aerospace

MDPI

Article

Multi-Fidelity Adaptive Sampling for Surrogate-Based Optimization and Uncertainty Quantification

Andrea Garbo *, Jigar Parekh, Tilo Rischmann and Philipp Bekemeyer

Institute of Aerodynamics and Flow Technology, Germany Aerospace Center (DLR), 38108 Braunschweig, Germany; t.rischmann@tu-braunschweig.de (T.R.); philipp.bekemeyer@dlr.de (P.B.)
* Correspondence: andrea.garbo@dlr.de

Abstract: Surrogate-based algorithms are indispensable in the aerospace engineering field for reducing the computational cost of optimization and uncertainty quantification analyses, particularly those involving computationally intensive solvers. This paper presents a novel approach for enhancing the efficiency of surrogate-based algorithms through a new multi-fidelity sampling technique. Unlike existing multi-fidelity methods which are based on a single multiplicative acquisition function, the proposed technique decouples the identification of the new infill sample from the selection of the fidelity level. The location of the infill sample is determined by leveraging the highest fidelity surrogate model, while the fidelity level used for its performance evaluation is chosen as the cheapest one within the "accurate enough" models at the infill location. Moreover, the methodology introduces the application of the Jensen–Shannon divergence to quantify the accuracy of the different fidelity levels. Overall, the resulting technique eliminates some of the drawbacks of existing multiplicative acquisition functions such as the risk of continuous sampling from lower and cheaper fidelity levels. Experimental validation conducted in surrogate-based optimization and uncertainty quantification scenarios demonstrates the efficacy of the proposed approach. In an aerodynamic shape optimization task focused on maximizing the lift-to-drag ratio, the multi-fidelity strategy achieved comparable results to standard single-fidelity sampling but with approximately a five-fold improvement in computational efficiency. Likewise, a similar reduction in computational costs was observed in the uncertainty quantification problem, with the resulting statistical values aligning closely with those obtained using traditional single-fidelity sampling.

Keywords: multi-fidelity sampling; surrogate-based optimization; uncertainty quantification; computational fluid dynamics

Citation: Garbo, A.; Parekh, J.; Rischmann, T.; Bekemeyer, P. Multi-Fidelity Adaptive Sampling for Surrogate-Based Optimization and Uncertainty Quantification. *Aerospace* **2024**, *11*, 448. https://doi.org/10.3390/aerospace11060448

Academic Editor: Alex Zanotti

Received: 26 April 2024
Revised: 28 May 2024
Accepted: 28 May 2024
Published: 31 May 2024

1. Introduction

Numerical simulations are one of the backbones of modern day aerospace engineering, spanning across structural and aerodynamic analyses. The level of accuracy that can be achieved, often referred to as level of fidelity, and its affordability have significantly increased over the years thanks to remarkable advances in software and hardware capabilities. For example, in the field of computational fluid dynamics (CFD), Reynolds-averaged Navier-Stokes (RANS) or even large eddy simulations are nowadays feasible also for complex configurations, whereas two decades ago, only Euler simulations were affordable for the same geometrical complexity. Hence, engineers frequently encounter situations where data from different fidelity levels are available. Traditionally, one would substitute lower-fidelity data with higher-fidelity, more accurate solutions as they are available. Alternatively, efforts may be directed towards improving lower-fidelity solutions through a correction scheme. A notable example is enhancing panel methods with CFD solutions for more accurate flutter predictions. Another option is to leverage multi-fidelity or variable-fidelity techniques to create continuous models that contain as much information as possible.

Multi-fidelity approaches describe methods that facilitate the integration of scalar quantities from different fidelity levels, such as lift coefficient values for varying angles of attack from different numerical simulations or even experimental measurements. Several methods have been proposed for this task, including bridge functions [1,2], co-Kriging [3–7], and hierarchical Kriging [8]. Even though all such models enable combining data from different levels of fidelity, their underlying formulations vary. Bridge functions necessitate the preprocessing of data to compute deltas between different fidelity levels, subsequently described by the variable-fidelity model. These differences can be accounted for through additive or multiplicative terms. Co-Kriging models take a slightly different path and can be viewed as a general extension of single-fidelity Kriging to a method that is assisted by auxiliary variables or secondary information, typically employing an autoregressive formulation [4]. Finally, hierarchical Kriging models use lower-fidelity data models as trend functions for the next higher level [8]. Lately, also neural networks have been employed to combine data from different fidelity levels [9,10]. Each modeling approach offers the capability to combine data from different levels of fidelity, with minor advantages and disadvantages when compared. Hence, there is no clear superior candidate, and the choice of a specific modeling type depends more on availability and the individual use case at hand.

Especially for tasks that require a significant amount of function evaluations, e.g., optimization or uncertainty quantification (UQ), relying on the highest available fidelity often remains unfeasible. Different approaches have been proposed and actively employed to tackle multi-query scenarios. The most simplistic one is to reduce the fidelity level that is investigated, i.e., simplify the governing equations that are solved or reduce the spatial and temporal resolutions of the simulation until the scenario becomes feasible. A more sophisticated approach uses surrogate models that treat the numerical simulation as a black box and emulate the input-to-output relation within certain bounds. Such models can then be employed during multi-query scenarios instead of the numerical simulation itself and lead to so-called surrogate-based or surrogate-assisted approaches that are the de facto state of the art for the optimization of cost-extensive black-box functions [3] as well as uncertainty propagation [11]. To ensure that models meet the accuracy requirements, it is common to iteratively refine them through adaptive sampling sometimes also labeled as infill or active learning. Various criteria are available, and the interested reader is referred to the literature for a more in-depth introduction [12–14]. Accounting for all levels of fidelity at hand during the adaptive sampling stage should result in reduced overall computational cost necessary to perform a certain investigation, as the available information is used efficiently.

Even though both aforementioned fields, variable-fidelity modeling as well surrogate-based techniques, are well established, the intersection between them has only become a research focus recently. Nevertheless, a few approaches have already been proposed. A straightforward approach is to collect data from all fidelity levels during the design-of-experiment (DoE) phase while restricting the adaptive sampling stage to only the highest fidelity level available [3,15]. In the last decade, some multi-fidelity sampling techniques have been proposed in the context of surrogate-based optimization [16–19]. The methods proposed by Huang in [16] and later extended by Di Fiore in [18] are based on a multiplicative acquisition function, where the expected improvement metric [20] is multiplied by other factors in order to take into consideration the evaluation cost and the accuracy of the different fidelity levels. Shu [17] introduces an alternative aggregate acquisition function. Here, the acquisition function is divided into two components based on whether the highest or lowest fidelity level is being considered, and the evaluation cost is taken into consideration as a multiplicative factor. Foumani in [19] proposes a similar approach, wherein the two components of the acquisition function serve two distinct objectives: exploration and exploitation. Specifically, the exploration and exploitation parts are active when either the low or the high fidelity level is under consideration, respectively. All these multi-fidelity sampling techniques identify the next infill sample and fidelity level by maximizing their acquisition function through the solution of a single-objective mixed-integer optimization

problem. In particular, all the aggregate acquisition functions (either sum or multiplication) are characterized by a strong coupling between the effectiveness of the infill point as well as the accuracy and evaluation cost of different fidelity levels. As a result, the algorithm may persistently sample at lower fidelity levels, especially in scenarios where a fidelity level is significantly less expensive than the higher-fidelity counterpart at the price of a limited reduction in prediction accuracy. The two-step multi-fidelity sampling criterion presented in this manuscript is expected to address this limitation by decoupling the identification of the next infill point from the selection of the fidelity level.

In this paper, we introduce a novel multi-fidelity infill criterion that can be combined with all single fidelity criteria, either for optimization or uncertainty quantification tasks. In particular, the selection of the fidelity level acts as a second step after the next sampling location has been computed and it relies on the Jensen–Shannon divergence. The efficacy of the proposed multi-fidelity sampling technique is successfully demonstrated on an aerodynamics application, addressing both optimization and uncertainty quantification challenges. Specifically, the application problem uses four aerodynamics solvers of increasing fidelity: a panel code, an Euler solver, and a RANS solver applied to two computational meshes of different sizes. The results show how the multi-fidelity sampling technique achieves results comparable to its the single-fidelity counterpart, albeit at a significant reduction in computational cost.

The manuscript is organized as follow: the two-step sampling methodology and its integration in optimization and uncertainty quantification schemes is described in Section 2. Section 3 presents the analysis and results of the application of the resulting algorithms on the aerodynamics problem. Finally, the conclusion section, Section 4, summarizes the key findings of the results and outlines potential future extensions of the presented algorithm.

2. Methodology

All surrogate-based techniques have the same iterative architecture in common which can be summarized in three phases: fit the surrogate models based on the available data, identify the next sample (or samples in case of batch-sampling techniques) via an infill criterion, and evaluate the resulting design by means of the selected analysis tools as shown in Figure 1. The process is repeated until a prescribed convergence criterion is reached, or the whole computational budget is consumed. Even though there might be some differences in the implementation of the iterative algorithm, the choice and definition of the infill criterion are the main distinctive characteristics of each surrogate-based technique [3]. In particular, the infill criteria specifically designed for multi-fidelity sampling are able to determine not only the next infill sample but also the fidelity level to consider for the design evaluation.

