

Computer Science Technical Report CSTR-22

November 11, 2015

Azam Moosavi, Razvan Stefanescu, Adrian Sandu

“Efficient Construction of Local Parametric
Reduced Order Models
Using Machine Learning Techniques”

Computational Science Laboratory
Computer Science Department
Virginia Polytechnic Institute and State University
Blacksburg, VA 24060
Phone: (540)-231-2193
Fax: (540)-231-6075
Email: rstefan@ncsu.edu
Web: <http://csl.cs.vt.edu>



Abstract

Reduced order models are computationally inexpensive approximations that capture the important dynamical characteristics of large, high-fidelity computer models of physical systems. This paper applies machine learning techniques to improve the design of parametric reduced order models. Specifically, machine learning is used to develop feasible regions in the parameter space where the admissible target accuracy is achieved with a predefined reduced order basis, to construct parametric maps, to choose the best two already existing bases for a new parameter configuration from accuracy point of view and to pre-select the optimal dimension of the reduced basis such as to meet the desired accuracy. By combining available information using bases concatenation and interpolation as well as high-fidelity solutions interpolation we are able to build accurate reduced order models associated with new parameter settings. Promising numerical results with a viscous Burgers model illustrate the potential of machine learning approaches to help design better reduced order models.

key words reduced order models, high-fidelity models, data fitting, machine learning, feasible region of parameters, local reduced order models.

1 Introduction

Many physical phenomena are described mathematically by partial differential equations (PDE), and, after applying suitable discretization schemes, are simulated on a computer. PDE-based models frequently require calibration and parameter tuning in order to provide realistic simulation results. Recent developments in the field of uncertainty quantification [52, 85, 35, 18] provide the necessary tools for validation of such models even in the context of variability and lack of knowledge on the input parameters. While uncertainty propagation techniques such as Markov chain [45] and perturbation methods [15, 16, 17] can measure the impact of uncertain parameters on some quantities of interest, they often become infeasible due to the large number of model realizations requirement. Similar difficulties are encountered when solving Bayesian inference problems since sampling from posterior distribution is required.

For large-scale simulations, the variational [79, 78, 21, 60, 61, 87] and ensemble [30, 96, 48, 43] based inference approaches are widely used in practice. Their efficiency decreases with increasing computational complexity of the underlying physical models. However, increasing model complexity is unavoidable as science fields progress. For example, finer space resolution of the underlying PDE models is one of the most important factors contributing to the one day/decade growth rate of the reliability of atmospheric weather predictions [13, 92].

The need for computational efficiency motivated the development of surrogate models such as response surfaces, low resolution, and reduced basis models, in order to facilitate optimization, inference, and uncertainty quantification.

Data fitting or response surface models [85] are constructed using only a data-driven angle. The underlying physics remains unknown and only the input-output behavior of the model is considered. Data fitting can use techniques such as regression, interpolation, radial basis function, Gaussian Processes, artificial neural networks and other supervised machine-learning methods. The latter techniques can automatically detect patterns in data, and one can use them to predict future data under uncertainty in a probabilistic framework [59]. While easy to implement due to the non-intrusive nature, the prediction abilities may suffer since the governing physics is not specifically accounted for.

Low fidelity models attempt to reduce the computational burden of the high-fidelity models by neglecting some of the physical aspects (e.g., replacing Navier-Stokes and Large Eddy Simulations with inviscid Euler's equations and Reynolds-Averaged Navier-Stokes [33, 76, 97], or decreasing the spatial resolution [20, 90]). The additional approximations, however, may considerably degrade the physical solution with only a modest decrease of the computational load.

Reduced basis [67, 9, 34, 65, 74, 24] and Proper Orthogonal Decomposition [49, 56, 42, 57, 58] are two of the popular reduced order modeling (ROM) strategies available in the literature. Data analysis is conducted to extract basis functions from experimental data or detailed simulations of high-dimensional systems (method of snapshots [81, 82, 83]), for subsequent use in Galerkin projections that yield low dimensional dynamical models. While these type of models are physics-based and therefore require intrusive implementations, they are usually faster and more robust than data fitting and low-fidelity models.

Robustness of ROM in a parametric setting can be achieved by constructing a global basis [40, 68], but this strategy generates large dimensional bases that may lead to slow reduced order models. Local approaches have been designed for parametric or time domains generating local bases for both the state variables [70, 23] and non-linear terms [28, 66]. *Here we address the robustness issue of POD reduced order models by charting the parametric domain with feasible regions of local reduced order models where the reduced solutions are accurate within an admissible prescribed threshold.* Two essential ingredients are used to construct the parametric map: a database of reduced order models, and a data-fitting probabilistic model for the reduced order model errors. Then an incremental procedure uses the newly created probabilistic model to sample the parametric domain and generates a feasible region where a specific reduced order model provides accurate solutions within a prescribed tolerance. We then use a greedy approach to sweep the parameter domain and cover it with such feasible regions. This methodology is applied to the viscous 1D-Burgers model, and a parametric map for the viscosity parameter is generated for various error thresholds. Once the map is constructed there is no need to run the high-fidelity model again, since for each parameter value μ^0 there exists a parameter μ , and the associated reduced order model (basis and reduced operators), in whose interval the current value μ^0 falls; the corresponding reduced solution error is accurately estimated a-priori. The dimension k of the local basis is usually small since it depends only on one high-fidelity model trajectory.

The database of reduced order models and the high-fidelity trajectories are used to statistically model the reduced order approximation errors. We solve the resulting non-linear regression problems using Gaussian processes (GP) [84, 55] and artificial neural networks

(ANN). Specifically, consider the reduced order model of dimension k constructed using the high-fidelity solution computed with parameter value μ . Let ε be the error of this reduced model when approximating the full solution at parameter configuration μ^0 . We use a GP or ANN approach to model the mapping $\{\mu^0, \mu, k\} \rightarrow \log(\varepsilon)$. Our approach is inspired by the multifidelity correction and ROMES methodologies available in the literature for estimation of surrogate models errors using a global basis. Multifidelity correction [2, 29, 33, 44] has been developed for low-fidelity models in the context of optimization. They simulate the input-output relation $\mu \rightarrow \varepsilon$, where ε is the low-fidelity model errors. The ROMES method [25] introduces the concept of error indicators for reduced order models and generalizes the ‘multifidelity correction’ framework by approximating the mapping $\rho(\mu) \rightarrow \log(\varepsilon)$. The error indicators $\rho(\mu)$ include rigorous error bounds and reduced order residual norms, while ε is the reduced order model error at μ using a reduced global basis. By estimating the log of the reduced order model error instead of the error itself the probabilistic map exhibits a lower variance as shown by our numerical experiments as well as those in [25].

The size of a feasible region directly depends on the location of parameter value μ within the parametric space. *For overlapping feasible regions one can combine the available bases and the high-fidelity model trajectories in an efficient manner to generate more accurate reduced order models.* Three different approaches are proposed here, i.e. bases interpolation, bases concatenation, and high-fidelity model solution interpolation. Assuming a linear dependence we perform a Lagrangian interpolation of the bases in the matrix space [54], or interpolate their projections onto some local coordinate systems [54, 4]. Following the idea of the spanning ROM introduced in [93], for a new parameter not found in the ROM database, we create a projection basis either by concatenating two of the available bases that generated the higher accurate reduced order solutions for the new configurations, or by interpolating the associated high-fidelity solutions and then extracting the singular vectors.

Finally, *we address the issue of a-priori selection of the reduced basis dimension for a prescribed accuracy of the reduced solution.* The standard approach is to analyze the spectrum of the snapshots matrix, and use the largest singular value removed from the expansion to estimate the accuracy level [91]. To take into account the error due to the integration in the reduced space, here we use data fitting models to approximate the mapping $\{\mu, \log(\varepsilon)\} \rightarrow k$. Numerical results obtained using Gaussian processing and artificial neural networks are very promising.

The remainder of the paper is organized as follows. Section 2 reviews the reduced order modeling parametric framework. The problems solved herein are formulated in Section 3. Gaussian processes and artificial neural networks are reviewed in Section 4. The proposed techniques for generating reduced order bases for new parameter configurations are introduced in Section 5. Section 6 presents the details of the data fitting models, constructs the parametric map for the viscous 1D-Burgers model, and analyses the probabilistic model’s performance. Conclusions are drawn in Section 7.

2 Parametrized Reduced Order Modeling

Proper Orthogonal Decomposition has been successfully applied in numerous applications such as compressible flow [72], computational fluid dynamics [50, 73, 98], and aerodynamics [12], to mention a few. It can be thought of as a Galerkin approximation in the spatial variable built from functions corresponding to the solution of the physical system at specified time instances. A system reduction strategy for Galerkin models of fluid flows leading to dynamic models of lower order based on a partition in slow, dominant, and fast modes, has been proposed in [64]. Closure models and stabilization strategies for POD of turbulent flows have been investigated in [77, 95].

In this paper we consider discrete inner products (Euclidian dot product), though continuous products may be employed as well. Generally, an atmospheric or oceanic computer model is described by the following semi–discrete dynamical system:

$$\frac{d\mathbf{x}(\mu^0, t)}{dt} = \mathbf{F}(\mathbf{x}, t, \mu^0), \quad \mathbf{x}(\mu^0, 0) = \mathbf{x}_0 \in \mathbb{R}^{N_{\text{state}}}, \quad \mu^0 \in \tilde{\mathcal{P}}. \quad (1)$$

The input-parameter μ^0 typically characterizes the physical properties of the flow. For a given parameter configuration μ^0 we select an ensemble of N_t time instances of the flow $\mathbf{x}_{t_1}^{\mu^0}, \dots, \mathbf{x}_{t_{N_t}}^{\mu^0} \in \mathbb{R}^{N_{\text{state}}}$, where N_{state} is the total number of discrete model variables, and $N_t \in \mathbb{N}$, $N_t > 0$. The POD method chooses an orthonormal basis $U_{\mu^0} = [\mathbf{u}_i^{\mu^0}]_{i=1, \dots, k} \in \mathbb{R}^{N_{\text{state}} \times k}$, $k > 0$, such that the mean square error between $\mathbf{x}(\mu^0, t_i)$ and the POD expansion $\mathbf{x}_{\text{POD}}^{\mu^0}(t_i) = U_{\mu^0} \tilde{\mathbf{x}}(\mu^0, t_i)$, $\tilde{\mathbf{x}}(\mu^0, t_i) \in \mathbb{R}^k$, is minimized on average. The POD space dimension $k \ll N_{\text{state}}$ is appropriately chosen to capture the dynamics of the flow. Algorithm 1 describes the reduced order basis construction procedure [88].

Algorithm 1 POD basis construction

- 1: Compute the singular value decomposition for the snapshots matrix $[\mathbf{x}_{t_1}^{\mu^0} \dots \mathbf{x}_{t_{N_t}}^{\mu^0}] = \bar{U}_{\mu^0} \Sigma_{\mu^0} \bar{V}_{\mu^0}^T$, with the singular vectors matrix $\bar{U}_{\mu^0} = [\mathbf{u}_i^{\mu^0}]_{i=1, \dots, N_{\text{state}}}$.
 - 2: Using the singular-values $\lambda_1 \geq \lambda_2 \geq \dots \lambda_n \geq 0$ stored in the diagonal matrix Σ , define $I(m) = (\sum_{i=1}^m \lambda_i) / (\sum_{i=1}^{N_{\text{state}}} \lambda_i)$.
 - 3: Choose k , the dimension of the POD basis, such that $k = \min_m \{I(m) : I(m) \geq \gamma\}$ where $0 \leq \gamma \leq 1$ is the percentage of total information captured by the reduced space $\mathcal{X}^k = \text{range}(U_{\mu^0})$. Usually $\gamma = 0.99$.
-

Next, a Galerkin projection of the full model state and equations (2) onto the space \mathcal{X}^k spanned by the POD basis elements is used

to obtain the reduced order model:

$$\frac{d\tilde{\mathbf{x}}(\mu^0, t)}{dt} = U_{\mu^0}^T \mathbf{F}\left(U_{\mu^0} \tilde{\mathbf{x}}(\mu^0, t), t, \mu^0\right), \quad \tilde{\mathbf{x}}(\mu^0, 0) = U_{\mu^0}^T \mathbf{x}(0). \quad (2)$$

The efficiency of the POD-Galerkin technique is limited to linear or bilinear terms, since the projected nonlinear terms at every discrete time step still depend on the number of variables of the full model. In case of polynomial nonlinearities the tensorial POD technique [88] can be employed to efficiently remove the dependence on the full dimension by manipulating the order of computations. A considerable reduction in complexity is achieved by the Discrete Empirical Interpolation Method (DEIM) [19, 86], a discrete variation of Empirical Interpolation Method [8], for any type of nonlinear terms.

While being accurate for the given parameter configuration, the reduced model (2) loses accuracy when moving away from the initial setting. Several strategies have been proposed to derive a basis that spans the entire parameter space. These includes the reduced basis methods combined with the use of error estimates [74, 69, 68], global POD [89, 80], Krylov-based sampling methods [22, 94], and greedy techniques [36, 63]. The fundamental assumption used by these approaches is that a smooth low-dimensional global manifold characterizes the model solutions over the entire parameter domain. However, in order to ensure high accuracy of the reduced solution across the parameter space, the dimension of the reduced basis has to be increased in practice, leading to high computational costs. To alleviate this drawback we propose an alternative approach based on local parametric reduced order models.

3 Problem Description and Solution Methodology

This work addresses the following problems in the construction of reduced order models: designing the parametric map, selecting the best two already existing bases for a new parameter configuration from accuracy point of view, and selecting the optimal dimension of the reduced basis. We formulate them in detail below.

3.1 Designing the parametric map

Problem 1 (Efficient approximate ROMs). *For an arbitrary parameter configuration $\mu^0 \in \tilde{\mathcal{P}}$ construct a reduced order model (2) that provides an accurate and efficient approximation of the high-fidelity solution (1):*

$$\|\mathbf{x}(\mu^0, t_i) - U\tilde{\mathbf{x}}(\mu^0, t_i)\|_2 < \bar{\varepsilon}, \quad i = 1, \dots, N_t, \quad (3)$$

for some prescribed admissible error level $\bar{\varepsilon} > 0$. The snapshots used to generate the basis U can be obtained with any parametric configuration that belongs to $\tilde{\mathcal{P}}$.

A simple solution is to solve the high-fidelity model for the specific configuration μ^0 , and then build the corresponding reduced order model. However, this approach is computationally expensive.

Our methodology proposes to select a small countable subset $\mathcal{I} = \{\mu_j, j = 1, \dots, M\} \subset \tilde{\mathcal{P}}$, $M > 0$ and for each μ_j , a reduced order basis U_{μ_j} along with the reduced operators are constructed for $j = 1, \dots, M$. We denote by \mathcal{U} the set of bases U_{μ_j} , $j = 1, \dots, M$. Then for each $\mu^0 \in \tilde{\mathcal{P}}$ we can find a basis $U_{\mu_j} \in \mathcal{U}$ and the suitable reduced order model such that its solution satisfies (3) for $U = U_{\mu_j}$.

