AIAA AVIATION Forum
June 27-July 1, 2022, Chicago, IL & Virtual
AIAA AVIATION 2022 Forum

10.2514/6.2022-3488

A Multi-Fidelity Approximation of the Active Subspace Method for Surrogate Models with High-Dimensional Inputs

Bilal Mufti*, Mengzhen Chen†, Christian Perron‡, and Dimitri N. Mavris§
Georgia Institute of Technology, Atlanta, Georgia, 30332

Modern design problems routinely involve high-dimensional inputs and the active subspace has been recognized as a potential solution to this issue. However, the computational cost for collecting training data with high-fidelity simulations can be prohibitively expensive. This paper presents a multi-fidelity strategy where low-fidelity simulations are leveraged to extract an approximation of the high-fidelity active subspace. Both gradient-based and gradient-free active subspace methods are incorporated with the proposed multi-fidelity strategy and are compared with the equivalent single-fidelity method. To demonstrate the effectiveness of our proposed multi-fidelity strategy, the aerodynamic analysis of an airfoil and a wing are used to define two application problems. The effectiveness of the current approach is evaluated based on its prediction accuracy and training cost improvement. Results show that using a low-fidelity analysis to approximate the active subspace of high-fidelity data is a viable solution and can provide substantial computational savings, yet this is counterbalanced with slightly worse prediction accuracy.

I. Introduction

THE modern aerospace design process routinely involves tasks such as design space exploration, numerical optimization, and uncertainty quantification, which often require engineers to evaluate the performance of a large number of concepts. As such, computer simulations have been widely used to facilitate and accelerate the appraisal of multiple designs. Among these tools, high-fidelity simulations, such as computational fluid dynamics (CFD) and computational structural mechanics (CSM), play an important role in shape optimization, sensitivity analysis, or other tasks, as they provide a high level of physical accuracy and reduce the need for expensive and time-consuming physical experiments. However, these advanced simulations come with a steep computational cost as individual analyses can require hours, or even days, to complete on large computing servers. This high computational burden is particularly problematic in design tasks that need to query these simulations repeatedly. For the specific case of numerical optimization, some have successfully alleviated the computational cost of CFD or CSM by efficiently evaluating the gradient of the solutions using the adjoint method [1]. However, this process only results in a point solution and does not easily allow for *what-if* scenarios. The efficient evaluation of the gradient also requires specialized methods which are not always available, especially with legacy codes.

As an alternative to directly evaluating high-fidelity simulations in design tasks, practitioners often rely on surrogate modeling to avoid long evaluation time and to rapidly explore the design space. Such methods use a design of experiment (DOE) to sample the high-fidelity simulation over the design space and to capture the associated response of interest, usually some performance metric. Note that each sample is independent and can be evaluated in parallel to further accelerate the sampling process. A statistical model is then constructed to approximate the simulation results given some new design parameters. While some resources must be invested upfront to train the surrogate model, it is compensated by the near-instantaneous evaluation of new designs. This is especially beneficial in scenarios where the same model can be reused for many design tasks. The literature contains many examples of surrogate-based workflow applied to optimization, design space exploration, and uncertainty quantification. The reviews of Forrester and Keane [2], Queipo et al. [3], and Yondo et al. [4], on surrogate modeling are valuable sources of information on that subject.

However, the increasing complexity of today's design problems has prompted engineers to expand their design space with additional design parameters. In some problems, the number of design variables can easily become hundreds or thousands for a single or multi-objective design optimization problem [5–7]. As the number of design variables grows,

*Graduate Researcher, ASDL, Daniel Guggenheim School of Aerospace Engineering
†Graduate Researcher, ASDL, Daniel Guggenheim School of Aerospace Engineering
‡Postdoctoral Fellow, ASDL, Daniel Guggenheim School of Aerospace Engineering
§S.P. Langley Distinguished Regents Professor and Director of ASDL, Daniel Guggenheim School of Aerospace Engineering, AIAA Fellow

the hypervolume of the design space increases exponentially, and consequently, the number of samples to explore this space adequately rapidly becomes prohibitive. This is further exacerbated by the fact that many surrogate modeling techniques rely on the Euclidean distance to define the input-output relation, and this measure loses its effectiveness in high-dimensional space [8]. Such behavior is often referred to as the *curse of dimensionality* [9]. Oftentimes, the answer to this issue is to exploit the intrinsic structure of the high-dimensional data such that the *effective* or *latent* dimension of the problem is much smaller than initially observed.

The task of discovering the special underlying structure of the data is addressed by a category of techniques called dimensionality reduction methods, and these have been widely investigated in different fields during the past several decades [10, 11]. A common example of dimensionality reduction is the principal component analysis (PCA) method, which is closely related to the proper orthogonal decomposition (POD) and the Karhunen-Loève expansion. This technique has been used in previous work to identify dominant structures in turbulent flows [12] and to compress the high-dimensional field responses of high-fidelity simulation to generate reduced-order models (ROM) [13, 14]. However, dimensionality reduction techniques such as PCA are generally *unsupervised*, meaning that only the similarities between samples within the same dataset, i.e., either the input or the output datasets, are leveraged to identify its lower-dimensional representation. In the context of computer experiments, the input space is usually uniformly sampled such that the set of samples is mostly uncorrelated. Therefore, by construction, the dataset of design parameters does not have an exploitable structure by itself. Also, unsupervised methods do not take into account the relation existing between the input space to a model and its corresponding response. As such, the solution of an unsupervised dimensionality reduction might produce a sub-optimal representation for surrogate modeling. This flaw is addressed by instead using a *supervised* approach where the reduction of the input space is aimed specifically at capturing the interrelations between the inputs and outputs of a model.

One such supervised technique is the active subspace (AS) method which was first introduced by Constantine [15]. The AS method reduces the high-dimensional input space of a notional function to a linear subspace, the so-called active subspace, by identifying directions in which the variation of the function gradient is maximal. In other words, the gradient information is used to capture the relation between the input and output space. This method has been successfully applied to multiple engineering problems, which are later described in Sect. II.

Even though the classical AS method appears to be effective, its reliance on the function gradient can be problematic for problems where this information cannot be obtained easily or when the response is noisy. To overcome this issue, other authors have recently proposed a gradient-free method, referred to as manifold optimized active subspace (MOAS), where the discovery of the linear subspace is built into the training of the surrogate model [8, 16, 17], typically a Gaussian process. It was shown that the MOAS method could recover a linear subspace equivalent to the classical gradient-based approach [17]. The main drawback of MOAS over its gradient-based counterpart is the training of the model which is both more challenging and computationally expensive as it will be shown later.

While the AS methods, both gradient-based and gradient-free, are effective solutions to the curse of dimensionality, the discovery of the active subspace still requires a substantial amount of samples of the original model. Since these techniques are aimed at replacing higher fidelity simulations, collecting the training data can still be a computationally intensive task. As such, this work proposed a multi-fidelity approach where a lower fidelity simulation is leveraged to extract an approximation of the high-fidelity active subspace. The assumption is that a low-fidelity simulation, despite being less accurate than its high-fidelity counterpart, would share similar behavior and potentially have an aligned active subspace. Furthermore, since the low-fidelity simulation is cheaper, an approximate active subspace could be extracted at a relatively low cost.

This paper is organized as follows. In Section II, a brief review of the relevant literature on AS and multi-fidelity methods is presented. This is followed in Section III with a discussion on the numerical approach used in this work. Section IV then describes the two application problems, i.e., the aerodynamic analysis of an airfoil and a wing, which are used to evaluate the effectiveness of the proposed multi-fidelity approach. The results of the study are presented in Section V and a conclusion summarizes our finding in Section VI.

II. Literature Review

Different methods exist in the literature to address the curse of dimensionality for the construction of surrogate models. Variable reduction methods, such as sensitivity analysis [18] and automatic relevance determination [19] are the simplest methods that rank the inputs based on their influence on the outputs. These methods are applicable for problems where the correlation between various input variables is not significant. Unsupervised learning techniques, such as principal component analysis (PCA) [20] or diffusion maps [21], exploit instead the correlations existing between

inputs. Yet, the dimension reduction is carried out without accounting for the relationship between inputs and outputs. Hence, these techniques cannot be applied to problems where the input space is uniformly sampled, e.g., with a DOE, as the input samples have no inherent structure to exploit in that case. Also, an unsupervised approach can provide a sub-optimal solution for surrogate modeling as the latter is a supervised model.

Alternatively, the partial least squares (PLS) is a supervised dimension reduction technique that leverages the relation between inputs and outputs, both of which could be high-dimensional [22]. However, PLS assumes that the connection between inputs and outputs is linear, which limits its application to problems having a relatively low complexity [23]. Bouhlel et al. [24] developed a model combining Kriging and PLs which was coined as KPLS. The intent behind KPLS is to reduce the computational time needed to construct a regression model. This method reduces the number of hyper-parameters in a Kriging model to the number of principal directions found by PLS. The model has been applied to several industrial problems and works well for problems with up to 100 dimensions, but runs into memory issues if dimensions are further increased [25].

In comparison, the AS method, which is the focus of this work, is another example of a dimension reduction technique based on the work done by Constantine [15]. The classical AS formulation uses gradient information to create a linear subspace of the input space [15]. Unlike PLS, no assumption is made regarding the linearity of the input and output relation. Lukaczyk et al. [26] applied AS method to a high-dimensional aerodynamic shape optimization problem. The study found the presence of a low dimensional subspace in a complex high-dimensional problem and was able to successfully identify the aerodynamic shape of the wing that best reduces drag. However, the method required a large initial sample of designs. Gradient-based AS methods have also been used in several other engineering applications. For instance, Tezzele et al. [27] leveraged AS for a naval engineering shape design problem that estimated the wave resistance of a hull advancing in calm water. Constantine et al. [28] applied this approach to a hypersonic scramjet engine uncertainty quantification problem.

Obtaining gradient information for complex real-world engineering problems requires a substantial amount of computational time and resources. This can offset the low computational cost advantage provided by dimensionality reduction techniques. Gradient information is also difficult to acquire if the response of a model is noisy. To address these difficulties, other authors have proposed gradient-free variants of the AS method where the subspace is identified using only direct function evaluations. These methods generally rely on probabilistic techniques, such as Gaussian process regression (GPR), to extract the subspace that optimally captures the non-linear dependence between the input and output space. The first example of this is the work of Tripathy et al. [8] who proposed a probabilistic version of the AS method using GPR with built-in dimensionality reduction. In their approach, the projection matrix was treated as additional hyper-parameters of the GPR model to be estimated from the given data. Lewis et al. [29] instead proposed a gradient-free AS method that approximates the gradient of the function by using an adaptive Morris screening algorithm. Rajaram et al. [30] improved on Tripathy's et al. approach by simultaneously optimizing the projection matrix and the GPR parameters with the use of a product manifold, and by leveraging an efficient manifold optimization library. More recently, Gautier et al. [17] developed a fully Bayesian and gradient-free AS method that was able to quantify uncertainty in both the surrogate model parameters and the low dimensional subspace.

Constructing an active subspace, whether using the gradient-based or gradient-free approach, still requires a substantial number of high-fidelity simulations [31]. When involving complex physics-based simulation, the upfront sampling cost for the AS discovery can be a significant computational burden. Multi-fidelity modeling or variable fidelity modeling addresses this issue by using one or multiple numerical simulations of lower fidelity to approximate the function of interest [32]. This reduces the number of high-fidelity simulations required, and hence, the computational cost to predict an output function of interest. The pioneering work for carrying out predictions using multi-fidelity simulations was done by Kennedy and O'Hagan [33]. In this work, the expensive to run high-fidelity simulation code was represented by a low fidelity approximation and a discrepancy term, which were modeled as independent Gaussian processes. Han and Görtz [34] also developed a slightly simpler multi-fidelity model, named hierarchical Kriging, to predict the aerodynamic data of an airfoil and an industrial transport aircraft model. Several similar multi-fidelity techniques were used for Bayesian optimization [35], uncertainty propagation [36], sensitivity analysis [37] and discovering response surfaces for dynamical systems [38].

Multi-fidelity techniques can also be combined with dimension reduction techniques to reduce the computational cost of extracting the active subspace. For instance, Lam et al. [39] created a multi-fidelity estimator for the covariance matrix used in the gradient-based AS calculation which uses both low- and high-fidelity simulation data. This formulation significantly reduces the number of gradient evaluations required to create an active subspace. Although, this method still depends on the availability of gradient information. Alternatively, Tsilifis et al. [31] developed a multi-fidelity and gradient-free AS technique that leverages an autoregressive GPR. As with the method of Tripathy et al. [8], the

projection matrix and the GPR are identified jointly, which Tsilifis et al. achieve with a Bayesian approach.

In this work, we present a computationally cheap and relatively simpler approach for combining dimension reduction techniques with the concept of multi-fidelity simulations. Our proposed technique computes the active subspace using the low fidelity simulation data only. This active subspace is then used as a low-dimensional representation of high-fidelity data. The approach assumes that the low fidelity active subspace is a good substitute for the high-fidelity active subspace. Constructing the active subspace in this manner saves considerable computational cost and time to create surrogate models and the need for running the high-fidelity simulation is minimized.

III. Methodology

The following section presents the numerical methods required to compute both the classical and gradient-free AS solutions. This is followed by a description of the proposed multi-fidelity approach to approximate the active subspace of a model.

A. Gaussian Process Regression

Gaussian process regression (GPR), which is equivalent to Kriging in other fields, is a commonly used surrogate modeling technique that allows the prediction of model uncertainty along with the desired output function [40]. A GPR model works by defining a probability distribution on the given function, and this distribution is consistent with the user's prior beliefs regarding the function behavior. The GPR model leverages Bayes' theorem to combine the prior beliefs with observed function values and infer a posterior probability distribution. The mean or median of the posterior distribution acts as a surrogate model that can be used to predict new outputs at unknown points in the design space. A high-level explanation of GPR is provided in this section, which is based on the formulation introduced in Rasmussen's book [41].