Consider a generic multi-fidelity sampling problem with a set of n_l fidelity levels $\mathbf{L} = \{l_0, \ldots, l_{n_l}\}$ (with l_0 and l_{n_l} as the highest and lowest fidelity levels, respectively). At a generic iteration i, the identification of the infill sample location and fidelity level for iteration $i+1$ can be formalized as:

$$\mathbf{x}^{(i+1)}, l^{(i+1)} = \arg\max_{\mathbf{x},l} A(\mathbf{x}, l) \tag{1a}$$

$$x_{\mathrm{LB},k} \leq x_k \leq x_{\mathrm{UB},k} \tag{1b}$$

$$l \in \mathbf{L} \tag{1c}$$

where $\mathbf{x}$ is the vector of the design variables, A is the acquisition function of the selected infill criterion, and $x_{\mathrm{LB}}, x_{\mathrm{UB}}$ are the lower and upper bounds, respectively, that define the design space. Given that the fidelity level is represented by the integer variable l, Equations (1a)–(1c) is a mixed integer optimization problem (even though it is simple), and it requires specific techniques to be efficiently solved.

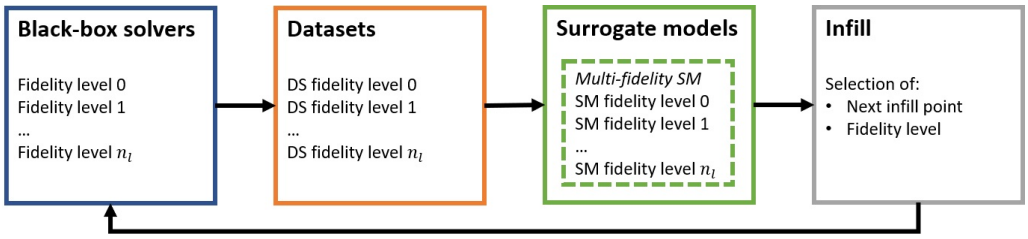


Figure 1. Generic multi-fidelity surrogate-based architecture.

Some important notes for the reader: henceforth, the term "multi-fidelity" is omitted before "surrogate model" unless explicitly necessary. Hence, any mention of "surrogate model" in the following text should be understood to refer to a multi-fidelity surrogate model. For instance, the term "highest fidelity surrogate model" denotes the multi-fidelity surrogate model trained by utilizing all available data from every fidelity level. In addition, the surrogate models are assumed to be statistical models, i.e., models that, when queried, return a statistical distribution representing the output uncertainty given the training data and the selected model functional form. Furthermore, different infill criteria may require either the minimization of the maximization of their respective acquisition functions. Without loss of generality, the acquisition function is assumed to be maximized in the remaining part of this manuscript.

2.1. Two-Step Multi-Fidelity Sampling

The multi-fidelity sampling technique presented in this manuscript is derived from the general formulation in Equations (1a)–(1c) and it is designed as an extension of any existing single-fidelity infill criterion. Specifically, the selection of the fidelity level is decoupled from the identification of the next sample, which is completed by optimizing the single-fidelity acquisition function on the highest fidelity surrogate model. The fidelity level is then chosen as the most cost-efficient one for which the associated surrogate model is judged "sufficiently accurate" at the previously identified infill location. Every iteration of the proposed surrogate-based multi-fidelity sampling scheme consists of the following:

1. Determining the location of the next infill sample by the maximization of the acquisition function over the highest fidelity surrogate model:

$$\mathbf{x}^{(i+1)} = \arg\max A(\mathbf{x})|_{\overline{\mathrm{SM}}_0} \tag{2}$$

   where $\overline{\mathrm{SM}}_0$ is the multi-fidelity surrogate model of the highest fidelity, i.e., level 0.
2. Identifying the set of "accurate enough" fidelity levels at the next infill location by means of the selected accuracy metric:

$$\mathbf{L}_{\mathrm{acc}}|_{\mathbf{x}^{(i+1)}} = \{l \in \mathbf{L} \,|\, l \text{ is accurate with respect to the accuracy metric at } \mathbf{x}^{(i+1)}\} \tag{3}$$

3. Selecting the fidelity level to use for the evaluation of the next infill sample as the fastest one within the "accurate enough" list:

$$l^{i+1} = \arg\min_j \left(\left[t_{\mathrm{eval},j} \right] \right) \qquad \text{for } j \in \mathbf{L}_{\mathrm{acc}}|_{\mathbf{x}^{(i+1)}} \tag{4}$$

   where t_{eval} is the fidelity level evaluation time.

Obviously, this is a generic description of the two-step multi-fidelity sampling scheme that has to be integrated in the specific surrogate-based implementation, e.g., surrogate-based optimization (SBO) in Section 2.4 and surrogate-based uncertainty quantification (SBUQ) in Section 2.5.

The sample location $\mathbf{x}^{(i+1)}$ resulting from the solution of Equation (2) has a couple of important properties. First of all, it is determined by leveraging the highest fidelity surrogate model, in other words, a surrogate model that is trained using all the available

data. This sample is, by the definition of a surrogate-based sampling scheme, the most effective location in the domain with respect to the infill metric. Moreover, the choice of the evaluation fidelity level (which occurs in the second step) does not affect the location of the next sample, unlike in other multi-fidelity schemes [16,18]. This characteristic of the sampling scheme is based on the assumption that, if the location of the next sample is determined by the most accurate surrogate model, there is no rationale for altering it depending on the evaluation fidelity level, and thus the evaluation cost and accuracy.

The design performance of the infill sample is then evaluated by using the fastest fidelity level (Equation (4)), for which the associated surrogate model is judged "sufficiently accurate" at the infill location with respect to an accuracy metric. In contrast to other multi-fidelity sampling techniques where the accuracy metric (or a function depending on it) is directly multiplied by the acquisition function [16,18], the proposed two-step approach uses the accuracy metric to perform a binary classification of the fidelity levels in "accurate" and "inaccurate" (Equation (3)). A popular approach to assess the expected accuracy of a generic fidelity level l at a given location $\mathbf{x}$ is by evaluating the surrogate model of level l at $\mathbf{x}$ and comparing the returned probability distribution with the one obtained from the evaluation of the surrogate model of the highest fidelity level (i.e., level 0). Several metrics have been proposed to compare the two distributions, with popular examples being the correlation factor and the Kullback–Leibler divergence (KLD or KL-divergence). However, both metrics have limitations. The correlation factor estimates the correlation between two sets of observation data rather than the distance and similarity of two probability distributions, like the KL-divergence. While the KL-divergence is a proper statistical distance, it is not symmetric and does not have an upper bound. The first limitation can be addressed by always computing the KL-divergence between the highest and the lower fidelity (and not vice versa), but the lack of an upper bound poses challenges in defining a threshold below which the considered fidelity level is considered "accurate". For these reasons, the Jensen–Shannon divergence (JSD or JS-divergence), a symmetric statistical distance bounded between 0 and 1, is adopted in this study to assess the accuracy of the multi-fidelity surrogate models as described in the next section (Section 2.2). While the need to specify the accuracy metric threshold (Equation (3)) may initially seem like a drawback of the method, it also provides users with the flexibility to adjust the algorithm's behavior according to their preferences. For example, the threshold might be actively reduced towards the end of the optimization in order to ensure that the majority of samples are evaluated at the highest fidelity. This adaptability allows users to tailor the algorithm to the specific application.

2.2. Jensen–Shannon Divergence as Accuracy Metric

The choice of the accuracy metric is crucial for the behavior of the multi-fidelity sampling schemes (Equation (3)). However, both of the previously mentioned options have some limitations: the correlation factor is not formally a distance metric between two distributions, and the KL-divergence is not symmetric (i.e., $\mathrm{KLD}(A, B) \neq \mathrm{KLD}(B, A)$), a property that is relevant for the assessment of the prediction accuracy between two models. For this reason, the multi-fidelity sampling technique presented in this manuscript adopts the JS-divergence as an accuracy metric.

The Jensen–Shannon divergence between two probability distributions A and B is a symmetric and smoothed version of the KL-divergence, and it is bounded between 0 (identical distributions) and 1 (infinitely different distributions):

$$\mathrm{JSD}(A||B) = \frac{1}{2}\mathrm{KLD}(A||M) + \frac{1}{2}\mathrm{KLD}(B||M) \tag{5}$$

$$0 \leq \mathrm{JSD}(A||B) \leq 1 \tag{6}$$

where M is a mixture distribution of P and Q (the upper bound in Equation (6) depends on the logarithm bases used in the evaluation of KLD in Equation (5)). Depending on the formulation of the mixture distribution, it is possible to derive the closed form of the

JSD for some probability distribution families [21]. However, a discrete approximation of the JSD is a valid and fast alternative to the closed form for the purposes of multi-fidelity sampling (Section 2.6). Combining Equation (3) with Equation (5), the list of expected accurate fidelity levels at the next infill sample $\mathbf{x}^{(i+1)}$ is:

$$\mathbf{L}_{\text{acc}} = \left\{ l \in \mathbf{L} \mid \text{JSD}\left(P(\overline{\text{SM}}_l) || P(\overline{\text{SM}}_{l_0})\right)\Big|_{\mathbf{x}^{(i+1)}} < \text{JSD}_{\text{lim}} \right\} \tag{7}$$

where $P(\overline{\text{SM}}_l)$ is the probability distribution returned by a query of the surrogate model of fidelity level l at $\mathbf{x}^{(i+1)}$, and JSD_{lim} is the maximum acceptable JSD value to consider the fidelity level l "accurate enough".