This strategy relies on the assumption that a reduced order basis and operators built for a specific parameter configuration $\mu_j \in \tilde{\mathcal{P}}$ can be used to design a reduced order model capable to accurately approximate the solution of the high-fidelity model (1) for all $\mu^0 \in \mathcal{B}(\mu_j, r_j) \cap \tilde{\mathcal{P}}$, where $\mathcal{B}(\mu_j, r_j)$ is the closed ball of radius $r_j \geq 0$ centered at μ_j . Specifically, for $\mu^0 \in \mathcal{B}(\mu_j, r_j) \cap \tilde{\mathcal{P}}$, the reduced order model is constructed by employing the basis U_{μ_j} and the reduced operators designed at μ_j , i.e.

$$\frac{d\tilde{\mathbf{x}}(\mu^0, \mu_j, t)}{dt} = U_{\mu_j}^T \mathbf{F}\left(U_{\mu_j} \tilde{\mathbf{x}}(\mu^0, \mu_j, t), t, \mu^0\right), \quad \tilde{\mathbf{x}}(\mu^0, 0) = U_{\mu_j}^T \mathbf{x}(0). \quad (4)$$

Then, by selecting a small radius r_j , one can be able to obtain

$$\|\mathbf{x}(\mu^0, t_i) - U_{\mu_j} \tilde{\mathbf{x}}(\mu^0, \mu_j, t_i)\|_2 < \bar{\varepsilon}, \quad i = 1, \dots, N_t, \quad (5)$$

for all $\mu^0 \in \mathcal{B}(\mu_j, r) \cap \tilde{\mathcal{P}}$.

The parametric map construction process ends as soon as the entire parameter domain $\tilde{\mathcal{P}}$ is covered with a finite union of overlapping balls $\mathcal{B}(\mu_j, r_j)$, $j = 1, \dots, M$, corresponding to different reduced order bases and local models

$$\tilde{\mathcal{P}} \subset \bigcup_{j=1}^M \mathcal{B}(\mu_j, r_j), \quad (6)$$

such that for each $j = 1, 2, \dots, M$ and $\forall \mu^0 \in \mathcal{B}(\mu_j, r) \cap \tilde{\mathcal{P}}$, the solution of the reduced order model (4) depending on the basis U_{μ_j} fulfils the accuracy condition (5).

This approach is inspired from the construction of local reduced order models where the time domain is split in multiple regions [70, 66]. In this way the reduced basis dimension is kept small allowing for fast on-line simulations. The cardinality of \mathcal{I} depends inversely proportional with the prescribed level of accuracy $\bar{\varepsilon}$. As the desired error threshold $\bar{\varepsilon}$ decreases, the map changes since usually the radii r_j are expected to become smaller, and more balls are required to cover the parametric domain, i.e. M is increased.

The construction of the parametric map (6) using the local reduced order models requires the following ingredients:

1. The ability to probe the vicinity of $\mu_j \in \tilde{\mathcal{P}}$ and to efficiently estimate the level of error

$$\varepsilon = \max_{i=1, \dots, N_t} \|\mathbf{x}(\mu^0, t_i) - U_{\mu_j} \tilde{\mathbf{x}}(\mu^0, \mu_j, t_i)\|_2, \quad (7)$$

2. The ability to find $r_j > 0$ such that $\varepsilon \leq \bar{\varepsilon}$ for all $\mu^0 \in \mathcal{B}(\mu_j, r) \cap \tilde{\mathcal{P}}$. This can be theoretically achieved by assuming that the error ε is monotonically increasing with larger distances $d(\mu^0, \mu_j)$. However, this is not necessarily true and in practice this is obtained by sampling.
3. The ability to identify the location of a new μ_ℓ (for the construction of a new local reduced order model) given the locations of the previous local parameters $\mu_j, j = 1, \dots, \ell - 1$, so that

$$\mathcal{B}(\mu_\ell, r_\ell) \not\subset \left(\bigcup_{i=1}^{\ell-1} \mathcal{B}(\mu_i, r_i) \right), \quad \mathcal{B}(\mu_\ell, r_\ell) \cap \left(\bigcup_{i=1}^{\ell-1} \mathcal{B}(\mu_i, r_i) \right) \neq \emptyset. \quad (8)$$

The implementation of the first ingredient does not employ a-posteriori error estimation formulas [63]. Inspired from the 'multi-fidelity correction' and ROMES methodologies we construct probabilistic models to approximate the level of error ε in (7). Gaussian process and artificial neural networks are used to build the probabilistic functions to model the mapping $(\mu^0, \mu_j, k) \rightarrow \log(\varepsilon)$. Since the dimension of basis determines the level of error we include it among the input features. To design probabilistic models with reduced variances we look to approximate the logarithm of the error as suggested in [25].

The above machine learning techniques allow to sample the vicinity of μ_j and estimate the error for each sample parameter value. Based on these error estimates we construct the ball $\mathcal{B}(\mu_j, r)$, or perhaps a larger set called a μ_j -feasible region, where the local reduced order model is accurate within the prescribed threshold $\bar{\varepsilon}$.

Next, a greedy algorithm is applied to identify the location of a new parametric configuration μ_ℓ (for the construction of a new basis) depending on the locations of the previous $\mu_i, i = 1, \dots, \ell - 1$. Constraint (8) is imposed so the entire parametric domain $\tilde{\mathcal{P}}$ satisfies (6) after the map construction is finished.

For the parametric area situated at the intersection of different feasible regions we can assign a new reduced order model based on the information required to construct the initial feasible regions. This is achieved by interpolation or concatenation of the underlying reduced bases or interpolation of the available high-fidelity solutions, as described in detail in Section 5.

3.2 Selecting the best two already existing bases for a new parameter configuration

Since the error of the reduced order solution at a new parameter location μ^0 is not necessarily smaller with the decrease of the distance $d(\mu^0, \mu_j)$, the following more general problem is posed.

Problem 2 (Selection of best bases). *For a new parameter configuration μ^0 find the best available bases (among the existing ones) that provide the most accurate reduced order model solution.*

The capability of the already proposed probability models can be used to estimate the error

$$\varepsilon = \|\mathbf{x}(\mu^0, t_i) - U_{\mu_j} \tilde{\mathbf{x}}(\mu^0, \mu_j, t_i)\|_2, \quad i = 1, \dots, N_t, \quad (9)$$

for all available $U_{\mu_j}, j = 1, 2, \dots$ in the database, and the bases that lead to the smallest estimated errors are selected. This approach is discussed in Section 6.3.

3.3 Selecting the dimension of the reduced basis

Problem 3 (Optimal basis dimension). *Find the optimal dimension of the basis U_μ for the parametric configuration μ such that the error is smaller than the prescribed threshold*

$$\|\mathbf{x}(\mu, t_i) - U_\mu \tilde{\mathbf{x}}(\mu, \mu, t_i)\|_2 \leq \bar{\varepsilon}, \quad i = 1, \dots, N_t. \quad (10)$$

By optimal we understand the smallest dimension that enables the reduced order model solution to satisfy the error constraint (10). The basis dimension represents one of the most important characteristics of a reduced order model. The reduced manifold size directly affects both the on-line computational complexity of the reduced order model and its accuracy [51, 39, 31]. By increasing the size of the basis the projection error decreases and the accuracy of the reduced order model is enhanced. Consequently, the spectrum of the

snapshots matrix offers guidance regarding the choice of the reduced basis size when some prescribed reduced order model error is desired. However the accuracy depends also on integration errors in the case of unsteady models as stated in [41].

We seek to estimate the optimal size of the reduced order model by accounting for both the projection and integration errors. For this we use data fitting models to approximate the mapping $\{\mu, \log \bar{\varepsilon}\} \rightarrow k$, as explained in Section 6.5.

For all the problems addressed in this study a general probabilistic framework is introduced in Section 4, along with the description of the supervised machine learning techniques used to construct the discussed probabilistic models. For each problem we discuss the dataset, the features, as well as the accuracy and stability of the predictions of the associated data fitting models.

4 Supervised Machine Learning Techniques

Consider a random vector \mathbf{z} . Neural networks and Gaussian processes are used to build a probabilistic model $\phi : \mathbf{z} \rightarrow \hat{y}$, where ϕ is a transformation function that learns through the input feature vector $\mathbf{z} \in \Omega$ (the sample space) to estimate the deterministic output $y \in \mathbb{R}$ [59]. The estimator \hat{y} is expected to have a low variance. The features of \mathbf{z} should be descriptive of the underlying problem at hand [11]. The accuracy and stability of estimations are assessed using the K -fold cross validation technique. The samples are split into K subsets (“folds”), where typically $3 \leq K \leq 10$. The machine is trained on $K - 1$ sets and tested on the K -th set in a round-robin fashion [59]. Each fold induces a specific error quantified as the average of the absolute values of the differences between the predicted and the K -th set values.

$$E_{\text{fold}} = \frac{\sum_{i=1}^N |\hat{y}_i - y_i|}{N}, \quad \text{VAR}_{\text{fold}} = \frac{\sum_{i=1}^N (\hat{y}_i - E_{\text{fold}})^2}{N}, \quad (11a)$$

where N is the number of test samples in the fold. The error is then averaged over all folds:

$$E = \frac{\sum_{\text{fold}=1}^K E_{\text{fold}}}{K}, \quad \text{VAR} = \frac{\sum_{\text{fold}=1}^K (E_{\text{fold}} - E)^2}{K}. \quad (11b)$$

The variance of the prediction results (11a) accounts for the sensitivity of the model to the particular choice of data set. It quantifies the stability of the model in response to the new training samples. A smaller variance indicates more stable predictions, however, this sometimes translates into a larger bias of the model. Models with small variance and high bias make strong assumptions about the data and tend to underfit the truth, while models with high variance and low bias tend to overfit the truth [62]. The trade-off between bias and variance in learning algorithms is usually controlled via techniques such as regularization or bagging and boosting [11].

In what follows we briefly review the Gaussian process and Neural network techniques.

4.1 Gaussian process kernel method

A Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution [71]. A Gaussian process is fully described by its mean and covariance functions

$$\phi(\mathbf{z}) \sim \text{gp}(m(\mathbf{z}), k(\mathbf{z}_i, \mathbf{z}_j)), \quad (12)$$

where [71]

$$m(\mathbf{z}) = \mathbb{E}[\phi(\mathbf{z})], \quad k(\mathbf{z}_i, \mathbf{z}_j) = \mathbb{E}[(\phi(\mathbf{z}_i) - m(\mathbf{z}_i))(\phi(\mathbf{z}_j) - m(\mathbf{z}_j))].$$

In this work we employ the commonly used squared-exponential-covariance Gaussian kernel with [71]:

$$k(\mathbf{z}_i, \mathbf{z}_j) = \sigma_\phi^2 \exp\left(-\frac{(\mathbf{z}_i - \mathbf{z}_j)^2}{2\ell^2}\right) + \sigma_n^2 \delta_{i,j}, \quad (13)$$

where \mathbf{z}_i and \mathbf{z}_j are the pairs of data points in training or test samples, and δ is the Kronecker delta symbol. The model (13) has three hyper-parameters. The length-scale ℓ governs the correlation among data points. The signal variance σ_ϕ^2 and the noise variance σ_n^2 govern the precision of variance and noise, respectively.

Consider a set of training data points $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n$, and the corresponding noisy observations y_1, y_2, \dots, y_n ,

$$y_i = \phi(\mathbf{z}_i) + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma_n^2), \quad i = 1, \dots, n. \quad (14)$$

Consider also the set of test points $\mathbf{z}_1^*, \mathbf{z}_2^*, \dots, \mathbf{z}_m^*$ and the predictions $\hat{y}_1, \hat{y}_2, \dots, \hat{y}_m$

$$\hat{y}_i = \phi(\mathbf{z}_i^*), \quad i = 1, \dots, m. \quad (15)$$

For a Gaussian prior the joint distribution of training outputs y and test outputs \hat{y} is:

$$\begin{bmatrix} y \\ \hat{y} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(\mathbf{z}) \\ m(\mathbf{z}^*) \end{bmatrix}, \begin{bmatrix} k(\mathbf{z}, \mathbf{z}) + \sigma_n^2 I & k(\mathbf{z}, \mathbf{z}^*) \\ k(\mathbf{z}^*, \mathbf{z}) & k(\mathbf{z}^*, \mathbf{z}^*) \end{bmatrix}\right) \quad (16)$$

The predictive distribution represents the posterior after observing the data [11] and is given by:

$$p(\hat{y}|\mathbf{z}, y, \mathbf{z}^*) \sim \mathcal{N}\left(k(\mathbf{z}^*, \mathbf{z}) (k(\mathbf{z}, \mathbf{z}) + \sigma_n^2 I)^{-1} y, k(\mathbf{z}^*, \mathbf{z}^*) - k(\mathbf{z}^*, \mathbf{z}) (k(\mathbf{z}, \mathbf{z}) + \sigma_n^2 I)^{-1} k(\mathbf{z}, \mathbf{z}^*)\right). \quad (17)$$

The prediction of Gaussian process will depend on the choice of the mean and covariance functions, and on their hyperparameters. The hyperparameters can be inferred from the data by minimizing the marginal negative log-likelihood function $\theta = \arg \min L(\theta)$, where

$$L(\theta) = -\log p(y|\mathbf{z}, \theta) = \frac{1}{2} \log |k(\mathbf{z}, \mathbf{z})| + \frac{1}{2} (y - m(\mathbf{z}))^T k(\mathbf{z}, \mathbf{z})^{-1} (y - m(\mathbf{z})) + \frac{n}{2} \log(2\pi).$$

4.2 Artificial Neural networks

The study of artificial neural networks (ANNs) begin in the 1910s in order to intimate human brain's biological structure. Pioneering work was carried out by Rosenblatt, who proposed a three-layered network structure, the perceptron [37]. ANNs detect the pattern of data by discovering the input–output relationships. Applications include the approximation of functions, regression analysis, time series prediction, pattern recognition, and speech synthesis and recognition [46, 7].

The architecture of ANNs is schematically represented in Figure 1. ANNs consist of neurons and connections between the neurons (weights). Neurons are organized in layers, where at least three layers of neurons (an input layer, a hidden layer, and an output layer) are required for construction of a neural network. The input layer distributes input signals $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n$ to the hidden layer. For a neural network with L hidden layers and m neurons in the hidden layer, let \hat{y}_j be the vector of outputs from layer ℓ , b^ℓ the biases at layer ℓ , and w_{kj}^ℓ the weight connecting the neuron j to the k th input. Then the feed-forward operation is:

$$\begin{aligned} \mathbf{x}_j^{\ell+1} &= \sum_{k=1} w_{kj}^{\ell+1} \hat{y}_k^\ell + b_j^{\ell+1}, \quad \hat{y}_j^0 = \mathbf{z}_j, \quad j = 1, \dots, m \\ \hat{y}_j^{\ell+1} &= \phi(\mathbf{x}_j^{\ell+1}), \quad \ell = 0, 1, \dots, L-1 \end{aligned}$$

The differentiable function ϕ is the transfer function and can be log-sigmoid, hyperbolic tangent sigmoid, or linear transfer function.

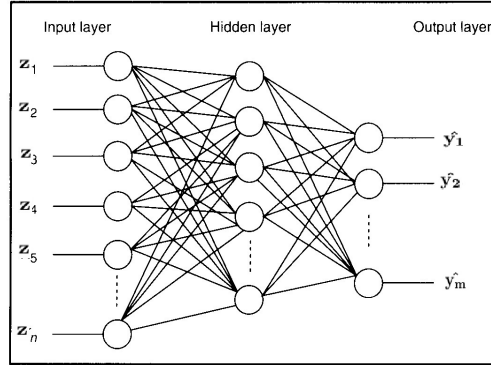


Figure 1: Schematic representation of the neural network structure.

The training process of ANN adjusts the weights $w_{i,j}$ in order to reproduce the desired outputs when fed the given inputs. The training process via the back propagation algorithm [75] uses a gradient descent method to modify weights and thresholds such that the error between the desired output and the output signal of the network is minimized [32]. In supervised learning the network is provided with samples from which it discovers the relations of inputs and outputs. The output of the network is compared with the desired output, and the error is back-propagated through the network and the weights will be adjusted. This process is repeated during several epochs, until the network output is close to the desired output [38]. In unsupervised learning a model of the data distribution is formed in order to extract significant data features. The network extracts data features without being shown a set of inputs and outputs [46].

