# Hierarchical Model Reduction Driven by Machine Learning for Parametric Advection-Diffusion-Reaction Problems in the Presence of Noisy Data

**Massimiliano Lupo Pasini**[1] · **Simona Perotto**[2]

## Abstract

We propose a new approach to generate a reliable reduced model for a parametric elliptic problem, in the presence of noisy data. The reference model reduction procedure is the directional HiPOD method, which combines Hierarchical Model reduction with a standard Proper Orthogonal Decomposition, according to an offline/online paradigm. In this paper we show that directional HiPOD looses in terms of accuracy when problem data are affected by noise. This is due to the interpolation driving the online phase, since it replicates, by definition, the noise trend. To overcome this limit, we replace interpolation with Machine Learning fitting models which better discriminate relevant physical features in the data from irrelevant unstructured noise. The numerical assessment, although preliminary, confirms the potentialities of the new approach.

**Keywords** Machine learning · Reduced order models · Partial differential equations · Numerical analysis · Numerical linear algebra

✉ Massimiliano Lupo Pasini
  lupopasinim@ornl.gov

  Simona Perotto
  simona.perotto@polimi.it

1   Computational Sciences and Engineering Division, Oak Ridge National Laboratory, 1 Bethel Valley Road, Oak Ridge, TN 37831, USA

2   MOX – Dipartimento di Matematica Politecnico di Milano, Piazza L. da Vinci, 32, I-20133 Milano, Italy

## 1 Introduction and Motivations

Numerical methods for solving parametric partial differential equations are relevant for all the engineering applications that can be framed into a multi-query or a real-time context. Recurrent instances range from the estimation of the parameters which govern physical phenomena (e.g., in biomedical engineering, where the known velocity of the blood and the Navier-Stokes equations for incompressible fluids are combined to draw conclusions about the blood viscosity inside a cardiovascular vessel) to the solution of optimal control problems (e.g., in environmental engineering, where experimental measurements of the concentration of a pollutant in the water and the Navier-Stokes equations for incompressible fluids with a parametrized forcing term are combined to compute the maximum amount of pollutant that can be released without compromising the ecosystem inside a river).

The curse of dimensionality characterizing full order models to simulate this type of physical systems raised the necessity to propose specific numerical methods in order to sustain the computational cost. Reduced order models have garnered interest in the scientific computing community as an effective way to compress complex partial differential models in a projection subspace, and thus make it more computationally convenient to solve [2, 4, 6, 14, 17, 35].

Here, we focus on a class of reduced order models conceived to describe flows in pipes or, more in general, phenomena with a privileged dynamics aligned with the centerline of the pipe, which may be locally modified by secondary dynamics evolving along the transverse sections. We are referring to the Hierarchical Model (HiMod) reduction [11, 27, 29, 30, 33, 36], and, in particular, to a parametric counterpart of such an approach, known as HiPOD [3, 21]. HiPOD offers a possible remedy to the well-known bottleneck of an offline/online paradigm, i.e., the computational burden characterizing the offline phase. The idea is to replace the "truth" model in the offline phase of the Proper Orthogonal Decomposition (POD) [18–20, 38] with a HiMod discretization, namely, more in general, with a reduced-order model characterized by a high accuracy and a contained computational demand. Then, the online phase recovers the HiMod approximation for a not yet sampled value of the selected parameter, after solving a problem of a very small dimensionality. The computational advantages led by a HiPOD approximation have been numerically investigated both on scalar and vector problems [3, 21].

In [21] two HiPOD model reduction procedures are presented, namely the basic and the directional HiPOD approaches. Here, we focus on the directional method, that combines HiMod and POD by exploiting the decoupling between leading and secondary dynamics at the basis of a HiMod formulation. Such an approach proved to outperform the basic one when dealing with phenomena which are of interest in this paper, i.e., characterized by a significant horizontal dynamics [21].

In this work, we show that the presence of noise in the data can make directional HiPOD unstable, due to an interpolation over the parametric space to estimate the online solution. Actually, interpolation reproduces, instead of filtering out, the noise affecting data. To overcome this issue, we propose to replace interpolation with Machine Learning (ML) fitting models which better discriminate relevant physical features in the data from irrelevant unstructured noise, thus allowing the model to retain and leverage the former and discard the latter. The numerical assessment we carried out shows that the ML methodology reduces the $L^2$- and the $H^1$-norm of the HiPOD relative error up to an order of magnitude with respect to state-of-the-art interpolation techniques, thus confirming the strong potentiality of the proposed approach.

The paper is structured as follows. Section 2 addresses the mathematical background exploited to upgrade the directional HiPOD procedure in [21], namely the basics to perform a HiPOD model reduction and some machine learning regression models. Section 3 introduces the new HiPOD approach and analyzes the effect of the noise propagation onto the response matrix. In Sect. 4, we carry out a numerical assessment of the proposed methodology, by comparing the state-of-the-art HiPOD approach with the new ML version, both in terms of accuracy and robustness to the level of noise. Finally, some conclusions and perspective are supplied in the last section.

## 2 Mathematical Background

HiPOD model reduction and Machine Learning (ML) models for a regression analysis represent the main methodological tools supporting the new approach proposed in this paper. The next sections are devoted to introduce such tools, for the reader completeness.

In particular, we choose as a reference problem a parametrized elliptic Partial Differential Equation (PDE), defined on a domain $\Omega \subset \mathbb{R}^2$, whose weak form is

$$\text{find } u(\alpha) \in V \quad \text{s.t.} \quad a(u(\alpha), v; \alpha) = f(v; \alpha) \quad \forall v \in V, \tag{1}$$

where $\alpha \in \mathcal{P} \subset \mathbb{R}^p$ is the selected parameter varying in the set $\mathcal{P}$ of the admissible values; $a(\cdot, \cdot; \alpha) : V \times V \times \mathcal{P} \to \mathbb{R}$ and $f(\cdot; \alpha) : V \times \mathcal{P} \to \mathbb{R}$ denote the parametrized bilinear and linear forms characterizing the differential problem at hand, the linearity being meant with respect to the variables different from $\alpha$, and with $V \subseteq H^1(\Omega)$ a suitable Hilbert space depending on the specific PDE problem as well as on the assigned boundary conditions, standard notation being adopted for function spaces [10]. To simplify the exposition, we focus here on the case of a standard scalar linear advection-diffusion-reaction (ADR) problem, completed with full homogeneous Dirichlet boundary conditions (see Sect. 4 for a more general setting), so that the bilinear and the linear forms in (1) are

$$a(w, z; \alpha) = \int_\Omega \mu \nabla w \cdot \nabla z \, d\Omega + \int_\Omega (\mathbf{b} \cdot \nabla w + \sigma w) z \, d\Omega, \quad f(z; \alpha) = \int_\Omega f z \, d\Omega, \tag{2}$$

with $w, z \in V = H_0^1(\Omega)$, and where parameter $\alpha$ coincides with some of the problem data, i.e., the viscosity $\mu$, the advective field $\mathbf{b} = [b_1, b_2]^T$, the reaction $\sigma$, the source term $f$ (or a boundary value when non homogeneous or more general boundary conditions are assigned).

Suitable assumptions are advanced on the problem data in order to guarantee the well-posedness of formulation (1), for any $\alpha \in \mathcal{P}$. Finally, we conjecture an affine parameter dependence [17, 35].

### 2.1 The HiPOD Approach

The HiPOD method provides the parametric counterpart of a Hierarchically Model (HiMod) reduction, by properly combining a HiMod discretization with the standard Proper Orthogonal Decomposition (POD) [18–20, 38]. In the next section we recap the main features characterizing a HiMod reduction, being instrumental in the setting of HiPOD procedures.

### 2.1.1 Hierachical Model Reduction

HiMod reduction proved to be an ideal method for the modeling of scenarios where a dominant direction is evident in the global dynamics of the considered phenomenon. This modeling property is, in general, mirrored by the geometric design of the computational domain, which is assumed to coincide with a pipe where the centerline is parallel to the dominant dynamics. Thus, $\Omega$, is identified by the fiber bundle $\bigcup_{x \in \Omega_{1D}} \{x\} \times \Sigma_x$, where $\Omega_{1D}$ is the one-dimensional (1D) supporting fiber paraller to the leading dynamics, while $\Sigma_x$ denotes the transverse section at the generic point $x$ along $\Omega_{1D}$ [11, 29, 30, 33]. For simplicity, we focus on rectilinear domains, so that $\Omega_{1D} \equiv (a, b) \subset \mathbb{R}$ (we refer the interested reader to [7, 28, 31] for pipes with a bent centerline).

HiMod reduction exploits the geometric requirement on $\Omega$ to discretize a PDE problem, so that the leading and the transverse dynamics are approximated with different methods, according to a separation of variable criterion. In this work, we adopt the approach proposed in [1] where finite elements model the main dynamics, while a basis of customized modal functions describes the dynamics parallel to the transverse sections of the pipe. This decoupling of leading and secondary dynamics allows us to commute the full problem into a system of coupled 1D problems, independently of the original model dimensionality, with a consequent considerable benefit in terms of computational effort [1, 7, 16, 23].

Additionally, as it is recurrent in several well-known contexts, computations are performed in a reference domain $\widehat{\Omega}$ (where, e.g., constants can be explicitly computed) and successively moved to the physical domain $\Omega$. This is performed by means of a map $\Psi : \Omega \to \widehat{\Omega}$ which is assumed to be differentiable with respect to both the independent variables $x$ and $y$. The domain $\widehat{\Omega}$ shares a fiber structure as $\Omega$, being $\widehat{\Omega} = \Omega_{1D} \times \widehat{\Sigma}$, with $\widehat{\Sigma}$ the reference transverse fiber and where the supporting fiber is the same as for $\Omega$. In particular, for any point $\mathbf{z} = (x, \mathbf{y}) \in \Omega$, there exists a point $\widehat{\mathbf{z}} = (\widehat{x}, \widehat{\mathbf{y}}) \in \widehat{\Omega}$, such that $\widehat{\mathbf{z}} = \Psi(\mathbf{z})$, with $\widehat{x} \equiv x$ and $\widehat{\mathbf{y}} = \psi_x(\mathbf{y})$, $\psi_x : \Sigma_x \to \widehat{\Sigma}$ being the map between the generic and the reference transverse fiber. Hereafter, we assume $\psi_x$ to be a $C^1$-diffeomorphism, for all $x \in \Omega_{1D}$ (more details about maps $\Psi$ and $\psi_x$ are available in [30]).

