Adaptive Switching of Variable-Fidelity Models in Population-Based Optimization

Ali Mehmani, Souma Chowdhury, Weiyang Tong and Achille Messac

Abstract This article presents a novel model management technique to be implemented in population-based heuristic optimization. This technique adaptively selects different computational models (both physics-based models and surrogate models) to be used during optimization, with the overall objective to result in optimal designs with high fidelity function estimates at a reasonable computational expense. For example, in optimizing an aircraft wing to obtain maximum lift-to-drag ratio, one can use low fidelity models such as given by the vortex lattice method, or a high fidelity finite volume model, or a surrogate model that substitutes the high-fidelity model. The information from these models with different levels of fidelity is integrated into the heuristic optimization process using the new adaptive model switching (AMS) technique. The model switching technique replaces the current model with the next higher fidelity model, when a stochastic switching criterion is met at a given iteration during the optimization process. The switching criterion is based on whether the uncertainty associated with the current model output dominates the latest improvement of the relative fitness function, where both the model output uncertainty and the function improvement (across the population) are expressed as probability distributions. For practical implementation, a measure of critical probability is used to regulate the degree of error that will be allowed, i.e., the fraction of instances where the improvement will be allowed to be lower than the model error, without having to change the model. In the absence of this critical probability, model man-

A. Mehmani · S. Chowdhury (✉)
Department of Mechanical and Aerospace Engineering, Syracuse University,
Syracuse, NY 13244, USA
e-mail: amehmani@syr.edu

S. Chowdhury
e-mail: chowdhury@bagley.msstate.edu

W. Tong · A. Messac
Department of Aerospace Engineering, Mississippi State University,
Mississippi State, MS 39762, USA
e-mail: wtong@syr.edu

A. Messac
e-mail: messac@ae.msstate.edu

N.D. Lagaros and M. Papadrakakis (eds.), *Engineering and Applied Sciences Optimization*, Computational Methods in Applied Sciences 38,
DOI 10.1007/978-3-319-18320-6_10

agement might become too conservative, leading to premature model-switching and thus higher computing expense. The proposed AMS-based optimization is applied to two design problems through Particle Swarm Optimization, which are: (i) Airfoil design, and (ii) Cantilever composite beam design. The application case studies of AMS illustrated: (i) the computational advantage of this method over purely high fidelity model-based optimization, and (ii) the accuracy advantage of this method over purely low fidelity model-based optimization.

1 Introduction

Population-based heuristic optimization algorithms, such as evolutionary algorithms and swarm optimization algorithms have been applied to diverse areas of science and engineering over the past few decades. They have been proven to be very effective in solving complex design optimization problems, especially those involving highly nonlinear functions. However, considering the computational cost of the high fidelity simulation models typically used to represent system behavior (e.g., CFD, FEA models), the large number of function evaluations often demanded by heuristic algorithms limit their applicability to practical complex system design (e.g., wing design of a high speed civil transport aircraft [1]). One approach to address this issue is *variable fidelity optimization*. In this approach, model management strategies adaptively integrate models of different fidelity and cost into the optimization process.

1.1 Variable Fidelity Models

Variable fidelity models refer to models with different levels of fidelity, where the computational cost of the model is generally related to the accuracy of the model estimation. In addition to low, medium, and high fidelity physics-based models, *surrogate models* (or mathematical approximation models) can also be used as candidates within a set of variable fidelity models. Surrogate models are purely mathematical models (i.e., not derived from the system physics) that are used to provide a tractable and inexpensive approximation of the actual system behavior. They are commonly used as an alternative to expensive computational simulations (e.g., CFD [2]) or to the lack of a physical model in the case of experiment-derived data (e.g., creation and testing of new metallic alloys [3]). Further description of the state of the art in surrogate modeling can be found in the following literature [4–6]. Major surrogate modeling methods include Polynomial Response Surfaces [7], Kriging [8, 9], Moving Least Square [10, 11], Radial Basis Functions (RBF) [12], Support Vector Regression(SVR) [13], Neural Networks [14] and hybrid surrogate models [15]. These methods have been applied to a wide range of disciplines, from aerospace design and automotive design to chemistry and material science [6, 16, 17].

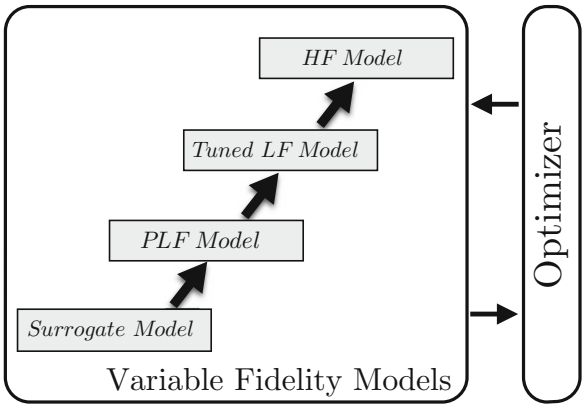

Fig. 1 Variable fidelity models

Besides direct implementation of a surrogate model as a black-box function (directly substituting a high fidelity model or data), low fidelity physic-based models can also be combined with a surrogate model to achieve a hybrid model of greater accuracy than its individual components (as illustrated in Fig. 1). Low fidelity physic-based models (e.g., the vortex lattice computational fluid dynamics method) are generally less complex than a high fidelity model and often provide a less faithful representation of the system behavior [18]. These models can be obtained by simplifying either the analysis model (e.g., using coarse finite element mesh) or the original physical formulation (e.g., using simplified boundary conditions or geometry). To their advantage, low fidelity physics-based models often inherit the major features of true models, while being significantly less expensive. Hence, these models could provide a reliable foundation for the construction of high-quality hybrid approximation models. These hybrid models, also called *tuned low fidelity models*, are expected to reflect the most prominent physical features of the system behavior, while preserving computational efficiency. Two well-known approaches for constructing tuned low fidelity (TLF) models are *multiplicative* and *additive* approaches, as given in Eqs. 1 and 2, respectively [19].

$$\text{Multiplicative approach:} \quad y_{TLF} = A \times y_{LF} \tag{1}$$

$$\text{Additive approach:} \quad y_{TLF} = B + y_{LF} \tag{2}$$

In both of these approaches, the tuning functions (A and B) are trained using the associated values of the high and low fidelity models for a given DoE, as shown below:

$$
\begin{aligned}
&A(X) = \frac{y_{HF}(X)}{y_{LF}(X)} \\
&B(X) = y_{HF}(X) - y_{LF}(X) \\
&\text{where } X = \{X_1, X_2, X_3, \ldots, X_{N_s}\} \\
&N_S\text{: Number of sample points}
\end{aligned} \tag{3}
$$

and where $y_{HF}(.)$ and $y_{LF}(.)$ respectively represent the functional responses of the low and the high fidelity models (where in the multiplicative scenario, y_{LF} is only allowed to take non-zero values). In surrogate-based tuned low fidelity models, the tuning (or correction) of a low fidelity model is performed using a surrogate model constructed through a DoE of the high fidelity model [20–22].

1.2 Model Management in Optimization

The major pitfall in using low fidelity models in optimization is that they can often mislead the search process, leading to suboptimal or infeasible solutions. To address this issue and provide optimum designs with high fidelity system evaluations, model management strategies can be applied. Different model management approaches have been reported in the literature, for integrating low fidelity models within optimization algorithms. One class of model management strategies are developed based on the *Trust-Region* methods [23–27]. The basic idea of the Trust-region is to solve an optimization problem, $\underset{x \in \mathbb{R}^P}{\text{Min}}\ f(x)$, using the high fidelity model ($f(x)$). In solving this optimization problem using a gradient-based algorithm, the kth iteration is computed as $x^{k+1} = x^k + \lambda\ \Delta x$, where λ is the step length and Δx is the decent direction. As Δx is fixed, the problem reduces to a one-dimensional optimization problem: $\underset{\lambda}{\text{Min}}\ f(x^k + \lambda \Delta x)$. To improve the computational efficiency of the problem, the low fidelity model, $\hat{f}(x)$, can be used in the latter optimization problem. Assuming the low fidelity model is only valid in the vicinity of x^k (e.g., $x^k + \gamma$), the optimization search for λ is changed to the following constrained optimization problem:

$$
\underset{\lambda}{\text{Min}}\ f(x + \lambda \Delta x), \quad \text{subject to: } ||\lambda \Delta x|| < \gamma \tag{4}
$$

where γ is the *trust-region radius*. In the *Trust-Region* based model management methods developed by Alexandrov et al. [28] and by Toropove and Alvarez in 1998 [29], the parameter γ is adaptively increased (or decreased) depending on how well the low fidelity model, $\hat{f}(x)$, predicts the improvement in the high fidelity model. This criterion is estimated by computing the ratio of the actual to the predicted improvement in the objective function, as given by

$$
\frac{f(x^k) - f(x^k + \lambda^k \Delta x^k)}{\hat{f}(x^k) - \hat{f}(x^k + \lambda^k \Delta x^k)} \tag{5}
$$

The *Trust-Region* method seeks the agreement of the function and its gradient values estimated by the low fidelity model with those estimated by the high fidelity model. However, these techniques may not be directly applicable in problems where gradients are expensive to evaluate, or where zero-order algorithms are being used for optimization.