5 Combining Available Information for Accurate ROMs at New Parametric Configurations

The POD method produces an orthogonal basis that approximately spans the state solution space of the model for a specific parameter configuration. Moving away from the initial parametric configuration requires the construction of new bases since the initial reduced order model is not accurate anymore. However, if states depend continuously on parameters, the POD basis constructed for one parameter configuration can approximately span the solution space at different parametric settings in a local vicinity.

Several methods to combine the available information to generate more accurate reduced order models for new parameter configurations have been proposed in the literature. One is the interpolation of the available reduced order bases U_{μ_j} computed for the parametric

configurations μ_j , $j = 1, \dots, M$. The dependence of the bases on the parameters has been modeled with various linear and nonlinear spatially-dependent interpolants.

Here we compare the performances of different strategies that involve Lagrange interpolation of bases in the matrix space and in the tangent space of the Grassmann manifold. In addition we propose to concatenate the available reduced bases followed by an orthogonalization process, and to interpolate the solutions of the high fidelity model as a mean to derive the reduced order basis for a new parameter configuration.

5.1 Basis interpolation

Lagrange interpolation of bases Assuming the reduced manifold $\mathbf{U} : \tilde{\mathcal{P}} \rightarrow \mathbb{R}^{N_{state} \times k}$ poses a continuous and linear dependency with respect to the parametric space, and if M discrete bases $U_{\mu_j} = \mathbf{U}(\mu_j)$ have been already constructed for various parametric configurations μ_j , $j = 1, 2, \dots, M$, then a basis corresponding to the new configuration μ^0 can be obtained using Lagrange's interpolation formula

$$U_{\mu^0} = \sum_{j=1}^M U_{\mu_j} L_j(\mu^0), \quad L_j(\mu) = \prod_{i \neq j} \frac{\mu - \mu_i}{\mu_j - \mu_i}. \quad (18)$$

Drawbacks of this approach include the orthogonalization requirement for the resulting interpolated basis vectors, and the lack of linear variation in the angles between pairs of subspace planes [54] spanned by the reduced bases U_{μ_j} . Differential geometry results can be employed to alleviate these deficiencies.

Grassmann manifold In the study proposed by Amsallem and Farhat [4] basis (matrix) interpolation in the tangent space of the Grassmann manifold at a careful selected point S representing a subspace spanned by one of the available reduced bases was performed. It has been shown that Grassmann manifold can be endowed with a differentiable structure [1, 27], i.e., at each point S of the manifold a tangent space exists. The mapping from the manifold to the tangent space at S is called the logarithmic mapping, while the backward projection is referred to as exponential mapping [10]. According to [4] the construction of a new subspace S_{μ^0} associated with a new parameter μ^0 can be obtained by interpolating the known subspaces $\{S_i\}_{i=1}^M$ spanned by the already computed reduced bases U_{μ_i} , $i = 1, \dots, M$. The steps required by this methodology [4] are described in the Algorithm2.

Algorithm 2 Interpolation in a tangent space to a Grassmann manifold algorithm [4]

- 1: Select a point S_{i_0} of the manifold to represent the origin point for the interpolation spanned by the basis $U_{\mu_{i_0}}$.
- 2: The tangent space $\mathcal{T}_{S_{i_0}}$ and the subspaces $\{S_i\}_{i=1}^M$ are considered. Each point S_i sufficiently close to S_{i_0} is mapped to a matrix Γ_i representing a point of $\mathcal{T}_{S_{i_0}}$ using the logarithm map $\text{Log}_{S_{i_0}}$ [10]

$$\begin{aligned} (I - U_{\mu_{i_0}} U_{\mu_{i_0}}^T) U_{\mu_i} (U_{\mu_{i_0}} U_{\mu_i})^{-1} &= R_i \Lambda_i Q_i^T \quad (\text{SVD factorization}), \\ \Gamma_i &= R_i \tan^{-1}(\Lambda_i) Q_i^T. \end{aligned}$$

- 3: Each entry of the matrix Γ_0 associated with the target parameter μ_0 is computed by interpolating the corresponding entries of the matrices $\Gamma_i \in \mathbb{R}^{N_{state} \times k}$ associated with the parameter points μ_i , $i = 1, \dots, M - 1$. A univariate or multivariate Lagrange interpolation may be chosen similar with the one introduced in (18).
- 4: The matrix Γ^0 representing a point in the tangent space $\mathcal{T}_{S_{i_0}}$ is mapped to a subspace S^0 on the Grassmann manifold spanned by a matrix U_{μ^0} using the exponential map [10]

$$\begin{aligned} \Gamma^0 &= R^0 \Lambda^0 Q^{0T} \quad (\text{SVD factorization}), \\ U_{\mu^0} &= U_{\mu_{i_0}} Q^0 \cos(\Lambda^0) + R^0 \sin(\Lambda^0). \end{aligned}$$

Amsallem and Cortial [3] proved that the subspace angle interpolation [54, 53] is identical to the interpolation in a tangent space to the Grassmann manifold of two reduced-order bases, thus the latter methodology can be viewed as a generalization of the former approach.

5.2 Basis concatenation

Basis concatenation idea was introduced in [93] and emerged from the notion of a global basis [89, 80]. In the global strategy, the existent high-fidelity snapshots corresponding to various parameter configurations are collected in a single snapshot matrix and then a matrix factorization is performed to extract the most energetic singular vectors. This global basis is then used to build reduced order models for parameter values not available in the initial snapshots set.

Assuming $\mathbf{x}_{\mu_1}, \mathbf{x}_{\mu_2} \in \mathbb{R}^{N_{\text{state}} \times N_t}$ are the snapshots corresponding to two high-fidelity model trajectories, the following error estimate holds [91, Proposition 2.16]:

$$\bar{\mathbf{x}} = [\mathbf{x}_{\mu_1} \mathbf{x}_{\mu_2}] = \bar{U} \Lambda \bar{V}^T, \quad (\text{SVD factorization}) \quad (19a)$$

$$\|\bar{\mathbf{x}} - \bar{U}(:, 1 : k) \bar{\mathbf{x}}\|_F = \sum_{i=k+1}^{N_t} \lambda_i = \mathcal{O}(\lambda_{k+1}), \quad (19b)$$

where λ_i is the i^{th} singular value of $\bar{\mathbf{x}}$, $\bar{U}(:, 1 : k)$ are the first k singular vectors of \bar{U} and $\bar{\mathbf{x}} \in \mathbb{R}^{k \times 2N_t}$. By $\|\cdot\|_F$ we refer to the Frobenius norm. The snapshot matrix typically stores correlated data and therefore contains linearly dependent columns. For rank deficient snapshot matrices, and in the case where the reduced order bases U_{μ_1} and U_{μ_2} corresponding to the trajectories μ_1 and μ_2 are available, we can construct a basis \hat{U} by simply concatenating columns of U_{μ_1} and U_{μ_2} such that the accuracy level in (19) is preserved.

Proposition 5.1. Consider the following SVD expansions of rank deficient snapshots matrices $\mathbf{x}_{\mu_1}, \mathbf{x}_{\mu_2} \in \mathbb{R}^{N_{\text{state}} \times N_t}$

$$\mathbf{x}_{\mu_j} = U_{\mu_j} \Lambda_j V_{\mu_j}^T, \quad j = 1, 2. \quad (20)$$

There are positive integers k_1, k_2 , and $\hat{\mathbf{x}} \in \mathbb{R}^{(k_1+k_2) \times 2N_t}$, such that $\bar{\mathbf{x}}$ defined in (19) satisfies

$$\|\bar{\mathbf{x}} - \hat{U} \hat{\mathbf{x}}\|_F = \mathcal{O}(\lambda_{k+1}), \quad (21)$$

where λ_{k+1} is the $(k+1)$ -st singular value of snapshots matrix $\bar{\mathbf{x}}$, and $\hat{U} = [U_{\mu_1}(:, 1 : k_1) \ U_{\mu_2}(:, 1 : k_2)] \in \mathbb{R}^{N_{\text{state}} \times (k_1+k_2)}$.

Proof. Since $\mathbf{x}_{\mu_1}, \mathbf{x}_{\mu_2} \in \mathbb{R}^{N_{\text{state}} \times N_t}$ are rank deficient matrices, there exist at least two positive integers k_1 and k_2 , such that the singular values associated with \mathbf{x}_{μ_1} and \mathbf{x}_{μ_2} satisfy

$$\lambda_{k_1+1}^1, \lambda_{k_2+1}^2 \leq \frac{\lambda_{k+1}}{2}, \quad \forall k = 0, \dots, N_t - 1. \quad (22)$$

Next, from [91, Proposition 2.16] and (20) we have the following estimates:

$$\|\mathbf{x}_{\mu_j} - U_{\mu_j}(:, 1 : k_j) \bar{\mathbf{x}}_j\|_F = \sum_{i=k_j+1}^{N_t} \lambda_i^j = \mathcal{O}(\lambda_{k_j+1}^j), \quad (23)$$

where λ_i^j is the i^{th} singular value of \mathbf{x}_{μ_j} and $\bar{\mathbf{x}}_j \in \mathbb{R}^{k_j \times N_t}$, for $j = 1, 2$.

By denoting

$$\hat{\mathbf{x}} = \begin{bmatrix} \bar{\mathbf{x}}_1 & \mathbf{0}_1 \\ \mathbf{0}_2 & \bar{\mathbf{x}}_2 \end{bmatrix},$$

where the null matrix $\mathbf{0}_j$ belongs to $\mathbb{R}^{k_j \times N_t}$, $j = 1, 2$, we have

$$\|\bar{\mathbf{x}} - \hat{U} \hat{\mathbf{x}}\|_F = \|[\mathbf{x}_{\mu_1} \ \mathbf{x}_{\mu_2}] - [U_{\mu_1}(:, 1 : k_1) \bar{\mathbf{x}}_1 \ U_{\mu_2}(:, 1 : k_2) \bar{\mathbf{x}}_2]\|_F \leq \quad (24)$$

$$\|\bar{\mathbf{x}}_{\mu_1} - U_{\mu_1}(:, 1 : k_1) \bar{\mathbf{x}}_1\|_F + \|\bar{\mathbf{x}}_{\mu_2} - U_{\mu_2}(:, 1 : k_2) \bar{\mathbf{x}}_2\|_F \leq \mathcal{O}(\lambda_{k_1+1}^1) + \mathcal{O}(\lambda_{k_2+1}^2) = \mathcal{O}(\lambda_{k+1}). \quad (25)$$

□

It is crucial for the proof that U_{μ_1} and U_{μ_2} are rank deficient since they have at least one null singular value. In practice usually $k_1 + k_2$ is larger than k thus more bases functions are required to form \hat{U} to achieve a similar level of precision as in (19) where \bar{U} is built using a global singular value decomposition. This matrix factorization of $\bar{\mathbf{x}}$ is more costly than both singular value decompositions of matrices \mathbf{x}_{μ_j} , $j = 1, 2$, (23) combined for large space dimension N_{state} . However the off-line stage of the concatenation method also includes the application of a Gramm-Schmidt-type algorithm to orthogonalize the overall set of vectors in \hat{U} .

For full rank matrices the precision is controlled by the spectra of snapshots matrices U_{μ_1} and U_{μ_2} but there is no guarantee that the concatenated basis \hat{U} can provide similar accuracy precision as \bar{U} in (19) for all $k = 1, 2, \dots, N_t$.

While the Lagrange interpolation of bases mixes the different energetic singular vectors in an order dictated by the singular values magnitude, this strategy concatenates the dominant singular vectors for each case and preserves their structure.

5.3 Interpolation of high-fidelity model solutions

The method discussed herein assumes that the model solution depends continuously on the parameters. Thus it is natural to consider constructing the basis for a new parameter configuration by interpolating the existent high-fidelity model solutions associated with

various parameter settings, and then performing a SVD factorization of the interpolated results. For example, the Lagrange solution interpolant is given by

$$\mathbf{x}_{\mu^0} = \sum_{j=1}^M \mathbf{x}_{\mu_j} L_j(\mu^0), \quad (26)$$

where $\mathbf{x}_{\mu_j} \in \mathbb{R}^{N_{\text{state}} \times N_t}$ is the model solution corresponding to parameter μ_j and the interpolation polynomials are defined in (18).

A new basis is constructed from the interpolated model solution (26) for the new parametric configuration μ_0 . By integrating the corresponding reduced order model the output projected solution will present variations in comparison with the high-fidelity interpolation solution \mathbf{x}_{μ^0} thus the nonlinear dynamics of the model influence the final solution too.

From computational point of view the complexity of the off-line stage of the solution interpolation method (26) is smaller than in the case of the bases concatenation and interpolation approaches. Only one singular value decomposition is required in contrast with the multiple factorizations needed in the latter two strategies where the involved matrices have the same size $N_{\text{state}} \times N_t$. Having only N_t snapshots the size of the outcome basis should be smaller than in the case of basis concatenation approach.

If for each model solution \mathbf{x}_{μ_j} a singular value decomposition is available such that

$$\mathbf{x}_{\mu_j} \approx U_{\mu_j} \tilde{\mathbf{x}}_j, \quad j = 1, 2, \dots, M, \quad (27)$$

then from (26) we get

$$\mathbf{x}_{\mu^0} \approx \sum_{j=1}^M L_j(\mu^0) U_{\mu_j} \tilde{\mathbf{x}}_j. \quad (28)$$

Now the basis U_{μ_0} associated with the new configuration is obtained following a matrix factorization, i.e..

$$\sum_{j=1}^M L_j(\mu^0) U_{\mu_j} \tilde{\mathbf{x}}_j = U_{\mu_0} S_{\mu_0} V_{\mu_0}^T. \quad (29)$$

In the basis interpolation method (18), the basis U_{μ_0} (denoted in (30) by \bar{U}_{μ_0} to differentiate it from the basis described in equation (29)) corresponding to the new parametric setting μ_0 is derived from the following

$$\sum_{j=1}^M L_j(\mu^0) U_{\mu_j} = \bar{U}_{\mu_0} \bar{S}_{\mu_0} \bar{V}_{\mu_0}^T. \quad (30)$$

While there is a close relationship between the philosophy behind the solution interpolation method and the bases interpolation strategy, the algebraic formulations of the bases U_{μ_0} (29) and \bar{U}_{μ_0} (30) are significantly different. Consequently an assumption of linear variation of the basis over the parametric interval does not imply that the high-fidelity solution varies linear over the parametric domain. The inverse proposition does not hold neither thus the choice of method depends on the model under study and its input.

6 Numerical Experiments

We illustrate the application of the proposed machine learning methodologies to the construction of error models for the reduced order models solutions for a one-dimensional Burgers model and their subsequent utilizations. The model proposed herein is characterized by two scalar parameters, but the envisioned parametric map is designed to cover variation in the viscosity coefficient space only. For each of the problems introduced in Section 3 (constructing the parametric map, selecting the best two already existing bases for a new parameter configuration from accuracy point of view, and selecting the dimension of the reduced basis) we present in detail the proposed solution approaches and the corresponding numerical results.

To assess the performance of probabilistic models we employ various cross validation tests. The dimensions of the training and testing data sets are chosen empirically based on the number of samples. For artificial neural network models the number of hidden layers and neurons in each hidden layer vary for each type of problems under study. The squared-exponential-covariance kernel (13) is used for Gaussian process models.