The separation of variables combined with the mapping to $\widehat{\Omega}$ leads us to define the parametric HiMod reduced space

$$V_m(\alpha) = \left\{ v_m(x, \mathbf{y}; \alpha) = \sum_{k=1}^{m} \sum_{j=1}^{N_h} \tilde{v}_{k,j}^{\alpha} \vartheta_j(x) \varphi_k(\psi_x(\mathbf{y})), \ (x, \mathbf{y}; \alpha) \in \Omega_{1D} \times \Sigma_x \times \mathcal{P} \right\},$$
(3)

where $m \in \mathbb{N}^+$ is the modal index setting the level of detail of the HiMod approximation in the hierarchy; $\mathcal{B}_1 = \{\vartheta_j\}_{j=1}^{N_h}$ is a basis for the 1D space, $V_{1D} \subset H_0^1(\Omega_{1D})$, of the finite element functions associated with the supporting fiber $\Omega_{1D}$ and vanishing at $a$ and $b$, with $\dim(V_{1D}) = N_h < +\infty$; $\mathcal{B}_2 = \{\varphi_k\}_{k \in \mathbb{N}^+}$ denotes the basis of modal functions defined on the reference transverse fiber $\widehat{\Sigma}$, orthonormal with respect to the $L^2(\widehat{\Sigma})$-scalar product and including, in an essential way [1], the data assigned on the lateral boundary $\Gamma_L = \cup_{x \in \Omega_{1D}} \partial \Sigma_x$ of $\Omega$ (the reader interested to different possible choices both for basis $\mathcal{B}_1$ and $\mathcal{B}_2$ may refer, e.g., to [7, 11, 16, 30, 31]). In particular, function $\tilde{v}_k(x; \alpha) = \sum_{j=1}^{N_h} \tilde{v}_{k,j}^{\alpha} \vartheta_j(x) \in V_{1D}$ identifies the frequency coefficient associated with the $k$-th modal function $\varphi_k$.

As far as the modal index $m$ is concerned, it may be assigned thanks to a trial-and-error procedure (see, e.g., [11, 30]) or starting from some preliminary (geometric or physic) information about the problem at hand (as, e.g., in [16]) or via an automatic selection based

on an a posteriori modeling error analysis (we refer, e.g., to [32, 34]). We adopt the same value for $m$ in the whole $\Omega$, although the modal index may be locally varied along the domain to match possible heterogeneities of the solution (see [29, 30, 32–34] for further details).

Thus, the HiMod approximation to problem (1), associated with the modal coefficient $m$, reads

$$\text{find } u_m(\alpha) \in V_m(\alpha) \quad \text{s.t.} \quad a(u_m(\alpha), v_m; \alpha) = f(v_m; \alpha) \quad \forall v_m \in V_m(\alpha), \tag{4}$$

with $u_m(\alpha) = u_m(x, \mathbf{y}; \alpha)$. To ensure the well-posedness of formulation (4), we endow the space $V_m(\alpha)$ both with a conformity and a spectral approximability assumption, while the convergence of the HiMod approximation $u_m(\alpha)$ to the full solution $u(\alpha)$ in (1) is guaranteed by introducing a standard density assumption on the discrete space $V_{1D}$.

After applying the HiMod representation in (3) to $u_m(\alpha)$ in (4), and choosing the test function $v_m$ as $\vartheta_t \varphi_q$, for $t = 1, \ldots, N_h$ and $q = 1, \ldots, m$, the HiMod formulation turns into the system of $m$ coupled 1D problems,

$$A_m(\alpha)\mathbf{u}_m(\alpha) = \mathbf{f}_m(\alpha), \tag{5}$$

with $A_m(\alpha) \in \mathbb{R}^{mN_h \times mN_h}$ denoting the HiMod stiffness matrix, $\mathbf{f}_m(\alpha) \in \mathbb{R}^{mN_h}$ representing the HiMod right-hand side, and where vector

$$\mathbf{u}_m(\alpha) = \left[ \tilde{u}_{1,1}^{\alpha}, \ldots, \tilde{u}_{1,N_h}^{\alpha}, \tilde{u}_{2,1}^{\alpha}, \ldots, \tilde{u}_{2,N_h}^{\alpha}, \ldots, \tilde{u}_{m,1}^{\alpha}, \ldots, \tilde{u}_{m,N_h}^{\alpha} \right]^T \in \mathbb{R}^{mN_h} \tag{6}$$

collects the modal coefficients $\{\tilde{u}_{k,j}^{\alpha}\}_{k=1,j=1}^{m,N_h}$, namely the actual unknowns of the HiMod discretization

$$u_m(x, \mathbf{y}; \alpha) = \sum_{k=1}^{m} \sum_{j=1}^{N_h} \tilde{u}_{k,j}^{\alpha} \vartheta_j(x) \varphi_k(\psi_x(\mathbf{y})) \tag{7}$$

[11, 30]. It has been numerically checked that, when the mainstream dominates the transverse dynamics (i.e., for small values of $m$), the HiMod approximation demands a considerably lower computational effort if compared with a standard (e.g., finite element) discretization of problem (1), without giving up the accuracy of the simulation [1, 7, 16, 23].

### 2.1.2 Directional HiPOD Reduction

A HiPOD approach offers a new way to yield a HiMod approximation, as an alternative to the resolution of the HiMod system (5). According to a data-driven procedure, the idea is to replace the HiMod discretization in (7) with a surrogate solution obtained by resorting to a POD reduced basis generated by HiMod approximations. A standard offline/online paradigm drives the computation of the HiPOD approximation. In particular, during the offline phase, we compute the HiMod solution in (5) for different choices of $\alpha$, to extract the POD basis. In the online step, such a basis is employed to approximate the HiMod solution in (5) associated with a value of the parameter not sampled during the offline phase.

As shown in [3, 21], a HiPOD procedure considerably lowers computational costs, without compromising the quality of the reduced solution.

Two HiPOD approaches have been explored so far. The basic method coincides with a straightforward application of a projection-based POD to HiMod solutions [3]. An advanced procedure, referred to as directional HiPOD, takes advantage of the separation of variables implied by a HiMod discretization [21], the SVD being used to remove the redundancy along the main stream and the transverse direction, separately. In addition, the online phase

is carried out by interpolation instead of projection, thus relieving us from assembling the HiMod stiffness matrix and the right-hand side associated with the online parameter, as expected by the basic HiPOD approach.

Below, we detail the directional HiPOD method, since instrumental to the new approach proposed in Sect. 3.

The goal of the offline phase is to identify the POD basis to be used in an online mode. To this aim, we build the response matrix by collecting the HiMod solution to problem (1) for $p$ different values, $\alpha_i$, of the parameter $\alpha$, with $i = 1, \ldots, p$. In more detail, the generic HiMod solution $u_m(x, \mathbf{y}; \alpha_i)$ is identified with the corresponding modal coefficients, $\{\tilde{u}_{k,j}^{\alpha_i}\}_{k=1, j=1}^{m, N_h}$, collected by mode into the $m$ vectors

$$\mathbf{U}^k(\alpha_i) = \left[\tilde{u}_{k,1}^{\alpha_i}, \tilde{u}_{k,2}^{\alpha_i}, \ldots, \tilde{u}_{k,N_h}^{\alpha_i}\right]^T \in \mathbb{R}^{N_h} \quad k = 1, \ldots, m, \tag{8}$$

so that the response matrix is assembled as

$$U = \left[\mathbf{U}^1(\alpha_1) \cdots \mathbf{U}^m(\alpha_1) | \mathbf{U}^1(\alpha_2) \cdots \mathbf{U}^m(\alpha_2) | \cdots\cdots\cdots | \mathbf{U}^1(\alpha_p) \cdots \mathbf{U}^m(\alpha_p)\right]$$

$$= \begin{bmatrix} \tilde{u}_{1,1}^{\alpha_1} & \cdots & \tilde{u}_{m,1}^{\alpha_1} & \tilde{u}_{1,1}^{\alpha_2} & \cdots & \tilde{u}_{m,1}^{\alpha_2} & \cdots\cdots & \tilde{u}_{1,1}^{\alpha_p} & \cdots & \tilde{u}_{m,1}^{\alpha_p} \\ \tilde{u}_{1,2}^{\alpha_1} & \cdots & \tilde{u}_{m,2}^{\alpha_1} & \tilde{u}_{1,2}^{\alpha_2} & \cdots & \tilde{u}_{m,2}^{\alpha_2} & \cdots\cdots & \tilde{u}_{1,2}^{\alpha_p} & \cdots & \tilde{u}_{m,2}^{\alpha_p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \tilde{u}_{1,N_h}^{\alpha_1} & \cdots & \tilde{u}_{m,N_h}^{\alpha_1} & \tilde{u}_{1,N_h}^{\alpha_2} & \cdots & \tilde{u}_{m,N_h}^{\alpha_2} & \cdots\cdots & \tilde{u}_{1,N_h}^{\alpha_p} & \cdots & \tilde{u}_{m,N_h}^{\alpha_p} \end{bmatrix}. \tag{9}$$

The blocks are associated with the different parameters, while, for each block, columns run over modes, rows over the finite element nodes. In order to extract the POD basis, we apply the Singular Value Decomposition (SVD) [15] to matrix $U$, so that

$$U = \varXi \varLambda K^T, \tag{10}$$

with $\varXi \in \mathbb{R}^{N_h \times N_h}$ and $K \in \mathbb{R}^{(mp) \times (mp)}$ unitary matrices, and $\varLambda \in \mathbb{R}^{N_h \times (mp)}$ a pseudo-diagonal matrix. Each column in $U$ can be expanded in terms of the left singular vectors $\{\boldsymbol{\xi}_j\}_{j=1}^{N_h}$ of $U$ since constituting an orthogonal basis for $\mathbb{R}^{N_h}$, so that

$$\mathbf{U}^k(\alpha_i) = \sum_{j=1}^{N_h} T_j^k(\alpha_i) \boldsymbol{\xi}_j \quad k = 1, \ldots, m, \; i = 1, \ldots, p. \tag{11}$$

