Adaptive Multi-fidelity Surrogate Modelling for High-quality Shape Optimization

Jeroen Wackers, Hayriye Pehlivan Solak, Michel Visonneau, Riccardo Pellegrini, Andrea Serani, Matteo Diez

Adaptive Multi-fidelity Surrogate Modelling for High-quality Shape Optimization

Jeroen Wackers, Hayriye Pehlivan Solak, Michel Visonneau
Ecole Centrale de Nantes
FRANCE

Riccardo Pellegrini, Andrea Serani, Matteo Diez
National Research Council
Institute of Marine Engineering
ITALY

Keywords: Workshop, Multi-fidelity surrogate, Active learning, Computational fluid dynamics, Adaptive grid refinement

ABSTRACT

A multi-fidelity surrogate modelling approach for shape optimization, which relies on adaptive techniques to obtain good performance for a large range of problems, is presented and critically evaluated. Furthermore, an approach to adaptive selection of the fidelity levels to be used is presented. Adaptation is shown to be effective for solving complex problems. Finally, potential improvements in the noise canceling, the uncertainty estimation, and the adaptive sampling are identified.

1 INTRODUCTION

Multi-fidelity surrogate modelling is a promising method for simulation-driven design optimization (SDDO), especially when high-quality optimization based on accurate physics-based models, such as CFD, is desired. The addition of inexpensive low-fidelity data can greatly reduce the total cost for the optimization [1]. However, the behavior of a multi-fidelity method depends strongly on the simulation data which form its input. Factors such as numerical noise, the number of dimensions, the optimum shape, and the number of local optima all have an influence on the best optimization strategy for a given problem. Thus, a consensus appears to be forming in the research community, that there is not one unique multi-fidelity method which is suited for all problems.

It is our conviction that the only way to obtain a general shape optimization methodology which does not require extensive tuning by the user, is through adaptivity. Ideally, each component of the optimization algorithm should have the capability to adjust its behavior to what the data require. This is the long-term objective of our research.

The current document is intended as a discussion paper on the subject of adaptivity. We briefly introduce our current multi-fidelity learning approach (section 2) and flow solver (section 3), focusing on their adaptive elements. Then, the behavior of the method for different analytical and SDDO problems is shown in section 4. Finally, in section 5 we discuss the strengths and weaknesses of the proposed adaptive algorithms and deduce new research directions for the further evolution of adaptive multi-fidelity SDDO.

2 MULTI-FIDELITY ACTIVE LEARNING METHOD

2.1 Multi-Fidelity Setting

Consider $\mathbf{x} \in \mathbb{R}^D$ as a design variables vector of dimension D. Let the true merit function to be optimized $f(\mathbf{x})$, be assessed by N fidelity levels: the highest-fidelity level is $f_1(\mathbf{x})$, the lowest-fidelity is $f_N(\mathbf{x})$, and the intermediate fidelity levels are $\{f_i\}_{i=2}^{N-1}(\mathbf{x})$. Using $\widetilde{\cdot}$ to denote surrogate model prediction and $\widehat{\cdot}$ for multi-fidelity prediction, the MF approximation $\widehat{f}_i(\mathbf{x})$ of $f_i(\mathbf{x})$ $(i = 1, \ldots, N-1)$ is the sum of the lowest-fidelity surrogate and surrogates of the errors (inter-level errors or bridge-functions, $\widetilde{\varepsilon}(\mathbf{x})$) between subsequent levels

$$\widehat{f}_i(\mathbf{x}) = \widetilde{f}_N(\mathbf{x}) + \sum_{k=i}^{N-1} \widetilde{\varepsilon}_k(\mathbf{x}). \tag{1}$$

For each i-th fidelity level the training set is $\mathcal{T}_i = \{\mathbf{x}_j, f_i(\mathbf{x}_j)\}_{j=1}^{J_i}$, with J_i the training set size. The resulting inter-level error training set is defined as $\mathcal{E}_i = \{\mathbf{x}_j, \varepsilon_i(\mathbf{x}_j)\}_{j=1}^{J_i}$, where

$$\varepsilon_i(\mathbf{x}_j) = f_i(\mathbf{x}_j) - \widehat{f}_{i+1}(\mathbf{x}_j). \tag{2}$$

The surrogate models are based on stochastic radial basis functions (SRBF) which provide both the prediction and its associated uncertainty [2]. If the uncertainty $U_{\widetilde{f}_N}$ of the lowest-fidelity prediction is uncorrelated with the uncertainty $U_{\widetilde{\varepsilon}_k}$ of the inter-level error predictions, the uncertainty $U_{\widehat{f}_i}$ of the MF prediction can be evaluated as $(i = 1, \ldots, N-1)$

$$U_{\widehat{f}_i}(\mathbf{x}) = \sqrt{U^2_{\widetilde{f}_N}(\mathbf{x}) + \sum_{k=i}^{N-1} U^2_{\widetilde{\varepsilon}_k}(\mathbf{x})}. \tag{3}$$