6.1 One-dimensional Burgers' equation

Burgers' equation is an important partial differential equation from fluid mechanics [14]. The evolution of the velocity u of a fluid evolves according to

$$\frac{\partial u}{\partial t} + \nu u \frac{\partial u}{\partial x} = \mu \frac{\partial^2 u}{\partial x^2}, \quad x \in [0, L], \quad t \in (0, t_f], \quad (31)$$

with $t_f = 1$ and $L = 1$. Here μ is the viscosity coefficient. The parameter ν has no physical significance and it is used to control the non-linear effect of the advection term. The model has homogeneous Dirichlet boundary conditions $u(0, t) = u(L, t) = 0$, $t \in (0, t_f]$. A smooth initial condition is used, described by a seventh degree polynomial and shown in Figure 2.

The discretization uses a spatial mesh of $N_s = 201$ equidistant points on $[0, L]$, with $\Delta x = L/(N_s - 1)$. A uniform temporal mesh with $N_t = 301$ points covers the interval $[0, t_f]$, with $\Delta t = t_f/(N_t - 1)$. The discrete velocity vector is $\mathbf{u}(t_j) \approx [u(x_i, t_j)]_{i=1,2,\dots,N_{state}} \in \mathbb{R}^{N_{state}}$, $N = j = 1, 2, \dots, N_t$, where $N_{state} = N_s - 2$ (the known boundaries are removed). The semi-discrete version of the model (31) is:

$$\mathbf{u}' = -\nu \mathbf{u} \odot A_x \mathbf{u} + \mu A_{xx} \mathbf{u}, \tag{32}$$

where \mathbf{u}' is the time derivative of \mathbf{u} , and $A_x, A_{xx} \in \mathbb{R}^{N_{state} \times N_{state}}$ are the central difference first-order and second-order space derivative operators, respectively, which take into account the boundary conditions too. The model is implemented in Matlab and the backward Euler method is employed for time discretization. The nonlinear algebraic systems are solved using Newton-Raphson method and the allowed number of Newton iterations per each time step is set to 50. The solution is considered accurate enough when the euclidian norm of the residual is less then 10^{-10} .

The viscosity parameter space is set to the interval $[0.01, 1]$. Smaller values of μ correspond to sharper gradients in the solution, and lead to a dynamics that is more difficult to approximate by a reduced order model.

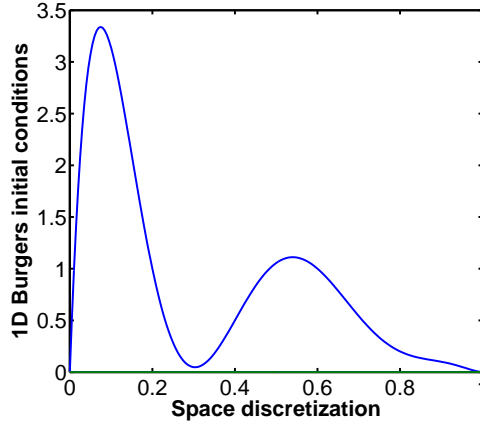


Figure 2: Seventh order polynomial used as initial conditions for 1D Burgers model.

6.2 Designing the parametric map

We seek to build a parameter map for the 1D-Burgers model for the viscosity domain consisting in the interval $[0.01, 1]$. The non-physical parameter ν is set to 1. As discussed in Section 3, we take the following steps. First we construct probabilistic models to approximate the error of reduced order solution. Next, we identify “ μ -feasible” intervals $[d_\ell, d_r]$ in the parameter space such that local reduced order model depending only on the high-fidelity trajectory at μ is accurate within the prescribed threshold for any $\mu^0 \in [d_\ell, d_r]$. Finally, a greedy algorithm generates the parametric map by covering the parameter space with a union of μ_i feasible intervals.

$$[0.01, 1] \subset \bigcup_{i=1}^M [d_\ell^i, d_r^i], \tag{33}$$

where each μ_i -feasible interval is characterized by an error threshold $\bar{\varepsilon}_i$ (which can vary from one interval to another). This relaxation is suggested since for intervals associated with small parameters μ , it is difficult to achieve small reduced order models errors similar to those obtained for larger parametric configurations. One way to maintain the error thresholds constant is to start with a larger proposal.

In existing reduced basis methods a global reduced order model depending on multiple high-fidelity trajectories is constructed. In contrast, our approach decomposes the parameter space into smaller regions where the local reduced order model solutions are accurate within some tolerance levels. Since the local bases required for the construction of the local reduced order models depend on only a single full simulation, the size of the reduced manifolds is small, leading to lower on-line computational complexity.

6.2.1 Error estimation of ROM solutions

We first construct probabilistic models

$$\phi : \mathbf{z} \rightarrow \hat{\varepsilon}, \tag{34}$$

where the input features \mathbf{z} include a new viscosity parameter value μ^0 , a parameter value μ associated with the full model run that generated the basis U_μ , and the dimension of the reduced manifold. The target $\hat{\varepsilon}$ is the estimated error of the reduced order model solution at μ^0 using the basis U_μ and the corresponding reduced operators computed using the Frobenius norm

$$\varepsilon = \|[x(\mu^0, t_i)]_{i=1,\dots,N_t} - U_\mu [\tilde{\mathbf{x}}(\mu^0, \mu, t_i)]_{i=1,\dots,N_t}\|_F. \tag{35}$$

The training data set includes equally distributed values of μ and μ^0 over the entire interval, $\mu \in \{0.1, 0.2, \dots, 0.9, 1\}$ and $\mu^0 \in \{0.01, 0.02, \dots, 0.99, 1\}$, respectively, reduced basis dimensions spanning the interval $[4, 5, \dots, 14, 15]$ and the reduced order model error ε . The entire training data set contains nearly 12,000 samples, and for each sample a high-fidelity model solution is calculated. Figure 3(b) shows isocontours of the reduced order model error ε for viscosity parameter values μ^0 and various POD basis dimensions. The design of the reduced order models relies on the high-fidelity trajectory for $\mu = 0.8$. Since target values ε vary over a wide range (from 300 to 10^{-6}) we consider the logarithms of the errors $\log(\varepsilon)$ to decrease the variance of the predicted results, i.e.

$$\phi : \mathbf{z} \rightarrow \widehat{\log(\varepsilon)}. \tag{36}$$

Figure 3(a) shows isocontours for the logarithms of the errors $\log(\varepsilon)$.

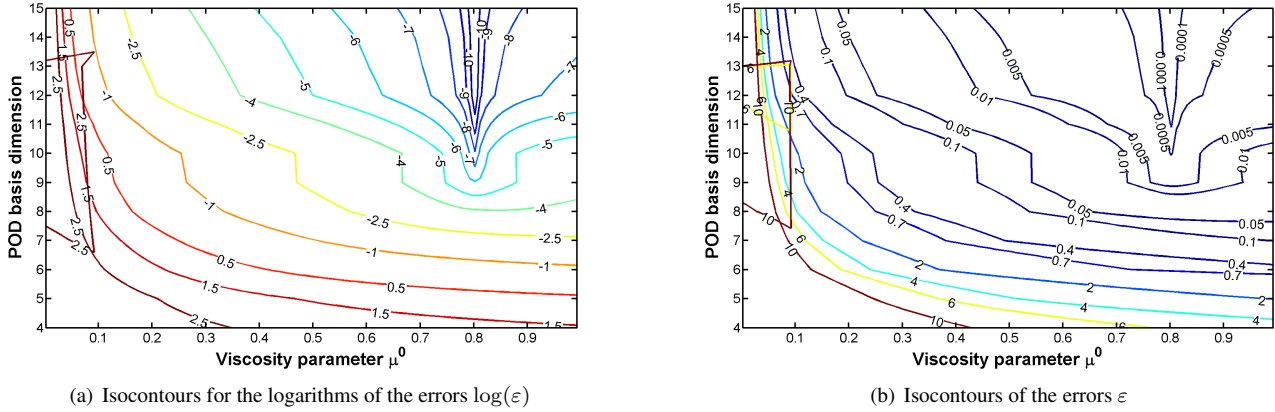


Figure 3: Isocontours of the reduced model errors for different POD basis dimensions and parameters μ^0 . The reduced order model uses a basis constructed from the full order simulation with parameter value $\mu = 0.8$.

We construct two probabilistic models for estimating the ROM model errors, the first one uses a Gaussian process with a squared-exponential covariance kernel (13) and the second one uses a neural network with six hidden layers and hyperbolic tangent sigmoid activation function in each layer. Tables 1 and 2 show the averages and variances of errors in prediction provided by GP and ANN for different sample sizes. The results are obtained using a conventional validation with 80% of the whole data set involved in training process and the remaining 20% employed for testing. The misfit is computed using the same formulas presented in (11a) to evaluate the prediction errors of one-fold set in the K-fold cross validation approach. Table 2 shows the prediction errors of (34) computed via equation (11a) with $y = \varepsilon$ and $\hat{y} = \hat{\varepsilon}$, i.e. no data scaling; the predictions have a large variance and a low accuracy. Scaling the data and targeting $\log(\varepsilon)$ results using (36), reduce the variance of the predictions, and increase the accuracy, as shown in Table 1. The same formula (11a) with $y = \log(\varepsilon)$ and $\hat{y} = \widehat{\log(\varepsilon)}$ was applied. In both cases ANN outperforms the GP. Moreover, as the number of data points grows, the accuracy increases and the variance decreases faster for ANN.

Sample size	GP		ANN	
	E_{fold}	VAR_{fold}	E_{fold}	VAR_{fold}
100	0.5076	0.3870	0.7237	0.9974
1000	0.2352	0.0746	0.0650	0.0397
3000	0.1555	0.0517	0.0110	0.0063
5000	0.0810	0.0176	0.0090	0.0006

Table 1: Average and variance of error in predictions of (36) for ANN and GP using logarithms of errors ($\log(\varepsilon)$) in training data for different sample sizes

Sample size	GP		ANN	
	E_{fold}	VAR_{fold}	E_{fold}	VAR_{fold}
100	8.1523	709.16	9.2341	536.71
1000	7.9558	593.36	4.4000	424.2900
3000	6.7521	524.98	4.2238	17.6300
5000	2.8229	12.63	2.8134	8.9430

Table 2: Average and variance of error in predictions of (34) for ANN and GP using raw errors (ε) in training data for different sample sizes

Figures 4 and 5 show the corresponding histogram of the predicted models errors (36) and (34) using 100 and 1000 training samples for both ANN and GP models. As the number of training samples increase, the uncertainty in the prediction decreases. The histograms can also assess the validity of GP assumptions (16), (12), (14). The difference between the true and estimated values should behave as samples from the distribution $\mathcal{N}(0, \sigma_n^2)$ [25]. In our case they are hardly normally distributed and this indicates that the data set for the problems we are working with, are not from Gaussian distributions.

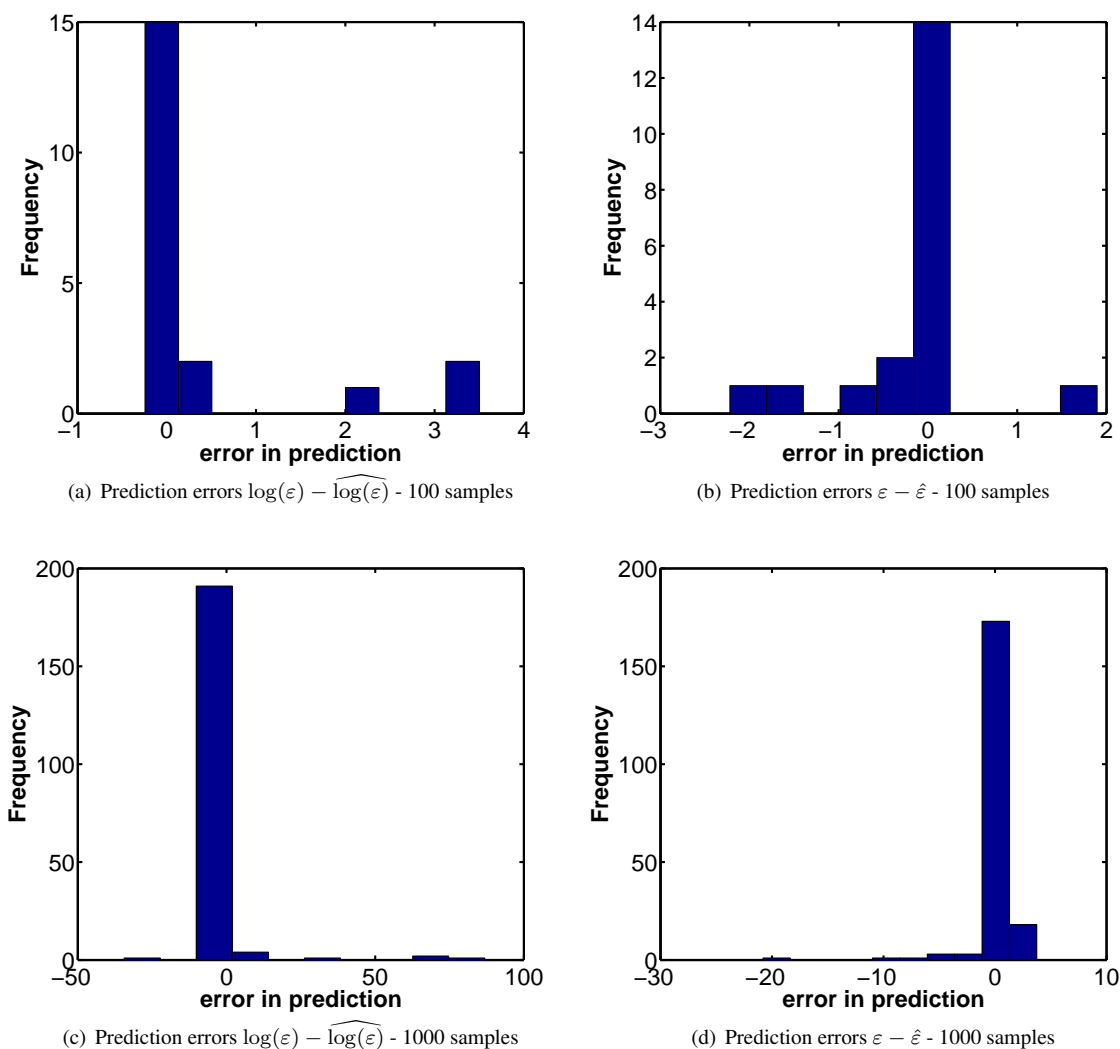


Figure 4: Histogram of errors in prediction using ANN.

Scaling the data and targeting $\log \varepsilon$ errors clearly improve the performance of our probabilistic models. Consequently for the rest of the manuscript we will only use model (36). To assess the quality of the probabilistic models a five-fold cross-validation process is also used. The results computed using formula (11b) are shown in Table 3. Neural network outperforms the Gaussian process and estimates the errors more accurately. It also has less variance than the Gaussian process which indicates it has more stable predictions.

