Chapter 7 – AIR BENCHMARK RESULTS

Dean E. Bryson, Daniel L. Clark Jr., Philip S. Beran
Air Force Research Laboratory
Wright-Patterson Air Force Base
UNITED STATES

Domenico Quagliarella
Italian Aerospace Research Centre
ITALY

Andrew S. Thelen, Markus P. Rumpfkeil
University of Dayton Research Institute
UNITED STATES

Melike Nikbay, Enes Cakmak, Sihmehmet Yildiz
Istanbul Technical University
TÜRKİYE

Francesco Di Fiore
Politecnico di Torino
ITALY

Laura Mainini
Politecnico di Torino and
Massachusetts Institute of Technology
ITALY

ABSTRACT

This chapter discusses the application of different MF methods to the L3 and L2 air benchmarks. The L3 benchmark problems consider drag and fuel burn minimization of a wing-fuselage-tail configuration using linear and compressible Euler aerodynamics, both with and without aeroelastic coupling. The L2 benchmark problems consider drag minimization of an airfoil using solutions of the Reynolds-averaged Navier-Stokes equations. Results are provided by Air Force Research Laboratory (AFRL), Istanbul Technical University (ITU), and University of Dayton Research Institute (UDRI) for the L3 benchmarks, whereas results are provided by Italian Aerospace Research Centre (CIRA), ITU, and Politecnico di Torino (PoliTO) for the L2 benchmarks. In the L3 and higher-dimensional L2 problems, the multi-fidelity methods effectively improved the predicted performance in the design problems to levels comparable with design using high-fidelity methods only. Performance in the lower-dimensional L2 problems also improved, but to a lesser degree compared to the high-fidelity, reference optimum. Some of the challenges and potential directions for future research are discussed.

7.1 INTRODUCTION

This chapter discusses MF analysis of the L3 and L2 air benchmark problems defined in Chapter 3. The L3 air benchmark challenges MF methods in four ways not challenged by the L3 sea benchmark (Chapter 8): (1) high-dimensionality of the design space for some benchmark variants (potentially hundreds of parameters); (2) a multidisciplinary design problem involving aeroelastic physics along with aerodynamic and/or structural design variables for some benchmark variants; (3) a challenging stress constraint for the flexible configuration, and (4) greater geometric freedom in some benchmark variants reflecting both the presence of planform and sectional design variables (such freedom increases the difficulty of robust analysis across the allowable design space, particularly for structurally flexible variants). The L2 air benchmark supports the L3 cases by capturing some essential modeling and analysis features; e.g., the presence of shocks at transonic speeds and the need for automated re-meshing (similar to the sea benchmarks).

Several MF methods contributed by different research organizations were applied to the air benchmarks,

generally assuming two levels of fidelity, for design optimization. These methods are described in detail in Chapter 2. Organizations contributing MF methods also supplied the results of MF analysis for the L2 and L3 air benchmarks. The reader is referred to a number of technical papers written or co-written by authors of this chapter: (1) overview of the L3 benchmark and low-dimensional results [1]; (2) high-dimensional aeroelastic design studies [2], [3], and (3) results of the L2 benchmark [4], [5].

UDRI and AFRL developed the L3 air benchmark and collected MF results using: (1) a Sparse Polynomial Chaos Expansion (SPCE) and kriging hybrid approach (SPCE-KG) [6]; (2) a gradient based, MF Broyden–Fletcher–Goldfarb–Shanno (BFGS) methodology [1], [7] for design optimization, and (3) a GP Regression Co-Kriging (RCK) method [1], [2]. ITU also computed results with a GP methodology with improved scalability through partial-least squares (GP-PLS) [8]. The RCK and GP-PLS approaches utilized Expected Improvement (EI) as an acquisition strategy (RCK-EI and GP-PLS-EI), while SPCE-KG made use of discrepancies between global and local regressions. Single fidelity results were also collected for comparison purposes. In the case of (RCK-EI), the high-fidelity reference method was regression kriging with expected improvement (RK-EI). Several of these methods were also applied to the L2 air benchmark contributed by CIRA, with the addition of CIRA's Asymmetric Multi-Objective Genetic algorithm (AMOGAe)[9] and PoliTO's variety of GP approaches [10]–[12] with different acquisition strategies.

L3 air benchmark results are presented first in Section 7.2, followed by L2 air benchmark results in Section 7.3. Each section will tabulate the benchmark variants executed and the applied methods. Concluding remarks are provided in 7.4, where attention is given to assessing the effectiveness and scalability of the different applied MF methods.

7.2 L3 BENCHMARK DESIGN OPTIMIZATION RESULTS

For the L3-Air problems, the emphasis is on comparison of multi-fidelity design optimization methods versus globally accurate surrogates. While accurate surrogate models are also of interest, the preponderance of methods contributed is in design optimization. Sections 7.2.1 and 7.2.2 present performance comparisons between multi- and single-fidelity, surrogate-based methods on the design problems of limited size (6 and 3 DVs, respectively), using Software Set 1 from Table 3-12. Section 7.2.3 presents results for multi- and single-fidelity, gradient-based optimization on larger scale design problems with 10's to 100's of design variables. These results utilize Software Set 2 due to easier computation of gradient information.

7.2.1 Planform Design for Fuel Burn Minimization (L3-Air-2 with 6 DVs)

Initially, for problems L3-Air-1 and -2, the contributing groups selected an initial design of experiments (DOE) best suited for their particular methods, and only one optimization run was conducted for each. Example results for L3-Air-2 are displayed in Figure 7-1, and the authors provide a complete set of results in [1]. While most methods converged to similar optima, one can observe an order-of-magnitude spread in computational cost, especially for the initial DOE (going from the leftmost point, which was the reference design, to the next data point in the figures). These results motivated the creation of common starting DOEs for all groups as well as reporting statistics taken over multiple runs due to the stochastic nature of these DOEs.

Thus, five different high-fidelity and multi-fidelity DOEs were created through Latin-hypercube sampling and distributed to the groups. The high-fidelity sample points in the multi-fidelity DOE were selected to the same as

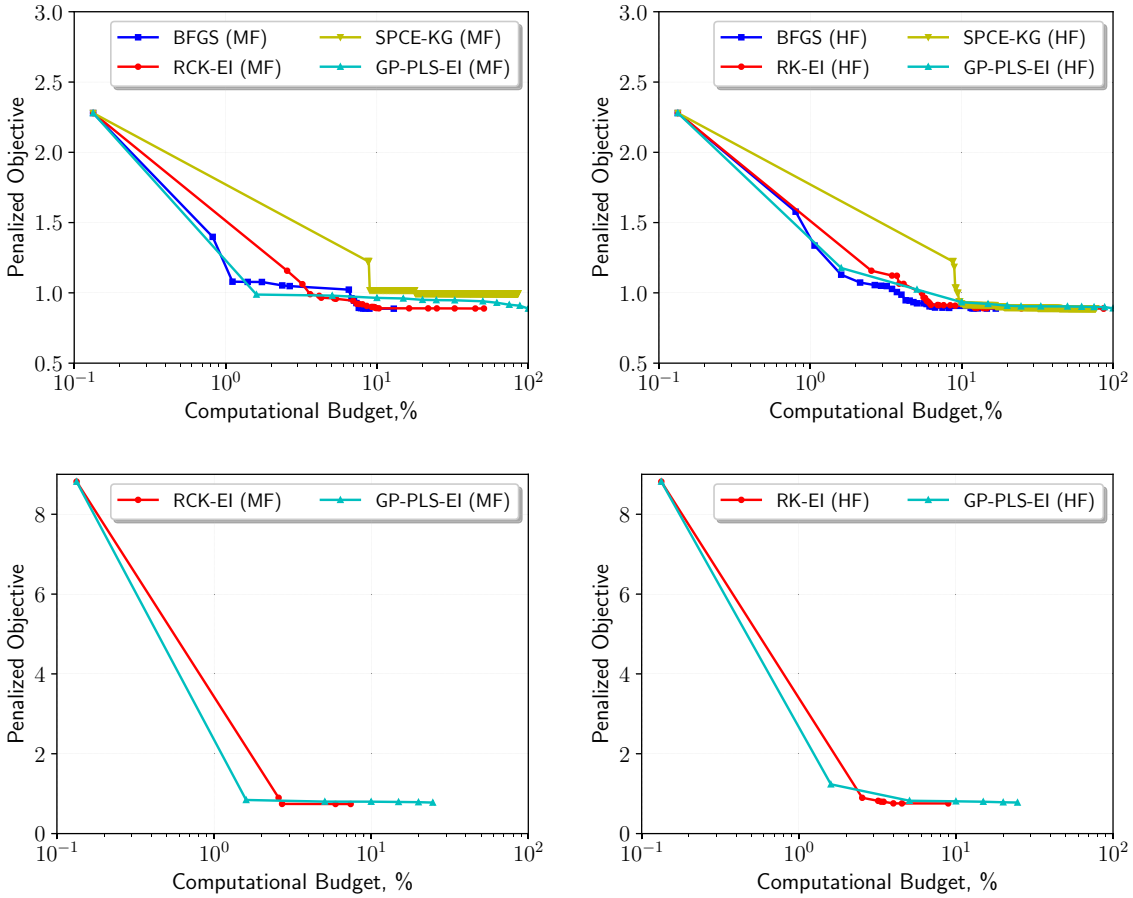


Figure 7-1: Optimization Performance for Rigid (top) and Aeroelastic (bottom) Fuel Burn Minimization Compared Between Multi-fidelity Methods (left) and Comparable High-fidelity Methods (right) Using Various Methods and Initial DOEs.

the high-fidelity DOE, with low-fidelity sample points added. The low-fidelity sample points were also a superset of the high-fidelity sample points. After performing the rigid and aeroelastic simulations, each group reported the mean of the best design found as well as the minimal and maximal best design found at each iteration. These results are shown in Figures 7-2 and 7-3 for the rigid and aeroelastic benchmark cases, respectively. Table 7-1 displays some details for the best design found by each optimization method over all 5 runs for both rigid and aeroelastic cases.