2.2 Stochastic Radial Basis Functions with Least Squares Approximation

Given a (single-fidelity) training set $\mathcal{T} = \{\mathbf{x}_j, f(\mathbf{x}_j)\}_{j=1}^{J}$, the SRBF surrogate model prediction $\widetilde{f}(\mathbf{x})$ is computed as the expected value (EV) over a stochastic tuning parameter of the surrogate model [2], $\tau \sim \text{unif}[1, 3]$

$$\begin{aligned} \widetilde{f}(\mathbf{x}) &= \text{EV}\left[g(\mathbf{x}, \tau)\right]_\tau, \\ g(\mathbf{x}, \tau) &= \text{EV}\left[\mathbf{f}\right] + \sum_{k=1}^{M} w_k ||\mathbf{x} - \mathbf{c}_k||^\tau, \end{aligned} \tag{4}$$

where $\mathbf{f}$ is the vector of training data, w_k are unknown coefficients, $||\cdot||$ is the Euclidean norm and $\mathbf{c}_k$ are the RBF centers, with $k = 1, \ldots, M$ and $M \leq J$. Noise reduction in the training set is achieved by choosing a number of RBF centers M smaller than the number of training points J, and the $\mathbf{c}_k$ coordinates are defined via k-means clustering [3] of the training points coordinates. w_k are determined with least squares regression by solving $\mathbf{w} = (\mathbf{A}^\mathsf{T}\mathbf{A})^{-1}\mathbf{A}^\mathsf{T}(\mathbf{f} - \text{EV}\left[\mathbf{f}\right])$, where the $M \times J$ matrix $\mathbf{A}$ contains the values of the M kernels in the J training points. The optimal number of stochastic RBF centers ($M^\star$) is defined by minimizing a leave-one-out cross-validation (LOOCV) metric [4].

2.3 Initial Training Set and Bounded Surrogate Model

The reduced initial training set (RS) [5] is used: except on the lowest fidelity level, where the domain center and the centers of the boundary faces are sampled, the surrogate models are initialized with only a point in the domain center. For the error surrogate models, this requires a SRBF surrogate which can handle extrapolation. Therefore, a bounded surrogate model prediction and uncertainty (both identified with the B subscript) are defined as

$$\widetilde{\varepsilon}_{Bi}(\mathbf{x}) = \widetilde{\varepsilon}_i(\mathbf{x})\,[1 - s_i(r)] + \mathrm{EV}[\varepsilon_i]s_i(r), \tag{5}$$

$$U_{\widetilde{\varepsilon}_{Bi}}(\mathbf{x}) = \min(U_{\widetilde{\varepsilon}_i}, \mathrm{EV}[\varepsilon_i]), \tag{6}$$

where a sigmoid-like function $s(r)$ is used to provide a smooth transition between the SRBF prediction and the bounded prediction:

$$s(r) = \frac{1}{1 + e^{\alpha(r-\gamma)}}. \tag{7}$$

For the present work, $\alpha = -75$ and $\gamma = 0.2$. To define r, the smallest hyperrectangle (whose edges are parallel to the Cartesian coordinated axis) containing the training points is defined and r is the Euclidean distance of $\mathbf{x}$ from the hyperrectangle boundaries.

The definition of $U_{\widetilde{\varepsilon}_{Bi}}(\mathbf{x})$ stems from the consideration that the error surrogates represent errors in the simulation results. Therefore the average error can be used as reference for the surrogate model prediction uncertainty when an extrapolation is performed. This is acceptable since great precision is not required far away from the high-fidelity points; the most important requirement is, that the uncertainty remains bounded.

2.4 Active Learning Method

The MF surrogate model is dynamically updated by adding new training points. First, a new training point $\mathbf{x}^\star$ is identified based on a modified version of the Lower Confidence Bound (LCB) [6] infill criterion presented in [7], which samples points with large prediction uncertainty and small objective function value:

$$\mathbf{x}^\star = \underset{\mathbf{x}}{\mathrm{argmin}} \left[\widehat{f}(\mathbf{x}) - U_{\widehat{f}}(\mathbf{x}) + P_x(\mathbf{x})\right]. \tag{8}$$

$P_x(\mathbf{x})$ is a penalization factor based on the distance from the existing training sets (considering all the fidelities) to prevent the sampling of already sampled points. Once $\mathbf{x}^\star$ is identified, the fidelity used for the evaluation of $f(\mathbf{x}^\star)$ is selected. The new training point is added to the k-th training set $\mathcal{T}_k$ and to the lower-fidelity sets from $k+1$ up to N, where $k = \mathrm{maxloc}(\boldsymbol{\phi})$ and the elements of $\boldsymbol{\phi}$ are:

$$\phi_i = \begin{cases} \frac{\sqrt{U^2_{\widehat{\varepsilon}_i} - \mathrm{MSE}_i}}{\beta_i}, & \text{if } \mathrm{MSE}_i < U^2_{\widehat{\varepsilon}_i}, \\ \frac{U_{\widehat{\varepsilon}_i}}{\beta_i}, & \text{if } \mathrm{MSE}_i \geq U^2_{\widehat{\varepsilon}_i}, \end{cases} \qquad \phi_N = \begin{cases} \frac{\sqrt{U^2_{\widetilde{f}_N} - \mathrm{MSE}_N}}{\beta_N} & \text{if } \mathrm{MSE}_N < U^2_{\widetilde{f}_N}, \\ \frac{U_{\widetilde{f}_N}}{\beta_N} & \text{if } \mathrm{MSE}_N \geq U^2_{\widetilde{f}_N}. \end{cases} \tag{9}$$

Here $\beta_i = c_i/c_1$ with c_i the computational cost associated to the i-th level and c_1 the computational cost of the highest-fidelity. MSE_i is the mean-squared error computed as

$$\mathrm{MSE}_i = \begin{cases} \frac{1}{J_i}\sum_{j=1}^{J_i} \left[\varepsilon_i(\mathbf{y}_j) - \widetilde{\varepsilon}_i(\mathbf{y}_j)\right]^2 & \text{if } i < N, \\ \frac{1}{J_N}\sum_{j=1}^{J_N} \left[f_N(\mathbf{y}_j) - \widetilde{f}_N(\mathbf{y}_j)\right]^2 & \text{if } i = N. \end{cases} \tag{10}$$

The inclusion of the MSE term aims to better distribute the available budget of function evaluations among the fidelity levels. Indeed, when the least-squares regression is accurate for a fidelity and the surrogate prediction uncertainty decreases below the average noise variance in the training set, it is not interesting to continue sampling that fidelity level, since the surrogate prediction is already accurate compared with the noise affecting the training set. Thus, adding more (noisy) training points will not improve the surrogate model accuracy.