In another class of model management strategies, developed for non-physics-based low fidelity models (e.g., surrogate model and tuned-low fidelity model) the accuracy of the surrogate model (or metamodels) is improved during the optimization process by adding infill points, where additional evaluations of the high fidelity model is then performed. Infill points are generally added in (i) the region where the optimum is located (local exploitation); and/or (ii) the entire design space to improve the global accuracy of the surrogate (global exploration) [20, 30, 31]. Trosset and Torczon in 1997 [32] proposed an approach where the balance between exploitation and exploration was considered using the aggregate *merit function*, $\hat{f}(x) - \rho d_{min}(x)$, where, $d_{min}(x) = \underset{x}{\text{Min}} \, \|x - x^i\|, \ \rho > 0$. It is important to note that, this technique is independent of the type of surrogate modeling technique being considered. Over the last two decades, different statistical model management strategies have been developed [33–36]. Among them, Jones et al. in 1998 [35] developed a well-known model management strategy that is based on an *Expected Improvement (EI)* criterion, and is called *Efficient Global Optimization (EGO)*. This powerful approach is however generally limited to surrogate models based on Gaussian processes. Assuming f_{min} is the objective function value of the optimum in the training data, the expected improvement in an infill point x is given by $\mathbb{E}(I(x)) = \mathbb{E}(max(f_{min} - F(x)), 0)$. In this case, $F(x)$ is a Gaussian distribution, $F(x) \sim \mathcal{N}(\hat{f}(x), \sigma^2(x))$, where the posterior mean, $\hat{f}(x)$, is used as a surrogate model, and the posterior variance $\sigma^2(x)$ gives an estimate of the uncertainty involved in the surrogate prediction. The expected improvement can be estimated by

$$\mathbb{E}(I(x)) = (f_{min} - \hat{f}(x))\Phi\left(\frac{f_{min} - \hat{f}(x)}{\sigma(x)}\right) + \sigma(x)\phi\left(\frac{f_{min} - \hat{f}(x)}{\sigma(x)}\right) \tag{6}$$

where $\Phi(.)$ and $\phi(.)$ denote the standard normal density and distribution functions, respectively [34]. Subsequently, an infill point can be found by maximizing the expected improvement, $X^{infill} = \arg\max_x (\mathbb{E}(I(x)))$.

The model management strategies used in heuristic optimization algorithms can be broadly classified into two different approaches which are (i) *individual-based evolution control*, and (ii) *generation-based evolution control* [37]. In the individual-based approach, selected individuals (*controlled individuals*) within a generation are evaluated using a high fidelity model. In the generation-based approach, the whole population at a certain generation (*controlled generation*) is evaluated using the high fidelity model. Graning et al. [38] explored different individual-based evolution frameworks such as (i) the Best Strategy [39], where the best individuals at each generation are selected as controlled individuals, (ii) the Pre-Selection method [40], where the offspring of the best individuals are selected as controlled individuals,

and (iii) the Clustering Technique [41], where the *k-means* clustering technique is used to find the "controlled individual cluster" based on the distance from the best individual.

In this section, a survey of existing model management strategies for integrating models with different levels of fidelity into an optimization process was provided. Several of the existing strategies are found to be defined for specific types of low fidelity model, e.g., EGO works primarily for Gaussian process-based surrogate models. On the other hand, existing techniques generally consider the combination of only two models of different fidelities (e.g., Trust-region methods, and individual- and generation-based techniques). This article seeks to address some of the above-stated crucial gaps in the variable-fidelity optimization paradigm. Specifically, the development of a model management strategy that can be coherently applied to different types of low fidelity models (i.e., physics-based and non-physics-based low fidelity models), and allows adaptive switching between more than two models is being pursued in this article.

1.3 A New Approach to Global Model Switching

The primary objective of this article is to investigate a new adaptive model management strategy that significantly reduces the computational cost of optimization while converging to the optimum with high fidelity model evaluation; in its current form, this method is designed to work with population-based optimization algorithms (e.g., GAs, PSOs). Additionally, this method assumes that models of different levels of fidelity are available to the user. Specifically, a new stochastic model switching metric, called Adaptive Model Switching (AMS), is formulated in this article. The AMS technique is implemented through a powerful version of the Particle Swarm Optimization (PSO) algorithm that involves explicit diversity preservation, called *Mixed-Discrete PSO* [42]. The effectiveness of this implementation is investigated by application to two engineering design optimization problems.

The remainder of the article is organized as follows: Sect. 2 presents the concept and the formulation of the new Adaptive Model Switching (AMS) metrics. Description of the model error quantification methods used in this article, including Predictive Error Estimation of Model Fidelity (PEMF), is provided in Sect. 2.3. Section 3 describes the practical problems to which AMS is applied; the numerical settings and case study results are illustrated and discussed in Sect. 3. Section 4 provides the concluding remarks.

2 Variable Fidelity Optimization with Adaptive Model Switching (AMS)

2.1 Major Steps in Optimization with AMS

In optimization based on variable fidelity models, the important question is when and where to integrate the models with different levels of fidelity. In this article, the "when to integrate" question is particularly addressed. Increasing fidelity too early in the design process can be computationally expensive while wasting resources to explore undesirable regions of the design domain. On the other hand, switching to a higher fidelity model too late might mislead the search process early on to suboptimal regions of the design domain (especially in multimodal problems), i.e., leading to scenarios where the global optimum is outside of the region spanned by the population of candidate solutions in later iterations. In this section, a novel model management strategy called, *Adaptive Model Switching (AMS)* metric is developed to avoid both these undesirable scenarios. AMS can be perceived as a decision-making tool for the timing of *model-switching* or *model integration*. The implementation of the proposed AMS in population-based algorithm involves the following five major steps:

Step 1 Assuming the available models are non-dominated w.r.t. each other in terms of fidelity and computational expense, the models are first ranked from the lowest fidelity to the high fidelity, based on the error associated with each model-M_i for $i = 1, \ldots, n$. where model M_1 has the lowest fidelity and model M_n has the highest fidelity. Assuming the distribution of model error is known for each model, the ranking is performed using the modal values of the error distributions.

Step 2 The initial population is then generated at $t = 1$, using M_1.

Step 3 At every iteration (t) of the heuristic optimization algorithm, the current model, M_i, is used to update the function values of the population, and then set $t = t + 1$. In this article, Particle Swarm Optimization is the chosen heuristic optimization algorithm.

Step 4 The following stopping criteria is checked after every iteration.
The optimization algorithm stops when the relative changes in the fitness function value is less than a predefined function tolerance, δ_F. To avoid termination before reaching the high fidelity model (M_n), the function tolerance must be specified to be less than the modal error of the last but one model (M_{n-1}).
IF the termination criteria is satisfied, the current optimum (the best global solution in the case of PSO) is identified as the final optimum and the optimization process is terminated.
ELSE, **Go To Step 5**

Step 5 The switching metric (AMS metric) is evaluated in this step.
IF the AMS metric is satisfied, a switching event occurs, and the algorithm

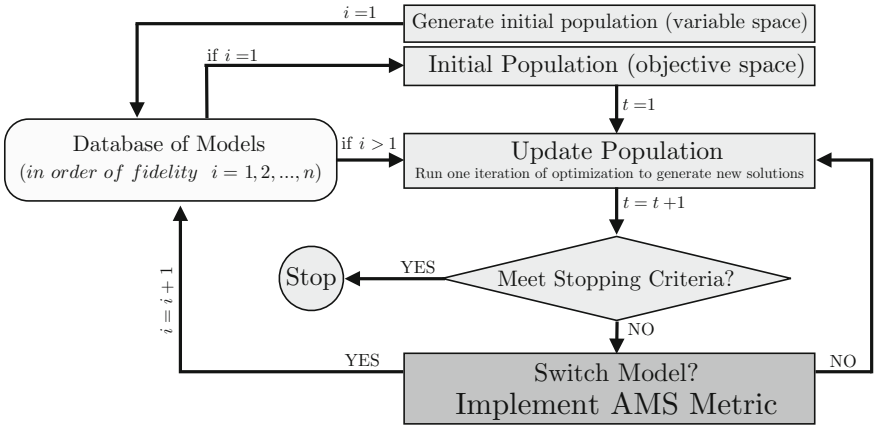

Fig. 2 Adaptive model switching in population-based optimization

switches from model M_i to M_{i+1}.
Go To Step 3

A flowchart of the algorithm for optimization with AMS is shown in Fig. 2. In practice, the AMS technique (Step 5) need not be applied at every iteration; the user can specify it to be applied after every K iteration (where K *is a small positive integer*). In the flowchart, AMS is shown to be applied at every iteration, for the sake of simplicity.

In the following subsection, the novel components of the AMS method (Fig. 2) are described. Subsequently, an overview of the Mixed-Discrete PSO algorithm, which is used for implementing and testing the AMS method, is provided.

2.2 The Adaptive Model Switching (AMS) Metric

In this article, it is assumed that the uncertainty associated with each model (M_i; $i = 1, \ldots, n$) is known or can be evaluated in the form of an error distribution, $\mathbb{P}_i$. Under this assumption, the fitness function values evaluated using the ith model can be related to the corresponding high fidelity estimation as

$$y^i_{HF} = \widehat{y}^i_{LF} + \varepsilon^i \tag{7}$$

In Eq. 7, $\widehat{y}^i_{LF}$ and ε^i respectively represent the response of the ith low fidelity model and the stochastic error associated with it; and y^i_{HF} is the corresponding high fidelity model response. The relative improvement in the fitness function value (Δf) can be considered to follow an unknown distribution, Θ, over the population of solutions. Here, Δf in the tth iteration ($t \geq 2$) can be expressed as

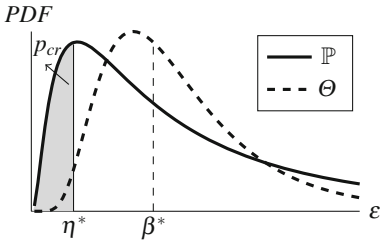

Fig. 3 The illustration of the AMS metric

$$\Delta {f_k}^t = \begin{cases} |\dfrac{{f_k}^t - {f_k}^{t-1}}{{f_k}^t}| & \text{if } {f_k}^t \neq 0 \\ |{f_k}^t - {f_k}^{t-1}| & \text{if } {f_k}^t = 0 \end{cases} \quad \text{where } k = 1, 2, 3, \ldots, N_{pop} \tag{8}$$

The model switching criteria is then defined based on "whether the uncertainty associated with a model response is higher than the observed improvement in the relative fitness function of the population". Due to the practical unavailability of reliable local measures of model error (i.e., ε as a function of x), the model switching criteria is designed using the stochastic global measures of model error and the distribution of solution improvement. Based on prior experience or practical design requirements, the designer is likely to be cognizant of what levels of global model error, η, is acceptable for a particular low fidelity model in an optimization process. Hence, η can be perceived as a user-preference. The *critical probability,* p_{cr} for that low fidelity model with an error distribution $\mathbb{P}$ is then defined as the probability of the model error to be less than η. This definition can be expressed as

$$p_{cr} = Pr[\varepsilon \leq \eta] = \int_0^{\eta} \mathbb{P}(\varepsilon^{'})\, d\varepsilon^{'} \tag{9}$$

The critical probability (p_{cr}) essentially indicates a critical bound in the error distribution $\mathbb{P}$ ($0 \leq \varepsilon \leq \eta$). If the predefined cut-off value (β) of the Θ distribution lies inside this region, the current low fidelity model is considered to be no more reliable for use in the optimization process. As illustrated in Fig. 3, assuming that Θ and $\mathbb{P}$ follow a log-normal distribution, $p_{cr} = Pr[\varepsilon \leq \eta^*]$; and β^* is the pre-computed cut-off value in the Θ distribution. The model with the $\mathbb{P}$ error distribution can be used in the optimization process provided that $\eta^* \leq \beta^*$.