Consider a Gaussian distributed function $f : \mathbf{x} \mapsto y$ which maps inputs $\mathbf{x} \in \mathbb{R}^d$ to scalar outputs $y \in \mathbb{R}$. The function can be represented as a Gaussian process of the form $f \sim N(\mu, \mathbf{\Sigma})$, where μ is the mean and $\mathbf{\Sigma}$ is the covariance matrix. Note that the observations $f(\mathbf{x})$ could be affected by the presence of uncertainty or noise. We can model this noise function as a zero-mean Gaussian white noise given as $\delta \sim N(0, \sigma_n^2)$. With the addition of noise, the Gaussian process can be represented by $f \sim N(\mu, \mathbf{\Sigma} + \sigma_n^2 \mathbf{I})$. We can assume that the observed data is centered such that the mean of our prior distribution is $\mu = \mathbf{0}$.

The elements of the covariance matrix $\mathbf{\Sigma}$ can be constructed using a positive-definite kernel function which correlates responses between outputs $f(\mathbf{x})$ and $f(\mathbf{x}')$ based on their corresponding input locations $\mathbf{x}$ and $\mathbf{x}'$. Different choices of kernel functions are available in the literature and this study relies on the widely used Gaussian kernel with automatic relevance determination (ARD) defined as

$$k(\mathbf{x}, \mathbf{x}') = \sigma_r \prod_{j=1}^{b} \exp\left(\frac{(x_j - x_j')^2}{2\theta_j^2}\right) \tag{1}$$

Here, σ_r is the process variance and θ_j are the length scales of the kernel function determined during the training process. For two input design matrices $\mathbf{X}$ and $\mathbf{X}'$ containing n and m observations respectively, the kernel matrix can be written as

$$\mathbf{K}_{\mathbf{X}\mathbf{X}'} = \begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1') & k(\mathbf{x}_1, \mathbf{x}_2') & \dots & k(\mathbf{x}_1, \mathbf{x}_m') \\ k(\mathbf{x}_2, \mathbf{x}_1') & k(\mathbf{x}_2, \mathbf{x}_2') & \dots & k(\mathbf{x}_2, \mathbf{x}_m') \\ \vdots & \vdots & \ddots & \vdots \\ k(\mathbf{x}_n, \mathbf{x}_1') & k(\mathbf{x}_n, \mathbf{x}_2') & \dots & k(\mathbf{x}_n, \mathbf{x}_m') \end{bmatrix}$$

The covariance matrix $\mathbf{\Sigma}$ can now be constructed using the above kernel matrix. By recalling that the underlying process f is Gaussian in nature, the predictions $f(\mathbf{X}^*)$ at the unknown locations $\mathbf{X}^*$ can be made using the joint probability

distribution

$$\begin{bmatrix} f(\mathbf{x}_1) \\ \vdots \\ f(\mathbf{x}_n) \\ f(\mathbf{X}^*) \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \mathbf{K_{XX}} + \sigma_n^2\mathbf{I} & \mathbf{K_{XX^*}} \\ \mathbf{K_{X^*X}} & \mathbf{K_{X^*X^*}} + \sigma_n^2\mathbf{I} \end{bmatrix} \right)$$

The posterior distribution for $f(\mathbf{X}^*)$ can be determined analytically using the Gaussian assumption and is given as

$$f(\mathbf{X}^*)|f(\mathbf{X}) \sim N(\mu^*, \mathbf{\Sigma}^*) \tag{2}$$

where μ^* and $\mathbf{\Sigma}^*$ are the posterior mean and variance respectively, and are given by

$$\mu^* = \mathbf{K_{X^*X}}(\mathbf{K_{XX}} + \sigma_n^2\mathbf{I})^{-1} f(\mathbf{x}) \tag{3}$$

$$\mathbf{\Sigma}^* = (\mathbf{K_{X^*X^*}} + \sigma_n^2\mathbf{I}) - (\mathbf{K_{X^*X}})(\mathbf{K_{XX}} + \sigma_n^2\mathbf{I})^{-1}\mathbf{K_{XX^*}} \tag{4}$$

Note that the above posterior distribution depends on the hyper-parameters of the GPR model, which are θ_j, σ_r, and σ_n. For simplicity, they can be combined into a single vector $\boldsymbol{\gamma} = [\theta_1, \ldots, \theta_n, \sigma_r, \sigma_n]^T$. Finding the correct values for these hyper-parameters is the main challenge behind the training of a GPR model. Arguably, the most popular way of determining $\boldsymbol{\gamma}$ is by the maximum likelihood estimator (MLE) approach. This is achieve by maximizing the likelihood function of the GPR, which is defined as follow

$$\mathcal{L}(\boldsymbol{\gamma}|\mathbf{X}, y) = \log p(f(\mathbf{X})|\boldsymbol{\gamma}) = \left[-\frac{1}{2} f(\mathbf{x})^T \mathbf{K_{XX}} f(\mathbf{x}) - \frac{1}{2}\log(\det(\mathbf{K_{XX}} + \sigma_n^2\mathbf{I})^{-1} - \frac{n}{2}\log(2\pi) \right] \tag{5}$$

and the optimization problem is given by

$$\hat{\boldsymbol{\gamma}} = \arg\max_{\boldsymbol{\gamma}} \ \mathcal{L}(\boldsymbol{\gamma}|\mathbf{X}, y) \tag{6}$$

where $\hat{\boldsymbol{\gamma}}$ is the MLE solution of the hyper-parameters. Note, the above optimization problem generally has multiple local minima. As such, this optimization problem can either be solved with a global optimization or with a gradient-based optimization with multiple restarts. This work uses the latter.

B. Gradient-Based Active Subspace

The following section summarizes the classic active subspace method using gradient information. To distinguish it from the gradient-free approach described later, this method is referred to as G-AS moving forward. A more detailed discussion of gradient-based AS can be found in [15]. Let us assume the dimension of the input space is d and there is n pairs of input and output measurements available such that

$$\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]^T \tag{7}$$

$$\mathbf{Y} = [y_1, \ldots, y_n]^T \tag{8}$$

where $\mathbf{x}_i \in \mathbf{R}^d$ is a high-dimensional input and $f(\mathbf{x}_i) = y_i$ is the corresponding output. The gradient of y_i with respect to $\mathbf{x}_i$ is denoted as $\nabla f_{\mathbf{x}}$. We then define a matrix $\mathbf{C}$, which is the average of the outer product of the gradient with itself

$$\mathbf{C} = \int (\nabla f_{\mathbf{x}})(\nabla f_{\mathbf{x}})^T p(\mathbf{x})\, dx \tag{9}$$

Note that the exact integration of Eq. (9) is very difficult to obtain. A practical solution is to approximate it with a finite number of gradients values measured through sample points such that

$$\mathbf{C} \approx \frac{1}{N}\sum_{i=1}^{N} \nabla f_{\mathbf{x},i} \left(\nabla f_{\mathbf{x},i}\right)^T \tag{10}$$

Since matrix $\mathbf{C}$ is symmetric positive definite, it can be diagonalized as

$$\mathbf{C} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^T \tag{11}$$

In the above equation, $\boldsymbol{\Lambda} = diag(\lambda_1, \lambda_1, \ldots, \lambda_d)$ is a diagonal matrix whose elements are the eigenvalues of $\mathbf{C}$ organized in descending order, i.e., $\lambda_1 \geq \cdots \geq \lambda_d \geq 0$. Also, $\mathbf{Q} = [\mathbf{w}_1, \ldots, \mathbf{w}_d]$ is a orthogonal matrix where the i-th column $\mathbf{w}_i$ is the eigenvector corresponding to λ_i. As higher eigenvalue means a greater variations in the response gradient, the AS approach suggests to separate the largest k eigenvalues with its corresponding eigenvectors such that

$$\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\Gamma} & 0 \\ 0 & \boldsymbol{\Gamma}_\perp \end{bmatrix}, \quad \mathbf{Q} = \begin{bmatrix} \mathbf{W} & \mathbf{W}_\perp \end{bmatrix} \tag{12}$$

The selected k eigenvectors $\{\mathbf{w}_i \,|\, i = 1, \ldots, k\}$ form the active subspace while the remaining $d - k$ eigenvectors are defined as the inactive subspace. The dimension k of the AS is usually determined by examining the eigenvalue decay and selecting a threshold value where a sharp drop is observed. With the above definition, the original function $f(\mathbf{x})$ can be written as

$$f(\mathbf{x}) = f\left(\mathbf{W}\mathbf{z} + \mathbf{W}_\perp \mathbf{z}_\perp\right) \tag{13}$$

where $\mathbf{z} = \mathbf{W}^T\mathbf{x}$ are the components of the input vector within the AS while $\mathbf{z}_\perp = \mathbf{W}_\perp^T\mathbf{x}$ are the components perpendicular to it. By neglecting the effect of $\mathbf{z}_\perp$, Eq. (13) can be approximated by:

$$f(\mathbf{x}) \approx g\left(\mathbf{W}^T\mathbf{x}\right) \tag{14}$$

where the function g is the link function connecting $\mathbf{z}$ with the response y. It can be statistically defined as the conditional expectation of f over the inactive subspace given a position in the AS

$$g(\mathbf{z}) = \mathbb{E}_{\mathbf{z}_\perp|\mathbf{z}}\left[f(\mathbf{W}\mathbf{z} + \mathbf{W}_\perp\mathbf{z}_\perp)\right] \tag{15}$$

C. Gradient-Free Active Subspace

To obtain the AS solution without any gradient information, this work uses the GPR with a built-in dimensionality reduction method as first proposed by Tripathy et al. [8]. In this method, the link function $g(\mathbf{z})$ from Eq. (14) is defined as a GPR model with hyper-parameters $\boldsymbol{\gamma}$. The projection matrix $\mathbf{W}$ is then extracted by including it as an extra hyper-parameter of the model, and its value is similarly obtained with the MLE approach. In other words, the optimal AS correspond to the matrix which maximizes, jointly with the hyper-parameters of the GPR, the following likelihood function

$$\mathcal{L}(\boldsymbol{\gamma}, \mathbf{W}|\mathbf{X}, y) = \log p\left(g\left(\mathbf{W}^T\mathbf{x}\right)\middle|\boldsymbol{\gamma}, \mathbf{W}\right) \tag{16}$$

such that the optimal $\hat{\mathbf{W}}$ and $\hat{\boldsymbol{\gamma}}$ are given by

$$\hat{\boldsymbol{\gamma}}, \hat{\mathbf{W}} = \arg\max_{\boldsymbol{\gamma}, \mathbf{W}} \mathcal{L}(\boldsymbol{\gamma}, \mathbf{W}|\mathbf{X}, y) \tag{17}$$

Tripathy et al. [8] performed the likelihood optimization by first optimizing $\mathbf{W}$ while keeping $\boldsymbol{\gamma}$ constant. Once a better estimate of $\mathbf{W}$ is obtained, then the algorithm proceeds with optimizing $\boldsymbol{\gamma}$ while freezing the current guess for $\mathbf{W}$. These two steps are repeated until the algorithm converges to an optimum. Note that $\mathbf{W}$ is an orthonormal matrix and that to preserve the orthogonality of the AS, the search over $\mathbf{W}$ must be restricted to the Stiefel manifold, i.e., the manifold of all orthonormal matrices. In their approach, the orthogonality of $\mathbf{W}$ was enforced using a Crank-Nicholson-like update based on the Cayley transformation.

In an attempt to improve the method of Tripathy et al., Rajaram et al. [30] proposed a strategy that leverages a state-of-the-art manifold optimization library named *pymanopt* [42]. Moreover, the authors have combined the optimization of both $\mathbf{W}$ and $\boldsymbol{\gamma}$ under a single step. This is achieved by defining a product manifold that joins the Euclidean space of the GPR hyper-parameters with the Stiefel manifold of the AS basis. This method was referred to as the manifold optimized active subspace (MOAS). Note that with this approach, the total number of parameters to optimize d_{opt} is

$$d_{\text{opt}} = \underbrace{2 + k}_{\boldsymbol{\gamma}} + \underbrace{d \times k - \frac{k(k+1)}{2}}_{\mathbf{W}} \tag{18}$$

where the first two terms are the number of unknown GPR hyper-parameters and the last two terms are the dimension of the Stiefel manifold of all orthonormal basis in $\mathbb{R}^{d\times k}$. This optimization process is further accelerated by using the *Jax* [43] library as a backend for fast linear algebra and automatic differentiation. This allows for the efficient computation of gradients for the maximization of Eq. (16). Since the likelihood function is non-convex, the optimization must be restarted multiple times with different initial guesses to avoid getting trapped into a local minimum. This approach was later reused by Gautier et al. [17] and the authors have made their framework publicly available*. This work uses the MOAS method and leverages the code developed by Gautier et al.

D. Proposed Approach

The methods aforementioned focus on building the surrogates from an algorithmic point of view. Both gradient-based and gradient-free AS methods have no requirement on the fidelity of the training data. However, former applications [26, 28, 44] of AS are mostly using high-fidelity simulations as training data. Though the number of cases needed and thus the computational time has been greatly reduced, the computational burden for even a single high-fidelity simulation still exists. In this section, we are exploring the possibility of further reducing computation time by generating training data in a multi-fidelity way.

As discussed in Sect. II, multi-fidelity strategies have been identified previously by other authors. Namely, Lam et al. [39] derived a multi-fidelity covariance matrix estimator for the gradient-based AS method, and Tislifis et al. [31] presented an autoregressive extension of the gradient-free AS methods. In both cases, the multi-fidelity AS is uncovered with a mixture of high- and low-fidelity samples.

This work proposes an alternative multi-fidelity strategy to complement the currently existing ones mentioned above. The proposed approach builds the surrogates in two steps:

1) Use a low-fidelity model to generate training data and extract the low-fidelity active subspace, i.e., the projection matrix $\mathbf{W}$ from Eq. 14.
2) Generate additional training samples using the high-fidelity simulation. Then, train the link function $g(\mathbf{z})$ using the high-fidelity samples and the low-fidelity matrix $\mathbf{W}$.