3 FLOW SOLVER AND ADAPTIVE SIMULATION

Simulations to generate the input data are performed with the Navier-Stokes solver ISIS-CFD [8] developed at ECN – CNRS, available in the FINE™/Marine computing suite from Cadence Design Systems. ISIS-CFD is an incompressible unstructured finite-volume solver for multifluid flow. The velocity field is obtained from the momentum conservation equations and the pressure field is extracted from the mass conservation constraint transformed into a pressure equation. Free-surface flow is simulated with a conservation equation for the volume fraction of water, discretized with specific compressive discretization schemes.

Adaptivity is essential to the simulations: the computational grids are created through adaptive grid refinement [9], which is used to take into account the need for several fidelities of the MF method (Fig. 1). The interest of this procedure is that different fidelity results can be obtained by running the same simulations and simply changing the refinement threshold, a parameter which determines the global mesh fineness.

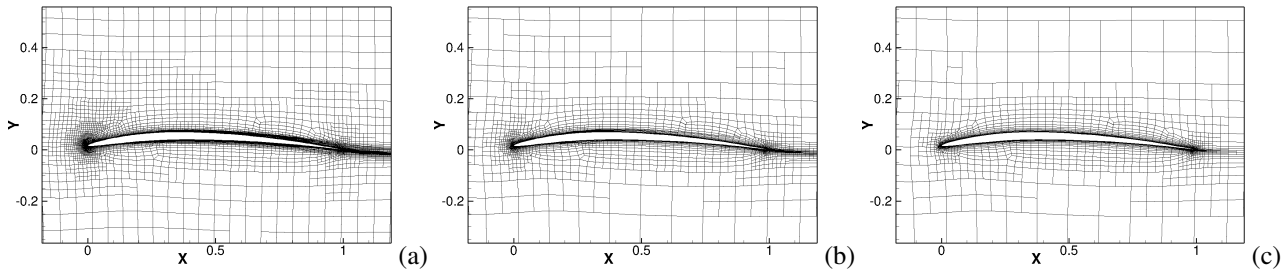


Figure 1: Adapted ISIS-CFD grids for the NACA problem of section 4.2: (a) high-fidelity, 12.8k cells, (b) medium-fidelity, 5.7k cells, and (c) low-fidelity, 3.6k cells.

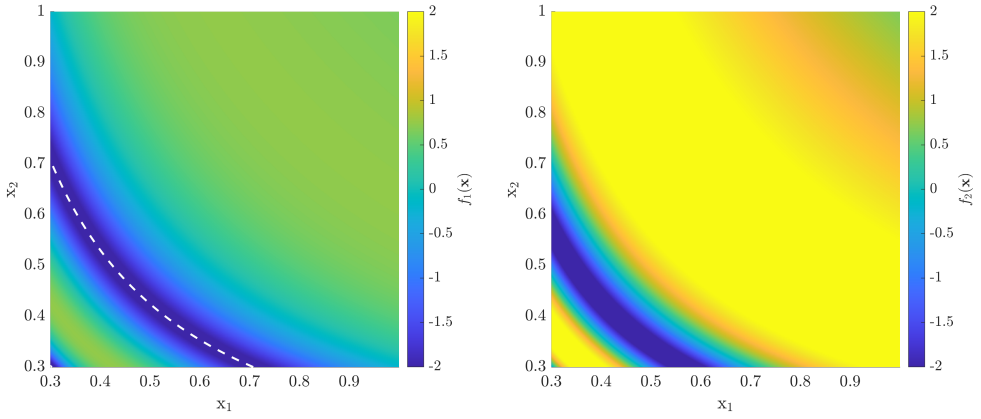
4 NUMERICAL TEST PROBLEMS

Numerical tests are presented to demonstrate the strengths of the current method, as well as its open issues. The MF SRBF method presented here (labeled RS-MSE) is compared with a method (labeled RS) where the correction with the MSE is removed from Eqs. (9) and with one (FS) where furthermore, the RS is replaced by a full startset on all levels. The optimization results are assessed by three error metrics [10], namely E_x and E_f evaluate design and goal accuracy, respectively. These comparisons are performed at the same computational cost CC which is taken proportional to the training set sizes $\mathcal{J}_l$: $CC = \mathcal{J}_1 + \sum_{l=2}^{N} \beta_l \mathcal{J}_l$.

All these test results have been published before [5]. For brevity, the detailed problem definitions and error metrics are not given here, but the original references are provided for those readers who are interested.

4.1 Analytical Test: Paciorek Problem

This two-dimensional optimization test is defined with two fidelity levels. This problem poses several challenges (see Fig. 2): first, it has multiple optima located on the loci $x_1x_2 = 2/(3\pi)$ and $x_1x_2 = 2/(7\pi)$, and second, the LF and HF optima are in different locations, which is a danger for multi-fidelity methods which rely too heavily on LF data. Finally, the LF data are perturbed by adding synthetic noise having a normal distribution with zero mean and variance equal to 10% of the high-fidelity function range.