The Adaptive Model Switching (AMS) metric is formulated as a hypothesis testing that is defined by a comparison between

(I) the distribution of the relative fitness function improvement (Θ) over the entire population, and
(II) the distribution of the error associated with the ith model ($\mathbb{P}_i$) over the entire design space.

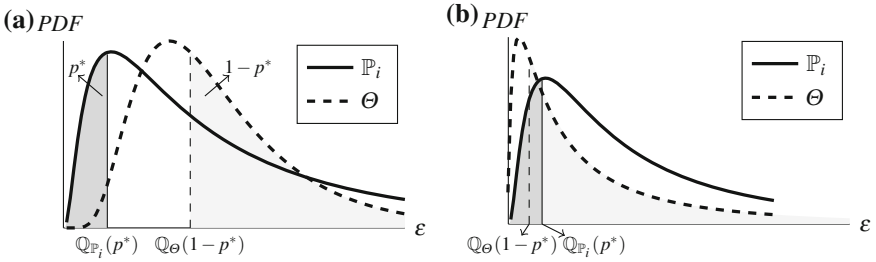

Fig. 4 The illustration of the AMS hypothesis test (comparing the model error distribution ($\mathbb{P}_i$) and the distribution of fitness function improvement (Θ)); **a** Rejection of the text; **don't change a model**. **b** Acceptance of the text; **change a model**

This statistical test for the ith model can be stated as

$$\begin{aligned} &H_0\colon \mathbb{Q}_{\mathbb{P}_i}(p_{cr}) \geq \mathbb{Q}_{\Theta}(1-p_{cr}) \\ &H_1\colon \mathbb{Q}_{\mathbb{P}_i}(p_{cr}) < \mathbb{Q}_{\Theta}(1-p_{cr}) \\ &0 < p_{cr} < 1 \end{aligned} \tag{10}$$

where $\mathbb{Q}$ represents a quantile function of a distribution; The p-quantile, for a given distribution function, Ψ, is defined as

$$\mathbb{Q}_{\Psi}(p) = inf\{x \in \mathbb{R}\colon p \leq \Psi_{(c.d.f.)}(x)\} \tag{11}$$

In Eq. 10, p_{cr} or the critical probability is an *Indicator of Conservativeness (IoC)*. The IoC is based on user preferences, and regulates the trade-off between optimal solution reliability and computational cost in the AMS-based optimization process. Generally, the higher the IoC (closer to 1), the higher the solution reliability and the greater the computational cost; under these conditions, model switching events will occur early on in the optimization process.

For the sake of illustration, assume Θ and $\mathbb{P}_i$ follow a log-normal distribution, and $p_{cr} = p^*$. In this case, the null hypothesis will be rejected, and the optimization process will use the current model (M_i) **if** $\mathbb{Q}_{\Theta} > \mathbb{Q}_{\mathbb{P}_i}$, as illustrated in Fig. 4a. Conversely, **if** $\mathbb{Q}_{\Theta} < \mathbb{Q}_{\mathbb{P}_i}$, the null hypothesis will be accepted, and the optimization process will switch to the next higher fidelity model (M_{i+1}), as shown in Fig. 4b.

In this article, Kernel Density Estimation (KDE) is adopted to model the distribution of the relative improvement in the fitness function over consecutive kt iterations. Since the distribution of fitness function improvement over the population (for different problems) may not follow any particular probability model, and is also observed to be multimodal at times, KDE is a suitable choice in this context. KDE is a standard non-parametric approach to estimate the probability density function of random variables. Here, it is assumed that $\Delta f = (\Delta f_1,\ \Delta f_2,\ \Delta f_3, \ldots,\ \Delta f_{N_{pop}})$ is an independent and identically distributed sample drawn from a distribution with an

unknown density $\Theta_{\Delta f}$. The kernel density estimator can then be used to determine $\Theta_{\Delta f}$, as given by

$$\tilde{\Theta}_{\Delta f}(x; H) = N_{pop}^{-1} \sum_{i=1}^{N_{pop}} K_H(x - x_i) \tag{12}$$

Here, the kernel $K(x)$ is a symmetric probability density function, H is the bandwidth matrix which is symmetric and positive-definite, and $K_H(x) = |H|^{-1/2} K(H^{-1/2}x)$. The choice of K is not as crucial as the choice of the H estimator for the accuracy of the KDE [43]. In this article, we consider $K(x) = (2\pi)^{-d/2} exp(\frac{-1}{2} x^T x)$, the standard normal throughout. The Mean Integrated Squared Error (MISE) method is used as a criterion for selecting the bandwidth matrix, H [44], where

$$MISE(H) = \mathbb{E}(\int [\tilde{\Theta}_{\Delta f}(x; H) - \Theta_{\Delta f}(x)]^2) \tag{13}$$

2.3 Quantifying Model Uncertainties

In this article, the uncertainties associated with surrogate models and surrogate-based tuned low fidelity models are determined using an advanced surrogate error estimation method, called *Predictive Estimation of Model Fidelity* or *PEMF* [45]. The PEMF method is derived from the hypothesis that "the accuracy of approximation models is related to the amount of data resources leveraged to train the model". A brief description of the PEMF method is provided in the following sub-section (Sect. 2.3.1). In the case of physics-based low fidelity (PLF) models, the uncertainty in their output is quantified through an inverse assessment process, by comparing the physics-based low fidelity model responses with the high fidelity model responses. In this case, the relative absolute error (RAE_{PLF}) of a PLF model is estimated as

$$RAE_{PLF_i} = \begin{cases} |\dfrac{HF_i - PLF_i}{HF_i}| & \text{if } HF_i \neq 0 \\ |HF_i - PLF_i| & \text{if } HF_i = 0 \end{cases} \tag{14}$$
$$\text{where } i = 1, 2, 3, \ldots, N_s \text{ (Number of sample points)}$$

A DoE of N_S high fidelity evaluations is used to perform the above-stated error quantification, and also to train a surrogate models and a tuned low fidelity models. The uncertainty of a low fidelity physics-based models is represented by a log-normal distribution, $\ln \mathcal{N}(\mu_{PLF}, \sigma_{PLF})$, where the p-quantile of this distribution is defined as

$$\mathbb{Q}_{\mathbb{P}_{PLF}}(p) = \mathbb{P}^{-1}_{PLF}(p|\mu_{PLF}, \sigma_{PLF}) = exp(\mu_{PLF} + \Phi^{-1}(p)\,\sigma_{PLF})$$

where

$$\mu_{PLF} = \ln\left(\frac{m^2_{RAR_{PLF}}}{\sqrt{v^2_{RAR_{PLF}} + m^2_{RAR_{PLF}}}}\right) \quad \sigma_{PLF} = \sqrt{\ln\left(1 + \frac{v_{RAR_{PLF}}}{m^2_{RAR_{PLF}}}\right)} \tag{15}$$

In Eq. 15, $\Phi^{-1}(.)$ is the inverse of the c.d.f of the standard normal distribution with zero mean and unit variance, and $m_{RAR_{PLF}}$ and $v_{RAR_{PLF}}$ are the mean and the variance of $RAE_{i=1,2,3,\ldots,N_s}$, respectively.

2.3.1 Predictive Estimation of Model Fidelity (PEMF)

In concept, the PEMF method [45] can be perceived as a novel sequential implementation of k-fold cross-validation, with carefully constructed error measures that are significantly less sensitive to outliers and the DoE (compared to Mean or Root Mean Square error measures). The PEMF method predicts the error by capturing the variation of the surrogate model error with an increasing density of training points (without investing any additional test points).

In the PEMF method, for a set of N_s sample points, intermediate surrogates are constructed at each iteration, r, using S^r heuristic subsets of n^r training points (called intermediate training points), where $n^r < N_s$. These intermediate surrogates are then tested over the corresponding remaining $N_s - n^r$ points (called intermediate test points). The median error is then estimated for each of the S^r intermediate surrogates at that iteration, and a parametric probability distribution is fitted to yield the modal value, $E^{mo,r}_{med}$, and the median value, $E^{med,r}_{med}$, of the model error at that stage. The smart use of the modal value of the median error significantly reduces the occurrence of oscillations in the variation of error with sample density, unlike mean or root mean squared error which are highly susceptible to outliers [46]. This approach gives PEMF an important advantage over conventional cross-validation-based error measures, as illustrated by Mehmani et al. [45–47]. It is important to note that all error quantifications are performed in terms of the relative absolute error (E_{RAE}), which is given by:

$$E_{RAE}(X_i) = \begin{cases} \left|\dfrac{F(X_i) - \hat{F}(X_i)}{F(X_i)}\right| & \text{if } F(X_i) \neq 0 \\ |F(X_i) - \hat{F}(X_i)| & \text{if } F(X_i) = 0 \end{cases} \tag{16}$$

where F is the actual function value at X_i, given by high fidelity model, and $\hat{F}$ is the function value estimated by the surrogate model.