The proposed strategy is based on the assumption that the low-fidelity simulation, despite being less accurate, has an AS that is relatively aligned to the one for the high-fidelity simulation. As such, this relation is exploited by finding the model AS using exclusively the low-fidelity samples. This simple multi-fidelity strategy can be advantageous as each high-fidelity sample exactly resides in the AS, which facilitates the training of the link function $g(\mathbf{z})$. Furthermore, the proposed approach applies to either gradient-based or gradient-free AS methods. The implementation of this strategy is also easier as it mostly reuses existing single-fidelity AS methods.

Compared to the multi-fidelity methods of Lam et al. and Tsilifis et al., the current approach is simpler and possibly more practical, although its effectiveness remains to be seen. As such, the remainder of this work is dedicated to testing and comparing the proposed multi-fidelity strategy with existing AS methods. The proposed workflow will be applied to both 2D airfoil and 3D wing shape optimization problems. Note that in both cases, the low-fidelity active subspace (step 1) will be computed for both gradient-based and gradient-free approaches.

IV. Application Problems

To demonstrate the effectiveness of our proposed multi-fidelity AS strategy, we apply it to two application problems which are detailed in the following section. This is followed by a description of the metrics used to quantify the accuracy and costs of our proposed method compared to current AS techniques.

A. Problem 1: RAE 2822 Airfoil

The first application problem considered in this work is the aerodynamic analysis of a transonic airfoil. Specifically, the geometry used corresponds to the RAE 2822 and is analyzed with compressible two-dimensional CFD simulations. Despite being a relatively simple problem, the CFD simulation of this airfoil still presents some of the physical challenges of more complex simulation involving three-dimensional geometries such as boundary layers, shock waves, flow separation, etc. It is worth noting that the RAE 2822 is also the starting geometry for the second benchmark problem of the AIAA aerodynamic design optimization group (ADODG). As such, many authors have used this geometry to quantify the performance of various optimization methods [45–47]. For this work, the CFD simulations are performed

*https://gitlab.com/raphaelgautier/bayesian-supervised-dimension-reduction

using the open-source SU2 code [†] [48]. The simulations are based on the Reynolds-averaged Navier-Stokes (RANS) equations and the effect of turbulence is modeled with the Spalart-Allmaras turbulence model [49]. Furthermore, the steady-state flow is resolved using a backward Euler scheme, and the Jameson-Smith-Turkel (JST) scheme [50] is used for the spatial discretization of the domain. Unless otherwise specified, the flight conditions considered in this study is a free-stream Mach number $M_\infty = 0.73$ and a Reynolds number $Re_\infty = 6.5 \times 10^6$. For the high-fidelity results, an unstructured grid of 96,913 nodes is used to represent the fluid domain around the airfoil and a close-up visualization of the grid used is shown in Fig. 1. The far-field boundary of the domain is located at a distance of roughly 100 chord lengths. As for the low-fidelity results, two additional grids with a coarser resolution are considered and are referred to as L1 and L2. These grids have a size of 7,485 and 19,062 nodes respectively, as detailed in Table 1. Also, unlike the other fidelity levels, the L1 grid is solved using an inviscid solver. As such, it does not have the near-wall refinement typically required to capture boundary layers, which explains its much smaller size compared to the L2 and L3 grids.

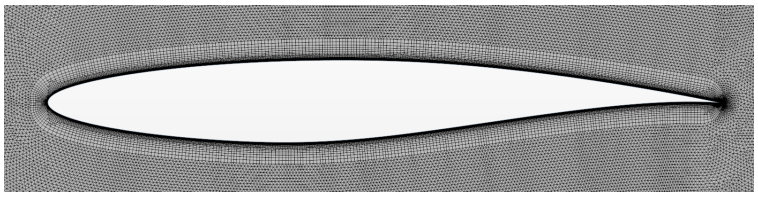


Fig. 1 Close-up of the high-fidelity L3 grid (96,913 nodes) of the RAE 2822 airfoil.

A grid convergence study has been conducted for the RAE 2822 test case, and the results are shown in Fig. 2. The results demonstrate that the lift and drag coefficients converge to an asymptotic result for mesh sizes roughly over 1×10^5 nodes which support our grid selection for the high-fidelity simulation.

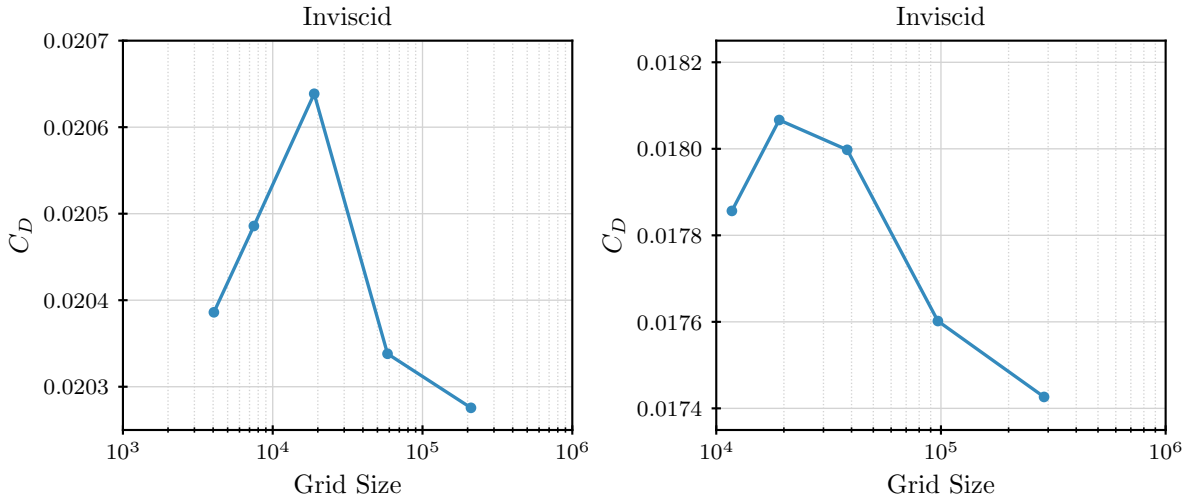


Fig. 2 Grid convergence study of the RAE 2822 airfoil ($M_\infty = 0.73$, $Re_\infty = 6.5 \times 10^6$, and $\alpha = 2.79°$).

B. Problem 2: CRM Wing

To complement the airfoil application problem, a second test case involving the aerodynamic analysis of a commercial aircraft wing is also considered. This allows us to evaluate the effectiveness of the proposed method on a scale more representative of an industrial application. In this problem, the wing geometry is extracted from NASA's common research model (CRM) [51] and is shown in Fig. 3. The CRM aircraft geometry is publicly available and has been specifically designed for academia as a challenging, yet realistic, problem to test and develop aerodynamic prediction methods. This work focuses solely on the wing of the CRM as it is the component responsible for the majority of the flow complex features. Similar to the RAE 2822, the CRM wing is also one of the ADODG benchmark problems and

[†]This work is using the release 7.2.0 of the SU2 code.

Table 1 Summary of RAE 2822 CFD simulations utilized for different fidelity levels.

Level	d	Grid Size [Num. Nodes]	Solver	Cost/Sample* [CPU-hr]	
				Flow Solver	Adjoint Solver
L1	40	7,485	Inviscid	0.125	0.147
L2	40	19,062	RANS	0.204	0.204
L3	40	96,913	RANS	1.776	2.483

*CPU times are obtained using Intel Xeon Gold 6226 CPUs

has been used by other authors in the context of aerodynamic shape optimization [6, 45, 46]. In fact, the same flow conditions as defined by the ADODG, i.e., $M_\infty = 0.85$ and $Re_\infty = 5 \times 10^6$, are considered for this study. For the CFD setup, the CRM wing simulation uses the same configuration as the RAE 2822 with the exception that the flow is now three-dimensional.

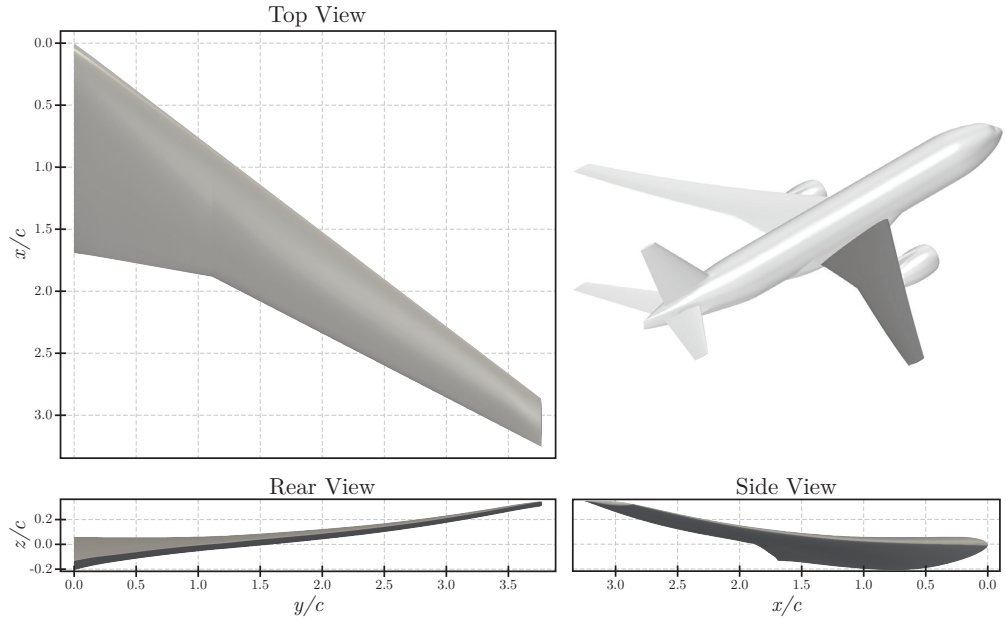


Fig. 3 Top, rear, and side view of the CRM wing as defined by the ADODG. The dimensions provided are normalized by the mean aerodynamic chord whose value is 275.8 in.

A multi-block structured grid of 3.7×10^6 nodes as shown in Fig. 4, is used to produce the high-fidelity results. Note, the grid used in this work was originally generated by the Multidisciplinary Design Optimization Laboratory [52] of the University of Michigan and was made public as part of the ADODG benchmark problem. This grid was also part of a previous study by Lyu et al. [6] in which a multi-level RANS-based optimization method was proposed. As for the low-fidelity results, a RANS-based CFD simulation with a coarse L2 grid and an inviscid simulation with an even coarser L1 grid are considered. The grid size for each level is described in Table. 2. Similar to the 2D airfoil test case, the result of a grid convergence study for the CRM wing problem is summarized in Fig. 5. In this study, the grid size of the midpoint in Fig. 5 is selected as the resolution for high-fidelity simulations.

C. High-Dimensional Parametrization

For both the airfoil and wing application problems, this work is evaluating the aerodynamic performance of these geometries as a function of their respective shape. Such aerodynamic shape can be relatively complex and many parameters are required to adequately meet some design objectives. For instance, Lyu et al. [6] used 720 design variables

Table 2 Summary of CRM wing CFD simulations utilized for different fidelity levels.

Level	d	Grid Size [Num. Nodes]	Solver	Cost/Sample* [CPU-hr]	
				Flow Solver	Adjoint Solver
L1	240	251,821	Inviscid	1.162	1.284
L2	240	464,325	RANS	3.384	10.07
L3	240	3,659,337	RANS	11.81	59.85

*CPU times are obtained using Intel Xeon Gold 6226 CPUs

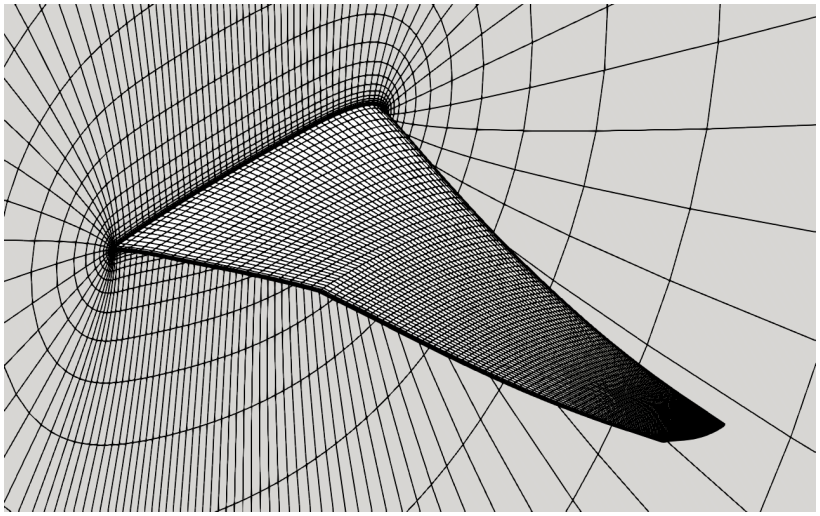


Fig. 4 CFD grid of the CRM wing surface and the symmetry plane. For visualization purposes, the coarser L2 grid is shown instead of the high-fidelity L3 grid.

to control the shape of the CRM wing for their optimization problem. Therefore, aerodynamic shape parametrization represents a scenario in which the input dimension is large and the AS method can be very beneficial.

There exist multiple methods to parametrize the shape of aerodynamic bodies such as splines [53], Hick-Hennes bump functions [54], and the class shape-function transformation (CST) method [55]. In this work, the geometry of the RAE 2822 airfoil and the CRM wing is controlled using a free-form deformation (FFD) approach [56]. Simply speaking, the FFD method encases a geometry into a lattice, which is often referred to as an FFD box. A mapping is then established between the FFD box and the geometry within such that moving the nodes of the box results in a deformation of the shape. With this technique, a complex shape can be easily modified without the need for CAD tools. Also, the granularity of the parametrization is easily controlled by the number of FFD nodes in the lattice. Note that this method only controls the shape of the RAE 2822 and CRM wing surface grid points. Once those shapes are altered, their deformation must be propagated through the rest of the volume grid. This is achieved using a linear elasticity approach [57], and the tools required for this are all included in the SU2 code suite.