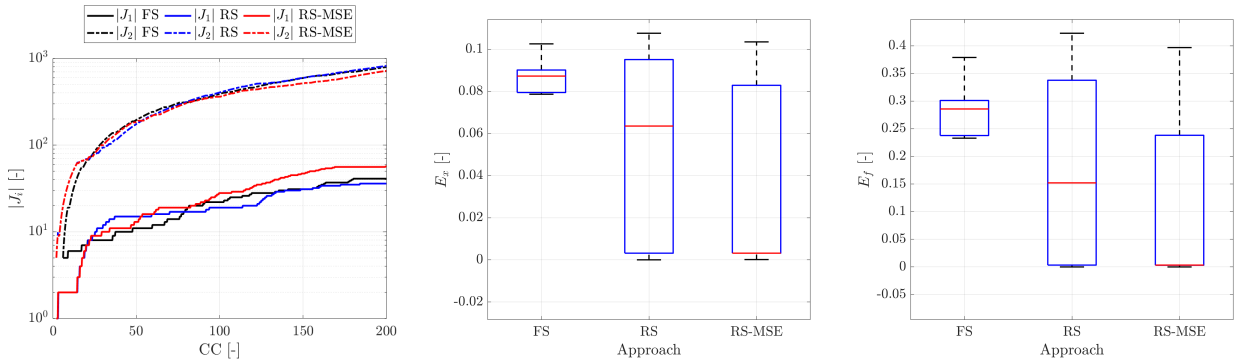
(a) High-fidelity $f_1(\mathbf{x})$ *(b) Low-fidelity $f_2(\mathbf{x})$*

Figure 2: Paciorek test problem. The dashed lines show the two loci with the lowest $f_1(\mathbf{x})$ value (the $x_1x_2 = 2/(7\pi)$ locus is in the neighborhood of the bottom-right corner).

In [5] we present a statistical analysis of results, created with 25 different realizations of the random noise; here the main results are reported. The evolution of the sample sizes over the course of the optimization (see Fig. 3a) shows the higher number of LF samples at the start (below $CC = 10$) for both RS and RS-MSE, indicating that the reduced startset not only decreases the cost of the initialization, but also encourages early adaptive exploration with low-fidelity samples only. This is followed by a larger increase of HF samples for RS-MSE, since LF sampling is discouraged once the LF uncertainty approaches the noise level.

Figure 3b-c shows box plots of the error in the optimum position E_x and value E_f. The RS and RS-MSE approaches achieve lower median values than the FS approach, while RS-MSE performs the best; its median result is an almost exact optimum. However, the FS approach has the smallest interquartile range. Thus, FS is consistent, but it is consistently wrong: the information in the large initial sample set forces the optimization into a fixed, but suboptimal direction. RS with its greater freedom performs better, but its reliance on LF data also leads to some bad results, since the LF optimum does not correspond to the HF one; hence the large inter-quartile range for this approach. Finally, RS-MSE with its initial LF sampling and final emphasis on HF samples provides the most consistently good results.

In Fig. 4 the collocation of HF samples with respect to x_1x_2 is studied: the quantity n_i is the number of elements in each bin of width 0.017. The figure shows that FS requests most of its high-fidelity training point close to the $x_1x_2 = 2/(3\pi)$ locus and only a small quantity in the neighborhood of the $x_1x_2 = 2/(7\pi)$ locus, negatively affecting the final performance. The RS approach requests almost the same quantity of high-fidelity training points between the two loci, in the position of the LF optimum. The RS-MSE approach requests for the highest number of high-fidelity training points in the $x_1x_2 = 2/(3\pi)$ locus and almost the same number of high-fidelity

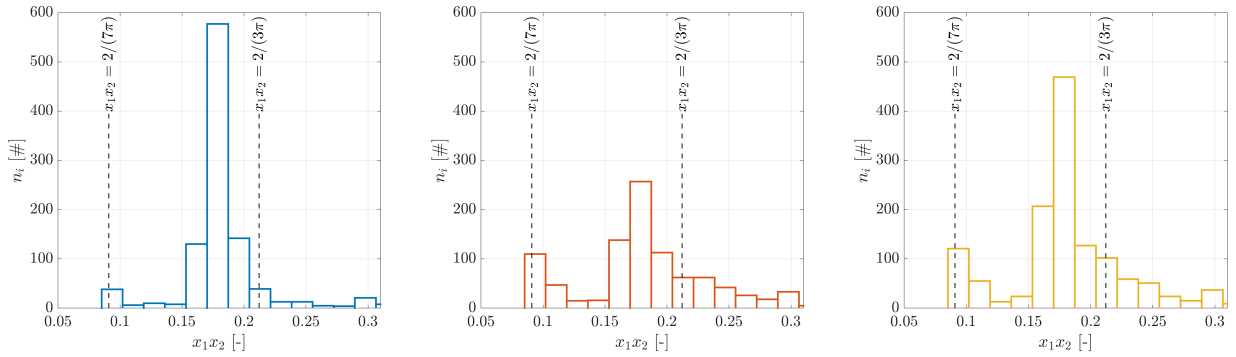


(a) Median of the training sets size versus CC

(b) E_x

(c) E_f

Figure 3: Analytical test, training sets sizes versus computational cost and box plots of the E_x and E_f metrics.



(a) x_1x_2 quantity for J_1 of FS

(b) x_1x_2 quantity for J_1 of RS

(c) x_1x_2 quantity for J_1 of RS-MSE

Figure 4: Analytical test, histogram of the x_1x_2 quantity for the high-fidelity training set (with n_i the number of elements in each bin).