In order to control the computational expense of PEMF, the lognormal distribution is used to represent the surrogate model error; this distribution has been previously observed (from numerical experiments) to be one of the most effective choice in

representing the surrogate model error distribution. The PDFs of the median error, p_{med}, can thus be expressed as

$$p_{med} = \frac{1}{E_{med}\sigma_{med}\sqrt{2\pi}} \exp\left(\frac{(\ln(E_{med} - \mu_{med}))^2}{2\sigma_{med}^2}\right) \tag{17}$$

In the above equation, E_{med} represents the median of the relative errors estimated over a heuristic subset of training points at any given iteration in PEMF. The parameters, $(\mu_{med}, \sigma_{med})$ are the generic parameters of the log-normal distribution. The modal and median values of the median error distribution at any iteration, r, can then be expressed as

$$\begin{aligned} E_{med}^{mo}|_r &= \exp(\mu_{med} - \sigma_{med}^2)|_r \\ E_{med}^{med}|_r &= \exp(\mu_{med})|_r \end{aligned} \tag{18}$$

Once the history of modal and median errors at different sample size ($<N_s$) are estimated, the variation of the modal and median values of the errors with sample density are then modeled using the multiplicative ($E = a_0 n^{a_1}$) or the exponential ($E = a_0 e^{a_1 n}$) regression functions (depending on the best least-square fit). These regression functions are then used to predict the modal and the median values of the error distribution in the final surrogate, where the final surrogate is trained using all the N_s sample points. The predicted modal and the median error values, ε_{mod} and ε_{med}, are then used to define the distribution of the error in the final surrogate model, or in other words the response uncertainty of the surrogate model. The location and scale parameters of the error distribution is then given by

$$\begin{aligned} \mu_{\varepsilon} &= \ln \varepsilon_{med} \\ \sigma_{\varepsilon} &= \sqrt{\ln\left(\frac{\varepsilon_{med}}{\varepsilon_{mod}}\right)} \end{aligned} \tag{19}$$

Subsequently, the p-quantile of the error distribution associated with the surrogate model is given by

$$\mathbb{Q}_{\mathbb{P}_{SM}}(p) = \mathbb{P}_{SM}^{-1}(p|\mu_{\varepsilon}, \sigma_{\varepsilon}) = exp(\mu_{\varepsilon} + \Phi^{-1}(p)\,\sigma_{\varepsilon}) \tag{20}$$

2.4 Optimization Algorithm: Particle Swarm Optimization

In the proposed model management methodology, optimization is performed using an advanced implementation of the Particle Swarm Optimization (PSO). PSO was originally developed for solving continuous nonlinear optimization problems by Eberhart and Kennedy in 1995 [48]. Several advanced versions of this algorithm have been reported in the literature since its inception. In this article, one particular advanced

implementation of the PSO algorithm called Mixed-Discrete PSO (MDPSO), which was developed by Chowdhury et al. [42], is used. The advantages that the MDPSO algorithm provides over a conventional PSO algorithm include: (i) an ability to deal with both discrete and continuous design variables, and (ii) an explicit diversity preservation capability that mitigates the possibility of premature stagnation of particles. Further description of the MDPSO algorithm can be found in the paper by Chowdhury et al. [42].

3 Numerical Case Studies

3.1 Aerodynamic Shape Optimization of 2D Airfoil

This section describes a 2D airfoil design problem where the ratio of the coefficients of lift and drag (C_L/C_D) of the Wortmann FX60.126 2D airfoil [49] is to be maximized. The lift-to-drag ratio (C_L/C_D) is expressed as a function of four design variables, which include the angle of incidence (ranging from 0 to 10) and the three normalized shape variables (each ranging from -0.01 to 0.01). As illustrated in Fig. 5, the three shape variables define the distances (i) between the middle of the suction side and the horizontal axis (x_1), (ii) between the middle of pressure side and the horizontal axis (x_2), and (iii) between the trailing edge and the horizontal axis (x_3). These three shape variables allow a modification of the un-deformed airfoil profile. With respect to the initial airfoil design, two cubic splines are added to the suction and the pressure sides. Each of these splines is characterized by 3 points, defined on the leading edge, the middle span, and the trailing edge. The chord length of the airfoil is equal to 1 m. The design constraints are the side constraints on the design variables which are listed in Table 1.

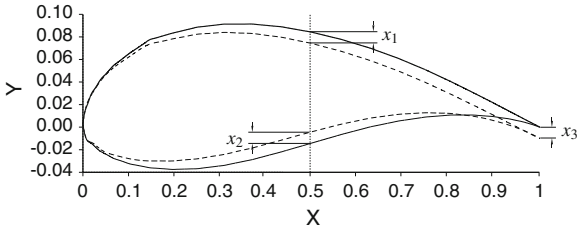

Fig. 5 Design variables governing the geometry of the airfoil

Table 1 Design variables in airfoil optimization problem

Description	Notation	Lower limit	Upper limit
Distance between the middle of suction side and horizontal axis	x_1	−0.01	0.01
Distance between the middle of pressure side and horizontal axis	x_2	−0.01	0.01
Distance between the trailing edge and horizontal axis	x_3	−0.01	0.01
Incidence angle	x_4	0°	10°

3.1.1 Aerodynamic Models with Different Level of Fidelity

To develop a high fidelity aerodynamic model for determining C_L and C_D ($\mathrm{M}^{\mathrm{A}}_{\mathrm{HF}}$), the commercial Finite Volume Method package, FLUENT, is used. The Reynolds-averaged Navier-Stokes (RANS) formulation is used along with a Reynolds model to represent the turbulence. The CFD mesh is constructed using quadrangular cells [49], characterized by 9,838 quadrangular cells and 10,322 grid points (Fig. 6a).

The low fidelity physics-based model ($\mathrm{M}^{\mathrm{A}}_{\mathrm{PLF}}$) is constructed based on the assumptions that the fluid is steady, incompressible, and irrotational. In this model, the Navier-Stokes equations are solved using the Finite Element method. Triangular T3 elements are used for demonstration, as shown in Fig. 6b. The incoming velocity in the analysis is set to 25 m/s. The computational time of the High and Low fidelity physics-based models are approximately 300 and 30 s, respectively (i.e., an order of magnitude apart). The pressure field around the airfoil for the low and high fidelity aerodynamic models at a baseline design ($x_1 = 0,\ x_2 = 0,\ x_3 = 0,$ and $\mathrm{x}_4 = 5°$) are illustrated in Fig. 7.

The third model is a surrogate model ($\mathrm{M}^{\mathrm{A}}_{\mathrm{SM}}$) constructed using a DoE of high fidelity evaluation involving 30 sample points. The fourth model is a tuned low

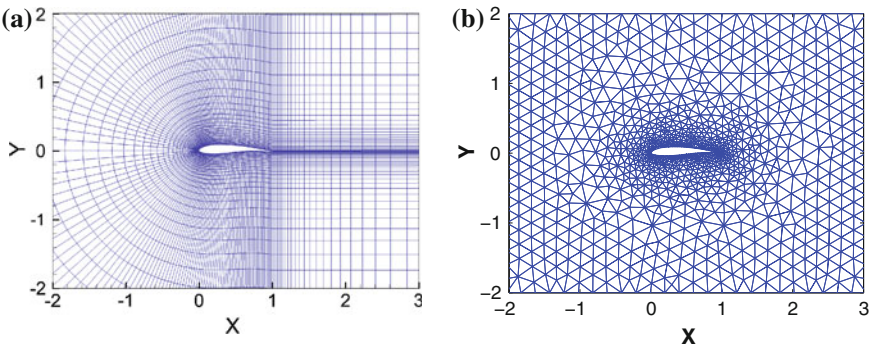

Fig. 6 Fine and coarse mesh for CFD of airfoil [49]; **a** High fidelity model mesh; **b** Low fidelity model mesh

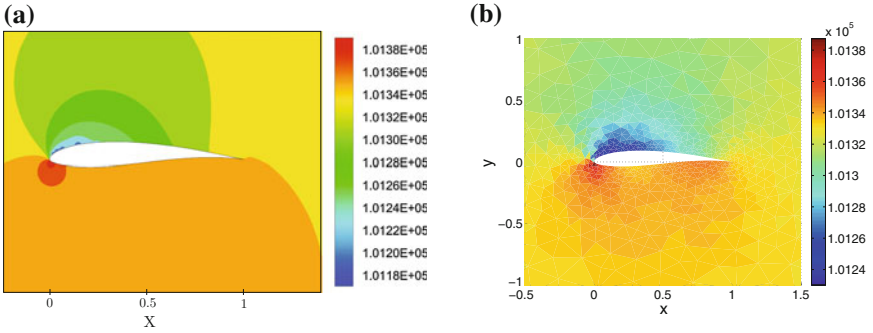

Fig. 7 Pressure field around the airfoil at a baseline design; **a** High fidelity model, **b** Low fidelity physics-based model

fidelity model (M^{A}_{TLF}). In this article, the tuned low fidelity model is constructed using the *Multiplicative approach*, as given by

$$\tilde{F}(x, a) = f(x) \times C(x) \tag{21}$$

where $\tilde{F}$ is a tuned low fidelity model; $f(x)$ is a low fidelity model; $C(x)$ is an explicit tuning surrogate constructed using the high fidelity samples, as shown below:

$$C(x) = \frac{\frac{C_L}{C_D}|_{\text{HF}}}{\frac{C_L}{C_D}|_{\text{PLF}}} \tag{22}$$

where C_L and C_D are respectively the lift and drag coefficients.

The surrogate model (M^{A}_{SM}) and the surrogate component of the tuned low fidelity model (M^{A}_{TLF}) are both constructed using Kriging with a Gaussian correlation function [8, 9]. Kriging is an interpolating method that is widely used for representing irregular data. Under the Kriging approach, the zero-order polynomial function is used as a regression model. In this article the Optimal Latin Hypercube is adopted to determine the locations of the sample points. The PEMF method is then applied to estimate the error in the surrogate models constructed using the high fidelity responses, and the tuned low fidelity model. To estimate the error in the physics-based low fidelity FEA model, the inverse assessment process defined in Sect. 2.3, is applied. Figure 8a–c illustrate the distributions of the error in the tuned low fidelity model, the surrogate model, and the physics-based low fidelity model. It is observed from Fig. 8 that the accuracy of the physics-based low fidelity model is less than that of the surrogate model. It is also readily evident that the computational cost of the physics-based low fidelity model is more than that of the surrogate model. Therefore, in this problem, the physics-based low fidelity model is dominated by the other three models and is hence not included as a model choice in the variable fidelity optimization.

