Power, M. Lees, and H. Morvan. The use of PDE centres in the local RBF Hermitian method for 3D convective-diffusion problems. *J. Comput. Phys.*, 228(12):4606–4624, 2009.
[26] G. E. Fasshauer. *Meshfree Approximation Methods with Matlab*, volume Interdisciplinary Mathematical Sciences: Volume 6. World Scientific, 2007.
[27] B. Šarler and R. Vertnik. Meshfree explicit local radial basis function collocation method for diffusion problems. *Comput. Math. Appl.*, 51(8):1269–1282, 2006.
[28] E. Divo and A. J. Kassab. An Efficient Localized Radial Basis Function Meshless Method for Fluid Flow and Conjugate Heat Transfer. *J. Heat Transf.*, 129(2):124–136, 2007.
[29] R. Zamolo and E. Nobile. Solution of incompressible fluid flow problems with heat transfer by means of an efficient RBF-FD meshless approach. *Numer. Heat Tr. B-Fund.*, 75(1):19–42, 2019.
[30] G. Kosec and J. Slak. Radial basis function-generated finite differences solution of natural convection problem in 3D. *AIP Conf. Proc.*, 2293(1):420094, 2020.
[31] A. J. Chorin. Numerical Solution of the Navier-Stokes Equations. *Math. Comput.*, 22(104):745–762, 1968.
[32] J.S. Hesthaven and S. Ubbiali. Non-intrusive reduced order modeling of nonlinear problems using neural networks. *Journal of Computational Physics*, 363:55–78, 2018.
[33] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 11 2005.
[34] P. Perdikaris, M. Raissi, A. Damianou, N. D. Lawrence, and G. E. Karniadakis. Nonlinear information fusion algorithms for data-efficient multi-fidelity modelling. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 473(2198):20160751, 2017.
[35] M.C. Kennedy and A. O'Hagan. Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, 87(1):1–13, 2000.
[36] L. L. Gratiet and J. Garnier. Recursive co-kriging model for design of computer experiments with multiple levels of fidelity. *Int J Uncertain Quantif*, 4(5):365–386, 2014.
[37] T. Hastie, R. Tibshirani, and J.H. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer series in statistics. Springer, 2009.