One can observe that all employed methods converge to a very similar normalized penalty function within roughly 10% of the allotted computational budget. Furthermore, the multi-fidelity methods converge slightly faster than their high-fidelity only counterparts with the exception of MF SPCE-KG. Both HF and MF-BFGS converge faster than the surrogate-rich methods due to the large number of data points required to train an accurate surrogate model in higher-dimensional spaces. Due to the high computational cost of the aeroelastic benchmark case only the GP-PLS-EI optimizer was able to be exercised. Figure 7-3 shows greater than 90% reduction in the normalized penalty function within 10% of the allotted computational budget, relative to

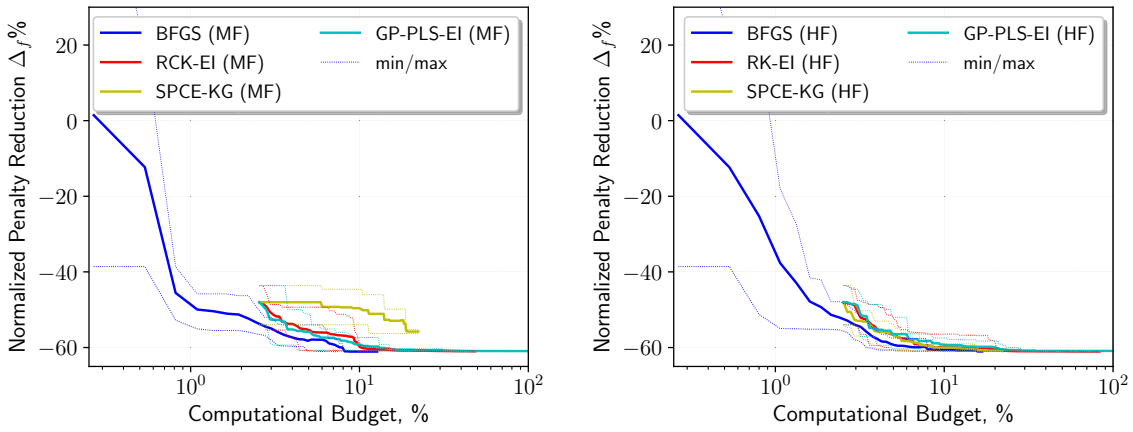


Figure 7-2: Optimization Performance for 6D Rigid Case Compared Between Multi-fidelity Methods (Left) and Comparable High-fidelity Methods (Right). Mean, minimum, and maximum observed over 5 repeated runs.

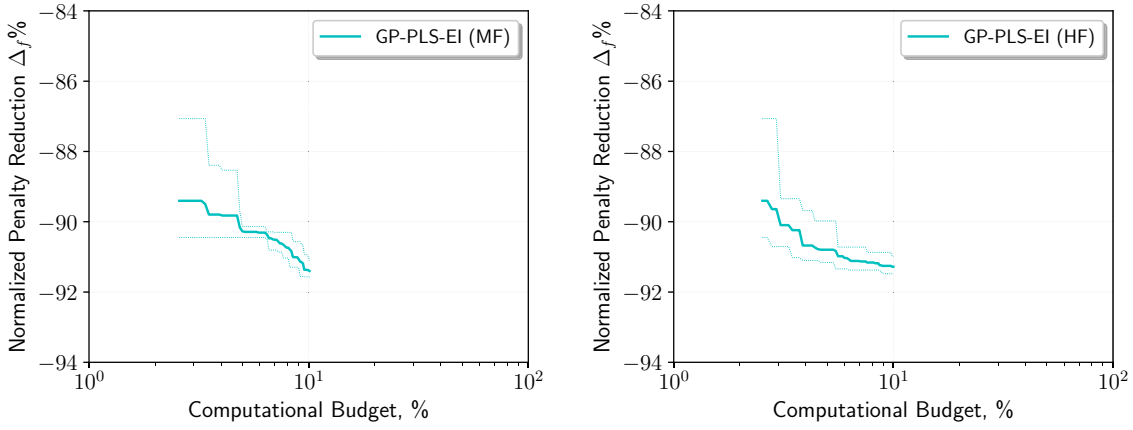


Figure 7-3: Optimization Performance for 6D Aeroelasatic Case Compared Between Multi-fidelity (Left) and High-fidelity (Right) Gaussian Process with Partial Least Squares. Mean, minimum, and maximum observed over 5 repeated runs.

Table 7-1: Best Design Found by Each Optimization Method for 6D Fuel Burn Minimization. † Indicates early termination of optimization. * Indicates best design from single optimization with independently selected initial design of experiments.

Optimization Method	x_1	x_2	x_3	x_4	x_5	x_6	f_{pen} $(\times 10^{-1})$	f_{obj} $(\times 10^{-1})$	g_{max} $(\times 10^{-2})$
Rigid Case									
BFGS (MF)	0.9310	1.0	0.7290	0.0	0.0	0.0	8.8690	8.2390	2.5398
BFGS (HF)	0.8750	1.0	0.7155	0.0	0.0	0.0	8.8829	8.3276	2.3776
SPCE-KG (MF)†	0.59560	1.0	0.5768	0.0203	1.0	0.0	9.9603	9.5078	2.0769
SPCE-KG (HF)	0.9048	1.0	0.7276	0.0	0.0	0.0	8.8862	8.2633	2.4384
GP-PLS-EI (MF)	0.9003	0.9850	0.7329	0.0044	0.0001	0.0033	8.8875	8.2651	2.5054
GP-PLS-EI (HF)	0.8945	0.9931	0.7337	0.0024	4.0×10^{-5}	0.0057	8.8989	8.2691	2.5283
RCK-EI (MF)	0.8997	0.9980	0.7300	0.0116	0.0011	0.0021	8.8874	8.2434	2.5182
RK-EI (HF)	0.9003	0.9978	0.7282	0.0114	0.0008	0.0038	8.8685	8.2565	2.5131
Aeroelastic Case									
GP-PLS-EI (MF)†	0.9302	0.9739	0.6510	0.9822	0.0906	0.4836	7.4349	7.3829	1.5424
GP-PLS-EI (HF)†	0.9608	0.9405	0.7212	0.9154	0.0595	0.5024	7.5146	7.3935	2.2960
GP-PLS-EI (MF)*	0.9998	0.6418	0.6814	0.6884	1.0	0.5159	7.7716	7.7716	0.0116
GP-PLS-EI (HF)*	1.0	0.5571	0.7846	0.9425	1.0	0.5865	7.7806	7.7175	0.4153
RCK-EI (MF)*	0.9967	0.8116	0.6456	0.9914	0.5712	0.4969	7.4121	7.3974	0.8074
RK-EI (HF)*	0.9936	0.6310	0.7337	0.9889	0.9107	0.5455	7.5593	7.4738	1.4865

the baseline design. Note that these optimizations were terminated early due to resource constraints and the similarity with results in Figure 7-1 and those previously reported in [1].

7.2.2 Planform Design for Drag Minimization (L3-Air-1 with 3 DVs)

Similar to the L3-Air-2 problems, Figures 7-4 and 7-5 show the mean, minimal, and maximal best designs found for the three-dimensional rigid and aeroelastic benchmark case, respectively. Again, all employed methods converge to a very similar normalized penalty function within roughly 10% of the allotted computational budget. In contrast to the six-dimensional L3-Air-2 case, the MF-BFGS convergence is similar to the surrogate based methods, and the single-fidelity BFGS is even slightly slower at the same cost as the initial DOEs. The difference between the L3-Air-1 and L3-Air-2 results highlights the relative scalability of the gradient- and surrogate-based methods. However, the wide bands between BFGS (both HF and MF) minimal and maximal designs within the first 5% of computational budget illustrate the sensitivity to initial design selection.

Once again, due to the high computational cost of repeated aeroelastic runs only the MF GP-PLS-EI optimizer was able to be exercised. From Figure 7-5, an 85% reduction in the normalized penalty function is realized within about 10% of the allotted computational budget. For the sake of methods comparison, Figure 7-6 shows the performance of individual aeroelastic optimization runs of GP-PLS-EI relative to RCK-EI and RK-EI. One can infer that both methods converge very quickly to a very similar optimum with no clear advantage in using multi-fidelity over high-fidelity data, reiterating the difference in initial training DOEs. However, both MF and HF GP-PLS-EI slightly outperform RCK-EI and RK-EI in these individual runs.

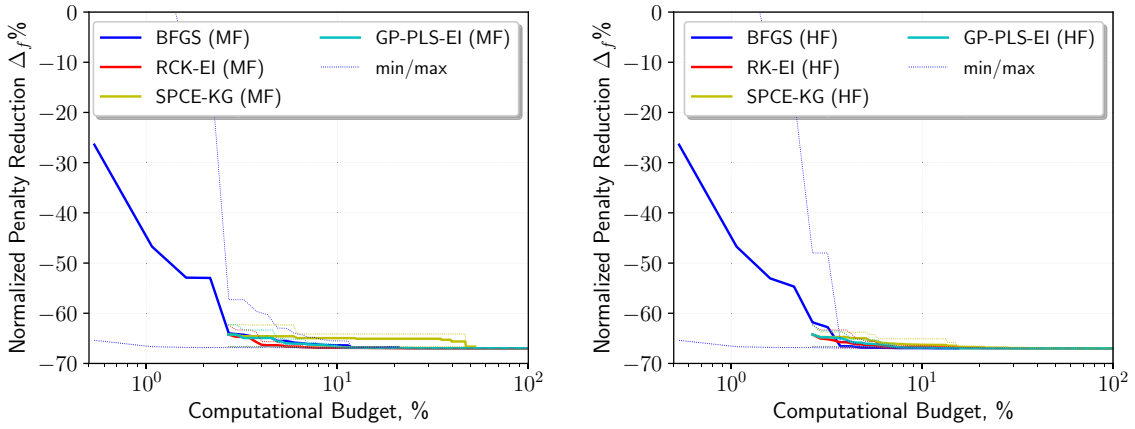


Figure 7-4: Optimization Performance for 3D Rigid Case Compared Between Multi-fidelity Methods (Left) and Comparable High-fidelity Methods (Right). Mean, minimum, and maximum observed over 5 repeated runs.