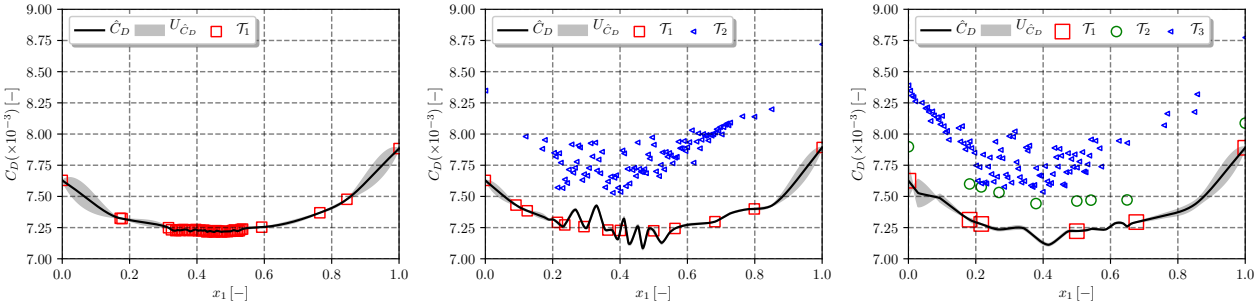
samples in the second locus as the RS approach. This explains why the RS-MSE approach achieves the best performance overall.

4.2 NACA 4-Digit Airfoil

The NACA 4-digit series is a family of airfoils defined by three design parameters: the thickness t, the maximum camber value m, and the position of the maximum camber p. We optimize the airfoil shape for minimum drag at a fixed lift. One- and two-parameter optimization problems can be defined by fixing some of the parameters as necessary. The challenge of the problem is that the optimum shape has a thin leading edge, whose resolution is highly dependent on the mesh. Therefore, the simulations are noisy, especially the low-fidelity ones.

Figure 5 shows the final iteration of a one-dimensional optimization (varying the maximum camber) from [4], comparing single-fidelity (HF) optimization with two-fidelity (HF – LF) and three fidelity levels, where an intermediate fidelity has been added. The three cases have the same total CC. Using a single-fidelity

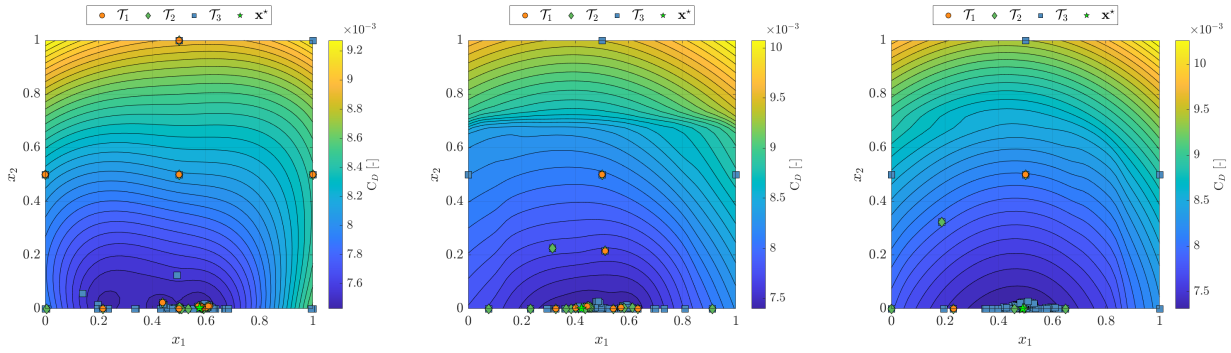
surrogate model (see Figure 5a) the noise in the CFD outputs is negligible. Differently, the use of two fidelities (see Figure 5b) introduces a significant amount of noise, negatively affecting the MF prediction. Finally, an intermediate-fidelity level (see Figure 5c) provides more effective filtering of the noise. Still, in this case, HF-only data produce the best surrogate model.



(a) $N = 1$ *(b)* $N = 2$ *(c)* $N = 3$

Figure 5: NACA airfoil MF surrogate models for $D = 1$.

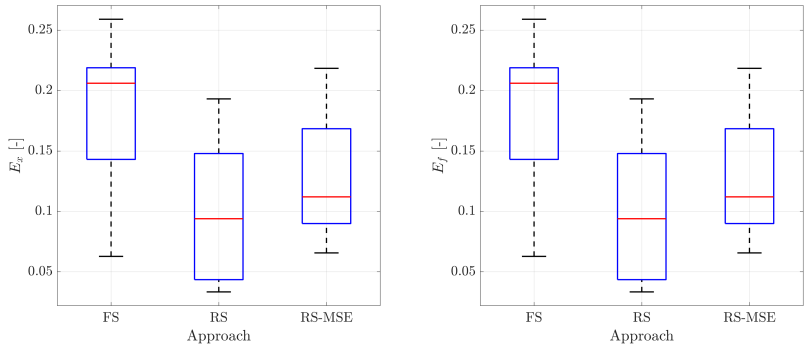
A statistical analysis based on 7 samples for a two-dimensional optimization (varying the maximum camber and thickness) using three-levels of fidelity is presented in [5]. Figure 6 shows the final iteration for the FS, RS and RS-MSE approaches. The FS approach has the same problem as for the analytical test case: the training points are clustered around an incorrect optimum position (the true optimum lies in $[0.3776, 0.0]$). RS again shows significant data clustering, but the points are better placed thanks to the less constrained initial exploration. Finally, RS-MSE surprisingly provides a concentration of LF points, which are not exactly in the optimum position but better placed than FS. This is confirmed by the box plots of the E_x and E_f metrics in Fig. 7: RS achieves the lowest errors and FS the highest.



(a) FS *(b) RS* *(c) RS-MSE*

Figure 6: $D = 2$ NACA problem, response surface and training sets for the three approaches.