It is customary to select the first, say $L$ with $L \leq N_h$, most meaningful singular vectors of $U$ to identify the reduced POD space $V_{\text{POD},1}^L = \text{span}\{\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_L\}$, with $\dim(V_{\text{POD},1}^L) = L$. As a consequence, the vectors in (11) are approximated as

$$\mathbf{U}^k(\alpha_i) \cong \sum_{j=1}^{L} T_j^k(\alpha_i) \boldsymbol{\xi}_j \quad k = 1, \ldots, m, \; i = 1, \ldots, p, \tag{12}$$

the equality being ensured for $L = N_h$. The directional HiPOD reduction involves a reorganization of coefficients $\{T_j^k(\alpha_i)\}$ first into the vectors $\mathbf{T}_j(\alpha_i) = [T_j^1(\alpha_i), \ldots, T_j^m(\alpha_i)]^T \in \mathbb{R}^m$ with $i = 1, \ldots, p$, and successively into the matrices

$$S_j = \left[\mathbf{T}_j(\alpha_1), \ldots, \mathbf{T}_j(\alpha_p)\right] = \begin{bmatrix} T_j^1(\alpha_1) & \ldots & T_j^1(\alpha_p) \\ \vdots & & \vdots \\ T_j^m(\alpha_1) & \ldots & T_j^m(\alpha_p) \end{bmatrix} \in \mathbb{R}^{m \times p},$$

with $j = 1, \ldots, L$, in order to associate a reduced POD basis with each index $j$. To this aim, we factorize the matrices $S_j$ via SVD, so that

$$S_j = R_j D_j P_j^T, \tag{13}$$

with $R_j \in \mathbb{R}^{m \times m}$ and $P_j \in \mathbb{R}^{p \times p}$ unitary matrices, and $D_j \in \mathbb{R}^{m \times p}$ the pseudo-diagonal matrix collecting the singular values of $S_j$. It follows that each column $\mathbf{T}_j(\alpha_i)$ of $S_j$ can be approximated by resorting to the POD orthogonal basis $\{\mathbf{r}_j^k\}_{k=1}^{\mu_j}$, with $\mu_j \leq m$, constituted by the most significant $\mu_j$ left singular vectors of $S_j$, namely

$$\mathbf{T}_j(\alpha_i) \cong \sum_{k=1}^{\mu_j} Q_j^k(\alpha_i)\mathbf{r}_j^k \quad j = 1, \ldots, L, \ i = 1, \ldots, p. \tag{14}$$

Thus, a POD space $V_{\text{POD},2,j}^{\mu_j} = \text{span}\{\mathbf{r}_j^1, \ldots, \mathbf{r}_j^{\mu_j}\}$ can be defined for each $j$, with $\dim(V_{\text{POD},2,j}^{\mu_j}) = \mu_j$.

The offline phase ends with the overall generation of $(L+1)$ POD reduced bases to be used in the online phase in order to predict the HiMod approximation to problem (1) for a value, $\alpha^*$, of the parameter, such that $\alpha^* \neq \alpha_i$ for $i = 1, \ldots, p$. The idea is to go backward through the directional procedure, starting from the coefficients $Q_j^k(\alpha^*)$ in (14), with $j = 1, \ldots, L$, $k = 1, \ldots, \mu_j$, which are approximated by interpolating the (known) values $Q_j^k(\alpha_i)$ for $i = 1, \ldots, p$. Coefficients $Q_j^k(\alpha^*)$, together with the $L$ POD bases $\{\mathbf{r}_j^k\}_{k=1}^{\mu_j}$, allow us to compute the vectors

$$\mathbf{T}_j(\alpha^*) = \left[T_j^1(\alpha^*), \ldots, T_j^m(\alpha^*)\right]^T = \sum_{k=1}^{\mu_j} Q_j^k(\alpha^*)\mathbf{r}_j^k \quad j = 1, \ldots, L \tag{15}$$

in $\mathbb{R}^m$ and, consequently, by exploiting the POD basis, $\{\boldsymbol{\xi}_j\}_{j=1}^L$, generated first, to assemble the $m$ vectors

$$\mathbf{U}_{\text{HiPOD}}^k(\alpha^*) = \left[u_{\text{POD},k,1}^{\alpha^*}, \ldots, u_{\text{POD},k,N_h}^{\alpha^*}\right]^T = \sum_{j=1}^L T_j^k(\alpha^*)\boldsymbol{\xi}_j \quad k = 1, \ldots, m \tag{16}$$

in $\mathbb{R}^{N_h}$, which represent the online counterpart of vectors in (12). The HiMod solution $u_m(\alpha^*)$ can thus be approximated by means of the expansion

$$u_{\text{HiPOD}}^{L,M_L}(\alpha^*) = \sum_{k=1}^m \left[\sum_{j=1}^{N_h} u_{\text{POD},k,j}^{\alpha^*} \vartheta_j(x)\right] \varphi_k(\psi_x(\mathbf{y})), \tag{17}$$

with $M_L = \{\mu_j\}_{j=1}^L$. In particular, coefficients $u_{\text{POD},k,j}^{\alpha^*}$ provide an approximation of the actual coefficient $\tilde{u}_{k,j}^{\alpha^*}$ in (8) for $\alpha_i = \alpha^*$.

The procedure adopted to predict the coefficients $Q_j^k(\alpha^*)$ plays an important role. When the offline solution data is not affected by noise, interpolation techniques are an effective tool. In [21], we assess different intepolations, namely, standard linear interpolation, a piecewise cubic Hermite (PCH) interpolant, and interpolating Radial Basis Functions (RBF), to infer that PCH and RBF interpolants slightly outperform the linear approach.

Vice versa, when the offline data is affected by noise, interpolation is not recommended because it does not discriminate between relevant features of the problem and noise, thus not being able to retain the former and discard the latter. This limit motivated us in the proposal

of a new tool, in order to guarantee a reliable HiMod approximation also in the presence of noisy data.

Finally, to select the dimension of the POD spaces $V_{\mathrm{POD},1}^L$ and $V_{\mathrm{POD},2,j}^{\mu_j}$, we resort to a control on the variance, i.e., after setting the tolerances $\epsilon_1$ and $\epsilon_2$, with $0 \leq \epsilon_1, \epsilon_2 \leq 1$, we keep the first $L$ left singular vectors $\boldsymbol{\xi}_j$ of $U$ and the first $\mu_j$ left singular vectors $\mathbf{r}_j^k$ of $S_j$, such that

$$\frac{\sum_{j=1}^{L} \lambda_j^2}{\sum_{j=1}^{N_h} \lambda_j^2} \geq \epsilon_1, \quad \frac{\sum_{j=1}^{\mu_j} d_{j,k}^2}{\sum_{j=1}^{m} d_{j,k}^2} \geq \epsilon_2, \tag{18}$$

respectively, with $\lambda_j$ the singular value of $U$ associated with $\boldsymbol{\xi}_j$, for $j = 1, \ldots, N_h$, and $d_{j,k}$ the singular value of $S_j$ corresponding to the $k$-th singular vector $\mathbf{r}_j^k$, with $k = 1, \ldots, m$.

## 2.2 Machine Learning Models for Regression

In this section we focus on Machine Learning (ML) models to approximate data distributions.

The final goal is to replace the interpolation step in the online phase of the directional HiPOD procedure with a regression technique, in order to address situations where the data in the offline phase may be noisy. In particular, to estimate the coefficients $Q_j^k(\alpha^*)$ in (15), we resort to a ML fitting model since the offline solution is related to parameter $\alpha_i$ by a highly nonlinear relation.

Now, in order to understand how the noise in the data may affect the accuracy of the predictions yielded by a ML fitting model, we have to make some preliminary assumption on the noise properties. The most common hypothesis leads us to consider an additive independent identically distributed (i.i.d.) Gaussian noise, $\tilde{\eta} \sim \mathcal{N}(0, \eta)$, in the output, with $\eta > 0$, namely, we assume to have

$$Q_j^k(\alpha_i) + \tilde{\eta} \quad k = 1, \ldots, m, \ j = 1, \ldots, L, \ i = 1, \ldots, p. \tag{19}$$

In the next sections, we consider two ML regression models that operate under the additive noise assumption, i.e., the polynomial [12, 13, 22, 37] and the Gaussian process regression [8, 9, 24, 25].

### 2.2.1 Polynomial Regression

Polynomial regression is a form of regression analysis where the relationship between the independent and the dependent variables is modeled as a polynomial in the independent variable, of a certain degree $n$. With reference to our specific context, a polynomial regression model of degree $n$ estimates the nonlinear dependence of $Q_j^k(\alpha_i)$ from the parameter $\alpha_i$ according to the formula

$$Q_j^k(\alpha_i) = \sum_{\ell=0}^{n} \beta_{j,\ell}^k \alpha_i^\ell + \tilde{\eta}, \quad k = 1, \ldots, m, \ j = 1, \ldots, L, \ i = 1, \ldots, p, \tag{20}$$

where $\beta_{j,\ell}^k$ are unknown parameters to be computed in order to optimize the matching between predictions and observations of the dependent variable, while $\tilde{\eta}$ denotes a zero-mean Gaussian

noise as in (19). Although relation (20) is nonlinear in the independent variable $\alpha_i$, this model is categorized as linear since the regression function is linear in terms of the unknown parameters $\beta_{j,\ell}^k$.

With a view to the directional HiPOD procedure, the quantity

$$\sum_{\ell=0}^{n} \beta_{j,\ell}^k [\alpha^*]^\ell \quad k = 1, \ldots, m, \; j = 1, \ldots, L \tag{21}$$

will be used as an estimate for the online coefficient $Q_j^k(\alpha^*)$.

### 2.2.2 Gaussian Process Regression

The goal of a Gaussian process regression is to determine the best set of random variables that describes the relation between input features and outputs. In the specific setting of interest, we denote the random variables we are looking for by $\mathcal{Q}_j^k$, while $\alpha_i$ and $Q_j^k(\alpha_i)$ represent the inputs and the outputs, respectively.

A Gaussian process is fully characterized by its mean function and covariance function. Therefore, Gaussian process regression reduces to calculating the best values for the mean and the covariance functions in order to minimize the mismatch between outputs ($Q_j^k(\alpha_i)$) and the predictions produced by the finite set of random variables selected from the Gaussian process, evaluated at the available inputs ($\alpha_i$).

Bayesian statistics combines prior Gaussian processes, that retain preliminary knowledge and information from the offline data samples (also known as likelihood), to construct an updated posterior Gaussian process. According to the Bayes' rule [5], a prior Gaussian process is iteratively updated using the information from the data till it converges to a stationary state defined by a posterior Gaussian process.