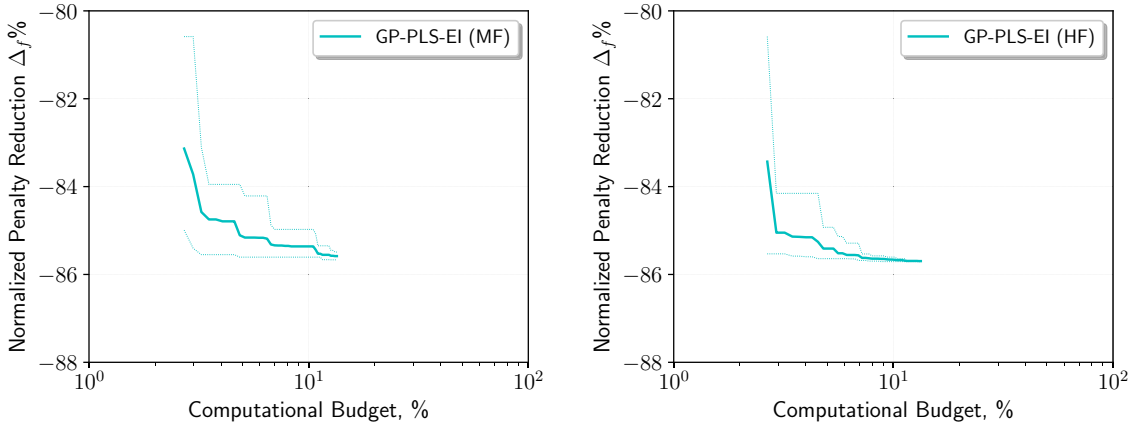


Figure 7-5: Optimization Performance for 3D Aeroelasatic Case Compared Between Multi-fidelity (Left) and High-fidelity (Right) Gaussian Process with Partial Least Squares. Mean, minimum, and maximum observed over 5 repeated runs.

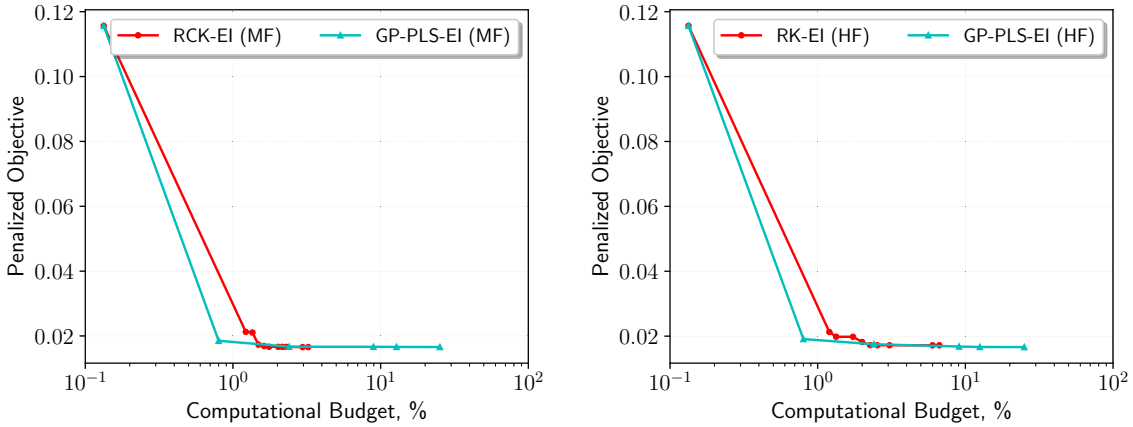


Figure 7-6: Optimization Performance for 3D Aeroelastic Case Compared Between Multi-fidelity Methods (Left) and Comparable High-fidelity Methods (Right).

Table 7-2: Best Design Found by Each Optimization Method for 3D Cases. † Indicates early terminiation of optimization. * Indicates best design from single optimization with independently selected initial design of experiments.

Optimization Method	x_1	x_2	x_3	f_{pen} $(\times 10^{-2})$	f_{obj} $(\times 10^{-2})$	g_{max} $(\times 10^{-2})$
Rigid Case						
BFGS (MF)	0.6070	0.8023	0.3075	2.0987	1.9497	3.8700
BFGS (HF)	0.5806	0.7841	0.2994	2.0928	1.9644	3.5936
SPCE-KG (MF)	0.9071	0.8044	0.3117	2.1140	2.0190	3.0820
SPCE-KG (HF)	0.6675	0.7778	0.2881	2.0925	1.9280	4.0550
GP-PLS-EI (MF)	0.6085	0.7876	0.2994	2.0932	1.9550	3.7274
GP-PLS-EI (HF)	0.7111	0.7735	0.2874	2.0910	1.9440	3.8435
RCK-EI (MF)	0.6860	0.7826	0.2936	2.0894	1.9419	3.8509
RK-EI (HF)	0.6974	0.7833	0.2924	2.0892	1.9414	3.8542
Aeroelastic Case						
GP-PLS-EI (MF)†	0.4509	0.7686	0.5754	1.6575	1.5668	3.0228
GP-PLS-EI (HF)†	0.4756	0.7663	0.5729	1.6528	1.5450	3.2932
GP-PLS-EI (MF)*	0.4022	0.8071	0.5835	1.6588	1.5363	3.5098
GP-PLS-EI (HF)*	0.4122	0.8008	0.5831	1.6626	1.5452	3.4363
RCK-EI (MF)*	0.4194	0.8081	0.5891	1.6542	1.5386	3.4099
RK-EI (HF)*	0.3585	0.99807	0.7236	1.7181	1.5892	3.6015

Details for the best design found over all five runs by each method (rigid and aeroelastic) are in Table 7-2.

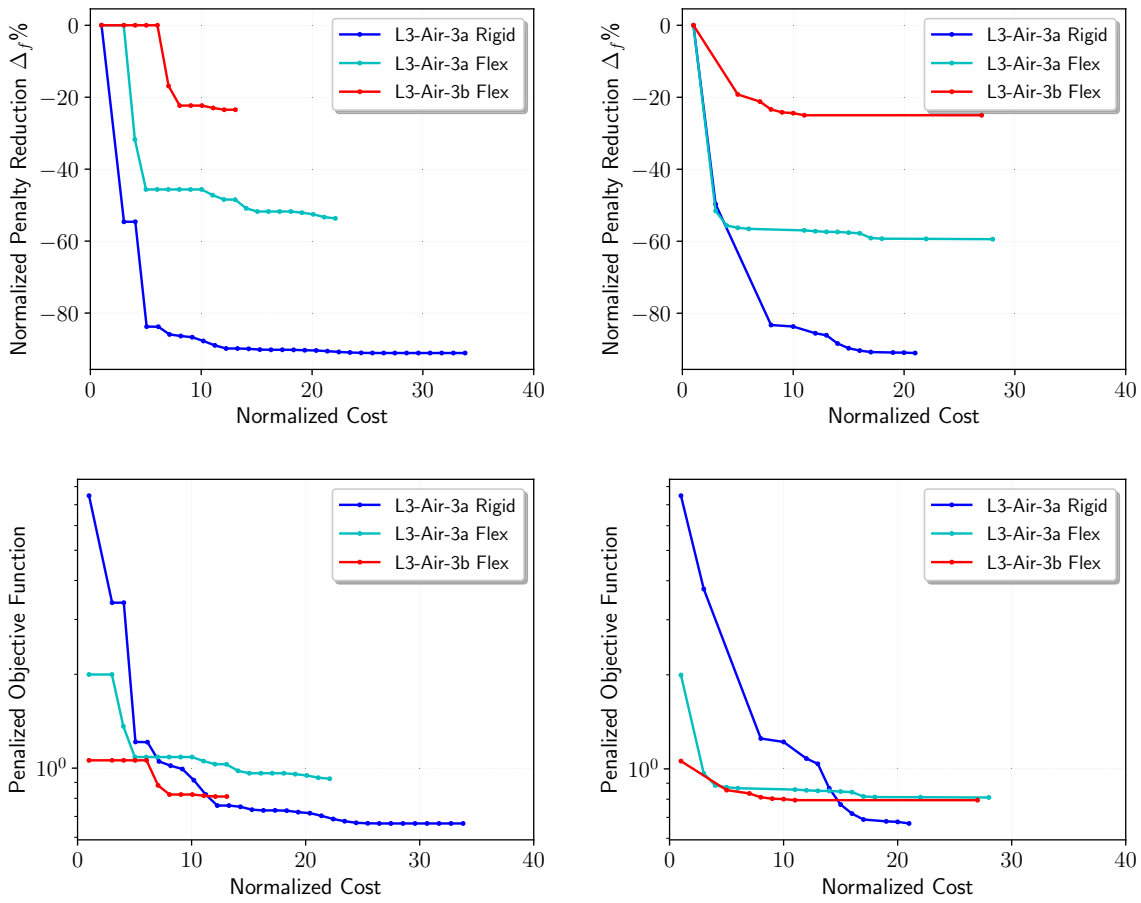


Figure 7-7: Optimization Performance for MF-BFGS (Left) and Single-fidelity BFGS (Right) for Rigid and Flexible Wing Shape Design.