The value of JSD_{lim} is a parameter that has to be set by the user and that controls the minimum accuracy requirement for a fidelity level to be considered "accurate enough" for the next sampling iteration. Lower values JSD_{lim} make the accuracy requirement more strict, therefore reducing the probability that the algorithm selects a lower fidelity level for the next sampling iteration. Even though the JSD_{lim} is a user-defined parameter, the property of JS-divergence of being bounded between 0 and 1 helps to identify some target values that are "empirically" meaningful and that are valid irrespective of the magnitude of the probability distributions that are compared. For instance, the JS-divergence between two normal distributions, each sharing the same standard deviation and with mean values differing by two times the standard deviation (Figure 2), approximates 0.7 regardless of the particular values assigned to the mean and standard deviation:

$$\text{JSD}(P_A || P_B) \approx 0.7 \qquad \text{with } P_A = N(\mu_a, \sigma), \quad P_B = N(\mu_a + 2\sigma, \sigma) \tag{8}$$

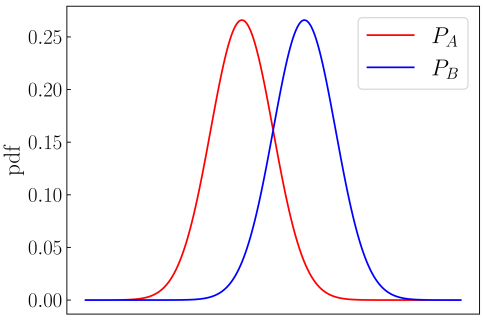


Figure 2. Normal distributions used to determine the JSD in Equation (8).

2.3. Comparison with Multiplicative Sampling Schemes

The decoupling between the identification of the next infill sample and the selection of the fidelity level distinguishes the proposed two-level approach from other methods based on the multiplicative multi-fidelity expected improvement [16,18]. In these sampling schemes, both the location and fidelity level of the next sample are determined simultaneously by maximizing a single acquisition function that is the multiplication of the single fidelity acquisition function and other factors that take into consideration the accuracy and evaluation cost of the fidelity levels. Therefore, the location and fidelity level of the next sample result from the solution of a single-objective optimization problem obtained from the multiplication of three distinct and conflicting objectives, i.e., the maximization of the single-fidelity acquisition function, the maximization of the accuracy between the selected and the highest fidelity levels, and the minimization of the evaluation time of the selected fidelity level. As a consequence, this approach shares the primary drawback of the "weighted sum" technique applied to solve a multi-objective optimization problem as a single-objective problem: if the multiplicative factors lack appropriate scaling, the effect of one factor may be dominant and drive the whole optimization. For instance, consider a multi-fidelity sampling scenario where a mid-fidelity level has a decent prediction accuracy and significantly lower evaluation costs compared to the highest fidelity level. In such a case, the multiplicative algorithm may consistently favor the lower-fidelity solver because

the multiplicative acquisition function tends to offset the penalty associated with the reduced accuracy with the benefit of decreased computational cost. This limitation is critical in a surrogate-based optimization application because a sample is considered optimal only if it is evaluated at the highest fidelity level. If no sample is evaluated at the highest fidelity level, it is impossible to update the knowledge about the current optimum.

The proposed two-level multi-fidelity sampling effectively addresses such a limitation of multiplicative techniques. The difference in magnitude of the infill metric, accuracy metric, and evaluation cost is irrelevant in the proposed criterion because these metrics are considered in different stages of the infill sample and fidelity level selection phase, therefore removing any risk of interaction. Consequently, the two-step sampling scheme is safeguarded against always selecting the same fidelity level as observed in multiplicative approaches [16,18].

2.4. Surrogate-Based Optimization

The popularity of surrogate-based optimization techniques has resulted in the development of numerous acquisition functions. Popular examples are probability of improvement (PI) [22], knowledge gradient (KG) [23], and expected improvement (EI) [24]. Out of these, expected improvement is arguably the most popular given its elegant and simple formulation, and for this reason, it is adopted in the SBO analyses presented in this manuscript. Even though a complete derivation of the expected improvement is out of the scope of this article (the whole derivation is available in [24]), a brief description is provided in the following paragraph. At a generic iteration i, the goal of the expected improvement technique is to identify the location in the domain that maximizes the expected reduction in the objective function value with respect to the current minimum, i.e., the expected improvement:

$$\mathrm{EI}(\mathbf{x}) = \mathrm{E}(\max(f_{\min} - Y, 0)) \tag{9}$$

where $f_{\min}$ is the current minimum and Y is the probability distribution returned by a query of the surrogate model at location $\mathbf{x}$. Under the assumption that Y is a normal distribution ($Y = N(\mu, \sigma)$), the closed form of Equation (9) is [24]:

$$\mathrm{EI}(\mathbf{x}) = (f_{\min} - \mu)\Phi\left(\frac{f_{\min} - \mu}{\sigma}\right) + \sigma\phi\left(\frac{f_{\min} - \mu}{\sigma}\right) \tag{10}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ represent the cumulative distribution and probability density functions or the standard normal distribution, respectively. Equation (10) is the acquisition function that is maximized at each optimization iteration in order to identify the location of the next infill point as described in Section 2, Equation (2).

Usually, real design optimization problems are characterized by several constraints that are introduced to guarantee a certain level of feasibility in the design space. Several techniques are available to handle constraints within surrogate-based optimization schemes [25,26], and popular choices are the penalty function and the probability of feasibility. The former is straightforward to implement, but its effectiveness is strongly dependent on the user-defined penalty values, which may artificially restrict the design space beyond what is necessary. For this reason, the constraints are handled in this manuscripts by means of the probability of feasibility [27]. Assuming that the problem has n_g constraints defined as:

$$g_i(\mathbf{x}) \leq 0 \qquad \text{with} \quad i = 1 \dots n_g \tag{11}$$

and a surrogate model is built for each of them, the probability of the feasibility of a generic design $\mathbf{x}$ with respect to all the constraints is computed as:

$$\mathrm{PoF}(\mathbf{x}) = \prod_{i=1}^{n_g} P(g_i(\mathbf{x}) \leq 0) \tag{12}$$

where $P(g_i(\mathbf{x}) \leq 0)$ is easily computed from the probability distribution returned by the evaluation of the constraint surrogate model. The resulting acquisition function based on expected improvement and probability of feasibility is obtained by multiplying Equation (10) and Equation (12):

$$A(\mathbf{x}) = \mathrm{EI}(\mathbf{x})\mathrm{PoF}(\mathbf{x}) \tag{13}$$

and represents the acquisition function that is used to obtain all optimization results presented in this manuscript.

Another important aspect to consider in multi-fidelity surrogate-based optimization is the potential presence of objective and constraint functions with a different number of fidelity levels. This scenario is quite common, especially in cases where multiple disciplines (e.g., aerodynamics and structures) are involved in the optimization process. It is important to remember that the two-step approach described in Section 2 independently selects the fidelity level for each function within the problem. Consequently, the correct implementation of a multi-fidelity surrogate-based optimization requires the definition of a logic to determine which fidelity level to use in situations where the selected fidelity levels are not identical. In a single-objective multi-fidelity problem, the minimum is obviously obtained by comparing only the feasible data obtained at the highest fidelity. Therefore, it is imperative that whenever the objective function is evaluated at the highest fidelity, all the constraints are also assessed at the highest fidelity. In contrast, when the objective function is not evaluated at the highest fidelity level, constraints can be evaluated at the fidelity level selected by the scheme presented in Section 2 Equation (4). Denoting with l_o and l_{g_i} the selected fidelity level for the objective and constraint functions, respectively, the combined logic for the selection of the constraint fidelity level can be formalized as:

$$l_{g_i} = \begin{cases} 0 & \text{if} \quad l_0 = 0 \\ l_{g_i} & \text{otherwise} \end{cases} \tag{14}$$

2.5. Surrogate-Based Uncertainty Quantification

Similar to SBO for optimization, surrogate-based uncertainty quantification (SBUQ) aims to transfer the uncertainties present in the inputs effectively to the Quantity of Interest (QoI), and subsequently, to accurately determine the relevant statistics of QoI [28]. Instead of employing the Monte Carlo method, which involves direct computation using complex and resource-intensive black-box functions, this approach applies the method to a more straightforward and computationally inexpensive surrogate model that approximates the behavior of the original function. In the present work, this is achieved by constructing a Kriging surrogate model, which is formulated based on an initial Design of Experiments (DoE) in the stochastic domain, coupled with its black-box solutions. To improve the accuracy of the surrogate, an active infill criterion based on the (approximated) statistics is adopted. Here, the infill criterion focuses on the mean value of the QoI, seeking to achieve high global accuracy of the surrogate model.