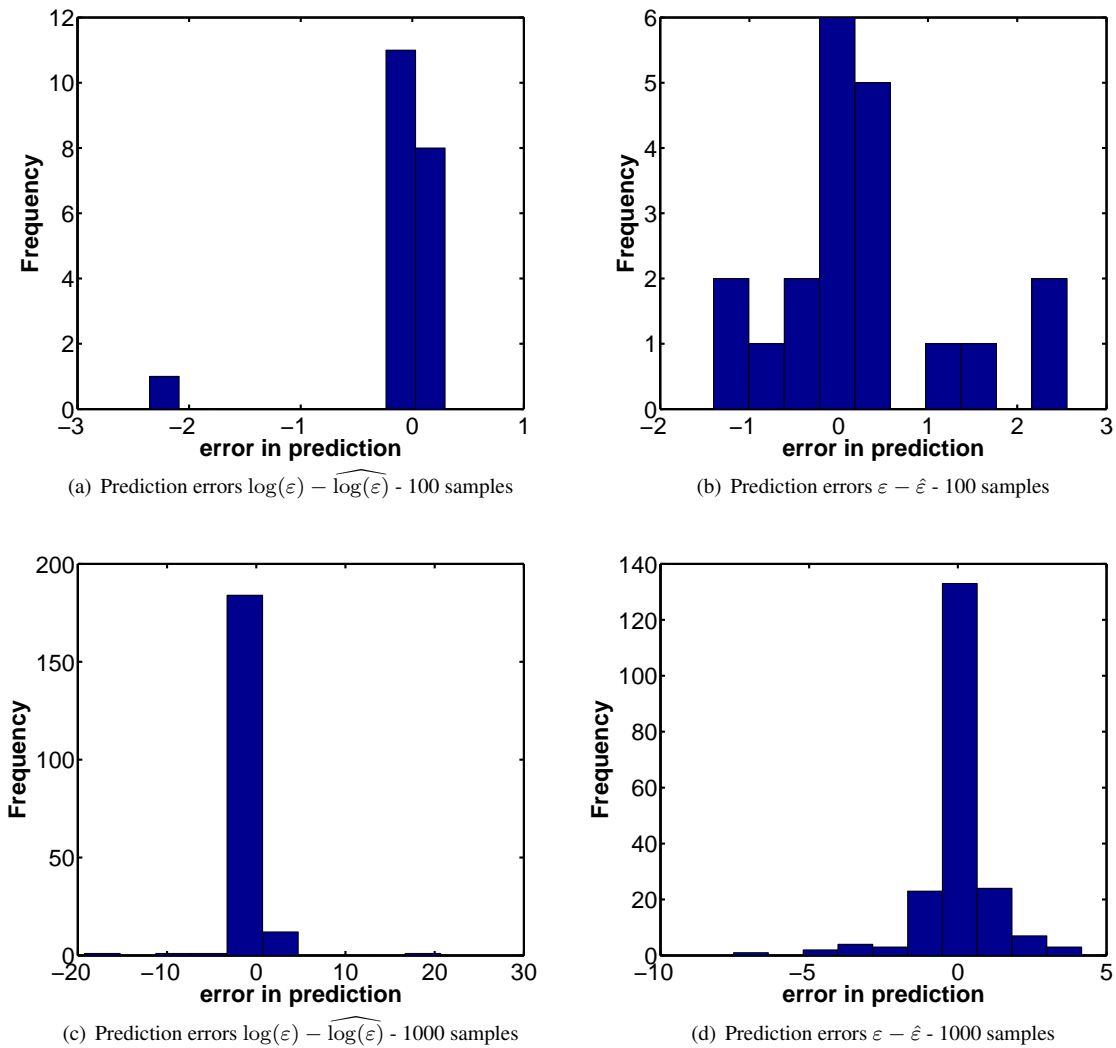


Figure 5: Histogram of errors in prediction using GP.

	E	VAR
ANN	0.004004	2.16×10^{-6}
GP	0.092352	1.32×10^{-5}

Table 3: ANN and GP statistical results over 5 fold cross validation.

Figure 6 illustrates the average of errors in predictions over five different ANN and GP configurations. In each configuration, the machine is trained on random 80% split of data set and tested on the fixed selected test data shown in figure 6. Getting the average of predictions on different trained models decreases the bias in predictions.

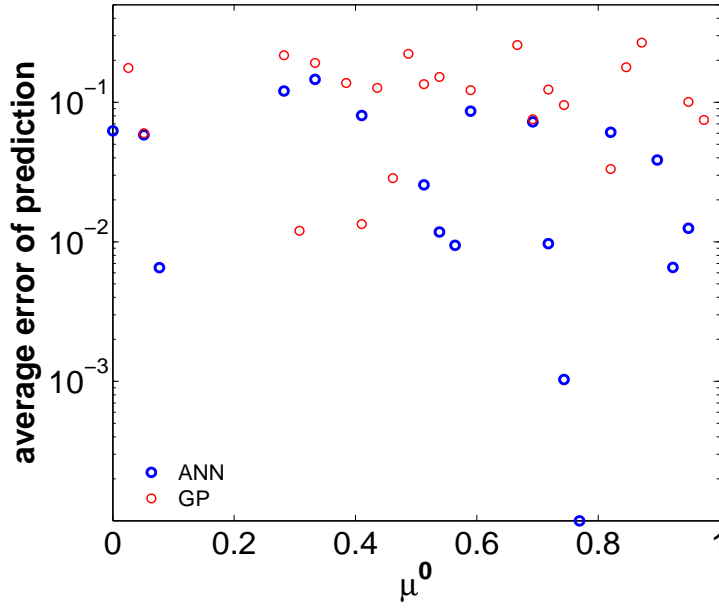


Figure 6: The average of errors in predictions using five different trained models

6.2.2 Construction of a μ -feasible interval

We saw in the previous subsection that probabilistic models can accurately estimate the error ε (35) associated with reduced order models. Thus we can employ them to establish a range of viscosity parameters around μ such that the reduced order solutions depending on U_μ satisfy some desired accuracy level. More precisely, starting from parameter μ , a fixed POD basis dimension and a tolerance error $\log(\bar{\varepsilon})$, we are searching for an interval $[d_l, d_r]$ such that the estimated prediction $\widehat{\log(\varepsilon)}$ of the true error $\log(\varepsilon)$ (35) meets the requirement

$$\widehat{\log(\varepsilon)} < \log(\bar{\varepsilon}), \forall \mu^0 \in [d_l, d_r]. \quad (37)$$

Our proposed strategy makes use of a simply incremental approach by sampling the vicinity of μ to account for the estimated errors $\widehat{\log(\varepsilon)}$ forecasted by the probabilistic models defined before. A grid of new parameters μ^0 is build around μ and the machines predict the errors outward of μ . Once the machines outputs are larger than the prescribed error $\log(\bar{\varepsilon})$, the previous μ^0 satisfying the constrain (37) is set as d_l , for $\mu^0 < \mu$ or d_r for $\mu^0 > \mu$.

Figure 7 illustrates the range of parameters estimated by the neural network and Gaussian process against the true feasible interval and the results show good agreement. For this experiment we set $\mu = 0.7$, dimension of POD=9 and error threshold $\bar{\varepsilon} = 10^{-2}$. Values of $\mu^0 = \mu \pm 0.001 \cdot i$, $i = 1, 2, \dots$ are passed to the probabilistic models. The average range of parameters obtained over five different configurations with neural network is $[0.650, 0.780]$ while in the case of Gaussian process we obtained $[0.655, 0.780]$. In each configuration, we train the model with 80% random split of the data set and test it over the fixed test set of figure 7. For this design, the true range of parameters is $[0.650, 0.785]$ underlying the predicting potential of machine learning models.

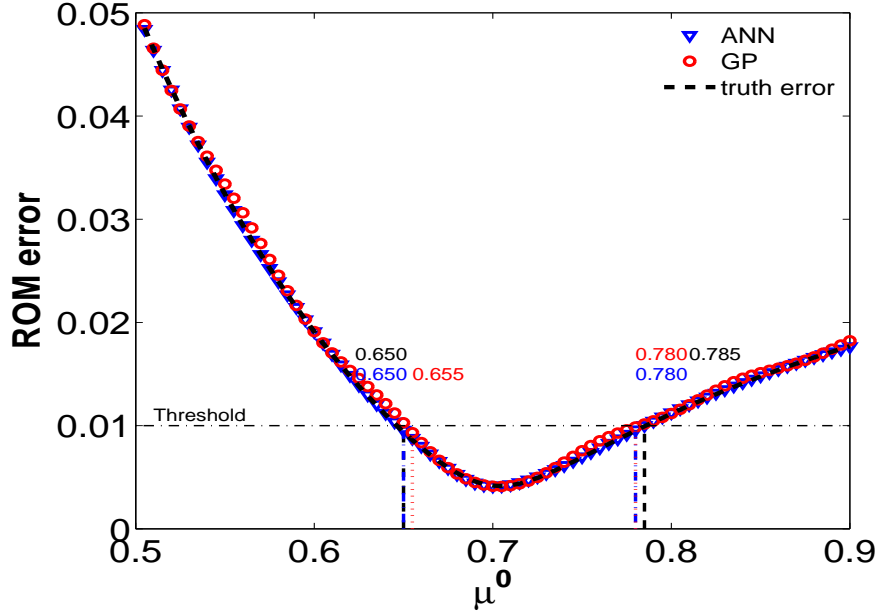


Figure 7: The average range of parameter μ^0 obtained with ANN and GP for dimension of POD=9 and $\mu = 0.7$. The desired accuracy is $\bar{\varepsilon} = 10^{-2}$. The numbers represent the left and the right edges of the predicted vs the true feasible intervals.

6.2.3 The parametric map as a reunion of μ -feasible intervals

A reunion of different μ_k -feasible intervals can be designed to cover a general entire 1D-parametric domain $[A, B]$. We refer to this reunion as a parametric map and once such construction is available will allow for reduced order simulations with a-priori error quantification for any value of viscosity parameter $\mu^0 \in [A, B]$.

A greedy strategy is described in Algorithm 3 and its output is a collection of feasible intervals $\cup_{k=1}^n [d_l^k, d_r^k] \supset [A, B]$. Each interval $[d_l^k, d_r^k]$ is associated with some accuracy threshold $\bar{\varepsilon}_k$. For small viscous parametric values we found out that designing μ_k -feasible intervals associated with higher precision levels (i.e. very small thresholds $\bar{\varepsilon}_k$) is impossible since the dynamics of parametric 1D-Burgers model solutions changes dramatically with smaller viscosity parameters. In consequence we decided to let $\bar{\varepsilon}_k$ vary along the parametric domain to accommodate the solution physical behaviour. Thus a small threshold $\bar{\varepsilon}_0$ will be initially set and as we will advance charting the parameter interval $[A, B]$ from right to left, the threshold $\bar{\varepsilon}_k$ will be increased.

The algorithm starts by selecting the first centered parameter μ_0 responsible for basis generation. It can be set to $\mu_0 = B$ but may take any value in the proximity of B , $\mu_0 \leq B$. This choice depends on the variability of parametric solutions in this domain region and by electing μ_0 to differ from the right edge of the domain, the number n of feasible intervals will be decreased.

The next step is to set the threshold $\bar{\varepsilon}_0$ along with the maximum permitted size of the initial feasible interval to be constructed. This is set to $2 \cdot r_0$, thus r_0 can be referred as the interval radius. Along with the radius, the parameter Δs will decide the maximum number of probabilistic model calls employed for the construction of the μ_0 -feasible interval. While the radius is allowed to vary during the algorithm iterations, Δs is kept constant. Finally the dimension of POD basis has to be selected together with three parameters β_1 , β_2 and β_3 responsible for changes in the threshold, radius and selecting a new parameter location μ_k encountered during the procedure.

The instructions between lines 5 and 20 generate the μ_k -feasible interval, for the case when the current centered parameter μ_k represents an interior point of $[d_l^k, d_r^k]$. For situation when $\mu_k = d_l^k$ or $\mu_k = d_r^k$, the threshold has to be increased (by setting $\bar{\varepsilon}_{k+1} = \beta_1 \bar{\varepsilon}_k$ at line 22), since the reduced order model solutions can not satisfy the desired precision according to the estimated probabilistic errors. At this stage the radius is decreased too since we reached a parameter region where model dynamics changes rapidly and a larger feasible interval is not possible. Once the new centered parameter μ_{k+1} is proposed the algorithm checks if the following constrain is satisfied

$$[d_l^{k+1}, d_r^{k+1}] \cap \left(\bigcup_{i=1}^k [d_l^i, d_r^i] \right) \neq \emptyset. \quad (38)$$

This is achieved by checking the estimated reduced order model solution error at d_l^k using the basis defined by the high-fidelity trajectory at μ_{k+1} (see instruction 25). If the predicted error is smaller than the current threshold, assuming a monotonically increasing error with larger distances $d(\mu^0, \mu_{k+1})$, the reduced order model solutions should satisfy the accuracy threshold for all $\mu^0 \in [\mu_{k+1}, d_l^k]$. In consequence the equation (38) will be satisfied for the current μ_{k+1} , if we set $r_{k+1} = \mu_{k+1} - d_l^k$ (see instruction 28). In the case the error estimate is larger than the present threshold, the centered parameter μ_{k+1} is updated to the middle point between old μ_{k+1} and d_l^k . For the situation where the monotonic property of the error does not hold in practice, a simply safety net is used at instruction 12. The entire algorithm stops when $\mu_k \leq A$.

For our experiments we set $A = 0.01$, $B = 1$, $\bar{\varepsilon}_0 = 1.e - 2$, $\Delta s = 5.e - 3$, $r_0 = 0.5$, $dim = 9$, $\beta_1 = 1.2$, $\beta_2 = 0.9$ and $\beta_3 = 1.4$. We initiate the algorithm by setting $\mu_0 = 0.87$, and the first feasible interval $[0.7700, 1.0500]$ is obtained. Next the algorithm selects $\mu_1 = 0.73$ with the associated range of $[0.6700, 0.8250]$ using the same initial threshold level. As we cover the parametric domain from right to left, i.e. selecting smaller and smaller parameters μ_k , the algorithm enlarges the current threshold $\bar{\varepsilon}_k$, otherwise the reduced order models would not satisfy the initial precision. We continue this process until we get the threshold 6.25 with $\mu_{32} = 0.021$ and the corresponding feasible interval $[0.00940, 0.039]$. The obtained parametric map is depicted in Figure 8 where the associated threshold varies with the parameter change.

Algorithm 3 Generation of parametric map for reduced order models usage

```

1: Select  $\mu_0$  as the right edge of the parameter interval, i.e.  $\mu_0 = B$ .
2: Set error threshold  $\bar{\varepsilon}_0$ , step size  $\Delta s$  for selection of new parameter locations  $\mu^0$ , the maximum search radius  $r_0$ , dimension of POD basis  $dim$  and  $\beta_1$ ,  $\beta_2$  and  $\beta_3$ .
3: Set  $k = 0$ .
4: DO
5: FOR  $i=1$  to  $int(\frac{r_k}{\Delta s})$ 
6: Set  $\mu_+^o = \mu_k + i\Delta s$ 
7: IF  $\phi(\mu_+^o, \mu_k, dim) > \log(\bar{\varepsilon}_k)$  THEN
8: Set  $d_r^k = \mu_k + (i - 1)\Delta s$ . EXIT.
9: END IF
10: END FOR
11: IF  $k > 0$  THEN
12: IF  $d_r^k < d_l^{k-1}$  THEN  $\mu_k = \frac{\mu_k + d_l^{k-1}}{2}$ . GOTO 5.
13: END IF
14: END IF
15: FOR  $j = 1$  to  $int(\frac{r_k}{\Delta s})$ 
16: Set  $\mu_-^o = \mu_k - j\Delta s$ 
17: IF  $\phi(\mu_-^o, \mu_k, dim) > \log(\bar{\varepsilon}_k)$  THEN
18: Set  $d_l^k = \mu_k - (j - 1)\Delta s$ . EXIT.
19: END IF
20: END FOR
21: IF (i=1).OR.(j=1) THEN
22: Set  $\bar{\varepsilon}_k = \beta_1 \cdot \bar{\varepsilon}_k$ ;  $r_k = \beta_2 \cdot r_k$ ; GOTO 5.
23: ELSE  $\mu_{k+1} = \mu_k - \beta_3(j - 1)\Delta s$ ;  $\bar{\varepsilon}_{k+1} = \bar{\varepsilon}_k$ .
24: END IF
25: WHILE  $\phi(d_l^k, \mu_{k+1}, dim) > \log(\bar{\varepsilon}_{k+1})$  DO
26:  $\mu_{k+1} = \frac{\mu_{k+1} + d_l^k}{2}$ .
27: END WHILE
28: Set  $r_{k+1} = \mu_{k+1} - d_l^k$ .
29:  $k = k + 1$ .
30: WHILE  $\mu_k \geq A$  THEN STOP.