For the RAE 2822 airfoil, a lattice of 20×2 is considered as shown in Fig. 6a. The FFD nodes are limited to moving in the vertical direction only, and their displacement is bounded by ± 0.03 chord length. Therefore, the dimension of the input space for the airfoil problem is 40. As for the CRM wing, the lattice is bigger with a shape of $15 \times 8 \times 2$ as shown in Fig. 6b. The FFD node displacement is also restricted to the vertical axis, and their movement amplitude is limited to ± 0.03 times the reference chord length. This brings the input space dimension of the CRM wing problem to 240.

D. Performance Metrics

To assess the effectiveness of the proposed multi-fidelity AS strategy, this work quantifies the accuracy and training cost improvement of the multi-fidelity approach when compared to an equivalent single-fidelity AS method. Specifically,

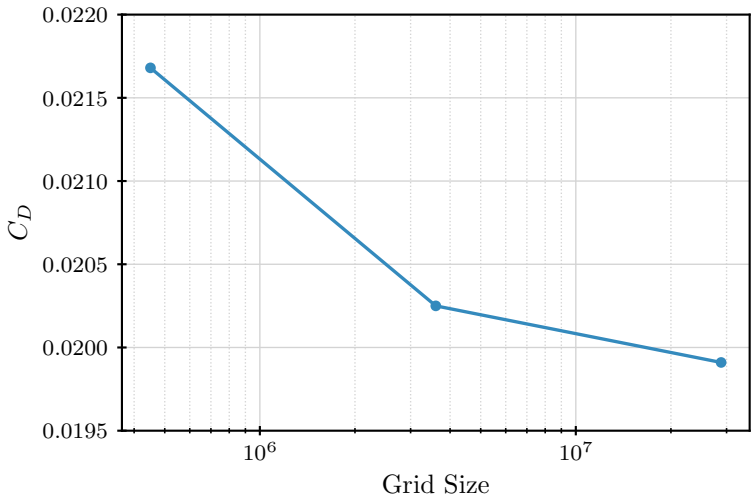
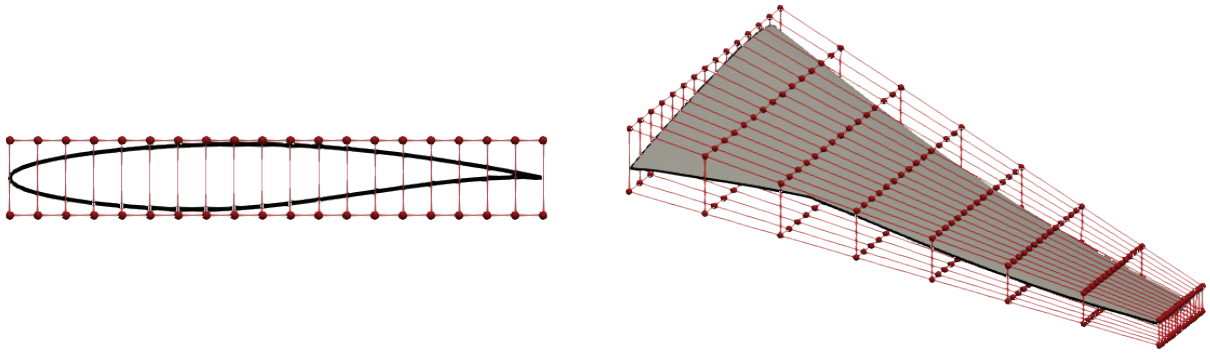


Fig. 5 Grid convergence study for the CRM wing ($M_\infty = 0.85$, $Re_\infty = 5 \times 10^6$, and $C_L = 0.5$).



a) RAE 2822 b) CRM wing

Fig. 6 FFD box used for the RAE 2822 airfoil and the CRM wing problems.

for the multi-fidelity strategy to be worthwhile, it must offer an equivalent model accuracy at a lower computational cost, or equivalently, provide a lower error for the same cost. A set of N samples were used to compute the AS, the number of samples used to train the surrogate model is given as n_t As such, the following section details the metrics used in this single- and multi-fidelity comparison.

1. Subspace Angles

Following Gautier et al. [17] example, the accuracy of the computed AS can be assessed by measuring its principal angles with respect to a reference AS. Let us consider, two matrices $\mathbf{A}$ and $\mathbf{B}$ whose column-spaces are the subspaces $\mathcal{A}$ and $\mathcal{B}$ in $\mathbb{R}^d$. As shown in [58], the principal angles' are defined recursively as

$$\cos(\theta_i) = \max_{\mathbf{a}\in\mathcal{A}} \max_{\mathbf{b}\in\mathcal{B}} \mathbf{a}_i^T \mathbf{b}_i \tag{19}$$

subject to

$$\|\mathbf{a}_i\| = \|\mathbf{b}_i\| = 1, \quad \mathbf{a}_i^T \mathbf{a}_j = 0, \quad \mathbf{b}_i^T \mathbf{b}_j = 0, \quad j = 1, \ldots, i-1 \tag{20}$$

where $\mathbf{a}_i$ and $\mathbf{b}_i$ are unit vectors in the subspaces $\mathcal{A}$ and $\mathcal{B}$ respectively. Note that the solution of Eq. 19 can be obtained with an SVD-based algorithm [58]. In the current context, the matrix $\mathbf{A}$ would be the computed AS basis while $\mathbf{B}$ would be a reference orthonormal basis. Ideally, the reference basis should be the exact AS, but since there exists no analytical solution for the current application problems, this is approximated by the AS solution computed with a large number of high-fidelity samples.

Given two subspaces $\mathcal{A}$ and $\mathcal{B}$, there exist a number of principal subspace angles equal to the smallest dimension between $\mathcal{A}$ and $\mathcal{B}$. However, we focus on two angles in particular: the smallest principal subspace angle (SPSA) and the largest principal subspace angle (LPSA). The former represents the best possible alignment between vectors in two subspaces, and the latter is indicative of the worst possible alignment.

2. Root Mean Squared Error

In addition to the accuracy of the computed AS, one can also consider the accuracy of the resulting surrogate model built with this active subspace. As such, using a set of n_v verification samples, i.e., samples not used during the training process, the root-mean-squared prediction error E of a given model can be estimated with

$$E(C_D) = \sqrt{\frac{1}{n_v}\sum_{i=1}^{n_v}\left(y_i^* - \tilde{y}_i^*\right)^2} \tag{21}$$

where y_i^* is an exact solution of the analysis used for verification and $\tilde{y}_i^*$ is the corresponding prediction from the surrogate model. Note that for both the airfoil and wing test cases, the quantity of interest being predicted is the drag coefficient C_D. Therefore, the prediction error is denoted $E(C_D)$ to explicitly identify which quantity the error is based on.

3. Mean Absolute Percentage Error

We are interested in knowing if an AS determined using low-fidelity data is a good approximation to the one obtained using the high-fidelity data. Mean absolute percentage error (MAPE) gives the relative error between two link functions $g(\tilde{\mathbf{z}}_{\mathrm{HF}})$ and $g(\tilde{\mathbf{z}_{\mathrm{LF}}})$

$$\mathrm{MAPE} = \frac{1}{n_v}\sum_{i=1}^{n_v}\frac{\left|g(\tilde{\mathbf{z}}_{\mathrm{HF}}) - g(\tilde{\mathbf{z}}_{\mathrm{LF}})\right|}{\left|g(\tilde{\mathbf{z}}_{\mathrm{HF}})\right|} \tag{22}$$

where $\tilde{\mathbf{z}}_{\mathrm{HF}} = \mathbf{W}_{\mathrm{HF}}^T\mathbf{x}$ is obtained using a design vector $\mathbf{x}$ and the high-fidelity AS $\mathbf{W}_{\mathrm{HF}}$, while $\tilde{\mathbf{z}}_{\mathrm{LF}} = \mathbf{W}_{\mathrm{LF}}^T\mathbf{x}$ is computed using the low-fidelity AS $\mathbf{W}_{\mathrm{LF}}$ instead.

4. Model Training Cost

Generally speaking, the computational cost for training a surrogate model can be divided between the cost of sampling the original model and the cost of making the surrogate model using the gathered data. Here, it is implicitly assumed that the training data are not already available and that many simulations must be performed beforehand. Depending on the type of surrogate model and the computer hardware used, creating a surrogate model from the training data can take anywhere between less than a minute to a handful of hours. In comparison, a single CFD simulation can need multiple hours to complete even when using multiple CPUs. Therefore, because the AS is applied to expensive high-fidelity simulations, the training cost of the AS is mostly reflected by the cost of generating all the training samples. Also, for multi-fidelity methods, this cost combines the computational resources used to generate both the low- and high-fidelity data. Since each CFD simulation is executed using multiple CPUs, and multiple simulations can be performed simultaneously, the computational cost can be more easily represented in terms of the total CPU time, which is usually measured in CPU-hr. This metric represents the total amount of computational time used by each processing unit involved. For both the airfoil and wing test cases, the average computational cost of each sample is listed in Tables 1 and 2 respectively. Note that for scenarios requiring gradient information, the computational cost of both the flow and adjoint solvers must be taken into account.

V. Results

In this section, the accuracy of the recovered AS with respect to the reference AS is discussed. A criterion is used to select the AS for the subsequent creation of surrogate models. Then, the results for the predictive performance of the surrogate models constructed with the selected AS are presented and explained, followed by a discussion on the trade-offs between cost and accuracy for the proposed multi-fidelity approach.

It should be noted that for both the airfoil and wing test cases, the output of interest is the drag coefficient C_D obtained for a fixed lift coefficient of $C_L = 0.5$. Also, for a given training data size, at least 25 distinct models are

trained using different training samples. Therefore, unless specified otherwise, the results shown are a distribution of the performance metrics of both the G-AS and MOAS methods and for both the airfoil and wing test cases.

A. Active Subspace Recovery

1. Problem 1: RAE 2822

Variations of SPSA and LPSA for different active subspace dimensions k and sizes of the training set N obtained with the G-AS method are shown in Fig. 7 for the airfoil test case. Adding more sample points to the training data results in the computed AS getting better aligned with the reference AS. This is reflected by the reduction in both SPSA and LPSA. Increasing the dimensions of the computed AS causes a reduction in SPSA, but this is also accompanied by an increase in LPSA. This suggests that with larger subspace dimensions, it becomes easier to find two directions in the computed and reference subspace that are aligned, and similarly, it is also easier to find vectors that are not. However, a large LPSA does not necessarily imply a bad model as the directions corresponding to large principal angles could correspond to directions having smaller importance on the prediction of the surrogate output.

For the MOAS method, the variation of SPSA and LPSA with k and N is shown in Fig. 8. The trends shown are similar to one seen for G-AS in Fig. 7, although the magnitude of the subspace angles is generally higher. In particular, the subspace angles are larger for relatively smaller values of N and this trend is consistent across all three fidelity levels. This points to a limitation of the MOAS method to recover the true AS when the training sample size is small. Note that the magnitude of the subspace angles for both the G-AS and MOAS solutions is comparable across all three fidelity levels. This supports our initial hypothesis that an AS obtained from low-fidelity data can be a good approximation to a high-fidelity AS.

2. Problem 2: CRM Wing

For the CRM wing test case, the AS recovered using the G-AS method has smaller values of SPSA showing good alignment with the reference AS. The active subspace recovered using the L2 fidelity level has subspace angles comparable to the AS corresponding to the L3 level. However, the MOAS method performs poorly and fails to recover the reference AS for a given range of k and N. The difficulty of the MOAS method to recover the reference AS for high-dimensional input space has also been observed in a previous study carried out by Raphael et. al. [17]. However, this does not necessarily result in poor prediction performance as will be later demonstrated in this study.

B. Prediction Capability of Resulting Surrogates

The following section assesses the prediction capability of surrogates using either the G-AS or MOAS methods to reduce the dimensionality of the input space. The method used to select an adequate AS is first described. Then, the prediction error or surrogates trained with the proposed multi-fidelity approach is presented.

1. Selection of AS Dimension

The dimensionality of the AS to train surrogate models was selected using Relative Information Criterion (RIC) [59] for the G-AS method. This criterion represents the ratio of the variance of the gradient data captured by the G-AS method, i.e., the variance of the gradient captured by the subspace, compared to the overall variance. For this study, a threshold RIC value of 95% is selected.

As for MOAS, the Bayesian Information Criterion (BIC) is used to determine the dimension of the AS [8]. The BIC is the difference between the maximum log-likelihood of a model and a penalty term that represents the model complexity. It is defined as

$$\mathrm{BIC} = d\ \log(N) - 2\mathcal{L} \tag{23}$$

where d is the number of parameters being trained, N is the number of training samples, and $\mathcal{L}$ is the model log-likelihood. Given a set amount of data, the BIC allows the user to balance the complexity of the chosen model while avoiding possibly overfitting the data. Note that for MOAS, the value of d is given by Eq. (18) and the log-likelihood is readily available as it is computed during the training process (see Eq. (16)). A decrease in the BIC with larger k would indicate that the AS is better capturing the response of interest, while an increasing BIC would suggest that the model is starting to overfit. Since the k value corresponding to the minimum BIC or the RIC threshold depends on N, the problem being considered, and the type of AS model, different combinations of k and N were selected for the creation of surrogate models.