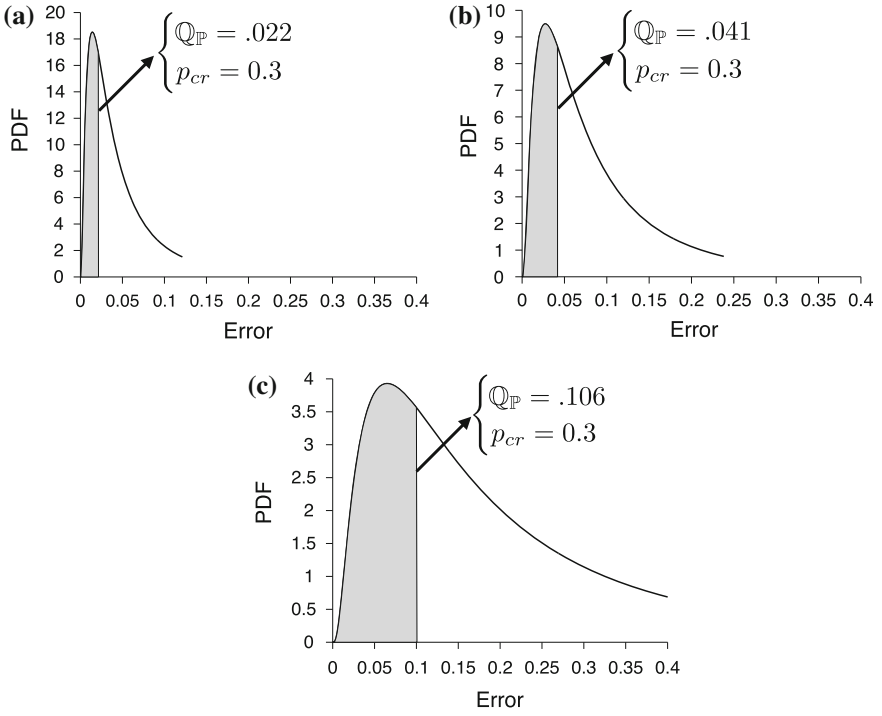

Fig. 8 Distribution of the model errors in evaluating the aerodynamic C_L/C_D ratio of the 2D airfoil: **a** Tuned LF model, **b** Surrogate model, **c** Physics-based LF model

3.1.2 Airfoil Optimization Problem: Results and Discussion

In the airfoil optimization problem, the initial population of particles is generated using the fastest model, which is the surrogate model. The AMS technique adaptively switches the model type twice during optimization (over a total of 22 iterations), resulting in an optimum design with a high fidelity function estimate.

The model types, the error distribution parameters associated with each model, and the number of calls made to each model in this optimization are listed in Table 2.

Table 2 Models with different levels of fidelity used in the airfoil optimization problem (the high fidelity model is assumed to be a true representation of the system behavior)

Model	Location parameter μ	Scale parameter σ	$\mathbb{Q}(p_{cr})$ $p_{cr} = 0.3$	No. of calls made $N_{pop} \times$ # Iter.
Surrogate	−2.6793	0.9628	0.0414	30 × 13
Tuned LF	−3.3197	0.9547	0.0219	30 × 6
High fidelity	–	–	–	30 × 3

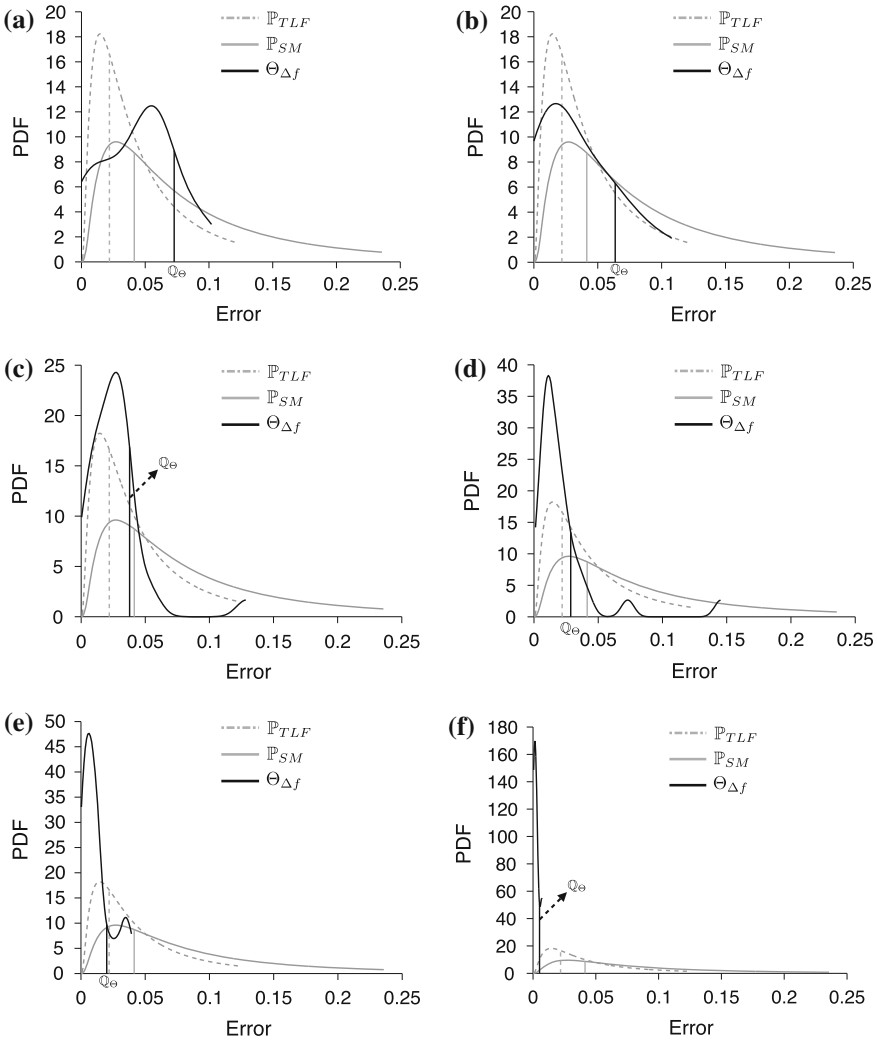

Fig. 9 Distribution of the fitness function improvements in different iterations of the airfoil optimization with PSO-AMS (also showing the model error distributions); **a** 5th iteration, **b** 10th iteration, **c** 15th iteration, **d** 18th iteration, **e** 20th iteration, **f** 22th iteration

The total number of calls made to each model is equal to the product of the particle population and the number of iterations during which that particular model is used for system evaluation. In this problem, the AMS technique is applied at every iteration.

Figure 9a–f illustrate the distribution of the fitness function improvement at different iterations during the optimization process. In these figures, $\mathbb{Q}$ represents $(1-p_{cr})$-quantile of the Θ distribution. The error distributions of the surrogate model and the tuned low fidelity model, which are determined apriori, are also shown in these fig-

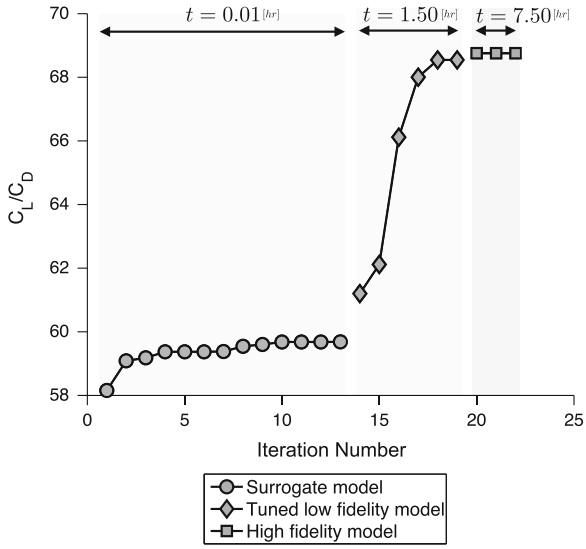

Fig. 10 Optimization history of the airfoil design problem

ures. Through AMS, model switching from the surrogate model to the tuned low fidelity model and from the tuned low fidelity model to the high fidelity model occur at the 13th and the 19th iteration, respectively.

The convergence history of the airfoil optimization is illustrated in Fig. 10. This figure also indicates which model is active at each iteration. It is observed that, from the first iteration till the 13th iteration the surrogate model (M_{SM}^{A}) is active, before switching to the tuned low fidelity model that remains active till the 19th iteration. Interestingly, most of the objective function improvement occurs under the tuned low fidelity model (more than 10 % increase in the CL/CD ratio). The optimization uses the high fidelity model in the last 3 iterations before reaching convergence. In this case, the algorithm converges by satisfying the predefined function tolerance, $\delta f = 10^{-5}$.