7.2.3 Airfoil Shape Design and Structural Sizing (L3-Air-3/4/5 with 10's-100's of DVs)

Due to the large number of training data required and the cost of training global surrogate models in higher dimensions, only the MF-BFGS method was applied to the L3-Air-3 (10's of design variables) and L3-Air-4/5 (100's of design variables) design problems. The objective of each of these three problem classes was fuel burn minimization under varying assumptions and design variables. The L3-Air-3 variants studied included the rigid and aeroelastic analyses with 17 wing shape design variables (3a rigid and flex, respectively), as well as aeroelastic analysis with 37 wing shape design variables (3b flex).

The single-fidelity BFGS method provides a comparison for the multi-fidelity results in Figure 7-7. The upper plots display results in terms of percent reduction in penalty function from the initial design, while the lower plots show raw penalized objective function. In the lower plots, the disparities in initial value arise both from the difference in physics (rigid vs. aeroelastic) as well as number of degrees of freedom in the airfoil parameterizations (7 vs. 17). In terms of percent reduction (upper plots), these differences manifest in the ultimate reduction obtained.

In general, both single- and multi-fidelity versions of BFGS achieved the majority of penalty function reduction within the first 5-10 iterations. These reductions result from both reducing constraint violations (thereby reducing the penalty), as well as reducing the fuel burn objective. Thelen et al. [3] explores the convergence of objective and constraints in finer detail. Notably, the MF-BFGS convergence histories for the aeroelastic cases (both 17 and 37 DVs) are flat for the first several iterations, whereas the single-fidelity BFGS reduces the penalty function immediately. This behaviour results from the multi-fidelity optimizer rejecting the first several candidate designs, suggesting that the starting trust region may aggressively large and the corrections to the low-fidelity aeroelastic model maybe be inadequate initially. In contrast, the rigid L3-Air-3a case does not suffer from this difficulty, and the MF-BFGS reduces the penalty function more rapidly than the single-fidelity version before convergence levels off.

The scalability of the gradient-based MF-BFGS method also enabled optimization of the higher-dimensional L3-Air-4 (structural sizing) and L3-Air-5 (shape and sizing) problems. Similar to the L3-Air-3 problems, the objective was fuel burn minimization, with additional constraints on structural stress and trim for a maneuver condition. The addition of structural sizing introduced 285 additional design variables.

Figure 7-8 provides a comparison between MF and HF BFGS convergence. Again Thelen et al. [2], [3] explore the convergence behavior of the objective and constraints in finer detail. Comparing the multi- and single-fidelity results, both approaches perform comparably. This similarity suggests that while the MF-BFGS is capable scaling to 100's of design variables, further work is required to take better advantage of low-fidelity information.

7.3 L2 BENCHMARK RESULTS

7.3.1 Reference Optimum

The reference optimization run used the fluid dynamic solver only in high fidelity mode, with a grid characterized by approximately 150,000 cells. `SU2` v 6.2.0 ("Spalart-Allmaras" T. Mod.) [13] runs at c_l constant required 20,000 iterations (using the implicit solver) to achieve the accuracy needed and took an average elapsed time of about 140 minutes on 8 Intel (R) Xeon (R) cores E5-2697 v4 @ 2.30GHz. `SU2` CFL adaptation strategy is used with a lower limit of 2, and an upper one equal to 40.

The Covariance Matrix Adaptation Evolution Strategy (CMA-ES) [14] evolutionary strategy was chosen for the optimization. The parameters used for the CMA-ES optimization algorithm are the maximum number of allowed evaluations, the population size λ, and the initial standard deviation σ. The parameters set for this problem are reported in Table 7-3.

Twenty shape functions modify the airfoil shape (ten reserved for the upper side and ten for the lower side). In particular, they are six polynomials, 12 Hicks-Henne bumps, and 2 Wagner functions. The allowed variation range for each shape design variable is [-6.0, 6.0].

Table 7-3: Air Benchmark: CMA-ES Parameters for the Reference High-fidelity Optimization Run.

Max. evaluations	λ	σ
3000	16	0.15

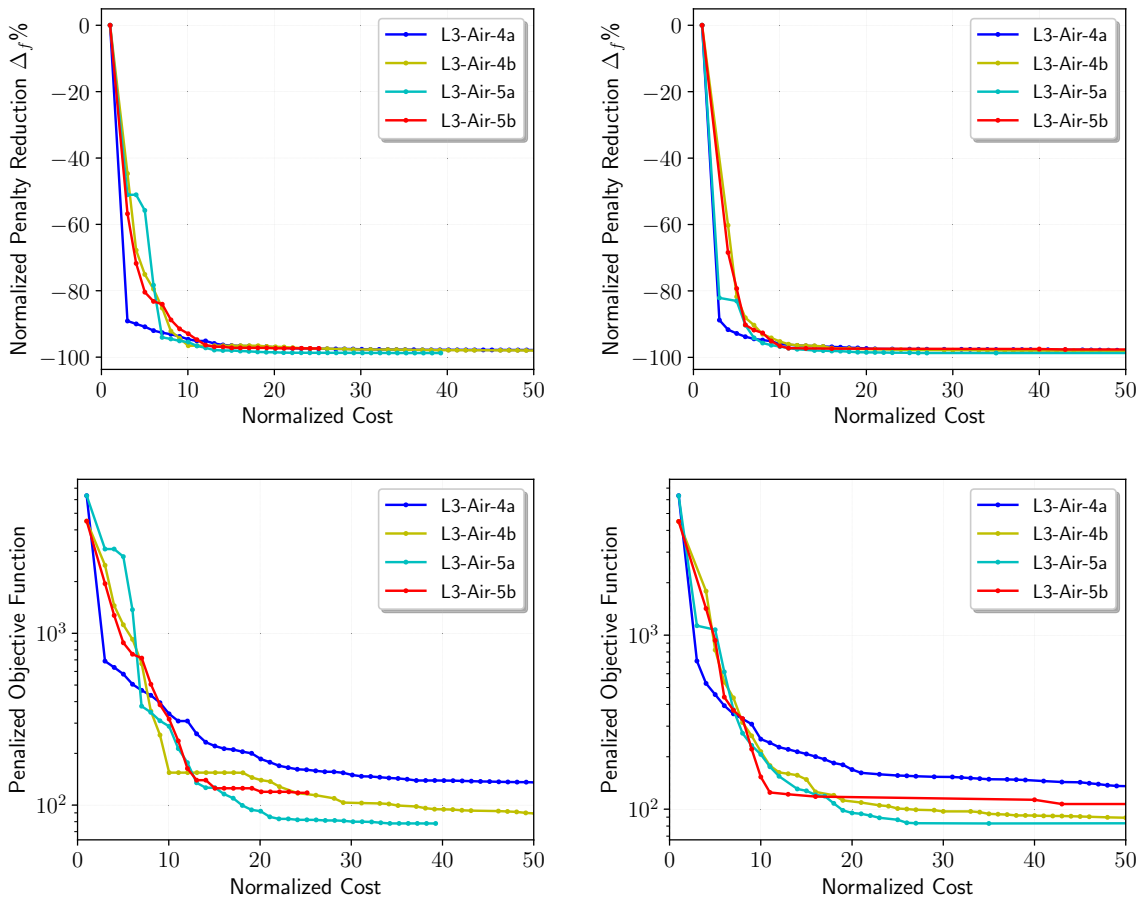


Figure 7-8: Optimization Performance for MF-BFGS (left) and Single-fidelity BFGS (right) for Structural Sizing with and Without Wing Shape Design.

Thanks to the fact that the optimization algorithm evaluates the population elements fully in parallel, the evaluation of each generation also takes about 140 minutes using 128 cores.

Figure 7-9 reports the baseline airfoil shape end the evolution history of the reference optimization run performed using the L2 driver in single, high-fidelity mode.

The design variable vector values for the initial and the optimal airfoil shapes are reported in table 7-4.

Table 7-5 shows a summary of the evaluation of the results of the reference optimization. The result obtained is particularly interesting in terms of single design point airfoil performance, although obtained at a relatively high computational cost.

Figure 7-10 compares the airfoils and the pressure coefficients between the baseline and the optimal solution found. Finally, Figure 7-11 shows the Mach number flow fields around the baseline RAE-2822 (left) and the optimal solution (right) that highlight a significant intensity reduction of the shock wave.

Table 7-4: Air Benchmark: Design Variable Values for RAE-2822 and Optimized Airfoil Shape.