To ensure a balanced space filling in the stochastic space, the infill criterion uses the prediction mean square error $\hat{s}(\boldsymbol{\xi})$ at any given point $\boldsymbol{\xi}$ in the Kriging surrogate. At every infill iteration, the surrogate model is (re)constructed at the new location $\boldsymbol{\xi}^*$ where the product of the joint probability distribution function of the input uncertainties PDF_{ξ} and the error estimate is maximized:

$$\boldsymbol{\xi}^* = \arg\min_{\boldsymbol{\xi}} \{ -\mathrm{PDF}_{\xi}(\boldsymbol{\xi})\, \hat{s}(\boldsymbol{\xi}) \} \tag{15}$$

The PDF_{ξ} component ensures sampling from the regions of high probability in the stochastic space, and the error term $\hat{s}(\boldsymbol{\xi})$ exploits regions where the surrogate model is inaccurate. We use differential evolution to search for the optimal location in the surrogate.

The statistic of the QoI is then approximated using a large number of Quasi Monte Carlo sample evaluations of the surrogate model.

2.6. Practical Implementation Details

The presented multi-fidelity surrogate-based sampling architecture is implemented in the Surrogate Modeling for AeRo-data Toolbox in python (SMARTy) [29] that is continually developed at the German Aerospace Center (DLR). While the implementation of the two-step multi-fidelity sampling technique is straightforward and it can be readily adapted to existing surrogate-based algorithms, there are a couple of specific aspects that should be highlighted.

As described in Section 2.1, the JS-divergence has a closed form for some families of probability density distributions [21] (e.g., normal distributions). However, the closed forms might be tedious to implement, therefore increasing the risk of coding errors. For this reason, the JS-divergence is numerically computed in this manuscript using the ready-to-use `spatial.distance.jensenshannon` method of Scipy python library [30]. The method accepts two probability vectors as input, representing the two probability distribution, for which JS-divergence is to be computed. In particular, if the two distributions are normal distribution ($P_A = N(\mu_A, \sigma_A)$ and $P_B = N(\mu_B, \sigma_B)$), the probability vectors are defined by evaluating the probability density function of the two probability distributions at linearly spaced locations in a range defined as:

$$\text{LB} = \min([\mu_A - 3\sigma_A, \mu_B - 3\sigma_B]) \tag{16}$$

$$\text{UB} = \min([\mu_A + 3\sigma_A, \mu_B + 3\sigma_B]) \tag{17}$$

The second aspect to specify is the algorithm adopted for the optimization of the acquisition function Equation (2). As described in Section 2, the two-step multi-fidelity sampling approach decouples the identification of the next infill point from the selection of the fidelity level, and the former results in the solution of a single objective global optimization problem. In all the applications presented in this manuscript, the acquisition function is optimized by means of a differential evolution algorithm. This choice is guided by the multi-modal nature of the optimization problem and the computationally inexpensive evaluation of the acquisition function, thus suggesting the utilization of a stochastic global algorithm for its optimization.

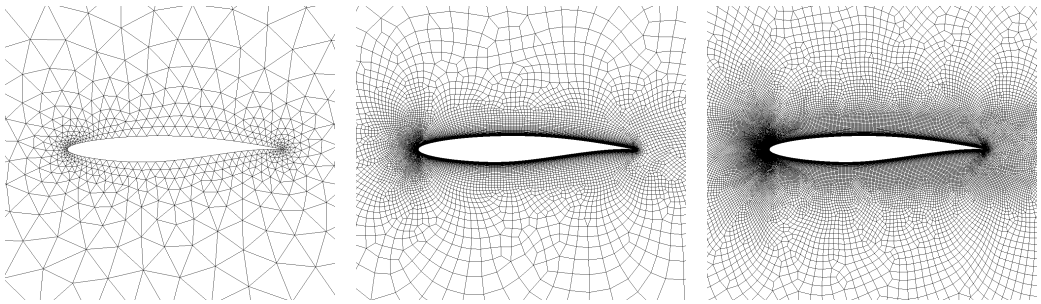
3. Application

The following sections showcase the newly proposed multi-fidelity sampling techniques for two aerodynamic challenges, namely, optimizing the aerodynamic performance of a configuration within a given design space and propagating known uncertainties within inputs through an aerodynamic analysis. The results are presented using four different fidelity levels ranging from potential flow up to RANS solutions. In particular, the CFD solver for the RANS simulations utilizes a Spalart–Allmaras turbulence model, a central scheme with scalar artificial dissipation for the inviscid fluxes discretization, and the Green–Gauss approach for the computation of the exact gradient of viscous and source terms. The simulations are considered converged once the density residual is below $1 \cdot 10^{-7}$.

3.1. Test Case

The supercritical RAE2822 airfoil serves as a baseline configuration for this investigation, and its aerodynamics performance is assessed by four different fidelity levels. The lowest fidelity level is obtained by solving the potential flow, including a compressibility correction employing the XFoil software v.6.99 [31,32]. The second lowest fidelity level solves the Euler equations, while the two highest fidelity levels are based on the RANS equations in conjunction with the Spalart Allmaras turbulence model, applied to two grids with varying mesh density. The CFD solutions based on a finite volume method, i.e.,

solutions to the Euler and RANS equations, are computed using the DLR TAU Code [33], and the corresponding computational grids are shown in Figure 3.



(**a**) Euler (**b**) Coarse RANS (**c**) Fine RANS

Figure 3. Different meshes for the CFD solver.

The flow condition for both the optimization and uncertainty quantification task is defined by the angle of attack, and the Mach and the Reynolds numbers listed in Table 1. The pressure coefficient distributions obtained by the four different fidelity levels are depicted in Figure 4. The RANS solutions are almost identical, with a shock wave on the suction side at approximately 10% of the chord length. Larger deviations in the pressure coefficient distribution are visible when analyzing the Euler solution: the shock is located further downstream, and a lower-pressure coefficient value is obtained upstream of the shock. Finally, as expected from the potential flow formulation, the XFoil result features a sole peak on the suction side without capturing any shock. The representative evaluation time (rounded average out of 50 simulations) for the four fidelity levels, along with the resulting integral aerodynamics coefficients (i.e., the lift coefficient (C_l), the drag coefficient (C_d), and their ratio C_l/C_d), are given in Table 2. Similarly to the pressure coefficient distributions, the RANS solutions exhibit only minor differences in both C_l and C_d, resulting in an accumulated error of roughly 6.6% in C_l/C_d. Because viscous effects are neglected during Euler simulations, larger differences emerge when comparing the highest-fidelity results with those obtained from solving the Euler equations, resulting in an error in C_l/C_d exceeding 100%. Interestingly, the integral aerodynamics coefficients determined by solving the potential flow equations with a compressibility correction exhibit closer agreement with the RANS solutions than with the Euler solver, despite the complete misrepresentation of the pressure distribution in the fore region of the suction side. When comparing the evaluation time of the different solvers, the anticipated trend is observed: decreasing fidelity simultaneously yields substantial speed-up.

Table 1. Flow conditions.

Flow Condition	Value
Angle of Attack	2.79 [°]
Mach Number	0.676
Reynolds Number	$5.7 \cdot 10^6$

The Euler solver has also been tested on meshes finer than the one shown in Figure 3a, resulting in negligible improvement in prediction accuracy at the cost of a substantial increase in computational time. Note that the goal of this manuscript is to propose a new multi-fidelity sampling technique capable of handling situations where some portions of the domain are better described by a lower fidelity level than a higher one. Consequently, the four fidelity levels selected for this study are representative of such a possible scenario.

Note that in the following analysis, the evaluation times reported in Table 2 are consistently utilized to determine the fidelity level of the next infill point (Equation (4)),

even though the actual evaluation times may be influenced by factors such as the underlying geometry and operating conditions of the simulation.

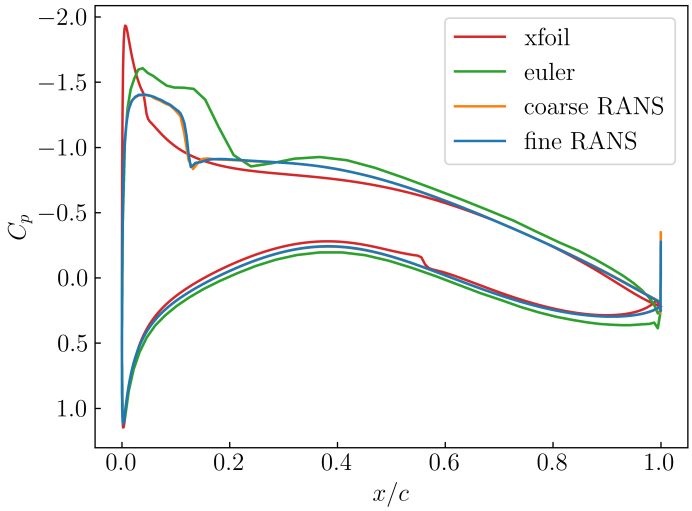


Figure 4. Pressure distribution resulting from the different CFD fidelity levels.

Table 2. Aerodynamics coefficient and evaluation time for different fidelity levels.