Next, the performance of the AMS method is investigated and compared with the performances of running optimizations that solely rely on a low fidelity model or a high fidelity model. The results yielded by the PSO-AMS thus compared with the results yielded by separately running MDPSO solely using the surrogate model (PSO-SM), solely using the tuned low fidelity model (PSO-TLF), and solely using the high fidelity model (PSO-HF). The optimum results thus obtained, the computational cost, and the total number of function evaluations in each case are reported in Table 3. The final column of this table shows the high fidelity function estimate at the optimum design obtained under each optimization run (e.g., $y^*_{HF}(x^*_{SM})$ and $y^*_{HF}(x^*_{TLF})$). It is observed that the PSO-AMS not only requires 185 % less computing time compared to PSO-HF, it also provides the best optimum value that is 5 % better that the next best value (where the 2nd best is obtained by PSO-TLF). It is also observed that, in the PSO-TLF approach, the optimum is located in the region where the TLF model

Table 3 2D Airfoil design: optimization results using single-fidelity and variable-fidelity optimization approaches

Approach	x_1^*	x_2^*	x_3^*	x_4^*	Optimum function C_L/C_D (f^*)	Model in last iteration	Computational time [hr] over function evaluation	HF response at optimum $(f_{HF}(x^*))$
PSO-SM	0.0003	0.0003	0.0003	4.8300	59.69	SM	0.275/990	59.43
PSO-TLF	−0.0028	−0.0058	−0.0014	2.7911	70.54	TLF	9.96/1380	65.20
PSO-HF	7.14E-5	−0.0021	−0.0018	4.8273	59.57	HF	25.7/360	59.57
PSO-AMS	−0.0020	−0.0004	−0.0009	2.8313	68.75	HF	9.01/660	68.75

PSO-SM optimization performed by MDPSO solely using the surrogate model
PSO-TLF optimization performed by MDPSO solely using the tuned low fidelity model
PSO-HF optimization performed by MDPSO solely using the high fidelity model
PSO-AMS optimization performed by MDPSO using AMS

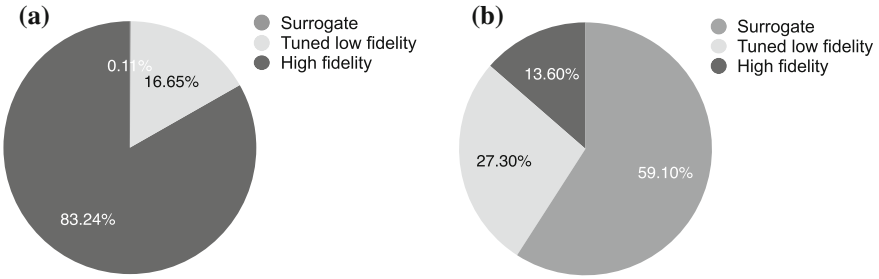

Fig. 11 Percentage of resources used by each model in the airfoil optimization problem performed through PSO-AMS: **a** Computing time resources; **b** Function evaluation resources

has more than 8 % error. This optimum is in the vicinity of the high fidelity optimum yielded by the AMS method. The optimization performed solely using the tuned low fidelity model (PSO-TLF) also incurs a slightly higher computational time in comparison with that performed using the AMS method, which is attributed to the high number of function evaluations invested to satisfy the termination criterion in the former (1380 evaluations vs. 660 evaluations).

Figure 11a, b illustrate the resources used in terms of computing time and function evaluations, by the three different models in the airfoil design optimization performed by PSO-AMS. These figures show that the overall computational cost is highly sensitive to the number of high fidelity model evaluations, which is expected. It is also observed that the surrogate model dominates the optimization process in terms of function calls, while the computational expense of this model is significantly lower than that of the tuned low fidelity and the high fidelity models. This observation supports the hypothesis that a probabilistic AMS technique can provide a significantly better balance between accuracy of the optimum and computational efficiency, compared to purely low fidelity or purely high fidelity optimizations.

3.2 Shape Optimization of a Cantilever Composite Beam

In the second optimization test problem, the maximum deflection of a cantilever composite beam (as shown in Fig. 12) is minimized. This beam is subjected to a parabolically-distributed load, $q(x) = q_0(1 - \frac{x^2}{L^2})$ [22]. In this problem, the fiber direction Young's modulus, E_L, and the composite weight density, ρ, are given by

$$\begin{aligned} &E_L = E_f v_f + E_m(1 - v_f) \\ &\rho = \rho_f v_f + \rho_m(1 - v_f) \\ &\text{where} \\ &v_f + v_m = 1 \end{aligned} \tag{23}$$

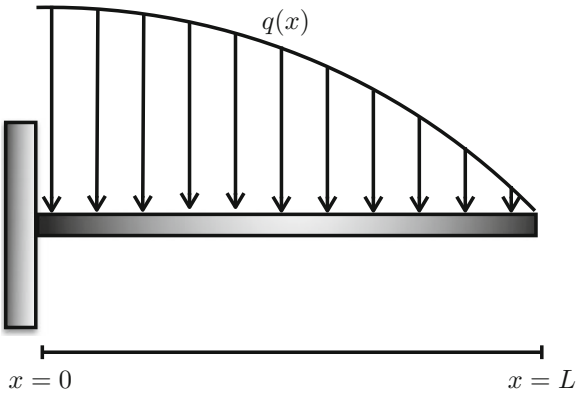

Fig. 12 Cantilever composite beam subjected to a parabolic distributed load

In Eq. 23, E_f and E_m are the elastic modulus for graphite and epoxy resin, respectively; ρ_f and ρ_m are the weight density of the graphite fiber and epoxy resin, respectively; and v_f and v_m respectively represent the fiber volume fraction and the matrix volume fraction in the continuous fiber composite material.

The design variables include (i) the second moment of area (x_1), (ii) the depth of the beam (x_2), and (iii) the fiber volume fraction (x_3). The side constraints on the design variables and the values of the prescribed design parameters are listed in Table 4 and Table 5, respectively.

The beam optimization problem is defined as

Table 4 Design variables for the beam design problem

Description	Notation	Lower limit	Upper limit
Second moment of area, I [mm^4]	x_1	$3.3E4$	$20.8E4$
Depth of the beam, h [mm]	x_2	20	50
Fiber volume fraction, v_f	x_3	0.40	0.90

Table 5 Prescribed design parameters for the beam design problem

Parameter	Value
Parabolic distributed load, q_0 [N/mm]	1
Length of the beam, L [mm]	1000
Elastic modulus of graphite fiber, E_f [N/mm^2]	$2.30E5$
Elastic modulus of epoxy resin, E_m [N/mm^2]	$3.45E5$
Weight density of graphite fiber, ρ_f [N/mm^3]	$1.72E-5$
Weight density of epoxy resin, ρ_m [N/mm^3]	$1.20E-5$

$$\text{Minimize: } \frac{\delta_{max}}{\delta_0}, \quad [\delta_0] = 12.93 \tag{24}$$

subject to

$$W/W_0 \leq 1, \quad [W_0] = 2.9E4 \tag{25}$$

$$\sigma_{max}/\sigma_0 \leq 1, \quad [\sigma_0] = 200 \tag{26}$$

$$\frac{x_2^4}{1.2E6x_1} \leq 1 \tag{27}$$

$$x_i^{min} \leq x_i \leq x_i^{max}, \quad i = 1, 2, 3 \tag{28}$$

In this optimization formulation, the inequality constraints (Eqs. 25, 26, and 27) are related to the allowable weight, the maximum stress, and a geometric restriction on the beam design ($depth \leq 10 \times width$). The weight and the maximum stress are given by

$$W = A\rho L = \frac{12I}{h^2} \times (12 + 5.2v_f)10^{-6} \times L = \frac{x_1}{x_2}(1440 + 624x_3) \tag{29}$$

$$\sigma_{max} = \frac{q_0 L^2 h}{8I} = \frac{1E6x_2}{8x_1} \tag{30}$$

The models used to estimate the maximum deflection, d_{max}, are described next.

3.2.1 Structural Models with Different Levels of Fidelity

To develop the high fidelity physics-based structural model (M^B_{HF}) and the low fidelity physics-based or PLF structural model (M^B_{PLF}), the Finite Element Analysis package ANSYS is used. In ANSYS, the PLF Finite Element model is constructed using 2 beam elements, while the HF Finite Elemet model comprises 1000 beam elements. The third model (M^B_{SM}) in this problem is a surrogate model constructed using Kriging with Gaussian correlation function. A set of 30 high fidelity function evaluations are used for this purpose. The fourth model (M^B_{TLF}) is a tuned low fidelity model constructed using the *Multiplicative form* where

$$C(x) = \frac{\delta_{max}|_{HF}}{\delta_{max}|_{PLF}} \tag{31}$$

The distribution of the error in the tuned low fidelity model (TLF) and the surrogate model (SM) are estimated using PEMF (Sect. 2.3.1) and are illustrated in Fig. 13a, c, respectively. The distribution of the error in the Physics-based low fidelity model (PLF) is estimated using the inverse assessment process, by leveraging the same 30 high fidelity samples that were used to construct the TLF and SM; the PLF error distribution is shown in Fig. 13b.

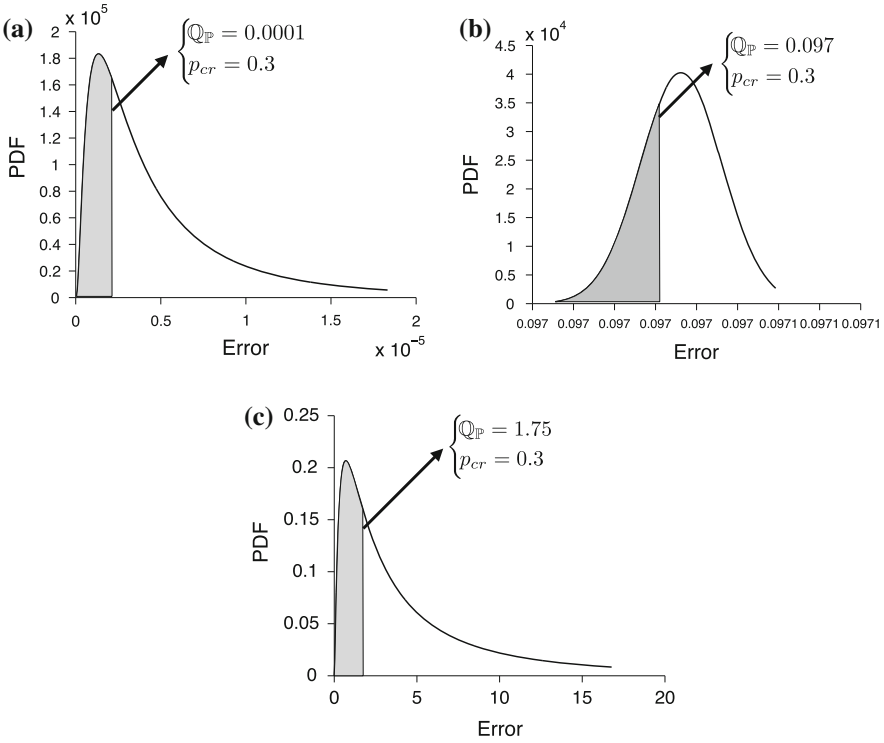

Fig. 13 Distributions of the model errors for the cantilever beam design problem; **a** Tuned LF model, **b** Physics-based LF model, **c** Surrogate model