Thus, for the NACA SDDO problem, the RS-MSE approach performs worse than RS. This could be caused by the limited number of low-fidelity samples in comparison with the analytical test problem. Another explanation is that the MSE becomes larger than the uncertainty due to the significant noise for this test case, which deactivates the procedure (Eq. (9)). As a preliminary check, this equation was changed to have zero uncertainty (*e.g.*, $U_{\varepsilon_i} = 0$ or $U_{\tilde{f}_N} = 0$) if $MSE_i > U_i^2$; the results show that the median numbers of high-, medium-, and low-fidelity samples change from 2, 13, 134 to 6, 15, 96, respectively. Thus, this modification seems effective in forcing



(a) E_x *(b)* E_f

Figure 7: $D = 2$ **NACA problem, box plots of the** E_x **and** E_f **metrics.**

the active learning towards higher fidelities in presence of noise. Nevertheless, when applied to the analytical test with two fidelities it completely prevented the sampling of the lower fidelity. The optimal formulation of RS-MSE is therefore an open problem.

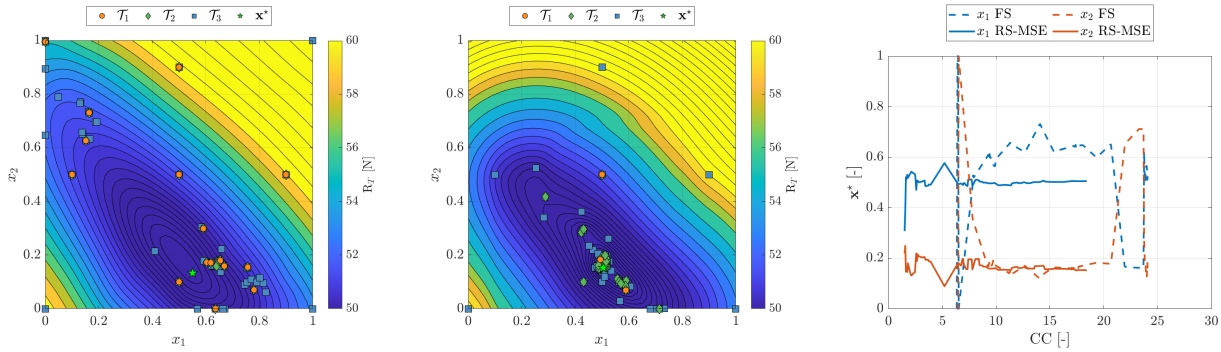
4.3 DTMB 5415 Model

Finally, the hull shape optimization of the DTMB 5415 destroyer for minimal resistance R_T [5] is presented. Since this optimization of 3D free-surface flow is characterized by high computational costs, the problem is solved only with the RS-MSE approach. The comparison with the FS approach is provided using the results presented in [4]. Since both the adaptive surrogate modelling strategy and the CFD simulations have changed, this does not provide a detailed assessment of one topic; rather, the comparison globally shows the progress that has been achieved in the last two years.

Figures 8a/b present the multi-fidelity surrogate models at the last iteration of the active learning approach, the MF training sets, and the predicted optima. The sampling strategies for the two approaches are significantly different: RS-MSE performed an exploration of the domain using only low-fidelity samples, correctly identifying the region of the minimum. The precision in this region is then increased using mainly medium-fidelity evaluations; only two high-fidelity points are sampled, one of which is almost in the optimum location. Near the end of the sampling, most points are added around the optimum. FS on the other hand, uses more HF points spread around the parameter space. Not all these points are useful; note for example the set of points in the top left corner, where a second minimum was suspected in the initial stages of the sampling. And while the data points are clustered, none are placed directly around the optimum.

These differences are also reflected in the $\mathbf{x}^\star$ convergence (see Fig. 8c). The CC of evaluating the startset is 7.35 for FS and only 1.24 for RS-MSE which, combined with the efficient initial LF exploration, means that RS-MSE has globally identified the optimum before FS finishes half its startset. The subsequent RS-MSE convergence is fast and without oscillations, as medium- and high-fidelity points are added around the optimum. The optimization has converged around $CC = 15$. The FS convergence is much more irregular, as it identifies two incorrect optima before finally settling on the correct one around $CC = 24$.

Table 1 summarizes the performance of the FS and RS-MSE approaches, where the prediction error $|E_p|$ is



(a) FS, R_T prediction *(b) RS-MSE, R_T prediction* *(c) $\mathbf{x}^\star$ convergence*

Figure 8: DTMB 5415 SDDO problem, multi-fidelity surrogate model prediction and $\mathbf{x}^\star$ convergence for the FS and RS-MSE approaches.

Table 1: DTMB 5415 SDDO problem, summary of the results.

Approach	CC	x_1	x_2	$\Delta_x\%$	$\Delta_f\%$	$\lvert E_p\rvert\%$	$\mathcal{J}_1$	$\mathcal{J}_2$	$\mathcal{J}_3$
FS	24.0	0.5506	0.1330	26.2	-4.5	1.73	16	18	72
RS-MSE	18.4	0.5043	0.1525	37.2	-4.9	0.87	3	44	103

computed between the MF surrogate model prediction of the minimum and its numerical validation. Although the CC is lower, the RS-MSE approach correctly identified the region of the minimum, using more low- and medium-fidelity data than the FS approach. The prediction error is twice lower, which confirms that the surrogate model is accurate around the optimum. Since a reference minimum is not available for this problem, two metrics are used instead of E_x and E_f, namely Δ_x and Δ_f that quantify the exploration of the design space and objective function reduction. The $\Delta_x\%$ value is larger for RS-MSE than for FS, meaning that the exploration for the identification of the minimum moved further. Finally, the RS-MSE approach achieves a lower resistance than the FS approach. Altogether, the RS-MSE optimization produces a similar optimum as the old FS result, in a more robust manner. Overall, the use of RS-MSE in combination with improvements made to the ISIS-CFD code [5], provided a reduction of the total wall clock time from 25 to about 11 days.