Now, we specify such a workflow onto our setting of interest. We denote a prior Gaussian process by $\mathcal{Q}_{j,\text{prior}}^k$, so that

$$\mathcal{Q}_{j,\text{prior}}^k \sim \mathcal{GP}\left(q_{j,\text{prior}}^k, z_{j,\text{prior}}^k + \eta\right) \quad k = 1, \ldots, m, \; j = 1, \ldots, L, \tag{22}$$

where $q_{j,\text{prior}}^k = q_{j,\text{prior}}^k(\alpha)$ is the mean function, $z_{j,\text{prior}}^k = z_{j,\text{prior}}^k(\alpha, \tilde{\alpha})$ is the covariance function, and $\eta$ is the noise function. The restriction of the prior Gaussian process in (22) at the input points $\alpha_i$, for $i = 1, \ldots, p$, is a multivariate $p$-dimensional Gaussian distribution $\mathcal{N}(\mathbf{m}_{j,\text{prior}}^k, Z_{j,\text{prior}}^k)$, where the mean vector, $\mathbf{m}_{j,\text{prior}}^k = [m_{j,\text{prior},r}^k] \in \mathbb{R}^p$, and the covariance matrix, $Z_{j,\text{prior}}^k = [Z_{j,\text{prior},rs}^k] \in \mathbb{R}^{p \times p}$, are identified by relations

$$m_{j,\text{prior},r}^k = q_{j,\text{prior}}^k(\alpha_r) = \frac{1}{4}\alpha_r^2 \quad r = 1, \ldots, p, \tag{23}$$

$$Z_{j,\text{prior},rs}^k = z_{j,\text{prior}}^k(\alpha_r, \alpha_s) = \exp\left[-\frac{1}{2}(\alpha_r - \alpha_s)^2\right] \quad r, s = 1, \ldots, p, \tag{24}$$

respectively. The Gaussian process $\mathcal{Q}_{j,\text{like}}^k$ associated with the data, is defined as

$$\mathcal{Q}_{j,\text{like}}^k \sim \mathcal{GP}\left(q_{j,\text{like}}^k, z_{j,\text{like}}^k\right) \quad k = 1, \ldots, m, \; j = 1, \ldots, L, \tag{25}$$

with $q_{j,\text{like}}^k$ denoting the mean function and $z_{j,\text{like}}^k$ the covariance function. The Gaussian process $\mathcal{Q}_{j,\text{like}}^k$ is not fully characterizable using the data, meaning that the mean and the covariance functions cannot be uniquely determined. However, the values attained

by $q_{j,\text{like}}^k$ and $z_{j,\text{like}}^k$ at the input points $\alpha_i$, for $i = 1, \ldots, p$, are known and correspond to a $p$-dimensional Gaussian distribution $\mathcal{N}(\mathbf{m}_{j,\text{like}}^k, Z_{j,\text{like}}^k)$, where the mean vector $\mathbf{m}_{j,\text{like}}^k = [m_{j,\text{like},r}^k] \in \mathbb{R}^p$ and the covariance matrix $Z_{j,\text{like}}^k = [Z_{j,\text{like},rs}^k] \in \mathbb{R}^{p \times p}$ are defined by

$$m_{j,\text{like},r}^k = Q_j^k(\alpha_r) \quad r = 1, \ldots, p,$$

$$Z_{j,\text{like},rs}^k = \exp\left[ -\frac{1}{2} \left( Q_j^k(\alpha_r) - Q_j^k(\alpha_s) \right)^2 \right] \quad r, s = 1, \ldots, p. \tag{26}$$

Now, the posterior can be used to make predictions for unseen values of $\alpha$ (namely, to predict coefficients $Q_j^k(\alpha^*)$ with $\alpha^*$ the online parameter). The joint distribution of the Gaussian processes evaluated at the points $\alpha_i$, for $i = 1, \ldots, p$, is a $2p$-dimensional Gaussian distribution

$$\mathcal{N}\left( \begin{bmatrix} \mathbf{m}_{j,\text{prior}}^k \\ \mathbf{m}_{j,\text{like}}^k \end{bmatrix}, \begin{bmatrix} Z_{j,\text{prior}}^k & Z_{j,\text{pl}}^k \\ Z_{j,\text{pl}}^k & Z_{j,\text{like}}^k \end{bmatrix} \right), \tag{27}$$

where the symmetric matrix $Z_{j,\text{pl}}^k \in \mathbb{R}^{p \times p}$ is the correlation matrix between prior and likelihood. Using Bayes' Theorem, the posterior $p$-dimensional Gaussian distribution is

$$\mathcal{N}\left( \mathbf{m}_{j,\text{prior}}^k - Z_{j,\text{pl}}^k \left[ Z_{j,\text{prior}}^k \right]^{-1} \mathbf{m}_{j,\text{like}}^k, \, Z_{j,\text{prior}}^k - Z_{j,\text{pl}}^k \left[ Z_{j,\text{like}}^k \right]^{-1} Z_{j,\text{pl}}^k \right). \tag{28}$$

The posterior Gaussian process associated with the distribution in (28) is consequently given by

$$\mathcal{Q}_{j,\text{posterior}}^k \sim \mathcal{GP}\left( q_{j,\text{posterior}}^k, z_{j,\text{posterior}}^k \right). \tag{29}$$

The evaluation of the mean function $q_{j,\text{posterior}}^k$ at the new input parameter value $\alpha^*$ is

$$q_{j,\text{posterior}}^k \left( \alpha^* \right) = q_{j,\text{prior}}^k \left( \alpha^* \right) - \left[ Z\left( \boldsymbol{\alpha}, \alpha^* \right) \right]^T \left[ Z_{j,\text{prior}}^k \right]^{-1} \mathbf{m}_{j,\text{like}}^k, \tag{30}$$

with $q_{j,\text{prior}}^k(\alpha^*)$ defined as in (23), $Z(\boldsymbol{\alpha}, \alpha^*) \in \mathbb{R}^p$ denoting the covariance vector between the offline data sampled at $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_p]^T \in \mathbb{R}^p$ and the new sampled parameter $\alpha^*$. The covariance function in (29) evaluated at $\alpha^*$ is

$$z_{j,\text{posterior}}^k \left( \alpha^* \right) = z_{j,\text{prior}}^k \left( \alpha^*, \alpha^* \right) - \left[ Z\left( \boldsymbol{\alpha}, \alpha^* \right) \right]^T \left[ Z_{j,\text{like}}^k \right]^{-1} Z\left( \boldsymbol{\alpha}, \alpha^* \right), \tag{31}$$

with $z_{j,\text{prior}}^k(\alpha^*, \alpha^*) = 1$. The value $q_{j,\text{posterior}}^k(\alpha^*)$ will be used as an estimate for $Q_j^k(\alpha^*)$.

## 3 HiPOD Reduction for Data Affected by Gaussian Noise

To address situations where the offline data is noisy, we propose here a variant of the directional HiPOD reduction presented in Sect. 2.1.2.

The idea is very straightforward. In the online phase we replace the initial interpolation step used to estimate coefficients $Q_j^k(\alpha^*)$ with a regression technique. In particular, formulas (21) and (30) provide us the desired estimate for the coefficients $Q_j^k(\alpha^*)$, when resorting to a polynomial or to a Gaussian process regression, respectively. Successively, the online phase

is performed exactly as in Sect. 2.1.2, going through the reconstructions (15)-(16), to obtain the final expansion in (17).

The choice for the ML regression models in Sects. 2.2.1, 2.2.2 is motivated by the highly nonlinear dependence of the offline HiMod solution onto the offline parameters.

The improvement led by the new HiPOD approach is numerically checked in Sect. 4. In the next section, we list some numerical quantities that can help us monitor the noise propagation throughout the HiPOD procedure.

## 3.1 Noise Propagation on the Response Matrix

In this section we compare the standard directional HiPOD procedure and the new variant proposed in this paper for different noise levels in the offline data. In particular, hereafter, we refer to the standard and to the new directional HiPOD approach as to the interpolation-HiPOD and the regression-HiPOD, respectively.

The data noise affects the HiPOD approximation by perturbing the $(L+1)$ SVD's involved in the offline phase. In [21], we showed that an inaccurate calculation of the first SVD (i.e., of the SVD of the response matrix $U$) compromises the reliability of the directional interpolation-HiPOD. For this reason, here we focus on the effect of the data noise onto the decomposition in (10).

Let us assume that a Gaussian noise affects the linear form in (1) (e.g., by perturbing the source term $f$ or the boundary data of the PDE problem at hand). As a consequence, the HiMod system in (5) is replaced by the perturbed problem

$$A_m(\alpha)\tilde{\mathbf{u}}_m(\alpha) = \mathbf{f}_m(\alpha) + \boldsymbol{\eta}_m, \tag{32}$$

where $\boldsymbol{\eta}_m \sim \mathcal{N}(\mathbf{0}_{mN_h}, \eta I_{mN_h})$ identifies the noise, with $\mathbf{0}_{mN_h} \in \mathbb{R}^{mN_h}$ the null vector, $I_{mN_h} \in \mathbb{R}^{mN_h} \times \mathbb{R}^{mN_h}$ the identity matrix, $\eta > 0$ the noise level, and $\tilde{\mathbf{u}}_m(\alpha)$ denotes the associated noisy HiMod discretization.

The solution to (32) can be regarded as a perturbation of the HiMod solution, $\mathbf{u}_m(\alpha)$, in (5) by an additive white noise $(A_m^{-1}(\alpha)\boldsymbol{\eta}_m)$, being

$$\tilde{\mathbf{u}}_m(\alpha) = A_m^{-1}(\alpha)\mathbf{f}_m(\alpha) + A_m^{-1}(\alpha)\boldsymbol{\eta}_m = \mathbf{u}_m(\alpha) + A_m^{-1}(\alpha)\boldsymbol{\eta}_m. \tag{33}$$

To simplify the notation, we define the random variable $\tilde{\boldsymbol{\eta}}_m(\alpha) = A_m^{-1}(\alpha)\boldsymbol{\eta}_m$, with

$$\tilde{\boldsymbol{\eta}}_m(\alpha) \sim \mathcal{N}\left(\mathbf{0}, \eta\left[A_m(\alpha)A_m^T(\alpha)\right]^{-1}\right), \tag{34}$$

so that the solution $\tilde{\mathbf{u}}_m(\alpha)$ to the perturbed HiMod linear system (32) can be recast as

$$\tilde{\mathbf{u}}_m(\alpha) = \mathbf{u}_m(\alpha) + \tilde{\boldsymbol{\eta}}_m(\alpha). \tag{35}$$