```

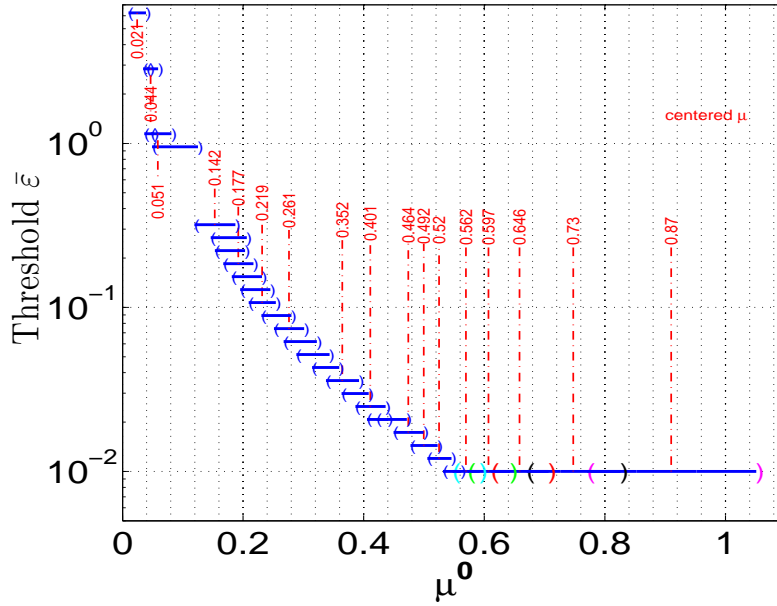


Figure 8: The diffusion parametric map defining the local feasible intervals and their corresponding errors. Associated with one feasible interval there is a centered parameter μ high-fidelity trajectory that guides the construction of a reduced basis and operators such that the subsequent reduced order model solutions along this interval are accurate within the threshold depicted by the Y-axis.

6.3 Select the best already existing ROMs for a new parameter value μ^0

An important and practical concern associated with the reduced order models is their lack of robustness with respect to parameter change. Here, we propose to solve another practical problem, i.e. giving a collection of reduced bases computed at various locations in the parameter space, find the one that proposes the most accurate reduced order solution for a new viscosity parameter μ^0 . We will rely on similar probabilistic models built in subsection 6.2.1. The input features for the GP and ANN are the new parameter μ^0 , a parameter μ whose corresponding trajectory is used as snapshots for generating basis U_μ and the dimension of the reduced manifold. The target random variable $\hat{y} = \widehat{\log \varepsilon}$ is the estimated log of error of the reduced order model solution at μ^0 using the basis U_μ . For our experiments, approximately 12,000 samples were generated and include different values of POD basis dimensions from 4 to 15, new viscosity parameter $\mu^0 \in [0.1, 1]$ and parameters $\mu_k \in [10^{-2}, 1]$ equally distributed with interval 0.1 and 0.01, respectively and the corresponding log of the Frobenius norm of true ROM errors $\log(\varepsilon)$ (35).

The data set constructed for this problem is similar to the one employed for designing the errors models. For that problem we fixed the parameter μ associated with the high-fidelity trajectory used for basis generation and computed the ROM errors corresponding to different μ^0 . Here we fixed μ^0 and then computed the ROM errors using different bases U_μ .

The data set is randomly partitioned into 5 equal size sub-samples and a 5-fold cross-validation process is used to assess the quality of the probabilistic models. A neural network with 6 hidden layers and hyperbolic tangent sigmoid activation function in each layer is used while for the Gaussian process we have employed the squared-exponential-covariance kernel (13). Table 4 shows numerical results obtained with Neural networks and Gaussian process, over five folds using (11b). The Gaussian process outperforms the neural network and has less variance which indicates the GP is more accurate and stable for this specific problem than neural network.

	E	VAR
ANN	0.002839	1.5496×10^{-5}
GP	0.001135	2.5922×10^{-8}

Table 4: Statistical results of error in predictions in ANN and GP over 5 fold cross validation

The results in Figure 9 illustrate the errors in prediction of the reduced order models errors for two viscosity parameters $\mu^0 = 0.35$ and 0.65 and various bases represented along the y axis. The mean of the estimates in the case of the ANN are closer than the true errors in comparison with the output of the Gaussian Process for this particular example.

Moreover we can notice that the estimation curves are crossing initially close to $\mu = 0.45$. It suggests that one can choose the high-fidelity trajectory $\mu = 0.45$ to construct a reduced order basis such that to obtain similar accuracy levels for both reduced order solutions computed at new viscosity parameters $\mu^0 = 0.35$ and 0.65. This reveals the non-monotonic property of the reduced order model error with respect to the distance $d(\mu^0, \mu)$.

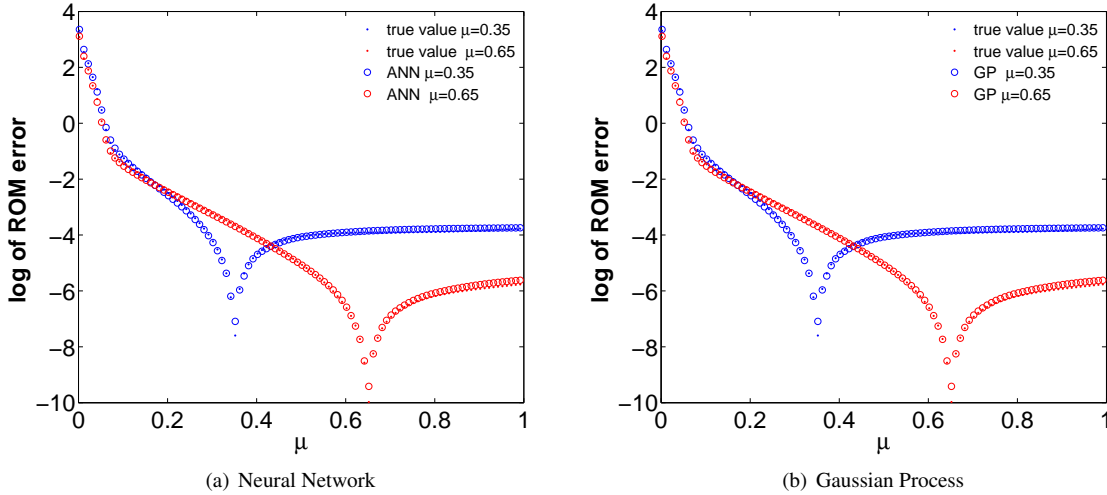


Figure 9: ANN and GP prediction errors ($\log(\varepsilon) - \widehat{\log(\varepsilon)}$) of the ROM errors for dimension of POD basis=12, $\mu^0 = 0.65$ and $\mu^0 = 0.35$ and different values of μ .

6.4 Combining Available Information for Accurate ROMs at New Parametric Configurations

Experiments in subsection 6.3 revealed the potential of probabilistic models to select a hierarchy of reduced manifolds that produces higher accurate solutions. Figure 9 depicts the accuracy of reduced order models for $\mu^0 = 0.35$ and 0.65 using POD basis computed at various locations in the parameter interval. Assuming only 10 existing POD subspaces constructed for parameters equally distributed along the interval $[0.1, 1]$, $\mu = 0.1, 0.2, 0.3, \dots$, figure 9 shows that the most accurate reduced solutions for $\mu^0 = 0.35$ are produced using the bases computed at $\mu_1 = 0.3$ and $\mu_2 = 0.4$. Consequently, the numerical experiments described here focus on the construction of a POD basis for $\mu^0 = 0.35$ by combining the data available for $\mu_1 = 0.3$ and $\mu_2 = 0.4$.

The performances of the discussed methods (bases concatenation, Lagrange interpolation of bases in the matrix space and in the tangent space of the Grassmann manifold, Lagrange interpolation of high-fidelity solutions) are shown in the case of three main experiments: variation in the final time t_f , in the non-linear advection coefficient ν and POD basis size. The first two experiments scale the time and space and modify the linear and nonlinear characteristics of the model. For example, in the case of a tiny small final time and advection coefficient, the diffusion linear part represents the main dynamical engine of the model thus it behaves linearly. The results are compared against reduced order models constructed using U_{μ_1} and U_{μ_2} , respectively.

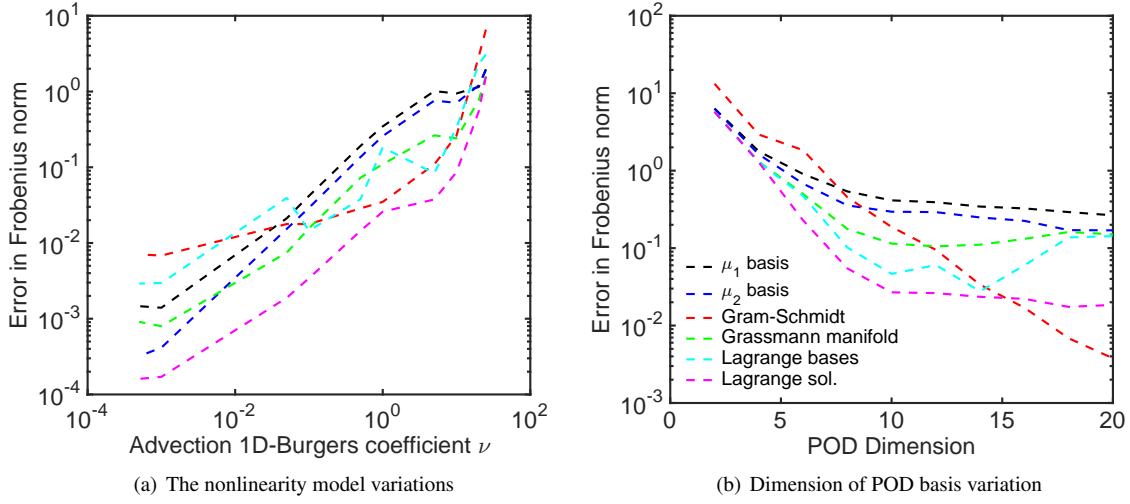
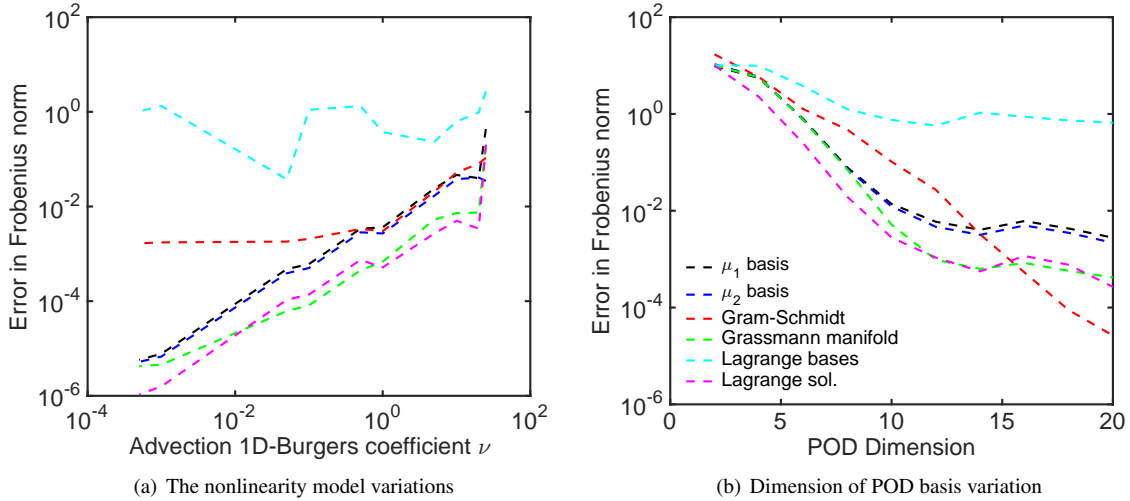
The experiments make use of a space mesh of 201 points while 301 time steps are used. Figure 10 illustrates the Frobenius norm error between the high fidelity and reduced order model solutions for the final time $t_f = 0.01$. Panel (a) presents the accuracy results as a function of the advection coefficient ν . Interpolating the high-fidelity solutions leads to the most accurate reduced order model. For large advection coefficients all of the methods suffer accuracy losses. Among the potential explanations we include the constant size of the POD basis and its linear dependence on the viscosity parameter assumed by all of the methods in various forms. Keeping the POD basis size constant represents a source of errors as seen in Figure 3 where the viscosity parameter is varied.

By interpolating the geometry characteristics of the reduced subspaces via subspace angle interpolation or Grassmann manifold approach (only one is shown in the figures since for interpolating two bases the methods are the same) we expect to obtain more accurate reduced order models. While Lagrangian interpolation of the bases is performed in both matrix space and tangent space of the Grassmann manifold (shown in cyan and green in Figure 10, the later approach performs better in this scenario confirming our expectations. The concatenation of bases using Gram-Schmidt was successful only for larger advection coefficients (red curve in Figure 10)(a) for a POD size set to 14.

Increasing the dimension of the basis enhances the so called Gram-Schmidt reduced order model solution accuracy for $\nu = 1$ (see Figure 10(b)). For this case Lagrange interpolation in the matrix space shows better performances in comparison with the output of the Grassmann manifold approach.

Next we increase the nonlinearity characteristics of the model by setting the final time to $t_f = 1$ and Figure 11 illustrates the Frobenius norm errors as a function of the advection coefficient ν and POD dimension size. The errors produced by the reduced order model derived via Grassmann manifold method are similar with the ones obtained by the surrogate model relying on a POD basis computed via the Lagrange interpolation of the high-fidelity model solutions.

The Lagrange interpolation of bases in the matrix space is not successful as seen in both panels of figure 11. Increasing the POD size to 20, the Gram-Schmidt approach enhance the accuracy of the solution (see Figure 11(b)).

Figure 10: Strategies comparison for generation of accurate ROMs for a new viscosity parameter $\mu^0 = 0.35$ and $t_f = 0.01$.Figure 11: Strategies comparison for generation of accurate ROMs for a new viscosity parameter $\mu^0 = 0.35$ and $t_f = 1$.

6.5 Optimal size of reduced order model

Here we propose two alternative approaches to select the reduced basis size that accounts for specified accuracy levels in the reduced order model solutions. These techniques employ construction of probabilistic models $\phi : \mathbf{z} \rightarrow \hat{y}$ via neural network and Gaussian process as stated in section 4. The input features \mathbf{z} for this problem consist of the viscosity parameter $\mu \in [0.01, 1]$ and the log of the Frobenius norm of the error between the high-fidelity and reduced order models $\log(\varepsilon)$ (35). The searched output \hat{y} is the dimension of the reduced manifold d . For the training phase, the data set is generated using several runs of the 1D-Burgers model with various viscosity parameters μ , different basis sizes d and the log of the Frobenius norms of the discrepancies between the full and the projected reduced order solutions $\log(\varepsilon)$. The machines will learn the sizes of reduced order basis d associated with the parameter μ and the corresponding $\log(\varepsilon)$. Later it will be able to estimate the proper size of reduced basis by providing it the specific viscosity parameter μ and the desired error $\log(\varepsilon)$. The computational cost is low once the probabilistic models are constructed. The output indicates the dimension of the reduced manifold for which the ROM solution satisfies the corresponding error threshold. Thus we do not need to compute the entire spectrum of the snapshots matrix in advance which for large spatial discretization meshes translates into important computational costs reduction.