5 CONCLUSION: TOWARDS ENHANCED ADAPTIVITY IN MULTI-FIDELITY MODELS

The tests in section 4 show that the adaptive elements presented here are mostly successful. Progress has been made with respect to our earlier work and the adaptive processes work as intended. Still, the development of a fully adaptive and generally efficient multi-fidelity surrogate modelling approach requires more work. The purpose of this section is to evaluate the current approach to determine which elements deserve to be retained and which would benefit from further improvement.

Reduced Startset The RS approach is successful in all the cases where it is tested, improving the identification of the minima and better distributing the high-fidelity samples in interesting regions of the domain. There are three reasons for this. First, RS makes available a larger budget for the adaptive sampling to produce the right training sets. Furthermore, the reduced startset encourages early exploration with low-fidelity samples, unconstrained by high-fidelity data which may wrongly indicate sub-optimal regions. This early exploration

leads to a reliable identification of the region of the minimum. Later on, thanks to the limits imposed on the extrapolated high-fidelity uncertainty, sampling of high-fidelity data is concentrated in the observed region of the minimum instead of the domain corners, which efficiently increases the precision in the region of the minimum.

While RS is not an adaptive technique, its success underlines the validity of the adaptive sampling; we can have confidence in the idea of adaptively choosing the training points. Furthermore, RS and RS-MSE encourage *temporal* as well as spatial adjustment of the MF sampling: optimizations are started with LF exploration and finished with HF data around the optimum. The tests indicate that this may be a generally correct procedure.

MSE correction A key issue in choosing new datapoints is evaluating the gain obtained when a point is sampled, usually in terms of the reduction in prediction uncertainty. In our classical approach for selecting the fidelity to be sampled (based on U_i/β_i only, compare with eq. (9)) it is implicitly assumed that the uncertainty after sampling becomes zero, so the gain in uncertainty is equal to the uncertainty itself. In the presence of noise however, this is no longer justified. There will always be a moment when a fidelity level is sampled so much that new training points no longer add useful information, so the expected gain in accuracy becomes zero. Therefore, a procedure like equation (9) where the sampling of a level is discouraged once the noise filtering is established by an effective regression, is mandatory for surrogate model efficiency. However, the current MSE procedure has mixed results in the tests and it may be possible to improve it.

Two issues could be addressed. The first one is the question of what to do when the noise variance is higher than the prediction uncertainty. Currently, eq. (9) switches off the MSE correction in this case, restoring the possibility to sample the 'noisy' level. According to section 4.2, it would be more logical to set ϕ to zero, but this choice requires testing. The second point is that good noise filtering ought to be able to bring the multi-fidelity surrogate model uncertainty below the noise level, so there is no reason why the noise level should be the cutoff point for the higher fidelities. An alternative would be to multiply ϕ by an 'impact factor' which estimates the gain in uncertainty when a point is added. Once this factor drops towards zero (i.e. the uncertainty stays the same after sampling), the level is no longer to be sampled.

Noise filtering Given the amount of noise which can appear in simulations (see Fig. 5), some sort of noise treatment is mandatory. As RBF surrogate models cannot use noise kernels like Gaussian processes, the least squares approach with a reduced number of kernels is a logical choice. Furthermore, this technique adapts itself to spatially varying noise, something that a standard GP formulation cannot do: regions with much noise and close-lying sample points are likely candidates for clustering several points into a single RBF kernel, so especially in these regions, the surrogate model will not go through all the data values.

However, the current clustering and cross-validation procedure is not optimal. First of all, it is slow since the cross-validation metric has to be computed for several numbers of kernels M. Also, the response of the cross-validation metric to changes in M is jittery, which means that the optimal M may change rapidly when data are added. This is of course not good for the stability of the algorithm, and there is a risk of overfitting. The introduction of a constraint to limit the variation of the number of centers among the iterations may trap the LOOCV metric in a local minimum and limits the adaptation possibilities of the methods. Finally, the k-means clustering may not be the ideal distribution of the RBF kernels. All these points will be addressed in the future.

Uncertainty estimation Since any adaptive measure is aimed at managing and reducing the prediction uncertainty, accurate estimation of this uncertainty is the backbone of any efficient adaptive procedure. The uncertainty estimation of Volpi et al. [2], based on the variance of a series of RBF fits with different kernels, works well for us in practice. Among others, this estimation takes into account the behavior of the data,

producing lower uncertainty in the 'easier' areas of the design space, where all the interpolations agree. Furthermore, it naturally suits the noise filtering approach, since it predicts non-zero uncertainty in the noisy training points because the LS surrogate models do not pass exactly through these points.

Still, in a noisy multi-fidelity context, there are issues with the current estimation. Notably, the multi-fidelity uncertainty estimation (3) supposes that the uncertainties of the LF and error surrogate models are uncorrelated. And while this is correct without noise, when the only uncertainty comes from the interpolation between the data, it is no longer valid when the training points themselves are considered uncertain. The reason for this is that the data for the error surrogate models are the difference between simulation results on two levels, so the uncertainty in each simulation result affects at least two surrogate models.