This decomposition finds a counterpart when assembling the response matrix, $\tilde{U}$, associated with the offline noisy data. Indeed, thanks to the additive property assumed for the noise $\boldsymbol{\eta}_m$, matrix $\tilde{U}$, which collects the perturbed HiMod solutions $\tilde{\mathbf{u}}_m(\alpha_i)$ in (32), for the $p$ values, $\alpha_1, \ldots, \alpha_p$, of the parameter $\alpha$, can be conceived as a perturbation of matrix $U$ in (9). Thus,

after introducing the noise matrix

$$E = \begin{bmatrix} \mathbf{E}^1(\alpha_1) \cdots \mathbf{E}^m(\alpha_1) | \mathbf{E}^1(\alpha_2) \cdots \mathbf{E}^m(\alpha_2) | \cdots \cdots | \mathbf{E}^1(\alpha_p) \cdots \mathbf{E}^m(\alpha_p) \end{bmatrix}$$

$$= \begin{bmatrix} \tilde{\eta}_{1,1}^{\alpha_1} & \cdots & \tilde{\eta}_{m,1}^{\alpha_1} & \tilde{\eta}_{1,1}^{\alpha_2} & \cdots & \tilde{\eta}_{m,1}^{\alpha_2} & \cdots\cdots & \tilde{\eta}_{1,1}^{\alpha_p} & \cdots & \tilde{\eta}_{m,1}^{\alpha_p} \\ \tilde{\eta}_{1,2}^{\alpha_1} & \cdots & \tilde{\eta}_{m,2}^{\alpha_1} & \tilde{\eta}_{1,2}^{\alpha_2} & \cdots & \tilde{\eta}_{m,2}^{\alpha_2} & \cdots\cdots & \tilde{\eta}_{1,2}^{\alpha_p} & \cdots & \tilde{\eta}_{m,2}^{\alpha_p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \tilde{\eta}_{1,N_h}^{\alpha_1} & \cdots & \tilde{\eta}_{m,N_h}^{\alpha_1} & \tilde{\eta}_{1,N_h}^{\alpha_2} & \cdots & \tilde{\eta}_{m,N_h}^{\alpha_2} & \cdots\cdots & \tilde{\eta}_{1,N_h}^{\alpha_p} & \cdots & \tilde{\eta}_{m,N_h}^{\alpha_p} \end{bmatrix}, \tag{36}$$

where the vectors

$$\mathbf{E}^k(\alpha_i) = \begin{bmatrix} \tilde{\eta}_{k,1}^{\alpha_i}, \tilde{\eta}_{k,2}^{\alpha_i}, \ldots, \tilde{\eta}_{k,N_h}^{\alpha_i} \end{bmatrix}^T \in \mathbb{R}^{N_h} \quad k = 1, \ldots, m, \ i = 1, \ldots, p,$$

gather, by mode, the noise affecting the modal coefficients $\{\tilde{u}_{k,j}^{\alpha_i}\}_{k=1,j=1}^{m,N_h}$ in (6) for $\alpha = \alpha_i$, the response matrix $\tilde{U} \in \mathbb{R}^{N_h \times (mp)}$ associated with the noisy HiMod solutions coincides with

$$\tilde{U} = U + E, \tag{37}$$

with $U$ the response matrix in (9). The SVD

$$\tilde{U} = \tilde{\Xi} \tilde{\Lambda} \tilde{K} \tag{38}$$

will consequently replace the factorization of matrix $U$ in (10), with $\tilde{\Xi} \in \mathbb{R}^{N_h \times N_h}$ and $\tilde{K} \in \mathbb{R}^{(mp) \times (mp)}$ the unitary matrices of the left and of the right singular vectors of $\tilde{U}$, and $\tilde{\Lambda} \in \mathbb{R}^{N_h \times (mp)}$ the pseudo-diagonal matrix of the singular values.

**Remark 1** The injection of a noise in the problem data involved in the bilinear form in (2) would still affect the HiMod solution $\mathbf{u}_m(\alpha)$, in (5), but not in an additive fashion. This would unavoidably make the perturbation propagation analysis more complex, and is beyond the purpose of this paper.

Some results are available in the literature which relate both the singular values and the singular vectors of matrices $U$ and $\tilde{U}$.

As for the singular values, we remind the Weyl theorem [40] and the Mirsky theorem [26]. The first result controls the discrepancy between the $i$-th singular value $\lambda_i$ of $U$ and the corresponding singular value $\tilde{\lambda}_i$ of $\tilde{U}$ in terms of the spectral norm, $\|E\|_S$, of the noise matrix, being

$$|\tilde{\lambda}_i - \lambda_i| \leq \|E\|_S \quad i = 1, \ldots, \min\{N_h, mp\}. \tag{39}$$

Mirsky theorem provides an upper bound on the sum of the quadratic deviations of values $\tilde{\lambda}_i$'s with respect to $\lambda_i$'s in terms of the Frobenius norm, $\|E\|_F$, of the noise matrix, given by

$$\sum_{i=1}^{\min\{N_h, mp\}} \left(\tilde{\lambda}_i - \lambda_i\right)^2 \leq \|E\|_F. \tag{40}$$

Of course, inequalities (39) and (40) are completely useless when norms $\|E\|_S$ and $\|E\|_F$ become larger and larger.

The reference result on the singular vectors is represented by the generalized $\sin \theta$ theorem [39]. To state such a result, it is instrumental to introduce an appropriate rewriting of the

SVD's in (10) and (38). By exploiting that matrices $\varXi$, $K$, and $\tilde{\varXi}$, $\tilde{K}$ are unitary, it follows that

$$U = U_1 + U_0, \quad \tilde{U} = \tilde{U}_1 + \tilde{U}_0,$$

where $U_s = \varXi_s \varLambda_s K_s^T$, $\tilde{U}_s = \tilde{\varXi}_s \tilde{\varLambda}_s \tilde{K}_s^T$, for $s = 0, 1$, with $K_1 = [\mathbf{k}_1, \ldots, \mathbf{k}_r]$, $K_0 = [\mathbf{k}_{r+1}, \ldots, \mathbf{k}_{mp}]$, $\varXi_1 = [\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_r]$, $\varXi_0 = [\boldsymbol{\xi}_{r+1}, \ldots, \boldsymbol{\xi}_{N_h}]$,

$$\varLambda_1 = \begin{cases} \begin{bmatrix} \mathrm{diag}(\lambda_1, \ldots, \lambda_r, 0, \ldots, 0) \\ \mathbf{0}_{(N_h-r) \times N_h} \end{bmatrix} & \text{if } N_h \geq mp \\ \begin{bmatrix} \mathrm{diag}(\lambda_1, \ldots, \lambda_r, 0, \ldots, 0), \mathbf{0}_{mp \times (mp-r)} \end{bmatrix} & \text{if } N_h < mp, \end{cases}$$

$$\varLambda_0 = \begin{cases} \begin{bmatrix} \mathrm{diag}(0, \ldots, 0, \lambda_{r+1}, \ldots, \lambda_{\min\{N_h, mp\}}) \\ \mathbf{0}_{(N_h-r) \times N_h} \end{bmatrix} & \text{if } N_h \geq mp \\ \begin{bmatrix} \mathrm{diag}(0, \ldots, 0, \lambda_{r+1}, \ldots, \lambda_{\min\{N_h, mp\}}), \mathbf{0}_{mp \times (mp-r)} \end{bmatrix} & \text{if } N_h < mp, \end{cases}$$

matrices $\tilde{K}_1$, $\tilde{K}_0$, $\tilde{\varXi}_1$, $\tilde{\varXi}_0$, $\tilde{\varLambda}_1$, $\tilde{\varLambda}_0$ being defined accordingly.

Moreover, we denote by $\mathcal{C}_1$ and $\mathcal{C}_2$ two generic Euclidean subspaces of $\mathbb{R}^{N_h}$, and by $P_{\mathcal{C}_1}$, $P_{\mathcal{C}_2} \in \mathbb{R}^{N_h}$ the associated orthogonal projection operators. Thus, the angle $\theta$ between a vector $\mathbf{x} \in \mathbb{R}^{N_h}$ and the subspace $\mathcal{C}_1$ (which, among all the mathematically equivalent formulations, by convention is always taken acute and positive) can be defined by

$$\sin \theta(\mathbf{x}, \mathcal{C}_1) = \min_{\mathbf{y} \in \mathcal{C}_1} \|\mathbf{x} - \mathbf{y}\|_2, \tag{41}$$

with $\|\cdot\|_2$ denoting the Euclidean norm of a vector, and with $\|\mathbf{x}\|_2 = 1$. From the projection theorem, it follows that

$$\min_{\mathbf{y} \in \mathcal{C}_1} \|\mathbf{x} - \mathbf{y}\|_2 = \|(I - P_{\mathcal{C}_1} \mathbf{x})\|_2. \tag{42}$$

When considering the angle between the subspaces $\mathcal{C}_1$ and $\mathcal{C}_2$, it is common to define

$$\sin \theta(\mathcal{C}_1, \mathcal{C}_2) = \|(I - P_{\mathcal{C}_2}) P_{\mathcal{C}_1}\|_S. \tag{43}$$

Now, if there exists a pair of values $\gamma > 0$, $\delta > 0$ such that

$$\lambda_{\min}(\tilde{U}_1) \geq \gamma + \delta, \quad \lambda_{\max}(U_0) \leq \gamma, \tag{44}$$

with $\lambda_{\min}(\tilde{U}_1)$ and $\lambda_{\max}(U_0)$ the minimum and the maximum eigenvalue of $\tilde{U}_1$ and $U_0$, respectively, the generalized $\sin \theta$ theorem [39] ensures that the following perturbation bound on the column space of $\tilde{\varXi}$ holds

$$\sin \theta \left( R(\tilde{\varXi}_1), R(\varXi_1) \right) \leq \frac{\|E\|_S}{\delta} = \mathcal{P}_{B1}, \tag{45}$$

with $R(W)$ the column space associated with the generic matrix $W$. The relation (45) is not always computationally convenient to provide meaningful insight on the actual perturbation triggered by the noise in the data. Indeed, computing the perturbation between the two column spaces $R(\tilde{\varXi}_1)$ and $R(\varXi_1)$ might be unpractical from a computational view point, due to the infinite vectors to be spanned.