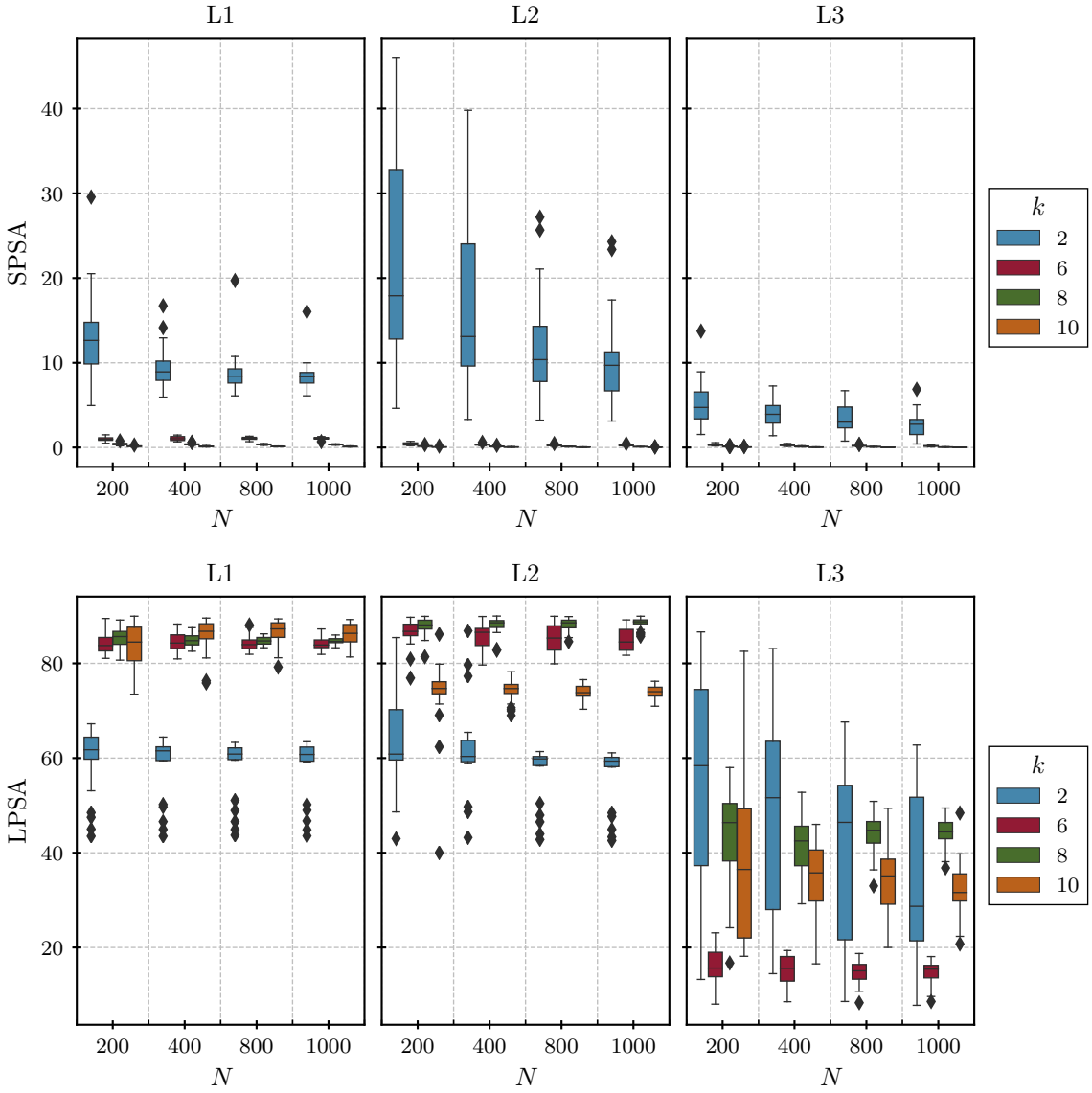


Fig. 7 RAE 2822: Distribution of smallest and largest principal subspace angles between computed and reference AS obtained for varying fidelity levels using G-AS

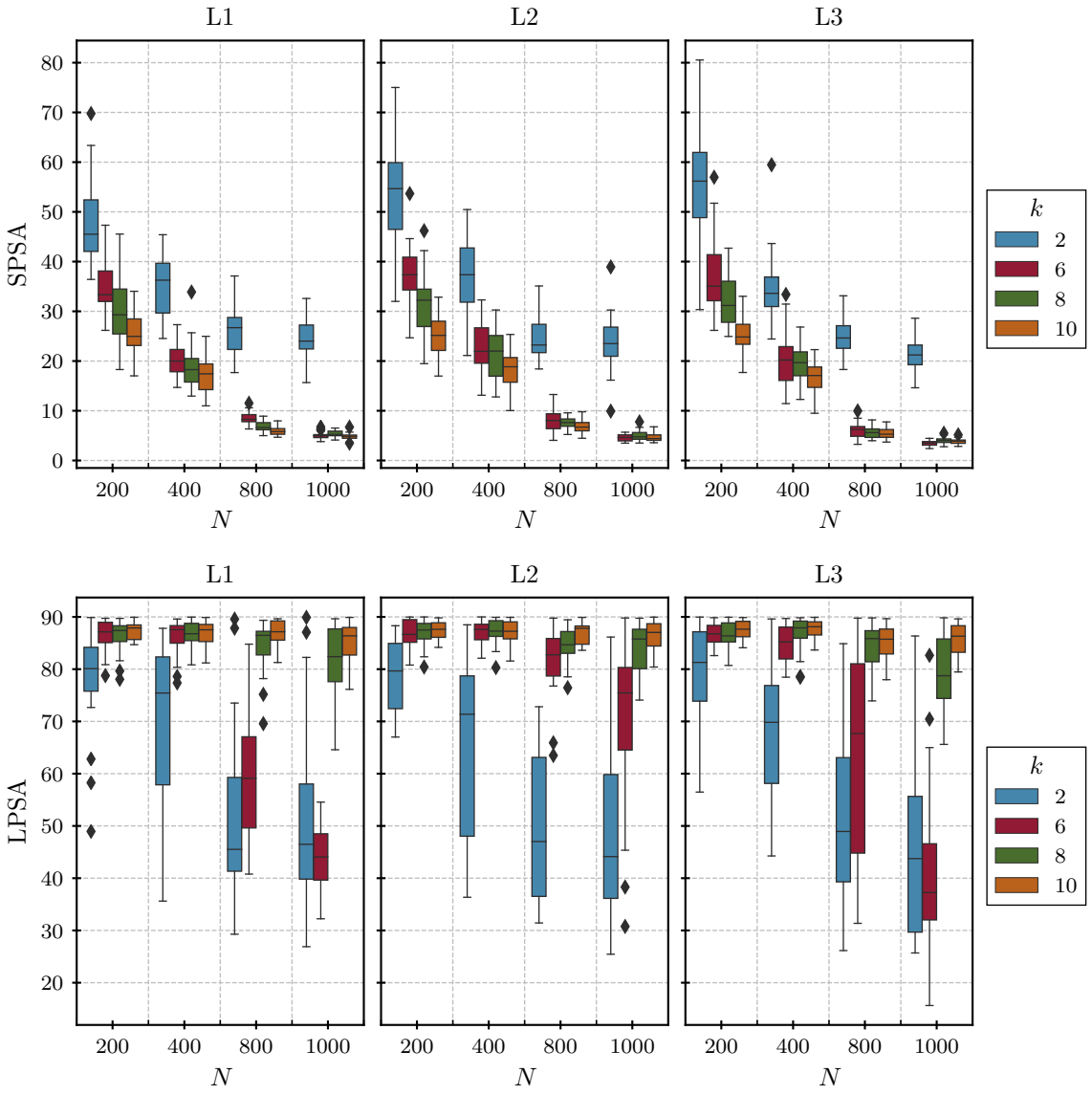


Fig. 8 RAE 2822: Distribution of smallest and largest principal subspace angles between computed and reference AS obtained for varying fidelity levels using MOAS

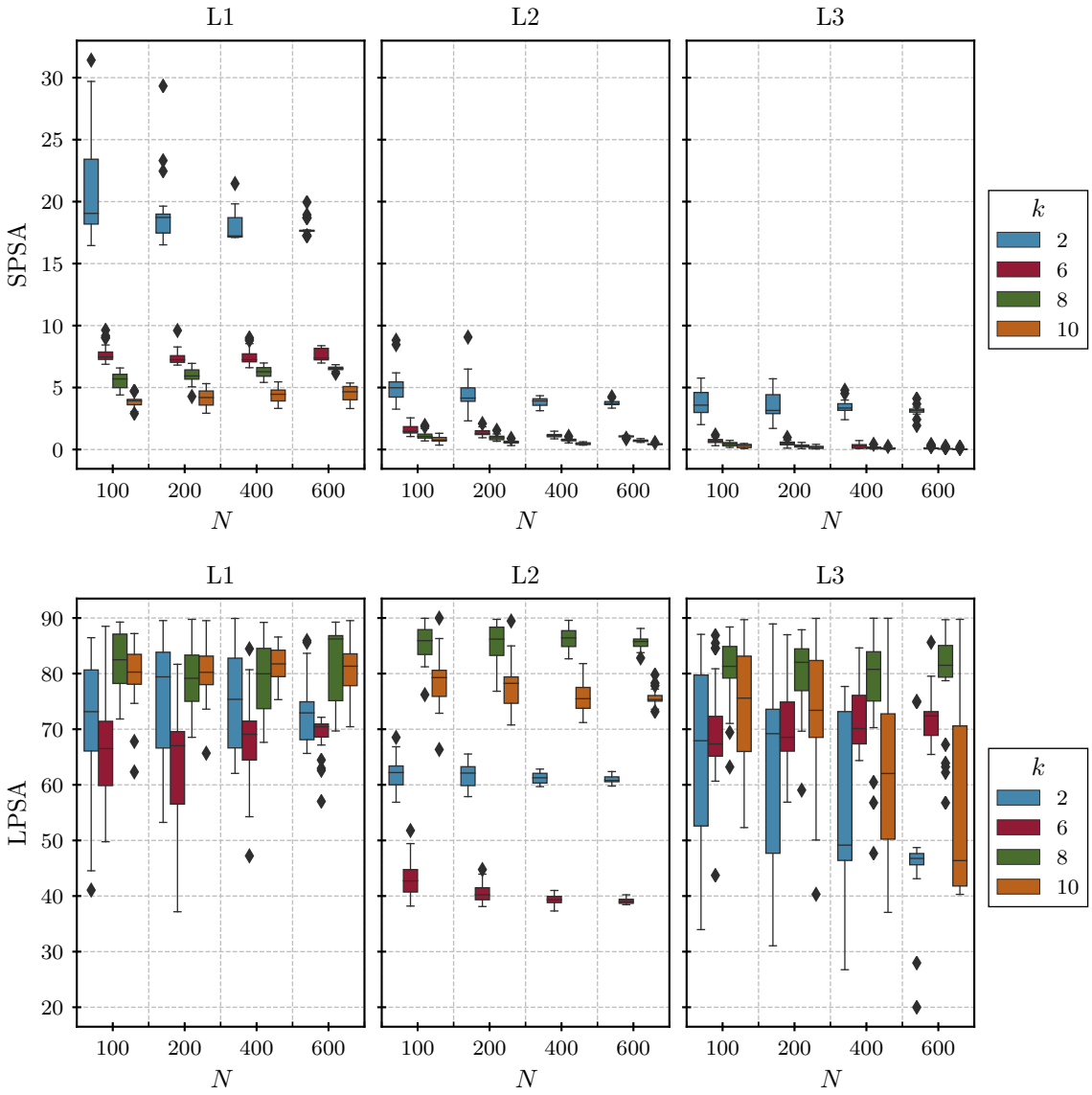


Fig. 9 CRM Wing: Distribution of smallest and largest principal subspace angles between computed and reference AS obtained for varying fidelity levels using G-AS

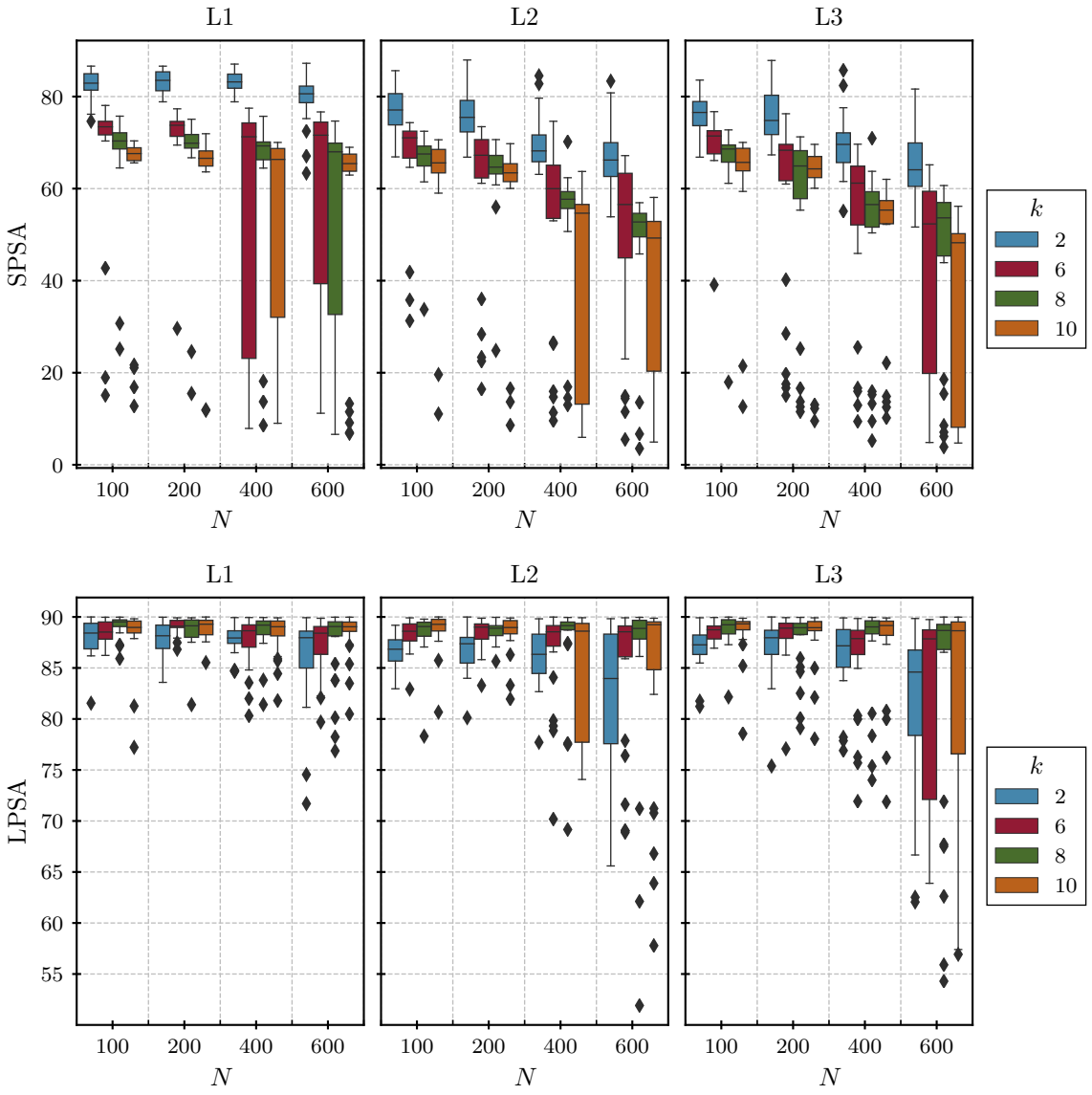


Fig. 10 CRM Wing: Distribution of smallest and largest principal subspace angles between computed and reference AS obtained for varying fidelity levels using MOAS