Fidelity Level	Solver	Eval. Time [s]	C_l	C_d	C_l/C_d
0	Fine RANS	1500.0	0.7057	0.0103	68.51
1	Coarse RANS	500.0	0.7068	0.0110	64.25
2	Euler	2.0	0.8377	0.0053	158.05
3	XFoil	0.3	0.6414	0.0077	83.29

3.2. SBO

The two-step multi-fidelity sampling presented in Section 2 is tested in a surrogate-based shape optimization of the airfoil described in Section 3.1. In particular, the class/shape function transformation technique (CTS) coefficients that define the airfoil shape are optimized in order to maximize the lift-to-drag ratio at a given flow condition. Moreover, a constraint on the minimum feasible maximum thickness is introduced in order to avoid unreasonably thin airfoils. The resulting shape optimization problem can be written in standard form as follows:

$$\begin{aligned} &\mathbf{x}^* = \arg\max_{\mathbf{x}}(C_l(\mathbf{x})/C_d(\mathbf{x})) \quad \text{with} \quad \mathbf{x} \in \mathbb{R}^9 \\ &\text{s.t.} \quad x_{i,l} \leq x_i \leq x_{i,u} \quad \text{and} \quad \theta_{\min} - \theta(\mathbf{x}) < 0 \end{aligned} \tag{18}$$

where $\mathbf{x}$ is the vector of the nine design variables bounded between $x_{i,l}$ and $x_{i,l}$, and θ is the maximum thickness of the airfoil shape that must be greater than $\theta_{\min}$. The lower and upper bounds are defined by subtracting and adding 0.075 to the design variables of the original shape Table 3. This choice both guarantees enough freedom in the domain space exploration and prevents the creation of a large number of self-intersecting or unreasonable airfoil shapes.

Table 3. Bounds of the design variables.

x_i	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
$x_{i,l}$	0.05372	0.06228	0.11768	0.11687	0.12716	0.06146	0.14980	0.00101	−0.11397
$x_{i,u}$	0.20372	0.21228	0.26768	0.26687	0.27716	0.21146	0.29980	0.15101	0.03602

Firstly, the optimization problem in Equation (18) is initially solved using a single-fidelity surrogate based optimizer. Such an optimization solution establishes a baseline performance, which will later serve as a point of comparison for the performance of the surrogate-based optimizer with the two-step multi-fidelity sampling. The initial surrogate model is a Gaussian process model with a Gaussian Exponential kernel and constant mean, and the training is completed using 50 initial samples generated via a Sobol sequence. The choices of the Gaussian process kernel and of the number of initial samples aim at achieving a root mean squared error below 10% and a maximum prediction error below 20% with respect to a test dataset made of 10 random samples. The convergence history of optimal C_l/C_d is visualized in Figure 5 as a function of unit time $\hat{t}$, where $\hat{t}$ represents the cumulative evaluation time divided by the evaluation time of the highest fidelity level, namely, the fine RANS simulation.

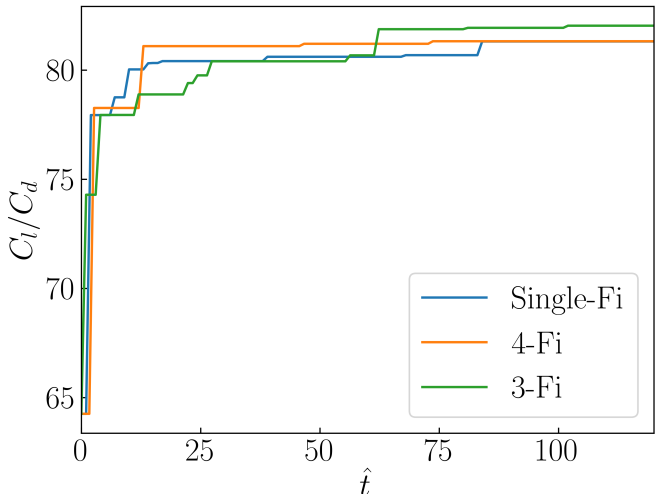
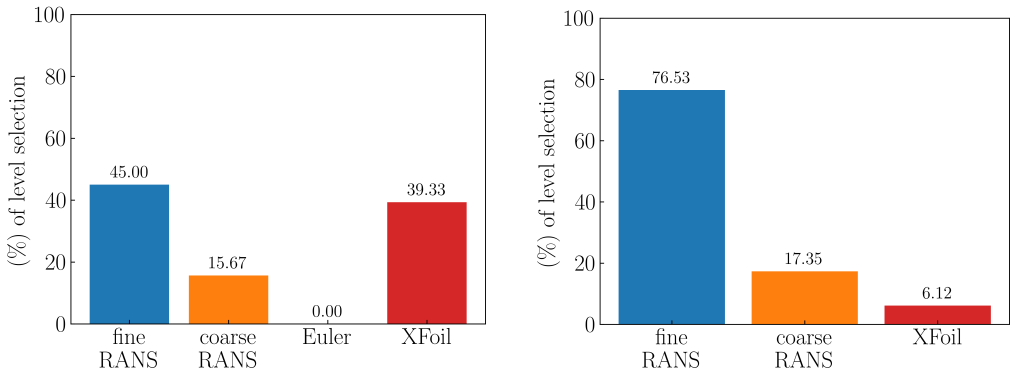


Figure 5. History of the optimum for the three optimizations with respect to the unit time, i.e., time divided by the evaluation time of the fine RANS.

A second optimization is completed by enabling the two-step multi-fidelity sampling technique presented in this manuscript. The selection of model kernels and the choice of the initial training dataset sizes for each fidelity level are calibrated to ensure that the resulting initial model has a comparable precision to that of the single-fidelity case (i.e., a root mean squared error below 10% and a maximum prediction error below 20% when evaluated against the same test dataset made of 10 random samples). As a result, all fidelity levels employ a Gaussian Exponential kernel with a constant mean, and the number of initial samples is set to 20, 30, 120, and 380 for fine RANS, coarse RANS, Euler, and XFoil, respectively. Taking into account the evaluation time of each fidelity level (Table 2), the total time required for evaluating the initial samples using the multi-fidelity model is 40% lower (45,354 s) compared to the single-fidelity approach (75,000 s). When considering the optimization process, the advantage of employing multi-fidelity sampling becomes apparent through the convergence history of the optimal C_l/C_d as shown in Figure 5. The proposed multi-fidelity sampling method converges to $C_l/C_d \approx 81.12$ in approximately 17 unit times, compared to the single-fidelity approach, which converges around 82 unit times: this translates to a speed-up factor of approximately 5 times. Further insights into the behavior of the multi-fidelity sampling can be obtained from the distribution of function calls across fidelity levels as depicted in Figure 6a. Notably, as expected, the Euler solver is consistently omitted from selection due to its poor prediction of C_l/C_d (refer to Table 2). Additionally, XFoil holds a significant portion of function calls, likely because of its low evaluation time, which proves advantageous during the exploration of the design space.

The initial multi-fidelity optimization was designed as an academic test to validate the consistency of the proposed multi-fidelity sampling method. Specifically, the preliminary

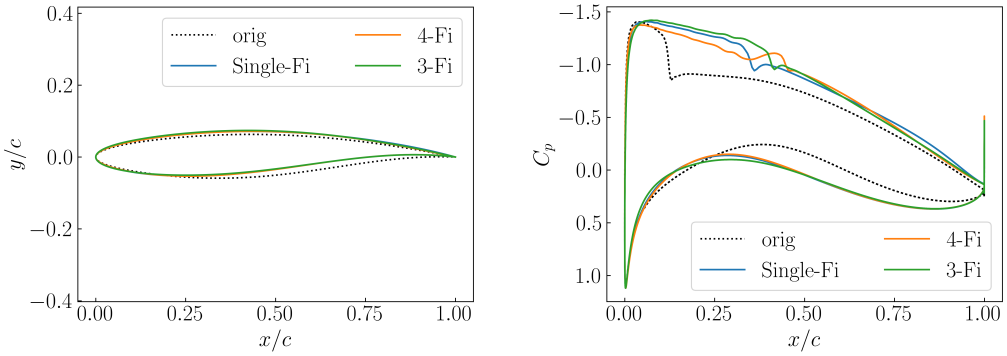
analysis of the prediction accuracy across the four fidelity levels (Table 2) clearly indicates that there may be no advantage in including the Euler solver because a faster fidelity level (XFoil) demonstrates superior prediction accuracy. Moreover, the negligible evaluation time required by XFoil suggests that the number of initial samples assessed with XFoil can be substantially augmented without noticeably increasing the evaluation time of the initial samples. For this reason, a second multi-fidelity optimization is conducted, omitting the Euler solver from the list of available fidelity levels and increasing the number of XFoil initial samples to 1500. This results in a total evaluation time for the initial sample of approximately 45,450 s, which is comparable to the value obtained for the first multi-fidelity optimization (45,354 s). Similar to the previous analysis, the convergence of the optimal C_l/C_d value is plotted in Figure 5, while the fidelity level selection percentage is illustrated in Figure 6b. As anticipated, the substantial number of initial XFoil samples yields an enhanced representation of the design space at this fidelity level, consequently reducing the need for XFoil queries during the optimization phase. Regarding the convergence of the optimal value, a slower initial reduction in optimal value is succeeded by a better final optimal value. This trend might arise from the finer sampling of the design space resulting from the increased number of initial XFoil samples. The resulting enriched representation of the design space may contain a greater number of promising regions that the algorithm must explore, thereby extending the initial exploration phase. However, the enhanced accuracy of the initial surrogate model ultimately guides the algorithm toward convergence to a better C_l/C_d value compared to both the single-fidelity and first multi-fidelity optimization.



(**a**) Four fidelity levels (**b**) Three fidelity levels

Figure 6. Percentage of fidelity level evaluations during the optimization for the 4 and 3 fidelity cases.