For our experiments, approximately 9000 samples were generated for different values of the viscosity parameter μ equally distributed within the interval $[10^{-2}, 1]$, various reduced basis dimensions from 4 to 15 and the corresponding $\log(\varepsilon)$. Figure 12 illustrates the contours of the log of reduced order model error, over the viscosity parameter domain and various POD sizes.

A neural network with 5 hidden layers and hyperbolic tangent sigmoid activation function in each layer is used while for the Gaussian process we have used the squared-exponential-covariance kernel (13). Table 5 show the average and variance of error in GP and ANN estimations using different sample sizes. The ANN outperforms the GP and as the number of data points grows, the accuracy increases

and the variance decreases. The results are obtained using a conventional validation with 80% of the sample size dedicated for training data and the other 20% for the test data. The employed formula is described in equation (11a).

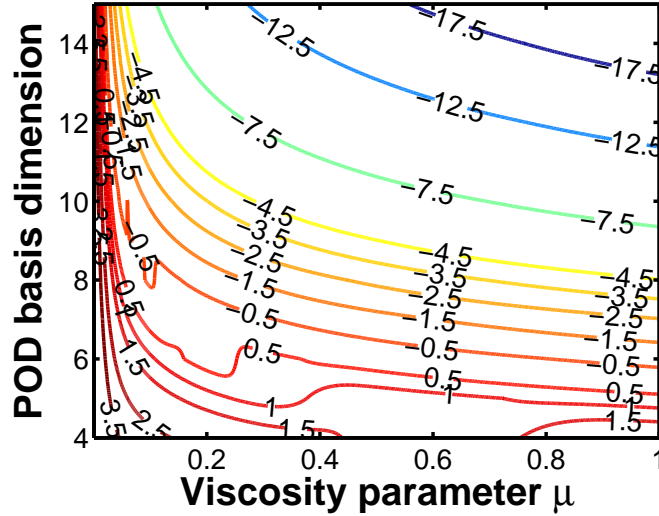


Figure 12: Isocontours of the reduced model errors for different POD basis dimensions and viscosity parameters μ .

sample size	GP		ANN	
	E_{fold}	VAR_{fold}	E_{fold}	VAR_{fold}
100	0.2801	0.0901	0.1580	0.02204
1000	0.1489	0.0408	0.0121	0.0015
3000	0.1013	0.0194	0.0273	0.0009
5000	0.0884	0.0174	0.0080	0.0002

Table 5: Average and variance of errors in prediction of optimal reduced basis size using ANN and GP probabilistic models for different sample sizes

Figures 13 and 14 shows the corresponding errors in estimation on 100 and 1000 training samples for both ANN and GP model. These histograms as stated before, can assess the validity of GP assumptions. The data set distribution shape is closer to a Gaussian profile than in the case of the data set distribution discussed in section 6.2.1 used for generation of reduced order model error probabilistic models.

To assess the accuracy of the probabilistic models, the data set is randomly partitioned into 5 equal size sub-samples, and 5– fold cross-validation test is implemented. The 5 results from the folds are averaged and they are presented in table 6. The neural network model correctly estimated the size of the reduced manifold in 87% cases. Gaussian process correctly estimates the POD size 53% of the times. The variance results shows that the GP model has more stable predictions indicating a higher bias in the outcomes.

Dimension discrepancies	zero	one	two	three	four	> four	VAR
ANN	87%	11%	2%	0	0	0	2.779×10^{-3}
GP	53%	23%	15%	5%	3%	1%	4.575×10^{-4}

Table 6: POD basis dimension discrepancies between the ANN and GP predictions and true values over 5 fold cross validation. The errors variance is also computed.

In figure 15, we compare the output of our probabilistic approaches against the eigenvalues estimation on a set of randomly selected

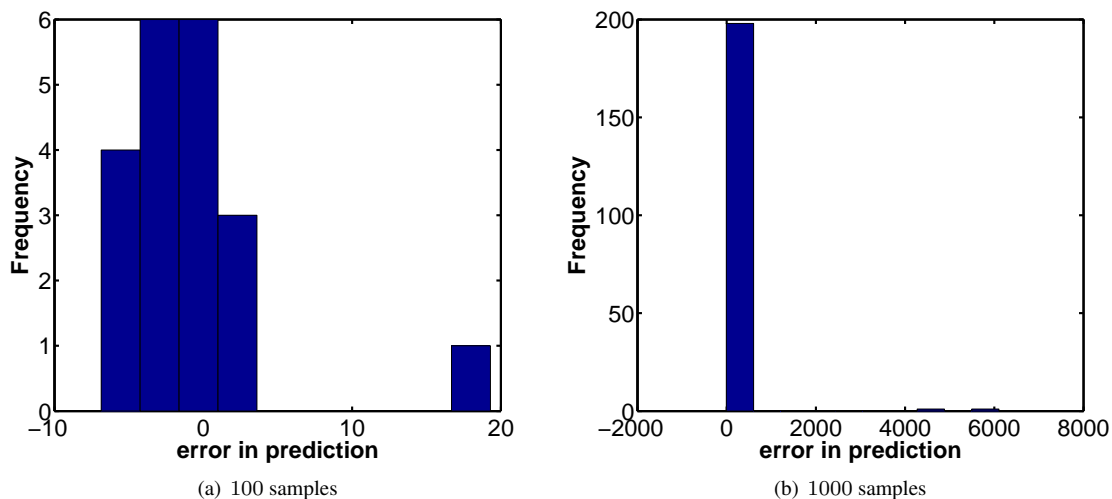


Figure 13: Histogram of errors in prediction of the optimal reduced basis dimension using ANN for different sample sizes

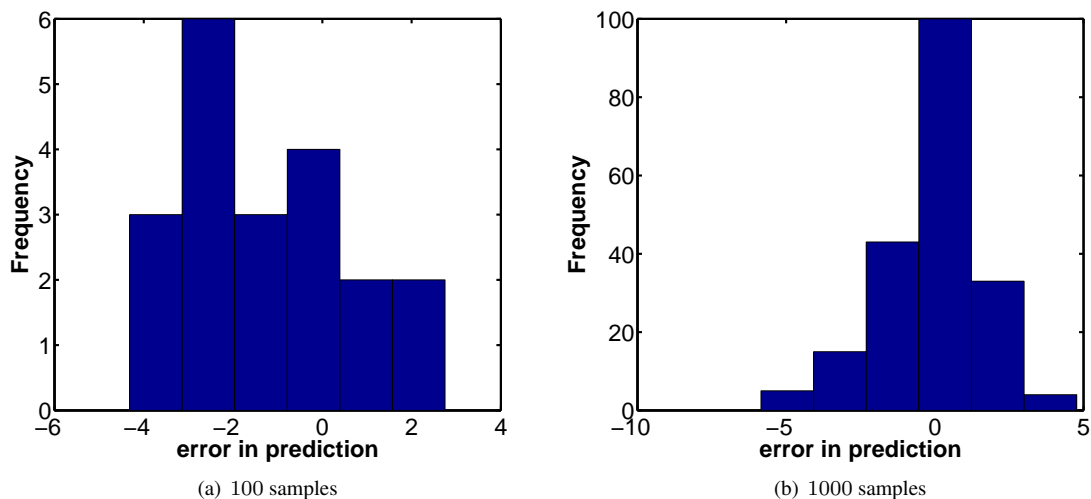


Figure 14: Histogram of errors in prediction of the optimal reduced basis dimension using GP for different sample sizes

test data. The eigenvalue estimation is the standard method for selecting the optimal reduced manifold dimension when a prescribed level of accuracy of the reduced solution is desired. Here the desired accuracy ε was set to 10^{-3} . The mismatches between the predicted and true dimensions are depicted in the figure 15. The predicted values are the averages over five different models where each model of ANN and GP are trained on random 80% split of dataset and tested on the fixed selected 20% test data. We notice that the snapshots matrix spectrum underestimates the true size of the manifold as expected since the errors due to integration in the reduced space are not accounted. The neural network predictions were extremely accurate for most of the samples while the Gaussian process usually overestimated the reduced manifold dimensions.

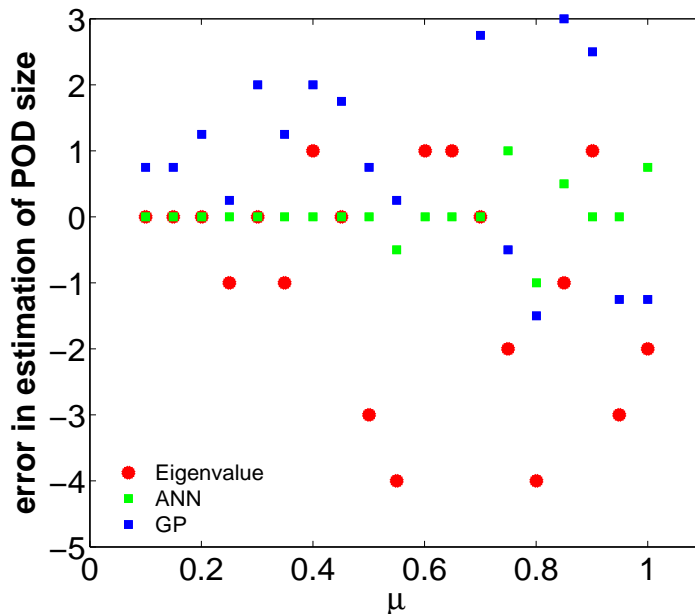


Figure 15: Average error of the POD size estimations on a randomly selected test data with desired accuracy of $\varepsilon = 10^{-3}$.

7 Conclusions

This work demonstrates the value of machine learning approaches to guide the construction of parametric space partitioning for the usage of efficient and accurate local reduced order models. While the current methodologies are defined in the sense of Voronoi tessellation [26] and rely on K-means algorithms [5, 6, 47], our approach delimitates sub-regions of the parametric space by making use of Gaussian Processing and Artificial Neural Networks models for the errors of reduced order models and parametric domain sampling. The employed machine learning models differ from the one proposed in [25] having more additional features such as reduced subspace dimension and are specially projected for accurate predictions of local reduced order models errors. For each sub-region, an associated reduced order basis and operators are constructed depending on a single representative high-fidelity trajectory and, the corresponding local reduced order models solutions have known precision levels. The novel methodology is applied for a 1D-Burgers model, and a parametric map covering the viscosity domain with parametric sub-intervals and associated errors thresholds is designed.

Our numerical experiments revealed the non-monotonic property of the reduced order model error with respect to the distance between the current parametric configuration and the one used to generate the reduced subspace. Thus we proposed machine learning models for selecting a hierarchy of reduced bases producing the most accurate solutions for a new parameter configuration. Based on this hierarchy, three already existing methods involving bases interpolation and concatenation and high-fidelity model solutions interpolation are applied to enhance the quality of the associated reduced order model solutions. It has been shown that the assumption of linear variation of the basis over the parametric space leads to a different reduced basis formulation than if the linear variation hypothesis of the high-fidelity solution over the parametric domain is followed. Several experiments were performed by scaling the time and space and modifying the nonlinear characteristics of the model. In most cases, interpolating the already existing high-fidelity trajectories generated the most accurate reduced order models for a new viscous parameter revealing that the solution behavior over the parametric region under study can be linearly approximated. Lagrange interpolation of bases in the tangent space of the Grassmann manifold and concatenation of bases for larger reduced subspaces showed also good performances.

Finally we addressed the problem of selecting the dimension of the reduced order model when its solution must satisfy a desired level of accuracy. Our approach based on learning better estimates the ROM basis dimension in comparison with the results obtained by truncating the spectrum of the snapshots matrix.

A future goal is to decrease the computational complexity of the parametric map design procedure. Currently the training data required by the probabilistic models relies on many high-fidelity simulations. By employing fast a-posteriori error estimation results [63], this dependency will be much decreased. In addition we plan to incorporate residual norm and rigorous error bounds among the data fitting models features to enhance their prediction capabilities as remarked in [25].

Acknowledgements

This work was supported in part and by the award NSF CCF 1218454 and by the Computational Science Laboratory at Virginia Tech.

References

- [1] P-A Absil, Robert Mahony, and Rodolphe Sepulchre. Riemannian geometry of grassmann manifolds with a view on algorithmic computation. Acta Applicandae Mathematica, 80(2):199–220, 2004.
- [2] Natalia M Alexandrov, Robert Michael Lewis, Clyde R Gumbert, Lawrence L Green, and Perry A Newman. Approximation and model management in aerodynamic optimization with variable-fidelity models. Journal of Aircraft, 38(6):1093–1101, 2001.
- [3] D. Amsallem and J. Cortial. The Grassmann Manifold and the Subspace Angle Interpolation Methods. FRG internal document, Stanford University, 2007.
- [4] D. Amsallem and C. Farhat. An Interpolation Method for Adapting Reduced-Order Models and Application to Aeroelasticity. AIAA Journal, 46(7):1803–1813, 2008.
- [5] David Amsallem, Matthew J Zahr, and Charbel Farhat. Nonlinear model order reduction based on local reduced-order bases. International Journal for Numerical Methods in Engineering, 92(10):891–916, 2012.
- [6] David Amsallem, Matthew J. Zahr, and Kyle Washabaugh. Fast local reduced basis updates for the efficient reduction of nonlinear systems with hyper-reduction. Advances in Computational Mathematics, pages 1–44, 2015.
- [7] Ramin Ayanzadeh, Azam S Zavar Mousavi, and Saeid Setayeshi. Fossil fuel consumption prediction using emotional learning in amygdala. In Electrical Engineering (ICEE), 2011 19th Iranian Conference on, pages 1–6. IEEE, 2011.
- [8] M. Barrault, Y. Maday, N. D. Nguyen, and A. T. Patera. An ‘empirical interpolation’ method: application to efficient reduced-basis discretization of partial differential equations. Comptes Rendus Mathématique. Académie des Sciences. Paris, 339(9):667–672, 2004.
- [9] M. Barrault, Y. Maday, N.C. Nguyen, and A.T. Patera. An ‘empirical interpolation’ method: application to efficient reduced-basis discretization of partial differential equations. Comptes Rendus Mathématique, 339(9):667–672, 2004.
- [10] Evgeni Begelfor and Michael Werman. Affine invariance revisited. In Proceedings of the 2006 IEEE Computer Society Conference of Computer Vision and Pattern Recognition, volume 2, pages 2087–2094, 2006.
- [11] C.M. Bishop. Pattern Recognition and Machine Learning. Information Science and Statistics. Springer, 2006.
- [12] T. Bui-Thanh, M. Damodaran, and K. Willcox. Aerodynamic data reconstruction and inverse design using proper orthogonal decomposition. AIAA Journal, pages 1505–1516, 2004.
- [13] Roberto Buizza. Horizontal resolution impact on short-and long-range forecast error. Quarterly Journal of the Royal Meteorological Society, 136(649):1020–1035, 2010.
- [14] Johannes Martinus Burgers. A mathematical model illustrating the theory of turbulence. Advances in Applied Mechanics, 1:171–199, 1948.
- [15] Dan Cacuci. Sensitivity and Uncertainty Analysis: Theory. Chapman and Hall/CRC Press, 2003.
- [16] Dan G. Cacuci. Second-order adjoint sensitivity analysis methodology (2nd-asam) for computing exactly and efficiently first- and second-order sensitivities in large-scale linear systems: I. Computational methodology. Journal of Computational Physics, 284(0):687 – 699, 2015.
- [17] Dan G Cacuci. Second-order adjoint sensitivity analysis methodology (2nd-asam) for computing exactly and efficiently first- and second-order sensitivities in large-scale linear systems: II. Illustrative application to a paradigm particle diffusion problem. Journal of Computational Physics, 284:700–717, 2015.
- [18] Dan G Cacuci, Mihaela Ionescu-Bujor, and Ionel Michael Navon. Sensitivity and uncertainty analysis, volume II: Applications to large-scale systems, volume 2. Chapman and Hall/CRC Press, 2005.
- [19] S. Chaturantabut and D.C. Sorensen. Nonlinear model reduction via discrete empirical interpolation. SIAM Journal on Scientific Computing, 32(5):2737–2764, 2010.
- [20] P. Courtier, J.N. Thepaut, and A. Hollingsworth. A strategy for operational implementation of 4D-VAR, using an incremental approach. Quarterly Journal of the Royal Meteorological Society, 120:1367–1387, 1994.
- [21] Philippe Courtier and Olivier Talagrand. Variational assimilation of meteorological observations with the adjoint vorticity equation. II: Numerical results. Quarterly Journal of the Royal Meteorological Society, 113(478):1329–1347, 1987.