Table 3 Cost vs. accuracy trade-off of the proposed multi-fidelity approach

Scenario		MOAS				G-AS			
		k/N	n_t	MAPE	Cost Reduction	k/N	n_t	MAPE	Cost Reduction
RAE2822	L1	6/1000	800	8.18%	99.90%	8/400	800	6.10%	94.07%
	L2	6/1000	800	7.67%	88.48%	8/400	800	5.16%	84.39%
CRM Wing	L1	6/100	400	2.30%	97.85%	8/200	400	1.60%	90.15%
	L2	6/100	400	2.01%	83.17%	8/200	400	1.03%	71.50%

2. Prediction Error

The AS created using different fidelity levels and different combinations of k and N, were then applied to samples from the high-fidelity simulation L3. Link functions $g(\mathbf{z})$ between reduced dimensional input variables $\mathbf{z}$ and high-fidelity outputs (i.e., C_D) were created using projection matrices $\mathbf{W}$ from each fidelity level. These surrogate models can be represented as $g(\mathbf{W}_*^T\mathbf{x})$, where $*$ represents the fidelity level used to compute the AS. For instance, $g(\mathbf{W}_{\text{L2}}^T\mathbf{x})$ represents a surrogate model of the high-fidelity data, but constructed using an basis $\mathbf{W}_{\text{L2}}$ recovered using L2 data.

Different numbers of high-fidelity training samples n_t are considered. As a reminder, n_t refers to the number of training samples used to train the link functions $g(\mathbf{z})$, while N is the size of data used to compute the AS. When both the AS and $g(\mathbf{z})$ are trained with a single level of fidelity, n_t and N would typically have the same value, but that is not necessarily the case with the current multi-fidelity approach. In all cases, the surrogate models were GPR with a squared-exponential kernel. Also, for a particular combination of n_t and k/N, 25 trial runs were carried out using a different subset of samples from the overall data. This allows us to consider a distribution of the possible prediction error and assess the performance of each model independently from the choice of training samples.

Fig. 11 shows a variation of $E(C_D)$ with training sample size for the AS recovered using both G-AS and MOAS methods. For a small number of training samples, the prediction error is high irrespective of the fidelity used for the AS, and in general, a smaller AS dimension appears to fair better. As we increase the size of the training set, the prediction error reduces. Also, the effect of a different combination of k/N diminishes once surrogates are trained using a larger number of sample points. This suggests that for a sufficiently large n_t, any AS that satisfies either the RIC or BIC criteria would produce similar prediction errors. Prediction errors for surrogate models using the AS computed with G-AS tends to be lower than with MOAS. Note that the magnitude of $E(C_D)$ for the three fidelity levels (i.e., for $g(\mathbf{W}_{\text{L1}}^T\mathbf{x})$, $g(\mathbf{W}_{\text{L2}}^T\mathbf{x})$, $g(\mathbf{W}_{\text{L3}}^T\mathbf{x})$) are comparable for both the G-AS and MOAS methods. This shows that for the airfoil test case, an AS found using low-fidelity data can be a good substitute for a high-fidelity AS.

Similarly, the variation of prediction error for surrogate models of the CRM wing drag is shown in Fig.12. Overall, the error trends for $E(C_D)$ follow a pattern similar to the airfoil case However, the magnitude of the prediction errors for the CRM wing is globally smaller than for the RAE 2822.

C. Discussion

To verify if low-fidelity data can be used to generate an AS that can be a reasonable approximation of high-fidelity AS, we look at the variation of mean absolute percentage error between predictions of the high-fidelity AS-based surrogate and a low-fidelity AS-based surrogate. In other words, the MAPE values between $g(\mathbf{W}_{\text{L3}}^T\mathbf{x})$ and either $g(\mathbf{W}_{\text{L2}}^T\mathbf{x})$ or $g(\mathbf{W}_{\text{L1}}^T\mathbf{x})$ are compared. It is important to remember that for all three cases, the surrogate models are created using the high-fidelity data only. The multi-fidelity aspect is due to the data used to compute the basis $\mathbf{W}$. Fig.13 shows the variation of MAPE with n_t. For G-AS, using a larger number of samples results in the MAPE for the L1-based surrogate to settle around 6%, whereas the L2-based surrogate shows an error of 4% for the same number of training samples. For the MOAS-based model, the MAPE errors are slightly higher than for G-AS-based models. This trend for MAPE is consistent with the AS recovery metric, where the G-AS method was able to recover the true AS better than the MOAS method. Again, we see that every AS that fulfills the RIC or BIC selection criteria results in similar MAPE values. For the CRM wing, Fig.14 shows changes in MAPE values with n_t for both models using both the G-AS and MOAS methods. We achieve a MAPE of 1% for L2-based AS computed using G-AS. The MOAS-based AS performs equally well for the CRM wing case and produces a low MAPE both for L1- and L2-based surrogates.

Table 3 gives a trade-off between the relative reduction in computational cost to recover an AS using low-fidelity

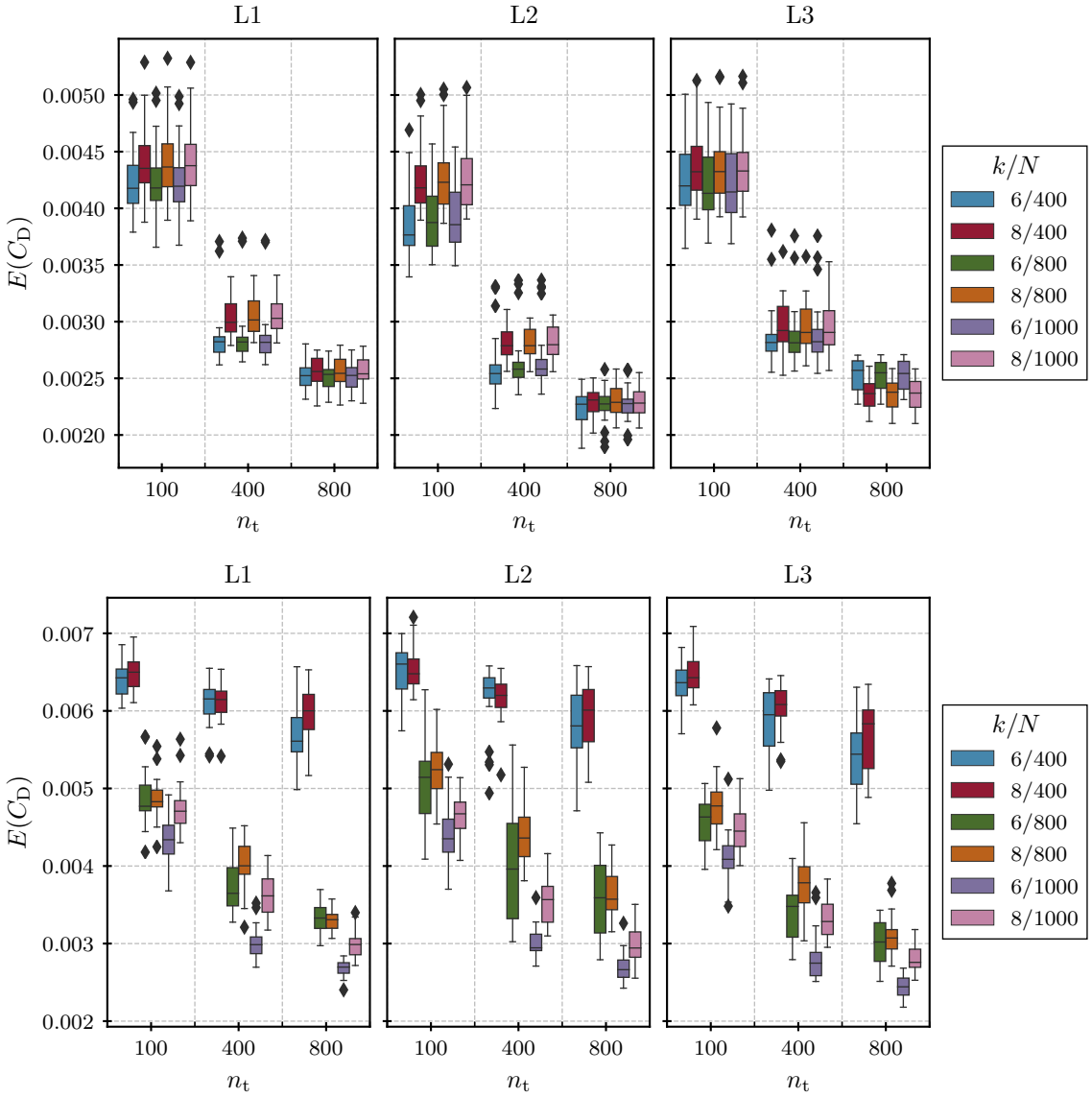


Fig. 11 RAE2822: Variation of prediction error with increasing number of training samples, and AS obtained by different combinations of N and k. G-AS (top) and MOAS (bottom)

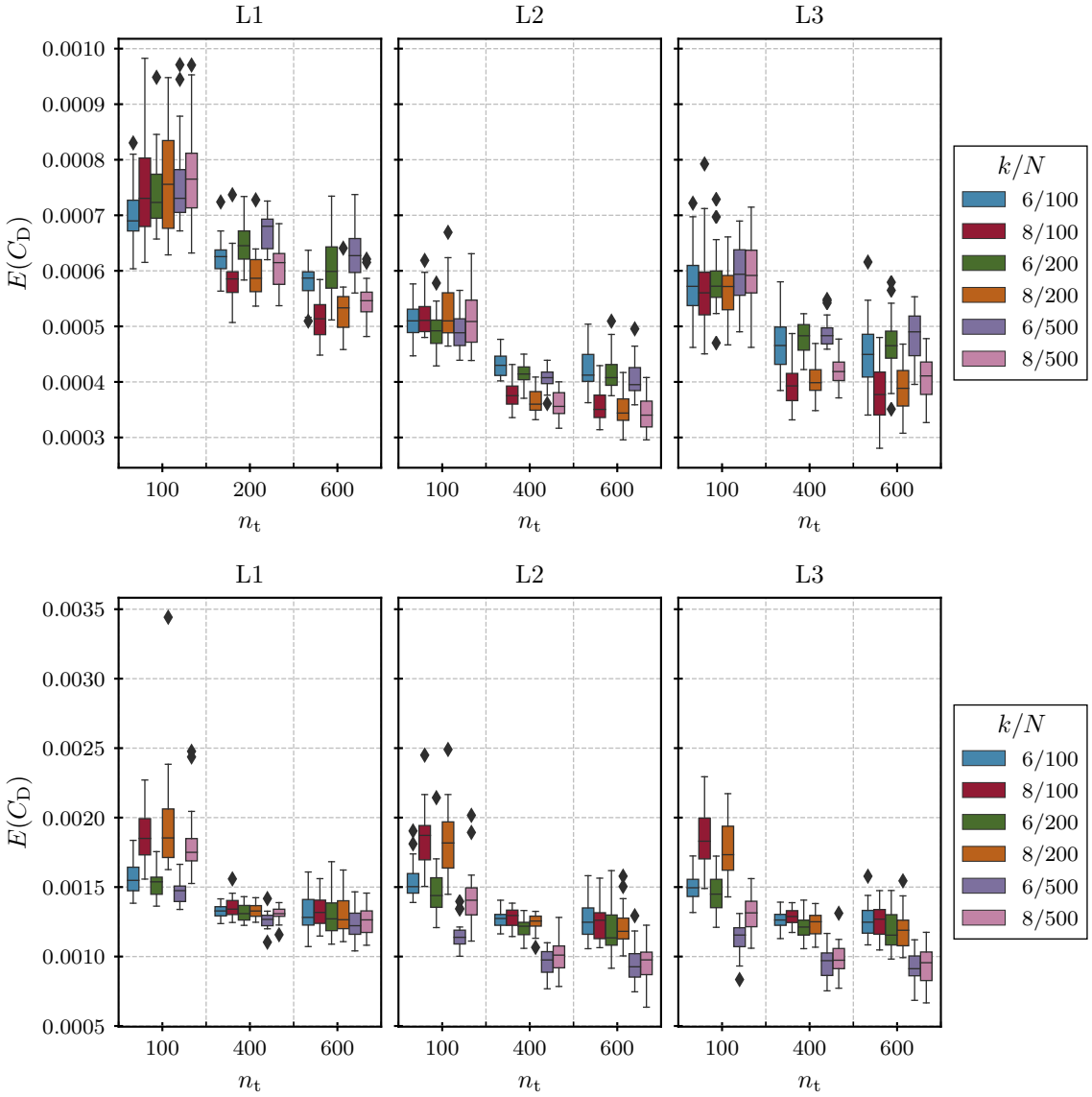


Fig. 12 CRM Wing: Variation of prediction error with increasing number of training samples, and AS obtained by different combinations of N and k. G-AS (top) and MOAS (bottom)