The resulting optimal design variables are presented in Table 4, while the optimal airfoil shapes and pressure distributions are compared against the original shape in Figure 7. All optimizations succeed in maximizing C_l/C_d by shifting the position and reducing the strength of the shock wave on the suction side, and by increasing the pressure on the second half of the pressure side. Interestingly, all the optimal airfoil shapes are extremely similar, indicating that all the conduced optimizations successfully identified the characteristics of the optimal shape. In case a finer optimization is desired, a gradient-based or a surrogate-based trust-region algorithm can be employed to further refine the optimal solution.



(**a**) Airfoil shapes (**b**) C_p distributions

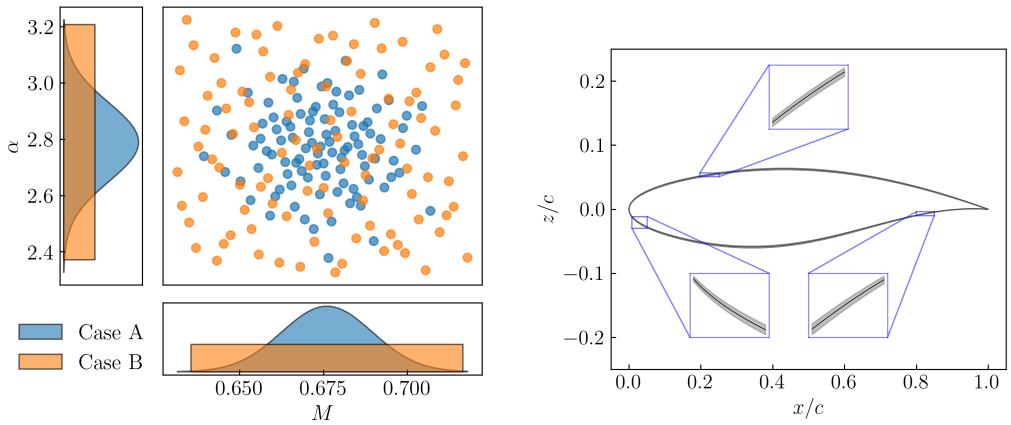
Figure 7. Comparison of optimal airfoil shapes and C_p distributions with the original airfoil.

Table 4. Optimal design variables.

Shape	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
Original	0.12872	0.13728	0.19268	0.19187	0.20216	0.13646	0.22480	0.07601	−0.03897
Single fidelity	0.14370	0.16171	0.21617	0.20375	0.27716	0.15377	0.14980	0.00101	−0.11397
4 fidelity levels	0.13766	0.16652	0.20038	0.244059	0.21019	0.159159	0.14981	0.00101	−0.11397
3 fidelity levels	0.15398	0.15974	0.23295	0.21245	0.23939	0.12518	0.15903	0.00101	−0.11397

3.3. SBUQ

The surrogate-based uncertainty quantification described in Section 2.5 is employed to propagate uncertainties in (i) operating conditions—Mach number M and angle of attack α, and (ii) geometrical representation of the airfoil using CST coefficients (Section 3.1). As shown in Figure 8, case A and B (later B1, B2) refer to low (informative) and high (uninformative) uncertainty levels (distributions) in operating conditions, respectively. In order to study the effect of varying JSD_{lim}, its default value of 0.7 is used for cases A and B1, and a value of 0.5 is used for case B2. The uncertainty in (CST) design vector $\mathbf{x} = [x_1, \ldots, x_9]^T$ results in about 1% variation in the airfoil geometry and is consistent for all three cases. The Quantity of Interest (QoI) considered is the lift-to-drag ratio which is the same as that considered for optimization. However, under the presence of uncertainties, a rather important quantity is the mean of QoI.



(**a**) Operational uncertainties (**b**) Uncertainty in airfoil geometry.

Figure 8. Uncertainties considered for forward UQ analysis.

To accumulate the information from all four levels of fidelity discussed in Section 3.1, along with a single-fidelity (highest) Gaussian process with a Gaussian Exponential kernel, a multi-fidelity surrogate model is constructed using a hierarchical approach (see

Appendix A). The mean lift-to-drag ratio is computed using 10,000 Quasi Monte Carlo samples in the surrogate model. In the single-fidelity analysis, the infill phase is considered to be converged when the value of the statistic, for a given threshold, does not change over two consecutive iterations. Similarly, for the multi-fidelity analysis, convergence is assumed when the statistic value, for two consecutive iterations, is close to that obtained from single-fidelity analysis, under a small threshold.

Figure 9 shows the infill history of multi-fidelity surrogate-based uncertainty propagation. Note that (i) iteration zeros refer to the last sample in the DoE stage, and (ii) the converged statistic values for single-fidelity investigation are shown with black squares. The multi-fidelity approach for Case A results in samples mostly from the highest fidelity with only a couple of samples from the lower levels. For relatively precise inputs in Case A, the surrogate errors are rather small. As a consequence, the Jensen–Shannon divergence (JSD) between these surrogates at different locations is mostly larger than the default value, resulting in samples predominantly from the highest fidelity level. As the input uncertainty is increased, i.e., Case B1, the surrogate errors become larger, resulting in the majority of the samples being from lower fidelity levels. Consequently, the convergence takes more iterations as compared to Case A. In Case B2, with the same level input uncertainty as Case B1, the $\mathrm{JSD}_{\mathrm{lim}}$ is lowered to a value of 0.5. This allows for a balanced sampling between fidelity levels with enough samples from the highest fidelity to make an accurate prediction of the mean lift-to-drag in a smaller number of infill iterations.

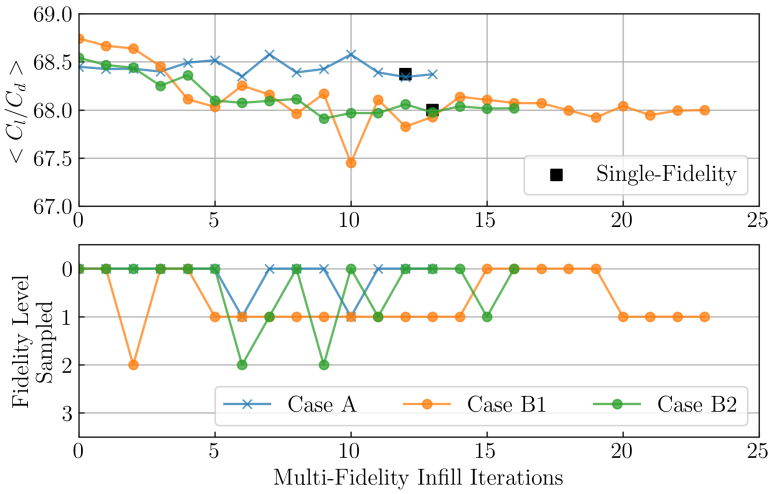


Figure 9. Infill iterations history for $< C_l/C_d >$ and fidelity level sampled for low (Case A) and high uncertainty levels (Cases B1 and B2). Iteration 0 refers to the last sample in the DoE stage.

Box plots indicating the variability in the lift-to-drag ratio for 10,000 realizations of the surrogate model under different fidelity levels and uncertainty cases are shown in Figure 10. As the uncertainty level increases, as expected, the single-fidelity analysis shows an increase in variance. In Case A, pertaining to lower uncertainty, the multi-fidelity approach estimates are fairly accurate. As the uncertainty is increased in Case B1, the multi-fidelity approximation of the variance becomes less accurate. This can be attributed to the inaccuracy in the surrogate model due to the majority of infill samples from the lower fidelity levels. This inaccuracy is, however, dealt with a lower $\mathrm{JSD}_{\mathrm{lim}}$ in Case B2, facilitating a balanced sampling which retains the accuracy of the model and thereby the statistics.

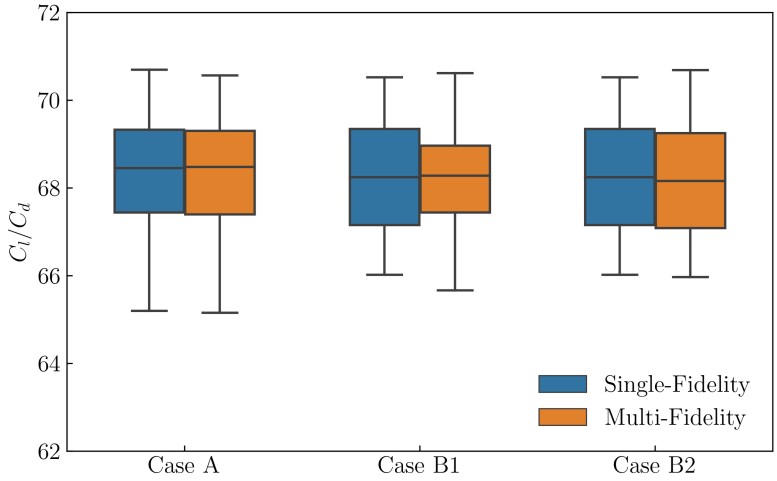


Figure 10. Box plots for 10,000 realizations of C_l/C_d using the final surrogate model based on high- and multi-fidelity infill strategies for low (Case A) and high uncertainty levels (Cases B1 and B2).