DV	RAE-2822	Optimized
w_0	0.00000000	-0.79727763
w_1	0.00000000	0.29257670
w_2	0.00000000	0.29650784
w_3	0.00000000	0.09121300
w_4	0.00000000	1.11069288
w_5	0.00000000	-0.55537671
w_6	0.00000000	0.79334328
w_7	0.00000000	-0.04209589
w_8	0.00000000	-0.59810476
w_9	0.00000000	0.38285159
w_{10}	0.00000000	0.51927560
w_{11}	0.00000000	-0.17718765
w_{12}	0.00000000	-0.27731534
w_{13}	0.00000000	1.26317646
w_{14}	0.00000000	0.50190416
w_{15}	0.00000000	-0.08360092
w_{16}	0.00000000	-0.47016052
w_{17}	0.00000000	-0.73298168
w_{18}	0.00000000	-0.19085208
w_{19}	0.00000000	-0.36425107

Table 7-5: Air Benchmark: Summary of the Optimization Results.

airfoil	c_l	α	c_d	c_m	r	τ	t_{85}/c	objective
baseline RAE-2822	0.824	2.77	0.02048	-0.0997	0.0079	8.6808	0.0318	0.02048
best hi-fi opt.	0.824	1.29	0.01230	-0.1093	0.0085	8.6979	0.0223	0.01230
6 DV hi-fi opt.	0.824	2.59	0.01272	-0.1090	0.0074	10.8651	0.0357	0.01272

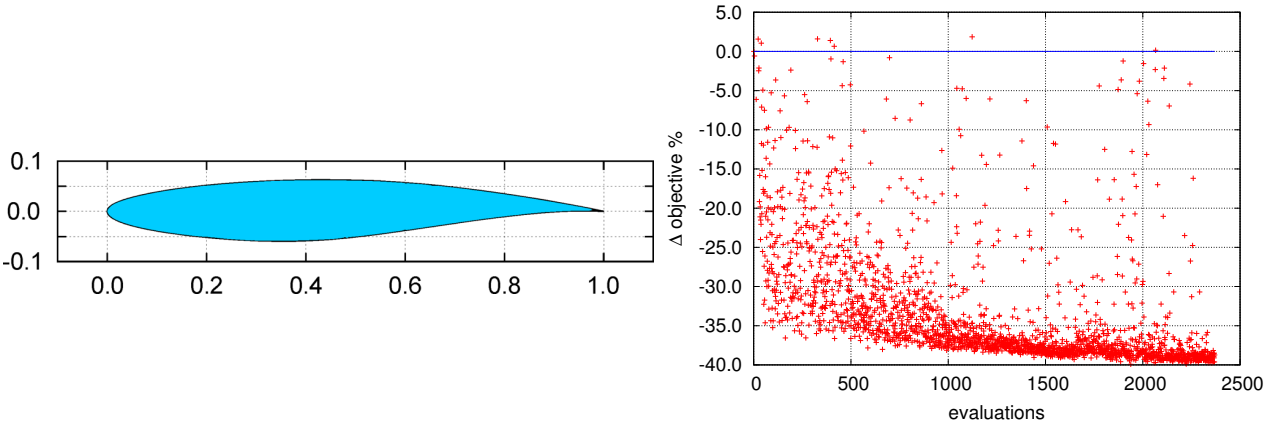


Figure 7-9: Air Benchmark L2: Baseline Airfoil and 20 DV Reference Run Evolution History.

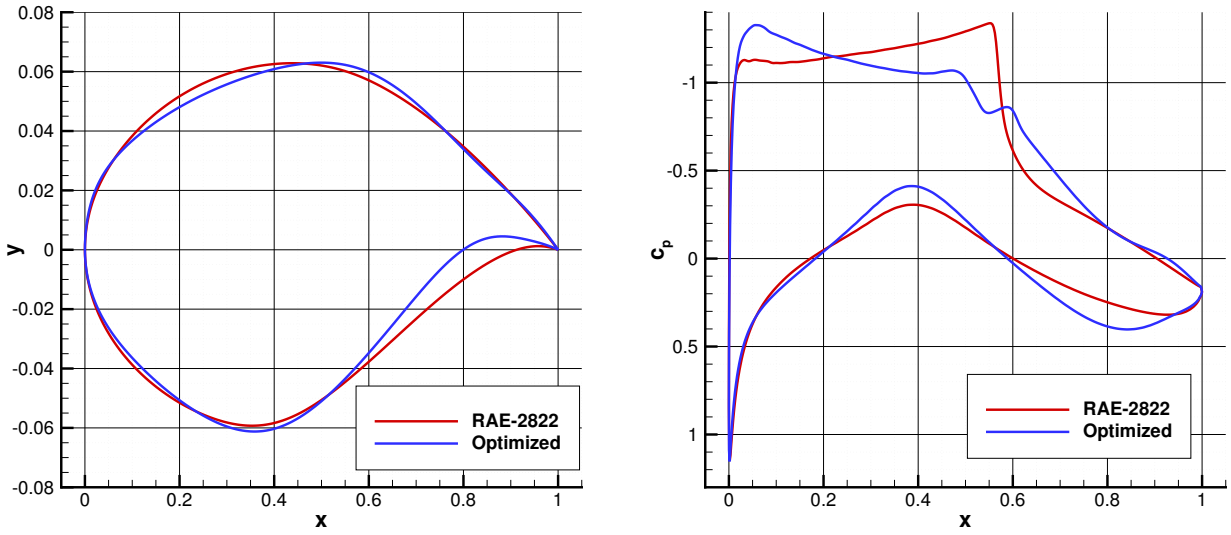


Figure 7-10: Air Benchmark L2: Comparison of Baseline RAE-2822 and Optimized Airfoil Shapes and Pressure Coefficients.

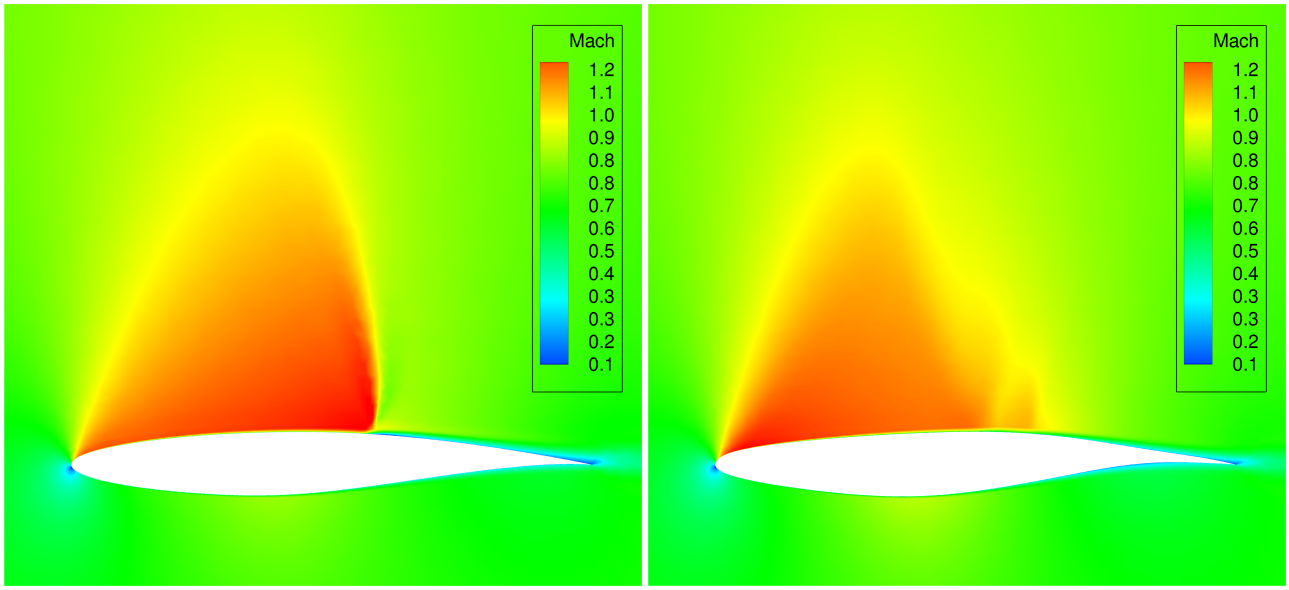


Figure 7-11: Air Benchmark L2: Comparison of Baseline RAE-2822 (Left) and Optimized Airfoil (Right) Mach Fields.

7.3.2 Six Design Variables Reference Optimum

A reference run was also carried out in the six design variables case to allow a fairer comparison in these conditions. In this run, the population size λ was reduced to eight elements, the initial σ was kept at 0.15, and the optimization run was stopped after 1201 objective function evaluations. The optimization history is reported in Figure 7-12 Again, Table 7-5 summarizes the evaluation of the six design variable reference optimization results.

It can be seen that the optimal solution, although slightly worse, is close in terms of c_d to that of the case with 20 design variables. However, the geometric characteristics of the airfoil are significantly different, as well as the distribution of the c_p, as can be seen from Figure XX reporting airfoil shape and pressure coefficient distribution comparisons, and from Table 7-6 showing the value of the six design variables.

Table 7-6: Air Benchmark: Design Variable Values for the Six Design Variables Optimized Airfoil Shape.

Design Variables			
w_0	-2.03621491	w_1	1.32237198
w_2	-0.13819503	w_3	0.33922546
w_4	2.42277353	w_5	-1.28732327

Figure 7-13 clearly shows that, although the difference between the c_d values between the two optimal solutions is not very large, the difference in shape between the two profiles is very significant. This implies that the objective function has quite a large plateau near the optimum point. This feature represents a considerable challenge in solving the benchmark with multi-fidelity methodologies because lower fidelity models may not be able to discriminate such differences.

7.3.3 Multi-Fidelity Design Optimization

Three different partners produced results for the L2 aero benchmark, namely, the Italian Aerospace Research Centre (CIRA), the Istanbul Technical University (ITU), and the Polytechnic University of Turin (POLITO). Each of the partners used different settings to generate low-fidelity solutions to best fit the feature of their algorithms resulting in varying levels of grid refinement and different elapsed computation times. These differences were considered by suitably balancing the low-fidelity computational cost when computing the equivalent total cost of each optimization run.

The description of the algorithms is reported in detail in Chapter 2, and here we will briefly summarize only the settings of the specific parameters for the resolution of the L2 aero benchmark.