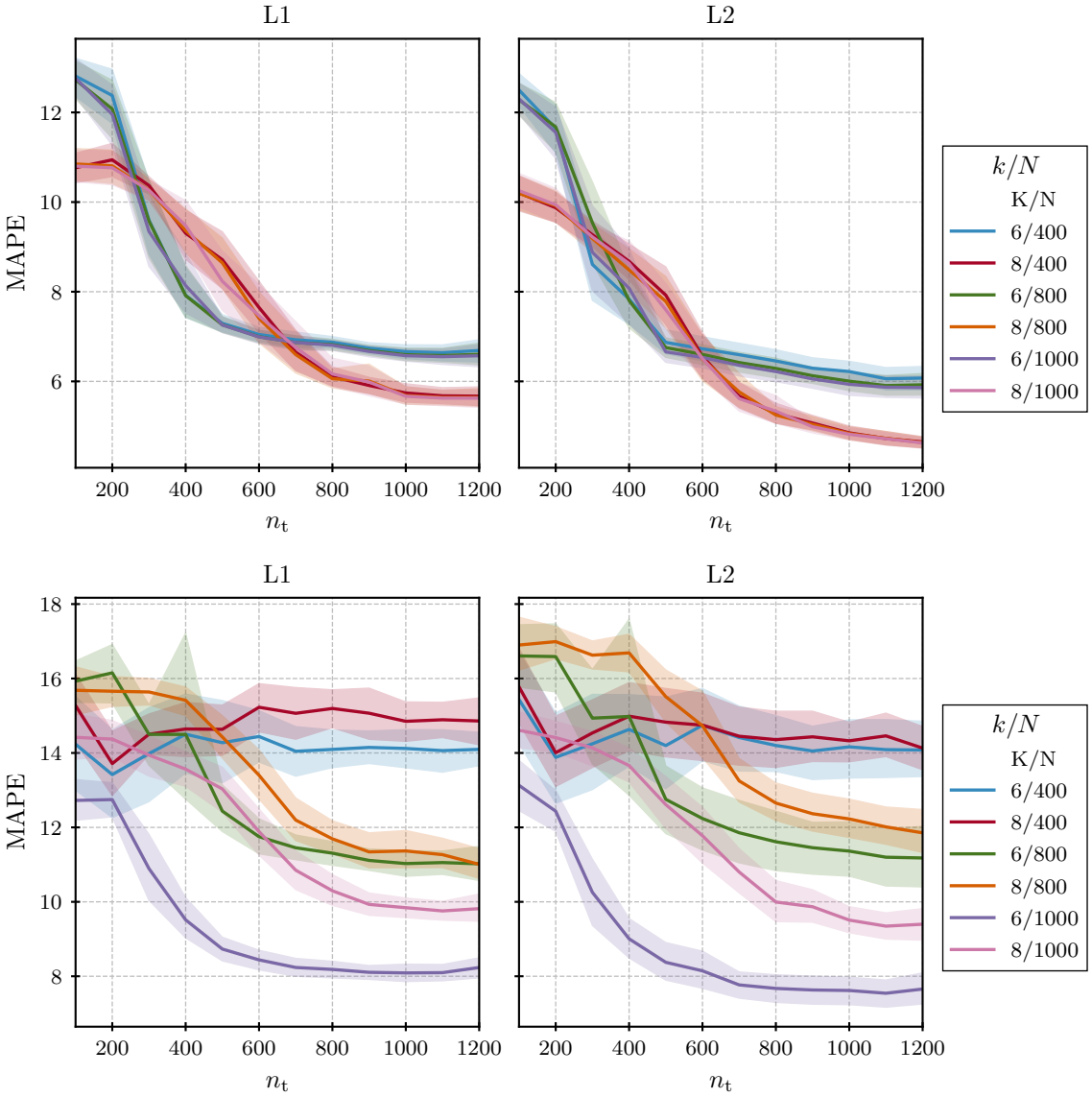


Fig. 13 RAE2822: Variation of MAPE with increasing number of training samples, and AS obtained by different combinations of N and k. G-AS (top) and MOAS (bottom)

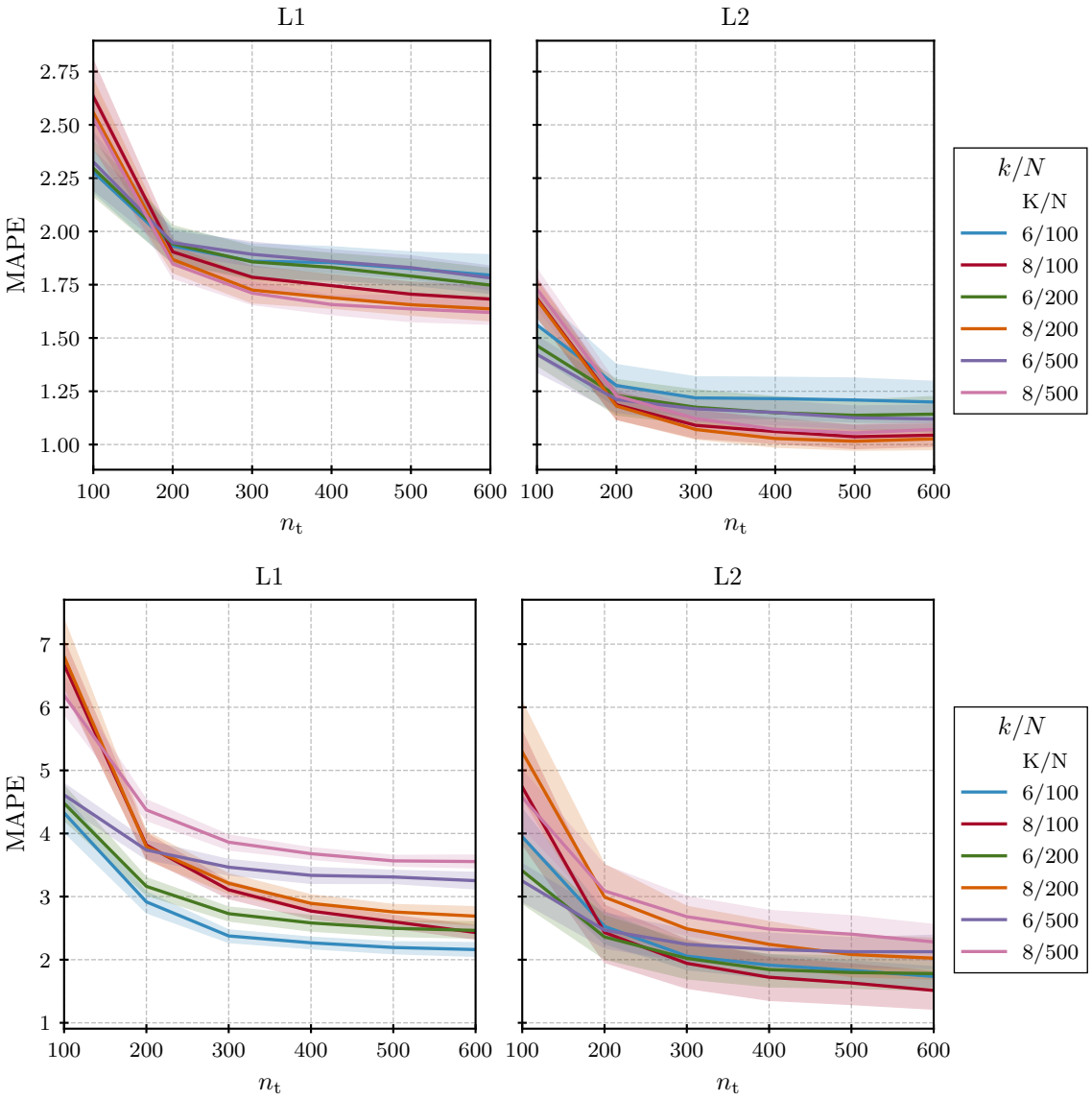


Fig. 14 CRM Wing: Variation of MAPE with increasing number of training samples, and AS obtained by different combinations of N and k. G-AS (top) and MOAS (bottom)

data and the accuracy of such an approach. The computational cost and time saving for creating an AS using low-fidelity data instead of high-fidelity data is given as a percent reduction in computational time, whereas the error associated with using this AS to construct a surrogate is given by MAPE. For the G-AS method, the computational cost includes the cost to run both the flow and adjoint solver. The k/N combination for the selected AS, the n_t to train the surrogates, and associated MAPE are also shown in table 3.

For the airfoil case, using the L2 fidelity level to create an AS results in computational savings of 88.48% and 84.39% for MOAS and G-AS-based AS respectively. The MAPE value associated with these methods is 7.67% and 5.16% respectively, which offers a very encouraging trade-off. In other words, using the L2 fidelity level to compute the AS results is a slightly worse model prediction, yet this is compensated by significant computational savings. For the CRM wing, this error reduces to 1-2% with a similar reduction in computational time. Using L1 fidelity level to create AS results in computational savings of over 90% for all cases. However, L1-based surrogates have higher MAPE compared to L2-based surrogates. A trade-off can also be established between computational cost and prediction error for using the G-AS or the MOAS method. On one hand, the gradient-based AS tends to offer smaller prediction errors. On the other hand, finding AS using this method is expensive, as both direct and adjoint solutions need to be obtained. There is also some computational cost associated with finding the AS through the MOAS method. That being said, this cost is low as compared to the cost of running the adjoint solver and is not affected by the fidelity of the data used. Hence this cost can be considered constant while carrying out the cost vs accuracy trade-off analysis.

VI. Conclusion

A multi-fidelity approach for the creation of an AS is proposed in this paper where the AS is computed using low-fidelity data only. The obtained AS is then used to sample high-fidelity sample space for the subsequent construction of a surrogate model. Gradient-based and gradient-free methods were both considered for the proposed approach when computing the AS. This multi-fidelity approach is applied to an airfoil and a wing test case, and significant savings, in terms of computational cost, are observed. The lower cost of the proposed approach is counterbalanced by a slight increase in error, but this trade-off would likely be worthwhile in many application problems. Overall, this study demonstrates that computing the AS of a high-fidelity analysis using low-fidelity data can be a viable approximation. The proposed approach is simpler and easier to implement when compared to other multi-fidelity approaches that instead combine both low- and high-fidelity for AS computation. Future improvements in this work include testing this approach with other gradient-free AS methods to improve prediction accuracy. Also, using a mixture of high- and low-fidelity data for the training of the link function in addition to the AS will also be considered.

VII. Acknowledgements and Disclaimer

This research was funded by the U.S. Federal Aviation Administration Office of Environment and Energy through ASCENT, the FAA Center of Excellence for Alternative Jet Fuels and the Environment, project 050 through FAA Award Number 13-C-AJFE-GIT-057 under the supervision of Chris Dorbian. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the FAA.

References

[1] Martins, J. R. R. A., and Hwang, J. T., "Review and Unification of Methods for Computing Derivatives of Multidisciplinary Computational Models," *AIAA Journal*, Vol. 51, No. 11, 2013, pp. 2582–2599. https://doi.org/10.2514/1.J052184, URL http://arc.aiaa.org/doi/abs/10.2514/1.J052184.

[2] Forrester, A. I., and Keane, A. J., "Recent advances in surrogate-based optimization," *Progress in Aerospace Sciences*, Vol. 45, No. 1-3, 2009, pp. 50–79. https://doi.org/10.1016/j.paerosci.2008.11.001.

[3] Queipo, N. V., Haftka, R. T., Shyy, W., Goel, T., Vaidyanathan, R., and Kevin Tucker, P., "Surrogate-based analysis and optimization," *Progress in Aerospace Sciences*, Vol. 41, No. 1, 2005, pp. 1–28. https://doi.org/10.1016/j.paerosci.2005.02.001.

[4] Yondo, R., Andrés, E., and Valero, E., "A review on design of experiments and surrogate models in aircraft real-time and many-query aerodynamic analyses," *Progress in Aerospace Sciences*, Vol. 96, No. November 2017, 2018, pp. 23–61. https://doi.org/10.1016/j.paerosci.2017.11.003, URL https://linkinghub.elsevier.com/retrieve/pii/S0376042117300611.

[5] Kenway, G. K. W., Kennedy, G. J., and Martins, J. R. R. A., "Scalable Parallel Approach for High-Fidelity Steady-

State Aeroelastic Analysis and Adjoint Derivative Computations," *AIAA Journal*, Vol. 52, No. 5, 2014, pp. 935–951. https://doi.org/10.2514/1.J052255, URL https://arc.aiaa.org/doi/10.2514/1.J052255.

[6] Lyu, Z., Kenway, G. K. W., and Martins, J. R. R. A., "Aerodynamic Shape Optimization Investigations of the Common Research Model Wing Benchmark," *AIAA Journal*, Vol. 53, No. 4, 2015, pp. 968–985. https://doi.org/10.2514/1.J053318, URL https://arc.aiaa.org/doi/10.2514/1.J053318.

[7] Chen, S., Lyu, Z., Kenway, G. K. W., and Martins, J. R. R. A., "Aerodynamic Shape Optimization of Common Research Model Wing–Body–Tail Configuration," *Journal of Aircraft*, Vol. 53, No. 1, 2016, pp. 276–293. https://doi.org/10.2514/1.C033328, URL https://arc.aiaa.org/doi/10.2514/1.C033328.

[8] Tripathy, R., Bilionis, I., and Gonzalez, M., "Gaussian processes with built-in dimensionality reduction: Applications to high-dimensional uncertainty propagation," *Journal of Computational Physics*, Vol. 321, 2016, pp. 191–223. https://doi.org/10.1016/j.jcp.2016.05.039.

[9] Koch, P. N., Simpson, T. W., Allen, J. K., and Mistree, F., "Statistical Approximations for Multidisciplinary Design Optimization: The Problem of Size," *Journal of Aircraft*, Vol. 36, No. 1, 1999, pp. 275–286. https://doi.org/10.2514/2.2435, URL https://arc.aiaa.org/doi/10.2514/2.2435.