For our approach, the choice of the error metamodel training set (2) is important. Isolating $\widehat{f}_{i+1}(\mathbf{x}_j)$ from this expression and substituting it in the identity $\widehat{f}_i(\mathbf{x}_j) = \widehat{f}_{i+1}(\mathbf{x}_j) + \tilde{\varepsilon}_i(\mathbf{x}_j)$ results in:

$$\widehat{f}_i(\mathbf{x}_j) = f_i(\mathbf{x}_j) + (\widetilde{\varepsilon}_i(\mathbf{x}_j) - \varepsilon_i(\mathbf{x}_j))\,.$$

This implies that the MF uncertainty $U_{\widehat{f}_i}(\mathbf{x}_j)$ depends on the simulation noise in $f_i(\mathbf{x}_j)$ and on how well the highest-level error surrogate model $\widetilde{\varepsilon}_i$ can filter this noise; none of the lower-fidelity surrogate models have any influence. Therefore, Eq. (3) is invalid in the neighborhood of the training points. An error estimation is needed which takes into account that the uncertainty close to training points only depends on the highest-fidelity data in that point, while the uncertainty far away from the training points depends on all the fidelity levels. A possible way to handle this, instead of using (4) for each separate surrogate model, is to construct entire noise-filtered multi-fidelity surrogate with standard RBF and different values of τ. The final surrogate model and its uncertainty are then deduced from the EV and 95% confidence interval of these MF non-stochastic surrogate models.

A second issue with the uncertainty estimation is the appearance of training point clustering, which is seen in the NACA test and, to a lesser extent, for the 5415. This is likely caused by two issues in the uncertainty estimation (which drives the adaptive sampling): a) with the noise filtering and cluster-based LS fitting, the uncertainty may not actually diminish when a training point is added, especially in a region where many training points are already available. b) if the estimated uncertainty is too low, the adaptive sampling (8) is mainly driven by the function value itself. In both cases, the addition of a training point will not change the behavior of the adaptive sampling criterion, so in the next iteration, the neighborhood of same training point is sampled again, leading to clusters of points.

Point a) may be improved by an uncertainty estimation which better evaluates the diminished uncertainty around training points. The MSE correction is also fundamental for reducing this issue. For point b), the uncertainty is underestimated if the true function lies outside the envelope of the RBF interpolations for different τ. For this reason, the LCB method can set a weighting > 1 for the uncertainty. This is a possibility to explore.

HF only It is a fact of life that HF-only surrogate modelling is sometimes better than MF. The NACA 1-parameter test is an example. Therefore, we believe that a fully adaptive MF surrogate model should be able to select the fidelities it uses, abandoning LF if necessary.

In principle, this is not that difficult. For any fidelity $i < N$, a single-fidelity surrogate model could be constructed based on the data available on that level. Then the uncertainty of that surrogate model could be compared with the MF surrogate model (3) to determine which of the two is the best choice. This optimal surrogate model for that level can then be used as LF model for the next level, or the next level may again require a single-fidelity model.

The difficulty of this approach is to locally vary the choice of the surrogate model combination. Since HF data are probably concentrated around the optimum, they cannot provide an effective surrogate model throughout the domain, so it is difficult to create HF-only surrogate models which are valid everywhere. However, if the LF surrogate model is used in some regions and not in others, then the resulting surrogate model may be discontinuous, which is a major problem for optimization. Maybe a gradual blending of MF and single-fidelity models could be envisaged?

ACKNOWLEDGMENTS

CNR-INM is partially supported by the Office of Naval Research through NICOP grant N62909-21-1-2042, administered by Dr. Elena McCarthy and by Dr. Woei-Min Lin of the Office of Naval Research Gobal and the Office of Naval Research, respectively. The work is conducted in collaboration with NATO STO AVT task group on "Goal-driven, multi-fidelity approaches for military vehicle system-level design" (AVT-331).

REFERENCES

1. Beran, P. S., Bryson, D. E., Thelen, A. S., Diez, M. & Serani, A. *Comparison of Multi-Fidelity Approaches for Military Vehicle Design* in *21th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference (MA&O)* (online, 2020).
2. Volpi, S. *et al.* Development and validation of a dynamic metamodel based on stochastic radial basis functions and uncertainty quantification. *Structural and Multidisciplinary Optimization* **51,** 347–368. ISSN: 1615-147X (2015).
3. Lloyd, S. Least squares quantization in PCM. *IEEE transactions on information theory* **28,** 129–137 (1982).
4. Wackers, J. *et al. Multi-Fidelity Machine Learning from Adaptive- and Multi-Grid RANS Simulations* in *33rd Symposium on Naval Hydrodynamics* (Osaka, Japan, 2020).
5. Wackers, J., Pellegrini, R., Diez, M., Serani, A. & Visonneau, M. *Improving Active Learning in Multi-fidelity Hydrodynamic Optimization* in *34th Symposium on Naval Hydrodynamics* (Washington, DC, 2022).
6. Cox, D. D. & John, S. *A Statistical Method for Global Optimization* in *IEEE International Conference on Systems, Man, and Cybernetics* (Chicago, IL, 1992), 1241–1246.
7. Serani, A. *et al.* Adaptive multi-fidelity sampling for CFD-based optimisation via radial basis function metamodels. *International Journal of Computational Fluid Dynamics* **33,** 237–255 (2019).
8. Queutey, P. & Visonneau, M. An Interface Capturing Method for Free-Surface Hydrodynamic Flows. *Computers & Fluids* **36,** 1481–1510 (2007).
9. Wackers, J. *et al.* Can adaptive grid refinement produce grid-independent solutions for incompressible flows? *Journal of Computational Physics* **344,** 364–380. ISSN: 0021-9991 (2017).
10. Serani, A. *et al.* Parameter selection in synchronous and asynchronous deterministic particle swarm optimization for ship hydrodynamics problems. *Applied Soft Computing* **49,** 313–334. ISSN: 1568-4946 (2016).