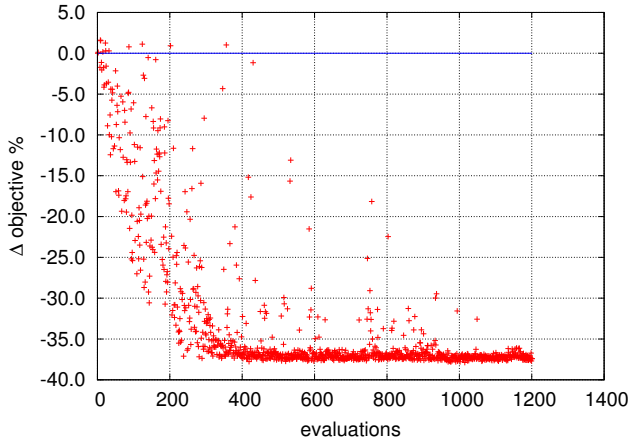


Figure 7-12: Air Benchmark L2: 6 DV Reference Run Evolution History.

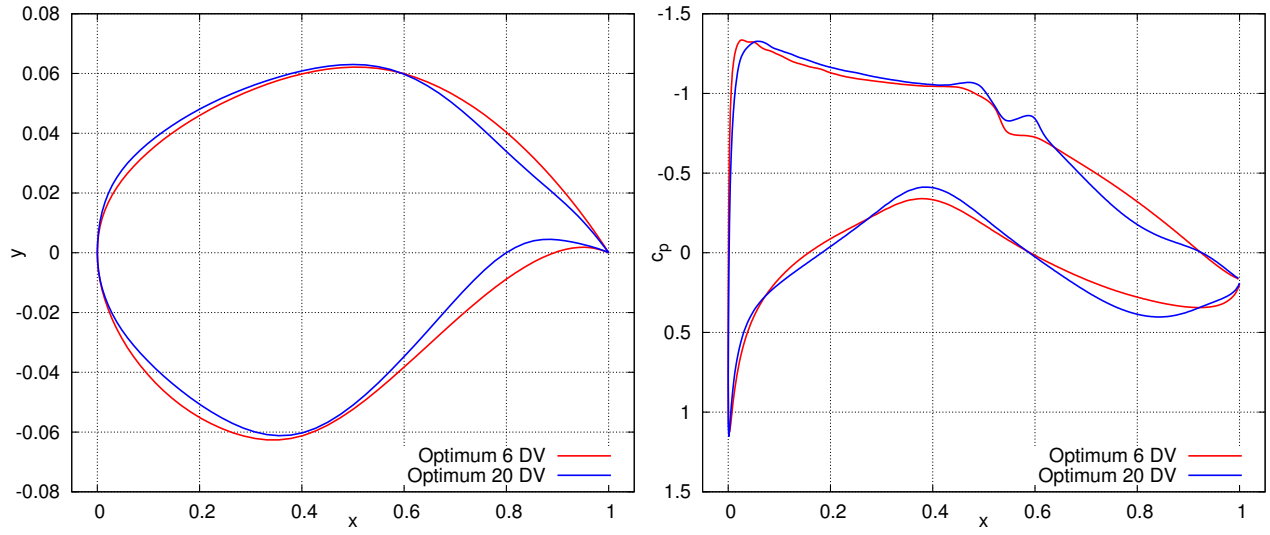


Figure 7-13: Air Benchmark L2: Comparison of Airfoil Shapes and Pressure Coefficients for 20 and 6 Design Variable Reference Optimization Problems.

7.3.3.1 AMOGAe

CIRA has used the AMOGAe algorithm, which does use surrogate models. The use of the ELEMENT_SCALE grid coarseness parameter equal to 12.5 characterizes low-fidelity runs. This has made it possible to obtain very coarse grids, which are often unsuitable for runs with an assigned lift coefficient. Consequently, it was decided to use the efficiency $E = c_l/c_d$ as the second objective (to be maximized), with a penalty if the resulting airfoil has a $c_l < 0.7$ at the prescribed angle of attack. The angle of attack chosen to calculate the second objective is that of the element selected for the evolutionary strategy $(1+1)$-ES which is used as the internal algorithm. This angle is known because it is calculated during the evaluation of the algorithm external cycle and is kept constant during the internal low-cost asymmetric optimization step.

The chosen ELEMENT_SCALE, together with the fact that the low-fi run ran at a prescribed angle of attack and not at assigned c_l, led to a ratio between low-fi and hi-fi running times equal to 0.04. Hi-fi runs were assigned 16 cores, and low-fi runs were assigned 8.

The optimization run with 20 design variables had a population of 4 elements in the outer genetic multi-objective loop that evolved over 40 generations. The bit-mutation operator had a mutation probability of 0.65%. The internal $(1+1)$-ES algorithm has a 50% activation probability and a 0.6% bit-mutation rate and is iterated eight times.

The optimization run required 238 hi-fi and 826 low-fi objective function evaluations.

The optimization run for the six-variable case has the same settings, with the only difference that the evolution process was stopped after the twentieth generation, for a total of 104 hi-fi and 412 low-fi evaluations.

7.3.3.2 GP-PLS-EI

ITU used their Auto-regressive Gaussian process (GP-PLS-EI) to run the 20 DV and the 6 DV problems.

The average computation time ratio between low and high-fidelity analyses was 0.012 with 28 CPU cores used both for low and high-fidelity runs.

High-fidelity analyses were performed at fixed c_l mode, with an element scale value of 2.5. Low-fidelity analyses were performed at a fixed angle of attack with an element scale value of 12.5.

7.3.3.3 GP-PI, GP-MVES, GP-EI

POLITO ran three different algorithms for the six and twenty design variables benchmarks, namely GP-PI, GP-MVES, GP-EI. The average computation time ratio between low and high-fidelity analyses was 0.2 because both high-fidelity and low-fidelity analyses were performed at fixed c_l mode. The high-fidelity element scale value was set to 2.5 and the low-fidelity one at 12. The number of HF/LF computations reported is the average of over ten experiments for each method/DV case. Instead, the c_d values reported in the table refer to the minimum values among all the experiments. For comparison with other implementations, it is essential to specify that the design space is defined as $\mathbf{x} = [-1.0, 1.0]^{\mathrm{NDV}}$.

7.3.3.4 L2 Air Benchmark Summary and Discussion

id	maker	algorithm	hi-fi eval.	lo-fi eval	total cost	c_d	$(c_d - c_{d_{\mathrm{ref}}})\%$	violated constr.
RAE2822 baseline	CIRA	n/a	1	0	1	0.0204746	66.4	no
20 DV — reference	CIRA	CMA-ES	2369	0	2369	0.0123039	0.0	no
20 DV — multi-fi	CIRA	AMOGAe	238	826	271.0	0.0128218	4.2	no
20 DV — multi-fi	ITU	GP-PLS-EI	250	550	256.6	0.0136646	11.1	no
20 DV — multi-fi	POLITO	GP-PI	134.3	828.7	300.0	0.0132743	7.9	no
20 DV — multi-fi	POLITO	GP-MVES	99.4	1003.6	300.1	0.0137308	11.6	no
20 DV — multi-fi	POLITO	GP-EI	87.5	1064.4	300.4	0.0142812	16.1	no
6 DV — reference	CIRA	CMA-ES	1201	0	1201	0.0127457	0.0	no
6 DV — multi-fi	CIRA	AMOGAe	104	412	120.5	0.0186789	46.6	no
6 DV — multi-fi	ITU	GP-PLS-EI	150	225	152.7	0.0183975	44.3	no
6 DV — multi-fi	POLITO	GP-PI	76.2	620.8	200.3	0.0187592	47.2	no
6 DV — multi-fi	POLITO	GP-MVES	54.3	730.1	200.3	0.0182511	43.2	no
6 DV — multi-fi	POLITO	GP-EI	58.5	709.3	200.4	0.0187482	47.1	no

Table 7-7: Summarizing and Comparative Table of the Results of the Multi-fidelity Optimizers On the L2 Air Benchmark.

Table 7-7 summarizes the results obtained from the multi-fidelity algorithms compared to the high single-fidelity reference ones. The first column indicates the type of result, i.e. whether it relates to the baseline (RAE2822 airfoil) or is a reference hi-fi run or a run of a multi-fidelity algorithm and how many design variables there

are. In particular, the "20 DV — reference" label indicates the hi-fi run with the maximum number of variables allowed in the L2 air benchmark that is conventionally adopted as the absolute optimum. The second column indicates the partner that produced the results. The third one is a label that identifies the algorithm used, which will also be used in Figure 7-14 to visually illustrate the differences in performance.

The fourth column shows the number of high-fidelity evaluations performed by the algorithm, the fifth column indicates the number of low-fidelity evaluations, and the sixth is the total computational cost of the aerodynamic analysis runs. The cost is obtained with the formula:

$$\text{total cost} = \text{number of hi-fi evaluations} + \text{ratio} \times \text{number of lo-fi evaluations} \tag{7-1}$$

where “ratio” represents the ratio between the elapsed time required for the low-fi and hi-fi run. Note that this time may vary according to the different approximation levels chosen by the partners (free parameter of the benchmark) and the number of cores used (for computational efficiency reasons in parallel, it may be more efficient to use fewer cores in the low-fi run).

The seventh column shows the value of the optimal c_d, and the eighth is the difference in percentage terms between the optimal c_d obtained by the multi-fidelity algorithm compared to the reference one obtained with the same number of design variables. Finally, the last column indicates whether the reported solution violates the problem constraints.