Sobol indices are used to attribute the variance in the QoI (C_l/C_d) to different input uncertainties, providing a sensitivity analysis. A high Sobol index for an input parameter indicates its significant impact on the output's variability. The bar charts in Figure 11 compare the sensitivity of the lift-to-drag ratio to the Mach number (M), angle of attack (α), and the CST design vector ($\mathbf{x}$) under single- and multi-fidelity strategies for the low and high uncertainty levels. For all the three cases, the single-fidelity analysis show that the variability in C_l/C_d is dominated by the uncertainty in the angle of attack. Although small, the impact of the Mach number and geometrical uncertainty is not negligible in Case A and Case B(1,2), respectively. Similar to the box-plot results, and for the same reasoning, the multi-fidelity approach estimates of the Sobol indices are fairly accurate for Case A, inaccurate for Case B1, and reasonably accurate for Case B2, indicating the correlation between the uncertainty level and $\mathrm{JSD}_{\mathrm{lim}}$.

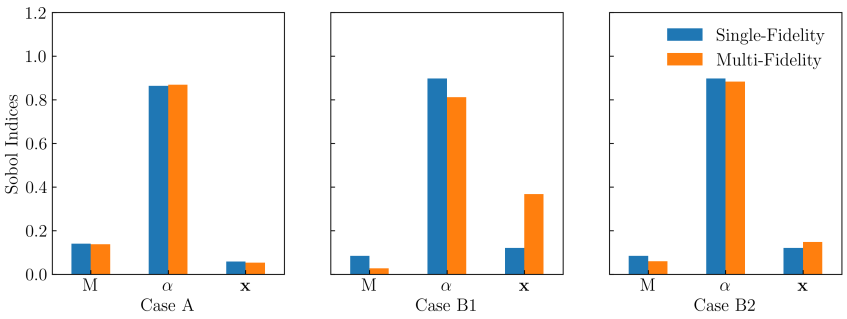


Figure 11. Sobol indices for C_l/C_d with respect to the Mach number (M), angle of attack (α) and design vector ($\mathbf{x}$), using the final surrogate model based on high- and multi- fidelity infill strategies for low (Case A) and high uncertainty levels (Case B1, B2).

Table 5 summarizes the computational cost for single- and multi-fidelity approaches for all the three cases. It is evident that, under different uncertainty levels and hyper-parameter settings, the multi-fidelity approach provides the same level of accuracy as the single-fidelity method at 25–30% lower computational cost. Therefore, multi-fidelity modeling may be an efficient approach for uncertainty quantification, especially when considering the computational resources and time constraints commonly encountered in engineering projects.

Table 5. Comparison of high-fidelity and multi-fidelity evaluations and total cost for different uncertainty levels.

Case	Single-Fidelity		Multi-Fidelity				
	No. of Evals. DoE, Infill	Total Eval. Time [s]	No. of Evals. DoE, Infill				Total Eval. Time [s]
			Fine RANS	Coarse RANS	Euler	XFoil	
A	15, 12	40,500	5, 11	6, 2	100, 0	400, 0	28,320
B1	15, 13	42,000	6, 8	9, 14	100, 1	400, 0	32,822
B2	15, 13	42,000	6, 11	9, 3	100, 2	400, 0	31,824

4. Conclusions

The manuscript introduces a novel infill technique for multi-fidelity sampling in the context of surrogate-based algorithms. In particular, the method decouples the identification of the new infill sample from the selection of the fidelity level that has to be used for its performance evaluation. This key characteristic distinguishes the proposed technique from the majority of existing multi-fidelity schemes which typically rely on a single multiplicative acquisition function. Consequently, the proposed algorithm avoids the risk of continuously sampling from a lower fidelity level in case its computational cost is significantly lower compared to others. Another relevant novelty introduced in this study is the use of the Jensen–Shannon divergence for quantifying the accuracy of the different fidelity levels.

The efficacy of the proposed sampling method was successfully tested in both surrogate-based optimization and uncertainty quantification scenarios. In the optimization test problem, a multi-fidelity surrogate-based optimization algorithm was used for aerodynamic shape optimization with the aim of maximizing the lift-to-drag ratio. The solution resulting from the proposed multi-fidelity sampling is compatible with the optimal solution obtained with a standard single-fidelity sampling. Notably, the multi-fidelity strategy exhibits approximately a five-fold reduction in computational cost compared to its single-fidelity counterpart. Similarly, the multi-fidelity sampling strategy is used for the propagation of operational and geometrical uncertainties to quantify the mean and standard deviation of the lift-to-drag ratio. The statistics obtained from the proposed multi-fidelity sampling approach is not only in accordance with the traditional single-fidelity sampling but also significantly cheaper to compute than the latter. The investigation also highlighted an interesting observation, where the correlation between the level of uncertainty in the inputs and JSD_{lim} can significantly impact the accuracy of the surrogate model and thereby the estimated statistics.

It is important to clarify that the research outlined in this manuscript focuses solely on single-objective optimization. Hence, the logic described in this paragraph is tailored to this specific context. The extension to multi-objective optimization is currently under development, and it will be probably presented in a future publication.

Author Contributions: Conceptualization, A.G., J.P., T.R. and P.B.; methodology, A.G., J.P., T.R. and P.B.; software, A.G.; validation, A.G. and J.P.; formal analysis, A.G. and J.P.; investigation, A.G. and J.P.; writing—original draft preparation, A.G., J.P. and P.B.; writing—review and editing, A.G., J.P. and P.B.; funding acquisition, P.B.; All authors have read and agreed to the published version of the manuscript.

Funding: Parts of this research were funded by the Luftfahrtforschungsprogramm VI-1 (LUFO VI-1) project SMARTfly (SMARTfly—Smart Modeling of flying Transport Vehicles—Entwicklung exakter und effizienter Modellierungen und Simulationsmethoden für den Entwurf von Fluggeräten und Triebwerken, FKZ: 20X1909A) by the German Federal Ministry for Economic Affairs and Climate Action (BMWK). Parts of this research were funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC 2163/1-Sustainable and Energy Efficient Aviation—Project-ID 390881007.

Data Availability Statement: Data Restrictions apply to the datasets: The datasets presented in this article are not readily available because corporate restrictions. Requests to access the datasets should be directed to the authors.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CFD	computational fluid dynamics
C_l	lift coefficient
C_d	drag coefficient
C_l/C_d	lift-to-drag ratio
DoE	design-of-experiment
EI	expected improvement
JSD	Jensen–Shannon divergence
KLD	Kullback–Leibler divergence
$\mathbf{L}$	list of available fidelity levels
$\mathbf{L}_{\text{acc}}$	list of fidelity levels whose surrogate models are sufficiently accurate at the infill sample location
PoF	probability of feasibility
RANS	Reynolds-averaged Navier-Stokes
SBO	surrogate-based optimization
SBUQ	surrogate-based uncertainty quantification
t_{eval}	evaluation time
UQ	uncertainty quantification

Appendix A. Gaussian Process Regression

We outline the process for constructing a Kriging surrogate model using SMARTy. Detailed steps can be found in the reference literature [34] and the references therein. Kriging assumes that the true functional relationship $y : \mathcal{D} \to \mathbb{R}$ between the inputs $x \in \mathcal{D} \subseteq \mathbb{R}^d$ and the corresponding scalar-valued output $y(x) \in \mathbb{R}$ is a realization of a random function,

$$\mathcal{Y}(x) = g(x)\beta + \epsilon(x), \tag{A1}$$

where $g : \mathbb{R}^d \to \mathbb{R}^p$ is a known regression model, β is the unknown regression parameter vector, and $\epsilon(x)$ is a Gaussian process with zero mean and known covariance function given by a known stationary spatial kernel:

$$\text{Cov}[\epsilon(x), \epsilon(x')] = \sigma^2 R_\theta(x, x'), \tag{A2}$$

where $R_\theta(x, x')$ is the correlation kernel, and σ^2 is the marginal variance. SMARTy consists of various kinds of covariance function implementations. In this work, we use one such kernel called the Gaussian Exponential kernel [35]:

$$R_\theta(x, x') = \prod_{k=1}^{m} \exp(-\theta_k |x_k - x'_k|^{p_k}), \quad \text{with } \theta_k \in \mathbb{R}_+,\ p_k \in [1, 2],\ k = 1, \ldots, m \tag{A3}$$

with θ as the unknown correlation parameters [3].

The predictor in Kriging is identified as the optimal linear unbiased estimator and can be reformulated as follows:

$$\hat{y}(x) = g(x)\hat{\beta} + r(x)^\top R^{-1}(Y - F\hat{\beta}), \tag{A4}$$

where R is the correlation matrix, Y is the vector of observed data, F is the matrix of regression vectors, $r(x)$ is the vector of correlations between sample locations and the new location x, and $\hat{\beta} = (F^T R^{-1} F)^{-1} F^T R^{-1} Y$.

The selection of correlation parameters θ significantly influences the Kriging predictor. Typically, these parameters are determined through the maximum likelihood estimate [36]. The mean squared error (MSE) of the Kriging estimator at a location $x \in \mathcal{D}$ is computed using the likelihood-optimal parameter:

$$\mathrm{MSE}(x) = \sigma^2(1 + u^\top (F^\top R^{-1} F)^{-1} u - r(x)^\top R^{-1} r(x)), \tag{A5}$$

with $u = F^\top R^{-1} r(x) - f(x)$ as an estimate for the prediction error variance at location x.