In order to make the control in (45) more computationally convenient, i.e., to restrict the dimensionality of the spaces to be spanned, we consider the alternative inequality

$$\sin \theta \left( R(\tilde{\varXi}_1), R(\varXi_1) \right) \leq \sin \theta \left( \tilde{\boldsymbol{\xi}}_1, \boldsymbol{\xi}_1 \right) = \mathcal{P}_{B2}, \tag{46}$$

with $\boldsymbol{\xi}_1$ and $\tilde{\boldsymbol{\xi}}_1$ denoting the first left singular vectors of the response matrix $U$ and $\tilde{U}$, respectively. The result in (46) can be easily proved, moving from the following chain of inequalities:

$$
\begin{aligned}
\sin\theta\left(\tilde{\boldsymbol{\xi}}_1, \boldsymbol{\xi}_1\right) &\geq \sin\theta\left(\tilde{\boldsymbol{\xi}}_1, \text{span}\{\boldsymbol{\xi}_1, \boldsymbol{\xi}_2\}\right) \geq \sin\theta\left(\tilde{\boldsymbol{\xi}}_1, \text{span}\{\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \boldsymbol{\xi}_3\}\right) \\
&\geq \ldots \geq \sin\theta\left(\tilde{\boldsymbol{\xi}}_1, R\left(\varXi\right)\right) \geq \sin\theta\left(\text{span}\{\tilde{\boldsymbol{\xi}}_1, \tilde{\boldsymbol{\xi}}_2\}, R\left(\varXi\right)\right) \\
&\geq \sin\theta\left(\text{span}\{\tilde{\boldsymbol{\xi}}_1, \tilde{\boldsymbol{\xi}}_2, \tilde{\boldsymbol{\xi}}_3\}, R(\varXi)\right) \geq \ldots \\
&\geq \sin\theta\left(R(\tilde{\varXi}_1), R\left(\varXi_1\right)\right).
\end{aligned}
$$

Bounds (39), (40), (45), together with the new one in (46), will be used in the next section to assess the sensitivity both of the interpolation-HiPOD and of the regression-HiPOD to the noise level.

## 4 Numerical Results

In this section we compare the interpolation-HiPOD with the regression-HiPOD methods. The comparison in carried out on two test cases, in terms of accuracy and robustness to the noise level. For the former method, we resort to the PCH interpolant, while the latter approach is assessed both with the polynomial and with the Gaussian process regression.

### 4.1 Test Case 1

We choose as reference setting the ADR problem (1)-(2) solved on the rectangular domain $\Omega = (0, 6) \times (0, 1)$, after setting the problem data to

$$
\begin{aligned}
\mu(x, y) &= 0.24, \quad \mathbf{b}(x, y) = [5, \sin(6x)]^T, \quad \sigma(x, y) = 0.1, \\
f(x, y) &= 10\chi_{E_1}(x, y) + 10\chi_{E_2}(x, y),
\end{aligned}
\tag{47}
$$

with $\chi_\omega$ the characteristic function associated with the generic region $\omega \subset \mathbb{R}^2$, $E_1$ and $E_2$ the ellipsoidal areas in $\Omega$ given by $\{(x, y) : (x - 0.75)^2 + 0.4(y - 0.25)^2 < 0.01\}$ and $\{(x, y) : (x - 0.75)^2 + 0.4(y - 0.75)^2 < 0.01\}$, respectively. The ADR problem is completed with a homogeneous Neumann data on $\Gamma_N = \{(x, y) : x = 6, 0 \leq y \leq 1\}$, while a homogeneous Dirichlet condition is assigned on $\Gamma_D = \partial\Omega \setminus \Gamma_N$, so that space $V$ in (1) coincides with $H^1_{\Gamma_D}(\Omega)$.
From a modeling viewpoint, this setting can be adopted to simulate the propagation of a pollutant released by two localized sources within a straight channel, under the effect of a sinusoidal horizontal convection.

We parametrize the ADR problem with respect to the diffusivity $\mu$. The offline phase is set up so that the coefficient $\mu$ uniformly spans the range $[0.2, 0.8]$. All the other problem data remain the same as in (47) through the entire offline phase. We hierarchically reduce 100 ADR problems, after discretizing the main dynamics with linear finite elements on a uniform partition of $\Omega_{1D}$ into 40 sub-intervals, while using 20 sinusoidal modal basis functions to approximate the transverse dynamics. To investigate the robustness to the noise of the interpolation- and of the regression-HiPOD procedures, we carry out different offline phases, where the source term $f$ is injected by different levels of white noise, thus changing $f$ into $f + \eta$, with $\eta = 0.01, 0.05, 0.1, 0.25$. Independently of the selected noise, the
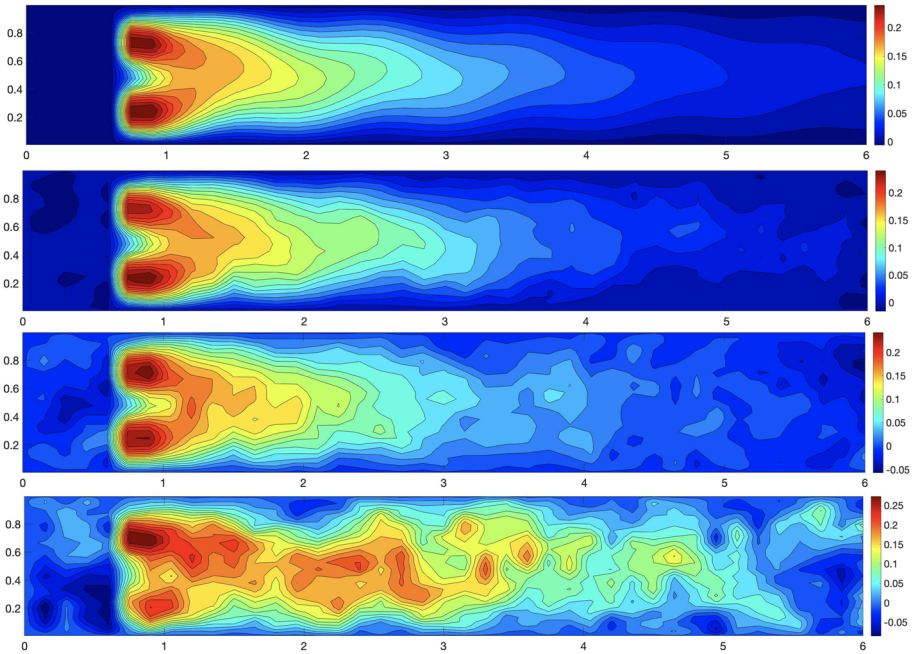
**Fig. 1** Test case 1. HiMod reference solution (top panel); interpolation-HiPOD approximation for different noise levels: $\eta = 0.01, 0.1, 0.25$ (second-fourth row)

online phase is used to recover the reference setting in (47), i.e., to reconstruct the HiMod solution for $\mu^* = 0.24$. The top panel in Fig. 1 shows the approximation yielded by the HiMod discretization adopted for the offline phase. The resolution of such a discretization is sufficiently accurate to capture the oscillatory dynamics induced by the sinusoidal field, together with the presence of the two localized sources in $E_1$ and $E_2$. Such a discretization represents the reference solution HiPOD approximations will be compared with.

A unique threshold tolerance, $\varepsilon$, is used to automatically select the number of POD modes to be preserved at the first and at the second stage of the directional method, which is equivalent to setting $\epsilon_1 = \epsilon_2 = \epsilon$ in (18) (we refer the reader to [21] for a thorough investigation about the interplay between tolerances $\epsilon_1$ and $\epsilon_2$). Table 1 collects the distribution of the number $\mu_j$ of the left singular vectors retained at each finite element node, for different choices of the tolerance $\varepsilon$ (by rows) and of the noise $\eta$ (by columns). As expected, number $\mu_j$ increases for larger and larger values both of $\varepsilon$ and $\eta$. Indeed, in the former case, the procedure is intrinsically requested to retain more information about the variability of the offline dataset. In the latter case, the quality of information retained by the offline solution deteriorates with increasing values of noise. This leads the HiPOD procedure to retain more singular vectors with respect to the case of a low (or of the absence of) noise, in order to capture the same amount of information from the offline data.

Figure 1 displays the interpolation-HiPOD approximation reconstructed from offline data with an increasing level of noise, i.e., $\eta = 0.01$, $\eta = 0.1$, $\eta = 0.25$ (the contourplot for $\eta = 0.05$ is omitted since it is very similar to the one associated with $\eta = 0.1$). As expected, the quality of the HiPOD solution deteriorates with increasing values of $\eta$. In particular, the wake behind the sources is completely lost by the HiPOD approximation for $\eta = 0.25$.

**Table 1** Test case 1



Distribution of the number of the POD modes retained at each finite element node for $\varepsilon = 0.9$ (first row), 0.99 (second row), 0.999 (third row), and for a different noise $\eta$. The supporting fiber is discretized by 41 uniformly distributed nodes
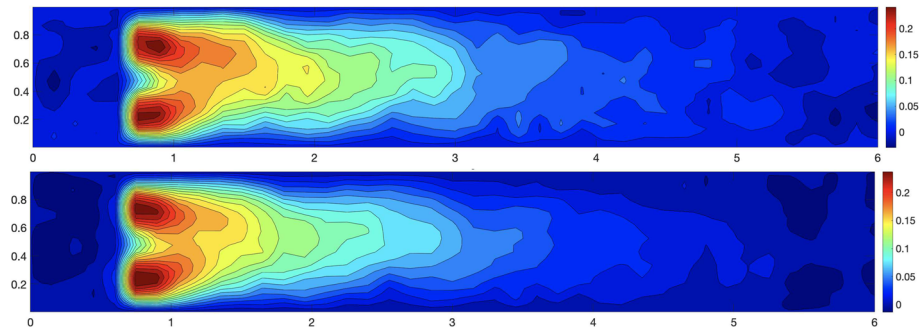


**Fig. 2** Test case 1. Regression-HiPOD approximation for a cubic polynomial fitting (top) and for a Gaussian process regression (bottom); the noise level is set to $\eta = 0.25$

Figure 2 highlights the benefits obtained by replacing the interpolation step in the HiPOD online phase with a regressione process. In particular, the two panels show the regression-HiPOD approximation when resorting to a cubic polynomial fitting and to a Gaussian process regression, for the largest level of noise analyzed in Fig. 1 (i.e., $\eta = 0.25$). Both fitting models clearly outperform the interpolation-HiPOD. In particular, the Gaussian process generates a solution of a better quality with respect to the one yielded by the cubic polynomial fitting.

This trend is confirmed by the values in Table 2, which gathers the $L^2(\Omega)$- and the $H^1(\Omega)$-norm of the relative modelig error obtained when replacing the reference HiMod discretization in Fig. 1 (top) with the interpolation- rather than the regression-HiPOD approximation. To ensure a high accuracy to the HiPOD solutions, we have set the thresholds driving the selection of the POD bases in (18) very close to 1, picking $\epsilon_1 = \epsilon_2 = \epsilon = 0.9999$. The interpolation-HiPOD scheme is compared with the regression-HiPOD approach when both a cubic polynomial fitting and a Gaussian process are used, for the four levels of noise considered above. Although the general trend confirms that the error increases with the noise, cubic polynomial fitting and Gaussian process regression outperform considerably the interpolation-HiPOD, in particular for high noise levels.