[10] Shan, S., and Wang, G. G., "Survey of modeling and optimization strategies to solve high-dimensional design problems with computationally-expensive black-box functions," *Structural and Multidisciplinary Optimization*, Vol. 41, No. 2, 2010, pp. 219–241. https://doi.org/10.1007/s00158-009-0420-2, URL http://link.springer.com/10.1007/s00158-009-0420-2.

[11] Van Der Maaten, L., Postma, E., Van den Herik, J., et al., "Dimensionality reduction: a comparative," *J Mach Learn Res*, Vol. 10, No. 66-71, 2009, p. 13.

[12] Lumley, J. L., "The Structure of Inhomogeneous Turbulent Flows," *Atmospheric turbulence and radio propagation*, edited by A. M. Yaglom and V. I. Tatarski, Nauka, Moscow, 1967, pp. 166–178.

[13] Lucia, D. J., Beran, P. S., and Silva, W. A., "Reduced-order modeling: new approaches for computational physics," *Progress in Aerospace Sciences*, Vol. 40, No. 1-2, 2004, pp. 51–117. https://doi.org/10.1016/j.paerosci.2003.12.001, URL https://linkinghub.elsevier.com/retrieve/pii/S0376042103001131.

[14] Yu, J., Yan, C., and Guo, M., "Non-intrusive reduced-order modeling for fluid problems: A brief review," *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, Vol. 233, No. 16, 2019, pp. 5896–5912. https://doi.org/10.1177/0954410019890721.

[15] Constantine, P. G., Dow, E., and Wang, Q., "Active Subspace Methods in Theory and Practice: Applications to Kriging Surfaces," *SIAM Journal on Scientific Computing*, Vol. 36, No. 4, 2014, pp. A1500–A1524. https://doi.org/10.1137/130916138, URL http://arxiv.org/abs/1304.2070http://dx.doi.org/10.1137/130916138http://epubs.siam.org/doi/10.1137/130916138.

[16] Seshadri, P., Yuchi, S., and Parks, G. T., "Dimension reduction via Gaussian ridge functions," *SIAM-ASA Journal on Uncertainty Quantification*, Vol. 7, No. 4, 2019, pp. 1301–1322. https://doi.org/10.1137/18M1168571.

[17] Gautier, R., Pandita, P., Ghosh, S., and Mavris, D., "A Fully Bayesian Gradient-Free Supervised Dimension Reduction Method using Gaussian Processes," *arXiv preprint arXiv:2008.03534*, 2020. URL http://arxiv.org/abs/2008.03534.

[18] Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S., *Global sensitivity analysis: the primer*, John Wiley & Sons, 2008.

[19] Neal, R., "Assessing relevance determination methods using DELVE," *Neural Networks and Machine Learning*, 1998, pp. 97–129. Cited By 62.

[20] Jolliffe, I., "Principal component analysis," *Encyclopedia of statistics in behavioral science*, 2005.

[21] Dietrich, F., Künzner, F., Neckel, T., Köster, G., and Bungartz, H.-J., "Fast and flexible uncertainty quantification through a data-driven surrogate model," *International Journal for Uncertainty Quantification*, Vol. 8, No. 2, 2018.

[22] Geladi, P., and Kowalski, B. R., "Partial least-squares regression: a tutorial," *Analytica Chimica Acta*, Vol. 185, 1986. https://doi.org/10.1016/0003-2670(86)80028-9.

[23] Wold, S., Sjöström, M., and Eriksson, L., "PLS-regression: a basic tool of chemometrics," *Chemometrics and Intelligent Laboratory Systems*, Vol. 58, 2001. https://doi.org/10.1016/S0169-7439(01)00155-1.

[24] Bouhlel, M. A., Bartoli, N., Otsmane, A., and Morlier, J., "An Improved Approach for Estimating the Hyperparameters of the Kriging Model for High-Dimensional Problems through the Partial Least Squares Method," *Mathematical Problems in Engineering*, Vol. 2016, 2016. https://doi.org/10.1155/2016/6723410.

[25] Bouhlel, M. A., Bartoli, N., Regis, R. G., Otsmane, A., and Morlier, J., "Efficient global optimization for high-dimensional constrained problems by using the Kriging models combined with the partial least squares method," *Engineering Optimization*, Vol. 50, 2018. https://doi.org/10.1080/0305215X.2017.1419344.

[26] Lukaczyk, T. W., Constantine, P., Palacios, F., and Alonso, J. J., "Active subspaces for shape optimization," *10th AIAA multidisciplinary design optimization conference*, 2014, p. 1171.

[27] Tezzele, M., Salmoiraghi, F., Mola, A., and Rozza, G., "Dimension reduction in heterogeneous parametric spaces with application to naval engineering shape design problems," *Advanced Modeling and Simulation in Engineering Sciences*, Vol. 5, No. 1, 2018, pp. 1–19.

[28] Constantine, P. G., Emory, M., Larsson, J., and Iaccarino, G., "Exploiting active subspaces to quantify uncertainty in the numerical simulation of the HyShot II scramjet," *Journal of Computational Physics*, Vol. 302, 2015, pp. 1–20. https://doi.org/https://doi.org/10.1016/j.jcp.2015.09.001, URL https://www.sciencedirect.com/science/article/pii/S002199911500580X.

[29] Lewis, A., Smith, R., and Williams, B., "Gradient free active subspace construction using Morris screening elementary effects," *Computers and Mathematics with Applications*, Vol. 72, No. 6, 2016, pp. 1603–1615. https://doi.org/https://doi.org/10.1016/j.camwa.2016.07.022, URL https://www.sciencedirect.com/science/article/pii/S0898122116304242.

[30] Rajaram, D., Gautier, R. H., Perron, C., Pinon-Fischer, O. J., and Mavris, D., "Non-intrusive parametric reduced order models with high-dimensional inputs via gradient-free active subspace," *AIAA AVIATION 2020 FORUM*, 2020, p. 3184.

[31] Tsilifis, P., Pandita, P., Ghosh, S., Andreoli, V., Vandeputte, T., and Wang, L., "Bayesian learning of orthogonal embeddings for multi-fidelity Gaussian Processes," *Computer Methods in Applied Mechanics and Engineering*, Vol. 386, 2021. https://doi.org/10.1016/j.cma.2021.114147.

[32] Forrester, A. I., Sóbester, A., and Keane, A. J., "Multi-fidelity optimization via surrogate modelling," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, Vol. 463, 2007. https://doi.org/10.1098/rspa.2007.1900.

[33] Kennedy, M., and O'Hagan, A., "Predicting the output from a complex computer code when fast approximations are available," *Biometrika*, Vol. 87, No. 1, 2000, pp. 1–13. https://doi.org/10.1093/biomet/87.1.1, URL papers2://publication/uuid/1A69EB40-F7E9-48E1-B609-FF3A7497DBC2http://biomet.oupjournals.org/cgi/doi/10.1093/biomet/87.1.1.

[34] Han, Z. H., and Görtz, S., "Hierarchical kriging model for variable-fidelity surrogate modeling," *AIAA Journal*, Vol. 50, 2012, pp. 1885–1896. https://doi.org/10.2514/1.J051354.

[35] Pang, G., Perdikaris, P., Cai, W., and Karniadakis, G. E., "Discovering variable fractional orders of advection–dispersion equations from field data using multi-fidelity Bayesian optimization," *Journal of Computational Physics*, Vol. 348, 2017. https://doi.org/10.1016/j.jcp.2017.07.052.

[36] Perdikaris, P., Venturi, D., and Karniadakis, G. E., "Multifidelity Information Fusion Algorithms for High-Dimensional Systems and Massive Data sets," *SIAM Journal on Scientific Computing*, Vol. 38, 2016. https://doi.org/10.1137/15M1055164.

[37] Qian, E., Peherstorfer, B., O'Malley, D., Vesselinov, V. V., and Willcox, K., "Multifidelity Monte Carlo estimation of variance and sensitivity indices," *SIAM/ASA Journal on Uncertainty Quantification*, Vol. 6, No. 2, 2018, pp. 683–706.

[38] Perdikaris, P., Venturi, D., Royset, J. O., and Karniadakis, G. E., "Multi-fidelity modelling via recursive co-kriging and Gaussian–Markov random fields," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, Vol. 471, 2015. https://doi.org/10.1098/rspa.2015.0018.

[39] Lam, R. R., Zahm, O., Marzouk, Y. M., and Willcox, K. E., "Multifidelity dimension reduction via active subspaces," *SIAM Journal on Scientific Computing*, Vol. 42, 2020, pp. A929–A956. https://doi.org/10.1137/18M1214123.

[40] Cressie, N., "The origins of kriging," *Mathematical geology*, Vol. 22, No. 3, 1990, pp. 239–252.

[41] Rasmussen, C. E., *Gaussian Processes in Machine Learning*, Springer Berlin Heidelberg, 2004. https://doi.org/10.1007/978-3-540-28650-9_4.

[42] Townsend, J., Koep, N., and Weichwald, S., "Pymanopt: A python toolbox for optimization on manifolds using automatic differentiation," *Journal of Machine Learning Research*, Vol. 17, 2016, pp. 1–5.

[43] Bradbury, J., Frostig, R., Hawkins, P., Johnson, M. J., Leary, C., Maclaurin, D., Necula, G., Paszke, A., VanderPlas, J., Wanderman-Milne, S., and Zhang, Q., "JAX: composable transformations of Python+NumPy programs," , 2018. URL http://github.com/google/jax.

[44] Tezzele, M., Salmoiraghi, F., Mola, A., and Rozza, G., "Dimension reduction in heterogeneous parametric spaces with application to naval engineering shape design problems," *Advanced Modeling and Simulation in Engineering Sciences*, Vol. 5, No. 1, 2018, p. 25. https://doi.org/10.1186/s40323-018-0118-3, URL https://amses-journal.springeropen.com/articles/10.1186/s40323-018-0118-3.

[45] Anderson, G. R., Nemec, M., and Aftosmis, M. J., "Aerodynamic Shape Optimization Benchmarks with Error Control and Automatic Parameterization," *53rd AIAA Aerospace Sciences Meeting*, American Institute of Aeronautics and Astronautics, Reston, Virginia, 2015, pp. 1–18. https://doi.org/10.2514/6.2015-1719, URL http://arc.aiaa.org/doi/10.2514/6.2015-1719.

[46] Lee, C., Koo, D., Telidetzki, K., Buckley, H., Gagnon, H., and Zingg, D. W., "Aerodynamic shape optimization of benchmark problems using jetstream," *53rd AIAA Aerospace Sciences Meeting*, 2015, p. 0262.

[47] Poole, D. J., Allen, C. B., and Rendall, T., "Control point-based aerodynamic shape optimization applied to AIAA ADODG test cases," *53rd AIAA Aerospace Sciences Meeting*, 2015, p. 1947.

[48] Economon, T. D., Palacios, F., Copeland, S. R., Lukaczyk, T. W., and Alonso, J. J., "SU2: An Open-Source Suite for Multiphysics Simulation and Design," *AIAA Journal*, Vol. 54, No. 3, 2015, pp. 1–19. https://doi.org/10.2514/1.J053813, URL http://dx.doi.org/10.2514/1.J053813.

[49] Spalart, P., and Allmaras, S., "A one-equation turbulence model for aerodynamic flows," *30th aerospace sciences meeting and exhibit*, 1992, p. 439.

[50] Jameson, A., Schmidt, W., and Turkel, E., "Numerical solution of the Euler equations by finite volume methods using Runge Kutta time stepping schemes," *14th fluid and plasma dynamics conference*, 1981, p. 1259.

[51] Vassberg, J., Dehaan, M., Rivers, M., and Wahls, R., "Development of a common research model for applied CFD validation studies," *26th AIAA Applied Aerodynamics Conference*, 2008, p. 6919.

[52] Lyu, Z., Kenway, G. K. W., and Martins, J. R. R. A., "CRM: Common Research Model," , 2021. Data retrieved from Mendeley Data, V1, https://doi.org/10.17632/7jnyjdbvf9.1.

[53] Masters, D. A., Poole, D. J., Taylor, N. J., Rendall, T., and Allen, C. B., "Impact of shape parameterisation on aerodynamic optimisation of benchmark problem," *54th AIAA Aerospace Sciences Meeting*, 2016, p. 1544.

[54] Hicks, R. M., and Henne, P. A., "Wing Design By Numerical Optimization," *Journal of Aircraft*, Vol. 15, No. 7, 1978, pp. 407–412. https://doi.org/10.2514/3.58379.

[55] Kulfan, B. M., "Universal Parametric Geometry Representation Method," *Journal of Aircraft*, Vol. 45, No. 1, 2008, pp. 142–158. https://doi.org/10.2514/1.29958.

[56] Kenway, G., Kennedy, G., and Martins, J., "A CAD-Free Approach to High-Fidelity Aerostructural Optimization," *13th AIAA/ISSMO Multidisciplinary Analysis Optimization Conference*, American Institute of Aeronautics and Astronautics, Reston, Virigina, 2010, pp. 1–18. https://doi.org/10.2514/6.2010-9231, URL http://arc.aiaa.org/doi/10.2514/6.2010-9231.

[57] Dwight, R. P., "Robust Mesh Deformation using the Linear Elasticity Equations," *Computational Fluid Dynamics 2006*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, pp. 401–406. https://doi.org/10.1007/978-3-540-92779-2_62, URL http://link.springer.com/10.1007/978-3-540-92779-2{_}62.

[58] Knyazev, A. V., and Argentati, M. E., "Principal Angles between Subspaces in an A -Based Scalar Product: Algorithms and Perturbation Estimates," *SIAM Journal on Scientific Computing*, Vol. 23, No. 6, 2002, pp. 2008–2040. https://doi.org/10.1137/S1064827500377332, URL http://epubs.siam.org/doi/10.1137/S1064827500377332.

[59] Berguin, S. H., Rancourt, D., and Mavris, D. N., "Method to Facilitate High-Dimensional Design Space Exploration Using Computationally Expensive Analyses," *AIAA Journal*, Vol. 53, No. 12, 2015, pp. 3752–3765. https://doi.org/10.2514/1.J054035, URL https://arc.aiaa.org/doi/10.2514/1.J054035.