A hierarchical surrogate model integrating multiple data fidelity levels resulting in a variable-fidelity model is implemented in SMARTy [8]. This approach efficiently utilizes high- and low-fidelity data. In the context of Kriging, a low-fidelity model serves as a global trend, formulated as $g(x) = \hat{y}_{lf}(x)$. The multi-fidelity Kriging predictor can be expressed as,

$$\hat{y}(x) = \hat{y}_{lf}(x)\hat{\beta} + r(x)^\top R^{-1}(Y - F\hat{\beta}), \tag{A6}$$

with $F = \left[\hat{y}_{lf}(x_1), \ldots, \hat{y}_{lf}(x_n)\right]^T$.

References

1. Keane, A.J. Wing Optimization Using Design of Experiment, Response Surface, and Data Fusion Methods. *J. Aircr.* **2003**, *40*, 741–750. [CrossRef]
2. Han, Z.H.; Görtz, S.; Zimmermann, R. Improving variable-fidelity surrogate modeling via gradient-enhanced kriging and a generalized hybrid bridge function. *Aerosp. Sci. Technol.* **2013**, *25*, 177–189. [CrossRef]
3. Forrester, A.I.J.; Sóbester, A.; Keane, A.J. *Engineering Design via Surrogate Modelling: A Practical Guide*; Wiley: Hoboken, NJ, USA, 2008. [CrossRef]
4. Kennedy, M.C.; O'Hagan, A. Predicting the output from a complex computer code when fast approximations are available. *Biometrika* **2000**, *87*, 1–13. [CrossRef]
5. Han, Z.H.; Zimmermann, R.; Görtz, S. A New Cokriging Method for Variable-Fidelity Surrogate Modeling of Aerodynamic Data. In Proceedings of the 48th AIAA Aerospace Sciences Meetings, Reston, VA, USA, 4–7 January 2010. [CrossRef]
6. Han, Z.H.; Zimmermann; Görtz, S. Alternative Cokriging Method for Variable-Fidelity Surrogate Modeling. *AIAA J.* **2012**, *50*, 1205–1210. [CrossRef]
7. Bertram, A.; Zimmermann, R. Theoretical investigations of the new Cokriging method for variable-fidelity surrogate modeling. *Adv. Comput. Math.* **2018**, *44*, 1693–1716. [CrossRef]
8. Han, Z.H.; Görtz, S. Hierarchical Kriging Model for Variable-Fidelity Surrogate Modeling. *AIAA J.* **2012**, *50*, 1885–1896. [CrossRef]
9. Meng, X.; Karniadakis, G.E. A composite neural network that learns from multi-fidelity data: Application to function approximation and inverse PDE problems. *J. Comput. Phys.* **2020**, *401*, 109020. [CrossRef]
10. Zhang, X.; Xie, F.; Ji, T.; Zhu, Z.; Zheng, Y. Multi-fidelity deep neural network surrogate model for aerodynamic shape optimization. *Comput. Methods Appl. Mech. Eng.* **2021**, *373*, 113485. [CrossRef]
11. Owen, N.E.; Challenor, P.; Menon, P.P.; Bennani, S. Comparison of Surrogate-Based Uncertainty Quantification Methods for Computationally Expensive Simulators. *SIAM/ASA J. Uncertain. Quantif.* **2017**, *5*, 403–435. [CrossRef]
12. Forrester, A.I.; Keane, A.J. Recent advances in surrogate-based optimization. *Prog. Aerosp. Sci.* **2009**, *45*, 50–79. [CrossRef]
13. Garbo, A.; German, B.J. Performance assessment of a cross-validation sampling strategy with active surrogate model selection. *Struct. Multidiscip. Optim.* **2019**, *59*, 2257–2272. [CrossRef]
14. Garbo, A.; German, B.J. A model-independent adaptive sequential sampling technique based on response nonlinearity estimation. *Struct. Multidiscip. Optim.* **2020**, *61*, 1051–1069. [CrossRef]
15. Parekh, J.; Bekemeyer, P. A Surrogate-based Approach for a Comprehensive UQ Analysis in CFD. In Proceedings of the AIAA SciTech 2024, Orlando, FL, USA, 8–12 January 2024; AIAA 2024-0707.
16. Huang, D.; Allen, T.T.; Notz, W.I.; Miller, R.A. Sequential kriging optimization using multiple-fidelity evaluations. *Struct. Multidiscip. Optim.* **2006**, *32*, 369–382. [CrossRef]
17. Shu, L.; Jiang, P.; Wang, Y. A multi-fidelity Bayesian optimization approach based on the expected further improvement. *Struct. Multidiscip. Optim.* **2021**, *63*, 1709–1719. [CrossRef]
18. Di Fiore, F.; Mainini, L. NM-MF: Non-Myopic Multifidelity Framework for Constrained Multi-Regime Aerodynamic Optimization. *AIAA J.* **2023**, *61*, 1270–1280. [CrossRef]
19. Foumani, Z.Z.; Shishehbor, M.; Yousefpour, A.; Bostanabad, R. Multi-fidelity cost-aware Bayesian optimization. *Comput. Methods Appl. Mech. Eng.* **2023**, *407*, 115937. [CrossRef]
20. Jones, D.R. A Taxonomy of Global Optimization MethodsBased on Response Surfaces. *J. Glob. Optim.* **2001**, *21*, 345–383. [CrossRef]
21. Nielsen, F. On the Jensen–Shannon Symmetrization of Distances Relying on Abstract Means. *Entropy* **2019**, *21*, 485. [CrossRef]

22. Lam, R.; Poloczek, M.; Frazier, P.; Willcox, K.E. Advances in Bayesian Optimization with Applications in Aerospace Engineering. In Proceedings of the 2018 AIAA Non-Deterministic Approaches Conference, Kissimmee, FL, USA, 8–12 January 2018. [CrossRef]
23. Balandat, M.; Karrer, B.; Jiang, D.; Daulton, S.; Letham, B.; Wilson, A.G.; Bakshy, E. BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization. In *Proceedings of the Advances in Neural Information Processing Systems*; Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H., Eds.; Curran Associates, Inc.: New York, NY, USA, 2020; Volume 33, pp. 21524–21538.
24. Jones, D.R.; Schonlau, M.; Welch, W.J. Efficient Global Optimization of Expensive Black-Box Functions. *J. Glob. Optim.* **1998**, *13*, 455–492. [CrossRef]
25. Bussemaker, J.H.; Bartoli, N.; Lefebvre, T.; Ciampa, P.D.; Nagel, B. Effectiveness of Surrogate-Based Optimization Algorithms for System Architecture Optimization. In Proceedings of the AIAA AVIATION 2021 FORUM, Virtual Event, 2–6 August 2021. [CrossRef]
26. Priem, R.; Bartoli, N.; Diouane, Y.; Sgueglia, A. Upper trust bound feasibility criterion for mixed constrained Bayesian optimization with application to aircraft design. *Aerosp. Sci. Technol.* **2020**, *105*, 105980. [CrossRef]
27. Schonlau, M.; Welch, W.J.; Jones, D.R. Global versus Local Search in Constrained Optimization of Computer Models. *Lect.-Notes-Monogr. Ser.* **1998**, *34*, 11–25.
28. Jolliffe, I.H.; Stephenson, D. *Surrogate Modeling: A Practical Guide*; John Wiley & Sons: Hoboken, NJ, USA, 2012.
29. Bekemeyer, P.; Bertram, A.; Hines Chaves, D.A.; Dias Ribeiro, M.; Garbo, A.; Kiener, A.; Sabater, C.; Stradtner, M.; Wassing, S.; Widhalm, M.; et al. Data-Driven Aerodynamic Modeling Using the DLR SMARTy Toolbox. In Proceedings of the AIAA AVIATION 2022 Forum, Chicago, IL, USA, 27 June–1 July 2022; p. 3899.
30. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [CrossRef]
31. Drela, M. XFOIL: An analysis and design system for low Reynolds number airfoils. In Proceedings of the Low Reynolds Number Aerodynamics: Proceedings of the Conference Notre Dame, Indiana, USA, 5–7 June 1989; Springer: Berlin/Heidelberg, Germany, 1989; pp. 1–12.
32. XFOIL. Version 6.99. 2013. Available online: https://web.mit.edu/drela/Public/web/xfoil/ (accessed on 10 May 2024).
33. Schwamborn, D.; Gerhold, T.; Heinrich, R. The DLR TAU-Code: Recent Applications in Research and Industry. In Proceedings of the European Conference on Computational Fluid Dynamics, Egmond aan Zee, The Netherlands, 5–8 September 2006.
34. Bishop, C.M. *Pattern Recognition and Machine Learning*, Corrected at 8th Printing 2009 ed.; Information Science and Statistics; Springer: New York, NY, USA, 2009.
35. Sacks, J.; Welch, W.J.; Mitchell, T.J.; Wynn, H.P. Design and analysis of computer experiments. *Stat. Sci.* **1989**, *4*, 409–423. [CrossRef]
36. Bertram, A.; Bekemeyer, P.; Held, M. Fusing Distributed Aerodynamic Data Using Bayesian Gappy Proper Orthogonal Decomposition. In Proceedings of the AIAA AVIATION 2021 FORUM, Virtual Event, 2–6 August 2021; American Institute of Aeronautics and Astronautics: Reston, VA, USA, 2021. [CrossRef]