Finally, we quantify the perturbation bounds in (45) and (46), for the considered four levels of noise. Actually, Table 3 shows that the new bound allows us to gain up to three

**Table 2** Test case 1

|      |                    | $\eta = 0.01$ | $\eta = 0.05$ | $\eta = 0.1$ | $\eta = 0.25$ |
|------|--------------------|---------------|---------------|--------------|---------------|
| PCH  | $L^2(\Omega)$-norm | 0.0222        | 0.0960        | 0.2258       | 0.5521        |
|      | $H^1(\Omega)$-norm | 0.0434        | 0.2209        | 0.4460       | 1.1431        |
| CP   | $L^2(\Omega)$-norm | 0.0060        | 0.0247        | 0.0648       | 0.1414        |
|      | $H^1(\Omega)$-norm | 0.0132        | 0.0592        | 0.1282       | 0.3913        |
| GP   | $L^2(\Omega)$-norm | 0.0096        | 0.0229        | 0.0742       | 0.1521        |
|      | $H^1(\Omega)$-norm | 0.0169        | 0.0559        | 0.1087       | 0.2730        |

Relative modeling error associated with the interpolation-HiPOD (PCH) and with the cubic polynomial (CP) and the Gaussian process (GP) regression-HiPOD, for different levels of noise

**Table 3** Test case 1

|                     | $\eta = 0.01$ | $\eta = 0.05$ | $\eta = 0.1$ | $\eta = 0.25$ |
|---------------------|---------------|---------------|--------------|---------------|
| $\mathcal{P}_{B1}$  | 2.2819        | 3.3340        | 2.6980       | 3.9770        |
| $\mathcal{P}_{B2}$  | 0.0032        | 0.0072        | 0.0173       | 0.0517        |

Perturbation bounds for the noise propagation in the response matrix

orders of magnitude with respect to quantifier $\mathcal{P}_{B1}$, thus offering an effective tool to evaluate the noise propagation on the response matrix $U$ in (9).

## 4.2 Test Case 2

As a second reference setting, we hierarchically reduce the ADR problem in (1)-(2) identified by the data

$$\mu(x, y) = 0.24, \quad \mathbf{b}(x, y) = [20, 2\sin(6x)]^T, \quad \sigma(x, y) = 0.1,$$
$$f(x, y) = 1000\chi_{R_1}(x, y) + 1000\chi_{R_2}(x, y), \tag{48}$$

on the same domain, $\Omega = (0, 6) \times (0, 1)$, as in Test case 1, with $R_1 = \{(x, y) : 1 < x < 2, \ 0 < y < 0.1\}$ and $R_2 = \{(x, y) : 1 < x < 2, \ 0.9 < y < 1.0\}$ two rectangular regions of interest. The problem is completed by homogeneous Neumann data on $\Gamma_N = \{(x, y) : x = 6, 0 \le y \le 1\}$ and by a homogeneous Dirichlet condition on $\Gamma_D = \partial\Omega \setminus \Gamma_N$, so that we have $V \equiv H^1_{\Gamma_D}(\Omega)$ in (1).

We can adopt this configuration to model, for instance, the transport of a drug released by a medical stent applied to the walls of a cardiovascular vessel, under the effect of incompressible fluid convection.

The HiMod discretization adopted to build the reference solution as well as to perform the offline phase of the HiPOD procedure employs linear finite elements along $\Omega_{1D}$ after subdiving the supporting fiber into 60 uniform subintervals, and a modal basis consisting of 20 sinusoidal functions to capture the dynamics along the transverse direction. The top panel in Fig. 3 displays the contour plot of the HiMod reference solution. We clearly distinguish the drug release in the regions $R_1$ and $R_2$, together with the transport of the medicine along the pipe. The offline phase is driven by the diffusivity, which is identified with parameter $\alpha$. In particular, we uniformly cover the range $[0.2, 0.8]$ with 100 samples, while keeping the same values as in (48) for all the other problem data.
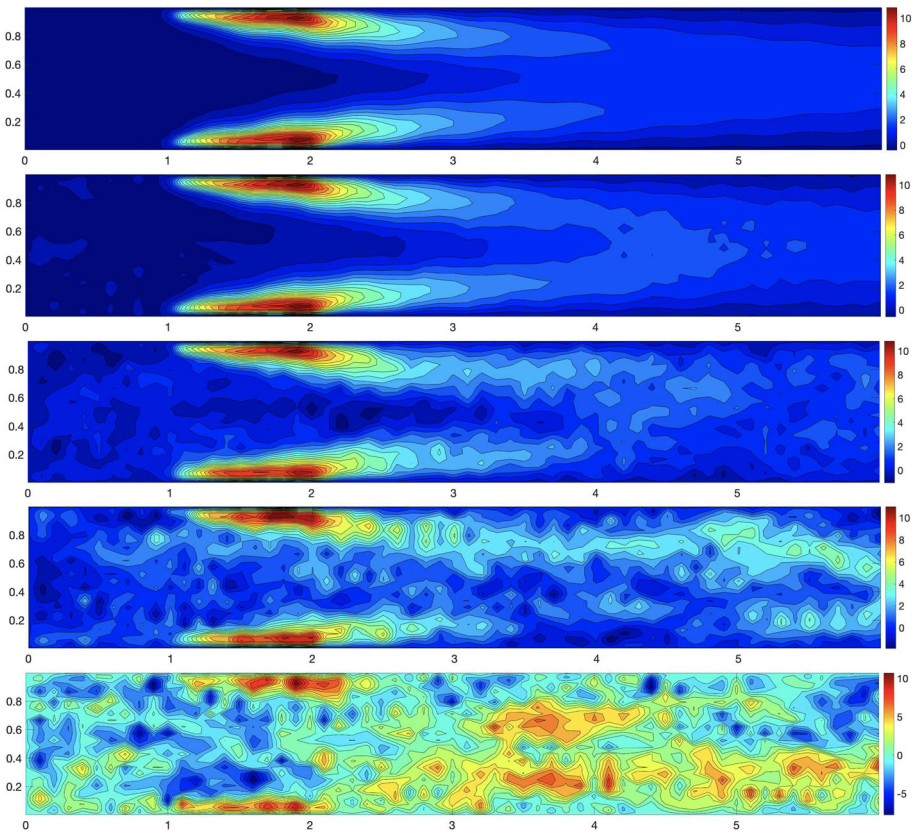
**Fig. 3** Test case 2. HiMod reference solution (top panel); interpolation-HiPOD approximation for different noise levels: $\eta = 0.01, 0.05, 0.1, 0.25$ (second-fifth row)

The performance of the interpolation- and of the regression-HiPOD approaches in the presence of noise is analyzed by affecting the forcing term in (2) with an increasing white noise $\eta$, set to $\eta = 0.01, 0.05, 0.1, 0.25$, respectively. For each choice of $\eta$, we replicate the offline phase set above, before predicting the HiMod approximation associated with parameter $\mu^* = 0.24$ (i.e., with the reference configuration) in the online phase.

The selection of the POD bases $\{\boldsymbol{\xi}_j\}_{j=1}^{L}$ and $\{\mathbf{r}_j^k\}_{k=1}^{\mu_j}$, for $j = 1, \ldots, L$, is driven by the variance-based criteria in (18), for a unique choice of the threshold tolerance (i.e., for $\epsilon_1 = \epsilon_2 = \epsilon$). The trend of the number $\mu_j$ of the left singular vectors selected by the HiPOD approach at the finite element nodes for different values of $\epsilon$ and $\eta$ is very similar to the one in Table 1 (and, consequently, skipped for shortness).

Figure 3 shows the interpolation-HiPOD solution for the four noise levels $\eta = 0.01$, $0.05, 0.1, 0.25$. As for Test case 1, the quality of the HiPOD approximation deteriorates very quickly when increasing the level of noise. In particular, for $\eta = 0.25$, the HiPOD solution is fully noisy, the problem dynamics being completely lost.

On the contrary, the regression-HiPOD discretization significantly outperforms the approximation quality provided by the interpolation. Figure 4 shows such an improvement when resorting to a cubic polynomial fitting and to a Gaussian process regression. Analo-
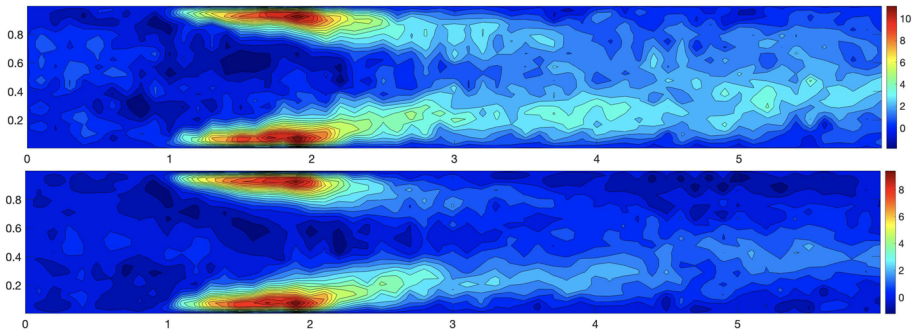
**Fig. 4** Test case 2. Regression-HiPOD approximation for a cubic polynomial fitting (top) and for a Gaussian process regression (bottom); the noise level is set to $\eta = 0.25$

**Table 4** Test case 2

|  |  | $\eta = 0.01$ | $\eta = 0.05$ | $\eta = 0.1$ | $\eta = 0.25$ |
|---|---|---|---|---|---|
| PCH | $L^2(\Omega)$-norm | 0.0446 | 0.1917 | 0.4310 | 1.0589 |
|  | $H^1(\Omega)$-norm | 0.0626 | 0.3177 | 0.6723 | 1.6280 |
| CP | $L^2(\Omega)$-norm | 0.0163 | 0.0887 | 0.1408 | 0.3067 |
|  | $H^1(\Omega)$-norm | 0.0213 | 0.0928 | 0.1875 | 0.4052 |
| GP | $L^2(\Omega)$-norm | 0.0137 | 0.0832 | 0.1617 | 0.3593 |
|  | $H^1(\Omega)$-norm | 0.0182 | 0.0982 | 0.1758 | 0.3703 |

Relative modeling error associated with the interpolation-HiPOD (PCH) and with the cubic polynomial (CP) and the Gaussian process (GP) regression-HiPOD, for different levels of noise

**Table 5** Test case 2

|  | $\eta = 0.01$ | $\eta = 0.05$ | $\eta = 0.1$ | $\eta = 0.25$ |
|---|---|---|---|---|
| $\mathcal{P}_{B1}$ | 2.4175 | 3.3268 | 3.7613 | 4.9623 |
| $\mathcal{P}_{B2}$ | 0.0030 | 0.0208 | 0.0841 | 0.4279 |

Perturbation bounds for the noise propagation in the response matrix

gously to Fig. 2, Gaussian process regression yields a solution of a better quality than cubic polynomial fitting.