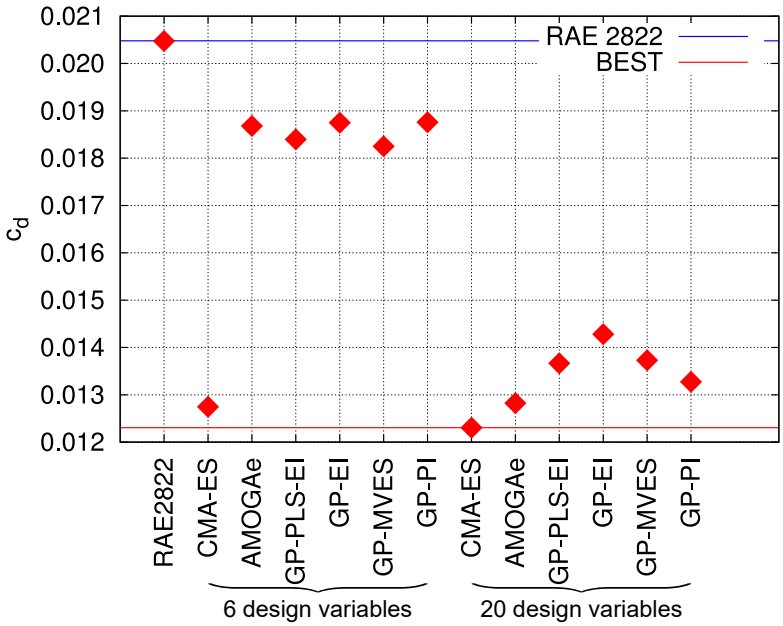


Figure 7-14: Air Benchmark L2: Summary of L2 Benchmark Results. Note that CMA-ES labels are related to the 6 DV and 20 DV hi-fi (single fidelity, no-metamodels) reference runs.

Table 7-7 and Figure 7-14 highlight an excellent behavior of the multi-fidelity algorithms in the 20-variable benchmark, where, against a computational cost ranging from 23 to 45 per cent of the high-fidelity run, an optimal value of the function was obtained that is only about 4 to 16 per cent worse than what the hi-fi run achieves.

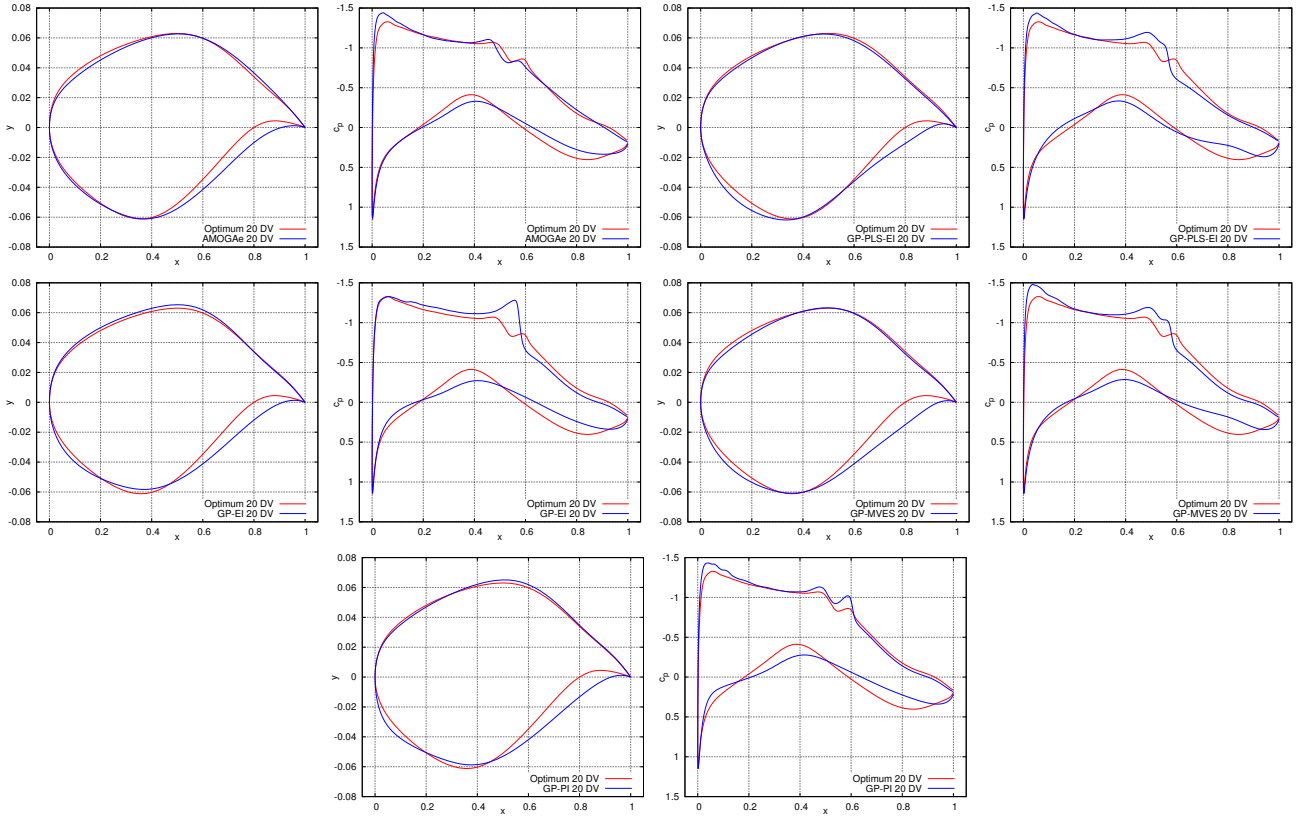


Figure 7-15: L2 Air Results Obtained for the Case with 20 Design Variables.

However, two critical features of the test case and setup should be highlighted: the first is that, as shown in Figure 7-9, which shows the evolution history of the reference 20DV run, the improvement in the objective function is under 5% from the "absolute best" it reaches well before the 1000 hi-fi evaluations; the second is that many of the algorithms used here in the benchmarks have a stochastic basis and therefore would require statistical analysis to be correctly evaluated which, at the moment, has not been possible due to the high computational resources required.

On the other hand, the case with six design variables presents some unexpected criticalities as all the multi-fidelity methods fail to obtain results less than 40% distant from the solution found by the hi-fi algorithm. This is even more surprising if we observe, as evident from Figure 7-12, that the algorithm that uses only hi-fi simulations manages to get close to the optimum after just about 300 evaluations. A possible explanation for this behavior could be that the low-fidelity optimum is distant from the high-fidelity optimum and that this low-fidelity optimum acts as a misleading attractor for multi-fidelity algorithms. However, this aspect requires further study and a complete multi-objective analysis of the problem, which will be the object of further investigation in future work.

Finally, for completeness, we report the airfoil shapes and pressure coefficients obtained using the multi-fidelity optimization algorithms compared to the reference obtained using only high-fidelity evaluations. Figure 7-15 highlights the goodness of the results obtained for the case with 20 design variables. In contrast, the observation

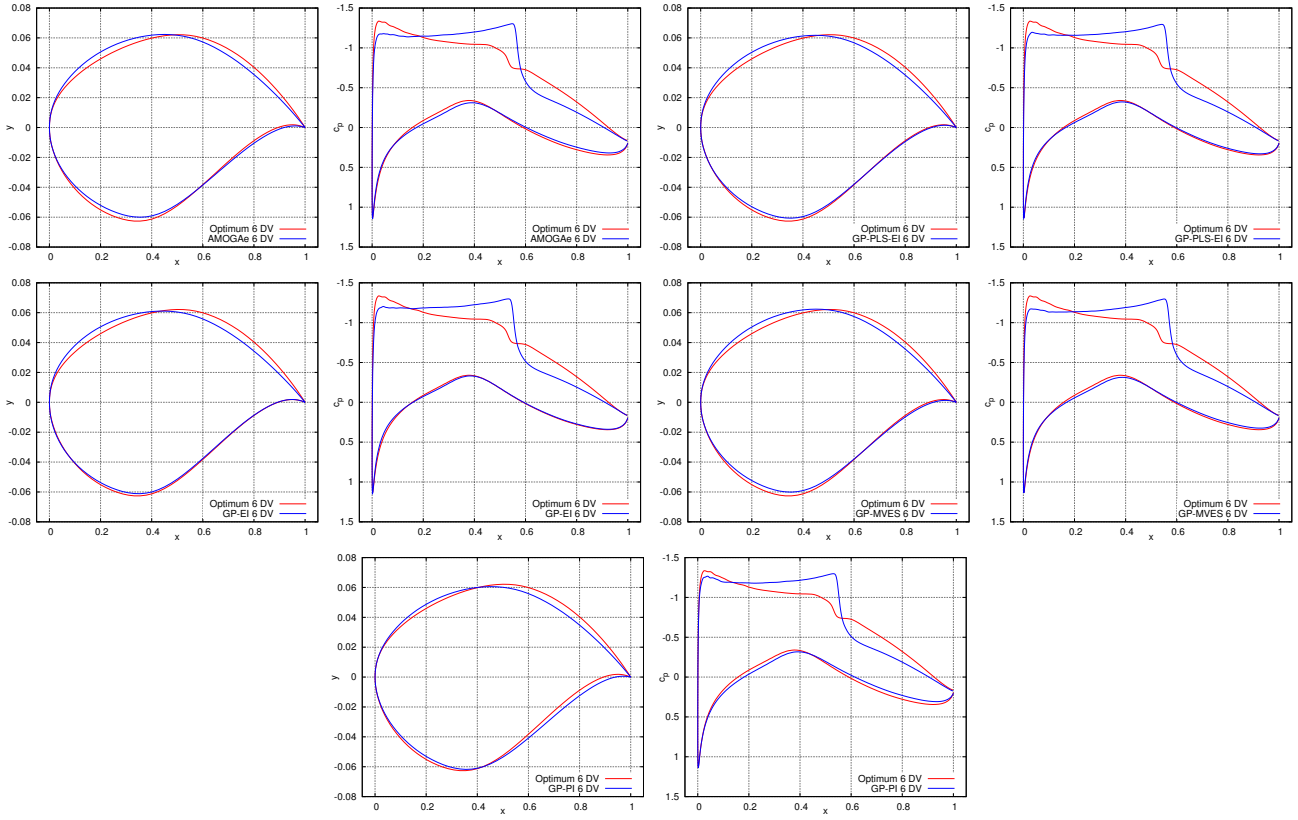


Figure 7-16: L2 Air Results Obtained for the Case with 6 Design Variables.

of Figure 7-16 shows how all the algorithms converge towards a largely sub-optimal solution in the case with six variables.