- [22] Luca Daniel, Ong Chin Siong, Low Sok Chay, Kwok Hong Lee, and Jacob White. A multiparameter moment-matching model-reduction approach for generating geometrically parameterized interconnect performance models. Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on, 23(5):678–693, 2004.
- [23] M Dihlmann, M Drohmann, and B Haasdonk. Model reduction of parametrized evolution problems using the reduced basis method with adaptive time-partitioning. Proc. of ADMOS, 2011, 2011.
- [24] M. Dihlmann and B. Haasdonk. Certified PDE-constrained parameter optimization using reduced basis surrogate models for evolution problems. Submitted to the Journal of Computational Optimization and Applications, 2013.
- [25] Martin Drohmann and Kevin Carlberg. The ROMES method for statistical modeling of reduced-order-model error. SIAM/ASA Journal on Uncertainty Quantification, 3(1):116–145, 2015.
- [26] Qiang Du and Max Gunzburger. Model reduction by proper orthogonal decomposition coupled with centroidal Voronoi tessellation. In Proc. Fluids Engineering Division Summer Meeting, FEDSM2002-31051, ASME, 2002.
- [27] Alan Edelman, Tomas A. Arias, and Steven T. Smith. The geometry of algorithms with orthogonality constraints. SIAM Journal of MATRIX ANALYSIS and APPLICATION, 20(2):303–353, 1998.
- [28] Jens L Eftang and Benjamin Stamm. Parameter multi-domain ?hp?empirical interpolation. International Journal for Numerical Methods in Engineering, 90(4):412–428, 2012.
- [29] MS Eldred, AA Giunta, SS Collis, NA Alexandrov, and RM Lewis. Second-order corrections for surrogate-based optimization with model hierarchies. In Proceedings of the 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, Albany, NY., Aug, pages 2013–2014, 2004.
- [30] Geir Evensen. Data assimilation: the ensemble Kalman filter. Springer Science & Business Media, 2009.
- [31] Marco Fahl and Ekkehard W Sachs. Reduced order modelling approaches to PDE-constrained optimization based on proper orthogonal decomposition. In Large-scale PDE-constrained optimization, pages 268–280. Springer, 2003.
- [32] Ken-Ichi Funahashi. On the approximate realization of continuous mappings by neural networks. Neural networks, 2(3):183–192, 1989.
- [33] Shawn E Gano, John E Renaud, and Brian Sanders. Hybrid variable fidelity optimization by using a kriging-based scaling function. AIAA Journal, 43(11):2422–2433, 2005.
- [34] M.A. Grepl and A.T. Patera. A posteriori error bounds for reduced-basis approximations of parametrized parabolic partial differential equations. ESAIM: Mathematical Modelling and Numerical Analysis, 39(01):157–181, 2005.
- [35] Mircea Grigoriu. Stochastic systems: Uncertainty quantification and propagation. Springer Science & Business Media, 2012.
- [36] Bernard Haasdonk and Mario Ohlberger. Reduced basis method for finite volume approximations of parametrized linear evolution equations. ESAIM: Mathematical Modelling and Numerical Analysis-Modélisation Mathématique et Analyse Numérique, 42(2):277–302, 2008.
- [37] M. Hagan, H. Demuth, M. Beale, and O. De Jesus. Neural Network Design (2nd Edition). Martin Hagan, 2014.
- [38] S.S. Haykin. Neural Networks and Learning Machines. Number v. 10 in Neural networks and learning machines. Prentice Hall, 2009.
- [39] M. Hinze and S. Volkwein. Error estimates for abstract linear-quadratic optimal control problems using proper orthogonal decomposition. Computational Optimization and Applications, 39:319–345, 2008.
- [40] Michael Hinze and Stefan Volkwein. Proper orthogonal decomposition surrogate models for nonlinear dynamical systems: Error estimates and suboptimal control. In Dimension Reduction of Large-Scale Systems, pages 261–306. Springer, 2005.
- [41] Chris Homescu, Linda R Petzold, and Radu Serban. Error estimation for reduced-order models of dynamical systems. SIAM Journal on Numerical Analysis, 43(4):1693–1714, 2005.
- [42] H. Hotelling. Analysis of a complex of statistical variables with principal components. Journal of Educational Psychology, 24:417–441, 1933.
- [43] Peter L Houtekamer and Herschel L Mitchell. Data assimilation using an ensemble Kalman filter technique. Monthly Weather Review, 126(3):796–811, 1998.
- [44] Deng Huang, TT Allen, WI Notz, and RA Miller. Sequential kriging optimization using multiple-fidelity evaluations. Structural and Multidisciplinary Optimization, 32(5):369–382, 2006.

- [45] Dean L Isaacson and Richard W Madsen. Markov chains, theory and applications, volume 4. Wiley New York, 1976.
- [46] J.S.R. Jang, C.T. Sun, and E. Mizutani. Neuro-fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence. MATLAB curriculum series. Prentice Hall, 1997.
- [47] Eurika Kaiser, Bernd R Noack, Laurent Cordier, Andreas Spohn, Marc Segond, Markus Abel, Guillaume Daviller, Jan Östh, Siniša Krajnović, and Robert K Niven. Cluster-based reduced-order modelling of a mixing layer. Journal of Fluid Mechanics, 754:365–414, 2014.
- [48] E. Kalnay. Atmospheric modeling, data assimilation, and predictability. Cambridge Univ Pr, 2003.
- [49] K. Karhunen. Zur spektraltheorie stochastischer prozesse. Annales Academiae Scientarum Fennicae, 37, 1946.
- [50] K. Kunisch and S. Volkwein. Galerkin Proper Orthogonal Decomposition Methods for a General Equation in Fluid Dynamics. SIAM Journal on Numerical Analysis, 40(2):492–515, 2002.
- [51] Karl Kunisch and Stefan Volkwein. Galerkin proper orthogonal decomposition methods for parabolic problems. Numerische Mathematik, 90(1):117–148, 2001.
- [52] OP Le Maître and OM Knio. Spectral methods for uncertainty quantification. Scientific computation, 2010.
- [53] Thuan Lieu, Charbel Farhat, and Michel Lesoinne. POD-based aeroelastic analysis of a complete F-16 configuration: ROM adaptation and demonstration. AIAA Paper, 2295:2005, 2005.
- [54] Thuan Lieu and Michel Lesoinne. Parameter adaptation of reduced order models for three-dimensional flutter analysis. AIAA Paper, 888:2004, 2004.
- [55] M Lilley. Gaussian processes as neural networks. PhD thesis, Honours thesis, Victoria University of Wellington, 2004. Available from <http://www.mcs.vuw.ac.nz/people/Marcus-Frean>, 2004.
- [56] M.M. Loève. Probability Theory. Van Nostrand, Princeton, NJ, 1955.
- [57] E.N. Lorenz. Empirical Orthogonal Functions and Statistical Weather Prediction. Technical report, Massachusetts Institute of Technology, Dept. of Meteorology, 1956.
- [58] John Leask Lumley. The structure of inhomogeneous turbulent flows. Atmospheric turbulence and radio wave propagation, pages 166–178, 1967.
- [59] Kevin P Murphy. Machine learning: a probabilistic perspective. MIT press, 2012.
- [60] I Michael Navon, Xiaolei Zou, J Derber, and J Sela. Variational data assimilation with an adiabatic version of the nmc spectral model. Monthly weather review, 120(7):1433–1446, 1992.
- [61] IM Navon. A review of variational and optimization methods in meteorology. Variational methods in geosciences, pages 29–34, 1986.
- [62] Andrew Ng. Stanford CS229 , Machine Learning. Technical report, Stanford, 2008.
- [63] Ngoc-Cuong Nguyen, Gianluigi Rozza, and Anthony T Patera. Reduced basis approximation and a posteriori error estimation for the time-dependent viscous burgers’ equation. Calcolo, 46(3):157–185, 2009.
- [64] B.R. Noack, M. Schlegel, M. Morzynski, and G. Tadmor. System reduction strategy for galerkin models of fluid flows. International Journal for Numerical Methods in Fluids, 63(2):231–248, 2010.
- [65] A.T. Patera and G. Rozza. Reduced basis approximation and a posteriori error estimation for parametrized partial differential equations, 2007.
- [66] Benjamin Peherstorfer, Daniel Butnaru, Karen Willcox, and Hans-Joachim Bungartz. Localized discrete empirical interpolation method. SIAM Journal on Scientific Computing, 36(1):A168–A192, 2014.
- [67] TA Porsching. Estimation of the error in the reduced basis method solution of nonlinear equations. Mathematics of Computation, 45(172):487–496, 1985.
- [68] Christophe Prud’homme, Dimitrios V Rovas, Karen Veroy, Luc Machiels, Yvon Maday, Anthony T Patera, and Gabriel Turinici. Reliable real-time solution of parametrized partial differential equations: Reduced-basis output bound methods. Journal of Fluids Engineering, 124(1):70–80, 2002.
- [69] Alfio Quarteroni, Gianluigi Rozza, and Andrea Manzoni. Certified reduced basis approximation for parametrized partial differential equations and applications. Journal of Mathematics in Industry, 1(1):1–49, 2011.

- [70] M.L. Rapún and J.M. Vega. Reduced order models based on local POD plus Galerkin projection. Journal of Computational Physics, 229(8):3046–3063, 2010.
- [71] C.E. Rasmussen and C.K.I. Williams. Gaussian Processes for Machine Learning. Adaptive computation and machine learning series. University Press Group Limited, 2006.
- [72] C. W. Rowley, T. Colonius, and R. M. Murray. Model reduction for compressible flows using POD and Galerkin projection. Physica D. Nonlinear Phenomena, 189(1–2):115–129, 2004.
- [73] C.W. Rowley. Model Reduction for Fluids, using Balanced Proper Orthogonal Decomposition. International Journal of Bifurcation and Chaos (IJBC), 15(3):997–1013, 2005.
- [74] Gianluigi Rozza, Dinh Bao Phuong Huynh, and Anthony T Patera. Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations. Archives of Computational Methods in Engineering, 15(3):229–275, 2008.
- [75] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning internal representations by error propagation. Technical report, DTIC Document, 1985.
- [76] Pierre Sagaut. Large eddy simulation for incompressible flows: an introduction. Springer Science & Business Media, 2006.
- [77] O. San and T. Iliescu. Proper orthogonal decomposition closure models for fluid flows: Burgers equation. Technical Report arXiv:1308.3276 [physics.flu-dyn], Cornell University, August 2013.
- [78] Yoshikazu Sasaki. An objective analysis based on the variational method. Journal of Meteorological Society in Japan, 36(3):77–88, 1958.
- [79] Yoshikazu Sasaki, P Gu, and L Yan. A fundamental study of the numerical prediction based on the variational principle. Journal of Meteorological Society in Japan, 33(6):262–275, 1955.
- [80] Ryan Schmit and Mark Glauser. Improvements in low dimensional tools for flow-structure interaction problems: using global pod. In APS Division of Fluid Dynamics Meeting Abstracts, volume 1, 2003.
- [81] L. Sirovich. Turbulence and the dynamics of coherent structures. I. Coherent structures. Quarterly of Applied Mathematics, 45(3):561–571, 1987.
- [82] L. Sirovich. Turbulence and the dynamics of coherent structures. II. Symmetries and transformations. Quarterly of Applied Mathematics, 45(3):573–582, 1987.
- [83] L. Sirovich. Turbulence and the dynamics of coherent structures. III. Dynamics and scaling. Quarterly of Applied Mathematics, 45(3):583–590, 1987.
- [84] Marek Słoiński. Bayesian neural networks and gaussian processes in identification of concrete properties. Computer Assisted Mechanics and Engineering Sciences, 18(4):291–302, 2011.
- [85] Ralph C Smith. Uncertainty Quantification: Theory, Implementation, and Applications, volume 12. SIAM, 2013.
- [86] R. Stefanescu and I.M. Navon. POD/DEIM Nonlinear model order reduction of an ADI implicit shallow water equations model. Journal of Computational Physics, 237:95–114, 2013.
- [87] R. Stefanescu, I.M. Navon, Henry Fuelberg, and Max Marchand. 1D+4D–VAR Data Assimilation of lightning with WRFDA model using nonlinear observation operators. submitted to Monthly Weather Review, 2013.
- [88] Răzvan Ștefănescu, Adrian Sandu, and Ionel M Navon. Comparison of POD reduced order strategies for the nonlinear 2D shallow water equations. International Journal for Numerical Methods in Fluids, 76(8):497–521, 2014.
- [89] JA Taylor and MN Glauser. Towards practical flow sensing and control via pod and lse based low-dimensional tools. Journal of Fluids Engineering, 126(3):337–345, 2004.
- [90] Yannick Trémolet. Incremental 4d-var convergence study. Tellus A, 59(5):706–718, 2007.
- [91] Stefan Volkwein. Proper orthogonal decomposition: applications in optimization and control. CEA-EDFINRIA Numerical Analysis Summer School, 2007.
- [92] Nils P Wedi. Increasing horizontal resolution in numerical weather prediction and climate simulations: illusion or panacea? Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 372(2018):20130289, 2014.

- [93] Gary Weickum, MS Eldred, and Kurt Maute. Multi-point extended reduced order modeling for design optimization and uncertainty analysis. In Proceedings of the 47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference (2nd AIAA Multidisciplinary Design Optimization Specialist Conference), Newport, RI, pages 2006–2145, 2006.
- [94] DS Weile, E Michielssen, Eric Grimme, and K Gallivan. A method for generating rational interpolant reduced order models of two-parameter linear systems. Applied Mathematics Letters, 12(5):93–102, 1999.
- [95] David Wells, Zhu Wang, Xuping Xie, and Traian Iliescu. Regularized reduced order models. arXiv preprint arXiv:1506.07555, 2015.
- [96] Jeffrey S Whitaker and Thomas M Hamill. Ensemble data assimilation without perturbed observations. Monthly Weather Review, 130(7):1913–1924, 2002.
- [97] David C Wilcox et al. Turbulence modeling for CFD, volume 2. DCW industries La Canada, CA, 1998.
- [98] K. Willcox and J. Peraire. Balanced model reduction via the Proper Orthogonal Decomposition. AIAA Journal, pages 2323–2330, 2002.