Table 6 Models with different levels of fidelity used in the cantilever beam optimization problem (the high fidelity model is assumed to be a true representation of the system behavior)

Model	Location parameter μ	Scale parameter σ	$\mathbb{Q}(p_{cr})$ $p_{cr} = 0.3$	Number of calls made $N_{pop} \times No.\ of\ Iter.$
Surrogate	1.22	1.20	1.75	30×3
Physics-based LF	−2.30	0.001	0.097	30×6
Tuned LF	−12.52	0.99	0.0001	30×7
High fidelity	–	–	–	30×4

3.2.2 Cantilever Beam Design: Results and Discussion

For the cantilever beam design problem, the four model types, the error distribution parameters and $\mathbb{Q}(p_{cr})$ associated with each model, and the number of calls made by AMS to each model are listed in Table 6. It can be seen from Fig. 13 and Table 6

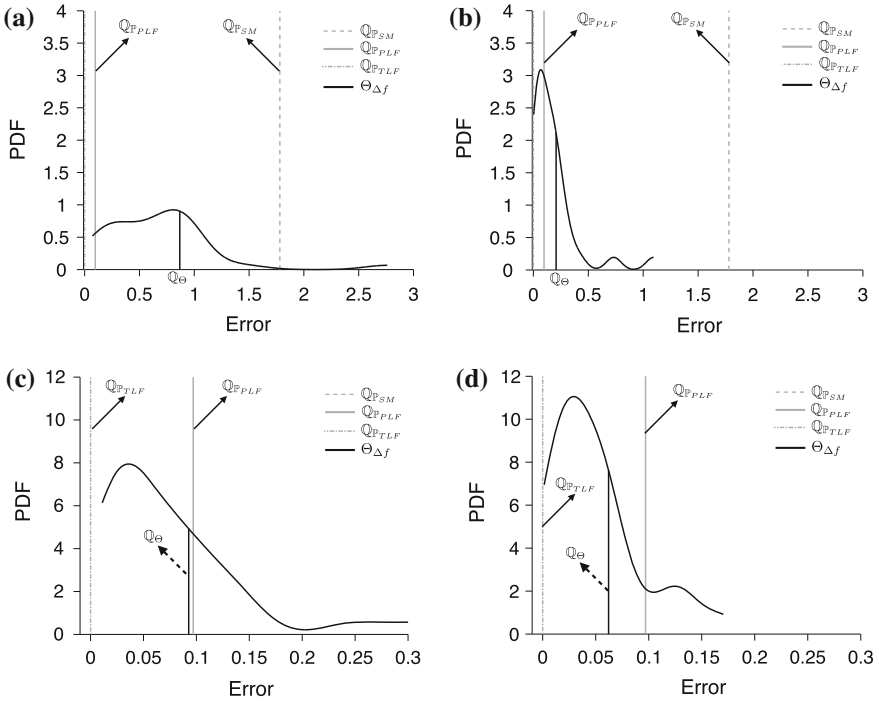

Fig. 14 Distribution of the fitness function improvements in different iterations of the beam optimization with PSO-AMS (also showing the model error distributions); **a** 3rd iteration, **b** 8th iteration, **c** 10th iteration, **d** 14th iteration

that the tuned low fidelity model provides the highest degree of accuracy and the surrogate model is the least accurate among the three low fidelity models. Hence, the initial population of particles is generated using the surrogate model in this case.

Figure 14a–d illustrate the distribution of the relative fitness function improvements ($\mathbb{Q}_{\Theta}$) at different iterations during the optimization process. The $(1 - p_{cr})$-quantile of the $\mathbb{Q}_{\Theta}$ distribution, and the p_{cr}-quantile of the error distributions of the tuned low fidelity model, the surrogate model, and the physics-based low fidelity model are also shown in these figures.

The convergence history of the cantilever beam optimization performed by PSO-AMS is illustrated in Fig. 15. The AMS technique adaptively switches the model type three times (SM -> LF -> TLF -> HF) during the optimization process at the 3rd, the 9th, and the 16th iteration, therefore resulting in an optimum design with a high fidelity function estimate. There is a substantial discontinuity in the estimated function value at the first switching event (3rd iteration), which can be attributed to the significant uncertainty in the surrogate model—the $\mathbb{Q}(p_{cr})$ value of the surrogate model (M^{B}_{SM}) is orders of magnitude higher than those of the other models (M^{B}_{PLF} and M^{B}_{TLF}). To avoid the termination of PSO before reaching the high fidelity model

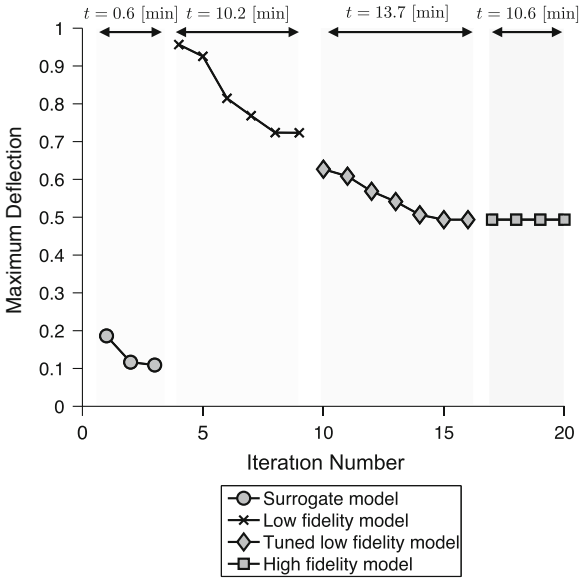

Fig. 15 Optimization history of the cantilever beam optimization with PSO-AMS

(M^B_{HF}), the relative function tolerance is set to $\delta = 10^{-5}$, which is smaller than the modal error of the tuned low fidelity model.

In Table 7, the optimization results obtained by PSO-AMS is compared with the results yielded by running MDPSO solely using the surrogate model (PSO-SM), solely using the physics-based low fidelity model (PSO-PLF), solely using the tuned low fidelity model (PSO-TLF), and solely using the high fidelity model (PSO-HF). Interestingly, the PSO-AMS, PSO-TLF, and PSO-HF arrive at the same optimum design with $f^* = 0.5435$. It is seen from Table 7 that PSO-AMS reaches this optimum design at a 33 % lower computational expense compared to PSO-TLF and a 119 % lower computational expense compared to PSO-HF (both expense differences are estimated with respect to PSO-AMS expense). It is important to note from Table 7 that the performance of the surrogate model-based optimization (PSO-SM) is significantly worse than that of the others. The error in the surrogate model (M^B_{SM}) at its optimum (X^*_{SM}) is more than 99 %, which is expected based on the predicted PEMF error of this model (Fig. 13c).

The resources used by the four different models, in terms of computing time and function calls, in the beam optimization performed by PSO-AMS are illustrated in Fig. 16a, b. It is observed that, unlike the airfoil problem, the surrogate model does not have a significant contribution in the beam optimization process in terms of function calls. Due to its high inaccuracy ($\mathbb{Q}(p_{cr}) = 1.75$), the fitness function improvement of the particles is quickly dominated by the error distribution of the surrogate model (in only 3 iterations). In this optimization process, the tuned low fidelity model (M^B_{TLF}) makes the highest contribution in terms of computing time and function calls. This case study again shows that the uncertainty in the lower fidelity

Table 7 Cantilever beam design: optimization results using single-fidelity and variable-fidelity optimization approaches

Approach	x_1^* $(e+4)$	x_2^*	x_3^*	Optimum function δ_{max}/δ_0 (f^*)	Model in last iteration	Computational time [min] over function evaluation	HF response at optimum $(f_{HF}(x^*))$
PSO-SM	2.82	43.28	0.71	0.0010	SM	5.7/990	0.8800
PSO-PLF	3.58	48.72	0.89	0.5011	PLF	18.4/330	0.5550
PSO-TLF	3.62	50.00	0.90	0.5435	TLF	46.8/720	0.5435
PSO-HF	3.62	50.00	0.90	0.5435	HF	76.85/870	0.5435
PSO-AMS	3.62	50.00	0.90	0.5435	HF	35.1/630	0.5435

PSO-SM optimization performed by MDPSO solely using the surrogate model
PSO-PLF optimization performed by MDPSO solely using the physics-based low fidelity model
PSO-TLF optimization performed by MDPSO solely using the tuned low fidelity model
PSO-HF optimization performed by MDPSO solely using the high fidelity model
PSO-AMS optimization performed by MDPSO using AMS

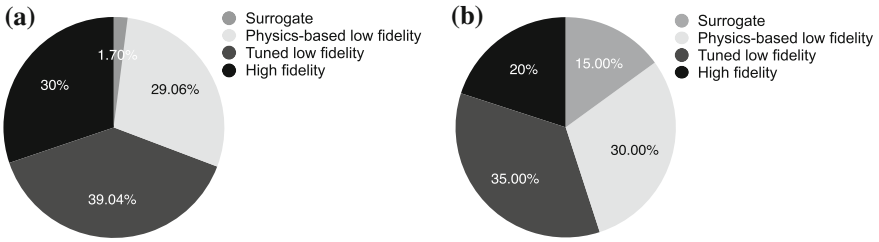

Fig. 16 Percentage of resources used by each model in the cantilever beam optimization problem using PSO-AMS : **a** Computing time resources; **b** Function evaluation resources

models could exceed the relative function improvement across constitutive iterations way ahead of reaching convergence in practical optimization, and this behavior is also highly problem dependent. Such likely scenarios make this variable fidelity optimization technique (AMS) a unique and essential tool for designing complex systems, where fast low fidelity models are almost indispensable.