7.4 CONCLUSIONS

The chapter presented the outcomes of applying multi-fidelity design optimization methods to L3 and L2 air benchmark problems. The L3 applications were a family of benchmark problems on aero-structural design of a wing-fuselage-tail transport configuration for minimum drag and fuel burn. Different combinations of airfoil shape, wing planform, and structural sizing variables were considered in combinations up to 321 design parameters for both rigid and aeroelastic cases using linear aerodynamics, Euler CFD, and coupling with linear structures. The L2 application was a drag minimization starting from the RAE-2822 airfoil at transonic conditions. Multiple classes of shape modification functions were considered up to 20 design parameters. The high-fidelity analysis was Reynolds-averaged Navier-Stokes solved for fixed lift coefficient on a relatively fine mesh. The lower-fidelity analyses utilized a coarser grid solved either for fixed lift coefficient or fixed angle of attack.

The design approaches applied to the L3 problems included three surrogate-based, global optimization methods and one gradient-based optimization method. Each method was implemented both in multi- and single-fidelity (highest-fidelity analysis). The multi-fidelity methods included auto-regressive Gaussian process with partial

least squares dimension reduction, kriging with sparse polynomial chaos trend function, and regression co-kriging. The gradient-based method was a BFGS-based approach. All the methods utilized a quadratic penalty function to enforce the design constraints. Initial benchmark comparisons highlighted the critical importance of the size and distribution of the initial training data; subsequent studies reported statistics of multiple optimizations initialized from shared designs of experiments.

The design approaches applied to the L2 problems included a multi-fidelity genetic algorithm, an auto-regressive Gaussian process with partial least squares dimension reduction, and a different auto-regressive Gaussian process implemented with varying data acquisition functions. Single-fidelity approaches using the highest fidelity level included a baseline evolutionary strategy as well as the Gaussian process with partial least squares. Similar to the L3 problems, the methods implement a quadratic penalty function to enforce the design constraints, with the exception of the design lift coefficient and airfoil thickness constraints, which are enforced implicitly by the analysis and geometry generation, respectively.

For the L3 problems, all the approaches achieved similar levels of penalty function reduction in the limit of computational budget, with the exception of multi-fidelity SPCE-KG, which was slow to converge. With the same exception, the Gaussian process/kriging-based methods also converged at comparable rates on average, starting from identical designs of experiments. For the three-dimensional cases, the gradient-based MF-BFGS had a comparable convergence rate as the surrogate-based methods, as well as comparable penalty function reduction, on average, for the same budget as the initial surrogate DoE. For the six-dimensional cases, however, the MF-BFGS method achieved significantly larger penalty function reduction than the surrogate-based methods at the initial DoE cost. This finding corroborated the relative scalability of the gradient-based approach versus the surrogate-based methods. The envelope of minimal and maximal best designs plotted versus computational cost indicated that the methods were not trapped by any potential local optima (or at least that any local minimal had similar performance). The relatively large envelopes on the BFGS methods (HF and MF) compared to the surrogate based methods highlighted their strong dependence in initial point selection.

For the L2 problems, the multi-fidelity methods achieved similar drag coefficients when compared to each other, but did not achieve the levels of reduction from the reference optimizations. For the twenty parameter cases, the drag coefficients were within 20% of the reference value, with reduced computational cost. However, the six parameter cases exceeded the best drag by more than 40%. This discrepancy in the latter case is an area for future investigation, but could be indicative that the low-fidelity optimum is far from that of the high-fidelity response.

The outcomes of this effort underscored several areas for future research. First are techniques for scaling multi-fidelity methods to higher-dimensional problems. In this study, the partial least squares approach to dimension reduction successfully enabled problems up to 20 design variables, and a gradient-based approach enabled optimization up to 321 design variables. The surrogate-based methods with partial least squares are capable of global optimization, but the dimension reduction only slows the curse of dimensionality. The dimension of the reduced subspace will continue to grow, requiring large sets of training data to explore the space (though the degree of dimension reduction depends on the behavior of the problem). In contrast, the gradient-based approaches are local optimization methods, and are more scalable to larger problems. However, many disciplinary analyses do not provide efficient computation of derivatives, and design including many disciplines often requires resorting to finite differences or approximate models. Further, the presence of multiple sources of numerical noise in the L3 problems proved challenging to achieving deep levels of optimization convergence. A second area for future research is the handling of design constraints in multi-fidelity

optimization. All the methods implemented utilized a penalty function to enforce the design constraints. However, discrepancies in the constraint boundaries between the high- and low-fidelity model could lead to poor model correlation and slow convergence. By handling constraints separate from the objective function, the discrepancies between fidelity levels may be corrected individually and more accurately.

ACKNOWLEDGMENTS

The first and second authors acknowledge support of the US Air Force Office of Scientific Research (Dr. Fariba Fahroo, Computational Mathematics Program Officer). Distribution Statement A: Approved for Public Release; Distribution is Unlimited. AFRL-2023-1834.

REFERENCES

[1] D. Bryson, P. S. Beran, A. Thelen, M. Nikbay, E. Cakmak, and S. Yildiz, "Comparison of multi-fidelity optimization methods using an aero-structural benchmark problem," in *AVT-354 Research Workshop on Multi-fidelity methods for military vehicle design, Varna, Bulgaria*, 2022.

[2] A. Thelen, D. Bryson, B. Stanford, and P. Beran, "Aeroelastic optimization studies of an aircraft jig shape using gradient-based multi-fidelity methods," in *International Forum on Aeroelasticity and Structural Dynamics (IFASD 2022)*, 2022.

[3] A. S. Thelen, D. E. Bryson, B. K. Stanford, and P. S. Beran, "Multi-fidelity gradient-based optimization for high-dimensional aeroelastic configurations," *Algorithms*, vol. 15, no. 4, p. 131, 2022.

[4] D. Quagliarella, P. S. Beran, D. Bryson, D. Clark, A. Thelen, L. Mainini, S. Yildiz, M. Nikbay, E. Minisci, P. Leyland, A. Serani, and M. Diez, "Reproducible industrial multifidelity optimization benchmark problems for air, space, and sea vehicles," in *AVT-354 Research Workshop on Multi-fidelity methods for military vehicle design, Varna, Bulgaria*, 2022.

[5] D. Quagliarella and M. Diez, "An open-source aerodynamic framework for benchmarking multi-fidelity methods," in *AIAA AVIATION 2020 FORUM*, 2020, p. 3179.

[6] M. P. Rumpfkeil, D. E. Bryson, and P. S. Beran, "Multi-fidelity sparse polynomial chaos and kriging surrogate models applied to analytical benchmark problems," *Algorithms*, vol. 15, no. 3, 2022, ISSN: 1999-4893.

[7] D. E. Bryson and M. P. Rumpfkeil, "Aerostructural design optimization using a multifidelity quasi-newton method," *Journal of Aircraft*, vol. 56, no. 5, pp. 2019–2031, 2019.

[8] S. Yildiz, H. Pehlivan-Solak, M. Diez, O. Goren, and M. Nikbay, "Advanced experiments on gaussian process-based multi-fidelity methods over diverse mathematical characteristics," in *8th European Congress on Computational Methods in Applied Sciences and Engineering*, AIAA Paper 2004-6514, ECCOMAS, Oslo, Norway, Jun. 2022.

[9] D. Quagliarella, P. Iannelli, P. L. Vitagliano, and G. Chinnici, "Aerodynamic shape design using hybrid evolutionary computation and fitness approximation," in *AIAA 1st Intelligent Systems Technical Conference*, AIAA Paper 2004-6514, American Institute of Aeronautics and Astronautics (AIAA), Chicago, IL, Sep. 2004. DOI: 10.2514/6.2004-6514.

[10] X. Ruan, P. Jiang, Q. Zhou, J. Hu, and L. Shu, "Variable-fidelity probability of improvement method for efficient global optimization of expensive black-box problems," *Structural and Multidisciplinary Optimization*, vol. 62, no. 6, pp. 3021–3052, 2020.

[11] D. Huang, T. T. Allen, W. I. Notz, and R. A. Miller, "Sequential kriging optimization using multiple-fidelity evaluations," *Structural and Multidisciplinary Optimization*, vol. 32, no. 5, pp. 369–382, 2006.

[12] S. Takeno, H. Fukuoka, Y. Tsukada, T. Koyama, M. Shiga, I. Takeuchi, and M. Karasuyama, "Multi-fidelity bayesian optimization with max-value entropy search and its parallelization," in *International Conference on Machine Learning*, PMLR, 2020, pp. 9334–9345.

[13] T. D. Economon, F. Palacios, S. R. Copeland, T. W. Lukaczyk, and J. J. Alonso, "SU2: An open-source suite for multiphysics simulation and design," *AIAA Journal*, vol. 54, no. 3, pp. 828–846, 2016.

[14] N. Hansen and A. Ostermeier, "Completely derandomized self-adaptation in evolution strategies," *Evolutionary computation*, vol. 9, no. 2, pp. 159–195, 2001.