These qualitative considerations are confirmed by the values in Table 4, which gathers the $L^2(\Omega)$- and the $H^1(\Omega)$-norm of the relative modeling error associated with both the interpolation- and the regression-HiPOD approximations, for the considered noise levels. The results corroborate what already remarked for the first test case, namely cubic polynomial fitting and Gaussian process regression outperform the interpolation-HiPOD for high noise levels.

As a last check, we compute the quantifiers $\mathcal{P}_{B1}$ and $\mathcal{P}_{B2}$ of the noise propagation in the response matrix defined in (45) and (46), respectively (see Table 5). Also for this test case, the bound in (45) is relatively large, and the discrepancy between $\mathcal{P}_{B1}$ and $\mathcal{P}_{B2}$ grows with the level of noise.

## 5 Conclusions and Future Developments

In this work we present a first-of-its-kind approach to make the directional HiPOD approach robust when the offline data is noisy. To this aim, we modify the interpolation-HiPOD proposed in [21] by replacing interpolation techniques with ML regression, which gives rise to what we name the regression-HiPOD method.

Numerical results, although preliminary, showcase the performance of the new HiPOD approach in ideal situations, where both the noisy and clean data is available. This allows for a practical validation of the quality of the reconstruction yielded by interpolation-HiPOD against regression-HiPOD.

In particular, when the level of noise in the data is large, regression-HiPOD outperforms interpolation-HiPOD in terms of accuracy, by gaining up to an order in the $L^2(\Omega)$- and $H^1(\Omega)$-norm of the relative modeling error. Moreover, the regression-HiPOD succeeds in accurately reproducing the HiMod solution also in the presence of localized abrupt dynamics (where interpolation-HiPOD plainly fails), as in the second test case where the forcing term is localized in narrow regions close to the boundary.

Additionally, we provide a new upper bound that estimates the effect of the noise level on the deformation of the subspace spanned by the left singular vectors of the response matrix. The new quantity we propose is more practical to compute and provides a more meaningful estimate with respect to some upper bounds available in the literature, as confirmed by the values in Tables 3 and 5.

Concerning possible future developments, we plain to include uncertainty quantification in the analysis, to extend the directional regression-HiPOD approach to handle multiple parameters simultaneously and to model vector PDEs such as the incompressible Stokes and Navier-Stokes equations, with a view to haemodynamics modeling [7].

**Data Availability** Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

# References

1. Aletti, M., Perotto, S., Veneziani, A.: HiMod reduction of advection-diffusion-reaction problems with general boundary conditions. J. Sci. Comput. **76**(1), 89–119 (2018)
2. Audouze, C., De Vuyst, F., Nair, P.: Reduced-order modeling of parameterized PDEs using time-space-parameter principal. Int. J. Numer. Methods Eng. **80**(8), 1025–1057 (2009)
3. Baroli, D., Cova, C., Perotto, S., Sala, L., Veneziani, A.: Hi-POD solution of parametrized fluid dynamics problems: preliminary results. In: Model Reduction of Parametrized Systems, *MS&A. Model. Simul. Appl.*, vol. 17, pp. 235–254. Springer, Cham (2017)
4. Baur, U., Beattie, C., Benner, P., Gugercin, S.: Interpolatory projection methods for parameterized model reduction. SIAM J. Sci. Comput. **33**(5), 2489–2518 (2011)
5. Bayes, T.: An essay towards solving a problem in the doctrine of chances. Philos. Trans. R. Soc. **53**, 370–418 (1763)
6. Benner, P., Gugercin, S., Willcox, K.: A survey of projection-based model reduction methods for parametric dynamical systems. SIAM Rev. **57**(4), 483–531 (2015)
7. Brandes Costa Barbosa, Y., Perotto, S.: Hierarchically reduced models for the Stokes problem in patient-specific artery segments. Int. J. Comput. Fluid Dyn. (2020)
8. Csato, L., Opper, M.: Sparse on-line Gaussian processes. Neural Comput. **14**(3), 641–668 (2002)
9. Dudley, R.M.: Sample functions of the Gaussian process. Ann. Probab. **1**(1) (1973)
10. Ern, A., Guermond, J.L.: Theory and practice of finite elements, applied mathematical sciences, vol. 159. Springer-Verlag, New York (2004)
11. Ern, A., Perotto, S., Veneziani, A.: Hierarchical model reduction for advection-diffusion-reaction problems. In: Numerical Mathematics and Advanced Applications, pp. 703–710. Springer, Berlin (2008)
12. Fan, J.: Local Polynomial Modelling and Its Applications: From Linear Regression to Nonlinear Regression. Monographs on Statistics and Applied Probability. Chapman & Hall/CRC (1996)
13. Gergonne, J.D.: The application of the method of least squares to the interpolation of sequences. Hist. Math. **1**(4), 439–447 (1815)
14. Ghattas, O., Willcox, K.: Learning physics-based models from data: perspectives from inverse problems and model reduction. Acta Numer. **30**, 445–554 (2021)
15. Golub, G., Van Loan, C.: Matrix computations, In: 4th edn. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD (2013)
16. Guzzetti, S., Perotto, S., Veneziani, A.: Hierarchical model reduction for incompressible fluids in pipes. Int. J. Numer. Methods Eng. **114**(5), 469–500 (2018)
17. Hesthaven, J., Rozza, G., Stamm, B.: Certified Reduced Basis Methods for Parametrized Partial Differential Equations. SpringerBriefs in Mathematics. Springer, Cham; BCAM Basque Center for Applied Mathematics, Bilbao (2016)
18. Kahlbacher, M., Volkwein, S.: Galerkin proper orthogonal decomposition methods for parameter dependent elliptic systems. Discuss. Math. Differ. Incl. Control Optim. **27**(1), 95–117 (2007)
19. Kerschen, G., Golinval, J., Vakakis, A., Bergman, L.: The method of Proper Orthogonal Decomposition for dynamical characterization and order reduction of mechanical systems: an overview. Nonlinear Dyn. **41**(1–3), 147–169 (2005)
20. Kunisch, K., Volkwein, S.: Galerkin proper orthogonal decomposition methods for a general equation in fluid dynamics. SIAM J. Numer. Anal. **40**(2), 492–515 (2002)
21. Lupo Pasini, M., Perotto, S.: Hierarchical model reduction driven by a proper orthogonal decomposition for parametrized advection-diffusion-reaction problems. Electronic Trans. Numer. Anal. **55**(187), 187–212 (2022)
22. Magee, L.: Nonlocal behavior in polynomial regressions. Am. Stat. **52**(1), 20–22 (1998)
23. Mansilla Alvarez, L., Blanco, P., Bulant, C., Dari, E., Veneziani, A., Feijóo, R.: Transversally enriched pipe element method (TEPEM) an effective numerical approach for blood flow modeling. Int. J. Numer. Methods Biomed. Eng. **33**(4), e02808 (2017)
24. Marcus, M.B.: Continuity of Gaussian processes. Trans. Am. Math. Soc. **151**(2), 377–391 (1970)
25. Marcus, M.B.: Sample behavior of Gaussian process. In: Proceedings of the sixth Berkeley Symposium on Mathematical Statistics and Probability, vol. 2, pp. 423–441. Project Euclid (1972)
26. Mirsky, L.: Symmetric gauge functions and unitarily invariant norms. Q.J. Math. **11**, 50–59 (1960)
27. Ohlberger, M., Smetana, K.: A dimensional reduction approach based on the application of reduced basis methods in the framework of hierarchical model reduction. SIAM J. Sci. Comput. **36**(2), A714–A736 (2014)
28. Perotto, S.: Hierarchical model (Hi-Mod) reduction in non-rectilinear domains. In: Domain Decomposition Methods in Science and Engineering XXI, *Lect. Notes Comput. Sci. Eng.*, vol. 98, pp. 477–485. Springer, Cham (2014)

29. Perotto, S.: A survey of hierarchical model (Hi-Mod) reduction methods for elliptic problems. In: Numerical simulations of coupled problems in engineering, *Comput. Methods Appl. Sci.*, vol. 33, pp. 217–241. Springer, Cham (2014)
30. Perotto, S., Ern, A., Veneziani, A.: Hierarchical local model reduction for elliptic problems: a domain decomposition approach. Multiscale Model. Simul. **8**(4), 1102–1127 (2010)
31. Perotto, S., Reali, A., Rusconi, P., Veneziani, A.: HIGAMod: a hierarchical isogeometric approach for model reduction in curved pipes. Comput. Fluids **142**, 21–29 (2017)
32. Perotto, S., Veneziani, A.: Coupled model and grid adaptivity in hierarchical reduction of elliptic problems. J. Sci. Comput. **60**(3), 505–536 (2014)
33. Perotto, S., Zilio, A.: Hierarchical model reduction: three different approaches. In: Numerical mathematics and advanced applications 2011, pp. 851–859. Springer, Heidelberg (2013)
34. Perotto, S., Zilio, A.: Space-time adaptive hierarchical model reduction for parabolic equations. Adv. Model. and Simul. in Eng. Sci. **2:25** (2015)
35. Quarteroni, A., Manzoni, A., Negri, F.: Reduced Basis Methods for Partial Differential Equations, Unitext, vol. 92. Springer, Cham (2016)
36. Smetana, K., Ohlberger, M.: Hierarchical model reduction of nonlinear partial differential equations based on the adaptive empirical projection method and reduced basis techniques. ESAIM Math. Model. Numer. Anal. **51**(2), 641–677 (2017)
37. Smith, K.: On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of the observations. Biometrika **12**(1/2), 1–85 (1918)
38. Volkwein, S.: Proper orthogonal decomposition: theory and reduced-order modelling. University of Konstanz, Lecture notes (2013)
39. Wedin, P.A.: Perturbation bounds in connection with singular value decomposition. BIT Numer. Math. **12**, 99–111 (1972)
40. Weyl, H.: Das asymptotische verteilungsgesetz der eigenwerte linearer partieller differentialgleichungen (mit einer anwendung auf die theorie der hohlraumstrahlung). Math. Ann. **71**(4), 441–479 (1912)

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