4 Conclusion

This article presented a novel model management technique that is implemented in population-based optimization algorithms to provide high fidelity optimum designs at a reasonable computational expense. The model pool is created with models that offer different (non-dominated) trade-offs between computational cost and fidelity. The optimization process is started using the model with the highest computational efficiency, which could be a physics-based low fidelity model or a surrogate model. A novel switching metric (called Adaptive Model Switching or AMS) is then used to determine when to switch to the next higher fidelity model during the optimization iterations. Assuming that the uncertainties associated with the lower fidelity models follow a probabilistic distribution (lognormal pdf is used here), the proposed model switching metric is defined as: "a probability estimate of whether the uncertainty associated with a model exceeds the improvement in the relative fitness function over the population of solutions". The new adaptive model switching technique (AMS) is applied to: (i) 2D Airfoil design and (ii) Cantilever composite beam design. A powerful version of the Particle Swarm Optimization (mixed-discrete PSO) algorithm is used to implement and investigate the performance of AMS. The results indicate that AMS along with Mixed Discrete PSO improve the efficiency of the optimization process significantly when compared to optimization performed solely using high fidelity models, with up to 185 % reduction in computing time, while reaching the same or a better optimum. The value of the optimum with AMS is also better than that accomplished using only single low fidelity models for optimization. The current version of AMS is implemented primarily for optimization problems where multiple physics-based and/or surrogate models exist to represent the physical system behav-

ior. Future work will focus on problems where only a high fidelity physics-based model or experimental data is available, which can be used to construct different surrogates. A related notion is that of Surrogate-based design optimization, where surrogate models are improved through adaptive or sequential sampling during the optimization process. A more intuitive definition of the Indicator of Conservativeness (IoC) as a function of user's preferences regarding computational expense and robustness would further establish the wide potential of AMS for optimizing complex practical systems.

Acknowledgments Support from the National Science Foundation Awards CMMI-1100948 and CMMI-1437746 is gratefully acknowledged. Any opinions, findings, conclusions, or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the NSF.

References

1. Hutchison MG, Unger ER, Mason WH, Grossman B, Haftka RT (1994) Variable-complexity aerodynamic optimization of a high-speed civil transport wing. J Aircr 31(1):110–116
2. Jeong S, Murayama M, Yamamoto K (2005) Efficient optimization design method using kriging model. J Aircr 42(2):413–420
3. Oktem H, Erzurumlu T, Kurtaran H (2005) Application of response surface methodology in the optimization of cutting conditions for surface roughness. J Mater Proces Technol 170(1):11–16
4. Simpson T, Booker A, Ghosh D, Giunta A, Koch P, Yang RJ (2004) Approximation methods in multidisciplinary analysis and optimization: a panel discussion. Struct Multidiscip Optim 27(5):302–313
5. Simpson T, Toropov V, Balabanov V, Viana F (2008) Design and analysis of computer experiments in multidisciplinary design optimization: a review of how far we have come or not. In: 12th AIAA/ISSMO multidisciplinary analysis and optimization conference, Victoria, Canada
6. Wang G, Shan S (2007) Review of metamodeling techniques in support of engineering design optimization. J Mech Des 129(4):370–381
7. Jin R, Chen W, Simpson TW (2000) Comparative studies of metamodeling techniques under multiple modeling criteria. AIAA (4801)
8. Forrester A, Keane A (2009) Recent advances in surrogate-based optimization. Prog Aerosp Sci 45(1–3):50–79
9. Simpson T, Korte J, Mauery T, Mistree F (2001) Kriging models for global approximation in simulation-based multidisciplinary design optimization. AIAA J 39(12):2233–2241
10. Choi K, Young B, Yang R (2001) Moving least square method for reliability-based design optimization. In: 4th world congress of structural and multidisciplinary optimization, Dalian, China, pp 4–8
11. Toropov VV, Schramm U, Sahai A, Jones RD, Zeguer T (2005) Design optimization and stochastic analysis based on the moving least squares method. In: 6th world congresses of structural and multidisciplinary optimization, Rio de Janeiro
12. Hardy RL (1971) Multiquadric equations of topography and other irregular surfaces. J Geophys Res 76:1905–1915
13. Clarke SM, Griebsch JH, Simpson TW (2005) Analysis of support vector regression for approximation of complex engineering analyses. J Mech Des 127(6): 1077–1087
14. Yegnanarayana B (2004) Artificial neural networks. PHI Learning Pvt, Ltd, New Delhi
15. Zhang J, Chowdhury S, Messac A (2012) An adaptive hybrid surrogate model. Struct Multidiscip Optim 46(2):223–238

16. Mehmani A, Chowdhury S, Messac A (2014) A novel approach to simultaneous selection of surrogate models, constitutive kernels, and hyper-parameter values. In: 55th AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and materials conference. National Harbor, MD, USA
17. Zhang J, Chowdhury S, Mehmani A, Messac A (2014) Characterizing uncertainty attributable to surrogate models. J Mech Des 136(3):031004
18. Barthelemy JF, Haftka R (1993) Approximation concepts for optimum structural design (in a review). Struct Optim 5(3):129–144
19. Haftka RT (1991) Combining global and local approximations. AIAA J 29(9):1523–1525
20. Keane A, Nair P (2005) Computational approaches for aerospace design: the pursuit of excellence. Wiley, Chichester
21. Zadeh PM, Mehmani A (2010) Multidisciplinary design optimization using variable fidelity modeling: application to a wing based on high fidelity models. In: Third international conference on multidisciplinary design optimziation, Paris, France
22. Zadeh PM, Toropov VV, Wood AS (2009) Metamodel-based collaborative optimization framework. Struct Multidiscip Optim 38(2):103–115
23. Alexandrov NM, Lewis RM, Gumbert C, Green L, Newman P (1999) Optimization with variable-fidelity models applied to wing design. Technical report, ICASE, Institute for Computer Applications in Science and Engineering. NASA Langley Research Center, Hampton, Virginia
24. Booker AJ, Dennis JE, Frank PD, Serafini DB, Torczon V, Trosset MW (1999) A rigorous framework for optimization of expensive functions by surrogates. Struct Optim 17(1):1–13
25. Marduel X, Tribes C, Trepanier JY (2006) Variable-fidelity optimization: efficiency and robustness. Optim Eng 7(4):479–500
26. Robinson TD, Eldred MS, Willcox KE, Haimes R (2008) Surrogate-based optimization using multifidelity models with variable parameterization and corrected space mapping. AIAA J 46(11):2814–2822
27. Rodriguez JF, Perez VM, Padmanabhan D, Renaud JE (2001) Sequential approximate optimization using variable fidelity response surface approximations. Struct Multidiscip Optim 22(1):24–34
28. Alexandrov NM, Dennis JE, Lewis RM, Torczon V (1998) A trust-region framework for managing the use of approximation models in optimization. Struct Optim 15(1):16–23
29. Toropov VV, Alvarez LF (1998) Development of mars-multipoint approximation method based on the response surface fitting. AIAA J 98: 4769
30. Forrester A, Sobester A, Keane A (2008) Engineering design via surrogate modelling: a practical guide. Wiley, Chichester
31. Sugiyama M (2006) Active learning in approximately linear regression based on conditional expectation of generalization error. J Mach Learn Res 7:141–166
32. Trosset MW, Torczon V (1997) Numerical optimization using computer experiments. Technicl report, DTIC Document
33. Bichon BJ, Eldred MS, Mahadevan S, McFarland JM (2013) Efficient global surrogate modeling for reliability-based design optimization. J Mech Des 135(1):011, 009
34. Duan Q, Sorooshian S, Gupta V (1992) Effective and efficient global optimization for conceptual rainfall-runoff models. Water Resour Res 28(4):1015–1031
35. Jones D, Schonlau M, Welch W (1998) Efficient global optimization of expensive black-box functions. J Glob Optim 13(4):455–492
36. Kleijnen JP, Beers WV, Nieuwenhuyse IV (2012) Expected improvement in efficient global optimization through bootstrapped kriging. J Glob Optim 54(1):59–73
37. Jin Y, Olhofer M, Sendhoff B (2002) A framework for evolutionary optimization with approximate fitness functions. IEEE Trans Evolut Comput 6(5):481–494
38. Graning L, Jin Y, Sendhoff B (2007) Individual-based management of meta-models for evolutionary optimization with application to three-dimensional blade optimization. In: Evolutionary computation in dynamic and uncertain environments, pp 225–250

39. Jin Y (2005) A comprehensive survey of fitness approximation in evolutionary computation. Soft Comput 9(1):3–12
40. Ulmer H, Streichert F, Zell A (2004) Evolution strategies with controlled model assistance. In: Evolutionary computation, 2004, IEEE congress on CEC2004, vol 2, pp 1569–1576
41. Jin Y, Sendhoff B (2004) Reducing fitness evaluations using clustering techniques and neural network ensembles. In: Genetic and evolutionary computation, GECCO 2004, pp 688–699
42. Chowdhury S, Tong W, Messac A, Zhang J (2013) A mixed-discrete particle swarm optimization algorithm with explicit diversity-preservation. Struct Multidiscip Optim 47(3):367–388
43. Epanechnikov V (1969) Non-parametric estimation of a multivariate probability density. Theory Probab Appl 14:153–158
44. Duong T, Hazelton M (2003) Plug-in bandwidth matrices for bivariate kernel density estimation. Nonparametric Stat 15(1):17–30
45. Mehmani A, Chowdhury S, Messac A (2015) Predictive quantification of surrogate model fidelity based on modal variations with sample density. Struct Multidiscip Optim (Accepted)
46. Mehmani A, Chowdhury S, Zhang J, Tong W, Messac A (2013) Quantifying regional error in surrogates by modeling its relationship with sample density. In: 54th AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and materials conference, Boston, MA, USA
47. Chowdhury S, Mehmani A, Messac A (2014) Concurrent surrogate model selection (cosmos) based on predictive estimation of model fidelity. In: ASME 2014 international design engineering technical conferences (IDETC), Buffalo, NY
48. Kennedy J, Eberhart RC (1995) Particle swarmoptimization. In: IEEE international conference on neural networks, vol 6, pp 1942–1948
49. Coelho F, Breitkopf P, Knopf-Lenoir C (2008) Model reduction for multidisciplinary optimization: application to a 2d wing. Struct Multidiscip Optim 37(1):29–48